

The advent of personal genome sequencing

Radoje Drmanac, PhD

Abstract: Rapid technological advances are decreasing DNA sequencing costs and making it practical to undertake complete human genome sequencing on a large scale for the first time. Disease studies that involve sequencing hundreds of patient genomes are underway. The all-inclusive sequencing price per genome is expected to reach \$1000 over the next few years and will likely decline further in the following years. This dramatic price decline will herald widespread personal genome sequencing and lead to significant improvements in human health and reduced health care costs. Key to realizing these benefits will be medical genomics' and systems biology's success in providing increasing contextual interpretation of biological and medical effects of the detected sequence variants in a genome. Given the substantial potential benefits and the manageability of the health and discrimination risks involved with the possible misuse of this information, we propose that governments and insurance companies support or even require personal genome sequencing. Critical to the widespread acceptance of personal genome sequencing, however, will be the need to educate physicians and the public about the realistic benefits and risks of such an analysis to prevent overinterpretation and misuse of this valuable information. *Genet Med* 2011;13(3):188–190.

Key Words: whole genome sequencing, personalized medicine, personal genomics, personal genome sequencing, DNA sequencing, genomic medicine

THE INCREASING AVAILABILITY AND EVER-DECREASING COST OF PERSONAL GENOME SEQUENCING

Complete human genome sequencing is becoming available at increasing scale and decreasing cost, thanks to massively parallel genomic micro- and nanoarrays (Ref. 1 and references therein). In 2010, multiple studies based on sequencing dozens to hundreds of complete human genomes were completed or initiated. With the more than 400 genomes/month sequencing capacity available at Complete Genomics, combined with the expanding capacity of National Human Genome Research Institute-funded US genome centers, the Wellcome-Trust Sanger Institute in Europe, and BGI in China, thousands of individual genome sequences are expected to be analyzed this year. The number of genomes sequenced has grown dramatically over the last few years from <100 in 2009 to >2000 in 2010 and is projected by the journal *Nature* to reach approximately 25,000 this year, including low-coverage genomes. It would not be surprising if within the next 5 years, we see the annual number of complete human genomes sequenced rise to over a million. Obtaining complete genetic and epigenetic information at this scale, coupled with routine transcriptome sequencing and var-

ious functional studies, will lead to an increasingly comprehensive understanding of disease development at the molecular level.^{2–5}

These DNA sequencing advances are making large-scale personal genome sequencing (PGS) a rapidly approaching reality.⁶ For example, using Complete Genomics' service based on its novel nanoarray technology,⁷ the price for sequencing and analyzing a complete human genome is now routinely below \$10,000 for 40× coverage and 99.999% accuracy. Complete Genomics will continue to increase its sequencing efficiency by further miniaturization to incorporate more DNA spots/mm² on each nanoarray, the use of faster imaging cameras, brighter dyes, haplotyping, and other improvements. Similarly, several sequencing instrument companies continue to improve efficiency and reduce assay cost of existing platforms or develop radically novel technologies such as single-molecule sequencing for targeted diagnostic applications.⁸

Experts predict that the consumer price to sequence a complete human genome will drop to \$1000 in 2014.⁹ In our opinion, this will be achieved with existing DNA nanoarray technologies. We further believe that the existing DNA nanoarray technologies, with expected engineering advances, are capable of driving the cost per genome to significantly below \$1000 in the following years. By 2020, with improved technology and reduced cost, we may expect tens of millions of personal genomes to be sequenced worldwide. It is important that society at large start preparing for this rapidly approaching genomic tsunami.

UNDERSTANDING THE HUMAN GENOME: DERIVING THE BIOLOGICAL CONTEXT OF GENETIC VARIANTS

Although large-scale genome sequencing is an exciting proposition, the need to convert the resulting data into actionable reports remains a daunting challenge. The genetic programs governing human development, adaptive functioning, and maintenance consist of complex regulatory and signal processing networks and pathways. Any segment of the genome sequence derives context from the other sequences in that genome (including parental sequences), environmental conditions, and stochastic events such as somatic mutations. These complex interactions explain why isolated genetic variants (e.g., single-nucleotide polymorphisms) that are statistically associated with various diseases most often show only incomplete penetrance. To understand the intricate regulatory networks and the biological context of any given variant, it will be necessary to obtain complete and accurate genomic, transcriptomic, and epigenetic sequence data from thousands of individuals: patients, family members, and healthy controls. Large-scale studies are underway to generate data of this type, such as those carried out by the 1000 Genomes Project,¹⁰ the Cancer Genome Atlas,^{11,12} and the National Institutes of Health Roadmap Epigenomics Mapping Consortium.¹³ By using a whole genome sequencing service, a number of institutions have already initiated several-hundred-sample genome projects.^{14,15} Similarly, understanding regulatory networks will also require additional systems ap-

From Complete Genomics, Inc., Mountain View, California.

Radoje Drmanac, PhD, Complete Genomics, Inc., 2071 Stierlin Court, Mountain View, CA 94043. E-mail: rdrmanac@completegenomics.com.

Disclosure: The author is cofounder, shareholder, and Chief Scientific Officer of Complete Genomics, Inc., a provider of complete human genome sequencing services to the research community.

Published online ahead of print February 9, 2011.

DOI: 10.1097/GIM.0b013e31820f16e6

proaches (e.g., studies of dynamic elements such as proteins and metabolites), many focused studies, collection of phenotypic data, and an enormous amount of computer modeling.

It is critical to achieve all four of these data measures: completeness, accuracy, volume, and diversity. For example, despite its utility, exon sequencing does not provide complete insight. For a molecular understanding and improved prevention and treatment of thousands of diverse diseases and conditions, it is critical that reliable and affordable services in genomic data generation and processing are available to the broad scientific and medical communities. The storage of these data in a digital format, preferably as part of an individual's electronic medical record, is essential to enable manageability and continuous analysis. We expect that advances in electronics will allow permanent lifelong storage of personal genetic variants (1 GB/person) for less than \$10.

MEDICAL BENEFITS AND REMAINING CHALLENGES

Large-scale PGS coupled with a proper interpretation of the results will permit a deep understanding of disease mechanisms, allowing for more rational interventions.¹⁶ This has already occurred to some extent using targeted gene studies. For example, companion diagnostics targeting relevant biomarkers have been approved by the US Food and Drug Administration for Gleevec® in the treatment of gastrointestinal stromal tumors; Erbitux® for metastatic colorectal cancer; and Herceptin® to treat metastatic breast cancer.¹⁷ PGS is also likely to permit increased identification of at-risk patients, such as those with mutation in a tumor suppressor gene, who should be monitored more frequently for disease development.

PGS information may also improve the drug development process by identifying genetically predisposed nonresponders and individuals who are at greater risk of experiencing a side effect from the treatment before they enter a clinical trial. By excluding those subjects, trial sponsors can greatly increase the likelihood that the study will be successful and achieve its endpoints. This genetic information will ultimately make drugs safer and more effective because they will be targeted at patients who are more likely to benefit from them and will be contraindicated for people more likely to develop adverse events.

For patients with cancer, PGS of hundreds of patients from each tumor type will lead to a detailed understanding of the diverse molecular processes in cancer development and metastasis^{5,18–20} and will enable the development of improved tumor diagnosis and the classification and selection of more effective treatments based on complete genome and transcriptome sequencing of each biopsy. This improved understanding of the disease pathways involved may also allow existing drugs to be repurposed for other indications. We have to be mindful of the complexities of these developments and the amount of time required to complete proper clinical studies before these advances are adopted as routine medical practice.

Similarly, PGS is expected to help diagnose, better understand, and select optimal treatment for children and other patients with undefined diseases.²¹ We believe that the initiation of a national project enabling immediate DNA sequencing and interpretation of the whole genomes of these affected children and their parents could be of great utility as one of the primary diagnostic procedures for these patients.

Finally, PGS can serve as a universal genetic test, carried out once, and used for life. PGS would combine tests for rare and metabolic diseases—such as predisposition to cancer and vari-

ous late-onset diseases, drug response and adverse reactions, carriers of recessive mutations, and human leukocyte antigen typing for immunological compatibility, to mention a few of the known biomarkers.

The limited understanding of the genome today does not mean that PGS should not be used today. There are at least 3000 genes for which interpretative information would be immediately useful. Furthermore, personal genome variants may be repeatedly reanalyzed in the light of new genomic and functional knowledge. For most people, no serious disease-causing genetic variants will be detected, even in advanced analysis. This is to be expected for any presymptomatic risk reduction test. It is important to treat this as a positive outcome, because it will still allow disease prevention recommendations to be made and better treatments prescribed for potentially millions of people. Furthermore, everyone tested could be provided with a report detailing dozens of drugs that would not work for them and a few that may cause adverse reactions.²² These reports, stored as part of an individual's electronic medical record, will allow medical professionals to select optimal personalized treatments for their patients.

However, with these benefits come certain risks. The most serious perhaps is the potential overinterpretation of results based on a limited understanding of contextual information. For example, a risk that is estimated at 1.2 times normal should probably not be reported. It could spur unnecessary medical actions and cause unwarranted psychological distress. Validated genome interpretation software using conservative reporting standards is a potential solution. To further minimize this risk, physician and patient education programs need to be introduced, so that the genotypic data are understood within a broader biological and statistical context—for example, personal medical history, family history, and other behavioral or molecular phenotypic data.

There is also the risk of genetic discrimination, which has just begun to be addressed by the Genetic Information Nondiscrimination Act. The implementation of this law and other supporting nondiscriminatory policies needs to be continued and reinforced.

COST BENEFITS

PGS as the single universal genetic test is a cost-effective solution that subsumes hundreds of individual tests and will facilitate some analyses, which are currently not performed due to high cost. For example, most cystic fibrosis tests do not cover approximately 20% of cases caused by less frequent mutations. Similarly, with the exception of *BRCA* genes, no other tumor suppressor genes are routinely sequenced to enhance tumor prevention. Furthermore, a recent study has shown that knowing the sequence of several genes implicated in mental disorders that frequently harbor de novo mutations would be highly beneficial.²³

Current annual health costs in the United States are estimated at \$2.6 trillion.²⁴ The cost of sequencing 26 million genomes per year (e.g., newborns, adults, and cancer biopsies) at \$1000 per genome would be only 1% of that amount at \$26 billion. Because most of the health care cost is generated by a small fraction of population,²⁵ we estimate that the benefits of that sequencing could reduce health care costs by at least 10% or \$260 billion. That would represent a potential annual savings of more than \$234 billion while potentially enabling better, more personalized health care. Because of the expected health benefits and large cost savings, we suggest that health insurance companies offer discounted premiums for people having their personal genome sequenced.

CONCLUSION

PGS is being enabled by unprecedented advances in complete genome sequencing technology. Medical genomics software advances are also occurring rapidly, driven by the need to interpret the influx of data from thousands of genome sequences. Together these advancements will enable a wider use of PGS in medical practice starting in a few years. It is our opinion that any revealed risk will be manageable through education, appropriate policies, and conservative data reporting standards. On the other hand, it is unlikely that the current increase in US health care costs will be sustainable even in the foreseeable future. These circumstances may work together to motivate decision makers and payers to adopt new methods of preventive and predictive personalized medicine based on complete genetic knowledge. We are witnessing the exciting and promising beginnings of genomic medicine.

ACKNOWLEDGMENTS

The author acknowledges and expresses appreciation to Ruth Mercado for her assistance in the preparation of this manuscript.

REFERENCES

- Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;327:78–81.
- Roach J, Glusman G, Smit AF, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010;328:636–639.
- Lupski J, Reid J, Gonzaga-Jauregui C, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010;362:1181–1191.
- Baranzini SE, Mudge J, van Velkinburgh JC, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 2010;464:1351–1356.
- Lee W, Jiang Z, Liu J, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 2010;465:473–477.
- Snyder M, Du J, Gerstein M. Personal genome sequencing: current approaches and challenges. *Genes Dev* 2010;24:423–431.
- Complete Genomics. Available at: <http://www.completegenomics.com/>. Accessed December 20, 2010.
- Pennisi E. Semiconductors inspire new sequencing technologies. *Science* 2010;327:1190.
- Metzger ML. Sequencing technologies—the next generation. *Nature Rev Genet* 2010;11:31–46.
- The 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altshuler DL, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–1073.
- Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 2009;458:719–724.
- Ledford H. The cancer genome challenge. *Nature* 2010;464:972–974.
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotech* 2010;28:1045–1048.
- UCSF Gladstone Institute Press Release, 2010. Available at: http://www.eurekalert.org/pub_releases/2010-06/gi-gai061510.php. Accessed December 20, 2010; and Complete Genomics Press Release, January 11, 2011.
- Complete Genomics Press Release, 2010. Available at: <http://www.medicalnewstoday.com/articles/200392.php>. Accessed December 20, 2010.
- Caskey CT. Using genetic diagnosis to determine individual therapeutic utility. *Annu Rev Med* 2010;61:1–15.
- Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med* 2010;363:301–304.
- Mardis ER. Cancer genomics identifies determinants of tumor biology. *Genome Biol* 2010;11:211.
- International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, et al. International network of cancer genome projects. *Nature* 2010;464:993–998.
- Verhaak RG, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010;17:98–110.
- Mayer AN, Dimmock DP, Arca MJ, et al. A timely arrival for genomic medicine. *Genet Med* 2011;13:195–196.
- Ashley EA, Butte AJ, Wheeler MT, et al. Clinical assessment incorporating a personal genome. *Lancet* 2010;375:1525–1535.
- Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. *Nat Genet* 2010;42:1109–1112.
- Center for Medicare and Medicaid Services, U.S. Department of Health and Human Services, National Health Expenditure Projections, 2010. Available at: http://www.cms.gov/NationalHealthExpendData/03_NationalHealthAccountsProjected.asp#TopOfPage. Accessed December 20, 2010.
- Cohen SB, Yu W. Medical Expenditure Panel Survey, 2010. Available at: http://www.meps.ahrq.gov/mepsweb/data_files/publications/st278/stat278.pdf. Accessed June 30, 2010.