

Accurate Whole Genome Sequencing as the Ultimate Genetic Test

Radoje Drmanac,^{1,2*} Brock A. Peters,^{1,2} George M. Church,³ Clifford A. Reid,¹ and Xun Xu²

Featured Article: Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;327:78–81.⁴

Even 30 years ago, it was obvious that Sanger sequencing had limited throughput, and a more efficient process could replace many tedious gene and genome mapping projects. It would take until the mid-2000s for massively parallel sequencing (MPS)⁵ technologies to demonstrate they could overtake the Sanger sequencing hegemony. Our paper was not the first description of a viable MPS technology, but it firmly established that human whole genome sequencing (WGS) could be done affordably (<US\$5000 in reagent cost), with high accuracy (<1 error in 100 kb), and with high throughput, thus heralding the arrival of personal genome sequencing.

This work was the result of a long effort that started in Serbia in 1987 with a proposal for sequencing by hybridization (SBH) on dot-blot DNA arrays. SBH on random bead microarrays prepared by emulsion PCR, in which a micron-sized bead replaced a millimeter-sized DNA dot, offered a first MPS solution (www.rdrmanac.com). The US Department of Energy and NIST grants funded most of these initial efforts. After the events of September 11, 2001, the NIH received an increase in biodefense funding, and our US\$2.3 million grant for ligation-based MPS technology on genomic DNA microarrays was awarded in 2003 by the National Institute of Allergy and Infectious Diseases. These grants and eventually >US\$250 million of private and public equity funding resulted in the formation (in 2005) and growth of Complete Genomics, Inc. (CGI), a Silicon Valley company with the audacious goal of sequencing millions of whole human genomes.

Before all these genomes could be sequenced, an efficient technology for arraying and imaging billions of clon-

ally amplified DNA fragments had to be developed. Rolling circle replication in solution was developed to form billions of discrete DNA nanoballs (DNBs), each comprising a concatemer with approximately 500 copies of the original circle. These DNBs were flooded, under optimized conditions, onto silicon chips patterned with densely packed sticky spots resulting in the first DNA nanoarrays for MPS; >90% of 1 billion spots were occupied by 1 DNB without overcrowding the array. These simple-to-prepare patterned DNA arrays provided low-volume (low-cost) sequencing reactions and efficient imaging by use of fewer camera pixels per spot. Building on our prior work, we further developed a low-cost ligation-based sequencing method that did not require engineered enzymes or cleaving labeled probes. A sophisticated mate-pair library with 4 synthetic adapters inserted into each circular DNB template allowed us to read up to 70 discontinuous bases per DNB. Analyzing the data required development of special statistical algorithms and software, including local de novo sequence assembly, to call all types of variants, including single-nucleotide variants, indels, structural variants, and copy number variants. The combination of these advanced algorithms, affordable high read coverage, and inherently unbiased sequencing chemistry resulted in the first demonstration of human WGS that had an error rate <1 in 100 kb. This established that short mate-pair reads could provide high-quality human WGS. We further refined this technology to take advantage of patterned DNB arrays by creating nanoarrays with 3–4 DNBs per square micrometer and achieving the ultimate imaging efficiency of a single pixel per DNA spot (1). These improvements allowed CGI to establish a human WGS service capable of analyzing >1000 samples per month at 50× read coverage.

Twenty thousand human genomes later, the payoffs from our MPS technology are becoming clear through hundreds of studies demonstrating the power of high quality WGS and advancing the understanding of the genetics behind many diseases. Some of the early success stories with our technology include the first demonstration that WGS in a nuclear family could determine the genetic underpinnings of mendelian disorders (2) and furthering the understanding of what causes cancer (3, 4). The most recent landmark study that used CGI WGS technology demonstrated that causative de novo mutations can be detected in >60% of children with intellectual disabilities (5). Our WGS service has also participated in a number of population studies such as

¹ Complete Genomics, Inc., Mountain View, CA; ² BGI-Shenzhen, Shenzhen, China; ³ Department of Genetics, Harvard Medical School, Boston, MA.

* Address correspondence to this author at: 2071 Stierlin Ct., Mountain View, CA 94043. E-mail rdrmanac@completegenomics.com.

Received October 24, 2014; accepted October 28, 2014.

Previously published online at DOI: 10.1373/clinchem.2014.224907

© 2014 American Association for Clinical Chemistry

⁴ This article has been cited more than 530 times since publication.

⁵ Nonstandard abbreviations: MPS, massively parallel sequencing; WGS, whole genome sequencing; SBH, sequencing by hybridization; CGI, Complete Genomics, Inc.; DNB, DNA nanoball.

1000 Genomes, the Welllderly Study, and the Personal Genome Project (www.personalgenomes.org), as well as sequencing and making freely available the whole genomes of 69 commonly studied cell lines (www.completegenomics.com/public-data/69-Genomes/).

In addition, our efficient WGS has enabled long fragment read technology, a process for highly accurate WGS (1 error in 10 Mb) and haplotyping from 10 cells (1). This process makes possible the use of individual WGS as the ultimate genetic test (6), as well as accurate WGS from samples in which only a small number of cells are available (e.g., microbiopsies). In molecular diagnostics, our MPS is the first next generation sequencing test in the world approved by a government agency. BGI, a genomics company that recently acquired CGI, registered a noninvasive prenatal test with China's Food and Drug Administration on the basis of a modified CGI sequencing platform.

This is not the end of the story for MPS and patterned DNA arrays. Advances in imaging, 2-color base labeling, and other improvements in sequencing chemistry, combined with our patterned DNB nanoarrays and large-scale manufacturing through BGI, have the potential to achieve dramatic cost reduction toward \$1 per gigabase of raw reads in the next few years. This level of efficiency, coupled with advanced computing, will enable, on a global scale, the highest-quality individual WGS tests as well as cancer WGS, transcriptome, metagenome, and other omics analyses performed as routine checkups. Ultimately, these widespread analyses will be used for both understanding the genetic basis of disease and implementing personalized disease prevention and treatment through genomic medicine.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 require-*

ments: (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 or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: R. Drmanac, Complete Genomics, Inc.; B.A. Peters, Complete Genomics, Inc.; C.A. Reid, Complete Genomics, Inc.

Consultant or Advisory Role: G.M. Church, CGI Scientific Advisory Board.

Stock Ownership: R. Drmanac, BGI; B.A. Peters, BGI (the parent company of Complete Genomics, Inc.); C.A. Reid, BGI; X. Xu, BGI.

Honoraria: None declared.

Research Funding: None declared.

Expert Testimony: None declared.

Patents: Rdrmanac, patent nos. 8785127 and 8763375; B.A. Peters, patent no. 8592150 B2.

Other Remuneration: G.M. Church, see <http://arep.med.harvard.edu/gmc/tech.html>.

References

1. Peters BA, Kermani BG, Sparks AB, Alferov O, Hong P, Alexeev A, et al. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature* 2012;487:190-5.
2. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010;328:636-9.
3. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 2010;465:473-7.
4. Molenaar JJ, Koster J, Zwijsen DA, van Sluis P, Valentijn LJ, van der Ploeg I, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neurogenesis genes. *Nature* 2012;483:589-93.
5. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014;511:344-7.
6. Drmanac R. The ultimate genetic test. *Science* 2012;336:1110-2.