

Next-Generation Sequencing and Its Impact on Pharmacogenetics

Huixiao Hong*

Division of Systems Biology, National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA

“Personalized medicine’ refers to the tailoring of medical treatment to the individual characteristics of each patient. It does not literally mean the creation of drugs or medical devices that are unique to a patient, but rather the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Preventive or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not.” [1].

Personalized medicine will improve health outcomes and patient satisfaction and will be a predominant challenge for the pharmaceutical industry and regulatory science in the current century. Implementing personalized medicine based on individuals’ biological information is far from simple. Pharmacogenetics is the key to transform personalized medicine into clinical practice for diagnosis and prognosis. The use of pharmacogenetics in clinical practice, drug development, regulatory decision making, and post-market surveillance can facilitate the development of safer and more effective medicines and enhance the benefit-risk profile of approved drugs and the implementation of personalized medicine.

Pharmacogenetics research dates back to 1957 when individual variation in the levels of activity of enzymes involved in the metabolism of a number of drugs was observed to be correlated with adverse reactions to those drugs [2]. Thereafter, methods and technology for pharmacogenetics have been advanced greatly. Progress in pharmacogenetics has led to the discovery of some genetic variants that can be used to improve the efficacy and safety of the drugs on the market [3]. Recently, the HapMap project determined genotypes of over 3.1 million common SNPs in human populations [4]. Concurrently, high-throughput SNP genotyping technology advanced to enable simultaneous genotyping of hundreds of thousands of SNPs. These advances made genome-wide association studies (GWAS) a feasible and a promising research field for pharmacogenetics. The published GWAS in pharmacogenetics fall in two categories. The first type of GWAS investigated variation in therapeutic response to drugs such as iloperidone [5], thiazide diuretic [6], and warfarin [7]. The second type of GWAS identified genetic variants associated with adverse events of drugs such as the elevation of serum alanine aminotransferase by ximelagatran [8] and drug-induced liver injury due to flucloxacillin [9]. However, GWAS typically focus on common variants that contribute a small portion to the total variation in the phenotype [10]. Recently, many rare variants of high penetrance have been identified to contribute to the phenotypes of interest, e.g., blood pressure [11]. Therefore, pharmacogenetic studies that interrogate all genetic variants (both common and rare) by sequencing the whole genome or candidate genes in each subject will provide the most information and move the field of personalized medicine forward.

The Sanger method, invented and published in 1977, has been used for DNA sequencing and involves DNA synthesis in the presence of chain-terminating inhibitors followed by electrophoresis [12]. While it has excellent accuracy and reasonable read length, the very

low throughput and expense makes it unsuitable for deciphering the human genome. The breakthrough came in 2005 when the next-generation sequencing (NGS), the sequencing-by-synthesis technology, developed by 454 Life Sciences was published [13]. Since then several NGS platforms have been developed and applied to various fields of biological and medical research including pharmacogenetics. Currently, the Illumina HiSeq-2000 and HiScan, the Roche 454 GS-FLX Titanium, and the Applied Biosystems SOLiD Analyzer 5500xl are commercially available and the most used platforms, but with two of these platforms (Illumina HiSeq-2000 and HiScan) dominating the market. In addition, new NGS platforms are under development mainly based on single DNA molecule sequencing technology (e.g., nanotechnology and electron microscopy) which can read through DNA templates in real time without amplification, thus providing accurate sequencing data with potentially long reads. Examples include the Pacific BioSciences RS system that produces reads of >1,000 bp and the nanoAnalyzer from BioNanomatrix (now BioNano Genomics) that generates reads of around 400,000 bp [14].

The cost of whole-genome sequencing by NGS has dropped substantially so more samples are being analyzed. At the same time, the amount of data generated per samples gets larger and larger. Therefore, storing, managing, analyzing, and interpreting these results become the crucial step in an NGS project. NGS opens many challenges for individuals skilled in bioinformatics, statistics and medicine to identify and interpret the genetic variations in pharmacogenetics. Extremely high performance computing and intensive bioinformatics support are essential to utilize fully the benefit of applying NGS in personalized medicine.

The first step in NGS data analysis is to align or assemble the huge amount of short reads to a reference genome. This is a crucial and basic requirement and a variety of algorithms and software packages have been specifically developed for dealing with millions of NGS short reads alignment or assembly [15]. Compared to reference-based assembly, *de novo* assembly with NGS data has more challenges. Currently it is generally limited to genomes small in size such as bacteria [16].

After alignment or assembly of NGS short reads, the next step is to identify genetic variants (mutations) in DNA-sequencing or to quantify expression levels of genes or transcripts in RNA-sequencing.

***Corresponding author:** Huixiao Hong, Division of Systems Biology, NCTR/ FDA, 3900 NCTR Road, Jefferson, AR 72079, USA, Tel: (870) 543-7296; E-mail: huixiao.hong@fda.hhs.gov

Received April 04, 2012; Accepted April 06, 2012; Published April 09, 2012

Citation: Hong H (2012) Next-Generation Sequencing and Its Impact on Pharmacogenetics. J Pharmacogenom Pharmacoproteomics 3:e119. doi:10.4172/2153-0645.1000e119

Copyright: © 2012 Hong H. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

This is very challenging. For identification of genetic variants, one of the challenges is how to distinguish the causal variants from the large number of apparently novel genetic variants present by chance in any human genome. The more challenging task in analyzing NGS data is to distinguish low-frequency alleles descended from ancient ancestors from the rare mutations introduced into the population [17]. For RNA-sequencing, as a read can be aligned equally to multiple genomic locations, quantification of expression of genes and transcripts depends on the methods of counting the mapped short reads. Furthermore, because of the short length of the reads from NGS and alternative splicing events, discovery of allelic-specific or isoform-specific genes remains challenging [18].

The field of pharmacogenetics is changing rapidly today. The “common disease-common variants hypothesis” based GWAS have not fulfilled the expectation in the scientific community after a great deal of effort. In contrast, since NGS can enable us to identify all the causative variants (including both common and rare genetic variants) this approach is driving the next wave in pharmacogenetics. It is anticipated that NGS will make significant contributions to our understanding of pharmacogenetics and will redefine the field of personalized medicine. It is likely that key advancements will be seen soon. First, a powerful infrastructure will be developed to support the storage, access, and management of the huge (and still growing) amount of NGS data [19]. Second, powerful bioinformatics tools will be developed for assessment of significance of rare variants identified in genes which has pharmacogenetic impact potentially in genome sequences of patients. Third, data mining approaches will be improved to mine the large genome sequence databases and drug response data of patients. This will lead to better understanding of how genetic variations affect responses to drugs (efficacy and adverse reactions) and for identifying and utilizing pharmacogenetic biomarkers in clinical practice [3]. Lastly, there is a growing awareness among federal agencies, such as the FDA and NIH, of the need to develop a strategic long-term plan to coordinate efforts of the public and private sectors in advancing NGS research and development relevant to pharmacogenetics for personalized medicine. As the pharmacogenetics knowledge base expands, application of NGS data of pharmacogenetics to regulatory decision making will most likely follow.

Together we can look to a brighter future wherein advances in pharmacogenetics leads to improved medical care.

References

- President's Council of Advisors on Science Technology (2008) Priorities for Personalized Medicine.
- Motulsky AG (1957) Drug reactions enzymes and biochemical genetics. *J Am Med Assoc* 165: 835-837.
- Hong H, Goodsaid F, Shi L, Tong W (2010) Molecular biomarkers: a US FDA effort. *Biomarkers Med* 4: 215-225.
- The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851-861.
- Lavedan C, Licamele L, Volpi S, Hamilton J, Heaton C, et al. (2009) Association of the NPAS3 gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. *Mol Psychiatry* 14: 804-819.
- Turner ST, Bailey KR, Fridley BL, Chapman AB, Schwartz GL, et al. (2008) Genomic association analysis suggests chromosome 12 locus influencing antihypertensive response to thiazide diuretic. *Hypertension* 52: 359-365.
- Takeuchi F, McGinnis R, Bourgeois S, Barnes C, Eriksson N, et al. (2009) A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. *PLoS Genet* 5: e1000433.
- Kindmark A, Jawaid A, Harbron CG, Barratt BJ, Bengtsson OF, et al. (2008) Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J* 8: 186-195.
- Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, et al. (2009) HLA-B*5701 genotype is a major determinant of drug induced liver injury due to flucloxacillin. *Nat Genet* 41: 816-819.
- Frazer KA, Murray SS, Schork NJ, Topol EJ (2009) Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10: 241-251.
- Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, et al. (2008) Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 40: 592-599.
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.
- Das SK, Austin MD, Akana MC, Deshpande P, Cao H, et al. (2010) Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. *Nucleic Acids Res* 38: e177.
- Li H, Homer N (2010) A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform* 11: 473-483.
- Chistoserdova L (2010) Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol Lett* 32: 1351-1359.
- van Oeveren J, Janssen A (2009) Mining SNPs from DNA sequence data; computational approaches to SNP discovery and analysis. *Methods Mol Biol* 578: 73-91.
- Luco RF, Allo M, Schor IE, Kornbliht AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* 144: 16-26.
- Richter BG, Sexton DP (2009) Managing and analyzing next-generation sequence data. *PLoS Comput Biol* 5: e1000369.