Genome diagnostics: next-generation sequencing, new genome-wide association studies and clinical challenges

Expert Rev. Mol. Diagn. 11(7), 663–666 (2011)

Dimosthenis E Ziogas

Department of Surgery, Ioannina University School of Medicine, Ioannina, TK 451 10, Greece

Dimitrios H Roukos

Author for correspondence
Department of Surgery, Ioannina University School of Medicine, Ioannina, TK 451 10, Greece
and
Personalized Cancer Genomic Medicine, Human Cancer Biobank Center, Ioannina University, Ioannina, TK 451 10, Greece
Tel.: +30 265 100 7423
Fax: +30 265 100 7094
droukos@uoi.gr

"The most rapidly evolving field of biomedical research is genomics, from the first draft of the complete human genome sequence a decade ago to the estimated 30,000 whole-genome sequencing at the end of 2011. The planned development of multiple next-generation sequencing machines worldwide reflects the high level of interest in genomics within the scientific community."

Next-generation sequencing (NGS) technologies have revolutionized biomedical research and provided a foundation for application of the human genome in health and disease. As all diseases arise from the accumulation of genetic and epigenetic alterations that are associated with dysregulation of gene expression, genomics research in both structure and function of the genome now provides reasonable hope for the future achievement of genomic medicine [1]. High-quality comparative effectiveness research in health and disease [2] can provide evidence for causative, underlying structural genome variation patients with complex common or rare disease. However, understanding gene regulation is still a daunting challenge. This article discusses whether whole-genome or exome sequencing is sufficient for understanding gene function in health and disease that can lead to the development of genome heterogeneity-based diagnostic tools and markers to identify high-risk asymptomatic individuals or disease progression in patients.

The genomic revolution

The most rapidly evolving field of biomedical research is genomics, from the first draft of the complete human genome sequence a decade ago to the estimated 30,000 whole-genome sequencing at the end of 2011. The planned development of multiple NGS machines worldwide [3] reflects the high level of interest in genomics within the scientific community. There are many reasons for this increase. First, genome sequencing appears to be essential for completing mutation catalogs and understanding the wide genetic variation involved in the pathogenesis of most diseases. Second, the costs of human whole-genome sequencing have been drastically reduced to US$5000 and could be further reduced in the coming years. However, the goal of reducing sequencing costs to US$1000 per genome is unlikely to become a reality given that bioinformatics analysis required for identifying causal single-nucleotide polymorphisms (SNPs) or copy-number variations (CNVs) and distinguishing them from passenger, also termed non-causal, variations. Third, although DNA sequencing alone may have substantial limitations to cure diseases, it will provide the background for elucidating how genome variation gives rise to genome dysfunction and pathogenesis.

EXPERT REVIEWS

KEYWORDS: common mutations • diagnostics • genome-wide association studies • next-generation sequencing technology • rare mutations • whole-genome sequencing

 10.1586/ERM.11.56 ISSN 1473-7159

The era of genome-wide association studies

The Human Genome Project published a reference of the human genome sequence [4] along with a reference database of 3.2 million human genome SNPs [5], allowing biotechnology companies to develop commercialized chips with approximately 1 million SNPs. The next step was easy – comparing patients' populations with controls in genome-wide association studies (GWAS) using mostly SNPs-chips could lead to the discovery of specific SNPs causatively related to various diseases. The hope was that the identification of novel genes and variants could result in whole genome-based diagnostic tools that not only would allow improved diagnosis of various diseases, but could also identify high-risk individuals without clinical symptoms, leading to early disease prevention [6].

NGS-based medical applications

DNA sequencing can also provide pharmacogenomic properties of individual patients in terms of therapeutic index response and adverse events profile predictions for specific drugs. Such an achievement is crucial for patient selection and personalized medicine. Indeed, the unique capacity of NGS technology for complete human genome analysis of both targeted sequencing of a protein-coding region (exome) and whole-genome sequencing, including the noncoding DNA, could identify genetic variants related to response and toxicity to cardiovascular or anticancer drugs [7–9].

Facing the reality

The current consensus based on available GWAS data is that GWAS have modest efficacy in identifying high-risk individuals and therefore the expectations of preventive medicine have been lowered in contrast with the initial enthusiasm. Despite the wealth of dozens of genes and hundreds of SNPs, the effective size of the susceptibility to disease provided by each SNP is too small that has no clinical application [6] and most inherited risk remains unexplained [10]. Approximately 20–25% of heritability can be explained with current GWAS [10]; even efforts to combine these newly identified genetic variants with environmental factors and clinicopathologic characteristics have failed to predict the risk of diseases such as breast cancer [11].

> "In cancer diagnostics and drug response prediction for patient selection, cancer genome sequencing has focused on the identification of specific mutations that characterize an individual patient's tumor."

Potential solutions to overcome this 'missing heritability' [12] include the use of SNPs-chips with over 1 million SNPs and CNVs for new GWAS in larger populations aiming at identifying not only common but also rare variation.

DNA sequencing-based diagnostics

The clinical implications of both new GWAS using SNPs and CNVs and whole-genome sequencing are limited. For example, rare CNVs involved in disease pathogenesis have recently been identified in a GWAS. This study using CNVs chips in 8290 patients with schizophrenia and in 7431 healthy controls showed that *VIPR2* duplications are involved in schizophrenia [13]. However, only 0.35% of these patients carry rare CNVs in the chromosomal locus 7q36.3. Similarly, another study found a link between CNVs and schizophrenia at this locus [14]. Therefore, these two studies [13,14] support testing for CNVs at this locus to predict individuals at high risk for developing schizophrenia. Collectively, evidence is accumulating that CNVs at certain chromosomal loci are risk factors for schizophrenia but these variants are rare and thus testing for CNVs can identify only 2–4% of schizophrenia patients [15].

In cancer diagnostics and drug response prediction for patient selection, cancer genome sequencing has focused on the identification of specific mutations that characterize an individual patient's tumor. Given the wide genetic heterogeneity of cancer among patients with the same type of cancer and same traditional established clinicopathologic features, mutations-based diagnostics derived from cancer tissues sequencing can guide treatment decision. The *V600E* mutation in the *BRAF* gene has been the most common defect identified by tumor screening in melanomas. The Cobas® 4800 BRAF V600 Mutation Test (Roche Molecular Systems) was developed when Phase I and II studies showed a treatment response rate to the kinase inhibitor vemurafenib (PLX4032) of up to 80% in patients with metastatic melanoma tested positive for *BRAF* V600E mutation [16]. A Phase III randomized trial used the Cobas 4800 BRAF V600 Mutation Test to enroll patients with metastatic melanoma that tested positive for this mutation, and demonstrated improved overall survival and progression-free survival among patients that received the targeted drug, PLX4032 compared with those treated with the standard chemotherapy, dacarbazine [17]. This improved overall survival, in contrast to other studies on targeted drugs that demonstrated only a progression-free survival without a survival prolongation, suggests that PLX4032 can now become a new standard therapy for patients with metastatic melanoma and has also established the BRAF V600 Mutation Test as an essential diagnostic real-time PCR assay for treatment decision.

DNA sequencing-based diagnostics limitations

Testing for *KRAS* mutation status has recently become popular for deciding whether to add cetuximab to chemotherapy treatments in patients with metastatic colorectal cancer. However, more recent Phase III randomized trials with a more appropriate design in which patients with KRAS wild-type colorectal cancer were enrolled, found, in contrast to previous studies, no significant overall survival benefit either in metastatic or adjuvant settings [18,19]. Another paradigm of sequencing clinical limitations is genotyping in lung cancer. A fundamental scientific discovery was made in 2004, with the characterization of a distinct mutation-based subset of patients with EGF receptor (EGFR)-mutant non-small-cell lung cancer. As a result of subsequent intensive research it was thought that testing for EGFR mutations is clinically relevant because patients with EGFR-mutant tumors have increased sensitivity to tyrosine kinase inhibitors (TKIs) [20]. But, as is very common in translating research into clinical medicine, there are now problems and clinical debate for the adoption of this EGFR mutation test to predict resistance to erlotinib and gefitinib [21]. Promises of complete sequencing-based

diagnostics and TKIs response prediction, provides a recent study [22]. Jones and colleagues support that genomic characterization of a rare tumor, such as adenocarcinoma of the tongue, can guide treatment by predicting sensitivity to TKIs [22]. This finding requires further studies for confirmation and clinical use.

Whole-genome sequencing application

Although exome and whole genome sequencing is a new area of research and the costs for genome-wide analyses have only recently have been reduced, there has been a critique on limited clinical applications. Here, two more recent prime paradigms of genome sequence-based approaches for the clinic are reported.

Bainbridge and colleagues recently report in *Science Translational Medicine*, the sequence of the genomes of Alexis and her twin brother [23]. Dopa-responsive dystonia, a rare genetic disorder associated with abnormal movements, was diagnosed at the age of 5 years in the twins, who were being treating with a dopamine precursor for their dystonia. At 13 years of age, Alexis developed a cough and a breathing problem. Following many inconclusive diagnostics tests to explain Alexis' breathing problem, the research team, led by Gibbs, sequenced the twins' genomes and found mutations in the *SPR* gene. This gene encodes the enzyme sepiapterin reductase and previously had been reported to be involved in some patients with dopa-responsive dystonia. Given that sepiapterin reductase enables the synthesis not only of the neurotransmitter dopamine but also serotonin, Gibbs and colleagues suggested additional treatment with a serotonin precursor, 5-hydroxytryptophan. The clinical success of this medication was impressive. Alexis' breathing problem had disappeared and her brother Noah's handwriting problem was improved [23].

> "...next-generation sequencing technology and emerging high-throughput genome-wide mapping techniques allow, for the first time, an in-depth understanding of life complexity and disease heterogeneity."

The fact that rare disorders can be caused by single gene defects suggests that in the next few years, the rapidly dropping costs of NGS technology-based exome or whole-genome sequencing and bioinformatics analysis may permit the identification of these causal mutations, leading to the solution of diagnostic and therapeutic challenges in the vast majority of these rare diseases. By contrast, for common chronic disorders such as cancer, heart disease, diabetes, neurodegenerative and psychotic diseases, it is difficult to use whole-genome sequencing alone to solve the problem of diagnostics and therapy. Pathobiology and genesis of these diseases is driven by multiple genetic and environmental factors and gene–gene and protein–protein complex interactions in a dynamic process, the inference of which will require genome sequencing in addition to genome function for the understanding and discovery of predictive models to elucidate genome functional regulatory networks in health and disease.

Awaiting genomic medicine

Currently, NGS technology and emerging high-throughput genome-wide mapping techniques allow, for the first time, an in-depth understanding of life complexity and disease heterogeneity. Next-generation GWAS using improved microarray chips with expanded numbers of SNPs and CNVs with methodological power to detect both common and rare genetic variants will provide insights into 'missing heritability'. Moreover, systematic genomic studies using the new technologies will complete the mutations catalog of human diseases. But the pressing, still unresolved, problem is to start to understand how genome variability in dysregulated gene function leads to disease pathogenesis. Breakthrough research in model organisms and humans is underway to understand how transcriptional activity, noncoding RNAs, epigenetics, nucleosomes remodeling chromatin dynamic states and their biomolecules interactions influence gene regulation causing disease. At the same time, rational network biology research is focused on how to discover predictive models that are able to predict the inference of complex dynamic gene–gene and protein–protein networks. The ultimate goal of all these efforts is to discover genome-based diagnostics. These innovative diagnostic tools based on a combination of both genetic heterogeneity and genome dysfunction may result in personalized diagnostics and disease prevention strategies based on an individual's risk prediction.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1 Green ED, Guyer MS. National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. *Nature* 470(7333), 204–213 (2011).

2 Tinetti ME, Studenski SA. Comparative effectiveness research and patients with multiple chronic conditions. *N. Engl. J. Med.* 364(26), 2478–2481 (2011).

3 [No author listed]. Human genome: genomes by the thousand. *Nature* 467 (7319), 1026–1027 (2010).

4 International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 431(7011), 931–945 (2004)

5 ENCODE Project Consortium; Birney E, Stamatoyannopoulos JA, Dutta A *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146), 799–816 (2007).

6 Manolio TA. Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* 363(2), 166–176 (2010).

7 Weinshilboum RM, Wang L. Pharmacogenetics and pharmacogenomics: development, science, and translation. *Annu. Rev. Genomics Hum. Genet.* 7, 223–245 (2006).

8 Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N. Engl. J. Med.* 364(12), 1144–1153 (2011).

9 de Bono JS, Ashworth A. Translating cancer research into targeted therapeutics. *Nature* 467(7315), 543–549 (2010).

10 Park JH, Wacholder S, Gail MH *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* 42(7), 570–575 (2010).

11 Wacholder S, Hartge P, Prentice R *et al.* Performance of common genetic variants in breast-cancer risk models. *N. Engl. J. Med.* 362(11), 986–993 (2010).

12 Manolio TA, Collins FS, Cox NJ. Finding the missing heritability of complex diseases. *Nature* 461(7265), 747–753 (2009).

13 Vacic V, McCarthy S, Malhotra D *et al.* Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* 471, 499–503 (2011).

14 Levinson DF, Duan J, Oh S *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am. J. Psychiatry* 168, 302–316 (2011).

15 Piggins HD. Schizophrenia: zooming in on a gene. *Nature* 471(7339), 455–456 (2011).

16 Flaherty KT, Puzanov I, Kim KB *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* 363, 809–819 (2010).

17 Chapman PB, Hauschild A, Robert C *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* 364(26), 2507–2516 (2011).

18 Maughan TS, Adams R, Smith CG *et al.*; MRC COIN Trial Investigators. Identification of potentially responsive subsets when cetuximab is added to oxaliplatin-fluoropyrimidine chemotherapy (CT) in first-line advanced colorectal cancer (aCRC): mature results of the MRC COIN trial. *J. Clin. Oncol.* 28(15 Suppl.) (2010) (Abstract 3502).

19 Alberts SR, Sargent DJ, Smyrk TC *et al.* Adjuvant mFOLFOX6 with or without cetuximab (Cmab) in KRAS wild-type (WT) patients (pts) with resected stage III colon cancer (CC): results from NCCTG Intergroup Phase III trial N0147. *J. Clin. Oncol.* 28(Suppl. 18) (2010) (Abstract CRA3507) .

20 Pao W, Chmielecki J. Rational, biologically based treatment of *EGFR*-mutant non-small-cell lung cancer. *Nat. Rev. Cancer.* 10(11), 760–774 (2010).

21 Keedy VL, Temin S, Somerfield MR *et al.;* For the American Society of Clinical Oncology Provisional Clinical Opinion. Epidermal growth factor receptor (EGFR) mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. *J. Clin. Oncol.* 29(15), 2121–2127 (2011).

22 Jones SJ, Laskin J, Li YY *et al.* Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biol.* 11(8), R82 (2010).

23 Bainbridge MN, Wiszniewski W, Murdock DR *et al.* Whole-genome sequencing for optimized patient management. *Sci. Transl. Med.* 3, 87re3 (2011).