Is Next-Generation Sequencing Appropriate for the Clinic?

Ramzi Hassouneh[1,2]

[1]Faculty of Medicine, University of Ottawa
[2]Kidney Research Centre, Cellular and Molecular Medicine

ABSTRACT

Next-generation sequencing (NGS) has ignited a revolution in genomic science and presents a new tool for use in the clinical setting. Previous sequencing technologies suffered from inherent limitations in throughput, scalability, and speed, thus restricting this tool to laboratory research. One can now perform exome sequencing as well as whole genome sequencing at a low cost and quickly enough so that it can be used in the clinic to make diagnoses or pre-screen for risk to future disease. Despite its clinical uses, many challenges exist before next-generation sequencing becomes a mainstay in the clinical setting. The largest challenge remains in data analysis. There is a lack of understanding of the impact of genetic variants and mutations on health and disease and how to best apply genetic information to patient care. Multidisciplinary teams including physicians, medical geneticists, and genetic counselors need to be established and be familiar with this technology, both with what it can and cannot offer as well as its technical and ethical challenges. Nevertheless, the translation of base pair reads to clinical applications has truly begun.

RÉSUMÉ

Le séquençage de prochaine génération a déclenché une révolution dans la science génomique et offre un nouvel outil qui pourra servir dans le milieu clinique. La technologie qui était utilisée précédemment pour le séquençage avait des limites inhérentes dans son débit de traitement, son extensibilité et sa vitesse, ce qui limitait l'utilisation de cet outil à la recherche en laboratoire. Il est maintenant possible de séquencer l'exome en plus de procéder au séquençage du génome complet, à faible coût et avec une rapidité suffisante pour que nous puissions l'utiliser en clinique pour établir des diagnostics ou pour le prédépistage des risques ultérieurs de maladie. Malgré son utilité dans le contexte clinique, il reste de nombreux défis à relever avant que le séquençage de prochaine génération ne soit utilisé de façon courante dans le milieu clinique. Le plus grand de ces défis demeure l'analyse des données. Il y a encore beaucoup d'incompréhension quant aux répercussions des variantes et des mutations génétiques sur la santé et sur la maladie. Il faut en outre élucider comment cette information génétique peut être appliquée au soin des patients. Il faut créer des équipes multidisciplinaires comprenant des médecins, des généticiens médicaux et des conseillers en génétique, puis les familiariser avec cette technologie, de même qu'avec ce qu'elle peut faire et ce qu'elle ne peut pas faire. Il faut également les informer des défis techniques et éthiques que cette technologie présente. Il n'en reste pas moins que la transcription des paires de bases pour l'utilisation clinique est véritablement lancée.

INTRODUCTION

DNA sequencing technologies have progressed rapidly in the last decade. The initial sequencing of the human genome cost approximately $70 million dollars and took about a decade to complete [1]. The technology used to complete the Human Genome Project was based on the sequencing method introduced by Sanger in the 1970s [2]. Although this project was a remarkable achievement, the limitations of the technology created a demand for rapid, cost-effective, and accurate DNA sequencing data. In 2004, the National Human Genome Research Institute (NHGRI) sought to fund the development of technologies capable of genome sequencing at a cost of $1,000 within a decade [3]. This initiative led to the emergence of next-generation sequencing (NGS) technologies. The ultimate goal of the initiative by the NHGRI is to reduce the cost of genome sequencing so that it can be implemented into the clinical practice of medicine. Despite ushering a critical and transformative period in the history of DNA sequencing with respect to genomics research, the readiness of NGS technologies for clinical application remains unclear. While studies have demonstrated the clinical application of NGS in diagnosing novel diseases, determining best treatment approaches, and screening for disease-relevant genes, the rapid evolution of these technologies have outpaced the development of other essential resources needed to achieve their full potential.

Keywords: next-generation sequencing; translational medicine; medical ethics

METHODOLOGY OF DNA SEQUENCING

The classical method of sequencing developed by Sanger required several arduous steps, starting with fragmenting the genomic DNA of interest and cloning it into a plasmid vector and transforming E. coli [2]. An amplified template is obtained from the bacterial colony and sequencing reactions for each DNA base are performed in the presence of chain-terminating dideoxynucleotides [2]. The resulting mixture of radiolabelled terminated products is subsequently denatured and run in separate lanes on polyacrylamide gels to separate strands based on size [2]. Finally,

the DNA sequence can be determined by viewing the gel using autoradiography and reading the banding pattern on the gel [2]. This process allows approximately 1,000 bases of sequence to be read per run, making it extremely slow when sequencing the human genome, which has almost 3 billion base pairs [2].

NGS is capable of generating data rapidly by sequencing large amounts of DNA in parallel using various approaches including Illumina sequencing, Roche 454 sequencing, Ion Torrent sequencing, and SOLiD sequencing. Although each platform differs in regards to sequencing biochemistry, they all share similarities in their conceptual design. Firstly, genomic DNA is fragmented and common adapters are ligated [1]. Then an array of immobilized PCR (polymerase chain reaction) colonies is generated according to each platform's unique protocol [1]. This array can then be enzymatically manipulated (i.e. addition of primers and extensions) in parallel and fluorescent labels can be detected with each extension to determine the sequence [1]. The array can be continuously manipulated to build up the entire DNA sequence, allowing sequencing of an entire human genome in less than a day [1].

EXOME VS. WHOLE GENOME SEQUENCING

Only a small percentage of the human genome's sequence is characterized, therefore it is often more effective and less expensive to sequence only the protein-coding regions (i.e. exome) or disease-relevant genes to screen for relevant mutations for the diagnosis and treatment of disease [4]. Due to the smaller region being sequenced in exome sequencing, specific sites of interest can be sequenced multiple times or at a higher depth of coverage [5]. A higher depth of coverage indicates higher accuracy. Although annotated regions are highly accurate, the shorter read lengths in exome sequencing provide less reliable information about the relative location of specific base pairs [5]. For the aforementioned reasons, exome sequencing is only limited to detecting coding and splice-site variants in annotated genes, making it most suitable for highly penetrant Mendelian disease gene identification [6]

While whole genome sequencing is still costly, it is capable of uncovering all genetic and genomic variations as well as discovering functional coding and non-coding variations amounting to approximately 3.5 million variants [5]. Compared to targeted sequencing, whole-genome sequencing has longer reads but with a lower depth of coverage [5]. This makes it more suited for identifying Mendelian and complex trait genes, as well as rare phenotypes caused by de novo single nucleotide variations or copy number variations. Table 1 summarizes the differences between targeted and whole-genome sequencing [7].

Table 1: Exome versus whole-genome sequencing.

	Cost (per sample)*	*Coverage*	*Depth*	*Read Length (10^6 bp)*
Exome	$1,000-$2,000	Small	60x-200x	30-100
Genome	$4,000-$20,000	Large	15x-70x	225-1050

*Presented in US dollars

CLINICAL APPLICATIONS: EXOME SEQUENCING

Several groups have demonstrated the effectiveness of targeted whole exome NGS approaches in diagnosing novel diseases in the clinic. Whole-exome sequencing was used to make genetic diagnoses of congenital chloride diarrhea in a subset of patients showing symptoms of Bartter syndrome, a renal salt-wasting disease [8]. Exome sequencing was performed on six patients who lacked mutations in known genes for Bartter syndrome on site-specific sequencing [8]. The data revealed homozygous deleterious mutations in the SLC26A3 (Solute Carrier Family 26, Member 3) gene for all six patients, suggesting a molecular diagnosis of congenital chloride diarrhea [8]. This diagnosis was later confirmed upon clinical evaluation [8]. Furthermore, whole-exome sequencing was used to make diagnoses in individuals with novel diseases such as childhood intractable inflammatory bowel disease and autoimmune lymphoproliferative syndrome [9,10].

Targeted sequencing strategies have also been successfully used to screen for diseases. For instance, NGS methods are capable of preconception screening of up to 448 severe recessive pediatric diseases that were formerly impossible under single gene testing models [11]. Additionally, targeted NGS have been used to screen panels of cancer genes, in particular for mutations in the breast cancer genes, BRCA1 and BRCA2, in individuals with a family history of breast and ovarian cancer [12].

Another application for targeted sequencing is to characterize the genetic basis of diseases. This approach was used to determine the genetic basis of 10 non-familial pancreatic neuroendocrine tumors [13]. The findings showed that the most common mutations occurred in the following cell cycle regulating genes: MEN1 (multiple endocrine neoplasia type 1), DAXX (death associated protein 6), ATRX (alpha thalassemia/ mental retardation syndrome X-linked), and mTOR (mechanistic target of rapamycin) [13]. These findings are clinically relevant due to the availability of drugs that inhibit mTOR pathways, therefore providing new avenues regarding treatment plans for this disease [13].

CLINICAL APPLICATIONS: GENOME SEQUENCING

Due to the significant decrease in sequencing costs made possible by NGS technologies, re-sequencing entire human genomes from clinical samples is becoming increasingly feasible. Whole genome NGS was used to identify a promyelocytic leukemia – retinoic acid receptor alpha (PML-RARA) gene fusion event, characteristic of acute promyelocytic leukemia, in a patient with an undifferentiated form of leukemia [14]. More significantly this fusion event was identified and confirmed within a week of obtaining a biopsy, thus allowing the physicians to alter the patient's treatment plan suitably [14]. Similarly, both whole-genome and targeted exome sequencing were used to develop and carry out a clinical protocol that identified individualized regions of mutations for tumors from patients with metastatic colorectal cancer and malignant melanoma [15]. Additionally, these tumor mutations were detected within 24 days

of biopsy and allowed the formation of biomarker-driven clinical trials in these types of cancer [15].

Whole genome sequencing can also be used pre-clinically for identifying novel cancer-related genetic abnormalities. Many cancer genomes have been sequenced using NGS providing a wealth of information on mutated processes and gene regulatory networks in different cancers. An exhaustive record of somatic mutations in a malignant melanoma genome revealed mutations linked to previous UV exposure, as well as 470 previously unknown somatic substitutions and 42 previously unknown mutations [16]. Similarly, NGS was used to sequence breast cancer genomes to elucidate the mutation and evolution of primary cancer to metastasis [17]. These findings showed a large deletion as a result of metastasis during disease progression, and 20 previously unknown mutations common to primary and metastatic cancer cells [17].

Cell-free DNA fragments circulating in an individual's bloodstream can also be sequenced using NGS. This non-invasive technique was applied to diagnose an acute cellular rejection following a heart transplant procedure [18]. A diagnosis was made after finding an increased level of cell-free DNA fragments from the donor's genome in the recipient bloodstream, and this was confirmed with an endomyocardial biopsy [18]. These findings relied on high coverage sequencing of cell-free DNA to identify minute amounts of DNA that belonged to the donor's heart [18]. Additionally, NGS was used to sequence cell-free DNA from a pregnant mother's bloodstream belonging to her fetus to detect severe chromosomal abnormalities such as Down syndrome, trisomy 13, and trisomy 18 [19].

ETHICAL CONSIDERATIONS

The biggest challenge in performing whole-exome and whole-genome sequencing on a clinical basis is dealing with the potential of identifying unexpected sequence variants that are unrelated to the primary reason for ordering the sequencing test. The American College of Medical Genetics (ACMG) recommends that laboratories report mutations of 24 highly medically actionable genetic diseases in all subjects, irrespective of age, but excluding fetal samples [20]. It is approximated that 1-3% of NGS tests present such findings and this may fall outside the expertise of the physician who ordered the test [20]. This situation presents an ethical dilemma, as the physician might be held liable if there is a failure to disclose this incidentally detected risk. This is particularly concerning when the patient is a child and the disease in question may only onset in adulthood. A child or their family should not be burdened with this knowledge, as a genetic mutation associated with an adult-onset disease does not accurately predict the development of said disease particularly due to other predictors of such disorders such as environmental factors. On the other hand, the parents and the child can take precautions to prevent the occurrence of this disease provided with preconceived knowledge. Therefore, physicians and staff must be trained and educated suitably to be able to use NGS appropriately and effectively. This necessitates a multidisciplinary team with genetic counselors who can provide medical guidance to patients. Patients must be made aware that NGS testing is not a comprehensive test for all disease-associated variants and that additional specific testing may be necessary to confirm the presence of incidental findings. Furthermore, since genetic information is shared among relatives, genetic variants in one individual may also affect that person's parents, siblings, or offspring. The patient -- and with consent, the patient's family – should be provided with information on the inheritance and penetrance of the condition in question.

Prenatal genetic testing is another area of ethical concern. These tests allow couples to determine if there is a risk that their unborn child will have a genetic condition. Although this may allow the couple to plan for delivery, it also allows couples to decide whether to continue a pregnancy. Similar to the problem with incidental findings, parents must be made aware that prenatal genetic testing does not identify all of the possible gene mutations that can cause a particular condition and that they may have limited predictive value.

PRACTICAL CHALLENGES

Although NGS technologies present many applications in the clinical setting, several practical challenges exist that must be addressed for these technologies to achieve their full potential. NGS approaches produce enormous sequence data sets that range from millions to billions of bases; therefore, their interpretation is not a simple undertaking. The nature of data produced by NGS technologies place substantial demands on computational infrastructure [17]. Generation, analysis, storage, and management of NGS data requires a high-end computing infrastructure with highly trained bioinformatics staff to maintain and run NGS data analysis tools [21]. Cloud computing presents a good alternative to investing in expensive advanced computing infrastructure and allows the easy sharing of information among different clinics, labs, and practices [21].

INHERENT LIMITATIONS OF DNA SEQUENCING

DNA sequencing has inherent limitations in that it does not provide information relating to gene expression or RNA transcript levels. A mutated gene identified via sequencing is not an accurate predictor of mutated proteins. Processes that may affect end-product protein include gene expression, post-transcriptional regulation, epigenetic regulation, protein expression, and post-translational modification. DNA microarray is one technology capable of measuring expression levels of a large number of previously sequenced genes simultaneously [22]. This technology has been mostly replaced by RNA sequencing which is capable of measuring the complete set of RNA transcripts that are produced by an un-sequenced genome either under specific circumstances or in a specific cell [23]. For instance, this technique can be used to distinguish healthy cells from pathological cells by identifying differentially expressed transcripts. Epigenetic regula-

tory systems, including DNA methylation, histone modification, and non-coding RNA associated gene silencing, are recognized as big determinants of gene expression [24]. Epigenetic changes have been shown to be predictors of several diseases and these changes can be detected using methylation sequencing or chromatin immunopreciptation sequencing (ChIP-Seq) [24]. previously sequenced genes simultaneously [22]. This technology has been mostly replaced by RNA sequencing which is capable of measuring the complete set of RNA transcripts that are produced by an un-sequenced genome either under specific circumstances or in a specific cell [23]. For instance, this technique can be used to distinguish healthy cells from pathological cells by identifying differentially expressed transcripts. Epigenetic regulatory systems, including DNA methylation, histone modification, and non-coding RNA associated gene silencing, are recognized as big determinants of gene expression [24]. Epigenetic changes have been shown to be predictors of several diseases and these changes can be detected using methylation sequencing or chromatin immunopreciptation sequencing (ChIP-Seq) [24].

DIRECT-TO-CONSUMER TESTING

As sequencing prices continue to plummet, commercial laboratories began to market to the public personalized testing of an individual's genome, regardless of whether there was a medical indication for such a test. These direct-to-consumer tests do not provide pre-test counseling which can help the individual understand the limitations of the test. Interpretation of results and post-test counseling are essential to provide meaning to the information provided to the individual. The ACGM strongly recommends that a certified medical geneticist or genetic counselor be involved in the process of ordering and interpreting such tests [25]. Unfortunately, the continued increase in the speed, decreases in cost, and the ability to perform it on samples easily collected at home will make direct-to-consumer genetic testing more prevalent in the near future [25]. Government oversight might be necessary to restrict or regulate the public's access to such services.

CONCLUSION

The evidence provided demonstrates the promise and dramatic effect of NGS on the diagnosis of genetic conditions. Because of advances in NGS techniques, parents may have the option in the foreseeable future to have their child's genome sequenced at birth to enable unprecedented health management and personalized care. The progress that has been made recently in the field of NGS is very encouraging, but certain challenges must be overcome in order for it to be used routinely in the clinic. The largest challenge is one of data interpretation. There is a lack of understanding of the impact of genetic variants and mutations on health and disease. Finally, there is lack of physician and patient understanding on how to implement genetic information for health benefits. Better education in genetics, as well as better tools are required for an effective integration of genetic data into the practice of medicine. Despite these obstacles, there are many examples of successful implementation of NGS in the clinic and it is only a matter of time before genetic medicine is an integral part of clinical care.

REFERENCES

1. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010; 11: 31–46.
2. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977; 74: 5463–7.
3. Mardis ER. Anticipating the 1,000 dollar genome. Genome Biol 2006; 7: 112.
4. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011; 12: 745–55.
5. Gonzaga-Jauregui C, Lupski J, Gibbs R. Human Sequencing in Health and Disease. Annu Rev Med 2012; 63: 35-61.
6. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medicine. Journal of Human Genetics 2014; 59: 5-15.
7. Biesecker L, Green R. Diagnostic Clinical Genome and Exome Sequencing. N Engl J Med 2014; 370: 2418-2425.
8. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci USA 2009; 106: 19096–101.
9. Bolze A, Byun M, McDonald D, et al. Whole-exome-sequencing-based discovery of human FADD deficiency. Am J Hum Genet 2010; 87: 873–81.
10. Worthey EA, Mayer AN, Syverson GD, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 2011; 13: 255–62.
11. Bell CJ, Dinwiddie DL, Miller NA, et al. Carrier testing for severe childhood recessive diseases by next- generation sequencing. Sci Transl Med 2011; 3: 65ra4.
12. Shen P, Wang W, Krishnakumar S, et al. High-quality DNA sequence capture of 524 disease candidate genes. Proc Natl Acad Sci USA 2011; 108: 6549–54.
13. Jiao Y, Shi C, Edil BH, et al. DAXX/ ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. Science 2011; 331: 1199–203.
14. Welch JS, Westervelt P, Ding L, et al. Use of whole-genome sequencing to diagnose a cryptic fusion onco-gene. JAMA 2011; 305: 1577–84.
15. Roychowdhury S, Iyer MK, Robinson DR, et al. Personalized oncology through integrative high-throughput sequencing: a pilot study. Sci Transl Med 2011; 3: 111ra21.
16. Pleasance ED, Cheetham RK, Stephens PJ, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010; 463: 191–6.
17. Ding L, Ellis MJ, Li S, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 2010; 464: 999–1005.
18. Snyder TM, Khush KK, Valantine HA, et al. Universal noninvasive detection of solid organ transplant rejection. Proc Natl Acad Sci USA 2011; 108: 6229–34.
19. Palomaki GE, Deciu C, Kloza EM, et al. DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. Genet Med 2012; 14: 296–305.
20. Green R, Berg J, Grody, W, et al. ACMG Recommendations for Reporting of Incidental Findings in Clincal Exome and Genome Sequencing. Genet Med 2013; 15(7): 565-574.
21. McPherson JD. Next-generation gap. Nat Methods 2009; 6: S2–5.
22. Alizadeh, A, Elsen, M, Davis, R, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000; 430: 503-511.
23. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10(1): 57-63.
24. Meaburn E, Schulz R. Next generation sequencing in epigenetics: insights and challenges. Semin Cell Dev Biol 2012; 23(2): 192-9.
25. Su, P. Direct-to-Consumer Genetic Testing: A Comprehensive View. Yale J Biol Med 2013; 86(3): 359-365.