

## Bringing Next-Generation Sequencing Oncology Tests into the Diagnostic Setting

Chee-Seng Ku\*

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

\*Corresponding author: Chee-Seng Ku, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, Tel: +46 8 31 11 01; E-mail: [cheeseng.ku@ki.se](mailto:cheeseng.ku@ki.se)

Rec date: Oct 5, 2015; Acc date: Jan 29, 2016; Pub date: Feb 1, 2016

Copyright: © 2016 Ku CS. 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.

### Introduction

Since the completion of the Human Genome Project (HGP) in 2003 and the appearance of the first next-generation sequencer (i.e. Roche 454 GS20) on the market [1], it has been widely anticipated that the rate of genomic discovery would increase exponentially. Genomic discovery in the context of human genetics encompasses the identification of new variants or mutations underlying human disease, from rare Mendelian disorders to common complex diseases including cancer, and pharmacogenomics for drug responses [2-4]. It has been envisaged that these 'disruptive technologies' will eventually be translated into the diagnostic setting. Thus, the term 'genomic medicine' has been widely used to describe this rapidly evolving field i.e. applying genomic information in the context of furthering the clinical care of patients. Pharmacogenomics is that branch of genomic medicine that specializes in applying genomic information in the context of therapeutic decision-making [5,6].

The completion of the HGP and the advent of multiple Next-Generation Sequencing (NGS) technologies in its wake have spawned a significant number of new developments in genomic medicine. One of the first examples was provided by the sequencing of an entire cancer genome by means of NGS to determine the genetic aberration in a patient with a diagnosis of Acute Myeloid Leukaemia (AML) of unclear subtype. This Whole-Genome Sequencing (WGS) analysis detected a novel insertional translocation on chromosome 17 which generated a pathogenic PML-RARA gene fusion, thereby confirming a diagnosis of acute promyelocytic leukemia (a subtype of AML with a favorable prognosis), and making possible the improved management of the patient [7]. The combination of NGS technologies with new sequence enrichment methods has also made Whole-Exome Sequencing (WES) feasible. The impact of WES on clinical management was first evidenced by the case of a patient, diagnosed with an intractable inflammatory bowel disease through the identification of a missense mutation in the XIAP gene, in whom allogeneic hematopoietic progenitor cell transplantation was subsequently performed [8].

In addition to WGS and WES, rapid advances have also been made in the application of NGS technologies in targeted sequencing of a panel of genes to detect somatic mutations in tumor tissue, which is important for therapeutic decision-making in the context of molecular targeted drugs. For example, somatic mutations in EGFR inform the prescription of drugs (such as erlotinib and gefitinib) targeting the gene product, a tyrosine kinase [9,10]. However, detecting rare (yet clinically actionable) somatic mutations has been very challenging using conventional methods such as Sanger sequencing. This is due to a combination of tissue heterogeneity and genetic heterogeneity, as well as temporal heterogeneity i.e. tumor evolution; taken together, these

account for somatic mutations often being characterized by low frequencies (<20%) in the tumor tissue. Whilst Sanger sequencing suffers from limited sensitivity in the context of detecting rare somatic mutations, quantitative PCR methods have limited scalability for multiplexing which could allow detection of up to hundreds or thousands of mutations in a cost- and time- effective manner. By contrast, by generating digital data, NGS technologies fulfill both these key requirements for their application in the diagnostic setting. Amplicon multiplexing is performed such that thousands of amplicons containing many different mutations are amplified and sequenced together, and sequencing depth is increased to enhance somatic mutation detection; as rare mutations can be detected at the level of single individual sequence reads (Table 1). Cost effectiveness may be further improved by sample indexing whereby multiple samples can be pooled together in the same sequencing run [11,12].

Methods	Sensitivity to detect rare mutations	Multiplexing of targets (Amplicons and Mutations)
Sanger sequencing	Limited	Limited scalability
Array genotyping	Limited (analog data)	High (thousands of mutations)
qPCR	High	Limited scalability
NGS	High (digital data)	High (and cost effective)

Table 1: Comparison of the common methods to detect rare somatic mutations.

The arrival of bench-top sequencers such as Illumina MiSeq and the Life Technologies Ion Torrent Personal Genome Machine (PGM) has further accelerated the pace of NGS transfer into the diagnostic setting. The bench-top sequencers are rather more suitable than the conventional high-throughput sequencing platforms such as Illumina HiSeq or Life Technologies 5500xl SOLiD System in terms of their adoption in the diagnostic setting. The reasons are manifold including the fact that the size of the targeted gene region is very much smaller than that required for either WES or WGS, and the requirement for smaller numbers of clinical samples to be run on a regular basis, which would hamper the batching of larger sample sizes. This leads to a requirement for a lower sequencing throughput in the diagnostic setting. In addition, the sequencing run time for bench-top sequencers is much more rapid, taking several hours rather than the days required for high-throughput NGS platforms [13]. Below we elaborate the challenges and preconditions to bring NGS into a diagnostic setting for therapeutic decision-making and highlight the first CE-IVD NGS diagnostic test for oncology that has been brought to the market.

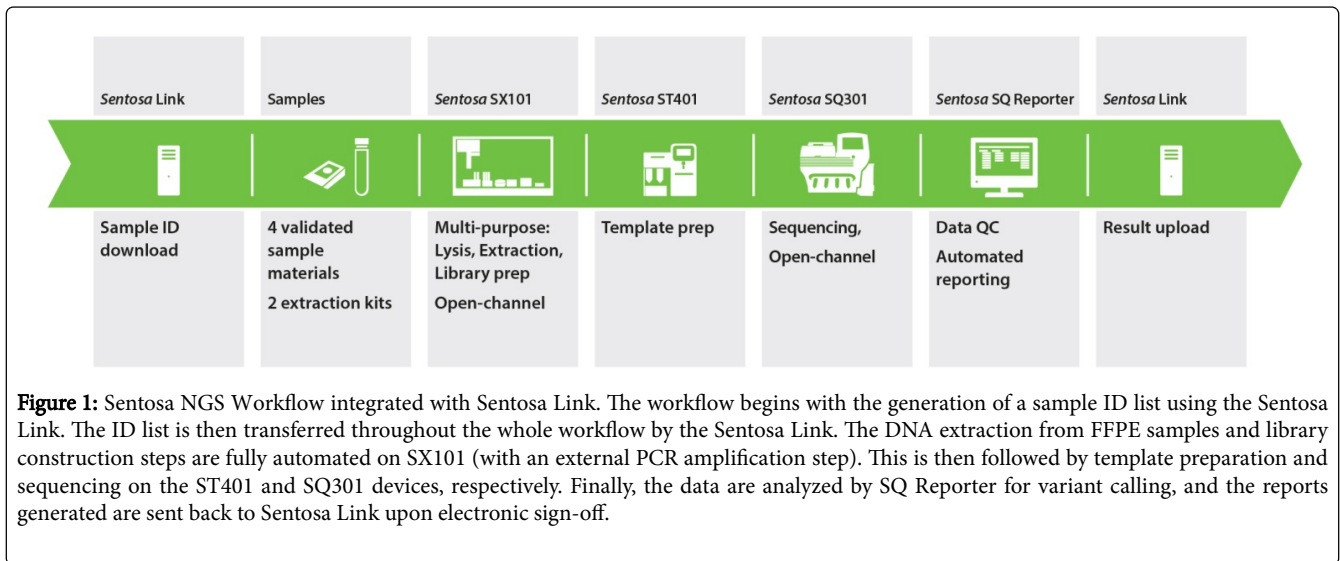
Although the gradual permeation of NGS into the diagnostic setting has been envisaged, there have been many challenges particularly in the context of detecting somatic mutations in oncology [14-17]. The challenges range from assay design to sample preparation, data analysis, variant calling, and automation of the entire workflow from samples to results. In terms of the assay design, it is critical to ensure that targets are selected so as to be both of clinical relevance and actionable. In addition, the tests should achieve the required levels of analytical and clinical sensitivity and specificity for diagnostic application.

The analytical sensitivity and specificity can be established by detecting known mutations in well-characterized reference materials. On the other hand, clinical performance is established by comparing the concordance and discordance rates of mutation detection in clinical or reference samples already tested by reference methods e.g. quantitative PCR and Sanger sequencing (the latter for mutations with frequencies >20%). As NGS tests are characterized by their high multiplexing capability (i.e. up to hundreds or thousands of mutations can be targeted), testing every single mutation to assess clinical performance represents quite a challenge. What hampers this validation is that a large number of samples would be required to establish the parameters and it is probably impracticable to obtain an adequate number of samples, especially for low population frequency mutations among cancer patients.

In addition to the selection of clinically relevant targets, and the requirements of analytical and clinical performance, an automated workflow from sample preparation to data analysis, variant calling and result reporting is also very critical in the diagnostic setting. The importance of this requirement is well exemplified in the NGS tests, where the workflow involves many steps from DNA extraction from tumor tissue to library construction, template preparation, sequencing and data analysis. Validation is needed to ensure the reproducibility of the workflow, and automation is a critical component for this. For example, robotic automation should ensure reproducibility in terms of quality and quantity of the DNA extraction from tumor tissue and library preparation as well as minimizing operator errors during the tedious manual steps. Automation also brings other significant advantages i.e. reducing the labor requirement for the work (i.e. improving the laboratory's manpower efficiency) and avoiding the need to repeat samples due to operator errors (increasing cost effectiveness). Other critical requirements in the diagnostic workflow lie in ensuring sample traceability, where automatic sample ID list is being transferred from one part of the workflow to another, this serves to avoid manual entry errors in the ID list.

Another challenge that hinders the widespread application of NGS in the diagnostic context is data analysis, which requires specialist expertise in bioinformatics. This includes the optimization of analysis parameters, alignment of sequence reads, variant calling and annotation. This is important to ensure high sensitivity and specificity with respect to detecting the target mutations. Building a robust and validated bioinformatic analysis pipeline to meet the strict requirements of analytical and clinical performance is resource-intensive. Furthermore, the analysis part should also be user friendly so that laboratory personnel without specialist bioinformatics training can perform it. The reports generated at the end of the analysis must also be designed specifically to meet requirements in the clinical context, where sufficient QC information must be included to allow assessment of the quality of the runs and samples, whilst the report for clinicians must be concise, summarizing only the mutations being sought or detected.

Although many commercial companies have been working towards bringing NGS oncology tests onto the market, Vela Diagnostics was the first by dint of launching its NGS oncology test and the automated workflow for melanoma with CE-IVD approval in September 2014 ([www.veladx.com](http://www.veladx.com)), and oncology tests for colorectal cancer, non-small cell lung cancer, and thyroid cancer have also been launched subsequently with CE-IVD mark, and all these tests have also obtained TGA (Therapeutic Goods Administration) approval in Australia. These tests are designed to detect somatic mutations in tumor samples aiding the prescription of molecular targeted drugs. This is an important milestone in the translational effort to bring NGS oncology test into the diagnostic setting. The workflow that was launched by Vela Diagnostics to support the NGS oncology tests (<http://www.veladx.com/products.html#ngs>) fulfilling the preconditions discussed above bringing NGS tests into the diagnostic setting become a reality. The workflow begins with automated genomic DNA extraction and library construction on a robotic liquid handling system (Figure 1). This automated workflow offers several significant advantages. For example, it ensures the reproducibility of libraries in terms of quality and quantity. Further, in the validated workflow, no QC steps are required to measure the concentration of extracted genomic DNA, or to assess the library's quality and quantity. After automated library construction is completed, the pool library is used for subsequent template preparation and sequencing to generate the sequencing data. The data are analyzed by the software i.e. Sentosa SQ Reporter automatically, where the steps such as alignment and variant calling take place. The automation of the entire workflow also makes it feasible to achieve a turn-around time of 2 days from DNA extraction to report generation for each clinical sample.



The widespread application of NGS tests in the diagnostic setting is growing apace. In addition to oncology tests designed to identify somatic mutations for therapeutic decision-making, NGS also has huge potential as a diagnostic tool in the context of Mendelian disorders. This is well exemplified by the Illumina MiSeqDx Cystic Fibrosis System, which is the first FDA-cleared IVD NGS platform for the comprehensive testing of cystic fibrosis [18]. These tests have so far used a targeted gene sequencing approach to sequence specifically those gene regions containing clinically relevant mutations. It is foreseeable that more NGS diagnostic tests will be developed and launched into the marketplace as the rapidly evolving discipline of genomic medicine is devoted to utilizing genomic information to further the clinical care of patients and to improve health outcomes.

## References

- 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.
- Mizzi C, Peters B, Mitropoulou C, Mitropoulos K, Katsila T, et al. (2014) Personalized pharmacogenomics profiling using whole-genome sequencing. *Pharmacogenomics* 15: 1223-1234.
- Rabbani B, Tekin M, Mahdih N (2014) The promise of whole-exome sequencing in medical genetics. *J Hum Genet* 59: 5-15.
- Ku CS, Cooper DN, Polychronakos C, Naidoo N, Wu M, et al. (2012) Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol* 71: 5-14.
- Manolio TA, Chisholm RL, Ozenberger B, Roden DM, Williams MS, et al. (2013) Implementing genomic medicine in the clinic: the future is here. *Genet Med* 15: 258-267.
- Green ED, Guyer MS (2011) National Human Genome Research Institute: Charting a course for genomic medicine from base pairs to bedside. *Nature* 470: 204-213.
- Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, et al. (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456: 66-72.
- Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, et al. (2011) Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med* 13: 255-262.
- Gillis NK, Patel JN, Innocenti F (2014) Clinical implementation of germ line cancer pharmacogenetic variants during the next-generation sequencing era. *Clin Pharmacol Ther* 95: 269-280.
- de Mello RA, Madureira P, Carvalho LS, Araújo A, O'Brien M, et al. (2013) EGFR and KRAS mutations, and ALK fusions: current developments and personalized therapies for patients with advanced non-small-cell lung cancer. *Pharmacogenomics* 14: 1765-1777.
- Mwenifumbo JC, Marra MA (2013) Cancer genome-sequencing study design. *Nat Rev Genet* 14: 321-332.
- Meyerson M, Gabriel S, Getz G (2010) Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 11: 685-696.
- Ku CS, Wu M, Cooper DN, Naidoo N, Pawitan Y, et al. (2012) Technological advances in DNA sequence enrichment and sequencing for germline genetic diagnosis. *Expert Rev Mol Diagn* 12: 159-173.
- Bennett NC, Farah CS (2014) Next-generation sequencing in clinical oncology: next steps towards clinical validation. *Cancers (Basel)* 6: 2296-2312.
- Salto-Tellez M, Gonzalez de CD (2014) Next-generation sequencing: a change of paradigm in molecular diagnostic validation. *J Pathol* 234: 5-10.
- Pant S, Weiner R, Marton MJ (2014) Navigating the rapids: the development of regulated next-generation sequencing-based clinical trial assays and companion diagnostics. *Front Oncol* 4: 78.
- Xuan J, Yu Y, Qing T, Guo L, Shi L (2013) Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* 340: 284-295.
- Bijwaard K, Dickey JS, Kelm K, Težak Ž (2015) The first FDA marketing authorizations of next-generation sequencing technology and tests: challenges, solutions and impact for future assays. *Expert Rev Mol Diagn* 15: 33-40.

This article was originally published in a special issue, entitled: "**Sequencing Technologies**", Edited by Jianping Wang