

Next Generation Sequencing: An Introduction for the Pathology Laboratory



Introduction

Molecular diagnostics play a key role in medicine in the diagnosis and classification of diseases, and increasingly in personalized cancer medicine (1). The recent expansion of knowledge about underlying genetic changes in cancer has revealed a set of actionable mutations and other genetic changes that can be therapeutically targeted (2). Notable examples are KIT mutations in gastrointestinal stromal tumors (3), EGFR, KRAS and ALK mutations in lung cancer (4), and BRAF mutations in melanoma (5). Gene copy variations and structural variants such as translocations are also important in both diagnosis and prognosis (6). Until recently, small-scale methods such as allele-specific polymerase chain reaction (AS-PCR), Sanger dideoxy sequencing, pyro sequencing, multiplex ligation-dependent probe amplification (MLPA), or mass spectrometry (MS) were the only methods used to identify gene mutations such as these. Gene copy number and structural variants have been frequently measured in a separate cytogenetics laboratory using fluorescence *in situ* hybridization (FISH).

What is Next Generation Sequencing?

The recent development of new molecularly targeted therapeutics needing companion diagnostics, as well as an increasing number of useful molecular-based biomarkers assisting diagnosis and prognosis, have increased the demand for testing. A decrease in biopsy size also results in a demand for more and more assays to be done on less and less tissue. Techniques that measure only one or a few biomarkers at a time are not useful in such cases. Next Generation Sequencing (NGS) allows the ability to sequence larger panels of genes in cancer with less tissue and as a result can meet this demand (7).

NGS, or massively parallel sequencing, is a technological innovation allowing for the sequencing of millions of small fragments of DNA at the same time, resulting in a massive increase in the amount of base pairs sequenced compared to the standard Sanger sequencing method. This new technology was based on technology developments starting in the mid to late 1990s that launched the first of the next generation platforms starting in 2000 (8).

NGS involves a few different methods that all result in much higher throughput and lower prices for sequencing DNA than with Sanger sequencing. It can produce in excess of one billion short reads per instrument run, delivering fast, inexpensive and accurate genome information (8). Currently the most popular and widely used platform, Illumina, sequences DNA by synthesis, using a combination of a modified shotgun sequencing approach and the addition of a fluorescent dye on the nucleobases. NGS technologies allow for sequencing of targeted regions, whole exome sequencing, or whole genome sequencing (7).

Next Generation Sequencing – Whole Genome Sequencing (WGS)

Whole genome sequencing (WGS) analyzes both protein-coding and non-coding regions in the human genome. WGS is the most comprehensive method, enabling the simultaneous detection of substitutions, duplications, insertions, deletions, gene and exon copy number changes, and chromosome inversions and translocations across the entire genome. However, WGS is expensive, with large computational demands for data storage and processing, and is low throughput. WGS typically produces lower depth of coverage therefore limiting the sensitivity of detection for low-abundance mutations (9). The clinical significance of most genomic alterations detected by WGS is unknown.

Next Generation Sequencing – Whole Exome Sequencing (WES)

The exome accounts for only 1.5% of the human genome, and yet includes 85% of all disease-causing mutations (10). Whole exome sequencing (WES) examines all protein-coding regions in the human genome. In this method, DNA fragments are hybridized in solution to sequence-specific capture probes corresponding to all protein-coding exons throughout the genome. WES enables the simultaneous detection of substitutions, duplications, insertions, deletions, and gene and exon copy number changes in many genes in a single assay. Compared to targeted gene panels, non-tumor specific, germline variants associated with disease will also be detected. WES typically produces lower depth of coverage, limiting the sensitivity of detection for low-abundance mutations compared to targeted gene panels. However, compared to WGS, WES has clear cost and speed advantages (10).

Next Generation Sequencing – Gene Panels, Hybridization Capture

With hybrid capture methods, DNA fragments are hybridized in solution to sequence-specific capture probes corresponding to targeted regions of the genome. Examples of hybridization capture technology include Agilent SureSelect, NimbleGen SeqCap, and Illumina TruSeq (11). These tests are designed to interrogate tissues for mutations of interest in specific genes, typically 50 to several thousand. Hybridization capture enables simultaneous detection of substitutions, duplications, insertions, deletions, and exon and gene copy number changes in many genes in a single assay. Probes can also be designed to capture specific translocation breakpoints in recurrently rearranged genes. When sequencing to high depth of coverage (500–1000x coverage), these assays are sensitive enough to detect low abundance mutations. Compared to Amplicon Capture (next section), this method is more specific without the artifacts seen with PCR. Compared with WES or WGS methods, these panels are usually cheaper and/or faster, since they are more limited in scope and are targeting smaller regions of interest. However, they will only detect mutations in the *targeted* regions.

Next Generation Sequencing – Gene Panels, Amplicon Capture

Amplicon sequencing enriches target genes by PCR with a set of primers for the exons of selected genes prior to NGS (12). Examples of amplicon capture technology include pure PCR-based methods such as Ion Torrent AmpliSeq, RainDance ThunderBolts or Illumina TruSeq amplicon, and hybridization and extension methods such as Agilent HaloPlex. These tests are designed to interrogate tissues for mutations of interest in specific genes, typically 1 to 100. Amplicon capture requires low inputs of DNA and enables the simultaneous detection of single-base substitutions as well as more complex mutations including duplications, insertions, and deletions in many genes in a single assay. The informatics analysis is relatively easy, as any read that does not map to a locus between primers can be disregarded. A downside of this simplicity is that the assay is inherently unable to detect unexpected fusions, because either the 5' or 3' primer would fail to bind the translocated DNA. Also, these methods can be prone to artifacts such as allele dropout, a problem associated with PCR resulting in the failure of the amplification of one of the two alleles

at a given locus. Similar to hybridization capture, amplicon capture is sensitive enough to detect low-abundance mutations and can be cheaper and/or faster.

Problems and Opportunities Specific to Cancer NGS

Sample Quality

Sample quality and type play an important role in the quality of NGS sequencing. Many factors can impact quality and/or quantity of the DNA extracted from samples, including pre-analytic factors such as cold ischemia, fixation and processing (e.g., fresh frozen versus FFPE samples, presence or absence of decalcification), length and conditions of slide and block storage, tumor size and cellularity, tumor fraction, and tumor viability (13, 14). It is very important to have the anatomic pathologist evaluate the section and confirm the diagnosis, quality of specimen, and other morphologic features mentioned before extracting (tumor size, cellularity, fraction and viability).

Tumor molecular profiling is generally performed on DNA extracted from formalin-fixed, paraffin-embedded (FFPE) tissue specimens due to the logistical complexities of preparing and storing fresh or frozen tissue in the clinical laboratory. Formalin fixation causes fragmentation and cross-linking of DNA, and storage of the block may further damage the DNA over time (13, 14, and 15). Because of this fragmentation and cross-linking, much of the NGS data on FFPE tissue have been derived from sequencing targeted amplicons (16). However, despite these issues, studies have shown the use of FFPE tissue in the clinical lab (17).

Sample and Tumor Heterogeneity

An important consideration in cancer NGS is specimen heterogeneity. Almost all excised tumor samples also include genetically normal tissue such as stroma or adjacent normal tissue. Prior to testing, the tissue specimen requires examination by a pathologist to confirm the presence of tumor, its viability and cellularity and to determine the tumor content/fraction in the specimen. As a result, many groups sequence heterogeneous cancer specimens to a higher median read depth than that used for constitutional DNA. The term 'read depth' or 'coverage' is a reflection of how often a specific region of the genome has been sequenced.

A tumor sample that contains 50% normal tissue would require double the read depth to detect the tumor mutations with the same confidence as a 100% pure tumor sample. If the tumor fraction is too low, manual dissection of non-tumor tissue can be done to enrich the specimen. Another consideration is tumor heterogeneity; where only a portion of the tumor cells contain the specific mutation. NGS can help detect minor clones when sequencing is done to a high read depth. The more clones that are present, the higher the read depth needs to be to represent each clone properly. NGS has helped to highlight the fact that cancers often do not comprise a single dominant clone, but may have multiple subclones at non-trivial frequencies that comprise part of the entire tumor (1). Clones representing a fraction of the original tumor have the potential to become the predominant clone during drug-resistant relapse. For example, a 1% clone will only be represented once in 100× coverage, assuming the tumor contains no normal tissue.

Liquid Biopsies

Liquid biopsies are an emerging method to assess tumor mutations from DNA circulating in the bloodstream. There are two sources of such DNA that can be less invasively assessed in the circulation: cell-free circulating tumor DNA (ctDNA) and circulating tumor cells (CTCs). ctDNA is composed of small fragments of nucleic acid that are not associated with cells or cell fragments (18). ctDNA and/or CTCs can potentially be used to screen for early-stage cancers, monitor responses to treatment and help explain why some cancers are resistant to therapies. They may also more accurately reflect the entire tumor genome than individual biopsies or blocks. Due to tumor heterogeneity, biopsies often suffer from sample bias. Tumor sampling for some cancer types, besides costly and risky to the patient, remains difficult resulting in inadequate amount of tissue available for genetic testing. Additionally, biopsies will only inform of the genotype at that specific point in time. NGS is becoming increasingly important in liquid biopsy applications. Although liquid biopsies would allow for longitudinal monitoring, they are not yet recommended by professional organizations. There are many reasons for this, in particular, low analytic sensitivity for certain mutations has been found in specific cases. It remains challenging to detect rare mutations in a background of wild-type sequences. In addition, pre-analytic methods

of specimen collection and processing, as well as quantitation methods need to be harmonized. While promising, more work on this technique needs to be done to fully validate the technique (19, 20).

Useful Professional Organizations that Offer Resources, Education, and Guidelines:

**College of American Pathologists
(general and molecular diagnostics):**
www.cap.org

**Association of Molecular Pathologists
(molecular diagnostics):**
www.amp.org

**American College of Medical Genetics
(inherited diseases)**
www.acmg.net

**American Society of Clinical Oncology
(cancer diagnostics and treatment):**
www.nccn.org

**American Society of Clinical Oncology
(cancer diagnostics and treatment):**
www.asco.org

References

- (1) Ulahannan *et al.* Technical and implementation issues in using next generation sequencing of cancers in clinical practice. *Br. J. Cancer* (2013) 109: 827-835.
- (2) Wong *et al.* Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours. *Sci. Rep.* (2013) 3: 3494.
- (3) Hirota *et al.* Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* (1998) 279: 577-580.
- (4) Paez *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* (2004) 304:1497-1500.
- (5) Davies *et al.* Mutations of the BRAF gene in human cancer. *Nature* (2002) 417: 949-954.
- (6) Mitelman *et al.* The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* (2007) 7: 233-245.
- (7) Types of molecular tumor testing. *My Cancer Genome* (2016) Updated February 8 <https://www.mycancergenome.org/content/molecular-medicine/types-of-molecular-tumor-testing/>
- (8) Barba *et al.* Historical perspective, development and applications of next generation sequencing in plant virology. *Viruses* (2014) 6: 106-136.
- (9) Gundry *et al.* Direct mutation analysis by high-throughput sequencing: from germline to low-abundant, somatic variants. *Mutat. Res.* (2012) 729: 1-15.
- (10) Choi *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* (2009) 106: 19096-19101.
- (11) Bodi *et al.* Comparison of commercially available target enrichment methods for next generation sequencing. *J. Biomol. Tech.* (2013) 24: 73-86.
- (12) Chang *et al.* Clinical application of amplicon-based next generation sequencing in cancer. *Cancer Genet.* (2013) 206: 413-19.
- (13) Bass *et al.* A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue. *Arch. Pathol. Lab. Med.* (2014) 138: 1520-1530.
- (14) Chen *et al.* Analysis of pre-analytic factors affecting the success of clinical next generation sequencing of solid organ malignancies. *Cancers* (2015) 7: 1699-1715.
- (15) Gilbert *et al.* The isolation of nucleic acids from fixed, paraffin-embedded tissues—which methods are useful when? *PLoS ONE* (2007) 6: e537.
- (16) Kerick *et al.* Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med. Genomics* (2011) 4: 68.
- (17) Spencer *et al.* Comparison of clinical targeted next generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *J. Mol. Diagn.* (2013) 15: 623-633.
- (18) Bettegowda *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* (2014) 6: 224.
- (19) Ma *et al.* "Liquid biopsy" ctDNA detection with great potential and challenges. *Ann. Transl. Med.* (2015) 16: 235.
- (20) Heitzer *et al.* Circulating tumor DNA as a liquid biopsy for cancer. *Clin. Chem.* (2015) 61: 112-123.

Trusted Answers. Together.



Agilent Pathology Solutions

www.dako.com

Represented in more than 100 countries

Australia
+61 2 9922 0700

Austria
+43 1 408 43 34 0

Belgium
+32 (0) 16 38 72 20

Brazil
+55 11 50708300

Canada
+1 905 335 3256

China
+86 21 3612 7091

Denmark
+45 44 85 97 56

Finland
+358 9 348 73 950

France
+33 1 64 53 61 44

Germany
+49 40 69 69 470

Ireland
+353 1 479 0568

Italy
+39 02 58 078 1

Japan
+81 3 5232 9970

Korea
+82 2 2194 9577

The Netherlands
+31 20 42 11 100

Norway
+47 23 14 05 40

Poland
+48 58 661 1879

Spain
+34 93 344 57 77

Sweden
+46 8 556 20 600

Switzerland
+41 41 760 11 66

United Kingdom
+44 (0)1 353 66 99 11

United States of America
+1 805 566 6655

PR7000-0184

© Agilent Technologies, Inc. 2016

Printed in USA, May 4, 2016

5991-6907EN

29144-2016MAY05



Agilent Technologies