Perspectives

Integrating next-generation sequencing into the diagnostic testing of inherited cancer predisposition


The clinical application of next-generation sequencing (NGS) as a diagnostic tool has become increasingly evident. The coupling of NGS technologies with new genomic sequence enrichment methods has made the sequencing of panels of target genes technically feasible, at the same time as making such an approach cost-effective for diagnostic applications. In this article, we discuss recent studies that have applied NGS in the diagnostic setting in relation to hereditary cancer.

Conflict of interest
Nothing to declare.

The advent of next-generation sequencing (NGS) technologies has contributed significantly to cancer genetics/genomics research through cancer genome sequencing studies (1, 2). In addition, the clinical applications of NGS, for example, as a diagnostic tool, have become increasingly evident (3–5). The coupling of NGS technologies with new genomic sequence enrichment methods has made the sequencing of panels of target genes technically feasible, at the same time as making such an approach cost-effective for diagnostic applications (6). This was amply shown by Pritchard et al. (7) who leveraged these technological advances to sequence the known causal genes for Lynch syndrome (hereditary nonpolyposis colon cancer) and familial adenomatous polyposis (FAP), namely, MLH1, MSH2, MSH6, PMS2, EPCAM, APC and MUTYH. Lynch syndrome is characterized by deficiencies in DNA mismatch repair caused by loss-of-function mutations in any one of the first five of these genes. Mutations in APC and MUTYH are responsible for FAP/attenuated FAP and autosomal recessive MUTYH-associated polyposis syndrome, respectively. In passing, it should be noted that NGS has already been assessed for its applicability as a diagnostic tool to detect known germline mutations for other hereditary cancers (breast and ovarian). NGS coupled with custom-designed oligonucleotides was used in a targeted sequencing approach to enrich 21 genes responsible for conferring an inherited
Integrating NGS into the diagnostic testing of hereditary cancer

Current molecular genetic testing approaches for most diseases employ traditional low-throughput PCR–Sanger sequencing based on a one-by-one selection approach in which the candidate gene deemed most likely to harbor causal deleterious mutations is sequenced first. This is followed by testing of additional genes if the initial results are negative (3). This stepwise approach is both time-consuming and costly, because multiple genes must often be analyzed sequentially using expensive and low-throughput methodology. However, recent technological advances in genomic sequence enrichment and NGS have allowed entire panels of genes to be sequenced simultaneously without the need for prioritization. Using the Agilent targeted sequence capture method followed by sequencing on the Illumina HiSeq2000 NGS platform, Pritchard et al. showed the robustness of this NGS-based method, referred to as the ColoSeq assay, by detecting all known pathogenic mutations in the seven genes implicated in hereditary colon cancer syndromes in 26 samples (23 cancer patients and 3 colon cancer cell lines) with known germline mutations (7). These pathogenic lesions included nonsense, missense, frameshift, in-frame deletions and splice site mutations, as well as large deletions and duplications. Although large deletions and duplications can be detected using normalized depth of coverage and split-read analysis, the precise breakpoints often cannot be determined because they are commonly located within Alu (or other repetitive DNA) elements or segmental duplications, which are not captured by the sequence enrichment design of the ColoSeq assay. Although the breakpoints could not be determined to single nucleotide resolution, the ColoSeq assay showed at least exon-level resolution for all large deletions and duplications detected in the patient samples. This resolution was comparable or better than that of traditional approaches such as multiplex ligation-dependent probe amplification (7).

The sequence enrichment design of the ColoSeq assay utilized oligonucleotides that were designed to capture all exons, introns and approximately 10 kb of the 5′ and 3′ flanking regions of the seven genes causally implicated in hereditary colon cancer. In addition, the sequence capture also targeted a further 24 genes that rarely harbor mutations causing colorectal cancer. A total of 31 genes were therefore captured, which covered a total of 1.1 Mb genomic DNA after the removal of repetitive DNA sequences. Each captured sample library was then amplified by PCR to incorporate a unique index barcode for sample multiplexing and up to 96 barcoded samples were pooled and run on a single lane of an Illumina HiSeq2000 instrument using paired-end sequencing (i.e. 2 × 101 bp) (7). The longer paired-end sequencing reads are important for the detection of small indels.

Attempts had already been made to incorporate custom genomic sequence enrichment and NGS methods into the genetic diagnostic testing of Lynch syndrome. A test was designed so as to capture every exon in a panel of 22 genes (most associated with hereditary colorectal cancer), followed by NGS using (Roche 454 GS-FLX, Branford, CT) and the (Illumina GA, San Diego, CA), to evaluate their performance (15). Although these technologies are promising in the context of cancer diagnostics, their technical limitations must also be recognized, e.g. GC-rich regions are difficult to enrich. In a worst-case scenario, these GC-rich regions would not be captured at all (16). Off-target sequencing is also an issue that leads to redundant sequencing beyond the targeted regions. Further, uneven sequencing coverage across the targeted regions (because of several factors such as uneven enrichment, uneven sequencing and difficulty in aligning the sequence reads to repetitive regions) might result in poor sequence coverage in some of the regions, which subsequently affects the sensitivity and specificity of variant detection (15).

The identification of known causal mutations in the respective disease genes is critical as it serves to confirm the original clinical diagnosis of hereditary colorectal cancer. Lynch syndrome and adenomatous polyposis syndromes have some overlapping clinical features; for example, both attenuated FAP and Lynch syndrome are characterized by a similar median age of onset and comparable numbers of colonic polyps. For suspected Lynch syndrome, the testing of tumor tissue for microsatellite instability can be used as a filter prior to the germline sequencing of mismatch repair genes. Immunohistochemical analysis of mismatch repair protein expression can also be used in parallel to help prioritize the candidate genes for sequencing. However, there are various practical problems associated with the routine testing of tumor tissues for microsatellite instability and mismatch repair protein expression, including the standardization, reporting and follow-up of test results (17). Simultaneous testing for germline mutations in all
Ku et al.

relevant genes using robust and inexpensive NGS methods may soon obviate the need for microsatellite instability and/or immunohistochemical assays. Compared with Sanger sequencing, the high-throughput production of hundreds of gigabases of DNA sequence data by NGS is very cost-effective.

In addition to the studies of hereditary colorectal cancer, NGS has also been assessed for its applicability as a diagnostic tool to detect known germline mutations for hereditary breast and ovarian cancers (8). In similar vein, by leveraging the technological advances in custom genomic sequence enrichment and NGS, Walsh et al. designed custom oligonucleotides in solution to capture 21 genes known to be responsible for an inherited risk of breast and ovarian cancer (8). The enrichment followed by NGS on an Illumina GA platform was tested in 20 women diagnosed with breast or ovarian cancer and with a known mutation in one of the genes responsible for inherited predisposition to these cancers. It generated encouraging results in that the known point mutations, small indels (ranging from 1 to 19 bp), and large genomic duplications and deletions (ranging from 160 to 101,013 bp) were detected in all the samples. The large deletions and duplications were detected using a read-depth strategy and concurred with the multiple ligation probe assay. In addition to being able to detect point mutations and small indels, the ability to detect larger deleted and duplicated regions is a major advantage of NGS when compared with Sanger sequencing; this is important in a diagnostic test setting because some causal genes are affected by copy number variants. Large deletions and duplications constitute a significant component of the mutational spectrum for many inherited disorders, including the DNA mismatch repair genes underlying Lynch syndrome and the BRCA1/BRCA2 genes in breast cancer (7, 8). The promise of NGS in the genetic diagnosis of familial breast cancer has also been shown by another study that sought to detect TP53, BRCA1 and BRCA2 mutations in tumor cell lines and DNA from patients with germline mutations. All the known pathological mutations (including point mutations and small indels of up to 16 nucleotides) were successfully identified (18).

Although the detection of germline variants using a targeted sequencing approach is the focus of our discussion in the context of integrating NGS in cancer diagnostics, it is noteworthy that WGS has also been shown to be a promising tool for the detection of complex chromosomal rearrangements in cancer. More specifically, WGS has been used to unravel the genetic rearrangement in a patient with a diagnosis of acute myeloid leukemia (AML) of unclear subtype (19). The underlying genetic cause was a novel insertional translocation on chromosome 17, which generated a pathogenic PML-RARA gene fusion that served to confirm a diagnosis of acute promyelocytic leukemia (a subtype of AML with a favorable prognosis). In similar vein, WGS has been used to unravel the genetic cause of a patient with a suspected cancer susceptibility syndrome based on the early onset of several primary tumors; WGS analysis identified a novel heterozygous deletion of three exons of the TP53 gene, the intact copy of TP53 having been lost in the leukemic cells as a consequence of uniparental disomy (20).

**Bench-top NGS in targeted sequencing**

In these studies, conventional high-throughput NGS platforms were used for the targeted sequencing of a panel of genes, achieving a high sequencing depth. For example, Pritchard et al. (7) achieved a median of 320-fold coverage per nucleotide (range 145- to 556-fold) across the entire targeted region when 96 barcoded samples were pooled for sequencing. The median coverage across the seven ColoSeq genes was 475-fold. Thus, a higher coverage would be expected for a smaller number of barcoded samples. Although these studies showed the feasibility of using a high-throughput NGS platform, the production of several hundred gigabases of DNA sequence data in a single run might render it less suitable for sequencing a panel of genes from a much smaller number of samples. It is common for a single patient or a small number of samples to be encountered in a clinical diagnostic context, thereby rendering the barcoding or multiplexing of a large number of patient samples impractical on grounds of cost. For this reason, bench-top NGS machines with a lower throughput are more suitable for smaller sample sizes and have the advantage of a more rapid sequencing turnaround time (3). Available bench-top NGS instruments include the Roche 454 Genome Sequencer (GS) Junior Sequencing System, the Ion Torrent Personal Genome Machine Sequencer (Life Technologies, Carlsbad, CA) and the MiSeq Personal Sequencing System (Illumina, San Diego, CA). The throughput of these bench-top NGS instruments ranges from 10 Mb to >1 Gb. For example, several Ion Torrent sequencing chips are available for different throughputs ranging from >10 Mb (chip314), >100 Mb (chip316) and >1 Gb (chip318). Similarly, the Illumina Miseq produces sequence data up to >1 Gb depending on the read length and whether single-end or paired-end sequencing is employed. By contrast, the 454 GS Junior has a much lower throughput (>35 Mb) per instrument run but has a longer read length of 400 bp on average compared with the other two platforms (11).

Bench-top NGS machines have enhanced technical and logistical flexibility (e.g. a smaller number of samples can be processed) but at the same time are able to avoid ‘redundant’ sequencing, i.e. a higher than required depth. Indeed, by avoiding redundant sequencing, these platforms display improved cost-effectiveness. An adequate coverage is critical for the accurate detection of large deletions and duplications using normalized depth of coverage and split-read analysis. The approximately 320-fold coverage used by Pritchard et al. (7) was sufficient to correctly identify all six of the known large deletions/duplications in the samples included in their study. In routine molecular diagnostics, it is common to sequence a panel of genes in a small number of samples. Thus,
Integrating NGS into the diagnostic testing of hereditary cancer

it is likely that the total length of sequenced genomic DNA will range from tens to hundreds of kilobases for the coding regions of a handful of genes. Hence, the capacity of high-throughput NGS technologies greatly exceeds requirements in the context of a few clinical samples. By contrast, Sanger sequencing does not meet the demand of this sequencing requirement, either in terms of efficiency or cost-effectiveness. Therefore, the development of several bench-top NGS instruments has more than adequately filled a niche in the market.

Pros and cons of different NGS-based approaches

NGS technology is rapidly changing the landscape of genetic testing and offers new opportunities for molecular diagnostics through the efficient sequencing of panels of specific disease target genes. In addition to the hereditary cancer syndromes investigated by the studies discussed here, other diagnostic applications of NGS in the form of targeted gene sequencing, WES and WGS have already been shown for other Mendelian disorders (3). There is ongoing debate about which sequencing approach will come to the fore in the diagnostic setting in the near future. Compared with targeted gene sequencing, WES and WGS are powerful information-gathering tools with great potential for generating novel findings that are sometimes even incidental to the original aims of the test (21–23). There are ongoing ethical concerns relating to whether those incidental findings that are not pertinent to the original purpose of the diagnostic test should be disclosed to the patients. The targeted gene sequencing approach employed by Pritchard et al. (7), Walsh et al. (8) and others (3) focuses exclusively on known causal genes; hence, the disclosure of incidental findings should be of less concern.

Nevertheless, WES and WGS can be considered as universal, all-inclusive diagnostic tools for heritable diseases caused by germline mutations in protein coding regions. Employing WES and WGS as a diagnostic tool should avoid the logistical problems associated with designing tests individually for each of the hundreds or thousands of inherited diseases that could in principle be subjected to genetic diagnostic testing. In addition, tests based on WES and WGS will not become ‘outdated’ in the event that novel disease mutations or genes are identified (assuming that the genes have been annotated in the case of WES). However, WGS would be the only realistic option in a situation where it had not been possible to identify the causative genetic defect(s) by WES, suggesting a possible extragenic location for the pathological lesions, e.g. mutations within gene flanking or regulatory regions, or deep within introns (24). Further, WES and WGS are also particularly robust techniques for diseases characterized by a high degree of clinical, locus or allelic heterogeneity and for diseases where the currently known disease genes explain only a minority of cases. This was well shown by a study that applied WES to identify casual mutations for Charcot–Marie–Tooth disease, a highly heterogeneous condition in which >30 genes have been implicated (25). At the same time, WES and WGS are analytically very challenging because a much larger set of genomic sequence variants is produced in comparison with the targeted gene sequencing. Although a larger set of mutation/variant data is generated by WES and WGS, the analysis can be restricted to the known disease genes by computational means, a process known as ‘in silico enrichment or filtering’. However, this also raises the issue of the generation of additional genetic data that are not ultimately to be analyzed. Irrespective of how the ongoing debate develops, all three sequencing approaches (targeted, WES and WGS) are likely to find applications in the routine diagnostic setting.

Summary and perspectives

In summary, Pritchard et al. evaluated in a clinical laboratory setting the performance of targeted DNA capture and NGS for the detection of germline mutations in genes implicated in hereditary colon cancer syndromes. They showed the robustness of their ColoSeq assay by accurately detecting all known pathogenic mutations in the samples sequenced. Even more importantly, a 100% between-run reproducibility of mutation detection was also attained for pathogenic mutations and variants of uncertain significance that were detected in the analysis. In addition, they reported that the ColoSeq assay can be performed for approximately the same cost as the sequencing and deletionduplication analysis of a single gene using traditional methods, with at least equivalent or even better sensitivity and accuracy (7). This study, therefore, strongly supports the clinical feasibility of using NGS in cancer diagnostics, which was first shown by Walsh et al. (8) using breast and ovarian cancers as a model.

Although incidental findings are a less important issue in the context of targeted gene sequencing, ColoSeq is still capable of identifying novel and potentially deleterious mutations of uncertain or unknown clinical significance. For example, by performing ColoSeq in prospectively collected blood samples from 31 patients with a clinical history suggestive of Lynch or adenomatous polyposis syndrome, but without a previously known mutation, Pritchard et al. identified six patients with pathogenic mutations and three additional patients with variants of uncertain significance. All these mutations were confirmed using alternative methods. This study defined variants of uncertain significance as being those missense variants that were not well characterized and which met the following criteria: (i) population frequency <5%, (ii) occurrence in evolutionarily conserved amino acid residues, (iii) predicted to be at least possibly damaging by in silico prediction tools, and (iv) confirmed by Sanger sequencing (7). With the information generating power of NGS technology, the number of variants of uncertain significance identified in known disease genes has increased dramatically. This has created a significant challenge for the interpretation of genome sequence data, because
the potential pathological consequences of these variants are difficult to assess. Biochemical and/or cellular assays are required to determine the functional significance (e.g. whether or not the variant in question produces dysfunctional or non-functional protein) or otherwise variants of uncertain significance (such as missense variants identified in mismatch repair genes or the \textit{BRCA1} and \textit{BRCA2} genes). Such assays can in turn provide valuable insights into the clinical significance of these variants. Although different approaches to functional analysis have been adopted, the challenge remains in applying the experimental method to examining the effect of variants of uncertain significance. Further, confirmation of the functional consequences (and ultimately the clinical significance) of these variants extends far beyond the clinical diagnostics arena, and clearly requires new research (26–29).

There is some debate as to whether data on sequence variants with unknown clinical significance (but which might nevertheless be found to have pathological significance in the future) should be released to patients and/or their attending physicians, and whether patients have a right to demand access to their entire genetic dataset. Clearly, proper consultation with medical geneticists will be important in order to interpret the mutational data prior to the disclosure of results to patients. Finally, while the widespread adoption of NGS in molecular diagnostics appears inevitable, tests conducted in a clinical setting must be properly regulated. They should be performed according to the Clinical Laboratory Improvement Amendment (CLIA) if the results are to be communicated to patients or if they are to impact upon the clinical decision-making process in any way. Molecular diagnostic assays that rely upon NGS technology must always be rigorously validated before their adoption in the clinical realm.

References