Next generation sequencing (NGS) in oncology: lights and shadows

Margherita Nannini[1], Maria A. Pantaleo[1,2]

Abstract
Advances in tumor genome sequencing using *next generation sequencing* (*NGS*) technologies have facilitated a greater understanding of the genetic abnormalities involved in cancer development and progression, dramatically changing oncology research. There are several different types of *NGS* technologies. *Whole genome sequencing* (*WGS*) determines the sequence of the complete genome, providing information on both coding and non-coding regions and structural variants. However, use is limited by the large volume of data generated, and associated time and resource costs. *Whole exome sequencing* (WES) determines the sequence of coding regions only, making it faster and cheaper than *WGS*, and the functional consequences of variants are easier to interpret. However, all variations in non-coding regions are missed. *WGS* and *WES* are often used together to maximize detection of variants. A less costly approach is the use of targeted sequencing, which focuses on particular regions of interest, based on their biological relevance. *NGS* technologies can also be used to sequence RNA, referred to as *RNA-Seq*. All these NGS technologies, individually or in combination, have a number of potential applications, including identification of biomarkers, and development of diagnostic and therapeutic strategies. However, although advances have been made, there are a number of limitations to be overcome before *NGS* technologies are routinely applied in both research and clinical practice.

Key words: cancer, DNA, molecular biology, mutations, next generation sequencing, research, RNA

Introduction

Cancer is a biologically complex disease, with characteristics acquired during the course of a multistep development process that allow cancer cells to survive, proliferate and disseminate by sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, deregulating cellular metabolism and evading the immune system [1]. Underlying all these features is instability in the tumor genome that defines the molecular fingerprint of each cancer type [2].

In this landscape, advances in tumor genome sequencing using *next generation sequencing* (*NGS*) technologies have dramatically changed oncology research, leading to a deeper understanding of most genetic abnormalities involved in cancer development and progression [3]. In the last decade, many *NGS*-based studies have provided a comprehensive molecular picture of several types of cancer, and have led to the identification of a large number of new genomic, transcriptomic and epigenomic alterations, expanding knowledge about complexity, heterogeneity and evolution of the neoplastic disease process [4-15].

NGS technologies offer different applications depending on the aim of the research, the type of material to be sequenced, the coverage, and finally the speed and cost of sequencing [16, 17]. *Whole genome sequencing* (*WGS*) determines the sequence of the complete genome, providing information on both coding and non-coding regions and structural variants. However, interpretation of the data generated is difficult due to large volume of information generated, meaning that this approach can be time consuming and expensive. Moreover, some variants can be missed because of variation in coverage across the genome. In contrast, *whole exome sequencing* (*WES*) determines the sequence of coding regions only, making it faster and cheaper than *WGS*, and the functional consequences of variants are easier to interpret. However, all variations in non-coding regions are missed. Therefore, both *WGS* and *WES* are frequently performed together in order to detect

[1]Department of Specialized, Experimental and Diagnostic Medicine, S. Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy.
[2]"Giorgio Prodi" Cancer Research Center, University of Bologna, Bologna, Italy.
Correspondence to: Maria A. Pantaleo, MD, PhD, Dipartimento di Ematologia e Scienze Oncologiche "L.A. Seragnoli", Policlinico S. Orsola-Malpighi, Via Massarenti 9, 40138 Bologna, Italy.
Phone: + 39 051 6364078 – Fax: + 39 051 6364037
E-mail: maria.pantaleo@unibo.it
CANCER BREAKING NEWS 2016;4(1):17-19
DOI: 10.19156/cbn.2016.0004

as many variants as possible. Finally, targeted sequencing determines the sequence of specific genes or parts of genes, focusing on particular regions of interest, selected according to their biological relevance. This approach is usually less costly than the others (although total cost depends on the size of the gene panel), has a wide sequence coverage and is preferred for a clinical application. *NGS* technologies can also be used to sequence RNA, referred to as *RNA-Seq*, providing data on gene expression, novel cryptic translocations or gene fusions [18-21]. Thus, the integrated analysis of *WGS* and *RNA-Seq* can facilitate the interpretation of a large number of genomic alterations detected in the cancer genome, and can increase mutation detection performance, especially for low purity tumors [22-24]. In addition to gene expression and fusions, *RNA-Seq* can provide a broader profile of all tumor transcriptome, including noncoding RNAs, such as microRNAs (miRNAs), small interfering RNAs, ribosomal RNAs, small nucleolar RNAs and long noncoding RNAs, that represent more than half of the cancer transcriptome and play an important and growing role in multiple biological and pathological cellular processes [25-28].

All these tumor genome sequencing assays, individually or in combination with each other, have a number of potential applications. This includes the identification of clinically useful prognostic and predictive biomarkers, and the development of increasingly precise diagnostics and targeted therapeutics for application in personalized medicine, driven by the molecular profile of each individual disease and patient [29].

However, despite the rapid progression of *NGS*-technologies and their interesting different applications, several challenges remain before these can be incorporated into clinical practice. Firstly, *NGS* generates a huge amount of data that are not always easy to interpret. Sophisticated and expensive software, as well as bioinformatics algorithms, are required for the functional naming of each genomic alteration, but the reproducibility and accessibility of these algorithms still needs to be enhanced to allow the output data to be presented in a transparent, reproducible and understandable manner [30]. Moreover, overall tumor complexity makes the interpretation of sequencing output data even more difficult. In particular, intratumor heterogeneity can result in underestimation of the tumor genomics landscape based on data from single tumor biopsy samples [31]. Circulating tumor DNA (ctDNA) and cell-free circulating DNA (cfDNA) may represent an alternative promising source for accessing the tumor genome, offering the possibility of non-invasive mutational assessment and reducing heterogeneity-related biases compared with single-site biopsies [32]. It may also enhance the understanding of clonal changes during treatment, and being useful for patient selection and dynamic monitoring of the response to targeted drug therapy over time [33].

Another limitation is the availability, quantity and quality of specimens for sequencing. Indeed, most *NGS* platforms have library preparations optimized for a specific DNA quantity and quality easily obtained from fresh-frozen (FF) samples. However, high-purity FF specimens are often not available. From a research point of view, this limits the size of samples analyzed and thus the statistical power of the studies, and from a clinical point of view the implementation of a sequencing workflow in clinical laboratories is reduced. Many efforts are being made to optimize sequencing protocols on low-quality DNA, derived from formalin-fixed paraffin-embedded (FFEP) specimens, facilitating the widespread application of *NGS* technologies both in research and clinical settings [34, 35].

Finally, there are also ethical issues to be considered and investigated. For example, should patients and family members be informed about incidental findings of novel variants that may be of clinical significance, especially those related to inherited susceptibility to cancer or to other diseases [36, 37]? Indeed, once sequencing data have been mapped, the tumor DNA sequence should be compared with the germ-line DNA sequence from the same patient in order to identify the somatic cancer-specific variants only. During this process, incidental variants in protein-coding genes, including some associated with unrecognized disease, future disease risks, drug response, carrier status, and variants of uncertain significance can be found.

In conclusion, *NGS* technologies have certainly shed light on many dark areas of cancer molecular biology and represent a milestone in recent oncology research. However, there are still many issues to be resolved before the application of this approach in clinical practice can be considered, and to shed light on the many areas that remain in the shadows.

Acknowledgments

The authors thank Nicola Ryan, an independent medical writer, who provided native English editing and journal styling on behalf of HPS. This editorial assistance was funded by PharmaMar, Spain.

Conflicts of Interest

The Authors declare there are no conflicts of interest in relation to this article.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144(5):646-74.
2. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. Nature 2013;502(7471):333-9.
3. Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application. Biol Proced Online 2013;15(1):4.
4. Ojesina AI, Lichtenstein L, Freeman SS, et al. Landscape of genomic alterations in cervical carcinomas. Nature 2014;506(7488):371-5.
5. Nannini M, Astolfi A, Urbini M, et al. Integrated genomic study of quadruple-WT GIST (KIT/PDGFRA/SDH/RAS pathway wild-type GIST). BMC Cancer 2014;14:685.
6. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 2012;486(7403):405-9.
7. Wang L, Tsutsumi S, Kawaguchi T, et al. Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. Genome Res 2012;22(2):208-19.
8. Totoki Y, Tatsuno K, Yamamoto S, et al. High-resolution characterization of a hepatocellular carcinoma genome. Nat Genet 2011;43(5):464-9.
9. Berger MF, Hodis E, Heffernan TP, et al. Melanoma genome sequencing reveals frequent PREX2 mutations. Nature 2012;485(7399):502-6.
10. Agrawal N, Frederick MJ, Pickering CR, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 2011; 333(6046):1154-7.
11. Pantaleo MA, Astolfi A, Indio V, et al. SDHA loss-of-function mutations in KIT-PDGFRA wild-type gastrointestinal stromal tumors identified by massively parallel sequencing. J Natl Cancer Inst 2011;103(12):983-7.
12. Shah SP, Köbel M, Senz J, et al. Mutation of FOXL2 in granulosa-cell tumors of the ovary. N Engl J Med 2009;360(26): 2719-29.
13. Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med 2009;361(11):1058-66.
14. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 2008;456(7218):66-72.
15. Campbell PJ, Stephens PJ, Pleasance ED, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet 2008;40(6):722-9.
16. Meynert AM, Ansari M, FitzPatrick DR, et al. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics 2014;15:247.
17. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013;31(11):1023-31.
18. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10(1):57-63.
19. Edgren H, Murumagi A, Kangaspeska S, et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. Genome Biol 2011;12(1):R6.
20. Zhou JB, Zhang T, Wang BF, et al. Identification of a novel gene fusion RNF213-SLC26A11 in chronic myeloid leukemia by RNA-Seq. Mol Med Rep 2013;7(2):591-7.
21. Qadir MA, Zhan SH, Kwok B, et al. ChildSeq-RNA: A next-generation sequencing-based diagnostic assay to identify known fusion transcripts in childhood sarcomas. J Mol Diagn 2014;16(3):361-70.
22. Zhao X, Wang A, Walter V, et al. Combined targeted DNA sequencing in non-small cell lung cancer (NSCLC) using UNCseq and NGScopy, and RNA sequencing using UNCqeR for the detection of genetic aberrations in NSCLC. PLoS One 2015;10(6):e0129280.
23. Shiraishi Y, Fujimoto A, Furuta M, et al. Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. PLoS One 2014;9(12):e114263.
24. Wilkerson MD, Cabanski CR, Sun W, et al. Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. Nucleic Acids Res 2014;42(13):e107.
25. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 2015;47(3):199-208.
26. White NM, Cabanski CR, Silva-Fisher JM, et al. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. Genome Biol 2014;15(8):429.
27. Cui W, Qian Y, Zhou X, et al. Discovery and characterization of long intergenic non-coding RNAs (lincRNA) module biomarkers in prostate cancer: an integrative analysis of RNA-Seq data. BMC Genomics 2015;16 Suppl 7:S3.
28. Ding X, Zhu L, Ji T, et al. Long intergenic non-coding RNAs (LincRNAs) identified by RNA-seq in breast cancer. PLoS One 2014;9(8):e103270.
29. Damodaran S, Berger MF, Roychowdhury S. Clinical tumor sequencing: opportunities and challenges for precision cancer medicine. Am Soc Clin Oncol Educ Book 2015;35:e175-82.
30. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet 2012;13(9):667-72.
31. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366(10):883-92.
32. Lebofsky R, Decraene C, Bernard V, et al. Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types. Mol Oncol 2015;9(4):783-90.
33. Frenel JS, Carreira S, Goodall J, et al. Serial next generation sequencing of circulating cell free DNA evaluating tumour clone response to molecularly targeted drug administration. Clin Cancer Res 2015;21(20):4586-96.
34. Hedegaard J, Thorsen K, Lund MK, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samplesof human cancer and normal tissue. PLoS One 2014;9(5):e98187.
35. Van Allen EM, Wagle N, Stojanov P, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat Med 2014;20(6):682-8.
36. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013;15(7):565-74.
37. Catenacci DV, Amico AL, Nielsen SM, et al. Tumor genome analysis includes germline genome: are we ready for surprises? Int J Cancer 2015;136(7):1559-67.