University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

Health & Biomedical Sciences Faculty Publications and Presentations

College of Health Professions

3-17-2017

Applications of Next-Generation Sequencing in Cancer Research and Molecular Diagnosis

KeSheng Wang
The University of Texas Rio Grande Valley

Chun Xu
The University of Texas Rio Grande Valley

Follow this and additional works at: https://scholarworks.utrgv.edu/hbs_fac

Part of the Diseases Commons

Recommended Citation

Review Article   Open Access

Applications of Next-Generation Sequencing in Cancer Research and Molecular Diagnosis

Wang K[1]* and Xu C[2]*

[1]Department of Biostatistics and Epidemiology, College of Public Health, East Tennessee State University, Johnson City, TN, USA

[2]Department of Health and Biomedical Sciences, College of Health Affairs, University of Texas Rio Grande Valley, Brownsville, TX, USA

*Corresponding authors: Kesheng Wang, Department of Biostatistics and Epidemiology, College of Public Health, East Tennessee State University, PO Box 70259, Lamb Hall, Johnson City, TN 37614-1700, USA, Tel: 1 423 439 4481; Fax: 1 423 439 4606; E-mail: wangk@etsu.edu and Chun Xu, Department of Health and Biomedical Sciences, College of Health Affairs, University of Texas Rio Grande Valley, Brownsville, Texas, USA, Tel: 956 882 4193; E-mail: chun.xu@utrgv.edu

Abstract

Next-generation sequencing (NGS) technologies including DNA sequencing and RNA sequencing provide "omics" approaches to reveal genomic, transcriptomic, and epigenomic landscapes of individual cancers. A variety of genomic aberrations can be screened simultaneously, such as common and rare variants, structural variations (e.g. insertions and deletions), copy-number variation, and fusion transcripts. NGS technologies together with bioinformatics analysis, which expand our knowledge, are increasingly used to simultaneously analyze multiple genes in a cost and time-effective manner and have been applied in analyzing clinical cancer samples and offering NGS-based molecular diagnosis. Therefore, NGS is increasingly valuable as a tool for diagnosis for a number of cancers. Here we briefly introduce NGS technologies and summarize the recent applications in cancer research and molecular diagnosis in breast and prostate cancers.

Introduction

In the United States (US), more than 1.69 million new cancer cases were diagnosed in 2016 [1]; while the global incidence of cancer was about 14 million new cases in 2012 and an annual 19.3 million cases will be expected by 2025 [2]. Family, twin, and adoption studies have indicated that genetic and environmental factors and their interactions contribute to the development of most cancers including breast, cervix, colon, endocrine glands, prostate, testis and thyroid cancers, with a heritability ranging from 20%-60% [3-6].

Genome-wide association study (GWAS) is a screening procedure to identify the location of pathogenically relevant variations and detects many common variants with small effects for cancers and other complex diseases [6,7]. It is estimated that more than 100 common variants with low risk may contribute to cancer susceptibility and other complex diseases [6,8]; on the other hand, it has been suggested that multiple rare variants may underlie susceptibility to common diseases/traits including cancers; whereas the allelic architecture of complex diseases/traits may be due to a combination of multiple common and rare variants [9,10]. However, GWAS using tagSNPs are underpowered for detecting associations with rare variants, although some rare variants and haplotypes were identified; whereas rare and potentially deleterious variants may not be detected by GWAS [11].

Next-generation sequencing (NGS) technology including DNA sequencing and RNA sequencing, also known as massively parallel sequencing, is increasingly used to detect sequence variations and has provided abundant genetic markers including common and rare variants and has been applied in the analysis of clinical cancer samples including NGS-based molecular diagnosis [12-16]. NGS technologies provide "omics" approaches to reveal genomic, transcriptomic, and epigenomic landscapes of individual cancers [14,17]. The human genomes obtain from the NGS technology have been a leap forward in the fundamental understanding of human genetic variation for a number of human diseases, treatment responses, including the genetic variation in cancers.

NGS Applications in Cancer Research

DNA sequencing includes whole-genome sequencing (WGS), whole-exome sequencing (WES), and targeted sequencing [14-16]. WGS allows sequencing of the entire cancer genomes (e.g., genomes of the patients with cancers and/or cancer tissues), which can be used to detect all types of somatic/germline including common and rare variants, nucleotide substitutions, insertions and deletions, copy number variations (CNVs), chromosomal rearrangements, as well as analysis of the non-coding regions [7,15-19]. Three years ago, a study reports that the sequencing of the genomes of all of the cancers in the US is now theoretically feasible [20]. Currently, WGS holds great potential for the rapid diagnosis for a number of cancers including breast cancer (BC) [21]. WES focuses on the coding regions (exons) of a genome, typically about 2% of the human genome, to discover rare or common variants associated with a disorder or phenotype; therefore, WES can be used to determine the sequence of the regulatory regions, such as promoters and non–protein-coding regions, defining the functional parts of the genome [15,18]; however, WES is not ideal for the identification of CNVs and other structural alterations in the genome [18]. Therefore, further targeted sequencing, focusing on a selection of genes of interest for a specific disease identify by WES, could be more accurate and accessible in terms of time and cost for clinical applications for more laboratories [7,15,19]. WES promises to

accelerate discovery of genetic causes and contributors of disease in both the research and clinical settings [22]. WES also provides more comprehensive in-depth exome sequencing to identify exact pathogenic mutations for the diseased with a low frequency of false-positive signals, which is a cost effective method (<$400-600/sample) compared with whole genome sequencing (WGS) ($3000-5000/sample) depending on the coverage [23]. In addition, precision cancer medicine, the use of genetic/genomic profiles (e.g. NGS information) of patient tumors at the point-of-care to inform treatment decisions, is rapidly changing treatment strategies across cancer types. With the knowledge of precision cancer medicine, we are able to increase treatment efficacy, reduce toxicity, and therefore decrease overall cost of cancer treatment for both individual family and society.

RNA sequencing (RNA-Seq) is to sequence the total RNA of the cell to obtain information regarding mutations/single-nucleotide polymorphisms (SNPs), splice variants, levels of gene expression, gene fusions, genomic rearrangements, allele-specific expression, posttranscriptional modifications, microRNAs, small and long noncoding RNAs [12,16-18]. Gene expression, either alone or in combination with mutational data, can also be used to investigate spatial and temporal tumor heterogeneity [24].

The development of NGS technologies have made large-scale projects such as The Cancer Genome Atlas (TCGA) [25] and the International Cancer Genome Consortium (ICGC) [26] feasible, and provided multi-platform data for thousands of tumors from a variety of cancer types and subtypes and integrative analyses of genomic, transcriptomic and epigenomic data and increase our understanding of cancer biology [17,25,26]. For example, WES and WGS studies have identified new high and moderate-risk genes in several types of cancers, such as the pancreatic cancer susceptibility genes PALB2 and ATM [27], and the hereditary colorectal cancer moderate-risk genes POLD1 and POLE [28].

Non-coding RNA species such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) may also play important roles in a variety of cellular processes, and have been shown to be widely dysregulated in cancer [29]. MiRNAs can regulate the expression of target genes and act as tumor suppressors or oncogenes and exert widespread gene- and pathway-level effects [30]. Long non-coding RNAs (lncRNAs) may play an important role in oncogenesis and cancer pathology such as guiding the site specificity of chromatin-modifying complexes or acting as regulators of protein signalling pathways involved in carcinogenesis [17].

Epigenome alterations were also seen in a number of cancers such as BC [31]. Among the epigenetic modifications, DNA methylation is well documented and well-studied in cancer. Whole-genome bisulphite sequencing (WGBS) or reduced representation bisulfite sequencing (RRBS) has been used to identify methylated cytosines, especially in the CpG islands, which are interspersed in the promoter and other regulatory regions of a gene. Such epigenetic regulation mediated through DNA methylation modulates the expression of certain genes [16-18].

NGS Applications in Cancer Diagnosis

Although NGS is extensively used for cancer research field, lately NGS technology has been applied in NGS-based cancer molecular diagnosis [14-16,32]. Early detection and diagnosis will allow effectively treatments, while NGS technology allows the simultaneous sequencing of a large number of target genes and provides rich early

diagnostic markers to develop NGS-based molecular diagnosis [13,15-17,33]. The advent of NGS has opened up many new frontiers. For example, large-scale projects such as TCGA and the ICGC have already made available data from thousands of tumors across major cancer types, and have helped us refine cancer classification systems and examine the interplay between DNA mutations, RNA expression, and epigenomic patterns, obtaining a comprehensive overview of cancer cells and providing enhanced diagnostic, prognostic, and therapeutic criteria [14,17].

Targeted genetic tests are currently used as diagnostic and prognostic tools in clinical oncology, and more extensive genomic tests seem likely to come into regular use in the near future [34,35]; while WES currently finds the most use in clinical diagnostics because it covers more than 95% of the exons, which contains 85% of disease-causing mutations [36]. WGS can be used to compare tumor progression, treatment efficacy, the mechanisms associated with resistance development; meanwhile, these can be the source for the future, as new discoveries in clinically actionable alterations are made; nevertheless, WGS incurs a high sequencing cost and computational burden due to the amount of data produced, and alternate approaches such as WES are also used for certain applications [15,18]. For example, targeted cancer panels are advantageous due to their low cost and relatively simple interpretability, and many exist both for specific cancers, such as prostate cancer (PC) [37].

Recent advances in NGS technology have improved the understanding of PC biology and clinical variability. Particularly, DNA-Seq, RNA-Seq, chromatin immunoprecipitation-Seq, and methyl-Seq experiments have identified some new recurrent alterations in PC (e.g. TMPRSS2-ERG translocation, ATM, SPOP and CHD1 mutations and chromoplexy) and have better elucidated the major pathways affecting prostate tumorigenesis, which are the AR-signaling, PI3K/PTEN/AKT, RB1, TP53 loss/mutations and RTK–Ras–MAPK pathways [15,18,38]. However, at present, the study of multiple genetic alterations in PC is not suggested for routine diagnostic purposes; due to its highly heterogeneous nature of genetics and phenotypes, PC continues to pose a tremendous challenge in terms of diagnosis and prognosis [15,18]. Moreover, although ncRNAs are considered as exciting new diagnostic, prognostic, and therapeutic tools, additional work is needed to characterize the RNA species, their functions, and their applicability to clinical practice in PC [39]. In addition, a recent study found that stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive PC using prostate epithelial transcriptomes (a type of deep RNA sequencing technology) [40].

Previous studies have shown that about 30% of BC cases are caused by BRCA1 and BRCA2 mutations. Genetic test using BRCA1 and BRCA2 mutations has been recommended; however, other genes such as ATM, CHEK2, PALB2 and TP53, have been shown to confer high BC risk [15,16]. For example, Lin et al developed a multiple gene sequencing panel using the NGS, which contained 68 genes including mutations in BRCA1 and BRCA2 genes, ATM, TP53; where the genes in the panel had cancer risk association for patients with early-onset or familial BC [33]. The multiple genes sequencing using the NGS is an effective method to increase detected rate of high-risk cases [33]. Using DNA methylation and miRNA profiles, a recent study reports that DNA methylation contributes to deregulation of 12 cancer-associated miRNAs and BC progression [41]. The authors also found a strong association between hypermethylation of MIR-127 and MIR-125b-1 and BC progression, particularly metastasis and concluded that

MIR-127 and MIR-125b-1 hypermethylation can be potential biomarkers of BC metastasis. Moreover, a recent study using NGS technology combined with protein expression identify PI3K pathway aberrations are among the most common in cancers (including BC and PC) by examining over 19784 consecutive tumor samples (>40 cancer types) from thousands of clinicians in 60 countries [42]. The author concluded that patterns of biomarker coalterations involving HER2 and hormone receptors may be important for optimizing combination treatments across cancer types.

Concluding Remarks

NGS technology can be used to identify single nucleotide variants (SNVs), multi nucleotide variants (MNVs), structural variations (SVs), CNV, gene transcripts, epigenetic variations, and has led to improvements in cancer classification systems and molecular diagnosis [7,13,15-18]. Furthermore, the TCGA, ICGC and other groups have conducted studies with multiple cancer types to integrate genomic, transcriptomic, and epigenomic data from several cancer types and have helped us refine classification systems as well as our general understanding of cancer biology [15,17,25,26]. However, due to the high costs, the need of complex bioinformatics pipelines, of large storage capacity and the high number of variants, the clinical utility of mutation discovery throughout the complete exome or genome analysis is not convenient yet; furthermore, genetic diagnostics would also benefit from RNA-seq approaches to allow the detection of a higher number of mutations involved in disease onset and will aid in improving personalised care and medical management [13,15,16]. The effective applications of NGS in routine diagnostics rely on the high discovery rate of new markers, increased clarity regarding the validation and implementation of NGS tests and technological improvements [13]. In the future, to better understand the genetic etiology of cancers, to improve effective molecular diagnosis and to apply for genome information in precision medicine, genomic medicine and cancer clinic care, it will be useful to combine the results of GWAS, gene-gene and gene-environment interactions, with the recent rapid advances in NGS technologies including whole exome sequencing, transcriptome sequencing, and whole genome sequencing.

References

1. (2016) American cancer society, Cancer facts and figures.

2. Gulland A (2014) Global cancer prevalence is growing at "alarming pace," says WHO. BMJ 348: g1338.

3. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, et al. (2000) Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343: 78- 85.

4. Czene K, Lichtenstein P, Hemminki K (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the swedish family-Cancer database. Int J Cancer 99: 260- 266.

5. Hjelmborg JB, Scheike T, Holst K, Skytthe A, Penney KL, et al. (2014) The heritability of prostate cancer in the Nordic twin study of cancer. Cancer epidemiol biomarkers Prev 23: 2303-2310.

6. Gomy I, Diz Mdel P (2016) Hereditary cancer risk assessment: insights and perspectives for the next-generation sequencing era. Genet Mol Biol 39: 184-188.

7. Wang Q, Lu Q, Zhao H (2015) A review of study designs and statistical methods for genomic epidemiology studies using next generation sequencing. Front Genet 6: 149.

8. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9: 356-369.

9. Iyengar SK, Elston RC (2007) The genetic basis of complex traits: Rare variants or "common gene, common disease"? methods. Mol Biol 376: 71-84.

10. Barc J, Koopmann TT (2011) Genome-wide association studies: providers of candidate genes for identification of rare variants? Europace 13: 911-912.

11. Wang KS, Zuo L, Luo X (2014) Genetic factors for alcohol dependence and schizophrenia: Common and rare variants. Austin J Drug Abuse Addict 1: id1001.

12. Li W, Zhao K, Kirberger M, Liao W, Yan Y (2015) Next generation sequencing technologies in cancer diagnostics and therapeutics: A mini review. Cell Mol Biol (Noisy-le-grand) 61: 91-102.

13. Luthra R, Chen H, Roy-Chowdhuri S, Singh RR (2015) Next-generation sequencing in clinical molecular diagnostics of cancer: advantages and challenges. Cancers (Basel) 7: 2023-2036.

14. Corless CL (2016) Next-generation sequencing in cancer diagnostics. J Mol Diagn 18: 813-816.

15. Serratì S, De Summa S, Pilato B, Petriella D, Lacalamita R, et al. (2016) Next-generation sequencing: Advances and applications in cancer diagnosis. Onco Targets Ther 9: 7355-7365.

16. Kamps R, Brandão RD, Bosch BJ, Paulussen AD, Xanthoulea S, et al. (2017) Next-generation sequencing in oncology: Genetic diagnosis, risk prediction and cancer classification. Int J Mol Sci 18.

17. LeBlanc VG, Marra MA (2015) Next-generation sequencing approaches in cancer: where have they brought us and where will they take us? Cancers (Basel) 7: 1925-1958.

18. Yadav SS, Li J, Lavery HJ, Yadav KK, Tewari AK (2015) Next-generation sequencing technology in prostate cancer diagnosis, prognosis, and personalized treatment. Urol Oncol 33: 267.

19. Suhaimi SS, Ab Mutalib NS, Jamal R (2016) Understanding molecular landscape of endometrial cancer through next generation sequencing: What we have learned so far? Front Pharmacol 7: 409.

20. Kricka LJ, Polsky TG, Park JY, Fortina P (2015) The future of laboratory medicine — A 2014 perspective. Clin Chim Acta 438: 284-303.

21. Desmedt C, Yates L, Kulka J (2016) Catalog of genetic progression of human cancers: Breast cancer. Cancer Metastasis Rev 35: 49-62.

22. Singleton AB (2011) Exome sequencing: A transformative technology. Lancet Neurol 10: 942-946.

23. Bras J, Guerreiro R, Hardy J (2012) Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. Nat Rev Neurosci 13: 453-464.

24. Parker NR, Khong P, Parkinson JF, Howell VM, Wheeler HR (2015) Molecular heterogeneity in glioblastoma: Potential clinical implications. Front Oncol 5: 51-59.

25. (2017) The Cancer genome atlas.

26. (2017) International cancer genome consortium.

27. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, et al. (2009) Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science 324: 217.

28. Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, et al. (2013) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 45: 136-144.

29. Esteller M (2011) Non-coding RNAs in human disease. Nat Rev Genet 12: 861–874.

30. Farazi TA, Hoell JI, Morozov P, Tuschl T (2013) MicroRNAs in human cancer. Adv Exp Med Biol 774: 1-20.

31. Walsh L, Gallagher WM, O'Connor DP, Ní Chonghaile T (2016) Diagnostic and therapeutic implications of histone epigenetic modulators in breast cancer. Expert Rev Mol Diagn 16: 541-551.

32. Gagan J, Van Allen EM (2015) Next-generation sequencing to guide cancer therapy. Genome Med 7: 80.

33. Lin PH, Kuo WH, Huang AC, Lu YS, Lin CH, et al. (2016) Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. Oncotarget 7: 8310-8320.

34. Johansen Taber KA, Dickinson BD, Wilson M (2014) The promise and challenges of next-generation genome sequencing for clinical care. JAMA Intern Med 174: 275-280.

35. 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.

36. Rabbani B, Tekin M, Mahdieh N (2014) The promise of whole-exome sequencing in medical genetics. J Hum Genet 59: 5-15.

37. Leyten GH, Hessels D, Smit F, Jannink S, de Jong H, et al. (2015) Identification of a candidate gene panel for the early diagnosis of prostate cancer. Clin Cancer Res 21: 3061-3070.

38. Chen J, Zhang D, Yan W, Yang D, Shen B (2013) Translational bioinformatics for diagnostic and prognostic prediction of prostate cancer in the next-generation sequencing era. Biomed Res Int 2013: 901578.

39. Ramalho-Carvalho J, Fromm B, Henrique R, Jerónimo C (2016) Deciphering the function of non-coding RNAs in prostate cancer. Cancer Metastasis Rev 35: 235-62.

40. Zhang D, Park D, Zhong Y, Lu Y, Rycaj K, et al. (2016) Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. Nat Commun 7: 10798.

41. Pronina IV, Loginov VI, Burdennyy AM, Fridman MV, Senchenko VN, et al. (2017) DNA methylation contributes to deregulation of 12 cancer-associated microRNAs and breast cancer progression. Gene 604: 1-8.

42. Millis SZ, Ikeda S, Reddy S, Gatalica Z, Kurzrock R (2016) Landscape of phosphatidylinositol-3-kinase pathway alterations across 19784 diverse solid tumors. JAMA Oncol 2: 1565-1573.