Sequencing approaches in cancer treatment

D. Sekar*, K. Thirugnanasambantham*[†], V. I. Hairul Islam* and S. Saravanan*[‡]

*Pondicherry Centre for Biological Sciences, Pondicherry, 605 005, India, [†]State Bio-control Laboratory, Perunthalaivar Kamaraj Krishi Vigyan Kendra (PKKV) Kurumbapet, Puducherry, 605 009, India and [‡]Division of Ethnopharmacology, Entomology Research Institute, Loyola College, Chennai, 600034, India

Received 18 April 2014; revision accepted 23 May 2014

Abstract

Use of sequencing approaches is an important aspect in the field of cancer genomics, where next-generation sequencing has already been utilized for targeting oncogenes or tumour-suppressor genes, that can be sequenced in a short time period. Alterations such as point mutations, insertions/deletions, copy number alterations, chromosomal rearrangements and epigenetic changes are encountered in cancer cell genomes, and application of various NGS technologies in cancer research will encounter such modifications. Rapid advancement in technology has led to exponential growth in the field of genomic analysis. The \$1000 Genome Project (in which the goal is to sequence an entire human genome for \$1000), and deep sequencing techniques (which have greater accuracy and provide a more complete analysis of the genome), are examples of rapid advancements in the field of cancer genomics. In this mini review, we explore sequencing techniques, correlating their importance in cancer therapy and treatment.

Introduction

Ascertaining the sequence of nucleotide or amino acid residues in a nucleic acid or protein is called sequencing, and thorough such analysis provides us with basic information of all sequences of nucleotides or amino acids present. There are many sequencing techniques available to analyse a gene or whole genome. However, next-generation sequencing or high-throughput technologies can produce thousands or millions of sequences

at lower cost in a short period, than previously (1). Currently, next-generation sequencing is being applied to cancer genomics study in various ways:

- genome-based sequencing (DNA-Seq), provides information on sequence variation, mutations, chromosomal rearrangement and copy number variation;
- transcriptome-based sequencing (RNA-Seq) provides quantitative information on transcribed regions (total RNA, mRNA, or noncoding RNA);
- interactome-based sequencing (ChIP-Seq), yielding information on protein binding sequences and histone modification;
- methylome-based sequencing (Methl-seq), yielding quantitative information on DNA methylation and chromatin conformation (2).

The \$1000 Genome Project, has the main aim to sequence an entire human genome for \$1000, and deep sequencing techniques (which have greater accuracy and provide a more complete analysis of the genome) (3), are examples of rapid advancements in the field of cancer genomics.

Deep sequencing

Deep sequencing is a next-generation approach in which nucleotide sequence or fragment can be repeatedly sequenced in a very short time; this delivers greatly increased sensitivity and accuracy. The approach brings new insights to the role of regulation of RNA complexity – the sum of unique isoforms of RNA in a cell, mRNA variants and non-coding RNAs including microRNAs (miRNAs), which play a role in generating organismal complexity from a relatively small number of genes (4). Deep sequencing has a number of advantages over other methods. It is driven largely by high depth of coverage for any library of nucleic acids. This allows estimates of alternative splicing and untranslated region utilization and also provides direct access to a sequence; it can be used on species for which a full-genome

Correspondence: Durairaj Sekar, Pondicherry Centre for Biological Sciences, Jawahar Nagar, Pondicherry 605 005, India. Tel.: +91 9791294770; Fax: +91 4132203530; E-mails: duraimku@gmail.com and pcbsresearch@gmail.com

sequence is not yet available. Moreover, junctions between exons can be assayed by deep sequencing without prior knowledge of the gene structure. RNA editing events can be detected, and knowledge of polymorphisms can provide direct measurement of allele-specific expression (5).

One recent report illustrates genetic characterization of 50 gastric adenocarcinoma samples. Affymetrix SNP arrays and Illumina mRNA expression arrays were used, as well as Illumina deep sequencing of coding regions, of 384 genes belonging to various pathways known to be altered in cancers (6). These techniques, thus, offer a more complete picture of a tumour's mutagenic profile and are even more informative in identifying sensitivity and resistance biomarkers. For example, to characterize chimaeric RNAs enriched in human prostate cancer, deep sequencing is used to enable a profound survey of the complex human genome, leading to validation of 32 recurrent chimaeric RNAs inclusive of 27 novel ones. Importantly, one of these chimaeras has appeared to be highly enriched in malignancy, being expressed at significantly higher levels in human prostate cancers, but of only very low levels in non-tumour prostate (2). That there are more chimaeric RNAs in malignant than in matched benign samples raises the possibility of using deep sequencing for characterization of chimaeric RNA (one of the molecular consequences of cancer), and deep sequencing can be used to interrogate protein–nucleic acid interactions (2). For example, it is useful to analyse promoter-binding proteins related to oncogenes or tumour suppressor genes where short DNA fragments isolated by chromatin-immunoprecipitation can be used to generate libraries, and provide interacting sequences as well as abundance (7).

Deep sequencing is used in the detection of different genotypes, DNA copy number variation and mRNA expression profiling, and is flexible for heterogeneous clinical samples. Ten somatic copy number alterations (SCNA) detection algorithms in both simulated and primary tumour deep sequencing data have been evaluated for applicability of exome sequencing data for SCNA detection. This revealed existence of clear differences in sensitivity and specificity between the algorithms (8). SCNA detection algorithms are able to identify most complex chromosomal alterations, and exome sequencing data are suitable for SCNA detection (8). Figure 1 is an outline schematic representation concerning deep sequencing technique in cancer research.

Deep sequencing in B-cell diseases

Information of deep sequencing can be used to elucidate small RNA transcriptomes of normal human B cells

Figure 1. Outline schematic representation about the deep sequencing technique in cancer research.

and human B-cell tumours (9). One recent investigation analysed more than 328, 934 149 separate sequences (6 billion bases) to identify expression of small non-coding RNAs including miRNA, transferRNA (tRNA), piwi-interacting RNA (piRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA). In all 333 mature miRNAs were identified, more than twice as many as previously reported in a single tissue type. 286 novel miRNA candidates that satisfied structural and expression criteria used to identify known miRNA (9) were seen, suggesting that by using deep sequencing, a comprehensive framework that spans the spectrum of identifying new miRNAs can be developed, and validating them, before applying real-time PCR to provide a clear path to clinical utility (9). Recent deep sequencing data provide conclusive evidence for an overlap of V_H but not V_L germlines between myelin basic protein and latent membrane protein 1 sublibraries. On this basis, it has been suggested that dual activity of auto antibodies to myelin basic protein and latent membrane protein 1 proteins is basically determined by variable regions of their heavy chains, while light chains appear to be responsible for fine-tuning of antibody specificity. This may be a hallmark of EBV-specific B-cell subpopulations involved in multiple sclerosis triggering (10).

Single cell exome sequencing

Exome sequencing is an efficient deep technique to selectively sequence coding regions (exons) of the genome; it is applied in analysing the exome to discover disease-related variations in them. Coupled with growing databases that contain known variants, exome

sequencing allows identification of genetic mutations and risk factors in families, and tumour samples that had previously been deemed insufficiently informative for genetic studies (11). Compared to other techniques, it is much quicker and cheaper to screen any number of exons in a tumour sample for identification of candidate genes and their variants.

Single-cell exome sequencing has been approved by prominent workers in the field of monogenic disorders and is increasingly employed as a diagnostic tool for specific genetic diseases, particularly in the context of disorders characterized by significant genetic and phenotypic heterogeneity (12). Recently, exome sequencing studies on melanoma cell lines (13,14) have identified multiple mutations in kinase genes that had not previously been linked to melanoma, including in *MAP3K5*, *MAP3K9*, *MAP2K1* and *MAP2K2*. Using exome sequencing, new investigation detected activating *MAP2K1* P124S mutation in A2058, known to result in demonstration of ERK1/2 phosphorylation in formalin-fixed tumour samples (15). Analysis of large-scale melanoma exome data discovered six novel melanoma genes (*PPP6C*, *RAC1*, *SNX31*, *TACCI*, *STK19* and *ARID2*), three of which – *RAC1*, *PPP6C* and *STK19* – harboured recurrent and potentially targetable mutations. Integration with chromosomal copy number data contextualized the landscape of driver mutations, providing oncogenic insights into BRAF- and NRAS-driven melanoma as well as those without known *NRAS/BRAF* mutations. This also clarified a mutational basis for RB and p53 pathway deregulation in melanomas (16). Exome sequencing uses DNA-enrichment methods and massively parallel nucleotide sequencing, to comprehensively identify mutations throughout the genome. Recent studies on exon resequencing in 218 prostate cancer tumours identified multiple somatic alterations in the androgen receptor (*AR*) gene as well as its upstream regulators and downstream targets (17). Robbins *et al.* (18) used NGS-based exome sequencing in 8 metastatic prostate tumours and revealed novel somatic point mutations in genes including *MTOR*, *BRCA2*, *ARHGEF12* and *CHD5*. Other investigations also used whole-exome sequencing of metastatic tumours and high-grade primary carcinomas where somatic mutations in some important genes like *TP53*, *DLK2*, *GPC6* and *SDF4* (19) were also observed.

Whole-genome sequencing

Whole-genome sequencing is a next-generation technique that determines the complete DNA sequence of a genome in an organism on a single occasion. The most significant impact of complete genome sequencing on cancer genomics is its ability to re-sequence, analyse

and compare matched tumour and normal genomes of a single patient (20). It provides clinical information in addition to its association between specific genetic variants and diseases. It is also used to detect germ line susceptibility loci, single-nucleotide polymorphism and small indel mutations, copy number alterations and structural variants (20). In 2008, the \$1000 Genome Project was initiated to provide deep characterization of human genome sequence variation as a foundation for investigating relationships between genotype and phenotype. Specifically, the goal of this project was to characterize over 95% of variants that are in genomic regions accessible to current high throughput sequencing technologies, that have allele frequency of 1% or higher in each of five major population groups. It includes populations in or with ancestry from Europe, East Asia, South Asia, West Africa and the Americas (21).

Whole-genome sequencing produces large amount of data which can be stored electronically and require bioinformatic tools to analyse. Cost-effective high-throughput sequencing necessary to accomplish full-genome sequencing is possible by using nanopore technology, where nucleotides are identified by different degrees and characteristics, when they are passed through nanopores (22). Recent studies have reported a new single-molecule assay for detection of DNA methylation, using solid-state nanopores. Nanopore-based methylation sensitive assay circumvents the need for bisulphite conversion, fluorescence labelling and PCR, and could therefore prove very useful in studying the role of epigenetics in human disease (23).

DNA nanoball technology is a further high throughput technology used to analyse entire genomic sequences, by rolling circle replication to amplify small fragments of DNA into DNA nanoballs (24). These next-generation platforms can produce data to characterize gene expression, methylation, histone packaging, transcription factor and other regulatory protein-binding positions, with which, data sets can be built to comprehensively characterize a broad spectrum of genomic alterations among sets of tumour samples (20).

To understand molecular changes leading to the pathogenesis of chronic lymphocytic leukaemia (CLL), whole-genome sequencing has been used in four cases of CLL, and has allowed identification of 46 somatic mutations that potentially affect gene function. Further whole-genome sequencing analyses of these mutations in 363 patients with CLL identified four genes recurrently mutated: notch 1 (*NOTCH1*), exportin 1 (*XPO1*), myeloid differentiation primary response gene 88 (*MYD88*) and kelch-like 6 (*KLHL6*). Mutations in *MYD88* and *KLHL6* are predominant in cases of CLL with mutated immunoglobulin genes, whereas *NOTCH1*

and *XPO1* mutations are mainly detected in patients with unmutated immunoglobulins. Patterns of somatic mutation, supported by functional and clinical analyses, strongly indicate that recurrent *NOTCH1*, *MYD88* and *XPO1* mutations are oncogenic changes that contribute to clinical evolution of the disease (25). This study highlighted usefulness of the whole-genome sequencing approach for identification of clinically relevant mutations in cancer. Using whole-genome sequencing, recent investigations have provided detailed backgrounds of genomic alterations in localized prostate cancers (18,26).

Currently, there are very many NGS platforms available commercially; these include (i) Genome Analyser/HiSeq 2000/MiSeq from Illumina, SOLiD/PGM/Proton from Life Sciences; (ii) GS-FLX (454)/GS Junior from Roche; (iii) Heliscope from Helicos Biosciences; and (iv) SMRT offered by Pacific Biosciences (1,2). All these provide digital information on DNA/RNA sequences enabling discovery of genetic variation mutations at high resolution with low cost.

It is widely acknowledged that cell proliferation of cancers is caused by point mutations, insertions and deletions, copy number alterations, chromosomal rearrangements and epigenetic changes (6). Figure 2 illustrates detailed application of various NGS technologies in cancer research. These methods well examine all altered information remaining cryptic within a genome; a variety of NGS technology, applications, advantages and disadvantages in prostate cancer have been discussed by Chen *et al.* (2).

\$1000 Genome Project

The \$1000 Genome Project provides a new era of predictive and personalized medicine, where cost of full-genome sequencing of an individual or patient drops to

\$1000 each (3,27). In this paradigm, cost of determining an individual genome sequence would fall to a price of around \$1000, placing it firmly in the realm of advanced clinical diagnostic tests. As a result, determining a person's genomic sequence might ultimately become an important first step on entering a health insurance network or a health care. Numerous institutes and companies have highlighted the importance the \$1000 Genome Project and its potential consequences. In January 2013, revamped Archon Genomics X PRIZE presented by Medco held a \$10-million grand prize competition for the team to reach or come closest to reaching the \$1000 genome. This grand prize goes to the team able to sequence 100 human genomes within 30 days, to an accuracy of 1 error per 1 000 000 bases, 98% complete, identification of insertions, deletions and rearrangements, and a complete haplotype at audited total cost of \$1000 per genome (28).

Conclusion

This mini review has described the advent and widespread availability of next-generation sequencing in cancer research. However, more applications of next-generation techniques, beyond those covered here, are yet to come. For example, genome re-sequencing will probably be used to characterize promoter sequence and microRNAs in cancer genomics. Studies of this type will identify and catalogue genomic variation on a wide scale, from single-nucleotide polymorphisms to copy number variations in large sequence blocks (>1000 bases). Ultimately, re-sequencing studies will help better characterize, for example, ranges of normal variation in complex genomes such as the human genome, and aid our ability to comprehensively view genome variation in clinical studies. Epigenomic variation, as an extension

Figure 2. Application of high-throughput sequencing in cancer research.

of genome re-sequencing applications, also will be employed for analysis of genome-wide patterns of methylation and how these patterns change through the course of an organism's development, in the context of cancer treatment or therapy, and under various other influences. Perhaps, the most exciting possibility engendered by ability to use DNA sequencing to rapidly read out experimental results, is enhanced potential to combine results of different experiments – correlative analyses of genome-wide methylation, histone binding patterns and gene expression. Thus, high throughput sequencing analyses may be the way to begin unlocking the last secrets of the cell.

Acknowledgements

All the authors are thankful to the Pondicherry Centre for Biological centre (PCBS) for providing the necessary facility to carry out the work. Financial support as start-up loan from the State Bank of India (RASMECC), Pondicherry, India, to establish the institute is also gratefully acknowledged. KT is a recipient of young scientist grant (SB/FT/LS-382/2012), Department of Science and Technology, Government of India.

References

- Meldrum C, Doyle MA, Tothill RW (2011) Next-generation sequencing for cancer diagnostics: a practical perspective. *Clin. Biochem. Rev.* **32**, 177–195.
- Chen J, Zhang D, Yan W, Yang D, Shen B *et al.* (2013) Translational bioinformatics for diagnostic and prognostic prediction of prostate cancer in the next-generation sequencing era. *Biomed. Res. Int.* **2013**, 901578.
- Mardis ER (2006) Anticipating the 1,000 dollar genome. *Genome Biol.* **7**, 112.
- Guffanti A, Simchovitz A, Soreq H (2014) Emerging bioinformatics approaches for analysis of NGS-derived coding and non-coding RNAs in neurodegenerative diseases. *Front. Cell. Neurosci.* **8**, 89.
- Malone JH, Oliver B (2011) Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* **9**, 34.
- Holbrook JD *et al.* (2011) Deep sequencing of gastric carcinoma reveals somatic mutations relevant to personalized medicine. *J. Transl. Med.* **9**, 119.
- Mardis ER (2008) Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* **9**, 387–402.
- Alkods A, Louhimo R, Hautaniemi S (2014) Comparative analysis of methods for identifying somatic copy number alterations from deep sequencing data. *Brief Bioinform.* 2014 (Epub). DOI:10.1093/bib/bbu004.
- Jima DD *et al.* (2010) Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood* **116**, e118–e127.
- Lomakin YA *et al.* (2014) Heavy-light chain interrelations of MS-associated immunoglobulins probed by deep sequencing and rational variation. *Mol. Immunol.* 2014 (Epub). DOI: 10.1016/j.molimm.2014.01.013.
- Singleton AB (2011) Exome sequencing: a transformative technology. *Lancet Neurol.* **10**, 942–946.
- Ku CS, Cooper DN, Polychronakos C, Naidoo N, Wu M, Soong R (2012) Exome sequencing: dual role as a discovery and diagnostic tool. *Ann. Neurol.* **71**, 5–14.
- Nikolaev SI *et al.* (2012) Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nat. Genet.* **44**, 133–139.
- Stark MS *et al.* (2012) Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing. *Nat. Genet.* **44**, 165–169.
- Wong SQ *et al.* (2013) Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours. *Sci. Rep.* **3**, 3494.
- Hodis E *et al.* (2012) A landscape of driver mutations in melanoma. *Cell* **150**, 251–263.
- Taylor BS *et al.* (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22.
- Robbins CM *et al.* (2011) Copy number and targeted mutational analysis reveals novel somatic events in metastatic prostate tumors. *Genome Res.* **21**, 47–55.
- Kumar A *et al.* (2011) Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc. Natl. Acad. Sci. USA* **108**, 17087–17092.
- Mardis ER, Wilson RK (2009) Cancer genome sequencing: a review. *Hum. Mol. Genet.* **18**, R163–R168.
- Hudson TJ *et al.* (2010) International network of cancer genome projects. *Nature* **464**, 993–998.
- Clarke J *et al.* (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270.
- Shim J *et al.* (2013) Detection and quantification of methylation in DNA using solid-state nanopores. *Sci. Rep.* **3**, 1389.
- Porreca GJ (2010) Genome sequencing on nanoballs. *Nat. Biotechnol.* **28**, 43–44.
- Puente XS *et al.* (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105.
- Beltran H *et al.* (2013) Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. *Eur. Urol.* **63**, 920–926.
- Service RF (2006) Gene sequencing. The race for the \$1000 genome. *Science* **311**, 1544–1546.
- Kedes L, Company G (2011) The new date, new format, new goals and new sponsor of the Archon Genomics X PRIZE competition. *Nat. Genet.* **43**, 1055–1058.