J Med Oncl Ther 2016; 1 (2): 76-80 *Journal of Medical Oncology and Therapeutics*

Next generation sequencing in clinical molecular diagnosis of cancer: A journey towards personalized diagnosis and therapeutics.

Tridip Chatterjee

Suraksha Genomics (R&D Division of Suraksha Diagnostic), Salt Lake, Kolkata, West Bengal, India.

Abstract

Most cancers are genetically complex and involve a series of patho-genomic DNA mutations or genetic hits. The order of development of these genetic hits is extremely diverse and this diversity in order of development of genomic mutations makes the diagnosis of most types of cancers very much challenging. So, each cancer needs to be treated as a single novel case. So, a new era has arrived when personalized diagnosis of cancer can be offered with development of massively parallel sequencing technology, popularly known as Next-Generation Sequencing. With this technology, it is quite possible to look into the substantial part of the genome of a cancer cell within a very short period of time and at an affordable cost.

Next Generation sequencing represents an effective way for capturing a substantially large amount of genetic information about a cancer. Most Next-Generation Sequencing technologies involve sequencing by synthesis, while other use sequencing by ligation. Each DNA fragment to be sequenced is actually bound to an array. DNA polymerase then adds labeled nucleotides sequentially to it. Then, a high-resolution camera captures the signal from each nucleotide, as it got integrated. The camera also takes notes on the spatial coordinates and time. Then the sequence at each spot can be inferred by a computer program/ software to generate a contiguous DNA sequence, which is termed as read. Multiple Next-Generation Sequencing technology uses different ways to capture the signal and make a read. Illumina GA/Miseq, HiSeq uses fluorescent chemistry, where as Life Technologies Ion PGM, Ion Proton uses semiconductor based chemistry to capture the signal. In nutshell, Next-Generation Sequencing has revolutionized the sequencing and cost has brought down to quite an affordable level, so that it can be implemented in clinical settings for regular diagnostic procedures in molecular pathology lab.

Keywords: Next generation sequencing, Molecular diagnosis, Cancer, Personalized diagnosis.

Accepted October 17, 2016

Introduction

The journey of a cell towards carcinogenesis is very unique and hence very interesting process to observe. Only for few cancer types like chronic myelogenous leukemia (CML) only single patho-genomic DNA mutation (often termed as genetic hit), is sufficient to take to the cell towards carcinogenesis. So, molecular detection of these types of malignancies is somehow easier and straight forward.

But, most of the cancers are involved with a series of patho-genomic DNA mutations or many genetic hits. For these types of cancer, which follows the multi-hit model, molecular detection becomes very difficult and challenging, because there is no specific order of development of the patho-genomic DNA mutations or genetic hits [1]. No one can predict which mutation will develop first, which second and which will develop at last, as the order of developing the patho-genomic mutations is highly diverse and differs from one another. This huge diversity in order of development of genomic mutations makes the diagnosis of most types of cancers very much challenging [2]. Most cancers are genetically complex and it is better to define it by the activation of specific signaling pathways, rather than relying upon a defined set of mutations [3]. So, each cancer needs to be treated as a single novel case. Any generalized strategy of diagnosis or prognosis will not be able to give exact information of the path of the journey of a cell towards carcinogenesis.

Projects for looking at various cancer genomes are inspired

by the success of Human Genome Project and increased affordability and reliability of different sequencing technologies and platforms. Affordable and reliable sequencing techniques have integrated the genome science with clinical diagnostic practice. So, we have moved into a new era of personalized diagnosis of cancers. Now-a-days it is quite possible to look into the substantial part of the genome of a cancer cell within a very short period of time, by the development of massively parallel sequencing technology, popularly known as Next-Generation Sequencing (NGS).

Next Generation Sequencing Technologies and Platforms

Frederick Sanger develop the di-deoxy or chain-termination sequencing (Sanger sequencing) in 1977. Since then, the sequencing technology has been continuously upgrading in very rapid space. Now, sequencing has arrived in the era of massively parallel sequencing technology, popularly known as 'Next Generation Sequencing'. All NGS technologies are basically based on the model of massively parallel, high-throughput sequencing. It was first successfully developed by the 454 sequencing platform (acquired by Roche) in 2005. Presently, the high throughput NGS platforms are represented by companies like Roche, Illumina, Life Technologies, etc. All of them have their own amplification technology for sequencing, detection technology and limitation of maximum output (Table 1). From the Table 1, it is quite evident that, the throughput of Roche 454 GS Junior (i.e., ~400 Mb/run) and 454 GS FLX (i.e., ~700 Mb/run) are substantially lower than that of the other platforms (i.e., >100 Gb). Though different NGS platforms follow different sequencing chemistries and throughputs, but all of them offer a sequencing cost at USD (0.07-10)/Mb sharply contrast to USD 2400/Mb by conventional Sanger sequencing [4].

The availability of compact low/medium throughput sequencers in the past few years, exemplified by the 454 GS Junior, MiSeq and Ion PGM launched by Roche, Illumina and Life Technologies (presently taken over by Thermo Fisher Scientific Corporation) respectively. Along with the high throughput NGS platforms, these bench-top sequencers, as mentioned above, offer a lower data throughput (from tens of MB to several GB) at low costs, which are well suited for clinical applications. More importantly, in terms of the sequencing turn-around time. For the some versions of Ion PGM versions, raw sequencing data are available within 3 hours, while high throughput NGS platforms usually take more than a week to complete a sequencing run. There is an increased amount of data evaluating the clinical applications of these bench-top sequencers for detecting mutations in human cancer genes with promising results [5-7] (Table 1).

Next Generation Sequencing in Cancer Diagnosis

As cancer is a genetic disease driven by heritable or somatic mutations, new DNA sequencing technologies will have significant impact on the detection, management and treatment of the disease. Next Generation sequencing is continuously enabling worldwide collaborative efforts, such as the International Genome Consortium [ICGC] [8] and The Cancer Genome Atlas [TCGA] project [9] to catalogue the genomic landscape of thousands of cancer genomes across many disease types. Several other studies contributing to the consortia have been published [10-12]. These discoveries from individual studies will ultimately lead to better understanding of pathogenesis of the disease, leading to a new era of molecular pathology and personalized medicine [13]. It is quite practical to imagine that every patient will sequence both their constitutional and cancer genomes to compare the changes in order to monitor disease progression, which will finally enable the clinician to perform an accurate molecular sub-typing of the disease and use molecularly guided therapies for accurate treatment.

Today the medical diagnosis has arrived to an era, when samples are no longer need to be handled differently

Table 1. *Different NGS platform and sequencing technologies*

Company	Platform	Sequencing	Amplification	Run time	Read length	Max. Output
Roche	454 GS Junior	SBS Pyro	emPCR	10 h	400 bp (SE, PE)	35 Mb
	454 GS FLX+	SBS Pyro	emPCR	10–23 h	700 bp (SE, PE)	700 Mb
Illumina	MiSeq	SBS RDT	Bridge PCR	4–27 h	36–151 bp (SE, PE)	>1 Gb
	GAIIx	SBS RDT	Bridge PCR	2–14 days	36–151 bp (SE, PE)	≤ 95 Gb
	HiSeq 1000	SBS RDT	Bridge PCR	1.5–8.5 days	36–101 bp (SE, PE)	≤ 150 Gb
	HiScanSQ	SBS RDT	Bridge PCR	1.5–8.5 days	100 bp (SE, PE)	≤ 150 Gb
	HiSeq 2000	SBS RDT	Bridge PCR	2.5–11 days	36–101 bp (SE, PE)	≤ 300 Gb
Life Technologies/ Thermo Fisher Scientific Corporation	Ion Torrent	SBS H+	emPCR	2 h+	316+318 chip >100 bp (SE)	316–>100 Mb 318–>1 Gb
	5500	SBL	emPCR	2–7 days	35–75 bp (SE, PE)	77 Gb
	5500XL (4hp)	SBL	emPCR	2–7 days	35–75 bp (SE, PE)	155 Gb

Note: PE: Paired-End Read; SE: Single-End Read; bp: base pairs; Mb: Megabases; Gb: Gigabases; SBL: Sequencing-By-Ligation; em PCR: emulsion PCR; Pyro: Pyro-sequencing; SBS: Sequencing By Synthesis; RDT: Reverse Dye Terminator Chemistry; H+: Hydrogen Ion Detection Chemistry

from standard diagnostic specimens and recent advances have even enabled complex genomic data to be derived from a patient's peripheral blood. The concept of precision medicine goes hand in hand with an accurate understanding of the cancer genome as determined by NGS. Many molecular pathology laboratories are now thinking of bringing the sequencing platforms, for making the transition to NGS from conventional approach.

Clinical Relevance of NGS Data in Cancer

Once the set of alterations are indentified within the tumor of any patient, many cases will yield only a small set of clinically relevant events. Along with that, a long list of sequencing variants will also appear which will have uncertain significance. Here come the roles of bioinformaticians, who can develop algorithms for interpretation of NGS data that can automate the clinical relevance of the alterations, which will enable more rapid clinical interpretation of genomic sequencing data of cancers. If such approaches get matured, the laboratory will be able to come up with better tumor-specific markers and susceptibility sequences from NGS data, enabling probabilistic approaches for ranking genomic alterations related to different tumors by clinical relevance.

Furthermore, there are many databases that can be evaluated and accessed to test the clinical significance of mutations. The first level approach of analysis can be the variation, observed in any clinical tumor sample has been seen reported before in published papers. Understandably, the driver mutations are likely to recur in multiple tumor types and patients. The common databases are listed in Table 2.

Among the databases listed in Table 2 Catalog of Somatic Mutations in Man (COSMIC) [14,15] and TCGA (available for data exploration at multiple sites) [16,17] are more common. Information about cancer therapies and prognostic information about development of cancers can be found at a number of databases. MD Anderson's Cancer Institute host a database for Personalized Cancer Therapy [18,19]. Broad Institute's develop their own database TARGET [15,16]. Vanderbilt holds database, named 'My Cancer Genome' [20,21]. Other databases include TARGET [22,23], IntOGen [24,25], DGIdb [26,27] and CIViC [28,29]. Each database has specialties in their organizations of clinically relevant research derived information links to relevant primary literatures. In nutshell, if someone is having a sequence data (by NGS), derived from some tumor cells of patients, he is not in a clueless position today. There are enough literature support and databases packed with clinically relevant information, which can help to annotate any sequence data derived from cancer genome. There can be novel variations, but that will also add up in the future literature and update is ever continuing process.

Finally, NGS technologies, that require both germ-line and somatic testing (e.g. whole-genome and whole-exome sequencing), the American College of Medical Genetics has released the guidelines after clearly mentioning which variants should be reported to patients always, regardless of whether these are relevant to the presenting illness [30] or not. This detailed reporting of list of genomic variations has developed the need of through genetic knowledge of the consultant oncologists, as finally they need to interpret the diagnosis report. Since most of reported genes are also involved non-cancer-related syndromes, there is an increasing demand for oncologists, who are prepared to receive the results that turn up unexpected inherited genetic issues [31].

Discussion

There are different ways that NGS can help the clinicians for better combating with cancers. It starts from diagnosis to choice of proper therapeutics to each specific patient. Tumor subtypes, which were defined by morphologic criteria only just a few years ago, they are now defined by genetic mutations, either exclusively or inclusively. For example, a study by Honeyman et al. fibrolamellar hepatocellular carcinoma patients showed recurrent in-frame fusion between DNAJB1 and PRKACA [32].

Next, NGS can also help in choosing the appropriate 'targeted therapy' for any particular patient or type of cancer, as an increasing number of therapies are available today. So, choice of therapy should be based on data available for therapeutic outcome correlated with DNA level genomic mutations, based on DNA sequencing results, as listed in Table 3. Patients, who do not get the facility for proper detection of mutations, can never be given targeted drug. For them, clinicians have to choose drugs based on assumption, which may benefit the patient, but can actually be also harmed by inappropriately targeted

Table 2. *Databases for interpretation of somatic mutations in cancer*

Database	Host Institute	Based on	Reference
COSMIC	Sanger	Gene	[15]
cBioPortal	MSK	TCGA diseases	[17]
Personalized Cancer Therapy (PCT)	MD Anderson	Gene	[19]
My Cancer Genome	Vanderbilt	Disease	[20]
TARGET	BROAD	Gene	[22]
IntOGen	University Pompeu Fabra	Gene	[24]
DGIdb	Washington University	Drug/gene interaction	[26]
CIViC	Washington University	Variant	[28]

Table 3. *FDA-approved drugs for Cancers correlated with DNA mutation*

Drug	Cancer Type	DNA mutation
Tramatenib	Melanoma	*BRAF* V600E/K
Cetuximab	Colon cancer	*KRAS* codon 12, 13
Vemurafenib, Dabrafenib	Melanoma	*BRAF* V600E
Imatinib, Nilotinib, Dasatinib, Bosutinib	Chronic myelogenous leukemia	*BCR-ABL1* fusion
Ponatinib	Chronic myelogenous leukemia	*BCR-ABL1* fusion
		T315I resistance mutation
Erlotinib, Afatinib	Lung adenocarcinoma	*EGFR*
		Exon 19 deletions
		L858R
Olaparib	Ovarian cancer	BRCA1 and BRCA2 mutations
Crizotinib	Lung cancer	ALK gene fusions

therapies [33]. So, proper diagnosis plays a key role in choice of specific targeted therapies (Table 3).

NGS data can also benefit, when a patient stops responding to a specific targeted therapy, after developing known resistance mutations. In some cases instances, the resistant mutations may be limited to one or a few loci. For example, resistance to EGFR targeted therapies in cancer quite frequently involves a single point mutation and can actually be overcome by merely switching to another different agent [34]. However, glioblastoma can become resistant to EGFR targeted therapies via a more complicated epigenetic regulation [35].

If a patient has failed to respond in conventional therapy, NGS can be immensely helpful to identify and enroll them into the appropriate clinical trial. The US National Cancer Institute has recognized the potential of the NGS adequately followed by targeted therapeutic approach by setting up Molecular Analysis for Therapy Choice (MATCH) Program [36]. In this program, biopsies from tumors from 3000 patients will undergo NGS to identify individuals whose tumors have genetic abnormalities that may respond to some selected targeted drugs. Then, as many as 1000 patients will be assigned to one of the phase II trials. The patients will be assigned based on the genetic abnormality or mutations that are thought to be driving their cancer, not on their cancer types [37]. In nutshell, NGS allows the clinicians to see more complete overview of tumor dynamics and decide accurate personalized therapeutics with lesser side effects, as compare to conventional therapy.

References

1. Jesse JS, Edward JF, Lawrence AL. Mutational heterogeneity in human cancers: Origin and consequences. Annual Rev Pathol 2010; 5: 51-75.
2. Michael RS, Peter JC, Andrew PF. Mutational signatures of ionizing radiation in second malignancies Nature 2009; 458: 719-724.
3. Jeffrey G, Van Allen EM. Next-generation sequencing to guide cancer therapy. Genome Medicine 2015; 7: 80.
4. www.nlm.nih.gov/medlineplus/news/fullstory_135866.html
5. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012; 11.
6. Jiang Q, Turner T, Sosa MX, et al. Rapid and efficient human mutation detection using a bench-top next-generation DNA sequences. Hum Mutat 2012; 33: 281-289.
7. Chan M, Ji SM, Yeo ZX, et al. Development of a next-generation sequencing method for BRCA mutation screening: A comparison between a high-throughput and a benchtop platform. J Mol Diagn 2012; 14: 602-612.
8. Hudson TJ, Anderson W, Artez A, et al. International network of cancer genome projects. Nature 2010; 464: 993-998.
9. http://cancergenome.nih.gov
10. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature 2011; 474: 609-615.
11. Verhaak RG, Hoadley KA, Purdom E, et al. Cancer genome atlas research network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR and NF1. Cancer Cell 2010; 17: 98-110.
12. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature 2011; 475: 101-105.
13. Harris TJ, Mccormick F. The molecular pathology of cancer. Nat Rev Clin Oncol 2010; 7: 251-265.
14. Forbes SA, Tang G, Bindal N, et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): A resource to investigate acquired mutations in human cancer. Nucleic Acids Res 2010; 38: D652-D657.
15. http://cancer.sanger.ac.uk/cosmic/
16. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012; 2: 401-404.
17. http://www.cbioportal.org/
18. Meric BF, Johnson A, Holla V, et al. A decision support framework for genomically informed investigational cancer therapy. J Natl Cancer Inst 2015; 107: djv098.
19. https://pct.mdanderson.org/
20. http://www.mycancergenome.org/

21. Yeh P, Chen H, Andrews J, et al. DNA-mutation inventory to refine and enhance cancer treatment: A catalog of clinically relevant cancer mutations to enable genome-directed anticancer therapy. Clin Cancer Res 2013; 19: 1894-1901.
22. https://www.broadinstitute.org/cancer/cga/target
23. 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: 682-688.
24. https://www.intogen.org/
25. Abel GP, Christian PL, Jordi DP, et al. IntOGen-mutations identifies cancer drivers across tumor types. Nature Methods 2013; 10: 1081-1082.
26. https://civic.genome.wustl.edu/#/home
27. Malachi G, Nicholas CS, Kilannin K, et al. CIViC: A knowledge base for expert-crowdsourcing the clinical interpretation of variants in cancer. bioRxiv 2015.
28. http://www.dgidb.org/
29. Malachi G, Obi LG, Adam CC, et al. DGIdb: Mining the drug able genome. Nature Methods 2013; 10: 1209-1210.
30. 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: 565-574.
31. Parsons DW, Roy A, Plon SE, et al. Clinical tumor sequencing: an incidental casualty of the American College of Medical Genetics and Genomics recommendations for reporting of incidental findings. J Clin Oncol 2014; 32: 2203-2205.
32. Honeyman JN, Simon EM, Robine N, et al. Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. Science 2014; 343: 1010-1014.
33. Douillard JY, Oliner KS, Siena S, et al. Panitumumab FOLFOX4 treatment and RAS mutations in colorectal cancer. N Engl J Med 2013; 369: 1023-1034.
34. Jänne PA, Yang JCH, Kim DW, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med 2015; 372: 1689-1699.
35. Nathanson DA, Gini B, Mottahedeh J, et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. Science 2014; 343: 72-76.
36. Barbara AC, James HD. Molecular analysis for therapy choice: NCI MATCH. Seminars in oncology 2014; 41: 297-299.
37. Abrams J, Conley B, Mooney M, et al. National cancer institute's precision medicine initiatives for the new national clinical trials network. Am Soc Clin Oncol Educ Book 2014; 34: 71-76.

Correspondence to:

Tridip Chatterjee,
Suraksha Genomics (R & D Division of Suraksha Diagnostic),
DD-18/1. Sector 1, Salt Lake,
Kolkata-700064,
West Bengal,
India.
Tel: +9198313252580; 91-033-6619-1000
E-mail: ctridip@gmail.com; tridip.academic@gmail.com