

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Explore the Novel Biomarkers through Next-Generation Sequencing

Peng Li

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.73205>

Abstract

Next-generation sequencing is being a robust technology for the practice of clinical diagnosis. The reason is that this technology offers the advantage of higher sensitivity and the potential to detect the full genome sequence of pathogens, including unknown pathogen species. In view of the exceptional advantages of next-generation sequencing, the technology can be used to improve and revolutionize conventional pathogenic detection methodologies. The technological result holds great possibilities in helping to support clinicians with richer insights into host's genomic features, including the appropriateness to drug resistance based on the sequencing mapping of the microorganism. Besides, the technology will help discover the source of infection and insights into treatment directions, furthermore lead to the advancements in diagnosis. Eventually, this technology will benefit the clinical community in infectious disease prediction and prevention.

Keywords: pathogens, clinical disease, diagnosis, next-generation sequencing, disease biomarker

1. Introduction

Bacteria and virus consist of the most of the microbial pathogens. The microbial pathogens are responsible for infectious diseases causes in human hosts [1]. The microbial infections can cause the serious clinical symptoms in the human host, such as the inflammation, fever, pain and septic shock. They can even lead to the host death if the patient is treated with a delay. Thus, the early and accurate identification of pathogen is very important in clinical practice [2], as the proper antimicrobial treatment can be used to prevent the infection effectively. However, the conventional diagnostics, like polymerase chain reaction (PCR), enzyme-linked immunosorbent assay (ELISA) and microbial cell culture, are lack of the ability for the

| Application | References |
|---------------------|------------|
| Biodiversity | [7] |
| Disease diagnostics | [8] |
| Nutrition | [9] |
| Environment | [10] |
| Aging disease | [11] |
| Forensic biology | [12] |
| Agriculture | [13] |

Table 1. The application of NGS in biotechnology.

unknown or the high-mutated pathogens detection. Therefore, the novel pathogen diagnostics is necessary for the clinical healthcare.

Next-generation sequencing (NGS) is the latest scientific technology till date to sequence the target gene or genome [3]. NGS technology refers to one high-throughput DNA sequencing method. In a single experiment, it can determine the sequence of the target gene or full genome with a total size of larger than millions of base pairs (bp) [4]. Sequencing thousands of genes or even genomes in one experiment is consequently made possible using this NGS technology.

Because of its robustness, NGS is widely being applied in biotechnologies, such as in forensic biology, plant science and environmental contamination, etc., (**Table 1**). For example, the genome of *Mycobacterium tuberculosis* had been determined by Genome Analyzer to study the pathogen epidemiology [5]. In 2013, US Food and Drug Administration (FDA) has cleared Illumina MiseqDx to be the first *in vitro* NGS diagnosis platform [6]. In the recent years, with more development in this technology, NGS will provide more comprehensive information for the clinicians in clinical studies. In particular, there is more potential to translate currently available NGS technology into the pathogen detection.

2. The pathogen mutation is a challenge for infectious disease diagnosis

The microorganism pathogens are mostly responsible for the infection disease in the human body. PCR, ELISA and cell culture are the conventional methods for the pathogens detection [14]. However in PCR, one pair of primers, including forwards and reverse primers are compulsory to design according to the target gene or genome spectrum. Meanwhile in ELISA, the functional antibodies are indispensable for the microorganism antigen detection. Finally for cell culture, the related culture medium is also required for the according microorganisms. Thus, these conventional methods will not be working for some unknown or high-mutated microorganism, as the species sequencing information is deficient.

Furthermore, the pathogens, like virus and bacteria, have a high mutation rate which makes microorganism gene or proteins easily mutated [15]. The mutated gene would lead to the dramatic change of the protein structure [16]. The changed protein may not be detected by the original antibody, as a result, it will be challenging for the identification of these pathogens using the conventional methods. Therefore, there is a limitation for the conventional methods to detect the unknown and high-mutated pathogenic species. These mutated species will lead to the drug resistance and other clinical problems. The problems will also compromise the success of current antimicrobial treatment, leading to further increase in pathogen infection incidence and host mortality. Hence, the more powerful sequencing technology is urgently needed in the clinic community.

3. The development of next-generation sequencing technology

The first method of sequencing DNA was developed by Sanger [17]. He first devised a method that allowed for the determination of small sequences in ribonucleic acids. The Sanger sequencing method was developed here using the chain termination technology. Because of his development, Sanger was awarded the Nobel Prize in 1980 [18]. At the beginning stage, the 3' end of the primer anneals closely to the target DNA sequence. The addition of nucleotides on the 3' end of the template could either be a usual unlabeled nucleotide or a fluorescently labeled nucleotide. After that, no more nucleotides are able to bind to that particular strand of DNA sequence. This happens to all the templates that have been loaded into the Sanger sequencing platform and denatured into single strand DNA, leading to various single strands with different lengths and with diverse labeled nucleotides. As soon as the sequencing has ended, the finished sample would be denatured once again to remove the sequenced strands from the original loaded DNA. The sequenced strands would then be inputted into an Agarose gel for observing under a UV light. While the sample is undergoing Agarose gel running, it can be noted that the varying lengths of sequenced DNA would run at diverse times across the Agarose electrophoresis. The longer fragments are able to travel much slower than their shorter counterparts. The bands that will be then analyzed at the end of the gel. The result will be arranged according to their fragment size and labeling, therefore giving a visual image of the base sequence in the inputted sample. Sanger sequencing had been regarded as a gold standard metric because of its high accuracy. The method of the Sanger sequencing had been applied to sequence the first *Homo sapiens* genome [19].

However, even there are technological advances in the Sanger sequencing method such as the automation, the Sanger sequencing was still time-consuming and very costly. The growing interest in the sequencing of the personal genomes fueled the development of new robust technology. The NGS technology has been introduced currently. NGS has many functions, such as the sequencing of an entire genome, deep sequencing for a target region of the genome, or even multiplex sequencing which allows many samples to be sequenced at one time. In the NGS workflow (**Figure 1**), the multiplex sequencing function is utilized so that various samples can be sequenced simultaneously. The principle behind NGS platform is alike to that of Sanger sequencing. The signals produced from fluorescently or radioactively



Figure 1. The NGS workflow consists of library construction, sequencing and data analysis.

labeled nucleotides are received, allowing the bases of the template DNA sequence to be read in order. But the difference between the two sequencing technologies is that NGS has the competence to handle numerous sequencing reactions simultaneously. Many templates of DNA are able to be processed at the same time, which makes the whole genome sequencing time to be much more rapid with this robustness. Thus, NGS will be the next important sequencing tool used in biological and clinical samples as it offers a super speed with a higher accuracy.

Over these years, NGS technologies have matured and thus lowered cost and dramatically increased throughput. This technology eliminates the time-consuming and labor-intensive step to generate single clones via bacterial cloning and gel electrophoresis, and using the parallel processing to simultaneously sequence a large number of DNA sample. Thus, instead of generating hundreds of longer reads (more than 1000 bp), NGS technology produces millions of shorter reads (100 ~ 600 bp) ranging on the order of gigabases per run (**Table 2**). Consequently, the major work of sequencing has shifted from the benchtop to the desktop. For example, a nanopore NGS instrument MinION had been developed and applied for pathogens detection [20, 21]. Moreover the analysis of the NGS data presents new advancements mainly due to the short read lengths and require significant investment in big data processing, including hardware, software and bioinformatics. Therefore, the high-throughput data is easily to be handled. As a result, what once took 1 week to sequence using Sanger sequencing can now be accomplished in a matter of days in a desktop NGS platform.

Superior to the traditional sequencing method, NGS is also able to sequence unknown DNA sequence. The unknown genome sequence can be *de novo* assembled. Depending on the platforms

| | Sanger sequencing | NGS |
|------------------------------|---------------------|----------------------|
| Experimental time | ~1 week | ~2 days |
| Cost per sample | Expensive | More cost-effective |
| Reads per sample | One read | Up to Millions |
| Reading length | Long reading length | Short reading length |
| Cloning vector required | Yes | No |
| Specific primer design | Yes | No |
| Gel electrophoresis required | Yes | No |

Table 2. Comparison of Sanger sequencing and NGS.

used, NGS can sequence from tens of thousands to more than a billion molecules in a single sequencing run. And it is independent of the known microorganism sequence. In addition, NGS obtains this feature because of its ability to sequence thousands of the inputted sample, in a parallel style, rather than sequencing a single DNA template (**Figure 1**). This particularly parallel analysis is achieved by the miniaturization of the volume of the individual sequencing reaction, which limits the size of the instrument and reduces the cost of reagents per reaction.

4. Identification of the novel biomarkers through NGS

NGS technology proves to be a cost-effective, rapid, yet a highly sensitive method to sequence large amounts of DNA at once. This can enhance infectious disease research which may eventually lead to new biomarkers discoveries. These discoveries can be translated into new diagnostic, prognostic and therapeutic targets. In previous studies, NGS technology was used to detect the common mutation in viral samples from infected patients. Across the various samples analyzed, two common viral mutations were identified in all of the samples. A silent mutation and a missense mutation was detected. These common mutations identified code for viral reverse transcriptase subunits p66 and p51. As reverse transcriptase is extremely important for the survival of the virus, the common mutations identified are possible novel biomarkers for virus among local strains. With the identification of these novel biomarkers, it would serve to improve diagnosis as well as treatment.

In the particular study for pathogen diagnosis, NGS had allowed for the sequencing and identification of even the smallest variants of the viral or bacterial genome. The NGS results were able to illustrate the various base mutations that occurred across all multiple samples provided. Therefore, NGS has the robust advantage over conventional diagnostics of having higher sensitivity, especially about low-frequent mutation or variants. This could be the reason that the traditional sequencing method counts on a given position when sequencing a determined DNA base. A minor mutation or variants will possibly have a low signal-to-noise ratio that is unclear from the background noise. But for NGS, it makes use of complete coverage over the full gene or the whole genome which provides a much higher sensitivity regarding minor mutant or variant. From these results, perhaps these mutations would prove to have a rather significant impact on the drug-resistant capabilities of the virus. Further studies would be also extended to other viral or bacterial genome research. These will provide the characteristics of the microbial pathogens and the disease transmission pathways.

The common mutations identified at virus are located on the pol gene that codes for reverse transcriptase subunits [22]. Unlike other silent mutations, the missense mutation has more implications. As a result, there is an alteration in the corresponding amino acid from leucine to phenylalanine. Aligned with the reference genome, the protein function could be altered because of the missense mutation. The missense mutation would result in changes to the protein structure of reverse transcriptase, causing the conformational changes. These changes to the protein have potential to cause resistance to antimicrobial drugs, allowing the virus to continue developing in the host.

Consequently drug resistance remains a challenge for the treatment of pathogen infection. It arises from the pathogen's ability to mutate rapidly. The infected patients can initially be infected with a drug-resistant virus or develop drug resistance after starting therapy. Studies have been conducted to identify the mutations due to the resistance to antiviral medicines. More than 50 and 40 reverse transcriptase mutations have been found to be associated with nucleoside reverse transcriptase inhibitor (NRTI) and non-nucleoside reverse transcriptase inhibitor (NNRTI) drugs respectively [23]. The viral reverse transcriptase is highly important as it catalyzes the conversion of single-stranded RNA to double-stranded viral DNA for integration into the host genome. This enzyme plays an important role in the life cycle of the virus and has been a good target for the development of antiretroviral drugs for the treatment of pathogens. Currently, there are two broad classes of drugs that target the viral reverse transcriptase; NRTI (like stavudine and emtricitabine) and NNRTI (like nevirapine and etravirine) [24].

The ability of the virus to mutate at the specific position may be due to the selective pressure. There is a high possibility that these mutations can be commonly observed in viral strains. The development of such mutations would allow for the survival and continuity of the local subtype virus. Hence, the identification of these common mutations could be used as novel biomarkers for the diagnosis. Studies have shown similar mutations identified in local viral strains. This further certifies that the mutation found is likely to be common in viral strains. Thus, the common mutations found through NGS technology can be applied as diagnostic biomarkers. Furthermore, it can be developed into the potential marker for the future drugs to defense against pathogens. This can significantly help the microbiological laboratories in large-scale studies of the virus, which aims to aid in the clinical management of pathogen infection [25]. Nevertheless, further studies should be done with a larger sample size to confirm if the identified common mutation is still observed in a larger population.

Compared with microbial culture, NGS technology is a culture-free detection methodology. Metagenomics sequencing will provide a fast, reliable tool for a rapid microbial diagnosis. The conservation region of 16 s ribosomal ribonucleic acid (rRNA) amplicons can be sequenced by a standard workflow. Thus, this methodology can identify hundreds or even thousands species at one time. The microbial system biology can be also investigated at the same time. The automated software or pipeline will help the metagenomics to be a standard microbial detection method.

The molecular epidemiology studies can be investigated deeply, using NGS results. And the investigation of subtype differences in clinical phenotypes and treatment outcomes can be achieved. It is fully predictable that NGS will be much better than the often used conventional diagnosis methods, for providing sequence information for the genomic, microbiological and clinical studies. The current standard of diagnosis methods, commercially available PCR, ELISA, cell culture or other assays target short sub-gene fragments for drug resistance determination. Furthermore, it is interesting to highlight that mutations outside these regions can influence drug resistance [26]. Additionally, NGS is more powerful to study of drug resistance, and disease transmission [27]. There are many emerging multi-drug resistant organisms globally, thus it is significant to investigate the molecular microbiology of the new pathogens.

Antimicrobial therapy is widely performed as a form of treatment in fighting against pathogen infection. Unfortunately, some microorganism has the ability to rapidly mutate which results in changes in its genetic or protein structure. This provides pathogens with the potential to develop resistance to existing antimicrobial drug or treatment. NGS can provide the solution of sequencing a pathogenic gene/genome and the identification of a common mutation in the targeted region. The decrease in cost and increase in accuracy, resolution and reproducibility of NGS allows large-scale sequencing of the virus to be performed efficiently. The study made use of NGS platform together with the usage of the library preparation to sequencing the mutated pathogenic samples. The advancement of NGS has brought about many benefits to the field of biological sciences and will continue to play a big role in the disease diagnostics.

5. Future perspectives

By using NGS, the advantages of this technology was able to be observed throughout the duration of the experiments. One of the really good advantages includes the deep sequencing protocol that occurs during NGS. Deep sequencing is the process of sequencing the same region several times, from hundreds to ten of thousands times coverage. When amplicons are able to be sequenced at a really high depth of coverage, the sequence mutations can be highlighted. It allows for the detection of multiple variants that are really low in number within its population. Somatic mutations that cannot be identified easily using the Sanger sequencing can be easily sequenced, making rare infectious diseases easier to study in a clinical environment.

However, the main NGS platforms used has a limited length of the sequence generated in individual reactions. The read length for the majority of platforms is in the range of hundreds of base pairs. In order to sequence DNA longer than the feasible read length, the material need fragmented before analysis. Following sequencing, the reads are analyzed through informatics to provide the information on the sequence of the whole target molecule. The short read, particularly the high-throughput sequencing methodology used in NGS was a different solution that developed sequencing competence. It enables population-scale sequencing and establishes a foundation for the novel genomic medicine as part of healthcare. NGS technologies are increasingly used for diagnosis and monitoring of infectious diseases such as virus infection. NGS is more powerful than other methods, such as Sanger sequencing, especially in the improved accuracy of the unknown region. The new technology is less costly and is more capable to detect the repeated fragments. Although the usage of NGS still has its setbacks, such as the relatively expensive price of the sequencing consumables required to conduct one sequencing run, perhaps in the near future the overall cost of NGS will be reduced; with increased popularity of this sensational sequencing method, NGS will eventually be as a cheap and standard method in biomarker discovery. Also, an adapted genotyping prediction informatics could be developed based on data acquired from whole-genome sequences of drug-resistant isolates. The predicted novel treatment resistance conferring mutations would be validated against phenotypic assay as well as clinical data acquired from the patients. The correlative application between the new solution and the conventional methods, such as PCR would be also determined together to assess the performance of the drugs.

Abbreviations

| | |
|-------|--|
| bp | Base pairs |
| ELISA | Enzyme-linked immunosorbent assay |
| NGS | Next-generation sequencing |
| NNRTI | Non-nucleoside reverse transcriptase inhibitor |
| NRTI | Nucleoside reverse transcriptase inhibitor |
| PCR | Polymerase chain reaction |
| rRNA | ribosomal ribonucleic acid |

Author details

Peng Li

Address all correspondence to: li_peng@sp.edu.sg

Singapore Polytechnic, Singapore

References

- [1] Gomez-Diaz E, Jorda M, Peinado MA, Rivero A. Epigenetics of host-pathogen interactions: The road ahead and the road behind. *PLoS Pathogens*. 2012;**8**:e1003007. DOI: 10.1371/journal.ppat.1003007
- [2] Dark PM, Dean P, Warhurst G. Bench-to-bedside review: The promise of rapid infection diagnosis during sepsis using polymerase chain reaction-based pathogen detection. *Critical Care*. 2009;**13**:217. DOI: 10.1186/cc7886
- [3] van Vliet AH. Next generation sequencing of microbial transcriptomes: Challenges and opportunities. *FEMS Microbiology Letters*. 2010;**302**:1-7. DOI: 10.1111/j.1574-6968.2009.01767.x
- [4] Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: From basic research to diagnostics. *Clinical Chemistry*. 2009;**55**:641-658. DOI: 10.1373/clinchem.2008.112789
- [5] Kato-Maeda M et al. Use of whole genome sequencing to determine the microevolution of *Mycobacterium tuberculosis* during an outbreak. *PLoS One*. 2013;**8**:e58235. DOI: 10.1371/journal.pone.0058235

- [6] FDA-approved next-generation sequencing system could expand clinical genomic testing: Experts predict MiSeqDx system will make genetic testing more affordable for smaller labs. *American Journal of Medical Genetics: Part A*. 2014;**164A**:x-xi. DOI: 10.1002/ajmg.a.36461
- [7] Huete-Perez JA, Quezada F. Genomic approaches in marine biodiversity and aquaculture. *Biological Research*. 2013;**46**:353-361. DOI: 10.4067/S0716-97602013000400007
- [8] Lecuit M, Eloit M. The potential of whole genome NGS for infectious disease diagnosis. *Expert Review of Molecular Diagnostics*. 2015;**15**:1517-1519. DOI: 10.1586/14737159.2015.1111140
- [9] Liu GE. Applications and case studies of the next-generation sequencing technologies in food, nutrition and agriculture. *Recent Patents on Food, Nutrition & Agriculture*. 2009;**1**:75-79. DOI: 10.2174/2212798410901010075
- [10] Wong K, Fong TT, Bibby K, Molina M. Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environment International*. 2013;**45**:151-164. DOI: 10.1016/j.envint.2012.02.009
- [11] Yang HJ, Ratnapriya R, Cogliati T, Kim JW, Swaroop A. Vision from next generation sequencing: Multi-dimensional genome-wide analysis for producing gene regulatory networks underlying retinal development, aging and disease. *Progress in Retinal and Eye Research*. 2015;**46**:1-30. DOI: 10.1016/j.preteyeres.2015.01.005
- [12] Yang Y, Xie B, Yan J. Application of next-generation sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics*. 2014;**12**:190-197. DOI: 10.1016/j.gpb.2014.09.001
- [13] Bansal U, Bariana H. Advances in identification and mapping of rust resistance genes in wheat. *Methods in Molecular Biology*. 2017;**1659**:151-162. DOI: 10.1007/978-1-4939-7249-4_13
- [14] Priyanka B, Patil RK, Dwarakanath S. A review on detection methods used for food-borne pathogens. *The Indian Journal of Medical Research*. 2016;**144**:327-338. DOI: 10.4103/0971-5916.198677
- [15] Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. *Journal of Virology*. 2010;**84**:9733-9748. DOI: 10.1128/JVI.00694-10
- [16] Hanna N. Advances in the treatment of second-line non-small-cell lung cancer. *Lung Cancer*. 2005;**50**(Suppl 1):S15-S17. DOI: 10.1016/S0169-5002(05)81554-5
- [17] Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*. 1975;**94**:441-448. DOI: 10.1016/0022-2836(75)90213-2
- [18] Sanger F. The early days of DNA sequences. *Nature Medicine*. 2001;**7**:267-268. DOI: 10.1038/85389

- [19] Tsiatis AC et al. Comparison of sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *Journal of Molecular Diagnostics*. 2010;**12**:425-432. DOI: 10.2353/jmoldx.2010.090188
- [20] Kilianski A et al. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. *Gigascience*. 2015;**4**:12. DOI: 10.1186/s13742-015-0051-z
- [21] Bates M, Polepole P, Kapata N, Loose M, O'Grady J. Application of highly portable MinION nanopore sequencing technology for the monitoring of nosocomial tuberculosis infection. *International Journal of Mycobacteriology*. 2016;**5**(Suppl 1):S24. DOI: 10.1016/j.ijmyco.2016.10.035
- [22] Shaw WH et al. Identification of HIV mutation as diagnostic biomarker through next generation sequencing. *Journal of Clinical and Diagnostic Research*. 2016;**10**:DC04-DC08. DOI: 10.7860/JCDR/2016/19760.8140
- [23] Shafer RW, Schapiro JM. HIV-1 drug resistance mutations: An updated framework for the second decade of HAART. *AIDS Reviews*. 2008;**102**:67-84
- [24] Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harbor Perspectives in Medicine*. 2012;**2**:a007161. DOI: 10.1101/cshperspect.a007161
- [25] Gall A et al. Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *Journal of Clinical Microbiology*. 2012;**50**:3838-3844. DOI: 10.1128/JCM.01516-12
- [26] Delviks-Frankenberry KA, Nikolenko GN, Pathak VK. The "connection" between HIV drug resistance and RNase H. *Virus*. 2010;**2**:1476-1503. DOI: 10.3390/v2071476
- [27] Islam MA et al. Emergence of multidrug-resistant NDM-1-producing Gram-negative bacteria in Bangladesh. *European Journal of Clinical Microbiology & Infectious Diseases*. 2012;**31**:2593-2600. DOI: 10.1007/s10096-012-1601-2