

High-throughput Sequencing Technology and Its Application in Epigenetics Studies

Zexuan Wang*

Nanjing Agricultural University, Nanjing, China, 210095

Abstract. DNA sequencing technology is one of the important tools in modern life science research, and with the development of science and technology, sequencing technology is constantly changing. High-throughput sequencing technology is known as the second generation sequencing technology. Compared with Sanger sequencing technology, high-throughput sequencing technology has the characteristics of high throughput, fast speed and low cost. In this paper, the researcher briefly introduced the emergence and development of DNA sequencing technology, and expounded the application of high-throughput sequencing in epigenetic research from the aspects of DNA methylation and histone modification.

1 Introduction

The first-generation sequencing technology is the dideoxy end termination method proposed by Sanger[1] in 1970, which is the most classic sequencing method, and also the sequencing method with the longest application time and the widest application range. Since then, DNA sequencing has taken off in a big way, bringing biology to a new level of genetic research and giving us the complete sequence of the human genome. After multiple generations of development, sequencing technology is now widely used in basic research, technology development and clinical practice. Epigenetics needs to be specialized because proteomics and proteomics are at the forefront of research, and many genes need to be sequenced to obtain data in the process. Therefore, it is necessary to study the application of high-throughput sequencing methods in epigenetic research

2 High-throughput sequencing

High-throughput sequencing technology is also known as deep sequencing technology, new generation sequencing technology or second generation sequencing technology. Compared with the traditional Sanger sequencing technology, this technology has two outstanding advantages. First, it can achieve large-scale parallel sequencing by using microarray configuration to improve sequencing efficiency. Second, there is no reliance on electrophoretic separation technology, thus reducing the cost, saving manpower and time[2,3].

2.1 Roche 454 (GS FLX Titanium System)

Roche 454 sequencing platform is the earliest and relatively mature second-generation sequencing platform in the world, which belongs to the cyclic microarray method platform, and its sequencing technology is based on the Sequencing by synthesis(SBS) technology[4].The system relies on bioluminescence to detect DNA sequences. After constructing the library, the template was amplified by PCR. In the synergistic action of DNA polymerase, ATP sulfase, luciferase and bisphosphatase, if the template is paired with deoxy-ribonucleoside triphosphate (dNTP), an equal amount of pyrophosphate group will be released[5]. The aggregation of each dNTP on the primer was coupled to a primary fluorescence signal release. By detecting the presence and intensity of the fluorescence signal released, the purpose of real-time DNA sequence determination can be achieved.

Compared with other high-throughput sequencing technologies, 454 pyrosequencing has the advantage of short time and a large number of fragments per unit time. In one sequencing work, 1 million sequences can be generated, the average length of sequences is 400bp, and the total amount of data is about 500M. However, because the length of the homomer cannot be accurately measured, insertion and miss errors may be introduced. With the increasingly fierce market competition and the update and iteration of technology, the sequencing method of Roche 454 has been gradually phased out. In 2013, Roche announced the closure of its 454 sequencing business and the termination of related services in 2016[6,7]

*11319202@njau.edu.cn

2.2 Illumina Solexa

Illumina's solexa technology is single-molecule array sequencing, and its core technologies are "bridge PCR" and "reversible terminal termination" [8,9]. Bridge PCR generates large clusters of molecules by amplifying DNA to achieve the intensity required for fluorescence detection. Solexa sequencing reaction is a reversible termination chemical reaction. DNA fragments can be randomly attached to the surface of the solid phase carrier and undergo bridge amplification on the surface of the solid phase. This creates thousands of identical single-molecule clusters, which are used as sequencing templates. Sequencing takes the method of Sequencing by synthesis, the ddNTP material paired with the template is added, and the unpaired ddNTP material is washed away, so that the imaging system can capture the fluorescently labeled nucleotides.

The difference between Solexa and other high-throughput sequencing technologies is that it only adds one DNTP per cycle in the synthesis process to solve the problem of homomer determination. The advantages of Solexa technology are its high accuracy, high speed, and low cost. This gives it an advantage in resequencing biological genomes, but it also has the big problem of having shorter read lengths than other sequencing techniques [10].

2.3 SOLiD

ABI's SOLiD sequencing technology is an evolution of ligase sequencing. SOLiD sequencing also used emulsion PCR similar to the 454 pyrosequencing technology. SOLiD sequencing technology is based on the continuous ligation synthesis of 8-base four-color fluorescently labeled oligonucleotides, which replaces the traditional polymerase ligation reaction, and can be used for large-scale amplification and high-throughput parallel sequencing of single-copy DNA fragments. SOLiD sequencing is unique in its "double base correction" technique for ligation reactions [11,12]. SOLiD sequencing platform supports two kinds of sequencing libraries, one is fragment library and the other is mate-paired library. The former is mainly used in RNA-seq, 3', 5'-RACE, methylation analysis, ChIP-seq, etc. The latter is mainly used in whole genome sequencing, SNP analysis, structural rearrangement, CNV and so on [13].

SOLiD sequencing's double-base correction technology, which allows each base to be determined twice, as well as the use of ligases and change of sequencing primer, greatly reduces the probability of incorrect sequencing, making it the most accurate platform among next-generation sequencing platforms. However, due to its short read length and slow running disadvantage compared with other technologies, it was finally withdrawn from the market.

3 Application of high-throughput techniques to epigenetics

Epigenetics refers to heritable changes in gene expression

without changes in DNA sequence. This change is the change of other heritable material in the cell except genetic information, and this change can be stably inherited during development and cell proliferation [14,15]. Epigenetics is a lot of things, DNA methylation, histone modification, Chromatin remodeling, non-coding RNAs Regulation, gene silencing, nucleolar dominance, dormant transposon activation and maternal effects are all typical epigenetic phenomena [16,17]. The application of high-throughput sequencing in epigenetic research is mainly focused on DNA methylation, histone modification and non-coding RNA regulation [18].

3.1 Application in the study of DNA methylation

In the study of DNA methylation, the commonly used high-throughput sequencing technology is Bisulfite sequencing (Bi-seq) and reduced representation Bisulfite sequencing (RRBS), methylated DNA immunoprecipitation sequencing (MEDIP-seq) and other techniques.

3.1.1 Application of Bi-Seq and RRBS in epigenetic studies

At present, bisulfite modification is recognized as the "gold standard" for DNA methylation detection. Both Bi-Seq and RRBS technologies are based on the above modifications. Whole-genome Bisulfite sequencing technology can obtain whole-genome methylation map with single base resolution. This technology was initially applied by Cokus et al. [19] to the whole gene methylation profile analysis of *Arabidopsis thaliana* (with the help of Solexa sequencing technology). In 2009, Lister et al. [20] published the first human genome-wide methylation map using whole-genome Bisulfite sequencing (using Solexa sequencing), and subsequently this technology began to be applied in the study of human diseases, such as obesity [21]. Based on Bisulfite processing of DNA samples from peripheral blood monocytes of the yellow race and deep sequencing combined with Solexa sequencing technology, Li et al. [22] successfully mapped the high-precision genome-wide methylation map of "Yanhuang No.1" in 2010. However, the high sequencing cost of bisulfite conversion has limited the wide application of this technology to a certain extent.

Reduced representation Bisulfite sequencing (RRBS), also known as bisulfite sequencing based on enzyme digestion, is another type of methylation sequencing method, which combines restriction length selection, bisulfite conversion, PCR amplification and cloning. RRBS is an accurate, efficient and economical method for DNA methylation research. It enriches promoter and CpG island regions by enzyme digestion, and performs Bisulfite sequencing. This method ensures the high resolution of DNA methylation status detection and improves the high utilization of sequencing data. In 2005, Meissner et al. [23] originally invented this technology in combination with Sanger sequencing, and successfully applied it to detect the methylation profile of mouse

embryonic stem cells before and after demethyltransferase. Then, the research group led by Meissner enriched nearly 90% of CpG island fragments in the whole mouse genome by switching to MspI digestion, and combined with Illumina's high-throughput sequencing technology, established and improved the method suitable for mammalian whole genome methylation sequencing analysis [24]. In 2010, this research group published another article on RRBS to further optimize RRBS technology and explore its potential in clinical application based on its previous work [25]. Later, Smallwood et al. [26] also applied this technology to the methylation detection of mammalian mature oocytes.

3.1.2 Application of MeDIP-seq in epigenetic studies

Methylated DNA immunoprecipitation sequencing (MeDIP-seq) mainly uses the principle of antibodies that specifically recognize 5-methylcytosine (5-MC) to enrich genome methylated fragments, and combines high-throughput sequencing technology [27]. The basic procedure is: (1) The whole genome DNA was extracted, and the fragments with length of 400-500 bp were interrupted by ultrasound; (2) Heat denaturation, and evenly divide the denatured single-stranded DNA samples into two parts; (3) One of them was added with methylated DNA specific antibody, and the other was used as Total input DNA sample; (4) The antibody complex of methylated DNA fragment in the previous step sample was separated by affinity chromatography, and the remaining unmethylated DNA fragment in the sample was eluted and purified to obtain methylated DNA fragment (MeDIP DNA). This method has been applied to the study of cancer pathogenesis. For example, Ruike et al. [28] chose the high-throughput sequencing technology of Illumina Company to conduct genome-wide methylation study on human breast cancer cell lines by MeDIP technology.

3.2 Application in the study of Histone modification

Histone modifications include methylation, acetylation, phosphorylation, ubiquitination and sumoylation. The current research mainly focuses on histone methylation and acetylation. At present, there are few methods to study histone modification, and the most commonly used method is chromatin immunoprecipitation.

Chromatin Immunoprecipitation (ChIP), proposed by O'Neill and Turner [29], is a technique used to study the interaction between protein and DNA *in vivo*. It is the basic principle of state fixed protein in living cells - DNA complexes, and then by ultrasound or enzyme treatment will be cut off for a certain length of chromatin within the scope of small fragments, and then through the immunological method to precipitate the complex, specific protein to enrichment and purpose of DNA fragments, fragments of purification and detection based on purpose, To get information about how proteins interact with DNA.

3.2.1 ChIP-chip

Chip-chip is a method that combines ChIP with biochip to analyze DNA binding sites or histone modifications in the whole genome or in a large region of the genome with high throughput [30]. The basic procedure of this technology is: firstly, histone modified DNA fragments are enriched by chromatin immunoco-deposition (ChIP), and then polymerase chain reaction (PCR) amplification is performed by adding universal adaptor, during which fluorophore groups are introduced. Since the length of enriched fragments is different, the amplification efficiency is different, and the bias is reduced by controlling the number of cycles. Finally, the amplified fragment was hybridized with the designed chip.

3.2.2 ChIP-seq

Chip-seq is a high-throughput method that combines ChIP and sequencing technology to detect DNA histone modifications on a genome-wide scale, which can be applied to any species with genome sequences and can obtain exact sequence information of each segment [30]. The basic procedure of this technology is: the target fragment is enriched by ChIP, purified, and then added to the universal adapter for PCR amplification, and finally added to the Solexa adapter for sequencing. At present, the technology is relatively mature, and the cost is gradually decreasing with the emergence and development of the new generation sequencing technology. The combination of ChIP and sequencing technology is more and more widely used in DNA and interacting protein analysis.

Compared with ChIP-chip, ChIP-seq technology has the advantages of higher accuracy and greater coverage [31]. In 2007, ChIP-seq technology was initially reported to be used in histone modification studies of mammalian T cells and embryonic stem cells [32-33]. After that, ChIP-seq technology was used to treat prostate cancer cells, and the whole gene localization of epigenetic mark nucleosomes was carried out. He et al. [34] studied nucleosome dynamics and gene transcriptional regulation. Hurtado et al. [35] studied breast cancer cells based on ChIP-seq technology and revealed that the determinant of estrogen receptor activity and endocrine response in breast cancer cells came from FOXA1 protein.

4 Conclusion

Lister et al. [36] used whole-genome Bisulfite sequencing technology to construct the human genome methylation map, which laid a solid foundation for the study of human genetic diseases. Zhang Xiaoli [37] used MediP-Seq technology to construct DNA library of 3 landwhite sows at 210 days of age, obtained highly matched 32.91 Gb DNA sequence, and identified differentially methylated regions (DMRs) in different tissues. The ratio of CpG in DMRs in exons was significantly higher than that in promoters, introns, and 2 KB upstream of transcription start sites, proving that methylation in promoters can regulate gene expression. This study provides epigenetic basic data for in-depth analysis of adipose function in

different animal tissues. Sun et al. [38] used ChIP-seq technology to study RNA Pol- ii promoters in mouse tissues and detected 38 639 POL- ii promoters and 12 270 new promoters, identifying POL- ii promoters of annotated genes in different tissues. 37% of the coding genes were found to be regulated by selective promoters.

Since the emergence of high-throughput sequencing technology, with the continuous improvement of molecular biology and science and technology, high-throughput sequencing technology has achieved rapid development in just 20 years. However, there are still many things to be improved in the high-throughput sequencing platform. The second-generation sequencing platform has a short sequence read length and relies on PCR amplification technology, which is easy to cause error and bias in sequence reading, which causes great difficulties in the later bioinformatics data analysis and processing. It will undoubtedly increase the time cost and economic cost of sequencing, so the second-generation sequencing platform needs to improve the read length and accuracy of sequences in the later stage [39].

Although high-throughput sequencing technology still has a long way to go, it has undoubtedly made a great contribution to the research field of epigenetics. With the development of sequencing technology and the emergence of new sequencing technologies, there are more and more overlapping research fields, which will surely provide the driving force for the progress of science and the development of mankind.

Reference

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*,1977, 74(12): 5463-7
2. Mitra RD, Church GM. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res*, 1999, 27(24): 34-9
3. Shendure J, Porreca GJ, Reppas NB, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 2005, 309(5471): 1728-32
4. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437: 376-80
5. Li Jixia, Gong Yanwen, Zhang Yuce, et al. The identification ability of pyrosequencing technology was evaluated by VITEK2Compact system for rare clinical pathogens with low accuracy[J].*Biomedical Engineering and Clinical*, 2016, 20 (6) :626-630.
6. Mardis ER. The impact of next-generation sequencing technology on genetics.*Trends Genet*, 2008, 24(3): 133-41
7. <http://www.454.com>.
8. Fedurco M, Romieu A, Williams S, et al. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*, 2006, 34: e22
9. Turcatti G, Romieu A, Fedurco M. et al. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res*, 2008, 36(4): e25
10. DEURENBERG R H, BATHOORN E,CHLEBOWICZ M A,et al.Application of next generation sequencing in clinical microbiology and infection prevention[J].*J Biotechnol.*,2017,243,16-24
11. Smith DR, Quinlan AR, Peckham HE, et al. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res*, 2008, 18 (10) : 1638-42
12. Zhou Xiaoguang, Ren Lufeng, Li Yuntao, et al. Next Generation Sequencing: Technology Review and prospects. *China science* ,, 2010, 40(1): 23-37
13. LIU L, LI Y, LI S, et al. Comparison of next generation sequencing systems[J].*J Biomed Biotchnol*, 2012,2012:251364
14. Egger G, Liang C, Aparicio A, et al. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, 2004, 429 (6990) : 457-63.
15. Bird A. Perceptions of epigenetics. *Nature*, 2007, 447 (7143) : 396-8. Bender J. DNA methylation and epigenetics. *Annu Rev Plant Biol*, 2004, 55: 41-68.
16. Paszkowski J, Whitham SA. Gene silencing and DNA methylation processes. *Curr Opin Plant Biol*, 2001, 4(2) : 123-9.
17. Yang Xu, Jiao Rui, Yang Lin, et al.Human disease-omics research strategies based on next-generation high-throughput technologies. *Genetic*, 2011, 33(8): 829-46
18. Cokus SJ, Feng SH, Zhang XY, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 2008, 452: 215-9
19. Lister R, Pellzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 2009, 462(7271): 315-22
20. Turcot V, Bouchard L, Faucher G, et al. DPP4 gene DNA methylation in the omentum is associated with its gene expression and plasma lipid profile in severe obesity. *Obesity*, 2011, 19(2): 388-95
21. Li Y, Zhu J, Tian G, et al. The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol*, 2010, 8(11): e1000533
22. Meissner A, Gnirke A, Bell GW, et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res*, 2005, 33(18): 5868-77
23. Smith ZD, Gu HC, Bock C, et al. High-throughput bisulfite sequencing in mammalian genomes. *Methods*, 2009, 48(3): 226-32
24. Gu HC, Bock C, Mikkelsen TS, et al. Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. *Nat Methods*, 2010, 7(2): 133-6

25. Smallwood S, Tomizawa S, Krueger F, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet*, 2011, 43: 811-5
26. Nair SS, Coolen MW, Stirzaker C, et al. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methylCpG binding domain(MBD) protein capture for genomewide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics*, 2011, 6(1): 34-4
27. Ruike Y, Imanaka Y, Sato F, et al. Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics*, 2010, 11(1): 137-47
28. O'Neill LP, Turner BM. Immunoprecipitation of chromatin. *Methods Enzymol*, 1996, 274: 189-97.
29. Li Minli, Wang Wei, Lu Zuhong. ChIP technology and its analysis of DNA and protein interactions at the genomic level. *Genetic*, 2010, 32(3) , 219-228.
30. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 2009, 10(10): 669-80
31. Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*, 2007,129(4): 823-37
32. Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 2007, 448(7153): 553-60
33. He HH, Meyer CA, Shin H, et al. Nucleosome dynamics define transcriptional enhancers. *Nat Genet*, 2010, 42(4): 343-7
34. Hurtado A, Holmes KA, Ross-Innes CS, et al. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet*, 2011, 43(1): 27-33
35. LISTER R, ECKER J R. Finding the fifth base: Genome-wide sequencing of cytosine methylation[J]. *Genome Res*, 2009,19(6) : 959-966.
36. Zhang Xiaoli. Genome-wide methylation of superficial and deep dorsal adipose tissue in pigs[D]. Yan'an: Sichuan Agricultural University, 2013.
37. SUN H, WU J J, WICK RAMASINGHE P, et al. Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-Seq[J]. *Nucleic Acids Res*, 2011, 39(1) : 190-01.
38. VAN DIJK E L, JASZCZYSZYN Y, DE THERMES C . Library preparation methods for next generation sequencing: tone down the bias[J]. *Exp Cell Res*,2014, 322(1):12-20