ejpmr, 2017,4(1), 216-222

SJIF Impact Factor 3.628

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

Review Article

ISSN 2394-3211

EJPMR

REVIEW STUDY OF METHODS USED IN NGS (NEXT GENERATION SEQUENCING) IN GENE EXPRESSION

***[1]Yusuf Talib and [2]Talib S.H.**

[1]MGM University of Health Sciences.
[2]MGM Medical College and Hospital, Aurangabad. 431001 (MS).

Corresponding Author: Yusuf Talib
MGM University of Health Sciences.

Article Received on 25/10/2016 **Article Revised on 15/11/2016** **Article Accepted on 05/12/2016**

ABSTRACT

Study of Central Dogma in biology at molecular level is much more important which helps us to understand the genetic arrangement and aspects of gene expression at different level and scenario, which can be useful in studying the expression level for disease and their cause and precautionary measures.DNA sequencinf has come a long way since the days of two-dimensional chromatography in 1970's,Genome sequencing is one of the method which helps to provide the details of every sequence and Base pair, which means to determine the primary structure (sometimes falsely called primary sequence) of an unbranched biopolymer. Sequencing results in a symbolic linear depiction known as a sequence which succinctly summarizes much of the atomic-level structure of the sequenced molecule. NGS (Next generation sequencing) makes sequence-based gene expression analysis a "digital" alternative to analog techniques. It lets you quantify RNA expression with the breadth of a microarray and the resolution of qPCR.

KEYWORDS: - sequencing, NGS, Microarray, Gene Expression, Sanger Method, Illumina.

INTRODUCTION

Since first introduced to the market in 2005, next-generation sequencing technologies have had a tremendous impact on genomic research. The next-generation technologies have been used for standard sequencing application, such as genome sequencing and re-sequencing, and for novel application previously unexplored by Sanger sequencing. In this review we first describe the three commercially available next generation sequencing technologies in comparison to a state-of-the-art Sanger sequencer, and follow this with a discussion of the novel kind of data produced by next-generation sequencers and the issue associated with it. We then turn our attention to the application of next generation sequencing technologies to functional genomics research, particularly focusing on transcriptomics and epigenomics. We end with a discussion of future prospects that next-generation technologies hold for functional genomics research.

Advances in DNA sequencing technologies

The landmark publication of the late 1970s by Sanger's and Gilbert's groups.[3,4] and notably the development of the chain termination method by Sanger and colleagues.[5] established the groundworks for decades of sequence driven research that followed. The chain termination method published in 1977.[5] also commonly reffered to as Sanger or dideoxy sequencing, has remained the most commonly used DNA sequencing technique to date and was used to complete human genome sequencing initiatives led by the international human genome sequencing consortium and celera genomics.[6-8] Very recently, the sanger method has been partially supplanted by several ''next-generation'' sequencing technologies that offer dramatic increases in cost-effective sequence throughput, albeit at the expense of read lengths. The next generation technologies commercially available today include the 454 GS20 pyrosequencing based instrument (Roche Applied Science), the Solexa 1G analyzer (illumina,Inc.), the SOLiD instrument from applied biosystems, and the Heliscope from helicos,Inc. As of this writing, information on the performance of the Heliscope in functional genomics applications is lacking, and so we have restricted our comments to the 454, SOLiD, and 1G sequencing platforms. These new technologies as well as the current state-of-the-art Sanger sequencing platform are. For a review of the history of DNA sequencing the reader is referred to,[9] for a more comprehensive review of emerging sequencing technologies see.[10]

Sanger sequencing: Since its initial report in 1977, the sanger sequencing method has remained conceptually unchanged. The method is based on the DNA polymerase-dependent synthesis of a complementary DNA strand in the presence of natural 2'-

deoxynucleotides (dNTPs) and 2',3'-dideoxynucleotides (ddNTPs) that serve as nonreversible synthesis terminators.[5] The DNA synthesis reaction is randomly terminated whenever a ddNTP is added to the growing oligonucleotide chain, resulting in truncated products of varying lengths with a appropriate ddNTP at their 3' terminus. The product are separated by size using polyacrylamide gel electrophoresis and the terminal ddNTPs are used to reveal the DNA sequence of the template strand. Originally, four different reactions were required per template, each reaction containing a different ddNTP terminator, ddATP,ddCTP, or ddGTP. However, advances in fluorescence detection have allowed for combining the four terminators into one reaction by having them labeled with fluorescent dyes of different colors.[11,12]. Subsequent advances have replaced the original slab gel electrophoresis with capillary gel electrophoresis, thereby enabling much higher electric fields to be applied to the separation matrix. One effect of this advance was to enhance the rate at which fragments could be separated.[13] The overall throughput of capillary electrophoresis was further increased by the advent of capillary arrays whereby many samples could be analyzed in parallel.[14] In addition, breakthroughs in polymer biochemistry, including the development of linear polyacrylamide.[15] and polydimethylacrylamide.[16] have allowed the reuse of capillaries in multiple electrophoretic runs, thus further increasing sequencing efficiency. For further reading on improvements in Sanger sequencing research the reader is referred to.[10] and.[17] These and many other advances in sequencing technology contributed to the relatively low error rate, long read length, and robust characteristics of modern Sanger sequencers. For instance, a commonly used automated high throughput Sanger sequencing instrument from applied biosystems, the ABI 3730xl, has a 96-capillary array format and is capable of producing 900 or more PHRED20.[18] bp per read for a total of up to 96kb for a 3-h run. However, despite the many advances in chemistries and the robust performance of instruments like the 3730xl, the application of relatively expensive Sanger sequencing to large sequencing projects has remained beyond the means of the typical grant-funded investigator. This is a limitation that has been apparently successfully addressed, to varying degrees, by all of the latest technology offerings.

454/Roche sequencing technology: pyro sequencing in high-density pico liter reactors.

An inherent limitation of Sanger sequencing is the requirement of in vivo amplification of DNA fragments that are to be sequenced, which is usually achieved by cloning into bacterial hosts. The cloning step is prone to host-related biases, is lengthy, and is quite labor intensive.[2] The 454 technology.[19] the first next-generation sequencing technology released to the market, circumvents the cloning requirement by taking advantage of a highly efficient in vitro DNA amplification method known as emulsion PCR.[20] In emulsion PCR, individual DNA fragment carry streptavidin beads, obtained through shearing the DNA and attaching the fragments to the beads using adapters, are captured into separate emulsion droplets. The droplets acts as individual amplification reactors, producing clonal copies of a unique DNA template per bead.[19] Each template-containing bead is subsequently transferred into a well of a picotiter plate and the clonally related templates are analyzed using a pyrosequencing reaction. The use of the picotiter plate allows hundreds of thousands of pyrosequencing reactions to be carried out in paralle, massively increasing the sequencing throughput.[19] The pyrosequencing approach.[21,22] is a sequencing-by-synthesis technique that measures the release of inorganic pyrophosphate (PPi) by chemiluminescence. The template DNA is immobilized, and solutions of dNTPs are added one at a time; the release of PPi, whenever the complementary nucleotide is incorporated, is detectable by light produced by a chemiluminescent signal intensity is proportional to the amount of pyrophosphate released and hence the number of bases incorporated, the pyrosequencing approach is prone to errors that result from incorrectly estimating the length of homopolymeric sequence stretches(i.e.,indels). The current state-of-the-art 454 platform marketed by Roche Applied Science is capable of generating 80-120 Mb of sequence in 200-to 300-bp reads in a 4-h run. The 454 technology has been the most widely published next-generation technology, having so far been featured in more than 100 research publications(Roche Applied Sciences).

Illumina: sequencing by synthesis of single-molecule arrays with reversible terminators.

The illumina/solexa approach.[23-25] achieves cloning-free DNA amplification by attaching single-stranded DNA fragments to a solid surface known as a single molecule array, or flow cell, and conducting solid-phase bridge amplification of single molecule DNA templates (illumina,Inc.). In this process, one end of single DNA molecule is attached to a solid surface using a adapter; the molecules subsequently bend over and hybridize to complementary adapters(creating the "bridge"), thereby forming the template for the synthesis of their complementary strands. After the amplification step, a flow cell with more than 40 million clusters is produced, wherein each cluster is composed of approximately 1000 clonal copies of a single template molecule. The templates are sequenced in a massively parallel fashion using a DNA sequencing-by-synthesis approach that employs reversible terminators by removable fluorescent moieties and special DNA polymerases that can incorporate these terminators into growing oligonucleotides chain. The terminators are labeled with fluors of four different colors to distinguish among the different bases at the given sequence position and the template sequence of each cluster is deduced by reading off the color at each successive nucleotide addition step.

Although the illumina approach is more effective at sequencing homopolymeric stretches than pyrosequencing, it produces shorter sequence reads.[25] and hence cannot resolve short sequence repeats. In addition, due to the use of modified DNA polymerases and reversible terminators, substitution errors have been noted in illumina sequencing data.[9]Typically the 1G genome analyzer from illumina,Inc., is capable of generating 35-bp reads and producing at least 1GB of sequence per run in 2-3 days.

ABI/SOLiD: massively parallel sequencing by ligation
Massively parallel sequencing by hybridization-ligation, implemented in the supported oligonucleotide ligation and detection system (SOLiD) from Applied Biosystems, has recently become available. The ligation chemistry used in SOLiD is based on the polony sequencing technique that was published in the same year as the 454 method.[26] Construction of sequencing libraries for analysis on the SOLiD instrument begins with an emulsion PCR single-molecule amplification step similar to that used in 454 technique. The amplification products are transferred onto a glass surface where sequencing occurs by sequential rounds of hybridization and ligation with 16 dinucleotides combinations labeled by four different fluorescent dyes (each dye used to labellel four dinucleotides). Using the four dye encoding scheme, each position is effectively probed twice, and the identity of the nucleotide is determined by analyzing the color that results from two successive ligation reactions. Significantly the two base encoading scheme enables the distinction between a sequencingerror and a sequencing polymorphism: an error would be detected in only a particular ligation reaction, whereas a polymorphism would be detected in both. The newly released SOLiD instrument is capable of producing 1-3 GB of sequence data in 35bp reads per an 8-day run.

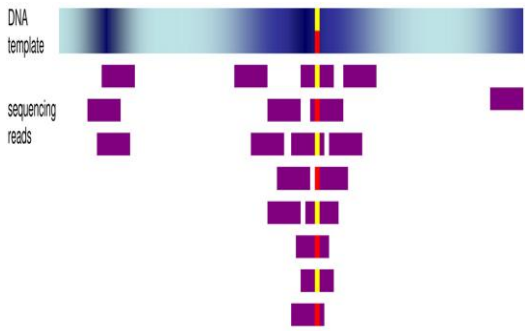

Fig.1. Sequence census approach. In the sequence census approach used in next-generation sequencing, short reads are mapped to the template molecule to provide three types of information. Sequence data are used to reveal sequence polymorphisms in the template, eSNPT(red and yellow), the abundance of reads is used as a quantitative measure of the abundance of the template, and the particular areas of the template covered by reads reveal the internal structure of the template, eg the presence of exons and introns.

POTENTIAL USES OF NGS IN CLINICAL PRACTICE.
There are numerous opportunities to use NGS in clinical practice to improve patient care, including.

NGS captures a broader spectrum of mutations than sanger sequencing.

The spectrum of DNA variation in a human genome comprises small base changes (substitutions), insertions and deletions of DNA, large genomic deletions of exons or whole genes and rearrangements such as inversions and translocations. Traditional sanger sequencing is restricted to the discovery of substitutions and small insertions and deletions. For the remaining mutations dedicated assays are frequently performed, such as florescence in situ hybridization for conventional karyotyping, or comparative genomic hybridization microarrays to detect submicroscopic chromosomal copy number changes such as microdeletions. However, these data can also be derived from ngs sequencing data directly, obviating the need for dedicated assays while harvesting the full spectrum of genomic variation in a single experiment. The only limitations reside in regions which sequence poorly or map erroneously due to extreme guanine/cytosine content or repeat architecture, for example, the repeat expansions underlying fragile x syndrome, or Huntington's disease.

GENOMES CAN BE INTERROGATED WITHOUT BIAS
Capillary sequencing depends on preknowledge of the gene or locus under investigation. However, NGS is completely unselective and used to interrogate full genomes or exomes to discover entirely novel mutations and disease causing genes. In paediatrics, this could be exploited to unravel the genetic basis of unexplained syndromes. For example a nationwide project, deciphering developmental disorders, running at the welcome trust sanger institute in collaboration with nhs clinical genetics services aims to unravel the genetic basis of unexplained developmental delay by sequencing affected children and their parents to uncover deleterious de novo variants. Allying these molecular data with detailed clinical phenotypic information has been successful in identifying novel genes mutated in affected children with similar clinical features.

The increased sensitivity of NGS allows detection of mosaic mutations.

Mosaic mutations are acquired as a postfertilization event and consequently they present at variable frequency within the cells and tissue of an individual. Capillary sequencing may miss these variants as they frequently present with the a subtlety which falls below the sensitivity of the technology. Ngs sequencing provides far more sensitive read-out and can therefore be used to identify variants which reside in just a few percent of the cell, including mosaic variation. In

addition, the sensitivity of ngs sequencing can be increased further, simply be increasing sequencing depth. This has seen ngs employed for very sensitive investigations such as interrogating foetal dna from maternal blood2 or tracking the lecels of tumour cells from the circulation of cancer patients.3.

The main utility of ngs in microbiology is to replace conventional characterization of pathogens by morphology, staining properties and metabolic criteria with a genomic definition of pathogens. The genomes of pathogens define what they are, may harbor information about drug sensitivity and inform the relationship of different pathogens with each other which can be used to trace sources of infection outbreaks. The last recently received media attention, when ngs was used to reveal and trace an outbreak of methicillin-resistant staphylococcus aureus(mrsa) on a neonatal intensive care unit in uk.4 what was most remarkable was that routine microbiology surveillance did not show that the cases of mrsa that occurred over several months were related. Ngs of the pathogens, however, allowed precise characterization of the marsa isolates and revealed a protracted outbreak of mrsa which could be traced to a single member of staff.

Oncology

The fundamental premise of cancer genomics is that cancer is caused by somatically acquired mutations, and consequently it is a disease of the genome. Although capillary based cancer sequencing has been ongoing for over a decade, these investigations were limited to relatively few samples and small number of candidate genes. With the advent of ngs, cancer genomes can now be systemically studied in their entirety, an endeavour ongoing through several large scale cancer genome projects around the world, including a dedicated paediatric cancer genome project. For the child suffering from cancer this may provide many benefits including a more precise diagnosis and classification of the disease, more accurate prognosis, and potentially the identification of ‘drug-able’ causal mutations. Individual cancer sequencing may, therefore, provide the basis of personalized caner management. Currently pilot projects are underway using ngs of cancer genomes in clinical practice, mainly aiming to identify mutations in tumours that can be targeted by mutation specific drugs.

Applications of next generation sequencing for the analysis of epigenetic modifications of histones and dna.

Epigenetics is the study of heritable gene regulation that does not involve the dna sequence. The two major types of epigenetic modifications regulating gene expression are dna methylation by covalent modification of cytosine-5’ and posttransilational modification of histone tails. Regulatory rna provide another means of epigenetic regulation of gene expression; however, the focus of the sections is on applications of next generation sequencing to the analysis of covalent modifications of dna and chromatin. Recent research has implicated such epigenetic modifications of prime importance in oncogenesis and development, setting the grounds of human epigenetic project(hep) initiative, which aims to catalogue dna mathylation on a genome wide scale. To date, these technologies have been applied in several epigenomic areas, including the characterization of dna methylation patterns, post translational modification of histones, and nucleosome positioning on a genome wide scale.

NGS methods: next generation sequencing platforms enable a wide variety of methods, allowing researchers to ask virtually any question related to the genome, transcriptome and epigenome of any organism. Sequencing methods differ primarily by how the dna or rna samples are obtained(eg, organism, tissue type, normal vs affected, experimental conditions) and by the data analysis options used. After the sequencing libraries are prepared, the actual sequencing stage remains fundamentally the same regardless of the method. there are a number of standard library preparation kits that offer protocols for whole genome sequencing, mrna sequencing, targeted sequencing(such as exome sequencing or 16s sequencing), custom-selected regions, protein binding regions and more. Although the number of ngs methods is constantly growing.

1. Genomics
   Whole genome sequencing
   Exome sequencing
   De novo sequencing
   Targeted sequencing
2. Transcriptomics
   Total rna and mrna sequencing
   Targeted rna sequencing
   Small gna and non coading rna sequencing
3. Epigenomics
   Methylation sequencing
   Chip sequencing
4. Ribosome profiling

REFERENCES

1. R.E. Green, et al., (Analysis of one million base pairs of Neanderthal DNA), Nature, 2006; 444: 330–336.
2. N. Hall, (Advanced sequencing technologies and their wider impact in microbiology), J. Exp. Biol, 2007; 210: 1518–1525.
3. F. Sanger, A.R. Coulson, A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, J. Mol. Biol. 1975; 94: 441–448.
4. A.M. Maxam, W. Gilbert, A new method for sequencing DNA, Proc. Natl. Acad. Sci. U. S. A. 1977; 74: 560–564.
5. F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. USA, 1977; 74: 5463–5467.

6. E.S. Lander, et al., Initial sequencing and analysis of the human genome, Nature, 2001; 409: 860–921.
7. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome, Nature, 2004; 431: 931–945.
8. J.C. Venter, et al., The sequence of the human genome, Science, 2001; 291: 1304–1351.
9. C.A. Hutchison III, DNA sequencing: bench to bedside and beyond, Nucleic Acids Res, 2007; 35: 6227–6237.
10. M.L. Metzker, Emerging technologies in DNA sequencing, Genome Res. 2005; 15: 1767–1776.
11. L.M. Smith, et al., Fluorescence detection in automated DNA sequence analysis, Nature, 1986; 321: 674–679.
12. J.M. Prober, et al., A system for rapid DNA sequencing with fluorescent chain terminating dideoxynucleotides, Science, 1987; 238: 336–341.
13. A.S. Cohen, D.R. Najarian, A. Paulus, A. Guttman, J.A. Smith, B.L. Karger, Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis, Proc. Natl. Acad. Sci. U. S. A, 1988; 85: 9660–9663.
14. X.C. Huang, M.A. Quesada, R.A. Mathies, DNA sequencing using capillary array electrophoresis, Anal. Chem. 1992; 64: 2149–2154.
15. M.C. Ruiz-Martinez, J. Berka, A. Belenkii, F. Foret, A.W. Miller, B.L. Karger, DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection, Anal. Chem. 1993; 65: 2851–2858.
16. R.S. Madabhushi, Separation of 4-color DNA sequencing extension products in non covalently coated capillaries using low viscosity polymer solutions, Electrophoresis, 1998; 19: 224–230.
17. E. Carrilho, DNA sequencing by capillary array electrophoresis and micro fabricated array systems, Electrophoresis, 2000; 21: 55–65.
18. B. Ewing, P. Green, Base-calling of automated sequencer traces using phred. II. Error probabilities, Genome Res, 1998; 8: 186–194.
19. M. Margulies, et al., Genome sequencing in microfabricated high-density picolitre reactors, Nature, 2005; 437: 376–380.
20. D.S. Tawfik, A.D. Griffiths, Man-made cell-like compartments for molecular evolution, Nat. Biotechnol, 1998; 16: 652–656.
21. P. Nyren, B. Pettersson, M. Uhlen, Solid phase DNA mini sequencing by an enzymatic luminometric inorganic pyrophosphate detection assay, Anal. Biochem, 1993; 208: 171–175.
22. M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen, P. Nyren, Real-time DNA sequencing using detection of pyrophosphate release, Anal. Biochem, 1996; 242: 84–89.
23. S. Bennett, Solexa Ltd, Pharmacogenomics, 2004; 5: 433–438.
24. S.T. Bennett, C. Barnes, A. Cox, L. Davies, C. Brown, Toward the 1,000 dollars human genome, Pharmacogenomics, 2005; 6: 373–382.
25. D.R. Bentley, Whole-genome re-sequencing, Curr. Opin. Genet. Dev. 2006; 16: 545–552.
26. J. Shendure, et al., Accurate multiplex polony sequencing of an evolved bacterial genome, Science, 2005; 309: 1728–1732.
27. B. Wold, R.M. Myers, Sequence census methods for functional genomics, Nat. Methods, 2008; 5: 19–21.
28. T.T. Torres, M. Metta, B. Ottenwalder, C. Schlotterer, Gene expression profiling by massively parallel sequencing, Genome Res. 2008; 18: 172–177.
29. M.N. Bainbridge, et al., Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach, BMC Genomics, 2006; 7: 246.
30. R.D. Morin, et al., Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells, Genome Res. 2008; 18: 610–621.
31. V.E. Velculescu, L. Zhang, B. Vogelstein, K.W. Kinzler, Serial analysis of gene expression, Science. 1995; 270: 484–487.
32. T.S. Mikkelsen, et al., Genome-wide maps of chromatin state in pluri potent and lineage-committed cells, Nature, 2007; 448: 553–560.
33. E.R. Mardis, ChIP-seq: welcome to the new frontier, Nat. Methods, 2007; 4: 613–614.
34. G. Robertson, et al., Genome-wide profiles of STAT1 DNA association using chromatin immuno precipitation and massively parallel sequencing, Nat. Methods, 2007; 4: 651–657.
35. S. Fields, Molecular biology: site-seeing by sequencing, Science, 2007; 316: 1441–1442.
36. D.S. Johnson, A. Mortazavi, R.M. Myers, B. Wold, Genome-wide mapping of in vivo protein–DNA interactions, Science, 2007; 316: 1497–1502.
37. A. Barski, et al., High-resolution profiling of histone methylations in the human genome, Cell. 2007; 129: 823–837.
38. D.S. Gilmour, J.T. Lis, Detecting protein–DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes, Proc. Natl. Acad. Sci. U. S. A. 1984; 81: 4275–4279.
39. O. Morozova, M.A. Marra, From cytogenetics to next-generation sequencing technologies: advances in the detection of genome rearrangements in tumors, Biochem. Cell Biol, 2008; 86: 81–91.
40. E. Tuzun, et al., Fine-scale structural variation of the human genome, Nat. Genet. 2005; 37: 727–732.
41. E.R. Mardis, Anticipating the 1,000 dollar genome, Genome Biol. 2006; 7: 112.
42. P. Ng, et al., Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes, Nucleic Acids Res. 2006; 34: 84.
43. J.O. Korbel, et al., Paired-end mapping reveals extensive structural variation in the human genome, Science, 2007; 318: 420–426.
44. Y. Ruan, et al., Fusion transcripts and transcribed retrotransposed loci discovered through

comprehensive transcriptome analysis using paired-end ditags (PETs), Genome Res. 2007; 17: 828–838.
45. W. Brockman, et al., Quality scores and SNP detection in sequencing-by-synthesis systems, Genome Res. 2008; 18: 763–770.
46. R.K. Thomas, et al., Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing, Nat. Med, 2006; 12: 852–855.
47. S. Tyagi, Taking a census of mRNA populations with microbeads, Nat. Biotechnol, 2000; 18: 597–598.
48. M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science, 1995; 270: 467–470.
49. S.A. Ness, Microarray analysis: basic strategies for successful experiments, Mol. Biotechnol, 2007; 36: 205–219.
50. S. Brenner, et al., Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays, Nat. Biotechnol, 2000; 18: 630–634.
51. S. Audic, J.M. Claverie, The significance of digital gene expression profiles, Genome Res. 1997; 7: 986–995.
52. K.L. Nielsen, A.L. Hogh, J. Emmersen, DeepSAGE—digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples, Nucleic Acids Res. 2006; 34: e133.
53. S.M. Wang, Understanding SAGE data, Trends Genet. 2007; 23: 42–50.
54. S. Saha, et al., Using the transcriptome to annotate the genome, Nat. Biotechnol. 2002; 20: 508–512.
55. M. Gowda, H. Li, J. Alessi, F. Chen, R. Pratt, G.L. Wang, Robust analysis of 5¢- transcript ends (5¢-RATE): a novel technique for transcriptome analysis and genome annotation, Nucleic Acids Res. 2006; 34: 126.
56. M. Gowda, C. Jantasuriyarat, R.A. Dean, G.L. Wang, Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis, Plant Physiol, 2004; 134: 890–897.
57. M. Seki, et al., Functional annotation of a full-length Arabidopsis cDNA collection, Science, 2002; 296: 141–145.
58. L.D. Hillier, et al., Generation and analysis of 280,000 human expressed sequence tags, Genome Res, 1996; 6: 807–828.
59. W.R.McCombie, et al., Caenorhabditis elegans expressed sequence tags identify gene families and potential disease gene homologues, Nat. Genet. 1992; 1: 124–131.
60. M. Sun, G. Zhou, S. Lee, J. Chen, R.Z. Shi, S.M. Wang, SAGE is far more sensitive than EST for detecting low-abundance transcripts, BMC Genomics, 2004; 5: 1.
61. A.P. Weber, K.L. Weber, K. Carr, C. Wilkerson, J.B. Ohlrogge, Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing, Plant Physiol, 2007; 144: 32–42.
62. M.W. Jones-Rhoades, J.O. Borevitz, D. Preuss, Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins, PLoS Genet. 2007; 3: 1848–1861.
63. F. Cheung, B.J. Haas, S.M. Goldberg, G.D. May, Y. Xiao, C.D. Town, Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology, BMC Genomics, 2006; 7: 272.
64. K. Ohtsu, et al., Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.), Plant J; 2007; 52: 391–404.
65. A.L. Toth, et al., Wasp gene expression supports an evolutionary link between maternal behavior and eusociality, Science, 2007; 318: 441–444.
66. J.C. Vera, et al., Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing, Mol. Ecol. 2008; 17: 1636–1647.
67. A.L. Eveland, D.R. McCarty, K.E. Koch, Transcript profiling by 3¢-untranslated region sequencing resolves expression of gene families, Plant Physiol. 2008; 146: 32–44.
68. W. Filipowicz, S.N. Bhattacharyya, N. Sonenberg, Mechanisms of posttranscriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 2008; 9: 102–114.
69. C. Lu, S.S. Tej, S. Luo, C.D. Haudenschild, B.C. Meyers, P.J. Green, Elucidation of the small RNA component of the transcriptome, Science, 2005; 309: 1567–1569.
70. B.C. Meyers, F.F. Souret, C. Lu, P.J. Green, Sweating the small stuff: microRNA discovery in plants, Curr. Opin. Biotechnol, 2006; 17: 139–146.
71. Z. Xie, et al., Genetic and functional diversification of small RNA pathways in plants, PLoS Biol. 2004; 2: E104.
72. R.C. Lee, R.L. Feinbaum, V. Ambros, The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14, Cell, 1993; 75: 843–854.
73. E. Berezikov, E. Cuppen, R.H. Plasterk, Approaches to microRNA discovery, Nat. Genet. 38 Suppl, 2006; S2–S7.
74. M.J. Axtell, C. Jan, R. Rajagopalan, D.P. Bartel, A two-hit trigger for siRNA biogenesis in plants, Cell, 2006; 127: 565–577.
75. I.R. Henderson, et al., Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning, Nat. Genet. 2006; 38: 721–725.
76. C. Lu, et al., MicroRNAs and other small RNAs enriched in the Arabidopsis RNAdependent RNA polymerase-2 mutant, Genome Res. 2006; 16: 1276–1288.
77. R. Rajagopalan, H. Vaucheret, J. Trejo, D.P. Bartel, A diverse and evolutionarily fluid set of microRNAs

in Arabidopsis thaliana, Genes Dev, 2006; 20: 3407–3425.
78. N. Fahlgren, et al., High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes, PLoS One, 2007; 2: 219.
79. K.D. Kasschau, et al., Genome-wide profiling and analysis of Arabidopsis siRNAs, PLoS Biol. 2007; 5: 57.
80. [80] M.D. Howell, et al., Genome-wide analysis of the RNA-dependent RNA polymerase 6/dicer-like 4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting, Plant Cell, 2007; 19: 926–942.
81. Y. Yao, et al., Cloning and characterization of microRNAs from wheat (Triticum aestivum L.), Genome Biol. 8 2007; R96.
82. A. Barakat, K. Wall, J. Leebens-Mack, Y.J. Wang, J.E. Carlson, C.W. Depamphilis, Large-scale identification of microRNAs from a basal eudicot (Eschscholzia californica) and conservation in flowering plants, Plant J, 2007; 51: 991–1003.
83. T. Zhao, et al., A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii, Genes Dev. 2007; 21: 1190–1203.
84. J. Burnside, et al., Marek's disease virus encodes microRNAs that map to meq and the latency-associated transcript, J. Virol. 2006; 80: 8778–8786.
85. E. Berezikov, et al., Diversity of microRNAs in human and chimpanzee brain, Nat. Genet. 2006; 38: 1375–1377.
86. N.C. Lau, et al., Characterization of the piRNA complex from rat testes, Science, 2006; 313: 363–367.
87. A. Girard, R. Sachidanandam, G.J. Hannon, M.A. Carmell, A germline-specific class of small RNAs binds mammalian Piwi proteins, Nature, 2006; 442: 199–202.
88. S. Houwing, et al., A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish, Cell, 2007; 129: 69–82.
89. M.R. Brent, Steady progress and recent breakthroughs in the accuracy of automated genome annotation, Nat. Rev. Genet; 2008; 9: 62–73.
90. M.R. Emmert-Buck, et al., Laser capture microdissection, Science, 1996; 274: 998–1001.
91. S.J. Emrich, W.B. Barbazuk, L. Li, P.S. Schnable, Gene discovery and annotation using LCM-454 transcriptome sequencing, Genome Res. 2007; 17: 69–73.
92. W.B. Barbazuk, S.J. Emrich, H.D. Chen, L. Li, P.S. Schnable, SNP discovery via 454 transcriptome sequencing, Plant J. 2007; 51: 910–918.
93. D. Hanahan, R.A. Weinberg, The hallmarks of cancer, Cell, 2000; 100: 57–70.
94. P.A. Callinan, A.P. Feinberg, The emerging science of epigenomics, Hum. Mol. Genet. 15 Spec. No. 1 (2006) R95–R101.
95. M. Esteller, The necessity of a human epigenome project, Carcinogenesis, 2006; 27: 1121–1125.
96. F. Eckhardt, et al., DNA methylation profiling of human chromosomes 6, 20 and 22, Nat. Genet. 2006; 38: 1378–1385.
97. K.H. Taylor, et al., Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing, Cancer Res. 2007; 67: 8511–8518.
98. S.J. Cokus, et al., Shotgun bisulfite sequencing of the Arabidopsis genome reveals DNA methylation patterning, Nature, 2008; 452: 215–219.
99. D.E. Schones, K. Zhao, Genome-wide approaches to studying chromatin modifications, Nat. Rev. Genet. 2008; 9: 179–191.
100. M.H. Kuo, C.D. Allis, In vivo cross-linking and immunoprecipitation for studying dynamic protein:DNA associations in a chromatin environment, Methods, 1999; 19: 425–433.
101. B. Ren, et al., Genome-wide location and function of DNA binding proteins, Science, 2000; 290: 2306–2309.
102. T.Y. Roh, W.C. Ngau, K. Cui, D. Landsman, K. Zhao, High-resolution genome-wide mapping of histone modifications, Nat. Biotechnol. 2004; 22: 1013–1016.
103. A.A. Bhinge, J. Kim, G.M. Euskirchen, M. Snyder, V.R. Iyer, Mapping the chromosomal targets of STAT1 by sequence tag analysis of genomic enrichment (STAGE), Genome Res. 2007; 17: 910–916.
104. I. Albert, et al., Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome, Nature, 2007; 446: 572–576.
105. S.M. Johnson, F.J. Tan, H.L. McCullough, D.P. Riordan, A.Z. Fire, Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin, Genome Res. 2006; 16: 1505–1516.
106. C.D. Schmid, P. Bucher, ChIP-Seq data reveal nucleosome architecture of human promoters, Cell 131 (2007) 831–832 author reply 832-833.
107. H. Boeger, J. Griesenbeck, J.S. Strattan, R.D. Kornberg, Nucleosomes unfold completely at a transcriptionally active promoter, Mol. Cell, 2003; 11: 1587–1598.
108. S. Nagel, et al., Activation of TLX3 and NKX2-5 in t(5;14)(q35;q32) T-cell acute lymphoblastic leukemia by remote 3¢-BCL11B enhancers and coregulation by PU.1 and HMGA1, Cancer Res. 2007; 67: 1461–1471.
109. J. Dostie, et al., Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements, Genome Res, 2006; 16: 1299–1309.
110. D.W. Deamer, M. Akeson, Nanopores and nucleic acids: prospects for ultra rapid sequencing, Trends Biotechnol, 2000; 18: 14.