doi:10.20944/preprints202011.0356.v1

# The Generations of Sequencing Technologies: An Overview

*[1]Ayobami A. L., [1]Kade E. A., [1]Iyebeye M. I., [2]Oladimeji K. A., [3]Kehinde S. and [4]Oladipupo T. A.

[1] Department of Microbiology, University of Lagos (UNILAG), Akoka, Lagos, Nigeria.

[2] National University of Ireland Galway, Ireland and University of Limerick

[3] Department of Cell biology and Genetics, University of Lagos (UNILAG), Akoka, Lagos, Nigeria.

[4] Department of Science and Technology Education, Obafemi Awolowo University (OAU), Ile-Ife, Nigeria.
* **Correspondence**: ayobamiadesiyan123@gmail.com; ORCID: 0000-0001-8240-5754

## Abstract

The world has now entered into a replacement era of genomics due to the continued advancements within the next generation high throughput sequencing technologies, which incorporates sequencing by synthesis-fluorescent in place sequencing (FISSEQ), pyrosequencing, sequencing by ligation using polony amplification, supported oligonucleotide detection (SOLiD), sequencing by hybridization alongside sequencing by ligation, and nanopore technology. Great impacts of those methods are often seen for solving the genome related problems of plant and Animalia which will open the door of a replacement era of genomics. This might ultimately overcome the Sanger sequencing that ruled for 30 years. NGS is predicted to advance and make the drug discovery process more rapid.

## Introduction

In 1977 Sanger et al. published two methodology based papers on the rapid determination of DNA sequences that helped to rework biology and provided a replacement tool for deciphering complete genes and later the whole genome [1, 2]. The methods dramatically improved existing DNA sequencing techniques developed by Maxam and Gilbert published within the same year and Sanger and Coulson's own "plus and minus" method published 2 years earlier [3, 4]. The advantage of handling less toxic chemicals and radioisotopes made "Sanger sequencing" the sole DNA sequencing method used for subsequent 30 years. The gel based sequencing technology has undergone dramatic improvement in throughput from one parallelization, automation, and refinement of sequencing methods and chemistry. Recent advances in microfibration have resulted in further improvements of Sanger sequencing by multiplexing and miniaturization. These advances are reviewed by Metzker [5]. Despite many improvements, the gel based Sanger sequencing technology still faces drawbacks within the name of cost and low throughput. For achieving high throughput, many commercial companies and scientific labs have come up with alternative ways of high throughput sequencing with an inexpensive cost. The technologies named together as next generation sequencing technologies include sequencing by synthesis developed by 454 Life-Science [6]; sequencing by ligation; sequencing by hybridization; single molecule DNA sequencing; nanopore sequencing; and multiplex polony sequencing of George Church's lab [7]. Next generation sequencing has revolutionary impacts on genetic applications like metagenomics, comparative genomics, high throughput polymorphism detection, analysis of small RNAs, mutation screening, transcriptome profiling, methylation profiling, and chromatin remodeling. The main measurements for the success of subsequent generation technology are sequence (read length), sequence quality, high throughput, and low cost.

## 1.1. Sequencing by Synthesis (SBS)

SBS using fluorophore-labeled, reversible-terminator nucleotides is the most common platform of sequencing by synthesis. It is sometimes named "fluorescent in situ sequencing" (FISSEQ). The pyrosequencing technology is another SBS technology developed by Ronaghi et al., at Stanford University [8, 9]. It is based on the detection of pyrophosphate (PPi) released during DNA synthesis when inorganic PPi is released after nucleotide incorporation by DNA polymerase. The released PPi is then converted to ATP by ATP sulfurylase. A luciferase reporter enzyme uses the ATP to generate light, which is then detected by a charged couple device (CCD) camera. Pyrosequencing has evolved into an ultrahigh throughput sequencing technology with the combination of several technologies such as template carrying microbeads deposited in microfabricated picoliter-sized reaction wells connecting to optical fibers [10]. 454 Life Sciences/Roche Diagnostic has Genome Sequencer 20 System and Genome Sequencer FLX System, two high throughput commercial sequencing platforms, and DNA helices are fractionated into 300–500 bp fragments and linkers are added to their 3' and 5' ends. Single stranded DNA is isolated and captured on beads. The beads with DNAs are then emulsified in a "water-in-oil" mixture with amplification reagents to create micro reactors for Emulsion PCR (emPCR). Finally beads with amplified DNAs are loaded onto a picotitre plate for sequencing.

## 1.2. Polony Sequencing

George Church's Laboratory at Harvard University has developed a high throughput technology based on polony amplification and FISSEQ [11]. Polony amplification is a method to amplify DNA in situ on a thin polyacrylamide film [12]. The DNA movement is limited in the polyacrylamide gel, so the amplified DNAs are localized in the gel and form the so-called "polonies," that is, polymerase colonies (Figure 1). Up to 5 million polonies (i.e., 5 million PCRs) can be formed on a single glass microscope slide. This sequencing technology has been tested in a bacterial genome and the sequence read length was found to be about 13 bases per colony [13]. Kim et al. developed a technology named Polony Multiplex Analysis of Gene Expression (PMAGE), which combines polony amplification and a sequence-by-ligation method, to sequence 14-base tags [14]. Up to 5 million polonies can be sequenced in parallel.

## 1.3. Single Molecule DNA Sequencing

Most current sequencing technologies are based on sequencing many identical copies of DNA molecules (often amplified). However, there are certain problems associated with sequencing amplified multiple copies of identical DNAs such as achieving the synchronous priming of each copy of the multiple DNA by the sequencing primers [15]. One way to solve such problems is to perform sequencing by synthesis method [16]. DNA can be attached to solid support to form single molecule arrays, and the single DNA molecule is then sequenced directly. Buzby at Helicos Bio-Science Corp. also invented a method of stabilizing a nucleic acid duplex on a surface for single molecule sequencing [17]. Applera Corp. invented fluorescent intercalators that are employed as a donor in fluorescence resonance energy transfer (FRET) for use in single molecule sequencing reactions [18].

## 1.4. Nanopore Sequencing

Nanopores are nanometer-scale pores and are considered one among the foremost promising technologies in achieving true real time, ultrafast, true single molecule DNA sequencing [19]. Nanopore has been used for the detection, counting, and characterization of single molecules by observing the changes in ionic current in nanopores when molecules traverse through a nanopore [20, 21]. The fabrication of nanopores like the alpha-hemolysin pore and artificial nanopore has been reviewed recently [22]. Recent progress has made an enormous step toward ultrafast sequencing using nanopore technologies. Lagerqvist et al. proposed a completely unique idea to live the electrical current perpendicular to the DNA backbone [23]. Zhao et al. reported that one nucleotide polymorphism are often detected by a change within the threshold voltage of a nanopore [24]. Biological nanopore sequencing systems have several fundamental characteristics that make them advantageous as compared with solid state systems- with each advantageous characteristic of this design approach stemming from the incorporation of proteins into their technology. Uniform pore structure, the precise control of sample translocation through pore channels, and even the detection of individual nucleotides in samples can be facilitated by unique proteins from a variety of organism types.

## 1.5. Sequencing by Hybridization (SBH)

SBH was given in the late 80s [25]. It is a method of reconstructing a DNA sequence based on its k-mer content. All possible k-nucleotides oligomers (k-mers) are hybridized to identify overlapping k-mers in an unknown DNA sample. These overlapping k-mers are subsequently aligned by algorithms to produce the DNA sequence. Drmanac et al. reviewed the technology of sequencing by hybridization [26]. In spite of having some problems, using SBH with predefined probe sets derived from a known sequence has been used to resequence a specific region of genomic DNA or cDNA for the identification of small deletions, insertions, polymorphisms, and mutations [27].

### 1.6. Sequencing by Ligation

The first high throughput sequencing by ligation was probably realized by Lynx Therapeutics Inc. in their early technology of Massively Parallel Signature Sequencing [28]. Church's laboratory at Harvard Medical School advanced the sequencing by ligation method for ultra-fast and high throughput sequencing [13, 14]. They converted an epifluorescence for rapid DNA sequencing. DNA molecules were amplified in parallel onto micro beads by emulsion polymerase chain reaction (Figure 1). Millions of beads were immobilized in a polyacrylamide gel and sequenced using sequencing by ligation method. Applied Biosystem Inc. acquired Agencourt Personal Genomics who developed the Supported Oligo Ligation Detection system (SOLiD) for high throughput DNA sequencing.

### 1.7. Bioinformatics Tools for Analyzing NGS Data

Sequence reads generated from NGS technologies are shorter than traditional Sanger sequences, which makes assembly and analysis of NGS data challenging. Some important bioinformatics tools are available which is helpful in analyzing NGS data.

### 1.8. Applications of NGS Techniques

NGS technologies have already been used for various applications, ranging from whole genome sequencing, resequencing, single nucleotide polymorphism (SNP), structural variation discovery, mRNA and noncoding RNA profiling, and protein-nucleic acid interaction assays. NGS technologies are becoming a potential tool for gene expression analysis; especially for those species having reference genome sequences already available (Figure 2). An overview of NGS applications that are relevant to drug discovery is discussed here.

### 1.2. Polony Sequencing

George Church's Laboratory at Harvard University has developed a high throughput technology based on polony amplification and FISSEQ [11]. Polony amplification is a method to amplify DNA in situ on a thin polyacrylamide film [12]. The DNA movement is limited in the polyacrylamide gel, so the amplified DNAs are localized in the gel and form the so-called "polonies," that is, polymerase colonies (Figure 1). Up to 5 million polonies (i.e., 5 million PCRs) can be formed on a single glass microscope slide. This sequencing technology has been tested in a bacterial genome and the sequence read length was found to be about 13 bases per colony [13]. Kim et al. developed a technology named Polony Multiplex Analysis of Gene Expression (PMAGE), which combines polony amplification and a sequence-by-ligation method, to sequence 14-base tags [14]. Up to 5 million polonies can be sequenced in parallel.

### 1.3. Single Molecule DNA Sequencing

Most current sequencing technologies are based on sequencing many identical copies of DNA molecules (often amplified). However, there are certain problems associated with sequencing amplified multiple copies of identical DNAs such as achieving the synchronous priming of each copy of the multiple DNA by the sequencing primers [15]. One way to solve such problems is to perform sequencing by synthesis method [16]. DNA can be attached to solid support to form single molecule arrays, and the single DNA molecule is then sequenced directly. Buzby at Helicos Bio-Science Corp. also invented a method of stabilizing a nucleic acid duplex on a surface for single molecule sequencing [17]. Applera Corp. invented fluorescent intercalators that are employed as a donor in fluorescence resonance energy transfer (FRET) for use in single molecule sequencing reactions [18].

## 1.4. Nanopore Sequencing

Nanopores are nanometer-scale pores and are regarded as one of the most promising technologies in achieving true real time, ultrafast, true single molecule DNA sequencing [19]. Nanopore has been used for the detection, counting, and characterization of single molecules by observing the changes in ionic current in nanopores when molecules traverse through a nanopore [20, 21]. The fabrication of nanopores such as the alpha-hemolysin pore and synthetic nanopore has been reviewed recently [22]. Recent progress has made a big step toward ultrafast sequencing using nanopore technologies. Lagerqvist et al. proposed a novel idea to measure the electric current perpendicular to the DNA backbone [23]. Zhao et al. reported that a single nucleotide polymorphism can be detected by a change in the threshold voltage of a nanopore [24].

## 1.5. Sequencing by Hybridization (SBH)

SBH was given within the late 80s [25]. It’s a way of reconstructing a DNA sequence supported its k-mer content. All possible k-nucleotides oligomers (k-mers) are hybridized to spot overlapping k-mers in an unknown DNA sample. These overlapping k-mers are subsequently aligned by algorithms to supply the DNA sequence. Drmanac et al. reviewed the technology of sequencing by hybridization [26]. In spite of getting some problems, using SBH with predefined probe sets derived from a known sequence has been wont to resequence a selected region of genomic DNA or cDNA for the identification of small deletions, insertions, polymorphisms, and mutations [27].

## 1.6. Sequencing by Ligation

The first high throughput sequencing by ligation was probably realized by Lynx Therapeutics Inc. in their early technology of Massively Parallel Signature Sequencing [28]. Church’s laboratory at

Harvard school of medicine advanced the sequencing by ligation method for ultra-fast and high throughput sequencing [13, 14]. They converted an epifluorescence for rapid DNA sequencing. DNA molecules were amplified in parallel onto micro beads by emulsion polymerase chain reaction. Many beads were immobilized during a polyacrylamide gel and sequenced using sequencing by ligation method. Applied Biosystem Inc. acquired Agencourt Personal Genomics who developed the Supported Oligo Ligation Detection system (SOLiD) for top throughput DNA sequencing.

## 1.7. Bioinformatics Tools for Analyzing NGS Data

Sequence reads generated from NGS technologies are shorter than traditional Sanger sequences, which makes assembly and analysis of NGS data challenging. Some important bioinformatics tools are available which is helpful in analyzing NGS data.

## 1.8. Applications of NGS Techniques

NGS technologies have already been used for various applications, ranging from whole genome sequencing, resequencing, single nucleotide polymorphism (SNP), structural variation discovery, mRNA and noncoding RNA profiling, and protein-nucleic acid interaction assays. NGS technologies are becoming a potential tool for gene expression analysis, especially for those species having reference genome sequences already available.

## 1.9. Single Nucleotide Polymorphisms (SNP) Discovery

SNPs are important genomic resources which may be utilized in a spread of analyses including physical characteristics like height and appearance also as less obvious traits like personality, behavior, and disease susceptibility. Sequence data generated for parental genotypes of the mapping population by using NGS technologies are often used for mining the SNPs at large scale [29]. SNPs also can significantly influence responses to pharmacotherapy and predicting whether the drugs will produce adverse reactions or not. The event of latest drugs are often made far cheaper and more rapid by selecting participants in drug trials supported their genetically determined response to the drugs [30].

## 1.10. Messenger RNA and Noncoding RNA Profiling

Apart from SNP discovery, expressed regions of a genome are often detected using NGS technologies. Subsequent generation sequencing platforms are capable of identifying expression levels of nearly all genes, including those rare and species specific transcripts. an identical approach are often applied to large genomes. RNA-seq data are often wont to characterize exon-exon splicing events namely cases of other splicing [31–33]. Noncoding RNA, like microRNA (miRNA), may be a broad class of regulatory RNA molecules. The NGS technologies are useful for discovery of noncoding RNA for his or her short lengths [34]. Most studies so far have used 454 technologies, due to its early availability to get new and different noncoding RNA classes in several species like Chlamydomonas, Drosophila, Arabidopsis, then forth [35–37]. A comprehensive study of miRNA in acute chronic myelocytic leukemia was performed using NGS, with novel findings of differentially expressed miRNAs [38]. Transcriptome sequencing has been found to be a strong tool for detecting novel gene fusions in neoplastic cell lines and tissues [39].

### 1.11. De Novo Sequencing and Resequencing

Metagenomics is defined because the application of genomics techniques to directly study the communities of microbial organisms without isolation and cultivation of individual species [40]. It involves the characterization of the genomes in these communities, also as their mRNA, protein, and metabolic products. Subsequent generation sequencing technologies have enabled to maneuver metagenomics from one organism type in isolation to the studies of whole communities. NGS enables the researchers to avoid the cloning formation and culture steps which are the main drawbacks of genomics. NGS strategies are straightforward within the following: (1) deep sequencing of DNA fragments is conducted on an uncultured sample and (2) short reads are compared against database of known sequences using bioinformatics tool like MEGAN [41] with or without assembly, and (3) these data are then wont to compute and explore their contents to infer relative abundances. Therefore, NGS technologies are often a potent tool for discovery of microorganisms and pathogens.

### 1.12. Regulatory Protein Binding

At low throughput, chromatin immunoprecipitation (ChIP) has enabled regulatory DNA-protein binding interactions to be elucidated [41]. It's a lengthy process like association of specific antibody to DNA-binding protein, followed by another protein-DNA cross linking agent, in order that any protein in close association with DNA becomes linked. Then, the cells are lysed, DNA is fragmented, and therefore the specific antibody is employed to precipitate the protein of interest alongside any associated DNA fragment. These DNA pieces are subsequently released by reversing the cross linking and identified by southern blotting or qPCR [42]. a search is employed to infer the DNA-binding site sequence of the protein under study. So, NGS helps in identifying the regulatory protein binding site for having shorter reads and specific binding site and it also provides higher resolution [43].

### 1.13. Exploring Chromatin Packaging

Chromatin packaging denotes the packing of DNA in histones. This packaging determines the transcription of a specific gene. Understanding of this DNA packing, that is, chromatin packaging is of an excellent interest. An initial 454-based study of genomic DNA packaging into nucleosomes was described for the C. elegans, by sequencing the DNA isolated from nucleosome cores after micrococcal nuclease digestion and mapping them to the reference genome sequence [44]. Mikkelsen et al. used the Illumina platform to demonstrate the connection between chromatin packaging and organic phenomenon in several different cell types. Mikkelsen et al. found that changes in chromatin state at specific promoters reflect changes in organic phenomenon for the genes they control. a far better understanding of the chromatin packaging will provide new strategies for the event of novel drugs [45]. NGS also can have a central role within the discovery of latest genomic biomarkers, detection of mutations, personalized medicine and pharmacogenetics, target identification and validation, clinical diagnostics, vaccine development, investigating drug resistance and lots of others

## 2. Conclusion

Because of their cost-effectiveness as compared to the Sanger sequencing method and with a good range of uses NGS approaches have emerged because the dominant genomics technology. Perhaps most importantly, these new sequencers have provided genome-scale sequencing capacity to individual laboratories additionally to larger genome centers. Compared to Sanger sequencing, advantages of subsequent generation technologies mentioned so far, including 454/Roche Illumina/Solexa and ABI/SOLiD, alleviate the necessity for in vivo cloning by clonal amplification of spatially separated single molecules using either emulsion PCR (454/Roche and ABI/SOLiD) or bridge amplification on solid surface (Illumina/Solexa). We are at the verge of a replacement genomic revolution with recent advances in next generation sequencing technologies. we'll see enormous impacts of those next generation sequencing methods in handling complex biological problems, for instance, the identification of all sequence changes in drug resistant HIV and drug resistant tuberculosis bacteria (TB), within the delineation of sequence changes for individual cells during cancer initiation and progression, and in global transcription factor mapping ChIP to sequencing [46]. Additionally, NGS technology will create an incredible impact in life science. It'll make routine sequencing of a person's genome sequence of a private human and therefore the initiation of the private Human Genome Project [47].

### Abbreviations

SBS: Sequencing by synthesis
FISSEQ: Fluorescent in situ sequencing
CCD: Charged couple device
emPCR: Emulsion PCR
PMAGE: Polony multiplex analysis of gene expression
FRET: Fluorescence resonance energy transfer
SBH: Sequencing by hybridization
k-mers: k-nucleotides oligomers
SOLiD: Supported oligo ligation detection
SNP: Single nucleotide polymorphism
ChIP: Chromatin immunoprecipitation.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

1. F. Sanger, G. M. Air, B. G. Barrell et al., “Nucleotide sequence of bacteriophage φX174 DNA,” Nature, vol. 265, no. 5596, pp. 687–695, 1977.
2. F. Sanger, S. Nicklen, and A. R. Coulson, “DNA sequencing with chain-terminating inhibitors,” Proceedings of the National Academy of Sciences of the United States of America, vol. 74, no. 12, pp. 5463–5467, 1977.
3. A.M. Maxam and W. A. Gilbert, “New method for sequencing DNA,” Proceedings of the National Academy of Sciences of the United States of America, vol. 74, no. 2, pp. 560–564, 1977.
4. F. Sanger and A. R. Coulson, “Nucleotide and amino acid sequences of Gene G of φX174,” Journal of Molecular Biology, vol. 94, pp. 441–448, 1975.
5. M. L. Metzker, “Emerging technologies in DNA sequencing,” Genome Research, vol. 15, no. 12, pp. 1767–1776, 2005.
6. M. Margulies, M. Egholm, E. William et al., “Genome sequencing in microfabricated high-density picolitre reactors,” Nature, vol. 437, pp. 376–380, 2005.
7. J. J. Shendure, G. J. Porreca, N. B. Reppas et al., “Molecular biology: accurate multiplex polony sequencing of an evolved bacterial genome,” Science, vol. 309, no. 5741, pp. 1728–1732, 2005.
8. M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlén, and P. Nyrén, “Real-time DNA sequencing using detection of pyrophosphate release,” Analytical Biochemistry, vol. 242, no. 1, pp. 84–89, 1996.
9. M. Ronaghi, M. Uhlén, and P. Nyrén, “A sequencing method based on real-time pyrophosphate,” Science, vol. 281, no. 5375, pp. 363–365, 1998.
10. J. Shendure, R. D. Mitra, C. Varma, and G. M. Church, “Advanced sequencing technologies: methods and goals,” Nature Reviews Genetics, vol. 5, no. 5, pp. 335–344, 2004.
11. J. Shendure and H. Ji, “Next-generation DNA sequencing,” Nature Biotechnology, vol. 26, no. 10, pp. 1135–1145, 2008.
12. R. D. Mitra and G. M. Church, “In situ localized amplification and contact replication of

many individual DNA molecules," Nucleic Acids Research, vol. 27, no. 24, article e34, 1999.

13. S. Myllykangas, J. Buenrostro, and H. P. Ji, "Overview of sequencing technology platforms," Bioinformatics For High Throughput Sequencing, pp. 11–25, 2012.

14. J. B. Kim, G. J. Porreca, L. Song et al., "Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy," Science, vol. 316, no. 5830, pp. 1481–1484, 2007.

15. B. Lin, J. Wang, and Y. Cheng, "Recent patents and advances in the next-generation sequencing technologies," Recent Patents on Biomedical Engineering, vol. 1, pp. 60–67, 2008.

16. S. N. Lapidus, P. R. Buzby, and T. Harris, US20077169560, 2007.

17. P. R. Buzby, US20077220549, 2007.

18. H. Sun, US20070202521A1, 2007.

19. M. Rhee and M. A. Burns, "Nanopore sequencing technology: research trends and applications," Trends in Biotechnology, vol. 24, no. 12, pp. 580–586, 2006.

20. A. Meller and D. Branton, "Single molecule measurements of DNA transport through a nanopore," Electrophoresis, vol. 23, pp. 2583–2591, 2002.

21. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton, "Rapid nanopore discrimination between single polynucleotide molecules," Proceedings of the National Academy of Sciences of the United States of America, vol. 97, no. 3, pp. 1079–1084, 2000.

22. M. Rhee and M. A. Burns, "Nanopore sequencing technology: nanopore preparations," Trends in Biotechnology, vol. 25, no. 4, pp. 174–181, 2007.

23. J. Lagerqvist, M. Zwolak, and M. di Ventra, "Fast DNA sequencing via transverse electronic transport," Nano Letters, vol. 6, no. 4, pp. 779–782, 2006.

24. Q. Zhao, G. Sigalov, V. Dimitrov et al., "Detecting SNPs using a synthetic nanopore," Nano Letters, vol. 7, no. 6, pp. 1680–1685, 2007.

25. D. Radoje, L. Ivan, B. Ivan, and C. Radomir, "Sequencing of megabase plus DNA by hybridization: theory of the method," Genomics, vol. 4, pp. 114–128, 1989.

26. R. Drmanac, S. Drmanac, G. Chui et al., "Sequencing by hybridization (SBH): advantages, achievements, and opportunities," Advances in Biochemical Engineering/Biotechnology, vol. 77, pp. 75–101, 2002.

27. A. Schirinzi, S. Drmanac, B. Dallapiccola et al., “Combinatorial sequencing-by-hybridization: analysis of the NF1 gene,” Genetic Testing, vol. 10, no. 1, pp. 8–17, 2006.
28. S. Brenner, S. R. Williams, E. H. Vermaas et al., “In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs,” Proceedings of the National Academy of Sciences of the United States of America, vol. 97, no. 4, pp. 1665–1670, 2000.
29. V. Thakur and R. Varshney, “Challenges and strategies for next generation sequencing (NGS) data AR,” Journal of Computer Science and Systems Biology, vol. 3, pp. 40–42, 2010.
30. J. Voisey and C. P. Morris, “SNP technologies for drug discovery: a current review,” Current Drug Discovery Technologies, vol. 5, no. 3, pp. 230–235, 2008.
31. M. Samuel, B. T. Wilhelm, and J. Bähler, “Next-generation sequencing: applications beyond Genomes,” Biochemical Society Transactions, vol. 36, pp. 1091–1096, 2008.
32. U. Nagalakshmi, Z. Wang, K. Waern et al., “The transcriptional landscape of the yeast genome defined by RNA sequencing,” Science, vol. 320, no. 5881, pp. 1344–1349, 2008.
33. M. Sultan, M. H. Schulz, H. Richard et al., “A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome,” Science, vol. 321, no. 5891, pp. 956–960, 2008.
34. M. Kiriakidou, P. T. Nelson, A. Kouranov et al., “A combined computational-experimental approach predicts human microRNA targets,” Genes and Development, vol. 18, no. 10, pp. 1165–1178, 2004.
35. J. Brennecke, A. A. Aravin, A. Stark et al., “Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila,” Cell, vol. 128, no. 6, pp. 1089–1103, 2007.
36. S. Alexander, K. Pouya, P. Leopold et al., “Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes,” Genome Research, vol. 17, no. 12, pp. 1865–1879, 2007.
37. K. D. Kasschau, N. Fahlgren, E. J. Chapman et al., “Genome-wide profiling and analysis of Arabidopsis siRNAs,” PLoS Biology, vol. 5, no. 3, article 57, 2007.
38. G. Ramsingh, D. C. Koboldt, M. Trissal et al., “Complete characterization of the microRNAome in a patient with acute myeloid leukemia,” Blood, vol. 116, no. 24, pp. 5316–5326, 2010.
39. A. Maher, C. Kumar-Sinha, X. Cao et al., “Transcriptome sequencing to detect gene

fusions in cancer," Nature, vol. 458, no. 7234, pp. 97–101, 2009.

40. K. Chen and L. Pachter, "Bioinformatics for whole-genome shotgun sequencing of microbial communities," PLoS Computational Biology, vol. 1, no. 2, article e24, 2005
41. H. Huson, A. F. Auch, J. Qi, and S. C. Schuster, "MEGAN analysis of metagenomic data," Genome Research, vol. 17, no. 3, pp. 377–386, 2007.
42. M. J. Solomon, P. L. Larsen, and A. Varshavsky, "Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene," Cell, vol. 53, no. 6, pp. 937–947, 1988.
43. R. Mardis, "The impact of next-generation sequencing technology on genetics," Trends in Genetics, vol. 24, no. 3, pp. 133–141, 2008.
44. S. M. Johnson, F. J. Tan, H. L. McCullough, D. P. Riordan, and A. Z. Fire, "Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin," Genome Research, vol. 16, no. 12, pp. 1505–1516, 2006
45. T. S. Mikkelsen, M. Ku, D. B. Jaffe et al., "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells," Nature, vol. 448, no. 7153, pp. 553–560, 2007.
46. S. Johnson, A. Mortazavi, R. M. Myers, and B. Wold, "Genome-wide mapping of in vivo protein-DNA interactions," Science, vol. 316, no. 5830, pp. 1497–1502, 2007.
47. S. Levy, G. Sutton, P. C. Ng et al., "The diploid genome sequence of an individual human," PLoS Biology, vol. 5, no. 10, article e254, 2007.