



## NEXT GENERATION SEQUENCING (NGS): POTENTIAL AND APPLICATION IN DRUG DISCOVERY

NAVNEET KUMAR YADAV, POOJA SHUKLA, \*SHRUTI PAREEK \*\*A.K. SRIVASTAVA, \*\*\*F.W. BANSODE AND #R. K. SINGH

Divisions of Toxicology, \*\*Computer and \*\*\*Endocrinology, CSIR-Central Drug Research Institute, Lucknow-226001, \*Jaipur National University, Jaipur, India

# Corresponding Author: Mailing address: Division of Toxicology, CSIR – Central Drug Research Institute, Lucknow, India – 226001. Phone:919984941595. E – mail: [rktox@yahoo.com](mailto:rktox@yahoo.com)

### ABSTRACT

We have now entered into a new era of genomics due to the continued advancements in the next generation high throughput sequencing technologies, which includes sequencing by synthesis including fluorescent *in situ* sequencing (FISSEQ) and pyrosequencing, sequencing by ligation including using polony amplification and supported oligonucleotide detection (SOLiD), sequencing by hybridization along with sequencing by nanopore technology. We will see great impacts of these next generation sequencing methods for solving the genome related problem of plant and animal kingdom that will usher a new era of genomics and will get past the Sanger- sequencing that ruled for 30 years. NGS is expected to advance the drug discovery process.

**KEY WORDS:** Sequencing by synthesis, sequencing by ligation, sequencing by hybridization, single molecule DNA sequencing, polony sequencing, pyrosequencing, supported oligonucleotide detection (SOLiD).

### INTRODUCTION

In 1977 Fred Sanger and Alan R. Coulson published two methodological papers on the rapid determination of DNA Sequences<sup>1,2</sup> that helped to transform biology and provided a new tool for deciphering complete genes and later entire genomes. The method dramatically improved earlier DNA sequencing techniques developed by Maxam and Gilbert<sup>3</sup> published in the same year, and Sanger and Coulson's own 'plus and minus' method published 2 years earlier<sup>4</sup>. Due to having advantage of handling less toxic chemicals and radioisotopes made 'Sanger sequencing' the only DNA sequencing method used for the next 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 microfabrication have resulted in further throughput improvements of Sanger sequencing by multiplexing and miniaturization. These advances have been reviewed by Metzker<sup>5</sup>.

Despite many improvements, the gel based Sanger sequencing technology has limitation in cost and throughput. For achieving high

throughput many commercial companies and scientific lab have come up with different ways of high throughput sequencing with reasonable cost.

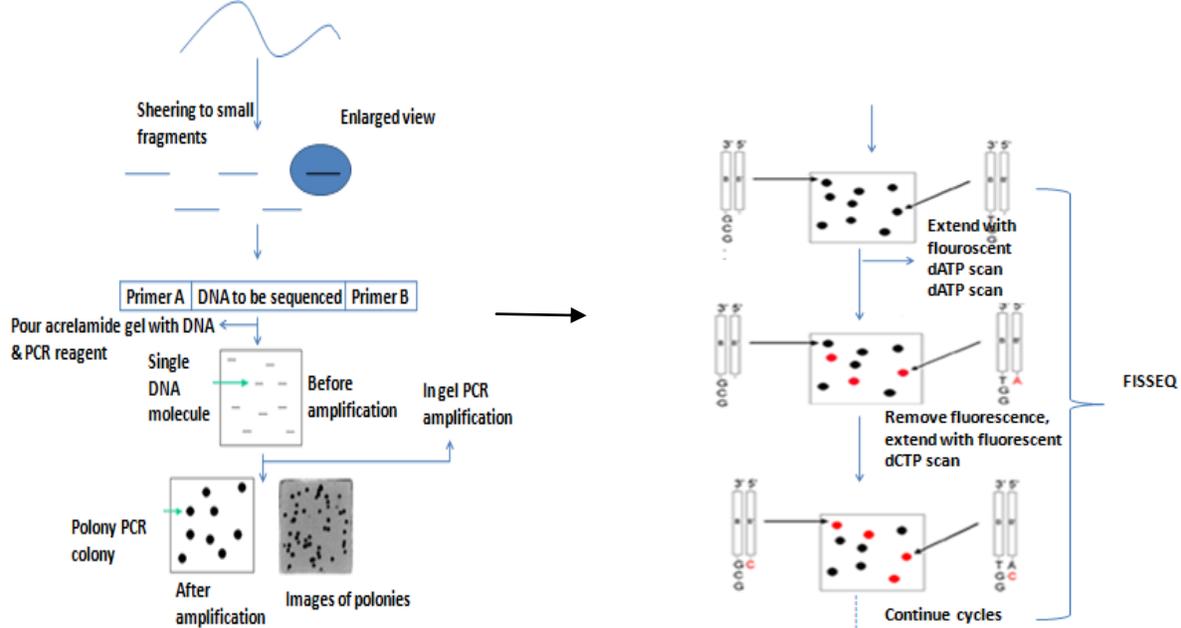
The technologies named together as next generation sequencing technologies include: sequencing by synthesis developed by 454 Life-Science<sup>6</sup> sequencing by ligation; sequencing by hybridization; single molecule DNA sequencing; nanopore sequencing and multiplex polony sequencing of George Church's lab<sup>7</sup>. Next generation sequencing have 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 major measurements for the success of the next generation technology are: sequence (read length), sequence quality, throughput and cost.

**Sequencing by synthesis (SBS):** Sequencing by synthesis 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

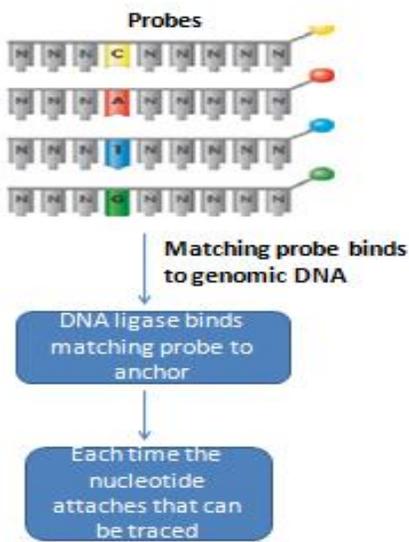
technology is another SBS technology University<sup>8,9</sup>. In brief, it is based on the detection of pyrophosphate (ppi) released during DNA synthesis when inorganic ppi is released after nucleotide incorporation by DNA polymerase. The released ppi is then converted to ATP by ATP sulfurylase. A luciferase reporter enzyme uses the ATP to generate light, which is then detected by a charged couple device (CCD) camera. Pyrosequencing has evolved

developed by Ronaghi *et al.* at Stanford

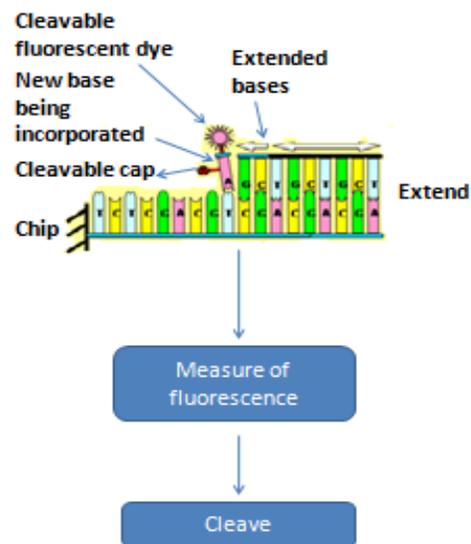
into an ultra-high throughput sequencing technology with the combination of several technologies such as template carrying micro beads deposited in micro fabricated Pico liter-sized reaction wells connecting to optical fibers<sup>10</sup>. 454 Life Sciences/ Roche Diagnostic has the Genome sequencer 20



(a)



(b)



(c)

Figure:1-Different techniques of Next Generation Sequencing (a) Polony sequencing (b) Sequencing by Ligation (c) Sequencing by Synthesis

System and the Genome Sequencer FLX system, two high-throughput commercial sequencing platforms, 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 emPCR (Emulsion PCR). Finally beads with amplified DNAs are loaded onto a pico-Titre plate for sequencing. The procedure is given in the Fig 1.

**Polony Sequencing:** George Church's Laboratory at Harvard University has developed a high throughput technology based on polony amplification and FISSEQ<sup>11</sup>. Polony amplification is a method to amplify DNAs in situ on a thin polyacrylamide film<sup>12</sup>. The DNA movement is limited in the polyacrylamide gel, so the amplified DNAs are localized in the gel and form the so called "polonies", polymerase colonies. Up to 5 million polonies (i.e. 5 million PCRs) can form 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<sup>13</sup>. Lim *et al.* developed a technology named PMAGE ("polony multiplex analysis of gene expression), which combines polony amplification and a sequence-by-ligation method, to sequence 14-base tags<sup>14</sup>. Up to 5 million polonies can be sequenced in parallel.

**Single molecule DNA sequencing:** Most current sequencing technologies are based on sequencing many identical copies of DNA molecules (often amplified). There are potential problems associated with sequencing amplified multiple copies of identical DNAs such as: How to achieve synchronous priming of each copy of the multiple DNAs by the sequencing primers?<sup>15</sup>

One way to solve these problems is to sequence by synthesis method<sup>16</sup>. DNA can be attached to solid support to form single molecule arrays, and the single DNA molecule be 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<sup>17</sup>. Applera Corp.<sup>18</sup>, fluorescent intercalators are employed as a donor in fluorescence resonance energy transfer

polyacrylamide gel and sequenced using sequencing by ligation method. Applied

(FRET) for use in single molecule sequencing reactions.

**Nanopore Sequencing:** Nanopores are nanometer-scale pores and are regarded as one of the most promising technologies in achieving true real time, ultra fast, true single molecule DNA sequencing<sup>19</sup>. 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<sup>20,21</sup>. The fabrication of nanopores such as the alpha-hemolysin pore and synthetic nanopore was reviewed recently<sup>22</sup>. Recent progress has made a big step toward ultra fast sequencing using nanopore technologies. Lagerqvist *et al.*<sup>23</sup> proposed a novel idea to measure the electric current perpendicular to the DNA backbone. Zhao *et al.* reported that a single nucleotide polymorphism can be detected by a change in the threshold voltage of a nanopore<sup>24</sup>.

**Sequencing by hybridization:** sequencing by hybridization (SBH) was given in the late 80's<sup>25</sup>. 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<sup>26</sup>. In spite of having some problems, using SBH with pre-defined 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<sup>27</sup>.

**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<sup>28</sup>. Church's laboratory at Harvard Medical School advanced the sequencing by ligation method for ultra fast and high throughput sequencing<sup>13,14</sup>. They converted an epifluorescence for rapid DNA sequencing. DNA molecules were amplified in parallel onto micro beads by emulsion polymerase chain reaction. Millions of beads were immobilized in a Biosystem Inc. acquired Agencourt personal Genomics who developed the SOLiD

(Supported Oligo Ligation Detection) system for high throughput DNA sequencing.

**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 help in analyzing NSG data. In table no.1 some bioinformatics tool is given.

Table No.1 ,Important bioinformatics tools for analysis of NGS data

<p><b>Alignment, assembly and visualization tools:</b> <b>Velvet</b> (<a href="http://www.ebi.ac.uk/~zerbino/velvet/">http://www.ebi.ac.uk/~zerbino/velvet/</a>): tool for de novo assembly of short and paired reads (Zerbino, D.R. &amp;Velvet.2008). <b>EULER</b> (<a href="http://euler-assembler.ucsd.edu/portal/">http://euler-assembler.ucsd.edu/portal/</a>): tool to generate short-read assembly and facilitate assembly of combined reads of NGS and Sanger sequencing (Chaisson,M.J. &amp; Pevzner, P.A.2008). <b>GMAP</b> (<a href="http://www.gene.com/share/gmap/">http://www.gene.com/share/gmap/</a>): program to map and align cDNA sequences to genome sequence using minimal time and memory, facilitates batch processing (Wu, T.D. &amp; Watanabe, C.K.2005).</p>
<p><b>Sequence variant discovery tools:</b> <b>SNPsniffer</b> (<a href="http://bioinformatics.bc.edu/marthlab/Polymorphism_Discovery_in_Next-Generation_Sequence_Data">http://bioinformatics.bc.edu/marthlab/Polymorphism_Discovery_in_Next-Generation_Sequence_Data</a>): tool for SNP discovery specifically designed for Roche/454 sequences <b>SeqMap</b> (<a href="http://biogibbs.stanford.edu/~jiangh/SeqMap/">http://biogibbs.stanford.edu/~jiangh/SeqMap/</a>): tool to map short sequences to a reference genome and detect multiple substitutions and indels (Jiang, H. &amp; Wong, W.H.2008.).</p>
<p><b>Integrated tools:</b> <b>NextGENE™</b> (<a href="http://www.softgenetics.com/NextGENE.html">http://www.softgenetics.com/NextGENE.html</a>): software to analyze NGS data for de novo assembly, SNP and indel detection and transcriptome analysis. <b>SeqMan genome analyzer</b> (<a href="http://www.dnastar.com/products/SMGA.php">http://www.dnastar.com/products/SMGA.php</a>): software with capacity to align NGS and Sanger data and detect SNPs; also facilitates visualization. <b>CLCbio Genomics Workbench</b> (<a href="http://www.clcbio.com">http://www.clcbio.com</a>): tool for de novo and reference assembly of Sanger and NGS sequence data, SNP detection and browsing.</p>

Source: <http://wiki.seqanswers.com>.

**Application of NGS techniques:** NGS technologies have already been used for various applications, ranging from whole genome sequencing and resequencing, single nucleotide polymorphism (SNP) and structural variation discovery, to mRNA and noncoding RNA profiling and protein- nucleic acid interaction assay NGS technologies are becoming a potential tool for gene expression analysis, especially for those species having reference genome sequences already available. An overview of NGS applications that are relevant to Drug Discovery are discussed below.

**SNP discovery:** Single nucleotide polymorphisms (SNPs) are important genomic resources which can be used in variety of analysis including physical characteristics like height and appearance as well as less obvious traits such as personality, behaviour and disease susceptibility. Sequence data generated for parental genotypes of the mapping population by using NGS technologies can be used for mining the SNPs at large scale<sup>29</sup>. SNPs can also significantly influence responses to pharmacotherapy and whether drugs will

produce adverse reactions. The development of new drugs can be made far cheaper and more rapid by selecting participants in drug trials based on their genetically determined response to drugs<sup>30</sup>.

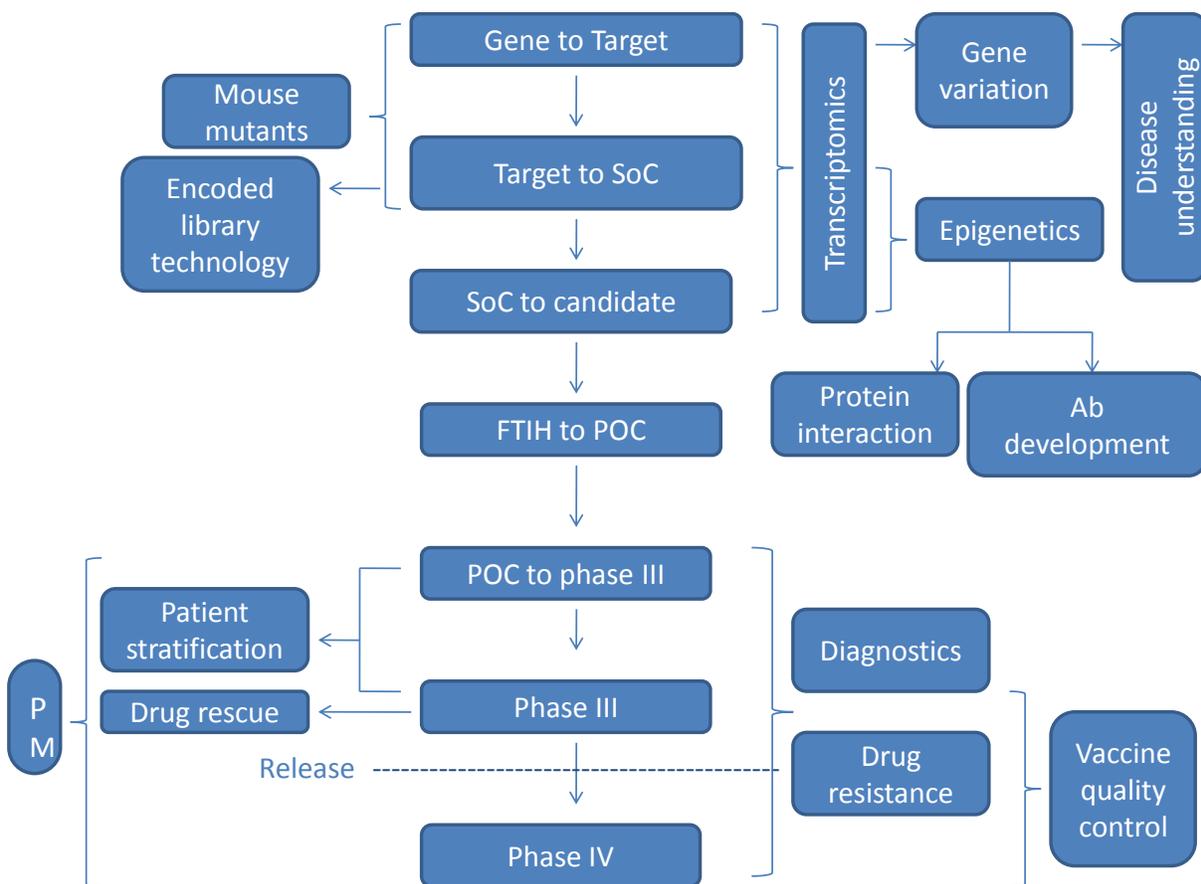
**Messenger RNA and non coding RNA profiling:** Apart from SNP discovery, expressed regions of a genome can be detected using NGS technologies. The next generation sequencing platforms are capable of identifying expression levels of nearly all genes, including those rare and species specific transcripts. A similar approach can be applied to large genomes . RNAseq data can be used to characterize exon-exon splicing events namely cases of alternative splicing<sup>31,32,33</sup>.

Non coding RNA, like microRNA (miRNA) is a broad class of regulatory RNA molecules. The NGS technologies are useful for discovery of non coding RNA for their short lengths<sup>34</sup>. *Chlamydomonas*, *drosophila*<sup>35,36</sup>. *Arabidopsis*<sup>37</sup>. A comprehensive study of miRNA in acute myeloid leukaemia was performed using NGS, with novel findings of differentially

expressed miRNAs<sup>38</sup>. Transcriptome sequencing has been found to be a powerful tool for detecting novel gene fusions in cancer cell lines and tissues<sup>39</sup>.

**Denovo Sequencing and Resequencing:** Metagenomics is defined as the application of genomics techniques to directly study communities of microbial organisms without isolation and cultivation of individual species<sup>40</sup>. It involves the characterization of the genomes in these communities, as well as their mRNA, protein and metabolic products. The next generation sequencing technologies have

enabled to move metagenomics from a single organism type in isolation to the studies of whole communities. NGS can avoid the cloning and culture steps, the main biases in genomics. NGS strategies are straight forward in that (1) deep sequencing of DNA fragments is conducted on an uncultured sample; (2) short reads are compared against database of known sequences using bioinformatics tool like MEGAN<sup>41</sup>. with or without assembly and (3) these data are then used to compute and explore their contents to infer relative abundances. Therefore NGS technologies can be a potent tool for discovery of microorganism and pathogen.



**Figure : 2 Application of Next Generation sequencing in Drug Discovery**  
**SoC-Standard of Care; POC-Proof Of Concept; FTIH-First Time In Human**

**Regulatory protein binding:** At low throughput , chromatin immuno precipitation (ChIP)<sup>41</sup> has enabled regulatory DNA-Protein binding interactions to be elucidated. Although it is a lengthy process like association of specific antibody to DNA-binding protein, followed by another protein- DNA cross linking agent, such

that any protein in close association with DNA becomes linked. The cells are lysed, the DNA is fragmented and the specific antibody is used to precipitate the protein of interest along with any associated DNA fragment. These DNA pieces are subsequently released by reversing the cross linking and identified by Southern blotting

or by qPCR<sup>42</sup>. Using a probe to infer the DNA binding site sequence of the protein being studied. So NGS can help in identifying the regulatory protein binding site for having shorter reads and specific binding site and it also provide higher resolution<sup>43</sup>.

**Exploring chromatin packaging:** Chromatin packaging denote how DNA is packed into histones. This packaging determines the transcription of a particular gene. So this understanding of DNA packing i.e chromatin packaging is of a great interest. An initial 454-based study of genomic DNA packaging into nucleosomes was described for the *C. elegans*<sup>44</sup> by sequencing the DNA isolated from nucleosome cores after micrococcal nuclease digestion and mapping them to the reference genome sequence. Mikkelsen *et al*<sup>45</sup> used the Illumina platform to demonstrate the connection between chromatin packaging and gene expression in several different cell types. Mikkelsen *et al.*<sup>45</sup> 2007 found that changes in chromatin state at specific promoters reflect changes in gene expression for the genes they control. A better understanding of the chromatin packaging will provide new strategies for the development of novel drugs.

NGS could have a central role in the discovery of new genomic biomarkers, Mutation detection, Personalised medicine and pharmacogenetics, Target identification and validation, Personalised medicine and pharmacogenetics, Clinical diagnostics, vaccine development, Investigating drug resistance and so on .

**Conclusion:** Due to their much improved cost effectiveness, compared to Sanger sequencing, and their many different uses, next-generation sequencing approaches are poised to emerge as the dominant genomics technology. Perhaps most significantly, these new sequencers have provided genome-scale sequencing capacity to individual laboratories in addition to larger genome centers. Compared to Sanger sequencing, advantages of the next-generation technologies mentioned thus far, including 454/Roche Illumina/Solexa, and ABI/SOLiD alleviate the need 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) (Morozova, O & Marra, A. M. 2008). We are at the verge of a new genomic revolution with recent advances in next generation sequencing

technologies. We will see enormous impacts of these next generation sequencing methods in dealing with complex biological problems, for example, the identification of all sequence changes in drug resistant HIV and drug resistance tuberculosis bacteria (TB), in the delineation of sequence changes for individual cells during cancer initiation and progression, and in global transcription factor mapping (CHIP to sequencing)<sup>46</sup>. In addition to, NGS technology will create a tremendous impact in medical science. It will make routine sequencing of a person's genome sequence of an individual human<sup>47</sup> and the initiation of the personal Human Genome Project.

## REFERENCES

1. Sanger, F., Air G. M., Barrell B. G., Brown N. L., Coulson A. R., Fiddes J. C., Hutchison C. A., Slocumbe P. M. and Smith M. (1977) Nucleotide sequence of bacteriophage  $\phi$ X174 DNA, *Nature*, 265, 687 – 695.
2. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
3. Maxam, A.M. and Gilbert, W. A. (1977) New method for sequencing DNA *Proc. Natl. Acad. Sci. USA* 74, 560–564.
4. Sanger, F. & Coulson, A.R. (1975). , [Nucleotide and Amino Acid Sequences of Gene G of  \$\phi\$ X174](#). *J. Mol. Biol.* 94, 441–448 .
5. Metzker, M. L., (2005) Emerging technologies in DNA sequencing, *Genome Res.* 15, 1767-1776.
6. Margulies, M., Egholm M., William E., Attiya A. S., Bader J. S., Bemben L. A., Berka J., et al., (2005) Genome sequencing in microfabricated high-density picolitre reactors , *Nature* 437, 376-380.
7. Yu P., Begley R. F. and Rothberg J. M., (2005) Genome sequencing in microfabricated high-density picolitre reactors, *Nature* 437, 376–380.
8. [Shendure](#) J. J., [Porreca](#) G. J., [Reppas](#) N. B, [Lin](#) X, [McCutcheon](#) J. P., [Rosenbaum](#) A. M., [Wang](#) M. D., et al, (2005) Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome , *Science* 309, 1728–1732.
9. Ronaghi M., Karamohamed S., Pettersson B., Uhle'n M., and Nyre'n P., (1996) Real-Time DNA Sequencing

- Using Detection of Pyrophosphate Release. *Anal Biochem*, 242, 84-89.
10. Ronaghi M, Uhlen M, Nyren PA., (1998) A sequencing method based on real-time pyrophosphate. *Science*; 281: 363, 365.
  11. Shendure J., Mitra R. D., Varma C. and Church G. C., (2004) Advanced sequencing technologies: methods and goals *Nature Reviews Genetics* 5, 335-344.
  12. Shendure J. and Ji H., (2008) Next-generation DNA sequencing, *Nature Biotechnology* 26, 1135 - 1145.
  13. Mitra RD, Church GM., (1999) In situ localized amplification and contact replication of mNucleic Acids Res; 27: 34-37.
  14. [Kim J. B.](#), [Porreca G. J.](#), [Song L.](#), [Greenway S. C.](#), [Gorham J. M.](#), [Church G. M.](#), [Christine E. Seidman](#) and [J. G. Seidman](#) (2007) Polony Multiplex Analysis of Gene Expression (PMAGE) in Mouse Hypertrophic Cardiomyopathy, *Science* 316, 1481-1484.
  15. Lin B., Wang J., and Cheng Y. (2008) Recent Patents and Advances in the Next-Generation Sequencing Technologies, *Recent patents on Biomedical Engineering*, 1,60-67.
  16. Lapidus, S.N., Buzby, P.R., Harris, T (2007) US20077169560
  17. Buzby, P.R. (2007) US20077220549
  18. Sun, H. (2007) US20070202521A1
  19. Rhee M. and Burns M. A. (2006) Nanopore sequencing technology: research trends and applications, *Trends Biotechnol.* 24, 580-586.
  20. Meller A, Branton D. (2002) Single molecule measurements of DNA transport through a nanopore, *Electrophoresis* 23, 2583-2591.
  21. Meller A., Nivon L., Brandin E., Golovchenko J., and Branton D. (2000) Rapid nanopore discrimination between single polynucleotide molecules, *Proc Natl Acad Sci USA* 97, 1079-1084.
  22. Rhee M., Burns M. A. (2007) Nanopore sequencing technology: nanopore preparations, *Trends in Biotechnology* 25(4), 174–181.
  23. Johan Lagerqvist, Michael Zwolak, and Massimiliano Di Ventra (2006) Fast DNA Sequencing via Transverse Electronic Transport, *Nano Lett* 6, 779-782.
  24. Q. Zhao, G. Sigalov, V. Dimitrov, B. Dorvel, U. Mirsaidov, S. Sligar, A. Aksimentiev, and G. Timp (2007) Detecting SNPs Using a Synthetic Nanopore, *Lett* 7, 1680-1685.
  25. Dramanac Radoje, Labat Ivan, Brukner Ivan, Crkvenjakov Radomir (1989) Sequencing of megabase plus DNA by hybridization: Theory of the method, *Genomics* 4, 114-128.
  26. Drmanac [Radoje](#), Drmanac [Snezana](#), Chui [Gloria](#), Diaz [Robert](#), Hou [Aaron](#), Jin [Hui](#), Jin [Paul](#), Kwon [Sunhee](#), Lacy [Scott](#) and Moeur [Bill](#), et al. (2002) Sequencing by Hybridization (SBH): Advantages, Achievements, and Opportunities, *Adv Biochem Eng Biotechnol* 77, 75-101
  27. Annalisa Schirinzi, Snezana Drmanac, Bruno Dallapiccola, Steve Huang, Kathryn Scott, Alessandro De Luca, Donald Swanson, Radoje Drmanac, Saul Surrey, and Paolo Fortina (2006) Combinatorial Sequencing-by-Hybridization: Analysis of the *NF1* Gene, *Genet Test* 10, 8-17.
  28. [Sydney Brenner](#), [Steven R. Williams](#), [Eric H. Vermaas](#), [Thorsten Storck](#), [Keith Moon](#), [Christie McCollum](#), [Jen-I Mao](#), et al. (2000) *In vitro* cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs, *Proc Natl Acad Sci USA* 97, 1665-1670.
  29. Thakur, V & Varsney (2010) Challenges and Strategies for Next Generation Sequencing (NGS) Data, *AR. J Comput Sci Syst Biol* 3, 40-42.
  30. [Voisey J](#), [Morris CP](#). (2008) SNP technologies for drug discovery: a current review, *Curr Drug Discov Technol.* 5(3), 230-5.
  31. Marguerat Samuel, Wilhelm Brian T. and Ju"rg Ba" hler (2008) Next-generation sequencing: applications beyond Genomes, *Biochem. Soc. Trans.* 36, 1091–1096.
  32. Nagalakshmi [Ugrappa](#), Wang [Zhong](#), Waern [Karl](#), Shou [Chong](#), Raha [Debasish](#), Gerstein [Mark](#) and Snyder [Michael](#) (2008) The Transcriptional Landscape of the Yeast Genome

- Defined by RNA Sequencing, *Science* 320, 1344–1349.
33. Sultan [Marc](#), Schulz [Marcel H.](#), Richard [Hugues](#), Magen [Alon](#), Klingenhoff [Andreas](#), Scherf [Matthias](#), Seifert [Martin](#), *et al* (2008) A Global View of Gene Activity and Alternative Splicing by Deep Sequencing of the Human Transcriptome, *Science* 321,956–960.
  34. Kiriakidou Marianthi, Nelson Peter T., Kouranov Andrei, Fitziev Petko, Bouyioukos Costas, Mourelatos Zissimos, and Hatzigeorgiou Artemis (2004) A combined computational-experimental approach predicts human microRNA targets, *Genes Dev* 18, 1165-1178.
  35. Julius Brennecke, Alexei A. Aravin, Alexander Stark, Monica Dus, Manolis Kellis, Ravi Sachidanandam, Gregory J. Hannon (2007) Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*, *cell* 128, 1089–1103.
  36. Stark Alexander, Kheradpour Pouya, Parts Leopold, Brennecke Julius, Hodges Emily, Hannon Gregory J. and Kellis Manolis (2007) Systematic discovery and characterization of fly, *Genome Res.* 17, 1865–1879.
  37. Kasschau Kristin D., Fahlgren Noah, Chapman Elisabeth J., Sullivan Christopher M., Cumbi Jason S. e, Givan Scott A., Carrington James C. (2007) Genome-Wide Profiling and Analysis of *Arabidopsis* siRNAs, *PLoS Biol.* 5(3), e57.
  38. Ramsingh G., Koboldt D.C., Trisal M, Chiappinelli K.B., Wylie T., Koul S., Chang Li-Wei (2010) Complete characterization of the microRNAome in a patient with acute myeloid leukemia, *Blood* 116, 5316–5326.
  39. Maher C.A., Sinha C.K., Cao Xuhong, Sundaram S.K., Han Bo, Jing Xiaojun, Sam Lee (2009) Transcriptome sequencing to detect gene fusions in cancer, *Nature* 458, 97–101.
  40. Nature 458, 97–101.
  41. Chen K, Pachter L (2005) Bioinformatics for whole-genome shotgun sequencing of microbial communities, *PLoS Comput Biol* 1, 106-112.
  42. Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data, *Genome Res.* 17, 377-386.
  43. Solomon MJ, Larsen PL and Varshavsky A. (1988) Mapping protein–DNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene, *Cell* 53, 937–947.
  44. Mardis, R. E (2007) The impact of next-generation sequencing technology on genetics, *Trends in Genetics* 24, 133-141.
  45. Johnson [Steven M.](#), Tan [Frederick J.](#), McCullough [Heather L.](#), Riordan [Daniel P.](#), and Fire [Andrew Z.](#) (2006) Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin, *Genome Res.* 16, 1505–1516.
  46. [Mikkelsen TS](#), [Ku M](#), [Jaffe DB](#), [Issac B](#), [Lieberman E](#), [Giannoukos G](#), [Alvarez P](#), *et al* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560.
  47. [Johnson DS](#), [Mortazavi A](#), [Myers RM](#), [Wold B](#), (2007) Genome-wide mapping of in vivo protein-DNA interactions, *Science* 316, 1497-1502
  48. Levy Samuel, Sutton Granger, Ng Pauline C., Feuk Lars, Halpern Aaron L., Walenz Brian P., Axelrod Nelson *et al* (2007) The Diploid Genome Sequence of an Individual Human, *PLoS Biol* 5, e254.