Middle Black Sea Journal of Health Science April 2018; 4(1):34-42

REVIEW

Genomic and Transcriptomic Sequencing and Analysis Approaches

Ümmügülsüm Tanman Ziplar[1], Demet Cansaran-Duman[1], Mine Turktaş[2]
[1]Ankara University, Biotechnology Institute, Ankara, Turkey
[2] Çankırı Karatekin University, Department of Biology, Çankırı, Turkey

Received: 12 February 2018 Accepted: 23 March 2018, Published online: 20 April 2018

Abstract
In this review, we explained that genomic and transcriptomic sequences and analysis assays. First of all, we detailed information of genomic and transcriptomic terms and related analyses. Genomics is aimed to elucidate the structural and functional properties of the genomes. Transcriptomics are used to express quantities of transcripts in a physiological state and specific developmental stage. The methods of genomic and transcriptomic analyses imply highly productive sequence analysis or microarray hybridization analysis as well as bioinformatics analyses. The sequencing technologies include set of methods such as preparing template, sequencing, imaging and data analysis. Firstly, the nucleotide information peculiar to DNA and RNA is obtained by means of the chosen technology in accordance with the goal and scope of the study. The obtained sequences are aligned with respect to a familiar reference sequence, or are combined as *de novo*. Subsequently, it is determined whether the distinct genomic sets are connected with the other genomic sets by overlapping distinct genomic sets, such as aligned sequence readings, gene annotation, EST, genetic polymorphism, and mobile elements. So, the structural variant to which the obtained sequence data are peculiar is determined. Within the scope of the study, giving information about sequencing technologies and the methods of analyses of the obtained sequences is aimed for researchers work on this subject.

Key words: Genomics, Transcriptomics, Bioinformatics, Next Generation Sequencing Technology

Address for correspondence/reprints:

Mine Türktaş

E-mail: mineturktas@karatekin.edu.tr

DOI: 10.19127/mbsjohs.370490

Introduction

Genomics and Transcriptomics

Genome is all DNA sequence information which are possessed by an organism. Each genome contains all the information necessary to create and maintain the organism. Genomic analysis is the identification, measurement, or comparison of genomic features such as DNA sequence, structural variation, expression at the gene expression or genomic position, and disclosure of functional element status. Transcriptomics is used to express quantities of transcripts in a physiological state and specific developmental stages. It is possible to sort out the main purposes of transcriptomics work as follows; cataloging of all transcriptomic species including mRNA, noncoding RNAs and small RNAs, identification of 5' and 3' ends, detection of the transcriptional structure of the gene, such as the presentation of spliced patterns, the discovery of the transcript level of each tissues in different processes. Transcriptomic sequencing is performed using the latest technologies of RNA-sequencing (RNA-Seq) technology. The RNA-seq technique is a very powerful and efficient new method that quantifies the state and organization of an RNA in a sample (Wang et al., 2009). Genomic and transcriptomic analysis methods generally require high-throughput sequencing or microarray hybridization and bioinformatic analysis.

What is bioinformatic?

Bioinformatics covers the fields of biological data collection and storage, data mining, database research, analysis and interpretation, modeling and product design, as well as an interdisciplinary field emerging in science and technology. Briefly, database creation is the process of creating and storing biological information. More specifically, we can define bioinformatics as a computational branch of molecular biology. Bioinformatics, an interface between modern biology and information technology, involves the discovery, development and implementation of computational algorithms and software tools that facilitate the understanding of biological processes. These software tools are intended primarily to serve the health and agriculture sectors. With these high-throughput data analysis methods, biological data can be analyzed and the organization of biological information is ensured. Most of the databases created in this area generate analyzes based on nucleic acids. Later, databases for the storage and organization of millions of nucleotide information obtained with these technologies are being created, and studies on the entry of new data into these databases are still continuing today (Luscombe et al., 2001).

Theoretical Bases

Generating the Nucleotide Information

Genetic information flow is transmitted by DNA molecules between generations except RNA viruses. The structure of DNA was elucidated by James Watson and Francis Crick in 1953. The DNA molecule is a polymeric nucleic acid macromolecule consisting of a five-carbon sugar (deoxyribose), a phosphate group and a nitrogen-rich purine (A: Adenine, G: Guanine) or pyrimidine (T: Thymine, C: Cytosine) bases (Watson and Crick, 1953). The characteristics of living organisms create with the genes which are found in DNA that are first formed by RNA and then protein transformation, which is called central dogma. In the RNA molecule, unlike the DNA molecule, there is a Uracil base instead of the Thymine, and the DNA molecule is double-stranded while the RNA molecule is the single-stranded (Shinkai et al., 2000).

The knowledge of nucleotides in DNA or RNA is obtained by sequencing. For this purpose, hierarchical approach and shotgun approach are used.

Hierarchical Approach

In this approach, firstly, DNA is cleaved into small fragments and then these fragments are cloned into BAC and YAC vectors. Fluorescence in situ hybridization (FISH) is usually used to determine which region of the genome is sequenced. For this purpose, colchicine is applied chemically on the cell in the cleavage stage and the cell is bursed, and then the cells captured at metaphase phase spread on a lamella. Fluorescently labeled probes are added to the clone that is cloned into the BAC and YAC vectors and incubated on the cell on this lamella for a certain period of time. If the part of the genome is sequenced, the fluorescent probe is expected to hybridize to that region, and then the fragments found in the BAC and YAC vectors are taken into smaller plasmid vectors. After sequencing operation, the small parts are read firstly then the contigs are combined and the contigs are combined to obtain the scaffolds. Then the scaffolds are overlapped on each other to determine the entire genome sequence information (Taylor and William, 1990).

Shotgun Approach

At this stage, DNA is first split into small fragments, then sequencing is performed for all the parts, and the resulting arrays are combined by overlapping each other. First, small contigs are obtained and scaffolds are then obtained by overlapping the contigs. And the integration of the scaffolds allows the complete genome (Myers et al., 1997).

Sequencing Technologies

First Generation Sequencing Technology

With this technology, also known as Sanger technology, target DNA is prepared with two approaches; the first is the transfer and replication of target DNA into plasmids, and the second is the amplification of the target DNA by the PCR reaction. The 'cycle sequencing' reaction is then performed. At this stage, denaturation of the template, primer binding, and elongation reactions are performed. The primers cover the target DNA starting from the immediate vicinity of the region of interest. At each step of primer extension, the chain is terminated by combining fluorescently labeled dideoxynucleotides (ddNTPs) and fragments generated by a nucleotide difference are amplified. The sequence gives the nucleotide identity of the ddNTP DNA sequence at the chain terminating point and thus the sequence is detected by high resolution electrophoretic separation of single-stranded and tagged dNTP extension products in a capillary-based polymer gel. The laser stimulation of fluorescent labels in nucleotides, as particles of apparent length, are extracted from the capillary and combined with the four-color detection of emission spectra to provide a reading represented as a 'trace'. While specific software for this area is converting these traces into DNA sequences, error probabilities are also generated for each basic search (Shendure and Ji, 2008).

Next Generation Sequencing Technology (NGS)

Template preparation, sequencing and imaging and data analysis are the main steps of a new generation of sequencing technology. The combination of these protocols distinguishes one technology from the other and identifies the type of data generated from each platform. The quality and accuracy estimations of the scores as a result of the sequencing obtained by the producers are provided. However, the quality of readings obtained from a platform is not equivalent to that obtained from another platform (Metzker, 2010).

Template Preparation

Today's technologies use randomly fragmented genomic DNA. Common to NGS technologies is the application of the template to a solid surface or support. Millions of sequencing reactions take place at the same spots where the templates to be arrayed are spatially fixed (Metzker, 2010).

Clonally replicated templates

Imaging systems are usually tuned to be determined by a single fluorescence. Therefore, the templates are multiplied. The two most common methods are emulsion PCR (emPCR) and solid phase amplification (Dressman et al., 2003; Fedurco et al., 2006).

EmPCR prevents the random loss of genomic sequences. In this study, sequencing patterns are prepared in a cell-free system. Create a library of trailers and connect adapters containing universal primer fields to the ends of the fragments. This structure allows genomes to be amplified by common PCR primers. The strands of DNA are separated from each other after ligation. Then one DNA molecule is attached to a bead. Once the EmPCR reaction has been successfully accomplished, the beads are immobilized on PicoTiterPlate (PTP) wells (Roche / 454), where NGS chemistry is performed (Shendure et al., 2005; Kim et al., 2007; Leamon et al., 2003).

Solid phase amplification is defined as the clusters cloned on a randomly distributed glass slide. The high-density forward and reverse primers are covalently attached to the plate and ratio of primers to template and support defines the surface density of reinforced clusters. Solid phase amplification to initiate the NGS reaction provides free ends where the universal sequencing primer can hybridize to initiate the reaction in 100-200 million spatially separated templates (Illumina / Solexa) (Metzker, 2010).

Single molecule template

While clonal propagation methods provide some advantages over bacterial cloning, implementing these protocols is very laborious and requires large amounts of genomic DNA material (3-20 µg). Preparation of single molecule templates is easier and requires less starting material (<1 µg). More importantly, these methods do not require a PCR run that generates mutations in clone-amplified templates, masked as PCR variants. At the same time, these methods do not require a PCR run to generate a mutation in clonal amplification templates. In addition, quantitative applications such as RNA-seq, which measure the number of mRNA molecules, provide more accurate results since the number of template is not multiplied (Wang et al., 2009).

With at least three different approaches, the single molecule templates are immobilized on the solid supports prior to the NGS reaction. In the first approach, the primer molecules are spatially distributed and covalently linked to solid support (Harris et al., 2008). Commonly, the bound adapters are hybridized with these primers previously fixed to the surface of template fragments prepared by random splitting of starting material to small sizes (e.g. ~ 200-250 bp). In the second approach, spatially distributed monomolecular patterns are covalently attached to solid support as a primer. A common primer template is then hybridized and sequenced. The DNA polymerase can be attached to the fixed template configuration in both approaches to initiate the NGS reaction. Helicos BioSciences uses both of these approaches. A third approach is that the spatially distributed single polymerase molecules are attached to the solid support and thus can be sequenced by clinging to the primer-bound template molecules. This approach is used by the Pacific Biosciences firm (Eid et al., 2009). With this technique, larger DNA molecules can be sequenced. Contrary to the first two approaches, this approach can be done by real-time sequencing (Metzker, 2010).

Sequencing and Imaging

There are fundamental differences in the sequencing of single-molecule templates and clonally amplification templates. Clonal amplification resulting in the population of the same templates can be sequenced at the same time. Observed signals are the consensus of nucleotides or probes attached to an identical template within a given cycle (Metzker, 2010).

Cyclic reversible termination (CRT)

Reversible terminators are used in the CRT method in steps involving addition of nucleotides, fluorescence imaging and cleavage. In the first step, the DNA polymerase inserts the nucleotide complementary to the template sequence only in exchange for a fluorophore. When the nucleotide is added, DNA synthesis is stopped by CRT. Once the corresponding nucleotide has been added, the nucleotides that do not participate in the reaction are washed. Then, imaging is performed to determine the identity of the inserted nucleotide. Subsequently, a cleavage step removes the termination / inhibiting group and fluorescein dye. After the addition of a nucleotide a fluorescence signal is released and a new nucleotide insertion is blocked by the CRT method. Illumina / Solexa uses a four-color CRT cycle and Helicos BioSciences performs a monochrome CRT cycle. The initial development of reversible blocking groups attached to the 3' end of the nucleotides has been based on the use of a dideoxynucleotide that acts as a chain terminator in the Sanger sequence. Blocking groups such as 3'-O-allyl-2'deoxyribonucleoside triphosphates (dNTPs) and 3'-O-azidomethyl-dNTPs in CRT have been used successfully (Metzker et al., 1994; Metzker, 2010).

Sequencing by ligation (SBL)

SBL is another cyclic method that uses single-base or two-base encoded probes that are used differently from CRT. In this method, a fluorescently labeled probe hybridizes to the complementary sequence of the primer template. The DNA-ligase then adds the dye-labeled probe primer. The unbound probes are then washed out and removed from the medium, and fluorescence imaging is performed to determine probe identity. In this sequencing cycle, splittable probes are used to remove fluorescent dye and a 5'-PO4 group is reconstituted for subsequent ligation cycles (Orita et al., 1989).

Single nucleotide addition

Roche 454, pyrosequencing

In this sequencing technology, the nucleotide variants are added individually to the sequencing medium. Pyrosequencing is an enzymatic reaction. When nucleotides are added, visible light is produced by the enzymes present in the medium. This method is a non-electrophoretic, bioluminescent method of measuring the release of inorganic pyrophosphate. Unlike other sequencing approaches that use modified nucleotides that terminate DNA synthesis in this method, the DNA polymerase is manipulated by monolithic addition of a dNTP in limiting amounts. The DNA polymerase extends and primes the primer by the addition of complementary dNTPs. DNA synthesis is resumed after the addition of the next complementary dNTP in the distribution cycle. And readings are recorded as light graphics (Ronaghi et al., 1996).

Ion Torrent

Nucleotides are also sent individually in this sorting technology. When a nucleotide is added, instead of using an enzymatic cascade, each dNTP detects released H + ions. The resulting pH change is detected by an ion sensitive field effect transistor (ISFET). The pH change detected by the sensor is proportional to the number of nucleotides detected (Goodwin et al., 2016).

Real time sequencing

It is a technology which influencing the commercial sector. Pacific Biosciences today uses this technology. Real-time nucleotides do not stop the DNA synthesis process, contrary to reversible terminators. During real-time sequencing, the dye-labeled nucleotides are continuously added to the nucleotide. On the Pacific Biosciences platform, zero-mode waveguide detectors (Zmw) are attached to the lower surface of DNA polymerases. Sequence information includes extended nucleotide primers incorporating phosphorylated nucleotides (Levene et al., 2003).

RNA-seq Technology

The recently developed RNA-seq technique uses a new generation of sequencing technologies. First, the RNAs are cleaved and translated into complementary DNA (cDNA) by adapters that attach to both ends of the fragments. In the next step those parts with or without amplification are sorted in a highly efficient manner. In the next step these amplified or unamplified segments are sequenced in a highly efficient manner. Reading lengths are typically in the 30-400 bp range, depending on the DNA sequencing technology used. After sequencing, the readings are aligned using both the reference genome and the transcriptome data (Wang et al., 2009). Then gene ontology analyzes of the readings are made. Classification and differences of genes expressed at least two folds are revealed on a tissue or cell basis.

Genome Alignment and Assembly

Once the NGS readings are complete, they are aligned to a known reference sequence or combined de novo. Identifying which strategy to use depends on factors such as the intended biological application, cost, labor and time. Sequence information of an organism detected by phylogenetic analysis closest to the material we use is used as the reference genome (Salzberg and Steven, 2009; Chaisson et al., 2009). Genomic sequencing studies of living organisms identified as model organisms such as Arabidopsis thaliana, C. elegans, Drosophila melanogaster and Saccharomyces cerevisiae have been completed. In this alignment approach, the presence of repeating regions in the reference genome, the absence of corresponding regions, the presence of gaps in the reference genome, or the presence of structural variables in the analyzed genome constitute a number of limitations (Frazer et al., 2009). Also, since each NGS platform produces a unique reproducible model for the variable sequence coverage, combining NGS reading types with alignment or aggregation may be incomplete (Aury et al., 2008; Reinhardt et al., 2009).

Sequence Analysis

In genomic research, the basic process is to investigate the association of different genomic clusters (eg, aligned sequence readings, gene annotations, ESTs, genetic polymorphisms, mobile elements, etc.) with other genomic clusters. With this method of comparison, the results of experiments can be characterized, causality and coincidence can be demonstrated and the biological effect of genetic findings can be evaluated (Quinlan and Hall, 2010). Implementing with large data clusters routinely produced with existing sequencing technologies is very complex to search for conflicts between existing web-based methods and features (Quinlan and Hall, 2010). For this reason, fast and flexible tools are needed to efficiently solve complex queries of these data. Genomic assays generally aim to compare features discovered in an experiment with known annotations for the same species. If the genomic features found in the different sequences match at least one base, these bases which are common are called overlapping or intersecting. For example, a typical question might be: 'Which of my new genetic variants coincide with exons?' and other structural variants such as deletions, insertions, duplications, translocations, transversions, multiallele copy numbers, transposon insertions, retrotransposon sites, satellit DNA and mitochondrial DNA variants are presented (Consortium, 2015).

Gen Ontology Analysis

The co-characterization of mutations and phenotypes of genes in genetic research has shown us that genes are common in many organisms. It is therefore clear that the functions of these genes can be understood in all organisms when we can understand the genes and proteins found in living organisms. In other words, when the role of a protein is understood in any organism, it means that it can be understood in other organisms. In other words, the role of a protein, when understood in any organism, means that it can be understood in other organisms. Gene ontology examines gene and protein roles and accumulations in the cell under three headings; these include biological processes, molecular functions and cellular components (Clarke, 2012). The biological purpose to which the gene contributes refers to the biological process. Any process can be performed through one or more regulated associations of molecular functions. Examples of these processes are cell growth, signal transduction. Cell growth, signal transduction is basically an example of these processes. More specifically, translation, pyrimidine metabolism, cAMP biosynthesis can be given as an example. The biochemical activity of a gene product (including ligands or specific binding to the construct) is defined as the molecular function. It only reveals what needs to be done without actually telling where and when it will happen. By way of example, the term 'enzyme', 'carrier' or 'ligand' may be given as an example. Examples of more narrow functional terms may be given as 'adenylate cyclase' or 'toll receptor ligand'. A cellular component refers to the location of a gene product that is active in the cell. These terms reflect our understanding of eukaryotic cell structure. Terms like "ribosome" or "proteasome" give the cellular component which indicates where the gene product is (Ashburner et al., 2000).

Conclusion

New generation sequencing technologies are used to classify genomes and discover genes with all genomic sequencing, transcriptomic sequencing, extraction of sequence-based profiles of epigenetic markers and chromatin structure, and metagenomic studies (Wang, 2009; Wold and Myers, 2008). Which technology to use for which approach depends on the characteristics of the platforms. Because of the large volumes of high quality bases are produced for each work. In addition, in the case of RNA-seq or direct RNA sequencing, the Helicos BioSciences platform is suitable for applications requiring quantitative information; because it is sequenced without having to convert RNA templates directly into cDNAs (Özsolak et al., 2009).

Compared with automatic Sanger sequencing, the new generation of sequencing technologies is cheaper, but still the cost of sequencing work is high. As a temporary solution to this problem, NGS platforms can be used to target only specific areas in the genome to reduce costs. With this workaround, genomic regions that cause disease or pharmacogenetic effects can be examined through all exons in the genome, specific gene families that constitute the drug targets, or genome-wide association studies (Altshuler et al., 2008; Wang and Weinshilboum, 2008). In addition, custom designed oligonucleotide microarrays are also used to determine the relevant gene regions (Singh-Gasson et al., 1999).

Establishing relationships with biological functions using individual sequences and the associated knowledge of these sequences is an important aspect of genetic data mining. For this, automatic functional annotation is performed. Functional description allows for the characterization of genes by understanding the physiological conception of an excess number of genes and the functional differences between subgroups of sequences. The gene ontology study provides a suitable framework for such analyzes. In this way, interpretations based on similarity can be made by comparing the ones known from the sequences with those unknown. For example, Blast2GO (B2G) is one of the programs used in this context (Conesa et al., 2005). In short, B2G uses the BLAST program to identify genes homologous to fasta formatted sequences and to perform sequence annotations with high efficiency (Conesa et al., 2005). At the same time, genome-wide assays have shown that approximate 40-60% of human genes have alternative splicing additions (Lee, 2002). At the same time, a lot of software has been developed nowadays from read-based alignment software like MAQ, BWA and SOAP to structural variable finding tools like BreakDancer, VarScan and MAQ.

In particular, the analysis has focused on illuminating both protein-coding and protein-encoding regions. However, different analyzes provide us with the understanding that different clusters of predicted genes exist when different additional criteria are used. This suggests that the number of genes that encode proteins are still not fully understood. There are also genes in the genome that belong to RNAs that do not encode proteins. Non-coding RNAs do not transcribe at high rates, but the genome is an important functional output. Non-coding RNAs generally play a greater role in the control processes such as genetic stresses and in the control of genetic networks, in addition to the roles in protein synthesis (ribosomal and transfer RNAs). Thus, transcriptomic readings cover all synthesized RNA, including protein-coding, protein-encoding, sense-antisense and RNA-regulated transcripts (Okazaki et al., 2002).

Today, with 1000 genome project, 2504 individuals from 26 different populations have been sequenced. As a result of this study, 88 million single nucleotide polymorphism (SNP) variants, 3.6 million insertions, 60,000 other structural variants have been identified and shared with the literature (Consortium, 2015). Genomic sequencing have been performed not only for humans but also for microorganisms, plants (rice, grapes, cucumber, corn, soybean, poplar etc.) and animals (chicken, pig, cow, sheep horse etc) (Michael and Jackson, 2013; Bai et al., 2012). Since 1998, 87 animal genomes and 55 plant genomes have been sequenced in non-human living species. Thus, functional elements and other structural variants of other living things will be able to be compared within themselves and have an idea of their evolutionary development (Song and Wang, 2013; Michael and Jackson, 2013). Today, the sequencing studies continue with the latest speed.

There are many ways in which genes measure expression levels comparatively in any developmental process or context. There are many ways in which genes measure expression levels comparatively in tissues that are in any developmental process or condition. For example, in a study conducted in Arabidopsis, it was observed that gene expression profiles were similar as a result of comparative transcriptomic analysis of developmental and dark aged leaf tissues (Buchanan-Wollaston et al., 2005). Today, however, transcriptomic studies are now being carried out by incorporating them into genomic studies together with developing technology. For example, in an effort to determine the alternative splicing status of mouse and human genes, all genomes were sequenced, resulting in similarity in mouse and human conserved alternative splice genes (Sugnet et al., 2003).

As a result, after the sequencing has been carried out, it is completed that genomic readings, read mapping, duplicate filtering, base quality value recalibration, INDEL realignment, Variant Site Discovery, genotype assignments and reporting of variants. Transcriptomic reading continues with alignment of the sequences and subsequent analysis of gene ontology after the assembly process. In this study, classification and differences of genes expressed at least two fold differences are revealed on tissue or cell basis. However, despite the rapid development of this area, there are still gaps in the processing and storage of data (Clarke, 2012).

Peer-review: Externally peer-reviewed.
Author Contributions: Concept UTZ, MT; Design UTZ, DCD, MT; Supervision MT; Literature review UTZ; Writing UTZ, MT; Critical review DCD, MT
Conflict of Interest: No conflict of interest was declared by the authors.
Financial Disclosure: The authors declared that this study hasn't received no financial support.

References

Altshuler D, Daly MJ, Lander ES. Genetic Mapping in Human Disease. Science 2008; 322: 881-88.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25: 25-9.

Aury JM, Cruaud C, Barbe V, Rogier O, Mangenot S, Samson G, et al. High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies. BMC Genomics 2008; 9: 603.

Bai Y, Sartor M, Cavalcoli J. Current status and future perspectives for sequencing livestock genomes. Journal of Animal Science and Biotechnology 2012; 3: 1-6.

Buchanan-Wollaston V, Page T, Harrison E, Breeze E, Lim PO, Nam HG, et al. Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. Plant J 2005; 42: 567-85.

Chaisson MJ, Brinza D, Pevzner PA. De novo fragment assembly with short mate-paired reads: Does the read length matter? Genome Res 2009; 19: 336-46.

Clarke L. The 1000 Genomes Project A History, Results and Tools. 2012.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005; 21: 3674-76.

Consortium. A global reference for human genetic variation. Nature 2015; 526: 68-74.

Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci 2003; 100: 8817-22.

Eid J, Adrian F, Gray J, Luong K, Lyle J, Otto G, et al. Real-Time DNA Sequencing from Single Polymerase Molecules. Science 2009; 23: 133-38.

Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res 2006; 34: e22.

Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nat Rev Genet 2009; 10: 241-51.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016; 17: 333-51.

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, et al. Single-molecule DNA sequencing of a viral genome. Science 2008; 320: 106-9.

Kim JB, Porreca GJ, Song L, Greenway SC, Gorham JM, George M, et al. Polony Multiplex Analysis of Gene Expression (PMAGE) in Mouse Hypertrophic Cardiomyopathy. Science 2007; 316: 1481-85.

Leamon JH, Lee WL, Tartaro KR, Lanza JR, Sarkis GJ, deWinter AD, et al. A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. Electrophoresis 2003; 24: 3769-77.

Lee BM, Christopher M. A genomic view of alternative splicing. Nature 2002; 30: 13-9.

Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science 2003; 299: 682-6.

Luscombe NM, Greenbaum D, Gerstein M. 2001. What is Bioinformatics? A Proposed Definition and Overview of the Field. Method Inform Med 2001; 40: 346-58.

Metzker ML. Sequencing technologies-the next generation. Nat Rev Genet 2010; 11: 31-46.

Metzker ML, Raghavachari R, Richards S, Jacutin SE, Civitello A, Burgess K, et al. Termination of DNA synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates. Nucleic Acids Research 1994; 22: 4259-67.

Michael TP, Jackson S. The First 50 Plant Genomes. The Plant Genome 2013; 6: 10.

Myers JL, Weber B, Eugene W. Human Whole-Genome Shotgun Sequencing. Genome Research 1997; 7: 401-09.

Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, et al. Analysis of the mouse transcriptome based on functional annotation of 60, 770 full-length cDNAs. Nature 2002; 420: 563-73.

Orita M, Suzuki Y, Sekiya T, Kenshi H. Rapid and sensitive detection of point mutations and dna polymorphisms using the polymerase chain reaction. Genomics 1989; 5: 874-79.

Özsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P. et al. Direct RNA sequencing. Nature 2009; 461: 814-8.

Reinhardt JA, Baltrus DA, Nishimura MT, Jeck WR, Jones CD, Dangl JL. De novo assembly using low-coverage short read sequence data from the rice pathogen Pseudomonas syringae pv. oryzae. Genome Res 2009; 19: 294-305.

Ronaghi M, Karamohamed S, Pettersson B, Uhle´n M, Nyre´n P. Real-time DNA sequencing using detection of pyrophosphate release. Analytical Biochemistry 1996; 242: 84-9.

Salzberg CT, Steven L. How to map billions of short reads onto genomes. Nature Biotechnology 2009; 27: 455-57.

Shendure J, Ji H. Next-generation DNA sequencing. Nature Biotechnology 2008; 26: 1135-45.

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 2005; 309: 1728-32.

Shinkai K, Sakurai S. Molecular recognition of adenine, cytosine, and uracil in a single-stranded RNA by a natural polysaccharide: Schizophyllan. J Am Chem Soc 2000; 122(18): 4520-4521.

Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR, Cerrina F. 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nature Biotechnology 1999; 17: 974-78.

Song L, Wang W. Genomes and evolutionary genomics of animals. Current Zoology 2013; 59: 87-98.

Sugnet CW, Kent WJ, Ares M, Haussler D. Transcriptome and genome conservation of alternative splicing events in humans and mice. Pac Symp Biocomput 2004;66-77.

Taylor R, William A. Hierarchical method to align large numbers of biological sequences. Elsevier 1990; 183: 456-74.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010; 26: 841-42.

Wang L, Weinshilboum RM. Pharmacogenomics: candidate gene identification, functional validation and mechanisms. Hum Mol Genet 2008; 17: 174-9.

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews and Genetics 2009; 10: 57-63.

Watson JD, Crick FHC. The Structure of DNA. Cold Spring Harbor Symposia on Quantitative Biology 1953; 18: 123-31.

Wold B, Myers RM. Sequence census methods for functional genomics. Nat Methods 2008; 5: 19-21.