Lecture 6. Genomic methods of research and diagnostics.

Learning outcomes:

1. Analyze Sanger and Maxam-Gilbert methods of DNA sequencing.
2. Compare Sanger method with several methods of Next Generation Sequencing (NGS), analyze their advantages and disadvantages.
3. Describe “short gun-sequencing” and “chromosome walking” methods of genome sequencing.
4. Characterize bioinformatical methods of genome analysis (genome assembling, genome annotation, finding of open-reading frames (ORFs) and prediction of genes, alignment of nucleotide sequences, protein structure prediction and etc.). Give specific examples of used programs and bioinformatical databases.
5. Explain the main traditional methods of molecular biology that are used in genomic research: polymerase chain reaction (PCR), gel-electrophoresis, northern blotting, Southern blotting, restriction analysis and etc.
6. Describe the methods of DNA genotyping, DNA diagnostics, DNA fingerprinting and DNA microarray.

Genomics is an interdisciplinary field of biology focusing on the structure, function, evolution, mapping, and editing of genomes. After an organism has been selected, **genome projects** involve three components: the **sequencing of DNA**, the **assembly of that sequence** to create a representation of the **original chromosome**, and the **annotation and analysis** of that representation.

DNA sequencing is the process of **determining the nucleic acid sequence** – the order of nucleotides in DNA.

Allan Maxam and **Walter Gilbert** published a DNA sequencing method in 1977 based on **chemical modification** of DNA and **subsequent cleavage at specific bases**. Also known as chemical sequencing, this method allowed purified samples of double-stranded DNA to be used without further cloning. This method's use of radioactive labeling and its technical complexity discouraged extensive use after refinements in the Sanger methods had been made.

Sanger sequencing is a method of DNA sequencing based on the **selective incorporation of chain-terminating dideoxynucleotides** by DNA polymerase during in vitro DNA replication. After first being developed by **Frederick Sanger** and colleagues in 1977, it became the most widely used sequencing method for approximately 40 years. It was first commercialized by Applied Biosystems in 1986.

Shotgun sequencing is a sequencing method designed for analysis of DNA sequences longer than 1000 base pairs, up to and including entire chromosomes. It is named by analogy with the rapidly expanding, quasi-random firing pattern of a shotgun. Since gel electrophoresis sequencing can only be used for fairly short sequences (100 to 1000 base pairs), longer DNA sequences must be broken into random small segments which are then sequenced to obtain **reads**. Multiple overlapping reads for the target DNA are obtained by performing several rounds of this fragmentation and sequencing. Computer programs then use the overlapping ends of different reads to assemble them into a continuous sequence.

Primer walking (or **Directed Sequencing**) is a sequencing method of choice for sequencing DNA fragments between 1.3 and 7 kilobases. Such fragments are too long to be sequenced in a single sequence read using the chain termination method. This method works by dividing the long sequence into several consecutive short ones. The DNA of interest may be a plasmid insert, a PCR product or a fragment representing a gap when sequencing a genome. The term "primer walking" is used where the main aim is to sequence the genome. The term **"chromosome walking"** is used instead when the sequence is known but there is no clone of a gene. For example, the gene for a disease may be located near a specific marker such as an RFLP on the sequence.

High-throughput sequencing, which includes **next-generation "short-read" and third-generation "long-read" sequencing methods**, applies to **exome sequencing, genome sequencing, genome resequencing, transcriptome profiling (RNA-Seq), DNA-protein interactions (ChIP-sequencing), and epigenome characterization**. The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that **parallelize** the sequencing process, producing **thousands or millions of sequences** concurrently. As of 2019, corporate leaders in the development of high-throughput sequencing products included **Illumina, Qiagen and ThermoFisher Scientific**.

In **bioinformatics**, **sequence assembly** refers to **aligning and merging fragments** from a longer DNA sequence in order to reconstruct the original sequence. This is needed as DNA sequencing technology cannot read whole genomes in one go, but rather reads small pieces of between 20 and 30,000 bases, depending on the technology used. Typically the short fragments, called reads, result from shotgun sequencing **genomic DNA**, or gene transcript (**ESTs**).

Genome annotation is the process of attaching biological information to sequences, and consists of **three main steps**:

1. identifying portions of the genome that do not code for proteins,
2. identifying elements on the genome, a process called gene prediction, and
3. attaching biological information to these elements.

Automatic annotation tools try to perform these steps ***in silico***, as opposed to manual annotation (a.k.a. curation) which involves human expertise and potential experimental verification. Ideally, these approaches co-exist and complement each other in the same annotation pipeline

The questions for self - control:

1. List the main genomic methods of scientific research and medical diagnostics.
1. Which classical and modern methods of DNA sequencing do you know? Describe each method.
3. List and explain the experimental methods of genetic analysis of completely sequenced genomes.
4. List and describe the computational (bioinformatical) methods of genome analysis.

Recommended readings:

1. "WHO definitions of genetics and genomics". World Health Organization.
2. Yu Liu. OMICS in Clinical Practice / 2014 by Apple Academic Press, Inc. – 456 p.
3. Barh D., Blum K., Madigan M.A. OMICS. Biomedical Perspectives and Applications / 2012 by Taylor & Francis Group, LLC. – 516 p.
4. "Introducing 'dark DNA' – the phenomenon that could change how we think about evolution".
5. Behjati S, Tarpey PS (December 2013). "What is next generation sequencing?". Archives of Disease in Childhood. Education and Practice Edition. 98 (6): 236–8. doi:10.1136/archdischild-2013-304340. PMC 3841808. PMID 23986538.
6. Maxam AM, Gilbert W (February 1977). "A new method for sequencing DNA". Proc. Natl. Acad. Sci. USA. 74 (2): 560–64. Bibcode:1977PNAS...74..560M. doi:10.1073/pnas.74.2.560. PMC 392330. PMID 265521
7. Sanger F; Coulson AR (May 1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase". J. Mol. Biol. 94 (3): 441–8. doi:10.1016/0022-2836(75)90213-2. PMID 1100841.
8. Sanger F; Nicklen S; Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors". Proc. Natl. Acad. Sci. U.S.A. 74 (12): 5463–7. Bibcode:1977PNAS...74.5463S. doi:10.1073/pnas.74.12.5463. PMC 431765. PMID 271968.
9. Adams, Jill U. (2008). "DNA Sequencing Technologies". Nature Education. Retrieved October 24, 2019.

10. Staden R (June 1979). "A strategy of DNA sequencing employing computer programs". Nucleic Acids Research. 6 (7): 2601–10. doi:10.1093/nar/6.7.2601. PMC 327874. PMID 461197.
11. Anderson S (July 1981). "Shotgun DNA sequencing using cloned DNase I-generated fragments". Nucleic Acids Research. 9 (13): 3015–27. doi:10.1093/nar/9.13.3015. PMC 327328. PMID 6269069.
12. Cann, Alan (1999). DNA Viruses: A Practical Approach. Oxford University Press. ISBN 978-0-19-963718-8.
13. Grada A (August 2013). "Next-generation sequencing: methodology and application". J Invest Dermatol. 133 (8): e11. doi:10.1038/jid.2013.248. PMID 23856935.
14. Hall N (May 2007). "Advanced sequencing technologies and their wider impact in microbiology". J. Exp. Biol. 210 (Pt 9): 1518–25. doi:10.1242/jeb.001370. PMID 17449817.open access
15. Church GM (January 2006). "Genomes for all". Sci. Am. 294 (1): 46–54. Bibcode:2006SciAm.294a..46C. doi:10.1038/scientificamerican0106-46. PMID 16468433.
16. Brent MR (January 2008). "Steady progress and recent breakthroughs in the accuracy of automated genome annotation" (PDF). Nature Reviews. Genetics. 9 (1): 62–73. doi:10.1038/nrg2220. PMID 18087260. S2CID 20412451.