Lecture 7. Transcriptomics. Methods of transcriptome investigation.

Learning outcomes:
1. Give the definition to the following terms: “transcript”, “transcriptome”, “transcriptomics”, “gene expression profile”.
2. Describe and analyze the different types of RNA by their structure and functions.
3. Explain the methods of different RNA extraction, amplification and sequencing.
4. Explain how the methods of gene expression profiling, RNA microarray and RNA-seq can be used for diagnostics of different diseases?

Transcriptomics technologies are the **techniques** used to study an organism's **transcriptome**, **the sum of all of its RNA transcripts**. The information content of an organism is recorded in the DNA of its genome and expressed through transcription. Here, **mRNA** serves as a transient **intermediary molecule** in the information network, whilst **non-coding RNAs** perform **additional diverse functions**. A **transcriptome** captures a snapshot in time of the total transcripts present in a cell. Transcriptomics technologies provide a broad account of which cellular processes are active and which are dormant. A major challenge in molecular biology lies in understanding how the same genome can give rise to different cell types and how gene expression is regulated. **Gene expression profiling** is the **measurement of the activity** (the expression) of thousands **of genes** at once, to create a global picture of cellular function. These profiles can, for example, distinguish between cells that are actively dividing, or show how the cells react to a particular treatment. Many experiments of this sort measure an entire genome simultaneously, that is, every gene present in a particular cell.
Several transcriptomics technologies can be used to generate the necessary data to analyse. **DNA microarrays** measure the relative activity of previously identified target genes. Sequence based techniques, like **RNA-Seq**, provide information on the sequences of genes in addition to their expression level.
A **non-coding RNA (ncRNA)** is an RNA molecule that is not translated into a protein. The DNA sequence from which a functional non-coding RNA is transcribed is often called **an RNA gene**. Abundant and functionally important types of non-coding RNAs include **transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), as well as small RNAs such as microRNAs, siRNAs, piRNAs, snoRNAs, snRNAs, exRNAs, scaRNAs and the long ncRNAs such as Xist and HOTAIR**.
The number of non-coding RNAs within the human genome is unknown; however, recent transcriptomic and bioinformatic studies suggest that there are thousands of them.] Many of the newly identified ncRNAs have not been validated for their function.[7] It is also likely that many ncRNAs are non-functional (sometimes referred to as junk RNA), and are the product of spurious transcription.
Non-coding RNAs are thought to contribute to diseases including **cancer and Alzheimer's**. Noncoding RNAs belong to several groups and are involved in many cellular processes. These range from ncRNAs of central importance that are conserved across all or most cellular life through to more transient ncRNAs specific to one or a few closely related species. The more conserved ncRNAs are thought to be molecular fossils or relics from the last universal common ancestor and the RNA world, and their current roles remain mostly in regulation of information flow from DNA to protein.
The first attempts to study whole transcriptomes began in the early 1990s. Subsequent technological advances since the late 1990s have repeatedly transformed the field, and made transcriptomics a widespread discipline in biological sciences. There are two key contemporary techniques in the field: **microarrays**, which quantify a set of predetermined sequences, and **RNA-Seq**, which uses high-throughput sequencing to record all transcripts. As the technology

improved, the volume of data produced by each transcriptome experiment increased. As a result, data analysis methods have steadily been adapted to more accurately and efficiently analyse increasingly large volumes of data. Transcriptome databases have grown and increased in utility as more transcriptomes are collected and shared by researchers. It would be almost impossible to interpret the information contained in a transcriptome without the context of previous experiments.

Measuring the expression of an organism's genes in different tissues or conditions, or at different times, gives information on how genes are regulated and reveal details of an organism's biology. It can also be used to infer the functions of previously unannotated genes. Transcriptome analysis has enabled the study of how gene expression changes in different organisms and has been instrumental in the understanding of human disease. An analysis of gene expression in its entirety allows detection of broad coordinated trends which cannot be discerned by more targeted assays.

The questions for self - control:

1. What are the “transcript”, “transcriptome”, “transcriptomics”, “gene expression profile”?
2. Which types of RNAs do you know and what are their functions?
3. Methods of different RNA extraction, amplification and sequencing.
4. Transcriptomic methods of molecular diagnostics.

Recommended readings:

1. "Transcriptomics technologies". PLOS Computational Biology. 13 (5): e1005457. 18 May 2017. doi:10.1371/JOURNAL.PCBI.1005457. ISSN 1553-734X. PMC 5436640. PMID 28545146. S2CID 3714586. Wikidata Q33703532.
2. Microarrays Factsheet". Retrieved 2007-12-28.
3. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, et al. (May 2005)."Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution". Science. 308 (5725): 1149–54. Bibcode:2005Sci...308.1149C. doi:10.1126/science.1108625. PMID 15790807. S2CID 13047538.
4. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, et al. (June 2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project". Nature. 447 (7146): 799–816. Bibcode:2007Natur.447..799B. doi:10.1038/nature05874. PMC 2212820. PMID 17571346.
5. Washietl S, Pedersen JS, Korbel JO, Stocsits C, Gruber AR, Hackermüller J, et al. (June 2007). "Structured RNAs in the ENCODE selected regions of the human genome". Genome Research. 17 (6): 852–64. doi:10.1101/gr.5650707. PMC 1891344. PMID 17568003.
6. Morris KV, ed. (2012). Non-coding RNAs and Epigenetic Regulation of Gene Expression: Drivers of Natural Selection. Caister Academic Press. ISBN 978-1-904455-94-3.
7. Shahrouki P, Larsson E (2012). "The non-coding oncogene: a case of missing DNA evidence?". Frontiers in Genetics. 3: 170. doi:10.3389/fgene.2012.00170. PMC 3439828. PMID 22988449.
8. van Bakel H, Nislow C, Blencowe BJ, Hughes TR (May 2010). Eddy SR (ed.). "Most "dark matter" transcripts are associated with known genes". PLOS Biology. 8 (5): e1000371. doi:10.1371/journal.pbio.1000371. PMC 2872640. PMID 20502517.
9. Hüttenhofer A, Schattner P, Polacek N (May 2005). "Non-coding RNAs: hope or hype?". Trends in Genetics. 21 (5): 289–97. doi:10.1016/j.tig.2005.03.007. PMID 15851066.
10. Brosius J (May 2005). "Waste not, want not--transcript excess in multicellular eukaryotes". Trends in Genetics. 21 (5): 287–8. doi:10.1016/j.tig.2005.02.014. PMID 15851065.

11. Palazzo AF, Lee ES (2015). "Non-coding RNA: what is functional and what is junk?". Frontiers in Genetics. 6: 2. doi:10.3389/fgene.2015.00002. PMC 4306305. PMID 25674102.
12. Jeffares DC, Poole AM, Penny D (January 1998). "Relics from the RNA world". Journal of Molecular Evolution. 46 (1): 18–36. Bibcode:1998JMolE..46...18J. doi:10.1007/PL00006280. PMID 9419222. S2CID 2029318.
13. Poole AM, Jeffares DC, Penny D (January 1998). "The path from the RNA world". Journal of Molecular Evolution. 46 (1): 1–17. Bibcode:1998JMolE..46....1P. doi:10.1007/PL00006275. PMID 9419221. S2CID 17968659.
14. Poole A, Jeffares D, Penny D (October 1999). "Early evolution: prokaryotes, the new kids on the block". BioEssays. 21 (10): 880–9. doi:10.1002/(SICI)1521-1878(199910)21:10<880::AID-BIES11>3.0.CO;2-P. PMID 10497339.