Gene technologies II

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Omics projects

High throughput methods in live sciences

1. Genomics
   Determination and analysis of DNA sequence of an organism

2. Transcriptomics and exomics
   Determination of mRNAs and their concentrations

3. Proteomics
   Determination of protein (including post-translational modifications) and their concentrations

4. Lipidomics, glycomics, ...
   Determination of lipid or carbohydrate composition or composition of other classes of molecules

5. Metabolomics
   Determination of low-molecular-weight compounds

Omics projects

1. Genomics

Determination and analysis of DNA sequence of an organism

First genome was sequenced in 1981 (human mitochondrial genome)

Genome of *Haemophilus influenzae* bacterium was sequence in 1995.

Yeast *S. cerevisiae* genome was sequenced in 1996.

Human genome was sequenced 1990 – 2003, but it was continuously improved till 2007.

Before introduction of next generation sequencing sequenced genomes included: human genome, genomes of important model organisms (*E. coli*, yeast, *Arabidopsis thaliana*, *Caenorhabditis elegans*), important pathogens and extremophiles.

Knowledge of human genomes make is possible to understand certain human diseases.

Knowledge of genomes of model organisms makes it possible to design experiment employing these species.

Knowledge of genomes of pathogens makes it possible to identify their weak points for antibacterial or antiviral therapy.

Knowledge of genomes of extremophiles makes it possible to identify, clone and produce novel enzymes and other proteins, for example thermostable ones.

Omics projects

1. Genomics

Determination and analysis of DNA sequence of an organism

Next-generation sequencing made genomics relatively cheap. The number of microbial genomes has grown to tens of thousands. It also introduced genomics of individuals or tissue-targeted and even single-cell genomics.

For human genome and genomes of common animals go to: http://www.ensembl.org

For other species go to:

http://metazoa.ensembl.org
http://plants.ensembl.org
http://fungi.ensembl.org
http://protists.ensembl.org
http://bacteria.ensembl.org

Omics projects

1. Genomics

    Determination and analysis of DNA sequence of an organism

Metagenomics

Conventional genomics isolates DNA from certain organism and sequences it. Metagenomcs was developed as an alternative. It isolates DNA from some environment and sequences it, no matter from which organism it comes from.

The advantage of metagenomics is that many organisms are hardly or not cultivable, nevertheless their DNA can be isolated.

Metagenomcs has been applied on various environments such as soil, ocean water, extreme environments, contaminated environments etc. This can be used to determine which organisms live in such environments and how microbial composition evolves and responds to external stimuli.

Metagenomics can also be used to find some new enzymes. It does not matter that we do not know the Latin name of the source.

Metagenomics also shapes the research in human microflora, especially gut microflora.

Omics projects

1. Genomics

Determination and analysis of DNA sequence of an organism

One DNA sequencing experiment can determine only a relatively short sequence (typically hundreds of base pairs in one “read”). In order to sequence longer DNAs it is necessary to use one of strategies described bellow.

Shotgun sequencing requires digestion of the DNA into overlapping fragments and their sequencing. The final sequence can be assembled from the fragments. Walking strategy tarts by sequencing a short read. The resulting sequence is used to design a primer. This primer serves for the next read and so forth until the whole sequence is determined. Sequencing by the former approach can lead to incompleteness. The later approach is low-throughput. Therefore, highly sophisticated strategies combining both approaches have been developed for large genomes.

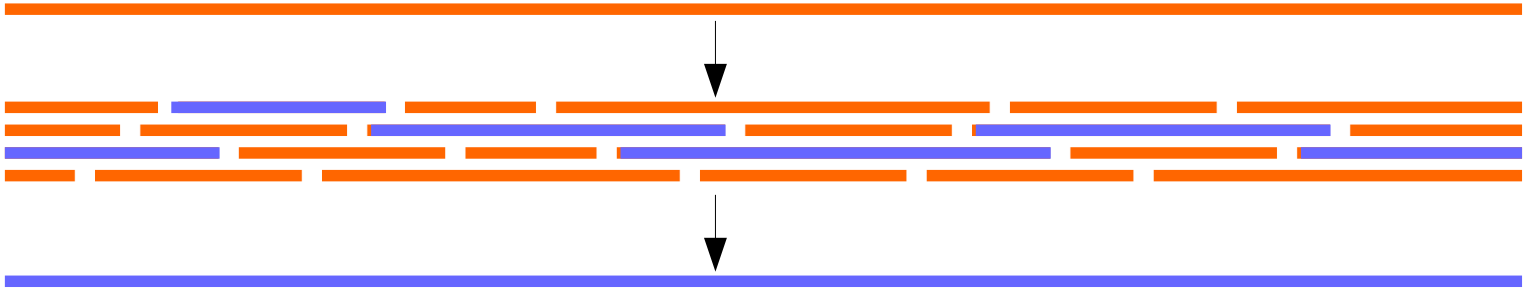
Since many genomes have been sequenced it is possible to map reads onto a reference genome instead assembling it *de novo*. For example, reads of a genome of a patient with a genetic disease can be mapped onto standard human genome.

Omics projects

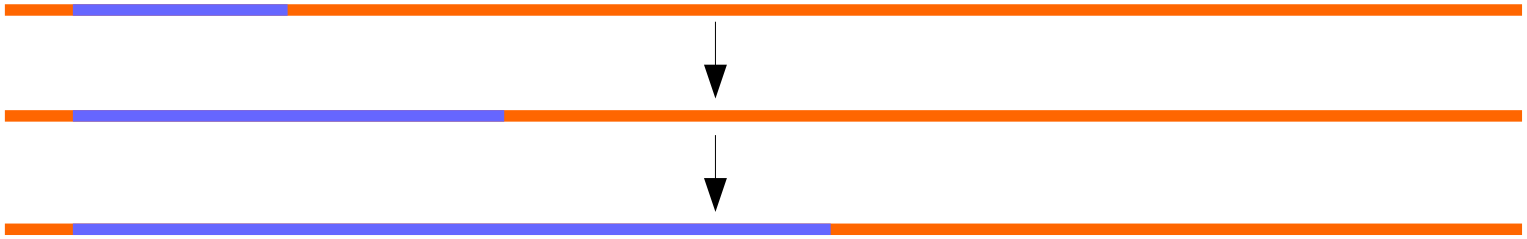
1. Genomics

Determination and analysis of DNA sequence of an organism

Shotgun



Walking



Omics projects

2. Transcriptomics and exomics

Determination of mRNAs and their concentrations

Transcriptomics and exomics study RNA (mostly mRNA). Exomics is focused on sequence (e.g. splicing variants of genes) whereas transcriptomics is focused on quantity of mRNA of individual genes.

Exomics determines sequence of all mRNA of certain cell types. This is useful for identification protein sequences and splicing variants.

Transcriptomics determines concentrations of individual mRNAs. It is used to:
find the difference between tissues, unstressed and stressed cells, sick and healthy cells etc. This can be used to understand the molecular basis of the tissue differentiation, stress response and disease. It can be also used in diagnostics to identify disease markers.

Traditional method for determination of mRNA concentration is Northern blot. More modern alternative is a quantitative PCR with reverse transcription.

However, these techniques can be applied on few genes. As a high throughput it is possible to use DNA microarrays. Recently, next-generation sequencing is used to sequence all mRNA. Concentration of mRNA is determined from the number of sequence reads on given gene.

Omics projects

3. Proteomics

Determination of protein variants and their concentrations

Proteomics determines sequences, post-translational modifications and quantities of proteins.

Concentration of an individual protein is usually determined by Western blot (gel electrophoresis followed by detection using a specific antibody). Determination of all proteins in cell can be done by 2-dimensional electrophoresis, with isoelectric focusing in one and SDS-PAGE in the second dimension. More modern are approaches based on mass spectrometry such as MALDI or electrospray, usually together with specific proteolytic sample digestion and/or liquid chromatography. Modern techniques make it possible to identify thousands of proteins in one sample.

The motivation for proteomics is in the fact that the state of the cell is not fully determined by mRNA levels. Exomics approach neglects the fact that a single mRNA can be translated into different variants of a proteins, differing mostly in posttranslational modifications. Transcriptomics approach neglects the fact that concentration of a protein is not fully correlated with the concentration of mRNA. There is a decent correlation, but there are important exceptions. For example, concentration of hypoxia inducible factor (HIF), involved in low oxygen response, is mostly controlled by its degradation and not by the level of mRNA.

Omics projects

4. Lipidomics, glycomics, ...

    Determination of lipid or carbohydrate composition or composition of other classes of molecules.

Healthy and sick cells differ in composition of other classes of molecules such as lipids or carbohydrates.

Lipids can be studied by chromatographic methods with mass spectrometry.

Similar methods can be used to determine carbohydrates. Many proteins in eukaryotic cells are glycosylated. Glycosylation patterns may differ between healthy and sick cells.

Omics projects

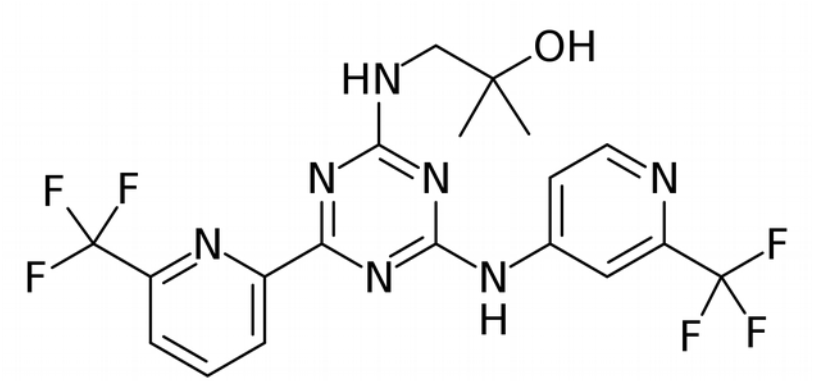
5. Metabolomics

Determination of low-molecular-weight compounds

Cells differ also in metabolite composition. This can be studied by chromatography, mass spectrometry, NMR and other techniques.

For example, it was found that some cancer cells contain high concentrations of D-2-hydroxyglutarate. Later it was found that the compound is produced as a side reaction by a mutant version of isocitrate dehydrogenase. This compound influences DNA and histone methylation and low-oxygen response, making cancer cells stronger.

This is a nice example how the omics project can identify a new target for therapy. Inhibitors of isocitrate dehydrogenase have been developed and subjected to clinical trials. Now the drug is marketed as Enasidenib.



Other fields of research

Bioinformatics
Systems biology
Epigenetics
Comparative genomics
Population genomics
Functional genomics
Interactomics
Synthetic biology

Other fields of research

Bioinformatics

Bioinformatics provides an informatics service for genomics, molecular biology and related fields. This discipline comprises:

- analysis of raw genomic data to determine complete genome (assembly, mapping)

- finding and identification of genes in the genome

- linking DNA, mRNA and protein sequences

- analysis of transcriptomics, proteomics and other omics data

- storing of sequences of genes and complete genomes in databases

- prediction of function of genes

- prediction of structural elements (e.g. domains) in protein structure

- prediction and analysis of 3D structures

- other analyses of sequences and 3D structures

Other fields of research

Systems biology

Systems biology is aimed at conversion of genomic information into functional models of cells and organisms.

For example, it is possible to identify all metabolic enzymes in the genome of an organism. This makes it possible to reconstruct metabolic pathways in the cell. Experiments providing time-resolved concentration changes in metabolites, enzyme kinetics etc. can be used to model time-course of metabolic processes. This makes it possible to simulate response to enzyme inhibition, enzyme mutation, excess of metabolite etc.

Moreover, it is possible to identify different signaling molecules (receptors, adaptor proteins, protein kinases and other signaling enzymes). Next, it is possible to reconstruct signaling pathways. Similarly to metabolism, signaling pathways can be simulated to predict the outcome of receptor activation or inhibition, mutation of a signaling proteins, inhibition of signaling enzymes etc.

Other fields of research

Epigenetics

It was found that there is heritable information that is not encoded in DNA sequence. These changes involve:

- covalent modification of DNA, especially DNA methylation

- covalent modification of histones, especially methylation or acetylation of Lys side chains

- presence of different RNAs

DNA methylation can be studied by bisulfite sequencing. Bisulfite (HSO_3^-) converts cytosine, but not methylated cytosine, to uracil. Parallel sequencing of treated and untreated sample provides information about DNA methylation.

Drugs altering methylation (epigenetic drugs) have been approved for cancer treatment.

Other fields of research

Comparative genomics

Similar organisms have similar genomes in terms of presence and sequences of genes. However, they can differ in genome architecture. For example, human and chimpanzee proteins typically differ in two amino acids, but some genes are moved on a chromosome and one human chromosomes is split into two in chimpanzee.

Comparison of genomes is therefore not straightforward and a whole new genomics discipline called comparative genomics has been developed.

Other fields of research

Population genomics

High throughput sequencing made it possible to sequence genomes at reasonable costs. First it was applied to genomes in populations of various species. The 1000 Genomes Project was launched in 2008 and in 2012 it provided 1,092 genomes of human individuals from various countries of the world.

Population genomics makes it possible to identify variability of a genome.

In medicine, the term mutation is used to describe a difference from the “standard” genome which causes a disease.

In contrast, the term polymorphism is used to describe a variability or a difference from the “standard” genome which causes differences between organisms, but not (directly) a disease.

Other fields of research

Functional genomics

Functional genomics is used to determine the function of genes. The most common tool is gene knockout. Knocking out of a gene leads to disruption of a metabolic or regulatory pathway or other response, which can be used to determine the function of the gene. Traditionally, homologous recombination is used to knockout genes. More recently, zinc-finger nucleases, Transcription activator-like effector nuclease (TALENs) and CRISPR have been developed.

The term functional genomics is also used to describe approaches aimed at integration of genomics, transcriptomics, interactomics etc.

Other fields of research

Interactomics

Many biological signaling processes involve interaction between proteins.

Affinity between two proteins can be determined by their co-isolation (pull-down assay). One protein is immobilized onto a suitable matrix. This is mixed with a cell extract. Proteins with the affinity can be isolated with the matrix and identified by chromatography with mass spectrometry.

As an alternative it is possible to use a yeast two-hybrid system. One gene is inserted into one vector, the second into second vector. If these proteins interact it starts a signal (e.g. enzymatic activity). Transcriptomic yeast two-hybrid studies of all pairs of genes have been carried out for several organisms.

It was found that a typical protein interacts with 10 – 100 other proteins.

Other fields of research

Synthetic biology

Building of new organisms is behind the idea of synthetic biology. One of most promising applications is cloning of whole metabolic pathways to “teach” some microorganism to produce some metabolite. Cloning multiple genes is similar to cloning a single gene, but much more complicated due to numerous reasons. Recently many of these obstacles have been solved.

Media reported in 2015 (*Nature*, **521**, 281-283, 2015) that two groups of researchers independently cloned the first and the second half of morphine biosynthesis pathway from poppies into yeast *S. cerevisiae*. This has raised concerns that leaking of these yeast strains to public would enable home “brewing” of opiates.

Today, most antibiotics and many other drugs are discovered as secondary metabolites of microorganisms. Researchers isolate microorganisms from environment. Next, they are cultivated and compounds are isolated from the medium. These compounds are tested for biological activity. This excludes microorganisms that cannot be cultivated. As an alternative, some researchers test sequencing of genomes and identification of microorganisms producing interesting compounds, cloning of their metabolic pathways and synthesis of compounds into a suitable organism.