Analysis of Genomic and Proteomic Data

Eva Budinská
Genomic and proteomic data challenges in MEDICINE

Human Genome Project decoded the human genome and brought new challenges in medicine

If we know genes, we can understand the function of their products (proteins) and their importance in biological pathways

We will be able to correlate the function of gene products with diseases (Pathobiology)

DISEASE ↔ GENE (S)

We will find optimal therapy, prevention and diagnostics of diseases
Genomic and proteomic data challenges in BIOLOGY

Decoding the genome of different species

We can apply high-throughput techniques in order to compare genome between different species.

We can study the evolutionary connections and construct phylogenetic trees.

We can understand the behaviour of parasites or reveal the mechanisms of survival of bacteria in extreme conditions.

...
High-throughput methods in genomics and proteomics

✔ **Genome analysis** (from nucleotide sequences to completely annotated genome)

- Analysis of structure – DNA sequencing, RNA-seq
- Expression analysis – Microarrays, SAGE, MPSS, Expressed sequence tags (ESTs)
- Comparative genomics – aCGH arrays, phylogenesis, SNP polymorphisms, fingerprinting
- Epigenetics – CHIP-on-chip, Chip-seq, alternative splicing arrays

✔ **Proteome analysis** (from mass spectra – through complex structures of protein clusters – to analysis of protein function)

- Mass spectrometry
- Protein microarrays
- Modelling of macromolecular systems – deriving the properties from atomic interactions

✔ **Modelling of complex systems** – protein interactions, signaling cascades, pathways...
High-throughput methodologies in genomics and proteomics

Modern „high-throughput“ technologies produce huge matrices of complex data

- **Microarrays**
  - Expression 10,000 – 100,000 genes, 100 – 1,000 samples

- **DNA sequencing**
  - Bilions of nucleotides

- **MASS spectrometry**
  - Thousands of protein spectra – GB datasets
Sources of genomic and proteomic data

- University hospital Bohunice Brno
- St. Anne's University Hospital Brno
- Masaryk Memorial Cancer Institute Brno

- Masaryk University
  - National Centre for Biomolecular Research
  - Department of Chemistry
  - Department of Biochemistry
  - Department of Biology
  - INBIT – biotechnological incubator
Analysis of genomic data – WITHOUT bioinformatics

"In principle, the string of genetic bits holds long-sought secrets of human development, physiology and medicine. In practice, our ability to transform such information into understanding remains woefully inadequate".

Analysis of genomic and proteomic data – WITH bioinformatics

1. Experiment
2. Extraction and processing of data
3. Data analysis
4. Biological and clinical interpretation
Sources of Data Analysts in Genomic and Proteomic Data

All over the world:

- Statistics & Applied Mathematics
- Physics
- Informatics
- Technical sciences
- Biologists

Czech Republic

- Mathematical biology
  - Institute of Biostatistics and Analyses, Masaryk University (15 students /year)

- Biomedical engineering
  - Brno
  - Prague
  - Kladno
  - Liberec
  - Olomouc
Rozvoj studijního oboru „Matematická biologie“ PřF MU Brno je finančně podporován prostředky projektu ESF č. CZ.1.07/2.2.00/07.0318 „Víceoborová inovace studia matematická biologie“ a státním rozpočtem České republiky

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ