Proteomics Course

LECTURE-3
Genomics and Transcriptomics: Why proteomics?

Lecture outline

- Genomics
- Transcriptomics
- Why proteomics?
Genomics

- Genome: The entire sequence of an organism's hereditary information, including both coding and non-coding regions, encoded in DNA is known as "genome".

- Studying genome of an organism by employing sequencing and genome mapping is known as "genomics".
Genome Sequencing: Traditional methods

DNA sequencing – Sanger’s method
**Shotgun Sequencing**

![Diagram of shotgun sequencing process]

**Traditional DNA Sequencing Methods**

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain termination method (Sanger)</td>
<td>Gold standard but time consuming</td>
<td>High</td>
</tr>
<tr>
<td>Pyrosequencing</td>
<td>Based on chemiluminescent detection</td>
<td>Very high</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Identifies variant alleles SNPs</td>
<td>Very high</td>
</tr>
</tbody>
</table>
Genome Sequencing Projects

- Bacterial genome sequencing
- *S. cerevisiae* sequencing
- *D. melanogaster* sequencing
- *A. thaliana* sequencing
- Mouse genome sequencing
- Human genome sequencing
Total genes 25,000

Average gene 3,000 bp

Genome 3.2 bp

Sequence is same in 99.9% people

2% of genome encodes instructions for protein synthesis

Insights from HGP

Potential Benefits of HGP

Molecular Medicine

Environment

Agriculture and Livestock

Microbial Genetics

Risk assessment of individuals to specific allergens

Bioarchaeology, Anthropology, Evolution, and Human Migration
Next Generation Sequencing (NGS)

- Next generation or second generation sequencing technology
  - sequencing by ligation or by synthesis, including pyrosequencing and reversible chain termination

- Third generation sequencing technology - to improve second-generation sequencing technology and lower the cost, use of scanning tunneling electron microscope, fluorescence resonance energy transfer, single-molecule detection and protein nanopores
Next Generation Sequencing: Nanopore sequencing

NGS Platforms (Commercial)

Illumina

Pyro-sequencing

Helicos

SOLiD

Ion Torrent

Ref: Am J Clin Pathol 2011;136:527-539
NGS vs. Sanger’s sequencing

- In NGS preparations are done *in vitro*
  - Transformation of *E. coli* and other limitations are avoided
- NGS methods based on arrays (not capillary)
  - Sequencing time is reduced
- Reduced cost

DNA Microarrays
DNA Microarrays

(II) Transcriptomics
Transcriptomics

• Transcriptome - analysis of all species of transcript, including mRNAs, non-coding RNAs and small RNAs

• Transcript analysis is used to quantify the expression level changes of each transcript during development and under different conditions

Techniques for evaluating gene expression

• Northern blotting
• Quantitative real-time polymerase chain reaction
• Differential display
• Serial analysis of gene expression
• Microarray
Reverse transcription PCR

- mRNA
- Reverse transcriptase
- cDNA
- Primer
- First cycle
- Taq polymerase
- Second cycle

Real-time PCR

- Primer
- Quencher
- Reporter probe
- Target ssDNA
- Annealing
- Taq polymerase
- Fluorescence quenched
- Polymerization & probe degradation
- Reporter fluorescence
- Relative fluorescence vs. cycles
cDNA Microarrays

1. DNA Array printing

2. Reverse transcription & labeling

3. cDNA hybridization

4. Array scanning

RNA Sequencing
RNA-Seq

- mRNA
- cDNA
- EST library with adaptors
- Short sequence reads
  - AGTTTGCCAATGACTAGGTACCAGGTAAA
  - GTTACCATGGATTCCATTGG
  - GTAGTCGAAGCTAGTACGTAGCTAA
  - GTGATACAGTCTAGTACAGTATCGATGATAA
- ORF
- Coding sequence
- Junction reads
- Exonic reads
- Poly A end reads
- Mapped sequence reads

Ref: Nat Review Genetics 2009;10:57-63

RNA-Seq (2)

Base resolution expression profile

RNA expression level

Nucleotide position
(III) Why proteomics?

**Proteomics**

- “Proteome” entire complement of proteins expressed by genome of an organism under defined conditions
- The study of entire compendium of proteins encoded by a genome is known as “proteomics”
Genomics Vs. Proteomics

• Availability of completed genome sequences of several species has shifted the focus towards identification and characterization of all gene products of organism.

• Genome represents only the starting point but products of gene expression, proteins, provide a much more meaningful insight into essential biological processes.
Single gene, multiple protein products indicates complexity of the proteome.
Genomics vs. Proteomics (3)

![Diagram showing post-translational modifications](image)

Gel-based proteomic techniques

![Diagram showing gel-based proteomic techniques](image)
**Mass Spectrometry**

Ionization → Mass Sorting (filtering) → Detection

- Ion Source: (charged molecules)
- Mass Analyzer: Sort ions by Mass (m/z)
- Data System: Mass Spectrum

**Data Processing**

- Relative Abundance

---

**Protein Microarrays**

1. Expression vector → Gene insert → Gene of interest → Protein expression → Protein purification
2. Functionalized array surface → Protein purification → SDS-PAGE → Chromatography

- Expression vector: Heterologous host (e.g. E. coli)
Label-free detection techniques: Surface Plasmon Resonance (SPR)

Summary

- Genomics
- Transcriptomics
- Why Proteomics?
- Need for Systems Level investigation
REFERENCES


REFERENCES