The course aims to give detailed insight into the techniques and technological trends in the fields of genomics and transcriptomics, building up the necessary foundations for further research in genetic association studies, population genetics, diagnostics, medicine and drug development. The course will be lecture and discussion based and will include a short introduction to conventional assays used in molecular biotechnology, descriptions of different methods for typing genetic variations, a variety of techniques for multiplex amplification, advanced techniques and technological platforms for DNA sequencing and whole genome sequencing, and different techniques for transcript and protein profiling.

In addition, the course involves a literature workshop of selected articles, which will be performed in groups. Each group presents one article and will oppose two other groups’ articles. This project aims to teach critical reading, interpretation and comparison of the most advanced techniques and platforms in the fields of whole genome sequencing, massive parallel genotyping, and transcript profiling. The project requires teamwork and planning, and participation as well as presence on the workshop days is compulsory.
The following topics will be covered:

• - introduction (repetition)

• - conventional methods and platforms for discovery and typing of single nucleotide polymorphisms (SNPs)

• - obstacles with multiplex PCR amplification

• - different techniques and platforms enabling multiplex PCR amplification

• - different techniques and platforms enabling whole genome SNP genotyping

• - conventional methods for DNA sequencing and its limitations

• - most recently developed techniques and platforms for whole genome sequencing and future trends in the field

• - different microarray-based techniques for parallel analysis of gene expression, including experimental design and difficulties with microarray-based approaches

• - alternative techniques and platforms for analysis of gene expression

• - techniques involved in proteomics
21,571 Known human genes
Technologies for genome, transcript and proteome analysis
The biomolecules of the cell

DNA
- Genetic variability (SNP)
- Copy number variations

RNA
- Transcript profiles (DNA-arrays)
- Splice variants
- Noncoding RNA

Protein
- Expression profiles (2-D gels)
- Structure
- Localization
- Enzymatic function
- Interactions
- Modifications

Cellular
- Models organisms
# Repetition basics - the molecular tool box

## Reagents
- **Enzymes**
  - restrictions enzymes
  - ligase
  - DNA-polymerase
  - reverse transcriptase
  - kinase
  - transferase
  - phosphatase

- **Oligonucleotidsyntes**
  - primers
  - probes
  - gene assembly

- **Vectorer**
  - plasmids
  - phage (M13)
  - cosmids
  - BACs/YACs
  - virus

- **Hostcells**
  - bacteria
  - yeast/fungi
  - plantcells
  - insectcells
  - mammalian cells

## Technologies
- **Blotting**
  - Northen
  - Southern
  - Western

- **Electrophores**
  - agarose
  - SDS-PAGE
  - RFLP

- **PCR**
  - *In vitro* mutagens

- **DNA sequencing**
  - Mass spectrometry
  - Pyrosequencing
  - Hybridization

- **SNP typing**

- **STR analysis**
  - Microsatellites

- **DNA chip**

- **RT- PCR/TaqMan**

- **In situ hybridization**
  - FISH

- **Transformation**
  - heat-shock
  - electroporation
  - microinjection
  - transfection

- **Selections methods**
  - antibiotic resistens
  - genetic complementation

- **Screening methods**
  - genotypic
  - phenotypic

- **Transgen technology**
  - transient
  - stable
  - knockout

## Reagents
- **Vectorer**
  - plasmids
  - phage (M13)
  - cosmids
  - BACs/YACs
  - virus

- **Hostcells**
  - bacteria
  - yeast/fungi
  - plantcells
  - insectcells
  - mammalian cells

- **Electrophores**
  - agarose
  - SDS-PAGE
  - RFLP

- **PCR**
  - *In vitro* mutagens

- **DNA sequencing**
  - Mass spectrometry
  - Pyrosequencing
  - Hybridization

- **SNP typing**

- **STR analysis**
  - Microsatellites

- **DNA chip**

- **RT- PCR/TaqMan**

- **In situ hybridization**
  - FISH

- **Transformation**
  - heat-shock
  - electroporation
  - microinjection
  - transfection

- **Selections methods**
  - antibiotic resistens
  - genetic complementation

- **Screening methods**
  - genotypic
  - phenotypic

- **Transgen technology**
  - transient
  - stable
  - knockout
Molecular tool box

Reagents - some examples...
Reverse transcriptase (RT)

- Reverse transcriptase (RT) synthesize a complementary DNA strand from a RNA template. The enzyme is cloned from retrovirus. A promiscuous enzyme.

- Applications areas:
  - cDNA cloning (cloning a gene)
  - quantification of RNA (RT-PCR)
  - partial DNA sequencing of cDNA libraries EST, expressed sequence tag
  - labelling of RNA for microarray analysis

![Diagram of mRNA and cDNA synthesis by reverse transcriptase](image)
Reverse transcriptase (RT)

- Reverse transcriptase (RT) synthesize a complementary DNA strand from a RNA template. The enzyme is cloned from retrovirus. A promiscuous enzyme.

- Applications areas:
  - cDNA cloning (cloning a gene)
  - quantification of RNA (RT-PCR)
  - partial DNA sequencing of cDNA libraries EST, expressed sequence tag
  - labelling of RNA for microarray analysis

\[
\text{mRNA} \quad 5'\text{Cap} \quad \text{AAAAA} \quad 3' \\
\text{Random hexamer} \\
\downarrow
\]

\[
\text{Reverse transcriptase} \\
+ \text{dNTPs} + \text{labelled nucleotides}
\]

\[
\text{mRNA} \quad 5'\text{Cap} \quad \text{AAAAA} \quad 3' \\
\text{cDNA}
\]

\[
\star \quad \star \quad \star \quad \star
\]
Cosmids

- Cosmids are hybrids between plasmids and phages and encode:
  - ori
  - gene for antibiotic resistance (e.g., Tet)
  - cos-sites
  - cloning sites

- Cosmids are used for cloning of large genome sequences. Phage lambda harbor 15 kbp DNA, cosmids can harbor 45 kbp.
DNA/cDNA library

Genomic DNA

Cleave with restrictions-enzymes

DNA-fragment

mRNA

RT-PCR

cDNA

Vector

RT = Reverse Transcription

DNA library or cDNA library
Different vectors are used to introduced recombinant DNA into different host cells. :

- **Bacteria**
  - *E. coli* is most common
  - High expression levels
  - Large scale production

- **Yeast**
- **Insect cells**
- **Mammalian cells**
- **Plant cells**

Eukaryotic cells

Post-translation modifications (ie glycosylations)
Molecular tool box

Technologies - some examples…
Gel electrophoresis

- Separation of DNA, RNA or proteins
- Charge/weight = constant
  DNA and RNA ⇒ neg. charged phosphate groups, proteins binds SDS ⇒ neg. charged
- Separation is achieved according to size

Small molecules travels faster through the gel as compared to larger molecules

Gelmatrix with pores, i.e. agarose or polyacrylamide
• **Etidiumbromid** intercalates between bases in DNA.
• Upon bindning to DNA ethidiumbromid becomes strongly fluorescent.
Visualisation of DNA on a gel

- DNA-fragments are depicted as an orange band when exposed on a UV light table

 ![Image of DNA visualization](image_url)
 Autoradiography

- DNA/RNA is labelled with radioactive nucleotides, ie $^{32}\text{P}$ or $^{35}\text{S}$.
- Radioactive nucleic acids are detected by xray film
Native protein + SDS → Denaturered polypeptid with many negative charges

Sodiumdodecylsulfat (SDS) binds strongly to proteins and cause denaturation of the protein
- approx one SDS per two aminoacids
- proteins are separated according to size
Western/Immunoblot

SDS-PAGE → Transfer → Nitrocellulose membrane → Probe membrane with antibody → Develop membrane

Note: The gel could also be stained with an unspecific stain such as Coomassie or AgNO₃. This would result in that all proteins in the gel will turn blue or "brown", respectively.
Southern blot

- Analysis of DNA

DNA-fragment

Gel electrophoresis → Agarose gel → Blotting → Nitrocellulosa-filter → $^{32}$P-labelled DNA probe → Autoradiogram
Northern blot

- Analysis of RNA

- mRNA

  Gel electrophoresis

  Agarose gel

  Blotting

  Nitrocellulose filter

  $^{32}$P-labelled DNA probe

  Autoradiography

  Autoradiogram
RFLP (Restriction Fragment Length Polymorphism)

- Cleave DNA-sample with restriction enzyme.
- Southern blot:

  ![Diagram of DNA cleavage and Southern blot process]

  - Compare pattern to judge similarities (family investigations)
  - Natural polymorphism in DNA results in different restriction patterns between non-related individuals.
Colony Blotting

Nitrocellulosa filter

Radioactive labelled specific DNA-probe

Xray film

Agarplate with DNA-library

Go back to plate to identify correct clone.
Co-immunoprecipitation

1. Lysis and immunoprecipitation
2. Detection of proteins
   - SDS-PAGE + autoradiography
3. Identification of proteins
   - mass spectrometry
   - Edman sequencing

Area of usage:
- check interaction between two known proteins
- identify unknown protein-protein interactions
DNA sequencing

Sangers’ dideoxymethod

5’-labelled oligonucleotid

ssDNA

+ DNA-polymeras

+ dNTPs

+ ddATP

+ ddCTP

+ ddGTP

+ ddTTP

A

C

G

T

A

C

G

T

2’,3’-dideoxyanalog

DNA sequencing

Sangers’ dideoxymethod
PCR (Polymerase Chain Reaction)

DNA-template

\[ \downarrow 94^\circ C \text{ separation of DNA-strands} \]

\[ \downarrow 45-65^\circ C \text{ annealing of specific primers} \]

\[ \downarrow 72^\circ C \text{ extension by Taq DNA polymerase} \]

Repeat 25-40 cycles (exponential increase of product)
SNP - single nucleotide polymorphism

Affect of SNP:
- Coding: different protein properties
- Regulatory: different amounts of mRNA (protein)
- Stability: stability of mRNA (or protein)
- Silent: neutral variations

One SNP per 1000 bp

Responsible for phenotypic differences between individuals

Can be identified by:
- Mass spectrometry
- Hybridization
- Pyrosequencing
  etc

Allele 1

<table>
<thead>
<tr>
<th>Allele 1</th>
<th>Allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G - C</td>
<td>G - C</td>
</tr>
<tr>
<td>C - G</td>
<td>C - G</td>
</tr>
<tr>
<td>A - T</td>
<td>A - T</td>
</tr>
<tr>
<td>A - T</td>
<td>A - T</td>
</tr>
<tr>
<td>T - A</td>
<td>C - G</td>
</tr>
<tr>
<td>C - G</td>
<td>C - G</td>
</tr>
<tr>
<td>G - C</td>
<td>G - C</td>
</tr>
<tr>
<td>G - C</td>
<td>G - C</td>
</tr>
<tr>
<td>T - A</td>
<td>T - A</td>
</tr>
<tr>
<td>C - G</td>
<td>C - G</td>
</tr>
<tr>
<td>T - A</td>
<td>T - A</td>
</tr>
</tbody>
</table>

Allele 2

p

q
STR- short tandem repeats / microsatellites

One microsatellite per 10 000 bp

Often in non-coding sequences

Can be identified through PCR and electrophoresis - DNA fingerprint
DNA chips - Analysis of gene expression

Compare gene expression from two samples (two tissues, healthy/disease etc)
DNA microarray

- Target (RNA/DNA) binds to probe (DNA) immobilized on chip surface
- Probe can be a PCR-product (cDNA), or synthetic oligonucleotides (50-70mer).
- The target is fluorescently labelled and binds through Watson Crick basepairing.
- Only relative (not absolute) amounts can be determined
cDNA microarray technology

**Probe**
- DNA clones (genes)
- PCR amplification
- Purification
- Robotic printing

**Target**
- Sample
- Reference
- mRNA
- Labelled cDNA
- Hybridize to microarray

**Image analysis**
- "red" raw data
- "green" raw data
- Image analysis

**Microarray / DNA chip**
Cytogenetics

Studies of chromosome morphology
(mainly in metaphase)

- number of chromosomes
- size
- shape

The karyotype differs between species even for closely related species.
Karyotyping

Principle: initiate cell arrest during mitosis and label the condensed chromosomes with Giemsa stain.

- Giemsa stains AT regions in particular and is detected as chromosome bands.
- Each chromosome have a unique staining pattern and morphology.
Fluorescence In Situ Hybridization (FISH)

**Approach:** Fluorescently labelled probes that binds to specific chromosome regions.

"**Probe-typer**":

- locus specific
  - gene localisation

- centromer specific
  - copy number estimation

- chromosome specific - (spectral karyotyping)
  - chromosome rearrangements (translocation, duplications, deletions)
Experimental systems for gene function investigation

• Mammalian cells (culture)
  – Transient or stable expression
  – Cancer cell lines

• Animal models
  – Transgenic animals by microinjection, retrovirus etc..
  – Knockout mice (yeast) by homolog recombination
  – Cloned animals by nuclear transfer
  – Gene silencing by inhibitory RNA (RNAi)
Animal models

- **Transgenic animal**
  - Retrovirus infections
  - Gene transfer via sperms
  - Injection into nucleus

- **Knockout animal (mouse and yeast)**
  - Homolog recombination and embryonic stem cells (ES)
  - Inactivation of specific genes - model system for diseases

- **Cloned animals**
  - Nuclear transfer - from somatic cells into embryonic cells (ES)
Expression in cells

- **Transient expression**
  - transfection of eukaryotic cells with vector
  - protein is expressed during 1-4 days then the vector is degraded/exported

- **Stable expression**
  - transfection of eukaryotic cells with vector
  - culture cells in presence of antibiotics (weeks)
  - a minority of cells incorporates vector DNA into the chromosome and survives in presence of antibiotics.
  - surviving cells are isolated and the gene function is investigated in following experiments.