Techniques in Molecular Biology
(to study the function of genes)

Analysis of nucleic acids:
- Polymerase chain reaction (PCR)
- Gel electrophoresis
- Blotting techniques (Northern, Southern)

Gene expression analysis:
- Real-time PCR
- Microarrays (DNA chips)

Recombinant DNA technology (Cloning of DNA fragments)
Sanger sequencing & next-generation sequencing

Methods to study gene function:
- Cell culture methods
- Generation of transgenic mice and knock-out mice
Polymerase chain reaction (PCR): Amplification of DNA

- Can be used to amplify rare specific DNA sequences from a complex mixture when the ends of the sequence are known
- PCR amplification of mutant alleles allows detection of human genetic diseases
- DNA sequences can be amplified by PCR for use in cloning, as probes, and in forensics
- Kary Mullis: Nobel price for Chemistry 1993

**PCR reaction mix:**

- Buffer
- Template DNA
- Primers
- Nucleotides (dNTPs)
- DNA polymerase
- Water
DNA molecules can be chemically synthesized

- Rather than using existing DNA pieces new DNA (up to several kilobases) can be designed and synthesized
- Usefull for:
  - * in vitro generation of new gene structures (eg. mutant genes)
  - * polylinker generation (with MCS)
  - * “Primers” and “probes” (short oligos matching a sequence of interest)
  - * Generating new “designer- genomes”?

Analysis of nucleic acids
Analysis of nucleic acids

Gel electrophoresis resolves DNA fragments of different size

Visualization of DNA/RNA samples separated by gel electrophoresis

DNA restriction fragments

Place mixture in the well of an agarose or polyacrylamide gel. Apply electric field

Well

Gel particle

Pores

Molecules move through pores in gel at a rate inversely proportional to their chain length
Analysis of nucleic acids

**Southern blotting** detects specific **DNA** fragments

**Northern blotting** detects specific **mRNAs**

Q: is a specific sequence (ie a gene) present in my DNA sample?

A: yes!

Q: is the Beta-globin gene expressed? (= is the mRNA present in my total RNA sample?)

A: yes!

And its expression increases over time
Gene expression analysis

Real-time PCR for gene expression analysis

- Allows real time monitoring of PCR amplification products by use of fluorescence
- First step: Reverse transcription of RNA into complementary DNA (cDNA)
- Either quantitative or semi-quantitative measurement (qPCR)
- Highly sensitive
- Used in research lab and clinical diagnostic labs
Gene expression analysis

Fluorescence chemistry for real-time PCR

SybrGreen intercalating dye:

TaqMan™ probe uses FRET (Fluorescent resonance energy transfer):
Gene expression analysis

Plotting PCR in real-time

<table>
<thead>
<tr>
<th>CYCLE NUMBER</th>
<th>AMOUNT OF DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>64</td>
</tr>
<tr>
<td>7</td>
<td>128</td>
</tr>
<tr>
<td>8</td>
<td>256</td>
</tr>
<tr>
<td>9</td>
<td>512</td>
</tr>
<tr>
<td>10</td>
<td>1,024</td>
</tr>
<tr>
<td>11</td>
<td>2,048</td>
</tr>
<tr>
<td>12</td>
<td>4,096</td>
</tr>
<tr>
<td>13</td>
<td>8,192</td>
</tr>
<tr>
<td>14</td>
<td>16,384</td>
</tr>
<tr>
<td>15</td>
<td>32,768</td>
</tr>
<tr>
<td>16</td>
<td>65,536</td>
</tr>
<tr>
<td>17</td>
<td>131,072</td>
</tr>
<tr>
<td>18</td>
<td>262,144</td>
</tr>
<tr>
<td>19</td>
<td>524,288</td>
</tr>
<tr>
<td>20</td>
<td>1,048,576</td>
</tr>
<tr>
<td>21</td>
<td>2,097,152</td>
</tr>
<tr>
<td>22</td>
<td>4,194,304</td>
</tr>
<tr>
<td>23</td>
<td>8,388,608</td>
</tr>
<tr>
<td>24</td>
<td>16,777,216</td>
</tr>
<tr>
<td>25</td>
<td>33,554,432</td>
</tr>
<tr>
<td>26</td>
<td>67,108,864</td>
</tr>
<tr>
<td>27</td>
<td>134,217,728</td>
</tr>
<tr>
<td>28</td>
<td>268,435,456</td>
</tr>
<tr>
<td>29</td>
<td>536,870,912</td>
</tr>
<tr>
<td>30</td>
<td>1,073,741,824</td>
</tr>
<tr>
<td>31</td>
<td>1,400,000,000</td>
</tr>
<tr>
<td>32</td>
<td>1,500,000,000</td>
</tr>
<tr>
<td>33</td>
<td>1,550,000,000</td>
</tr>
<tr>
<td>34</td>
<td>1,580,000,000</td>
</tr>
</tbody>
</table>
Gene expression analysis

Amplification curve to determine relative expression values

Output: relative fold-change
(for semiquantitative qPCR)

\[
R = \frac{(E_{\text{target}})^{\Delta C_{\text{P target}} (\text{control} - \text{sample})}}{(E_{\text{ref}})^{\Delta C_{\text{P ref}} (\text{control} - \text{sample})}}
\]
DNA microarrays consist of thousands of individual gene sequences bound to closely spaced regions on the surface of a glass microscope slide or synthesized sequences on a chip surface.

DNA microarrays allow the simultaneous analysis of the expression of thousands of genes.

The combination of DNA microarray technology with genome sequencing projects enables scientists to analyze the complete transcriptional program of an organism during specific physiological response or developmental processes.
Gene expression analysis

Scan image of a DNA microarray

Each spot contains many copies of a unlabeled DNA-probe corresponding to part of one specific mRNA. The labeled cDNA hybridizes to spots where there is a sequence match (red+green=yellow; identical expression level in both samples)
Cluster analysis of µArray data can reveal coexpressed genes

Each column represents a different gene at times after addition of serum

Bioinformatic analysis (*in silico*) can then be applied to investigate common functions of clustered genes (=gene ontologies) or if several genes can be mapped to common biological pathways (glycolysis, cell cycle control, etc.)
Investigating genes with biotechnological approaches is important because:

**Applied therapeutic cloning:**
- Producing big amounts of identical copies of a DNA of interest → “gain-of-function” analysis
- Expression of genes in other organisms (→ transgenic organisms)
  * to obtain large amounts of a protein (example protein drugs such as insulin are produced in bacteria or yeast from a human gene)
  * to investigate the function of a protein (coded by its gene) in cell-culture models or in model animals (e.g. disease-related genes or gene mutations (oncogenes) can be studied in detail in order to find and test the effects of drug-treatments)
  * generating functional DNA molecules for gene therapy in gene-related diseases

**Genetically modified organisms (GMO):**
- GMO's are living organisms where a gene has been deleted (knocked out), added (transgenic organism), or changed (mutant transgene) for the purpose of:
  * Generating improved agricultural crops with resistance against vermin or have increased content of “healthy protein”
  * Generating transgene animals. In basic medical research transgene animals (mainly mice) are extremely important for the understanding of diseases and possible treatments.
  * In farming: Generation of healthy animals with a improved production potential
Recombinant DNA technology enables to produce large numbers of identical DNA molecules (DNA-cloning). Clones are typically generated by placing a DNA fragment of interest into a vector DNA molecule, which can replicate in bacterial host cells. When a single vector containing a single DNA fragment is introduced into a host cell, large numbers of the fragment are reproduced (cloned) along with the vector. Two common vectors are *E. coli* plasmid vectors and bacteriophage λ vectors.

Plasmids are extrachromosomal, self-replicating DNA molecules.
Cloning

The general procedure for cloning with recombinant DNA plasmid vectors:

1. **Plasmid vector**
   - Enzymatically insert DNA into plasmid vector

2. **Recombinant plasmid**
   - Mix *E. coli* cells with plasmids in presence of CaCl₂
   - Culture on nutrient agar plates containing ampicillin

3. **Bacterial chromosome**
   - Transformed *E. coli* cell survives
   - Cells that do not take up plasmid die on ampicillin plates

4. **Cell multiplication**
   - Colony of cells each containing copies of the same recombinant plasmid
Plasmid cloning permits isolation of DNA fragments from complex mixtures.

Each colony represents one clone

(= a group of organisms with identical DNA)
Restriction enzymes cut DNA molecules at specific sequences

- Restriction enzymes are site-specific DNAses generating double-strand breaks
- Until now more than 600 naturally occurring restriction enzymes identified
- They recognize 4-6 basepair sequence patterns
- Restriction cuts are most often asymmetric: generating single-strand overhangs (sticky ends)
- Restriction cuts are less frequently symmetric (Blunt end)

Restriction enzyme recognition sites are usually palindromic

Examples of palindromes:
Anna, otto, radar, kayak
## Cloning

#### Selected restriction enzymes

<table>
<thead>
<tr>
<th>ENZYME</th>
<th>SOURCE MICROORGANISM</th>
<th>RECOGNITION SITE</th>
<th>ENDS PRODUCED</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BamHI</strong></td>
<td><em>Bacillus amyloliquefaciens</em></td>
<td>↓ -G-G-A-T-C-C- ↑ -C-C-T-A-G-G-</td>
<td>Sticky</td>
</tr>
<tr>
<td><strong>Sau3A</strong></td>
<td><em>Staphylococcus aureus</em></td>
<td>↓ -G-A-T-C- ↑ -C-T-A-G-</td>
<td>Sticky</td>
</tr>
<tr>
<td><strong>SmaI</strong></td>
<td><em>Serratia marcescens</em></td>
<td>↓ -C-C-C-G-G-G- ↑ -G-G-G-C-C-C-</td>
<td>Blunt</td>
</tr>
<tr>
<td><strong>NcoI</strong></td>
<td><em>Nocardia otitidis-caviarum</em></td>
<td>↓ -G-C-G-G-C-C-G-C- ↑ -C-G-C-G-G-C-G-</td>
<td>Sticky</td>
</tr>
</tbody>
</table>
Restriction fragments with complementary “sticky ends” are ligated easily.

Recombinant DNA:
Cloning

Polylinkers facilitate insertion of restriction fragments into plasmid vectors

(a) Sequence of polylinker

GAATTCTCGAGCTCGGTACCCGGGGATCTCTTAGAGTCGACCTGCAGCGATGCAAGCTT

EcoRI  Kpnl  BamHI  Sall  PstI  HindIII

(b) Insertion of EcoRI restriction fragments

![DNA digestion and ligation process]
Producing high levels of proteins from cloned cDNAs

- Many proteins are normally expressed at very low concentrations within cells, which makes isolation of sufficient amounts for analysis difficult.
- To overcome this problem, DNA expression vectors can be used to produce large amounts of full length proteins.

Example: *E. coli* expression systems can produce full-length proteins from recombinant genes.
Cloning the transcriptome:

Complementary DNA (cDNA) libraries are prepared from isolated mRNAs.

**Preparation of a bacteriophage λ cDNA library**

1. **Hybridize with oligo-dT primer**
2. **Transcribe RNA into cDNA**
3. **Remove RNA with alkali**
4. **Add poly(dG) tail**
5. **Synthesize complementary strand**
6. **Protect cDNA by methylation**
7. **Ligate cDNA to linkers**
8. **Cleave with EcoRI**
9. **Ligate to λ arms**
10. **Package in vitro**
11. **Infect E. coli**
12. **Individual λ cDNA clones**

**Mix under hybridization conditions**

**Mix under hybridization conditions**

**Wash away rRNA and tRNA**

**Elute column in low-salt buffer**

**Purified mRNA preparation**

**Single-stranded cDNA**

**Double-stranded cDNA**

**Mixture of cytoplasmic RNAs**
Identifying, analyzing, and sequencing cloned DNA

- The most common approach to identifying a specific clone involves screening a library by hybridization with radioactively labeled DNA or RNA probes.

- Labeled nucleotide probes are also used to identify specific nucleotide sequences in a complex mixture of DNA or mRNA (Southern and Northern blot).

- Direct sequencing is also often applied
  - Single nucleotide sequencing (i.e., sequencing a cloned DNA or cDNA)
  - Large scale sequencing: sequencing whole libraries of DNA or cDNA (i.e., sequencing a whole genome or a whole transcriptome)
DNA sequencing: the Sanger (di-deoxy) method

5' T A G C T G A C T C 3'  
3' A T C G A C T G A G T C A A G A A C T A T T G G G C T T A A ...

DNA polymerase  
+ dATP, dGTP, dCTP, dTTP  
+ ddGTP in low concentration

5' T A G C T G A C T C A G 3'  
3' A T C G A C T G A G T C A A G A A C T A T T G G G C T T A A ...

5' T A G C T G A C T C A G T T C T T G 3'  
3' A T C G A C T G A G T C A A G A A C T A T T G G G C T T A A ...

5' T A G C T G A C T C A G T T C T T G A T A A C C C G 3'  
3' A T C G A C T G A G T C A A G A A C T A T T G G G C T T A A ...

Deoxyribonucleoside triphosphate (dNTP)  
Dideoxyribonucleoside triphosphate (ddNTP)
DNA sequencing: the Sanger (dideoxy) method

Primer 5’ ———— DNA polymerase + dNTPs (100 μM) ———— 5’

Template 3’  

+ ddATP (1 μM)  
+ ddGTP (1 μM)  
+ ddTTP (1 μM)  
+ ddCTP (1 μM)

A  
G  
T  
C

etc.  

denature and separate daughter strands by electrophoresis

ATGCTCTCGGCAAATATA
Next generation sequencing

Technologies/Platforms:
- Roche/454 FLX: 2004
- Illumina Solexa Genome Analyzer: 2006
- Applied Biosystems SOLiD System: 2007
- Helicos Heliscope: 2010
- Pacific Biosciencies SMRT: 2010
- LifeTechnologies Ion Torrent: 2011

Parameters:
- Cost (device, cost/Mb)
- Read length
- Speed
- Accuracy
- Preparation time
- Manipulation steps (amplification needed)
<table>
<thead>
<tr>
<th>Platform</th>
<th>Library/template preparation</th>
<th>NGS chemistry</th>
<th>Read length (bases)</th>
<th>Run time (days)</th>
<th>Gb per run</th>
<th>Machine cost (US$)</th>
<th>Pros</th>
<th>Cons</th>
<th>Biological applications</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roche/454 GS FLX Titanium</td>
<td>Frag, MP/ emPCR</td>
<td>PS</td>
<td>330</td>
<td>0.25</td>
<td>0.45</td>
<td>500,000</td>
<td>Longer reads improve mapping in repetitive regions; fast run times</td>
<td>High reagent cost; high error rates in homopolymer repeats</td>
<td>Bacterial and insect genome de novo assemblies; medium scale (&lt;3 Mb) exome capture; 16S in metagenomics</td>
<td>D. Muzny, pers. comm.</td>
</tr>
<tr>
<td>Illumina/Solexa's GA</td>
<td>Frag, MP/ solid-phase</td>
<td>RTs</td>
<td>75 or 100</td>
<td>4, 9</td>
<td>18, 35</td>
<td>540,000</td>
<td>Currently the most widely used platform in the field</td>
<td>Low multiplexing capability of samples</td>
<td>Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics</td>
<td>D. Muzny, pers. comm.</td>
</tr>
<tr>
<td>Life/454's SOLID 3</td>
<td>Frag, MP/ emPCR</td>
<td>Cleaveable probe SBL</td>
<td>50</td>
<td>7, 14</td>
<td>50, 50</td>
<td>595,000</td>
<td>Two-base encoding provides inherent error correction</td>
<td>Long run times</td>
<td>Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics</td>
<td>D. Muzny, pers. comm.</td>
</tr>
<tr>
<td>Polonator G.007</td>
<td>MP only/ emPCR</td>
<td>Non-cleaveable probe SBL</td>
<td>26</td>
<td>5</td>
<td>12</td>
<td>170,000</td>
<td>Least expensive platform; open source to adapt alternative NGS chemistries</td>
<td>Users are required to maintain and quality control reagents; shortest NGS read lengths</td>
<td>Bacterial genome resequencing for variant discovery</td>
<td>J. Edwards, pers. comm.</td>
</tr>
<tr>
<td>Helicos BioSciences HeliScope</td>
<td>Frag, MP/ single molecule</td>
<td>RTs</td>
<td>32</td>
<td>8</td>
<td>37</td>
<td>000,000</td>
<td>Non-bias representation of templates for genome and seq-based applications</td>
<td>High error rates compared with other reversible terminator chemistries</td>
<td>Seq-based methods</td>
<td>01</td>
</tr>
<tr>
<td>Pacific Biosciences</td>
<td>Frag only/ single molecule</td>
<td>Real-time</td>
<td>964</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Has the greatest potential for reads exceeding 1 kb</td>
<td>Highest error rates compared with other NGS chemistries</td>
<td>Full-length transcriptome sequencing; complements other resequencing efforts in discovering large structural variants and haplotype blocks</td>
<td>S. Turner, pers. comm.</td>
</tr>
</tbody>
</table>

*Average read-lengths. 1Fragment run. 2Mate-pair run. Frag, fragment; GA, Genome Analyzer; GS, Genome Sequencer; MP, mate-pair; N/A, not available; NGS, next-generation sequencing; PS, pyrosequencing; RT, reversible terminator; SBL, sequencing by ligation; SOLID, support oligonucleotide ligation detection.
## Next generation sequencing

<table>
<thead>
<tr>
<th></th>
<th>Roche (454)</th>
<th>Illumina</th>
<th>SOLiD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>Pyrosequencing</td>
<td>Polymerase-based</td>
<td>Ligation-based</td>
</tr>
<tr>
<td>Amplification</td>
<td>Emulsion PCR</td>
<td>Bridge Amp</td>
<td>Emulsion PCR</td>
</tr>
<tr>
<td>Paired ends/sep</td>
<td>Yes/3kb</td>
<td>Yes/200 bp</td>
<td>Yes/3 kb</td>
</tr>
<tr>
<td>Mb/run</td>
<td>100 Mb</td>
<td>1300 Mb</td>
<td>3000 Mb</td>
</tr>
<tr>
<td>Time/run</td>
<td>7 h</td>
<td>4 days</td>
<td>5 days</td>
</tr>
<tr>
<td>Read length</td>
<td>250 bp</td>
<td>32-40 bp</td>
<td>35 bp</td>
</tr>
<tr>
<td>Cost per run (total)</td>
<td>$8439</td>
<td>$8950</td>
<td>$17447</td>
</tr>
<tr>
<td>Cost per Mb</td>
<td>$84.39</td>
<td>$5.97</td>
<td>$5.81</td>
</tr>
</tbody>
</table>
Roche (454) Workflow

Roche (454) GSFLX Workflow:

Library construction → Emulsion PCR → PTP loading

Signal image

DNA capture bead containing millions of copies of a single clonally amplified fragment

Sulfurylase

Luciferase

Pyrosequencing reaction

TRENDS in Genetics
Sequencing

Illumina (Solexa) Workflow

Prepare genomic DNA sample
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

Attach DNA to surface
Bind single-stranded fragment randomly to the inside surface of the flow cell channels.

Bridge amplification
Add unincorporated nucleotides and enzyme to initiate solid-phase bridge amplification.

Denature the double stranded molecules

First chemistry cycle; determine first base
To initiate the first sequencing cycle, add all four labeled reversible terminators, primers, and DNA polymerase enzyme to the flow cell.

Image of first chemistry cycle
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

Before initiating the next chemistry cycle
The blocked 3' terminus and the fluorophore from each incorporated base are removed.

Sequence read over multiple chemistry cycles
Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

Mardis ER. 2008.
Sequencing

ABI SOLiD Workflow

(a) Solid sequencing process

(b) Principles of two base encoding
An integrated semiconductor device enabling non-optical genome sequencing

Sequencing

Applications

- Genomes
- Re-sequencing Human Exons (Microarray capture/amplification)
- small (including mi-RNA) and long RNA profiling (including splicing)
- ChIP-Seq:
  - Transcription Factors
  - Histone Modifications
  - Effector Proteins
- DNA Methylation
- Polysomal RNA
- Origins of Replication/Replicating DNA
- Whole Genome Association (rare, high impact SNPs)
- Copy Number/Structural Variation in DNA
- ChIA-PET: Transcription Factor Looping Interactions
- The $1000$ genome

Current bottle neck: Data management!!!
Methods to study gene function

Studying the function of a gene/protein: Some basic principles and nomenclature

- **In vivo**: in a living organism (living bacteria, yeast, monkey, christmas tree)
- **Ex vivo**: in an organ explant, e.g. a tissue biopsy
- **In vitro**: in a reagent tube; a synthetic setup
- **In silico**: computer-based simulation and/or data-analysis

---

**Diagram:**

- **Mutant organism/cell**
  - Comparison of mutant and wild-type function

- **Gene inactivation**

- **Cloned gene**
  - DNA sequencing

- **Database search to identify protein-coding sequence**
  - PCR isolation of corresponding gene

- **Expression in cultured cells**

- **Protein**
  - Localization
  - Biochemical studies
  - Determination of structure
Advantages of working with cultured cells over intact organisms

- More homogeneous than cells in tissues
- Can control experimental conditions
- Can isolate single cells to grow into a colony of genetically homogeneous clone cells
- Ethically non-controversial

Commonly used cell-cultures are derived from:

- Bacteria
- Yeasts
- Vertebrates (Mouse, human, rat, hamster)
- Plants (Arabidopsis Thaliana)
Growth of microorganisms in culture

- Examples: *E. coli* and the yeast *S. cerevisiae*
- Have rapid growth rate and simple nutritional requirements
- Can be grown on semisolid agar
- Mutant strains can be isolated by replica plating
Methods to study gene function

Replica plating

- Sterile, velvet-covered stamp
- Colonies
- Missing colony (Arg⁻) requires arginine
  - Minimal medium (lacks arginine)
  - Rich medium (master plate)
  - Arg⁻ mutant colony
  - Minimal medium plus arginine
Methods to study gene function

Growth of animal cells in culture

- Requires rich media including essential amino acids, vitamins, salts, glucose, and serum
- Most grow only on special solid surfaces (others, eg. Bloodcells, grow in suspension)

A single mouse cell  A colony of human cells  Many colonies in a petri dish
Primary cells and cell lines

- Primary cell cultures are established from animal tissues
- Certain types of cells are easier to culture than others
- Most cells removed from an animal contain a fraction of cells that grow and divide for a limited period of time (about 50 doublings), then eventually die of old age.
- Certain “transformed cells” may arise that are **immortal** and can be used to form a **cell line**
- Transformed cells may be derived from tumors, from manipulation (**immortalisation** with oncogene) or may arise spontaneously
- The rate of spontaneous transformation varies for different species

**Establishment of a cell culture**

(a) Human cells

(b) Mouse cells
Fibroblasts can be cultured (primary/immortal) and differentiated
Fibroblasts can be cultured (primary/immortal) and differentiated.
Generation of transgenic mice and knock-out mice

- Loss of function
  - Knock-out mice
  - siRNA

- Gain of function
  - Transgenic mouse

- Flavours
  - Conditional (inducible system)
  - Tissue-specific
Mutations: types and causes

- An organism’s **genotype** is its entire set of genes and may denote whether an individual carries mutations in one or more genes. (genotype = basic genetic make-up of an individual organism)

- An organism’s **phenotype** is its function and physical appearance, and depends on that individual’s genotype.

- Normal (wild-type) organisms may develop changes termed **mutations** in their DNA sequence thereby altering their genotype and perhaps their phenotype.

- Organisms that develop such mutations are termed “mutants”.

- Mutations may be part of the genotype or occur as acquired (somatic) mutations.
Haploids and diploids

- A **haploid** organism has a single copy of each chromosome and its phenotype is a consequence of that one copy.

- A **diploid** organism has two copies of each chromosome and thus two copies of each gene.
  - Diploids that carry identical alleles are termed **homozygous**.
  - Diploids that carry different alleles are termed **heterozygous**.

- The two copies of each gene may be the same or the copies may be different. Different forms of each gene are termed **alleles** (allele = naturally occurring mutation).

- Alleles may be **dominant** or **recessive**.

<table>
<thead>
<tr>
<th>Haploids</th>
<th>Diploids</th>
</tr>
</thead>
<tbody>
<tr>
<td>A haploid organism has a single copy of each chromosome and its phenotype is a consequence of that one copy.</td>
<td>A diploid organism has two copies of each chromosome and thus two copies of each gene.</td>
</tr>
<tr>
<td>Diploids that carry identical alleles are termed homozygous.</td>
<td>Diploids that carry different alleles are termed heterozygous.</td>
</tr>
<tr>
<td>The two copies of each gene may be the same or the copies may be different. Different forms of each gene are termed alleles (allele = naturally occurring mutation).</td>
<td></td>
</tr>
<tr>
<td>Alleles may be dominant or recessive.</td>
<td></td>
</tr>
</tbody>
</table>

Methods to study gene function

- **Methods to study gene function**

- **Haploids and diploids**

- A **haploid** organism has a single copy of each chromosome and its phenotype is a consequence of that one copy.

- A **diploid** organism has two copies of each chromosome and thus two copies of each gene.
  - Diploids that carry identical alleles are termed **homozygous**.
  - Diploids that carry different alleles are termed **heterozygous**.

- The two copies of each gene may be the same or the copies may be different. Different forms of each gene are termed **alleles** (allele = naturally occurring mutation).

- Alleles may be **dominant** or **recessive**.
**Methods to study gene function**

### Geno-type linked diseases (non-cancerous)

<table>
<thead>
<tr>
<th>TABLE 5-2</th>
<th>Common Inherited Human Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DISEASE</strong></td>
<td><strong>MOLECULAR AND CELLULAR DEFECT</strong></td>
</tr>
<tr>
<td>AUTOSOMAL RECESSIVE</td>
<td></td>
</tr>
<tr>
<td>Sickle-cell anemia</td>
<td>Abnormal hemoglobin causes deformation of red blood cells, which can become lodged in capillaries; also confers resistance to malaria.</td>
</tr>
<tr>
<td>Cystic fibrosis</td>
<td>Defective chloride channel (CFTR) in epithelial cells leads to excessive mucus in lungs.</td>
</tr>
<tr>
<td>Phenylketonuria (PKU)</td>
<td>Defective enzyme in phenylalanine metabolism (tyrosine hydroxylase) results in excess phenylalanine, leading to mental retardation, unless restricted by diet.</td>
</tr>
<tr>
<td>Tay-Sachs disease</td>
<td>Defective hexosaminidase enzyme leads to accumulation of excess sphingolipids in the lysosomes of neurons, impairing neural development.</td>
</tr>
<tr>
<td>AUTOSOMAL DOMINANT</td>
<td></td>
</tr>
<tr>
<td>Huntington's disease</td>
<td>Defective neural protein (huntingtin) may assemble into aggregates causing damage to neural tissue.</td>
</tr>
<tr>
<td>Hypercholesterolemia</td>
<td>Defective LDL receptor leads to excessive cholesterol in blood and early heart attacks.</td>
</tr>
<tr>
<td>X-LINKED RECESSIVE</td>
<td></td>
</tr>
<tr>
<td>Duchenne muscular dystrophy (DMD)</td>
<td>Defective cytoskeletal protein dystrophin leads to impaired muscle function.</td>
</tr>
<tr>
<td>Hemophilia A</td>
<td>Defective blood clotting factor VIII leads to uncontrolled bleeding.</td>
</tr>
</tbody>
</table>
Segregation of geno-type linked diseases

Methods to study gene function

(a) Autosomal dominant: Huntington’s disease

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A^{HD}/A^+$</td>
<td>$A^+/A^+$</td>
</tr>
<tr>
<td>Affected</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Males and females

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A^{HD}/A^+$</td>
<td>Affected</td>
<td></td>
</tr>
<tr>
<td>$A^+/A^+$</td>
<td>Not affected</td>
<td></td>
</tr>
</tbody>
</table>

(b) Autosomal recessive: Cystic fibrosis

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A^{CFTR}/A^+$</td>
<td>$A^{CFTR}/A^+$</td>
</tr>
<tr>
<td>Carrier</td>
<td></td>
<td>Carrier</td>
</tr>
</tbody>
</table>

Males and females

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A^{CFTR}/A^{CFTR}$</td>
<td>Affected</td>
<td></td>
</tr>
<tr>
<td>$A^{CFTR}/A^+$</td>
<td>Carrier</td>
<td></td>
</tr>
<tr>
<td>$A^+/A^{CFTR}$</td>
<td>Carrier</td>
<td></td>
</tr>
<tr>
<td>$A^+/A^+$</td>
<td>Noncarrier</td>
<td></td>
</tr>
</tbody>
</table>

(c) X-linked recessive: Duchenne muscular dystrophy

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X^+/Y$</td>
<td>$X^{DMD}/X^+$</td>
</tr>
<tr>
<td>Carrier</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Males and females

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{DMD}/Y$</td>
<td>Affected</td>
<td></td>
</tr>
<tr>
<td>$X^+/Y$</td>
<td>Unaffected</td>
<td></td>
</tr>
<tr>
<td>$X^{DMD}/X^+$</td>
<td>Carrier</td>
<td></td>
</tr>
<tr>
<td>$X^+/X^+$</td>
<td>Noncarrier</td>
<td></td>
</tr>
</tbody>
</table>
Methods to study gene function

Example: mutants in *Drosophila*

Mutations may cause only subtle changes or may produce significant changes in development, cellular function, appearance and/or behavior.
Gene replacement and transgenic animals

- It is possible to replace an organism’s wild type gene with an inactive gene to create a “gene knockout”

- It is also possible to introduce additional genes (transgenes) to create a transgenic organism

- Gene knockout and transgenic techniques can involve mutagenesis of cloned genes prior to transfer into the organism (= *in vitro* mutagenesis)
Creation of mice embryonic stemm cells (ES cells) carrying a knockout mutation

Specific targeting of gene of interest (GOI)

Formation of ES cells carrying a knockout mutation

- Homologous recombination
- Nonhomologous recombination

Positive and negative selection of recombinant ES cells

- Recombinants with random insertion
- Recombinants with gene-targeted insertion

Treat with G-418 (positive selection)
Treat with ganciclovir (negative selection)

ES cells with targeted disruption in gene X
Gene knockout in mice

Methods to study gene function

1. Inject ES cells into blastocoeal cavity of early embryos
   - Brown mouse: $A/A, X^{+}/X^{+}$
   - Black mouse: $a/a, X^{+}/X^{+}$
   - 4.5-day blastocyst

2. Surgically transfer embryos into pseudopregnant female
   - Foster mother
   - Possible progeny
     - Chimeric
     - Black

3. Select chimeric mice for crosses to wild-type black mice
   - Possible germ cells: $A/X^{+}; A/X^{-}; a/X^{+}$
   - All germ cells: $a/X^{+}$
   - ES cell-derived progeny will be brown
     - $A/a, X^{+}/X^{+}$
     - $A/a, X^{-}/X^{+}$
     - $a/a, X^{+}/X^{+}$

4. Screen brown progeny DNA to identify $X^{-}/X^{+}$ heterozygotes
5. Mate $X^{-}/X^{+}$ heterozygotes
6. Screen progeny DNA to identify $X^{-}/X^{-}$ homozygotes
   - Knockout mouse
Cell-type-specific gene knockouts in mice

**Advantage:**
Effect on specific tissue/process can be studied

Embryonal lethal knock-outs can be studied in tissues where the knock out is not lethal

Knock-out can be induced chemically

Methods to study gene function
Methods to study gene function

**Gene-silencing with small interfering RNAs (siRNA)**

(a) In vitro production of double-stranded RNA

Dicer is an endogenous enzyme which cleaves specific RNA structures (part of the miRNA system)

(b) Noninjected Injected

(c) In vivo production of double-stranded RNA

siRNAs are designed to basepair with specific mRNA, which will then be degraded
Production of transgenic mice

The foreign DNA contains a promoter+GOI

If the promoter is a constitutive active promoter the GOI will be expressed in all cells

If the promoter is regulated by proteins only present in certain cells, a tissue-specific expression can be achieved
Methods to study gene function

Example: transgenic mouse (GH+)