Molecular pathology - effect and detection of mutations, single gene disorders, dideoxysequencing and NGS

What is molecular pathology?

- **Pathology**  
  Study of diseases

- **Molecular biology**  
  Study of macromolecules essential for life, such as nucleic acids and proteins, and their role in cell replication and transmission of genetic information

Many technologies applied

- **Molecular pathology**  
  Integrated knowledge and techniques applied in molecular biology to clinical question in pathology

Also, the clinical laboratory in which molecular diagnostic test are preformed
Molecular pathology – interpretation of genetic changes

- Seems to explain why a given genetic change should result in a particular clinical phenotype
- Requires us to work out the effect of a mutation on the quantity or function of the gene product
- Explains why the change is or is not pathogenic for any particular cell, tissue or stage of development
- Deals with inherited and acquired genetic changes
- Tries to explain genotype-phenotype correlation
Origins and types of mutations

• A mutation can be defined as any change in the primary nucleotide sequence of DNA regardless of its functional consequences.

• In clinical genetics – mutations are pathological genetic changes.

Some mutations can be lethal, others less deleterious, and some may confer as evolutionary advantage.

Mutations can occur in germine (sperm or oocytes); these can be transmitted to progeny.

Alternatively, mutations can occur during embryogenesis or in somatic tissues.

• Mutations that occur during development lead to mosaicism, a situation in which tissues are composed of cells with different genetic constitution.

• Other somatic mutations are associated with neoplasia because they confer a growth advantage to cells.

Epigenetic events, heritable changes that do not involve changes in gene sequence (e.g., altered DNA methylation) may influence gene expression or facilitate genetic damage.

Mutation – arise in single individuals – in germline or in somatic cells

Germline mutations

• Allelic sequence variation – DNA polymorphism if more than one variant (allele) occur in human population with frequency greater than 1% (lecture 2).

• Phenotypic abnormality.

• Increased susceptibility to disease; disease-causing mutations.

Low level of mutation may therefore be viewed as a balance between permitting occasional novelty at the expense of causing disease or death in a portion of a species.

• Somatic mutations in cells, e.g. tumor cells

Cancer as a microevolutionary process - it is a disease in which individual mutant clones of cells begin by prospering at the expense of their neighbors, but in the end destroy the whole cellular society.
Deciding whether a DNA sequence change is pathological can be difficult

Not every sequence variant seen in an affected person is necessarily pathogenic. How can we decide whether a sequence change we have discovered is sought for pathogenic mutation or harmless variant? In descending order of reliability, the criteria are:

1. Functional studies showing that the change is pathogenic;
2. The nature of the sequences change;
3. Precedent: that change has been seen before in patients with this disease (and not in ethnically matched controls) – *mutations databases*

DNA in human genome is not a static entity – it is subjected to a variety of different types of heritable changes (mutation)

- Whole genome such as triplodies
- Large-scale chromosome abnormalities
  - Numerical - Loss or gain of chromosomes
  - Structural - Breakage and rejoining of chromatides
- Small-scale mutations can be grouped into different mutation classes according to the effect on the DNA sequence
  - Base substitution - involve replacement of usually a single base
  - Deletion - one or more nucleotides are eliminated from a sequence
  - Insertions - one or more nucleotide are inserted into a sequence
DNA in human genome is not a static entity – it is subjected to a variety of different types of heritable changes (mutation)

- Small-scale mutations can be grouped into different mutation classes according to the effect on the DNA sequence
  
  Base substitution – involve replacement of usually a single base
    
    ....CTGACCTTT......
    ....CTGTCCTTT......
  
  Deletion – one or more nucleotides are eliminated from a sequence
    
    ....CTGACCTTT......
    ....CTGCTTTTT......
  
  Insertion – one or more nucleotide are inserted into a sequence
    
    ....CTGACCTTTTT......
    ....CTGACCTTTTTTT......

**Variation in the genome**

- The human genome is polymorphic, that is it shows variation
  
  - polymorphism is more frequent in non-coding regions of the genome, where it has few consequences

- Variation can occur at
  
  - The level of the chromosome
  - The level of individual base sequence
DNA polymorphism

- strictly - the existence of two or more variants at significant frequencies in the populations;

1) Any sequence variants present at a frequency more than 1% in a population
2) Any non-pathogenic sequence variant, regardless of frequency

DNA polymorphism

- **Single nucleotide polymorphism** (SNP) – single base substitution public dbSNP
  - Designed by numbering beginning with rs (reference SNP)

- **Variable number of tandem repeats** (VNTR) – describes alleles at loci containing tandemly repeated runs of a simple sequence
  - **Microsatellites** – STR (short tandem repeats)
  - **Minisatellite** DNA polymorphism

- **Copy number variations** - variation between individuals in the number of copies of a particular DNA sequences in their genome
Three types of single nucleotide polymorphisms
SNP

Substitution
A

Substitution
resulting in
RSP
B

Insertion/
deletion
C

Short tandem repeats (microsatellites) – important for family and forensic studies

1  tctaaaattttaa gtgtggttc ccagataatc tctactaat aatagc agatcagct
61 atgtgacaag ggtgattttc ctctttggta tccttatgta atattttgaa gata
gata
gata
tactaataaa agtatatgtt aatagcagct
121 tagatagata gtagatagata tagatagata gtagatagata tagatagata tagaggtgtaa aataaggata gata
181 cagatagata tctactaatgtaaactgtg gctatgattg gaatcacttg gctaaaaagc
gctataagc tctctctgnga gaggcaatct tttttnct taggnacnct ctcancagtc
gctaaagcct ttttctgnga gaggcaatct tttttnct taggnacnct ctcancagtc
gctaaagcct ttttctgnga gaggcaatct tttttnct taggnacnct ctcancagtc
gctaaagcct ttttctgnga gaggcaatct tttttnct taggnacnct ctcancagtc

• 1989 (short tandem repeats STR)
• about 150,000 STR have been identified
• Usually di-, tri- and tetra-nucleotide repeats; mostly (CA)n repeats
• Many alleles, highly informative
• Can type by automated multiplex PCR
• Easy physical localization
• Distributed through genome
Single nucleotide and other small-scale changes are a common type of pathogenic change

- **The main classes of mutations**
  - Deletions
  - Insertion
  - Single base substitutions:
    - Missense mutations replace one amino acid with another in the gene product
    - Nonsense mutations replace an amino acid codon with a stop codon
  - Combination of the main classes can lead to
    - Splice site mutations create or destroy signals for exon-intron splicing
    - Frameshift can be produced by deletions, insertions or splicing errors
    - Dynamic mutations (tandem repeats that often change size on transmission to children)

**A.** Examples of mutations. The coding strand is shown with the encoded amino acid sequence. **B.** Chromatograms of sequence analyses after amplification of genomic DNA by polymerase chain reaction. Obr. Harrisson

<table>
<thead>
<tr>
<th>Wild-type</th>
<th>Silent mutation</th>
<th>Missense mutation</th>
<th>Nonsense mutation</th>
<th>1 bp Deletion with frameshift</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA AAG CTC CTA GCA CCG GGT GAC GAG AAG AGG</td>
<td>DNA AAG CTC CTA GCA CCG GGT GAC GAG AAG AGG</td>
<td>DNA AAG CTC CTA GCA CCG GGT GAC GAG AAG AGG</td>
<td>DNA AAG CTC CTA GCA CCG GGT GAC GAG AAG AGG</td>
<td>DNA AAG CTC CTA GCA CCG GGT GAC GAG AAG AGG</td>
</tr>
</tbody>
</table>

A point mutation occurring within the coding region leads to an amino acid substitution if the codon is altered. Point mutations that introduce a premature stop codon result in a truncated protein. Large deletions may affect a portion of a gene or an entire gene, whereas small deletions and insertions alter the reading frame if they do not represent a multiple of three bases. These ‘frameshift’ mutations lead to an entirely altered carboxy terminus. Mutations occurring in regulatory or intronic regions may result in altered expression or splicing of genes.
Missense mutations aminoacid substitution—

p.E6V- substitution of charged glutamic acid by uncharged valin

Effect depends from the type of the substituted aminoacid

a) Conservative substitution
b) Non-conservative substitution

Example - known mutation – missense mutation c.17A – T in sickle cell disease

Strategy – exactly defined molecular change and estimation of its presence/absence
Metódy: ARMS or RSP-PCR
5’-CCTNAGG-3’ (MstII)
allelic discrimination
PCR and successive sequencing
Nonsense and frameshift mutations

Substitution – nucleotide change resulting in nonsense codon

Small deletion or insertion – "frameshift" resulting in nonsense codon

Nonsense and frameshift mutations - nonsense mediated decay (NMD)

Mechanisms in cell that detects mRNAs containing premature termination codons and degrades them; thus usual result of a nonsense mutation is to prevent any expression of the gene.

If NMD does not work – truncated proteins are synthetized which are potentially more pathogenic than a simple absence of the protein.
Summary

• Mutation are divided at the level of the sequence changes in nucleotide substitutions, small deletions and insertion.
• According the effect of the sequence change are mutations synonymous, non-synonymous (missense, nonsense, splicing), frameshift and dynamic
• Strategy for the detection of known mutations is – ARMS, RSP PCR, allelic discrimination, or endpoint PCR and the sequencing of the PCR product.
• Nonsense and frameshift mutations can result into a truncated protein.

Nomenclature for describing sequence changes

Nucleotide substitution

\( g \) (genomic) or \( c \) DNA,
the A of the initiator ATG codon is +1; the immediately preceding base is -1.
mutation c.17A – T in sickle cell disease
Nomenclature for describing sequence changes

**Nucleotide substitution**

g (genomic) or c DNA,
the A of the initiator ATG codon is +1; the immediately preceding base is -1.

Give the nucleotide number followed by the change
g.1162G>A replace guanine at position 1162 by adenine
For changes within intron, when only the cDNA sequence is known in full, specify the intron number by IVSn or the number of the nearest exon position

c.621+1G>T or IVS4+1G>T Replace G by T at the first base of intron 4; exon 4 ends at nt 621

**Amino acid substitution**

Start with „p“ to indicate protein

Use the one-letter codes (e.g. A, alanin; C, cystein; D, aspartic acid)

Three-letter codes are also acceptable (e.g. Ala; alanin, Cys, cystein; asp, aspartic acid)

- p.R117H or Arg117His replace arginine 117 by histidine (the initiator methioninen is codon 1)

- p.G542X or Gly542Stop – glycine 542 replaced by a stop codon.
mutation c.17A – T in sickle cell disease

c.17A-T
p.Glu6Val (E6V) – typical sickle cell disease mutation in
Nomenclature for describing sequence changes

Deletions and insertion

Use \texttt{del} for deletion and \texttt{ins} for insertion.

For \textbf{DNA} changes the nucleotide position or interval comes first:
- \texttt{g.409\_410insC} – insert B between nt 409 and 410
- \texttt{c.6232\_6236del} or \texttt{c.6232\_6236depATTAG} – delete 5 nucleotides (which can be specified) starting with nt 6232 of the cDNA

For \textbf{amino acid} changes the amino acid symbols comes first:
- \texttt{p.F508del} – delete phenylalanine 508

Basic terminology

- An observed trait is referred to as a \textit{phenotype}
- The genetic information defining the phenotype is called the \textit{genotype}
- Alternative forms of a gene or a genetic marker are referred to as \textit{alleles}.
  1. Polymorphic variants of nucleic acids that have no apparent effect on gene expression or function
  2. Polymorphic variants of nucleic acids that have subtle effects on gene expression, thereby conferring the adaptive advantages associated with genetic diversity
  3. Allelic variants may reflect mutations in a gene that clearly alter its function
Basic terminology

- An observed trait is referred to as a **phenotype**.
- The genetic information defining the phenotype is called the **genotype**.
- Alternative forms of a gene or a genetic marker are referred to as **alleles**.
  1. Polymorphic variants of nucleic acids that have no apparent effect on gene expression or function.
  2. Polymorphic variants of nucleic acids that have subtle effects on gene expression, thereby conferring the adaptive advantages associated with genetic diversity.
  3. Allelic variants may reflect mutations in a gene that clearly alter its function.

Examples: Glu6Val (E6V) sickle cell mutation in the beta-globin gene.
          the F508 deletion of phenylalanine (F) in the CFTR gene.
Both are examples of allelic variants of these genes that result in disease.
- Pathogenic mutations.

Basic terminology

- An observed trait is referred to as a **phenotype**.
- The genetic information defining the phenotype is called the **genotype**.
- Alternative forms of a gene or a genetic marker are referred to as **alleles**.
  1. Polymorphic variants of nucleic acids that have no apparent effect on gene expression or function.
  2. Polymorphic variants of nucleic acids that have subtle effects on gene expression, thereby conferring the adaptive advantages associated with genetic diversity.
  3. Allelic variants may reflect mutations in a gene that clearly alter its function.

Because each individual has two copies of each chromosome (one inherited from the mother and one inherited from the father), he or she can have only two alleles at a given locus.
However, there can be many different alleles in the population.
The normal or common allele is usually referred to as **wild type**.
Common allele is usually referred to as *wild type*

- **Wild-type homozygote** - when alleles at a given locus are identical, the individual is *homozygous*
- If the alleles are different on the maternal and the paternal copy of the gene, the individual is *heterozygous* at this locus
- **Homozygous point mutations** – inheriting identical copies of a mutant allele occurs in many autosomal recessive disorders, particularly in circumstances of consanguinity
- If two different mutant alleles are inherited at a given locus, the individual is said to be a *compound heterozygote*
- **Hemizygous** is used to describe males with a mutation in an X chromosomal gene or a female with a loss of one X chromosomal locus.

**Classification of mutation according the the effect on the gene product - the biggest distinction in molecular pathology is between**

- **in loss of function mutations** the product has reduced or no function (most often produce recessive phenotypes because heterozygotes often function perfectly normally) specific condition - **haploinsufficiency inherited AD**
- **in gain of function mutations** the product does something positively abnormal (usually cause dominant phenotypes – i.e. they result in phenotypic alterations when a given single allele is affected, because the presence of normal allele does not prevent the mutant allele from behaving abnormally)**
Loss of function mutations can occur in all domains of a gene

Point mutations causing beta-thalassemia as example of allelic heterogeneity. The beta-globin gene is located in the globin gene cluster. Point mutations can be located in the promoter, the CAP site, the 5'-untranslated region, the initiation codon, each of the three exons, the introns, or the polyadenylation signal. Many mutations introduce missense or nonsense mutations, whereas others cause defective RNA splicing. Not shown here are deletion mutations of the -globin gene or larger deletions of the globin locus that can also result in thalassemia.

Loss of function is likely when point mutations in a gene produce the same pathological change or the same clinical phenotype as deletions of large larts of the gene

- mutational allelic heterogeneity
Specific cases of loss of function mutation can be inherited in autosomal dominant manner

- haploinsufficiency a 50% reduction in the level of the gene expression causes an abnormal phenotype which is then inherited in autosomal dominant manner; certain form of dosage sensitivity –

Gene products are transcription factors
Gene products are parts of quantitative signalling
Gene products that compete with each other

Specific cases of loss of function mutation can be inherited in autosomal dominant manner

- Dominant negative effect occurs when a mutant polypeptide not only loses its own function, but also interferes with the product of a normal allele in heterozygote

Mutant protein may interfere with the function of a multimeric protein complex (COLIA1, COLIA2 in osteogenesis imperfecta
Mutant protein may occupy binding sites on proteins or promoter response elements – thyroid hormone resistance, disorder, in which inactivated hormone receptor binds to target sequence and functions as an antagonist of normal receptor

- Epigenetic modifications can abolish gene function without a DNA sequence change –
DNA imprinting
Promoter methylation
Gain of function mutations

- Specific change
- Mutation spectrum is limited to few mutations causing the same clinical phenotype (deletions are not involved)
- **Mutational homogeneity** (few specific mutated alleles) is a typical characteristic
- Rare in inherited diseases
- Common in somatic cancer

**Loss of function vs. Gain of function**

- Loss of function is likely when point mutations in a gene produce the same pathological change as deletions – mutational heterogeneity
- Gain of function is likely when only a specific mutation in a gene produces a given pathology
- **Mutational homogeneity** is an indicator of a gain of function
  - Rare in monogenic diseases
  - Often in somatic cancer
Deciding whether a DNA sequence change is pathogenic can be difficult

Not every sequence variant seen in an affected person is necessarily pathogenic. How can we decide whether a sequence change we have discovered is sought for pathogenic mutation or harmless variant? In descending order of reliability, the criteria are:

1. Functional studies showing that the change is pathogenic;
2. The nature of the sequences change
3. Precedent: that change has been seen before in patients with this disease (and not in ethnically matched controls) – mutations databases
4. A de novo mutation, not present in the parents, in a person with de novo disease
5. A novel sequence change that is absent in a panel of, say 100 normal controls
Genetic Testing Categories

- **Research genetic testing**
  - Primary objective to advance understanding of disease pathophysiology and further therapeutic development process
  - Results cannot be reported for clinical use because labs not certified and proper quality control mechanisms may not be in place

- **Clinical genetic testing**
  - Performed for fee at request of MD or counselor
  - Labs must be certified to perform testing on human specimens for quality control
  - Patient gets a report and documented in the medical record

Genetic testing using methods of molecular biology

- Molecular detection of inherited diseases
- Molecular oncology
- Detection and identification of microorganisms and viruses
- DNA-based Tissue typing
- Pharmaco- and Nutrigenetics
Molecular detection of inherited diseases

- Field of medical genetics, based on the analysis of molecular nature of monogenic diseases
- Patterns of inheritance are determined by examination of family histories
  *Pedigree is a diagram of the inheritance pattern of a phenotype of family members*
  - **Autosomal-dominant**
  - **Autosomal-recessive**
  - **X-linked or sex-linked**

**Penetrance**

- refers to the proportion of individuals with a mutant genotype that express the phenotype.
- If all carriers of a mutant express the phenotype, penetrance is **complete**, whereas it is said to be **incomplete** or **reduced** if some individuals do not have any features of the phenotype.
- Dominant conditions with incomplete penetrance are characterized by skipping of generations with unaffected carriers transmitting the mutant gene. For example, hypertrophic obstructive cardiomyopathy breast cancer caused by mutations in the BRCA 1 or BRCA 2 genes is a dominant disorder with clinical features in only a subset of patients who carry the mutation.
- Patients who have the mutation but no evidence of the disease can still transmit the disorder to subsequent generations.

- When describing penetrance, one has to specify age. For example, for disorders such as Huntington disease, which present late in life, the rate of penetrance is influenced by the age at which the clinical assessment is performed.
- **Cumulative cancer risks for mutation carriers BRCA 1 at age 70 years were as follows: breast cancer risk of 57% and BRCA 2 of 49%**
Molecular basis of single-gene disorders - OMIM

14000 genes putative disease-association
3500 genes with disease-causing mutations

Molecular testing – confirmation of the clinical diagnosis
- differential diagnosis
- presymptomatic diagnosis
- diagnosis in the family
- prognosis und severity

Molecular diagnostics – 2 sequencing strategies

• Dideoxysequencing according Sanger
• Next generation sequencing (NGS)
Dideoxysequencing according Sanger - revision

Chain-termination method – based on amplification of the DNA fragment to be sequenced by DNA-polymerase and incorporation of modified nucleotides specifically, dideoxynucleotides
Both alleles sequenced in one reaction – fragments separated in capillary electrophoresis
Out-put - electropherogram

Confirmation of allelic discrimination results for rs1800782 by dideoxysequencing method
Sequencing technologies in molecular diagnostics

Dideoxysanger-Sequencing -
For small genes with relative low number of exones (Hemoglobinopaties – HBB 3 exones, Hereditäre hemochromatose – 6 exons

For „true“ single gene disorders
BRCA1/2 – different complicated strategies, because each amplicon is separated by CE; starting with mutational hot spots, possible 112 runs for all 112 amplicons

NGS - next-gen sequencing
Genetic diseases with multiple genes involved (genetic heterogeneous disorders)
Congenital hearing loss (> 20 genes)
BRCA1/2 (2 genes) - 112 amplicons; each amplicon has a patients specific ID-label; more patients in a run
Dilative cardymyopathie (40 genes)

Crucial – the interpretation of results – pathophysiological effect of detected, the mutations
Molecular basis of single-gene disorders - OMIM

Single gene disorder affect

- structural proteins
- Cell surface receptors
- Growth regulators
- enzymes

Structural protein and inherited single-gene disorders OMIM

**Hemoglobimopathies**
Sickle cell anemia - gene: **HBB** (hemoglobin beta)
  - type of mutation: missense p.E6V
  - examples of molecular methods: sequencing, RSP-PCR, AD

**Connective tissue disorders**
Marfan syndrome – **FBN1** (fibrillin)
  - missense, deletions
  - sequencing, next-gen sequencing, detection of deletions

**Cell-membran associated protein disfunction**
Muscular dystrophy – **DMD** (dystrophin)
  - deletion
  - detection of deletions

Cystic fibrosis – **CFTR**
  - missense, nonsense, splicing
  - DNA sequencing,