University of Maribor

Faculty of Chemistry and Chemical Engineering

Katja Repnik, Helena Sabina Čelešnik, Staša Jurgec, Uroš Potočnik

BIOCHEMISTRY AND MOLECULAR BIOLOGY

Instructions for laboratory exercises

Maribor, 2017

BIOCHEMISTRY AND MOLECULAR BIOLOGY

Instructions for laboratory exercises

Second-cycle (master's) programme in CHEMISTRY

Proofreader: Dr. Victor Kennedy

CONTENTS

GENERAL DIRECTIONS and SAFETY in LABORATORY

Precautions and laboratory safety procedures

- No eating and drinking in the laboratory
- Wear suitable personal safety equipment (laboratory coat, safety glasses, gloves)
- Tie back long hair
- Clean the working area before starting an experiment
- Read about the basic properties of the chemicals you will be using (toxicity, flammability - bottle label or safety data sheet), before the experiment
- Harmful volatile substances should always be processed in a fume hood (lab ventilation system)
- Liquid waste should be collected separately in an appropriate container
- Instruments/apparatus should be used according to instructions and always under the supervision of technical staff
- In case of an accident, immediately report it to the supervisor
- Wash your hands when work is done

Special safety procedures for the purpose of the exercises

- Read the instructions for the exercise before you begin the laboratory work.
- Always use a laboratory coat and gloves when handling biological material. Use grips and glasses, as well.
- Label your experiments with your name and the date.
- Make precise notes of procedures and results. After the exercise, a report must be submitted.
- Watch the demonstrations carefully.
- Handle the equipment carefully in accordance with the given instructions; no excessive force should be used.
- Before leaving the laboratory, clean the work surface and check that the gas supply has been closed.
- Do not eat or drink in the laboratory. While working, do not put anything in the mouth (pencil, pens, glassware, etc.).
- Use an aseptic technique and disinfect your hands before and after work.
- When working with the burner, make sure that the flame does not burn your skin, hair or coat. Long hair must be tied back.
- During work with microbes or infectious material, the doors and windows must be closed so that the airflow does not carry the microbes through the laboratory.
- During work, ensure that microbial cultures are not poured on the table, on the floor or on the clothes. Nor do we do not touch colonies and suspensions of living microbes with our hands. If the infectious material comes into contact with the skin or the working surface, we inform the assistant or technical staff.

- Cover the contaminated working surface with the absorbent material and pour the disinfectant onto it. Leave the disinfectant for at least 20 minutes. Disinfect any contamination on the body or clothing and rinse with water.
- Contaminated material is disposed of in the disposers or in the designated place and then autoclaved.
- Metal objects that come into contact with the infectious material (tweezers, bacteriological loops, etc.) are disinfected by burning in flame. We also burn the necks of tubes, conical flasks and microbial bottles before they are opened and after they are used.
- The work space must be disinfected before and after work (using disinfectant, 5% sodium hypochlorite, ethanol).
- Residues of biological material should be disposed of in specially designed containers (red bags suitable for autoclaving) and immediately discarded after the work is done in a yellow container.
- The ethidium bromide solution may only be prepared by a qualified researcher or technical assistant. Working with ethidium bromide may only take place in the digester.
- Agarose gels are always transported in a dedicated tub, and the gels are always thrown into a specific yellow container after analysis. The working surface where electrophoresis is carried out and the transilluminator must be cleaned regularly with 75% ethanol.
- When using the UV transilluminator, always use a protective cover.
- With gloves, it is forbidden to touch the handle on the door or surfaces where others are working.
- Dangerous chemicals are stored in a fireproof closet located below the digestor.
- First aid equipment is located in a visible position above the sink.
- In case of extraordinary events (fire, chemical spillage, burglary and other disasters), responsible persons should be contacted immediately. The list of phone numbers is located on the bulletin board.

INSTRUCTIONS for LABORATORY REPORT PREPARATION

Read the instructions for the exercise before you start the laboratory work. During the exercise, take precise notes of procedures and results. Results should be written inside the instructions for each exercise separately. Laboratory report should be completed after each exercise and should be checked before the next exercise.

The most important parts of each exercise are the **RESULTS** and **DISCUSSION**. The results section should include all results obtained during an exercise and all calculations, as well as graphs or sketches. In the Discussion section, you should explain in detail what the results mean and what are the main conclusions, as well as all the limitations or deviations that were observed during the exercise.

All exercises should be checked and signed by the assistant prior to accession to the colloquium.

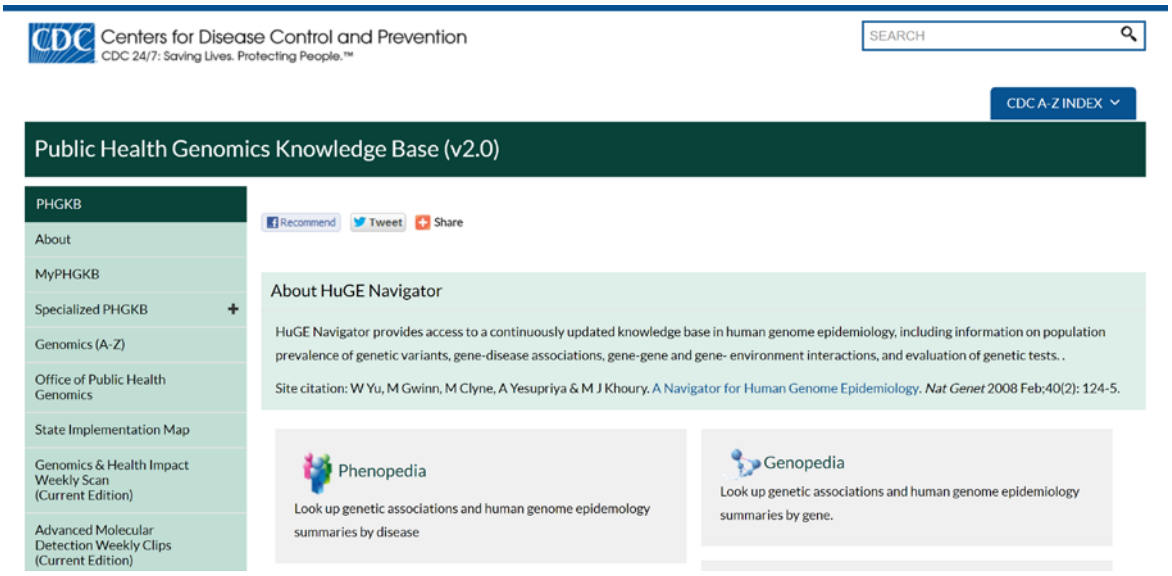
1 DATABASES in BIOCHEMISTRY and MOLECULAR BIOLOGY

Biochemistry and molecular biology are strongly complementary areas that have an important role in understanding the onset and course of diseases. In the context of this exercise, students will become familiar with databases and online web tools for searching data on the genetics of complex diseases. Complex diseases are genetic diseases resulting from an interaction between genetic and environmental factors. The most common genetic markers in the human genome are single nucleotide polymorphisms (SNPs), which are also very important biomarkers of complex diseases, as well as pharmacogenetic markers - that is, genetic markers of response to therapy. When looking for genes or polymorphisms that are associated with complex diseases, association studies are very successful. Association studies seek associations between polymorphisms and disease in cohort studies of unrelated individuals, where we compare allele frequencies between a group of patients and a group of healthy individuals, serving as a control group (case-control studies). Compared to linkage analysis, which detects larger segments of genomic regions (many genes at once), association analyses study only a few candidate genes at once. Significant progress in technology in recent years has enabled the study of a large number of SNPs across an entire genome simultanepisly in genome-wide association studies (GWAs). GWAs constitute a powerful tool for discovering polymorphisms associated with complex diseases. GWAs are based on comparing the frequency of a disease-susceptible allele between the control group and a group of patients, where we genotype thousands of polymorphisms across the entire genome simultaneously in a large group of patients and a large group of healthy individuals. GWA studies were developed as a result of technology advances, especially by microarrays. The majority of SNPs, associated with disease in GWAs are usually in linkage disequilibrium (LD) with parts of the genome that are actually involved in the onset of disease. For more accurate identification of a gene and the SNP associated with disease pathogenesis, association studies of candidate genes are still needed as independent confirmation of GWA discoveries and fine mapping of the causal variant.

When setting up a new genetic study or association analysis, results of previous studies should be taken into consideration. Today, the most important source of information has become articles published in scientific journals and abstracts at conferences. There is a huge number of studies published daily, which is why bioinformatics tools and online databases are needed to filter only those studies that are of interest to us.

In the chapters 1.1. to 1.4, we describe freely available online databases that we can use when looking for SNPs and genes associated with diseases, or we can set up studies, to help us find new candidate genomic loci associated with a particular disease.

1.1 HuGE Navigator



HuGE Navigator (version 2.0) is available at https://phgkb.cdc.gov/PHGKB/hNHome.action and provides access to a continuously updated knowledge base in human genome epidemiology, including information on population prevalence of genetic variants, gene-disease associations, gene-gene and gene-environment interactions, and evaluation of genetic tests. This database allows us to look for associations according to a specific gene (Genopedia) or according to phenotype (disease) (Phenopedia). It also contains other bioinformatics tools, like a literature finder, a gene browser, which sorts, ranks and shows information about genes associated with different phenotypes, risk factors and other genotypes, as well as a tool for mapping common gene variant names to their corresponding rs numbers (Variant Name Mapper). HuGE Navigator is transparent and user-friendly. Using HuGE Navigator, we can find studies that seek correlations between a particular gene and disease, the number of GWA studies and meta-analyses for various diseases or genes, etc.

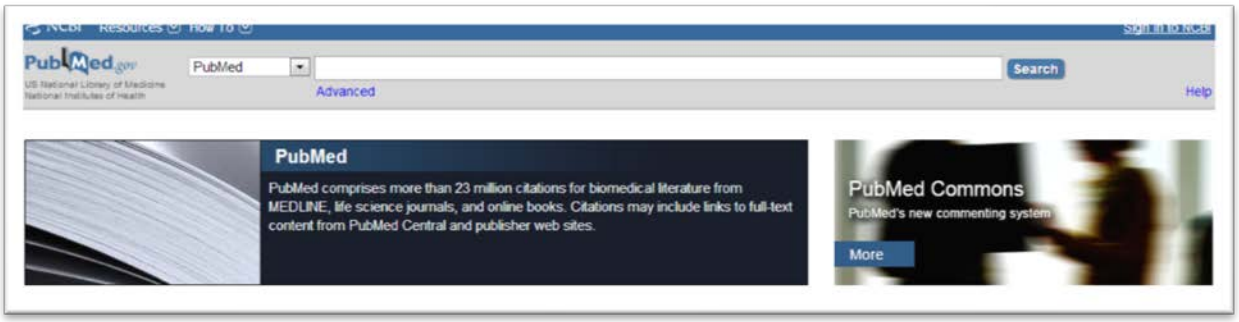
1.2 Ensembl

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotates genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species. Access to Ensembl is through http://www.ensembl.org/index.html.

1.3 National Center for Biotechnology Information (NCBI)

NCBI (National Center for Biotechnology Information) is one of the most important databases for browsing the human genome. NCBI contains several other biotechnological and biomedical databases. The biggest are GenBank, which contains DNA sequences, and PubMed, which is bibliographic database for biomedical litratue. All databases are available online through Entrez.

1.3.1 PubMed



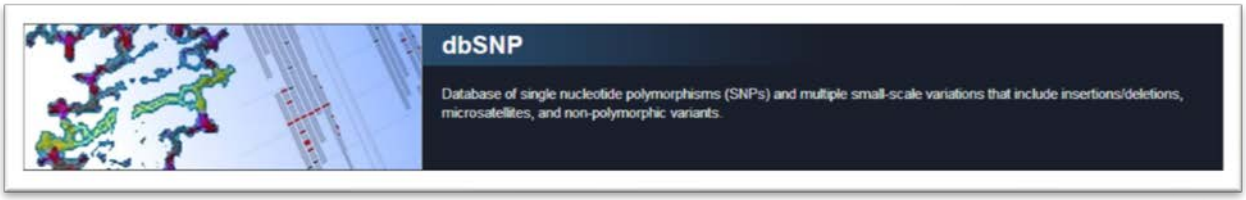
PubMed database (http://www.ncbi.nlm.nih.gov/pubmed) comprises more than 27 million citations for biomedical literature from MEDLINE, life science journals and online books. Citations may include links to full-text content from PubMed Central and publisher web sites. An example of search results from the literature using PubMed is shown in Figure 1.

Figure 1: Search results in PubMed. On the left, articles can be filtered according to type or according to text availability.

After a relevant article has been selected, an abstract is shown and a link to a journal or to a full text article on the right is displayed (Figure 2).

Figure 2: An example of an abstract and a link to a journal/full text.

1.3.2 NCBI dbSNP



Changes in DNA sequence are common in the human genome and are responsible for specific phenotypic characteristics, including the predisposition of an individual to develop complex diseases like cancer or cardiovascular disease. The dbSNP database is in the public domain and comprises a collection of a large number of simple genetic polymorphisms. This collection includes single nucleotide polymorphisms (SNP), deletions, insertions, retroposons or microsatellites. The database contains data on a sequence around the polymorphism of interest, the frequency of the polymorphism in different populations and the location for all variants.

1.3.3 NCBI Gene

The Gene database offers detailed information about known and also predicted genes, which are defined by nucleotide sequence or position. The database contains more than 14 million entries and data from all important taxonomic groups. Each entry corresponds to one gene and contains data on gene nomenclature, reference sequence (RefSeq), location, pathways, variations and phenotype, as well as genotype-phenotype and locus-specific references all over the world.

1.4 HapMap

The HapMap project brought a breakthrough in GWA study development. The goal of this international project was to develop a haplotype map of the human genome, which will help to explain common patterns of gene variation in the human genome and become a key source of information for discovering genes associated with disease, as well as response to treatment and environmental factors.

The HapMap project included genotyping of 1 million SNPs of the human genome in Phase 1 and 3.1 million SNPs in Phase 2. Until June 2016, data on SNP frequencies and their correlations through linkage disequilibrium (LD) were available through the www.hapmap.org web page. Now, access to the best current data (genotypes, sequences and genome mapping) is available through 1000 Genomes Project resources at http://www.1000genomes.org/, or through the NCBI Browser: https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/.

Phases of the HapMap project:

• **Phase 1** – sequence data for 4 populations CEU (Northern Europe), YRI (Yoruba, Nigeria), CHB (China, Beijing), JPT (Tokyo, Japan) - 269 individuals.
#SNP = 1 million, approximately 1 per 5 kb through whole genome, MAF > 0.05

• **Phase 2** – more sequencing in populations from Phase 1
#SNP = 3.1 million and more, approximately 1 kb through whole genome, MAF > 0.05

• **Phase 3** – more sequencing in previous populations and sequencing of 7 more populations
#SNP = 1.6 million and more

HapMap populations:

- ASW***** Descendants of Africans in Northwest USA
- CEU***** Residents of Utah, descendants of northern and western Europeans
- CHB Han Chinese, Beijing, China
- CHD Chinese in Denver, Colorado
- GIH Gujarati Indians in Houston, Texas, USA
- JPT Japanese in Tokyo, Japan
- LWK Luhya in Webuye, Kenya
- MEX***** Mexican ancestry in Los Angeles, California
- MKK***** Maasai in Kinyawa, Kenya
- TSI Tuscans in Italy
- YRI***** Yoruba in Ibadan, Nigeria

** Population are trios (mother, father and child)*

1.4.1 Linkage Disequilibrium (LD)

Linkage disequilibrium (LD) is a type of direct association study for mapping causal genes. The success of LD depends on its understanding and other effects like recombination-associated mutations and population history. Over generations, mutations are separated, owing to recombination from their specific alleles of the original haplotype and are mixed with others, while some DNA variations remain part of the original haplotype through many generations. This kind of combination of alleles that happens by chance is termed linkage disequilibrium (LD) and represents the basis of most association strategies.

To describe linkage disequilibrium, several parameters are used, like Lewontin parameter D' and index r^2. Lewontin D' is also termed association probability and represents an important parameter for identification of genomic regions with a small number of recombination events. If D'=1, we say that a region is in linkage disequilibrium. Index r^2 shows the power of tests for indirect associations and can

vary, although D'=1. The index r^2 value is associated with allele frequencies and with a position of corresponding mutation.

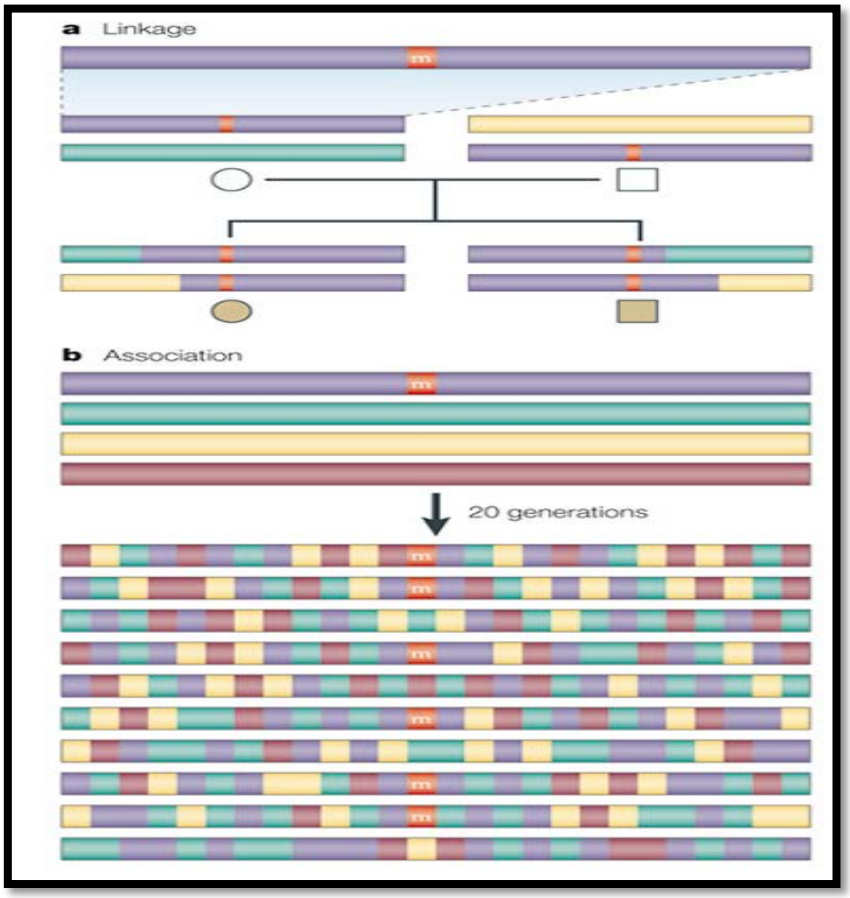


Figure 3: An example of a mutation "m". Mutation "m" separates from the original haplotype over time because of recombination and is mixed with other haplotypes.

The drop in LD depends on chromosome location (regions with low recombination frequencies – centromeres, have high LD).

LD statistics:

- D' can be between 1 and -; what is important is the absolute value of |D'|;
- D' = 1 represents an event **without** recombination;
- Value D' < 1 represents recombination;
- intermediate values for D' are hard to interpret (they are higher in cases of small sample size or in cases of low allele frequency);
- Squared correction coefficient $\mathbf{r^2}$ is also used. We can calculate it by dividing **D'** by allele frequencies at both loci [**D'/(p1p2qp1q2)**];
- If $\mathbf{r^2=1}$, this indicates no recombination and allele frequencies are the same;
- The great advantage of r^2 is the inverse relationship with sample size, which we need in order to find genetic associations between markers;

- r^2 is sensitive to allele frequency and is difficult to interpret in cases when two loci have different allele frequencies.

- ***Allele frequencies can be specific tofor different populations:***
  - ***Allele frequencies of SNP can serve to distinguish between different populations;***
  - ***LD can distinguish between populations;***
  - ***Genotyping results for one population do (not) represent genotypes of any other population.***

Exercise: *Using these databases and tools, find genes associated with lactose intolerance. Find the 4 genes most significantly associated with the disease. Find 2 SNPs that can be used in the diagnosis of lactose intolerance. Where are these two SNPs located? Write the chromosome, gene and part of gene (intron, exon etc.) for these two SNPs and possible alleles for SNPs. Record the frequency of both SNPs in the European population (HapMap CEU).*

1.5 Association analysis

Association analysis compares genotype and allele frequencies between two groups, mainly patients (cases) and healthy individuals (controls) also known as case-control studies. When comparing genotype and allele frequencies, we can use Fisher's exact test, which is available through statistical packages like SPSS.

1.5.1 Genotype and allele frequencies

First, we determine the genotype and allele frequencies for analyzed polymorphism separately for a group of patients and for a control group. Genotype frequencies are a percentage of each genotype according to the total number of genotypes in each group. Allele frequency is calculated according to the following equation:

- assuming a polymorphism A/G – possible genotypes are AA, AG and GG
  - frequency of allele A = (2 x n(AA) + n(AG))/2/n(AA+AG+GG)
  - frequency of allele G = (2 x n(GG) + n(AG))/2/n(AA+AG+GG)
  - (n – number of individuals with determined genotype)

1.5.2 Hardy Weinberg Equilibrium (HWE)

When genotype and allele frequencies for our samples are available, we need to verify if they match with frequencies for other populations, which can be found in other databases. We always have to compare with populations that are in the closest relationship with our population (e. g. HapMap CEU).

Genotype frequencies are not random in nature, but instead are genotypes in Hardy-Weinberg equilibrium, which means that if the frequency of allele A = q, and the frequency of allele G = p, then the frequency of genotype AA = p^2, genotype AG = 2pq and genotype GG = q^2.

From the calculated allele frequencies, we calculate p^2, q^2 and 2pq. Then, p^2, q^2 and 2pq values should be multiplied by the number of all genotypes available for the analyzed group. Those values should be compared with actual values of genotyping, and this should be statistically analysed, and the p-value should be near 1 (between 0.5 and 1). Hardy-Weinberg calculators are freely available online, e.g. http://quantpsy.org/chisq/chisq.htm.

Exercise: *When genotyping SNP rs1050152, we get following genotypes:*

CC – N = 32

CT – N = 66

TT – N = 39

Calculate the genotype and allele frequencies. Are the genotype frequencies in Hardy-Weinberg equilibrium?

2 DNA EXTRACTION FROM BLOOD SAMPLES

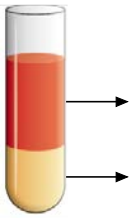
2.1 THEORETICAL BASICS

For any DNA analysis (such as sequencing or amplification of a DNA fragment), we first need to isolate it from the sample. In medical research the most commonly isolated DNA is from peripheral blood, more precisely from lymphocytes. In the isolation process, the cells must first be broken to precipitate the DNA. In the next step, the DNA needs to be cleaned of impurities and finally dissolved for further use.

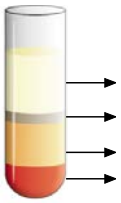
2.2 EXERCISE PROTOCOL

Collection of lymphocytes

- in a 1.5 mL tube transfer 650 µL of blood from an EDTA tube
- add 650 µL of PBS buffer
- mix
- in a fresh sterile 2 mL tube put 650 µL of Lympholyte®
- carefully layer 1300 µL of the mixed blood sample on top of the Lympholyte® (when layering the sample, do not mix the Lympholyte® and the diluted blood sample)



- centrifuge at 1800 rpm for 30 min at 18°C
- remove the upper phase (serum and thrombocytes)
- transfer the second phase (lymphocytes) with a pipette to a new sterile 1.5 mL tube



- add 1 mL of PBS buffer
- centrifuge at 1400 rpm for 15 min at 18°C
- remove supernatant
- add 650 µL of PBS buffer
- centrifuge at 1400 rpm for 15 min at 18°C

- remove all supernatant (use a pipette)

DNA isolation using TRI Reagent®

- add 260 µL of TRI Reagent and lyse cells (pellet of lymphocytes) with pipetting
- incubate for 5 min at room temperature
- add 52 µL of chloroform
- close the tube tightly and vigorously shake it for 15 seconds
- incubate for 2 – 15 min at room temperature
- centrifuge at 12.000 x g (rcf) for 15 min at 4°C
- centrifugation separates the mixture into 3 phases:
    - lower red organic phase – proteins
    - white interphase – DNA
    - colorless upper aqueous phase – RNA
- remove the aqueous phase (for RNA isolation)
- to precipitate the DNA from the interphase and the organic phase, add 100 mL of 100% ethanol
- mix by inversion and allow to stand for 2 – 3 min at room temperature
- centrifuge at 2.000 x g for 5 min at 4°C
- remove the supernatant (for protein isolation)
- wash the DNA pellet twice in 260 µL 0.1 M sodium citrate/10% ethanol solution
- during each wash, allow the DNA pellet to stand in this solution for 10 min (gently mix a few times)
- centrifuge at 2.000 x g for 5 min at 4°C
- remove the supernatant and resuspend the DNA pellet in 430 µL of 75% ethanol
- allow to stand for 10 – 20 min at room temperature
- centrifuge at 2.000 x g for 5 min at 4°C
- remove the supernatant and dry the DNA pellet in the air
- dissolve the DNA pellet in 20 µL H_2O by repeated pipetting

2.3 RESULTS and DISCUSSION

3 LACTOSE INTOLERANCE: ANALYSIS OF POLYMORPHISM IN THE *LCT* GENE USING THE PCR-RFLP TECHNIQUE

3.1 THEORETICAL BASICS

3.1.1 Lactose intolerance

Lactose intolerance is an autosomal recessive metabolic disorder, which affects a large part of the adult population around the world. Lactose intolerance is an inability to digest lactose, the main sugar in milk.

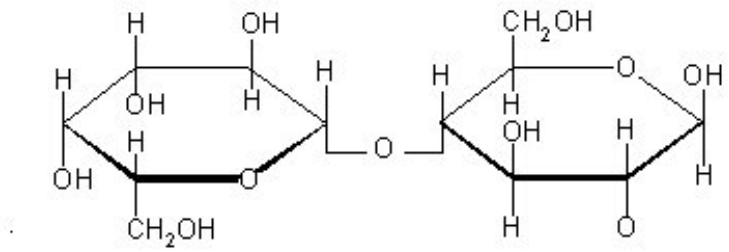


Figure 4: Lactose – disaccharide composed of glucose and galactose

Lactose intolerance is due to a lack of the enzyme lactase (full name: lactase – phlorizin hydrolase or LPH) in the small intestines to break lactose down into glucose and galactose. People lacking that enzyme are unable to fully digest lactose in milk, which causes symptoms like indigestion, diarrhea, sweating, rashes and abdominal pain. Lactose that remains undigested cannot be directly absorbed through the wall of the small intestine into the bloodstream. In the small intestine, lactose absorbs in the gut a large amount of water. Bacteria in the colon metabolise the lactose, and the resulting fermentation produces large amounts of gas (a mixture of hydrogen, carbon dioxide and methane), which causes the various abdominal symptoms. Some hydrogen is passed through the blood to the lungs, where it can be detected in the air breathed out. Because of the unpleasant symptoms, people with lactose intolerance give up milk and dairy products, which can lead to other problems, e.g. lack of calcium, magnesium and vitamin D.

3.1.2 Genetics of lactose intolerance

LCT is the gene encoding for the enzyme lactase (LPH) and is located on chromosome 2. Lactose intolerance in adulthood results in reduced expression of this gene and occurs in a significant proportion of the population.

The *LCT* expression controls the part of the DNA-regulatory element - located approximately 14 kbp upstream (in the direction of 5') from the ATG initiation codon of this gene. The *LCT* gene provides the instructions for making lactase. The *LCT* regulatory site lies in the intron of the *MCM6* gene, the product

of which is a cell cycle regulator and is not functionally related to lactose. The specific DNA sequence in the MCM6 gene helps control whether the *LCT* gene is turned on or off.

In 2002, polymorphism was discovered in the regulatory element for the *LCT* gene, the intron of the gene MCM6, which is 100% associated with the development of adult lactose intolerance. For polymorphism, two alleles – C or T are possible. Allele C is linked to low and allele T to high expression of the *LCT* gene, consequently with the amount of the LPH enzyme. Alelle T is dominant, which means that lactose intolerance can only be expected in homozygotes for allele C (individuals with the CC genotype, but not for those with CT or TT genotypes).

Genotyping is the most reliable way of diagnosing lactose intolerance, but it can only be used as a screening test. People who have the CT or TT genotype cannot develop lactose intolerance, and if they have the symptoms described, we need to look for causes somewhere else. People with the CC genotype have a risk of lactose intolerance, which may occur either in puberty, in adulthood, at a late age or never.

3.1.3 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is a technique to amplify a single or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.
The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers (DNA fragments around 20 bp long) containing sequences complementary to the target region, along with a DNA polymerase (after which the method is named) are the key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. Almost all PCR applications employ a heat-stable DNA polymerase, such as Taq polymerase, an enzyme originally isolated from the bacterium Thermus aquaticus. This DNA polymerase enzymatically assembles a new DNA strand from DNA building blocks, the nucleotides, by using single-stranded DNA as a template and DNA oligonucleotides (also called DNA primers), which are required for initiation of DNA synthesis. Thermal cycling steps are necessary first to physically separate the two strands in a DNA double helix at a high temperature in a process called DNA melting. At a lower temperature, each strand is then used as the template in DNA synthesis by the DNA polymerase to selectively amplify the target DNA. The PCR usually consists of a series of 35-45 repeated temperature changes called cycles.

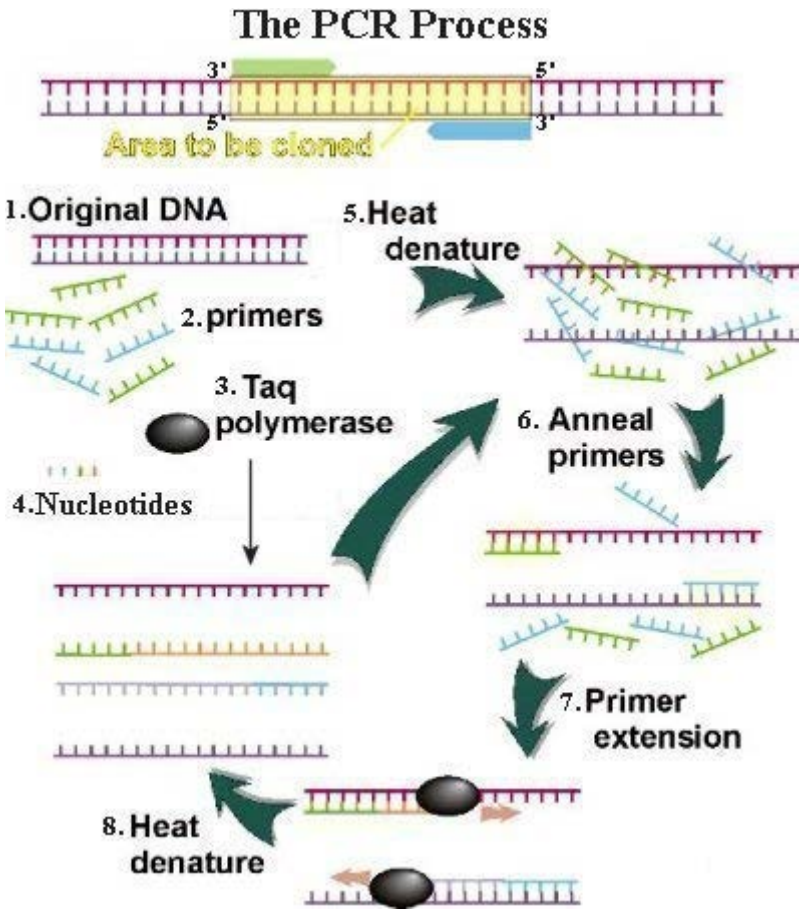


Figure 5: PCR reaction

Denaturation step: This step consists of heating the reaction to 92°C or more. It causes DNA melting of the DNA template by disrupting the hydrogen bonds between complementary bases, yielding single strands of DNA.

Annealing step: The reaction temperature is lowered to 50–65°C allowing annealing of the primers to the single-stranded DNA template. The polymerase binds to the primer-template hybrid and begins DNA synthesis.

Extension/elongation step: At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in 5' to 3' direction. Optimal temperature for Taq polymerase is at 72°C.

Final elongation: This single step is occasionally performed at a temperature of 70–74°C for 5–15 minutes after the last PCR cycle to ensure that any remaining single-stranded DNA is fully extended.

Final hold: This step at 4°C for an indefinite time may be employed for short-term storage of the reaction.

Since in each cycle the number of molecules is doubled, the number of molecules of the desired section of the DNA can be calculated using the following formula:

$N=N_0*2^n$

(N=number of molecules, N_0= initial number of molecules, n= nuber of cycles)

3.1.4 Calculating the quantity of reagent needed for a successful PCR reaction

For the PCR reaction, the following reagents are needed:

- DNA template
- Deoxynucleoside triphosphates (dNTPs)
- $MgCl_2$ (some buffers already contain it)
- Buffer solution
- Primer 1
- Primer 2
- Taq polymerase

For a successful PCR reaction, we must choose the optimal amounts of each reagent. The amounts of the necessary template DNA, dNTP, buffer (with $MgCl_2$) and the Taq polymerase are the same for each PCR reaction, while the optimum amount of both primers should be determined experimentally.

The amount of the particular reagent required in the reaction is expressed by concentration - the amount of substance in the volume unit.

- Mass concentration: g/L (or g/mL, µg/µL, ng/µL...)
  - DNA
- Molar concentration: mol/L = M (mmol/L = mM, mmol/ml = M...)
  - dNTP, $MgCl_2$, primers
- Concentration of enzymes: U/L (U/µL...)

DNA concentration

The desired DNA concentration depends on the type of further analysis. A concentration of 50 ng/µL of the matrix DNA is required for the PCR reaction.

After isolating the DNA, we measure the concentration by the spectrophotometric method by measuring the absorbance at a wavelength of 260 nm. Depending on the measured concentration, we dilute the DNA to the desired final concentration, most often with water or a suitable buffer. When measuring the concentration by the spectrophotometric method, we acquire information about the purity of the isolated DNA (ratio A_{260} / A_{280}), where A_{260} represents the wavelength where DNA absorbs, and A_{280} represents the wavelength where proteins absorbs. We have pure DNA if that ratio is between 1.8 and 2.

When diluting the DNA to the desired concentration, we use the following formula:

$c*V = c_1*V_1 + c_2*V_2$ $\quad\quad$ $V = V_1 + V_2$

c... final concentration (25ng/µL)
V... final volume

c_1... concentration of isolated DNA (initial concentration)
V_1... volume of the isolated DNA solution (which must be added to obtain the desired final volume with the desired concentration)
c_2... final desired DNA concentration
V_2... volume of water (which must be added to obtain the desired final volume with the desired concentration)

Buffer, dNTP and enzyme polymerase

The buffer in the reaction mixture maintains the appropriate pH and concentration of the salt in which the enzyme polymerase works optimally. To some buffers, $MgCl_2$ is added.

dNTPs must be present to build a new DNA chain. A sufficient concentration of dNTPs for the reaction is 200 µM for each individual nucleotide (ATP, CTP, GTP, TTP). To make work easier, a concentrated mixture of all four dNTPs is prepared and used for the PCR reaction mixture.

The enzyme concentration is always given in units (U) per unit of volume. The amount of polymerase that is sufficient for a successful PCR reaction with the final volume of the reaction mixture of 10 µL is 0.5 U (or 0.05U/µL) and is accordingly higher if we work with a larger volume.

Optimization of temperature and quantity of oligonucleotides (primers)

We begin the optimization of PCR by searching for the most appropriate temperature for primer annealing. In this case, we put a maximum amount of primers in the reaction mixture and run the reaction at different annealing temperatures with the same DNA template.
The optimum concentration of the primers is also determined experimentally. The PCR reaction is more effective if the concentration is higher, but an excessive amount of the primers leads to non-specific bindings, leading to the formation of non-specific products and dimers.

3.1.5 Gel electrophoresis

For gel electrophoresis, we need an agarose or polyacrylamide gel and buffer that provides stable pH and electrical conductivity. With agarose gel electrophoresis, we separate the DNA fragments by length, while the DNA (negatively charged) travels toward the positive electrode. Longer DNA molecules travel more slowly along the gel than shorter ones.
Usually, we add to the gel a size marker to estimate the size of our PCR products.

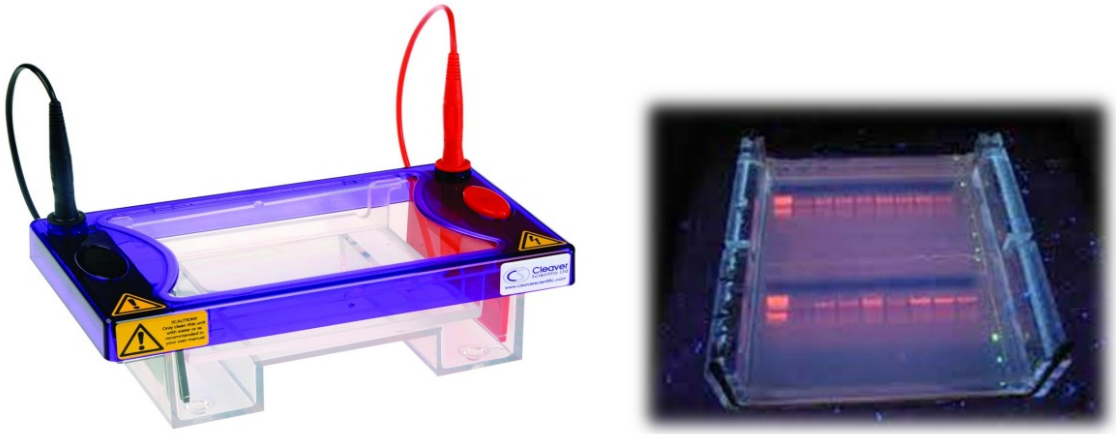


Figure 6: Detection of DNA on agarose gel

3.1.6 Restriction fragment length polymorphism (RFLP)

When we want to analyze mononucleotide polymorphisms (SNPs), the PCR-RFLP technique (polymerase chain reaction - restriction fragment length polymorphism) is often used. The desired DNA section, which includes the SNP that is being investigated, is first amplified by the PCR reaction, and then the restriction enzyme is added to the product, which specifically cuts the product in the case of one allele, and does not cut the product in the case of the second allele.

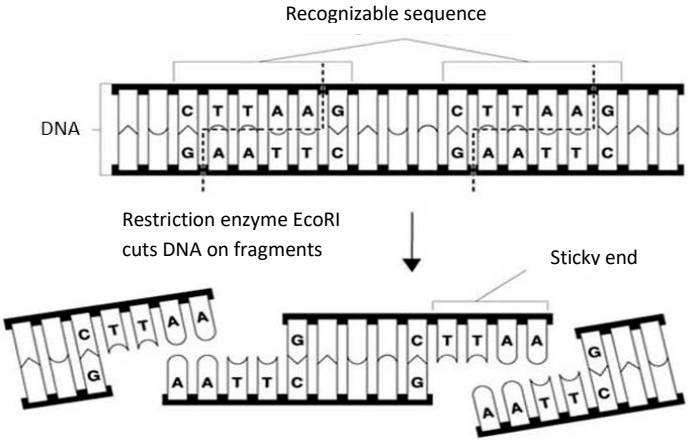


Figure 7: Schematic presentation of the EcoRI enzyme activity

The natural role of restriction enzymes is primarily the identification and removal (cutting) of foreign DNA. These enzymes recognize the characteristic sequence, and at the site of this sequence, the molecule is cut. Using bioinformatics tools, we select an enzyme that cuts the product only in the case of one allele or that cuts them differently.
These products are then separated by gel electrophoresis and, based on the size of the fragments, the genotype is determined.

For example: With PCR we multiply a 243 bp long sequence around polymorphism rs4988235, responsible for the regulation of the *LCT* gene.

The DNA sequence around the SNP rs4988235 (obtained from the NCBI SNP database) is:

ATAAGACTCCTTAAGTTCTCTACAATGTACGACCCGGAAAGGTTGCACAACTTAGACCCTACAATGTACTAGTA
GGCCTCTGCGCTGGCAATACAGATAAGATAATGTAG**[C/T]**CCCTGGCCTCAAAGGAACTCTCCTCCTTAGGTTG
CATTTGTATAATGTTTGATTTTTAGAACAAAGGTTGCATGTACTAGTTCTCCTCCTTAAGTTCTCGTTCTCCTCCA
ATAAGAATAAGATAATGTT

Encim Faq I recognizes the **TAGTCC** sequence and cuts the molecule into two parts. Since this sequence is found only in those products having allele T at the site of polymorphism rs4988235, we get after incubation with the enzyme, products of different lengths.

Allele C: CAATACAGATAAGATAATGTAG**[C]**CCCTGGCCTCAAAGGAACTCTCC

↓

Allele T: CAATACAGATAAGATAATGTAG**[T]**CCCTGGCCTCAAAGGAACTCTCC

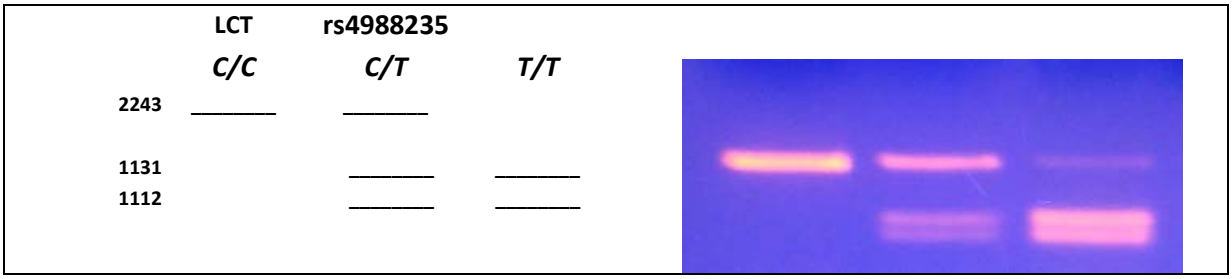


Figure 8: Detection of products by the PCR-RFLP technique for rs4988235

3.2 EXERCISE PROTOCOL

MATERIALS

Template DNA from exercise 2, reagents for the PCR reaction (dNTP, buffer, pair of primers, Taq polymerase), pipettes, 8-tube strip, thermocycler, electrophoresis system, reagents for restriction (enzyme Faq I, SAM, buffer).

PROTOCOL

- In the tubes from the strip, put 2 µL of the folowing: template DNA (1 - 12; if you have 12 samples of DNA), all three positive controls (13 – 15; DNA with known genotype) and H_2O for negative control (16).

Tube in strip 1	1	2	3	4	5	6	7	8
Sample	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**

Tube in strip 2	9	10	11	12	13	14	15	16
Sample	**V9**	**V10**	**V11**	**V12**	**CC**	**CT**	**TT**	**H_2O**

- Defrost the reagents (except the polymerase), mix on vortex, short spin on centrifuge and put on ice. The polymerase must only be centrifuged and keep it on ice the whole time.

- Prepare the reaction mixture with the following composition:

Reagent	**Quantity for 1 sample (µL)**	**Quantity for 20 samples (µL)**
H_2O	12.5	
dNTP	0.4	
buffer	2	
Primer 1	2	
Primer 2	2	
Taq polymerase	0.1	
Total volume:	18	
Total volume of reaction:	20	/

- Add 18 µL of reaction mixture to each sample (1 – 16); put the strip on the vortex and short spin it on the centrifuge.

- On the thermocycler, set the number of cycles to 35 and the annealing temperature to 63°C.

- During the PCR reaction, prepare the 2% agarose gel for product detection.
  - Weigh 1 g of agarose and dissolve it in 50 mL of 1X TBE buffer by heating in a microwave oven.
  - To the hot gel, add 2.5 µL of ethidium bromide, gently mix and pour it into model.

- After the PCR reaction, use gel electrophoresis to check whether the reaction was successful – take 10 µL of PCR product and place it on gel; store another 10 µL for restriction.

- Mix the sample with xylencyanol dye on parafilm and place it a pocket on the gel. Run the electrophoresis for 150 min at 160 V.
- Detect products under UV light.

- Defrost the reagents for restriction (except the enzyme), mix on the vortex, centrifuge and put on ice. The enzyme is only centrifuged and kept it on ice the whole time.

- Prepare the restriction mixture with the following composition:

Reagent	**Quantity for 1 sample (µL)**	**Quantity for 17 samples (µL)**
H_2O	17.27	
buffer	2	
SAM	0.6	
Enzym: Faq I	0.13	
Total volume:	20	
Total volume of reaction:	30	/

- Agitate the restriction mixture on the vortex, centrifuge briefly and add 18 µL of it to each sample (PCR product). Mix the strip on the vortex and place it in the centrifuge for a few seconds.

- Incubate samples with added enzyme overnight at 37°C.

- After incubation, detect the products on the agarose gel and read the genotypes of the individual samples.

- Calculate the frequencies (percentages) of individual genotypes by taking into account the results for all groups.

3.3 RESULTS

Your group:

Sample	V1	V2	V3	V4	V5	V6	V7	V8
Result								

Sample	V9	V10	V11	V12	CC	CT	TT	H_2O
Result								

Number of successfully genotyped samples: _____

Genotype	Number of samples	Percentage
CC		
CT		
TT		

3.4 DISCUSSION

4 RHEUMATOID ARTHRITIS: GENOTYPING of SNP IN THE *PTPN22* GENE USING HIGH RESOLUTION MELTING CURVE ANALYSIS (HRM)

4.1 THEORETICAL BASICS

4.1.1 Rheumatoid Arthritis

Rheumatoid arthritis (RA) is chronic systemic, inflammatory and heterogeneous disease, which affects approximately 1% of the adult population in Slovenia and is more common in women. RA primarily affects joints/synovial membranes, which results in progressive decay of joint cartilage. In addition to joints, RA also affects other body parts, like the eyes lungs and skin.

Figure 9: Rheumatoid arthritis

Autoimmunity is an aberrant immune response of an organism against its own healthy cells and tissues. Several diseases are a result of autoimmunity and are called autoimmune diseases (e.g. systemic lupus erythematosus, multiple sclerosis, rheumatoid arthritis, inflammatory bowel disease, etc.).
The causes of rheumatoid arthritis are still unknown, but we do know that autoimmunity plays an important role in disease progression. The causes for the immune system turning against itself are also unknown, but we do know that this is because of a combination of genetic and environmental factors.

4.1.2 Rheumatoid arthritis genetics and the *PTPN22* gene

In recent years, it has been found that there are several different genes contributing to the development of RA, and these are located at 30 different genome regions. One of the most important genes is *PTPN22*, which is located at chromosome 1 and the code for protein tyrosin phosphatase nonreceptor type 22. Together with HLA-DRB1, *PTPN22* represents 40% of the genetic risk for RA development.

Diseases for which *PTPN22* is a promising candidate gene all have autoimmunity as an important feature. The genetic variant most significantly associated with RA development is nonsynonymous missense single nucleotide polymorphism (SNP) rs2476601, which changes the genetic code so that the amino acid composition at position 620 changes from arginine (R) to tryptophan (W) (R620W).
It is not known how alleles affect the disease; however, it has been speculated that a protein with a disease-susceptible variant decreases the inhibition of immune system T cells, which leads to autoimmunity.
In the case of variant 620W in protein or allele A on DNA, there is an increased risk for RA development. It has been shown that individuals with genotype AA (homozygote individuals) of SNP rs2476601 have a four times higher risk of developing RA compared to the general population. Heterozygote individuals have a risk less than two times higher, while for wild type individuals, who are homozygote for allele G, the risk is even lower.

Genotype	Genotype frequency	Rheumatoid arthritis risk
AA	1.4%	4.13%
AG	20.7%	1.75%
GG	77.9%	0.75%

4.1.3 High Resolution Melting (HRM) Analysis

In addition to gene expression analysis, real time quantitative PCR (qPCR) also enables genotyping of DNA samples. One application for genotyping is high resolution melting (HRM) analysis.

Similar to classic end point PCR and quantitative real time PCR (qPCR), with HRM analysis we also amplify DNA fragments of interest (the part of the DNA with polymorphism or a mutation of interest). The reaction mixture of primers, HRM master mix (which includes HRM dye, dNTPs, buffer, $MgCl_2$ and enzyme polymerase) and the template DNA, are heated to 95°C (denaturation), cooled to 60°C (annealing) and again heated to 72°C (elongation) in several cycles in a thermal cycler. Formation of the product is detected in real time based on a fluorescence signal emitted by the HRM dye, which is bound to the DNA during elongation and when bound, emits a signal.

After approximately 40 cycles of amplification in the PCR reaction, another step is added for HRM analysis. The reaction mixture is slowly heated from 50 to 95°C. At some point or at a specific temperature, the melting temperature of the amplicon is reached, and the two strands of DNA separate or “melt” apart – hydrogen bonds between nucleotides are separated, and double-stranded DNA is separated into single-stranded DNA. The HRM dye binds only to double-stranded DNA, which is why denaturation of DNA can be detected based on the fluorescence signal. When half the molecules have been separated, the fluorescence signal decreases by half. The temperature at which this happens is called the melting temperature T_m (Figure 10).

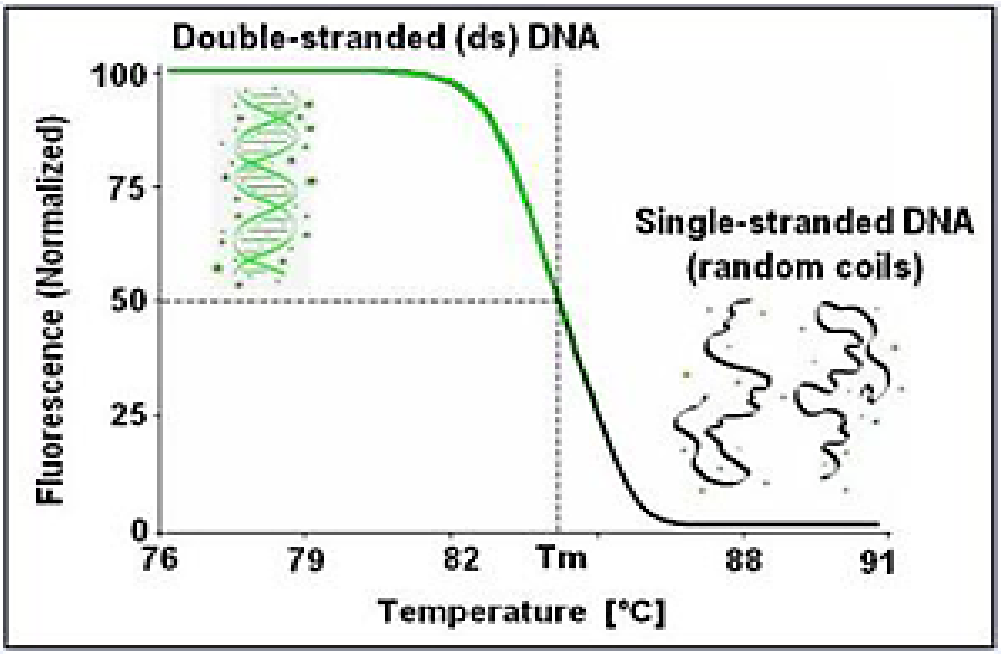


Figure 10: Melting of double-stranded DNA in HRMA.

Since mononucleotides are bound with a different number of hydrogen bonds (adenine and thymine – double bond, cytosine and guanine – triple bond), the energy needed to separate amplicons with different nucleotides is different. If the amplicon after a PCR reaction resembles only one base – e.g. if we have a SNP with two possible alleles, A and C, the molecules that contain allele A and thus form a double bond will melt at a lower temperature compared to that for the molecules that contain allele C and form a triple bond. Since each person, for each SNP, has 2 possible alleles, three different genotypes are possible – AA, AC and CC, so we get three curves of different shape in HRM analysis (Figure 11).

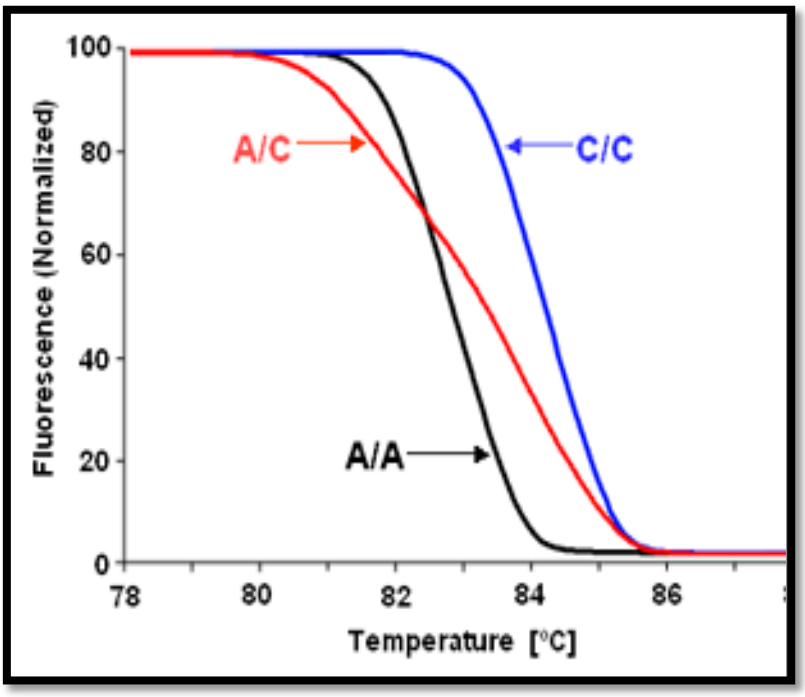


Figure 11: Different melting curves for different genotypes.

Homozygote AA and CC genotypes resemble Tm, while heterozygote genotypes resemble more the shape of a curve, because the drop in fluorescence happens twice. First, half the molecules (those with allele A) are separated, and later the other half of the molecules (those with allele C) are separated. If the y-axis represents the derivative of fluorescence, we get a graph as shown on Figure 12.

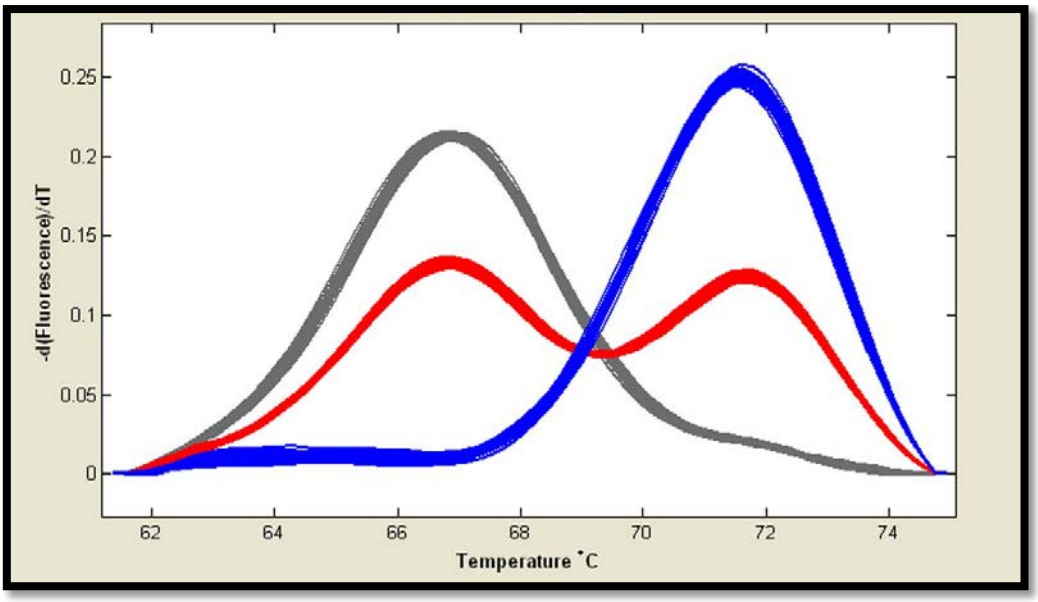


Figure 12: Derivative of fluorescence in HRM analysis.

4.2 EXERCISE PROTOCOL

MATERIALS

Template DNA, HRM-PCR reagents (HRM master mix, primers), pipettes, PCR plate, qPCR instrument

PROTOCOL

Using HRM, we will genotype 20 different samples for SNP rs2476601 in the *PTNP22* gene.

- Put 2 µL of samples (DNA) into a PCR plate in the order shown in the table below. Add three control samples with known genotype and no template control (water).

	1	2	3	4	5
A	V1	V2	V3	V4	V5
B	V6	V7	V8	V9	V10
C	V11	V12	V13	V14	V15
D	V16	V17	V18	V19	V20
E	AA	AG	GG	H_2O	H_2O

- Prepare the reaction mixture according to the table below:

Reagent	**Amount for 1 sample [µL]**	**Amount for 27 sample [µL]**
DNA	2	/
PCR grade H_2O	1	27
HRM master mix	5	135
Primer 1	1	27
Primer 2	1	27
Total volume:	10	/

- To each DNA sample on the plate, add 8 µL of the reaction mixture and perform the PCR reaction according to template temperature protocol on the qPCR instrument. Perform the reaction in 40 cycles.

- After the reaction is completed, determine the genotype for each sample from the melting curves.

- Which sample has a higher risk of developing rheumatoid arthritis? Why?

- Calculate the frequency of risk genotype and the frequency of heterozygotes for analyzed SNP.

4.3 RESULTS

Samples genotypes:

	1	2	3	4	5
A					
B					
C					
D					
E	*AA*	*AG*	*GG*	/	/

Number and percentage of samples with genotype AA: _____ (_____ %)
Number and percentage of samples with genotype AG: _____ (_____ %)
Number and percentage of samples with genotype GG: _____ (_____ %)
Frequency of allele A: _____ %
Frequency of allele G: _____ %

4.4 DISCUSSION

5 TUMOR NECROSIS FACTOR (TNF): MEASURING *TNFα* GENE EXPRESSION BY REAL-TIME QUANTITATIVE PCR (qPCR)

5.1 THEORETICAL BASICS

5.1.1 Immune system and cytokines

The immune system is a network of specialized cells, organs and processes that safeguard the body and protect it against disease. To function properly, the immune system must recognize many different agents, such as viruses, bacteria and other pathogens, and distinguish them from the organism's own cells and healthy tissues.

The immune system can be classified into two subsystems. Microorganisms or toxins that successfully enter the organism first encounter protective cells and mechanisms of the innate immune system. This primary response is non-specific and represents a systemic response against pathogens. On the other hand, the secondary or adaptive immune system recognizes specific antigens and mounts a stronger and more specific immune response.

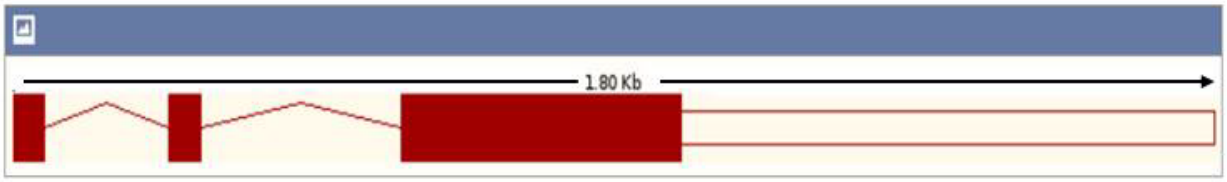
Cells of the immune system communicate with each other via small protein molecules called cytokines.

Cytokines are a group of small regulatory proteins secreted by most cell types, particularly lymphocytes (Th cells) and macrophages. They function as intercellular messengers that elicit biological activities in their target cells by binding to cell receptors. They participate in numerous physiological responses, for example humoral and cell-mediated immune responses, inflammation, haemopoesis, cell proliferation and differentiation and wound healing.

By binding to specific receptors on target cell membranes, cytokines trigger an intracellular signal that results in altered gene expression. Owing to high binding affinity, biological effects are achieved by picomolar concentrations of cytokines. Most commonly, they function in a paracrine manner (by affecting the cells near the secreting cell), but they can also work in an autocrine way (by binding to the surface receptors of the same cell that secretes them), or an endocrine way (by binding to cells in a distant location in the body). Correct and adequate cytokine production is therefore essential for proper functioning of the immune system. Abnormal functioning of the immune system can lead to autoimmune diseases, chronic inflammation and cancer.

5.1.2 Tumor necrosis factor: a cytokine of primary immune response

Cytokines of primary immune system are secreted mostly by mononuclear phagocytes as a response to infection. Secretion can be stimulated by bacterial products (lipopolysaccharides) or viruses (double-stranded RNA). Most cytokines of this type function in endothelial cells and leukocytes by stimulating early inflammatory responses to pathogens. TNF participates in numerous biological processes, such as proliferation, differentiation, apoptosis, lipid metabolism and coagulation, and is a key mediator of acute inflammatory response, responsible for many systemic complications linked to severe infections. Small concentrations stimulate expression of surface adhesion molecules (cell membrane molecules important for connections to extracellular matrix components) on vascular endothelial cells (inner layer of blood vessels). Endothelial cells become adhesive, causing leukocyte recruitment and accumulation. Furthermore, TNF stimulates other immune cells (especially mononuclear phagocytes) to secrete inflammatory cytokines (interleukin-1 (IL-1), interleukin-6 (IL-6), TNF and chemokines).
When larger amounts of TNF are secreted, it enters the bloodstream where it functions as an endocrine hormone (as a pyrogen, which induces fever). In liver cells (hepatocytes), it increases secretion of some acute-phase proteins, and in bone marrow it inhibits division of stem cells.
The TNF gene is located on chromosome 6 and is 2770 bp long. It spans three exons and encodes a 171-amino acid long protein.



Exons: 3 Transcript length: 1315 bp Protein: 171 amino acid residues

Figure 13: Schematic representation of the TNF gene.

5.1.3 TNF and disease

TNF is linked to a number of diseases, including autoimmune diseases, insulin resistance and cancer. Chronic exposure to TNF can result in certain conditions, such as lymphopenia and immune deficiency. Mouse knockout studies have also demonstrated that TNF affects the nervous system.
Association studies that compared TNF-gene polymorphisms among different groups of patients and healthy individuals found connections with asthma, rheumatoid arthritis, type 1 and 2 diabetes, Crohn's disease and other diseases. Polymorphisms located in the promoter region (which regulates gene expression) appear to have the most pronounced consequences. Consequently, increased TNF expression resulting from promoter polymorphism has been observed in all these diseases. Increased expression leads to increased cytokine concentration, which in turn contributes to disease development. Gene expression (measured as mRNA levels) can be determined by real-time quantitative PCR (qPCR).

5.1.4 Real-time quantitative PCR

When performing traditional PCR, amplification products are usually detected on agarose gel after the final PCR cycle. On the other hand, real-time quantitative PCR (also known as real-time PCR) allows the PCR products to be detected during the cycles and the amplification reactions to be monitored in real time. For detection, fluorescent molecules are included in amplification reactions. The intensity of fluorescence directly correlates with the amount of the amplified product. Fluorescent dyes or fluorescent probes (that bind specifically to complementary DNA) can be used. The fluorescent signal is detected by the PCR apparatus (i.e. the thermal cycler instrument), which is linked to a computer. Special software enables florescence signal monitoring after each cycle and subsequent data analysis. Real-time PCR can be used for qualitative measurements (to determine the presence or absence of individual sequences) or quantitative measurements (to determine DNA copy number). When the method is used for quantitative analyses, it is also called quantitative PCR (qPCR).
In contrast to traditional PCR, product detection does not require visualization on a gel. Rather, the system automatically delivers results, thereby shortening the total processing time and reducing the chance of contamination.
The reaction is graphically represented in Figure 14. The x-axis shows the cycle number and the y-axis shows the fluorescence emitted by the amplified product (the fluorescence intensity is directly correlated with the product amount). At low cycle numbers, amplification is exponential (product amount doubles in each cycle), then it reaches a plateau (non-exponential phase).

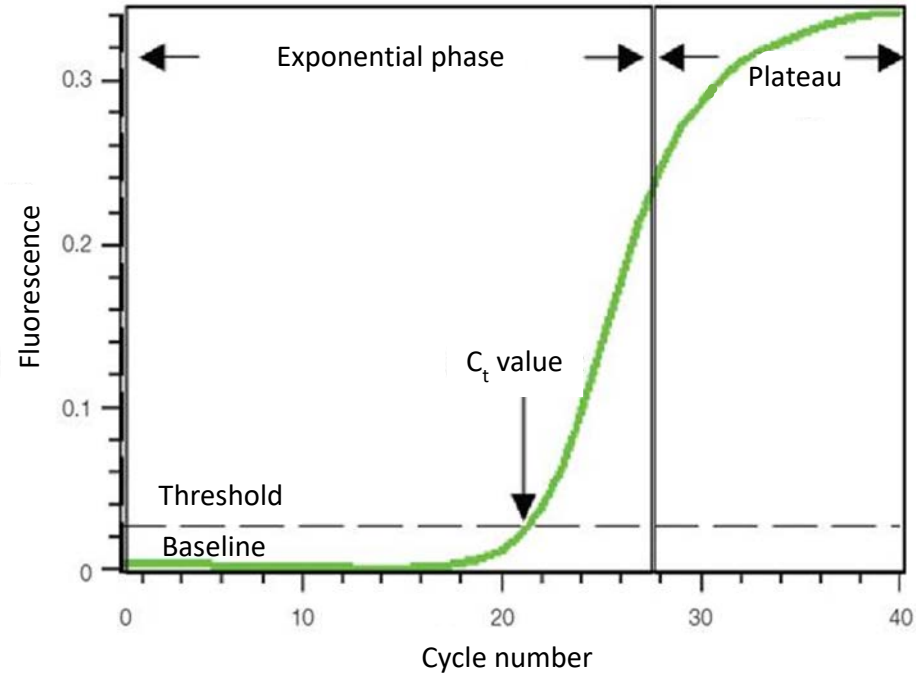


Figure 14: Product amplification in RT-PCR.

During the initial cycles the product amount is very low despite exponential amplification, and fluorescence cannot be measured above the background. The PCR cycle number at which the reporter fluorescence is greater than the detection threshold is called the threshold cycle (Ct). The Ct value depends on the initial DNA amount; higher initial DNA amounts will require fewer amplification cycles to reach the threshold of detection → low Ct value. Conversely, amplification of low amounts of DNA

will require more cycles to reach the same signal intensity → high Ct value. This is the basic principle of RT-PCR.

Theoretically, all samples with the same signal intensity will contain the same number of product copies. The threshold line is the level of fluorescence at which the Ct value is recorded → all samples at a particular Ct cycle have an equal number of specific copies.

The graph can also be represented by plotting the log of relative fluorescence against the cycle number, as depicted in Figure 15.

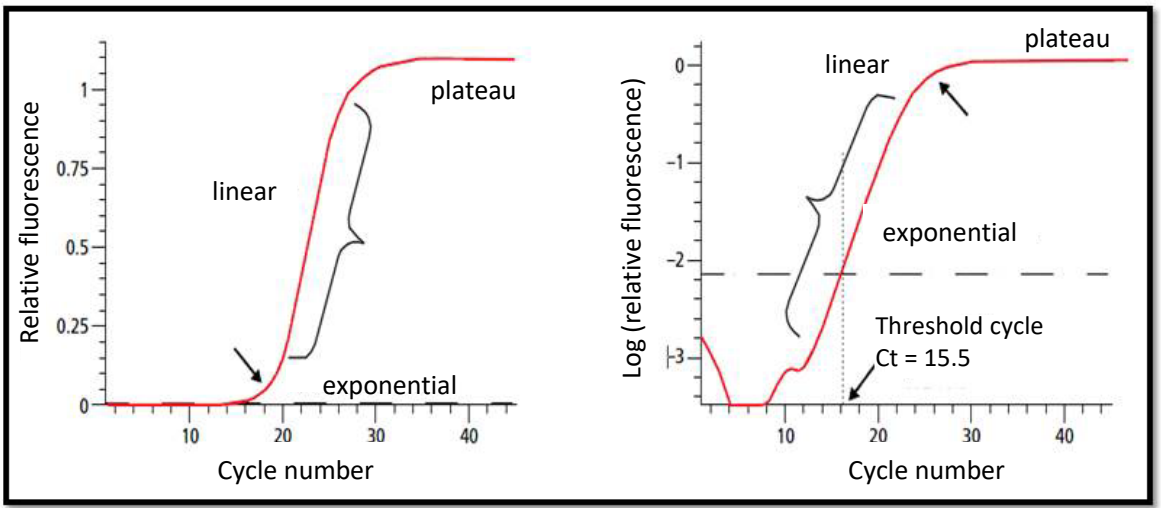


Figure 15: Product amplification in RT-PCR.

In summary, the RT-PCR reaction consists of three phases:

I. Background phase: the signal cannot be detected, thus the product amount cannot be measured
II. Exponential phase: optimal kinetics. This phase is used for quantification.
III. Plateau phase: PCR amplification dies out (due to excess product concentration relative to primer concentration, decreased enzyme activity, decreased mononucleotide concentration etc.)

The Ct value represents the beginning of the exponential phase.

Each cycle in RT-PCR has the same steps as in traditional PCR: denaturation (95°C), primer annealing (50-64°C) and polymerization (72°C).

Figure 9 shows a classic diagram of RT-PCR reaction for 3 different samples.

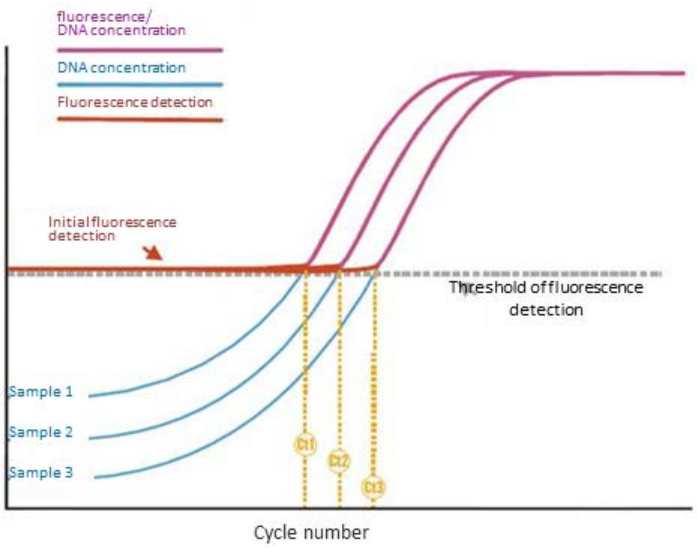


Figure 16: Three samples and their Ct values.

5.1.5 Detection methods in qPCR

Two methods can be employed to observe qPCR amplification products. One is based on the use of fluorescent dyes (nonspecific detection), and the other is based on the use of fluorescent probes (specific detection). Both techniques allow quantification of the products.

The first method uses fluorescent dyes that bind DNA molecules generated during the PCR. The dyes bind any double-stranded DNA nonspecifically, thereby generating a fluorescent signal. This allows the determination of initial DNA concentration in the sample. The most commonly used dye is SYBR Green. The reaction mixture contains the polymerase, dNTPs (mononucleotides), the buffer and the dye. The dye does not bind to single-stranded DNA in the sample. In Figure 10, SYBR Green is depicted as green spheres. When the dye binds to the newly generated double-stranded amplification products during the PCR reaction, it starts emitting fluorescence. When enough product accumulates and the fluorescence signal extends over the background, it can be detected. The cycle at which this happens is called the threshold cycle.

Figure 17: Fluorescent dye based qPCR (SYBR Green). SYBR Green binds to all newly generated double-stranded DNAs and emits fluorescence. Fluorescence accumulates as the PCR reaction progresses and is recorded at the end of each PCR cycle.

The fluorescent probe method of quantification employs probes that bind specifically to particular DNAs. Products can also be evaluated qualitatively; for example, a different allele variant will not generate a signal. The probes have a covalently attached reported dye (e.g. FAM, VIC) and a quencher. During the synthesis of the nascent strand, the polymerase degrades the hydrolysis probe annealed to the template, thereby breaking the close proximity of the reporter and the quencher, which causes fluorescence emission. Figure 18 shows specific annealing of the probe to the template DNA. The synthesis of the new strand displaces the probe and separates the reporter from the quencher, causing fluorescence emission. The emitted fluorescence during each cycle is directly proportional to the amount of the PCR product.

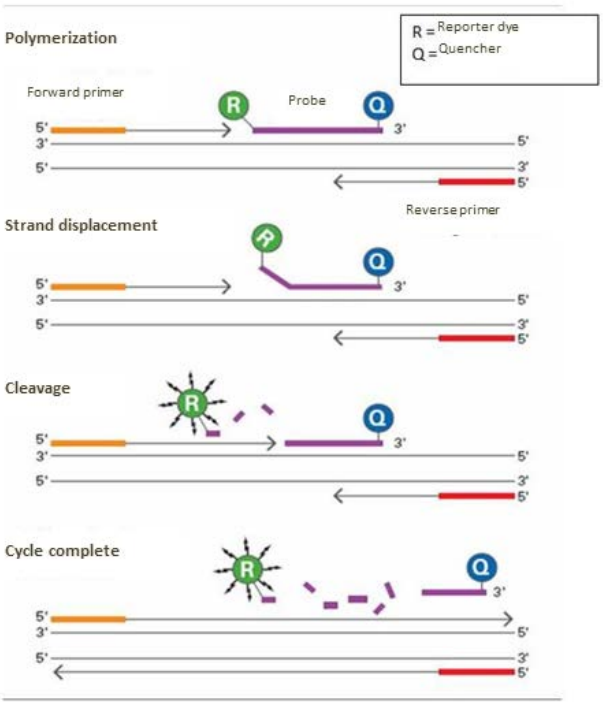


Figure 18: RT-PCR with a TaqMan hydrolysis probe.

This method can also be used to detect several products simultaneously. Several probes (each emitting a different signal) are used together (multiplex). The most commonly used signals are green and red.

5.1.6 Reverse transcription

In order to measure gene expression, the amount of mRNA in the sample is determined. Since PCR can only amplify DNA, mRNA needs to be reverse transcribed into single-stranded complementary DNA (cDNA). This is accomplished by using reverse transcriptase, an enzyme naturally found in viruses (retroviruses employ reverse transcriptase to synthesize DNA from their genetic material (RNA) for integration into the host genome as a provirus). In research, this enzyme is used for synthesis of cDNA for downstream applications.

5.1.7 Data analysis

Qualitative and quantitative analyses

The most basic genetic analyses only require the answer YES or NO – is a certain gene or a specific mRNA present in the sample or not. The qPCR method enables such analysis by using specific probes, which either anneal or not. This method also allows the amount of a specific DNA in the sample to be quantified.
Quantitative methods can be divided into absolute or relative quantification methods. The former enable determination of the absolute quantity of the target molecule. This is often used in virology or microbiology research. The latter, relative quantification, is used to compare the expression levels of two target sequences in the same sample. The result is given as the ratio of expression of the two targets. Relative quantification is often used, since ratios can be compared between different samples (whereas comparing absolute quantities can lack significance, given the differences in the quality of samples (e.g. quality of isolates)).

Why relative quantification?

Relative quantification is the most commonly used quantification method in genetic research (e.g. for gene expression analysis). Ct values are normalized to a reference gene known to have comparable expression in both cell types. Absolute values are not relevant in this case, since the samples do not contain identical amounts of mRNA.

Example: expression of p53 gene in cancer and normal cells

The level of p53 protein is known to change in cancer cells. Scientists were interested in p53-mRNA level change, as well.

a) *The first group of researchers isolated mRNA from 1000 cancer cells and 1000 healthy cells and measured the amount of p53 mRNA in each sample type. A statistically significant difference in p53 mRNA levels was observed after normalization to the number of cells in the sample.*
b) *The second group of researchers measured the levels of GAPDH mRNA in healthy and cancer cells by using equal cell numbers, and observed that the levels were the same. Then qPCR was employed to measure both p53 and GAPDH mRNAs in individual samples of each cell type. The result of relative quantification was given as the ratio of p53 mRNA and GAPDH mRNA, which revealed a statistically significant difference between healthy and cancer cells.*

In both cases, a similar outcome was observed – by using normalization to cell number in the first experiment, or by using normalization to GAPDH levels in the second experiment (taking into account prior observation of equal GAPDH levels in both cell types).

Relative quantification normalized to a reference gene

Because samples used in research generally contain different numbers of cells, normalization to a reference gene (e.g. GAPDH, beta-actin) is typically employed. This technique is fast and simple, but requires knowing a reference gene that has equal expression intensity in all sample types and whose expression is not influenced by therapeutic intervention.
When several samples are compared by relative quantification, one of the samples is selected as the calibrator (the calibrator can also be the average of all samples). Expression of target genes is then presented as an increase or a decrease relative to the calibrator.

Example: expression of p53 gene in cancer and normal cells

We have 3 samples of cancer tissue (B1-3) and 3 samples of healthy tissue (Z1-3). As the calibrator, we can use the average value of all healthy/tested tissues (we could also take only 1 sample or the average value of healthy tissues).

To determine the relative expression of a target gene to a reference gene in a sample, Ct values of the target and reference genes must be determined by qPCR.

Example: expression of p53 gene in cancer and normal cells

	Z1	Z2	Z3	B1	B2	B3
Target gene	Ct_{Z1T}	Ct_{2T}	Ct_{3T}	Ct_{B1T}	Ct_{B2T}	Ct_{B3T}
Reference gene	C_{tZ1R}	Ct_{2R}	Ct_{3R}	Ct_{B1R}	Ct_{B2R}	Ct_{B3R}

Once we know the Ct values, expression levels can be calculated by using several methods. The most commonly used is the "Livak method", also known as the $2^{-\Delta\Delta Ct}$ method.

$2^{-\Delta\Delta Ct}$ method

The Ct values are first calculated relative to the reference gene.

$$\Delta Ct_{sample} = Ct_{sample,T} - Ct_{sample,R}$$

$$\Delta Ct_{calibrator} = Ct_{calibrator,T} - Ct_{calibrator,R}$$

The ΔCt values of samples are then normalized to the calibrator.

$$\Delta\Delta Ct_{sample} = \Delta Ct_{sample} - \Delta Ct_{calibrator}$$

Finally, the expression level is calculated:

$$2^{-\Delta\Delta Ct} = \text{normalized expression level}$$

This result determines whether there is an increase or a decrease of target gene expression relative to the calibrator and normalized to the reference gene. The result is completely independent of the amount of starting material or the number of cells in the sample.

Example: expression of p53 gene in cancer and normal cells

The following values have been measured for p53 mRNA (target gene) and GAPDH mRNA (reference gene).

	Z1	Z2	Z3	B1	B2	B3
Ct – target gene	15.0	16.8	15.9	12.0	11.5	12.2
Ct - reference gene	17.2	17.9	17.7	17.8	17.3	17.7
ΔCt	-2.2	-2.1	-1.8	-5.8	-5.8	-5.5

As calibrator, average ΔCt values can be used: ΔCt = -3.9

ΔΔCt is calculated.

e.g. for the first sample Z1: ΔΔCt = (-2.2) – (-3.9) = 1.7 $2^{-\Delta\Delta Ct}$ = 0.308

And for other samples:

	Z1	Z2	Z3	B1	B2	B3
ΔΔCt	1.7	1.8	2.1	-1.9	-1.9	-1.6
$2^{-\Delta\Delta Ct}$	0.308	0.287	0.233	3.732	3.732	3.031

We can see that p53 gene expression is significantly higher in cancer tissues relative to controls, which indicates that increased p53 mRNA levels are associated with carcinogenesis.

5.2 EXERCISE PROTOCOL

MATERIALS

Sample cDNA, PCR reagents (SybrGreen Mastermix, primers), pipettes, microtiter plate, qPCR thermocycler.

METHOD

Expression of TNF gene will be measured by qPCR in 5 different samples. As a reference gene, B2M will be used.

- Use one of the samples for generating a standard curve. Prepare the following dilutions: 1:1, 1:2, 1:5, 1:10, 1:20, 1:40, 1:80. Pipette these dilutions onto the microtiter plate, as shown in the table below:

	1	2	3	4	5	6	7	
A	V1 1:1	V1 1:2	V1 1:5	V1 1:10	V1 1:20	V1 1:40	V1 1:80	→ *TNF*
B	V1 1:1	V1 1:2	V1 1:5	V1 1:10	V1 1:20	V1 1:40	V1 1:80	→ *TNF*
C	V1 1:1	V1 1:2	V1 1:5	V1 1:10	V1 1:20	V1 1:40	V1 1:80	→ *B2M*
D	V1 1:1	V1 1:2	V1 1:5	V1 1:10	V1 1:20	V1 1:40	V1 1:80	→ *B2M*
E	V1	V2	V3	V4	V5	V6	V7	→ *TNF*
F	V1	V2	V3	V4	V5	V6	V7	→ *B2M*

- To measure gene expression, prepare two reaction mixtures:

Target gene

TNF	1	x
2*SYBR GREEN	5	
primer 1	1.2	
primer 2	1.2	
cDNA	2	
H_2O	0.6	
	10	

Reference gene

B2M	1	x
2*SYBR GREEN	5	
primer 1	0.5	
primer 2	0.5	
cDNA	2	
H_2O	2	
	10	

- Add 8 µL of the appropriate reaction mix to each sample, following the table.
- Determine the Ct values for all dilutions for each gene and draw the standard curve.
- Determine the Ct values for all samples and calculate relative expression by using the $2^{-\Delta\Delta Ct}$ method.

5.3 RESULTS

Standard curve for target TNF gene:

Dilution	C	Log(C)	Ct_1	Ct_2	$Ct_{average}$
1:1	1				
1:2					
1:5					
1:10					
1:20					
1:40					
1:80					

Place for graph:

Standard curve for the reference B2M gene:

Dilution	C	Log(C)	Ct_1	Ct_2	$Ct_{average}$
1:1	1				
1:2					
1:5					
1:10					
1:20					
1:40					
1:80					

Place for graph:

Sample	Ct_{TNF}	Ct_{B2M}	ΔCt	$\Delta Ct_{average}$	ΔΔCt	$2^{-\Delta\Delta Ct}$
S1						
S2						
S3						
S4						
S5						
S6						
S7						

5.4 DISCUSSION

5.5 COMPUTATIONAL EXERCISE

Calculate how to prepare dilutions of your samples (how much water and concentrated stock solution you would mix) to obtain the necessary mixtures (dilutions) for the experiments below.

	H_2O	Starting solution	Final volume
1	/		
1:2			
1:5			
1:10			
1:20			
1:40			
1:80			

Primer stock solution is 5 µM

250nM primer		
	1 sample	___ samples
H_2O		
SYBR Green Mix	5	
P1		
P2		
DNA	2	
Total volume	10	

300 nM primer		
	1 sample	___ samples
H_2O		
SYBR Green Mix	5	
P1		
P2		
DNA	2	
Total volume	10	

350 nM primer		
	1 sample	___ samples
H_2O		
SYBR Green Mix	5	
P1		
P2		
DNA	2	
Total volume	10	

6 LITERATURE

1. William S. Klug, Michael R. Cummings, Charlotte A. Spencer, Michael A. Palladino: Concepts of Genetics, 10th edition, Pearson, 2012
2. William S. Klug, Michael R. Cummings, Charlotte A. Spencer, Michael A. Palladino: Essentials of Genetics, 8th edition, Pearson, 2012
3. Harvey Lodish, Arnold Berk, Chris A. Kaiser, Monty Krieger, Anthony Bretscher, Hidde Ploegh, Angelika Amon, Matthew P. Scott: Molecular Cell Biology, 7th edition, Freeman, 2012
4. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402-408 (2001).
5. Wittwer CT. High-resolution DNA melting analysis: advancements and limitations. Hum Mutat. 2009, 30(6):857-9.
6. Vossen RH, Aten E, Roos A, den Dunnen JT. High-resolution melting analysis (HRMA): more than just sequence variant screening. Hum Mutat. 2009 Jun; 30(6):860-6.