Introduction into molecular medicine

. László Bálint Bálint
 . Bertalan Meskó
 . László Nagy
 . Árpád Lányi
 . Beáta Scholtz
 . Lajos Széles
 . Tamás Varga

Edited by . László Nagy and . László Bálint Bálint
 , University of Debrecen
Introduction into molecular medicine
by László Bálint Bálint, Bertalan Meskó, László Nagy, Árpád Lányi, Beáta Scholtz, Lajos Széles, Tamás Varga, László Nagy, László Bálint Bálint, and
Copyright © 2011 The project is funded by the European Union and co-financed by the European Social Fund., Manuscript completed: 17 November 2011

“Manifestation of Novel Social Challenges of the European Union in the Teaching Material of Medical Biotechnology Master’s Programmes at the University of Pécs and at the University of Debrecen” Identification number: TAMOP-4.1.2-08/1/A-2009-0011
# Table of Contents

Index .......................................................................................................................... 1
1. 1. Introduction into Molecular medicine ................................................................. 2
    1. .......................................................................................................................... 2
2. 2. The Genome ........................................................................................................ 3
    1. The nucleus ...................................................................................................... 3
3. 3. Sequencing of the genome ................................................................................ 9
    1. .......................................................................................................................... 9
4. 4. Investigation of the human genome with microarrays .................................... 13
    1. .......................................................................................................................... 13
5. 5. Genes and Diseases ........................................................................................... 16
    1. 5.1. How does the genome functions? .............................................................. 16
6. 6. Nuclear Receptors .............................................................................................. 23
    1. .......................................................................................................................... 23
7. 7. Personalized genetics ........................................................................................ 25
    1. .......................................................................................................................... 25
8. 8. Immunodeficiencies ......................................................................................... 29
    1. .......................................................................................................................... 29
9. 9. Cancers ............................................................................................................... 33
    1. .......................................................................................................................... 33
10. 10. Molecular mechanisms of cancer development ........................................ 36
    1. .......................................................................................................................... 36
11. 11. Obesity: introduction .................................................................................... 39
    1. .......................................................................................................................... 39
12. 12. Genomics of obesity ....................................................................................... 42
    1. .......................................................................................................................... 42
13. 13. Genes involved in development of obesity .................................................... 43
    1. .......................................................................................................................... 43
14. 14. Therapeutic approaches to obesity ............................................................... 45
    1. .......................................................................................................................... 45
15. 15. Interconnected mechanisms in lipid metabolism ......................................... 46
    1. .......................................................................................................................... 46
List of Figures

2.1. Figure 2.1. Localization of genetic material in the cell .......................................................... 3
2.2. Figure 2.2 Representation of chromosomes .............................................................................. 3
2.3. Figure 2.3. Euchromatin, heterochromatin and transcription regulation. Chromosomes in a non dividing cell ......................................................................................................................... 4
2.4. Figure 2.4. The nucleus .................................................................................................................. 5
2.5. Figure 2.5. Chromosomes and epigenomes ................................................................................ 6
2.6. Figure 2.6. Structuring of the genome ......................................................................................... 6
2.7. Figure 2.7. Complex organization of the nuclei .......................................................................... 7
3.1. Figure 3.1. Sanger sequencing ..................................................................................................... 9
3.2. Figure 3.2. The fluorescent automated capillary sequencing ....................................................... 9
3.3. Figure 3.3. Shotgun sequencing ................................................................................................ 10
3.4. Figure 3.4. The first high throughput massive parallel sequencing technology ...................... 11
4.1. Figure 4.1. The SNP-s .................................................................................................................. 14
5.1. Figure 5.1. Genomes and their regulation .................................................................................... 16
5.2. Figure 5.2. Components of the human genome .......................................................................... 17
5.3. Figure 5.3. Genes encoded on opposite strands ......................................................................... 17
5.4. Figure 5.4. The origin of repetitive elements .............................................................................. 18
5.5. Figure 5.5. Transposones ............................................................................................................ 18
5.6. Figure 5.6. Genomic variations and diseases .............................................................................. 19
5.7. Figure 5.7. Duchenne and Becker type muscular dystrophy ....................................................... 20
5.8. Figure 5.8. Globin recombinations ............................................................................................. 20
5.9. Figure 5.9. Repetitive transcriptional units .................................................................................. 21
5.10. Figure 5.10. Mutations in genes present in several Osteogenesis imperfecta and Ehlers Danlos syndrome ................................................................................................................................. 22
6.1. Figure 6.1. Molecular mechanisms of nuclear receptor driven transcriptional regulation ........ 23
6.2. Figure 6.2. Evolution of nuclear receptors ................................................................................... 24
7.1. Figure 7.1. Personalized therapies .............................................................................................. 27
8.1. Figure 8.1. Duplication time of pathogens is short ...................................................................... 29
8.2. Figure 8.2. X linked autoimmune diseases .................................................................................. 29
8.3. Figure 8.3. NK cells are regulated by activating and inhibitory signals .................................... 30
8.4. Figure 8.4. In the absence of SLAM-Associated Protein (SAP) adapter adapter protein, in the NK cells of XL patients the activating 2B4 receptors act as inhibitory receptors ........................................ 30
8.5. Figure 8.5. The iNKT cells, contrary to conventional T cells recognize glycolipids ................. 31
9.1. Figure 9.1. Cancers can be of viral origin .................................................................................. 33
9.2. Figure 9.2. Viruses can cause cancers in cell cultures ............................................................... 34
9.3. Figure 9.3. Viruses can be integrated into the genome .............................................................. 34
10.1. Figure 10.1 The mechanism of self replicating proliferation .................................................... 36
10.2. Figure 10.2 Mechanism of autocrine growth ............................................................................ 36
10.3. Figure 10.3. Is ps3 an oncogene or a tumor suppressor? ....................................................... 37
10.4. Figure 10.4 Structure and function of p33 ................................................................................. 38
10.5. Figure 10.5. A p53 és apoptózis kapcsolat .................................................................................. 38
11.1. Figure 11.1. Obesity: environmental cultural and internal factors regulate energy input. Energy input and consumption has to be in balance .............................................................................. 39
11.2. Figure 11.2 Twin studies and the genetics of obesity .............................................................. 40
13.1. Figure 13.1. The role of UCP-s and thermogenesis ................................................................... 43
15.1. Figure 15.1 Molecular mechanismos of adipocyte differentiation .......................................... 46
15.2. Figure 15.2. PPARgamma is modulating metabolism in several tissues .................................. 46
15.3. Figure 15.3. Circulation of cholesterol ...................................................................................... 47
15.4. Figure 15.3. PPARg and lipid uptake ....................................................................................... 48
15.5. Figure 15.4. Uptake of cholesterol into the cells ........................................................................ 48
15.6. Figure 15.5. Differentiation of embryonic stem cells ............................................................... 49
Index

Table of Figures

• Figure 2.1. Localization of genetic material in the cell
• Figure 2.2. Representation of chromosomes
• Figure 2.3. Euchromatin, heterochromatin and transcription regulation. Chromosomes in a non dividing cell
• Figure 2.4. The nucleus
• Figure 2.5. Chromosomes and epigenomes
• Figure 2.6. Structuring of the genome
• Figure 2.7. Complex organization of the nuclei
• Figure 3.1. Sanger sequencing
• Figure 3.2. The fluorescent automated capillary sequencing
• Figure 3.3. Shotgun sequencing
• Figure 3.4. The first high throughput massive parallel sequencing technology
• Figure 4.1. The SNP-s
• Figure 5.1. Genomes and their regulation
• Figure 5.2. Components of the huma genome
• Figure 5.3. Genes encoded on opposite strands
• Figure 5.4. The origin of repetitive elements
• Figure 5.5. Transposones
• Figure 5.6. Genomic variations and diseases
• Figure 5.7. Duchenne and Becker type muscular dystrophy
• Figure 5.8. Globin recombinations
• Figure 5.9. Repetitive transcriptional units
• Figure 5.10. Mutations in genes present in several Osteogenesis imperfecta and Ehlers Danlos syndrome
• Figure 6.1. Molecular mechanisms of nuclear receptor driven transcriptional regulation
• Figure 6.2 Evolution of nuclear receptors
• Figure 7.1. Personalized therapies
• Figure 8.1. Duplication time of pathogens is short
• Figure 8.2. X linked autoimmune diseases
• Figure 8.3 NK cells are regulated by activating and inhibitory signals
• Figure 8.4. In the absence of SLAM-Associated Protein (SAP) adapter adaptor protein, in the NK cells of XLPatients the activating 2B4 receptors act as inhibitory receptors
• Figure 8.5 The iNKT cells, contrary to conventional T cells recognize glycolipids
• Figure 9.1. Cancers can be of viral origin
• Figure 9.2. Viruses can cause cancers in cell cultures
• Figure 9.3. Viruses can be integrated into the genome
• Figure 10.1 The mechanism of self replicating proliferation
• Figure 10.2 Mechanism of autocrine growth
• Figure 10.3. Is p53 an oncogene or a tumor suppressor
• Figure 10.4 Structure and function of p53
• Figure 10.5. The p53 and apoptózis connection
• Figure 11.1. Obesity: environmental cultural and internal factors regulate energy input. Energy input and consumption has to be in balance
• Figure 11.2. Twin studies and the genetics of obesity
• Figure 15.1. Molecular mechanisms of adipocyte differentiation
• Figure 15.2. PPARgamma is modulating metabolism in several tissues
• Figure 15.3. Circulation of cholesterol
• Figure 15.3. PPARg and lipid uptake
• Figure 15.4. Uptake of cholesterol into the cells
• Figure 15.5. Differentiation of embryonic stem cells
Chapter 1.1. Introduction into Molecular medicine

1.

A timeline

I. Conceptual discoveries


II. Development of methodologies


III. Modern era (1980-2001)

Chapter 2. 2. The Genome

1. The nucleus

Every nucleus of our body contains the same DNA. In each cell, DNA is located in several places, in nuclei and in mitochondria. The DNA in the nuclei is made of several molecules of DNA. Two similar molecules, named sister chromatids are connected and form the chromosomes. The mitochondrial DNA is different, and is made up of tens, hundreds of copies of circular DNA molecules.

**Figure 2.1. Localization of genetic material in the cell**

On the first picture we can see the nucleus and the mitochondria as the organells of DNA.

The organelle which contains the genome is the nucleus. Each cell of our body contains the same information encoded in the DNA. In a cell DNA is located beside the nucleus in other organelles too, in humans it is located in the mitochondria. The genetic material of the nucleus is made up of several DNA molecules. Each type of DNA molecule is present in two copies one of maternal and one of paternal origin. These are called chromatids and the two sister chromatids form the chromosome. The mitochondrial genome is very different from the nuclear genome and consists of small circular DNA molecules. Each cell contains tens, hundreds or even thousands of similar copies of mitochondrial DNA molecules.

Chromosomes were identified by light microscope studies and named based on their size. We have 22 somatic and two sex chromosomes, in total 46 chromosomes.

Chromosomes are rod-like structures only during cell division. In all other situations, the chromosomes are unwrapped. By transmission electronmicroscopic studies, we can distinguish hetero and euchromatin. The chromosome has four arms and is made up of two DNA molecules united at the centromeres. These two molecules are called sister chromatids. There is only one specific cell state when these can be seen in the rod-like structure, namely during the cell division.

**Figure 2.2. Representation of chromosomes**
In the picture, we can see the classical representation of the human chromosomes. Chromosomes were identified based on studies performed with light microscopes. The chromosomes show this specific banding if stained with special dyes.

The chromosomes were numbered based on their size, the first chromosome is the largest one. We have 22 so-called somatic chromosomes and two sex chromosomes. The two sex chromosomes in females there are two x chromosomes and in males there is an x and a Y chromosome. In total we have forty-four plus two equals forty-six chromosomes. Above this we have the mitochondrial DNA that can be present in a cell in several hundreds of copies. In all other states, the DNA in the nucleus is in a loose state.

**Figure 2.3. Euchromatin, heterochromatin and transcription regulation. Chromosomes in a non dividing cell**

In the figure, we can see how the DNA looks like if it is stained with a dye and visualized through a microscope. In a non-dividing cell, the chromosomes look like an unwound clew, while in a dividing cell they look more like rods.
In an electronmicroscopic image, the heterochromatin is darker and is composed of transcriptionally silent DNA regions. On the other hand, the light gray material seen in the electronmicroscopic picture is the euchromatin, the regions involved in transcription, in the generation of RNA from DNA. Each cell has identical DNA (with a few exceptions) but the genes transcribed in one or another cell are very different in each individual cell.

If investigated by transmission electron microscopy we can distinguish an electron-dense region called heterochromatin and a less electron-absorbent region called euchromatin. Euchromatin shows up light on electronmicroscopic images and represents the regions where gene transcription is present, while heterochromatin is electron dense, dark and is the region which is mute from a transcriptional point of view. Each cell has the same DNA (with some exceptions) but the transcribed regions are very different. The DNA in a cell is 4 meters long, meaning that in the nucleus it has to coil up. This structure is not random, but rather tightly controlled. DNA and proteins together form the chromatin. Nuclei have dual membranes, pores, nucleolus and other structures.

**Figure 2.4. The nucleus**

In the nuclei, we can observe interchromosomal areas where very intensive trafficking is going on. The DNA is wrapped on histone octamers and forms solenoid structures with 30 nm diameter.

How is the DNA coiled on proteins?

The DNA wrapped on the histone octamer complexes forms the units called nucleosomes. Each nucleosome contains two and a half turns of DNA which is 146-152 base-pairs long. The diameter of a nucleosome is 11 nanometers, which means that it is one ten-thousands part of a millimeter and can be visualized only with specific atomic force microscopes.

What is the function of this unit?

The environment is communicating with the genome through the histone tails. The core of the nucleosomes is made of two of each histone variants: H2A, H2B, H3, H4. The tails of the histones reach out beyond the DNA and contain very reactive residues, which can be post-translationally modified in a variety of ways. The epigenome is made of the totality of the modifications that occur on the DNA and the histone tails. A genome can be presented through as many epigenomes as the types of cells that exist. The epigenome is the totality of the possibilities of transcriptional events, meaning that each epigenome can produce various transcriptomes depending on the incoming signals. This means that various cells can give various responses to the same signal.
based on the epigenetic translation of the genomic material. The epigenetic signatures will determine what response will be turned on by a specific incoming signal and what is the transcriptomic effect produced by an input.

What is the role of nucleosomes?

**Figure 2.5. Chromosomes and epigenomes**

Nucleosomes are the sites where environmental signals communicate with the genome, through the histone tails. Histone octamers are constituted by two molecules of each H2A, H2B, H3 and H4 histones, each having a globular core and a protruding tail, which are highly reactive. The DNA has to be coiled in order to fit in the nucleus. Each cell contains approximately four meters of DNA, which is coiled four hundred thousand fold. This coiling is not random but tightly regulated by proteins. These proteins together with the DNA form the chromatin.

The structure of the genome

The sequenced human genome can be considered as a string of unstructured letters. Like all texts, the genome is also structured. The DNA methylation for example is similar to breaking the string of letters into words. Capitalization and punctuation as ways of structuring a text, in turn, are similar to the way in which histone tail modifications are structuring the genome. Beside this, we have to take into consideration that the DNA in the nucleus is not linear, it has a rhythm in this case like that of a sonnet. The picture is complete if take into consideration that nucleosomes can form strings and are structuring the genome into closed heterochromatic and open euchromatic regions.

What is the relevance of these modifications from a medical point of view? The majority of these modifications are performed by enzymes, which can be modified by chemical inhibitors. DNA methylation can be blocked by azacytidine, histone tail modifications by TSA, pargyline or valproic acid. By introducing chemical substances, we can modify the way in which the genome functions.

The tail of the histones contains various modifications that can be modified by drugs, whereby theses chemical signals directly influence the context of transcription.

**Figure 2.6. Structuring of the genome**
Each genome is represented in the form of several epigenomes, which all give rise to different transcriptomes based on the actual environmental signals. The epigenome represents the totality of the modification seen on DNA and the histone tails that together determine what transcription factors will bind to that genomic area.

The genome can be conceived of as a long string of unstructured letters, structured into words by DNA modifications. The histone tail modifies them to form sentences, and the rhythm of the nucleosomes organizes them in a recognizable textual pattern, such as a sonnet.

We might think that the chromatin in the nucleus is disorganized, but this is not true. If we stain different chromosomes with different colors, we can see that different chromosomes occupy different territories. Even the replication of different chromosomes is happening at different time points. If we stain the RNA with red and the chromatin is marked by a GFP tagged histone, we can see that the processing of RNA is happening in chromatin free territories. If we block the transcription, the regions where the RNA processing is occurring, the so called Speckle regions, are becoming larger, showing us that RNA processing and transcription are linked.

In the nucleus, we can observe the interchromosome territories, which contain nucleoplasm and are involved in the active transport of proteins and RNA molecules.

The nucleus itself has a double nuclear membrane, and it is communicating with other parts of the cell through the nuclear pores. Through the pores, there is a well-controlled transport of materials in both directions. RNA molecules are exported with the help of exportins and in the cytoplasm they are translated into proteins.

In the nucleus, beside chromatin, we can distinguish the nucleolus where a special RNA transcription is performed namely the production of the ribosomal RNA.

**Figure 2.7. Complex organization of the nuclei**
The nucleolus itself contains the more than 200 tandem repeats of the ribosomal genes located on chromosomes 13, 14, 15, 21 and 22, the large precursor molecules of the ribosomal RNA, RNA processing enzymes and snoRNA molecules which guide the enzymes to their site of action. The assembly of the ribosome subunits involves additional rRNA molecules and proteins. The subunits are exported through the nuclear pores and remain free or are assembled during the translation.

Besides this, we have other formations in the nuclei that are not fully characterized, for example the speckles where the processing of the RNA molecules is performed.

In the case of nuclei, one might believe that the chromosomes are arranged in a random fashion, which, however, is not the case. If each chromosome was stained specifically with probes which recognize that single chromosome, one could observe that each chromosome occupies a distinct territory. These territories have some degree of independence, for example, they divide separately. Other regions of the nuclei are, for example, the speckles. These are RNA processing territories, where splicing is performed. If the transcription is blocked by chemical agents, these regions are increased.
Chapter 3. 3. Sequencing of the genome

1.

The most well-known technology for sequencing is the so called Sanger sequencing.

Figure 3.1. Figure 3.1. Sanger sequencing

In this case DNA is sequenced in four parallel lanes with a nucleotide in each, by using radioactive primers and a small amount of dideoxi NTP-s (different ones in each lane), which causes chain termination at a low frequency. This technology allows sequencing up to more than 100 bp-s.

Figure 3.2. Figure 3.2. The fluorescent automated capillary sequencing
The basic technology developed by Fred Sanger was soon transferred to an automatic system developed by Applied Biosystems and that mixed the fluorescently labeled samples in one sample and run them in a capillary, in front of a color detector. The result was an instrument which could run in several samples in a parallel fashion and this technology was used for the sequencing of the human genome. By eliminating steps of sample preparation, the speed of the sequencing and the reagent cost was decreased.

The task was performed by several research groups: it was started by National Human Genome Research Institute under the guidance of James Watson, who, together with Francis Crick, described the structure of DNA in Nature in 1953, and it was completed in 2001. An important person in the rapid finalizing of the project was Craig Venter, who said that by eliminating the labor-intensive planning phase, he could speed up the sequencing by increasing the coverage and finding overlapping sequences.

Figure 3.3. Shotgun sequencing
The increased coverage would allow computer scientists to generate the map much faster. By combining the two methods, the genome was completed in 2001. The approach of shotgun sequencing became the standard, replacing hierarchical sequencing.

Recently, the speed of sequencing has been further increased by eliminating other sample preparation steps. The recent developments eliminated the bacteria from the clonal amplification of the individual DNA molecules, and this amplification is performed on solid phase (Illumina) or bead in an emulsion PCR reactions (454 technology). The beads in the small buffer droplet allow the clonal amplification of a single DNA molecule in a couple of hours and later in a pyrosequencing reaction the DNA is sequenced by light emission in a luciferase reaction, combining the images in a bioinformatic way.

**Figure 3.4.** The first high throughput massive parallel sequencing technology
“454” sequencing technology

DNA library
↓
1 DNA/ bead
↓
Oil/water emulsion microreactors
↓
Beads on a plate
↓
Millions of parallel sequencing reactions
Chapter 4. 4. Investigation of the human genome with microarrays

1.

Microarrays are tools that contain DNA molecules with known sequences hybridized on a solid support, usually a glass surface. Microarrays can be considered the “microscopes” of the genome. According to their method of production, they can be printed or synthesized and according to the labelling approach, they can be single color or two colored microarrays.

Their working principles are quite simple and rely on the double helix structure of the DNA published in 1953 by Watson and Crick. Based on this model, the two strands of the DNA molecule are complementary and bound by hydrogen bonds. If the two strands are separated (e.g. by heat), they will rejoin if the conditions are optimal for this reaction. The basic reaction is called hybridization. If one strand is labelled with a fluorophore, the hybridization can be visualized by the presence of the dye. In the case of the microarrays one strand is fixed on a solid surface, the sample is labelled and heated, the single stranded sample will find its counterpart probes on the solid surface. Hybridization is used in various DNA based technologies from PCR to in situ hybridization.

The generation of microarrays follows the same logic. A target sequence library is amplified and linked to a solid surface. After the samples are labelled and hybridized, washing and scanning are performed and pixel intensities are further interpreted by bioinformatic analysis.

Classically, microarrays were printed with multi-channel printers. The tips of these printers were able to spot nanoliter volumes of DNA solutions. The same sequence was printed on hundreds of microarrays, thereby lowering the price of microarrays. The small volume (0.2-0.6 microliters) of spotted DNA solution was linked to the surface of the glass slide. Since the order of the various sequences on the glass surface was known, the linked probes on the solid surface allowed the researchers to compare the nucleic acid composition of two samples: e.g. healthy and diseased. RNA is extracted from both samples and the two samples are labelled with two different dyes and hybridized to the arrays. Classically the two dyes used are Cy3 and Cy5 and if excited they emit a signal in red or green colour respectively. The arrays are scanned twice with different filters and the high resolution images are superposed. If the two colors are similar in intensity, the result is yellow, meaning that the two samples express the same amount of mRNA of that particular gene. If any colour is predominant, the corresponding sample is expressing more of that particular gene. If the particular gene is not expressed, the spot will not show any signal. By knowing the exact sequences on each location on the array, we can interpret the results in the context of the spotted cDNA clones on the array.

These types of classical spotted microarrays are not used any more due to the higher reproducibility and better specificity of the the oligo-based microarrays. The best known array of this group is the so called Affymetrix array. These arrays are not produced by printing but by a technique called photolithography. In this case, the oligos are synthesized in situ on the glass surface with the help of masks and UV light. Tiny holes on the masks are created with lasers and through these holes the UV light is deprotecting light sensitive oligonucleotides. Deprotection makes these molecules reactive and the deprotected oligos serve as substrate for the next step of the synthesis reaction. By repeating the reaction several times on the glass surface, a single stranded DNA molecule is synthesized, which serves as a probe for the steps to follow. Technologies acquired from computer technology allow the generation of probes in very high density, namely millions of probes per single array, with a feature in the range of micrometers. By now, the resolution of the scanners is the limitation of the probe densities.

Other developments of computer technology have also been incorporated into the production of microarrays. For example, the HP derived Agilent Technologies use the technology used in printing to deliver the reagents for in situ oligo synthesis. Another company, Roche-Nimbegen uses the technology used in the DLP projectors, which consists of individually controlled micromirrors, to deprotect by UV light the protected bases, similar to the Affymetrix type photolithography. This production does not need special masks and thereby the arrays can be customized. These arrays are longer and have a much better controlled melting temperature, and due to this they can be used for the so called array CGH: comparative genomic hybridization. These investigations allow the identification of the insertion, deletions. Currently these questions are addressed by cytogenetics and FISH but with a much smaller resolution.
Currently, microarrays can be considered to be a basic technology, with kits available that work in everyone’s hands based on well elaborated protocolls, and the prices of instruments and reagents is decreasing year by year. The technology has become well standardized and the results are intercomparable among labs and systems, as demonstrated by the MAQC study (The MicroArray Quality Control project, 2006) supervised by FDA.

New technologies

In the last couple of years, we can hear about the continuous decrease of the price of sequencing. Will these technologies replace the microarray technology? In the field of research, microarrays will be probably considered as straightforward as the PCR is today. The analysis of the sequencing data for diagnostics needs to generate results within hours be available in almost every major diagnostic unit. This is a challenge that will probably be met only in future phases of technological development.

Medical applications: what can be studied?

Almost any sample with nucleic acid structure can be studied by microarrays if it is not a repetitive sequence. As a result, genexpression (from RNA samples), copy number variations (CGH-from DNA samples), single nucleotide polimorphisms (from DNA) or microRNA-s can be investigated. The SNP-s are considered to be responsible for the individual differences among patients. One of the most important array of this type is the Roche CYP 450 genotyping array which covers all the variants of the CYP 450 enzymes involved in drug metabolism. By surveying these variants, one can predict the speed of the degradation of various drugs, and thereby the time used to find the right dose for an individual patient can be shortened. The probes on the arrays can be of human origin but patogens can also be investigated.

The Single Nucleotide Polymorphism (or SNP) is a variation in the DNA sequence which occurs when one nucleotide changes in the genome, but the variation is only considered SNP when it appears in 1% of the population. SNPs represent 90% of human genetic variations which means they appear every 100-300 basepairs in the genome. As the changes in the DNA affect the way the human organism reacts to diseases, external factors such as infections or chemicals, these are used more and more often in medical research, drug developments and diagnostics.

Craig J. Venter who is one of the leaders of the Human Genome Project had 3,213,401 SNPs after his genome was sequenced.

**Figure 4.1.** The SNP-s

Craig Venter had 3,213,401 SNPs.

The road to medical applications
In order to develop medical applications, patentability has to be investigated. Basic patents in the field of microarrays cover the technology inclusive of probe density. Other types of patents cover disease type gene expression patterns. A new type of patents are those that use the Affymetrix microarrays and describe specific gene expression patterns identified on this array type. The Affymetrix arrays thereby become a platform where others can develop their own intellectual property. Such developments are already used by biotech companies that developed technologies that are able to identify translocations without copy number variations based on the specific gene expression changes or even SNP-s from RNA samples.

One of the obstacles in front of the diagnostic applications of the microarrays is their relatively high price. One of the possibilities to bypass this obstacle is the multiplexing of the samples, a possibility that has become available for all the major microarray companies (Agilent, Affymetrix, Illumina and Roche-Nimblegen). Some of the vendors split up the array into smaller areas to allow multiplexing while others developed special high throughput, fully automated platforms in 96 well format.

One of the important competing technologies in diagnostics is real-time quantitative PCR, which gives more precise results and is much faster compared to the microarray technology. On the other hand, this technology is relatively low throughput, or if the throughput is increased the price becomes higher than the price of microarrays. The QPCR technology can be considered a validation method of the microarray results since these two methods are totally independent.

How will microarrays be used in clinical practice?

The processing of the samples will be done probably in core laboratories. The time frame of sample processing will be difficult to reduce to under a couple of days. Areas of application will be probably creating of subgroups for diagnosis, clarification of the genetic background of heritable diseases, genetic serotyping of pathogens, identification of biomarkers. What will a typical result look like? Data processing will be performed in a closed bioinformatic pipeline, and questions will be answered probably with “yes” or “no” meaning the presence or absence of a specific marker, pattern etc. Other types of answers will look like regular laboratory results where actual parameters and reference values will be given. What is clearly a demand is that no bioinformatic knowledge should be necessary to interpret the results.

What kind of samples will be sent to the laboratory? DNA based samples will be probably stored and shipped under the protection of EDTA. DNA will be extracted from blood, cytological aspirates, blood fraction or biopsy. The processing of RNA is a more difficult issue. RNA has to be processed under the protection of stabilizing agents, or snap frozen in liquid nitrogen.
Chapter 5. 5. Genes and Diseases

1. 5.1. How does the genome functions?

5.1.1. Repetitive regions and the genome

If we compare the size of different species we can see that humans have 30k genes similar to mice. Drosophila has 14k genes, but its total genome is less than one tenth in size. While the size of the genome increased in the course of evolution, the number of genes does not show a spectacular increase.

Figure 5.1. Genomes and their regulation

If we compare the size of the genome of different species and we look at the number of genes in those organisms we can see that humans have 30 k genes like mice, while fruitflies have 13,6 k genes but their genome is substantially smaller 180 million bases compared to the three billion bases in human. We can see that in the course of evolution, while the size of the genome increased, the number of genes increased at a smaller rate

What can be the explanation of this?

One of the explanations is that the way in which the genome functions is not determined by the number of genes, but by the way they are regulated. Between genes, there are large territories deficient in genes. Little is known about the function of these gene-deficient regions, which used to be called junk DNA. Today we know that these are involved in gene regulation, but we do not know the exact mechanisms thereof.

The DNA in the genome can be grouped in unique sequences and repetitive regions. More than half of our genome is made up of non-gene regions. But what is a gene? Traditionally, the genetic content associated with a property of the organism can be considered one gene. Exons and introns of the gene with the core promoter make up only up to 5% of the genome.

The DNA of the nucleus can be characterized as unique sequences or repetitive regions. The unique regions code for genes and gene regulatory regions. Almost half of our genome consists of repetitive regions with unknown functions.
What is a gene?

The concept of the gene was determined based on properties of biological organisms. The current approach states that a gene is the totality of coding and non-coding regions that together are responsible for the transcription of a protein. Based on today’s knowledge, less than five percent of the genome can be considered to be such genes. Regions that directly code for the amino acid sequences of proteins are less than 1 percent of the whole genome.

**Figure 5.2. Components of the human genome**

![Diagram of the Human Genome](image)

What is the origin of the repetitive regions?

One type of repetitive regions are the so called pseudoegenes. These originate from the genes transcribed. If a reverse transcriptase is present in the cell (e.g., by an infection with a retrovirus), the genes transcribed and spliced will be reverse transcribed, but without their introns. This cDNA can be integrated into other locations, usually without regulatory and promoter regions. The result is that a pseudo gene will be generated.

**Figure 5.3. Genes encoded on opposite strands**

![Diagram of Gene Encoding on Opposite Strands](image)

To understand the complexity of the genome and how exons and introns are related, we have to understand that genes can be encoded on both strands even in an overlapping fashion or as in the example of NF1 gene. The region between the exons 26 and 27, which is the intron 26 on the opposite strand as the encoding of NF1 gene, we can see three regular genes OGMP, EVI2A and EVI2B.

There are repetitive regions originating from transposons. There are several well documented cases which report that transposons can be active in humans. How was their activity identified? By careful investigation of families...
where some inherited disease was identified, and finding that the real biological parents do not have that transposon like sequence in their genomes. Like Line 1 transposon in hemophylic patients, in thalassemic and muscular dystrophic patients. It is highly likely that in other cases these transposons are also active, but unless they do some harm, we do not find the signs of their activity. Chromosomal breaks and rearrangements might be responsible for some of the cancer cases meaning that these kinds of rearrangements could be later identified in cancer samples by deep sequencing.

**Figure 5.4. The origin of repetitive elements**

Repetitive sequences might cause other diseases too, like X linked metal retardation where the repetitive triplets instead of 30 copies can be up to thousands of copies or in Huntington disease where the CAG repeats produce long glutamine stretches and change the activity of the proteins.

Other repetitive regions are for example the genes in the globin cluster. Originally, we had probably one single copy, now we have two clusters on chromosomes 16 and 11-th namely the alpha and globin gene clusters. Every gene duplication opens the potential for a new function and this can be clearly seen in the case of globin genes, e.g., in fetal hemoglobin or in sickle cell anaemia.

Another type of repetitive regions are the tandem repeat transcriptional units, such as the ribosomal genes, which can give rise to extremely high levels or RNA in a short period of time.

One part of the repetitive regions originates from mobile elements called the transpozones. We know of at least three documented cases where it was shown that the transposon was active in humans.

In one of the cases, a hemophylic patient was investigated and the transposone was found in the gene of factor VIII, in another case, a patient suffering from muscular dystrophy was shown to contain a LINE element in the dystrophine gene. The third case is that of a patient with beta thalassemia, where it was shown that a repetitive element was inserted into the globine gene. In all these cases, the patients were investigated because of a severe genetically encoded disease. In these cases, the biological parents were also investigated and their genome did not contain the respective mutations, suggesting that the genomic change occurred in the children. Although these elements are active, we do not find their traces mainly because systematic genetic investigation is performed only in the case of genetically encoded diseases. If there is no disease, no such investigation is performed, meaning that we will not find the signs of their activity.

**Figure 5.5. Transposones**
Repetitive regions in part contain transposones. These can be autonomous or non-autonomous LINE-s, which are made of two ORF-s and are up to 8 kb length. Retrovirus like elements have LTR-s on their ends and they contain gag, pl, env sequences. Some fossils in the genome are the 2-3 kb long DNA transposases. The non-autonomous elements are the SINE elements and some truncated versions of the previous ones that do not encode the necessary enzymes for their activity but use the enzymes produced by other transposases. LINE-s work as transposons.

**Figure 5.6. Genomic variations and diseases**

Single mutations can also cause specific diseases. In the case of Waardenburg disease the PAX3 gene can be modified by insertions deletions non sense and in-frame mutations.

Genomic variants might also be responsible for some of the neoplasms. In the case of cancer chromosomal rearrangements can generate new proteins or proteins with a modified regulation that are changing the cell division in a manner that the cell will start an uncontrolled cell division and this will cause the sheath of the organism. These chromosomal rearrangements are not random but they occur in hot spots that are involved in the genomic rearrangements which occurred in the course of the course of evolution.

One part of the repetitive regions is called pseudogenes and are products of our own genome. What is a pseudo-gene? A pseudo gene is formed when the mature mRNA, which contains no introns any more, is reverse transcribed by a reverse transcriptase, probably of viral or transposon origin. The cDNA produced is corresponds to a gene and can be randomly integrated into the genome, but will not be transcribed since it will
not contain a promoter or other regulatory sequence that is needed for transcription. Due to these considerations, we call it a pseudogene. In the figure we can see that the gene encoded on chromosome X is the source of two pseudo genes integrated on the two arms of chromosome 4.

**Figure 5.7. Duchenne and Becker type muscular dystrophy**

Duchenne and Becker type muscular dystrophy

These diseases are caused by mutations in DMD gene. Since the gene itself has 79 exons, hundreds of mutations have been identified. Some of these produce truncated proteins, other mutations cause insertions and deletions. The Becker type mutations cause a less severe disease.

Repetitive sequences can be causative of other diseases too like fragile X chromosome.

In this case, the normal allele consists of no more than 30 alleles of CGG repeats. In the five prime untranslated region in the diseased patients we can find repeats in the number of hundreds or even one thousand.

A similar mechanism can be seen in Huntington's disease, where the CAG microsatellite is longer, and as a result more glutamin residues will be found in the protein that will form proteins with altered behavior and formation of precipitates in the affected neurons.

**Figure 5.8. Globin recombinations**
In the genome, we have repetitive regions that contain functional genes such as the globin gene family. Variants of the globin gene family enable the structure of hemoglobin to have, for example, two copies of alpha and two copies of beta globins.

Before birth, the gamma subtype makes possible the better oxygenation of the embryo. In sickle cell disease, a point mutation in the beta globin gene offers protection against malaria. Various types of globin genes are the result of duplications of an ancestral globin gene resulting in two copies. These were the origin of the alpha and beta globin genes. Humans have two clusters of globin genes on chromosomes 16 and 11. Each gene duplication is in fact allowing the species to develop a new function. Such duplications occurred in the course of evolution. One of the largest gene duplication event occurred in fishes, which might explain their variability in shape colour and other properties. A similar mechanism is used during plant improvement where polyploidyism is generated and after this, plants are selected for some novel properties.

What other repetitive regions are found in the genome?

In Alpha thalassemia the alpha globin gene is mutant and as a result the equilibrium between alpha and beta globin proteins is disrupted. Beta globins are becoming predominant and they form unstable tetramers with impaired oxygen transporting capabilities. On each allele, we have two copies of alpha globine genes, four in total. The severity of the disease is greater if more copies of the alpha globin gene are affected.

**Figure 5.9. Repetitive transcriptional units**
Repetitive transcriptional units.

Another type of repetitive region is the tandemly repeated transcriptional units like the genes encoding ribosomal genes. If large amount is needed of a gene at one time point it is possible that the genome is encoding it in several copies in hundreds or even thousands of copies. In the genome, we have such regions on chromosomes 13, 14, 15, 21 and 22 in such a manner that whole short arms are made of ribosomal genes.

Mutations in repetitive genes: Osteogenesis imperfecta and Ehlers Danlos syndrome

**Figure 5.10.** Mutations in genes present in several Osteogenesis imperfecta and Ehlers Danlos syndrome
Chapter 6.6. Nuclear Receptors

1. Molecular mechanisms of Nuclear Receptor driven transcriptional activation

Nuclear receptors primarily act through direct association with specific DNA sequences known as hormone response elements. These cis-elements can be found in the promoters and enhancers/silencers of target genes. Alternatively, nuclear receptors might interfere with the activity of inflammatory transcriptional factors, such as NF-κB and AP-1, in which case DNA binding is not required. (This mechanism, known as transrepression, is not indicated in the figure.) The regulatory capacity of nuclear receptors occurs through their relationship with the RNA polymerase II (Pol II) holocomplex and the chromatin environment that surrounds the regulated genes. Gene regulation is mediated by protein complexes that interact specifically with (ligand-bound and ligand-free) nuclear receptors. These proteins are called either co-repressor or co-activator proteins, depending on the transcriptional outcome. In general, co-repressor complexes (containing histone deacetylases) bind to ligand-free nuclear receptors and are displaced by activating ligands. After ligand-binding and displacement of co-repressor, co-activator proteins associate with the receptor. Cofactors can exhibit enzymatic activity, and/or function as a platform protein (e.g. SMRT), which facilitating the recruitment of other proteins. Enzymes of co-activator and co-repressor complexes can be divided into two generic classes. The first class consists of enzymes that are capable of covalently modifying histone tails, while the second class includes components of a family of remodeling complexes, which alters the structure of the nucleosomes by modifying the histone–DNA interface, and often causes nucleosome sliding. Histone modifying enzymes consist of acetylating/deacetylating enzymes (HATs and HDACs), methylating/demethylating enzymes, protein kinases and phosphatases, poly(ADP)ribosylases, ubiquitin, and SUMO ligases. (Acetylation of the histones enhance the transcription, while different types of histone methylation can be involved in both activation and repression.) Co-activator complexes typically include SWI/SNF, CBP/SRC-1/p/CAF and TRAP/DRIP/ARC. The SWI/SNF complex possesses ATP-dependent chromatin remodeling activities. The CBP and p/CAF complexes have histone acetyltransferase activities. These complexes may act in concert to relieve chromatin-mediated repression, with the TRAP/DRIP/ARC complex functioning to recruit core transcription factors. Co-repressor complexes include the SIN3/HDAC complex, which has been proposed to be recruited via the NR corepressors NCoR or SMRT. This complex possesses histone deacetylase activity and is thought to reverse actions of histone acetyltransferase-containing complexes.

Figure 6.1. Molecular mechanisms of nuclear receptor driven transcriptional regulation

Evolution of Nuclear Receptors
Most nuclear receptors are activated by a specific hormone or small lipid soluble molecule, but some do not require a ligand, and still others are incapable of activating gene expression and so act primarily as repressors. A central question concerning the evolutionary origin of the nuclear receptor family is whether the ancestral nuclear receptor was ligand-dependent or the ligand-binding evolved independently. There are two alternative scenarios for the evolution of nuclear receptors’ ligand binding ability. Based on the first scenario, first, nuclear receptors were orphans that have gained ligand binding several times independently during Metazoan evolution. In contrast, the second scenario proposes that ancestral nuclear receptor was a ligand-activated transcriptional activator that existed in the earliest period of animal evolution. According to this scenario, the ligand-independence of orphan receptors could be derived states rather than ancestral.

**Figure 6.2. Evolution of nuclear receptors**
Chapter 7.7. Personalized genetics

1.

In 2007, an oncologist, Eric Lester, M.D., from Michigan, USA used DNA microarray technology that enables scientists to examine how active thousands of genes are at a given time to analyze the expression of genes associated with positive response to anti-cancer drugs in the tumors of seven patients with advanced, incurable cancer. Then he based his drug treatment plans on the results which resulted in four of seven patients being reported to have had a better outcome than expected. This is one of the very first examples of personalized medicine.

In the last decade, personalized medicine clearly started to change the way healthcare is delivered. We are genetically different (cc. 0.5% of our genome) therefore there is a clear rationale behind the observation that people can not benefit from every drug and in not all dosages. In several medical conditions, we need the specific drug in a specific dosage that is the most suitable based on our own genetic background.

As defined by the US President’s Council on Advisors on Science and Technology, “Personalized Medicine refers to the tailoring of medical treatment to the individual characteristics of each patient to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Preventative or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not.”

As global population has recently passed the 7 billion milestone and the cost of human genome sequencing is rapidly declining, sequence data of the 3 billion basepairs of billions of people should be accessible in the very near future. Theoretically, we should be able to sequence 3 billion times 7 billion basepairs soon. Moreover, pharma companies have to switch to the more expensive, but still more cost-effective concept of personalized medicine. Another observation is the decreasing number of the new FDA (Food and Drug Administration) approved drugs each year which shows innovation is very much needed in this area.

The advance of personalized medicine in the last couple of years has been unquestionable as described by the Personalized Medicine Coalition that was launched in 2004 to educate the public and policymakers, and to promote new ways of thinking about healthcare. Today, it represents more than 200 academic, industry, patient, provider and payer communities, and they seek to advance the understanding and adoption of personalized medicine concepts and products for the benefit of patients. In their report, it was mentioned that the cost of sequencing a human genome was $300,000,000 in 2001 and it is around $5000 in 2011. This rapid and enormous decline in the cost of sequencing might lead to an era when everyone’s genomic data are available.

Moreover, 60% of all treatments in preclinical development rely on biomarker data 2 and 10% of marketed drugs inform or recommend genetic testing for optimal treatment. While there were only 13 prominent examples of personalized medicine drugs, treatments and diagnostics products available in 2006, 72 examples are known in 2011. This continuous development seems to be constant as there is a 75% increase in personalized medicine investment by industry over the last 5 years according to the Personalized Medicine Coalition.

The concept of personalized medicine requires key elements such as advanced methods for genomic analysis; extended bioinformatics with biobanks and online data storages available; and close collaboration between physician, geneticist and the patient. These elements are all needed for the proper initiation of personalized medicine.

From the genomic perspective, those medical conditions that have strong genetic background can be divided into two groups. One is for monogenic disorders in which a gene mutation leads to a disease. Examples include hemophilia or colour-blindness. The other group is for complex disorders in which several gene variants and environmental factors such as smoking, diet, pollutants, etc. can cause a disease risk, but only the risk. Genes load the gun, lifestyle pulls the trigger. These conditions include heart disease, gout, diabetes and many other complex conditions.

In the case of such conditions, genomic technologies hold great promise for the near future as by analyzing gene variants or sets of gene expression changes, we might be able to predict the risks and progression for different diseases in time to be able to prevent or modify the condition itself.
In a possible future scenario, the patient goes to the doctor, they take a peripheral blood sample, isolate DNA, sequence it with large genome sequencer centres in a few hours’ time and analyze data and determine risks for the above mentioned complex diseases in close collaboration with the physician, geneticist and the patient. This is the ideal scenario as the patient know the most about his/her medical background, family history and symptoms; the geneticist was trained to interpret the pure data of genome sequences; and the physician make the medical decision. This triad should serve as a basis for personalized medicine.

In the first years of the 21st century, several companies in the USA and Iceland were launched with the mission of analyzing DNA obtained from saliva samples sent by consumers who purchased these genomic tests online. These patients can go to the internet, order the service on the website of the company; receive the sampling package in which they have to provide a few milliliters of saliva; send it back to the lab and wait for the results of the analysis. Such companies claim to predict the risks for different medical conditions based only a few single nucleotide polymorphisms (SNPs); and also determine carrier status for metabolism-related monogenic disorders or identify sensitivity to therapies and compounds as well as visualizing the genetic archeology of the customer. The so-called direct-to-consumer genetic testing has received wide criticism and authors, in some cases, could not even compare the results of the same DNA analyzed by different companies. While the scientific background behind these can easily be questionable (the analysis might show I’m susceptible to a specific disease based on a few SNPs, but there could be new SNPs discovered next year which actually protect me from that condition, therefore the results of the analysis represents the state of science and not totally my own disease risks), the number of these companies will definitely rise in the near future.

The Human Genome Project that initiated this whole era of customized medicine was one of the largest collaborative research projects in human history. Hundreds of co-authors appeared in the paper published in Nature and Science. The Human Genome Project made the final announcement of the successful sequencing of the human genome in 2001. This project cost around 3 billion US dollars and came to a few important conclusions which seemed to be well established that time.

1) The human genome contains around 24,000 genes (now the number is somewhere between 25,000 and 30,000). 2) The genetic diversity between two individuals is about 0.1% (this number is now 0.5%). 3) Most mutations are found in men. This project was only the first step towards more sequenced human genomes and less and less expensive sequencing methods. By 2011, data of the genomes of dozens of individuals are available including Craig. J. Venter, the leader of the Human Genome Project or George Church, head of the Personal Genome Project. After the genome of Craig J. Venter was sequenced, it turned out he had 3,213,401 SNPs.

SNPs represent 90% of human genetic variations which means they appear every 100-300 basepairs in the genome. As the changes in the DNA affect how the human organism reacts to diseases, external factors such as infections or chemicals, these are used more and more often in medical research, drug developments and diagnostics.

One example for the SNP’s role in medical conditions is the association between ApoE4 and Alzheimer’s disease. 2 single nucleotide polymorphisms lead to four potential gene variants in the gene coding for Apolipoprotein E. These variants can lead to a change in the amino acids, therefore the variant E4 causes a higher risk for Alzheimer’s disease; while E2 means a lower risk. This is a good example for what kind of genomic data cannot be revealed for insurer companies or future employers.

In the Human Genome Project, 3 billion basepairs of a particular human genome were sequenced in 15 years’ time and the cost was above 3 billion dollars. By 2011, the cost of a human genome is estimated to be less than 5000 USD and this cost is going to decline rapidly in the next few years. While the Personal Genome Project aimed at sequencing first 10, then a hundred individuals’ genomes, now Chinese sequencing centers focus on the million genome project.

The data obtained by sequencing human genomes is too huge to be stored and analyzed at this point. There is no solution for sequencing 7 billion people’s genomes right now. Only one human genome, if interpreted as letters would fill 200 telephone books counting for 30 terabyte of data. But if only those genomic variants are stored that can have medical relevancy or can be used to assist in making a medical decision, it would take around 20 megabytes therefore a whole family’s medically relevant genomic data could be stored on a single CD. It represents the problems geneticists face now about the analysis of genomic data and what we can do with this huge amount of sequences in medicine and healthcare. Even if we sequence everyone’s data, we are not sure whether we can include these in medical decision making.
One of the first practical examples of personalized medicine was related to the anticoagulant Coumadin that contains warfarin. Two known SNPs modify the metabolism of this compound. A variant of the gene VKORC1 makes someone more sensible to warfarin, while the variant of the gene CYP2C9 metabolizes it faster therefore those people having these variants metabolize the drug which they are more sensible for faster than other although they all received the same drug and similar or the same dosage. This is because we are genetically different. Only in the US, Coumadin is prescribed 30 million times each year and the non-expected side effects which are due to the different metabolization rate and sensibility to the drug lead to 43,000 emergency admissions every year.

Trastuzumab (under the name Herceptin) is a monoclonal antibody that interferes with the HER2/neu receptor that is embedded in the cell membrane and communicates molecular signals from outside the cell to inside the cell controlling genes. In some breast cancers, HER2 is over-expressed, and, among other effects, causes breast cells to reproduce uncontrollably. Trastuzumab is an antibody that binds selectively to the HER2 protein and blocks the cancer cells in the breast to reproduce uncontrollably. This increases the survival of people with cancer.

In cancer cells that do not over-express HER2, Herceptin cannot bind to the cell surface, but in cancer cells that overproduce HER2, it can bind to the protein and block the uncontrollable cell growth.

Mothers who have recently given birth to their child often take painkillers containing chodein. Chodein is metabolized into morphine by the enzyme coded by the gene CYP2D6. Morphine, in very small amounts, can get into the baby through breast milk without causing any real effects. But some mothers have a gene variant in CYP2D6, therefore they metabolize the same drug in the same amount faster into morphine which is transmitted in a larger amount into the baby causing slower breathing, somnolence and sometimes death. This is only because we are genetically different.

In a few years’ time, it is going to be possible to store the genomic data of each one of us on a USB drive or a chip card similarly to our bank accounts. When the doctor prescribes a drug for the patient, they also check the genomic components, specific enzyme activities in order to prescribe the most suitable drug in the most tolerable dosage based on the patient’s own genomic data with medical history and other relevant parameters as well. Electronic medical records will certainly contain genomic data.

Genomics can never be a single area of medicine. It has to be incorporated into several branches of medicine. The Human Genome Project was the initiator and the basic point, now we have to deal with ethical, legal, social issues such as who owns the information of my DNA (I own it as the DNA belongs to me or the company owns it that makes the data of my genome available); or whether my insurer or future employer can see any of the data of my genome. With proper education, more and more resources, the developments of computational biology and bioinformatics, we should be able to make personalized medicine an integrated part of the healthcare system in which everyone gets the most suitable therapy based on their own genetic background.

Trastuzumab (under the name Herceptin) is a monoclonal antibody that interferes with the HER2/neu receptor that is embedded in the cell membrane and communicates molecular signals from outside the cell to inside the cell controlling genes. In some types of breast cancer, HER2 is over-expressed, and, among other effects, causes breast cells to reproduce uncontrollably. Trastuzumab is an antibody that binds selectively to the HER2 protein and blocks the cancer cells in the breast to reproduce uncontrollably. This increases the survival of people with cancer.

Figure 7.1. Personalized therapies
In cancer cells that do not over-express HER2, Herceptin cannot bind to the cell surface (on the left), but in cancer cells that overproduce HER2, it can bind to the protein and block the uncontrollable cell growth (on the right).
Chapter 8. 8. Immunodeficiencies

1. Figure 8.1. Duplication time of pathogens is short

The two most powerful weapons of pathogens against immunosurveillance is the generation of an enormous number of progeny with the largest possible diversity over the shortest time possible. Compared to humans (where it takes 15-30 years), all classes of pathogens including bacteria, viruses or parasites have a short duplication time. A single viral particle generates thousands of virions within a few hours. Several of them carry mutations changing surface antigens or the biochemical machinery of the virus. Thus, the human body needs to be protected (for a long time) in a fairly hostile, ever changing environment. Against the large number and diversity of pathogens, the immune system generates and deploys a large number of innate cells and also lymphocytes with receptors that in diversity match the set of the pathogens.

Figure 8.2. X linked autoimmune diseases

The two most powerful weapons of pathogens against immunosurveillance is the generation of an enormous number of progeny with the largest possible diversity over the shortest time possible. Compared to humans (where it takes 15-30 years), all classes of pathogens including bacteria, viruses or parasites have a short duplication time. A single viral particle generates thousands of virions within a few hours. Several of them carry mutations changing surface antigens or the biochemical machinery of the virus. Thus, the human body needs to be protected (for a long time) in a fairly hostile, ever changing environment. Against the large number and diversity of pathogens, the immune system generates and deploys a large number of innate cells and also lymphocytes with receptors that in diversity match the set of the pathogens.

Figure 8.2. X linked autoimmune diseases
Approximately 200 gene defects are known to cause or contribute to autoimmune disease. A rather painful approach, called positional cloning, was used to identify some of the X chromosome-linked genes inactivated in severe immunodeficiencies. Positional cloning is based on the construction of a detailed physical map of the chromosomal region in which the gene defect should reside (called the candidate region), followed by the identification of all genes in the region. The candidate region is determined by linkage analysis (close chromosomal markers tend to link together with the disease phenotype with a frequency that is proportional to the distance of the marker from the disease causing gene). The Human Genome Project greatly facilitated the gene identification project generating a nucleotide resolution physical map of the genome and an incredible number of genetic markers. With these resources at hand, the candidate region and all the genes within the candidate region can be identified. The last step of gene identification, screening for mutations in the candidate gene sequence among patients showing the symptoms of the disease, is the same for the positional cloning and the candidate approach described above.

**Figure 8.3.** NK cells are regulated by activating and inhibitory signals

Activating and inhibitory signals mediated by receptor-ligand interactions regulate Natural Killer (NK) cells.

NK cells are indispensable components of the anti-viral response. NK cell function is controlled by both activating and inhibitory receptors present in the cell membrane. Activating receptors include the NKR-P1 family of receptors, which are triggered by polysaccharide ligands. Other receptors e.g. 2B4 and its ligand CD48 also facilitate NK-cell mediated killing of virus-infected cells. 2B4 and CD48 are members of the signaling lymphocyte activation molecule family (SLAMF) that consists of nine members with diverse immune functions regulating both innate and adaptive immunity. Killer inhibitory receptors (KIR) bind to MHC molecules on target and transmit strong inhibitory signals to ensure resistance for the host cell against lysis by NK cells.

**Figure 8.4.** In the absence of SLAM-Associated Protein (SAP) adapter adaptor protein, in the NK cells of XLPPatients the activating 2B4 receptors act as inhibitory receptors
Activation signals are converted to inhibitory signals in the absence of SLAM-Associated Protein (SAP) in NK-cells of patients with X-Linked Lymphoproliferative disease (XLP).

XLP is known as a primary immunodeficiency, in which patients show extreme sensitivity to infection with Ebstein-Barr virus (EBV). However, the mechanism behind this sensitivity has been a mystery for a long time. Identification of the genetic defect has shed some light on the possible mechanism of the mysterious, exuberant EBV-specific immune response seen in XLP boys.

In healthy individuals, activation signals generated by NKR-P1 family receptors are augmented by interaction between 2B4 and its natural ligand CD48 present on the surface of the NK cell and the target cell respectively. 2B4 signals are mediated by the single SH2-domain protein SAP. The cytoplasmic tail of 2B4 contains two tyrosine based motifs that are targets for phosphorylation by Src-family kinases, Fyn, in particular. Upon phosphorylation of these sites SAP as well as Fyn are recruited and induce positive signals by a mechanism that is not completely understood.

In XLP patients the SAP protein is absent or not functional due to various mutations, thus the phosphorylated docking sites of 2B4 become available for other SH2-domain containing signaling molecules. One such molecule, SHP-1 a tyrosine phosphatase was shown to be recruited to 2B4 in the NK cells of XLP patients. Once recruited, SHP-1 can inhibit activation pathways by dephosphorylating phosphotyrosine residues on activating receptors or other signaling molecules.

In sum, the small single SH2-domain protein SAP seems to work as a molecular switch that converts 2B4 from an inhibitory molecule into an activator. iNKT cells are powerful immunoregulator cells that unlike conventional T-cells recognize glycolipid molecules.

**Figure 8.5.** The iNKT cells, contrary to conventional T cells recognize glycolipids
NK cell deficiency is only one factor contributing to the immuno-regulatory defect present in XLP. Recently, three groups have independently described the absence of iNKT cells in XLP patients.

iNKT cells harbor both the NK cell receptor NK1.1 and TCR with an α and a beta chain. Compared to regular T-cells, iNKT cells have TCRs with a limited diversity. The alpha chain is formed by the canonical rearrangement of the Vα24 gene segment to the Jα18 gene segment. The polymorphic beta chain consists of the vβ11 segment joined with various Jβ segments and they are polymorphic due to diversity in their CDR3 region.

Unlike conventional T-lymphocytes that recognize peptides in combination with self-MHC, these cells recognize the lipid receptor CD1d, in complex with self or bacteria-derived glycolipids. Upon activation, depending on the stimuli, iNKT-cells instantly produce pro-or anti-inflammatory cytokines that regulate practically all cells of innate and acquired immunity.

Production of Th2 cytokines by iNKT cells prevents Th1/Th17-induced autoimmune disease in multiple animal models including diabetes, EAE (an animal model of multiple sclerosis), DSS-induced colitis or collagen induced arthritis.

All the above shows clearly that iNKT cells are central players of immunoregulation and contribute greatly to the maintenance of immune-homeostasis as well as to the clearance of pathogens from the body.
One of the interesting early observations about cancers was that some viral infections in certain animal species could contribute to tumorigenesis. Such a virus was the Rous sarcoma virus. Chickens with rous sarcome virus infection often developed sarcomas. Moreover, the cancer showed signs of a typical infectious disease, i.e., cancer could be passed from one animal to the other. Not only could cancer be passed from one animal to the other, but it was also found that meticulous break-up of cancerous tissue did not abrogate the passage of cancer. This suggested that that the passage of the cancer was not due to a simple cell transplantation-like event, but a tiny infectious agent (the virus) was responsible for the cancer. Let’s take a closer look at rous sarcoma virus (RSV). The presence of RSV in chicken sarcomas does not unquestionably prove that the virus is the tumor causing agent.

**Figure 9.1. Cancers can be of viral origin**

When cells were infected in vitro with RSV, it was found that in vitro infection led to cellular phenotypes (such as loss of anchorage dependence and contact inhibition which together result in growing in foci rather in a single layer) commonly found in cancer cell cultures but not in healthy cell cultures. Hence, it was proved that RSV was the agent that directly caused tumoigenesis. This early observation led to the hunt for “the” virus of human cancers. It was assumed that most cancers were caused by (a group of) certain viral infection. The hunt was not successful. Now we know that the reason for the failure is that these cases in which infectious agents directly contribute to cancers are the exceptions, rather the rules.

There are two ways in which viral infections can contribute to tumorigenesis. In one mechanism, the virus carries a gene in its genome which can dramatically alter the phenotype of the infected cell in a way that results in increased proliferation and acquisition of other cancer related traits. One of the most important discoveries in cancer research have come from the study of such viruses. It has been found that the genes that these cancer causing viruses carried were counterparts of normal human genes. It meant that the ability to trigger tumorigenesis was not an inherent trait of these viruses, rather, it was an acquired phenotype. The fact that cancer causing genes were counterparts of endogenous genes revealed that these viruses perturbed a few key functions in the infected cells. These key functions are carried out by the endogenous genes in normal cells. The introduction of extra copies or mutated versions of these genes into healthy cells is what led to malignant transformations.
Very few such viruses affect humans. Human papilloma virus is such a virus. This virus carries genes in its genomes which can perturb the normal proliferation of host cells. Infections by these viruses in humans are associated with tumorigenesis, but the infection itself is not sufficient to drive tumorigenesis without additional events.

There are other viruses that are associated with an increased tendency of tumorigenesis. These viruses do not seem to contain viral versions of cellular proto-oncogenes. Why do they contribute to tumorigenesis then? During the life-cycle of these viruses, the viral genome is integrated in the genome. The integration of viral DNA into human chromosomes at random locations can have deleterious effects. Sometimes the integration of genes that are robustly expressed can perturb the normal pattern of gene expression of nearby genes. Such an event might lead to the overexpression of an otherwise normal proto-oncogene (shown in the slide). It is also
possible that the integration of viral DNA happens within the coding region of important cellular genes. In these cases, the host gene is mutated (by insertion) and its activity is lost. If such a random insertion event affects an important tumor-suppressor gene, then the resultant loss of expression can contribute to tumorigenesis.

Viral integration or chemical mutagenesis can lead to cancer because both processes are mutagenic. This reinforces our notion that cancer is a genetic disease. The genome is exposed to mutagenic effects all the time, yet cancers arise slowly and not often. The reason for this is the mutation in some key genes that are truly relevant for tumorigenesis, while mutations in the rest of the genome are irrelevant from the perspective of tumorigenesis.

What are these key genes whose integrity is important and whose perturbation leads to cancer? These genes can be divided into two major groups. One group is proto-oncogenes. These genes regulate cell growth and proliferation. If these genes are overexpressed or become over-active, an increased proliferation potential will contribute to malignant transformation. The other main group of genes is that of the tumor-suppressors. Tumor-suppressors, as their name suggests, negatively regulate proliferation and cell survival. If the expression or activity of these genes is lost, this will also contribute to tumorigenesis.
Chapter 10. Moleculer mechanisms of cancer development

1.

Let us take a look at a few selected examples which demonstrate how proto-oncogenes and tumor-suppressors function in normal and in cancer cells. Cell surface proteins that regulate proliferation are very often cell surface receptors, which can detect the presence of growth factors in the environment of cells. If growth factors are present, the receptors bind their ligand and transduce the signal into the inside of cells. Typically, upon ligand binding, they will acquire a new activity (such as a kinase activity) that will transduce the signal of ligand binding. In healthy cells, ligand binding will cause a strictly temporary signal transduction. Without further ligand binding events, the cell will not start to proliferate. This will ensure that a cell will move towards proliferation only if a persistent pro-proliferation milieu is present. In cancers, cell surface receptors are often mutated in a manner that results in a receptor molecule which will transduce a permanent pro-proliferation signal even without a persistent presence of growth factors. As a result of this, cells bearing mutated cell surface receptors will proliferate without the (normally) necessary amount of growth factors in the vicinity of cells.

Figure 10.1. Figure 10.1 The mechanism of self replicating proliferation

Another mechanism which can contribute to increased cell proliferation in cancer cells is when cells become self-sufficient in providing a milieu required for proliferation. Cells that produce one type of growth factor usually do not express the receptor for the given growth factor. This is important, since in the presence of the receptor the cell could drive itself into an autocrine proliferative cycle. In some cancers, such autocrine cycles are created by an altered expression of growth factor receptors. In some cancers cells that produce TGF-β, a powerful growth factor also express the cognate receptor, EGF-R. This will make cells become responsive to TGF-β and will drive proliferation.

Figure 10.2. Figure 10.2 Mechanism of autocrine growth
Tumor-suppressors are also mutated in cancers. Tumor-suppressors usually do not regulate cell growth but act on other critical pathways. Such a pathway is apoptosis. Normal cells possess a potent mechanism to reign in unwanted proliferation. With the help of apoptosis, cells that are stressed and possibly diseased can commit suicide. This is a powerful counter-mechanism antagonizing proliferation. Cells with intact apoptotic pathways cannot become fully malignant. In cancer cells, the pathways of apoptosis is almost always impaired. This ensures that cells that proliferate at an increased rate will not die.

**Figure 10.3. Figure 10.3. Is p53 an oncogene or a tumor suppressor?**

One of the main tumor-suppressors in humans is p53. This protein is a regulator of apoptosis. When cells are exposed to stress that endanger the normal functions of the cell and might have long lasting consequences (such as genotoxic events), cells stop proliferating and allow themselves enough time to repair damages. Once the damages are corrected/repaired, cells will ensue proliferation. If the damages are serious enough to be irreparable, the cell dies by apoptosis.
Figure 10.4. Figure 10.4 Structure and function of p53

P53 is a key regulator of the decision making during stress. In the absence of p53, cells will not stop for repair. Instead, p53 deficient cells proliferate even when their genome is seriously damaged. This leads to a high degree of genetic instability in p53 deficient cells. Moreover, cells whose functions are seriously perturbed will not die by apoptosis, if p53 activity is lost in cells. These features make p53 a very powerful tumor-suppressor and will lead to increased tumorigenesis in p53-/- animals.

Figure 10.5. Figure 10.5. A p53 és apoptózis kapcsolat
Chapter 11.11. Obesity: introduction

1. Obesity is a modern endemic. This is a well-known fact, but little is known about the root of the problem, namely that younger and younger children are affected by this disease every year.

Similarly, it is known that obesity can cause several diseases from diabetes to hypertension, but it is extremely important to understand that obesity is a disease in itself. As any other disease, it can be classified, and needs medical treatment.

Obesity is also a disease of a distorted of perception of the body or body-scheme. It is the opposite of anorexia, when the patient feels herself obese with normal body parameters and as a result a pathologic fasting is destroys normal physiological processes in a multi-organic fashion. The patient has a distorted perception of his/her own body scheme.

Obesity is a disease that affects large populations. Hungary is strongly affected by this problem. For a deeper understanding of the underlying mechanisms of obesity, let us see what the body is made of.

On tissue-level, in the body of a healthy middle aged man the most abundant tissue is the muscle and right after that comes the adipose tissue. If we investigate on cell compartment level, after water, the largest pool of molecules is not that of the proteins and enzymes, but the lipids.

Obesity is controlled by quite simple mechanisms. For this we have to understand the regulation of appetite, which is performed in the nucleus arcuatus, involving two types of neurons: orexigenic and anorexigenic ones.

In the course of a year, a person consumes hundreds, or even thousands of kilograms of food. If there is a slight deregulation in the controlling appetite and body weight, this can have severe consequences. Imagine a 20-kilogram increase or decrease in weight per year for a person who is not overweight. This might be a sign of a serious disease and could have severe consequences.

Orexigenic and anorexigenic neurons together regulate the appetite: by comparing our self-perception to our actual body, our appetite is increased or decreased accordingly. The inputs of this system are important and we can differentiate basically among four types of input. Leptin, released by the adipose tissue is a signaling molecule that translates the total level of adipose tissue to the neurons. Leptin provides the input to a so called adipostat. The adipostat regulates the total level of adipose tissue available in the body. In obesity the adipostat is set to a higher value and the patient will do everything to maintain its adipose tissue levels intact. The mechanism works in a similar way to a thermostat. If it is set to a higher temperature, you will overheat your apartment and thereby consume a lot of energy. The energy waste increases in proportion to the increase of the external temperature.

Another protein involved in the regulation of the appetite is insulin. Insulin is decreasing body temperature. If we eat and the blood sugar level increases, our appetite will decrease.

Another two important signaling molecules are Ghrelin, released from the stomach, and PYY, released from the colon.

As a general principle, we can state that weight gain and weight loss follows the laws of thermodynamics. Energy input has to equal the consumption to be in a steady state level, thus, any change can produce weight gain or weight loss.

Figure 11.1. Obesity: environmental cultural and internal factors regulate energy input. Energy input and consumption has to be in balance
Obesity: environmental and social effects. Balance of energy input and output. Weight increase obeys the laws of thermodynamics. The balance of energy intake and energy output produces increase or decrease in weight. The energy output consists of 3 components:

1. basic metabolism;
2. TEF: thermic effect of food;
3. energy consumed by physical exercise.

Energy consumption consists of several components. Basal metabolism plus TEF (thermic effect of food) and physical activity together represent the total energy consumption. The energy provided by food is delivered in the form of three different types of molecules: sugars, lipids, and proteins. Other factors also influence the energy balance. Energy consumption is proportional to the total amount of lipids. Lower sympathetic activity favors weight gain, while beta agonists can increase heat production through the UCP proteins. The transcription factor PPAR gamma is responsible for terminal differentiation of adipocytes and lipid accumulation.

The energy is supplied to the organism in the form of 3 type of molecules: proteins, carbohydrates, and lipids. The energy balance is related to other factors as well. 1. The energy consumption is directly proportional to the amount of available triglycerids. 2. Lower sympathetic activity is causing weight increase. 3. Beta agonists increase the release of heat through UCP proteins. 4. The transcription factor PPAR gamma facilitates the terminal differentiation of adipocytes.

Twin studies have shown that from all cases of obesity, 60% is not genetically inherited, 30% has a genetic component, and 10% is culturally determined. In the genetically determined overweight cases, the prevalence of extreme obesity is more frequent. The non-genetically determined overweight cases are more commonly associated with increased appetite because of indirect effects. From the point of view of molecular medicine, these cases are most relevant.

Twin studies

It has become evident from twin studies that 60% of overweight cases are not genetically inherited, 30% are genetically inherited, while 10% show cultural transmission. In the genetically determined overweight cases, the prevalence of extreme obesity is more frequent. The non-genetically determined overweight cases are more commonly associated with increased appetite because of indirect effects. From the point of view of molecular medicine, the cases with genetic transmission are of special interest (stress and psychological disorders).

**Figure 11.2. Figure 11. 2 Twin studies and the genetics of obesity**
Twin studies

- Genetic: 30%
- Cultural transmission: 10%
- Non-transmissible: 60%
Chapter 12. 12. Genomics of obesity

1.

To quantify obesity, we can consider several methods. The most basics ones are the calculation of BMI (weight(kg)/(height(m) on the power of two)), filling in questionnaires, and more complex ones, such as assessing fat distribution by CT or MR, metabolic profiles, molecular markers. Their price increases with their complexity, but the relevance of the information they provide is also increasing.

Some of the genes involved in the development of obesity were clearly identified, such as leptin, leptin receptor, glucocorticoid receptors, pargamma, UCP proteins, adrenergic receptors.

The identification of these genes has been made possible by population genetics studies, where we compare two cohorts of samples that are highly identical except for the investigated disease.

If these groups of patients are available, by using the appropriate methods, we can identify the genomic regions where differences can be seen. These differences, if consistent, can be causative of the disease. The next step is to dig into these regions and identify the genes affected that might be responsible for the development of the disease. The larger the cohorts are, the more likely it is that we can identify relevant genomic variations for the studied disease. If we study individual parameters like SNP-s, tens of thousands of samples are needed. If the sampling is more complex and we can compare other parameters beside SNP-s, such as gene expression profiles, the number of samples can be reduced to a few of hundred.

Such studies can sometimes generate surprising outputs. One of the sodium channels was identified as a gene being related to the obese phenotype. How is a sodium transporter linked with obesity? It turned out that this is a triptophane sodium co-transporter and tryptophane is needed for the production of serotonin, which in turn will regulate the appetite.

Through these kinds of investigations, several regions of the genome have been identified as being involved in obesity and some of these are linked specifically with extreme obesity.

Normal BMI (body mass index) is between 18.5 and 24.9. Below 18.5 we speak about leanness above 25 about overweight. A second relevant parameter is waist circumference. Higher than 102 cm for men and 88 cm for women are the two values that are showing increased health risk. Another relevant parameter is the location of the adipose tissue in the body. Several studies have shown that subcutaneous adipose deposits are less problematic than the deposits in visceral fat. Increase in visceral fat definitely increases health risk. This is relevant information, since some drugs, like the PPARgamma activators, might increase body weight as a side effect but they still do not increase from this point of view the risk of disease, since it is only subcutaneous fat that has been increased, adipose fat has not. They might increase the risk of disease in terms of other problems, such as edema and cardiac failure as well.

There are various afferent signals, such as gustative signals, leptin, insulin etc. These signals are processed in the nucleus arcuatus in the hypothalamus and the neurotransmitters are histamine, epinephrine, serotonin, dopamine, norepinephrine. Some peptides increase appetite like Neuropeptide Y, AGRP or agouti related peptide, ghreline, orexin and the opioids. Some neuropeptide decrease appetite like alpha MSH, cocaine, amphetamine, corticotropine. There are various efferent signals, such as growth hormone, testosterone, estrogen, and besides other effects, they are able to influence the distribution of the lipid deposits in the body. Glucocorticoids, insulin control the differentiation of adipocytes and adipogenesis.
Chapter 13. 13. Genes involved in development of obesity

1. The best known animal models for obesity are the leptin and leptin receptor deficient mice and rats.

The role of leptin and leptin receptors has been clarified in parabiosis experiments. These experiments are not used any more due to animal welfare reasons. Ob/Ob or Db/DB mice were connected in their circulations to wild type mice. The observation of the change in body weight of the wild type and mutant animals due to the increased leptin levels (in leptin receptor mutant) or mutant type of leptin clarified the role of leptin and the role of the two different proteins in leptin signaling.

Leptin has been known as the obesity gene, its role being to translate the amount of fat deposit to an interpretable signal by the central nervous system. If the leptin molecule or the leptin receptor are mutant or absent, in both cases the animals will start increasing in weight and will become extremely obese. This mouse is the Ob/Ob mouse.

Leptin can be used in humans for the therapy of the leptin mutant type of early onset obesity. A child with leptin mutation at the age of 3 was 42 kg-s while at the age of 7 after leptin therapy he was only 32 kg-s.

Leptin deficiency is a rare disease which causes extreme obesity.

Some facts about leptin: it is produced in the adipocytes and placenta. Women have higher levels of leptin. More adipose tissue means more leptin. Receptors are located in the CNS, ovary, liver, and hematopoietic stem cells. Leptin deficiency causes obesity. Leptin can be found in mother’s milk. Leptin levels at birth are connected with the menarcheal age. Choriocarcinoma can secrete leptin. Low levels of leptin predict obesity.

The role of UCP-s and thermogenesis

The energy balance is related to other factors as well. After a meal, heat is produced in the organism by the UCP proteins and with a similar mechanism betaagonists also increase thermogenesis. What happens if we create genetically modified mice that lack the UCP proteins? We have three UCP proteins. If these are knocked out, the mice will not become fat. The UCP1 knock-out mouse is very heat-sensitive. That means if a wild type mouse is put in an environment of 4 degrees for an overnight period, it will survive. On the contrary, the UCP1 KO mouse will die. The UCP2 KO will not become fat either, but it is showing an increased level of reactive oxigen species. UCP3 has a similar phenotype.

Figure 13.1. The role of UCP-s and thermogenesis
Some other animal models

Other models include the MC4R receptor mutants PCI mice, etc. Agouti mice are mutant in their melanocortin receptor. Some animal models are resistant to obesity, like the FOXC2 overexpressed mouse. This mouse is producing more heat than the control mouse and this is the cause of its protection against obesity. A similar phenotype can be seen in the Perilipin KO mouse, with the difference that in this case it is not heat production, but lipolysis that is increased.
Chapter 14. Therapeutic approaches to obesity

1.

How can we treat obesity medically? Our basal metabolism is responsible for 70% of our energy consumption, 15% is used for locomotion and 15% by thermogenesis. Of these three components, exercise can be increased very easily. The two other components can be modulated by medication. Depending on BMI, different strategies have to be considered. If BMI is higher than 40, surgical treatment might be necessary.

Medical treatments are not a real success story. One of the first therapy was the thyroid extract in 1893. In 1921 dinitrophenol was used but it caused cataract. In the 1930s, amphetamine was used but it caused severe addiction. In the 1970s, the so called Rainbow pill was introduced, which contained diuretics, digitalis, dinitrophenol and amphetamine, but caused sudden death. Later on, aminorex was introduced but this caused pulmonary hypertension. Very low caloric diets also caused death. Phentermine caused valvulopathies, phenil propanolamine caused stroke in women if they stopped taking it.

Some of the currently used drugs are reducing appetite in the CNS or others block the absorption of lipids (pancreatic lipase blockers.)

We have to mention the importance of obesity as a side effect of some medications. Of special interest, beside PPARg activators, steroids, antidiabetics and thyroid blockers, are antidepressants, antiepileptics and neuroleptics.

The first operation for the treatment of obesity was performed in 1954 as a jejuno-ileal bypass. Today, laparoscopic gastric bandages are the most frequently used surgical interventions for the treatment of obesity (not to mention plastic surgery).
Chapter 15. 15. Interconnected mechanisms in lipid metabolism

1. The transcription factor PPARgamma is involved in the differentiation of adipocytes and well-established cellular models are available to study this differentiation pathways. With these models, high throughput screens can be performed, where hundreds or even thousands of ligands can be screened to assess their role in adipose tissue differentiation, whereby one can estimate if these ligands as drugs cause weight gain or not as a side effect of the adipose differentiation.

The molecular mechanisms of adipocyte differentiation

The PPAR gamma is a well-characterized transcription factor and its role in adipocyte differentiation is well-known. There are established cell models of adipocyte differentiation, such as 3T3L1 mouse fibroblasts or embryonic stem cells. Immature adipocytes can be differentiated from fibroblasts adipoblasts, preadipocyteadipocyte by an endpoint of differentiation in the adipocyteadipocyte, which is filled with lipid droplets. In such a cell model several tens, hundreds or even hundreds of thousands of ligands can be tested in order to see if they influence adipocyte adipocyte differentiation, and thereby to establish if they increase weight if administered as a drug.

Figure 15.1. Figure 15.1 Molecular mechanisms of adipocyte differentiation

PPARgamma is modulating the metabolism through several tissues. Since PPARgamma is a transcription factor that forms heterodimers with RXR, it can directly cause changes in gene expression. The three main tissues where it is influencing metabolism are the hepatic, the adipose tissues and the muscle. Other effects of PPARgamma include changes in the tolerogenic activity of immune cells. Synthetic PPAR gamma ligands are administered in type two diabetes due to the fact that they reduce the insulin resistance of the peripheral tissues, mainly the muscles. Another effect of them is the facilitation of terminal differentiation of adipocytes. This effect is producing weight gain in the subcutaneous adipose tissues and redistribution from visceral fat to subcutaneous fat.

Figure 15.2. Figure 15.2. PPARgamma is modulating metabolism in several tissues
Through CD36 and ap2 or FABP4 it is increasing the production of triglycerides. In hepatocytes PPARgamma is increasing beta oxidation in mitochondria and peroxisomes.

PPARg and lipid uptake

Cholesterol is transported in the circulation in the form of lipoprotein complexes. LDL is able to leave the circulation and is responsible for the cholesterol supply of the peripheral cells through the LDL receptors. In some of the cases, the lipids of the LDL are present in an oxidized form, because of smoking or because of the fried oils of dietary origin. In these cases, the LDL is forming oxidized LDL, which is taken up by scavenger receptors and by CD36. Oxidized LDL contains PPARg ligands, which, if activated, will further increase the production of CD36, whereby a self upregulated loop is initiated, which will lead to the filling of the macrophages with cholesterol, the formation of foamy cells and the death of the macrophage with a local inflammation. The result is the formation of the atherosclerotic plaque.

Figure 15.3. Circulation of cholesterol
There is a nuclear receptor driven pathway that is able to export the cholesterol from the cell. The PPARg agonists from the oxidized LDL are turning on the expression of LXR. If LXR ligands are present, LXR will turn on the synthesis of ABCA1 the main ATP driven cholesterol pump of the cells. As a result, cholesterol is pumped into the HDL molecules and transported back to the liver.

**Figure 15.4. Figure 15.3. PPARg and lipid uptake**

LXR will turn on the synthesis of ABCA1 the main ATP driven cholesterol pump of the cells. As a result cholesterol is pumped into the HDL molecules and transported back to the liver.

Cholesterol has a very low water solubility, in the circulation it is transported in the form of lipoprotein complexes. Some of the lipoprotein complexes are quite large and involved mainly in intravascular transport, like chylomicrons and VLDL other, smaller complexes like LDL and HDL can leave the circulation and are involved in direct transfer from and to the cells. Chylomicrons can be conceived as huge lipid droplets that leave the enterocyte and arrive into the lymph. After initial processing, they get into the circulation and the chylomicron remnants are taken up by the hepatocytes. In the hepatocytes, the lipids are repacked into VLDL. VLDL is transformed into IDL and LDL in the circulation and LDL leaves the circulation. HDL is produced mainly in hepatocytes and to a lesser extent in enterocytes. As an empty complex arrives at the periphery where it is filled with cholesterol by the ATP dependent transporter ABCA1. The function of HDL is the transport of retrograde cholesterol from the periphery to the liver, the supply of cholesterol of the endocrine organs and the moving of lipoproteins between the lipoprotein complexes in the circulation. Bile acids produced in the liver arrive at the intestine with the bile and they are involved in the emulsification of lipids in the intestine and thus in the uptake of the lipids in the intestine.

**Uptake of cholesterol into the cells**

Cholesterol is taken up by the cells in the form of LDL. The highest number of LDL receptors are on the hepatocytes and the expression of LDL receptors is controlled by the level of cholesterol in the cells through the cholesterol sensing molecule SREBP1. The first step of the uptake is the binding of LDL by the LDL receptors expressed on the cell surface. Mutations in the LDL receptors are causing an increased level of circulating cholesterol. In the case of homozygotic form of the mutation, the patients will have extremely high levels of cholesterol and atherosclerosis at a very young age, which can manifest itself even as myocardial heart infarction. In this case, the single chance to prolong the patient’s life at present is liver transplantation. The new liver will have healthy LDL receptors whereby it will be able to remove the LDL from circulation.

**Figure 15.5. Figure 15.4. Uptake of cholesterol into the cells**
The LDL is taken up through coated pits and it arrives in lysosomes where the proteins are hydrolyzed and the cholesterol enters the cell membranes. LDL receptors recirculate every 10-20 minutes. Their half-life is in the range of 10 to 30 hours and through their life cycle they are involved in the uptake of hundreds of cholesterol molecules.

Through this property of theirs, they can play a role in regenerative medicine, but, more importantly, through this feature we can have insight into the details of differentiation pathways. This knowledge can be used for the differentiation of the induced pluripotent cells (iPS) as well. iPS cells have two other very important properties. First: they are immunologically identical with the organism which they are originating from and through this they stand a good chance of being used in regenerative medicine applications.

Second: because they are genetically identical with the organism they originate from, they can serve as a source of cellular models of these genetic diseases, whereby we can gain an insight into the molecular details of these diseases.

In the figure, we can see the differentiation steps of embryonic stem cells. In the first step ES cells generate mesenchymal stem cells under the control of BMP and FGF signals. Mesenchymal stem cells are able to differentiate into myoblasts, adipoblasts and osteoblasts. From this step on, tissue specific transcription factors such as MyoD+Myogenin, PPARg+CEBPa, or RunX2+Osx are involved in the terminal differentiation of the previously mentioned cell types.

**Figure 15.5. Differentiation of embryonic stem cells**
Embryonic stem cells are able to differentiate along all differentiating pathways.