Pharmacogenetic significance of non-coding gene variants in Hungarian and Roma population

PhD thesis

Dalma Varszegi, M.D.

Department of Medical Genetics

Medical School, University of Pecs

Program leader: Bela Melegh, M.D., Ph.D, D.Sc.

Pecs, 2014

1. INTRODUCTION

The growing number of genomic sequence, the description of genetic information from different species provides useful data to understand non-coding genes region evolutionary importance. When a mutation in the human genetic background is looked by researchers, they focus on the genes and their neighbouring control elements. The results of recent year supported, that the non-coding regions of genes have important role in the development of several diseases, they can causes genetic abnormalities, they may predispose to disorders or on the contrary, they can be protective agent too.

1.1. The genetic background of *IL23R* gene

Interleukins (IL) take part in development and activation of many cell. These molecules play a crucial role in the maturation and differentiation of cells. The name refers to the establishment of interactions between leukocytes.

The interleukin-23 (IL23) which belongs to the IL12 familiy is primarily intracellular connection mediator glycoproteins with low molecular weight, which play important role during immune response in the transmission and regulation of the information. It was identified in 2000; it has a heterodimer structure, where its subunits are linked trough disulfide connections. IL23 is composed of the p19 domain, which is typical only for IL23; and of p40 domain, which is also a part of interleukin-12 (IL12).

The IL23R tasks are very diverse. In chronic inflammation, antigen-stimulated dendritic cells and macrophages produce the IL23, which affects naive $CD4^+$ T cells and promotes the development of Th17/Th_{IL-17} cells. These cells produce IL17, which enhances T cell priming and triggers potent inflammatory responses by inducing the production of a variety of inflammatory mediators.

Cells expressing the receptor complex built from IL23R and IL-12Rβ1 are able to respond to IL23. The human IL23R gene locates on the short arm of chromosome 1 (1p31.3). The basic form of IL23R is coded by 11 exons; however through alternate splicing at least six different isoforms can be generated (IL-23R1-6). The most common deletions are deletion of exon 7 and/or 10. These variations result early termination of the protein leading to different forms of the extracellular domain, or to frameshift of the open reading frame, resulting different length of intracellular domains.

In a genome-wide association study Duerr et al found strong association between Crohn's disease and IL23R polymorphisms, furthermore between IL23R gene and polymorphism of the neighboring intergenic region of the IL-12Rß1 gene. The authors published ten different single nucleotide polymorphisms (SNP), which represented strong significant association with this type of inflammatory bowel disease. Five polymorphisms, the rs10889677 in the 3'-untranslated region (3'-UTR), the intronic rs1004819 and rs2201841, furthermore the intergenic rs11209032 and rs1495965 were found as susceptibility factors in the development of the disease. However some others were found to be protective against CD: Arg381Gln (rs11209026) in the cytoplasmic domain, the intronic rs7517847, rs10489629, rs11465804 and rs1343151 mutations.

Not much later, relationship between several autoimmune diseases and the examinded IL23R polymorphisms were found. The most studied diseases: psoriasis (PS), Sjögren's syndrome (SS), ankylosing spondylitis (AS) and systemic lupus erythematosus (SLE).

1.2. The genetic background of *IL28B* és *IL10R* genes

IL-28B belongs to the interleukin 10 familiy. It is expressed by peripheral blood mononuclear cells, dendritic cells, and hepatocytes upon infection with viruses or stimulation with double-stranded RNA. IL28B in turn activates signal transduction through the JAK-STAT pathway, exerts antiviral activity and has an impact on natural clearance of HCV. IL28B exhibits fewer IFN-like adverse effects because IL28B receptors are expressed on a limited number of cell types. The receptor for IL28 is composed of a unique IL-28R-α chain that pairs with the IL-10R-β chain.

IL28B gene on human chromosome 19q (19q13.13) was discovered using genomic screening process in which the entire human genome was scanned for putative functional variants. It consists of 5 exons.

Ge et al. identified first a SNP (rs12979860) located only 3 kilobases upstream from the IL28B gene, that encodes IL-28B. They applied genome-wide association study (GWAS), retrospectively analysed samples of 1,137 HCV infected individuals participating in a clinical trial with pegylated interferon plus ribavirin (P/R) therapy, and found that the high IL-28B production CC genotype was associated with approximately a threefold greater rate of SVR to P/R when compared with the TT genotype. This finding was confirmed by others. At the same time it was also shown that the frequency of the CC genotype was significantly lower in HCV patients when compared to matched controls; thus, suggesting that this variant may be linked to a higher rate of natural clearance of HCV. This hypothesis was soon strengthened by Thomas et al., who reported that the IL28B CC genotype enhanced the spontaneous resolution rate of HCV infection. Thus, it became evident that HCV patients who harbor the CC genotype at rs12979860 are more prone to respond to P/R treatment and to clear the virus than patients who do not possess this genetic polymorphism.

Interleukin-10 (IL10), whic belongs to the interleukin 10 family is produced by monocytes, macrophages and T cells, inhibits both the activation of CD4+ T-helper cells and the function of cytotoxic CD8 + T, NK and antigen-presenting cells, and also modulates hepatic stellate cell collagen synthesis. IL-10 plays a regulatory role in immune reaction and suppresses inflammatory responses by inhibiting the production of pro-inflammatory cytokines. Its effect is mediated through the IL-10 receptor (IL-10R), which is a heterodimer that consists of both IL-10R1 required for binding, and IL-10R2 required for signaling.

The human *IL10R* gene is located on chromosome 11 (11q23), and it consists of 7 exons.

Increased IL-10 production was reported in association with the -1082G, -819C, -592GCC IL-10 promoter (ATA) haplotype. If macrophages secrete large amount of IL-10, it may decrease circulating TFN-α and IL-6 levels, thus, reducing their harmful effects. A higher frequency of IL10 -1082 GG genotype was found in older healthy controls than in patients with myocardiac infarction; high IL-10 production was protective for longevity.

1.3. The genetic background of *PRDM1-ATG5* genes

The human *PRDM1* (positive regulatory domain I protein) also called *BLIMP1* (B lymphocyte-induced maturation protein 1) gene is a a protein, which functionate as a repressor transcription factor. It play an important role in the regulation of different cellular processes such as cell proliferation, differentiation and apoptosis. It participates the final specialization of lymphocytes, epidermal cells, stem cells and other cell types, and it can be associated with the pathogenesis of human lymphoma.

The transcript of the *PRDM1* gene results two dissimilar mRNA types forming two different isoforms: the PRDM1 alpha and beta which share different types of cells. The

PRDM1 alpha is larger and more often expressed isoform, it contains 825 amino acids with a molecular mass of 92 kDa. The N-terminal part of the protein contains PR domain, while the C terminal region comprises five C2H2 type zinc finger, which provides the DNA binding domain. In the middle of this we can find the proline/serine-rich region. In contrast of this, the PRDM1 beta is shorter on the N-terminal side with 101 amino acid, thereby the PR domain is truncated. The PRDM1 beta transcriptional repressor functional activity is reduced.

The PRDM1 expressed in many different tissues and organs, such as the hematopoietic system (plasma cells, B and T lymphocytes, NK cells, monocytes, granulocytes and dendritic cells), skin, central nervous system, testis, and intestine. In B cell population the expression will begin when it commits themselves to the final differentiation. It expressed in normal lymphoid cells, multiple myeloma and different lymphomas, which are under regulation of several transcriptional activator and repressor. Pasqualucci and his collegaues established that *BLIMP1* operates as a tumor supressor gene, which inactivation leads to lymphoma's developing. The two examined SNPs (rs4946728 and rs1040411) are located between the *PRDM1* and the *ATG5* (autofágia-related 5) genes in non-coding region.

The *ATG5* gene is one of the genes, which plays an essential function in the autophagosome late endosome and lysosome formation. The human ATG5 gene is located on the long arm of the 6. chromosome (6q21), it consists of eight exons (www.ensembl.org). Az The conjugation of ATG5-ATG12-vel is a key regulator of the autophagy process of natural anti-viral immune response.

1.4. The genetic background of *FCER2* gene

The human *FCER2* gene is located on chromosome 19 in position 19p13.3. This gene is responsible for the encoding FcεRII /CD23 (low-affinity Immunoglobulin E receptor) integral membrane glycoprotein. The protein product of FCER2 gene is a 45 kDa protein which consists of 321 amino acids.

The FcεRII receptor mediates various cellular biological processes; such as protection from apoptosis, the release of cytotoxic mediators and cellular adhesion. The principal role of the receptor is the synthesis of IgE down-regulation (through B lymphocytes), further it is responsible for the antigen presentation, maturation of T and B lymphocytes Proteolytic cleavage results FCER2 soluble form, which makes it unable to bind to the cell surface, and thereby terminate the negative feedback inhibition mechanism, resulting in increasing IgE levels. Presumably, this process will form the basis of various allergic reactions.

The rs28364072 (T2206C) polymorphism in *FCER2* gene is associated with elevated IgE level, severe asthmatic exacerbations and decreased gene expression. This polymorphism may play a role in the ineffective drug treatment because decreased therapeutic response was experienced using inhaled corticosteroids in patients carrying 2206C variant of FCER2 gene.

1.5. Autoimmune disorders

Ankylosins spondylitis (AS)

The ankylosing spondilitis also known Bechtrew-disease is a chronic, inflammatory and progressive disease of the spinal and sacroiliac joints, especially characterized with calcification of the joints and ligaments. As a consequence of these results a rigid spine. The peripheral joins are affected in the one third of the cases. It can develop in 20 to 30 years, both sexes are afftected; however it is more common in man, than women. As the joint surfaces die due to the inflammation, the joins calcificates, the vertebrae begin to grow together, and beaklike lumps are formed on their edges. SPA may affect the entire body. It can develop extraintestinal manifestation, the most common is the uveitis anterior, that causes continuous inflammation of the eyes, sensitivity to light, increased tear production is also specific for the disease.

Systemic lupus erythematosus (SLE)

The SLE is an autoimmune disorder with chronin inflammation, the symptoms can affect the whole body system. In the background of the disease are several factors, genetic, immunolological, environmental and hormonal. Known risk factors of ultraviolet radiation, smoking and a number of environmental toxins. In males and females occur both but in women of child-bearing period most often, when the hormon system changes. General symptoms include fever, fatigue, joint pain, loss of appetite and weight loss. The skeletal system, skin and respiratory system is primarily affected. The redness may occur in the face, which form a characteristic feature in lupus "butterfly" rash. Manifestations are furthermore: sores around the mouth and nose, the Raynaud syndrome and alopecia. Muscosceletal symptoms are arthritis in the muscles, and pain in the joints. Respiratory symptoms are cough, shortness of breath. When the cardiovascular system is affected pericarditis, miocarditis, endocarditis, when the nervous system is affected, headaches, depression, anxiety, stroke, when the excretory organ system is affected glomerulonephritis can occur.

Sjögren’s syndrome (SS)

Sjögren's syndrome or Sjögren syndrome is a chronic autoimmune disease in which the body's white blood cells destroy the exocrine glands, specifically the salivary and lacrimal glands, that produce saliva and tears, respectively. The immune-mediated attack on the salivary and lacrimal glands leads to the development of xerostomia and keratoconjunctivitis sicca, which takes place in association with lymphocytic infiltration of the glands. That inflammatory process eventually severely damages or destroys the glands. Common symptoms of burning, itching and foreign body sensation in the eye, dysgeusia, tongue edge cracking. The SS is an multifactorial disease affecting origin, such as environmental, hormonal and genetic factors in a cross. Genetic predisposition study revealed that in individuals with genetic predisposition a viral or other infections can be the cause of the disease.

Psoriasis (PS)

Psoriasis is a genetically determined, polygenic inheritance skin disease, associated with increased proliferation of keratinocytes, paraceratosis, dermal inflammation, vascular and immunological abnormalities and T cell proliferation in the skin. Recent research emphasize the role of IL17 and IL23 cytokines. The inhibition of these treatment options come in through. The disease has polygenic inheritance, but enviromental factors are also important in the development of the disease, stress, triggering climatic conditions may be included as a factor in subjects with a predisposition to the disease. In the beginning, just at certain predilection areas show wax-white, stout and flaky deposits covered papules, this can remind us to dried candle wax. In other cases, especially in pediatric cases, millet size, in the later phase scaly, small and bright red papules occur on the whole body (psoriasis eruptiva, psoriasis guttata). Predilection sites of the symptoms are the surfaces areas exposed to irritation: elbows, knees, scalp, areas under the breasts, genitofemoral bend, perianal area and the external genitalias. The PS can develop in any ages, but there are stages of life, when the disease is more common (in 15 to 20 and in 55 to 66).

Inflammatory bowel disease (IBD)

The pathogenesis of IBD is very complex and both environmental and genetic factors contribute to its etiology. Environmental factors may be infectious agents, drugs, poisons, smoking, alcohol, and bowel bacteria as well. The genetic background has a particular importance, people may have genetic susceptibility and show reduced resistance of the mucosa or abnormal activation of the immune system in the intestines. These together with environmental factors cause inflammation. IBD may develop in a susceptible individual when the normal hostmicrobial interactions are dysregulated. The normal host-microbial flora consists of 300-400 different strains of bacteria. This balance can overturn, when people have chronic inflammatory bowel disease. This time the concentration of certain bacteria, such as *Bacteroides* increases causing inflammation. External bacteria can be also responsible for IBD.

There are two main types of inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC). 25-30.000 people in Hungary have IBD. The disease may occur in persons adults between ages 20 and 40 in both sex, although women are more frequently affected. There are many extraintestinal manifestations associated with inflammatory bowel disease, for instance eye and mouth inflammation, skin rash, gallstones, kidney stones, pancreatitis and joint pain.

Crohn’s disease is a chronic inflammation of uncertain etiology that can affect any portion of the digestive tract from mouth to anus. Inflammation occurs primarily in the ileum and colon, although any portion of the intestinal tract can be affected. Ileocolitis is the most common form affecting the ileum and the colon. Ileitis only affects the ileum, while granulomatous colitis only affects the colon. The intestinal inflammation can appear in the terminal ileum (terminal ileitis) or might occur in some parts of the ileum (regional enteritis) or the colon (Crohn-colitis). Symptoms include abdominal pain, diarrhea, bloody stools, fistulas, fever, loss of appetite, vomiting, weight loss, growth retardation, indisposition and late of puberty.

Ulcerative colitis is a disease that causes inflammation and ulcers in the lining of the rectum and colon. When the inflammation occurs in the rectum it is called proctitis. If the entire colon is affected it is called pancolitis. Proctosigmoiditis involves inflammation of the rectum and the sigmoid colon, while left–sided colitis involves inflammation that starts at the

rectum and extends up the left colon (sigmoid colon and the descending colon). Symptoms range from abdominal pain, cramps, bloody diarrhea, tenesmus, fever, anemia, loss of body fluids and nutrients, to decreased level of protein and iron.

Rheumatoid arthritis (RA)

It is a disease with unknown etiology, affecting the joins with chronic inflammatory, where genetic factors may contribute to the development of the disease. Bacterial or viral infection is presumed (Myobacterium, Proteus, Streptococcus, Chlamydia species, parvovirus B19, rubella, HIV viruses).

In systemic, autoimmune diseases, the body treats theirs own connective tissue elements as foreign, defending against them with inflammatory reaction. During this process, the intima of the joint thickens due to the inflammation and synovial fluid builds up. All these prevent the normal function, preventing the movement of the joint. Muscle spasms occur and at a later stage the angled form of the joint will be stabilized. During inflammation different lesions occur: degradation of the joints and surrounding areas, tendonitis is very common, rheumatoid nodules may appear in the skin, and different bone deformations may also develop. After disease manifestation, painful joint distorsion of the connective tissue or bony joint stiffness develops in the later stage.

Based on epidemiological data, RA is a very common disease, it affects 0.5-0.5% of the population, in Hungary approx. 100,000 people. Over 16 years of age, it can develop at any age, but most often in 40 to 50 years of age. Practically, it occurs everywhere on the world, but the incidence, the course and the severity of the disease may be different among ethnic groups.

1.6. Hepatitic C virus infection (HCV)

Hepatitis C virus (HCV) is a major global health problem: currently 170 million subjects are suffering from HCV infection worldwide, incubation period is from 2 to 25 week. The outcome of HCV infection ranges from spontaneous viral clearance to hepatitis, cirrhosis and hepatocellular carcinoma. Some individuals have rapidly progressive liver disease, while others remain symptom-free virus carriers. The exact causes of the different disease courses are not known. Immune mechanisms, as well as environmental factors are responsible for the various HCV-related events.

In case HCV infection outcome the main determinant is host's immune response, beside the specificity of virus and certain enviromental factors. The elimination of HCV is natural (innate) immune response, on the other hand the adaptive function, polyclonal CD4 and CD8 T-cell responses. In this process the major agent are the major histocompatibility complex (MHC) molecules and the cytokines produced by immun cells (this is under control by genetic factors).

Previously, mainly viral hepatitis in human leukocyte antigens (HLA) and disease association correlations tested, however, the coding gene variants of cytokines are also studied. Genome-wide association study (GWAS)) method and the SNPs analysis are became suitable to examine genetic variants in the neighbour region of IL28B gene and promoter region of IL10R gene.

1.7. Hodgkin’s lymphoma (secondary malignancy)

Hodgkin's lymphoma, also called Hodgkin's disease mainly affecting the lymph nodes, it is a lymphocyte-derived malignant disease (called. lymphoma), which was given the name of Thomas Hodgkin, who first described the characteristics of the disease.

Hodgkin's disease usually begins in one lymph node group, often in the head and neck, less frequently in the armpit or in the groin area. This disease has specific, predictable pattern. It spreads from lymph node to the neighboring lymph nodes in the body. Extranodal involvement occurs only in the advanced stage of the disease (eg. marrow, liver). The general symptoms are: fever, weight loss, night sweats, itching all over the body. Characteristics may be also decreasing the performance, and after the consumption of alcohol pain can be detected in lymph nodes. The dysfunction of T-lymphocytes may result that the patients are susceptible to TBC, fungal and viral infections can develop. Furthermore, there may be neurological disorders, hormonal disorders, bone, lung involvement and urogenital complaints.

Determining the outcome is very important the stage of the disease and the presence of the risk factors. The good prognosis in turn reduces the radiation therapy and late toxicity of chemotherapy. Here we can mention the increased risk of secondary tumors (particularly breast and thyroid cancer) 15 years after the radiation (11%), appearance of acute myeloid leukemia in the first 10 years (1%) and development of non-Hodgkin's lymphoma.

A recent GWAS study identified two non-coding SNPs (rs4946728 and rs1040411) between the PRDM1 and the ATG5 regions, which are risk factor for secondary malignancy after radiation therapy in childhood Hodgkin's lymphoma patients.

1.8. Asthma

Bronchial asthma is a complex etiology, chronic, multifactorial inflammatory disease which affects the respiratory system. The prevalence of the disease is high, more than 300 million people are affected worldwide. Typical symptoms are wheezing, cough, shortness of breath.

The disease usually starts (70-80% of the cases) in child or young adulthood as a result of ordinary, everyday allergens such as pollen, dust, animal dander, foods or drugs. Epidemiological studies have revealed that in addition to genetic factors, a number of other risk factors may affect disease progression, such as ethnicity, gender, diet, occupation, breast-feeding, viral- and microbial infections, domestic animals, exposure to tobacco smoke.

With using the genetic informations the primary goal of the pharmacogenetic research are to tailor the drug therapy to individual with maximize the impact and minimizing the toxic side effects. In recent years asthma pharmacogenetic studies great emphasis has been placed among other things to study the low-affinity IgE receptor gene (FCER2) polymorphisms, which may be responsible for the symptoms in patients with asthma aggravation.

1.9. Description and the main characteristics of Roma population

The Roma population size is estimated to be about 12–15 million in the world. From this, 10–12 million people lives in Europe. The largest number (70%) of European Roma population is concentrated in Central and South-Eastern Europe. Hungary is the fourth in Europe considering the estimated size of the Romas, with about 550-600.000 people. In Roma population the general morbidity rate is elevated, the infant mortality is fourfold increased, they have specific private disease-associated mutations, and their life expectancy is less with ten years compared to Central and Eastern European populations.

As the Roma people have clear genetic diversity compared with the surrounding nations with their relatively highly conserved gene pool deriving ultimately from India, similar to other pharmacogeneticaly relevant polymorphisms we supposed differences between the Hungarian and Roma population examining several non-coding genetic polymorphisms.

2. AIMS

The aim of our experiments was to highlight the importance of non-coding region polymorphisms of several genes.

Our aim was to analyze four susceptibility *IL23R* genetic variants: the rs10889677, which is located in the 3'UTR region of the gene and three intronic variants, rs1004819 and rs2201841, the intergenic rs11209032. We also studied one protective polymorphisms of the *IL23R* gene too, the intronic rs7517847. We wanted to examine, if there are significant genetic differences between Hungarian and Roma group, that can be susceptibility or protective indicator factor in the given population.

A further aim was to examine the rs4946728 and rs1040411 noncoding SNPs located between PRDM1 and ATG1 genes on chromosome 6q21 as risk factors for secondary malignancies in patients formerly treated with radiotherapy for pediatric Hodgkin disease. Our aim was to draw conclusions which of the two ethnic groups are more vulnerable to secondary malignancy.

Our aim was also to investigate the non-coding rs28364072 SNP of the *FCER2* gene in Roma and Hungarian population, which is responsible for the symptoms exacerbation of asthma and determine genetic difference between the two group.

Another goal was to analyze the rs12979860 SNP of the *IL28B* and the rs1800896 SNP of the *IL10R* gene, and ascertain if these variants have influence on HCV patients comparing the results with healthy controls.

3. PATIENTS AND METHODS

3.1. Study populations

For the genotype analysis of IL23R gene (rs10889677, rs1004819, rs2201841, rs11209032 and rs7517847) we used 273 DNA samples from Romas (103 man, 170 female) and 253 healthy control DNA (104 male, 149 female).

A total of 748 HCV1 infected patients (365 men, 383 female, ranging in age from 18 to 82 years, mean 54 ± 10 years) were enrolled. Of the 748 HCV1 patients, 420 were treated with pegylated interferon alfa 2a/2b (Pegasys, Hoffmann-La Roche Inc./Pegintron, SP Labo N.V. Belgium) 135–180 μg/1.0-1.5 μg/kg subcutaneously per week, and ribavirin (Copegus, Hoffmann-La Roche Inc./Rebetol, SP Labo N.V. Belgium) 1000–1200 mg orally per day for 24–72 weeks, then followed up for 24 weeks. Subjects who had undetectable HCV RNA levels following 24 weeks of therapy were designated as sustained virological response (*SVR*), undetectable serum HCV RNA at week 4 after starting treatment as rapid virological response (*RVR*), undetectable HCV RNA at week 12 after starting therapy as early virological response (EVR). In our paper, patients who did not achieve SVR were regarded as "non-responders" (*non-SVR*). One hundred and five healthy individuals (64 men, 41 female, mean age 45 ± 3 years), who were consecutive voluntary blood donors with normal liver function tests and negative for HBV, HCV and HIV serology, served as controls.

For rs4946728 and rs1040411 SNPs localized between the *PRDM1* and the *ATG5* gene region DNA samples from a total of 289 Hungarians (166 males and 123 females, mean age: 37 ± 1.00, range: 18-64 yrs) and 293 Roma subjects (124 males and 169 females, mean age: 55 ± 1.02, range: 13-90 yrs) were tested.

In addition, discovery, replication and combined cases and control groups taken from the study of Best et al at Nature Medicine 2011 were enrolled to our analyses.[4] The discovery set consisted of 96 SMN cases (19 males, 77 females) and 82 SMN-free controls (31 males, 51 females). All cases and controls were individuals of European descent diagnosed with Hodgkin's lymphoma as children (median age: 15.6, range: 8–20) and treated with 25–44 Gy radiation therapy with or without alkylating chemotherapy. Cases developed SMNs with a mean latency of 20.0 years (s.d. = 5.8 years, range: 6–34). Controls were followed for at least 27 years (median: 32 years, range: 27–38) to ensure that the maximal contamination of controls by future cases was < 2%. To replicate their findings, an independent set of 62 cases with SMNs (6 males, 56 females) and 71 SMN-free controls (14 males, 57 females) were enrolled, all treated for Hodgkin's lymphoma in childhood with 25–44 Gy mediastinal radiation therapy. The replication cases and controls were self-identified as white, non-Hispanic. The genotypes and allele frequencies of the PRDM1 variants were also analyzed in combined set of discovery and replication collection of patients.

For the rs28364072 SNP of the *FCER2* gene 458 Roma (206 male, 252 female) and 397 Hungarian (222 male, 175 female) DNA samples were used.

3.2. Molecular biological methods

The molecular analyses were performed using DNA extracted from peripheral blood leukocytes with a routine salting out procedure. PCR–RFLP methods were applied to test the alleles of the *IL23* receptor gene, using the following forward and reverse primers: for rs10889677 forward primer: 5'-ATC GTG AAT GAG GAG TTG CC-3' and reverse primer: 5'-TGT GCC TGT ATG TGT GAC CA-3'; for rs1004819 forward primer: 5'-GCA TTC TAG GAC CGT TTT GG-3' and reverse primer: 5'-ATC TGG TGG AAA TAT GTG AAA CCT A-3'; for rs2201841 forward primer: 5'-GGC AAA AGG GAA TTG AGA

GG-3’ and reverse primer: 5’-GGC CTA TGA TTA TGC TTT TTC CTG-3’; for rs11909032 forward primer: 5’-TTG TTA CTG GAG TTA AAC CTC TTG C-3’ and reverse primer: 5’-AGG AAT AAT TGC TGA GAT GCA ATG-3’; for rs7517847 forward primer: 5’-AAA CAT TGA CAT TCC CTT CAT AC-3’ and reverse primer: 5’-GAA ATG AGT CAC CAA TAA TCC AC-3’. The following restriction endonucleases were used for the digestion, MnlI for rs10889677, TaaI for rs1004819, HpyF3I for rs2201841, BseMI for rs11209032 , BseMII for the rs7517847 SNP.

The *IL28B* rs12979860 SNP was determined using Custom Taqman SNP Genotyping Assays (Applied Biosystems, Life Technologies, Foster, CA, USA). The IL10R −1087 (also known as IL10R −1082) *(rs1800896)* promoter region was required for the formation of the EcoNI recognition sequence 5′-AAGACAACACTACTAAGGCT-3′; the lower primer was 5′-TAAATATCCTCAAAGTTCC-3′. EcoNI restriction enzyme was used for digestion.

The *PRDM1-ATG5* gene regions SNPs, rs4946728 and rs1040411 were determined two predesigned TaqMan SNP Genotyping Assays (Applied Biosystems, Life Technologies, Foster, CA, USA).

The *FCER2* rs28364072 SNP was also determined using Custom Taqman SNP Genotyping Assays (Applied Biosystems, Life Technologies, Foster, CA, USA) with specified conditions.

3.3. Statistical analysis

To the evaluation of the relationship between the diseases, the genetic variants and the haplotypes, χ^2-test and regression analysis was used. Analyses were carried out SPSS 20.0 package for Windows.

4. RESULTS

4.1. *IL23R* gene

All IL23R genotype and allele frequencies were in Hardy–Weinberg equilibrium both in Hungarian and in Roma subjects. The rs10889677 AA (24.5% vs. 5.93%, $p < 0.05$), rs1004819 AA (24.2% vs. 9.49%, $p < 0.05$), rs2201841 CC (20.5% vs. 6.32%, $p < 0.05$) and rs11209032 AA (19.0% vs. 9.49%, $p < 0.05$) homozygous genotype frequencies were significantly increased in the Roma group compared to the Hungarian samples. The rs10889677A (48.5% vs. 29.8%, $p < 0.05$), rs1004819A (47.9% vs. 30.2%, $p < 0.05$), rs2201841C (46.1% vs. 28.9%, $p < 0.05$) and rs11209032A (43.2% vs. 33.2%, $p < 0.05$) allele frequencies were also significantly over-represented in the Roma population compared to the healthy Hungarian controls. The rs7517847 showed significantly decreased GG homozygous genotype frequency in Roma samples (5.13%), while it was 16.2% in the Hungarian subjects ($p < 0.05$). The frequency of allelic variants and genotypes of IL23R in the Roma group and Hungarian controls are shown in Table 1.

Table 1.: *Genotypes and minor allele frequencies of IL23R polymorphisms.*

	Roma (273)	**Hungarian (253)**
IL23R genotype		
IL23R rs10889677		
CC	75 (27.5%)	117 (46.2%)
CA	131 (48.0%)	121 (47.8%)
AA	67 (24.5%)*	15 (5.93%)
A allele frequency	48.5%*	29.8%
IL23R rs1004819		
GG	77 (28.2%)	124 (49.0%)
GA	130 (47.6%)	105 (41.5%)
AA	66 (24.2%)*	24 (9.49%)
A allele frequency	47.9%*	30.2%
IL23R rs2201841		
TT	77 (28.2%)	123 (48.6%)
TC	140 (51.3%)	114 (45.1%)
CC	56 (20.5%)*	16 (6.32%)
C allele frequency	46.1%*	28.9%
IL23R rs11209032		
GG	89 (32.6%)	109 (43.1%)
GA	132 (48.4%)	120 (47.4%)
AA	52 (19.0%)*	24 (9.49%)
A allele frequency	43.2%*	33.2%
IL23R rs7517847		
TT	129 (47.3%)	74 (29.2%)
TG	130 (47.6%)	138 (54.5%)
GG	14 (5.13%)*	41 (16.2%)
G allele frequency	28.9%*	43.5%

* $p<0.05$ vs. Hungarian

4.2. *IL28B* and *IL10R* genes

When comparing the *IL28B* genotype frequencies between groups of healthy controls and patients, the CC genotype occurred with lower frequency in HCV patients than in controls, thus, suggesting its protective effect against chronic hepatitits C. On the other hand, CT heterozygosity and T alleles were more prevalent in patients, and may, therefore, convey susceptibility for the disease (Table 2.). Subjects that received P/R therapy and who had the *IL28B* CC genotype achieved higher rates of SVR than subjects who had CT genotype (58.6% vs 40.8%) (OR 2.057, 95% CI: 1.305-3.058, $p = 0.002$), or those carrying the T allele (41.8%, OR 1.976, 95% CI: 1.263-3.058, $p = 0.002$) (Table 3.).

IL10R GG genotype occurred with lower frequency in patients (31.8%) than in controls (52.2%) (OR: 0.428, $p < 0.001$). The prevalence of IL10R AA allele was 68.15% in patients and 47.8% in healthy individuals (OR: 2.335, $p < 0.001$). Among P/R treated patients with SVR, the IL10R GG genotype occurred with higher frequency than the AA genotype, (57/178, 32.0% vs 31/178, 17.4%, (OR: 1.84, 95% CI 1.13-2.98, $p = 0.013$). The SVR rate in patients with IL10R GG genotype was 42.2% (59/125); in those with GA 47.4% (88/178) and in AA patients 39.7% (31/78). HCV patients with A allele (non-GG genotype) achieved SVR in 46.4% (119/256) (Table 4.).

Table 2.: *Prevalence of IL28B genotypes in Hungarian HCV patients and healthy controls.*

		HCV1 (n=748)	**Controls (n=105)**
***IL28B* genotype rs12979860**	CC	195 (26.1%)*	54 (51.4%)
	CT	411 (54.9%)*	39 (37.1%)
	TT	142 (19.0%)	12 (11.4%)
	T allele frequency	553 (73.9%)*	51 (48.6%)

Table 3.: *Sustained virological response (SVR) rate stratified by IL28B genotypes in Hungarian HCV patients.*

	Treated	**SVR**	
***IL28B* genotype**	**n**	**Number of patients**	**%**
CC	116	68	58,6
CT	228	93	40,8
TT	76	34	44,7
T allele (non-CC)	304	127	41,8

ILB28 CC vs. CT	OR: 2.057 (1.305–3.236), p = 0.002*
IL28B CC vs. TT	OR: 1.751 (0.975–3.134), p = 0.059
IL28B CC vs. T (non-CC)	OR: 1.976 (1.263–3.058), p = 0.002

Table 4.: *Prevalence of IL10R −1087 genotypes in HCV patients and healthy controls.*

		HCV1 (n=672)	Controls (n=92)
***IL10R* genotype**			
rs1800896	GG	214 (31.8%)	48 (52.2%)
	GA	333 (49.6%)	32 (34.8%)
	AA	125 (18.6%)	12 (13.0%)
	A allele frequency	458 (68.15%)	44 (47.8%)

4.3. *PRDM1-ATG5* gene region

The profile of allelic variants for both of rs4946728 and rs1040411 SNPs of the PRDM1-ATG5 gene region, including the risk-associated minor allele frequencies did not significantly differ between the Hungarian and Roma samples. The Roma and the Hungarian samples predisposing variants frequencies did not differ statistically from the same variants of the replication, discovery and combined conrols (Best et al.).

The homozygous carriers of the PRDM1 rs4946728 risk C allele were much frequent in the Hungarian and Roma subgroups compared to the discovery and combined controls presented by Best et al. In the Roma subgroup, the heterozygous carriers of the PRDM1 rs4946728 risk C allele were significantly prevalent compared to all three control groups. In addition, considerable accumulation of the PRDM1 rs4946728 risk C allele was observed in both the Hungarian and Roma populations compared to the discovery and combined controls (79.4 and 83.5% vs. 59.1 and 63.7%; $p<0.05$).

The homozygous carriers of the PRDM1 rs1040411 risk A allele were significantly frequent in the Hungarian and Roma subgroups compared to the discovery and combined controls. Statistically significant accumulation of the PRDM1 rs1040411 risk A allele was detected in the Hungarian and Roma populations compared to the discovery controls (56.4 and 56.8% vs. 39.6%; $p<0.05$). Besides, in the Hungarian subgroup, the frequency of the rs1040411 risk A allele was higher than in combined controls, as well (56.4% vs. 44.1%; $p<0.05$).

The risk alleles of PRDM1 variants (rs4946728 C; rs1040411 A) did not show accumulation in neither the Hungarian nor the Roma groups compared to the discovery, replication and combined cases. The allele frequencies observed among the “Cases” groups by Best et al. did not also significantly differ from those of our sample groups for either PRDM1 variant.

Table 5.: *Distribution of the PRDM1 rs4946728 genotypes in Hungarian and Roma population samples, with cases and controls from Best et al.*

rs4946728				
Group	**Genotype**			Risk allele frequency **(%)**
	A/A (n)	**A/C (n)**	**C/C (n)**	
Roma	7[abc]	78[abc]	194[ac]	83.5[ac]
Hungarian	7[abc]	100	170[ac]	79.4[ac]
Discovery Cases[f]	2	23	71	86.0
Replication Cases[f]	1	17	43	84.4
Combined Cases[f]	3	40	114	85.3
Discovery Controls[f]	12	43	27	59.1
Replication Controls[f]	6	32	33	69.1
Combined Controls[f]	18	75	60	63.7

[a] $p<0.05$ vs. „discovery" Controls
[b] $p<0.05$ vs. „replication" Controls
[c] $p<0.05$ vs. „combined" Controls
[d] $p<0.05$ vs. „discovery" Cases
[e] $p<0.05$ vs. „combined" Cases
[f] Calculated from genotype count data of table 1 of Timothy Best et al. (Best és mtsai 2011).

Table 6.: *Distribution of the PRDM1 rs1040411 genotypes in Hungarian and Roma population samples, with cases and controls from Best et al.*

rs1040411				
Group	**Genotype**			Risk allele frequency **(%)**
	A/A (n)	**A/G (n)**	**G/G (n)**	
Roma	84[ac]	140	52[acde]	55.8[a]
Hungarian	84[ac]	141	49[acd]	56.4[ac]
Discovery Cases[f]	42	47	7	68.2
Replication Cases[f]	22	30	9	60.6
Combined Cases[f]	64	77	16	65.3
Discovery Controls[f]	10	45	27	39.6
Replication Controls[f]	18	33	19	49.3
Combined Controls [f]	28	78	46	44.1

[a] $p<0.05$ vs. „discovery" Controls
[b] $p<0.05$ vs. „replication" Controls
[c] $p<0.05$ vs. „combined" Controls
[d] $p<0.05$ vs. „discovery" Cases
[e] $p<0.05$ vs. „combined" Cases
[f] Calculated from genotype count data of table 1 of Timothy Best et al. (Best és mtsai 2011).

4.4. *FCER2* gene

Examining the intronic *FCER2* T2206C polymorphism, we found significant differences for CC homozygous genotype (p=0.032) compared the Roma population with Hungarian controls. The homozygous presence of the minor 2206C allele in Roma participants is more than twofold decreased than in Hungarians (2.8% vs 5.8%). The predisposing C allele frequency was similar in the two populations (24.8% vs 24.6%).

Table 7.: *Genotypes and minor allele frequencies of FCER2 polymorphism in Roma and Hungarian populations.*

		Roma (n=458)	**Hungarian (n=397)**
***FCER2* genotypes rs28364072**	TT	244 (53.3%)	225 (56.7%)
	TC	201 (43.9%)	149 (37.5%)
	CC	13 (2.8%)*	23 (5.8%)
	C allele frequency	24.8%	24.6%

* p<0.05 vs. Hungarian

5. DISCUSSION OF THE RESULTS AND CONCLUSIONS

After the recognition of the importance of DNA, the researchers focused on the coding region of DNA, but 98% their part is non-coding sequence. It revealed that these "junk DNAs have an important significance, the larger part of the human genome is non-coding (pseudogene, gene fragments, non-translated region), several are outside of the gene or between the genes (extragenic sequence).

The 80% of the extrageneic sequences is unique or low copy numbers, smaller part is strongly iterative. These can be tandem repeats and dispersed repetitive sequences.
The tandem repeats are short DNA sequences that occur only in certain chromosome segments. Is subdivided into three classes: satellite DNA, minisatellite DNA, microsatellite DNA.

Tandem repeats phases influence the activity of the neighboring genes. The repetitions determined how closely the DNA twists into a special shape called nucleosomes, and that the tunneling mechanism that determines the extent to which genes can be activated.

Level of genes very important are the introns and the non-translated regions. The largest part of DNA are the introns, based on these that introns have diverse biological functions, regulatory and structural targets. The 5'UTR region is necessary to linkage to mRNA ribosome, thereby these are key player in the normal protein synthesis. When a mutation occurs in this region, that can change the whole process or certain parts.

In case of gene mutation not only the extent of the mutation (the length of the DNA sequence) is important, but also the location. No matter whether there are mutations in the coding or non-coding region or even the border of the two. These is splicing mutation exon-intron boundary sequences play an important role in several process. As a result of a splicing mutation an exon can lost or an intron can translate, these can result in defective protein.

It can be said that the non-coding regions have an important role, their mutations can be the cause of development of several diseases. In the dissertation 10 SNP were investigated, the rs10889677 of IL23R located in the 3'UTR region; six intronic polymorphisms (rs1004819, rs2201841, rs7517847 of the *IL23R*, rs12979860 of the *IL28B*, rs1800896 of the *IL10R*, rs28364072 of the *FCER2*) and three intergenic variant (rs11209032 between the *IL23R* and the *IL12Rß1* region, rs4946728, rs1040411 between the *PRDM1* and the *ATG5* region). Our aim was to examine these variants in Roma nad Hungarian population (we studied HCV patients and controls in case of *IL28B* and *IL10R*), and draw conclusion, which population is more susceptible to the studied disease.

Examining the IL23R variants, the A allele and AA homozygous genotype frequencies of rs10889677, rs1004819, and rs11209032 SNPs and the C allele and homozygous CC genotype frequencies of rs2201841SNP were significantly increased in Roma group, so carrying these SNPs means higher risk to develop autoimmune diseases. Evaluate the results of rs7517847 SNP of the IL23R gene the G allele and the GG homozygous genotype frequencies were significantly decreased in the Roma group comparing the result to the healthy controls, so carrying this variant is probably a protective factor against develop autoimmune diseases.

Examining IL28B rs rs12979860 SNP the C allele and CC homozygous genotype frequencies were significantly decreased in the HCV patients's group, so carrying this variant is a protective factor against HCV infection. The rs1800896 of IL10R gene showed significantly decreased G allele and GG homozygous genotype frequencies, so carrying this variant protect against HCV disease.

We investigated the rs4946728 and the rs1040411 noncoding SNPs located between the PRDM1 and the ATG1 genes on chromosome 6q21, which are risk factors for secondary malignancies in patients formerly treated with radiotherapy for pediatric Hodgkin disease and

we could not detect significant difference between the allele and the genotype frequencies compare the Roma group to Hungarians. So it can concluded that the groups are exposed the same to secondary malignancy after radiation therapy.

The pharmacogeneticcally important FCER2 rs28364072 CC homozygous genotype freqeuncy was significantly decreased in Roma population compare the results to Hungarians. So we can expect decreased expression of the receptor protein in Hungarian population that causes the icreased IgE level which can be responsible for the exacerbation of inflammatory symptoms. Results of these the conventional drug therapy can be reduced in the Hungarian population than in the Roma group.

The non-coding region are very important, they have special significance for certain diseases susceptibility, as well as pharmacogenetic point of view. The variants can cause individual differences in drug metabolism and thereby influence the therapeutic response. There are differences between nations, the common message of the the dissertation is this.

6. SUMMARY OF THE RESULTS

1. Our aim was to investigate autoimmune diseases associated *IL23R* gene variants (rs10889677, rs1004819, rs2201841, rs7517847 és rs11209032 SNPs) examining 273 Roma and 253 healthy Hungarian controls. *IL23R* rs10889677, rs1004819, rs2201841 és rs11209032 SNPs showed increased minor allele and homozygous genotype frequencies in the Roma group, so carrying these varaints means susceptibility factor for autoimmune diseases in Romas. *IL23R* rs7517847 SNP showed decreased G allele and homozygous GG genotype frequencies in the Roma cohort, so carrying this variant is probably a protective factor against develop autoimmune diseases.

2. During the examination of rs12979860 SNP of IL28B we analyzed 748 HCV patients and 105 controls, while in case of *IL10R* rs1800896 SNP 672 HCV patients and 92 controls were the subjects of the investigation. *IL28B* rs12979860 SNP showed decreased C allele and CC homozygous genotype frequencies in HCV group, so carrying this variant means a protective agent against HCV infection. The rs1800896 SNP os *IL10R* gene the G allele and the GG homozygous genotype showed decreased frequencies, so carrying this variant is a protective factor against HCV infection.

3. Examining the rs4946728 and the rs1040411 noncoding SNPs located between the PRDM1 and the ATG1 genes on chromosome 6q21, which are risk factors for secondary malignancies in patients formerly treated with radiotherapy for pediatric Hodgkin disease on 293 Roma and 289 Hungarian DNA samples we could not detect significant difference between the allele and the genotype frequencies compare the two groups. So it can concluded that the groups are exposed the same to secondary malignancy after radiation therapy.

4. Investigating the pharmacogeneticcally important FCER2 rs28364072 SNP (which is responsible for the symptoms exacerbation of asthma) on 458 Roma and 397 Hungarian DNA samples, we could detect that the CC risk homozygous genotype frequency is lower in the Roma group. In Hungarian population the increased CC risk homozygous genotype freqeuncy may predict higher asthmatic disease incidence.

7. PUBLICATIONS

7.1. Publications supporting the dissertation

Varszegi D, Duga B, Melegh BI, Sumegi K, Kisfali P, Maasz A, Melegh B. Hodgkin Disease Therapy Induced Second Malignancy Susceptibility 6q21 Functional Variants in Roma and Hungarian Population Samples. Pathol Oncol Res. 2013 Dec 5. [Epub ahead of print] **IF: 1.555**

Magyari L, **Varszegi D**, Sarlos P, Jaromi L, Melegh BI, Duga B, Kisfali P, Kovesdi E, Matyas P, Szabo A, Szalai R, Melegh B. Marked differences of haplotype tagging SNP distribution, linkage, and haplotype profile of IL23 receptor gene in Roma and Hungarian population samples. Cytokine. 2014 Feb;65(2):148-52. **IF: 2.518**

Pár A, Pár G, Tornai I, Szalay F, **Várszegi D**, Fráter E, Papp M, Lengyel G, Fehér J, Varga M, Gervain J, Schuller J, Nemes Z, Péterfi Z, Tusnádi A, Hunyady B, Haragh A, Szinku Z, Vincze A, Szereday L, Kisfali P, Melegh B. IL28B and IL10R -1087 polymorphisms are protective for chronic genotype 1 HCV infection and predictors of response to interferon-based therapy in an East-Central European cohort. BMC Res Notes. 2014 Jan 8;7(1):12. **IF: 1.390**

Pár A, Pár G, Tornai I, Szalay F, **Várszegi D**, Fráter E, Papp M, Lengyel G, Fehér J, Varga M, Gervain J, Schuller J, Nemes Z, Péterfi Z, Tusnádi A, Hunyady B, Haragh A, Szinku Z, Pálinkás L, Berki T, Vincze A, Kisfali P, Melegh B. [IL28B CC genotype: a protective factor and predictor of the response to interferon treatment in chronic hepatitis C virus infection]. Orv Hetil. 2013 Aug 11;154(32):1261-8.

Szalai R, Matyas P, **Varszegi D**, Melegh M, Magyari L, Jaromi L, Sumegi K, Duga B, Kovesdi E, Hadzsiev K, Melegh B. Admixture of beneficial and unfavourable variants of GLCCI1 and FCER2 in Roma samples can implicate different clinical response to corticosteroids. Mol Biol Rep. 2014 Nov;41(11):7665-9. **IF: 2.506**

7.2. Other publications

Sarlos P, **Varszegi D**, Csongei V, Magyari L, Jaromi L, Nagy L, Melegh B. Susceptibility to ulcerative colitis in Hungarian patients determined by gene-gene interactions. World J Gastroenterol. 2014 Jan 7;20(1):219-27. **IF: 2.547**

Marek E, Dergez T, D'cruz G, Bozsa S, Cseh A, Szilard I, Benczik M, Kiss I, **Varszegi D**, Vilagi S, Ember I, Gocze P. Human papillomavirus infections among Hungarian female sex workers. Eur J Cancer Care (Engl). 2014 Jan;23(1):65-75. **IF: 1.308**

Szigeti R, Chao SC, **Várszegi D**, Czakó M, Kosztolányi G, Kellermayer R. [The first genetically supported case of chronic benign pemphigus (Hailey-Hailey disease in Hungary]. Orv Hetil. 2005 Sep 11;146(37):1933-5.

Harangi F, **Várszegi D**, Schneider I, Zombai E. Complete recovery from juvenile pemphigus vulgaris. Pediatr Dermatol. 2001 Jan-Feb;18(1):51-3. **IF: 1.041**

Nyul Z, Harangi F, **Várszegi D**, Zombai E. Vesicobullous lesions in a child. Bullous pemphigoid (BP). Arch Dermatol. 1997 Jun;133(6):776-7, 779-80. **IF: 2.714**

Harangi F, **Várszegi D**, Szücs G. Asymmetric periflexural exanthem of childhood and viral examinations. Pediatr Dermatol. 1995 Jun;12(2):112-5. **IF: 1.041**

Cumulative impact factors: 16.620