Allele distribution of the interleukin 23 receptor gene polymorphisms, its interactions with susceptibility genes in ulcerative colitis, and haplotype variants in healthy Roma population

Doctoral (Ph.D.) theses

Author: Patrícia Sarlós, M.D.

Department of Medical Genetics
Medical School, University of Pécs

Pécs, 2014

Head of the Doctoral School: Prof. Balázs Sümegi, Ph.D, D.Sc
Program leader: Prof. Béla Melegh, M.D., Ph.D, D.Sc.
Project leaders: Prof. Béla Melegh, M.D., Ph.D, D.Sc.
Prof. Nagy Lajos, M.D., Ph.D, D.Sc.

1. Abbreviations

ATG16L1	autophagy-related protein 16-1
CARD15	caspase recruitment domain-containing protein 15
CD	Crohn's disease
DLG5	disks large homolog 5
EDTA	ethylenediaminetetraacetic acid
GWAS	genome-wide association study
ht	haplotype
IBD	inflammatory bowel disease
IL	interleukin
IL23R	interleukin 23 receptor
LD	linkage disequilibrium
MDR	multidimensional data reduction
MIM	Mendelian Inheritance in Man
NOD	nucleotide-binding oligomerization domain-containing protein
OCTN1/2	organic cation transporter 1 and 2
OR	Odds ratio
PCR	polymerase chain reaction
RAF	risk allele frequency
RFLP	restriction fragment length polymorphism
SLC22A4	solute carrier family 22, member 4
SLC22A5	solute carrier family 22, member 5
SNP	single nucleotide polymorphism
Th cell	T helper cell
UC	ulcerative colitis
UTR	untranslated region

2. Introduction

Single nucleotide polymorphisms (SNPs) are the most common variations at a single position in a DNA sequence among individuals, but they only have modest effect individually in complex genetical disorders. Analyses of genome-wide association studies (GWAS) often focus on identifying individual SNPs which modify the risk of a phenotype, assuming the underlying association of an individual SNP without considering the involvement of any other SNPs. Today, single SNP studies are more often replaced with the combined analysis of SNP effects through interaction and haplotype analysis.

Inflammatory bowel diseases (IBD) – clinically classified as Crohn's disease (CD; MIM 26600) or ulcerative colitis (UC; MIM 191390) – are common chronic, relapsing inflammatory disorders of the gastrointestinal tract with complex, multifactorial etiology. GWAS have resulted in the identification of many novel loci in IBD; to date, the number of known risk loci has expanded to 163. The identified separate IBD loci have only modest effects individually on IBD susceptibility. They account together for only 20-25% of the heritability, suggesting that gene-gene interactions as well as gene-environmental interactions could play a key role in IBD pathogenesis and fill the so called "genetic vacuum" of polygenic diseases.

The most studied SNPs (*SLC22A4*, *SLC22A5*, IGR2096_a, IGR2198_a, IGR2230_a) on inflammatory bowel disease-5 (IBD5) locus (chromosome 5q31, OMIM 606348) have been reported to confer susceptibility to CD. Peltekova reported two novel functional CD associated SNPs: the C1672T substitution in exon 9 in the *SLC22A4* gene (*OCTN1*) and the G-207C transversion in the promoter region in the *SLC22A5* gene (*OCTN2*), and the so called TC haplotype. Noble et coworkers assessed that not only the *OCTN* variants are susceptible for the disease but also a number of surrounding SNPs from the list of SNPs comprising the IBD5 risk haplotype (IGR2096_a, IGR2198_a, IGR2230_a). The TC homozygous carriers increased the risk for CD by 3,4-5,1-fold, while the elevated risk attributed to the *OCTN* TC haplotype and *NOD2/CARD15* mutations was additive with an Odds ratio of 7,2-10,5 in double carriers. In some studies, including a GWAS meta-analysis, association with UC has also been established. One theory of why *OCTN1* and *OCTN2* are related to disease activity has to do with their role in maintaining barrier function in the intestine.

Interleukin-23 (IL23) is a heterodimeric, proinflammatory cytokine that shows similar functions to IL12 in promoting cellular immunity and enhancing lymphocyte proliferation. However, unlike IL12, IL23 develops CD4+ T cells into IL17 producing Th17 cells, instead of Th1 cells. IL23 is a regulatory cytokine produced by activated macrophages and dendritic cells. The interleukin-23 receptor (*IL23R*) gene (OMIM 607562) was originally described as a CD susceptibility gene, but recently the association with UC has been also confirmed in three separate GWA studies. The *IL23R* gene investigation was initiated by Duerr in 2006, who found an association between the *IL23R* gene and IBD, and reported several independent functional SNPs of the gene and its neighboring region, which are susceptible (rs10889677, rs1004819, rs2201841, rs11805303, rs11209032) to or protective (rs7517847) against CD and UC in non-Jewish subjects. Two variants showed no disease association, the rs7530511 and rs1884444.

The attention of recent IBD studies has focused on multi-locus analysis, especially involving the *NOD2*, *IL23R*, *ATG16L1*, *DLG5* genes and IBD5 locus.

Haplotype analysis permits simultaneous analyses of multiple SNPs. Haplotypes are a combination of alleles at different markers along the same chromosome which are inherited as a unit. Since the sequence of the ancient haplotypes was conserved early in the process of evolution, today there are haplotypes the distribution of which marks ethnic groups.

As the Roma people have clear genetic difference compared with the surrounding nations with their relatively highly conserved gene pool deriving ultimately from India, similar to other pharmacogenetically relevant polymorphisms we supposed differences of the *IL23R* structure as well. The Roma population size is estimated to be about 12-15 millions in the world. From this, 10-12 million people live 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 700000 - 1 million people. In the 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 ten years less.

3. Aims

1. The main objective of our work was to study the allele distribution of the interleukin 23 receptor (*IL23R*) gene polymorphisms and its interactions with susceptibility genes in Hungarian ulcerative colitis patients.

2. In single-locus association study we analysed two of the susceptibility *IL23R* gene risk SNPs (rs1004819 and rs2201841) and five variants of the IBD5 locus (IGR2096a_1 (rs12521868), IGR2198a_1 (rs11739135), IGR2230a_1 (rs17622208), *SLC22A4* (rs1050152), *SLC22A5* (rs2631367)) in Hungarian UC patients.

3. We performed a combined genetic analysis, stratifying the two susceptibility *IL23R* gene variants, rs2201841 and rs10004819 by the IBD5 markers (*SLC22A4, SLC22A5,* IGR 2096_a, IGR2198_a, IGR2230_a) to test for possible gene-gene interactions in UC population.

4. We also aimed to identify the occurring *IL23R* haplotypes in healthy Hungarian and Roma population samples.

5. Our aim was to determine the genetic variability of the major haplotype tagging polymorphisms, and the haplotype profile of *IL23R* between the two ethnically different, Hungarian and Roma population both living in Hungary.

4. Methods

4.1. Study population

The study population (n = 636) was comprised of 320 UC patients and 316 healthy, unrelated controls. All patients and controls were Caucasian and of Hungarian origin. The diagnosis of UC was determined according to established guidelines based on clinical, endoscopic, radiological and histopathological criteria.

For the *IL23R* haplotype analyses we used the DNA samples of a total of 273 Roma and 253 Hungarian, healthy blood donor subjects. The origin of DNA samples was the central Biobank governed by the University of Pecs, as part of the National Biobank Network of Hungary (www.biobanks.hu), which belongs also to the pan-European Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) preparatory phase project (http://bbmri.eu/bbmri/). The donors were real representatives of the entire population.

4.2. DNA isolation and genotyping

Genomic DNA was isolated from peripheral blood leukocytes with routine salting out method. For genotyping the variants of IBD5 locus [IGR2198a_1 (rs11739135), IGR2096a_1 (rs12521868), IGR2230a_1 (rs17622208) and *SLC22A5* (rs2631367)] and *IL23R* (rs1004819, rs2201841, rs10889677, rs11805303, rs11209032, rs7517847, rs7530511, rs1884444) gene PCR-RFLP (restriction fragment length polymorphism) methods were applied, for the *SLC22A4* (rs1050152) direct DNA sequencing was used by BigDye Terminator labeling with ABI 3100 automatic sequencer (Foster City, CA, United States). The PCR amplifications were performed on MJ Research PTC-200 thermal cyclers (Bio-Rad, Hercules, CA, United States). The amplicon contained an obligate cleavage site of the restriction enzyme for the suitable visual control of the efficacy of the digestion. The restriction fragments were separated by electrophoresis on 3% agarose gels containing ethidium bromide and visualized by UV transillumination.

4.2.1 Determination of IBD5 variants

Amplification included an initial denaturation step (95 °C for 2 min) followed by 35 cycles of denaturation (95 °C for 30 s), annealing for 45 s at 54 °C (rs17622208, rs1050152), 58 °C (rs11739135, rs12521868, rs2631367), primer extension at 72 °C for 45 s and final extension at 72 °C for 5 min.

Each polymerase chain reaction contained 200 μmol/L of each dNTP, 1 unit of Taq polymerase, 5 μL of reaction buffer (100 mmol/L Tris HCl, pH = 9.0; containing 500 mmol/L KCl, 15 mM MgCl2), 0.2 μmol/L of each primer and 1 μL DNA to be amplified in a final volume of 50 μL. The amplicons were digested by allelespecific restriction endonucleases *Hin1II* (rs11739135), *TrulI* (rs12521868), *DdeI* (rs17622208), *HpaII* (rs2631367). The restriction fragments were separated by electrophoresis on 3% agarose gels containing ethidium bromide and visualized by UV transillumination.

4.2.2. Determination of *IL23R* variants

For the rs10889677, rs1004819, rs2201841, rs11805303, rs11209032, rs7517847, rs7530511 and rs1884444 of the *IL23R* gene we designed forward and reverse primers. The following conditions were used: initial denaturation at 96 ºC for 3 min followed by 35 cycles of denaturation at 96 ºC for 45 s, annealing at 60 ºC for 45 s (rs10889677 and rs7530511); 54 ºC for 45 s (rs1004819); 55 ºC for 30 s (rs2201841, rs11209032); 59 ºC for 30 s (rs11805303); 55 ºC for 45 s (rs7517847); 58 ºC for 45 s (rs1884444) and extension at 72 ºC for 45 s and final extension at 72 ºC for 10 min.

The amplicons were digested by allele-specific restriction endonucleases *HpyF3I* (rs2201841), *TaaI* (rs1004819), *MnlI* (rs10889677 and rs11805303), *BseMI* (rs11209032), *BseMII* (rs7517847), *HphI* (rs7530511), *PscI* (rs1884444). Each restriction endonuclease had an obligate cleavage site on the amplicon to enable us to control the sure-efficacy of the digestion. The restriction fragments were separated by electrophoresis on 3% agarose gels containing ethidium bromide and visualized by UV transillumination.

4.3. Statistical analysis

Each genetic marker was tested for Hardy-Weinberg equilibrium in the control population. Statistical analysis was carried out using SPSS 20.0 package for Windows (SPSS

Inc, Chicago, IL, United States). Genotype and allele frequency differences between cases and controls were evaluated using Pearson's χ 2-test. Haploview 4.1 was used to test linkage disequilibrium. Haplotype frequencies were estimated using PHASE version 2.1.

Binary logistic regression analysis was applied to observe the individual contributions of IBD5 and *IL23R*, and to test for pairwise statistical interaction. An association was considered significant if a P value of < 0.05 was attained. The *IL23R* genotypes were stratified by IBD5 genotypes. The odds ratios and confidence intervals for these specific combinations of IBD5 and *IL23R* were derived from χ^2 in 2×2 contingency tables.

5. Results

5.1. Linkage analysis in ulcerative colitis patients

All of the investigated SNPs were in Hardy-Weinberg equilibrium in controls. The r^2 values for the tested IBD5 loci (IGR2198a_1, IGR2096a_1, IGR2230a_1, *SLC22A4*, *SLC22A5*) and for *IL23R* (rs1004819 and rs2201841) were below 0.8, for *SLC22A*4 and IGR2096a_1 the r^2 value was 0.9.

5.2. Single SNP marker association analysis of IBD5

No significant association for any variants of IBD5 region and UC was observed (Table 1).

5.3. Single SNP marker association analysis of *IL23R* rs1004819 G/A and rs2201841 T/C

For the *IL23R* rs1004819 A allele we found significantly higher allele frequency (0.343 versus 0.287; $P = 0.032$) in UC patients compared to control subjects. The SNP rs1004819 showed significant association with UC risk for carriers (heterozygotes and homozygotes together, $P = 0.004$, OR = 1.606; 95%CI:1.160-2.223) and the SNP rs2201841 for homozygotes ($P = 0.030$, OR = 1.983; 95%CI: 1.069-3.678) (Table 1).

5.4. Gene-gene interaction (IBD5-*IL23R*) analysis in ulcerative colitis patients

We analyzed the possible statistical interactions by pairs of *IL23R* variants and IBD5 with binary logistic regression. No evidence of interactions between these seven markers was found. None of the *P* values was significant; the lowest *P* value was 0.084.

Next, we stratified the *IL23R* genotypes by IBD5 genotypes and observed these specific genotype combinations of single markers by pairs, the combined odds ratios are shown in Table 2. The *IL23R* rs1004819 A variant did not show significant association with UC on the background of all wild type IBD5 genotypes, respectively. We could detect significantly elevated high odds ratios for rs1004819 A variant only in carriers of *SLC22A4* T allele, *SLC22A5* C, IGR2198a_1 C or IGR2096_a T allele. The combined OR seen in rs1004819 A and *SLC22A5* C carriers ($P = 0.048$, OR = 1.691; 95%CI: 1.003-2.821) and the odds ratio for rs1004819 A in single gene analysis ($P = 0.004$, OR = 1.606; 95%CI: 1.160-2.223) were nearly equal while in combination with IGR2198a_1 C ($P = 0.020$, OR = 1.803; 95%CI: 1.096-2.966) and IGR2096_a T ($P = 0.010$, OR = 1.911; 95%CI: 1.162-3.143) the rs1004819 A variant showed higher disease risk. The highest OR value was calculated in the presence of *SLC22A4* T allele ($P = 0.005$, OR = 2.015; 95%CI:1.230-3.300). There was no association with UC for any combinations of rs1004819 and IGR2230a_1.

For the combinations of IBD5 loci and *IL23R* rs2201841 we could detect significantly elevated high odds ratios only in carriers of rs2201841 homozygotes and wild type IBD5 genotypes ($P = 0.018$, OR = 3.413; 95%CI: 1.169-9.965 for *SLC22A4*; $P = 0.014$, OR = 3.946; 95%CI: 1.232-12.645 for *SLC22A5*; $P = 0.018$, OR = 3.777; 95%CI: 1.181-12.084 for

IGR2230a_1; *P* = 0.027, OR = 3.165; 95%CI: 1-088-9.206 for IGR2198a_1; *P* = 0.026, OR = 2.977; 95%CI: 1.099-8.066 for IGR2096a_1 background). The *IL23R* rs2201841 homozygous genotype and IBD5 carrier status together did not confer susceptibility for UC.

5.5. Linkage analysis in healthy Roma and Hungarian population

All *IL23R* genotype and allele frequencies (rs1004819, rs2201841, rs10889677, rs11805303, rs11209032, rs7517847, rs7530511, rs1884444) were in Hardy–Weinberg equilibrium both in Hungarian (n = 253) and in Roma (n = 273) subjects.

Using the Hapmap LD map, the strongest linkage disequilibrium was shown between the SNPs rs11805303–rs7517847 in the Roma population ($r^2 = 1$) and rs1004819–rs11805303 ($r^2 = 0.96$) in Hungarian population (Figure 1).

5.6. *IL23R* haplotype analysis in healthy Roma and Hungarian population

We examined five susceptibility (rs10889677, rs1004819, rs2201841, rs11805303, rs11209032), one protective (rs7517847) and two neutral variants (rs7530511, rs1884444) of the *IL23R* gene in pooled DNA of healthy Roma (Gipsy) and Hungarian population samples. The major haplotypes (ht) can be seen in Table 3. The haplotype frequencies are shown in Table 4. The ht2 (14.3% vs. 85.7%, $p < 0.05$), ht3 (25.0% vs. 75.0%, $p < 0.05$), ht4 (87.1% vs. 12.9%, $p < 0.05$), ht5 (67.8% vs. 37.2%, $p < 0.05$) haplotype frequencies significantly differed in the two groups. The ht1 (51.6% vs. 48.4%), ht6 (47.5% vs. 52.5%), ht7 (56.4% vs. 43.6%), ht8 (57.1% vs. 42.9%) showed no remarkable differences between the Roma and the Hungarian populations.

Table 1. Case-control genotypes and allele frequencies of variants in *IL23R* and IBD5

	UC (*n* = 320)	**Controls** (n=316)	**OR (95%CI)***	***P***
IL23R (rs1004819)				
GG	126 (39.4%)	158 (50.0%)		
GA	168 (52.5%)	134 (42.4%)		
GA+AA	194 (60.6%)	158 (50.0%)	**1.606 (1.160-2.223)**	**0.004**
AA	26 (8.1%)	24 (7.6%)	1.254 (0.696-2.261)	0.452
RAF	0.343	0.287		**0.032**
IL23R (rs2201841**)**				
TT	140 (43.8%)	155 (49.1%)		
TC	150 (46.9%)	143 (45.3%)		
TC+CC	180 (56.3%)	161 (51.0%)	1.268 (0.920-1.749)	0.147
CC	30 (9.4%)	18 (5.7%)	**1.983 (1.069-3.678)**	**0.030**
RAF	0.328	0.283		0.242
SLC22A4 (rs1050152)				
CC	93 (29.1%)	110 (34.8%)		
CT	159 (49.7%)	148 (46.8%)		
CT+TT	227 (71.0%)	206 (65.2%)	1.319 (0.935-1.86)	0.115
TT	68 (21.3%)	58 (18.4%)	1.150 (0.768-1.723)	0.498
RAF	0.460	0.417		0.120
SLC22A5 (rs2631367)				
GG	83 (25.9%)	89 (28.2%)		
GC	163 (50.9%)	156 (49.4%)		
GC+CC	237 (74.0%)	227 (71.9%)	1.138 (0.794-1.631)	0.481
CC	74 (23.1%)	71 (22.5%)	0.982 (0.669-1.440)	0.925
RAF	0.485	0.471		0.607
IGR2230a_1 (rs17622208)				
GG	87 (27.2%)	90 (28.5%)		
AG	160 (50.0%)	157 (49.7%)		
AG+AA	233 (72.8%)	226 (71.5%)	1.073 (0.751-1.532)	0.698
AA	73 (22.8%)	69 (21.8%)	0.990 (0.673-1.457)	0.960
RAF	0.478	0.466		0.685
IGR2198a_1 (rs11739135)				
GG	105 (32.8%)	117 (37.0%)		
GC	159 (49.7%)	150 (47.5%)		
GC+CC	215 (67.2%)	199 (63.0%)	1.260 (0.900-1.763)	0.179
CC	56 (17.5%)	49 (15.5%)	1.169 (0.760-1.797)	0.477
RAF	0.423	0.392		0.260
IGR2096a_1 (rs12521868)				
GG	101 (31.6%)	117 (37.0%)		
GT	164 (51.3%)	147 (46.5%)		
GT+TT	219 (68.5%)	199 (63.0%)	1.256 (0.897-1.760)	0.185
TT	55 (17.2%)	52 (16.5%)	1.045 (0.680-1.608)	0.840
RAF	0.428	0.397		0.262

[1]Adjusted for age and gender.
Associations significant at $P < 0.05$ vs controls are shown in bold.
RAF: Risk allele frequency; UC: Ulcerative colitis.

Table 2. Genotype-specific ulcerative colitis odds ratios (with 95%CI) for combinations of variants in *IL23R* and IBD5

	SLC22A4		***SLC22A5***		**IGR2230a_1**		**IGR2198a_1**		**IGR2096a_1**	
IL23R rs1004819	CC	CT+TT	GG	GC+CC	GG	AG+AA	GG	GC+CC	GG	GT+TT
GG	1	1.298 (0.782-2.156) *P* = 0.313	1	1.064 (0.623-1.815) *P* = 0.821	1	0.941 (0.556-1.593) *P* = 0.822	1	1.033 (0.623-1.711) *P* = 0.900	1	1.104 (0.666-1.831) *P* = 0.702
GA+AA	1.527 (0.872-2.673) *P* = 0.138	**2.015 (1.230-3.300) *P* = 0.005**	1.424 (0.776-2.614) *P* = 0.253	**1.691 (1.003-2.821) *P* = 0.048**	1.300 (0.716-2.359) *P* = 0.388	1.549 (0.927-2.588) *P* = 0.093	1.263 (0.736-2.166) *P* = 0.396	**1.803 (1.096-2.966) *P* = 0.020**	1.281 (0.744-2.206) *P* = 0.372	**1.911 (1.162-3.143) *P* = 0.010**
IL23R rs2201841										
TT+TC	1	1.328 (0.989-1.927) *P* = 0.058	1	1.254 (0.868-1.812 *P* = 0.228	1	1.184 (0.823-1.704) *P* = 0.363	1	1.296 (0.922-1.822) *P* = 0.136	1	1.385 (0.982-1.953) *P* = 0.063
CC	**3.413 (1.169-9.965) *P* = 0.018**	1.716 (0.788-3.739 *P* = 0.171	**3.946 (1.232-12.645) *P* = 0.014**	1.474 (0.679-3.200) *P* = 0.324	**3.777 (1.181-12.084) *P* = 0.018**	1.411 (0.652-3.056) *P* = 0.381	**3.165 (1.088-9.206) *P* = 0.027**	1.592 (0.735-3.449) *P* = 0.236	**2.977 (1.099-8.066) *P* = 0.026**	1.701 (0.765-3.784) *P* = 0.189

Associations significant at $P < 0.05$ *vs* controls are shown in bold.

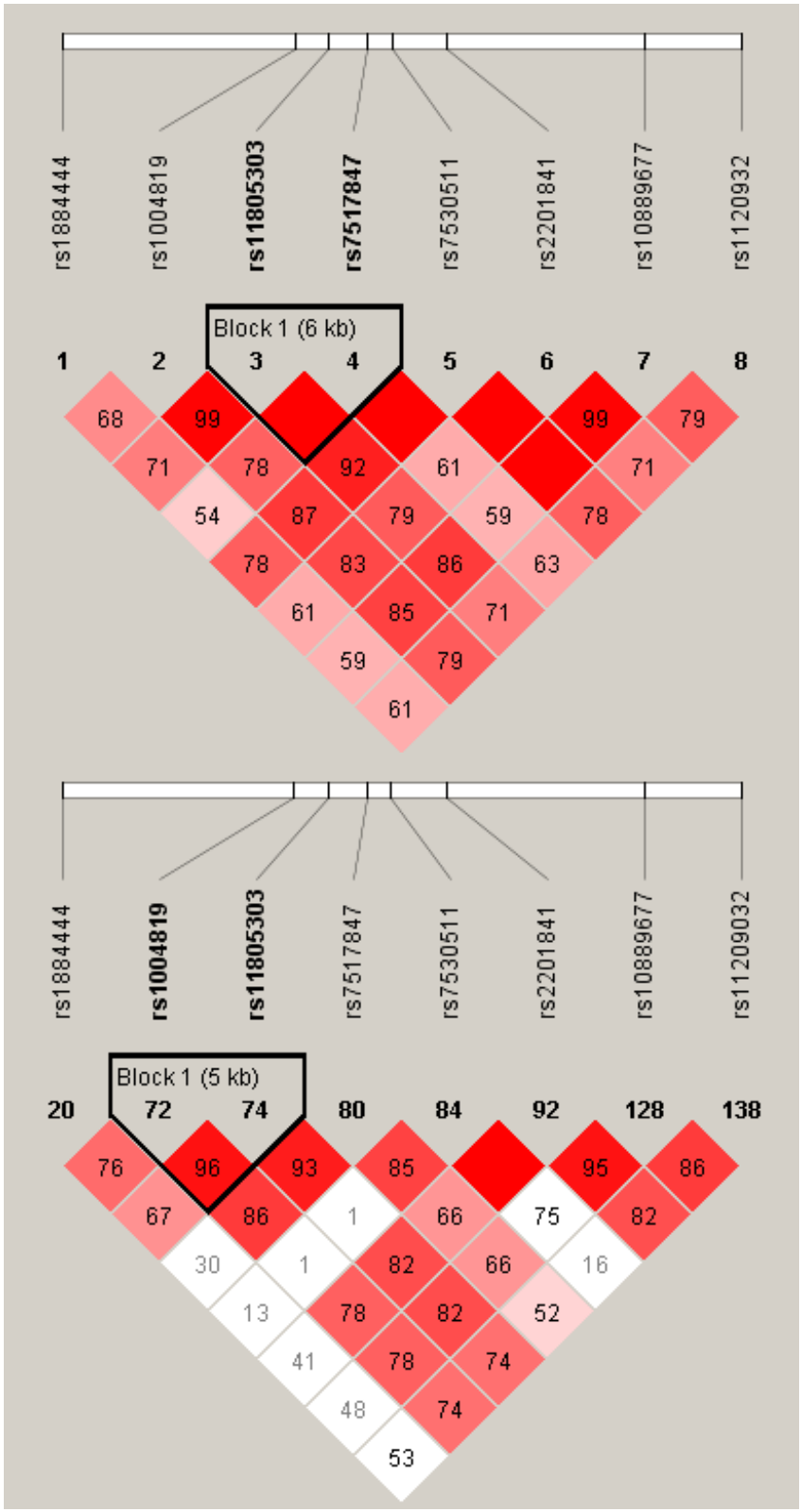

Figure 1. *IL23R* linkage disequilibrium map in Roma population (upper panel), and in Hungarian population samples (lower panel).
Diamonds represents pairwise linkage disequilibrium (D') for each SNP combinations. Shading represents the magnitude of pairwise LD; red, D'<1; bright red, D' = 1; white, D'< 1. The pentagons represent the haplotype blocks, the boldfaced SNPs are htSNPs. LD data show that there was 1–1 haplotype block within the *IL23R* gene region both in Romas and in Hungarians.

Table 3. Major haplotypes (ht) created by the examined *IL23R* variants

	rs1884444	rs1004819	rs11805303	rs7517847	rs7530511	rs2201841	rs10889677	rs11209032
ht1	G	G	C	T	C	T	C	G
ht2	G	G	C	T	T	T	C	G
ht3	G	G	C	G	C	T	C	G
ht4	G	A	T	T	C	C	A	G
ht5	G	A	T	T	C	C	A	A
ht6	T	G	C	T	C	T	C	G
ht7	T	G	C	T	T	T	C	G
ht8	T	G	C	G	C	T	C	G

Table 4. Haplotype frequencies of the examined *IL23R* variants

	Roma (%)	Controls (%)
ht1	51.6	48.4
ht2	14.3*	85.7
ht3	25.0*	75.0
ht4	87.1*	12.9
ht5	67.8*	37.2
ht6	47.5	52.5
ht7	56.4	43.6
ht8	57.1	42.9

* $P < 0.05$ vs controls

6. Discussion

6.1. Gene-gene interaction (IBD5-*IL23R*) analysis in Hungarian ulcerative colitis patients

Most of the identified genes have only modest effects on IBD susceptibility individually, suggesting that complex interactions are more important. Epistasis, defined generally as gene-gene interactions, has become a hot topic in complex disease genetics in recent years and can explain the lack of replication of single-locus results. Different statistical methods are used to investigate gene-gene interactions, for example one traditional approach still widely used today is the logistic regression analysis or the 2x2 allele-based χ^2 test.

Both the IBD5 and the *IL23R* genes have been identified originally as CD susceptibility genes, but their association with UC has also been confirmed. In previous Hungarian single-locus association studies our research group found no significant differences in the allele frequencies of *SLC22A4* and *SLC22A5* genes either in CD in pediatric and adult patients or in UC. The TC haplotype was not associated with a higher risk of CD and UC in the Hungarian population. The IGR2096a_1 and IGR2198a_1 markers on IBD5 locus were found to confer susceptibility to CD but not for UC. The distribution of IGR2230a_1 was not significantly different in the CD or the UC group compared with the controls. We observed an increased prevalence of the *IL23R* rs2201841 and rs1004819 in CD in previous Hungarian studies. In our recent work, besides confirming the negative association for IBD5 loci in UC we could detect significantly higher allele frequency for the *IL23R* rs1004819 and increased prevalence of the homozygous rs2201841 CC genotypes in UC patients compared to controls in Hungarian population.

In CD the interactions between the main susceptibility genes are better characterized than in UC. Multifactor dimensionality reduction analysis (MDR) suggested an interaction between IBD5, *ATG16L1*, and *IL23R* risk alleles in CD patients from Manitoba IBD Research Registry. Weersma et coworkers observed multiple gene combinations (*ATG16L1*, *IL23R*, *CARD15*, IBD5 and *DLG5*). According to their results an association between the increase in the number of risk alleles and an increased risk for the development of CD and a more severe disease course was found. In the study of Cummings et al with the exception of rs11209026, *IL23R* risk polymorphisms showed significant CD association only in the subgroup of persons positive for the IGR2060a_1 variant in IBD5. This result may suggest that the *IL23R* gene influences CD in tandem with effects of the IBD5 haplotype.

When we analyzed the possible statistical interactions by pairs of risk-conferring *IL23R* variants and IBD5 using binary logistic regression, no evidence of interaction was found between these seven markers, suggesting that all the examined loci are independent factors. In contrast with the single gene effects, after genotype stratification, the *IL23R* rs2201841 CC variant confers risk for UC only on a wild-type IBD5 background and the rs1004819 A allele in combination with IBD5 carrier status except of IGR2230a_1.

The genetic variants identified during GWAS individually confer relatively small risk with estimated odds ratios typically in the range of less than 1.5 and explain only a small proportion of heritability (only 20-25%), suggesting that gene-gene interactions as well as gene-environmental interactions could play a key role in IBD pathogenesis and fill the so called “genetic vacuum” of polygenic diseases.

The future lies in understanding the etiology of epistasis at the biological level and understanding pathways underlying disease and using that knowledge to develop strategies for therapy and prevention. In conclusion, our study has shown that UC susceptibility genes are likely to act in a complex interactive manner similar to CD. Our results play important role in the understanding of the pathogenesis of UC and areas of overlap with CD but further studies are needed to confirm them.

6.2. *IL23R* haplotype analysis in healthy Roma and Hungarian population

The *IL23R* gene investigation was initiated by Duerr, who found an association between the *IL23R* gene and IBD, and reported several independent functional SNPs of the gene and its neighboring region, which confer susceptibility (rs10889677, rs1004819, rs2201841, rs11805303, rs11209032) or protection (rs7517847) against CD and UC in non-Jewish subjects. Two variants, the rs7530511 and rs1884444, showed no disease association. After the primary publication, numerous replication studies have been published in different ethnical populations. Later, the associations were extended to other autoimmune diseases (psoriasis, ankylosing spondylitis, Sjögren's syndrome and systemic lupus erythematosus). Very recently, a study of association of *IL23R* haplotypes with some of these diseases has also been initiated.

Here we examined five susceptible, one protective and two neutral variant of the *IL23R* gene in single-locus association analysis, and found significant increased genotype and allele frequencies for rs10889677, rs1004819, rs2201841, rs11805303, rs11209032 in Roma samples compared with the Hungarian population, and the rs7517847 showed significantly decreased genotype and allele frequencies in the Roma samples compared to the Hungarians. Hypothetically, while the former variants associate with disease risk based on observations of other ethnicities, and the later one often confer protection against diseases, these changes independently and in combination are strongly suggestive for increased disease-risk in Roma people. However, at the moment there is no evidence available for the higher prevalence rate of any of the known autoimmune disease in Roma people.

Here we also determined the major haplotypes, and the ht4, ht5 showed significantly increased haplotype frequencies, while ht2, ht3 showed decreased haplotype frequencies when comparing the results of the Roma population to the Hungarians. In addition to this, linkage analysis revealed major differences, including strong linkage disequilibrium between rs11805303-rs7517847 in the Roma population, and rs1004819-rs11805303 in the Hungarian population. Taken together the unique distribution of SNPs, the linkage of them, the haplotype profiles strongly suggest unique susceptibility of Roma people, this needs further investigations.

In summary, our results emphasize that complex interactions are more important than the independent main effects of any single susceptibility gene among common human diseases like IBD. Identification of gene-gene interactions and determination of genotype-phenotype associations might contribute to the recognition of risk factors, to early diagnosis and efficient prevention of the disease. Haplotypes may also offer additional information for risk predictions.

In the future, simultaneous analysis of SNPs hopefully will lead to better explanation of the so-called "missing heritability" of common traits. The knowledge of the complex genetic background may prove to be greatly useful in the development of novel therapies and to implement personalized medicine based on the individual's genetics.

7. Thesis

1. In our recent work, we identified the *IL23R* rs1004819 as susceptibility factor for UC in Hungarian patients. For the *IL23R* rs1004819 A allele we found significantly higher allele frequency in UC patients compared to control subjects and it showed significant association with UC risk for carriers.

2. We could detect significantly increased prevalence of the homozygous rs2201841 CC genotypes in UC patients compared to controls in Hungarian population.

3. In single-locus analysis, no significant association for any variants of IBD5 region (IGR2096a_1 rs12521868 G/T, IGR2198a_1 rs11739135 G/C, IGR2230a_1 rs1762208 G/A, *SLC22A5* rs263136 G/C, *SLC22A4* rs1050152 C/T) and UC was observed.

4. During the search for possible statistical interaction between the risk-conferring *IL23R* and IBD5 variants using binary logistic regression analysis by pairs, no evidence of interaction was found between these seven markers, suggesting that all the examined loci are independent factors.

5. We could detect a positive association on the background of *IL23R* rs1004819 A allele for *SLC22A4* T allele, *SLC22A5* C, IGR2198a_1 C or IGR2096a_1 T allele, when stratifying the *IL23R* genotypes by IBD5 genotypes. There was no association with UC for any combinations of *IL23R* rs1004819 and IGR2230a_1.

6. The *IL23R* rs1004819 A variant (GA heterozygous and AA homozygous together) did not show significant association with UC on the background of all wild type IBD5 genotypes during combined interaction analysis in Hungarian UC population.

7. For the combinations of IBD5 loci and *IL23R* rs2201841 we could detect significant association only in carriers of rs2201841 CC homozygote alleles and wild type IBD5 genotypes. The IBD5 carrier status did not confer susceptibility for UC either in the presence of *IL23R* rs2201841 TT + TC or CC genotypes.

8. We found difference in the prevalence rate of GGCTTTCG, GGCGCTCG, GATTCCAG and GATTCCAA haplotypes of *IL23R* rs1004819, rs2201841, rs10889677, rs11805303, rs11209032, rs7517847, rs7530511 and rs1884444 between the healthy Roma and the Hungarian populations.

9. We could not detect difference in the distribution of the GGCTCTCG, TGCTCTCG, TGCTTTCG and TGCGCTCG haplotypes between the two, ethnically different healthy Roma and the Hungarian populations.

8. Publications

8.1 Publications supporting the dissertation

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

2. 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;65(2):148-52. **IF: 2.518**

8.2. Other publications

1. Csizmadia Cs, **Sarlos P**, Nagy L, Kiraly A: Sósavfüggő kórképek. *Granum* 2004;7(4):37-40.

2. **Sarlos P**, Kiraly A, Nagy L: A Peutz-Jeghers szindrómáról családvizsgálatok kapcsán. *Orvosi Hetilap* 2007;148(6):255-258.

3. Lakner L, Csongei V, Magyari L, Varga M, Miheller P, **Sarlos P**, Orosz P, Bari Z, Takács I, Járomi L, Sáfrány E, Sipeky C, Bene J, Tulassay Z, Döbrönte Z, Melegh B: Possible role of selected IGR and SLC22A4/SLC22A5 loci in development of inflammatory bowel diseases. *Orvosi Hetilap* 2009;150(29):1375-80.

4. Lakner L, Csongei V, **Sarlos P**, Jaromi L, Safrany E, Varga M, Orosz P, Magyari L, Bene J, Miheller P, Tulassay Z, Melegh B: IGR2096a_1 T and IGR2198a_1 C alleles on IBD5 locus of chromosome 5q31 region confer risk for Crohn's disease in Hungarian patients. *Int J Colorectal Dis.* 2009;24(5):503-7. **IF: 2.102**

5. Csongei V, Jaromi L, Safrany E, Sipeky C, Magyari L, Faragó B, Bene J, Polgár N, LaknerL, **Sarlos P**, Varga M, Melegh B. Interaction of the major inflammatory bowel disease susceptibility alleles in Crohn's disease patients. *World J Gastroenterol*. 2010;16(2):176-83. **IF: 2.240**

6. Csongei V, Jaromi L, Safrany E, Sipeky C, Magyari L, Polgar N, Bene J, **Sarlos P**, Lakner L, Szabo M, Rappai G, Melegh B. Interaction between CTLA4 gene and IBD5 locus in Hungarian Crohn's disease patients. *Int J Colorectal Dis*. 2011;26(9):1119-25. **IF: 2.385**

7. Par G, Trosits A, Pakodi F, Szabo I, Czimmer J, Illes A, Godi S, Bajor J, **Sarlos P**, Kenyeres P, Miseta A, Vincze A, Par A. Transient elastography as a predictor of oesophageal varices in patients with liver cirrhosis. *Orvosi Hetilap* 2014;155(7):270-276.

8. Magyari L, Kovesdi E, **Sarlos P**, Javorhazy A, Sumegi K, Melegh B. Interleukin and interleukin receptor gene polymorphisms in inflammatory bowel diseases susceptibility. *World J Gastroenterol.* 2014;20(12):3208-3222. **IF: 2.547**

8.3. Citable abstracts

1. Nagy Zs, Koszo F, Par A, Emri G, Horkay I, Horányi M, Karadi O, **Sarlos P**, Morvay M, Varga V, Dobozy A, Mozsik Gy: Haemochromatosis (HFE) gene mutations and hepatitis C virus (HCV) infection as risk factors for porphyria cutanea tarda. *Gastroenterology* 2002;122(Suppl. 1):308. **IF: 13.44**

2. Nagy Zs, Par A, **Sarlos P**, Nagy A, Karadi O, Mozsik Gy: Haemochromatosis (HFE) gén mutáció vizsgálata hepatitis C vírus (HCV) infekcióban. *Magyar Belorvosi Archivum* 2001;54(S2):80.

3. **Sarlos P**, Kiraly A, Nagy L: Klinikai jellemzők, malignomák kockázata és a gondozás jelentősége Peutz-Jeghers szindrómában (PJS). *Magyar Belorvosi Archivum* 2004;57(S1):115.

4. Mozsik Gy, **Sarlos P**, Racz I, Szolcsanyi J: Evidence for the gastric protective effect of capsaicine in human subjects. *Gastroenterology* 2003;124(Suppl. 1):A454. **IF:12.718**

5. **Sarlos P**, Rumi Gy, Szolcsanyi J, Mozsik Gy, Vincze A: Capsaicin prevents the indomethacin-induced gastric mucosal damage in human healthy subjects. *Gastroenterology* 2003;124(Suppl. 1):A511. **IF:12.718**

6. **Sarlos P**, Kiraly A, Nagy L: Peutz-Jeghers syndrome: clinical characteristics, risk of malignancies and importance of surveillance. *Zeitschrift für Gastroenterolgie* 2004;42(5):128. **IF:1.000**

7. **Sarlos P**: A HSPCO34/humán placentáris protein 25 kimutatása colorectalis carcinomában. *Magyar Belorvosi Archivum* 2005;58 (S1):52-53.

8. Lukacs M, Illes A, Undi S, Csizmadia Cs, **Sarlos P**, Weninger Cs, Kassai M, Kiraly A: Gender differences in the symptoms and findings os patients with chronic constipation. *Magyar Belorvosi Archivum,* 2005;58 (S1):53-54.

9. Illes A, Undi S, Csizmadia Cs, Nagy L, **Sarlos P**, Kiraly A: Effect of dietary fat to the gastric emptying in patients with gastroesophageal reflux disease. *Magyar Belorvosi Archivum,* 2005;58 (S1):105-106.

10. **Sarlos P**, Szigeti A, Bellyei Sz, Sumegi B, Nagy L, Kiraly A: The presence of HSPCO34 in colorectal cancer. *Zeitschrift für Gastroenterolgie* 2005;43(5):112. **IF:0.800**

11. Lukacs M, Csecsei P, Csizmadia Cs, Illes A, Hegedüs D, **Sarlos P**, Nagy L, Kiraly A: Gastrointestinal tract haemorrhage during concomittant long-lasting oral anticoagulant therapy in the elderly. *Magyar Belorvosi Archivum* 2006;59(S2):108-109.

12. Szelestei T, Vas T, Szigeti N, Degrell P, Berki T, **Sarlos P**, Wittmann I, Nagy J, Kovacs T: Anti-saccharomyces cerevisiae antibodies are raised in IgA nephropathy. *Nephrol Dial Transplant* 2007;22(S6):285. **IF: 3.154**

13. Lakner L, Csongei V, **Sarlos P**, Jaromi L, Safrany E, Varga M, Magyari L, Miheller P, Tulassay Zs, Dobronte Z, Melegh B: Az 5Q31 régióban elhelyezkedő IBD5 gén IGR2096_1 T és IGR2198A_1 C allélek hajlamosító szerepe Chron-betegség kialakulásában. *Magyar Belorvosi Archivum* 2008;61(S3):79.

14. **Sarlos P**, Illes A, Solt J, Nagy L, Kiraly A: Thymus carcinoma–associated motility disorders: achalasia and gastroparesis caused by myenteric ganglionitis. *Zeitschrift für Gastroenterolgie* 2008;46(05):A89. **IF:0.880**

15. **Sarlos P**, Acel P, Csizmadia Cs, Illes A, Kiraly A, Nagy L: Secondary aortoenteric fistula: three 3 case reports and review. *Zeitschrift für Gastroenterolgie* 2009;47(05):480. **IF:1.188**

16. **Sarlos P**, Pakodi F, Szabó I, Nemes Zs, Peterfi Z, Vincze A: Role of colonoscopic decompression in the treatment of toxic megacolon secondary to clostridium difficile colitis *Zeitschrift für Gastroenterolgie* 2013;51(05):A65. **IF:1.408**

17. Magyari L, Varszegi D, **Sarlos P**, Jaromi L, Bene J, Duga B, Hadzsiev K, Kisfali P, Komlosi K, 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 population samples. Eur J of Human Genetics 2013,21(2):565. **IF:4.319**

18. Par G, Trosits A, Pakodi F, Szabo I, Czimmer J, Illes A, Godi S, Bajor J, **Sarlos P**, Kenyeres P, Vincze A, Par A: Liver stiffness measurement selects patients with chronic liver diseases at risk of bearing large oesophageal varices. *Zeitschrift für Gastroenterolgie* 2013;51(05):A52. **IF:1.408**

Summary:
Impact factor of publicated papers: 14.339
Impact factor of citable abstracts: 53.033

9. Acknowledgements

The research, which serves as a basis of the doctoral dissertation, was carried out at the Department of Medical Genetics and the 3rd Department of Internal Medicine at the University of Pécs.

First of all, I would like to express my greatest thanks to my supervisor, Professor Béla Melegh for rendering it possible for me to join the Ph.D. program; for his confidence and his supervision of my work, for the leadership and useful advices.

I am grateful to my co-supervisor, Professor Lajos Nagy for his guidance and advices during my scientific and clinical work, and for his continuous support.

I would like to thank for each colleague in the laboratory especially for Lili Magyari, Veronika Csöngei and Erzsébet Kövesdi who helped me with their adept and conscientious work and professional experience and provide some unforgettable evening during performing statistical analyses and writing publications.

I owe a special thank to all the co-authors in our publications for their help and suggestions during DNA collection and result evaluation. I am really thankful to Jermásné Margitka for her help and assistance in the laboratory work.

Finally, I express my warmest thanks to my parents, my husband, my children and my whole family for their love and encouraging support.