SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D.)

Genomic and Genetic Analysis of Chronic Obstructive Pulmonary Disease and Inflammatory Bowel Diseases

By Szilárd Póliska

Supervisor: Professor László Nagy MD, PhD

UNIVERSITY OF DEBRECEN

DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

DEBRECEN, 2011

Genomic and Genetic Analysis of Chronic Obstructive Pulmonary Disease and Inflammatory Bowel Diseases

By Szilárd Póliska biologist M.Sc.

Supervisor: Professor László Nagy M.D. Ph.D

Doctoral School of Molecular Cell and Immune Biology, University of Debrecen

Head of the **Examination Committee**:	Prof. Dr. Róza Ádány
Members of the Examination Committee:	Prof. Dr. Margit Balázs
	Dr. Zoltán Prohászka

The Examination takes place at Department of Biochemistry and Molecular Biology, Medical and Health Science Center, University of Debrecen 6th of September 2011.

Head of the **Defense Committee**:	Prof. Dr. Róza Ádány
Reviewers:	Prof. Dr. Margit Zeher Dr. Vladimir Benes
Members of the Defense Committee:	Prof. Dr. Margit Balázs
	Dr. Zoltán Prohászka

The Ph.D. Defense takes place at the Lecture Hall of the 1st Department of Medicine, Institute for Internal Medicine, Medical and Health Science Center, University of Debrecen 6th of September 2011.

1 INTRODUCTION

Chronic inflammatory diseases have shown increasing incidence in the past thirty years, afflicting millions of people worldwide and placing an elevated burden on health-care services in both industrialized and developing countries. Allergic conditions (asthma, eczema), as well as disorders with autoimmune origin (rheumatoid arthritis (RA), psoriasis and inflammatory bowel diseases (IBD)), chronic obstructive pulmonary disease (COPD) and neurodegenerative disease are part of this group of conditions. The general cause of these conditions is shared, it is an unregulated inflammatory process, which affects specific organ systems or cause systemic inflammation.

Genetic predisposition is one of the major causes of the development of chronic inflammatory diseases; therefore it is very important to identify genetic susceptibility loci. This knowledge may lead to the identification of the pathomechanism of the inflammatory process, to improve new effective therapies and better prediction of the development of the disease. Sequencing of the human genome allows identification of the genetic variation such as single nucleotide polymorphisms (SNP) and copy number variations. Recent development in genomic technologies, including the extensive use of microarray technology, provided an opportunity to perform global gene expression analyses. Microarray experiments allow us to compare the transcriptome of tissue samples or distinct cell types in order to establish correlations between disease states and/or to stratify disease.

1.1 COPD

COPD is an increasing global health problem and it is predicted to become the third most common cause of death and the fifth most common cause of disability in the world by 2020. It is one of the main causes of mortality in Europe and Hungary is one of most affected country. The age-adjusted rates for men are the highest in Hungary among the European Union and the Central East European countries according to data published by the European Respiratory Society

However, COPD and asthma involve chronic inflammation of the respiratory tract, but there are many differences in the site of inflammation and in the involved inflammatory cell types. Although in the past few years the pathomechanism of asthma has been examined widely resulting development of new effective therapies, but in the case of COPD the exact molecular

mechanisms which leading to the development of the disease and effective treatment are still remained unknown yet.

COPD is characterized by slowly progressive development of airflow limitation, which is mostly irreversible, in contrast to asthma where airway obstruction is usually reversible. The airflow limitation is associated with an abnormal inflammatory response to noxious particles or gases in the lungs. COPD is accompanied by emphysema, chronic bronchitis with airway obstruction and small airway disease, these are distinct phenotypes of COPD, but most patients show a combination of these phenotypes.

There are several risk factors of COPD including life and work circumstances, environmental pollution and aging, but the major risk factor of the development of COPD is cigarette smoking. Due to the fact, that only 15-20% of smokers develop COPD suggests the involvement of an additional risk factor, the genetic susceptibility.

Due to the complexity of COPD as an inflammatory disease, several types of inflammatory cells are involved in the development of the disease. Increased number of implicated cell types, such as alveolar macrophages (AM), neutrophils, dendritic cells, T-lymphocytes and B-lymphocytes are observed in alveoli and small airways. AMs seem to play central role in the pathogenesis of COPD and the maintenance of the inflammatory process ad they are the major host defense cells in the lower airspace. Cigarette smokes exposure causes more than fivefold increase in total cells in brochoalveolar lavage fluid (BALF), 95-98% of these cells are macrophages. The precursors of AMs are monocytes (MO). MOs are released from bone marrow and are homing to the lungs in response to cigarette smoke, but higher rate of MO recruitment from peripheral blood was also observed. AMs secrete large number of inflammatory mediators including chemokines, cytokines, growth factors, lipid mediators, reactive oxygen and nitrogen species.

Although inflammation is beneficial in the defense response against pathogens and external stimuli, without control it can lead to the pathogenesis of common chronic inflammatory diseases, such as RA, type 2 diabetes, multiple sclerosis and COPD. Complex network of signaling mediators are involved in this process, many of them are under the transcriptional control of transcription factors (TF); nuclear factor κB (NF-κB), activator protein 1 (AP-1) glucocorticoid receptor (GR) and peroxisome proliferator-activated receptor gamma (PPARγ).

PPARγ is a member of nuclear receptor superfamily, abundantly expressed in adipose tissue and in many other tissues and different types of cells including the lungs and AMs. Four variants of the PPARγ mRNA are known, the PPARγ1, PPARγ2, PPARγ3 and PPARγ4, which arise due to alternative splicing/promoter usage, and they show tissue specidfic distribution. PPARγ can be activated by fatty acid derivatives, such as 15-deoxy-$\Delta^{12,14}$ -prostaglandin J_2 (15d-PGJ_2), which are natural ligands of PPARγ and by synthetic ligands such as thiazolidinediones (TZDs), for example rosiglitazone. It plays a central role in adipogenesis, intracellular insulin signaling and regulation of cellular energy homeostasis and it is involved in alternative macrophage activation and DC biology. PPARγ also has an anti-inflammatory effect, inhibiting the expression of pro-inflammatory genes by interfering with NF-κB and AP-1 or inducing transcription of anti-inflammatory genes.

Development of complex genetic diseases, such as COPD is associated with the interaction of environmental factors and genetic susceptibility. Genetic association and genome-wide studies have identified several genes and their SNPs that might be involved in the pathogenesis of COPD. Since the protease-antiprotease and oxidant-antioxidant balance are crucial in COPD, most of the associated genes functions correlate with these two pathways. However SNPs of other candidate genes, which are part of the inflammatory process or play role in immune response, have also been found to show association with COPD.

SERPINE2 gene, a member of serpin peptidase inhibitors, was identified as having potential role in COPD. Previously multiple SNPs of this gene were examined in two huge COPD case-control studies from North America, in the Boston early-onset COPD cohort and in patients from the NETT cohort. Several SNPs of SERPINE2 were significantly associated with the disease in these studies. However case-control studies could not replicate these associations with COPD in European patients, which questioned the validity of the results found and raised a geographical distribution of specific SNPs.

Microsomal epoxide hydrolase (EPHX1) is an enzyme associated with the metabolism and detoxification of xenobiotic chemicals; it plays an important role in the general oxidative defense of lung. Several polymorphisms are known in EPHX1 including two relatively common SNPs, the exon 3 Tyr113His (rs1051740) and exon 4 His139Arg (rs2234922) variants. These two variants of alleles have been suggested to be associated with altered EPHX1 enzyme activity. Substitution of Tyr113 for His decreases (slow allele), whereas substitution of His139

for Arg (fast allele), increases EPHX1 activity. Several studies found the slow metabolizing form of EPHX1 to be associated with an increased risk for COPD.

Tumor necrosis factor alpha (TNF-α) is a key factor to initiate and mediate the inflammation in both the respiratory tract and systemic features of COPD. Several polymorphisms of TNF-α were examined in case-control studies but the most studied is a promoter variant (-308G>A), which increases the production of TNF-α through directly affecting the gene transcription. The -308G>A SNP is linked to increased airway inflammation in Taiwanese patients with chronic bronchitis and to airflow obstruction in subjects without chronic bronchitis and severity of emphysema in a Japanese study, but not in Caucasians.

The surfactant proteins contribute to the regulation of surface tension in the alveoli and also have a role in the host defense and control of inflammation. A SNP in surfactant protein B, which causes amino acid change (Thr131Ile), has been associated with COPD in Boston Early-onset COPD cohort and in a case-control study from Mexico.

Since PPARγ is involved in the regulation of inflammatory signaling pathways of various autoimmune disorders, therefore polymorphisms of PPARγ could be implicated in the susceptibility to these diseases. Association of PPARγ SNP with type 2 diabetes was reported in several studies. The minor allele of Pro12Ala, which causes lower binding affinity of PPARγ protein to peroxisome proliferator response element, was associated with lower body mass index, improved insulin sensitivity and decreased risk of type 2 diabetes. Three SNPs of PPARγ were examined in asthma (Pro12Ala, His447His, and C-681G). The combination of the major alleles of Pro12Ala and His447His was associated with increased risk for asthma exacerbations; in another study this haplotype combination was associated with aspirin sensitivity in asthmatic patients.

Several recent studies attempt to identify novel pathways or gene sets involved in the pathogenesis of COPD using gene expression profiling analyses. Strikingly there is only a minimal overlap between the differentially expressed genes among the different datasets. This variability could be derived from several factors, such as differences in sample collection and size, disease severity, to origin of the examined tissue or cell components and using different expression platforms. Beyond the variability of the studies, these findings have presented useful information to the pathogenesis of COPD.

1.2 IBD

IBD is a chronic inflammatory disorder with autoimmune origin affecting the gastrointestinal tract. It is clinically classified into two major phenotypes: ulcerative colitis (UC) and Crohn's disease (CD). UC is characterized by diffuse mucosal inflammation, which extends proximally from the rectum to a varying degree, and by the presence of a significant number of neutrophils within the lamina propria and the crypts, where they cause formation of micro-abscesses. The severe inflammatory process with the production of inflammatory mediators leads the development of extensive superficial mucosal ulceration. Whilst, CD may be patchy and segmental, the inflammation is usually transmural. Any site of the gastrointestinal tract could be affected, but the involvement of the terminal ileum is the most common and the earliest mucosal lesions often appear over Peyer's patches. It is characterized by the aggregation of macrophages that form non-caseating granulomas.

In recent years CD and UC have shown increasing frequency similar to COPD, imposing a significant burden on the healthcare system. Although the etiology of these diseases still remains unknown, the cause of the disease is better understood, a dysregulated mucosal immune response to the normal intestinal microflora in a genetically susceptible host is involved in the pathogenesis.

Several studies indicate that genetic factors appear to predispose some individuals to the development of IBD. Recent single locus genetic association and genome-wide association studies (GWAS) have identified several new susceptibility genes and chromosomal loci, such as NOD2/CARD15, IL23R or ATG16L1 that were associated with CD or UC. NOD2/CARD15 is an important receptor for the intracellular recognition of bacterial products such as lipopolysaccharid and peptidogylcan and has a role in signaling innate defense response signaling in the host. IL23 pathway might cause an aberrant immune response to microbial encounter and showed that the depletion of IL23 was associated with decreased proinflammatory responses in the intestine. ATG16L1 plays role in autophagy that is important in the trafficking of antigens for immune recognition.

PPARγ is strongly expressed in the colon, mostly by the epithelial cells, indicating that it may have an important function in the normal and pathological processes of the colon. Animal experiments also provided support for a role in the inflammation of the intestine for PPARγ. 5-aminosalicylic acid, an anti-inflammatory agent, that is widely used to treat IBD, is a PPARγ

ligand as well. Rosiglitazone, another PPARγ activator, was tested in clinical trials and was found to be effective in the treatment of UC. Despite its suggested role as a repressor of inflammatory responses, very few studies have assessed so far the association between the known polymorphisms of PPARγ and IBD so far.

2 AIMS

Because of the lack of systematic genomic analysis of COPD and IBD in Hungary, we decided to perform global gene expression and genetic association studies in the Hungarian population.

The objectives of our studies

- To expand and validate the list of genes associated with smoking-induced COPD in AMs.
- To find and validate potential biomarkers in peripheral blood MOs that reflect COPD-specific gene expression patterns identified in AMs.
- To correlate gene expression patterns with disease parameters.
- To investigate the association of EPHX1, SERPINE2, TNF-α and SFTBP polymorphisms to COPD in a Hungarian population.
- To assess possible association between SNPs of PPARγ (rs10801282 (Pro12Ala), rs3856806 (His447His) and rs1800571 (Pro113Gln)), a new candidate gene, and COPD outcome.
- To examine the association of four polymorphisms of PPARγ (rs10865710 (C-681G), rs2067819, rs3892175 and Pro12Ala) with CD and UC in a cohort of Hungarian patients.

3 METHODS

3.1 Informed consent

The Research Ethics Committee of the University of Debrecen Medical and Health Science Center approved the clinical protocol and the study. All healthy and diseased individuals of this study have been recruited from Hungary (Caucasians) and signed informed consent was obtained from all patients.

3.1.1 INCLUSION/EXCLUSION CRITERIA FOR COPD STUDY

Inclusion and exclusion criteria of diseased and healthy patients were defined before starting sample collection. All subjects were between 40 and 65 years old, COPD patients had FEV_1 predicted < 80% and FEV_1/FVC% < 70%, control patients had FEV_1 predicted ≥ 80% and FEV_1/FVC% ≥ 70%. Control patients had no history of cancer, IBD and other chronic inflammatory or autoimmune diseases (e.g. RA, psoriasis, SLE).

3.1.2 INCLUSION/EXCLUSION CRITERIA FOR IBD STUDY

675 well-characterized IBD patients were recruited from four locations by four centers of the Hungarian IBD Study Group. The diagnosis was based on the Lennard-Jones criteria and the disease phenotype was determined according to the Montreal Classification. Only those patients were enrolled who had a confirmed diagnosis for more than 1 year.

486 sex- and age-matched healthy control blood donors were recruited, without history of cancer, IBD and other chronic inflammatory or autoimmune diseases (e.g. RA, psoriasis, SLE).

3.2 AM and peripheral blood MO collection

BALF samples of healthy and COPD patients were collected by fiber-optic bronchoscopy. AMs were separated by Percoll gradient centrifugation. Total cell number was determined by counting in hemocytometer.

50 ml heparin treated venous blood was collected from healthy and diseased patients. MOs were separated by a two-step procedure, 1) PBMCs were separated by Ficoll gradient centrifugation,

2) MOs were separated from PBMCs by positive selection using anti-CD14 conjugated magnetic beads.

3.3 RNA and microarray processing

Total RNA was extracted from alveolar macrophages and peripheral blood monocytes using the RNeasy Mini Kit. RNA integrity was checked on Agilent Bioanalyser 2100, NanoDrop ND-1000 was used to determine the concentration. Affymetrix HG-U133A arrays containing 22283 probe sets were used to perform gene expression analysis. Complementary DNA (cDNA) was generated from 5 µg total RNA, using SuperScript choice system. Biotin labeled cRNA was synthesized using BioArray™ HighYield™ RNA Transcript Labeling Kit (T7), 20 µg labeled cRNA samples were hybridized and scanned by EMBL Genomics Core Facility (Heidelberg). AM samples were hybridized separately, but in the case of MOs a control and a COPD RNA mix from 5-5 patients were created. AM and MO microarrays were analyzed in separate experiments using GeneSpring 7.3 software.

3.4 TaqMan mRNA analysis by RT-QPCR

In order to confirm the results of the microarray analysis RT-QPCR measurements were performed. Reactions were run in ABI Prism HT 7900 instrument. Relative gene expression levels were calculated by ΔΔ Ct method with GAPDH as the internal control. Statistical analysis was performed by GraphPad Prism software.

3.5 Canonical variates analysis (CVA)

Separation between predefined groups of objects is best revealed by CVA. It is used to determine whether the groups of controls and COPD patients are separable in the multidimensional space spanned by the genetic variables, and if so, which gene subsets have the best discriminatory power. The results of CVA are the so-called canonical scores obtained from the canonical functions derived through eigenanalysis, which serve as coordinates of observations in the canonical space.

3.6 DNA isolation

Venous blood samples were collected in tubes containing EDTA. Genomic DNA was extracted from 0.5 ml whole blood using Roche MagNa Pure LC instrument and MagNa Pure

LC DNA Isolation Kit according to the manufacturer's protocol. DNA concentration was checked by UV photometry using NanoDrop ND1000 instrument.

3.7 SNP genotyping

SNPs were genotyped using TaqMan SNP genotyping assays. Measurements were performed on ABI Prism 7900 HT instrument and genotypes were assigned using the allelic discrimination option of SDS 2.1 software. The average genotyping success rate of at least 95% was attained for each SNP.

Differences between cases and the control group concerning demographic and main clinical data were analyzed by the Mann-Whitney *U*-test or by the Pearson $\chi 2$ test. Genotype data for each SNP were tested for departures from Hardy-Weinberg equilibrium (HWE) separately in case and control populations using a goodness-of-fit χ^2-test or the exact test to estimate P values. HWE calculations were done by using the HWE tool offered on http://ihg.gsf.de/cgi-bin/hw/hwa1.pl.

Haplotype frequencies were reconstructed from unphased genotypes for control and patient groups separately with the Full-Precise-Iteration algorithm implemented in the SHEsis software (http://analysis.bio-x.cn/myAnalysis.php). The extent of linkage disequilibrium (LD) between pairs of biallelic markers was determined using both the standardized disequilibrium and correlation coefficients (given as Lewontin's D' and r^2, respectively) and global association of haplotypes with CD and UC was assessed by a χ^2-test or the exact test as implemented in the program SHEesis. We have used conservative Bonferroni correction to correct for multiple testing. *A posteriori* estimates of the power of the study were assessed by the Quanto software (http://hydra.usc.edu.gxe).

4 RESULTS

4.1 Gene expression analysis of COPD

BALF samples were collected from 5 controls and 5 COPD patients all of whom were active smokers. The control and COPD groups were homogenous regarding age and smoking habits, but differed in lung function parameters and CRP levels.

AMs isolated from BALF were subjected to global gene expression analysis using microarrays. A list of 671 probe sets representing 389 annotated genes, which showed significant correlation with the disease were identified. Hierarchical cluster analysis was performed and these 671 probe sets separated completely healthy controls from COPD patients.

Comparison of the transcripts expressed differently in disease versus control samples revealed some previously identified COPD associated genes, including CYP1B1, a cytochrome P450 enzyme known to be associated with smoking, and induced by cigarette smoke. Chemokine ligand 2, osteopontin (SPP1), IL-8 and IL-1β genes are involved in immune response, chemotaxis, and cell migration. These findings validated our data set and suggest that our novel findings can complement already existing markers and very likely to link additional genes and pathways to COPD.

We carried out gene ontology (GO) analysis of the 389 genes and found that the identified genes have fallen into several of those pathways that are believed to be related to COPD pathogenesis and inflammatory processes, such as response to stress, immune response, cell death, apoptosis.

In order to confirm and validate the microarray results, the gene expression levels of 96 selected genes were measured using RT-QPCR in a new cohort of 46 patients including 26 COPD and 20 healthy controls. All individuals were active or ex-smokers in the validation set. Our custom TLDAs contained 96 genes including 4 housekeeping genes, 23 candidate genes obtained from the literature and 69 genes chosen from our exploratory microarray experiment.

Twenty-four genes showed significantly different expression patterns between control and COPD patients. Twenty-three genes changed in the same direction both on the microarrays and in the RT-QPCR measurements**.** We performed an independent statistical approach, CVA using these 23 validated genes in order to test whether this gene set can indeed separate COPD

patients from controls. Canonical correlation for this gene set was 0.8 (χ^2 =34.25, d.f. =20, significant at p=0.05, Wilks λ =0.365), which suggests that such a set can differentiate between healthy and COPD patients.

Next we decided to turn our attention to peripheral monocytes and explore their expression changes between COPD and control samples. MOs were separated from peripheral blood samples that were collected simultaneously and pooled separately from 5 controls and also from 5 COPD patients following RNA isolation. Microarray analysis was performed on this new sample set and the results revealed a list of 555 probe sets with 356 annotated genes showing at least 1.5 fold change differences between COPD and control samples in peripheral blood MOs. We compared that list to the AM microarray result. The comparison resulted in a list of 54 probe sets representing 35 annotated genes. GO analysis was also performed and interestingly the identified 356 genes have fallen into the response to stress and stimuli and cell death categories, as it was the case with AMs samples. Remarkably, the majority of the identified categories overlapped with those found in disease effected AMs

We validated commonly expressed genes in AMs and peripheral blood MOs by measuring the expression levels of these 35 genes using RT-QPCR. A new cohort of 22 COPD and 16 control patients was recruited/assembled and their MO RNA samples were examined for this validation, 13 COPD patients and 7 controls were newly recruited individuals. RT-QPCR analyses validated 5 of the 35 genes that showed significant differences between COPD and control patients. ADAM10, GK, IFRD1, PHLDA1 and SEC14L1 showed higher expression levels in diseased patients. CVA showed that the combination of these 5 genes separated fairly well the COPD and control patients (canonical correlation was 0.62, χ^2 =14.75, d.f. =5, significant at p=0.05, Wilks λ =0.617).

Remarkably, these experiments confirmed our hypothesis that peripheral monocytes also carry a COPD related gene expression pattern and moreover there is some overlap between AM and MO COPD signatures.

A key issue in identifying biomarkers in general and disease specific expression signatures in particular is the ability to correlate these with clinical parameters relevant in the disease. In the case of COPD, the key clinical feature is lung function decrease measured by lung function tests. Therefore, as the next step of our studies the possible correlation between FEV1% values and expression data of the RT-QPCR validated genes were examined. We calculated Pearson correlation between lung function parameter FEV1% and all the 23 genes

and found that higher expression of CCR1, GDF15 and SLC2A3 from the AM validation set obtained from BAL samples correlated with lower FEV1% values. However, higher expression of FN1 correlated with increased values of FEV1%. ADAM10, GK, IFRD1 and SEC14L1 from the peripheral MO validation set showed correlation with lower FEV1% values, but PHLDA1 did not show correlation.

4.2 SNP analyses in chronic inflammatory diseases

4.2.1 COPD SNP STUDY

573 subjects, including 272 patients with COPD and 301 healthy controls were genotyped. COPD patients had been exposed to more tobacco smoke as evidenced by the difference in pack-years but the difference was not significant, cases had a much larger reduction in lung function, typical for a COPD population.

Tests of HWE of SNPs of SERPINE2, TNF-α and SFTPB genes were carried out for all loci among cases and controls separately; all SNPs were found to be in HWE. Genotype and allele frequencies of these three SNPs did not differ significantly between the COPD and the control group; none of the previously reported positive associations between the analyzed SNPs and COPD were replicated in our study.

SNPs of EPHX1 were in HWE in cases and controls. The assessment of the association of individual SNPs with COPD showed that homozygosity for the minor allele increased the risk of disease in case of Tyr113His polymorphism (“slow” allele) and reduced it in case of the His139Arg SNP (“fast” allele). However, none of these SNPs were significantly associated with COPD even after adjusting the model for gender, age and pack-years in logistic regression.

Alleles of the two loci are in complete LD. We have assessed the association of the predicted (“rapid”, “normal”, “slow” and “very slow”) EPHX1 phenotypes with the development of COPD. The distribution of the predicted EPHX1 activity was significantly different between control subjects and COPD patients (P= 0.041). The COPD group had higher proportion of the predicted “slow” phenotype. Consequently the slow phenotype significantly raises the risk of developing COPD (P= 0.021) in our case control study.

We have genotyped three SNPs (rs10801282 (Pro12Ala), rs3856806 (His447His) and rs1800571 (Pro113Gln)) in the PPARγ gene, but the rs1800571 locus was left out from the

analysis because it was homozygous for the major C allele in all individuals. The other two SNPs were in HWE and showed modest LD. The extent of LD between Pro12Ala and His447His was found to be D' = 0.673, although the pair showed lower LD with respect to their correlation coefficient (r^2=0.42).

The single loci allelic and genotypic analysis found no significant association between the two coding variants of PPARγ and COPD. In logistic regression applying a model adjusted for both SNPs, age and pack-years, the rare variant of His447His polymorphism was significantly associated with increased odds for COPD.

We have estimated the frequency of the possible two-SNP haplotypes of PPARγ and assessed the association between haplotypes and COPD development. There was a significant difference in the frequency of the GC haplotype involving the rare G variant of the Pro12Ala locus between the two groups. This finding suggests a protective effect of the GC (12Ala/447His) haplotype of the PPARγ gene for COPD outcome.

4.2.2 IBD SNP STUDY

CD and UC patients were recruited from four centers of Hungarian IBD Study Group. 572 CD, 103 UC and 486 healthy control patients were examined in the study. Four SNPs of PPARγ were tested; three of them localized in the first intron and the fourth the Pro12Ala in the exon B.

Tests of HWE were carried out among CD, UC and control groups separately for all loci and all polymorphisms were found to be in equilibrium in all three groups. The allele and genotype frequencies did not differ significantly among the cohorts. Although the frequency of the rare GG genotype of the Pro12Ala polymorphism occurred with lower frequencies in CD patients compared to controls, this distribution was still not significant. When the distribution of individuals carrying the rare Ala/Ala genotype were compared with Pro/Pro homozygotes between CD cases and controls, the association was found to be significant, suggesting that the rare Ala allele confers a reduced risk of CD, when homozygous. In contrast, no significant statistical association was detected between the PPARγ Pro12Ala and UC susceptibility. We would like to note here that despite the difference in the size of the CD and UC cohort, the allele frequencies of the genotyped SNPs did not differ significantly between these groups.

To determine the non-random association of alleles on the four loci in PPARγ, pair-wise LD was measured among the four polymorphisms. LD analysis revealed that the four SNPs were in strong LD. Since genetic association analysis based on haplotypes considered more powerful than single SNP analysis, we have assessed the association between haplotypes and CD or UC.

In the case of the CD cohort, the GAGG haplotype, containing the rare Ala variant of the Pro12Ala polymorphism, showed borderline significant association to the disease ($P_{corrected}$ = 0.11), suggesting that it confers a protective effect.

In the case of the UC group three haplotypes showed significant association with the development of the disease. Interestingly, the most common four-SNP haplotype, CGGC was found to be strongly associated with UC, it has a protective effect due to its significantly higher frequency in controls compared to cases ($P_{corrected}$ = 0.025). The GAGG haplotype has a protective effect, like in the case of CD ($P_{corrected} = 1.94 \times 10^{-4}$). However, having the common C allele of the Pro12Ala variant instead of the rare G allele– resulting the GAGC haplotype – significantly increases the risk of UC ($P_{corrected} = 1.94 \times 10^{-9}$).

5 DISCUSSION

Inflammation is a physiological response to infections and tissue injury, and is part of the immune defense mechanisms, which helps to restore homeostasis at damaged sites. During acute inflammation negative feedback mechanisms are activated such as production of anti-inflammatory cytokines or activation of regulatory cells. Thus, to maintain the healthy homeostasis, the regulation of inflammatory responses is essential. Dysregulated inflammatory responses lead to chronic inflammation, which are believed to underlie the pathophysiolgy of several disorders. The increased levels of inflammatory mediators amplify the inflammatory response, which are destructive and contribute to clinical symptoms and progression of the disease. Determining disease specific markers and differentiating between disease states are critically important in chronic diseases.

Environmental factors and genetic predisposition act together in the development of complex diseases. The development of genomic technologies, including the microarray and the next generation sequencing technology provides an opportunity to perform global gene expression and genome wide association analyses.

In our study we performed global gene expression analysis of AMs, which are one of the most important cells in the inflammatory process of COPD. We expanded the list of COPD associated genes, identified and validated a list of COPD specific genes which overlap between AMs and their precursors peripheral MOs, and correlated gene expression patterns with lung function parameter. Genetic association studies were also performed and SNPs of an inflammatory transcription regulator gene, PPARγ were examined in association with COPD and IBD.

COPD is a prime example of a long-lasting, chronic disease with devastating outcome with essentially no validated clinical parameter or biomarker other than lung function tests available. Previously several studies examined the pathomechanism of COPD at the mRNA level. They analyzed animal models of the disease, performed cell culture experiment and collected primary clinical samples to obtain gene expression data.

In our study largely independent sets of patients for screen/exploration (microarray) and validation (RT-QPCR) were collected, stringent inclusion/exclusion criteria were

determined for the sample collection. Our results suggest that gene sets specific for COPD can be identified in both AMs and MOs based on their expression patterns. Several previously identified COPD related genes such as CYP1B1, PLA2G7, CCL2, SPP1, IL-8 and IL-1β were validated our microarray experiment; in addition we could identify a list of COPD specific probe sets that were expressed differently in diseased and healthy patients in AMs. From this list, 23 genes were confirmed by RT-QPCR on a larger cohort of independent samples and thirteen novel candidate genes were identified, including CCR1, TREM2, GADD45A, PHLDA1 and PHLDA2, EMR3, GDF15, FPRL2, HOMER3, SLC2A3, IGSF6, PLEKHB1 PAPSS2. GO analysis of these data showed that the differentially expressed genes related to such pathways, which may contribute to the pathogenesis of COPD.

We identified commonly expressed genes in AM and peripheral MOs. What we consider as a key aspect of our work is the identification and validation of a set of 5 genes (ADAM10, PHLDA1, IFDR1, GK, SEC14L1) showing common expression patterns in AM and MO. It suggests that these two related cell types are both affected in the disease and this is manifested in peripheral gene expression signatures.

Importantly, our results also showed, that the validated COPD-specific gene expression signatures of AMs (CCR1, FN1, GDF15, SLC2A3) and MOs (ADAM10, GK, IFRD1, SEC14L1) correlate with lung function parameter (predicted FEV1%) which may suggest that disease progression can be determined and monitored by analyzing either disease-specific or peripheral blood gene expression patterns using a panel of genes.

The genetic predisposition plays a major role in the development of chronic inflammatory and autoimmune diseases. Identification of disease-susceptible SNPs of candidate genes could be used as primary target of therapy, however further validation of all genes are needed. Validation is divided into two types. Validation of association by additional association tests, such as replicate the study on new independent cohort either from the same race or different race. The frequencies of common SNPs could show huge differences between races, which could lead to controversial interpretation of disease-susceptibility of SNPs. Validation of the association could be performed by using biological assay. In this case the function of the identified susceptible gene is examined whether this function is relevant to the pathophysiology of the disease and how the disease-associated variant could affect this function.

PPARγ has role in negative regulation of inflammatory process, previous studies and clinical trials reported beneficial effect of the PPARγ ligand on chronic inflammatory diseases. Polymorphisms of PPARγ have been examined in metabolic syndrome, diabetes and obesity, and also in disorders with autoimmune origin such as psoriasis and psoriatic arthritis. However, only a few studies have examined the association of SNPs of PPARγ with IBD. Based on these findings we have investigated the association certain PPARγ gene polymorphisms and common inflammatory diseases, COPD and IBD.

Several polymorphisms were identified as COPD-susceptibility loci, from which the gene product might functionally contribute to the pathophysiology of COPD. Although the incidence of COPD shows an increasing rate in Hungary, our study was the first genetic association study, which examined associations between polymorphisms of candidate genes and COPD in a Hungarian population. We chose several previously tested SNPs, which showed association to COPD in Caucasian cohorts and performed replicate experiments to validate our cohort. In the case of SNPs of SERPINE2, TNF-α and SFTPB we did not find any associations with COPD. But the distribution of predicted phenotypes of EPHX1 showed significant difference between the COPD and control group and the "slow" metabolizing phenotype significantly increased the risk for COPD in our study.

Three SNPs of PPARγ were genotyped (Pro12Ala, His447His and rs1800571), the rs1800571 was excluded since all individuals carried an identical homozygous genotype. Our single-marker tests for the other two coding variants yielded a significant association for the minor allele of His447His polymorphism in logistic regression adjusted for both SNPs, age, sex and pack-years. However the His447His variant did not cause amino acid change, several papers pointed out that synonymous SNPs could affect mRNA splicing, stability and protein folding. These changes can cause a significant effect on the protein function therefore there is a reason to investigate it further.

A modest pair-wise LD was found between Pro12Ala and His447His. Since the use of SNP-based haplotypes in genetic association studies may offer a more powerful approach than the use of individual SNPs, a haplotype analysis was also performed. A significant difference was found in the frequency of GC haplotype (containing the minor G allele of Pro12Ala and major C for His447His variant) between the control and COPD group and the association of this haplotype to COPD outcome were also determined. The GC haplotype

might contribute to a significantly lower risk for COPD, pointing to a potentially functional protective effect of this haplotype.

The incidence of IBD is rapidly increasing in Eastern Europe including Hungary and has reached the level found in Western European countries. We felt it important to study genetic risk factors that might influence susceptibility to IBD in the Eastern European population. Although only limited data are available from this part of Europe, Hungary has a large and well-characterized cohort of IBD patients. In previous studies, several IBD associated SNPs such as polymorphisms in NOD2/CARD15, TLR4, ATG16L1 and IL-23R genes were genotyped and their contributions to susceptibility to IBD of this cohort was examined.

We have further analyzed this Hungarian IBD cohort extending the characterization of genetic factors by genotyping four SNPs (C-681G, rs2067819, rs3892175 and Pro12Ala) of a potential candidate gene, PPARγ. We did not find any significant differences between CD/UC patients and controls in the case of the allele frequency of C-681G, rs2067819 and rs3892175 polymorphisms. Interestingly, the homozygous variant of the minor allele (Ala/Ala) of Pro12Ala was found to be significantly less common in CD patients compared to controls. Strong linkage disequilibrium was identified between the four examined loci and the associations of potential haplotypes were investigated. The GAGG haplotype, containing the rare Ala variant showed higher frequency in controls compared to CD and also UC patients, suggesting a potential protective effect of this variant against the development of IBD. In contrast, the GAGC haplotype was significantly more common in UC patients than in healthy controls, increasing the probability of UC phenotype. Surprisingly the combination of the four major alleles was significantly more common in UC patients than in controls.

We are aware of the fact that our study has some limitations such as the relatively low sample size; and that the population stratification was not investigated. Further studies using larger populations are needed in a different cohort to replicate our results.

6 SUMMARY

We used genomic and genetic analyses to characterize chronic inflammatory diseases. We performed gene expression analysis and sought to define gene sets showing differential expression patterns between healthy controls and patients with COPD. Gene expression patterns of AMs and peripheral MOs were examined in order to evaluate whether the disease affected either one or both cell types and if there was an overlap in altered gene expression. We identified and validated COPD-specific overlapping signatures of expression profiles of peripheral blood MOs and AMs demonstrating that peripheral MOs are also affected by COPD. The discriminating power of the set of 23 genes validated in AM samples; and the set of 5 genes validated in MO samples were also demonstrated by using CVA that could separate COPD and healthy patients in both cases in which the χ^2 test of canonical correlation was significant at p=0.05, suggesting that the gene lists do have strong discriminating power even though separation of the two groups was not perfect. We found that some of the validated expression changes correlate with lung function parameter FEV1%. The presented data can provide a basis for more mechanistic studies of disease pathomechanisms and some of the differentially expressed genes might serve as biomarkers of disease progression and effectiveness of therapy.

We performed the first genetic association study of Hungarian patients with COPD. In a replicate study five SNPs of four genes were genotyped, which had been previously reported as COPD-associated polymorphisms in Caucasian population. This experiment validated our cohort, however we could not found any associations with COPD in the case of SERPINE2, TNF-α and SFTPB, but our study provided support for the suggested causative role of EPHX1 polymorphisms and phenotypes imputed from exon 3 and exon 4 genotype data in COPD outcome in a Hungarian population. Furthermore we have carried out the first investigation of PPARγ gene polymorphisms in a COPD case-control study and characterized the association between individual SNPs and haplotypes of PPARγ and the susceptibility to COPD. The GC haplotype variant showed a modest protective effect. Documentation of PPARγ haplotype association with COPD identifies this important gene as a target of further investigation for the pathogenesis of COPD and as a potential target of therapy.

In addition we have investigated four polymorphisms of PPARγ in a well-characterized Hungarian IBD cohort. We found association between IBD and PPARγ (Pro12Ala) at single loci level suggesting a potential protective effect of the minor Ala allele in CD and we also found haplotypes, which have significantly different frequencies between CD/UC patients and controls. However we could not replicate previous results, where the minor alleles of the two non-coding polymorphisms in intron 1 (rs2067819 and rs3892175) have been found to have significantly higher frequencies in CD cases than in controls. Nonetheless these data provide further support for the notion that a nuclear hormone receptor is linked to human intestinal inflammation and warrant further studies on the exact mechanism contributing to disease progression and also as a potential therapeutic target.

7 PUBLICATIONS

7.1 Publications related to the thesis

Poliska S, Csanky E, Szanto A, Szatmari I, Mesko B, Szeles L, Dezso B, Scholtz B, Podani J, Kilty I, Takacs L, Nagy L. (2011) Chronic Obstructive Pulmonary Disease-Specific Gene Expression Signatures of Alveolar Macrophages as well as Peripheral Blood Monocytes Overlap and Correlate with Lung Function. **Respiration.** 81:499-510 (2011) IF: 1.935

Penyige A, **Poliska S**, Csanky E, Scholtz B, Dezso B, Schmelczer I, Kilty I, Takacs L, Nagy L. (2010) Analyses of association between PPAR gamma and EPHX1 polymorphisms and susceptibility to COPD in a Hungarian cohort, a case-control study. **BMC Medical Genetics.** 11:152, (2010) IF: 2.84

Poliska S, Penyige A, Lakatos PL and the Hungarian IBD Study Group, Papp M, Palatka K, Lakatos L, Molnar T, Nagy L. Association of peroxisome proliferator-activated receptor gamma polymorphisms to inflammatory bowel diseases in a Hungarian cohort. **Inflammatory Bowel Diseases.** 2011;000:000–000 (DOI 10.1002/ibd.21798) IF: 4.64

7.2 Other publications

Mesko B, **Poliska S**, Nagy L. (2011) Gene expression profiles in peripheral blood for the diagnosis of autoimmune diseases. Trends Molecular Medicine. 17 (4), 223-233, (2011)

IF: 11.049

Horvath, E., Sikovanyecz, J., Pál, A., Kaiser, L., Bálint, B.L., **Póliska, S.,** Kozinszky, Z., Szabó, J., Cystic Dilation of the Aqueductus Sylvii in Case of Trisomy 17p11.2-pter with the Deletion of the Terminal Portion of the Chromosome 6. Case Reports in Medicine. 2010, 1-5, 2010.

Mesko B, **Poliska S**, Szegedi A, Szekanecz Z, Palatka K, Papp M, Nagy L. (2010) Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets. BMC Medical Genomics. 3 (1) 15, (2010)
IF: 2.661

Simandi Z, Balint BL, **Poliska S**, Ruhl R, Nagy L. (2010) Activation of retinoic acid receptor signaling coordinates lineage commitment of spontaneously differentiating mouse embryonic stem cells in embryoid bodies. FEBS Letters. 584(14): 3123-30. (2010) IF: 3.541

Szanto A, Balint BL, Nagy ZS, Barta E, Dezso B, Pap A, Szeles L, **Poliska S**, Oros M, Evans RM, Barak Y, Schwabe J, Nagy L. (2010) STAT6 transcription factor is a facilitator of the nuclear receptor PPARγ-regulated gene expression in macrophages and dendritic cells. Immunity. 33(5): 699-712. (2010) IF: 20.589

Széles L, **Póliska S**, Nagy G, Szatmari I, Szanto A, Pap A, Lindstedt M, Santegoets SJ, Rühl R, Dezsö B, Nagy L. (2010) Research resource: transcriptome profiling of genes regulated by RXR and its permissive and nonpermissive partners in differentiating monocyte-derived dendritic cells. Molecular Endocrinology. (11): 2218-31 (2010) IF: 5.257

Töröcsik D, Baráth M, Benko S, Széles L, Dezso B, **Póliska S**, Hegyi Z, Homolya L, Szatmári I, Lányi A, Nagy L. (2010) Activation of liver X receptor sensitizes human dendritic cells to inflammatory stimuli. Journal of Immunology. 184(10): 5456-65. (2010)
IF: 5.646

Széles L, Keresztes G, Töröcsik D, Balajthy Z, Krenács L, **Póliska S**, Steinmeyer A, Zuegel U, Pruenster M, Rot A, Nagy L. (2009) 1,25-dihydroxyvitamin D3 is an autonomous

regulator of the transcriptional changes leading to a tolerogenic dendritic cell phenotype. Journal of Immunology. 182(4): 2074-83. (2009) IF: 5.646

7.3 First author posters on international meetings

Poliska S, Penyige A, Palatka K, Papp M, Nagy L.: Analysis of single nucleotide polymorphisms of PPARγ in chronic inflammatory diseases. EMBO Conference on Nuclear Receptors. Dubrovnik, Croatia. 2009.

Poliska S, Scholtz B, Csanky E, Szatmari I, Szanto A, Nagy L.: Global Gene Expression Analysis of a Complex Human Disease. Functional Genomics & Disease, Innsbruck, Austria, 2008.

Póliska S, Szatmári I, Szántó A, Csánky E, Scholtz B, and Nagy L.: Global Gene Expression Analysis of a Complex Human Disease. EMBO Workshop on Chromatin structure, organization and dynamics, Prague, Czech Repaublic, 2006.

7.4 Presentations

Póliska Szilárd, Csánky Eszter, Scholtz Beáta and Nagy László: A COPD globális génexpressziós analízise. Magyar Tűdőgyógyász Társaság 54. Konferenciája, Szeged, 2006.

Register Number: DEENKÉTK /138/2011.
Item Number:
Subject: Ph.D. List of Publications

Candidate: Szilárd Póliska

Neptun ID: IJFYJP

Doctoral School: Doctoral School of Molecular Cell and Immune Biology

List of publications related to the dissertation

1. **Póliska, S.**, Csánky, E., Szántó, A., Szatmári, I., Meskó, B., Széles, L., Dezső, B., Scholtz, B., Podani, J., Kilty, I., Takács, L., Nagy, L.: Chronic Obstructive Pulmonary Disease-Specific Gene Expression Signatures of Alveolar Macrophages as well as Peripheral Blood Monocytes Overlap and Correlate with Lung Function.
Respiration. Epub ahead of print (2011), 12p, 12p.
DOI: http://dx.doi.org/10.1159/000324297
IF:1.935 (2009)

2. Penyige, A., **Póliska, S.**, Csánky, E., Scholtz, B., Dezső, B., Schmelczer, I., Kilty, I., Takács, L., Nagy, L.: Analyses of association between PPAR gamma and EPHX1 polymorphisms and susceptibility to COPD in a Hungarian cohort, a case-control study.
BMC Med. Genet. 11, 152, 2010.
DOI: http://dx.doi.org/10.1186/1471-2350-11-152
IF:2.84 (2009)

List of other publications

3. Meskó, B., **Póliska, S.**, Nagy, L.: Gene expression profiles in peripheral blood for the diagnosis of autoimmune diseases.
Trends Mol. Med. 17 (4), 223-233, 2011.
DOI: http://dx.doi.org/10.1016/j.molmed.2010.12.004
IF:11.049 (2009)

4. Horváth, E., Sikovanyecz, J., Pál, A., Kaiser, L., Bálint, B.L., **Póliska, S.**, Kozinszky, Z., Szabó, J.: Cystic Dilation of the Aqueductus Sylvii in Case of Trisomy 17p11.2-pter with the Deletion of the Terminal Portion of the Chromosome 6.
Case Reports in Medicine. 2010, 1-5, 2010.
DOI: http://dx.doi.org/10.1155/2010/354170

5. Meskó, B., **Póliska, S.**, Szegedi, A., Szekanecz, Z., Palatka, K., Papp, M., Nagy, L.: Peripheral blood gene expression patterns discriminate among chronic inflammatiory diseases and healthy controls and identify novel targets.
BMC Med. Genomics. 3 (1), 15, 2010.
DOI: http://dx.doi.org/10.1186/1755-8794-3-15
IF:2.661 (2009)

6. Simándi, Z., Bálint, B.L., **Póliska, S.**, Rühl, R., Nagy, L.: Activation of retinoic acid receptor signaling coordinates lineage commitment of spontaneously differentiating mouse embryonic stem cells in embryoid bodies.
FEBS Lett. 584 (14), 3123-3130, 2010.
DOI: http://dx.doi.org/10.1016/j.febslet.2010.05.052
IF:3.541 (2009)

7. Szántó, A., Bálint, B.L., Nagy, Z., Barta, E., Dezső, B., Pap, A., Széles, L., **Póliska, S.**, Oros, M., Evans, R.M., Barak, Y., Schwabe, J., Nagy, L.: STAT6 Transcription Factor Is a Facilitator of the Nuclear Receptor PPAR?-Regulated Gene Expression in Macrophages and Dendritic Cells.
Immunity. 3 (5), 699-712, 2010.
DOI: http://dx.doi.org/10.1016/j.immuni.2010.11.009
IF:20.589 (2009)

8. Széles, L., **Póliska, S.**, Nagy, G., Szatmári, I., Szántó, A., Pap, A., Lindstedt, M., Santegoets, S.J.A.M., Rühl, R., Dezső, B., Nagy, L.: Research resource: Transcriptome profiling of genes regulated by RXR and its permissive and nonpermissive partners in differentiating monocyte-derived dendritic cells.
Mol. Endocrinol. 24 (11), 2218-2231, 2010.
DOI: http://dx.doi.org/10.1210/me.2010-0215
IF:5.257 (2009)

9. Törőcsik, D., Baráth, M., Benkő, S., Széles, L., Balázs, D., **Póliska, S.**, Hegyi, Z., Homolya, L., Szatmári, I., Lányi, Á., Nagy, L.: Activation of liver X receptor sensitizes human dendritic cells to inflammatory stimuli.
J. Immunol. 184 (10), 5456-5464, 2010.
DOI: http://dx.doi.org/10.4049/jimmunol.0902399

IF:5.646 (2009)

10. Széles, L., Keresztes, G., Törőcsik, D., Balajthy, Z., Krenács, L., **Póliska, S.**, Steinmeyer, A., Zuegel, U., Pruenster, M., Rot, A., Nagy, L.: 1,25 dihidroxyvitamin D3 Is an autonomous regulator of the transcriptional changes leading to a tolerogenic dendritic cell phenotype.
J. Immunol. 182 (4), 2047-2083, 2009.
DOI: http://dx.doi.org/10.4049/jimmunol.0803345
IF:5.646