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

Peripheral Blood Gene Expression Profiling as a Tool in Exploring the Pharmacogenomics of Autoimmune Diseases

by Dr. Bertalan Meskó

Supervisor: Prof. Dr. László Nagy

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

Debrecen, 2012

Peripheral Blood Gene Expression Profiling as a Tool in Exploring the Pharmacogenomics of Autoimmune Diseases

By Bertalan Meskó, MD

Supervisor: László Nagy, MD, DSc

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

Head of the **Examination Committee**:	Róza Ádány, MD, DSc
Members of the Examination Committee:	Edit Bodolay, MD, DSc
	Márta Széll, DSc

The Examination takes place at the Research Centre for Molecular Medicine, Medical and Health Science Center, University of Debrecen, 2012

Head of the **Defense Committee**:	Róza Ádány, MD, DSc
Reviewers:	Joel Dudley, PhD
	Margit Zeher, MD, DSc
Members of the Defense Committee:	Edit Bodolay, MD, DSc
	Márta Széll, DSc

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, 19th of November, 2012

1 Introduction

The future of medicine will clearly be centered around personalized care and clinical genomics meaning that every individual will receive prevention tips, diagnosis and therapy based on the combination of their own genomic background and lifestyle. According to the US President's Council on Advisors on Science and Technology, "Personalized Medicine refers to the tailoring of medical treatment to the individual characteristics of each patient…to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Preventative or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not."

The first announcements of the completion of the Human Genome Project in 2001 opened up new opportunities, and also raised important questions concerning data privacy, genetic discriminations and medical implications of genome-related variables. With this huge amount of genomic data available, scientific communities have to tackle new challenges such as data storage capacities, lowering the cost of human genome sequencing; and the most important one, making genomic knowledge useful and accessible in everyday medicine. A key component of this is translating the science of pharmacogenomics into clinical practice.

1.1 Autoimmunity

The overall goal of personalized medicine is to identify new ways for the prediction, diagnosis and treatment of diseases including rare conditions as well as conditions affecting large populations. The latter group also refers to autoimmune and chronic inflammatory disorders which have been in the focus in our studies.

In autoimmune diseases, an inappropriate immune response, that can be restricted to certain organs or involve a particular tissue in different places, appears against substances and tissues that are normally present in the body. It can be defined as the breakdown of mechanisms responsible for self tolerance as well as the induction of an immune response against components of the self.

1.2 The molecular and immunological background of autoimmunity

Under normal circumstances, several T cells and antibodies react with "self" peptides, but one of the functions of specialized cells located in the thymus and bone marrow is to eliminate cells that recognize self-antigens, thus preventing autoimmunity. During normal development of mature B and T cells, although antibodies highly attracted to self-antigens are eliminated, antibodies that recognize self-antigens with low affinity can be kept in the periphery. As lymphocytes are transformed into B-cells in the bone marrow and into T-cells in the thymus, self-reactive cells undergo apoptosis or become unreactive. This process is called the central tolerance.

On the other hand, mature lymphocytes that encounter self antigens in secondary lymphoid organs and undergo anergy, deletion or suppression in order to prevent disease (peripheral tolerance).

Failure or breakdown either in the central or peripheral tolerance can lead to autoimmune reactions and diseases.

1.3 Examples of autoimmune conditions

1.3.1 Rheumatoid arthritis

RA is a chronic, systemic autoimmune condition causing inflammation and tissue damage in joints and tendon sheaths, but it can also produce inflammation in the lungs, pericardium, pleura, and the skin. Its autoimmune origin is explained by a genetic link with major histocompatibility complexes (MHC); environmental factors and the presence of autoantibodies, known as rheumatoid factors (RF). Lymphocytes are activated and chemical messengers (cytokines, such as TNF and IL-1) are expressed in the inflamed areas.

The most recent therapies include biological agents (e.g. TNFa blockers - etanercept, infliximab, adalimumab) that can improve symptoms by reducing inflammatory response. Though, 30-40% of patients are non-responders and cannot tolerate the side effects. In these cases, future gene profiling based on peripheral blood mononuclear cells might prevent non-responders from starting the unefficient therapy.

1.3.2 Psoriasis

Psoriasis is a chronic inflammation of the skin and joints with scaly patches or psoriatic plaques which are areas of inflammation and excessive skin production. Such plaques frequently occur on the skin of the elbows and knees, but can affect any other areas including the scalp and genitals.

1.3.4 Inflammatory Bowel Diseases

IBD is a group of conditions including CD and UC that cause inflammation in the intestines. It affects about 1.4 million people in the United States and 2.2 million people in Europe. The common feature of IBD is the inflammation of the intestines and the autoimmune origin. All layers of the intestines are affected in CD, but it is restricted to the mucosa in UC.

1.4 Gene expression profiling

Gene expression profiling has been widely and more reliably used to generate reproducible expression patterns for obtaining disease-specific information studying the pathomechanism and therapy responsiveness of different diseases either by using peripheral blood mononuclear cells (PBMC), whole blood, tissue biopsy samples or other sources of mRNA. Although, gene expression signatures have not yet been applied in the clinical practice despite the fact that the major technical and methodological issues have been worked out in the last decade; and the conversion of gene expression data into practical diagnostic tools could be a turning point in translational medicine and personalized genomics.

Obtaining PBMCs is less invasive compared to biopsies and contain cells affected by inflammation, such as circulating monocytes, T- and B-lymphocytes. Therefore gene expression profiles of PBMC may reflect pathomechanisms of the disease, but not necessarily, that is why extending pharmacogenomic markers to clinical application through the development of assays based on gene panels is the key challenge.

PBMCs consist of cell types such as CD4+ and CD8+ T-lymphocytes (70% of all cells), B-lymphocytes (15%), natural killer cells (10%), monocytes (5%) and dendritic cells (≈1%).

Basically, there are three methods for obtaining gene expression data. The most sensitive and targeted way is real-time polymerase chain reaction (RT-QPCR) and the global analysis is performed by microarrays. RT-QPCR measurements could be low-throughput when only few genes are detected or could be low- to medium-throughput using 96- and 384-well assay plates or TaqMan Low Density Array (TLDA).

1.5 Gene expression profiling of autoimmune diseases

A huge number of transcriptome studies focusing on chronic inflammatory or autoimmune diseases are available in the literature with examples including colon tissue in IBD; skin lesions in psoriasis and synovial tissue in RA.

It was pointed out before that the molecular signature of disease across tissues is more prominent than the signature of tissue expression across diseases. It underscores the notion that it will be possible to detect disease-specific gene expression signatures in PBMCs.

Similarly to other areas, mouse models have been used to predict gene expression changes in human disease, but first, significant variation was observed between different murine autoimmune models; and second, gene expression profiles of lymphocytes from common murine models of autoimmune disease and the profiles of patients with autoimmune diseases show little overlap.

2. Aims

Identification of gene panels differentiating between autoimmune conditions and healthy cohorts; as well as development of models predicting responses to biologic therapies were in the focus of our work.

The objectives of our studies:

- Finding PBMC gene expression differences between autoimmune conditions (RA, psoriasis and IBD) compared to healthy patients. (chronic inflammatory diseases – CID study)

- Comparing PBMC gene expression profiles of autoimmune conditions. (CID study)
- Designing PBMC gene panels that predict response to tocilizumab therapy in RA. (Tocilizumab study)
- Designing PBMC gene panels that predict response to infliximab treatment in RA and CD; and also validate the results in independent cohorts. (Infliximab study)
- Comparing the gene expression patterns RA and CD patients show following infliximab therapy. (Infliximab study)
- Identifying gene networks underlying autoimmunity.

3. Methods

3.1 Patient cohorts

The Institutional Review Board of University of Debrecen Medical and Health Science Center approved the clinical protocol and the study that was in compliance with the Helsinki Declaration. Signed informed consent was obtained from all individuals providing blood sample.

All blood samples were obtained from Caucasian patients after the subjects fasted overnight for 12 hours locally between 8:00 AM and 9:00 AM in the CID study; before the first admission of infliximab at week 0 (baseline) and the second one at week 4 in the tocilizumab study; and before the first admission of infliximab at week 0 (baseline) and the second one at week 2 in the infliximab study. Samples were processed within one hour after sample collection.

Thirteen patients (nine females, four males) who met the American College of Rheumatology criteria for RA were included in the tocilizumab study.

40 CD patients (16 females, 24 males) diagnosed by clinicians; and 34 RA patients (28 females, 6 males) who met the 2010 EULAR/ACR classification criteria for RA were included in the infliximab study; all of whom had active disease at the time blood was drawn. Regarding the study design, 20 CD and 19 RA patients were included in the first test cohort for microarray experiments sampling at baseline and week 2. For the validation cohort, samples from 15 RA patients and 20 CD patients at baseline were

included in the RT-QPCR experiments. Medication remained unchanged during the study and co-medication was given after blood was taken.

Responder status is determined by a CDAI decrease of 100 points compared to baseline in CD; and by ACR categories at week 14 in RA (ACR0% and ACR 20% improvement represent the non-responder; ACR50% and ACR70% represent the responder status).

3.2 PBMC separation and RNA isolation

10 ml of venous peripheral blood samples were collected by clinicians in Venous Blood Vacuum Collection Tubes containing EDTA (BD Vacutainer K2EDTA) and another 10 ml in native tubes for the extraction of serum. PBMCs were separated by Ficoll gradient centrifugation. Total RNA was extracted from PBMCs using Trizol reagent (Invitrogen), according to the manufacturer's protocol.

3.3 Microarray experiment

Affymetrix GeneChip Human Gene 1.0 ST array was used to analyze global expression pattern of 28869 well-annotated genes. Ambion WT Expression Kit (Life Technologies) and GeneChip WT Terminal Labeling and Control Kit (Affymetrix) were used for amplifying and labeling 250 ng of RNA samples. Samples were hybridized at 45 Celsius degrees for 16 hours and then standard washing protocol was performed using GeneChip Fluidics Station 450 and the arrays were scanned on GeneChip Scanner 7G (Affymetrix).

3.4 TaqMan mRNA analysis by RT-QPCR

In the CID study and the validation step of the infliximab study, gene expression data was obtained using TLDA containing 96 pre-selected genes. Relative gene expression levels were calculated by comparative Ct method that results in normalizing to GAPDH expression in the CID study, and PPIA expression for each sample in the infliximab study.

Statistical analyses of the normalized gene expression data were performed in Prism (GraphPad). As our data did not follow normal distributions, the gene expression

levels in groups were compared separately using the non-parametric Mann-Whitney U test. Test statistics P-values < 0.05 were considered to be statistically significant.

In the tocilizumab study, RT-QPCR was performed to validate a subset of differently expressed transcripts identified by microarray analysis. Individual gene expression assays (Life Technologies) of 12 genes selected for validation were used. Relative gene expression levels were calculated by comparative Ct method that results in normalizing to GAPDH expression for each sample. Unpaired and paired T-tests were used for statistical analysis (p value < 0.05 was considered significant).

3.5 Univariate data analysis

In the infliximab study, microarray data were analyzed with Genespring GX10. Affymetrix data files were imported using the RMA algorithm and median normalization was performed. Regarding the responder vs. non-responder comparison, 20% of probe sets with the lowest expression levels were filtered out in the first step, then the list of remaining probe sets was filtered by fold change (1.2 fold cut off) and statistical analysis was performed using unpaired T-test with Benjamini-Hochberg correction for multiple-testing.

In the tocilizumab study, microarray data were also analyzed with Genespring GX10. Affymetrix data files were imported using RMA algorithm and median normalization was performed. Regarding the baseline vs. week 4 comparison, 26 samples (13-13 samples from baseline and week 4, respectively) were used and 20% of probe sets with the lowest expression levels were filtered out in the first step (5733 probe sets filtered out*)*. Then the list of 23136 probe sets was filtered by fold change (1.2 fold cut off) and statistical analysis was performed using paired Mann-Whitney U-test with Benjamini-Hochberg multiple-testing correction.

Regarding the responder vs. non-responder comparison, 13 samples (13 samples from baseline) were used; and 20% of probe sets with the lowest expression levels were filtered out in the first step (5679 probe sets filtered out*)*. Then the list of 23190 probe sets was filtered by fold change (1.2 fold cut off) and statistical analysis was performed using unpaired T-test with Benjamini-Hochberg correction for multiple-testing.

3.6 Multivariate data analysis

3.6.2 Canonical Variates Analysis (CVA)

Given many genes, separation between predefined groups of patients is best revealed by the multivariate technique of CVA. The method was used to determine whether the groups of responders and non-responders are separable in the multidimensional space spanned by the genetic variables, and if so, which gene subsets have the best discriminatory power.

3.6.3 Multivariate data analysis: Canonical variate analysis (CVA) or Linear discriminant analysis (LDA)

Separation between predefined groups of objects is best revealed by CVA. CVA is the generalization of linear discriminant analysis (LDA), the two terms are used equivalently in the study.

Linear discriminant analysis (LDA) was used for automatically generating gene panels that show 100% segregation between responders and non-responders in both conditions and in both cohorts (test and validation) according to the following algorithm (40 genes in CD and 41 genes in RA were used that were pre-filtered during the experiments with the test cohorts and validated in the validation cohorts).

3.7 ELISA

We measured the serum levels of IL-6, IL-8, IL-12, and IFNg by using enzyme-linked immunosorbent assays (ELISA) in the infliximab study. The concentrations of IL-6, IL-1b and IL-8 in the tocilizumab study were determined in the serum by an ELISA kit (Amersham, G.B.) and the results were given in pg/ml by the Regional Immunology Laboratory, Third Department of Internal Medicine, Medical and Health Science Centre, University of Debrecen.

3.8 Measuring the decrease in the degree of galactosylation of IgG N-glycans

IgG was isolated by the Horváth Laboratory of Bioseparation Sciences, Medical and Health Sciences Center, University of Debrecen from 9 out of the 13 RA patient

samples in the tocilizumab study using Protein A affinity pulldown. The aim of this part of the study was to investigate the changes in the relative amount of agalactosylated (G0) glycans before and after the treatment. We analyzed data by using paired and unpaired t tests in GraphPad Prism ($p<0.05$ was considered statistically significant).

4. Results

4.1 Comparison of gene expression profiles of RA, psoriasis and IBD patients

We performed the PBMC gene expression profiling of RA, psoriasis and IBD patients; and compared them to healthy controls with a goal of identifying expression signatures determining chronic inflammation, disease progression and subtypes and showing which genes separate autoimmune conditions from the control subjects was also a priority.

4.1.2 Overlapping or differential gene expression patterns among chronic inflammatory conditions

Differential gene expression patterns between a disease and control samples or among diseases are shown in a Venn diagram highlighting those genes that differentiate between a particular disease and the control group.

4.1.3 PBMC gene expression profiling identified universal markers of chronic inflammation

We also identified five genes, ADM, AQP9, CXCL2, IL10 and NAMPT showing significant differences between all the three conditions and the control group. These genes might be considered universal markers of chronic inflammation in PBMCs.

4.2 Peripheral blood gene expression profiling of tocilizumab response in RA

4.2.1 Clinical characteristics of patients

Clinical responder status was assessed at week 14 using a binary outcome variable: patients with ACR0 or ACR20 scores were classified as ‘‘non-responders’’; and patients with ACR50 or ACR70 scores were classified as ‘‘responders’’. Within 4 and 14

weeks of tocilizumab therapy, disease activity of all patients decreased significantly if all patients were considered as a single group.

4.2.2 Global gene expression analyses and validation by RT-QPCR

Microarray analysis of 26 samples at baseline and week 4 resulted in a list of 59 genes showing significant differences between baseline and week 4 after correction for multiple-testing accounting for the effects of the therapy itself.

Regarding the gene expression differences determining clinical response, global expression profiling of samples at baseline (n=13) identified 787 probe sets with statistically significant differences between responders (n=9) and non-responders (n=4). We sought to exclude gender-derived differences caused by the disequilibrium in gender-specific gene expression (female/male ratio was 1/3 in the group of non-responders); therefore probe sets differentiating between males and females were removed from the list of probe sets separating responders from non-responders leading to a list of 686 probe sets devoid of gender differences. Four genes including CCDC32, DHFR, EPHA4 and TRAV8-3 remained statistically significant after correction for multiple-testing, thus future analyses should confirm the prediction value of these genes.

As our study used an exploratory approach, technical validation was performed by RT-QPCR in order to determine the expression changes of 12 genes (4 genes from the NR vs. R; and 8 genes from the baseline vs. week 4 comparisons) for each sample (n=26). In this analysis, the normalized mRNA levels showed significant differences validating the microarray data in 10 out of 12 genes selected (CCDC32, EPHA4 and TRAV8-3 between NR and R; ALAS2, CLU, GMPR, ITGB3, ITGA2B, SH3BGRL2 and TREML1 between baseline and week 4).

4.2.3 CVA identified gene panels that determine tocilizumab response

While individual genes cannot separate the two groups of patients, the same genes used simultaneously can provide perfect segregation in the multidimensional space. As we aimed at identifying gene panels that can be potentially used as discriminators

between the two groups of patients, CVA was chosen to detect a set of genes with the highest discriminatory power.

Nine gene lists were selected for CVA including a list containing IL-6 pathway-related genes; four lists obtained from the set of genes showing significant differences between responders and non-responders.

4.3 Predicting infliximab response in RA and CD by global gene expression profiling and validation by RT-QPCR

4.3.1 Clinical characteristics

In CD, 14 responders and 6 non-responders in the test cohort; and 13 responders and 7 non-responders in the validation cohort were identified by clinicians; and there were no significant differences regarding age, CDAI, CRP, hemoglobin, leukocytes or neutrophils between the non-responder and responder groups.

4.3.2 Global gene expression profiling identifies differentially expressed genes between responders and non-responders both in RA and CD

Global gene expression profiling of CD samples at baseline resulted in a list of 48 probe sets significantly differentiating responders from non-responders. Analysis of samples obtained at week 2 showed probe sets with statistically significant differences between responders and non-responders, out of which genes such as ABCC4, BMP6 and THEM5 were significantly changing at baseline as well; while others were new findings at week 2 such as CA2, CADM2, GPR34, IL1RL1, MMD, PRDM1, RAD23A and SLC7A5.

Analysis of baseline RA samples yielded a list of 30 probe sets showing statistically significant differences between responders and non-responders. Out of this list, some of these probe sets such as ELOVL7, FCGR3A, GPAM, MICA, PF4 and RGS1 were significantly changing ones at week 2 as well, while others were new findings at week 2 such as EPSTI1, IFI44, IFIT1, IFIT2, IFIT3, RFC1 and RSAD2.

4.3.3 Differentially expressed genes between responders and non-responders were validated by RT-QPCR in an independent cohort

Genes that showed statistically significant differences between responders and non-responders at baseline in the microarray experiment of the test cohort were added to the gene list of the TLDA cards, excluding probe sets without annotation or probe sets representing small nucleolar RNAs and microRNAs. Validation in the independent cohort yielded a list of 4 genes showing statistically significant differences between non-responders and responders including TMEM176A, TMEM176B, UBE2H and WARS; and CYP4F3, DHRS9, PF4 and MGAM in CD and RA, respectively.

4.3.4 Biostatistical analysis of gene expression data

Genes in the microarray experiment were pre-filtered resulting in a list of 41 genes that was used for the validation analysis as well as the automatic gene panel generation in RA; and a list of 40 genes in CD. A biostatistical algorithm for automatically identifying gene panels discriminating between responders and non-responders was designed.

The algorithm was run once using the deterministic min_F model. It was run with both stochastic models 10,000 times. 9536 and 9657 combinations of gene panels showing 100% segregation between responders and non-responders both in the test and validation cohorts (using microarray and RT-QPCR data, respectively) were produced in CD and RA, respectively. Using the min_F model led to a perfect segregation but stochastic models produced more profound segregation regarding accuracy indicators in both conditions. The high number of gene panels providing 100% segregation after 10,000 runs underscores the notion that there are other gene panels with perfect segregation. Our estimation is that there are over 50 000 such panels in each condition.

We chose 3-3 gene panels with the best discriminatory power taking p values (showing the possibility of producing gene panels with such accuracy by chance) and margins between the segregated groups into consideration for visualization. A list of

genes with the highest p values obtained from the microarray experiment served as negative controls showing slightly over 50% segregation.

5. Discussion

Gene expression profiling has been extensively used for the analysis of autoimmune conditions including disease stratification, prognosis or prediction to biologic therapies. Identifying biomarkers and hints of mechanisms of autoimmunity in conditions that affect large populations in order to personalize treatments and accurately monitor disease progression is of critical importance.

One could expect to detect disease-specific gene expression signatures in PBMC samples which are easy-to-access as the molecular signature of disease across tissues is more prominent than the signature of tissue expression across diseases. We reviewed several autoimmune conditions that have been investigated in different types of samples and with different methods and it seems multiple sclerosis, SLE and RA are in the focus of the highest number of gene expression investigations with Affymetrix global expression profiling being the most frequently used method. Regarding studies focusing on the response to biologic therapies from the gene expression point of view which were also assigned to research groups and geographical locations of their cohorts showing a lack of collaboration in this area.

5.1 Gene expression profiles of chronic inflammatory disorders

When we compared the PBMC gene expression profiles of psoriasis, RA and IBD and compared them to healthy controls, gene panels identifying the pathogenesis of the particular disease were found of which several genes have been linked to these conditions although either at a different genetic level or by using different model. Examples include PTGS2, CCR1 in psoriasis; ADAM10, PTPN22, IL8 in RA; and GZMK, MMP9 and IFNG in IBD. It means we managed to validate some of the genes which have been already pointed out in the literature highlighting the power of our gene panels.

We consider the identification of five candidate markers of chronic inflammation including ADM, AQP9, CXCL2, IL10 and NAMPT that differentiate between samples from patients with chronic inflammation and healthy controls a key aspect of our work. ADM plays role in response to wounding and was found to be distributed on the surface

of the human colonic mucosa and in the synovial tissue of RA patients. IL10 is a known marker of psoriasis, RA and IBD patients in PBMCs.

5.2 Predicting response to tocilizumab in RA by gene expression profiling of PBMC

Turning our attention to predicting response to the expensive biologic therapies to which approximately 30% of patients do not respond well, we designed and used an experimental approach in which global gene expression profiling of PBMCs of RA patients led to the identification of a gene panel predicting response. The power of gene panels was demonstrated with a mathematical method, CVA that shows the differences between two groups of patients. Fifty-nine genes showed significant differences between baseline and week 4 correlating with treatment; out of which 4 genes such as DHFR, TRAV8-3, EPHA4 and CCDC32 determined responders after correction for multiple-testing and 10 of the 12 genes with the most significant changes were technically validated using RT-QPCR. SNP of DHFR (dihydrofolate reductase) was identified as putative predictor for MTX response, efficacy and side effects suggesting that response to tocilizumab is related to response to MTX. TRAV8-3 which is a T cell receptor alpha variable 8-3 was related to CD8+ T-cell response against an HIV-1 epitope; and EPHA4 (ephrin receptor A4) that plays a role in the nervous system and CCDC32 are not associated with RA.

These data suggest that CVA is a powerful mathematical tool for identifying gene panels with the highest discriminatory power; and while a limitation of our approach was the relatively small patient group, such results should be further validated in independent and larger sample sets.

5.3 Predicting infliximab response in RA and CD with validation in an independent cohort

Similarly to the case of tocilizumab efficacy, clinical parameters alone cannot clearly separate future responders from non-responders, although such a gene panel would have significant health and economical benefits. We used the approach that had been tested in the tocilizumab study but added a validation step as well. Global gene expression profiling of RA and CD PBMC samples in a test cohort and the following

independent validation step with a more sensitive method yielded genes relevant to CD and RA based on the literature.

It has been previously known in RA that the regulation of gene activity in the IFN response in infliximab therapy is strongly associated with treatment response based on data derived from whole blood gene expression profiling; therefore it was not surprising to identify genes of IFN pathways such as IFI44, IFI44L, IFIT1, IFIT2 and IRF2. PTGS2 that discriminated RA patients from healthy subjects in our study; and reduced folate carrier (RFC1) of which genetic polymorphism can modify MTX transport thus influencing response to treatment were found to be significant.

Validation by RT-QPCR in an independent cohort was a key step and although we focused on the identification of gene panels predicting response, genes with significant differences between responders and non-responders were also found. In CD, genes include TMEM176B and TMEM176A which are targets of DC function by forming multimers and restraining DC maturation; UBE2H of which TNFa is a regulator; and WARS with no known role in CD pathogenesis. Validation in RA uncovered single genes differentiating responders from non-responders such as CYP4F3 that is associated with the pathomechanism of ostheoarthritis; as well as DHRS9, MGAM; and PF4 that was found to be a predictor of non-response for infliximab in RA in a proteomic study.

LDA was used in order to identify gene panels with the highest discriminatory power as compared to univariate analysis that may disregard potential interactions among genes; it can reveal underlying differences by analyzing genes simultaneously as a gene panel providing perfect segregation in the multidimensional space.

Outcome-related gene panels identified by similar gene expression studies showed only a few genes in common which might be attributed to the different methods of sample preparation, mRNA extraction or analysis of the data and, as well as individual variations and heterogeneities even in a clinically homogenous cohort of patients. The importance of the gene in the pathogenetics of the disease or therapy does not necessarily means it has to be included in the list of genes with statistically significant differences between responders and non-responders. Accordingly, the entire list of outcome-related genes should be taken into consideration in order to detect the potential targets for prediction of treatment.

We built a biostatistical algorithm that made the generation of gene panels with a high discriminatory power between responders and non-responders in both cohorts and diseases automatic which resulted in the identification of over 9500 gene panels in each condition with a 100% segregation regarding responder status.

Due to the relatively low number of subjects in the groups in both diseases, several gene panels with 100% segregation may not represent general correlations and therefore are based on observations of special phenomena existing only among these individuals. Although, the high number of perfectly segregating panels shows it is highly probable that 100% segregation panels will be found when tested with higher subject numbers. In order to facilitate this process, we assigned F values to each gene by summing F values of the lists in which the gene is represented which gave sophisticated scoring that makes possible to select the strongest genes.

6. Conclusions: Future implications and directions of clinical genomics

Gene expression analysis whether it is microarray or RT-QPCR-focused, requires a network-based approach. In clinical genomics, instead of single gene differences, only gene panels could solve unmet needs in the clinical settings by determining early diagnosis, disease progression, subtypes; or whether a patient would respond to a specific therapy before even starting it by analyzing the gene expression patterns of the least invasively obtained peripheral blood samples or tissue biopsies.

Next-generation sequencing (NGS), particularly RNAseq which provides an efficient way to measure transcriptome data including how alleles of a gene are expressed, detect post-transcriptional mutations or identify gene fusions, is going to be in the focus regarding the study of autoimmune diseases as well. Although, it will generate more data and more phenotype-genotype interactions for analysis. Improvements in bioinformatics, data analysis and biobanking will clearly be required paving the way for the developments of personalized medicine which is becoming an integral part of healthcare and the key challenge is to establish a strategic focus on biomarker-based non-invasive or minimally invasive clinical tests. These could give insights into the pathogenesis of autoimmune conditions, identify new targets for future therapies of

predict response to treatments available now therefore reducing healthcare routes and costs.

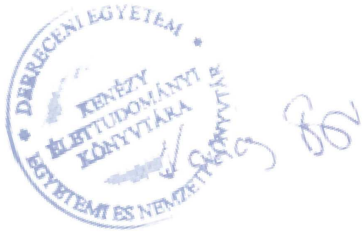
7. Publications related to the thesis

List of publications related to the dissertation

1. **Meskó, B.**, Póliska, S., Szamosi, S., Szekanecz, Z., Podani, J., Váradi, C., Guttman, A., Nagy, L.: Peripheral blood gene expression and IgG glycosylation profiles as markers of tocilizumab treatment in rheumatoid arthritis.
   J. Rheumatol. 39 (5), 916-928, 2012.
   DOI: http://dx.doi.org/10.3899/jrheum.110961
   IF:3.695 (2011)

2. **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:10.355

3. **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:3.766



7.2 Other publications

List of other publications

4. **Meskó, B.**, Zahuczky, G., Nagy, L.: The triad of success in personalised medicine: Pharmacogenomics, biotechnology and regulatory issues from a Central European perspective.
New Biotechnology. 29 (6), 741-750, 2012.
DOI: http://dx.doi.org/10.1016/j.nbt.2012.02.004
IF:2.756 (2011)

5. Heilman, J.M., Kemmann, E., Bonert, M., Chatterjee, A., Ragar, B., Beards, G.M., Iberri, D.J., Harvey, M., Thomas, B., Stomp, W., Martone, M.F., Lodge, D.J., Vondracek, A., de Wolff, J.F., Liber, C., Grover, S.C., Vickers, T.J., **Meskó, B.**, Laurent, M.R.: Wikipedia as a Key Tool for Global Public Health Promotion.
J. Med. Internet Res. 13 (1), 1-12, 2011.
IF:4.409

6. Kollár, J., **Meskó, B.**: Revolution in Education: New Possibilities in Education of Medical Students.
Med. Teach. 33 (8), 685-686, 2011.
DOI: http://dx.doi.org/10.3109/0142159X.2011.600102

7. 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. 81 (6), 499-510, 2011.
DOI: http://dx.doi.org/10.1159/000324297
IF:2.258

8. Csősz, É., **Meskó, B.**, Fésüs, L.: Transdab wiki: The interactive transglutaminase substrate database on web 2.0 surface.
Amino Acids. 36 (4), 615-617, 2009.
DOI: http://dx.doi.org/10.1007/s00726-008-0121-y
IF:3.877

9. **Meskó B.**, Dubecz A.: Az orvostudomány és a világháló nyújtotta új lehetőségek.
Orv. Hetil. 148 (44), 2095-2099, 2007.
DOI: http://dx.doi.org/10.1556/OH.2007.28162

10. **Meskó B.**: A genetika másfél évszázada.
Orv. Hetil. 145 (32), 1971-1672, 2004.

Total IF: 31.116

Total IF (publications related to the dissertation): 17.816

The Candidate's publication data submitted to the Publication Database of the University of Debrecen have been validated by Kenezy Life Sciences Library on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

10 October, 2012

H-4032 Debrecen, Egyetem tér 1. e-mail: publikaciok@lib.unideb.hu

7.3 First author posters on international meetings

1. Gene expression patterns of chronic inflammatory diseases (European Dermatology Congress, Budapest, 2009)
2. Digital Literacy in Medical Education: An Elective Course (Medicine 2.0 Congress, Maastricht, 2010)
3. Digital Literacy in Medical Education: An Elective Course (Medicine 2.0 Congress, Stanford, 2011)

7.4 Presentations

1) Personalized Genomics; Genomic National Technological Platform Assembly, Hungarian Academy of Sciences, Budapest, Hungary, 2009
2) Personalized Genomics and the Online World, MKBE Pharmacobiochemistry Group, Balatonöszöd, 2010
3) Genomics of Chronic Inflammatory Diseases, MKBE Pharmacobiochemistry Group, Balatonöszöd, 2011
4) Direct-to-Consumer Genomic Tests, Hungarian Congress of Genetics, Siófok, Hungary, 2011

8. ACKNOWLEDGEMENTS

I would like to thank my supervisor, Prof. László Nagy, for guiding my work and giving me constructive pieces of advice. I'm thankful to Prof. László Fésüs for giving me a chance to work at the department; and I'm grateful to Dr. Szilárd Póliska for his technical suggestions and guidance.

I would like to thank all the clinicians who participated in our studies.

I would like to thank the technical assistance of Ibolya Fürtös; and I thank all my colleagues in the Nuclear Hormone Receptor research laboratory and in the Department of Biochemistry and Molecular Biology for their help.