

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

Dr. Bertalan Meskó

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

Supervisor: Prof. Dr. László Nagy



UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

Debrecen, 2012

TABLE OF CONTENTS

<u>1. Abbreviations</u>	4
<u>2. Introduction</u>	10
2.1 Autoimmunity	11
2.2 The molecular and immunobiological background of autoimmunity	14
2.3 Economic aspects and the diversity of autoimmune conditions	18
2.5 Examples of autoimmune conditions	20
2.5.1 Rheumatoid arthritis	20
2.5.2 Ankylosing spondylitis	25
2.5.3 Psoriasis	25
2.5.4 Psoriatic arthritis	26
2.5.5 Inflammatory Bowel Diseases	27
2.6 Gene expression profiling	28
2.7 Gene expression profiling of autoimmune diseases	31
<u>3. Aims</u>	34
<u>4. Methods</u>	35
4.1 Patient cohorts	35
4.2 PBMC separation and RNA isolation	38
4.3 Microarray experiment	39
4.4 TaqMan mRNA analysis by RT-QPCR	39
4.5 Gene List Selection Process of CID study	40
4.6 Univariate data analysis	41
4.7 Multivariate data analysis	42
4.7.1 Principal Components Analysis (PCA)	42
4.7.2 Canonical Variates Analysis (CVA)	43
4.7.3 Multivariate data analysis: Canonical variate analysis (CVA) or Linear Discriminant analysis (LDA)	43
4.8 ELISA	45
4.9 Measuring the decrease in the degree of galactosylation of IgG N-glycans	46
<u>5. Results</u>	46
5.1 Comparison of gene expression profiles of RA, psoriasis and IBD patients	46
5.1.1 Identification of gene panels discriminating chronic inflammatory disease patients from healthy controls	46
5.1.2 Overlapping or differential gene expression patterns among chronic inflammatory conditions	50
5.1.3 Differentially expressed genes identify disease subtypes and prognosis	54
5.1.4 PBMC gene expression profiling identified universal markers of chronic inflammation	56
5.2 Peripheral blood gene expression profiling of tocilizumab response in RA	58
5.2.1 Clinical characteristics of patients	58
5.2.2 Global gene expression analyses and validation by RT-QPCR	62
5.2.3 Gene expression patterns correlate with clinical parameters	64
5.2.4 CVA identified gene panels that determine tocilizumab response	65
5.2.5 The degree of galactosylation of IgG N-glycans increases after treatment and in responders versus non-responders	68

5.3 Predicting infliximab response in RA and CD by global gene expression profiling and validation by RT-QPCR.....	70
5.3.1 Clinical characteristics.....	70
5.3.2 Global gene expression profiling identifies differentially expressed genes between responders and non-responders both in RA and CD.....	71
5.3.3 Differentially expressed genes between responders and non-responders were validated by RT-QPCR in an independent cohort	77
5.3.4 Biostatistical analysis of gene expression data	79
5.3.5 Determination of serum cytokine levels.....	85
6.	
<u>Discussion</u>	88
6.1 Gene expression profiles of chronic inflammatory disorders.....	90
6.2 Predicting response to tocilizumab in RA by gene expression profiling of PBMC.....	92
6.3 Predicting infliximab response in RA and CD with validation in an independent cohort.....	92
6.4 Implications for autoimmune conditions and future studies.....	96
7. Conclusions: Future implications and directions of clinical genomics.....	99
8. Acknowledgement	101
9. References	102
10. Publications.	110
10.1 Publications related to the thesis	110
10.2 Other publications.....	111
10.3 First author posters on international meetings	113
10.4 Presentations.....	113
11.	
<u>Appendix</u>	113

1 ABBREVIATIONS

ABCC4: ATP-binding cassette, sub-family C (CFTR/MRP), member 4.

ACPA: anti-citrullinated protein antibody

ACR: American College of Rheumatology

ACTB: Beta-actin

ADAM10: ADAM metallopeptidase domain 10.

ADM: Adrenomedullin

AIDA: axin interactor, dorsalization associated

AIRE: Autoimmune regulator

ALAS2: aminolevulinate, delta-, synthase 2

ANXA3: annexin A3

APTS: aminopyrene-trisulfonate

AQP9: Aquaporin 9

ARHGEF12: Rho guanine nucleotide exchange factor (GEF) 12

AS: ankylosing spondylitis

B4GALT1: Beta-1,4-galactosyltransferase 1

BMP6: Bone morphogenetic protein 6

BTN3A2: butyrophilin, subfamily 3, member A2

C1orf101: chromosome 1 open reading frame 101

CA2: carbonic anhydrase II

CADM2: cell adhesion molecule 2

CCDC32: coiled-coil domain containing 32

CCL20: Chemokine (C-C motif) ligand 20

CCL4: Chemokine (C-C motif) ligand 4

CCL5: Chemokine (C-C motif) ligand 5

CD: Crohn's disease

CDAI: Crohn's disease activity index

CGE-LIF: capillary gelelectrophoresis with laser induced fluorescence detection

CLU: clusterin

COPD: Chronic obstructive pulmonary disease

CRP: C-reactive protein

CUGBP1: CUG triplet repeat, RNA binding protein 1

CUGBP2: CUG triplet repeat, RNA binding protein 2

CVA: Canonical Variates Analysis

CXCL2: Chemokine (C-X-C motif) ligand 2

CXCR4: C-X-C chemokine receptor type 4

CYP2C9: Cytochrome P450 2C9

CYP4F3: cytochrome P450, family 4, subfamily F, polypeptide 3

DAS28: disease activity score evaluated in 28 joints

DC: dendritic cell

DHFR: dihydrofolate reductase

DHRS9: dehydrogenase/reductase (SDR family) member 9

DMARD: disease modifying anti-rheumatic drugs

DUSP1: Dual specificity protein phosphatase 1

DYRK3: Dual specificity tyrosine-phosphorylation-regulated kinase 3

EDTA: ethylenediaminetetraacetic acid

EIF2AK2: eukaryotic translation initiation factor 2-alpha kinase 2

EIF5B: Eukaryotic translation initiation factor 5B

ELOVL7: ELOVL fatty acid elongase 7

ENDOD1: endonuclease domain containing 1

EPB42: erythrocyte membrane protein band 4.2

EPHA4: EPH receptor A4

EPSTI1: epithelial stromal interaction 1 (breast)

ETS2: v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)

EULAR: the European League Against Rheumatism

FAM108A1: family with sequence similarity 108, member A1

FAM47B: family with sequence similarity 47, member B

FCGR3A: Fc fragment of IgG, low affinity IIIa, receptor (CD16a)

FMN1: formin 1

FOS: FBJ murine osteosarcoma viral oncogene homolog

FOXP3: forkhead box P3
FUT8: Alpha-(1,6)-fucosyltransferase
GAPDH: Glyceraldehyde 3-phosphate dehydrogenase
GMPR: guanosine monophosphate reductase
GPAM: Glycerol-3-phosphate acyltransferase
GPR34: G protein-coupled receptor 34
GZMK: granzyme K (granzyme 3; tryptase II)
HAQ: Health Assessment Questionnaire
HBB: hemoglobin, beta
HLA: Human Leukocyte Antigene
HORMAD1: HORMA domain containing 1
HPRT1: hypoxanthine phosphoribosyltransferase 1
IBD: Inflammatory bowel disease
IFI44: interferon-induced protein 44
IFI44L: interferon-induced protein 44-like
IFIT1: interferon-induced protein with tetratricopeptide repeats 1
IFIT2: interferon-induced protein with tetratricopeptide repeats 2
IFIT3: interferon-induced protein with tetratricopeptide repeats 3
IFN: interferon
IFNG: interferon gamma
IGJ: immunoglobulin J
IGKC: Immunoglobulin kappa constant
IL-10: Interleukin-10
IL-13: Interleukin-13
IL-1b: Interleukin-1 beta
IL-1ra: IL-1 receptor antagonist
IL1RL1: interleukin 1 receptor-like 1
IL-4: Interleukin-4
IL-6: Interleukin-6
IL-8: Interleukin-8
IPEX: immune dysregulation, polyendocrinopathy, enteropathy, and X-linked inheritance

ITGA2B: Integrin alpha-IIb
ITGB3: integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)
KAT2B: K(lysine) acetyltransferase 2B
KLRF1: killer cell lectin-like receptor subfamily F, member 1
LDA: Linear discriminancy analysis
MAP1LC3B: microtubule-associated protein 1 light chain 3 beta
MGAM: maltase-glucoamylase (alpha-glucosidase)
MGAT3: Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase
MHC: major histocompatibility complexes
MICA: MHC class I polypeptide-related sequence A
MMD: monocyte to macrophage differentiation-associated
MMP8: matrix metalloproteinase 8 (neutrophil collagenase)
MMP9: matrix metalloproteinase 9 (neutrophil collagenase)
MS: Multiple Sclerosis
MTX: methotrexate
MYH2: myosin, heavy chain 2, skeletal muscle, adult
NAMPT: Nicotinamide phosphoribosyltransferase
NARG2: NMDA receptor regulated 2
NGS:Next-generation sequencing
NK: Natural killer cell
NOD2: Nucleotide-binding oligomerization domain-containing protein 2
NR: non-responder
NSAID: Non-steroidal anti-inflammatory drug
ODC1: ornithine decarboxylase 1
OLR1: oxidized low density lipoprotein (lectin-like) receptor 1
OMIM: Online Mendelian Inheritance in Man
OR2A9P: olfactory receptor, family 2, subfamily A, member 9 pseudogene
PADI4: Peptidyl arginine deiminase, type IV
PBMC: Peripheral blood mononuclear cell
PCA: Principal Components Analysis
PF4: Platelet factor 4

PGRMC1: Progesterone receptor membrane component 1
 PIP4K2A: phosphatidylinositol-5-phosphate 4-kinase, type II, alpha
 PPIA: peptidylprolyl isomerase A (cyclophilin A)
 PRDM1: PR domain containing 1, with ZNF domain
 PsA: Psoriatic arthritis
 PSORS: psoriasis Susceptibility genes
 PTGS2: prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)
 PTPN22: Protein tyrosine phosphatase, non-receptor type 22 (lymphoid)
 PTX3: Pentraxin-related protein
 RA: rheumatoid arthritis
 RAD23A: RAD23 homolog A (*S. cerevisiae*)
 RF: rheumatoid factor
 RFC1: Replication factor C subunit 1
 RGL1: ral guanine nucleotide dissociation stimulator-like 1
 RGS1: regulator of G-protein signaling 1
 RNF2: ring finger protein 2
 RPLP0: ribosomal protein, large, P0
 RSAD2: radical S-adenosyl methionine domain containing 2
 RT-QPCR: real-time quantitative polymerase chain reaction
 RUNDC3A: RUN domain containing 3A
 S100P: Protein S100-P
 SDF-1: Stromal cell-derived factor-1
 SERPINB10: serpin peptidase inhibitor, clade B (ovalbumin), member 10
 SH3BGRL2: SH3 domain binding glutamic acid-rich protein like 2
 SLC25A39: solute carrier family 25, member 39
 SLC7A5: solute carrier family 7 (amino acid transporter light chain, L system) member 5
 SLE: Systemic lupus erythematosus
 THEM5: thioesterase superfamily member 5
 TIMP1: Tissue inhibitor of matrix metalloprotease-1
 TLDA: TaqMan Low Density Array

TMEM176A: transmembrane protein 176A
TMEM176B: transmembrane protein 176B
TMEM56: transmembrane protein 56
TMOD1: tropomodulin 1
TNF: tumor necrosis factor
TNFAIP6: tumor necrosis factor, alpha-induced protein 6
TRAV8-3: T cell receptor alpha variable 8-3
TREML1: triggering receptor expressed on myeloid cells-like 1
UBE2H: ubiquitin-conjugating enzyme E2H
UC: ulcerative colitis
VEGF: Vascular endothelial growth factor
VKORC1: Vitamin K epoxide reductase
WARS: Tryptophanyl-tRNA synthetase, cytoplasmic
YME1L1: YME1-like 1 (*S. cerevisiae*)
ZNF687: zinc finger protein 687

2 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." [1]

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 [2].

The success of the approach of personalized care based on genomic findings has been demonstrated several times. Examples include genotyping for single nucleotide polymorphisms in genes involved in warfarin (Coumarin) metabolism which is used as an anticoagulant in clinical practice but requires monitoring and is associated with adverse outcomes. Variants of genes CYP2C9 encoding Cytochrome P450 enzyme which metabolizes warfarin, and VKORC1 encoding the Vitamin K epoxide reductase, a target of coumarins, led to the development of commercially-available genetic tests enabling more accurate dosing based on algorithms that includes the race, medication, weight, and genotype of an individual [3].

Another well-known example is the use of Herceptin (trastuzumab), a genetically engineered humanized monoclonal antibody, targeting the human epidermal growth factor receptor-2 protein (HER-2/neu) which is over-expressed on the surface of breast

cancer cells in approximately 25–30% of breast cancer patients. HER-2/neu positive patients receiving trastuzumab besides adjuvant chemotherapy regimens have improved outcomes in early breast cancer [4]. As of 2012, there are about 72 commercially available, prominent products in personalized medicine [5].

2.1 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.

The term "Horror autotoxicus" ("the horror of self-toxicity") was coined by the German immunologist Paul Ehrlich (1854-1915) in order to describe the body's innate aversion to immunological self-destruction. The concept of autoimmunity as the cause of human illnesses was not accepted into the mainstream of medical thinking until the 1950s and 1960s.

The genomic background, immune system regulations and environmental factors have to contribute synergistically to form autoimmune diseases (Figure 1).

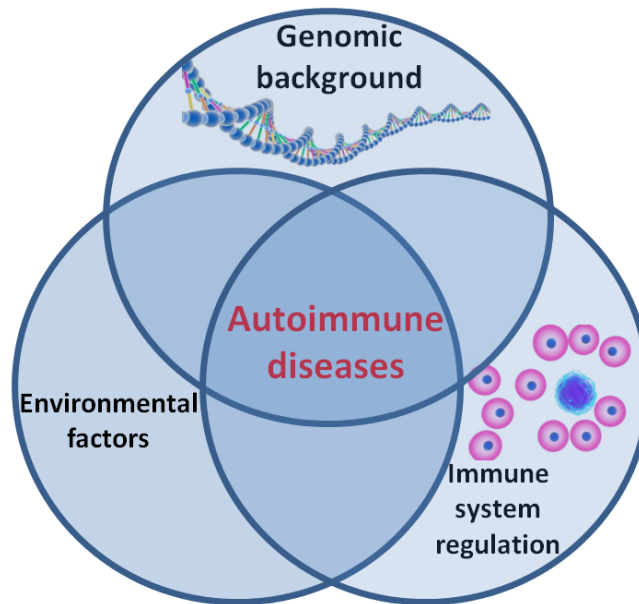


Figure 1: Schematic outline of factors contributing to the development of autoimmune conditions.

Although autoimmune conditions are referred to as multifactorial ones in which several components along with genomic ones can cause a disease risk, in some exceptions single gene differences can also lead to the development of autoimmune disorders. The first report of a single-gene defect causing a systemic human autoimmune disease, thus providing a tool for exploring the molecular basis of autoimmunity, was the isolation of the AIRE gene encoding for a putative nuclear protein featuring two PHD-type zinc-finger motifs of which five mutations were reported in individuals with autoimmune polyglandular syndrome type 1 [6]. In another case, it was shown that immune dysregulation, polyendocrinopathy, enteropathy, and X-linked inheritance (IPEX), a syndrome of systemic autoimmunity, was caused by mutations of FOXP3, a critical regulator of T-cell homeostasis [7].

Regarding the multifactorial approach, in other autoimmune conditions, susceptibility genes can be identified such as the HLA-B27 allele in case of ankylosing spondylitis (AS). However, only 1%–5% of HLA-B27-positive individuals develop AS suggesting that other genes must also be involved [8].

Several environmental factors such as pathogens, drugs, hormones and toxins can trigger autoimmune reactions. Procainamide an antiarrhythmic agent used for the medical treatment of cardiac arrhythmias is known to induce SLE [9]. The toxic oil syndrome, that caused the death of over 600 people in Spain in 1981 as the consumption of colza oil was intended for industrial rather than personal use, is an example of an exogenously induced autoimmune reaction [10]. This unique illness was marked by lung disease, eosinophilia, and excessive IgE levels. Some pathogens can even molecularly mimic autoimmune reactions. Rheumatic fever caused by *Streptococcus pyogenes* through antibody cross-reactivity involving the heart, joints, skin and brain is one of the relevant examples [11].

There are clear differences in several autoimmune conditions regarding gender (Figure 2). Related to this, there are two major clinical observations: 1) females are more susceptible to autoimmune conditions than males; and 2) pregnancy reduces relapses in females. A protective role for high levels of androgens in young males was found as well as little role for fluctuating levels of ovarian hormones in females [12].

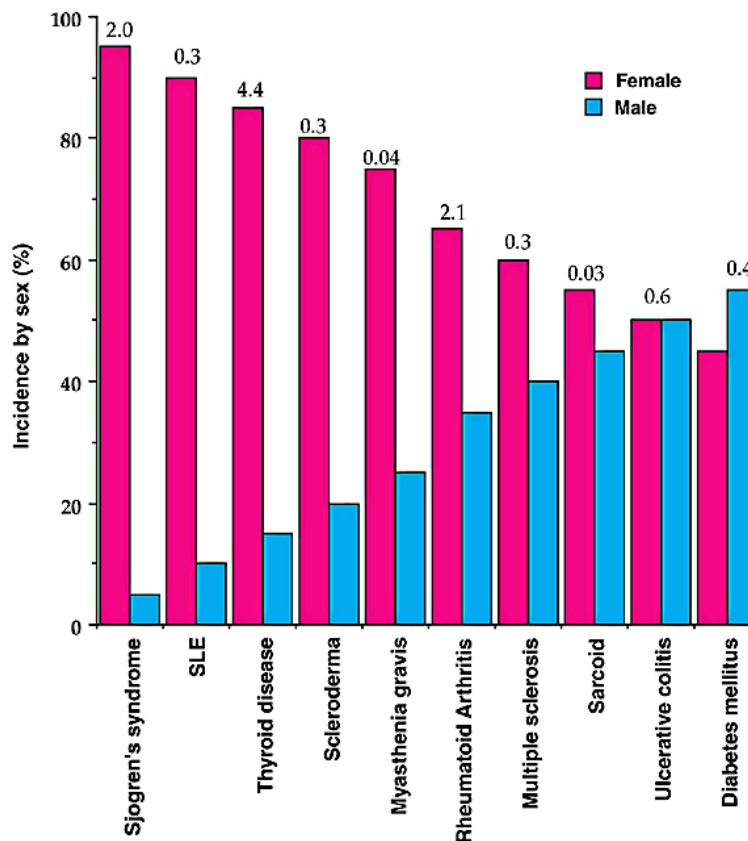


Figure 2: Gender differences in autoimmune conditions [13].

2.2 The molecular and immunological background of autoimmunity

The immune system consists of an innate and an adaptive component (Figure 3-4). The innate response, the dominant system of host defense in most organisms, is triggered when microorganisms or toxins are identified by pattern recognition receptors [14]. It consists of soluble factors such as complement proteins; and cellular components including granulocytes (basophils, eosinophils and neutrophils), mast cells, macrophages, dendritic cells (DC) and natural killer cells (NK cells).

The adaptive immune response is slower to develop, but manifests as increased antigenic specificity and memory consisting of antibodies, B cells, CD4⁺ and CD8⁺ T-lymphocytes.

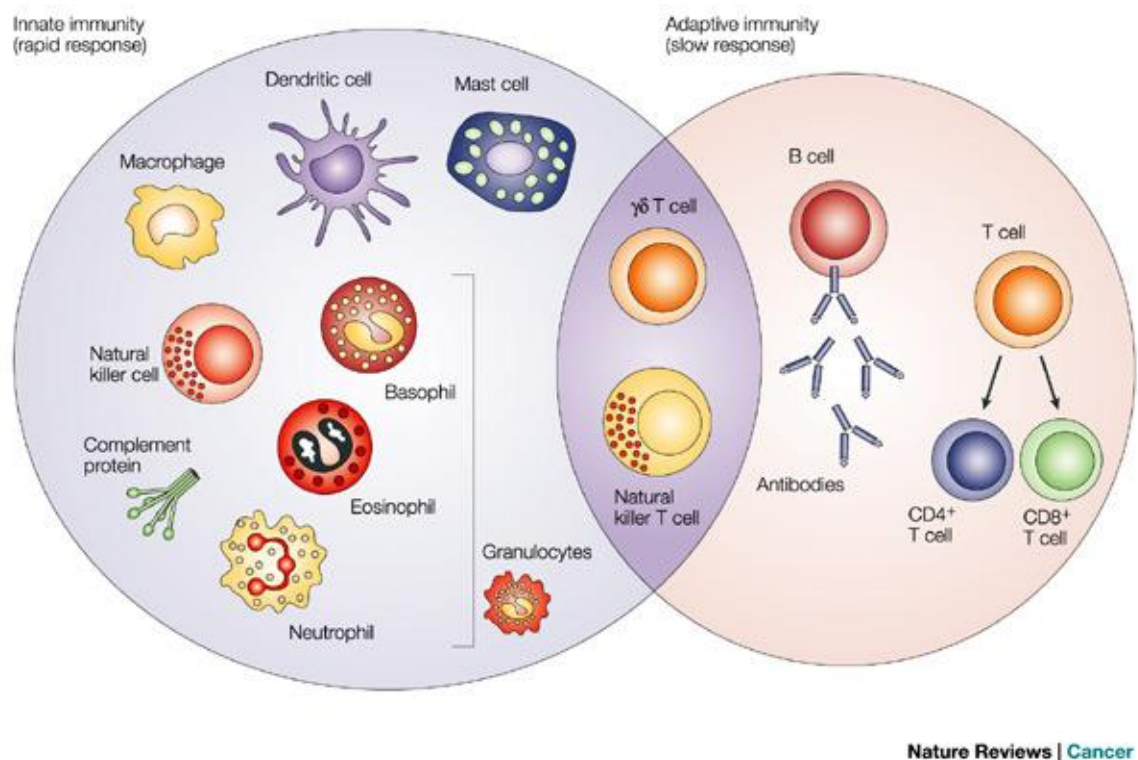


Figure 3: Components of the innate and adaptive immunity [15].

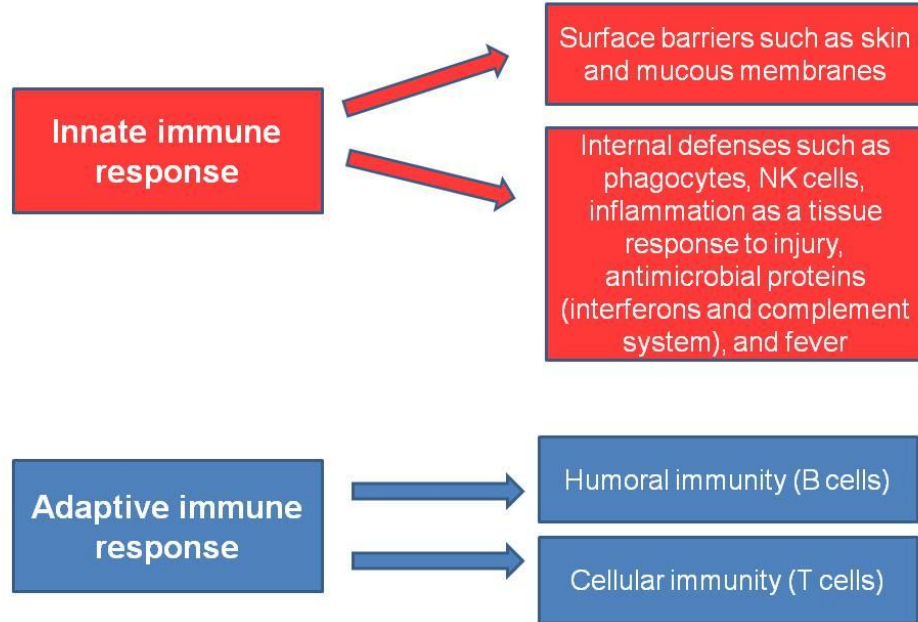


Figure 4: Details of the innate and adaptive immune defenses.

Under normal circumstances, several T cells and antibodies react with "self" peptides [16], 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 [17]. 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) (Figure 5).

Failure or breakdown either in the central or peripheral tolerance can lead to autoimmune reactions and diseases [18] [19].

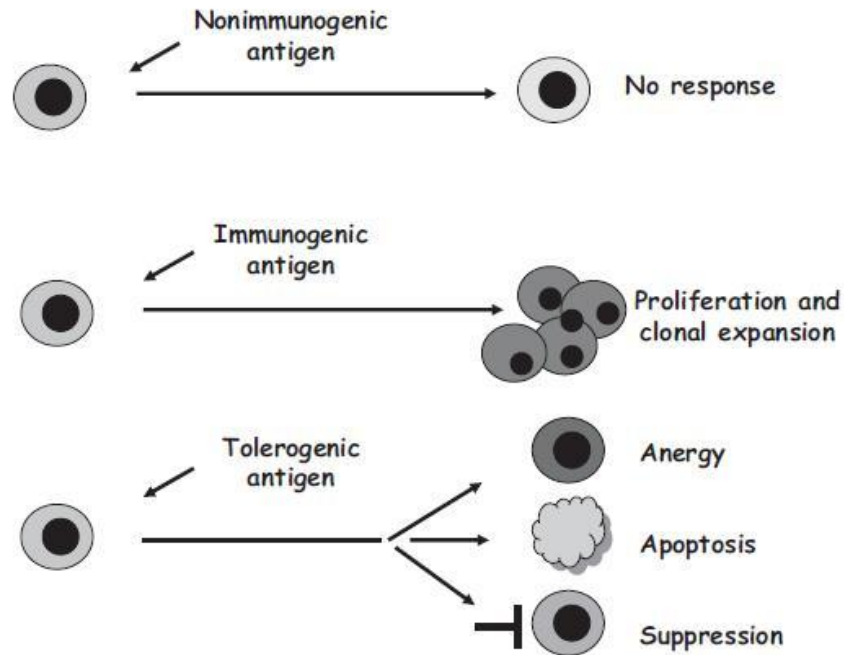


Figure 5: Possible events following the encounter between antigen and the immune system [18].

A thoroughly documented example of the cascade of steps occurring in the development of an autoimmune condition is how rheumatoid arthritis (RA) develops from the interaction of genomic and environmental factors, through the creation of neo-antigens activating lymphocytes leading to acute, then chronic synovitis which will result in joint destruction and later on systemic manifestations [20] (Figure 6).

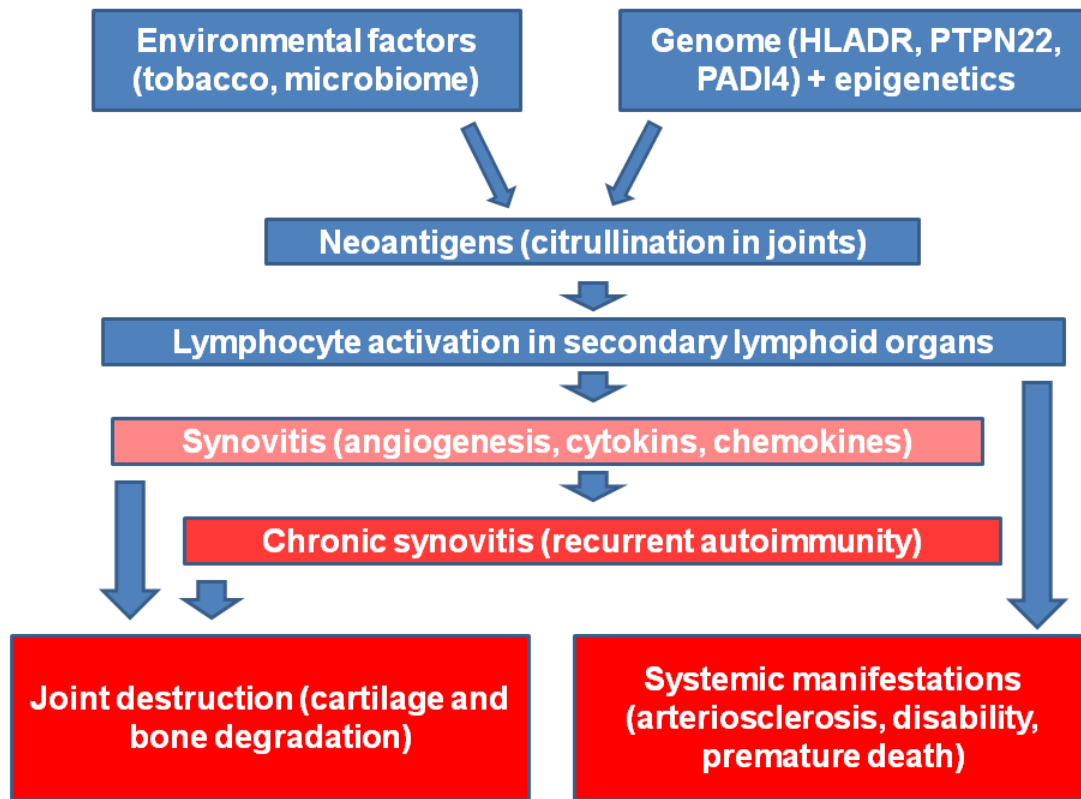


Figure 6: Cascade of events occurring during the development of rheumatoid arthritis.

Figure is a modified version of Figure 1 of Boissier et al. [20].

The common feature of autoimmune conditions is the presence of autoantibodies and the consequential inflammation that involves phagocytes, autoreactive T-lymphocytes and plasma cells which are autoantibody producing B cells [21]. Tissues and organs can be damaged through autoantibody-, and also cell-mediated ways. As most autoimmune diseases are chronic conditions, autoantibodies are present for a long time before clinical symptoms arise, inducing damage by binding to self-tissues, activating the complement cascade, inducing lysis, removal of cells by phagocytosis; interacting with cell-surface receptors (e.g. autoantibodies binding to the acetylcholine receptor block transmission at the neuromuscular junction and lead to myasthenia gravis).

The cell-mediated damage can be caused by phagocytic macrophages, neutrophils, self-reactive CD4⁺ T helper cells and self-reactive CD8⁺ cytolytic T cells, with smaller quantities of NK cells, mast cells and DCs. The mechanism of tissue or

organ damage can be based on killing cells directly or indirectly by releasing cytotoxic cytokines, prostaglandins, reactive nitrogen or oxygen intermediates (macrophages and neutrophils release cytotoxic nitric oxide and hydrogen peroxide). Macrophages can act as killer cells through antibody-dependent cell-mediated cytotoxicity. Also these released cytokines and mediators such as tumour necrosis factor (TNF) or interleukin (IL)-1 released by macrophages can recruit other inflammatory cells to the site of inflammation expanding the cytotoxic arsenal.

2.3 Economic aspects and the diversity of autoimmune conditions

A medical condition must meet the Witebsky's postulates in order to be classified as an autoimmune disease. These postulates include direct evidence from transfer of pathogenic antibody or pathogenic T cells; indirect evidence based on reproduction of the autoimmune disease in experimental animals; and circumstantial evidence from clinical clues [22]. Prominent examples of autoimmune diseases are Type I diabetes mellitus, SLE, Sjögren's syndrome, Hashimoto's thyroiditis, Graves' disease, Addison's Disease, RA, Inflammatory Bowel Diseases (IBD) including CD and UC; and allergies.

Autoimmune conditions can be grouped into organ-specific; and systemic conditions (Table 1); as well as conditions can be grouped based on the type of the autoantigen causing the disease (Table 2).

Table 1: Groups of autoimmune conditions.

Organ-specific autoimmune diseases	Systemic autoimmune diseases
Type 1 Diabetes Mellitus	Rheumatoid arthritis
Multiple Sclerosis	Scleroderma
Graves' disease	Systemic lupus erythematosus
Hashimoto's thyroiditis	Sjögren's syndrome
Crohn's disease	Allergy
Myasthenia gravis	Polymyositis

Table 2: Groups of autoimmune conditions based on the type of autoantigen involved in the disease.

Type of autoantigens	Disease
Intracellular DNA, RNA	SLE
Cell receptors	
Thyrotropin receptor	Grave's disease
Acetyl choline receptor	Myasthenia gravis
Insulin receptor	Diabetes mellitus
Cell membranes	
Platelets	Immune thrombocytopenic purpura
Proteins	
Immunoglobulins	RA, SLE
Hormons	
Insulin	Diabetes mellitus
Thyroid hormones	Thyreoiditis

Autoimmune disease as a category affects 50 million people in the United States being one of the top ten causes of death in women under the age of 65 [23], the top cause of morbidity in women in the United States; and the second highest cause of chronic illness. Hungarian data and analyses show that approximately 5-6% of the Hungarian population have a type of autoimmune condition [24].

Autoimmunity represents a substantial economic and social burden for societies and for public and private payers as the annual costs of only seven of the 100+ known autoimmune diseases (Crohn's disease - CD, ulcerative colitis - UC, Systemic lupus erythematosus - SLE, multiple sclerosis - MS, rheumatoid arthritis - RA, psoriasis and scleroderma) are estimated through epidemiological studies to a total of \$51.8-\$70.6 billion annually (Table 3).

Table 3: Estimated costs of selected autoimmune diseases in the United states.

Autoimmune Disease	Total estimated costs
---------------------------	------------------------------

Crohn's disease	10,9-15,5 billion USD
Ulcerative colitis	8,1-14,9 billion USD
SLE	2,2-9,6 billion USD
RA	19,3 billion USD
Psoriasis	11,2 billion USD
Scleroderma	1,5 billion USD

Moreover chronic inflammation leads to an elevated number of doctor visits, rising healthcare expenses and dysfunction of the infected tissues, and therefore more severe health complications. Patients often miss extended periods of employment, and as on average patients spend upwards of four years seeking an effective diagnosis, the time to which patients begin to obtain proper treatment and care for their illness extends which leads to major consequences such as organ damage and physical disability resulting in earnings losses and financial distress due to high out-of-pocket costs [25].

2.5 Examples of autoimmune conditions

2.5.1 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 prevalence is 1% worldwide, women are 3-5 times more susceptible to the disease [26], and the disease incidence appears to be highest for women between 40 and 50 years of age, and somewhat later for men [27].

RA is diagnosed mainly based on symptoms and signs, but also with blood tests (e.g.: RF) and radiological images (X-rays). In 1987, the American College of Rheumatology defined the exact criteria for the classification of rheumatoid arthritis [28], although the 2010 American College of Rheumatology (ACR) / the European League Against Rheumatism (EULAR) Rheumatoid Arthritis Classification Criteria were introduced in 2010 [29] overruling the old one. The criteria establish a point value between 0 and 10 based on which every patient with a point total of 6 or higher is classified as an RA patient, provided synovitis is present in at least one joint and given that there is no other diagnosis better explaining the synovitis. Four areas are covered in the diagnosis following the new criteria:

- joint involvement, designating the metacarpophalangeal joints, proximal interphalangeal joints, the interphalangeal joint of the thumb, second through fifth metatarsophalangeal joint and wrist as small joints, and shoulders, elbows, hip joints, knees, and ankles as large joints:
 - Involvement of 1 large joint gives 0 points
 - Involvement of 2-10 large joints gives 1 point
 - Involvement of 1-3 small joints (with or without involvement of large joints) gives 2 points
 - Involvement of 4-10 small joints (with or without involvement of large joints) gives 3 points
 - Involvement of more than 10 joints (with involvement of at least 1 small joint) gives 5 points
- serological parameters – including the rheumatoid factor as well as ACPA – "ACPA" stands for "anti-citrullinated protein antibody":
 - Negative RF and negative ACPA gives 0 points
 - Low-positive RF or low-positive ACPA gives 2 points
 - High-positive RF or high-positive ACPA gives 3 points
- acute phase reactants: 1 point for elevated erythrocyte sedimentation rate, ESR, or elevated CRP value (C-reactive protein)
- duration of arthritis: 1 point for symptoms lasting six weeks or longer

X-rays may demonstrate juxta-articular osteopenia, soft tissue swelling and loss of joint space in the early stages of the disease, but in later stages, there may be bony erosions and subluxation (Figure 7).



Figure 7: X-rays of the hands of an RA patient with deformities.(Image source: [30])

Regarding differential diagnosis, conditions such as osteoarthritis, SLE, PsA, Lyme disease, reactive arthritis and AS can resemble RA, among others.

Symptoms include the arthritis of joints known as synovitis affecting multiple joints (polyarthritis); rheumatoid nodule, a necrotizing granuloma, a type of inflammatory reaction on the skin; fibrosis of the lungs; renal amyloidosis; atherosclerosis, the risk of myocardial infarction and stroke, among others.

Deformities caused by late-stage RA can be ulnar deviation of the bones of the hands, swan-neck deviation of the fingers due to the muscles and tendons on one side of the joint can overpower those on the other side, pulling the bones out of alignment (Figure 8).

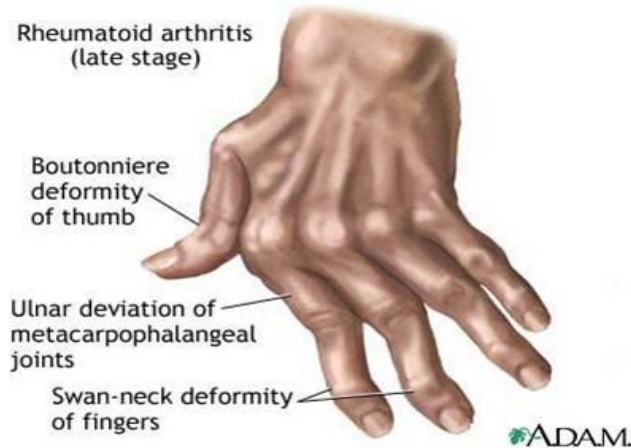


Figure 8: Hand deformities caused by late stage RA (Image source: Adam Inc.)

In the joint's synovial membrane, fluid is secreted, the synovium becomes inflamed, and the cartilage becomes rough and pitted (Figure 9).

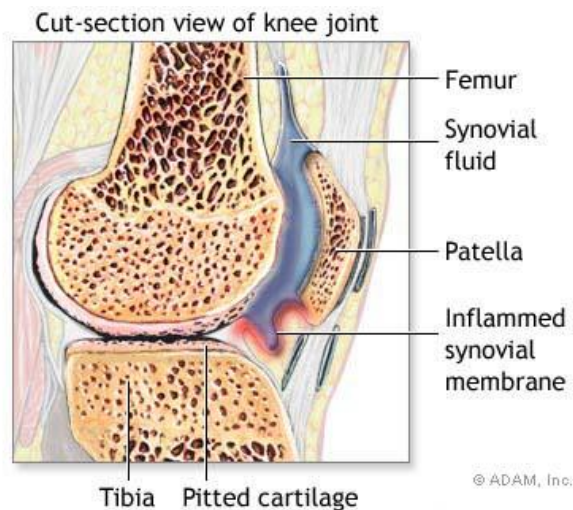


Figure 9: Fluid is secreted following inflammation of the synovial membrane which leads to the degradation of bone and joints. (Image source: Adam Inc.)

The causes in the pathogenesis of the formation of autoimmunity in RA are not completely known yet, but several details have been described. There is a genetic link with HLA-DR4 and related allotypes of MHC Class II and the T cell-associated protein PTPN22 [31]; and a link to the pathogenesis of vascular diseases such as rheumatoid vasculitis [32]. RA involves abnormal interaction between B and T cells as well as the

presentation of antigens by B cells to T cells via HLA-DR eliciting T cell help and consequent production of RF and antibodies to citrullinated peptides (ACPA).

There is no cure for RA, but the goal of minimizing clinical symptoms can be reached by using one type of drug or the combination of these as the ACR suggests. Such treatment options include analgesics and non-steroid anti-inflammatory drugs; and disease modifying anti-rheumatic drugs (DMARDs). DMARDs can reduce the rate of damage to bone and cartilage. Combinations of DMARDs are methotrexate – hydroxychloroquine, methotrexate – sulfasalazine, sulfasalazine – hydroxychloroquine, and methotrexate – hydroxychloroquine – sulfasalazine.

The most recent therapies include biological agents (e.g. TNF α 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 [33]. In these cases, future gene profiling based on peripheral blood mononuclear cells might prevent non-responders from starting the inefficient therapy.

Infliximab, a genetically constructed IgG1 murine-human chimeric monoclonal antibody binding to the soluble subunit and the membrane-bound precursor of TNF α forming stable non-disassociating immune complexes [34], has proven to be a successful treatment for RA [35]. Infliximab regulates the cytokine network causing reductions in C-reactive protein (CRP), IL-6, IL-1ra (IL-1 receptor antagonist) and soluble TNF receptor concentrations. It also regulates cell recruitment through a reduction in the cellularity of the synovium which leads to the reduction in the swelling and tenderness of joints [36]. Moreover, it regulates a major angiogenic factor (VEGF) and angiogenesis; and prevents cartilage catabolism and bone erosion [33].

Besides the different types of TNF α inhibitors [37], emerging biotherapies such as inhibitors of the IL-1 or interleukin-6 (IL-6) pathways have also been added to the therapy options in RA. IL-6 can activate cells through binding to membrane-bound and soluble receptors which leads to acute and chronic inflammation; joint destruction, pannus development and increased bone resorption [38]; as well as inflammatory cell migration [39]. Tocilizumab which is a humanized anti-IL-6R monoclonal antibody that can block IL-6 mediated signal transduction is approved as a biological therapy in

combination with methotrexate (MTX) for patients with moderate to severe RA who showed inadequate response to DMARDs or TNF α inhibitors [40] or to those being intolerant to that therapy. Tocilizumab was proved to cause a significant reduction in disease activity [41-42].

2.5.2 Ankylosing spondylitis

Many of the disease characteristics of RA are similar to AS (previously known as Bekhterev's disease) which is a chronic inflammatory arthritis and autoimmune disease affecting joints in the spine and the sacroilium in the pelvis. The overall prevalence is 0.25%, approximately 90% of AS patients express the HLA-B27 genotype, however, only 5% of individuals with the HLA-B27 genotype contract the disease [43]. TNF α and IL-1 also implicated in this disease. There is no cure for AS, but NSAIDs, corticosteroid therapy and medications suppressing the immune system can be used, as well as TNF-inhibitors such as etanercept, adalimumab and infliximab [44].

2.5.3 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 (Figure 10-11).



Figure 10: Typical skin signs on the hands in psoriasis. (Image source: Adam Inc.)



Figure 11: Skin signs in psoriasis on the elbow. (Image source: [45])

Its prevalence varies considerably affecting approximately 2% of the population only in the United States [45]. Both women and men can be affected at any ages. Diagnosis is based on the skin symptoms. In this process, a score system (PASI) assists clinicians in classifying patients by the severity and location of skin symptoms [46].

The cause of the disease is not fully understood. According to one of the hypotheses, T cells that become active migrate to the dermis and trigger the release of cytokines such as TNF α causing inflammation and rapid production of skin cells.

While several factors are thought to aggravate psoriasis including stress, excessive alcohol consumption, and smoking, there are thoroughly documented genetic associations such as 7 genetic loci associated with the condition (PSORiasis Susceptibility genes: PSORS-1, -2, -3, -4, -5, -6, -7) [47].

It is not curable, but treatable, though different subtypes of the disorder lead to different sensitivity of pharmacotherapy. Besides topical, systemic and phototherapy, biologics are included in the management of the condition.

2.5.4 Psoriatic arthritis

PsA is a type of inflammatory arthritis that develops in about 30% of patients with psoriasis. PsA is a seronegative spondyloarthropathy and occurs more commonly in HLA-B27 positive patients [48] [49].

Common symptoms include pain, swelling, or stiffness in one or more joints; dactylitis, tendinitis in the Achilles tendon or Plantar fasciitis in the sole of the foot. The exact cause of the disease is not known, and regarding treatment, NSAIDs, DMARDs and biologics are generally considered.

2.5.5 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 [50]. 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 (Figure 12).

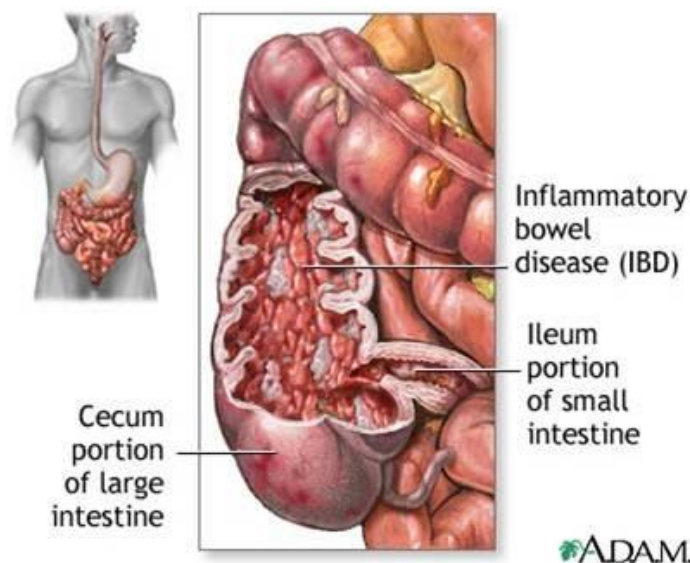


Figure 12: Inflamed colon mucosa. (Image source: Adam Inc.)

CD can affect any region in the gastrointestinal system, but 50% of cases are located in the ileum and the large intestine. It may invade deeper tissues of the intestinal wall and spread to involve more areas of the bowel. Ulcers may form at the sites of the most intense inflammation and are usually separated by areas of healthy tissue with little or no inflammation. Sometimes, first symptoms are caused by anal abscess or/and fistula.

The exact cause of CD remains undiscovered, but innate or acquired predisposition and an external trigger (infection, stress) might be a possible explanation. In details, impaired cytokine secretion by macrophages leads to impaired innate immunity which will result in a sustained microbial-induced inflammatory response in the colon [51].

The first gene to be associated with CD was NOD2 (also known as CARD15) [52].

UC is the inflammation of the large intestines that includes characteristic ulcers, or open sores in the colon. It occurs in 35-100 people for every 100,000 in the United States [8]. It is normally continuous from the rectum up the colon. In the most severe form, the whole bowel is involved (pancolitis). It is an intermittent disease, with periods of exacerbated symptoms, and periods that are relatively symptom-free. Its exact cause is also unknown.

2.6 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 ($\approx 1\%$) [53]; but the amount of mRNA derived from platelets is extremely small making it inadequate for gene expression analysis [54].

Whole blood represents another major sampling option for the transcriptome analysis of peripheral blood. PAXgene tubes containing nucleic acid stabilization solution enable to examine whole blood gene expression pattern. Using PBMCs or whole blood samples have advantages and disadvantages (Table 4).

Table 4: Comparison of PBMC and whole blood in gene expression profiling.

Whole blood		PBMC	
Advantages	Disadvantages	Advantages	Disadvantages
RNA gets stabilized immediately	High cost of special collection tubes (i.e Paxgene)	More homogenous cell populations	Potential ex vivo changes due to processing and/or shipping
Can be frozen, shipped overnight	Contaminated by red blood cell RNA	Lymphocytes and monocytes are the most transcriptionally active cells in blood	Influence of changing blood cell count
No need to separate cells	High levels of globin transcripts	Low cost of collection tubes (Tubes containing EDTA)	Requires processing (Ficoll gradient centrifugation)

A major limitation of whole blood sampling is the variation of gene expression it shows regarding cellular diversity. Another one is a decrease in the number of genes truly present which is related to a bias of reticulocyte-derived globin transcripts. Globin mRNAs may accumulate to more than 95% of total cellular mRNA as red blood cells represent the major type of cells investigated with the PAXgene system leading to a conclusion that globin-related transcript amplification is likely to occur in gene

expression profiling. This might suggest that PBMC is the more suitable choice of source [55].

The first thoroughly documented examples demonstrating that gene expression profiling could have clinical utility analyzed the distinct subtypes of diffuse large B-cell lymphoma [25, 56] and human breast cancer progression by combining laser-capture microdissection and DNA microarrays in order to predict clinical outcome [57]. They showed that the stages of breast cancer progression such as premalignant, preinvasive and invasive stages could be distinguished at the transcriptome level with a set of genes associated with high tumor grade which correlated with the transition from preinvasive to invasive growth.

Basically, there are three methods for obtaining gene expression data. The most sensitive and targeted way is real-time polymerase chain reaction (RT-QPCR); a laboratory technique based on PCR and allows for the quantification of mRNA molecules. RT-QPCR measurements could be low-throughput when only few genes are detected [58] or could be low- to medium-throughput using 96- and 384-well assay plates or TaqMan Low Density Array (TLDA), which is a 384-well micro fluidic card that enables to perform 384 simultaneous real-time PCR runs [59].

When the goal is to perform a global screen of gene expression patterns, microarray is the chosen platform which is a collection of microscopic DNA spots containing picomoles of a specific DNA sequence known as probe attached to a solid surface. These probes are representing distinct genes and used to hybridize fluorescently labeled cDNA or cRNA samples under high-stringency conditions; and the probe-target hybridization is detected and quantified (Figure 13). Examples include the global expression profiles of subclasses of RA patients [60].

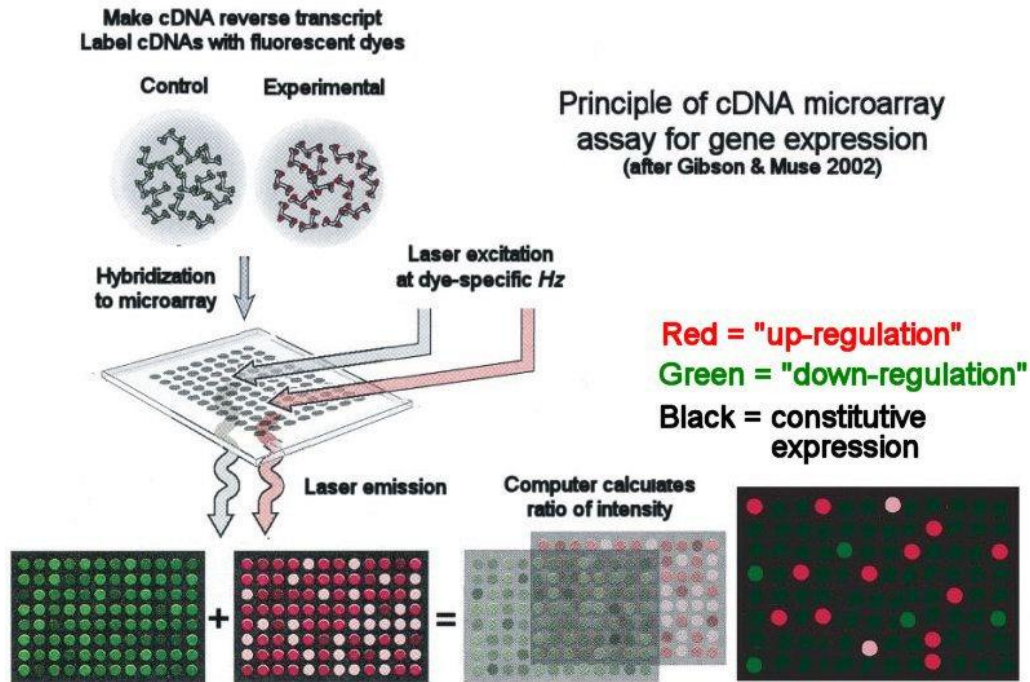


Figure 13: Schematic outline of microarray experiments. (Image source: [61])

Microarray analyses of PBMC of healthy individuals have proven that gene expression profiles appear to be consistent allowing for the determination of a „normal” expression profile [62]. Evidence for distinct patterns reflecting inter-individual variation due to differences in the cellular composition of the blood sample, gender, age and time of day was also shown. A genome-wide gene expression signature interrogating regional population differences such as lifestyle, geography and biotic factors found in peripheral blood leukocyte samples of desert nomadic, mountain agrarian and coastal urban Moroccan Amazigh individuals suggested that gene expression variation among and within populations could complicate the process of designing gene arrays for clinical practice [63].

2.7 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 [64]; skin lesions in psoriasis [65] and synovial tissue in RA [66].

It was pointed out before that the molecular signature of disease across tissues is more prominent than the signature of tissue expression across diseases [67]. 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 [68]; 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 [69].

A summary of the types of autoimmune conditions investigated most in gene expression profiling, as well as the method used by these studies is available in Figure 14.

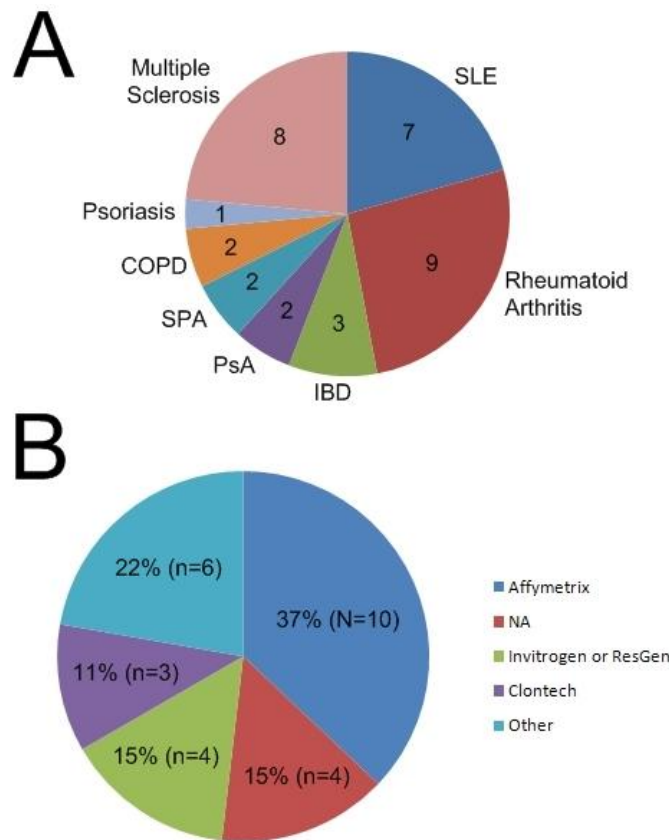


Figure 14. Distribution of PBMC gene expression profiling studies based on (a) disease and (b) gene expression platform. Studies included in Table 2 were analyzed and the number of studies focusing on each autoimmune disease was visualized as well as the number of different platforms used for global gene expression analyses in these studies. Abbreviations: ResGen, Research Genetics which was acquired by Invitrogen; NA, not described.

As RA is considered a homogeneous disorder with tissues such as synovial fluid and PBMC relatively easily available for analysis, it has been in the focus of gene expression studies.

Global gene expression analysis on synovial tissue biopsies and PBMC samples from the same RA patients revealed differential tissue pathology which was not reflected in PBMC at the single gene level [13]. MHC-related genes [70] and interferon (IFN) mediated immunity [59] have been traditionally associated with this condition.

Due to the numerous disease characteristics of AS that are similar to those of RA, these conditions are usually studied together as in the evaluation of PBMCs which identified CXCR4 and SDF-1 as parts of a potential pro-inflammatory axis for AS, RA and PsA [71].

In psoriasis, gene expression profiles are primarily studied in skin biopsy samples with the inclusion of lesional and nonlesional skin [72], involved and uninvolved skin of patients with psoriasis versus healthy controls [73], although global gene expression analysis in PBMCs of psoriasis patients resulted in a list of 18 differentially expressed genes [74].

In PsA, the expression levels of 313 genes showed significant differences compared to healthy control subjects in PBMC [75].

The transcriptome of IBD is usually examined at tissue level by obtaining colon biopsy [64]. Global gene expression profiling with microarrays of CD and UC patients revealed 12 genes distinguishing these patients from each other with high accuracy [76]. On the single gene level, CCL20 expression in PBMC seemed to correlate with the degrees of disease severity, erythrocyte sedimentation rate and CRP [77].

3. 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.

4. Methods

4.1 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 (Figure 15). Samples were processed within one hour after sample collection.

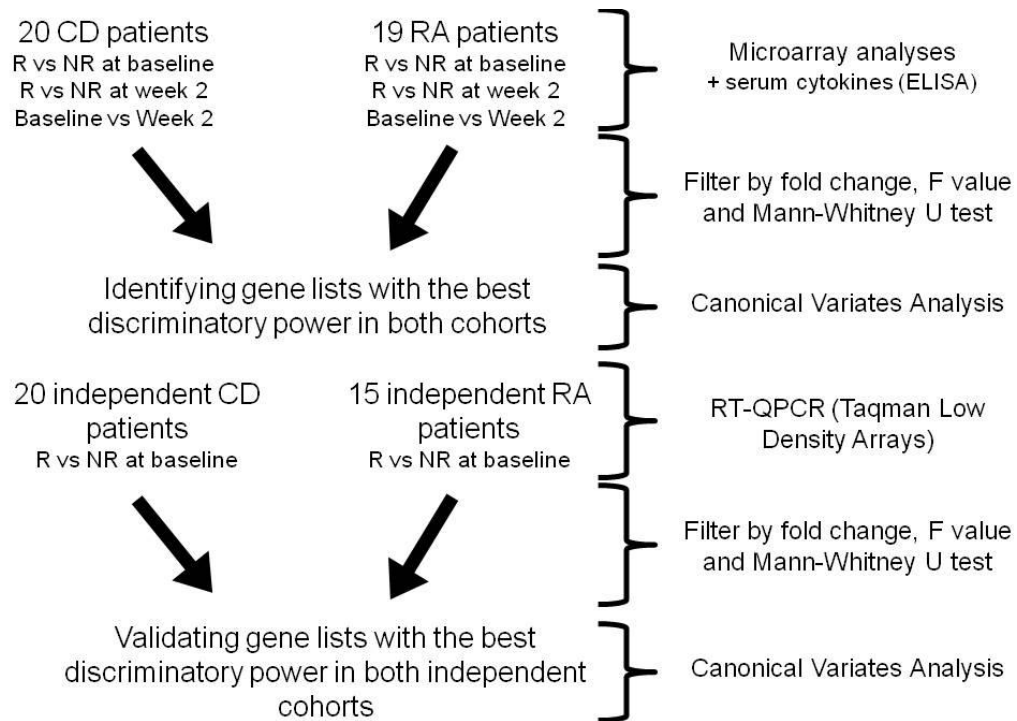


Figure 15: Schematic outline of the infliximab study.

Inclusion and exclusion criteria for CID study subjects were determined by clinicians (Table 5).

Table 5: Inclusion and exclusion criteria for the CID study.

	Inclusion criteria	Exclusion criteria
IBD patients (n=13)	Clinical diagnosis of Crohn's disease or ulcerative colitis Age between 15 and 40 Newly diagnosed, treatment-free patients	Pregnant or breastfeeding women Current or previously known cancer Clinically essential co-morbidity
Psoriasis patients (n=15)	Active psoriasis (chronic plaque form) PASI score >10 Age between 15-70 Haven't received systemic therapy for at least 4 weeks (retinoids, methotrexate, cyclosporin, azathioprin) or light therapy Have no other inflammation Have no other skin symptoms	Pregnant or breastfeeding women Current or previously known cancer Clinically essential co-morbidity
Rheumatoid arthritis patients (n=12)	Clinical diagnosis of rheumatoid arthritis Age between 25 and 70 Haven't received systemic therapy for at least 4 weeks (antiallergic drugs or immunosuppressant agents) DAS28 (Disease Activity Score) > 5.1 Sex and age-matched patients Have no inflammation	Pregnant or breastfeeding women Current or previously known cancer Clinically essential co-morbidity
Healthy controls (n=18)	Haven't received anti-inflammatory drugs for at least 4 weeks	Pregnant or breastfeeding women Current or previously known cancer Clinically essential co-morbidity

Thirteen patients (nine females, four males) who met the American College of Rheumatology criteria for RA were included in the tocilizumab study. The inclusion criteria in our study were 1) fulfillment of the 1987 American Rheumatism Association criteria; 2) receiving concomitant MTX treatment of maximum 20 mg/wk; 3) age between 30 and 60; 4) failure to respond to at least two DMARDs; 5) active disease defined as having a disease activity score evaluated in 28 joints (DAS28) >3.2; 6) having stable MTX, prednisolone and NSAID doses during the previous 4 weeks to the inclusion in the study 7) having discontinued previous DMARDs at least 4 weeks prior to the inclusion. Exclusion criteria were 1) pregnancy or breastfeeding; 2) current or recent cancer; 3) active infectious disease; 4) patients with a history of an acute inflammatory joint disease of different origin. All patients were TNF blocking therapy-naïve.

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.

The inclusion criteria in RA were fulfillment of the 2010 EULAR/ACR classification criteria; age between 20 and 60, failure to respond to at least two DMARDs, active disease (DAS28 >3.2), anti-TNF therapy-naïve patients or previous anti-TNF use at least 3 months prior to blood sampling, prednisone therapy ≤10mg per day was allowed provided that the dosage has been stable for at least 2 months before entry and non-steroidal anti-inflammatory drugs were allowed in doses stable for at least 1 month before baseline, patients were on maximal tolerable methotrexate treatment (5–30mg per week), which had to be stable for at least 4 weeks before baseline. Exclusion criteria were 1) pregnancy or breastfeeding; 2) current or recent cancer; 3) active infectious disease; 4) patients with a history of an acute inflammatory joint disease of different origin, and 5) smoking.

Inclusion criteria in CD were clinically diagnosed Crohn's disease, age between 20 and 60, Crohn's disease activity index (CDAI) >250, anti-TNF therapy naive patients, prednisolone therapy with a dosage of less than 10 mg/day. Exclusion criteria were 1) pregnancy or breastfeeding; 2) current or recent cancer; 3) active infectious disease; and 4) smoking.

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).

4.2 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. RNA quality was checked on Agilent Bioanalyzer 2100 (Agilent Technologies), all samples had a 28S/18S ratio between 1.5 and 2.0 and the RNA Integrity Number was between 9 and 10. Quantity was determined by NanoDrop (Thermo Scientific).

4.3 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).

4.4 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. Our custom designed TLDA card allows for 2 samples to be run in parallel against 96 TaqMan gene expression assays. 91 genes were chosen based on our previous experiments and the remaining 5 genes were housekeeping genes for normalization. (ACTB, GAPDH, HPRT1, PPIA, RPLP0). cDNA was generated with High Capacity cDNA Reverse Transcription Kit (Life Technologies) according to manufacturer's protocol. RT-QPCR amplification was performed using an Applied Biosystems Prism 7900HT sequence detection system with the following thermal cycler conditions: 2 min at 50°C and 10 min at 94.5°C, followed by 40 cycles of 30 s at 97°C and 1 min at 59.7°C. 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. This method does not include correction for multiple comparisons and is widely used in the analysis of TLDA data [78] [79] and is explained in details [80]. 96 genes included in the TLDA cards were selected from the microarray experiments, also genes selected from the literature and 5 housekeeping genes. Mann-Whitney U-test was used in the validation phase.

Gene interactions were analyzed by using the Direct Interactions and Biological Processes functions of GeneSpring GX10 (Agilent Technologies) and GeneMania [81].

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. Reactions were performed in an ABI Prism HT 7900 machine in triplicate and all samples ($n = 26$) were included in the validation set. 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).

4.5 Gene List Selection Process of CID study

A database containing 400 genes which are associated with the 3 chronic inflammatory diseases and inflammation was constructed by using the data derived from the Human Genome Navigator that lists all the genes related to a specific disease and the evidence the relation is based on; the current medical literature and international databases (e.g. OMIM, Entrez Gene). Genes were also selected from genome-wide association studies as well as microarray analyses focusing on skin biopsy in psoriasis, colon biopsy in IBD and synovial fluid in RA.

In the second step, genes were put in order by using a score list that was based on the number of publications mentioning the disease-gene association, relation to other diseases and expected expression in PBMC according to previous studies and our pilot experiment in which we analyzed selected genes in PBMC samples through individual RT-QPCR assays. This selection process resulted in 150 genes.

In the third step, genes that did not show expression (missing signal in QPCR measurements) in the pilot study were removed, leading to a final list of 96 genes (Figure 16).

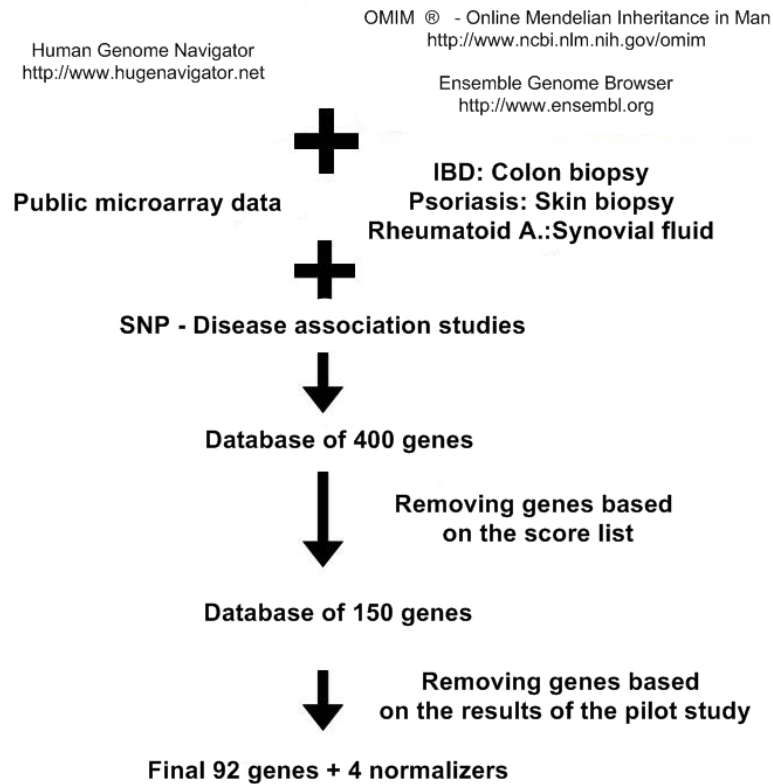


Figure 16: Schematic outline of the CID study. A flow chart of the gene selection process shows the steps through which we chose the final 96 genes out of the database containing 400 genes.

4.6 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 baseline vs. week 2 samples 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 paired Mann-Whitney U-test with Benjamini-Hochberg multiple-testing correction.

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.

Functional categorization of genes was performed with Panther Classification System (<http://www.pantherdb.org/>).

4.7 Multivariate data analysis

4.7.1 Principal Components Analysis (PCA)

PCA reduces the dimensionality of the data in such a way that relatively few artificial and uncorrelated dimensions, principal components, will explain as much of the variation as possible. The success of variance compression is data-dependent and is measured by the relative percentages pertaining to each component. Often, 2-3 components explain 60-70% of the variation, allowing graphical display of results by a biplot, which is a simultaneous arrangement of study objects and original variables for a given pair of components. The role of variables in influencing data structure can be evaluated on the basis of length and directionality of arrows pointing to variable positions. These are obtained after arbitrary rescaling of component correlations to allow for effective visualization.

4.7.2 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. 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.

Since the maximum number of canonical axes is one less than the number of groups, in the present study CVA did not allow graphical display, and separation of responders and non-responders is expressed merely by a list of scores for observations on a single canonical axis. If the observations are taken at random and the variables satisfy multivariate normality, then statistical procedures are available to test the significance of group separation. Nevertheless, if these criteria are not met, as in our case, examination of the two groups whether they overlap on the canonical axis or not, provides equally meaningful information. A partial limitation of CVA is that the number of variables cannot exceed the number of observations (patients). Therefore, the genes were examined by F-test to filter out those not separating the two groups of patients at all, thereby reducing the number of genes used below the total number of patients.

Computations were performed by the SYN-TAX 2000 package [82].

4.7.3 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. CVA 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. 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.

Linear discriminant analysis (LDA) [83] 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). (Figure 17)

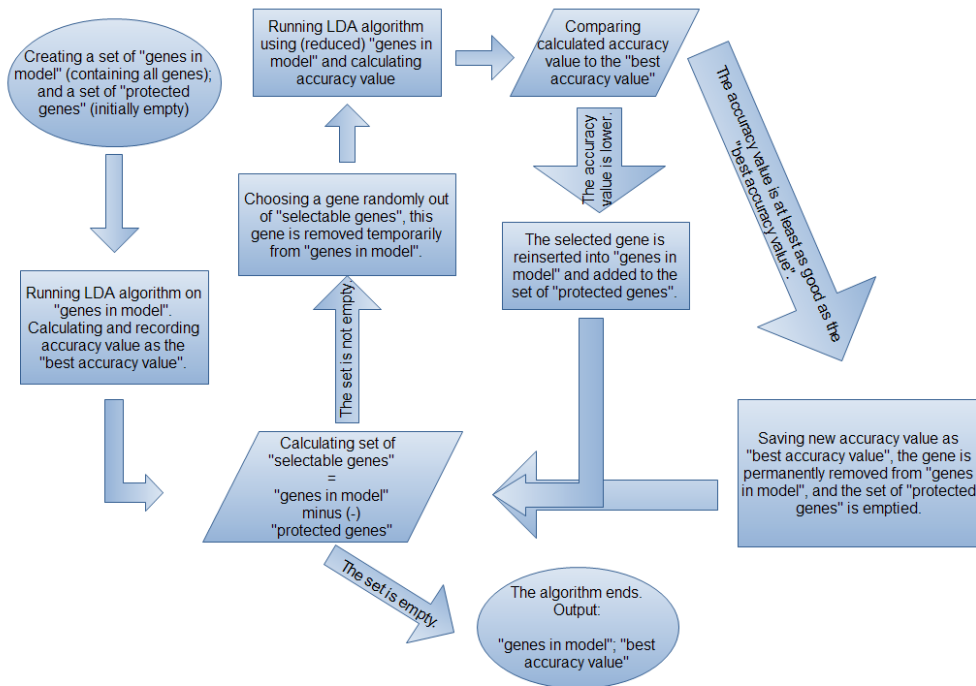


Figure 17: Schematic outline of the automatic gene panel generation.

- 1) The set of 'genes in model' is created. Initially, this set contains all genes. A set of genes with so-called 'protected genes' is also created. Initially, this set is empty.
- 2) F-value that is the ratio of between-group variability and within-group variability is calculated for each gene.
- 3) The classifier algorithm (LDA) is run using the set of 'genes in model' both in test and validation cohorts. In both cases an accuracy percentage value is recorded as the 'best accuracy values'.

4) The set of ‘selectable genes’ is defined as: ‘selectable genes’:= ‘genes in model’ minus ‘protected genes’. If the group of ‘selectable genes’ is not empty, the algorithm is continued in step 5. Otherwise, the algorithm skips to step 7.

5) Genes are selected from the set of ‘selectable genes’ according to the following models:

- a) Randomly with equal probabilities (uniform model);
- b) Randomly with a probability that is inversely proportional to their F-value (F_prop model);
- c) Genes with the lowest F-values (min F model).

In either case, the selected gene is temporary removed from the set of ‘genes in model’. The advantage of using stochastic models instead of min F model is that those can provide better segregation of patient groups. Uniform and F_prop models represent stochastic algorithms while min F model is deterministic.

6) The classifier is run using the (temporary reduced) set of ‘genes in model’.

- a) If any of the two accuracy percentage values becomes lower, the selected gene is reinserted into the set of ‘genes in model’ and added to the set of ‘protected genes’.
- b) If both accuracy percentage values are at least as good as the ‘best accuracy values’, the selected gene is permanently removed from ‘genes in model’ and the set of ‘protected genes’ is emptied. The ‘best accuracy values’ are overwritten with the calculated accuracy values. The algorithm returns to step 4.

7) The algorithm ends. The outputs are the set of ‘genes in model’ and ‘best accuracy values’.

Linear discriminant analysis was performed by using R software (R Development Core Team (2008) with package MASS[84].

4.8 ELISA

We measured the serum levels of IL-6, IL-8, IL-12, and IFN γ by using enzyme-linked immunosorbent assays (ELISA) in the infliximab study. Quantikin ELISA kits (R&D Systems) were used according to manufacturer’s protocol and the results were

given in pg/ml. Statistical analysis was performed with Prism using Mann Whitney U-test regarding the NR vs. R comparisons, and paired T test regarding the baseline vs. week 2 comparisons.

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.

4.9 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 N-glycans were released by peptide-N-glycanase F (PNGase F). The released glycans were then fluorescently labeled with aminopyrene-trisulfonate (APTS) and analyzed by capillary gelelectrophoresis with laser induced fluorescence detection (CGE-LIF). 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).

5. Results

5.1 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.

5.1.1 Identification of gene panels discriminating chronic inflammatory disease patients from healthy controls

Fifty-three genes show statistically significant differences between diseased and healthy samples after the expression levels of 96 selected genes were measured by using TLDA. Regarding the comparisons of one particular disease and the healthy cohort, 25 genes were found to be significant in case of IBD; 16 genes in case of psoriasis; and 33 genes in case of RA (Figure 18).

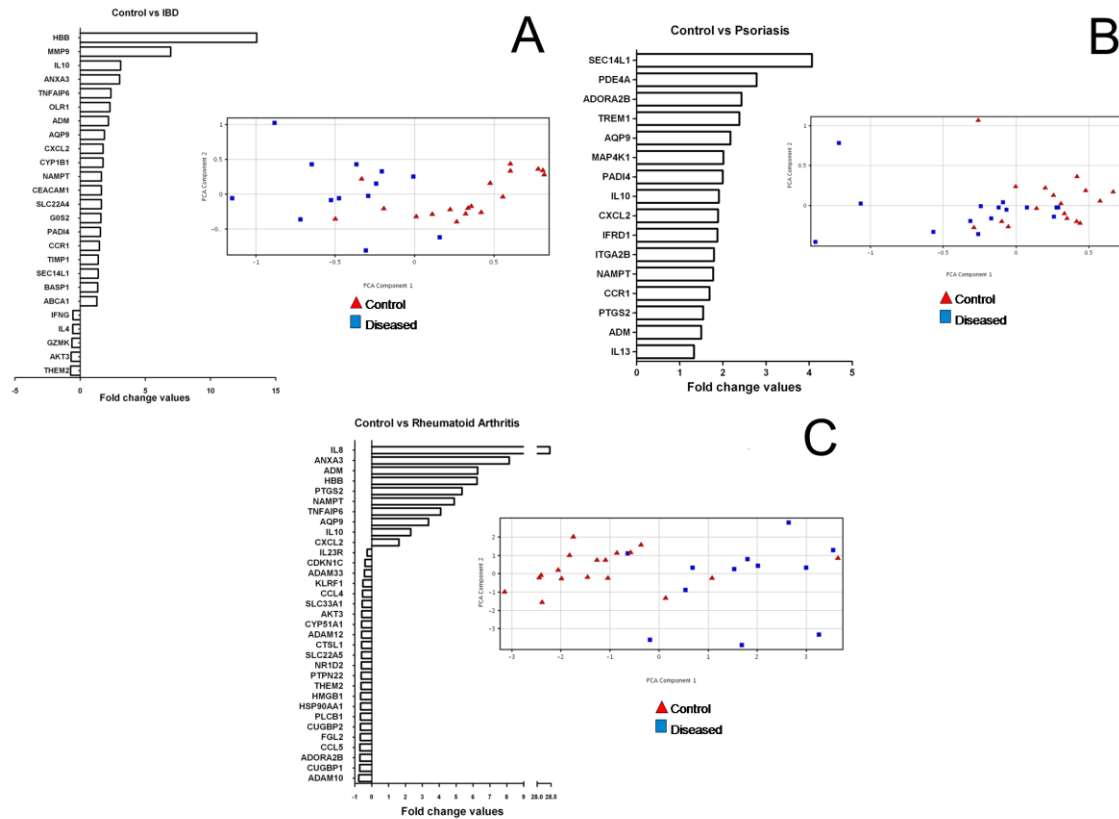


Figure 18: Fold change values of genes, showing statistically significant (Mann-Whitney U test) differential expression between diseased and control patients, were generated from RT-QPCR measurements and represent the difference of the means of the diseased and control groups. Principal component analysis was performed and separates the two groups of samples. Figure 18 A represents the IBD, B the psoriasis and C the RA gene panel.

Principal component analysis was performed to show how these gene sets can separate the two groups of samples, and gene interaction analyses highlighted genes such as IFNG, IL4, IL10, MMP9 and TIMP1 in IBD; IL10, IL13 and PTGS2 in psoriasis; IL8, IL10 and PTGS2 in RA as genes with the most direct interactions that might give a hint about their key role in the pathogenetic background of these diseases (Figure 19a-c).

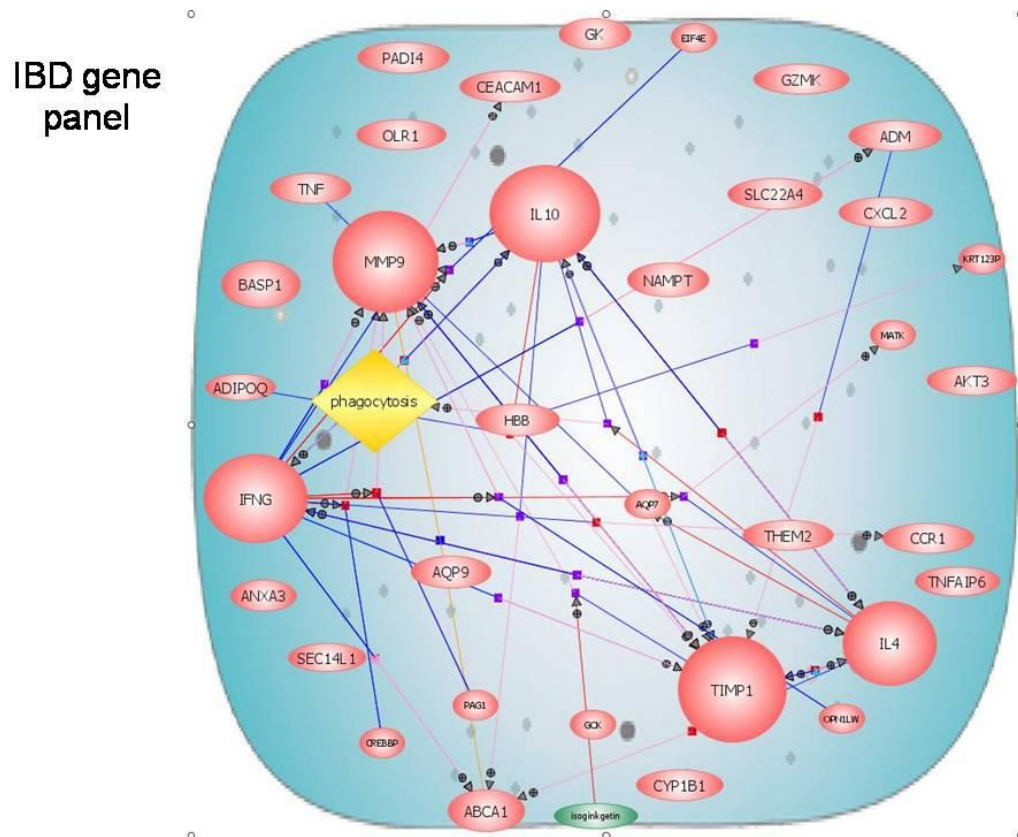


Figure legends:

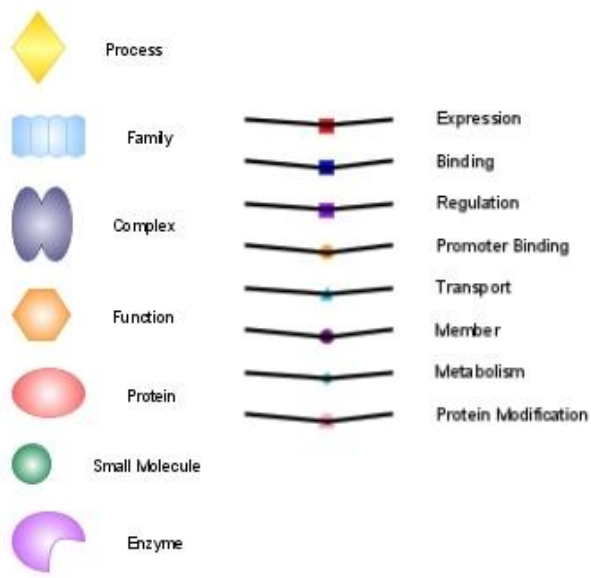


Figure 19a: Gene interaction analysis of the genes differentiating between IBD and controls.

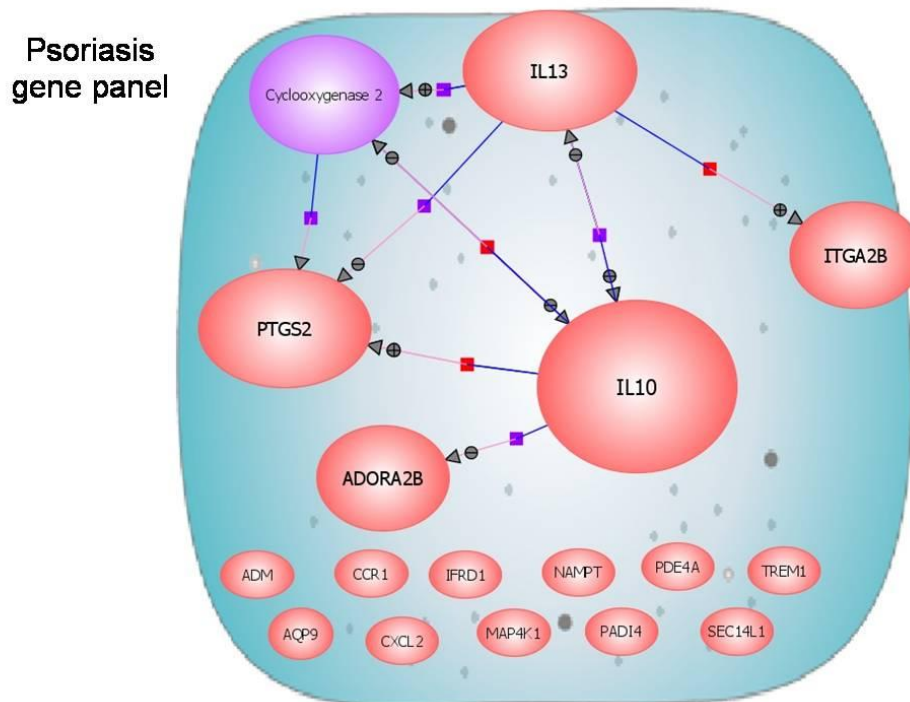


Figure 19b: Gene interaction analysis of the genes differentiating between Psoriasis and controls.

RA gene
panel

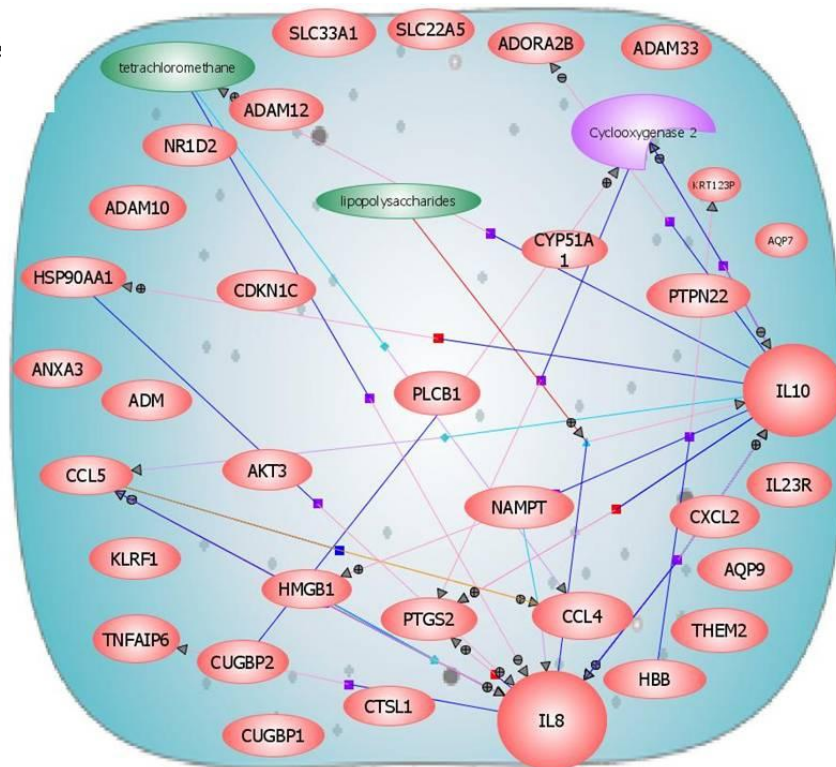


Figure 19c: Gene interaction analysis of the genes differentiating between RA and controls.

5.1.2 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 (Figure 20).

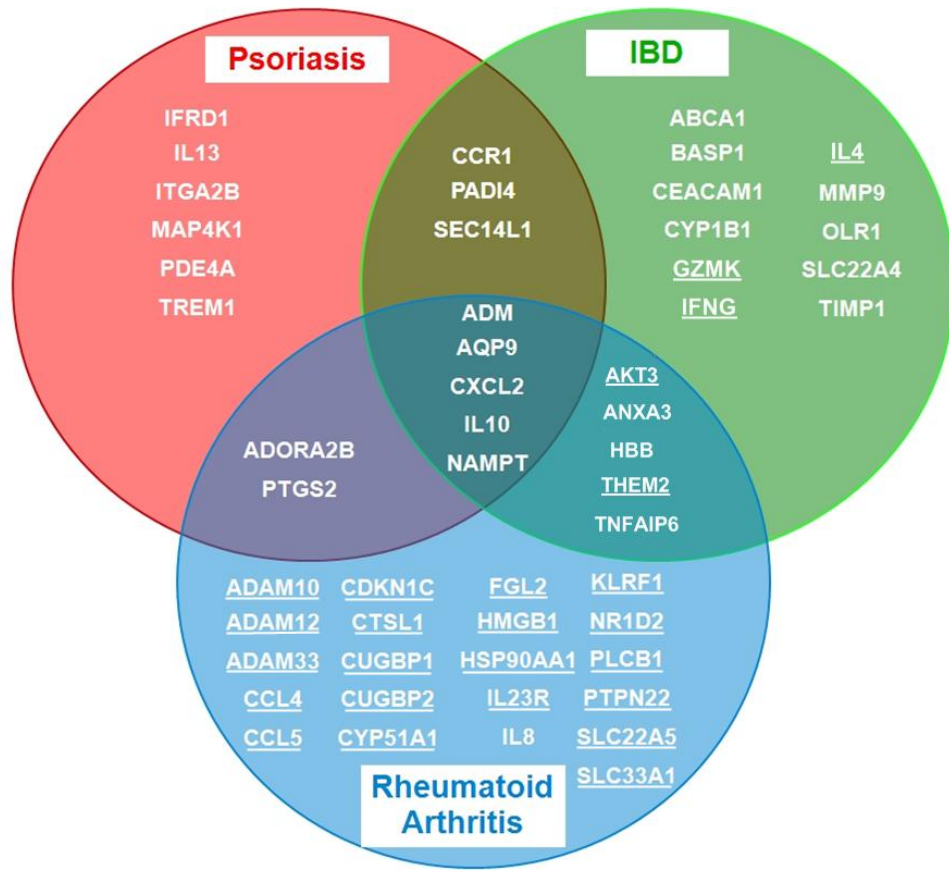


Figure 20: Venn diagram shows all of the genes with significant differences between a particular disease group and control samples. Underlined genes were down-regulated compared to healthy controls.

Gene Interaction Analysis identified a network of genes such as CCL5, IFNG, IL4, IL8, IL10, IL13, MMP9, PTGS2 and TIMP1 highlighting the interactions among them (Figure 21). It underscores the notion that entire networks are impacted and can be identified. As these genes with significant differences between diseased and healthy samples form a network together, it might represent the common pathogenetic background of these conditions.

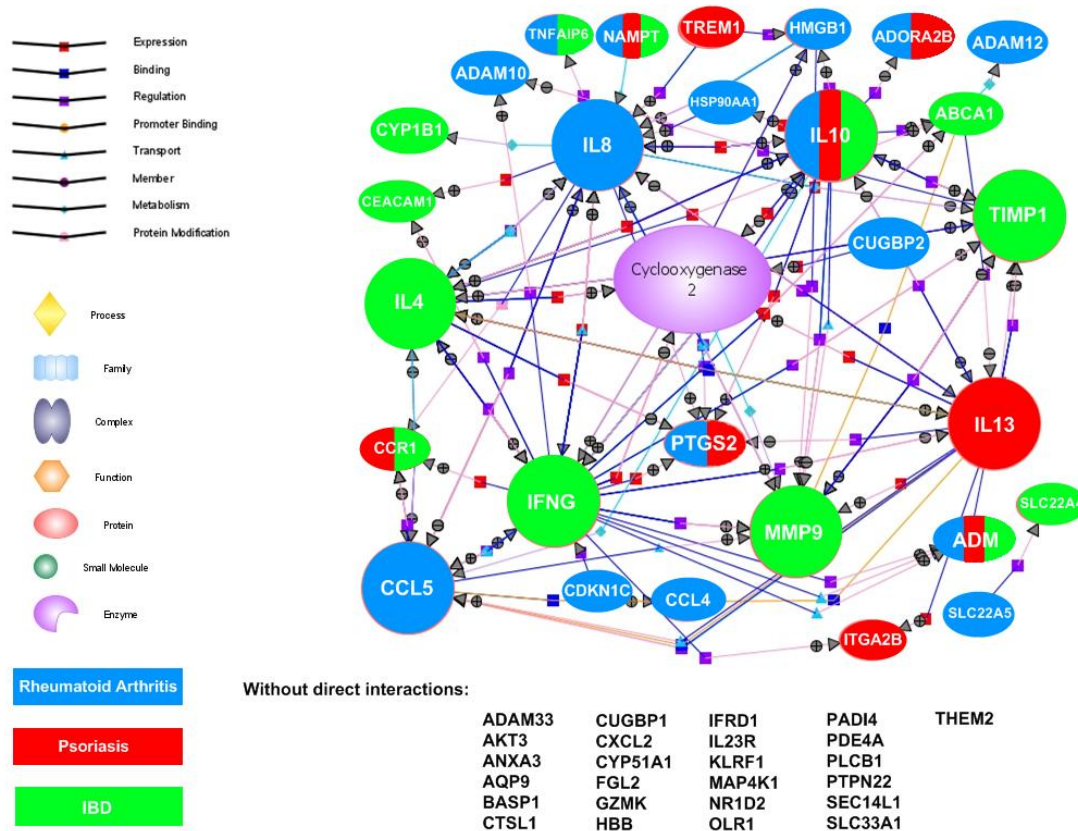


Figure 21: Gene interaction analysis highlights 28 genes that have direct interactions with each other while 25 genes have no direct interactions. Genes with the highest number of interactions are shown in extended size. Cyclooxygenase 2 enzyme is located in the middle of this pathway network. The genes that showed significant differences between one of the diseased groups and healthy controls have color codes in which blue represents RA, green represents IBD and red codes for psoriasis-related genes.

In order to reveal additional information in this network of genes, significantly changing genes were assigned to functional categories based on EASE, the Expression Analysis Systematic Explorer, that provides statistical methods for discovering biological themes within gene lists [85]. Genes were grouped into different functional categories such as inflammatory response; cell growth and maintenance; proteolysis; metabolism and other or unclassified (which included genes in unique categories or genes without categories) with 21, 11, 7, 9 and 7 genes respectively. Illustration of the functional categories, number of genes and disease-specific gene panels led to the conclusion that

there are similarities in the pathogenetic background of these conditions as the number of genes in the same functional categories in each of the diseases were similar, although the number of unique genes (11 in IBD, 6 in psoriasis and 21 in RA) separating only a disease from the control group was also high (Figure 22).

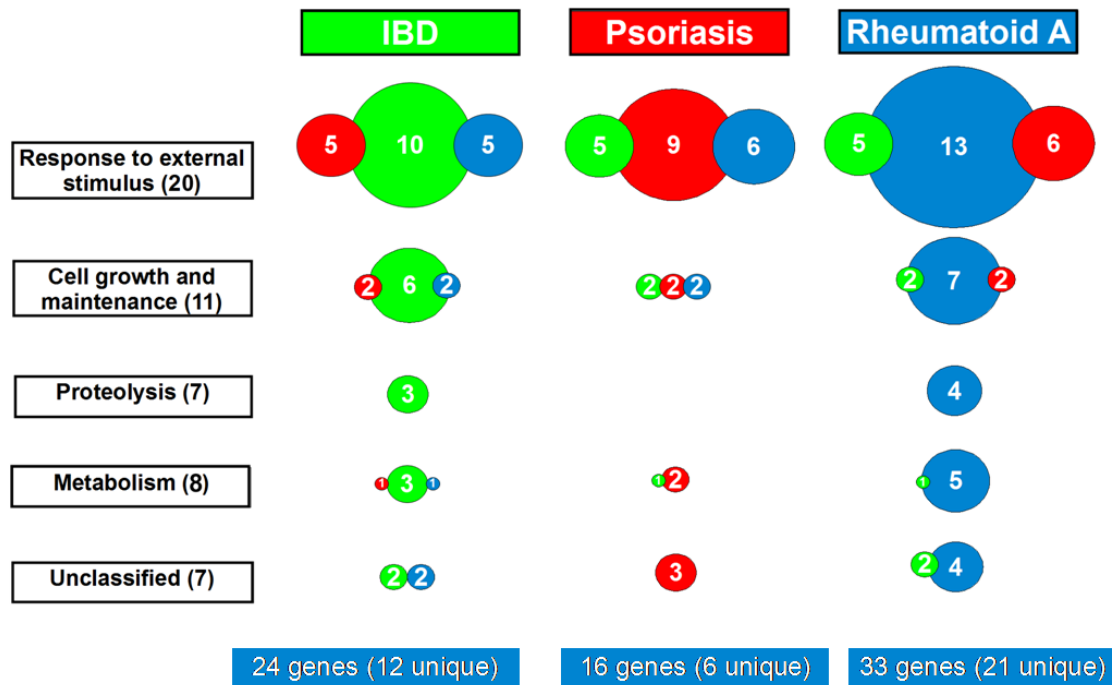


Figure 22: Fifty-three genes that show significantly differences between chronic inflammatory diseases and control samples were grouped into functional categories made by EASE. Genes were grouped in the category that had the highest relevance according to the EASE software. Colors represent disease groups. The size of the diagrams correlates with the number of genes in each set. Overlapping sets of genes are shown as overlapping diagrams.

5.1.3 Differentially expressed genes identify disease subtypes and prognosis

Identifying potential gene expression markers that differentiate between subtypes or states of prognosis in each of the three diseases has clear clinical advantages. ADAM10, CUGBP1, CUGBP2, DUSP1, ETS2, FOS, IL6, IL18 and PGRMC1 showed significantly higher expression levels in patients with CD compared to UC in the IBD cohort (Figure 23a).

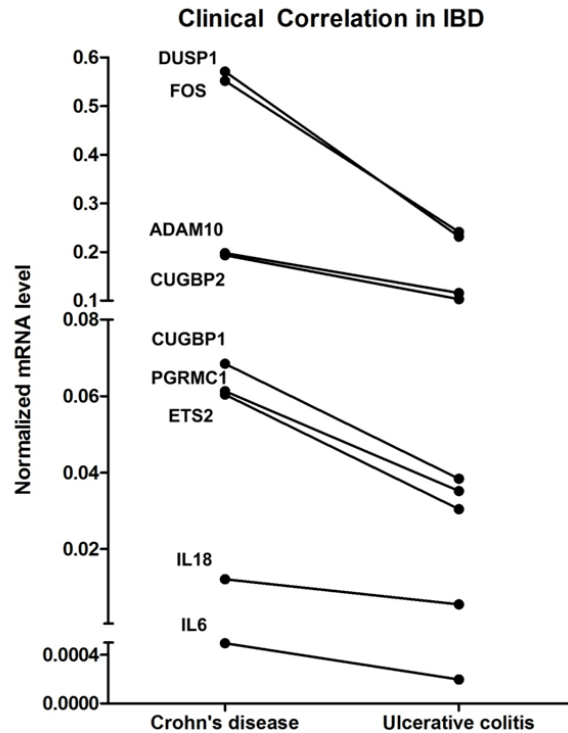


Figure 23a: Normalized mRNA levels of genes showing significant differences between patients with CD and UC were generated from RT-QPCR measurements and represent the means of the expression levels of the diseased and control groups.

CCL4, GZMK and KLRF1 showed significantly higher, while ADM, ANXA3, IL4, MMP9 and OLR1 show significantly lower expression levels in patients with arthritis negative psoriasis compared to patients with symptoms of arthritis (Figure 23b).

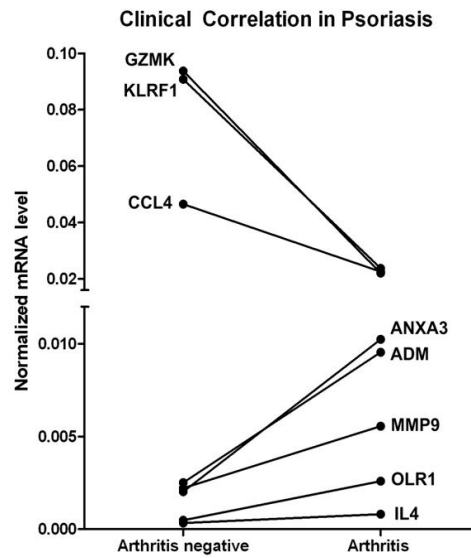


Figure 23b: Normalized mRNA levels of genes showing significant differences between patients with psoriasis patients with and without arthritis were generated from RT-QPCR measurements and represent the means of the expression levels of the diseased and control groups.

HBB and PADI4 showed lower; GZMK showed higher expression in RA patients with bone erosions compared to those without such symptoms (Figure 23c).

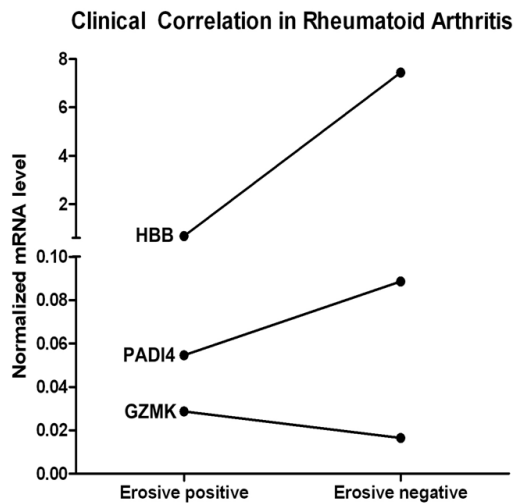


Figure 23c: Normalized mRNA levels of genes showing significant differences between patients with RA patients with and without MR-confirmed bone erosion were generated from RT-QPCR measurements and represent the means of the expression levels of the diseased and control groups.

5.1.4 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 (Figure 24). These genes might be considered universal markers of chronic inflammation in PBMCs.

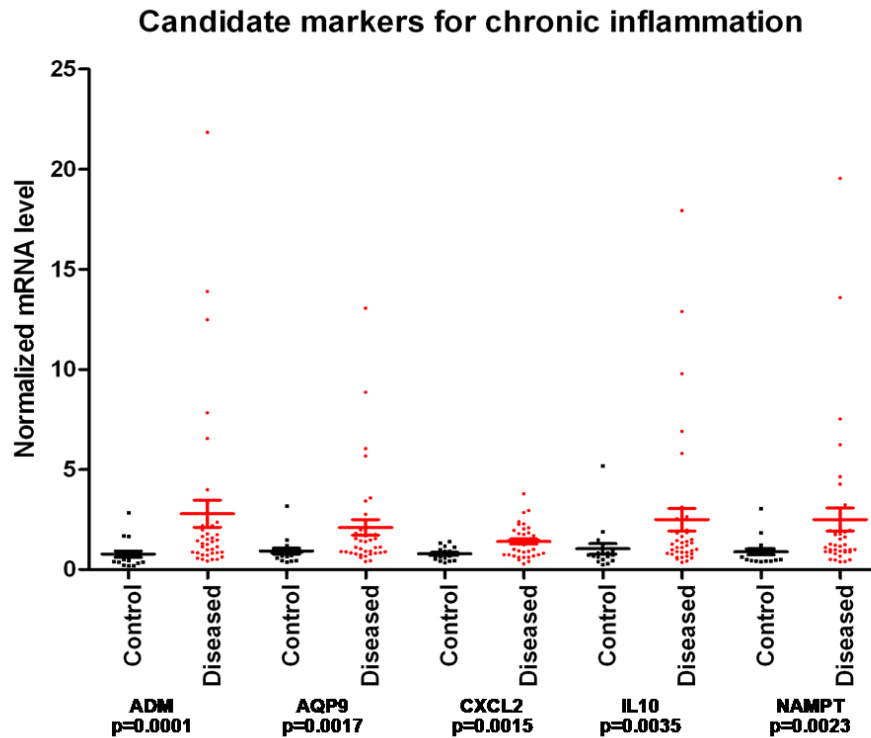


Figure 24: Normalized mRNA levels of genes ($\pm$ SD) showing significant differences between all the samples of chronic inflammatory diseases and healthy controls were generated from RT-QPCR measurements.

Pathway analysis featuring biological processes the genes are associated with identified IL10 as the gene with the highest number of related biological processes including lymphocyte, T and B cell proliferation, interleukin production and macrophage activation. ADM is related to processes such as activation of MAPK-activity and interleukin-6 production; while AQP9, CXCL2 and NAMPT are only associated with urea transport, interleukin-18 production and neutrophil apoptosis, respectively (Figure 25).

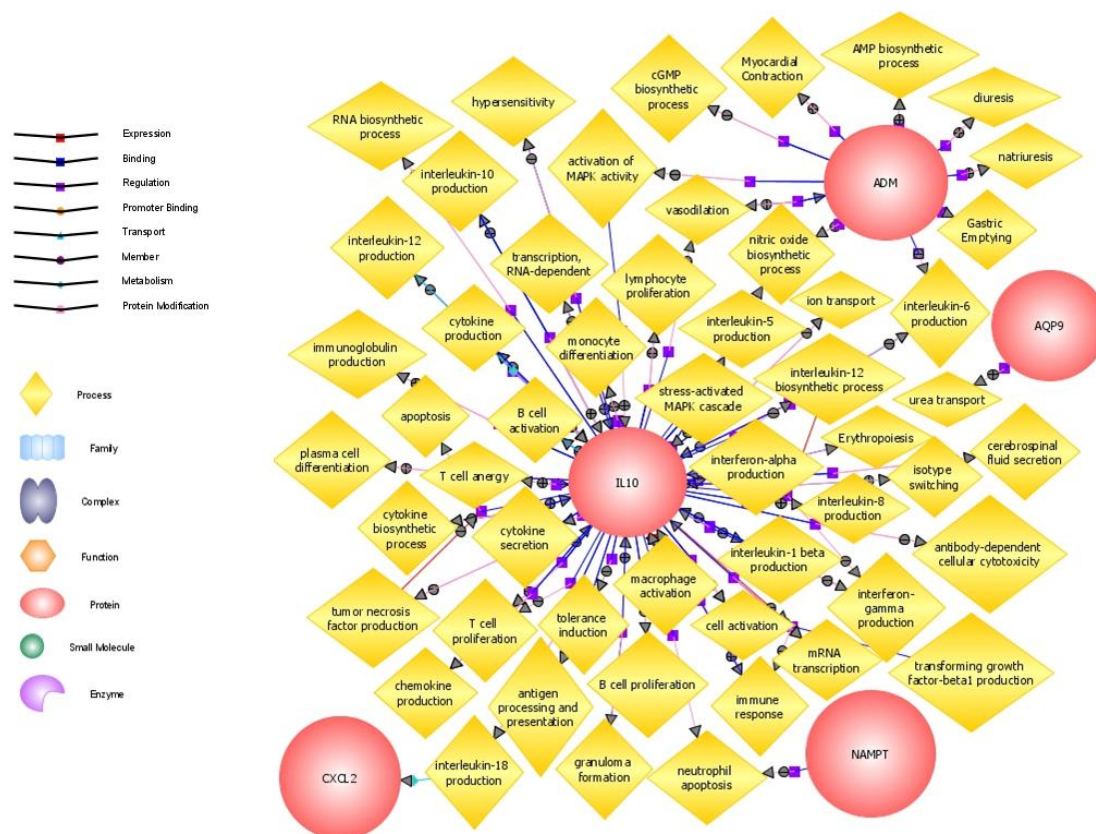


Figure 25: Pathway analysis revealed all the biological processes the genes are related to. Circles represent the 5 genes; deltoids represent biological processes.

5.2 Peripheral blood gene expression profiling of tocilizumab response in RA

5.2.1 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 (Table 6).

Table 6: Patient parameters per groups of responders and non-responders. Brinkmann index is calculated by multiplying the number of cigarettes a day by the years spent smoking.

Patient	Total (mean±SD)	Responders (range)	Non-responders (range)
Response at week 14	n=13	n=9	n=4
ACR Response at week 14	9/4		
Gender	Female/Male	8/1 Female/Male	1/3 Female/Male
Age	47.31±8.81	48.2 (37-60)	45.2 (35-53)
Duration of disease (months)	150.46±130.13	141.3 (24-480)	171 (48-300)
Smoking (Brinkmann index)	58.85±121.01	0	191.2 (0-375)
DAS 28 at baseline	5.42±0.72	5.25 (4.36-6)	5.81 (4.62-7)
DAS 28 at week 4	3.66±1.4	3.68 (2.39-6)	3.62 (1.21-6.21)
DAS 28 at week 14	2.79±1.41	2.27 (0.16-3.82)	3.97 (2.72-5.82)
HAQ at baseline	1.74±0.66	1.83 (1-2.75)	1.53 (0.75-1.875)
HAQ at week 4	1.55±0.65	1.48 (0.5-2.62)	1.71 (1.25-2.25)
HAQ at week 14	1.12±0.66	0.93 (0.12-2)	0.56 (1.25-1.87)
CRP at baseline	27.9±47.17	13.8 (0.4-40.2)	59.6 (2.34-178)
CRP at week 4	19.47±44.67	4.62 (0.5-22.81)	52.89 (0.5-161.81)
CRP at week 14	16.83±37.6	2.89 (0.5-16.7)	48.21 (0.5-129.56)
ESR at baseline	32.08±22.96	28.4 (8-50)	40.2 (10-90)
ESR at week 4	15.38±20.87	9.22 (2-28)	29.25 (2-76)
ESR at week 14	14.46±21.3	5.77 (1-18)	34 (8-80)
Previous DMARDs	3.46±1.27	3.33 (2-5)	3.75 (2-6)
MTX dose (mg per week)	14.04±4.39	14.44 (7.5-20)	13.13 (10-20)
Prednisone use (mg per week)	1.23±2.52	0.89 (0-8)	2 (0-4)

The distribution of the patient population is visualized by using the combination of ACR categories, DAS28 improvement between baseline and week 14, and DAS28 at week 14 (Figure 26a). PCA of clinical parameters did not differentiate between groups of responders and non-responders with a clear cut difference and PCA results were most influenced by two patients (Figure 26b). There were no significant differences in DAS28 values between responders and non-responders neither at baseline, nor at week 4 (Figure 27) suggesting that clinical characteristics of patient groups do not allow clear differentiation regarding the responsiveness to therapy.

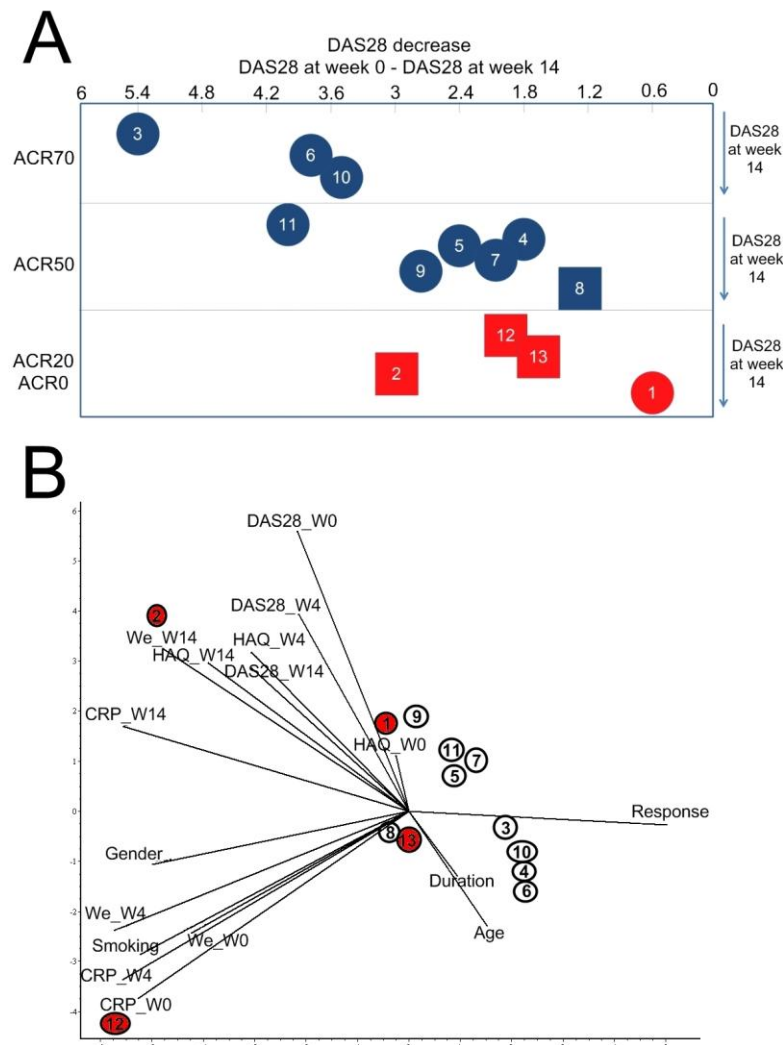


Figure 26a: Blue color represents responders, red represents non-responders, circle shows female, and square shows male patients. Numbers represent individual patient based on assessment included in Table 1. Patients are classified by ACR criteria (from

ACR0-20 to ACR70 from bottom to top); by DAS28 improvement between baseline and week 14 (from 0 to 6 from right to left); and by DAS28 value at week 14 (from the lowest to the top values from top to the bottom in each horizontal section)

Figure 26b: PCA was performed on data from all patients as one group using parameters such as responder status, age, gender, smoking status, duration of disease, DAS28, ESR (erythrocyte sedimentation rate), CRP and HAQ at week 0, 4 and 14. Red circles represent non-responder patients.

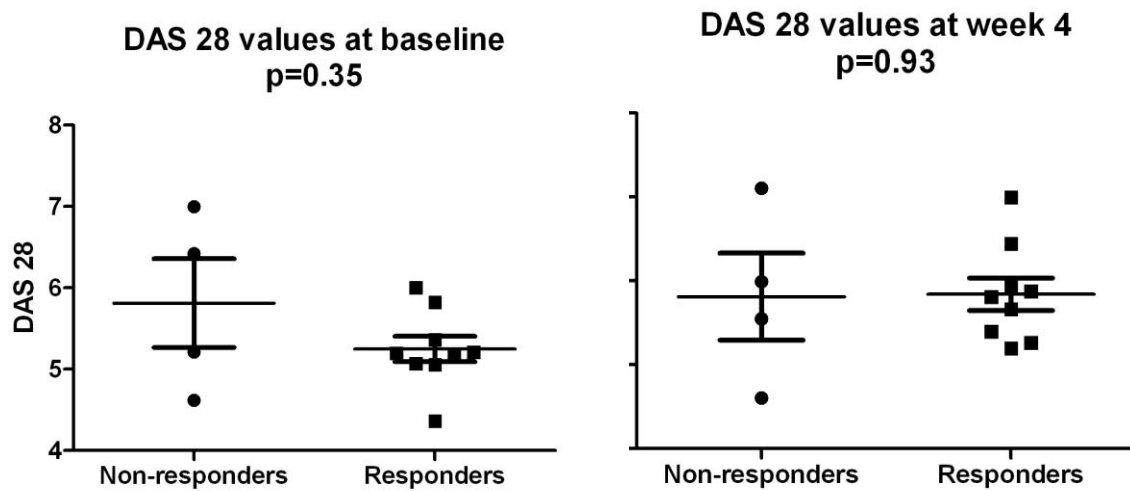


Figure 27: DAS 28 values at baseline and at week 4 regarding the comparison of responders and non-responders.

ELISA analyses of serum did not reveal significant differences in IL-6 or IL-8 serum cytokine levels (Figure 28), and data were not used for further analyses (cytokine levels of IL-1b were below measurement limit).

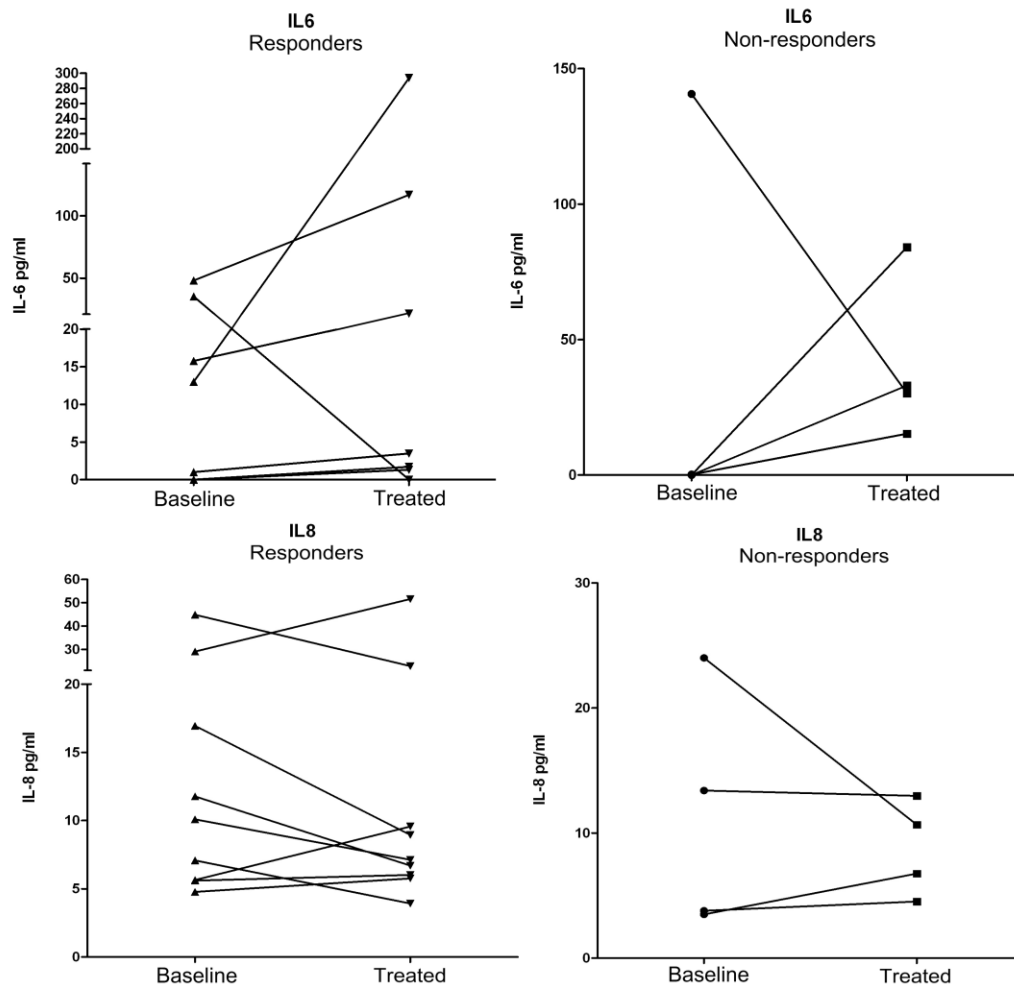


Figure 28: Serum cytokine levels of IL-6, IL-1b and IL-8, samples were obtained from all patients at baseline and week 4. Cytokine levels of IL-6 and IL-8 did not show significant changes neither in the comparison of responders and non-responder, nor in the comparison of samples at baseline and week 4.

5.2.2 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 (Figure 29a). 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) (Figure 29b) (Figure 30).

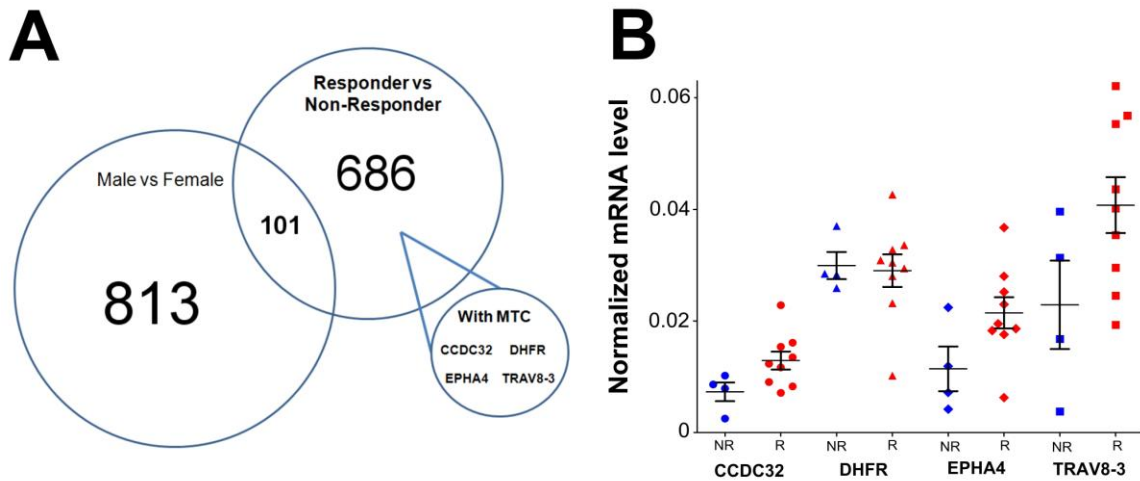


Figure 29

A: Venn diagram of gene expression differences based on the microarray experiments between responders and non-responders. Gender specific gene expression differences have been subtracted leading to a list of 686 probe sets differentiating between responder statuses without differences caused by gender. Changes in 4 genes: CCDC32, DHFR, EPHA4 and TRAV8-3 remained statistically significant after correction for multiple testing (Benjamini-Hochberg FDR).

B: Normalized mRNA levels of genes CCDC32, DHFR, EPHA4 and TRAV8-3 ($\pm$ SD) showing significant differences between Responders and Non-Responders. Normalized mRNA levels of the indicated genes for each patient are shown.

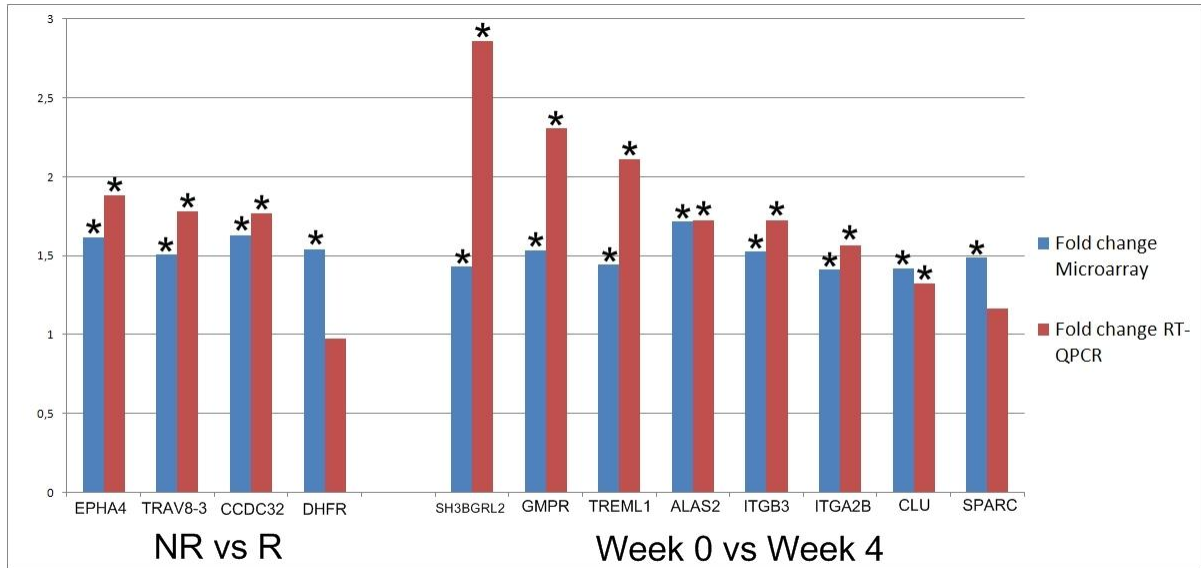


Figure 30: Fold change between responders and non-responders (first 4 sets of bars); baseline and week 4 (second 8 sets of bars) by microarray experiments (blue bars) and RT-QPCR (red bars). Asterisk demonstrates whether the comparison was statistically significant (only gene expression changes regarding DHFR and SPARC by RT-QPCR were not significant).

5.2.3 Gene expression patterns correlate with clinical parameters

Comparison of gene expression data with clinical parameters such as DAS28 yielded a list of nine genes including ALAS2, DYRK3, EPB42, PTX3, RGL1, RUNDC3A, SLC25A39, TMOD1 and TMEM56, which showed significant correlation with DAS28 scores at baseline ($r > 0.74$, $p \leq 0.0002$). Significant correlation was found in other comparisons as well. (Table 7)

Table 7: Pearson correlations (r) and p values of genes used in correlation analyses. Red numbers represent significant correlation.

DAS28 and gene expression				
	Baseline		Week 4	
	p value	r coefficient	p value	r coefficient
ALAS2	0.0001	0.78	0.42	0.05
DDX4	0.0001	0.75	0.29	0.09
DYRK3	0.0001	0.75	0.84	0.003
EPB42	0.0002	0.74	0.59	0.02
GMPR	0.001	0.64	0.25	0.11
ITGB3	0.35	0.07	0.04	0.32
PTX3	0.0001	0.75	0.87	0.002
RGL1	0.0001	0.77	0.05	0.29
RUNDC3A	0.0001	0.74	0.61	0.02
SLC25A39	0.0001	0.76	0.47	0.04
TMEM56	0.0002	0.74	0.08	0.24
TMOD1	0.0001	0.75	0.28	0.1

DAS28 (week 14) and gene expression (baseline)		
	p value	r coefficient
MAP7D1	0.0003	0.7
PDAP1	0.0002	0.73
ZIC3	0.0001	0.8

5.2.4 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. Columns 1-3 showed a remarkable discriminatory power, while gene lists compiled by other authors such as PBMC studies in infliximab treatment for RA and genes related to synovium showed moderate separation, and lists of randomly selected genes served as negative controls (Figure 31, Table 8).

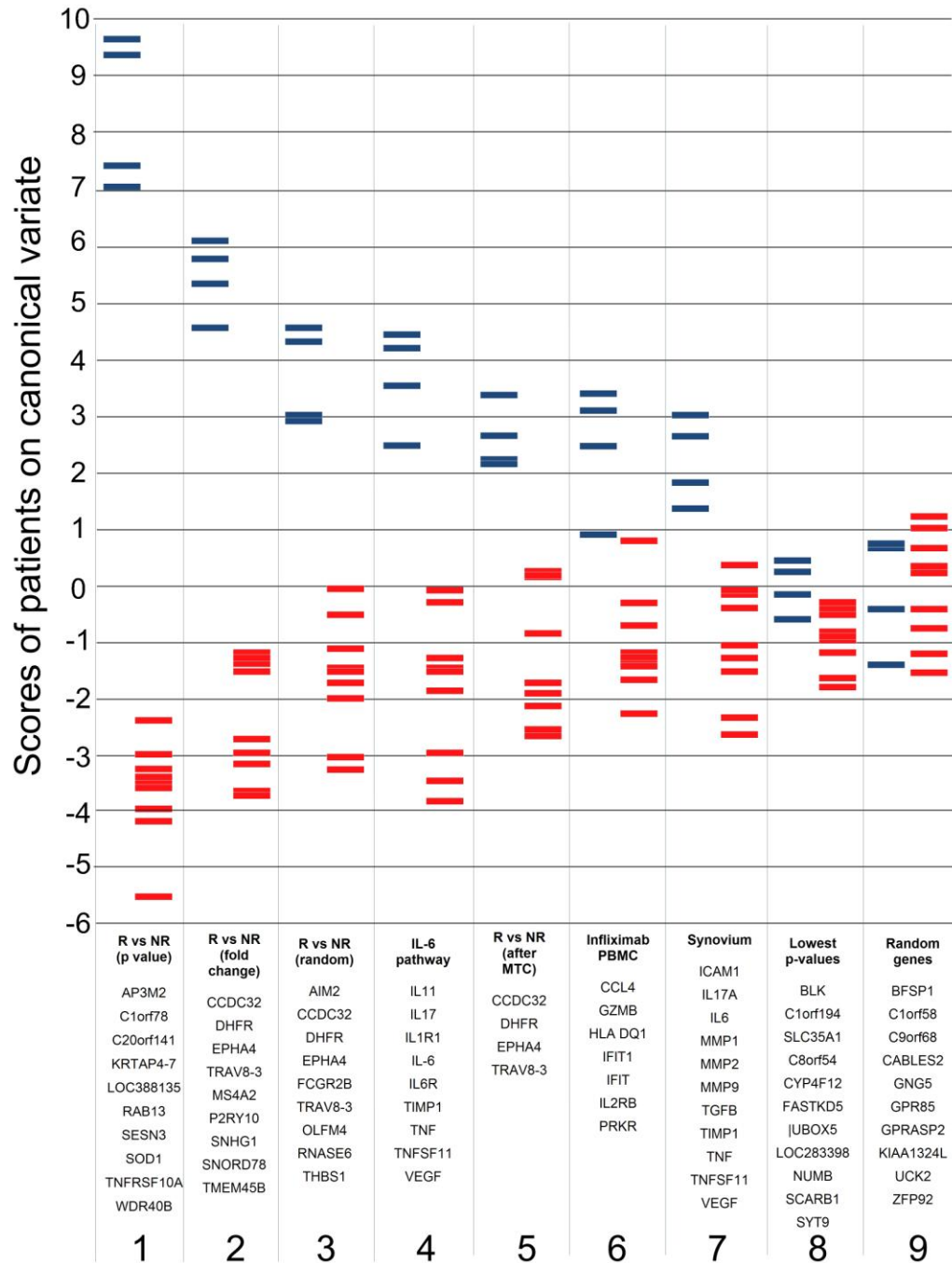


Figure 31: 9 gene lists evaluated by CVA. The power of canonical correlation decreases from left to right. Red bars represent responders, blue bars represent non-responders. The bigger the space between the groups and the less the overlap is among samples, the higher the power of separation of the gene list is.

Table 8: Canonical correlation, CHISQ and WILKS lambda values of gene lists used for CVA (Figure 3). The higher the Canonical Correlation and the CHISQ are as well as the lower the WILKS lambda is, the stronger the gene list is.

List	Canonical		
	Correlation	CHISQ	WILKS lambda
R vs. NR (p value)	0.99	21.83	0.02
R vs. NR (fold change)	0.97	18.33	0.05
R vs. NR (random)	0.94	13.74	0.12
IL-6 pathway	0.94	13.68	0.12
R vs. NR (after MTC)	0.88	13.70	0.22
Infliximab PBMC	0.88	11.14	0.23
Synovium	0.86	7.22	0.27
Lowest p-values	0.193	0.23	0.96
Random genes	0.014	0	0.99

5.2.5 The degree of galactosylation of IgG N-glycans increases after treatment and in responders versus non-responders

Based on the significant differences we found in the expression levels of genes such as B4GALT1, FUT8 and MGAT3 related to glycosylation of IgG, the degree of galactosylation of the N-glycans of IgG was measured. We analyzed and compared the glycosylation patterns in samples in order to reveal a potential correlation between IgG1 N-glycan profile and the pathogenesis of RA.

The percentage of the area under the curve of IgG G0 seems to be lower in responders compared to non-responders at baseline (Figure 32A), as well as after treatment (Figure 32B). IgG G0 area% (Figure 32C) showed a significant decrease after treatment compared to the baseline values in the comparison of all baseline samples and the treated ones. These data suggest that the degree of galactosylation increases in responders.

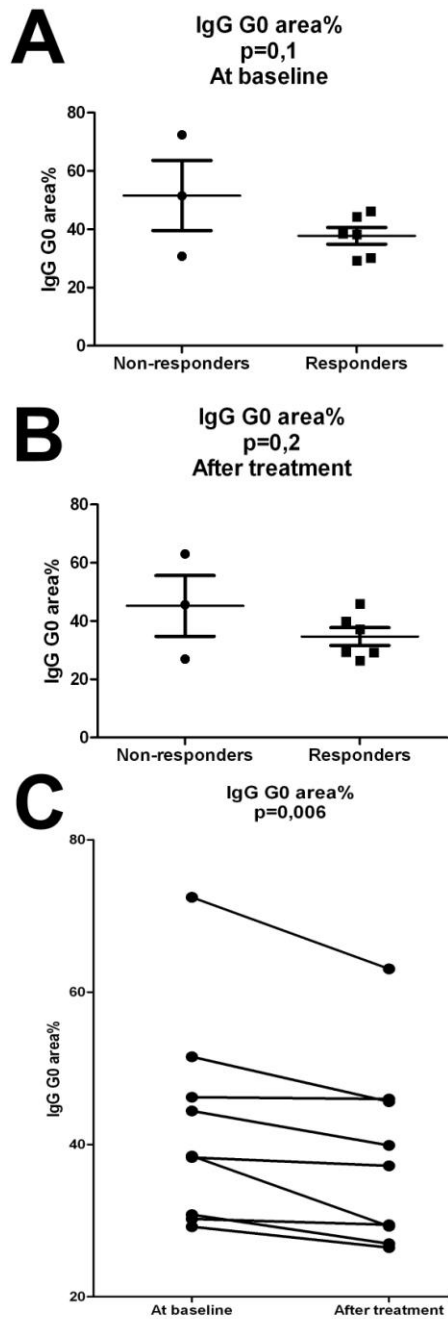


Figure 32

A: Percentage of the area under the curve of IgG G0 at baseline shows a slight decrease in responders compared to non-responders. B: Percentage of the area under the curve of IgG G0 after treatment at week 4 shows a slight decrease in responders compared to non-responders. C: Percentage of the area under the curve of IgG G0 at baseline shows a significant decrease in samples obtained after treatment at week 4 compared to those

obtained at baseline. Matching samples from the same patients are connected to each other between baseline and after treatment.

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

5.3.1 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 (Table 9a).

In RA, a binary outcome variable was used for assessing clinical responder status: patients with ACR0 or ACR20 scores were classified as “non-responders”; and patients with ACR50 or ACR70 scores were classified as “responders”. Thus, 6 responders and 13 non- or moderate responders in the test cohort; and 4 responders and 11 non or moderate responders in the validation cohort were identified by clinicians; and there were no significant differences regarding age, DAS28, HAQ, CRP, Rheumatoid factor, anti-CCP status or DMARDs between the non-responder and responder groups (Mann Whitney U test) (Table 9b).

Table 9a: Summary of the clinical parameters of CD patients. Data are shown as mean $\pm$ SD. NS means non significant.

	Non- Responders		Non- responders		
	Test cohort		Validation cohort		Difference
At baseline	14	6	13	7	
Gender					
(Male/Female)	8/6	4/3	8/5	4/4	
Age	36,2 $\pm$ 14,6	36 $\pm$ 15,4	26,8 $\pm$ 8,7	30,2 $\pm$ 5,1	ns*
CDAI	319,6 $\pm$ 41,3	351,5 $\pm$ 33,1	338,3 $\pm$ 61,7	370,7 $\pm$ 48,12	ns

CRP	22,78±20,2	13,59±28,9	27,2±36,2	16,5±11,6	ns
Hemoglobin	125,1±17,6	120,6±27,6	130±17,9	127±12	ns
Leukocytes	9±3,3	8±3,1	7,58±2,38	9,24±2,8	ns
Neutrophils	70±8,9	74,5±8,9	74±8,1	72,83±5,7	ns

Table 9b: Summary of the clinical parameters of RA patients. Data are shown as mean±SD.

	Responders	Non- or moderate responders	Responders	Non- or moderate responders	
	Test cohort		Validation cohort		Difference
At baseline	6	13	4	11	
Gender					
(Male/Female)	1/5	2/11	0/4	3/8	
Age	44,33±9,3	47,08±10,7	54±17	56,2±6,8	ns
DAS28	5,63±0,3	5,26±0,7	5,23±0,06	5,44±0,5	ns
HAQ	1,27±0,7	2,06±0,6	1,59±1,1	1,93±0,7	ns
CRP	16,83±18,3	28,31±23,8	18,9±18,3	9,58±10,3	ns
DMARDs	2,83±0,9	2,69±0,7	3±0,8	2,7±1,4	ns

5.3.2 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 (Table 10a).

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 (Table 10b).

Table 10a: Gene symbol, gene description, whether the gene is associated with CD based on the literature, the time point at which it showed significant differences between responders and non-responders in the test cohort, and whether it showed significant differences between responders and non-responders in the validation cohort are shown. Genes marked with green are included in the gene panel with the best discriminatory power. Literature represents PMID numbers, $p < 0,05$ represents whether that gene was significantly changing.

Gene symbol	Description	Literature	Test cohort		Validation cohort
			Timepoint	$p < 0,05$	$p < 0,05$
ABCC4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	20393862	Baseline + week 2	Yes	No
AIDA	axin interactor, dorsalization associated		Baseline	Yes	No
ARHGEF12	Rho guanine nucleotide exchange factor (GEF) 12		Baseline	Yes	No
BMP6	bone morphogenetic protein 6	21351217	Baseline + week 2	Yes	No
BTN3A2	butyrophilin, subfamily 3, member A2		Baseline	Yes	No
CA2	carbonic anhydrase II		Week 2	Yes	No
CADM2	cell adhesion molecule 2		Week 2	Yes	No
CD300E	CD300 antigen like family member E		Baseline	Yes	No
CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1	20444268	Baseline	Yes	No
ENDOD1	endonuclease domain containing 1		Baseline	Yes	No
FCGR1A	Fc fragment of IgG, high affinity Ia, receptor (CD64)	20016944	Baseline	Yes	No
FMN1	formin 1		Baseline	Yes	No
GCLC	glutamate-cysteine ligase, catalytic subunit		Baseline	Yes	No

GPR34	G protein-coupled receptor 34		Week 2	Yes	No
HORMAD1	HORMA domain containing 1		Baseline	Yes	No
IGF2BP2	insulin-like growth factor 2 mRNA binding protein 2		Baseline	Yes	No
IL18R1	interleukin 18 receptor 1	19249008	Baseline	Yes	No
IL1RL1	interleukin 1 receptor-like 1		Week 2	Yes	No
KAT2B	K(lysine) acetyltransferase 2B		Baseline	Yes	No
MAP1LC3B	microtubule-associated protein 1 light chain 3 beta		Baseline	Yes	No
MMD	monocyte to macrophage differentiation-associated		Week 2	Yes	No
MS4A4A	membrane-spanning 4-domains, subfamily A, member 4		Baseline	Yes	No
MS4A7	membrane-spanning 4-domains, subfamily A, member 7		Baseline	Yes	No
ODC1	ornithine decarboxylase 1		Baseline	Yes	No
PBX1	pre-B-cell leukemia transcription factor 1		Baseline	Yes	No
PCYT1B	phosphate cytidylyltransferase 1, choline, beta isoform		Baseline	Yes	No
PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha		Baseline	Yes	No
PIP5K1B	phosphatidylinositol-4-phosphate 5-kinase, type I, beta		Baseline	Yes	No
PRDM1	PR domain containing 1, with ZNF domain	19760754	Week 2	Yes	No
PSME4	proteasome (prosome, macropain) activator subunit 4		Baseline	Yes	No
RAD23A	RAD23 homolog A		Week 2	Yes	No
RIOK3	RIO kinase 3 (yeast)		Baseline	Yes	No
RNASE2	ribonuclease, RNase A family, 2 (liver, eosinophil-derived neurotoxin)	17453718	Baseline	Yes	No
RNF11	ring finger protein 11		Baseline	Yes	No
SLC7A5	solute carrier family 7 member 5		Week 2	Yes	No
THEM5	thioesterase superfamily member 5		Baseline + week 2	Yes	No
TMEM176A	transmembrane protein 176A		Baseline	Yes	Yes
TMEM176B	transmembrane protein 176B		Baseline	Yes	Yes
UBE2H	ubiquitin-conjugating enzyme E2H (UBC8 homolog, yeast)		Baseline	Yes	Yes
WARS	tryptophanyl-tRNA synthetase		Baseline	Yes	Yes

Table 10b: Gene symbol, gene description, whether the gene is associated with RA based on the literature, the time point at which it showed significant differences between responders and non-responders in the test cohort or studies from the literature based on which genes were included in the analysis, and whether it showed significant differences between responders and non-responders in the validation cohort are shown. Genes marked with green are included in the gene panel with the best discriminatory power.

			Test cohort		Validation cohort
Gene symbol	Description	Literature	Timepoint	p<0,05	p<0,05
APOBEC3A	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A		Baseline	Yes	No
AQP9	aquaporin 9	20444268	Baseline vs Week 2	Yes	No
CCL4	chemokine (C-C motif) ligand 4		Sekiguchi et al	No	No
CNTNAP3	contactin associated protein-like 3		Baseline	Yes	No
CYP4F3	cytochrome P450, family 4, subfamily F, polypeptide 3		Baseline	Yes	Yes
DHRS9	dehydrogenase/reductase (SDR family) member 9		Baseline	Yes	Yes
EIF2AK2	eukaryotic translation initiation factor 2-alpha kinase 2		Sekiguchi et al	No	No
ELOVL7	ELOVL fatty acid elongase 7		Week 2	Yes	No
EPSTI1	epithelial stromal interaction 1 (breast)		Baseline + week 2	Yes	No
FCGR3A	Fc fragment of IgG, low affinity IIIa, receptor (CD16a)		Week 2	Yes	No
GPAM	glycerol-3-phosphate acyltransferase, mitochondrial		Week 2	Yes	No
GPR15	G protein-coupled receptor 15		Baseline	Yes	No
GZMB	granzyme B		Sekiguchi et al	No	No
IFI35	interferon-induced protein 35		van Baarsen et al	No	No
IFI44	interferon-induced protein 44	20096109	Baseline + week 2	Yes	No
IFI44L	interferon-induced protein 44-like	20096109	Baseline	Yes	No
IFI6	interferon, alpha-inducible protein 6		van Baarsen et al	No	No

IFIT1	interferon-induced protein with tetratricopeptide repeats 1	20096109	Baseline + week 2	Yes	No
IFIT2	interferon-induced protein with tetratricopeptide repeats 2	20096109	Baseline + week 2	Yes	No
IFIT3	interferon-induced protein with tetratricopeptide repeats 4		Baseline + week 2	Yes	No
IFITM1	interferon induced transmembrane protein 1 (9-27)		van Baarsen et al	No	No
IL2RB	interleukin 2 receptor, beta		Sekiguchi et al	No	No
IRF2	interferon regulatory factor 2	20096109	van Baarsen et al	No	No
IRF7	interferon regulatory factor 7		Sekiguchi et al	No	No
MGAM	maltase-glucoamylase (alpha-glucosidase)		Baseline	Yes	Yes
MICA	MHC class I polypeptide-related sequence A		Week 2	Yes	No
MME	membrane metallo-endopeptidase		Baseline	Yes	No
MX1	myxovirus resistance 1, interferon-inducible protein p78	16736505	Baseline	Yes	No
OR2A9P	olfactory receptor, family 2, subfamily A, member 9 pseudogene		Baseline	Yes	No
PF4	platelet factor 4		Week 2	Yes	Yes
PTGS2	prostaglandin-endoperoxide synthase 2	20444268	Baseline	Yes	No
RAVER2	ribonucleoprotein, PTB-binding 2		Baseline	Yes	No
RFC1	replication factor C (activator 1) 1, 145kDa	18607581	Baseline + week 2	Yes	No
RGS1	regulator of G-protein signaling 1		Week 2	Yes	No
RSAD2	radical S-adenosyl methionine domain containing 2	20096109	Baseline + week 2	Yes	No
S100P	S100 calcium binding protein P		Baseline	Yes	No
SERPINB10	serine (or cysteine) proteinase inhibitor, clade B, member 10		Baseline	Yes	No

SERPING1	serine (or cysteine) proteinase inhibitor, clade G, member 1,	20096109	Baseline	Yes	No
SIGLEC1	sialic acid binding Ig-like lectin 1, sialoadhesin		Baseline	Yes	No
TNF	tumor necrosis factor			No	No
TNFAIP6	tumor necrosis factor, alpha-induced protein 6	20096109	Baseline vs Week 2	Yes	No

Comparing baseline and week 2 samples resulted in a list of 5 genes including AQP9, IGKC, MGAM, MMP8 and TNFAIP6 in CD (Figure 33a); and a list of 3 genes such as AQP9, IGJ and TNFAIP6 with statistically significant differences with Benjamini-Hochberg correction for multiple testing in RA. These genes represent the effects of the therapy over time (Figure 33b).

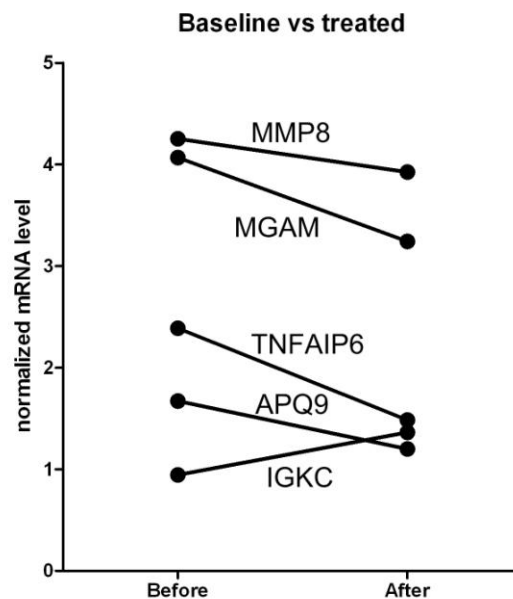


Figure 33a: Genes showing a statistically significant change between baseline and week 2 microarray data in CD: Dots account for the mean of the normalized mRNA levels of a certain gene in the patient groups.

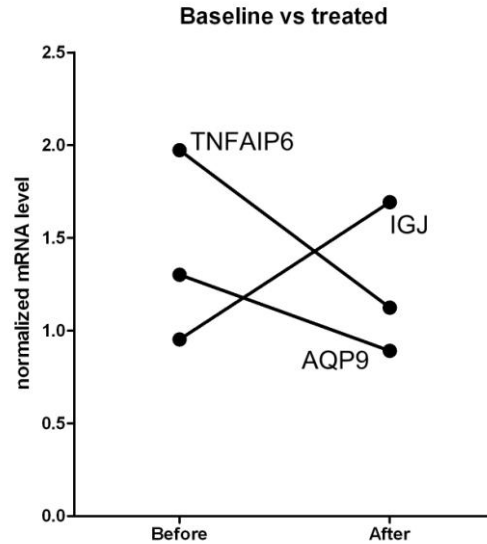


Figure 33b: Genes showing a statistically significant change between baseline and week 2 microarray data in CD (A) and in RA (B). Dots account for the mean of the normalized mRNA levels of a certain gene in the patient groups.

5.3.3 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 (Figure 34a-b).

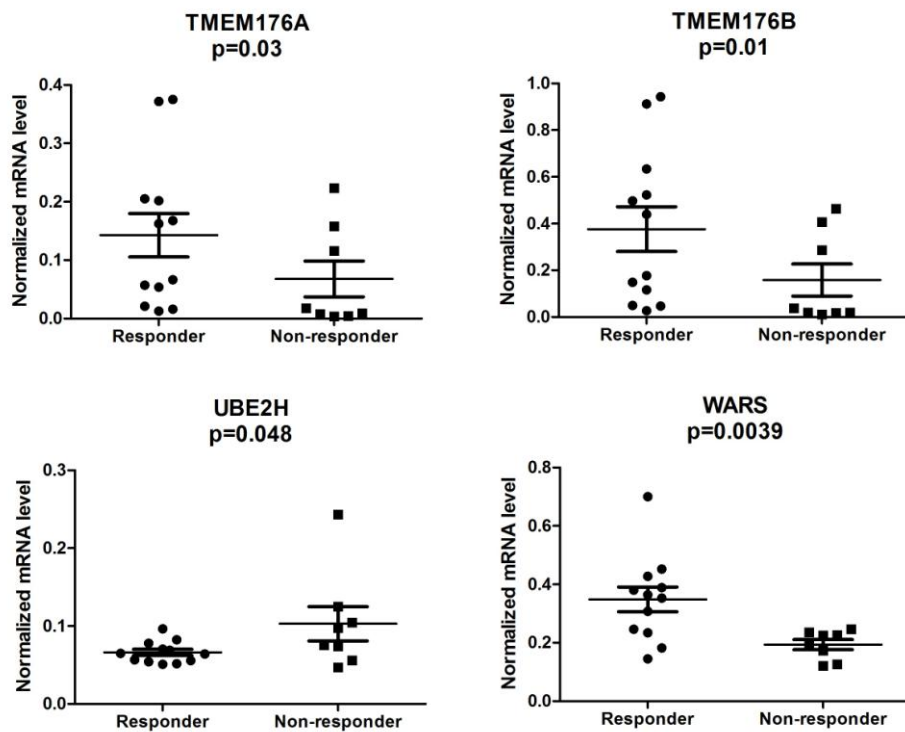


Figure 34a: Scatter plots of normalized mRNA levels in genes showing statistically significant differences between responders and non-responders in the validation cohort in CD.

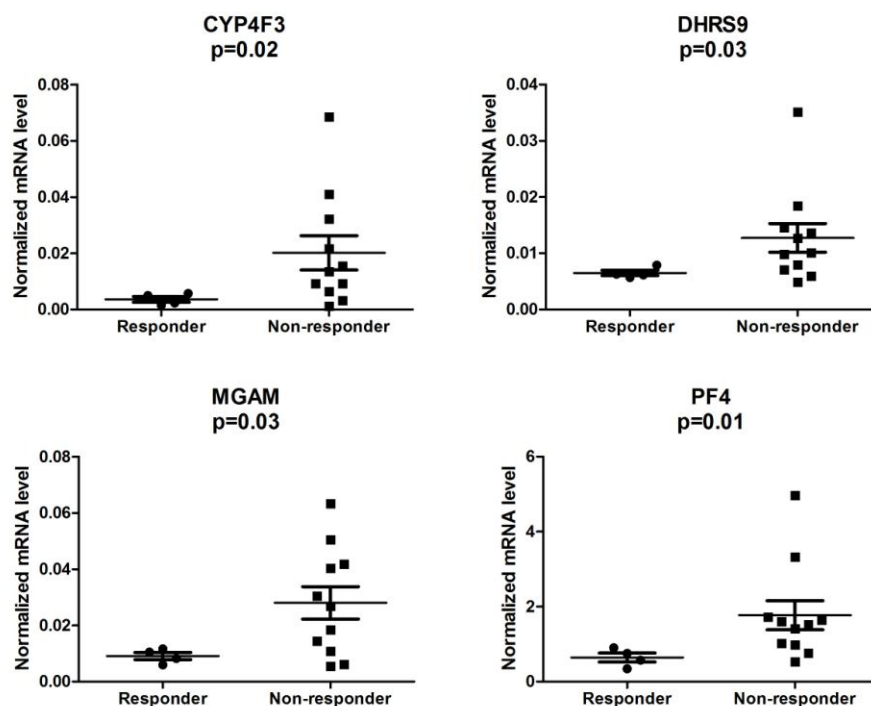


Figure 34b: Scatter plots of normalized mRNA levels in genes showing statistically significant differences between responders and non-responders in the validation cohort in .

5.3.4 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.

In CD, the gene panel with the best discriminatory power included ABCC4, BMP6, CADM2, CD300E, CYP1B1, FMN1, IGF2BP2, MAP1LC3B, MS4A4A, PBX1, PIP4K2A, PIP5K1B, PRDM1, PSME4, RAD23A, SLC7A5, THEM5 and WARS. (Figure 35a, Table 11)

The gene panel with the best discriminatory power in RA contained genes such as CCL4, DHRS9, ELOVL7, IFI35, IFI44, IFI6, IFIT1, IFIT3, IRF7, MICA, OR2A9P, PF4 and RAVR2. (Figure 35b, Table 11)

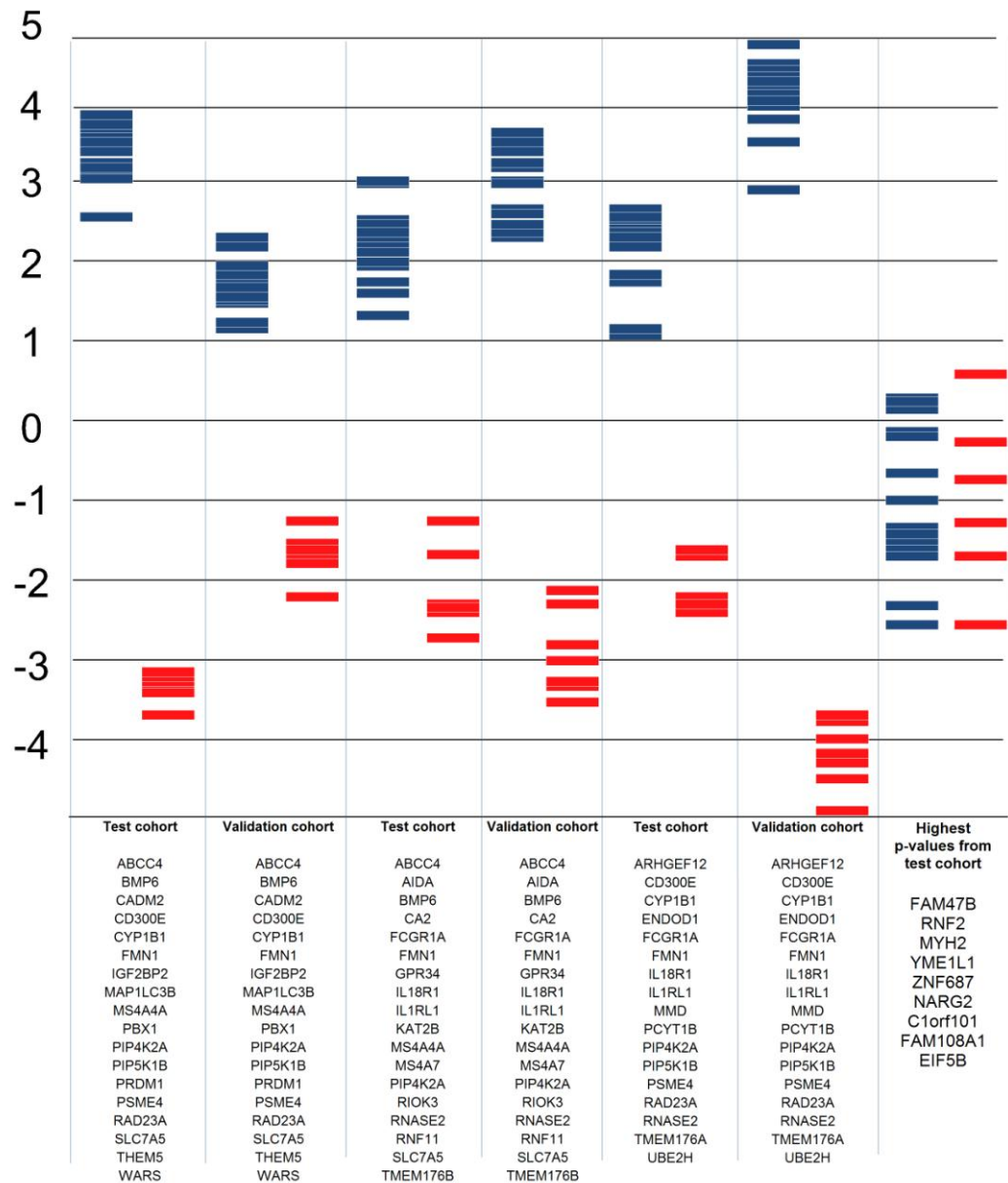


Figure 35a: 3 gene lists evaluated by LDA in CD. Red bars represent non-responders, blue bars represent responders. The bigger the space between the groups and the less the overlap is among samples, the higher the power of separation of the gene list is. Percentages on the top of the lists show the rate of segregation with a focus on identifying the highest number of non-responders. The gene panel from the microarray experiment (test cohort) is on the left, the gene panel from the RT-QPCR experiments

(validation cohort) is in the middle, and a list of random genes from the genome serving as a negative control based on the microarray data is on the right.

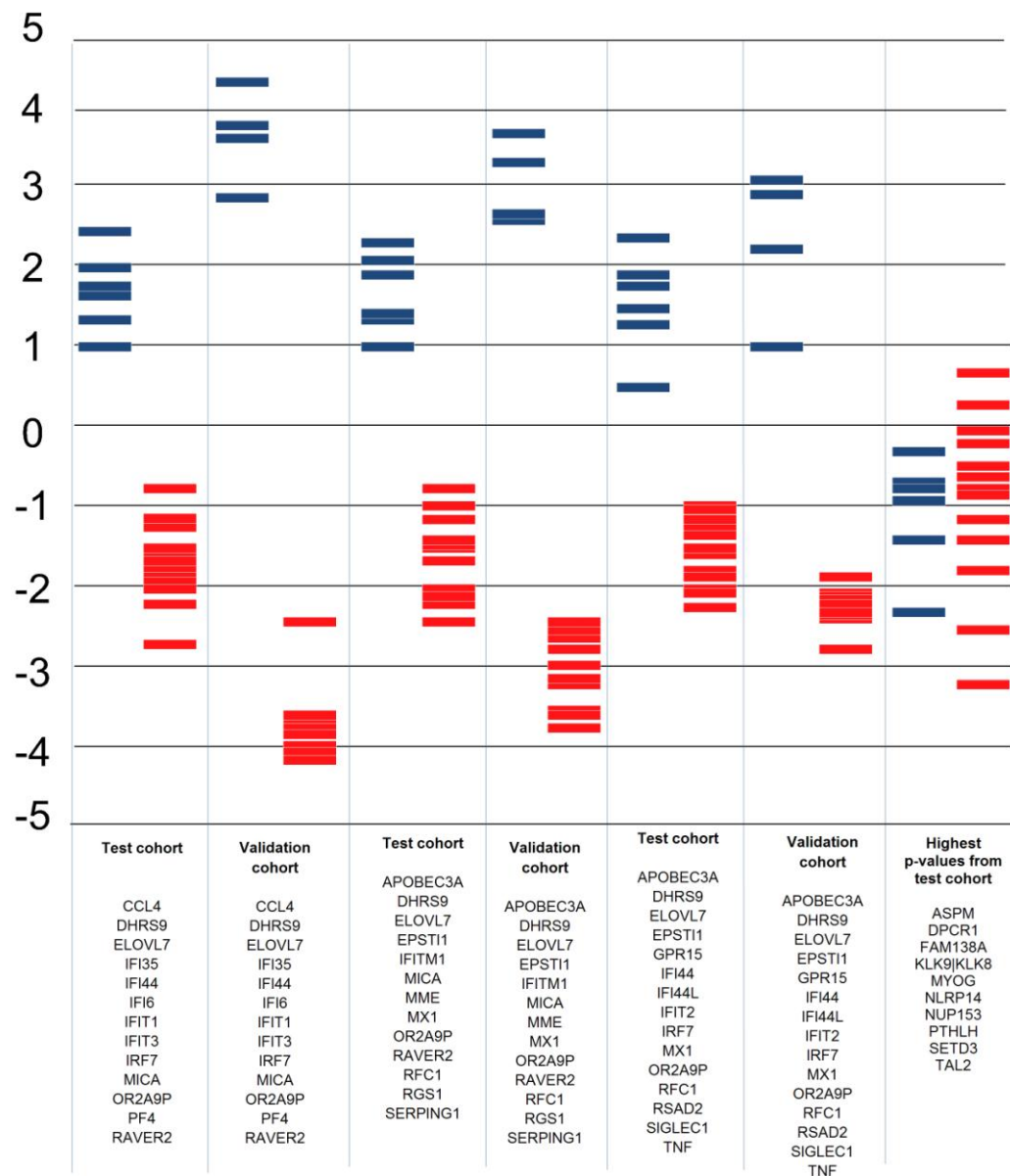


Figure 35b: 3 gene lists evaluated by LDA in RA. Red bars represent non-responders, blue bars represent responders. The bigger the space between the groups and the less the overlap is among samples, the higher the power of separation of the gene list is. Percentages on the top of the lists show the rate of segregation with a focus on

identifying the highest number of non-responders. The gene panel from the microarray experiment (test cohort) is on the left, the gene panel from the RT-QPCR experiments (validation cohort) is in the middle, and a list of random genes from the genome serving as a negative control based on the microarray data is on the right.

Table 11b: Accuracy indicators of the gene panels selected for visualization. Segregation and p value data are presented of each selected gene panel and the negative control.

		Segregation of the groups	p value	Source of data
Test cohort	CD gene panel 1	100%	3,78E-14	Microarray
	RA gene panel 1	100%	9,69769E-11	Microarray
Validation cohort	CD gene panel 1	100%	3,78E-14	RT-QPCR
	RA gene panel 1	100%	9,69769E-11	RT-QPCR
Test cohort	CD gene panel 2	100%	6,33E-13	Microarray
	RA gene panel 2	100%	1,99305E-10	Microarray
Validation cohort	CD gene panel 2	100%	6,33E-13	RT-QPCR
	RA gene panel 2	100%	1,99305E-10	RT-QPCR
Test cohort	CD gene panel 3	100%	6,72E-13	Microarray
	RA gene panel 3	100%	8,6969E-10	Microarray
Validation cohort	CD gene panel 3	100%	6,72E-13	RT-QPCR
	RA gene panel 3	100%	8,6969E-10	RT-QPCR
Negative control	CD	55%	-	Microarray
	RA	53%	-	Microarray

The number of genes within each panel showing a segregation of 100% between responders and non-responders in both cohorts and conditions are visualized sorted by the

p value in order to reveal the estimated number of gene panels with the highest discriminatory power (Figure 36a-b).

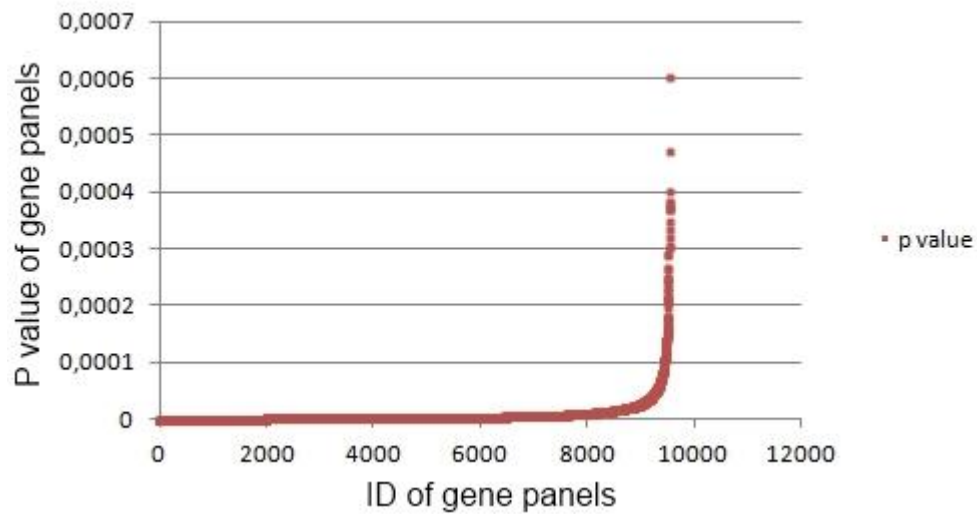


Figure 36a: 9536 gene panels are shown in CDon the X axis. P values showing the possibility of producing gene panels with such accuracy by chance are shown on the y axis.

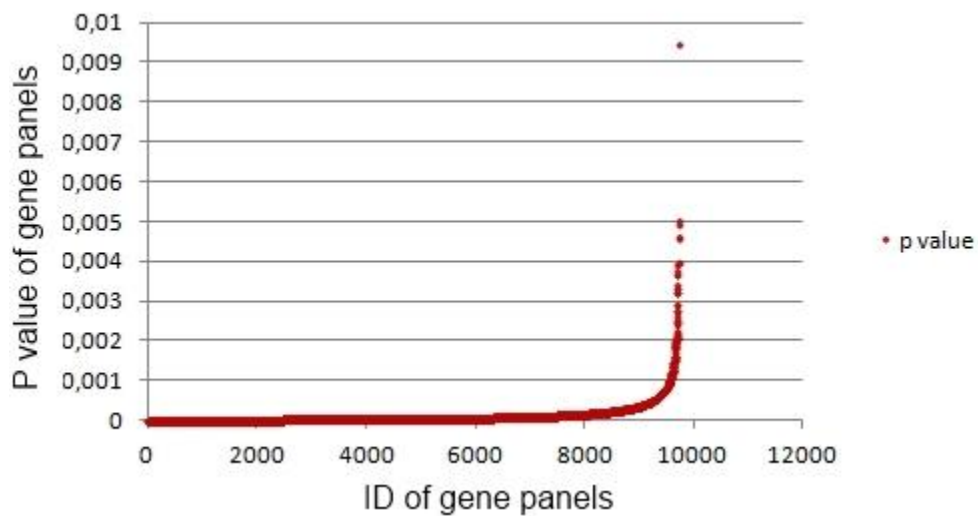


Figure 36b: 9657 gene panels are shown in R on the X axis. P values showing the possibility of producing gene panels with such accuracy by chance are shown on the y axis.

5.3.5 Determination of serum cytokine levels

Serum levels of IL6, IL8, IFN γ and IL12 were measured by ELISA in which several comparisons were found to be statistically significant such as IL12 regarding the responder status; IL6, IL8 and IL12 regarding baseline vs week 2 comparison in RA; IFN γ regarding the responder status; and IL6 regarding baseline vs week 2 comparison in CD (Figure 37a-c).

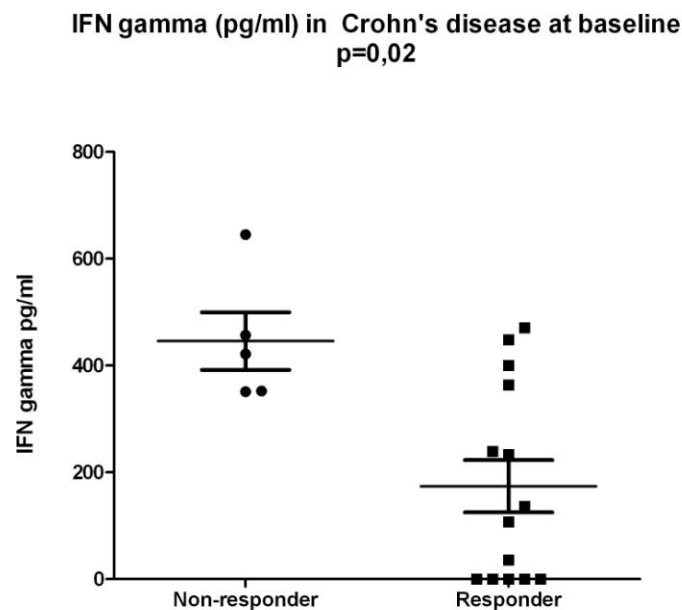


Figure 37a: ELISA results in IFN γ in CD at baseline regarding the responder vs non-responder comparison.

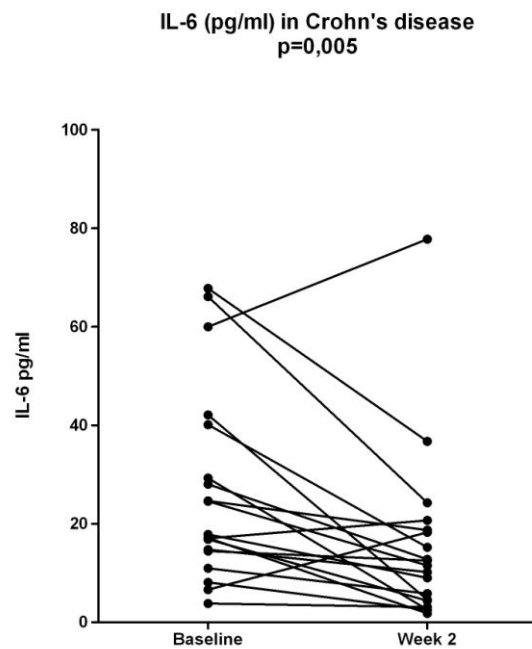


Figure 37b: ELISA results in IL-6 in CD at baseline regarding the responder vs non-responder comparison.

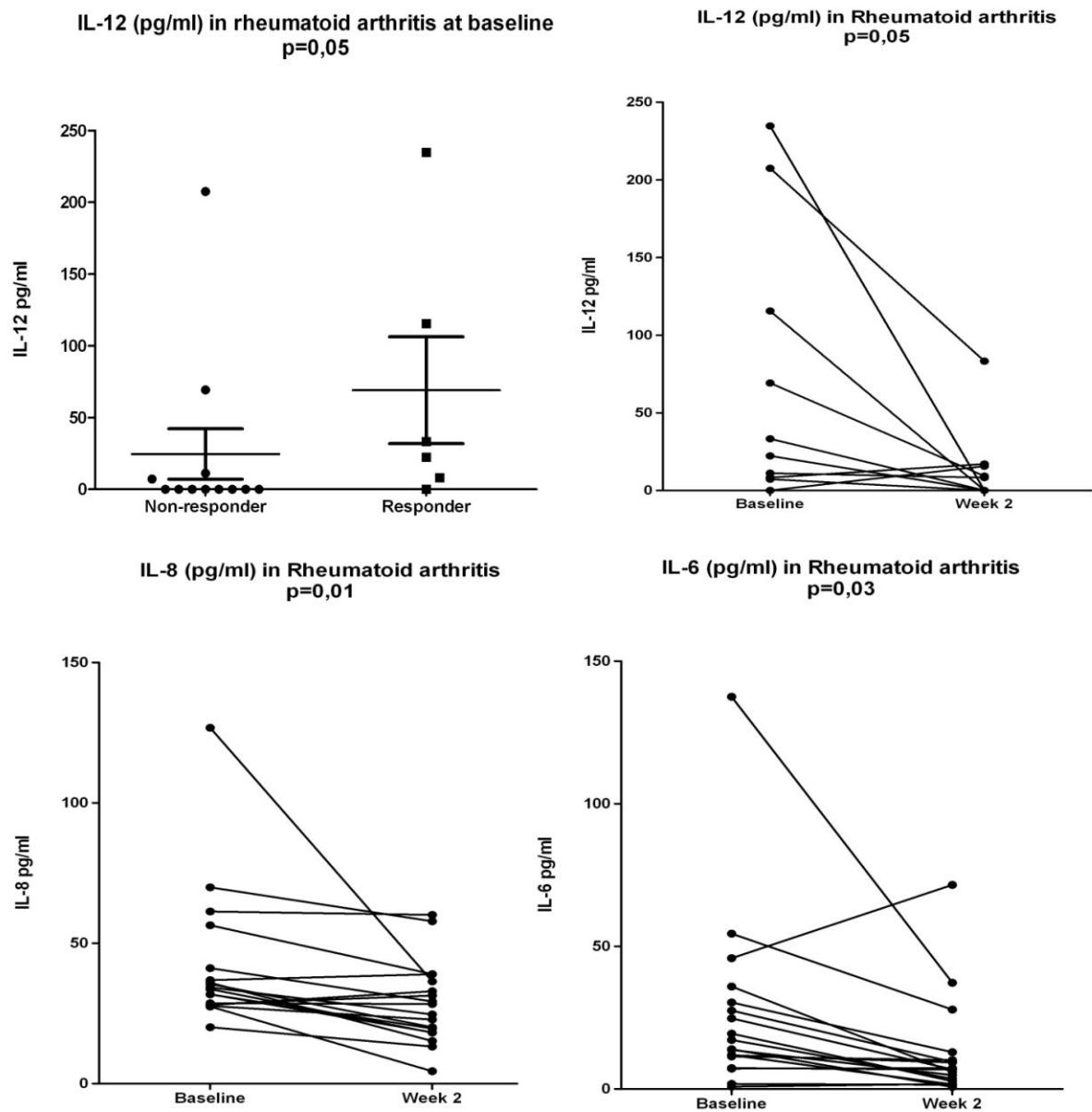


Figure 37c: ELISA results in IL-12, IL-6 and IL-8 in RA at baseline regarding the responder vs non-responder; and baseline vs week 2 comparisons.

6. 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 [67]. We reviewed several autoimmune conditions that have been investigated in different types of samples and with different methods [55] 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 (Figure 14). Regarding studies focusing on the response to biologic therapies from the gene expression point of view (Figure 38) which were also assigned to research groups and geographical locations of their cohorts (Figure 39a-b) showing a lack of collaboration in this area [86].

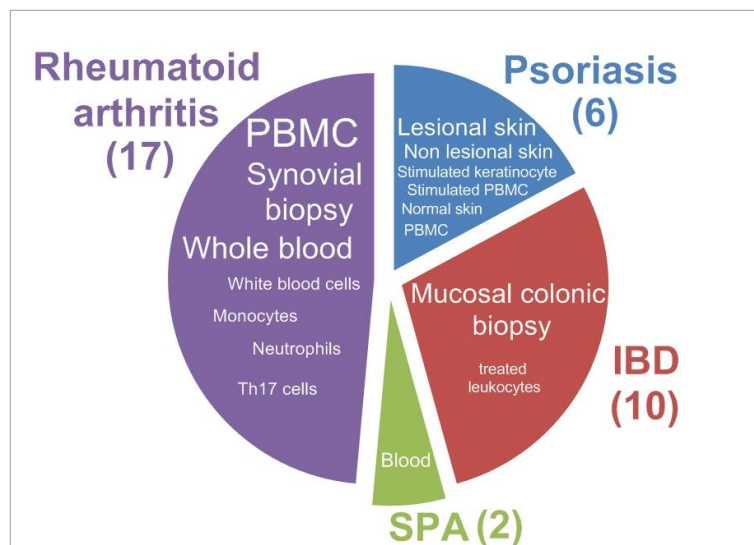


Figure 38: The number of diseases covered in studies focusing on gene expression profiling and the techniques used. Pie chart of diseases covered by studies focusing on gene expression profiling in pharmacogenomics (number of studies are shown). Word

cloud represents the techniques studies used for each disease; and the size of the words correlate with the number of studies using that technique.

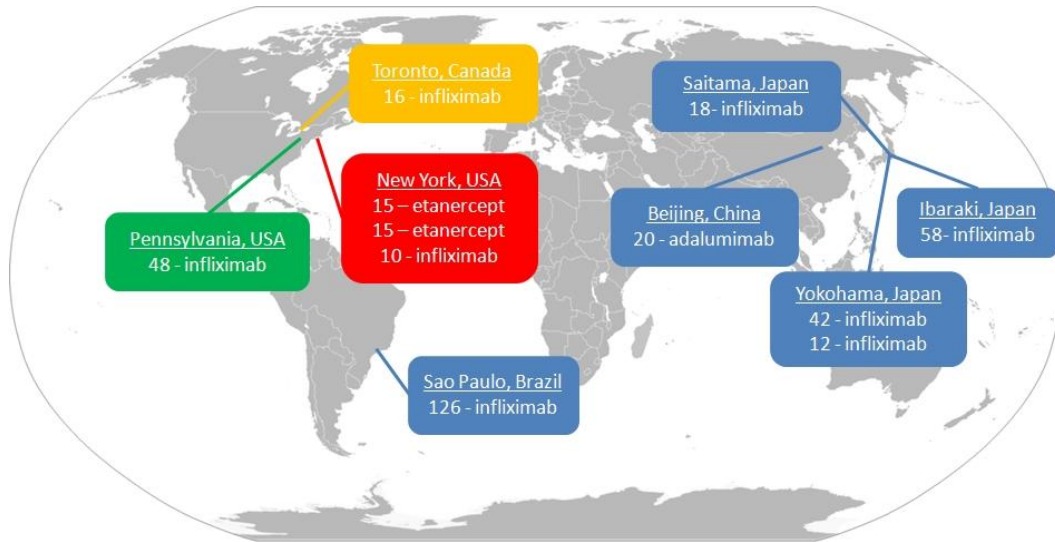


Figure 39a: Global map of studies focusing on gene expression profiling of the response to biologic therapies in autoimmune diseases. Boxes point to the city in which the first author of the particular study was affiliated. Colors represent diseases (blue - RA, red - psoriasis, yellow - AS and green - IBD). Each line in a box represents a study in which the number of patients and the therapy used are indicated.

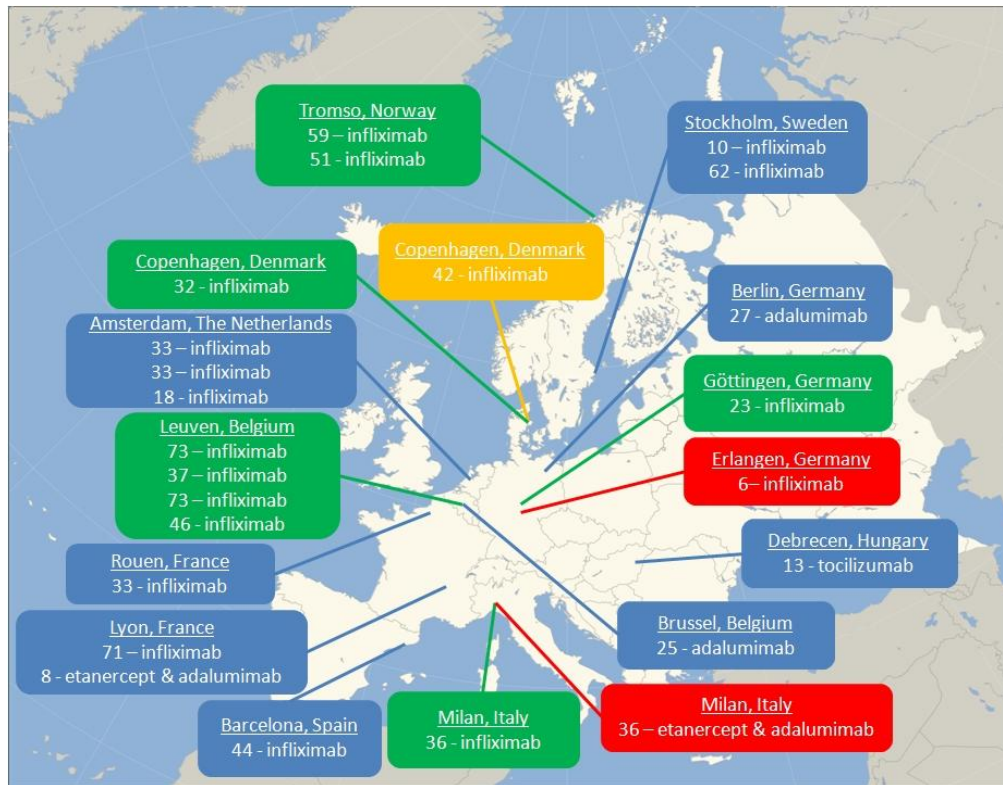


Figure 39b: European map of studies focusing on gene expression profiling of the response to biologic therapies in autoimmune diseases. Boxes point to the city in which the first author of the particular study was affiliated. Colors represent diseases (blue - RA, red - psoriasis, yellow - AS and green - IBD). Each line in a box represents a study in which the number of patients and the therapy used are indicated.

6.1 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 [87], CCR1 [88] in psoriasis; ADAM10 [89], PTPN22 [90], IL8 [91] in RA; and GZMK [76], MMP9 [92] and IFNG [93] 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.

Another important step was using gene expression profiling from peripheral blood for detecting different stages of disease prognosis or subtypes of a particular disease. The

comparison of CD and UC as subtypes of IBD yielded a list of genes out of which FOS [94] and ETS2 [95] have already been related to IBD while IL-18 may serve as an indicator of acute phase reactivity in CD as opposed to UC that correlates with our findings [96].

8 genes were identified showing significant differences between psoriasis patients with and without arthritis. Of these genes, IL4 on the protein level [97]; as well as MMP2 and MMP9 have been implicated in several features of inflammatory arthritis [98].

Separating different states of prognosis in RA regarding MR-confirmed bone erosion resulted in a gene set including HBB, GZMK and PADI4. PADI4 has been associated with RA several times being a target of autoantibodies; having an increased expression in synovia of RA patients [99], also related to the disease as an SNP [100].

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 [101] and in the synovial tissue of RA patients [102]. IL10 is a known marker of psoriasis [103], RA [104] and IBD patients [105] in PBMCs. The expression of NAMPT, a pre-B cell-enhancing factor, is increased in colonic biopsy of IBD patients compared to healthy subjects [106], as well as in plasma and synovial fluid of RA patients [107]; and in PBMC of psoriasis [91]. Despite the fact that AQP9 has an activity in immune response, it has never been associated with any of these medical conditions.

Although, when we put these genes in a network analyzing the biological processes they are related to, two conclusions could have been drawn. First, as expected, IL10 is the key gene in this network; and second, there might be other potential markers that play role in the pathomechanism of the conditions as the number of biological processes related to ADM is relatively high without direct interactions with the genes we examined. These all together may suggest a similar pathogenetic background all of these chronic inflammatory conditions as reflected by peripheral blood gene expression profiling.

6.2 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 [108] 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 [109]; 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.

In our study, we demonstrated the first example that (1) gene expression changes of tocilizumab therapy can be derived from PBMCs and can determine clinical outcome particularly if combined with clinical parameters; 2) significant increase of the degree of galactosylation of N-glycans of IgG occurs after tocilizumab therapy corroborating previous reports [110].

6.3 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.

Analysis of CD identified CYP1B1 that was marked as an IBD marker compared to healthy controls in our previous study [111]; FCGR1A which is related to inflammation was up-regulated in the PBMCs of UC and CD patients [112]; IL18R1 [113] and PRDM1 [114] were associated with CD as SNP candidates; and RNASE2 showed a reduced expression in IBD patients compared to healthy subjects in peripheral polymorphonuclear leucocytes [115].

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 [116]. PTGS2 that discriminated RA patients from healthy subjects in our study [111]; and reduced folate carrier (RFC1) of which genetic polymorphism can modify MTX transport thus influencing response to treatment [117] were found to be significant.

The comparisons of baseline and week 2 samples by microarray in both conditions demonstrated the effects of the therapy. From this perspective, analysis of CD samples resulted in a list of 5 genes including AQP9 and TNFAIP6 that were IBD markers in our previous work [111]; IGKC; MGAM that was found to be an SNP predictor of response to TNFa blocking therapy in pediatric IBD [118]; and MMP8 that is present in the actively inflamed area of colonic samples [119] and in the stroma as well [120].

The same analysis in RA detected AQP9 and TNFAIP6 that discriminated RA patients from healthy controls in our work similarly to CD; and immunoglobulin J that showed higher gene expression levels in twins with RA compared to their healthy co-twins [121] meaning that each gene is relevant to the pathogenesis of RA.

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 [122]; UBE2H of which TNF α is a regulator [123]; 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 osteoarthritis [124]; 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 [125].

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 [126]. The importance of the gene in the pathogenetics of the disease or therapy does not necessarily mean 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 (Table 12).

These gene groups containing 24 genes were tested with 100 runs using the uniform model as previously described and both produced 97-97 100% segregation gene panels underscoring the notion that these sets can serve as good starting points for further developments.

Table 12: 24 genes with the best discriminatory power ranked by summarized F scores in CD and RA

CD		RA	
Gene	F Score	Gene	F Score
1 CADM2	240952	RFC1	167054
2 ENDOD1	215519	SERPINB10	156077
3 ODC1	191135	DHRS9	151149
4 AIDA	191056	EPSTI1	144760
5 UBE2H	190080	FCGR3A	143228
6 FMN1	167706	GPR15	136897
7 CD300E	161073	OR2A9P	135391
8 PCYT1B	160930	PF4	133786
9 MAP1LC3B	160566	IFI44	132858
10 KAT2B	153562	APOBEC3A	131955
11 MMD	150816	RSAD2	124992
12 CA2	150714	S100P	118786
13 BMP6	149255	IFIT1	118722
14 BTN3A2	149142	MICA	117776
15 RIOK3	135968	EIF2AK2	116605
16 IL1RL1	128245	CCL4	115561
17 GCLC	127954	IFIT3	113547
18 RNF11	127952	RAVER2	112869
19 WARS	126943	SERPING1	109882
20 ARHGEF12	125960	IFIT2	106275
21 CYP1B1	125714	SIGLEC1	106138
22 PBX1	125236	IFI44L	102521
23 IL18R1	123542	IRF7	102067
24 PIP5K1B	123073	MX1	96453

It could mean that on a larger scale, these gene panels should be tested in international cohorts to detect the potentially additional underlying and yet uncovered differences solving unmet needs in the clinical settings. Determination of a response to a

specific and fairly expensive biologic therapy before starting the therapy by analyzing the gene expression patterns of the least invasively obtained PBMCs and thus preventing the patient from receiving an inefficient therapy has major advantages.

6.4 Implications for autoimmune conditions and future studies

Overlapping gene expression signatures were found in the PBMCs of patients with RA and SLE [127], as well as with RA, psoriasis and IBD in our work [128] which could be explained in two ways. 1) It is caused by a common underlying mechanism (the “cause” scenario). 2) It results from the common inflammatory component of the diseases (the “consequence” scenario). We predicted that a balanced mix of the two scenarios leads to the best explanation, notably early gene expression changes in the disease fall into the “cause”, while changes in a later stage fall into the “consequence” scenario.

Although overlap is not always present. Examples include when gene expression signatures of patients with PsA did not show significant overlaps with profiles from patients with RA or SLE [129] leading to the observation that the degree of such overlapping signatures might be disease specific and indicative of the existence of disease specific signatures.

We decided to analyze significantly changing genes considered functionally important for autoimmune conditions such as psoriasis, RA, PsA, AS, COPD, IBD, SLE and MS in order to detect functionally significant genes that contribute to overlapping and distinct gene expression changes. This analysis resulted in the identification of a gene network with three layers of genes based on PBMC gene expression profiling studies. The first layer that centers around TNF, IL-8 and IFN is linked to rheumatic and psoriatic conditions which have a common pathogenetic background including psoriasis and PsA [130]; or psoriasis and RA [128] (Figure 40a). The second layer highlights how conditions with organ-specificity such as COPD [131], IBD [76] and MS [132] [133] are connected to the first layer through TNF, IL-8, IFN and IL-1B (Figure 40b). The third layer adds SLE that manifests in systemic characteristics such as skin lesions, lung

inflammation or joint pain having connections to all the genes (e.g. TNF, IL-8, IFN and IL-1B) that play a central role in the pathogenesis of these diseases (Figure 40c).

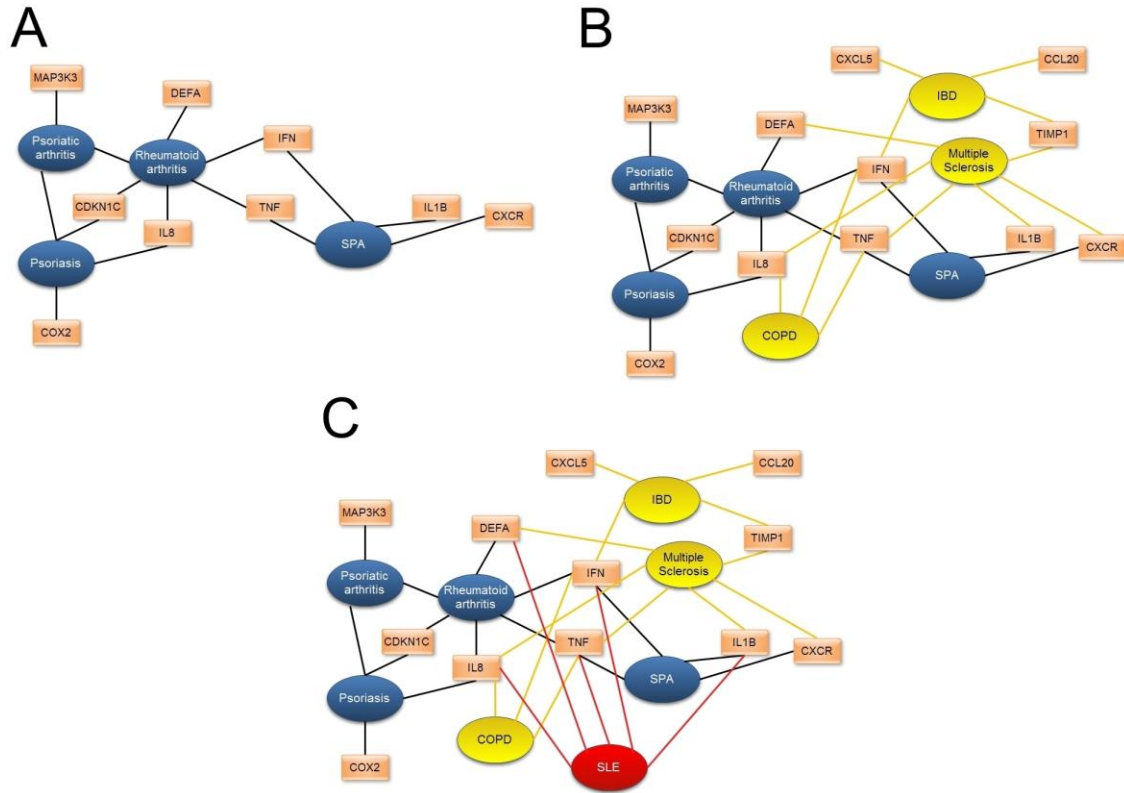


Figure 40: Genes associated with the diseases form a network of three layers. The connections between genes and diseases are meant to represent that that specific gene was found statistically significant regarding changes between healthy patients and patients with the particular disease according to the cited study. Ellipses represent diseases, boxes represent genes. The color of the lines shows which genes have interactions with which diseases. The first layer of clustered genes is linked to rheumatic and psoriatic disorders that have a common pathogenetic background centered around TNF, IL-8 and IFN (a). A second layer shows how the diseases with distinct organ-specificity such as COPD, IBD and MS can be connected to this network through TNF, IL-8, IFN and IL-1B (b). The third layer represents the systemic characteristics of SLE, having connections to all the genes that play a central role in these conditions (c).

These gene clusters or networks might suggest that the comparison of different autoimmune disorders can lead to the detection of new gene panels determining disease progression or subtypes. The limitations include the use of different platforms and screens by the various studies, therefore important links might be missing for technical reasons.

In order to facilitate standardization and overcome this problem, we proposed a general protocol for performing gene expression profiling for autoimmune diseases, as well as the use of a decision tree describing how gene expression studies for identifying disease specific gene expression changes should be performed (Figure 41). This flowchart was designed in order to assist researchers without previous experience with gene expression profiling in getting a global view of the structure of such studies, hoping it would open new doors for future interdisciplinary collaborations.

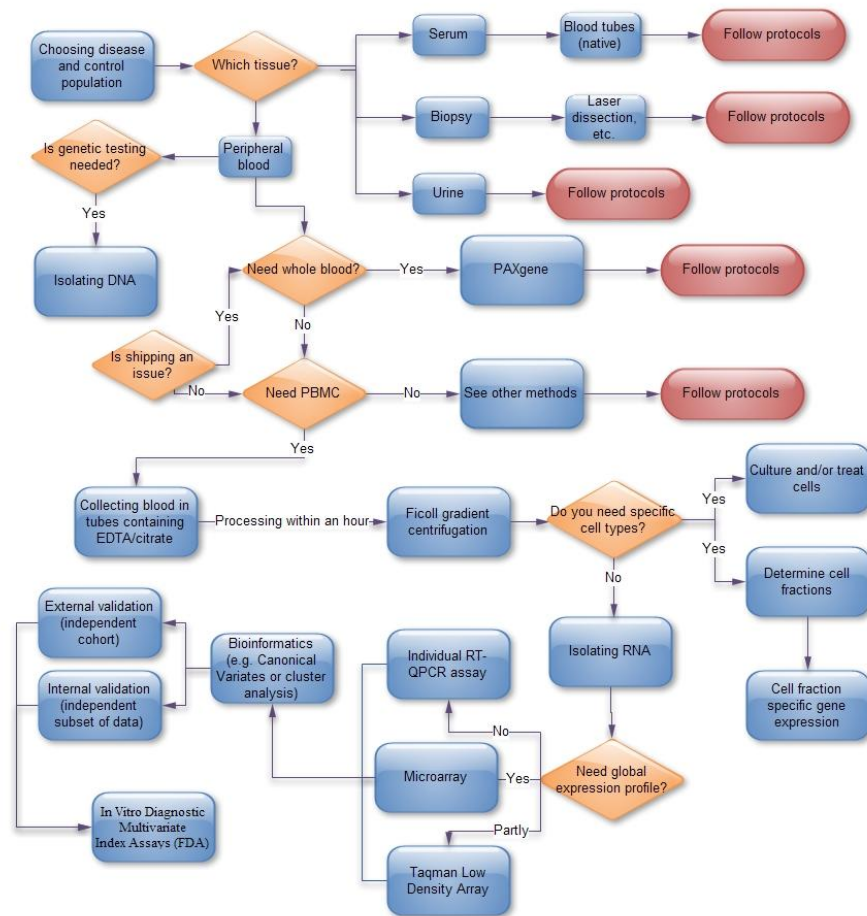


Figure 41: This decision tree was designed to help researchers gain a global view of the structure of such studies. Blue boxes represent process elements; yellow boxes represent questions.

7. 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.

The network approach also applies to the design of the studies as labs should combine efforts in analyzing blood samples drawn in different parts of the world which could serve as the immunological alternative (or complement) of the Human Genome Project [134]. Combining the networks and capabilities of different labs might result in better understanding of autoimmune conditions and more efficient prediction models in determining drug responsiveness or disease progression. Peripheral blood gene expression profiling will be a valuable research tool and play an important and later indispensable role in developing genomic assays in clinical diagnostics.

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.

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.

The work was supported by TAMOP-4.2.2/08/2 and TAMOP-4.2.1/B-09/1KONV-2010-0007 implemented through the New Hungary Development Plan co-financed by the European Social Fund and the European Regional Development Fund.

9. REFERENCES

1. *The Personalized Medicine Coalition*.
<http://www.personalizedmedicinecoalition.org/about>, 2012.
2. Issa, A.M., *Personalized medicine and the practice of medicine in the 21st century*. McGill J Med, 2007. **10**(1): p. 53-7.
3. Klein, T.E., et al., *Estimation of the warfarin dose with clinical and pharmacogenetic data*. N Engl J Med, 2009. **360**(8): p. 753-64.
4. Costa, R.B., et al., *Efficacy and cardiac safety of adjuvant trastuzumab-based chemotherapy regimens for HER2-positive early breast cancer*. Ann Oncol, 2010. **21**(11): p. 2153-60.
5. Greco, J.A., 3rd, et al., *A microarray analysis of temporal gene expression profiles in thermally injured human skin*. Burns, 2010. **36**(2): p. 192-204.
6. *An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains*. Nat Genet, 1997. **17**(4): p. 399-403.
7. Gambineri, E., T.R. Torgerson, and H.D. Ochs, *Immune dysregulation, polyendocrinopathy, enteropathy, and X-linked inheritance (IPEX), a syndrome of systemic autoimmunity caused by mutations of FOXP3, a critical regulator of T-cell homeostasis*. Curr Opin Rheumatol, 2003. **15**(4): p. 430-5.
8. Brown, M.A., et al., *Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis*. Ann Rheum Dis, 2000. **59**(11): p. 883-6.
9. Ladd, A.T., *Procainamide-induced lupus erythematosus*. N Engl J Med, 1962. **267**: p. 1357-8.
10. Bell, S.A., *The toxic oil syndrome: an example of an exogenously induced autoimmune reaction*. Toxic oil syndrome. Mol Biol Rep, 1996. **23**(3-4): p. 261-3.
11. Bisno, A.L., *Group A streptococcal infections and acute rheumatic fever*. N Engl J Med, 1991. **325**(11): p. 783-93.
12. Voskuhl, R., *Sex differences in autoimmune diseases*. Biol Sex Differ, 2011. **2**(1): p. 1.
13. Whitacre, C.C., *Sex differences in autoimmune disease*. Nat Immunol, 2001. **2**(9): p. 777-80.
14. Medzhitov, R., *Recognition of microorganisms and activation of the immune response*. Nature, 2007. **449**(7164): p. 819-26.
15. Dranoff, G., *Cytokines in cancer pathogenesis and cancer therapy*. Nat Rev Cancer, 2004. **4**(1): p. 11-22.
16. Miller, J.F., *Self-nonsel discrimination and tolerance in T and B lymphocytes*. Immunol Res, 1993. **12**(2): p. 115-30.
17. Sproul, T.W., et al., *A role for MHC class II antigen processing in B cell development*. Int Rev Immunol, 2000. **19**(2-3): p. 139-55.
18. Romagnani, S., *Immunological tolerance and autoimmunity*. Intern Emerg Med, 2006. **1**(3): p. 187-96.

19. Thomas, R., *The balancing act of autoimmunity: central and peripheral tolerance versus infection control*. Int Rev Immunol, 2010. **29**(2): p. 211-33.
20. Boissier, M.C., et al., *Rheumatoid arthritis: From autoimmunity to synovitis and joint destruction*. J Autoimmun, 2012.
21. Fairweather, D., *Autoimmune Disease: Mechanisms*. ENCYCLOPEDIA OF LIFE SCIENCES 2007: p. 3-4.
22. Rose, N.R. and C. Bona, *Defining criteria for autoimmune diseases (Witebsky's postulates revisited)*. Immunol Today, 1993. **14**(9): p. 426-30.
23. Walsh, S.J. and L.M. Rau, *Autoimmune diseases: a leading cause of death among young and middle-aged women in the United States*. Am J Public Health, 2000. **90**(9): p. 1463-6.
24. Szegedi, G., *Characteristics of autoimmune conditions*. LAM, 2004.
25. Zhou, X., et al., *Lentivirus-mediated gene transfer and expression in established human tumor antigen-specific cytotoxic T cells and primary unstimulated T cells*. Hum Gene Ther, 2003. **14**(11): p. 1089-105.
26. Silman, A.J. and J.E. Pearson, *Epidemiology and genetics of rheumatoid arthritis*. Arthritis Res, 2002. **4 Suppl 3**: p. S265-72.
27. Alamanos, Y., P.V. Voulgari, and A.A. Drosos, *Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review*. Semin Arthritis Rheum, 2006. **36**(3): p. 182-8.
28. Arnett, F.C., et al., *The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis*. Arthritis Rheum, 1988. **31**(3): p. 315-24.
29. Aletaha, D., et al., *2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative*. Arthritis Rheum, 2010. **62**(9): p. 2569-81.
30. Spector, T.D., *Rheumatoid arthritis*. Rheum Dis Clin North Am, 1990. **16**(3): p. 513-37.
31. van Zeben, D., et al., *Association of HLA-DR4 with a more progressive disease course in patients with rheumatoid arthritis. Results of a followup study*. Arthritis Rheum, 1991. **34**(7): p. 822-30.
32. Albano, S.A., E. Santana-Sahagun, and M.H. Weisman, *Cigarette smoking and rheumatoid arthritis*. Semin Arthritis Rheum, 2001. **31**(3): p. 146-59.
33. Maini, R.N. and M. Feldmann, *How does infliximab work in rheumatoid arthritis?* Arthritis Res, 2002. **4 Suppl 2**: p. S22-8.
34. Knight, D.M., et al., *Construction and initial characterization of a mouse-human chimeric anti-TNF antibody*. Mol Immunol, 1993. **30**(16): p. 1443-53.
35. Smolen, J.S., et al., *New therapies for treatment of rheumatoid arthritis*. Lancet, 2007. **370**(9602): p. 1861-74.
36. Maini, R.N., et al., *Monoclonal anti-TNF alpha antibody as a probe of pathogenesis and therapy of rheumatoid disease*. Immunol Rev, 1995. **144**: p. 195-223.
37. Tracey, D., et al., *Tumor necrosis factor antagonist mechanisms of action: a comprehensive review*. Pharmacol Ther, 2008. **117**(2): p. 244-79.
38. Dayer, J.M. and E. Choy, *Therapeutic targets in rheumatoid arthritis: the interleukin-6 receptor*. Rheumatology (Oxford), 2010. **49**(1): p. 15-24.

39. Suzuki, M., et al., *Anti-inflammatory mechanism of tocilizumab, a humanized anti-IL-6R antibody: effect on the expression of chemokine and adhesion molecule*. Rheumatol Int, 2010. **30**(3): p. 309-15.
40. Oldfield, V., S. Dhillon, and G.L. Plosker, *Tocilizumab: a review of its use in the management of rheumatoid arthritis*. Drugs, 2009. **69**(5): p. 609-32.
41. Nishimoto, N., et al., *Treatment of rheumatoid arthritis with humanized anti-interleukin-6 receptor antibody: a multicenter, double-blind, placebo-controlled trial*. Arthritis Rheum, 2004. **50**(6): p. 1761-9.
42. Smolen, J.S., et al., *Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial*. Lancet, 2008. **371**(9617): p. 987-97.
43. Reveille, J.D., *Major histocompatibility genes and ankylosing spondylitis*. Best Pract Res Clin Rheumatol, 2006. **20**(3): p. 601-9.
44. Sidiropoulos, P.I., et al., *Evidence-based recommendations for the management of ankylosing spondylitis: systematic literature search of the 3E Initiative in Rheumatology involving a broad panel of experts and practising rheumatologists*. Rheumatology (Oxford), 2008. **47**(3): p. 355-61.
45. Langley, R.G., G.G. Krueger, and C.E. Griffiths, *Psoriasis: epidemiology, clinical features, and quality of life*. Ann Rheum Dis, 2005. **64 Suppl 2**: p. ii18-23; discussion ii24-5.
46. Schmitt, J. and G. Wozel, *The psoriasis area and severity index is the adequate criterion to define severity in chronic plaque-type psoriasis*. Dermatology, 2005. **210**(3): p. 194-9.
47. Nestle, F.O., D.H. Kaplan, and J. Barker, *Psoriasis*. N Engl J Med, 2009. **361**(5): p. 496-509.
48. Farragher, T.M., et al., *Benefit of early treatment in inflammatory polyarthritis patients with anti-cyclic citrullinated peptide antibodies versus those without antibodies*. Arthritis Care Res (Hoboken), 2010. **62**(5): p. 664-75.
49. Liu, Y., et al., *A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci*. PLoS Genet, 2008. **4**(3): p. e1000041.
50. Loftus, E.V., Jr., *Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences*. Gastroenterology, 2004. **126**(6): p. 1504-17.
51. Dessein, R., M. Chamaillard, and S. Danese, *Innate immunity in Crohn's disease: the reverse side of the medal*. J Clin Gastroenterol, 2008. **42 Suppl 3 Pt 1**: p. S144-7.
52. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 603-6.
53. Abbas, A.K., and A. H. Lichtman *Cellular and molecular immunology*. Vol. 5th Edition. 2003: Elsevier Science, Philadelphia, Pa.
54. Rox, J.M., et al., *Gene expression analysis in platelets from a single donor: evaluation of a PCR-based amplification technique*. Clin Chem, 2004. **50**(12): p. 2271-8.
55. Mesko, B., S. Poliska, and L. Nagy, *Gene expression profiles in peripheral blood for the diagnosis of autoimmune diseases*. Trends Mol Med, 2011. **17**(4): p. 223-33.

56. Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling*. Nature, 2000. **403**(6769): p. 503-11.
57. Aziz, H., A. Zaas, and G.S. Ginsburg, *Peripheral blood gene expression profiling for cardiovascular disease assessment*. Genomic Med, 2007. **1**(3-4): p. 105-12.
58. Zhang, J., et al., *Role of TL1A in the pathogenesis of rheumatoid arthritis*. J Immunol, 2009. **183**(8): p. 5350-7.
59. van Baarsen, L.G., et al., *Gene expression profiling in autoantibody-positive patients with arthralgia predicts development of arthritis*. Arthritis Rheum, 2010. **62**(3): p. 694-704.
60. van der Pouw Kraan, T.C., et al., *Discovery of distinctive gene expression profiles in rheumatoid synovium using cDNA microarray technology: evidence for the existence of multiple pathways of tissue destruction and repair*. Genes Immun, 2003. **4**(3): p. 187-96.
61. Gibson, M., http://www.mun.ca/biology/scarr/cDNA_microarray_Assay_of_Gene_Expression.html, 2002.
62. Whitney, A.R., et al., *Individuality and variation in gene expression patterns in human blood*. Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1896-901.
63. Idaghdour, Y., et al., *A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs*. PLoS Genet, 2008. **4**(4): p. e1000052.
64. Wu, F., et al., *Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis*. Inflamm Bowel Dis, 2007. **13**(7): p. 807-21.
65. Nomura, I., et al., *Distinct patterns of gene expression in the skin lesions of atopic dermatitis and psoriasis: a gene microarray analysis*. J Allergy Clin Immunol, 2003. **112**(6): p. 1195-202.
66. van Baarsen, L.G., et al., *Synovial tissue heterogeneity in rheumatoid arthritis in relation to disease activity and biomarkers in peripheral blood*. Arthritis Rheum, 2010. **62**(6): p. 1602-7.
67. Dudley, J.T., et al., *Disease signatures are robust across tissues and experiments*. Mol Syst Biol, 2009. **5**: p. 307.
68. Liu, Z., K. Maas, and T.M. Aune, *Comparison of differentially expressed genes in T lymphocytes between human autoimmune disease and murine models of autoimmune disease*. Clin Immunol, 2004. **112**(3): p. 225-30.
69. Rus, V., et al., *Gene expression profiling in peripheral blood mononuclear cells from lupus patients with active and inactive disease*. Clin Immunol, 2004. **112**(3): p. 231-4.
70. Edwards, C.J., et al., *Molecular profile of peripheral blood mononuclear cells from patients with rheumatoid arthritis*. Mol Med, 2007. **13**(1-2): p. 40-58.
71. Gu, J., et al., *A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthropathy patients*. Rheumatology (Oxford), 2002. **41**(7): p. 759-66.
72. Lee, E., et al., *Increased expression of interleukin 23 p19 and p40 in lesional skin of patients with psoriasis vulgaris*. J Exp Med, 2004. **199**(1): p. 125-30.

73. Bowcock, A.M., et al., *Insights into psoriasis and other inflammatory diseases from large-scale gene expression studies*. Hum Mol Genet, 2001. **10**(17): p. 1793-805.
74. Rihl, M., et al., *Technical validation of cDNA based microarray as screening technique to identify candidate genes in synovial tissue biopsy specimens from patients with spondyloarthropathy*. Ann Rheum Dis, 2004. **63**(5): p. 498-507.
75. Batliwalla, F.M., et al., *Microarray analyses of peripheral blood cells identifies unique gene expression signature in psoriatic arthritis*. Mol Med, 2005. **11**(1-12): p. 21-9.
76. Burczynski, M.E., et al., *Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells*. J Mol Diagn, 2006. **8**(1): p. 51-61.
77. He, C., et al., *Higher levels of CCL20 expression on peripheral blood mononuclear cells of chinese patients with inflammatory bowel disease*. Immunol Invest, 2010. **39**(1): p. 16-26.
78. Steinbach, D., et al., *Identification of a set of seven genes for the monitoring of minimal residual disease in pediatric acute myeloid leukemia*. Clin Cancer Res, 2006. **12**(8): p. 2434-41.
79. Schobesberger, M., et al., *Gene expression variation between distinct areas of breast cancer measured from paraffin-embedded tissue cores*. BMC Cancer, 2008. **8**: p. 343.
80. Kosa, J.P., et al., *Postmenopausal expression changes of immune system-related genes in human bone tissue*. J Clin Immunol, 2009. **29**(6): p. 761-8.
81. Warde-Farley, D., et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function*. Nucleic Acids Res, 2010. **38**(Web Server issue): p. W214-20.
82. Podani, J., *SYN-TAX 2000. Computer programs for data analysis in ecology and systematics. User's Manual*. Scientia, Budapest, 2001.
83. Hamadeh, H.K., et al., *Prediction of compound signature using high density gene expression profiling*. Toxicol Sci, 2002. **67**(2): p. 232-40.
84. Venables, W.N.R., B. D, ed. *Modern Applied Statistics with S*. Fourth ed. 2002, Springer, New York.
85. Hosack, D.A., et al., *Identifying biological themes within lists of genes with EASE*. Genome Biol, 2003. **4**(10): p. R70.
86. Mesko, B., G. Zahuczky, and L. Nagy, *The triad of success in personalised medicine: pharmacogenomics, biotechnology and regulatory issues from a Central European perspective*. N Biotechnol, 2012.
87. Koczan, D., et al., *Gene expression profiling of peripheral blood mononuclear leukocytes from psoriasis patients identifies new immune regulatory molecules*. Eur J Dermatol, 2005. **15**(4): p. 251-7.
88. Rottman, J.B., et al., *Potential role of the chemokine receptors CXCR3, CCR4, and the integrin alphaEbeta7 in the pathogenesis of psoriasis vulgaris*. Lab Invest, 2001. **81**(3): p. 335-47.
89. van der Voort, R., et al., *Elevated CXCL16 expression by synovial macrophages recruits memory T cells into rheumatoid joints*. Arthritis Rheum, 2005. **52**(5): p. 1381-91.

90. Batliwalla, F.M., et al., *Peripheral blood gene expression profiling in rheumatoid arthritis*. Genes Immun, 2005. **6**(5): p. 388-97.
91. Koczan, D., et al., *Molecular discrimination of responders and nonresponders to anti-TNF alpha therapy in rheumatoid arthritis by etanercept*. Arthritis Res Ther, 2008. **10**(3): p. R50.
92. Fireman, E., et al., *Induced sputum-retrieved matrix metalloproteinase 9 and tissue metalloproteinase inhibitor 1 in granulomatous diseases*. Clin Exp Immunol, 2002. **130**(2): p. 331-7.
93. Gonsky, R., R.L. Deem, and S.R. Targan, *Distinct Methylation of IFNG in the Gut*. J Interferon Cytokine Res, 2009. **29**(7): p. 407-14.
94. Alexander, R.J., et al., *Expression of protooncogene-encoded mRNA by colonic epithelial cells in inflammatory bowel disease*. Dig Dis Sci, 1996. **41**(4): p. 660-9.
95. van der Pouw Kraan, T.C., et al., *Acute experimental colitis and human chronic inflammatory diseases share expression of inflammation-related genes with conserved Ets2 binding sites*. Inflamm Bowel Dis, 2009. **15**(2): p. 224-35.
96. Haas, S.L., et al., *Interleukin-18 serum levels in inflammatory bowel diseases: correlation with disease activity and inflammatory markers*. Swiss Med Wkly, 2009. **139**(9-10): p. 140-5.
97. Partsch, G., et al., *T cell derived cytokines in psoriatic arthritis synovial fluids*. Ann Rheum Dis, 1998. **57**(11): p. 691-3.
98. Hitchon, C.A., et al., *Gelatinase expression and activity in the synovium and skin of patients with erosive psoriatic arthritis*. J Rheumatol, 2002. **29**(1): p. 107-17.
99. Chang, X., et al., *The expression of PADI4 in synovium of rheumatoid arthritis*. Rheumatol Int, 2009. **29**(12): p. 1411-6.
100. Li, Y. and A.B. Begovich, *Unraveling the genetics of complex diseases: susceptibility genes for rheumatoid arthritis and psoriasis*. Semin Immunol, 2009. **21**(6): p. 318-27.
101. Marutsuka, K., et al., *Adrenomedullin and proadrenomedullin N-terminal 20 peptide (PAMP) are present in human colonic epithelia and exert an antimicrobial effect*. Exp Physiol, 2001. **86**(5): p. 543-5.
102. Chosa, E., et al., *Increased plasma and joint tissue adrenomedullin concentrations in patients with rheumatoid arthritis compared to those with osteoarthritis*. J Rheumatol, 2003. **30**(12): p. 2553-6.
103. Van Der Fits, L., et al., *Polymorphisms in the interferon regulatory factor-1 promoter are not associated with psoriasis and do not influence IFN-alpha-induced Th1 polarization*. J Interferon Cytokine Res, 2007. **27**(10): p. 841-6.
104. Bucht, A., et al., *Expression of interferon-gamma (IFN-gamma), IL-10, IL-12 and transforming growth factor-beta (TGF-beta) mRNA in synovial fluid cells from patients in the early and late phases of rheumatoid arthritis (RA)*. Clin Exp Immunol, 1996. **103**(3): p. 357-67.
105. Gasche, C., et al., *IL-10 secretion and sensitivity in normal human intestine and inflammatory bowel disease*. J Clin Immunol, 2000. **20**(5): p. 362-70.
106. Moschen, A.R., et al., *Visfatin, an adipocytokine with proinflammatory and immunomodulating properties*. J Immunol, 2007. **178**(3): p. 1748-58.

107. Nowell, M.A., et al., *Regulation of pre-B cell colony-enhancing factor by STAT-3-dependent interleukin-6 trans-signaling: implications in the pathogenesis of rheumatoid arthritis*. Arthritis Rheum, 2006. **54**(7): p. 2084-95.
108. Sharma, S., et al., *Purine biosynthetic pathway genes and methotrexate response in rheumatoid arthritis patients among north Indians*. Pharmacogenet Genomics, 2009. **19**(10): p. 823-8.
109. Miyazaki, E., et al., *Highly restricted T-cell receptor repertoire in the CD8+ T-cell response against an HIV-1 epitope with a stereotypic amino acid substitution*. AIDS, 2009. **23**(6): p. 651-60.
110. Alavi, A. and J. Axford, *Evaluation of beta 1,4-galactosyltransferase in rheumatoid arthritis and its role in the glycosylation network associated with this disease*. Glycoconj J, 1995. **12**(3): p. 206-10.
111. Mesko, B., et al., *Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets*. BMC Med Genomics, 2010. **3**: p. 15.
112. Chae, S.C., et al., *Identification of single nucleotide polymorphisms in the TNFRSF17 gene and their association with gastrointestinal disorders*. Mol Cells, 2010. **29**(1): p. 21-8.
113. Wang, K., et al., *Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease*. Am J Hum Genet, 2009. **84**(3): p. 399-405.
114. Torkvist, L., et al., *Analysis of 39 Crohn's disease risk loci in Swedish inflammatory bowel disease patients*. Inflamm Bowel Dis, 2010. **16**(6): p. 907-9.
115. Bovin, L.F., et al., *Gene expression profiling in autoimmune diseases: chronic inflammation or disease specific patterns?* Autoimmunity, 2007. **40**(3): p. 191-201.
116. van Baarsen, L.G., et al., *Pharmacogenomics of infliximab treatment using peripheral blood cells of patients with rheumatoid arthritis*. Genes Immun, 2010. **11**(8): p. 622-9.
117. Bohanec Grabar, P., et al., *Genetic determinants of methotrexate toxicity in rheumatoid arthritis patients: a study of polymorphisms affecting methotrexate transport and folate metabolism*. Eur J Clin Pharmacol, 2008. **64**(11): p. 1057-68.
118. Dubinsky, M.C., et al., *Genome wide association (GWA) predictors of anti-TNFalpha therapeutic responsiveness in pediatric inflammatory bowel disease*. Inflamm Bowel Dis, 2010. **16**(8): p. 1357-66.
119. Arihiro, S., et al., *Vascular smooth muscle cells and pericytes express MMP-1, MMP-9, TIMP-1 and type I procollagen in inflammatory bowel disease*. Histopathology, 2001. **39**(1): p. 50-9.
120. Pirila, E., et al., *Gelatinase A (MMP-2), collagenase-2 (MMP-8), and laminin-5 gamma2-chain expression in murine inflammatory bowel disease (ulcerative colitis)*. Dig Dis Sci, 2003. **48**(1): p. 93-8.
121. Haas, C.S., et al., *Identification of genes modulated in rheumatoid arthritis using complementary DNA microarray analysis of lymphoblastoid B cell lines from disease-discordant monozygotic twins*. Arthritis Rheum, 2006. **54**(7): p. 2047-60.
122. Condamine, T., et al., *Tmem176B and Tmem176A are associated with the immature state of dendritic cells*. J Leukoc Biol, 2010. **88**(3): p. 507-15.

123. Martin, I., et al., *Association study of the ubiquitin conjugating enzyme gene UBE2H in sporadic ALS*. Amyotroph Lateral Scler, 2009. **10**(5-6): p. 432-5.
124. Kumarasinghe, D.D., et al., *An update on primary hip osteoarthritis including altered Wnt and TGF-beta associated gene expression from the bony component of the disease*. Rheumatology (Oxford), 2011. **50**(12): p. 2166-75.
125. Trocme, C., et al., *Apolipoprotein A-I and platelet factor 4 are biomarkers for infliximab response in rheumatoid arthritis*. Ann Rheum Dis, 2009. **68**(8): p. 1328-33.
126. Ein-Dor, L., et al., *Outcome signature genes in breast cancer: is there a unique set?* Bioinformatics, 2005. **21**(2): p. 171-8.
127. Olsen, N., et al., *A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells*. Ann Rheum Dis, 2004. **63**(11): p. 1387-92.
128. Mesko, B., et al., *Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets*. BMC Med Genomics, 2010. **3**(1): p. 15.
129. Stoeckman, A.K., et al., *A distinct inflammatory gene expression profile in patients with psoriatic arthritis*. Genes Immun, 2006. **7**(7): p. 583-91.
130. Ko, J.M. and A.W. Qureshi, *Skin and bone: the pathogenetic relationship between psoriasis and psoriatic arthritis*. G Ital Dermatol Venereol, 2010. **145**(3): p. 393-405.
131. Kent, L., et al., *Cigarette smoke extract induced cytokine and chemokine gene expression changes in COPD macrophages*. Cytokine, 2008. **42**(2): p. 205-16.
132. Hong, J., et al., *Gene expression profiling of relevant biomarkers for treatment evaluation in multiple sclerosis*. J Neuroimmunol, 2004. **152**(1-2): p. 126-39.
133. Mandel, M., et al., *Autoimmunity gene expression portrait: specific signature that intersects or differentiates between multiple sclerosis and systemic lupus erythematosus*. Clin Exp Immunol, 2004. **138**(1): p. 164-70.
134. Leslie, M., *Biomedical research. Immunology uncaged*. Science, 2010. **327**(5973): p. 1573.

10. Publications

10.1 Publications related to the thesis



UNIVERSITY AND NATIONAL LIBRARY UNIVERSITY OF DEBRECEN
KENÉZY LIFE SCIENCES LIBRARY

Register Number: DEENKÉTK/296/2012.

Item Number:

Subject: Ph.D. List of Publications

Candidate: Bertalan Meskó

Neptun ID: BOODSE

Doctoral School: Doctoral School of Molecular Cell and Immune Biology

List of publications related to the dissertation

1. **Meskó, B.**, Pólska, 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.**, Pólska, 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.**, Pólska, S., Szegedi, A., Szekanecz, Z., Palatka, K., Papp, M., Nagy, L.: Peripheral blood gene expression patterns discriminate among chronic inflammatory 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



10.2 Other publications



UNIVERSITY AND NATIONAL LIBRARY UNIVERSITY OF DEBRECEN
KENÉZY LIFE SCIENCES LIBRARY

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.**, Fésűs, 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), 1971-1972, 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



10.3 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)

10.4 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

11. Appendix