DEPARTMENT OF MEDICINE, SOLNA
Karolinska Institutet, Stockholm, Sweden

Genomic Profiling and Gene-Gene Interaction in Rheumatoid Arthritis

Klementy Shchetynsky

Karolinska Institutet

Stockholm 2015

Published by Karolinska Institutet.
Printed by åtta.45 tryckeri AB

ISBN 978-91-7549-956-7

Genomic profiling and gene-gene interaction in rheumatoid arthritis
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Klementy Shchetynsky

Principal Supervisor:
Associate Professor Leonid Padyukov
Karolinska Institutet
Department of Medicine, Solna
Rheumatology Unit

Co-supervisor(s):
Professor Lars Klareskog
Karolinska Institutet
Department of Medicine, Solna
Rheumatology Unit

Opponent:
Professor Brage Storstein Andersen
University of Southern Denmark
Department of Biochemistry and Molecular Biology

Examination Board:
Professor Peter Zaphiropoulos
Karolinska Institutet
Department of Biosciences and Nutrition

Professor Lars Wieslander
Stockholm University
Department of Molecular Biosciences
The Wenner-Gren Institute

Professor Martin Schalling
Karolinska Institutet
Department of Molecular Medicine and Surgery
Division of Neurogenetics

"If you do not expect the unexpected, you will not find it; for it is hard to be sought out, and difficult"

Heraclitus

"There is no subject so old that something new cannot be said about it."

Fyodor Dostoevsky -- *A Diary of a Writer* (1876)

To my family

ABSTRACT

Complex disease is characterized by the interplay of multiple genetic and environmental factors. Rheumatoid arthritis (RA) is a complex autoimmune disease with a pronounced genetic component, mainly due to HLA-DRB1 gene, but also a multitude of loci outside the HLA region. In this work we strive to contribute to the understanding of the functional involvement of these susceptibility loci in the pathogenesis of RA.

This study is based on a large material of whole blood samples and peripheral blood mononuclear cells (PBMCs) from RA patients and matched healthy controls from Sweden. The main methods used in this work included probe-based genotyping and gene-expression assays, cell cultures, RNA-sequencing, gene-gene interaction and pathway analysis, as well as a plethora of common molecular genetics and bioinformatics methods.

We investigated the role of expression of known genetic risk factors PTPN22 and PTPN2 in RA, with a special attention to the splicing profile of these genes. Our data indicates significant differences in the expression ratio of splice variants for PTPN22 in whole blood samples from RA patients and healthy controls. For PTPN2 we demonstrate a significant difference in the relative mRNA expression of' transcript TC48 in PBMCs of healthy controls and RA patients. Additionally, we identified new susceptibility SNPs in the *PTPN2* locus: rs657555 and rs11080606, by addressing the interaction of PTPN2 variants with *HLA-DRB1* shared-epitope (SE) alleles in autoantibody positive RA patients in two independent cohorts.

In this work, we also address the functional genetic role of the members of the MAP signaling pathway upstream of p38 and JNK – crucial enzymes in RA – with a regard to splicing profile and their connection to HLA-DRB1. We found a significant statistical interaction for rs10468473 from MAP2K4 locus with SE alleles in autoantibody-positive RA. Importantly, individuals heterozygous for rs10468473 demonstrated higher expression of total MAP2K4 mRNA in blood, compared to A-allele homozygous. We also describe a novel, putatively translated RNA splice form of MAP2K4, that is differentially expressed in peripheral blood mononuclear cells from 88 RA cases and controls, and is modulated in response to TNF in Jurkat cell line.

Finally, we performed an expression analysis of multiple validated RA risk loci, and pathway analysis to assess functional relationship between RA susceptibility genes and predict new potential study candidates. New candidate molecules suggested by the pathway analysis, genes ERBB2 and HSPB1, as well as HLA-DRB1, were differentially expressed between RA patients and healthy individuals in RNA-seq data. ERBB2 expression profile was similar in whole blood of both treated and untreated patients compared to healthy individuals. A similar expression profile was replicated for ERBB2 in PBMCs in an independent material.

In this work, we approached the task of elucidating the functional aspects of genetic susceptibility of RA, by integrating genetic epidemiology, transcriptomics, proteomics, cell-models, and bioinformatics. We maintain, that such integrative approach provides the rationale to prioritize genes and genetic events for further functional studies. Our findings also outline the need to consider potential clinical significance of alternative splicing in gene expression studies.

CONTENTS

LIST OF SCIENTIFIC PAPERS

I. Ronninger, M., Guo, Y. J., **Shchetynsky, K.,** Hill, A., Khademi, M., Olsson, T., Reddy, P. S., Seddighzadeh, M., Clark, J. D., Lin, L. L., O'Toole, M., & Padyukov, L. (2012). The balance of expression of PTPN22 splice forms is significantly different in rheumatoid arthritis patients compared with controls. *Genome Medicine, 4*.

II. Houtman, M., **Shchetynsky, K.,** Catrina, A., Padyukov, L. Strengthening evidence from association studies: PTPN2 genetic variations interact with *HLA-DRB1* shared epitope alleles and *PTPN2* is differentially expressed in rheumatoid arthritis

Manuscript

III. **Shchetynsky, K**., Protsyuk, D., Ronninger, M., Diaz-Gallo, L. M., Klareskog, L., & Padyukov, L. (2015). Gene-gene interaction and RNA splicing profiles of MAP2K4 gene in rheumatoid arthritis. *Clin Immunol, 158*, 19-28.

IV. **Shchetynsky, K.,** Diaz-Gallo, L.M., Klareskog, L., Padyukov L. Expression pathway analysis for genes associated with rheumatoid arthritis

Manuscript

PUBLICATIONS NOT INCLUDED IN THE THESIS

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., **Shchetynsky, K**., Scheynius, A., Kere, J., Alfredsson, L., Klareskog, L., Ekstrom, T. J., & Feinberg, A. P. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology, 31*, 142-47.

Shchetynsky K., Padyukov L. Functional Studies of Gene–Gene Interaction of Autoimmune Diseases. In: Between the Lines of Genetic Code. Genetic Interactions in Understanding Disease and Complex Phenotypes, 2014, Elsevier, *Edited by Leonid Padyukov*, ISBN: 978-0-12-397017-6, *Pages 137-150.*

LIST OF ABBREVIATIONS

AP	Attributable proportion
ACPA	Anti-citrulinated protein antibodies
bp	Base pairs
CD	Cluster of differentiation
cDNA	Complimentary DNA
CI	Confidence interval
CRP	C-reactive protein
DC	Dendritic cell
DCIR	Dendritic cell immunoreceptor
DMARs	Disease-modifying antirheumatic drugs
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
EIRA	Epidemiological Investigation of Rheumatoid Arthritis
Erk	Extracellular signal-regulated kinase
ESR	Erythrocyte sedimentation rate
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GWAS	Genome-wide association study
HIFs	Hypoxia-inducible factors
HLA	Human leukocyte antigen
IFN	Interferon
IL	Interleukin
JNK	c-Jun N-terminal kinase
LD	Linkage disequilibrium
MAF	Minor allele frequency
MAP2K4	Mitogen-activated protein kinase kinase 4
MAPK	Mitogen activated protein kinase
MHC	Major histocompatibility complex
MMPs	Matrix metalloproteinases

mRNA	Messenger ribonucleic acid
NARAC	North-American Rheumatoid Arthritis Consortium
nt	Nucleotides
OR	Odds ratio
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PTP	Protein tyrosine phosphatase
PTPN2	Protein tyrosine- phosphatase, non-receptor type 2
PTPN22	Protein tyrosine phosphatase, non-receptor type 22
qPCR	Quantitative polymerase chain reaction
RA	Rheumatoid arthritis
RF	Rheumatoid factor
RNA	Disoxyribonucleic acid
RQ	Relative quantification
SE	Shared epitope
SNP	Single nucleotide polymorphism
snRNPs	Small nuclear ribonucleoproteins
STAT	Signal transducer and activator of transcription
TNF	Tumor necrosis factor
tRNA	Transport RNA
UTR	Untranslated region
VCAM	Vascular cell adhesion molecule
VEGF	Vascular endothelial growth factor

INTRODUCTION

In the following introduction, I want to provide an overview of relevant concepts in molecular genetics, genetic epidemiology, and rheumatoid arthritis research that served as a scientific context for my work.

1 TRANSCRIPTION

The absolute majority of living organisms, from unicellular to highly complex multicellular, pass the information to their offspring through DNA. DNA is a macromolecule with a sequence of four nucleotides which are used to store and reproduce numerous and diverse features of a living organism. According to the central dogma of molecular biology, DNA molecules need to be transcribed into RNA, which, in turn, gets translated into proteins – the main effectors in a cell.

Transcription of DNA-encoded information into RNA form starts with transcription factors binding to the regulatory elements of the genes. This causes recruitment of the RNA polymerase to the transcription initiation site. To initiate the transcription, transcriptional enhancers cause the conformational change of RNA polymerase and allow it to unwind a part of a double-stranded DNA helix in order to copy one of the strands into an mRNA molecule (messenger RNA). During the elongation stage, RNA polymerase proceeds alongside the DNA strand occasionally proof-reading the synthetized transcript. The newly synthetized mRNA promptly undergoes non-template related modifications followed by the recruitment of alternative splicing machinery and mRNA maturation. The process ends by the regulated release of the RNA polymerase. After that, tRNAs translate mRNA from the transcript matrix into amino acid sequence making the ultimate functional elements (proteins).

Canonically, a DNA motif responsible for a certain feature is called a gene and the collectivity of the genes define an individual species' genome. To crack the code that makes us unique, Human Genome Project set out in 1990 towards the goal of annotating the sequences of all human genes. It culminated in a historical publication in 2004, opening the human genome to the public. Those years were also marked by the emergence of the HapMap project - a large-scale effort to catalogue the differences and similarities of the human genetic code between individuals and world populations, providing invaluable insight

into genetic variation contributing to disease, response to treatment and environmental factors, etc.

However, one of the less expected revelations brought about by these monumental studies was the great similarity of the genetic code between individuals. Moreover, the number of human genes was found to be nearly identical to the one in mice, and very few protein families could be characterized as human-specific. These findings implied that there are additional levels of complexity in human genome, not necessarily explainable by the nucleotide sequence alone. To address this question, ENCODE (Encyclopedia of DNA Elements) project was launched in order to annotate all functional elements in the human genome by high-throughput methods. The initial result based on RNA and DNA sequencing, combined with chromatin immunoprecipitation (ChiP-seq) and deoxyribonuclease-based (DNase) footprinting demonstrated that transcription is not limited to protein-coding genes and is more universal than was thought before. Numerous novel transcripts (spliced and unspliced) were found both in protein-coding and previously considered as transcriptionally silent loci. This broadens both the central dogma and the classical gene definition and implies the existence of a complex multi-dimensional regulatory network determining the phenotype. It has also fast become clear that new types of non-protein coding RNAs are produced from relatively disregarded parts of the genome previously labeled as “gene deserts” and “junk DNA”. These were shown to play an important role in the embryo development as well as in the pathogenesis of different diseases.

The multifaceted functionality of the transcriptome presents both exciting opportunities to explain the complexity of biological process, but also challenges us with a whole new level of complexity.

1.1 RNA SPLICING MECHANISMS

The process of mRNA splicing takes place between transcription and translation. The transcribed pre-mRNA still contains non-protein coding elements – introns – that need to be removed - “spliced-out” - to produce mature mRNA, containing only the sequence necessary for protein construction (Figure 1). Splicing is dependent on splice sites - the sequence motifs, recognized by the specialized organelle, called the spliceosome.

Spliceosome components - small nuclear ribonucleoproteins (snRNPs) – recognize the splice sites, generally represented by GU sequence at the intronic 5’ end and AG sequence at the 3’

end. Another motif, that guides splicing is the so-called "branch point", located in the proximity of the 3' end of the intron.

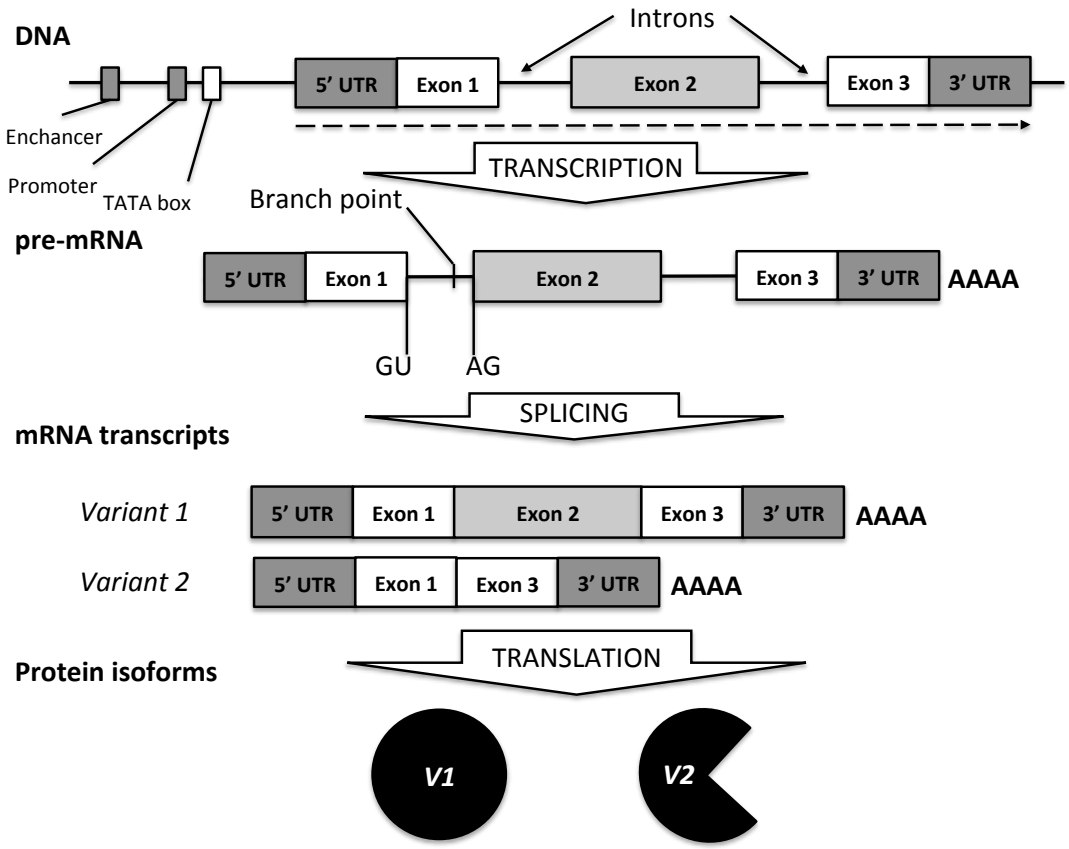

Figure 1. Schematic representation of the mRNA-maturation processes.

Splicing can be influenced by other regulatory elements in the genome, such as enhancers and silencers. Importantly, variations in the genome, such as deletions, insertions, and single nucleotide polymorphisms (SNPs), can modify the structure of these elements, causing changes in the regulation of the transcriptome.

1.2 ALTERNATIVE SPLICING

Alternative splicing is a process that makes possible production of different mRNA products from a single gene and ultimately – different proteins from the same gene. The phenomenon of alternative splicing was described almost simultaneously with the discovery of splicing itself (1). Generally, carrying a large number of introns that may then be utilized to achieve different combinations of exons in the final mRNA product may have some evolutionary advantage. The changes introduced to a translated protein product by alternative splicing could be drastic compared to single amino acid substitutions, resulting from non-synonymous SNPs. By regulating the functional domain composition of the final product, alternative splicing could affect its ability to carry out its function or form complexes with other proteins.

As transcription starts, splicing machinery gets activated and the primary transcripts are being spliced as soon as the splicing sites become available (2). The process is mediated by snRNPs U1, U2, U4, U5, U6 and by a number of complementary proteins like U2AF65 and U2AF35. The catalytically active spliceosome, regulated by altered protein-protein interaction as well as by RNA-dependent ATPases and helicases, performs a two-step splicing reaction (5' of the intron first, followed by the 3') leading to intron excision (3, 4).

The pre-mRNA regions at the beginning and at the end of introns contain a specific consensus sequence recognized by a spliceosome. The splice sites can be strong or weak depending on their affinity to a spliceosome determined by the sequence deviations from the consensus sequence. Strong sites are almost always used, whereas the usage of weak sites is determined by additional regulation thus providing the basis for the alternative splicing. The regulators of weak sites include *cis*-regulatory sequences (both intronic and exonic, both splicing enhancers and splicing silencers). *Trans*-regulatory factors can be universal (e.g., heterogeneous RNPs) or tissue-specific (NOVA, FOX, PTB, etc.)(5-8) and they can bind to the regulatory sequences thus providing additional level of splicing regulation. In addition, relative positioning of strong and weak sites, the rate of transcription (9, 10), small RNAs (11) also regulate this process what taken together with other factors, create vast possibilities for alternative splicing.

Indeed, alternative splicing appears in eukaryotes and is more pervasive in more complex organisms, hinting on its protective and regulatory roles in the homeostasis (e.g., coding larger protein variety by a genome, extensive regulatory opportunities, protection from mutations, etc.) (12, 13). In humans, alternative splicing is perceived to be extremely common, as about 95% of all human genes are thought to have multiple isoforms (13, 14),

whereas the corresponding value in *C.elegans* is estimated to be 13-25% (15, 16). As the gene number is approximately similar in humans and *C.elegans,* it is tempting to speculate that alternative splicing reflects and may largely contribute to the complexity of organization.

Notably, different tissues of the same organism may have vastly different transcript profiles, raising important questions about the involvement of alternative splicing in cell-differentiation and regulation of cell-type specific functions.

1.3 ALTERNATIVE SPLICING: CLINICAL SIGNIFICANCE

Given the abundance and tight regulation of mRNA splicing, it is reasonable to assume that it can become deregulated. Indeed, it has been demonstrated for multiple pathological conditions that alternatively spliced isoforms might be of importance (reviewed in(17, 18)). Mutations in core splicing machinery are rare implying that they might not be compatible with life due to extreme importance of splicing in all cellular processes. On the other hand, alternatively spliced isoforms can be tolerated by an organism, but might be the cause or the consequence of certain diseases. Therefore, the most frequent deregulations in alternative splicing include mutations in the splicing sites, thus promoting generation of novel isoforms (with diverse functions) and impacting the isoform balance by exon skipping, using a pseudo splicing site or retaining the mutated intron (mutation affecting in *cis*)(19, 20). Also, mice experiments demonstrate that splicing can be altered in *trans* by mutations in the regulatory factors (e.g., mutation in NOVA-1(21)). The role of the alternatively splicing deregulation in autoimmunity is being constantly addressed shedding light on additional details of disease mechanisms (exemplified in(22-24) and in the papers included in this thesis). The accumulation of data on alternative splicing also suggests that it could have significant contribution in development of RA by acting on multiple immunological pathways. For instance, alternatively spliced forms of CD44, CXCL12, hRasGRP4 were shown to effect the invasive capacity of synovyocites, T-cell activation, and cell signaling, respectively (25-27). Interestingly, a separate study comparing RA and osteoarthritis patients found that naturally occurring IL-32γ has an alternative protein isoform IL-32β, which is a less potent proinflammatory mediator and thus may be a safety switch in controlling the effects of IL-32γ and thereby decrease chronic inflammation (28). Finally, the study of Jin et al. shows that expression of human VEGF receptor type 1 splicing isoform in murine joints ameliorate collagen-induced arthritis (29).

Interestingly, most of the mutations causing phenotype changes are single nucleotide interchanges or SNPs, what gives another evidence of the importance and potential implications of alternative splicing. The rapidly accumulating knowledge on the involvement of alternative splicing in human diseases opens new therapeutic and prognostic opportunities.

2 GENETIC RISK

2.1 MENDELIAN AND POLYGENIC DISORDERS

Classical experiments of Gregor Mendel with *Pisum sativum* carried out on the backyard of the church during prayer-free time made a basis for the laws of inheritance and became the beginning of modern genetics. Currently, most of human diseases are considered to be genetic to some extend, even those which were believed to be due to purely environmental factors. Some disorders occur due to a rare genetic variant of a single gene and are, therefore, called single-gene disorders (e.g., sickle-cell anemia, Huntington's disease, etc., about 4000 disorders in total). These diseases have a classical inheritance pattern according to Mendelian laws and therefore, knowing the variation, it is possible to identify individuals at risk and to predict the occurrence of the disease.

Most of human disorders are polygenic (complex) meaning that multiple genes as well as environmental factors and a lifestyle contribute to their pathogenesis. Although in some cases (e.g., RA, asthma, pancreatic cancer, obesity) there is a clear clustering in the families, it is not clear which genes are involved, what genetic variant are important and, hence, which individuals are at risk. Moreover, we might not always have a complete understanding of how much inheritance plays role in disease development. The involvement of the environmental factors is equally important but more difficult to detect in studies of complex diseases. Genotype-phenotype association studies are carried out in order to decipher genetic and environmental components of complex diseases for potential risk prediction and for development of early interventions.

2.2 SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

SNPs are the variations of DNA sequence commonly occurring in the general population. Basically, a SNP is a mutation (interchange) of a single nucleotide in a particular DNA locus. These different nucleotide variants at a given position are referred to as different alleles, with the vast majority of common SNPs having just two alleles. In 2014, the number of validated SNPs in humans, registered in RefSeq reached 41,735,528 (30, 31).

SNPs in the exonic regions can be either synonymous (when the nucleotide alteration does not cause the change of the amino acid sequence) or non-synonymous (when a substitution in

the polypeptide change occurs). Non-synonymous SNPs often cause phenotype change and might be clearly associated with a certain disease, whereas the role of synonymous SNPs is usually more difficult to assess. Genetic variants in the promoter region of genes may affect transcription activation or interfere with silencing elements that ultimately will change activity of transcription. SNPs within introns may influence splicing elements and affect splicing mechanisms. Additionally, SNPs may influence mRNA stability, conformation of mRNA, change level of methylation and associated with miRNA function (32).

The distribution of SNPs throughout the genome is not uniform. The most common regions of SNPs occurrence are non-translated regions (both 3'- and 5'-UTRs, as well as intergenic regions) where the selective pressure has not eliminated variations throughout the evolution. The latter were previously believed to be of no to little impact on the transcription and gene expression (33). However, it is becoming increasingly clear that non-translated regions may play a substantial regulatory role in gene expression.

Technology development made possible genome-wide investigation of SNPs and comparison of the SNP profiles between large groups of individuals.

2.3 GWAS

Sequencing of a human genome as well as annotating a human haplotype facilitated development and near-ubiquitous use of genome-wide studies in complex genetic diseases. These studies usually compare two or more groups of individuals with different phenotypes and aim at identification of specific genomic features associated with each of the conditions (case-control studies).

Genome-wide association study (GWAS) examines common variants (SNPs) across the human genome. The array includes a large number of SNPs, thus allowing a hypothesis-free investigation of the allele frequency across the groups. To facilitate the analysis of large numbers of SNPs, GWAS make use of linkage disequilibrium (LD) in the genome. LD is an association between alleles from different loci that were inherited as a continuous sequence of DNA without recombination and that would not be expected to differ from their individual allele frequencies by random sampling of SNPs. Practically, this means that testing all known SNPs would be redundant, since it is possible to use a "tagging" SNP to represent an LD block, and to infer the associated genotypes of other SNPs in this block. This approach allows high cost-effectiveness of genotyping and analysis (34).

The variants with higher frequency in a disease group are referred to as disease-associated variants and odds ratios are normally used to present effect for a certain group of individuals carrying genetic variant (35).

The first large GWAS published in 2007 by Wellcome Trust Case Control Consortium genotyped 14,000 cases of seven common diseases (including 2,000 RA patients) and 3,000 controls (36). This study identified novel genes involved into the pathogenesis of several common diseases. In case of RA, it confirmed the involvement of HLA-DRB1 and PTPN22 variants, but also unveiled the involvement of other SNPs. For example, SNP rs11761231 located in chromosome 7 was found to be associated with RA in women whereas carrying no association in men. Also, SNPs in IL2RA (rs2104286), IL2B (rs743777), as well as tumor necrosis factor activating protein 2 *TNFAIP2* (rs2771369) were described as having association to RA.

For RA, approximately 50 percent of heritability remains unexplained, and a subject of debate. Although GWAS helps to get insights into previously underpinned genetic factors, for most of the associated variants outside of HLA-DRB1 and PTPN22 loci, the odd ratios are about 1.3. One of the implications may be that common genetic variants most often play rather small role in explaining heritability of RA and rare variants, which might play a larger role in the disease, are more difficult to identify. A study of 500 RA cases and 650 controls, discovered rare disease-associated variants in within the protein-coding portion of *IL2RA*, *IL2RB and CD2* – loci, previously identified by GWAS as contributing to the risk of RA (37). A recent report indicates an enrichment for rare/low-frequency variants associated with RA in the loci responsible for mitochondria respiratory chain-related proteins, providing a link between reactive oxygen species and RA pathogenesis (38). Conversely, other studies of rare variance in autoimmune disease suggest that unexplained heritability for common autoimmune diseases may be a result of multiple common-variant loci of weak effect, rather then the rare coding-region variant portion of the allelic spectrum (39, 40). In order to address these issues, on one hand, the cohort sizes were increased, and, on the other hand, a more carefully chosen material was genotyped for a smaller subset of SNPs. However, even very significant and carefully controlled for confounders associations only imply the involvement of a given SNP into a disease, but functional studies are needed to investigate the biological mechanisms behind the associations.

2.4 GENE-GENE AND GENE-ENVIRONMENT INTERACTION

In complex diseases, the interactions between genes and environment play a definitive role. Naturally, the disease association of SNPs from different loci does not provide evidence of gene-gene interaction. However, many methods for calculating gene-gene interaction exist, (e.g. additive and multiplicative interaction, multifactor dimensionality reduction, genetic vectors, and network biology approaches) that could directly utilize genotype information from GWAS studies. If the risk of developing a condition in a carrier of both susceptibility genotypes (as defined by a relevant dominance model) is significantly higher than a combination of their individual risk effects would imply, we can assume interaction between the two (41). This is referred to as the additive model of interaction. The percentage of the additional risk that cannot be explained by the sum of individual risks can be attributed to interaction effect size (attributable proportion of risk). The calculation relies on the logistic regression model that includes three groups: individuals carrying both genetic risk factors and those, who carry either one of the two (while the unexposed group is assigned the baseline risk of 1) (42). The relative excessive risk due to interaction (RERI) is defined as:

$$RERI=RR_{11} - RR_{10} - RR_{01}+1;$$

Where RR_{11} is the relative risk in a group, exposed to both risk factors, while RR_{10} and RR_{11} represent relative risk in groups exposed to either risk factor 1 or 2. The attributable proportion (AP) is then defined as:

$$AP= RERI/ RR_{11}$$

The development of high-throughput computational-based methods around this approach allows for pairwise calculation of interactions between large numbers of genetic risk factors in GWAS data (43, 44).

Notably, approaches based on multifactor dimensionality reduction (MDR) (45, 46) have seen an increase in usage for assessment of gene-gene and gene-environment interactions (46).

A classification of multi-locus genotypes into high-risk and low-risk classes makes it possible to reduce the dimensionality of genomic risk predictors to 1. The result modifies the contingency table in a way that facilitated the computation of statistics. The multiple modifications of the MDR have been used for resolving gene-gene interaction between specific genes in several disorders. In its current state, MDR-based interaction analysis can be

extremely taxing computationally, when employed on a whole-genome scale, warranting an optimization of the existing algorithms (47).

Although, interaction analysis has proven to be a useful tool in assessing potential functional connecting in existing GWAS data, it is still not representative of the underlying biology, and extensive functional studies are warranted to elucidate the statistical connection (48).

3 RHEUMATOID ARTHRITIS

Rheumatoid arthritis (RA) is a debilitating common complex autoimmune disease characterized by persistent inflammation of the joint, resulting in gradual erosion of the synovial tissue by host immune cells, leading to pain and dysfunction. It is a systemic disease, and can affect any joint in the body. Characteristically for autoimmunity in general, the disease pathogenesis is propagated by the host adaptive immune system. However, the factor that initially tips the host immune system out of balance is still unresolved. Several genetic and environmental factors have been described to contribute to susceptibility, making RA a disease with complex etiology.

3.1 CLINICAL MANIFESTATION OF RA

RA is a complex disease which includes both joint-specific and more general systemic symptoms. Joint-specific manifestations include swelling, pain, morning stiffness (which lasts for more than 2h) and limited motion in the joint. In principal, RA might affect any of the joints in the body, but smaller joints are usually more frequently affected. The long-lasting morning stiffness is a special characteristic of RA which allows clinicians to suspect that the patient presents with RA rather than other arthritic diseases (for instance, osteoarthritis has similar symptoms with regards to pain and swelling, but prolonged morning stiffness is not usually present) (49). American College of Rheumatology first developed and later reconsidered the score-based algorithm for the diagnosis of RA. Below the clinical parameters for the diagnostics of RA are presented according to 2010 ACR-EULAR classification criteria (50). Importantly, this update introduced serological tests for anti-citrullinated peptide antibodies (ACPA) and rheumatoid factor (RF) as an integral part of differential diagnosis for RA.

Target Population

1. patients, having at least 1 joint with definite clinical synovitis (swelling)
2. synovitis cannot be explaind by another disease

Classification criteria for RA

A. Joint involvement	**Score**
1 large joint	0
2-10 large joints	1
1-3 small joints (with or without invovement of large joints)	2
4-10 small joints (with or without invovement of large joints)	3
>10 joints (at least 1 small joint)	5
Serology (at least 1 test result is needed for classification)	
Negative RF *and* negative ACPA	0
Low-positive RF *or* low-positive ACPA	2
High-positive RF *or* high-positive ACPA	3
C. Acute phase reactants (at least 1 test is needed for classification)	
Normal CRP *and* normal ESR	0
Abnormal CRP *or* abnormal ESR	1
D. Duration of symptoms	
<6 weeks	0
≥6 weeks	1

Table 1. The 2010 ACR-EULAR classification criteria for RA. Large joints refer to shoulders, elbows, hips, knees, and ankles. The duration of the symptoms is self-reported by patients. RF - rheumatoid factor; CRP - C-reactive protein; ESR - erythrocyte sedimentation rate.

To be diagnosed with RA, a patient should have a score of ≥ 6/10. The patients with the lower score or with borderline cases should be followed-up, since the criteria might be fulfilled as the disease progresses.

More general symptoms of RA include fatigue/loss of energy, low fever, loss of appetite, firm nodules growing underneath the skin. Although not necessary for the diagnosis, these symptoms greatly contribute to the debilitating nature of RA.

3.2 PATHOGENESIS

Although our knowledge about RA has greatly advanced during the last thirty years, the trigger of the disease remains unknown. The pathogenesis of the disease, however, is better understood which made possible the development of the symptomatic therapy.

It is believed that before the disease onset, an individual is exposed to certain environmental factors which "prepare the ground" for further events leading to RA (described in more detail in the corresponding section of this thesis). These environmental factors are not specific for RA and are normally managed well by the general population. However, some individuals are more predisposed to the development of autoimmunity due to their genetic background (MHC class II genes, PTPN22, polymorphisms in cytokine promoters and other yet unidentified genes; described in the section "Genetic factors of RA"). Even long before the symptoms occur, RF and ACPA can be detected in the blood as evidence for autoimmune reactions (51, 52).

For the reasons which are not yet completely understood, the synovial tissue is primarily targeted in RA. The potential antigens include citrullinated peptides, immunoglobulins, heat-shock proteins as well as cartilage antigens (e.g., type II collagen, aggrecan, proteoglycans, etc.), however, there is no evidence of a specific rheumatoid antigen. Antigens are processed by dendritic cells, which then present peptides to naïve T-cells in the local and central lymphoid organs. In turn, T-cells activate B-cells to produce pathogenic antibodies and/or to move to the synovial tissue and engage into a complex interaction network with other cells present on the site. Once this happens, the antigen is not necessary anymore and the cells cross-activate each other through non-specific cell-to-cell communication mechanisms (53). This leads to the beginning of a destructive process involving pro-inflammatory cytokine production, cartilage destruction by matrix metalloproteinases and increased levels of reactive oxygen and nitrogen. The organism's inflammation-limiting capacity is not capable to dampen this process through the production of suppressive cytokines. Further, synovial intimal hyperplasia occurs which is associated with deficient apoptosis, deregulation of a key tumor suppressor p53 and even upregulation of pro-angiogenic factors (e.g. HIF-1α, VEGF, soluble VCAM) (54).

Clearly, multiple cellular types of the synovium participate in the pathogenesis of RA. These include macrophage-like type A synoviocytes, fibroblast-like type B synoviocytes (the absolute number of both these subpopulations is increased in RA, skewed towards type A), highly activated macrophages, lymphocytes which can actively infiltrate or form aggregates, CD4+ T-cells which have memory phenotype, antigen-induced and antibody-producing synovial B-cells, mast cells producing large number of mediators of inflammation, dendritic cells, etc (55). Apart from that, it has been demonstrated that mesenchymal cells from the bone marrow can migrate and reside in the synovial tissue, interact with other cells and contribute to the development of synovitis (56).

Also, peripheral lymphocytes of RA patients are notably different from those of healthy individuals. Firstly, the counts of CD4+ helper T-cells are drastically increased in circulation (whereas CD8+ T-cells are accordingly decreased), and the cells show the features of activation (57). B-cells capable of producing RF are also significantly elevated in the peripheral blood and synovial tissue of RA patients (58, 59). Whether these cells escape from the synovial site of inflammation or they are produced in the lymphoid organs in response to antigens is not clear, but it is recognized that the peripheral blood profile is different from that of healthy individuals (60). These changes make it possible to use peripheral blood of RA patients for diagnostic, as well as prognostic studies (61).

The repertoire of the cells considered for the development and progression of RA is wide and new probable players are continuously added to the list. The description provided here is far from being complete, but identifies some key cell types.

3.3 CYTOKINES IN RA

Cytokines are small proteins allowing communication between different cell types of the immune system. Each cell type is capable of producing a certain set of cytokines which are regulated by a certain set of transcription factors. Since multiple types of immune cells are implicated into the pathogenesis of RA, abnormal cytokine production is also a marked characteristic of RA. Th1 cells mainly produce IFNγ which is a potent inducer of MHC class II. IL-2, another Th1-produced cytokines, serves as a growth factor for T-cells. Apart from that, TNF, GM-CSF and IL-6 can be produced by both Th1 and Th2 cells (controlled by STAT4 in the former case and by GATA-3 and STAT6 transcription factors in the latter case) (62). Th17 cells produce different isoforms of IL-17 under control of the transcription factor STAT3 (63). The main part of massive cytokine burden comes from activated macrophages and fibroblasts, which produce a wide range of cytokines (e.g., IL-1, IL-6, IL-10, IL-15, IL-18, TNF, RANKL, etc).

TNF has been defined as one of the key pro-inflammatory cytokines in RA (64). It can be detected both in the synovial tissue and in the serum of RA patients which led to the development of targeted anti-TNF therapy (65). It is a pleiotropic cytokine produced mainly by monocytes in a membrane-bound form and being cleaved by a specific MMP. Through binding to a set of own receptors, TNF is able to regulate the production of other pro-inflammatory cytokines, as well as to increase the expression of adhesion molecules and MMPs in synoviocytes (66). The success of anti-TNF therapy underlines its crucial role in

RA and, on the other hand, points to the heterogeneity of the disease, as ⅓ of RA patients do not respond to the therapy (67). Also, continuous therapy is needed to reach a clinically noticeable improvement which suggests that TNF is an intermediate, although crucial, player in the pathogenesis of RA, and the main players are yet to be determined.

3.4 CURRENT TREATMENT STRATEGIES

RA is a chronic disease, for which there is currently no cure. Partially, the reason for that is that our knowledge about the causative agent for RA still limited and therefore, it cannot be targeted. Given that, the current treatment aims at symptom relieve and slowing down the disease progression. The early stage of RA (usually 3 months to 2 years after the first symptoms appear) is considered to be "a window of opportunity" meaning that application of the therapy on this stage can halt or significantly slow down the disease progression (but not to revert the disease).

Disease-modifying anti-rheumatic drugs (DMARDs) are a group of otherwise unrelated classes of medicines which are normally prescribed as soon as RA is diagnosed. The mechanisms of their action are not always completely understood, but they all exhibit anti-inflammatory effects. Methotrexate is often a drug of choice for the initial treatment, but other drugs (e.g., leflunomide, azathioprine, sulfasalazine etc.) are also available and can be combined to reach the clinical effect and to suit the individual toleration. The latter is of extreme importance due to usually a life-long prescription of DMARDs. DMARDs can be safely combined with NSAIDs, that provide analgesic effect and can be effective in reducing stiffness (68). Profound inflammation is often treated with short courses of cortisone and other cortisone-like medicines (69)

If DMARDs do not suit the patient or stop working eventually, biological therapies are commonly applied. The first biological therapeutics developed for the use in RA were TNF receptor and anti-TNF antibodies which revolutionized the rheumatology field (70). Other biological therapies (IL6 blockers, removers of B- and T-cells) are also available (71, 72). Also, a number of clinical trials for other drugs (both biologicals and non-biologicals) are currently under way (73-75). All of them are directed towards dampening inflammation.

If medications fail to help a patient, joint surgery of different level of invasiveness is being applied.

To summarize, a number of relatively effective medications are available for the treatment of RA patients. However, medication withdrawal after the remission has not been proven beneficial for the patients, making RA a life-long condition that requires constant and expensive management. Therefore, search for causative agents must be intensified to develop a potentially curative treatment.

3.5 GENETIC RISK FACTORS FOR RA

The results from a number of twin studies characterize RA as a disease with relatively high heritability of around 60%, hinting at a pronounced genetic component of the disease (76), however other studies give lower estimates, especially for ACPA-negative RA (77). A number of known genetic susceptibility loci for RA has greatly increased with the advance of genome-wide association studies. However, the most considerable risk of RA development stems from HLA-DRB1 locus. The higher prevalence of RA among females suggests that gender-related genes may also play a role in susceptibility. Close to 50% of RA heritability cannot be currently explained by known genetic associations (77).

Below some of the common genes associated with the RA risk are described in more detail.

HLA-DRB1

HLA genes code for major histocompatibility complex molecules (MHC) which play an important role in the immune system of vertebrates. There are two distinct classes of HLA (I and II), each of these coded by specific gene regions located in chromosome 6. The function of HLA class I and HLA class II is to present the processed antigens to the immune cells and to modulate the interaction between the immune cells. Additionally, the area between HLA class I and II, which is sometimes denoted as HLA class III, represents a heterogeneous set of molecules involved into immune responses.

In the context of autoimmunity and RA, the role of HLA class II is one of the most influential findings to date. It was recognized before genome-wide arrays became available and still remains the strongest genetic factor in RA (78). The region is comprised of HLA-DR, HLA-DP and HLA-DQ loci, encoding α- and β-chains of HLA class II molecules which are expressed as heterodimers on the antigen-presenting cells. When a naïve T-cell encounters an antigen coupled to HLA, its TCR is primed by a peptide-HLA II complex and a cell gets polarized according to the cellular context. Generally, HLA class II were found to be upregulated in the synovial tissue of RA patients. Initially, genotyping and later sequencing

of HLA-DRB1 (encoding polymorphic β-chain of the DR molecule) indicated that different variants of this molecule differ between populations. This led to proposing a hypothesis of shared epitope (SE), which points out the importance of a highly conserved region in the amino acid residues 70-74 of the third hypervariable β-chain (i.e., QKRAA/QRRAA/RRRAA) (79). However, it was later demonstrated that SE is mainly associated with ACPA-positive RA implying a role of SE in the presentation of citrullinated peptides. Mechanistically, crystallographic analysis showed that the risk-associated variant creates a positively charged groove that binds negatively charged citrullinated antigens. Conversely, the non-associated variant makes a negatively charged pocket of DRB1, thus preventing the interaction with the citrullinated peptides (80).

Although being in the strongest association, HLA-DRB1/SE explains only a fraction of ACPA-positive RA cases. Additionally, HLA-DRB1 locus may not be the only one responsible for susceptibility effects. A study of 1,352 patients with RA and 922 healthy individuals demonstrated that DRB1*13 allele plays a dual role in the development of RA, by protecting against ACPA-positive RA but, in combination with DRB1*03, increasing the risk of ACPA-negative RA. (81)

Later studies demonstrated that the amino acid valine at position 11 at HLA-DRB1 gene has highest association with RA, while amino acids lysine and alanine within SE sequence (positions 71 and 74) have an independent effect on risk (82).

The relative impact of each mechanism and the involvement of other genes are still being investigated.

PTPN22

Being the largest risk factor for RA, outside of HLA region, a SNP in PTPN22 (a gene coding for lymphoid phosphotase LYP) is also an important contributor to pathogenesis in other immune diseases (83). A single SNP - R620W (rs2476601) was initially discovered to be in association with type 1 diabetes (T1D), systemic lupus erythematosus, and later as a risk genotype in multiple autoimmune diseases (reviewed in (84).

Structurally, the disease genotype is a substitution of arginine (R) to tryptophan (W) in the non-catalytic part of LYP. The current understanding of the functional impact of R620W is still incomplete, with some reports of gain of function, as well as a loss of function, in lymphocytes (85). Mouse models demonstrated an increase in germinal center activity and serum levels of immunoglobulins in both PTPN22 knockouts and mice expressing the risk

allele (86). Increased germinal center activity could potentially further exacerbate the autoimmune condition driven by other genetic loci. Additionally, the presence of the R620W allele could influence the substrate specificity of LYP, possibly resulting in a loss of function with some substrates and a gain of function with others.

PTPN2

PTPN2 is a comparatively new entry in the list of genetic risk factors for RA. Genetic variants in this gene have been reported to be associated with RA, juvenile idiopathic arthritis, and Crohn's disease in several studies (36). Intronic PTPN2 SNP rs16939895 has been shown to be associated with Crohn's disease and celiac disease (87).

PTPN2 gene codes for an enzyme tyrosine-protein phosphatase non-receptor type 2, a member of the protein tyrosine kinases (PTP) superfamily. Two alternatively spliced isoforms were identified: TC45 (which contains a nuclear localization signal) and TC48 (located primarily in the endoplasmic reticulum) and the existence of several more spliced isoforms was predicted (88). The isoforms share the crucial catalytic domain, but the TC45 lacks C-terminal domain that was found to be inhibitory for the phosphatase activity. TC45 has been demonstrated to have various targets involved into the pathogenesis of RA (JAKs, STAT1, STAT4, STAT3, p42/44, ERK) (89-91). PTPN2 is assumed to participate in the regulation of TNF as well as IFNγ signaling hinting on its role in autoimmunity. Several studies addressed this change of function and have indeed found evidences for immune regulatory effects, suggesting that cells with the disease associated allele may have a gain of function for PTPN2 resulting in a stronger negative regulation of T cell activation and B cell activation (92, 93). However, the exact mechanisms by which PTPN2 influences RA development are not completely understood.

PADI4

PADI4 represents the peptidyl argenine deaminase enzyme, responsible for the posttranslational modification of argenine into citruline. Initially discovered as an RA susceptibility locus in Asian population (94), PADI4 has a stronger effect on RA risk in the individuals of Asian, rather than European ethnicity (95). It is speculated, that gene-environment interaction (namely, the much higher prevalence of smoking in East Asian males, compared to Europeans) could account for inter-population discrepancies in effect size (96).

TNFAIP3

The tumor necrosis factor-α-induced protein 3 (TNFAIP3) locus has been found associated with the increase in risk for both RA (97, 98) and systemic lupus erythematosus (99, 100). One of the known functional variants is a missense substitution of phenylalanine with cysteine at amino acid position 127, impairing function of A20 - the ubiquitin-editing enzyme (100). The other associated variant is a TT>A polymorphic dinucleotide located downstream of the *TNFAIP3* promoter that results in reduced *TNFAIP3* expression (101). A20 is known to be inactivated by somatic mutations in B-lineage lymphomas (102), and similar events have been observed for other genes associated with lymphomas, such as *REL*, *FCRL3*, and *DDX6*, in which common variants are associated with RA (103), possibly pointing out the difference in underlying pathogenic mechanisms.

STAT4

The signal transducer and activator of transcription 4 (STAT4) was identified as a susceptibility locus for RA, SLE and Sjogren's syndrome (104, 105). The risk allele genotype is associated with the overexpression of STAT4, suggesting that the product of the gene may contribute to autoimmunity (106). It is possible that the major role of STAT4 in type I interferon signaling, as well as in differentiation of T-helper cells underlie the association of this locus with several autoimmune diseases.

MAP2K4

In our work, MAP2K4 (mitogen-activated protein kinase kinase 4) was discovered as an RA susceptibility locus in interaction with HLA-DRB1 (107). MAPKs have attracted considerable attention as potential targets for autoimmunity because they can alter the production of key inflammatory mediators (108). There are three major families of MAPKs—namely, p38, extracellular-regulated protein kinase (ERK) and c-Jun N-terminal kinase (JNK). MAPK activation is mediated by upstream MAPK kinases (MKKs or MAP2Ks), which in turn, are activated by MKK kinases (MKKK or MAP3K) (109). p38 activation and phosphorylation is regulated by two upstream kinases, MKK3, MKK6 and, reportedly, MAP2K4, which are all phosphorylated by multiple upstream MAP3Ks. MAP2K4 (in tandem with MKK7) is also involved in co-activation of JNK, which has a regulatory effect on the expression of HLA class II genes and CIITA, directly related to antigen presentation (110). The p38-mediated signaling cascade culminates in increased expression of proinflammatory molecules like TNF, IL-6, IL-1, cyclo-oxygenase 2 (COX-2) and metalloproteinases (MMPs) (111).

Being among the major regulators of proliferation and differentiation of T- and B- cells, JNK and p38 were considered to be very promising drug targets for rheumatoid arthritis. However, clinical trials of direct inhibitors were largely disappointing with undesirable side effects. It has since been proposed, that targeting molecules involved in activation and regulation up-stream of JNK and p38, like MAP2K4, may allow for fine regulation of the activity of these kinases (112).

3.6 ENVIRONMENTAL FACTORS

Most of the known environmental factors for RA are described for the ACPA-positive subset of the disease.

ACPA-negative RA patients are a much smaller group, which makes it less studied. Additionally, the two major subgroups of RA differ profoundly in the gene expression profiles as well as in the possible influence of certain environmental factors(113).

Multiple epidemiological studies indicate that smoking is an environmental factor best associated with ACPA-positive RA(114-116). By controlling genetic background in twin studies, it was demonstrated that the smoker was much more prone to develop RA than the non-smoker. The risk of developing RA has been described to have a linear relationship with the intensity of cigarette use, being twice as high for a heavy smoker compared to a non-smoker. Interestingly, the risk of developing RA can remain elevated for decades after secession of smoking(117).

Another well-confirmed environmental factor responsible for higher RA risk is chronic exposure to crystalline silica (118), which can occur in many industrial settings, such as mining, construction work and electronics production. Similarly to smoking, long-term exposure to silica occurs primarily through inhalation, and can contribute to a two-fold increase in risk, after accounting for the confounding effects of smoking (119).

Among other occupational exposures that could contribute to the development of RA, extended contact with mineral oil was found to be associated with disease susceptibility in a study on Scandinavian population (120).

Alcohol consumption was shown to be associated with the protection from ACPA-positive RA in some epidemiological studies (121, 122). Protective effect of alcohol was found to be dose dependent, with individuals having 5 or more drinks per week being 40-50% less likely

to develop RA, compared to non-drinkers or individuals with low alcohol consumption. Alcohol administration was also correlated negatively with the disease severity in animal models of arthritis.

Some dietary aspects are suspected to be an influence in RA, since fasting and vegetarian diet were shown to decrease disease activity. However, some of the larger and more recent studies struggled to find an influence of consumed amounts of protein, poultry, red meat, and fish on RA risk, while others reported that a diet containing fatty fish could be protective, as was vitamin D intake (123).

Additionally, some data also point out the protective role of hormone replacement therapy in the post-menopausal period, suggesting the effect of estrogen (124). On the other hand, usage of oral contraceptives was not convincingly associated with RA onset or development.

Other possible factors include weight at birth, breastfeeding, occupation, socioeconomic status, region of birth, etc., however the exact mechanism of their influence remains yet to be determined.

3.7 GENE-GENE AND GENE-ENVIRONMENT INTERACTION IN RA

Several notable interactions between RA risk factors have been identified. Smoking has been well-established as a major environmental factor for RA, leading to a number of studies addressing smoking as factor in gene-environment interaction studies of RA (125-127). This led to a discovery of a pronounced interaction between smoking habits and the number of carried HLA-DRB1 alleles in regard to RA risk. Specifically, carriers of two HLA-DRB1 SE alleles, had a 21-fold increase (and a 40-fold increase in case of heavy smoking) in risk of developing ACPA-positive RA, compared to non-smokers, who do not have HLA-DRB1 risk alleles. Considering that individually, SE alleles and smoking account only for 5-fold and 2-fold increases in risk, respectively, it is very likely that gene-environment interaction plays a major role in RA etiology (128).

The gene-gene interaction between the 2 major genetic risk factors in RA – HLA-DRB1 SE and PTPN22 polymorphism R620W – was confirmed in three independent study materials from North America, the Netherlands, and Sweden (129, 130). Individuals, carrying at least one copy of each risk genotype were almost 10-fold more susceptible to RA compared to individuals, carrying none of the susceptible genotypes. Notably, in the absence of HLA-

DRB1 alleles, PTPN22 risk allele alone did not produce an increase in risk for ACPA-positive RA in carriers.

Finally, it was found that there is an environment-environment interaction between exposure to silica and smoking, leading to higher RA susceptibility (131). The risk of developing RA, is almost 8 times higher in silica exposed current smokers, while in non-smoking individuals, exposed to silica, it is around twice the risk of the unexposed.

4 CONCLUSIONS OF INTRODUCTION

This introduction served to convey the following points to provide context for my work:

Alternative splicing is an important and often overlooked mechanism that is an integral part in the regulation of many biological processes on both mRNA and protein level. There is multiple evidence that alternative splicing may consequently also play an important role in the development of human diseases.

GWA studies have provided an invaluable insight into the genetic component of various diseases by addressing disease-associated variants across individuals and populations. However, the functional implications of these associations are often insufficiently understood.

Studies of gene-gene and gene-environment interactions can greatly expand on the application of GWAS data in pinpointing functionally relevant candidate genes and signaling pathways.

RA is common autoimmune disease with complex etiology and debilitating inflammation of the joints. The disease is sub-divided by its serological profile, determined by the presence of autoantibodies to citrullinated peptides. A host of environmental and genetic factors play a role in the development of the disease.

Gene-gene and gene-environment interaction can potentially account for unexplained heritability and population effect-size differences of known susceptibility loci in RA.

5 STUDY DESIGN

5.1 AIMS

The general aim of this thesis was to bridge genetic association and functional processes behind RA pathogenesis, and to better understand functional role, which the genetic component plays in the development of the disease. While we specifically look at susceptibility loci outside the well-established HLA-DRB1, we strive to understand the cumulative effects and possible interplay of the latter with the former. We also maintain that addressing the balance of alternative mRNA products provides a more telling picture than bulk gene expression.

More specifically, in paper I we aimed to address the effect of the risk-associated polymorphism in PTPN22 on the expression of this gene, with regard to its known alternative mRNA isoforms. In paper II we continue to apply a similar approach to another member of the phosphatase family, PTPN2, which is also characterized by a polymorphism with a strong association to RA, while also addressing its potential interaction with HLA-DRB1. In paper III we explore the possibility of interaction of the members of the MAPK pathway with SE, and detailing clinical relevance of splicing in the gene MAP2K4. In paper IV, using RNA-seq data and pathway analysis, we attempt to establish new candidates for molecular studies of RA pathogenesis.

5.2 METHODS

5.2.1 Patient cohorts

In this work, the population-based EIRA (Swedish Epidemiological Investigation of Rheumatoid Arthritis) cohort was included with overall 2010 RA cases and 2280 control individuals. All cases met the American College of Rheumatology (ACR) 1987 criteria for RA(50). This cohort was built up from the incident cases of RA documented in the middle region of Sweden starting from year 1996. To each patient in the cohort, 1-2 healthy control individuals were matched based on gender, age and the residence area. All individuals participating in the study filled in a comprehensive questionnaire on their background, health issues, life style, etc (132).

The North American Rheumatoid Arthritis Consortium (NARAC) study comprised 873 ACPA positive RA cases and 1196 controls. The NARAC study included ACPA positive RA patients of self-reported white ancestry that were randomly drawn from four different groups of samples. All cases met the ACR 1987 criteria for RA(50). Control subjects from the New York Cancer Project were enrolled during a 2-year period by means of general advertising and point-of-service solicitation, as described previously(133). Information on phenotypic and genotypic data from NARAC was used by courtesy of Prof. Peter Gregersen from Feinstein Institute for Medical Research and Dr. Robert Plenge from Broad Institute.

Several sample materials were used for expression studies:

The RNA expression experiments utilized PBMC-samples from 44 RA patients and 44 controls, and an independent source of whole blood samples from 80 RA patients and 80 controls from the Swedish population; controls were selected with consideration to gender, age and ethnicity for the patient group. RA patients were selected at the Rheumatology Clinic at Karolinska University Hospital on two occasions, and all correspond to ACR 1987 criteria for RA. Information about anti-citrullinated protein antibody (ACPA) status, smoking habits and medication (for RA patients) was obtained from medical records.

In paper I, two additional independent cohort were used, (II) including 47 RA patients and 19 controls from Sweden, and 48 RA patients and 48 controls from the US respectively; PBMCs from 60 multiple sclerosis (MS) patients from the same geographical area as cohorts I and II were also used in the study (Cohort IV).

RNA-seq experiments in paper IV were performed on whole-blood RNA samples from 12 healthy female controls, 6 untreated (“naïve”) female RA patients, and 6 female RA patients receiving standard treatment with methotrexate. The replication cohort consisted of PBMC samples from an independent material of 48 untreated female RA patients, methotrexate treated female RA patients, and 35 healthy female controls.

Informed consent was obtained from all participants in all cohorts, in compliance with the latest version of the Helsinki Declaration. Regional ethical committees at all sites have approved the study.

5.2.2 Cell culture models

To address the effect of methotrexate on the expression of PTPN22 isoforms in Paper I, we used Jurkat (immortalized T-lymphocytes) and Daudi (B-lymphoblasts) cell lines. To assess the expression of PTPN22 in different cellular systems, the following cell lines were used in addition to the above mentioned: Raji (lymphoblast-like B-lymphocytes), U266, RPMI 8226 (both are multiple myeloma cell lines) and HEK293 (a transformed human embryonic kidney cell line).

In Paper III the following cell lines were used to address the expression of MAP2K4 isoforms: Jurkat, Raji and U937 (a monocyte cell line). Also, HEK293 cell line was used in this study for overexpression of alternatively spliced isoforms of MAP2K4.

All the cell lines were maintained according to the recommendations of ATCC.

5.2.3 Transcript discovery by polymerase-chain reaction (PCR)

To investigate the expression of novel alternatively spliced products, we first amplified the canonical isoforms using conventional PCR. PCR allows fast and profound amplification of a gene product existing in a cell in a very few copies into the quantities which are possible to visualize. It allowed us to assess whether using the primers for the canonical isoforms we were able to amplify other isoforms as well. After the preliminary assessment of potential transcript diversity, PCR product was inserted into a plasmid and transfected into a bacterial system. Screening by electrophoresis was used to identify transformants of interest, and the sequence of prospective alternative transcripts was confirmed by sequencing. The sequences were compared to the annotated isoforms with the use of bioinformatics tools.

5.2.4 Gene expression analysis

As a proxy of the gene expression, the abundance of the transcript (number of mRNA copies) is commonly measured. A modification of the conventional PCR described above has been made to allow for simultaneous amplification and quantification of the transcript (134). cDNA (complementary DNA) is first being generated from RNA using a viral RNA-dependent DNA polymerase (reverse transcriptase). Thereafter, there are two main ways to quantify the amount of a specific cDNA region, but both of them rely on using fluorophores and translation fluorescence intensity into relative quantities of cDNA (and hence, mRNA).

We used the variety of qRT-PCR based on a pair of template-specific primers and a template-specific probe coupled to a fluorescent reporter (Taqman assay). The probe-based design of TaqMan assays allowed for more precise targeting of specific exon-junctions, making it possible to quantify alternative transcripts from a single locus. In all the papers of this thesis, we used relative quantification method which relies on detecting the difference in the expression level between groups of samples (e.g., from patients and controls), with normalization against the expression of the reference gene within each group.

5.2.5 Gene-gene interaction analysis

In our work, we used additive model of interaction. To assess genetic interaction, relative excessive risk (departure from additivity) was calculated for the group of individuals, carrying both prospective risk markers, and represented as the attributable proportion of risk due to interaction (AP) with 95% CI. All SNPs were tested for interaction with SE alleles using the binary dominant/recessive model. Additional stratification criteria could be applied (e.g. presence of ACPA-autoantibodies). Gender was treated as a potential confounder. When testing multiple SNPs for interaction, we applied Bonferroni correction to p-values for AP in order to circumvent the multiple comparisons problem.

6 RESULTS

6.1 PAPER I: THE BALANCE OF EXPRESSION OF PTPN22 SPLICE FORMS IN RHEUMATOID ARTHRITIS

PTPN22 has long been known as the most important genetic risk factor in rheumatoid arthritis, outside of HLA locus. The PTPN22 gene encodes an 807 amino acid protein called lymphoid tyrosine phosphatase (LYP), which belongs to the proline-, glutamic acid-, serine-, and threonine-rich (PEST) group of non-receptor classical class I protein tyrosine phosphatases. An amino acid substitution, resulting from a SNP rs2476601, resulted in altered signaling in T and B cells, has been confirmed to be associated with RA in multiple populations. Importantly, PTPN22 is strongly implicated in a number of other autoimmune diseases, such as type 1 diabetes, systemic lupus erythematosus, juvenile idiopathic arthritis, and Graves' disease.

The current understanding of the functional impact of 620W is still incomplete, with some reports of gain of function (135, 136), as well as a loss of function (137), in lymphocytes. These discrepancies may be attributed to the difference in function between human PTPN22 and its murine analogue.

PTPN22 expresses several splice forms, but there is surprisingly little known about the function and regulation of these products (Figure 2). At least two of them are translated into proteins, LYP1 (PTPN22_v1) and LYP2 (PTPN22_v4)(138).

To evaluate expression of the PTPN22 splice forms and their possible relation to RA, we measured the expression levels of PTPN22-long isoforms and the short isoform PTPN22_v4, relative to a ZNF592 expression in PBMC from 44 RA patients and 44 controls from Sweden (Cohort I).

.

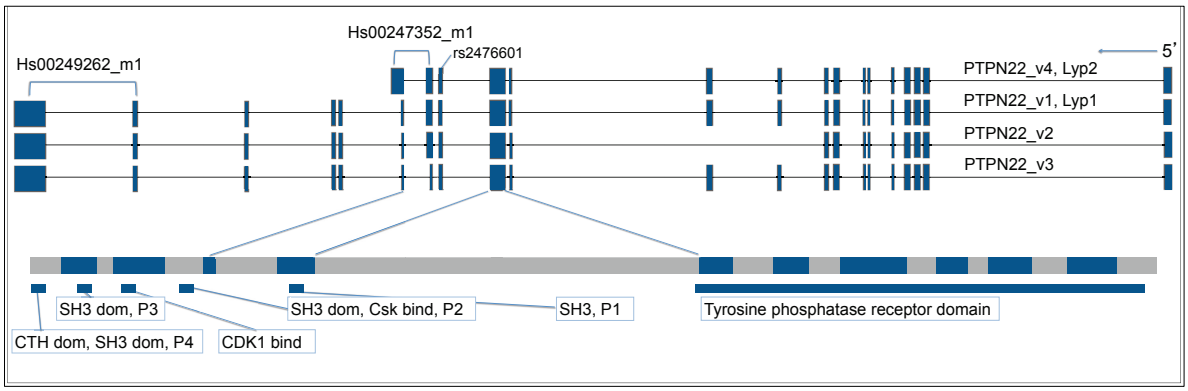

Figure 2. Exonic-intronic structure of splice forms and domain representation of PTPN22. A genomic illustration of PTPN22 in human chromosome 1p32. Hs00247352_m1 and Hs00249262_m1 are assay numbers (Applied Biosystems). The four SRC homology 3 (SH3) domains are denoted by P1 to P4 and are typically involved in protein-protein interaction. Known binding sites for carboxy-terminal Src kinase (Csk) and CD2-binding protein 1 (CD2BP1) are located in P2 and P4 (carboxy-terminal homology (CTH) domain), respectively. The carboxyl terminus of Lyp1 contains a consensus motif for cyclin-dependent kinase 1 (CDK1). The exonic and basic domain structure is deduced from Ensembl (Ensembl 64, September 2011) and UCSC Genome Browser (GRCh37/hg19 assembly).

By comparing the ratio of expression, we found that the profile of PTPN22 isoform expression in RA cases is significantly different compared with that in controls ($p = 6 \times 10^{-9}$, Mann-Whitney U test) (Figure 3 A), which is a result of significantly higher expression of PTPN22-long isoforms and simultaneously slightly lower expression of PTPN22_v4 in patients. We found that the average expression of PTPN22 splice forms in IFNγ-stimulated PBMCs from the same healthy individuals and RA patients was similar for RA patients and controls for the ratio between PTPN22_v4 and PTPN22-long forms (Figure 3c).

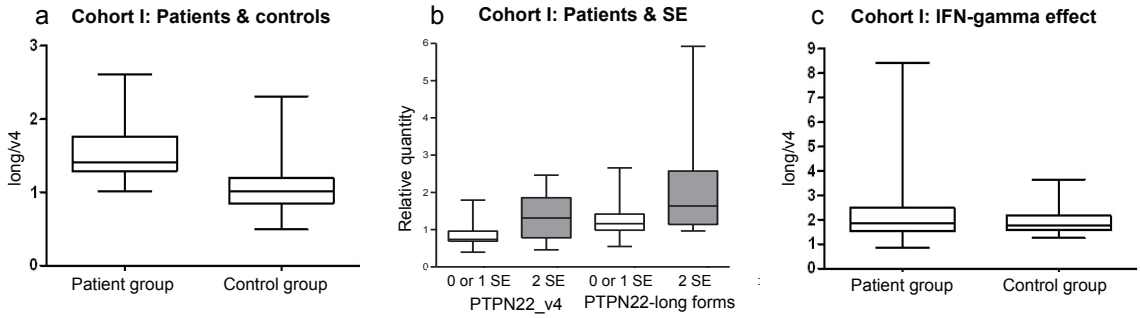

Figure 3. Expression of PTPN22 splice forms. (a) The ratio of PTPN22-long forms and PTPN22_v4 for rheumatoid arthritis (RA) patients and controls in cohort I. (b) Expression of PTPN22_v4 and PTPN22-long forms for RA patients stratified by copies of the shared epitope (SE), 0 or 1 (white boxes) versus 2 (grey boxes). (c) The effect of IFNγ *ex vivo* stimulation on expression ratio.

The HLA-DRB1 SE was found to be moderately associated with the expression of PTPN22. RA patients carrying two copies of the SE had moderately higher expression of both

PTPN22_v4 and PTPN22-long forms ($p = 0.04$ and $p = 0.02$, respectively, Mann-Whitney U test) compared to individuals with no SE or only one copy (Figure 3b).

The RA risk SNP rs2476601 (PTPN22 R620W) did not associate with the expression of splice forms in our study, nor did any other polymorphism in the PTPN22 locus show a consistent association.

We also tested how relevant inflammatory drugs and disease-modifying antirheumatic drugs (DMARDs) influence the expression of PTPN22 mRNA. The direct effect of pharmacological levels of methotrexate on PTPN22 expression was investigated in two cell lines and in PBMCs from healthy donors, and was found to be moderate. Additionally, we compared the expression levels of PTPN22 between RA patients with and without methotrexate treatment at the time of blood donation and found no differences.

To control for disease specificity of differential expression of PTPN22 splice forms, we analyzed independent collection of mRNA from PBMCs from individuals with multiple sclerosis (MS) and, separately, asthma. The MS patient group showed significantly higher expression of PTPN22_v4 compared to the control group and consequently a lower ratio of splice forms compared to controls. The asthma group did not show differential expression of PTPN22 transcripts when compared to healthy controls.

Our data show that RA patients and healthy controls differ in terms of expression ratio of splice forms derived from the PTPN22 gene. This balance shift does not appear to be caused by standard treatment and may be of importance during immune response due to significant difference in structural features of the encoded PTPN22 proteins.

6.2 PAPER II: PTPN2 VARIATIONS INTERACT WITH HLA-DRB1 SHARED EPITOPE ALLELES AND PTPN2 IS DIFFERENTIALLY EXPRESSED IN RHEUMATOID ARTHRITIS

Different studies have identified PTPN2 as a susceptibility locus associated with RA. The Wellcome Trust Case Control Consortium reported an association of the single nucleotide polymorphism (SNP) rs2542151, located 5.5 kb downstream of PTPN2 with Crohn's disease, celiac disease and RA (87). The association to RA was reported for SNPs rs1893217, rs7234029, and rs62097857, located in the intronic sequence of PTPN2 (36). In addition, the intronic SNP rs657555 was associated with RA in a Korean population. Even though

multiple SNPs within the PTPN2 region have been discovered, the functional consequences of these variations for the pathogenesis of RA remain unclear.

We identified 11 SNPs available in GWA study in the PTPN2 locus, both in the EIRA and in NARAC study. These SNPs were tested for interactions with HLA-DRB1 SE using a dominant genetic model. For the EIRA study, an additive interaction was observed between HLA-DRB1 SE and three SNPs in the PTPN2 region regarding risk of developing RA: rs2847297, rs657555, and rs11080606 (Figure 4).

The PTPN2 SNP rs657555 was additionally tested for interaction with HLA-DRB1 SE alleles in the extended EIRA study (2019 ACPA positive RA cases, 1126 ACPA negative RA cases, and 2247 healthy controls). We found that this SNP was in interaction with HLA-DRB1 SE in ACPA positive RA in extended cohort too.

For the NARAC study, interaction was observed between HLA-DRB1 SE and three SNPs in the PTPN2 region: rs1893217, rs657555, and rs11080606. These results leave PTPN2 SNPs rs657555 and rs11080606 as most reliable candidates for interaction with HLA-DRB1 SE regarding risk of developing of ACPA positive RA in EIRA and NARAC studies. Our analysis revealed a significant association of PTPN2 SNP rs657555 with RA in EIRA while rs11080606 was not associated with the disease.

We measured the expression levels of total PTPN2 and, separately, the transcripts TC41, TC45, and TC48, relative to expression of the reference house-keeping gene ZNF592 in whole blood and PBMCs of RA patients and healthy controls. We found, however, that TC41 mRNA was not expressed at detectable levels in whole blood samples and PBMCs. In whole blood samples (67 RA patients and 70 healthy controls), we found no differences in average expression of total PTPN2 and TC45 between healthy controls and RA patients. However, we found marginally lower average expression of TC48 in RA patients ($p = 0.046$) and also the ratio of expression between TC45 and TC48 was significantly different between healthy controls and RA patients.

To identify more specifically the source of differential expression, we also tested the PTPN2 expression pattern in an independent set of samples from RA patients and controls with mRNA extracted from PBMCs. We found a significant difference in expression of total PTPN2 and TC48 in healthy controls and RA patients with a higher average expression of

total PTPN2 and TC48 observed in RA patients. We detected no difference in TC45 expression in PBMCs between the groups of healthy controls and RA patients.

In whole blood samples, we detected no differences in PTPN2 expression with regard to HLA-DRB1 SE alleles and ACPA status. In PBMCs, we found that the mRNA expression levels of total PTPN2 and PTPN2 transcripts are not different in carriers and non-carriers of HLA-DRB1 SE alleles, neither in the control group nor in the RA group. We observed a trend in difference between total PTPN2 expression between groups of individuals with ACPA positive and ACPA negative RA, however not in PBMCs.

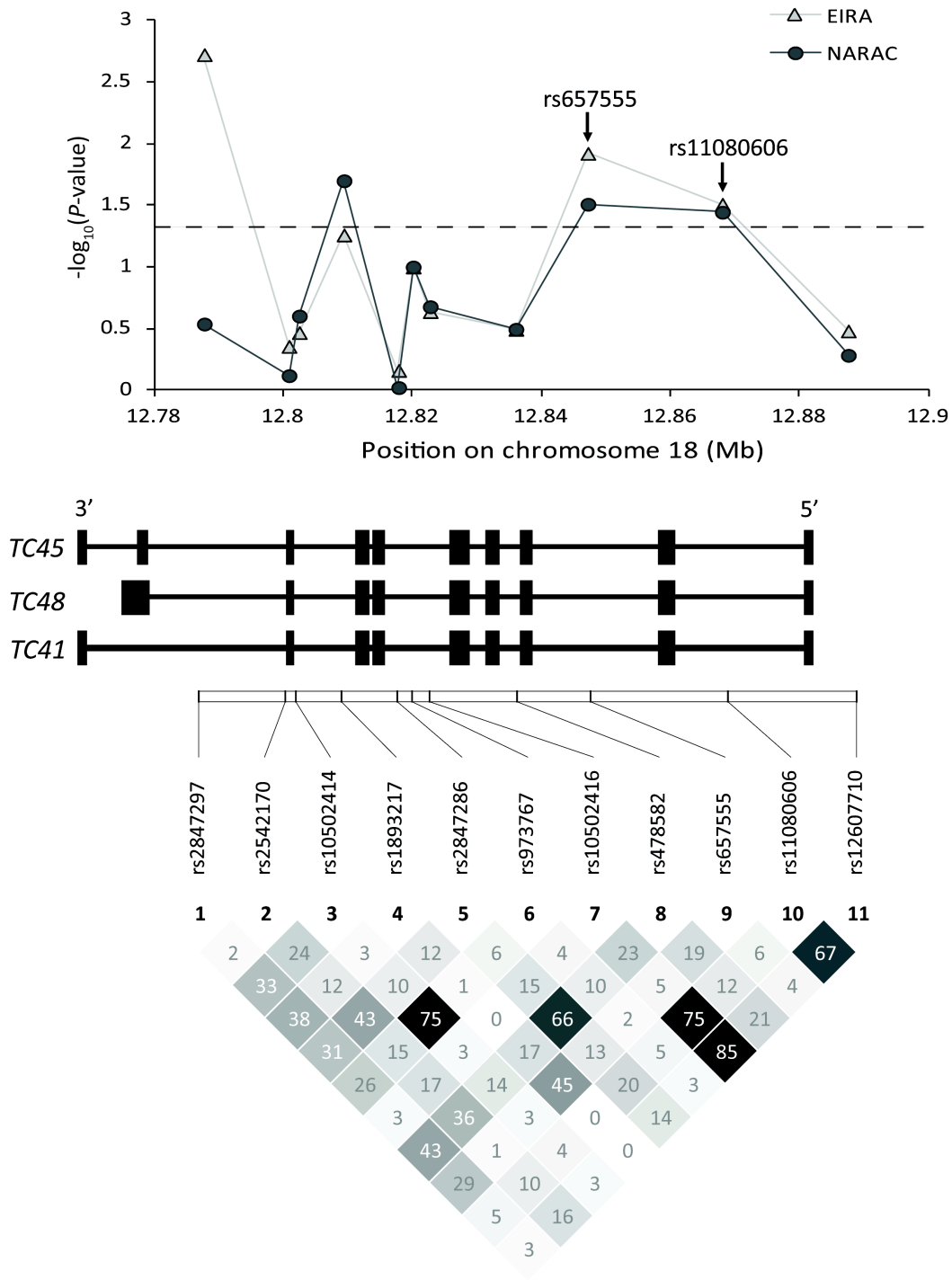

Figure 4. Interaction between *PTPN2* SNPs and *HLA---DRB1* SE alleles. Unadjusted *P-values* (-log10) for interaction (AP) between *HLA - DRB1* SE and the different SNPs in the *PTPN2* region of the EIRA (triangle) and NARAC (solid circle) cohort plotted against the genomic positions of each SNP on chromosome 18. The dominant genetic model was used for the interaction studies. The horizontal dashed line corresponding to a threshold of $P < 0.05$. The LD - plot (r2) is shown for variants in the *PTPN2* region.

6.3 PAPER III: GENE-GENE INTERACTION AND RNA SPLICING PROFILES OF MAP2K4 GENE IN RHEUMATOID ARTHRITIS

One of the recent developments in the biological therapy for RA is targeting mitogen-activated protein kinases (MAPKs), in particular p38 and JNK. MAPKs are considered to be among the key inflammatory mediators for autoimmune diseases due to their involvement in the regulation of immune response to stress stimuli. MAPK signaling plays a major role in proliferation and differentiation of T- and B- cells, which are known to be fundamental in the development of autoimmune reaction and sustained inflammatory response. Attempts to inhibit the activity of p38 and JNK to treat RA patients in clinical trials were unsuccessful, because of unacceptable level of side effects. It has been suggested, that upstream (MAPKKs and MAPKKKs) are responsible for fine regulation of JNK and p38 activity, and may be functionally and clinically relevant for RA development.

We performed interaction analysis on pairwise combinations of HLA-DRB1 SE alleles and 195 SNPs from 12 genetic loci encoding known MAP2K, MAP3K and MAP4K upstream of JNK1 and p38 (including JNK1), in 1921 RA cases and 1079 healthy controls (EIRA cohort). Our data indicate that a single SNP, rs10468473, is in a relatively strong interaction with HLA-DRB1 SE alleles in the development of ACPA-positive RA (AP = 0.79, $p = 4.4 \times 10^{-6}$, 95% CI [0.51–1.0]).

To improve the representation of the individuals, positive exclusively for rs10468473 risk allele in our interaction analysis, we performed additional genotyping for this SNP in 861 RA patients and 1200 controls from EIRA study (not previously included in GWAS). In the extended cohort, rs10468473 from the MAP2K4 locus remained in significant interaction with HLA-DRB1 SE alleles (AP = 0.434, p = 0.041, 95% CI [0.018–0.85]) (Figure 5).

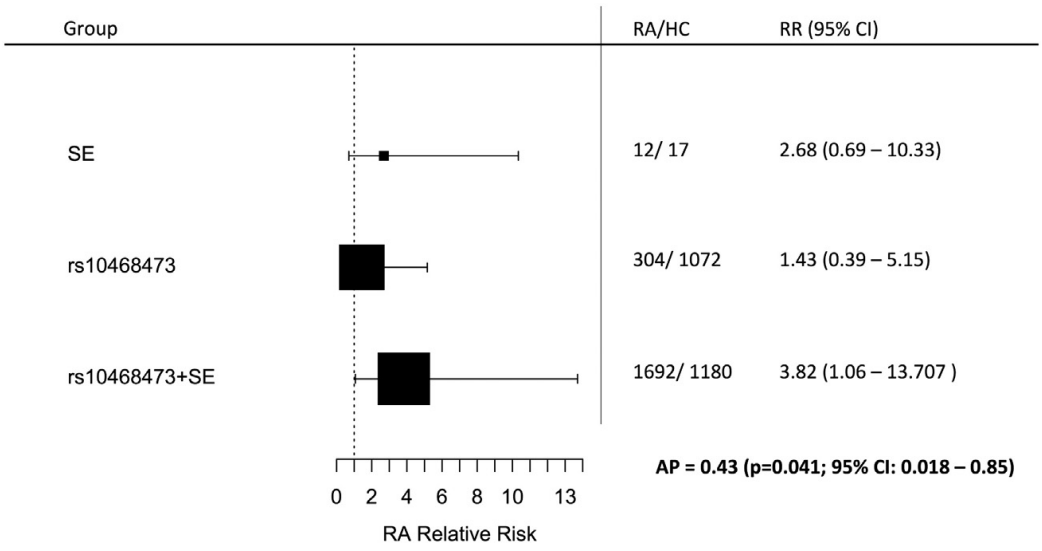

Figure 5. SNP rs10468473 from the MAP2K4 locus, when combined with SE, provides a significant, non-additive increase in risk for ACPA-positive RA. The statistical interaction is represented by attributable proportion (AP). The analysis was based on 2010 ACPA-positive cases and 2280 healthy controls. The unaffected reference group included 11 individuals. Data presented as relative risk (RR) for carriers of different risk genotypes: SE — individuals with shared epitope risk genotype; rs10468473 — individuals with MAP2K4 rs 10468473_A risk genotype; – rs10468473_A + SE — individuals with both MAP2K4 rs1046847 and SE risk genotypes; reference line (dashed) — unaffected individuals; the area of the squares represents the log-size of the genotype group; whiskers show RR confidence intervals.

To find possible mechanisms, explaining the risk association of genetic variation in the MAP2K4 locus, we investigated the allelic expression of the MAP2K4 gene in mRNA from PBMCs of 44 RA patients and 44 matched controls with known rs10468473 genotype. Our data showed significantly elevated MAP2K4 expression in RA patients heterozygous for rs10468473, compared to individuals homozygous for A allele ($p = 0.045$; Mann–Whitney U test) (Figure 6a).

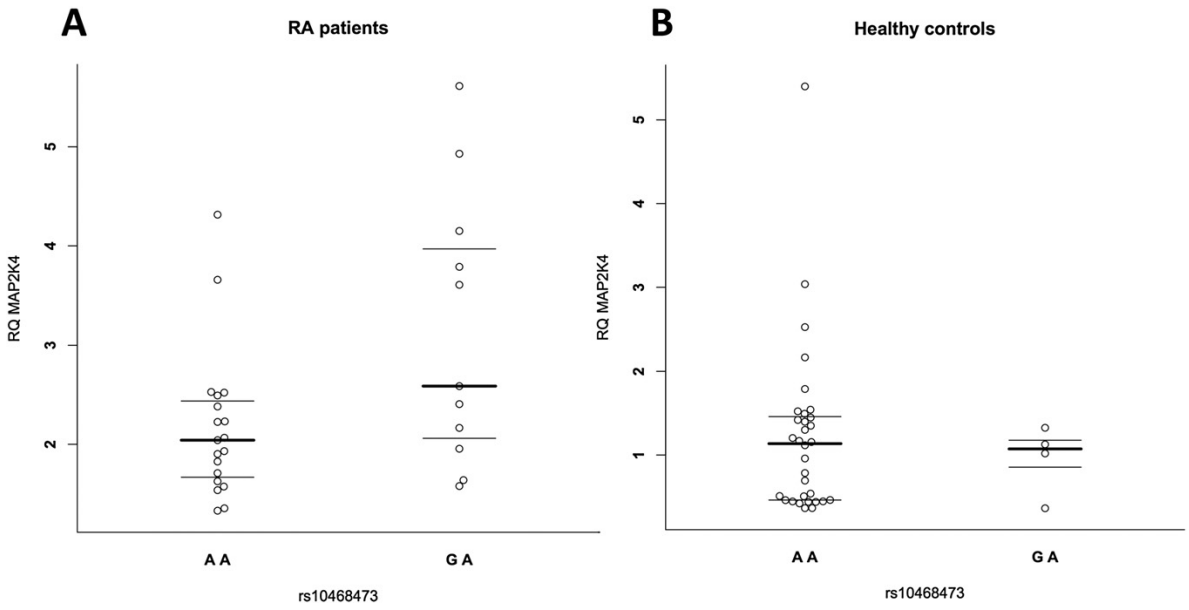

Figure 6. **rs10468473 genotype correlates with MAP2K4 gene expression in RA patients**. (a) The expression of MAP2K4 gene is significantly higher in PBMC of 13 RA patients, with a heterozygous G/A genotype for rs10468473, compared to 22 patients homozygous for A allele ($p = 0.045$). (b) In healthy controls, no significant change in MAP2K4 expression was observed between individuals with G/A ($n = 4$) and A/A ($n = 32$) rs10468473 genotypes. Shown are the relative quantification values (RQ) for MAP2K4 expression in PBMCs of RA patients and healthy controls with known rs10468473 genotype. The expression of MAP2K4 was measured by TaqMan qPCR using predesigned gene-expression assay Hs00387426_m1, spanning exons 3–4 of the original sequence. Mann–Whitney U test with a significance threshold of $p = 0.05$ was used for statistical analysis.

We performed an exploratory study by using generic PCR to identify a spectrum of MAP2K4 mRNA products, detectable in PBMC. We identified a series of transcripts with altered exonic composition, and a prevalence of "cassette-exon"-type splicing events in the coding part of the Serine/Threonine protein kinase catalytic domain. Specifically, exon 5 skipping had been reported previously for various transcripts of human (ENST00000538465; ENST00000538465) and mouse (ENSMUST00000125598) origin, and was featured in our material (Figure 7a).

To assess the possibility of translation for MAP2K4 transcript with skipped exon 5 (assigned as V2), we transfected HEK293 cells with canonical MAP2K4 sequence (V1) and V2-containing plasmids. Both V1 and V2 transfection products could be detected by using commercially available mouse monoclonal anti-MAP2K4 anti-bodies in lysates of transfected HEK293 cells. Proteins with similar molecular weight were also detectable in non-transfected T-lymphocyte cell lines by Western blot (Figure 7b).

We investigated the production of MAP2K4 transcript with skipped exon 5 (assigned as MAP2K4_V2) in relation to the expression of canonic full isoform (assigned as MAP2K4_V1), in PBMC from 38 RA cases and 43 controls. When taken as a ratio (V2/V1), the expression of the transcripts was significantly different for RA patient and healthy control groups.

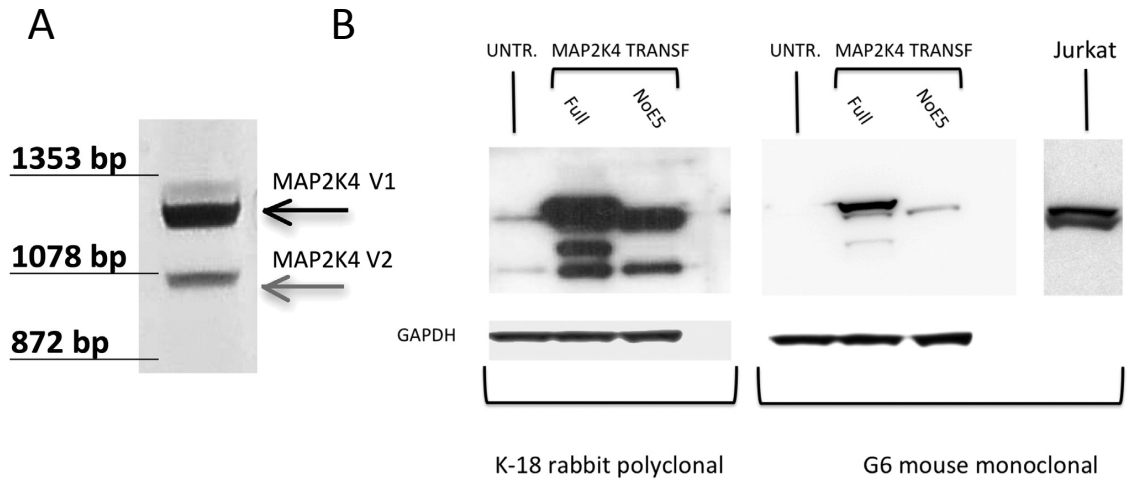

Figure 7. MAP2K4 transcript, lacking exon 5 may undergo translation into protein

(a) A "skipped-exon" variant MAP2K4 was characterized in whole-blood mRNA. Shown are the PCR amplification spanning exons 3–11 of the original MAP2K4 sequence. The canonical MAP2K4 transcript (black arrow) and a "skipped exon" splice form (gray arrow) were confirmed through sequencing. **(b)** MAP2K4 NoE5 protein isoform can be detected with monoclonal anti-MAP2K4 G6 and rabbit polyclonal G6 IgGs in lysates of in HEK293 cells, transfected with cDNA of NoE5 splicing variant of MAP2K4. The produced band is similar to the lower molecular weight product in 2-product pattern, detectable with the same antibody in both HEK293 cells, transfected with canonical MAP2K4 cDNA and untransfected Jurkat cells (right). Full-MAP2K4-transfected and untransfected HEK293 also exhibit lower molecular weight products that are not present in the MAP2K4 NoE5 HEK293 transfection.

In RA patients, we identified a significant correlation and a dosage effect between the number of carried SE alleles and the MAP2K4 V2/V1 expression ratio. Within the same group, we could also observe a correlation between the MAP2K4 variant expression ratio and RA-related serological factors, i.e. the presence of ACPA and presence of rheumatoid factor.

We studied the effect of TNF stimulation on the transcript expression ratio of MAP2K4 in lymphatic and myeloid cell-lines. The expression of MAP2K4 V1/V2 was measured at four different time points in Jurkat, Raji, and U937 cell lines treated with recombinant TNF. The transcripts of MAP2K4 were differentially expressed in T-cell-like Jurkat cell-line at 24 h of stimulation. There was no significant difference in transcript ratio for Raji and U937.

Our data suggest a presence of the genetic interaction between MAP2K4 and HLA-DRB1, and the importance of rs10468473 and MAP2K4 splice variants in the development of autoantibody-positive RA.

6.4 PAPER IV: EXPRESSION PATHWAY ANALYSIS FOR GENES ASSOCIATED WITH RHEUMATOID ARTHRITIS

The rapid decline in cost of RNA-sequencing, combined with a dramatic increase in throughput turned into an indispensible tool for transcriptome research.

However, there are multiple ways of data analysis for RNA-seq derived results, but little consensus about a unified analysis and interpretation approach. Integration of multiple RNA-seq data with immense amounts of data from GWAS studies, and protein interaction databases may ameliorate the interpretation program, and contribute to the discovery of new molecular targets, supported on multiple levels of experimental evidence. Importantly, integration with whole-transcriptome analysis methods allows for a more informative view of candidate gene expression, as expression of multiple transcripts from different loci can be assessed in a single experiment.

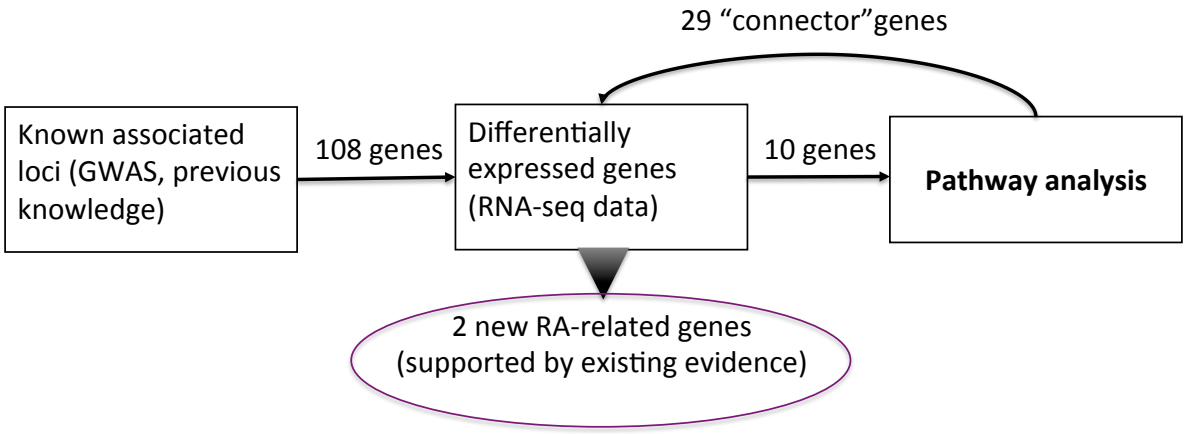

Figure 8. Out of the initial set of genes in direct proximity to reported RA variants, those DE in our RNA-seq data were used in Ingenuity Pathway analysis sofware. Interaction molecules, suggested by IPA were again compared to the RNA-seq differential expression results.

The general design of the study is presented in Figure 8. Candidate genes were chosen based on proximity to validated genetic variants. RNA-seq was performed on whole-blood samples of 12 healthy controls, 6 RA patients before methotrexate treatment, and 6 RA patients undergoing methotrexate treatment (all individuals were of female sex). The consideration that treatment may have a pronounced effect gene expression informed our selection of differentially expressed genes – only genes with similar expression in treated and untreated RA patients were selected. Out of 108 genes from RA-associated loci (4, 103, 139-142), 28 genes remained differentially expressed by RNA-seq between any of the three groups after correction, but only 10 were differentially expressed in both non-treated RA and treated RA compared to healthy controls (p=0.05, Fisher combined p-value). As expected, clustering based on the expression pattern of these 10 genes results in reasonably good separation of healthy controls and RA patients, but does not provide a clear distinction for the treatment status.

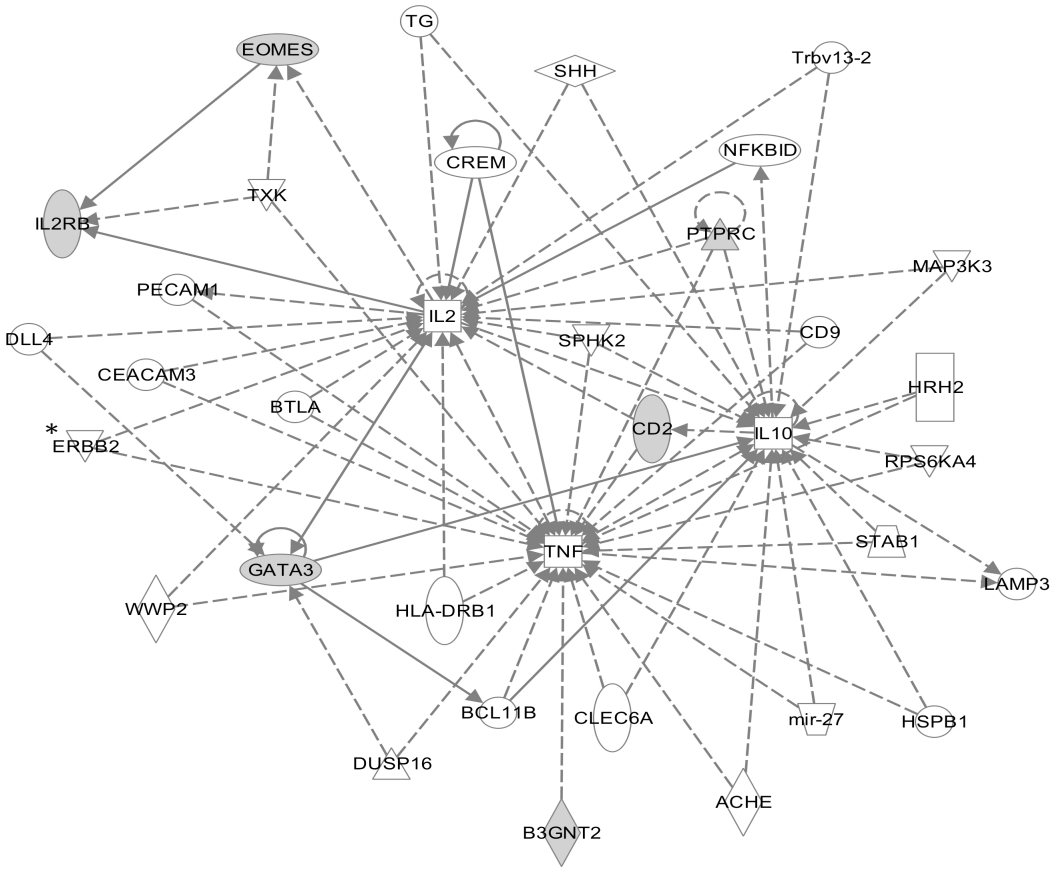

Figure 9. A network, containing HLA-DRB1 and 9 other genes, significantly DE in whole blood RNA-seq data (Ingenuity Pathway Analysis (IPA)). Gray shapes represent genes, previously associated in RA, that are also DE in RNA-seq data (IPA-input), White shapes represent interaction molecules, derived from IPA database; solid lines – direct interaction evidence; dashed lines – indirect interaction evidence; asterisk - significant difference in expression in whole blood from both naïve RA vs. HC and treated RA vs. HC RNA-seq data comparisons.

With the use of IPA software, 10 differentially expressed genes were grouped into a functional molecular network with an IPA score of 12, involving 29 additional interacting molecules, based on experimental data from immune cells and cell lines (Figure 9).

HLA-DRB1, HSPB1 and ERBB2 retained a significant DE profile in both untreated and treated RA compared to healthy controls after multiple testing correction of the combined p-values from both comparisons (data for ERBB2 presented at Figure 10a).

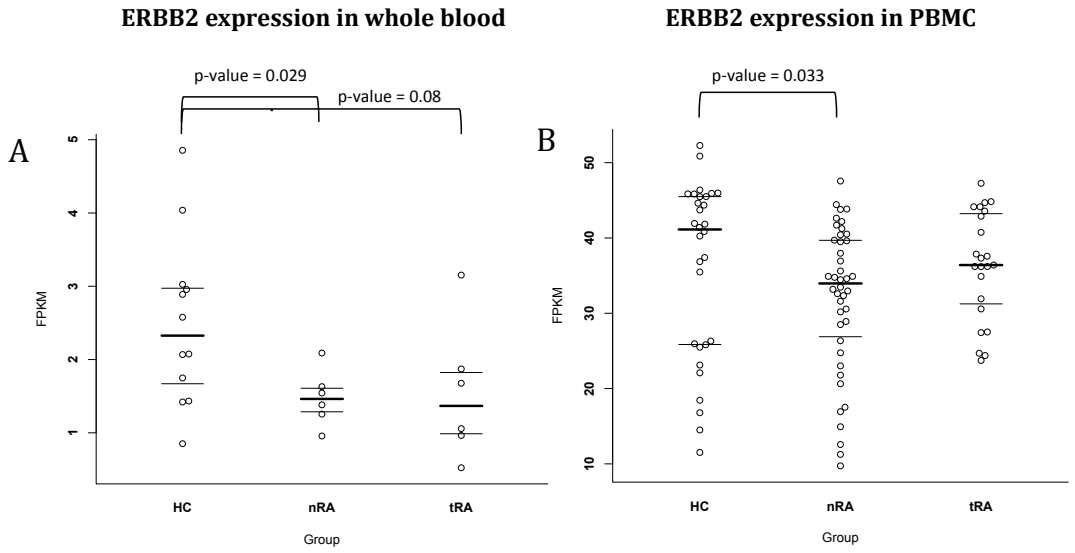

Figure 10. (a) ERB22 was significantly DE in whole blood from 6 pre-treatment RA patients and 12 healthy controls (p=0.029), and a trend towards DE in treated RA patients and 12 healthy controls (p=0.086); Bonferroni corrected Cuffdiff p---value of 0.05 was used as threshold. (b) The DE of ERBB2 was replicated in PBMCs samples from an independent cohort of 48 untreated RA patients, 30 methotrexate treated RA patients, and 35 healthy controls (p=0.033). Mann Whitney test p-value of 0.05 was used as a significance threshold.

HC – healthy controls: nRA – "naïve" RA patients: tRA – treated RA patients;

To replicate our findings we employed analysis of RNA-seq data from independent cohort of PBMCs from 48 untreated female RA patients, 30 methotrexate treated RA female patients, and 35 healthy female controls. ERBB2 expression remained significantly different in PBMC samples, with the gene expression pattern similar to that, observed in whole blood, and a significant difference in expression between healthy controls and untreated patients (Figure 10b).

In this study, an integration of RNA-seq data with prior association study data and with following pathway analysis allowed us to infer new candidate genes and molecular pathways that are potentially involved in RA pathogenesis. Integrated pathway analysis implies a potential functional link between several genes previously associated with RA, and new candidates, ERBB2 and HSPB1.

7 DISCUSSION

7.1 OVERVIEW

From the era of the GWAS studies, we have discovered multiple genetic variants associated with disease. In genetic studies, association allows us to attribute a degree of causation to genetic variance, due to the role of the genome as a basis for structure and function of a biological system. However, association remains a purely statistical concept that does not provide us with enough information to infer the function of a particular genetic variant.

In case of RA, great progress has been made in pinning down the genetic variance that could contribute to disease development. However, the understanding of how this genetic contribution is realized is far from complete.

Assessing the functional importance of a given genetic variant includes a vast set of methods, spanning from bioinformatics and gene expression analysis to proteomics and genetically modified model organisms.

The intention of thesis work was to fill the gaps in our understanding of the genetic events that form the genetic component of RA. Specifically, a number of loci associated with RA were studied in terms of their expression, splicing profile and interaction with HLA-DRB1.

To achieve this, we resorted to a combination of distinct methodological approaches, discussed below.

Gene-gene interaction analysis can be a useful tool to get a glimpse of the potential functional connections between associated loci, while using available GWAS data. It allows for identification of genetic associations, which are otherwise missed due to methodological constraints of GWAS. Interaction analysis enabled us to report a novel associated polymorphism in PTPN2 gene, which we initially detected through interaction with HLA-DRB1. Furthermore, our work on MAP2K4 was also prompted by a gene-gene interaction analysis of the members of MAPK signaling pathway with HLA-DRB1.

The measurement of total gene expression for candidate genes has been a routine approach testing the functional involvement of a genetic locus in disease pathogenesis. However, with our current understanding of the complexity of the transcriptome, this approach may no longer be sufficient. Detection methods, relying on a consensus part of the transcripts, will

not yield complete information in cases, when alternative mRNA splice forms of a gene are differentially regulated. In our work on PTPN22 and PTPN2, we try to circumvent this issue, by differentiating between the existing isoforms of these genes in our expression studies. Further more, in the study MAP2K4 we include transcript discovery stage to create a detailed picture of the gene expression profile, prior to expression analysis in clinical samples and cell lines.

Lastly, to find new functionally relevant candidates in RA we required an approach that would integrate validated susceptibility loci from GWAS, gene expression genome-wide profiling, and existing experimental evidences. This prompted us to take advantage of the increased efficiency and availability of RNA-seq technology, which was used to analyze mRNA expression for genes, proximal to verified susceptibility loci for RA, in clinical samples from RA patients and healthy individuals. We used pathway analysis to look for existing evidence of the possible functional interactions for genes that are both proximal to RA-associated variants and differentially expressed in case-control expression study. We proceeded to investigate the expression of the functionally relevant genes, suggested by pathway analysis.

7.2 EXPRESSION PROFILES OF PTPN22 AND PTPN2 IN RA

The polymorphisms from genes PTPN22 and PTPN2 were previously found in association with RA, with PTPN22 being second strongest risk factor for the disease. Both genes belong to the family of protein tyrosine phosphatases.

PTPN22 SNP rs2476601 was previously reported to be in interaction with HLA-DRB1 SE in determining the risk of RA. Therefore, it was interesting to also see an association of PTPN22 expression to HLA-DRB1.

Notably, PTPN22 protein isoforms, corresponding to the transcripts in this study, have distinct structural differences, reported previously. For instance, not all functional motifs, common between Lyp1 and other members of the PEST family of phosphatases, PTP-PEST (PTP with PEST domain, genomic designation PTPN12) and PTP-HSCF (PTP hematopoietic stem-cell fraction, genomic designation PTPN18) are preserved in Lyp 2. These differences include the absence of the carboxy-terminal homology (CTH) domain, which is a binding site for CD2-binding protein 1 (CD2BP1), cyclin-dependent kinase 1 (CDK1) recognition motif, and 2 out of 3 SRC kinase binding sites.

These differences could very well stand behind the association of PTPN22 with the RA phenotype. For example, CD2BP1 is expressed in T cells and has been proposed to be an important regulator of T-cell behavior by modulating CD2 activity, which in turn regulates the binding of T-cell receptors to a major histocompatibility complex (MHC) molecule with an antigenic peptide. Cyclin-dependent kinase 1 (CDK1), a cell cycle regulatory kinase, which may imply a difference in cell-cycle dependent regulation of Lyp2 regulation compared to Lyp1 (138). c-Src is known to negatively regulate leukocyte-specific protein tyrosine kinase - an important mediator in cytokine production signaling.

These possible functional differences between Lyp1 and Lyp2 were not appreciated previously and may suggest that a shift in the balance of splice forms could have a considerable impact on the inflammatory signaling. This difference in expression seems not to be significantly regulated by allelic variants and although it cannot be totally excluded according to our cell line experiments, the expression seems not to be affected by medication of the patients. Also, in the presence of a strong inflammatory environment (IFNγ in our study) the balance in transcript expression could be well compensated. Interestingly, our study of PTPN22 splice form expression in MS patients did not show a similar effect, which is in line with the fact that this disease is not associated with polymorphisms in the PTPN22 gene and supports disease specificity. To resolve if this difference in expression is related to genetic association with PTPN22 variants, expression of PTPN22 splice forms in systemic lupus erythematosus and type 1 diabetes, diseases shown to be associated with variations in PTPN22, should be tested in a similar way.

In the present study, we were able to identify an additional susceptibility SNP in PTPN2 that distinguish a subgroup of RA cases from healthy controls, through the analysis of interactions between SNPs in PTPN2 and HLA-DRB1 SE. Although the size of the subgroup is moderate, the effect of this combination on RA risk is relatively high.

Evidence of interaction between the PTPN2 SNP rs657555 and HLA-DRB1 SE was observed in both the EIRA and NARAC study for ACPA positive disease, when an additive model was used. In addition, our analysis revealed a significant association of rs657555 with RA. A possible biological background for the interaction between HLA-DRB1 SE and PTPN2 might be that PTPN2 regulates T-cell receptor signaling. However, the potential link between PTPN2 and HLA-DRB1 SE should be further investigated in functional studies in relation to different autoimmune diseases.

There was no experimentally determined proof that the PTPN2 SNP rs657555 had any regulatory features. In addition, we found that rs657555 does not correlate with the expression of total PTPN2, TC45, and TC48 mRNA levels in whole blood samples and PBMCs of healthy controls and RA patients. Given that rs657555 is in strong LD ($r2 > 0.8$) with 20 intronic SNPs, it can be hypothesized that one of these SNPs is the actual causal variant. We investigated the 20 intronic SNPs in PTPN2 using the databases HaploReg and RegulomeDB and discovered that rs7241016 and rs11875687 are likely to affect the binding of STAT1 and/or STAT3, which might in turn regulate the expression of genes after cytokine or growth factor stimulation. The SNP rs534911 may affect the binding of CTCF, which is a crucial regulator of transcription and may also influence mRNA splicing. Because these SNPs might have regulatory features, they provide candidates for future investigation.

The current study also detected that average levels of total PTPN2 and TC48 expression are significantly higher in PBMCs of healthy controls and RA patients. The opposite results for TC48 expression in whole blood samples and PBMCs of RA patients might be explained by the differences in cell composition between whole blood and PBMCs. It will be of interest to study the expression of PTPN2 transcripts in specific types of blood cells to reveal the function of PTPN2 in RA.

TC45 and TC48 are expressed in different cellular compartments and interact with different cellular targets. Many of the TC48-interacting proteins are involved in intracellular vesicle transport and cell adhesion. However, further investigation is needed to elucidate the role of TC48 in the pathogenesis of RA.

7.3 RELEVANCE OF MAP2K4 GENE IN RA

Analysis of genetic variations in MAPKK and MAPKKK related loci imply an interaction between a single member of the MAP signaling pathway, MAP2K4, and HLA-DRB1 SE loci, that may contribute to the risk of autoantibody-positive RA that was overlooked in GWAS studies. These results constitute that a gene-gene interaction attributes to increased risk of ACPA-positive rheumatoid arthritis and can be helpful for clinical prediction in the future. Based on the interaction of MAP2K4 SNP with the major risk factor for RA, it is possible that MAP2K4 plays a role in RA pathogenesis only in a specific subgroup of individuals with HLA-DRB1 SE alleles.

Importantly, we observed association of the rs10468473 genotype with MAP2K4 mRNA expression in our material of PBMC from RA patients. This allows us to assume that rs10468473, or other SNPs in LD with it, is potentially related to the regulation of expression in the MAP2K4 locus, and could be involved in inflammatory responses in RA.

In this study we characterized splicing events in MAP2K4, resulting in partial lack of catalytic domain coding sequence. Specifically, MAP2K4_V2, characterized by the skipping of exon 5, could be effectively translated into a protein product, as confirmed by our transfection experiments could be consistently identified with a specific primer-set in mRNA from PBMC of RA patients and controls. The functional implications from the lack of catalytic part of MAP2K4 could be substantial: while not crucial for activation of either p38 or JNK, MAP2K4 works in tandem with MKK6/3 and MKK7, ensuring optimal activation of these MAPKs (108). An inadequate upstream regulation of p38 and JNK may influence their role as transducers of inflammation and joint destruction. Furthermore, altered regulation of JNK could potentially have an effect on the expression of genes, involved in antigen presentation (HLA class II and CIITA) (143).
Additionally, our results show differential expression of the canonical and alternatively spliced MAP2K4 mRNAs in RA patients, compared to healthy controls. It also depends on ACPA status, RF positivity, and number of SE alleles. The decreased V2/V1 ratio in RA patients could be chiefly attributed to the increased expression of canonical MAP2K4 transcript, while the correlations with SE alleles and serological markers within the RA group are driven by V2 (NoE5).
It has been reported that SNPs located in untranslated areas may contribute to the regulation of gene splicing (144). However, we could not see a clear association between the rs10468473 or 21 other common MAP2K4 SNP genotypes with the expression of individual splicing variants (canonical and lacking exon 5) of the MAP2K4 gene.

TNF signaling leads to the activation of JNK, p38 and ERK MAPKs (111, 145). The observed effect of TNF stimulation on the transcript ratio in our experiments in the Jurkat cell-line, but not in Raji or U937, may indicate a cell-type specific splicing profile of MAP2K4 in pro-inflammatory conditions — and a topic for further analyses of mechanisms of involvement in RA and development of treatment strategy. This is in line with the established effect of anti-TNF treatment in RA and suggests how T-cell specific expression of MAP2K4 isoforms could be functionally involved in RA pathogenesis through the regulation of proliferation in the T-cell population.

Alternative mRNA transcripts can have a regulatory function by competing with the canonical transcript and shifting the amount of the main variant in the total gene output. This knowledge can support the functional significance of splicing in MAP2K4 transcripts, even when the translation efficiency for non-canonical products is lower. It also seems reasonable to consider the possibility of NMD-subjected transcripts potentially retaining mRNA-level regulatory capacity in disease pathogenesis. It remains to be determined whether, and if so, how the balance of MAP2K4 splice variants may functionally affect the activation of downstream JNK and p38 in the context of autoimmunity. The quantity of an alternative MAP2K4 isoform could potentially interfere with complex formation and activation of downstream p38 and JNK MAPKs. If the gene expression of MAP2K4 is constitutively effected by a specific genotype it could potentially contribute to the development of autoimmunity over time.

7.4 108 RA ASSOCIATED LOCI: EXPRESSION AND PATHWAY ANALYSIS

Genetic association and gene-gene interaction studies were instrumental in detection of important candidate genes in the development of RA. However, it may show only a reflection of possible functional involvement of genomic mechanisms and alternative approaches, like study of gene expression are important. However, despite the decrease in the cost of RNA-seq, we have a limitation in the number of samples for the analysis with relatively low statistical power for genome-wide comparison. To address potentially unknown disease associated pathways, we integrated validated GWAS markers with whole transcriptome expression data and pathways analysis thus limiting our selection to well validated genetic risk factors with the goal to link it with other members of the disease pathways.

We identified that 10 of the genes, proximal to validated RA-associated variance, were differentially expressed in our material of whole blood samples from 12 RA patients and 12 healthy controls. Notably, samples could be reasonably grouped into RA and non-RA based on this expression profile alone. This particular gene set, however, does not provide a distinctive clustering between treated and untreated RA patients. This circumstance aligns with our intention to avoid genes, displaying heterogeneity of expression of genes depending on response to treatment.

Using the Ingenuity Pathway Analysis service, we obtained information about the functional interactions for the DE genes from RA-associated loci in primary immune cells and immune cell-lines, based exclusively on available experimental data (i.e. the prediction features of the software were not employed). As a result, 6 out of 10 input genes were grouped into a single

network, where IL2, IL10 and TNF served as connecting hubs. Importantly, this suggested network also contained HLA-DRB1 – the major genetic risk factor of rheumatoid arthritis, which was not manually included within the input genes dataset.

Addressing the expression of 28 new "connector" genes, suggested by the pathway analysis, revealed that some of them were DE in whole blood from 6 untreated, 6 treated RA patients and 12 healthy controls. However, with the exception of HLA-DRB1, ERBB2, and HSPB1 DE was limited to the comparison of healthy individuals to either treated or untreated RA groups, but not both. Notably, replicating ERBB2 expression in an independent material of PBMC samples from methotrexate treated and naive RA patients versus healthy controls revealed a similar expression pattern, and a significant difference in expression between the latter two groups.

ERBB2 (HER2/neu) is a receptor tyrosine-protein kinase erbB-2, previously implicated in promoting hyperproliferative growth in arthritic synovial tissue (146). ERB22 protein was shown to be highly expressed in synovial joint(147). However, the ERBB2 expression has not been studied before in either whole blood or PBMCs in context of RA. Notably, it is known that ERBB2 plays an important role in the activation of the NFkB pathway and, potentially, an important mediator of TNF signaling (148, 149), which are both strongly implicated in RA pathogenesis. Therefore, assessing the effects of existing ERBB2 inhibitors in an experimental model of RA may be a promising endeavor.

Integration of GWAS data, whole-transcriptome RNA-seq analysis with pathway analysis may suggest new relevant genes and interaction networks that can be used for functional studies. However, one should be careful to avoid potential annotation mismatches when integrating data from different sources. Specifically, for highly computational analysis of whole-transcriptome data, it is important to consider biases that may be inherent in a specific analysis pipeline.

Additionally, attribution of associated SNPs to proximal genes could be a limiting factor of this study, since it does not take into account potential trans-regulatory effects of SNPs on gene expression in other loci.

8 CONCLUSIONS

With this work we attempted to elucidate the molecular workings behind genetic association in rheumatoid arthritis – a common complex autoimmune disease. In terms of clinical significance, genetic epidemiology has now reached a point of diminishing returns from genome-wide association studies, with unprecedented cohort sizes utilized to detect extremely modest effects on disease risk. Currently, a lot of ongoing work is focused on combining different levels of biological knowledge, and breathing new life into available GWAS data. In our view, integration of GWAS data, gene-gene interaction studies, gene expression profiles, and model cell-systems allows for functionally meaningful interpretation of associated variance.

We believe it is essential to account for a specific transcript profile in genes from susceptibility loci. Based on our studies of RA associated genes PTPN22 and PTPN2, the transcript ratio of a gene may play a role in disease pathogenesis, and reflect other genetic and phenotypic features of the disease.

Moreover, as we observed in our work on MAP2K4, it may be important to consider, that the functionally relevant alternative splice forms have not yet been reported for the gene of interest.

Gene-gene interaction analysis can be an effective way to utilize GWAS data to find new risk variants and functionally associated genes (studies of PTPN2 and MAP2K4, respectively).

Finally, integration of GWAS data, whole-transcriptome RNA-seq analysis with pathway analysis may suggest new relevant genes and interaction networks, that can be used for functional studies.

According to previous studies, genetic risk factors, discussed in this work may participate in the pathogenesis of RA on multiple levels (Figure 11). However, more information about the functional interplay between HLA-DRB1 and other involved factors is necessary to better understand the pathogenesis of RA and, possibly, other autoimmune diseases. We believe, this work can be helpful for conducting informed functional studies of genes, involved in RA pathogenesis – an essential next step.

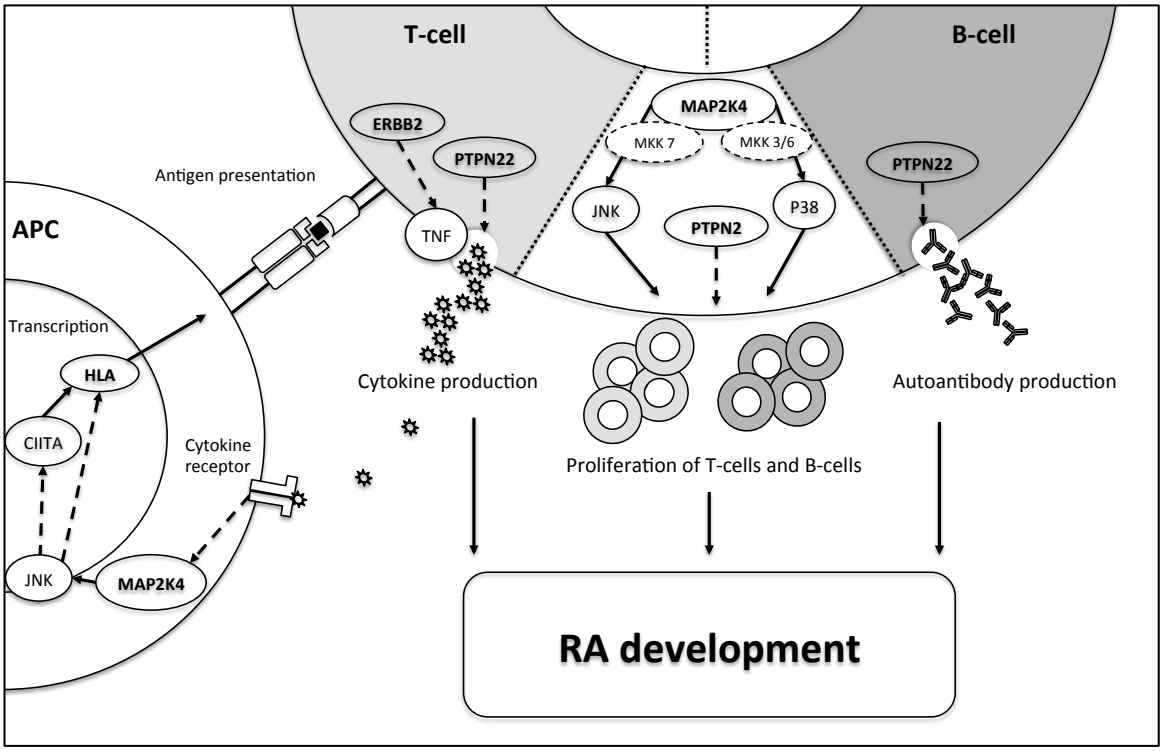

Figure 11. A schematic of the potential involvement of the discussed genes in the pathogenesis of RA and a sustained of autoimmune reaction. APC – antigen presenting cell. Solid lines – direct interactions; dashed lines – indirect interactions.

A common trait of all the studies, included in this thesis, is an attempt to build up from a genetic susceptibility variant to a functional gene, relevant to disease pathogenesis, then to a protein and its interactions and functions, and, finally, to a signaling network that would help us better understand and contain rheumatoid arthritis and other complex autoimmune diseases.

9 AKNOWLEDGEMENTS

This work would be hardly imaginable without these people. I would like to express my deep gratitude to:

Leonid Padyukov – my main supervisor and mentor. Thank you for sharing your invaluable science and life knowledge with me – these discussions will continue to guide me long after this work is over. Working with you made it seem there is no goal our group cannot achieve.

Lars Klareskog – it was a great privilege to be a part of the wonderful working environment that you have built in your group, and in CMM as whole. Thank you for sharing your vision and inspiring us all through the years.

Ingrid Lundberg – thank you for motivating us, leading us forward and keeping the whole lab working so seamlessly.

All my **co-authors**: thank you for your involved contributions, reliable feedback, and all the mental and physical resources that went into our common work – it is impossible to do this alone!

My dear group-mates: **Miranda Houtman**, **Espen Hesselberg**, **Lina-Marcella Diaz-Gallo –** for creating a wonderful and productive working environment. Your ideas, your critique and your support made the cornerstone for most of this work. I feel lucky and privileged I got to know all of you as both my co-workers and friends.

Marcus Rönninger, **Emeli Lunström** and **Maria Seddighzadeh**, who showed me the ropes during my first month in the group. Working with you was a great pleasure, but knowing you outside of the lab is an even greater one!

Alexander Espinosa and **Maria Sjöstrand** – a bright, friendly and helpful team!

Lasse Folkersen and **Loise Berg –** thank you for help with and coordination of the enormous Combine 2 project**.**

Darya Protsyuk – for all the energy and effort that you put into our work, and for the wonderful friendship our families share**.**

Chun Too – your kind and positive attitude could always brighten up an unlucky day in the lab!

Barbro Larsson, **Eva Jemseby, Gull-Britt Almgren, Gloria Rostvall** and **Julia Boström** – thank you for making our lab universe go round through all these years! Without you there would be chaos and we could barely find anything we could do research on!

Veronica Ebbersten Lindholm, Gunnel Bemerfeldt and **Susanne Karlfeldt –** for doing all the administrative heavy lifting**.** Thank you for always being there for us, making and maintaining our connection with the outside world!

Maria Rastas and all **the CMM administration –** for making sure we can get into work (even in the middle of the night)**!**

CMM-IT – please, don't think that we only appreciate you when things don't work! Thank you for keeping our digital hive buzzing securely!

Gustavo Nader – the life of the party! This lab will always miss your energy.

Lars Alfredsson's group – thank you for giving us the indispensible tools for unraveling the epidemiological part of our work, and teaching us how to use them. Your guidance has been instrumental to us through the years.

Ingrid Kockum's group – for all the productive discussions and valuable input.

Vivi Malmström and her group – thank you for the keen immunology expertise!

Per Johan's group – for providing good advice and proteomics knowledge.

Marie Wahren's group – you are all wonderful people!

Helena Erlandsson Harris – for making sure we all have a place in CMM**.**

Mohsen Khademi – for a fruitful collaboration and for your light-hearted character**.**

My students **Ivonne Pelaez** and **Caitlin Chasaide** – I learned so much together with you!

My dear **office-mates**, past and present – one just can't have a boring day in the office when you are around!

All **members of Rheumatology** and **Neroimmunology units** – since I became part of this wonderful working environment.

All my friends – for always so masterfully balancing out my work life!

My parents: Дорогие мои! Без ваших усилий, я не только не смог бы пройти этот немалый путь - я не мог бы и начать его! Наперекор нашим многочисленным испытаниям, вы всю дорогу поддерживали меня, не давая упасть, как бы сильно я не оступился. Я благодарю вас за ваши неисчерпаемые любовь и заботу и за свой каждый прожитый день!

My wife Iryna – the most dedicated collaborator, my fairest peer reviewer, and the most ruthless editor! This work demanded no less patience from you than it did from me. I have always been feeling that we were in it together.

My daughter Anna – for all the happiness you bring to my each and every day.

10 REFERENCES

1 Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, **12**, 1-8.

2 Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R. and Guigo, R. (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res*, **22**, 1616-1625.

3 Wang, Y., Ma, M., Xiao, X. and Wang, Z. (2012) Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat Struct Mol Biol*, **19**, 1044-1052.

4 Danoy, P., Wei, M., Johanna, H., Jiang, L., He, D., Sun, L., Zeng, X., Visscher, P.M., Brown, M.A. and Xu, H. (2011) Association of variants in MMEL1 and CTLA4 with rheumatoid arthritis in the Han Chinese population. *Ann Rheum Dis*, **70**, 1793-1797.

5 Jelen, N., Ule, J., Zivin, M. and Darnell, R.B. (2007) Evolution of Nova-dependent splicing regulation in the brain. *PLoS genetics*, **3**, 1838-1847.

6 Kafasla, P., Mickleburgh, I., Llorian, M., Coelho, M., Gooding, C., Cherny, D., Joshi, A., Kotik-Kogan, O., Curry, S., Eperon, I.C. *et al.* (2012) Defining the roles and interactions of PTB. *Biochem Soc Trans*, **40**, 815-820.

7 Lee, J.A., Tang, Z.Z. and Black, D.L. (2009) An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. *Genes Dev*, **23**, 2284-2293.

8 Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T. and Chabot, B. (2007) hnRNP proteins and splicing control. *Adv Exp Med Biol*, **623**, 123-147.

9 Kadener, S., Cramer, P., Nogues, G., Cazalla, D., de la Mata, M., Fededa, J.P., Werbajh, S.E., Srebrow, A. and Kornblihtt, A.R. (2001) Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *Embo J*, **20**, 5759-5768.

10 Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J. and Nogues, G. (2004) Multiple links between transcription and splicing. *Rna*, **10**, 1489-1498.

11 Allo, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S.A. *et al.* (2009) Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol*, **16**, 717-724.

12 Keren, H., Lev-Maor, G. and Ast, G. (2010) Alternative splicing and evolution: diversification, exon definition and function. *Nature reviews. Genetics*, **11**, 345-355.

13 Merkin, J., Russell, C., Chen, P. and Burge, C.B. (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science*, **338**, 1593-1599.

14 Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, **40**, 1413-1415.

15 Kent, W.J. and Zahler, A.M. (2000) The intronerator: exploring introns and alternative splicing in Caenorhabditis elegans. *Nucleic Acids Res*, **28**, 91-93.

16 Spartz, A.K., Herman, R.K. and Shaw, J.E. (2004) SMU-2 and SMU-1, Caenorhabditis elegans homologs of mammalian spliceosome-associated proteins RED and fSAP57, work together to affect splice site choice. *Mol Cell Biol*, **24**, 6811-6823.

17 Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev*, **17**, 419-437.

18 Buratti, E., Baralle, M. and Baralle, F.E. (2006) Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res*, **34**, 3494-3510.

19 Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dork, T. and Baralle, F.E. (2002) A new type of mutation causes a splicing defect in ATM. *Nat Genet*, **30**, 426-429.

20 Cartegni, L., Chew, S.L. and Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature reviews. Genetics*, **3**, 285-298.

21 Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y. and Darnell, R.B. (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, **25**, 359-371.

22 Roescher, N., Vosters, J.L., Alsaleh, G., Dreyfus, P., Jacques, S., Chiocchia, G., Sibilia, J., Tak, P.P., Chiorini, J.A., Mariette, X. *et al.* (2014) Targeting the splicing of mRNA in autoimmune diseases: BAFF inhibition in Sjogren's syndrome as a proof of concept. *Mol Ther*, **22**, 821-827.

23 Lamas, J.R., Rodriguez-Rodriguez, L., Tornero-Esteban, P., Villafuertes, E., Hoyas, J., Abasolo, L., Varade, J., Alvarez-Lafuente, R., Urcelay, E. and Fernandez-Gutierrez, B. (2013) Alternative splicing and proteolytic rupture contribute to the generation of soluble IL-6 receptors (sIL-6R) in rheumatoid arthritis. *Cytokine*, **61**, 720-723.

24 Chang, H.H., Tai, T.S., Lu, B., Iannaccone, C., Cernadas, M., Weinblatt, M., Shadick, N., Miaw, S.C. and Ho, I.C. (2012) PTPN22.6, a dominant negative isoform of PTPN22 and potential biomarker of rheumatoid arthritis. *PLoS One*, **7**, e33067.

25 Tolboom, T.C., Huidekoper, A.L., Kramer, I.M., Pieterman, E., Toes, R.E. and Huizinga, T.W. (2004) Correlation between expression of CD44 splice variant v8-v9 and invasiveness of fibroblast-like synoviocytes in an in vitro system. *Clin Exp Rheumatol*, **22**, 158-164.

26 Santiago, B., Izquierdo, E., Rueda, P., Del Rey, M.J., Criado, G., Usategui, A., Arenzana-Seisdedos, F. and Pablos, J.L. (2012) CXCL12gamma isoform is expressed on endothelial and dendritic cells in rheumatoid arthritis synovium and regulates T cell activation. *Arthritis Rheum*, **64**, 409-417.

27 Kojo, S., Tsutsumi, A., Goto, D. and Sumida, T. (2003) Low expression levels of soluble CD1d gene in patients with rheumatoid arthritis. *J Rheumatol*, **30**, 2524-2528.

28 Heinhuis, B., Koenders, M.I., van de Loo, F.A., Netea, M.G., van den Berg, W.B. and Joosten, L.A. (2011) Inflammation-dependent secretion and splicing of IL-32{gamma} in rheumatoid arthritis. *Proc Natl Acad Sci U S A*, **108**, 4962-4967.

29 Jin, P., Zhang, J., Sumariwalla, P.F., Ni, I., Jorgensen, B., Crawford, D., Phillips, S., Feldmann, M., Shepard, H.M. and Paleolog, E.M. (2008) Novel splice variants derived from the receptor tyrosine kinase superfamily are potential therapeutics for rheumatoid arthritis. *Arthritis Res Ther*, **10**, R73.

30 (2014). United States National Library of Medicine, in press.

31 (2014). United States National Library of Medicine, in press.

32 Barreiro, L.B., Laval, G., Quach, H., Patin, E. and Quintana-Murci, L. (2008) Natural selection has driven population differentiation in modern humans. *Nat Genet*, **40**, 340-345.

33 Nachman, M.W. (2001) Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet*, **17**, 481-485.

34 Marchini, J. and Howie, B. (2010) Genotype imputation for genome-wide association studies. *Nat Rev Genet*, **11**, 499-511.

35 Clarke, G.M., Anderson, C.A., Pettersson, F.H., Cardon, L.R., Morris, A.P. and Zondervan, K.T. (2011) Basic statistical analysis in genetic case-control studies. *Nat Protoc*, **6**, 121-133.

36 (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661-678.

37 Diogo, D., Kurreeman, F., Stahl, E.A., Liao, K.P., Gupta, N., Greenberg, J.D., Rivas, M.A., Hickey, B., Flannick, J., Thomson, B. *et al.* (2013) Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. *Am J Hum Genet*, **92**, 15-27.

38 Mitsunaga, S., Hosomichi, K., Okudaira, Y., Nakaoka, H., Suzuki, Y., Kuwana, M., Sato, S., Kaneko, Y., Homma, Y., Oka, A. *et al.* (2015) Aggregation of rare/low-frequency variants of the mitochondria respiratory chain-related proteins in rheumatoid arthritis patients. *J Hum Genet*, in press.

39 Bang, S.Y., Na, Y.J., Kim, K., Joo, Y.B., Park, Y., Lee, J., Lee, S.Y., Ansari, A.A., Jung, J., Rhee, H. *et al.* (2014) Targeted exon sequencing fails to identify rare coding variants with large effect in rheumatoid arthritis. *Arthritis Res Ther*, **16**, 447.

40 Hunt, K.A., Mistry, V., Bockett, N.A., Ahmad, T., Ban, M., Barker, J.N., Barrett, J.C., Blackburn, H., Brand, O., Burren, O. *et al.* (2013) Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature*, **498**, 232-235.

41 Ahlbom, A. and Alfredsson, L. (2005) Interaction: A word with two meanings creates confusion. *Eur J Epidemiol*, **20**, 563-564.

42 Andersson, T., Alfredsson, L., Kallberg, H., Zdravkovic, S. and Ahlbom, A. (2005) Calculating measures of biological interaction. *Eur J Epidemiol*, **20**, 575-579.

43 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**, 559-575.

44 Ding, B., Kallberg, H., Klareskog, L., Padyukov, L. and Alfredsson, L. (2011) GEIRA: gene-environment and gene-gene interaction research application. *Eur J Epidemiol*, **26**, 557-561.

45 Hahn, L.W., Ritchie, M.D. and Moore, J.H. (2003) Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics*, **19**, 376-382.

46 Chung, Y., Lee, S.Y., Elston, R.C. and Park, T. (2007) Odds ratio based multifactor-dimensionality reduction method for detecting gene-gene interactions. *Bioinformatics*, **23**, 71-76.

47 Yang, C.H., Lin, Y.D., Yang, C.S. and Chuang, L.Y. (2015) An efficiency analysis of high-order combinations of gene-gene interactions using multifactor-dimensionality reduction. *BMC genomics*, **16**, 489.

48 Shchetynsky K., P.L. (2014) Padyukov, L. (ed.), In *Between the Lines of Genetic Code. Genetic Interactions in Understanding Disease and Complex Phenotypes*. Elsevier, in press., pp. 137-150.

49 Majithia, V. and Geraci, S.A. (2007) Rheumatoid arthritis: diagnosis and management. *Am J Med*, **120**, 936-939.

50 Aletaha, D., Neogi, T., Silman, A.J., Funovits, J., Felson, D.T., Bingham, C.O., 3rd, Birnbaum, N.S., Burmester, G.R., Bykerk, V.P., Cohen, M.D. *et al.* (2010) 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum*, **62**, 2569-2581.

51 van Zeben, D., Hazes, J.M., Zwinderman, A.H., Cats, A., van der Voort, E.A. and Breedveld, F.C. (1992) Clinical significance of rheumatoid factors in early rheumatoid arthritis: results of a follow up study. *Ann Rheum Dis*, **51**, 1029-1035.

52 Rantapaa-Dahlqvist, S., de Jong, B.A., Berglin, E., Hallmans, G., Wadell, G., Stenlund, H., Sundin, U. and van Venrooij, W.J. (2003) Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum*, **48**, 2741-2749.

53 Edwards, J.C., Cambridge, G. and Abrahams, V.M. (1999) Do self-perpetuating B lymphocytes drive human autoimmune disease? *Immunology*, **97**, 188-196.

54 Marrack, P., Kappler, J. and Kotzin, B.L. (2001) Autoimmune disease: why and where it occurs. *Nat Med*, **7**, 899-905.

55 Alunno, A., Manetti, M., Caterbi, S., Ibba-Manneschi, L., Bistoni, O., Bartoloni, E., Valentini, V., Terenzi, R. and Gerli, R. (2015) Altered immunoregulation in rheumatoid arthritis: the role of regulatory T cells and proinflammatory th17 cells and therapeutic implications. *Mediators of inflammation*, **2015**, 751793.

56 Boissier, M.C., Semerano, L., Challal, S., Saidenberg-Kermanac'h, N. and Falgarone, G. (2012) Rheumatoid arthritis: from autoimmunity to synovitis and joint destruction. *Journal of autoimmunity*, **39**, 222-228.

57 Quandt, D., Rothe, K., Scholz, R., Baerwald, C.W. and Wagner, U. (2014) Peripheral CD4CD8 double positive T cells with a distinct helper cytokine profile are increased in rheumatoid arthritis. *PLoS One*, **9**, e93293.

58 Ma, L., Liu, B., Jiang, Z. and Jiang, Y. (2014) Reduced numbers of regulatory B cells are negatively correlated with disease activity in patients with new-onset rheumatoid arthritis. *Clinical rheumatology*, **33**, 187-195.

59 Doorenspleet, M.E., Klarenbeek, P.L., de Hair, M.J., van Schaik, B.D., Esveldt, R.E., van Kampen, A.H., Gerlag, D.M., Musters, A., Baas, F., Tak, P.P. *et al.* (2014) Rheumatoid arthritis synovial tissue harbours dominant B-cell and plasma-cell clones associated with autoreactivity. *Ann Rheum Dis*, **73**, 756-762.

60 Schellekens, G.A., Visser, H., de Jong, B.A., van den Hoogen, F.H., Hazes, J.M., Breedveld, F.C. and van Venrooij, W.J. (2000) The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. *Arthritis Rheum*, **43**, 155-163.

61 van Jaarsveld, C.H., ter Borg, E.J., Jacobs, J.W., Schellekens, G.A., Gmelig-Meyling, F.H., van Booma-Frankfort, C., de Jong, B.A., van Venrooij, W.J. and Bijlsma, J.W. (1999) The prognostic value of the antiperinuclear factor, anti-citrullinated peptide antibodies and rheumatoid factor in early rheumatoid arthritis. *Clin Exp Rheumatol*, **17**, 689-697.

62 Pober, J.S., Gimbrone, M.A., Jr., Lapierre, L.A., Mendrick, D.L., Fiers, W., Rothlein, R. and Springer, T.A. (1986) Overlapping patterns of activation of human endothelial cells by interleukin 1, tumor necrosis factor, and immune interferon. *J Immunol*, **137**, 1893-1896.

63 Lubberts, E. (2015) The IL-23-IL-17 axis in inflammatory arthritis. *Nat Rev Rheumatol*, in press.

64 Brennan, F.M., Maini, R.N. and Feldmann, M. (1992) TNF alpha--a pivotal role in rheumatoid arthritis? *Br J Rheumatol*, **31**, 293-298.

65 Williams, R.O., Feldmann, M. and Maini, R.N. (1992) Anti-tumor necrosis factor ameliorates joint disease in murine collagen-induced arthritis. *Proc Natl Acad Sci U S A*, **89**, 9784-9788.

66 Osborn, L., Hession, C., Tizard, R., Vassallo, C., Luhowskyj, S., Chi-Rosso, G. and Lobb, R. (1989) Direct expression cloning of vascular cell adhesion molecule 1, a cytokine-induced endothelial protein that binds to lymphocytes. *Cell*, **59**, 1203-1211.

67 Nam, J.L., Winthrop, K.L., van Vollenhoven, R.F., Pavelka, K., Valesini, G., Hensor, E.M., Worthy, G., Landewe, R., Smolen, J.S., Emery, P. *et al.* (2010) Current evidence for the management of rheumatoid arthritis with biological disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of RA. *Ann Rheum Dis*, **69**, 976-986.

68 Tarp, S., Bartels, E.M., Bliddal, H., Furst, D.E., Boers, M., Danneskiold-Samsoe, B., Rasmussen, M. and Christensen, R. (2012) Effect of nonsteroidal antiinflammatory drugs on the C-reactive protein level in rheumatoid arthritis: a meta-analysis of randomized controlled trials. *Arthritis Rheum*, **64**, 3511-3521.

69 Scott, D.L., Wolfe, F. and Huizinga, T.W. (2010) Rheumatoid arthritis. *Lancet*, **376**, 1094-1108.

70 Aaltonen, K.J., Virkki, L.M., Malmivaara, A., Konttinen, Y.T., Nordstrom, D.C. and Blom, M. (2012) Systematic review and meta-analysis of the efficacy and safety of existing TNF blocking agents in treatment of rheumatoid arthritis. *PLoS One*, **7**, e30275.

71 Edwards, J.C., Szczepanski, L., Szechinski, J., Filipowicz-Sosnowska, A., Emery, P., Close, D.R., Stevens, R.M. and Shaw, T. (2004) Efficacy of B-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. *N Engl J Med*, **350**, 2572-2581.

72 Singh, J.A., Christensen, R., Wells, G.A., Suarez-Almazor, M.E., Buchbinder, R., Lopez-Olivo, M.A., Ghogomu, E.T. and Tugwell, P. (2010) Biologics for rheumatoid arthritis: an overview of Cochrane reviews. *Sao Paulo medical journal = Revista paulista de medicina*, **128**, 309-310.

73 Miossec, P. and Kolls, J.K. (2012) Targeting IL-17 and TH17 cells in chronic inflammation. *Nature reviews. Drug discovery*, **11**, 763-776.

74 Burmester, G.R., Weinblatt, M.E., McInnes, I.B., Porter, D., Barbarash, O., Vatutin, M., Szombati, I., Esfandiari, E., Sleeman, M.A., Kane, C.D. *et al.* (2013) Efficacy and safety of mavrilimumab in subjects with rheumatoid arthritis. *Ann Rheum Dis*, **72**, 1445-1452.

75 O'Shea, J.J., Laurence, A. and McInnes, I.B. (2013) Back to the future: oral targeted therapy for RA and other autoimmune diseases. *Nat Rev Rheumatol*, **9**, 173-182.

76 MacGregor, A.J., Snieder, H., Rigby, A.S., Koskenvuo, M., Kaprio, J., Aho, K. and Silman, A.J. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum*, **43**, 30-37.

77 Frisell, T., Holmqvist, M., Kallberg, H., Klareskog, L., Alfredsson, L. and Askling, J. (2013) Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number and type of affected relatives, sex, and age. *Arthritis Rheum*, **65**, 2773-2782.

78 Stastny, P. (1978) Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. *N Engl J Med*, **298**, 869-871.

79 Gregersen, P.K., Silver, J. and Winchester, R.J. (1987) The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum*, **30**, 1205-1213.

80 Scally, S.W., Petersen, J., Law, S.C., Dudek, N.L., Nel, H.J., Loh, K.L., Wijeyewickrema, L.C., Eckle, S.B., van Heemst, J., Pike, R.N. *et al.* (2013) A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *The Journal of experimental medicine*, **210**, 2569-2582.

81 Lundstrom, E., Kallberg, H., Smolnikova, M., Ding, B., Ronnelid, J., Alfredsson, L., Klareskog, L. and Padyukov, L. (2009) Opposing effects of HLA-DRB1*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum*, **60**, 924-930.

82 Raychaudhuri, S., Sandor, C., Stahl, E.A., Freudenberg, J., Lee, H.S., Jia, X., Alfredsson, L., Padyukov, L., Klareskog, L., Worthington, J. *et al.* (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet*, **44**, 291-296.

83 Zheng, J., Ibrahim, S., Petersen, F. and Yu, X. (2012) Meta-analysis reveals an association of PTPN22 C1858T with autoimmune diseases, which depends on the localization of the affected tissue. *Genes Immun*, **13**, 641-652.

84 Stanford, S.M. and Bottini, N. (2014) PTPN22: the archetypal non-HLA autoimmunity gene. *Nat Rev Rheumatol*, **10**, 602-611.

85 Zheng, J., Petersen, F. and Yu, X. (2014) The role of PTPN22 in autoimmunity: learning from mice. *Autoimmun Rev*, **13**, 266-271.

86 Zhang, J., Zahir, N., Jiang, Q., Miliotis, H., Heyraud, S., Meng, X., Dong, B., Xie, G., Qiu, F., Hao, Z. *et al.* (2011) The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. *Nat Genet*, **43**, 902-907.

87 Festen, E.A., Goyette, P., Green, T., Boucher, G., Beauchamp, C., Trynka, G., Dubois, P.C., Lagace, C., Stokkers, P.C., Hommes, D.W. *et al.* (2011) A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. *PLoS genetics*, **7**, e1001283.

88 Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M. *et al.* (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*, **31**, 142-147.

89 Tiganis, T., Kemp, B.E. and Tonks, N.K. (1999) The protein-tyrosine phosphatase TCPTP regulates epidermal growth factor receptor-mediated and phosphatidylinositol 3-kinase-dependent signaling. *J Biol Chem*, **274**, 27768-27775.

90 Simoncic, P.D., Lee-Loy, A., Barber, D.L., Tremblay, M.L. and McGlade, C.J. (2002) The T cell protein tyrosine phosphatase is a negative regulator of janus family kinases 1 and 3. *Current biology : CB*, **12**, 446-453.

91 van Vliet, C., Bukczynska, P.E., Puryer, M.A., Sadek, C.M., Shields, B.J., Tremblay, M.L. and Tiganis, T. (2005) Selective regulation of tumor necrosis factor-induced Erk signaling by Src family kinases and the T cell protein tyrosine phosphatase. *Nature immunology*, **6**, 253-260.

92 Wiede, F., Shields, B.J., Chew, S.H., Kyparissoudis, K., van Vliet, C., Galic, S., Tremblay, M.L., Russell, S.M., Godfrey, D.I. and Tiganis, T. (2011) T cell protein tyrosine phosphatase attenuates T cell signaling to maintain tolerance in mice. *J Clin Invest*, **121**, 4758-4774.

93 Doody, K.M., Bourdeau, A. and Tremblay, M.L. (2009) T-cell protein tyrosine phosphatase is a key regulator in immune cell signaling: lessons from the knockout mouse model and implications in human disease. *Immunological reviews*, **228**, 325-341.

94 Suzuki, A., Yamada, R., Chang, X., Tokuhiro, S., Sawada, T., Suzuki, M., Nagasaki, M., Nakayama-Hamada, M., Kawaida, R., Ono, M. *et al.* (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet*, **34**, 395-402.

95 Kochi, Y., Suzuki, A., Yamada, R. and Yamamoto, K. (2010) Ethnogenetic heterogeneity of rheumatoid arthritis-implications for pathogenesis. *Nat Rev Rheumatol*, **6**, 290-295.

96 Kochi, Y., Thabet, M.M., Suzuki, A., Okada, Y., Daha, N.A., Toes, R.E., Huizinga, T.W., Myouzen, K., Kubo, M., Yamada, R. *et al.* (2011) PADI4 polymorphism predisposes male smokers to rheumatoid arthritis. *Ann Rheum Dis*, **70**, 512-515.

97 Scherer, H.U., van der Linden, M.P., Kurreeman, F.A., Stoeken-Rijsbergen, G., Cessie, S., Huizinga, T.W., van der Helm-van Mil, A.H. and Toes, R.E. (2010) Association of the 6q23 region with the rate of joint destruction in rheumatoid arthritis. *Ann Rheum Dis*, **69**, 567-570.

98 Thomson, W., Barton, A., Ke, X., Eyre, S., Hinks, A., Bowes, J., Donn, R., Symmons, D., Hider, S., Bruce, I.N. *et al.* (2007) Rheumatoid arthritis association at 6q23. *Nat Genet*, **39**, 1431-1433.

99 Graham, R.R., Cotsapas, C., Davies, L., Hackett, R., Lessard, C.J., Leon, J.M., Burtt, N.P., Guiducci, C., Parkin, M., Gates, C. *et al.* (2008) Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet*, **40**, 1059-1061.

100 Musone, S.L., Taylor, K.E., Lu, T.T., Nititham, J., Ferreira, R.C., Ortmann, W., Shifrin, N., Petri, M.A., Kamboh, M.I., Manzi, S. *et al.* (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat Genet*, **40**, 1062-1064.

101 Adrianto, I., Wen, F., Templeton, A., Wiley, G., King, J.B., Lessard, C.J., Bates, J.S., Hu, Y., Kelly, J.A., Kaufman, K.M. *et al.* (2011) Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat Genet*, **43**, 253-258.

102 Kato, M., Sanada, M., Kato, I., Sato, Y., Takita, J., Takeuchi, K., Niwa, A., Chen, Y., Nakazaki, K., Nomoto, J. *et al.* (2009) Frequent inactivation of A20 in B-cell lymphomas. *Nature*, **459**, 712-716.

103 Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. *et al.* (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376-381.

104 Remmers, E.F., Plenge, R.M., Lee, A.T., Graham, R.R., Hom, G., Behrens, T.W., de Bakker, P.I., Le, J.M., Lee, H.S., Batliwalla, F. *et al.* (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med*, **357**, 977-986.

105 Zheng, J., Yin, J., Huang, R., Petersen, F. and Yu, X. (2013) Meta-analysis reveals an association of STAT4 polymorphisms with systemic autoimmune disorders and anti-dsDNA antibody. *Human immunology*, **74**, 986-992.

106 Sigurdsson, S., Nordmark, G., Garnier, S., Grundberg, E., Kwan, T., Nilsson, O., Eloranta, M.L., Gunnarsson, I., Svenungsson, E., Sturfelt, G. *et al.* (2008) A risk haplotype of STAT4 for systemic lupus erythematosus is over-expressed, correlates with anti-dsDNA and shows additive effects with two risk alleles of IRF5. *Hum Mol Genet*, **17**, 2868-2876.

107 Shchetynsky, K., Protsyuk, D., Ronninger, M., Diaz-Gallo, L.M., Klareskog, L. and Padyukov, L. (2015) Gene-gene interaction and RNA splicing profiles of MAP2K4 gene in rheumatoid arthritis. *Clin Immunol*, **158**, 19-28.

108 Thalhamer, T., McGrath, M.A. and Harnett, M.M. (2008) MAPKs and their relevance to arthritis and inflammation. *Rheumatology (Oxford)*, **47**, 409-414.

109 Sweeney, S.E. and Firestein, G.S. (2007) Primer: signal transduction in rheumatic disease--a clinician's guide. *Nat Clin Pract Rheumatol*, **3**, 651-660.

110 Sundarrajan, M., Boyle, D.L., Chabaud-Riou, M., Hammaker, D. and Firestein, G.S. (2003) Expression of the MAPK kinases MKK-4 and MKK-7 in rheumatoid arthritis and their role as key regulators of JNK. *Arthritis Rheum*, **48**, 2450-2460.

111 Kant, S., Swat, W., Zhang, S., Zhang, Z.Y., Neel, B.G., Flavell, R.A. and Davis, R.J. (2011) TNF-stimulated MAP kinase activation mediated by a Rho family GTPase signaling pathway. *Genes Dev*, **25**, 2069-2078.

112 Hammaker, D. and Firestein, G.S. (2010) "Go upstream, young man": lessons learned from the p38 saga. *Ann Rheum Dis*, **69 Suppl 1**, i77-82.

113 Abdel-Nasser, A.M., Rasker, J.J. and Valkenburg, H.A. (1997) Epidemiological and clinical aspects relating to the variability of rheumatoid arthritis. *Semin Arthritis Rheum*, **27**, 123-140.

114 Uhlig, T., Hagen, K.B. and Kvien, T.K. (1999) Current tobacco smoking, formal education, and the risk of rheumatoid arthritis. *J Rheumatol*, **26**, 47-54.

115 Heliovaara, M., Aho, K., Aromaa, A., Knekt, P. and Reunanen, A. (1993) Smoking and risk of rheumatoid arthritis. *J Rheumatol*, **20**, 1830-1835.

116 Silman, A.J., Newman, J. and MacGregor, A.J. (1996) Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. *Arthritis Rheum*, **39**, 732-735.

117 Stolt, P., Bengtsson, C., Nordmark, B., Lindblad, S., Lundberg, I., Klareskog, L. and Alfredsson, L. (2003) Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis*, **62**, 835-841.

118 Stolt, P., Kallberg, H., Lundberg, I., Sjogren, B., Klareskog, L. and Alfredsson, L. (2005) Silica exposure is associated with increased risk of developing rheumatoid arthritis: results from the Swedish EIRA study. *Ann Rheum Dis*, **64**, 582-586.

119 Costenbader, K.H., Feskanich, D., Mandl, L.A. and Karlson, E.W. (2006) Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *Am J Med*, **119**, 503 e501-509.

120 Sverdrup, B., Kallberg, H., Bengtsson, C., Lundberg, I., Padyukov, L., Alfredsson, L. and Klareskog, L. (2005) Association between occupational exposure to mineral oil and rheumatoid arthritis: results from the Swedish EIRA case-control study. *Arthritis Res Ther*, **7**, R1296-1303.

121 Maxwell, J.R., Gowers, I.R., Moore, D.J. and Wilson, A.G. (2010) Alcohol consumption is inversely associated with risk and severity of rheumatoid arthritis. *Rheumatology (Oxford)*, **49**, 2140-2146.

122 Kallberg, H., Jacobsen, S., Bengtsson, C., Pedersen, M., Padyukov, L., Garred, P., Frisch, M., Karlson, E.W., Klareskog, L. and Alfredsson, L. (2009) Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. *Ann Rheum Dis*, **68**, 222-227.

123 Pattison, D.J., Symmons, D.P. and Young, A. (2004) Does diet have a role in the aetiology of rheumatoid arthritis? *Proc Nutr Soc*, **63**, 137-143.

124 Vessey, M.P., Villard-Mackintosh, L. and Yeates, D. (1987) Oral contraceptives, cigarette smoking and other factors in relation to arthritis. *Contraception*, **35**, 457-464.

125 Michou, L., Teixeira, V.H., Pierlot, C., Lasbleiz, S., Bardin, T., Dieude, P., Prum, B., Cornelis, F. and Petit-Teixeira, E. (2008) Associations between genetic factors, tobacco smoking and autoantibodies in familial and sporadic rheumatoid arthritis. *Ann Rheum Dis*, **67**, 466-470.

126 Lee, H.S., Irigoyen, P., Kern, M., Lee, A., Batliwalla, F., Khalili, H., Wolfe, F., Lum, R.F., Massarotti, E., Weisman, M. *et al.* (2007) Interaction between smoking, the shared epitope, and anti-cyclic citrullinated peptide: a mixed picture in three large North American rheumatoid arthritis cohorts. *Arthritis Rheum*, **56**, 1745-1753.

127 Pedersen, M., Jacobsen, S., Garred, P., Madsen, H.O., Klarlund, M., Svejgaard, A., Pedersen, B.V., Wohlfahrt, J. and Frisch, M. (2007) Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. *Arthritis Rheum*, **56**, 1446-1453.

128 Klareskog, L., Stolt, P., Lundberg, K., Kallberg, H., Bengtsson, C., Grunewald, J., Ronnelid, J., Harris, H.E., Ulfgren, A.K., Rantapaa-Dahlqvist, S. *et al.* (2006) A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared

epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum*, **54**, 38-46.

129 Kallberg, H., Padyukov, L., Plenge, R.M., Ronnelid, J., Gregersen, P.K., van der Helm-van Mil, A.H., Toes, R.E., Huizinga, T.W., Klareskog, L., Alfredsson, L. *et al.* (2007) Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. *Am J Hum Genet*, **80**, 867-875.

130 Qiao, B., Huang, C.H., Cong, L., Xie, J., Lo, S.H. and Zheng, T. (2009) Genome-wide gene-based analysis of rheumatoid arthritis-associated interaction with PTPN22 and HLA-DRB1. *BMC proceedings*, **3 Suppl 7**, S132.

131 Stolt, P., Yahya, A., Bengtsson, C., Kallberg, H., Ronnelid, J., Lundberg, I., Klareskog, L. and Alfredsson, L. (2010) Silica exposure among male current smokers is associated with a high risk of developing ACPA-positive rheumatoid arthritis. *Ann Rheum Dis*, **69**, 1072-1076.

132 Plenge, R.M., Seielstad, M., Padyukov, L., Lee, A.T., Remmers, E.F., Ding, B., Liew, A., Khalili, H., Chandrasekaran, A., Davies, L.R. *et al.* (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. *N Engl J Med*, **357**, 1199-1209.

133 Mitchell, M.K., Gregersen, P.K., Johnson, S., Parsons, R. and Vlahov, D. (2004) The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J Urban Health*, **81**, 301-310.

134 VanGuilder, H.D., Vrana, K.E. and Freeman, W.M. (2008) Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques*, **44**, 619-626.

135 Vang, T., Congia, M., Macis, M.D., Musumeci, L., Orru, V., Zavattari, P., Nika, K., Tautz, L., Tasken, K., Cucca, F. *et al.* (2005) Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet*, **37**, 1317-1319.

136 Arechiga, A.F., Habib, T., He, Y., Zhang, X., Zhang, Z.Y., Funk, A. and Buckner, J.H. (2009) Cutting edge: the PTPN22 allelic variant associated with autoimmunity impairs B cell signaling. *Journal of immunology*, **182**, 3343-3347.

137 Zhang, J., Zahir, N., Jiang, Q., Miliotis, H., Heyraud, S., Meng, X., Dong, B., Xie, G., Qiu, F., Hao, Z. *et al.* (2011) The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. *Nat Genet*, **43**, 902-907.

138 Cohen, S., Dadi, H., Shaoul, E., Sharfe, N. and Roifman, C.M. (1999) Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp. *Blood*, **93**, 2013-2024.

139 Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K. *et al.* (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet*, **44**, 1336-1340.

140 Stahl, E.A., Raychaudhuri, S., Remmers, E.F., Xie, G., Eyre, S., Thomson, B.P., Li, Y., Kurreeman, F.A., Zhernakova, A., Hinks, A. *et al.* (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet*, **42**, 508-514.

141 Okada, Y., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Kawaguchi, T., Stahl, E.A., Kurreeman, F.A., Nishida, N. *et al.* (2012) Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet*, **44**, 511-516.

142 Freudenberg, J., Lee, H.S., Han, B.G., Shin, H.D., Kang, Y.M., Sung, Y.K., Shim, S.C., Choi, C.B., Lee, A.T., Gregersen, P.K. *et al.* (2011) Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis Rheum*, **63**, 884-893.

143 Valledor, A.F., Sanchez-Tillo, E., Arpa, L., Park, J.M., Caelles, C., Lloberas, J. and Celada, A. (2008) Selective roles of MAPKs during the macrophage response to IFN-gamma. *J Immunol*, **180**, 4523-4529.

144 Halvorsen, M., Martin, J.S., Broadaway, S. and Laederach, A. (2010) Disease-associated mutations that alter the RNA structural ensemble. *PLoS genetics*, **6**, e1001074.

145 Davis, R.J. (2000) Signal transduction by the JNK group of MAP kinases. *Cell*, **103**, 239-252.

146 Kitamura, T., Sekimata, M., Kikuchi, S. and Homma, Y. (2005) Involvement of poly(ADP-ribose) polymerase 1 in ERBB2 expression in rheumatoid synovial cells. *Am J Physiol Cell Physiol*, **289**, C82-88.

147 Hallbeck, A.L., Walz, T.M., Briheim, K. and Wasteson, A. (2005) TGF-alpha and ErbB2 production in synovial joint tissue: increased expression in arthritic joints. *Scand J Rheumatol*, **34**, 204-211.

148 Papanikolaou, V., Iliopoulos, D., Dimou, I., Dubos, S., Kappas, C., Kitsiou-Tzeli, S. and Tsezou, A. (2011) Survivin regulation by HER2 through NF-kappaB and c-myc in irradiated breast cancer cells. *J Cell Mol Med*, **15**, 1542-1550.

149 Yamaoka, T., Yan, F., Cao, H., Hobbs, S.S., Dise, R.S., Tong, W. and Polk, D.B. (2008) Transactivation of EGF receptor and ErbB2 protects intestinal epithelial cells from TNF-induced apoptosis. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11772-11777.