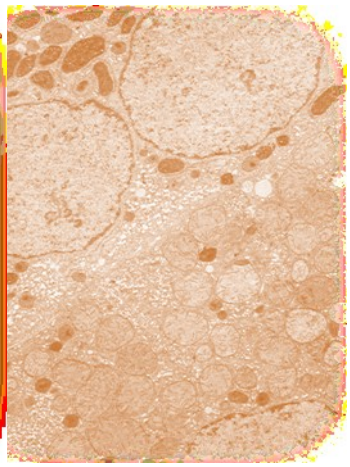


Investigations of genetic and environmental risk factors in rheumatoid arthritis

Yngvild Voldhaug Storlykken

2012/2013

Oslo University Hospital

OSLO AND AKERSHUS UNIVERSITY COLLEGE OF APPLIED SCIENCES

Thesis submitted for the Master degree (60 ECTS):

By

Yngvild Voldhaug Storlykken

Master of Biomedicine

Faculty of Health Sciences

2012

Department of medical genetics,

Division of diagnostics and intervention,

Institute of clinical medicine

Oslo University Hospital

Acknowledgements

The project presented in this thesis was carried out at the Institute for Clinical Medicine, Department of Medical Genetics, Oslo University Hospital (Ullevål), during the period August 2012 to May 2013, for Master of Biomedicine at Oslo and Akershus University College of Applied Science, Faculty of Health Sciences. This project has been founded by the Research council of Norway.

First of all, special thanks to my supervisor Benedicte Lie, who included me in the research group and gave me the opportunity to carry out this exciting project. Thank you for all the good advice, and useful and rapid feedbacks, even on holidays. I want you to know that I find your capacity and commitment to research very inspiring!

I will also thank my co-supervisor Marthe Mæhlen, for introducing me to the present knowledge of rheumatoid arthritis and for all help and support. Siri T. Flåm, thank for your guidance in the laboratory, for always being available to help me, and for your joyful and contagious humour! Marte K. Viken, thank you for all help, technical guidance and for interesting discussions. Ingvild Gabrielsen, thank you for always cheering for me! I will also express my gratitude for the warm welcome I got at the Department of Medical Genetics.

When it seemed like it could not have been any better at Ullevål, I was lucky enough to share an office with two wonderful master students, Aase Marie C. Kiel and Maria D. Vigeland. Thank you for frequent breaks filled with interesting conversations!

Least but not at last, thanks to my family and friends for encouragement and support during all the years I have been studying.

I want to dedicate my thesis to my grandmother, Oddbjørg Helene Voldhaug, who passed away 28th of December 2012. As a little girl she was very clever, and her teacher recommended that she should continue school after finishing the obligatory years. Even though she wanted this, her working force was needed at her parents little farm, as she was the oldest among the siblings. I know my grandmother felt sorry for this, and it reminds me that I should not take my opportunity to study for granted. Thank you "mor", for reminding me of this, and thank you for all our great moments!

Abstract

Rheumatoid arthritis is an autoimmune disorder where the immune system attacks the small joints in hands and feet, often leading to bone destruction. This leads to pain and reduced life quality for the patients, and comorbidities like cardiovascular disorders, other autoimmune disorders, fatigue and depression are often observed.

Now, we know that ~60% of disease development is caused by genetic contribution. 45 riskpolymorphisms are demonstrated, with small impact on their own and together their contribution is estimated to 15% of the genetic risk. Polymorphisms in the *HLA* gene region are estimated to contribute to additional 11-37% of the genetic risk. Environmental factors are also necessary for disease development, but except for smoking, little is known about the environmental risk factors.

There is still much we do not know about the development of rheumatoid arthritis, and in this thesis we wanted to find out more about the genetic and environmental contribution in the Norwegian RA population.

By the use of different cohorts, 950 patients and 1121 controls were included, and genotyping of 35 single nucleotide polymorphisms newly reported associated in other rheumatoid arthritis populations was carried out, and a questionnaire was analysed to increase our knowledge regarding environmental risk factors. In this study, 11 polymorphisms were confirmed to be associated with rheumatoid arthritis. Smoking, periodontitis and coffee were found to increase the risk of develop rheumatoid arthritis, and alcohol, pets and domestic animals during childhood, mononucleosis and breastfeeding 13 months or more were found to significantly decrease the risk.

Further investigation is needed to map out genetic changes and environmental risk factors for rheumatoid arthritis. Studies of immunological functions to increase knowledge regarding interplay between genetics and environmental factors will also be of importance. This knowledge is important for development of new and better medicines, and to diagnose patients at an earlier stage, so that treatment can be started earlier. It is also important to increase the knowledge of how risk of disease development can be reduced by avoiding certain environmental risk factors, especially for genetic susceptible individuals.

Sammendrag

Leddgikt er en autoimmun sykdom, hvor immunforsvaret i hovedsak går til angrep på bein og brusk i små ledd i fingre og føtter. Dette fører til smerte og redusert livskvalitet hos pasientene, og i tillegg er sykdommer som hjerte-karsykdommer, andre autoimmune sykdommer, tretthet og depresjoner ofte observert.

I dag vet man at ~60% av sykdommen skyldes genetisk disposisjon. Det er funnet 45 risiko-polymorfismer i gener som hver for seg har liten effekt på risikobidraget, og er til sammen beregnet til å forklare15% av det genetiske risikobidraget. I tillegg bidrar polymorfismer i *HLA* med om lag 11-37% av det genetiske risikobidraget. Miljøfaktorer er også nødvendig for utvikling av leddgikt, men bortsett fra at røyk er vist å gi økt risiko for sykdomsutvikling, har man begrenset kunnskap på dette feltet.

Det er mye man ennå ikke vet om utviklingen av leddgikt, og i denne oppgaven har vi forsøkt å finne ut mer om hvilke genetiske risikobidrag som finnes i den norske leddgikt populasjonen, og hvilke miljøfaktorer som bidrar til utvikling av sykdommen.

Vi benyttet datamateriale bestående av 950 pasienter og 1121 kontroller, for genotyping av 35 enkle nukleotid polymorfismer nylig funnet assosiert med leddgikt i andre populasjoner, og analysering av spørreskjema for å øke kunnskapen angående miljøfaktorer som bidrar til leddgikt. I denne studien ble 11 polymorfismer funnet signifikant assosiert i den norske leddgikt populasjonen. Røyking, periodontitt og kaffe ble funnet å øke risikoen for leddgikt, mens alkohol, dyrehold under oppveksten, mononukleose og å amme i 13 måneder eller lengre ble funnet assosiert med signifikant redusert risiko.

Videre studier for å kartlegge flere genetiske endringer og miljøfaktorer som øker risikoen for leddgikt er nødvendig. Det blir også viktig å studere de immunologiske funksjonene for å finne sammenhengen mellom arv og miljø. Denne kunnskapen vil være nyttig for utviklingen av nye og bedre medisiner for leddgiktspasienter, for å fange opp pasientene tidligere slik at behandling kan påbegynnes på tidligere stadium, og for å øke kunnskapen om hvordan utviklingen av leddgikt kan forhindres ved å unngå enkelte miljøfaktorer.

ABBIREVATIONS

ACPA Anti-citrullinated protein antibodies

ACR American College of Rheumatology

AID Autoimmune disease

APC Antigen presenting cell

bp base pair

CNV Copy number variation

CPA Citrullinated protein/peptide antigen

ddNTP dideoxy nucleotide tri-phosphate

DMARD disease-modifying antirheumatic drug

DNA deoxyribonucleic acid

EULAR European League Against Rheumatism

Exo-SAP Exonuclease1-Shrimp alkaline phosphatase

GEMS Genes and Environment in multiple sclerosis

GSR Genotype success rate

GWAS Genome wide association study

HLA Human leukocyte antigen

HUNT Nord-Trøndelag Health Study

HWE Hardy-Weinberg Equilibrium

Ig Immuoglobulin

IL Interleukin

kb kilo base

LD Linkage disequilibrium

MAF Minor allele frequency

MALDI-TOF-MS Matrix- assisted laser desorption ionization- time of flight mass spectrometry

NSAID Non-steroid anti-inflammatory drugs

OR Odds ratio

PAD Peptidyl arginine deimenase

PCR Polymerase chain reaction

RA Rheumatoid arthritis

REK Regional Ethical Committee

RF Rheumatoid factor

SE Shared epitope

SNP Single nucleotide polymorphism

STR Short tandem repeats

TE Tris-EDTA

WGA Whole-genome amplified

Table of Contents

ABBIREVATIONS 1
1. INTRODUCTION 5
1.1 The immune system 5
1.2 Autoimmune disorders 6
1.3 Rheumatoid arthritis 6
1.3.1 Diagnostic criteria 7
1.3.2 Autoantibodies 8
1.3.3 Etiology 9
1.3.4 Pathogenesis 13
1.3.5 Treatment 14
1.4 Human genetic variation 15
1.4.1 Single nucleotide polymorphism 16
1.4.2 Linkage disequilibrium 16
1.5 Studying AID genetics 17
2. AIM OF STUDY 19
3. MATERIAL AND METHODS 20
3.1 Materials 20
3.2 Investigating genetic risk factors 21
3.2.1 SNP selection 21
3.2.2 TaqMan genotyping 23
3.2.3 Genotyping using MassARRAY 27
3.3 Sequencing to resolve unexpected genotype result 29
3.4 Questionnaire- Investigate environmental risk factors 34
3.5 Statistical analysis 36
3.6 Bioinformatic tools used 38
4. RESULTS 39
4.1 Control of genotyping quality 39
4.2 Genotyping results of association analysis 42
4.3 Sequencing revealed SNP interfering with TaqMan result 47
4.4 Environmental risk factors associated with RA 48
5. DISCUSSION 54
5.1 Several RA risk loci confirmed in the Norwegian RA population 54

5.2 Environmental factors associated with RA.....58
5.3 Methodological considerations.....62
5.3.1 Quality of genotyping results.....62
5.3.3 Sequencing.....64
5.3.3 Obtaining information by the use of questionnaire.....64
5.4 Missing heritability.....66
5.5 Conclusion.....67
5.6 Further investigation.....67
6. REFERENCE LIST.....68
Appendix.....74

1. INTRODUCTION

1.1 The immune system

The immune system consists of surface barriers and inner barriers. The surface barriers consist of mechanical barriers like the skin and mucous membranes, and chemical barriers like enzymes, saliva and tears. Not many microorganisms manage to get through there barriers, but if they do- the inner defence will fight to get rid of them (1).

The inner defence consists of the innate- and the adaptive immune system (Figure 1). The innate immune system consists of complement proteins and effector cells like phagocytic cells, dendritic cells, mast cells and natural killer cells. Complement proteins marks pathogens, which is thereby recognized and eliminated by effector cells.

The adaptive immune system consists of B- lymphocytes which produce antibodies and T-lymphocytes with receptors that can identify and bind foreign structures on different pathogens. Antigen presenting cells (APC) present antigens from pathogens to T-cells, and thereby initiate activation of the adaptive immune response, specifically aimed towards the pathogen. Some of the lymphocytes also develop to memory T-cells, which provide a more rapidly response to the same pathogen next time it enters the body (1).

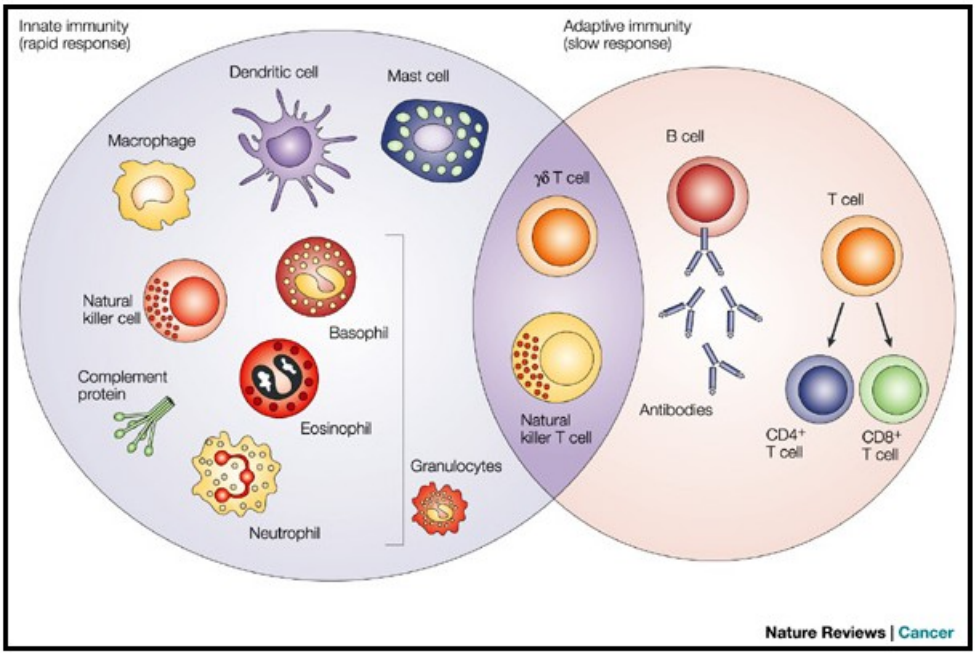


Figure 1. Cells of the innate- and adaptive immune system. (2)

1.2 Autoimmune disorders

Autoimmune disorders (AID) are caused by an inappropriate immune response towards self; the immune system interprets healthy tissue and cells of the body as foreign structures, and initializes an immune response. AID vary widely in the symptoms they cause and the tissues being attacked, some targeting a particular organ or cell type (e.g. Addison`s disease in which the adrenal gland do not produce sufficient steroid hormones and Type 1 diabetes caused by destruction of insulin-producing beta-cells of the pancreas) while others act systemically (e.g. systemic lupus erythematosus- a connective tissue disease that can affect any part of the body and rheumatoid arthritis (RA) that primarily affect joints but may affect many tissues and organs) (3). The prevalence of AID was estimated to 5.3% in Western countries, published by Eaton et al in 2007(4), and the underestimation was corrected by Cooper et al in 2009 who estimated the prevalence to be 7.6-9.4% (5). The incidence may have increased due to more precise diagnostic criteria which improve the number of patients correctly diagnosed with AID. Comorbities as additional AID is often observed in patients with AID, and this might indicate that different AID share genetic risk loci, which are confirmed by studies of genetic contribution to AID.

1.3 Rheumatoid arthritis

Rheumatoid arthritis is characterized by joint inflammation which strike synovial membrane, bone and cartilage. The chronic inflammation is thought to be initialised by an environmental factor and/or trauma- most likely years before symptoms of RA can be seen. Environmental risk factors are thought to trigger and/or catalyse the presentation of autoantigens and thereby activate the immune response towards self. APC with human leukocyte antigen (HLA)-class II molecules, present antigen to- and thereby activate naïve T-cells. Activated T-cells produce cytokines and chemokines, and stimulates B-cell activation (6). B-cells contribute to RA development by producing antibodies and cytokines. Pro-inflammatory cytokines (e.g. interleukins) are involved in RA pathogenesis by activating genes associated with inflammatory responses, through complex signal pathways. T- and B- cell activation results in increased production of cytokines and chemokines, leading to a vicious circle of additional activation of T-cell, machrophage and B-cell thought to cause a chronic inflammation in the joints and development of RA.

The inflammation leads to stiffness, pain and swelling in the area surrounding the joints. It mainly affects small joints in hands and feet, but may also affect larger joints like shoulders

and elbows, and inflammatory tissue leads to erosion and destruction when it grows into cartilage and bone.

RA usually appears in the age of 40-60, but can appear in all ages, and 2/3 of the affected are women (7). The prevalence of RA is 0.5-1% in the Norwegian population (7) as in Northern Europe and North America (8). The expansion varies geographically; a tendency of a North-to-South gradient in Europe is observed, with the highest prevalence in Northern countries (8). Higher prevalence has been reported for some Native American tribes (9), and lower prevalence has been reported for African and Asian populations (10, 11). It is also important to remember that in addition to the actual variation in prevalence and incidence across populations, some of the variation reported may be due to underestimation, statistical methods and differences in disease definitions.

Even though there are huge individual differences, it is observed that 1/3 of RA patients are work disabled within the first five years (12), many patients suffer from comorbidities (13), and death caused by cardio vascular diseases is 50% greater in RA patients than in the general population (14).

1.3.1 Diagnostic criteria

RA was classified as a disease in 1859 (1, 15), and has until recently been classified with seven criteria produced by American rheumatologists in the mid 1980s (Table A1, appendix) (16). Because at least two of the seven criteria (nodules and erosions) are generally not present at the best time for early diagnosis and initiation of treatment (17), new RA criteria were recently developed by the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR) (Table 1) (18) to ensure early diagnosis and treatment which is important to reduce long term damage and disability.

Table 1. The 2010 American College of Rheumatology / European League Against Rheumatism classification criteria for rheumatoid arthritis*

	Score
A. Joint involvement	
1 large joint¶	0
2-10 large joints	1
1-3 small joints (with or without involvement of large joints)#	2
4-10 small joints (with or without involvement of large joints)	3
>10 joints (at least 1 small joint)	5
B. 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 result is needed for classification)‡‡	
Normal C-reactive protein *and* normal erythrocyte sedimentation rate	0
Abnormal C-reactive protein *or* abnormal erythrocyte sedimentation rate	1
D. Duration of symptoms§§	
<6 weeks	0
≥6 weeks	1

Patients who have at least one joint with definite clinical synovitis (swelling), not better explained by another disease should be tested. *Classification criteria for RA: score-based algorithm: add score of categories A–D; a score of ≥6/10 is needed for classification of a patient as having definite RA. ¶Shoulders, elbows, hips, knees, and ankles. #Metacarpophalangeal joints, proximal interphalangeal joints, second through fifth metatarsophalangeal joints, thumb interphalangeal joints, and wrists. ‡‡Normal/abnormal is determined by local laboratory standards. §§Patient self-report of the duration of signs or symptoms of synovitis (e.g. pain, swelling, tenderness) of joints that are clinically involved at the time of assessment, regardless of treatment status.

1.3.2 Autoantibodies

Rheumatoid factor (RF) and anti-citrullinated protein antibodies (ACPA), are both antibodies produced by the immune system, directed against individuals own proteins. Growing evidence show that grouping RA patients with regard to antibody-status is of great importance (17). Autoantibodies are good markers for RA, as they appear in the body years before the degradation of bone and cartilage starts. Presence or absence of RF was the classic way of dividing patients groups. RF are antibodies produced by the immune system directed against the Fc fragment of antibodies of immunoglobulin G (IgG) class. Most of the RF positive patients are also positive for antibodies to citrullinated proteins. Citrullination is a

posttranslational process, by which peptidylarginine is deaminated to peptidylcitrulline by the enzyme peptidyl arginine deiminase (PAD) (19). One positive charge is lost for every arginine residue converted to a neutral citrulline, and could lead to altered protein folding and exposure of cryptic epitopes.

ACPA seems more specific and sensitive for diagnosis and prognostic features than RF (20), and is found in approximately 60% of RA patients. ACPA is quite rare in other inflammatory diseases and exists in only ~2% of the general population (21), and RF is detected in 1-4% of young people and increase with age in the general population (22).

Genetic studies demonstrate differences in the genetic risk profile between ACPA positive and ACPA negative patients. Most loci correlated to RA have been identified in ACPA positive patient populations (among these human leukocyte antigen (*HLA)-DRB1-SE* and *PTPN22*), and less is known about the genetic contribution to ACPA negative disease. ACPA status is also associated with specific environmental risk factors (e.g. smoking) (23). In addition to differences in genetic and environmental risk factors, the two subgroups differ clinically with regard to severity, and ACPA positive patients have a more rapid disease course with progressive joint damage and low remission rate (23). This indicates that ACPA positive and ACPA negative patients probably differ in pathogenesis, and therefore should be studied separate in both genetic and functional studies.

1.3.3 Etiology

RA is a complex disorder, in which multiple genes and environmental risk factors are necessary for disease development. The etiology is largely unknown, but several studies have demonstrated that genetic factors are important contributors for RA development. λ_s is a measure of familial clustering, and is calculated to be 8 for RA (24). This means that a sibling of an affected individual has eight times greater risk of developing RA than a member of the general population. The genetic contribution of RA is also confirmed by twin studies, by showing an elevation in concordance ratio in monozygotic twins (~15%) compared to dizygotic twins (~4%) (25). Estimates based on data from twin concordance rate studies show that genetic factors account for ~ 60% of the risk of RA development (26).

Genetic contribution

The "common disease/common variation" hypothesis state that alleles common in the general population at a handful of loci, interact to cause disease (27). Many risk alleles predisposing to RA are fairly common in the general population, and have modest effect on disease development on their own (28). Except *HLA*-genes which contribute to 11-37% of the estimated heritability (29), non-HLA loci are estimated to ~15% of the heritability of RA.

The *HLA* complex is located on chromosome 6, and contains genes coding for immunologically important molecules, including the antigen-presenting *HLA*-molecules. *HLA-class II* molecules expressed on APC, present antigen-peptides from outside the cell to T-lymphocytes with CD4 proteins on the cell surface. Binding of T-lymphocyte receptor to *HLA-class II* molecules stimulates development of T helper cells, which in turn increase the production of cytokines and activates B-cells to produce antibodies.

Gregersen et al have described a shared amino acid sequence at position 70-74 of the *HLA-DRB1* protein (*HLA-class II*) (30). This is known as the "shared epitope" (SE) because of the related sequence composition of the third hyper-variable region of all RA predisposing *DRB1* alleles. Much of the risk attributed to *HLA* is associated with variations at *HLA-DRB1 SE*.

In addition to *HLA*, 45 loci associated with RA have been reported at genome-wide significance level ($p<5*10^{-8}$)(31). Most of these polymorphisms are non-coding variants (32), labelled with names of the most compelling candidate gene(s) from each region of linkage disequilibrium (LD). These genes are generally immune-related genes, involved in for example innate immune pathways, T-cell differentiation and immune cell signalling. Several SNPs associated with RA are also associated with other AID (33).

PTPN22 was the second confirmed RA susceptibility gene. *PTPN* encodes a tyrosine phosphate and plays a part in T- and B-cell intracellular signalling. Except for *PTPN22*, confirmed non-HLA loci do not contribute much to the genetic load on their own, but together they are estimated to explain 15% of the heritability (31).

Most risk alleles found associated to RA are associated with ACPA positive RA, including *HLA-DRB1 SE* and *PTPN22*. Because only ~50% of the heritability can be explained by confirmed loci, more common risk alleles with modest effect sizes remain to be identified, particularly for ACPA negative disease. In addition to investigate such risk alleles, the roles of rare variants, copy number variants and epigenetic modifications will need to be explored.

Environmental contribution

An observed concordance rate less than 100% in monozygotic twins (25) is taken as evidence that genetics do not account for RA development alone. Smoking is the only environmental risk factor for RA development that has been extensively studied and widely accepted. Several studies have been carried out, investigating multiple other potential environmental factors, but few conclusive results have been obtained.

Smoking is hence the best established environmental risk factor, and cigarette smokers have an increased risk of developing ACPA positive (and RF positive (34)) RA, compared with never-smokers (23). The risk of RA increases in a dose dependent manner, and furthermore smokers who carry the SE have much higher risk compared to non-smokers who do not carry the SE (35). Hence, there appears to be a gene-environment interaction between smoking and the *HLA* encoded risk.

The "hygiene hypothesis" was first put forward in 1989 (36). Early-life infection was proposed to reduce allergic diseases, and the hypothesis has extended to include AID. The mechanism of how exposure to infection protects against allergy and AID is not known, but it is thought that early childhood exposure to infectious agents, microorganisms and parasites better "prime" the immune system, and lack of such infections might suppress natural development of the immune system (37).

In contrast to the expanded "hygiene hypothesis", observations of transient arthropathy caused by infectious agents (38) and of viral presence in synovial fluid of RA patients (39), make infectious agents also interesting candidates as environmental risk factors. The theoretical possibility that foreign structures can cause cross-activation of autoreactive T- or B cells because of sequence similarities between microbial agents and self- antigens is called "molecular mimicry" (40). It is believed that the "peptide mimic" is responsible for the production of autoantibodies, thus leading to autoimmunity. The role of infectious agents to cause RA is still controversial. Recent years, the interest regarding *Porphyromonas gingivalis,* has increased. This is due to the reported increased prevalence of periodontitis mainly caused by *P. Gingivalis,* in RA patients compared to the general population (41). This bacterium expresses the PAD enzyme, which citrullinates proteins, (as observed locally activated in some individuals during smoking, Figure 2 on page 14), and might influence RA development by similar mechanisms. Another infectious agent proposed in RA is the Epstein- Barr virus causing mononucleosis, which is also proposed as a risk factor for multiple sclerosis (42).

As two thirds of the RA patients are women and disease improvement during pregnancy is reported, it is suggested that hormones may play a role in the pathogenesis (39). Studies on the use of oral contraceptives have reported ambiguous results and extended breastfeeding (≥13 months) has been found associated with a significant reduction of the risk of RA (23, 43). One study noted that early menarche decreases the risk of RA (39), and another study indicated that early menopause increases risk of RA (44). No difference in the sex hormone levels in women with RA and healthy controls has been observed, but male sex hormone levels in men with RA have been found decreased (39). No conclusive results regarding hormone levels and the underlying mechanisms have been able to explain why females are more affected than males, and this theme needs further investigation.

Among factors suggested to decrease RA risk is alcohol consumption, where a dose-dependent effect has been observed (increased alcohol consumption decreases the risk of develop RA), and carriers of the SE was found to have a more pronounced risk reduction (45).

Another important aspect is geography, as a North-to-South gradient in Europe is observed, with more people affected by RA in Northern countries. This may be due to limited access to the sun, and thereby limited amount of Vitamin D, which has been proposed as a risk factor for RA development. One might expect increased risk of RA with low vitamin D consumption and/or production, because of its essential role for bone and mineral homeostasis, and as a suppressor of pro-inflammatory response, but Vitamin Ds role in decreasing the risk of RA remains equivocal (39).

A period of fasting followed by a regimented vegetarian diet can decrease disease activity (23), and this led to the investigation of intake of protein- and red meat. No association between RA risk and amount of protein, red meat, poultry and fish consumption has been shown (23).

Studies on socioeconomic status (education and occupational class), implicate that people with longest education, compared with those with the lowest level of education, have reduced risk of RA. It has also been observed an increased risk for patients whose occupation required manual labour, compared with non-manual workers (23).

In relation to the environmental factors contributing to RA development, time of exposure is of importance for the effect of the risk factor. The inflammation and disease development is

believed to start several years before the symptoms of RA can be seen, and could result from exposure of a risk factor when you were younger.

Epigenetic contribution

A classic way of defining epigenetic is "heritable changes in gene expression patterns that are not caused by changes in the primary DNA sequence" (46). Post-translational modifications determine the accessibility of the DNA, and therefore the ability of transcription factors to bind and initiate gene expression. The changes include a variety of modifications on histones, like acetylation which loosens up the chromatin structure and activate expression. DNA methylation is another epigenetic change, where methylation of cytosine located in CpG islands represents a biological mechanism for reduce gene expression. These marks are not stable and can rapidly change in response to stimulus, and dysfunction of this system can lead to disease.

Influences of environmental factors and ageing on the epigenome are possible explanations for age related autoimmune diseases like RA (46). Study of chronic exposure to cigarette smoke in rats showed reduced expression of histone deacetylases (47), which removes acetyl groups from lysine residues in the histone tails (46). Reduced removal of acetyl groups can lead to increased transcription of pro-inflammatory cytokines, involved in RA development. Another example of epigenetic changes reported in RA is methylation of CpG islands surrounding the transcription start site of death receptor 3. The methylation probably explains the reduced expression of death receptor 3, observed in synovial cells from RA patients compared to patients suffering from osteoarthritis (48). This provides a link between altered DNA methylation pattern and resistance to apoptosis seen in synovial cells of RA patients.

All together, these are examples of how our knowledge about epigenetic influence increases the understanding of RA pathogenesis. More detailed studies on how disease related genes are epigenetically regulated, and the interactions of environmental factors are needed.

1.3.4 Pathogenesis

The pathogenesis of RA is largely unknown, and is thought to vary between the two autoantibody defined subgroups, in regard to environmental factors and genetic disposition. Our knowledge of underlying risk factors is very limited for ACPA negative disease, but

researchers have established a hypothetical development of ACPA positive RA (Figure 2) (49).

Long-term smoking is thought to locally activate PAD enzymes, which further causes citrullination of proteins in the lungs. Activation of APC in genetically predisposed individuals (carrying *HLA-DRB1 SE*), leads to presentation of citrullinated proteins and thereby activation of T-cells, which further activates B-cells and produce of antibodies to citrullinated proteins. A second inflammatory event possibly triggered by an additional factor occurs in the synovium. A vicious circle of local activation of PAD enzymes which citrullinates proteins and increases the activation of T- and B- cells is thought to cause a chronic inflammation in the joints, and development of RA.

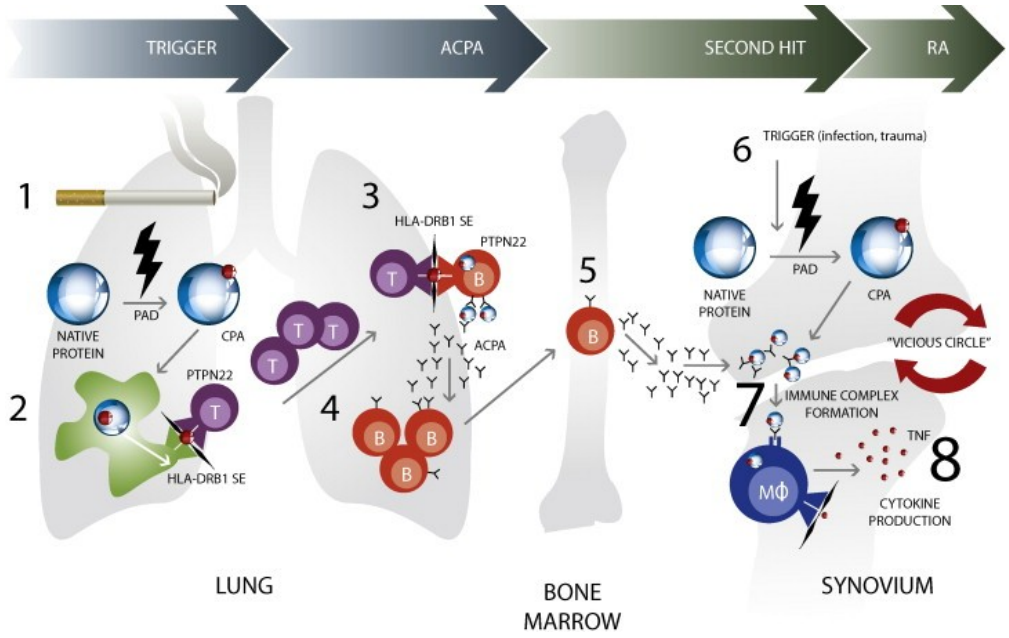


Figure 2. Hypothetical evolution of ACPA positive RA triggered by long-time smoking. (49) CPA= citrullinated protein/peptide antigen

1.3.5 Treatment

The ultimate goal of treatment is remission or sustained low disease with reduced pain and maintenance of function. For patients to have the best possible effect and to increase the probability for remission, early diagnosis and treatment is extremely important.

There are several possible treatments of RA, including analgesics and non-steroid anti-inflammatory drugs (NSAIDs) which reduce pain and stiffness, disease-modifying antirheumatic drugs (DMARDs), biological agents, and non-drug treatments (28). New

insight into various molecular pathways in the RA development have contributed to develop efficient treatment approaches, but we still need to find even better ways to specifically target these drugs to the right individuals at the right time (17).

1.4 Human genetic variation

The human genome is organized in a molecular structure called deoxyribonucleic acid (DNA), which is built up of four different bases; adenine, thymine, cytosine and guanine (A, T, C and G, respectively), sugar and phosphate (Figure 3). All human cells (with a few exceptions, e.g. red blood cells) contain DNA, packed in 46 chromosomes; 22 autosome pairs (1-22) and two sex chromosomes (XX/XY). Each chromosome consists of genes with exons (coding) and introns (noncoding), and intergenic regions (previously called "junk") which are thought to play a role in regulation of transcription.

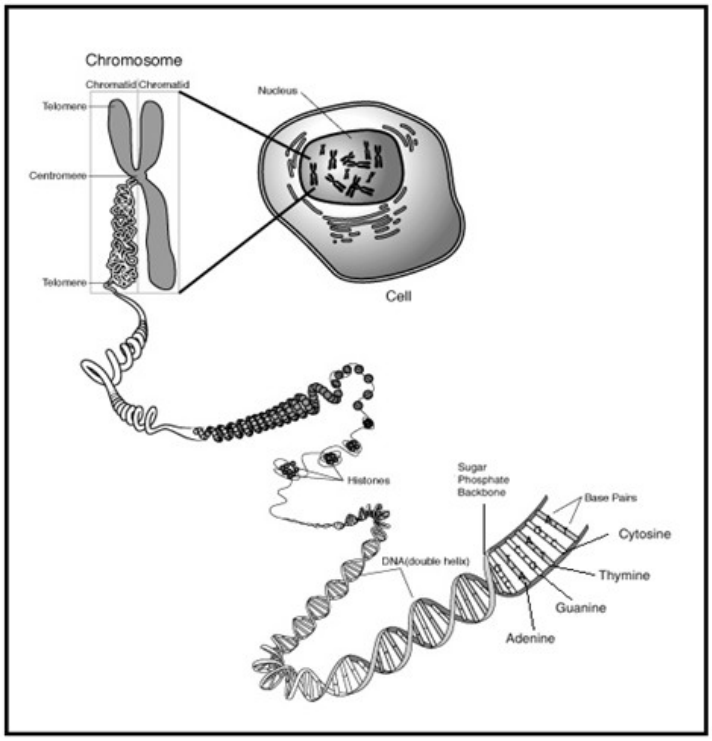


Figure 3. **DNA structure and organisation of the human genome**. DNA built up by four different bases, sugar and phosphate, packed in chromosomes located in the nucleus of the cell (50).

The human genome consists of 3000 megabases, and 21-23000 genes. Less than 1% of the DNA sequence varies between two individuals, and the variations can be classified with regard to their size and composition (Table 2).

Table 2. Some forms of human genetic variation. Reference sequence in the first row, variations seen in the DNA listed with a description in the following rows.

	Sequence	Description
Reference sequence	A-B-a-t-c-t-t-g-a-a-t-g-c-c	Single stranded DNA sequence with two large genome segments A and B (> 1kb) and a detailed view of the nucleotide sequence
SNP	A-B-a-t-c-t-t/c-g-a-a-t-g-c-c	Only one base pair (bp) is changed
Microsatellite (STR-short tandem repeats)	A-B-a-t-c-t-t-g-a-t-g-a-(t-g-a)$_n$-a-t-g-c-c	Short sequence (two to five bp) of repetitive DNA
Minisatellite (VNTR-variable number of tandem repeats)	A-B-a-t-c-t-t-g-a-a-t-g-c-c-(a-t-c-t-t-g-a-a-t-g-c-c)$_n$	Large sequence (10-60 bp) of repetitive DNA
Bi-allelic copy number variation (CNV)	A-B-B-a-t-c-t-t-g-a-a-t-g-c-c	Larger segments of DNA is repeated
Multi-allelic CNV	A-B-B-B-(B)$_n$-a-t-c-t-t-g-a-a-t-g-c-c	Larger segments of DNA is repeated multiple times
Insertion	A-B-a-t-c-t-t-a-g-t-c-c-g-g-a-a-t-g-c-c	a DNA segment has been inserted
Deletion	A-B-a-------t-g-a-a-t-g-c-c	a DNA segment has been deleted

Other variations include inversions by which a segment is turned around, and translocation where a segment is either copied or cut from one location and "pasted" in to a new location. Variations ≥1kb is called structure variations.

1.4.1 Single nucleotide polymorphism

Single nucleotide polymorphisms (SNPs) are variants of a base where two alternative alleles have a frequency of ≥1% in the population (51). It was found more than 3 million SNPs in the human genome by the first sequencing of the human genome in 2007. Most of the SNPs in the genome are bi-allelic (with two alternative variants), and tri- and tetra-allelic SNPs (with three and four alternative variants) occur less frequently.

1.4.2 Linkage disequilibrium

High degree of correlation between alleles of for example SNPs in at the same chromosome, is called linkage disequilibrium (LD). It means that alleles in an area (LD-block) tend to be inherited together more often than expected by chance in a population. Testing of one SNP will therefore be enough to survey disease influence for all the SNPs in LD with the first SNP. This is called "tagging". This reduces the amount of work by reducing the number of SNPs

that needs to be tested. However it makes it more difficult to identify the causal variant, and this is one of the greatest challenges in genetics of complex disease as RA. LD is an advantageous tool to find minimum number of SNPs needed for genotyping, and still catch the genetic variation in an area, and is a prerequisite to being able to pick up an association through screening (e.g. genome-wide association studies (GWAS)).

There are multiple ways to calculate LD. One is R square (r^2), which is bidirectional and measures the degree of correlation, for instance $r^2 = 1$ = perfect LD and $r^2 = 0$ = no LD. The alleles at two SNPs need to have exactly the same frequency for the r^2 between them to be 1. Another measure is D', which is unidirectional; D' = 1 if one or more haplotypecombinations never is observed, and D' = 0 if the loci observed are independent of each other (no LD).

1.5 Studying AID genetics

The investigation of genetics in multifactorial AID with susceptibility controlled by multiple genetic and environmental factors is often performed by association studies. The first polymorphisms investigated by association studies in AID were variations in the *HLA*-genes. Association studies are preformed by searching for genetic variation that differs in frequency between a patient- and a control group (52). Later, investigation of candidate genes, chosen on the background of published studies and knowledge of biological functions, identified only a handful of non-HLA genes, like *PTPN22* in RA. Development of new technologies (high-throughput genotyping (microarray)), mapping of the humane genome and collection of large study populations), have increased the knowledge concerning genetic contribution to RA and other AID. Combinations of new insight and establishment of international collaborations have been important in developing GWAS, which has proven efficient in finding genetic risk factors of common diseases.

Based on the "common disease/common variant" hypothesis, GWAS screen, by genotyping 0.5-1 million SNPs, more than 80% of common genetic variations for their contribution to disease susceptibility (52). Because of low prior probability of association among the multiple tests performed, genetic variants with an association showing p-value of $<5x10^{-8}$ have a high likelihood of being true positives. SNPs showing less significant association are replicated in independent case-control panels in order to achieve this convincing genome-wide significant p-value. GWAS is built on the paradigm "LD can be utilized to indirect decide untyped

variation nearby a genotyped SNP". Unfortunately, no genotype array covers all SNPs in the human genome. Only SNPs with minor allele frequency >5% can be caught by LD on the GWAS analysis, which is approximately 2.8 million out of the 3.1 million SNPs found by the HapMap project, where genetic similarities and differences in human being have been identified and catalogued. GWAS has revealed important conclusions regarding number of risk variants, their frequency in the population and the risk of disease. Much data is generated and analyses are therefore computational and resource demanding.

Theoretical models based on currently identified loci, predict that many more risk variants with very small effect sizes are hidden below the genome-wide significance threshold in current GWAS (52). GWAS meta-analysis (collection of multiple case-controls studies) increases the strength to reveal novel genetic risk factors. It is important to remember that ethnic diversity can lead to reduced association, because of variation of SNPs frequencies between ethnical groups, when choosing commercial genotyping array (53).

Immunochip is a custom made chip that include approximately 200.000 SNPs, and can utilize the fact that different AID share genetic background. An Illumina Infinium High-Density array has been designed by the Immunochip consortium to include 186 susceptibility loci collected from 12 different AID. In addition all sample variants from the 1000 Genome Project low-coverage pilot Caucasian population in 0.1 centiMorgan recombinant blocks around each GWAS region were submitted. This array has been applied in large International cohorts to densely genotype immune-mediated disease loci to explore the remarkable genetic overlap identified across a range of AID.

2. AIM OF STUDY

Novel genetic risk factors for rheumatoid arthritis have been identified by genome wide association studies. The aim of this thesis was to investigate the association of novel risk polymorphisms not previously investigated in the Norwegian rheumatoid arthritis population.

The knowledge regarding the effect of different environmental risk factors on rheumatoid arthritis development is limited, and throughout this thesis we wanted to generate more information related to this subject, by sending out a questionnaire. The aim was to get an overview of putative environmental risk factors, which should be considered for further work in the research group on this subject.

3. MATERIAL AND METHODS

The analyses performed in this thesis are based on a case- control study design, in which an affected "case" group is compared with an unaffected "control" group, to identify factors that may contribute to a medical condition. Comparing the RA patients and controls has been carried out to obtain knowledge regarding the contribution of genetics and environmental factors to RA development. For this purpose, precise clinical classification and ethnically matching is important to avoid confounding effect.

3.1 Materials

The controls were recruited through the Norwegian Bone Marrow Registry at the Institute of Immunology, Oslo University Hospital. Patients diagnosed with RA, classified by the criteria developed by the American Rheumatism Association (Table A1, appendix), were recruited through different Norwegian cohorts, at different time points (Table 3).

Table 3. Cohorts the RA patients are included from.

The cohorts	Number of patients recruited from each cohort	Description of the cohorts	The article first describing the cohorts
EURIDISS cohort European Research on Incapacitating Disease and Social Support	216	Established in 1992. All the patients had < 4 years of disease duration when they were included. The patients were followed with clinical inspection, x-ray of hands and blood-tests year 0, 1, 2, 5 and 10.	(54)
ORAR cohort Oslo Rheumatoid Arthritis Register	619	The patients answered questionnaires in 1994, 96, 2001, 2004 and 2009. We have blood samples from the patients that have been to clinical inspection in 1996/97 (totally 636 patients the first round) and most of them had a 2 years follow- up in 1998/99.	(55)
Early RA MRI cohort Early Rheumatoid Arthritis, examined by Magnetic Resonance Imaging	81	Cohort with 84 patients collected in 2002-2004, who had disease duration less than 12 months at the time of inclusion. They were followed up with clinical inspection, x-ray and MR of hands at the baseline, after 3 months, 6 months, 12 months, 3 years and 5 years. Blood samples were taken for analyses of biomarkers and genetic analysis each time.	(56)

TNF cohort Tumour Necrosis Factor-inhibitor therapy	34	These patients were included when they started their first treatment with biological DMARDs (disease modifying anti rheumatic drugs). They have been following the same protocol as the MRI cohort, except that they were only followed up one year.	(57)

The patients and controls gave informed consent for their participation in the study. Ethical permits for the studies were obtained from the Regional Ethical Committee (REK) at the sites were patients were recruited. To exclude false positive results due to population stratification non- ethical Norwegian samples were eliminated, as far as possible. Hence, all patients and controls were of Norwegian origin.

DNA

Whole-genome amplified-DNA (WGA-DNA) was used instead of genomic DNA, because the access to genomic DNA was limited. Increasing the amount of DNA by whole-genome amplification has been validated for the genotyping methods used (58). 50µl WGA was obtained by whole-genome amplification with REPLI-g Midi Kit, Qiagen (Hilden, Germany). 1µl of this was diluted with 19µl 1x Tris-EDTA(TE)buffer. Qiagen guarantees that 3µl of this 1/19 dilution is enough for sequencing and genotyping.

3.2 Investigating genetic risk factors

3.2.1 SNP selection

In total, 35 SNPs was selected for genotyping in the Norwegian RA population (Table 4). These represent SNPs, not previously tested in the Norwegian RA population. SNPs were chosen on the background of different publications;

- 18 risk alleles associated with RA of European ancestry, investigated by GWAS meta-analysis from 2010 (15)
- 12 risk alleles associated with RA of European ancestry, investigated by using Immunochip custom SNP array combined in a meta-analysis with GWAS data, from 2012 (31)
- Three *ERAP* SNPs, based on functional findings in our research group and which has shown independent association with ankylosing spondylitis (chosen based on several reports (59-61)
- One SNP located in the *IL6R* gene associated with asthma (62).
- One SNP associated with RA in a Japanese GWAS (63)

Table 4. All SNPs investigated in this thesis.

SNP	Gene	Chromosome	Published p-value	Minor/major allele	Genotyping method
Risk alleles associated with RA in European ancestry, investigated by GWAS meta-analysis					
rs874040	*RBPJ*	4	$1.0*10^{-16}$	C/G	MassARRAY
rs11676922	*AFF3*	2	$1.0*10^{-14}$	T/A	MassARRAY
rs6859219	*ANKRD55/IL6ST*	5	$9.6*10^{-12}$	A/C	MassARRAY
rs3093023	*CCR6*	6	$1.5*10^{-11}$	A/G	MassARRAY
rs706778	*IL2RA*	10	$1.4*10^{-11}$	A/G	MassARRAY
rs10488631	*IRF5*	7	$4.2*10^{-11}$	C/T	MassARRAY
rs951005	*CCL21*	9	$3.9*10^{-10}$	C/T	MassARRAY
rs934734	*SPRED2*	2	$5.3*10^{-10}$	C/T	MassARRAY
rs13315591	*PXK*	3	$4.6*10^{-8}$	C/T	MassARRAY
rs26232	*C5orf30*	5	$4.1*10^{-8}$	T/C	MassARRAY
rs5029937	*TNFAIP3*	6	$7.5*10^{-8}$	T/G	MassARRAY
rs2736340	*BLK*	8	$1.5*10^{-5}$	T/C	MassARRAY
rs6822844	*IL2_IL21*	4	0.0007	T/G	MassARRAY
rs3218253	*IL2RB*	22	0.002	T/C	MassARRAY
rs7155603	*BATF*	14	$1.1*10^{-7}$	G/A	TaqMan
rs2872507	*IKZF3*	17	$9.4*10^{-7}$	A/G	TaqMan
rs13119723	*IL2, IL21*	4	$6.8*10^{-7}$	G/A	TaqMan
rs7543174	*IL6R*	1	$1.2*10^{-5}$	C/T	TaqMan
Risk alleles associated with RA of European ancestry, investigated by using Immunochip custom SNP array combined with GWAS meta-analysis data					
rs34536443	*TYK2*	19	$2.3*10^{-14}$	C/G	MassARRAY
rs13397	*IRAK1*	X	$1.2*10^{-12}$	A/G	MassARRAY
rs12764378	*ARID5B*	10	$4.5*10^{-10}$	A/G	MassARRAY
rs8026898	*TLE3*	15	$1.4*10^{-10}$	A/G	MassARRAY
rs8043085	*RASGRP1*	15	$1.4\text{-}10^{-10}$	T/G	MassARRAY
rs9979383	*RUNX1*	21	$5.0*10^{-10}$	C/T	MassARRAY
rs12936409	*IKZF3*	17	$2.8*10^{-9}$	T/C	MassARRAY
rs13330176	*IRF8*	16	$4.0*10^{-8}$	A/T	MassARRAY
rs883220	*POU3F1*	1	$2.1*10^{-8}$	T/G	MassARRAY
rs2275806	*GATA3*	10	$4.6*10^{-8}$	G/A	MassARRAY
rs2834512	*RCAN1*	21	$2.1*10^{-8}$	A/G	MassARRAY
rs595158	*CD5*	11	$3.4*10^{-8}$	G/T	MassARRAY
Risk alleles associated with other AID					

rs4129267	*IL6R*	1	10^{-8}	T/C	TaqMan
rs10050860	*ERAP1*	5	NA	T/C	MassARRAY
rs30187	*ERAP1*	5	NA	T/C	MassARRAY
rs2248374	*ERAP2*	5	NA	A/G	MassARRAY
Risk alleles associated with RA in Japanese population investigated by GWAS meta-analysis					
rs10821944	*ARID5*	10	NA	G/T	MassARRAY

NA- not available; not previously reported with RA

Genotyping was carried out at CiGene (Ås, Norway) on a Sequenom MassARRAY by the use of iPLEX GOLD assays two separate times, for 32 SNPs in total. The remaining five SNPs, was genotyped with TaqMan allele discrimination assay at Ullevål. The genotyping procedures are described below. Genotyping was carried out successfully for 32 SNPs, and case/control association analyses was carried out to calculate differences between minor allele frequency (MAF) between cases and controls.

There are several alternative technologies for genotyping, and the choice is based on the number of SNPs, sample size, access to the method, costs and users preference. There is no technology or platform that satisfies all users or study design (64).

3.2.2 TaqMan genotyping

TaqMan allele discrimination is a polymerase chain reaction (PCR) based method for SNP genotyping. Allele- specific fluorescently labeled probes anneals specific to the complementary strand. The probes do not fluorescence, because they are attracted to a quencher that absorbs fluorescence from the reporter (65). Allelic discrimination use probes specific for each allele (66), distinguished by labelling with two different fluorescent reporter dyes, in our concern VIC and FAM. Taq polymerase extends the primers attached to the template, and degenerates probes that are hybridized to the target, by the polymerases 5`nuclease activity (Figure 4). This separates the reporter from the quencher, and causes fluorescence that can be measured at the end of the PCR (65). The fluorescence signals generated by the PCR amplification indicate which allele is present in the sample.

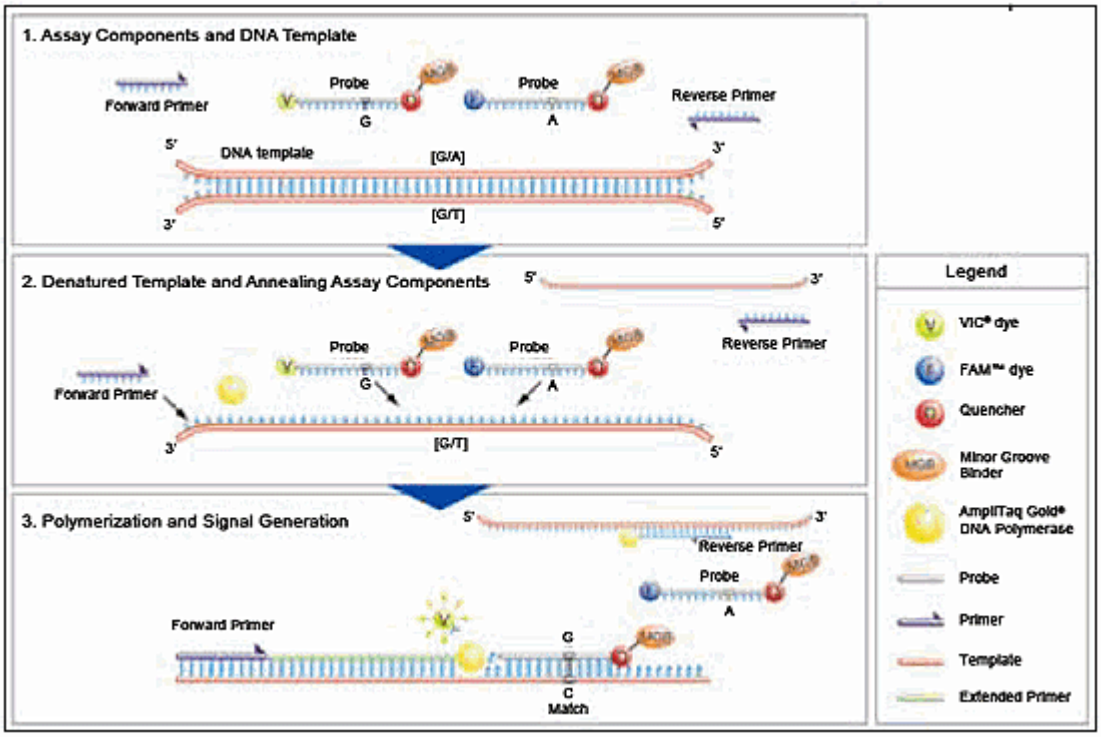


Figure 4. 5´ nuclease activity during TaqMan allele discrimination (67).

Fluorescence measurements are being made after PCR by SDS software v.2.4, which automatically processes the fluorescence data to make genotype calls. A plot is generated, and three clusters of samples represent different genotypes; two homozygous- and one heterozygous clusters (Figure 5).

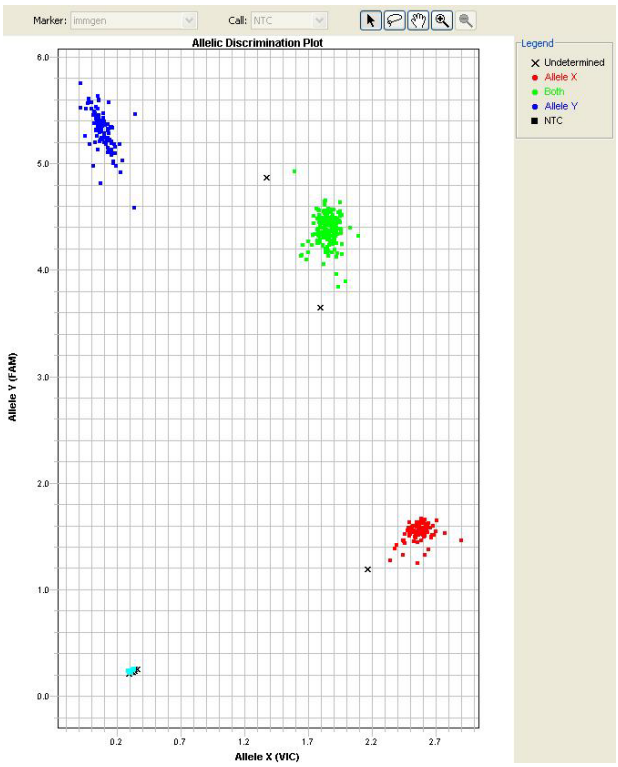


Figure 5. TaqMan plot. Y axis= FAM signal (blue) X axis= VIC signal (red). Blue and red clusters demonstrate homozygous samples and green cluster demonstrates heterozygous samples.

Materials

- 10ng WGA in 1xTE buffer
- 2xTaqMan Universal® PCR Master Mix, LifeTechnologies (Foster City, CA, USA)
- ABsolute QPCR ROX mix, Thermo Scientific (Epsom, Surrey, UK)
- 40x SNP genotyping assay, LifeTechnologies
- 384-well optical plates and plate sealers, LifeTechnologies
- Biomex FX, Beckman Coulter Genomics (Brea, CA, USA)
- Heraeus Megafuge 16R Centrifuge (4000rpm), Thermo Scientific
- Thermal cycler 9700 384, LifeTechnologies
- ABI PRISM 7900 HT sequence detection system for 384-well format, LifeTechnologies (SDS software version 2.4)

TaqMan genotyping was performed with a few alterations in regard to the protocol;
The recommended template for TaqMan SNP Genotyping Assays is purified genomic DNA or complentaryDNA (cDNA). We used WGA, verified for genotyping and sequencing, and do not observe any errors caused by the use of WGA in stead of genomic DNA (or cDNA). TaqMan assay are designed and optimized to work with TaqMan® Universal mix, but as you can see to the left in Figure A3 (appendix), the genotyping results did not cluster very well. Genotyping with ABsolute QPCR ROX mix (Figure A3, to the right, appendix) carryed out much more narrow clusters. Individuals in these clusters are more easily genotyped correctly, and we therefore chose to further use the Absolute-mix instead of TaqMan® Universal-mix.

1. 2µl WGA (5ng/µl) from 96-well plates was delivered to the bottom surface of 384-wells plate using an automated liquid handler, Beckman Coulter Biomex FX and dried down completely by evaporation at room temperature in a dark location.
   The 384-well plates contain 376 sample DNAs, 6 positive control wells (MOU) and 6 negative control wells (dH_2O). Six 384-plates makes a set of WGA from all the patients (n=950) and controls (n=1121)
2. 5µl 80% assay-mix was dispensed per well, using the volumes indicated in Table 5.

Table 5. 80% TaqMan allele discrimination-mix.

	400 samples
TaqMan mix (2x) (lot1105058) or ABsolute QPCR ROX mix (2x) AB1138 (lot090318 and lot00083730)	1000µl
40x TaqMan SNP genotyping assay (differ for each SNP, see Table 6)	40µl
dH_2O	960µl
Total volume	2000µl

Table 6. 40x TaqMan SNP genotyping assay. Two lot numbers are listed for each SNP, because assay was ordered two times.

SNP ID	**Lot**	**Assay ID**	**VIC**	**FAM**
rs7543174	P121022-033 E06 P121113-000E06	C_29898806_10	C (25.776%)	T (74.224%)
rs4129267	P121022-003 E07 P121113-000E07	C_26292282_10	C (73.238%)	T (26.762%)
rs13119723	P121022-003 E08 P121113-000E08	C_26404981_10	A (93.017%)	G (6.983%)
rs7155603	P121022-033 E09 P121113-000 E09	C_2676689_10	A (72.626%)	G (27.374%)
rs2872507	P121022-033 E10 P121113-000 E10	C_11630970_20	A (31.939%)	G (68.061%)

3. The plates were covered using the plate sealer, and centrifuged in a plate centrifuge (4000 rpm).

4. Samples were amplified by PCR, using conditions listed in Table 7.

Table 7. PCR conditions for TaqMan allele discrimination.

Temperature	Time	Description
95ºC	10min	**Denaturation:** Separation of the two DNA strands
95º C	15sec	**Denaturation**
60ºC	1min } x 40	**Annealing/extension:** Primers and probes anneals specific to complementary sequences Extension of the primers by Taq-polymerase and cleaves probes thereby separating reporter dye from quencher
4ºC	∞	

5. Endpoint detection of fluorescence by the use of ABI PRISM 7900 HT sequence detection system for 384-well format
6. Reading fluorescence by "Allelic Discrimination" format, using Allele-calling software supplied with ABI PRISM 7900 HT sequence detection, SDS software v2.4
7. Genotypes were exported as tab-delimited (.txt) for further processing in genetic analysis software.

Reanalysing failed individuals

A few individuals (~5-15) failed genotyped per 384-plate, in the initial round, and therefore needed to be typed again. 4µl WGA (in stead of 2µl) was dried down for individuals showing low signals. Controls for minor allele were included when reanalysing, to make sure each genotype was represented in the plot. The genotyping procedure was otherwise the same as mentioned above.

3.2.3 Genotyping using MassARRAY

CiGene is a core facility in Ås (Norway), who has served several external users in its capacity as a SNP genotyping service for Norway's academic community (68). The two techniques offered at CiGene are Sequenom MassARRAY and Affymetrix platform. MassARRAY is best suitable for genotype relatively few SNPs (<50), in opposite to Affymetrix which suites

to study many SNPs (500000 in a single human sample). In this study, genotyping at CiGene was carried out twice, first for 18 SNPs and thereafter for 12 SNPs, and based on the number of SNPs for genotyping, we choose MassARRAY.

1) UCSC Genome Browser (69) was used to find the SNP sequences and mask repeated area. This data was send to CiGene for primer design.
   UCSC Genome Browser show stricter mask-criteria than RepeatMasker Web Server, and masking was therefore carried out by the use of UCSC Genome Browser.

2) A set of 384 wells plates (see italic text "TagMan genotyping step 1") with 10µl WGA (10ng/µl) was delivered to CiGene on ice.

3) Sequenom MassARRAY was carried out at CiGene according to the general steps described (64):

The Sequenom MassARRAY software designs automatic PCR and extension primers for each SNP, and avoid primer-combinations and nontemplate extension products that can lead to nonspecific extension. iPLEX assays adjust the concentration of extension primers, so that the intensity is as equal as possible. Sequenom real SNP software scans PCR primers to validate that only unique amplification product which consist of target for extended probeprimer is produced.

Specific, individual loci of DNA fragments are evenly amplified with minimal nonspecific by- products, for genotyping on the Sequenom platform. The amplicon is rinsed with shrimp alkaline phosphatase (SAP), which removes remaining non-incorporated dNTPs from amplification product, by cleaving a phosphate group from the 5`termin of the dNTPs. Rinsed amplicons are used as template for primer extension reaction.

iPLEX GOLD reaction (primer extension) is carried out by extend the primer by one mass-modified nucleotide depending on the allele and the design of the assay. To optimise spectrometry analysis it is important to remove salts such as Na^+, K^+ and Mg^+ ions, or else this can result in high background noise in the masspectra. Primer extension products are spotted on SpectroCHIPs, to incorporate oligonucelotides with the appropriate matrix for MALDI-TOF.

Each spot on the chip is shot with laser under vacuum, by matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) method in the mass spectrometer. The matrix

absorbs parts of the light energy and parts of the illuminated substrate vaporize (from fluid to gas). The illuminated and ionized substrate transfers electrostatically into a time of flight mass spectrometer (TOF-MS), where they separates from matrix ions, which is detected individually based on mass-to-charge (m/z) ratio and get analysed.

The detection at the end of the tube is based on flight time which is proportional to $\sqrt{(m/z)}$. Traces are available for viewing imminently after detection, using a suite of software tools, SpectroTYPER-RT (Sequenom).

3.3 Sequencing to resolve unexpected genotype result

Sequencing was carried out because we observed an extra cluster in about 100 of the 2071 genotyped individuals, when analysing the TaqMan results for the rs13119723 SNP (Figure A2 appendix). Looking into the area surrounding the SNP (UCSC Genome Browser), we observed another SNP (rs114092637) three bp upstream from rs13119723. We wanted to find out if appearance of the rs114092637 SNP, caused the extra cluster observed.

Sanger sequencing give us information about the order of the nucleotides, by random incorporation of chain terminating dideoxy-nucleotides (ddNTPs). After isolation of DNA, region of interest is amplified by PCR. The sequencing reaction is carried out by cycle sequencing, where ddNTPs added in the sequencing mix will terminate the amplification, resulting in DNA fragments of different lengths. Because four different ddNTPs labelled with different fluorescent dyes are utilised (dye terminators), the sequencing reaction can be performed in one tube. After rinsing the sequencing products to remove fluorescent ddNTPs not incorporated in the fragments, the products are injected electrokinetically into capillaries filled with polymer. DNA is negatively charged, and will therefore travel towards the positively charged anode when electricity is turned on. High voltage is applied, so that DNA fragments are separated by size, given that small fragments travel more rapidly through the polymer. The fragments are detected by a laser/ camera system, and a sample file of the raw data is created after electrophoration, by Data Collection software.

Materials

- Amplification- and sequencing primers Eurogentec, (Liège, BE)
- WGA (selected individuals and a control sample called MOU)
- BigDye Terminator 1.1 Ready reaction mix (polymerase) LifeTechnologies
- BigDye Terminator 1.1 5x Sequencing Buffer LifeTechnologies
- Gel-electrophoresis: 1% TBE gel with 1kb ladder, Thermo Scientific
- SAP, Thermo Scientific
- Exo, New England BioLabs (Ipswich, MA, UK)
- Data program: Primer3 (version 4.0.0) (70) and UCSC In-Silico PCR (71)
- Veriti Thermal cycler, LifeTechnologies
- Biomex FX, with Agencourt Cleanseq, Beckman Coulter Genomics
- 3730XL DNA Analyzer, LifeTechnologies
- SeqScape Software, LifeTechnologies

1) Primers were design by the program Primer3, and ordered from Eurogentec (Table 8).

   Our criteria for the region of amplification (demonstrated by [ ] in Prime3) was that the sequence carried both SNPs of interest. We designed three primers, due to the SNPs of interest placed in a repetitive area. Placing the reverse- amplification primer in a non-repetitive area increased the specificity, to avoid any unspecific amplification products. The forward primer was used both as amplification- and sequencing primer, but moving the reverse primer outside the repetitive area makes the distance to the SNPs too long for sequencing (the PCR-product is 1308 bp). Therefore we designed a reverse sequencing primer that made it possible to sequence the area we are interested in, both ways. Checking the primer-pair using UCSC In-silico PCR, we found that they amplify a specific PCR product.

Table 8. Sequencing primers, used to sequence the region around the ra13119723

Primer	Primer sequence
Forward, amp and seq	GGCACCAGCAAAGATTTCAT
Reverse, amp	CAAGAGCATGGTGCAGGTTA
Reverse, seq	TAGCCATCCTGACTGGTGTG

2) Optimisation was carried out by PCR with temperature-gradient (54ºC, 56ºC, 58ºC, 60ºC, 62ºC and 64ºC) (Table A2, appendix), and mixes with different $MgCl_2$ concentrations (Table A3, appendix), using MOU as template. The PCR products were analysed by gel-electrophoresis (1% TBE gel with 1kb ladder).

Unfortunately, due to technical difficulties with the camera, we were unable to get a visual picture of the gel. However, the manual inspection of the gel visualised by UV light showed that 1.5mM $MgCl_2$ (Table 9) and 58^{o}C annealing temperature was optimal for amplification.

Table 9. Sequencing reaction mix used after optimalisation of $MgCl_2$ concentration

No. of Wells:	**8**
Sterile Water	71.3µl
10X PCR Buffer (no $MgCl_2$)	9.2 µl
MgCl2 25mM	5.5 µl
dNTP 20 mM	2.5 µl
Fwd Primer*	0.8 µl
Rev Primer*	0.8 µl
Taq Polymerase 5U/uL	0.6 µl
Total Volume	90.9 µl

* 20 pmol/ul. 20µl reaction mix adds to 2µl WGA

Optimal $MgCl_2$ concentration and annealing temperature was demonstrated by analysing one negative control and eight individuals by gel-electrophoresis (Figure 6).

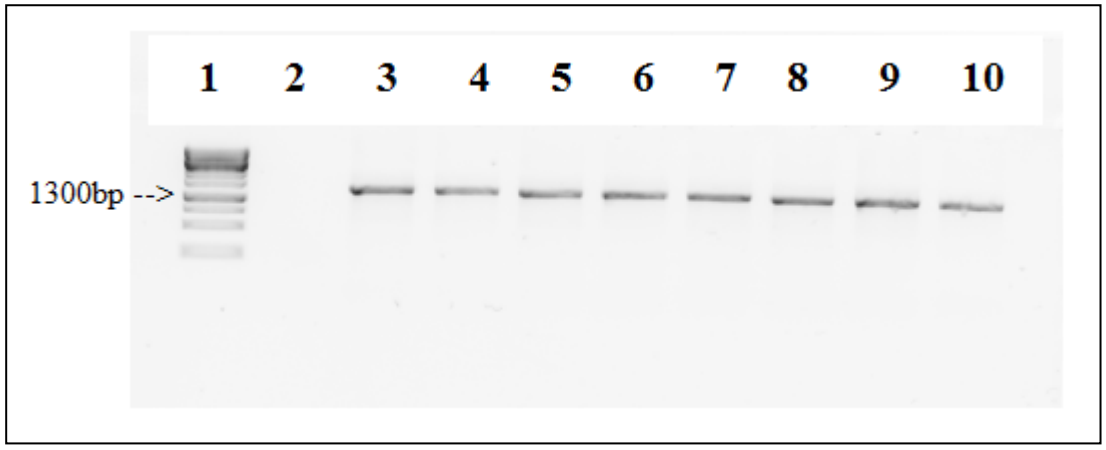


Figure 6. Gel-electrophoresis demonstrating optimal $MgCl_2$ concentration and annealing temperature.
First well= 1kb ladder, second well= negative control, well 3-10= sequencing products for eight individuals.

3) Individuals were selected based on TaqMan plots from the SNP genotyping (Table 10). They were selected to represent individuals from all clusters.

Table 10. Individuals selected for sequencing.

Selected individuals for the 96 wells plate
1 Blank (negative control)
1 MOU (positive control)
34 Patients "extra cluster"
34 Controls "extra cluster"
6 Patents and controls, homozygous GG
10 Patients and controls homozygous AA
10 Patients and controls heterozygous GA

4) Amplification of samples by PCR, using 20µl mix (Table 9) and 2µl WGA. The PCR conditions are listed in Table 11.

Table 11. PCR conditions for amplification of DNA sequence for sequencing.

Temperature	Minutes	
95°C	5min	
95°C	30sec	
58°C	30sec	} x30
72°C	1min30sec	
72°C	7min	
4°C	∞	

5) 12µl ExoSAP-mix (Table 12) was added to 8µl PCR product. The sample was heated to 37°C for 30 minutes for activation of the enzymes and thereafter inactivated at 80°C in 20 minutes.

ExoSAP-mix was used to cleanup the PCR product, to ensure clean and readable DNA sequence. Exonuclease1(EXO1) degrades excess primers and SAP degrades remaining dNTPs from the PCR mixture.

Tabell 12. ExoSAP-mix for PCR cleanup.

	Pr. sample
Exo1(20U/µl) BioLabs lot.0151005	0.5µl
SAP(1U/µl) Thermo Scientific lot.00054687	1.7µl
dH_2O	9.8µl
Total volume	12µl

6) 8µl BigDye sequencing mix (Table 13) adds to 2µl PCR product. One mix with forward primer and one mix with reverse primer (Table 8) were made for sequencing opposite directions in different wells for each individual. Cycle sequencing was carried out at the conditions listed in Table 14, by the use of Veriti Thermal Cycler.

Table 13. BigDye sequencing mix.

	100 samples
BigDye Terminator 1.1 Ready reaction mix (polymeras) lot.1004047M	50µl
BigDye Terminator 1.1 5x Sequencing Buffer lot.1103132	175µl
Sequencing primers 20µM (fwd or rev)	25µl
dH_2O	550µl
Total	800µl

Table 14. Cycle sequencing program on the PCR-machine.

Temperature	**Minutes**	
96°C	1min	
96°C	10sec	
50°C	5sec	} x25
60°C	4min	
4°C	∞	

7) Biomex FX, with Agencourt Cleanseq with magnetic beads was used to clean the sequencing product.

The beads are coated with polynucleotides that bind and separate the DNA from contaminants by magnetic field. The beads are washed with 85% ethanol before they elutes from the magnetic beads and transfers to new 96-wells plate with 85% ethanol and 0,05mM EDTA.

8) 3730XL DNA Analyzer with 96 capillaries filled with pop7 (polymer) was used for analyzing the sequencing products. The capillaries are located in wells with sequencing-products for 15 seconds with electricity turned on, so that the negatively charged DNA is moving towards the anode.

9) Sequences were visualized by SeqScape, and compared with a reference sequence (found by UCSC Genome Browser).

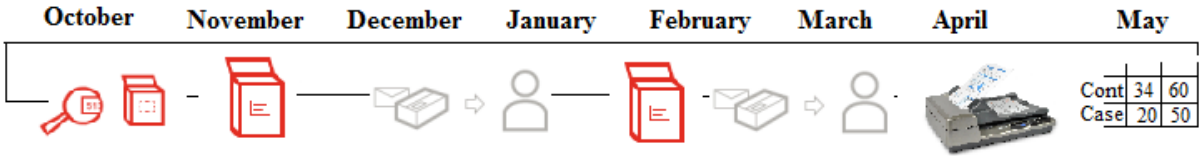
3.4 Questionnaire- Investigate environmental risk factors

Investigation of the influence of environmental factors on the risk of developing RA was carried out by sending out a questionnaire patients and controls in Norway (translated version of the questionnaire in the appendix). The questions was collected from other validated studies, among these the HUNT (Nord-Trøndelag Health Study) and the GEMS study (Genes and Environment in MS). A local validation was carried out by a test-retest including 50 hospital employers, to test this quality aspect of the questionnaire. The participants filled in the questionnaire twice with an interval of four weeks, 33 responded both times. In general, the concordance between the results given the first and second time was good. Only three of the questions in the questionnaire had a high degree of variation between the first and the second response, this was the questions concerning sun exposure, and these were changed in order to increase their reliability (question number 28-30). The questionnaire was designed and approved by The Norwegian Ethical Committee, before I joined the research group.

To send out the questionnaire, the following steps were performed (See Figure 7 for a schematic overview of the process I was involved in); First, our research database was linked with the clinical database containing personal ID numbers. Next, we eliminated the patients who were dead, and retrieved the addresses of the living patients by checking the National Register. The questionnaire was sent by mail to the patients in November 2012. As this questionnaire is also utilized for studying other AID in addition to RA and we share the same

control group, the controls had already answered the questionnaire. The participants gave written informed consent. After a few months, a reminder was sent to the patients who did not return the questionnaire. In April, answered questionnaires were scanned to a regular scanner by the use of the TeleForm Scan Station software. We used the time in April and May to get an overview of results from the questionnaire. Based on previous knowledge about environmental risk factors, we chose to analyse the questions regarding smoking (question 14), periodontitis (q10), alcohol (q20 and 21), coffee (q22 and 23), presence of pets and domestic animals (q12), mononucleosis (q7) and breastfeeding (q41).

Figure 7. A schematic overview of the process of sending out, receive, scanning and analysing the questionnaires from the RA patients.



In October 2012, the patients addresses were obtained from the National Register, and the questionnaire was sent out by mail in November 2012. We received answers during the winter, and the questionnaire was resent to individuals who did not respond the first time, in February 2013. We finished scanning the questionnaires 9th of April 2013, and analyses was carried out in April and May.

The limited amount of time, restrict my engagement in this project. We therefore wanted to get a brief overview of the received questionnaires in regard to response-rate, the population homogeneity and some of the questions that we were most interested in. This will be used as a background for further and deeper investigation.

Prior to this project, information regarding the patients smoking habits had been collected (mostly in 1994), 285 of these patients have also answered the questionnaire sent out in 2012/2013. We wanted to use this data to find out to which degree the answers correlate between the two times they answer the question.

3.5 Statistical analysis

For statistical analysis, I mainly used two tools; HaploView (72) and PLINK (73) (74). HaploView is a bioinformatic-software, which can be utilized to visualize LD patterns, perform association studies, choosing tagSNPs and estimating haplotype frequencies. PLINK is a genetic analysis toolset designed to handle large data sets, focusing on analysis of genotype- and phenotype data.

Genotype success rate (GSR) gives us the percentage of individuals that have been genotyped successfully per SNP. High GRS (>95%) tell us that genotyping of less than 5% of the individuals failed. To avoid skewing due to on genotype being difficult to score, a high GSR is necessary. GSR was calculated to check the genotype quality, by using HaploView v4.2, Number of successfully genotyped samples was divided by the number of samples undergoing genotyping, and indicates the percentage of successfully genotyped samples.

Hardy Weinberg equilibrium (HWE) states that allele- and genotype- frequencies in a population will remain constant over time if there are no disturbing influences. Aberrancy from HWE can indicate problems with genotyping. Hardy-Weinberg equilibrium was calculated by using PLINK v1.07, with $p<0.05$ as significant threshold. PLINK v.1.07 was also used to remove individuals that failed genotyping for more than 75% of the SNPs, before calculations of GSR and HWE.

Case/control association analysis for each SNP was perform by using PLINK v1.07. The frequency of the minor allele in patients was compared with the frequency of minor allele in controls, and to see if there was significant differences between the minor allele frequencies between the two groups, with significant threshold $p<0.05$. The association analysis was performed by analysing all the patients versus controls, and also by stratification regarding ACPA- status and RF- status.

Analyses of environmental risk factors were preformed by comparing patients ever and never exposed to an environmental riskfactor with controls ever and never exposed to the same risk factor. A two by two table were used to calculate χ^2, odds ratios with 95% confidence interval and p-values with significant threshold $p\leq0.05$.

Bonferroni corrections:

To counteract the problem of multiple testing, we used Bonferroni correction. Multiple testing is a problem, because the likelihood of witnessing a rare event, and therefore reject the null hypotheses when it is true (type I error) increases, as the number of hypotheses in a test increases. Bonferroni correction is performed by multiplying the p-values obtained after an association test with the total number of tests performed, and will make sure that the total likelihood of type I error remains 0.05.

k (number of tests for analysing the questionnaire) = (7 questions*3tests)+(2questions*1test) = 23
α (the probability that we will falsely reject the null hypothesis for one single test) = 0.05
Bonferroni corrected p-value (α/k) = 0.05/23 = 0.002

A p-value ≤ 0.002 ensures that the total likelihood of type I error still is 0.05, when analysing the environmental factors, with 23 as the total number of tests performed.

Power calculations:

The software PS- Power and Sample Size Calculation v.3.0.43(75) was used for power calculations. It is a computer program which can determine the sample size needed to detect a specific alternative hypothesis, the power with which a specific alternative hypothesis can be detected with a given sample size, or the alternative hypothesis that can be detected with a given power and sample size. We utilized the latter option to determine the alternative hypothesis in terms of odds ratio, for dichotomous, case-control study and uncorrected χ^2 test.

Listed below are the terms entered into the program:

α (the probability that we will falsely reject the null hypothesis) = 0.05
power (the probability of correctly rejecting the null hypothesis) = 0.8
p_0 (probability of exposure in controls) = 0.1-0.5
n (the number of case patients) = 287
m (the ratio of control to experimental subjects) = (922/287) 3.213

3.6 Bioinformatic tools used

- UCSC Genome Browser (http://genome.ucsc.edu/)
- dbSNP (http://www.ncbi.nlm.nih.gov/snp/)
- GWAS catalogue (http://www.genome.gov/gwastudies/)
- Prime3 (http://primer3.wi.mit.edu//)
- UCSC In-Silico PCR (http://genome.ucsc.edu/cgi-bin/hgPcr?hgsid=336261293)
- Gene (http://www.ncbi.nlm.nih.gov/gene/)

4. RESULTS

Genotyping was carried out to investigate 35 novel RA risk loci in the Norwegian population. As listed in Table 15, 950 patients and 1121 controls were included in the study, of these ~79% of the patients and 55% of the controls were women. Genotyping was carried out on two separate occasions at CiGene (different SNPs at different times), and therefore the analysis was carried out at two separate time points (first time did also include the genotypes generated by TaqMan allele discrimination). The number of individuals passing >75% quality control (each individual had to be genotyped for ≥75% of the SNPs) varied between the genotyping rounds, this is why the number of patients and controls included for genotype-analyses are listed twice in Table 15.

Table 15. Demographic characteristics of patients and controls included in the thesis.

	Cases	Controls
Included in the study, n	950	1121
Females, n (%)	735 (79.3)	617 (55.0)
ACPA positivity[a], n (%)	548 (61.6)	
RF-positivity[b], n (%)	478 (53.6)	
First round of genotyping n*	937	1113
Second round of genotyping n*	889	1092

* Individuals passed quality control (75%). [a]ACPA data missing n = 60. [b]RF data missing n = 59.

Sequenom MassARRAY was, as mentioned above carried out at CiGene twice. The last time, technical failure occurred, and a few individuals failed genotyping per plate. This is why the number of genotyped individuals was a bit low in the second round (Table 15).

4.1 Control of genotyping quality

Three of the SNPs genotyped were removed before additional analyses for different reasons; rs934737 and rs5029937 failed genotyping and rs706778 showed several individuals clustering between the AG and AA clusters (Figure A1, appendix), and therefore could not be confidently called. To further ensure good quality of the genotyping, GSR and HWE were calculated (Table 16).

Table 16. Calculated genotyping success rate and Hardy-Weinberg equilibrium, for the SNPs successfully genotyped.

	GSR	n failed		HWE	
SNP	**%**	**Cases (tot 950)**	**Controls (tot 1121)**	**p cases**	**p controls**
rs4129267	99.5	9	1	0.4055	0.1002
rs7543174	99.1	11	7	1	0.6328
rs11676922	97.6	19	31	0.6913	0.395
rs13315591	99.4	2	11	0.3826	1
rs874040	99.7	4	2	0.937	0.2702
rs6822844	99.8	5	0	0.8156	0.6951
rs6859219	99.7	5	1	0.4224	0.9252
rs10050860	99.8	4	0	0.8328	0.5764
rs30187	99.9	2	0	1	0.5979
rs2248374	99.9	3	0	0.4714	0.1342
rs26232	99.7	3	3	1	0.3259
rs3093023	99.7	3	3	0.9479	1
rs10488631	99.9	2	1	0.07054	0.9001
rs2736340	99.9	3	0	0.7358	0.2754
rs951005	99.9	3	0	0.5036	0.7103
rs10821944	98.2	6	31	0.4716	0.9389
rs7155603	99.5	8	2	1	0.9273
rs2872507	99.6	8	0	0.7428	0.6747
rs3218253	99.6	8	0	0.3858	0.8246
rs13119723	**93.9**	64	62	0.1121	0.2344

rs883220	99.7	4	2	0.8557	0.8038
rs12764378	99.0	4	15	0.6305	1
rs2275806	99.7	6	0	0.4962	0.2181
rs595158	99.9	2	0	0.5914	**0.02503**
rs8026898	99.8	4	0	0.382	0.8137
rs8043085	99.8	3	0	0.7655	0.9294
rs13330176	99.2	1	15	0.6606	0.872
rs12936409	99.2	15	1	0.946	0.4674
rs34536443	99.9	1	0	1	0.0527
rs2834512	99.6	7	1	0.3964	0.3729
rs9979383	99.9	2	0	0.6139	**0.02593**
rs13397	99.8	3	1	0.32	0.94

Bold numbers: GSR<95% and HWE<0.05

GSR was above 95% for all SNPs, except rs13119723 (GSR= 93.9%), caused by the extra cluster observed in genotyping-plots generated by TaqMan allele discrimination (Figure A2, appendix), by which about 100 of the heterozygous individuals did not genotype successfully. This was further investigated by sequencing (see 4.3 "Sequencing revealed SNP interfering with TaqMan result").

All SNPs were in HWE given $p<0.05$, and hence were included for further association analysis (Table 16). However two SNPs showed a tendency towards deviation from HWE, rs595158 and rs9979383 ($p\approx0.03$) and the genotype distribution was therefore inpected by calculations based on the Hardy-Weinberg proportions ($p^2 + 2pq + q^2 = 1$) (Table 17 and 18).

Table 17. Observed and expected genotype distribution among cases and controls for rs595158, based on HWE proportions.

	GG	GT	TT	p-value
Controls O (E)	272 (291)	583 (545)	237 (255)	0.03
Cases O (E)	223 (227)	452 (443)	212 (216)	0.56

O= observed E= expected

Table 18. Observed and expected genotype distribution among cases and controls for rs9979383, based on HWE proportions.

	TT	CT	CC	p-value
Controls O (E)	423 (405)	484 (519)	184 (166)	0.03
Cases O (E)	349 (353)	421 (413)	117 (120)	0.57

O= observed E= expected

For rs595158, there were more heterozygote individuals among the controls than expected given HWE (Table 17). For rs9979383, there were less heterozygous individuals among the controls than expected given HWE (Table 18). These differences were however neither very big nor skewed and not caused by genotyping problems of a particular genotype, and the SNPs were therefore kept for further analysis.

4.2 Genotyping results of association analysis

Association analyses of the genotyping data by comparing all patients versus controls (Table 19), showed five significant SNPs ($p<0.05$).

The A allele at rs3093023 (*CCR6*) was found at a higher frequency in patients compared to controls (OR = 1.28, p = 0.002), the C allele at rs34536443 (*TYK2*) was found at a reduced frequency (OR = 0.682, p = 0.026), the A allele at rs6859219 (*ANKRD55/IL6ST*) was found at a reduced frequency (OR = 0.834, p = 0.025), the A allele at rs8026898 (*TLE3*) was found at a increased frequency (OR = 1.239, p = 0.003) and the G allele at rs874040 (*RBPJ*) was found at increased frequency (OR = 1.202, p = 0.009).

Table 19. Results from association analysis of the whole RA material.

SNP	Minor allele	MAF case	MAF control	OR (95 %CI)	p-value
rs10050860	T	0.191	0.201	0.939 (0.804-1.096)	0.426
rs10488631	C	0.149	0.139	1.087 (0.912-1.295)	0.351
rs10821944	G	0.287	0.273	1.073 (0.935-1.231)	0.319
rs11676922	T	0.471	0.497	0.901 (0.795-1.02)	0.099
rs12764378	A	0.225	0.218	1.046 (0.8988-1.216)	0.564
rs12936409	T	0.481	0.495	0.948 (0.836-1.075)	0.404
rs13119723	G	0.15	0.172	0.848 (0.713-1.007)	0.06
rs13315591	C	0.061	0.067	0.908 (0.706-1.17)	0.456
rs13330176	A	0.257	0.253	1.02 (0.883-1.178)	0.789
rs13397	A	0.12	0.12	1.000 (0.81-1.234)	0.997
rs2248374	G	0.491	0.5	0.963 (0.851-1.09)	0.543
rs2275806	G	0.449	0.434	1.062 (0.9361-1.205)	0.35
rs26232	T	0.297	0.31	0.941 (0.823-1.08)	0.372
rs2736340	T	0.263	0.291	0.873 (0.760-1.001)	0.052
rs2834512	A	0.112	0.115	0.977 (0.802-1.191)	0.819
rs2872507	A	0.484	0.494	0.9613 (0.85-1.088)	0.53
rs30187	T	0.356	0.349	1.031 (0.907-1.173)	0.64
rs3093023	**A**	**0.488**	**0.439**	**1.218 (1.077-1.378)**	**0.002**
rs3218253	T	0.296	0.283	1.062 (0.928-1.217)	0.382
rs34536443	**C**	**0.03**	**0.044**	**0.682 (0.486-0.958)**	**0.026**
rs4129267	T	0.385	0.387	0.992 (0.874-1.126)	0.901
rs595158	G	0.494	0.484	1.04 (0.9176-1.179)	0.539
rs6822844	T	0.168	0.188	0.870 (0.741-1.023)	0.091
rs6859219	**A**	**0.172**	**0.2**	**0.834 (0.711-0.978)**	**0.025**
rs7155603	G	0.227	0.207	1.126 (0.97-1.307)	0.12
rs7543174	C	0.196	0.196	1.002 (0.857-1.17)	0.984
rs8026898	**A**	**0.303**	**0.26**	**1.239 (1.078-1.424)**	**0.003**
rs8043085	T	0.215	0.218	0.983 (0.844-1.144)	0.824
rs874040	**G**	**0.293**	**0.256**	**1.202 (1.047-1.379)**	**0.009**
rs883220	T	0.245	0.24	1.026 (0.8864-1.188)	0.73
rs951005	C	0.142	0.141	1.011 (0.848-1.206)	0.903
rs9979383	C	0.369	0.39	0.915 (0.804-1.041)	0.178

Dividing cases in groups regarding ACPA status indicated nine significantly associated SNPs. Five of these were also significant associated when analysing all the patients compared to controls (Table 19), described above. rs3093023 (*CCR6*), rs8026898 (*TLE3*) and rs874040 (*RBPJ*) was associated with ACPA positive RA, and rs34536443 (*TYK2*) and rs6859219 (*ANKRD55/IL6ST*) was associated with ACPA negative RA. Among these, rs6859219 (*ANKRD55/IL6ST*) had a much reduced p-value for analyses of ACPA negative versus controls (p=0.0000936) compared to analyses of all patients versus controls (p=0.025).

In addition to the five SNPs already mentioned, stratification of patients regarding ACPA status showed four additional significantly associated SNPs (Table 20). One SNP was significantly associated with ACPA positive disease; the G allele of rs7155603 (*BATF*) was observed at an increased level for ACPA positive patients compared to controls (OR = 1.198, p = 0.042). Three SNPs was significantly associated with ACPA negative disease, by which minor allele for the three SNPs was observed at an reduced frequency for ACPA negative patients compared to controls; rs8043085 (*RASGRP1*) (OR = 0.791, p = 0.041), rs2736340 (*BLK*) (OR = 0.791, p = 0.02) and rs11676922 (*AFF3*) (OR = 0.819, p = 0.025).

Table 20. Results form case/control association analysis when grouping patients regarding ACPA status.

					ACPA pos vs. controls		**ACPA neg vs. controls**	
SNP	**Minor allele**	MAF ACPA pos	MAF ACPA neg	MAF control	OR (95 %CI)	p-value	OR 95 %CI	p-value
rs10050860	T	0.177	0.204	0.201	0.853 (0.707-1.029)	0.097	1.016 (0.820-1.259)	0.883
rs10488631	C	0.158	0.142	0.139	1.165 (0.951-1.428)	0.14	1.026 (0.801-1.313)	0.84
rs10821944	G	0.274	0.305	0.273	1.006 (0.854-1.186)	0.939	1.169 (0.968-1.413)	0.105
rs11676922	**T**	0.481	**0.447**	**0.497**	0.939 (0.811-1.088)	0.403	**0.819 (0.688-0.976)**	**0.025**
rs12764378	A	0.212	0.23	0.218	0.968 (0.807-1.161)	0.723	1.072 (0.87-1.323)	0.517
rs12936409	T	0.511	0.452	0.495	1.068 (0.92-1.241)	0.388	0.842 (0.705-1.005)	0.056
rs13119723	G	0.157	0.148	0.171	0.905 (0.737-1.11)	0.336	0.844 (0.659-1.08)	0.177
rs13315591	C	0.06	0.067	0.067	0.9 (0.666-1.216)	0.492	1.001 (0.709-1.414)	0.995
rs13330176	A	0.259	0.254	0.253	1.031 (0.869-1.223)	0.726	1.005 (0.821-1.229)	0.965
rs13397	A	0.123	0.117	0.12	1.027 (0.804-1.314)	0.829	0.967 (0.721-1.297)	0.822
rs2248374	G§	0.506	0.481	0.5	1.021 (0.883-1.181)	0.782	0.924 (0.777-1.098)	0.369
rs2275806	G	0.455	0.452	0.434	1.089 (0.938-1.265)	0.263	1.077 (0.902-1.285)	0.412
rs26232	T	0.304	0.289	0.31	0.973 (0.830-1.139)	0.73	0.908 (0.752-1.098)	0.32
rs2736340	**T**	0.28	**0.245**	**0.291**	0.950 (0.809-1.117)	0.536	**0.791 (0.649-0.964)**	**0.02**

rs2834512	A	0.111	0.112	0.115	0.967 (0.794-1.224)	0.781	0.976 (0.739-1.289)	0.865
rs2872507	A	0.511	0.458	0.494	1.073 (0.927-1.241)	0.346	0.868 (0.73-1.032)	0.108
rs30187	T	0.372	0.333	0.349	1.105 (0.949-1.285)	0.198	0.932 (0.777-1.118)	0.45
rs3093023	**A**	**0.494**	0.478	**0.439**	**1.249 (1.079-1.445)**	**0.003**	1.168 (0.983-1.388)	0.078
rs3218253	T	0.311	0.277	0.283	1.141 (0.973-1.338)	0.105	0.97 (0.8-1.175)	0.753
rs34536443	**C**	0.036	**0.026**	**0.044**	0.819 (0.556-1.205)	0.31	**0.588 (0.348-0.992)**	**0.044**
rs4129267	T	0.379	0.391	0.387	0.967 (0.832-1.124)	0.662	1.015 (0.850-1.212)	0.868
rs595158	T	0.494	0.494	0.484	1.041 (0.897-1.208)	0.593	1.04 (0.873-1.24)	0.66
rs6822844	T	0.167	0.173	0.188	0.865 (0.713-1.048)	0.138	0.9 (0.717-1.129)	0.361
rs6859219	**A**	0.2	**0.133**	**0.2**	1.002 (0.835-1.203)	0.981	**0.616 (0.482-0.787)**	**0.0000936**
rs7155603	**G**	**0.238**	0.225	**0.207**	**1.198 (1.007-1.426)**	**0.042**	1.11 (0.902-1.367)	0.325
rs7543174	C	0.193	0.194	0.196	0.981 (0.815-1.181)	0.843	0.991 (0.797-1.232)	0.937
rs8026898	A	**0.305**	0.285	**0.26**	**1.247 (1.058-1.47)**	**0.008**	1.133 (0.932-1.378)	0.211
rs8043085	T	0.239	**0.181**	**0.218**	1.128 (0.946-1.346)	0.179	**0.791 (0.632-0.991)**	**0.041**
rs874040	**G**	**0.307**	0.27	**0.256**	**1.286 (1.095-1.511)**	**0.002**	1.075 (0.884-1.306)	0.469
rs883220	T	0.247	0.237	0.24	1.037 (0.872-1.233)	0.683	0.983 (0.8-1.209)	0.873
rs951005	C	0.135	0.159	0.141	0.950 (0.769-1.175)	0.638	1.149 (0.905-1.459)	0.253
rs9979383	C	0.355	0.386	0.39	0.862 (0.738-1.006)	0.059	0.982 (0.820-1.176)	0.843

§Different minor allele for ACPA positive and ACPA negative, but we chose to define G as minor allele.

Dividing cases in groups regarding RF status indicated seven significantly associated SNPs, five of these (rs3093023, rs8026898, rs874040, rs34536443 and rs6859219) was also significantly association when comparing all patients versus controls (Table 19) and when analysing patient groups stratified regarding ACPA status versus controls (Table 20). Among these, SNPs associated with ACPA positive disease was associated with RF positive disease (rs3093023 (*CCR6*), rs8026898 (*TLE3*) and rs874040 (*RBPJ*)), and SNPs associated with ACPA negative disease was associated with RF negative disease (rs34536443 (*TYK2*) and rs6859219 (*ANKRD55/IL6ST*)). rs3493023 (*CCR6*) had a much reduced p-value comparing RF positive patients versus controls (p = 0.000349), than comparing ACPA positive- or all patients versus controls (p = 0.003 and p = 0.002, respectively).

Additional two SNPs showed significant association when stratifying for RF status; the C allele of rs10488631 (*IRF5*) was observed with an increased frequency in RF positive patients compared to controls (OR = 1.262, p = 0.029), and the G allele of rs10821944 (*ARID5*) was observed with an increased frequency in RF negative patients compared to controls (OR = 1.237, p = 0.018) (Table 21).

Table 21. Results form case/control association analysis when grouping patients regarding RF status.

					RF pos vs. controls		RF neg vs. controls	
SNP	**Minor allele**	MAF* RF pos	MAF* RF neg	MAF* controls	OR (95 %CI)	p-value	OR (95 %CI)	p-value
rs10050860	T	0.183	0.202	0.201	0.889 (0.731-1.08)	0.236	1.004 (0.822-1.227)	0.965
rs10488631	**C**	**0.169**	0.131	**0.139**	**1.262 (1.024-1.554)**	**0.029**	0.935 (0.739-1.184)	0.579
rs10821944	**G**	0.264	**0.317**	**0.273**	0.956 (0.804-1.137)	0.611	**1.237 (1.038-1.475)**	**0.018**
rs11676922	T	0.47	0.467	0.497	0.898 (0.77-1.047)	0.17	0.889 (0.756-1.046)	0.156
rs12764378	A	0.198	0.249	0.218	0.888 (0.731-1.078)	0.231	1.194 (0.985-1.446)	0.07
rs12936409	T	0.488	0.471	0.495	0.972 (0.831-1.137)	0.726	0.912 (0.773-1.075)	0.271
rs13119723	G	0.147	0.156	0.171	0.841 (0.676-1.046)	0.12	0.901 (0.719-1.129)	0.365
rs13315591	C	0.062	0.058	0.067	0.918 (0.671-1.257)	0.594	0.855 (0.61-1.2)	0.365
rs13330176	A	0.254	0.264	0.253	1.005 (0.84-1.202)	0.958	1.059 (0.879-1.276)	0.545
rs13397	A	0.121	0.125	0.12	1.004 (0.775-1.301)	0.976	1.04 (0.796-1.359)	0.775
rs2248374	G§	0.49	0.5	0.5	0.961 (0.825-1.119)	0.606	0.998 (0.850-1.172)	0.983
rs2275806	G	0.46	0.434	0.434	1.111 (0.95-1.299)	0.189	1.003 (0.851-1.184)	0.968
rs26232	T	0.302	0.302	0.31	0.963 (0.816-1.137)	0.657	0.964 (0.81-1.148)	0.683
rs2736340	T	0.263	0.265	0.291	0.872 (0.735-1.035)	0.118	0.879 (0.734-1.053)	0.162
rs2834512	A	0.119	0.11	0.115	1.039 (0.815-1.324)	0.757	0.958 (0.739-1.242)	0.744
rs2872507	A	0.49	0.474	0.494	0.987 (0.848-1.15)	0.865	0.924 (0.786-1.086)	0.336
rs30187	T	0.367	0.337	0.349	1.085 (0.926-1.271)	0.315	0.948 (0.800-1.123)	0.538
rs3093023	**A**	**0.509**	0.468	**0.439**	**1.321 (1.134-1.539)**	**0.000349**	1.123 (0.956-1.32)	0.158
rs3218253	T	0.29	0.308	0.283	1.032 (0.872-1.222)	0.711	1.128 (0.947-1.345)	0.177
rs34536443	**C**	0.038	**0.022**	**0.044**	0.864 (0.58-1.288)	0.472	**0.482 (0.286-0.813)**	**0.005**
rs4129267	T	0.389	0.39	0.387	1.007 (0.861-1.178)	0.927	1.01 (0.856-1.192)	0.903
rs595158	T	0.491	0.504	0.484	1.029 (0.880-1.202)	0.723	1.083 (0.92-1.275)	0.34
rs6822844	T	0.169	0.169	0.188	0.876 (0.716-1.071)	0.197	0.878 (0.710-1.085)	0.227
rs6859219	**A**	0.202	**0.148**	**0.2**	1.012 (0.836-1.224)	0.905	**0.695 (0.558-0.866)**	**0.001**
rs7155603	G	0.229	0.224	0.207	1.135 (0.945-1.364)	0.176	1.106 (0.910-1.343)	0.311
rs7543174	C	0.195	0.194	0.196	0.997 (0.822-1.209)	0.976	0.991 (0.808-1.214)	0.929
rs8026898	**A**	**0.318**	0.277	**0.26**	**1.326 (1.118-1.573)**	**0.001**	1.089 (0.906-1.308)	0.363
rs8043085	T	0.237	0.197	0.218	1.111 (0.924-1.337)	0.263	0.88 (0.718-1.078)	0.218
rs874040	**G**	**0.311**	0.267	**0.256**	**1.313 (1.11-1.553)**	**0.001**	1.06 (0.88-1.267)	0.558
rs883220	T	0.253	0.233	0.24	1.075 (0.898-1.288)	0.431	0.965 (0.796-1.169)	0.715
rs951005	C	0.146	0.146	0.141	1.043 (0.84-1.294)	0.706	1.038 (0.826-1.305)	0.747
rs9979383	C	0.368	0.383	0.39	0.911 (0.776-1.071)	0.259	0.969 (0.819-1.146)	0.713

§Different minor allele for RF positive and RF negative, but we chose to define G as minor allele.

4.3 Sequencing revealed SNP interfering with TaqMan result

Sequencing was carried out for the rs13119723 SNP, because of the extra cluster observed from genotyping with TaqMan allele discrimination (Figure A2, appendix). Our hypothesis was that the extra cluster observed was due to the SNP rs114092637, located two bp upstream for the rs13119723 SNP. Hence, the individuals in the extra cluster would be heterozygous for the rs114092637 SNP, and potentially show a reduction of the FAM signal was caused by weaker binding of the probe to the template when minor allele for the rs114092637 was present. Sequencing results for the two positions of interest confirmed the hypothesis (Table 22).

Figure 8 illustrates the difference in SNP- position rs114092637, by showing the sequencing results for an individual heterozygous for both SNPs to the left and for an individual homozygous for the rs114092637 SNP and heterozygous for the rs13119723 SNP to the right.

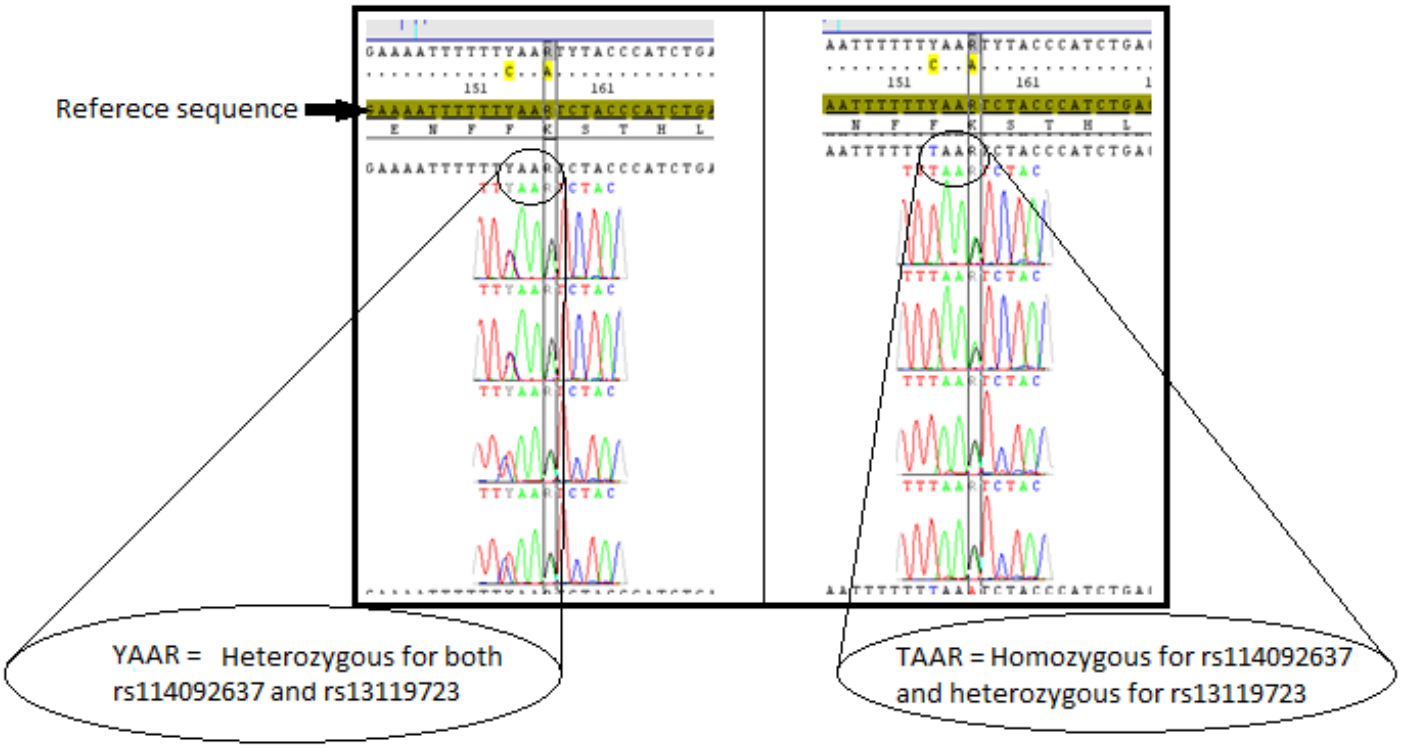


Figure 8. Visualization of the sequencing results. In the sequence "YAAR" and "TAAR" Y (C/T) and T are the alleles for rs114092637 position and R (A/G) are the alleles for rs13119723 position.

All individuals in the extra cluster were heterozygous for the SNP two bp up stream for rs13119723, while none of the individuals in the other genotype clusters covered the C-allele (all were TT homozygous) (Table 22).

Table 22. Sequencing results for the two positions of interest.

		Sequencing result	
TaqMan results for rs13119723	**Number of individuals sequenced**	**rs13119723**	**rs114092637**
Extra cluster	68	R*(68)	Y§ (68)
Homozygous TT	6	GG (6)	TT (6)
Homozygous AA	10	AA (10)	TT (10)
Heterozygous AG	10	R (10)	TT (10)

*R=A/G §Y=C/T

4.4 Environmental risk factors associated with RA

The questionnaire was sent to 562 RA patients. 12 of these were returned unopened by the Postal Service, and therefore 550 patients apparently received the questionnaire. In total 416 patients responded, but 114 of these checked for not wanting to answer the questionnaire. The response rate among the cases was therefore 75.6%, and 54.9% completed the questionnaire (Figure 9). The questionnaire was sent to 1100 controls, ten were returned unopened by the Postal Services, and of the 1090 who received the questionnaire, 933 responded and 11 of these checked for not wanting to answer the questionnaire. The response rate was therefore 85.6% among the controls, and 84.6% completed the questionnaire (Figure 9).

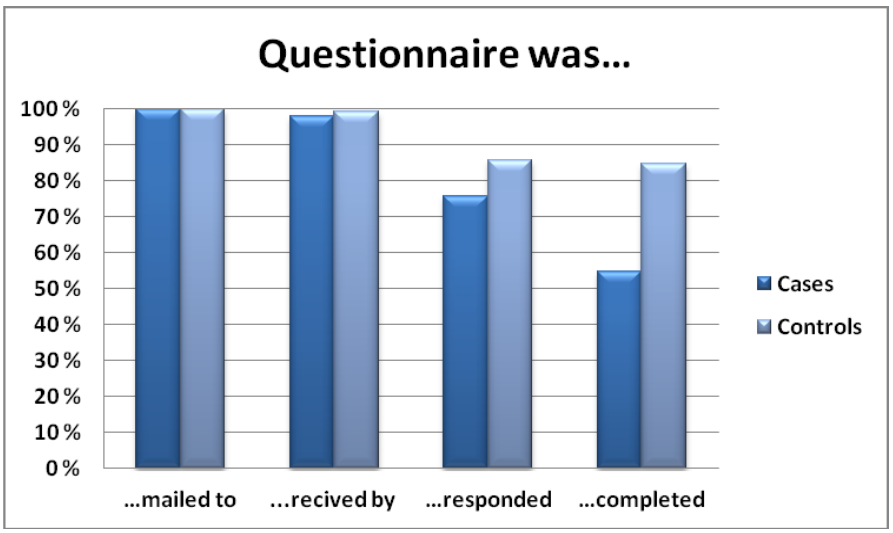


Figure 9. Collection of data regarding environmental factors by sending out questionnaires.

Even though the RA cohort had been gone through by the clinicans to only include Norwegians, the questionnaire (question 4 and 5) reveald that six individuals had non-Norwegian background. Individuals reporting three or more parents and grandparents from other countries in Southeast-Europe (or other continents) were removed, and individuals reporting one ore more parent(s) or grandparent(s) coming from other continents (except Europe and USA) were removed.

We got data regarding autoantibody status and shared epitope for most of the patients who answered the questionnaire, but nine individuals missed data on both antibodies, and therfore 287 cases were included for analysis. The demographic charachteristics is shown in Table 23. The patients mean age was almost 20 years higher than for the controls, and the percentage of women was much higher among the patients than the controls.

The data material initially contained 548 ACPA positive patients (Table 15 on page 39), but questionnaire was sent out to only 283 of these, because several had passed away. In comparison, questionnaire was sent out to 222 of the 342 ACPA negative patients from the initial data material. This is why the percentage of ACPA positive patients was reduced among patients included for analysis of environmental risk factors (55.3%) compared to the initial data material used for genotyping (61.6%).

Table 23. Title Demographic characteristics of patients and controls included for analyses of environmental risk factors.

	Cases (total n= 287)	**Controls (total n=922)**
Mean age, years (lowest and highest)	64.9 (33-92)	45.9 (24-63)
Females n (%)	231 (80.2)	529 (58.3)
Regarding the cases		
Mean age at onset	39	
ACPA positive n (%)[a]	152 (55.3)	
RF positive n (%)[b]	139 (50.9)	

[a] missing ACPA status n=12. [b] missing RF status n=14.

Effects (odds ratios) that could be detected with 80% power in our sample consisting of 287 patients and 922 controls, in regard to the probability of exposure in controls (10-50%) was calculated. Figure 10 show that we were able to detect predisposing effect of OR >1.5-1.8, for 50%-10% exposure among controls respectively, and protective effects of OR <0.7-0.4, for 50%-10% exposure among controls respectively.

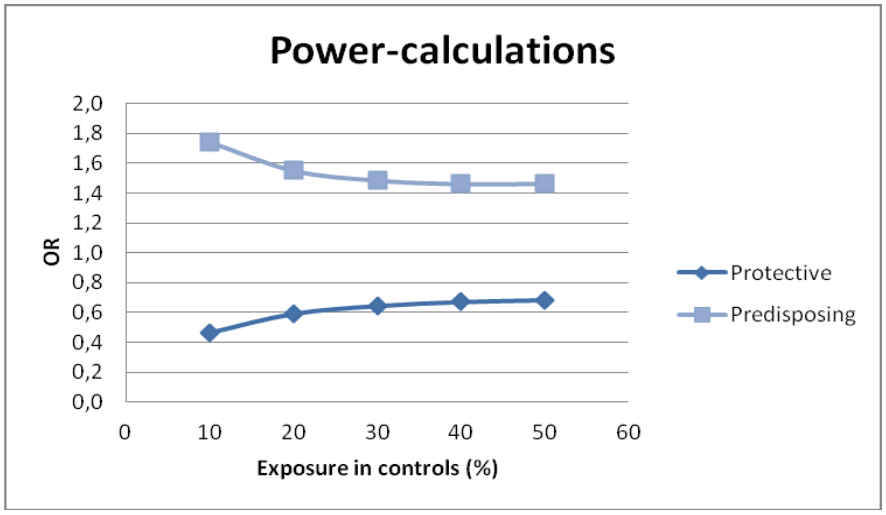


Figure 10. True OR for disease that we were able to detect in exposed subjects relative to unexposed subjects with 80% probability (power).

The questionnaire covered a number of previous diseases and environmental risk factors, but because of limited amount of time, we only had the opportunity to analyse some of them within the scope of this thesis. Based on previous knowledge about environmental risk factors, we chose to analyse the questions regarding smoking (question 14), periodontitis (q 10), alcohol consumption (q 20 and 21), coffee consumption (q 22 and 23), presence of pets and domestic animals during childhood (q 12), mononucleosis (q 7) and breastfeeding own children (q 41) (Table 24). These analyses gave us an overview of some of the most interesting environmental factors regarding RA development, and a clue about which factors should be investigated more comprehensively.

The effect of smoking, alcohol- and coffee consumption, periodontitis and presence of pets and domestic animals during childhood were analysed by stratification of ACPA status, as risk factors often differ between these RA subgroups. Stratification of ACPA positive and ACPA negative patients with regard to mononucleosis and breastfeeding own children, would resulted in low (<10) number of individuals for each group, these effects on RA development

were therefore analysed by including all RA patients together. For the question regarding breastfeeding own children, ≥13 months was used as a cut off when considering the duration of breastfeeding as long-term, as this has been used by other researchers, e.g. Pikwer M et al (43). Only women were included when analysing this question.

Results from the analyses of environmental risk factors (Table 24), showed that smoking, periodontitis and coffee consumption (current and at the age of 18 years) seemed to predispose for RA in the Norwegian RA population. Alcohol consumption (current and at the age of 18 years), pets and domestic animals during childhood, mononucleosis and breastfeeding among women seemed to have a protective effect in the Norwegian RA population.

Questions regarding alcohol- and coffee consumption were asked twice: i.e. at two time points (current and at the age of 18 years) in the questionnaire, as the risk factors early in life are of greater importance regarding RA development, than current consumption (after disease development). Even more significant p-values were detected when analysing the questions regarding exposure to the risk factors at 18 years old, compared to now a days.

Some of the risk factors were significantly associated in both ACPA positive- and ACPA negative patients, including periodontits, alcohol consumption at the age of 18 and pets and domestic animals during childhood. Pets and domestic animals during childhood were significantly more protective among ACPA positive patients compared to ACPA negative patients. Current alcohol consumption seemed like a protective environmental factor among ACPA positive patients and smoking seemed like a predisposing factor among ACPA negative patients. Except for the question regarding pets and domestic animals during childhood, we did not observe any significant differences for the two subgroups.

Mononucleosis appared to have protective effect and breastfeeding for at least 13 months among women showed a significant protective effect (Table 24).

Bonferroni-corrected p-values were calculated based on number of tests performed (see 3.5 Statistical analyses). Bonferroni- correction is quite strict (p= 0.0022 in our concern) but still, many of the risk factors investigated in this thesis, reached this significant threshold (Table 24). Smoking, current coffee consumption- and at the age of 18 for ACPA positive patients, and current alcohol consumption, coffee consumption to day and at 18 years in ACPA

negative patients, and mononucleosis did not reach the Bonferroni-corrected significance threshold.

Calculations of effects (OR) that can be detected with 80% power in our data material (Figure 10), demonstrated that the only risk factors with too weak effect (OR) was current alcohol consumption, the presence of pets and domestic animals for ACPA negative patients and mononucleosis (Table 24).

Table 24. Result from analyses of the questionnaire regarding selected environmental factors (see questionnaire, appendix)

	ACPA+ Total = 152 n(%)	ACPA- Total = 139 n(%)	Controls Total = 911 n(%)	ACPA pos vs ACPA neg OR (95%CI) p-value	ACPA pos vs controls OR (95%CI) p-value	ACPA neg vs controls OR (95%CI) p-value
Smoking (q14)	103(67.3)	87(70.7)	504(54.6)	0.85 (1.42-0.51) 0.547	1.70 (1.19-2.44) 0.0036	1.99 (1.33-2.99) 0.0009
Periodontitis (q10)	26(17.6)	28(23.3)	54(5.97)	0.70 (1.27-0.39) 0.24	3.38 (2.05-5.56) $1.63*10^{-06}$	4.81 (2.93-7.92) $6.27*10^{-10}$
Alcohol consumption now (q20)	111(72.5)	98(79.7)	809(87.7)	0.68 (1.19-0.39) 0.1734	0.37 (0.25-0.55) $1.12*10^{-06}$	0.54 (0.34-0.87) 0.0113
Alcohol consumption 18 years old (q21)	87(57.2)	66(53.7)	737(80)	1.16 (1.87-0.72) 0.5517	0.33 (0.23-0.48) $1.87*10^{-09}$	0.29 (0.20-0.43) $3.26*10^{-10}$
Coffee consumption now (q22)	127(87.6/)	104(92)	733(81.6)	0.63 (1.41-0.28) 0.2568	1.56 (0.93-2.59) 0.0901	2.48 (1.27-4.85) 0.0079
Coffee consumption 18 years old (q23)	74(52.1)	65(55.6)	371(41.5)	0.87 (1.42-0.54) 0.581	1.53 (1.08-2.18) 0.0177	1.76 (1.2-2.58) 0.0041
Pets and domestic animals during childhood (q12)	89(60.1)	81(66.9)	715(79.5)	0.39 (0.56-0.27) $3.53*10^{-7}$	0.29 (0.21-0.71) $1.32*10^{-12}$	0.52 (0.34-0.78) 0.0016

	Cases n (%)	Controls n (%)	Cases vs Controls OR (95%CI) p-value
Mononucleosis (q7)	18(6.6)	107(12.3)	0.52(0.31-0.86) 0.0106
Breastfeeding of own children ≥13months vs <13months (women only) (q41)	8(4.7)	71(16.2)	0.27(0.13-0.55) 0.0003

We also obtained data regarding smoking habits for842 cases, collected years ago (most of them, n>700, in 1994). 275 who answered the first time, did also answer the questionnaire (including questions regarding their smoking habits through life) sent out during the work of this thesis (2012/2013). We wanted to use the answers from these questions, to get an impression of the reliability of the answering in general. 250 individuals did answer the same now as they did previously. Among the 25 that gave a different answer in 2012/2013, 9 explained this by also stating that they had quit smoking. Therefore, 94.2% (259/275) gave reliable answers on the questions regarding smoking after ~20 years.

5. DISCUSSION

5.1 Several RA risk loci confirmed in the Norwegian RA population

Most of the SNPs significantly associated with RA in this study, are thought to be of immunological importance, as most of them are located near or in genes involved in the complex regulation of immunological pathways (e.g. *CCL6*, *RBPJ*, *IRF5*). Several of the SNPs are also reported, or in LD with reported SNPs, associated with other AID. This indicates their involvement in immunological processes, and strengthens their susceptibility to involvement in RA.

11 of the 32 successfully genotyped SNPs were significantly associated with RA development in the Norwegian population (Table 19- 21 on page 43-46). By stratification of autoantibody status, we observed three significant SNPs; rs3093023 (*CCR6*), rs8026898 (*TLE3*) and rs874040 (*RBPJ*) for autoantibody positive RA, and two significant SNPs; rs34536443 (*TYK2*) and rs6859219 (*ANKRD55/IL6ST*) for autoantibody negative RA. Additionally six SNPs were significant associated in either group of the autoantibody positive or autoantibody negative RA. For ACPA negative RA, the additional significant SNPs were rs11676922 (*AFF3*), rs2736340 (*BLK*) and rs8043085 (*RASGRP1*). The SNP rs7155603 (*BATF*) was significantly associated with ACPA positive RA. rs10488631 (*IRF5*) was significantly associated in RF positive disease and rs10821944 (*ARID5*) was significantly associated in RF negative disease. None of the SNPs genotyped were significantly associated in both autoantibody positive and negative disease in the Norwegian population.

SNPs significantly associated in the Norwegian RA population in this study were selected from studies where they showed association in ACPA positive patient groups (15, 31). Therefore, SNPs associated in ACPA negative disease in our study, are most likely associated in both subgroups. Limited power is a likely explanation for why we did not observe an association also in the ACPA positive group. Most previous studies have notably focused on ACPA positive RA, and hence ACPA negative RA has been much less explored. SNPs associated with ACPA positive disease in our study (and not ACPA negative), might only be associated with ACPA positive RA, and these findings support the idea that different genetic background contribute to development of the two RA subgroups, as suggested by Padyokov et al (76). Some SNPs were only associated with ACPA positive RA and not RF positive RA, and others were only associated with ACPA negative RA and not RF negative RA, and visa

versa. One might think that because ACPA and RF define largely overlapping populations of patients, the different associations might be due to coincidence for the association s was detected in one group and not in the other. But from Table 20 and 21 on page 43-46, we see that these SNPs did not even seem to have a tendency to be associated for the other autoantibody. ACPA is more specific for RA than RF (17), and this implicate that there is differences between the two subgroups, and they therefore might also have different genetic contributions.

The RA risk loci are labelled with names of the most compelling candidate gene(s) from each region of LD, based upon analysis of published connections among genes and/or knowledge of RA pathogenesis (15). The majority of the newly validated loci associated with RA susceptibility included in this study contain genes that are strongly linked to immune function. Next, I will present the significant associated SNPs from Table 19-21 (on page 43-46), and their related genes. Information regarding the genes has been collected from NCBIs Gene-database (http://www.ncbi.nlm.nih.gov/gene/). However, it should be noted that the causal genes within each region have not yet been identified, and fine mapping and functional studies are needed to determine which gene carries the directly involved risk variant and how that variant biologically is involved in the RA pathogenesis.

SNPs significantly associated with ACPA- and RF positive RA

rs3093023 is located in a gene called *Chemokine receptor 6* (*CCR6*). The receptor is expressed by T helper-cells that produce IL-17, and is involved in IL-17 driven inflammation observed in RA and other chronic inflammations (29). rs3093023 is almost in complete LD (r^2>0,99) with rs3093024, associated with RA in the Japanese RA population (77). The rs3093023 SNP is also to some degree in LD (r^2=0.48) with rs2301436, associated with Crohn`s disease (15). These findings provide strong evidence for association of the *CCR6* locus with risk of RA.

rs874040 is located near a gene called *recombinant binding protein for immunoglobulin kappa J region* (*RBPJ*). This gene encodes a transcription factor within the Notch signalling pathway, and acts as a repressor by binding to Notch proteins. Notch proteins are trans-membrane proteins, important for cell-cell communication and plays important roles during T cell-mediated immune response (78). RBPJ has also been reported to be associated with Type I diabetes (rs10517086, r^2 = 1 with rs874040) (15).

rs8026898 is located near a gene called *transducin-like enhancer of split 3*(*TLE3*). TLE genes encodes corepressors thought to negatively regulate transcription and play critical roles in developmental and cellular pathways (79).

SNPs significantly associated with ACPA- and RF negative RA

rs34536443 is located in a gene called *tyrosine kinase 2* (*TYK2*), which encodes an enzyme that transfers phosphate groups from high-energy donor molecules (e.g. ATP) to tyrosine on proteins, and function as an "on" or "off" switch in many cellular functions. The SNP cause a nonsynonymous mutation within exonic region (31), by the change from the amino acid proline to alanine when C is replaced with G. A mutation in the *TYK2* gene has been associated with a primary immunodeficiency characterized by elevated serum immunoglobulin E. The SNP is also associated in multiple sclerosis, at genome wide significance level (80).

rs6859219 is located in a gene called *ankyrin repeat domain-containing gene* (*ANKRD55*) near *interleukin 6 signal transducer* (*IL6ST*). *ANKRD55* is a gene of unknown function, but *IL6ST* is a more plausible immunological candidate, which lies ~150kb proximal to rs6859219, but outside the region of LD with associated SNPs (15). *IL6ST* encodes glycoprotein130, which associate with e.g. IL6 bound to IL6receptor to form a complex which can produce downstream signals, involved in inflammation and maturation of B-cells.

SNP significantly associated with ACPA positive RA

rs7155603 is located close to a gene named *basic leucine zipper transcription factor, ATF-like* (*BATF*), which encodes a transcription factor that mediates dimerization with proteins. It is thought to be a negative regulator of AP-1/ATF transcriptional events in T-cells, in response to distinct stimuli, including cytokines, and in turn controls a number of cellular processes like differentiation, proliferation and apoptosis (81).

SNPs significantly associated with ACPA negative RA

rs8043085 is located in a gene called *RAS guanyl releasing protein 1* (*RASGRP1*). It function by activating Ras through the exchange of bound GDP to GTP, it activates Erk/MAP kinase cascade and regulates T- cell and B-cell development, homeostasis and differentiation.

rs2736340 is located near a gene called *B lymphocyte kinase* (*BLK*), a non receptor tyrosine kinase, which has a role in B-cell receptor signalling and B-cell development. The SNP is

associated as a susceptibility factor for systemic sclerosis (82), and Zhou et al provide evidence for possible gene-gene interactions of *BLK*, *TNFAIP3*, *REL*, *TNFSF4* and *TRAF1* in systemic lupus erythematous in Chinese, which may represent a synergic effect on T-and B-cells in determining immunological aberration (83).

rs11676922 is located near *AF4/FMR2 family, member 3* (*AFF3*) which is a tissue restricted nuclear transcriptional activator expressed in lymphoid tissue, and may function in lymphoid development. *AFF3* has previously been implicated with RA (rs10865035, rs1160542 and rs9653442) (84) and the latter with equivocal evidence for association with Type I diabetes (15).

SNP significantly associated with RF positive RA

rs10488631 is located near a gene called *interferon regulatory factor 5* (*IRF5*), which is a member of a group transcription factors (IRF) with roles including virus-mediated activation of interferon, modulation of cell growth, differentiation, apoptosis and immune system activity, and *IRF5* is a common susceptibility factor for several rheumatic and autoimmune diseases, like systemic lupus erythematosus and juvenile idiopathic arthritis (15, 85). IRF5 variants are correlated with expression of alternative IRF5 transcripts in thymus, which imply a regulatory role (85). The investigation carried out by Nordang et al did not manage to confirm association between RA and the IRF5 polymorphism rs2004640, and reported that this most likely was due to lack of power caused by limited sample size (85), even though the association was carried out in the same patient and control- material as for this thesis, but the RA patients was divided into two cohorts. The rs2004640 SNP might be significantly associated with RA if the cases are not split, or it might implicate that rs2004640 is not the causal variant in regard to RA, since we managed to confirm association of rs10488631 in the current study (r^2=0.15 and D`=1).

SNP significantly associated with RF negative RA

rs10821944 is located in a gene called *AT rich interactive domain 5B* (*ARID5B*), encoding a DNA binding protein, and plays a role in cell growth and differentiation B-lymphocyte progenitors. The SNP has been associated with RA in samples of Japanese ancestry, and is moderately correlated with rs12764378 (r^2 = 0.52) (31). We have also genotyped rs12764378, but it did not show significantly association in our study (Table 19-21 on page 43-46), but showed a trend towards association (p = 0.07) in RF negative patients (Table 21), which

indicate an association between this SNP and RA, which might get significantly associated with larger sample size.

Several of the SNPs did not show significant ($p<0.05$) association in the Norwegian RA population. Because they are reported associated with RA in European populations we would expect them to be associated with RA in the Norwegian population. It can be hard to detect associations due to the relatively small sample size (Table 15 on page 39), and increasing the sample size and thereby the power, would most likely have revealed more significantly associated SNPs. Differenced in LD pattern between populations can make it hard to the catch some of SNPs associated in some populations, if the causal variants are not tested themselves as the strength of the LD between the tested markers and the true risk loci might differ.

5.2 Environmental factors associated with RA

From our questionnaire study, smoking, periodontitis and coffee consumption seemed to increase the risk of RA, whereas periodontitis, alcohol consumption, presence of pets and domestic animals during childhood, mononucleosis and breastfeeding own children among women seemed to decrease the risk of RA development (Table 24 on page 52-53).

Ever-smoking showed significantly increased risk of RA development among ACPA negative patients (OR=1.99, p=0.0009) in our study. The Bonferroni correction is quite strict (p=0.002), and even though smoking did not reach this significant threshold for ACPA positive RA, it seemed like smoking also increase the RA risk for this subgroup (OR 1.7, p= 0.0036). As stated in the introduction, smoking is the best established environmental risk factor and interactions between smoking, genotype and autoantibodies are of fundamental importance for the hypothetical pathogenesis of RA (Figure 2 on page 14) (34). In the literature, smoking is associated with an increased risk of ACPA-positive, and not ACPA-negative, RA (86). Our findings may be influenced by the skewing regarding percentage of ACPA positive patients who answered the questionnaire (55.3%) compared to the initial percentage of ACPA positive patients included in this study (61.6%). This might reduce the power to detect associations within the ACPA positive subgroup. The observed significant increased risk of RA due to smoking might be biased by the difference in age between patients and controls, as smoking habits have overall changed in the population. Age should be taken into account as a confounder in further analysis. Unfortunately, there was not enough

time to analyse all data obtained regarding the use of and/or exposure to tobacco, in regard to pack years, passive smoking and the use of snuff (question 15-19). Smoking has been shown to increase RA risk in a dose dependent manner (34), and analyses of this should therefore be carried out. Since there are many unsolved questions regarding passive smoking and the effect of snuff usage on disease development, investigations regarding these questions should be addressed out by the use of our data.

Periodontitis significantly increased the risk of RA development, for both ACPA positive and ACPA negative disease, in our study. This may be due to *P. gingivalis,* a bacterium which is present abundantly in periodontal tissue and expresses PAD enzymes that citrullinate proteins. Increased citrullination in patients with periodontitis may result in break of immune tolerance to citrullinated proteins, and play a causal role in the initiation of ACPA positive RA, but the mechanisms whereby periodontitis leads to an ACPA response and RA are unknown (87). Pablo et al showed that the antibody response in periodontitis was directed to uncitrullinated peptides of RA autoantigens, and proposed that loss of tolerance predominantly directed to uncitrullinated peptide of RA autoantigens could lead to epitope spreading to citrullinated epitopes (87). The observed significant increased risk of RA due to periodontitis in our study might be biased by the difference in age between cases and controls, as the cases in general are almost 20 years older than the controls and the chance of being exposed to periodontitis increase with age, and stratification for age should therefore be carried out.

We observed significant decreased risk of RA among individuals who drink alcohol (nowadays), especially among ACPA positive patients. This is consistent with the findings by Kallberg et al (45). Alcohol consumption was analysed in a never/ever- manner, but based on the observations previously reported (45), this should also be tested in a dose-dependent manner. Such an analysis is possible based on the data collected through the questionnaire, but the strength of this analysis might be weakened due to small sample size for each group. Further investigation of alcohol consumption based on our material should take in to account that alcohol consumption possibly can be biased by gender and age, and most likely also smoking.

The reduced risk of RA within patients versus controls due to alcohol consumption was more significant and with stronger reduced effect at the age of 18 than nowadays (Table 24). Even though answers regarding the past are more uncertain, alcohol consumption at the age of 18

years (before disease onset) is more likely to directly affect RA development than alcohol consumption nowadays (after disease development). Alcohol consumption at the age of 18 years should therefore be analysed with more advanced statistics, taking smoking (and other possible confounders like gender and year of birth since alcohol habits might have changed over time) in to account as a covariable.

Our study indicated that coffee consumption increase the risk of RA development, even though it did not reach the Bonferroni- corrected p-value threshold. Coffee appeared stronger associated with risk of ACPA negative RA than ACPA positive RA, but no significant difference in coffee consumption was observed when we compared ACPA positive patients versus ACPA negative patients. Studies regarding coffee consumption and RA risk are inconclusive (86, 88). In a study of 515 patients recently diagnosed wit RA and 769 controls, Pedersen et al showed that coffee consumption (10 years before interview) significantly increased risk of ACPA positive RA, but not ACPA negative RA (86). Further research is needed to determine the effect of coffee consumptions on RA, and whether it is a subtype-specific risk factor or not. As for alcohol, we observed a stronger association between coffee consumption at the age of 18 years than nowadays. Because exposure before disease onset is more likely to affect disease development, we recommend that analysing data regarding consumption at 18 years old should be emphasised more than analysing consumption nowadays.

Exposure to pets and domestic animals during adolescence showed significantly decreased disease development in our study. Looking at this data the other way around; lack of exposure to animals during childhood might increase RA risk. Even though there is an ongoing debate whether infection prevents or predisposes autoimmune diseases, our findings support the hygiene hypothesis, by which early childhood exposure to infections is thought to better "prime" the immune system, and lack of such infections might suppress natural development of the immune system (37). Findings in another research group that we have been cooperating with regarding development of the questionnaire, also implicated that pets and domestic animals during childhood significantly reduced the risk of multiple sclerosis, also after age and sex stratification (Gustavsen MW, unpublished). It is therefore reason to believe that our findings also persist significant after stratification, but needs to be performed.

Mononucleosis implicated decreased risk of RA in our material, even though it did not reach Bonferroni-corrected p-value threshold. Because the number of individuals exposed to

mononucleosis was low, the analysis was carried out by calculating the difference between all cases and controls (not stratified by ACPA status). Mononucleosis is caused by the Epstein Barr virus, and because the severity varies between individuals, it is often underestimated. It is plausible that mononucleosis has been more underestimated in among patients than controls in our study, because of their older age and that the disease was less frequently detected in the past. It is also likely that more patients have had the disease, as chance of being exposed to the virus increases with age (as for periodontitis). Overall, the data regarding mononucleosis are rather uncertain, but if further analysis is carried out, one should at least stratification for age.

Because 2/3 of the affected RA patients are women, we included some questions related to pregnancy and reproductive health among women in the questionnaire. We observed that breastfeeding for at least 13 months, significantly reduced the risk of RA, as observed in previous studies (23). The protective effect has been proposed to be related to long-term increased immune tolerance (43) but this needs further investigation. As noted in the introduction, many factors can participate to the increased risk among women, and it is thought that hormones play a key role. Little is yet known, and it can be hard to investigate- and separate the effects of the different factors (e.g. childbirth and breastfeeding). Among the questions regarding pregnancy and reproductive health among women, only data regarding breastfeeding have been analysed in this thesis. The questionnaire also include questions regarding menarche, the use of oral contraceptives and menopause (question 35-41), which should be investigated, as there are still many unsolved questions regarding the high representation of women among RA patients. For this study, we only included breastfeeding-data reported regarding the first childbirth. Further investigations should include breastfeeding-data reported from all childbirths.

The risk factors smoking, alcohol- and coffee consumption at the age of 18 years, periodontitis, presence of pets and domestic animals and breastfeeding own children among women should be analyzed with more advanced statistics, taking age and sex into account. Further investigations of environmental risk factors based on the questionnaire and answers obtained from this study, should include analysis regarding the use of- and exposure to tobacco (question 14-19), and regarding pregnancy and reproductive health among women (question 35-41). Among questions not considered in this thesis are socioeconomic status (question 31-33) which has shown to be associated with risk of RA, questions regarding food

and vitamin supplement habits (question 24-37) which have shown no or equivocal results, (23) and questions regarding sun exposure, (question 28-30) as exposure to ultraviolet light has been proposed to explain the North-to-South gradient observed (89). Analysis including these questions should be carried out.

The accuracy of the answers given in a questionnaire study is of uttermost importance for the reliability of the results. We see that more than 94% of the patients answered the same regarding smoking when they answered the question this year as they did almost 20 years ago (Table 23 on page 49). Less than 6% gave different answers, and among those who did, some answered "never" the first time and "ever" in 2012/2013, and almost the same amount answered "ever" the first time and "never" in 2012/2013. Taking into account that the patients are quite old (on average ~65 years), and most likely have not started smoking since the first time they were asked the question, I think the correspondence in the answers they gave was satisfying. Furthermore, one cannot exclude that some of the patients actually changed their smoking status in the time period between answering the two questionnaires. Overall, the consistency is quite interesting regarding how trustworthy information obtained by questionnaires is.

5.3 Methodological considerations

5.3.1 Quality of genotyping results

High quality control is essential to avoid false results and to ensure reliable results, which reflect the actual genotypes. Quality control can be carried out by manually looking at the generated plots, and by statistical calculations of GSR and HWE.

Not all the SNPs and genes that we wanted to investigate could be tested, due to different reasons. Three of the SNPs genotyped were removed before additional analyses; rs934734 and rs5029937 failed genotyping and rs706778 show several individuals clustering between two clusters of AG and AA (Figure A1, appendix), and therefore cannot be confidently called.

rs934734 (*SPRED2*) and rs5029937 (*TNFAIP3*) failed genotyping for all individuals. This may be due to failed primer design, and therefore no PCR product which can be genotyped. rs934734 (*SPRED2*) is located < 50 bp upstream from a repeated area, which might cause difficulties regarding primer-design and amplification of the area.

Several individuals (~130) could not be confidently called when genotyping rs706778 (*IL2RA*) (Figure A1, appendix), as they cluster between the heterozygous cluster and one of the homozygous clusters. We did not observe any SNPs near by, but the SNP is located only 30 bp downstream of a repetitive area, which can cause difficulties with primer design. Another explanation could be that the SNP region is duplicated in the genome and fixed at one duplicated chromosome (e.g. AA) and variable at the other (e.g. AG). This is frequently seen in the genome as a result from historical duplication events. The SNP appeared significantly associated in the Norwegian RA population ($p=2.45*10^{-5}$), and sequencing should therefore be carried out to figure out these individuals genotype, and thereafter the true allele frequencies in cases versus controls.

HWE and GSR

We would expect the SNPs related to autoimmune diseases to be in HWE for both the patient group and the control group, since the risk effect exerted by the SNPs are low. This is in line with our observation, as all our SNPs were in HWE. Hence, this indicated that our genotyping was of good quality.

We observed two SNPs with a tendency towards deviate from HWE for the controls, rs595158 and rs99799383 (Table 16 on page 40-41). We therefore controlled to what extent the genotypes differed from HWE relative to the expected distribution-, and more heterozygous genotypes was observed among the controls than expected for rs595158 (Table 17), while less heterozygous genotypes than expected for rs99799383 (Table 18). These differences were not considered large enough to have skewed the results, and we concluded that the genotype quality was satisfactory. The SNPs were not among the SNPs we found associated with RA predisposition.

All of the SNPs (except for the rs13119723 SNP) had a GSR above 95% (Table 16 on page 40-41). This means that more than 95% of the individuals had been successfully genotyped for each SNP, and ensured that not any specific group of genotype was lost during genotyping, to such an extent that a skewing could give unreliable results.

As an extra control, in addition to HWE and GSR, we checked that the minor allele frequencies listed in the articles the SNPs were selected from (15, 31, 62, 63) were approximately the same as the frequencies we observed in our population (Table 4 on page 22-23).

5.3.3 Sequencing

We observed an "extra cluster" between the heterozygote cluster and the homozygote cluster closest to the X-axis when analysing the TaqMan results for the rs13119723 SNP (Figure A2, appendix). Checking the sequence surrounding this SNP by UCSC Genome browser, we observed another SNP (rs114092637) two base pairs upstream of the rs13119723. The probes used for TaqMan allele discrimination have the major allele (T) in the position of rs114092637, and will therefore not bind sequences with C in this position. Since the C allele at this new SNP was in LD with the G allele at our SNP, the presence of the C allele will reduce the FAM signal (as observed), and thereby cause the "extra cluster". Based on this observation, our hypothesis was therefore that the individuals clustering in "the extra cluster" is heterozygous for both SNPs.

The sequencing-results (Table 22 on page 48) supported this hypothesis, because all individuals clustering in the extra cluster were heterozygous for the two positions (rs114092637 and rs13119723), and the rest of the individuals had the major allele in the rs114092637-position, as expected. When analysing the sequencing result, we also looked at the rest of the sequences, but could not find any other polymorphisms that could explain the extra cluster.

The frequency of individuals in our material (0.05) observed in the extra cluster (meaning carrying minor allele for rs114092637) was almost the same as the reported frequency of minor allele for rs114092637 (0.049) (dbSNP buildt137), further supporting our hypothesis.

Both patients and controls were represented in the extra cluster, and the combination of the two SNPs can not be considered associated with RA based on our findings.

5.3.3 Obtaining information by the use of questionnaire

Studies of the impact of environmental factors on risk of developing RA are associated with several methodological and practical challenges. The main acceptable methods for providing knowledge on environmental factors are population-based case-control studies and cohort studies (23). The drawback for case-control studies is the risk of bias caused in recruitment of patients, and recall bias in responses, especially for cases diagnosed years ago. Cohort studies often suffer from low power, but are less subject to recruitment- and recall bias (23).

287 patients and 922 controls were included for analysing the questionnaire (Table 23 on page 49). The effects (OR) we are able to detect with 80% power given our data vary according to percentage of exposed cases (Figure 10 on page 50). Differences in mean age between the patients and the controls, and the relative overrepresentation of males in the control croup compared to the patient group, may have influenced the results. Except that one group is affected by disease and the other is not, we would prefer homogeneity between the two groups, and differences (e.g. age and gender) between the two groups should be stratified or corrected for when analysing risk factors thought to be affected by differences in age and gender.

In the process of sending out questionnaire, we observe that more ACPA positive patients than ACPA negative patients had passed away. Only ~53% of the patients who answered the questionnaire were ACPA positive (Table 23 on page 49), this was lower than what we would expected from the literature where ~60-70% of the RA patients are ACPA positive. Approximately 62% of the initial patient-group was ACPA positive (Table 15 on page 39), but because quite a few had passed away, the questionnaire was sent to only 56% ACPA positive patients. ACPA positive RA is reported to have a more rapid disease course with progressive joint damage and low remission rate (23), and this is supported by our observation. The differences observed between the two subgroups might indicate that different genetic background, as well as environmental risk factors, underlie the two subgroups, and study them separate increase the power to detect associations between the groups. Reduced ACPA positive patients compared to ACPA negative patients might bias some of the risk factors analysed, and should be taken into account when analysing and interpreting the results.

We got feedback from caregivers of some of the RA patients, with the message that the patients were not able to answer the questionnaire due to cognitive impairments caused by the disease. This is reflected by lower response rate for the patient group (Figure 9 on page 48).

In addition, answering the questions regarding everyday habits decades a go, might cause some uncertainty in the answers, probably in both groups, but in particular for the patients, considering their higher age. On the other hand, it is likely that patients with chronic disorders focus more on possible risk factors, and therefore thought more carefully about this when they answered the questionnaire.

We should also consider the fact that the control group are chosen from a bone marrow donor registry, and therefore could be a somewhat selected group of healthy individuals, and this might indicate that they are more focused on healthcare than the general population.

5.4 Missing heritability

Throughout this thesis we have identified 11 significantly associated RA risk loci and suggested association between RA development and environmental risk factors (e.g. smoking, periodontitis and pets and domestic animals during childhood) in the Norwegian RA population. The link between genes and environmental risk factors are largely unknown, and studies indicate that interactions between genetic risk loci and environmental risk factors might increase the effect size of single risk genes.

About 60% of the risk of developing RA is caused by genetic predisposition. *HLA* risk alleles are estimated to explain 11-37% of the genetic risk, and confirmed non-HLA RA risk loci are estimated to explain 15%. All together confirmed risk loci (*HLA* + non-HLA) explain 51% of the estimate of heritability. Researchers propose different explanations for the missing heritability in RA and other complex diseases (90).

One explanation could be that the effect size of single risk genes are underestimated due to gene-environment and gene-gene interactions, which might result in higher heritability estimates than for the gene alone. The effect size estimated might be reduced when the causal variant is not localised, and LD between the variant detected and causal variant could lead to estimation of indirect association. Rare variants (MAF<5%) are poorly detected by GWAS which is built up on the "common disease/ common variant" hypothesis. Therefore, rare variants (possibly with large effect) might explain some of the missing heritability. Structural variation (Figure 3 on page 15) may account for some of the "missing heritability", as it has been largely unexplored in relation to complex traits. RA is a heterogeneous disease, and grouping patients regarding ACPA status has shown different genetic contribution in each group, which would not be discovered if the two groups were merged together and compared with controls. Grouping patients regarding autoantibody status might not be the only way to separate RA patients, and this might increase the possibility to discover other genetic factors.

5.5 Conclusion

We confirmed 11 risk loci in the Norwegian RA population, five of these were associated in autoantibody positive RA and six were associated with autoantibody negative RA. The SNPs were collected from studies carried out for ACPA positive RA, indicating that the SNPs associated with autoantibody negative disease in this thesis, is genetic risk factors in both subgroups. The SNPs only associated in ACPA positive disease in this thesis, support the idea of different genetic predisposition in the two autoantibody subtype of disease.

We suggest that smoking, periodontitis and coffee consumption predispose RA, and that alcohol consumption, pets and domestic animals during childhood and breastfeeding own children protect against RA development. Further analyses based on the questionnaire should adjust for gender and age, and include the variables being analysed as covariables.

5.6 Further investigation

Further research of RA should include fine mapping, to identify the causal variant. Population-based case-control studies and cohort studies should be carried out to increase the understanding of environmental risk factors. Grouping patients regarding ACPA status is proved to be of great importance, and should persist for both genetic- and environmental studies. The importance of gene-environmental interactions should be investigated, and more detailed studies of how certain disease-related genes are epigenetically regulated and how environmental or internal influences might contribute to such changed epigenotypes are needed. Next-generation sequencing provide an opportunity to complete the analysis of the relationship between genomic variation and genotype, and not using only partial information as GWAS and LD, which will not be able to explain the full range of genetic susceptibility alone (91). Increased knowledge regarding biological functions of disease development and sustained RA is important for development of better treatment strategies, and lead us one step closer to the ultimate goal; remission or sustained low disease with reduced pain and maintenance of function.

6. REFERENCE LIST

1. Lea T. Immunologi og immunologiske teknikker2006.
2. Dranoff G. Cytokines in cancer pathogenesis and cancer therapy. Nature reviews Cancer. 2004 Jan;4(1):11-22. PubMed PMID: 14708024.
3. Rai E, Wakeland EK. Genetic predisposition to autoimmunity--what have we learned? Seminars in immunology. 2011 Apr;23(2):67-83. PubMed PMID: 21288738. Epub 2011/02/04. eng.
4. Eaton WW, Rose NR, Kalaydjian A, Pedersen MG, Mortensen PB. Epidemiology of autoimmune diseases in Denmark. Journal of autoimmunity. 2007 Aug;29(1):1-9. PubMed PMID: 17582741. Pubmed Central PMCID: 2717015.
5. Cooper GS, Bynum ML, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. Journal of autoimmunity. 2009 Nov-Dec;33(3-4):197-207. PubMed PMID: 19819109. Pubmed Central PMCID: 2783422.
6. Choy E. Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis. Rheumatology. 2012 Jul;51 Suppl 5:v3-11. PubMed PMID: 22718924.
7. Uhlig T, Kvien TK, Glennas A, Smedstad LM, Forre O. The incidence and severity of rheumatoid arthritis, results from a county register in Oslo, Norway. The Journal of rheumatology. 1998 Jun;25(6):1078-84. PubMed PMID: 9632067. Epub 1998/06/19. eng.
8. Tobon GJ, Youinou P, Saraux A. The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. Journal of autoimmunity. 2010 Aug;35(1):10-4. PubMed PMID: 20080387.
9. Harvey J, Lotze M, Stevens MB, Lambert G, Jacobson D. Rheumatoid arthritis in a Chippewa Band. I. Pilot screening study of disease prevalence. Arthritis and rheumatism. 1981 May;24(5):717-21. PubMed PMID: 7236326.
10. Silman AJ, Ollier W, Holligan S, Birrell F, Adebajo A, Asuzu MC, et al. Absence of rheumatoid arthritis in a rural Nigerian population. The Journal of rheumatology. 1993 Apr;20(4):618-22. PubMed PMID: 8496853.
11. Lau E, Symmons D, Bankhead C, MacGregor A, Donnan S, Silman A. Low prevalence of rheumatoid arthritis in the urbanized Chinese of Hong Kong. The Journal of rheumatology. 1993 Jul;20(7):1133-7. PubMed PMID: 8371205.
12. Sokka T, Kautiainen H, Pincus T, Verstappen SM, Aggarwal A, Alten R, et al. Work disability remains a major problem in rheumatoid arthritis in the 2000s: data from 32 countries in the QUEST-RA study. Arthritis research & therapy. 2010;12(2):R42. PubMed PMID: 20226018. Pubmed Central PMCID: 2888189.
13. Michaud K, Wolfe F. Comorbidities in rheumatoid arthritis. Best practice & research Clinical rheumatology. 2007 Oct;21(5):885-906. PubMed PMID: 17870034.
14. Avina-Zubieta JA, Choi HK, Sadatsafavi M, Etminan M, Esdaile JM, Lacaille D. Risk of cardiovascular mortality in patients with rheumatoid arthritis: a meta-analysis of observational studies. Arthritis and rheumatism. 2008 Dec 15;59(12):1690-7. PubMed PMID: 19035419.
15. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nature genetics. 2010 Jun;42(6):508-14. PubMed PMID: 20453842. Epub 2010/05/11. eng.
16. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis and rheumatism. 1988 Mar;31(3):315-24. PubMed PMID: 3358796.
17. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet. 2009 Feb 21;373(9664):659-72. PubMed PMID: 19157532. Epub 2009/01/23. eng.
18. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against

Rheumatism collaborative initiative. Annals of the rheumatic diseases. 2010 Sep;69(9):1580-8. PubMed PMID: 20699241. Epub 2010/08/12. eng.
19. Vossenaar ER, Zendman AJ, van Venrooij WJ, Pruijn GJ. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. BioEssays : news and reviews in molecular, cellular and developmental biology. 2003 Nov;25(11):1106-18. PubMed PMID: 14579251.
20. van der Linden MP, van der Woude D, Ioan-Facsinay A, Levarht EW, Stoeken-Rijsbergen G, Huizinga TW, et al. Value of anti-modified citrullinated vimentin and third-generation anti-cyclic citrullinated peptide compared with second-generation anti-cyclic citrullinated peptide and rheumatoid factor in predicting disease outcome in undifferentiated arthritis and rheumatoid arthritis. Arthritis and rheumatism. 2009 Aug;60(8):2232-41. PubMed PMID: 19644872. Epub 2009/08/01. eng.
21. Klareskog L, Ronnelid J, Lundberg K, Padyukov L, Alfredsson L. Immunity to citrullinated proteins in rheumatoid arthritis. Annual review of immunology. 2008;26:651-75. PubMed PMID: 18173373.
22. Aggarwal R, Liao K, Nair R, Ringold S, Costenbader KH. Anti-citrullinated peptide antibody assays and their role in the diagnosis of rheumatoid arthritis. Arthritis and rheumatism. 2009 Nov 15;61(11):1472-83. PubMed PMID: 19877103. Pubmed Central PMCID: 2859449.
23. Liao KP, Alfredsson L, Karlson EW. Environmental influences on risk for rheumatoid arthritis. Current opinion in rheumatology. 2009 May;21(3):279-83. PubMed PMID: 19318947. Pubmed Central PMCID: 2898190.
24. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell. 1996 May 3;85(3):311-8. PubMed PMID: 8616887.
25. Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A, et al. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. British journal of rheumatology. 1993 Oct;32(10):903-7. PubMed PMID: 8402000. Epub 1993/10/01. eng.
26. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis and rheumatism. 2000 Jan;43(1):30-7. PubMed PMID: 10643697.
27. Smith DJ, Lusis AJ. The allelic structure of common disease. Human molecular genetics. 2002 Oct 1;11(20):2455-61. PubMed PMID: 12351581.
28. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. Lancet. 2010 Sep 25;376(9746):1094-108. PubMed PMID: 20870100.
29. Kurko J, Besenyei T, Laki J, Glant TT, Mikecz K, Szekanecz Z. Genetics of Rheumatoid Arthritis - A Comprehensive Review. Clinical reviews in allergy & immunology. 2013 Jan 5. PubMed PMID: 23288628.
30. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis and rheumatism. 1987 Nov;30(11):1205-13. PubMed PMID: 2446635. Epub 1987/11/01. eng.
31. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nature genetics. 2012 Dec;44(12):1336-40. PubMed PMID: 23143596. Pubmed Central PMCID: 3605761.
32. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. Nature biotechnology. 2012 Nov;30(11):1095-106. PubMed PMID: 23138309.
33. Voight BF, Cotsapas C. Human genetics offers an emerging picture of common pathways and mechanisms in autoimmunity. Current opinion in immunology. 2012 Oct;24(5):552-7. PubMed PMID: 23041452.
34. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Annals of the rheumatic diseases. 2003 Sep;62(9):835-41. PubMed PMID: 12922955. Pubmed Central PMCID: 1754669.

35. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, et al. 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 and rheumatism. 2006 Jan;54(1):38-46. PubMed PMID: 16385494. Epub 2005/12/31. eng.
36. Strachan DP. Hay fever, hygiene, and household size. Bmj. 1989 Nov 18;299(6710):1259-60. PubMed PMID: 2513902. Pubmed Central PMCID: 1838109.
37. Ellis JA, Munro JE, Ponsonby AL. Possible environmental determinants of juvenile idiopathic arthritis. Rheumatology. 2010 Mar;49(3):411-25. PubMed PMID: 19965974.
38. Franssila R, Hedman K. Infection and musculoskeletal conditions: Viral causes of arthritis. Best practice & research Clinical rheumatology. 2006 Dec;20(6):1139-57. PubMed PMID: 17127201.
39. Ruiz-Esquide V, Sanmarti R. Tobacco and other environmental risk factors in rheumatoid arthritis. Reumatologia clinica. 2012 Nov-Dec;8(6):342-50. PubMed PMID: 22609003.
40. Lis J, Jarzab A, Witkowska D. [Molecular mimicry in the etiology of autoimmune diseases]. Postepy higieny i medycyny doswiadczalnej. 2012;66:475-91. PubMed PMID: 22922148. Rola mimikry molekularnej w etiologii schorzen o charakterze autoimmunizacyjnym.
41. de Pablo P, Dietrich T, McAlindon TE. Association of periodontal disease and tooth loss with rheumatoid arthritis in the US population. The Journal of rheumatology. 2008 Jan;35(1):70-6. PubMed PMID: 18050377.
42. Disanto G, Hall C, Lucas R, Ponsonby AL, Berlanga-Taylor AJ, Giovannoni G, et al. Assessing interactions between HLA-DRB1*15 and infectious mononucleosis on the risk of multiple sclerosis. Multiple sclerosis. 2013 Feb 14. PubMed PMID: 23413297.
43. Pikwer M, Bergstrom U, Nilsson JA, Jacobsson L, Berglund G, Turesson C. Breast feeding, but not use of oral contraceptives, is associated with a reduced risk of rheumatoid arthritis. Annals of the rheumatic diseases. 2009 Apr;68(4):526-30. PubMed PMID: 18477739.
44. Pikwer M, Bergstrom U, Nilsson JA, Jacobsson L, Turesson C. Early menopause is an independent predictor of rheumatoid arthritis. Annals of the rheumatic diseases. 2012 Mar;71(3):378-81. PubMed PMID: 21972241.
45. Kallberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, Garred P, et al. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. Annals of the rheumatic diseases. 2009 Feb;68(2):222-7. PubMed PMID: 18535114. Pubmed Central PMCID: 2937278.
46. Trenkmann M, Brock M, Ospelt C, Gay S. Epigenetics in rheumatoid arthritis. Clinical reviews in allergy & immunology. 2010 Aug;39(1):10-9. PubMed PMID: 19707891.
47. Yang SR, Wright J, Bauter M, Seweryniak K, Kode A, Rahman I. Sirtuin regulates cigarette smoke-induced proinflammatory mediator release via RelA/p65 NF-kappaB in macrophages in vitro and in rat lungs in vivo: implications for chronic inflammation and aging. American journal of physiology Lung cellular and molecular physiology. 2007 Feb;292(2):L567-76. PubMed PMID: 17041012.
48. Takami N, Osawa K, Miura Y, Komai K, Taniguchi M, Shiraishi M, et al. Hypermethylated promoter region of DR3, the death receptor 3 gene, in rheumatoid arthritis synovial cells. Arthritis and rheumatism. 2006 Mar;54(3):779-87. PubMed PMID: 16508942.
49. Klareskog L, Malmstrom V, Lundberg K, Padyukov L, Alfredsson L. Smoking, citrullination and genetic variability in the immunopathogenesis of rheumatoid arthritis. Seminars in immunology. 2011 Apr;23(2):92-8. PubMed PMID: 21376627.
50. NCBI. [updated September 28, 2000]. Available from: http://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Genetics/chromosome.html
51. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921. PubMed PMID: 11237011. Epub 2001/03/10. eng.
52. Goris A, Liston A. The immunogenetic architecture of autoimmune disease. Cold Spring Harbor perspectives in biology. 2012 Mar;4(3). PubMed PMID: 22383754.

53. Karlsen TH, Melum E, Franke A. The utility of genome-wide association studies in hepatology. Hepatology. 2010 May;51(5):1833-42. PubMed PMID: 20232293. Epub 2010/03/17. eng.
54. Syversen SW, Gaarder PI, Goll GL, Odegard S, Haavardsholm EA, Mowinckel P, et al. High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from a 10-year longitudinal study. Annals of the rheumatic diseases. 2008 Feb;67(2):212-7. PubMed PMID: 17526555.
55. Uhlig T, Kvien TK, Jensen JL, Axell T. Sicca symptoms, saliva and tear production, and disease variables in 636 patients with rheumatoid arthritis. Annals of the rheumatic diseases. 1999 Jul;58(7):415-22. PubMed PMID: 10381485. Pubmed Central PMCID: 1752918.
56. Haavardsholm EA, Boyesen P, Ostergaard M, Schildvold A, Kvien TK. Magnetic resonance imaging findings in 84 patients with early rheumatoid arthritis: bone marrow oedema predicts erosive progression. Annals of the rheumatic diseases. 2008 Jun;67(6):794-800. PubMed PMID: 17981915.
57. Halvorsen EH, Haavardsholm EA, Pollmann S, Boonen A, van der Heijde D, Kvien TK, et al. Serum IgG antibodies to peptidylarginine deiminase 4 predict radiographic progression in patients with rheumatoid arthritis treated with tumour necrosis factor-alpha blocking agents. Annals of the rheumatic diseases. 2009 Feb;68(2):249-52. PubMed PMID: 18723564.
58. Barker DL, Hansen MS, Faruqi AF, Giannola D, Irsula OR, Lasken RS, et al. Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. Genome research. 2004 May;14(5):901-7. PubMed PMID: 15123587. Pubmed Central PMCID: 479118.
59. Lee YH, Choi SJ, Ji JD, Song GG. Associations between ERAP1 polymorphisms and ankylosing spondylitis susceptibility: a meta-analysis. Inflammation research : official journal of the European Histamine Research Society [et al]. 2011 Nov;60(11):999-1003. PubMed PMID: 21877190.
60. Andres AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin SQ, Hurle B, et al. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. PLoS genetics. 2010 Oct;6(10):e1001157. PubMed PMID: 20976248. Pubmed Central PMCID: 2954825.
61. Coulombe-Huntington J, Lam KC, Dias C, Majewski J. Fine-scale variation and genetic determinants of alternative splicing across individuals. PLoS genetics. 2009 Dec;5(12):e1000766. PubMed PMID: 20011102. Pubmed Central PMCID: 2780703.
62. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souef P, et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet. 2011 Sep 10;378(9795):1006-14. PubMed PMID: 21907864. Pubmed Central PMCID: 3517659.
63. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nature genetics. 2012 May;44(5):511-6. PubMed PMID: 22446963.
64. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. Current protocols in human genetics / editorial board, Jonathan L Haines [et al]. 2009 Jan;Chapter 2:Unit 2 12. PubMed PMID: 19170031. Epub 2009/01/27. eng.
65. Hui L, DelMonte T, Ranade K. Genotyping using the TaqMan assay. Current protocols in human genetics / editorial board, Jonathan L Haines [et al]. 2008 Jan;Chapter 2:Unit 2 10. PubMed PMID: 18428424. Epub 2008/04/23. eng.
66. Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genetic analysis : biomolecular engineering. 1999 Feb;14(5-6):143-9. PubMed PMID: 10084106. Epub 1999/03/20. eng.
67. Corporation LT. 2013. Available from: http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/PCR/real-time-pcr/qpcr-education/what-can-you-do-with-qpcr/genotyping.html.
68. CIGENE. 2013. Available from: http://www.cigene.no/snp-services.html.
69. Human (Homo sapiens) Genome Browser Gateway: UCSC Genome Bioinformatics Available from: http://genome.ucsc.edu/cgi-bin/hgGateway.

70. Skaletsky SRaHJ. Primer3Web.
71. Kent J. In-Silico PCR: Genome Bioinformatics UCSC. Available from: http://genome.ucsc.edu/cgi-bin/hgPcr?command=start.
72. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263-5. PubMed PMID: 15297300.
73. PLINK v.1.07 2009. Available from: http://pngu.mgh.harvard.edu/~purcell/plink/.
74. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007 Sep;81(3):559-75. PubMed PMID: 17701901. Pubmed Central PMCID: 1950838.
75. Dupont WD, Plummer WD, Jr. Power and sample size calculations. A review and computer program. Controlled clinical trials. 1990 Apr;11(2):116-28. PubMed PMID: 2161310.
76. Padyukov L, Seielstad M, Ong RT, Ding B, Ronnelid J, Seddighzadeh M, et al. A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. Annals of the rheumatic diseases. 2011 Feb;70(2):259-65. PubMed PMID: 21156761. Pubmed Central PMCID: 3015094.
77. Kochi Y, Okada Y, Suzuki A, Ikari K, Terao C, Takahashi A, et al. A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nature genetics. 2010 Jun;42(6):515-9. PubMed PMID: 20453841.
78. Radtke F, Fasnacht N, Macdonald HR. Notch signaling in the immune system. Immunity. 2010 Jan 29;32(1):14-27. PubMed PMID: 20152168.
79. Chen G, Courey AJ. Groucho/TLE family proteins and transcriptional repression. Gene. 2000 May 16;249(1-2):1-16. PubMed PMID: 10831834.
80. Mero IL, Lorentzen AR, Ban M, Smestad C, Celius EG, Aarseth JH, et al. A rare variant of the TYK2 gene is confirmed to be associated with multiple sclerosis. European journal of human genetics : EJHG. 2010 Apr;18(4):502-4. PubMed PMID: 19888296. Pubmed Central PMCID: 2987240.
81. Hess J, Angel P, Schorpp-Kistner M. AP-1 subunits: quarrel and harmony among siblings. Journal of cell science. 2004 Dec 1;117(Pt 25):5965-73. PubMed PMID: 15564374.
82. Gourh P, Agarwal SK, Martin E, Divecha D, Rueda B, Bunting H, et al. Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. Journal of autoimmunity. 2010 Mar;34(2):155-62. PubMed PMID: 19796918. Pubmed Central PMCID: 2821978.
83. Zhou XJ, Lu XL, Nath SK, Lv JC, Zhu SN, Yang HZ, et al. Gene-gene interaction of BLK, TNFSF4, TRAF1, TNFAIP3, and REL in systemic lupus erythematosus. Arthritis and rheumatism. 2012 Jan;64(1):222-31. PubMed PMID: 21905002.
84. Barton A, Eyre S, Ke X, Hinks A, Bowes J, Flynn E, et al. Identification of AF4/FMR2 family, member 3 (AFF3) as a novel rheumatoid arthritis susceptibility locus and confirmation of two further pan-autoimmune susceptibility genes. Human molecular genetics. 2009 Jul 1;18(13):2518-22. PubMed PMID: 19359276. Pubmed Central PMCID: 2694689.
85. Nordang GB, Viken MK, Amundsen SS, Sanchez ES, Flato B, Forre OT, et al. Interferon regulatory factor 5 gene polymorphism confers risk to several rheumatic diseases and correlates with expression of alternative thymic transcripts. Rheumatology. 2012 Apr;51(4):619-26. PubMed PMID: 22179739.
86. Pedersen M, Jacobsen S, Klarlund M, Pedersen BV, Wiik A, Wohlfahrt J, et al. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. Arthritis research & therapy. 2006;8(4):R133. PubMed PMID: 16872514. Pubmed Central PMCID: 1779386.
87. de Pablo P, Dietrich T, Chapple IL, Milward M, Chowdhury M, Charles PJ, et al. The autoantibody repertoire in periodontitis: a role in the induction of autoimmunity to citrullinated proteins in rheumatoid arthritis? Annals of the rheumatic diseases. 2013 Mar 16. PubMed PMID: 23434568.
88. Karlson EW, Mandl LA, Aweh GN, Grodstein F. Coffee consumption and risk of rheumatoid arthritis. Arthritis and rheumatism. 2003 Nov;48(11):3055-60. PubMed PMID: 14613266.

89. Arkema EV, Hart JE, Bertrand KA, Laden F, Grodstein F, Rosner BA, et al. Exposure to ultraviolet-B and risk of developing rheumatoid arthritis among women in the Nurses' Health Study. Annals of the rheumatic diseases. 2013 Apr;72(4):506-11. PubMed PMID: 23380431.
90. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747-53. PubMed PMID: 19812666. Pubmed Central PMCID: 2831613.
91. Kilpinen H, Barrett JC. How next-generation sequencing is transforming complex disease genetics. Trends in genetics : TIG. 2013 Jan;29(1):23-30. PubMed PMID: 23103023.

Appendix

Table A1. Criteria produced by the American Rheumatism Association. Rheumatoid arthritis is defined by the presence of 4 or more criteria, and no further qualifications (classic, definite, or probable) or list of exclusions are required. (16)

Morning stiffness in and around joints lasting at least 1 hour before maximal improvement*
Soft tissue swelling (arthritis) of 3 or more joint areas observed by a physician*
Swelling (arthritis) of the proximal interphalangeal, metacarpophalangeal, or wrist joints*
Symmetric swelling (arthritis)*
Rheumatoid nodules
The presence of rheumatoid factor
Radiographic erosions and/or periarticular osteopenia in hand and/or wrist joints

*Must have been present for at least 6 weeks.

Table A2. Conditions for PCR with temperature gradient, for optimisation of amplification of DNA sequence for sequencing

Temperature	Minutes	
95°C	5min	
95°C	30sec	
58°C	30sec	} x30
72°C	1min30sec	
72°C	7min	
4°C	∞	

Table A3. Sequencing reactionmix for optimaliastion of MgCl2 concentration

	3mM MgCl	2mM MgCl	1,5mM MgCl	1mM MgCl	0,5mM MgCl	0,0mM MgCl	Bioline 3mM MgCl	4mM MgCl
No. of Wells:	Mix for 8	8	8	8	8	8	8	8
Sterile Water	65,8	69,2	71,3	72,9	74,6	76,7	63,8	62,1
10X PCR Buffer (no MgCl2)	9,2	9,2	9,2	9,2	9,2	9,2	11,7	9,2
MgCl2 25mM	11,0	7,3	5,5	3,7	1,8	0,0	11,0	14,7
dNTP 20 mM	2,5	2,5	2,5	2,5	2,5	2,5	1,7	2,5
Fwd Primer*	0,8	0,8	0,8	0,8	0,8	0,8	0,9	0,8
Rev Primer*	0,8	0,8	0,8	0,8	0,8	0,8	0,9	0,8
Taq Start	0,0	0,0	0,0	0,0	0,0	0,0		0,0
Taq Polymerase 5U/uL	0,6	0,6	0,6	0,6	0,6	0,6	0,9	0,6
Total Volume	91,0	90,7	90,9	90,8	90,6	90,8	90,8	90,9
* 20 pmol/uL								
10 uL mix + 1ul repliG (1+19)							11	

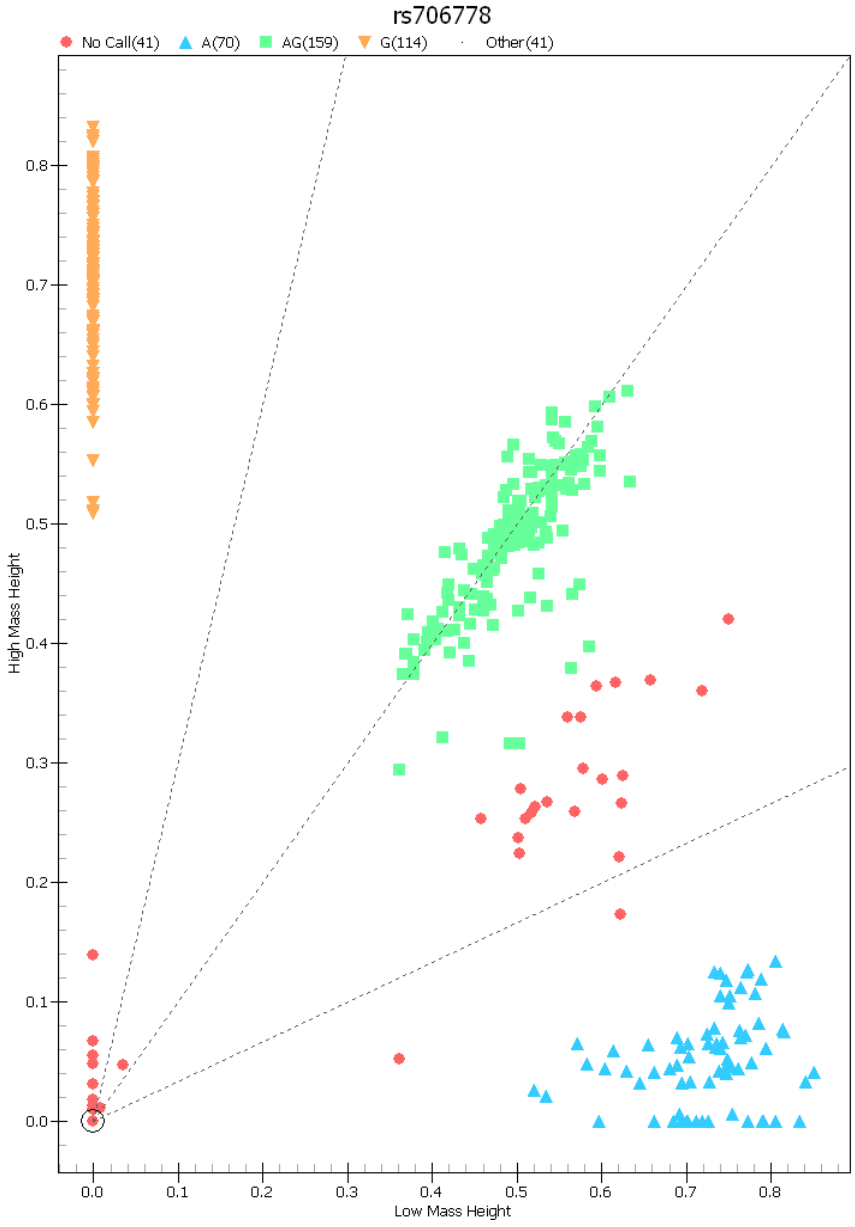


Figure A2. Genotyping result, rs706778. Several individuals clustering between AG and AA.

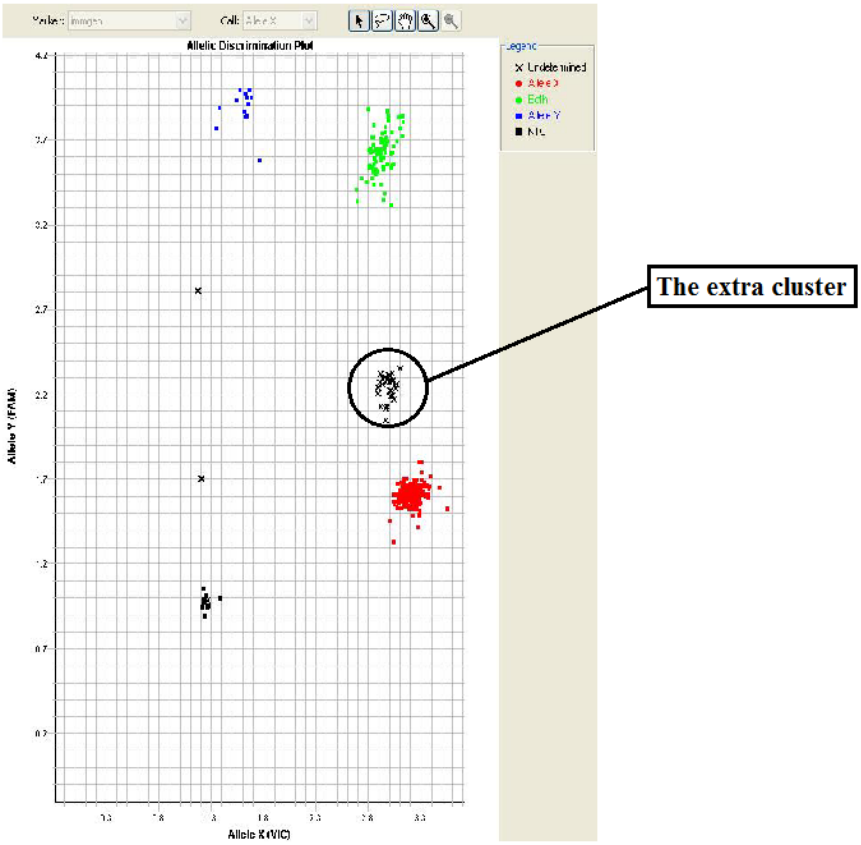


Figure A2. Genotyping result, rs13119723

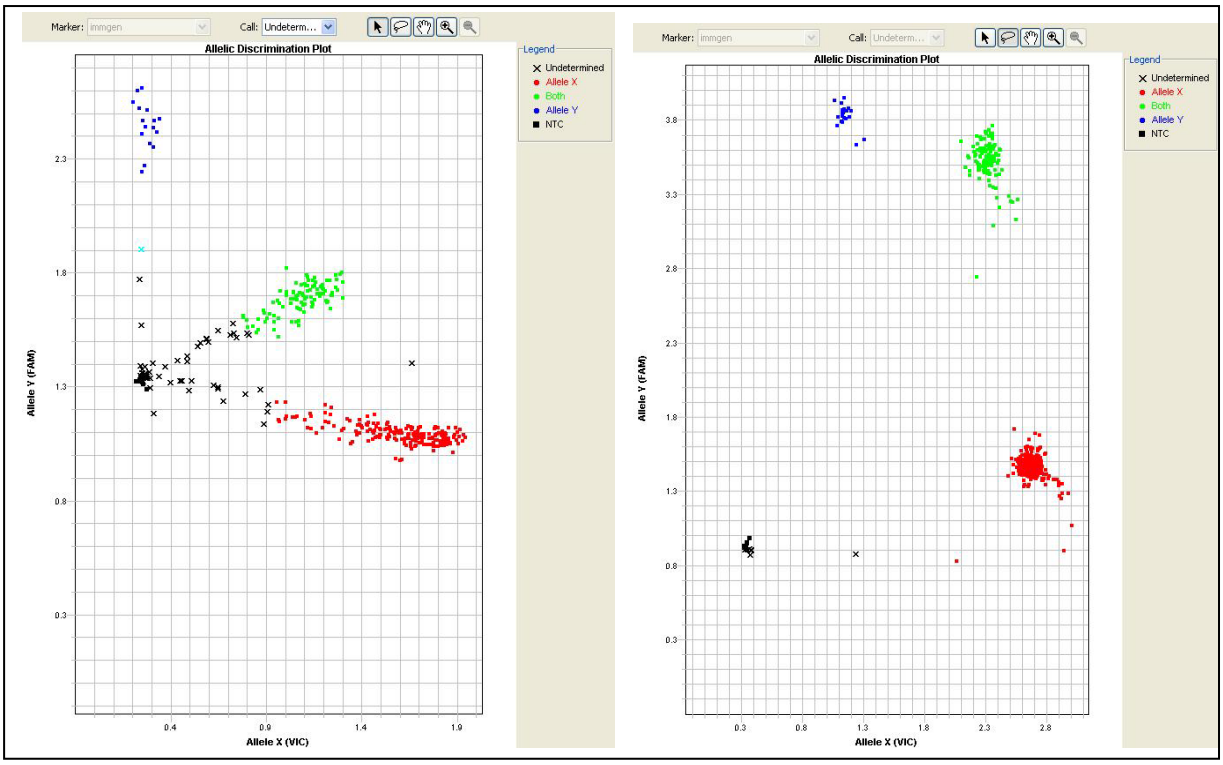


Figure A3. TaqMan genotyping resoult for rs 7155603, genotyping the same individuals with different mixes; TaqMan mix to the left and ABgene mix to the right, respectively.

The questionnaire (translated to English)

9446436354

Oslo universitetssykehus

Code | | | | | | |

Questionnaire of environmental factors

Dear participant.

This questionnaire contains questions about your family history, education and work, diet and other lifestyle factors, and diseases you have or have had. See attached information letter for supplementary information.

About filling in the questionnaire:
The questionnaire is to be read optically. Please use blue or black ballpoint pen. If you mistype, cross it out with a straigth line and check off the correct box. We ask you to check off in the middle of the squares and write with CAPITAL letters in the squares like shown below. If anything is written outside the marked fields, it will not be registered. If you have comments or feedbacks, we ask you to use the commentary area on the last page.

Like this: ☒ | O | S | L | O | |

Not like this: ☒ | O | S L | | O |

Thanks in advance for your valuable contribution!

If you do not wish to answer the questionnaire, check off in the squares below and return the questionnaire in the attached response envelope. Then you´ll avoid receiving duns!

☐ I don´t wish to answer the questionnaire

☐ Woman ☐ Man Age: | | |

1

2590436357

Background

1. Sex: ☐ Woman ☐ Man

2. Month of birth:
(01 for jan, 02 for feb etc.) ☐☐ Year of birth: ☐☐☐☐

3. Height and weight:

How tall are you? ☐☐☐ cm

How much do you weigh? ☐☐☐ kg

Approximately how tall were you when you were 18? ☐☐☐ cm ☐ Don´t remember

Approximately how much did you weigh when you were 18? ☐☐☐ kg ☐ Don´t remember

4. Your country of birth: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

If you were not born in Norway, in what year did you move here? ☐☐☐☐

5. Country of birth of your biological parents and grandparents:

Mother: ☐☐☐☐☐☐☐☐☐☐☐☐

Maternal grandmother: ☐☐☐☐☐☐☐☐☐☐☐☐

Maternal grandfather: ☐☐☐☐☐☐☐☐☐☐☐☐

Father: ☐☐☐☐☐☐☐☐☐☐☐☐

Paternal grandmother: ☐☐☐☐☐☐☐☐☐☐☐☐

Paternal grandfather: ☐☐☐☐☐☐☐☐☐☐☐☐

Diseases or surgical treatment

6. Have you, or have you ever had, any of these diseases/illnesses:
(Check once per line)

	Yes	No	If yes, how old were you the first time?
Heart attack	☐	☐	☐☐ Years
Angina pectoris	☐	☐	☐☐ Years
Heart failure	☐	☐	☐☐ Years
Other heart disease	☐	☐	☐☐ Years
Cerebral infarction/intracranial hemorrhage	☐	☐	☐☐ Years
Kidney disease	☐	☐	☐☐ Years
Asthma	☐	☐	☐☐ Years
Chronic bronchitis, emphysema, COPD	☐	☐	☐☐ Years
Diabetes type 1	☐	☐	☐☐ Years
Diabetes type 2	☐	☐	☐☐ Years
Psoriasis	☐	☐	☐☐ Years
Rheumatoid arthritis	☐	☐	☐☐ Years
Bechterews disease	☐	☐	☐☐ Years
Sarcoidosis	☐	☐	☐☐ Years
Osteoporosis	☐	☐	☐☐ Years
Osteoarthritis	☐	☐	☐☐ Years
Myasthenia Gravis	☐	☐	☐☐ Years
MS (multiple sclerosis)	☐	☐	☐☐ Years
PSC (primary sclerosing cholangitis)	☐	☐	☐☐ Years

Continuation question 6:

	Yes	No	If yes, how old were you the first time?
Coeliac disease	☐	☐	☐☐ Years
Inflammatory bowel disease (Ulcerative colitis, Crohns disease)	☐	☐	☐☐ Years
SLE (systemic lupus erythematosus)	☐	☐	☐☐ Years
Sjøgren´s disease	☐	☐	☐☐ Years
Hypothyreosis	☐	☐	☐☐ Years
Hyperthyreosis	☐	☐	☐☐ Years
Tonsillectomy	☐	☐	☐☐ Years
Appendectomy	☐	☐	☐☐ Years
Migraine	☐	☐	☐☐ Years

Infectious diseases

7. Have you ever had mononucleosis? ☐ Yes ☐ No ☐ Don´t know

If yes, how old were you when you had the disease? ☐☐ Years

8. Have you had any other infectious disease(s) before the age of 18 that demanded hospitalization? ☐ Yes ☐ No ☐ Don´t know

If yes, what kind of infection did you have?

☐ Pneumonia

☐ Kidney/urinary tract infection

☐ Stomach/bowel infection

☐ Infection of the brain/meningitis

☐ Other infection

Dental health

9. Have you ever had infections of the teeth (root infection)? ☐ Yes ☐ No

If yes, in what year(s)? ☐☐☐☐ ☐☐☐☐

10. Have you ever had infecions of the gums (periodontitis)? ☐ Yes ☐ No

If yes, in what year(s)? ☐☐☐☐ ☐☐☐☐

Vaccinations

11. Did you follow the normal vaccination program as a child?

☐ Yes ☐ No ☐ Partly/interrupted ☐ Don´t know

Pets during childhood

12. Did you have pets during your childhood? ☐ Yes ☐ No

If yes, what kind(s) of animal(s) did you have?

☐ Cat ☐ Dog ☐ Horse ☐ Other: ☐☐☐☐☐☐☐☐☐☐ ☐☐☐☐☐☐☐☐☐☐ ☐☐☐☐☐☐☐☐☐☐

Nail polish

13. How many times have you been using nail polish the last year?

☐ Never ☐ 1--10 ☐ 11--20 ☐ 21--30 ☐ 31--40 ☐ 41--50 ☐ More than 50 times

Tobacco

14. Do you smoke?

☐ Yes, cigarettes sometimes (parties/holidays, not daily)

☐ Yes, cigarettes daily

☐ Yes, cigars/cigarillos/pipe sometimes

☐ Yes, cigars/cigarillos/pipe daily

☐ No, I have stopped smoking

☐ No, I have never smoked

If you've never smoked, go directly to question 17.

15. Answer this question if you now smoke daily, or earlier have smoked daily:

How many cigarettes do or did you normally smoke daily? _____ ☐☐ Cigarettes pr. day

How old were you when you started smoking daily? ☐☐ Years

If you have smoked daily earlier, how old were you when you quitted? ☐☐ Years

16. Answer this question if you smoke or have smoked sometimes, but not daily:

How many cigarettes do or did you normally smoke per month? ☐☐☐ Cigarettes pr. month

How old were you when you started smoking sometimes? ☐☐ Years

If you have smoked sometimes earlier, how old were you when you quitted? ☐☐ Years

17. Have you ever lived with one person or more who smoked daily in your surrondings? (Inside the house etc)

☐ Yes ☐ No

If yes, in what time period? (f.ex. from 1980 through 1985)

From: ☐☐☐☐ — Through: ☐☐☐☐

From: ☐☐☐☐ — Through: ☐☐☐☐

From: ☐☐☐☐ — Through: ☐☐☐☐

From ☐☐☐☐ — Through: ☐☐☐☐

18. Do you use, or have you ever used, snuff?

☐ No, never

☐ I have used snuff in the past, but I have stopped using snuff.

☐ Yes, sometimes

☐ Yes, daily

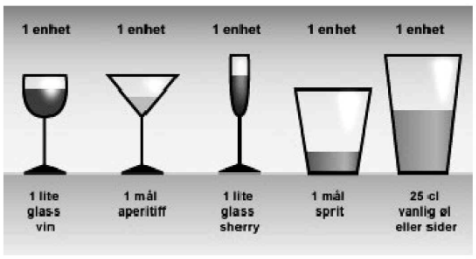
If you have never used snuff, go directly to question 20.

19. If you use/have ever used snuff:

How old were you when you started using snuff? ☐☐ Years

How many boxes of snuff do/did you use per month? ☐☐ Number of boxes

Alcohol



This figure demonstrate what one unit of alcohol signifies.

20. How many units of alcohol do you drink per day? (Think about how much you drink during 4 weeks, and divide this amount in number of days)

☐ I don't drink alcohol

☐ 0--1 units

☐ 1--3 units

☐ More than 3 units

21. How many units of alcohol did you drink per ~~day when~~ you were 18 Years? (Think consumption over 4 weeks and divide with number of days)

☐ Didn't drink alcohol

☐ 0--1 units

☐ 1--3 units

☐ More than 3 units

Coffee/tea

22. How many cups of coffee/tea do you drink per day?
(Answer 0 if you don't drink coffee/tea)

Coffee: ☐☐ Tea:: ☐☐

23. How many cups of coffee/tea did you drink per day when you were 18 Years?
(Answer 0 if you didn't drink coffee/tea)

Coffee: ☐☐ Tea: ☐☐

Eating habits/supplements

24. How often do you eat the following types of food nowadays?

	0--3 times per month	1--3 times per week	4--6 times per week	1 time per day	2 times or more per day
Fruit/berries	☐	☐	☐	☐	☐
Vegetables	☐	☐	☐	☐	☐
Red meat (cattle/sheep/pig)	☐	☐	☐	☐	☐
Fat fisk (salmon,trout,herring, mackerel, redfish)	☐	☐	☐	☐	☐
Lean fish (cod,pollock,haddock,vitting)	☐	☐	☐	☐	☐

25. How often did you normally eat the following types of food in your childhood up to the age of 18?

	0--3 times per month	1--3 times per week	4--6 times per week	1 time per day	2 times or more per day
Fruit/berries	☐	☐	☐	☐	☐
Vegetables	☐	☐	☐	☐	☐
Red meat (storfe, får, svin)	☐	☐	☐	☐	☐
Fat fish (salmon,trout,herring, mackerel, redfish)	☐	☐	☐	☐	☐
Lean fish (cod,pollock,haddock,vitting)	☐	☐	☐	☐	☐

26. Do you use the following supplements?

	Yes, daily	Sometimes	No
Cod oil	☐	☐	☐
Omega 3 capsules	☐	☐	☐
Vitamine or mineral supplements	☐	☐	☐

27. Did you use the following supplements when you were 18 Years?

	Yes, daily	Sometimes	No
Cod oil	☐	☐	☐
Omega 3 capsules	☐	☐	☐
Vitamine or mineral supplements	☐	☐	☐

Sun bathing habits

28. How many weeks per year have you been taking sunbaths/ doing activities in the sun in areas with very strong sun radiation (such as in Africa or in the Southern parts of Europe)? (Check once per line.)

	Never	1--2 weeks	3--5 weeks	6 weeks or more
During the last 12 months	☐	☐	☐	☐
Before the age of 10	☐	☐	☐	☐
Between the ages of 10 and 19	☐	☐	☐	☐
20 years or older	☐	☐	☐	☐

29. How often have you been sunbathing or doing activity in the sun in Scandinavia during spring/summer/fall? (Check once per line)

	Never	A few hours per month	A few hours per week	A few hours per day
During the last 12 months	☐	☐	☐	☐
Before the age of 10	☐	☐	☐	☐
Between the ages of 10 and 19	☐	☐	☐	☐
20 years or older	☐	☐	☐	☐

30. How often have you sunbathing in a solarium? (Check once per line)

	Never	A few times per year	A few times per month	Several times per month
During the last 12 months	☐	☐	☐	☐
Before the age of 10	☐	☐	☐	☐
Between the ages of 10 and 19	☐	☐	☐	☐
20 years or older	☐	☐	☐	☐

Education and working situation

31. What is the highest education you have completed? (Check one)

- ☐ Less than 7 years of primary school
- ☐ Primary 7-10 years, the continuation school, elementary school
- ☐ High school, lower secondary school, vocational school (10-13 years)
- ☐ College / university 4 years or less
- ☐ College / university 4 years or more

32. What kind of work/job do you have now? (Check one or more boxes)

- ☐ Employed in paid work
- ☐ Selfemployed
- ☐ Fulltime housework
- ☐ Education, military service
- ☐ Unemployed
- ☐ Retired

33. What is your current or most recent job? (Check one or more boxes. Also respond if you are unemployed or retired)

- ☐ Administrative director, politician, director
- ☐ Academic profession (engineer, doctor, lawyer, economist, lecturer)
- ☐ Occupation with shorter college / university education and technicians (nursing, engineering, police, physiotherapi, optometrist)
- ☐ Office and customer service (secretary, finance staff, receptionist)
- ☐ Sales, service and care work (customer service representative, "helping nurse", hairdresser)
- ☐ Occupation in the agriculture, forestry and fishing
- ☐ Home Entertainment Systems/Artisan? (mechanic, electrician, carpenter, painter, baker, butcher)
- ☐ Process and machine operators, transport workers, etc..
- ☐ Military occupation (private, officer)
- ☐ Other occupation/job:

34. Have you, on your current or previous occupation, been significantly exposed over time for any of the following?

	Yes	No	If yes, how old were you when the exposure started	If yes, how many years have you been exposed?
Motor oil	☐	☐	☐☐ Years	☐☐ Years of exposure
Cutting oil/ Skjæreolje	☐	☐	☐☐ Years	☐☐ Years of exposure
Oil/ Formolje	☐	☐	☐☐ Years	☐☐ Years of exposure
Hydraulic oil	☐	☐	☐☐ Years	☐☐ Years of exposure
Turbine oil	☐	☐	☐☐ Years	☐☐ Years of exposure
Asfalpht (asfalt)	☐	☐	☐☐ Years	☐☐ Years of exposure
Oil/ Råolje	☐	☐	☐☐ Years	☐☐ Years of exposure
Anesthetic gases	☐	☐	☐☐ Years	☐☐ Years of exposure
Organic solvents*	☐	☐	☐☐ Years	☐☐ Years of exposure

*Organic solvents such as degreasers, trichlorethylene, white spirit, thinner, toluene, styrene, xylene etc. (Med organiske løsemidler menes f.eks. avfettingsmidler, trikloroetylen, white spirit, tynnere, toluen, styren, xylen e.l.)

Women

35. How old were you when you had your period for the first time? ☐☐ Years

36. Has your menstruation ceased? ☐ Yes ☐ No

If yes, at what age? ☐☐ Years

37. Have you used birth control pills (not mini pills) , vaginal ring or patch? ☐ Yes ☐ No

If yes, approximately for how long? (Years) ☐☐ Years

38. Have you ever used another form of hormonal contraceptive? For instance mini pill, birth control shot or IUS (not copper IUD) ☐ Yes ☐ No

If yes, approximately for how long? (Years) ☐☐ Years

39. Have you had a pregnancy that ended in a miscarriage or an abortion? ☐ Yes ☐ No

If yes, how many? ☐☐

40. Have you had hormonal therapy for infertility? ☐ Yes ☐ No

41. Have you ever been pregnant? ☐ Yes ☐ No

☐ I'm currently pregnant

If yes, fill in, for each child you have born, the following information about birth year and how many months you breast-fed. (Also to be filled in for stillborn or children who died later in life).

Child	Year of birth	Number of months with breastfeeding	Child	Year of birth	Number of months with breastfeeding
1	☐☐☐☐	☐☐	5	☐☐☐☐	☐☐
2	☐☐☐☐	☐☐	6	☐☐☐☐	☐☐
3	☐☐☐☐	☐☐	7	☐☐☐☐	☐☐
4	☐☐☐☐	☐☐	8	☐☐☐☐	☐☐

Thank you for your participation

Comments and feedback: