Norges miljø- og biovitenskapelige universitet
Fakultet for veterinærmedisin og biovitenskap
Institutt for kjemi, bioteknologi og matvitenskap

Masteroppgave 2015
30 studiepoeng

Studier av AIRE genet med hensyn til disposisjon for autoimmune sykdommer

Studies of the involvement of the AIRE gene in predisposition to autoimmune diseases

Marthe Flatø Vestby

Forord

Denne masteroppgaven er utført ved Oslo universitetssykehus, Ullevål, ved avd. for medisinsk genetikk i perioden januar 2015 til mai 2015. Oppgaven er den avsluttende delen av master i Teknologi (sivilingeniør) – Kjemi og bioteknologi ved Norges miljø- og biovitenskapelige universitet.

Jeg vil først og fremst takke min hovedveileder Benedicte A. Lie, for at jeg fikk muligheten til å skrive denne oppgaven og ble inkludert i deres forskningsgruppe. Tusen takk for all den gode veiledningen, alle raske tilbakemeldinger og din positive innstilling til alt, uansett hva. Det har vært noen svært spennende og lærerike måneder hos dere.

Vil også takke min veileder Ingvild Gabrielsen for all hjelp og gode råd under denne prosessen, din tålmodighet og din evne til å alltid kunne forklare ting på en forståelig måte. Siri Tennebø Flåm, tusen takk for all hjelp på laboratoriet, for at du har delt kontoret ditt, at du har vært tilgjengelig for å hjelpe og for ditt gode humør. Det har betydd mye. Fatemeh Kaveh, takk for all den gode hjelpen med analyseringen av RNA sekvenseringsdataene og for at du har delt din faglige kompetanse med meg. Kari Guderud, takk for at du har delt av dine gode datakunnskaper og triks slik at både mine beregninger og analyseringer har gått litt lettere. Jeg vil rette en stor takk til alle på avdelingen for medisinsk genetikk for at jeg har blitt tatt så godt i mot og for all god hjelp.

Takk til min interne veileder på Ås, Tor Lea, for raske tilbakemeldinger og hjelp med det praktiske rundt å skrive en masteroppgave.

Til slutt vil jeg takke min familie og min samboer for uvurderlig støtte gjennom flere år med skolegang, oppmuntring og at dere alltid har stilt opp for meg. Uten dere hadde jeg ikke klart dette.

Ås, 15.05.2015

Marthe Flatø Vestby

Sammendrag

Revmatoid artritt (RA) er en autoimmun, kronisk leddsykdom hvor immunsystemet angriper små ledd (som f.eks. fingre og tær) i kroppen, som fører til leddbetennelser og ødeleggelse av bein og brusk. Sykdommen er mest frekvent i kvinner mellom 45-60 år og reduserer både livskvalitet og livslengde. Det autoimmune angrepet involverer T celler. Tymus er et lymfatisk organ bak brystbenet hos mennesker og er hovedorgan i immunsystemet hvor modning og utvikling av T-celler finner sted. Autoimmun regulator (AIRE) er en transkripsjonsfaktor som uttrykkes i epitelcellene i medulla av tymus hvor den regulerer uttrykket av selv-peptidene som presenteres for T-lymfocyttene. I dag vet man at ~60 % av sykdomsetiologien til RA skyldes genetiske faktorer og 101 RA risiko loci er identifisert. På tross av dette er det fortsatt en stor del av etiologien og patogenesen til sykdommen som er ukjent, og i denne oppgaven har vi forsøkt å finne ut mer om den genetiske bakgrunnen til sykdommen i den norske befolkningen ved å undersøke ulike polymorfismer i genet *AIRE* for assosiasjon til RA. *AIRE* polymorfismer har tidligere vært funnet assosiert i Japan, men ikke i europeiske populasjoner. Vi brukte et prøvemateriale bestående av 944 pasienter og 1098 kontroller for genotyping av seks selekterte enkle nukleotid polymorfismer og utførte bioinformatiske- og statistiske analyser for å se nærmere på assosiasjon mellom de enkle nukleotid polymorfismene og RA. Vi brukte også datamateriale fra 42 tymusprøver for å se på de enkle nukleotid polymorfismenes regulatoriske evne ved å assosiere disse med uttrykket av *AIRE* i tymus. I tillegg ble RNA sekvenseringsdata brukt for å se etter nye transkripter av genet i tymus. I denne studien ble to polymorfismer funnet signifikant assosiert med utvikling av RA i den norske leddgikts populasjonen, og disse ble hovedsakelig funnet til RA pasienter positive for antistoffet ACPA. Det ble ikke funnet noen assosiasjon mellom *AIRE* polymorfismene og nivået av *AIRE* i tymus, men det ble funnet evidens for flere transkripter av *AIRE* enn det som allerede er rapporterte i tymus. Resultatene fra oppgaven kan tyde på at det er assosiasjon mellom RA og *AIRE* regionen også i europeisk befolkning og at repertoaret av transkripter fra *AIRE* genet bør kartlegges nærmere i tymus.

Abstract

Rheumatoid arthritis (RA) is an autoimmune, chronic joint-disease where the immune system attacks small joints like fingers and toes in the body, which leads to inflammation in the joints and destruction of bones and cartilage. The disease is most frequent in women between 45-60 years and reduces both life quality and life length. The autoimmune attack involves T cells. Thymus is in human a lymphatic organ behind the chest bone and a principal organ in the immune system where maturation and education of T-cells takes place. Autoimmune regulator (AIRE) is a transcription factor expressed in epithelial cells in the medulla of the thymus, where it regulates the expression of self-peptides presented to the T-lymphocytes. Today we know that ~60 % of RA disease etiology is due to genetic factors and 101 RA risk loci have been identified. Despite the large number of risk variants, there is still a big part of the disease etiology and pathogenesis which is unknown, and in this thesis we have tried to uncover more of the genetic background of the disease in the Norwegian population by examining different polymorphisms in the *AIRE* gene for association with RA. *AIRE* polymorphisms have previously been found associated in Japan, but not in the European population. We used different cohorts, 944 patients and 1098 controls for the genotyping of six selected single nucleotide polymorphisms and performed bioinformatic- and statistical analysis to investigate the association between the single nucleotide polymorphisms and RA. We also used data material from 42 thymus samples to look at the regulatory ability of the single nucleotide polymorphisms by associating them with the expression of *AIRE* in the thymus. In addition, RNA sequencing data was used to look for new transcripts of the gene in thymus. In this thesis, two polymorphisms were found associated with the development of RA in the Norwegian arthritis population, and these were primarily found in RA patients positive for anti-citrullinated protein antibodies (ACPA). No association was found between *AIRE* polymorphisms and the expression level of *AIRE* in thymus, but there was evidence of several new transcripts of *AIRE* that has not previously been reported in thymus. The results from the thesis may indicate that an association between RA and the *AIRE* region is also present in the European population and that the repertoire of transcripts from the *AIRE* gene should be mapped in more detail in the thymus.

Forkortelser

ACPA	Anti-citrullinerte protein antistoff
AIRE	Autoimmun regulator protein
APECED	Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy
APS	Autoimmune polyglandular syndrome
Bp	Basepar
CI	Konfidensintervall
cTEC	Kortikale tymiske epitelceller
Df	Frihetsgrader
DNA	Deoksyribonukleinsyre
eQTL	Expression quantitative trait locus
GSR	Genotypesuksessrate
GWAS	Genome Wide Association Study
HGP	Human Genome Project
HLA	Humant leukocytt antigen
HWE	Hardy Weinberg likevekt
IBD	Inflammatorisk tarmsykdom
LD	Koblingsulikevekt
MAF	Minor allel frekvens
MHC	Major histocompability complex

mTEC	Medullære tymiske epitelceller
NK-celler	Natural killer-celler
NSC	Norwegian Sequencing Center
OR	Odds ratio
PCR	Polymerase kjede reaksjon
RA	Revmatoid artritt
RF	Revmatoid faktor
RNA	Ribonukleinsyre
SE	Shared epitope
SLE	Systemisk lupus erythematosus
SNP	Enkel nukleotid polymorfisme
T1D	Type 1 Diabetes
TCR	T-celle reseptor
TE	Tris-EDTA
TNF	Tumor nekrose faktor
UC	Ulcerøs kolitt
WGA	Hel-genom amplifisering

Innhold

1. Introduksjon

1.1 Autoimmune sykdommer

Immunsystemet er et nettverk av celler i kroppen som samarbeider for å forsvare deg mot mikrober og infeksjon. Systemet består av to deler, det ene er et medfødt system som er førstelinjen i immunforsvaret hvor de fleste mikroorganismer blir uskadeliggjort vha. blant annet fagocytose og «natural killer» – celler (NK-celler) gjennom uspesifikk gjenkjennelse. Det andre er et dynamisk system som kontinuerlig identifiserer nye fremmedlegemer og som utvikler et bredere forsvar ved å lage spesifikke antistoffer. B celler i lymfevevet produserer antistoff. Antistoffenes hovedfunksjon er å beskytte kroppen mot fremmedlegemer som bakterier og virus. Bakteriene/virusene blir identifisert av T celler når de kommer inn i kroppen fordi de består av fremmede peptider (antigen) og B celler starter å produsere spesifikke antistoffer som angriper disse. T cellene og B cellene samarbeider om å tilintetgjøre mikroben. Hvis viruset/bakterien kommer tilbake inn i kroppen ved en senere anledning «husker» immuncellene at de har møtt dette før og vil gå til et angrep med det samme. Ved autoimmune sykdommer reagerer immunapparatet feilaktig mot en bestanddel av kroppens egne friske vev/celler og kan produsere autoantistoff. Hvilke egne proteiner i kroppen som setter i gang den autoimmune responsen er som regel ikke kjent. (1)

Autoimmune sykdommer rammer ca. 5 % av Norges befolkning. En autoimmun prosess fører til betennelser og ofte svekket organ funksjon. Det finnes så mange som 80 ulike klassifiserte autoimmune sykdommer og mange av disse deler de samme symptomene, noe som gjør de vanskelige å diagnostisere. Autoimmune sykdommer rammer vanligvis oftere kvinner.

De autoimmune sykdommene kan deles inn i to hovedgrupper, en organspesifikk og en ikke-organspesifikk gruppe. Organspesifikke autoimmune sykdommer rammer et spesifikt organ i kroppen, et eksempel på dette er type 1 diabetes (T1D) hvor bukspyttkjertelen angripes og de insulin-produserende β-cellene ødelegges. Ikke-organspesifikke autoimmune sykdommer rammer kroppen systemisk og fører ofte til betennelse i ledd, muskler eller organer. Revmatoid

artritt (RA) er et eksempel på dette, hvor pasienten får vonde og hovne ledd, men andre organer kan også rammes. (1-3)

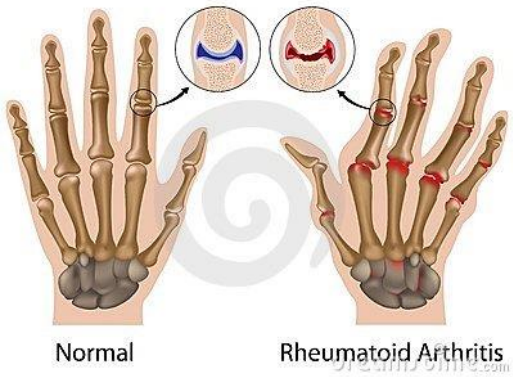
Årsaken til hvordan og hvorfor autoimmune sykdommer oppstår er fortsatt ikke helt kjent, men man har sett at slike sykdommer har en tendens til å akkumulere innad i familier, noe som tyder på at en arvelig komponent spiller inn. Det er et avvikende, mer komplekst arvemønster i autoimmune sykdommer i forhold til det man ser ved monogene sykdommer. En populasjonsstudie utført av Hemminki et. al. (4) undersøkte den familiære fordelingen av immunrelaterte sykdommer i pasienter som ble innlagt på sykehus for sin første autoimmune sykdom. Funnene indikerte at det er en økt relativ risiko for å utvikle en hvilken som helst autoimmun sykdom for barn av foreldre med en autoimmun sykdom. Det er også funnet at pasienter med autoimmune sykdommer selv har økt risiko for å utvikle andre autoimmune sykdommer i tillegg.

En hypotese for hvorfor en autoimmun sykdom trigges går ut på at det skjer en molekylær mimikk gjennom at visse mikrobielle antigener kan forårsake kryss-aktivering av autoreaktive T- eller B celler, på grunn av sekvenslikheter mellom mikrobielle antigener og selv-antigen. Autoimmune sykdommer ser ut til å være vanligere hos eldre enn hos yngre mennesker og det er en teori om at dette har en sammenheng med at immunsystemet svekkes med alderen. (1, 5)

Det finnes ennå ingen kur for autoimmune sykdommer, men pasienter blir behandlet med medisiner for å prøve og holde sykdommen under kontroll og med mål om å bedre livskvaliteten deres. Ikke-organspesifikke autoimmune sykdommer som fører til betennelsesreaksjoner i spesifikke områder i kroppen som for eksempel hos RA pasienter, kan få betennelsesdempende medisiner for og forebygge sine plager. Det er en utfordring for behandlere av autoimmune sykdommer å finne en behandlingsmetode som gir en balansegang mellom det å svekke sykdommen mest mulig uten å svekke det normale immunsystemet for mye. (6)

1.2 Revmatoid artritt

Revmatoid artritt er en kronisk leddsykdom hvor mye av patogenesen, og det som utløser selve sykdommen fortsatt er ukjent. Sykdommen er karakterisert ved vonde, stive og hovne ledd derav leddbetennelser (artritt), gir nedsatt livskvalitet og redusert livslengde. Uten behandling kan leddbetennelsene føre til ødeleggelse av leddene. Fingre og tær (perifere ledd) er de mest utsatte områdene sammen med leddnært vev, når disse blir angrepet kan det over tid føre til ødeleggelse av både beinvev og brusk (figur 1). (7)



Figur 1: Illustrasjon av en hånd hos en frisk person og en hånd hos en person med RA. (8)

Revmatoid faktor (RF) og anti-citrullinerte protein antistoff (ACPA) er begge autoantistoff som er til stede hos over halvparten av RA pasienter. RF er rettet mot Fc-fragmentet til IgG (og danner uoppløselige immunkomplekser) og er til stede i 50-80 % av pasienter med RA og bare 5-10 % hos friske personer.(9) ACPA er antistoffer som er rettet mot citrullinerte proteiner. Proteiner kan gjennomgå post-translasjonelle modifikasjoner og ved citrullinering blir aminosyren arginin konvertert til citrullin. Omkring 60 % av alle RA pasienter er positive for ACPA, mens mindre enn 2 % av den vanlige befolkningen er positive for det. (10) Gruppering av pasienter med hensyn på tilstedeværelsen av antistoff er brukt fordi antistoffene er gode diagnostiske markører for RA, da de dukker opp i kroppen flere år før degradering av ben og brusk starter, noe som også støtter en mulig patogen rolle i RA utvikling.

RA er en heterogen sykdom som kan deles inn i flere undergrupper. En klinisk viktig oppdeling av sykdommen er definert ved tilstedeværelsen eller fraværet av ACPA, kalt hhv. ACPA+ eller ACPA-. Disse to undergruppene skiller seg fra hverandre ved at ACPA+ RA har en dårligere prognose med høyere forekomst av erosiv leddskade, hvor det er viktig å starte med en kraftig behandling tidligst mulig. ACPA+ RA har andre risikofaktorer enn ACPA- RA (11), og de fleste kartlagte genetiske assosiasjonene og miljørisikofaktorene (12) (for eksempel røyking) er hovedsakelig knyttet til ACPA+ sykdom. Mindre er kjent om det genetiske bidraget til ACPA- sykdom. Dette tyder igjen på ulik patogeneser ved de to subfenotypene. På bakgrunn av dette vil mest sannsynlig de ulike ACPA undergruppene av RA reagere forskjellig på behandling og må derfor behandles med ulike medikamenter og eventuelt ulike doser.(13, 14)

RA er mest utbredt hos kvinner mellom 45-60 år og ca. 0,5-1,0 % av den voksne befolkningen rammes av denne sykdommen. Det er en kompleks sykdom hvor flere gener og miljø risikofaktorer er nødvendig for sykdomsutvikling. Det meste av etiologien er ukjent, men flere studier har vist at genetiske faktorer er viktige bidrag i RA utvikling.(15) MacGregor et.al. forsket på arvelighet i RA i to kohorter av tvillinger og estimerte det relative bidraget av genetiske faktorer til 60 % (og 40 % miljø).(16) Et mål på familiær gruppering, nemlig søskenrisiko (λs), ble kalkulert til 8 for RA, som betyr at et søsken av en syk person har åtte ganger høyere risiko for å utvikle RA enn en tilfeldig person i den generelle populasjonen.(17) Silman og kolleger estimerte RA konkordansratene til monozygote tvillinger til å være ~15 % og for dizygote tvillinger til å være ~4 %. Dette vil si at hos monozygote tvillinger med tilnærmet likt arvemateriale, har en tvilling av en person med RA nesten 4 ganger større sjanse for å utvikle sykdommen enn mellom dizygote tvillinger (ca. 50 % felles arvemateriale) hvor den ene er rammet. (18) En populasjonsstudie utført av Hemminki et. al. viste at barna til pasienter med RA har en økt standardisert insidensrisiko for å utvikle andre autoimmune sykdommer som for eksempel T1D og cøliaki.(4) De arvelige årsakene til RA er multifaktorielle hvor en rekke genetiske- og miljømessige faktorer øker mottakeligheten for sykdommen. Det er hittil identifisert 101 RA risiko loci (se delkap. 1.3.1 for nærmere beskrivelse). (12) Hver risikovariant har en liten effekt på sykdomsdisposisjonen.

Den disponerende faktoren som bidrar aller mest til sykdomsrisiko hos RA er visse varianter av humant leukocytt antigen (HLA)-DRB1 alleler. *HLA* genene er lokalisert på kromosom 6 og

koder antigen presenterende molekyler som presenterer peptider til T celler. *DRB1* allelene som disponerer for RA har en delt epitop (shared epitope; SE). SE henviser til en spesifikk sekvens med aminosyrer i posisjon 70-74 i den peptidbindende gropa på HLA-DRB1 proteinet**,** hvor risikovariantene har delt epitop ved at det tredje hyper-variable området til alle RA disponerende *DRB1* alleler har tilnærmet lik sekvens sammensetning.(19) Mye av risikoen som kan tilskrives HLA er assosiert med variasjoner ved *HLA-DRB1* SE. Den best etablerte miljøfaktoren som påvirker risiko for RA er røyking, og tobakk røykere har en økt risiko for utvikling av ACPA + RA sammenliknet med de som ikke røyker. I 2006 publiserte Klareskog og kolleger en studie hvor de viste at SE alleler bare er assosiert med ACPA + RA, og at det er en gen-miljø interaksjon mellom SE og røyking. Individer som hadde en forhistorie med røyking og to kopier av SE allelene, hadde 21 ganger økt risiko for å utvikle ACPA + RA sammenliknet med ikke røykere uten SE alleler.(20) Den økte risikoen for utviklingen av RA ved røyking avhenger også av hvert individs forbruk.

Det er fremstilt en hypotese om at langtids røyking kan aktivere enzymer som bidrar til citrullinering av proteiner i lungene hos individer som er genetisk disponert og bærer *HLA-DRB1* SE. Tilstedeværelsen av de citrullinerte proteinene fører til aktivering av T-celler, som videre aktiverer B-celler og produksjonen av antistoff mot de citrullinerte proteinene. En annen inflammatorisk hendelse som trolig har blitt trigget av en tilleggsfaktor som for eksempel en infeksjon eller traumer skjer så i leddhulen(økt leddvæske) hvor aktivering av enzymer som citrullinerer proteiner og øker aktiveringen av T- og B-celler trolig fører til en kronisk inflammasjon i leddene og utvikling av RA. (14, 21)

1.3 Komplekse sykdommer og Genome-Wide Association Study (GWAS)

Mendelske- eller monogene sykdommer er sykdommer som er forårsaket av mutasjoner i ett enkelt gen som for eksempel cystisk fibrose og Huntingtons sykdom. Disse sykdommene kan nedarves fra foreldre ved enten recessiv eller dominant arv, hvor du henholdsvis må arve en sykdomsmutasjon fra mor og en fra far eller bare arve en fra enten mor eller far for å bli syk.

Komplekse- eller polygene sykdommer som for eksempel diabetes, kreft, og autoimmune sykdommer, er sykdommer som ikke følger et mendelsk arvemønster, men er forårsaket av en kombinasjon av flere gener, miljø- og livsstilsfaktorer. Sykdommene er heterogene fordi kombinasjonen av risikofaktorer hos hver pasient varierer, og de er vanskelig å behandle fordi man har lite informasjon om hva som forårsaker dem pga. at det er kompliserte kombinasjoner av alle faktorene som må slå inn samtidig. I tillegg kan en person ha den genetiske forutsetningen for å utvikle en kompleks sykdom, men personen blir ikke syk så lenge ikke de riktige miljøfaktorene spiller inn. Det er gjort få studier på gen-miljø interaksjoner som påvirker sykdom.(22) En ny metode innen forskningen på autoimmune sykdommer ser derfor nå på forholdet mellom immunsystemet, pasientens genetikk og pasientens mikrobiom, hvor de har sett antydning til at det er endring av mikrobiomet i pasienter med RA, cøliaki og T1D mellitus i forhold til friske kontroller.(23)

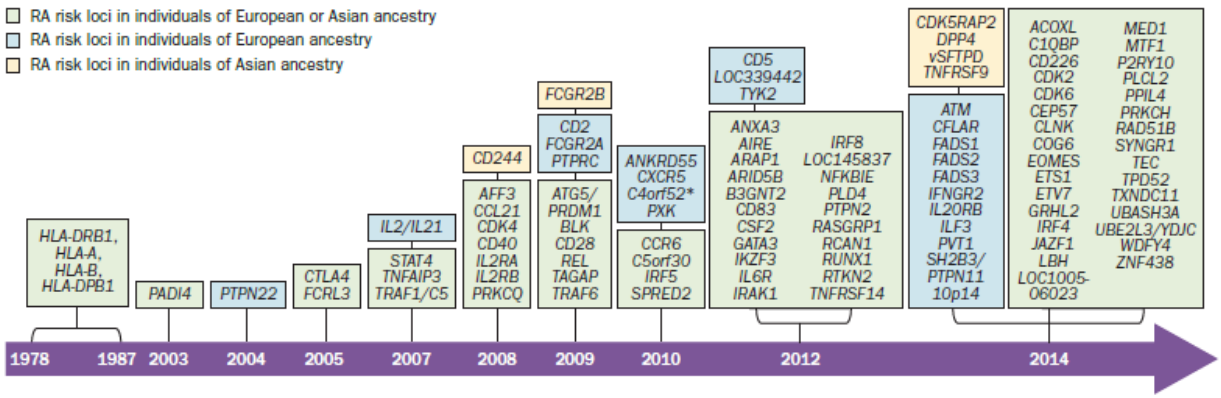
Assosiasjonsstudier brukes ofte i forskning på genetikken i multifaktorielle sykdommer. Før GWAS ble startet opp (rundt 2005), var identifiseringen av sykdomsassosierte genvarianter basert på kandidatgen metoden eller screening ved familiebaserte koblingsstudier. Som oftest endte disse med motstridende resultater og etablering av lite ny kunnskap. En svakhet ved kandidatgenstudiene var at kunnskapen om både sykdomsmekanismene og geners funksjoner generelt var begrenset. I tillegg fantes det begrensete opplysninger om polymorfismer i menneskets genom. Men mest avgjørende for både kandidatgen studiene og for koblingsstudiene var det at styrken på studiene ofte var for lave til å detektere risikovariantene. Kohortene som ble undersøkt besto gjerne av et par hundre individer. I kandidatgenstudiene ble det som regel ikke sett på mer enn en eller noen få polymorfismer av gangen, mens i koblingsstudiene var ofte det genetiske mangfoldet stort (stor genetisk heterogenitet) og det var få familier med gitt

risikofaktor. På tross av begrensningene ved kandidatgenstudiene endte det i noen funn, som for eksempel av genene *HLA* og *PTPN22* assosiert med RA. (24)

Det Humane Genom Prosjektet (HGP) resulterte i 2003 med kartleggingen av mosaikk sekvensen av et humant genom. Det internasjonale HapMap prosjektet fra 2005 ga en oversikt over koblingsulikevekt arkitekturen mellom polymorfismer i det humane genomet. Begge prosjektene har blitt viktige verktøy i kartleggingen av sykdomsgener og er bakgrunnen for at GWAS var mulig å utføre, i tillegg til den teknologiske utviklingen av mikromatriser som muliggjør genotyping av tusenvis av polymorfismer. 1000 Genomes prosjektet ble startet for å se på mangfoldet mellom personer av genetisk variasjon, og bidra til en katalog over polymorfismer i menneskets genom. Dette prosjektet har vært et svært viktig bidrag for GWAS i de senere årene. (25) GWAS undersøker SNPer i hele genomet i tusenvis av personer for å se etter assosiasjon med en spesifikk sykdom. Microarray brukt til genotyping i GWAS inneholder mellom 0,5-1,0 million SNP markører og er designet slik at de fanger opp mesteparten av den vanlige SNP variasjonen i den kaukasiske populasjonen, enten direkte eller indirekte gjennom koblingsulikevekt (LD) ved å markere felles haplotyper (såkalte merke-SNPs). For å gjennomføre en GWAS trengs det to grupper med deltakere, en gruppe med personer som har den aktuelle sykdommen og en gruppe med friske personer for å se på forskjeller i allelfrekvens for SNPene. Det er viktig at begge gruppene har samme etniske bakgrunn for å nulle ut allelfrekvens forskjeller som skyldes ulik populasjonsbakgrunn og dermed unngå populasjonsstratifikansjon. Det finnes ingen genotype array som dekker alle SNPene i det menneskelige genomet og bare SNPer med minor allelfrekvens (MAF) > 5 % kan fanges opp av koblingsulikevekt i GWAS analyser. (24) GWAS katalogen er en offentlig tilgjengelig database med publiserte GWAS data. Katalogen inneholder 2111 publikasjoner og 15396 SNPer per 20.02.15. (26)

1.3.1 GWAS i RA og autoimmune sykdommer

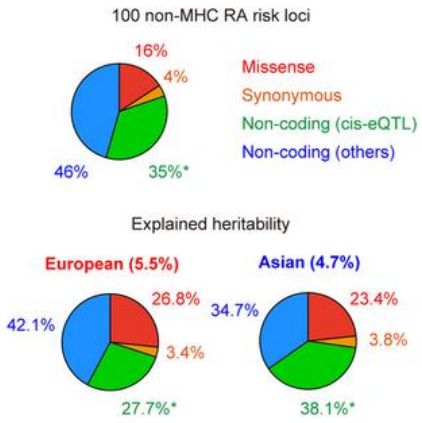
GWAS har de siste 10 årene bidratt til identifiseringen av de totalt 101 RA risiko loci funnet. Under 10 risikogener var kjent før GWAS ble initiert (figur 2). Etter GWAS ble startet med hypotesefrie undersøkelser av opptil en million SNPer pr. studie og dermed stor fare for falske positive funn, ble det satt krav om et «genome wide» signifikansnivå på $p < 5.0 * 10^{-8}$ før et locus ble erklært å være et risikolocus. HLA komplekset eller «major histocompability complex» (MHC) viste seg i alle GWAS i RA å være den sterkeste genetiske determinanten, som forklarer omtrent halvparten av arveligheten. Okada og kolleger regnet ut at de øvrige 100 RA risikoloci (utenfor MHC) forklarer 5,5 % og 4,7 % av arveligheten i hhv. europeere og asiatere. De fant også at 2/3 av RA risiko loci viste pleiotropi med andre fenotyper som immunrelaterte sykdommer, inflammasjonsrelaterte og hematologiske biomarkører og andre komplekse egenskaper. For de 101 risikoloci har 98 biologisk kandidat gener blitt identifisert, hvor nitten av loci inkluderte flere gener. De fleste loci er knyttet til immunologiske gener. Okada et. al. undersøkte den potensielle rollen til RA genetikk med hensyn til legemidler og fant at 27 gener kodet for proteiner som er mål for RA medisiner viste signifikant overlapp med de 98 biologiske RA risikogenene. RA risiko genene ble også testet mot mål for godkjente medisiner for andre sykdommer hvor spesielt tre medisiner for ulike typer kreft skilte seg ut. (12)



Figur 2: Historisk oversikt over gener peket ut av sykdomsmottakelighetspolymorfismer i RA. Hvor risiko loci er selektert med en genom-vid signifikans terskel på $p < 5.0 * 10^{-8}$. (27)

GWAS har også bidratt til å avdekke at mange av RA risikoloci er delt med risiko loci for andre autoimmune sykdommer, og generelt er overlappet mellom risiko loci stort for alle autoimmune sykdommer. Flere sykdommer deler mer enn 50 % av sine assosierte loci med andre sykdommer, som for eksempel cøliaki og RA, og cøliaki og T1D. Dette støttes opp under det faktum at personer med en autoimmun sykdom har økt risiko for å utvikle en annen autoimmun sykdom**,** som for eksempel at personer med psoriasis har en økt risiko for å utvikle RA og motsatt. (23) Et av de første eksemplene på overlappende autoimmune sykdoms assosiasjoner var til kromosom 4q27 som inneholder *IL2* og *IL21* genene og som viser assosiasjon med cøliaki, RA, T1D og Ulcerøs kolitt (UC). Det er også vist at sykdomsassosierte alleler i noen loci kan ha motsatt effekt, som betyr at det samme allelet utøver risiko for en sykdom, men er beskyttende for en annen. Et eksempel på dette er hos *PTPN22* hvor det assosierte allelet har motsatt effekt i Crohn`s sykdom i forhold til i T1D og RA. (24) Etter at det ble kjent at mange genetiske faktorer deles på tvers av flere immun-relaterte sykdommer ble Immunochip utviklet for og nøyaktig kartlegge immunrelaterte loci og for å teste risikoloci på tvers av autoimmune sykdommer. Immunochip er et SNP array som er utviklet for replikasjon og nøyaktig kartlegging. (24)

Genvarianter assosiert med RA og andre autoimmune sykdommer befinner seg i stor grad nær immungener og antas hovedsakelig å være regulatoriske polymorfismer da de er lokalisert i introner eller intergene regioner. Hos RA ligger 80 % av risiko SNPene utenfor kodende områder (figur 3). (12) Hele 90 % av GWAS autoimmune sykdoms funn har ikke kunnet forklares av protein-kodende varianter som bekrefter at det er mest av de regulatoriske variantene.(28) For en del av polymorfismene er det funnet assosiasjon til at genuttrykket blir opp- eller nedregulert. (24)



Figur 3: Funksjonell annotering av SNPene i 100 ikke-MHC RA risikoloci, inkludert den relative andelen av arvelighet forklart ved SNP annoteringer. (12)

Flere ulike immun-relaterte sykdommer er ikke bare assosiert med samme gen, men med den samme biologiske veien («pathway»). På den måten åpner muligheten seg for at flere immun-relaterte sykdommer kan behandles med samme legemidler som hemmer den biologiske veien. Visse biologiske veier som er funnet å være assosiert med flere vanlige autoimmune sykdommer har blitt kjente medisinske mål og brukes i dag aktivt i behandling. TNF (tumor nekrose faktor) reseptor veien er et eksempel på dette. TNF superfamilien av cytokiner representerer en multifunksjonell gruppe av proinflammatoriske cytokiner som aktiverer signalveier for celleoverlevelse, apoptose, inflammatorisk respons og cellulær differensiering. TNF-hemmere brukes som anti-inflammatoriske medisiner og anti-TNF medikamenter brukes i dag for å behandle Crohn`s sykdom, psoriasis, RA og systemisk lupus erythematosus (SLE) med varierende resultater (23 % av RA pasienter responderer ikke). Flere av de kjente cytokinveiene har også blitt assosiert med autoimmune sykdommer og medisiner er utviklet som har disse spesifikke cytokinene som mål, som for eksempel medisiner mot psoriasis og RA. (23, 29)

For enkelte autoimmune sykdommer kan man se at det forekommer en ganske stor forskjell i den genetiske variasjonen og effekten på bakgrunn av hvor i verden en person kommer fra. I 2003 ble de første RA assosierte SNPene utenfor *HLA* gen locuset rapportert i *PADI4* i asiatisk befolkning, men de fikk ikke de samme resultatene for assosiasjon med polymorfismene i europeisk befolkning. Flere år senere ble det allikevel bekreftet at polymorfismer i *PADI4* genet kunne assosieres med risiko for utvikling av RA i europeisk befolkning, men effekt størrelsen av de identifiserte variantene var mindre enn i de asiatiske populasjonene. (27) Et annet eksempel er

at de i Japan har funnet assosiasjon mellom RA og transkripsjonsfaktoren autoimmun regulator (AIRE) for japanske pasienter. Det samme har ikke kunnet påvises i europeisk befolkning, men flere SNPer ved kromosom 21q22.3 hvor *AIRE* befinner seg har blitt assosiert med cøliaki, RA og inflammatorisk tarmsykdom (IBD). (23, 30)

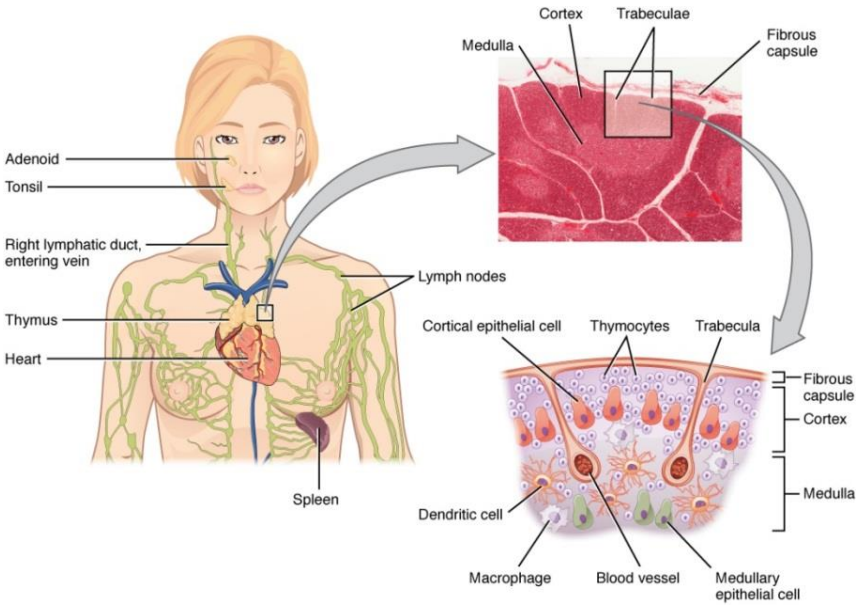
Ved å utføre GWAS på immunrelaterte sykdommer kan det bidra til utviklingen av nye teorier om de underliggende sykdomsmekanismene. Denne informasjonen kan være med på å utvikle bedre metoder for å oppdage, behandle og forebygge slike sykdommer. Et viktig mål er og etter hvert kunne utvikle persontilpasset medisinering som trolig vil bedre mange menneskers helse da de får en mer tilpasset behandling i forhold til deres unike genetiske oppbygging. (24)

1.4 Tymus

Tymus er et lymfatisk organ som ligger bak brystbenet hos mennesker (figur 4). Den inneholder kjertelvev og produserer flere hormoner, som for eksempel thymosin som stimulerer utviklingen og produksjonen av T-celler. På tross av dette er tymus i høyere grad assosiert med immunsystemet enn endokrinsystemet. Årsaken til dette er at tymus er der hvor modning og utvikling av T-celler finner sted.

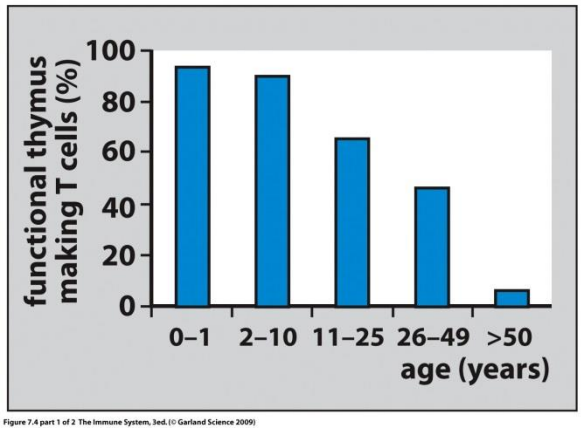
I tymus skjer det først positiv og deretter en negativ seleksjon av T-celler. Positiv seleksjon er prosessen som sikrer at T-cellene gjenkjenner peptider presentert av kroppens egne HLA molekyler. Umodne T-celler (thymocytter) fra beinmargen entrer cortex av tymus hvor de kommer i kontakt med epitelceller som bærer HLA ligander på celleoverflaten. De umodne T-cellene (er dobbelt negative og uttrykker verken CD4 eller CD8) tester sine T-cellereseptorer (TCR) og de som binder seg til HLA molekylene med medium affinitet får et positivt overlevelsessignal og blir dobbelt positive thymocytter med $CD4^+$ og $CD8^+$. Interaksjoner mellom TCR og HLA-peptid kompleks på de kortikale epitelcellene (cTEC) avgjør om thymocytten blir en enkelt positiv $CD4^+$ eller $CD8^+$ thymocytt ved at den mottar ulike modningssignaler. Dersom T-cellen har en TCR som gjenkjenner HLA-klasse I molekyl vil den slutte å uttrykke $CD4^+$ og bare uttrykke $CD8^+$, mens om den gjenkjenner HLA-klasse II vil den bare uttrykke $CD4^+$, modnes og migrerer til medulla. Negativ seleksjon er prosessen som sikrer

at T-cellene som er autoreaktive, dvs. reagerer på kroppens egne peptider blir fjernet. I medulla kommer de overlevende T-cellene i kontakt med de medullære epitelcellene (mTEC), som presenterer peptider fra kroppens egne proteiner. De T-cellene som binder til disse peptidene med for høy affinitet vil elimineres ved apoptose. De strenge seleksjonsprosessene bidrar til at de T-cellene som slippes ut i periferien kun reagerer på fremmede antigener fra patogener som bakterier og virus, ikke på peptider fra kroppens egne organer og vev. (31, 32)



Figur 4: Lokalisasjon, histologi og struktur av tymus. (33)

Tymus er størst og mest aktiv før du blir født og i årene før puberteten. Organet vil deretter minske med alderen. Ved fødselen veier tymus mellom 10-15 g og er 5 cm lang, mens ved puberteten når den sin maksimale størrelse på ca. 30-40 g. Etter puberteten erstattes tymus av bindevev og fett, men selv om den skrumper inn til et betydelig mindre organ er den fortsatt litt aktiv store deler av livet på tross av at immunsystemet har produsert største delen av T-cellene under barndommen (figur 5).(34)

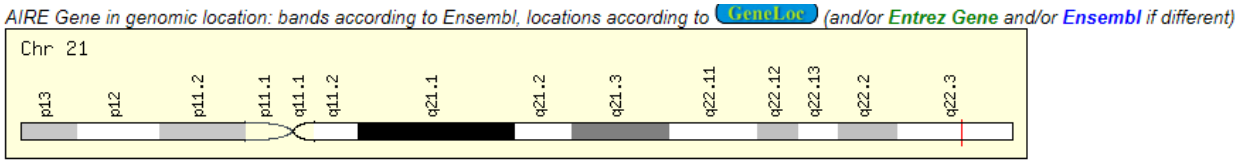


Figur 5: Produksjonen av T-celler (%) i en funksjonell tymus ved ulik alder. (35)

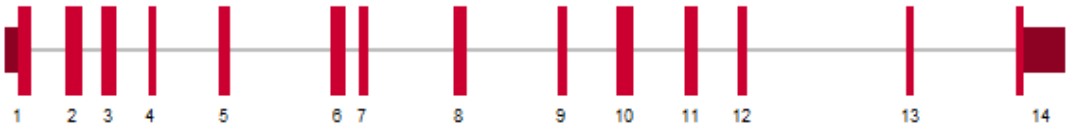
1.5 AIRE

Autoimmun regulator (AIRE) er en transkripsjonsfaktor som er uttrykt av mTEC i tymus og her regulerer den uttrykket av vevsspesifikke proteiner slik at selv-peptidene kan presenteres for thymocyttene. Genet er også uttrykt i sekundært lymfevev, men dette er ikke like godt forsket på i mennesker.(36) AIRE har en struktur typisk for proteiner som kan bindes til kromatin og regulere gentranskripsjon. *AIRE* genet tilhører genfamilien PHF (PHD-type zinc fingers). Det er lokalisert på kromosom 21, mer eksakt på den lange q armen ved posisjon 22.3, og strekker seg fra basepar 44,285,838 til 44,298,219 (Hg38) og består av 14 eksoner (figur 6). (30, 37)

A



B



Figur 6: *AIRE* genet med A) lokalisasjon på kromosom 21, (38) og B) skjematisk skisse av genstrukturen med 14 ekson (røde streker).(39)

Mutasjoner i *AIRE* kan føre til at immunsystemet får en funksjonssvikt, som videre kan resultere i autoimmunitet. «Autoimmune polyglandular syndrome» (APS) type 1/ «Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy» (APECED) er et eksempel på dette. Sykdommen er sjelden og forårsaket av mer enn 60 mutasjoner i *AIRE* genet. Noen av de genetiske endringene fører til produksjon av en unormalt kort, ikke-funksjonell versjon av autoimmun regulator proteinet. Andre mutasjoner kan endre enkelte aminosyrer på kritiske regioner i proteinet. Det er en arvelig sykdom (autosomal recessiv) og den påvirker mange av kroppens organer. Sykdommen kan karakteriseres med i hvert fall 2 av de 3 følgende funnene; svakt fungerende skjoldbruskkjertel som kontrollerer kalsium, soppinfeksjon og svakt fungerende binyrer. Av uklare grunner tror man at defekter i AIRE proteinet primært påvirker de hormonproduserende kjertlene som ligger til grunn for mange av de viktigste egenskapene hos APS1/APECED.(37) Klinisk presentasjon av APS1 omfatter T1D, vitiligo, Addisons sykdom og andre fenotyper som også har en kompleks genetikk. Gener ved monogene tilstander kan også vise seg å ha mer frekvente genetiske varianter som gir økt risiko for komplekse sykdommer. *AIRE* har (også) blitt assosiert med RA i japanere. Mens flere SNPer ved 21q22.3, nære *AIRE* genet, har blitt funnet å være assosiert med cøliaki, RA og IBD i europeisk befolkning. (23)

2. Formål med oppgaven

Målet med oppgaven var å undersøke om det er assosiasjon mellom polymorfismer i *AIRE* genet og kromosomområdet og RA i den norske populasjonen.

Videre ønsket vi å studere om de undersøkte polymorfismene viste korrelasjon med genuttrykk, samt få innblikk i spleisevarianter av *AIRE* i tymus.

3. Material og metoder

Analysene som er utført i denne masteroppgaven er basert på et pasient-kontroll studie design. Her sammenliknes en pasientgruppe med en frisk kontrollgruppe fra samme populasjon for å prøve og identifisere faktorer som kan bidra til sykdom.

Alle databasene/verktøyene brukt i dette forskningsprosjektet er beskrevet i kapittel 3.8.

3.1 Materialer

Kontroller (N=1098) er hentet fra «Det norske beinmargsregisteret» ved avdeling for immunologi og transfusjonsmedisin, Oslo universitetssykehus avdeling Rikshospitalet. RA prøvematerialet (N=944) er samlet ved Diakonhjemmet sykehus. Alle pasienter og kontroller har gitt skriftlig samtykke og etisk godkjennelse ble innhentet fra den regionale etiske komiteen. Alle pasienter og kontroller er av norsk opprinnelse for å unngå populasjonsstratifikasjon.

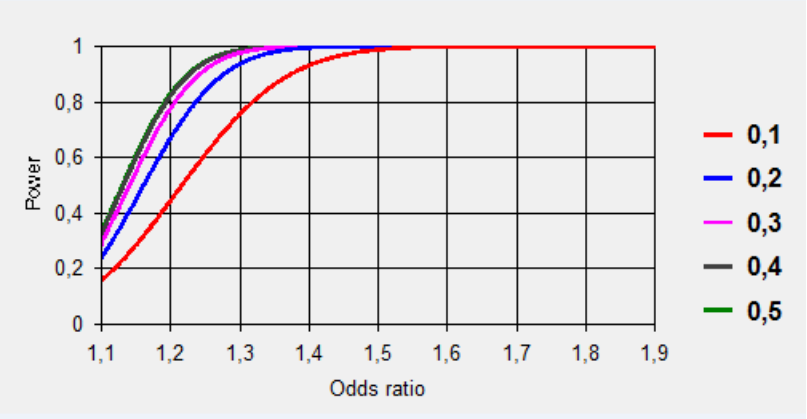
Vevsprøver fra tymus har blitt samlet inn fra 42 norske barn under 13 år som skulle hjerte opereres. Kjønnsfordelingen var på 22 jenter og 20 gutter. Prosjektet er godkjent av den regionale etiske komiteen og skriftlig samtykke ble gitt av foreldre. Alle vevsprøvene ble anonymisert.

DNA og RNA ekstraksjon for alt prøvematerialet var allerede utført før denne masteroppgaven ble påbegynt.

Hel-genom amplifisert-DNA (WGA-DNA) ble benyttet i stedet for genomisk DNA av den grunn at tilgangen til genomisk DNA var begrenset. Det genomiske DNAet ble derfor WGA amplifisert med repliG Midi Kit (Qiagen, Hilden, Tyskland). 1 µl av WGA-DNA ble fortynnet med 19 µl 1xTris-EDTA(TE)buffer.

3.2 Styrkeberegning

Styrkeberegninger, for vår case-kontrollstudie med 944 pasienter og 1098 kontroller, ble gjort i programmet PS power and Sample size calculation (http://ps-power-and-sample-size-calculation.software.informer.com/). Prosjektet vårt tar utgangspunkt i *AIRE* assosiasjon rapportert i en japansk studie (30). Gitt at risiko SNPen rs2075876 har samme allelfrekvens (0,34) og OR (1,2) i vår populasjon som i Japan, har vi 81 % power til å detektere en assosiasjon med signifikansnivå p<0,05. Styrken vil variere for ulike SNPer inkludert vårt SNP panel, ved at den svekkes med lavere allelfrekvens. Figur 7 viser hvordan styrken påvirkes av odds ratio (OR) og allelfrekvensen.



Figur 7: Grafisk fremstilling av hvordan styrken påvirkes av odds ratio og allelfrekvens (ulike farger på kurvene er for ulike allelfrekvenser) i vår case-kontrollstudie.

3.3 SNP seleksjon

For å undersøke om genetisk variasjon i *AIRE* genet bidrar til risiko for RA i vår populasjon, ble det tatt utgangspunkt i tidligere rapporterte assosiasjoner mellom SNPer i RA. Det fantes en tidligere studie av Terao et.al.(30) hvor de har rapportert tre SNPer, rs2075876, rs760426 og rs1800520, i *AIRE* genet som var assosiert med RA hos japanere. Disse tre SNPene ble derfor utgangspunktet i vår SNP seleksjon.

Det har ikke tidligere blitt påvist at polymorfismer i *AIRE* genet hos nordmenn/europeere er assosiert med RA, av den grunn ble det valgt ut flere SNPer for å undersøke om polymorfismer i omkringliggende områder viste assosiasjon. Det ble gjort et søk på «rheumatoid arthritis» og kromosomområdet 21q22.3, hvor *AIRE* er lokalisert, i GWAS katalogen, som er en offentlig tilgjengelig database med publiserte GWAS data fra både små og store genetiske assosiasjonsstudier. (https://www.genome.gov/26525384) Søket ble utført 09.01.15, da inneholdt katalogen 2087 publikasjoner og 15177 SNPer fra siste oppdatering 18.12.14, hvorav 21 publikasjoner var på RA. Søket resulterte i tre studier (12, 30, 40). En av studiene rapporterte om SNP assosiert med RA i *AIRE* i japansk befolkning (artikkelen som vår hypotese er basert på) (30), mens de to andre studiene rapporterte om SNPer assosiert med RA i genet *UBASH3A* hvor den ene studien (40) bare fant dette hos europeere og den andre (12) fant det hos både europeere og øst-asiatere.

Siden risikogener for autoimmune sykdommer viser stor grad av genetisk overlapp, ble søket utvidet til å inkludere andre autoimmune sykdommer i samme kromosomområde som *AIRE*. Alle publikasjoner (N=47) med rapporterte assosiasjoner i 21q22.3 ble derfor gjennomgått og de som omhandlet autoimmune fenotyper og var hel-genom signifikante med $p<5*10^{-8}$ ble selektert (søket ble utført 09.01.15). De selekterte publikasjonene hadde detektert SNPer (N=7) med assosiasjon til de autoimmune sykdommene cøliaki, T1D, Crohn`s sykdom, IBD, vitiligo og UC i genene *ICOSLG* og *UBASH3A*.

En annen database kalt Immunobase (https://www.immunobase.org/page/Welcome/display**)** ble også benyttet til å søke på RA assosiasjon i kromosomområdet 21q22.3 for å undersøke om den

inneholdt informasjon som ikke var registrert i GWAS katalogen. Immunobase er en web-basert kilde som fokuserer på genetikk relatert til immunologiske sykdommer.

rs1800520 er en ekson-polymorfisme som gir en aminosyreforandring i *AIRE* genet ved at aminosyren serin byttes ut med arginin. SNPen er i sterk koblingsulikevekt (r^2= 0,94) med rs2075876, men den viser ikke sterk nok assosiasjon til RA (p= 0,0071) i forhold til kriteriet som ble satt på $p<5*10^{-8}$ for seleksjonen. SNPen ligger derfor ikke i GWAS katalogen, men den ble tatt med som en funksjonell kandidat fra artikkelen av Terao et. al. (30).

Vi ønsket å undersøke koblingsulikevektsmønsteret til SNPene fra GWAS-søket både i japansk og europeisk befolkning for å kunne ekskludere/inkludere flere SNPer. Vår hypotese var at SNPer som var i koblingsulikevekt med de to RA risiko SNPene (rs2075876 og rs760426) hos japanere, men ikke i koblingsulikevekt med disse SNPene hos europeere kunne forklare at det ikke var funnet noen assosiasjon mellom RA risiko SNPene og RA hos europeere før. Slike SNPer ble derfor også inkludert etter følgende utvelgelseskriterier: 1) SNPer med koblingsulikevekt (proxy), $r^2 \geq 0,8$, med de to japanske sykdoms SNPene, rs2015876 og rs760426, i japansk befolkning ved å bruke datasett 1000 Genomes Pilot 1 populasjonspanel CHBJPT (Han kinesere og japanere i Tokyo). 2) Deretter ble de av de SNPene fra trinn 1 som hadde koblingsulikevekt $r^2<0,8$ med rs2075876 og rs760426 hos europeere (populasjonspanel CEU («Utah residents with Northern and Western European ancestry»)) selektert (N=2). For å oppsummere: SNPene som var i koblingsulikevekt med RA risk SNPene (rs2075876 og rs760426) i både japanere og europeere ble ekskludert, men dersom de kun var i koblingsulikevekt hos japanere og ikke hos europeere ble de inkludert.

For å kunne ekskludere «redundant» SNPer (SNPer i høy koblingsulikevekt med hverandre) fra hele SNP panelet, da disse gir veldig like genotypingsresultater, ble koblingsulikevektsmønsteret kartlagt mellom de utvalgte SNPene. SNP datasettet som ble brukt i søket var 1000 Genomes Pilot 1, og populasjonspanel ble satt til CEU, og SNPer med $r^2 \geq 0,8$ ble ekskludert ved at en vilkårlig SNP ble valgt å gå videre med i seleksjonen.

Snap Proxy Search med data fra 1000 Genomes ble brukt som søkeverktøy for koblingsulikevektsanalysene og begge er nærmere beskrevet i bioinformatikk kap.3.8.

Den siste delen av seleksjonsprosessen ble utført for å undersøke frekvensen til SNPene i CEU og dermed bare inkludere SNPer som var polymorfe i europeere med en allelfrekvens >0,05. SNPer med tre antatte alleler ble ekskludert da det bl.a. ville gjøre genotyping vanskelig. Databasen dbSNP med data fra både HapMap og 1000 Genomes katalogene ble brukt til å utføre dette søket og er nærmere beskrevet i kap.3.8.

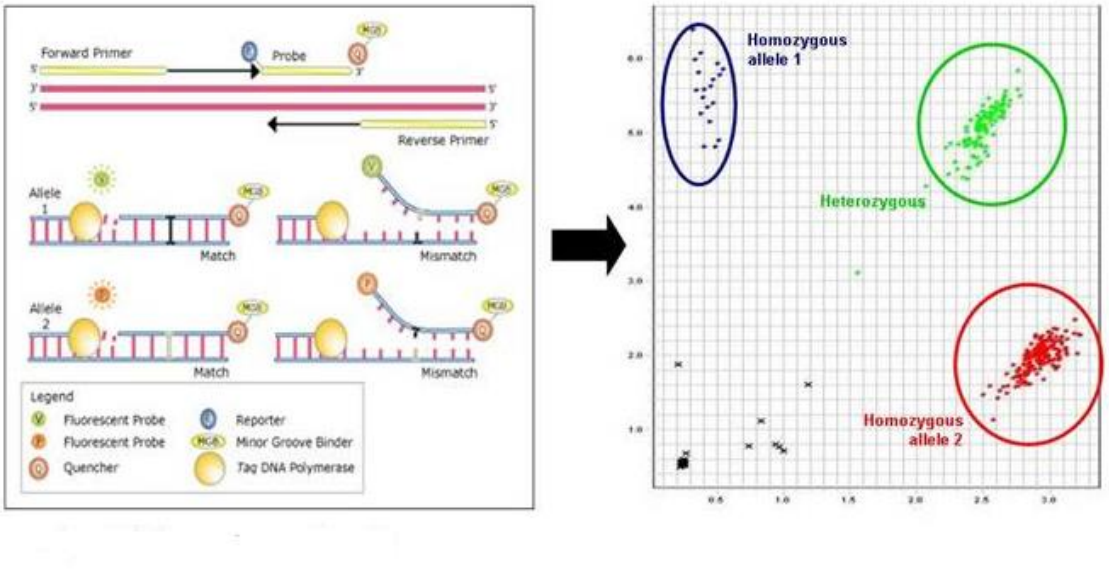
Seleksjonen resulterte i seks SNPer rs2075876, rs760426, rs1800520, rs3788113, rs7282490 og rs4819388 (tabell 4).

3.4 TaqMan genotyping

Alleldiskriminerings analyser detekterer ulike varianter av SNPer ved hjelp av sekvensspesifikke oligonukleotider og denne metoden ble brukt til genotypingen av de seks SNPene. TaqMan assayene ble bestilt fra Thermo Fisher Scientific (Waltham, MA, USA) og er listet i tabell 1. Fem av de seks TaqMan assayene ble bestilt ferdig designet, mens TaqMan assayet til SNPen rs1800520 ble spesial designet med Thermo Fisher Scientific sine prosedyrer beskrevet på deres hjemmeside. (https://www.lifetechnologies.com/no/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/snp-genotyping-taqman-assays.html)

TaqMan alleldiskriminering er en polymerase kjede reaksjon (PCR) basert metode for SNP genotyping. Allelspesifikke fluoressens merkede prober «annealer» spesifikt til den komplementære DNA tråden, men fluoreserer ikke på grunn av at en «quencher» er bundet i andre enden av proben og absorberer fluoressens fra reporteren. Alleldiskriminering bruker prober som er spesifikke for hvert SNP allel som merkes med to ulike fluoressens reporter farger, som i denne studien var VIC og FAM. Primerne som er festet til templatet forlenges av Taq polymerasen og degenererer prober ved polymerase 5`nuklease aktivitet. Reporteren blir dermed separert fra «quencheren» og det oppstår fluoressens som kan måles på slutten av PCR (figur 8). Fluoressenssignalene som genereres fra PCR amplifiseringen indikerer hvilket allel som er til stede i prøven. (41)

Fluoressens målingene (av rådata) ble utført etter PCR med SDS 2.4, som er en automatisert programvare som behandler fluoressens dataene og lager genotype plott. Plottet kan bestå av tre cluster av prøver som representerer ulike genotyper, to homozygote- og et heterozygot cluster (figur 8). (41)



Figur 8: Til venstre er TaqMan SNP genotyping illustrert, hvor TaqMan polymerasen under amplifisering kløyver «quenceren» fra reporteren som gir fra seg allelspesifikt fluoressens signal. Til høyre er det et genotypingsplott som kan lages etter PCR med SDS 2.4, med y-akse = FAM signal(blå) og x-akse = VIC signal(rød). Blått og rødt «cluster» består av homozygote prøver og grønt representerer prøver med signal fra både FAM og VIC, og dermed er heterozygote prøver. (42)

3.4.1 Materialer

Kjemikalier

- 10 ng WGA i 1xTE buffer
- 2xTaqMan® Genotyping Master Mix, Applied Biosystems (Foster City, CA, USA), lot.1203070, part.no. 4371355
- 2xABsolute ^{TM} QPCR ROX mix, Thermo Scientific (Epsom, Surrey, UK), lot.00083730, part.no. AB-1138/B
- 40xTaqMan assay mix (SNP genotyping assay), Thermo Fisher Scientific
- Positiv kontroll: MOU (International Histocompatibility Working Group (IHWG) reference panel #IHW09050)
- Negativ kontroll: dH_2O

Utstyr

- 384-brønners optiske plater, Life Technologies, part.no. 4309849
- Optisk film, Micro Amp, Optical Adhesive Film, Applied Biosystems, part.no. 4311971
- Gene Amp, PCR Systems 9700 db384, Applies Biosystems
- Heraeus, Function Line, Labofuge 400R
- Centrifuge 5810R, Eppendorf
- ABI Prism, 7900 HT Sequence Detection System, Applied Biosystems (SDS programvare versjon 2.4)
- Biomex FX, Beckman Coulter Genomics (Brea, CA, USA)

Taqman genotypingen ble utført med noen få endringer i forhold til protokollen;

- Vi brukte WGA (verifisert for genotyping og sekvensering), i stedet for genomisk DNA eller komplementær DNA som var de anbefalte templatene for TaqMan SNP Genotyping Assays.

- I tillegg til 2xTaqman® Genotyping Master Mix som TaqMan assayene er designet og optimalisert til, ble 2xABsolute ™ QPCR ROX mix brukt for noen av TaqMan assayene. Alle assayene ble først testet med 2xABsolute ™ QPCR ROX mix, og de som «clustret» dårlig med denne mastermiksen ble testet med 2xTaqMan® Genotyping Master Mix. Hvilken av miksene som ble brukt for ulike TaqMan assayer er oppgitt i tabell 1 med ekstra opplysninger om assayene.

SNPene rs760426, rs7282490 og rs4819388 for 42 tymusprøver var allerede genotypet med Immunochip som er et Illumina HumanWG-6 v3 array med 200 000 SNPer som dekker områder som viser assosiasjon med autoimmune sykdommer. De tre resterende SNPene fra seleksjonen, rs2075876, rs3788113 og rs1800520, for de 42 tymusprøvene ble derfor genotypet på lik linje med alle RA prøver og kontroller.

Av de seks TaqMan assayene var det kun rs3788113 som var funksjonelt testet.

Tabell 1: Oversikt over 40xTaqMan SNP Genotyping Assay.

SNP ID	Gen	Lot	Assay ID	Mastermiks	VIC	FAM	Kons.*	PCR program
rs2075876	*AIRE*	P150112-002C12	C__15863141_10	2xABsolute ™ QPCR ROX mix	A	G	80 %	A
rs760426	*PFKL, AIRE*	P150112-002D01	C____588344_10	2xABsolute ™ QPCR ROX mix	A	G	70 %	A
rs3788113	*AIRE, PFKL*	P150112-002D02	C___2978255_10	2xABsolute ™ QPCR ROX mix	A	G	80 %	A
rs7282490	*ICOSLG*	P150112-002D03	C__29159235_10	2xTaqman® Genotyping Master Mix	A	G	70 %	B
rs4819388	*ICOSLG*	P150112-002D04	C__26539794_20	2xTaqman® Genotyping Master Mix	C	T	70 %	B
rs1800520	*AIRE*	1397910	AHS1QEY	2xABsolute ™ QPCR ROX mix	C	G	80 %	A

* probe/primer konsentrasjon

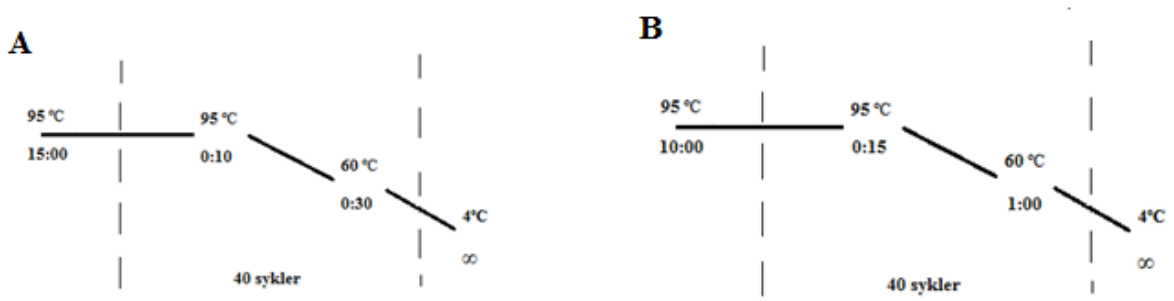
2 µl WGA (5 ng/µl) ble overført fra 96-brønners brett til bunnen av 384-brønners brett ved hjelp av pipetteringsrobot (Biomex FX, Beckman Coulter Genomics). Så ble prøvene fullstendig tørket ved fordampning/luft tørking i et mørkt skap ved romtemperatur. Hver 384 plate inneholdt en positiv kontroll (MOU), en negativkontroll (dH_2O) og resten prøvemateriale, slik at seks 384 plater ble genotypet pr SNP.

5 µl 70 % eller 80 % assaymiks ble tilsatt i hver brønn. 70 % TaqMan alleldiskrimineringsmiks inneholdt 1000 µl mastermiks, 35 µl 40xTaqMan assay mix og 965 µl dH_2O. Mens 80 % TaqMan alleldiskrimineringsmiks inneholdt 1000 µl mastermiks, 40 µl 40xTaqMan assay mix og 960 µl dH_2O. For rs1800520 ble det for noen av prøvene som skulle genotypes (4 av 6 plater) ikke tørket WGA før kjøring, men det ble kjørt vått sammen med assaymiksen. Da ble bare 1 µl WGA og 4 µl assaymiks tilsatt i hver brønn. Assaymiksen for 400 prøver vått, skilte seg fra blandingsforholdet for 80 % assaymiks (tørt) med at det ble tilsatt 400,0 µl mindre dH_2O, men ellers med like volum.

Det ble laget mikser med både 70 % og 80 % av primer/probe konsentrasjon av den grunn at miksene skulle holde til flest mulig prøver og det ble godt nok resultat ved bruk av en lavere 40xprimer/probe miks konsentrasjon enn 100 % (som er anbefalt fra leverandør). Alle TaqMan assayene ble først testet med 70 % assaymiks, noen av assayene «clustret» litt for dårlig ved at man ikke så tre tydelige grupper av prøver i genotypingsplottet, men en stor gruppe (gjør det umulig å genotype prøvene, blir dårlig kvalitet) ved 70 % og derfor ble det laget assaymiks med 80 % primer/probe konsentrasjon for disse TaqMan assayene (tabell 1).

384 platene ble forseglet med optisk film og en kort sentrifugering ble utført for å få bort eventuelle luftbobler og for å få alt prøvematerialet ned i brønnen.

Prøvene ble amplifisert ved PCR med betingelsene for enten program A eller B vist i figur 9. *Det ble brukt to ulike PCR program av den grunn at de to mastermiksene som ble brukt hadde forskjellige optimaliseringstemperaturer og for at de skulle fungere best mulig måtte to ulike program benyttes.*



Figur 9: PCR-program for assay kjørt med A) 2xABsolute™ QPCR ROX mix og B) 2xTaqMan® Genotyping Master Mix.

PCR reaksjonen ble etterfulgt av en post-PCR avlesning (endepunktsdeteksjon av fluoressens) på ABI Prism 7900 HT Sequence Detection System og fluoressens data ble analysert på en automatisert programvare, SDS 2.4, ved «Allelic discrimination» format.

En del prøver (~80-100) fra hvert SNPassay som ikke lot seg genotype ved første kjøring ble forsøkt genotypet en gang til med samme metode som før med unntak av at det ble tilsatt 2 µl (10 ng/µl) WGA (dobbelt) til tørking siden de tidligere prøvene som ikke lot seg genotype hadde vært så svake. Kontroller fra hvert genotypings-«cluster» ble inkludert i alle kjøringene for å forsikre oss om at hver genotype var representert i plottet. (Ellers lik prosess som beskrevet over). Et tilfeldig utvalg av genotypingsplott for hvert SNP assay er vist i vedlegg 1.

3.5 Genekspresjon i tymus

Mikromatriseanalyse av total RNA fra de 42 tymusprøvene ble utført i 2005 ved bruk av Illumina HumanWG-6 v3 array (Illumina, San Diego, CA, USA). Alle RNA prøvene hadde vært analysert på Bioanalyzer 2100 instrument (Agilent Technologies, Santa Clara, CA, USA) for mikroarrayanalysene for å sikre renhet og RNA integritet. Genekspresjonsanalysene ble gjort på Radiumhospitalet på kjernefasiliteten for mikromatrise (The Norwegian Microarray Consortium, NMC; www.microarray.no). Rådata fra probeintensitetsmålingene ble ekstrahert ved hjelp av BeadStudio software, og normaliserte data ble log2 transformert i programmet J-express.

Genekspresjonsdata for *AIRE* (tre prober**)** og *ICOSLG* (én probe) ble trukket ut for de 42 tymusprøvene for å analyseres i denne masteroppgaven. Genekspresjonsfilen som ble brukt fra microarrayen hadde probekoordinater fra Hg18.

Microarray er et 2D array hvor et høyt antall DNA-prober festes i et tett rutemønster over et lite areal på et fast underlag for å se på genekspresjon. Aktive gener i prøvematerialet merkes med fluoressens, hybridiseres til microarrayet og blir analysert ved sensitiv fluorescensdeteksjon. Denne metoden kan analysere flere gener i ett og samme forsøk. (43)

3.5.1 Kvalitetskontroll av probene for *AIRE* og *ICOSLG*

Det ble utført en kvalitetskontroll av de fire probene, ILMN_1670282, ILMN_2261519, ILMN_1791236 og ILMN_1675671 for *AIRE* og *ICOSLG* sekvensert fra Illumina Human WG-6 v3 arrayet. Probene ble blastet med BLAT (https://genome.ucsc.edu/cgi-bin/hgBlat?command=start) som er et verktøy i UCSC Genome Browser databasen (søket ble utført 25.03.15). BLAT ble også brukt til å finne de oppdaterte koordinatene til probene, hvilke transkripter som binder probene og til å se på polymorfismer i probesekvensene. Hg38 ble valgt som referanse og RefSeq gener ble brukt under analysen. Det ble satt en begrensning på at SNPene måtte være funnet i mer enn 1 % av CEU da det skulle søkes etter polymorfismer i probene.

3.6 RNA sekvensering av fire tymusprøver

RNA-sekvensering er en metode som kan brukes til transkriptom profilering, som gir en nøyaktig måling av transkriptnivå og deres isoformer. Metoden brukes ofte til å identifisere (og kvantifisere) både sjeldne og vanlige transkripter og til å «aligne»/sammenstille sekvenserings «reads» på tvers av «spleise junctions» og detektere nye transkripter. (44)

RNA sekvensering var nylig blitt utført av noen andre i forskningsgruppen og data fra dette er brukt i masteroppgaven.

Vi ville undersøke hvilke transkripter/alternative spleisevarianter som fantes for *AIRE* og *ICOSLG* i forhold til rapporterte transkripter i to RNA-sekvenserte tymusprøver (*T02a* (gutt, 2 mnd. og 23 dager) og *T27a* (jente, 3 mnd. og 3 dager)).

RNA ble ekstrahert fra 50 mg tymusvev ved å bruke TRIzol® reagenser (Invitrogen, Carlsbad, CA) etter produsentens anbefalinger.

Prøvepreparingen før RNA sekvensering ble utført ved at 1000 ng RNA per prøve ble behandlet med kitet TruSeq Stranded Total RNA with Ribo-Zero Gold, Illumina, etter leverandørens anbefalinger.

De preparerte tymusprøvene ble sendt til «Norwegian Sequencing Center» (NSC) for RNA sekvensering på HiSeq™ 2500, Illumina. Fra sekvenseringen hadde det blitt generert 135420176 «reads» for begge trådene for T02a og 131433304 «reads» for begge trådene for T27a og med sekvens lengde 125 bp paired end for begge prøvene.

3.7 Statistiske analyser

For statistiske analyser ble i hovedsak verktøyene: Haploview 4.2 (45), PLINK (46) og GraphPad (47) brukt.

3.7.1 Kvalitetssikring av genotypedata

En kvalitetskontroll av alle genotypingsdataene ble utført ved å beregne genotypingssuksessrate (GSR) og Hardy-Weinberg likevekt (HWE) før de kunne bli tatt i bruk i videre analyser. I tillegg ble clustering for hver assayanalyse kontrollert ved en manuell inspeksjon av alle genotypingsplottene.

3.7.1.1 Genotypesuksessrate

GSR kan brukes som en kvalitetskontroll for å se på hvor stor andel av prøver/assays som har blitt suksessfullt genotypet. Det er viktig med høy genotypesuksessrate for å hindre at det har blitt skjevhet i materialet pga. at noen genotyper kan være vanskeligere og genotype enn andre.

Det ble satt en grense på minimum 95 % GSR for andel prøver genotypet for hvert enkelt SNP assay. SNPassayer med GSR lavere enn dette ble vurdert som upålitelige og ekskludert.

For kontroll av hvor mange SNPer som ble genotypet for hver enkelt prøve ble en GSR grense på minimum 80 % satt, noe som betyr at prøver som feilet på mer enn 20 % av SNPassayene ble ekskludert.

3.7.1.2 Hardy Weinberg likevekt

HWE prinsippet forteller at allel og genotype frekvensene i en populasjon er i likevekt fra generasjon til generasjon under visse forutsetninger. Forstyrrende faktorer kan for eksempel være nye mutasjoner, ikke-tilfeldig partnervalg, seleksjon eller migrasjon. Ved et avvik fra HWE kan det tyde på at det er seleksjon for spesifikke genotyper. De fleste polymorfismer i genomet vårt vil være i HWE til tross for at forutsetningen ikke er fullstendig oppfylt i populasjonen. Ved

komplekse sykdommer som RA kan man vanligvis forvente at SNPene som undersøkes er i HWE da effekten av risikovariantene er så lav at det ikke medfører seleksjon. Et avvik fra HWE kan derfor indikere problemer med genotyperesultatene som for eksempel at en del prøver fra en spesiell genotype har blitt feiltypet. Avvik fra HWE ble satt til $p<0,05$. Det statistiske analyseverktøyet PLINK ble brukt til å utføre kvalitetskontrollen og er nærmere beskrevet i kap. 3.8.

3.7.2 Koblingsulikevekt

Det bioinformatiske verktøyet Haploview 4.2 ble brukt på kontrollmaterialet genotypet i denne studien for å kartlegge koblingsulikevekt (både D' og r^2) i vår populasjon, se kapittel 3.8 for nærmere verktøysbeskrivelse.

3.7.3 Assosiasjonsanalyse

Vi ønsket å undersøke om de selekterte SNPene var assosiert med sykdom i prøvematerialet vårt av RA pasienter og kontroller. Det var også av interesse å se på om det var noen sykdomsassosiasjon i ulike grupper av RA basert på tilstedeværelsen av sirkulerende antistoff (ACPA+ og ACPA-) mot kontroller og om det var noen forskjeller i sykdomsassosiasjon mellom gruppene ved å sammenlikne ACPA + og ACPA -.

Det ble utført både kjikvadrat test og Fishers eksakt test som begge er standard assosiasjonstester for pasient/kontroller.

Assosiasjonstestene ble brukt til å sammenlikne allelfrekvenser mellom pasienter og kontroller og undersøke om disse frekvensene var signifikant forskjellige. Testen kjøres for SNPassay med prøver genotypet for alle de tre genotypevariantene. OR og 95 % konfidensintervallet (CI) ble også beregnet.

Fishers eksakt test er en statistisk signifikans test som ser på assosiasjon når det er observert mindre enn 5 prøver av et spesifikt allel eller genotype. Denne testen ble utført på to av SNPassayene i ACPA- RA prøvemateriale mot kontroller fordi det var for noen av SNPassayene

observert bare 5 prøver/varianter av en spesifikk genotype (rs760426) og bare to av tre mulige genotypevarianter (rs2075876).

Genetiske modeller for å se på om risikovarianter følger et dominant eller recessivt mønster ble også utført. Kjikvadrattester ble benyttet med 2 frihetsgrader (2 df) for genotypeassosiasjonstest og 1 df for recessiv og dominant modell. For trend test eller dosemodell ble «Cochran-Armitage trend test» brukt. Signifikansnivået for alle testene ble satt til $p<0,05$ siden genregionen allerede har vist assosiasjon til RA i tidligere studier.

Det statistiske analyseverktøyet PLINK ble brukt til å utføre alle analysene og er nærmere beskrevet i kap. 3.8. Felles for alle disse testene i PLINK er at de tester bare en fenotype av gangen.

3.7.3.1 Haplotypeanalyse

Vi ville undersøke haplotypefrekvensen i genotypingsdataene av RA pasienter mot kontroller til noen av de SNPene vi fant assosiert med RA for å se hvilke haplotyper av risiko-allelene som var assosiert med sykdom. Vi ønsket å undersøke om en av SNPene synes å styre assosiasjonen eller om haplotypeanalysen pekte mot at en annen utypet SNP var den primære SNPen assosiert til sykdom.

Haplotypeanalyse av assosierte SNPer ($p<0,05$) ble utført i UNPHASED (v.3.1.7), for nærmere beskrivelse av programvaren se kap. 3.8.

3.7.3.2 Shared epitope (SE)

SE alleler på *DRB1* er etablerte og sterke risikovarianter for RA. Vi ville undersøke om det var samspillseffekter mellom SE og *AIRE*. Til dette ble stratifiserte analyser for SNP assosiasjonsanalysene gjennomført. Data for SE forelå basert på tidligere Sangersekvensering av ekson 2 for HLA-DRB1. Individer som hadde minst en kopi av (DRB1*0101, 0102, 0401, 0404, 0405, 0408, 1001 eller 1402) var klassifisert som SE positive. PLINK ble brukt som statistisk verktøy for å utføre disse analysene, se kap. 3.8.

3.7.4 eQTL analyse av SNPene og uttrykk av *AIRE* eller *ICOSLG*

For å kunne se etter assosiasjon mellom de selekterte SNPene og uttrykk av *AIRE* og *ICOSLG* ble genotypen til 42 tymusprøver for hver av de seks SNPene korrelert med genuttrykket for de ulike tymusprøvene fra microarray (Illumina HumanWG-6 v3 array). Om noen av SNP genotypene viser signifikant assosiasjon ($p<0,05$) med genuttrykket av *AIRE* eller *ICOSLG* kan det bety at de er eQTL (expression Quantitative Trait Loci) i tymus. Hver av de seks SNPene ble analysert mot hver av de fire genekspresjons probene.

Genuttrykket (ved hver av de 4 probene) og genotypen til de 42 tymusprøvene i hver av de seks selekterte SNPene ble lagt inn og analysert i det statistiske programmet GraphPad Prism 6.0 for å utføre assosiasjonsanalysene. GraphPadPrism 6.0 er et statistikk program som brukes til å analysere og presentere vitenskapelige data. Det ble opprinnelig designet for eksperimentelle biologer på medisinskoler og spesielt for medisin firmaer innen farmakologi og fysiologi.

Variansanalysen Kruskal-Wallis test (ikke-parametrisk) ble utført på SNPene med tre ulike genotypevarianter (for eksempel AA, AG, GG), mens analysen Mann-Whitney U-test (ikke-parametrisk) ble utført på SNPene med bare to ulike genotyper tilstede i tymusmaterialet. En p-verdi på $p<0,05$ ble satt som grense for at resultatet fra testene skulle være signifikant.

Mann-Whitney U-test er en ikke-parametrisk test som sammenlikner to uparede grupper. I Prism blir først alle verdiene rangert fra lav til høy, hvor hvilken gruppe hver verdi egentlig hører til ikke tas i betraktning. De laveste tallene blir rangert som 1, mens de store rangeres som n, hvor n er totalt antall verdier i de to gruppene. Så beregnes gjennomsnittet i hver gruppe. Hvis middelverdiene av rangeringene i de to gruppene er veldig ulike, vil p-verdien være liten. (Tester om fordelingen av rangering er forskjellig.).

Kruskal-Wallis test er en ikke-parametrisk enveis varians analyse, som baserer seg på samlet rangordning og gruppevise rangberegninger. Tre eller flere umatchede grupper blir sammenliknet i denne testen, hvor alle verdiene først blir rangert fra lavt til høyt, og hvilken gruppe verdiene hører til blir ikke tatt i betraktning. De minste tallene blir rangert til 1, mens de største tallene blir rangert til N, hvor N er totalt antall verdier i alle gruppene. Forskjellene mellom de rangerte verdiene slås sammen slik at en enkel verdi kalt Kruskal-Wallis statistikk

kan beregnes. En høy Kruskal-Wallis statistikk viser til et stort avvik mellom de rangerte verdiene. Testen forutsetter at den overordnede fordelingen av observasjoner er ganske lik mellom gruppene, med unntak av medianverdien i gruppen.

3.7.5 RNA sekvensering av fire tymusprøver

Etter RNA sekvenseringen ble det utført en kvalitetskontroll og en kartlegging av «readsene». Disse stegene ble utført av forskningsgruppens bioinformatiker, som også utførte et «alignment» steg hvor sekvenseringsdataene ble «alignet» opp mot det humane referanse genom, Hg38 (med kromosom 21 som fokusområde).

Oppstillingsverktøyet Bowtie 2 ble brukt til å «aligne» opp alle «readsene» til referanse genomet og derfra ble dataene lagt over i visualiseringsverktøyet IGV for å se etter alternative spleisevarianter, for nærmere beskrivelse av disse verktøyene se kap. 3.8.

3.8 Bioinformatiske og statistiske programmer

Under følger en oversikt over databaser, programvarer og bioinformatiske- og statistiske verktøy brukt i denne masteroppgaven.

HapMap

HapMap er en katalog med oversikt over vanlige genetiske varianter. Katalogen beskriver allelfrekvenser i ulike populasjoner og koblingsulikevektsmønsteret mellom SNPene. Det internasjonale HapMap prosjektet er et samarbeid mellom forskere fra ulike land og hensikten var å lage en database for å gi informasjon som andre forskere kan bruke i genetiske assosiasjonsstudier. Katalogen ble brukt som kilde for SNP seleksjonen for å se på allelfrekvensen og koblingsulikevektsmønsteret av de selekterte SNPene i europeere for design av noen av GWAS genotypingsarrayene. (http://hapmap.ncbi.nlm.nih.gov/)

1000Genomes

1000 Genomes prosjektet er et internasjonalt forskningsprosjekt som hadde som mål å sekvensere genomene til et stort antall mennesker (mer enn 1000) fra ulike etniske grupper for å etablere en omfattende ressurs over genetiske variasjoner i mennesket. Prosjektet hadde også som mål å bidra til en bedre forståelse av rollen til genetiske variasjoner i sykdom, historie og evolusjon. 1000 Genomes katalogen er i dag en av de mest detaljerte, offentlige og kostnadsfrie katalogene på genetiske variasjoner i mennesket. Data fra katalogen ble brukt under SNP seleksjonen til å undersøke koblingsulikevekt mellom de ulike SNPene i både asiatisk og europeisk befolkning. Den ble også brukt til å finne allelfrekvensen av noen SNPer i seleksjonen som det ikke fantes data på i HapMap. (http://www.1000genomes.org/home)

Snap Proxy Search

Snap Proxy Search er et databaseverktøy som brukes for å finne «proxy» SNPer basert på koblingsulikevekt, fysisk avstand (og tilstedeværelse på utvalgte genotyping arrayer). Her blir parvis koblingsulikevekt kalkulert basert på fasede genotyping data fra HapMap prosjektet og

1000 Genomes prosjektet. Verktøyet ble brukt i SNP seleksjonen til å se på koblingsulikevekt hos SNPer i japanere og europeere. (https://www.broadinstitute.org/mpg/snap/ldsearch.php)

dbSNP

dbSNP er en database som katalogiserer korte variasjoner (SNPer, korte nukleotid insersjoner og delesjoner, korte tandem «repeats» og mikrosatelitter) i nukleotidsekvenser fra et stort spekter av organismer. Databasen ble brukt i SNP seleksjonen til å se på allelfrekvensen hos SNPene i europeere. (http://www.ncbi.nlm.nih.gov/SNP/)

PLINK

PLINK er et «open-source» verktøy for utførelse av hel-genom assosiasjonsanalyser. Programmet fokuserer på analyse av fenotype/genotype data for case/kontroll studier. I denne oppgaven ble PLINK brukt til både å utføre en kvalitetskontroll av alle genotypedataene og til å utføre assosiasjonsanalysene. (http://pngu.mgh.harvard.edu/~purcell/plink/)

Haploview

Haploview 4.2 er en bioinformatisk programvare som er designet for å analysere og visualisere koblingsulikevekt mønster i genetiske data. Programvaren kan også utføre assosiasjonsanalyser og estimere haplotype frekvenser. Programmet ble brukt for å se på koblingsulikevekt mønsteret hos de selekterte SNPene i norsk populasjon. (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview)

Unphased

Unphased er en statistisk programvare til å utføre genetiske assosiasjonsanalyser i familier og ubeslektede individer. Noen av de mest vanlige utførte analysene er koblingsulikevektstester, globale og individuelle tester for haplotyper, tester som ser på gen-gen interaksjoner og sammenlikninger av risiko mellom ulike haplotyper. Programvaren ble brukt til å utføre en haplotypeanalyse av gitte SNPer i ulike grupper. (https://dsgweb.wustl.edu/aldi/software/manuals/unphased/Unphased_manual.pdf)

Bowtie 2

Bowtie 2 er et raskt og minne-effektivt verktøy for og "aligne" sekvens-"reads" til lange referanse sekvenser (f.eks. human genom). Verktøyet støtter paired-end alignment. Bowtie 2 ble brukt til og "aligne" RNA sekvenserings "reads" for de to tymusprøvene til menneske referanse genom (Hg38)). (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml)

IGV

"Integrative Genomics Viewer" (IGV), er et visualiseringsverktøy som brukes til utforskning av store, integrerte genomiske datasett. Programvaren støtter et bredt spekter av datatyper som f.eks. array-baserte og neste-generasjons sekvenseringsdata og genomiske annoteringer. Verktøyet ble brukt til å se på resultatet av «aligningen» av RNA sekvenserings dataene som Bowtie 2 hadde utført til referanse genomet, for å se etter nye transkripter. (https://www.broadinstitute.org/igv/)

RegulomeDB

For å undersøke om de seks selekterte SNPene var regulatoriske ble det gjort et søk i RegulomeDB versjon 1.1 den 26.03.15.

RegulomeDB er en database som annoterer SNPer med kjente og forutsatte regulatoriske elementer i intergene regioner av det humane genomet (Hg19). Databasen scorer regulerende evidens for søke SNP vha. en skala som kalles RegulomeDB score, som går fra 1 til 6. 1 betyr at SNPen er veldig regulerende/eQTL SNP og 6 betyr lite regulerende. Dataene brukt i RegulomeDB er hentet fra offentlige datasett som fra GEO, ENCODE prosjektet og publisert litteratur. (http://www.regulomedb.org/)

4. Resultater

For å undersøke om *AIRE* genet eller kromosomområdet rundt er assosiert med RA i vår befolkning, ble først en SNP seleksjon utført for å finne et relevant sett SNPer for denne studien ved hjelp av databaser og bioinformatiske verktøy. Det ble testet om *AIRE* og/eller *ICOSLG* kunne assosieres med sykdom og subfenotyper av RA ved å utføre assosiasjonsanalyser ved bruk av spesialiserte genetiske statistiske verktøy. Deretter ble de utvalgte polymorfismene testet med hensyn til mulig regulatorisk funksjon ved å se på korrelasjon med genekspresjon av *AIRE* og *ICOSLG* gjennom eQTL-analyser. Transkripter av *AIRE* og *ICOSLG* i tymus ble også kartlagt gjennom RNA sekvenseringsdata.

4.1 SNP seleksjon

Søket på kromosomområdet 21q22.3 i GWAS katalogen ga 47 studier med signifikant assosierte SNPer. Fenotypene for disse 47 studiene ble gjennomgått og de 11 SNPene som var assosiert med immunrelevante sykdommer ble inkludert for videre vurdering (tabell 2).

Immunobase søket resulterte i to SNPer som allerede var blitt plukket ut i GWAS katalogsøket, rs1893592 og rs11203203, og dermed ble ikke dette søkeverktøyet brukt videre i seleksjonen.

Tabell 2: SNPer rapportert per 09.01.15 assosiert med immunrelaterte sykdommer på kromosom 21q22.3 fra GWAS katalogen. Lokalisasjon til SNPene er hentet fra UCSC (Hg38) og frekvensene er hentet fra GWAS katalogen.

Sykdom	**Gen**	**SNP**	**Frekvens**	**Lokalisasjon**	**Populasjon**	**Avstand fra AIRE (rs2075876) (bp)**
Revmatoid artritt	*UBASH3A*	rs1893592	0,73	Intron	Europeere og øst-asiatere	1854313
Kronisk inflammatorisk tarmsykdom	*ICOSLG*	rs7282490	0,391	Intergen	Europeere	93412
Ulcerøs kolitt	*ICOSLG*	rs2838519	0,39	Intergen	Europeere	94130
Crohn's sykdom	*ICOSLG*	rs2838519	0,391	Intergen	Europeere	94130
Revmatoid artritt	*UBASH3A*	rs11203203	0,37	Intron	Europeere	1873193
Cøliaki og Revmatoid artritt	*UBASH3A*	rs11203203	-	Intron	Europeere	1873193
Vitiligo	*UBASH3A*	rs11203203	0,373	Intron	Europeere	1873193
Cøliaki	*ICOSLG*	rs4819388	0,72	Intron	Europeere	61732
Type 1 diabetes	*UBASH3A*	rs11203203	-	Intron	Europeere	1873193
Type 1 diabetes	*UBASH3A*	rs9976767	-	Intron	Europeere	1872989
Crohn's sykdom	*ICOSLG*	rs762421	0,39	Intergen	Europeere	93592

For å unngå å selektere SNPer fra et annet gen som vil gi veldig lav koblingsulikevekt med SNPer i *AIRE* genet ble det sett på autoimmune risikogener for et intervall på 500 kb oppstrøms og nedstrøms for *AIRE* genet i Hg38. Risikogener innenfor intervallet ble inkludert for videre vurdering. I sammenligning med tabell 2 fra GWAS søket ble genet *UBASH3A* ekskludert på grunn av at det lå utenfor det gitte intervallet, mens genet *ICOSLG* ble inkludert.

Snap Proxy Search ble brukt til å kartlegge koblingsulikevektsmønsteret mellom de fire SNPene fra GWAS søket som lå ved *ICOSLG* genet for å kunne ekskludere SNPer i sterk koblingsulikevekt ($r^2 \geq 0,8$) med hverandre da de gir tilnærmet samme assosiasjon pga. sterk korrelasjon. Det viste seg at rs2838519, rs762421 og rs7282490 var i koblingsulikevekt med hverandre ($r^2 > 0,90$), de parvise koblingsulikevektverdiene var: $r^2 = 0,967$ mellom rs7282490 og rs2838519, og $r^2 = 0,904$ mellom rs7282490 og rs762421. Dermed ble rs7282490 tilfeldig valgt for å representere de tre SNPene. Den fjerde SNPen, rs4819388, ble inkludert da den ikke var i koblingsulikevekt ($r^2 < 0,8$) med noen av de andre *ICOSLG* SNPene.

Neste seleksjonstrinn var å se etter SNPer som var i høy koblingsulikevekt ($r^2 \geq 0,8$) i asiatisk populasjon med SNPene funnet assosiert med RA i japansk befolkning (N=8), men med lav koblingsulikevekt ($r^2 < 0,8$) i europeere (N=2) for eventuelt å kunne forklare diskrepansen mellom

assosiasjonsstudiene i de to populasjonene (tabell 3). SNPene rs9974362 og rs2075877 ble ekskludert på grunnlag av at det ikke finnes informasjon om disse for europeere i HapMap og 1000Genomes datasettene. Dette ga oss grunn til å tro at SNPene ikke er polymorfe i europeere.

Tabell 3: SNPer i koblingsulikevekt ($r^2 \geq 0,8$) med de to japanske sykdoms SNPene, rs2075876 og rs760426, hos asiatere og deres koblingsulikevekt hos europeere.

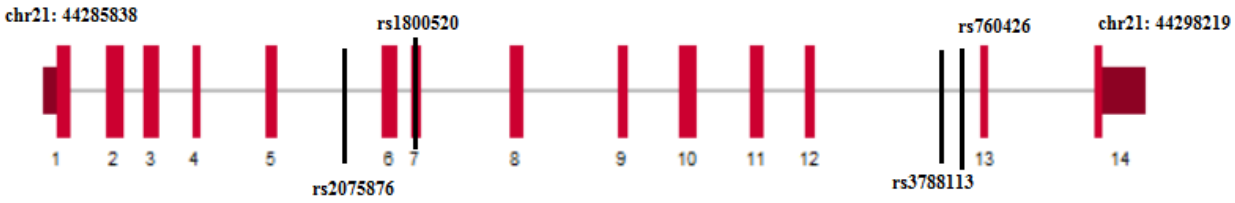
					JAPAN	**EUROPA**	
SNP	**Proxy**	**Distanse**	**Kromosom**	**Koordinat_HG18**	**r^2**	**r^2**	**Ekskludert/ Inkludert**
rs2075876	rs2075875	12	21	44533569	1,000	1,000	Ekskludert
rs2075876	rs62220374	18	21	44533599	0,967	1,000	Ekskludert
rs2075876	rs9974362	131	21	44533450	0,935	-	Ekskludert
rs2075876	rs2075877	899	21	44534480	0,803	-	Ekskludert
rs2075876	rs2073610	1375	21	44534956	0,803	0,001	Inkludert
rs760426	rs3788113	194	21	44540048	0,967	0,673	Inkludert
rs760426	rs1800526	1558	21	44538684	0,875	0,898	Ekskludert
rs760426	rs760427	177	21	44540419	0,846	1,000	Ekskludert

dbSNP ble brukt som siste seleksjonsverktøy til å finne allelfrekvensen til de selekterte SNPene i europeisk befolkning. Da vi ikke har styrke til å detektere assosiasjon til lavfrekvente alleler, ble bare SNPer hvor det mest sjeldne allelet hadde en frekvens over 5 % inkludert for genotyping. SNP, rs2073610, allel C hadde en frekvens på 1,7 % og ble dermed ekskludert. Se tabell 4 for de utvalgte SNPene inkludert i studien og figur 10 og 11 for de selekterte SNPene i *AIRE*- og *ICOSLG* genet.

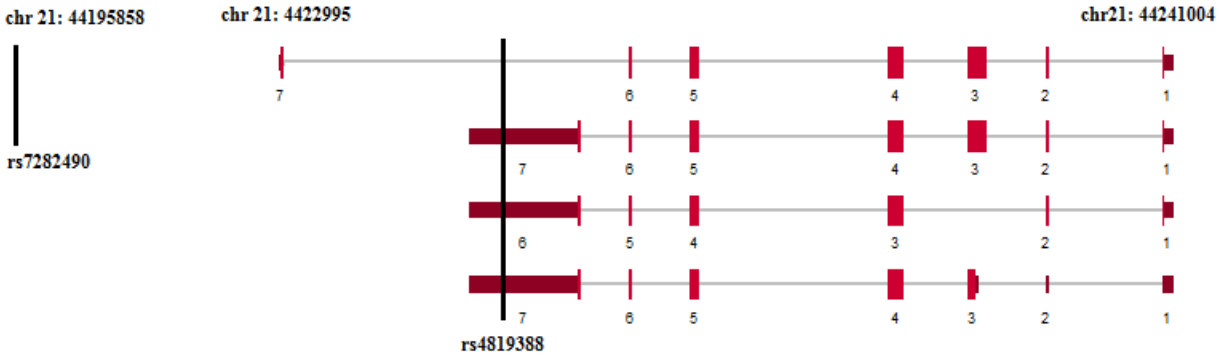
Tabell 4: Oversikt over de utvalgte SNPene som skal brukes videre i assosiasjonsanalysen.

Gen	**SNP**	**Lokalisasjon**	**MAF* (%)**	**Minst frekvent allel**	**Mest frekvent allel**	**Rapportert sykdomsassosiasjon**	**Referanse**
AIRE	rs2075876	Intron	13,3	A	G	Revmatoid artritt	(30)
AIRE	rs760426	Intron	15,9	G	A	Revmatoid artritt	(30)
AIRE	rs3788113	Intron	25,0	G **	A	-	-
ICOSLG	rs7282490	Intergen	42,9	G	A	Kronisk inflammatorisk tarmsykdom	(48)
AIRE	rs1800520	Ekson 7	7,5	G **	C	Revmatoid artritt	(30)
ICOSLG	rs4819388	Intron	24, 8	T	C	Cøliaki	(49)

*Hentet fra dbSNP/HapMap (50)
**Fra 1000 Genomes, pilot 1 (resten fra HapMap).



Figur 10: Fire av de selekterte SNPene i *AIRE* genet.(røde fylte streker: ekson, svarte tynne streker: SNP, tynn grå strek: intron). (39)



Figur 11: To av de selekterte SNPene som hører til *ICOSLG* genet. .(røde fylte streker: ekson, svarte tynne streker: SNP, tynn grå strek: intron) (51)

4.2 Sykdomsassosiasjon

4.2.1 Kvalitetskontroll av genotypedata

GSR til fem av de seks SNPene var på 98 % eller bedre (tabell 5), noe som betyr at resultatene kunne brukes i videre analyser. Den sjette SNPen (rs1800520) hadde en GSR < 95 % og ble ekskludert for videre analyser.

Tabell 5: Oversikt over genotypesuksessrate for hver av de seks SNPassayene.

SNP ID	Ant. prøver forsøkt genotypet	Ant. ind. uten SNP genotyperesultat	GSR (%)
rs7282490	1765	19	98,9
rs4819388	1765	25	98,6
rs2075876	1765	6	99,7
rs3788113	1765	30	98,3
rs760426	1765	7	99,6
rs1800520	1506	299	80,2

16 pasient- og kontrollprøver ble ekskludert pga. at de hadde en GSR< 80 % for de fem SNP assayene som ble suksessfullt genotypet.

TaqMan assayet for rs1800520 ble bare genotypet på 1506 individer på grunn av manglende assayreagenser. Det passerte ikke kvalitetskontrollen fordi den både hadde en GSR < 95 % og HWE<0,05.

HWE analysen viste at SNPene, rs2075876, rs760426, rs3788113, rs4819388 og rs7282490 var i HWE >0,05 i testen av kontroller og pasienter sammen (alle) og i testene av kontroller (friske) og pasienter (syke) hver for seg selv. Ved monogene tilstander kan pasientgrupper vise avvik i HWE, men RA er en multifaktoriell sykdom hvor risikoallelet ikke har så stor effekt på sykdom at det lager skjevhet i populasjonen. Det er derfor forventet at pasientgenotypene også skal være i HWE. Dermed kunne disse fem SNPassayene inkluderes for videre analyse (tabell 6).

Tabell 6: Oversikt over HWE (p-verdi) for de seks SNP(assayene) i totalmaterialet av både pasienter og kontroller (alle), bare i kontrollgruppen (frisk) og bare i pasientgruppen (syk).

SNP ID	Test	Minor allel	Major allel	Genotypetellinger*	Observert heterosygositet	Forventet heterosygositet	HW p-verdi
rs7282490	ALLE	G	A	264/845/637	0,484	0,477	0,581
rs7282490	SYK	G	A	147/438/338	0,475	0,479	0,784
rs7282490	FRISK	G	A	117/407/299	0,495	0,476	0,272
rs4819388	ALLE	T	C	124/670/946	0,385	0,388	0,711
rs4819388	SYK	T	C	68/347/503	0,378	0,388	0,444
rs4819388	FRISK	T	C	56/323/443	0,393	0,389	0,858
rs2075876	ALLE	A	G	16/345/1398	0,196	0,191	0,382
rs2075876	SYK	A	G	10/176/746	0,189	0,188	1,000
rs2075876	FRISK	A	G	6/169/652	0,204	0,195	0,211
rs3788113	ALLE	G	A	62/566/1107	0,326	0,319	0,366
rs3788113	SYK	G	A	27/287/605	0,312	0,302	0,382
rs3788113	FRISK	G	A	35/279/502	0,342	0,336	0,678
rs760426	ALLE	G	A	22/374/1362	0,213	0,210	0,571
rs760426	SYK	G	A	11/176/746	0,189	0,190	0,863
rs760426	FRISK	G	A	11/198/616	0,240	0,231	0,364
rs1800520	ALLE	G	C	4/238/965	0,197	0,183	0,004
rs1800520	SYK	G	C	3/144/550	0,207	0,192	0,048
rs1800520	FRISK	G	C	1/94/415	0,184	0,171	0,071

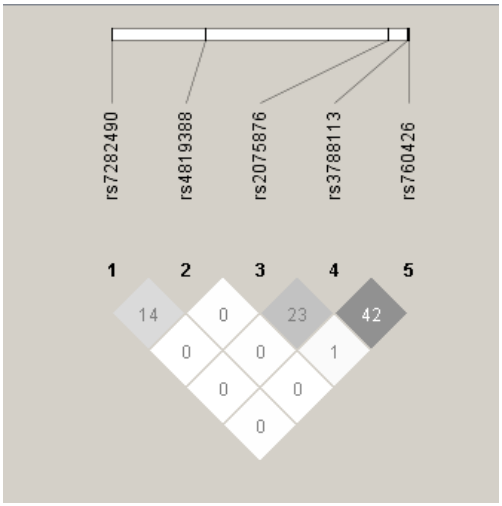
* Homozygot for minst frekvente allel/ heterozygot/ homozygot for mest frekvente allel

4.2.2 Koblingsulikevektanalyser

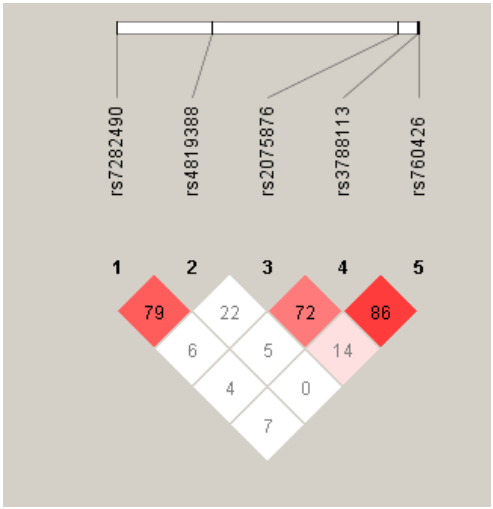
Vi ønsket å få et overblikk over koblingsulikevekten mellom de testede SNPene i genotypingsdataene for norsk befolkning derfor ble det utført en koblingsulikevektanalyse. Siden risikogener for autoimmune sykdommer viser stor grad av genetisk overlapp ble det også sett på om *ICOSLG* variantene som er rapportert for IBD (rs7282490) og cøliaki (rs4819388) kunne tilskrives *AIRE* SNPer som er rapportert for RA. Koblingsulikevektsanalysen (figur 12 og 13) av de genotypede kontrolldataene for de fem SNPene som passerte kvalitetskontrollen viste at det ikke var noen koblingsulikevekt ($r^2 = 0$), mellom de to *ICOSLG* SNPene og de tre *AIRE* SNPene, noe som indikerer at SNP allelene er i fullstendig likevekt og dette gjør at *ICOSLG* SNPene og *AIRE* SNPene ikke kan være samme assosiasjon. Var også lave koblingsulikevekt-verdier *AIRE* SNPene i mellom ($r^2 \leq 0{,}42$) og *ICOSLG* SNPene i mellom ($r^2 = 0{,}14$) som viser at det er lav korrelasjon mellom SNPene. Koblingsulikevektmønstrene (D` og i hovedsak r^2) viser at de selekterte SNPene har gitt unik informasjon om assosiasjon.

Det ble også undersøkt om vår hypotese om at SNPen vi hadde selektert med $r^2 > 0{,}8$ i asiatisk befolkning, faktisk hadde lavere koblingsulikevekt i norsk befolkning. Rs3788113 viste seg å ha lavere koblingsulikevekt ($r^2 = 0{,}42$) i norsk befolkning enn i japansk befolkning ($r^2 = 0{,}967$) med rs760426 som forventet.

D` viste høyere koblingsulikevekt-verdier for nesten alle SNPene enn r^2, noe som er vanlig, da r^2 er et mer stringent mål. Det var lav D' mellom SNPene i de to genene, men innbyrdes i *ICOSLG* og *AIRE* var D' relativt markant D'>0,7, unntatt mellom AIRE SNPene rs2075876 og rs760426 (D'=0,14). Dette kan tyde på at det er haplotypestrukturer i genene.



Figur 12: Oversikt over koblingsulikevekt (r^2) mellom fem av de selekterte SNPene i *AIRE* og *ICOSLG* for friske kontrollprøver i norsk befolkning.



Figur 13: Oversikt over koblingsulikevekt (D`) mellom fem av de selekterte SNPene i *AIRE* og *ICOSLG* for friske kontrollprøver i norsk befolkning.

4.2.3 Assosiasjonsanalyse inkludert stratifisert for ACPA og SE

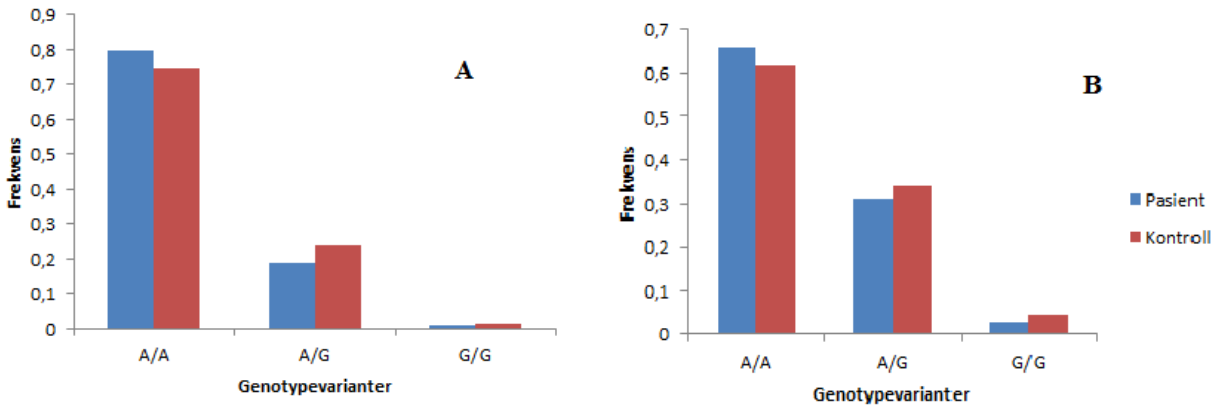
En assosiasjonstest ble utført i PLINK for de fem SNPene som passerte kvalitetskontrollen for å sammenlikne allelfrekvensene mellom pasienter og kontroller og se om en allelvariant var signifikant assosiert med sykdom (tabell 7). SNPene rs3788113 og rs760426 i *AIRE* genet hadde signifikante p-verdier <0,05, noe som kan tyde på at SNPene kan påvirke risikoen for RA i norsk befolkning. For SNP rs3788113 var frekvensen av det minst frekvente («minor») allelet (G) lavere hos RA pasienter enn hos kontroller med hhv. 18,6 % mot 21,4 %. SNP rs760426 viste også lavere allelfrekvens av «minor» allelet (G) hos RA pasienter enn hos kontroller, hhv.10,7 % mot 13,4 %. Disse to allelene viser dermed redusert risiko for å utvikle RA (rs3788113*G: OR(95 % CI) = 0,84 (0,71-0,99) og rs760426*G: OR(95 % CI) = 0,77 (0,63-0,95)). Og de allelene som er mest frekvente («major») kan dermed trolig være assosiert med økt risiko for RA. De resterende SNPene viste ikke signifikant assosiasjon med RA i vår kohort.

Tabell 7: Assosiasjonsanalyse av SNPene i *AIRE/ICOSLG* med p-verdi, odds ratio og konfidensintervall.

SNP ID	Gen	Minor allel	Frekvens RA	Frekvens ktr.	OR	95 % CI	p-verdi
rs7282490	*ICOSLG*	G	0,397	0,388	1,04	(0,90-1,19)	0,61
rs4819388	*ICOSLG*	T	0,263	0,265	0,99	(0,85-1,15)	0,92
rs2075876	*AIRE*	A	0,106	0,109	0,96	(0,78-1,20)	0,74
rs3788113	*AIRE*	G	0,186	0,214	0,84	(0,71-0,99)	0,04
rs760426	*AIRE*	G	0,107	0,134	0,77	(0,63-0,95)	0,01

En assosiasjonsanalyse med genetisk modell ble utført for å undersøke om allelvariantene for de ulike SNPene opptrer dominant eller recessivt (vedlegg 2). Bare rs760426 viste signifikant assosiasjon på genotypenivå (p=0,03), og assosiasjonen ga signifikant utslag på dominant modell (p=0,008). Genotypefordelingen til rs760426 (figur 14A) viser at det beskyttende G-allelet følger en dominant modell ved å gi redusert risiko ved bærerskap av allelet.

Den andre assosierte SNPen, rs3788113, viste assosiasjon med trend test (p=0,04), noe som kan tyde på at risikoen forbundet med SNPen er allel-doseavhengig. Genotypefordelingen til rs3788113 (figur 14B) viser at den prosentvis største forskjellen mellom pasienter og kontroller for risikoallelet er når det er tilstede i dobbelt dose (G/G) i forhold til i enkelt dose (A/G).



Figur 14: Genotypefordelingen til A) rs760426 og B) rs3788113 for pasienter og kontroller i genotypedataene for de fem selekterte SNPene.

4.2.3.1 ACPA + vs. Kontroll

RA pasienter kan klassifiseres som ACPA + eller ACPA -, og genetiske assosiasjoner er ofte ulike i de to gruppene. Derfor var det av interesse å se på data stratifisert for ACPA status for å undersøke om SNP assosiasjonene var begrenset til en spesifikk pasientsubgruppe.

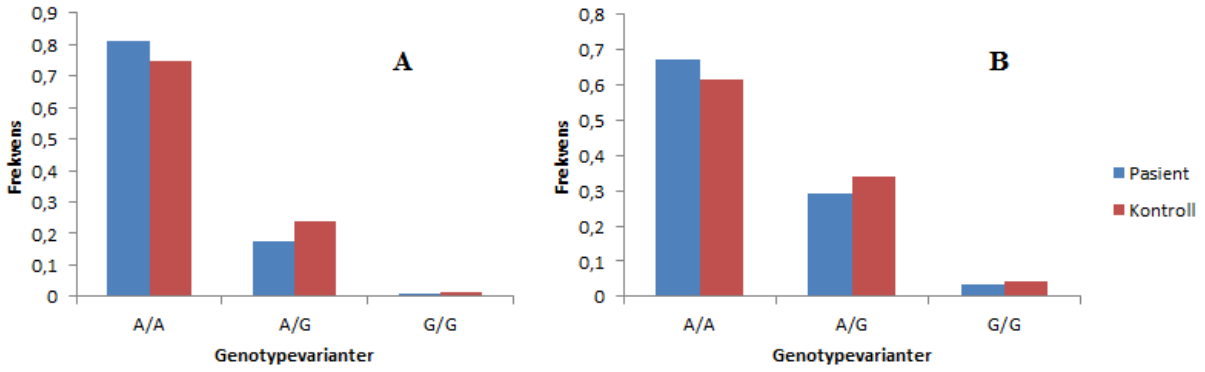
Assosiasjonsanalyse av de ACPA+ RA pasientene (539 stk.) mot kontroller for de fem SNPene viste at de samme SNP assosiasjonene som i totalanalysen, rs3788113 og rs760426, ga signifikante p-verdier ($p<0,05$) (tabell 8) med antydning til enda mer uttalt assosiasjon da signifikansen ble mer uttalt til tross for lavere styrke i den stratifiserte analysen. MAF til rs3788113 var lavere hos ACPA+ RA pasienter (18,0 %) enn hos kontroller (21,4 %), noe den også var for rs760426 med (10,0 %) for RA pasienter og (13,3 %) for kontroller. Disse to allelene viser dermed redusert risiko for å utvikle ACPA+ RA (rs3788113*G: OR(95 % CI) = 0,81 (0,66-0,98) og rs760426*G: OR(95 % CI) = 0,72 (0,56-0,92)). De resterende SNPene viste ikke signifikant assosiasjon med RA i vår kohort.

Tabell 8: Assosiasjonsanalyse av SNPene i *AIRE/ICOSLG* hos ACPA+ individer mot kontroller med p-verdi, odds ratio og konfidensintervall.

SNP ID	Gen	Minor allel	Frekvens ACPA + RA	Frekvens ktr.	OR	95 % CI	p-verdi
rs7282490	*ICOSLG*	G	0,405	0,389	1,07	(0,91-1,25)	0,42
rs4819388	*ICOSLG*	T	0,255	0,265	0,95	(0,80-1,13)	0,57
rs2075876	*AIRE*	A	0,102	0,109	0,92	(0,72-1,18)	0,52
rs3788113	*AIRE*	G	0,180	0,214	0,81	(0,66-0,98)	0,03
rs760426	*AIRE*	G	0,100	0,133	0,72	(0,56-0,92)	0,01

En assosiasjonsanalyse med genetisk modell ble utført for å undersøke om allelvariantene for de ulike SNPene opptrer dominant eller recessivt (vedlegg 3). Bare rs760426 viste signifikant assosiasjon på genotypenivå (p=0,02), og assosiasjonen ga signifikant utslag på dominant modell (p=0,005). Genotypefordelingen til rs760426 (figur 15A) bekrefter også her at G-allelet har en assosiasjon som følger en dominant modell.

Den andre assosierte SNPen, rs3788113, viste assosiasjon med trend test (p=0,03), som i totalanalysen, men ga i tillegg signifikant utslag på dominant modell (p=0,03). Genotypefordelingen til rs3788113 (figur 15B) støtter ikke i like stor grad en doseavhengighet som i totalanalysen, noe som kan forklare hvorfor den dominante modellen ble signifikant i denne stratifiserte analysen.



Figur 15: Genotypefordelingen til A) rs760426 og B) rs3788113 for ACPA + RA pasienter og kontroller i genotypedataene for de fem selekterte SNPene.

4.2.3.2 ACPA – vs. Kontroll

Assosiasjonsanalysen for ACPA- RA pasienter (394 stk.) mot kontroller for de fem selekterte SNPene viste ikke signifikant assosiasjon med RA i vår kohort (alle p>0,3) (tabell 9). Heller ikke assosiasjonsanalyser med ulike genetiske modeller viste noen signifikante assosiasjoner (vedlegg 4).

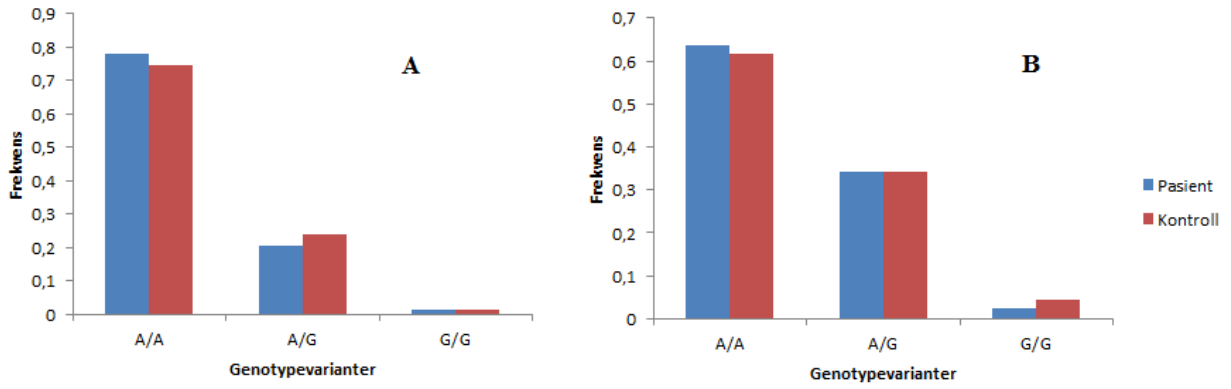
MAF mellom ACPA- RA og kontroller er ganske lik for de fleste SNPene, men for SNPene rs760426 og rs3788113 har kontrollene også her litt høyere frekvens enn pasientene som for de tidligere analysene, men forskjellene er ikke statistisk signifikante.

Tabell 9: Assosiasjonsanalyse av SNPene i *AIRE/ICOSLG* hos ACPA- RA mot kontroller med p-verdi, odds ratio og konfidensintervall.

SNP ID	Gen	Minor allel	Frekvens ACPA - RA	Frekvens ktr.	OR	95 % CI	p-verdi
rs7282490	*ICOSLG*	G	0,388	0,389	0,99	(0,83-1,20)	0,95
rs4819388	*ICOSLG*	T	0,280	0,265	1,08	(0,88-1,32)	0,47
rs2075876	*AIRE*	A	0,109	0,109	1,00	(0,75-1,33)	1,00*
rs3788113	*AIRE*	G	0,195	0,214	0,89	(0,71-1,11)	0,30
rs760426	*AIRE*	G	0,118	0,133	0,87	(0,66-1,15)	0,34*

*Fishers eksakt p-verdi, da disse SNPassayene ble analysert med Fishers eksakt og ikke kjikvadrat test.

Genotypefordelingen til SNPen rs760426 og rs3788113 er vist i figur 16A og B.



Figur 16: Genotypefordelingen til A) rs760426 og B) rs3788113 for ACPA - RA pasienter og kontroller i genotypedataene for de fem selekterte SNPene.

4.2.3.3 ACPA+ vs. ACPA-

Siden de genetiske assosiasjonene bare ble detektert i ACPA+ RA pasienter ble fordelingen av SNP allelfrekvenser sammenliknet mellom ACPA+ og ACPA- pasienter. Ingen av SNPene viste signifikante ($p<0,05$) forskjeller mellom de to pasientgruppene (tabell 10). MAF ser ut til å være litt lavere for SNPene i ACPA+ enn for de i ACPA- RA pasienter, bortsett fra for rs7282490 som har høyere allelfrekvens i ACPA+ enn i ACPA-.

Tabell 10: Assosiasjonsanalyse av SNPene i *AIRE/ICOSLG* hos ACPA- RA mot ACPA+ RA med p-verdi, odds ratio og konfidensintervall.

SNP ID	Gen	Minor allel	Frekvens ACPA + RA	Frekvens ACPA- RA	OR	95 % CI	p-verdi
rs7282490	*ICOSLG*	G	0,405	0,388	1,08	(0,88-1,31)	0,48
rs4819388	*ICOSLG*	T	0,255	0,28	0,88	(0,71-1,10)	0,26
rs2075876	*AIRE*	A	0,102	0,109	0,92	(0,67-1,26)	0,63*
rs3788113	*AIRE*	G	0,18	0,195	0,91	(0,71-1,16)	0,45
rs760426	*AIRE*	G	0,10	0,118	0,83	(0,61-1,13)	0,23*

*Fishers eksakt p-verdi

4.2.3.4 Haplotypeanalyse

Vi ville undersøke haplotypefrekvensen i RA pasienter mot kontroller til de to assosierte SNPene (rs760426 og rs3788113) for å se hvilke haplotyper som viste assosiasjon med sykdom, samt fordelingen av allelene fra de to assosierte SNPene på ulike haplotyper.

Haplotypeanalysen for RA pasienter mot kontroller for de to assosierte SNPene, rs3788113 og rs760426, viste signifikant assosiasjon ($p=0,05$) for at det var skjevheter i frekvensfordelingen av haplotypene (tabell 11). De to disponerende «major» allelene forekommer på samme haplotype A-A, som er signifikant ($p=0,02$) mer frekvent hos pasienter (81,5 %) enn kontroller (77,2 %). De to beskyttende «minor allelene» på de to SNPene befinner seg ofte på samme haplotype G-G som viser signifikant ($p=0,03$) redusert frekvens hos pasienter (9,4 %) mot kontroller (12,1 %). Det beskyttende allelet for rs378813*G befinner seg også ofte på haplotype med A-allelet på rs760426, men haplotypen gir da ikke signifikant beskyttelse (8,8 % hos pasienter mot 9,3 % hos kontroller).

Tabell 11: Haplotypefrekvens i RA pasienter mot kontroller til de to assosierte SNPene rs760426 og rs3788113.

Haplotype	RA (n)	Kontroll (n)	Frekvens RA	Frekvens ktr.	p-verdi
A-A	1476	1258	0,815	0,772	0,02
A-G	161	152	0,088	0,093	0,61
G-A	16	23	0,009	0,014	0,15
G-G	179	197	0,094	0,121	0,03

Det var også av interesse å undersøke haplotypefrekvensen til de assosierte SNPene i ACPA + RA pasienter opp mot kontroller for å se nærmere på om den ene pasientsubgruppen ville ha høyere frekvens av en spesiell haplotype enn kontrollene.

Haplotypeanalysen mellom ACPA + RA pasienter og kontroller for de to assosierte SNPene viser enda mer signifikant skjevfordelingen av frekvens enn i totalanalysen (p=0,02) (tabell 12). Frekvensen av den disponerende haplotypen A-A var signifikant (p=0,01) høyere hos pasienter (81,3 %) enn kontroller (77,2 %). Tilsvarende viser den beskyttende haplotype G-G signifikant (p=0,02) redusert frekvens hos pasienter (9,2 %) mot kontroller (12,1 %). Det beskyttende allelet for rs378813*G befinner seg her også ofte på haplotype med A-allelet på rs760426 (8,8 % hos pasienter mot 9,3 % hos kontroller) uten at haplotypen da er signifikant redusert. Tilstedeværelse av den beskyttende rs760426*G forekommer på haplotype med rs378813*A hos 0,7 % av pasientene mot 1,4 % av kontrollene (p=0,1).

Tabell 12: Haplotypefrekvens i genotypingsdataene av ACPA + RA pasienter mot kontroller til de to assosierte SNPene, rs760426 og rs3788113.

Haplotype	RA (n)	Kontroll (n)	Frekvens ACPA+ RA	Frekvens ktr.	p-verdier
A-A	863	1258	0,813	0,772	0,01
A-G	94	152	0,088	0,093	0,67
G-A	8	23	0,007	0,014	0,1
G-G	97	197	0,092	0,121	0,02

4.2.3.5 Shared epitope

En assosiasjonstest ble utført i PLINK på genotypingsdataene for de fem SNPene som passerte kvalitetskontrollen for en SE-negativ gruppe av pasienter/kontroller og en SE-positiv gruppe av pasienter/kontroller for å finne ut om assosiasjonen for *AIRE* ble influert av tilstedeværelse av HLA risikovarianter.

Assosiasjonsanalysen av den SE-negative gruppen med RA pasienter mot kontroller for de fem SNPene viste at ingen av SNPene hadde noen signifikant forskjell ($p<0,05$) mellom RA pasienter og kontroller (tabell 13).

Tabell 13: Assosiasjonsanalyse av SNPene i *AIRE/ICOSLG* hos SE negative RA prøver mot kontrollprøver med p-verdi, odds ratio og konfidensintervall.

SNP ID	Gen	Minor allel	Frekvens RA	Frekvens ktr.	OR	95 % CI	p-verdi
rs7282490	*ICOSLG*	G	0,387	0,396	0,96	(0,75-1,24)	0,78
rs4819388	*ICOSLG*	T	0,26	0,278	0,91	(0,69-1,21)	0,52
rs2075876	*AIRE*	A	0,113	0,119	0,94	(0,64-1,38)	0,77*
rs3788113	*AIRE*	G	0,202	0,221	0,89	(0,66-1,21)	0,47
rs760426	*AIRE*	G	0,128	0,14	0,90	(0,62-1,30)	0,58*

*Fisher eksakt p-verdi

Assosiasjonsanalysen av den SE-positive gruppen med RA pasienter mot kontroller for de fem SNPene viste også at ingen av SNPene hadde signifikante p-verdier ($p<0,05$) (tabell 14).

Tabell 14: Assosiasjonsanalyse av SNPene i *AIRE/ICOSLG* hos SE positive RA prøver mot kontrollprøver med p-verdi, odds ratio og konfidensintervall.

SNP ID	Gen	Minor allel	Frekvens RA	Frekvens ktr.	OR	95 % CI	p-verdi
rs7282490	*ICOSLG*	G	0,40	0,382	1,08	(0,90-1,30)	0,40
rs4819388	*ICOSLG*	T	0,263	0,262	1,00	(0,82-1,23)	0,98
rs2075876	*AIRE*	A	0,102	0,104	0,98	(0,73-1,31)	0,90
rs3788113	*AIRE*	G	0,18	0,206	0,85	(0,68-1,06)	0,15
rs760426	*AIRE*	G	0,107	0,125	0,84	(0,64-1,10)	0,21

4.3 Assosiasjon mellom polymorfismer og genekspresjon

For å kunne undersøke om det var assosiasjon mellom noen av de seks selekterte SNPene og genuttrykket til *AIRE* og/eller *ICOSLG*, ble genotypingsdata for 42 tymusprøver plottet mot genekspresjonsdata hentet fra Illumina Human WG-6 v3 arrayet for tre prober i *AIRE* og én i *ICOSLG*.

4.3.1 Kvalitetskontroll av probene for AIRE og ICOSLG

Det ble utført en kvalitetskontroll av de fire probene, ILMN_1670282, ILMN_2261519, ILMN_1791236 og ILMN_1675671, (fra Illumina Human WG-6 v3 arrayet) for hhv. *AIRE* (3 stk.) og *ICOSLG* (1 stk.) for å undersøke hvor de lå i forhold til de aktuelle genene, hvilke transkripter probene binder og om de hadde polymorfismer i probesekvensen.

På microarrayet var det tre prober for å fange opp *AIRE* og én for *ICOSLG*. To av probene for *AIRE*, ILMN_1791236 og ILMN_1670282, viser genuttrykkssignal mellom 5-6, mens probe ILMN_2261519 (*AIRE*) viser et signal på 13 (tabell 15). Proben for *ICOSLG*, ILMN_1675671**,** viste et genuttrykk på 5,85 (tabell 15).

Tabell 15: Oversikt over de ulike microarray probene for *AIRE* og *ICOSLG*, hvilke transkripter de binder til og om det finnes polymorfismer i probesekvensen.

Gen	Probe	Koordinater (Hg38)	Sekvens	Transkripter	Genuttrykk**	Polymorfismer (MAF >1 %)
AIRE	ILMN_1670282	chr21:44297893 -44297942	GTGCCTGGAAATTAAAC CCTGCCCCACTTCTCTAC TCTGGAAGTCCCCGG	NM 000383.3	6,58	Nei
AIRE	ILMN_2261519	chr21:44298080 -44298129	TGAGATTGCGCCACTGC ACTCCAGTCTGGTCGGCA AGAGTGAGACTCCGT	NM 000383.3, *NM_021148**	13,54	Nei
AIRE	ILMN_1791236	chr21:44289698 -44289747	ACTCCCAGCAAGTTCGA AGACTCCGGCAGTGGGA AGAACAAGGCCCGCAG	NM 000383.3	5,29	Nei
ICOSLG	ILMN_1675671	chr21:44226963 -44227012	TGAAGCCCCCTCAGAAG CCCTGCCTGTCACGTCGG CATTTGTGAGACCTA	NM_001283052.1, NM_001283051.1, NM_015259.5	5,85	Nei

*Ikke validert transkript og ikke RefSeq gen
**Median av genuttrykk for hver probe

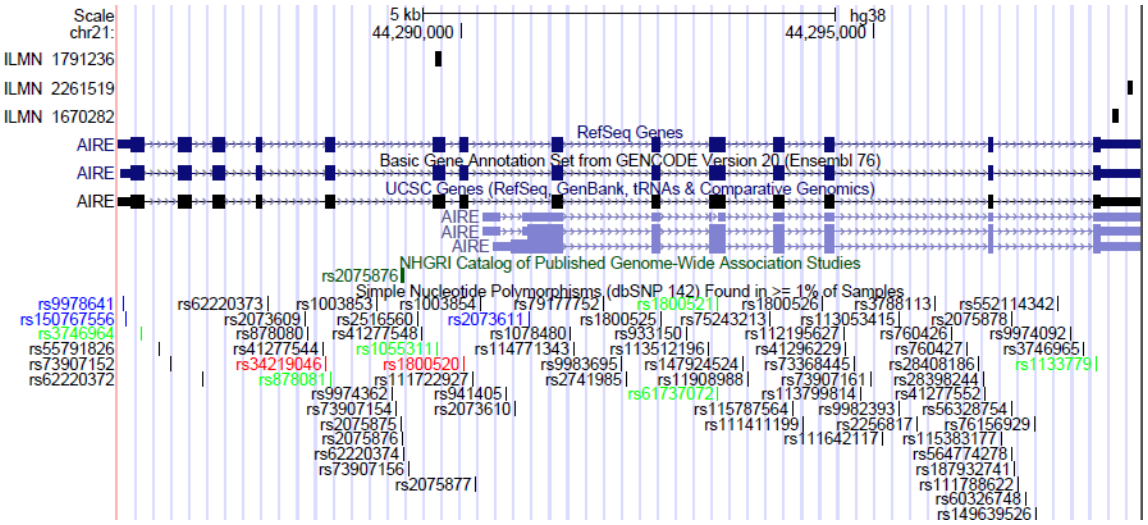
I følge RefSeq har *AIRE* (figur 17) ett godkjent og validert transkript. Når vi ser på andre databaser i UCSC browser er det opptil fire transkripter for *AIRE* som er predikert, men tre av disse er ikke validert og den fjerde er den samme som hos RefSeq.

ICOSLG (figur 18) viste seg å ha fire ulike transkripter i følge RefSeq. Når vi ser på andre databaser i UCSC browser ser vi at det er et ekstra transkript, i tillegg til de fire for *ICOSLG*, som er predikert, men ikke validert.

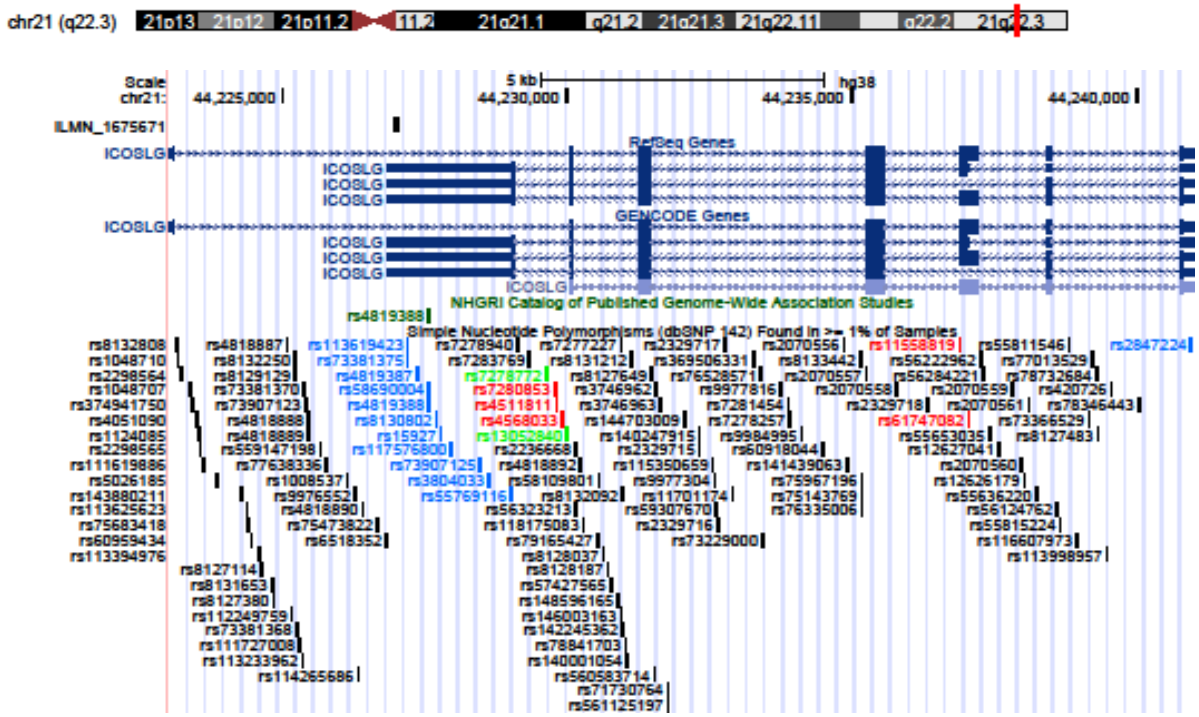
Alle de tre *AIRE* probene viste seg å binde til det ene validerte transkriptet som finnes av *AIRE* med 100 % (tabell 15 og figur 17). Probe, ILMN_2261519, ga mange treff på ulike sekvenser. Alle sekvensene som kunne binde seg 100 % (20 stk.) med hele probesekvensen ble undersøkt nærmere. Det resulterte i et annet gen som proben kan binde seg til med 100 %, *ZNF273* (chr.7: (64897368-64897396), men dette genet er bare predikert og ikke validert, det er heller ikke rapportert i RefSeq. Allikevel kan det bety at proben fanger opp dette genet i tillegg til *AIRE*. For de tre transkriptene av *AIRE* som ikke er validert var det bare to av probene, ILMN_2261519 og ILMN_1670282, som festet seg til alle de fire *AIRE* transkriptene, mens ILMN_1791236 bare festet seg til det validerte transkriptet (figur 17). Dette kan føre til at de to probene som binder fire transkripter får høyere intensitet fordi de dekker mer av transkriptene.

Proben for *ICOSLG*, ILMN_1675671, fanget opp tre av de fire rapporterte transkriptene som den binder til med 100 % (tabell 15 og figur 18). Probesekvensen ligger på siste ekson (nr.7) og fanger ikke opp noen andre sekvenser.

Polymorfismer i probene kan føre til at proben ikke fester seg godt nok og dermed vil det bidra til dårlige analyseresultater. I søket etter polymorfismer i de fire ulike probene vha. BLAT var det verken SNPer i *ICOSLG* eller *AIRE* probene.



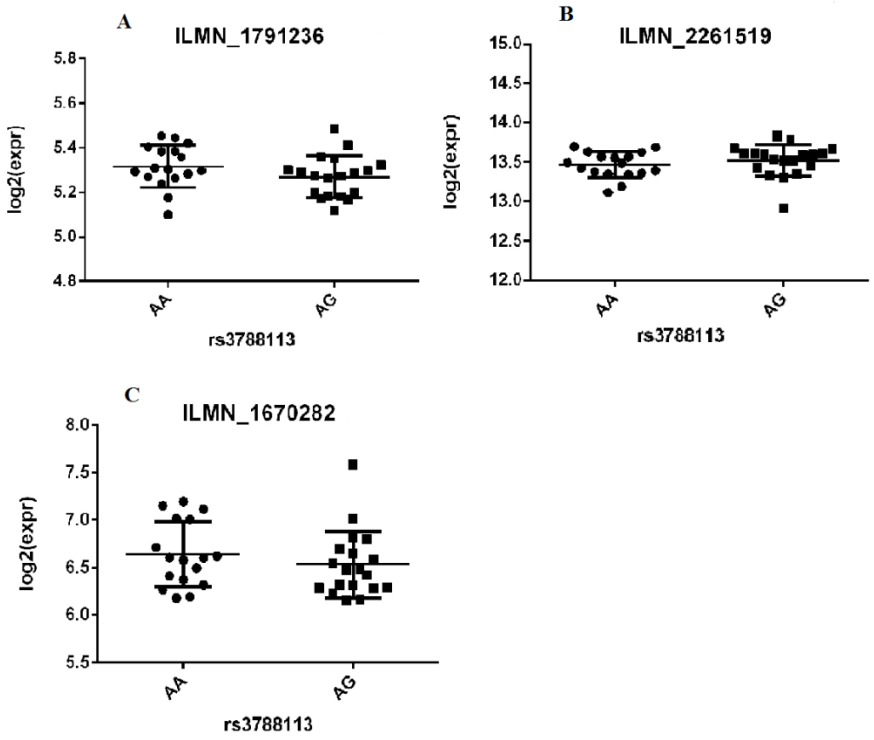
Figur 17: Oversikt over *AIRE* transkriptene (mørke blå – RefSeq validert og godkjent transkript, lyse blå – ikke validerte transkripter), hvor probene sitter på transkriptet (svarte rektangler øverst i bilde med probenavn øverst til venstre) og polymorfismer i genet.



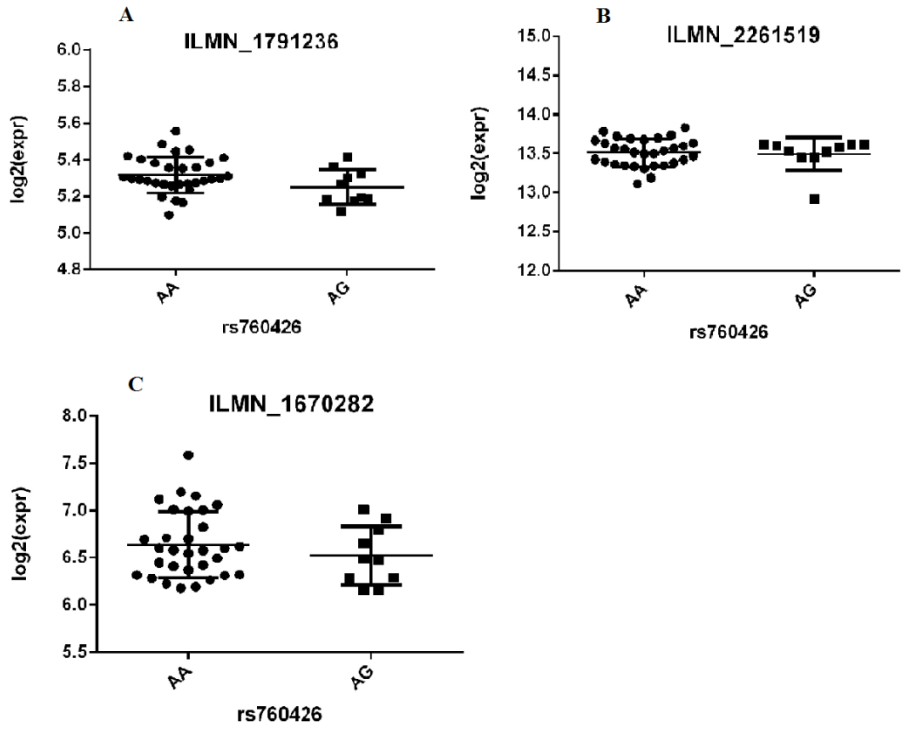
Figur 18: Oversikt over *ICOSLG* transkriptene (mørke blå – RefSeq validert og godkjent transkript, lyse blå – ikke validerte transkripter), hvor proben sitter på transkriptet (svart rektangel øverst i bildet med probenavn øverst til venstre) og polymorfismer i genet.

4.3.2 eQTL analyse

Analyse av genekspresjonsdataene for de 42 tymusprøvene mot genotypingsdataene for de seks selekterte SNPene for å se etter assosiasjon mellom SNPene og uttrykk av *AIRE* og *ICOSLG* viste at ingen av SNPene var signifikant assosiert med genekspresjonsnivå (p>0,05). Figur 19 og 20 viser en grafisk fremstilling av genekspresjonsnivået for *AIRE* probene for SNPene som var signifikant assosiert med sykdom, rs3788113 og rs760426. Resten av resultatene kan leses av i vedlegg 5.



Figur 19: Log2 transformerte genuttrykksdata plottet mot genotypedata for SNP rs3788113 i 42 tymusprøver for probe A) ILMN_1791236 (*AIRE*) med p-verdi = 0,1369, B) ILMN_2261519 (*AIRE*) med p-verdi = 0,2886, C) ILMN_1670282 (*AIRE*) med p-verdi = 0,2830.



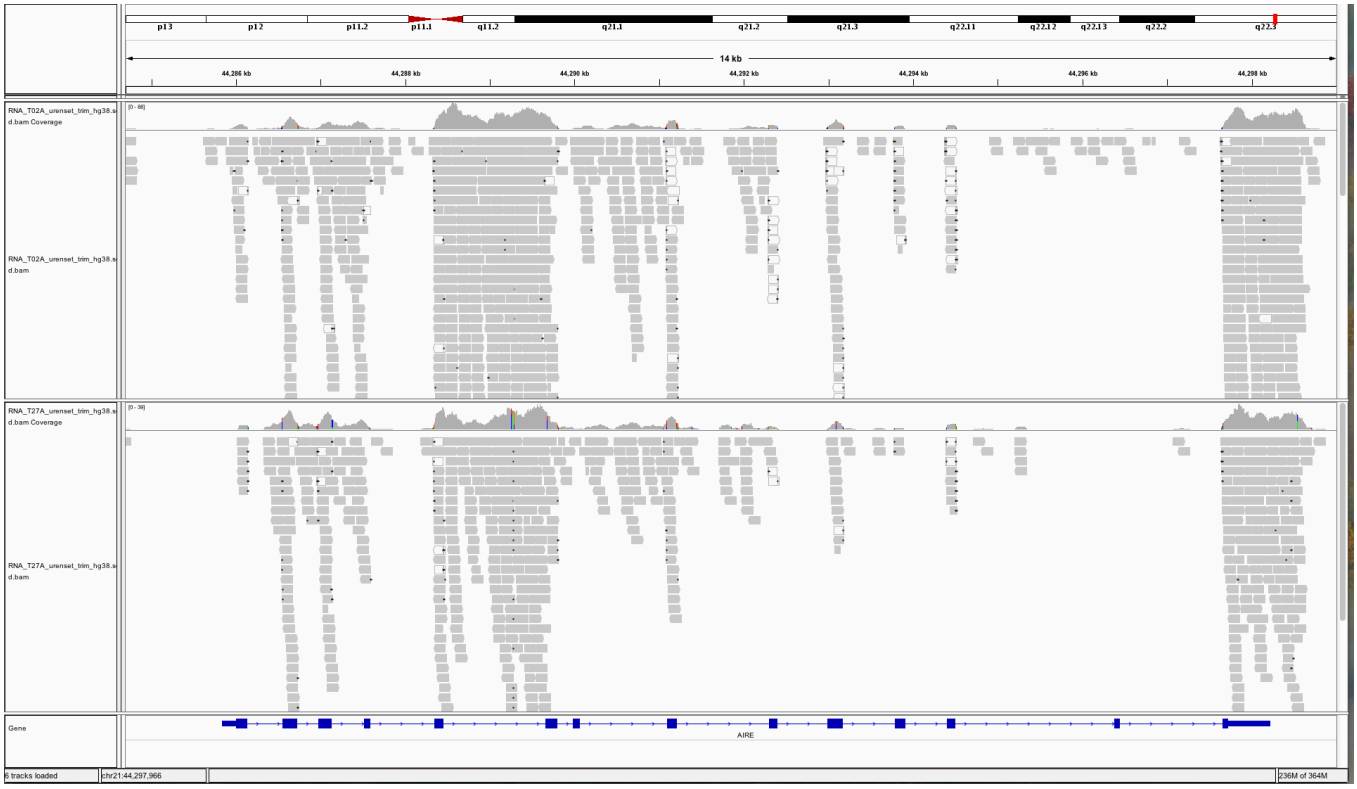
Figur 20: Log2 transformerte genuttrykksdata plottet mot genotypedata for SNP rs760426 i 42 tymusprøver for probe A) ILMN_1791236 (*AIRE*) med p-verdi = 0,1157, B) ILMN_2261519 (*AIRE*) med p-verdi = 0,8872, C) ILMN_1670282 (*AIRE*) med p-verdi = 0,3356.

4.3.3 RNA sekvensering av tymusprøver

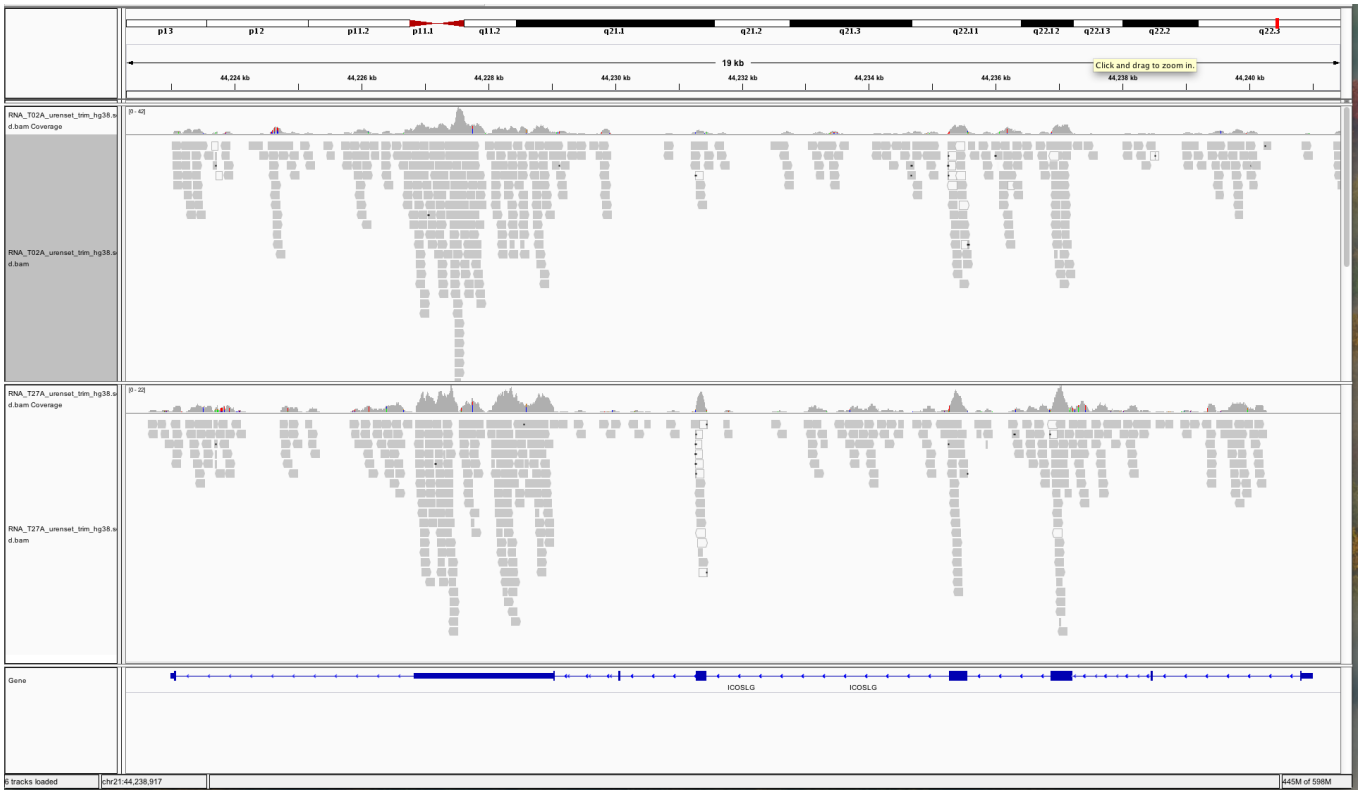
RNA-sekvenseringsdata for de to tymusprøvene, T02a og T27a, «alignet» til referanse genom Hg38 ved hjelp av Bowtie 2 og visualisert i IGV, kan ses i figur 21 og 22 for hhv. *AIRE* og *ICOSLG*.

De «alignede» RNA-sekvenseringsdataene for tymusprøvene T02a og T27a i *AIRE* (figur 21) viser at det mulig er flere spleisevarianter av genet i tymus som ikke er rapportert. Det ser slik ut fordi det er «alignet» veldig mange «reads» i intron 5, som ikke er rapportert å inngå i kjente transkripter. Det er også flere «reads» i intron7 og 3'UTR kan se ut til å være lenger enn det som er oppgitt. Det kan se ut som om at ekson 7 er spleiset ut fordi det er svært få «reads» «alignet» der. For T27a er det absolutt ingen «reads» «alignet» til ekson 13 som skaper en stor mulighet for at eksonet er spleiset ut i denne prøven. Generelt er det stor variasjon i antall «reads» i de ulike eksonene, noe som kan indikere flere ulike transkripter tilstede i tymusprøvene.

De «alignede» RNA-sekvenseringsdataene for tymusprøvene T02a og T27a i en sammenslått versjon av de fire transkriptene av *ICOSLG* (figur 22) viser at det trolig er flere spleisevarianter av genet i tymus som ikke er rapportert. Det er både variasjoner i antall «reads» over de ulike eksonene og det kan også se ut til at deler av intronene kan være transkribert for alle transkriptene av genet. For prøve T02a er det «alignede» «reads» til intron 2, og for T27a er det til intron 6. For begge prøvene kan det se ut til at det ikke er «alignet» noen «reads» til ekson 8 som kan tyde på at dette er spleiset ut og det er dannet en ny spleisevariant av genet for disse tymusprøvene.



Figur 21: «Alignede» RNA sekvenserings «reads» til referanse genom, Hg38, for RNA-sekvenseringsdataene til tymusprøvene T02a (øverst i bildet) og T27a (nederst i bildet) i *AIRE*.



Figur 22: «Alignede» «reads» til referanse genom, Hg38, for RNA sekvenseringsdataene til tymusprøvene T02a (øverst i bildet) og T27a (nederst i bildet) i *ICOSLG*.

4.3.4 RegulomeDB for å sjekke potensiell regulatorisk rolle

For å undersøke om de seks selekterte SNPene kunne ha en regulatorisk rolle ble det gjort et søk i RegulomeDB (tabell 16). Det viste seg å være lite evidens for at noen av SNPene var regulerende, da alle fikk en RegulomeDB score på 4 eller 5 (hvor 6 er veldig lite bevis for regulerende effekt og 1 er mye evidens for regulerende funksjon). En RegulomeDB score på 4 kan bety at SNPen sitter i et transkripsjonsfaktorssete og i en DNase Peak, mens en score på 5 kan bety at SNPen sitter i enten et transkripsjonsfaktorsete eller i en DNase Peak.

Tabell 16: Oversikt over RegulomeDB score for de seks selekterte SNPene i analysen som forteller noe om graden av den regulatoriske evnen til SNPen. På en skala fra 1-6 blir SNPene rangert, hvor 1 er svært regulerende, mens 6 er lite regulerende.

SNP	**RegulomeDB score [52]**
rs2075876	5
rs760426	4
rs7282490*	-
rs3788113	4
rs4819388	4
rs1800520	5

*Fantes ingen informasjon om rs7282490 i RegulomeDB.

5. Diskusjon

Hensikten med denne masteroppgaven var å undersøke om polymorfismer i *AIRE* genet var assosiert med RA i norsk befolkning og det ble funnet at to av de selekterte SNPene, rs760426 og rs3788113, viste assosiasjon med RA. Stratifiserte subfenotypeanalyser i RA ga indikasjoner på at assosiasjonen kunne knyttes til ACPA+ RA. Et annet mål med masteroppgaven var å undersøke om polymorfismer i *AIRE* hadde en regulatorisk funksjon på genuttrykket i tymus, noe som vi ikke fant for noen av de selekterte SNPene. RNA sekvensering av to tymusprøver viste at *AIRE* genet trolig har flere transkripter i tymus enn tidligere rapportert.

I en studie utført i Japan (30) ble det funnet polymorfismer i genet *AIRE* som kunne assosieres med RA i den japanske befolkningen, men det har ikke blitt rapportert noen assosiasjon i europeisk befolkning. Før assosiasjonsanalysene ble startet ble en SNP seleksjon utført med utgangspunkt i SNPene som ble assosiert med RA i japansk befolkning og SNPer ble selektert ved ulike kriterier ut i fra dette.

5.1 Assosiasjon mellom RA og SNPer i *AIRE* og *ICOSLG*

I denne assosiasjonsstudien er det selektert rapporterte SNPer som er assosiert med autoimmune sykdommer. Dette indikerer deres medvirkning i immunologiske prosesser og styrker deres sannsynlighet for medvirkning i RA.

To av de fem suksessfullt genotypede SNPene var signifikant assosiert med RA i norsk befolkning. rs3788113og rs760426 (i *AIRE* genet) hadde begge signifikante p-verdier ($p<0,05$). Grunnen til at man ser en assosiasjon her er fordi allelet det er minst av, G, observeres i høyere frekvens hos kontroller enn hos pasienter og dette gir en redusert risiko for sykdom med OR 0,84 for rs3788113 og 0,77 for rs760426. Dette tyder på at den andre allelvarianten av disse to *AIRE* SNPene, A (som er varianten det er mest av), kan være disponerende for RA i vår befolkning. Før man kan si sikkert om allelvarianten A er sykdomsdisponerende må flere analyser utføres. Det ble valgt å rapportere videre for det beskyttende allelet (G), som er det allelet som har blitt rapportert i andre studier (30).

Assosiasjonen til *AIRE* (rs2075876 og rs760426) i asiatisk populasjon er bekreftet i andre studier (53, 54). Terao et. al. fant at rs760426*G var signifikant assosiert ($p<0,001$) med RA i japansk befolkning med $OR\approx1,21$. (30) Vi har altså funnet assosiasjon til samme SNP, men allel G gir økt risiko i japansk befolkning, mens vi har funnet tegn på at det reduserer risiko. rs760426 viste seg å ha signifikant forskjell i allelfrekvens av G, med 0,11 for RA pasienter og 0,13 for kontroller i norsk befolkning. Assosiasjon med motsatt fortegn er funnet for SNPen i artikkelen for opptil flere ulike kohorter av japansk befolkning, der allelfrekvensen til G i de japanske kohortene er 0,39 hos RA pasienter og 0,36 hos kontroller. Ut i fra disse dataene kan det virke som om at nordmenn og japanere har motsatt risikoallel. Hos nordmenn ser det ut som at den andre allelvarianten, A, som det er mest av, er sykdoms disponerende (risikoallelet), mens hos japanere er risiko forbundet med G-allelet. MAF for rs760426 hos japanere er mye høyere enn den for nordmenn, noe som betyr at det er mer av denne allelvarianten hos både kontroller og RA pasienter generelt i den japanske populasjonen enn i den norske.

For rs3788113 var frekvensen til G, som var allelet det var minst av, også lavere hos RA pasienter enn hos kontroller (hhv. 0,19 mot 0,21). Noe som betyr at den mest sjeldne allelvarianten av SNPen dukker oftere opp hos friske personer enn hos RA pasienter og at den da kan virke beskyttende mot RA i vår befolkning.

Haplotypefrekvensen mellom RA individer og kontroller viste at de to risikoallelene på SNPene rs760426 og rs3788113 forekom på den mest frekvente haplotypen A-A (81 % hos RA individer mot 77 % hos kontroller). De beskyttende allelene befinner seg på haplotype G-G som var redusert i RA pasienter (10 %) mot kontroller (12 %). Noe som styrker vår antakelse om at G er den beskyttende allelvarianten og A den som er disponerende for RA, da haplotype A-A er mest frekvent hos RA individer og G-G hos kontroller. Det at haplotypene med beskyttende allel på bare den ene SNPen ikke ga signifikant beskyttelse, kan bety at ingen av de to testede SNPene er kausale, men en annen variant i koblingsulikevekt med disse.

Grunnen til at vi finner ulikt risiko/beskyttende allel mellom Japan og Norge kan skyldes ulikt koblingsulikevektmønster mellom de genotypede SNPene og den kausale risiko-SNPen. Det at allelfrekvensene er så forskjellige kan forklares med at asiatere og europeere generelt er ganske ulike genetisk, og at polymorfismer og deres koblingsulikevektmønster har utviklet seg ulikt. SNPen vi inkluderte basert på ulikt koblingsulikevektmønster i asiatisk og europeisk populasjon var blant de to SNPene vi fant assosiasjon til (rs3788113), noe som støtter ulikt koblingsulikevektmønster og assosiasjonsmønster mellom de to populasjonene. I tillegg ble det funnet en moderat koblingsulikevekt (r^2 = 0,63) mellom rs760426 og rs2075876 i japanere (30), som kan tyde på at det er en viss korrelasjon mellom disse SNPene, mens for kontrollprøver i vår befolkning fant vi full koblingslikevekt (r^2=0,01) og ingen korrelasjon mellom SNP genotypene.

Koblingsulikevektsanalysen mellom de fem suksessfullt genotypede SNPene i de norske kontrolldataene viste at det ikke var noen koblingsulikevekt (r^2= 0, D`<0,22) mellom SNPene i *ICOSLG* (rs4819388 og rs7282490) og i *AIRE* (rs2075876, rs760426 og rs3788113). Noe som betyr at sykdomsassosiasjonene rapportert for *ICOSLG* SNPene (48, 49) antagelig ikke kan tilskrives RA assosiasjonen til *AIRE* SNPene som er rapportert (30). Det er allikevel mulig at assosiasjonen funnet i kromosomområdet til RA og IBD/cøliaki skyldes samme risikofaktor, til en SNP som ennå ikke er studert. Denne hypotesen styrkes av funnene gjort i studien utført av Farh K.K-H. et. al (28) hvor de har kommet frem til at det bare er en gjennomsnittlig sjanse på 5 %

at markør-SNPer som har plukket opp assosiasjon i GWAS katalogen representerer kausale SNPer. De kom også frem til at markør-SNPer som har plukket opp assosiasjon i GWAS ofte er et stykke unna den kausale SNPen (median 14 kb) og mange er ikke i sterk koblingsulikevekt. Dermed er det en sjanse for at de to SNPene som de fant assosiert med RA i japansk befolkning (30) ikke er de kausale SNPene og at de kausale SNPene ennå ikke er studert eller rapportert.

Koblingsulikevektmønsteret mellom de tre ulike *AIRE* SNPene viste at det var lav korrelasjon SNPene i mellom og den høyeste koblingsulikevekten (r^2= 0,42) var mellom rs3788113 og rs760426. Mellom de to *ICOSLG* SNPene var det også lav korrelasjon med koblingsulikevekt (r^2= 0,14). Dette betyr at ingen av de selekterte SNPene gir tilnærmet like resultater, men at de gir unik informasjon om assosiasjon noe som styrker vår seleksjon da vi hadde som mål å velge SNPer som ikke var i koblingsulikevekt i europeisk befolkning, noe som også stemte overens med norsk befolkning.

D` viste litt høyere koblingsulikevekt verdier for nesten alle SNPene enn $r^2$, noe som indikerer haplotypestrukturer og tyder på at det er begrenset antall haplotyper mellom SNPene i regionen. *ICOSLG* SNPene, rs4819388 og rs7282490,viste lav koblingsulikevekt, alle D`≤0,22, med alle *AIRE* SNPene. D` for de assosierte *AIRE* SNPene, rs760426 og rs3788113, var på 0,86, som kan tyde på en skjev haplotypefordeling. Ut fra haplotypeanalysen så vi at tre av de fire haplotypene mellom disse SNPene er vanlige i norsk befolkning, mens én (G-A) er sjelden (1,4 % i kontrollmaterialet).

Vi ble oppmerksomme på en studie på *AIRE* i europeisk befolkning (55), som ikke var rapportert i GWAS katalogen og ble dermed ikke vurdert i vår SNP seleksjon. De hadde brukt en annen strategi enn oss for å undersøke hvorfor *AIRE* er funnet assosiert i Japan, og ikke i Europa. De selekterte *AIRE* SNPer som ikke var representert på GWAS genotypingsarray. En av disse SNPene, rs878081, var signifikant assosiert med RA i spansk populasjon (OR= 1,41). Det er ingen koblingsulikevekt mellom denne SNPen og våre to assosierte SNPer (rs3788113 og rs760426) i verken europeisk (r^2<0,1) eller asiatisk befolkning (r^2<0,01). Generelt viser den RA assosierte SNPen fra Japan (rs2075876) høyere koblingsulikevekt med andre *AIRE* SNPer i asiatisk populasjon i forhold til europeisk, noe som peker på at det er mange aktuelle kandidatSNPer som kan forklare assosiasjonen rapportert i Japan (se vedlegg 6). Likeledes for SNPen funnet assosiert i Spania (rs878081), ser vi at den er i koblingsulikevekt med andre

SNPer i Europa, men ikke i Asia. Hvis risiko locuset er det samme på tvers av populasjonene, er det lite sannsynlig at det er en av disse to SNPene. Når det gjelder de to SNPene som bare er funnet assosiert i Europa (rs878081 i Spania og rs3788113 i Norge) er det ingen felles merke-SNPer (r^2>0,4) i 1000 Genomes datasettet for CEU (data ikke vist). Samlet sett tyder våre analyser og de publiserte studiene på at det kan være flere *AIRE* SNPer som er involvert i RA disposisjonen, men at ingen av de kausale RA SNPene i *AIRE* ennå er funnet.

5.1.1 ACPA stratifiserte analyser

Assosiasjonsanalysene som ble utført på hele prøvematerialet ble også kjørt på subgruppene av RA (ACPA+ (539 stk.) og ACPA- (394stk.)) mot kontroller for å undersøke om noen SNPer var assosiert med RA i en av gruppene og om det var store forskjeller mellom dem.

Før analysen av den ACPA + og ACPA – gruppen ble utført, ble 57 RA individer fjernet fra datasettet da vi manglet informasjon på deres ACPA status. Dette antallet individer var såpass lite i forhold det totale antallet og det skal være tilfeldig fordelt, så det er ikke grunn til å tro at det har laget skjevhet i resultatene. Det hadde allikevel styrket analysen om vi kunne inkludert flere individer med definert ACPA status.

De to SNPene rs3788113 og rs760426 viste signifikante p-verdier (p<0,05) i ACPA+ RA individer mot kontroller, akkurat som i totalanalysen. Mens ingen av SNPene viste signifikant assosiasjon med RA i vår kohort (alle p>0,05) hos ACPA- RA individer mot kontroller. Dette kan tyde på at SNPene disponerer for ACPA+ RA.

rs760426 hadde en MAF som var registrert lavere for ACPA + RA individer (0,10) enn for kontroller (0,13) og det samme gjaldt for rs3788113 med MAF hos ACPA + RA individer (0,18) og kontroller (0,21). OR var sterkere i subgruppen ACPA+ RA enn i totalanalysen for begge de to signifikant assosierte SNPene. Resultatet viste også nesten samme allelfrekvensmønster som for totalanalysen, men med litt lavere allelfrekvens registrert for ACPA + RA enn for RA som inneholder prøver fra begge subgruppene. Det at det er en større forskjell i allelfrekvensen

mellom kontroller og ACPA + RA, og at ingen av SNPene viste signifikant assosiasjon med RA i ACPA- individer og kontroller, kan tyde på at assosiasjonen er knyttet til ACPA + RA.

MAF hos ACPA- RA individer og kontroller hadde dog litt lavere allelfrekvens for de beskyttende allelene på rs760426 og rs3788113 for ACPA – RA individer enn kontrollene. I vår kohort hadde vi flere ACPA + RA individer enn ACPA- RA individer, noe som gjør at de ACPA+ har mer styrke enn de ACPA-. Da vi ser samme tendens for allelfrekvensene i de ulike subgruppene, kan det være en mulighet for at vi kunne sett en signifikant assosiasjon for den ACPA- subgruppen også dersom prøvematerialet hadde vært større. Vi så ingen signifikante forskjeller ($p<0,05$) mellom ACPA+ RA individer og ACPA- RA individer, så vi kan ikke konkludere med at det er forskjeller mellom de to pasientpopulasjonene. I tidligere studier (30) av polymorfismer i *AIRE* genet assosiert med RA i japanere har de heller ikke funnet noen signifikant forskjell mellom de to kliniske fenotypene av RA. De undersøkte dette for SNPen rs2075876 hvor allelfrekvensene var 0,39 for ACPA+ RA og 0,40 for ACPA- RA (mot 0,34 i kontroller). I vår befolkning var allelfrekvensene for denne SNPen 0,11 for ACPA- RA og 0,10 for ACPA+ RA (mot 0,11 i kontroller). Før det kan trekkes en konklusjon rundt om assosiasjonen er knyttet til ACPA + RA må analysene utføres på flere kohorter i befolkningen, fordi konfidensintervallene mellom de to analysegruppene overlapper og det ikke ble påvist signifikante forskjeller mellom ACPA+ og ACPA- individer.

Haplotypefrekvensen mellom ACPA + RA individer og kontroller for de to assosierte SNPene rs760426 og rs3788113 viste nesten akkurat den samme fordelingen av frekvens som i totalanalysen. Haplotype A-A var uttrykt ved høyere frekvens i ACPA + RA individer (0,81) enn hos kontroller (0,77), som vil si at denne haplotypen er mer frekvent hos individer med ACPA + RA enn hos kontroller i norsk befolkning. Og haplotype G-G var uttrykt ved høyere frekvens i kontroller (0,12) enn i ACPA+ RA individer (0,09), hvor det var enda litt lavere frekvens av haplotypen i ACPA + RA individer enn for RA i totalanalysen.

Det ble undersøkt om assosiasjonen med SNPene fulgte SE stratifisering. Det var ingen signifikant forskjell ($p<0,05$) mellom RA individer og kontroller for de fem selekterte SNPene i verken SE-positive eller SE-negative strata. Dette tyder på at assosiasjonen følger ACPA status og ikke SE status.

5.1.2 Genotypefordeling

Genotypeanalysen både i totalmaterialet og i de stratifiserte analysene viste entydig at rs760426 fulgte en dominant modell for det beskyttende allelet G. For rs3788113 viste det beskyttende allelet kun assosiasjon med trend test i totalmaterialet, mens det i ACPA+ RA ga utslag både for dominant modell og trend test. Frekvensfordelingen av genotypene for de to SNPene støttet den dominante modellen. I tillegg viste haplotypeanalysen at de to beskyttende allelene er på samme haplotype, noe som gjør at man kan anta at de representerer samme effekt og dermed må følge samme modell. Samlet sett, tyder analysene på at *AIRE* assosiasjonen vi har funnet i den norske befolkningen følger en dominant modell for det beskyttende allelet.

5.2 Genekspresjon

Dersom en av de selekterte SNPene har en regulatorisk funksjon på genekspresjonen til *AIRE* og/eller *ICOSLG* og dette fører til at uttrykket av *AIRE* og/eller *ICOSLG* blir forandret, vil dette styrke vår antagelse om at SNPene er assosiert med RA da funksjonen til *AIRE* og/eller *ICOSLG* blir svekket noe som kan føre til autoimmunitet og sykdom.

Assosiasjonsanalysen av genekspresjonsdataene for de 42 tymusprøvene mot genotypingsdataene for de seks selekterte SNPene viste at ingen av SNPene var signifikant assosiert med genekspresjonsnivå (alle $p>0,05$). Dette kan tyde på at ingen av de selekterte SNPene har en regulatorisk funksjon på genekspresjonen til *AIRE* eller *ICOSLG*.

I den japanske studien (30) så de også på genekspresjon, av rs760426 og rs2075876 i 210 lymfoblastoid celler i japansk befolkning. De fant at transkripsjonen av *AIRE* ble redusert av risikoallelet A i rs2075876 ($p = 6,8*10^{-5}$), men de så ingen assosiasjon med genekspresjonsnivå for rs760426. Genekspresjon er celle- og vevsspesifikk, slik at ekspresjonsnivåer og regulering av disse vil variere mellom celletyper. eQTL er funnet å variere mellom celletyper både når det gjelder hvilke SNPer som er assosiert med ekspresjonsnivåene og hvorvidt SNPene øker, reduserer eller har ingen effekt på geneskpresjonsnivået (56). *AIRE* er hovedsakelig uttrykt i mTEC, mens våre analyser var gjort på helt tymusvev, som er en blanding av ulike celletyper.

Den mest frekvente cellepopulasjonen er tymocytter, og epitelcellene utgjør under 5 % av cellene. Dette kan forklare hvorfor *AIRE* geneskpresjonsnivået er relativt lavt (signalet ligger i nedre sjikt av ekspresjonssignalene fra alle probene på arrayet).

De fire probene fra Illumina Human WG-6 v3 arrayet som ble brukt til å se på genekspresjonen av *AIRE* og *ICOSLG* kan ha begrenset vår analyse mht. hvilke transkripter de detekterer. AIRE proben, ILMN_2261519, bandt seg uspesifikt til et transkript fra et annet gen (*ZNF273*) i tillegg til *AIRE* transkriptet og vi har da fått ut en upålitelig måling av genuttrykket til *AIRE* for denne proben. ICOSLG proben detekterer ikke alle transkriptene av *ICOSLG* genet og vi kan derfor ha fått lavere genekspresjonsdata for genet i visse individer ved at dette transkriptet ikke har blitt plukket opp. Generelt diskriminerer ikke probene mellom ulike spleisevarianter fra de to genene.

Dersom vi hadde funnet tegn til at noen av SNPene var assosiert med genuttrykket av *AIRE* eller *ICOSLG*, kunne vi verifisert dette med qPCR spesifikke for de ulike transkriptene fra de to genene.

5.2.1 Spleisevarianter av *AIRE* og *ICOSLG* i tymus

De "alignede" RNA-sekvenseringsdataene for tymusprøvene T02a og T27a i *AIRE* viste at det mulig er flere spleisevarianter av genet i tymus som ikke er rapportert. Bakgrunnen for dette er at det har blitt «alignet» mange «reads» til intron 5 og intron 7 for begge prøvene, hvor det i utgangspunktet ikke skal havne noen RNA data da disse bare består av kodende sekvenser (ekson). Og at det kan se ut som om opptil flere eksoner (ekson 7, 13) er spleiset ut da det er få eller ingen «reads» «alignet» til disse.

For T02a og T27a i *ICOSLG* så det også ut som om at det trolig finnes flere spleisevarianter av genet i tymus som ikke er rapportert. Dette pga. at begge prøvene har «alignede» «reads» til intron hvor i utgangspunktet RNA data ikke skal kunne «alignes» i alle de fire transkriptene av genet. Fordelingen av «reads» i intron for *ICOSLG* var mer spredt utover, og disse kan også representere pre-mRNA før intronsekvensen er spleiset ut. For *ICOSLG* kan vi også se en mulighet for at ekson kan ha blitt spleiset ut da det ikke er «alignet» noen «reads» til ekson 8 for noen av prøvene.

Ulike celletyper har sitt repertoar av spleisevarianter og transkripter noe som betyr at dette må kartlegges for hver enkelt celletype. Det at vi ser at det trolig er flere spleisevarianter av *AIRE* og *ICOSLG* i tymus som ennå ikke er rapportert og at transkripter i tymus generelt er lite studert, skaper behov for å kartlegge variantene av disse i tymus.

5.3 Metodologiske betraktninger

5.3.1 SNP seleksjon

En viktig forutsetning for funnene i assosiasjonsstudien er valget av polymorfismer for genotyping. SNP seleksjonen i dette arbeidet resulterte i seks SNPer som ble genotypet.

Utgangspunktet for SNP seleksjonen baserte seg på en tidligere studie av Terao et.al.(30) hvor de fant assosiasjon mellom polymorfismer i *AIRE* og RA i japanere. Vi valgte å se etter SNPer i *AIRE* genet for å undersøke om disse kunne ha assosiasjon til RA i vår befolkning. Selv om det er funnet polymorfismer i dette genet assosiert med RA hos japanere, vet vi at det kan være store genetiske forskjeller i ulike etniske grupper, noe som kan bety at det kan være andre genvarianter i vår befolkning enn i asiatere som er assosiert til RA. Det kan også være en sannsynlighet for at SNPene som ble funnet i assosiasjon med RA i japanere ikke er de kausale variantene og at ulikt koblingsulikevektmønster i ulike populasjoner kan gjøre at det er forskjellige merke-SNPer og markører som plukker opp assosiasjon fra den kausale varianten i ulike populasjoner.

SNPer fra genet *ICOSLG* ble inkludert i seleksjonen fordi risikogener for autoimmune sykdommer viser stor grad av genetisk overlapp, og vi ville dermed undersøke om *ICOSLG* og *AIRE* kunne ha samme assosiasjon til RA, da genet lå i nærheten av *AIRE* i kromosomområdet 21q22.3. I seleksjonen valgte vi bare å se etter SNPer assosiert med autoimmune sykdommer i andre gener enn *AIRE* innenfor et bestemt intervall på 500 kb, noe som kan ha begrenset våre resultater ved at vi har gått glipp av eventuelle SNPer som ennå ikke har blitt rapportert til å være assosiert med autoimmune sykdommer (men er det) da vi ikke så på annet enn risikogenene i kromosomområdet. Samtidig så er det en sjanse for at vi valgte å begrense oss til et for lite

område rundt *AIRE*, slik at vi kan ha gått glipp av risikogener for autoimmune sykdommer med SNPer som kan ha en assosiasjon til RA selv om de ligger langt unna i kromosomområdet. Ideelt sett burde vi ha testet alle genetiske varianter i området og i hvert fall merke-SNP for alle genetiske varianter i kromosomområdet. En grunn til at intervallet ble begrenset til dette området var fordi koblingsulikevekten svekkes med avstand og vi skulle bruke opplysninger om koblingsulikevekt videre i seleksjonen til å ekskludere/inkludere SNPer ved ulike kriterier.

De nærmeste andre genene i området er *DNMT3L* (ligger mellom *ICOSLG* og *AIRE*) og *PFKL* (som ligger telomert for *AIRE*, mens *ICOSLG* ligger centromert for *AIRE*). *PFKL* koder for subenheten i et enzym som katalyserer omdannelsen av D-fruktose 6-fosfat til D-fruktose 1,6-bisfosfat, som er et viktig trinn i glukosemetabolismen, mens *DNMT3L* koder for en DNA-metyltransferase (GeneCards; http://www.genecards.org/). Ingen av disse genene er åpenbare kandidatgener for RA og autoimmune sykdommer, og det har heller ikke vært rapportert autoimmun sykdomsassosiasjon til polymorfismer i disse genene.

Hypotesen vår for et trinn i seleksjonsprosessen var at vi skulle se etter SNPer som var i koblingsulikevekt ($r^2 \geq 0{,}8$) med de assosierte SNPene, rs760426 og rs2075876, i asiatisk befolkning, men med lav koblingsulikevekt til disse SNPene i vår befolkning, pga. det ikke er påvist at polymorfismer i *AIRE* genet er assosiert med RA hos europeere som hos japanere. Det som kan være tilfelle er at selv om det ikke er rapportert at samme polymorfismer i *AIRE* er assosiert med RA hos europeere som hos japanere, kan det allikevel vise seg å være samme kausale polymorfisme hos europeere og japanere, men at ulike merke-SNPer vil fange opp assosiasjonen pga. ulikt koblingsulikevektmønster. I tillegg har vi bare testet tre *AIRE* SNPer, mens det er flere hundre SNPer rapportert i dette genet (dbSNP) og vi har således ikke kunnet fange opp all genetisk variasjon som eksisterer for dette genet. Det kunne også vært en mulighet å ha en litt lavere koblingsulikevektsgrense for seleksjonen av SNPer i koblingsulikevekt med de assosierte SNPene, rs760426 og rs2075876, i asiatisk befolkning for å inkludere enda litt flere SNPer som allikevel viser en moderat koblingsulikevekt til de to SNPene.

I seleksjonen ble SNPer med MAF <5 % ekskludert da vi ikke hadde styrke til å detektere assosiasjon til lavfrekvente alleler. Dette kan ha ført til at vi har ekskludert lavfrekvente SNPer som kan være assosiert med RA.

5.3.2 Kvalitetskontroll

Kvalitetskontroll av rådata (før selve assosiasjonsanalysene settes i gang) er svært viktig for å sikre seg pålitelige – og for å unngå falske resultater.

5.3.2.1 Kvalitetssikring av genotyperesultater

I denne forskningsstudien ble kvalitetskontroll av TaqMan genotypingsdataene utført både manuelt ved å se på genotypingsplottene og ved statistiske beregninger av GSR og HWE.

Flere individer lot seg ikke genotype, dette gjaldt både identiske prøver som gikk igjen i alle de ulike SNPassayene og individer som ikke lot seg genotype i absolutt alle SNPassayene. Grunnen til dette var fordi de havnet i mellom genotype «clusterne»/klyngene eller fordi de hadde for svak konsentrasjon og dannet en klynge rundt den negative kontrollen (vann).

Dette gjaldt spesielt for rs1800520 som var det eneste SNPassayet som var bestilt spesiallaget fra Thermo Fisher Scientific (det ble sjekket for feilbestilling og det var ikke en realitet). Vi observerte to SNPer, en -19 bp og en + 17 bp unna rs1800520, begge var i svært lav frekvens og såpass langt unna at vi ikke regnet med at dette ville skape noen problemer for oligodesignet. Det viste seg at tre av de andre selekterte SNPene også hadde flere SNPer i nærheten av seg. For rs2075876 ble det observert to SNPer, en +12 bp og en – 9 bp unna, noe som kan bidra til problemer rundt designet ved at oligoen ikke får korrekt feste hvis den legger seg over SNPen. Det ble observert en SNP (rs183796400) bare + 7 bp fra rs7282490, SNPen var registrert med svært lav allelfrekvens, men den kan allikevel ha gjort det vanskelig med design av oligoen. For rs4819388 ble det observert en SNP (rs58690004) + 7 bp unna, men denne gikk vi ut i fra at ikke skulle ha noe å si for festing av oligo og genotyping da den ikke er polymorf hos kaukasere i 1000 Genomes i følge SNAP Proxy Search.

Alle individene som ikke lot seg genotype ved første forsøk, ble reanalysert. For rs1800520 ble en reanalysering ikke utført og heller ikke alle kontrollprøver genotypet på grunn av manglende assayreagenser. Antall genotypede individer ble derfor også mye mindre for rs1800520 (N=1506) enn de andre SNPassayene (N=1765). For at kvalitetskontrollen av genotypingsdataene til hvert SNPassay skulle få så godt som samme utgangspunkt for utregning av GSR ble de individene

som ikke lot seg genotype i et flertall av SNPassayene fjernet, da mye tydet på at vi hadde en altfor lav konsentrasjon eller DNA kvalitet av disse prøvene fra start av.

For å være sikre på at genotypingen stemmer bør noen prøver analyseres flere ganger, for å se at resultatet ikke endres. Eventuelt bør noen prøver analyseres med en annen metode, men av tidsmessige grunner ble ikke dette gjort. Siden ingen prøver ble analysert flere ganger (etter de hadde gitt et genotypingsresultat) kan vi ha en viss feilrate i rådataene.

5.3.2.2 GSR og HWE av genotypedata

Alle SNPene med unntak av rs1800520 hadde en GSR> 95 %. Dette betyr at mer enn 95 % av individene har blitt suksessfullt genotypet for hver SNP. For rs1800520 var GSR 80,2 %, noe som kan medvirke til at det er en skjevhet i genotypingsresultatene våre. Denne SNPen passerte ikke kravet til GSR >95 %.

For GSR på tvers av individer ble en grense på GSR ≥ 80 % satt, som betyr at alle individer som var suksessfullt genotypet for minst fire av de fem SNPassayene ble inkludert. Totalt 16 individer ble ekskludert for de fem SNPassayene (med unntak av rs1800520).

Vi forventer at SNPer relatert til autoimmune sykdommer er i HWE for både pasient- og kontrollgrupper fordi den utøvde risikoeffekten fra SNPene er lav. Dette stemmer overens med våre resultater, med unntak av rs1800520, da alle de andre SNPene bortsett fra denne er i HWE. At fem av SNPene var i HWE indikerer at genotypingen vår var av god kvalitet for disse SNPene.

Kontrollprøvene var i HWE for rs1800520 med p = 0,071, men ikke RA prøvene noe som kan tyde på feil under genotyping. Totalmaterialet viste avvik fra HWE. Blant de 299 prøvene som ikke lot seg genotype var 80 % av prøvene RA prøver. Dette kan forklare hvorfor det var et mer uttalt avvik fra HWE i denne gruppen. Det ble kontrollert i hvilken grad genotypene skilte seg fra HWE relativt til den forventede fordelingen og flere heterozygote genotyper enn forventet ble observert for både RA individer og kontroller. Ved manuell inspeksjon så vi at det var lite tydelige skiller mellom klyngene av prøver, noe som tyder på at SNPassayet kan ha fungert litt dårlig og at «clustringen» ikke var spesielt god. Det ble konkludert med at genotypingkvaliteten for denne SNPen ikke var tilfredsstillende. Dermed ble det bestemt at på bakgrunn av ikke

tilfredsstillende resultater for både GSR og HWE at denne SNPen ikke kunne brukes i videre analyser. En ting som kunne vært gjort for og mulig forbedre dette resultatet er at alle prøvene som ikke lot seg genotype for dette assayet og de resterende kontrollprøvene som ikke ble genotypet i det hele tatt ble genotypet (på nytt), slik at det hadde vært flere resultater. En annen ting som kunne blitt undersøkt nærmere er SNPassayet, om det er feildesignet eller at det er noe som forstyrrer annealingen av oligoene. Alternativt kunne SNPen blitt genotypet med en annen metode.

5.3.2.3 Kvalitet på ekspresjonsprobene

Kvalitetskontroll av de fire probene ILMN_1670282, ILMN_2261519, ILMN_1791236 og ILMN_1675671, (fra Illumina Human WG-6 v3 arrayet) for hhv. *AIRE* (3 stk.) og *ICOSLG* (1 stk.) ble utført manuelt med søk i databasen UCSC Genome Browser og ved bruk av verktøyet BLAT. Alle de fire probene lå i et ekson i enten *AIRE* eller *ICOSLG*, noe som er nødvendig for at korrekte ekspresjonsdata skal kunne registreres.

Alle de tre *AIRE* probene viste seg å binde til det ene validerte transkriptet som finnes av *AIRE* med 100 %. I tillegg fikk probe ILMN_2261519, 20 treff i BLAT på ulike genomiske områder med 100 % komplementær sekvens. Da disse ble undersøkt, viste det seg at alle sekvensene, bortsett fra en i *ZNF273* genet, var introns og ble derfor utelukket. Eksonsekvensen i *ZNF273* (chr.7: (60903242-64931577)) kunne binde seg 100 % til hele probesekvensen, men dette genet var bare predikert, ikke validert og ikke funnet i RefSeq. På grunn av dette ble proben brukt i analysen, noe som kan ha bidratt til usikkerhet i dataene da vi kan ha fått ut upålitelige genuttrykksmålinger siden vi ikke vet om proben har bundet både *AIRE* transkripter og *ZNF273* transkripter.

For de tre transkriptene av *AIRE* som ikke var validert var det to av probene, ILMN_2261519 og ILMN_1670282, som festet seg til alle disse, mens ILMN_1791236 bare festet seg til det validerte transkriptet. Dette betyr at de to probene som binder til alle de fire transkriptene av *AIRE* (validerte og ikke) høyst sannsynlig vil få høyere intensitet av genuttrykket hvis alle transkriptene befinner seg i det vevet man studerer. I våre tymusdata så vi at ILMN_1670282 og ILMN_2261519, som begge binder 3-UTR i *AIRE*, fikk hhv. 6 og 13 i gjennomsnittlig

genekspresjon. Det er flere muligheter til at ILMN_2261519 viste et høyere genuttrykk enn ILMN_1670282. En viktig mulighet er, som nevnt ovenfor, at ILMN_226159 i tillegg plukker opp transkripter fra *ZNF273*, men en annen mulighet er at det i tymus kanskje befinner seg en variant som har en kortere 3-UTR (som ILMN_1670282 binder) som utrykkes lavere, og en variant med lengre 3-UTR (som ILMN_2261519 binder) som utrykkes høyere. Dette ville vi imidlertid sett på dypsekvenseringsdataene av *AIRE*, noe som ikke var tilfellet.

Proben for *ICOSLG*, ILMN_1675671, viste seg å binde fullstendig (100 %) til tre av de fire rapporterte transkriptene. Det som kan være en svakhet ved dette er at proben vi har valgt å bruke ikke fanger opp alle variantene av genet vi er interessert i å se nærmere på og at vi da kan gå glipp av verdifull informasjon om genuttrykk for prøvene våre i *ICOSLG*. Derfor kan ikke resultatene av denne analysen fastslå noe med 100 % sikkerhet, men den kan allikevel gi indikasjoner på hvordan ting fungerer.

Det at *ICOSLG* viste seg å ha fire ulike transkripter i følge RefSeq, som alle var validerte, betyr at det finnes flere ulike varianter av *ICOSLG* transkriptet hvor kanskje den ene varianten kan ha et litt annet ekson enn den andre. Dette er grunnet alternativ spleising, en biologisk prosess som øker repertoaret av ulike proteiner som kan produseres fra et gen.

Det ble ikke funnet polymorfismer i de fire ulike probene, noe som betyr at det er stor sannsynlighet for at proben klarer å feste seg godt og gi gode analyseresultater. Noe som allikevel kan bidra til å gjøre resultatene litt usikre er at det ble bare sett etter polymorfismer i probesekvensene med MAF > 1 %. Det vil si at det allikevel er en liten sannsynlighet for at det kan ha vært polymorfismer med MAF < 1 % i probesekvensen som har gjort slik at proben ikke har klart å feste seg ordentlig og da ikke har klart å fange opp de aktuelle transkriptene, som gjør at vi får et svakere resultat enn det som egentlig er realiteten.

5.3.3 Styrke i analyse og prøvematerialet

5.3.3.1 Assosiasjonsanalyser av stratifisert og ikke-stratifisert material

Den statistiske styrken vår ble beregnet før prosjektstart til å ligge på ca. 80 % for analyser av alle RA pasienter mot kontroller. Vår totalanalyse av RA individer mot kontroller inneholder et stort prøvemateriale som gir relativt gode estimater for frekvensen av de undersøkte genetiske variantene. Antall RA individer reduseres veldig når dataene blir stratifisert i subgruppeanalyse (ACPA og SE), dette vil redusere styrken i materialet vårt betraktelig. På den annen side kan en inndeling i subfenotype gjøre at man studerer en mer homogen pasientgruppe og hvis den genetiske risikovarianten bare bidrar i den ene subgruppen vil effektstørrelsen (OR) øke i de stratifiserte analysene, noe som igjen kan virke positivt på styrken og gjøre at en slik analyse faktisk gir mer styrke enn en totalanalyse.

Assosiasjonsresultatene og det statistiske signifikansnivået er ikke korrigert for multippel testing. En Bonferroni korreksjon av antall SNPer analysert ville redusert signifikansnivået til $p_c<0,01$. Bare rs760426 ville da blitt funnet signifikant assosiert i den initielle analysen. I tillegg utførte vi en rekke stratifiserte analyser, slik at det totale antall tester som er utført er betraktelig høyere. Generelt har man i feltet gått bort fra korreksjon i enkeltstudier, men velger heller å sette en signifikansgrense på $p<5 *10^{-8}$ for at en variant skal erklæres å være et risikolocus. Siden dette prosjektet tok utgangspunkt i en studie som allerede hadde vist en slik genom-vid signifikans, ble p-verdiene ikke korrigert.

For å undersøke om resultatene er reproduserbare burde assosiasjonen med RA blitt analysert for de samme SNPene og i lik befolkning, men for flere ulike datasett med individer. Om da alle de ulike datasettene ga samme resultat som for våre prøver vil man ha større styrke rundt resultatet.

Videre kunne man prøve å finne de kausale allelvariantene og utføre funksjonelle studier for å kartlegge polymorfismenes funksjon.

5.3.3.2 eQTL-analyse/genekspresjon

Det at prøvematerialet i eQTL-analysen bare består av 42 tymusprøver gjør at muligheten for at noen av de selekterte SNPene fortsatt kan ha en regulatorisk funksjon på genekspresjonen til *AIRE* eller *ICOSLG* er der. Prøvematerialet er ikke stort, men det kan allikevel brukes til å gi en indikasjon på SNPenes funksjon og deres assosiasjon. For å styrke analysen og resultatet burde mer prøvemateriale bli inkludert i analysen og den burde bli utført på flere ulike kohorter for å se om resultatene er reproduserbare.

5.3.3.3 RNA sekvensering

Det faktum at det kun er brukt to tymusprøver til denne undersøkelsen, som er et lite prøvemateriale, bør ikke gi upålitelige resultatet i forhold til at vi bare skulle se på repertoaret av transkripter. Det som derimot kan være verdt å tenke på er at repertoaret kan vise seg å være enda større om man undersøker flere prøver. Av den grunn burde vi ha brukt et mye større prøvemateriale for å kunne få et større overblikk over om det dukker opp enda flere varianter hos andre individer enn de vi har fanget opp, om resultatene er reproduserbare til den grad at man ser at noen av de samme spleisevariantene dukker opp hos flere individer eller om disse varierer veldig fra person til person.

5.4 Bioinformatiske analyser av de selekterte SNPene

Ingen av de selekterte SNPene viste seg å ha en kjent genregulerende rolle ut i fra RegulomeDB. Det betyr derimot ikke at de ikke er regulerende, da alle SNPene endte opp med en RegulomeDB score på 4-5 hvor 1 er veldig regulerende og 6 er lite regulerende. Ut i fra databasen kan dette bety at SNPene er lokalisert i et transkripsjonsfaktorsete og/eller i en DNase Peak, noe som kan føre med seg at de forstyrrer transkripsjonen i genet. I tillegg er det viktig å ta i betraktning at mengden informasjon er begrenset og at stadig ny kunnskap rundt regulatoriske funksjoner til genomet avdekkes.

At ingen av SNPene viste seg å være regulatoriske generelt kan derimot også stemme overens med våre resultater fra tidligere hvor vi kom fram til at ingen av SNPene var regulatoriske i verken *AIRE* eller *ICOSLG*.

Det at ingen av SNPene viste seg å være regulerende er ikke utslagsgivende på om de kan assosieres med autoimmune sykdommer eller ikke, fordi de kan vise seg og bare være en liten regulerende del i en stor sammenheng som er nødvendig for at man utvikler en sykdom (som f.eks. i en multifaktoriell sykdom som RA).

6. Konklusjon

To polymorfismer i *AIRE* genet, rs760426 og rs3788113, viste seg å ha en allelvariant (G) som kan assosieres med redusert risiko for å utvikle RA enn den andre allelvarianten (A) i norsk befolkning. Assosiasjonsanalyser av de to subgruppene i RA viste at odds ratioen var sterkere i ACPA + RA individer for begge de assosierte SNPene og at ingen av de selekterte SNPene viste assosiasjon med ACPA- RA, kan tyde på at assosiasjonen er knyttet til ACPA+ RA individer.

Ingen av de selekterte SNPene viste assosiasjon med genuttrykket av *AIRE* eller *ICOSLG* i 42 tymusprøver. Men RNA sekvenseringsdata av to tymusprøver tydet på at det finnes flere transkripter av både *AIRE*- og *ICOSLG* genet i tymus som ennå ikke er rapportert.

7. Veien videre

Videre arbeid med forskningsprosjektet bør gå ut på å få genotypet SNPen rs878081 som er funnet signifikant assosiert med RA i spansk populasjon (55), samt å få en komplett genotyping av ekson SNPen rs1800520 for å undersøke om det er noen assosiasjon mellom disse og RA i norsk befolkning.

Eventuelt inkludere flere SNPer i en ny seleksjon med litt andre kriterier for å fortsette undersøkelsen av om polymorfismer i *AIRE* kan assosieres med RA.

Funnene gjort i denne masteroppgaven må bekreftes i større studier. Ideelt sett bør det utføres en screening av all genetisk variasjon i kromosomområdet både i japanske og europeiske/norske individer for å finne den kausale varianten. En større studie ville også gi svar på hvorvidt assosiasjonen er knyttet til bare ACPA+ RA eller om den også er tilstede i ACPA- RA. Til slutt vil funksjonelle studier være viktig for å se på hvordan variantene påvirker den biologiske funksjonen.

8. Litteratur

1. Helseinformatikk N. Autoimmune sykdommer: Norsk Helseinformatikk; 2013 [09.09.13]; [cited 11.03.15]. Available from: http://nhi.no/pasienthandboka/sykdommer/allergi/autoimmune-sykdommer-2527.html?page=3.
2. Om autoimmune sykdommer: Diavista; 2010 [November 2010]; [cited 12.03.15]. Available from: http://www.diavista.com/om-autoimmune-sykdommer/.
3. J. Roddick. Autoimmune Disease Healthline; [cited 11.03.15]. Available from: http://www.healthline.com/health/autoimmune-disorders#Treatment6.
4. Hemminki K., Li X., Sundquist K., Sundquist J. Shared familial aggregation of susceptibility to autoimmune diseases. Arthritis & Rheumatism. 2009;60(9):2845-7.
5. Cusick M.F. Libbey J.E., Fujinami R.S. Molecular Mimicry as a Mechanism of Autoimmune Disease. Clin Rev Allergy Immunol.: PMC; 2013.
6. roarp. Kroppen angriper: Lommelegen; 2000 [01.01.00]; [cited 12.03.15]. Available from: http://www.lommelegen.no/artikkel/kroppen-angriper.
7. Helseinformatikk N. Leddgikt (revmatoid artritt): Norsk Helseinformatikk; 2012 [updated 27.02.12]; [cited 02.03.15]. Available from: http://nhi.no/pasienthandboka/sykdommer/muskel-skjelett/leddgikt-oversikt-3184.html?page=all.
8. Dreamstime. Rheumatoid arthritis of hand: Dreamstime; [cited 05.05.15]. Available from: http://www.dreamstime.com/royalty-free-stock-photo-rheumatoid-arthritis-hand-image22366005.
9. Waaler E. On the occurrence of a factor in human serum activating the specific agglutintion of sheep blood corpuscles. 1939. APMIS. 2007;115(5):422-38.
10. Mierau R. Genth E. Diagnosis and prognosis of early rheumatoid arthritis, with special emphasis on laboratory analysis. Clin Chem Lab Med. 2006;44(2):138-43.
11. Maehlen M.T., Olsen I.C., Andreassen B.K., Viken M.K., Jiang X., Alfredsson L., Källberg H., Brynedal B., Kurreeman F., Daha N., Toes R., Zhernakova A., Gutierrez-Achury J., de Bakker P.I.W., Martin J., Teruel M., Gonzalez-Gay M.A., Rodríguez-Rodríguez, Balsa A., Uhlig T., Kvien T.K, Lie B.A. Genetic risk scores and number of autoantibodies in patients with rheumatoid arthritis. Ann Rheum Dis. 2015;74:762-8.
12. Okada Y., Wu D., Trynka G., Raj T., Terao C., Ikari K., Kochi Y., Ohmura K., Suzuki A., Yoshida S., Graham R.R., Manoharan A., Ortmann W., Bhangale T., Denny J.C., Carroll R.J., Eyler A.E., Greenberg J.D., Kremer J.M., Pappas D.A., Jiang L., Yin J., Ye L., Su D-F., Yang J., Xie G., Keystone E., Westra H-J., Esko T., Metspalu A., Zhou X., Gupta N., Mirel D., Stahl E.A., Diogo D., Cui J., Liao K., Guo M.H., Myouzen K., Kawaguchi T., Coenen M.J.H., van Riel P.L.C.M., van de Laar M.A.F.J., Guchelaar H-J., Huizinga T.W.J., Dieudé P., Mariette X., Bridges Jr S.L., Zhernakova A., Toes R.E.M., Tak P.P., Miceli-Ricard C., Bang S-Y., Lee H-S., Martin J., Gonzales-Gay M.A., Rodríguez-Rodríguez L., Rantapää-Dahlqvist S., Ärlestig L., Choi H.K., Kamatani Y., Galan P., Lathrop M., the RACI consortium, the GARNET consortium, Eyre S., Bowes J., Barton A., de Vries N., Moreland L.W., Criswell L.A., Karlson E.W., Taniguchi A., Yamada R., Kubo M., Liu J.S., Bae S-C., Worthington J., Padyukov L., Klareskog L., Gregersen P.K., Raychaudhuri S., Stranger B.E., Jager P.L.D., Franke L., Visscher P.M., Brown M.A., Yamanaka H., Mimori T., Takahashi A., Xu H., Behrens T.W., Siminovitch K.A., Momohara S., Matsuda F., Yamamoto K., Plenge R.M. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376-81.

13. Seegobin S.D., Ma M.H.Y., Dahanayake C., Cope A.P., Scott D.L., Lewis C.M., Scott I.C. ACPA-positive and ACPA-negative rheumatoid arthritis differ in their requirements for combination DMARDs and corticosteroids: secondary analysis of randomized controlled trial. *Arthritis Research & Therapy*. 2014;16.
14. Liao K.P., Alfredsson L., Karlson E.W. Environmental influences on risk for rheumatoid arthritis. Curr Opin Rheumatol.2009.
15. Kåss E., Kvien T.K. Leddgikt I Store medisinske leksikon, 2009 [updated 13.02.09]; [cited 03.03.15]. Available from: https://sml.snl.no/leddgikt.
16. MacGregor A.J., Snieder H., Rigby A.S., Koskenvuo M., Kaprio J., Aho K., Silman A.J. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis & Rheumatism. 2000;43(1):30-7.
17. Vyse T.J., Todd J.A. Genetic Analysis of Autoimmune Disease. Cell. 1996;85:311-8.
18. Silman A.J., MacGregor A.J., Thomson W., Holligan S., Carthy D., Farhan A., Ollier W.E. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. Br J Rheumatol. 1993;32(10):903-7.
19. Gregersen P.K., Silver J., Winchester R.J. The shared epitope hypthesis. An Approach to Understanding The Molecular Genetics of Susceptibility to Rheumatoid Arthritis. Arthritis & Rheumatism. 1987;30(11).
20. Padyukov L., Silva C., Stolt P., Alfredsson L., Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. Arthritis Rheum. 2004;50(10):3085-92.
21. Klareskog L., Malmström V., Lundberg K., Padyukov L., Alfredsson L. Smoking, citrullination and genetic variability in the immunopathogenesis of rheumatoid arthritis. Seminars in Immunology. 2011;23(2):92-8.
22. Genetic Analysis Tools Help Define Nature and Nurture in Complex Disorders National Human Genome Research Institute: National Human Genome Research Institute; [updated 03.05.13]; [cited 16.03.15]. Available from: http://www.genome.gov/10000865.
23. Zhernakova A., Withoff S., Wijmenga C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases. NatRevEndocrinol. 2013;9:646-59.
24. Ricaño-Ponce I., Wijmenga C. Mapping of Immune-Mediated Disease Genes. *Annual Review Genomics and Human Genetics*. 2013;14:325-53.
25. What are genome-wide association studies? Genetics Home Reference: U.S. National Library of Medicine; [cited 10.03.15]. Available from: http://ghr.nlm.nih.gov/handbook/genomicresearch/gwastudies.
26. Hindorff L.A., MacArthur J., Morales J., Junkins H.A., Hall P.N., Klemm A.K., Manolio T.A. Catalog of Published Genome-Wide Association Studies: National Human Genome Reasearch Institute; [cited 20.02.15]. Available from: http://www.genome.gov/gwastudies/.
27. Yamamoto K., Okada Y., Suzuki A., Kochi Y. Genetics of rheumatoid arthritis in Asia-present and future. NatRevRheumatol. 2015.
28. Farh K.K-H., Marson A., Zhu J., Kleinewietfeld M., Housley W.J., Beik S., Shoresh N., Whitton H., Ryan R.J.H., Sishkin A.A., Hatan M., Carrasco-Alfonso M.J., Mayer D., Luckey C.J., Patsopoulos N.A., Jager P.L.D., Kuchroo V.K., Epstein C.B., Daly M.J., Hafler D.A., Bernstein B.E. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518:337-43.

29. Technologies L. Tumor Necrosis Factor (TNF) Overview: Life Technologies; [cited 09.04.15]. Available from: http://www.lifetechnologies.com/no/en/home/life-science/cell-analysis/signaling-pathways/tumor-necrosis-factor-tnf/tumor-necrosis-factor-tnf-overview.html.
30. Terao C., Yamada R., Ohmura K., Takahashi M., Kawaguchi T., Kochi Y., Human Disease Genomics Working Group, RA Clinical and Genetic Study Consortium, Okada Y., Nakamura Y., Yamamoto K., Melchers I., Lathrop M., Mimori T., Matsuda F. . The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. Human Molecular Genetics 2011;20(13):2680-5.
31. Taylor T. Thymus Gland: InnerBody; [cited 07.03.15]. Available from: http://www.innerbody.com/image_endoov/lymp04-new.html.
32. Bates K. Chapter 7: The Development of T Lymphocytes: Winona State University 2005 [updated 10.10.05]; [cited 15.03.15]. Available from: http://course1.winona.edu/kbates/Immunology/chapter7-09_000.htm.
33. imgbuddy.com. Human Thymus Gland: imgbuddy.com; [cited 07.02.15]. Available from: http://imgbuddy.com/human-thymus-gland.asp.
34. Holck P. Thymus I Store medisinske leksikon2015; [cited 05.03.15]. Available from: https://sml.snl.no/thymus.
35. The Immune System: Garland Science; 2009; [cited 01.05.15]. Available from: http://course1.winona.edu/kbates/Immunology/images/figure_07_04_01.jpg.
36. Eldershaw S.A., Sansom D.M., Narendran P. Expression and function of the autoimmune regulator (Aire) gene in non-thymic tissue. Clin Exp Immunol. 2011;163(3):296-308.
37. Genetics Home Reference. AIRE: Genetics Home Reference; 2007; [cited 15.02.15]. Available from: http://ghr.nlm.nih.gov/gene/AIRE.
38. GeneCards. Autoimmune regulator: GeneCard; [cited 05.02.15]. Available from: http://www.genecards.org/cgi-bin/carddisp.pl?gene=AIRE.
39. Dessen P. Le M.S. AIRE (autoimmune regulator): Atlas of Genetics and Cyogenetics in Oncology and Haematology; 2002 [September 2002]; [cited 22.04.15]. Available from: http://atlasgeneticsoncology.org/Genes/GC_AIRE.html.
40. Stahl E.A., Raychaudhuri S., Remmers E.F., Xie G., Eyre S., Thomson B.P., Li Y., Kurreeman F.A., Zhernakova A., Hinks A., Guiducci C., Chen R., Alfredsson L., Amos C.I., Ardlie K.G.; BIRAC Consortium, Barton A., Bowes J., Brouwer E., Burtt N.P., Catanese J.J., Coblyn J., Coenen M.J., Costenbader K.H., Criswell L.A., Crusius J.B., Cui J., de Bakker P.I., De Jager P.L., Ding B.,Emery P., Flynn E., Harrison P., Hocking L.J., Huizinga T.W., Kastner D.L., Ke X., Lee A.T., Liu X., Martin P., Morgan A.W., Padyukov L., Posthumus M.D., Radstake T.R., Reid D.M., Seielstad M., Seldin M.F., Shadick N.A., Steer S., Tak P.P., Thomson W., van der Helm-van Mil A.H., van der Horst-Bruinsma I.E., van der Schoot C.E., van Riel P.L., Weinblatt M.E., Wilson A.G., Wolbink G.J., Wordsworth B.P.; YEAR Consortium, Wijmenga C., Karlson E.W., Toes R.E., de Vries N., Begovich A.B., Worthington J., Siminovitch K.A., Gregersen P.K., Klareskog L., Plenge R.M. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nature Genetics. 2010;42(6):508-14.
41. Hui L., DelMonte T., Ranande K. Genotyping Using the Taqman Assay. Current Protocols in Human Genetics. 2008.
42. Dublin T.C. Genotyping, 2009 [updated 23.03.09]; [cited 0.03.15]. Available from: https://medicine.tcd.ie/neuropsychiatric-genetics/functional-genetics-genomics/genotyping.php.

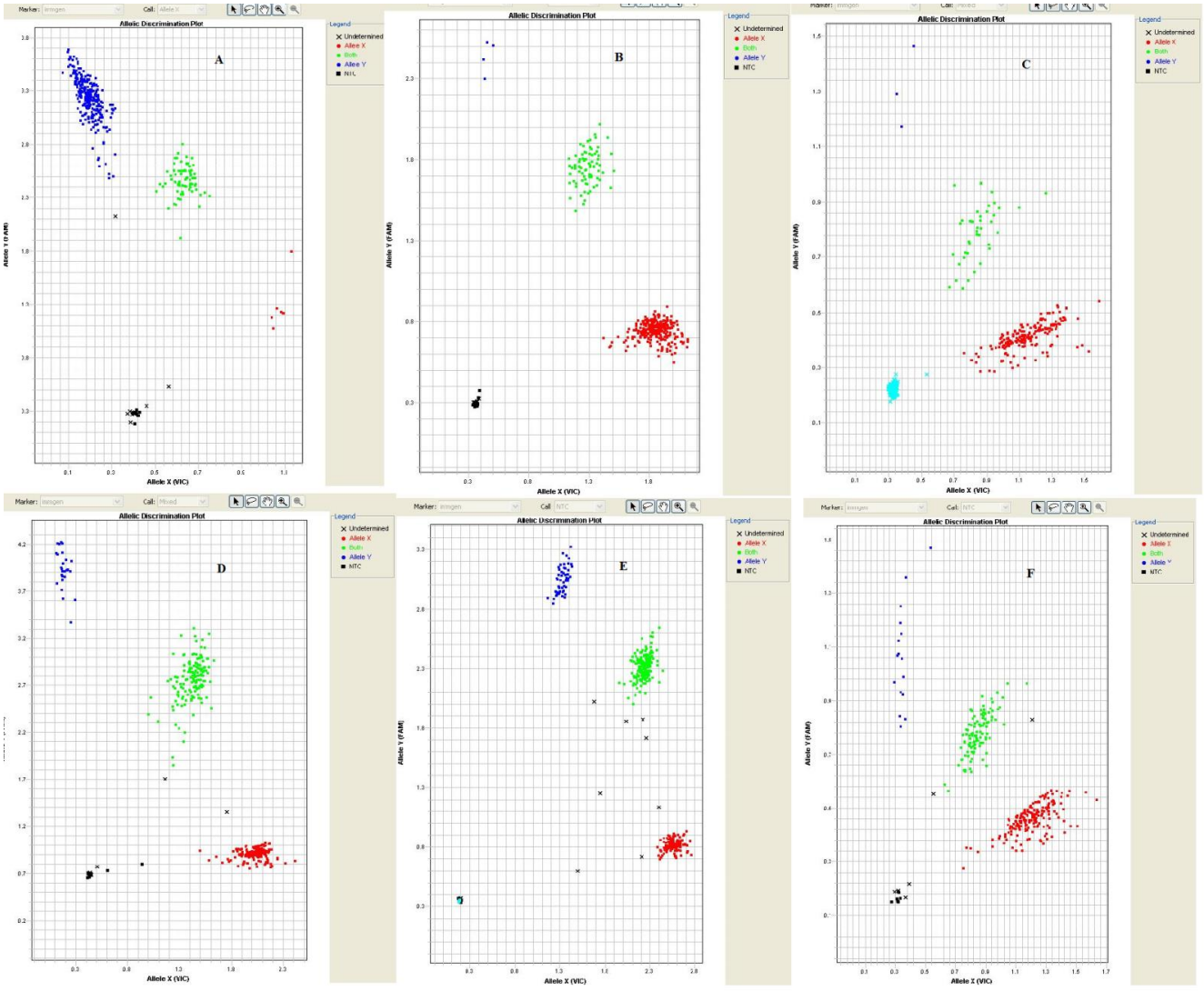
43. Sandvik A.K., Støren O., Nørsett K., Lægreid A., Børresen-Dale A.L., Myklebost O. Måling av genaktivitet med DNA-mikromatriser. Tidsskrift for Den norske Lægeforening. 2001;121(10):1225-8.
44. Wang Z., Gerstein M., Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009;10(1):57-63.
45. Barrett J.C., Fry B., Maller J., Daly M.J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics Oxford journals. 2004;21(2):263-5.
46. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. PLINK : a toolset for whole-genome association and population-based linkage analysis Am J Hum Genet. 2007;81(3):559-75.
47. Motulsky H. GraphPad Software Inc. 1984. GraphPad Prism 6.0; [cited 11.03.15]. Available from: http://www.graphpad.com/scientific-software/prism/?tab=5.
48. Jostins L., Ripke S., Weersma R.K., Duerr R.H., McGovern D.P., Hui K.Y., Lee J.C., Schumm L.P., Sharma Y., Anderson C.A., Essers J., Mitrovic M., Ning K., Cleynen I., Theatre E., Spain S.L., Raychaudhuri S., Goyette P., Wei Z., Abraham C., Achkar J.P., Ahmad T., Amininejad L., Ananthakrishnan A.N., Andersen V., Andrews J.M., Baidoo L., Balschun T., Bampton P.A., Bitton A., Boucher G., Brand S., Büning C., Cohain A., Cichon S., D'Amato M., De Jong D., Devaney K.L., Dubinsky M., Edwards C., Ellinghaus D., Ferguson L.R., Franchimont D., Fransen K., Gearry R., Georges M., Gieger C., Glas J., Haritunians T., Hart A., Hawkey C., Hedl M., Hu X., Karlsen T.H., Kupcinskas L., Kugathasan S., Latiano A., Laukens D., Lawrance I.C., Lees C.W., Louis E., Mahy G., Mansfield J., Morgan A.R., Mowat C., Newman W., Palmieri O., Ponsioen C.Y., Potocnik U., Prescott N.J., Regueiro M., Rotter J.I., Russell R.K., Sanderson J.D., Sans M., Satsangi J., Schreiber S., Simms L.A., Sventoraityte J., Targan S.R., Taylor K.D., Tremelling M., Verspaget H.W., De Vos M., Wijmenga C., Wilson D.C., Winkelmann J., Xavier R.J., Zeissig S., Zhang B., Zhang C.K., Zhao H.; International IBD Genetics Consortium (IIBDGC), Silverberg M.S., Annese V., Hakonarson H., Brant S.R., Radford-Smith G., Mathew C.G., Rioux J.D., Schadt E.E., Daly M.J., Franke A., Parkes M., Vermeire S., Barrett J.C., Cho J.H. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24.
49. Dubois P.C., Trynka G., Franke L., Hunt K.A., Romanos J., Curtotti A., Zhernakova A., Heap G.A., Adány R., Aromaa A., Bardella M.T., van den Berg L.H., Bockett N.A., de la Concha E.G., Dema B., Fehrmann R.S., Fernández-Arquero M., Fiatal S., Grandone E., Green P.M., Groen H.J., Gwilliam R., Houwen R.H., Hunt S.E., Kaukinen K., Kelleher D., Korponay-Szabo I., Kurppa K., MacMathuna P., Mäki M., Mazzilli M.C., McCann O.T., Mearin M.L., Mein C.A., Mirza M.M., Mistry V., Mora B., Morley K.I., Mulder C.J., Murray J.A., Núñez C., Oosterom E., Ophoff R.A., Polanco I., Peltonen L., Platteel M., Rybak A., Salomaa V., Schweizer J.J., Sperandeo M.P., Tack G.J., Turner G., Veldink J.H., Verbeek W.H., Weersma R.K., Wolters V.M., Urcelay E., Cukrowska B., Greco L., Neuhausen S.L., McManus R., Barisani D., Deloukas P., Barrett J.C., Saavalainen P., Wijmenga C., van Heel D.A. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010;42(4):295-302.
50. dbSNP: Database for Short Genetic Variations: National Center for Biotechnology Information; [cited 15.01.15]. Available from: http://www.ncbi.nlm.nih.gov/SNP/.
51. Dessen P., Le M.S. ICOSLG (inducible T-cell co-stimulator ligand): Atlas of Genetics and Cytogenetics in Oncology and Haematology; 2002 [September 2002]; [cited 23.04.15]. Available from: http://atlasgeneticsoncology.org/Genes/GC_ICOSLG.html.

52. Boyle A.P. Hong E.L, Hariharan M., Cheng Y., Schaub M.A., Kasowski M., Karczewski K.J., Park J., Hitz B.C., Weng S., Cherry J.M., Snyder M. Annotation of functional variation in personal genomes using RegulomeDB: Genome Research 2012; [cited 27.03.15]. Available from: http://regulomedb.org/.
53. Shao S., Li X-R., Cen H., Yin Z-S. Association of *AIRE* Polymorphisms with Genetic Susceptibility to Rheumatoid Arthritis in a Chinese Population. *Inflammation*. 2014;37(2).
54. Feng Z-J., Zhang S-L., Wen H-F., Liang Y. Association of rs2075876 polymorphism of AIRE gene with rheumatoid arthritis risk. Human Immunology. 2015:281-5.
55. García-Lozano J.R., Torres-Agrela B., Montes-Cano M.A., Ortiz-Fernández L., Conde-Jaldón M., Teruel M., García A., Núñez-Roldán A., Martín J., González-Escribano M.F. Association of the AIRE gene with susceptibility to rheumatoid arthritis in a European population: a case control study. *Arthritis Research & Therapy*. 2013;15(1).
56. Fu J., Wolfs M.G., Deelen P., Westra H.J., Fehrmann R.S., Te Meerman G.J., Buurman W.A., Rensen S.S., Groen H.J., Weersma R.K., van den Berg L.H., Veldink J., Ophoff R.A., Snieder H., van Heel D., Jansen R.C., Hofker M.H., Wijmenga C., Franke L. Unraveling the Regulatory Mechanisms Underlying Tissue-Dependent Genetic Variation of Gene Expression. PLoS Genetics. 2012;8(1).

9. Vedlegg

Vedlegg 1: Genotypingsplott

Et tilfeldig utvalg av genotypingsplott for hvert SNPassay fra TaqMan genotypingen er vist i figur 23.



Figur 23: Et tilfeldig utvalg av genotypingsplott fra TaqMan genotypingen for A) SNP assay rs2075876 (plate 2), B) SNP assay rs760426 (plate 4), C) SNP assay rs1800520 (plate 1), D) SNP assay rs4819388 (plate 3), E) SNP assay rs7282490 (plate 4), F) SNP assay rs3788113 (plate 1).

Vedlegg 2: Assosiasjonsanalyse med genetisk modell for de fem suksessfullt genotypede SNPene

Resultatet fra assosiasjonsanalysen med genetisk modell for de fem selekterte SNPene er gitt i tabell 17.

Tabell 17: Oversikt over allelvarianter for fem av de selekterte SNPene og om disse er dominante eller recessive hos RA individer mot kontroller.

SNP ID	Minor allel	Major allel	Test	Genotyper i RA	Genotyper i ktr.	p-verdi
rs7282490	G	A	GENO	147/435/337	115/406/298	0,46
rs7282490	G	A	TREND	729/1109	636/1002	0,61
rs7282490	G	A	ALLELIC	729/1109	636/1002	0,61
rs7282490	G	A	DOM	582/337	521/298	0,90
rs7282490	G	A	REC	147/772	115/704	0,26
rs4819388	T	C	GENO	68/346/502	56/321/441	0,78
rs4819388	T	C	TREND	482/1350	433/1203	0,92
rs4819388	T	C	ALLELIC	482/1350	433/1203	0,92
rs4819388	T	C	DOM	414/502	377/441	0,71
rs4819388	T	C	REC	68/848	56/762	0,64
rs2075876	A	G	GENO	10/176/741	6/167/647	0,59
rs2075876	A	G	TREND	196/1658	179/1461	0,74
rs2075876	A	G	ALLELIC	196/1658	179/1461	0,74
rs2075876	A	G	DOM	186/741	173/647	0,59
rs2075876	A	G	REC	10/917	6/814	0,45
rs3788113	G	A	GENO	27/286/603	35/279/502	0,10
rs3788113	G	A	TREND	340/1492	349/1283	0,04
rs3788113	G	A	ALLELIC	340/1492	349/1283	0,04
rs3788113	G	A	DOM	313/603	314/502	0,06
rs3788113	G	A	REC	27/889	35/781	0,13
rs760426	G	A	GENO	11/176/740	11/198/611	0,03
rs760426	G	A	TREND	198/1656	220/1420	0,01
rs760426	G	A	ALLELIC	198/1656	220/1420	0,01
rs760426	G	A	DOM	187/740	209/611	0,01
rs760426	G	A	REC	11/916	11/809	0,77

Vedlegg 3: Assosiasjonsanalyse med genetisk modell for de fem suksessfullt genotypede SNPene stratifisert for ACPA status (ACPA+)

Resultatet fra assosiasjonsanalysen med genetisk modell for de fem selekterte SNPene stratifisert for ACPA + RA status mot kontroller er gitt i tabell 18.

Tabell 18: Oversikt over allelvarianter for fem av de selekterte SNPene og om disse er dominante eller recessive hos ACPA + RA individer mot kontrollprøver.

SNP ID	Minor allel	Major allel	Test	Genotyper i ACPA + RA	Genotyper i ktr.	p-verdi
rs7282490	G	A	GENO	93/245/194	117/407/299	0,22
rs7282490	G	A	TREND	431/633	641/1005	0,41
rs7282490	G	A	ALLELIC	431/633	641/1005	0,42
rs7282490	G	A	DOM	338/194	524/299	0,96
rs7282490	G	A	REC	93/439	117/706	0,10
rs4819388	T	C	GENO	36/198/296	56/323/443	0,76
rs4819388	T	C	TREND	270/790	435/1209	0,57
rs4819388	T	C	ALLELIC	270/790	435/1209	0,57
rs4819388	T	C	DOM	234/296	379/443	0,48
rs4819388	T	C	REC	36/494	56/766	0,99
rs2075876	A	G	GENO	10/89/437	6/169/652	0,04
rs2075876	A	G	TREND	109/963	181/1473	0,52
rs2075876	A	G	ALLELIC	109/963	181/1473	0,52
rs2075876	A	G	DOM	99/437	175/652	0,23
rs2075876	A	G	REC	10/526	6/821	0,06
rs3788113	G	A	GENO	18/155/358	35/279/502	0,09
rs3788113	G	A	TREND	191/871	349/1283	0,03
rs3788113	G	A	ALLELIC	191/871	349/1283	0,03
rs3788113	G	A	DOM	173/358	314/502	0,03
rs3788113	G	A	REC	18/513	35/781	0,41
rs760426	G	A	GENO	6/95/435	11/198/616	0,02
rs760426	G	A	TREND	107/965	220/1430	0,01
rs760426	G	A	ALLELIC	107/965	220/1430	0,01
rs760426	G	A	DOM	101/435	209/616	0,01
rs760426	G	A	REC	6/530	11/814	0,73

Vedlegg 4: Assosiasjonsanalyse med genetisk modell for de fem suksessfullt genotypede SNPene stratifisert for ACPA status (ACPA-)

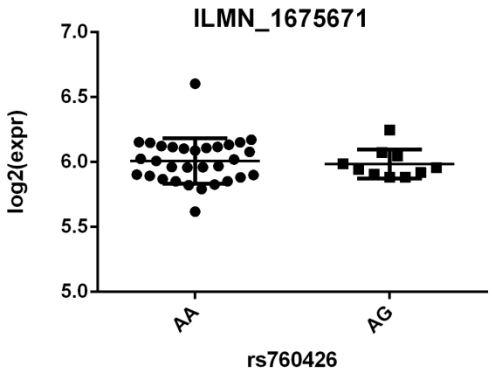
Resultatet fra assosiasjonsanalysen med genetisk modell for de fem selekterte SNPene stratifisert for ACPA - RA status mot kontroller er gitt i tabell 19.

Tabell 19: Oversikt over allelvarianter for fem av de selekterte SNPene og om disse er dominante eller recessive hos ACPA – RA individer mot kontrollprøver. (NA vil si at PLINK ikke har klart å beregne noe resultat).

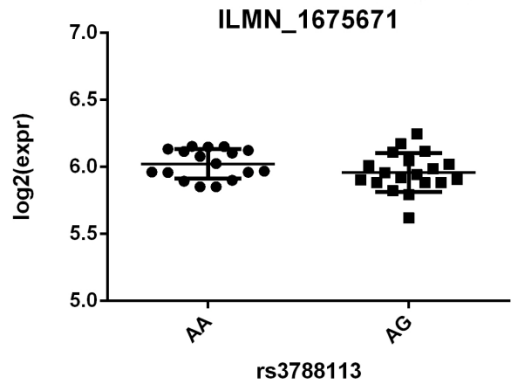
SNP ID	Minor allel	Major allel	Test	Genotyper i ACPA- RA	Genotyper i ktr.	p-verdi
rs7282490	G	A	GENO	47/162/121	117/407/299	0,99
rs7282490	G	A	TREND	256/404	641/1005	0,94
rs7282490	G	A	ALLELIC	256/404	641/1005	0,95
rs7282490	G	A	DOM	209/121	524/299	0,91
rs7282490	G	A	REC	47/283	117/706	0,99
rs4819388	T	C	GENO	26/133/172	56/323/443	0,75
rs4819388	T	C	TREND	185/477	435/1209	0,47
rs4819388	T	C	ALLELIC	185/477	435/1209	0,47
rs4819388	T	C	DOM	159/172	379/443	0,55
rs4819388	T	C	REC	26/305	56/766	0,53
rs2075876	A	G	GENO	0/73/261	6/169/652	NA
rs2075876	A	G	TREND	73/595	181/1473	0,99
rs2075876	A	G	ALLELIC	73/595	181/1473	0,99
rs2075876	A	G	DOM	73/261	175/652	NA
rs2075876	A	G	REC	0/334	6/821	NA
rs3788113	G	A	GENO	8/112/209	35/279/502	0,32
rs3788113	G	A	TREND	128/530	349/1283	0,29
rs3788113	G	A	ALLELIC	128/530	349/1283	0,30
rs3788113	G	A	DOM	120/209	314/502	0,53
rs3788113	G	A	REC	8/321	35/781	0,13
rs760426	G	A	GENO	5/69/260	11/198/616	0,47
rs760426	G	A	TREND	79/589	220/1430	0,32
rs760426	G	A	ALLELIC	79/589	220/1430	0,33
rs760426	G	A	DOM	74/260	209/616	0,25
rs760426	G	A	REC	5/329	11/814	0,83

Vedlegg 5: Grafisk fremstilling av genekspresjonsnivå

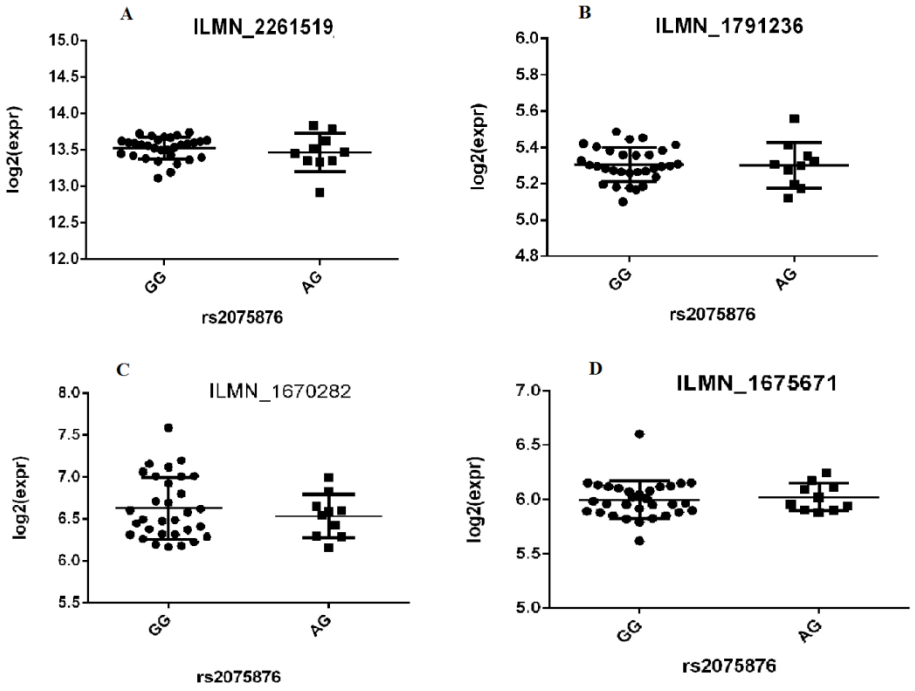
Under er de grafiske fremstillingene av genekspresjonsnivået for *AIRE* og *ICOSLG* probene for de resterende SNPene på GraphPad Prism 6.0, figur 24-29.



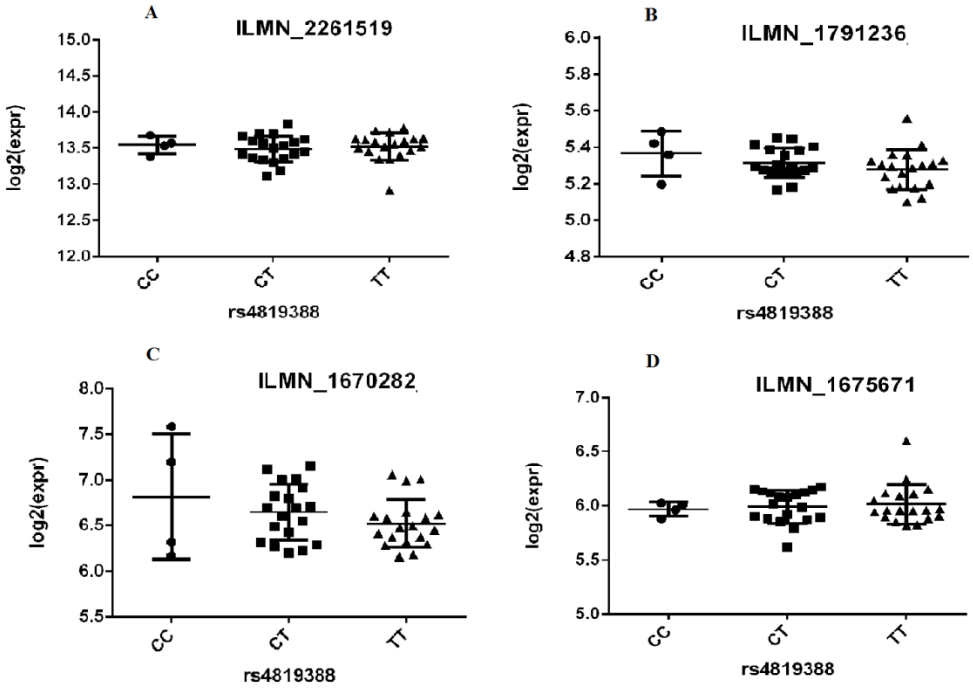
Figur 24: Log2 transformerte genuttrykksdata (for probe ILMN_1675671(*ICOSLG*)) plottet mot genotypedata for SNP rs760426 i 42 tymusprøver (p-verdi = 0,6717).



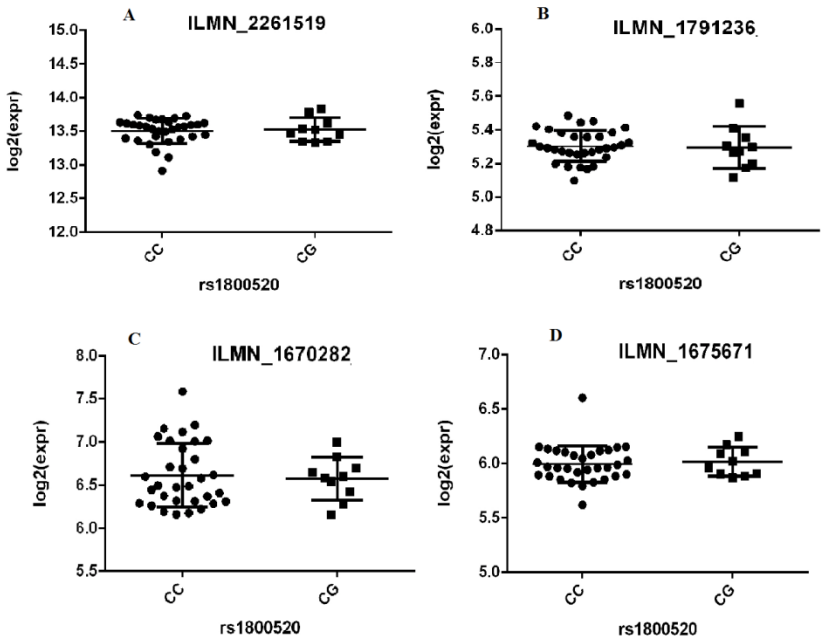
Figur 25: Log2 transformerte genuttrykksdata (for probe ILMN_1675671(*ICOSLG*)) plottet mot genotypedata for SNP rs3788113 i 42 tymusprøver (p-verdi = 0,1546).



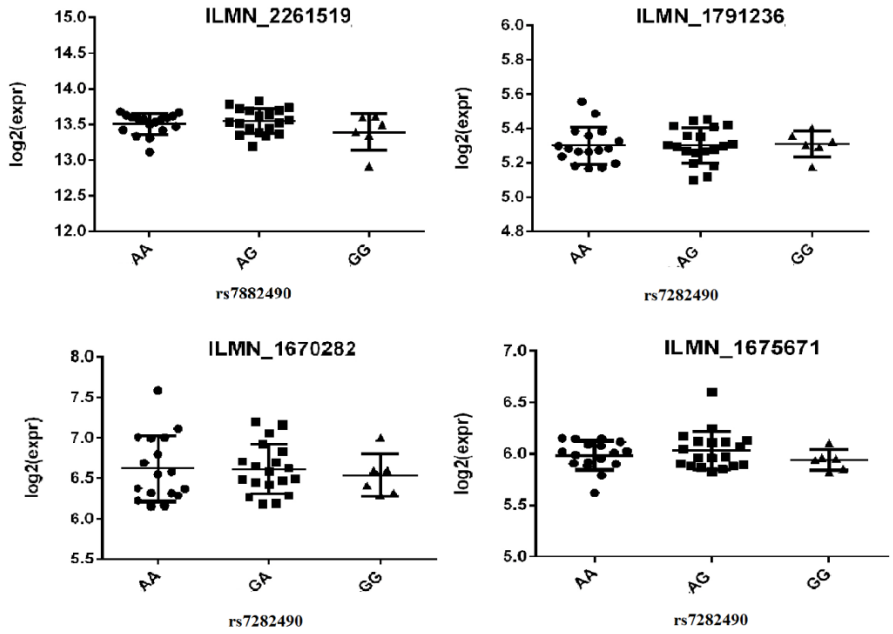
Figur 26: Log2 transformerte genuttrykksdata plottet mot genotypedata for SNP rs2075876 i 42 tymusprøver for probe A) ILMN_ 2261519 (*AIRE*) med p-verdi = 0,4054, B) ILMN_1791236 (*AIRE*) med p-verdi = 0,9233, C) ILMN_1670282 (*AIRE*) med p-verdi = 0,6078, D) ILMN_1675671 (*ICOSLG*) med p-verdi = 0,5793.



Figur 27: Log2 transformerte genuttrykksdata plottet mot genotypedata for SNP rs4819388 i 42 tymusprøver for probe A) ILMN_ 2261519 (*AIRE*) med p-verdi =0,6379, B) ILMN_1791236 (*AIRE*) med p-verdi = 0,2451, C) ILMN_1670282 (*AIRE*) med p-verdi = 0,4707, D) ILMN_1675671 (*ICOSLG*) med p-verdi = 0,9434.



Figur 28: Log2 transformerte genuttrykksdata plottet mot genotypedata for SNP rs1800520 i 42 tymusprøver for probe A) ILMN_ 2261519 (*AIRE*) med p-verdi = 0,7880, B) ILMN_1791236 (*AIRE*) med p-verdi = 0,6799, C) ILMN_1670282 (*AIRE*) med p-verdi = > 0,9999, D) ILMN_1675671 (*ICOSLG*) med p-verdi = 0,6724.

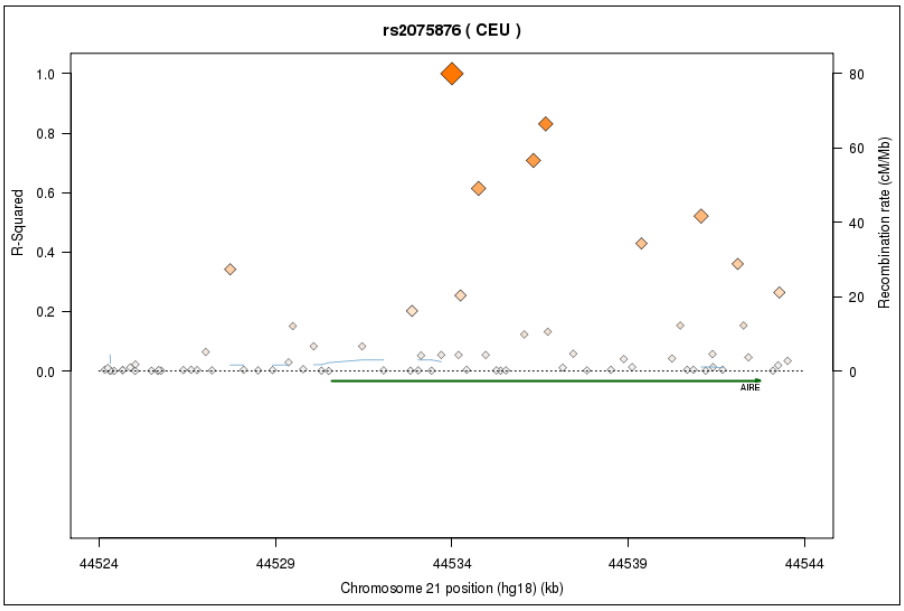


Figur 29: Log2 transformerte genuttrykksdata plottet mot genotypedata for SNP rs7282490 i 42 tymusprøver for probe A) ILMN_ 2261519 (*AIRE*) med p-verdi = 0,3823, B) ILMN_1791236 (*AIRE*) med p-verdi = 0,7728, C) ILMN_1670282 (*AIRE*) med p-verdi = 0,8698, D) ILMN_1675671 (*ICOSLG*) med p-verdi = 0,4450.

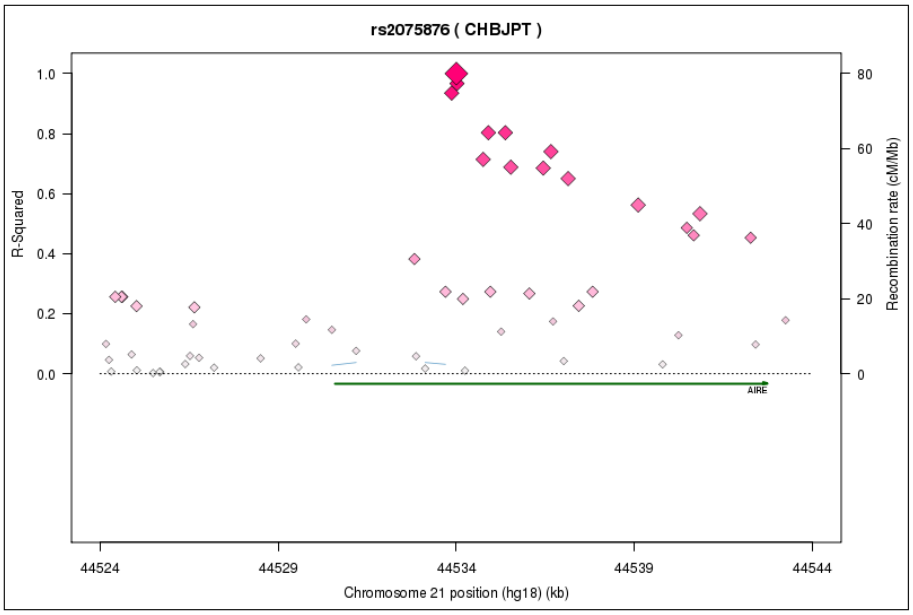
Vedlegg 6: Koblingsulikevektmønster

Koblingsulikevekt mellom SNPen rs2075876 (funnet assosiert med RA i japansk populasjon) og SNPer i 1000 Genomes, og SNPen rs878081 (funnet assosiert med RA i spansk populasjon) og SNPer i 1000 Genomes i hhv. europeisk og asiatisk befolkning, se figur 30 og 31.

A

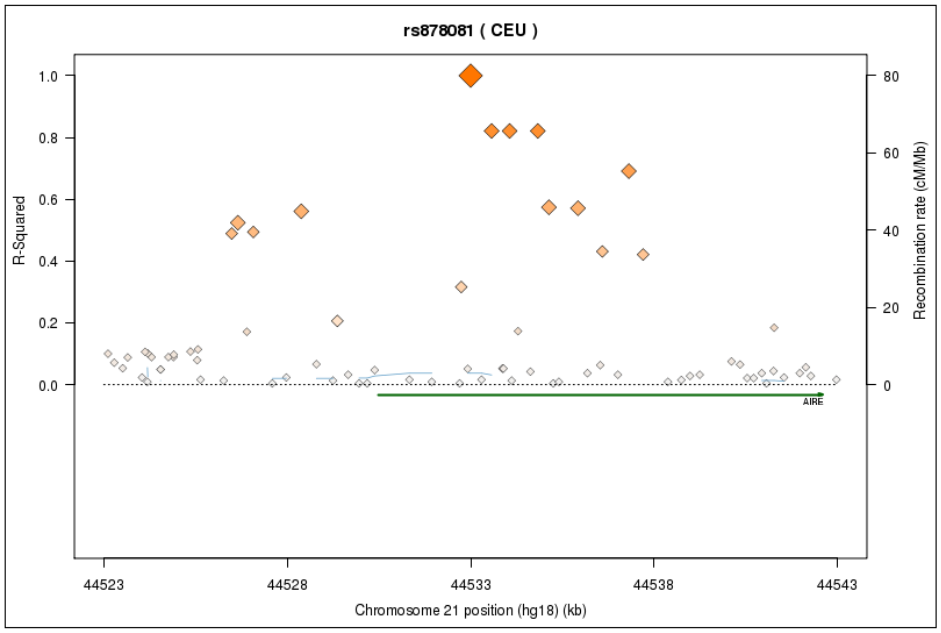


B

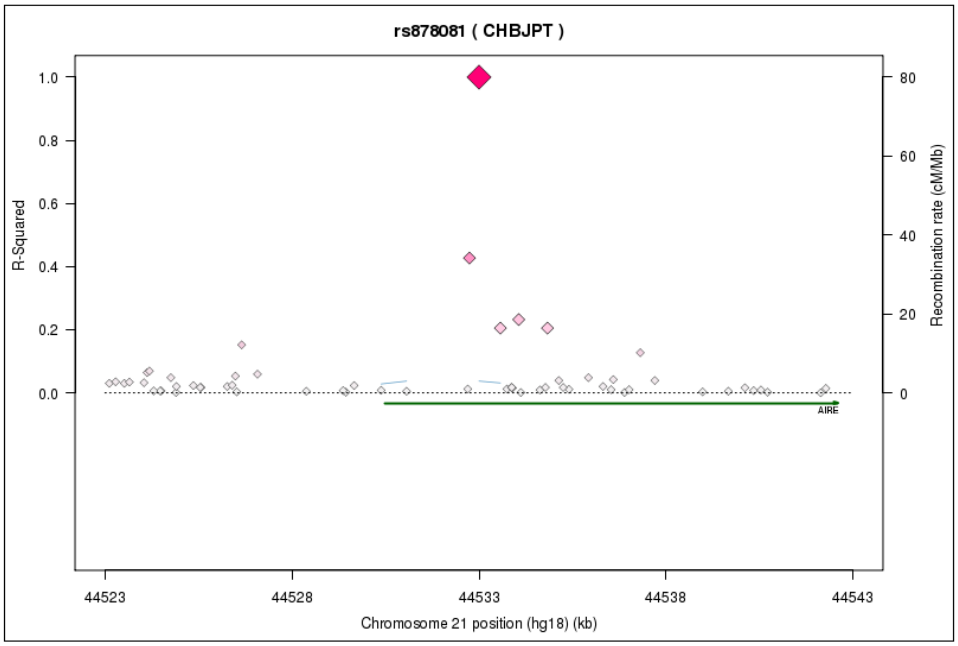


Figur 30: Koblingsulikevekt mellom rs2075876 og SNPer i 1000 Genomes for hhv. A) europeisk og B) asiatisk befolkning.

A



B



Figur 31: Koblingsulikevekt mellom rs878081 og SNPer i 1000 Genomes for hhv. A) europeisk og B) asiatisk befolkning.

Norges miljø- og
biovitenskapelige
universitet

Postboks 5003
NO-1432 Ås
67 23 00 00
www.nmbu.no