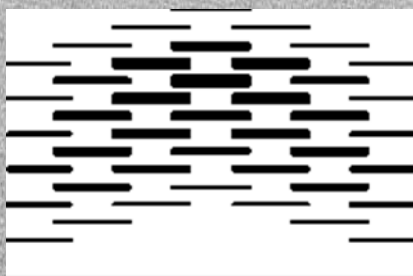


Undersøkelse av genekspresjon i tymus korrelert med risikovarianter assosiert med autoimmune sykdommer

Nimo Hatinoor

2012



OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES



Undersøkelse av genekspresjon i tymus korrelert med risikovarianter assosiert med autoimmune sykdommer

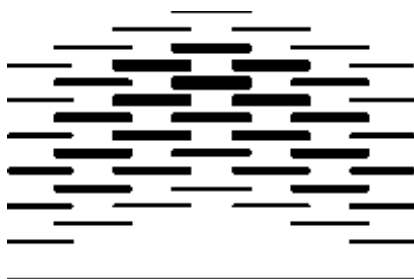
Av

Nimo Hatinoor

Master i Biomedisin
Fakultet for Helsefag
2012

Avhandling levert for Mastergrad i Biomedisin (60 ECTS)
Oslo universitetssykehus, Institutt for Klinisk Medisin
Avdeling for Medisinsk Genetikk

Veileder: Prof. Benedicte A. Lie



OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES



Forord

Prosjektet som presenteres i denne avhandlingen er utført på Institutt for Klinisk Medisin, Avdeling for Medisinsk Genetikk, Oslo Universitetssykehus (Ullevål) fra august 2011 til august 2012, for Mastergrad i Biomedisin på Høgskolen i Oslo og Akershus, Fakultet for Helsefag.

Først vil jeg takke prosjektgruppen på Medisinsk Genetikk, Immungenetikk ved autoimmune sykdommer, for å ha inkludert meg i sine daglige aktiviteter ved avdelingen, og for å ha delt sin kunnskap med meg i arbeidet med denne oppgaven.

Jeg vil takke spesielt min veileder Benedicte A. Lie for utmerkede timer med veiledning, uvurderlige innspill, råd og for all sin vennlighet. Ved å gi meg rask og konstruktiv tilbakemelding, selv i sin fritid, har det bidratt til å danne denne oppgaven. Takk for at du delte din kunnskap med meg, og lærte meg hva forskning virkelig handler om.

Jeg vil også uttrykke min takknemlighet til Marte K. Viken og Silja S. Amundsen, for sin tålmodighet gjennom alle de bioinformatiske verktøyene og for å ha delt sin kompetanse med meg både i og utenfor laboratoriet, noe som har ført til mange fruktbare diskusjoner.

En stor takk til Siri T. Flåm som ønsket meg velkommen til Medisinsk genetikk, og for all hjelp på laboratoriet. Jeg vil også takke Kristian Holm for å ha delt sin kunnskap om SNP exp., og utførelsen av analysen.

Til slutt vil jeg takke min familie for all støtte og oppmuntring. En spesiell takk til min mor, som har støttet meg gjennom hele livet, takk for all din oppmuntring og kjærlighet. Min hjerteligste takk går til min mann Ahmed for uendelig tålmodighet og oppmuntring. Takk for din støtte og din tro på meg.

Oslo, 15. August 2012

Nimo Hatinoor

Sammendrag

Autoimmune sykdommer (AIS) er en gruppe sykdommer hvor immunsystemet feilaktig går til angrep på eget vev. Immunsystemet er ideelt utdannet i tymus til ikke å angripe eget vev. AIS rammer omtrent 5 % av befolkningen, sykdommene er et resultat av en virkning mellom flere risikofaktorer fra arv og miljø. Dette prosjektet går ut på å karakterisere genekspressjonen i tymus, og undersøke hvordan genetiske varianter som disponerer for AIS påvirker denne genekspressjonen. De autoimmune sykdommene som ble fokusert på i dette prosjektet var type 1 diabetes, reumatoid artritt, systemisk lupus erythematosus, multippel sklerose, ankyloserende spondylitt (AS), cøliaki, inflammatorisk tarmsykdom, crohns sykdom, ulcerøs kolitt og psoriasis.

I studiet gjennomført i dette prosjektet ble 42 tymusprøver brukt. Alle kjente AIS risiko SNPs ble selektert (N=256) og genotypet med Immunochip, en array designet spesielt for å genotype AIS risiko genregioner. Helgenom genekspressjon ble undersøkt ved Illumina human WG-6v3. Genotypedata for AIS risikoSNPene ble korrelert mot transkripsjonsnivå av Microarray prober innenfor et område på +/- 1 Mb. 10 risiko SNP'er utenfor kromosom 6 (HLA komplekset) viste signifikante assosiasjoner ($p < 5 \times 10^{-5}$) til forskjell i ekspresjonsnivå mellom ulike genotyper. Alle transkriptene lå mindre enn 100 Kb unna SNP'en. ERAP1/ERAP2-genregionen viste konsistente funn ved at flere SNP'er i regionen var korrelert med flere prober i disse genene. For å undersøke ERAP1/ERAP2-genekspressjonen ble det utført TaqMan genekspressjonsanalyser, som videre ble korrelert mot genotypene. Korrelasjonen ble beregnet med Mann-Whitney U-test og Kruskal Wallis test.

Risiko SNPene i ERAP1/ERAP2-regionen er rapportert å være assosiert med AS (ERAP1 SNPene rs30187 og rs27434) og Ps (ERAP1 SNP'en rs27524). Korrelasjon mellom genekspressjon og genotype viste at ERAP1 og ERAP2 genene er signifikant assosiert med risiko SNPene rs30187 ($p < 0,0001$), rs27524 ($p < 0,0001$), rs27434 ($p = 0,0010$) og rs2248374 ($p < 0,0001$). På grunn av ufullstendig koblingsulikevekt mellom ERAP1 SNPene (rs27524, rs27434) og ERAP2 SNP'en (rs2248374), vil tymus-prøvene for denne proben, fordele seg ulikt i de ulike genotypegruppene avhengig av hvilken SNP man ser på. ERAP1 SNP'en rs30187 gir mest signifikant ekspresjonsforskjeller, og er mest assosiert med AS.

Resultatene førte til en bedre forståelse for samspillet mellom ERAP1/ERAP2 SNP polymorfismene ved avdekking av deres biologiske og patologiske virkning, og hvordan sammenhengen (LD) mellom SNPene kan regulere genekspressjonen av gener assosiert med AIS. Studien viste at AIS risikopolymorfismer ser ut til å påvirke ekspresjonsnivå av visse immunologiske gener i tymus.

Abstract

Autoimmune diseases (AID) are diseases where the immune system wrongly attacks its own tissues. The immune system is ideally educated in thymus not to attack own tissues. AID affect approximately 5% of the population, and the diseases are a result of an effect of several risk factors both genetic and environmental. This project aimed to characterize gene expression in thymus, and explore how genetic variants that predispose to AID affect this gene expression. The AID that were focused on were type 1 diabetes, rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, ankylosing spondylitis (AS), celiac disease, inflammatory bowel disease, Crohn's disease, ulcerative colitis and psoriasis.

There was 42 thymus samples used in the study conducted in this project. All known AID risk SNPs were selected (N = 256) and genotyped by Immunochip, an array designed specifically to genotype AID risk gene regions. Whole genome gene expression was assessed by Illumina Human WG-6v3. Genotype data for AID risk SNPs were correlated with the transcription level of Microarray Probes within a range of + / - 1 Mb. 10 risk SNPs outside chromosome 6 (HLA complex) showed significant associations ($p < 5 \times 10^{-5}$) to differences in expression levels between different genotypes at AID loci. All transcripts were less than 100 Kb away from the SNP. SNPs in the ERAP1/ERAP2- gene region showed consistent findings as several SNPs in the region were correlated with multiple probes in these genes. Further investigation of ERAP1/ERAP2- gene expression was performed by TaqMan gene expression assays. The correlation was calculated using Mann-Whitney U test and Kruskal Wallis test.

Risk SNPs in the ERAP1/ERAP2- region has been reported to be associated with AS (ERAP1 rs30187 and rs27434) and Ps (ERAP1 rs27524). In our study, ERAP1 and ERAP2 gene expression were significantly associated with risk SNPs rs30187 ($p < 0.0001$), rs27524 ($p < 0.0001$), rs27434 ($p = 0.0010$) and rs2248374 ($p < 0, 0001$). Because of incomplete linkage disequilibrium (LD) between ERAP1 SNPs (rs27524, rs27434) and ERAP2 SNP (rs2248374), the thymus samples distribute differently for the genotypes for the various SNPs. The ERAP1 SNP, rs30187, provided the most significant expression differences, and has been reported to show the strongest association with AS.

The results shed light on the biological effects of ERAP1/ERAP2 and their role in development of autoimmunity by revealing that risk SNPs were associated with regulation of gene expression. The study showed that AID risk polymorphisms appear to influence the expression levels of certain immune genes in the thymus.

Forkortelser

AIS	Autoimmune sykdommer
AS	Ankyloserende Spondylitt
B2M	Beta-2 Microglobulin
CD	Crohns sykdom (Crohn's Disease)
cDNA	complementary DNA
CeD	Cøliaki (Celiac Disease)
eQTL	expression Quantitative Trait Locus
ERAP1	Endoplasmatisk Retikulum Aminopeptidase 1
ERAP2	Endoplasmatisk Retikulum Aminopeptidase 2
FDR	False Discovery Rate
FRET	Fluorescence Resonance Energy Transfer
GSR	Genotype Suksessrate
GWAS	Genome-wide association studies
HGP	Human Genome Project
HLA	Human Leukocytt Antigen
IBD	Inflammatorisk tarmsykdom (Inflammatory Bowel Disease)
ICMB	Institute for Clinical Molecular Biology
LD	Linkage Disequilibrium
LINE	Long Interspersed Nuclear Elements
MALT	Mucosa Associated Lymphoid Tissue
MHC	Major Histocompatibility Complex
MS	Multippel Sklerose
NMC	The Norwegian Microarray Consortium
PCR	Polymerase Chain Reaction
Ps	Psoriasis
QPCR	Quantitative Polymerase Chain Reaction
RA	Reumatoid Artritt
REK	Regional etisk forskningskomite
RNA	Ribonukleinsyre (Ribonucleic Acid)
rsSNP	reference SNP
SINE	Short Interspersed Nuclear Elements
SLE	Systemisk Lupus Erythematosus

	Enkelt basepar polymorfisme (Single Nucleotide Polymorphism)
SNP	
STR	Short Tandem Repeats/Simple Tandem Repeats
T1D	Type 1 diabetes
UC	Ulcerøs kolitt
VNTR	Variable Number of Tandem Repeats
WGA	Whole Genome Amplified

INNHold

FORORD	II
SAMMENDRAG	III
ABSTRACT	IV
FORKORTELSER	V
INNHold	VII
FIGURER	X
TABELLER	XI
1 INTRODUKSJON	1
1.1 Immunologi	2
1.1.1 Det medfødte, uspesifikke immunforsvaret.....	3
1.1.2 Det ervervete, spesifikke immunforsvaret	4
1.2 De lymfoide organer	6
1.2.1 Beinmargen	6
1.2.2 Tymus	7
1.2.3 De sekundære lymfoide organene	9
1.3 Årsak til autoimmune sykdommer	9
1.3.1 Hovedvevsforlikelighetsantigen	10
1.3.2 Autoimmune sykdommer	11
1.4 Genetisk disposisjon	15
1.4.1 Genetisk variasjon	16
1.4.2 Identifikasjon av sykdomsgener ved komplekse sykdommer	18
2 FORMÅL MED OPPGAVEN	20
3 MATERIAL OG METODER	21
3.1 Seleksjon av SNPer	21
3.2 Tymus vev	23
3.2.1 RNA isolasjon og cDNA syntese	24
3.2.2 DNA isolasjon.....	24

3.3	Metoder	26
3.3.1	Mikromatriser	26
3.3.1.1	Illumina human WG-6v3	26
3.3.1.2	ImmunoChip	27
3.3.2	TaqMan genespresjon	27
3.3.2.1	TaqMan genespresjons assay.....	30
3.3.3	TaqMan Genotyping	31
3.4	Bioinformatisk og statistiske verktøy.....	33
3.4.1	SNP exp v 1.1	33
3.4.2	mRNA by SNP Browser v 1.0.1	33
3.4.3	GraphPad Prism 4	34
3.4.4	DataAssist	35
3.4.5	Ingenuity Pathway Analysis (IPA)	35
3.4.6	HaploView v 4.1	36
3.4.7	Unphased v 2.404	36
4	RESULTATER	37
4.1	Korrelasjon mellom genespresjon og autoimmune risikopolymorfismer	37
4.1.1	eQTL analyser av alle AIS SNPer mot nærliggende transkripter	37
4.1.2	13 eQTL genområder	38
4.2	Kvalitetssikring av genotype data	40
4.2.1	Genotypingssuksessrate og Hardy Weinberg Equilibrium (HWE)	40
4.2.2	Retyping av rs30187 og genotyping av rs2248374	41
4.3	Genespresjonsforskjeller for de signifikante eQTL i ERAP1/ERAP2-regionen	42
4.4	ERAP1 og ERAP2 genespresjon.....	48
4.4.1	TaqMan genespresjonsassay vs Genotype	49
4.4.2	ERAP2 regionen	53
4.4.3	SNPer i ERAP1/ERAP2-regionen	53
4.5	LD mellom SNPene i ERAP1/ERAP2 regionen	55
4.6	Korrelasjon mellom ERAP1/ERAP2 genotyper og HLA klasse I genotyper	57
4.7	Ingenuity Pathway Analysis	57
4.8	Genespresjonen i ERAP1/ERAP2-regionen i forhold til genuttrykket i tymus	60
4.9	Blod korrelasjon - mRNA by SNP browser	62
5	DISKUSJON.....	63
5.1	eQTL assosiasjoner med AIS	63
5.2	Vurdering av signifikansnivå	64
5.2.1	Genotype suksessraten	65

5.3	Replikasjon og kvalitetssikring av eQTL funn i ERAP1 og ERAP2 regionen	65
5.3.1	Microarray genekspressjon vs TaqMan genekspressjon	66
5.4	eQTL og regulerende sekvenser	67
5.4.1	ERAP2 Spleisevarianter	69
5.5	ERAP1/ERAP2 og AIS.....	71
5.5.1	ERAP1 og ERAP2	71
5.6	Ingenuity Pathway Analyse.....	73
5.7	Feilkilder.....	74
6	KONKLUSJON	76
7	REFERANSER	77
8	VEDLEGG	82
8.1	Vedlegg A	82
8.2	Vedlegg B.....	91
8.3	Vedlegg C.....	94
8.4	Vedlegg D	96
8.5	Vedlegg E.....	97

FIGURER

FIGUR 1: SELEKSJONSPROESSEN I TYMUS(11).	5
FIGUR 2: TYMUS(12).	7
FIGUR 3: HLA REGIONEN PÅ KROMOSOM 6(18).....	11
FIGUR 4: GENOMPHI™ AMPLIFIKASJONS KIT (GE HEALTHCARE, LITTLE CHALFONT, UK)(63).	25
FIGUR 5: TAQMAN GENEKSPRESJON(69).	28
FIGUR 6: ILLUSTRASJON AV QPCR ANALYSE PÅ ABI7900HT (ERAP1 TAQMAN ASSAY, HS00372304).	29
FIGUR 7: TAQMAN SNP GENOTYPING(69).	32
FIGUR 8: ILLUSTRERER NEGATIV LOGARITME (-LOG10) AV P-VERDIEN FOR ALLE AID RISK GENOTYPER MOT EKSPRESJONSNIVÅER (INNENFOR +/- 1MB), HVER SPOTT REPRESENTERER P-VERDI FOR ET SNP GENOTYPE-EKSPRESJONSPROBE PAR.....	38
FIGUR 9: PLOTT OVER AIS SNPENE RS30187, RS27524 OG RS27434 OG DERES KORRELASJON MED GENEKSJON AV ERAP1 OG ERAP2. Y-AKSEN VISER – LOG10 (P-VERDI), X-AKSE VISER SNP-POSISJON PÅ KROMOSOMET. PILENE PÅ HVER GRAF INDIKERER HVOR DEN TESTEDE SNPEN LIGGER.....	39
FIGUR 10: IMMUNOCHIP GENOTYPE VS MICROARRAY GENEKSPRESJON FOR RS30187	43
FIGUR 11: RETYPING AV RS30187. TAQMAN GENOTYPE VS MICROARRAY GENEKSPRESJON.	44
FIGUR 12: IMMUNOCHIP GENOTYPE VS MICROARRAY GENEKSPRESJON FOR RS27524	45
FIGUR 13: IMMUNOCHIP GENOTYPE VS MICROARRAY GENEKSPRESJON FOR RS27434	46
FIGUR 14: TAQMAN GENOTYPE VS MICROARRAY GENEKSPRESJON FOR RS2248374.	47
FIGUR 15: ERAP1-ASSAY (HS00372304) VS ERAP1 OG ERAP2 SNPS	50
FIGUR 16: ERAP2-ASSAY (HS01073631) VS ERAP1 OG ERAP2 SNPS	51
FIGUR 17: ERAP1-ASSAY (HS00429970) VS ERAP1 OG ERAP2 SNPS	52
FIGUR 18: UTTRYKK AV ERAP2-PROBEN ILMN_1743145.....	53
FIGUR 19: OVERSIKT OVER LD MELLOM SNPENE I ERAP1/ERAP2-REGIONEN (R SQUARE). A: TYMUS, B: HAPMAP.	56
FIGUR 20: LD MELLOM ALLE SNPER GENOTYPET I HAPMAP PRØVER I HELE REGIONEN.	56
FIGUR 21: OVERSIKT OVER GENUUTTRYKKET I TYMUS ANGITT SOM PROBEINTENSITET FRA HØYEST TIL LAVEST FOR HVER PROBE. Y-AKSE: EKSPRESJONUTTRYKK, X-AKSE: ANTALL PROBER.	60
FIGUR 22: GJENNOMSNITTLIG GENEKSPRESJONUTTRYKK MÅLT VED PROBEINTENSITET I ERAP1/ERAP2- REGIONEN.	61
FIGUR 23: ILLUSTRASJON AV ERAP2 SNPEN RS2248374 ENDRER SPLEISEMØNSTER OG GIR ET PREMATURT STOP KODON(82).	70
FIGUR 24: ERAP1 STRUKTUR OG DOMENE(89).....	72
FIGUR 25: GRAF OVER RISIKO SNPENE RS13003464, RS27434 OG RS30187.....	91
FIGUR 26: GRAF OVER RISIKO SNPENE RS4788084, RS2281808 OG RS2292239.....	92
FIGUR 27: GRAF OVER RISIKO SNPENE RS27524, RS7804356, RS1008953 OG RS1297265.....	93

TABELLER

TABELL 1: OVERSIKT OVER DE UTVALGTE AUTOIMMUNE SYKDOMMENE (AIS).	13
TABELL 2: OVERSIKT OVER SNP SELEKSJON STUDIER FOR AIS.	22
TABELL 3: OVERSIKT OVER TAQMAN GENEKSPRESJONS ASSAY.	30
TABELL 4: B2M TAQMAN ENDOGENE KONTROLL.....	30
TABELL 5: OVERSIKT OVER TAQMAN SNP GENOTYPINGS ASSAY.	32
TABELL 6: GENOTYPINGSKVALITETSSIKRING AV DE10 SIGNIFIKANTE SNPENE ($P < 5 \times 10^{-5}$), GSR OG HWE (HWE=1,0 BETYR FULL LIKEVEKT).	41
TABELL 7: GENOTYPINGSRESULTATER FOR RS30187.....	41
TABELL 8: TAQMAN TYPING AV RS2248374.	42
TABELL 9: OPPSUMMERING AV QPCR RESULTATENE.	48
TABELL 10: OVERSIKT OVER ERAP1 OG ERAP2 POLYMORFISMER OG PLASSERING.	54
TABELL 11: LD MELLOM ERAP1 OG ERAP2 SNPS	55
TABELL 12: LD MELLOM SNP GENOTYPE OG HLA KLASSE I.....	57
TABELL 13: SIGNALVEI ANALYSEN FOR GENENE FRA DE 100 MEST SIGNIFIKANTE AIS SNP – GENEKSJONS-PROBE PARENE ($P < 1 \times 10^{-3}$).	58
TABELL 14: GENEKSJONSUTTRYKK I TYMUS FOR ASSOSIERTE PROBER I ERAP1/ERAP2-REGIONEN.	61
TABELL 15: GENUUTTRYKK I BLOD VS GENUUTTRYKK I TYMUS.....	62
TABELL 16: OVERSIKT OVER AIS 10 EQLNENE ER ASSOSIERT MED NI UTVALGTE AIS.....	64
TABELL 17: OVERSIKT OVER 10 EQTL LOKALISERING.	68
TABELL 18: OVERSIKT OVER HVILKE TRANSKRIPTER TAQMAN - OG MICROARRAY EKSPRESJONSMETODENE FANGER OPP.....	69
TABELL 19: ERAP2-PROBEN FANGER OPP BEGGE SPLEISEVARIANT TRANSKRIPTENE PÅ BÅDE TAQMAN OG MICROARRAY GENEKSJONEN.	71
TABELL 20: AUTOIMMUNE RISIKO SNPS (N=256).	82

1 Introduksjon

Autoimmune sykdommer (AIS) rammer omtrent 5 % av befolkningen, sykdommene er et resultat av en virkning mellom flere risikofaktorer som arv og miljø. AIS er en gruppe sykdommer hvor immunsystemet feilaktig går til angrep på eget vev. Når eget vev blir angrepet vil immunsystemet starte å produsere antistoffer mot bestemte deler av kroppen(1). Årsaken til AIS en kombinasjon av arvelige og miljømessige faktorer. Miljømessige faktorer som diett, toksiner og/eller infeksjoner med patogeniske organismer, har vært forstått som mulige årsaker til initiering av AIS hos genetisk disponerte individer. Det antas at genetisk disponerte individer har genomer med en rekke av følsomme alleler som influerer immunsystemet på en måte som predisponerer for AIS, når de utsettes for de utløsende miljømessige faktorene(2). AIS følger ikke en typisk Mendelian nedarvingsmønster, da det er flere gener og miljø faktorer som er involvert i utviklingen.

Inntil nylig har identifikasjonen av genetiske varianter, som har ført til komplekse sykdommer, som AIS, gått tregt. Med helgenom assosiasjons studier (Genome-wide association studies, GWAS) har hundretusener av enkel nukleotid polymorfismer (SNPs, single nucleotide polymorphisms) blitt undersøkt hos flere tusen personer, og det vil gi større muligheter til å undersøke den genetiske oppbyggingen av komplekse sykdommer. I løpet av de siste årene har disse studiene identifisert flere hundre genetiske varianter assosiert med AIS, og har gitt en innsikt i kompleksiteten i den genetiske arkitekturen(3).

Det er vist at AIS sykdommer ofte blir utløst av virusinfeksjoner. Likevel er det vanskelig å forklare hvordan en slik infeksjon kan utløse en organspesifikk autoimmunitet. Et annet epidemiologisk funn er at hos genetisk disponerte individer, kan infeksjoner i barndommen disponere dem til MS eller T1D, flere år etter klinisk debut(4).

Genetiske faktorer spiller viktig rolle i utviklingen av AIS, den arvelige komponenten av disse sykdommene viser at konkordansen mellom monozygote tvillinger er høyere enn hos dizygote tvillinger. Tvillingmetoden blir brukt for å vurdere i hvor stor grad en egenskap skyldes arv eller miljø. Tvillingforskning ble først benyttet av Francis Galton (1876). Dersom det finnes monozygote tvillinger som er diskordante, må en slik egenskap være påvirket av miljøfaktorer. En slik effekt kan studeres ved og kun utsette den ene tvillingen for denne påvirkningen, og bruke den andre som kontroll (co-twin-metoden). Det kan også sees på fra en annen vinkel. Dersom en egenskap (f.eks. sykdom) oftere finnes konkordant hos monozygote enn dizygote tvillinger, må en slik egenskap være påvirket av arv faktor(5, 6).

På grunn av høy arvelighet bland AIS, så har det ført til mye forskning for å identifisere underliggende gener. Det er derimot en stor sannsynlighet for at det er forandringer i forskjellige alleler, som sammen bidrar til utviklingen av et sykdoms allel(5).

Jakten på genetiske risiko faktorer var frem til 2009 begrenset til koblingsanalyser og assosiasjonsstudier ved å bruke enkelte kandidat gener. De fleste assosiasjonsstudier var kun basert på gener involvert i T-celle signaleringen i den adaptive immunresponsen. Det har blitt gjort flere assosiasjoner mellom HLA systemet og AIS. Den første assosiasjonen ble oppdaget i 1970-tallet, lenge før molekylære teknikker ble oppfunnet. I 2009 var det kun to gener som var assosiert med flere AIS, CTLA4 (cytotoxic T-lymphocyte-associated 4) og PTPN22 (protein tyrosine phosphatase, non-receptor type 22 (lymfoid))(5).

Bottini et al. var de første til å publisere en assosiasjon mellom PTPN22 620W allelet og type 1 diabetes. Videre kom det en rapport fra Begovich et al., som viste til en lignende assosiasjon med reumatoid artritt (RA), i både familiær og sporadisk sykdoms populasjoner. Assosiasjon med PTPN22 620W allelet med RA er også vist i den norske befolkningen(7).

PTPN22 620W allelet ser ut til å være fraværende i asiatiske og afrikanske populasjoner. Begovich et al. viser at dette allelet ikke var detektert hos 100 Han Kineser og 21 Afrikanere. Derfor har det ikke vært vellykket å replikere en assosiasjon med AIS i disse populasjonene. Den høyst sannsynlige assosiasjonen hos kaukasiere og mangelen på PTPN22 620W polymorfismen i andre populasjoner, vil naturlig øke mulighetene for at flere risikoalleler finnes i populasjonene, og disse kan også være til stede hos kaukasiere(8).

1.1 Immunologi

Immunologi er læren om immunforsvaret. En av de viktigste oppgavene til kroppens immunforsvar er evnen til å skille mellom kroppens egne strukturer og fremmede strukturer. Fremmede strukturer omfatter også kroppens egne celler, hvis cellene endrer seg(9). Immunresponsen består av et effektivt samarbeid mellom alle cellene og flere ulike løselige forbindelser. Samarbeidet baseres på direkte kontakt mellom cellene, cellene kan skille ut løselige forbindelser som diffunderer ut i nærmiljøet slik at nabocellenes aktivitet påvirkes, eller bidra til å gjøre mikroorganismer mer synlige slik at ulike spesialiserte celler kan lettere gjøre jobben sin (10).

Immunforsvaret har viktige funksjoner. Det beskytter kroppen mot sykdomsfremkallende mikroorganismer, og angriper kroppsfremmede molekyler. Immunforsvaret fjerner skadede og døde celler og vev, og angriper og fjerner kreftceller. Disse funksjonene utøves av hvite blodceller og flere

ekstracellulære proteiner, som er fordelt i to grupper. Det medfødte, uspesifikke immunforsvaret, og det ervervete, spesifikke immunforsvaret(9).

1.1.1 **Det medfødte, uspesifikke immunforsvaret**

Det uspesifikke immunforsvaret skal forhindre at smittestoffer kommer inn i kroppen, hvis ikke det lykkes bidrar det til at smittestoffer ikke sprer seg i kroppen. Det uspesifikke immunforsvaret består av to deler, det ytre forsvaret og det indre forsvaret.

Det ytre, uspesifikke immunforsvaret består av huden og av celler som kler slimhinner, forsvaret dekker kroppens overflate. Det ytterste laget på huden, epidermis, består av flerlaget plateepitel. Slimhinnene kler de hulrommene i kroppen som har direkte åpning til utsiden, det vil si fordøyelsessystemet, respirasjonsveiene, urinveiene og reproduksjonskanalen, Slimhinnene er dekket av slim som fanger opp mikroorganismer, i tillegg skyller den vedvarende væskestrømmer og fjerner sykdomsfremkallende mikroorganismer i fordøyelsessystemet og urinveiene før de får fotfeste. På slimhinnene er det flere ufarlige bakterier som hindrer vekst av sykdomsfremkallende bakterier. I slimhinnene er det oftest enlaget plateepitel, og her kan derfor mikroorganismer trenge lettere gjennom enn huden(9).

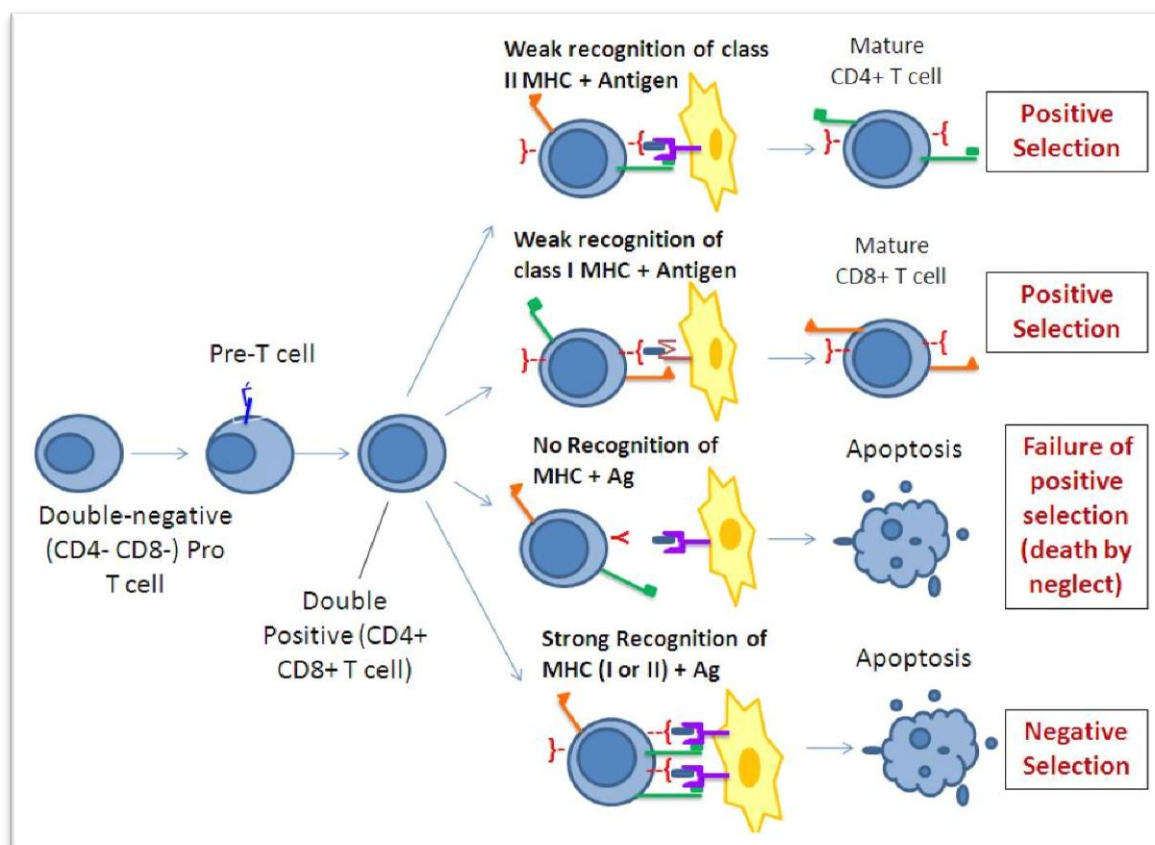
Det indre, uspesifikke immunforsvaret utøves av leukocytter, de hvite blodcellene. Leukocytene kjenner igjen, og skader fremmede partikler sammen med fagocytene. Fagocytene dannes i benmargen, og differensieres til monocytter og nøytrofile granulocytter. Fagocytene angriper fremmede partikler ved blant annet fagocytose, de svelger fremmede partikler og bryter de ned ved å bruke degraderende enzymer i lysosomer som befinner seg i cytoplasma. Noen av disse cellene tar vare på molekyler fra overflaten av patogenet, proteinet og karbohydratstrukturer, og frakter molekylene til overflaten slik at de presenteres for det ervervete immunforsvaret(9, 10).

1.1.2 Det ervervete, spesifikke immunforsvaret

Det spesifikke immunforsvaret utvikles det første leveåret. Denne delen av immunforsvaret utøves av to typer lymfocytter, T- og B-lymfocytter som har ulike oppgaver.

B-lymfocytterne dannes og modnes i beinmargen, deretter går de over i lymfatisk vev hvor de fortsetter å dele seg så lenge de lever (15-20 dager). B-cellenes antigenbindende reseptorer er antistoffmolekyler, det er B-cellene som produserer antistoff. Cellene stimuleres til differensiering ved kontakt med antigen. De differensieres til antistoffproduserende celler, plasma celler og hukommelsesceller. Når det dannes en binding mellom antigen og B-cellenes antistoff, vil det føre til differensiering til mange plasmaceller som produserer en stor mengde av det spesifikke antistoffet. Det vil føre til en binding av alle inntrengende antigener av denne sorten. I løpet av livet vil vi møte på mange forskjellige antigener, og dermed utvikle mange ulike kloner av B-celler og hukommelsesceller. På den måten erverver vi immunitet(9).

T-lymfocytterne forlater beinmargen umodne, og modnes i tymus (Figur 1). De differensierer til T-hjelpeceller og T-angrepsceller (cytotoksiske T-celler). T-cellene har antigene reseptorer på overflaten, de utgjør kloner som stimuleres til deling ved antigenbinding. Reseptorene på overflaten gjenkjenner bare antigenepitoper som er bundet til et vevstypemolekyl (Major Histocompatibility Complex, MHC) på overflaten av en annen celle. Vevstypemolekylene er proteiner som er delt inn i to hovedgrupper, MHC klasse I og MHC klasse II. MHC klasse I, CD8+ T-celler, fins på overflaten til alle kjerneholdige celler, alle celler utenom erytrocytter (røde blodceller) og trombocytter (plater). MHC klasse II, CD4+ T-celler, fins kun på overflaten til antigenpresenterende celler, makrofager, B-lymfocytter og dendrittiske celler(9).



Figur 1: Seleksjonsprosessen i tymus(11).

T-angrepsceller gjenkjenner antigenepitopen når den presenteres på overflaten til MHC klasse I. Antigenepitopen som danner kompleks med MHC I, oppstår inne i cellen. T-celle forsvaret er viktig for å bekjempe kreftceller og virus. T-angrepsceller identifiserer den virusinfiserte cellen ved å binde seg til MHC I-molekyler, som er koblet til en antigenepitop fra viruset. På denne måten vil T-angrepscellen aktiveres. T-cellen vil frigjøre stoffer som danner porer i cellemembranen til den virusinfiserte cellen, slik at cellen sprekker og dør. Etter at vertscellen er død, ødelegges viruset både ved at antistoff bindes til virusantigener, og ved at viruset tas opp og ødelegges av fagocytterende celler(9).

T-hjelpeceller gjenkjenner antigenepitoper kun når de presenteres på MHC klasse II. De aktiverte T-hjelpecellene produserer cytokiner, det fører til stimulering av både T-hjelpecellen og den aktiverte klonen av T-angrepscelle til økt celledeling og celledifferensiering. T-hjelpecellenes viktigste funksjoner er økt fagocytose i makrofager, opp forming av spesifikke T-angrepsceller eller plasma celler og opp forming av hukommelsesceller(9).

1.2 De lymfoide organer

Mesteparten av kroppens immune celler finnes i spesielle vevsstrukturer, de lymfoide organene. Organene er plassert i forhold til fordøyelseskanalen og luftveiene på en siden, og transportsystemene for blod og lymfe på den andre siden. Med en slik plassering vil forsvarscellene kunne komme i kontakt med og identifisere fremmede eller infeksjøs materialer over hele kroppen. Deretter kan de forflytte seg videre til et lymfoid organ, hvor andre viktige celler ligger(10).

Lymfoide vev finnes i form av beinmarg, tymus, lymfeknuter, halsmandler, milten og lymfoid vev knyttet til tarmen (Peyerske flekker). De lymfoide organene deles inn i primære og sekundære. De primære lymfoide organene er ansvarlig for at immunapparatets celler skal få den spesialopplæringen som trengs, for og senere kunne ivareta sin immunologiske funksjon. De sekundære lymfoide vevene har sin funksjon i å skape kontakt mellom immunsystemets celler og fremmede antigener slik at immunrespons initieres(10).

1.2.1 Beinmargen

Immunforsvarets celler finnes i forskjellige varianter med hensyn til modningsnivå. De eksisterer både som umodne, og som svært spesialiserte celletyper. Modning og spesialisering er ofte to prosesser som går parallelt i utviklingen av infeksjonsforsvarets celler.

Den røde beinmargen finner vi i de store rørknoklene våre. Den inneholder et stort antall celler på forskjellige modningstrinn, cellene vokser hele tiden på grunn av celledelingen. Utgangspunktet for celledelingen er en bestemt type umodne morceller, stamceller (pluripotente). Alle cellene i blodet dannes fra de samme stamcellene i beinmargen, og de utgjør langt mindre enn en promille av alle cellene som er til stede her(10).

Stamcellene er umodne celler som ser ut til å ha en nærmest ubegrenset evne til å dele seg. De er selv ikke særlig spesialiserte, men de samme cellene gir derimot opphav til de ulike celletypene. Dattercellene som dannes fra stamcellene, fortsetter å dele seg. Samtidig utvikler de andre, og etter hvert også ulike egenskaper. Det medfører at utviklingen av cellene tar hver sin retning, det er samtidig forklaringen på at alle ulike celletyper stammer fra de samme stamcellene(10).

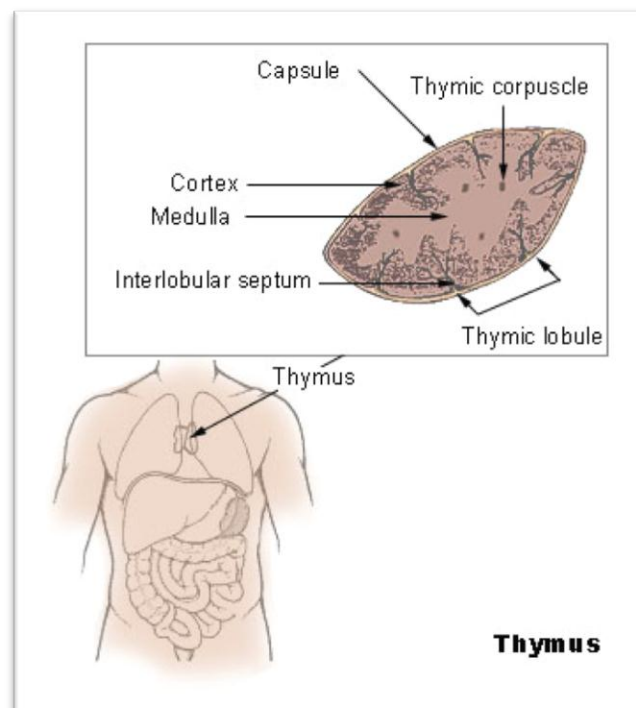
Forsvarsceller blir produsert i et stort tempo i beinmargen, helt fra tidlig i foster tilværelsen. Dersom kroppen skulle ha behov for en spesiell celletype, kan beinmargen svare på kroppens signaler og selektivt øke produksjonen av den celletypen. Det er en påvirkning fra løselige, hormonlignende forbindelser som fører til dannelsen av blodceller ved behov. Erythropoietin (EPO) er et slikt stoff som har en markert effekt på produksjonen av erytrocytter(10).

Oppholdet i beinmargen er begrenset til cellene når et modningsnivå, deretter slippes lymfocytene ut i blod -og lymfesystemet(10).

1.2.2 Tymus

Tymus, brisselen på norsk, er kroppens andre primære lymfoide organ (Figur 2). Det består av epitelialt vev og er et tolappet organ som befinner seg i halsregionen, like bak brystbeinet. Tymus består av to hoveddeler, en bark (cortex) og en sammenhengende marg (medulla). På grensen mellom cortex og medulla følger blodkarene, det gir et nett av kapillærer til cortex-området.

Tymus er størst ved fødselen, da er den ca. 5 cm lang og veier 10-15 gram. Den når sin maksimale størrelse på 30-40 gram rundt puberteten, men vokser langsomt frem til da. Etter puberteten skrumper organet inn, og erstattes av bindevev og fettvev, denne prosessen kan fremskyndes ved sykdom, forgiftning eller hormonelle endringer.



Figur 2: Tymus(12).

På samme måte som beinmargen, har tymus sin funksjon knyttet til utviklingen av forsvarsceller, men bare en bestemt variant. Tymus utvikles hos fosteret fra epitelet i tredje gjellelomme. I løpet av

fosterlivet fraktes umodne lymfocytter fra beinmargen til tymus, de vil hurtig formere seg til thymocytter for å videreutvikles til en høyt spesialisert cellype, T-lymfocytter. Tymus består av en ytre bark hvor lymfocytene ligger tettere sammen enn i den sentrale margen, her er det opphoping av epiteliale celler (tymus-legemer, Hassels legemer) med ukjent funksjon(13).

Noe av det viktigste som skjer med thymocytene under oppholdet er at de utstyres med reseptorer for antigen i cellemembranen, hver thymocyt med sin type reseptor. Det blir dannet en thymocyt populasjonen, som ved hjelp av reseptorene blir i stand til å gjenkjenne alle tenkelige strukturer, inkludert kroppens egne. Dette innebærer at man i neste omgang må kvitte seg med de thymocytene som kan tenkes å gjenkjenne og reagere med kroppsegne molekyler. Det vil også foregå en spesiell utvelgelse av de thymocytene som er i stand til å identifisere og reagere mot fremmede strukturer på en bestemt måte. Vi sier at det foregår både positiv og negativ seleksjonsprosesser i tymus. Rundt 95 % av thymocytene som dannes i beinmargen, blir drept i tymus. Utvelgelsen av T-lymfocytter i tymus er av en sentral betydning for immunsystemets funksjon, og for eventuelle feilreaksjoner (immunsvikt, autoimmunitet). Thymocytene utstyres med reseptorer av vilkårlig spesifisitet, deretter elimineres thymocytter som reagerer mot eget vev. På den måten er man garantert at flest mulig spesifisiteter blir igjen(13).

En moden lymfocyt og alle de lymfocytene som gjennom celledeling stammer fra denne lymfocytten, har identiske reseptorer. Lymfocytter med identiske reseptorer utgjør til sammen en klon, som kun kan bekjempe ett fremmedstoff. I kroppen fins det derimot millioner av forskjellige kloner av lymfocytter.

Etter modningsprosessen i de primære lymfoide organene, sitter kroppen igjen med millioner av lymfocytter, der hver lymfocyt kan identifisere og angripe sitt antigen. Etter modning vil lymfocytene vandre fra beinmargen og tymus, til de sekundære lymfoide organene som er lymfeknutene, milten, tonsillene og det lymfoide vevet i slimhinnene i fordøyelseskanalen, urinveiene og luftveiene. Dersom T-lymfocytene klarer å komme seg gjennom tymus uten denne seleksjonen, vil det kunne føre til immunreaksjoner mot kroppens eget vev, det er et fenomen som ses ved autoimmune sykdommer(13).

Tymus er som sagt relativt størst før eller ved puberteten, og minsker deretter i størrelse med alder. Mye tyder på at nedsatt aktivitet i tymus henger sammen med generelt nedsatte immunfunksjoner etter hvert som vi blir eldre. Dette kan forklare både økt forekomst av infeksjoner, utviklingen av autoimmune sykdommer og kreft med økende alder. Dersom tymus skades tidlig i livet, kan viktige funksjoner i immunsystemet bli påvirket. Hos mennesker kan dette medføre nedsatt evne til å bekjempe virus og soppinfeksjoner(13).

1.2.3 De sekundære lymfoide organene

Disse organene har helt andre funksjoner enn beinmarg og tymus. De er viktige, men ikke absolutt livsnødvendige. De sekundære, perifere, lymfoide organene viser mange likheter, med hensyn til oppbygning og funksjon. Her vil forskjellige forsvarsceller i fellesskap representere en funksjonell enhet for bekjempelse av fremmede inntrengere. Normalt virker milten som et effektivt filter for bakterier som kommer med blodstrømmen. Når mikroorganismer blir holdt igjen her, vil de bli møtt av både lymfocytter, makrofager og andre spesialiserte celler.

På samme måte er det med lymfeknutene, det finnes mellom 500 og 1000 lymfeknuter i kroppen vår med størrelse fra 1-2 med mer og opptil 25 med mer. De fungerer både som trafikkknutepunkter og kontrollstasjoner i nettverket av lymfeknuter. Der vil ulike forsvarsceller gå løs på det fremmede. Enkelte av disse vil også finne veien til det opprinnelige infeksjonsstedet og bidra til å uskadeliggjøre mikrobene.

Sjansen for at mikroorganismene skal finne vei inn i kroppen gjennom slimhinnene, er vesentlig økt i forhold til andre steder. Derfor er immunsystemet rundt mage-tarm-kanalen betydelig forsterket, bl.a. i form av større ansamlinger av forsvarsceller i såkalt mucosa-assosiert-lymfoid vev (Mucosa Associated Lymphoid Tissue, MALT).

Det er to til tre ganger så mange forsvarsceller knyttet til fordøyelsessystemet som det befinner seg i resten kroppen til sammen.

Tarmen er kroppens viktigste lymfoide organ. Mellom epitelcellene finnes en viktig, spesialisert celletype som kalles M-celler. De har som oppgave å bidra til kontakt mellom det lymfoide vevet og tarmlumen. Det gjør de blant annet ved å frakte komponenter fra tarminnholdet over epitelcellelaget og inn til det lymfoide vevet, slik at immunsystemet har muligheter for å kontrollere hva som til enhver tid befinner seg i tarmen.

1.3 Årsak til autoimmune sykdommer

AIS er som sagt sykdommer hvor immunsystemet forårsaker en immunreaksjon mot eget vev og celler. Det antas at AIS oppstår ved at kroppen immunsystemet ikke klarer å skille mellom eget vev og fremmed vev, og ofte starter å produsere antistoffer mot bestemte deler av kroppens vev. De viktigste elementene som er med på å regulere angrepet av vev eller organsystemer er kjente genetiske faktorer som HLA (Human Leukocyte Antigen region) -gener, det medfødte – og det ervervede immunsystemets regulering og utløsende miljømessige faktorer.

Ved AIS kan det være flere genetiske faktorer som er involvert i utviklingen av sykdom, og det kan være flere felles risikovarianter mellom de ulike sykdommene. Det er ved tidligere forskning blitt oppdaget at det er mange immunologiske gener som er involvert, og at det er ofte SNPer i immunologisk regulatoriske områder som HLA-regionen (14, 15).

1.3.1 Hovedvevsforlikelighetsantigen

Hovedvevsforlikelighetsantigen(Human leukocytantigen, HLA) komplekset inneholder et stort antall immunologiske gener som disponerer for AIS. Oppbyggingen av HLA-komplekset, med en stor grad av koblingsulikevekt og mange potensielle kandidatgener, gjør det vanskelig å finne spesifikt sykdomsdisponerende gener(16).

Hos mennesker består HLA-komplekset av mer enn 200 gener, som ligger tett sammen på kromosom 6. HLA-komplekset kan deles inn i flere regioner eller kromosomsegmenter. HLA-komplekset er inndelt i klasse I, II og III (Figur 3)(17).

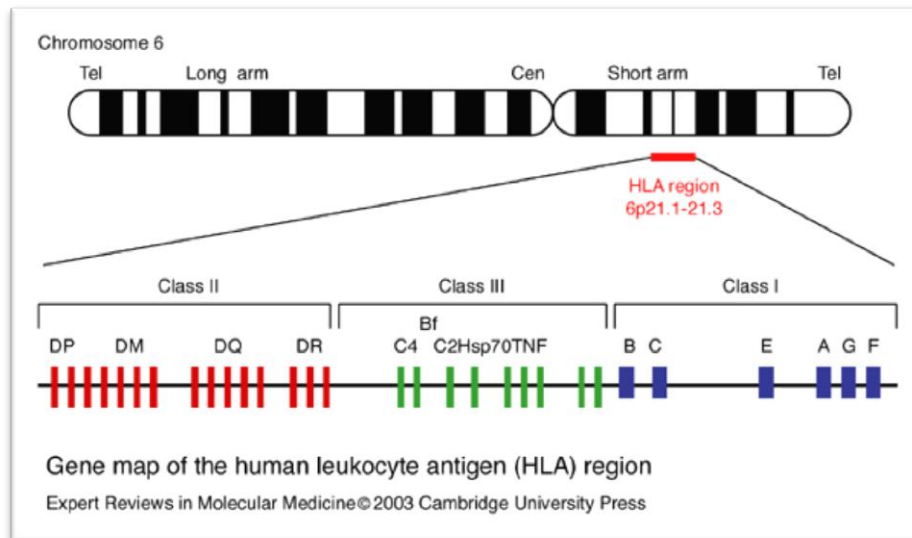
I klasse I regionen finnes det tre loci, A, B og C. Genene i klasse I-regionen koder for glykoproteiner som finnes på overflaten til alle kjerneholdige celler. HLA klasse I-molekylene består av en tung α -kjede og lett β 2-mikroglobulin kjede. Det er kun α -kjedene som kodes for i regionene, den lette kjeden derimot ligger på et annet kromosom(10).

HLA klasse II molekylene består av polypeptidkjedene α og β , som er nesten like tunge. I motsetning til HLA klasse I, befinner begge kjedene seg i HLA klasse II regionen. HLA klasse II-molekylene består av HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA og HLA-DRB1. HLA klasse II-molekylene finnes kun på antigenpresenterende celler (APC), som makrofager, monocytter og B-lymfocytter(10).

Proteinene som blir produsert i HLA klasse III regionen har en litt annen funksjon enn de øvrige regionene. Disse proteinene er involvert i betennelsesprosesser og andre aktiviteter som foregår i immunsystemet. Funksjonen til noen gener i HLA-regionen er ukjent(17)

For hvert locus i HLA klasse I og klasse II finnes det en stor grad av genetiske variasjoner, og HLA genene regnes som noen av de mest polymorfe genene i det humane genom. I noen sykdommer forekommer bestemte HLA-molekyler i mye høyere frekvens hos pasienter enn hos kontroller; e.g. 95 % av AS pasienter har HLA-B27 mot 9% av befolkningen forøvrig. I dag (2012) vet man om cirka 500 ulike kliniske tilstander som er assosiert med gener i HLA-komplekset. Det er ukjent hvorfor kun noen bestemte HLA-molekyler er assosiert med bestemte sykdommer. Det kan skyldes HLA-genene

og proteinene de koder for, eller så kan det kobles med andre gener i HLA-komplekset som er assosiert med immunsystemet(16).



Figur 3: HLA regionen på kromosom 6(18).

1.3.2 Autoimmune sykdommer

Kliniske og epidemiologiske studier viser at mer enn en AIS er observert oftere enn forventet hos nær beslektet familiemedlemmer, og i tillegg rammes en pasient oftere av flere AIS enn prevalensen av sykdommen skulle forutsi. Dette funnet indikerer at det er en tilstedeværelse av felles genetiske faktorer som disponerer for autoimmunitet generelt, og ikke bare for bestemte sykdommer. Spekteret av AIS kan involvere nesten alle kroppens systemer. Det kan strekke seg fra en tilstand som kun rammes et enkelt organ (f.eks. T1D), til en annen som involverer alle systemene i kroppen (f.eks. SLE)(19).

Genetikere ved universitetet Medical Center Groningen (UMCG)(5) ser ut til å ha kommet med resultater som viser til at de avdekkede risikogenene som er involvert i utviklingen av 11 AIS, som T1D, Cøliaki, Crohns og RA stort sett disponerer også for andre AIS. F.eks. er det pr i dag ikke funnet et eneste risikogen for RA som ikke også disponerer for andre AIS. Analysene viser at selv om sykdommene har forskjellig klinisk utgangspunkt, så kan de ha en felles biologisk opprinnelse. Kjennskap til dette fører til en bedre forståelse av hvorfor AIS oftere oppstår i familier, og hvorfor pasientene kan ha flere sykdommer samtidig (5).

Zherakov et al. tok for seg 23 gener, som har vist assosiasjon med to eller flere AIS. Ved å bruke forskjellige verktøy til å analysere signalveiene, kom de frem til et begrenset antall signalveier som var felles for de assosierte genene. De 23 genene var involvert i tre viktige immunologiske signalveier (T-celle differensiering, T-celle signalering og det ervervete, spesifikke immunforsvaret), de 23 genene kunne også inndeles i to funksjonelle grupper; en gruppe bestående av gener som er assosiert med ulcerøs kolitt og Crohns sykdom (IBD), og en annen gruppe bestående av kodete cytokiner og chemokiner. Det var derimot noen sykdomsspesifikke gener som ikke var assosiert med noen av de nevnte signalveiene, disse genene kan være assosiert med andre immunologiske signalveier som fører til AIS(5).

Grovt sett kan AIS deles inn i to hovedtyper, organspesifikke og systemiske. Ved organspesifikke AIS er immunresponsen direkte rettet mot et antigen, som er karakteristisk for et vev eller organ. I dette tilfellet kan det være både cellulære og antistoff medierte mekanismer, og da vil antistoffer eller T-lymfocytter føre til skade på vev eller organer. Antistoffer fører til skade ved at de binder seg til antigen-overflaten slik at komplement kaskaden aktiveres, deretter vil det føre til lysering av cellen og det vil oppstå en betennelsestilstand. Aktiverte T-lymfocytter vil også føre til den samme reaksjonen. Etter hvert vil organstrukturen bli ødelagt og erstattes av bindevev, noe som vil føre til en nedsatt funksjon av organet(10).

T1D er en organspesifikk autoimmun sykdom, sykdommen rammer omtrent 0,4 % av Europas befolkninger og den rammer gjerne visse familier. Ved T1D rammes de insulinproduserende betacellene i bukspyttkjertelens Langerhanske øyer. Aktiverte T-lymfocytter sprer seg i bukspyttkjertelen, og det vil oppstå en betennelsesprosess. Vevsskade oppstår ved at T-hjelperceller frigjør cytokiner som aktiverer makrofager til frigjøring av proteaser. Det har også blitt påvist autoantistoffer mot betaceller, det fører til et ytterligere bidrag til betennelsen. T1D viser sterk HLA assosiasjon med HLA klasse II loci, HLA-DQB1 og HLA-DRB1 på kromosom 6p21. Det er i tillegg andre immunologiske gener involvert, som ikke befinner seg i HLA-regionen(6, 10, 20).

Cøliaki er en T-celle mediert AIS, som rammer tyntarmen. Tarmtottene blir ødelagt som følge av betennelsen som oppstår. Betennelsesreaksjonen i tyntarmen skyldes en unormal reaksjon på gluten, som finnes blant annet i bygg og hvete. Cøliaki rammer omtrent 0,1 % av individer med nord-europeisk opprinnelse. Cøliaki viser assosiasjon med HLA-DQB1(21).

Man finner oftere T1D pasienter med cøliaki enn i befolkningen generelt, avhengig av alderen. Det er for eksempel rapportert at 10 % av barn og 2 % av voksne med T1D er positive, samtidig som det er et økende tilfelle. Det er også blitt antatt at gluten forbruk og inflammasjon, kan være faktorer i utviklingen av T1D(20).

Ved systemiske AIS er målantigenet utbredt, det kan være flere antigener som befinner seg i mange vev og organer. I sykdomsutviklingen kan det være cellemedierte mekanismer, antistoffmedierte cytotoxicitet og det kan være komplement aktiverende immunkomplekser. Eksempler på systemiske AIS er AS, MS, RA og SLE.

Ved multipel sklerose (MS) antas det at kombinasjonen av arveanlegget og virusinfeksjon får immunforsvaret til å angripe myelinet i hjernecellene. Myelinet er kjeden som beskytter nervetrådene.

Nedenfor vises en oversikt over de AIS det har vært fokusert på i denne studien (Tabell 1). Årsaken til de ulike AIS er ennå ikke fullstendig kartlagt, men en del av de immunologiske prosessene har blitt karakterisert. En avdekking av genetiske risikofaktorer og en forståelse av hvordan disse virker biologiske og patologisk vil komplementere vår forståelse av AIS.

Tabell 1: Oversikt over de utvalgte autoimmune sykdommene (AIS).

	Sykdom	Klinisk manifestasjon	Patogenese	Prevalens	Referanse
1	Ankyloserende spondylitt (AS) (OMIM 106300)	Kronisk betennelsesykdom som hovedsakelig rammer de små leddene i ryggraden, det resulterer i stivhet av ryggen og forkalkning av leddene langs virvelsøylen.	Spesifikke auto-antistoffer kan ikke detekteres.	0,025 %	(22)
2	Cøliaki (CeD) (OMIM 212750)	Kronisk betennelsesykdom, atrofi av tarmtøtten og tarmens overflate reduseres.	Tarmtøtten ødelegges av gluten peptider, det fører til en T-cellemediert autoimmun reaksjon mot transglutaminase.	0,5-2 %	(23), (24)

3	Crohns (CD) (OMIM 266600)	Inflammatorisk mage- og tarm sykdom som kan ramme hele fordøyelsessystemet. Angriper alle lag i tarmveggene.	Immunologiske mekanismer som reagerer feil mot strukturer i egen tarm.	0,25-0,65 %	(25), (26)
5	Multipel sklerose (MS) (OMIM 126200)	Rammer sentralnervesystemet. Immunforsvaret angriper myelinet i hjernecellene. Myelinet er den skjeden som beskytter nervetrådene.	T-celle mediert autoimmun reaksjon stimulerer andre immunceller, cytokiner og antigener til å angripe myelinet, det fører til en inflammatorisk prosess.	0,02 %	(27), (28)
6	Psoriasis(Ps) (OMIM 177900)	Betennelses prosess, fører til rask vekst og reproduksjon av hudceller.	Aktiverte T-celler migrerer til dermis og aktiverer utgivelse av cytokiner (TNF α), som videre fører til betennelse.	0,5-1 % (verdensbasis), 2 % (Europa)	(29)
7	Reumatoid artritt (RA) (OMIM 180300)	Kronisk leddsykdom, som angriper særlig perifere ledd (fingre og tær).	Autoimmun reaksjon mot ledd komponenter.	0,5-1 %	(30), (31)
8	Systemisk lupus erythematosus (SLE) (OMIM 152700)	Kronisk betennelse som kan ramme hele kroppen. Angriper hovedsakelig hjerte, ledd, hud, lunger, blodårer, lever, nyrer og nervesystemet	Autoimmun reaksjon som rammer nukleære proteiner, som videre fører til dannelsen av immune	0,04-0,12 %	(32)

			komplekser.		
9	Type 1 diabetes (T1D) (OMIM 222100)	Ødeleggelse av pankreas-celler, fører til utilstrekkelig produksjon av insulin.	Rammer insulinproduserende betacellene i bukspyttkjertelens Langerhanske øyer. Aktiverte T-lymfocytter sprer seg i bukspyttkjertelen, og det oppstår en betennelsesprosess.	0,2-0,3 %	(33)
10	Ulcerøs colitt (UC) (OMIM 191390)	Kronisk betennelse, fører til sår i øverste lag av slimhinnen i tykktarmen.	Immunologiske mekanismer som reagerer feil mot strukturer i tarmen.		(25), (34), (35)

1.4 Genetisk disposisjon

Begrepet gen er ikke helt klart definert, og det har endret innhold i løpet av genetikkens historie. Et gen er en selvstendig enhet av et kromosom, som kreves for produksjon av et funksjonelt produkt. De fleste gener koder for et protein. Det humane genom prosjektet (Human Genome Project, HGP) fullførte i 2003 et referanse-genom, som består av 25000 protein kodende gener, en fjerdedel av det antallet som var anslått tidligere. HGP ble lansert i 1990 med et mål om å oppnå en nøyaktig sekvens av det menneskelige genom(36).

Hvert enkelt proteins funksjon bestemmes av aminosyresammensetningen, og den er bestemt av genene. Hvis det skjer endringer i gensekvensen, så kan det føre til endringer i proteinsekvensen (aminosyresammensetningen). I verste fall kan endringer i aminosyresammensetningen føre til sykdommer, fordi proteinene ikke kan utføre oppgavene sine slik de skal.

Gener er bygget opp forskjellig i bakterier og eukaryote celler. Bakterier inneholder hovedsakelig DNA som koder for proteiner. Gener som koder for proteiner utgjør kun 1-2 % av genomet hos

mennesker. Vi har derimot mange flere proteiner, fordi et gen kan kode for proteindeler som kan settes sammen på ulike måter og kan gi ulike proteiner. En annen egenskap er at proteiner kan modifiseres ved at det settes på andre molekyler, for eksempel suktermolekyler.

1.4.1 Genetisk variasjon

Variasjoner i DNA mellom individer er som regel et resultat av germline mutasjoner, det vil si at det oppstår en mutasjon i kjønncellelinjen og finnes i alle kroppens celler i de enkelte individer. Alle celler i et individ kommer fra den samme befruktete eggcellen, og det er en slik mutasjon som fører til arvelige sykdommer. Det forventes da at DNA er identisk i alle celler. Når det ikke er slik kalles de nye variantene somatiske mutasjoner, forandring som inntreffer i andre celler. En slik mutasjon vil ikke være arvelig, og er begrenset til dattercellene av den muterte cellen. De fleste mutasjoner som forekommer i kreftvev er somatiske mutasjoner(37).

Den enkleste typen variasjon er en gitt posisjon der det forekommer ulike basepar. Dette kalles single nucleotide polymorphisms (SNPs). Kjente SNPs er samlet i en SNP database, og har alle fått et nummer som starter med rs (reference SNP). Den mest kjente SNP database er dbSNP, dette offentlige arkivet inneholder polymorfismer som SNPs og STRs (short tandem repeats), som jeg kommer tilbake til senere. Per dags dato er antallet rsSNPs som er registrert i dbSNP 52 327 221, av disse er det 41 740 143 rsSNPs som er validerte(38).

For å finne regioner i genomet som har polymorfismer som styrer transkripsjonen innenfor populasjoner, brukes genuttrykket som en kvantitativ fenotypisk egenskap. Dette betegnes som uttrykt kvantitativ trait loci (eQTL), og har blitt brukt i GWAS til å kartlegge SNPs til regioner som påvirker genuttrykket(39). I de fleste tilfeller har en SNP to alternative alleler, for eksempel T eller C i en gitt posisjon. For å kunne kalles polymorfisme, må en slik variasjon forekomme med hyppighet på over 1 %.

DNA variasjon kan være ulik DNA sekvens i et område, eller det kan være ulik fragmentlengde. Repetert DNA kan kvalifiseres etter lengden på den repeterte sekvensen, og lokaliseringen på kromosomet. Når den repeterte sekvensen ligger tett etter hverandre i den samme retningen, kalles de for tandem repeterte. Den første typen tandem repetert DNA som ble karakterisert var satellitt DNA som var lokalisert rundt centromeren. Etter hvert oppdaget man også tandem repetert DNA, som ikke var lokalisert rundt centromeren og sekvensene var ikke like lange som satellitt DNA. Tandem repetert DNA ble inndelt i minisatelitt DNA (9-100bp) og mikrosatelitt DNA(1-6bp). Minisatelitter viste en stor lendevariasjon, den totale lengden på et allel varierte mellom 1-20kb, og de var mindre en satellitt DNA. Minisatelitter har høy heterozygositet (mer enn 90 %), dersom man

sammenliknet to alleler i en person ville det vært en stor lengdevariasjon. På grunn av høy heterozygositeten ble det innført et nytt navn, Variable Number of Tandem Repeats (VNTR). Mikrosatelitter er også tandem repetert DNA, men med kortere repetisjoner på 1-6 bp. De ble oppdaget etter minisatelitter, og blir også kalt Short/Simple Tandem Repeats (STR). Lengdevariasjonen mellom alleler i STR skyldes også her et varierende antall repeats, men de er mindre variable enn VNTR og det er færre ulike alleler(37).

Det er også bestemte DNA sekvenser som det fins mange kopier av, men kopiene ligger spredt på ulike kromosomer og kromosomdelere. Den type DNA kalles spredt repetert DNA eller innskutt repetert DNA. Hovedtypene av spredt repetert DNA er: Short Interspersed Nuclear Elements (SINEs) på 260-300bp, Long Interspersed Nuclear Elements (LINEs) på 6-7kb, retrovirus liknende elementer på 1-11kb og DNA transposoner på størrelser mellom 80bp-3kb. De tre første typene er retrotransposoner, de kopierer seg selv og flytter kopien inn et annet sted i genomet. DNA transposoner «klipper» hele sekvensen ut og limer den inn ett annet sted i genomet. Det er ganske få elementer av denne typen i vårt genom(37).

Det er små forskjeller i gensekvensen mellom individer innen en art. Hos mennesker er forskjellen ca. en forskjell per 1000 basepar. Små endringer i baserekkefølgen i arvestoffet over tid, vil føre til genetiske variasjoner fra individ til individ innen en art, og det vil føre til større genetiske variasjoner mellom arter. Oftest har mutasjoner ingen effekt på individet, men det kan slå uheldig ut eller gi individet en fordel. Dersom det skjer en mutasjon i et område som koder for et protein, eller et genregulerende område, så vil man få påvist en fenotype. I de fleste slike tilfeller vil en slik mutasjon føre til at cellene blir drept ved programmert celledød. I noen tilfeller kan det føre til at proteinet får endret egenskaper, og da vil det kunne føre til utvikling av kreft. I andre sjeldne tilfeller vil en mutasjon i et genregulerende område føre til noe positivt, som dannelsen av et nytt allel og det kan gi cellene bedre egenskaper(40).

Genetisk variasjon skyldes ikke kun mutasjoner. Individuell variasjon skyldes også at vi vokser opp i forskjellige miljøer og blir utsatt for ulike påvirkninger. Epigenetikk er læren om hvordan genene forandrer seg ved at blant annet miljøet påvirker aktiviteten og funksjonen til genene, denne genreguleringen er arvelig. Cellebiologen Lipton definerer epigenetikk slik: Epigenetikk er vitenskapen om hvordan signaler fra omgivelsene utvelger, modifierer og regulerer genaktivitet(41).

1.4.2 Identifikasjon av sykdomsgener ved komplekse sykdommer

Identifisering av disponerende loci i AIS har blitt gjort ved hjelp av tre grunnleggende strategier: (i) koblingsanalyser av flere familier med AIS, (ii) assosiasjonsstudier gjort på kandidat gener hvor man bruker pasienter og kontroller, (iii) helgenom assosiasjonsstudier gjort på samlinger av pasienter og kontroller som regel utført av konsortier av forskere. Mange tidligere studier av mottakeligheten av AIS ble utført med koblingsanalyser, det vil utnytte co-segregering av kromosomale regioner med sykdom hos disponerte familier, slik at man kan identifisere sykdomsgener. Slike analyser har identifisert mange loci, men de fleste disponerende lociene ble detektert med lav statistisk signifikans og funnene var ofte dårlig replikert i uavhengige studier(2).

Familiestudier og ulike dyremodeller har indikert at det er flere gener som bidrar til autoimmunitet, og man antar at små og additive effekter fører til en sykdomsfenotype. Forandringer i forskjellige gener kan føre til samme sykdomsfenotype, og det vanskeliggjør deteksjonen av sykdomsgener som predisponerer for komplekse sykdommer. Som tidligere nevnt er det ofte miljøfaktorer involvert. Det er en mye større utfordring å påvise gendefekter som forårsaker komplekse sykdommer som autoimmunitet. Man tar i bruk koblingsanalyser og assosiasjonsanalyser for å bestemme gener som predisponerer for monogene sykdommer og komplekse sykdommer.

Koblingsanalyser utføres vanligvis for å kartlegge monogene sykdommer med klar mendelsk arvegang. Her er det spesielt viktig med penetrans, dersom man får et gen som bidrar til sykdom og fører til sykdomsfenotypen. Koblingsanalyser blir utført i familiemateriale, målet er å finne ut om to loci er koblet og om sykdommen er koblet til denne markøren, og samtidig finner man ut av hvor ofte to loci blir separert av meiotisk rekombinasjon. Dette undersøker man ved å observere hvordan allelene fordeles i barna i forhold til hvordan de var i foreldregenerasjonen. Hvis to loci ligger på hvert sitt kromosom, vil de alltid nedarves uavhengig av hverandre og det vil si at de ikke er koblet. Loci som ligger nær hverandre på samme kromosom, vil sjeldent bli separert av meiotisk rekombinasjon, og de vil som oftest være koblet og nedarves sammen. Ved å undersøke om et bestemt gen nedarves sammen med sykdommen i store sykdomsbelastede familier, kan en se om genet og sykdommen oftere enn forventet nedarves sammen. Dette vil være et tegn på at genet som undersøkes (markøren) og sykdomsgenet sitter ganske nær hverandre på samme kromosom. Slike studier er godt egnet til å finne gener som forårsaker monogene sykdommer(37).

For å kunne kartlegge hvor på kromosomet et gen som disponerer for AIS befinner seg, så brukes det genetiske markører. Det er i de senere år blitt utviklet genetiske markører for hele genomet, for mus, rotte og menneske. Ved å bruke å bruke genkartene har man klart å identifisere kromosomlokaliseringen til flere gener som predisponerer for autoimmunitet både i dyremodeller og

hos mennesker. Mikrosatelittmarkørene er ofte dinukloetidrepetisjoner som er jevnt fordelt over hele genomet. Mikrosatelittene inneholder høye nivåer av variasjon mellom individer, med repetisjoner på 1-6bp. Lengdevariasjonen mellom alleler i mikrosatelitter skyldes et varierende antall repeats, de er derfor multi alleliske polymorfismer. Den store fordelingen med mikrosatelittloci er at totalstørrelsen på et allel oftest er mindre enn 1 kb. De kan dermed analyseres ved hjelp av PCR-teknikk (Polymerase Chain Reaction). Siden mikrosatelittene er svært polymorfe, er det stor sannsynlighet for at parentale kromosomer bærer ulike alleler for en gitt markør, noe som gjør det mulig og oppdage det genetiske bidraget fra foreldre til sykt avkom(37).

For komplekse sykdommer der flere gener gir opphav til sykdommen, har det vist seg å være vanskeligere å påvise gener som er involvert i sykdomsutviklingen med koblingsanalyser. En metode for å identifisere gener ved komplekse sykdommer er assosiasjonsstudier. Denne studien baserer seg på sammenlikning av en gruppe pasienter med en kontrollgruppe fra samme populasjon. Disse må ikke være i slekt med hverandre. I en befolkning vil en kunne se at to genvarianter opptrer sammen oftere enn forventet ut ifra hver av allelenes frekvens i befolkningen. Dette kalles koblingsulikevekt (Linkage Disequilibrium, LD). Dette skyldes også rekombinasjon.

En økt forekomst av et bestemt allel i en sykdomsbefolkning sammenliknet med kontrollgruppen fra samme populasjon, betyr derfor at det enten er det undersøkte allelet som selv bidrar til sykdomsdisposisjon eller det er et allel på en nært koblet locus som medfører sykdomsdisposisjon. Assosiasjonsanalyser baserer seg på om et SNP-allel forekommer oftere hos pasientene enn kontrollene i en gitt populasjon. SNP-markører benyttes fordi den haplotypen som viser assosiasjon (på grunn av LD) vil være relativt liten (1-2 cM). Mikrosatelittene ligger ikke så tett på kromosomet og kan dermed ikke brukes som markører. SNP-markører derimot er svært hyppige, sitter med ca. 300 bp avstand(37).

2 Formål med oppgaven

Mål for oppgaven er å kombinere genetiske og funksjonelle studier for å kartlegge risikogener i tymus, som kan føre til autoimmune sykdommer. Vi skal studere rollen til immunologiske gener i patogenesen og deres primære biologiske faktorer. Vi ønsker å:

- Karakterisere genekspressjonen i tymus.
- Korrelere risikoalleler for AIS og ekspresjonsnivå av nærliggende gener.
- Fordyping i en av regionene som viser AIS SNPer som er assosiert med genekspressionsnivåer.

3 Material og metoder

3.1 Seleksjon av SNPer

Ved oppstart av oppgaven ble det gjort en SNP seleksjon ved å gjøre et litteratursøk i PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>), for å se på de studiene som er gjort på AIS. De autoimmune sykdommene som ble fokusert på i dette prosjektet er type 1 diabetes (T1D), reumatoid artritt (RA), systemisk lupus erythematosus (SLE), multippel sklerose (MS), ankyloserende spondylitt (AS), cøliaki (CeD), inflammatorisk tarmsykdom (IBD), crohns sykdom (CD), ulcerøs kolitt (UC) og psoriasis (Ps) (Tabell 2). I PubMed ble det spesielt sett nærmere på de studiene hvor man undersøkte flere AIS, og man muligens hadde funnet en assosiasjon. I litteratursøket ble det for det meste brukt review artikler. Litteratursøket var viktig for å få en oversikt over alle risiko SNPs, i de ni AIS sykdommene som er valgt ut til studien (Tabell 2). Det ble selektert ut risiko SNPs som hadde en immunologisk rolle, og det innsamlede materialet måtte være av Europeisk herkomst.

I tillegg ble databasen "A catalogue of genome-wide association studies" (<http://genome.gov/GWASudies/>) brukt for å søke opp gener og polymorfismer som har vært rapportert å påvirke risikoen for de aktuelle AIS. De studiene som publiseres i GWAS katalogen omfatter kun de studiene som analyserer minst 100 000 SNPer i den første fasen. Studiene er identifisert gjennom ukentlige PubMed litteratursøk, samlinger med nyheter og media rapporter, og sammenlikninger med eksisterende database med GWAS litteratur.

De SNPene som ble valgt ut til å inkludere i studien er etablerte risiko SNPer i GWAS, og SNP-fenotype assosiasjonene er begrenset til de SNPene ved at vi brukte genome-wide signifikans nivå ($p < 1.0 \times 10^{-5}$).

Disse seleksjonskriteriene resulterte i at det til slutt ble selektert ut 256 SNP (Vedlegg A)

Tabell 2: Oversikt over SNP seleksjon studier for AIS.

	Sykdom	GWAS studier	Region	Initial studie størrelse(caser/kontroll) ¹	Oppfølging studier (caser/kontroll)	Referanse
1	Ankyloserende sponylititt (AS) (OMIM 106300)	2	16	3023/8779 2053/5140	2111/4483 898/1518	(42), (43)
2	Cøliaki(CeD) (OMIM 212750)	3	41	4533/10750 767/1422 778/1422	4918/5684 1643/3406 991/1489	(44), (24), (23)
3	Crohns (CD) (OMIM 266600)	10	70	16179 6333/15056 896/3204 3230/4829 382 ³ 393/399 1748/2938 1748/2938 946/977 547/928	IR ² 15694/14026/414 ³ 1174/357 2325/1809/1339 ³ 750/828/521 ³ 2920/1961/1248 ³ (see Parkes 2007) 1182/2024 353/207/530 ³ 1266/559/428 ³	(45), (46), (47), (26), (48), (33, 49), (50), (51), (52)
4	Multippel sklerose (MS) (OMIM 126200)	4	68	9772/16846 5545/12153 1470/IR ² 1618/3413	4218/7296 IR ² IR ² 2256/2310	(53), (54), (55), (56)
5	Psoriasis(Ps) (OMIM 177900)	2	17	233/519 2178/	IR ² 3174/5464	(29), (57)

¹ Europeiske avstammings caser/ europeiske avstammings kontroller

² IR: Ikke rapportert

³ Europeisk avstamning (Foreldre-avkom trioer)

6	Reumatoid artritt (RA) (OMIM 180300)	2	32	2768/4198 5539/20169	IR ² 6768/8806	(58), (15)
7	Systemisk lupus erythematosus (SLE) (OMIM 152700)	1	24	811+906 ⁴ /4813	IR ²	(59)
8	Type 1 diabetes (T1D) (OMIM 222100)	3	13	7514/9045 9934/16956 16179 ⁵	4840/4152 1120 ³ IR ²	(60), (61),(45)
9	Ulceriøs colitt (UC) (OMIM 191390)	3	52	6687/19718 2693/6791 2361/5417	9628/12917 2009/1580 2321/4818	(62), (63), (64)

3.2 Tymus vev

Det er samlet inn tymus-prøvemateriale fra 42 norske barn (22 jenter og 20 gutter), under 13 år (26 barn under 1. år, 16 barn 1-13 år). Prøvematerialet ble innsamlet under hjerteoperasjon, tymus skulle bli fjernet uansett, og pasientene er helt anonyme for forskningsprosjektet. Foreldrene ga informert samtykke og prosjektet er godkjent av Regional etisk forskningskomite (REK S-04101) og Norsk samfunnsvitenskapelige datatjeneste (NSD, 11135).

⁴ Anti-dsDNA positive + anti-dsDNA negative

⁵ Individuelle europeiske avstammere

3.2.1 RNA isolasjon og cDNA syntese

Det innsamlede tymus vevet ble kuttet i mindre biter rett etter operasjonen (<5 mm tykke), og samlet i rør som inneholdt RNeasy lysis solution (Qiagen, Crawley, UK), for å stabilisere og beskytte RNA fra degradering. Deretter ble det isolert total RNA ved å bruke TRIzol® Reagent (Invitrogen, CA, USA). TRIzol® RNA isolasjons metoden er en guanidine isothiocyant-fenol-kloroform væske-væske ekstraksjon. Metoden separerer RNA fra komplekse biologiske prøver og genererer en aqua-fase bestående rent RNA.

cDNA syntesen ble derimot utført ved først behandling med RNase-free DNase I (New England Biolabs, Ipswich, UK) og hexamerer, deretter med SuperScript™ III Revers Transkriptase (Invitrogen, Cat. No: 18080-051). cDNA syntesen ble utført etter leverandørens anbefaling, ved å bruke 1000ng total RNA som input. Etter isolering av total RNA ble konsentrasjonen og kvaliteten sjekket på BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Bioanalyzer er en microfluidics-basert kapillær elektroforese for dimensjonering, kvantifisering og kvalitetskontroll av blant annet DNA og RNA.

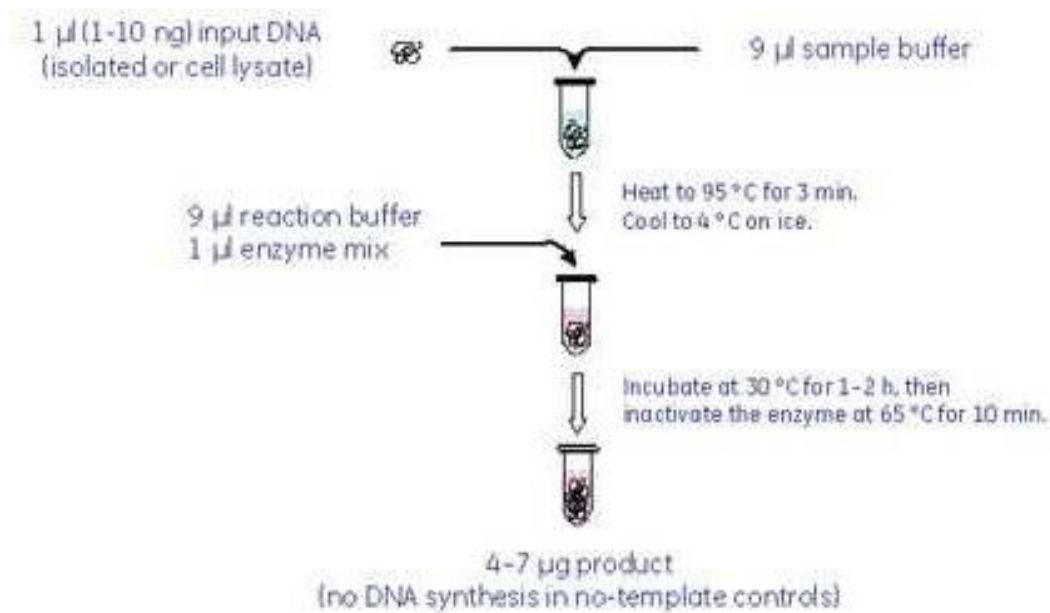
Alle RNA prøvene ble testet på Bioanalyzer 2100 instrument (Agilent technologies, CA, USA) for å sjekke renhet og RNA integritet (RIN; dvs at de ikke er degradert). RIN (RNA integrity number) var over 7,5 for alle prøver unntatt to (T15 og T33). Det var ingen tegn på DNA forurensning i noen av prøvene.

Konsentrasjonen på cDNA fra blodprøvene var 15 ng/μl. cDNA syntesen ble gjort ved å bruke Tempus rør (Applied Biosystems, Foster city, CA, USA) og medfølgende Tempus spin RNA kit. Det ble brukt AbsolutRNA Wash solution (Applied Biosystems) til DNase-behandlingen.

3.2.2 DNA isolasjon

DNA ble også isolert fra alle tymus-prøver, og deretter ble genomisk-DNA amplifisert ved whole – genom amplified (WGA) ved å bruke GenomiPhi™ amplifikasjons kit (GE Healthcare, Little Chalfont, UK). GenomiPhi metoden bruker bakteriofagen Phi 29 DNA polymerase, som eksponentielt amplifiserer enkelt – og dobbeltrådet lineært DNA templat. DNA polymerasen senker reaksjonstiden og reduserer uspesifikke amplifikasjoner(65). GenomiPhi™ amplifikasjons kit (GE Healthcare, Little Chalfont, UK) består av prøve buffer (tilfeldige hexamerer), reaksjons buffer (salt og deoxynukleotider) og enzym mix (Phi 29 DNA polymerase og tilfeldige primere). Reaksjons bufferen er justert til en pH-verdi som er optimal for DNA polymerasen syntesen.

DNA amplifikasjonen starter med å tilsette prøve buffer til templat-DNA, deretter inkuberes røret i 95°C i 3 min for at DNA skal denaturere (Figur 4). Tilfeldige hexamer primere annealer til denaturert templat-DNA. Røret nedkjøles og blir tilsatt reaksjons buffer og enzym mix, blandingen blir inkubert i 30°C over natten (16-18 timer). Phi 29 DNA polymerase starter replikasjon på flere steder på denaturert lineært DNA. Mens polymeriseringen fortsetter vil det skje en forskyvning av nedstrøms replikert DNA, som genererer en ny enkelt-trådet DNA. For hver ny syntetisert enkelt-trådet DNA, vil det bindes flere primere. Dermed vil genomisk-DNA amplifiseres eksponentielt. Etter amplifikasjon vil DNA polymerasen inaktiveres ved 10 minutters inkubering i 65°C. GenomPhi amplifisert DNA gir høy molekylvekt, nøyaktig korrektur lesing og dobbelt-trådet DNA. DNA kvaliteten er viktig i genetiske studier, da dårlig kvalitet kan føre til en dårlig genotype suksessrate (GSR).



Figur 4: GenomPhi™ amplifikasjons kit (GE Healthcare, Little Chalfont, UK)(63).

3.3 Metoder

3.3.1 Mikromatriser

DNA mikromatriseteknologien er en molekylær biologisk metode som brukes for å påvise DNA eller cDNA fragmenter i en prøve. Det kan altså brukes både til SNP genotyping og til ekspresjonsanalyser. Det er en metode som kan analysere flere gener i ett og samme forsøk.

Med dagens teknologi er det mulig å plassere millioner av små distinkte spotter av ulikt materiale innen et område på bare noen få cm². Data blir analysert ved å bruke spesifikke programmer som samler gener med likt ekspresjonsmønster (cluster), her antar man at genene deler felles biologiske funksjoner og reguleringer. Mikromatrisene har høy sensitivitet, og man trenger ikke å ha kjennskap til DNA-sekvens.

En mikromatrise er et ordnet mønster med mikroskopiske spotter (<250 µm) som kan inneholde DNA, protein eller annet materiale, der det er mulig å analysere komplekse biologiske prøver (Definisjon: Shena et al, Science 1995). Det faste underlaget som brukes til å feste spottene kan være vanlig glass slides, silisium brikker eller nylonfilter.

3.3.1.1 Illumina human WG-6v3

Illumina human WG-6v3, som ble brukt i dette prosjektet, består av små metallkuler som sitter i små brønner. BeadChip kittet består av hybridisering buffer, vaske buffer og vaske brett. Hver metallkule binder mange identiske kopier av gen-spesifikke prober, og det er gjennomsnittlig 30 tilfeldig plasserte replikater på hver metallkule(66).

Det kan analyseres seks prøver på hver BeadChip, og det er > 45 000 mål-sekvenser per prøve. Mikromatrisen ble brukt til å utføre en ekspresjonsanalyse på total RNA. Illumina human WG-6v3 består av 48804 prober totalt, av disse er det kun 27455 som er velkarakteriserte transkripter. Noen prober er komplementære med en gensekvens, mens noen gensekvenser er komplementære med flere prober. Det spesifikke antall gener som er dekket er vanskelig å si, men det skal være de fleste velkarakteriserte transkripter(67).

Prøvene ble analysert på Radiumhospitalet, den norske mikromatrise konsortiet (The Norwegian Microarray Consortium, NMC)(www.microarray.no).

Den mest vanlige bruken av mikromatriser er for genekspresjonsanalyser. Mikromatrisene inneholder spotter med DNA som koder for ulike gener (mRNA). Mikromatrisen vil detektere

mengden av DNA transkripsjon i de biologiske prøvene. Spottene på cDNA mikromatriser inneholder genspesifikke cDNA fragmenter, med en lengde på 0,2-2 kb.

3.3.1.2 *ImmunoChip*

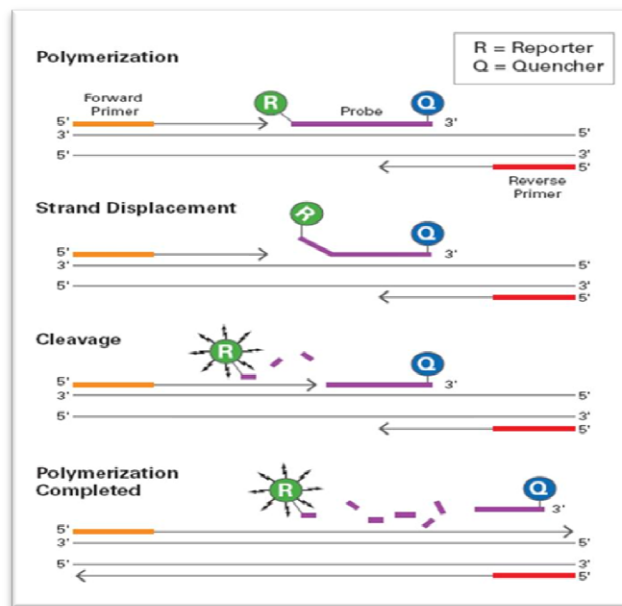
ImmunoChip ble i dette prosjektet brukt til å genotype de selekterte AIS risiko SNPene. ImmunoChip er en custom Illumina Infinium HD DNA genotypingsarray (Illumina, San Diego, CA, USA) som består av mer enn 200 000 forskjellige genetiske prober i regioner som er funnet assosiert med AIS. Spesifikt er chipen spesielt laget for å brukes til genetisk forskning mot autoimmune sykdommer, og særlig for å studere overlapp av risikofaktorer på tvers av ulike AIS. I tillegg genotyper ImmunoChipen alle kjente SNPer i områdene rundt de etablerte risikofaktorene slik at den også kan brukes til fin-mapping. Denne chipen er et resultat av et internasjonalt konsortium kalt immunoChip konsortium. ImmunoChip har blitt brukt i en rekke internasjonale studier for å teste på tvers av de autoimmune sykdommene, og for å se etter felles faktorer. Ved og selv designe en ImmunoChip direkte rettet mot AIS, vil det gi lavere kostnader i forhold til kommersielle plattformer som er rettet mot global screening av genetiske polymorfismer. Med et utvalg av sekvenser av interesse, kan man fokusere på viktige signalveier og molekylære mekanismer i AIS (68). Genotypingen for dette prosjektet ble utført i Kiel, på instituttet for klinisk molekylær biologi (Institute for clinical molecular biology, ICMB).

3.3.2 *TaqMan genekspressjon*

QPCR er en metode som blir brukt til å amplifisere og kvantifisere et mål-DNA molekyl. Den blir brukt til å detektere en spesifikk sekvens i en DNA/RNA-prøve. QPCR har en større sensitivitet enn vanlig PCR, og ved analysering vil man kunne kvantifisere produktet for hver syklus av PCR-reaksjonen. For å kvantifisere produktet blir det brukt TaqMan®prober.

TaqMan genekspressjonsassay består av et primer-par og TaqMan®prober, som er lineære oligonukleotider. Dette er hydrolyserende prober, som er designet for å øke spesifisiteten til qPCR. Proben er bundet med en reporter-fluorofor (grønn, kort bølglengde) på 5' enden, og en Quencher (hemmer-rød, lang bølglengde) i 3' enden. Mekanismen er basert på FRET (Fluorescence Resonance Energy Transfer), som er hemming av ett fargestoff (FAM) forårsaket av en annen (VIC) uten utslipp av ett proton, og kan oppstå over en lang avstand (10 nm). Probene er designet til å ha en høyere T_m enn primerne. TaqMan®proben er komplementær med deler av målsekvensen(templat-DNA) og den vil binde seg etter denaturering mellom bindingssetene til forward primer og revers primer, deretter vil primerne starte hybridisering til cDNA. Taq-polymerasen har en 5' – 3' eksoklease aktivitet, og under forlengelsen av forward primeren vil polymerasen treffe på å kutte proben. Dermed vil ikke fluoroforen lenger være intakt med quencheren, og det vil bli avgitt et fluorescens

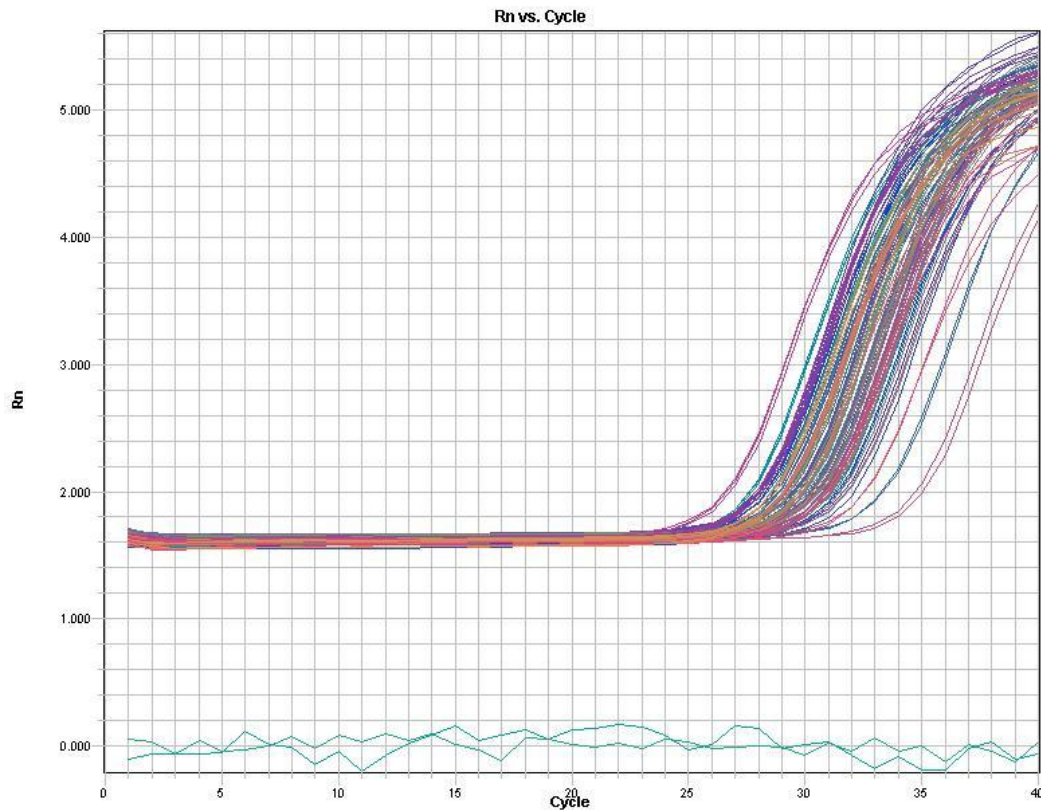
signal (Figur 5). Så lenge fluoroforen er intakt med quencheren, så vil quencheren hemme at det avgis et fluorescens signal. Fluorescensen leses av spektrofluorometrisk når reaksjonen er ferdig, det vil kvantifiseres ved hjelp av en datamaskin, og da vil man se hvor mange sykluser som må før det er påviselig mengde i prøven. Under PCR-syklusen vil flere TaqMan-prober binde seg, og emittert lys vil detekteres. Mengden fluorescerende signal er direkte proporsjonalt med antall PCR-produkt molekyler som er generert i den eksponentielle fasen av reaksjonen (37).



Figur 5: Taqman genekspressjon(69).

Fordelene med qPCR er at det blir mulig å kvantifisere lav konsentrasjon av mRNA, og i tillegg kvantifisere relativ genekspressjon i en prøve.

QPCR ble utført på fem ferdig designet TaqMan genekspressjons assay, som ble utført på ABI7900HT RT-PCR maskin og SDS v2.3 (Applied Biosystems). Amplifikasjons grense er det punktet den eksponentielle fasen (dobling av PCR produkt per cycle) starter for hver prøve, den farge intensiteten der Ct-verdien (Ct = tiden i cykler) avleses på x-aksen (Figur 6). Denne grensen bestemmes automatisk for hver analyse som blir kjørt. Dette blir gjort ved å se hver enkelt kurve og benytte alle til å beregne et middelpunkt det er en eksponentiell økning i produkt (fargeintensitet), som passer best for alle prøver(70).



Figur 6: Illustrasjon av qPCR analyse på ABI7900HT (ERAP1 Taqman assay, Hs00372304).

På triplikatene for de analyserte TaqMan genekspressjonsresultatene, ble det satt en standardavvik (STDEV) cut-off grense = 0,165.

For videre analysering av TaqMan genekspressjonsresultatene ble Ct-verdien brukt. Ct-verdien reflekterer antall sykluser som må til for at en prøve har nådd treshold-grensen, og mengden fluorescerende signal er direkte proporsjonalt med antall PCR-produkt molekyler som er generert i den eksponentielle fasen av reaksjonen. Desto mer templat som er til stede i en prøve, jo færre PCR-sykluser skal det til for å lage påviselige mengder av PCR-produkt. For normalisering av genuttrykket ble Ct-verdier for kandidatgenet (ERAP1/ERAP2) og Ct-verdier for housekeepinggen (B2M) brukt for å finne δ (delta) Ct-verdien.

$$\frac{Ct - verdi (Kandidatgen)}{Ct - verdi (Housekeepinggen)} = \delta Ct$$

Delta Ct-verdien ble brukt til korrelasjon av TaqMan genekspressjonen med genotypen (ImmunoChip og Taqman).

3.3.2.1 TaqMan genekspressjons assay

TaqMan genekspressjonsassay ble valgt på bakgrunn av informasjon fra Applied Biosystems, og for å kunne detektere de fleste kjente transkripter på genene ERAP1 og ERAP2 (Endoplasmatiske Retikulum Amino-peptidase 1 og 2). Det er ulike transkripter som fanges opp av ulike assays. For å unngå DNA kontaminasjon ble det valgt assays med primere som dekket ekson-ekson overgang (Tabell 3).

Tabell 3: Oversikt over TaqMan genekspressjons assay.

	Gen	Assay	Ekson	Antall transkripter	Fragment lengde
1	ERAP1	Hs00429970_m1	3-4	3 Refseq	66
2	ERAP1	Hs00372304_m1	19-20	1 Refseq	60
3	ERPA2	Hs01073631_m1	18-19	2 Refseq	116
4	ERAP2	Hs01082025_m1	1-2	1 Refseq	148
5	ERAP2	Hs01073630_m1	17.18	2 Refseq	118

Beta-2 microglobulin (B2M) som er et housekeeping gen, ble brukt som endongen kontroll. B2M tillater relativ genekspressjons kvantifisering i cDNA prøver, når den er brukt sammen med andre TaqMan genekspressjons assay (Tabell 4). TaqMan®proben er merket med 6FAM™ farge.

Tabell 4: B2M TaqMan Endogene kontroll.

Endogen kontroll	Assay ID	Ekson	Konsentrasjon/ Enhet	Fragmentlengde
Human B2M (FAM / MGB Probe, Non- Primer Limited)	4333766F- 0505006	2-3	20x/200 rxn	75

Konsentrasjonen på cDNA var 20 ng/μl. Fortynning: 5 μl cDNA + 45 μl H₂O.

Det var fem genekspressjonsassay som ble brukt til undersøkelse av ERAP1/ERAP2-regionen. I hver mix ble det tilsatt 1250 μl 2x TaqMan genekspressjon mastermix (Applied Biosystems, Foster City, CA, USA), 125 μl 20x TaqMan assay mix (Applied Biosystems, Foster City, CA, USA) og 375 μl sterilt H₂O.

Det ble brukt 3 μl fortynnet cDNA + 7 μl mix i hver brønn, totalt 10 μl i hver brønn. Det ble brukt en plate med 384 brønner, og pipetter med filter-spiss.

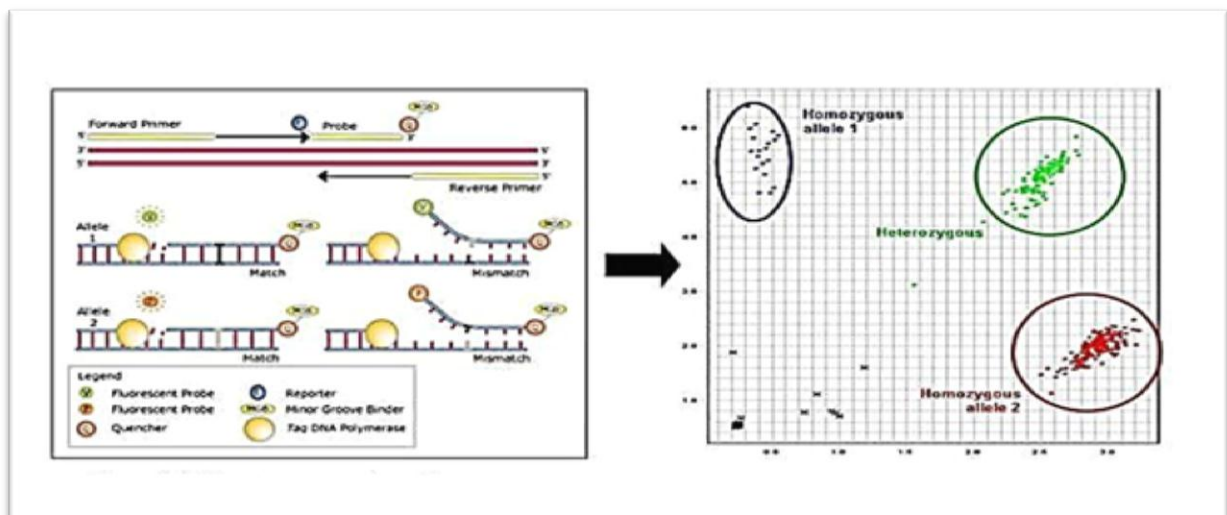
Det ble brukt samme fortynnet cDNA til analysering av B2M, men tillaging av mix var ulikt. I hver mix ble det tilsatt 1100 μl 2x TaqMan genekspressjon mastermix (Applied Biosystems, Foster City, CA, USA), 110 μl 20x TaqMan assay mix (Applied Biosystems, Foster City, CA, USA) og 375 μl sterilt H₂O. Mixen ble tilsatt i brønnene med lik mengde volum, totalt 10 μl i hver brønn.

3.3.3 TaqMan Genotyping

Allel-diskrimineringsanalyser er også kjent som ende-punktsanalyser, fordi data samles inn på slutten av PCR-reaksjonen. Analysen detekterer varianter av enkelt nukleotid polymorfismer (SNPs).

Tilstedeværelsen av ett primer par/prober i hver reaksjon gjør det mulig å kunne detektere mulige varianter av en SNP. Hver prøve i allelisk diskriminerings analyse inneholder et par fluorescerende fargestoff detektorer, dette er prober som binder seg til et område en SNP ligger.

Taqman er en SNP genotypingsmetode som er probe-basert. Assayet består av to primere, en forward og revers primer på hver side av området en SNP ligger, to TaqMan[®] MGB (Minor Groove Binder) prober i PCR-mixen som er farget med en fluoriserende farge (VICog FAM) for hvert av de to allelene. For eksempel for en A/G SNP kan man ha en VIC-probe som binder seg til A og en FAM-probe som binder seg til G. Prøver som er homozygote for A-allelet vil kun bli farget med VIC, tilsvarende vil de med GG kun ha FAM, mens de som er heterozygote (AG) vil ha både VIC og FAM. På instrumentet som leser av dette med en laser vil dermed prøvene grupper seg i tre forskjellige grupper på et plott med en X-akse og en Y-akse (Figur 7). Genotyping ble gjort ved å bruke TaqMan assay, som vi bestilte av Applied Biosystems (Foster City, CA, USA).



Figur 7: TaqMan SNP Genotyping(69).

Konsentrasjonen på genomisk DNA (gDNA) var 20 ng/μl. Fortynnet gDNA med 2x2, siden pipettene ikke gikk ned til 0,5 μl.

Fortynning: 1 μl gDNA + 1 μl H₂O.

Det var to SNPer som skulle genotypes, og dermed var det to TaqMan genotype assay mixer (40x assay mix, Applied Biosystem)(Tabell 5).

Tabell 5: Oversikt over TaqMan SNP genotypings assay.

SNP	Gener	Major/minor allele	LOT	Assay ID
rs2248374	ERAP2	A/G	P120228-005 B07	C_25649529_10
rs30187	ERAP1	C/T	P120301-007 A08	C_3056885_10

I hver mix ble det tilsatt 200 μl 2x PCR mastermix (Applied Biosystems, Foster City, CA, USA), 8 μl 40x TaqMan assay mix (Applied Biosystems, Foster City, CA, USA) og 152 μl sterilt H₂O.

Brukte 1 μl fortynnet gDNA + 4,5 μl mix i hver brønn, totalt 5,5 μl i hver brønn. Det ble brukt en plate med 384 brønner, og pipetter med filter-spiss.

PCR reaksjonen besto av et denaturerings steg på 95 °C i 15 minutter, etterfulgt av 40 syklers hybridisering steg på 60 °C i 1 minutt og deretter polymerisering på 72 °C. PCR ble utført på Geneamp® PCR System 9700, versjon 3.08, serienummer RH 32851 (Applied Biosystems, Foster City, CA, USA).

PCR reaksjonen ble etterfulgt av post-PCR avlesning på 7900HT Fast real-time PCR system (Applied Biosystems). Fluorescens data ble analysert på en automatisert programvare (SDS v2.3, Applied Biosystems).

3.4 Bioinformatisk og statistiske verktøy

3.4.1 SNP exp v 1.1

I løpet av SNP seleksjonen ble det selektert ut totalt 256 risiko SNPs fra de AIS. Det ble brukt en modifisert form av SNP exp programmet(71). Dette programmet ble brukt til å korrelere de 256 SNPene med Immunochip genotyping i tymus og med ekspresjon av nærliggende gener (Microarray data). Det ble korrelert med gener innen 2 Mb. Siden de fleste cis-regulerende SNPer ligger innenfor et vindu på +/- 100 kb, så vil det være ulik sjans for å finne eQTL innenfor dette området og innenfor et område på +/- 1 Mb. På det grunnlaget så er det brukt to ulike p-verdier på 200 kb vindu og 2 Mb vindu. Bonferroni korreksjon gir en p-verdi på 5×10^{-5} innenfor 200 kb vindu rundt SNPene, og 5×10^{-6} for 2 Mb vindu rundt SNPene(52). Bonferroni korreksjonen reduserer sjansen for at det skal forekomme sammenhenger av tilfeldige årsaker, når man utfører korrelasjonsanalyser mellom mange tallgrupper(52).

3.4.2 mRNA by SNP Browser v 1.0.1

mRNA by SNP Browser er et program som gir grafiske oversikter over helgenom-assosiasjons studier med en mengde fenotypisk informasjon om genekspresjon. Programvaren inneholder en eQTL database, som undersøker assosiasjon mellom 54675 transkripter og 406912 SNPs i data fra 400 individer. For hvert transkript vil programmet plote test statistikken (p-verdi <0,001), estimere effekt størrelsen og allel informasjon på tvers av genomet. I tillegg til å søke på spesifikke transkripter og posisjon, kan man også søke direkte på en spesifikk SNP. Nettleseren kobler automatisk resultatene til <http://genome.ucsc.edu/>, hvor det er mulig og undersøke hvert transkript eller SNP i en genomisk sammenheng. LD og tag informasjon på SNPs er gitt av det internasjonale HapMap konsortiet.

Dette programmet ble brukt til å søke opp SNPs i ERAP1/ERAP2-regionen, som ble funnet igjen i både immunochip-data og microarray-data før SNP exp analysen ble utført. Da ble det søkt direkte på SNP,

for å finne ut hvilke gener som korrelerte med hver SNP(72). Det ble brukt en signifikans grense $\alpha = 0,05$.

3.4.3 GraphPad Prism 4

GraphPad Prism 4 (GraphPad Software Inc, San Diego, CA) er et statistikk program som brukes til å analysere og presentere vitenskapelig data. Programmet ble brukt til å se på data fra ImmunoChip genotyping, for å se hvordan genotypen fordelte seg for de mest signifikante SNPs fra SNP eksp programmet (13 SNPs). Dette blir gjort ved å sette opp plots for hver SNP, prøvene grupper seg deretter i tre forskjellige grupper (AA, AG, GG) på et plott med en X-akse og en Y-akse(73).

I GraphPad ble variansanalysen One way ANOVA (Analysis of Variance, ikke-parametrisk) utført. Mann-Whitney U-test (two-tailed) ble brukt på 2-gruppers analyse, og Kruskal-Wallis test på 3-gruppers analyse. ANOVA ble utviklet av statistikeren Fisher, for å kunne studere mange tallgrupper samtidig. Analysen tester om tallgruppens middelerverdier kommer fra samme populasjon, og metoden kan betraktes som en utvidelse av t-testen for uavhengige tallgrupper(74).

Det er utviklet hypotesetester for å sammenlikne avhengige og uavhengige tallgrupper dersom gruppene er ikke-parametriske, Wilcoxon's og Mann-Whitney's test. Testen for uavhengige grupper har to varianter, Wilcoxon's rangtest (rang sum test) og Mann-Whitney's U-test

Mann-Whitney's U-test utføres forskjellig fra Wilcoxon's rangtest, men gir samme p-verdi. Mann-Whitney's U-test og Wilcoxon's test har forskjellige testobservatorer og disse sammenliknes med teoretiske verdier i tabeller. Sammenlikningen med teoretiske verdier blir utført automatisk på GraphPad.

Kruskal-Wallis test er en ikke-parametrisk ensidig varians analyse, som baserer seg på samlet rangordning og gruppevise rangberegninger. Når man bruker Kruskal-Wallis test samler man alle observasjonene og sorterer og tildeler en rangverdi, basert på ordinal rangering av observerte verdier. Man skiller gruppene og beregner gjennomsnittlig rangverdi for hver gruppe.

Kruskal-Wallis testen har sin styrke i sin anvendelighet som ikke-parametrisk test av gruppeforskjeller mellom flere enn 2 grupper. Den forutsetter at den overordnede fordeling av observasjoner er ganske lik mellom gruppene, med unntak av median verdien i gruppen(73).

3.4.4 DataAssist

DataAssist er en programvare som inneholder verktøy for sammenligning av prøvemateriale ved bruk av komparativ CT ($\Delta\Delta\text{CT}$) metode for å beregne relativ kvantifisering av genuttrykk. Ct-verdien oppgir antall cykler (Ct) den enkelte prøve bruker for å nå en treshold-grense, som ligger i området den eksponentielle fasen starter. Ct-verdien vil være ulik for hver prøve avhengig av hvor mye templat som var i prøven i utgangspunktet, mengden fluorescerende signal er direkte proporsjonalt med antall PCR-produkt molekyler som er generert i den eksponentielle fasen av reaksjonen.

Programvaren importerer rådata fra opptil hundrevis av prøver, utført på TaqMan array plater (356 brønner), og normaliserer prøvemateriale ved å bruke endogen kontroll (B2M). Analyse resultatene blir fremstilt i tabeller og grafiske plott-diagrammer.

RQ Manager 1.2.1 ble brukt til å formatere filene, og laste opp filene i DataAssist(75).

3.4.5 Ingenuity Pathway Analysis (IPA)

Ingenuity Pathways Analysis (IPA) er en programvare som hjelper forskere med å analysere og forstå komplekse biologiske og kjemiske systemer. Dette er en database som inneholder millioner av individuelle modulerte sammenhenger mellom proteiner, gener, celler, vev og sykdommer. IPA gjør det enkelt å relatere informasjonen i Ingenuity Knowledge Base (IKB) til eksperimentelle datasett, for å identifisere signalering og metabolske reaksjonsveier, nettverk av sammenhengende molekyler og biologiske mekanismer og sykdomstilstander av mest betydning for datasettet(76).

IPA krever en genliste som består av karakteriserte gener for å kunne utføre en analyse på datasettet. For å kunne identifisere probene som representerer genene, må probe-data microarray legges inn (Illumina human WG-6v3). IPA ble brukt til å utføre en signalvei analyse på de 100 mest signifikante Microarray-dataene ($p < 5 \cdot 10^{-5}$).

Pathway analysen ble utført ved å bruke right-tailed Fishers eksakte test og Benjamin-Hochberg (B-H) metoden, som justerer p-verdiene ved flere tester. B-H metoden blir brukt til å beregne falsk funnrater (FDR-False Discovery Rate) verdiene. FDR er en statistisk metode som brukes etter flere hypotesetester til å korrigere for flere sammenligninger(77). Under biologiske funksjoner er det oppgitt fra en p-verdi – til en p-verdi. Dette er fordi man ved biologiske funksjoner tester flere underfunksjoner, og derfor får flere p-verdier.

3.4.6 HaploView v 4.1

HaploView er en bioinformatisk programvare som analyserer og visualiserer linkage disequilibrium (LD) mønster. Programmet er utviklet for å gi en omfattende pakke med verktøy for haplotype analyse, og for å gi et bredt utvalg av datasett størrelser. HaploView genererer markør kvalitet statistikk, LD informasjon, haplotype blokker og populasjonsdynamikk haplotype frekvenser. Alle beregninger på HaploView utføres i sanntid, selv for datasett med hundrevis av individer og hundrevis av markører. Programmet ble brukt for å se på sammenhengen mellom SNPene ERAP1/ERAP2-regionen.

Det er mulig å hente inn genotype data direkte fra Human HapMap nettsiden (<http://www.hapmap.org>), ved å kun definere posisjonen av interesse. Dette ble gjort for alle ERAP1/ERAP2 SNPene, utenom for ERAP2 SNPen rs2248374 som ikke var genotypet på HapMap. Datakilden som ble brukt er HapMap Data Rel 28 phase II + III, Aug. 10, on NCBI B36 assembly, dbSNP b126 (ERAP1 posisjon: 96122004-96175000)(78).

LD mellom SNP polymorfismene var oppgitt som D prime (D') som angir verdien mellom de to loci, og R square (r^2) som er korrelasjonskoeffisienten mellom dem. LD mellom SNPene er basert på graden av korrelasjonen mellom sammenkoblede markører ved å måle forskjellen mellom den observerte haplotype frekvens og forventet haplotype frekvens hvis allelene segregere tilfeldig(79).

3.4.7 Unphased v 2.404

Unphased er et program som kan brukes innenfor mange områder for å utføre genetiske analyser. Programmet tilbyr mange av de vanligste utførte analysene som f.eks LD tester, globale og individuelle tester for haplotyper, tester som står for sammenslutninger av nærliggende loci, tester av gen-gen interaksjon, genotype tester og sammenligning av risiko mellom ulike haplotyper. Unphased ble tatt i bruk for å undersøke om det er en sammenheng mellom HLA klasse I og risiko SNPene i ERAP/ERAP2-regionen. Korrelasjonen kunne ikke utføres på HaploView (punkt 3.3.6) på grunn av visse begrensninger. HaploView tar ikke polymorfismer med mer enn to alleler, og kan dermed ikke brukes på HLA data.

Cramers V er en statistisk måleverdi i Unphased, og gir en overordnet korrelasjon mellom allelene i genotype dataen som lastes opp. Det er en måte å beregne korrelasjonen i tabeller som har mer enn 2x2 rader og kolonner. Verdien brukes som en post-test for å avgjøre assosiasjons styrken etter utførelse av chikvadrat-test (X^2 -test). Cramers V varierer mellom 0 og 1. En verdi nær 0 viser nesten ingen sammenheng, mens en verdi nær 1 indikerer en sterk sammenheng(80).

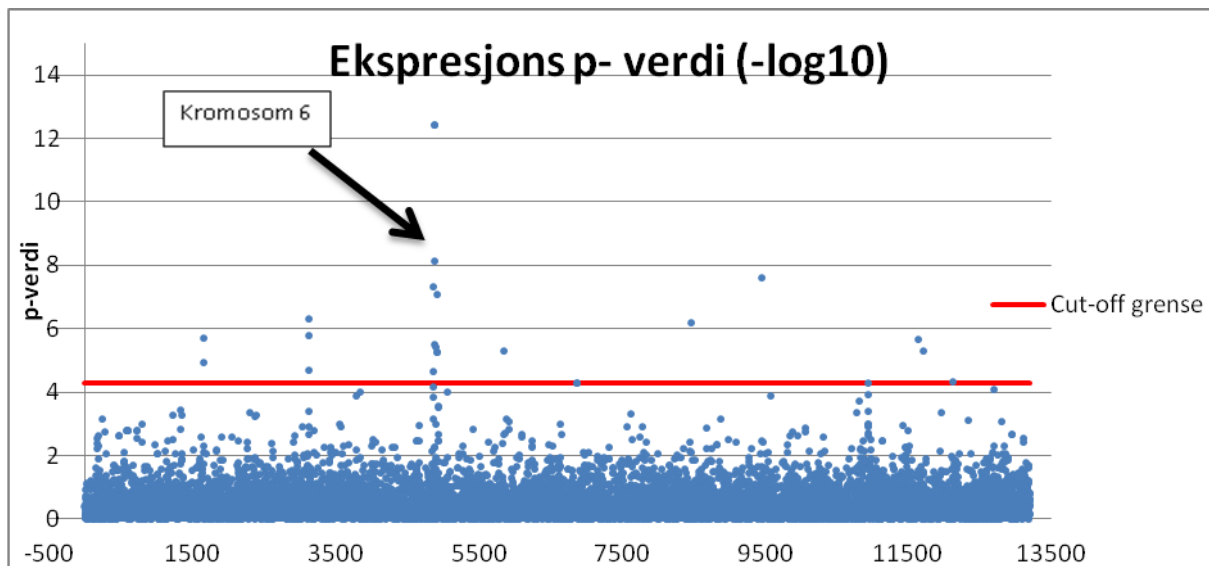
4 Resultater

4.1 Korrelasjon mellom genekspresjon og autoimmune risikopolymorfismer

Ved å bruke en modifisering av programmet SNP exp fikk vi korrelert Immunochip genotyping i tymus med ekspresjon av nærliggende gener ved bruk av probeintensitet. Etter korrelasjon ble det laget en fil som bestod av 256 mapper, en for hver SNP. I hver mappe lå det informasjon om p-verdi, ekspresjonsverdi for nærliggende gener og probe-id. Denne informasjonen ble brukt for å få oversikt over korrelasjonen mellom genekspresjon og SNPgenotyper (se under).

4.1.1 eQTL analyser av alle AIS SNPer mot nærliggende transkripter

Genotypene for alle de 256 AIS (Vedlegg A) risiko SNPene ble korrelert mot transkripsjonsnivå av Microarray prober innenfor et område på +/- 1 Mb. Et plott over signifikansnivået for korrelasjon mellom alle SNP genotype-genekspresjonsprobeparene er vist i figur 8. 13 risiko SNPer utenfor kromosom 6 (HLA komplekset) viste signifikante avvik ($p < 5 \times 10^{-5}$) i ekspresjonsnivå mellom ulike genotyper (Figur 8). Kromosom 6 ble fjernet fordi disse AIS SNPene ligger i HLA komplekset, hvor genene er svært polymorfe og ekspresjonsprobene som dekker disse genene ikke fanger opp alle allelvarianter. Signifikansnivået ble satt til $p < 5 \times 10^{-5}$ noe som representerer en Bonferroni korreksjon av antall eQTL (expression Quantitative Trait Locus) analysesammenlikninger innen et +/- 100 kb område. Dette vinduet ble valgt da det har vært vist at de fleste cis-regulerende SNPer ligger innenfor nettopp et vindu på +/- 100 kb(52). Det vil være større sjanse for å finne eQTL innenfor dette området enn innenfor et område på +/- 1 Mb. Vi brukte derfor to ulike p-verdigrenser, en for korrelasjon innenfor 200 kb vinduet og en annen for korrelasjon innenfor 2 Mb vinduet (52). Bonferroni korreksjon gir en p-verdi på 5×10^{-5} innenfor 200 kb vindu rundt SNPen, og 5×10^{-6} for 2 Mb vindu rundt SNPen. Siden vi hadde satt det initielle signifikansnivået til $p < 5 \times 10^{-5}$, gikk vi derfor videre og undersøkte SNP-genekspresjonskorrelasjonene i hver enkelt av disse 13 regionene for å avgjøre avstanden mellom SNPen og proben.



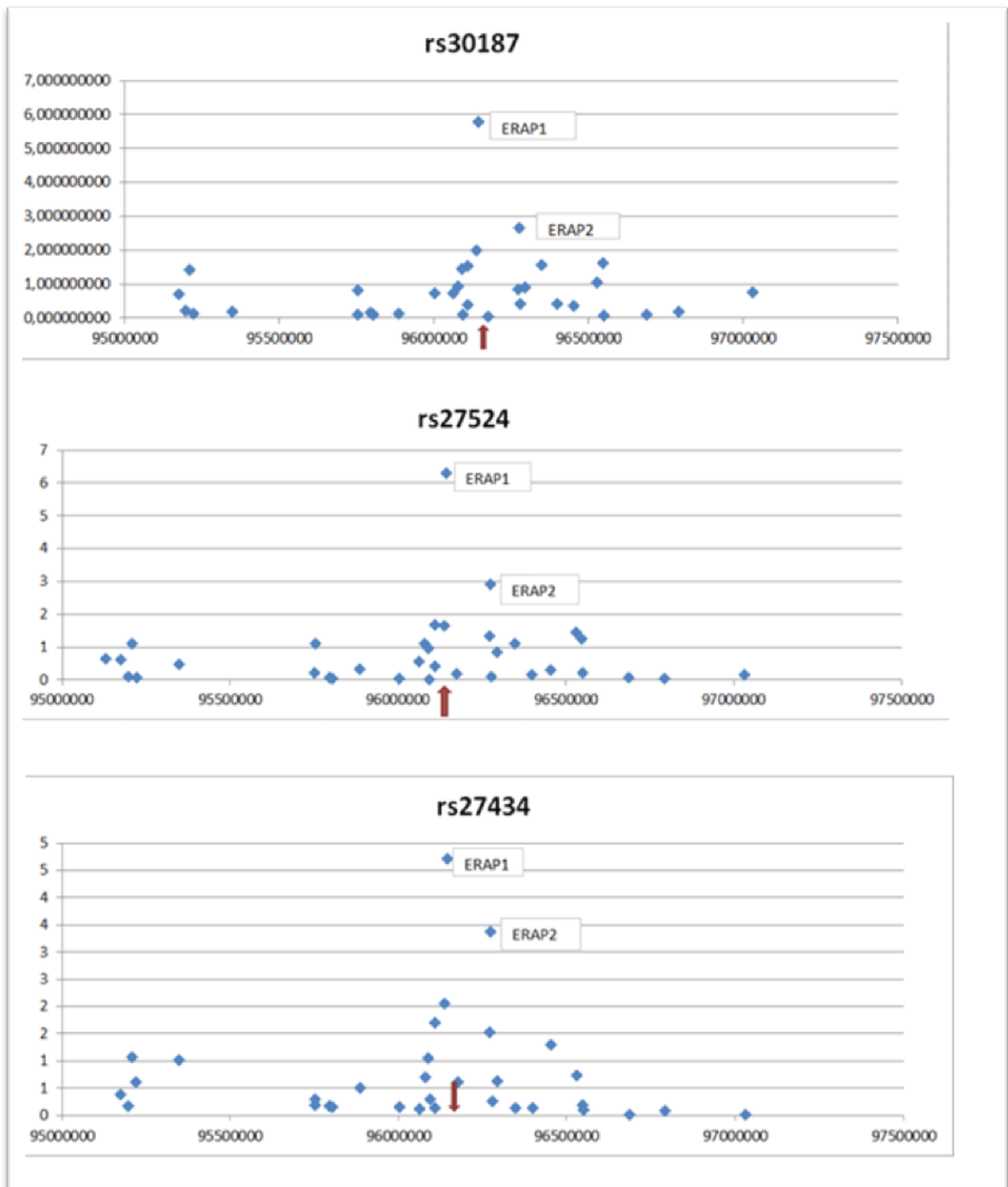
Figur 8: Illustrerer negativ logaritme (-log₁₀) av p-verdien for alle AID risk genotyper mot ekspresjonsnivåer (innenfor +/- 1Mb), hver spott representerer p-verdi for et SNP genotype-ekspresjonsprobe par.

4.1.2 13 eQTL genområder

For å kartlegge hvor langt fra hver av disse SNPene probene/genene som viser differensiell ekspresjon ligger, ble hvert enkelt område undersøkt (Vedlegg B og Figure 9). Bare for fem av de 13 Microarray-probene som viste korrelasjon med SNP genotype (rs45450798, rs10781500, rs7608910, rs13003464, rs4788084), lå transkriptet mer enn 100 kb unna SNPen, og disse nådde ikke den mer stringente signifikans-grensen ($p < 5 \times 10^{-6}$), utenom rs13003464 ($p = 2 \times 10^{-6}$) og rs4788084 ($p = 2.5 \times 10^{-8}$), som ble satt for dette ytre vinduet.

Blant de resterende 8 signifikante eQTLene, var det en genregion (ERAP1/ERAP2) som viste særlig konsistente funn ved at flere SNPer i regionen var korrelert med flere prober i disse genene (Figur 9). På plottene for de tre SNPene (rs30187, rs27524, rs27434), er det tydelig at det er to gener som er mest assosiert, ERAP1 og ERAP2, med p-verdier under 5×10^{-5} . ERAP1 ekspresjonen viser gjennomgående den aller mest signifikante assosiasjonen med alle de tre AIS SNPene. Samtidig viser SNPene, særlig rs27434, viser assosiasjon med intensiteten av flere prober i området, selv om disse ikke når signifikansnivået satt for screeningen. Det at både flere prober og flere SNPer viser assosiasjon i dette området styrker funnet, og sannsynligheten for at det ikke bare er en tilfeldig observasjon.

De øvrige eQTL viser assosiasjon til andre gener som, CAST, LNPEP, RIOK2, PCSK1, GLRX, LRAP, LIX1, FIS, RHOBTB3 og LOC642716 (Vedlegg B).



Figur 9: Plott over AIS SNPene rs30187, rs27524 og rs27434 og deres korrelasjon med genekspressjon av ERAP1 og ERAP2. Y-aksen viser $-\log_{10}(p\text{-verdi})$, X-akse viser SNP-posisjon på kromosomet. Pilene på hver graf indikerer hvor den testede SNP'en ligger.

4.2 Kvalitetssikring av genotype data

Genotype dataene, som ble brukt som utgangspunkt for å korrelere mot genekspressjonen i tymus, ble utført på Immunochip som er en Illumina chip. Immunochipen inneholder 200 000 SNP'er og en overordnet kvalitetssikring ble gjennomført før dataen ble tatt i bruk. For å sikre kvaliteten på de 10 eQTLene, ble genotypedataene for disse SNPene i tillegg undersøkt nærmere.

4.2.1 Genotypingssuksessrate og Hardy Weinberg Equilibrium (HWE)

Det ble sett på genotype suksessraten (GSR) for de 10 SNP polymorfismene. GSR sier noe om hvor stor andel av de 42 tymus prøvene, som har blitt suksessfullt genotypet. En høy GSR (ideelt sett over 95%) er viktig for å sikre at ikke visse genotyper (allelkombinasjoner) har blitt mistet under genotypingen og at noen genotyper har vært vanskeligere å detektere slik at det oppstår en skjevhet i materialet. Immunochip genotypingen gav et godt resultat (GSR>95 %) for alle 10 SNP'er (Tabell 6), utenom rs30187 som hadde 8 prøver uten genotyperesultat, og dermed hadde en GSR= 81 %. Denne SNP'en ble derfor genotypet på nytt ved å bruke TaqMan genotyping assay (C_3056885_10).

Etter at vi hadde funnet at ERAP1 og ERAP2 genene ga interessante korrelasjonsfunn mot AIS SNPene, ble vi oppmerksomme på ERAP2 SNP'en rs2248374 i litteraturen (81, 82). SNP'en ligger i et intron av ERAP2 genet, veldig nær et ekson og kan dermed være funksjonell mht genekspressjon og alternative transkripter. Denne var ikke var tatt med i den opprinnelige listen over AIS SNP'er som ble testet i studien, og ble derfor genotypet med TaqMan alleldiskriminering.

Resultater for alleldiskrimineringsgenotypingen av disse to SNPene vil bli omtalt under punkt 4.2.2.

Hardy Weinberg likevektsprinsippet (HWE) sier at begge allelene og genotype frekvensene i en populasjon er i likevekt fra generasjon til generasjon, med mindre spesifikke forstyrrende faktorer er påvirkende. Påvirkende faktorer kan for eksempel være nymutasjoner, ikke-tilfeldig partnervalg, seleksjon eller innflytting. Et avvik fra HWE kan blant annet bety at det er seleksjon for spesifikke genotyper. Generelt kan man forvente at SNPene man undersøker ved komplekse sykdommer er i HWE og et avvik fra HWE kan indikere mulige problemer med genotyperesultatene, f.eks at en del av prøvene for en genotypekombinasjon har blitt feiltypt, f.eks ved alleltap.

I tabell 6 vises en oversikt over p-verdiene for avvik fra HWE for de 10 eQTL SNPene. Alle de testede SNPene var i HWE, foruten en av SNPene (rs4788084) som viste et signifikant avvik fra HWE ($p=0,01$), men med 100 % GSR.

Tabell 6: Genotypingskvalitetssikring av de10 signifikante SNPene ($p < 5 \times 10^{-5}$), GSR og HWE (HWE=1,0 betyr full likevekt).

	SNP	Antall mislykkete prøver	Genotype suksessrate (%)	HWE (p-verdi)
1	rs30187	8	80,95	0,39
2	rs13003464	2	95,2	0,23
3	rs27524	0	100	0,35
4	rs27434	0	100	0,45
5	rs7804356	1	97,6	0,57
6	rs2292239	0	100	0,46
7	rs4788084	0	100	0,01
8	rs2281808	0	100	0,76
9	rs1008953	1	97,6	0,97
10	rs1297265	2	95,2	0,67

4.2.2 Retyping av rs30187 og genotyping av rs2248374

På grunn av lav genotypesuksessrate (81 %) ble SNP rs30187 retypet ved TaqMan alleldiskriminering. Alle prøver unntatt en lot seg genotype i denne retypingen (GSR = 97,6 %). Alle genotypene som ble suksessfullt genotypet for hver enkelt prøve ved begge teknikker, var i 100 % overensstemmelse (Se vedlegg 2). De prøvene med manglende genotypedata (N=8) fra ImmunoChip-typingen viste seg å være jevnt fordelt mellom ulike genotyper (Tabell 7).

Tabell 7: Genotypingsresultater for rs30187

rs30187	Mislykket prøver	A/A-genotype	A/G-genotype	G/G-genotype	GSR (%)	HWE
ImmunoChip	8	5	13	16	80,95	0,39
TaqMan	1	8	15	18	97,6	0,16

Rs2248374 har ikke vært rapportert assosiert med autoimmune sykdommer, og det er derfor den ikke kom med på SNP-listen tidligere. Rs2248374 er en SNP som er funksjonell i henhold til genespresjonen av ERAP2, den er koblet til alternativ spleisevarianten på G-allelet(81, 82). Etter genotyping med TaqMan assay (C_25649529_10) lot alle prøver seg type utenom fire (Tabell 8). For rs2248374 var genotype suksessrate 90,5 %, og SNPen var i HWE ($p=0,36$).

Tabell 8: TaqMan typing av rs2248374.

SNP	Mislykket prøver	A/A-genotype	A/G-genotype	G/G-genotype	GSR (%)	HWE
rs2248374	4	9	16	13	90,5	0,36

4.3 Genespresjonsforskjeller for de signifikante eQTL i ERAP1/ERAP2-regionen

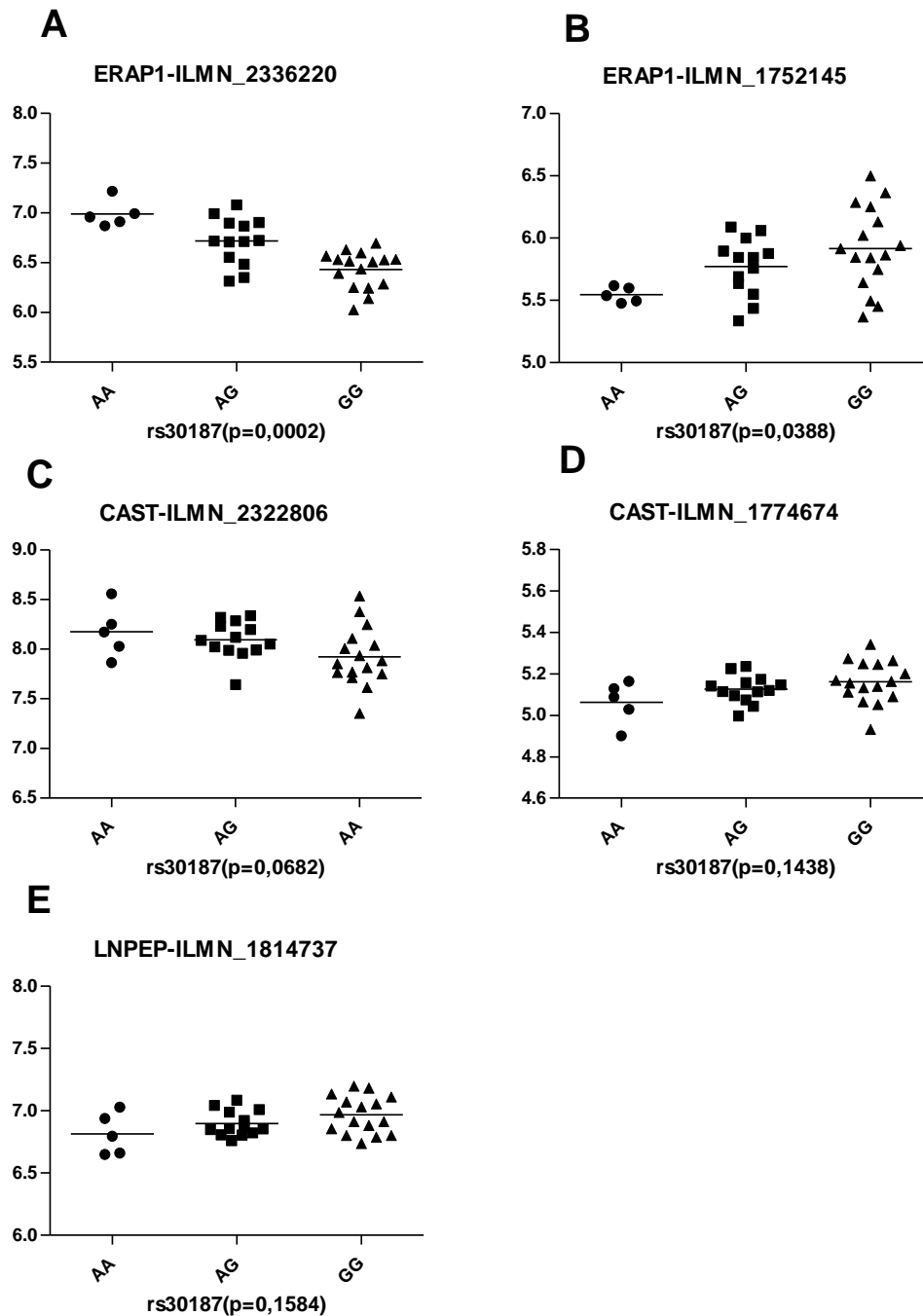
Siden de initielle screeningresultatene bare ga p-verdier, og ikke viste de faktiske ekspresjonsnivåene og forskjellene i disse for de tre ulike genotypene for hver SNP, ønsket vi å se nærmere på dette. For å karakterisere assosiasjon mellom SNP genotypen og genespresjonen, ble genotype data fra Immunochip og/eller Taqman alleldiskriminering korrelert med genespresjonsdata fra Microarray.

Siden rs30187 ble genotypet med to metoder, der Immunochip genotypingen (som ble brukt i den initielle screenen) hadde lavere GSR enn TaqMan genotypingen, ble begge genotyperesultatene brukt til korrelasjon med genespresjonsdataene. Etter screening var det 8 prober som var signifikant assosierte, i henhold til signifikans-grensen ($p < 5 \times 10^{-6}$) som ble satt for det ytre vinduet, av disse var det de fem mest assosierte probene som ble tatt med på korrelasjonen. For å illustrere forandringer i signifikansnivået på rs30187 (med en signifikans grense på $p=0,05$) ble det sett på korrelasjon mellom de fem probene, som initielt ble viste tegn til assosiasjon i screeningen, mot genotype data fra Immunochip (Figur 10) og Taqman (Figur 11), uten å begrense det til det stringente korrigerede signifikansnivået.

Det var to prober i ERAP1 genotom som var signifikant assosiert med rs30187 genotype data fra både Immunochip og Taqman, nemlig probene ILMN_2336220 (Immunochip $p=0,0002$, Taqman $p < 0,0001$) og ILMN_1752145 (Immunochip $p=0,0388$, Taqman $p=0,0057$). Det var dermed tre av fem prober som ikke var signifikant ($p < 0,05$) assosierte med rs30187 (Figur 10C, 10D, 10E), etter retyptingen (Figur 11D, 11E), disse to probene er assosierte med genene CAST og LNPEP. Etter retypting var det en

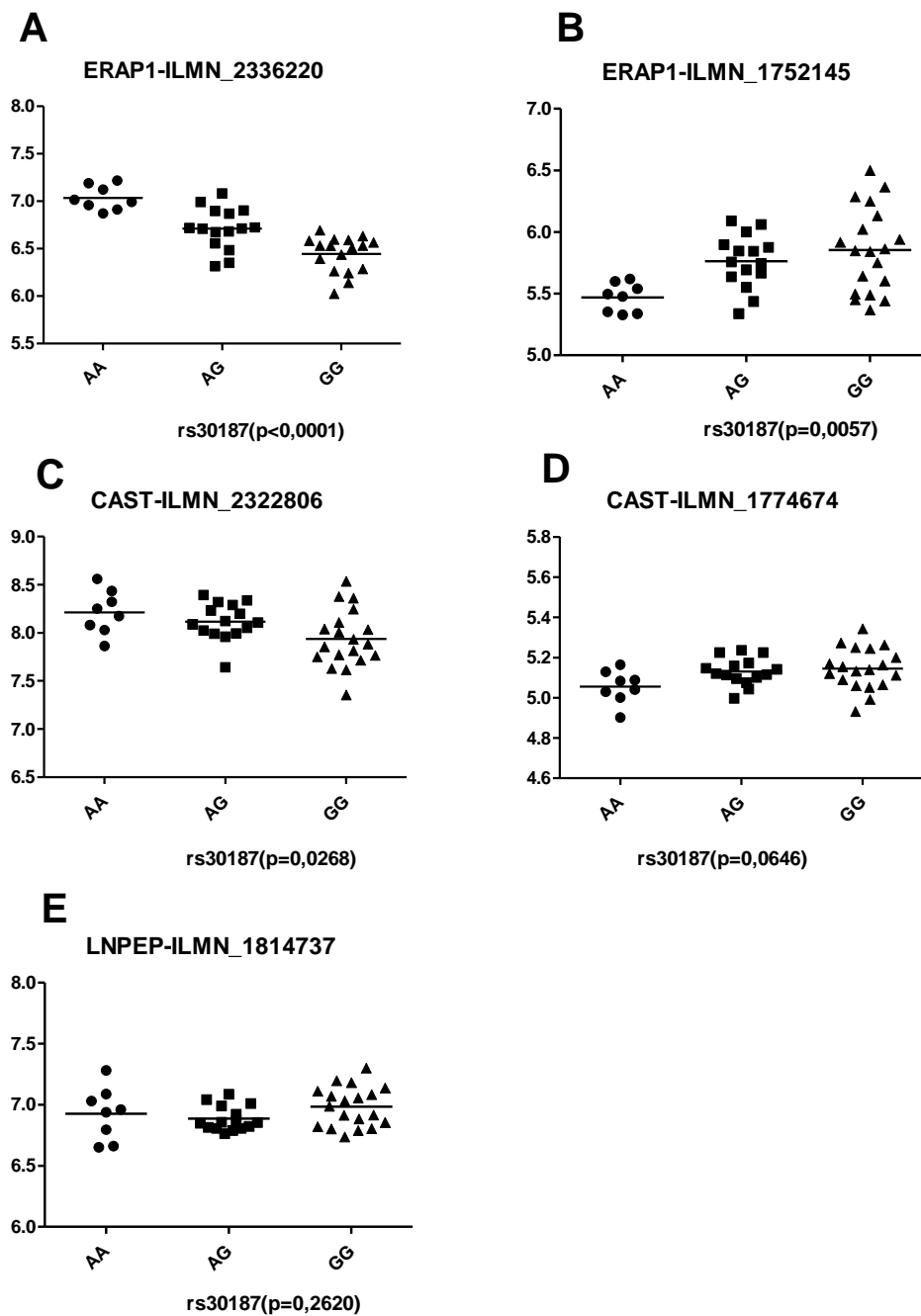
probe som ble signifikant assosiert, ILMN_2322806 (ImmunoChip $p=0,0682$, Taqman $p=0,0268$) er assosiert med genet CAST som overlapper ERAP2.

Genekspresjon vs Genotype (8 failet)



Figur 10: ImmunoChip genotype vs Microarray genekspresjon for rs30187

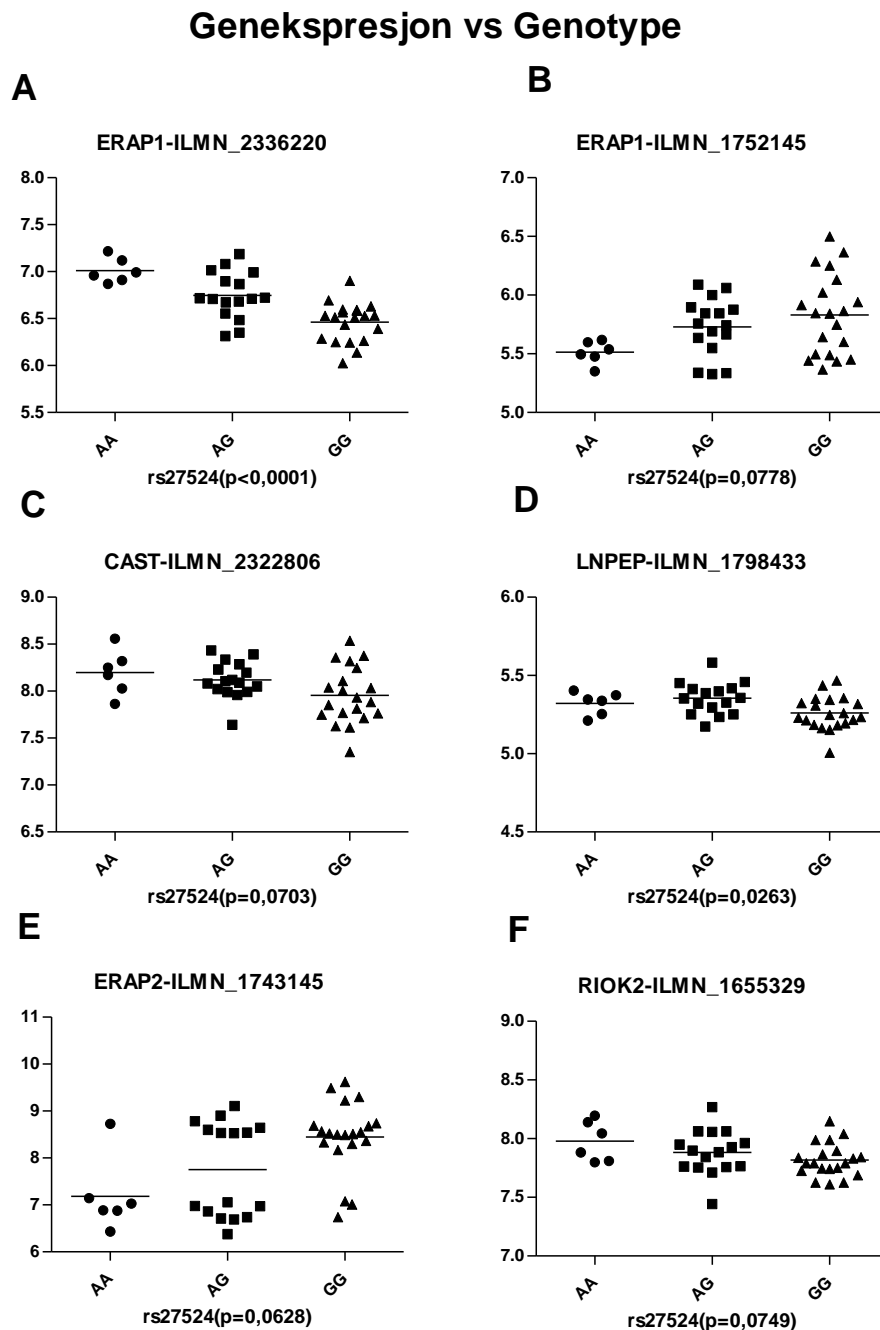
Genekspresjon vs Genotype(retyping)



Figur 11: Retyping av rs30187. TaqMan genotype vs Microarray genekspresjon.

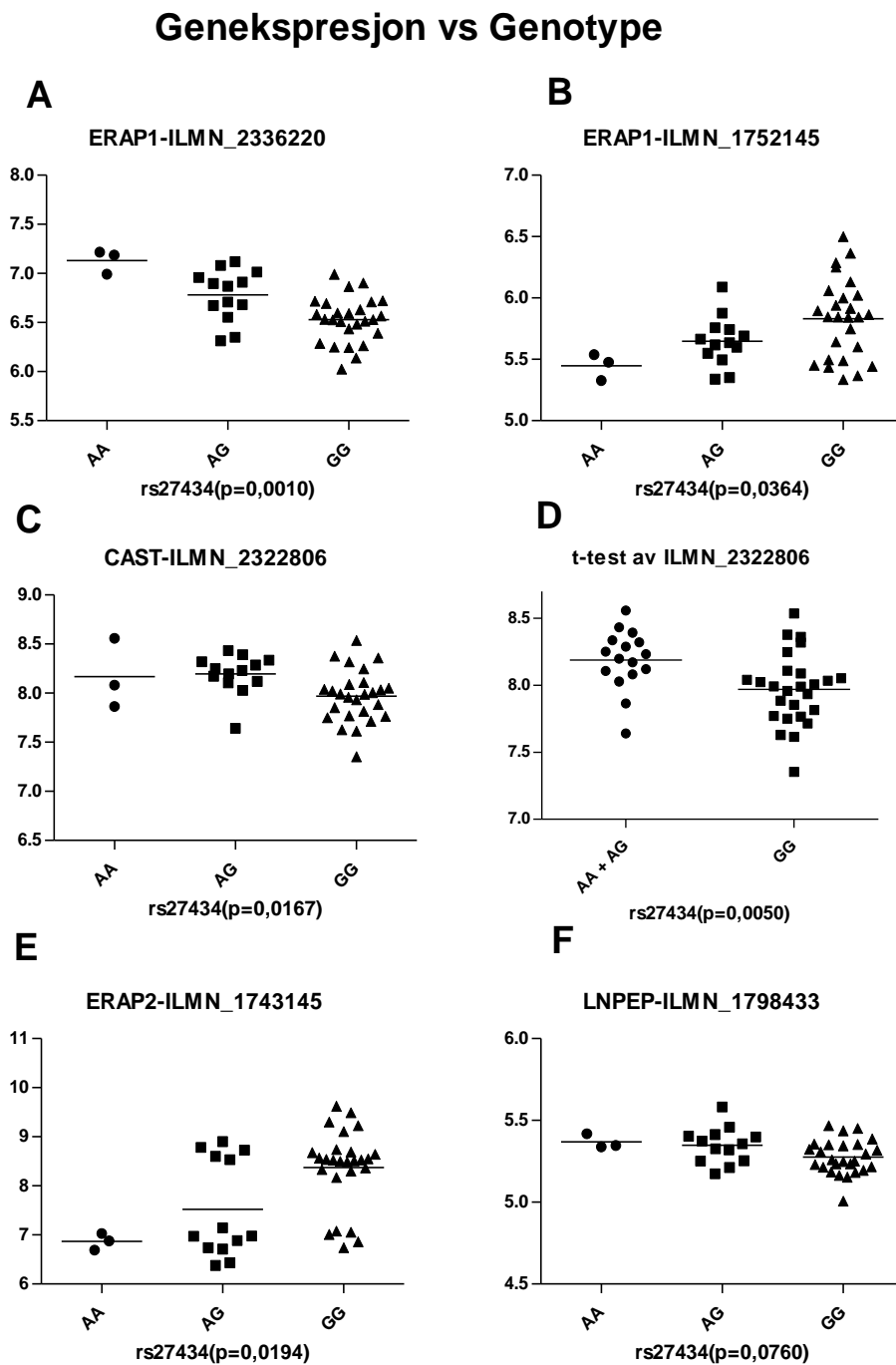
En annen SNP i den samme ERAP/ERAP2-genregionen, rs27524, ble testet mot ekspresjonsnivåer av genene ERAP1, ERAP2, LNPEP og RIOK2 (Figur 12). To prober viste signifikant assosiasjon, ERAP1 proben ILMN_2336220 ($p < 0,0001$) og LNPEP proben ILMN_1798433 ($p = 0,0263$).

ERAP2 proben ILMN_1743145 viste en høyere signalstyrke enn de andre probene (Figur 12E), det er i tillegg en fordeling innad i de tre genotype-gruppene. Fordelingen av genotypene kan blant annet bety at rs27524 er i LD med en annen SNP, og kan indikere at genekspressjonsassayet kan fange opp flere ulike spleisevarianter.



Figur 12: ImmunoChip genotype vs Microarray genekspressjon for rs27524

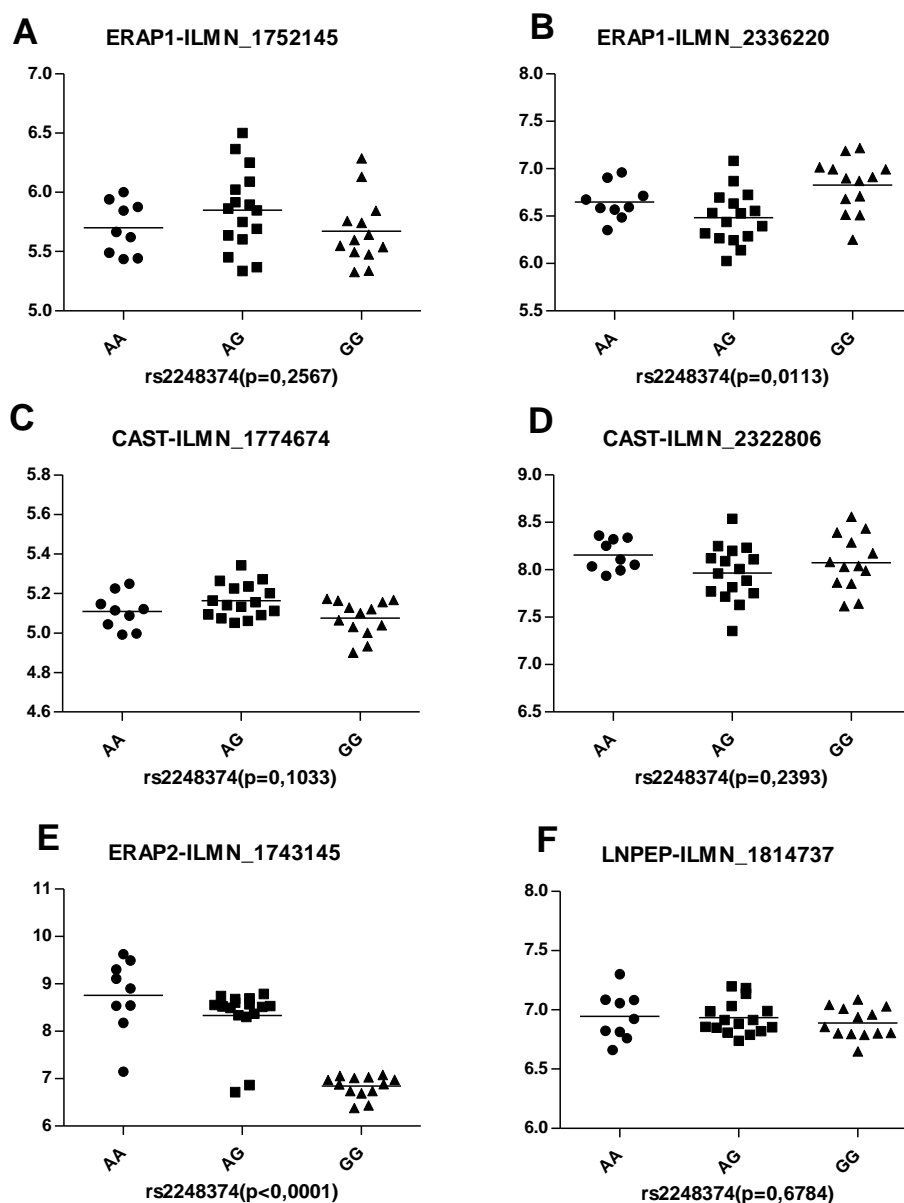
rs27434 ligger innenfor ERAP1/ERAP2-regionen. Alle probene var signifikant assosierte med rs27434, utenom en LNPEP proben ILMN_1798433 ($p=0,0760$). ERAP2 proben (ILMN_1743145) viste derimot en signifikant assosiasjon til rs27434, med det samme mønsteret med fordeling innad i genotype-gruppene (Figur 13E).



Figur 13: ImmunoChip genotype vs Microarray genekspresjon for rs27434

SNPen rs2248374 er funksjonell i henhold til genekspressjonen av ERAP2. rs2248374 genotyper ble korrelert mot 6 prober i regionen (Figur 14), oversikt over alle probene ligger i vedlegg D. På rs2248374 er genotypen godt korrelert med den samme ERAP2-proben (ILMN_174314)(Figur 14E) som viste nivåforskjeller innad i genotype-gruppene for SNPene rs27524 og rs27434, og den er sterkt signifikant ($p < 0,0001$). Dette kan komme av at proben er komplementær med det transkriptet denne ERAP2 SNP polymorfismen dirigerer alternativ spleising av. I tillegg til denne ERAP2-proben ILMN_174314 ($p < 0,0001$), er det kun ERAP1 proben ILMN_2336220 ($p = 0,0113$)(Figur 14B) som viser signifikant assosiasjon med ERAP2 SNPen rs2248374.

Genekspressjon vs Genotype



Figur 14: TaqMan genotype vs Microarray genekspressjon for rs2248374.

4.4 ERAP1 og ERAP2 genekspressjon

For nærmere undersøkelse av genekspressjonen i ERAP1 - og ERAP2-regionen, ble det utført en genekspressjonsanalyse ved å bruke TaqMan genekspressjonsassay. De fem TaqMan genekspressjonsassayene (Tabell 3) ble valgt på Applied Biosystems (Foster City, CA, USA), for å detektere de fleste kjente transkriptene på ERAP1 og ERAP2 genene, samt spesielle enkelt transkripter. Isolert cDNA fra 42 tymus-prøver ble fortynnet, og beta-2 microglobulin (B2M) ble brukt som endogen kontroll (Tabell 4).

Begge ERAP1 assayene gav gode resultater (Tabell 9). Det første ERAP1 assayet (Hs00429970_m1) dekker alle tre transkriptene, og gav vellykkede ekspresjonsresultater. Det andre ERAP1 assayet (Hs00372304_m1) dekket kun ett av tre transkripter, ekspresjonsresultatene var gode på de fleste tymus-prøvene, men siden det var dårlig triplikater på 18 tymus-prøver måtte dette assayet rekjøres. Rekjøringen av ERAP1 assayet gav ingen bedring i resultatene, dermed ble den avvikende triplikaten fjernet og duplikatene brukt til videre analysing.

ERAP2 assayene hadde kun ett vellykket assay (Hs01073631_m1), hvor Ct verdiene havnet innenfor godkjent område, men også der var triplikater på noen tymus-prøver dårlige, dermed ble det assayet også rekjørt. Rekjøring av ERAP2 assayet gav heller ingen bedring resultater, og også her ble duplikatene brukt til videre analyse.

Tabell 9: Oppsummering av qPCR resultatene.

	Gen	Assay	Transkript	Resultat	Kommentar
1	ERAP1	Hs00429970_m1	NM_001040458.1	Vellykket	Godt uttrykk på alle prøvene, ingen av triplikaterne måtte tas ut.
			NM_001198541.1	analysering.	
			NM_016442.3	Ct-verdier: 30.	
2	ERAP1	Hs00372304_m1	NM_016442.3	Ok kjøring. Ct-verdier: 25-30.	Dårlig triplikater på 18 tymus-prøver, brukte duplikatene.
3	ERPA2	Hs01073631_m1	NM_001130140.1	Vellykket	Kun 5 tymus-prøver med
			NM_022350.3.	analysering.	

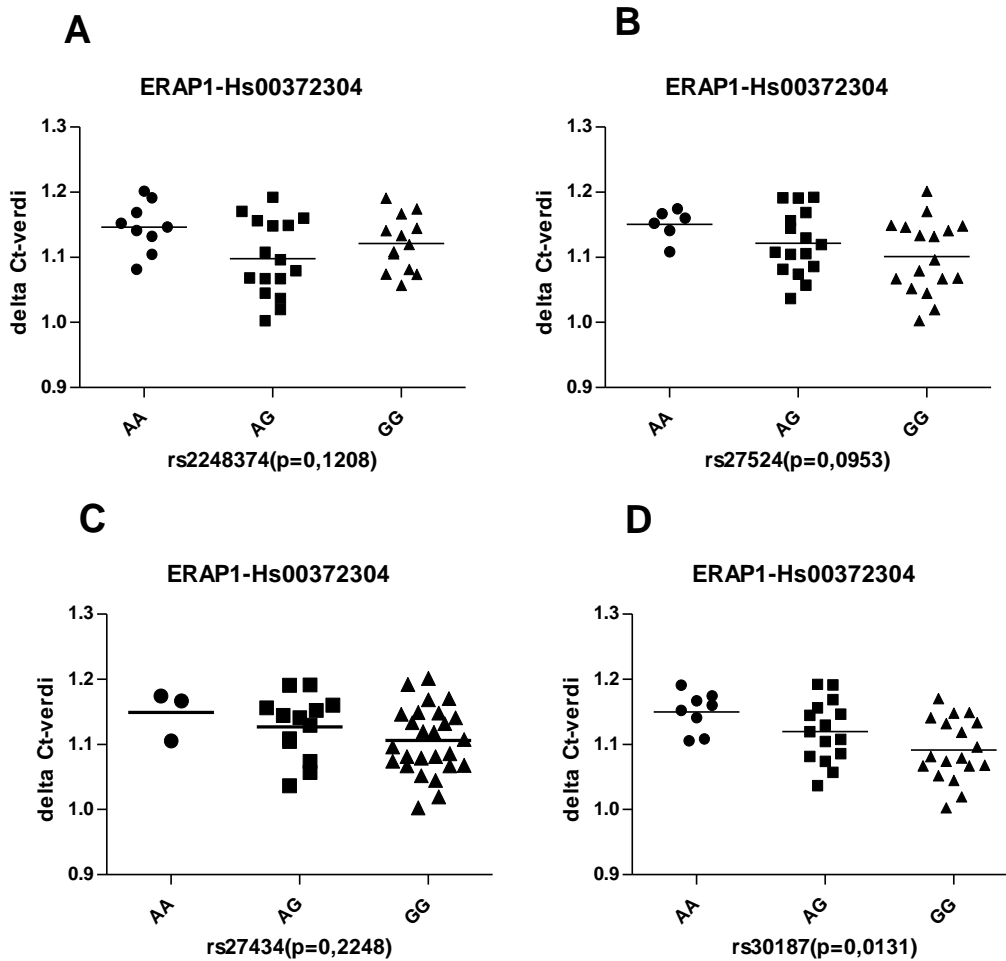
				Ct-verdier: 29-33.	dårlig triplikater, pga. stdev>0,167.
4	ERAP2	Hs01082025_m1	NM_022350.3.	Svak kjøring med høye Ct-verdier: 33-35	Kun en prøve som gir ok resultat, T34.
5	ERAP2	Hs01073630_m1	NM_001130140.1 NM_022350.3.	Svak kjøring med høye Ct-verdi: 33-35	Veldig dårlig triplikater for 22 av 42 tymus-prøver.

Etter en gjennomgåelse av de analyserte TaqMan genekspressjons resultatene, med standardavvik (STDEV) cut-off grense = 0,165, var det kun tre av fem velfungerende assay (ERAP1: Hs00429970_m1, ERAP1: Hs00372304_m1, ERAP2: Hs01073631_m1) som ble brukt til videre analysering av resultatene.

4.4.1 TaqMan genekspressjonsassay vs Genotype

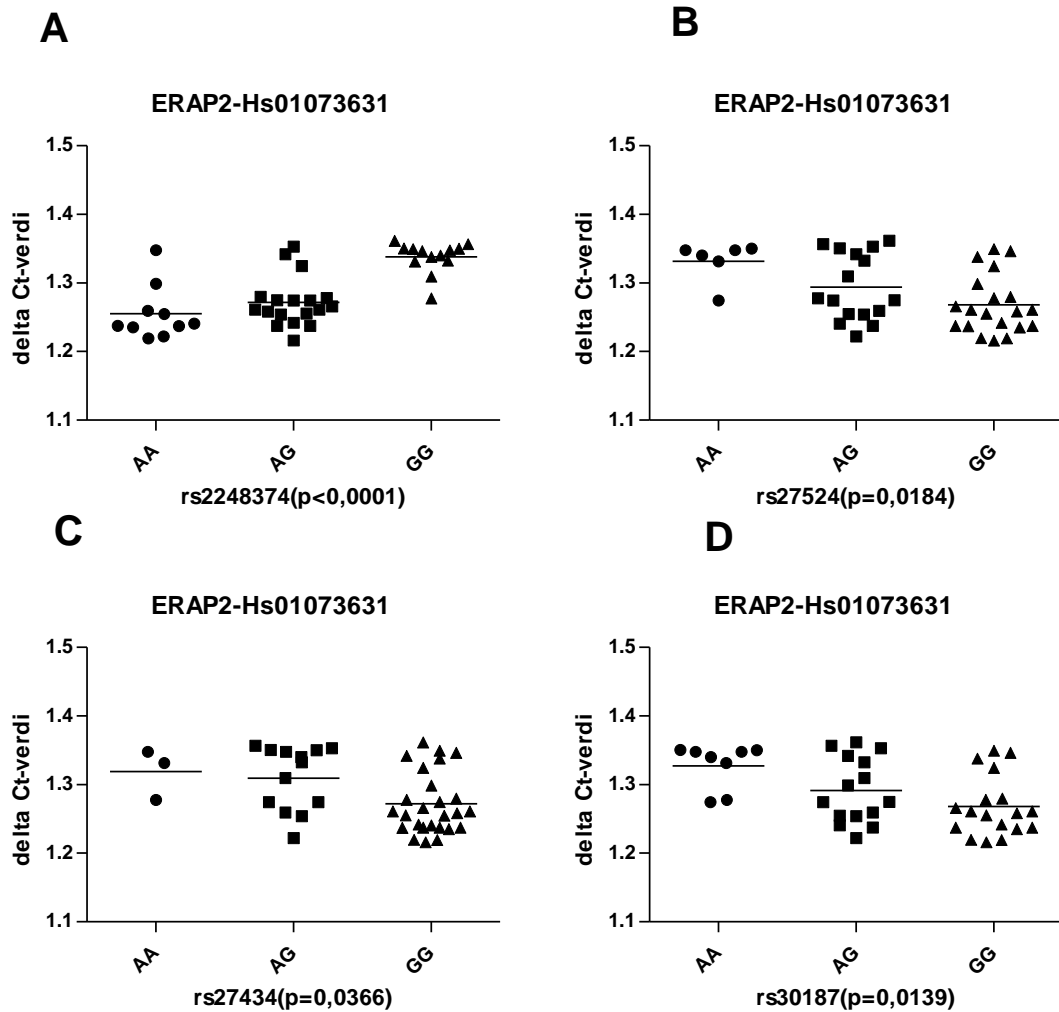
For å undersøke om assosiasjonen detektert med microarray var robust, ble TaqMan genekspressjonen og genotypen (ImmunoChip og TaqMan) også korrelert. Siden det har blitt gjort genekspressjonsanalyser på Microarray og TaqMan på SNPene i ERAP1/ERAP2-regionen, så skal også forandringer på signifikansnivået sees nærmere på for de to genekspressjonsmetodene.

ERAP1-assayet Hs00372304 (Figur 15) viste generelt mye mindre signifikante forskjeller i ekspressjonsnivå mellom de ulike genotypene enn det vi hadde sett ved microarray dataene, og det var kun rs30187 (Figur 15D), med en signifikans-grense $p=0,05$, som var signifikant assosiert ($p=0,0131$). Hs00372304 fanger kun opp det ene (NM_016442.3) av tre kjente transkripter for ERAP1-genet. For Microarray genekspressjonen var rs30187 (Figur 10, Figur 11) mest signifikant assosiert med ERAP1-probene (Figur 10A og 10B, Figur 11A og 11B), både før og etter retyping.



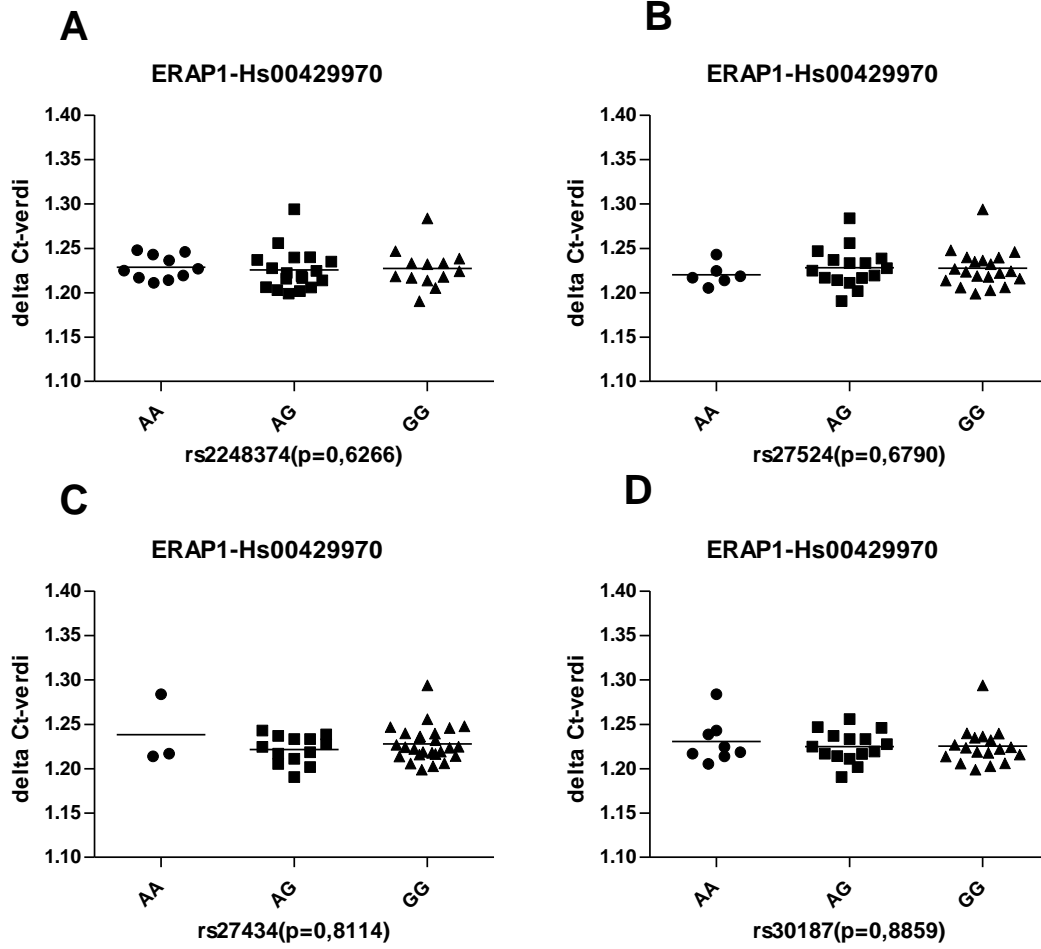
Figur 15: ERAP1-assay (Hs00372304) vs ERAP1 og ERAP2 SNPs

ERAP2-assayet Hs01073631 (Figur 16) var signifikant assosiert med alle SNPene i ERAP1/ERAP2-regionen (signifikans grense $p=0,05$). Hs01073631 var derimot mest assosiert med ERAP2 SNPen rs2248374 ($p<0,0001$)(Figur 16A), ERAP2-assayet fanger opp to av to kjente transkripter for ERAP2-genet (Tabell 10). På Microarray genespresjonen var rs2248374 mest assosiert med ERAP2-proben ILMN_1743145, med et tilsvarende signifikans-nivå på $p<0,0001$.



Figur 16: ERAP2-assay (Hs01073631) vs ERAP1 og ERAP2 SNPs

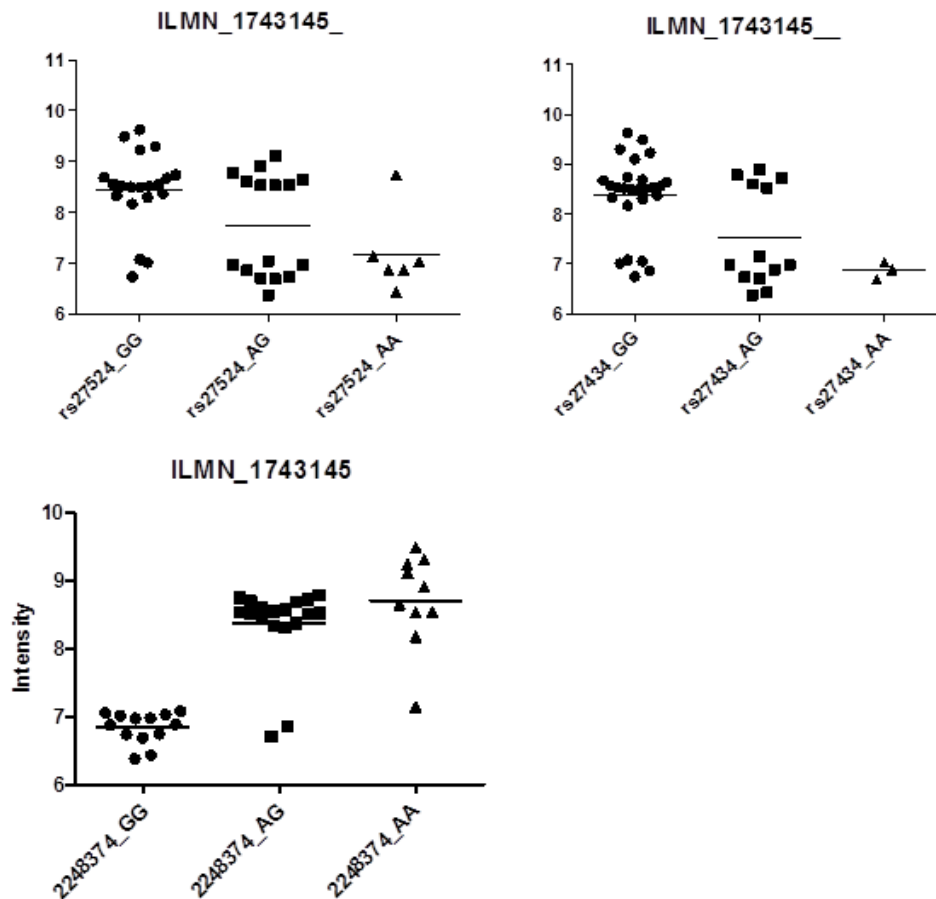
For ERAP1-assayet Hs00429970 (Figur 17) var det ingen av SNPene i ERAP1/ERAP2-regionen som viste signifikant assosiasjon, gitt en signifikant-grense $p < 0,05$. Assayet fanget opp tre av tre kjente transkripter for ERAP1-genet (Tabell 9), og gav gode Ct-resultater etter qPCR.



Figur 17: ERAP1-assay (Hs00429970) vs ERAP1 og ERAP2 SNPs

4.4.2 ERAP2 regionen

De lavt uttrykte tymus-prøvene som fordelte seg som AA og AG for rs27524 og rs27434 på ERAP2-proben (ILMN_1743145), grupperer seg i homozygot rs2248374*GG gruppen på den samme proben. Samtidig som at homozygote rs27524 og rs27434 *GG individene med høyt ERAP2 uttrykk grupperer seg i rs2248374*AA gruppen (Figur 18).



Figur 18: Uttrykk av ERAP2-proben ILMN_1743145.

4.4.3 SNPer i ERAP1/ERAP2-regionen

For å kunne si noe om den potensielle funksjonelle rollen av SNPene i ERAP1/ERAP2 regionen, ble plasseringen av SNPene kartlagt. Plassering av SNPene ble funnet ved å søke på hvert enkelt SNP i dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>). SNPene i ERAP1/ERAP2-regionen er plassert i intron eller ekson av ERAP-transkriptene (Tabell 10).

SNP polymorfismen for rs30187 er lokalisert i ekson 12 av ERAP1 og er en missense mutasjon som fører til endring av en aminosyre i proteinet fra lysin til arginin i posisjon 528. SNPen rs27524 er

lokalisert i intron av ERAP1 (NM_016442.3) og CAST (NM_001042440.2, NM_001190442.1, NM_173060.3). Det betyr at genene overlapper, og det er ikke rapportert aminosyre endring..

For rs27434 har det oppstått en stille mutasjon, dette er mutasjoner i kodende-regioner (ekson 7, nær intron 7), som ikke fører til forandring i aminosyresekvensen til proteinet (her for posisjon 356).

ERAP2 SNPen rs2248374 er lokalisert i intron 10 (posisjon 56) av ERAP2 transkriptet NM_001130140.1, svært nær det foregående eksonet (69, 82).

Tabell 10: Oversikt over ERAP1 og ERAP2 polymorfismer og plassering.

SNP	Gen	SNP posisjon/(Posisjon i ekson/intro)	SNP lokalisasjon	Aminosyre forandring	Kommentar
1	ERAP1	96124330/(1930 ⁶ ,1688 ⁷)	Ekson 12	Lys --> Arg	Mis-sense mutasjon, Transisjon substitusjon
2	ERAP1, CAST	96101944/(52905 ⁸)	Intron 30	IR	Transisjon substitusjon
3	ERAP1	96129512/(1415 ⁶ , 1173 ⁷)	Ekson 7	Ala --> Ala	Stille mutasjon
4	ERAP2	96235896/(IR ⁹)	Intron 10	IR	Transisjon substitusjon

⁶ ERAP1 transkript 1: NM_001040458.1

⁷ ERAP1 transkript 2: NM_001198541.1

⁸ ERAP1 transkript 3: NM_001042440.2

⁹ ERAP2 transkript: NM_001130140.1

4.5 LD mellom SNPene i ERAP1/ERAP2 regionen

LD (D' og r^2) ble brukt for å måle graden av koblingsulikevekt mellom SNPene i ERAP1/ERAP2 gen regionen som viste konsistente funn. Dette ble gjort for å se i hvor stor grad genotypene på de ulike SNPene korrelerer. Graden av LD mellom SNPene blir spesifisert både med D' og r^2 (83). En verdi på 0 indikerer at de to SNPene er i fullstendig likevekt, og en verdi på 1 representerer høyest grad av koblingsulikevekt som kan forekomme for de to målene (83).

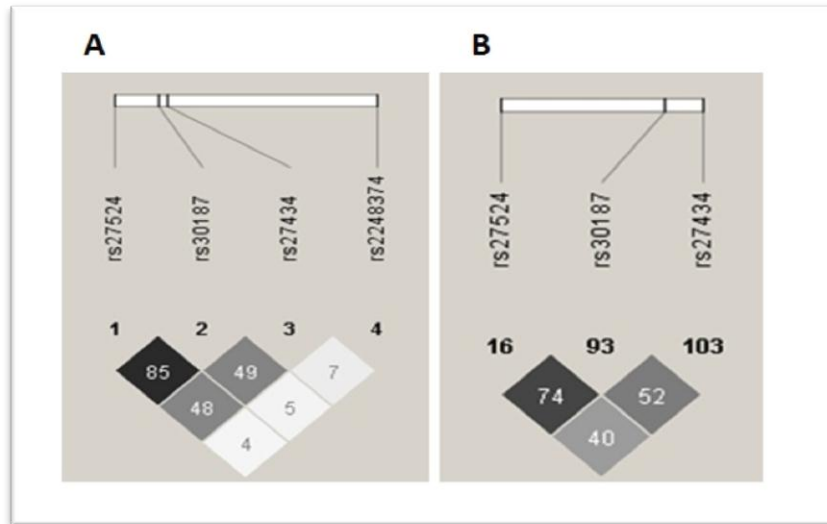
LD målingene viste en korrelasjon mellom tre av fire SNPene (Tabell 11), noe som indikerer at SNPene ikke er nedarvet uavhengig av hverandre i populasjonen. For rs27524 og rs30187 er det sterkest grad av LD, $r^2=0,85$ (Figur 19). rs27524 og rs2248374 viser nærmest ingen korrelasjon ($r^2=0,046$), rs2248374 er nesten ikke korrelert med noen av de øvrige tre SNPene (Figur 19A).

Tabell 11: LD mellom ERAP1 og ERAP2 SNPs

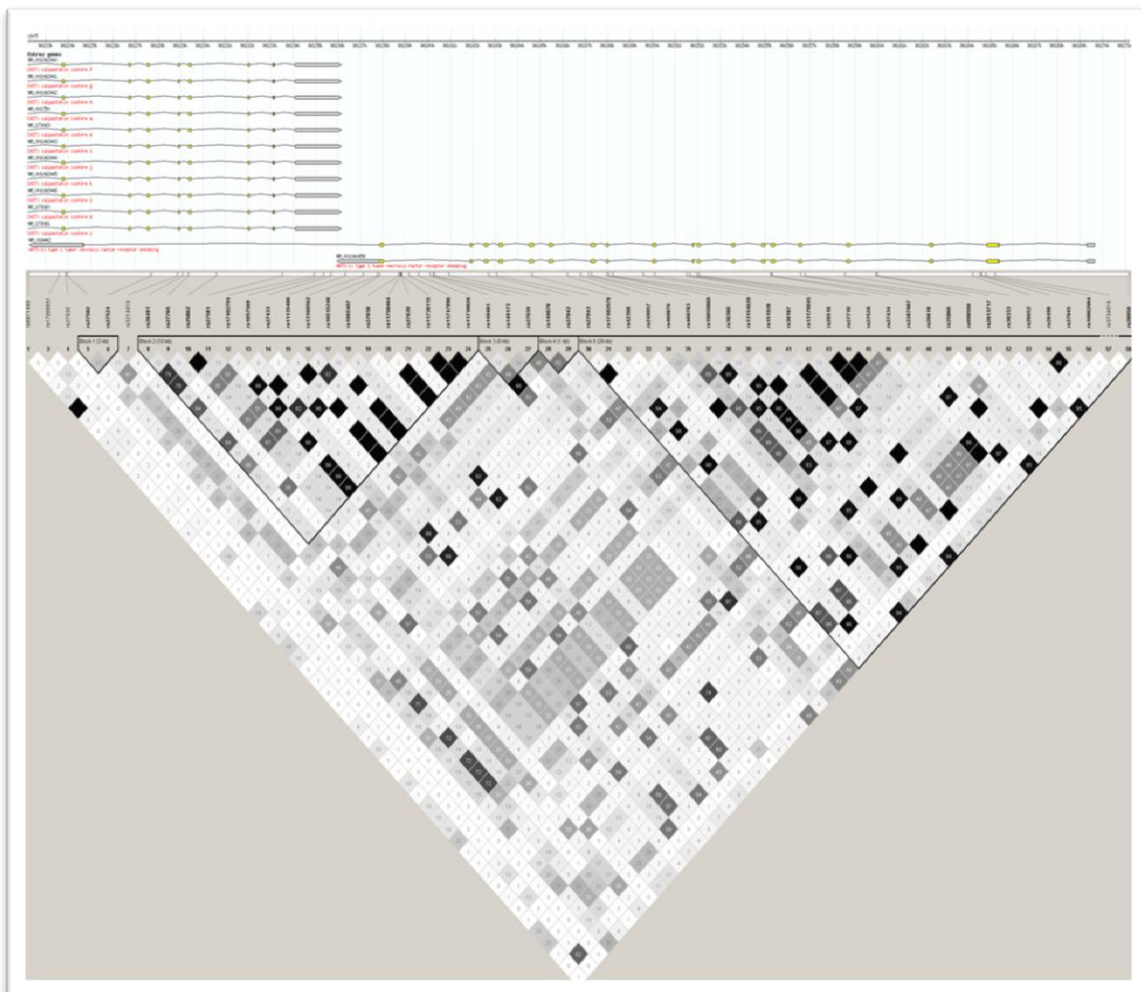
	Koblingsulikevekt mellom SNPs	r^2	D'
1	rs27524 og rs30187	0,85	1,0
2	rs27524 og rs27434	0,49	0,91
3	rs27524 og rs2248374	0,046	0,34
4	rs27434 og rs30187	0,50	1,0
5	Rs27434 og rs2248374	0,08	0,56
6	rs30187 og rs2248374	0,05	0,33

Figur 19 viser en grafisk fremstilling av LD (målt som r^2) mellom SNPene i genregionen (ERAP1/ERAP2). I figur 19B er ImmunoChip genotypedata for våre 42 tymusprøver brukt som grunnlag for LD beregningene, mens i 19A er det brukt data fra det internasjonale HapMap konsortiet (HapMap Data Rel 28 phase II + III, Aug. 10, on NCBI B36 assemble, dbSNP b126). SNPen rs2248374 var ikke genotypet i HapMap (Figur 19B). Graden av LD er lik mellom disse to datasettene. I figur 19A ser man klart at det er sterk grad av LD mellom SNPene i ERAP1 genet, og dermed stor grad av korrelasjon mellom genotypene. Mens disse ikke viser LD med ERAP2 SNPen (rs2248374).

Et LD kart over alle SNPene i regionen som er genotypet i HapMap prøvene (HapMap Data Rel 28 phase II + III, Aug. 10, on NCBI B36 assemble, dbSNP b126) viste at regionen som helhet ikke viser stor grad av LD.



Figur 19: Oversikt over LD mellom SNPene i ERAP1/ERAP2-regionen (R square). A: Tymus, B: HapMap.



Figur 20: LD mellom alle SNPer genotypet i HapMap prøver i hele regionen.

4.6 Korrelasjon mellom ERAP1/ERAP2 genotyper og HLA klasse I genotyper

Korrelasjon mellom ERAP genotyper og HLA klasse I genotyper ble utført, siden det i et tidligere studie (2011) ble vist at polymorfismer i ERAP1 påvirker AS risiko bare hos HLA-B27 positive individer(42).

For å undersøke korrelasjon mellom SNP genotype og HLA klasse I genotyper (HLA-A, -B og -C loci), ble det gjort en overordnet korrelasjon mellom allelene (Cramer's V), siden antall mulige kombinasjoner mellom HLA klasse I er stor, da disse lociene er så polymorfe. Det var ingen signifikant sammenheng mellom SNP genotype og HLA klasse I genotype (Tabell 12), da det var ingen p-verdier lavere enn $p=0,55$.

Tabell 12: LD mellom SNP genotype og HLA klasse I.

Cramer's V			
SNP	HLA-A	HLA-B	HLA-C
rs27524	0,41	0,5	0,52
rs30187	0,286	0,536	0,548
rs27434	0,411	0,538	0,451
rs2248374	0,452	0,508	0,437

4.7 Ingenuity Pathway Analysis

Pathway analysen ble utført på genene fra de 100 mest signifikante SNP-probe korrelasjonene ($p < 1 \times 10^{-3}$), for å se om genene som er differentielt uttrykt relativt til AIS genotype klustrer i visse biologiske veier. For å få et større antall gener som basis for pathway analysen, ble genene knyttet til de 100 SNP-probe parene med lavest p-verdi. Dette ga resultater med en p-verdigrense på $p < 1 \times 10^{-3}$. Genene i ERAP-regionen var også blant disse.

De mest signifikante pathway analyse resultatene er oppgitt i tabell 13, fordelt på de biologiske funksjoner (sykdommer og lidelser, molekylære og cellulære funksjoner), canonical signalveier og oppstrøms regulatorer. Canonical signalveier er vedtatte signalveier IPA har tatt med i sin database, og som i mest mulig grad matchet genlisten som ble generert på basis av de 100 mest signifikante eQTL. Antall molekyler sier noe om hvor mange gener i genlisten, som IPA har koblet til de faktiske

biologiske funksjonene. Canonical signalvei resultatene som er oppgitt i ratio, sier noe om antall molekyler av interesse fra datasettet som er representert i ulike signalveier relatert til det totale antallet molekyler som inngår i signalveien. Signalvei analysen gir en oversikt over de biologiske funksjonene (sykdom og lidelser, molekylære og cellulære funksjoner) i IKB (Ingenuity Knowledge Base), som genene i datasettet er signifikant assosiert med (Tabell 13). Canonical signalveier viser signalveier som gir mest signifikant assosiasjon med datasettet, og denne assosiasjonen er målt ved å se på rasion (forklart over), og ved at en beregnet p-verdi bestemmer sannsynligheten for at sammenhengen mellom genene i datasettet og canonical signalveien er ved en tilfeldighet(84).

Resultatene ble oppgitt i p-verdi, ratio og falsk funnrate (FDR- False Discovery Rate). Den oppgitte p-verdien ble kalkulert ved Fishers eksakte test og estimerer om sammenhengen mellom en gitt biologisk funksjon i IKB og en gruppe molekyler i datasettet er tilfeldig. Statistisk signifikans grense er satt til $p=0,05$.

Interessant nok kom arvelige og inflammatoriske sykdommer ut som signifikante av pathway analysen. I tillegg til skjelett og muskelsykdommer, bindevevssykdommer og neurologiske sykdommer. Dette stemmer godt overens med de AIS som vi brukte i SNP seleksjonen før analysen. Likeledes var inflammatorisk respons en av kategoriene som dukket opp som signifikante under "molekylære og cellulære funksjoner". Under "canonical signalveier" ble immunologisk interessante prosesser som cytotoksisk T lymfocyt-mediated apoptose av målceller og antigen-presentasjon pekt ut.

Analyseresultatene viser at ERAP1 og ERAP2 er signifikant assosierte med de topp fem sykdommene og lidelsene, og molekylære og cellulære funksjonene (Tabell 13). Dette er funn som gir en bedre illustrasjon over de signalveiene og molekylære funksjonene som blir påvirket av en SNP mutasjon i ERAP1/ERAP2-regionen.

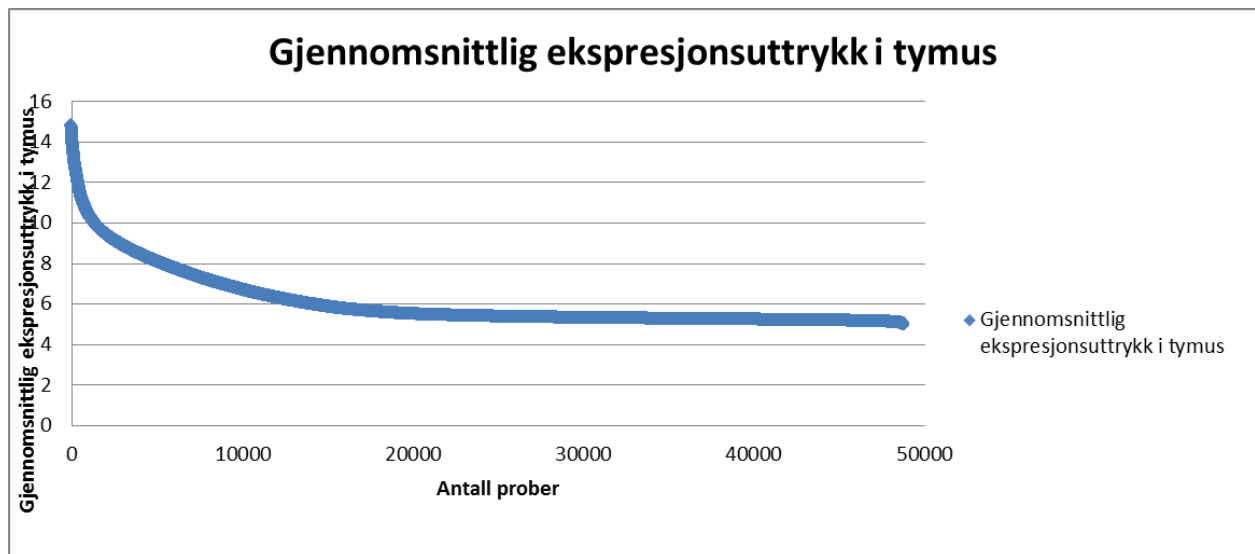
Tabell 13: Signalvei analysen for genene fra de 100 mest signifikante AIS SNP – genespresjonsprobe parene ($p < 1 \times 10^{-3}$).

Sykdom og lidelser	B-H p-verdi	#Molekyler (Inkludert)
Arvelig sykdom	3,98E-05-1,29E-01	14 (ERAP2)
Skjelett og muskelsykdommer	3,98E-05-1,19E-01	22 (ERAP1)
Bindevevssykdommer	1,31E-03-1,19E-01	17 (ERAP1)

Inflammatorisk sykdom	1,31E-03-8,59E-02	15 (ERAP1)
Nevrologisk sykdom	5,41E-03-1,11E-01	18 (ERAP1 og ERAP2)
Molekylære og cellulære funksjoner	p-verdi	#Molekyler (Inkludert)
Protein degradering	6,52E-04 - 1,96E-02	7 (ERAP1)
Protein syntese	6,52E-04 - 2,57E-02	9 (ERAP1)
Molekylær biokjemi	1,03E-04 - 4,96E-02	13 (ERAP1)
Molekylær transport	1,03E-04 - 4,96E-02	7 (ERAP1)
Inflammatorisk respons	5,53E-02-1,29E-01	12 (ERAP1)
Canonical signalvei	p-verdi	Ratio
Allograft Rejection Signaling	2,46E-09	6/95 (0,063)
Cytotoksisk T lymfocyt-mediatede apoptose av målceller	4,54E-09	6/85 (0,071)
OX40 signalveien	9,04E-09	6/94 (0,064)
Antigen presenterende signalveier	9,95E-08	5/40 (0,125)
Cdc42 Signalering	2,62E-06	6/178 (0,034)

4.8 Genekspresjonen i ERAP1/ERAP2-regionen i forhold til genuttrykket i tymus

Siden ERAP1/ERAP2-regionen har gitt konsistente funn, ønsket vi å se nærmere på genekspresjonen i denne regionen i forhold til det generelle genekspresjonsuttrykket i tymus. For å kunne illustrere dette ble gjennomsnittlig intensitet for alle tymusprøvene (N=42) for hver probe (N=48804 prober) plottet fra høyest til lavest (Figur 21).



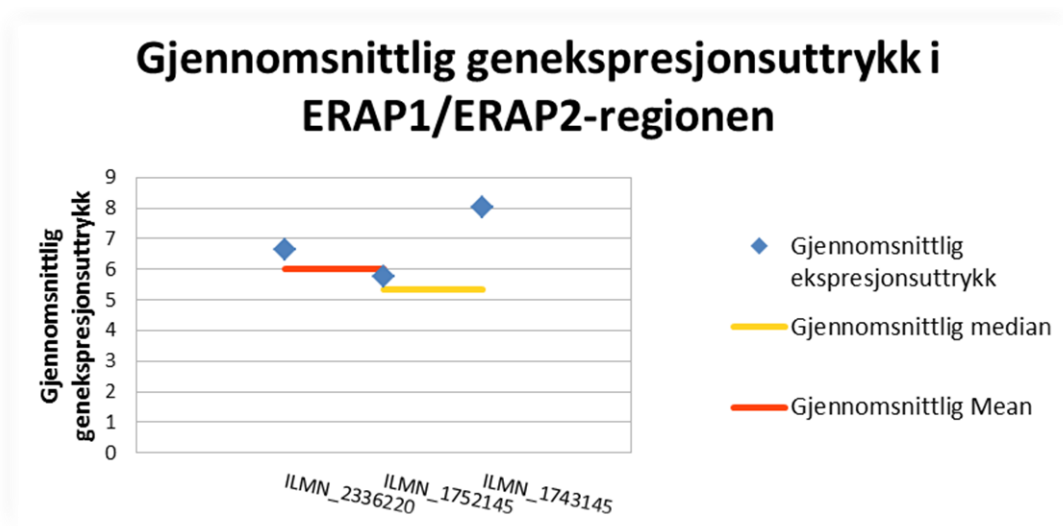
Figur 21: Oversikt over genuttrykket i tymus angitt som probeintensitet fra høyest til lavest for hver probe. Y-akse: ekspresjonsuttrykk, X-akse: antall prober.

Probene som var signifikant assosiert med SNPene i ERAP1/ERAP2-regionen (signifikans grense $p < 5 \times 10^{-5}$), var forskjellig uttrykt i tymus, men var alle blant de 10 000 høyest uttrykte probene (Tabell 14). Det er høyest ekspresjonsintensitet for proben som fanger opp transkripter fra ERAP2-genet (ILMN_1743145).

Tabell 14: Genekspresjonsuttrykk i tymus for assosierte prober i ERAP1/ERAP2-regionen.

Gen	SNP	Probe	Probe posisjon/(SNP posisjon)	Gjennomsnittlig Ekspresjonsuttrykk
ERAP1	Rs30187/rs27524	ILMN_2336220	96144589- 96144638/(96124330/96127700)	6,65
ERAP1	Rs30187/rs27434	ILMN_1752145	96136694-96136703:96137864- 96137903/(96124330/96155268)	5,75
ERAP2	Rs2248374/rs27434/rs27524	ILMN_1743145	96274820-96274869/ (96235896/96155268/96127700)	8,00

For å kunne illustrere ekspresjonsintensiteten til probene for ERAP1/ERAP2-regionen, i forhold til probene for øvrige gener i tymus, er disse sammenliknet med gjennomsnittlig og median av intensitet av alle probene på Microarray ekspresjons chipen (Figur 22). Det er kun ERAP1 proben, ILMN_1752145, som har et lavere ekspresjonsuttrykk, enn det gjennomsnittlige tymus ekspresjonsuttrykket. De to andre probene ERAP1 (ILMN_2336220) og ERAP2 (ILMN_1743145), har et høyere uttrykk enn både gjennomsnittlig median og gjennomsnittlig mean-verdi.



Figur 22: Gjennomsnittlig genekspresjonsuttrykk målt ved probeintensitet i ERAP1/ERAP2-regionen.

4.9 Blod korrelasjon - mRNA by SNP browser

Det kan være ulike gener som er uttrykt i tymus og som viser korrelasjon med AIS risikovarianter, enn det er i blod. For å se etter forskjeller og likheter ble det gjort en undersøkelse av genekspressionsforskjeller i ERAP1/ERAP2-regionen i blod og dette ble sammenliknet med funnene fra tymus. Hele den 2 Mb regionen ble undersøkt for å se om eventuelt andre gener påvirkes av AIS SNPene i blod enn i tymus. Det var mange av de samme genene som viste signifikant forskjellig uttrykk avhengig av AIS SNP genotype i både blod og tymus (Tabell 15). For rs30187 var både ERAP1 (ARTS-1 $p=0,00015$) og ERAP2 (LRAP $p=2,5 \times 10^{-27}$) signifikant assosierte i blod og tymus. Rs27524 ($p=1,1 \times 10^{-27}$) og rs27434 ($p=0,00012$) viser kun assosiasjon med ERAP1-genet i blod, men er derimot assosiert med både ERAP1- og ERAP2 i tymus. ERAP2 SNPen rs2248374 fantes ikke i databasen (mRNA by SNP browser) som ble brukt til dette formålet.

Tabell 15: Genuttrykk i blod vs genuttrykk i tymus

SNP	Gen i blod	p-verdi (blod)	Kromosom	Gen i tymus
Rs30187	ARTS-1 (ERAP1)	0,00015	5	ERAP1, RIOK2, ERAP2, LNPEP
	LRAP (ERAP2)	$2,5 \times 10^{-27}$	5	GLRX, CAST, ERAP1
rs27524	ARTS-1	$1,1 \times 10^{-27}$	5	CAST, ERAP1, LNPEP, ERAP2, RIOK2
rr27434	ARTS-1	0,00012	5	CAST, ERAP1, LNPEP, ERAP2

5 Diskusjon

5.1 eQTL assosiasjoner med AIS

Ved å utføre korrelasjon mellom genekspressionsdata (mål ved prober på microarray) og genotypedata (bestemt ved Immunochip array genotyping) for en liste med selekterte autoimmune risiko SNPs (totalt 256 SNPs), ble det oppdaget flere autoimmune SNPs som signifikant påvirket ekspresjon av omkringliggende gener i tymus.

De 10 eQTLene som viste signifikante forskjeller ($p < 5 \times 10^{-6}$) i ekspresjonsnivå mellom ulike genotyper, er assosierte med seks ulike AIS (Tabell 16). Blant SNPene som ble testet fantes det risikovarianter for ni autoimmune sykdommer. Ingen av risikovariantene for RA, SLE eller MS kom opp med korrelert genekspressjon blant de 10 mest signifikante funnene. Dette til tross for at pathway analysene der gener fra de 100 mest signifikante funnene ble inkludert, dukket både leddsykdommer, bindevevssykdommer og nevrologiske sykdommer opp på listen. Det er meget mulig at risiko SNPene for disse tre sykdommene befinner seg noe lenger ned på listen (under de 10 sterkeste funnene), og at dette er grunnen til at disse sykdomskategoriene dukket opp i pathway analysen. Vi kan ikke anta at tymus eQTL knyttet til risiko SNPene er begrenset til de seks AIS i tabell 16. Det er også fremdeles mye uavklart når det gjelder genetisk overlapp mellom ulike AIS og Internasjonale data fra store case-control assosiasjonsstudier, der man tester om Immunochip genotypedata i en rekke AIS vil gi oss et bedre svar på hvilke AIS som faktisk er assosiert med de eQTL vi har avdekket i denne studien.

SNPene har blitt avdekket som eQTL i tymus er også lokalisert på ulike kromosomer, utenom SNPene i ERAP1/ERAP2-regionen på kromosom 5. T1D er en AIS som er assosiert med fire av ti SNPene i tabellen (rs7804356, rs2292239, rs4788084, rs2281808), alle fire SNPene er lokalisert på fire forskjellige regioner. SNPene i ERAP1/ERAP2-regionen er assosiert med AS (rs39187, rs27434) og Ps (rs27524), i tillegg er SNPen rs1008953 som ikke befinner seg i ERAP1/ERAP2-regionen, også assosiert med Ps. De øvrige eQTLene er assosierte med CeD, UC og CD (rs13003464) og UC (rs1297265). Dette viser at det er spredning av AISs risiko SNPs, som viste signifikante avvik ($p < 5 \times 10^{-5}$) i ekspresjonsnivå mellom ulike genotyper. Dette er forventet da hypotesen om eQTL i tymus knyttet til autoimmune sykdommer ikke var antatt å være sykdomsspesifikk, men knyttet til autoimmunitet generelt.

Tabell 16: Oversikt over AIS 10 eQTLNene er assosiert med ni utvalgte AIS.

	SNP	Autoimmune Sykdommer	Gen	Region
1	Rs30187	AS	ERAP1	5q15
2	rs13003464	CeD, CD, UC	PUS10	2p16.1
3	rs27524	Ps	CAST, ERAP1	5q15
4	rs27434	AS	ERAP1	5q15
5	rs7804356	T1D	SKAP2	7p15.2
6	rs2292239	T1D	ERBB3	12q13.2
7	rs4788084	T1D	IL27, NUPR1	16p11.2
8	rs2281808	T1D	SIRPB1	20p13
9	rs1008953	Ps	SDC4, SYS1	20q13.12
10	rs1297265	UC	IR ¹⁰	21q21

Pathway analyse resultatene bekrefter også dette, da disse viste en signifikant assosiasjon mellom de 100 mest signifikante SNP-probe korrelasjonene, og sykdommer og lidelser som kan føre til AIS (Tabell 13). Det var særlig interessant at inflammasjon og antigenpresentasjon dukket opp som biologiske veier genene var involvert i.

5.2 Vurdering av signifikansnivå

For de 13 risiko SNPene ble signifikansnivået satt til $p < 5 \times 10^{-5}$ innenfor et område på +/- 100 Kb rundt SNPen, og $p < 5 \times 10^{-6}$ innenfor et område på +/- 1 Mb rundt SNPen. Bonferroni korreksjonen representerer antallet eQTL analyse sammenligninger rundt disse områdene. Det ble brukt to p-verdier så det ville være ulik sjans for å finne eQTL rundt disse to områdene. En tidligere studie har vist at cis-regulerende sykdoms assosierte SNPer kan ligge innenfor et +/- 100 Kb vindu, og på bakgrunn av det studiet ble det samme vinduet brukt for de 13 risiko SNPene (52). Dette var overens med våre funn, hvor 8 av 13 eQTL befant seg nærmere enn 100 kb fra genet som det påvirket transkripsjonen av, , mens 5 befant seg >100 kb og <1 Mb unna, hvorav 2 hadde p-verdi lavere enn signifikansnivået satt for dette ytre vinduet.

Selv om de fleste SNPer som påvirker ekspresjon befinner seg nær genet, er det kjent at SNPer kan påvirke uttrykk av gener som befinner seg langt unna og til og med på andre kromosomer. Siden en

¹⁰ IR: Ikke registrert

undersøkelse hvor hver SNP ble korrelert mot alle transkripter i hele genomet ville medføre et svært høyt antall sammenlikninger og dermed en rekke falske positive (eller en så lavt signifikansgrense at vi ikke hadde hatt styrke til å detektere forskjeller med vårt material på 42 tymuser), valgte vi å begrense det utvidete søket til 2 Mb rundt SNPen. Denne tilnærmingen medførte at det ble brukt to ulike p-verdigrensener, en for prober som fanget opp transkripter innenfor et 200 kb område, og en mer stringent for prober innenfor 2 Mb område. De fem av de 13 SNPene(rs45450798, rs10781500, rs7608910, rs13003464, rs4788084) som nådde den initielle grensen beregnet på 200 kb vinduet ($p < 5 \times 10^{-5}$) lå faktisk lenger unna, og befant seg i 2 Mb vinduet men bare to (rs13003464, rs4788084) av dem nådde signifikansnivået basert på korrigerings av antall sammenlikninger innenfor dette vinduet og de tre andre ble derfor ekskludert fra videre undersøkelser.

Hvis det hadde blitt valgt et enda større vindu enn 2 Mb, så kunne det bli oppdaget flere autoimmune risiko SNPs som påvirker genekspresjon i tymus, men for å ha styrke til å detektere signifikante forskjeller gitt de da enda mer stringente p-verdigrensene måtte vi ha økt materialstørrelsen. Det er dermed svært sannsynlig at vi har mistet noen sanne assosiasjoner ved å begrense korrelasjonen med de nærmeste transkriptene, samt å operere med relativt strenge korreksjonsfaktorer. Men på den annen side, er det grunn til å tro at disse kravene har gjort at de 10 eQTLene vi har funnet i tymus knyttet til risiko for autoimmunitet er reelle.

5.2.1 Genotype suksessraten

Alle SNPene hadde en genotype suksessrate over 90 % (Tabell 7), utenom rs30187, som etter retypering fikk en GSR på 97,6 % (Tabell 7). Ved en høy GSR er det sikret at ingen spesifikke genotyper er mistet under ImmunoChip - og TaqMan genotyping, slik at det kunne ha oppstått en skjevhet i materialet. Høy genotype suksessrate betyr også at DNA kvaliteten på prøvene var god.

5.3 Replikasjon og kvalitetssikring av eQTL funn i ERAP1 og ERAP2 regionen

Det ble utført korrelasjon mellom TaqMan genekspresjonsassay og ImmunoChip genotypene for alle SNPene utenom ERAP2-SNPen 2248374 som kun ble genotypet ved bruk av TaqMan assay. Alle SNPene ga signifikante resultater for noen prober, utenom rs30187 som måtte retypes på TaqMan genotype.

For ERAP1 SNPen rs30187 var det to prober som viste signifikant assosiasjon med ImmunoChip og Taqman genotyping, ILMN_2336220 (ImmunoChip $p=0,0002$, Taqman $p<0,0001$) og ILMN_1752145.

Etter retyping ble CAST-proben signifikant assosiert, ILMN_2322806 (ImmunoChip $p=0,0682$, TaqMan $p=0,0268$). rs30187 viste signifikant assosiasjon med det ene ERAP1-assayet Hs00372304 ($p=0,0131$) og ERAP2-assayet Hs01073631 ($p=0,0139$).

rs27524 viste signifikant assosiasjon med to av seks prober, ERAP1-proben ILMN_2336220 ($p<0,0001$) og LNPEP-proben ILMN_1798433 ($p=0,0263$), og viste kun assosiasjon med ERAP2-assayet Hs01073631 ($p=0,0184$). For rs27434 viste alle probene, utenom en LNPEP-proben ILMN_1798433 ($p=0,0760$), signifikant assosiasjon. ERAP2-SNPen rs2248374 var signifikant assosiert med kun ERAP1-proben (ILMN_2336220 $p=0,0113$), og ERAP2-proben (ILMN_1743145 $p<0,001$).

Ved korrelasjon mellom TaqMan genekspressjonsassay og ImmunoChip/TaqMan (rs30187, rs2248374) genotyping viste alle ERAP-SNPene assosiasjon med ERAP2-assayet Hs01073631.

5.3.1 Microarray genekspressjon vs TaqMan genekspressjon

Microarray genekspressjon ble utført på total RNA, og bestod av 27455 velkarakteriserte transkripter, hvor hver transkript kodet for et gen, totalt bestående av 48804 prober hvor hver probe kan være komplementær med flere gensekvenser. Det var fem TaqMan genekspressjonsanalyser som ble utført på cDNA. Assayene dekket ERAP1/ERAP2-genregionen, med to ERAP1-assay og tre ERAP2-assay.

Videre undersøkelser i blod viste veldig sterk korrelasjon mellom rs30187 og ERAP2 ($p=2,5 \times 10^{-27}$), og rs27524 viste sterk korrelasjon med ERAP1 ($p=1,1 \times 10^{-27}$). rs27434 viste samme assosiasjonen til ERAP1/ERAP2-regionen ($p=1,2 \times 10^{-4}$).

SNPene i ERAP1/ERAP2-regionen ga konsistente funn, i både Microarray genekspressjon - og TaqMan genekspressjonsuttrykk, utenom ERAP2-SNPen (rs2248374) som kun viste signifikant assosiasjon med ERAP2-assayet (Hs01073631 $p<0,0001$), og var mest assosiert med ERAP2-proben (ILMN_1743145 $p<0,0001$). TaqMan assayene for både ERAP1 og ERAP2-genene fanget opp ulike transkripter (Tabell 9), det ene ERAP1-assayet (Hs00429970_m1) fanget opp alle tre transkripter og det ene ERAP2-assayet (Hs01073631_m1) fanget opp begge transkripter for dette genet. ERAP1-genet er assosiert med tre transkripter (NM_001040458.1, NM_001198541.1, NM_016442.3), men dataene for plasseringen av (Tabell 10) ERAP1-SNPen rs27524, viste lokalisering i ett transkript (NM_001042440.2) som ikke blir fanget opp i TaqMan genekspressjonsassayene.

Det var de samme probene i Microarrayen som var assosiert med ERAP1-SNPene (rs30187, rs27524 og rs27434), og det var kun en probe (ILMN_1743145) som var assosiert med ERAP2-genet og ERAP2-SNPen (rs2248374).

5.4 eQTL og regulerende sekvenser

eQTL analyser av alle 10 autoimmune risiko SNPs ble utført for å studere forholdet mellom genomet og transkriptom(85).

Forskjellen mellom eQTL kartlegging og tradisjonell QTL mapping, er at QTL studier fokuserer på ett eller noen få egenskaper (traits) mens de fleste eQTL studier vil analysere flere tusen ekspresjons egenskaper. eQTL analyser identifiserer genekspresjons regulatorer som er enten cis-regulerende faktorer som påvirker proksimale gener, eller trans-regulerende faktorer som fungerer gjennom en rekke mekanismer for å påvirke transkripsjonen av gener på andre kromosomer.

Det er nylig (2010) blitt vist at SNPs som er assosiert med en fenotypisk egenskap, mest sannsynlig er eQTLs. Ved å forstå bidraget av genetiske variasjoner på uttrykket av gener, vil det ha en stor påvirkning på den måten legemidler blir individuelt personlig og hvordan komplekse sykdommer blir undersøkt(86).

Enkelte forskere har foreslått at introner er viktige i uttrykk av gener, ved å fungere som cis-regulerende elementer. Denne funksjonen er derimot ingen generell funksjon i introner, da flere gener som mangler introner uttrykker seg normalt. Introner kan også være viktige i prosessen av alternativ spleising, som fører til produksjonen av flere typer mRNA fra et enkelt gen. Forskjellene i ekspresjonsnivå som vi har oppdaget i denne studien, kan skyldes at bare et av de alternative transkriptene for genet viser ekspresjonsnivåer som påvirkes av hvilket allel som er tilstede på AIS risiko SNPen. Dessverre fikk vi ikke resultater for alle de ulike TaqMan genekspresjonsassayene som skulle hjelpe oss å belyse forskjeller i uttrykk av ulike ERAP1 og ERAP2 transkripter. Videre studier er nødvendig for å avdekke dette nærmere.

Cis-regulerende sekvenser ligger inntil transkripsjonsenheten de regulerer, 1000 basepar fra transkripsjonsstart. Trans-virkende faktorer er derimot transkripsjonsfaktorer som bindes til de cis-virkende sekvensene. Transkripsjonsfaktorene kalles aktivator eller repressor, avhengig av om faktorene aktiverer eller hemmer transkripsjonen. Cis-virkende faktorer kan føre til både positiv (økere, enhancere) og negativ regulering av promotor, det er også mulig å kontrollere promotor via kontroll elementer som ligger flere tusen basepar nedstrøm/oppstrøms fra transkripsjonsstart(87).

De 10 SNPene som ble funnet å ha eQTL egenskaper i denne studien var lokalisert i intron, utenom ERAP1-SNPene rs30187 og rs27434 som er lokalisert i ekson av ERAP1-genet (Tabell 17).

En enkel basepolymorfisme i intron kan påvirke translasjonen av mRNA, og kan fungere som en regulerende variant for et gen som kan være lokalisert i intron av et annet gen. Dette kan gjelde SNPen rs2292239 som er lokalisert i intron av ERBB3, men er assosiert med reguleringen av RPS26.

Den intergene SNP'en rs4788084 er derimot ikke lokalisert i noe gen, ligger mellom NUPR1 og IL27, men kan regulere genekspressjonen av LOC728734. Det kan også være at de genotypede risiko SNPene ikke selv er de SNPene som biologisk påvirker genekspressionsnivået, men at disse er i LD med de regulatoriske SNPene. Graden av LD varierer mye fra genomisk region til region, og teoretisk sett kan den kausale SNP'en ligge langt unna markør risiko SNPene brukt her. I noen studier er risiko locuset bedre mappet, som for eksempel for ERAP1, slik at det er større sannsynlighet for at SNP'en testet faktisk er den sykdomsinvolverte enn i andre studier.

En enkelt basepolymorfisme i ekson som ikke forandrer aminosyre kan påvirke transkripsjonsnivå, for eksempel kan dette gjelde rs27434.

For å kunne forstå det genetiske uttrykket av et gen og eventuelt utfall ved en SNP polymorfisme, er det viktig å få en oversikt over blant annet SNP plasseringen i genet. Intron polymorfismer har en annen effekt enn ekson polymorfismer. De fleste SNPene som er eQTL er lokalisert i det samme genet, som det er vist sykdomsassociering til (Tabell 17).

Tabell 17: Oversikt over 10 eQTL lokalisering.

	SNP	Region	Alleler	Kb fra SNP til probe	SNP Plassering (Gen)	Sykdomsassociering (Gen ¹¹)
1	Rs30187	5q15	C/T	Samme gen	Ekson (ERAP1)	AS (ERAP1)
2	rs13003464	2p16.1	A/G	203 Kb ¹²	Intron (PUS10)	CeD (LOC33980)
3	rs27524	5q15	A/G	16,9 Kb	Intron (ERAP1)	Ps (ERAP1)
4	rs27434	5q15	A/G	Sammen gen	Ekson (ERAP1)	AS (ERAP1)
5	rs7804356	7p15.2	C/T	Samme gen	Intron (SKAP2)	T1D (SKAP2)
6	rs2292239	12q13.2	G/T	45,9 Kb	Intron (ERBB3)	T1D (RPS26)

¹¹ Microarray gen assosiasjon for autoimmune risiko SNPs.

¹² rs13003464 er signifikant assosiert innefor +/- 1Mb.

7	rs4788084	16p11.2	T/C	120 Kb	Intergen (mellom NUPR1 og IL27)	T1D (LOC728734)
8	rs2281808	20p13	C/T	Samme gen	Intron (SIRPG)	T1D (SIRPG)
9	rs1008953	20q13.12	C/T	16,8 Kb	Intergen (mellom SDC4 og SYS1)	Ps (SYS1)
10	rs1297265	21q21	A/G	73,9 Kb	Intergen (mellom CYCSP42 og NPIP1)	UC (LOC38881)

5.4.1 ERAP2 Spleisevarianter

Ved korrelasjon mellom Microarray genekspressjon og genotypedata (ImmunoChip og TaqMan), viste ERAP2-proben (ILMN_1743145) et høyere ekspresjonsuttrykk enn de andre probene som viste assosiasjon med SNPene i ERAP1/ERAP2-regionen. ERAP2-proben ILMN_1743145 viste også en fordeling innad i genotypene (AA, AG, GG) på korrelasjonsplottene, for alle fire SNPene (rs30187, rs27524, rs27434 og rs2248374).

Det er en forskjell mellom TaqMan ekspresjons data og Microarray ekspresjonsdata, da Microarray proben ikke fanger opp alle transkriptene som er assosiert med ERAP1 og ERAP2 genene (Tabell 18).

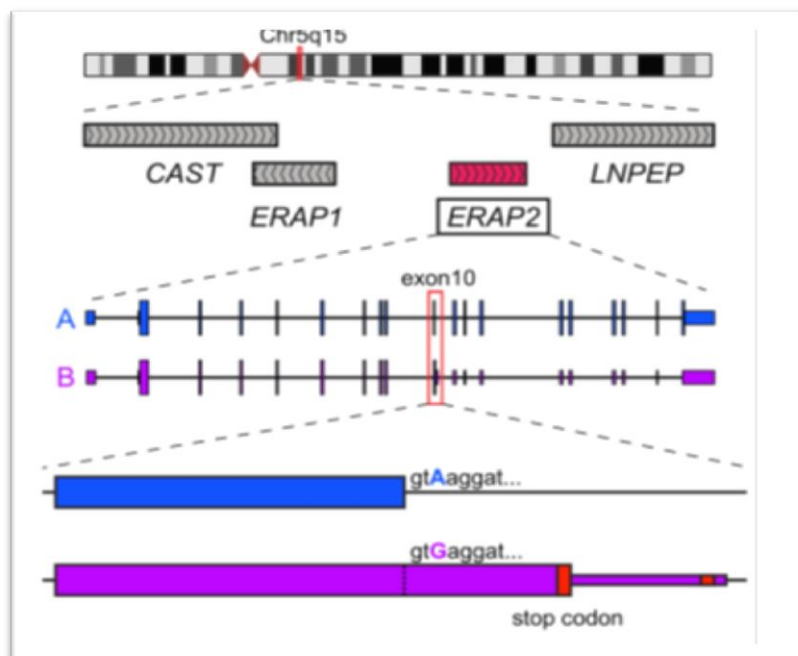
Tabell 18: Oversikt over hvilke transkripter Taqman - og Microarray ekspresjonsmetodene fanger opp.

Gen	TaqMan assay	TaqMan transkript	Microarray probe	Microarray transkript
ERAP1	Hs00429970_m1	NM_001040458.1 NM_001198541.1 NM_016442.3	ILMN_2336220	NM_001040458.1
ERAP1	Hs00372304_m1	NM_016442.3	ILMN_1752145	NM_016442.3
ERAP2	Hs01073631_m1	NM_001130140.1 NM_022350.3.	ILMN_1743145	NM_022350.2

På grunn av ufullstendig koblingsulikevekt mellom ERAP1 SNPene (rs30187, rs27524, rs27434) og ERAP2 SNPen (rs2248374), vil prøvene fordele seg ulikt i de ulike genotypegruppene avhengig av hvilken SNP man ser på.

De lavt uttrykte tymus-prøvene som fordelte seg som AA og AG for rs27524 og rs27434 på ERAP2-proben (ILMN_1743145), grupperer seg i homozygot rs2248374*GG gruppen på den samme proben. De homozygote GG*individene med høyt ERAP2 uttrykk, grupperte seg i rs2248374*AA gruppen.

Dette stemmer med at ERAP2 SNPen rs2248374 endrer spleisemønster og gir et prematurt stop kodon (Figur 23), som fører til et lavere genuttrykk for homozygote rs2248374*GG individer og tilsvarende høyere uttrykk for homozygot AA individer(81).



Figur 23: Illustrasjon av ERAP2 SNPen rs2248374 endrer spleisemønster og gir et prematurt stop kodon(82).

ERAP2 TaqMan genespresjonsassayet ,Hs01073631, fanger opp begge spleisevariantene (NM_001130140 og NM_0022350) (Tabell 19), mens Microarray ERAP2-proben (ILMN_1743145), fanger kun opp det ene transkriptet (NM_022350.2). Dette kan forklare forskjellene mellom assayene.

Tabell 19: ERAP2-proben fanger opp begge spleisevariant transkriptene på både TaqMan og Microarray genekspresjonen.

ERAP2-probe	TaqMan probe sekvens	TaqMan transkript	Microarray probe sekvens	Microarray transkript
ILMN_174314 5	AACTTGGCAGCTCTCCT	NM_001130140.1	AACTTGGCAGCT	NM_022350.2
	TCATGCGATTGCCAGA	NM_022350.3	CTCCTTCATGCG	
	CGTCCAAAGGGGCAGC		ATTGCCAGACGT	
	A		CCAAAGGGGCA	
			GCA	

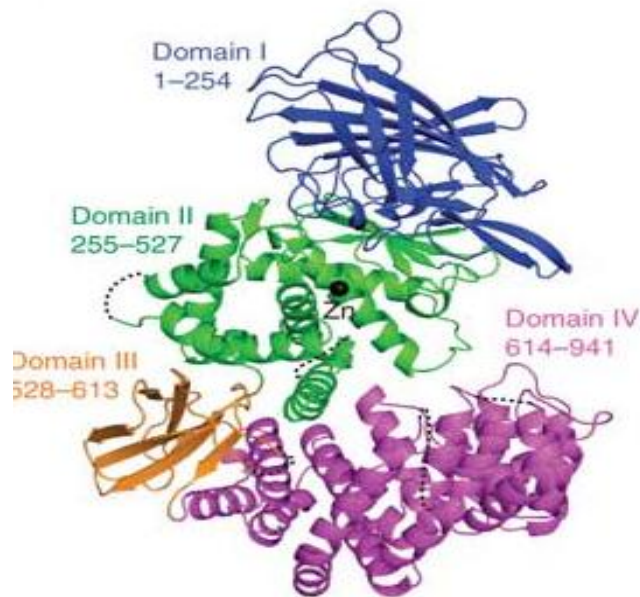
5.5 ERAP1/ERAP2 og AIS

De fire SNPene som viste assosiasjon med ERAP1/ERAP2-regionen, var assosierte med AS og PS, utenom rs2248374 som til nå ikke har vist noen assosiasjon med AS. Upubliserte data har derimot vist at både ERAP1 og ERAP2 er assosiert med AS, og at disse assosiasjonene ((Brown et al.) er uavhengige. Dette passer overens med bildet vi så på lite LD mellom SNPer i disse to genene.

Sterkest assosiasjon i ERAP1/ERAP2-regionen er funnet til å være ERAP1-SNPen rs30187 for AS(88). Dataene viser til at det er rs30187 som har den mest kausale assosiasjonen med AIS. Denne SNPen er den mest sannsynlige primære som er assosiert med AS, og er lokalisert på kromosom 5q15. rs30187 er mest signifikant assosiert med ERAP1-proben ILMN_2336220 ($p < 0,0001$), og var sterkt assosiert med kun det ene TaqMan ERAP1-assayene Hs00372304 ($p = 0,0131$) og ERAP2-assayet Hs01073631 ($p = 0,0139$).

5.5.1 ERAP1 og ERAP2

ERAP1, også kjent som ARTS 1 (Aminopeptidase Regulating Tumor necrosis factor receptor [TNFR] Shedding 1) er et gen som koder for en aminopeptidase i endoplasmatisk retikulum 1, involvert i trimming av HLA klasse I-bindende forløpere slik at de kan presenteres på MHC klasse I molekyler. Det kodete proteinet fungerer som en monomer eller som en heterodimer med ERAP2. Dette proteinet kan også være involvert i blodtrykket regulering ved inaktivering av angiotensin II. En annen funksjon er at ERAP1 kløyver cytokin reseptorer på celleoverflaten, ved dette reduseres deres evne til å overføre kjemiske signaler inn i cellen, noe som påvirker inflammasjons prosessen.



Figur 24: ERAP1 struktur og domene(89).

ERAP1 og ERAP2 genene, som også er uttrykt i blod, var signifikant assosierte (lod=3) med risiko SNPene i ERAP1/ERAP2-genregionen (Tabell 15). Dette styrker funnene av at de fungerer som eQTL ytterligere.

ERAP1 genvarianter er sterkt assosiert med AS (rs30187 som er sterkest), men ERAP1 assosiasjon er også rapportert hos pasienter med T1D(90), og MS(91). Polymorfismer i ERAP1 genotyper påvirker kun individer med AS, som er HLA-B27 positive(42, 88). Flere polymorfismer i ERAP1 genotyper har blitt funnet å påvirke risikoen til å utvikle AS, og hver av disse polymorfismene fører til en aminosyre- endring. Det er uklart i hvor stor grad en slik endring vil føre til sykdom, siden andre genetiske og miljømessige faktorer også påvirker sjansen for utvikling, men forskere mener at endringer i proteinstrukturen vil påvirke en av de to hoved funksjonene til ERAP1(92).

En polymorfisme i ERAP1-genet kan påvirke genets funksjon i tymus, ved at trimmingen av HLA klasse-I bindende forløpere blir mindre effektiv. Det at HLA klasse-I bindende forløpere blir delvis presentert på MHC klasse I molekyler, kan være innvirkende i utviklingen av AIS. Hattori A. et.al fant ut i sin studie at ERAP1 var utskilt fra makrofager i respons til aktivering av behandling med lipopolysakkarid (LPS) og interferon (IFN)- γ og forbedret sin fagocytterende aktivitet(93).

ERAP1 er også assosierte med psoriasis, og der også viser det HLA klasse I interaksjon, men da med HLA-C som er hoved HLA genet i psoriasis(42). Vi fant ikke at det var en tendens for at visse HLA

genotyper forekom sammen med visse ERAP1 og/eller ERAP2 genotyper. Dette er ikke så rart siden lociene befinner seg på ulike kromosomer, og dermed ikke nedarves sammen.

Det har blitt antydnet at ERAP2 trimmer «rester» som har blitt ineffektivt fjernet av ERAP1. SNPen rs2248374 (A/G) som er lokalisert innenfor 5' spleisesetet til ekson 10, resulterer i en alternativ spleising av ERAP2 mRNA som videre fører til degradering. Dette påvirker antigenpresentasjonen. Våre funn av at ERAP2 SNPen er en eQTL i tymus kan tyde på at denne SNPen påvirker effektiviteten av antigenpresentasjon i tymus i forhold til opplæring av T celler til å ikke gjenkjenne eget vev(94). En mindre effektiv antigenpresentasjon av autoantigener kan være en faktor i disponeringen av AIS. Denne ERAP2 genotypen er funnet assosiert med redusert overflate MHC klasse I uttrykk på B-celler.

5.6 Ingenuity Pathway Analyse

IPA er en analyse som ble brukt for å undersøke assosiasjonen mellom genekspressjon og biologiske funksjoner. Pathway analysen ga resultater for sykdom og lidelser, som er assosierte med signalveier som påvirker mottakelighet for AIS. Noe som styrket dette funnet, i forhold til genekspressjon i ERAP1/ERAP2-regionen, er at genene i denne regionen kom opp i de fem mest signifikante assosiasjonene (sykdom og lidelser, molekylære og biologiske funksjoner). ERAP1 var blant annet signifikant, signifikans grense $p=0,05$, assosiert med skjelett og muskelsykdommer ($3,98E-05-1,19E-01$), og det kan dermed bli spennende å se om de Internasjonale Immunochip assosiasjonsstudiene vil vise at dette genet også er assosiert med den autoimmune sykdommen RA, en kronisk leddsykdom som fører til autoimmun reaksjon mot ledd komponenter. Under biologiske og molekylære funksjoner var ERAP1 assosiert med inflammatorisk respons ($5,53E-02-1,29E-01$) og proteinsyntesen ($6,52E-04 - 2,57E-02$). En polymorfisme i ERAP1-genet kan dermed føre til endringer i den inflammatoriske responsen på flere nivåer og i flere ulike vev, og påvirke translasjonen av mRNA og/eller alternativ spleising.

Fishers eksakte test ble brukt til å kalkulere om sammenhengen mellom biologisk funksjon i IKB og en gruppe molekyler i datasettet er tilfeldig. Testen ble brukt fremfor en annen statistisk test, for å kunne være åpen i forhold til hva man finner, og deretter gå videre med andre metoder hvis det ønskes sterkere kriterier. Testen som blir brukt til å utføre analyse er avhengig av hvor stort data materialet er, Fisher hadde en grense på minst 100 gener.

Benjamin-Hochberg (B-H) metoden ble utført i tillegg for å undersøke om det ville forekomme p-verdi endringer, etter beregning av FDR og det ble korrigert for flere sammenlikninger etter flere

hypotesetester. Det utgjorde ingen forskjell på p-verdien/ratio å analysere datasettet ved B-H metoden, dette er positivt i forhold til resultatenes signifikans.

Fordelen med IPA er at analysen tar i bruk IKB bestående av kurrert data, som er vitenskaplige artikler med signifikante funn som er lagt inn manuelt. En svakhet ved analysen kan være at utfallet kun er basert på datasettet som er opplastet i IKB, og andre signifikante gener som kan være assosierte med de biologiske funksjonene ikke tas i betraktning.

En utfordring ved IPA kan være størrelsen av biologiske funksjoners utfall, da det kan være vanskelig å trekke konklusjoner og man er begrenset av de kunnskapene man har.

5.7 Feilkilder

Det innsamlede prøvematerialet besto av tymus-vev fra 42 norske barn (22 jenter og 20 gutter). Prøvematerialet ble behandlet som "friskt", mens faktumet er at det ble innsamlet under en hjerteoperasjon, noe som viser at dette var hjerte-syke pasienter. Denne faktoren kan påvirke genuttrykket i tymus, og dermed forstyrre funnene. En annen faktor som kan påvirke er om tymus-biten, som er tatt videre til DNA isolering er representativ, i forhold til om biten er fra cortex eller medulla (Figur 2) og dermed hvor stor andel av ulike cellepopulasjoner som biten inneholder. Videre var vi avhengige av at kirurgene rask skar opp vevsbiten og puttet den på RNAlater slik at ikke degradering av RNA satt i gang. Forskjeller i tidsrammen for dette, kan ha gitt skjevheter i materialet. Det er dog liten grunn til å tro at slike mulige skjevheter skal ha rammet visse genotyper og ikke tilfeldig. Materialstørrelsen var begrenset, da det ikke var noen annet tilgang på tymus prøvemateriale enn å bruke de tymus-bitene som ble innsamlet under hjerteoperasjonen.

ERAP1-assayet Hs00429970 fanget opp tre av tre transkripter for ERAP1-genet (Tabell 9), og gav gode Ct-resultater etter qPCR. Det var derimot ingen av SNPene i ERAP1/ERAP2-regionen som viste signifikant assosiasjon med dette assayet, med alle p-verdier over signifikant-grensen $p=0,05$ (Figur 17). Årsaken til dette kan være at transkriptene ikke er helt komplementære med gensekvensene i ERAP1/ERAP2-regionen.

De andre ERAP1-assayet Hs00372304_m1, var assosiert med kun ett av tre transkripter (NM_016442.3), og assayet viste signifikant assosiasjon med kun rs30187. Dette kan være et falskt positivt resultat, da det er ingen andre SNPer i ERAP1/ERAP2-regionen som var signifikant assosiert med dette assayet, og det ikke fanger opp alle tre transkriptene for ERAP1.

Alle de 10 eQTL SNPene var i HWE, foruten rs4788084 som viste et avvik fra HWE ($p=0,01$), med 100 % GSR (Tabell 6). Dette kan blant annet skyldes at denne SNPen er vanskelig å få korrekt genotype på, og dermed får en genotype som ikke er pålitelig. En annen årsak kan være at metoden man genotyper med kan være påvirket av for eksempel en annen SNP, som kun påvirker ett av allelene og dermed gir denne skjevheten i henhold til observert vs forventet.

6 Konklusjon

Konklusjonen baserer seg på funksjonelle eQTL studiene som er gjennomført i denne oppgaven, kombinert med genetiske assosiasjonstudier fra AIS. Prøvematerialet fra tymus ga oss muligheten til å undersøke genekspressjonen i tymus, og korrelere det mot risiko SNPs i ni autoimmune sykdommer. Prosjektet viste at risikovarianter for AIS er assosiert med ekspresjonsnivå av visse gener i tymus. Spesifikt, førte studien til en bedre forståelse for samspillet mellom ERAP1/ERAP2 SNP polymorfismene, og hvordan SNPene assosiert med AIS kan regulere genekspressjonen.

Funnene bør replikeres da det er begrenset med tymus prøvemateriale. For å kunne helt forstå endringene de genetiske forandringene fører til, kan videre analyse av ERAP1/ERAP2 spleisevariantene hjelpe. Videre ville det være viktig å avdekke om SNPene påvirker transkripsjonsbindingssteder eller spleiseseter.

Det kunne også vært mulig å undersøke om microRNA er en funksjon som spiller inn i reguleringen av ERAP1 og ERAP2 genene, som videre kan føre til AIS.

7 Referanser

1. Coutinho A CI, Seixas E, Demengeot J. Coutinho A, Caramalho I, Seixas E, Demengeot J. 2005.
2. Rai E, Wakeland EK. Genetic predisposition to autoimmunity – What have we learned? *Seminars in immunology*. 2011;23(2):67-83.
3. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-53.
4. Merkler D, Horvath E, Bruck W, Zinkernagel RM, Del la Torre JC, Pinschewer DD. "Viral déjà vu" elicits organ-specific immune disease independent of reactivity to self. *The Journal of clinical investigation*. 2006;116(5):1254-63. Epub 2006/04/11.
5. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nature reviews Genetics*. 2009;10(1):43-55. Epub 2008/12/19.
6. Bell JT, Saffery R. The value of twins in epigenetic epidemiology. *International journal of epidemiology*. 2012;41(1):140-50. Epub 2012/01/19.
7. Viken MK, Amundsen SS, Kvien TK, Boberg KM, Gilboe IM, Lilleby V, et al. Association analysis of the 1858C>T polymorphism in the PTPN22 gene in juvenile idiopathic arthritis and other autoimmune diseases. *Genes and immunity*. 2005;6(3):271-3. Epub 2005/03/11.
8. Gregersen PK, Lee HS, Batliwalla F, Begovich AB. PTPN22: setting thresholds for autoimmunity. *Seminars in immunology*. 2006;18(4):214-23. Epub 2006/05/30.
9. Olav Sand ØVS, Egil Haug, Jan G. Bjælie. *Menneskekroppen, fysiologi og anatomi*. Oslo: Gyldendal Norsk Forlag AS; 2009.
10. Lea T. *Immunologi og immunologiske teknikker*. Bergen: John Grieg AS; 2006.
11. DokuWiki. DokuWiki.
12. Wikipedia. Wkipedia.
13. Nishino M, Ashiku SK, Kocher ON, Thurer RL, Boiselle PM, Hatabu H. The thymus: a comprehensive review. *Radiographics : a review publication of the Radiological Society of North America, Inc*. 2006;26(2):335-48. Epub 2006/03/22.
14. Zanelli E, Gonzalez-Gay MA, David CS. Could HLA-DRB1 be the protective locus in rheumatoid arthritis? *Immunology Today*. 1995;16(6):274-8.
15. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet*. 2010;42(6):508-14. Epub 2010/05/11.
16. Undlien DE, Lie BA, Thorsby E. HLA complex genes in type 1 diabetes and other autoimmune diseases. Which genes are involved? *Trends in Genetics*. 2001;17(2):93-100.
17. Medicine USNLo. Genetics home reference. Genetics home reference2012.
18. Kaur NKMoG. *Expert Reviews in Molecular Medicine*. Expert Reviews in Molecular Medicine2003.
19. Somers EC, Thomas SL, Smeeth L, Hall AJ. Autoimmune Diseases Co-occurring Within Individuals and Within Families: A Systematic Review. *Epidemiology*. 2006;17(2):202-17 10.1097/01.ede.0000193605.93416.df.
20. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JHM, et al. Shared and Distinct Genetic Variants in Type 1 Diabetes and Celiac Disease. *New England Journal of Medicine*. 2008;359(26):2767-77.
21. Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *The Journal of experimental medicine*. 1989;169(1):345-50. Epub 1989/01/01.
22. (TASC) WTCCCTA-A-ASC. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet*. 2007;39(11):1329-37.
23. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet*. 2007;39(7):827-9.

24. Hunt KA, Zhernakova A, Turner G, Heap GAR, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet.* 2008;40(4):395-402.
25. Fernando MMA, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, et al. Defining the Role of the MHC in Autoimmunity: A Review and Pooled Analysis. *PLoS genetics.* 2008;4(4):e1000024.
26. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet.* 2008;40(8):955-62.
27. Fernando MMA, Stevens CR, Sabeti PC, Walsh EC, McWhinnie AJM, Shah A, et al. Identification of Two Independent Risk Factors for Lupus within the MHC in United Kingdom Families. *PLoS genetics.* 2007;3(11):e192.
28. Consortium TIMSG. Risk Alleles for Multiple Sclerosis Identified by a Genomewide Study. *New England Journal of Medicine.* 2007;357(9):851-62.
29. Liu Y, Helms C, Liao W, Zaba LC, Duan S, Gardner J, et al. A Genome-Wide Association Study of Psoriasis and Psoriatic Arthritis Identifies New Disease Loci. *PLoS genetics.* 2008;4(4):e1000041.
30. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burt NP, et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat Genet.* 2008;40(10):1216-23.
31. Bowes J, Barton A. Recent advances in the genetics of RA susceptibility. *Rheumatology.* 2008;47(4):399-402.
32. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PTK, KIAA1542 and other loci. *Nat Genet.* 2008;40(2):204-10.
33. The International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN)11 JBH, 2, Marta E Alarcón-Riquelme3, Lindsey A Criswell4, Chaim O Jacob5, Robert P Kimberly6, Kathy L Moser1,7, Betty P Tsao8, Timothy J Vyse9 & Carl D Langefeld10. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007;447(7145):661-78.
34. Franke A, Balschun T, Karlsen TH, Svantoraityte J, Nikolaus S, Mayr G, et al. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet.* 2008;40(11):1319-23.
35. Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat Genet.* 2008;40(6):710-2.
36. Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. *Nature.* 2004;431(7011):931-45.
37. Read TSaA. *Human Molecular Genetics*, Fourth edition: GS Publisher; 2010.
38. NCBI. NCBI, dbSNP Summery.
39. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 2003;422(6929):297-302.
40. Aasland R. *Genetisk variasjon: Alleler.* 2007.
41. Lipton B. Bruce H. Lipton, PhD. 2012.
42. Evans DM, Spencer CCA, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet.* 2011;43(8):761-7.
43. Australo-Anglo-American Spondyloarthritis Consortium (TASC) RJ, Sims AM, Danoy P, Evans DM, Leo P, Pointon JJ, Jin R, Zhou X, Bradbury LA, Appleton LH, Davis JC, Diekman L, Doan T, Dowling A, Duan R, Duncan EL, Farrar C, Hadler J, Harvey D, Karaderi T, Mogg R, Pomeroy E, Pryce K, Taylor J, Savage L, Deloukas P, Kumanduri V, Peltonen L, Ring SM, Whittaker P, Glazov E, Thomas GP, Maksymowych WP, Inman RD, Ward MM, Stone MA, Weisman MH, Wordsworth BP, Brown MA. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat Genet.* 2010;42(2):123-7.
44. Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet.* 2010;42(4):295-302.

45. Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *European journal of human genetics : EJHG*. 2012;20(7):801-5. Epub 2012/02/02.
46. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010;42(12):1118-25.
47. McGovern DPB, Jones MR, Taylor KD, Marciante K, Yan X, Dubinsky M, et al. Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Human Molecular Genetics*. 2010;19(17):3468-76.
48. Raelson JV, Little RD, Ruether A, Fournier H, Paquin B, Van Eerdewegh P, et al. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proceedings of the National Academy of Sciences*. 2007;104(37):14747-52.
49. Franke A, Hampe J, Rosenstiel P, Becker C, Wagner F, Häsler R, et al. Systematic Association Mapping Identifies *NELL1* as a Novel IBD Disease Gene. *PLoS ONE*. 2007;2(8):e691.
50. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet*. 2007;39(7):830-2.
51. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet*. 2007;39(5):596-604.
52. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, et al. Novel Crohn Disease Locus Identified by Genome-Wide Association Maps to a Gene Desert on 5p13.1 and Modulates Expression of *PTGER4*. *PLoS genetics*. 2007;3(4):e58.
53. International Multiple Sclerosis Genetics Consortium;Wellcome Trust Case Control Consortium 2 SS, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Oturai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S, Giannoulatou E, D'alfonso S, Blackburn H, Boneschi FM, Liddle J, Harbo HF, Perez ML, Spurkland A, Waller MJ, Mycko MP, Ricketts M, Comabella M, Hammond N, Kockum I, McCann OT, Ban M, Whittaker P, Kempainen A, Weston P, Hawkins C, Widaa S, Zajicek J, Dronov S, Robertson N, Bumpstead SJ, Barcellos LF, Ravindrarajah R, Abraham R, Alfredsson L, Ardlie K, Aubin C, Baker A, Baker K, Baranzini SE, Bergamaschi L, Bergamaschi R, Bernstein A, Berthele A, Boggild M, Bradfield JP, Brassat D, Broadley SA, Buck D, Butzkueven H, Capra R, Carroll WM, Cavalla P, Celius EG, Cepok S, Chiavacci R, Clerget-Darpoux F, Clysters K, Comi G, Cossburn M, Cournu-Rebeix I, Cox MB, Cozen W, Cree BA, Cross AH, Cusi D, Daly MJ, Davis E, de Bakker PI, Debouverie M, D'hooghe MB, Dixon K, Dobosi R, Dubois B, Ellinghaus D, Elovaara I, Esposito F, Fontenille C, Foote S, Franke A, Galimberti D, Ghezzi A, Glessner J, Gomez R, Gout O, Graham C, Grant SF, Guerini FR, Hakonarson H, Hall P, Hamsten A, Hartung HP, Heard RN, Heath S, Hobart J, Hoshi M, Infante-Duarte C, Ingram G, Ingram W, Islam T, Jagodic M, Kabesch M, Kermode AG, Kilpatrick TJ, Kim C, Klopp N, Koivisto K, Larsson M, Lathrop M, Lechner-Scott JS, Leone MA, Leppä V, Liljedahl U, Bomfim IL, Lincoln RR, Link J, Liu J, Lorentzen AR, Lupoli S, Macciardi F, Mack T, Marriott M, Martinelli V, Mason D, McCauley JL, Mentch F, Mero IL, Mihalova T, Montalban X, Mottershead J, Myhr KM, Naldi P, Ollier W, Page A, Palotie A, Pelletier J, Piccio L, Pickersgill T, Piehl F, Pobywajlo S, Quach HL, Ramsay PP, Reunanen M, Reynolds R, Rioux JD, Rodegher M, Roesner S, Rubio JP, Rückert IM, Salvetti M, Salvi E, Santaniello A, Schaefer CA, Schreiber S, Schulze C, Scott RJ, Sellebjerg F, Selmaj KW, Sexton D, Shen L, Simms-Acuna B, Skidmore S, Sleiman PM, Smestad C, Sørensen PS, Sørensgaard HB, Stankovich J, Strange RC, Sulonen AM, Sundqvist E, Syvänen AC, Taddeo F, Taylor B, Blackwell JM, Tienari P, Bramon E, Tourbah A, Brown MA, Tronczynska E, Casas JP, Tubridy N, Corvin A, Vickery J, Jankowski J, Villoslada P, Markus HS, Wang K, Mathew CG, Wason J, Palmer CN, Wichmann HE, Plomin R, Willoughby E, Rautanen A, Winkelmann J, Wittig M, Trembath RC, Yaouanq J, Viswanathan AC, Zhang H, Wood NW, Zuvich R, Deloukas P, Langford C, Duncanson A, Oksenberg JR, Pericak-Vance MA, Haines JL, Olsson T, Hillert J, Ivins AJ, De Jager PL, Peltonen L, Stewart GJ, Hafler DA, Hauser SL, McVean G, Donnelly P,

- Compston A. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011;476(7359):214-9.
54. Patsopoulos NA, Esposito F, Reischl J, Lehr S, Bauer D, Heubach J, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Annals of neurology*. 2011;70(6):897-912. Epub 2011/12/23.
 55. International Multiple Sclerosis Genetics C. Genome-wide association study of severity in multiple sclerosis. *Genes and immunity*. 2011;12(8):615-25.
 56. (ANZgene). AaNZMSGC. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet*. 2009;41(7):824-8.
 57. 2 GAOPTWTCCC. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat Genet*. 2010;42(11):985-90.
 58. Eleftherohorinou H, Hoggart CJ, Wright VJ, Levin M, Coin LJM. Pathway-driven gene stability selection of two rheumatoid arthritis GWAS identifies and validates new susceptibility genes in receptor mediated signalling pathways. *Human Molecular Genetics*. 2011;20(17):3494-506.
 59. Chung SA, Taylor KE, Graham RR, Nititham J, Lee AT, Ortmann WA, et al. Differential Genetic Associations for Systemic Lupus Erythematosus Based on Anti-dsDNA Autoantibody Production. *PLoS genetics*. 2011;7(3):e1001323.
 60. Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat Genet*. 2010;42(1):68-71.
 61. Bradfield JP, Qu H-Q, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci. *PLoS genetics*. 2011;7(9):e1002293.
 62. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*. 2011;43(3):246-52.
 63. McGovern DPB, Gardet A, Torkvist L, Goyette P, Essers J, Taylor KD, et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet*. 2010;42(4):332-7.
 64. 2 TUIGCTWTCCC. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet*. 2009;41(12):1330-4.
 65. Healthcare G. Illustra GenomiPhi V2 DNA Amplification Kit.
 66. Pan Du WAKaSML. lumi: a pipeline for processing Illumina microarray2008. Available from: <http://bioinformatics.oxfordjournals.org/content/24/13/1547.full.pdf+html>.
 67. Illumina. Illumina mRNA Expression Analysis. Available from: http://www.illumina.com/documents/products/datasheets/datasheet_mrna_expression.pdf.
 68. Nikula T, West A, Katajamaa M, Lönnberg T, Sara R, Aittokallio T, et al. A human ImmunoChip cDNA microarray provides a comprehensive tool to study immune responses. *Journal of Immunological Methods*. 2005;303(1-2):122-34.
 69. Biosystems A. Applied Biosystems.
 70. Caraguel CGB, Stryhn H, Gagné N, Dohoo IR, Hammell KL. Selection of a Cutoff Value for Real-Time Polymerase Chain Reaction Results to Fit a Diagnostic Purpose: Analytical and Epidemiologic Approaches. *Journal of Veterinary Diagnostic Investigation*. 2011;23(1):2-15.
 71. Sun Kim KNaTH. mCpG-SNP-EXPRESS.
 72. Liming Liang WCoGA. Center for Statistical Genetics.
 73. GraphPad software I. GraphPad software.
 74. Arild Bjørndal DH. Statistikk for Helse - og Sosialfag. Gyldendal.
 75. Technologies L. Life Technologies.
 76. Ingenuity Systems I. Ingenuity IPA software.
 77. Thissen D, Steinberg L, Kuang D. Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. *Journal of Educational and Behavioral Statistics*. 2002;27(1):77-83.

78. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263-5.
79. Devlin B, Risch N. A Comparison of Linkage Disequilibrium Measures for Fine-Scale Mapping. *Genomics*. 1995;29(2):311-22.
80. Cramer H. *Mathematical Methods of Statistics*. Princeton University Press.
81. Coulombe-Huntington J, Lam KC, Dias C, Majewski J. Fine-scale variation and genetic determinants of alternative splicing across individuals. *PLoS genetics*. 2009;5(12):e1000766. Epub 2009/12/17.
82. Andrés AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin S-Q, Hurle B, et al. Balancing Selection Maintains a Form of *ERAP2* that Undergoes Nonsense-Mediated Decay and Affects Antigen Presentation. *PLoS genetics*. 2010;6(10):e1001157.
83. Sham SPP. *Genetic Power Calculation*. 2001-2009.
84. Systems I. *Ingenuity Systems Citation Guidelines*. 2011.
85. Holloway B, Luck S, Beatty M, Rafalski J-A, Li B. Genome-wide expression quantitative trait loci (eQTL) analysis in maize. *BMC Genomics*. 2011;12(1):336.
86. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS genetics*. 2010;6(4):e1000888.
87. Aarnes H. *Cellebiologi og genetikk 2004*. Available from: <http://www.mn.uio.no/bio/tjenester/kunnskap/plantefys/zoologi/cellebio.pdf>.
88. Haroon N, Tsui FW, Uchanska-Ziegler B, Ziegler A, Inman RD. Endoplasmic reticulum aminopeptidase 1 (ERAP1) exhibits functionally significant interaction with HLA-B27 and relates to subtype specificity in ankylosing spondylitis. *Annals of the Rheumatic Diseases*. 2012;71(4):589-95.
89. Nguyen TT, Chang S-C, Evnouchidou I, York IA, Zikos C, Rock KL, et al. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat Struct Mol Biol*. 2011;18(5):604-13.
90. Fung EYMG, Smyth DJ, Howson JMM, Cooper JD, Walker NM, Stevens H, et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes and immunity*. 2008;10(2):188-91.
91. Guerini FR, Cagliani R, Forni D, Agliardi C, Caputo D, Cassinotti A, et al. A Functional Variant in *ERAP1* Predisposes to Multiple Sclerosis. *PLoS ONE*. 2012;7(1):e29931.
92. Reference GH. *Genetics Home Reference*.
93. Hattori A, Goto Y, Tsujimoto M. Exon 10 Coding Sequence Is Important for Endoplasmic Reticulum Retention of Endoplasmic Reticulum Aminopeptidase 1. *Biological and Pharmaceutical Bulletin*. 2012;35(4):601-5.
94. Evnouchidou I, Kamal RP, Seregin SS, Goto Y, Tsujimoto M, Hattori A, et al. Cutting Edge: Coding Single Nucleotide Polymorphisms of Endoplasmic Reticulum Aminopeptidase 1 Can Affect Antigenic Peptide Generation In Vitro by Influencing Basic Enzymatic Properties of the Enzyme. *The Journal of Immunology*. 2011;186(4):1909-13.

8 Vedlegg

8.1 Vedlegg A

Tabell 20: Autoimmune risiko SNPs (N=256).

Immunochip ID	SNP	Pos	Chr	Sykdom
1kg_1_159733666	rs12746613	159733666	1	RA
1kg_1_159746369	rs1801274	159746369	1	SLE
1kg_1_171576336	rs10798269	171576336	1	SLE
1kg_1_93076191	rs6604026	93076191	1	MS
imm_1_114179091	rs2476601	114179091	1	RA
imm_1_116839810	rs12025416	116839810	1	MS
imm_1_116902480	rs1335532	116902480	1	MS
imm_1_116905738	rs2300747	116905738	1	MS
imm_1_117064661	rs11586238	117064661	1	RA
imm_1_170977623	rs859637	170977623	1	CeD
imm_1_190803436	rs2816316	190803436	1	CeD
imm_1_199144185	rs7554511	199144185	1	CD, CeD
imm_1_199158760	rs296547	199158760	1	CeD
imm_1_199202489	rs11584383	199202489	1	CeD, CD, MS, Ps, SLE, T1D
imm_1_199226930	rs2297909	199226930	1	AS
imm_1_205006527	rs3024505	205006527	1	T1D
imm_1_2503076	rs734999	2503076	1	CeD
imm_1_2516606	rs3748816	2516606	1	CeD
imm_1_25176163	rs10903122	25176163	1	CeD
imm_1_2543484	rs3890745	2543484	1	RA
imm_1_67454257	rs7517847	67454257	1	UC
imm_1_67475114	rs11465804	67475114	1	CeD, CD, MS, Ps, SLE, T1D
imm_1_67478546	rs11209026	67478546	1	AS
imm_1_67497708	rs10889677	67497708	1	CeD, CD, MS, Ps, SLE, T1D

imm_1_67512680	rs11209032	67512680	1	CeD, CD, MS, Ps, SLE, T1D
imm_1_7969259	rs12727642	7969259	1	CeD
rs10492972	rs10492972	10275699	1	MS
rs12141391	rs12141391	72047061	1	SLE
rs1801274	rs1801274	159746369	1	SLE
rs2201841	rs2201841	67466790	1	CeD, CD, MS, Ps, SLE, T1D
rs2269241	rs2269241	63881359	1	T1D
rs3748816	rs3748816	2516606	1	CeD
rs4112788	rs4112788	150817900	1	Ps, RA
rs4649203	rs4649203	24392507	1	Ps
rs6691768	rs6691768	61564451	1	CeD
rs864537	rs864537	165678008	1	CeD
seq-t1d-1-196967065- G-A	rs10919563	196967065	1	RA
1kg_2_241227781	rs4676406	241227781	2	UC
imm_2_100173372	rs11676922	100173372	2	RA
imm_2_100202166	rs10865035	100202166	2	RA
imm_2_102030060	rs2310173	102030060	2	AS
imm_2_102338297	rs13015714	102338297	2	CeD
imm_2_162832297	rs1990760	162832297	2	T1D
imm_2_162968937	rs17716942	162968937	2	Ps
imm_2_181716045	rs1018326	181716045	2	AS
imm_2_191611003	rs3821236	191611003	2	SLE
imm_2_191672878	rs7574865	191672878	2	SLE
imm_2_204318641	rs1980422	204318641	2	RA
imm_2_204447164	rs3087243	204447164	2	RA
imm_2_204510823	rs4675374	204510823	2	CeD
imm_2_60935046	rs702873	60935046	2	Ps
imm_2_60945454	rs842636	60945454	2	Ps
imm_2_60989633	rs13031237	60989633	2	RA
imm_2_61040333	rs13003464	61040333	2	CeD

imm_2_61058360	rs7608910	61058360	2	RA, CD, CeD
imm_2_62404976	rs10865331	62404976	2	AS
imm_2_68452459	rs17035378	68452459	2	CeD
rs11676348	rs11676348	218718391	2	UC
rs1534422	rs1534422	12558192	2	T1D
imm_3_120601486	rs11712165	120601486	3	CeD
imm_3_161147744	rs17810546	161147744	3	CeD
imm_3_161179692	rs9811792	161179692	3	CeD
imm_3_161181639	rs4680534	161181639	3	MS
imm_3_189595248	rs1464510	189595248	3	CeD
imm_3_32990473	rs13314993	32990473	3	CeD
imm_3_46210205	rs13098911	46210205	3	CeD
imm_3_46327388	rs6441961	46327388	3	CeD
imm_3_49694733	rs9822268	49694733	3	CD
imm_3_58345217	rs6445975	58345217	3	SLE
rs10936599	rs10936599	170974795	3	CeD
rs6806528	rs6806528	69335589	3	CeD
imm_4_102970099	rs10516487	102970099	4	SLE
imm_4_123334952	rs13151961	123334952	4	CeD
imm_4_123351942	rs4505848	123351942	4	T1D
imm_4_123548812	rs17388568	123548812	4	CeD, T1D
imm_4_123596932	rs2069763	123596932	4	CeD, RA, T1D
imm_4_123728871	rs6822844	123728871	4	RA
rs10517086	rs10517086	25694609	4	T1D
1kg_5_647083	rs11739663	647083	5	UC
ccc-5-96150086-T-C	rs30187	96150086	5	AS
imm_5_131798704	rs2188962	131798704	5	CeD, CD, MS, Ps, SLE, T1D
imm_5_132023863	rs20541	132023863	5	Ps
imm_5_158650367	rs2082412	158650367	5	Ps
imm_5_158683347	rs3213094	158683347	5	Ps
imm_5_158696755	rs7709212	158696755	5	CeD, CD, MS, Ps, SLE, T1D

imm_5_158751323	rs6556416	158751323	5	AS
imm_5_158759370	rs6871626	158759370	5	CD, Ps, SLE
imm_5_35839334	rs931555	35839334	5	MS
imm_5_35910332	rs6897932	35910332	5	MS
imm_5_35912031	rs3194051	35912031	5	MS
imm_5_40428485	rs4613763	40428485	5	CeD, CD, MS, Ps, SLE, T1D
imm_5_40446692	rs6451493	40446692	5	CD, MS
imm_5_40460183	rs6896969	40460183	5	MS
imm_5_40526547	rs10440635	40526547	5	AS
imm_5_96127700	rs27524	96127700	5	Ps
imm_5_96155268	rs27434	96155268	5	AS
rs17728338	rs17728338	150458511	5	Ps
rs254560	rs254560	134471505	5	UC
imm_6_106628720	rs6911490	106628720	6	CD, RA, SLE
imm_6_106674727	rs548234	106674727	6	RA
imm_6_126740412	rs9388489	126740412	6	T1D
imm_6_128320491	rs802734	128320491	6	CeD
imm_6_138014761	rs2327832	138014761	6	CeD
imm_6_138044330	rs10499194	138044330	6	RA
imm_6_138048197	rs6920220	138048197	6	RA
imm_6_138236844	rs5029937	138236844	6	RA
imm_6_138237416	rs5029939	138237416	6	SLE
imm_6_138241110	rs610604	138241110	6	Ps
imm_6_159385965	rs1738074	159385965	6	CeD
imm_6_159389562	rs182429	159389562	6	CeD, RA
imm_6_159402509	rs394581	159402509	6	RA
imm_6_328546	rs1033180	328546	6	CeD
imm_6_90983333	rs10806425	90983333	6	CeD
rs11962089	rs11962089	105718913	6	MS
rs12191877	rs12191877	31360904	6	Ps
rs2187668	rs2187668	32713862	6	SLE
rs2395029	rs2395029	31539759	6	Ps

rs240993	rs240993	111780407	6	Ps
rs3129934	rs3129934	32444165	6	MS
rs3131379	rs3131379	31829012	6	SLE
rs3134792	rs3134792	31420305	6	Ps
rs3135388	rs3135388	32521029	6	MS
rs477515	rs477515	32677669	6	IBD
rs6910071	rs6910071	32390832	6	RA
rs7743761	rs7743761	31444079	6	AS
rs9268853	rs9268853	32537621	6	CD, CeD, MS, RA, T1D
rs9271366	rs9271366	32694832	6	MS
rs943072	rs943072	43903946	6	UC
imm_7_128361203	rs4728142	128361203	7	CeD, CD, MS, Ps, SLE, T1D
imm_7_128381419	rs10488631	128381419	7	SLE
imm_7_128505142	rs12537284	128505142	7	SLE
imm_7_26858190	rs7804356	26858190	7	CeD, CD, MS, Ps, SLE, T1D
imm_7_50276409	rs4917014	50276409	7	CeD, CD, MS, Ps, SLE, T1D
imm_7_50994688	rs4948088	50994688	7	T1D
rs4728142	rs4728142	128361203	7	CeD, CD, MS, Ps, SLE, T1D
rs4917014	rs4917014	50276409	7	CeD, CD, MS, Ps, SLE, T1D
rs6974491	rs6974491	37341035	7	CeD
imm_8_11381382	rs2736340	11381382	8	RA
imm_8_11386595	rs13277113	11386595	8	SLE
imm_8_11389950	rs2618476	11389950	8	SLE
imm_8_129333771	rs9792269	129333771	8	CeD
imm_9_116593070	rs4246905	116593070	9	CD
imm_9_116608587	rs6478109	116608587	9	IBD
imm_9_122730060	rs3761847	122730060	9	RA
imm_9_138386226	rs10781499	138386226	9	CD

imm_9_138389159	rs10781500	138389159	9	AS
imm_9_34700260	rs2812378	34700260	9	RA
imm_9_34733681	rs951005	34733681	9	RA
imm_9_4281747	rs7020673	4281747	9	T1D
imm_9_4971602	rs10758669	4971602	9	CeD, CD, MS, Ps, SLE, T1D
rs3780792	rs3780792	135825164	9	MS
imm_10_35594060	rs12261843	35594060	10	CD
imm_10_49789060	rs1913517	49789060	10	CeD, CD, MS, Ps, SLE, T1D
imm_10_6137289	rs12722495	6137289	10	CeD, MS, RA
imm_10_6138955	rs706778	6138955	10	RA
imm_10_6139051	rs2104286	6139051	10	MS
imm_10_6142018	rs12722489	6142018	10	MS
imm_10_6162015	rs11594656	6162015	10	CeD, MS, RA
imm_10_6163501	rs12251307	6163501	10	T1D
imm_10_6430456	rs947474	6430456	10	CeD, CD, MS, Ps, SLE, T1D
imm_10_6433266	rs4750316	6433266	10	RA
imm_10_6512897	rs11258747	6512897	10	T1D
imm_10_80706013	rs1250540	80706013	10	MS
imm_10_80728033	rs1250552	80728033	10	CeD
imm_10_90013013	rs10509540	90013013	10	T1D
rs11101442	rs11101442	49606342	10	SLE
rs1913517	rs1913517	49789060	10	CeD, CD, MS, Ps, SLE, T1D
imm_11_127816269	rs6590330	127816269	11	CeD, CD, MS, Ps, SLE, T1D
imm_11_127886184	rs11221332	127886184	11	CeD
imm_11_2138800	rs689	2138800	11	T1D, RA
imm_11_2169742	rs7111341	2169742	11	T1D
imm_11_579564	rs4963128	579564	11	SLE
imm_11_75976842	rs2155219	75976842	11	CD

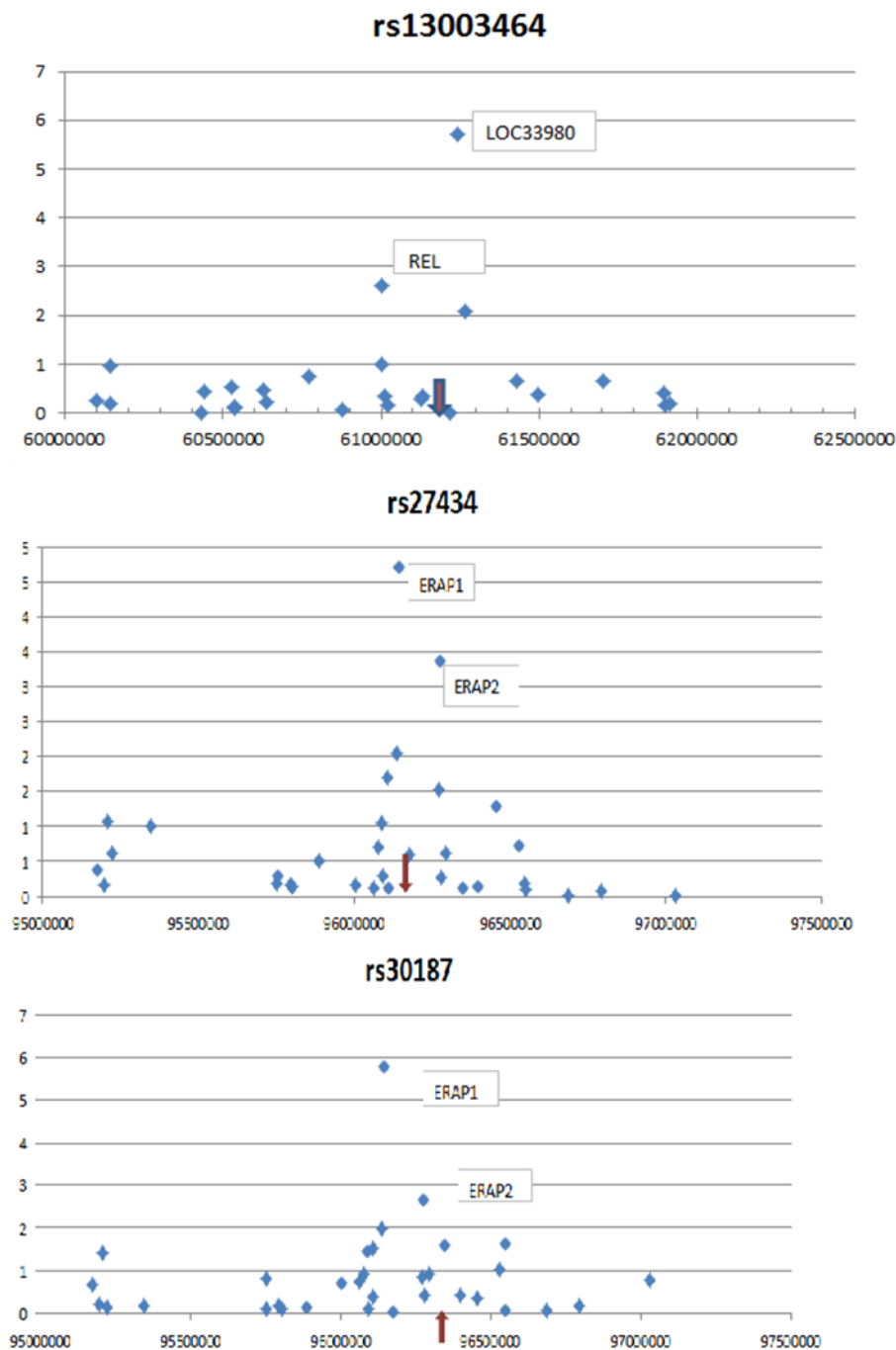
rs678170	rs678170	113937166	11	UC
rs907611	rs907611	1830648	11	UC
ccc-12-6373003-G-A	rs11616188	6373003	12	AS
imm_12_110368991	rs3184504	110368991	12	T1D
imm_12_110492139	rs653178	110492139	12	CeD
imm_12_122222678	rs1790100	122222678	12	MS
imm_12_54725476	rs12580100	54725476	12	Ps
imm_12_54768447	rs2292239	54768447	12	T1D
imm_12_55024240	rs2066808	55024240	12	Ps
imm_12_56254982	rs1678542	56254982	12	RA
imm_12_56449006	rs703842	56449006	12	MS
imm_12_6310270	rs1800693	6310270	12	MS
imm_12_6312904	rs4149584	6312904	12	MS
imm_12_66786342	rs7134599	66786342	12	UC
imm_12_9801431	rs4763879	9801431	12	T1D
rs17085007	rs17085007	26429267	13	UC
rs2762051	rs2762051	49733716	13	CeD
rs7993214	rs7993214	39248912	13	Ps
rs941823	rs941823	39911977	13	UC
1kg_14_34902417	rs8016947	34902417	14	Ps
imm_14_100375798	rs941576	100375798	14	T1D
imm_14_68333352	rs1465788	68333352	14	T1D
imm_14_97568704	rs4900384	97568704	14	T1D
imm_15_77022501	rs3825932	77022501	15	T1D
imm_16_11074189	rs11865121	11074189	16	CeD, CD, MS, Ps, SLE, T1D
imm_16_11087374	rs12708716	11087374	16	T1D
imm_16_11156830	rs6498169	11156830	16	MS
imm_16_11311394	rs12928822	11311394	16	CeD
imm_16_28447349	rs4788084	28447349	16	T1D
imm_16_30550368	rs7197475	30550368	16	CeD, CD, MS, Ps, SLE, T1D
imm_16_30850126	rs10782001	30850126	16	Ps

imm_16_31220754	rs9888739	31220754	16	SLE
imm_16_31241737	rs11150610	31241737	16	SLE
imm_16_49314275	rs5743289	49314275	16	IBD
imm_16_49314382	rs2076756	49314382	16	IBD
imm_16_67232289	rs6499188	67232289	16	UC
imm_16_73804746	rs7202877	73804746	16	T1D
imm_16_84575164	rs17445836	84575164	16	MS
rs11574637	rs11574637	31276375	16	SLE
rs12708716	rs12708716	11087374	16	T1D
rs6498169	rs6498169	11156830	16	MS
imm_17_35294289	rs2872507	35294289	17	CD, T1D
imm_17_35319766	rs2290400	35319766	17	T1D
imm_17_36023812	rs7221109	36023812	17	T1D
imm_17_37767727	rs744166	37767727	17	CeD, CD, MS, Ps, SLE, T1D
rs16956936	rs16956936	7574417	17	T1D
rs1975974	rs1975974	21631187	17	Ps
rs2074404	rs2074404	42220599	17	CeD
rs4795067	rs4795067	23130802	17	Ps
imm_18_12769947	rs2542151	12769947	18	CeD, CD, MS, Ps, SLE, T1D
imm_18_12782940	rs45450798	12782940	18	CeD, T1D
imm_18_12799340	rs1893217	12799340	18	T1D
imm_18_12825976	rs478582	12825976	18	CeD, T1D
imm_18_65682622	rs763361	65682622	18	CeD, CD, MS, Ps, SLE, T1D
imm_19_10330975	rs12720356	10330975	19	Ps
imm_19_10333933	rs280519	10333933	19	Ps
imm_19_10336652	rs2304256	10336652	19	T1D
imm_19_51900321	rs425105	51900321	19	T1D
rs12720356	rs12720356	10330975	19	Ps
rs2304256	rs2304256	10336652	19	T1D
rs3745672	rs3745672	12007370	19	MS

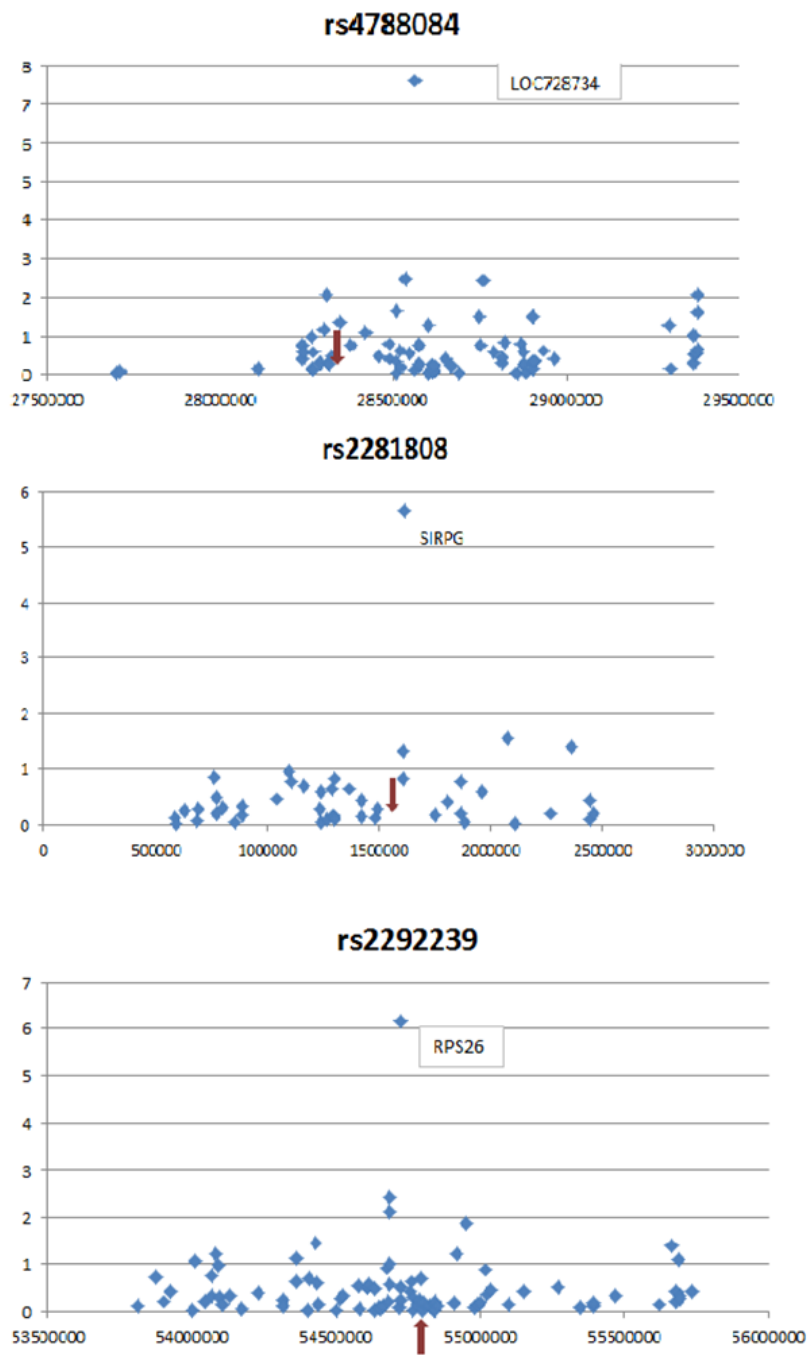
1kg_20_43414140	rs1008953	43414140	20	Ps
imm_20_1558551	rs2281808	1558551	20	T1D
imm_20_44173603	rs6074022	44173603	20	MS
imm_20_44181354	rs4810485	44181354	20	RA
imm_20_61798026	rs2297441	61798026	20	CD
imm_20_61814400	rs2315008	61814400	20	IBD
imm_21_15738922	rs1297265	15738922	21	CD
imm_21_39387048	rs2242944	39387048	21	AS
imm_21_39387404	rs2836878	39387404	21	IBD
imm_21_39391390	rs378108	39391390	21	AS
imm_21_42709255	rs11203203	42709255	21	T1D
imm_21_42714397	rs3788013	42714397	21	T1D, RA
imm_21_44439451	rs2838519	44439451	21	CD
imm_21_44471849	rs4819388	44471849	21	CeD
imm_22_20139185	rs463426	20139185	22	CeD, CD, MS, Ps, SLE, T1D
imm_22_20312892	rs2298428	20312892	22	CeD
imm_22_28911722	rs5753037	28911722	22	T1D
imm_22_35874756	rs3218253	35874756	22	RA
imm_22_35921264	rs229541	35921264	22	T1D
1kg_X_153598796	rs2664170	153598796	X	T1D
rs2664170	rs2664170	153598796	X	T1D
rs5979785	rs5979785	12881445	X	CeD

8.2 Vedlegg B

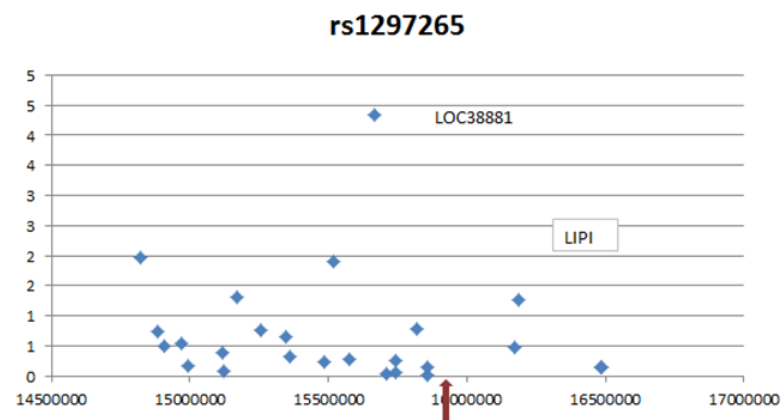
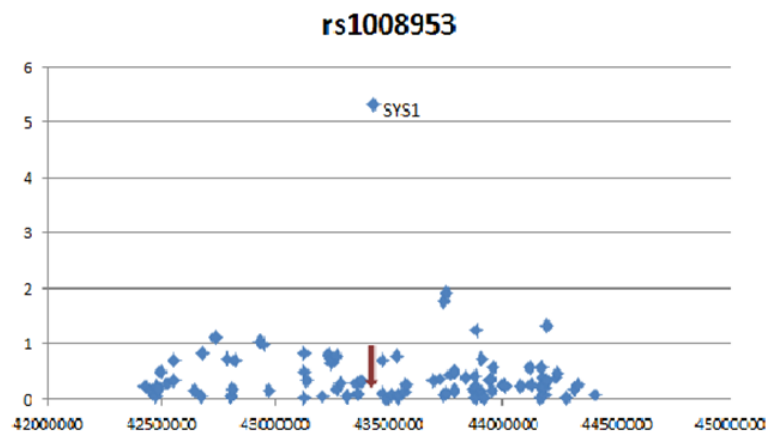
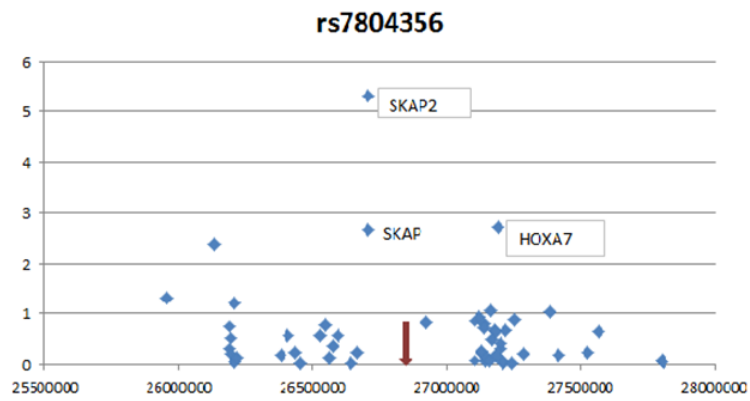
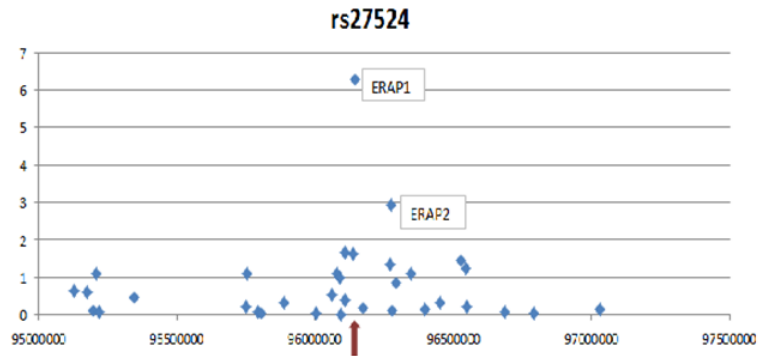
Under dette vedlegget finnes de øvrige 10 SNPene, utenfor ERAP1/ERAP2-regionen, som viste signifikante avvik ($p < 5 \times 10^{-5}$) i ekspressionsnivå mellom ulike genotyper. Pilene på hver graf indikerer hvor den enkelte nukleotid polymorfismen ligger. Disse SNPene fikk vi ikke sett nærmere på, fordi vi valgte å gå videre med de SNPs (rs30187, rs27434 og rs27434) som viste assosiasjon til ERAP1 og ERAP2.



Figur 25: Graf over risiko SNPene rs13003464, rs27434 og rs30187.



Figur 26: Graf over risiko SNPene rs4788084, rs2281808 og rs2292239.



Figur 27: Graf over risiko SNPene rs27524, rs7804356, rs1008953 og rs1297265.

8.3 Vedlegg C

ImmunoChip og TaqMan genotype dataene for ERAP1-SNPen rs30187.

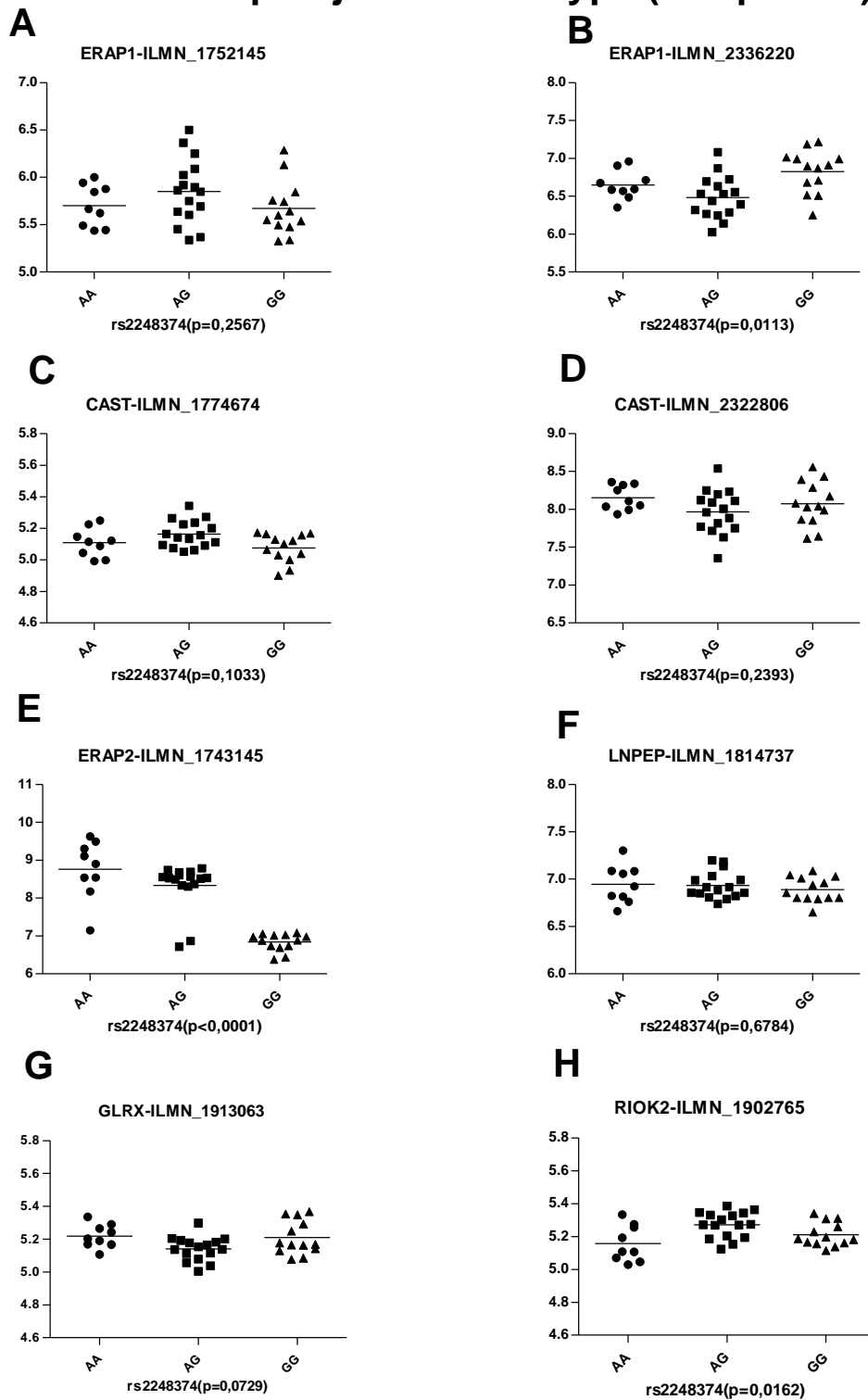
Sample	ImmunoChip genotype	Sample	TaqMan genotype
T01	G/G	T01	G/G
T02	0	T02	A/A
T05	G/G	T05	G/G
T06	0	T06	A/G
T09	G/G	T09	G/G
T10	A/G	T10	A/G
T11	G/G	T11	G/G
T12	0	T12	A/A
T13	0	T13	G/G
T15	A/A	T15	A/A
T16	A/G	T16	A/G
T18	A/G	T18	A/G
T19	G/G	T19	G/G
T20	A/A	T20	A/A
T22	G/G	T22	G/G
T23	A/G	T23	A/G
T24	0	T24	G/G
T25	G/G	T25	G/G
T27	0	T27	G/G
T28	A/G	T28	A/G
T30	A/A	T30	A/A
T31	A/G	T31	A/G
T32	G/G	T32	G/G
T33	G/G	T33	G/G
T34	G/G	T34	0
T36	A/G	T36	A/G
T37	0	T37	A/A
T38	A/A	T38	A/A
T39	A/G	T39	A/G
T40	0	T40	A/G
T43	G/G	T43	G/G
T44	G/G	T44	G/G
T45	G/G	T45	G/G
T47	A/G	T47	A/G
T48	A/G	T48	A/G
T51	A/G	T51	A/G
T52	A/A	T52	A/A

T53	G/G	T53	G/G
T55	A/G	T55	A/G
T56	G/G	T56	G/G
T57	G/G	T57	G/G
T58	A/G	T58	A/G

8.4 Vedlegg D

Oversikt over alle probene som ble korrelert med ERAP2-SNPen rs2248374 (N=8).

Genekspresjon vs Genotype (alle prober)



8.5 Vedlegg E

Oversikt over ekspresjons p-verdien for de 13 eQTL som viste signifikante avvik ($p < 5 \times 10^{-5}$) i ekspresjonsnivå mellom ulike genotyper.

SNP	Gen	Kromosom	Probe	Exp. p-verdi
rs7608910	REL	2	ILMN_1766085	0,02407
rs7608910	LOC339804	2	ILMN_1754501	0,00001141
rs7608910	AHSA2	2	ILMN_1798308	0,005174
rs13003464	REL	2	ILMN_1766085	0,00253
rs13003464	LOC339804	2	ILMN_1754501	0,000001897
rs13003464	AHSA2	2	ILMN_1798308	0,007975
rs27434	CAST	5	ILMN_2322806	0,01951
rs27434	ERAP1	5	ILMN_1752145	0,008893
rs27434	ERAP1	5	ILMN_2336220	0,0000198
rs27434	LNPEP	5	ILMN_1798433	0,03022
rs27434	ERAP2	5	ILMN_1743145	0,000419
rs30187	GLRX	5	ILMN_1913063	0,03921
rs30187	CAST	5	ILMN_1774674	0,03629
rs30187	CAST	5	ILMN_2322806	0,0297
rs30187	ERAP1	5	ILMN_1752145	0,01035
rs30187	ERAP1	5	ILMN_2336220	0,0000016
rs30187	LNPEP	5	ILMN_1814737	0,02674
rs30187	ERAP2	5	ILMN_1743145	0,002235
rs30187	RIOK2	5	ILMN_1902765	0,02333
rs27524	CAST	5	ILMN_2322806	0,02129
rs27524	ERAP1	5	ILMN_1752145	0,02309
rs27524	ERAP1	5	ILMN_2336220	0,000000504
rs27524	LNPEP	5	ILMN_1798433	0,04535
rs27524	ERAP2	5	ILMN_1743145	0,001199
rs27524	RIOK2	5	ILMN_1655329	0,0367
rs7804356	NFE2L3	7	ILMN_1912827	0,04785
rs7804356	NFE2L3	7	ILMN_1831436	0,00429
rs7804356	SKAP2	7	ILMN_2125010	0,002157
rs7804356	SKAP2	7	ILMN_1657129	0,000004804
rs7804356	HOX A7	7	ILMN_1706478	0,001948
rs10781500	CARD9	9	ILMN_1712532	0,02555
rs10781500	FUT7	9	ILMN_1724236	0,02929
rs10781500	QSOX2	9	ILMN_1789384	0,04454
rs10781500	INPP5E	9	ILMN_1811301	0,00005266
rs2292239	GDF11	12	ILMN_1745132	0,0348
rs2292239	SUOX	12	ILMN_2383455	0,00379
rs2292239	SUOX	12	ILMN_1803745	0,007266

rs2292239	RPS26	12	ILMN_2209027	0,000000664
rs2292239	COQ10A	12	ILMN_1666364	0,01352
rs2292239	LOC390332	12	ILMN_1652967	0,04138
rs4788084	LOC653352	16	ILMN_1660942	0,008443
rs4788084	LOC728689	16	ILMN_1700723	0,04537
rs4788084	CCDC101	16	ILMN_1684789	0,02211
rs4788084	SULT1A1	16	ILMN_1657302	0,00345
rs4788084	LOC728734	16	ILMN_1687571	0,000000025
rs4788084	ATXN2L	16	ILMN_2354334	0,03219
rs4788084	TUFM	16	ILMN_1738369	0,003772
rs4788084	SPNS1	16	ILMN_1681016	0,03203
rs4788084	SULT1A4	16	ILMN_2336130	0,02387
rs4788084	SULT1A4	16	ILMN_2336133	0,00882
rs45450798	MPPE1	18	ILMN_1667744	0,01996
rs45450798	IMPA2	18	ILMN_2094061	0,007217
rs45450798	SLMO1	18	ILMN_1654671	0,001066
rs45450798	SLMO1	18	ILMN_2232157	0,00005163
rs45450798	SPIRE1	18	ILMN_1757845	0,002258
rs45450798	SEH1L	18	ILMN_2381138	0,01248
rs45450798	SEH1L	18	ILMN_1705032	0,02792
rs45450798	PSMG2	18	ILMN_1832943	0,04799
rs2281808	SIRPG	20	ILMN_2383058	0,04853
rs2281808	SIRPG	20	ILMN_1676780	0,000002226
rs2281808	STK35	20	ILMN_1727809	0,02793
rs2281808	TGM6	20	ILMN_1776214	0,04146
rs1008953	SYS1	20	ILMN_1756590	0,000004977
rs1008953	WFDC5	20	ILMN_2079042	0,01718
rs1008953	WFDC12	20	ILMN_2231576	0,01222
rs1008953	CD40	20	ILMN_1779257	0,04977
rs1297265	LOC388813	21	ILMN_1857771	0,01096
rs1297265	NRIP1	21	ILMN_1845019	0,0481
rs1297265	LIPI	21	ILMN_1805609	0,01231
rs1297265	LOC388814	21	ILMN_1729831	0,00004536