

TARTU ÜLIKOOL
Loodus- ja täppisteaduste valdkond
Arvutiteaduse instituut
Informaatika õppekava

Peep Kolberg

Atoopilise dermatiidiga seotud geenide
tuvastamine geneetilise kolokalisatsiooni
abil

Bakalaureusetöö (9 EAP)

Juhendaja(d): Kaur Alasoo, PhD
Ana Rebane, PhD

Tartu 2021

Atoopilise dermatiidiga seotud geenide tuvastamine geneetilise kolokalisatsiooni abil

Lühikokkuvõte:

Atoopiline dermatiit on pärilik põletikuline nahahaigus. Haiguse tunnusteks on muutused naharakkudes, naha mikroobioomis ja immuunsüsteemi aktiivsuses, kuid pole selge, kuidas need DNA põhjal kujunevad. Genoomiuuringud seostavad geneetilisi variante fenotüüpidega. Kombineerides GWAS ja eQTL analüüside andmestikke on võimalik tuvastada geneetilised variandid, mis seostuvad haigusega ja mõjutavad ka geeniekspressiooni. Mõjutatud geenid mängivad tõenäoliselt rolli haiguse tekkel ja nende signaaliradade uurimine aitab kirjeldada atoopilise dermatiidi molekulaarseid mehhanisme. Bakalaureusetöös kolokaliseeriti andmestikud kahest GWAS uuringust ja 112-st eQTL analüüsist. Haigusega seostati 148 geeni, millest lähemalt vaadati interleukiin-1 tsütokiinide perekonna geenide ekspressiooniandmeid. Välja toodi geenid *IL1RL1*, *IL18*, *IL18R1*, *IL18RAP*. Tulemused kinnitavad, et interleukiin-1 perekonna valgud mõjutavad atoopilise dermatiidi teket. Tulemused on aluseks edasistele uuringutele konkreetsete geenide panuse kohta haiguses.

Võtmesõnad:

atoopiline dermatiit, kolokalisatsioon, GWAS, eQTL

CERCS: P175 Informaatika, süsteemiteooria

B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

Identifying Genes Associated With Atopic Dermatitis Using Genetic Colocalisation

Abstract:

Atopic dermatitis is a multifactorial inflammatory skin disease with a strong genetic component. The disease is characterized by abnormalities in skin cells, the skin microbiome and the activity of the immune system. However, it is not fully understood how these disease-related changes are associated with genetic variation. Genome studies associate genetic variants with phenotypes. By combining datasets from GWAS and eQTL studies, it is possible to identify variants that are associated with the disease and also alter gene expression. The affected genes likely act in the development of atopic dermatitis and studying their signaling pathways helps elucidate the molecular mechanisms of the disease. In this thesis, datasets from two GWAS and 112 eQTL studies were colocalised. 148 genes were associated with the disease, from which expression data of genes in the interleukin-1 cytokine family were examined in more detail. The genes *IL1RL1*, *IL18*, *IL18R1*, *IL18RAP* were identified in multiple immune-system related cell types. The results suggest that proteins of the interleukin-1 family affect the development of atopic dermatitis. The results are the basis for further research into the contribution of specific

genes in the disease.

Keywords:

atopic dermatitis, colocalisation, GWAS, eQTL

CERCS: P175 Informatics, systems theory

B110 Bioinformatics, medical informatics, biomathematics, biometrics

Sisukord

1	Sissejuhatus	5
2	Taust	6
2.1	Atoopiline dermatiit	6
2.2	Genoomiülesed assotsiatsiooniuringud	6
2.3	Ekspressiooni kvantitatiivse tunnuse lookuste analüüs	7
2.4	Geneetiline kolokalisatsioon	8
3	Meetodid	9
3.1	Andmestikud	9
3.2	Kolokalisatsiooni töövoog HPC keskses	9
3.3	Funktsioon coloc.abf	10
3.4	Analüüs	11
4	Tulemused	12
4.1	Ülevaade tuvastatud geenidest	12
4.2	Interleukiin-1 perekonna geenid	12
5	Arutelu	18
6	Kokkuvõte	20
	Viidatud kirjandus	24
	Litsents	25

1 Sissejuhatus

Geneetikaalased uuringud on aidanud selgitada inimese genoomi mõju fenotüübile, kaasa arvatud selle rolli erinevate haiguste kujunemisel. Samas on paljude haiguste molekulaarsed mehhanismid siiani ebaselged. Selgitades haiguste kujunemist, on võimalik leida uusi reaktsiooniradasid, mida ravimitega mõjutada ja nii haiguste sümptomeid leevendada.

Atoopiline dermatiit on multifaktoriaalne haigus, mis põhjustab põletikulisi sügelevaid lööbeid nahal [1]. Haigus on väga heterogeenne ja selle tekkel mängivad rolli nii geneetilised, immunoloogilised kui keskkonnategurid [1]. Haigusega on seostatud mutatsioone geenid *FLG* ja geenid *IL4* ning *IL13* lähedalt [2, 3], kuid interleukiin-1 valkude perekonna gene on atoopilise dermatiidi kontekstis vähe uuritud.

Komplekshaigusi aitavad uurida genoomiüleused assotsiatsiooniuuringud (GWAS), näidates, millised geneetilised variandid on haigust põdevatel inimestel sagedamini esindatud kui tervetel inimestel. Kuna aga enamik tuvastatud variante asub genoomi mittekodeerivas alas [4] ega mõjuta seega otseselt valkude järjestusi, on ainult variantide põhjal keeruline teha järeldusi haiguse molekulaarsete mehhanismide kohta. Seega on kasulik GWAS andmestikke kombineerida näiteks geeniekspressiooni või molekulaarbioloogiliste uuringutega.

Kolokalisatsioon on geenandmete analüüsimeetod, mis toob eri andmestikest välja ühised signaalid. Kolokaliseerides GWAS ja ekspressiooni kvantitatiivse tunnuse lookuste (eQTL) analüüside andmestikud, leiab geneetilised variandid, mis on haigusega seostatud ja mõjutavad ka geeniekspressiooni. Seega näitab kolokalisatsioon, milliste geenide diferentsiaalne ekspressioon võib viia haiguse tekkeni.

Bakalaureusetöö eesmärk oli tuvastada uusi atoopilise dermatiidiga seotud gene ja leida, millistes rakutüüpides need ekspresseeritud on. Selleks kolokaliseeriti kahe genoomiülese assotsiatsiooniuuringu andmestikud 112-e eQTL analüüsi andmestikuga. Eraldi toodi välja interleukiin-1 perekonna valkude ekspressioonandmed. Tulemuste põhjal saab püstitada hüpoteese haiguse patofüsioloogia kohta ning viia läbi edasisi täpsustavaid uurimusi.

Töös kasutatud GWAS andmestikud pärinesid L. Paternoster jt tööst [5] ja UK Biobanki kogutud andmetest, mida analüüsis organisatsioon Neale Lab [6]. eQTL andmestikud pärinesid eQTL Catalogue projektist [7]. Kolokalisatsiooni läbiviimiseks kasutati Nurlan Kerimovi ja Kaur Alasoo töövoogu, mille lähtekood on saadaval GitHubis [8]. Töövoog käivitati Tartu Ülikooli Teadusarvutuste keskus (HPC keskus).

Tausta peatükis kirjeldatakse atoopilise dermatiidi olemust. Tutvustatakse GWAS ja eQTL uuringuid ning nende väärtust komplekshaiguste uurimisel. Lisaks antakse ülevaade geneetilisest kolokalisatsioonist ning varasematest töödest, kus meetodit on rakendatud atoopilise dermatiidiga seotud geenide tuvastamiseks. Meetodite peatükis kirjeldatakse kasutatud andmestikke, tarkvara ning töökäiku. Tulemuste ja järelduste peatükkides tuuakse välja analüüsis tuvastatud geenid, võrreldakse nende kattuvust seni tuntud geenidega ning pakutakse välja edasisi uurimissuundi.

2 Taust

Peatükis kirjeldatakse atoopilise dermatiidi kliinilisi tunnuseid, riskitegureid ning haigusega seostatud geene ja kudesid. Lisaks tutvustatakse genoomi uurimise meetodeid GWAS (ingl. *genome-wide association studies*) ja eQTL (ingl. *expression quantitative trait loci*) analüüs ning nende põhjal tehtud töid atoopilise dermatiidi uurimisel. Viimasena selgitatakse kolokalisatsiooni põhimõtet ning kuidas meetodit kasutada haiguste uurimiseks.

2.1 Atoopiline dermatiit

Atoopiline dermatiit on krooniline põletikuline nahahaigus, mis esineb arenenud riikides kuni 10% täiskasvanutest ning 20% lastest [1]. Haiguse peamiseks tunnuseks on allergilise reaktsiooni tulemusena tekkivad sügelevad lööbed nahal [1, 9]. Atoopilise dermatiidi patofüsioloogia on kompleksne ning haiguse tekkemehhanismid vajavad veel selgitamist. Atoopilise dermatiidi tekkel mängivad rolli nii geneetilised, immunoloogilised kui ka keskkonnategurid [1], kusjuures suurimaks riskiteguriks on haiguse esinemine suguvõsas [10]. Seega on indiviididevahelise geneetilise variatsiooni uurimine paljulubav viis haiguse molekulaarsete mehhanismide selgitamiseks.

Atoopilises dermatiidis on peamiseks muutusteks nahabarjääri kahjustused, naha mikrobioomi abnormaalsus ning immuunsüsteemi aktivatsioon [1]. Üheks olulisemaks nahabarjääri ebanormaalsuseks loetakse valgu filagriin puudumist, mis tuleneb mutatsioonidest *FLG* geenis. Nahabarjääri kahjustused võimaldavad allergeenidel infiltreeruda, mis põhjustab immuunrakkude (peamiselt Th1, Th2, Th17, Th22) aktivatsiooni. Th rakud (ingl. *T helper cell*) sekreteerivad signaalmolekule, mis aktiveerivad teisi immuunsüsteemi rakke, põhjustades immuunvastuse.

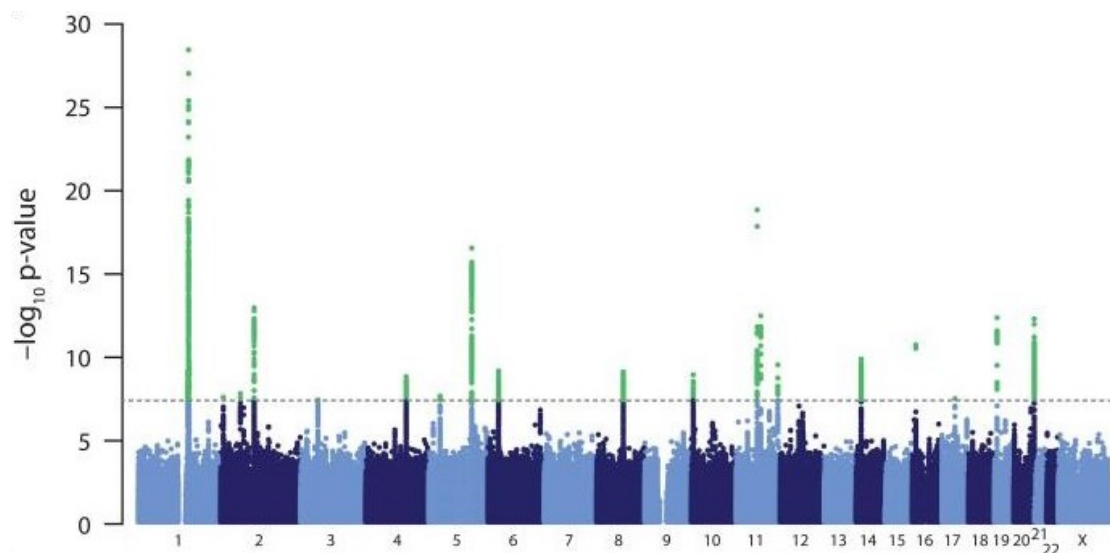
Järgnev lõik põhineb M. Akdis' e jt töö [11]. Interleukiinid on signaalmolekulid, mis reguleerivad valgete vererakkude tööd. Interleukiinidel on oluline roll ka atoopilise dermatiidi tekkel. Hiljuti on näidatud, et lisaks paljudele teistele tsütokiinidele võib atoopilise dermatiidi kujunemisse panustada ka interleukiin-1 valkude perekond. Interleukiin-1 perekonda kuulub 11 valku: IL-1 α , IL-1 β , IL-18, IL-33, IL-36 α , IL-36 β , IL-36 γ , IL-1Ra, IL-36Ra, IL-37, IL-38. Neist esimeset seitse soodustavad põletiku teket ning ülejäänud suruvad põletikku maha.

2.2 Genoomiülesed assotsiatsiooniuuringud

Genoomiülesed assotsiatsiooniuuringud seostavad geenivariandid fenotüübiliste tunnustega, kaasa arvatud haiguste esinemine. Tuvastatud variantide põhjal saab hinnata haiguste kujunemise riski tervetel inimestel ja pakkuda kandidaatgeene, mis võivad haiguste tekkesse panustada. Põhjalikult uuritud riskitegurid on filagriini kodeeriva geeni funktsioonikaoga mutatsioonid [12, 13, 14]. Täiendavalt on haigusega seostatud variante

epidermise rakkude jagunemisel ja diferentseerumisel osalevate geenide lähedalt, mis kinnitab epiteelbarjääri olulisust atoopilises dermatiidis [2]. Uuemad uuringud on lisaks tuvastanud variante immuunsüsteemi rakkudes ekspresseeritud geenide lähedalt, viidates immuunsüsteemi rollile haiguse kujunemisel [5]. Näiteks interleukiin-1 perekonna tsütokiinide ja nende retseptorite geenides esinevad variatsioonid võivad olla seotud atoopilise dermatiidiga [11].

Ainult GWAS uuringute alusel on aga keeruline teha järeldusi haiguste tekkemehhanismide kohta, kuna umbes 90% tuvastatud variantidest asuvad genoomi mittekodeerivas alas [4]. Seega ei mõjuta need otseselt valkude järjestust, vaid reguleerivad geeniekspressiooni läbi muude mehhanismide. Lisaks raskendab tulemuste tõlgendamist aheldatus. Analüüsis tuvastatud variant ei pruugi ise olla põhjuslik, kuid aheldatuse tõttu esineb sageli põhjusliku variandiga koos. Seda on näha L. Paternoster jt töö [5] tulemustest, mis on kujutatud joonisel 1. Seega on vaja täiendavaid uuringuid, mis aitavad genoomiüleste assotsiatsiooniuringute tulemusi mõtestada.



Joonis 1. Analüüsitud geneetiliste variantide p-väärtused [5, kohandatud]. Variandid moodustavad lookuseid, kus lähestikku asub palju variante, mille p-väärtused on madalad. Samas, kõik lookuses asuvad variandid pole tegelikult põhjuslikud.

2.3 Ekspressiooni kvantitatiivse tunnuse lookuste analüüs

eQTL analüüs tuvastab koos diferentsiaalselt ekspresseerunud geenid ja seostab need põhjuslike variantidega. On näidatud, et enamik sellistest variantidest asuvad mõjutavate geenide lähedal, kuid see järeldus võib tuleneda meetodi piirast [15]. eQTL

analüüs aitab seletada GWAS uuringutest leitud variantide mõju, näidates haigusega seostatud variandi kohta, milliste geenide ekspressiooni see sama variant mõjutab.

A. C. Nica ja E. T. Dermitzakis [15] kirjeldavad eQTL analüüsi tulemuste koespetsiifilisust. Korduvalt on näidatud, et eQTL analüüsi tuvastatud lookused erinevad rakutüüpide vahel. Peale selle võib üks ja seesama geenivariant seostuda koest sõltuvalt erinevate geenidega. Autorid kinnitavad, et eQTL analüüsides tuleb tähelepanu pöörata analüüsitud rakutüübile. Haigusi uurides on tähtis silmas pidada ekspressiooni koespetsiifilisust, et mitte teha valejärelusi haiguse patofüsioloogia kohta [15].

2.4 Geneetiline kolokalisatsioon

Kolokalisatsioon on andmeanalüüsimeetod, kus joondatakse eri andmestikud, et näha korraga avalduvaid signaale. C. Giambartolomei jt [16] kirjeldavad statistilist kolokalisatsioonimeetodikat, mis tuvastab kahest seostusuuringust (ingl. *association study*) ühised põhjuslikud variandid. Autorid töötasid välja Bayesi testi, mis hindab kolokalisatsiooni ja väljastab tulemused lihtsalt interpreteeritavate tõenäosustena. Testi kirjeldatakse täpsemalt bakalaureusetöö Meetodite peatükis.

GWAS ja eQTL analüüsi andmestikke kolokaliseerides leiab, millised haigusega seostatud variandid mõjutavad ka geeniekspressiooni. Teades diferentsiaalselt ekspresseerunud gene, saab lihtsamalt selgitada haiguste molekulaarseid mehhanisme. M. K. Sobczyk jt [17] viisid hiljuti läbi töö atoopilise dermatiidiga seotud geenide tuvastamiseks. Selleks valiti 25 haigusega seostatud GWAS lookust. Vaadati nende lookuste kattuvust geeniekspressiooni, valkude ning DNA metülatsiooni QTL andmestikega ja igas lookuses määrati kandidaatgeenid. Töötati välja ADGAPP (Atopic Dermatitis GWAS Annotation & Prioritisation Pipeline) töövoog, mis järjestas kandidaatgeenid arvutatud prioriteediskoori järgi, kusjuures üks analüüsi osa oli ka kolokalisatsioon. Enamik tuvastatud geenidest olid juba varasemast teada, kuid osade lookuste puhul eelistas töövoog gene, mida varem polnud nendes lookustes põhjuslikeks peetud. Seega endiselt tuvastatakse atoopilise dermatiidiga seotud uusi gene.

3 Meetodid

Peatükis tuuakse välja kasutatud andmestikud ja kirjeldatakse töövoogu. Täpsemalt selgitatakse töövoogu kolokalisatsiooni osa ning saadud tulemuste tõlgendamist. Lisaks kirjeldatakse, kuidas töövoogu käivitati ja milliseid tulemusi vaadati.

3.1 Andmestikud

Bakalaureusetöös kasutati L. Paternoster jt tööst [5] ning UKBiobankist [6] pärit GWAS andmestikke. Mõlemad andmestikud on kättesaadavad MRC IEU OpenGWAS andmebaasis [18]. L. Paternoster jt andmestikus on kokku 11059641 tuvastatud üksiku nukleotiidi polümorfismi (SNP). Analüüsitud haigete ja kontrollgrupi suurused olid vastavalt 10788 ning 30047 indiviidi. UKBiobanki andmestikus on kokku 9851867 tuvastatud SNP-d, mis pärinevad 11819-lt patsiendilt ja 451114-lt tervelt isikult. eQTL andmestikena kasutati eQTL Catalogue projektis [7] kogutud geeniekspressiooni andmestikke [7].

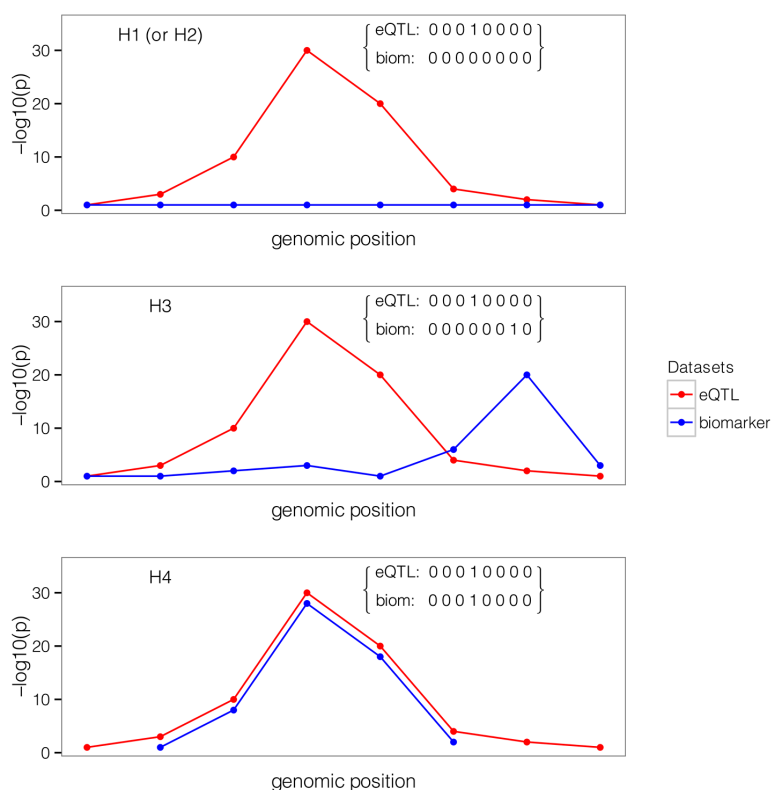
3.2 Kolokalisatsiooni töövoog HPC keskses

Bakalaureusetöös kasutati Nurlan Kerimovi ja Kaur Alasoo GWAS ja eQTL andmete kolokalisatsiooni töövoogu [8]. Kolokalisatsiooniks peavad eQTL andmestikud olema samas vormingus, mis eQTL Catalogue projektis esitletud andmestikud, ning GWAS andmestikud *Variant Call Format* (VCF) vormingus. Töövoog on ehitatud kasutades Nextflow keelt [19], mis vastutab töövoogu täitmise eest HPC keskses. Andmestike kolokaliseerimiseks kasutab töövoog programmeerimiskeeles R kirjutatud paketti *coloc*, mille lähtekood on avalik GitHubis [20].

Töövoos võetakse sisendiks GWAS ja eQTL andmestikud ning fail juhtpaaridega (*lead pairs*). Ühe juhtpaari moodustavad fenotüübiline tunnus ja geneetiline variant, millevahelisi seoseid uuritakse. Kui juhtpaaride asemel antakse sisendiks permutatsioonifailid, valitakse permutatsioonifailidest juhtpaarideks sellised kirjed, millel on piisavalt tugev seos geeni ekspressiooniga. Permutatsioonifailid saadi FastQTL tarkvaraga [21] indiviidide genotüüpe iga geeni jaoks 1000 korda permuteerides ja saadud andmete järgi p-väärtused ümber hinnates. GWAS andmestikud peavad olema versioonis GRCh37 ning töövoog teisendab need versiooni GRCh38, kasutades tarkvara CrossMap [22]. Edasi jaotatakse juhtpaarid rühmadesse (*batches*), mida käsitletakse rööbiti. Kõigis rühmades võetakse GWAS ja eQTL andmestikest alamhulga, mis sisaldavad ainult juhtpaari variandiga samasse genoomi regiooni (*window*) jäävaid variante. GWAS ja eQTL alamhulkade peal viiakse läbi kolokalisatsioon, kasutades *coloc* paketi funktsiooni *coloc.abf*.

3.3 Funktsioon coloc.abf

Funktsioon coloc.abf viib andmestikel läbi Bayesi analüüsi, milles arvutatakse tinglikud tõenäosused (ingl. *posterior probability*) variandi konfiguratsiooni hüpoteesidele [16]. Võimalikest hüpoteesidest neli on toodud joonisel 2. Lisaks on veel hüpotees H_0 , mille puhul ei esine signaali kummaski andmestikus.



Joonis 2. Võimalikud hüpoteesid kolokalisatsiooni Bayesi testis [16]. Graafikud kujutavad olukordi, millele iga hüpotees vastab. Hüpotees H_4 kirjeldab olukorda, kus mõlemas andmestikus annab tugeva signaali üks ja seesama variant. Hüpoteesi H_3 korral on põhjuslikud variandid andmestikes erinevad. Hüpoteeside H_1 ja H_2 puhul on signaal ainult ühes andmestikus kahest.

Funktsioon arvutab iga juhtpaari geneetilise variandi jaoks tinglikud tõenäosused kõigi hüpoteeside toetuseks. Näiteks, suur tõenäosus hüpoteesi H_3 kohta viitab, et andmestikes põhjustavad signaale erinevad variandid. Kolokaliseerunud signaalide korral on tõenäosus suur hüpoteesi H_4 jaoks. Bakalaureusetöös loeti signaalid kolokaliseerunuks, kui tõenäosus hüpoteesi H_4 toetuseks oli suurem kui 0,8.

3.4 Analüüs

Bakalaureusetöös kasutatud töövoog käivitati Tartu Ülikooli Teadusarvutuste keskus (HPC keskus). Selleks anti autorile ligipääs keskuse arvutitele. Kasutatud GWAS andmestikud laeti alla OpenGWAS andmebaasist [18], mis sisaldab VCF-failideks konvertitud andmestikke. Andmestikud olid ka töövoole sobivas versioonis: GRCh37. eQTL andmestikud koos permutatsioonifailidega olid varasemast HPC keskusel olemas.

GitHubist laeti HPC keskuse arvutisse töövoog koodi kaust [8]. Esimesel korral käivitati töövoog juhtpaaridega faili asemel permutatsioonifailidega. Selleks määrati Nextflow konfiguratsioonifailis failiteed kasutatavate andmestike jaoks ning muudeti vastav argument: `use_permutation = true`. Järgnevalt käivitati töövoog vastavalt juhistele koodi repositooriumis [8].

Tulemusfailides jäeti alles vaid kirjed, mille puhul tinglik tõenäosus hüpoteesi H_4 toetuseks (PP H_4) oli üle 0,8. Alles jäänud variandid jagati 500000 aluspaari suurusesse lookustesse, kus iga lookuse esindajaks (juhtvariant) valiti variant, mille PP H_4 oli lookuses suurim. Edasi koostati juhtpaaride fail nii, et iga juhtvariandi kohta võeti geenide metaandmete tabelitest [23] kõik geenid, mis asusid juhtvariandist 1000000 aluspaari kaugusel. Kaugus arvutati metaandmete tabelis tulba `phenotype_pos` järgi. Ühe juhtpaari moodustasid üks juhtvariant ja üks geen. Kõik juhtpaarid tõsteti kokku üheks juhtpaaride failiks. Järgmiseks kustutati juhtpaaridest kõik kirjed, mille juhtvariant asus kromosoomi 6 positsioonide vahemikus 30000000 - 39999999. Selles piirkonnas on palju aheldatustasakaalutust (ingl. *linkage disequilibrium*) ning kolokalisatsiooni tulemused pole usaldusväärsed.

Teisel korral käivitati töövoog permutatsioonifailide asemel juhtpaaride failiga. Konfiguratsioonifailis määrati ka vastav argument: `use_permutation = false`. Juhtpaaride kasutamine permutatsioonifailide asemel võimaldas tulemustes võrrelda geeni kolokaliseerumise tugevust eri rakutüüpide vahel. Saadud tulemustest eemaldati veel sellised variandid, mille puhul polnud üheski eQTL andmestikus PP H_4 üle 0,8.

4 Tulemused

Peatükis vaadatakse analüüsi tulemusi. Esiteks võetakse kokku kolokaliseerunud geenid. Seejärel vaadatakse tuvastatud interleukiin-1 perekonna geene lähemalt, tuues igauhe kohta välja kolokalisatsiooni tugevuse erinevates kudedes.

4.1 Ülevaade tuvastatud geenidest

Näidatakse tulemusi vaid enda defineeritud juhtpaaridega sooritatud kolokalisatsioonist. Pärast töövoost saadud tulemuste filtreerimist jäi L. Paternosteri jt [5] GWAS-i peal sooritatud kolokalisatsiooni tulemustest järele 105 geeni, UK Biobanki GWAS-ist [6] 66 geeni. Ühiseid geene kahe GWAS-i vahel oli 23. Suurt erinevust kummagi GWAS-i põhjal tuvastatud geenides põhjendab GWAS-ide väike statistiline võimsus [24]. Mõlemas uuringus kaasati vaid umbes 10000 haigust põdevat patsienti. Seega mõned geneetilised variandid ei saavuta kolokaliseerumiseks piisavalt suurt kaalu. Tuvastatud geenid ja jaotus on toodud tabelis 1. Lisaks näidatakse joonisel 7, mitu kolokaliseerunud geeni igas eQTL andmestikus tuvastati.

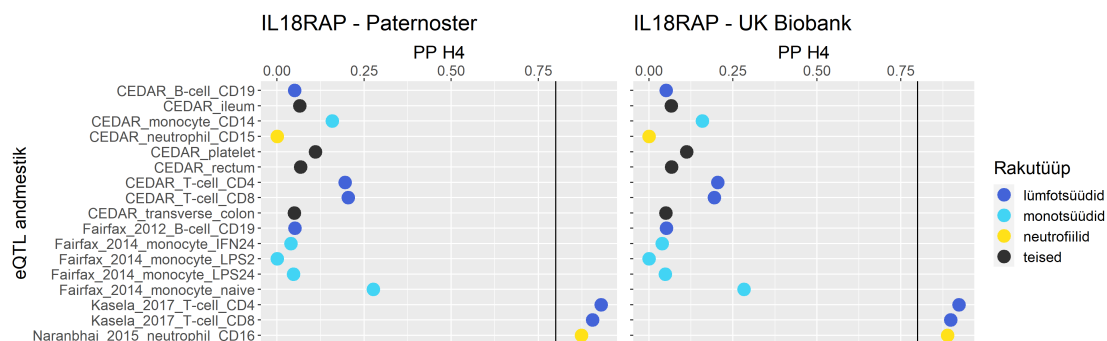
Kolokaliseerusid mitmed immuunsüsteemiga seotud geenid, näiteks *CD86* (dendriitrakkude aktivatsiooni määrav marker), *STAT3* (transkriptsioonifaktor, mis vastutab Th17 tüüpi T-rakkude diferentseerumise eest), *TRAF3* (vajalik immuunvastuse tekki-miseks paljudes erinevates rakutüüpides) ning tsütokiinide või nende retseptorite geenid. Interleukiin-1 perekonna geenidest tuvastati *IL18* ise ning lisaks kaks interleukiin-18 retseptori geeni: *IL18R1*, *IL18RAP*. Peale nende tuli analüüsist välja geen *ILIRL1*. Enim tuvastati geene vere, lümfotsüütide ja monotsüütide andmestikes, kuid lähemalt vaadatud interleukiin-1 perekonna geenid kolokaliseerusid mõnel juhul ka kudedes, mida pole atoopilise dermatiidi jaoks oluliseks peetud.

4.2 Interleukiin-1 perekonna geenid

Analüüsi tulemustest vaadati lähemalt interleukiin-1 perekonna tsütokiinide ja nende retseptorite geene. Interleukiin-1 perekonna valgugeenidest tuli kolokalisatsioonis välja ainult geen *IL18* ja vaid UK Biobanki GWAS-ist. Sellele geenile lisaks kolokaliseerusid interleukiin-1 perekonna retseptorite geenid *ILIRL1*, *IL18R1*, *IL18RAP*. Joonised 3, 4, 5, 6 näitavad kolokalisatsiooni tulemusi loetletud geenide kohta. Geen loetakse kolokaliseerunuks, kui tinglik tõenäosus hüpoteesi H_4 jaoks (PP H_4) on üle 0,8. Geenide, mis kolokaliseerusid mõlemas GWAS-is, puhul on näha, et analüüsis arvatud tinglikud tõenäosused erinevad genoomiüleste assotsiatsiooniuringute vahel väga vähe. Peale interleukiin-1 perekonna geenide tuleb analüüsist välja geen *IL4* (joonistel pole kujutatud), mille produkt osaleb samuti atoopilise dermatiidi tekkel [25].

Ainult Paternoster				Kattuvad	Ainult UK Biobank	
ABCA3	DCTN5	MRPS21	SOX15	AC007278.1	ACSL6	OPRL1
AC006252.1	DDX6	MTMR11	STAT3	AC007278.2	AL118506.1	P4HA2
AC010266.2	DHX58	MXD3	STK4	AL121845.3	AL356417.1	P4HA2-AS1
AC012467.2	DNAJC27-AS1	MYRF	TARS2	AP001107.8	AL589765.1	PCSK7
AC016876.1	ECM1	NCF4-AS1	TARSL2	AP5B1	AP3D1	PDCD1
AC138819.1	EFEMP2	NDFIP1	TEF	IL18R1	ARNT	PIP5K1A
AES	EIF1AD	NEPRO	TM2D3	IL18RAP	ATP8B2	POLR2K
AL121845.4	FADS1	NINJ2	TMEM258	IL1RL1	BCO2	PRKCQ
AL135818.1	FADS2	P2RY14	TOMM34	IL6R	CCDC18	PRKCQ-AS1
AL356218.1	FADS3	PICART1	UBAP1	KIAA0391	CD200R1	PSMC3IP
AL592430.1	FALEC	PLK1	ULBP3	LIME1	CFL2	PSMD4
AMN	FCGR1A	PUS10	WNK1	MIR3936HG	CTSK	RNF19A
ANK3	FEZ2	PVALB	ZGPAT	NCF4	DR1	SENP7
ANK3-DT	GLS	RAB5C		OVOL1	FAM177A1	SHE
AP001107.4	GPR171	REL		OVOL1-AS1	GMEB2	TAGLN
AQP3	GTPBP8	RGS14		PPP2R3C	HORMAD1	TDRKH-AS1
ARFRP1	ITIH3	RIN1		RAD50	IL18	TMED5
CD200R1L	KLC2	RPRD2		RTEL1-TNFRSF6B	IL4	TMEM181
CD200R1L-AS1	LINC00200	SAMD3		SLC22A4	ITGB8	WTAP
CD86	LINC02207	SART1		SLC22A5	LINC01238	Z82188.2
CDK1	LMAN2	SCYL1		STMN3	LINC02648	
COG4	MATN4	SLC2A4RG		TH2LCRR	LINGO4	
CRAT	MDM1	SNRNP35		TRAF3	MACC1	

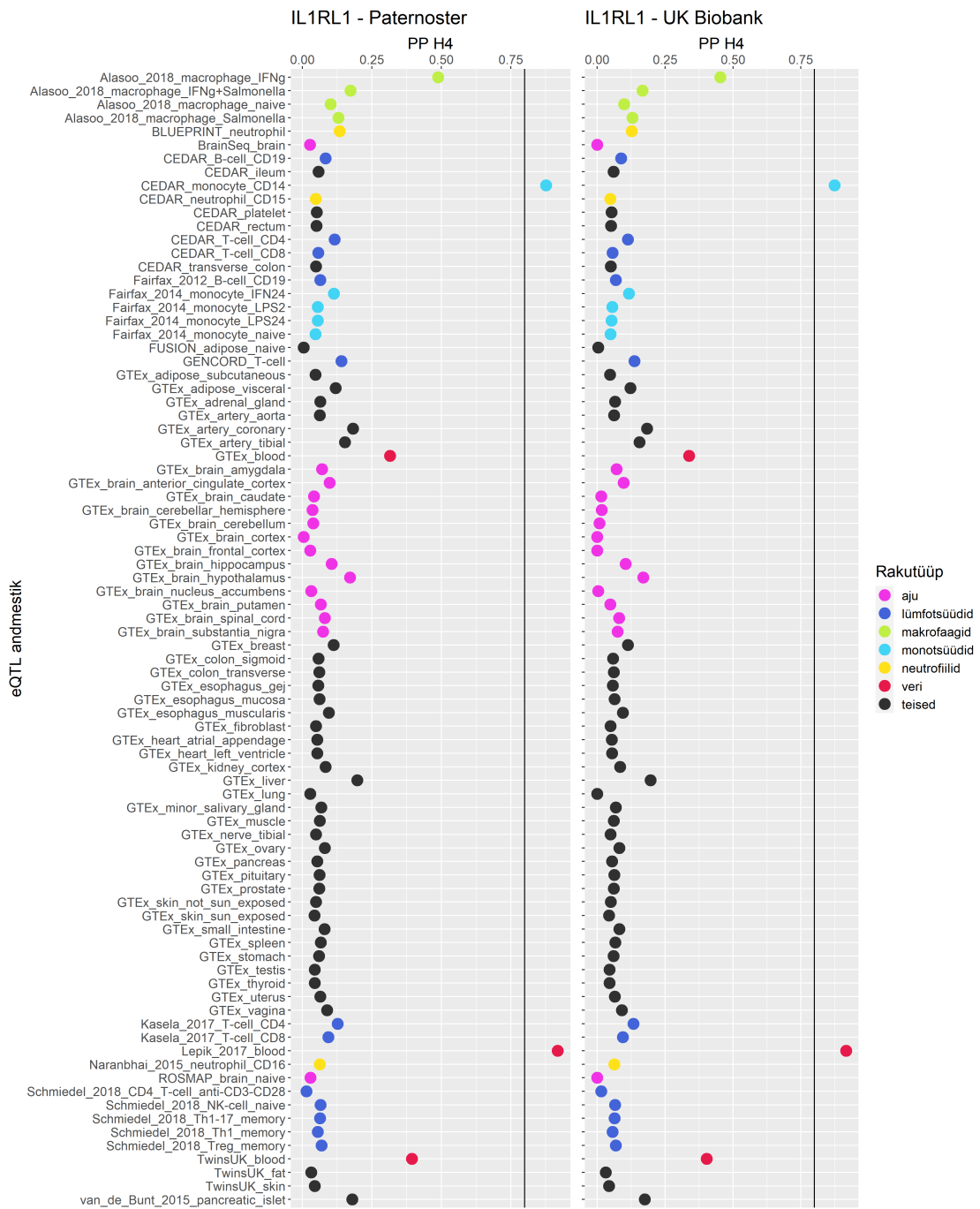
Tabel 1. Kolokalisatsioonid tuvastatud geenid pärast tulemuste filtreerimist. Näidatud on, millised geenid tuvastati kummagi GWAS andmestiku põhjal. Kattuvad geenid tulid välja mõlemast GWAS-ist.



Joonis 3. Geeni *IL18RAP* kolokalisatsiooni tulemused eQTL andmestike lõikes kummagi GWAS-i kohta. Geen kolokaliseerus lümfootsüütides ja ühes neutrofiilide andmestikus, kusjuures selle geeni ekspressiooniandmed esinesid ainult ekspressioonikiibi *microarray* meetodil kogutud eQTL andmestikes.



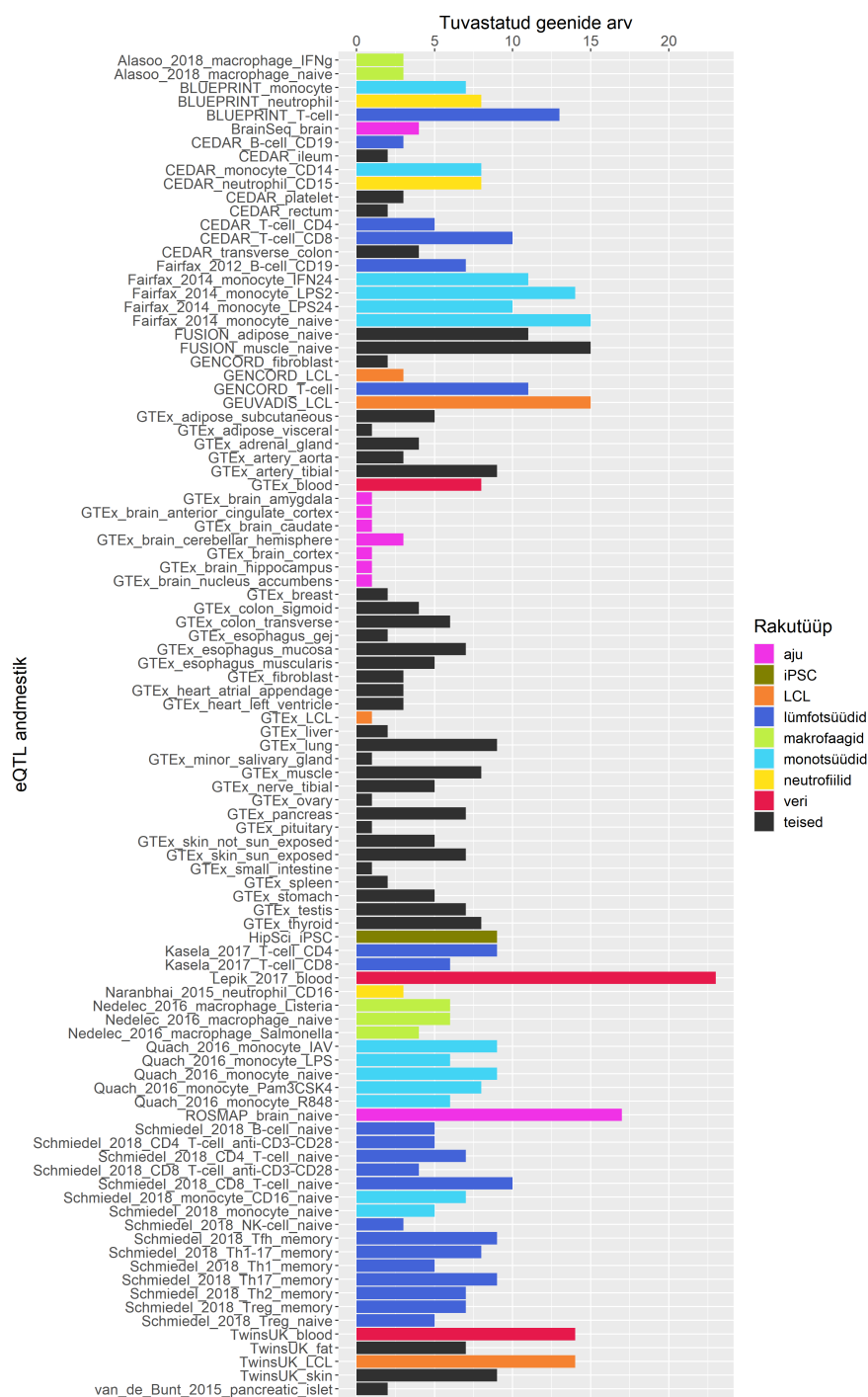
Joonis 4. Geeni *IL18R1* kolokalisatsiooni tulemused eQTL andmestike lõikes kummagi GWAS-i kohta. Geen kolokaliseerus lihaskoes, kopsus ja lümfotsüütides. Seatud 80% lävendi lähedale tuleb geen veel ühes vere andmestikus.



Joonis 5. Geeni *IL1RL1* kolokalisatsiooni tulemused eQTL andmestike lõikes kummagi GWAS-i kohta. Geen kolokaliseerus veres ja monotsüütides.



Joonis 6. Geeni *IL18* kolokalisatsiooni tulemused eQTL andmestike lõikes. Geen tuli välja vaid UK Biobanki GWAS-ist ja kolokaliseerus veres, monotsüütides, neutrofiilides ja pankreases.



Joonis 7. Kolokaliseerunud geenide arv igas eQTL andmestikus. Geenid loeti mõlema GWAS-i peale kokku, kusjuures kattuvad geenid (vt tabel 1) loeti ühe korra. Rakutüübid, kus kolokaliseerus pidevalt rohkem gene, on veri, lümfotsüüdid ja monotsüüdid. See on oodatud tulemus, kuna atoopilise dermatiidi tekkel on oluline roll immuunrakkudel [1]

5 Arutelu

Atoopilise dermatiidi üheks põhjustajaks on muutused immuunsüsteemi rakkude geeniekspressioonis. Analüüs tuvastas mitmel juhul, et haiguse puhul on mõjutatud interleukiini retseptori ekspressioon, mitte tsütokiin ise. See viitab olukorrale, kus haiguse tekkel on oluline retseptorite hulk rakus. Näiteks proinflammatoorsete tsütokiinide retseptorite ülekspressioon tekitab kergemini põletikulise reaktsiooni, nagu ka anti-inflammatoorsete retseptorite alaekspressioon.

M. K. Sobczyk jt [17] viisid hiljuti läbi töö atoopilise dermatiidiga seotud geenide tuvastamiseks. Autorid valisid L. Paternoster jt [5] GWAS-ist 25 haigusega seostatud lookust ja vaatasid nende kattuvust erinevate QTL andmestike lookustega, et leida põhjuslikke kandidaatgeene. Bakalaureusetöös kasutati analüüsil lisaks sellele samale GWAS andmestikule veel UK Biobanki [6] GWAS-i. Täiendavalt analüüsiti bakalaureusetöös rohkem erinevaid kudesid ja rakutüüpe.

M. K. Sobczyk jt [17] töös ja bakalaureusetöös tuli välja samu immuunsüsteemi rakkudes ekspresseeritud geene, näiteks *TRAF3*, *STAT3*. Interleukiinidest tuvastati mõlemas töös geenid *IL1RL1*, *IL6R*, *IL18R1*, *IL18RAP*. M. K. Sobczyk jt identifitseerisid veel geenid *IL2RA*, *IL7R*, *IL22*, mida bakalaureusetöö tulemused ei näita. Samas, bakalaureusetöös tuvastati ka geenid *IL4*, mida on varem haigusega seostatud [1], ning *IL18*. Viimane tuli välja just tänu UK Biobanki GWAS-i kaasamisele ja pole sageli atoopilise dermatiidi uurimisel esimene huviobjekt. Küll aga on teada selle tsütokiini põletikku tekitav efekt [26]. M. K. Sobczyk jt [17] töö ning bakalaureusetöö näitavad, et uuringute kombineerimise teel atoopilise dermatiidiga seotud geene otsides sõltuvad tulemused kasutatud andmestikest ja analüüsimeetodist. See tähendab, et endiselt tuvastatakse uusi haigusega seotud geene, mis võivad aidata paremini mõista atoopilise dermatiidi molekulaarseid mehhanisme.

Kuna atoopilise dermatiidiga kaasnevad sageli ka teised allergilised haigused, näiteks astma [1], pole bakalaureusetöös tuvastatud geenide kolokaliseerumine kopsus üllatav. Küll aga on tulemustest näha interleukiinide kolokaliseerumist ka lihaskoes ja pankreases, millel ei paista olevat otsest seost haigusega. Kolokaliseerumine nendes kudedes võibki olla juhuslik, sest osad eQTL-id on aktiivsed paljudes rakutüüpides, kuid haiguse riski mõjutavad vaid mõnes üksikus [15]. Samas on ka võimalik, et arusaam atoopilisest dermatiidist ei ole veel piisav nende kudede selgitamiseks. Ilma täiendavate uuringuteta pole võimalik kahel olukorral vahet teha.

Edasi saab uurida juba üksikute tuvastatud geenide signaaliradu, et haiguse tekkemehhanisme paremini kaardistada. Bakalaureusetöös sooritatud kolokalisatsiooni võib ka laiendada, kaasates splaissimise ja konkreetsete transkriptide ekspressiooni andmestikke. Täiendavalt peaks vaatama, mis suunas analüüsitud geneetilised variandid geeniekspressiooni mõjutavad: kas polümorfism põhjustab valgu hulga vähenemist või suurenemist. Lisaks on kasulik võrrelda selliste analüüsitud tulemusi erinevate allergiliste haiguste vahel (näiteks atoopiline dermatiit, allergiline riniit, astma), kuna need haigused esinevad

patsientidel sageli koos ning jagavad molekulaarseid mehhanisme.

6 Kokkuvõte

Atoopiline dermatiit on põletikuline nahahaigus, mida põhjustavad mitmed faktorid. Haiguse tunnusteks on nahabarjääri kahjustused, muutused naha mikroobioomis ning immuunsüsteemi aktivatsioon, kuid täpsed tekkemehhanismid vajavad veel selgitamist. Saadaval on suur hulk bioloogilisi andmeid, mida analüüsid on võimalik valgustada komplekshaiguste tagamaid ja nii leida uusi reaktsiooniradasid, mida ravimitega mõjutada.

Bakalaureusetöö eesmärk oli tuvastada uusi atoopilise dermatiidiga seotud gene. Selleks kolokaliseeriti kahe genoomiülese assotsiatsiooniuuringu andmestikud ekspressiooni kvantitatiivse tunnuse lookuste andmestikega. Analüüsiks kasutati olemasolevat kolokalisatsiooni töövoogu, mis käivitati Tartu Ülikooli Teadusarvutuste keskus.

Kokku tuvastati 148 unikaalset geeni. Enim kolokaliseerus gene veres, lümfotsüütides ja monotsüütides. Täpsemalt uuriti interleukiin-1 tsütokiinide perekonna ekspressiooniandmeid. Identifitseeriti üks interleukiini-1 perekonna geen: *IL18*. Vastaval tsütokiinil on küll põletikku tekitav mõju, kuid seostusuuringud seda geeni tihti atoopilise dermatiidiga ei seosta. Täiendavalt tulid analüüsis välja interleukiin-1 perekonna retseptorite geenid *IL1RL1*, *IL18R1*, *IL18RAP*. Lähemalt vaadatud geenid kolokaliseerusid lisaks oodatud kudedele ka näiteks pankreases ja lihaskoes, mis pole küll esmapilgul atoopilise dermatiidi jaoks olulised.

Edasistes uurimustes võib analüüsi kaasata ka kolmandat tüüpi andmestikke, et veel paremini haigusega seostatud geneetiliste variantide mõju selgitada. Lisaks saab uurida bakalaureusetöös tuvastatud geenide signaaliradasid atoopilise dermatiidi tekkemehhanismide valgustamiseks. Haiguse taust on kirju ning bioinformaatika tööriistad aitavad tuvastada uusi uurimisobjekte.

Viidatud kirjandus

- [1] Langan SM, Irvine AD, Weidinger S. Atopic dermatitis. *The Lancet*. 2020 Aug;396(10247):345–360. Available from: <http://www.sciencedirect.com/science/article/pii/S0140673620312861>.
- [2] Paternoster L, Standl M, Chen CM, Ramasamy A, Bønnelykke K, Duijts L, et al. META-ANALYSIS OF GENOME-WIDE ASSOCIATION STUDIES IDENTIFIES THREE NEW RISK LOCI FOR ATOPIC DERMATITIS. *Nature genetics*. 2011 Dec;44(2):187–192. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3272375/>.
- [3] Weidinger S, Willis-Owen SAG, Kamatani Y, Baurecht H, Morar N, Liang L, et al. A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. *Human Molecular Genetics*. 2013 Dec;22(23):4841–4856.
- [4] Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 2014 Oct;1842(10):1910–1922. Available from: <https://www.sciencedirect.com/science/article/pii/S0925443914000714>.
- [5] Paternoster L, Standl M, Waage J, Baurecht H, Hotze M, Strachan DP, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nature Genetics*. 2015 Dec;47(12):1449–1456.
- [6] Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank;. Available from: <http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank>.
- [7] Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, et al. eQTL Catalogue: a compendium of uniformly processed human gene expression and splicing QTLs. *bioRxiv*. 2021 Jan:2020.01.29.924266. Publisher: Cold Spring Harbor Laboratory Section: New Results. Available from: <https://www.biorxiv.org/content/10.1101/2020.01.29.924266v2>.
- [8] Kerimov N. eQTL-Catalogue/colocalisation. eQTL Catalogue; 2021. Original-date: 2020-09-29T10:03:26Z. Available from: <https://github.com/eQTL-Catalogue/colocalisation>.

- [9] Guttman-Yassky E, Krueger JG. Atopic dermatitis and psoriasis: two different immune diseases or one spectrum? *Current Opinion in Immunology*. 2017 Oct;48:68–73. Available from: <https://www.sciencedirect.com/science/article/pii/S0952791517301061>.
- [10] Apfelbacher CJ, Diepgen TL, Schmitt J. Determinants of eczema: population-based cross-sectional study in Germany. *Allergy*. 2011;66(2):206–213. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1398-9995.2010.02464.x>. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1398-9995.2010.02464.x>.
- [11] Akdis M, Aab A, Altunbulakli C, Azkur K, Costa RA, Cramer R, et al. Interleukins (from IL-1 to IL-38), interferons, transforming growth factor β , and TNF- α : Receptors, functions, and roles in diseases. *Journal of Allergy and Clinical Immunology*. 2016 Oct;138(4):984–1010. Available from: <https://www.sciencedirect.com/science/article/pii/S0091674916307151>.
- [12] Palmer CNA, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nature Genetics*. 2006 Apr;38(4):441–446. Number: 4 Publisher: Nature Publishing Group. Available from: <https://www.nature.com/articles/ng1767>.
- [13] Rodríguez E, Baurecht H, Herberich E, Wagenpfeil S, Brown SJ, Cordell HJ, et al. Meta-analysis of filaggrin polymorphisms in eczema and asthma: Robust risk factors in atopic disease. *Journal of Allergy and Clinical Immunology*. 2009 Jun;123(6):1361–1370.e7. Available from: <https://www.sciencedirect.com/science/article/pii/S0091674909005247>.
- [14] Irvine AD, McLean WHI, Leung DYM. Filaggrin Mutations Associated with Skin and Allergic Diseases [review-article]; 2011. Archive Location: world Publisher: Massachusetts Medical Society. Available from: <https://www.nejm.org/doi/10.1056/NEJMra1011040>.
- [15] Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2013 Jun;368(1620):20120362. Publisher: Royal Society. Available from: <https://royalsocietypublishing.org/doi/10.1098/rstb.2012.0362>.
- [16] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics*. 2014 May;10(5):e1004383.

Publisher: Public Library of Science. Available from: <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004383>.

- [17] Sobczyk MK, Richardson TG, Zuber V, Min JL, Gaunt TR, Paternoster L. Triangulating molecular evidence to prioritize candidate causal genes at established atopic dermatitis loci. *Journal of Investigative Dermatology*. 2021 Apr. Available from: <https://www.sciencedirect.com/science/article/pii/S0022202X2101160X>.
- [18] Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv*. 2020 Aug:2020.08.10.244293. Publisher: Cold Spring Harbor Laboratory Section: New Results. Available from: <https://www.biorxiv.org/content/10.1101/2020.08.10.244293v1>.
- [19] Tommaso PD, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nature Biotechnology*. 2017 Apr;35(4):316–319. Number: 4 Publisher: Nature Publishing Group. Available from: <https://www.nature.com/articles/nbt.3820>.
- [20] Wallace C. chr1swallace/coloc; 2021. Original-date: 2013-03-22T16:49:20Z. Available from: <https://github.com/chr1swallace/coloc>.
- [21] Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics (Oxford, England)*. 2016 May;32(10):1479–1485.
- [22] Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics*. 2014 Apr;30(7):1006–1007. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3967108/>.
- [23] Alasoo K. Metadata for various molecular traits included in the eQTL Catalogue. Zenodo; 2019. Type: dataset. Available from: <https://zenodo.org/record/3366011#.YJf6yLUzbBU>.
- [24] Caballero A, Tenesa A, Keightley PD. The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses. *Genetics*. 2015 Dec;201(4):1601–1613. Available from: <https://doi.org/10.1534/genetics.115.177220>.
- [25] Danso MO, van Drongelen V, Mulder A, van Esch J, Scott H, van Smeden J, et al. TNF- α and Th2 Cytokines Induce Atopic Dermatitis-Like Features on Epidermal Differentiation Proteins and Stratum Corneum Lipids in Human

Skin Equivalents. *Journal of Investigative Dermatology*. 2014 Jul;134(7):1941–1950. Available from: <https://www.sciencedirect.com/science/article/pii/S0022202X15368731>.

- [26] Lee JH, Cho DH, Park HJ. IL-18 and Cutaneous Inflammatory Diseases. *International Journal of Molecular Sciences*. 2015 Dec;16(12):29357–29369. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute. Available from: <https://www.mdpi.com/1422-0067/16/12/26172>.

Litsents

Lihlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Peep Kolberg**,
(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihlitsentsi) minu loodud teose
Atoopilise dermatiidiga seotud geenide tuvastamine geneetilise kolokalisatsiooni abil,
(lõputöö pealkiri)
mille juhendaja(d) on Kaur Alasoo ja Ana Rebane,
(juhendaja nimi)
reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Peep Kolberg
07.05.2021