

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Ralf Tambets

**Beetainterferooni signaaliraja aktiivsust
mõjutavate geneetiliste variantide
tuvastamine**

Bakalaureusetöö (9 EAP)

Juhendaja: PhD Kaur Alasoo

Tartu 2021

Beetainterferooni signaaliraja aktiivsust mõjutavate geneetiliste variantide tuvastamine

Lühikokkuvõte:

Beetainterferoon on inimorganismi põletikuvastuses tähtsal kohal olev biomolekul. See on osa laiaast põletikuvastasest geenivõrgustikust, mille toimemehhanism ei ole veel lõplikult selge. Käesolevas töös püüti nelja populatsiooni andmetele tuginedes ülegenoomsete assotsiatsioonuringute abil tuvastada selle signaaliraja aktiivsust mõjutavaid seni välja selgitamata geneetilisi variante. Selleks käsitleti beetainterferooni kodeerivas *IFNBI* lookuses olevaid geneetilisi variante *trans*-eQTLidena ja selgitati välja nendega kõige tugevamas seoses olev geen. Leitud geeni beetainterferooni signaaliraja markergeeninena kasutades üritati tuvastada täiendavaid signaaliraja aktiivsust mõjutavaid *trans*-eQTLi väljaspool *IFNBI* lookust. *IFNBI* lookuse seos *IFI6* geeniga õnnestus replitseerida nii Fairfax_2014 andmestikus, kus see esmakordselt tuvastati, kui ka ülejäänud kolmel populatsioonil. Lõpliku valimi väiksuse ($n=685$) tõttu uusi seoseid tuvastada ei õnnestunud.

Võtmesõnad: beetainterferoon, *trans*-eQTL, geeniekspressioon

CERCS: B110 - Bioinformaatika, meditsiiniinformaatika, biomatematika, biomeetrika

Determining genetic variants modifying the activity of the beta interferon signaling pathway

Abstract:

Interferon beta is a biomolecule integral to the inflammatory response of the human organism. It is known to be a part of a wide anti-inflammatory gene network, the mechanism of action of which is not yet entirely known. The aim of this thesis was to use the data regarding four human populations for genome-wide association studies in order to determine previously unknown genetic variants that play a part in interferon beta's signaling pathway. To achieve this, variants in the *IFNBI* locus, which codes for interferon beta, were treated as *trans*-eQTLs and the gene most strongly associated with them was found. By using that gene as a marker gene for the beta interferon signaling pathway, additional *trans*-eQTLs modifying the activity of the beta interferon signaling pathway were sought for. The association between the *IFNBI* locus and the *IFI6* gene was replicated in the Fairfax_2014 data-

set, where it was first discovered, and in three other populations. Due to a relatively low sample size (n=685), no novel associations were found.

Keywords: interferon beta, *trans*-eQTL, gene expression

CERCS: B110 - Bioinformatics, medical informatics, biomathematics, biometrics

Sisukord

1.	Sissejuhatus	6
2.	Mõisted ja terminid	8
3.	Kirjanduse ülevaade	10
3.1	DNA struktuur	10
3.2	Geeniekspressioon	10
3.3	Geneetilise mitmekesisuse alused	11
3.4	Genoomi sekveneerimine	13
3.5	GWAS ehk ülegenoomne assotsiatsiooniuring	13
3.6	eQTL olemus ja nende tuvastamine	14
3.7	Beetainterferoon ja selle roll geeniekspressioonis	15
4.	Kasutatud tarkvara	17
4.1	R	17
4.1.1	MatrixEQTL.....	17
4.1.2	SNPrelate.....	17
4.1.3	dplyr	17
4.1.4	ggplot2.....	17
4.1.5	qqman.....	17
4.2	BCFtools.....	18
4.3	METAL	18
4.4	PuTTY	18
4.5	FileZilla	18
5.	Lahenduskäigu kirjeldus	19
5.1	Uuritavad andmestikud.....	19
5.2	Lahenduskäigu põhiidee	19
5.3	Beetainterferooni vastuse markergeeni otsimine Fairfax_2014 andmestikust	20

5.4	Otsingu kordamine Fairfax_2018 ja Quach_2016 andmestikel	21
5.5	Analüüsi tulemuste usaldusväarsuse hindamine	22
5.6	Quach_2016 andmestiku täiendav töötlus.....	24
5.7	Metaanalüüsi läbiviimine nelja populatsiooni andmetel	25
6.	Tulemused	26
7.	Kokkuvõte	29
	Tänuavaldused.....	30
	Viidatud kirjandus	31
	Lisad	34
I.	<i>IFI6</i> tugevaimad assotsiatsioonid andmestike kaupa	34
II.	<i>IFI6</i> tugevaimad assotsiatsioonid metaanalüüsil.....	38
III.	Litsents	41

1. Sissejuhatus

Kõigi päristuumsete organismide elutegevuseks vajalike bioloogiliste protsesside toimimiseks tarvilik informatsioon on salvestatud rakutuumas ja mitokondrites paiknevatesse DNA molekulidesse. Organismi kokkupuude mutageensete välisteguritega või vead informatsiooni lugemisel võivad viia muutusteni DNA ahelas ja geneetilise info teisenemiseni, kusjuures võib piisata juba ühe nukleotiidi muutusest, et geeni poolt kodeeritava biomolekuli struktuuri või funktsiooni muuta [1,2,3]. Info teisenemine sugurakkudes võib viia muutuse pärandumiseni järglastele ja selle juurdumiseni populatsiooni genofondis [4].

Geeniekspressiooniks nimetatakse geeni baasil valgu sünteesimise protsessi, mis kätkeb endas kahte suurt alamprotsessi – DNA abil RNA tootmist ehk transkriptsiooni ja RNA baasil valgu sünteesimist ehk translatsiooni [3,5]. Geeniekspressiooni aktiivsust hinnatakse geeni lõpp-produkti kontsentratsiooni kaudu [6]. Erinevate indiviidide genoomi ja geeniekspressiooni uurides on meetodite nagu ülegenoomse assotsiatsiooniuringu abil võimalik kindlaks teha, millised mutatsioonid ja polümorfismid valgusünteesi aktiveerivad või pärsvivad, ja jõuda mitmete patoloogiliste protsesside tekkepõhjusteni [6,7,8].

Kindlaid DNA regioone, mis mõjutavad ühe või mitme geeni produktide hulka, nimetatakse eQTLideks (geeniekspressiooni kvantitatiivse tunnuse lookusteks). Need jaotatakse mõjutatava geeni asukoha alusel kaheks - *cis*-eQTLid paiknevad sellega samas kromosoomis, *trans*-eQTLid avaldavad mõju ka teise homologse kromosoomi geenidele [7]. Käesoleva töö eesmärgiks on lipopolüsahhariidiga mõjutatud immuunrakkudes toimuva geeniekspressiooni tasemega statistiliselt oluliselt korreleeruvate geneetiliste variantide leidmine, kusjuures keskendutakse just *trans*-eQTLidele. See võimaldab leida beetainterferooni, põletikuvastuse kontrollimisel olulise valgu signaaliraja aktiivsust mõjutavad lookused, millest võib olla kasu inimkehas toimuvate põletikuliste protsesside mõistmisel [9].

“Mõistete ja terminite” peatükis tutvustatakse lugejale teemaga seotud põhilisi bioloogilisi mõisteid. “Kirjanduse ülevaates” selgitatakse käsitletavaid protsesse ja antakse lühiülevaade valdkonnas tehtud uuringutest. „Kasutatud tarkvara“ peatükis on üles loetletud töö käigus kasutatud tehnilised lahendused. „Lahenduskäigu kirjeldus“ annab ülevaate töö meetodikast ja tehtud arvutustest. Analüüsil leitu on välja toodud peatükis „Tulemused“. Töös kasutatud Tartu Ülikoolile usaldatud konfidentsiaalsed genoomiandmestikud on olnud aluseks kahe artikli ilmumisele [10,11].

Lisadena on tabelkujul tööga kaasas tugevaimad tuvastatud seosed iga andmestiku kaupa ja meta-analüüsil. Töö läbiviimisel ette tulnud kalkulatsioonid on läbi viidud Tartu Ülikooli teadusarvutuste keskuse arvutusmahtu kasutades. Töö käigus kasutatud R ja Linux shell skriptid on avalikult üles laetud GitHubi koodivaramusse ralf-tambets/bachelor-thesis [12].

2. Mõisted ja terminid

DNA ehk desoksüribonukleinhape on pärilikkusaine kaheaahelaline vorm, mis sisaldab geneetilisi juhiseid kõikide elusorganismide ja paljude viiruste arengu, elutegevuse ja paljunemisega seonduva toimimiseks.

Kromosoom (ingl *chromosome*) on kompaktne DNA molekul, mis sisaldab kõike või osa organismi geneetilisest materjalist. Inimese normaalses keharakkudes on 46 kromosoomi.

Geen (ingl *gene*) on kindla kromosomaalse paiknevusega DNA osa, mis kodeerib valgusünteesi.

Geneetiline variant ehk alleel (ingl *variant, allele*) on konkreetsel indiviidil esinev kindla geeni DNA järjestus, mis ei pruugi kattuda liigikaaslaste omaga.

Polümorfism (ingl *polymorphism*) on geneetilise varieeruvusega kaasnev mitme spetsiifilise geneetilise variandi esinemine ühe liigi geneetilise materjali piires¹. Antud töös tuleks eraldi välja tuua ka **SNP ehk ühe nukleotiidi polümorfism** (ingl *single nucleotide polymorphism*), mille puhul erinevad geneetilised variandid ainult ühe DNA ehitusploki ehk nukleotiidi poolest.

Genoom (ingl *genome*) on kogu kromosoomides leiduva geneetilise materjali kogum. Inimese genoomi moodustavad hinnanguliselt 20000-25000 geeni ja kromosoomide mittekodeeriv osa.

GWAS ehk ülegenoomne assotsiatsiooniuring (ingl *genome-wide association study*) on kogu genoomi uurimine mingi tunnusega seotud polümorfismide tuvastamiseks². Seda tehnikat kasutatakse sageli haigust tekitavate geneetiliste variantide leidmiseks.

Kvantitatiivne tunnus (ingl *quantitative trait*) on organismil avalduv tunnus, mida on võimalik mõõta pideval skaalal, tihti on nende avaldumisega seotud mitu erinevat geeni³. **QTL**

¹ <https://www.britannica.com/science/polymorphism-biology> (külastatud 05.05.2021)

² <https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/gwas> (külastatud 05.05.2021)

³ <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/quantitative-traits> (külastatud 05.05.2021)

ehk kvantitatiivse tunnuse lookus (ingl *quantitative trait locus*) on genoomi regioon, mis põhjustab kvantitatiivsete tunnuste varieeruvust⁴.

mRNA ehk infokandja RNA (ingl *messenger RNA*, kasutatakse ka eesti kirjanduses) on DNast transkriptsiooni teel saadav molekul, mis kannab geeni kodeeritud informatsiooni rakutuumast ribosoomidesse ja osaleb sel viisil valgusünteesis⁵.

eQTL ehk geeniekspressiooni kvantitatiivse tunnuse lookus (ingl *expression quantitative trait locus*) on genoomi regioon, mis põhjustab erisusi ühe või mitme geeni ekspresioonis⁶.

Trans-eQTL on eQTL, mis mõjutab geeniekspressiooni kahel homoloogsel kromosoomil⁶. **Cis-eQTL** suudab mõjutada vaid endaga samal kromosoomil paiknevat geeni⁶.

Interferoonid (ingl *interferons*) on tsütokiinide hulka kuuluvad viirusevastase toimega biomolekulid ja immuunomodulaatorid, mis jaotatakse kaheks tüübiks⁷. Esimese tüübi moodustavad alfa- ja beetainterferoon, mida suudavad vastusena viirustele toota peaaegu kõik inimese rakutüübid⁷. Teise tüübi ainus esindaja on gammainterferoon, mida toodetakse ainult kindlates immuunrakkudes (NK-rakud ja T-lümfotsüüdid) ja mille eesmärk on alarmeerida muid rakke võõra materjali olemasolust⁷.

Lipopolüsahhariid on Gram-negatiivsete bakterite rakuseinast eraldatud endotoksiin, mis käitub inimorganismi sattudes süsteemse põletikureaktsiooni vallandajana ja ka septilise šoki põhjustajana⁸.

⁴ <https://www.sciencedirect.com/topics/medicine-and-dentistry/quantitative-trait-locus> (külastatud 05.05.2021)

⁵ <https://www.britannica.com/science/messenger-RNA> (külastatud 05.05.2021)

⁶ <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/expression-quantitative-trait-loci> (külastatud 05.05.2021)

⁷ <https://www.britannica.com/science/interferon#ref283735> (külastatud 05.05.2021)

⁸ <https://www.sciencedirect.com/topics/neuroscience/lipopolysaccharide> (külastatud 05.05.2021)

3. Kirjanduse ülevaade

3.1 DNA struktuur

DNA molekulis sisalduv elusorganismi arenguks, elutegevuseks ja elu taastootmiseks vajalik geneetiline info on sinna salvestatud, kasutades nelja nukleotiidi [1]. Need pärilikkuse ehitusplokid koosnevad suhkrust (desoksüriboosist), fosfaatrühmast ja molekulile nime andvast lämmastikalusest, milleks võib olla adeniin, guaniin (puriinid), tsütosiin või tümiin (pürimidiinid), mida tähistatakse vastavalt tähtedega A, G, C ja T [1]. Nukleotiididest moodustuv lineaarne DNA polümeer salvestatakse kahe antiparalleelse ja komplementaarse aheelana, mis puriin-pürimidiinsidemete abil topeltheeliksi kuju võtavad [3]. Saadud molekul pakitakse spetsiaalsete valkude (histoonide) abil kompaktselt kolmedimensionaalseks struktuuriks ehk kromosoomiks, kaitsmaks seda välismõjutuste eest [1,3]. Inimese geneetiline info on salvestatud keharakkude tuumadesse 23 kromosoomide paari näol.

3.2 Geeniekspressioon

DNAs sisalduvad juhised viiakse teooriast praktikasse spetsiaalsete valkude, näiteks ka biokatalüsaatoriteks kutsutavate ensüümide abil. Geeniekspressiooni all peetakse silmas DNA lõigu põhjal valgusünteesimist, mis jaguneb kaheks suureks osaks: transkriptsiooniks ja translatsiooniks [1].

DNA valke kodeerivaid lõike nimetatakse geenideks ja nende osakaal kogu DNAs on inimestes vähem kui 5% [3]. Transkriptsiooni käigus moodustatakse konkreetse kodeeriva DNA lõiguga komplementaarne üheaheelaline mRNA molekul, mis kannab geneetilise info rakutuumast väljas paiknevasse ribosoomi, valgusünteesi eest vastutavasse organelli [5]. Olulised erinevused mRNA ja üheaheelalise DNA vahel on desoksüriboosi muutumine riboosiks ja tümiini (T) asendumine uratsiiliga (U) [3].

Translatsiooni käigus aheldatakse ribosoomis aminohappeid RNA poolt määratud järjestuses üksteise külge ning moodustub valk. Selleks loetakse mRNA nukleotiide kolme kaupa, kuni jõutakse järjestuseni AUG, mida nimetatakse startkoodoniks ja mis kodeerib aminohapet metioniin [1,3]. Eluslooduses on geneetiline kood, mis seab 61 erinevat aminohapet kodeerivat nukleotiidikolmikut ehk koodonit vastavusse 20 erineva aminohappega, praktiliselt

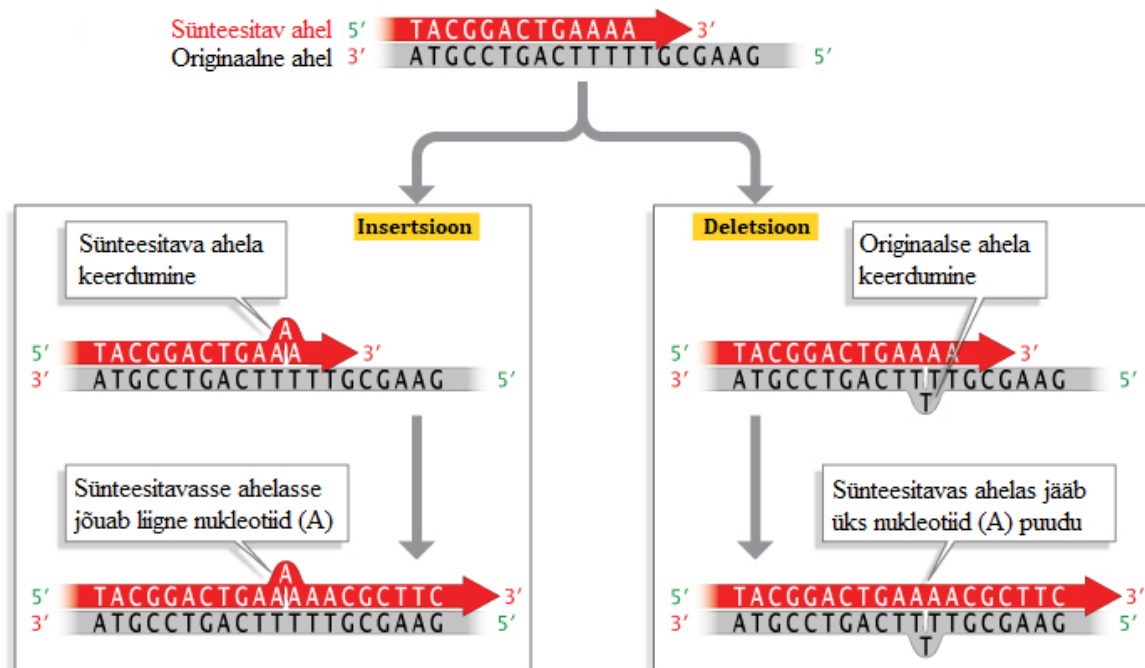
universaalne [5,3]. mRNA lugemine ja aminohapete järjestamine lõppeb, kui kohtutakse stoppkoodoniga ehk järjestusega UAA, UAG või UGA [3].

3.3 Geneetilise mitmekesisuse alused

Liigisisese geneetilise mitmekesisuse aluseks on muutused pärilikkusaines, mis tekivad mutatsioonide tagajärjel või rekombinatsiooni käigus. Nende protsesside puudumisel oleksid kõik sama liigi organismid teineteise kloonid ja erinevused fenotüübis oleksid puhtalt elukeskkonna eripäradest tingitud.

Kuigi vigu võib esineda nii translatsiooni, transkriptsiooni kui ka DNA replikatsiooni käigus, on viimased neist kõige ohtlikumad, sest parandamata jääv funktsionaalne muutus geenis viib selle geeni info kasutamisel peaaegu kindlasti veani [4]. Korrigeerimata jäänud replikatsioonivigadest saavad esimese raku jagunemise järel püsivad mutatsioonid, nende esinemine sugurakkudes võimaldab mutatsioonil pärilikuks saada [2].

Vigu DNA replikatsioonis võivad teoreetiliselt põhjustada prootoni asukoha muutumine nukleotiidi lämmastikaluses, mis muudab võimalikuks tavaliste puriin-pürimidiinsidemete ($A=T$, $G=C$) asemel ka puriin-puriin- ($A=C$) ja pürimidiin-pürimidiinsidemete ($G=T$) esinemise ja vastavate aluspaaride moodustumise [2]. Rohkem esineb aga DNA ahelate vahelist minimaalset ruumilist nihet või lämmastikalustesse prootonite lisandumist, mis samuti mainitud paardumishälveteni viia võivad [2]. Kirjeldatud lämmastikaluse vahetumist replikatsioonil nimetatakse substitutsiooniks, kuid see pole ainus mutatsioonide tekkimise mehhanism: näiteks DNA keerdumised võivad viia deletsioonideni ehk nukleotiidide vahele jätmiseni või insertsioonideni ehk nukleotiidide lisamiseni (joonis 1), millel on tihti markantsed tagajärjed, sest need muudavad translatsiooni lugemisakent ja võivad põhjustada korraga paljude aminohapete muutuseid [2]. Mutatsioone esineb ka väljaspool DNA replikatsiooni, näiteks hapniku vabade radikaalide või teiste mutageenide (UV-kiirgus, kemikaalid) toimel [3]. Lisaks esineb ka kromosoom- ja genoommutatsioone, kuid need ei tule selles töös vaatluse alla [4].



Joonis 1. Insertsioonide ja deletsioonide üks võimalik tekkemehhanism [2].

Niimoodi tekkinud kvantitatiivsed või kvalitatiivsed eripärad DNAs võivad vastavalt geneetilisele koodile viia valede aminohapete kaasamiseni valku ja muuta sel teel selle struktuuri ja funktsionaalsust. Teatud juhtudel võib juba ühe nukleotiidi mutatsioonist piisata tõsise haiguse tekkeks [4,13]. Vahetunud nukleotiidiga variandi sagedus määrab ära selle käsitluse: populatsioonis harva esinevaid eripärasid nimetatakse punktmutatsioonideks, sagedamini kui 1% indiviididest leiduvaid variante aga ühe nukleotiidi polümorfismideks (SNP – *single nucleotide polymorphism*) [14].

Paljude levinud haiguste tekkepõhjused on kompleksed ja nende patogeneesi uurides ei saa tihti peale osutada ühele kindlale DNA muutusele. Ent on ka erandeid: sirprakuline aneemia on üks sagedamatest ja enim uuritud monogeensetest haigustest maailmas ja selle enim levinud vormi põhjustab β -globiini geeni 17. nukleotiidi punktmutatsioon [13]. Tümiini muutumine adeniiniks toob kaasa kuuenda aminohappe muutumise valiinist glutamaadiks, mis omakorda põhjustab sünteesitava hemoglobiini polümeriseerumise ja toob kaasa struktuuraalseid häireid: kahjustatud saavad selle hapnikusidumisvõime ja ka erütrotsüüdi üldine funktsioon [13]. Hemoglobiinipolümeeri suurenemine pressib lõpuks erütrotsüüdist vee välja ja muudab selle kujult sirpjaks [13]. Sellised praakverelibled võivad veresoontes puruneda või trombe tekitada, mis põhjustavad vastavalt aneemiat ja infarkte [13]. Seega võib ka väikseim muutus geenis kaasa tuua markantseid tagajärgi.

3.4 Genoomi sekveneerimine

1990. aasta oktoobris alustati tööd *Human Genome Projecti* raames eesmärgiga välja selgitada täielik inimese DNA aluspaaride järjestus [15]. 2003. aastal jõudis projekt eduka lõpuni: „inimese ehitamise joonis“ sai välja üldjoontes selgitatud ja avas paljud uued fenotüübi ja patoloogiate kujunemise uurimise ja kirjeldamise võimalused [15]. Alguspäevadega võrreldes on genoomi sekveneerimine muutunud oluliselt kiiremaks ja ökonoomsemaks: kui esimese selline protseduur maksis ligikaudu 300 miljonit USA dollarit, siis aastaks 2016 oli hind langenud alla 1000 dollari [16]. Tehnoloogilised edasimineked on olnud personaalmeditsiini progresseerumisel märgilise tähtsusega ja võimaldanud koguda geneetilisi andmeid üha suurematelt populatsioonidelt.

3.5 GWAS ehk ülegenoomne assotsiatsiooniuring

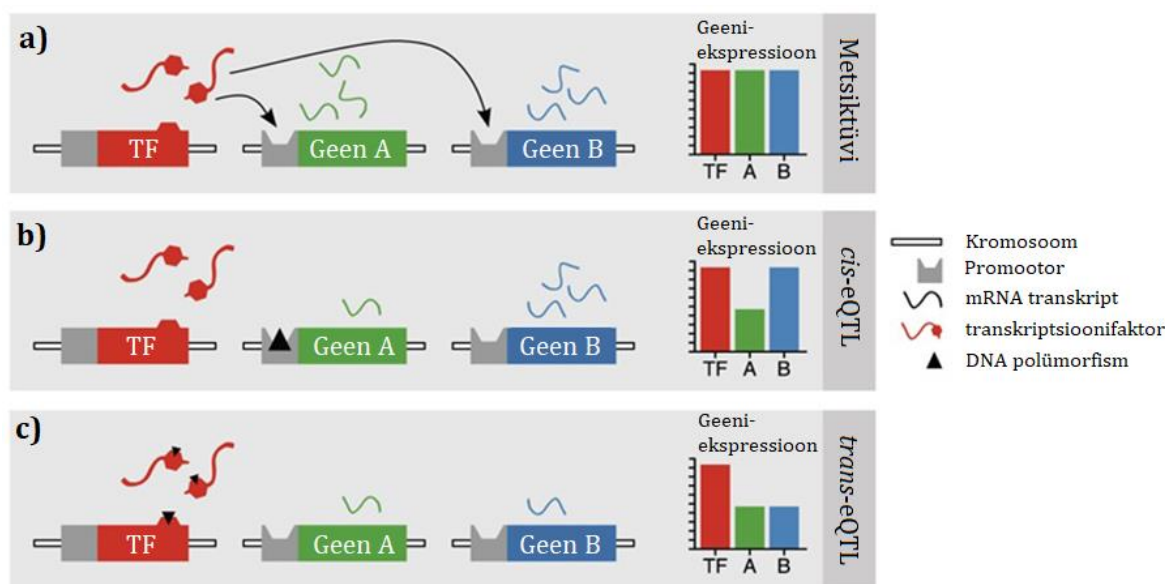
Üks põhilisi uurimismeetodeid, mille abil üritatakse välja selgitada komplekshaiguste geneetilisi tegureid, on ülegenoomne assotsiatsiooniuring ehk GWAS. Selle käigus testitakse rohkete katsealuste geneetilisel materjalil sadu tuhandeid ühe nukleotiidi polümorfisme (SNP), et leida geneetilised variandid, mis on uuritava (näiteks patoloogilise) seisundiga tihedalt assotsieerunud [8]. Näiteks olid juba kümme aastat tagasi tuvastatud viis geneetilist varianti, mis suurendavad vanusega seotud makulaarse degeneratsiooni, tsentraalse nägemisvälja kahjustusega kulgeva levinud silmahaiguse riski 2-3 korda, ja kolm sagedast polümorfismi, mis on seotud peamiselt põletikulise soolehaigusena avalduva Crohni tõve riski 1,5-4 kordse tõusuga [8]. Tänapäevaks on õnnestunud kaardistada juba tuhandeid komplekshaigustega seotud geneetilisi lookuseid, kuid arusaam nende haiguse kulgu soodustavatest mehhanismidest on veel sageli puudulik [6,7].

Et inimestes on geneetiline info salvestatud kromosoomide paaridena, kusjuures erinevates kromosoomides paiknev info on reeglina päritud eri vanematelt, võib katsealune olla uuritava lookuse suhtes homosügoot (mõlemas kromosoomis antud lookuses sama nukleotiid) või heterosügoot (lookuse nukleotiidid erinevad kromosoomiti). Nende kombinatsioonide ja tunnuse väljendumise korrelatsiooni leidmine ongi GWASi aluseks [6,17]. Kuna ühe GWASi käigus testitakse tuhandeid SNPe, on valepositiivsete tulemuste leidmise oht suur. Nende minimeerimiseks kasutatakse Bonferroni korrektsiooni: usaldusnivoo määramiseks võetakse klassikaline p-väärtus 0,05 ja jagatakse see testitud SNPide arvuga [17]. Standardseks GWASil kasutatavaks usaldusnivooks on $5E-8$ [17].

GWASile on teadusringkonnas ka vastaseid tekkinud. Kriitikute sõnul on SNPide roll komplekshaiguste tekkes väheoluline ja GWASi käigus tuvastatud seoste põhjuslikkuses on raske kindel olla – paljud assotsiatsioonisignaaliid on seotud mittekodeerivate DNA järjestustega ja seetõttu on nende bioloogilise efekti hindamine keeruline [18]. Lisaks pälvib kriitikat fakt, et mastaapsete järelduste tegemiseks peab uuritav populatsioon olema piisavalt mitmekesine ja seda on reaalsuses keeruline saavutada [18].

3.6 eQTL olemus ja nende tuvastamine

Alberti ja Kruglyaki (2015) definitsiooni kohaselt on eQTLid genoomi regioonid, mis sisaldavad DNA järjestuse variante, mis mõjutavad ühe või mitme geeni ekspressiooni ja mis tuvastatakse geneetiliselt populatsiooniuringutel [7]. Katsealustelt kogutakse geneetiline info kas kogu genoomi või selle kindla regiooni järjestamise abil ja mõõdetakse siis iga geeni ekspressiooni (geeniproducti kontsentratsiooni) indiviidi kaupa [7]. Seejärel hinnatakse statistilise analüüsi käigus DNA variandi mõju geeniekspressioonile [7]. Kui variantiga seotud geeniproducti hulk on statistiliselt olulisel määral erinev populatsiooni keskmisest, järeldatakse, et vaadeldav variant ise või sellega geneetiliselt aheldunud variant põhjustab muutusi geeniekspressioonis ehk on eQTL [7].



Joonis 2. *cis*- ja *trans*-eQTLide toime illustatsioon [19]. a) illustreerib tavalist stsenaariumi, kus TF on geen, mis toodab transkriptsioonifaktorit, mis seostub homologsetes kromosoomides asuvate geenide A ja B promootoritega ja aktiveerib nende transkriptsiooni. Kõiki transkripte tekib normaalsel hulgal. b) puhul pärsib A promootoris asuv polümorfism

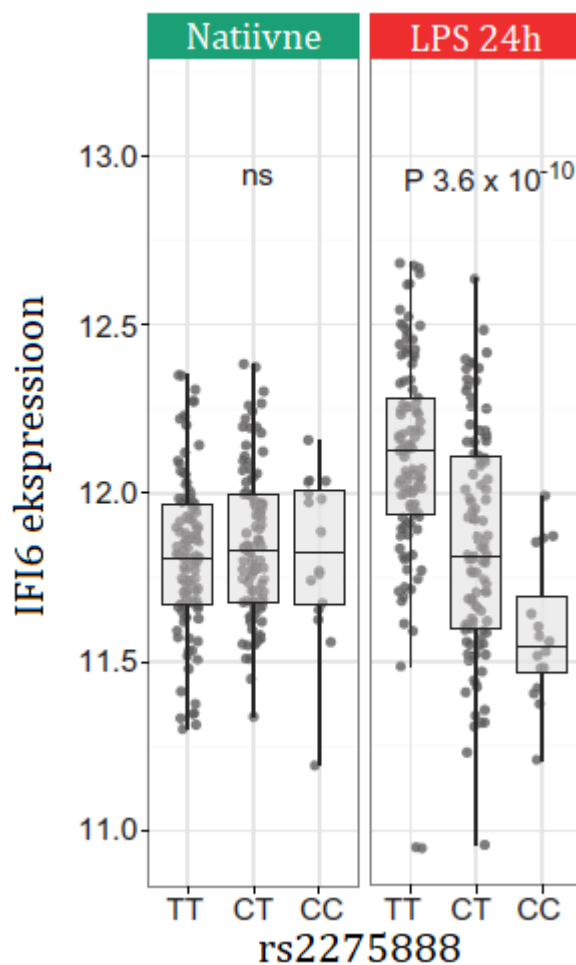
transkriptsioonifaktori seostumist. A ekspressioon langeb, B oma mitte. Antud polümorfismi näol on tegemist *cis*-eQTLiga. c) illustreerib *trans*-eQTLi: TFis paikneva polümorfismi tõttu on transkriptsioonifaktori seostumine mõlema sihtmärkgeeniga takistatud ja mõlema geeni ekspressioon langeb.

eQTLid saab toime sihtpunkti järgi jagada *cis*-eQTLideks, mis mõjutavad vaid samal kromosoomil paiknevaid gene, ja *trans*-eQTLideks, mis avaldavad mõju ka teisel homoloogsel kromosoomil paiknevatele geenidele (joonis 2) [7]. Käesolev töö keskendub beetainterferooni (IFN- β) ekspressiooni muutustega seotud *trans*-eQTLidele.

3.7 Beetainterferoon ja selle roll geeniekspressioonis

Beetainterferoon on tsütokiinide, täpsemalt I klassi interferoonide hulka kuuluv viirusvastase, rakukasvu pärssiva ja immunomoduleeriva toimega signaalmolekul, mida toodetakse immuunrakkude poolt vastusena bioloogilistele ja keemilistele mõjutustele [9]. Selle sünteesi kontrollib 9. kromosoomis asuv geen *IFNB1* [20].

2014. aastal ilmunud artiklis kirjeldasid Fairfax jt, et monotsüütide kahetunnisel eksponeerimisel põletikuvastust stimuleerivale lipopolüsahhariidile (LPS) õnnestus tuvastada *IFNB1* vastust mõjutav *cis*-eQTL [10]. Lisaks täheldati, et *IFNB1* ekspressiooniga seotud geneetilised variandid ise on *trans*-eQTLiks 19 geenist koosnevale regulatoorsele võrgustikule, mille tulemusena leidis aset geenide *IRF9*, *IFI6*, *EPSTI1*, *LY6E* ja *MX1* ekspressiooni statistiliselt oluline tõus pärast 24 tunni pikkust LPSi mõjutust [10]. Variandi rs2275888 seost *IFI6* ekspressiooniga on illustreeritud joonisel 3.



Joonis 3. Variandi rs2275888 genotüübi ja *IFI6* ekspressiooni vahel on T-alleeli olemasolu tugev seos geeniekspressiooni tõusuga tuvastatav pärast 24tunnist LPS mõjutust, kuid mitte natiivolekus [10].

Kaks aastat hiljem läbi viidud laiahaardelises uuringus kinnitati varasemat leidu: nii aafrika kui ka euroopa populatsiooni uurimisel leiti, et *INF1* lookusel oli tugevaim *trans*-regulaatorne mõju ning see oli aluseks viirusvastase toimega geenivõrgustiku toimimisele [11]. Lisaks märgati, et *INF1* ennast mõjutavatel *cis*-eQTLidel on geenivõrgustiku mõjutamisel tagasihoidlik roll ja järeldati, et *trans*-regulatsiooni taga peavad olema seni tuvastamata faktorid [11]. Nende leidmine ongi käesoleva töö põhieesmärk.

4. Kasutatud tarkvara

4.1 R

Andmete töötlemiseks kasutati programmeerimiskeele R versiooni 4.0.3. Järgneb ülevaade töö käigus kasutatud teekidest.

4.1.1 MatrixEQTL

MatrixEQTL on suurtest andmestikest eQTLide tuvastamiseks mõeldud efektiivse algoritmiga R teek [21]. MatrixEQTL on võimeline arvestama kovariaatidega, lubab määrata maksimaalse p-väärtuse, mille puhul tulemus tagastatakse, ja täpsustada vähima aluspaaride hulga uuritava SNPi ja uuritava geeni vahel [21]. Tagastab relevantsuse järgi reastatud SNPide ja geenide paarid, neile vastava p-väärtuse, beeta-väärtuse, t-statistiku ja valepositiivsete osakaalu kõigi positiivsete proovide seas [21].

4.1.2 SNPrelate

SNPrelate on genoomiüleste assotsiatsioonuuringute tulemuste töötlemiseks mõeldud Xiwen Zhengi poolt välja töötatud vabavaraline teek [22], mida kasutati käesoleva lõputöö raames Quach_2016 populatsioonil peakomponentanalüüsi teostamiseks.

4.1.3 dplyr

dplyr on tidyverse'i kuuluv andmetega manipuleerimiseks (tabelite muutmise, andmete filtreerimise) mõeldud teek [23]. Töös kasutati seda tabelite liitmisel, proovide järjekorra korrastamisel ja andmete populatsioonide kaupa filtreerimisel.

4.1.4 ggplot2

ggplot2 on populaarne graafikute loomise teek, mis loodi rohkem kui 10 aastat tagasi [24]. Sarnaselt dplyr'ile on ka ggplot2 osa tidyverse'ist [24]. Antud töös kasutati seda peakomponentanalüüsi graafikute ja histogrammide kuvamiseks.

4.1.5 qqman

Stephen Turneri poolt loodud qqman on teek kvantiil-kvantiil- ja manhattani graafikute loomiseks. Kasutatud andmestike analüüsi kvaliteedi hindamiseks ja lõpptulemuste graafiliseks kujutamiseks [25].

4.2 BCFtools

BCFtools on VCF- (*Variant Call Format*) ja BCF-failide (*BIM Collaboration Format*) töötlemiseks mõeldud tarkvarapakett, mis võimaldab neid faile konverteerida, filtreerida, soovitud osas kuvada, indekseerida ja annotatsioone lisada, eemaldada või muuta [26]. Kasutatud kogu genoomi kirjeldavast VCF-failist IFNB1 lookuse välja lõikamiseks ja VCF-failidest info lugemiseks.

4.3 METAL

METAL on 2008. aastal välja töötatud ja laialdast kasutust leidnud genoomiüleste assotsiatsiooniuuringute metaanalüüsiks kombineerimise tööriist, mis vajab sisendiks standardvigu või p-väärtusi [27]. Antud töös rakendati seda nelja genoomiülese assotsiatsiooniuuringu kombineerimisel ja lõpptulemuste saavutamiseks.

4.4 PuTTY

PuTTY on laialt levinud vabavaraline SSH, Telnet ja Rlogin klient Windowsi süsteemidele [28]. Antud töös kasutati seda Tartu Ülikooli teadusarvutuste keskuse süsteemiga suhtlemiseks.

4.5 FileZilla

FileZilla on vabavaraline FTP klient [29], mis leidis töö käigus rakendust Tartu Ülikooli teadusarvutuste keskuse ja autori arvuti vahel failide liigutamisel.

5. Lahenduskäigu kirjeldus

5.1 Uuritavad andmestikud

Fairfax_2014 - artikli „Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression“ ilmumiseni viinud andmestik sisaldab 325 eurooplase geneetilist informatsiooni [10].

Fairfax_2018 – veel avaldamata andmestik koosneb 176 eurooplasest indiviidist.

Quach_2016 - andmestikku kuulub 96 eurooplast ja 88 aafriklast. Andmestikku kasutati artikli „Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations“ loomiseks [11].

Kõigi andmestike kohta oli olemas genotüüpe kirjeldav VCF-fail, mis sisaldab infot kõigi uuringus käsitletud SNPide kohta. Iga andmerida hõlmab muuhulgas infot kromosoomi kohta, milles uuritav SNP asub, positsiooni selles kromosoomis, SNPi nimetust, referentsalleeli ja alternatiivse alleeli nukleotiidi ja iga uuringus osalenud indiviidi genotüüpi selles lookuses. Ükski andmestik ei kätke infot sugukromosoomide kohta. Geeniekspressiooni tasemed, metainformatsioon uuringus osalenud inimeste kohta ja kogu populatsiooni põhjal genereeritud kuus põhilist peakomponenti olid kaasas eraldi failidena. Fairfax_2014 ja Fairfax_2018 andmestikku kuuluvad monotsüüdid olid lipopolüsahhariidile eksponeeritud 24h, Quach_2016 omad 6h.

5.2 Lahenduskäigu põhiidee

Lipopolüsahhariidiga põletikureaktsiooni esile kutsumine tõstab beetainterferooni signaalirajas osalevate geenide ekspressioonitaset ja võimaldab hinnata, milliste geenide ekspressioon on *IFNBI* genotüübiga kõige tugevamalt seotud [10]. Selleks tuleb genoomiülese assotsiatsiooniuuringu käigus kõrvutada *IFNBI* lookuse geneetilisi variante kõigi uuringu käigus mõõdetud geeniekspressiooni tasemetega. GWASi käigus tugevate seoste leidmine geeniga väljaspool 9. kromosoomi viitab *IFNBI* variantide *trans*-reguleerivale toimele, selle geeni võimalikule osalusele beetainterferooni signaalirajas ja tõstab tõenäosust, et antud geen oleks sobilik markergeen beetainterferooni vastusele.

Markergeeni valimise järel on otstarbekas teostada teine GWAS, mille puhul uuritakse selle geeni korrelatsiooni kõigi andmestikku kuuluvate SNPidega. Tugevad seosed väljaspool

IFNB1 lookust annavad tunnistust võimalike *trans*-eQTLide kohta, mis beetainterferooni signaaliraja aktiivsust mõjutavad.

5.3 Beetainterferooni vastuse markergeeni otsimine Fairfax_2014 andmestikust

Fairfax_2014 andmestik valiti beetainterferooni vastuse markergeeni otsimiseks välja suurima valimi tõttu. Andmetöötluse esimese sammuna filtreeriti metaandmete faile, et kaasata uuringusse ainult proovid, mis olid võetud pärast monotsüütide 24-tunnist ekspositsiooni lipopolüsahhariidile ja mis olid eelnevalt läbinud RNA ja genotüübiandmete kvaliteedikontrolli. Valimisse mitte mahtunud proovid jäeti kõrvale kõiki ülejäänud faile töödeldes. Järgmisena kasutati BCFtoolsi, et lõigata VCF-failist välja beetainterferooni tootmise eest vastutav *IFNB1* lookus ja selle lähiümbrus. Lähiümbruseks valiti 9. kromosoomis positsioonil 20818520 paiknev variant rs13296842 ja sellest mõlemas suunas kuni 500000 aluspaari kaugusele jäävad SNPid.

Beetainterferooni vastuse markergeeni leidmiseks tuleb testida kõiki *IFNB1* lookuse variante kõigi geenide vastu Fairfax_2014 andmestikus ja leida tugevaimates seostes olevad paarid. Selleks kasutati MatrixEQTL teeki, mis nõuab arvutuse läbiviimiseks andmeid geeniekspressiooni, genotüübi, SNPide asukohtade, geenide asukohtade ja soovitatavalt ka kovariaatide kohta. Asukohainfo ekstraheeriti VCF-failist BCFtoolsi kasutades ning kõik andmed viidi analüüsiks sobivale, kas maatriks- või andmetabelkujule, kasutades programmeerimiskeelt R ja selle teeki dplyr. *Trans*-eQTLidele keskendumiseks arvati uurimishulgest välja geenid, mis paiknesid lookusest vähem kui 200000 aluspaari kaugusel. Valepositiivsete tulemuste vähendamiseks ei kaasatud uuringusse variante, mille keskmine alleelisagedus (ingl k *mean allele frequency*, MAF) populatsioonis oli väiksem kui 1%.

Analüüsi tulemusena selgus, et *IFNB1* lookusega annab selgelt kõige rohkem tugevaid seoseid 1. kromosoomil paiknev geen *IFI6*: 100 kõige tugevamast seosest 99 olid just selle geeniga seotud, kusjuures vähim p-väärtus oli 2.23E-17.

Leiud on kooskõlas 2014. aastal Fairfax jt poolt sama andmestiku põhjal avaldatud artikliga, mis tuvastas *IFNB1* seotuse *IFI6*, *IRF9*, *EPST11*, *LY6E* ja *MX1* ekspressiooniga. Kõigi nende geenidega tuvastati tugevamad seosed, kui genoomiülel uuringul juhuslikuks pidada saab (p-väärtused väiksemad kui 5E-8). Tugevaimad seosed nende geenidega on välja toodud tabelis 1.

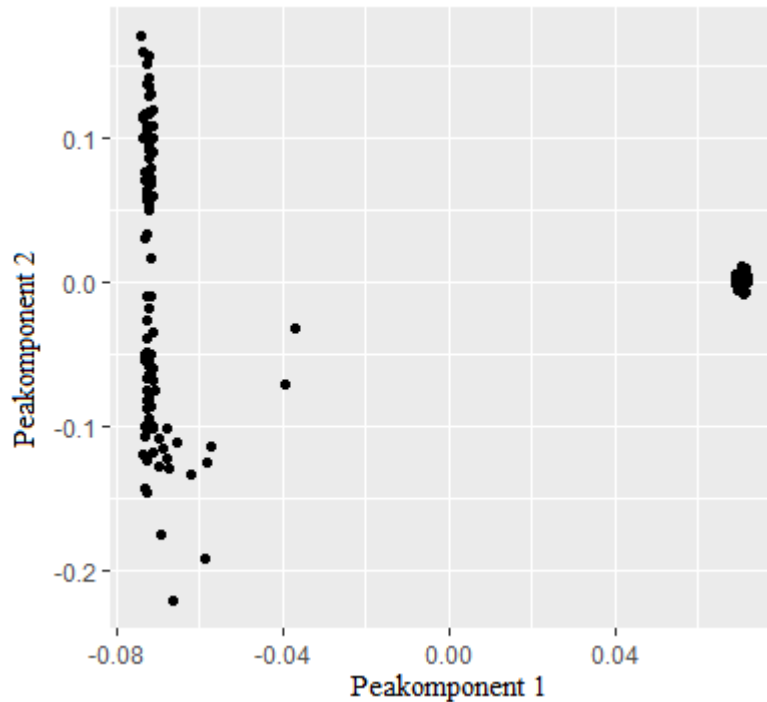
Tabel 1. Tugevaimad seosed Fairfax jt artiklis välja toodud geenidega. Tulp „SNPi kood“ sisaldab infot SNPi kromosoomi, positsiooni, referents- ja alternatiivse alleeli kohta.

<i>Geen</i>	<i>SNPi kood</i>	<i>p-väärtus</i>
IFI6	chr9_20827115_G_C	2.2287E-17
EPSTI1	chr9_20818520_A_G	4.4152E-13
IRF9	chr9_20818520_A_G	6.5357E-12
LY6E	chr9_20976060_A_G	1.4261E-12
MX1	chr9_20840034_C_T	3.9897E-11

IFI6 markergeeniks sobivuse kinnitamiseks teostati uus eQTL-analüüs, mille käigus testiti *IFI6* geeni kõigi SNPide vastu Fairfax_2014 andmestikus (Lisa 1, tabel 2). Analüüs paljastas, et *IFI6* ekspressiooni mõjutavad geneetilised variandid paiknevad väga kitsas DNA piirkonnas: ühegi variandiga väljaspool *IFNBI* lookust ei esinenud seoseid, mille p-väärtus oleks väiksem kui 5E-8. Selle põhjal valiti *IFI6* potentsiaalseks beetainterferooni vastuse markergeeniks ja asuti uurima, kas sama seos on leitav ka teistes andmestikes.

5.4 Otsingu kordamine Fairfax_2018 ja Quach_2016 andmestikel

Fairfax_2018 andmestiku töötlemise ja uuringu läbiviimise sammud olid Fairfax_2014 omaga identsed ja neid siin ei korrata. Quach_2016 andmestik nõudis täiendavat töötlust, sest see sisaldas lisaks Euroopa uurimisalustele ka Aafrika päritolu indiviide. Et mitte uurida abstraktset keskmist populatsiooni, eraldati kahe kontinendi inimeste andmed teineteisest, toetudes uuringuülelele peakomponentanalüüsile (joonis 4). Selline teguviis nõudis saadud VCF-failide populatsiooni kaupa poolitamist ja populatsiooniti uute peakomponentanalüüside teostamist ning annotatsioonide uuendamist.



Joonis 4. Quach_2016 kogu andmestiku peakomponentanalüüsi tulemus. Graafiku paremasse serva on koondunud indiviidid Euroopast, vasakusse Aafrikast.

Ei Fairfax_2018 ega Quach_2016 kummaski populatsioonis polnud *IFNBI* lookuse variantide tugevaimad seosed *IFI6* geeniga. Fairfax_2018 analüüsil oli *IFNBI* lookusega tugevaimas seoses 2. kromosoomil paiknev *MRPL30*, Quach_2016 Euroopa populatsioonis *MSLN* kromosoomilt 16 ja Quach_2016 Aafrika populatsioonis *AL445490* 1. kromosoomilt.

Tugevaim seos *IFNBI* ja *IFI6* geeni vahel oli Fairfaxi populatsioonis kõigest $3.85E-4$ (pingereas kohal 21470). Quachi Euroopa populatsioonis oli sama seose tugevuseks $5.84E-9$ (pingereas 4515.) ja Aafrika populatsioonis küllaltki tugev $1.10E-11$ (355. koht).

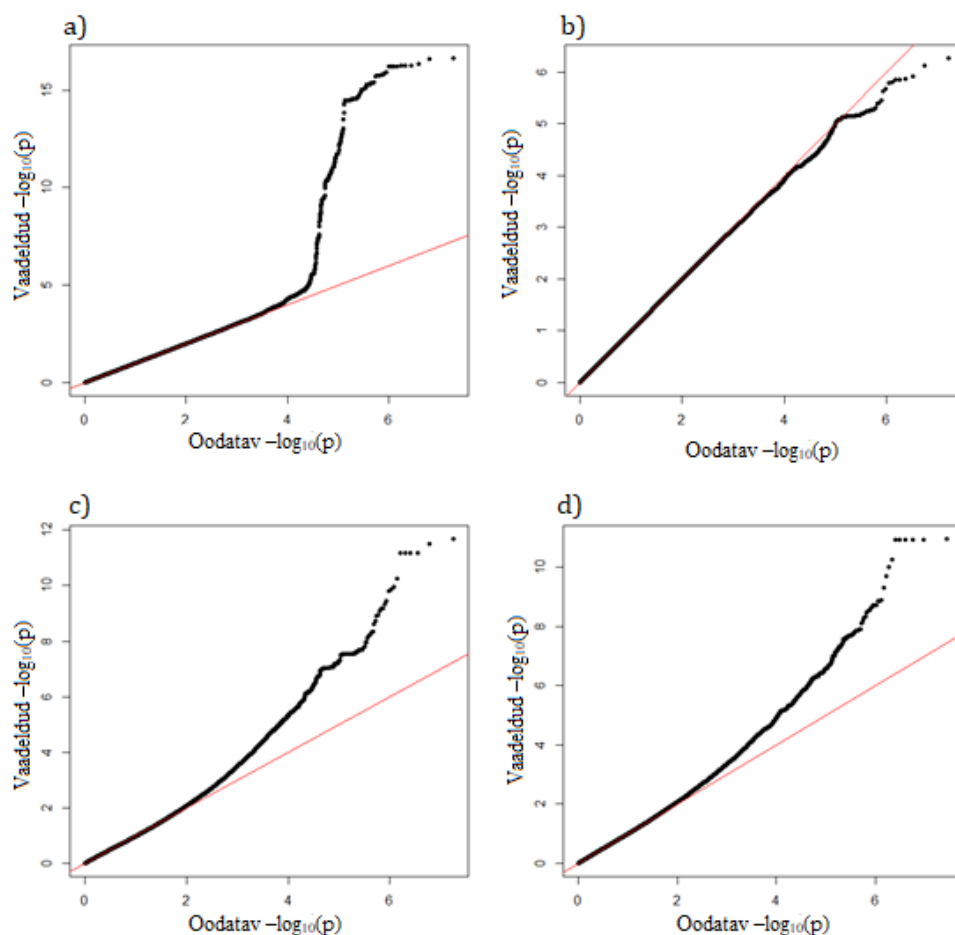
IFI6 geeni testimisel kõigi geneetiliste variantide vastu domineerisid Quachi Aafrika populatsioonis seosed *IFNBI* lookusega, kuid ülejäänud kahes populatsioonis polnud tulemused nii selged, vaid sarnase tugevusega seotud variante võis leida mitmetes kromosoomides (Lisa 1, tabelid 3-5).

5.5 Analüüsi tulemuste usaldusvääruse hindamine

Kõigil neljal populatsioonis erinevate tulemuste saamine tõstatab küsimuse: kas tulemused olid erinevad valimite erinevuse tõttu või on analüüsi käigus tehtud vigu? Selles selgusele jõudmiseks on üks võimalik lähenemine p-väärtuste jaotuse uurimine. Statistiliselt oluliste

seoste puudumisel järgivad p-väärtused teoorias ühtlast jaotust, madalate p-väärtuste suur osakaal annab tunnistust tugevate seoste olemasolust [30].

Käesoleva töö puhul, nagu GWASide puhul üldse, on alternatiivse hüpoteesi kehtimisel oodatavaks tulemuseks vaid väheste väärtuste kõrvalekalle ühtlasest jaotusest, sest vaikivaks eelduseks on, et ainult üksikud geneetilised variandid on kindla geeni ekspressiooniga seotud. Seetõttu kasutatakse GWASide p-väärtuste visualiseerimiseks ja seeläbi uuringu usaldusväarsuse hindamiseks tüüpiliselt kvantiil-kvantiil-graafikuid, mitte histogramme nagu väiksemate uuringute puhul. Kvantiil-kvantiilgraafik (edaspidi QQ-graafik) näitab eksperimendi käigus kohatud p-väärtuste kõrvalekallet nullhüpoteesi kehtimisel oodatavast, võrreldes eksperimendi tulemuste kvantiile simuleeritud ühtlase jaotuse kvantiilidega [31]. Miljonitele andmepunktidele tugineval histogrammil ei põhjustaks väike hulk ülimaldalaid p-väärtusi märgatavat muutust, samas kui QQ-graafikul on need selgesti näha.

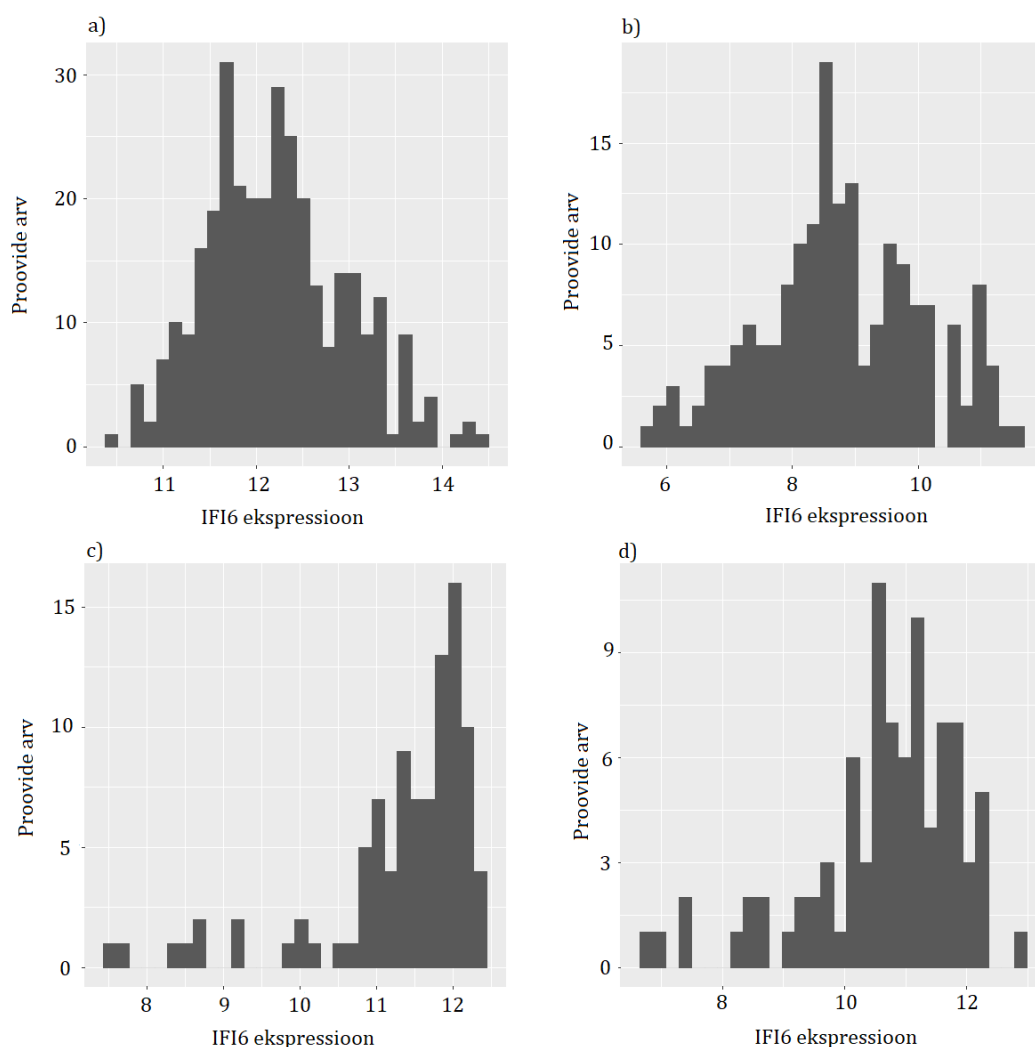


Joonis 5. QQ-graafikud *IFI6* assotsiatsioonidest kõigi levinud geneetiliste variantidega (MAF > 0.01) andmestike kaupa. a) Fairfax_2014; b) Fairfax_2018; c) Quach_2016, populatsioon 1; d) Quach_2016, populatsioon 2.

Joonis 5 sisaldab kõigi nelja andmestiku tulemuste QQ-graafikuid ja näitab, et Quach_2016 mõlema populatsiooni puhul toimub kõrvalekalle oluliselt varem kui Fairfaxi andmestikel. Selline tulemuste jaotus võib viidata populatsiooni kihistumisele [31] ja tähendab, et antud kujul ei tohiks Quach_2016 andmeid metaanalüüsi kaasata, kuna see tõstaks valepositiivsete seoste hulka.

5.6 Quach_2016 andmestiku täiendav töötlus

Quach_2016 andmestikul üllatavate tulemuste saamise võimalikuks põhjuseks peeti *IFI6* ekspressiooni jaotuse kõrvale kaldumist normaaljaotusest. Hüpooteesi kontrollimiseks koostati kõigi andmestike *IFI6* ekspressiooni histogrammid (joonis 6).



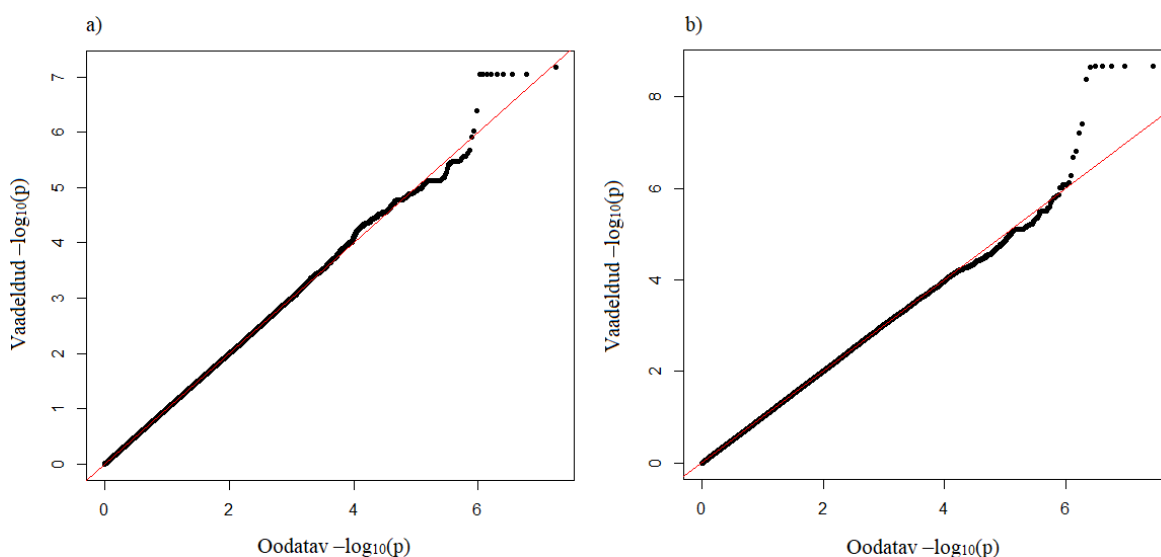
Joonis 6. *IFI6* ekspressiooni histogrammid andmestike kaupa. a) Fairfax_2014; b) Fairfax_2018; c) Quach_2016, Euroopa populatsioon; d) Quach_2016, Aafrika populatsioon.

Histogrammid kinnitasid juba tekkinud teooriat: Quach_2016 andmestikus ei järgi IFI6 ekspressioon normaaljaotust, vaid on kõrgete väärtuste poole kaldu. Probleemi lahendamiseks lähendati andmestiku mõlema populatsiooni ekspressiooniväärtused normaaljaotusele, kasutades Yang jt (2012) poolt välja töötatud R koodilõiku [32]:

```
y <- qnorm((rank(x, na.last="keep") - 0.5) / sum(!is.na(x)))
```

, mis asendab ekspressiooniväärtuse selle asukohaga ekspressiooniväärtuste kasvavas järjendis ja asetab tulemused normaaljaotusesse.

Töödeldud andmetel korraldati analüüsi ja koostati uued QQ-graafikud (joonis 7).



Joonis 7. QQ-graafikud Quach_2016 populatsioonidest pärast täiendavat andmetöötlust. a) Euroopa populatsioon; b) Aafrika populatsioon.

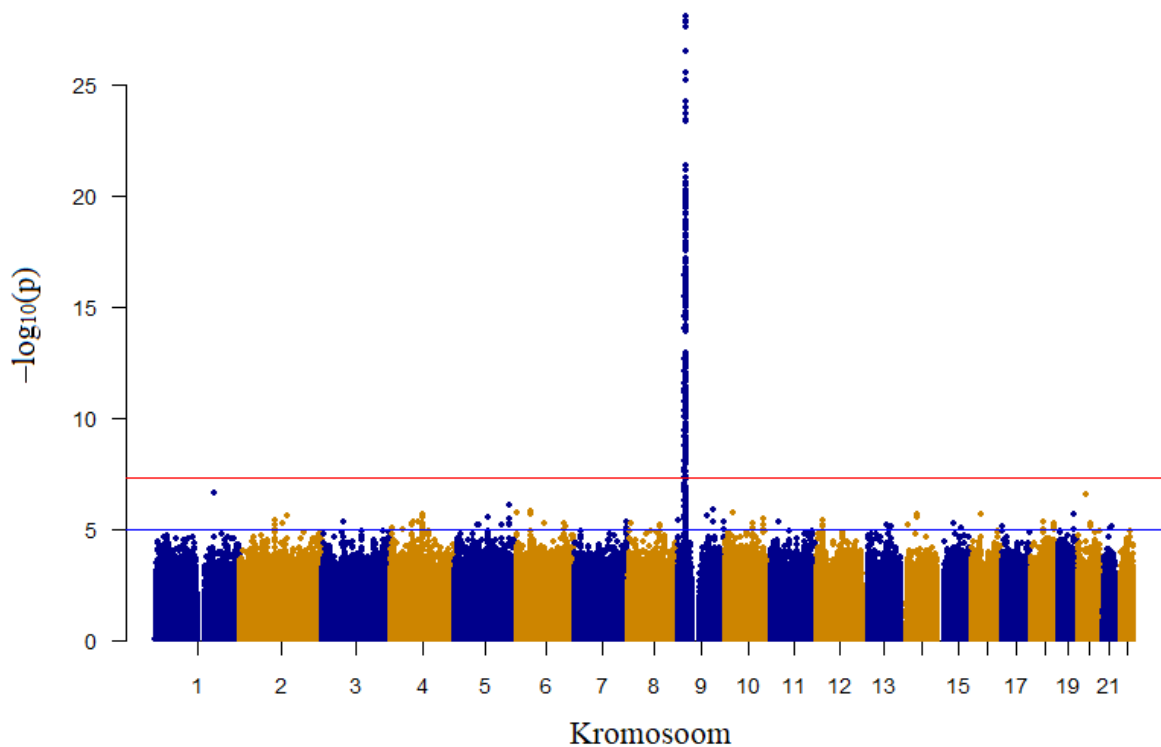
Seekord järgisid tulemused oluliselt paremini oodatavat kuju. Et kõik neli uuritavat populatsiooni on teineteisest sõltumatud ja et tulemuste QQ-graafikutes anomaaliaid ei esinenud, otsustati need kõik metaanalüüsi kaasata.

5.7 Metaanalüüsi läbiviimine nelja populatsiooni andmetel

Statistilise võimsuse suurendamiseks kaasati kõigi 685 indiviidi eQTL-analüüsi tulemused (kokku 4986494 p-väärtust) metaanalüüsi. Selleks anti METAL tarkvara sisendiks SNPide nimed, referents- ja alternatiivne alleel igas lookuses, statistiline olulisus p-väärtuse näol, efekti suund (ekspressiooni võimendav või pärssiv) beeta-väärtuse kujul ja iga p-väärtuse olulisus ehk kaal. Kaalu all mõeldakse antud kontekstis indiviidide hulka, mille geneetilise materjali põhjal vaatluse all olev seos välja arvatud sai.

6. Tulemused

Metaanalüüsi tulemusena selgus, et 285 *IFI6*ga kõige tugevamalt seotud geneetilist varianti paiknevad *IFNB1* lookuses, kusjuures 266 puhul neist oli p-väärtus väiksem kui GWASide puhul tavaliselt statistilise olulisuse näitajaks peetav $5E-8$. Statistiliselt olulisim seos *IFNB1* lookuses asuva variandiga oli oluliselt kõrgem kui ühegi üksikuuringu puhul, $9.074E-29$ (Lisa 2, tabel 6), samas kui väljaspool seda asuvate variantidega oli tugevaimaks seoseks $2.2580E-7$. Tulemuste visualiseerimiseks genereeriti qqman teeki kasutades manhattani graafik, võimaldamaks seoseid kromosoomide kaupa näha (joonis 8).

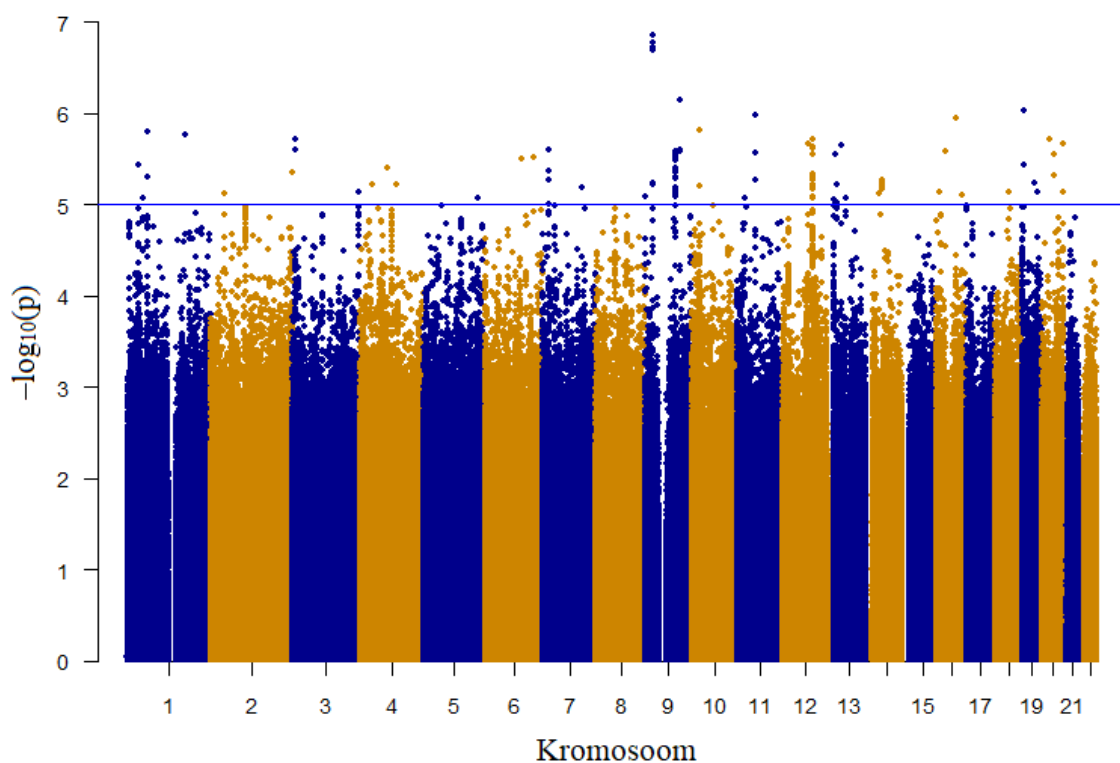


Joonis 8. Manhattani graafik nelja andmestiku metaanalüüsi tulemustest. 9. kromosoomis paiknev piik kattub *IFNB1* lookusega.

Tulemused peegeldavad *IFNB1* märkimisväärset mõju *IFI6* ekspressioonile, kuid ei paljasta ühtegi teist statistiliselt olulist *trans*-eQTLi. See äratas kahtlust, et *IFNB1* lookuse tugev seos *IFI6* ekspressiooniga lisas analüüsi müra, mistõttu langetati otsus uuringut korrata andmestike peal, milles *IFNB1* mõju on välja regresseeritud.

IFNB1 mõju vähendamiseks lisati seni MatrixEQTLi analüüsil kovariaatidena kasutatud kuue peakomponendi hulka ka juhtvariandi (Fairfax_2014 andmestikul *IFI6*ga tugevaimat

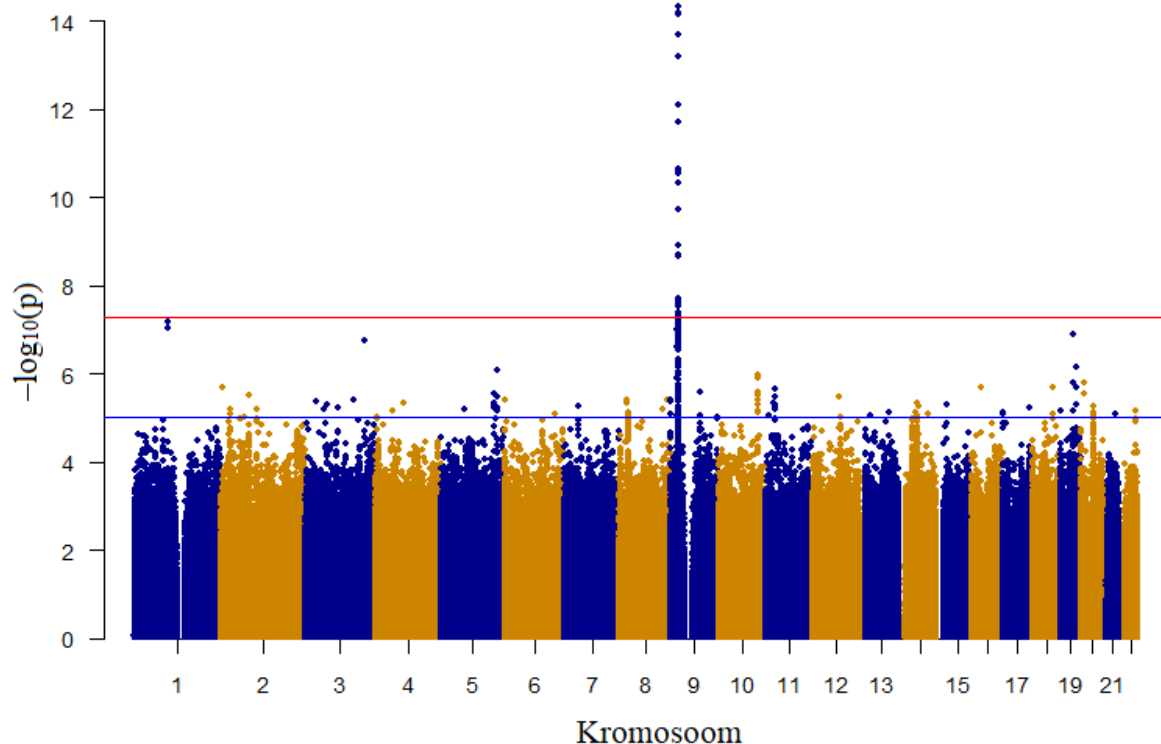
seost näidanud geneetilise variandi chr9_20827115_G_C) doos igas proovis. Analüüsi tulemused on jäädvustatud joonisele 9.



Joonis 9. Manhattani graafik nelja andmestiku metaanalüüsi tulemustest pärast *IFNBI* mõju regresseerimist.

Selgus, et ka pärast *IFNBI* mõju regresseerimist ei õnnestunud metaanalüüsil tuvastada täiendavaid statistiliselt olulisi *trans*-eQTLs, mis beetainterferooni signaaliraja aktiivsuse kontrollis *IFI6* geeniekspressiooni mõjutamise kaudu osaleksid (Lisa 2, tabel 7).

Ühe võimaliku probleemide allikana nähti Fairfax_2014 liigtugevat mõju: tegemist oli suurima kaaluga andmestikuga ning *IFNBI* lookuse ja *IFI6* vaheline signaal oli seal mitme suurusjärgu võrra tugevam kui ülejäänud andmestikel. Seega otsustati metaanalüüsi korrata ka ilma Fairfax_2014 andmestiku kaasamata. Ka see analüüs kinnitas juba varem leitud: ainsad *IFI6* ekspressiooniga tugevalt seotud *trans*-eQTLid paiknevad *IFNBI* lookuses (joonis 10; lisa 2, tabel 8).



Joonis 10. Manhattani graafik kolme andmestiku tulemustest (Fairfax_2014 välja arvatud).

Saavutatud tulemuste põhjal ei saa välja tuua ühtegi seni tuvastamata beetainterferooni signaalirada mõjutavat geneetilist varianti, mis võib näidata markergeeni ebaõnnestunud valikut: pole võimalu, et *trans*-regulatsioon mängib olulisemat rolli mõne teise Fairfax jt uuringus tuvastatud beetainterferooni signaalirajas osaleva geeni ekspressioonil [10].

Tulemuste analüüsil ei saa mainimata jätta *IFNBI* lookuse *trans*-seoste olulist erinevust andmestike kaupa: igas andmestikus oli tugevaim seos erineva geeniga. Erinevuste põhjusteks võivad olla LPSi erinev kontsentratsioon, mõõtmiste ajapunkt, populatsioonide erinevus, mõõtevead või süstemaatilised vead uuringu käigus, aga ka „võitja needuse“ (ingl *winner's curse*) fenomen – esmane positiivne leid geneetilistel uuringutel kipub seose olulisust üle hindama ja järgmistel uuringutel, mis eeldavad seose kehtimist, ei õnnestu sageli sama seost reprodutseerida [33].

7. Kokkuvõte

Käesoleva töö eesmärgiks oli välja selgitada seni tundmata beetainterferooni signaaliraja aktiivsust mõjutavaid geneetilisi variante. Lugejale tutvustati teemaga seonduvaid mõisteid, geeniekspressiooni ja genoomiüleste uuringute olemust ja antud töö eesmärgi saavutamiseks kasutatud metodoloogiat.

Osaliselt Tartu Ülikooli teadusarvutuste keskusel läbi viidud töö käigus töötati nelja populatsiooni ja 685 inimese geneetilise materjaliga, et leida statistiliselt olulised seosed beetainterferooni sünteesiva *IFNBI* geeni lookuses asuvate geneetiliste variantide ja nendega *trans*-seoses olevate geenide vahel. Ülegenoomse assotsiatsiooniuuringu käigus tuvastati, et antud lookuse geneetilised variandid funktsioneerivad *trans*-eQTLidena, mõjutades kõige tugevalt *IFI6* geeni ekspressiooni.

Täiendav ülegenoomne assotsiatsiooniuuring, mille käigus prooviti leida teisi *IFI6* ekspressiooni mõjutavaid ühe nukleotiidi polümorfisme, kinnitas *IFNBI* ja *IFI6* vahelist tugevat seost, kuid ei paljastanud ühtegi *IFNBI* lookuse välist geneetilist varianti, mis oleks nelja andmestiku ülesel metaanalüüsil statistiliselt oluliselt markergeeni ekspressiooniga seotud. Ka *IFNBI* mõju regresseerimine ning tugevaima signaali eest vastutava andmestiku uurimist väljajätmine ei viinud uute leidudeni. Tulevikus on võimalik valimit täiendada stimuleeritud makrofaagide (Alasoo jt 2018, Nedelec jt 2016) ja dendriitrakkude (Ye jt 2018) andmestikega, aga kuna neis andmestikes erinevad nii rakutüüp kui ka stiimul, ei pruugi see tulemust parandada [34,35,36].

Ehkki töö käigus ei õnnestunud leida uusi beetainterferooni signaalirajas osalevaid geneetilisi variante, saab töös kasutatud metoodikat tulevikus rakendada täiendavateks uuringuteks, näiteks testides teisi signaaliraja genee kogu genoomi SNPide vastu.

Tänuavaldused

Soovin tänada oma juhendajat Kaur Alasood kiirete ja konstruktiivsete vastuste eest kõigile minu küsimustele ning töö edenemise vastu pidevalt huvi tundmise eest.

Tänan ka Tartu Ülikooli teadusarvutuste keskust mulle arvutusmahu eraldamise eest.

Viidatud kirjandus

- 1 Hartl, Daniel L and Jones, Elizabeth W. *Genetics: Analysis of Genes and Genomes. Fifth Edition*. Jones & Bartlett Pub., Sudbury, Massachusetts, 2001.
- 2 Pray, Leslie A. DNA Replication and Causes of Mutation. *Nature Education*, 1, 1 (2008), 214.
- 3 Lodish, Harvey, Berk, Arnold, Matsudaira, Paul et al. *Molecular Cell Biology. Fifth Edition*. W. H. Freeman, New York, 2008.
- 4 Clancy, Suzanne. Genetic Mutation. *Nature Education*, 1, 1 (2008), 187.
- 5 Clancy, Suzanne and Brown, William. Translation: DNA to mRNA to protein. *Nature Education*, 1, 1 (2008), 101.
- 6 Cano-Gamez, Eddie and Trynka, Gosia. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics*, 11 (2020).
- 7 Albert, Frank W and Kruglyak, Leonid. The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, 16 (2015), 197-212.
- 8 Manolio, Teri A. Genomewide association studies and assessment of the risk of disease. *New England journal of medicine*, 363, 2 (2010), 166-176.
- 9 Haji Abdolvahab, Mohadeseh, Mofrad, Mohammad RK, and Schellekens, Harriet. Interferon Beta: From Molecular Level to Therapeutic Effects. *International Review of Cell and Molecular Biology*, 326 (2016), 343-372.
- 10 Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., Knight, J. C. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, 343, 6175 (March 07, 2014), 3.
- 11 Quach, H., Rotival, M., Pothlichet, J., Loh, Y.H.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M. ja Deschamps, M. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell*, 167, 3 (2016), 643-656.
- 12 Tambets, Ralf. <https://github.com/ralf-tambets/bachelor-thesis> (kõlastatud 06.05.2021).

- 13 Rees, David C, Williams, Thomas N, and Gladwin, Mark T. Sickle-cell disease. *The Lancet*, 376, 9757 (2010), 2018-2031.
- 14 Karki, Roshan, Pandya, Deep, Elston, Robert C, and Ferlini, Cristiano. Defining "mutation" and "polymorphism" in the era of personal genomics. *BMC Med Genomics*, 8, 37 (2015).
- 15 Chial, Heidi. DNA Sequencing Technologies Key to the Human Genome Project. *Nature Education*, 1, 1 (2008), 219.
- 16 Park, Sang Tae and Kim, Jayoung. Trends in next-generation sequencing and a new era for whole genome sequencing. *International neurology journal*, 20, Suppl 2 (2016), S76.
- 17 Marees, Andries T, de Kluiver, Hilde, Stringer, Sven, Vorspan, Florence, Curis, Emmanuel, Marie-Claire, Cynthia, and Derks, Eske M. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, 27, 2 (February 2018).
- 18 Tam, Vivian, Patel, Nikunj, Turcotte, Michelle, Bossé, Yohan, Paré, Guillaume, and Meyre, David. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 20 (May 2019), 467-484.
- 19 Wolen, Aaron R and Miles, Michael F. Identifying Gene Networks Underlying the Neurobiology of Ethanol and Alcoholism. *Alcohol Research : Current Reviews*, 34, 3 (November 2012), 306-317.
- 20 Stelzer, G, R, Rosen, I, Plaschkes et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis. *Current Protocols in Bioinformatics* (2016).
- 21 Shabalin, Andrey A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, 28, 10 (May 2012), 1353-1358.
- 22 Zheng, Xiuwen, Levine, David, Shen, Jess, Gogarten, Stephanie M, Laurie, Cathy, and Weir, Bruce S. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 24 (December 2012), 3326-3328.
- 23 dplyr. <https://dplyr.tidyverse.org/> (kūlastatud 05.05.2021).
- 24 ggplot2. <https://ggplot2.tidyverse.org/> (kūlastatud 05.05.2021).
- 25 qqman. <https://CRAN.R-project.org/package=qqman> (kūlastatud 05.05.2021).

- 26 BCFtools. <http://www.htslib.org/doc/bcftools.html> (külastatud 05.05.2021).
- 27 Willer, Cristen J, Li, Yun, and Abecasis, Goncalo R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26, 17 (September 2010), 2190-2191.
- 28 PuTTY. <https://documentation.help/PuTTY/> (külastatud 05.05.2021).
- 29 FileZilla. <https://wiki.filezilla-project.org/Documentation> (külastatud 05.05.2021).
- 30 Wang, Bokai, Zhou, Zhirou, Wang, Hongyue, Tu, Xin M, and Feng, Changyong. The p-value and model specification in statistics. *General Psychiatry*, 32, 3 (July 2019).
- 31 Ehret, Georg B. Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension. *Curr Hypertens Rep*, 12, 1 (2011), 17-25.
- 32 Yang, J, Loos, Ruth JF, Powell, Joseph E, Medland, Sarah E et al. FTO genotype is associated with phenotypic variability of body mass index. *Nature*, 490 (September 2012), 267-272.
- 33 Xiao, Rui and Boehnke, Michael. Quantifying and correcting for the winner's curse in genetic association studies. *Genet Epidemiol.*, 33, 5 (July 2010), 453-362.
- 34 Alasoo, Kaur, Rodrigues, Julia, Mukhopadhyay, Subhankar et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nature Genetics*, 50 (January 2018), 424-431.
- 35 Nédélec, Yohann, Sanz, Joaquin, Baharian, Golshid et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*, 167, 3 (October 2016), 657-669.
- 36 Ye, Jimmie C, Chen, Jenny, Villani, Alexandra-Chloé et al. Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. *Genome Research*, 28 (November 2018), 1812-1825.

Lisad

I. *IFI6* tugevaimad assotsiatsioonid andmestike kaupa

Tabel 2. *IFI6* tugevaimad assotsiatsioonid Fairfax_2014 andmestikus.

<i>nr</i>	<i>Geen</i>	<i>SNPi kood</i>	<i>beta</i>	<i>t-statistik</i>	<i>p-väärtus</i>	<i>FDR</i>
1	IFI6	chr9_20827115_G_C	-0.50335	-8.99364	2.2287E-17	1.1308E-10
2	IFI6	chr9_20818520_A_G	-0.50430	-8.97608	2.5312E-17	1.1308E-10
3	IFI6	chr9_20916829_A_C	-0.52033	-8.89337	4.6007E-17	1.1308E-10
4	IFI6	chr9_20976060_A_G	-0.51581	-8.87681	5.1832E-17	1.1308E-10
5	IFI6	chr9_20999771_T_C	-0.51607	-8.86818	5.5153E-17	1.1308E-10
6	IFI6	chr9_20975007_C_A	-0.51470	-8.86712	5.5577E-17	1.1308E-10
7	IFI6	chr9_20976299_C_T	-0.51433	-8.86278	5.7335E-17	1.1308E-10
8	IFI6	chr9_20840034_C_T	-0.51487	-8.85740	5.9600E-17	1.1308E-10
9	IFI6	chr9_20978874_A_G	-0.51432	-8.85736	5.9614E-17	1.1308E-10
10	IFI6	chr9_20977813_G_T	-0.51382	-8.85679	5.9860E-17	1.1308E-10
11	IFI6	chr9_20974347_T_C	-0.52254	-8.75707	1.2229E-16	2.0064E-10
12	IFI6	chr9_20927341_A_C	0.51517	8.74691	1.3149E-16	2.0064E-10
13	IFI6	chr9_20931306_G_A	0.51376	8.72393	1.5490E-16	2.0064E-10
14	IFI6	chr9_20910318_T_C	-0.51290	-8.72337	1.5552E-16	2.0064E-10
15	IFI6	chr9_20916403_C_T	-0.51285	-8.70996	1.7111E-16	2.0064E-10
16	IFI6	chr9_20961907_C_G	-0.50880	-8.70261	1.8029E-16	2.0064E-10
17	IFI6	chr9_20963576_T_A	-0.50779	-8.70240	1.8056E-16	2.0064E-10
18	IFI6	chr9_20926792_A_G	0.51135	8.69064	1.9631E-16	2.0602E-10
19	IFI6	chr9_20987878_A_G	-0.49847	-8.59002	4.0042E-16	3.6466E-10
20	IFI6	chr9_20983817_A_G	-0.49787	-8.58293	4.2097E-16	3.6466E-10
21	IFI6	chr9_20984804_A_T	-0.49786	-8.58218	4.2318E-16	3.6466E-10
22	IFI6	chr9_20985827_A_T	-0.49789	-8.58168	4.2469E-16	3.6466E-10
23	IFI6	chr9_20822667_A_C	-0.48641	-8.56087	4.9179E-16	3.7593E-10
24	IFI6	chr9_20822816_AATGATC_A	-0.48683	-8.55785	5.0237E-16	3.7593E-10
25	IFI6	chr9_20991795_A_C	-0.49721	-8.55771	5.0288E-16	3.7593E-10
26	IFI6	chr9_20822079_G_A	-0.48563	-8.55366	5.1742E-16	3.7593E-10
27	IFI6	chr9_20832096_G_T	-0.48777	-8.54813	5.3796E-16	3.7638E-10
28	IFI6	chr9_20945208_G_C	0.50975	8.50362	7.3553E-16	4.8701E-10
29	IFI6	chr9_20941310_G_C	0.50997	8.50129	7.4765E-16	4.8701E-10
30	IFI6	chr9_20905385_A_G	-0.50845	-8.47991	8.6858E-16	5.2076E-10

Tabel 3. *IFI6* tugevaimad assotsiatsioonid Fairfax_2018 andmestikus.

<i>nr</i>	<i>Geen</i>	<i>SNPi kood</i>	<i>beta</i>	<i>t-statistik</i>	<i>p-väärtus</i>	<i>FDR</i>
1	IFI6	chr7_118802262_G_T	2.0985	5.2149	5.3543E-07	0.92528
2	IFI6	chr7_118781404_G_A	2.1562	5.1446	7.4080E-07	0.92528
3	IFI6	chr6_150403950_TA_T	0.7266	5.0390	1.2003E-06	0.92528
4	IFI6	chr7_118863479_G_C	2.0310	5.0102	1.3671E-06	0.92528
5	IFI6	chr10_62687658_T_A	-1.0346	-5.0084	1.3786E-06	0.92528
6	IFI6	chr7_118830264_C_T	2.0058	5.0064	1.3907E-06	0.92528
7	IFI6	chr7_118976360_T_C	2.0328	4.9834	1.5432E-06	0.92528
8	IFI6	chr7_118849983_A_T	1.9192	4.9713	1.6290E-06	0.92528
9	IFI6	chr16_28509678_T_TTG	1.4908	4.9123	2.1222E-06	0.92528
10	IFI6	chr14_95364764_T_C	-1.0626	-4.8925	2.3180E-06	0.92528
11	IFI6	chr14_95364320_G_A	-1.0143	-4.8017	3.4619E-06	0.92528
12	IFI6	chr6_1530882_T_C	-2.2689	-4.7693	3.9892E-06	0.92528
13	IFI6	chr22_45100474_C_T	-0.7482	-4.7683	4.0070E-06	0.92528
14	IFI6	chr7_119095886_G_A	2.0367	4.7211	4.9215E-06	0.92528
15	IFI6	chr7_119109679_A_AT	2.0318	4.7087	5.1925E-06	0.92528
16	IFI6	chr7_119107086_G_A	2.0303	4.7083	5.2025E-06	0.92528
17	IFI6	chr20_38781455_C_CTTT	1.0182	4.7004	5.3829E-06	0.92528
18	IFI6	chr19_43759167_T_C	1.1418	4.6975	5.4509E-06	0.92528
19	IFI6	chr20_38766975_G_A	0.8578	4.6883	5.6711E-06	0.92528
20	IFI6	chr12_121329685_G_A	-2.0180	-4.6878	5.6851E-06	0.92528
21	IFI6	chr20_38752919_C_CT	0.8631	4.6854	5.7439E-06	0.92528
22	IFI6	chr7_119084636_C_T	2.0197	4.6786	5.9144E-06	0.92528
23	IFI6	chr20_38770845_C_T	0.8554	4.6699	6.1417E-06	0.92528
24	IFI6	chr20_38770922_A_AGGGAGGT	0.9485	4.6615	6.3675E-06	0.92528
25	IFI6	chr20_38750439_G_A	0.8532	4.6611	6.3785E-06	0.92528
26	IFI6	chr20_38784893_G_A	0.9500	4.6602	6.4013E-06	0.92528
27	IFI6	chr10_63640091_T_C	-0.7966	-4.6508	6.6669E-06	0.92528
28	IFI6	chr2_196061474_C_A	-1.2921	-4.6486	6.7298E-06	0.92528
29	IFI6	chr10_63639087_GA_G	-0.8462	-4.6472	6.7696E-06	0.92528
30	IFI6	chr20_38773752_C_T	0.8528	4.6468	6.7828E-06	0.92528

Tabel 4. *IFI6* tugevaimad assotsiatsioonid Quach_2016 Euroopa populatsioonis.

<i>nr</i>	<i>Geen</i>	<i>SNPi kood</i>	<i>beta</i>	<i>t-statistik</i>	<i>p-väärtus</i>	<i>FDR</i>
1	IFI6	chr1_96055738_T_C	1.3994	5.8990	6.6607E-08	0.08911
2	IFI6	chr1_96054048_G_A	1.3912	5.8329	8.8694E-08	0.08911
3	IFI6	chr1_96082512_G_A	1.3900	5.8326	8.8837E-08	0.08911
4	IFI6	chr1_96084044_C_T	1.3900	5.8326	8.8837E-08	0.08911
5	IFI6	chr1_96098014_C_T	1.3900	5.8326	8.8837E-08	0.08911
6	IFI6	chr1_96113258_G_A	1.3900	5.8326	8.8837E-08	0.08911
7	IFI6	chr1_96056311_C_T	1.3901	5.8325	8.8871E-08	0.08911
8	IFI6	chr1_96071726_A_G	1.3901	5.8323	8.8930E-08	0.08911
9	IFI6	chr1_96129354_C_T	1.3899	5.8319	8.9081E-08	0.08911
10	IFI6	chr1_95825717_G_A	1.3954	5.4750	4.0855E-07	0.36784
11	IFI6	chr9_21009192_G_A	-0.8345	-5.2732	9.4841E-07	0.65826
12	IFI6	chr9_20832096_G_T	-0.7538	-5.2157	1.2028E-06	0.65826
13	IFI6	chr1_96123116_T_G	0.8432	5.0728	2.1581E-06	0.65826
14	IFI6	chr9_20814143_C_T	-0.7516	-5.0452	2.4140E-06	0.65826
15	IFI6	chr9_20822816_AATGATC_A	-0.7327	-5.0128	2.7529E-06	0.65826
16	IFI6	chr9_20822079_G_A	-0.7313	-5.0124	2.7575E-06	0.65826
17	IFI6	chr9_21030852_G_A	-0.8134	-4.9897	3.0219E-06	0.65826
18	IFI6	chr11_72653103_G_C	0.8509	4.9772	3.1782E-06	0.65826
19	IFI6	chr9_21021953_G_A	-0.7943	-4.9652	3.3349E-06	0.65826
20	IFI6	chr9_21014629_T_C	-0.7943	-4.9649	3.3389E-06	0.65826
21	IFI6	chr9_21017885_T_C	-0.7943	-4.9649	3.3389E-06	0.65826
22	IFI6	chr9_21018551_G_T	-0.7943	-4.9649	3.3389E-06	0.65826
23	IFI6	chr9_21018812_C_A	-0.7943	-4.9649	3.3389E-06	0.65826
24	IFI6	chr11_72653033_T_C	0.8412	4.9580	3.4328E-06	0.65826
25	IFI6	chr9_21017241_A_G	-0.7935	-4.9566	3.4519E-06	0.65826
26	IFI6	chr9_20822667_A_C	-0.7247	-4.9555	3.4681E-06	0.65826
27	IFI6	chr12_101066503_T_C	-1.4724	-4.9351	3.7644E-06	0.65826
28	IFI6	chr9_20827115_G_C	-0.7049	-4.9341	3.7793E-06	0.65826
29	IFI6	chr4_16981523_G_A	1.8437	4.8893	4.5203E-06	0.65826
30	IFI6	chr9_20828706_T_A	-0.7553	-4.8440	5.4134E-06	0.65826

Tabel 5. *IFI6* tugevaimad assotsiatsioonid Quach_2016 Aafrika populatsioonis.

<i>nr</i>	<i>Geen</i>	<i>SNPi kood</i>	<i>beeta</i>	<i>t-statistik</i>	<i>p-väärtus</i>	<i>FDR</i>
1	IFI6	chr9_21008726_A_G	-0.89831	-6.7499	2.1366E-09	0.00533
2	IFI6	chr9_21012171_G_A	-0.89831	-6.7499	2.1366E-09	0.00533
3	IFI6	chr9_21012746_G_A	-0.89831	-6.7499	2.1366E-09	0.00533
4	IFI6	chr9_21014629_T_C	-0.89831	-6.7499	2.1366E-09	0.00533
5	IFI6	chr9_21018551_G_T	-0.89831	-6.7499	2.1366E-09	0.00533
6	IFI6	chr9_21017241_A_G	-0.90013	-6.7365	2.2660E-09	0.00533
7	IFI6	chr9_21017885_T_C	-0.84958	-6.5966	4.1791E-09	0.00843
8	IFI6	chr9_21018812_C_A	-0.84061	-6.0836	3.8264E-08	0.06754
9	IFI6	chr9_21009192_G_A	-0.88712	-5.9674	6.2702E-08	0.09838
10	IFI6	chr6_4562395_T_C	-0.69421	-5.7512	1.5587E-07	0.22010
11	IFI6	chr9_21053045_G_T	-0.83769	-5.6737	2.1532E-07	0.27641
12	IFI6	chr9_21052927_A_G	-0.80492	-5.4595	5.2174E-07	0.61395
13	IFI6	chr9_21045063_C_T	-0.83416	-5.3690	7.5545E-07	0.73850
14	IFI6	chr5_161485329_TA_T	-1.89289	-5.3476	8.2414E-07	0.73850
15	IFI6	chr1_36936812_A_AT	-0.85086	-5.3474	8.2472E-07	0.73850
16	IFI6	chr9_21040184_C_A	-0.83155	-5.3438	8.3678E-07	0.73850
17	IFI6	chr9_21008378_T_C	-0.71821	-5.3072	9.7073E-07	0.76520
18	IFI6	chr16_846679_T_C	0.85045	5.3061	9.7542E-07	0.76520
19	IFI6	chr9_21048542_A_C	-0.77996	-5.2241	1.3580E-06	0.95740
20	IFI6	chr9_21044022_T_G	-0.81437	-5.2089	1.4440E-06	0.95740
21	IFI6	chr9_21021953_G_A	-0.79769	-5.1991	1.5021E-06	0.95740
22	IFI6	chr10_99788711_G_A	0.71232	5.1862	1.5815E-06	0.95740
23	IFI6	chr10_99787906_T_A	0.70911	5.1853	1.5875E-06	0.95740
24	IFI6	chr9_21048412_G_A	-0.61933	-5.1400	1.9031E-06	0.95740
25	IFI6	chr19_44918507_A_G	-1.94080	-5.1182	2.0763E-06	0.95740
26	IFI6	chr9_131919363_G_A	-0.76684	-5.0605	2.6129E-06	0.95740
27	IFI6	chr9_21047632_G_T	-0.60990	-5.0580	2.6383E-06	0.95740
28	IFI6	chr9_84467451_CTGTGTGTGG_C	1.22113	5.0514	2.7084E-06	0.95740
29	IFI6	chr22_17835055_C_T	0.81513	5.0165	3.1106E-06	0.95740
30	IFI6	chr9_21023447_C_T	-0.71605	-5.0098	3.1936E-06	0.95740

II. *IFI6* tugevaimad assotsiatsioonid metaanalüüsil

Tabel 6. *IFI6* tugevaimad assotsiatsioonid nelja andmestiku metanalüüsil ilma *IFNBI* mõju regresseerimata.

<i>nr</i>	<i>Geen</i>	<i>SNPi kood</i>	<i>Kaal</i>	<i>z-skoor</i>	<i>p-väärtus</i>	<i>Suund</i>
1	<i>IFI6</i>	chr9_21014629_T_C	685	-11.129	9.0740E-29	----
2	<i>IFI6</i>	chr9_21017241_A_G	685	-11.084	1.4950E-28	----
3	<i>IFI6</i>	chr9_21017885_T_C	685	-11.072	1.7200E-28	----
4	<i>IFI6</i>	chr9_21009192_G_A	685	11.028	2.7960E-28	++++
5	<i>IFI6</i>	chr9_21018551_G_T	509	10.804	3.2820E-27	+?++
6	<i>IFI6</i>	chr9_21018812_C_A	509	10.601	2.9380E-26	+?++
7	<i>IFI6</i>	chr9_21020115_T_A	685	10.528	6.4300E-26	++++
8	<i>IFI6</i>	chr9_21021247_T_C	685	-10.528	6.4470E-26	----
9	<i>IFI6</i>	chr9_21022501_A_C	685	-10.527	6.4890E-26	----
10	<i>IFI6</i>	chr9_21023447_C_T	685	10.527	6.4980E-26	++++
11	<i>IFI6</i>	chr9_21024671_G_A	685	10.526	6.5480E-26	++++
12	<i>IFI6</i>	chr9_21021953_G_A	509	10.313	6.1830E-25	+?++
13	<i>IFI6</i>	chr9_21028637_C_T	685	10.311	6.2540E-25	++++
14	<i>IFI6</i>	chr9_21029331_T_C	685	-10.252	1.1540E-24	----
15	<i>IFI6</i>	chr9_21020233_GAC_G	685	10.195	2.0910E-24	++++
16	<i>IFI6</i>	chr9_21021386_A_G	509	-10.125	4.2900E-24	-?--
17	<i>IFI6</i>	chr9_21022870_A_C	509	-10.121	4.4380E-24	-?--
18	<i>IFI6</i>	chr9_20999771_T_C	685	-9.656	4.6580E-22	----
19	<i>IFI6</i>	chr9_20832096_G_T	685	9.654	4.7220E-22	++++
20	<i>IFI6</i>	chr9_20822816_AATGATC_A	685	9.608	7.4000E-22	++++
21	<i>IFI6</i>	chr9_20822079_G_A	685	9.535	1.5040E-21	++++
22	<i>IFI6</i>	chr9_20827115_G_C	685	9.486	2.3900E-21	++++
23	<i>IFI6</i>	chr9_20978874_A_G	685	-9.483	2.4600E-21	----
24	<i>IFI6</i>	chr9_20822667_A_C	685	-9.479	2.5620E-21	----
25	<i>IFI6</i>	chr9_20908657_C_T	685	9.469	2.8370E-21	++++
26	<i>IFI6</i>	chr9_20829417_GT_G	685	9.468	2.8530E-21	++++
27	<i>IFI6</i>	chr9_20818520_A_G	685	-9.467	2.8740E-21	----
28	<i>IFI6</i>	chr9_21030852_G_A	509	9.444	3.5940E-21	+?++
29	<i>IFI6</i>	chr9_20975007_C_A	685	9.398	5.5560E-21	++++
30	<i>IFI6</i>	chr9_20828706_T_A	685	9.369	7.3540E-21	++++

Tabel 7. *IFI6* tugevaimad assotsiatsioonid nelja andmestiku metaanalüüsil pärast *IFNB1* mõju regresseerimist.

<i>nr</i>	<i>Geen</i>	<i>SNPi kood</i>	<i>Kaal</i>	<i>z-skoor</i>	<i>p-väärtus</i>	<i>Suund</i>
1	IFI6	chr9_21014629_T_C	685	-5.268	1.3760E-07	----
2	IFI6	chr9_21009192_G_A	685	5.230	1.6990E-07	++++
3	IFI6	chr9_21017241_A_G	685	-5.208	1.9110E-07	----
4	IFI6	chr9_21017885_T_C	685	-5.196	2.0410E-07	----
5	IFI6	chr9_104701523_C_T	597	-4.954	7.2860E-07	---?
6	IFI6	chr19_7114716_T_C	597	-4.907	9.2410E-07	---?
7	IFI6	chr11_57161535_G_A	509	-4.879	1.0650E-06	-?--
8	IFI6	chr16_59086126_A_G	184	4.866	1.1410E-06	??++
9	IFI6	chr10_22205724_T_C	685	4.803	1.5600E-06	++++
10	IFI6	chr1_62139120_G_C	685	4.801	1.5770E-06	++++
11	IFI6	chr1_170651137_AT_A	325	-4.785	1.7090E-06	-???
12	IFI6	chr12_90165264_CTT_C	509	4.763	1.9070E-06	+?++
13	IFI6	chr20_23985833_C_T	325	-4.759	1.9430E-06	-???
14	IFI6	chr3_7801543_G_C	509	4.758	1.9560E-06	+?++
15	IFI6	chr20_62442759_C_T	685	-4.736	2.1820E-06	----
16	IFI6	chr12_78582973_C_A	96	-4.736	2.1830E-06	??-?
17	IFI6	chr13_40632799_G_A	501	-4.732	2.2250E-06	--??
18	IFI6	chr12_90165821_G_A	509	-4.721	2.3460E-06	-?--
19	IFI6	chr12_90151443_C_T	509	-4.710	2.4720E-06	-?--
20	IFI6	chr9_104704444_C_G	597	4.706	2.5310E-06	+++?
21	IFI6	chr3_7799848_A_T	509	-4.703	2.5620E-06	-?--
22	IFI6	chr7_20084480_A_G	597	4.703	2.5690E-06	+++?
23	IFI6	chr9_91518688_C_T	685	4.701	2.5830E-06	++++
24	IFI6	chr16_28509678_T_TTG	176	4.699	2.6090E-06	?+??
25	IFI6	chr9_91511064_G_A	685	4.699	2.6140E-06	++++
26	IFI6	chr9_104771686_C_A	597	-4.696	2.6580E-06	---?
27	IFI6	chr11_57153112_C_T	685	-4.690	2.7330E-06	----
28	IFI6	chr13_22691798_CA_C	509	-4.686	2.7800E-06	-?--
29	IFI6	chr12_90153130_G_T	509	-4.686	2.7910E-06	-?--
30	IFI6	chr12_90153022_G_A	509	-4.686	2.7920E-06	-?--

Tabel 8. *IFI6* tugevaimad assotsiatsioonid kolme andmestiku metaanalüüsil (Fairfax_2014 eemaldatud).

<i>nr</i>	<i>Geen</i>	<i>SNPi kood</i>	<i>Kaal</i>	<i>z-skoor</i>	<i>p-väärtus</i>	<i>Suund</i>
1	IFI6	chr9_21014629_T_C	360	-7.8400	4.489E-15	---
2	IFI6	chr9_21017241_A_G	360	-7.7960	6.372E-15	---
3	IFI6	chr9_21017885_T_C	360	-7.7860	6.912E-15	---
4	IFI6	chr9_21009192_G_A	360	7.6520	1.977E-14	+++
5	IFI6	chr9_21018551_G_T	184	7.4980	6.464E-14	?++
6	IFI6	chr9_21018812_C_A	184	7.1610	8.034E-13	?++
7	IFI6	chr9_21021247_T_C	360	-7.0360	1.971E-12	---
8	IFI6	chr9_21022501_A_C	360	-7.0360	1.971E-12	---
9	IFI6	chr9_21023447_C_T	360	7.0360	1.971E-12	+++
10	IFI6	chr9_21020115_T_A	360	7.0360	1.976E-12	+++
11	IFI6	chr9_21024671_G_A	360	7.0360	1.976E-12	+++
12	IFI6	chr9_21021953_G_A	184	6.6850	2.307E-11	?++
13	IFI6	chr9_21028637_C_T	360	6.6730	2.502E-11	+++
14	IFI6	chr9_21029331_T_C	360	-6.6580	2.767E-11	---
15	IFI6	chr9_21020233_GAC_G	360	6.5770	4.807E-11	+++
16	IFI6	chr9_21021386_A_G	184	-6.3700	1.892E-10	?--
17	IFI6	chr9_21022870_A_C	184	-6.3700	1.893E-10	?--
18	IFI6	chr9_21008726_A_G	360	-6.0800	1.204E-09	---
19	IFI6	chr9_21012171_G_A	360	6.0020	1.954E-09	+++
20	IFI6	chr9_21012746_G_A	360	5.9910	2.090E-09	+++
21	IFI6	chr9_20832096_G_T	360	5.6180	1.928E-08	+++
22	IFI6	chr9_21016782_C_CT	360	-5.6090	2.030E-08	---
23	IFI6	chr9_21019651_T_G	360	5.6080	2.047E-08	+++
24	IFI6	chr9_21016930_C_G	360	5.6060	2.066E-08	+++
25	IFI6	chr9_21028135_C_CA	184	5.5950	2.203E-08	?++
26	IFI6	chr9_21056086_T_C	360	-5.5890	2.277E-08	---
27	IFI6	chr9_20908657_C_T	360	5.5760	2.465E-08	+++
28	IFI6	chr9_20829417_GT_G	360	5.5700	2.553E-08	+++
29	IFI6	chr9_20822816_AATGATC_A	360	5.5470	2.910E-08	+++
30	IFI6	chr9_21008378_T_C	360	-5.4900	4.030E-08	---

III. Litsents

Lihlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Ralf Tambets**,

1. Annan Tartu Ülikoolile tasuta loa (lihlitsentsi) minu loodud teose „**Beetainterferooni signaaliraja aktiivsust mõjutavate geneetiliste variantide tuvastamine**“, mille juhendaja on dr. Kaur Alasoo, 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.

Ralf Tambets

06.05.2021