

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT

BIOTEHNOLOOGIA ÕPPETOOL

Timo Tõnis Sikka

Immuunoglobuliin G glükosüleerimist mõjutavad DNA variatsioonid Eesti populatsioonis

Magistritöö (30 EAP)

Juhendaja Tõnu Esko, PhD

TARTU 2016

INFOLEHT

Immuunoglobuliin G glükosüleerimist mõjutavad DNA variatsioonid Eesti populatsioonis

Valgud on suured kompleksed molekulid, mis omavad mitmeid olulisi ja elutähtsaid rolle organismis. Erinevate ülesannete täitmiseks läbivad valgud pärast sünteesi ka erinevaid modifikatsioone, nagu näiteks glükosüleerimine. Antud töö eesmärgiks oli analüüsida geneetiliste järjestusvariantide mõju immuunoglobuliin G glükosüleerimisele. Töö tulemusena leiti hulgaliselt ülegenoomselt statistiliselt olulisi assotsiatsioone, mis omasid seost immuunoglobuliini glükolüsatsiooniga. Uuringud kinnitasid ülegenoomsete assotsiatsiooniuuringute efektiivsust komplekstunnuste uurimisel.

Märksõmad: GWAS, immuunoglobuliin G, glükaan, glükosüleerimine

B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

The role of genetic variants on immunoglobulin G glycosylation in Estonian population

Proteins are complex molecules that have several functions. The vast repertoire of functions is obtained through postranslational modifications, such as glycosylation. The aim of this study was to analyse the role of genetics variants on immunoglobulin G glycosylation. Many genetics marker associations of genome-wide significance with immunoglobulin G glycosylation were found. This study also showed the suitability of genome-wide association studies for analysing complex traits.

Key words: GWAS, immunoglobulin H, glycan, glycosylation

B110 Informatics, medical informatics, biomathematics, biometrics

SISUKORD

KASUTATUD LÜHENDID.....	4
SISSEJUHATUS.....	5
1. KIRJANDUSE ÜLEVAADE.....	6
1.1. Valkude glükosüleerimine.....	6
1.2. Immuunoglobuliin G.....	6
1.2.1. IgG glükosüleerimine ja selle mõju.....	7
1.3. Ülegenoomsed assotsiatsiooniuuringud.....	8
1.3.1 Markerid ülegenoomsetes assotsiatsiooniuuringutes.....	9
1.3.2. Imputeerimine.....	9
1.3.3. Valimi suurus ja võimsus.....	10
1.3.4. Leiud ülegenoomsetest assotsiatsiooniuuringutest.....	11
1.3.5. Uute populatsioonipõhiste referentspaneelide roll GWASides.....	12
2. EKSPERIMENTAALOSA.....	14
2.1. Töö eesmärgid.....	14
2.2. Materjal ja meetodika.....	14
2.2.1. Valim.....	14
2.2.2. Genotüpiseerimine ja kvaliteedikontroll.....	15
2.2.3. Imputeerimine.....	16
2.2.5 Assotsiatsioonianalüüsid.....	17
2.3. Tulemused.....	18
2.3.1. GWAS tulemused kasutades HapMap2 referentspaneeli.....	18
2.3.1. GWASi tulemused kasutades 1000 Genoomi referentspaneeli.....	20
2.3.3. GWASi tulemused kasutades Eesti referentspaneeli.....	22
2.3.4. Sekveneerimisandmete assotsiatsioonanalüüside tulemused.....	23
2.4. Arutelu.....	24
KOKKUVÕTE.....	28
SUMMARY.....	29
TÄNUAVALDUSED.....	30
KIRJANDUSE LOETELU.....	31
KASUTATUD VEEBIAADRESSID.....	36
LISAD.....	37
Lisa 1.....	37
Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.....	39

KASUTATUD LÜHENDID

<i>CR</i>	<i>Call rate</i> , edukalt genotüpiseeritud indiviidide protsentuaalne osakaal markeril
<i>EAF</i>	<i>Efect allele frequency</i> , efektalleeli sagedus
<i>GWAS</i>	<i>Genome-wide association study</i> , ülegenoomne assotsatsiooniuring
<i>HWE</i>	<i>Hardy-Weidenberg equilibrium</i> , Hardy-Weidenbergi tasakaalustatus
<i>IgG</i>	Immuunoglobuliin G
<i>LD</i>	<i>Linkage disequilibrium</i> , ahelduse tasakaalustamatus
<i>MAF</i>	<i>Minor allele frequency</i> , minoorse alleeli sagedus
<i>OR</i>	<i>Odds ratio</i> , šansside suhe

SISSEJUHATUS

Valgud on suured kompleksed molekulid, mis omavad mitmeid olulisi ja elutähtsaid rolle organismis. Valke on tarvis organismi kudede ja organite struktuuri, funktsiooni ning regulatsiooni hoidmiseks ja täitmiseks.

Erinevate ülesannete täitmiseks läbivad valgud pärast sünteesi ka erinevaid modifikatsioone. Üks olulisi post-translatsioonilisi modifikatsiooniprotsesse on glükosüleerimine, kus valgu polüpeptiidsele ahelale lisatakse oligosahhariide ehk glükaane. Glükosüleerimine tulemusena muutub valgu struktuur ja funktsioon – sellest tulenevalt on glükosüleerimisel tähtis biofüsioloogiline roll organismides.

Üheks enimuuritud valguks, mille glükosüleerimise mõju on analüüsitud, on immuunoglobuliin G. Seoseid erinevate haiguste ja glükosüleerimise vahel on raporteeritud juba 30 aastat tagasi. Samas pole seni väga hästi uuritud glükosüleerimise geneetiline taust, mis on olnud tingitud piirangutest glükaanide taseme mõõtmisel bioloogilistest proovidest. Viimase 10 aasta jooksul toimunud arengud tehnoloogias võimaldavad nüüdseks usaldusväärset ja kõrge läbilaskevõimega glükaanide taset määrata, mis ühtlasi on viinud arenguteni glükoomika valdkonnas.

Eelmainitust tulenevalt on käesoleva magistritöö eesmärgiks hinnata immuunoglobuliin G glükaanide taseme ning geneetiliste järjestusvariantide seost Eesti kohordis kasutades Tartu Ülikooli Eesti Geenivaramu geenidoonorite andmeid ja ülegenoomseid geneetilisi profiile.

Märksõnad: immuunoglobuliin G, glükaan, glükosüleerimine, GWAS

1. KIRJANDUSE ÜLEVAADE

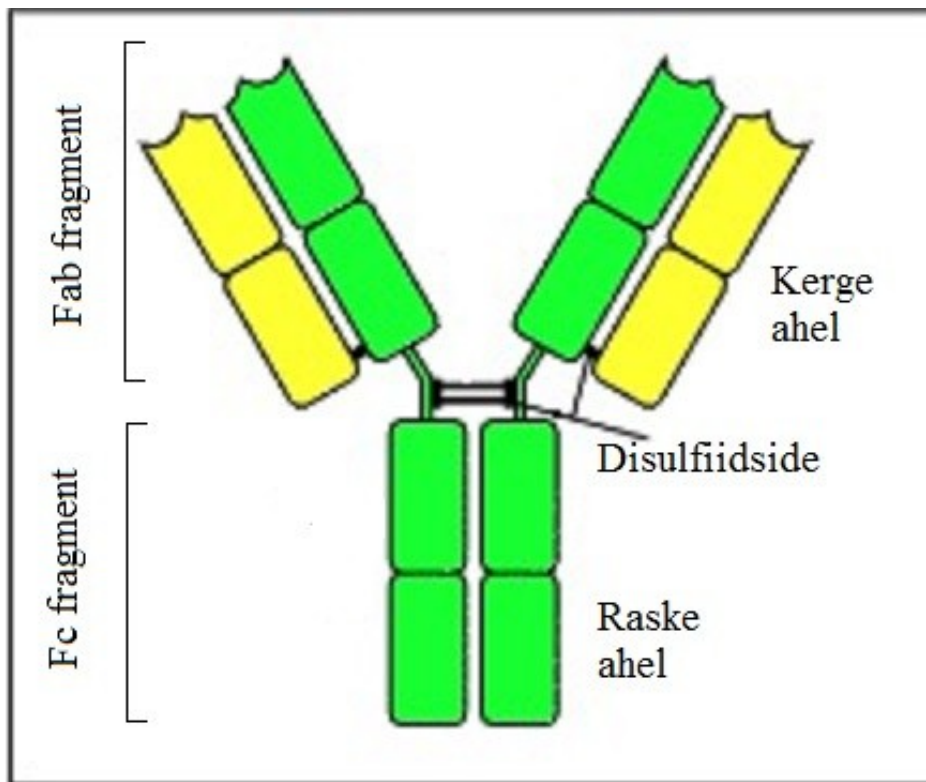
1.1. Valkude glükosüleerimine

Glükosüleerimine on laialt levinud valkude post-translatsiooniline ensümaatilise protsess rakkudes, mille käigus modifitseeritakse glükoproteiinide polüpeptiidsete osade struktuuri ja sellest tulenevalt nende funktsiooni (Opdenakker *et al.*, 1993; Skropeta, 2009). See on tugevalt reguleeritud protsess, mis on omane kõigile eukarüootidele (Marek *et al.*, 1999) ja mida vahendavad glükosültransferaasid (Nairn *et al.*, 2008; J. Wang *et al.*, 2011). Enamik membraani- ja sekreteeritud valke on glükosüleeritud ning glükosüleerimises osaleb sadu glükosültransferaase (Gornik *et al.*, 2012). Seni on kirjeldatud üle 700 nn glükogeeni, kusjuures erinevate glükaanide biosüntees on koe- ja rakuspetsiifiline (Nairn *et al.*, 2008; Yamamoto-Hino *et al.*, 2010).

Üheks glükosüleerimise tüübiks on asparagiin-seotud glükosüleerimine (N-glükosüleerimine), kus polüpeptiidahela asparagiini jäägile lisatakse kovalentselt oligosahhariid(id), mida nimetatakse glükaani(de)ks (Dorner *et al.*, 1987). N-glükosüleerimine on hädavajalik mitmerakuliste organismide eluks ning selle täielik puudumine on embrüonaalselt letaalne (Marek *et al.*, 1999). Samas on mutatsioonid glükaani ahelate modifitseerimise eest vastutavates geenides levinud ning põhjustavad fenotüübilist variatsiooni indiviidide glükaanide kompositsioonis populatsiooni tasemel (Knezevic *et al.*, 2009).

1.2. Immuunoglobuliin G

Üks paremini uuritud glükoproteiine on immuunoglobuliin G (IgG). IgG on suur, ligikaudu 150 kDa suurune kompleksne valk, mis koosneb kahest raskest ja kahest kergest ahelast. Rasked ahelad on omavahel seotud disulfiidsidemega ning kumbki raske ahel on disulfiidsidemega ühendatud kerge ahelaga. Korrektselt pakitud IgG antikeha moodustab Y-kujulise molekuli, kus eristuvad kaks osa – kaks Fab fragmenti (*Fragment antigen binding*) ja üks Fc fragment (*Fragment crystallizable*). Fab regioonid IgG antikehas vastutavad antigeenidega seondumise eest. Fc regioon interakteerub raku pinnal paiknevate Fc retseptorite kaudu rakuga ning võimaldab seeläbi antikehadel immuunsüsteemi aktiveerida. Immuunoglobuliin G skemaatiline struktuur on kujutatud joonisel 1. (Janeway *et al.*, 2001)



Joonis 1. Kujutatud on immuunoglobuliin G (IgG) struktuursed osad. Rohelisega on märgitud IgG rasked ahelad ning kollasega kerged ahelad. Lisaks on joonisel ära toodud Fab ning Fc fragmendid. (Janeway *et al.*, 2001 järgi)

Glükosüleerimise kohalt on oluline just Fc regioon, mille moodustavad rasked ahelad. Kumbki IgG raske ahel kannab Fc regioonis asuva kõrgelt konserveerunud Asn297 jäägi kaudu ühte kovalentselt seotud kaheharulise N-glükaani, mis mõjutavad antikeha stabiilsust ja funktsioone (Skropeta, 2009).

1.2.1. IgG glükosüleerimine ja selle mõju

IgG glükosüleerimine on olulise reguleeriva funktsiooniga – vastavalt sellele, kuidas on glükosüleeritud IgG Fc regioon, muutub ka tema funktsioon. On leitud, et siaalhappe lisamine muudab IgG põletikku soodustava toime põletikuvastaseks (Anthony & Ravetch, 2010; Kaneko *et al.*, 2006). 2011. aastal näitasid Anthony jt, et sialüleeritud glükaanidega IgG-d suruvad T_H2 signaaliraja kaudu maha põletikuvastust ja loovad sellega olulise mehhanismi, kuidas hoida immuunhomeöstaasi ning seeläbi pakkuda terapeutilist kasu autoimmuunhaiguste raviks (Anthony *et al.*, 2011).

Lisaks on leitud, et IgG-de fukosüleerimine aitab ära hoida antikehast sõltuvat rakulist

tsütotoksilisust (*antibody dependant cell-mediated cytotoxicity*, ADCC) (Scanlan *et al.*, 2008). Tavaliselt on ligi 95% tsirkuleerivatest IgG-dest fukosüleeritud glükaani tuumaga, mis muudab Fc regiooni konformatsiooni ning seega selle võimet seonduda FcγRIIIa retseptorile ADCC algatamiseks (Pucic *et al.*, 2011). Fukosüleerimise korral suudavad IgG antikehad seonduda oma retseptoriga ning käivitub ADCC.

Eeltoodu põhjal on näha, et glükosüleerimine väga oluline bioloogiline protsess, kuid teadmised selle geneetilise regulatsiooni kohta on seni olnud kesised, mis on peamiselt tingitud piirangutest glükaanide mõõtmisega bioloogilistest proovidest ja glükaanide matriitsõltumatust sünteesist (Lauc *et al.*, 2010). Üks esimesi põhjalikke glükaanide uuringuid viidi läbi Lauc jt poolt 2010. aastal, kus analüüsiti 2705 indiviidi glükoomi kolmest populatsioonist. Uuringu tulemusena leiti, et kolm geeni – *HNF1α*, *FUT6* ja *FUT8* – mis mõjutasid fukosüültransferaaside ekspressiooni regulatsiooniga N-glükaanide taset inimese vereplasmas (Lauc *et al.*, 2010).

2013. aastal viisid Lauc jt läbi teise ülegenoomse assotsiatsiooniuringu IgG-ga seonduvate glükaanide geneetilise tausta uurimiseks. Antud uuringusse kaasati 4095 indiviidi proovid ning leiti 9 geneetilist lookust, mis olid ülegenoomselt olulised glükaanidega. Neli lookust olid ensüümid, mis mõjutavad otse IgG glükosüleerimist ning sellega tõestati assotsieerunud lookuste bioloogiline roll. Ühtlasi tõestas uuring, et ülegenoomsete assotsiatsiooniuringutega on võimalik leida uusi lookuseid, mis kontrollivad ühe valgu glükosüleerimist. (Lauc *et al.*, 2013)

1.3. Ülegenoomsed assotsiatsiooniuringud

Inimese genoomis paiknevad geneetilised järjestusvariandid DNAs mõjutavad tema tunnuseid ja fenotüüpi ning võivad seeläbi suurendada indiviidi haigestumise riski (Zondervan & Cardon, 2007). Vastavalt järjestusvariantide mõju suurusele sõltub erinevate haiguste ja tunnuste uurimise keerukus. Harvaesinevate mendeliaalsete haiguste puhul on haiguste tuvastamine lihtsam, sest kindlale haigusele vastab kindel genotüüp – mutatsioon(id) paiknevad ühes geenis ja efekt fenotüübile on suur (Ng *et al.*, 2010). Vastandina mõjutavad komplekshaiguste kujunemist väga paljud väikese efektiga DNA järjestusvariandid ning keskkond. See muudab laiemalt esinevate komplekshaiguste ja -tunnuste geneetilise tausta uurimise keerukamaks (Zondervan & Cardon, 2007).

Viimase 10 aasta jooksul on komplekshaiguste ja -tunnuste edukaks analüüsimise meetodiks kasutatud ülegeomseid assotsiatsiooniuuringuid (*Genome-Wide Association Studies*, GWAS) (Manolio *et al.*, 2009). GWAS on juht-kontrolluuring, kus võrreldakse kahe grupi (haigust põdevate ehk juhtude ning tervete ehk kontrollide) indiviidide DNAd. Uuring põhineb hüpoteesil, et geneetilise variandi ehk markeri ühe alleeli suurem esinemissagedus juhtude hulgas viitab haiguse ja järjestusvariandi seosele (Kruglyak, 1999).

1.3.1 Markerid ülegeomsetes assotsiatsiooniuuringutes

Geneetiliste markeritena kasutatakse ülegeomsetes assotsiatsiooniuuringutes ühenukleotiidseid polümorfisme (*single nucleotide polymorphism*, SNP), mis on enimlevinud variatsioonid inimese geoomis. Keskmiselt esineb SNP üks kord iga 300 aluspaari kohta (The 1000 Genomes Project Consortium, 2010).

Lisaks kasutatakse assotsiatsiooniuuringutes ära aheldatuse tasakaalustamatust (*linkage disequilibrium*, LD), mis iseloomustab geoomis lähestikku paiknevaid markereid, mille alleelide kombinatsioonide esinemissagedus populatsioonis on kõrgem (või madalam), kui individuaalse alleelisageduste põhjal eeldada võiks (Weiss *et al.*, 2002). LD struktuuri teadmine üle geoomi võimaldab efektiivsemalt assotsiatsiooniuuringuid teostada, sest LD olemasolu põhjal saab kasutada nn *tag*-SNPsid, millega on võimalik kirjeldada haplotüüpe ehk SNPde grupe ning seeläbi on vaja vähem genotüpiseerida, mis on rahaliselt soodsam ning tehniliselt vähem mahukas (Gabriel *et al.*, 2002).

Aastate jooksul on SNPde ja teiste struktuursete variantide tuvastamiseks läbi viidud erinevaid projekte. 2008. aastal HapMap-projekti teise faasi lõpuks, kus olid sekveneeritud 270 inimese geoomid neljast populatsioonist, oli tuvastatud üle 3 miljoni SNP (Frazer *et al.*, 2007). 2015. aasta lõpuks, mil avaldati 1000 Geoomi Projekti raames 26 populatsioonist sekveneeritud 2504 inimese geoomi analüüsi tulemused, oli teadaolevate SNPde arvuks ligi 85 miljonit (Auton *et al.*, 2015; Sudmant *et al.*, 2015).

1.3.2. Imputeerimine

Praegu kasutusel olevad kommertsiaalsed genotüpiseerimiskiibid, mis sisaldavad mitusada tuhat kuni mitu miljonit SNPd. Lisaks on saadaval vastavalt projekti tüübile spetsiaalseid kiipe, mis sisaldavad geneetilisi variante kindlatest regioonidest, näiteks Illumina

ImmunoArray, mis sisaldavad immuunspetsiifilisi markereid (www.illumina.com) Sellele vaatamata ei pruugi need sisaldada uuritavat SNPd (Spencer *et al.*, 2009). Selle probleemi ühe võimaliku lahendusena on kasutusel geneetiliste markerite imputatsioon, kus üritatakse ennustada neid genotüüpe, mida pole uuritavas valimis otse genotüpiseeritud (Delaneau *et al.*, 2013; Marchini & Howie, 2010). Imputeerimisel on kasutusel olnud eelmainitud HapMap ja 1000 Genoomi projektide raames koostatud referentspaneelid.

Imputeerimise tulemusena suureneb SNPde arv, mida GWASide käigus on võimalik testida. Sellest tulenevalt tõuseb uuringu võimsus ning seega paraneb võimalus leida fenotüübilist efekti põhjustavaid geneetilisi järjestusvariante. Lisaks on võimalik tänu imputatsioonile võrrelda erinevaid uuringuid. (Spencer *et al.*, 2009)

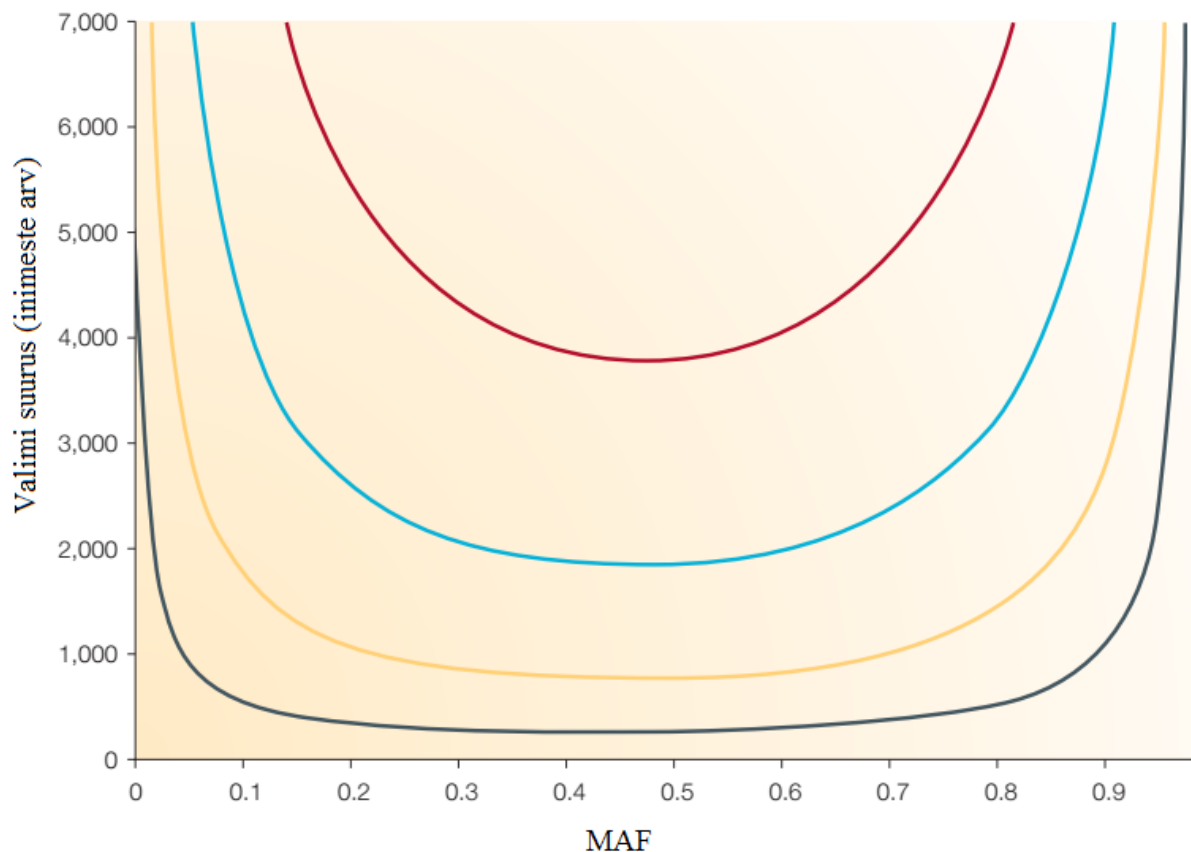
1.3.3. Valimi suurus ja võimsus

Assotsiatsioonianalüüside juures on väga tähtis valim, sest komplekstunnustega seotud geneetilised variandid on enamasti väikese efektiga ning assotsiatsiooni tuvastamiseks SNP ja uuritava fenotüübi vahel on tarvis piisav statistiline võimsus (80%) (Cardon & Bell, 2001).

GWASides leitud assotsiatsioonide tugevuse hindamiseks kasutatakse parameetrit šansside suhe (odds ratio, OR). Kui OR on 1, siis näitab see, et geneetilisel variandil puudub efekt riskile. GWASides tuvastatud sagedaste variantide OR on tavaliselt vahemikus 1,1 kuni 1,4 (Hindorff *et al.*, 2009).

Samuti mõjutab GWASi võimsust efekti omava variandi vähemesineva alleeli sagedus (*minor allele frequency*, MAF). Joonisel 2 on näidatud vajaminev juhtude arv vastavalt erinevate MAFi ja ORi suhete puhul. Sagedaste variantide (MAF > 0.05) puhul peaks 80%-lise statistilise võimsusega valimis juhtude arv jääma vahemikku 2000-7000. Madala sagedusega (MAF < 0.05) ja harvade (MAF < 0.01) variantide leidmiseks peaks juhtude arv olema juba kuni mitukümmend tuhat. (Wang *et al.*, 2005)

Lisaks MAFile mõjutavad valimi võimsust mitmete statistiliste testide tegemine, mis võib kaasa tuua valepositiivsete tulemuste avaldumise (Altshuler *et al.*, 2008). Selle vältimiseks on GWASide puhul määratud p-väärtuse nivoo 5×10^{-8} (Risch & Merikangas, 1996).



Joonis 2. Kujutatud on vajalik juhtude arv assotsiatsioonide (p -väärtusega vähemalt 1×10^{-6}) leidmiseks erinevate šansi suhete juures, et uuringu võimsus oleks vähemalt 80%. Joontega on tähistatud erinevad OR väärtused: 1,2 (punane); 1,3 (sinine); 1,5 (kollane) ja 2,0 (must). (Wang et al., 2005 põhjal)

1.3.4. Leiud ülegenoomsetest assotsiatsiooniuuringutest

Alates 2007. aastast, kui viidi läbi üks esimesi suuri GWASe seitsme levinud komplekshaiguse uurimiseks (Wellcome Trust Case Control Consortium, 2007), on GWASidega uuritud mitmeid komplekshaiguseid ja -tunnuseid (tüüp 1 ja 2 diabeet, reumatoidartriit, eesnäärme- ja rinnavähk, Crohn'i haigus; inimese pikkus, lipiidide ja valkude glükosüleerituse tase vereplasmas) (Lauca *et al.*, 2013; Welter *et al.*, 2014; Wood *et al.*, 2014). Mainitud uuringute käigus on leitud ka hulgalist geneetilisi lookuseid, millel arvatavaks olevat seos haiguste ja tunnuste väljakujunemiseks.

Läbi on viidud ka väga suurte valimitega uuringuid, kui leitud seoste puuduseks on see, et need kirjeldavad enamasti ära väikese osa tunnuse varieeruvusest või haiguse riskist. Näiteks 2014. aastal Wood jt poolt läbiviidud metaanalüüsis ($N \sim 250000$), kus uuriti inimese pikkuse varieeruvuse geneetilist tausta, leiti ligi 700 geneetilist varianti, mis kokku seletasid ära kuni

20% tunnuse varieeruvusest (Wood *et al.*, 2014).

1.3.5. Uute populatsioonipõhiste referentspaneelide roll GWASides

Seda osa, mida GWASides leitud sagedaste geneetiliste variantidega ei õnnestu ära seletada, on hakatud nimetama nn peidetud pärilikkuseks (*hidden heritability*) (Gibson, 2010; Manolio *et al.*, 2009). Seda kujundavad harvad variandid, mille esinemine populatsioonis on tunduvalt harvem ($MAF < 0.01$), kuid mille efekti suurus on suurem kui sagedastel variantidel (Manolio *et al.*, 2009). Harvade variantide uurimine on nende väga madala esinemissageduse tõttu aga väga keerukas, kuid on teada, et nad mängivad haiguste kujunemisel suurt rolli (Gibson, 2011). Üheks võimalikuks lahenduseks madala sagedusega ja harvade variantide analüüsimiseks on pidevalt kasvavad täisgenoomide sekveneerimise andmestikud (Lee *et al.*, 2014).

Viimase viie aasta jooksul on teostatud mitmeid populatsioonipõhiseid sekveneerimisprojekte. 2014. aastal Genome of the Netherlands Consortium'i poolt avaldatud uuringus, kus analüüsiti 250 vanemad-laps perekonna sekveneeritud andmeid (Genome of the Netherlands, GoNL), koostati haplotüübi kaart, mis koosnes 20,4 miljonist ühenukleotiidses variantist ja 1,2 miljonist insertioonist ning deletsioonist (The Genome of the Netherlands Consortium, 2014). Uuringust selgus, et ainult 39% harvadest ühenukleotiidses variantidest, mida GoNL andmestikus vaadeldi, olid olemas ka 1000 Genoomi Projekti Euroopa populatsiooni paneelis. See näitas ilmekalt, kui suure väärtusega on eri populatsioonide põhjalikum uurimine. Samuti selgus antud projektist, et populatsioonipõhine referentspaneel parandab tunduvalt madala sagedusega ja harvade variantide imputatsiooni kvaliteeti (Deelen *et al.*, 2014).

Populatsioonipõhise referentsi olemasolu tähtsust tõestas ka Sardiinia populatsiooni sekveneerimisprojekti tulemused (Sidore *et al.*, 2015). Projekti käigus sekveneeriti üle 2100 indiviidi genoomid ning leiti 19,6 miljonit SNPd. Analüüside käigus leiti kaks lipiidide tasemega ja kaks põletikuliste markerite lookust, mis oleksid 1000 Genoomi Projekti andmestikku kasutades jäänud nägemata.

Sarnased järeldused on ilmnunud Islandi täisgenoomi sekveneerimisprojektist, kus uuriti 2600 täisgenoomi andmeid ning leiti 19,6 miljonit SNPd (Gudbjartsson *et al.*, 2015).

Antud sekveneerimisprojektide tulemuste põhjal saab järeldada, et uued

populatsioonispetsiifilised referentspaneelid parandavad võimalusi harvade variantide seoste leidmiseks ning kinnitamiseks.

2. EKSPERIMENTAALOSA

2.1. Töö eesmärgid

Käesoleva magistritöö eesmärgiks oli teostada Eesti populatsioonis replikatsioonuring varem avaldatud immuunoglobuliin G glükosüleerimist mõjutavate geneetiliste järjestuvariantide mõju kohta, kasutades HapMap projekti teise faasi referentspaneeli.

Lisaks oli eesmärgiks viia läbi täiendavad ülegenoomsed assotsiatsioonianalüüsid, hindamaks, kuidas muutuvad tulemused kasutades referentspaneelina uuemaid, 1000 Genoomi Projekti ja Eesti täisgenoomide sekveneermise projekti andmestikke.

Viimaks teostati assotsiatsioonanalüüsid harvade variantide leidmiseks Eesti täisgenoomide andmestikust

2.2. Materjal ja metoodika

2.2.1. Valim

Antud töö raames kasutati TÜ EGV üle Eesti kogutud geenidonorite proove. Valimisse kaasati geenidonorid, kellel oli teostatud vereplasmast IgG N-glükosüleerituse taseme mõõtmise analüüsid, kasutades ultraefektiivset vedelikkromatograafiat (*ultra performance chromatography*, UPLC). UPLC analüüs viidi läbi Gordon Lauc'i laboris Zagrebis, Horvaatias.

Uuringutes kasutati 77 kvantitatiivset fenotüüpi, millest 23 olid otse mõõdetud ning 54 tuletatud eelnevalt mõõdetud tulemustest. Uuringus kasutatud fenotüüpide täielik nimekiri ning kirjeldus on ära toodud lisa 1.

Genotüpiseerimiskiibiga saadud andmete analüüsis kasutati 575 geenidoonori ning täissekveneerimisandmete analüüsis 119 geenidoonori andmeid. Uuringutesse kaasatud indiviidide sooline ja ealine jaotus on toodud ära vastavalt tabelites 1 ja 2.

Tabel 1. Genotüpiseerimiskiibi andmete assotsiatsioonianalüüsidest kasutatud indiviidide sooline ja ealine jaotus.

	Mehed	Naised	Kokku
Arv	285	290	575
Keskmine vanus (standardhälve)	65,86 (10,94)	70,59 (10,77)	
Vanusevahemik	41-88	31-88	

Tabel 2. Täissekveneerimisandmete assotsiatsioonianalüüsidest kasutatud indiviidide sooline ja ealine jaotus.

	Mehed	Naised	Kokku
Arv	72	47	119
Keskmine vanus (standardhälve)	63,90 (8,19)	62,14 (11,10)	
Vanusevahemik	42-78	31-78	

Käesoleva magistritöö raames teostatavate uuringute läbiviimiseks on geenidoonorid andnud informeeritud nõusoleku ning olemas on Tartu Ülikooli inimuuringute eetika komitee kooskõlastus.

2.2.2. Genotüpiseerimine ja kvaliteedikontroll

Kõikide indiviidide ülegenoomne genotüpiseerimine teostati Illumina Infinium II tehnoloogiat (Illumina Inc., San Diego, CA, USA) kasutades. Geenidoonorite genotüpiseerimiseks kasutati 770 000 markeriga HumanOmniExpress kiipi. Genotüpiseerimine viidi läbi TÜ EGV tuumiklaboris.

Genotüpiseeritud andmed läbisid kvaliteedikontrolli järgnevate parameetrite osas:

- 1) edukalt genotüpiseeritud markerite protsentuaalne osakaal (*Call rate*, CR) nii indiviidi kui iga üksiku SNP kohta;
- 2) minoorse alleeli sagedus (*Minor Allele Frequency*, MAF);
- 3) Hardy-Weidenbergi tasakaalustatus (*Hardy-Weidenberg equilibrium*, HWE);
- 4) indiviidide vaheline sugulus;
- 5) genotüübi põhjal leitud soo mitte-kokkulangevus proovi fenotüübi informatsiooniga;

6) heterosügootsus.

Analüüsidest jäid välja individid, kelle CR oli alla 99% ning kelle fenotüübis märgitud sugu ei vastanud genotüübis ilmnenud soole. Lähisugulusastme tuvastamiseks hinnati ühiselt eellaselt päritud alleelide (*identical by descent*, IBD) osakaalu genoomis. Sugulaspaaridest, kus IBD osakaal oli suurem või võrdne 20% ($PIHAT \geq 0,2$) genoomist, eemaldati üks sugulastest. Eemaldati ka ebanormaalselt kõrge või madala heterosügootsusega individid (Wright's inbreeding coefficient F – kolm standardhälvet keskmisest). Lisaks eemaldati SNPd, mis ei vastanud järgmistele kriteeriumitele:

- 1) $CR < 95\%$;
- 2) HWE testi p -väärtus $< 10^{-6}$;
- 3) $MAF < 1\%$;
- 4) mitte-autosomaalsed markerid ning markerid, mille alleelideks olid A/T või C/G.

Kvaliteedikontrolli genotüüpiseerimiskiibi ülegenoomseteks assotsiatsioonianalüüsiks läbis 575 geenidonorit ja 606871 SNPd. Andmete kvaliteedikontrolli ja filtreerimise läbiviimiseks kasutati programme Illumina GenomeStudio 3.1 (www.illumina.com), PLINK v.1.07 (<http://pngu.mgh.harvard.edu/purcell/plink/>; (Purcell *et al.*, 2007)) ja R-3.0.2 (Free Software Foundation, Boston, MA, USA; www.fsf.org).

TÜ EGV geenidonorite genoomse DNA täielik järjestamine ehk sekveneerimine viidi teostati Broad Instituudis (Boston, MA, USA). Sekveneerimine toimus kasutades teise põlvkonna sekveneerimisel põhinevat Illumina HiSeq X platvormi tehnoloogiat. Iga positsiooni katvuseks oli keskmiselt 30 korda.

Sekveneeritud andmete kvaliteedikontroll viidi läbi kasutades Broad Instituudi poolt välja töötatud standardset protokollit. Täiendav proovidepõhine bioinformaatiline kvaliteedikontroll viidi läbi TÜ EGV genoomika töögrupi poolt. Sekveneerimisandmete kvaliteedikontrolli assotsiatsioonianalüüsiks läbis 119 geenidonorit ja 8697038 ühenukleotiidset varianti.

2.2.3. Imputeerimine

Illumina HumanOmniExpress kiibi puuduvate markerite imputeerimiseks kasutati referentsgenoomina andmeid kolmest erinevast paneelist:

- 1) HapMap2 (*phase 2*; november 2008)

- 2) 1000 Genoomi Projekt (*phase 3*; oktoober 2014);
- 3) Eesti referentspaneel (aprill 2016; avaldamata andmed).

Imputeerimiseks kasutati programme SHAPEIT v2 (Delaneau *et al.*, 2013) ja IMPUTE v2.2.2 (Howie *et al.*, 2009). Imputeerimise järgselt eemaldati SNPd imputeerimise kvaliteediskoori alusel (<0.4) Imputeerimise järgselt jäi andmestikku *ca* 600000 ja lisaks imputeeritud markereid 1909453 (HapMap2), 10431697 (1000 Genoomi Projekt), 6973762 (Eesti referentspaneel). Imputatsiooni HapMap2 ja 1000 Genoomi projekti andmetele teostas TÜ EGV teadur Evelin Mihailov ning Eesti referentspaneeli andmetele TÜ EGV teadurid Reedik Mägi, Maris Alver ja Mart Kals.

2.2.5 Assotsiatsioonianalüüsid

Kiibiga genotüpiseeritud ja imputeeritud geneetiliste markerite ning fenotüübi vahelised assotsiatsioonianalüüsid viidi läbi programmiga SNPTEST v2.5 (Marchini & Howie, 2010). Iga referentspaneeli andmeid kasutades analüüsiti geneetiliste markerite seost glükaani tasemete kui pidevate tunnuste fenotüüpidega, rakendades lineaarse regressiooni alleeliefektide aditiivset mudelit. Kõigis analüüsides kasutati kovariaatidena sugu, vanust ja suguluste maatriksi põhjal leitud viite esimest peakomponenti, et vähendada võimalikku populatsiooni stratifikatsiooni.

Sekveneermisandmete analüüs teostati programmiga EFACTS (<http://genome.sph.umich.edu/wiki/EFACTS>). Andmete analüüs viidi läbi EFACTS programmi kuuluva testiga EMMAX (Efficient Mixed Model Association eXpedited), mis võtab arvesse populatsiooni struktuuri ja peidetud sugulust. Antud test sai valitud, kuna see põhineb lineaarsel mudelil ja sobib kvantitatiivsete tunnuste, nagu seda on glükaanide glükosüleerituse tase, ning ühenukleotiidsete variantide seoste analüüsimiseks (Kang *et al.*, 2010).

Nii kiibi kui sekveneermisandmete analüüsides kasutati fenotüübi tunnustena tunnuse pöördnormaaljaotuse transformeeritud jääkide väärtusi, tagamaks, et tunnuste jaotus oleks võimalikult sarnane normaaljaotusele. Glükaanide väärtuste kohandamiseks kasutati kovariaatidena vanust ja sugu.

2.3. Tulemused

2.3.1. GWAS tulemused kasutades HapMap2 referentspaneeli

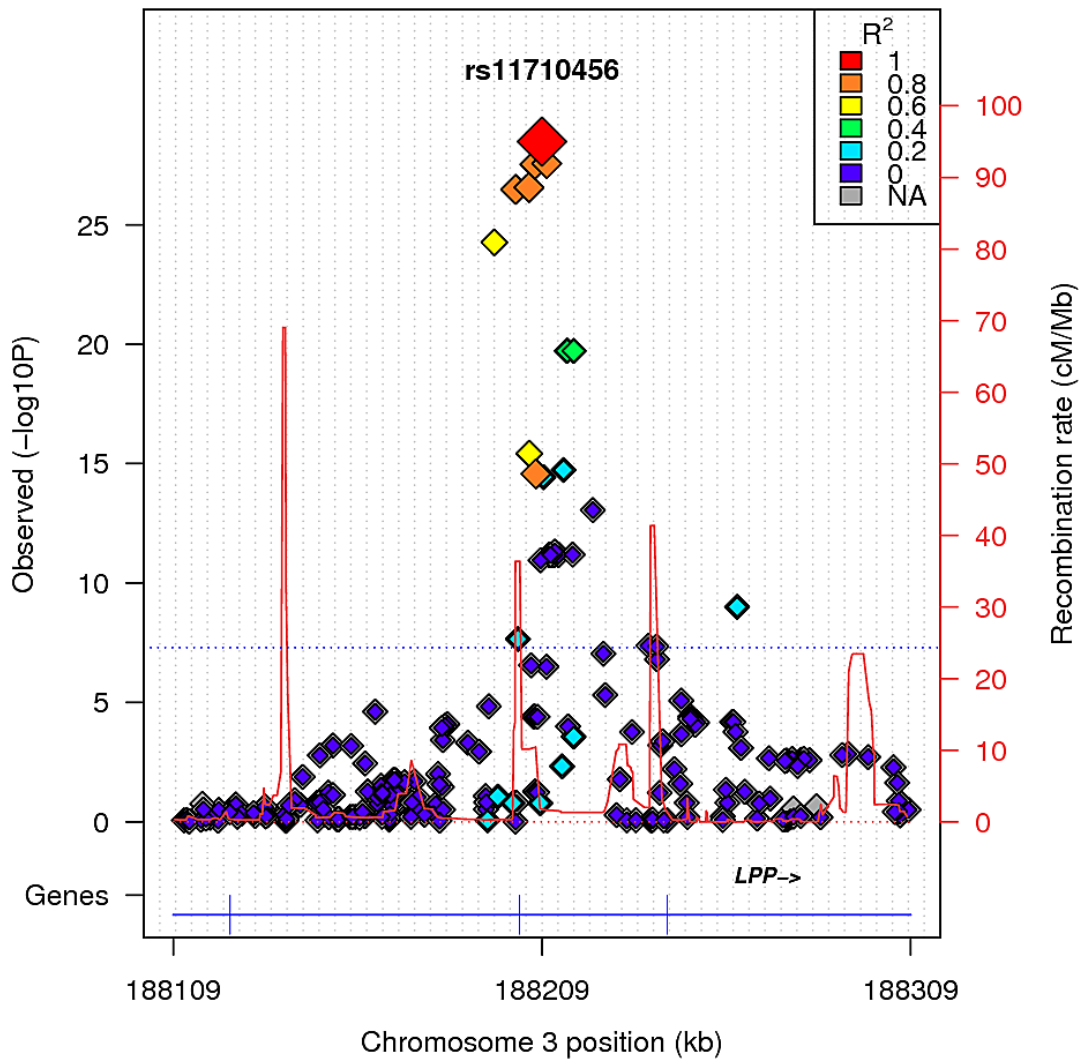
Assotsiatsioonianalüüsidesse, kus kasutati HapMap2 referentspaneeli, kaasati 575 indiviidi (285 meest ja 290 naist). Analüüside käigus leiti 125 markerit kolmest genoomsest lookusest, mille p-väärtus oli üle genoomselt statistilise olulisusega. Iga lookuse parima p-väärtusega SNP on ära toodud tabelis 3.

Tabel 3. GWASide tulemused, kus kasutati HapMap2 referentspaneeli. Lähim geen on määratud, kui selle kaugust seost näidanud SNPst on väiksem kui 100 kb. Tabelis on toodud iga lookuse kõige tugevama assotsiatsiooniga markerid kolmest kromosoomist. Chr – kromosoom; EAF – efektalleeli sagedus. Tunnus tähistab fenotüüpi, millega marker oli kõige enam assotsieerunud.

Chr	SNP	EAF	Efekti suurus (std.viga)	p-väärtus	Lähim geen	Tunnus
1	rs16830188	0,0162	-1,3887 (0,2273)	1,83x10 ⁻⁹	<i>CLIC4</i>	IGP76
3	rs11710456	0,7177	0,7145 (0,0602)	3,29x10 ⁻⁹	<i>ST6GALI</i>	IGP29
7	rs1554716	0,3809	-0,3350 (0,0586)	1,75x10 ⁻⁹	<i>SLC13AI</i>	IGP9

Leitud variandid asusid kromosoomides 1, 3 ja 7 (tabel 3). Variandid kromosoomides 1 ja 7 asuvad järgmiste geenide läheduses: *CLIC4* (lookus 1p36.11) ja *SLC13AI* (7q31.32). Kromosoomis 3 leitud variant rs11710456 (joonis 3) paikneb geeni *ST6GALI* (3q27.3) intronis.

rs11710456



Joonis 3. Markeri rs11710456 regionaaljoonis. HapMap2 referentspaneelil põhinevate andmete assotsiatsioonanalüüsil tunnusega IGP29 leitud kõige tugevama assotsiatsiooniga marker rs11710456 on märgitud punase rombina, teised markerid muud värvi rombidenä. Markeri värvusega on esindatud nende LD variandiga rs11710456. X-teljel on markerite füüsilised asukohad kolmandas kromosoomis ja Y-teljel on markerite p-väärtuste negatiivsed kümnenlogaritmid. Punased jooned märgivad rekombinatsioonimäära.

2.3.1. GWASi tulemused kasutades 1000 Genoomi referentspaneeli

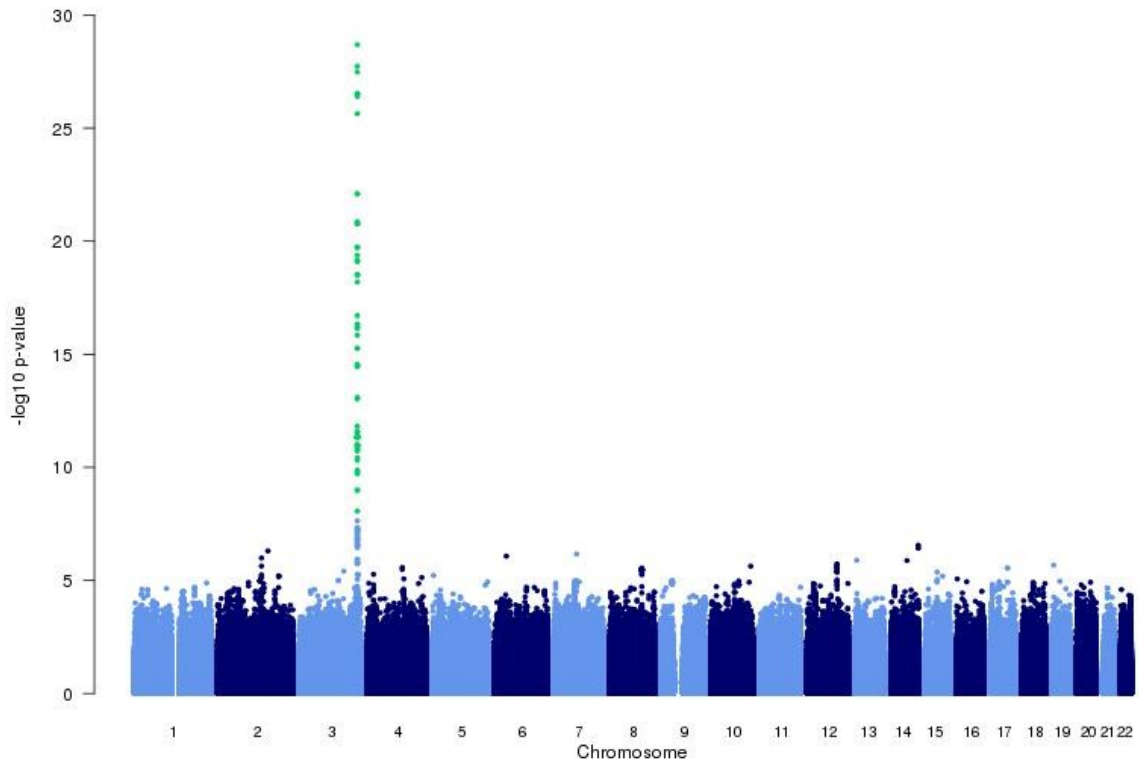
Assotsiatsioonianalüüsidesse, kus kasutati 1000 Genoomi projekti faas 3 referentspaneeli, kaasati 575 indiviidi (285 meest ja 290 naist). Analüüside käigus leiti 327 markerit seitsmest genoomsest lookusest, mille p-väärtus oli ülevenoomselt statistilise olulisusega. Iga lookuse parima p-väärtusega SNP on ära toodud tabelis 4.

Tabel 4. GWASide tulemused, kus kasutati 1000 Genoomi referentspaneeli. Lähim geen on määratud, kui selle kaugust seost näidanud SNPst on väiksem kui 100 kb. Tabelis on toodud iga lookuse kõige tugevama assotsiatsiooniga markerid seitsmest kromosoomist. Chr – kromosoom; EAF – efektalleeli sagedus. Tunnus tähistab fenotüüpi, millega marker oli kõige enam assotsieerunud.

Chr	SNP	EAF	Efekti suurus (std.viga)	p-väärtus	Lähim geen	Tunnus
1	rs188468174	0,0338	-2,6224 (0,411)	$3,54 \times 10^{-10}$	<i>RUNX3</i>	IGP76
3	rs7621161	0,3231	-0,7062 (0,059)	$2,04 \times 10^{-29}$	<i>ST6GAL1</i>	IGP29
7	rs1554716	0,6191	0,3345 (0,059)	$1,86 \times 10^{-8}$	<i>IQUB</i>	IGP9
9	rs10982623	0,0490	-0,8104 (0,138)	$6,82 \times 10^{-9}$	<i>DECI</i>	IGP62
14	rs113756886	0,4400	0,4499 (0,073)	$1,55 \times 10^{-9}$	<i>ELK2AP</i>	IGP8
15	rs71125923	0,1427	-0,4922 (0,088)	$3,91 \times 10^{-8}$	-	IGP68
21	rs11313672	0,4269	0,3699 (0,063)	$5,62 \times 10^{-9}$	<i>D21S2088E</i>	IGP77

Leitud variandid paiknesid kromosoomides 1, 3, 7, 9, 14, 15 ja 21. Variandid kromosoomides 1, 3 ja 9 paiknesid lookustes *RUNX3* (1p36), *ST6GAL1* (3q27.3) ja *DECI* (9q32), geenide intronites. Variandid kromosoomides 7, 14, 15 ja 21 paiknesid vastavalt lookustes *IQUB* (7q31.32), 14q32.3, 15q22.2 ja 21q21.2.

Kromosoomis 3 paiknev variant rs7621161 oli väga tugevalt assotsieerunud tunnusega IGP29 (p-väärtus $2,04 \times 10^{-29}$) (joonis 4).



Joonis 4. IGP29 assotsiatsiooniuringu Manhattan joonis. X-teljel on tunnusega seostunud markerite genoomsed positsioonid ning Y-teljel nende p-väärtuse negatiivne logaritm. Ülegenoomselt statistilise olulisusega markerid asuvad kromosoomis 3.

2.3.3. GWASi tulemused kasutades Eesti referentspaneeli

Assotsiatsioonianalüüsidesse, kus kasutati Eesti populatsiooni referentspaneeli, kaasati 575 indiviidi (285 meest ja 290 naist). Analüüside käigus leiti 105 markerit neljast genoomsest lookusest, mille p-väärtus oli üle genoomselt statistilise olulisusega. Iga lookuse parima p-väärtusega SNP on ära toodud tabelis 5.

Tabel 5. GWASide tulemused, kus kasutati Eesti referentspaneeli. Lähim geen on määratud, kui selle kaugust seost näidanud SNPst on väiksem kui 100 kb. Tabelis on toodud iga lookuse kõige tugevama assotsiatsiooniga markerid neljast kromosoomist. Chr – kromosoom; EAF – efektalleeli sagedus. Tunnus tähistab fenotüüpi, millega marker oli kõige enam assotsieerunud.

Chr	SNP	EAF	Efekti suurus (std.viga)	p-väärtus	Lähim geen	Tunnus
3	rs6764279	0,2696	-0,6970 (0,0599)	3,31x10 ⁻²⁸	<i>ST6GALI</i>	IGP29
7	rs1554716	0,6191	0,3345 (0,0586)	1,86x10 ⁻⁸	<i>IQUB</i>	IGP9
9	rs79504765	0,4879	-0,4162 (0,0732)	2,96x10 ⁻⁸	<i>OLFMI</i>	IGP42
11	rs182241801	0,1987	-0,431128	1,52x10 ⁻⁸	-	IGP34

Leitud variandid paiknesid kromosoomides 3, 7, 9 ja 11. Variandid paiknesid järgmistes lookustes: *ST6GALI* (3q27.3), *IQUB* (7q31.32) ja *OLFMI* (9q34.3). Lookuses 3q27.3 asunud variant oli vastava geeni intronis.

2.3.4. Sekvenerimisandmete assotsiatsioonanalüüside tulemused

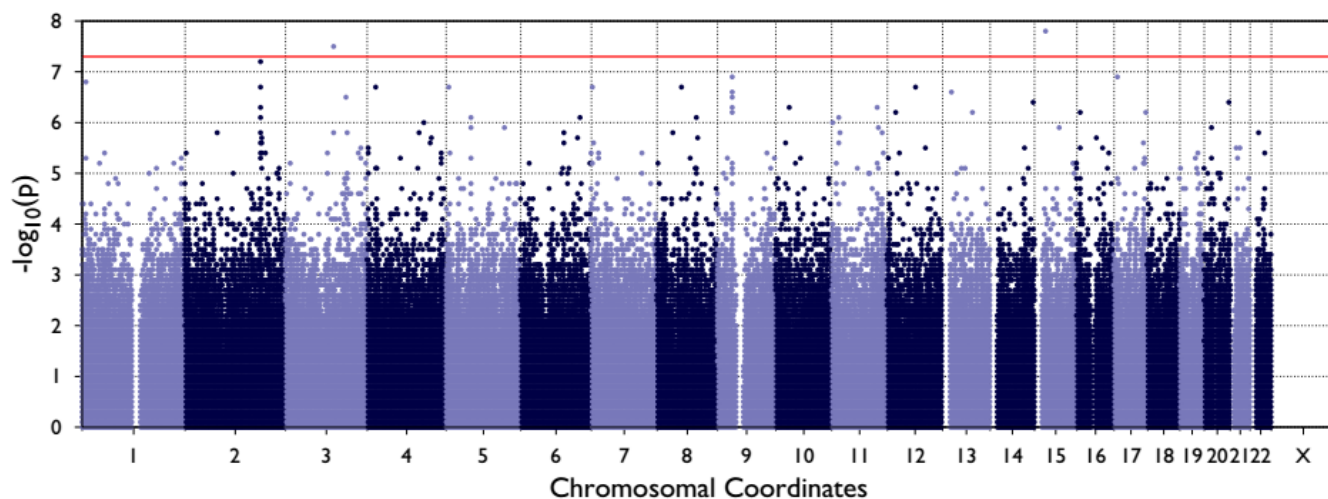
Sekvenerimisandmete assotsiatsioonianalüüsidesse kaasati 119 indiviidi (52 meest ja 47 naist). Analüüside käigus leiti 26 markerit üheksast genoomsest lookusest, mille p-väärtus oli ülegenoomselt statistilise olulisusega. Iga lookuse parima p-väärtusega SNP on ära toodud tabelis 6.

Tabel 6. GWASide tulemused, kus kasutati Eesti referentspaneeli. Geen on määratud, kui vastav variant asub antud geenis. Tabelis on toodud iga lookuse kõige tugevama assotsiatsiooniga markerid kaheksast kromosoomist. Chr – kromosoom; EAF – efektalleeli sagedus. Tunnus tähistab fenotüüpi, millega marker oli kõige enam assotsieerunud.

Chr	Marker	EAF	Efekti suurus (std.viga)	p-väärtus	Geen	Tunnus
1	1:55861614_T/C	0,0252	2,262 (0,3823)	3,77x10 ⁻⁸	-	IGP46
3	3:117429607_C/A	0,0168	2,572 (0,433)	3,41x10 ⁻⁸	<i>LSAMP</i>	IGP2
3	3:186722362_T/C	0,2437	-0,7366 (0,1256)	4,78x10 ⁻⁸	<i>ST6GAL1</i>	IGP29
6	6:143584682_T/A	0,0126	3,082 (0,5235)	4,33x10 ⁻⁸	<i>AIG1</i>	IGP6
7	7:2650389_C/T	0,0168	-2,784 (0,4628)	2,39x10 ⁻⁸	<i>IQCE</i>	IGP58
8	8:85748669_C/T	0,0378	1,585 (0,2489)	4,56x10 ⁻⁸	<i>RALYL</i>	IGP27
11	11:36125631_C/T	0,0126	3,122 (0,5174)	2,20 x10 ⁻⁸	<i>LDLRAD3</i>	IGP2
15	15:29613867_G/A	0,0378	1,780 (0,2906)	1,44x10 ⁻⁸	<i>FAM13A1</i>	IGP42
17	17:8904448_A/G	0,0294	-2,094 (0,3292)	4,73x10 ⁻⁹	-	IGP59

Leitud variandid paiknesid kromosoomides 1, 3, 6, 7, 8, 11, 15 ja 17. Variandid kromosoomides 1 ja 17 olid intergeensed. Variandid kromosoomides 3, 6, 7, 8, 11 ja 15 asusid järgnevas geenilookustes: *LSAMP* (3q13.31), *ST6GAL1* (3q27.3), *AIG1* (6q24.2), *IQCE* (7p22.3), *RALYL* (8q21.2), *LDLRAD3* (11p13), *FAM13A1* (4q22.1).

Lookuses 3q13.31 paikneva üle genoomselt statistilise olulisusega markeri 3:117429607_C/A (joonis 5) ja uuritavate tunnuste vahel seost kiibipõhistes assotsiatsioonanalüüsides ei leitud, mis teeb sellest huvitava leiu sekveneermisandmetest.



Joonis 5. IGP2 assotsiatsiooniuringu Manhattan joonis. X-teljel on tunnusega seostunud markerite genoomsed positsioonid ning Y-teljel nende p-väärtuse negatiivne logaritm. Huvipakkuv variant asub kromosoomis 3.

2.4. Arutelu

Valkude modifitseermise protsessid on olulised organismi normaalseks talitluseks. Üks olulisemaid valkude modifikatsioone on nende post-translatsiooniline glükosüleerimine, mille käigus lisatakse polüpeptiidahelale oligosahhariide ehk glükaane. Lisatud oligosahhariidid muudavad valgu struktuuri ja funktsiooni (Opdenakker *et al.*, 1993; Skropeta, 2009). Glükosüleerimise uurimine on näidanud selle olulist mõju organismile, aga selle geneetilise tausta kohta on seni limiteeritud teadmised.

Eelmainitust tulenevalt oli käesoleva magistritöö eesmärgiks uurida immuunoglobuliin G glükosüleerituse taset mõjutavaid geneetilisi variante. Immuunoglobuliin G (IgG) on üks enamuuritud glükoproteiine, mis on olemuselt antikeha, ning on seetõttu oluline immuunsüsteemi talitluses (Janeway *et al.*, 2001). IgG on kõige levinum immuunoglobuliin inimese veres (Vidarsson *et al.*, 2014) ning edasiarendused tehnoloogiates glükaanide taseme määramiseks verest on teinud sellest suurepärase uurimisobjekti.

Käesoleva töö viidi läbi assotsiatsiooniuringud 77 glükaanide glükosüleerituse taset iseloomustava tunnusega (lisa 1) 575 geenidoonori andmetega kasutades HapMap projekti teise faasi, 1000 Genoomi Projekti faas 3 ja äsjaloodud Eesti populatsioonis põhinevat referentspaneeli.

HapMap projekti teise faasi referentspaneel sai valitud lähtuvalt faktist, et üks esimesi assotsiatsiooniuringuid glükaanide geneetilise tausta uurimiseks Lauc jt poolt 2013. aastal kasutas samuti referentspaneelina just HapMap2 andmestikku (Lauc *et al.*, 2013). Antud töö üheks eesmärgiks oli vaadata, kas leiud antud tööst replitseeruvad ka Eesti populatsioonis. Ülejäänud referentspaneelid sai valitud eesmärgiga kontrollida, kas uute referentspaneelidega õnnestub leida uusi variante ning varasemaid leida kinnitada.

Lisaks teostati assotsiatsiooniuring 119 täisgenoomi sekveneermisandmete uurimiseks, leidmaks uusi harvasid variante ($MAF < 0.01$).

HapMap2 andmestiku analüüsides leiti kõigi analüüsitud tunnuste peale 125 markerit kolmest lookusest, mis olid ülegenoomselt statistilise olulisusega ($P < 5 \times 10^{-8}$). Esimeses kromosoomis leitud marker rs16830188 oli kõige tugevalt assotsieerunud tunnusega IGP76, mis iseloomustab fukosüleeritud digalaktüseeritud mittehargnevate N-atsetüülglükosamiini (GlnNAc) ja digalaktüseeritud kaheks hargnenud GlnNAc struktuuride suhet. Antud variant paikneb geeni *CLIC4* (*chloride intracellular channel 4*) lähedal, mis kodeerib samanimelist valku, mis osaleb mitmete kudede vahelises signalisatsioonis ning diferentseerumisprotsessides, lisaks on teda kirjeldatud kui tuumorsupressorit lamerakulise kartsinoomi vastu (Suh *et al.*, 2012). Seni puudub variandi seos glükosüleerimisega. Kromosoomis 3 leitud variant rs1171046 asub *ST6GAL1* geeni intronis. *ST6GAL1* kodeerib sialüültransferaasi 6, mis lisab sialhapet mitmetele glükoproteiinidele, kaasaarvatud IgG-le. Erinevates uuringutes on näidatud *ST6GAL1* ekspressiooni tõusu erinevat tüüpi kasvajatel (Antony *et al.*, 2014; Swindall & Bellis, 2011).

Mõlema geeni põhjal võib järeldada, et IgG glükosüleerimisel võib olla mõju vähitekke regulatsioonis.

1000 Genoomi projekti referentspaneeli kasutades leiti kõigi analüüsitud tunnuste peale 327 markerit seitsemest lookusest, mis olid ülegenoomselt statistilise olulisusega ($P < 5 \times 10^{-8}$). Esimeses kromosoomis leitud marker rs188468174 asub geeni *RUNX3* intronis. Geeni poolt

kodeeritud valgul on näidatud tuumorsupressori roll erinevates kasvajates, näiteks mitteväikerakk-kopsuvähis ning munasarjavähis (Barghout *et al.*, 2015).

Lisaks leiti üks marker rs7621161 kromosoomist 3, mis paikneb eelmainitud *ST6GAL1* intronis.

Kromosoomis 7 paiknev marker rs1554716 asub *IQUB* (*IQ motif and ubiquitin domain containing*) geenis, mida ekspresseeritakse ajus ning mida varem on seostatud bipolaarse häirega (Kerner *et al.*, 2013).

Kromosoomis 9 paiknev marker rs10982623 asub geenis *DECI* (*deleted in esophageal cancer 1*), mis on transkriptsioonifaktor, millel on apoptoosi reguleeriv roll. Antud valku on seostatud pankrease vähi epiteel-mesenhümaalsel transitsioonil (Wu *et al.*, 2012).

Taaskord näitavad tulemused seost glükosüleerimise ja vähi teket mõjutavate valkude vahel.

GWASides, kus kasutati Eesti populatsioonil põhinevat referentspaneeli, leiti 105 markerit neljast lookusest, mis olid ülegenoomselt statistilise olulisusega ($P < 5 \times 10^{-8}$). Kõige tugevamat assotsiatsiooni IGP9ga näitas taas marker rs1554716 kromosoomist 7. Samuti leiti kolmas väga tugeva assotsiatsiooniga marker rs6764279 lookusest 3q27.3, kus asub *ST6GAL1* geen.

Lisaks variant rs79504765 üheksanda kromosoomi lookusest 9q34.3, kus asub geen *OLFM1* (*olfactomedin 1*). Valgul on kirjeldatud roll epiteel-mesenhümaalsel transitsioonil kanaembrüo südames (Lencinas *et al.*, 2013).

Sekveneerimisandmete assotsiatsioonianalüüside tulemusena leiti 26 markerit üheksast lookusest, mis olid ülegenoomselt statistilise olulisusega ($P < 5 \times 10^{-8}$). Geenisesed variandid leiti lookustest kromosoomidest 3, 6, 7, 8, 11 ja 15. Kromosoomist kolm leiti kaks lookust, millest üks oli varem kirjeldatud 3q27.3, kus asub *ST6GAL1* geen. Uue lookusena leiti 3q13.31, kus asub geen *LSAMP* (*limbic system-associated membrane protein*). Antud geen kodeerib immunoglobuliin LAMP, OBCAM ja neurotrimiin perekonda kuuluvat valku. Valk osaleb limbilise süsteemi arengus, lisaks oletatakse sel olevat roll tuumorsupressorina ning psühhiaatriliste haiguste (näiteks skisofreenia ning bipolaarne häire) juures (Koido *et al.*, 2014; Nurnberger *et al.*, 2014; Petrovics *et al.*, 2015).

Kromosoomist 6 leitud marker asub geenis *AIG1* (*androgen-induced 1*), mille poolt kodeeritud valgul on kirjeldatud raku surma vastuvõtlikkuse suurenemine oksüdatiivse stressi korral (Nickel *et al.*, 2016).

Kromosoomist 11 leitud geeni *LDLRAD3* (*low density lipoprotein receptor class A domain containing 3*) poolt kodeeritud valk on transmembraanne retseptor, mis on ekspresseerunud

neuronites. On kirjeldatud, et LRAD3 vähendab oluliselt amüloid prekursorvalkude eluiga (Ranganathan *et al.*, 2011).

Kromosoomist 15 leitud geeni *FAM13A1* (*family with sequence similarity 13 member A*) on varem seostatud erinevate kopsu funktsiooni näitajatega ning krooniliste kopsuhaigustega (Corvol *et al.*, 2014).

Eelmainitud analüüsides leiti hulgaliselt seoseid geneetiliste variantide ning IgG glükosüleerituse taset iseloomustavate tunnuste vahel. Paljud tugeva assotsiatsiooniga lookused on seotud neuroloogiliste ja immunoloogiliste haigustega ning kasvajatega. Leitud tulemused kattuvad varem kirjanduses näidatud glükosülatsiooni efektiga inimese immuunhaiguste ning kasvajate tekke riskile (Lauc *et al.*, 2010, 2013). Käesolevas töös replitseerus üks varem kirjeldatud ((Lauc *et al.*, 2013)) geneetiline variant *ST6GAL1* geenis – rs11710456. See aspekt toetab GWAS kui katsedisaini sobviust komplekssete tunnuste analüüsiks.

Lisaks saab käesoleva töö tulemuste põhjal järeldada, et sekveneerimisandmed annavad geneetilisteks assotsiatsioonanalüüsideks juurde palju informatsiooni.

KOKKUVÕTE

Valgud on elutähtsad molekulid organismides, mille funktsioon ja struktuuri on võimalik muuta post-translatsiooniliste modifikatsiooniprotsessidega nagu glükosüleerimine. Üheks olulisemaks glükoproteiiniks inimeses on immuunoglobuliin G (IgG), mis osaleb immuunvastuse kujunemisel. Paraku pole seni palju teada IgG glükosüleerimise geneetilise tausta kohta.

Käesoleva magistritöö eesmärgiks oli uurida, mis mõju omavad geneetilised järjestusvariandid IgG glükosüleerimisele ning võrrelda neid seni avaldatud töödega. Samuti oli eesmärgiks leida uusi geneetilisi markereid, mis võivad IgG glükosüleeritust mõjutada. Selleks teostati kolm genotüpiseerimiskiipide andmetel põhinevat ülegenoomset assotsiatsiooniuuringut, kuhu kaasati 575 Tartu Ülikooli Eesti Geenivaramu geenidoonorit. Referentsandmestikuna kasutati HapMap, 1000 Genoomi projekti ning äsja koostatud Eesti populatsioonipõhiseid referentspaneeli. Lisaks viidi läbi assotsiatsiooniuuring 119 geenidoonori sekveneeritud täisgenoomi andmetega.

Töö tulemusena leiti hulgaliselt ülegenoomselt statistiliselt olulise assotsiatsiooniga markereid. Kasutades referentsandmestikke, leiti vastavalt 125 markerit kolmest lookusest (HapMap), 327 markerit seitmest lookusest (1000 Genoomi projekt), 105 markerit neljast lookusest, mille p-väärtus ületas ülegenoomselt statistilise olulisuse nivoo. Lisaks leiti 26 markerit üheksast lookusest sekveneermisandmete analüüsist. Kõigis neljas analüüsis leiti seos IgG glükosüleerimist iseloomustavate tunnuste ning kolmandas kromosoomis paikneva geeni *ST6GAL1* vahel, mida on ka varem kirjeldatud.

Ühtlasi ilmes töö tulemusena, et uuemad referentspaneelid parandavad tunnusega assotsieerunud markerite leidmist ning sekveneermisandmed annavad võimaluse leida uusi ja harvamaid markereid.

SUMMARY

The role of genetic variants on immunoglobulin G glycosylation

Timo Tõnis Sikka

Summary

Proteins have many essential roles in organism. One mechanism to have a wide repertoire of functions in the organism is the post-translational modification of proteins, such as glycosylation. Most membrane and secreted proteins are glycosylated, however, little is known about the genetic background of the glycosylation.

The aim of this study was to analyse the effect of genetic variants on immunoglobulin G (IgG) glycosylation. Three genome-wide association studies with 77 IgG glycosylation traits were conducted using three different reference panels and samples of 575 individuals from Estonian population who had IgG glycosylation traits measured. As a results of forementioned analysis, we found 125, 327, and 105 associated genetic markers from HapMap, 1000 Genome Project and Estonian population reference panels, respectively. Additionally, association studies using whole genome sequence data of 119 individuals were conducted, and 16 markers of genome-wide significance were found.

The markers in *ST6GALI* gene on chromosome 3 were found in all four conducted studies. The gene codes the enzyme sialyltransferase 6 which adds sialic acid to glycoproteins. This locus was also found in a study by Gordon Lauc and others in 2013. This confirms GWASes as a way of analysing complex traits such as glycosylation.

The results from this study also indicate the power of sequence data to find new genetic markers. However, as the sample size for whole-genome data analysis was small, additional studies should be conducted to confirm the association between the markers and glycosylation traits.

TÄNUAVALDUSED

Soovin tänada kõiki inimesi, kes mind antud magistritöö kirjutamisel toetasid nii nõu kui jõuga. Tänan oma juhendajat Tõnu Eskot. Samuti tahaksin tänan Andreast.

Lisaks tänan Tartu Ülikooli Eesti Geenivaramut hea koostöö eest ning töös uuritud geenidoonoreid, kellela poleks olnud seda magistritööd võimalik teostada.

KIRJANDUSE LOETELU

- Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic Mapping in Human Disease. *Science*, (November), 881–888.
- Anthony, R. M., Kobayashi, T., Wermeling, F., & Ravetch, J. V. (2011). Intravenous gammaglobulin suppresses inflammation through a novel T(H)2 pathway. *Nature*, 475(7354), 110–113.
- Anthony, R. M., & Ravetch, J. V. (2010). A novel role for the IgG Fc glycan: The anti-inflammatory activity of sialylated IgG Fcs. *Journal of Clinical Immunology*, 30(SUPPL. 1), 9–14.
- Antony, P., Rose, M., Heidenreich, A., Knuchel, R., Gaisa, N. T., & Dahl, E. (2014). Epigenetic inactivation of ST6GAL1 in human bladder cancer. *BMC Cancer*, 14, 901.
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., ... Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74.
- Barghout, S. H., Zepeda, N., Vincent, K., Azad, A. K., Xu, Z., Yang, C., ... Fu, Y. (2015). RUNX3 contributes to carboplatin resistance in epithelial ovarian cancer cells. *Gynecologic Oncology*, 138(3), 647–655.
- Cardon, L. R., & Bell, J. I. (2001). Association study designs for complex diseases. *Nature Reviews. Genetics*, 2(2), 91–9.
- Corvol, H., Hodges, C. A., Drumm, M. L., & Guillot, L. (2014). Moving beyond genetics: is FAM13A a major biological contributor in lung physiology and chronic lung diseases? *Journal of Medical Genetics*, 51(10), 646–9.
- Deelen, P., Menelaou, A., van Leeuwen, E. M., Kanterakis, A., van Dijk, F., Medina-Gomez, C., ... Swertz, M. A. (2014). Improved imputation quality of low-frequency and rare variants in European samples using the “Genome of The Netherlands”. *European Journal of Human Genetics*, 22(11), 1321–6.
- Delaneau, O., Zagury, J.-F., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods*, 10(1), 5–6.
- Dorner, a J., Bole, D. G., & Kaufman, R. J. (1987). The relationship of N-linked glycosylation and heavy chain binding protein associated with the secretion of glycoproteins. *J. Cell Biol.*, 105(6), 2665–2674.
- Frazer, K. a, Ballinger, D. G., Cox, D. R., Hinds, D. a, Stuve, L. L., Gibbs, R. a, ... Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), 851–61.
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., ... Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science*

- (New York, N.Y.), 296(5576), 2225–9.
- Gibson, G. (2010). Hints of hidden heritability in GWAS. *Nature Genetics*, 42(7), 558–560.
- Gibson, G. (2011). Rare and common variants: twenty arguments. *Nature Reviews. Genetics*, 13(2), 135–45.
- Gornik, O., Pavić, T., & Lauc, G. (2012). Alternative glycosylation modulates function of IgG and other proteins - Implications on evolution and disease. *Biochimica et Biophysica Acta - General Subjects*, 1820(9), 1318–1326.
- Gudbjartsson, D. F., Helgason, H., Gudjonsson, S. a, Zink, F., Oddson, A., Gylfason, A., ... Stefansson, K. (2015). Large-scale whole-genome sequencing of the Icelandic population. *Nature Genetics*, 47(5), 435–444.
- Hindorff, L. a, Sethupathy, P., Junkins, H. a, Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. a. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*, 106(23), 9362–7.
- Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5(6), e1000529.
- Kaneko, Y., Nimmerjahn, F., & Ravetch, J. V. (2006). Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science (New York, N.Y.)*, 313(5787), 670–673.
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., ... Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), 348–354.
- Kerner, B., Rao, A. R., Christensen, B., Dandekar, S., Yourshaw, M., & Nelson, S. F. (2013). Rare Genomic Variants Link Bipolar Disorder with Anxiety Disorders to CREB-Regulated Intracellular Signaling Pathways. *Frontiers in Psychiatry*, 4(November), 154.
- Knezevic, A., Polasek, O., Gornik, O., Rudan, I., Campbell, H., Hayward, C., ... Lauc, G. (2009). Variability, Heritability and Environmental Determinants of Human Plasma N-Glycome. *Journal of Proteome Research*, 8(1), 694–701.
- Koido, K., Janno, S., Traks, T., Parksepp, M., Ljubajev, I., Veiksaar, P., ... Vasar, E. (2014). Associations between polymorphisms of LSAMP gene and schizophrenia. *Psychiatry Research*, 215(3), 797–798.
- Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, 22(2), 139–44.
- Lauc, G., Essafi, A., Huffman, J. E., Hayward, C., Knežević, A., Kattla, J. J., ... Rudan, I. (2010). Genomics meets glycomics-the first gwas study of human N-glycome identifies

- HNF1A as a master regulator of plasma protein fucosylation. *PLoS Genetics*, 6(12), 1–14.
- Lauc, G., Huffman, J. E., Pucic, M., Zgaga, L., Adamczyk, B., Mućinić, A., ... Rudan, I. (2013). Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. *PLoS Genetics*, 9(1).
- Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. (2014). Rare-variant association analysis: Study designs and statistical tests. *American Journal of Human Genetics*, 95(1), 5–23.
- Manolio, T. a, Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. a, Hunter, D. J., ... Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–53.
- Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Rev Genetics*, 11(7), 499–511.
- Marek, K. W., Vijay, I. K., & Marth, J. D. (1999). A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality. *Glycobiology*, 9(11), 1263–1271.
- Nairn, A. V., York, W. S., Harris, K., Hall, E. M., Pierce, J. M., & Moremen, K. W. (2008). *Regulation of glycan structures in animal tissues: Transcript profiling of glycan-related genes*. *Journal of Biological Chemistry* (Vol. 283).
- Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., ... Bamshad, M. J. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*, 42(1), 30–35.
- Nickel, N., Cleven, A., Enders, V., Lisak, D., Schneider, L., & Methner, A. (2016). Androgen-inducible gene 1 increases the ER Ca(2+) content and cell death susceptibility against oxidative stress. *Gene*, 586(1), 62–68.
- Nurnberger II, Koller DL, Jung J, et al. Identification of Pathways for Bipolar Disorder A Meta-analysis. *JAMA psychiatry*. 2014;71(6):657-664
- Opdenakker, G., Rudd, P. M., Ponting, C. P., & Dwek, R. A. (1993). Concepts and principles of glycobiology. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, 7(14), 1330–7.
- Petrovics, G., Li, H., Stümpel, T., Tan, S. H., Young, D., Katta, S., ... Srivastava, S. (2015). A novel genomic alteration of LSAMP associates with aggressive prostate cancer in African American men. *EBioMedicine*, 2(12), 1957–1964.
- Pucic, M., Knezevic, A., Vidic, J., Adamczyk, B., Novokmet, M., Polasek, O., ... Lauc, G. (2011). High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol Cell Proteomics*, 10(10), M111.010090.

- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. a R., Bender, D., ... Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575.
- Ranganathan, S., Noyes, N. C., Migliorini, M., Winkles, J. A., Battey, F. D., Hyman, B. T., ... Strickland, D. K. (2011). LRAD3, A Novel Low-Density Lipoprotein Receptor Family Member That Modulates Amyloid Precursor Protein Trafficking. *The Journal of Neuroscience*, *31*(30), 10836–10846.
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science (New York, N.Y.)*, *273*(5281), 1516–7.
- Scanlan, C. N., Burton, D. R., & Dwek, R. a. (2008). Making autoantibodies safe. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4081–4082.
- Sidore, C., Busonero, F., Maschio, A., Porcu, E., Naitza, S., Zoledziewska, M., ... Abecasis, G. R. (2015). Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nature Genetics*, *47*(11), 1272–1281.
- Skropeta, D. (2009). The effect of individual N-glycans on enzyme activity. *Bioorganic and Medicinal Chemistry*, *17*(7), 2645–2653.
- Spencer, C. C. a, Su, Z., Donnelly, P., & Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genetics*, *5*(5), e1000477.
- Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., ... Korbel, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, *526*(7571), 75–81.
- Suh, K. S., Malik, M., Shukla, A., Ryscavage, A., Wright, L., Jividen, K., ... Yuspa, S. H. (2012). CLIC4 is a tumor suppressor for cutaneous squamous cell cancer. *Carcinogenesis*, *33*(5), 986–995.
- Swindall, A. F., & Bellis, S. L. (2011). Sialylation of the Fas death receptor by St6Gal-I provides protection against Fas-mediated apoptosis in colon carcinoma cells. *Journal of Biological Chemistry*, *286*(26), 22982–22990.
- The 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–73.
- The Genome of the Netherlands Consortium. (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics*, *46*(8), 818–25.
- Vidarsson, G., Dekkers, G., & Rispen, T. (2014). IgG subclasses and allotypes: From structure to effector functions. *Frontiers in Immunology*, *5*(OCT), 1–17.

- Wang, J., Balog, C. I. a, Stavenhagen, K., Koeleman, C. a M., Scherer, H. U., Selman, M. H. J., ... Wuhler, M. (2011). Fc-glycosylation of IgG1 is modulated by B-cell stimuli. *Molecular & Cellular Proteomics : MCP*, 10(5), M110.004655.
- Wang, W. Y. S., Barratt, B. J., Clayton, D. G., & Todd, J. a. (2005). Genome-wide association studies: theoretical and practical concerns. *Nature Reviews. Genetics*, 6(2), 109–18.
- Weiss, K. M., Weiss, K. M., Clark, A. G., & Clark, A. G. (2002). Linkage disequilibrium and the mapping of complex human traits. *Trends in Genetics : TIG*, 18(1), 19–24.
- Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145), 661–78.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., ... Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, 42(Database issue), D1001–6.
- Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., ... Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46(11), 1173–86.
- Wu, Y., Sato, F., Yamada, T., Bhawal, U. K., Kawamoto, T., Fujimoto, K., ... Kijima, H. (2012). The BHLH transcription factor DEC1 plays an important role in the epithelial-mesenchymal transition of pancreatic cancer. *International Journal of Oncology*, 41(4), 1337–1346.
- Yamamoto-Hino, M., Kanie, Y., Awano, W., Aoki-Kinoshita, K. F., Yano, H., Nishihara, S., ... Goto, S. (2010). Identification of genes required for neural-specific glycosylation using functional genomics. *PLoS Genetics*, 6(12), 1–11.
- Zondervan, K. T., & Cardon, L. R. (2007). Designing candidate gene and genome-wide case-control association studies. *Nature Protoc*, 2, 2492–2501.

KASUTATUD VEEBIAADRESSID

<http://www.ensembl.org/index.html>

<http://genome.sph.umich.edu/wiki/EPACTS>

<http://gump.qimr.edu.au/general/gabrieC/LocusTrack/>

https://mathgen.stats.ox.ac.uk/impute/impute_v2.html

https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html

https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html

<http://ncbi.nlm.nih.gov/>

<http://pngu.mgh.harvard.edu/~purcell/plink/>

LISAD

Lisa 1.

Assotsiatsioonianalüüsid kasutatud fenotüübid koos kirjeldusega.

IgG glycans INITIAL PROFILE	IGP1	<i>The percentage of FA1 glycan in total IgG glycans</i>
	IGP2	<i>The percentage of A2 glycan in total IgG glycans</i>
	IGP3	<i>The percentage of FA2 glycan in total IgG glycans</i>
	IGP4	<i>The percentage of M5 glycan in total IgG glycans</i>
	IGP5	<i>The percentage of FA2B glycan in total IgG glycans</i>
	IGP6	<i>The percentage of A2G1 glycan in total IgG glycans</i>
	IGP7	<i>The percentage of FA2[6]G1 glycan in total IgG glycans</i>
	IGP8	<i>The percentage of FA2[3]G1 glycan in total IgG glycans</i>
	IGP9	<i>The percentage of FA2[6]BG1 glycan in total IgG glycans</i>
	IGP10	<i>The percentage of FA2[3]BG1 glycan in total IgG glycans</i>
	IGP11	<i>The percentage of A2G2 glycan in total IgG glycans</i>
	IGP12	<i>The percentage of A2BG2 glycan in total IgG glycans</i>
	IGP13	<i>The percentage of FA2G2 glycan in total IgG glycans</i>
	IGP14	<i>The percentage of FA2BG2 glycan in total IgG glycans</i>
	IGP15	<i>The percentage of FA2G1S1 glycan in total IgG glycans</i>
	IGP16	<i>The percentage of A2G2S1 glycan in total IgG glycans</i>
	IGP17	<i>The percentage of FA2G2S1 glycan in total IgG glycans</i>
	IGP18	<i>The percentage of FA2BG2S1 glycan in total IgG glycans</i>
	IGP19	<i>Structure not determined</i>
	IGP20	<i>The percentage of A2G2S2 glycan in total IgG glycans</i>
	IGP21	<i>The percentage of A2BG2S2 glycan in total IgG glycans</i>
	IGP22	<i>The percentage of FA2G2S2 glycan in total IgG glycans</i>
	IGP23	<i>The percentage of FA2BG2S2 glycan in total IgG glycans</i>
IgG CHARGED glycans - derived parameters	IGP24	<i>The percentage of sialylation of fucosylated galactosylated structures without bisecting GlcNAc in total IgG glycans</i>
	IGP25	<i>The percentage of sialylation of fucosylated galactosylated structures with bisecting GlcNAc in total IgG glycans</i>
	IGP26	<i>The percentage of sialylation of all fucosylated structures without bisecting GlcNAc in total IgG glycans</i>
	IGP27	<i>The percentage of sialylation of all fucosylated structures with bisecting GlcNAc in total IgG glycans</i>
	IGP28	<i>The percentage of monosialylation of fucosylated monogalactosylated structures in total IgG glycans</i>
	IGP29	<i>The percentage of monosialylation of fucosylated digalactosylated structures in total IgG glycans</i>
	IGP30	<i>The percentage of disialylation of fucosylated digalactosylated structures in total IgG glycans</i>
	IGP31	<i>The percentage of monosialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans</i>
	IGP32	<i>The percentage of disialylation of fucosylated digalactosylated structures with bisecting GlcNAc in total IgG glycans</i>
	IGP33	<i>Ratio of all fucosylated (+/- bisecting GlyNAc) monosialylated and disialylated structures in total IgG glycans</i>
	IGP34	<i>Ratio of fucosylated (without bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans</i>
	IGP35	<i>Ratio of fucosylated (with bisecting GlcNAc) monosialylated and disialylated structures in total IgG glycans</i>
	IGP36	<i>Ratio of all fucosylated sialylated structures with and without bisecting GlcNAc</i>
	IGP37	<i>Ratio of fucosylated monosialylated structures with and without bisecting GlcNAc</i>
	IGP38	<i>The incidence of bisecting GlcNAc in all fucosylated monosialylated structures in total IgG glycans</i>
	IGP39	<i>Ratio of fucosylated disialylated structures with and without bisecting GlcNAc</i>
	IGP40	<i>The incidence of bisecting GlcNAc in all fucosylated disialylated structures in total IgG glycans</i>

IgG NEUTRAL glycans	IGP41	<i>The percentage of GP1 glycan in total neutral IgG glycans (GPn)</i>
	IGP42	<i>The percentage of GP2 glycan in total neutral IgG glycans (GPn)</i>
	IGP43	<i>The percentage of GP4 glycan in total neutral IgG glycans (GPn)</i>
	IGP44	<i>The percentage of GP5 glycan in total neutral IgG glycans (GPn)</i>
	IGP45	<i>The percentage of GP6 glycan in total neutral IgG glycans (GPn)</i>
	IGP46	<i>The percentage of GP7 glycan in total neutral IgG glycans (GPn)</i>
	IGP47	<i>The percentage of GP8 glycan in total neutral IgG glycans (GPn)</i>
	IGP48	<i>The percentage of GP9 glycan in total neutral IgG glycans (GPn)</i>
	IGP49	<i>The percentage of GP10 glycan in total neutral IgG glycans (GPn)</i>
	IGP50	<i>The percentage of GP11 glycan in total neutral IgG glycans (GPn)</i>
	IGP51	<i>The percentage of GP12 glycan in total neutral IgG glycans (GPn)</i>
	IGP52	<i>The percentage of GP13 glycan in total neutral IgG glycans (GPn)</i>
	IGP53	<i>The percentage of GP14 glycan in total neutral IgG glycans (GPn)</i>
	IGP54	<i>The percentage of GP15 glycan in total neutral IgG glycans (GPn)</i>
	IgG NEUTRAL glycans - derived parameters	IGP55
IGP56		<i>The percentage of monogalactosylated structures in total neutral IgG glycans</i>
IGP57		<i>The percentage of digalactosylated structures in total neutral IgG glycans</i>
IGP58		<i>The percentage of all fucosylated (+/- bisecting GlcNAc) structures in total neutral IgG glycans</i>
IGP59		<i>The percentage of fucosylation of agalactosylated structures</i>
IGP60		<i>The percentage of fucosylation of monogalactosylated structures</i>
IGP61		<i>The percentage of fucosylation of digalactosylated structures</i>
IGP62		<i>The percentage of fucosylated (without bisecting GlcNAc) structures in total neutral IgG glycans</i>
IGP63		<i>The percentage of fucosylation (without bisecting GlcNAc) of agalactosylated structures</i>
IGP64		<i>The percentage of fucosylation (without bisecting GlcNAc) of monogalactosylated structures</i>
IGP65		<i>The percentage of fucosylation (without bisecting GlcNAc) of digalactosylated structures</i>
IGP66		<i>The percentage of fucosylated (with bisecting GlcNAc) structures in total neutral IgG glycans</i>
IGP67		<i>The percentage of fucosylation (with bisecting GlcNAc) of agalactosylated structures</i>
IGP68		<i>The percentage of fucosylation (with bisecting GlcNAc) of monogalactosylated structures</i>
IGP69		<i>The percentage of fucosylation (with bisecting GlcNAc) of digalactosylated structures</i>
IGP70		<i>Ratio of fucosylated structures with and without bisecting GlcNAc</i>
IGP71		<i>The incidence of bisecting GlcNAc in all fucosylated structures in total neutral IgG glycans</i>
IGP72		<i>Ratio of fucosylated non-bisecting GlcNAc structures and all structures with bisecting GlcNAc</i>
IGP73		<i>Ratio of structures with bisecting GlcNAc and all fucosylated structures (+/- bisecting GlcNAc)</i>
IGP74		<i>Ratio of fucosylated digalactosylated structures with and without bisecting GlcNAc</i>
IGP75		<i>The incidence of bisecting GlcNAc in all fucosylated digalactosylated structures in total neutral IgG glycans</i>
IGP76		<i>Ratio of fucosylated digalactosylated non-bisecting GlcNAc structures and all digalactosylated structures with bisecting GlcNAc</i>
		<i>Ratio of digalactosylated structures with bisecting GlcNAc and all fucosylated digalactosylated structures (+/- bisecting GlcNAc)</i>

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Timo Tõnis Sikka

(sünnikuupäev: 01.06.1991)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

„Immuunoglobuliin G glükosüleerimist mõjutavad DNA variatsioonid Eesti populatsioonis”,

mille juhendaja on Tõnu Esko,

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 27.05.2016