

TARTU ÜLIKOOL  
LOODUS- JA TÄPPISTEADUSTE VALDKOND  
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT

**Ülegenoomne assotsiatsioonuring viirushepatiidiga seonduvate geneetiliste markerite  
leidmiseks**

Bakalaureusetöö (12 EAP)

Siim Kurvits

Juhendaja Toomas Haller, PhD

TARTU 2018

# INFOLEHT

## **Ülegenoomne assotsiatsiooniuuring viirushepatiidiga seonduvate geneetiliste markerite leidmiseks**

Viirushepatiiti põhjustavad mitmed erinevad levinud viirused. Haigus on maailmas väga levinud, WHO hinnangul oli 2015. aastal 325 miljonit kroonilisse hepatiiti haigestunut. Haiguse tagajärjed võivad olla väga tõsised põhjustades maksatsirroosi ja maksavähki. Antud töö eesmärgiks oli analüüsida erinevate geenivariantide mõju viirushepatiiti haigestumisele. Töös kasutati ülegenoomse assotsiatsiooniuuringu (GWAS) meetodikat ja selle tulemusena leiti mitmeid statistiliselt olulisi assotsiatsioone.

Märksõnad: GWAS, viirushepatiit

B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

## **Genome-Wide Association Study of Viral Hepatitis**

Wide range of viruses can cause viral hepatitis. WHO has estimated that 325 million people world-wide had chronic viral hepatitis in the year 2015, making it a very prevalent infectious disease. Viral hepatitis can lead to cirrhosis and liver cancer. The aim of this study was to analyze the role of genetic variants on viral hepatitis. Genome-wide association studies (GWAS) were conducted and many statistically relevant associations were found.

Keywords: GWAS, viral hepatitis

B110 Bioinformatics medical informatics biomathematics, biometrics

## SISUKORD

INFOLEHT .....	2
KASUTATUD LÜHENDID.....	5
SISSEJUHATUS.....	6
1. KIRJANDUSE ÜLEVAADE.....	7
1.1 Viirushepatiit.....	7
1.1.1 Viirushepatiidi sümptomid.....	7
1.1.2 Viirushepatiidi diagnoosimise kriteeriumid.....	8
1.1.3 Viirushepatiidi ennetamine ja ravivõimalused.....	8
1.1.4 Varasemad viirushepatiidi ülegenoomsed assotsiatsiooniuuringud .....	8
1.2 Geneetiline varieeruvus.....	10
1.2.1 Mõju haigestumisele.....	10
1.2.2 Tuvastamine.....	12
1.3 Ülegenoomne assotsiatsiooniuuring (GWAS).....	12
1.3.1 Statistilised meetodid.....	13
1.3.2 Tulemused.....	14
1.3.3 Probleemid.....	14
1.3.2 SNPTEST .....	15
2. UURIMUS .....	17
2.1 Töö eesmärgid .....	17
2.2 Materjalid ja meetodika .....	17
2.2.1 Valim .....	17
2.2.2 Genotüpiseerimine, imputeerimine ja kvaliteedikontroll.....	19
2.2.3 Assotsiatsioonianalüüsid.....	20
2.2.5 Metaanalüüs .....	20
2.3 Tulemused.....	21
2.3.1 B15 (äge A-hepatiit).....	22
2.3.2 B16 (äge B-hepatiit) .....	24
2.3.3 B18 (krooniline B- või C-hepatiit) .....	27
2.3.4 B19 (täpsustamata hepatiit).....	31
2.4 Arutelu .....	34
2.4.1 Tulevikuplaanid.....	36
Kokkuvõte.....	38
Genome-Wide Association Study of Viral Hepatitis .....	39

Tänuõnad.....	40
Kasutatud kirjandus.....	41
Kasutatud veebiaadressid.....	45
Lisad.....	46
Lisa 1 Viirushepatiidi RHK-10 klassifikatsioon.....	46
Lisa 2 Q-Q joonised.....	47
Lih litsents lõputöö reprodutseerimiseks ja lõputööüldsusele kättesaadavaks tegemiseks.....	51

## KASUTATUD LÜHENDID

<b>CI</b>	usaldusvahemik,	<i>confidence interval</i>
<b>GWAS</b>	ülegenoomne assotsiatsiooniuuring,	<i>Genome-Wide Association Study</i>
<b>HAV</b>	A-hepatiidi viirus,	<i>hepatitis A virus</i>
<b>HBV</b>	B-hepatiidi viirus,	<i>hepatitis B virus</i>
<b>HCC</b>	hepatotsellulaarne kartsinoom,	<i>hepatocellular carcinoma</i>
<b>HCV</b>	C-hepatiidi viirus,	<i>hepatitis C virus</i>
<b>HDV</b>	D-hepatiidi viirus,	<i>hepatitis D virus</i>
<b>HEV</b>	E-hepatiidi viirus,	<i>hepatitis E virus</i>
<b>KMI</b>	kehamassi indeks,	<i>bodymass index</i>
<b>LD</b>	alleelne aheldatus,	<i>linkage disequilibrium</i>
<b>MAF</b>	harvema alleeli sagedus,	<i>minor allele frequency</i>
<b>OR</b>	šansside suhe,	<i>odds ratio</i>
<b>PCA</b>	peakomponent analüüs,	<i>principal component analysis</i>
<b>PC</b>	peakomponent,	<i>principal component</i>
<b>Q-Q</b>	kvantiil-kvantiil,	<i>quantile-quantile</i>
<b>RHK-10</b>	rahvusvaheline haiguste klassifikatsioon 10,	<i>The International Classification of Diseases 10, ICD-10</i>
<b>SNP</b>	ühenukleotiidne variant,	<i>single nucleotide variant</i>

## SISSEJUHATUS

Viirushepatiiti ehk viiruslikku maksapõletikku põhjustavad üksteisest fülogeneetiliselt kauged viirused. Näiteks A-hepatiidi viirus kuulub *Picornaviridae* sugukonda, B-hepatiidi viirus sugukonda *Hepadnaviridae* ja C-hepatiidi viirus sugukonda *Flaviviridae*. A- ja C- hepatiidi viirused on üheaheelalise positiivse polaarsusega RNA viirused, samas B-hepatiidi viirus on pöördtranskriptsiooniline kaheaheelaline DNA viirus (ICTV *International Committee on Taxonomy of Viruses*). Vaatamata viiruste bioloogilistele erinevustele põhjustavad nad kõik haigustekitajana sarnaseid maksakahjustusi. (Samji, 2017)

Inimesed pole kõikidele haigustekitajatele ühtmoodi vastuvõtlikud. Malaaria eest kaitset pakkuv sirprakulise aneemia alleel ja HIV-1 residentsust andev CCR5 $\Delta$ 32 alleel on tuntud näited nakkushaiguste eest kaitsvatest geenivariantidest (Dean *et al.*, 1996). Ülegenoomne assotsiatsiooniuuring (GWAS, *Genome-wide association study*) võimaldab otsida sarnaseid geenivariante, mis kaitseksid hepatiidiviiruste eest või vastupidiselt süvendaksid haiguse kulgu. Geneetiliste markerite leidmine võimaldab hinnata riski haigestuda (Läll *et al.*, 2017) ja annab täiendavat informatsiooni haiguse täpsemate molekulaarsete mehhanismide kohta.

Käesoleva töö eesmärgiks on identifitseerida viirusliku hepatiiti haigestumist mõjutavaid geneetilisi markereid (SNP, *single nucleotide polymorphism*) kasutades selleks Tartu Ülikooli Genoomika Instituudi (TÜGI) Eesti Geenivaramu geenidoonorite andmestikku.

# 1. KIRJANDUSE ÜLEVAADE

## 1.1 VIIRUSHEPATIIT

Viirushepatiidi peamisteks põhjustajateks on viis viirust: A-hepatiidi viirus (HAV), B-hepatiidi viirus (HBV), C-hepatiidi viirus (HCV), D-hepatiidi viirus (HDV) ja E-hepatiidi viirus (HEV). Ainult HBV, HCV ja HDV põhjustavad kroonilist hepatiiti, mis omakorda põhjustab maksatsirroosi ja vähki. Samuti on HBV ja HCV suurimad viirushepatiiti suremuse põhjustajad moodustades 96% surmajuhtudest. Aastas sureb viirushepatiidi poolt põhjustatud maksakahjustuste: maksatsirroosi, hepatotsellulaarse kartsinoomi (HCC) tõttu 1,4 miljonit inimest.(GBD 2013 *Mortality and Causes of Death Collaborators*, 2015) Viirushepatiiti võib haigestuda igas eas ja hepatiidiviirused on levinud praktiliselt kõikjal. WHO hinnangul elas 2015. aastal maailmas 325 miljonit kroonilisse viirushepatiiti haigestunut (WHO, 2017). Suurim kroonilise B- ja C-hepatiidi levimus on Aasias ja Aafrikas, kus nakatumine toimub peamiselt sünnil või varases lapsepõlves. Euroopas ja Põhja-Ameerikas on seevastu peamiseks nakatumise põhjuseks süstitavate uimastite kasutamine ja riskantne seksuaalkäitumine. (Lok ja McMahon, 2007) Mitte kõik viirushepatiidid ei levi kehavedelikega, HAV-i ja HEV-i peamiseks levikuvektoriteks on saastunud vesi ja toit.

### 1.1.1 VIIRUSHEPATIIDI SÜMPTOMID

Viirushepatiit esineb 2 vormis: äge ja krooniline. Äge viirushepatiit võib kulgeda sümptomiteta, põhjustada vaid gripilaadset haigestumist või tekitada eluohtliku maksakooma. Sümptomid sõltuvad nii inimese immuunvastusest kui ka konkreetsest viirusest. A- ja C- hepatiidiga kaasnevad enamasti kerged sümptomid. Raskemat haiguskulgu põhjustavad B- ja E-hepatiit, kusjuures B-hepatiidiga võib kaasneda D-hepatiidi koinfektsioon, mis tekitab raskemaid komplikatsioone ja suurenda maksakooma riski. Ägeda viirushepatiidiga kaasnevateks sümptomiteks on söögiisu langus, iiveldus ja oksendamine, palavik, valu maksapiirkonnas. Samuti võib hilisemalt tekkida kollatõbi – naha ja silmavalgete kollakaks muutmine, sest maks ei ole enam võimeline veres olevat bilirubiini eemaldama (MSD *manual*).

Isegi ilma ravimiteta taastatakse ägedast hepatiidist enamasti 4-8 nädala jooksul. C-hepatiiti haigestunudel on endiselt kõrge risk jääda viiruse kandjateks (B-hepatiidi korral on risk madalam), mis tähendab, et nad levitavad viirust edasi ning võivad haigestuda kroonilisse hepatiiti (MSD *manual*). Krooniline viirushepatiit suurendab aga oluliselt maksatsirroosi ja HCC-sse haigestumise riski (Chemin ja Zoulim, 2009).

### 1.1.2 VIIRUSHEPATIIDI DIAGNOOSIMISE KRITERIUMID

Viirushepatiidi diagnoosimiseks on kasutusel mitmed erinevad meetodid. Levinud diagnostiliseks vahendiks on seroloogilised uuringud, mis põhinevad veres sisalduvate viirusele spetsiifiliste antikehade (ja antigeenide) tuvastamisel. Veres leiduvad antikehad ja antigeenid on erinevad, sõltudes nii hepatiiti põhjustavast viirusest kui ka viirusinfektsiooni kulust. Samuti on võimalik, et positiivne seroloogiline test tuleneb varasemast haigestumisest. Täpsema vahendina on lisaks kasutusel molekulaardiagnostiline meetod: qRT-PCR. Eelnimetatud meetodikaga on võimalik tuvastada millise hepatiidiviirusega on tegemist aga ka määrata viiruse tiitrit (osakeste arvu). (Thomas *et al.*, 2014)

### 1.1.3 VIIRUSHEPATIIDI ENNETAMINE JA RAVIVÕIMALUSED

Viirushepatiidi vastu vaktsineerimine on võimalik vaid osaliselt, sest välja on töötatud ainult HAV, HBV ja HBE vastased vaktsiinid. Haiguse leviku kontrollimiseks on osutunud edukaks vahendiks vastsündinute vaktsineerimine HBV vastase vaktsiiniga (Nelson *et al.*, 2016). Kroonilise B-hepatiidi vastase ravimina on kasutusel Tenofovir. Antiviraalsed teraapiad on olemas ka C-hepatiidi raviks, kuid oluline on raviga varakult alustada. Õigeaegse alustamise raviga teeb keeruliseks asjaolu, et hinnanguliselt on vähem kui 5% HCV-ga nakatunutest teadlikud oma positiivsest staatusest. Teadlikkus kahaneb madala sissetulekuga kogukondades veelgi jäädes alla 1% (WHO, 2017).

### 1.1.4 VARASEMAD VIIRUSHEPATIIDI ÜLEGENOOMSED ASSOTSIATSIOONIUURINGUD

Nakkushaiguste nagu näiteks viirusliku hepatiidi ülegenoomsed assotsiatsiooniuringud on kasulikud mitmeti. Esmalt on võimalik tuvastada geenivariante, mis suurendavad või vähendavad haigusrisiki. B-hepatiidi haigusriskiga on tugevalt seostatud mitmetes varasemates uuringutes HLA-DP lookust (tabel 1). Kusjuures leidub nii riski suurendavaid kui vähendavaid

variante. (Kamatani *et al.*, 2009) Kuid lisaks annavad uuringu leiud täpsustavat informatsiooni infektsiooni molekulaarsete mehhanismide kohta. Näiteks on eksperimentaalselt tõestatud, et GWAS-i leid INST10 geeni lähedal on omakorda seotud maksas ekspresseeritava IRF3-ga, mille tulemusel supresseeritakse HBV replikatsiooni maksarakkudes (Li *et al.*, 2016).

**Tabel 1 HBV GWAS varasemad tulemused**

SNP	Asukoht	Lähim geen	Riskifaktor OR (CI 95%)	P- väärtus	Viide
rs3077	6 (33141000)	<i>HLA-DPA1</i>	0,56 (-)	2,31E-38	Katamani <i>et al.</i> , 2009
rs9277535	6 (33162839)	<i>HLA-DPB1</i>	0,57 (-)	6,34E-39	Katamani <i>et al.</i> , 2009
rs2856718	6 (32778233)	<i>HLA-DQB1</i>	1,56 (1,45 – 1,67)	3,99E-37	Mbarek <i>et al.</i> , 2011
rs7453920	6 (32837990)	<i>HLA-DQB2</i>	1,81 (1,62 – 2,01)	5,98E-28	Mbarek <i>et al.</i> , 2011
rs3130542	6p21.33	<i>HLA-C</i>	1,33 (1,23–1,44)	9,49E-14	Hu <i>et al.</i> , 2013
rs652888	6 (-)	<i>EHMT2</i>	1,26 (1,07-1,47)	2,78E-6	Kim <i>et al.</i> , 2013
rs1419881	6 (-)	<i>TCF19</i>	0,76 (0,68-0,86)	4,51E-11	Kim <i>et al.</i> , 2013
rs12614	6 (31914179)	<i>CFB</i>	1,89 (1,69-2,08)	1,28E-34	Jiang <i>et al.</i> , 2015
rs1883832	20 (44746982)	<i>CD40</i>	1,19 (1,14-1,25)	2,95E-15	Jiang <i>et al.</i> , 2015
rs422951	6 (32188383)	<i>NOTCH4</i>	1,27 (1,20-1,35)	5,33E-16	Jiang <i>et al.</i> , 2015
rs7000921	8p21.3	<i>INTS10</i>	0,78 (-)	3,2E-12	Li <i>et al.</i> , 2016
rs7453920	6p21.32	-	0.53 (0.48–0.59)	4,93E-37	Hu <i>et al.</i> , 2013
rs4821116	22q11.21	-	0.82 (0.77–0.87)	1,71E-12	Hu <i>et al.</i> , 2013

Kroonilise B-hepatiidi infektsiooniga on seostatud ka INF- $\gamma$ , vitamiin D retseptorit, östrogeen retseptor-alfat (Khanh ja Lan, 2012).

HCV-ga seonduvad GWAS-id on suunitletud peamiselt paremate ravivõimaluste leidmiseks. Interferoon lambda-3 kodeeriv IL28B lookuses leiduv geenivariant on tugevalt assotsieerunud parema ravivastusega. Geenivariandi kandjatel on 2-3 korda suurem tõenäosus viirusest vabaneda pegüleeritud interferoon alfa ja ribaviriin teraapiaga.(Ge *et al.*, 2009) Lisaks on seostaud veel HLA-d, KIR-i, interleukiine ja mitmeid kemokiine C-hepatiidi infektsiooniga.

## 1.2 GENEETILINE VARIEERUVUS

Inimesed on DNA järjestuse alusel 99,5% ulatuses identsed (Pettersson *et al.*, 2009). Leidmaks geneetilisi eelsoodumusi haigestumiseks või haiguse raskemaks kuluks keskendume genoomis olevatele erinevustele. Üheks suurimaks (üle 90%) geneetilise varieeruvuse põhjustajaks on ühenukleotiidilised muutused (SNP-id) DNA järjestuses (Brookes, 1999).*International HapMap Project* on rahvusvaheline projekt, mille eesmärk on kaardistada geneetilist varieeruvust ja määrata esinevate alleelide sagedusi inimpopulatsioonides. Leitud SNP-id on koondatud vabalt kättesaadavasse dbSNP andmebaasi. Aastal 2008 alustanud 1000G projekt võttis eesmärgiks veelgi täpsema geneetilise varieeruvuse kaardistamise. Selle tulemusel kirjeldati 84,7 miljonit SNP-i (1000 Genome Project Consortium, 2015). Geneetilise varieeruvuse kaardistamine on vajalik edasiste spetsiifiliste uuringute läbiviimiseks. Mõlemad projektid on seega olulised alutalad edasisteks inimesegeneetika teadustöödeks.

### 1.2.1 MÕJU HAIGESTUMISELE

Alleelisagedust ja selle mõju suurust hindav hüpotees (CD-CV, *common disease-common variant*) väidab, et sagedaste haiguste geneetilise tausta määravad mitmed populatsioonis levinud geenivariandid. Teisisõnu tähendab see, et sagedased haigused on polügeensed olles mõjutatud mitmete genoomipiirkondade poolt ja üksiku geenivariandi mõju haiguse tekkeks on sealjuures vähene (Bush ja Moore, 2012). Kuid mitte kõik haigused pole polügeensed. Leidub ka Mendeliaalseid ehk monogeenseid haigusi, mis kujunevad üksiku geenivariandi mõju tõttu.

		<b>Mõju suurus</b>	
		<b>väike</b>	<b>suur</b>
<b>Alleeli sagedus</b>	<b>haruldane</b>	Raske tuvastada	<b>Haruldased Mendeliaalsed haigused</b>
	<b>levinud</b>	<b>Sagedased polügeensed haigused</b>	Harva esinev

**Joonis 1** Geneetiliste haiguste jaotus lähtuvalt alleeli sagedusest ja mõjust

Väikse mõjuga väga harvade alleelide tuvastamine on keeruline. Samuti ei ole teada juhtumeid, kus populatsioonis väga levinud alleelid põhjustaksid rasket haigestumist. Geneetiliste haiguste jaotust lähtuvalt alleeli sagedusest ja mõjust illustreerib joonis 1 (Bush ja Moore, 2012).

Nakkushaigused on geneetiliselt keerukad, sest haigestumisel mängib rolli nii peremeesorganismi geneetika, keskkond kui ka patogeeni enda genoom. Uuringud kaksikutega on näidanud, et ka nakkushaigused nagu näiteks tuberkuloos, leepira, malaaria on mõjutatud peremehegeneetika poolt (Jepson *et al.*, 1995; Kallman *et al.*, 1943; Chakravarti *et al.*, 1973). Teada on ka mitmeid Mendeliaalseid geenivariante, mis pakuvad kaitse haigustekitajate eest. Tuntuimaks on ilmselt CCR5del32 homosügootid, kes on HIV-1 immuunsed (Dean *et al.*, 1996). Nimelt tekitab deletsioon enneaegse stop-koodoni nii, et raku pinnal ei ekspresseeru kemokiini retseptor 5, mida viirus kasutab koretseptorina, et peremeesrakku siseneda. Konkreetsete molekulaarse mehhanismi avastamine on kasulik, sest see annab võimalusi uute ravimite väljatöötamiseks. Näiteks Maraviroc on CCR5 antagonist, mida saab seega kasutada HIV-1 vastases ravis (Farkenhauer *et al.*, 2005).

### 1.2.2 TUVASTAMINE

DNA-mikrokiip on tahke kandja, millele on kovalentselt kinnitatud oligonukleotiidsed järjestused, mis hübriidiseerivad komplementaarsete DNA järjestustega. Hübriidiseerumise käigus tekkiva valgussignaali abil on võimalik kiiresti määrata sadu tuhandeid (miljoneid) SNP-e. Paljud biopangad, sealhulgas Tartu Ülikooli Genoomika Instituut (TÜGI) kasutavad inimeste genotüpiseerimiseks mikrokiipide tehnoloogiat. Mikrokiipide eeliseks on suhteliselt odav hind ja kõrge genotüpiseerimise kiirus. Peamiseks puudujäägiks aga asjaolu, et mikrokiipidega pole võimalik määrata kogu geneetilist varieeruvust. Mikrokiip on sobilik vaid tootja poolt määratud spetsiifiliste genoomiosade tuvastamiseks. Alternatiivina on võimalik huvipakkuv või kogu genoom sekveneerida, mille tulemusel tekib pidev nukleotiidide järjestus. Sekveneerimisel saadakse oluliselt rohkem informatsiooni, kuid meetod on ka oluliselt kallim.

Aheldatuse tasakaalustamatus (LD, *linkage disequilibrium*) on pidevale genoomijärjestusele paigutatud SNP-ide omadus, mis kirjeldab kuivõrd ühe SNP-i pärandumine on korrelatsioonis teise populatsioonis esineva SNP-i pärandumisega (Weiss *et al.*, 2002). LD määr sõltub populatsioonist, mida suurem on aset leidnud rekombinatsioonisündmuste arv, seda väiksem on LD määr. Seega näiteks Aafrika populatsioonides, mis on võrreldes Euroopa või Aasiaga vanemad on väiksemad LD regioonid.

LD olemasolu võimaldab kasutada nõ kaudseid SNP-e (*tag-SNP*), mis kirjeldavad haplotüüpe ehk SNP-ide grupe. See tähendab, et on vaja vähem positsioone genotüpiseerida, mis muudab uuringu teostamise odavamaks ja vähendab ka tehnilist mahukust. (Gabriel *et al.*, 2002) Hinnanguliselt piisab 500 000 – 1 000 000 SNP-i määramisest, et katta kogu genoom.

### 1.3 ÜLEGENOOMNE ASSOTISATSIOONIUURING (GWAS)

Ülegenoomne assotsiatsiooniuring (GWAS) on hüpoteesivaba uuringutüüp, millega kirjeldatakse uuritava tunnuse seotust kindla genoomi piirkonnaga. Esialgu kasutati GWAS-i sagedaste polügeensete komplekshaiguste riskifaktorite tuvastamiseks. (Bush ja Moore, 2012) Enamlevinumaks vormiks on juht-kontrolluuring, mis tuvastab haigusega assotsieeruvaid SNP-e. Hilisemalt on lisandunud mitmed GWAS-i vormid, mis võivad erineda nii kasutatava geneetilise markeri (nt. koopiaarvu variatsioonide, CNV) kui ka uuritavate fenotüüpide poolest.

Näiteks on võimalik kasutada GWAS-i ka pidevate tunnustena väljenduvate fenotüüpide (nt. pikkus, metaboliitide tase veres) analüüsil. Olenevalt uuritavast tunnusest valitakse GWAS-i jaoks vastav statistiline test.

GWAS-i saame olemuselt jagada neljaks etapiks. Esiteks tuleb moodustada valim, mis koosneb piisava suurusega juhtude grupist ja kontrollgrupist (pidevate tunnuste korral on tegemist vaid juttude grupiga). Seejärel on vajalik valimi genotüpiseerimine. Kolmandaks viiakse läbi statistilised testid uuritava tunnuse ja geneetiliste markerite vahel. Viimaks tuleks uuringut korrata sõltumatu valimiga. (Bush ja Moore, 2012)

### 1.3.1 STATISTILISED MEETODID

Uuringus kasutatavad statistilised meetodid on määratud lähtuvalt uuritavast fenotüübist. GWAS-i on võimalik läbi viia nii pidevate (nt. pikkus) kui ka diskreetsete (nt. diabeeti haigestumine) tunnustega. Haiguslike fenotüüpide uurimiseks on levinud logistilise regressioonimudeli kasutamine, sest uuritav tunnus on binaarne ja mudel võtab arvesse ka teisi tunnuseid (nt. sugu, vanus, KMI). Mudeli põhjal arvutatakse iga SNP-i kohta olulisuse tõenäosus ehk p-väärtus.

Klassikaliselt on vaja null-hüpoteesi ümberlukkamiseks p-väärtust, mis oleks väiksem kui olulisuse nivoo (*confidence level*) 0,05. Sisuliselt tähendab see, et saadud tulemus esineks juhuslikult vähem kui viiel korral sajast. GWAS-i läbiviimisel arvutatakse sadu tuhandeid või isegi miljoneid p-väärtusi, mistõttu annaks olulisuse nivoo 0,05 palju valepositiivseid tulemusi. Selle vältimiseks tuleks mitmete hüpoteeside statistilisel testimisel korrigeerida usaldusnivoo väärtust. Üks lihtsamaid meetodeid selleks on Bonferroni korrigeerimine (Miller, 1981), mille käigus jagatakse soovitud valepositiivsete tulemuste tõenäosus läbi sooritatud statistiliste testide arvuga, mille tulemusel saadakse korrigeeritud olulisuse nivoo. Seega kasutatakse uuringus tavaliselt oluliselt madalamat olulisuse nivood:  $5 \times 10^{-8}$ , mis on vastavuses 5% valepositiivsete leidudega 1 000 000 sõltumatu assotsiatsioonitesti läbiviimisel (Risch ja Merikangas, 1996; Pe'er *et al.*, 2008).

Kohandamiseks mudelit geneetilisele stratifikatsioonile (süsteemiline alleelisageduste erinevus kontroll ja uuritavas grupis) on levinud peakomponentanalüüsi (PCA, *principal component analysis*) kasutamine. Analüüsi tulemusel saadakse genotüübiandmete põhjal peakomponendid,

mis kirjeldavad võimalikult hästi iga uuritava indiviidi geneetilist variatsiooni võrreldes teistega. See tähendab, et peakomponendid on justkui suurused, mis kirjeldavad alleelisageduste jaotust indiviidide seas, ning kasutades peakomponente on võimalik sarnaseid indiviide grupeerida. Analüüsi tulemusel saadavaid peakomponente saab kasutada logistilises regressioonimudelil analoogselt teistele kovariaatidele (Price *et al.*, 2006). Sel viisil väldime valeassotsiatsioonide saamist geneetiliste mustrite tõttu uuringu- ja kontrollgruppides, mis võivad tuleneda indiviidide etnilisest päritolust, kus alleelisagedused varieeruvad alampopulatsioonides oluliselt (Bush ja Moore, 2012).

### 1.3.2 TULEMUSED

GWAS-i tulemuseks on maatriks, mis sisaldab SNP-e ja neile vastavaid p-väärtusi ja muid statistilisi näitajaid. Kandes SNP-id lähtuvalt nende positsioonile genoomis x-teljele, ning p-väärtuse kümnendlogaritmi y-teljele saame Manhattanijoonise (*Manhattan plot*). Leitud assotsiatsioon on jooniselt näha kõrge pilvelõhkujat meenutava punktide sambana.

Tulemuse valiidsuse hindamiseks kasutatakse kvantiil-kvantiil (Q-Q, *quantile-quantile*) joonist. Q-Q joonise x-teljel on oodatavad p-väärtused ja y-teljel saadud p-väärtused. Korrektsel GWAS-i iseloomustab joonis, mida suuremas osas kirjeldab sirge  $y = x$ , mis aga lõpus tõuseb teravalt. See tähendab, et uuringus kasutatud juhtude ja kontrollide gruppidel ei esinenud üldist geneetilist erinevust ja vähene arv uuritavaid SNP-e tõepoolest assotsieerub uuritava tunnusega.

GWAS-i tulemusel tuvastatud SNP-id ei pruugi olla kausaalsed fenotüübi suhtes. Näiteks võib uuringu tulemusel leida hoopis *tag*-SNP-i, mis on LD kaudu assotsieerunud tegelikku bioloogilist mõju avaldava lookusega. Loomulikult on võimalik ka variant, et uuringu leid asub geenis ja on otseselt fenotüüpi põhjustav. Assotsieeruvad on seega nii markerid, mis otseselt põhjustavad fenotüüpi, kui ka markerid, mis lubavad ennustada uuritava fenotüübi esinemist. (Bush ja Moore, 2012)

### 1.3.3 PROBLEEMID

GWAS-is kasutavad statistilised testid määravad tunnuse assotsiatsiooni geenipiirkonnaga ning arvutatakse ka šansside suhe (OR), kuid määratud seos ei pruugi otseselt fenotüüpi põhjustada.

Lisaks kausaalse seose puudumisele on GWAS-i meetodikal veel mõned kitsaskohad, millest sõltub uuringu kvaliteet.

Võimalike assotsiatsioonide tuvastamiseks on oluline, et uuringu võimsus oleks piisav ehk teisisõnu valimi suurus ei oleks liiga väike. Mitmete väikeste valimite ühendamine metaanalüüsi kasutades on üks võimalus võimsuse kasvatamiseks (Winkler *et al.*, 2014). Lisaks väikesele valimi mahule võib probleeme tekitada ka väike MAF.

Juht-kontrollgrupi uuring puhul on oluline fenotüübi klassifitseerimine. Seega oleks hea kui diagnoosikriteeriumid oleksid üheselt määratud kogu uuringus. (Pearson ja Manolio, 2008) Probleemseteks võivad osutuda näiteks rahvusvahelised uuringud, kus on riigiti erinevad diagnostilised kriteeriumid või ka fenotüübid, mille tõlgendus on väga lai.

### 1.3.2 SNPTEST

SNPTEST on GWAS-i läbiviimise tarkvara, mis loodi Oxfordi Ülikooli teadlaste poolt (. (Marchini ja Howie, 2010) Uuringu teostamiseks tuleb SNPTEST sisendina varustada genotüübi- ja fenotüübiandmetega. Genotüübiandmed hoitakse GEN formaadis, kus invidiidid on esitatud veergudes ja geneetilised markerid ridades. Iga indiviidi ja vastava markeri kohta on toodud kolm veergu (AA, AB, BB), mis näitavad genotüüpide esinemise tõenäosusi. Fenotüübiandmed koondatakse SAMPLE faili, kus iga indiviidi kohta on üks rida. SAMPLE failiformaati lisatakse lisaks uuritavale tunnusele teised analüüsis kasutatavad kovariaadid (nt. sugu, vanus, PC). SAMPLE ja GEN failid peavad olema omavahel ühilduvad ehk isikude järjekord peab mõlemas failis kattuma.

```

ID_1 ID_2 missing cov_1 cov_2 cov_3 cov_4 pheno1 bin1
0 0 0 D D C C P B
1 1 0.007 1 2 0.0019 -0.008 1.233 1
2 2 0.009 1 2 0.0022 -0.001 6.234 0
3 3 0.005 1 2 0.0025 0.0028 6.121 1
4 4 0.007 2 1 0.0017 -0.011 3.234 1
5 5 0.004 3 2 -0.012 0.0236 2.786 0

```

**Joonis 2 SAMPLE faili näidis.** Esimesed kaks veergu sisaldavad isikule unikaalset ID-tunnust, kolmas *missing* rida täidetakse puuduvate andmete korral. Neile järgneb 4 erinevat kovariaati (näiteks peakomponendid). Fenotüüp võib olla nii pidev kui ka binaarne tunnus. Teise rea ülesanne on inditseerida, mis laadi tunnusega on tegemist.

```

SNP1 rs1 1000 A C 1 0 0 1 0 0
SNP2 rs2 2000 G T 1 0 0 0 1 0
SNP3 rs3 3000 C T 1 0 0 0 1 0
SNP4 rs4 4000 C T 0 1 0 0 1 0
SNP5 rs5 5000 A G 0 1 0 0 0 1

```

**Joonis 3 GEN faili näidis.** Vajalik on iga SNP-i nimetamine näiteks rs-koodiga. Sellel järgneb SNP-i asukoht genomis. Seejärel on näha alleele A ja B. Kolm järgnevat numbrit näitavad vastavalt AA, AB ja BB genotüüpide esinemise tõenäosusi indiviidil. Seega on iga uuringus osaleva indiviidi kohta 3 veergu.

## 2. UURIMUS

### 2.1 TÖÖ EESMÄRGID

Käesoleva töö eesmärgiks on leida viirushepatiiti haigestumisega seonduvaid või ka seda ennustavaid geenivariante kasutades Tartu Ülikooli Genoomika Instituudi (TÜGI) andmestikku. Võimalike geenivariantide leidmiseks viidi läbi 4 GWAS-i, mille tulemuste põhjal sooritati metaanalüüs. Uuritavateks fenotüüpideks valiti **B15** (ägedasse A-hepatiiti haigestunud), **B16** (ägedasse B-hepatiiti haigestunud) , **B18** (kroonilisse B- või C-hepatiiti haigestunud) ja **B19** (täpsustamata viirushepatiiti haigestunud) (tabel 2).

### 2.2 MATERJALID JA METOODIKA

#### 2.2.1 VALIM

Töö raames viidi läbi neli juht-kontrolluuringut, mille valimid koosnesid vastavalt 644, 92, 109 ja 150 viirushepatiidihaigest ja 13097 tervest kontrollindiviidist. Juhtgrupid koostati vastavalt tabel 2 RHK-10 kirjeldustele. Grupp B15 koosneb ägedasse A-hepatiiti haigestunutest, B16 ägedasse B-hepatiiti haigestunutest, B18 kroonilisse B- või C-hepatiiti haigestunutest ja B19 täpsustamata viirushepatiiti haigestunutest. Töös kasutati Tartu Ülikooli Genoomika Instituuti Eesti Geenivaramu proove, mis on kogutu üle terve Eesti. Vaatama sellele, et TÜGI valimi pole kogutud lähtuvalt klassikalisele valikuuringule on tegu siiski Eesti rahvast esindava valimiga (Leitsalu *et al.*, 2015). Uuringus osalevate gruppide ealine ja sooline jaotuvus on väljendatud joonisel 4. Uuringus osalejad jaotati gruppidesse vastavate RHK-10 koodi olemasolul. RHK-10 kood võis pärineda ühest või mitmest järgnevast kaheksast registrist (andmed 2015. aasta seisuga):

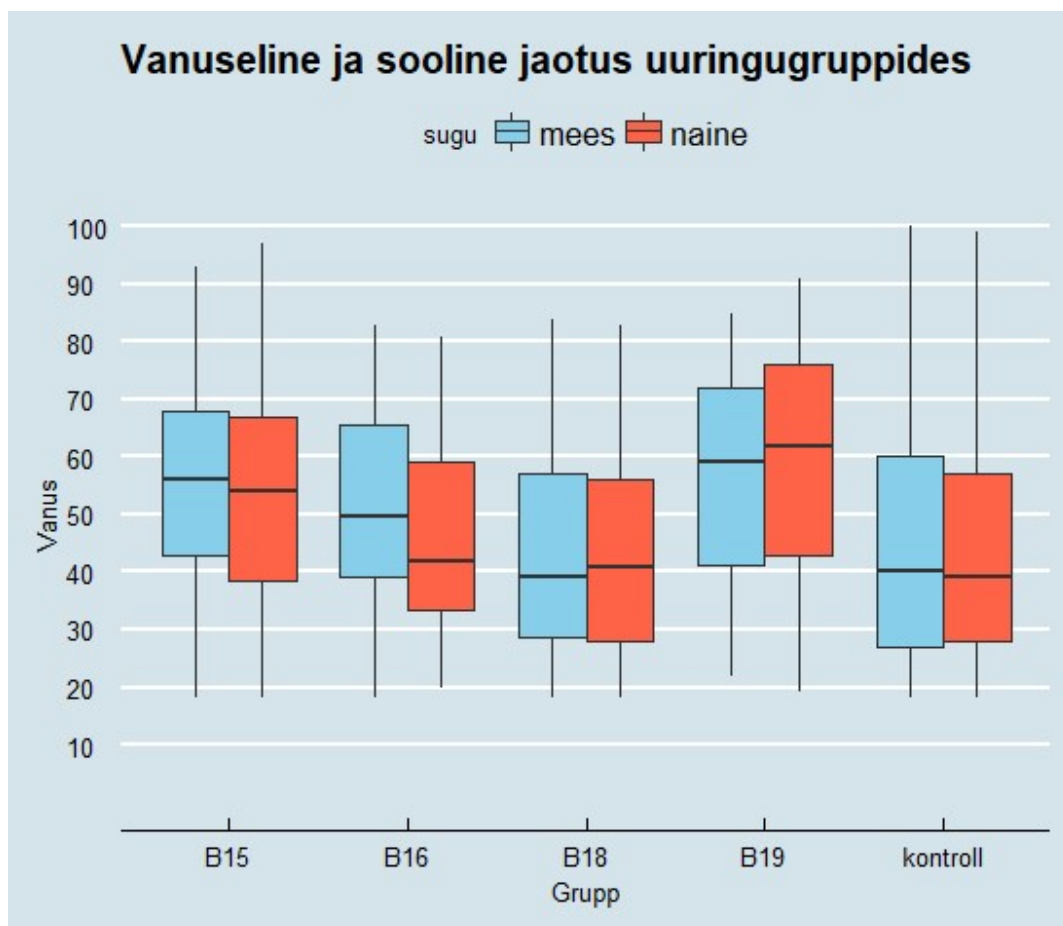
- Eesti Haigekassa
- Müokardiinfarktiregister
- Põhja-Eesti Regionaalhaigla
- Surma põhjuste register
- Tuberkuloosiregister
- Tartu Ülikooli Kliinikum

- Eesti Geenivaramu küsimustik
- Vähiregister

Seega on kasutusel kõik diagnoosi usaldusväärsuse nivood alates eriarsti poolt määratud kuni inimese enda poolt raporteeritule. Viirushepatiiti haigestunuks loeti kõik, kellel vähemalt ühes eelnimetatud registrist oli vastava RHK-10 koodiga sissekanne.

**Tabel 2 GWAS grupid.** Tabelis on toodud uuringugrupi kood, gruppi kuuluvate haiguste nimetused ja RHK-10 koodid, ning grupi suurus.

<b>Grupp</b>	<b>Kirjeldus (RHK-10 kood)</b>	<b>Isikute arv</b>
<b>B15</b>	Hõlmab maksakoomata A-hepatiidiga (B15.0) ja maksakoomaga A-hepatiidiga (B15.9) isikuid.	644
<b>B16</b>	Hõlmab B-hepatiiti delta-agensiga ja maksakoomaga (B16.0), B-hepatiiti delta-agensita ja maksakoomata (B16.1), ägeda B-hepatiidi ja maksakoomaga (B16.2), ägeda B-hepatiidid ja maksakoomata (B16.9) isikuid.	92
<b>B18</b>	Hõlmab järgnevate krooniliste viirushepatiitidega isikuid: B-hepatiit delta-agensiga (B18.0), B-hepatiit delta-agensita (B18.1), C-hepatiit (18.2), muud ja täpsustamata viirushepatiidid (B18.8 ja B18.9)	109
<b>B19</b>	Hõlmab täpsustamata viirushepatiite koomaga (B19.0) ja koomata (19.9).	150
<b>Kontroll</b>	Kõik kellel ei esinenud RHK-10 koode: B15.0, B15.9, B16.0, B16.1, B16.2, B16.9, B18.0, B18.1, B18.2, B18.8, B18.9, B19.0 ja B19.9	13097



**Joonis 4** Uuringus osalenud juht- ja kontrollgruppide vanuseline ja sooline jaotuvus.

## 2.2.2 GENOTÜPISEERIMINE, IMPUTEERIMINE JA KVALITEEDIKONTROLL

Uuringus osalenud isikud genotüpiseeriti kasutades kolme erinevat mikrokiipi: *IlluminaHumanCoreExome*, *IlluminaHumanHap370CNV* ja *IlluminaHumanOmniExpress* (Illumina Inc. San Diego, CA, USA). Genotüpiseerimine viidi läbi Eesti Geenivaramu tuumiklaboris. Seejärel läbisid genotüpiseeritud andmed kvaliteedikontrolli järgnevate parameetrite osas:

- edukalt genotüpiseeritud markerite protsentuaalne osakaal (*Call Rate*, CR);
- harvema alleeli sagedus (*Minor Allele Frequency*, MAF);
- Hardy-Weinbergi tasakaalustatus (*Hardy-Weinberg equilibrium*, HWE);
- genotüübi põhjal leitud soo mitte-kokkulangevus proovi fenotüübi informatsiooniga;

Kontrollitav SNP eemaldati kui:

1. CR < 95%
2. HWE testi p-väärtus <  $10^{-6}$
3. MAF < 1%

Imputeerimine teostati Eesti Geenivaramu teadlaste poolt, kes kasutasid kiipidel puuduolevate markerite imputeerimiseks EGV referentsgeneoomist pärinevaid andmeid. Imputeerimiseks kasutati tarkvara IMPUTE2 (Howie *et al.*, 2009).

### 2.2.3 ASSOTSIATSIOONIANALÜÜSID

Viirushepatiidi ja geneetiliste markerite vaheliste assotsiatsioonide tuvastamiseks kasutati programmi SNPTEST versioon 2.5.2 (Marchini ja Howie, 2010). Geneetiliste markerite seose uurimiseks kasutati logistilist regressioonimudelit. Analüüsis lisandusid kovariaatidena sugu ja vanus. SNPTEST käsurea näidis:

```
/snptest-2.5.2/snptest -data *.gen *.sample -method ml -pheno * -frequentist 1 -o *.snptest -cov_names age sex -hwe
```

Ülegenoomselt statistiliselt olulisteks markeriteks loeti neid, mille p-väärtus jääb alla  $5 \times 10^{-8}$  (Pe'er *et al.*, 2008). Analüüsid viidi läbi eraldi kõigi kolme mikrokiibi andmetega.

### 2.2.5 METAANALÜÜS

Metaanalüüsi teostamiseks kasutati tarkvara GWAMA (Mägi ja Morris, 2010). GWAMA käsurea näidis:

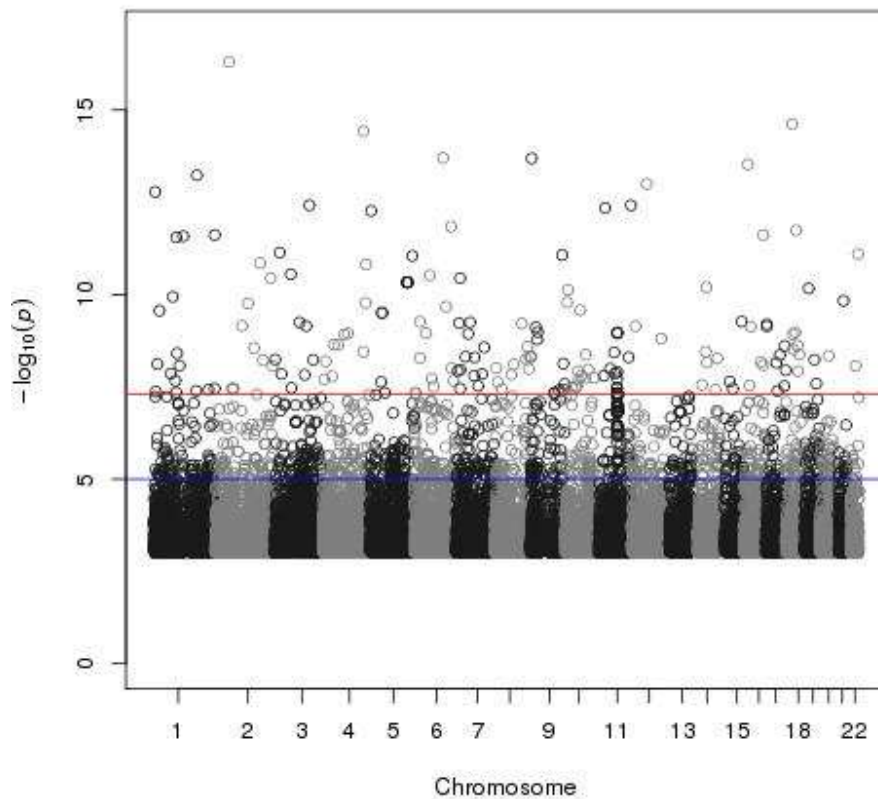
```
/GWAMA --filelist *.gwamain --output *.gwamaout -gc -gco --indel_alleles --name_marker MARKER --map *.map
```

Metaanalüüsiks valiti välja vaid SNP-id, mille kolme kiibi keskmine infoskoor oli kõrgem kui 0,8. Metaanaalüüsi tulemuste põhjal joonistati Manhattan ja Q-Q joonised. Edasistes tabelites tulemuste osas võeti arvesse SNP-id, mis vähemalt kahel kiibil olid samasuunalise assotsiatsiooniga, ning kus ei tekkinud assotsiatsiooni suunas kiipidevahelist konflikti.

## 2.3 TULEMUSED

Käesolevas töös on SNP positsioonide jaoks kasutatud GRCh37 (*Genome Reference Consortium Human Build 37*) koordinaate. Manhattan joonistel on toodud ära kõik metaanalüüsil saadud SNP-id (üksikud kõrged punktid on ilmselt valepositiivsed tulemused, mis on tingitud madalast MAF-ist ja/või info skoorist), samuti võib visuaalset segadust suurendada asjaolu, et iga SNP on kujutatud ringikesena, mis võimendab üksikuid erandeid). Tabelites on toodud igast lookusest väikseima p-väärtusega SNP, kusjuures vähemalt kahel kiibil esines samapidine assotsiatsioon, ning puudusid kiipidevahelised konfliktid. Tabelites paksus kirjas esitatud SNP-id kuulusid selgelt Manhattan joonisel piigi koosseisu. Q-Q joonised on esitatud lisana.

### 2.3.1 B15 (ÄGE A-HEPATIIT)



**Joonis 5 B15 assotsiatsiooniuringu Manhattan joonis.** X-teljel on assotsieerunud markerite genoomsed positsioonid ning Y-teljel nende p-väärtuse negatiivne kümnendlogaritm. Punane horisontaaljoon märgib p-väärtust  $5 \times 10^{-8}$ .

Leiti 9 SNP-i kus vähemalt kahel kiibil esines samasuunaline assotsiatsioon ja p-väärtus jäi alla  $5 \times 10^{-8}$ . Iga lookuse parima p-väärtusega SNP on ära toodud tabelis 3.

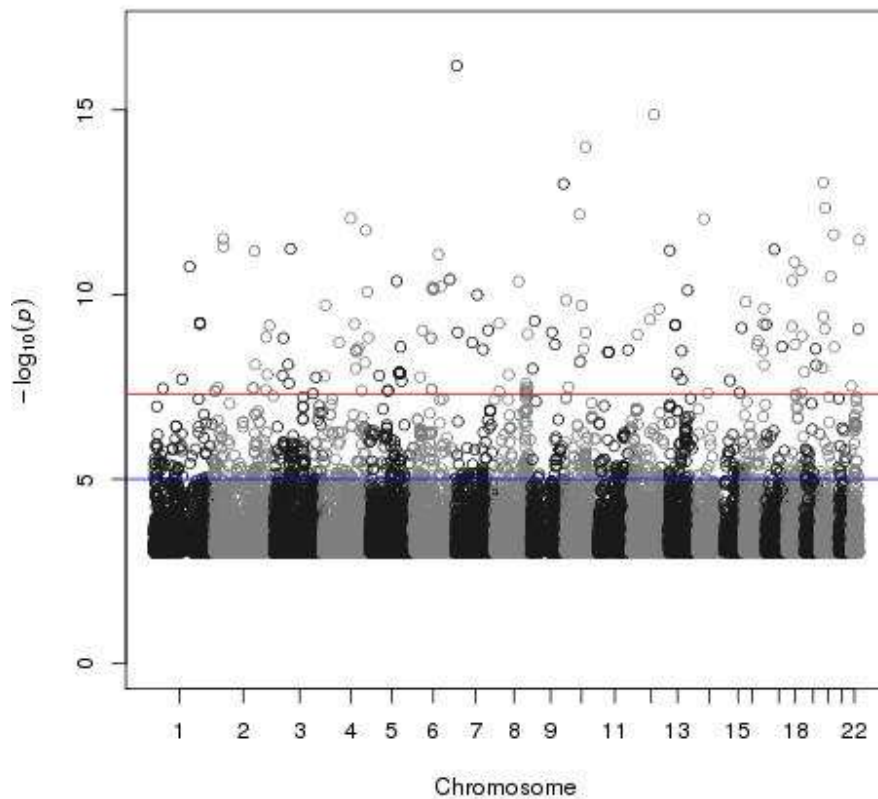
**Tabel 3 B15 assotsieerunud lookuste madalaima p-väärtusega SNP-id.** Tabelis on SNP-i kromosoom, positsioon, alleelid, OR, p-väärtus ja geeni, mille läheduses SNP paikneb.

Kromosoom	Positsioon	Efektalleel/ teine alleel	OR (CI 95%)	P- väärtus	Geen
1	8242709	G / A	0,220957 (0,132584 – 0,368232 )	7,06E-09	-
1	78797751	G / T	0,150925 (0,077581 – 0,293608 )	2,62E-08	<i>RP11-</i>

					<i>183M13.1</i>
<b>4</b>	43469651	G / T	0,068894 (0,028779 – 0,164929 )	1,95E-09	<i>RP11-324J13.2</i>
<b>12</b>	8954057	C / T	0,138511 (0,069742 – 0,27509 )	1,68E-08	<i>A2ML1-AS1</i>
<b>14</b>	39127379	T / C	0,275252 (0,17398 – 0,435474 )	3,64E-08	-
<b>14</b>	52287037	C / T	0,046474 (0,015769 – 0,13697 )	2,69E-08	-
<b>15</b>	101650359	C / T	0,072035 (0,028701 – 0,180796 )	2,16E-08	-
<b>16</b>	26513066	G / C	0,1838 (0,101884 – 0,331576 )	1,88E-08	-

*RP11-183M13.1* ja *RP11-324J13.2* on lncRNA-d (*long non-coding RNA*) ja *A2ML1-AS1* on ncRNA (*non-coding RNA*) (RefSeq). Seega üksi leitud assotsiatsioonidest ei asu valku kodeerival järjestusel ega selle vahetus läheduses.

### 2.3.2 B16 (ÄGE B-HEPATIIT)



**Joonis 6 B16 assotsiatsiooniuringu Manhattan joonis.** X-teljel on assotsieerunud markerite genoomsed positsioonid ning Y-teljel nende p-väärtuse negatiivne kümnendlogaritm. Punane horisontaaljoon märgib p-väärtust  $5 \times 10^{-8}$ .

Leiti 63 SNP-i kus vähemalt kahel kiibil samasuunaline assotsiatsioon ja p-väärtus alla  $5 \times 10^{-8}$ . Iga lookuse parima p-väärtusega SNP on ära toodud tabelis 4.

**Tabel 4 B16 assotsieerunud lookuste madalaima p-väärtusega SNP-id.** Tabelis on SNP-i kromosoom, positsioon, alleelid, OR, p-väärtus ja geeni, mille läheduses SNP paikneb.

Kromosoom	Positsioon	Efektalleel/ teine alleel	OR (CI 95%)	P- väärtus	Geen
1	35311165	G / A	0.038052 (0.012285 - 0.117864)	1.49E-08	-
1	88660298	C / T	0.146351 (0.073962 - 0.289587)	3.49E-08	-
1	166338740	C / A	0.024287 (0.007443 - 0.079253)	7.44E-10	-

1	218404867	T / C	0.041275 (0.014875 - 0.114532)	9.55E-10	-
2	74208753	G / A	0.063223 (0.023953 - 0.166873)	2.52E-08	-
2	74610435	A / G	0.038252 (0.012212 - 0.119822)	2.17E-08	<i>DCTN1</i>
2	74661196	T / C	0.038252 (0.012212 - 0.119822)	2.17E-08	<i>RTKN</i>
2	226718784	G / C	0.030406 (0.009492 - 0.097399)	4.19E-09	-
2	230164811	C / T	0.040891 (0.013482 - 0.124022)	1.67E-08	-
3	106571957	G / A	0.044785 (0.015017 - 0.133563)	2.59E-08	<i>LINC00882</i>
3	197103073	C / T	0.060925 (0.022534 - 0.16472)	3.59E-08	-
4	135302067	T / C	0.022214 (0.005767 - 0.08556)	3.22E-08	-
5	173137502	G / A	0.076492 (0.033447 - 0.174935)	1.16E-09	-
5	179112616	C / T	0.160836 (0.083839 - 0.308548)	3.94E-08	-
6	45700744	C / G	0.048735 (0.018346 - 0.129462)	1.39E-09	-
6	107453271	G / A	0.028782 (0.00934 - 0.088692)	6.64E-10	-
7	2761578	C / T	0.10696 (0.047941 - 0.238637)	4.89E-08	-
7	4811764	C / T	0.038417 (0.012399 - 0.119029)	1.66E-08	-
7	45089503	A / T	0.036219 (0.014294 - 0.091771)	2.75E-12	<i>CCM2</i>
7	45141395	G / T	0.05321 (0.021274 - 0.133085)	3.68E-10	<i>TBRG4</i>
7	45219226	A / G	0.0374 (0.014812 - 0.094433)	3.70E-12	<i>RAMP3</i>
7	51894133	C / T	0.126899 (0.062742 - 0.25666)	9.46E-09	-
7	153769342	T / G	0.031973 (0.01037 - 0.098584)	2.12E-09	<i>DPP6</i>
8	<b>128701496</b>	<b>A / T</b>	<b>0.423239 (0.311563 - 0.574943)</b>	<b>3.86E-08</b>	<b><i>BC042052</i></b>
9	22026737	C / G	0.027341 (0.008397 - 0.089026)	2.35E-09	<i>MTAP,</i> <i>CDKN2B-</i> <i>ASI</i>
13	29020745	T / C	0.04572 (0.015203 - 0.137498)	4.07E-08	<i>FLT1</i>
14	52062925	T / C	0.014519 (0.004146 - 0.05084)	3.74E-11	<i>FRMD6-</i> <i>AS2</i>
14	63667572	G / T	0.042753 (0.014291 - 0.127902)	1.76E-08	-
14	85241324	G / A	0.049109 (0.018148 - 0.132888)	3.04E-09	-
15	46197968	A / G	0.014991 (0.00403 - 0.055757)	3.79E-10	-
15	61701818	G / C	0.037038 (0.012232 - 0.11215)	5.67E-09	-
16	82301202	C / G	0.040344 (0.014886 - 0.109338)	2.86E-10	-
17	8598817	T / G	0.059484 (0.022374 - 0.158151)	1.58E-08	-
17	64865930	G / A	0.041102 (0.01343 - 0.125793)	2.29E-08	<i>CACNG5</i>
19	44916092	G / A	0.034593 (0.011144 - 0.107384)	6.01E-09	-
20	7929448	C / T	0.056748 (0.021568 - 0.149312)	6.31E-09	-
21	19293516	A / G	0.047231 (0.017586 - 0.12685)	1.43E-09	<i>CHODL</i>

*LINC00882*, *BC042052*, *FRMD6-AS2* ja *CDKN2B-ASI* ekspresseerivad RNA-d, kuid ei kodeeri valku (RefSeq). Varasemalt on *CDKN2B-ASI* ncRNA lookust, mis inerakteerub PRC1

(*polycomb repressive complex-1*) ja PRC2-ga vaigistades epigeneetiliselt teisi geene, seostatud südame- ja veresoonkonna haiguste, vähi, diabeedi ja veel paljude muude haigustega (RefSeq).

*DCTN1* (*Dynactin Subunit 1*) täidab rakus mitmeid funktsioone osaledes endoplasmaatilise retiikulumi ja Golgi kompleksi vahelises transpordisüsteemis, lüsoosoomide, endosoomide ja kromosoomide liikumises (RefSeq). Sünteesitava valgul on täheldatud ka makrofaagides interaktsioone HIV-1 viirusevalgu *Vpr*-iga (Dumas *et al.*, 2015))

*TBRG4* (*Transforming Growth Factor Beta Regulator 4*) Täpne roll pole küll teada, kuid geeni seostatakse rakutsükli reguleerimisega (RefSeq). Samuti on täheldatud inimese rakuliinides (HEK293 ja Jurkat), et valk interakteerub HIV-1 *Vpu* valguga (Jäger *et al.*, 2011).

*FLT1* (*fms related tyrosine kinase 1*) ekspresseerib retseptor türosiinkinaasi, mis on oluline angio- ja vaskulogeneesil (RefSeq). Lisaks interakteerub mitmete HIV-1 valkudega: gp120, Nef ja tat.

HIV-1 viiruse valkudega interakteeruvate geenide *DCTN1*, *TBRG4* ja *FLT1* assotsiatsioon võib viidata mõningale bioloogilise rollile infektsioonhaigustega toimetulekul.

*RTKN* (*Rhotekin*) osaleb Rho signaaliradades, mille regulatsiooniprobleeme on seostatud vähiga. Rakus täidab funktsioone tsütokineesil, transkriptsioonil, raku kasvamisel ja silelihaste kontraktsioonil (RefSeq).

*MTAP* (*methylthioadenosine phosphorylase*) on olulise tähtsusega polüamiinide metabolismis, adeniini ja metioniini kasutamises. Ensüüm puudub tihti vähirakkudes, sest geen deleteerub koos tuumorsupressor p16-nega (RefSeq).

Kuna viirushepatiit võib põhjustada HCC-d, siis on ka võimalik, et vähiga seonduvate geenide *RTKN* ja *MTAP* assotsiatsioonidel olla bioloogiline põhjus.

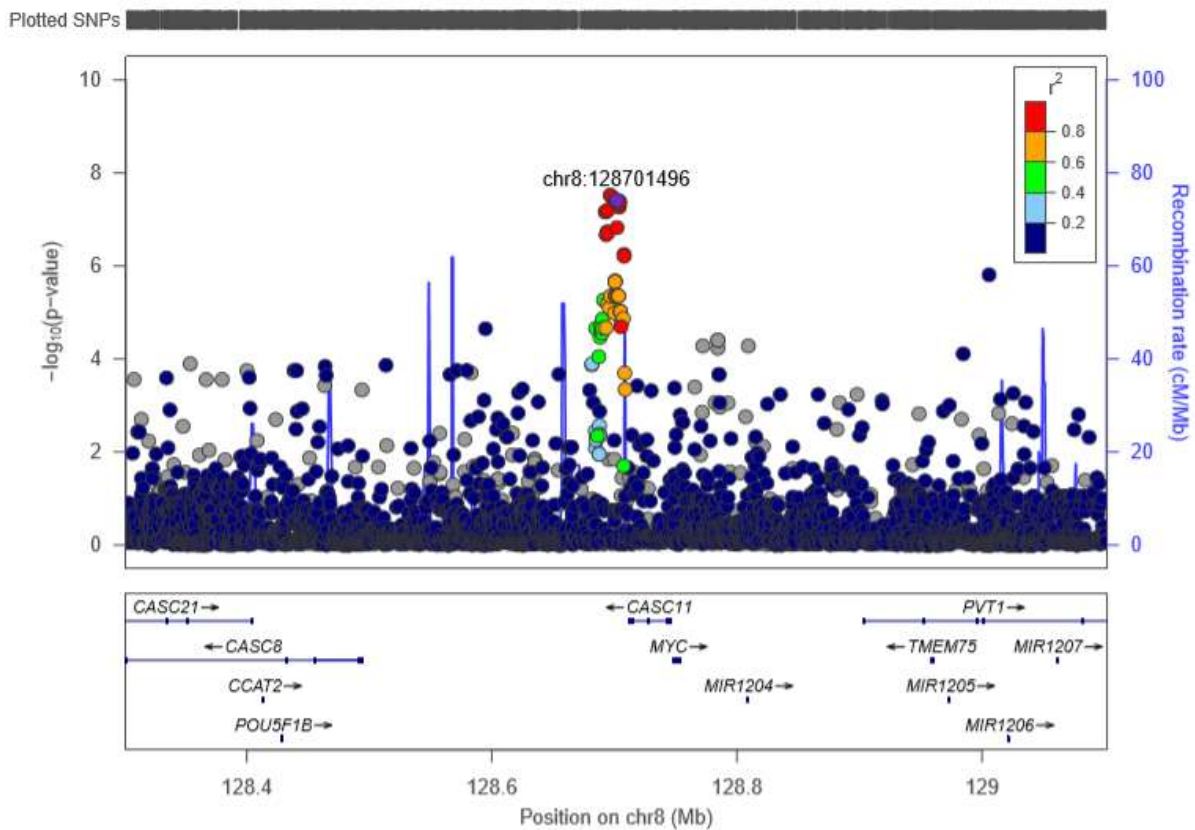
*CCM2* (*Scaffolding Protein*) on vajalik normaalseks raku tsütoskeleti toimimiseks, rakk-rakk interaktsioonides ja endoteelis luumeni tekkeks (RefSeq).

*RAMP3* (*Receptor Activity Modifying Protein 3*) annab transmembraansele valgule CRCL (*calcitonin-receptor-like receptor*) adrenomedulliini retseptori funktsionaalsuse (RefSeq).

*DPP6* (*Dipeptidyl Peptidase Like 6*) ei oma vaatamata oma nimele proteaasest aktiivsust. Ekspresseeritav valk kinnitub seevastu *KCDN2* kaaliumi kanalitele reguleerides nende ekspressiooni ja biofüüsikalisi omadusi (RefSeq).

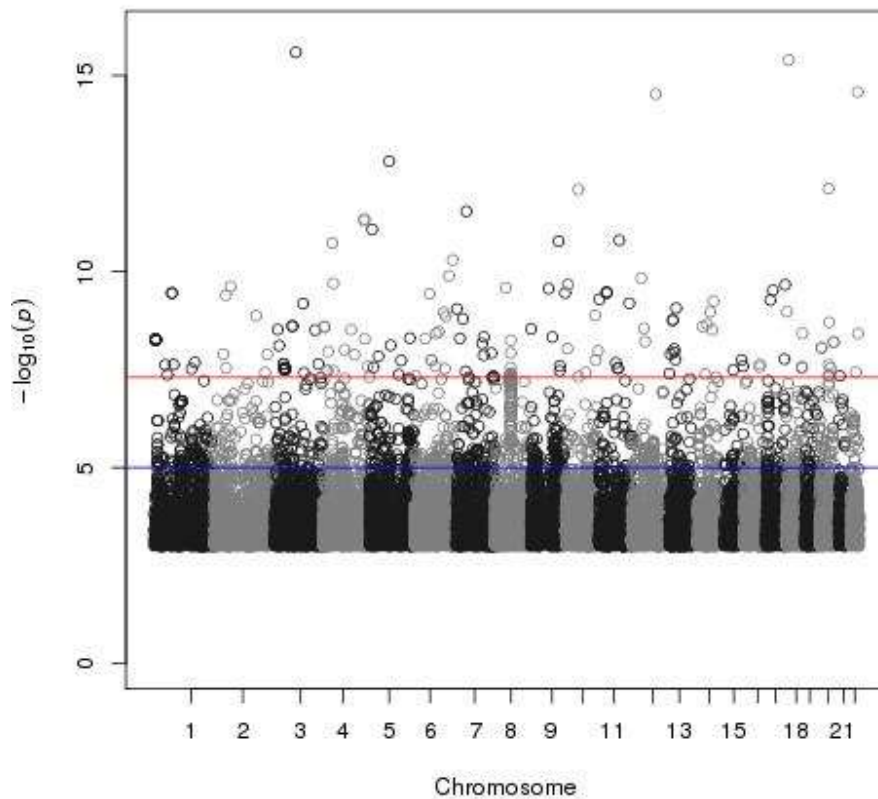
*CACNG5* (*calcium voltage-gated channel auxiliary subunit gamma 5*) reguleerib AMPA retseptoreid. Geeni on seostatud skisofreenia ja bipolaarse häirega (RefSeq).

*CHODL* (*chondrolectin*) kodeerib süsivesikute äratundmisega tegelevat C-tüüpi lektiini domeeniga membraanivalku. Eelnimetatud domeen on seotud glükoproteiinide ja patogeenide endotsütoosiga (RefSeq).



**Joonis 7 Markeri chr8:128701496 regionaaljoonis.** Marker chr8:128701496 on märgitud lilla rombina, ning teised markerid ringikestena. Erineva värvusega on esitatud markerite LD uuritava markeriga. X-teljel on markerite füüsiline asukoht 8. kromosoomis. Y-teljel markerite p-väärtuse negatiivne kümnendlogaritm. Sinine joon märgin rekombinatsioonimäära. Joonise allosas on näidatud regioonis asuvate geenide asukohad.

### 2.3.3 B18 (KROONILINE B- VÕI C-HEPATIIT)



**Joonis 8 B18 assotsiatsiooniuringu Manhattan joonis.** X-teljel on assotsieerunud markerite genoomsed positsioonid ning Y-teljel nende p-väärtuse negatiivne kümnendlogaritm. Punane horisontaaljoon märgib p-väärtust  $5 \times 10^{-8}$ .

Leiti 95 SNP-i kus vähemalt kahel kiibil samasuunaline assotsiatsioon ja p-väärtus alla  $5 \times 10^{-8}$ . Iga lookuse parima p-väärtusega SNP on ära toodud tabelis 5.

**Tabel 5 B18 assotsieerunud lookuste madalaima p-väärtusega SNP-id.** Tabelis on SNP-i kromosoom, positsioon, alleelid, OR, p-väärtus ja geeni, mille läheduses SNP paikneb.

Kromosoom	Positsioon	Efektalleel/ teine alleel	OR (CI 95%)	P- väärtus	Geen
2	178799463	G / A	0.07554 (0.03121 - 0.182837)	1.04E-08	<i>PDE11A</i>
3	136049538	G / C	0.056047 (0.021182 - 0.148299)	6.63E-09	<i>PCCB</i>

3	136089591	C / G	0.056047 (0.021182 - 0.148299)	6.63E-09	<i>STAG1</i>
3	136530031	T / A	0.057303 (0.021697 - 0.15134)	8.10E-09	-
4	88613048	G / A	0.220401 (0.128209 - 0.378886)	4.58E-08	-
5	64433779	A / C	0.159747 (0.083482 - 0.305685)	3.11E-08	-
8	5648673	G / A	0.033995 (0.010519 - 0.10987)	1.65E-08	-
8	<b>63884091</b>	<b>T / C</b>	<b>0.338165 (0.232423 - 0.492015)</b>	<b>1.49E-08</b>	<b><i>NKAIN3</i></b>
8	63945258	A / G	0.350759 (0.24105 - 0.510398)	4.49E-08	<i>GGH</i>
10	14845290	C / A	0.140252 (0.071192 - 0.276301)	1.40E-08	-
10	121161291	G / A	0.042324 (0.014065 - 0.127361)	1.89E-08	<i>GRK5</i>
11	5015163	A / C	0.05181 (0.019628 - 0.136759)	2.33E-09	-
11	55667728	C / A	0.065702 (0.026984 - 0.159975)	2.07E-09	-
11	59333908	C / T	0.033775 (0.010994 - 0.10376)	3.38E-09	<i>TRNA_Phe</i>
13	21716984	G / A	0.051221 (0.018896 - 0.138842)	5.34E-09	<i>SAP18</i>
14	33218472	G / A	0.039697 (0.014221 - 0.110818)	7.49E-10	<i>AKAP6</i>
17	71503828	A / G	0.098678 (0.042999 - 0.226456)	4.76E-08	<i>SDK2</i>
20	37667148	T / C	0.177071 (0.096075 - 0.326351)	2.93E-08	<i>DHX35</i>
22	36901613	C / T	0.048874 (0.018378 - 0.129974)	1.50E-09	<i>FOXRED2</i>

*TRNA\_Phe* on fenüülalaniini transpordi-RNA ja seega ei kodeeri valku (RefSeq).

Ka B18 grupis leiti geene, mis interakteeruvad HIV-1 valkudega. *PDE11A* (*phosphodiesterase 11A*) osaleb cAMP ja cGMP signalisatsiooniradade regulatsioonis (RefSeq) ja interakteerub HIV-1 Tat valguga (Zauli *et al.*, 2001). *SAP18* (*Sin3A associated proteiin 18*) osaleb histoonide deatsetüleerimises ja interakteerub ka HIV-1 valkude: Pr55(gag), Tat ja gag-pol-iga.

*PCCB* (*propionyl-CoA carboxylase subunit beta*) osaleb propionüül-CoA katabolismis (RefSeq).

*STAG1* (*stromal antigene 1*) eksprepeerib rakutuumas leiduvat valku, mis osaleb mitoosiprotsessis tütarchromatiidide kohesioonis (RefSeq).

*GGH* (*gamma-glutamyl hydrolase*) kodeerib ensüümi, mis katalüüsib gamma-glutamaatide hüdrolüüsi (RefSeq).

*AKAP6* (*A-kinase anchoring protein 6*) ekspresseeritakse kõrgelt ajus, südame- ja vöötlihaskoes. Rakus on valk lokeeritud sarkoplasmaatilisse retiikulumi ja tuumamembraani, kus toimib PKA (*protein kinase-A*) ankruna (RefSeq).

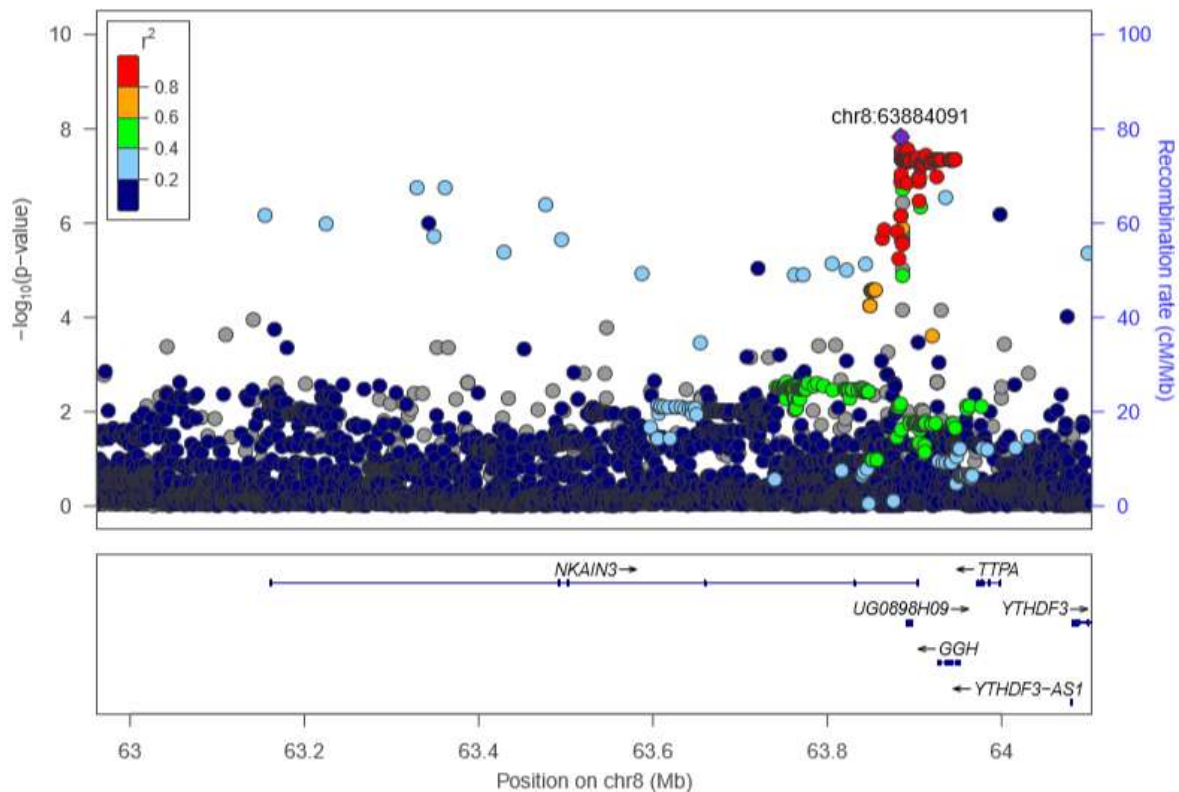
*DHX35* (*DEAH-box helicase 35*) geeni funktsioon pole teada, kuid sarnased valgud osalevad embrü- ja spermatogeneesil ning rakkude kasvamisel ja jagunemisel (RefSeq).

*FOXRED2* (*FAD dependent oxidoreductase domain containing 2*) geen, mis ilmselt osaleb endoplasmaatilise retiikulumiga assotsieerival degradatsioonil (Riemer *et al.*, 2009).

*SDK2* (*Sidekick cell adhesion molecule 2*) geeni funktsioon pole teada, kuid sisaldab kaht immunoglobuliini domeeni ja kolmeteist fibronektiini domeeni, mis sisaldavad kohti DNA, hepariini või rakupinnaga seondumiseks (RefSeq).

*GRK5* (*G proteiin-coupled receptor kinase 5*) on seotud G-valguga seonduvate retseptorite deaktiivatsiooniga ja polümorfonukleaarsete leukotsüütide liikuvuse reguleerimisega (RefSeq).

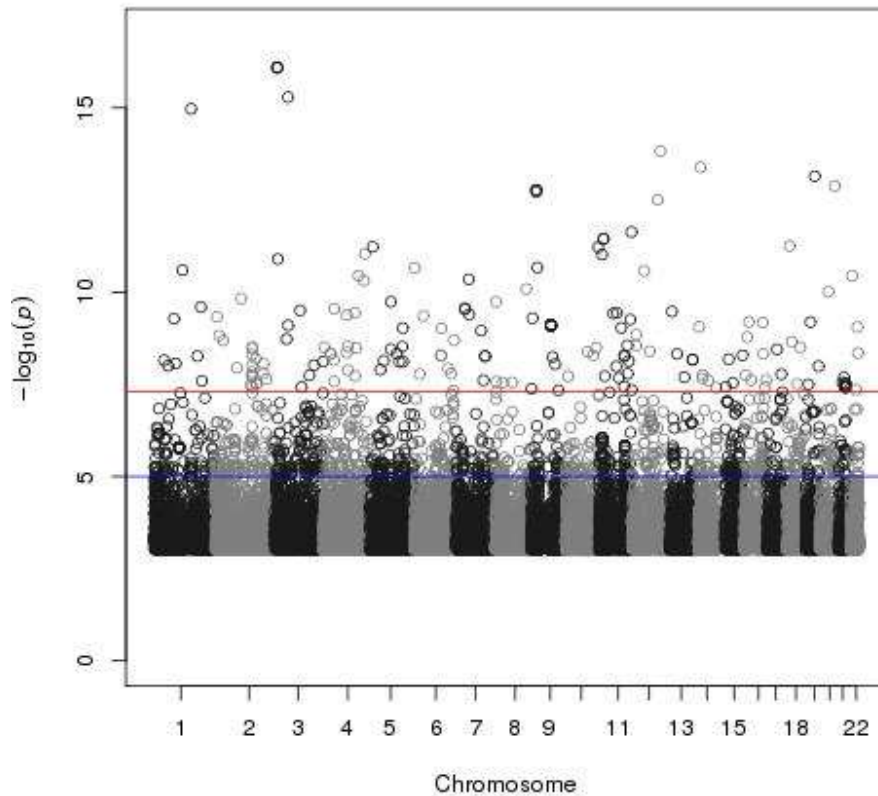
*NKAIN3* (*sodium/potassium transport ATPase interacting 3*) reguleerib naatriumi ja kaaliumi transpordi (RefSeq). Kõikidest lookustest on kõige selgema piigi ja rohkemate usaldusnivoo ületanud SNP-idega lookus just selles geenis (joonis 8).



**Joonis 9 Markeri chr8:63884091 regionaaljoonis.** Marker chr8:63884091 on märgitud lilla rombina, ning teised markerid ringikestena. Erineva värvusega on esitatud markerite LD uuritava markeriga. X-teljel on markerite füüsiline asukoht 8. kromosoomis. Y-teljel markerite p-väärtuse

negatiivne kümnendlogaritm. Sinine joon märgib rekombinatsioonimäära. Joonise allosas on näidatud regioonis asuvate geenide asukohad.

### 2.3.4 B19 (TÄPSUSTAMATA HEPATIIT)



**Joonis 10 B19 assotsiatsiooniuringu Manhattan joonis.** X-teljel on assotsieerunud markerite genoomsed positsioonid ning Y-teljel nende p-väärtuse negatiivne kümnendlogaritm. Punane horisontaaljoon märgib p-väärtust  $5 \times 10^{-8}$ .

Leiti 49 SNP-i kus vähemalt kahel kiibil samasuunaline assotsiatsioon ja p-väärtus alla  $5E-08$ . Iga lookuse parima p-väärtusega SNP on ära toodud tabelis 5.

**Tabel 5 B19** assotsieerunud lookuste madalaima p-väärtusega SNP-id.

Tabelis on SNP-i kromosoom, positsioon, alleelid, OR, p-väärtus ja geeni, mille läheduses SNP paikneb.

Kromosoom	Positsioon	Efektalleel/ teine alleel	OR (CI 95%)	P- väärtus	Geen
1	19611683	C / G	0.137376 (0.068884 - 0.273972)	1.78E-08	<i>AKR7A3</i>
1	19656164	T / A	0.146893 (0.073807 - 0.29235)	4.81E-08	<i>PQLC2</i>
2	151121387	C / A	0.188329 (0.109236 - 0.324689)	1.93E-09	-
2	207882262	G / A	0.104547 (0.049087 - 0.222667)	4.93E-09	-
3	114677355	T / C	0.018247 (0.004362 - 0.076328)	4.26E-08	<i>ZBTB20</i>
4	75351138	A / C	0.057123 (0.022159 - 0.147256)	3.21E-09	<i>AREG</i>
5	2399728	T / C	0.133157 (0.064539 - 0.27473)	4.97E-08	-
5	16489713	T / C	0.018247 (0.004538 - 0.073373)	1.75E-08	<i>FAM134B</i>
6	159651324	G / A	0.183408 (0.100687 - 0.334089)	3.04E-08	<i>FNDC1</i>
11	116652298	T / G	0.027612 (0.008542 - 0.089255)	2.07E-09	<i>ZNF259</i>
11	116775308	G / A	0.020609 (0.006185 - 0.068668)	2.67E-10	<i>SIK3</i>
12	9559433	G / A	0.036505 (0.01173 - 0.113607)	1.13E-08	<i>RP11-599J14.2</i>
13	36492283	T / C	0.110832 (0.053503 - 0.229588)	3.31E-09	<i>DCLK1</i>
13	49851023	A / G	0.15761 (0.083485 - 0.297549)	1.24E-08	<i>CDADCI</i>
13	49979807	G / A	0.153415 (0.081219 - 0.289784)	7.80E-09	<i>CAB39L</i>
14	95859403	C / T	0.058917 (0.021809 - 0.159166)	2.40E-08	-
15	34319493	A / G	0.032235 (0.009664 - 0.107523)	2.35E-08	<i>AVEN, CHRM5</i>
15	101817887	G / C	0.074976 (0.029728 - 0.189096)	4.14E-08	-
16	34297847	A / T	0.078643 (0.032288 - 0.191546)	2.21E-08	<i>CCNYL3</i>
16	34608614	G / T	0.080201 (0.032969 - 0.195099)	2.71E-08	<i>RP11-488I20.3</i>
16	34983057	G / A	0.081843 (0.033677 - 0.198899)	3.38E-08	<i>FLJ26245</i>
16	55210062	G / A	0.099702 (0.048625 - 0.204433)	3.21E-10	-
18	40287433	C / T	0.039864 (0.014854 - 0.106984)	1.63E-10	-
19	41788173	C / T	0.055361 (0.019667 - 0.15584)	4.34E-08	<i>HNRNPUL1</i>
19	51988054	T / A	0.038112 (0.012295 - 0.118144)	1.55E-08	<i>CEACAM18</i>

*RP11-488I20.3*, *RP11-599J14.2* ja *FLJ26245* ei kordeeri valku.

HIV-1 valkudega interakteervuaid geene oli 2. *AREG* (*amphiregulin*), mille poolt ekspresseeritav valk on mitogeen astrotsüütidele, Scwanni rakkudele ja fibroplastidele. Geen assotsieerub psoriaasi-laadse fenotüübi, vähi ja põletikuliste seisunditega. Samuti interakteerub HIV-1 valguga Vif. (Kim *et al.*, 2013) Teine geen, mis interakteerub HIV-1 oli *HNRNPUL1*

(*heterogeneous nuclear ribonucleoprotein U like 1*), mis kodeerib valku, mis kinnitub adenoviiruse onkovalgule. Interakteerub HIV-1 gag-pol valguga. (Wagner *et al.*, 2014)

Leidusid ka mitmed närvisüsteemiga seonduvad geenid: *ZBTB20* (*Zinc finger and BTB domain containing 20*) toimib transkriptsiooni repressorina ja on oluline neurogenesil ja glükoosi homöostaasil. *FAM134B* (*reticulophagy regulator 1*) kodeerib transmembraanset cis-Golgi valku, mis on võib olla vajalik notsitseptiivsete ja autonoomsete ganglionineuronite pikaajaliseks elumuseks. *ZNF259* (*ZPR1 zinc finger*) on seotud neuronite diferentseerumisega. *DCLK1* (*doublecortin like kinase 1*) kodeerib mikrotuubulite polümeerisatsiooni reguleerivat valku, mis osaleb mitmetes rakulistes protsessides (neuronite migratsioonis, apoptoosis ja neurogenesis). Geen assotsieerub ka mälu ja koginitiivsete võimetega. *CHRM5* (*cholinergic receptor muscarinic 5*) kodeerib muskariinireseptorit, mis mõjutab atsetüülkoliini kesk- ja piirdenärvisüsteemis. Retseptori stimulatsioon tõstab cAMP taset. Neuroloogilisi sümptomeid on täheldatud C-hepatiidi korral (Forton *et al.*, 2006), kuigi selgemat bioloogilist mehhanismi, mis neid gene fenotüübiga seoks on keeruline välja tuua.

*PQLC2* (*PQ loop repeat containing 2*) on seotud aminohapete (arginiin, histidiin, lüsiin) transpordiga (UniProt).

*AKR7A3* (*aldo-keto reductase family 7 member A3*) osaleb aldehüüdide ja ketoonide detoksifikatsioonis (RefSeq).

*FNDC1* (*fibronectin type III domain containing 1*) Võib olla G-valgu signaaliraja aktivaator (UniProt).

*SIK3* (*SIK family kinase 3*) on seotud valkude fosforüleerimisega (UniProt).

*CDADC1* (*cytidine and dCMP deaminase domain containing 1*) võib osaleda spermatogeneesis ja munandite arengus (Liu *et al.*, 2006).

Apoptoosi ja vähiga on veel seotud *CAB39L*, (*calcium binding proteiin 39 like*) mis aktiveerib tuumorsupressorit *LKB1* (*liver kinase B 1*) (UniProt) ja *AVEN* (*apoptosis and caspase activation inhibiitor*), mis reguleerib apoptoosiprotsesse (UniProt).

*CCNYL3* (*cyclin Y-like 3*) reguleerib seriin/threoniin kinaasi aktiivsust (UniProt).

*CEACAM18* (carcinoembryonic antigen related cell adhesion molecule 18) geeni paraloog *CEACAM1* on seotud raku adhesiooniga ja mõjutab paljusid olulisi bioloogilisi protsesse (kudede mitmemõõtmeline struktuur, angiogenees, apoptoos, vähi teke ja immuunsüsteemi) (RefSeq).

## 2.4 ARUTELU

GWAS standardiks on uuringutulemuste kinnitamiseks kordusuuringu teostamine sõltumatu valimiga. Käesolevas töös pole seda tehtud ja seega tulemuste kinnitamiseks tuleks sooritada sõltumatu valimiga kordusuuring. Tulenevalt kasutatava mudeli lihtsusest esineb töös kasutatavatel Manhattan joonistel artefakte ja lisas 2 toodud Q-Q joonistel mõningast inflatsiooni, mida võeti metaanalüüsil arvesse.

Varasem uuringud on leidnud assotsiatsioone *SDK2* geenis. Näiteks HIV ja HCV koinfektsiooni korral on leitud seos positsioonis 17:73383923 maksafibroosiga (Ulveling *et al.*, 2016) ja B-hepatiiti haigestumisega positsioonis 17:73706567 (Tian *et al.*, 2017). Töös tuvastati B18 grupis SNP *SDK2* geeni positsioonis 17:71503828, mis osutus ülegenoomselt statistiliselt oluliseks. Kahjuks pole geeni täpne funktsioon teada, kuid kodeeritav valk omab sarnaseid domeene immunsüsteemis osalevate valkudega, seega on võimalik, et *SDK2* geenil on bioloogiline roll viirushepatiidi infektsiooni takistamisel.

Käesolevas töös tuvastati mitmeid SNP-e *NKAIN3* geenis, mis assotseerusid tugevalt B18 grupiga ehk kroonilise B- ja C-hepatiidiga. Varasemad GWAS-id on leidnud samas geenis variante, mis assotseeruvad buliimiaga (Wade *et al.*, 2013), Alzheimeri tõvega (Herold *et al.*, 2016) ja allergilise nohuga (Bunyavanich *et al.*, 2014). Viirushepatiidiga antud lookust varasemalt seostatud pole.

Kuigi erinevalt *NKAIN3* geeni lookusest ei tuvastati teisi samavõrd tähendusrikkaid Manhattan joonise piike, esines siiski huvitavates geenides paiknevate SNP-idega assotsiatsioone. Näiteks esinesid tulemustes mitmed HIV-1 viirusevalkudega interakteeruvad geenid: *DCTN1*, *TBRG4* ja *FLT1*, *PDE11A*, *SAP18*, *AREG*, *HNRNPUL1*. HIV-1 seoseliste geenide tuvastamine võib tuleneda nii sellest, et tegemist on infektsioonhaigusega ja seega mõjutavad neid sarnased mehhanismid, kui ka sellest, et HIV ja viirushepatiit võivad esineda koinfektsioonina, sest riskigrupid on Eestis osaliselt kattuvad. Teine oluliselt esindatud rühm genee olid seotud vähiga: *CDKN2B-AS1*, *RTKN*, *MTAP*, *AREG*, *HNRNPUL1*, *CAB39L*. Seos vähiga võib ilmneda ka kroonilise hepatiidi tagajärjel tekkivast HCC-st. Üllatuslikult tõusid B19 grupis esile ka mitmed närvisüsteemis olulist rolli mängivad geenid: *ZBTB20*, *FAM134B*, *ZNF259*, *DCLK1*, *CHRM5*.

Töös esines ka arvukalt leide, mis ei paiknenud ühegi geeni läheduses. Vaatamata bioloogilise tõlgenduse puudumisele, miks antud geenivariant võiks viirushepatiitidid assotsieeruda saaks analoogseid markereid siiski kasutada haigusrisiki hindamiseks (Läll *et al.*, 2017). Viirushepatiiti haigestumise riski hindamisel on muidugi lisaks geneetikale veel oluline ka keskkonnast tulenev risk (viiruse levimus populatsioonis).

Vaatamata suurele hulgale statistilise olulisusega leidudest markeerivad vaid üksikud töös olevad SNP-id (chr8:128701496, chr8:63884091) selgeid lookuseid, millel on assotsiatsioon viirushepatiidiga.

#### 2.4.1 TULEVIKUPLAANID

Eesti Geenivaramu kogub üle Eesti 100 000 uut geenidoonorit, kes genotüüpiseeritakse Illumina GSA mikrokiipidel, mis määravad suurema hulga SNP-e. Seega TÜGI biopanga andmed kasvavad, sisaldades suuremat hulka isikuid kui ka rohkem SNP-e isiku kohta. Eelnimetatud asjaolud tõstatavad vajaduse kordusuuringu sooritamiseks. Võib eeldada, et parem uuringu võimsus annab ka paremaid tulemusi. Käesolev töö teeb mõned ettepanekud, kuidas parandada järgmist viirushepatiidi GWAS-i.

Edaspidistes töödes võiks kontrollgrupi valikul olla konservatiivsem (eelistades näiteks 1:4 juhtude ja kontrollide suhte), ning välistada kõik lisas 1 toodud viirushepatiidiga seonduvad RHK-10 koodid, ning samuti teised maksahaigustega seotud RHK-10 koodid K70-K77. Nii saaks olla kindel, et geneetilised variandid on assotsieerunud just viirusliku hepatiidi, mitte üldiste maksahaigustega. Kontrollgrupp võiks ka pisut paremini vanuselisele jaotusele vastata kui käesolevas töös (joonis 4). Parema tulemuse saavutamiseks tuleks eemaldada ka kõrge sugulusastmega indiviidid, ning lisada peakomponente.

Täpsustaval uuringul võiks fenotüüpe jaotada tabel 6 põhjal (lähtudes haiguse kulust ja kaasates suurema arvu RHK-10 koode). Sel juhul saaksime sooritada GWAS-i võttes arvesse viirushepatiidi haiguse kulgu. Samuti oleks põnev uurida ka konkreetse viiruse poolt põhjustatud viirushepatiiti, kuid sellisel juhul tuleb saada andmeid täpsustatud andmeid. Kahjuks vaid RHK-10 koode kasutades pole võimalik viirust täpselt määrata.

**Tabel 6 Edasisteks uuringuteks uuringurühmade moodustamine.**

x tähistab kõiki võimalikke nii algavaid RHK-10 koode.

Rühma kirjeldus	RHK-10 koodid
<b>Viirushepatiidi kandjad</b>	Z22.5
<b>Viirushepatiidi jääknähtudega</b>	B94.2
<b>Viirushepatiidi kroonilised vormid</b>	B18.x
<b>Viirushepatiidid ägedad vormid</b>	B15.x, B16.x, B17.x, B19.x
<b>Viirushepatiidi ägedad vormid koos maksakoomaga</b>	B15.0, B16.0, B16.2, B19.0
<b>Viirushepatiidid ägedad vormid ilma maksakoomata</b>	B15.9, B16.1, B16.3, B16.9, B19.9
<b>Viirushepatiidi vormid, mis pole kulu järgi defineeritud</b>	B25.1, B00.8, O98.4, P35.3
<b>Viirushepatiidid ägedad ja kulu järgi defineerimata vormid</b>	B15.x, B16.x, B17.x, B19.x, B25.1, B00.8, O98.4, P35.3

Lisaks tabelis 6 toodud gruppidele võiks ühe grupina lisada ka kõik Lisa 1 RHK-10 klassifikatsioonis nimetatud koodidega grupi, mis sisaldaks endast kõiksuguste viiruslike hepatiidi diagnoosidega indiviide.

## KOKKUVÕTE

Viirusliku hepatiiti põhjustavad mitmed viirused (HAV, HBV, HCV, HDV, HEV) ja haigus on levinud praktiliselt kõikjal. Haiguse kulg ja tõsidus sõltuvad nii haigustekitajast, patsiendist kui ka keskkonnast. Lähtuvalt haiguse kulust jaotatakse viirushepatiit ägedaks ja krooniliseks.

Käesolevas töös viidi läbi nelja uuringugrupiga (B15: äge A-hepatiit, B16: äge B-hepatiit, B18: krooniline B- või C-hepatiit ja B19: täpsustamata hepatiit) logistilise regressiooni mudeliga GWAS. Uuringus kasutati kolme genotüpiseerimiskiibi andmeid, mis pärinesid TÜGI-ist ja fenotüüpide määramisel kasutati mitmete Eestis olevate haigusjuhtude registrite koondandmeid. Seejärel sooritati metaanalüüs.

Statistiliselt oluliseks ( $p < 5 \times 10^{-8}$ ) indentifitseeriti 9 SNP-i B15 grupis, 63 SNP-i B16 grupis, 95 SNP-i B18 grupis ja 49 SNP-i B19 grupis. Vaid üks lookus asus varasemalt viirushepatiidiga assotsieerunud geenis *SDK2*. Mitmed leiud paiknesid HIV-1 viiruse valkudega interakteeruvates geenides või nende vahetus läheduses ja varasemalt vähiga seostatud geenides. Töös tuvastati ka tugevalt assotsieerunud lookus *NKAIN3* geenis, mida pole varasemalt sarnaste haigustega seostatud. Lisaks leidis ka geeniväliseid ja geenidest kaugeid SNP-e, mis assotsieerusid fenotüübiga. Vaatamata suurele hulgale SNP-ide moodustasid vaid mõned (chr8:128701496, chr8:63884091) töös kirjeldatutest selged lookused, mis viitavad assotsiatsioonile viirushepatiidiga.

Töös leitud lookuste assotsiatsiooni kinnitamiseks tuleks sõltumatus kohordis läbi viia täiendavad uuringud.

# GENOME-WIDE ASSOCIATION STUDY OF VIRAL HEPATITIS

Siim Kurvits

## Summary

Viral hepatitis is caused by different viruses, primarily HAV, HBV, HCV, HDV and HEV. The disease is very prevalent. Based on WHO data in year 2015 there were globally more than 325 million chronically infected. The disease can be divided into acute and more damaging chronic condition. Chronical hepatitis can lead to cirrhosis and hepatocellular carcinoma.

Previous GWAS have found many loci linked to viral hepatitis. The main findings have been in immunological genes e.g. HLA, CD40, but various regions have been associated with the disease.

The aim of this study was to identify genetic variants that can be associated with viral hepatitis in Estonian population. In the current study 4 case groups were defined based on the ICD-10 codes. Group B15 (acute hepatitis A), group B16 (acute hepatitis B), group B18 (chronic hepatitis B or C) and group B19 (unspecified viral hepatitis). Four GWA studies with logistic regression model were conducted and meta-analyzed.

More than 200 SNPs with whole genome statistical significance ( $p < 5 \times 10^{-8}$ ) were found (9 in B15, 63 in B16, 95 in B18 and 49 in B19). In previous studies SDK2 gene locus has been associated with HIV and HBV co-infection. SNPs from the same gene were found statistically significant in the current study as well. No more previously known associations were replicated. However, a novel discovery of SNPs in the NKAIN3 gene were made with high statistical association. Statistically relevant SNPs were located in genes interacting with HIV-1 proteins. Some SNPs in cancer associated genes were found as well. Although many statistically significant SNPs were found only few of them (chr8:128701496, chr8:63884091) were correctly positioned to indicate association with viral hepatitis.

Further research and replicative studies are needed to confirm the associations described in this study.

# TÄNUSÕNAD

Tahan tänada suurepärast juhendajat, Toomas Hallerit ja Eesti Geenivaramu geenidoonoreid, ilma kelleta poleks selle töö tegemine võimalik olnud. Tänan ka Anu Reigot, kes oli RHK-10 koodide valimisel abiks.

Suur aitäh perekond ja sõbrad, kes te toetasite mind lõputöö valmimisel.

## KASUTATUD KIRJANDUS

- 1000 Genome Project Consortium (2015) A global reference for human genetic variation. *Nature*. 526(7571):68-74.
- Brookes, A.J., (1999) The essence of SNPs. *Gene*, 234(2), pp.177–186
- Bunyavanich S., Schadt E.E., Himes B.E., ... Weiss S.T. (2014) Integrated genome-wide association, coexpression network, and expression single nucleotide polymorphism analysis identifies novel pathway in allergic rhinitis. *BMC Med Genomics* 7:48
- Bush, W.S. ja Moore, J.H., 2012. Chapter 11: Genome-wide association studies. *PLoS computational biology*, 8(12), p.e1002822.
- Chakravarti M.R., Vogel F. (1973) A twin study on leprosy. In: Becker PE, Lenz W, Vogel F, Wendt GG (eds). *Topics in Human Genetics*. Stuttgart: Georg Thieme. 1–123.:399
- Chemin I., Zoulim F. (2009) Hepatitis B virus induced hepatocellular carcinoma, *Cancer Lett.* vol. 286 (pg. 52-59)
- Dean M., Carrington M., Winkler C., ..., O'Brien S.J. (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CCR5* structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science*. 273(5283):1856-62
- Dumas A., Le-Bury G., Marie-Anais F., ... Niedergang F. (2015) The HIV-1 protein Vpr impairs phagosome maturation by controlling microtubule-dependent trafficking. *J. Cell. Biol.* 211(2):359-72
- Fatkenheuer G., Pozniak A.L., Johnson M.A. (2005) Efficacy of short-term monotherapy with maraviroc, a new *CCR5* antagonist, in patients infected with HIV-1. *Nat. Med.* 11:1170–2
- Forton D.M., Taylor-Robinson S.D., Thomas H.C. (2006) Central nervous system changes in hepatitis C virus infection. *Eur. J. Gastroenterol Hepatol.* 18(4):333-8.
- 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* 31 (New York, N.Y.), 296(5576), 2225–9.
- GBD 2013 Mortality and Causes of Death Collaborators (2015). Global, regional and national age specific all-cause and cause-specific mortality for 240 causes of death 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 385:117–171
- Ge D., Fellay J., Thompson A.J., ..., Goldstein D.B. (2009) Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature*. 461(7262)
- Herold C., Hooli B.V, Mullin K., Liu T., Roehr J.T., Mattheisen M., Parrado A.R., Bertram L., Lange C., Tanzi R. E. (2016) Family-based association analyses of imputed genotypes reveal genome-wide significant association of Alzheimer's disease with *OSBPL6*, *PTPRG*, and *PDCL3*. *Mol Psychiatry*. 21(11):1608-1612.

- 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
- Hu Z., Liu Y., Zhai X., ... Shen H., (2013) New loci associated with chronic hepatitis B virus infection in Han Chinese. *Nature Genetics* 45, 1499–1503
- J. Marchini, B. Howie, S. Myers, G. McVean, P. Donnelly (2007) A new multipoint method for genome-wide association studies via imputation of genotypes. *Nature Genetics* 39: 906-913
- Jäger S., Cimermancic P., Gulbahce N., ... Kroan N. J. (2011) Global landscape of HIV-human protein complexes. *Nature*. 481(7381):365-70.
- Jepson A.P., Banya W.A.S., Sisay-Joof F.,(1995). Genetic regulation of fever in *Plasmodium falciparum* malaria in Gambian twin children. *J.Infect.Dis.*172:316–9
- Jiang D.K., Ma X.P., Yu H., ... Yu L. (2015) Genetic variants in five novel loci including CFB and CD40 predispose to chronic hepatitis B. *Hepatology* 62, 118-128
- Kallman F.J., Reisner D. (1943) Twin studies on genetic variations in resistance to tuberculosis. *J.Hered.* 34:293–301
- Katamani Y., Wattanapokayakit S., Ochi H., ..., Matsuda K. (2009). A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nature Genetics* 41, 591-595
- Khanh V. Q. L., Lan T. H. N., (2012) Theoretical basis of a beneficial role for vitamin D in viral hepatitis. *World J. Gastroenterol.* 18(38):5338-5350
- Kim D.Y., Kwon E., Hartley P.D., Crosby D.C, Mann S., Krogan N.J., Gross J.D. (2013) CBF $\beta$  stabilizes HIV Vif to counteract APOBEC3 at the expense of RUNX1 target gene expression. *Mol. Cell.* 49(4):632-44.
- Kim Y.J., Kim H.Y., Lee J.H., ... Shin H.D. (2013) A genome-wide association study identified new variants associated with the risk of chronic hepatitis B. *Human Molecular Genetics* 22, 4233–4238.
- Leitsalu L., Haller T., Esko T., ... Metspalu A. (2015) Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology* 44, 1137–1147
- Li Y., Si L., Zhai Y., ... Zhou G., (2016) Genome-wide association study identifies 8p21.3 associated with persistent hepatitis B virus infection among Chinese. *Nature Communications* 7, 11664
- Liu Q., Liu J., Cao Q., Sha J., Zhou Z., Wang H., Li J. (2006) NYD-SP15: a novel gene potentially involved in regulating testicular development and spermatogenesis. *Biochem Genet.* 44(7-8):409-23
- Läll K., Mägi R., Morris A., Metspalu A., Fischer K. (2017) Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. *Genet Med.* 19(3):322-329
- Lok A. S., McMahon B.J., (2007) Chronic hepatitis B. *Hepatology.* 45(6):1347

- Mägi R., Morris A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 11:288
- Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Rev Genetics*, 11(7), 499–511.-401
- Mbarek H., Ochi H., Urabe Y., ... Matsuda K. (2011) A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. *Human Molecular Genetics* 20, 3884–3892
- Miller R. G. (1981) *Simultaneous Statistical Inference* New York Wiley
- Nelson N.P., Easterbrook P.J., McMahon B.J. (2016) Epidemiology of Hepatitis B Virus Infection and Impact of Vaccination on Disease. *Clin Liver Dis.* 20(4):607-628
- Pe'er I., Yelensky R., Altshuler D., Daly M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* 32(4):381-5
- Pearson T.A., Manolio T.A. (2008) How to interpret a genome-wide association study. *JAMA.* 19;299(11):1335-44
- Pettersson F. H., Anderson C.A, Clarke G.M, Cardon L.R., Morris A.P., Zondervan K.T (2009) Marker selection for genetic case-control association studies. *Nat. Protoc.* 4(5): 743-752.
- Riemer J., Appenzeller-Herzog C., Johansson L., Bodenmiller B., Hartmann-Petersen R., Ellgaard L. (2009) A luminal flavoprotein in endoplasmic reticulum-associated degradation. *Proc Natl Acad Sci USA* 106(35):14831-6.
- Risch N., Merikangas K. (1996). The future of genetic studies of complex human diseases. *Science.* 273(5281):1516-7
- Thomas H.C., Anna S.F, Lok S., Locarnini A., Zuckerman A.J. 2014. *Viral Hepatitis, Fourth Edition*, John Wiley & Sons, Ltd
- Tian C., Hromatka B.S., Kiefer A.K., Eriksson N., Noble S.M., Tung J.Y., Hinds D.A. (2017) Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun.* 19;8(1):599
- Ulveling D., Le Clerc S., Cobat A., ... Zagury J.F. (2016) A new 3p25 locus is associated with liver fibrosis progression in human immunodeficiency virus/hepatitis C virus-coinfected patients. *Hepatology* 64(5):1462-1472
- Wade T.D., Gordon S., Medland S., Bulik C.M., Heath A.C., Montgomery G.W., Martin N.G. (2013) Genetic variants associated with disordered eating. *Int. J. Eat. Disord.* 46(6):594-608
- Wagner T.A., McLaughlin S., Garg K., ... Frenkel L.M. (2014) HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science.* 345(6196):570-3.
- Weiss, K. M., Clark, A. G.(2002). Linkage disequilibrium and the mapping of complex human traits. *Trends in Genetics: TIG*, 18(1), 19–24.

WHO. Global hepatitis raport 2017. p 1. lehekülg 400.co.uk. Prantsusmaa.

Winkler A.M., Ridgway G.R., Webster M.A. Smith S.M., Nichols T.E. (2014) Permutation inference for the general linear model. *Neuroimage*. 15;92:381:97

Zauli G., Milani D., Mirandola P., Mazzoni M., Secchiero P., Miscia S., Capitani S. (2001) HIV-1 Tat protein down-regulates CREB transcription factor expression in PC12 neuronal cells through a phosphatidylinositol 3-kinase/AKT/cyclic nucleoside phosphodiesterase pathway. *FASEB J*. 15(2):483-91.

Zondervan, K.T., Cardon, L.R. (2007) Designing candidate gene and genome-wide case-control association studies. *Nature Protocols* 2, 2492–2501

## KASUTATUD VEEBIAADRESSID

<https://emedicine.medscape.com/article/775507-overview>

<https://talk.ictvonline.org/>

<https://www.msdmanuals.com>

<https://www.ncbi.nlm.nih.gov/refseq/>

<http://www.uniprot.org/>

<http://locuszoom.org/>

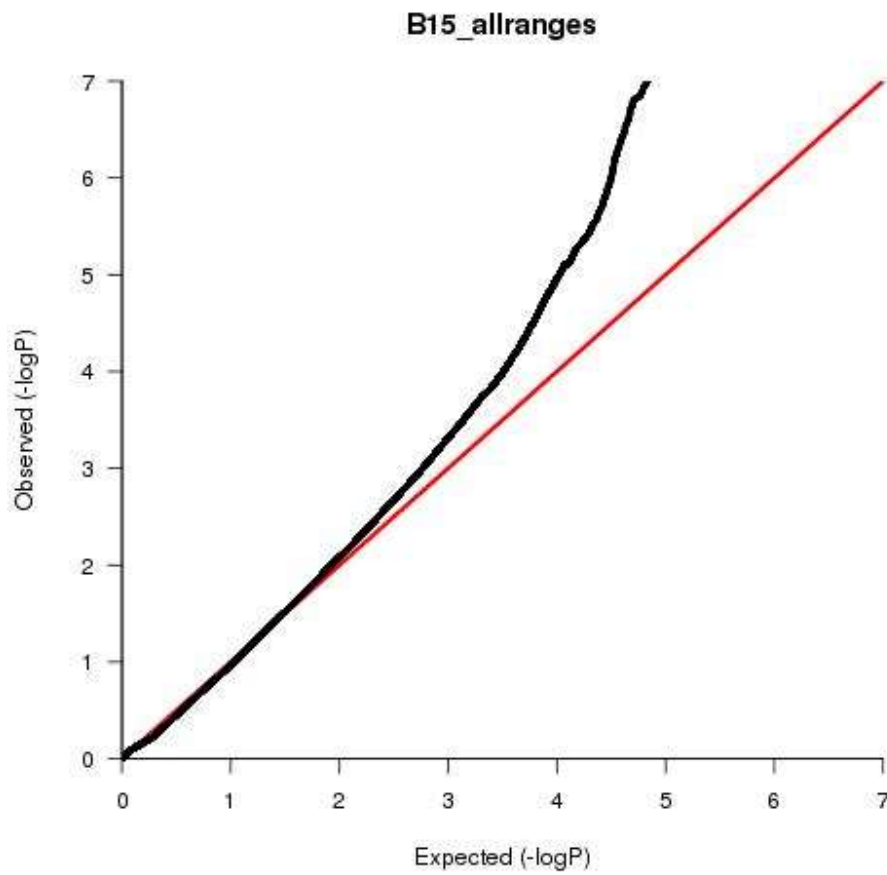
Külastatud 27. mai 2018.

# LISAD

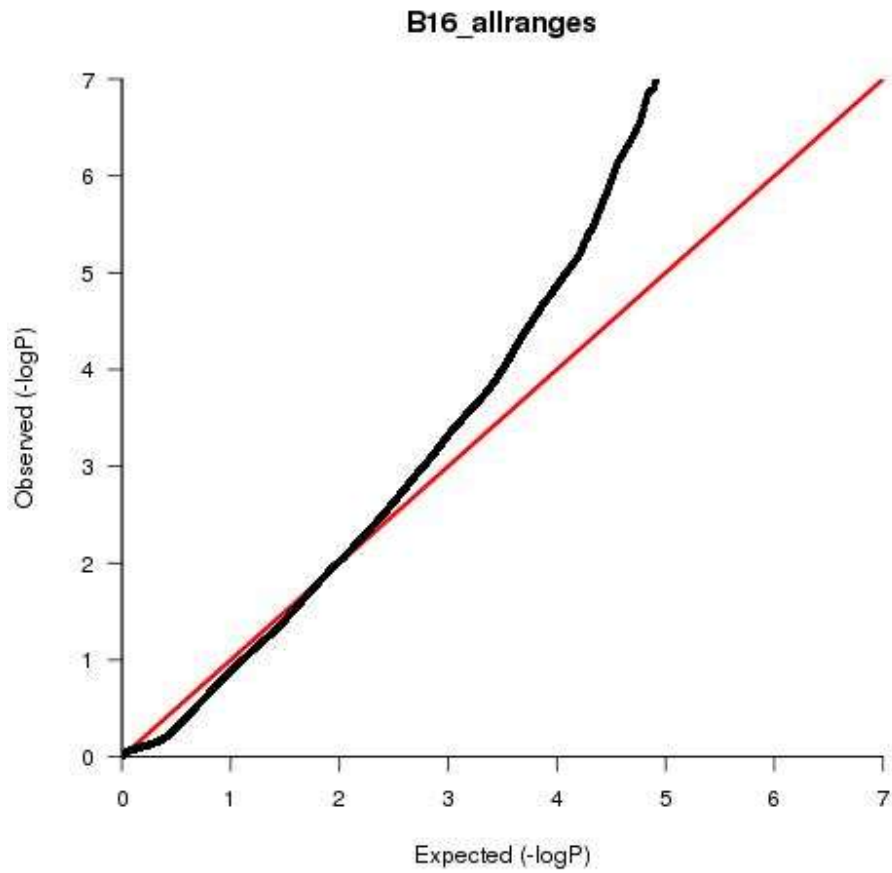
## LISA 1 VIIRUSHEPATIIDI RHK-10 KLASSIFIKATSIOON

<b>KOOD</b>	<b>NIMETUS</b>
<b>B00.8</b>	Herpesviirus hepatiit
<b>B15.0</b>	A-viirushepatiit maksakoomaga
<b>B15.9</b>	A-viirushepatiit maksakoomata
<b>B16.0</b>	Äge B-viirushepatiit delta-agensiga (kaasnakkus) maksakoomaga
<b>B16.1</b>	Äge B-viirushepatiit delta-agensiga (kaasnakkus) maksakoomata
<b>B16.2</b>	Äge B-viirushepatiit delta-agensita maksakoomaga
<b>B16.9</b>	Äge B-viirushepatiit delta-agensita maksakoomata
<b>B17.0</b>	B-hepatiidi viiruskandja äge delta- (super)infektsioon
<b>B17.1</b>	Äge C-hepatiit
<b>B17.2</b>	Äge E-hepatiit
<b>B17.8</b>	Muu äge täpsustatud viirushepatiit
<b>B17.9</b>	Muu äge täpsustamata viirushepatiit
<b>B18.0</b>	Krooniline B-viirushepatiit delta-agensiga
<b>B18.1</b>	Krooniline B-viirushepatiit delta-agensita
<b>B18.2</b>	Krooniline C-viirushepatiit
<b>B18.8</b>	Muud kroonilised viirushepatiidid
<b>B18.9</b>	Täpsustamata krooniline viirushepatiit
<b>B19.0</b>	Täpsustamata viirushepatiit koomaga
<b>B19.9</b>	Täpsustamata viirushepatiit koomata
<b>B25.1</b>	Tsütomegaloviirus-hepatiit
<b>B94.2</b>	Viirushepatiidi jääknähud
<b>O98.4</b>	Rasedust, sünnitust ja sünnitusjärgset perioodi tüsisstav viirushepatiit
<b>P35.3</b>	Kaasasündinud viirushepatiit
<b>Z22.5</b>	Viirushepatiidi kandja

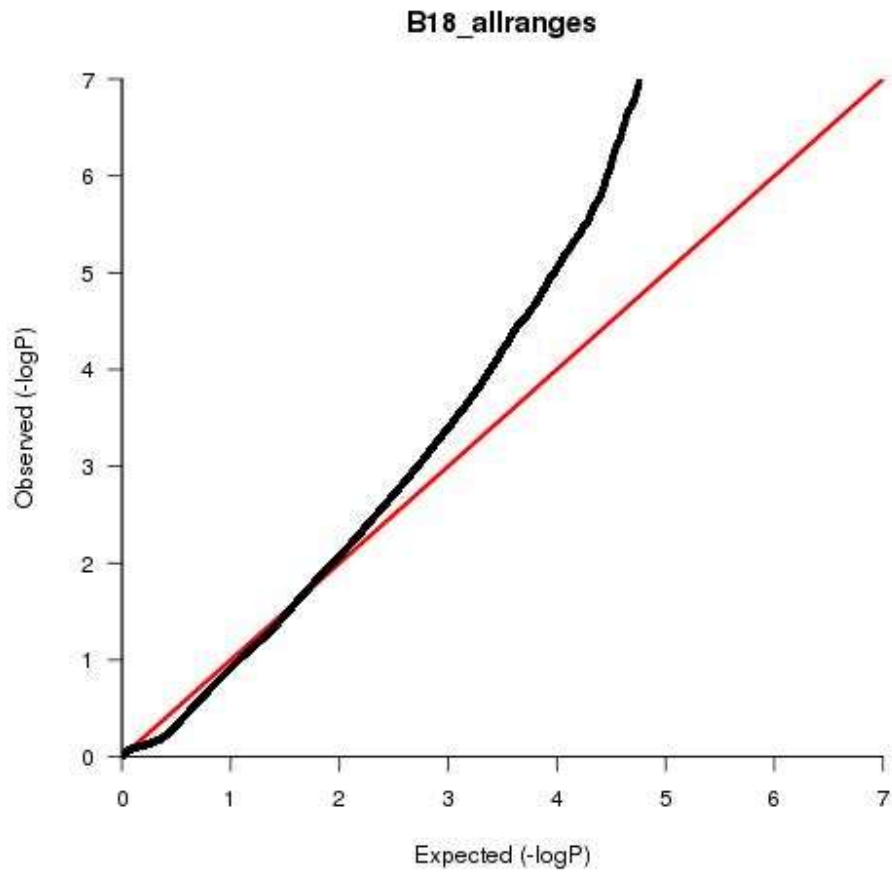
## LISA 2 Q-Q JOONISED



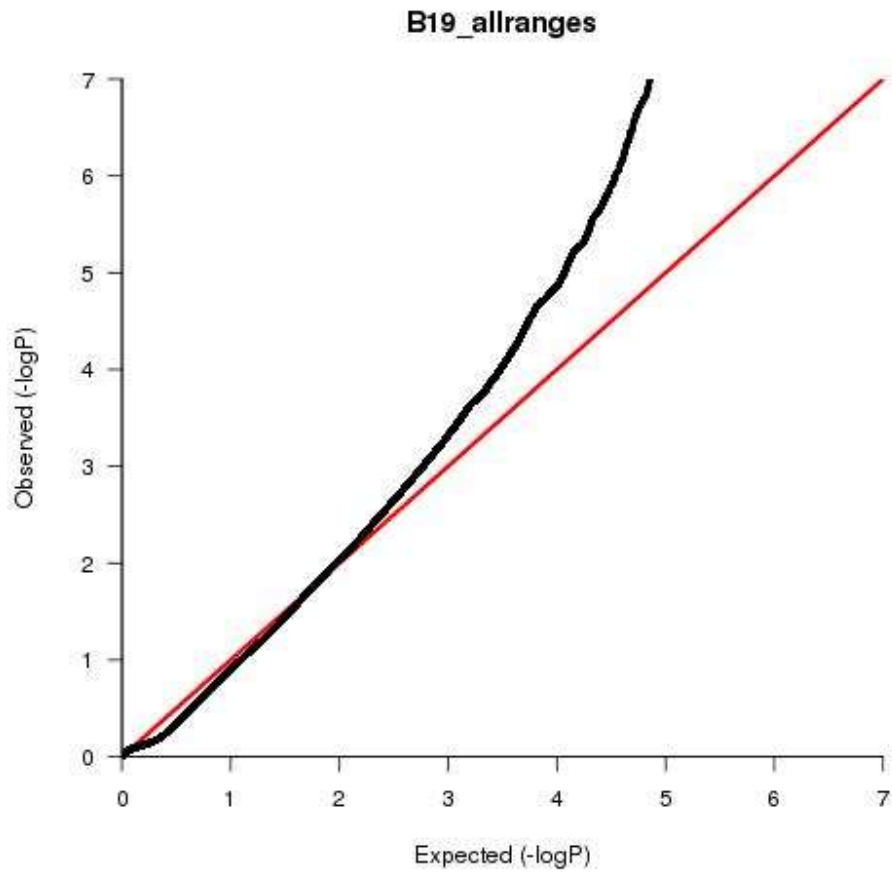
**Joonis B15** eeldatud ja vaadeldud assotsiatsioonide Q-Q joonis. X-teljel on toodud eeldatav p-väärtuse negatiivne kümnendlogaritm ning Y-teljel vaadelda p-väärtuse negatiivne kümnendlogaritm.



**Joonis B16** eeldatud ja vaadeldud assotsiatsioonide Q-Q joonis. X-teljel on toodud eeldatav p-väärtuse negatiivne kümnendlogaritm ning Y-teljel vaadelda p-väärtuse negatiivne kümnendlogaritm.



**Joonis B18** eeldatud ja vaadeldud assotsiatsioonide Q-Q joonis. X-teljel on toodud eeldatav p-väärtuse negatiivne kümnendlogaritm ning Y-teljel vaadelda p-väärtuse negatiivne kümnendlogaritm.



**Joonis B19 eeldatud ja vaadeldud assotsiatsioonide Q-Q joonis.** X-teljel on toodud eeldatav p-väärtuse negatiivne kümnendlogaritm ning Y-teljel vaadelda p-väärtuse negatiivne kümnendlogaritm.

## **LIHTLITSENTS LÕPUTÖÖ REPRODUTSEERIMISEKS JALÕPUTÖÖÜLDSUSELE KÄTTESAADAVAKS TEGEMISEKS**

Mina, Siim Kurvits

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

Ülegenoomne assotsiatsioonuring viirushepatiidiga seonduvate geneetiliste markerite leidmiseks

*(lõputöö pealkiri)*

mille juhendaja on Toomas Haller,

*(juhendaja nimi)*

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.

Tartu, **28.05.2018**