

Tartu Ülikool

Loodus- ja täppisteaduste valdkond

Molekulaar- ja rakubioloogia instituut

Biotehnoloogia õppetool

Kreete Lüll

**Harvade variantide analüüs polütsüstiliste munasarjade
sündroomiga seotud geneetiliste variantide tuvastamiseks**

Magistritöö (30 EAP)

Geenitehnoloogia

Juhendaja PhD Reedik Mägi

TARTU

2016

Infoleht

Harvade variantide analüüs polütsüstiliste munasarjade sündroomiga seotud geneetiliste variantide tuvastamiseks

Polütsüstiliste munasarjade sündroom (ingl k *Polycystic Ovary Syndrome, PCOS*) on maailmas laialt levinud endokriinne komplekshaigus reproduktiivses eas olevatel naistel. Haigust iseloomustavad paljud erinevad tunnused nagu näiteks hüperandrogenism, munasarjade tsüstiline morfoloogia, insuliinresistentsus ning ülekaal või rasvumine. Haiguse etioloogia on siiani jäänud ebaselgeks, selle leidmiseks on läbi viidud mitmeid sagedastele variantidele keskenduvaid ülegenoomseid assotsiatsiooniuringuid, mis on pakkunud mitmeid kandidaatgeene. Üheks uuemaks viisiks leida siiani tabamatuks jäänud *PCOS*-i peidetud pärilikkust on harvade variantide analüüs, mis on läbiviidud ka käesoleva magistritöö raames, kasutades selleks Eesti populatsioonist pärit indiviide.

Märksõnad: polütsüstiliste munasarjade sündroom, *PCOS*, harvade variantide analüüs

CERCS kood: T490, biotehnoloogia

Rare variant analysis to find genetic variants associated with polycystic ovary syndrome

Polycystic ovary syndrome (PCOS) is a common complex disease among women in reproductive age. PCOS is characterized by different traits among which are hyperandrogenism, polycystic ovarian morphology, insulin resistance and overweight or obesity. The etiology of the disease remains unknown, but different genome-wide association studies, using common variants, have revealed many candidate genes. New approach in order to find the so-called missing heritability of PCOS is rare variant analysis. Under this Master's thesis, a rare variant analysis using Estonian cohort is conducted in order to find possible associations between rare variants and PCOS.

Keywords: polycystic ovary syndrome, PCOS, rare variant analysis

CERCS code: T490, biotechnology

Sisukord

Infoleht	2
Kasutatud lühendid	4
Sissejuhatus	6
1. Kirjanduse ülevaade	7
1.1 Polütsüstiliste munasarjade sündroom.....	7
1.1.1. PCOS-i fenotüüp	7
1.1.1.1. PCOS-i diagnoosimise kriteeriumid.....	11
1.1.2 PCOS-i pärilikkus.....	12
1.1.3 PCOS-i geneetika	13
1.2 Peidetud pärilikkus	17
2. Ekperimentaalosa	19
2.1. Töö eesmärk.....	19
2.2. Materjal ja meetodika	19
2.2.1. Valim	19
2.2.2. Genotüpiseerimine ja andmete kvaliteedikontroll.....	20
2.2.3. Andmete ettevalmistus analüüsiks	22
2.2.4. Geneetiline stratifikatsioon.....	23
2.2.5. Statistiline analüüs.....	23
2.3. Tulemused.....	24
2.4. Arutelu	31
Kokkuvõte	35
Resümee	37
Tänuavaldused.....	39
Kasutatud kirjandus	40
Kasutatud veebiaadressid	47
Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	48

Kasutatud lühendid

<i>ASRM</i>	ingl k <i>American Society for Reproductive Medicine</i>
<i>CI</i>	usaldusintervall (ingl k <i>Confidence Interval</i>)
<i>CR</i>	edukalt genotüpeeritud indiviidide protsentuaalne osakaal markeril (ingl k <i>Call Rate</i>)
<i>GWAS</i>	ülegenoomne assotsiatsiooniuuring (ingl k <i>Genome-Wide Association Study, GWAS</i>)
<i>EMMAX</i>	ingl k <i>Efficient Mixed-Model Association eXpedited</i>
<i>EPACTS</i>	ingl k <i>Efficient and Parallelizable Association Container Toolbox</i>
<i>ESHRE</i>	ingl k <i>European Society for Human Reproduction and Embryology</i>
<i>FSH</i>	folliikuleid stimuleeriv hormoon (ingl k <i>Follicle Stimulating Hormone</i>)
<i>GnRH</i>	gonadotropiine vabastav hormoon (ingl k <i>Gonadotrophin Releasing Hormone</i>)
<i>HLA</i>	inimese leukotsüüdi antigeen (ingl k <i>Luman Leukocyte Antigen</i>)
<i>HWE</i>	Hardy-Weinbergi tasakaalustatus (ingl k <i>Hardy-Weinberg Equilibrium</i>)
<i>kb</i>	1000 aluspaari (ingl k <i>kilobase</i>)
<i>LH</i>	luteiniseeriv hormoon (ingl k <i>Luteinizing Hormone</i>)
<i>LOF</i>	funktsioonikadu (ingl k <i>Loss-of-Function</i>)
<i>MAF</i>	harvema alleeli sagedus (ingl k <i>Minor Allele Frequency</i>)
<i>MDS</i>	multidimensionaalne hindamine (ingl k <i>Multidimensional Scaling</i>)
<i>MS</i>	mittesünonüümne
<i>NIH</i>	ingl k <i>National Institutes of Health</i>
<i>OR</i>	šansside suhe (ingl k <i>Odds Ratio</i>)
<i>PCOS</i>	polütsüstiliste munasarjade sündroom (ingl k <i>Polycystic Ovary Syndrome</i>)
<i>SHBG</i>	suguhormoone siduv globuliin (<i>Sex Hormone-Biding Globulin</i>)
<i>SKAT</i>	ingl k <i>SNP-set (Sequence) Kernel Association Test</i>

SKAT-O ingl k *Optimized Sequence Kernel Association Test*

SNP üksiknukleotiidne polümorfism (ingl k *Single Nucleotide Polymorphism*)

Sissejuhatus

Polütsüstiliste munasarjade sündroom on komplekshaigus, mis esineb kuni kümnel protsendil reproduktiivses eas olevatel naistel, olles seega üks enamlevinud endokriinseid haiguseid. Sündroomi iseloomustavad mitmed sümptomid, mille hulka kuuluvad näiteks polütsüstilised munasarjad, menstruaaltsüklihäired ja hormonaalsed häired.

PCOS-i geneetilisi põhjuseid on varasemalt uuritud nii aheldusanalüüsides kui ka ülegenoomsetes assotsiatsiooniuuringutes, mis on küll paljastanud mitmeid seotud lookuseid, kuid sellest hoolimata on suur osa *PCOS*-i geneetilisest taustast seni tuvastamata (Day *et al.*, 2015; Hayes *et al.*, 2015).

Üks võimalikke ja seni väheuuritud *PCOS*-i geneetilise eelsoodumuse allikaid on harvad genoomsed variandid. Lisaks on võimalik, et ka teadaolevad assotsiatsioonid sagedaste geneetiliste variantidega on põhjustatud samades piirkondades leiduvate ja sagedaste variantidega aheldunud harvade kausaalsete variantide poolt. Seetõttu uuringi antud töös kodeerivad harvu variante, millel võib olla otsene põhjuslik mõju *PCOS*-i tekkimisel. Leides geneetilisi variante, mis mõjutavad *PCOS*-iga seotud gene ja nende funktsioone, oleks võimalik edasi uurida just nimelt assotsieerunud geenide funktsioone. See aitaks kaasa haiguse bioloogilise tausta paremale mõistmisele ja võimaldaks tulevikus hakata otsima täiendavaid diagnostika ja ravivõimalusi.

Käesolev magistritöö on valminud Tartu Ülikooli Eesti Geenivaramus ja selle eesmärgiks on läbi viia hüpoteesivaba harvade variantide analüüs, mille sihiks on leida assotsiatsioone polütsüstiliste munasarjade sündroomiga.

1. Kirjanduse ülevaade

1.1 Polütsüstiliste munasarjade sündroom

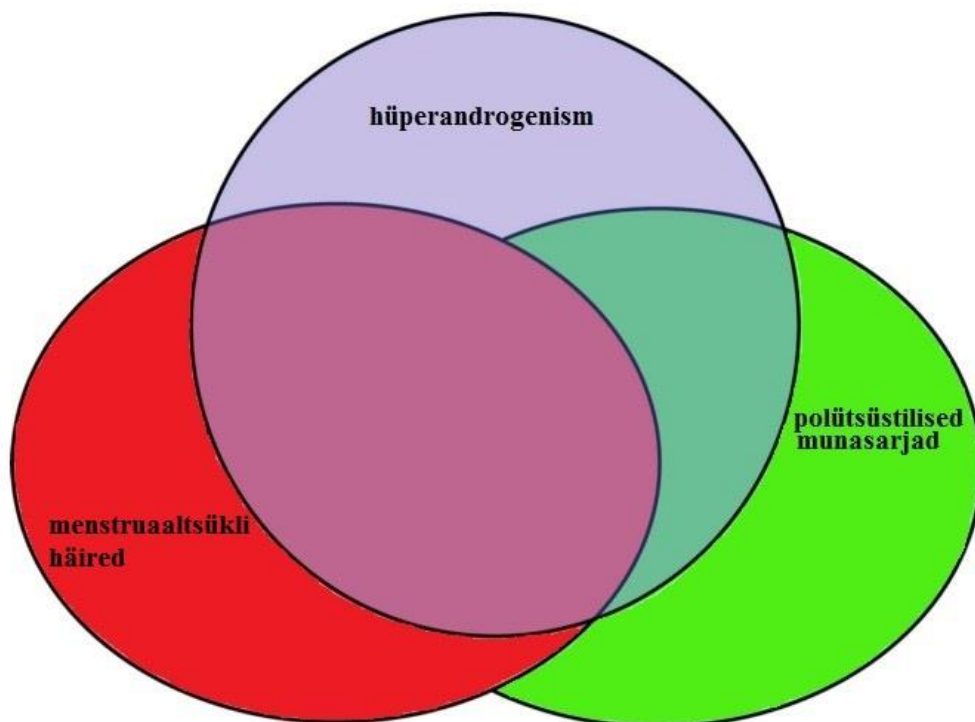
1.1.1 PCOS-i fenotüüp

Polütsüstiliste munasarjade sündroom (ingl k *Polycystic Ovary Syndrome, PCOS*) on haigus, mille esmakirjeldus jääb aastasse 1935 ning mis on üks enimlevinud kompleksseid endokriinseid metaboolseid haigusi reproduktiivses eas naistel (Azziz *et al.*, 2009). PCOS-i põdejaid on kirjanduse andmetel tänasel päeval kogu maailmas ligikaudu 7–15%, sõltuvalt kasutatavast diagnoosikriteeriumist. Diagnoosimisel kasutatakse kolme peamise tunnusega menstruaaltsükli häireid, hüperandrogenismi ehk kõrgeenenud meessuguhormoonid taset ning tsüstilise morfoloogiaga munasarjade esinemist (Diamanti-Kandarakis ja Dunaif, 2012).

PCOS-i kõige iseäralikumaks tunnuseks on polütsüstilised munasarjad, kuid lisaks leidub veel teisi kliinilisi ja biokeemilisi tunnuseid, mis sündroomi iseloomustavad: nagu näiteks viljatus, androgeenide, insuliini ja luteiniseeriva hormooni kõrgeenenud kontsentratsioon seerumis ning vähenenud insuliintundlikkus (Crosignani ja Nicolosi, 2001). Hüperandrogenism, mis on oluline sündroomi tunnus, diagnoositakse PCOS-i puhul juhul, kui haigel esineb hirsutism ehk liigne karvakasv, androgeenne alopeetsia ehk meestüüpi juuksekadu, akne või kõrgeenenud androgeenide tase (Diamanti-Kandarakis *et al.*, 1999; Sirmans ja Pate, 2013; Zhang *et al.*, 2013). PCOS-i metaboolne fenotüüp sisaldab eeskätt selliseid omavahel tihedalt seotud sümptomeid nagu metaboolne sündroom, insuliinresistentsus, ülekaal või rasvumine ja suurenenud risk tüüp II diabeediks (Casarini ja Brigante, 2014; Diamanti-Kandarakis *et al.*, 1999; Diamanti-Kandarakis ja Dunaif, 2012). On näidatud, et metaboolne fenotüüp on levinud rohkem Kesk-Aasias ning Ameerikas, samal ajal kui hüperandrogeenne fenotüüp on rohkem esindatud Euroopas ning teistes maailma piirkondades (Casarini ja Brigante, 2014).

PCOS-ile iseloomulikud anovulatsioon ja oligo-ovulatsioon ning hüperandrogenism, on peamiselt põhjustatud ebakorrapärasest gonadotropiinide, mille perekonda kuuluvad ka luteiniseeriv hormoon (ingl k *Luteinizing Hormone, LH*) ja folliikuleid stimuleeriv hormoon (ingl k *Follicle-Stimulating Hormone, FSH*), sekretsioonist (Diamanti-Kandarakis ja Dunaif, 2012). Kõrgeenenud LH tase stimuleerib munasarjades testosterooni tootmist ning PCOS-i puhul ongi kõrgeenenud androgeenide produktsioon osaliselt LH-sõltuv. Teisel gonadotropiinil FSH-il on keskne roll munasarja follikulogeneesis. PCOS-iga naistel on follikulogenees häiritud tingituna eeskätt just FSH sekreteerimise vaegusest (S Franks ja Hardy, 2010; Franks *et al.*, 2008). Kõrgeenenud testosteroonitaseme tagasiside hüpotaalamusele põhjustab

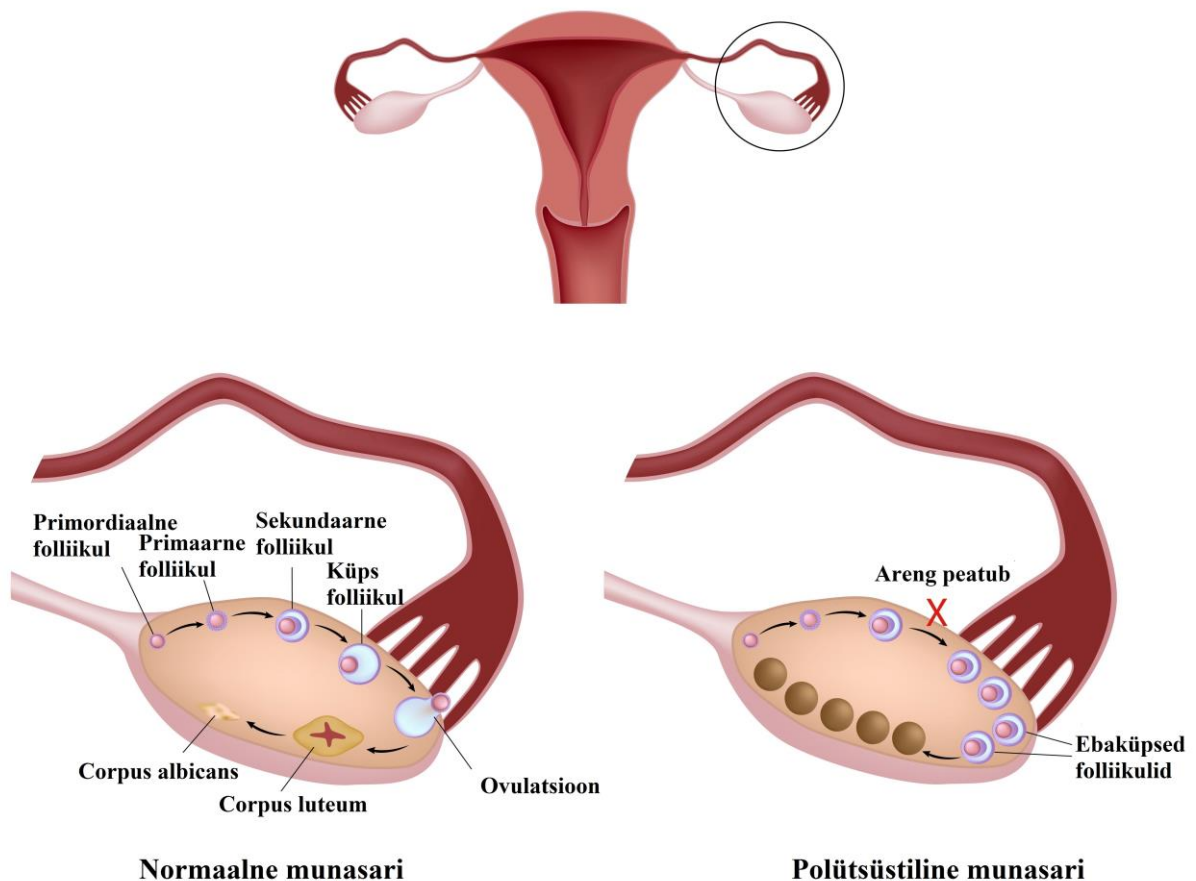
omakorda ebakorrapärasest gonadotropiini vabanemist (Hayes *et al.*, 2015). Katseloomade puhul on näidatud, et *PCOS*-ile omane fenotüüp tekkis juhul, kui neile kriitilisel arenguhetkel täiendavalt androgeene manustati, millest võiks järeldada, et sarnane mehhanism võib toimida ka inimorganismis (Abbott *et al.*, 2002). Lisaks leidub *PCOS*-i puhul gonadotropiin-sõltuvas prenataalsete folliikulite arengus anormaalsusi, mille tulemuseks on primaarsete folliikulite liig võrrelduna normaalsete munasarjadega (Hayes *et al.*, 2015). Siiski on oluline välja tuua, et rohkem kui 20 protsendil regulaarse menstruaaltsükliga naistel, eriti just noortel naistel, esinevad munasarjades tsüstid, ehkki nad ei põe *PCOS*-i (Johnstone *et al.*, 2010). Peamised *PCOS*-i sümptomid, mida kasutatakse ka haiguse diagnoosimiseks, on välja toodud joonisel 1.



Joonis 1. *PCOS*-i peamised tunnused. *PCOS*-i diagnoosimise kriteeriumid sisaldavad kahte või rohkemat joonisel esitatud tunnustest: hüperandrogenism, menstruaaltsükli häired ja polütsüstilised munasarjad. (mugandatud Diamanti-Kandarkis ja Dunaif järgi, 2012).

Uuringutes on täiendavalt näidatud, et *PCOS*-iga kaasnev hüperandrogenism on ülekaalulistel ja normaalkaalulistel patsientidel tingitud erinevatest aspektidest. Nimelt on püstitatud hüpotees, et ülekaalulistel *PCOS*-i patsientidel on hüperandrogenismi põhjuseks hüperinsulineemia (olukord, kus kehas ringleb liigselt insuliini), mis toob kaasa suurenenud androgeenide produktsiooni. Normaalkaalulistel aga liigest kasvuhormooni ja LH tasemest (Inslar *et al.*, 1993). Nimelt normaalne munasarja funktsioon on kontrollitud peamiselt LH ja FSH poolt, mida toodetakse ajuripatsis ja mille sekreteerimist reguleerib gonadotropiine

Lisaks on oluliseks *PCOS*-i sümptomiks viljatuse, mida esineb ligikaudu 40% haigetest naistest (Teede *et al.*, 2010). Kui tervel naisel folliikulid küpsevad ja munarakk vabastatakse igas menstruaaltsükli, siis *PCOS*-i puhul ei toodeta piisavalt hormoone selleks, et folliikul lõpuni areneks, mis põhjustab olukorra, kus munarakk ei vabane ning ovulatsiooni ei teki, mis omakorda toobki kaasa viljatuse (joonis 3) (Hart *et al.*, 2004).



Joonis 3. Follikukogeenes tervetel ja *PCOS*-iga naistel. *PCOS*-i puhul folliikuli areng peatub ning ovulatsiooni ei toimu. Ebaküpsed folliikulid jäävad munasarja ning põhjustavad tsüstilise munasarja. (mugandatud <http://www.clearpassage.com/what-you-should-know-about-pcos/> (kasutatud 06.05.2016)).

PCOS-i sümptomite leevendamise üheks kõige efektiivsemaks viisiks on elustiili muutmine – eelkõige kaalukaotus. Ligikaudu 50% *PCOS*-iga naistest on ülekaalulised või rasvunud (Domecq *et al.*, 2013). Lisaks on teada, et rasvumine süvendab märgatavalt *PCOS*-i sümptomeid ja sellel võib olla oluline roll hüperandrogenismi ja kroonilise anovulatsiooni tekkimisel (Yildiz *et al.*, 2008). Suurenenud rasvkoe hulka on seostatud arvukate suguhormoonide metabolismihäiretega, sealhulgas suurenenud androgeenide tootmisega.

Täiendavalt on täheldatud, et rasvunud *PCOS*-i patsientidel on kardiovaskulaarste ja metaboolsete haiguste risk märgatavalt kõrgem kui neil, kes on normaalses kaaluvahemikus (Cattrall ja Healy, 2004; Daan *et al.*, 2014). Sellest tulenevalt on kaalulangetamine üheks kõige efektiivsemaks haiguse leevendamise meetodiks, see küll ei ravi haigusest terveks, kuid leevendab tunduvalt olemasolevaid sümptomeid (Domecq *et al.*, 2013; Moran *et al.*, 2013).

1.1.1.1. PCOS-i diagnoosimise kriteeriumid

Tulenevalt asjaolust, et *PCOS*-i täpsed põhjused on siiani jäänud ebaselgeks ja haiguse kliiniline pilt on väga varieeruv ning mitmekesine, on ka diagnoosimiseks kasutusel erinevad kriteeriumid (Diamanti-Kandarakis ja Dunaif, 2012). Esimene kriteerium, mille järgi *PCOS*-i diagnoosima hakati, on *National Institutes of Health* ehk *NIH* kriteerium. Nimetatud kriteeriumi kohaselt peavad *PCOS*-i diagnoosi panemiseks inimesel esinema kas kliiniline või biokeemiline hüperandrogenism ja krooniline an- või oligo-ovulatsioon ning välistatud peavad olema haigused, millel on sarnased sümptomid (näiteks kaasasündinud neerupearliste hüperplaasia, Cushingi sündroom ning androgeene eritavad kasvajad) (Azziz *et al.*, 2004; Diamanti-Kandarakis ja Dunaif, 2012; Teede *et al.*, 2010). Põhjus, miks *NIH* kriteeriumi hulka ei lisatud polütsüstilisi munasarju, seisnes selles, et 20–30 protsendil naistest, kellel oli regulaarne menstruaaltsükkel, leiti ultrahelivaatluse käigus munasarjadest tsüste (Polson *et al.*, 1988). *NIH*-le järgnes *ESHRE/ASRM* ehk Rotterdami kriteerium, mis lisas diagnoosimiseks võimalike sümptomite hulka polütsüstilised munasarjad. Teisisõnu peavad *PCOS*-i diagnoosimiseks Rotterdami kriteeriumi järgi esinema vähemalt kaks järgnevast kolmest sümptomist: kliiniline või biokeemiline hüperandrogenism (defineeritud kui kõrgenenud seerumi androgeenide tase või hirsutism ja/või akne), an- või oligo-ovulatsioon ning polütsüstilised munasarjad. Rotterdami kriteerium kasutuselevõtmine tõi kaasa selle, et *PCOS*-i diagnoosi alla koondusid mitmed erinevad fenotüübid. Näiteks võib sama diagnoos olla naisel, kellel on polütsüstilised munasarjad ja hüperandrogenism, aga mitte anovulatsioon või naised, kellel on anovulatsioon ja polütsüstilised munasarjad, kuid ei esine hüperandrogenismi (Teede *et al.*, 2010). Teades, et haigete fenotüüp on varieeruv, ei ole siiani olnud võimalik kindlalt öelda, kas erinevate *PCOS*-i fenotüüpide geneetiline taust on samuti erinev või mitte (Hayes *et al.*, 2015).

2006. aastal tuli välja kolmas niinimetatud *Androgen Excess and PCOS Society* kriteerium, mis nõuab diagnoosimiseks seda, et esineks nii hüperandrogenism, polütsüstilised munasarjad kui ka an- või oligo-ovulatsioon (Azziz *et al.*, 2006). *Androgen Excess and PCOS Society* kriteerium on neist kolmest kõige rangem ning sellele tuginedes on sündroomi määramine kõige täpsemalt piiritletud ja puudub võimalus, et sündroomi kandjatel on üksteisest erinev fenotüüp. Lisaks on jõutud konsensussele, et *NIH* kriteeriumit järgides on kõige suurem risk määratleda *PCOS*-ina insuliinresistentsusega fenotüüpi, mis on ennekõike seotud ka metaboolsete tunnusjoontega, samal ajal kui teise kahe kriteeriumi korral on *PCOS*-i patsientide metaboolne risk väiksem (Diamanti-Kandarakis ja Dunaif, 2012). Täiendavalt, *PCOS*-i esinemissagedus reproduktiivses eas olevate naiste seas sõltub samuti sellest, missugust diagnoosikriteeriumit on kasutatud (Sirmans ja Pate, 2013). Ülevaade kolmest diagnoosikriteeriumist asub tabelis 1.

Tabel 1. *PCOS*-i diagnoosikriteeriumid

<i>NIH</i>	<i>ESHRE/ASRM</i> ehk Rotterdam	<i>Androgen Excess and PCOS Society</i>
<i>PCOS</i> diagnoositakse kui kõik kolm järgnevat punkti on täidetud:	<i>PCOS</i> diagnoositakse, kui esineb kaks kolmest tunnusest:	<i>PCOS</i> diagnoositakse kui kõik kolm järgnevat punkti on täidetud:
<ul style="list-style-type: none"> kliiniline või biokeemiline hüperandrogenism 	<ul style="list-style-type: none"> kliiniline või biokeemiline hüperandrogenism 	<ul style="list-style-type: none"> kliiniline või biokeemiline hüperandrogenism
<ul style="list-style-type: none"> an- või oligo-ovulatsioon 	<ul style="list-style-type: none"> an- või oligo-ovulatsioon 	<ul style="list-style-type: none"> an- või oligo-ovulatsioon
<ul style="list-style-type: none"> väljastatud on sarnaste sümptomitega haigused 	<ul style="list-style-type: none"> polütsüstiliste munasarjade morfoloogia 	<ul style="list-style-type: none"> polütsüstiliste munasarjade morfoloogia

1.1.2 *PCOS*-i pärilikkus

Hoolimata sellest, et *PCOS* on laialt levinud haigus, ei ole selle etioloogia senini lõplikult selge. Uuringutest on küll selgunud, et haigusel on olemas päritav komponent ja välja on pakutud mitmeid kandidaatgeene, kuid sajaprotsendiliselt ei ole haiguse päritolu osas konsensussele jõutud (Day *et al.*, 2015; Hayes *et al.*, 2015; Shi *et al.*, 2012; Welt *et al.*, 2012).

2006. aastal viisid Vink jt Hollandis läbi kaksikute uuringu eesmärgiga hinnata *PCOS*-i päritavust. Uuringus kasutati 1332 ühemunakaksikute ja 1873 kahemunakaksikute paari andmeid ning *PCOS*-i diagnoosimisel lähtuti Rotterdami kriteeriumist (Vink *et al.*, 2006). Uuringu tulemusena leiti, et *PCOS*-i päritavus on kuni 79 protsenti. (Vink *et al.*, 2006).

Kaks aastat hiljem, 2008. aastal, avaldasid Franks jt töö, mille hüpotees oli, et munasarjade morfoloogia on pärilike biokeemiliste tunnuste markeriks. Munasarjade morfoloogia pärilikkust uuriti sellepärast, et nii hüperandrogenismil kui ka metaboolsetel anormaalsetel on *PCOS*-i puhul olemas perekondlik koosinemine (Yildiz *et al.*, 2003). Uuringusse kaasatud katseisikutel (n=125) oli *PCOS* diagnoositud juhul, kui neil esines anovulatsioon, ja/või hüperandrogenism ja polütsüstilised munasarjad. Lisaks kuulusid valimisse õed, kellel olid polütsüstilised munasarjad, sõltumata sellest, kas neil esines muid *PCOS*-ile iseloomulikke sümptomeid või mitte ja õed, kellel oli normaalne munasarja morfoloogia. Töö tulemustest järeldasid Franks jt, et androgeenide kontsentratsiooni ja munasarja morfoloogia vahel on väga tugev seos, sõltumata teistest kliinilistest tunnustest (Franks *et al.*, 2008). *PCOS*-i perekondlikku klasterdumist uurivates töodes keskendutakse peamiselt munasarjade morfoloogiale, menstruaaltsükli häiretel ja hüperandrogenismi sümptomitele. Samuti on selliste uuringute käigus täheldatud, et *PCOS*-iga naiste meessoost sugulastel esineb sagedamini enneaegset kiilanemist ehk alopeetsiat (Crosignani ja Nicolosi, 2001).

1.1.3 *PCOS*-i geneetika

Mitmetest uuringutest on selgunud, et *PCOS*-il on olemas tugev geneetiline komponent. Geneetiliste uuringute kaudu on võimalik tuvastada peamisi haigusega seotud geene ja signaaliradu aidates seeläbi *PCOS*-i etioloogiat paremini mõista (Day *et al.*, 2015). Läbi viidud kandidaatgeenuuringuid on välja pakkunud erinevaid geene, mis võiksid haigusega seotud olla. Paraku ei ole enamik neist leitud seostest hilisemates uuringutes replitseerunud, seda eelkõige uuringute madala võimsuse ja väikese valimi tõttu (Barber ja Franks, 2013). Assotsiatsiooniuuringute käigus on uuritud vähemalt 70 erinevat kandidaatgeeni, mis on peamiselt seotud suguhormoonide, insuliinresistentsuse ja põletikuliste protsessidega, nagu näiteks folliikuleid stimuleeriva hormooni retseptor (*FSHR*), insuliini retseptor (*INSR*) ja interleukiin-6 (*IL6*) (Chen *et al.*, 2011). Üleüldiselt assotsiatsiooniuuringud on olnud heaks meetodiks *PCOS*-i uurimisel seetõttu, et need on hüpoteesivabad ja ei eelda varasemat haiguse patogeneesi täielikku mõistmist. See võimaldab esile kerkida uut patogeneesi

radadel ja mehhanismidel, mis võivad *PCOS*-i kujunemisega seotud olla (Barber ja Franks, 2013).

Ühed suurematest ülegenoomsetest assotsiatsioonianalüüsides, mis seni seoses *PCOS*-iga teostatud, on läbi viidud han-hiinlaste seas. Chen jt teostasid uuringu, kus uurisid 744 *PCOS*-i juhtu ja 895 kontrollindiviidi ning viisid seejärel läbi 2-astmelise replikatsiooni: esimese põhja han-hiinlaste seas (2840 juhtu ja 5012 kontrolli) ning teise lõuna ja kesk han-hiinlaste seas (498 juhtu ja 780 kontrolli). Kõik sealsed *PCOS*-i juhud olid diagnoositud lähtudes 2003. aasta Rotterdami kriteeriumist. Uuringu tulemusena leiti kolm uut *PCOS*-iga assotsieeruvat lookust (Chen *et al.*, 2011). Teine suuremahuline uuring han-hiinlaste seas kinnitas eelnevalt leitud kolme lookust ning lisaks tuvastati kaheksa uut assotsiatsiooni, mis asusid geenide *C9orf3*, *YAP1*, *RAB5B*, *HMGA2*, *TOX3*, *INSR*, *SUMO1P1* ja *FSHR* läheduses (Shi *et al.*, 2012). Siiski ei ole enamik tulemusi, mis on leitud han-hiinlastega läbi viidud uuringutes, seni veel Euroopa päritolu valimiga uuringutes replitseerunud (Day *et al.*, 2015; Goodarzi *et al.*, 2012; Welt *et al.*, 2012). Põhjus võib peituda asjaolus, et han-hiinlaste kohordis saadud tulemused on omased just sealsele piirkonnale ega avaldu teistes kohortides, sealhulgas Euroopa päritolu valimis.

2015. aastal avaldasid Hayes jt töö, mille eesmärk oli leida sündroomiga võimalikus seoses olevaid uusi lookuseid, kasutades *NIH* kriteeriumit, mille kasutamisel on *PCOS*-i patsientide hulgas enim metaboolsete haiguste ja kõrgenenud suguhormoonide tasemega haigeid (Hayes *et al.*, 2015). Uuringus viidi esmalt läbi ülegenoomne assotsiatsiooniuring (ingl k *genome-wide association study*, *GWAS*), milles kasutati 984 *PCOS*-i juhtu ja 2964 kontrolli. Sarnaselt han-hiinlaste suuremahulisele uuringule viisid ka Hayes jt oma töös läbi kaheetapilise replikatsiooni, mille esimeses etapis kasutati 1799 *PCOS*-i juhtu ning 1231 kontrolli. Teises replikatsioonis kasutati 24 parimat tulemust, mis saadi põhiuuringu ja esimese replikatsiooni meta-analüüsist ning see viidi läbi täiendava 217 *PCOS*-i juhu ja 1335 kontrolliga (Hayes *et al.*, 2015). Teostatud *GWAS*-ist ilmnis kolm lookust, mis olid ülegenoomselt olulised ja need paiknesid geenide *DGKI*, *CASC2* ja *FSHB/ARL14EP* läheduses. Nimetatud geenid on vastavalt seotud testosterooni, folliikuleid stimuleeriva hormooni ja luteiniseeriva hormooni tasemetega (Hayes *et al.*, 2015). Kombineerides kolm analüüsi (*GWAS* ja replikatsioonianalüüsid), leiti kaks uut lookust, mis ületasid ülegenoomse statistilise olulisuse ning need asusid geenide *GATA4/NEIL2* ja *FSHB/ARL14EP* läheduses. Täiendavalt oli ülegenoomselt oluline veel üks lookus, mis oli üles kerkinud ka Shi jt poolt läbi viidud han-hiinlaste töös ning selleks oli lookus *C9orf3/FANCC* läheduses. Need leiud viitavad sellele, et

FSH geneetiline varieeruvus omab olulist rolli *PCOS*-i etioloogias Euroopa päritoluga naistel, kes on diagnoositud *NIH* kriteeriumi põhjal (Hayes *et al.*, 2015). Teine uus lookus, mis Euroopa päritolu valimis *PCOS*-iga assotsieerus, asus *GATA4* ja *NEIL2* läheduses. *GATA4* on geen, mis kodeerib tsinksõrme transkriptsioonifaktorit, mis omakorda reguleerib gonaadide arengut ja steroidogeensete (ingl k *steroidogenic*) geenide transkriptsiooni.

Hayes jt kohandasid oma uuringut kehamassiindeksile, kuna see on *PCOS*-iga tugevas seoses. Oluline tähelepanek nende tööst oli see, et nad ei tuvastanud ühtegi ülegenoomse olulisusega assotsiatsiooni *PCOS*-i ja teadaolevate rasvumise ega tüüp 2 diabeediga seotud geenide vahel. Sellest võib järeldada, et geenid, mis on olulised kehamassiindeksi seisukohast, ei pruugi omada *PCOS*-i geneetikas olulist mõju. (Hayes *et al.*, 2015).

Järgmine mahukas uuring, mille valimisse kuulusid Euroopa päritolu invidiidid, viidi läbi Day jt poolt. Valim koosnes 5184 ise raporteeritud (ingl k *self-reported*) *PCOS*-i juhust ja 82 759 kontrollist ning kontrolluuringu valimisse kuulus 2000 kliiniliselt kinnitatud *PCOS*-i juhtu ja 100 000 kontrolli. Uuringu tulemusena leiti kuus signaali, mis olid ülegenoomset olulised: neli uut leitud ning kaks varasemalt teistes uuringutes kirjeldatud lookust (Day *et al.*, 2015). Kõik leitud signaalid olid vähemalt nominaalselt olulised (p -väärtus $<0,05$), kasutades ainult kliiniliselt valideeritud *PCOS*-i juhte. Tugevaima uudse signaali andis Day jt töös lookus *ERBB4/HER4* intronis, mis kodeerib epidermaalse kasvufaktori retseptori perekonna (*EGFR*) liikmeid. Võib-olla kõige huvipakkuvam assotsiatsioon, mis uuringust ilmnes, oli signaal, mis asus geeni *FSHB* läheduses. Huvitav on tulemus seetõttu, et Day jt lugesid seda küll uudseks signaaliks, kuid tegelikkuses olid seda vahetult enne artikli ilmumist kirjeldanud Euroopa päritolu valimi seas Hayes jt (Day *et al.*, 2015; Hayes *et al.*, 2015). Tulenevalt sellest, et *FSHB* ilmnes mõlemas suures Euroopa päritolu valimiga läbiviidud uuringus ning see on seotud follikulogeneesi ja kõrgenenud *LH/FSH* suhtega, võib seda lugeda tugevaks *PCOS*-i kandidaatgeeniks. Teised varasemalt kirjeldamata signaalid, mis ilmnesid, asusid geenide *RAD50* ja *KRR1* läheduses, mis on vastavalt seotud DNA kaksikahelalise katke parandamise ja ribosoomi assambleerimise (ingl k *assembly*) faktori kodeerimisega (Day *et al.*, 2015).

Day jt võrdlesid oma tulemusi lisaks veel kahe suure han-hiinlaste seas läbiviidud uuringuga (Chen *et al.*, 2011; Shi *et al.*, 2012). Üheteistkümnest han-hiinlaste valimis leitud assotsiatsioonist ilmnes Euroopa päritoluga uuringus vaid kaks geeni, milleks olid *YAPI* ja *THADA*. Ülejäänud üheksa tugevat assotsiatsiooni, mis Chen jt ning Shi jt töödes esile kerkisid, ei näidanud Day jt töös ülegenoomset statistilist olulisust. Sellised tulemused võivad

peegeldada Hiina ja Euroopa päritolu gruppide alleelse struktuuri erinevusi, mida on varasemalt ka teistes töödes kirjeldatud (Mutharasan *et al.*, 2013). Lisaks võib sellest järeldada, et alleelsetest erinevustest tulenevalt võib olla keeruline nende kahe populatsiooni vahel läbi viia tavapärasest meta-analüüsi (Day *et al.*, 2015). Veel üks väga oluline järeldus, milleni artiklis jõuti, on see, et haigusjuhtude ise raporteerimine on uuringu ülesehituse seisukohast vaadatuna väga tõhus meetod. See võimaldab kasutada suuremal hulgal haigusjuhte kui ainult kliiniliselt kinnitatud juhud ja annab seeläbi uuringule piisava võimsuse, millega *PCOS*-iga assotsieeruvaid geneetilisi signaale otsida. Lisaks täheldati, et ise raporteeritud juhtude ja kliiniliselt kinnitatud juhtudega teostatud analüüsid andsid samaväärseid tulemusi. (Day *et al.*, 2015). Suuremate *PCOS*-iga seotud uuringute kokkuvõtte on toodud tabelis 2.

Tabel 2. Suuremate *PCOS*-i uuringute koondtabel. Välja on toodud uuring, selle esimene autor, valimi suurus ning töödest ilmnenuid kandidaatgeenid. Juhtude ja kontrollide arvudes on esimesena välja toodud algse uuringu indiviidide arv ning peale pluss märki sama uuringu replikatsioonis olnud indiviidide arv (Chen jt töö raames viidi läbi kaheastmeline replikatsioon).

Uuring	Han-hiina 1 (Chen <i>et al.</i> , 2011)	Han-hiina 2 Shi <i>et al.</i> , 2012)	Euroopa 1 (Hayes <i>et al.</i> , 2015)	Euroopa 2 (Day <i>et al.</i> , 2015)
Juhtude arv	744+2840+498	1510+8226	5187+2000	984+1799
Kontrollide arv	895+5012+780	2016+7578	82 759+100 000	2964+1231
Ilmnenuid kandidaatgeenid	<i>FSHR</i> <i>THADA</i> <i>DENND1A</i>	<i>FSHR</i> <i>THADA</i> <i>DENND1A</i> <i>LHCGR</i> <i>C9orf3</i> <i>YAP1</i> <i>RAB5B/SUOX</i> <i>HMGA2</i> <i>TOX3</i> <i>INSR</i> <i>SUMO1P1</i>	<i>GATA4/NEIL2</i> <i>C9orf3</i> <i>FSHB</i>	<i>THADA</i> <i>ERBB4</i> <i>RAD50</i> <i>FSHB</i> <i>YAP1</i> <i>KRR1</i>

1.2 Peidetud pärilikkus

Ülegenoomsed assotsiatsiooniuuringud on võimaldanud küll saada väärtuslikke teadmisi inimese geneetiliste haiguste kohta (Hindorff *et al.*, 2009), kuid samas on komplekshaiguste päritavusest siiani suur osa leidmata (Eichler *et al.*, 2010; Manolio *et al.*, 2009). Selles vallas on näha jõudsat edasiminekut, näiteks inimese pikkusel, mis on samuti suure päriliku komponendiga (päritavus umbes 80%) komplekstunnus, suudeti 2008. aastaks kirjeldada 40 tolleks hetkeks leitud geneetilise variandi abil varieeruvusest ligikaudu 5 protsenti (Visscher, 2008). Aastaks 2010 oli leitud aga juba 180 pikkusega seostuvat lookust (Lango Allen *et al.*, 2010). Kasutades kõiki sagedasi variante, suudeti kirjeldada ligikaudu 45% pikkuse geneetilisest varieeruvusest ning spekulēriti teemadel, et suur osa seni kirjeldamata varieeruvusest võiks olla seostunud harvade variantidega (Yang *et al.*, 2010).

Ülegenoomsetes assotsiatsiooniuuringutes leitud geneetilised variandid on enamasti väikese efektsuurusega (Altmüller *et al.*, 2001; Manolio *et al.*, 2009), mis on viinud vaidlusele selle üle, kus võiks peituda komplekshaiguste peidetud pärilikkus (ingl k *missing heritability*) (Eichler *et al.*, 2010). Peidetud pärilikkuse olulisus seisneb selles, et märkimisväärne osa haiguse vastuvõtlikkuse individuaalsetest erinevustest tuleneb geneetilisest faktoritest ja selle geneetilise varieeruvuse mõistmine võib aidata kaasa haiguste paremale ennetamisele, diagnoosile ja ravile (Manolio *et al.*, 2009).

Peidetud pärilikkusele on välja pakutud mitmeid erinevaid seletusi, näiteks seni leidmata arvukad väikese efektsuurusega geneetilised variandid, mida on olemasolevate genotüüpiseerimiskiiptidega raske tuvastada; struktuuralsed variandid, mida on samuti seniste kiipidega raske leida; geen-geen interaktsioonide tuvastamise vähene võimsus ning ebapiisav keskkonnafaktorite arvesse võtmine sugulaste vahel (Manolio *et al.*, 2009). Ehkki konsensus neile seletustele puudub, loetakse üheks peidetud pärilikkuse aluspõhimõtteks harva variandi – sagedase haiguse hüpoteesi, mis ütleb, et geneetilise haiguse, mis on populatsioonis sage (levimus populatsioonis üle 1–5%), geneetilised põhjused ei pruugi peituda sagedaste geneetiliste variantide hulgas nagu väidab sagedase variandi – sagedase haiguse hüpotees. See-eest võivad aga haiguse põhjused hõlmata arvukaid riskialleele, mis kõik on populatsioonis haruldased ehk sagedast haigust võivad samaaegselt põhjustada mitmed erinevad harvad variandid erinevates geenides. (Gibson, 2012; Schork *et al.*, 2009). Harva variandi – sagedase haiguse hüpotees märgib, et sage polügeenne haigus võib peegeldada

mitmete harvade variantide konvergentsi samas geenis (alleelne heterogeensus) või mitmetes erinevates geenides (lookuse heterogeensus) (El-Fishawy, 2013).

Selle üle, kas peidetud pärilikkus võib peituda harvades variantides ($MAF < 0,5\%$), on palju spekuleeritud. Juhul kui MAF on väga väike, muutub assotsiatsiooni leidmise tõenäosus seniste genotüpiseerimiskiipidega väga väikeseks, lisaks ei pruugi nende efektiivsus olla piisav selleks, et need ilmneks klassikalise perekonna aheldatusanalüüsi käigus (Manolio *et al.*, 2009; McCarthy ja Hirschhorn, 2008). Võimalik on assotsiatsioone leida vaid juhul kui nende variantide efektiivsus on väga suur nagu näiteks monogeensete haiguste puhul (Manolio *et al.*, 2009).

Hetkel on peamine ning kõige tõhusam viis harvade variantide leidmiseks sekveneerimine, mis on suunatud kas kindlale regioonile või kogu genoomile. Harvade variantide identifitseerimine põhineb selliste genoomsete regioonide sekveneerimisel, mille lähedalt on varasemate ülegenoomsete assotsiatsiooniuringute käigus leitud tugevaid ja korduvaid sagedaste variantide assotsiatsioone, lisaks on sekveneeritud äärmuslike fenotüüpidega inimeste genoomist suuremaid fragmente (Manolio *et al.*, 2009). Juhul kui puuduvad varasemalt leitud assotsiatsioonid, mille ümbrust uurida, on võimalik sekveneerida ekstreemsete kvantitatiivsete tunnustega indiviidide kandidaatgeene, mille kaudu on võimalik tuvastada nii sagedasi kui ka harvu variante (Kotowski *et al.*, 2006; Manolio *et al.*, 2009). Alternatiivne võimalus küllaltki kallile sekveneerimisele on genotüpiseerimiskiipide kasutamine. Nende puhul on aga problemaatiliselt asjaolu, et need on eelkõige disainitud sagedaste variantide tarbeks ega suuda harvu variante piisavalt hästi tabada (Morris ja Zeggini, 2010).

Teiseks oluliseks peidetud pärilikkuse allikaks on keskkond. Geeni ja keskkonna interaktsioonide suurim mõju võib avalduda harvade alleelide puhul, mida on aga väga raske leida juhtumi kaupa teostatavates uuringutes. Heaks näiteks on uuringud, kus vaadatakse, kuidas keskkond mõjutab transkriптоomi ja metabooli. (Eichler *et al.*, 2010). Veel võiks peidetud pärilikkust otsida geneetilistest variatsioonidest, mis mõjutavad mittekodeerivate RNA-de ekspressiooni. Näiteks 2010. aastal ilmunud Nicoloso jt töös näidati, et SNP -d, mis assotsieeruvad rinnavähi riskiga, võivad olla osalised mikro-RNA geeniregulatsiooni muutmises (Nicoloso *et al.*, 2010).

2. Ekperimentaalosa

2.1. Töö eesmärk

Käesoleva magistritöö eesmärgiks on harvu geneetilisi variante analüüsides uurida polütsüstiliste munasarjade sündroomi geneetilist tagapõhja. Antud uuringu käigus teostasin harvade geneetiliste variantide ($MAF \leq 1\%$) nii üksiku markeri põhiseid kui ka geenipõhiseid analüüse.

2.2. Materjal ja meetoodika

2.2.1. Valim

Töös kasutasin kahte andmestikku: DNA proove, mis on kogutud prof Andres Salumetsa töögrupi poolt Tartu Ülikooli Kliinikumi Naistekliinikus ning lisaks Tartu Ülikooli Eesti Geenivaramu andmestikku.

Tartu Ülikooli Eesti Geenivaramu kohorti kuulub ligikaudu 52 000 täiskasvanut (vanuses ≥ 18), mis moodustab 5% Eesti täiskasvanute populatsioonist. Kõik proovid on kogutud vabatahtlikelt geenidoonoritelt kõikidest maakondadest, tänu millele on kaetud Eesti inimeste vanuseline, sooline ja geograafiline jaotumus. Biopanka kuuluvad ligikaudu 5,5% kõigist täiskasvanud Eesti naistest ning meesdoonorite hulk Biopangas moodustab umbkaudu 3,4% Eesti täisealistest meestest. (Leitsalu *et al.*, 2014).

Kvaliteedikontrolli läbinud valimisse kuulus 169 PCOS-i diagnoosiga naispatsienti. Kontrollgrupp koosnes 4437 indiviidist, kelle hulka kuulus lisaks naistele (64%) ka mehi (36%). Uuringus kasutati inimesi vanuses 18 kuni 69. Kontrollgrupi keskmiseks vanuseks oli 30,8 ja juhtude keskmine vanus oli 30,9. Kehamassiindeks oli vahemikus 14,5 kuni 58,4, keskmise väärtusega 23,6 kontrollidel ja 26,2 juhtudel. Valimisse kuulunud indiviidide iseloomustus on esitatud tabelis 3.

Tabel 3. Valimi iseloomustus. Tabelis on väljatoodud uuringus osalenute arv, naiste hulk juhtude ja kontrollide seas, valimi keskmine kehamassiindeks ja vanus.

	Juhud	Kontrollid
Indiviidide arv	169	4437
Naiste osakaal	100%	64%
Keskmine kehamassiindeks (std hälve)	26,2 (6,7)	23,6 (5,3)
Keskmine vanus (std hälve)	30,9 (8,5)	30,8 (7,4)

Tartu Ülikooli Kliinikumi Naistekliiniku proovide seas (n=110) oli *PCOS* fenotüüp defineeritud Rotterdami kriteeriumi järgi, kus *PCOS* diagnoositakse juhul, kui patsiendil esineb järgnevatest sümptomitest vähemalt kaks:

1. an- või oligoovulatsioon,
2. hüperandrogenism,
3. polütsüstilised munasarjad.

Tartu Ülikooli Eesti Geenivaramu andmestikust pärit juhtude (n=59) näol oli tegemist ise raporteeritud juhtudega, mille kasutamine on geneetilistes uuringutes aktsepteeritav (Day *et al.*, 2015). Kontrollgrupi liikmed valiti geenidoonorite hulgast juhuslikult ja sellesse kuulusid indiviidid, kellel teadaolevalt polnud *PCOS*-i diagnoositud.

Antud magistritöö raames läbiviidud uuringutesse kaastatud geenidoonorid on selleks andnud oma nõusoleku, samuti on uuring kooskõlastatud Tartu Ülikooli inimuuringute eetika-komiteega (luba 234/T-12).

2.2.2. Genotüpiseerimine ja andmete kvaliteedikontroll

Selleks, et uurida kodeerivate variantide panust *PCOS*-i patogeneesi, genotüpiseeriti Tartu Ülikooli Eesti Geenivaramu tuumiklaboris kommertsiaalse Illumina eksoomi kiibiga 169 juhtu ning 4437 kontrolli. Genotüpiseerimine viidi läbi Illumina Infinium CoreExome-24 v1.1 BeadChip kiibiga¹, mis võimaldab teostada suureskaalalisi genotüpiseerimisuuringuid. Nimetatud kiibi kogumarkerite arv on 551 839 ning ühele kiibile on võimalik kanda kuni 24

¹ http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_human_core_exome_beadchip.pdf

proovi. Lisaks on *RefSec*² geenide hulgast antud kiibile kantud 374 258 markerit. Genotüpiseerimise käigus järgiti tootjapoolset laboriprotokollit ja kasutati tootja poolt pakutavaid reaktiive.

Enne edasist analüüsimist ning haiguse võimalike põhjuslike variantide analüüside läbiviimist teostasin genotüpiseerimisandmetele nii indiviidi- kui ka markeripõhise kvaliteedikontrolli, milles võtsin arvesse järgnevalt väljatoodud näitajaid.

Indiviidipõhine kvaliteedikontroll:

- edukalt genotüpiseeritud markerite protsentuaalne osakaal indiviidil (ingl k *Call Rate*, *CR*)
- heterosügootsus
- *singleton SNP*-de jaotus proovi kohta
- genotüübijärgse soo mittevastavus fenotüübi soole
- indiviidide omavaheline sugulus

Markerite kvaliteedikontroll:

- edukalt genotüpiseeritud indiviidide protsentuaalne osakaal markeril
- Hardy-Weinbergi tasakaalustatus (ingl k *Hardy-Weinberg Equilibrium*, *HWE*)
- klastrite eraldusskoor
- *GenTrain* skoor

Uuringus analüüsisin vaid autosomaalseid ja X-kromosoomi. Kromosoomist Y, mitokondri DNA-st ning X- ja Y-kromosoomi jagatavast pseudoautosomaalsest regioonist (XY) pärinevaid variante kaasatud ei ole. Indiviidide kvaliteedikontrolli järgselt eemaldasid järgnevatest analüüsides need, kelle *CR* oli väiksem kui 98% ja kelle heterosügootsuse määr oli oluliselt suurem keskmisest heterosügootsusest (keskmine $+3 \times \text{std}$ hälve), lisaks eemaldasid analüüsides *singleton*-id, ehk geneetilised markerid, mis märgivad regioonis vaid iseend ega seostu ülejäänud regiooniga nagu seda teevad näiteks *tag-SNP*-d (Ke *et al.*, 2008). Juhul kui genotüpiseerimise järgne sugu ei vastanud fenotüübijärgsele soole, eemaldasid indiviid(id) uuringust. Selleks, et teha kindlaks, kas valimis leidub sugulasi, hindasin iga indiviidide paari jaoks, kui suur on genoomis nende regioonide osakaal, mis on päritud ühiselt eellaselt (ingl k *Identity by Descent*, *IBD*). Sugulussuhteid omavateks on loetud indiviidide paarid, kelle *pi_hat* skoor oli $>0,3$. Sugulaste eemaldamisel lähtusin sellest, et eemaldatud

² <http://www.ncbi.nlm.nih.gov/refseq/>

saaksid indiviidid, kes on suguluses rohkema arvu teiste indiviididega, juht-kontroll paaride vahel eemaldas eelistatavalt kontrolli. Sugulussuhteid omavate juht-juht ja kontroll-kontroll paaride korral jätsin uuringust välja madalaima *CR*-iga indiviidi.

Lisaks eemaldas edasistest uuringutest geneetilised variandid, mille *CR* jäi alla 95%, *HWE* p -väärtus $<1 \times 10^{-4}$. Täiendavalt jäid uuringutest välja sellised markerid, mille genotüpiseerimisel ei paigutunud need kindlatesse klastritesse (eraldusskoor $<0,4$) ning markerid, mille *GenTrain* skoor oli $<0,6$. *GenTrain* skoor on kasutusel eesmärgiga eemaldada sellised markerid, mille puhul ei suudeta genotüüpe üksteisest eristada. Täiendavalt eemaldas uuringust halvasti genotüpiseeritavad *HLA* (inimese leukotsüüdi antigeen, ingl *Human Leukocyte Antigen*) regioonid, inversioone sisaldavad piirkonnad ja markerid, mille *MAF* $\geq 1\%$. Kvaliteedikontroll teostasin kasutades programmi *PLINK* v1.07 (Purcell *et al.*, 2007).

Indiviide, kes läbisid kvaliteedikontrolli, oli 4606 (169 juhtu ja 4437 kontrolli). Lisaks kaasasin uuringusse kvaliteedikontrolli läbinud 197 892 markerit.

2.2.3. Andmete ettevalmistus analüüsiks

Peamised genotüübi variantide määramise meetodid on disainitud sagedasete variantide tarbeks. See toob kaasa olukorra, kus harvade variantide puhul on hästi esindatud vaid sagedaste alleelide homosügootide klaster, mis limiteerib klasterdamise algoritmide tõhusust (Goldstein *et al.*, 2012). Illumina poolt pakutav geneetiliste variantide määramine ei suuda samuti niivõrd hästi harvu variante määrata, mistõttu on käesolevas magistritöös täiendavaks variantide määramiseks kasutatud *zCall*-i.

Analüüsieelselt teostasin *VCF* (ingl *k Variant Call Format*) annotatsiooni, kasutades selleks *encode* V14 andmebaasi, mida pakub *EPACTS*. *VCF* ise on failiformaat, mis sisaldab endas andmeid genoomsete positsioonide kohta. Annotatsioon on oluline üksiku nukleotiidi variantide, insertioonide ja deletsioonide interpreteerimiseks. Annoteerimine võimaldab uurida nende muutuste funktsionaalseid tagajärgi geenidele, identifitseerida geneetilisi variante konserveerunud regioonides ning 1000 Genoomi Projektis väljatoodud geneetilistes variantides (Wang *et al.*, 2010) .

2.2.4. Geneetiline stratifikatsioon

Geneetiline stratifikatsioon on süstemaatiline alleelisageduste erinevus populatsiooni sees või populatsioonide vahel, mis tuleneb erinevast geneetilisest taustast (Kiezun *et al.*, 2012). Populatsioonistratifikatsioon võib assotsiatsiooniuuringutes tekitada valepositiivseid tulemusi. Lisaks võivad populatsioonistratifikatsiooni tõttu ilmned assotsiatsioonid mitte uuritava tunnuse ja geneetilise variandi vahel, vaid assotsiatsioonid, mis eristavad näiteks üht populatsiooni teisest juhul, kui nende populatsioonide proportsioonid on juhtude ja kontrollide vahel erinevad (Abraham ja Inouye, 2014). Seda seetõttu, et erinevatel populatsioonidel võib olla erinev alleelisagedus tulenevalt nende erinevast demograafilisest ajaloost (Kiezun *et al.*, 2012). Uuringus võimaliku populatsioonistratifikatsiooni eemaldamiseks saab kasutada näiteks peakomponentanalüüsi. Peakomponentide väljaarvutamise aluseks on varasem ühiste alleelide osakaalu leidmine genoomis (ingl k *Identity by State, IBS*). *IBS* on termin, mida geneetikas kasutatakse kahe identse alleeli või kahe identse DNA järjestuse kirjeldamiseks (Stevens *et al.*, 2011). Käesolevas magistritöös on *IBS* leitud *PLINK* tarkvara kasutades (Purcell *et al.*, 2007). Selle leidmise järgselt oli võimalik välja arvutada peakomponendid eesmärgiga identifitseerida etnilisest päritolust tulenevad sugulussuhted ja hinnata kohordisest populatsiooni stratifikatsiooni. Peakomponendid leidsin samuti *PLINK* tarkvara abil, milles kasutasin lisaks multidimensionaalse hindamise (ingl k *Multidimensional Scaling, MDS*) analüüsi, mis aitab detekteerida erinevusi populatsioonide vahel ja kohandada populatsiooni nende tunnuste suhtes, mis võivad stratifikatsiooni põhjustada (Tzeng *et al.*, 2008). Tulemuste visualiseerimiseks ja jooniste genereerimiseks kasutasin R tarkvara³. Saadud peakomponentidest esimest nelja kasutasin hilisema statistilise analüüsi käigus kovariaatidena populatsiooni stratifikatsiooni arvesse võtmiseks.

2.2.5. Statistiline analüüs

Töö käigus viisin läbi nii harva variandi analüüsi ühe markeri kaupa kui ka geenipõhised analüüsid. Analüüside läbiviimiseks kasutasin programmi *EPACTS* v.3.2.4. Kovariaatidena võtsin analüüsides arvesse kehamassiindeksit, vanust ning nelja esimest peakomponenti. Autosomaalsete variantide analüüsimisel olid kontrollgruppi kaasatud nii naised kui mehed,

³ <https://www.r-project.org/>

kuid võrdlusena teostasin analüüsi, milles kasutasin vaid naiskontrolle. X-kromosoomi analüüsides kasutasin samuti kontrollidena vaid naisi.

Markerikaupa analüüsi viisin läbi, kasutades *EPACTS* tarkvaras olevat *EMMAX* testi⁴ ning kokku kasutasin üksiku markeri kaupa läbiviidud analüüsides 197 892 üksikut markerit.

Geenikaupa analüüsimisel kasutasin *EPACTS* programmis *SKAT-O* testi⁵. *SKAT* test lubab kasutada analüüsides kovariaate ning võimaldab arvutada geenide mõju olukorras, kus osad geenid on harvad variandid suurendavad ja teised vähendavad haigusrisiki (Ionita-Laza *et al.*, 2013). Geenipõhised analüüsid teostasin kahe erineva mudeli alusel: (1) kasutades geenides olevaid mittesünonüümseid variante ja (2) kasutades geenides olevaid funktsioonikaupa variante (ingl k *Loss-of-Function*, *LOF*). Mõlemal juhul kaasasin uuringusse sellised variandid, mille harvema alleeli sagedus <1%. Käesolevas analüüsis on geeni *LOF* variandid defineeritud kui *nonsense*, splaiss-saidi ning stopp- ja startkoodonis esinevad mutatsioonid. Mittesünonüümsete variantide analüüsis kasutasin kokku 16 480 geeni (geenide arv, mis sisaldas vähemalt ühte mittesünonüümset varianti), *LOF* variantide korral 4299 geeni (geenide arv, mis sisaldas vähemalt ühte *LOF* varianti).

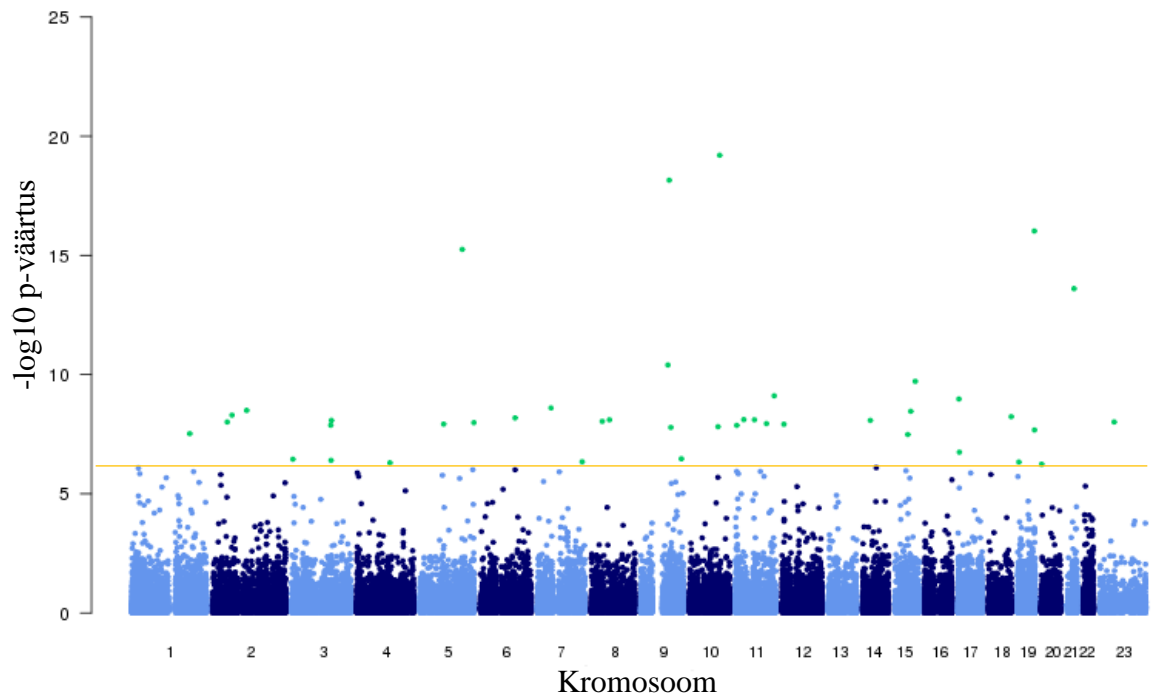
2.3. Tulemused

Käesoleva töö raames otsisin Eesti populatsioonist pärit *PCOS*-i diagnoosiga indiviidide baasil harvu geneetilisi variante, mis võiksid uurimise all oleva haigusega assotsieeruda. Varasemalt olen Eesti populatsiooni baasil teostanud ülegenoomse assotsiatsioonianalüüsi sagedaste markeritega. Uurimistöös teostasin harvade variantide analüüsi nii üksiku markeri kui ka geeni kaupa. Üksiku markeri kaupa analüüsides (geneetilised variandid sagedusega alla 1%) kasutasin kokku 65 789 mitte-monomorfset markerit, mille keskmine harvema alleeli sagedus oli 0,01%.

Üksiku markeri kaupa analüüsil ilmnis uuringust 43 geneetilist varianti, mille seos *PCOS*-iga oli statistiliselt oluline peale mitmese testimise korrektsiooni ($0,05/65\,789 = 7,6 \times 10^{-7}$). Leitud variantidest 39 olid mittesünonüümseid, sünonüümseid variante oli kaks (geenides *CHD7* ja *EGF*), üks variantidest asus geenidevahelisel alal (ingl k *intergenic*) ning üks geeni *PRICLE3* intronis. Üksiku markeri põhise analüüsi tulemused on esitatud tabelis 2 ning täiendavalt visualiseeritud Manhattan joonise abil joonisel 4.

⁴ <http://genome.sph.umich.edu/wiki/EMMAX>

⁵ <http://genome.sph.umich.edu/wiki/EPACTS>



Joonis 4. Üksiku markeri analüüsi tulemused Manhattan joonisel. Ülegenoomselt statistilise olulisuse nivoo on $p < 7,6 \times 10^{-7}$ (märgitud kollase joonega), millest madalamad olid 43 markeri p-väärtused. X-teljel on toodud assotsieerunud markerite geoomsed positsioonil ja Y-teljel p-väärtuse negatiivne kümnendlogaritm.

Tabel 4. Harvade variantide üksiku markeri kaupa läbiviidud analüüsi tulemused. Tabelis on esitletud statistiliselt olulise p-väärtusega ($<7,6 \times 10^{-7}$) markerid. **OR* märgib markerite šansside suhet (ingl k *Odds Ratio*) ning ***CI* (ingl k *Confidence Interval*) selle 95%-list usaldusintervalli. P-väärtus näitab, kui tugevalt kaks tunnust omavahel seotud.

Markeri ID	Positsioon	Geen	Efektalleel/ teine alleel	MAF	OR* (95% CI**)	P-väärtus
exm1283088	17:5036249	<i>USP6</i>	G/T	0,001	1,92 (1,71-2,16)	$1,64 \times 10^{-28}$
exm849225	10:102054300	<i>PKD2L1</i>	C/T	0,0005	2,14 (1,82-2,52)	$6,27 \times 10^{-20}$
exm760920	9:94794758	<i>SPTLC1</i>	C/T	0,0003	2,6 (2,11-3,21)	$7,01 \times 10^{-19}$
exm1509699	19:56156032	<i>ZNF581</i>	A/C	0,0006	1,88 (1,62-2,18)	$9,48 \times 10^{-17}$
exm488613	5:140762533	<i>PCDHGA7</i>	C/A	0,0023	1,38 (1,28-1,49)	$5,58 \times 10^{-16}$
exm1568790	21:37442658	<i>CBR1</i>	G/A	0,0004	2,04 (1,70-2,45)	$2,48 \times 10^{-14}$
exm759493	9:90500747	<i>C9orf79</i>	G/A	0,0033	1,25 (1,17-1,33)	$3,96 \times 10^{-11}$
exm1188448	15:90216467	<i>PLIN1</i>	G/A	0,0009	1,49 (1,32-1,68)	$1,92 \times 10^{-10}$
exm969699	11:128850539	<i>ARHGAP32</i>	T/C	0,0006	1,6 (1,38-1,86)	$7,77 \times 10^{-10}$
exm1277716	17:3324574	<i>OR3A3</i>	G/A	0,0006	1,59 (1,37-1,85)	$1,06 \times 10^{-09}$
exm619209	7:45726156	<i>ADCY1</i>	G/A	0,0003	1,9 (1,54-2,35)	$2,52 \times 10^{-09}$
exm221956	2:113676187	<i>IL1F7</i>	C/A	0,0022	1,26 (1,17-1,36)	$3,19 \times 10^{-09}$
exm1177316	15:75146366	<i>SCAMP2</i>	G/T	0,0003	1,89 (1,53-2,34)	$3,47 \times 10^{-09}$
exm197788	2:64780206	<i>AFTPH</i>	A/G	0,0003	1,88 (1,52-2,32)	$5,02 \times 10^{-09}$
exm1394139	18:76753726	<i>SALL3</i>	G/A	0,0003	1,88 (1,52-2,32)	$5,81 \times 10^{-09}$
exm573403	6:116574371	<i>TSPYL4</i>	C/T	0,0003	1,87 (1,51-2,31)	$6,59 \times 10^{-09}$
exm897432	11:27389898	<i>LGR4</i>	G/T	0,0003	1,86 (1,51-2,30)	$7,74 \times 10^{-09}$

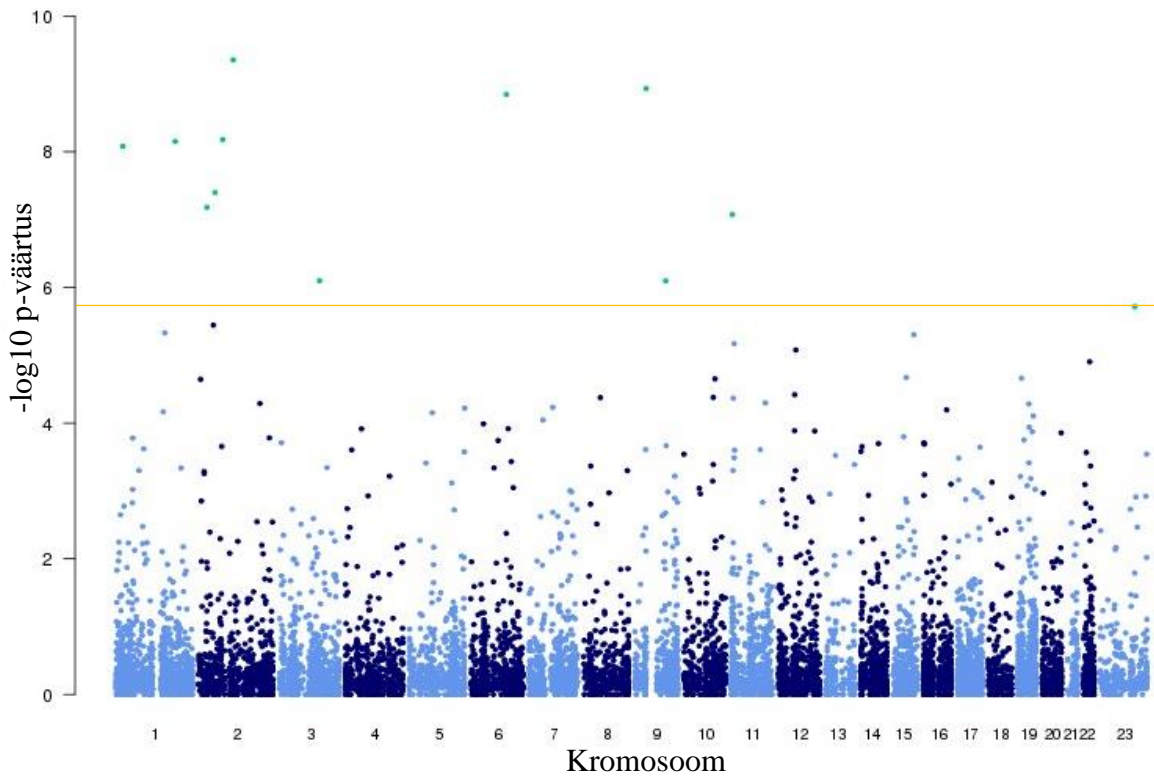
exm920300	11:62910891	<i>SLC22A24</i>	C/A	0,0003	1,86 (1,51-2,30)	$7,85 \times 10^{-09}$
exm2258913	8:61777937	<i>CHD7</i>	C/T	0,0003	1,87 (1,51-2,31)	$7,86 \times 10^{-09}$
exm351958	3:133169282	<i>BFSP2</i>	G/A	0,0003	1,86 (1,51-2,30)	$8,32 \times 10^{-09}$
exm1098829	14:45496534	<i>FAM179B</i>	A/G	0,0003	1,86 (1,51-2,30)	$8,38 \times 10^{-09}$
exm695493	8:37690601	<i>GPR124</i>	G/A	0,0003	1,86 (1,51-2,30)	$9,16 \times 10^{-09}$
exm2262822	X:49041906	intron: <i>PRICKLE3</i>	G/A	0,0003	1,86 (1,51-2,30)	$9,73 \times 10^{-09}$
exm193071	2:48914951	<i>LHCGR</i>	T/C	0,0003	1,86 (1,51-2,30)	$9,77 \times 10^{-09}$
exm510832	5:179743819	<i>GFPT2</i>	C/T	0,0003	1,86 (1,51-2,30)	$1,04 \times 10^{-08}$
exm951838	11:103043949	<i>DYNC2H1</i>	A/G	0,0003	1,85 (1,50-2,29)	$1,14 \times 10^{-08}$
exm463845	5:79034727	<i>CMYA5</i>	C/T	0,0003	1,85 (1,50-2,29)	$1,19 \times 10^{-08}$
exm977282	12:6458149	<i>SCNNIA</i>	G/A	0,0003	1,85 (1,50-2,29)	$1,21 \times 10^{-08}$
exm351068	3:131415435	<i>CPNE4</i>	T/C	0,0003	1,85 (1,50-2,28)	$1,33 \times 10^{-08}$
exm880340	11:3849236	<i>RHOG</i>	C/T	0,0003	1,85 (1,50-2,28)	$1,35 \times 10^{-08}$
exm843903	10:96535173	<i>CYP2C19</i>	T/C	0,0025	1,24 (1,15-1,33)	$1,53 \times 10^{-08}$
exm765101	9:100075554	<i>C9orf174</i>	A/G	0,0018	1,29 (1,18-1,41)	$1,64 \times 10^{-08}$
exm1511362	19:56733294	<i>ZSCAN5A</i>	C/T	0,0013	1,35 (1,22-1,51)	$2,11 \times 10^{-08}$
exm132950	1:193044958	<i>TROVE2</i>	A/G	0,0007	1,48 (1,29-1,70)	$2,99 \times 10^{-08}$
exm1169722	15:65350832	<i>RASL12</i>	C/T	0,0007	1,48 (1,29-1,70)	$3,27 \times 10^{-08}$
exm1892945	17:4927205	<i>KIF1C</i>	G/A	0,0036	1,18 (1,11-1,26)	$1,80 \times 10^{-07}$
exm790905	9:135140020	<i>SETX</i>	A/G	0,0014	1,3 (1,18-1,44)	$3,39 \times 10^{-07}$

exm286753	3:4842276	<i>ITPR1</i>	G/A	0,0008	1,4 (1,23-1,59)	$3,56 \times 10^{-07}$
exm351260	3:132196983	<i>DNAJC13</i>	G/A	0,0008	1,4 (1,23-1,59)	$3,97 \times 10^{-07}$
exm670140	7:149522165	<i>Intergenic</i>	G/A	0,003	1,2 (1,12-1,28)	$4,55 \times 10^{-07}$
exm1409414	19:4501243	<i>AC011498.1</i>	G/C	0,0004	1,6 (1,33-1,92)	$4,68 \times 10^{-07}$
exm2256278	4:110884420	<i>EGF</i>	C/T	0,0004	1,6 (1,33-1,92)	$5,05 \times 10^{-07}$
exm1519012	20:1459244	<i>SIRPB2</i>	C/A	0,0004	1,59 (1,33-1,91)	$5,72 \times 10^{-07}$

Geenipõhisel analüüsil kasutasin kahte mudelit, millest üks oli suunatud geenides olevatele mittesünonüümsetele variantidele ning teine geenides olevatele funktsioonikaoga variantidele. Filtreerisin tulemusi nii, et uuritavates geenides oleks vähemalt kaks analüüsitud markerit. Mittesünonüümsete variantide korral analüüsiti kokku 14 056 geeni, keskmise variantide arvuga 5,2 geeni kohta (minimaalne 2 ja maksimaalne 272). Geene, milles oli antud variante ainult üks, ei kaasatud, sest sellised variandid on kasutusel juba üksiku markeri analüüsi käigus. Peale mitmese testimise korrektsiooni oli statistilise olulisuse nivooks $0,05/14\ 056 = 3,56 \times 10^{-6}$. Leitud nivoost madalama p-väärtusega on kokku kaksteist erinevat varianti kaheteistkümnes erinevas geenis, mis on välja toodud tabelis 5. Nende 12 geeni seas jäid mittesünonüümsete variantide arvud vahemikku kaks kuni viis. Madalaima väärtusega oli geneetiline marker geenis *IL1F7*, mis on tuntud ka kui *IL37* (interleukiin 37).

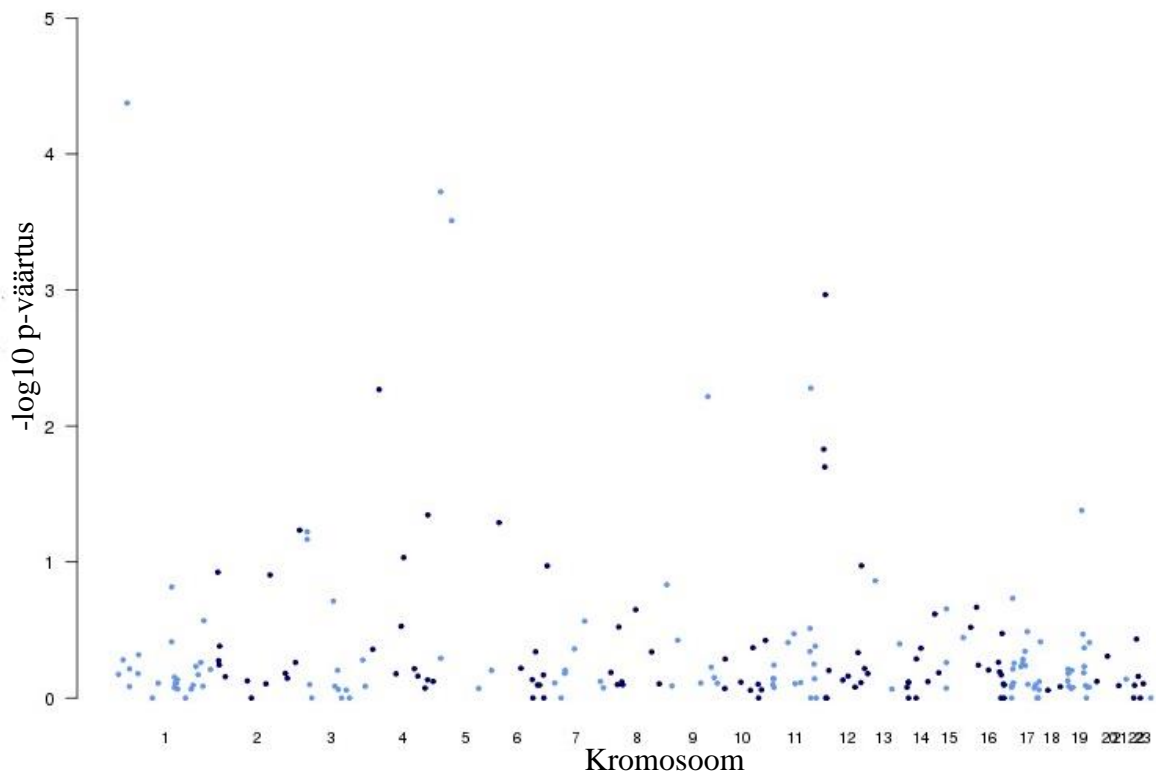
Tabel 5. Geenipõhise analüüsi mittesünonüümsete (MS) variantide mudeli puhul statistiliselt olulise p-väärtusega geenide nimistu. *Singleton*-id on geneetilised markerid, mis märgivad regioonis vaid iseend ega seostu ülejäänud regiooniga. P-väärtus märgib kahe tunnuse omavahelise seose tugevust.

Positsioon	Geeni nimi	MS variantide arv geenis	<i>Singleton</i> -ide arv	p-väärtus
2:113670627-113676381	<i>IL1F7</i>	2	0	$4,4 \times 10^{-10}$
9:37441383-37442329	<i>ZBTB5</i>	2	1	$1,2 \times 10^{-9}$
6:116720448-116758473	<i>DSE</i>	3	2	$1,4 \times 10^{-9}$
2:79253283-79255377	<i>REG3G</i>	2	2	$6,6 \times 10^{-9}$
1:193038254-193053716	<i>TROVE2</i>	4	0	$7,0 \times 10^{-9}$
1:21924217-21943804	<i>RAP1GAP</i>	5	2	$8,3 \times 10^{-9}$
2:54482136-54483137	<i>TSPYL6</i>	3	1	$4,0 \times 10^{-8}$
2:27886669-27908127	<i>SLC4A1AP</i>	3	2	$6,6 \times 10^{-8}$
11:3659993-366124	<i>ART5</i>	3	1	$8,4 \times 10^{-8}$
3:133119363-133191385	<i>BFSP2</i>	2	0	$7,9 \times 10^{-7}$
9:101570283-101602372	<i>GALNT12</i>	2	0	$8,0 \times 10^{-7}$
X:114357331-114414073	<i>LRCH2</i>	2	1	$1,9 \times 10^{-6}$



Joonis 5. Geenipõhise analüüsi mittesünonüümsete markerite tulemused Manhattan joonisel. Ülegenoomselt statistilise olulisuse nivoo on $p < 3,56 \times 10^{-6}$ (märgitud kollase joonega), millest madalamad olid 12 markeri p-väärtused. X-teljel on toodud assotsieerunud markerite genoomsed positsioonid ja Y-teljel p-väärtuse negatiivne kümnendlogaritm.

Funktsioonikaoga harvade markeritega läbiviidud geenipõhises analüüsis kasutasin kokku 1716 geeni. Funktsioonikaoga geneetilisi variante oli uuritavates geenides vahemikus kahest viieni (geene, kus oli vaid üks funktsioonikaoga variant, ei analüüsitud). Mitmese testimise korrigeerimise järgselt ($p < 2,91 \times 10^{-6}$) ei esinenud ühtegi geeni, mis oleksid leitud statistilise olulisuse piirist väiksema p-väärtusega. Madalaima p-väärtusega marker funktsioonikaoga harvade variantidega analüüsis asub esimeses kromosoomis geen *CNKSRI*, mille p-väärtus on $4,23 \times 10^{-5}$ ja selles asub kokku 3 erinevat funktsioonikaoga varianti.



Joonis 6. Geenipõhise analüüsi funktsioonikaoga markerite tulemused Manhattan joonisel. Ülegenoomselt statistilise olulisuse nivoo on $p < 2,91 \times 10^{-6}$, millest madalamat tulemust ei omanud ükski geneetiline marker. X-teljel on toodud assotsieerunud markerite genoomsed positsioonid ja Y-teljel p-väärtuse negatiivne kümnendlogaritm.

Esines kolm geeni, mille variandid omasid peale mitmese testimise korrektsiooni nii üksiku markeri kui ka mittesünonüümsete variantide geenipõhise analüüsi tulemuste seas statistiliselt olulist p-väärtust. Need geenid on *TROVE2*, *BFSP2* ja *ILIF7*. Neist esimese p-väärtused on üksiku markeri ja mittesünonüümsete variantide analüüsis vastavalt $2,99 \times 10^{-8}$ ja $7,0 \times 10^{-9}$. *BFSP2* p-väärtus on üksiku markeri analüüsis $8,32 \times 10^{-9}$ ja mittesünonüümsete variantide uuringus $7,9 \times 10^{-7}$.

2.4. Arutelu

PCOS on reproduktiivses eas olevatel naistel esinev komplekshaigus, mis on oluliseks viljatuse põhjustajaks. Kaksikute uuringud on näidanud, et *PCOS*-il on tugev geneetiline komponent (Vink *et al.*, 2006), mistõttu on haiguse tagapõhja väljaselgitamiseks läbi viidud mitmeid ülegenoomseid assotsiatsioonianalüüse, mille käigus uuritakse sündroomi geneetikat sagedaste variantide seas. Eelnevate uurimistööde käigus on esile kerkinud erinevaid genee, kuid *PCOS*-i täpne geneetiline taust on, sarnaselt teistele polügeensetele haigustele, tänaseni

lõplikult välja selgitamata (Dunaif, 2016). Antud magistritöö raames teostas in Eesti populatsiooni kasutades analüüsi, leidmaks harvu geneetilisi variante, mis võivad mõjutada PCOS-i teket. Kasutatud valim koosnes 169 PCOS-i juhust ja 4437 kontrollist. Töö tulemusena leiti üksiku geneetilise variandi ja geenipõhise analüüsi tulemusena 55 erinevat geeni, mis Eesti populatsioonis omasid PCOS-iga statistiliselt olulist seost.

Ühekaupa geneetiliste variantide analüüsist ilmnis kokku 43 erinevat markerit, millest kahte on varasemalt PCOS-iga seostatud, on *PLIN1* ja *LHCGR* (Kawai *et al.*, 2010; Louwers *et al.*, 2013; Thathapudi *et al.*, 2015). Geenid *ILIF7*, *TROVE2* ja *PKD2L1* on leidudest samuti ühed kõige rohkem huvipakkuvad, sest nendega seoses on kirjeldatud tunnuseid, mis on iseloomulikud muuhulgas PCOS-ile.

Perilipiinid on hormonaalselt reguleeritavad fosfovalgud, mida leidub adipotsüütide rasvaladustamise piiskades ning neerupealsete, testisete ja munasarjade teatud rakkudes (Greenberg *et al.*, 1991). Perilipiini geeni (*PLIN1* või ka *PLIN*) variante on varasemates uuringutes seostatud inimese rasvumise ja lipiidide tasemega, mille põhjal püstitati hüpotees, et *PLIN* võiks olla seotud PCOS fenotüübiga. Selle uurimiseks viisid Kawai jt 2009. aastal läbi uuringu, milles analüüsiti seitset SNP-i, mis asusid *PLIN* geenist 2 kb üles- või allavoolu ja mille *MAF* oli >0,01. Nende uuring kinnitas varasemate uuringute tulemusi, mis on näidanud, et *PLIN* teatud variandid on seotud ainevahetuse tasakaalutusega, mis on omakorda seotud ülekaalu, insuliin- ja glükoosiresistentsusega. Samas oli kõikide uuritud variantide efektisuurus Kawai jt poolt teostatud uuringus oli PCOS patsientide kui kontrollide vahel ligikaudu samaväärne, mis viitab sellele, et seotus PCOS-iga puudub (Kawai *et al.*, 2010).

2008. aastal viisid Qi jt naiste seas läbi juht-kontroll uuringu, milles uurisid, kas *PLIN* geeni variandid on seotud ülekaaluga, mis on üks tüüp 2 diabeedi suurimatest riskifaktoritest. Tulemused viitasid sellele, et ülekaal, eriti kõhupiirkonna rasvumine, mõjutab naistel oluliselt assotsiatsioone *PLIN* geneetiliste variantide ja tüüp 2 diabeedi vahel (Qi *et al.*, 2008).

Luteniseeriva hormooni/kooriongonadotropiini retseptorgeeni *LHCGR* seos PCOS-iga ilmnis esmalt 2011. aastal, mil han-hiinlaste seas viidi läbi ülegenoomne assotsiatsiooni-uuring (Chen *et al.*, 2011), muuhulgas on leitud kinnitanud ka kaukaasia päritolu inimeste seas läbi viidud replikatsioonianalüüsid (Mutharasan *et al.*, 2013). *LHCGR* puhul on tegemist geeniga, mis on olulisel kohal imetajate follikulogeneesis, mispärast on seda loetud ka üheks tõenäoliseks PCOS-i kandidaatgeeniks (Thathapudi *et al.*, 2015; Wang *et al.*, 2014; Yung *et*

al., 2014). *LHCGR*-i ekspresseeritakse munasarja teekarakkudes ja munandi Leydigi rakkudes, aga ka rasvkoos ning see kodeerib nii LH kui ka koorionigonadotropiini retseptoreid (Thathapudi *et al.*, 2015). Arvatakse, et *PCOS*-i ja *LHCGR*-i vaheline seos on tingitud *LHCGR*-i hüpomütatsioonist, mis viib *PCOS*-i väljakujunemiseni (Wang *et al.*, 2014).

Üksiku geneetilise variandi kaupa läbiviidud uuringus oli huvipakkuv veel geneetiline marker, mis asus geenis *IL1F7*, mis on tuntud ka kui *IL37* (interleukiin 37) ja mille ekspressiooni ülesreguleerimist on näidatud tuberkuloosi patsientidel (Huang *et al.*, 2015). Sama geen omas muuhulgas mittesünonüümsete variantide geenipõhises analüüsis kõige madalamat p-väärtust, mis annab alust arvata, et *IL1F7* võiks kirjeldada osa *PCOS*-i geneetilisest komponendist. *PCOS*-i uurides on *IL37* puhul huvitav veel see, et seda on seostatud kaitsega ülekaalulisusest tingitud põletiku ning insuliinresistentsuse vastu ja see on võimalik sihtmärk, kui soovitakse ravida ülekaalust tulenevat insuliinresistentsust ning tüüp 2 diabeeti (Ballak *et al.*, 2014). Ülekaalulisuse korral hakkavad adipotsüüdid tootma tsütokiine ja kemokiine, mis võivad omakorda aktiveerida makrofaage ja teisi immuunrakke, tuues sellega kaasa kroonilise põletiku. Jätkuv põletikuline seisund, mida iseloomustab proinflammatoorsete tsütokiinide tootmine, segab otseselt insuliini signaalrada, mis lõpuks võib viia insuliinresistentsuseni ning hiljem tüüp 2 diabeedi kujunemiseni (Ballak *et al.*, 2014; Donath ja Shoelson, 2011; Jager *et al.*, 2007). Ballack jt viisid 2014. aastal läbi uuringu (kasutades nii hiiri kui inimesi), kus näitasid, et *IL37* neutraliseerib ülekaalust tingitud rasvkoe põletikku ning kaitseb insuliinresistentsuse väljakujunemise eest. Hiirtel, kellel ekspresseeriti inimese *IL37* geeni, täheldati kaalutõusu peatumist, väiksemat adipotsüütide suurust ja massi, vähem rasvkoe põletikku ning insuliinresistentsust võrrelduna samasugust dieeti järgivate metsik-tüüpi hiirtega (Ballak *et al.*, 2014). Ehkki *IL37*-t ei ole hetkeseisuga *PCOS*-iga otseselt seostatud, on see siiski huvitavaks kandidaatgeeniks, sest üsna mitmed *PCOS*-ile iseloomulikud tunnused on assotsieerunud *IL37*-iga (insuliinresistentsus, tüüp 2 diabeet).

Geeni *CBRI*, mis samuti on üksiku variandi analüüsis statistiliselt olulise väärtusega, ei ole kirjanduse andmetel varasemalt *PCOS*-iga seostatud, küll on aga näidatud, et see võib olla seotud munasarjavähiga (Osawa *et al.*, 2015). Karbonüülreduktaas 1 (*CBRI*) on *NADPH*-sõltuv oksidoreduktaas, mis aitab vähendada aldehüüdide ja ketoonide hulka (Gonzalez-Covarrubias *et al.*, 2007). Vähenenud *CBRI* ekspressiooni taset seostatakse vähi progresseerumisega. 2015. aastal uurisid Osawa jt *CBRI* ekspressiooni ja pahaloomulise

munasarjavähi vahelist seost. Katse teostati hiirmudelil ning tulemustes näidati, et hiirtel, kelle *CBRI* üle ekspresseeriti, vähenes rakujagunemine kontrollhiirtega võrrelduna märgatavalt. Samuti suurenes hiirtel, kellel *CBRI* ekspressiooni pärsiti, kontrollhiirtega võrreldes rakujagunemine tunduvalt. (Osawa *et al.*, 2015).

Veel üks üksikute geneetiliste variantide analüüsist ilmnenuv huvipakkuv geen seoses *PCOS*-iga on *PKD2L1* (tuntud ka kui *PKDL* või *PKD2L*), mille mutatsioonid põhjustavad polütsüstilist neeruhaigust (Basora *et al.*, 2002; Guo *et al.*, 2000; Veldhuisen *et al.*, 1999). *PKD2L1* kodeerib polütsüstiini perekonna liikmeid ning just mutatsioonid polütsüstiin-2 ja polütsüstiin-1, mida kodeerib *PKD1*, põhjustavad inimesel autosomaalset dominantset polütsüstilist neeruhaigust, mis viib progresseeruva neerupuudulikkuseni (Guo *et al.*, 2000). Siiski on nimetatud valkude täpne funktsioon selgusetu (Basora *et al.*, 2002; Veldhuisen *et al.*, 1999). Huvitav oleks teada nende funktsioone, sest kõne all olev geen on seotud tsüstide tekitamisega neerudes ning *PCOS*-i puhul on üheks haiguse sümptomiks tsüstid munasarjas. On võimalik spekuloida teemal, et nende kahe haiguse puhul võiks mingi osa tsüstide tekkest toimude sarnaste mehhanismide läbi.

Järgmine geen, mis kerkis esile nii üksiku markeri kui mittesünonüümsete harvade variantide analüüsil on *TROVE2*, mis on tuntud ka kui *RO60* ja mida on seostatud Sjögreni sündroomiga (Lindop *et al.*, 2013; Reed *et al.*, 2010), kuid millel ei ole *PCOS*-i ega selle sümptomitega tänase seisuga seost leitud. Ülejäänud ühteist geeni, mis geeni kaupa tehtud analüüsid *PCOS*-iga statistiliselt olulises suhtes olid, ei ole kirjanduse andmetel varasemalt selle haigusega seostatud. Küll aga on mitmeid neist geenidest seostatud erinevate vähiliikidega. Näiteks on vähiga leitud seos *ZBTB5*, *DSE* ja *RAP1GAP* geenidel (Dong *et al.*, 2012; Koh *et al.*, 2009; Qiu *et al.*, 2012; Thelin *et al.*, 2012; Tsygankova *et al.*, 2013).

Geen, mis täiendavalt geenidele *TROVE2* ja *IL1F7* mitmes analüüsis oluline oli, on *BFSP2*. Nimetatud geeni on tugevalt seostatud varajases nooruses tekkiva katarakti ehk hallkaega, mille võivad kaasa tuua missenssmutatsioonid *BFSP2* geenis (Jakobs *et al.*, 2000; Ma *et al.*, 2008; Zhang *et al.*, 2006).

Funktsioonikaoga harvade variantide analüüsis ei ilmnenuv ühtegi statistiliselt olulise tähtsusega geeni. Madalaim p-väärtus selles analüüsis oli geenil *CNKSRI*, mis on kirjanduses tuntud veel kui *CNK* või *CNK1*. Nimetatud geeni osaleb insuliini signaalrajas, mis teeb sellest huvitava *PCOS*-i kandidaatgeeni. On näidatud, et *CNK1/cytohesin* interaktsioonid

soodustavad membraanseoselise *Arf GTP*Paasi signalisatsiooni plasmast fosfatidülinositol 4-fosfaat 5-kinaasile (*PIP5K*), et genereerida *PIP2*-rikas mikrokeskkond, mis on hädavajalik insuliini retseptori substraati 1 (*IRS1*) värbamiseks membraani ja signaali edastamiseks *PI3K/AKT* kaskaadi. (Lim *et al.*, 2010). *PI3K/AKT* signalisatsioonirada on seostatud ka vähi ning tüüp II diabeediga (Mackenzie ja Elliott, 2014).

Käesoleva magistritöö raames valminud harvade variantide analüüs polütsüstiliste munasarjade sündroomiga seotud geneetiliste variantide tuvastamiseks on tehtud väikese andmestiku põhjal, mistõttu on oluline saadud tulemusi replitseerida teistes kohortides ja suurema andmestikuga. Samas on töös kasutatud andmed osaks *PCOS* kontsortsiumi suuremahulisest meta-analüüsist, mille käigus loodetavasti selgub suurem osa haiguse geneetilisest komponendist.

Kokkuvõte

Polütsüstiliste munasarjade sündroom on üle maailma naiste seas laialt levinud komplekshaigus, mis võib kaasa tuua viljatust ja suurenenud riski erinevatele kaasuvatele haigustele nagu tüüp 2 diabeet ja insuliinresistentsus. *PCOS*-ile on iseloomulikud nii metaboolsed kui ka endokriinsed sümptomid, mille alla kuuluvad muuhulgas ülekaal ja rasvumine, hüperandrogenism ning insuliinresistentsus. Ehkki haigust on ulatuslikult uuritud, puudub tänaseni lõplik selgus haiguse etioloogia kohta. Küll aga on teada, et *PCOS*-il on olemas tugev geneetiline komponent. Selle uurimiseks on läbi viidud mitmeid ülegenoomseid assotsiatsiooniuuringuid, millest on ilmnenu erinevaid võimalikke kandidaatgeene. Tihtilugu on aga uuringute piiranguks olnud väike valim ning asjaolu, et leiud ei ole replitseerunud teistes populatsioonides.

Kui enamik senini teostatud uuringutest on keskendunud sagedaste geneetiliste variantide uurimisele, siis antud töös viidi läbi analüüs harvade geneetiliste variantidega. Töö raames analüüsiti 4606 indiviidi (169 *PCOS*-i juhtu ning 4437 kontrollindiviidi). Analüüsimine toimus nii üksiku markeri kaupa, mille *MAF* oli alla 1% kui ka geenikaupa, kus uuriti kas funktsioonikaoga või mittesünonüümsete geneetiliste variantidega geenipiirkonda, kus variantide *MAF* oli samuti alla 1%.

Töö raames saadud tulemustes ilmnis üksiku markeri kaupa tehtud analüüsides kokku 43 geneetilist varianti, mille p-väärtused olid madalamad ülegenoomse statistilise olulisuse piirist peale mitmese korrektsiooni rakendamist. Markeripõhises analüüsis asus madalaima p-

väärtusega geneetiline variant geenis *USP6* (p-väärtus $1,64 \times 10^{-28}$). Geenipõhistes analüüsid, kus kasutati kahte mudelit – mittesünonüümsete ja funktsioonikaoga geneetiliste variantide mudel – ilmnes peale mitmese testimise korrektsiooni ülegenoomse statistilise olulisuse piirist madalamate p-väärtustega geene vaid mittesünonüümsete variantide uuringus. Kokku selgus mittesünonüümsete uuringust 12 erinevat geeni, neist madalaima p-väärtusega geen *ILIF7* (p-väärtus $4,4 \times 10^{-10}$). Nagu öeldud, ei ilmnenu funktsioonikaoga variantide uuringust ühtegi geeni, mille p-väärtus oleks olnud madalam statistilise olulisuse piirist; madalaima p-väärtusega geen sealses analüüsis oli *CNKSRI* (p-väärtus $4,23 \times 10^{-5}$).

Käesolevas töös leiti mitmeid geneetilisi variante, mis võivad olla *PCOS*-iga seotud. Kirjanduse uurimisel selgus, et osasid siinsetest leidudest on juba varasemalt haigusega seostatud. Kindlasti esineb leidude seas ka valepositiivseid tulemusi ja seda eeskätt valimi väiksest suurusest tulenevalt. Ilmnenu geneetilised variandid, mis antud uuringus *PCOS*-iga assotsieeritud, vajavad veel edasist valideerimist.

Tulenevalt sellest, et *PCOS*-il on olemas tugev pärilik komponent, aitab selle uurimine jõuda lähemale haiguse bioloogilise tausta mõistmiseni. See omakorda võimaldaks paremat ülevaadet haiguse tekkemehhanismist ja paremate *PCOS*-i ennetus- või ravimeetodite arendamist, mis aitaksid selle laialt levinud komplekshaiguse põdejaid kogu maailmas.

Resümee

Rare variant analysis to find genetic variants associated with polycystic ovary syndrome

Kreete Lüll

Summary

Polycystic ovary syndrome (PCOS) is an endocrine highly heritable complex disorder of unknown aetiology that affects up to 10% of women in reproductive age. PCOS is characterized by menstrual irregularities (an- or oligo-ovulation), clinical or biochemical hyperandrogenism and polycystic ovarian morphology. Women who have PCOS may also suffer from acne, hirsutism, insulin resistance, overweight or obesity and infertility. Moreover, people with PCOS have increased risk for type II diabetes and cardiovascular diseases.

As said, the aetiology of PCOS remains unknown. Many genetic studies have been conducted that could identify underlying genes and pathways. Genome-wide association studies among different populations have found many possible candidate genes that could provide insights into the origin of PCOS. However, the candidate gene studies have not been conclusive and only a small portion of loci that were found in Han Chinese population have been replicated in European population. A possible reason for this could be that some of the studies have been underpowered, also it is probable that PCOS has, to some extent, different causes in European and Han Chinese population. As common variants have not managed to explain the underlying causes of PCOS, it is believed that rare variants could clarify the missing heritability of PCOS.

In this study, I conducted a rare variant analysis in Estonian population with 169 PCOS cases and 4437 controls. 110 of the cases were collected in the Tartu Ülikooli Kliinikumi Naistekliinik based on Rotterdam criteria, where cases presented at least 2 out of these 3 symptoms: clinical and/or biochemical signs of hyperandrogenism, polycystic ovarian morphology and an- or oligo-ovulation. 59 of the cases were self-reported and genedonors of Tartu Ülikooli Eesti Geenivaramu. I conducted single variant analysis and gene-based analysis in which I used two different models (1) non-synonymous variants in a gene and (2) loss-of-function variants in a gene.

In single variant analysis, the statistically significant p-value after adjusting for multiple analysis was 7.6×10^{-7} . In total 43 variants had lower p-values that, 39 non-synonymous, 2 synonymous, 1 in inter genic region and 1 in a gene intron. In the gene-based analysis, where non-synonymous variants in a gene were analysed, the statistically significant p-value after multiple testing adjusting was 3.56×10^{-6} . In this analysis, 12 variants in 12 different genes appeared that had lower p-values. Lowest of them was a variant in gene *ILIF7* that has previously been associated with insulin resistance and tuberculosis. In the second gene-based analysis where loss-of-function variants were used, the statistically significant p-value was 2.91×10^{-6} but none of the variants that were analyzed had lower p-values.

Overall, many rare variants in different genes were found in the current study. Not all of them were previously associated with PCOS. There still are some interesting findings, for example *PLIN* that has been associated with obesity, which is also highly prevalent in PCOS patients. Another interesting gene is *LHCGR*, it has been shown that it has a role in human folliculogenesis and it has been previously shown that it might have a role in PCOS development.

In this study, different rare variants were found in Estonian population that could explain the aetiology of PCOS. It is important to keep in mind that these are preliminary results that need to be replicated in a study with bigger sample size. This study is also a part of a bigger PCOS consortia meta-analysis that is hopefully able to describe the genetic component of PCOS.

Tänuavaldused

Soovin tänada oma juhendajat Reedik Mägi parima võimaliku juhendamise eest ning Triin Laisk-Podarit omapoolsete nõuannete ning töö lugemise eest.

Lisaks tänan pere ja sõpru, kes on mulle magistriõppe ajal suureks toeks olnud.

Kasutatud kirjandus

Abbott, D., Dumesic, D. ja Franks, S. (2002). Developmental origin of polycystic ovary syndrome—a hypothesis. *Journal of Endocrinology*. 174: 1–5.

Abraham, G. ja Inouye, M. (2014). Fast principal component analysis of large-scale genome-wide data. *PLoS ONE*. 9(4): 1–5.

Altmüller, J., Palmer, L. J., Fischer, G., Scherb, H. ja Wjst, M. (2001). Genomewide scans of complex human diseases: true linkage is hard to find. *American Journal of Human Genetics*. 69(5): 936–50.

Azziz, R., Carmina, E., Dewailly, D., Diamanti-Kandarakis, E., Escobar-Morreale, H. F., Futterweit, W. ja Witchel, S. F. (2006). Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *The Journal of Clinical Endocrinology and Metabolism*. 91(11).

Azziz, R., Carmina, E., Dewailly, D., Diamanti-Kandarakis, E., Escobar-Morreale, H. F., Futterweit, W. ja Witchel, S. F. (2009). The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertility and Sterility*. 91(2): 456–88.

Azziz, R., Tarlatzis, R., Dunaif, A., Ibanez, L., Pugeat, M., Taylor, A. ja Fauser, C. J. *Fertility and Sterility*. (2004). Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertility and Sterility*. 81(1): 19–25.

Ballak, D. B., van Diepen, J., Moschen, A. R. jansen, H. J., Hijmans, A., Groenhof, G.-J. ja Stienstra, R. (2014). IL-37 protects against obesity-induced inflammation and insulin resistance. *Nature Communications*. 5: 4711.

Barber, T. M. ja Franks, S. (2013). Genetics of polycystic ovary syndrome. *Frontiers of Hormone Research*. 40: 28–39.

Basora, N., Nomura, H., Berger, U. V, Stayner, C., Guo, L., Shen, X. ja Zhou, J. (2002). Tissue and Cellular Localization of a Novel Polycystic Kidney Disease – Like Gene Product , Polycystin-L. *Journal of the American Society of Nephrology*. 13: 293–301.

Casarini, L. ja Brigante, G. (2014). The polycystic ovary syndrome evolutionary paradox: A genome-wide association studies-based, in silico, evolutionary explanation. *Journal of Clinical Endocrinology and Metabolism*. 99(11): E2412–E2420.

Catrrall, F. R. ja Healy, D. L. (2004). Long-term metabolic, cardiovascular and neoplastic risks with polycystic ovary syndrome. *Best Practice ja Research. Clinical Obstetrics ja Gynaecology*. 18(5): 803–12.

Chen, Z.-J., Zhao, H., He, L., Shi, Y., Qin, Y., Shi, Y. ja Zhao, Y. (2011). Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nature Genetics*. 43(1): 55–9.

- Crosignani, P. G. ja Nicolosi, E. (2001). Polycystic ovarian disease: heritability and heterogeneity. *Human Reproduction Update*. 7(1): 3–7.
- Daan, N. M. P., Louwers, Y. V, Koster, M. P. H., Eijkemans, M. J. C., de Rijke, Y. B., Lentjes, E. W. G. ja Laven, J. S. E. (2014). Cardiovascular and metabolic profiles amongst different polycystic ovary syndrome phenotypes: who is really at risk? *Fertility and Sterility*. 102(5): 1444–1451.
- Day, F. R., Hinds, D. a., Tung, J. Y., Stolk, L., Styrkarsdottir, U., Saxena, R. ja Perry, J. R. B. (2015). Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nature Communications*. 6, 8464.
- Diamanti-Kandarakis, E. ja Dunaif, A. (2012). Insulin resistance and the polycystic ovary syndrome revisited: An update on mechanisms and implications. *Endocrine Reviews*. 33(6): 981–1030.
- Diamanti-Kandarakis, E., Kouli, C. R., Bergiele, A. T., Filandra, F. A., Tsianateli, T. C., Spina, G. G. ja Bartzis, M. I. (1999). A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal and metabolic profile. *The Journal of Clinical Endocrinology and Metabolism*. 84(11): 4006–4011.
- Domecq, J. P., Prutsky, G., Mullan, R. J., Hazem, A., Sundaresh, V., Elamin, M. B. ja Murad, M. H. (2013). Lifestyle modification programs in polycystic ovary syndrome: Systematic review and meta-Analysis. *Journal of Clinical Endocrinology and Metabolism*. 98(12): 4655–4663.
- Donath, M. Y. ja Shoelson, S. E. (2011). Type 2 diabetes as an inflammatory disease. *Nature Reviews. Immunology*, 11(2): 98–107.
- Dong, X., Tang, W., Stopenski, S., Brose, M. S., Korch, C. ja Meinkoth, J. L. (2012). RAP1GAP inhibits cytoskeletal remodeling and motility in thyroid cancer cells. *Endocrine-Related Cancer*. 19(4): 575–588.
- Dunaif, A. (2016). Perspectives in Polycystic Ovary Syndrome: From Hair to Eternity. *The Journal of Clinical Endocrinology ja Metabolism*. 101(3): 759–768.
- Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H. ja Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics*. 11(6): 446–450.
- El-Fishawy, P. (2013). Encyclopedia of Autism Spectrum Disorders, p 719–720. *In* F. R. Volkmar (ed.), New York, NY: Springer New York.
- Franks, S. ja Hardy, K. (2010). Aberrant follicle development and anovulation in polycystic ovary syndrome. *Annales D’endocrinologie*. 71(3): 228–230.
- Franks, S., Stark, J. ja Hardy, K. (2008). Follicle dynamics and anovulation in polycystic ovary syndrome. *Human Reproduction Update*. 14(4): 367–378.

Franks, S., Webber, L. J., Goh, M., Valentine, A., White, D. M., Conway, G. S. ja McCarthy, M. I. (2008). Ovarian morphology is a marker of heritable biochemical traits in sisters with polycystic ovaries. *Journal of Clinical Endocrinology and Metabolism*. 93(9): 3396–3402.

Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics*. 13(2): 135–145.

Goldstein, J. I., Crenshaw, A., Carey, J., Grant, G. B., Maguire, J., Fromer, M. ja Neale, B. M. (2012). Zcall: A rare variant caller for array-based genotyping. *Bioinformatics*. 28(19): 2543–2545.

Gonzalez-Covarrubias, V., Ghosh, D., Lakhman, S. S., Pendyala, L. ja Blanco, J. G. (2007). A functional genetic polymorphism on human carbonyl reductase 1 (CBR1 V88I) impacts on catalytic activity and NADPH binding affinity. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*. 35(6): 973–980.

Goodarzi, M. O., Jones, M. R., Li, X., Chua, A. K., Garcia, O. A., Chen, Y.-D. I. ja Urbanek, M. (2012). Replication of association of DENND1A and THADA variants with polycystic ovary syndrome in European cohorts. *Journal of Medical Genetics*. 49(2): 90–95.

Greenberg, A. S., Egan, J. J., Wek, S. A., Garty, N. B., Blanchette-mackieej, E. J. ja Londos, C. (1991). Perilipin, a Major Hormonally Regulated Adipocyte-specific Phosphoprotein Associated with the Periphery of Lipid Storage Droplets. *The Journal of Biological Chemistry*. 266(17): 11341–11346.

Guo, L., Chen, M., Basora, N. ja Zhou, J. (2000). The human polycystic kidney disease 2-like (PKDL) gene: exon/intron structure and evidence for a novel splicing mechanism. *Mammalian Genome*. 50(11): 46–50.

Hart, R., Hickey, M. ja Franks, S. (2004). Definitions, prevalence and symptoms of polycystic ovaries and polycystic ovary syndrome. *Best Practice ja Research. Clinical Obstetrics ja Gynaecology*. 18(5): 671–683.

Hayes, M. G., Urbanek, M., Ehrmann, D. a., Armstrong, L. L., Lee, J. Y., Sisk, R. ja Dunaif, A. (2015). Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nature Communications*. 6(May): 7502.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S. ja 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): 9362–9367.

Huang, Z., Gao, C., Chi, X., Hu, Y. W., Zheng, L., Zeng, T. ja Wang, Q. (2015). IL-37 Expression is Upregulated in Patients with Tuberculosis and Induces Macrophages Towards an M2-like Phenotype. *Scandinavian Journal of Immunology*. 82(4): 370–379.

Insler, V., Shoham, Z., Barash, A., Koistinen, R., Seppala, M., Hen, M., Lunenfeld, B. ja Zadik, Z. (1993). Polycystic ovaries in non-obese and obese patients: possible

pathophysiological mechanism based on new interpretation of facts and findings. *Human Reproduction*. 8(3): 379-384

Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D. ja Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. *American Journal of Human Genetics*. 92(6): 841–853.

Jager, J., Grémeaux, T., Cormont, M., Le Marchand-Brustel, Y. ja Tanti, J.-F. (2007). Interleukin-1 β -induced insulin resistance in adipocytes through down-regulation of insulin receptor substrate-1 expression. *Endocrinology*. 148(1): 241–251.

Jakobs, P. M., Hess, J. F., FitzGerald, P. G., Kramer, P., Weleber, R. G. ja Litt, M. (2000). Autosomal-Dominant Congenital Cataract Associated with a Deletion Mutation in the Human Beaded Filament Protein Gene BFSP2. *American Journal of Human Genetics*. 66(4): 1432–1436.

Johnstone, E. B., Rosen, M. P., Neril, R., Trevithick, D., Sternfeld, B., Murphy, R. ja Cedars, M. I. (2010). The polycystic ovary post-Rotterdam: A common, age-dependent finding in ovulatory women without metabolic significance. *Journal of Clinical Endocrinology and Metabolism*, 95(11): 4965–4972.

Kawai, T., Ng, M. C. Y., Hayes, M. G., Yoshiuchi, I., Tsuchiya, T., Robertson, H. ja Ehrmann, D. A. (2010). Variation in the Perilipin Gene (PLIN) Affects Glucose and Lipid Metabolism in Non-Hispanic White Women With and Without Polycystic Ovary Syndrome. *Diabetes Research and Clinical Practice*. 86(3): 186–192.

Ke, X., Taylor, M. S. ja Cardon, L. R. (2008). Singleton SNPs in the human genome and implications for genome-wide association studies. *European Journal of Human Genetics: EJHG*. 16(4): 506–515.

Kiezun, A., Garimella, K., Do, R., Stitzel, N. O., Benjamin, M., McLaren, P. J. ja Magnusson, P. (2012). Exome sequencing and the genetic basis of complex traits. *Nature Genetics*. 44(6): 623–630.

Koh, D. I., Choi, W. Il, Jeon, B. N., Lee, C. E., Yun, C. O. ja Hur, M. W. (2009). A novel POK family transcription factor, ZBTB5, represses transcription of p21CIP1 gene. *Journal of Biological Chemistry*. 284(30): 19856–19866.

Kotowski, I. K., Pertsemlidis, A., Luke, A., Cooper, R. S., Vega, G. L., Cohen, J. C. ja Hobbs, H. H. (2006). A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *American Journal of Human Genetics*. 78(3): 410–22.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F. ja Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 467(7317): 832–838.

Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H. ja Metspalu, A. (2014). Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology*. 44(4): 1137–1147.

- Lim, J., Zhou, M., Veenstra, T. D. ja Morrison, D. K. (2010). The CNK1 scaffold binds cytohesins and promotes insulin pathway signaling. *Genes and Development*. 24(14): 1496–1506.
- Lindop, R., Arentz, G., Bastian, I., Whyte, A. F., Thurgood, L. a., Chataway, T. K. ja Gordon, T. P. (2013). Long-term Ro60 humoral autoimmunity in primary Sjögren's syndrome is maintained by rapid clonal turnover. *Clinical Immunology*. 148(1): 27–34.
- Louwers, Y. V., Stolk, L., Uitterlinden, A. G. ja Laven, J. S. E. (2013). Cross-ethnic meta-analysis of genetic variants for polycystic ovary syndrome. *The Journal of Clinical Endocrinology and Metabolism*. 98(12): E2006–E2012.
- Ma, X., Li, F.-F., Wang, S.-Z., Gao, C., Zhang, M. ja Zhu, S.-Q. (2008). A new mutation in BFSP2 (G1091A) causes autosomal dominant congenital lamellar cataracts. *Molecular Vision*. 14(August): 1906–1911.
- Mackenzie, R. W. ja Elliott, B. T. (2014). Akt/PKB activation and insulin signaling: a novel insulin signaling pathway in the treatment of type 2 diabetes. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*. 7: 55–64.
- Manolio, T. a, Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L., Hunter, D. J. ja Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*. 461(7265): 747–753.
- McCarthy, M. I. ja Hirschhorn, J. N. (2008). Genome-wide association studies: potential next steps on a genetic journey. *Human Molecular Genetics*. 17(R2): R156–65.
- Moran, L. J., Ko, H., Misso, M., Marsh, K., Noakes, M., Talbot, M. ja Teede, H. J. (2013). Dietary composition in the treatment of polycystic ovary syndrome: a systematic review to inform evidence-based guidelines. *Journal of the Academy of Nutrition and Dietetics*. 113(4): 520–545.
- Morris, A. P. ja Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genetic Epidemiology*. 34(2): 188–193.
- Mutharasan, P., Galdones, E., Peñalver Bernabé, B., Garcia, O. A., Jafari, N., Shea, L. D., Woodruff, T. K., Legro, R. S. ja Urbanek, M. (2013). Evidence for chromosome 2p16.3 polycystic ovary syndrome susceptibility locus in affected women of European ancestry. *The Journal of Clinical Endocrinology and Metabolism*. 98(1): E185–E190.
- Nicoloso, M. S., Sun, H., Spizzo, R., Kim, H., Wickramasinghe, P., Shimizu, M. ja Calin, G. A. (2010). Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Research*. 70(7): 2789–2798.
- Osawa, Y., Yokoyama, Y., Shigeto, T., Futagami, M. ja Mizunuma, H. (2015). Decreased expression of carbonyl reductase 1 promotes ovarian cancer growth and proliferation. *International Journal of Oncology*. 46(3): 1252–1258.

- Polson, D. W., Wadsworth, J., Adams, J. ja Franks, S. (1988). Polycystic ovaries - a common finding in normal women. *The Lancet*. 331(8590): 870–872.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D. ja Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. 81(3): 559–75.
- Qi, L., Zhang, C., Greenberg, A. ja Hu, F. B. (2008). Common Variations in Perilipin Gene, Central Obesity, and Risk of Type 2 diabetes in US Women. *Obesity (Silver Spring)*. 16(5): 1061–1065.
- Qiu, T., Qi, X., Cen, J. ja Chen, Z. (2012). Rap1GAP alters leukemia cell differentiation, apoptosis and invasion in vitro. *Oncology Reports*. 28(2): 622–628.
- Reed, J. H., Dudek, N. L., Osborne, S. E., Kaufman, K. M. Jackson, M. W., Purcell, A. W. ja Gordon, T. P. (2010). Reactivity with dichotomous determinants of Ro 60 stratifies autoantibody responses in lupus and primary Sjögren's syndrome. *Arthritis and Rheumatism*. 62(5): 1448–1456.
- Schork, N. J., Murray, S. S., Frazer, K. ja Topol, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. *Current Opinion in Genetics and Development*. 19(3): 212–219.
- Shi, Y., Zhao, H., Shi, Y., Cao, Y., Yang, D., Li, Z. ja Chen, Z.-J. (2012). Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nature Genetics*. 44(9): 1020–1025.
- Sirmans, S. M. ja Pate, K. (2013). Epidemiology, diagnosis, and management of polycystic ovary syndrome. *Clinical Epidemiology*. 6(pcos): 1–13.
- Sharpe, R. ja Franks S. (2002). Environment, lifestyle and infertility: an intergenerational issue. *Nature Cell Biology*. 4: 33-40.
- Stevens, E. L., Heckenberg, G., Roberson, E. D. O., Baugher, J. D., Downey, T. J. ja Pevsner, J. (2011). Inference of relationships in population data using identity-by-descent and identity-by-state. *PLoS Genetics*. 7(9).
- Zhang, H. Y., Guo, C. X., Zhu, F. F., Qu, P. P., Lin, W. J. ja Xiong, J. (2013). Clinical characteristics, metabolic features, and phenotype of Chinese women with polycystic ovary syndrome: a large-scale case-control study. *Archives of Gynecology and Obstetrics*. 287(3): 525–531.
- Zhang, L., Gao, L., Li, Z., Qin, W., Gao, W., Cui, X. ja Liu, P. (2006). Progressive sutural cataract associated with a BFSP2 mutation in a Chinese family. *Molecular Vision*. 12(December): 1626–1631.
- Teede, H., Deeks, A. ja Moran, L. (2010). Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. *BMC Medicine*. 8: 41.

- Thathapudi, S., Kodati, V., Erukkambattu, J., Addepally, U. ja Qurratulain, H. (2015). Association of luteinizing hormone chorionic gonadotropin receptor gene polymorphism (rs2293275) with polycystic ovarian syndrome. *Genetic Testing and Molecular Biomarkers*. 19(3): 128–132.
- Thelin, M. a., Svensson, K. J., Shi, X., Bagher, M., Axelsson, J., Isinger-Ekstrand, A. ja Malmström, A. (2012). Dermatan sulfate is involved in the tumorigenic properties of esophagus squamous cell carcinoma. *Cancer Research*. 72(8): 1943–1952.
- Tsygankova, O. M., Wang, H. ja Meinkoth, J. L. (2013). Tumor cell migration and invasion are enhanced by depletion of Rap1 GTPase-activating protein (Rap1GAP). *Journal of Biological Chemistry*. 288(34): 24636–24646.
- Tzeng, J., Lu, H. H. ja Li, W. H. (2008). Multidimensional scaling for large genomic data sets. *BMC Bioinformatics*. 9, 179.
- Wang, K., Li, M. ja Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* 38(16): e164.
- Wang, P., Zhao, H., Li, T., Zhang, W., Wu, K., Li, M. ja Chen, Z. (2014). Hypomethylation of the LH/Choriogonadotropin Receptor Promoter Region Is a Potential Mechanism Underlying Susceptibility to Polycystic Ovary Syndrome. *Endocrinology*. 155(4): 1445–1452.
- Veldhuisen, B., Spruit, L., Dauwerse, H. G., Breuning, M. H. ja Peters, D. J. (1999). Genes homologous to the autosomal dominant polycystic kidney disease genes (PKD1 and. *European Journal of Human Genetics: EJHG*. 7(June): 860–872.
- Welt, C. K., Styrkarsdottir, U., Ehrmann, D. a, Thorleifsson, G., Arason, G., Gudmundsson, J. ja Stefansson, K. (2012). Variants in DENND1A are associated with polycystic ovary syndrome in women of European ancestry. *The Journal of Clinical Endocrinology and Metabolism*. 97(7): E1342–7.
- Vink, J. M., Sadrzadeh, S., Lambalk, C. B. ja Boomsma, D. I. (2006). Heritability of polycystic ovary syndrome in a Dutch twin-family study. *The Journal of Clinical Endocrinology and Metabolism*. 91(6): 2100–4.
- Visscher, P. M. (2008). Sizing up human height variation. *Nature Genetics*. 40(5): 489–90.
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E. ja Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*. 42(7): 565–569.
- Yildiz, B. O., Knochenhauer, E. S. ja Azziz, R. (2008). Impact of obesity on the risk for polycystic ovary syndrome. *The Journal of Clinical Endocrinology and Metabolism*. 93(1): 162–8.

Yildiz, B. O., Yarali, H., Oguz, H. ja Bayraktar, M. (2003). Glucose intolerance, insulin resistance, and hyperandrogenemia in first degree relatives of women with polycystic ovary syndrome. *The Journal of Clinical Endocrinology and Metabolism*. 88(5): 2031–6.

Yung, Y., Maman, E., Rubinstein, N., Avivi, C., Orvieto, R. ja Hourvitz, A. (2014). Localization of luteinizing hormone receptor protein in the human ovary. *Molecular Human Reproduction*. 20(9): 844–849.

Kasutatud veebiaadressid

<http://www.clearpassage.com/what-you-should-know-about-pcos/> (kasutatud 06.05.2016))

http://www.illumina.com/content/dam/illumina/marketing/documents/products/datasheets/data_sheet_human_core_exome_beadchip.pdf (kasutatud 06.03.2016)

<http://www.ncbi.nlm.nih.gov/refseq/> (kasutatud 06.03.2016)

<https://www.r-project.org/> (kasutatud 06.03.2016)

<http://genome.sph.umich.edu/wiki/EMMAX> (kasutatud 12.03.2016)

<http://genome.sph.umich.edu/wiki/EPACTS> (kasutatud 12.03.2016)

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kreete Lüll,

(sünnikuupäev: 21.10.1992)

annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Harvade variantide analüüs polütsüstiliste munasarjade sündroomiga seotud geneetiliste variantide tuvastamiseks,

mille juhendaja on PhD Reedik Mägi,

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

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

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

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

Tartus, 27.05.2016