

TARTU ÜLIKOOL  
LOODUS- JA TEHNOLOOGIATEADUSKOND  
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT  
BIOTEHNOLOOGIA ÕPPETOOL

Kreete Lüll

**Ülegenoomne assotsiatsiooniuring polütsüstiliste munasarjade sündroomiga seotud  
geneetiliste variantide leidmiseks**

Bakalaureusetöö

Juhendajad Reedik Mägi, PhD

Prof. Andres Metspalu, M.D., PhD

TARTU 2014

# Sisukord

Sisukord.....	2
Kasutatud lühendid.....	4
Sissejuhatus .....	5
1. Kirjanduse ülevaade .....	6
1.1. Polütsüstiliste munasarjade sündroom.....	6
1.1.1. Polütsüstiliste munasarjade sündroomi sümptomid .....	6
1.1.2. Polütsüstiliste munasarjade sündroomi diagnoosimise kriteeriumid.....	8
1.1.3. Polütsüstiliste munasarjade sündroomi ravivõimalused.....	9
1.1.4. Polütsüstiliste munasarjade sündroomi geneetika ja varasemad uuringud.....	10
1.2. Ülegenoomsed assotsiatsiooniuuringud.....	11
1.2.1. Genotüübi imputatsioon .....	12
1.2.2. Ülegenoomsete assotsiatsiooniuuringute probleemid .....	13
2. Uurimus.....	15
2.1. Töö eesmärk.....	15
2.2. Materjalid ja meetodika .....	15
2.2.1. Valim .....	15
2.2.2. Genotüpiseerimine ja kvaliteedikontroll .....	16
2.2.3. Imputeerimine.....	16
2.2.4. Assotsiatsioonianalüüs .....	17
2.2.5. Replikatsioonianalüüs.....	17
2.2.6. Varasemalt teadaolevate markerite võrdlus.....	17
2.3. Tulemused.....	18
2.4. Arutelu .....	24
Kokkuvõte .....	27
Resümees/Summary.....	28
Tänuavaldused.....	30
Kasutatud kirjandus .....	31

Kasutatud veebileheküljed.....	36
Lisad .....	37
Lihtlitsents .....	39

## Kasutatud lühendid

<b>ASRM</b>	<i>American Society for Reproductive Medicine</i>
<b>CI</b>	Usaldusintervall ( <i>Confidence Interval</i> )
<b>CR</b>	Edukalt genotüpileeritud indiviidide protsentuaalne osakaal markeril ( <i>Call Rate</i> )
<b>EGCUT</b>	Tartu Ülikooli Eesti Geenivaramu ( <i>Estonian Genome Center, University of Tartu</i> )
<b>ESHRE</b>	<i>European Society for Human Reproduction and Embryology</i>
<b>GWAS</b>	Ülegenoomne assotsiatsiooniuuring ( <i>Genome – Wide Association Study</i> )
<b>HMM</b>	Peidetud Markovi mudel ( <i>Hidden Markov Model</i> )
<b>HWE</b>	Hardy – Weinbergi tasakaalustamatus ( <i>Hardy – Weinberg Equilibrium</i> )
<b>LD</b>	Aheldatuse tasakaalustamatus ( <i>Linkage Disequilibrium</i> )
<b>MAF</b>	Harvema alleeli sagedus ( <i>Minor Allele Frequency</i> )
<b>NIH</b>	<i>National Institutes of Health</i>
<b>NICHD</b>	<i>National Institute of Child Health and Human Disease</i>
<b>OR</b>	Šansside suhe ( <i>Odds Ratio</i> )
<b>PCO</b>	Polütsüstilised munasarjad ( <i>Polycystic Ovary</i> )
<b>PCOS</b>	Polütsüstiliste munasarjade sündroom ( <i>Polycystic Ovary Syndrome</i> )
<b>SNP</b>	Ühenukleotiidiline polümorfism ( <i>Single Nucleotide Polymorphism</i> )
<b>SWH</b>	<i>Salvador-Warts-Hippo</i> signaalülekanerada
<b>WTCHG</b>	Oxfordi Ülikooli <i>Wellcome Trust</i> Inimgeneetika Instituut ( <i>Wellcome Trust Centre for Human Genetics, University of Oxford</i> )

## Sissejuhatus

Polütsüstiliste munasarjade sündroom (*Polycystic Ovary Syndrome, PCOS*) on üheks enamlevinumaks hormonaalseks komplekshaiguseks, esinedes 5 – 10 protsendil reproduktiivses eas olevatel naistel. *PCOS*-i põhilisteks sümptomiteks on polütsüstilised munasarjad, häired menstruaaltsükliks, hüperandrogenism ning ülekaalulisus. Lisaks peetakse polütsüstiliste munasarjade sündroomi üheks peamiseks viljatuse põhjustajaks naistel.

Tänini puudub lõplik selgus, mis võiks olla polütsüstiliste munasarjade sündroomi tekkepõhjuseks. Selle välja selgitamiseks on viimasel ajal on üha rohkem hakatud läbi viima ülegenoomseid assotsiatsiooniuuringuid (*Genome – Wide Association Study, GWAS*), mille käigus otsitakse geneetilisi variante, mis võiksid haigusega assotsieerunud olla. *GWAS*-ide tulemuste analüüsimise tagajärjena on esile kerkinud mitmeid potentsiaalseid *PCOS*-i põhjustajaid geene, üheks võimalikuks geeniks on näiteks D-vitamiini retseptori geen *DENNDIA*.

Käesoleva töö eesmärgiks on Tartu Ülikooli Eesti Geenivaramu (*Estonian Genome Center, University of Tartu, EGCUT*) doonorite andmetele toetudes viia läbi hüpoteesivaba ülegenoomne assotsiatsiooniuuring, et leida võimalikke assotsiatsioone polütsüstiliste munasarjade sündroomiga ning seejärel saadud tulemuste replitseerimine sõltumatus kohordis. Töö teiseks eesmärgiks on varasemalt assotsieerunud markerite väärtuste hindamine Eesti Geenivaramus läbiviidud analüüsis.

Töö koostati Tartu Ülikooli Eesti Geenivaramus ning see on osa polütsüstiliste munasarjade sündroomi konsortsiumi suuremast uuringust. Selle uuringu raames oli minu osaks kirjandusest teadaolevate assotsiatsioonide välja otsimine ja analüüs Eesti Geenivaramu andmetest, *GWAS* tulemustest assotsieerunud markerite leidmine, nende lähedal paiknevate geenide tuvastamine ja kirjandusest antud geenide funktsioonide leidmine.

# 1. Kirjanduse ülevaade

## 1.1. Polütsüstiliste munasarjade sündroom

Polütsüstiliste munasarjade sündroom, mida kannab keskmiselt kuni 10% reproduktiivses eas olevatest naistest, on kompleksne endokriinne häire, mis on ka üheks olulisemaks viljatuse põhjustajaks naistel (Domecq jt., 2013; Franks jt., 1997). *PCOS*-i kirjeldati esmakordselt 1935. aastal Irving F. Stein Sr. ja Michael L. Leventhali poolt, mistõttu on haigust nimetatud ka Stein – Leventhal tõveks (Azziz jt., 2009; Knochenhauer jt., 1998).

### 1.1.1. Polütsüstiliste munasarjade sündroomi sümptomid

*PCOS*-i puhul on tegemist komplekshaigusega, mille juures mängivad rolli nii pärilikkus kui keskkond. Haiguse kandjatel esineb mitmeid erinevaid sümptomeid, millest kõik aga ei pruugi esineda igal patsiendil (Welt ja Carmina, 2013). Samuti on *PCOS*-ile iseloomulikuks asjaolu, et sümptomid püsivad stabiilsena vaid varases täiskasvanueas ehk vanusevahemikus 18 – 30, hilisemas elus aga võivad haiguse sümptomid, tulenevalt muutustest munasarjades ja metabolismis, muutuda (Welt ja Carmina, 2013). *PCOS*-i mõju on täheldatud metabolismis (II tüüpi diabeet, insuliinresistentsus), reproduktiivsuses (hüperandrogenism, viljatus) ning lisaks mõjutab haigus patsiente psühholoogilisel tasandil (depressioon, ärevushood) (Teede jt., 2010). Enimnimetatud *PCOS*-i sümptomite alla kuuluvad: hüperandrogenism, häired menstruaaltsükli, polütsüstilised munasarjad ning ülekaalulisus ja rasvumine (Domecq jt., 2013; Teede jt., 2010). Kuigi *PCOS*-ile omaseid sümptomeid on võimalik määrata juba noorukieas, kinnitatakse diagnoos reeglina siiski alles täiskasvanuikka jõudes. Põhjus seisneb selles, et mõned *PCOS*-i sümptomid kattuvad murdeeaie iseloomulike nähtudega nagu näiteks menstruaaltsükli ebaregulaarsus, mis ei pruugi olla regulaarne ka 2 aasta möödumisel esimesest menstruatsioonist (Welt ja Carmina, 2013).

Androgeenide ehk meessuguhormoonide liig *PCOS*-iga naiste hulgas on kõrge, täheldatud on, et 60% kuni 80% naistest, kellel on androgeenide ülehulk, põevad polütsüstiliste munasarjade sündroomi (Sirmans ja Pate, 2013; Teede jt., 2010). Hüperandrogenism toob endaga kaasa mitmeid sümptomeid nagu näiteks hirsutism ehk liigne karvakasv, akne ja androgeenne alopeetsia ehk meestüüpi juuksekadu (Badawy ja Elnashar, 2011). Neist kõige levinumaks, kuni 70% ulatuses, on hirsutism, mida määratakse toetudes Ferriman – Gallwey hindamissüsteemile, kus hinnatakse karvakasvu reitel, kätel, alakõhul, seljal, ülahuulel, näol ning rinnal (Sirmans ja Pate, 2013). Aknet esineb *PCOS*-iga patsientidel ligikaudu 35%-l juhtudest, kuid seda ei loeta nii spetsiifiliseks tunnuseks kui hirsutismi. Siiski uuritakse vahel tõsise aknega patsientidel täiendavaid märke hüperandrogenismist ning küsitakse teavet nende

menstruaaltsükli kohta, sest 40% naistest, kellel esineb tugev akne, diagnoositakse *PCOS* (Sirmans ja Pate, 2013). Alopeetsiat esineb *PCOS*-iga naistel samuti vähem kui hirsutismi ning selle puhul on oluliseks teguriks pärilikkusest tulenev eelsoodumus (Sirmans ja Pate, 2013; Teede jt., 2010).

Polütsüstiliste munasarjade sündroomiga naiste seas on levinud häired menstruaaltsükli. Oligomenorröa (ebaregulaarne menstruaaltsükkel, menstruatsioon 35 ja enama päeva tagant) ja ka amenorröa (menstruaaltsükli puudumine) esinevad kuni 80%-l *PCOS*-iga naistest. Oligomenorröaga naistel diagnoositakse *PCOS*-i 80% kuni 90% ulatuses, kuid amenorröa puhul on diagnoosimääraks kuni 40% (Sirmans ja Pate, 2013; Teede jt., 2010). Selle põhjuseks on asjaolu, et amenorröad kutsuvad tavaliselt esile hüpotaalamuse düsfunktsioon. Samas ei pruugi menstruaaltsüklihäired esineda kõigil *PCOS*-iga naistele (Sirmans ja Pate, 2013; Teede jt., 2010).

*PCOS* on oluliseks viljatuse põhjustajaks naiste seas, mõjutades ligikaudu 40%-i patsientidest (Teede jt., 2010). Enim esineb viljatuseprobleeme ülekaalulistel haigetel, kelle puhul on raskendatud ka viljatuse ravi. Samuti juhul, kui ülekaalulisel patsiendil õnnestub rasestuda, on neil võrreldes normaalkaalus olevate naistega raseduse katkemise oht suurem (Teede jt., 2010). Ehkki menstruaaltsüklihäired ja viljatuse on olulisteks sümptomiteks polütsüstiliste munasarjade sündroomi juures, ei esine need kõigil *PCOS*-i patsientidel. 2003. aastal avaldatud *ESHRE/ASRM* kriteeriumist on näha, et kõrvalekalded menstruatsioonil ei ole diagnoosimise juures obligatoorsed, mis tähendab, et *PCOS* võib esineda ka naistel, kelle menstruaaltsükkel on normaalne (Azziz jt., 2004; Sirmans ja Pate, 2013).

*PCOS* on oma nime saanud polütsüstiliste munasarjade järgi, mis samuti sümptomite hulka kuuluvad. Munasarjad loetakse polütsüstilisteks juhul kui kas ühes või mõlemas munasarjas leidub 12 või rohkem 2 – 9 millimeetri suurust folliikulit ehk munarakualget, mille areng on mingil põhjusel peatunud (Hart jt., 2004). Viljatuse peamine põhjus *PCOS*-i puhul on asjaolu, et folliikulid ei ole lõpuni arenenud, mis toob kaasa selle, et ei teki ovulatsiooni, mille käigus tavaolukorras valminud munarakk vabaneks (Hart jt., 2004).

*PCOS*-i diagnoosimise põhilisteks näidustusteks loetakse hüperandrogenismi, oligo-ovulatsiooni ja teiste haiguste välistamist, mis oleks tingitud androgeenide liiast kuid olulisteks tunnusteks on lisaks veel insuliinresistentsus, hüperinsulineemia, II tüüpi diabeet ning rasvumine, mis kõik omavahel ka tugevalt korreleerunud on (Kosova ja Urbanek, 2013; Teede jt., 2010). Insuliini toodetakse pankrease  $\beta$ -rakkudes ja selle mõjul toodavad munasarjad androgeene ning vastutavad selle eest, et ei toimuks apoptoosi folliikulites, mis

võiks viia tsüstide tekkeni munasarjas. Kuna insuliinil on tugev seos *PCOS*-iga patsientide hulgas, on sellega seotud geenid olnud arvestatavateks kandidaatideks polütsüstiliste munasarjade sündroomile (Kosova ja Urbanek, 2013).

Lisaks on uuringutes hakatud pöörama tähelepanu ka *PCOS*-i psühholoogilisele mõjule. Ehkki enamasti suunatakse peardõhk füsioloogilisele ja bioloogilisele tasandile, on täheldatud, et naised, kes põevad *PCOS*-i, on vastuvõtlikumad ärevushäiretele ja depressioonile ning neil on sageli tervete naistega võrreldes madalam enesehinnang (Teede jt., 2010).

### **1.1.2. Polütsüstiliste munasarjade sündroomi diagnoosimise kriteeriumid**

Polütsüstiliste munasarjade sündroomi diagnoosimise kriteeriumeid on kolm: *NIH/NICHD* (*National Institutes of Health/National Institute of Child Health and Human Disease*), *ESHRE/ASRM* (*European Society for Human Reproduction and Embryology/American Society for Reproductive Medicine*), tuntud ka kui Rotterdami kriteerium ning *Androgen Excess and PCOS Society* poolt välja antud kriteerium (Tabel 1) (Azziz jt., 2004, 2006; Sirmans ja Pate, 2013).

1992. aastal anti *PCOS*-i diagnoosimiseks välja esimene kriteerium, *NIH/NICHD* kriteerium, mille järgi diagnoositi polütsüstiliste munasarjade sündroom juhul kui patsiendil esines (1) oligo-ovulatsioon, (2) märgid hüperandrogenismist ning (3) olid välistatud muud haigused, mis oleksid võinud põhjustada liigset androgeenide hulka veres nagu näiteks Cushingi sündroom, androgeene sekreteerivad kasvajad, idiopaatiline hüperandrogenism või kilpnäärme talitlushäired (Lujan jt., 2008; Teede jt., 2010). Üksteist aastat hiljem, aastal 2003, anti välja *ESHRE/ASRM* kriteerium, mis on tuntud ka Rotterdami kriteeriumi nime all (Azziz jt., 2004). Rotterdami kriteeriumis lisati diagnoosimiseks vajalike sümptomite nimistusse veel polütsüstiliste munasarjade esinemine, mille määramiseks kasutatakse ultraheliuuringut. Nimetatud muudatus viis selleni, et *PCOS*-i hakati diagnoosima senisest rohkematel naistel, seda põhjustas ka asjaolu, et Rotterdami kriteerium näeb ette, et haiguse diagnoosiks ei pea esinema kõik kolm sümptomit vaid kolmest kaks (Teede et jt., 2010). Lisaks kahele eelnevalt olemasolevale kriteeriumile andis *Androgen Excess and PCOS Society* 2006. aastal omalt poolt välja uue kriteeriumi, kus on olemas küll kõik seninimetatud sümptomid, kuid vastupidiselt Rotterdami kriteeriumile, peavad *PCOS*-i diagnoosiks esinema väljatoodud sümptomitest kõik kolm (Azziz jt., 2006, 2009).

*PCOS*-i määramise teeb keeruliseks haiguse heterogeensus ning asjaolu, et diagnoosimise juures kasutatakse teiste haiguste väljaarvamist. Kuigi haigus on nime saanud tsüstide järgi munasarjas, ei ole see kunagi olnud peamine ega ainus sümptom haiguse määratlemiseks, sest



kuni 30%-l tervetel naistel esineb samuti polütsüstilisi munasarju (Azziz jt., 2006). Olemasoleva kolme kriteeriumi juures on nimetatud küll peamised sümptomid, milleks on kliinilised ja/või biokeemilised märgid hüperandrogenismist, polütsüstilised munasarjad ning oligo- ja/või anovulatsioon, kuid tuleb mees pidada, et lisaks neile on *PCOS*-il veel palju iseloomulikke tunnuseid, mis hetkel olemasolevates kriteeriumites ei kajastu. Rahvusvaheliseks polütsüstiliste munasarjade sündroomi kriteeriumiks tänase seisuga on *ESHRE/ASRM* ehk Rotterdami kriteerium (Teede jt., 2010).

**Tabel 1** Polütsüstiliste munasarjade sündroomi diagnoosimise kolm kriteeriumit (Teede jt., 2010).

<i>NIH/NICHD (1992)</i>	<i>ESHRE/ASRM (Rotterdam) (2003)</i>	<i>Androgen Excess and PCOS Society (2006)</i>
Diagnoosimiseks peavad esinema alljärgnevatest kõik kolm:	Diagnoosimiseks peavad esinema vähemalt kaks kolmest:	Diagnoosimiseks peavad esinema alljärgnevatest kõik kolm:
<ul style="list-style-type: none"> <li>kliinilised ja/või biokeemilised märgid hüperandrogenismist</li> </ul>	<ul style="list-style-type: none"> <li>kliinilised ja/või biokeemilised märgid hüperandrogenismist</li> </ul>	<ul style="list-style-type: none"> <li>kliinilised ja/või biokeemilised märgid hüperandrogenismist</li> </ul>
<ul style="list-style-type: none"> <li>oligo-ovulatsioon</li> </ul>	<ul style="list-style-type: none"> <li>oligo-ovulatsioon ja/või anovulatsioon</li> </ul>	<ul style="list-style-type: none"> <li>oligo-ovulatsioon ja/või anovulatsioon ning polütsüstilised munasarjad</li> </ul>
<ul style="list-style-type: none"> <li>teiste androgeenide liiaga tulevate haiguste välistamine</li> </ul>	<ul style="list-style-type: none"> <li>polütsüstilised munasarjad</li> </ul>	<ul style="list-style-type: none"> <li>teiste androgeenide liiaga tulevate haiguste välistamine</li> </ul>

### 1.1.3. Polütsüstiliste munasarjade sündroomi ravivõimalused

Unikaalset ravivõimalust *PCOS*-ile ei ole, seda taaskord tulenevalt haiguse heterogeensusest. *PCOS*-i puhul sõltub ravi otseselt haigel esinevatest sümptomitest, mida saab jagada kolme gruppi: androgeenide liiaga seotud sümptomid (hirsutism, akne, alopeetsia), viljatusprobleemid ja menstruatsiooniga seotud häired (Badawy ja Elnashar, 2011). Ravivõimalusteks on nii ravimite kasutamine, kirurgilised protseduurid kui ka eluviiside muutmine.

Üheks soovitud ravimeetodiks *PCOS*-i puhul on kehakaalu langetamine toitumisharjumuste muutmise ja kehaliste harjutuste kaudu (Badawy ja Elnashar, 2011). Ülekaal esineb *PCOS*-i patsientide seas kuni 60%-l ning kaalukaotus viib reeglina märkimisväärsete sümptomite paranemiseni. On täheldatud, et kui suudetakse kaotada esialgsest kehakaalust 5%, toob see kaasa menstruaaltsükli normaliseerumise ning ovulatsiooni, mille tulemusena paraneb ka võimalus rasestuda (Badawy ja Elnashar, 2011; Sirmans ja Pate, 2013). Lisaks aitab kaalukaotus kaasa androgeenide hulga vähenemisele ja aitab parandada insuliinresistentsust. Kuid tuleb märkida, et kehakaalu langetamine on soovituslik vaid patsientidele, kelle kehamassiindeks on suurem kui  $25 - 27 \text{ kg/m}^2$  (Badawy ja Elnashar, 2011).

Kaalukaotusele lisaks on androgeenide liiast tulenevate sümptomite raviks mitmeid võimalusi. Üks esimesi valikuid siin on suukaudsete rasestumisvastaste tablettide kasutamine (*Oral Contraceptive Pill, OCP*), mida kasutatakse juhul, kui patsiendil ei ole antud hetkel soovi lapseootele jääda. *OCP*-de mõju põhjustab negatiivse tagasiside luteiniseeriva hormooni sekretsioonile, mis viib selleni, et munasarjad toodavad väiksemal hulgal androgeene (Badawy ja Elnashar, 2011).

Lisaks on ravimeetodina kasutusel laparoskoopia – üldnarkoosis teostatav operatsioon, mille käigus munasarjadest eemaldatakse tsüste ja kudesid, mis võivad olla vastutavad liigsete androgeenide tootmise eest. Laparoskoopia käigus tehakse naba lähedale lõige, mille kaudu viiakse kõhuõõnde valgusallikaga varustatud peenike kaamera – laparoskoop. Lisaks tehakse paar lõiget vaagnapiirkonda, mille kaudu viiakse kõhuõõnde operatsiooniks tarvilikud instrumendid. (Healthwise, 2012).

#### **1.1.4. Polütsüstiliste munasarjade sündroomi geneetika ja varasemad uuringud**

*PCOS* on kompleksne haigus ning juba üle 60 aasta on arvatud, et sellel on tugev geneetiline komponent (Govind jt., 1999). Seda arvamust on aegade jooksul läbi viidud uuringud ka kinnitanud, kuid samas ollakse veendunud, et lisaks geneetikale on *PCOS*-i puhul oluliseks teguriks keskkonnafaktor (Crosignani ja Nicolosi, 2001; Kosova ja Urbanek, 2013).

Erinevad uuringud on näidanud *PCOS*-i perekondlikku klasterdumist (Franks jt., 1997; Govind jt., 1999). Govind jt. poolt läbi viidud uuring näitas, et *PCOS*-i kandjate esimese astme naissoost järeltulijatel on 61%-line võimalus haigestumiseks (Govind jt., 1999). Kõrget pärilikkust näitas samuti Hollandis teostatud kaksikute uuring, mis hindas *PCOS*-i pärilikkuse määra lausa 79%-ni (Vink jt., 2006). Geneetilisele taustale annab kinnitust veel Legro jt. uuring, milles uuriti hüperandrogenismi pärilikkust *PCOS*-i kandjate seas. Tulemuseks saadi,

et hüperandrogenism on *PCOS*-i puhul perekondlikus seoses, esinedes 46%-l juhtudel õdedel (Legro jt., 1998).

Varasemalt on *PCOS*-i geneetilise komponendi leidmiseks läbi viidud peamiselt kandidaatgeeni uuringuid, mis eeldavad häid teadmisi haiguse bioloogia ja uuritava geeni funktsionaalsuse kohta (Kosova ja Urbanek, 2013). Viimasel ajal on aga üha enam hakatud rõhku panema ülegenoomsele assotsiatsiooniuuringule, mis võimaldavad korraga uurida suurt hulka sagedasi geneetilisi variante kogu genoomi ulatuses (Marchini ja Howie, 2010). Lisaks ei eeldata neis uuringutes eelteadmisi leitud geenide funktsiooni kohta (Kosova ja Urbanek, 2013). Tänapäevaks läbi viidud mitmete *GWAS*-ide tulemuste analüüsimise järgselt on kokku leidnud rohkem kui 70 *PCOS*-i kandidaatgeeni. Probleem on aga selles, et paljud neist geenidest ei ole replitseerunud teistes uuringutes (Welt jt., 2012).

Chen jt. Hiina kohordi seas läbi viidud *GWAS* andis tulemuseks mitu *PCOS*-iga tugevalt assotsieerunud lookust (Chen jt., 2011). Hiljem viisid Shi jt. Hiina kohordis läbi uue *GWAS*-i ja meta-analüüsi ning kinnitasid tuvastatud lookuste seost *PCOS*-iga. Lisaks leiti veel 8 uut võimalikku assotsieerunud lookust (Shi jt., 2012). Welt jt. viisid Euroopa päritoluga kohordi seas läbi uuringu, et näha, kas Hiinas saadud tulemused Euroopa päritolu indiviidide seas samuti ilmnevad. Uuringu tulemused näitasid küll tugevat korrelatsiooni kahe geneetilise variandiga, mis mõlemad asusid geenis *DENNDIA*, kuid ülejäänud lookused Euroopa päritoluga inimeste hulgas antud uuringus ei replitseerunud (Welt jt., 2012). Selle põhjuseks võib olla asjaolu, et mõned ilmnunud assotsiatsioonidest on iseloomulikud just Hiina populatsioonile ja ei avaldu Euroopa päritoluga inimestel (Welt jt., 2012).

*DENNDIA* seost *PCOS*-iga on näidanud veel teisedki uuringud. Eriksen jt. uurisid *DENNDIA* geeni ning leidsid sealt kaheksa *PCOS*-iga assotsieerunud geneetilist markerit. Samas toodi välja, et tulenevalt väikesest valimi suurusest on leiu kinnitamiseks vajalik läbi viia täiendavaid uuringuid (Eriksen jt., 2013).

## **1.2. Ülegenoomsed assotsiatsiooniuuringud**

Juba mitukümmend aastat on edukalt suudetud kindlaks määrata geene, milles toimuvad mutatsioonid viivad mendeliaalsete ehk ühe geeni haigusteni nagu näiteks Huntingtoni tõbi või tsüstiline fibroos (Hardy ja Singleton, 2009). Vastupidiselt ühe geeni haigustele, on komplekshaigused mõjutatud mitmete geenide, keskkonnategurite ning nende omavahelistest koosmõjudest, mis kõik üksikuna võttes omavad haigusele väikest efekti (Hardy ja Singleton, 2009; Manolio, 2010). Lähiajalos on üha rohkem hakatud läbi viima ülegenoomseid assotsiatsiooniuuringuid leidmaks geneetilisi põhjuseid kompleks- ehk multifaktoriaalsetele

haigustele (Pearson ja Manolio, 2008). Komplekshaiguste geneetilise tagapõhja uurimise teebki keeruliseks asjaolu, et need on mõjutatud mitmete geenide poolt, mis kõik haigestumisel rolli mängivad (Hardy ja Singleton, 2009; Manolio, 2010).

*GWAS* on uuring, mille käigus püütakse leida assotsiatsioone geneetiliste markerite (enamasti ühenukleotiidiliste polümorfismide (*Single Nucleotide Polymorphism, SNP*)) ja haiguse vahel (Pearson ja Manolio, 2008). Reeglina teostatakse uuring, kasutades vähemalt 100 000 *SNP* markerit üle kogu genoomi (Welter jt., 2014). Klassikaliselt koosneb *GWAS* neljast osast: (1) valimi ehk haigete ning tervete kontrollindiviidide seleksioon, (2) genotüpiseerimine ning kvaliteedikontroll, (3) statistiline test assotsiatsioonide leidmiseks markeri ja haiguse vahel, (4) leitud assotsiatsioonide replitseerimine sõltumatus kohordis (Pearson ja Manolio, 2008). Alates aastast 2005 on läbi viidud hulgaliselt *GWAS*-e, mille kaudu on õnnestunud leida mitmeid lookuseid, mis on seoses erinevate komplekshaigustega (Burton jt., 2007; Sladek jt., 2007).

Edukate uuringute läbiviimiseks on oluline, et varasemalt oleks koostatud optimaalne uuringu kavand, mis komplekshaiguste uurimise korral võib osutada küllaltki keerukaks, sest silmas tuleb pidada mitmeid erinevaid aspekte, mis kõik analüüsi tulemusi mõjutavad (Spencer jt., 2009). Uuringu optimaalseks statistiliseks võimsuseks võiks olla vähemalt 80%, mida assotsiatsiooniuringute puhul saab vaadelda kui tõenäosust, et avastatakse tõene assotsiatsioon markeri ja haiguse vahel (Cardon ja Bell, 2001; Spencer jt., 2009). Wang jt. on välja toonud, et leidmaks 80%-lise statistilise tõenäosusega assotsiatsiooni markeri ja haiguse vahel, eeldusel, et harvema alleeli sagedus (*Minor Allele Frequency, MAF*) on  $< 10\%$  ning šansside suhe (*Odds Ratio, OR*) on  $< 1,3$ , oleks vajalik uuringusse kaasata valim, millesse kuulub vähemalt 10 000 juhtu ja sama palju kontrollindiviide (Wang jt., 2005).

### **1.2.1. Genotüübi imputatsioon**

Genotüübi imputatsioon on protsess, mille kaudu püütakse ennustada genotüüpe, mida ei ole indiviididel veel analüüsitud (Marchini ja Howie, 2010). Genotüübi imputatsiooni võib rakendada nii kogu genoomile kui ka mingile kindlale regioonile genoomis, kuid *GWAS*-i puhul kasutatakse imputeerimist terve genoomi ulatuses (Marchini ja Howie, 2010). Imputeerimisel kasutatakse genotüpiseeritud andmestikku, milles võib leiduda puuduoleva genotüübiga markereid. Kui sellise andmestikuga läbi viia *GWAS*, ei pruugi uuringust välja tulla olulised assotsiatsioonid. Selle ennetamiseks haplotüpiseeritakse olemasolevad andmed ja puuduolevate markerite genotüüpide leidmiseks kasutatakse erinevaid referentshaplotüüpe, mille tulemusena saadakse imputeeritud genotüübid. Imputeeritud genotüübiga

assotsiatsiooniuringute läbi viimine võib suurendada uuringute võimsust kuni 10%, mis aitab kaasa suurema hulga assotsiatsioonide ilmnemisele. Imputeerimisest on kasu ka meta-analüüside koostamisel, kui eesmärgiks on võrdsustada erinevate uuringute andmeid (Marchini ja Howie, 2010).

Genotüübi imputatsiooni on võimalik läbi viia näiteks programmiga IMPUTE, mis põhineb peidetud Markovi mudelil (*Hidden Markov Model, HMM*) (Howie jt., 2009). *HMM* on statistikal põhinev mudel, mis võimaldab omavahel siduda varasemalt täheldatud protsessi genoomis senitundmatu, kuid huvipakkuva protsessiga (Yoon, 2009). Referentshaplotüüpe on imputeerimise läbiviimiseks võimalik saada HapMap2, HapMap3 ja 1000 Genoomi Projekti referentspaneelidest (Marchini ja Howie, 2010).

### **1.2.2. Ülegenoomsete assotsiatsiooniuringute probleemid**

*GWA* uuringud on osutunud edukaks lähenemisviisiks leidmaks seoseid geneetiliste variantide ning erinevate haiguste vahel (Manolio, 2010). Hoolimata *GWAS* eduloost, leidub sel mõningaid murekohti, millest suurimateks on loetud populatsiooni stratifikatsiooni, mis võib kaasa tuua vale-positiivseid tulemusi ja väike valimi suurus, millest tulenevalt võib uuringu võimsus jääda liiga väikseks ning olulised assotsiatsioonid jääda märkamatuks (Hardy ja Singleton, 2009; McCarthy ja Hirschhorn, 2008).

Populatsiooni stratifikatsioon on alleelisageduste erinevus kontrollide ja juhtude vahel, mis võib eelkõige tuleneda erinevast põlvnemisest. Suuremahuliste assotsiatsiooniuringute juures on tihtilugu probleemiks asjaolu, et teadmata on populatsiooni struktuur ja sellest tulenev stratifikatsioon. Kui aga valimis leidub sisemine struktuur võib see viia vale-positiivsete tulemusteni või takistada tõelise assotsiatsiooni leidmist (Freedman jt., 2004). Selle ärahoidmiseks viiakse valmi seas läbi kohandamine nende tunnuste hulgas (näiteks vanus, sugu, etniline grupp), mis võivad fenotübile mõju avaldada (Freedman jt., 2004; Marchini jt., 2004).

Populatsiooni kohandamise juures on oluline hoolikalt jälgida, et ei toimuks nii öelda ülekohandamist, mis võib langetada uuringu statistilist jõudlust. Populatsiooni struktuuri hindamiseks üks olemasolev meetod on genoomne kontroll (*Genomic Control*), mille kaudu on võimalik vähendada vale-positiivsete tulemuste hulka (Devlin jt., 2001). Genoomse kontrolli võtmeks on *genomic inflation* faktor  $\lambda$ , mille väärtus sõltub uuritava populatsiooni struktuurist. Genoomse kontrolli eelduseks on, et juhud ja kontrollid pärinevad samast populatsioonist, sest sellisel juhul on faktor  $\lambda$  väärtus kogu genoomi ulatuses muutumatu (Freedman jt., 2004). Faktor  $\lambda$  väärtus arvutatakse  $\chi^2$  testi statistiku mediaani jagamisel

eeldatava  $\chi^2$  testi mediaanväärtusega, mille väärtus on 0,456. Kui arvutuse tulemusena on  $\lambda$  väärtus suurem kui üks, siis esineb populatsioonis stratifikatsioon (Price jt., 2006). Genoomse kontrolli läbiviimisel on oluline, et kaasatud oleks piisaval hulgal lookuseid, sest vastasel juhul võib populatsiooni struktuur jääda ebaadekvaatselt kohandatuks (Devlin jt., 2001; Freedman jt., 2004; Marchini jt., 2004).

Lisaks populatsiooni stratifikatsioonile on *GWAS*-ide juures tihti probleemiks asjaolu, et haigusega seotud geneetiliste variantide esinemissagedused on küllaltki väikesed, mis omakorda tähendab, et *GWAS*-i teostamisel peab valim nende variantide tuvastamiseks olema üpriski suur (Hardy ja Singleton, 2009). Juhul kui valim on liialt väike, võib uuringu võimsus assotsieerunud markerite leidmiseks jääda liialt madalaks (Winkler jt., 2014). Selleks, et lahendada väikesest valimist tulenevat probleemi, on üha enam hakatud läbi viima ülegenoomsete assotsiatsiooniuringute meta-analüüsi (*Genome – Wide Association Meta Analysis, GWAMA*), milles on kombineeritud mitmete väiksemate *GWAS*-ide statistilised tulemused, saades tulemuseks analüüsi, millel on suurem statistiline võimsus (Winkler jt., 2014).

Lähtuvalt valimist on *GWAS*-i juures veel üks murekoht. Nimelt tulenevalt uuringu iseloomust, kus võrreldakse terveid indiviide haigetega, on oluline, et juhtude määramisel ei esineks vigu nende klassifitseerimisel, mis võib taaskord vähendada uuringu võimsust ja tingida olukorra, kus uuringust ei ilmne tõeseid assotsiatsioone (Pearson ja Manolio, 2008). Sellest johtuvalt on oluline, et uuringutes kasutatavad juhud oleksid diagnoositud ühest ja samast usaldusväärsest diagnoosikriteeriumist lähtuvalt (Pearson ja Manolio, 2008).

Valimist selekteerimisest järgmine samm *GWAS*-i juures on genotüpiseerimine, mille juures võib samuti ilmned mõningaid probleeme, mille hulka kuuluvad: genotüpiseerimise õnnestumise protsent, mida määratakse *Call Rate (CR)* abil; testitava geneetilise markeri minoorse alleeli sagedus ja Hardy – Weinberg tasakaalustamatus (Cardon ja Bell, 2001; Pearson ja Manolio, 2008).

Assotsiatsiooniuringute piiranguks on veel asjaolu, et analüüsi tulemusena identifitseeritakse *SNP* ja genoomne asukoht, kuid uuring ei võimalda saada otsest informatsiooni võimalikus assotsiatsioonis oleva geeni ja selle funktsiooni(de) kohta (Pearson ja Manolio, 2008). See ei ole aga probleemiks, kui on assotsieerunud *SNP*-de kohta on teada nende võimalikud mõjud geeniekspressioonis (Pearson ja Manolio, 2008).

## 2. Uurimus

### 2.1. Töö eesmärk

Käesoleva bakalaureusetöö praktilise osa eesmärgiks oli polütsüstiliste munasarjade sündroomiga assotsieerunud geneetiliste variantide leidmine kasutades professor Andres Salumetsa kogutud DNA proove koos Tartu Ülikooli Eesti Geenivaramu andmestikuga ning saadud tulemuste replitseerimine sõltumatus kohordis. Töö teiseks eesmärgiks oli varasemalt teadaolevate markerite p-väärtuste hindamine *EGCUT* andmetega teostatud analüüsis.

### 2.2. Materjalid ja meetodika

#### 2.2.1. Valim

Uuringu raames viidi läbi juht – kontroll uuring, mille valim koosnes 169 *PCOS*-i patsiendist ning 7001 tervest kontrollindiviidist. Töös kasutati Tartu Ülikooli Eesti Geenivaramu proove, mis olid kogutud kõigist Eesti maakondadest. Uuringus olevad *PCOS* patsiendid valiti *ESHRE/ASRM* ehk Rotterdami kriteeriumile toetudes, kus *PCOS* määratakse juhul kui indiviidil esineb vähemalt kaks kolmest järgnevast sümptomist: (1) kliinilised ja/või biokeemilised märgid hüperandrogenismist, (2) oligo-ovulatsioon ja/või anovulatsioon, (3) polütsüstilised munasarjad. Hüperandrogenism määrati, kui Ferriman ja Gallwey skoor andis tulemuseks 8 ja rohkem. Oligo-ovulatsioon määrati, kui menstruaaltsükli pikkuseks oli  $\geq 35$  päeva ning kui menstruaaltsükli ei olnud esinenud  $> 3$  kuu vältel eeldusel, et patsient ei olnud rase. Tsüstide esinemine munasarjas määrati kindalaks ultraheliuuringu käigus.

Kontrollgruppi kuulusid terved naised, kellel oli regulaarne menstruaaltsükkel (26 – 35 päeva) ning ei esinenud hüperandrogenismi ega polütsüstilisi munasarju. Lisaks kasutati kontrollidena ka mehi. Uuringus osalenud indiviidide arv, ealine ja sooline jaotus ning keskmine kehamassiindeks on esitatud tabelis 2.

**Tabel 2 Uuringus osalenud indiviidide omadused.** Tabelis on toodud valimi arv, naiste osakaal haigete ja kontrollide hulgas, ealine jaotuvus ning valimi keskmine kehamassiindeks.

	Juhud	Kontrollid
<b>Indiviidide arv</b>	169	7001
<b>Naiste protsent</b>	100%	56%
<b>Keskmine vanus (standardhälve)</b>	30,1 (8,5)	51,5 (20,1)
<b>Keskmine kehamassiindeks (standardhälve)</b>	26,2 (6,7)	26,7 (5,3)

Kõik uuringus osalevad doonorid on andnud kirjaliku nõusoleku teadustöös osalemiseks ning uuring on kooskõlastatud Tartu Ülikooli inimuuringute eetika komiteega.

### 2.2.2. Genotüpiseerimine ja kvaliteedikontroll

Kokku genotüpiseeriti 172 juhtu ja kasutati 8131 eelnevalt genotüpiseeritud kontrollindiviidi. Patsientide genotüpiseerimisel kasutati Illumina HumanCoreExome BeadChip (Illumina Inc., San Diego, CA, USA) kiipi ning kontrollid olid genotüpiseeritud Illumina HumanOmniExpress BeadChip kiibiga. Genotüpiseerimine viidi läbi Eesti Geenivaramu tuumiklaboris. Analüüsiks kasutati kahe genotüpiseerimispaneeli kokkulangevaid markereid, mida oli kokku 234 962.

Kvaliteedikontroll genotüpiseeritud andmetele viidi läbi järgnevate tunnussuuruste osas:

- harvema alleeli sagedus
- edukalt genotüpiseeritud indiviidide ja *SNP*-de protsentuaalne osakaal markeril (*Call Rate, CR*)
- Hardy – Weinbergi tasakaalustatus
- Tartu Ülikooli Eesti Geenivaramu proovide fenotüübi soo mitte-kokkulangevus genotüübi järgi leitud sooga
- indiviidide sugulus
- heterosügootsus

Uuringutest jäid välja indiviidid, kelle *CR* oli alla 95% ning inimesed, kelle fenotüübis märgitud sugu ei vastanud genotüübist ilmnenule soole. Indiviidide vaheline sugulus leiti ühiste alleelide osatähtsuse (*Identity by State, IBS*) hindamise järgi genoomis, millele toetudes omakorda ennustati iga sugulaspaari päritolult identse (*Identical by Descent, IBD*) DNA osatähtsust. Juhul kui *IBD* osatähtsust kahe inimese vahel oli  $\geq 10\%$  genoomist, eemaldati analüüsist üks sugulastest. Lisaks jäid uuringust välja *SNP*-d, millel oli puudulik  $CR \leq 95\%$ ,  $MAF \leq 1\%$  ja *HWE*  $p$ -väärtus  $\leq 10^{-5}$ . Kvaliteedikontrolli läbisid 169 juhtu, 7001 kontrolli ja 228 922 *SNP*-i, mida seejärel assotsiatsiooniuuringu läbiviimiseks kasutati. Andmete kvaliteedikontrolli läbiviimiseks kasutati arvutiprogrammi PLINK (Purcell jt., 2007).

### 2.2.3. Imputeerimine

Imputeerimise tarbeks muudeti tööriista GTOOL (<http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html>) abil LINKAGE vormis



olevad genotüpiseerimisandmed IMPUTE vormi. Selleks, et imputeerida markereid, mida ei olnud Illumina HumanOmniExpress ja Illumina HumanCoreExome kiibi ühisosas, kasutati referentsgenoomina 1000 Genoomi Projektist (*1000 Genomes Phase I integrated haplotypes, Dec. 2013*) pärinevaid andmeid. Imputeerimiseks kasutati programmi IMPUTE v2 (Howie jt., 2009). Imputeerimise järgselt eemaldati analüüsist *SNP*-d *MAF*-i ja imputeerimise kvaliteediskoori (*proper\_info*) alusel. Genotüpiseerimise ja imputeerimise järgselt jäi andmestikku genotüpiseeritud markereid 228 922 ja lisaks imputeeritud markereid 7 042 499.

#### **2.2.4. Assotsiatsioonianalüüs**

Assotsiatsioone geneetiliste markerite ja haiguse vahel analüüsiti programmi SNPTEST v2 kasutades (Marchini ja Howie, 2010). Geneetiliste markerite seost *PCOS*-iga analüüsiti kasutades logistilise regressiooni alleeliefektide aditiivset mudelit. Analüüsides kasutati kovariaatidena proovide suguluste maatriksi põhjal leitud nelja esimest peakomponenti, et kahandada võimalikku populatsiooni stratifikatsiooni. Mitmesest testimisest tulenevate vigade vältimiseks, loetakse analüüsis ülegenoomselt statistiliselt oluliseks markereid, mille *p*-väärtus oli  $< 5 \times 10^{-8}$  (Pe'er jt., 2008). Analüüs viidi läbi rahvusvahelise *PCOS* konsortsiumi analüüsiplaani kohaselt.

#### **2.2.5. Replikatsioonianalüüs**

Geneetilised markerid, mis *GWA*-uuringu käigus olid tugevas assotsiatsioonis ( $p$ -väärtus  $\leq 1 \times 10^{-6}$ ), valiti välja ning saadeti replikatsioonianalüüsiks Oxfordi Ülikooli *Wellcome Trust* Inimgeneetika Instituuti (*Wellcome Trust Centre for Human Genetics, University of Oxford, WTCHG*). Kokku telliti replikatsioon 30 markeri kohta. Replikatsioonianalüüsi valimis oli kokku 3542 indiviidi, kellest 687 juhtu ja 2855 kontrolli, kes olid valitud samade kriteeriumite alusel, mida kasutati põhianalüüsis. Replikatsioonianalüüsis rakendati samasugust logistilise regressiooni mudelit, mida põhianalüüsiski, kasutades kovariaatidena samuti suguluse põhjal arvutatud peakomponente. Replikatsioonianalüüsi genotüpiseerimine viidi läbi Illumina HumanCoreExome kiibiga.

#### **2.2.6. Varasemalt teadaolevate markerite võrdlus**

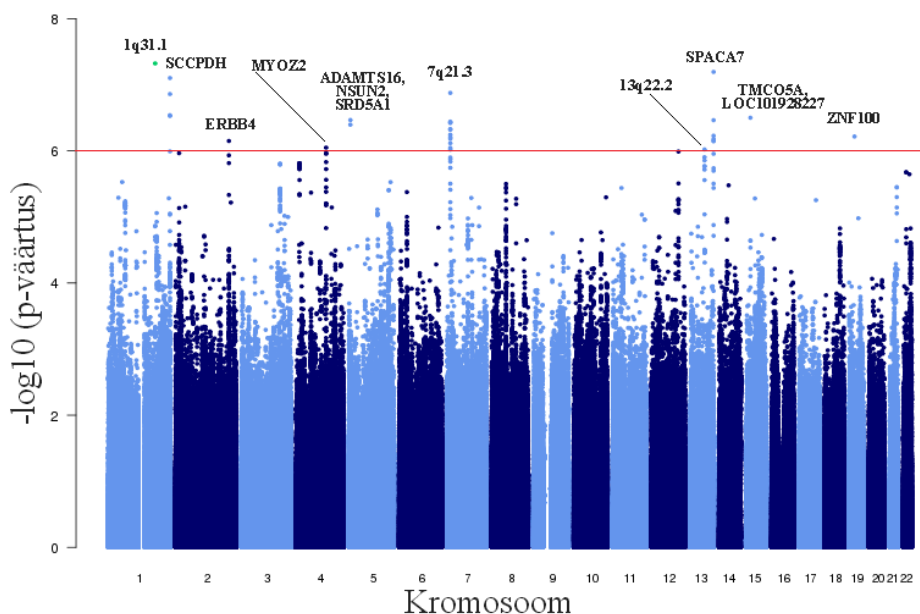
Töö käigus otsiti kirjandusest markereid, mida on juba varasemalt seostatud *PCOS*-iga. Kokku leiti edasiseks analüüsiks 33 teadaolevat markerit kuuest erinevast kirjandusallikast (Amal jt., 2013; Cui jt., 2013; Eriksen jt., 2013; Hwang jt., 2012; Shi jt., 2012; Wang jt., 2013). Seejärel otsiti vastavad markerid välja *EGCUT* andmetega teostatud analüüsist ning kontrolliti nende markerite assotsiatsioonide *p*-väärtusi.

## 2.3. Tulemused

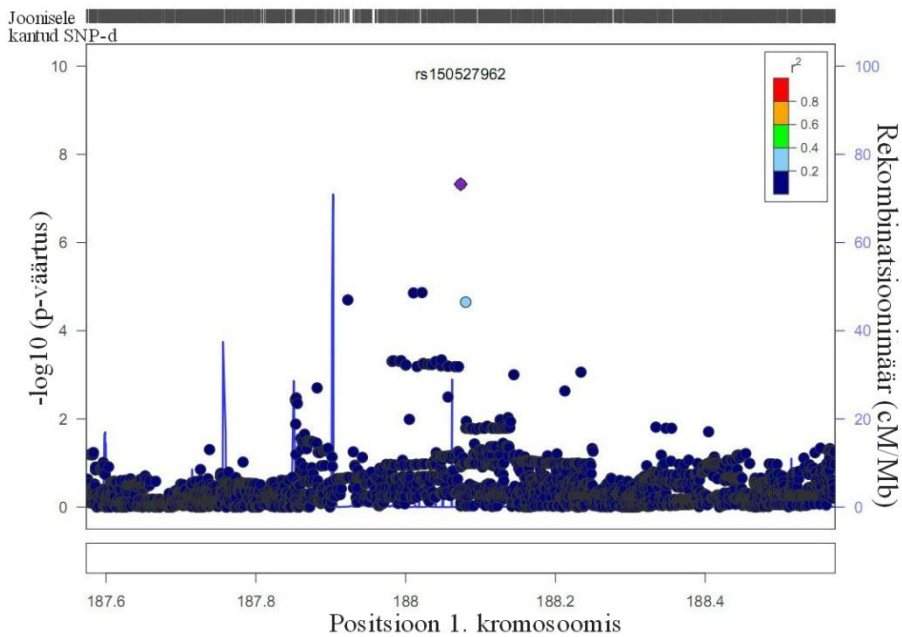
Antud töö raames otsiti ülegenoomse assotsiatsiooniuringus geneetilisi variante, mis oleksid seostunud *PCO* sündroomiga. Selleks kasutati logistilise regressiooni mudelit, kus kovariaatidena kasutati proovide omavahelise suguluse põhjal arvatud nelja esimest peakomponenti. *EGCUT*-s läbi viidud analüüsi käigus identifitseeriti kolmkümmend markerit  $p$ -väärtusega  $< 1 \times 10^{-6}$ , kokku kümnest genoome lookusest (Joonis 1, Tabel 3), mis valiti välja sõltumatuks replikatsioonianalüüsiks. Enamik leitud regioonidest asus lähimast geenist vähem kui 100 kb kaugusel. Erandiks oli kolm regiooni: 1q31.1, 7q21.3 ja 13q22.2. Geenide sees asusid markerid kuuest analüüsitud regioonist: kromosoomides 1 (geen *SCCPDH*), 2 (geen *ERBB4*), 13 (geen *SPACA7*) ja 19 (geen *ZNF100*). Geeniregioonidest kaugemal kui 100 kb asusid neliteist markerit kolmest regioonist: kromosoomides 1, 7 ja 13.

Geenialleelisageduste põhjal ilmnas analüüsist, et leitud assotsiatsioonidest kolm olid harvad variandid,  $MAF < 5\%$  (kromosoomides 4, 15 ja 19) ning seitse regiooni sisaldas sagedasi variante (kromosoomides 1, 2, 5, 7 ja 13).

Lookuses 1q31.1 asuv marker rs150527962 oli ainus, mille  $p$ -väärtus andis ülegenoomselt olulise tulemuse:  $p$ -väärtus =  $4,76 \times 10^{-8}$ . Tegemist oli imputeeritud harva variandiga ( $MAF = 1\%$ ,  $proper\_info = 0,81$ ) geenivabast piirkonnast, mille lähedal ei asu teisi tugevalt assotsieerunud markereid (Joonis 2).

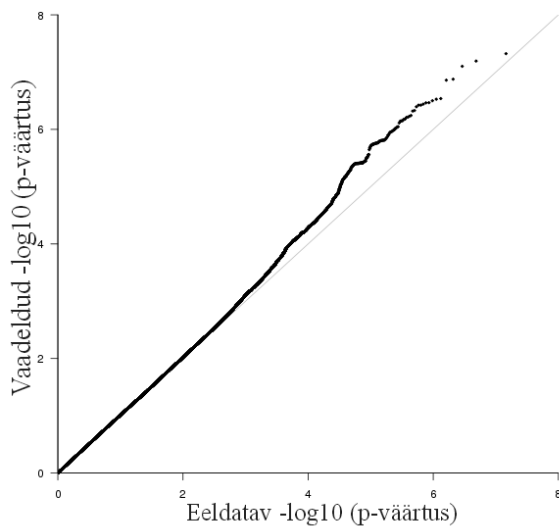


**Joonis 1** PCOS-i assotsiatsiooniuringu Manhattan joonis. X-teljel on assotsieerunud markerite genoomsed positsioonid ning Y-teljel nende  $p$ -väärtuste negatiivne kümnendlogaritm. Punane horisontaaljoon märgib  $p$ -väärtust  $1 \times 10^{-6}$ .



**Joonis 2 Markeri rs150527962 regionaaljoonis.** Ülegenoomse olulisusega marker rs150527962 on märgitud lilla rombina, teised markerid on märgitud ringidena. Markerite värvusega on esitatud nende *LD* markeriga rs150527962. X-teljel on esitatud markerite füüsilised asukohad 1. kromosoomis ja Y-teljel on markerite p-väärtuste negatiivsed kümnendlogaritmid. Sinine joon märgib rekombinatsioonimäära.

Hindamaks populatsiooni stratifikatsiooni kasutati genoomse kontrolli meetodit. Parameetri  $\lambda$  hinnanguks oli 1,03 (Joonis 3), mis näitab, et analüüsitavas andmestikus ei olnud olulist stratifikatsiooni.

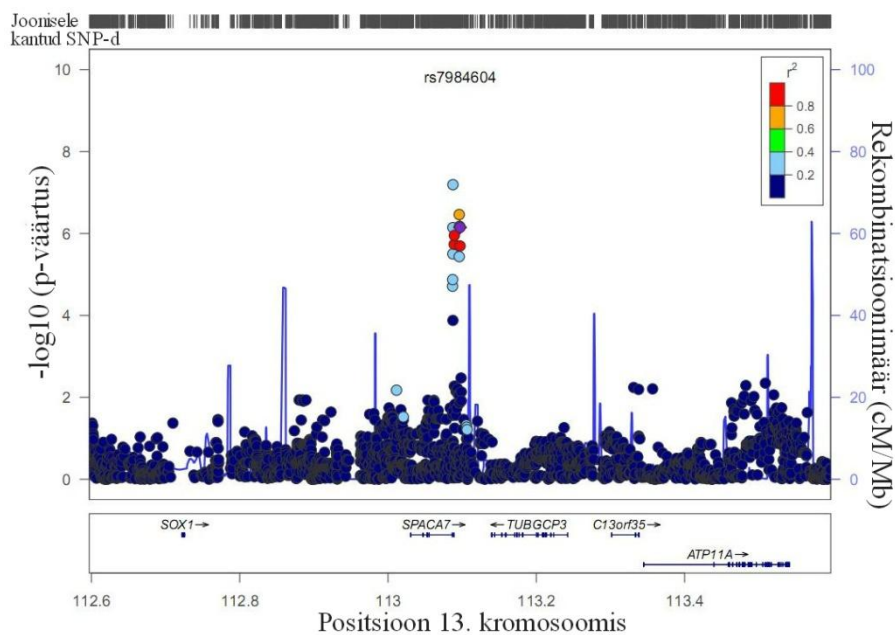


**Joonis 3 Eeldatud ja vaadeldud assotsiatsioonide p-väärtuste kvantilide joonis.** X-teljel on toodud eeldatav p-väärtuste negatiivne kümnendlogaritm ning Y-teljel vaadeldud p-väärtuste negatiivne kümnendlogaritm.

Kolmkümmend markerit, mille p-väärtused olid alla  $1 \times 10^{-6}$ , replitseeriti sõltumatus uuringus Oxfordis *WTCHG* keskuses, kasutades sama analüüsiplaani. Replikatsioonianalüüsis osutusid

statistiliselt oluliseks peale mitmese testimise korrektsiooni ( $p$ -väärtus  $< 5/30$ ) kümnest markerist neli, millest üks osutus hilisemal analüüside kombineerimisel ka ülegenoomselt oluliseks. Kõik neli markerit (chr13:113090656:D, rs9577365, rs9577367, rs7984604) asusid kolmeteistkümnendas kromosoomis ning paiknesid geeni *SPACA7* (*Sperm Acrosome Associated 7*) läheduses.

Meta-analüüsi järgselt, kui kombineeriti *EGCUT* ja *WTCHG* uuringu tulemused, osutus ülegenoomselt statistiliselt oluliseks marker rs7984604 *SPACA7* geeniregioonist (Joonis 4, Tabel 3). Markeri šansside suhe kombineeritud analüüsis oli 1,69, 95%-lise usaldusintervalliga 1,40 – 2,04 ning  $p$ -väärtus oli  $4,04 \times 10^{-8}$ . *EGCUT* ja *WTCHG* eraldiseisvate analüüside tulemustes olid rs7984604  $p$ -väärtused vastavalt *EGCUT* analüüsis  $7,04 \times 10^{-7}$  ja *WTCHG* analüüsis  $3,10 \times 10^{-4}$ . Esimeses kromosoomis asuv marker rs150527962 ei replitseerunud, ning analüüside kombineerimisel jäi selle  $p$ -väärtuseks  $4,13 \times 10^{-6}$ .



**Joonis 4 Markeri rs7984604 regionaaljoonis.** *EGCUT* ja *WTCHG* kombineeritud analüüsis ülegenoomse olulisusega marker rs7984604 on märgitud lilla rombina, teised markerid on märgitud ringidena. Markerite värvusega on esitatud nende *LD* markeriga rs7984604. X-teljel on esitatud markerite füüsilised asukohad 13. kromosoomis ja Y-teljel markerite  $p$ -väärtuste negatiivsed kümnendlogaritmid. Sinine joon märgib rekombinatsioonimäära. Joonise all on näidatud regioonis asuvate geenide asukohad. Marker rs7984604 ei olnud antud regioonis enne replikatsiooni kõige madalama  $p$ -väärtusega ent muutus selleks replikatsioonijärgselt.

**Tabel 3** PCOS-i assotsiatsiooniuuringu tulemused. Lähim geen määrati, kui selle kaugus assotsieerunud markerist oli < 100 kb. Paksus kirjas on välja toodud markerid ülegenoomselt olulise p-väärtusega < 5 x 10<sup>-8</sup> ning paksus kursiivis on märgitud replikatsioonimarkerid, mis olid statistiliselt olulised peale mitmese testimise korrigeerimise (p-väärtusega < 0,05/30 = 0.0017). \* CI märgib šansside suhte 95%-list usaldusintervalli (*Confidence Interval, CI*).

Marker	Kromosoom	Positsioon	Lähim geen	Efekti/teine alleel	Efekti alleeli sagedus	EGCUT analüüs	WTCHG analüüs	Kombineeritud analüüs
						OR (95% CI*) P-väärtus	OR (95% CI*) P-väärtus	OR (95% CI*) P-väärtus
rs150527962	1	188073526	1q31.1	G/A	0,01	29,14 (8,69-97,77) <b>4,76 x 10<sup>-8</sup></b>	1,17 (0,20-6,9) 0,86	10,51 (3,86-28,58) 4,13 x 10 <sup>-6</sup>
rs61852473	1	246887131	SCCPDH	T/C	0,03	5,72 (2,99-10,94) 1,38 x 10 <sup>-7</sup>	1,09 (0,75-1,58) 0,65	1,63 (1,19-2,18) 0,0027
rs140725729	1	246901319	SCCPDH	A/G	0,04	4,47 (2,52-7,93) 2,90 x 10 <sup>-7</sup>	1,07 (0,75-1,54) 0,70	1,61 (1,19-2,18) 0,0022
rs61852486	1	246913471	SCCPDH	T/C	0,04	4,46 (2,52-7,91) 2,97 x 10 <sup>-7</sup>	1,10 (0,77-1,58) 5,85 x 10 <sup>-1</sup>	1,63 (1,20-2,20) 0,0015
rs144230454	1	246970689	SCCPDH	T/C	0,03	6,63 (3,32-13,23) 7,91 x 10 <sup>-8</sup>	1,04 (0,72-1,50) 0,84	1,56 (1,13-2,16) 0,007
rs35180566	2	213135048	ERBB4	C/T	0,07	2,89 (1,90-4,41) 7,10 x 10 <sup>-7</sup>	1,15 (0,93-1,42) 0,18	1,38 (1,15-1,67) 6,88 x 10 <sup>-4</sup>
rs1809264	4	120051252	MYOZ2	G/A	0,55	0,57 (0,46-0,72) 8,96 x 10 <sup>-7</sup>	0,94 (0,82-1,06) 0,31	0,83 (0,74-0,93) 8,54 x 10 <sup>-4</sup>
rs187356748	5	5988160	ADAMTS16, NSUN2, SRD5A1	T/C	0,01	13,16 (4,86-35,64) 4,03 x 10 <sup>-7</sup>	0,99 (0,36-2,78) 0,10	3,76 (1,84-7,69) 2,84 x 10 <sup>-4</sup>

rs140230529	5	6003468	ADAMTS16, NSUN2, SRD5A1	T/C	0,01	13,48 (4,96-36,63) $3,42 \times 10^{-7}$	0,89 (0,33-2,38) 0,82	3,39 (1,68-6,83) $6,47 \times 10^{-4}$
chr7:1335774 4:D	7	13357744	7q21.3	C/CAAT	0,20	2,03 (1,54-2,68) $4,69 \times 10^{-7}$	0,94 (0,82-1,09) 0,41	1,11 (0,98-1,26) 0,11
rs4721213	7	13358803	7q21.3	A/G	0,20	1,97 (1,50-2,58) $9,58 \times 10^{-7}$	0,94 (0,82-1,09) 0,44	1,11 (0,98-1,26) 0,11
rs17713172	7	13362581	7q21.3	G/C	0,20	2,02 (1,53-2,66) $5,72 \times 10^{-7}$	0,95 (0,82-1,10) 0,49	1,12 (0,98-1,27) 0,09
rs17713411	7	13368437	7q21.3	G/A	0,20	2,01 (1,52-2,66) $9,02 \times 10^{-7}$	0,94 (0,81-1,09) 0,41	1,11 (0,97-1,26) 0,12
rs17713738	7	13378131	7q21.3	C/A	0,20	2,03 (1,54-2,68) $6,20 \times 10^{-7}$	0,95 (0,82-1,09) 0,45	1,11 (0,98-1,26) 0,11
rs148365756	7	13391814	7q21.3	T/C	0,17	2,28 (1,68-3,10) $1,33 \times 10^{-7}$	0,93 (0,80-1,09) 0,37	1,12 (0,97-1,29) 0,11
rs17657354	7	13391997	7q21.3	A/C	0,17	2,20 (1,62-2,98) $3,77 \times 10^{-7}$	0,93 (0,80-1,08) 0,36	1,11 (0,97-1,27) 0,14
rs62445824	7	13392242	7q21.3	A/G	0,17	2,20 (1,63-2,99) $3,62 \times 10^{-7}$	0,93 (0,80-1,09) 0,37	1,11 (0,97-1,27) 0,14
rs17657390	7	13392263	7q21.3	C/T	0,17	2,20 (1,62-2,98) $3,76 \times 10^{-7}$	0,93 (0,80-1,08) 0,36	1,11 (0,97-1,27) 0,14
rs4141998	7	13394623	7q21.3	G/A	0,37	1,81 (1,43-2,29) $7,69 \times 10^{-7}$	0,98 (0,87-1,12) 0,79	1,13 (1,01-1,26) 0,03
chr7:1340305 5:D	7	13403055	7q21.3	A/AC	0,20	2,10 (1,58-2,81) $4,78 \times 10^{-7}$	0,94 (0,81-1,09) 0,39	1,11 (0,97-1,27) 0,12

rs2262866	7	13421173	7q21.3	C/G	0,30	1,90 (1,47-2,44) $6,60 \times 10^{-7}$	0,96 (0,84-1,09) 0,52	1,11 (0,99-1,25) 0,08
rs4885393	13	76871760	13q22.2	A/T	0,06	3,19 (2,01-5,08) $9,53 \times 10^{-7}$	1,25 (0,95-1,65) 0,11	1,60 (1,26-2,02) $1,10 \times 10^{-4}$
rs7332708	13	113087440	SPACA7	C/G	0,15	2,19 (1,61-2,99) $7,15 \times 10^{-7}$	1,18 (0,99-1,41) 0,07	1,37 (1,18-1,60) $5,54 \times 10^{-5}$
rs145924721	13	113087583	SPACA7	G/A	0,14	2,49 (1,79-3,46) $6,40 \times 10^{-8}$	1,22 (1,02-1,47) 0,03	1,44 (1,23-1,69) $7,01 \times 10^{-6}$
chr13:113090656:D	13	113090656	SPACA7	T/TA	0,08	2,78 (1,86-4,15) $5,92 \times 10^{-7}$	1,41 (1,14-1,73) <b>0,0015</b>	1,63 (1,35-1,96) $2,93 \times 10^{-7}$
rs9577365	13	113096019	SPACA7	T/C	0,07	3,25 (2,07-5,12) $3,44 \times 10^{-7}$	1,46 (1,18-1,82) <b><math>6,48 \times 10^{-4}</math></b>	1,70 (1,40-2,07) $1,29 \times 10^{-7}$
rs9577367	13	113096464	SPACA7	T/G	0,14	2,22 (1,62-3,03) $6,65 \times 10^{-7}$	1,31 (1,09-1,56) <b>0,0036</b>	1,49 (1,27-1,74) $5,97 \times 10^{-7}$
rs7984604	13	113097084	SPACA7	G/A	0,09	2,79 (1,86-4,18) $7,06 \times 10^{-7}$	1,48 (1,19-1,82) <b><math>3,10 \times 10^{-4}</math></b>	1,69 (1,40-2,04) <b><math>4,04 \times 10^{-8}</math></b>
rs7177901	15	38319711	TMCO5A, LOC101928227	G/A	0,55	0,56 (0,45-0,70) $3,16 \times 10^{-7}$	0,95 (0,82-1,10) 0,46	0,81 (0,72-0,91) $5,96 \times 10^{-4}$
rs7252798	19	21908053	ZNF100	T/C	0,45	1,77 (1,41-2,21) $6,08 \times 10^{-7}$	0,96 (0,83-1,12) 0,60	1,16 (1,02-1,31) 0,02

Töö raames analüüsiti lisaks veel senileitud *PCO* sündroomiga seostatud markereid *EGCUT* andmetest. Kokku leiti kirjandusest 33 markerit, millest 30 olid olemas *EGCUT* andmestikus. Ükski nendest markeritest ei näidanud olulist p-väärtust peale mitmese testimise korrektsiooni, ent kaks uuritavatest markeritest olid p-väärtusega  $< 0,05$  (Lisa 1). Üks kahest leitud markerist rs1894116 jäi kromosoomi 11 ning asus geenis *YAPI*. Markeri p-väärtus oli 0,03 ning šansside suhe meie analüüsis 1,48, 95%-lise usaldusintervalliga 1,03 – 2,13. Teine marker rs4784165 paiknes kuueteistkümnendas kromosoomis, kuid ei asunud geenis, lähimaks geeniks oli *TOX3*. Markeri p-väärtus oli 0,04 ning šansside suhe 1,31, 95%-lise usaldusintervalliga 1,00 – 1,71.

## 2.4. Arutelu

Varasemalt on mitmed uuringud kinnitanud, et polütsüstiliste munasarjade sündroomil on olemas geneetiline komponent (Franks jt., 1997; Govind jt., 1999). Antud töö raames viidi sarnaselt eelnevatele *GWAS*-idele läbi hüpoteesivaba ülegenoomne uuring Eesti populatsioonis, kus valimi hulka kuulus 169 haiget ning 7001 kontrolli. Uuriti, kas Eesti andmeid kasutades leitakse lisaks juba teadaolevatele lookustele uusi *PCOS*-iga assotsieerunud lookuseid. Uuringu tulemustes keskenduti markeritele, mille p-väärtus jäi alla  $1 \times 10^{-6}$ . Selliseid markereid leiti kolmkümmend, kokku kümnest erinevast lookusest, mis seejärel saadeti sõltumatuks replikatsioonianalüüsiks Oxford *WTCHG*-i. Lisaks otsiti kirjandusest seni teadaolevaid markereid ning uuriti, kas ja mil määral need avalduvad Eesti populatsiooni hulgas läbiviidud uuringus.

*EGCUT* analüüsitulemusena leiti 2 olulist regiooni, kromosoomides 1 ja 13. Esimeses kromosoomis asunud marker rs150527962 andis koheselt ülegenoomselt olulise p-väärtuse  $4,76 \times 10^{-8}$ . Antud marker ei asunud üheski geenis, samuti ei leidunud genee selle läheduses. Lisaks ei replitseerunud rs150527962 Oxford *WTCHG* analüüsis, mis aga ei tähenda, et tegemist on kindlasti vale-positiivse tulemusega. Mittereplitseerumise põhjus võib peituda asjaolus, et tegemist oli harva variandiga, mis võis olla Eesti kohordispetsiifiline ning ei andnud seetõttu olulist p-väärtust teistes kohortides. Markeri rs150527962 osalus *PCOS*-i geneetilise tagapõhja juures vajaks edasist uurimistööd.

Teine oluline leitud regioon asus kolmeteistkümnendas kromosoomis, kus paiknes marker rs7984604, mis *EGCUT* analüüsis andis p-väärtuseks  $7,06 \times 10^{-7}$  ning replikatsioonianalüüsiga kombineerides muutus ülegenoomselt oluliseks p-väärtusega  $4,04 \times 10^{-8}$ . Marker ei asunud geeni sees, kuid paiknes geeni *SPACA7* läheduses. *SPACA7*-s ja selle



läheduses asus veel teisigi markereid, mis olid analüüsis olulise p-väärtusega. Sellest tulenevalt võib *SPACA7*-t lugeda huvipakkuvaks *PCOS*-i kandidaatgeeniks.

*SPACA7* on esmakordselt antud nimega määratletud Korfanty jt. töös, kus see identifitseeriti kui spermi akrosoomi valk, mis ekspresseerub vaid testistes ning mille ortoloogid on olemas näiteks inimestes, šimpansides ja hiirtes (Korfanty jt., 2012; Nguyen jt., 2014). Korfanty jt. leidsid oma töös, et *SPACA7*-l esineb märgatavaid järjestuse erinevusi võrrelduna teiste *SPACA* perekonna liikmetega, kuid tänini ei olda kindlad selle täpses funktsioonis. Välja on toodud, et küpsed *SPACA7* valgud võivad omanda proteolüütilist aktiivsust, mis võib viia raku surmani (Korfanty jt., 2012). *SPACA7* asumine akrosoomis on andnud põhjust arvata, et *SPACA7* võib omada olulist rolli reproduktiivsuse juures. Akrosoomi puhul on tegu membraanseoselise organelliga, mis katab spermatoosoidi pead ning millesse on akumulunud hüdrofüütilised ensüümid, mis omakorda on vajalikud selleks, et spermatoosoid suudaks tungida läbi munaraku membraani (Primakoff ja Myles, 2002). Nguyen jt., töötasid oma uuringus välja antikeha (anti-*SPACA7* IgG), et selle abil uurida *SPACA7*-e funktsiooni. Tulemuseks saadi, et anti-*SPACA7* ei pärssinud spermatoosoidi penetratsiooni ega viljastumist juhul kui puudusid *Cumulus* rakud. Sellest järeldati, et *SPACA7* võib viljastumist hõlbustada, vallandades *Cumulus* rakkudest teguri/tegureid, mis suurendavad spermatoosoidi pädevust viljastumise protsessis (Nguyen jt., 2014).

Publitseeritud artiklitest leiti 33 teadaolevat assotsieerunud markerit, millest 30 olid olemas ka *EGCUT* uuringu tulemustes. Neist markeritest kaks olid *EGCUT* analüüsis nominaalse p-väärtusega alla 0,05 (markerid rs1894116 ja rs4784165). Mõlemad markerid olid varasemalt kirjeldatud *PCOS*-iga assotsieeruvateks Shi jt. uuringus, kus valim koosnes 1510-st haigest ja 2016 kontrollindiviidist, kes kõik olid Hiina päritolu (Shi jt., 2012). Üks kahest markerist asus geenis *YAP1* (*Yes-associated protein 1*) ning teine geeni *TOX3* (*TOX high mobility group box family member 3*) läheduses. Põhjuseks, miks kirjandusest leitud markerid käesolevas uuringus olulisi p-väärtusi ei andnud, võib peituda asjaolus, et *EGCUT* valimi suurus ei olnud piisavalt suur.

*TOX3*, tuntud ka kui *TNRC9*, on kuueteistkümnendas kromosoomis asuv geen, mis koosneb kolmest domäänist, milleks on: tuuma lokalisatsiooni signaaliga (*Nuclear Localization Signal*) N-terminaalne domään, *HMG-box* (*High Mobility Group, HMG*) ning polüglutamiinse järjestusega C-terminaalne domään. *HMG-box*'i olemasolu viitab sellele, et *TOX3* võib DNA painutamise ning lahti hargnemise teel olla seotud kromatiini struktuuri muutusega ja lisaks võib *TOX3* olla transkriptsioonifaktoriks (Dittmer jt., 2011). *TOX3* geeni

on mitmetes uuringutes seostatud kõrgeenenud rinnavähi riskiga, kuid selle täpne bioloogiline roll vähi juures on hetkel selgusetu (Dittmer jt., 2011; Easton jt., 2007; Jones jt., 2013). Elematore jt. uuring näitas, et vähemalt üks *TOX3*-s asuvatest markeritest oli perekondlikult klasterdunud, assotsieerudes suurel määral rinnavähiga. Kusjuures TT homosügootide seas oli assotsiatsiooni tõenäosus suurem kui CT heterosügootidel (Elematore jt., 2014). Dittmeri jt. uuringust selgus, et *TOX3* ekspreseeerub valdavalt ajus ja ei täheldatud selle esinemist normaalses rinnanäärme koes. Siiski arvati, et on võimalik *TOX3* geeni ekspreseeerumine ebanormaalses rinnanäärme koes (Dittmer jt., 2011).

*YAP1*, tuntud ka kui *YAP* ja *YAP65*, on *Src* perekonna kinaaside hulka kuuluv onkogeen, mille üleekspreseeerimist on seostatud mitmete erinevate vähivormide tekke ja progresseerumisega (Sudol, 1994; Zhang jt., 2011). *YAP1* on transkriptsiooni koaktivaator *Salvador-Warts-Hippo (SWH)* signaalülekanerajas, mis kontrollib imetajates kudede kasvu ning mida on täheldatud kui olulist signaalrada vähiarengus. Sellest tulenevalt peetakse nii *YAP1* kui ka *SWH* signaalrada potentsiaalseks vähiravi sihtmärgiks (Zhang jt., 2011; Zhao jt., 2010). *YAP1* seotust polütsüstiliste munasarjade sündroomiga on näidanud veel Louwers jt. poolt läbi viidud meta-analüüs, mille valimisse kuulusid indiviidid Hollandist, Hiinast ning Ameerika Ühendriikidest. Meta-analüüsi tulemused näitasid tugevat assotsiatsiooni *YAP1* ning *PCOS*-i vahel, andes ülegenoomselt olulise p-väärtuse  $1,0 \times 10^{-9}$  (Louwers jt., 2013).

Polütsüstiliste munasarjade sündroomi tagapõhja selgitamiseks on läbi viidud arvukalt uuringuid, millest on välja kerkinud mitmeid huvipakkuvaid kandidaatgeene. Seni teadaolevad assotsieerunud markerid Eesti andmetes tugevas assotsiatsioonis ei esinenud, kuid siinsest analüüsist ilmnisid kaks uut signaali, kromosoomides üks ja kolmteist. Kolmeteistkümnendas kromosoomis asuva markeri kaudu kerkis esile veel uus ja huvitav kandidaatgeen, milleks on *SPACA7*. Leitud geeni täpse funktsiooni ja seose selgitamiseks *PCOS*-iga tuleks läbi viia täiendavaid uuringuid.

## Kokkuvõte

Polütsüstiliste munasarjade sündroom on üks sagedasemaid haiguseid naiste seas, olles seejuures ka oluliseks viljastuse põhjustajaks naistel. Ehkki *PCOS*-i etioloogia ei ole lõplikult selge, omab haiguse juures suurt rolli geneetiline taust. Selle kindlaks määramiseks on läbi viidud hulgaliselt uuringuid, millest on ilmnenu nii mitmeid seoseid geneetiliste markeritega kui ka olulisi kandidaatgeene.

Käesoleva töö raames teostati 7170 Eesti päritolu indiviidi seas logistilise regressiooni mudelit kasutades ülegenoomne assotsiatsiooniuuring eesmärgiga leida geneetilisi variante, mis võiksid suurendada *PCOS*-i haigestumise riski. Puuduolevate markerite genotüüpide leidmiseks kasutati referentspaneelina 1000 Genoomi Projekti imputeeritud andmestikku. Lisaks saadeti leitud markeritest osa replikatsioonianalüüsi ning uuriti varasemalt täheldatud assotsiatsioonis olevate variantide osatähtsust Eesti andmetes.

Tulemustena identifitseeriti 30 markerit, mille p-väärtuse jäi alla  $1 \times 10^{-6}$ , nende hulgast üks andis koheselt ülegenoomselt olulise tulemuse, kuid ei replitseerunud. Replikatsioonijärgselt muutus ülegenoomselt oluliseks veel teinegi variant, mille kaudu õnnestus identifitseerida võimalik *PCOS*-i kandidaatgeen. Selleks oli spermi akrosoomiga seotud geen *SPACA7*, mis võib rolli omada reproduktiivsuse juures, kuid mille täpsem funktsioon on hetkel selgusetu.

Kirjandusest leitud senituntud *PCOS*-iga assotsieerunud markerid uuringus peale mitmese testimise korrektsiooni olulist p-väärtust ei näidanud, kuid kaks varianti andsid p-väärtuse alla 0,05. Nendeks olid *TOX3* ja *YAPI*, mis mõlemad on seostatud kõrgegenenud vähi riskiga.

Identifitseeritud markerite ja lookuste assotsiatsioonide kinnitamiseks polütsüstiliste munasarjade sündroomiga tuleks sõltumatus kohordis läbi viia veel täiendavaid replikatsioonianalüüse.

## Resümee/Summary

### Genome – Wide Association Study in order to find genetic variants for polycystic ovary syndrome

Kreete Lüll

#### Summary

Polycystic ovary syndrome (PCOS) is one of the most common multifactorial disorders, occurring up to 10% of women in reproductive age. PCOS is described to have many different symptoms such as hyperandrogenism, polycystic ovaries, disorders in menstrual cycle, overweight and obesity. PCOS is also one of the main reasons for female infertility, it has been said that approximately 40% of women with PCOS are infertile. The etiology of PCOS is unknown; however for a long time it has been clear that combined with environmental factors PCOS has a genetic background.

The aim of this study was to identify genetic loci that can be associated with PCOS in the Estonian population and replicate the findings in an independent cohort data. Present study also had a second aim which was to estimate P-values of previously detected markers of PCOS in the Estonian population.

In the current study 169 cases and 7001 controls gathered from all over Estonia were tested for the association with PCOS by using a logistic regression method. In order to find genotypes for missing markers a 1000 Genome Project imputed data was used. Also 33 markers from 13 different loci were selected from previous studies and their P-values were tested in Estonian cohort.

As a result of Genome-Wide Association Study 30 genetic variants in 10 different loci with P-values less than  $1 \times 10^{-6}$  were found, out of which one marker (rs150527962) in chromosome 1 had a genome wide significance with P-value  $4.76 \times 10^{-8}$ . After replicating Estonian results in the Wellcome Trust Centre for Human Genetics in the University of Oxford and combing the two results, another marker showed a genome wide significance with P-value  $4.04 \times 10^{-8}$ . This marker (rs7984604) was located in the chromosome 13 and it was near to the sperm acrosome associated gene *SPACA7*, which can be a novel candidate gene for PCOS. *SPACA7* is expressed only in testis and has been associated with the development of the acrosome of sperm and that has given a reason to believe that *SPACA7* is related with the reproduction. However, it has not been clarified and further studies are needed to validate this.

30 out of 33 markers that were found in previous studies were also present in our dataset but none of them showed significant P-values after multiple testing corrections. However, two variants had P-values less than 0.05. These markers were rs1894116 in chromosome 11 and rs4784165 in chromosome 16. The marker in chromosome 11 belonged to gene *YAP1* and marker in chromosome 16 was near to gene *TOX3*. Both of these genes are associated with higher risk of cancer development.

In order to verify the results of present study further studies should be considered in an independent cohort.

## Tänuavaldused

Soovin tänada kõiki inimesi, kes mind lõputöö kirjutamise juures toetasid. Eelkõige tänan oma juhendajat Reedik Mägi suure abi ning tarkade õpetussõnade eest. Lisaks tänan oma kaasjuhendajat professor Andres Metspalu, tänu kellele sain võimaluse tegeleda niivõrd põneva uurimusega.

Suured tänusõnad *PCOS*-i DNA proovide kogumise eest professor Andres Salumetsale ning Tartu Ülikooli Eesti Geenivaramule. Replikatsioonianalüüsi teostamise eest soovin tänada Oxfordi Ülikooli *Wellcome Trust* Inimgeneetika Instituuti.

Suur aitäh minu perele ja sõpradele igakülgse toetuse eest lõputöö kirjutamisel.

## Kasutatud kirjanduss

**El-Shal**, AS., Shalaby, SM., Aly, NM., Rashad NM., Abdelaziz, AM. (2013). Genetic Variation in the Vitamin D Receptor Gene and Vitamin D Serum Levels in Egyptian Women with Polycystic Ovary Syndrome. *Molecular biology reports* 40(11):6063–73.

**Azziz**, R., Tarlatzis, R., Dunaif, A., Ibanez, L., Pugeat, M., Taylor, A., Fauser, CJM. (2004). Revised 2003 Consensus on Diagnostic Criteria and Long-Term Health Risks Related to Polycystic Ovary Syndrome. *Fertility and Sterility* 81(1):19–25.

**Azziz**, R., Tarlatzis, R., Dunaif, A., Ibanez, L., Pugeat, M., Taylor, A., Fauser, C. J. M. (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):4237–45.

**Azziz**, R., *et al.* (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

**Badawy**, A., Elnashar, A. (2011). Treatment Options for Polycystic Ovary Syndrome. *International journal of women's health* 3:25–35.

**Burton**, PR., Clayton, DG., Cardon, LR., The Wellcome Trust Case Control Consortium.. (2007). Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls. *Nature* 447(7145):661–78.

**Cardon**, LR., Bell, JI. (2001). Association Study Designs for Complex Diseases. *Nature reviews. Genetics* 2(2):91–99.

**Chen**, ZJ., Zhao, H., He, L., *et al.* (2011). Genome-Wide Association Study Identifies Susceptibility Loci for Polycystic Ovary Syndrome on Chromosome 2p16.3, 2p21 and 9q33.3. *Nature genetics* 43(1):55–59.

**Crosignani**, PG., Nicolosi, E. (2001). Polycystic Ovarian Disease: Heritability and Heterogeneity. *Human reproduction update* 7(1):3–7.

**Cui**, L., Zhao, H., Zhang, B., *et al.* (2013). Genotype-Phenotype Correlations of PCOS Susceptibility SNPs Identified by GWAS in a Large Cohort of Han Chinese Women. *Human reproduction (Oxford, England)* 28(2):538–44.

**Devlin**, B., Roeder K., Wasserman, L. (2001). Genomic Control, a New Approach to Genetic-Based Association Studies. *Theoretical population biology* 60(3):155–66.

**Dittmer**, S., Kovacs, Z., Yuan, Sh., *et al.* (2011). TOX3 Is a Neuronal Survival Factor That Induces Transcription Depending on the Presence of CITED1 or Phosphorylated CREB in the Transcriptionally Active Complex. *Journal of cell science* 124(Pt 2):252–60.

**Domecq**, JP., Prutsky, G., Mullan., RJ., *et al.* (2013). Adverse Effects of the Common Treatments for Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis. *The Journal of clinical endocrinology and metabolism* 98(12):4646–54.

- Easton, DF., Pooley, KA., Dunning AM., et al. (2007).** Genome-Wide Association Study Identifies Novel Breast Cancer Susceptibility Loci. *Nature* 447(7148):1087–93.
- Elematore, I., Gonzalez-Hormazabal, P., Reyes, JM., et al. (2014).** Association of Genetic Variants at TOX3, 2q35 and 8q24 with the Risk of Familial and Early-Onset Breast Cancer in a South-American Population. *Molecular biology reports*. February 2014: 1-8
- Eriksen, MB., Nielsen, MFB., Brusgaard, K., Tan, Q., Andersen, MS., Glintborg, D., Gaster, M. (2013).** Genetic Alterations within the DENND1A Gene in Patients with Polycystic Ovary Syndrome (PCOS). *PloS one* 8(9):e77186.
- Franks, S., Gharani, N., Waterworth, D., Batty, S., White, D., Williamson, R., McCarthy, M. (1997).** The Genetic Basis of Polycystic Ovary Syndrome. *Human reproduction (Oxford, England)* 12(12):2641–48.
- Freedman, ML., Reich, D., Penney, KL et al. (2004).** Assessing the Impact of Population Stratification on Genetic Association Studies. *Nature genetics* 36(4):388–93.
- Govind, A., Obhrai, MS., Clayton, RN. (1999).** Polycystic Ovaries Are Inherited as an Autosomal Dominant Trait: Analysis of 29 Polycystic Ovary Syndrome and 10 Control Families. *The Journal of clinical endocrinology and metabolism* 84(1):38–43.
- Hardy, J., Singleton, A. (2009).** Genomewide Association Studies and Human Disease. *New England Journal of Medicine* 1759–68.
- Hart, R., Hickey, M., Franks, S. (2004).** Definitions, Prevalence and Symptoms of Polycystic Ovaries and Polycystic Ovary Syndrome. *Best practice & research. Clinical obstetrics & gynaecology* 18(5):671–83.
- Healthwise. 2012.** Laparoscopic Ovarian Drilling (Ovarian Diathermy) for PCOS. Retrieved (<http://www.webmd.com/women/laparoscopic-ovarian-drilling-ovarian-diathermy-for-pcos>).
- Howie, BN., Donnelly, P., Marchini, J. (2009).** A Flexible and Accurate Genotype Imputation Method for the next Generation of Genome-Wide Association Studies. *PLoS genetics* 5(6):e1000529.
- Hwang, JY., Lee, EJ., Go, MJ., et al. 2012.** Genome-Wide Association Study Identifies GYS2 as a Novel Genetic Factor for Polycystic Ovary Syndrome through Obesity-Related Condition. *Journal of human genetics* 57(10):660–64.
- Jones, JO., Chin, SF., Wong-Taylor, LA., Leaford, D., Ponder, PAJ., Caldas, C., Maia, AT. (2013).** TOX3 Mutations in Breast Cancer. *PloS one* 8(9):e74102.
- Knochenhauer, ES., Key, TJ., Kahsar-Miller, M., Waggoner, W., Boots, LR., Azziz, R. (1998).** Prevalence of the Polycystic Ovary Syndrome in Unselected Black and White Women of the Southeastern United States: A Prospective Study. *The Journal of clinical endocrinology and metabolism* 83(9):3078–82.
- Korfanty, J., Toma, A., Wojtas, A., Rusin, A. (2012).** Identification of a New Mouse Sperm Acrosome-Associated Protein. *Reproduction* 143:749–57.



- Kosova, G., Urbanek, M.** (2013). Genetics of the Polycystic Ovary Syndrome. *Molecular and cellular endocrinology* 373(1-2):29–38.
- Legro, RS., Driscoll, D., Strauss III, JF., Dunaif, A.** (1998). Evidence for a Genetic Basis for Hyperandrogenemia in Polycystic. *Proceedings of the National Academy of Sciences of the United States of America* 95(December):14956–60.
- Li, T., Zhao, H., Zhao, X., Zhang, B., Cui, L., Shi, Y., Li, G., Wang, P., Chen, ZJ.** (2012). Identification of YAP1 as a novel susceptibility gene for polycystic ovary syndrome. *Journal of medical genetics* 49(4):254-7.
- Louwers, YV., Stolk, L., Uitterlinden, AG., Laven, JSE.** (2013). Cross-Ethnic Meta-Analysis of Genetic Variants for Polycystic Ovary Syndrome. *The Journal of clinical endocrinology and metabolism* 98(12):E2006–12.
- Lujan, ME, Chizen DR., Pierson, RA.** (2008). Diagnostic Criteria for Polycystic Ovary Syndrome: Pitfalls and Controversies. *Journal of obstetrics and Gynaecology Canada* 30(8):671–79.
- Manolio, TA.** (2010). Genomewide Association Studies and Assessment of Risk of Disease. *The New England journal of medicine* 363(21):2076–7; author reply 2077.
- Marchini, J., Cardon, LR., Phillips, MS., Donnelly, P.** (2004). The Effects of Human Population Structure on Large Genetic Association Studies. *Nature genetics* 36(5):512–17.
- Marchini, J., Howie, B.** (2010). Genotype Imputation for Genome-Wide Association Studies. *Nature reviews. Genetics* 11(7):499–511.
- McCarthy, MI., Hirschhorn, JN.** (2008). Genome-Wide Association Studies: Past, Present and Future. *Human molecular genetics* 17(R2):R100–1.
- Nguyen, EB., Westmuckett, AD., Moore, KL.** (2014). SPACA7 Is a Novel Male Germ Cell-Specific Protein Localized to the Sperm Acrosome That Is Involved in Fertilization in Mice. *Biology of reproduction* 90(1):16.
- Pearson, TA., Manolio, TA.** (2008). How to Interpret a Genome-Wide Association Study. *JAMA: the journal of the American Medical Association* 299(11):1335–44.
- Pe'er, I., Yelensky, R., Altshuler, D., Daly, MJ.** (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic Epidemiology* 32(4):381-5.
- Price, AL., Patterson, NJ., Plenge, RM., Weinblatt, ME., Shadick, NA., Reich, D.** (2006). *Nature Genetics* 38(8):904-9.
- Primakoff, P., Myles, DG.** (2002). Penetration, Adhesion, and Fusion in Mammalian Sperm-Egg Interaction. *Science (New York, N.Y.)* 296(5576):2183–85.
- Shi, Y., Zhao, H., Shi, Y., et al.** (2012). Genome-Wide Association Study Identifies Eight New Risk Loci for Polycystic Ovary Syndrome. *Nature genetics* 44(9):1020–25.

- Sirmans, SM., Pate, KA. (2013).** Epidemiology, Diagnosis, and Management of Polycystic Ovary Syndrome. *Clinical epidemiology* 6:1–13.
- Sladek, R., Rocheleau, G., Rung, J., et al. (2007).** A Genome-Wide Association Study Identifies Novel Risk Loci for Type 2 Diabetes. *Nature* 445(7130):881–85.
- Spencer, CCA., Su, Z., Donnelly, P., Marchini, J. (2009).** Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. *PLoS genetics* 5(5):e1000477.
- Sudol, M. (1994).** Yes-Associated Protein (YAP65) Is a Proline-Rich Phosphoprotein That Binds to the SH3 Domain of the Yes Proto-Oncogene Product. *Oncogene* 9(8):2145–52.
- Zhang, X., George, J., Deb, S., Degoutin, JL., Takano, EA., Fox, SB., Bowtell, DDL., Harvey, KF. (2011).** The Hippo Pathway Transcriptional Co-Activator, YAP, Is an Ovarian Cancer Oncogene. *Oncogene* 30(25):2810–22.
- Zhao, B., Li, L., Lei, Q., Guan, KL. (2010).** The Hippo-YAP Pathway in Organ Size Control and Tumorigenesis: An Updated Version. *Genes & development* 24(9):862–74.
- Teede, H., Deeks, A., 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.
- Vink, JM., Sadrzadeh, S., Lambalk, CB., Boomsma, DI. (2006).** Heritability of Polycystic Ovary Syndrome in a Dutch Twin-Family Study. *The Journal of clinical endocrinology and metabolism* 91(6):2100–2104.
- Visscher, PM., Brown, MA., McCarthy, MI., Yang, J. (2012).** Five Years of GWAS Discovery. *American journal of human genetics* 90(1):7–24.
- Wang, Z., Li, T., Xing, X., Gao, X., Zhang, X., You, L., Zhao, H., Ma, J., Chen, ZJ. (2013).** Replication Study of RAD54B and GREB1 Polymorphisms and Risk of PCOS in Han Chinese. *Reproductive biomedicine online* 27(3):316–21.
- Wang, WYS., Barratt, BJ., Clayton, DG., Todd, JA. (2005).** Genome-Wide Association Studies: Theoretical and Practical Concerns. *Nature reviews. Genetics* 6(2):109–18.
- Welt, CK., Carmina, E. (2013).** Lifecycle of Polycystic Ovary Syndrome (PCOS): From In Utero to Menopause. *The Journal of Clinical Endocrinology & Metabolism* 98(12):4629–38.
- Welt, CK., Styrkarsdottir, U., Ehrmann, DA., et al. (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.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., Parkinson, H. (2014).** The NHGRI GWAS Catalog, a Curated Resource of SNP-Trait Associations. *Nucleic acids research* 42(Database issue):D1001–6.
- Winkler, TW., Day, FR., Croteau-Chonka, DC., et al. (2014).** Quality Control and Conduct of Genome-Wide Association Meta-Analyses. *Nature protocols* 9(5):1192–1212.

**Yoon, B.J.** (2009). Hidden Markov Models and Their Applications in Biological Sequence Analysis. *Current genomics* 10(6):402–15.

## **Kasutatud veebileheküljed**

[https://mathgen.stats.ox.ac.uk/impute/impute\\_v2.html](https://mathgen.stats.ox.ac.uk/impute/impute_v2.html)

[https://mathgen.stats.ox.ac.uk/genetics\\_software/snptest/old/snptest\\_v2.3.0.html](https://mathgen.stats.ox.ac.uk/genetics_software/snptest/old/snptest_v2.3.0.html)

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

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

<http://www.webmd.com/women/laparoscopic-ovarian-drilling-ovarian-diathermy-for-pcos>

<http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html>

## Lisad

**Lisa 1 Kirjandusest leitud teadaolevad PCOS-iga assotsieerunud markerid, mis kajastusid ka meie uuringus.** P-väärtus < 0,05 on märgitud paksus kirjas. \* CI märgib šansside suhte 95%-list usaldusintervalli.

Marker	Positsioon	Kromo- soom	Lähim geen	Efekti/teine alleel	Efekti alleeli sagedus	OR (95% CI*) p-väärtus
rs11686574	11684007	2	GREB1	G/C	0,73	0,86 (0,67-1,10) 0,22
rs6740248	11706892	2	GREB1	G/C	0,05	0,61 (0,37-1,00) 0,05
rs12470971	11725241	2	GREB1	A/G	0,35	1,00 (0,80-1,26) 0,99
rs12468394	43561161	2	THADA	A/C	0,52	1,11 (0,88-1,39) 0,37
rs13429458	43638838	2	THADA	C/A	0,15	1,12 (0,82-1,53) 0,47
rs12478601	43721508	2	THADA	T/C	0,54	0,98 (0,79-1,23) 0,88
rs13405728	48978159	2	LHCGR, STON1 – GTF2AIL	G/A	0,12	0,80 (0,56-1,14) 0,22
rs2268361	49201612	2	FSHR	T/C	0,64	0,90 (0,72-1,14) 0,38
rs2349415	49247832	2	FSHR	C/T	0,66	0,81 (0,64-1,03) 0,08
rs2930961	95443806	8	FSBP, RAD54B	C/T	0,38	0,95 (0,76-1,19) 0,67
rs4385527	97648587	9	C9orf3	A/G	0,44	0,82 (0,66-1,02) 0,08
rs3802457	97741336	9	C9orf3	A/G	0,01	0,83 (0,33-2,08) 0,69
rs10739633	126202551	9	DENND1A	C/A	0,66	1,14 (0,90-1,43) 0,28
rs3829851	126219706	9	DENND1A, LOC102723374	G/A	0,06	0,83 (0,52-1,31) 0,43
rs61736953	126220114	9	DENND1A, LOC102723374	T/C	0,07	0,93 (0,60-1,45) 0,75
rs12377595	126439100	9	DENND1A	G/T	0,31	0,94 (0,75-1,20) 0,65

rs10818854	126446778	9	DENND1A	A/G	0,04	1,32 (0,79-2,19) 0,29
rs9785285	126520068	9	DENND1A	C/T	0,30	0,96 (0,76-1,22) 0,75
rs2479106	126525212	9	DENND1A	G/A	0,43	0,88 (0,70-1,09) 0,25
rs1778890	126531755	9	DENND1A	C/T	0,30	0,97 (0,76-1,23) 0,79
rs10986105	126549955	9	DENND1A	G/T	0,04	1,60 (0,95-2,71) 0,08
rs1894116	102070639	11	YAP1	G/A	0,10	1,48 (1,03-2,13) <b>0,03</b>
rs10841843	21691891	12	GYS2	C/T	0,80	1,08 (0,83-1,41) 0,57
rs6487237	21699928	12	GYS2	A/C	0,81	1,05 (0,81-1,38) 0,70
rs7485509	21702005	12	GYS2	T/C	0,81	1,05 (0,81-1,37) 0,71
rs705702	56390636	12	RAB5B, SUOX	G/A	0,30	1,09 (0,85-1,39) 0,49
rs2272046	66224461	12	HMGA2	C/A	0,01	0,53 (0,21-1,36) 0,19
rs4784165	52347819	16	TOX3	G/T	0,23	1,31 (1,01-1,71) <b>0,04</b>
rs2059807	7166109	19	INSR	G/A	0,54	1,24 (0,98-1,56) 0,07
rs6022786	52447303	20	BCAS1	G/A	0,61	0,97 (0,77-1,21) 0,77

## **Lihtlitsents**

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.

Mina, Kreete Lüll

(sünnikuupäev: 21. 10. 1992)

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

Ülegenoomne asotsiatsiooniuuring polütsüstiliste munasarjade sündroomiga seotud geneetiliste variantide leidmiseks,

mille juhendajad on Reedik Mägi ja Andres Metspalu,

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

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

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

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

Tartus, 26. 05. 2014