

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

Tsütokroom P450 2C19 deletsioonide esinemine Eesti populatsioonis

Magistritöö

30 EAP

Kadri Maal

Juhendajad prof. PhD Lili Milani
ja MSc Tarmo Puurand

TARTU 2021

Tsütokroom P450 2C19 deletsioonide esinemine Eesti populatsioonis

Lühikokkuvõte. Tsütokroom P450 on üheks põhiliseks ensüümide perekonnaks inimestel, mis on seotud ravimite lagundamisega. Indiviididel, kellel on polümorfismid tsütokroom P450 geenides, esineb oht kogeda ravimite standardiseeritud annustamise korral kõrvaltoimeid või ravimite ebatõhusat toimet (Meyer, 2004). Personaalmehitsiini eesmärgiks on leida igale inimesele võimalikult individuaalne ennetus- või raviplaan, analüüsides inimese geeniandmeid koos keskkonna-, tervisekäitumise ja terviseandmetega (Personaalmehitsiin, 2019). Tsütokroom P450 2C on üks olulisematest P450 ensüümide alamperekondadest, kuna sinna kuuluvad ensüümid (CYP2C8, CYP2C9, CYP2C18, CYP2C19) lagundavad üle 20% kõigist terapeutilistest ravimitest (Rendic, 2002). Antud töö eesmärgiks oli leida ja katseliselt kinnitada *CYP2C19* geeni deletsioone Tartu Ülikooli Eesti Geenivaramu (TÜ EGV) andmete põhjal, määrata kindlaks leitud deletsioonide katkekohad genoomis ning määrata nende deletsioonide esinemissagedused Eesti populatsioonis. GSA mikrokiibi genotüpiseerimise andmete analüüsimisel selgus, et *CYP2C19* geeni deletsioonidega indiviidide esinemissagedus Eesti populatsioonis on ligikaudu 1,68%.

CERCS teaduseriala kood: B790 (Kliiniline geneetika)

Märksõnad: Tsütokroom P450, CYP2C19, personaalmehitsiin, farmakogeneetika

Cytochrome P450 2C19 deletions in the Estonian population

Abstract. Cytochrome P450 enzymes in humans are one of the major catalysts involved in drug degradation. Individuals with genetic polymorphisms in the cytochrome P450 genes are at risk of adverse drug reactions or therapeutic failure at standard drug doses (Meyer, 2004). The goal of personalised medicine is to find an individualised prevention or treatment plan for each person by analysing genetic data in combination with the environment, health behaviour and health data of the person (Personaalmehitsiin, 2019). Cytochrome P450 2C is one of the most important subfamilies of P450s, because its enzymes (CYP2C8, CYP2C9, CYP2C18, CYP2C19) degrade more than 20% of all therapeutic drugs (Rendic, 2002). The aim of this study was to find and experimentally confirm deletions of *CYP2C19* based on the data from the University of Tartu's Estonian Biobank and to determine the frequencies of these deletions in the Estonian population. Analysis of genotyping data revealed that the frequency of individuals with the *CYP2C19* gene deletion in the Estonian population is approximately 1,68%.

CERCS research specialisation: B790 (Clinical genetics)

Keywords: Cytochrome P450, CYP2C19, personalised medicine, pharmacogenetics

SISUKORD

KASUTATUD LÜHENDID	5
SISSEJUHATUS	7
1. KIRJANDUSE ÜLEVAADE.....	8
1.1. Farmakogeneetika ja farmakogenoomika.....	8
1.1.1. Tsütokroom P450 geenid ja ensüümid	9
1.1.2. CYP geenide evolutsioon.....	11
1.1.3. Geneetiline varieeruvus	13
1.1.4. Haplotüübid ja star-alleelid	15
1.1.5. Inimese tsütokroom P450-d.....	15
1.2. Personaalmeditsiin.....	18
1.2.1. Farmakokineetika ja farmakodünaamika.....	19
1.2.2. Ravimite kõrvaltoimed	20
1.2.3. Ravimi metaboliseerimise tüübid	21
1.2.4. Ravimite lagundamine ja seos tsütokroom P450-tega.....	24
1.3. CYP2C19.....	25
1.3.1. CYP2C19 star-alleelid	26
1.3.2. P450 2C19 substraadid	28
1.4. Järgmise põlvkonna sekveneerimine ja mikrokiiptehnoloogiad	30
2. EKSPERIMENTAALOSA	32
2.1. Töö eesmärk	32
2.2. Materjalid ja meetodika.....	32
2.2.1. Andmestik.....	32
2.2.2. CYP2C19 deletsioonide katkekohtade leidmine.....	33
2.2.3. Alleelispetsiifiliste PCR praimerite disainimine	34
2.2.4. PCR ja geelelektroforees	35
2.2.5. Deletsiooniga indiviidide diplotüüpide määramine.....	36
2.2.6. Deletsioonidega aheldunud SNP-ide leidmine	37
2.3. Tulemused	39

2.3.1. <i>CYP2C19</i> deletsioonide katkekohad	39
2.3.2. Leitud <i>CYP2C19</i> deletsioonide katseline valideerimine	41
2.3.3. Leitud <i>CYP2C19</i> deletsioonide esinemissagedused Eesti populatsioonis.....	44
2.3.4. <i>CYP2C19</i> diplotüüpide määramine WGS andmetest leitud deletsiooniga indiviididel.....	45
2.3.5. Deletsioonidega potentsiaalselt aheldunud SNP-id.....	46
2.4. Arutelu	49
KOKKUVÕTE	58
SUMMARY.....	60
TÄNUSÕNAD	63
KASUTATUD KIRJANDUSE LOETELU	64
Kasutatud veebiaadressid	76
LISA 1	78
LIHTLITSENTS.....	79

KASUTATUD LÜHENDID

ADR	Ravimi kõrvaltoime (<i>adverse drug reaction</i>)
BP	Aluspaar (<i>base pair</i>)
CN	Geeni koopiaarv (<i>copy number</i>)
CNV	Geeni koopiaarvu variatsioon (<i>genetic copy number variation</i>)
CO	Süsinikmonooksiid
CPIC	<i>Clinical Pharmacogenetics Implementation Consortium</i>
CYP	Tsütokroom P450 geen
CYPOR	Tsütokroom P450 oksüdüreduktaas
DNA	Desoksüribonukleiinhape
DNP	Dinitrofenüül
EM	Normaalne metaboliseerija (<i>extensive metaboliser</i>)
ER	Endoplasmaatiline retiikulum
FAD	Flaviin-adeniin-dinukleotiid
FDX	Ferredoksiin
FDXR	Ferredoksiin reduktaas
FMN	Flaviin mononukleotiid
FMO	Flaviini sisaldav monooksügenaas
GoF	Geenimutatsioon, mis põhjustab uue funktsiooniga geeniproducti teket või valgu suurenenud aktiivsust (<i>gain-of-function</i>)
GRCh37	Inimese referents genoomi versioon 37
GRCh38	Inimese referents genoomi versioon 38
GSA	Illumina Infinium Global Screening Array mikrokiip
GWAS	Ülegenoomne assotsiatsiooniuuring (<i>genome-wide association study</i>)
HGP	Inimese genoomi projekt (<i>Human Genome Project</i>)
IM	Keskmine metaboliseerija (<i>intermediate metaboliser</i>)
LD	Aheldatuse tasakaalutus (<i>linkage disequilibrium</i>)
LoF	Funktsioonikaoga mutatsioon (<i>loss-of-function</i>)
LOVD	<i>Leiden Open Variation Database</i>
MAF	Vähemesineva alleeli sagedus (<i>minor allele frequency</i>)
MAO	Monoamiini oksüdaas
MRCA	Viimane ühine esivanem (<i>most recent common ancestor</i>)
NADH	Nikotiinamiid adeniin dinukleotiidi resutseeritud vorm
NADPH	Nikotiinamiid adeniin dinukleotiid fosfaadi redutseeritud vorm

NGS	Järgmise põlvkonna sekveneerimine (<i>next generation sequencing</i>)
NO	Lämmastikmonooksiidi
NAT	N-atsetüültransferaas
Omni	Illumina InfiniumOmniExpress mikrokiip
P450	Tsütokroom P450 ensüüm
PharmVar	<i>Pharmacogene Variation Consortium</i>
PM	Aeglane metaboliseerija (<i>poor metaboliser</i>)
RM	Kiire metaboliseerija (<i>rapid metaboliser</i>)
ROS	Reaktiivsed hapnikühendid (<i>reactive oxygen species</i>)
SNP	Ühenukleotiidne polümorfism (<i>single nucleotide polymorphism</i>)
SNV	Ühenukleotiidne variatsioon (<i>single nucleotide variation</i>)
SuperCYP	<i>Cytochrome P450 database</i>
TÜ EGV	Tartu Ülikooli Eesti Geenivaramu
UGT	UDP-glükuronosüültransferaas
UM	Ülikiire metaboliseerija (<i>ultrarapid metaboliser</i>)
VA	Variatsioonide annotatsioonid (<i>variant annotation</i>)
WES	Kogu eksoomi sekveneerimine (<i>whole exome sequencing</i>)
WGS	Kogu genoomi sekveneerimine (<i>whole genome sequencing</i>)
WT	Metsiktüüp (<i>wild type</i>)

SISSEJUHATUS

Arstid on juba pikka aega teadnud, et patsientide ravimvastused erinevad. Kuigi individuaalsed erinevused ravimvastuses võivad tuleneda vanusest, soost, eelnevalt põetud haigustest või ravimite koostoimetest, mõjutavad geneetilised tegurid nii ravimi efektiivsust kui ka kõrvaltoimete esinemise tõenäosust (Weinshilboum, 2003). Ravimite lagundamisega seotud polümorfismide uurimise valdkonda nimetatakse farmakogeneetikaks (Gonzalez ja Nebert, 1990). Farmakogeneetikal on potentsiaali tuvastada konkreetne ravim ning ravimiannus, mis on suure tõenäosusega kindla patsiendi jaoks efektiivne ja ohutu. See on kaasaegse ravimiteraapia üks peamisi eesmärke ja seda kutsutakse personaalmeditsiiniks (Meyer, 2004).

Tsütokroom P450 on üheks põhiliseks ensüümide perekonnaks inimestel, mis on seotud ravimite lagundamisega. Indiviididel, kellel esinevad geneetilised polümorfismid, sealhulgas geene kodeerivate järjestuste mutatsioonid, geenide duplikatsioonid, deletsioonid ja regulatiivsed mutatsioonid tsütokroom P450 perekonna geenides, on oht kogeda ravimite standardsete annuste korral kõrvaltoimeid või ravimite ebatõhusust (Meyer, 2004). Enamik ravimvastustest on multifaktoriaalsed, see tähendab, et mitmed geenid ja mitmed keskkonnategurid koos mõjutavad indiviidi ravimvastust. Personaalmeditsiini eesmärgiks on leida igale inimesele võimalikult individuaalne ennetus- või raviplaan, analüüsides inimese geenandmeid koos keskkonna-, tervisekäitumise ja terviseandmetega (Personaalmeditsiin, 2019).

Tsütokroom P450 2C on üks olulisematest P450 ensüümi alam perekondadest, kuna sinna kuuluvad ensüümid (CYP2C8, CYP2C9, CYP2C18, CYP2C19) lagundavad üle 20% kõigist terapeutilistest ravimitest ning lisaks ka mitmeid endogeenseid ühendeid (Shimada jt., 1994; Rendic, 2002). Käesolevas töös uurisin Tartu Ülikooli Eesti Geenivaramu (TÜ EGV) andmete põhjal *CYP2C19* geeni deletsioonide esinemist Eesti populatsioonis. P450 2C19 on kliiniliselt oluline ensüüm, mis lagundab muuhulgas krambivastast ravimit mefenütoiini, haavandivastaseid ravimeid nagu omeprasool, teatud antidepressante nagu näiteks estsitalopraam ja malaariavastast ravimit proguaaniil (Blaisdell jt., 2002).

Minu töö üldeesmärk oli leida *CYP2C19* geeni deletsioone Eesti populatsioonis, määrata nende deletsioonide esinemissagedused ning pakkuda välja võimalikke viise, kuidas uuritud *CYP2C19* deletsioonide esinemist oleks tulevikus lihtsam ja kiirem kinnitada.

1. KIRJANDUSE ÜLEVAADE

1.1. Farmakogeneetika ja farmakogenoomika

Võib öelda, et farmakogeneetika ajalugu ulatub tagasi aastasse 510 eKr, kui Pythagoras märkas, et põldubade (*Vicia faba*) söömine põhjustas potentsiaalselt surmavat hemolüütilist aneemiat osades, kuid mitte kõigis inimestes (Arslan, 2015). Täna on teada, et selle põhjuseks oli glükoos-6-fosfaadi dehüdrogenaasi (G6PD) puudulikkus (Simoons, 1998).

Termin farmakogeneetika on kasutuses alates 1959. aastast (Vogel, 1959), mil seda kasutati esmakordselt seoses ravimite metabolismi ja kindlate ravimite ravimvastuse fenotüübi varieerumisega. Vogel mainib, et ravimite lagundamise võime on geneetiliselt määratud ning esinevad erinevad pärandumistüübid. Ühe näitena tõi ta kirurgias kasutatava lühiajalise toimega lihasrelaksandi, suksinüülkoliinkloriid, mille mõju kestis paljudel patsientidel tavapärasest kauem, mis oli omadus, mis esines perekonniti. (Vogel, 1959)

Pärast seda kui geenide sekveneerimine kloonimise põhiste meetoditega asendus ülegenoomse sekveneerimise ja genotüpiseerimise tehnoloogiaga, võeti kasutusse termin farmakogenoomika (Daly, 2017). Farmakogenoomika terminit kasutas esimest korda ingliskeelses kirjanduses „Nature Biotechnology” ajakirja toimetaja Andrew Marshall 1997. aasta septembris. Ta kirjutas Genset (Pariis, Prantsusmaa) ja Abbott Laboratories (Abbott Park, Illinois, USA) firmade vahelisest lepingust, et analüüsida geneetilist varieeruvust, mille tulemuseks on patsienditi erinevad ravimvastused – nii nimetatud farmakogenoomika. See arvati olevat esimene mitmetest tulevastest tehingutest spetsialiseerunud genoomika firmade ja ravimifirmade vahel selles valdkonnas (Marshall, 1997).

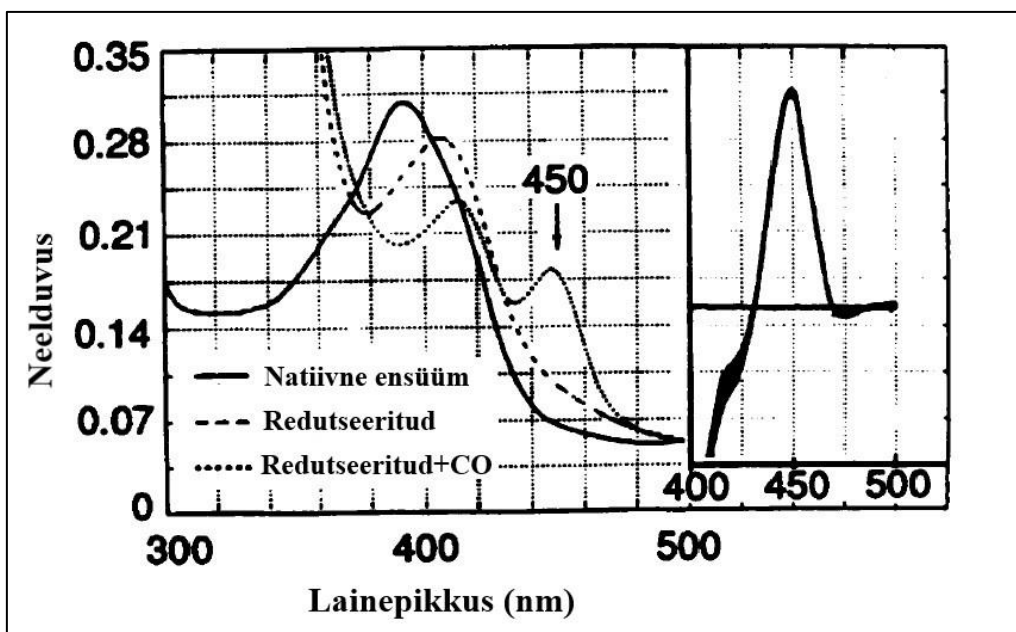
1997. aasta oktoobrikuus avaldati ka ajakirjas „Médecine/sciences” artikkel, milles toodi välja 1997. aasta maikuu Monte Carlos Euroopa Inimese Genoomi Projekti sümposiumil esinenud inimesed ja ette kantud teemad (HGP Europe, 1997). Üks esinejatest oli meditsiinilise geneetika professor Daniel Cohen (Genset firma esindaja), kes tegi esitluse teemal *Genomics and Pharmacogenomics*. Kuigi ettekande sisu pole hetkel võimalik lugeda, võib ainuüksi pealkirja ning toimumise aja järgi öelda, et Daniel Cohen oli tõenäoliselt esimene, kes kasutas terminit farmakogenoomika.

Termineid farmakogeneetika ja farmakogenoomika võib kasutada samaväärsetena, kuigi farmakogenoomika on laiem ning see hõlmab ka konkreetsetele haigusgeenidele suunatud uute ravimite väljatöötamist (Daly, 2017).

Farmakogeneetika keskendub peamiselt inimese pärandatava (*germline*) DNA variatsioonidele, kuid hiljuti on hakatud uurima ka kasvajate DNA varieerumist seoses ravimite loomisega, mis on suunatud nendes kasvajates leiduvate muteerunud geenide vastu (Daly, 2017). Farmakogenoomikat saab defineerida kui valdkonda, milles uuritakse, kuidas ravimid interakteeruvad kogu genoomiga, mõjutamaks bioloogilisi protsesse ja metaboolseid radu (Daxinger ja Whitelaw, 2010).

1.1.1. Tsütokroom P450 geenid ja ensüümid

Üheks geeniperekonnaks, mille variatsioone farmakogeneetikas uuritakse, on tsütokroom P450 (*CYP*), mille geeniproduktid osalevad elektronide transpordis ning inimese puhul on täheldatud seoseid teatud geenivariandi ja kindla komplekshaiguse suurenenud riski vahel, nagu näiteks seos *CYP4A11* geeni variantide ja koronaararterite haiguse vahel (Nebert jt., 2013). Tsütokroomi korrektne definitsioon on: “rauda sisaldav valk, mis osaleb rakkude hingamises oksüdatsiooni redutseerimise katalüsaatorina” (Nebert jt., 2013). Tsütokroom P450 (P450) ensüümid moodustavad hemoproteiini perekonna, mis hetkel teadaolevalt sisaldab üle 35 000 ensüümi ning esinevad erinevates liikides alates bakteritest kuni taimede ja imetajateni välja. (Gilardi, 2018) Mõiste tsütokroom P450 võtsid kasutusele Omura ja Sato 1962. aastal, et kirjeldada membraaniga seotud ja heemi sisaldavat valku, mille pigment neelab valgust süsinikmonooksiidiga seotud olekus kõige rohkem lainepikkusel 450 nm (joonis 1) (Debose-Boyd, 2007).



Joonis 1. Hemoproteiini neeldumisspekter natiivses, redutseeritud ning redutseeritud ja CO-ga seotud olekutes. CO-ga seotud redutseeritud ensüümi neeldumisspekter näitab eeldatavat neeldumismaksimumi lainepikkusel 450 nm, mis on tsütokroom P450 ensüümide üldtuntud omadus. (Tõlgitud ja kohandatud, Song ja Brash, 1991).

P450 ensüümid on heem-tiolaat valgud. Nad kannavad väliskeskkonna eest kaitstult hüdrofoobses taskus protoporfüriini IX, mis on seotud apoproteiiniga heemi rauatsentri ja konserveerunud tsüsteinüüljäägi väävli vahelise sideme kaudu. P450 ensüüme on leitud kõigist taimeosadest ja kõigist uuritud rakutüüpidest. (Kahn ja Durst, 2000)

Bakterites ja eukariootsetes mitokondrites leiduvaid tsütokroom P450 ensüüme nimetatakse “tüüp I” ensüümideks ning eukariootide endoplasmaatilises retiikulumis (ER; mikrosoomid) leiduvaid nimetatakse “tüüp II” P450 ensüümideks (Nebert jt., 2013). Mõlema P450 tüübi puhul võtab neis olev heemi raud elektronid vastu ning annab need üle molekulaarsele hapnikule ehk elektronide terminaalsele aktseptorile, vahendades sellega katalüüsi (Nebert jt., 2013).

Tsütokroom P450 geenide nimetamisel on iga geeninime lihttüveks “CYP”, millele järgneb numbriga tähistatud geeniperekond, seejärel tähega märgistatud alamperikond (kui see eksisteerib) ning lõpuks konkreetse geeni number (Gonzalez ja Nebert, 1990). Individuaalsete CYP-de paigutamine perekondadesse ja alamperikondadesse põhineb geeni järjestuse homoloogial. CYP geenide samasse perekonda paigutamise aluseks on $\geq 59\%$ homoloogia esinemine ning samasse alamperikonda paigutamise aluseks on nõutav sarnasuse määr ligikaudu 70% (Lewis jt., 1998). Ortoloogsed CYP geenid on üldiselt 71% – 80% ulatuses

sarnased ning on leitud, et üks geeniperekond on tavaliselt <36% sarnane teiste geeniperekondade P450 järjestustega (Nelson jt., 1996). *CYP*-de nomenklatuur põhineb ainult geneetilise järjestuse sarnasusel, geenide jaotumine perekondadesse ei kajasta üksikute *CYP* geenide omadusi ega funktsiooni (Gonzalez ja Gelboin, 1992).

1.1.2. *CYP* geenide evolutsioon

CYP-de klassifitseerimine perekondadesse ja alamperikondadesse järjestuste võrdlemise põhjal on osutunud kasulikuks fülogeneetiliste puude koostamisel ning muude analüüsivormide jaoks, mis uurivad *CYP* geenide vahelisi evolutsioonilisi seoseid. Fülogeneetilise puu näiteks on lisas 1 välja toodud Lewis jt. 1998 aasta artikli illustratsioon, mis näitab, milliste evolutsiooniliste sündmustega *CYP* geenide hargnemised võiksid korreleeruda.

Arvatakse, et esimesed *CYP* geenid tekkisid prokarüootses eellasliigis juba umbes 3,5 miljardit aastat tagasi (Nelson jt., 1993). Umbes 3,3 miljardit aastat tagasi, mil atmosfääri hakkas ilmuma vähestes kogustes hapnikku (Canfield, 2005), võis P450 võimalik funktsioon olla seda mürgist hapnikku neutraliseerida, katalüüsides kontsentreeritud 4-elektronilise O₂ reduktsiooni H₂O-ks, vältides sellega reaktiivsete hapnikuühendite (*reactive oxygen species*, ROS) teket. Umbes 2 miljardit aastat tagasi tõusis O₂ kontsentratsioon atmosfääris, mis põhjustas atmosfääri nihkumist redutseerivalt oksüdeerivale. Sarnaselt võis P450 funktsioon koos sellega nihkuda hapniku redutseerimiselt hapniku aktiveerimisele, katalüüsides hapniku aatomi ülekandmist molekulaarselt hapnikult endogeensetele ja hiljem ka ksenobiootilistele substraatidele. (Nebert ja Feyereisen, 1994)

400 miljonit aastat tagasi algas plahvatuslik *CYP* geenide superperikonna areng, mis korreleerub ajaga, mil taimed hakkasid koloniseerima maad ning arvatakse, et sellise hüppelise *CYP*-geenide mitmekesisustumise üheks juhtivaks jõuks oli nii nimetatud “keemiline sõda” taimede ja loomade vahel (Wöll jt., 2013). Röövloomade ja patogeenide tõrjumiseks on taimed välja töötanud üha keerukamaid kemikaale, millel on mürgised, toitumist takistavad (*anti-feedant*) või käitumist muutvad omadused. Seevastu P450 kiire areng, põhiliselt *CYP1*, *CYP2* ja *CYP3* perekondades, andis loomadele võime neid ühendeid mürgitustada (Kahn ja Durst, 2000). Teisisõnu, loomadel tekkis võime muuta taimede poolt toodetud toksilised keemilised ühendid enda jaoks kahjutuks. Molekulaarsed muutused loomapopulatsioonis võisid viia uute alleelide tekkimiseni, mis lubasid toiduks uusi taimi kasutada. Iga uue *CYP* geeni või geenivariandi tekkimise ja levimise tagajärjel võis suureneda populatsiooni võime kohaneda

keskkonnaga (Gonzalez ja Nebert, 1990). Loomade ja taimede vahelise “sõja” mõiste võib seega tegelikult olla pigem molekulaarse koevolutiooni näide.

Võib arvata, et inimese evolutsiooni käigus kadus vajadus mürgitustada lämmastikku sisaldavaid taimemürke (mida teeb näiteks CYP2D6), kuna inimeste toitumine muutus ning see andis võimaluse muidu juhuslikult tekkinud alleelivariantidel inimese genoomis püsima jääda, kuna puudus koevolutiooniline surve neid ära kaotada (Lewis jt., 1998). On pakutud, et toitumisest tulenev selektiivne surve on tõstnud erinevate *CYP* alleelide esinemise sageduse inimpopulatsioonis kõrgemaks kui see oleks ainult juhusliku geenitriivi tulemusena (Nebert, 1997). Järjestuse sarnasused *CYP* geenide superperekonnas viitavad evolutsioonile eellas-*CYP* geeni järjestuse korduvate duplitseerumiste ja varieerumiste tulemusena, mis on aset leidnud vähemalt viimased 1,4 miljardit aastat (Paquette jt., 2000). Geeniplokki, mis koosneb paarist samas järjekorras paiknevast geenist, nimetatakse mikrosünteeniaks ning erinevate liikide samane mikrosünteenia tähendab, et esineb ka samane geenide järjestus nende ühises (genoomses) eellases (Housworth ja Postlethwait, 2002).

Nelson jt. (2013) leidsid, et *CYP* geenide teke oli evolutsiooniliselt suhteliselt kiire protsess, kuid edasised muutused erinevates geeniperekondades tekkisid aeglaselt. Nad mainivad ka, et *CYP* geenide järjestuste võrdlemisest ainuüksi ei piisa, et selgitada välja nende geenide evolutsioonilised suhted. Näiteks geenide paiknemise ehk sünteenia uurimine on andnud uusi teadmisi *CYP* geenide pärinemise kohta. Nelson jt. (2013) analüüs näitab, et algsed *CYP* perekonnad jagasid ühist genoomset naabruskonda, mida nad nimetavad “tsütokroom P450 geneesi lookuseks”.

CYP51 on ainus tsütokroom P450 geen, mis esineb kõikides bioloogilistes riikides, ning kuna see on evolutsiooni käigus väga hästi säilinud (Rozman jt., 1996), arvatakse, et *CYP51* oli esimene *CYP* geen, mille levimise ja mitmekesisistumise tagajärjel tekkisid kõik teised *CYP* geenid (Yoshida jt., 1997). Senini on esinenud kaks vastuolulist hüpoteesi *CYP51*-e evolutsiooni kohta:

- 1) Kõige laialdasemalt on aktsepteeritud hüpotees, et *CYP51* perekond võis tekkida prokarüootse evolutsiooni ajastul ning on hiljem vertikaalselt kandunud üle kõikidesse eukarüootide riikidesse. Teisisõnu, steroolide ja *CYP51* biosünteesil on bakteriaalne päritolu lahustuvate (membraaniga mitteseotud) ensüümide näol, mis on potentsiaalseks evolutsiooniliseks esivanemaks (membraaniga seotud) eukarüootsetele *CYP51*-dele (Yoshida

jt., 1997; Nelson, 1999; Yoshida jt., 2000). Kuna *CYP51* geeni ei leidu kõikides bakteriliikides, on võimalik, et see geen on läinud kaduma enamikust bakteritest, mis ei sünteesi steroole.

2) Teiseks hüpoteesiks on, et *CYP51* geen kandus taimedelt horisontaalselt üle bakteritele. See viitab, et bakteriaalne *CYP51* ei pruugi olla vanim *CYP51* perekonna liige, vaid hoopis arenenud eukarüootsel eellasel ja kandunud horisontaalselt evolutsiooni käigus taimedelt mõnede bakteriliikidele (Debeljak jt., 2003; Režen jt., 2004).

Lamb jt. (2020) kinnitavad bioinformaatiliste analüüside tulemustega teooriat, et varajased eukarüoodid võisid siiski saada enda *CYP51*-e bakteritelt. Nende analüüsid hõlmasid enam kui 247 000 bakteriaalset genoomi ning nad leidsid, et *CYP51* geeni leidub enam kui 1000-el bakteriliigil, 9-st erinevast hõimkonnast, mis nende sõnul muudab horisontaalse geeni ülekande äärmiselt ebatõenäoliseks (Lamb jt., 2020).

Arvatakse, et esimene oksügenaasi tüüpi funktsioon P450-l oli steroolide ja rasvhapete oksüdeerimine, kuna steroolid ja rasvhapped on tihedalt seotud membraanidega ning üldse rakuelu kõigi aspektidega (Kahn ja Durst, 2000). Sterool-14-demetülaas, *CYP51*, katalüüsib reaktsiooni, mis on vajalik loomades kolesterooli, seentes ja algloomades ergosterooli, ning taimedes ja vetikates stigmasterooli biosünteesiks (Lamb jt., 2020). Steroole peetakse pigem eukarüootsete organismide määravaks osaks ning neid leidub harva bakterites, kus steroolide rolli täidavad hopanoidid (Jackson jt., 2002). Sarnaselt steroolidele on hopanoidid samuti tsüklilised isoprenoidsed lipiidid, kuid seejuures ei vaja nende sünteesimine atmosfäärset molekulaarset hapnikku (Abe, 2007). Endiselt pole teada bakteriaalse sterooli biosünteesi mehhanismi päritolu ning see, kuidas suhtestuvad bakteriaalsed *CYP51* valgud eukarüootsete *CYP51*-tega (Lamb jt., 2020). Seega pole endiselt päris täpselt teada kust *CYP* geenid pärinevad.

1.1.3. Geneetiline varieeruvus

Üldjuhul on igal geenil kromosoomipaaris kaks alleeli – üks kummaltki vanemalt. Geneetiline polümorfism on kahe või enama geeni alleeli esinemine populatsioonis, misjuures harvaesinevate alleelide sagedus on suurem kui vahetu mutatsiooniline teke seda võimaldaks (tavaliselt esinemissagedusega $\geq 0,01$). Ühenukleotiidiline polümorfism (*single nucleotide polymorphism*, SNP) on DNA järjestuse variatsioon, mis on tekkinud DNA ühe nukleotiidi (A, T, C või G) vahetumisel. (Lackie, 2007)

Vähemesinevat (*minor*) alleeli, mille sagedus on $\geq 0,05$ (5%), kutsutakse tavaliseks alleelivariandiks. Alleeli, mille esinemissagedus on vahemikus $0,05 \geq 0,01$, kutsutakse polümorfseks alleelivariandiks ning alleeli, esinemissagedusega alla 0,01 (1%), kutsutakse haruldaseks variandiks. Vähemesineva alleeli sagedus (*minor allele frequency*, MAF) on kõigi bialleelsete SNP alleelide esinemissagedus kindlas populatsioonis, välja arvatud peamiselt esinevate (*major*) alleelide esinemissagedused (Nebert ja Vesell, 2013). Tuleb silmas pidada, et niinimetatud peamiselt esinev (*major* või metsiktüüpi) alleel ühes etnilises populatsioonis võib teises etnilises populatsioonis olla hoopis vähemesineva alleeli staatuses (Stephens jt., 2001).

SNP-id esinevad genoomis umbes üks iga 300 – 3000 aluspaari kohta, kui võrrelda omavahel kahe mitte suguluses oleva indiviidid genome. Seejuures moodustavad need SNP-id 90 – 95% kõikidest varieeruvatest DNA saitidest (Meyer, 2004). Iga kaks üksikisikut erinevad teineteisest seega umbes 3 – 10 miljoni aluspaari võrra, mis on vähem kui 1% kogu haploidsest genoomist, umbes 3,2 miljardit aluspaari kokku.

Null alleel või funktsiooni kaoga alleel (*loss-of-function*, LoF) on geeni mittefunktsionaalne alleel (geeni variant), mis on põhjustatud geneetilisest mutatsioonist (Ganten ja Ruckpaul, 2006). Selliste mutatsioonide puhul võib täielikult puududa geeniproducti tootmine või on toodetud geeniproducti funktsionaalsus puudulik, mõlemal juhul peetakse null alleeli mittefunktsionaalseks (LoF). Null alleeli ei saa eristada deletsioonidest põhinedes ainult fenotüübilistele vaatlustele.

Geenide koopiaarvu erinevused (*genetic copy number variation*, CNV) võivad tuleneda geeni deletsioonidest või duplikatsioonidest. Deletsioon on mutatsioon, mille tulemusena on kadunud osa kromosoomist (Clancy ja Shaw, 2008). Geenide deletsioonid võivad olla osalised, mille puhul on ainult osa geeni järjestusest deleteerunud või täielikud geeni deletsioonid, mille puhul on deleteerunud kogu geen. Mõlemal juhul on tegemist geeni funktsiooni kaoga (LoF) ning geeni koopiaarv (*copy number*) on $cn \leq 1$. Duplikatsioon on mutatsioon, mille tulemusena on osa kromosoomist kahekordistunud (Clancy ja Shaw, 2008). Geeni ühe alleeli duplikatsiooni puhul on geeni koopiaarv $cn = 3$, kuid mõlema alleeli duplikatsiooni puhul või sama alleeli korduvate duplikatsioonide tagajärjel võib kindla geeni koopiaarv olla ka rohkem kui 3.

1.1.4. Haplotüübid ja star-alleelid

Kromosoomilõigu alleelne koosseis ehk geeni variandid või variantide kombinatsioonid, mis paiknevad ühes kromosoomis, on aheldatud ning päranduvad üheskoos järglastele, moodustavad haplotüübi. Haplotüüpi võib defineerida ka kui koos päritavate SNP-ide kogumit. Farmakogeneetiliste testide tulemused esitatakse tihti diplotüüpina, mis on haplotüüpide paarid ehk emalt ja isalt saadud haplotüübid. Haplotüübid või diplotüübid määratakse tavaliselt testitud geneetiliste variantide genotüüpide põhjal. Vaadeldud variantide grupeerimine alleelideks ja haplotüüpideks võimaldab diplotüüpide sidumist ennustatavate fenotüüpidega (Kalman jt., 2016).

Farmakogeneetikas kasutatakse farmakogeenide variantide nimetamisel star-alleelide (*star-allele*, *) süsteemi, mille puhul on enamasti referents alleel või haplotüüp märgistatud *1-ga (Kalman jt., 2016). *1 alleel on seega nii öelda metsiktüüpi (*wild type*, wt) ehk täielikult funktsionaalne alleel ning teised märgistused, nagu näiteks *2 või *3 tähistavad alleele/haplotüüpe, mis kannavad ühte või enamat geenivarianti (Robarge jt., 2007). *1 alleeli määratlus põhineb tavaliselt alampopulatsioonil, kus seda geeni algselt uuriti, mis aga ei tähenda, et see *1 alleel on kõigi populatsioonide enam levinud alleel. Näiteks *NAT2* geeni referents alleeliks on *NAT2*4*, kuna see on kõige levinum funktsionaalne alleel üle kõigi inimpopulatsioonide (Hein jt., 2008).

Star-alleelide kohta teabe saamiseks on hea teadmiste varamu farmakogeenide variatsioonide andmebaas PharmVar (*Pharmacogene Variation Consortium*). See keskendub haplotüübi struktuurile ja alleeli variantidele. Antud ressursis olev informatsioon lihtsustab farmakogeneetiliste testide tulemuste tõlgendamist täppismeditsiini suunamiseks. Selles andmebaasis keskendutakse geenide (mis on seotud ravimite lagundamisega, jaotumisega ja ravimvastusega) alleelide variatsiooni kataloogimisele ning ühese nomenklatuuri pakkumisele. (PharmVar, 2021)

1.1.5. Inimese tsütokroom P450-d

Inimese genoom sisaldab 18 *CYP* perekonda, mis on jagatud 41-ks valku kodeerivaks alamperekonaks, mis kodeerivad kokku 57 geeni (Guengerich, 2015). Inimese *CYP*-d on teadaolevalt geneetiliselt varieeruvad (Vesell, 1968) ja need mängivad paljudes kriitilistes eluprotsessides kesket rolli, mistõttu võivad defektid nendes geenides põhjustada

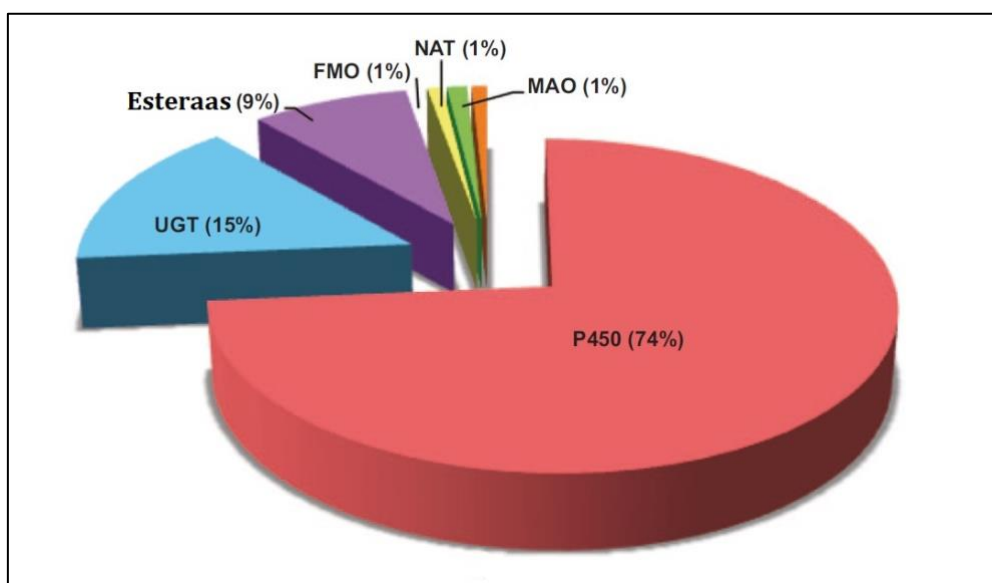
mitmesuguseid raskeid haigusi. On ilmunud mitmeid artikleid ülekenoomsete seoseuuringute kohta, milles kirjeldatakse seoseid *CYP* perekondade geeni variantide ja inimeses kindla komplekshaiguse suurenenud riski vahel (Nebert jt., 2013). Näiteks kui esinevad *missense*-mutatsioonid mõlemas *CYP* geenialleelis või kui mõlemad alleelid on deleteerunud, võivad kujuneda embrüoletaalsus (seotud *CYP51A1* ja *CYP26A1* geenidega), kaasasündinud ainevahetuslikud haigused (seotud *CYP2R1*, *CYP4F22*, *CYP5A1*, *CYP27A1* ja *CYP27B1* geenidega), hüpertensioon (seotud *CYP4A11* ja *CYP8A1* geenidega) ja vanusega seotud kollatähni kärbumine (seotud *CYP46A1* geeniga) (Nebert jt., 2013). Geneetiliste polümorfismide kontekstis on kõige ulatuslikumalt uuritud *CYP* geeniks *CYP2D6* ning on näidatud, et see lagundab umbes 20% tänapäeval kasutusel olevatest ravimitest (Taylor jt., 2020).

Lisaks *CYP* geenide modifikatsioonidele esineb muid viise, mis mõjutavad, enamasti inhibeerivad, P450 ensüümide aktiivsust ning seeläbi ravimite lagundamist (Guengerich, 2015). P450-te funktsiooni inhibeerimine võib olla otsene või kaudne. Otsese keemilise inhibeerimise suhtes eriti tundlikud P450 katalüütilise tsükli etapid on: substraadi seondumine P450 valguga, molekulaarse hapniku seondumine redutseeritud P450-ga ning hapniku aatomi substraati sisestamise etapp. Ained, mis konkureerivad substraatidega P450 aktiivsaidiga seondumises, on pöörduvad inhibiitorid. Pöördumatu inhibitsioon esineb, kui ühendid seonduvad P450 aktiivsaidiga kovalentselt, ilma reaktiivse vaheühendi vabastamiseta, seeläbi takistades substraadi seondumist ning valgu edasist funktsioneerimist. Teoreetiliselt võib otsene inhibitsioon toimuda ka pärast P450 valgu pinna posttranslatsioonilisi modifikatsioone oksüdantide, alküülivate, nitreerivate või atsetüülivate ainete poolt, mis takistavad P450 valgu ja tema redoks partnerite vahelist interakteerumist (Aguiar jt., 2005).

Kaudse toimega inhibiitorid võivad takistada P450 katalüütilise tsükli teisi etappe, nagu näiteks tsütokroom P450 oksüdüreduktaasi elektronide ülekandmise etappi (Rodrigues jt., 1991), suunates elektronid eemale P450 hemoproteiinist või inaktiveerides tsütokroom P450 oksüdüreduktaasi flavoproteiini ennast (Zhukov ja Ingelman-Sundberg, 1999). P450 funktsionaalset inhibeerimist võivad esile kutsuda ka ained, mis kahjustavad valgu või heemi sünteesi või kiirendavad valgu või heemi lagunemist (näiteks metallid nagu Co^{+2}), ning seeläbi vähendavad P450 hemoproteiini hulka organismis (Correia jt., 2011).

P450-ed on põhilised ensüümid inimestel, mis on seotud ravimite lagundamisega. Inimese maksas ainevahetust vahendavad P450-d on substraatide suhtes märkimisväärselt

mittespetsiifilised, lagundades peaaegu kolm neljandikku kõikidest kliiniliselt kasutatavatest ravimitest (Behrendorff jt., 2015). Kõikidest ravimite lagundamisega seotud ensüümidest moodustavad umbes 75% P450-ed (joonis 2) (Guengerich, 2015).



Joonis 2. Erinevate ensüümide panus ravimite metabolismi. UGT – UDP-glükuronosüültransferaas, FMO – flaviini sisaldav monooksügenaas, NAT – N-atsetüültransferaas, MAO – monoamiini oksüdaas. (Tõlgitud ja kohandatud, Guengerich, 2015).

Ravimeid lagundavad P450-d näivad olevat suure konformatsioonilise mitmekesisusega, mis võimaldab neil ensüümidel kohanduda erinevate substraatide struktuuride jaoks, peenelt ümber paigutades enda sekundaarseid struktuurielemente (Yano jt., 2004; Zhao jt., 2006). Üldiselt on P450 struktuur üsna konserveerunud. Mida lähemal heemile, seda konserveerunum on ehitus. P450 struktuuri kõige rohkem konserveerunud elemendid on seotud heem-tiolaadi hapniku aktivatsioonikeemiaga ning sellest tulenev struktuurne paigutus on ühine kõikidele P450-tele (Guengerich, 2015). Ootuspäraselt on substraadi spetsiifikat kontrollivad valguregioonid kõige enam teineteisest erinevad, kui võrrelda mitmete P450-te struktuure. P450-tele esineb kaks konformatsiooni: avatud ja suletud. Huvitavaks struktuurseks iseärasuseks on võimekus kohaneda erineva suuruse ja kujuga substraatidega, seejuures säilitades üldist P450 kuju ning elektronide ülekande ja O₂ aktivatsiooni keemiat (Guengerich, 2015). P450-tele aktiivsite mitmekesisus on aluseks nende ainulaadsetele rollidele imetajate füsioloogias ning kõigi P450-tele koosinemine indiviidis annab piisavalt mitmekülgse kaitse erinevate keskkonnas leiduvate potentsiaalselt ohtlike ainete vastu võitlemiseks ja neutraliseerimiseks organismis.

1.2. Personaalmehitsiin

Inimese Genoomi Projekti (*Human Genome Project*, HGP) lõpule viimise järel peaaegu 20 aastat tagasi arvati, et on algamas personaalmehitsiini ajastu, mil hakatakse laialdasemalt kasutama genotüübist sõltuvat ravimite välja kirjutamist patsientidele (Carr jt., 2014). Terminid personaalmehitsiin ja täppismehitsiin on tähenduselt lähedased farmakogeneetikale, kuid on laiemad ja hõlmavad lisaks mitte-geneetilisi faktoreid (Daly, 2017).

Mõni ravim võib mõne inimese puhul toimida tõhusalt, põhjustamata raskemaid kõrvaltoimeid patsiendi organismile, kuid mõne teise inimese puhul võib sama ravim olla ebatõhus ja/või tekitada organismis lagundamise käigus kõrvaltoimeid (Kalow, 1962; Weber, 1997; Licinio ja Wong, 2002). Personaalmehitsiini eesmärgiks on leida igale inimesele võimalikult individuaalne ennetus- või raviplaan, analüüsides inimese geeniandmeid koos keskkonna-, tervisekäitumise ja terviseandmetega (Personaalmehitsiin, 2019).

Personaliseeritud farmakogenoomikas on eesmärgiks püüda leida geenide ja ravimvastuse seoste kasutamiseks praktilisi väljundeid. Uuritakse, kuidas oleks võimalik efektiivselt, kiiresti, mõistlike kuludega ning võimalikult suure täpsusega prognoosida ühe isiku genoomse järjestuse põhjal, enne ravi alustamist, kas mingi ravim on patsiendi jaoks ohutu ja tõhus (Relling ja Evans, 2015). Farmakogenoomika võib parendada ravimite avastamist ja väljatöötamist kahel viisil: tuvastades ravimite sihtmärke ja aidates kaasa alampopulatsioonide spetsiifiliste ravimite väljatöötamisel (Evans ja Relling, 2004). Farmakogenoomikat saab kasutada, tuvastamaks geneetilisi polümorfisme, mille puhul esineb patsientidel eelsoodumus ravimite kõrvaltoimete tekke suhtes. Paljude farmakogeneetiliste polümorfismide genotüübiliste alamrühmade sagedused erinevad rassiti olulisel määral (Holden, 2003). Kuigi genoomne varieeruvus rassiliste/geograafiliste/etniliste gruppide siseselt on suurem kui gruppide vaheliselt (Rosenberg jt., 2002), esineb spetsiifilisi geneetilisi polümorfisme, mille sagedus varieerub rassiliste gruppide vaheliselt nii ulatuslikult, et see tähendab mõnede ravimite puhul efektiivsuse või toksilisuse erinevuste olulist esinemist (Burchard jt., 2003). Eriti oluline on see uute ravimite kliiniliste katsete etapis. Polümorfismid võivad küll esineda väga madala sagedusega patsientidel, kes uut ravimit saavad, kuid neil esinevad kõrvaltoimed võivad olla nii tõsised, et selle ravimi edasine arendamine peatatakse kõikide patsientide jaoks (Evans ja Relling, 2004).

Polümorfismide ja muude geeni variatsioonide osas on teada mitmed, mis võivad mõnede ravimite manustamisel tekitada toksilist reaktsiooni, mistõttu nende ravimite kasutamine võib osutuda ebapraktiliseks. Näiteks perekiliin põhjustab indiviidides perifeerset neuropaatiat CYP2D6 metabolismi puudumisel (Shah jt., 1982). Samuti võivad geeni polümorfismid mõjutada raviks vajalikku ravimiannust, näiteks CYP2C9 mõjutab varfariini doseerimist (Steward jt., 1997; Aithal jt., 1999) ning CYP2C19 omeprasooli doseerimist (Karam jt., 1996). Mõnedel juhtudel esinevad sellised geeni polümorfismid populatsioonide vaheliselt erineva sagedusega. Näiteks HLA-B*15:02 geenivariant on haruldane Eurooplaste populatsioonides, alleelisagedusega alla 0,1% (McCormack jt., 2011), kuid seda esineb laialdaselt Aasia populatsioonides, keskmise alleelisagedusega üle 10% (Chang jt., 2017).

Enamasti on terapeutiline efektiivsus ja ravimi kõrvaltoimed multifaktoriaalsed tunnused (tunnused, mis on mõjutatud paljude geenide koosmõjust, lisaks ka keskkonnamõjudest, põlvkondadeülelestest ja muudest epigeneetilistest mõjudest). Ülegenoomsete assotsiatsiooniuringutega (GWAS) saab leida seoseid geneetiliste variantide ja patsientide võimalike ravivastuste vahel enne kliinilise uuringu alustamist. Näiteks tuvastati ühes GWAS uuringus, et klopidogreeli toime ADP-indutseeritud trombotsüütide agregatsioonile on seotud ainult ühe alleelivariandiga, *CYP2C19*2* (Shuldiner jt., 2009), mida oli võimalik arvesse võtta kliinilise uuringu planeerimisel. Selliste seoste leidmise järel on võimalik kliinilise uuringu planeerimisel keskenduda ainult vajaliku uurimisele, hoides kokku nii aja ja ressursside kuludelt. Üldiselt võib öelda, et laialdasem geneetiline profileerimine võib statistiliselt parendada ravimite ohutust, kuid ei saa seda kunagi tagada (Kalow, 2006).

1.2.1. Farmakokineetika ja farmakodünaamika

Farmakokineetika ja farmakodünaamika on farmakoloogia peamised harud, mis keskenduvad endogeensete ja eksogeensete keemiliste ainete koostoimete uurimisele elusorganismides. Eelkõige uurib farmakodünaamika, kuidas ravim mõjutab organismi, samas kui farmakokineetika uurib, kuidas organism mõjutab ravimit. Mõlemad koos mõjutavad ravimi annustamist, efektiivsust ja ka kahjulike kõrvaltoimete teket. Geneetilised tegurid võivad mõjutada nii ravimi farmakokineetikat kui ka farmakodünaamikat, põhjustades muutusi ravimi lokaalses ja süsteemses toimes või ravimi sihtmärgi funktsioonis, muutes nii ravivastust (Pirmohamed, 2014).

Farmakodünaamikat saab defineerida kui ravimite ja bioloogiliste süsteemide koostoimest tulenevate bioloogiliste mõjude uurimist (Holford ja Sheiner, 1981). Farmakokineetika on ravimite omastamise (kehaomaseks muutmise) protsess organismis, ravimite poolt läbitav biotransformatsioon, ravimite ja nende metaboliitide jaotumine kudedes ning ravimite ja nende metaboliitide elimineerimine (Meyer, 2004). Teisisõnu, farmakokineetika on ravimite imendumine, jaotumine, biotransformatsioon ja eritumine. Ravimi eliminatsioon on reaktsioon, mis on suunatud ravimi toime või ravimi enese organismist eemaldamisele. Biotransformatsioon on ravimi keemiline muutmine organismis ensüümsüsteemide osavõtul, mille käigus muutub ravim ka kergemini erituvaks (Evans, 2006). Peamiseks biotransformatsiooni toimumise kohaks on maks.

Farmakokineetilistel interaktsioonidel on mitu molekulaarset alust. Üheks neist on ensüüminduktsioon, mille tulemuseks on tavaliselt biosaadavuse vähenemine (Guengerich, 2015). Ravimi vähenenud biosaadavus võib olla tingitud selle ravimi või mõne muu ravimi indutseerimisest. Teine ravimite koostoimete aspekt hõlmab endas P450 inhibitsiooni. Inhibitsioon võib siinkohal olla konkureeriv, ehk kaks substraati konkureerivad piiratud koguse P450 suhtes, või võib inhibitsioon olla mehhanismi põhine, mille korral substraadi oksüdeerimise järel P450 inaktiveerub (Guengerich jt., 2019).

Paljudel juhtudel on ksenobiootilise molekuli metabolismi esimeseks sammuks oksüdatsioonireaktsioon. Kuigi teised ensüümid, nagu näiteks peroksüdaasid või fenolaasid, näivad olevat võimelised kemikaale oksüdeerima, on enamikul juhtudel tõestatud P450 ensüümi osalus (ksenobiootiliste molekulide oksüdeerimisel). Tsütokroom P450 poolt katalüüsitud reaktsioonid on pöördumatud ja aeglased. Tüüpiliselt reageerib umbes 100 mooli substraati ühe mooli P450 kohta ühes minutis, kuigi paljud ksenobiootilised oksüdatsioonid on sellest veel aeglasemad. Sellest tulenevalt määrab P450 poolt katalüüsitud etapp eksogeense ühendi üldise farmakokineetika. (Guengerich, 2002)

1.2.2. Ravimite kõrvaltoimed

Eesti ravimiseaduse järgi on ravim igasugune aine või ainete kombinatsioon, mis on mõeldud haiguse või haigussümptomi vältimiseks, diagnoosimiseks, ravimiseks või haigusseisundi kergendamiseks (Ravimiseadus, 2004). Mürk on aine, mis organismi sattudes tekitab tugevaid, organismile kahjulikke talitlushäireid või kutsus esile surma. Iga ravim võib olla mürk kui seda ei kasutata õiges annuses ja õigel viisil.

Ravimid võib jagada kahte põhirühma: aktiivsed ravimid ja eelravimid (*prodrug*). Aktiivsete ravimite lagundamise tulemuseks on enamasti teisendatud ravim sellisel kujul, mis on vees lahustuv ja seega ka kergemini organismist elimineeritav (Weinshilboum, 2003). Eelravimid on toimeteta või vähese farmakoloogilise toimega kemikaalid, mis muutuvad läbi biotransformatsiooni terapeutiliselt aktiivseks metaboliidiks (Albert, 1958). Eelravimite metabolismi korral on vajalik, et manustatud ravim muundatakse esmalt organismis aktiivseks ravimi vormiks, enne selle elimineerimist kehast. Ravimi organismist väljutamise kaks peamist rada on ravimite lagundamine maksa ensüümide poolt ning ravimite eritumine neerude kaudu (Scott ja Scott, 2013). Selleks, et ravim omaks soovitud mõju, peab seda sihtmärgini jõudma piisavas koguses, ilma et see liigselt kuhjuks teistes kudedes (DeGorter jt., 2012).

Ravimi kõrvaltoime (*adverse drug reaction, ADR*) on ravimi kahjulik ja soovimatu toime, mis tekib ravimi profülaktiliste, diagnostiliste ja terapeutiliste annuste kasutamisel (Ravimite kõrvaltoimetest..., 2020). Ravimite kõrvaltoimete hulka kuuluvad annusest sõltuvad reaktsioonid, annusest sõltumatud reaktsioonid (sealhulgas idiosünkraatilised ravimreaktsioonid ja allergilised reaktsioonid), annusest ja ajast sõltuvad reaktsioonid (kumulatiivsed) ja võõrutusreaktsioonid (Edwards ja Aronson, 2000). Üldjuhul töötatakse ravimi soovituslikud annused välja, võttes sihtrühmaks normaalsed metaboliseerijad (*extensive metaboliser, EM*) (Guengerich, 2015), mis tähendab, et teiste metaboliseerimise tüüpide puhul ei pruugi soovitatav annus olla sobiv. Enamike ravimite metabolism on seotud rohkem kui ühe geeniga ning on raske eristada geneetilisi ja mitte geneetilisi faktoreid (nagu ravimite koosmõjud, dieet ja suitsetamine) (Evans ja Relling, 2004). Farmakogeneetika ja farmakogenoomika edenemine aitavad vähendada ravimi kõrvaltoimete tekkimist, kui patsientide genoomi informatsiooni arvestatakse ravi määramisel (Lauschke jt., 2018).

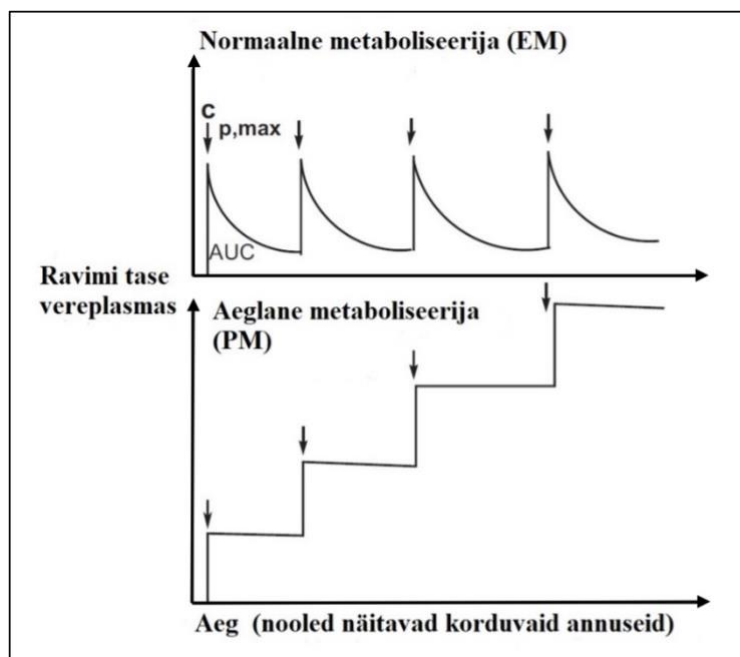
1.2.3. Ravimi metaboliseerimise tüübid

Metaboliseerimise tüüp sõltub ravimist ja seda mõjutavate ensüümide aktiivsusest. Ravimi metaboliseerimise kiiruse ehk indiviidi fenotüübi põhjal saab inimesed jaotada järgnevasse rühmadesse:

- 1) ülikiired metaboliseerijad (*ultrarapid metaboliser, UM*) on inividid, kellel on märkimisväärselt suurenenud metaboolne aktiivsus kindla ravimi suhtes, kuna neil esinevad geenimutatsioonid, mis annavad valgule suurenenud aktiivsuse (*gain-of-function, GoF*);

- 2) kiired metaboliseerijad (*rapid metaboliser*, RM) on invidiidid, kellel on suurenenud metaboolne aktiivsus;
- 3) normaalsetel metaboliseerijatel (EM) esineb normaalne metaboolne aktiivsus, mis tavaliselt esindab suuremat osa elanikkonnast;
- 4) keskmistel metaboliseerijatel (*intermediate metaboliser*, IM) on vähenenud metaboolne aktiivsus, kuna neil esineb ainult üks normaalne või üks funktsionaalselt puudulik alleel;
- 5) aeglasel metaboliseerijatel (*poor metaboliser*, PM) on invidiidid, kellel esineb puudulik või olematu metaboolne aktiivsus, mis viitab homosügootsele või ühendheterosügootsele funktsiooni kaole (LoF) (Zanger ja Schwab, 2013).

Selle spektri kaks äärmust on aeglasel metaboliseerijatel (PM) ja ülikiired metaboliseerijatel (UM) (Meyer, 2004). Invidiid võib olla ühe ravimi puhul UM ja mõne teise ravimi puhul PM. Näitlik ravimi annustamisele järgnev ravimi hulga muutuse (kontsentratsioon) muster vereplasmas on illustreeritud joonisel 3 normaalse (EM) ja aeglase (PM) metaboliseerija puhul.



Joonis 3. P450-te (või teiste ensüümide) vähenenud metaboolse aktiivsuse olulisus aktiivsete ravimite lagundamisel. “Tüüpiline” muster on toodud välja ülemises paneelis (EM), mille puhul hoitakse konkreetse korduva annuse määramisel ravimi vereplasma taset teatud vahemikus. Aeglase metaboliseerija (alumine paneel, PM) puhul tõuseb ravimi tase vereplasmas iga järgneva annuse manustamisel. $C_{p, \max}$ – maksimaalne plasmakontsentratsioon, AUC – kõvera alune pindala. (Tõlgitud ja kohandatud, Guengerich, 2015).

Normaalse metaboliseerija (EM) puhul ravimi kontsentratsioon vereplasmas tõuseb pärast esimest annust maksimaalsele tasemele ($C_{p,max}$) ning langeb seejärel madalamale tasemele enne järgmise annuse manustamist (joonis 3). Järjestikuste annuste manustamisel jääb ravimi kontsentratsioon vereplasmas püsima kindlas vahemikus, kutsudes esile soovitud farmakoloogilise efekti (Guengerich, 2015).

Kui aeglasele metaboliseerijale (PM) anda sama suuri ravimiannuseid, toimuks annuste manustamiste vahel väga piiratud ravimi lagundamine ning ravimi kontsentratsioon vereplasmas (ning arvatavasti ka sihtmärk koes) tõuseks lõpuks ootamatult kõrgele tasemel, mida illustreerib joonisel 3 kõvera aluse pindala (AUC) suurenemine (Guengerich, 2015). See kutsuks esile liiga tugeva (ja ilmselt ka ebasoovitav) farmakoloogilise vastuse.

Kui aga tegemist oleks ülikiire metaboliseerijaga (UM), oleks nende ravimi annustamisele järgnev ravimi hulga muutumise muster vereplasmas sarnane normaalse metaboliseerija tüübile. Kuid nende puhul oleks graafikul $C_{p,max}$ ja AUC väiksemad ning sellest tulenevalt on oodatav ka vähenenud ravimi efektiivsus (Guengerich, 2015). Seega vajavad aeglase metaboliseerija (PM) fenotüübiga invidiidid väiksemat ravimiannust ning ülikiire metaboliseerija (UM) fenotüübiga invidiidid vajavad suuremat ravimiannust, et saavutada sama ravimi kontsentratsioon vereplasmas kui normaalse metaboliseerija (EM) fenotüübiga invidiididel (Meyer, 2004).

Aeglase (PM) ja keskmise (IM) metaboliseerija fenotüübiga invidiidide puhul on ravimi organismist elimineerimise kiirus vähenenud ning ravimi kontsentratsioon veres on suurenenud. Kiire (RM) ja ülikiire (UM) metaboliseerija fenotüübiga invidiidide puhul on ravimi organismist elimineerimise kiirus suurenenud ning ravimi kontsentratsioon veres on seega vähenenud. Aktiivravimi korral võivad suurenenud ravimi kontsentratsioonid kehas põhjustada toksilisust ja ravimi kõrvaltoimeid ning vähenenud kontsentratsioon võib viia väiksema või puuduva ravimõjuni.

Eelravimid vajavad toimimiseks esmalt aktiivsesse ravimivormi muundamist (biotransformatsiooni), mistõttu on metaboliseerija tüüpide puhul siis vastupidised efektid – kiire metaboliseerija (RM) fenotüübiga invidiididel tõuseb kõrvaltoimete risk, kuna veres tekib suuremas kontsentratsioonis aktiivseid metaboliite ning aeglase metaboliseerija (PM) fenotüübiga invidiididel on aktiivsete metaboliitide kontsentratsioon väiksem, mis võib viia vähenenud või puuduva ravimõjuni. (Zanger ja Schwab, 2013)

1.2.4. Ravimite lagundamine ja seos tsütokroom P450-tega

Paljud kehavõõrad ühendid, nagu ravimid, taimsed metaboliidid ja keskkonna saastained on mittepolaarsed ja lahustuvad lipiidides kergesti. Ravimeid lagundavate ensüümide peamine ülesanne on muuta need kemikaalid vees hästilahustuvateks produktideks, mida on seejärel võimalik kehast lihtsamini eemaldada (Wilkinson, 2001). Tsütokroom P450 ensüümid on seotud enamiku ravimite ja ksenobiootikumide lagundamisega, sealhulgas ka eelravimite metaboliseerimisega (Ortiz de Montellano, 2013). Inimestes näiteks on need ensüümid väga tähtsad maksas ravimite esimese faasi metabolismis. Need ensüümid on võimelised katalüüsima aktiveerimata C–H sidemete regio- ja stereospetsiifilisi oksüdatiivseid reaktsioone. (Gilardi, 2018)

Kui farmakogeneetika uurib geenide mõju ravimi vastusele, siis farmakogenoomika uurib muuhulgas ka, kuidas ravimid mõjutavad geenide funktsiooni (Kalow, 2006). Näiteks võib ravimi regulaarne tarbimine suurendada ekspressiooni geenil, mis kontrollib ravimit lagundavat ensüümi (Conney, 1967), see võib mõnel juhul viia ka ravimi suhtes sõltuvuse tekkimiseni (Rhodes ja Crabbe, 2005).

CYP geenide evolutsiooni üks peamisi kliinilisi tagajärgi on erisused ravimite metabolismis. *CYP* geenivariantide ja ka metabolismitüüpide esinemissagedused on erinevates populatsioonides erinevad. *CYP2D6* aeglast metaboliseerija fenotüüpi leidub umbes 5 – 10% kaukaaslastest, kuid idamaade populatsioonides esineb seda oluliselt vähem (Evans, 1993). Seevastu *CYP2C19* funktsiooni häiretega seotud aeglaste metaboliseerijate esinemine idamaalaste populatsioonides on 15 – 20%, samas kui kaukaaslaste seas on see esinemissagedus ainult 3 – 5% (Inaba jt., 1995; Goldstein ja Blaisdell, 1996). Arvatakse, et alleelvariantide esinemissageduste erinevused võivad evolutsioonilisest aspektist lähtudes olla seotud toitumise erinevustega populatsioonide vahel.

Ravimid, mille peamised metaboliseerivad ensüümid on polümorfse geneetilise regulatsiooni all, võivad põhjustada märkimisväärset individuaalset farmakokineetilist varieeruvust, mis võib omakorda põhjustada kliiniliselt olulisi erinevusi mõnede patsientide ravimvastuses ning viia ravimi arendamise peatamiseni (Balant jt., 1989). Ravimikandidaadid, mida lagundab ainult üks ning seejuures väga varieeruva geenijärjestusega P450 (nagu seda on *CYP2D6* või *CYP2C19*), langevad tihti edasisest ravimiarendusest välja. Kliiniliselt oluliste geen-ravim seoste andmebaasid on näiteks CPIC (*Clinical Pharmacogenetics Implementation*

Consortium), mis haldab ning teeb avalikult kättesaadavaks detailseid geen-ravim kliinilise praktika juhendeid, et lihtsustada farmakogeneetilistest testidest saadavate tulemuste kliinilist kasutamist (CPIC, 2019) ning SuperCYP (*Cytochrome P450 database*), mis sisaldab 2785 teadaoleva koostoime kirjeldust tsütokroom ensüümide ja ravimite vahel (SuperCYP, 2021).

Katseloomades ja inimestes on tuvastatud erinevaid *CYP* genee ja vastavaid ensüüme, mida kummaski teises liigis ei leidu (Nebert jt., 1989). Sellised liigilised erinevused võivad põhjustada raskusi katseloomadele toksiliste või vähki tekitavate uurimuste tulemuste ekstrapoleerimisel kliinilisse meditsiini (Balant jt., 1989).

1.3. CYP2C19

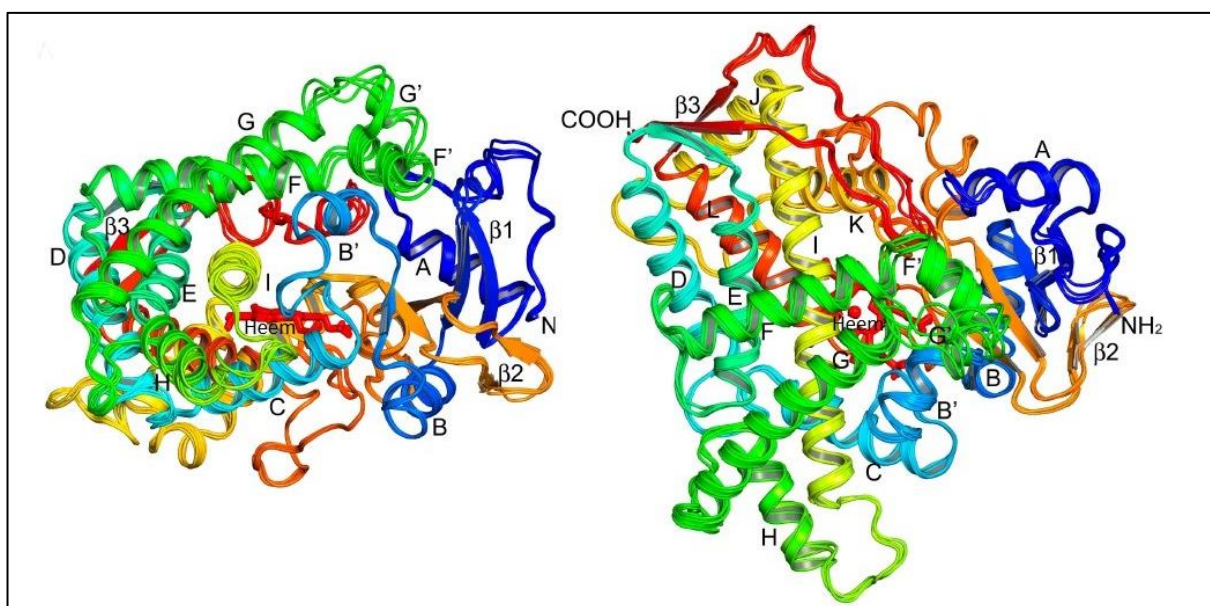
CYP2C19 geen (ENSG00000165841) on tsütokroom P450 perekonna 2, alam perekonna C liige 19 (HUGO Gene Nomenclature Committee: CYP2C19, 2019). Tsütokroom P450 2C alam perekond moodustab umbes 20% kõigist inimese maksas leiduvatest P450-test ja on üks olulisematest P450 alam perekondadest, kuna sinna kuuluvad ensüümid lagundavad üle 20% kõigist terapeutilistest ravimitest ning lisaks ka mitmeid endogeenseid ühendeid (Shimada jt., 1994; Rendic, 2002). *CYP2C* alam perekonnas on neli liiget: *CYP2C8*, *CYP2C9*, *CYP2C18*, *CYP2C19*; ning kuigi nende nelja geeni poolt kodeeritud ensüümid on üle 77% identsed oma aminohappejärjestustes, on need siiski üsna erinevad substraadi spetsiifilisuse poolest (Ridderström jt., 2001). *CYP2C8*, *CYP2C9* ja *CYP2C19* geneetilisi polümorfisme on laialdaselt uuritud, kuid kõige enam tuntud on *CYP2C19* polümorfismid (Niwa ja Yamazaki, 2012). Oma töös keskendungi just sellele geenile ja selle geeni produktile ning polümorfismidele.

CYP2C19 geen paikneb tsütokroom P450 geenide klastris kromosoomil 10q23.33, positsioonis (GRCh38) 94 762 681 – 94 855 547 (NCBI: CYP2C19, 2019), mis teeb geeni pikkuseks 920 866 aluspaari (bp) ehk 92,8 kbp. Referents genoomi GRCh37 põhiselt paikneb *CYP2C19* positsioonis 10: 96 447 911 – 96 613 017, mis teeb geeni pikkuseks aga 165 106 bp ehk 165,1 kbp. GRCh37 referents genoomi järgselt on *CYP2C19*-l kattuvus *CYP2C18*-ga ligikaudu 48 kbp, kuid GRCh38 puhul on need geenid üksteisest eraldi. *CYP2C19* geenil on 9 eksonit ja 8 intronit, ning see kodeerib tsütokroom P450 ensüümide superperekonna polüpeptiidi 2C19.

S-mefenütoiin-4-hüdroksüülimine on tüüpiline P450 2C19 valgu poolt läbiviidav reaktsioon (Guengerich, 2015), kuid esineb palju teisi demetüülamise, oksüdeerimise, hüdroksüülamise ja muid reaktsioone, milles P450 2C19 osaleb (Rendic, 2002). Tavaliselt stimuleerib P450 2C19

ja teiste 2C alamperekonna ensüümide aktiivsust tsütokroom b₅ (Yamazaki jt., 2002). 2C19 puhul ei sõltu stimulatsioon tsütokroom b₅-e heemist ja pole seega seotud elektronide ülekandega (Yamazaki jt., 2002).

P450 2C19 valgust on olemas röntgenkristallstruktuur (Reynald jt., 2012). Reynald jt. (2012) võrdlevad enda artiklis P450 2C19 struktuuri 2C8 ja 2C9 valkude struktuuriga, ning nad väidavad, et P450 2C19 valgu selgroo konformatsioon on kõige sarnasem 2C8 omaga, kuigi 2C8 substraati siduv õõnsus (aktiivsaite) on palju suurem kui 2C19 oma ning 2C19 ja 2C9 aktiivsaidid on suuruselt sarnasemad, kuid konformatsiooniliselt erinevad. P450 2C19 kompleksi tertsiaarstruktuuris (joonis 4) on 12 α -heeliksit ja 3 β -lehte, mis on tavaline P450-te puhul. Lisaks esineb 2C19 ensüümil B – C heeliksite vahelisel lingul heeliks B', ning F – G heeliksite vahelisel lingul esinevad heeliksid F' ja G'. Neid lisa heelikseid on varasemalt nähtud pigem teiste P450 perekonna 2 ensüümide ehituse juures (Reynald jt., 2012).



Joonis 4. P450 2C19 tertsiaarstruktuur. Asümmeetrilise üksuse neli ahelat paiknevad üksteise peal ning iga ahel on eri värvusega, alates N-terminusest (sinine) kuni C-terminuseni (punane). Joonisel on α -heeliksid tähistatud tähtedega (A – L) ja β -lehed on nummerdatud (1 – 3). Heemi proteesigrupp on kujutatud pulgakujulisena ja punase värvusega. (Tõlgitud ja kohandatud, Reynald jt., 2012).

1.3.1. *CYP2C19* star-alleelid

CYP2C19 kohta on teada vähemalt 39 star-alleeli, mis on kõik välja toodud PharmVar andmebaasis (PharmVar: CYP2C19, 2021). Tuvastatud alleelivariantidel on üsna palju

erinevusi, mis hõlmab muu hulgas ka initsiatsioonikoodoni mutatsioone (Ferguson jt., 1998) ja ensümaatiliste omaduste muutusi (Blaisdell jt., 2002).

Kõige enam levinud *CYP2C19* null alleel on *CYP2C19*2*, mille alleelisagedus kaukaaslastel on ligikaudu 15%, afroameeriklastel 15% ja aasialastel 29 – 35% (Scott jt., 2011). *CYP2C19* geeni alleelivariant *2 põhjustab aeglase metaboliseerija (PM) fenotüüpi. Selle fenotüübiga seotud defekt tuvastati esmalt kaukaaslastes ja jaapanlastes kui *CYP2C19* ekson 5 mutatsioon, mis viib hälbiva splaissimissaidini ja seeläbi ka lühenenud valguproduktini (De Morais jt., 1994). See mutatsioon esineb enam kui 75% *CYP2C19* geenist tuleneva PM fenotüüpiga indiviididest, kuid mitte kõigil, seega esinevad ka teised *CYP2C19* mutatsioonid, mis põhjustavad PM fenotüüpi (Richardson ja Johnson, 1996).

Janha jt. (2014) leidsid mõningaid tõendeid selle kohta, et *CYP2C19*2* alleel on olnud positiivse selektsiooni all viimase 10 000 aasta jooksul, mis korreleerub süsteemsete põllumajandustavade arenguga. Põhjuseks pakuvad nad, et varased inimesed puutusid kokku üldlevinud uudse toksiiniga, mida aktiveerib *CYP2C19*, kuid selektiivse surve eest vastutavad kindlad bioloogilised tegurid on praegu teadmata (Janha jt., 2014).

CYP2C19 geeni alleelivariandi *3 puhul on tekkinud eksonis 4 enneaegne stoppkoodon, mis põhjustab samuti aeglase metaboliseerija (PM) fenotüüpi (Mizutani, 2003). Mõned teised tuvastatud star-alleelid, mis põhjustavad PM fenotüüpi on näiteks *CYP2C19*4*, *CYP2C19*7*, *CYP2C19*5*, *CYP2C19*6* ja *CYP2C19*8*. Star-alleel *4 puhul on tegemist initsiatsioonikoodoni mutatsiooniga (Ferguson jt., 1998), *7 puhul on tegemist intron 5-e splaissmutatsiooniga (Ibeanu jt., 1999), *5 põhjustab aminohappelist muutust konserveerunud heemiga seondumise piirkonnas (Xiao jt., 1997) ning *6 ja *8 põhjustavad mõlemad erinevaid kodeerimismuutusi ekson 3-s (Ibeanu jt., 1998; Ibeanu jt., 1999). Kõik need null alleelid toodavad valguprodukte, millel on madalam katalüütiline aktiivsus kui metsiktüüpi (*CYP2C19*1*) valkudel. Kõik need polümorfismid avastati *CYP2C19* sekveneerimisega inimestel, keda arstid olid esmalt määranud ravimite mefenütoiini või omeprasooli suhtes PM-ideks (Blaisdell jt., 2002).

Seni on *CYP2C19* geenil tuvastatud ainult üks alleelivariant, mille korral on valgul suurenenud katalüütiline aktiivsus. Selleks on promotori variant 806>T (*CYP2C19*17*), mis näib suurendavat ensüümi ekspressiooni ja selle aktiivsust mefenütoiini ja omeprasooli lagundamise suhtes seni ebaselge mehhanismi kaudu (Sim jt., 2006). Regulaatorne star-alleel *CYP2C19*17*

võib suurendada transkriptsiooni määra umbes kahekordselt (Sanford jt., 2013). Seda alleeli varianti on seostatud peptilise haavandtõvega (Musumba jt., 2013) ning omeprasooli ja estsitalopraami tavapärasest kiirema lagundamisega (Baldwin jt., 2008; Rudberg jt., 2008). Kliiniliselt oluliste ravimite puhul, mis metaboliseeruvad peamiselt P450 2C19 kaudu, võib *CYP2C19*17* alleel olla seotud ravimi subterapeutilise kontsentratsiooniga ja põhjustada ravi ebaõnnestumise ning selle mutatsiooniga indiviidide puhul on vaja suuremat ravimiannust (Baldwin jt., 2008).

Antud töö kontekstis on oluline rääkida ka star-alleelist *CYP2C19*37*, mis koondab enda alla kõik osalised *CYP2C19* geeni deletsioonid. Kõigi nende struktuurivariantide ühiseks jooneks on esimese eksoni puudumine (vaata tabel 1) (Botton jt., 2019). Kuna ekson 1 kodeerib aminohappeid, mis suunavad valgusünteesi endoplasmaatilisse retiikulumi, ei kodeeri need variandid tõenäoliselt funktsionaalset valku (Botton jt., 2019).

Tabel 1. *CYP2C19* osalised geeni deletsioonid (Botton jt., 2019). (Tõlgitud ja kohandatud).

Alleeli nimetus	Geenid	Deleteerunud eksonid
*37.001	<i>CYP2C18</i>	–
	<i>CYP2C19</i>	1 – 5
*37.002	<i>CYP2C18</i>	8 – 9
	<i>CYP2C19</i>	1 – 4
*37.003	<i>CYP2C18</i>	1 – 9
	<i>CYP2C19</i>	1
*37.004	<i>CYP2C18</i>	5 – 9
	<i>CYP2C19</i>	1 – 7
*37.005	<i>CYP2C18</i>	2 – 9
	<i>CYP2C19</i>	1 – 7

1.3.2. P450 2C19 substraadid

P450 2C19 valku leidub endoplasmaatilises retiikulumis ning see on kliiniliselt oluline ensüüm, mis lagundab paljusid erinevaid ravimeid, sealhulgas krambivastast ravimit nimega mafenütoiin, haavandivastaseid ravimeid nagu omeprasool, teatud antidepressante nagu näiteks estsitalopraam ja malaariavastast ravimit proguaaniil (Blaisdell jt., 2002). Järgnevalt toon välja konkreetsed näited mõnede ravimite kohta, mille metabolismis P450 2C19 osaleb.

CYP2C19 S-mefenütoiini polümorfism avastati 1980-ndate alguses ja see on üheks kliiniliselt kõige olulisemaks P450 polümorfismiks (Küpfer ja Preisig, 1984; De Morais jt., 1994; Meyer ja Zanger, 1997; Goldstein, 2001). See polümorfism on seotud indiviidi vähenenud võimekusega lagundada mefenütoiini, mis on teada kui aeglase (PM) metaboliseerija fenotüüp. PM fenotüübiga indiviididel on puudulik võime 4'hüdoksüülida mefenütoiini S-enantiomeeri, mis põhjustab ratseemilise segu lagundamise aeglasemat kiirust (Goldstein ja Blaisdell, 1996).

Antidepressandid nagu tsitalopraam ja esitsitalopraam metaboliseeruvad peamiselt P450 2C19 kaudu ning indiviididel, kellel on vähenenud ensüümi funktsionaalsus, esineb kõrgem algravimi tase kehas (Rudberg jt., 2006). Seega PM fenotüübiga patsiendid vajavad aktiivravimite, nagu esitsitalopraam, puhul madalamat annust kui EM-id. Indiviididel, kelle on *CYP2C19*17/*17* genotüüp, on suurenenud risk ravi ebaõnnestumiseks (Rosenborg jt., 2008).

P450 2C19 on peamine ensüüm malaariavastase ravimi proguanili aktiveerimisel tsükloguaniliks, kuid puuduvad kliinilised andmed, mis viitaks sellele, et PM-del on suurem oht malaaria profülaktika või ravi ebaõnnestumiseks (Desta jt., 2002).

Üks põhjalikumalt uuritud substraate on ravim klopidogreel, mis muundatakse aktiivseks vormiks kahes etapis, seejuures mõlemad etapid on katalüüsitud suures osas P450 2C19 abil (Kazui jt., 2010). Klopidogreel, mida kasutatakse adenosiindifosfaadi poolt indutseeritud trombotsüütide agregatsiooni pärssimiseks, on eelravim, mis vajab aktiivseks metaboliidiks transformeerimisel mitme CYP valgu abi (Mega jt., 2009). Vähenenud P450 2C19 funktsiooniga ja klopidogreeliga ravitud indiviididel on nähtud märkimisväärselt madalamat klopidogreeli aktiivse metaboliidi taset kehas, mistõttu on neil isikutel vähenenud organismi võimekus pärssida trombotsüütide teket ning neil on suurenenud risk raskete kardiovaskulaarsete tüsistuste tekkele võrreldes normaalse P450 2C19 funktsiooniga isikutega (Mega jt., 2009). Seevastu *CYP2C19*17/*17* genotüübiga isikud võivad kogeda suurenenud klopidogreeli aktivatsiooni, mis on seotud parema terapeutilise tulemusega (Tiroch jt., 2010), kuid ka suurenenud riskiga verejooksu tekkeks (Sibbing jt., 2010).

P450 2C19 ulatus ravimite metabolismis on üsna märkimisväärne. Erilist huvi pakkuv on haavandiravim omeprasool (ja sellega seotud ühendid), kuna madala ensüümiaktiivsusega isikud reageerivad paremini haavandite ravile (Chiba jt., 1993). On teada, et PM-de puhul esineb suurem paranemismäär omeprasooli manustavates patsientides maohaavandite ja sellega kaasnevate *Helicobacter pylori* infektsioonide korral, kuna neil indiviididel on kõrgem ravimi

kontsentratsioon veres kui EM-idel (Goldstein, 2001). See võiks aga viidata sellele, et EM-ide puhul on vaja kasutada kõrgemat ravimiannust. UM fenotüübi puhul on omeprasooli lagundamine kiirenenud, mille tulemuseks on subterapeutiline ravimi kontsentratsioon (Baldwin jt., 2008). Omeprasool samuti inhibeerib P450 2C19 aktiivsust ja vähendab seeläbi ravimite klopidoogreeli (Scott ja Scott, 2013) ja proguaaniili (Funck-Brentano jt., 1997) lagundamist.

Esinevad ka ravimid, mis indutseerivad 2C19 mRNA transkribeerimist. Uuringud inimese hepatotsüütidega näitasid 2C19 mRNA transkribeerimise indutseerimist ravimite rifampitsiin, deksametasoon ja fenobarbitaal poolt (Rauchy jt., 2002).

Paljud ravimid, mis on P450 2C19 substraatideks, on ka teistele P450 ensüümidele substraadiks (Rendic, 2002), kuid siinkohal võib ensüümide kineetiline aktiivsus nende substraatide suhtes erineda (Niwa ja Yamazaki, 2012).

1.4. Järgmise põlvkonna sekveneerimine ja mikrokiiptehnoloogiad

Järgmise põlvkonna sekveneerimise tehnoloogiaid (*next generation sequencing*, NGS) saab kasutada kogu genoomi sekveneerimiseks (*whole genome sequencing*, WGS), kogu eksoomi sekveneerimiseks (*whole exome sequencing*, WES) või suunatud sekveneerimiseks, mis tähendab ainult huvipakkuvate geenide või genoomi regioonide sekveneerimist (Rapoport jt., 2020). NGS töövoog algab DNA eraldamisega, millele järgnevad raamatukogu valmistamine (genoomse DNA fragmenteerimine ja sekveneerimise adapterite ligeerimine DNA fragmentide külge), DNA fragmentide amplifikatsioon ning lõpuks sekveneerimine ja bioinformaatiline analüüs.

Antud töös kasutatud kogu genoomi andmete saamiseks kasutati Illumina HiSeq X sekvenaatorit. Illumina järgmise põlvkonna sekveneerimise tehnoloogia kasutab kiire ja täpse sekveneerimise võimaldamiseks klonaalset amplifitseerimist ja töötab järjestuse sünteesimise põhimõttel (Illumina: Introduction to DNA sequencing, 2021). Selle tehnoloogia kasutamise puhul sisaldavad sekveneerimise adapterid spetsiifilisi järjestusi, mis on loodud seondumaks tahke kandja mikroplaadil (*flowcell*) olevate komplementaarsete oligonukleotiididega (Morey jt., 2013). Raamatukogu klonaalne amplifitseerimine toimub identse järjestusega DNA klastrite genereerimise teel. Sekveneerimisel kasutatakse nelja modifitseeritud pöördterminaatorit, mis sisaldavad eemaldatavat fluorestsentsmärgist, et iga sekveneerimise tsükliga detekteerida

sünteesitavast ahelast ühte järgmist nukleotiidi (Metzker, 2010). Illumina tehnoloogia võimaldab sekveneerimist raamatukogu sisestuse mõlemast otsast (*paired-end sequencing*).

Farmakogenoomika keskendub peamiselt SNP-ide uurimisele, kuna see on kõige levinum ja ka tehniliselt kõige kättesaadavam geneetiliste variatsioonide informatsiooni kogum. SNP-ide tuvastamine ja analüüsimine on lihtsam ja praktilisem kui näiteks kogu genoomi sekveneerimine ja sellest saadava andmekogu analüüsimine. SNP-ide määramiseks ja indiviidide genotüpiseerimiseks, kaasaarvatud geeni koopiaarvu variatsioonide (CNV) tuvastamiseks, saab kasutada mikrokiipe.

Antud töös kasutatud genotüpiseerimise andmed saadi kasutades Illumina InfiniumOmniExpress (Omni) või Infinium Global Screening Array (GSA) mikrokiipe. Mõlemad eelnimetatud mikrokiibid töötavad BeadArray tehnoloogial, kasutades ränidioksiidi mikrokerakesi. Iga mikrokiibi pinnal saab analüüsida sadu tuhandeid kuni miljoneid genotüüpe ühe indiviidi kohta korraga (Illumina: Microarray Technology, 2021). Ränidioksiid kerakesed on mikrokiibil paigutatud hoolikalt süvendatud mikroaukudesse ning on kaetud sadade tuhandete kovalentselt seotud oligonukleotiidsondi koopiatega (Shen jt., 2005). Iga oligonukleotiidsond on suunatud konkreetsele genoomi lookusele. Kui DNA fragmendid liiguvad üle mikrokiibi, seondub iga sondi külge sellega komplementaarset järjestust omav DNA fragment, misjuures sondi oligonukleotiidjärjestus lõppeb ühe aluse võrra enne huvipakkuvat lookust ehk SNP positsiooni (Illumina: Microarray Technology, 2021). Polümeraas pikendab sondi küljes olevat järjestust ühe märgistatud nukleotiidi võrra, vastavalt komplementaarsetele DNA järjestusele, mis on sondiga seotud. A ja T nukleotiidid on märgistatud dinitrofenüülga (DNP) ning G ja C on märgistatud biotiiniga (Illumina: Microarray Technology, 2021). Järgneva värvimise protsessi käigus kinnitub märgistuste (DNP ja biotiin) külge kindel fluorofoor. Laseriga ergastamisel kiirgab nukleotiidi märgistuse küljes olev fluorofoor signaali, mille intensiivsust on võimalik detekteerida ja tõlgendada ümber alleelisuhte kohta käivaks informatsiooniks antud lookuses. Näiteks on T/T homosügooti puhul näha peamiselt punast signaali ning C/C homosügooti puhul on näha peamiselt rohelist signaali. Kui aga uuritav indiviid on kindla huvipakkuva SNP-i suhtes heterosügoot, paistab signaal kollasena enda intensiivsuses. Seega on võimalik eristada, milline nukleotiid asub kindlas SNP positsioonis ning mis genotüüpi omab uuritav indiviid selle SNP-i suhtes. Mikrokiipidega saab tuvastada ka geeni koopiaarvu variatsioone (CNV), mis avaldub signaali intensiivsuses ning mille kohaselt on võimalik tuvastada geeni deletsioone ($cn \leq 1$) või duplikatsioone ($cn \geq 3$).

2. EKSPERIMENTAALOSA

2.1. Töö eesmärk

Käesoleva magistratöö eksperimentaalse osa kolm eesmärki on:

1. Määrata *CYP2C19* geeni deletsioonide esinemissagedused Eesti populatsioonis Tartu Ülikooli Eesti Geenivaramu (TÜ EGV) andmete põhjal;
2. Määrata WGS andmete põhjal *CYP2C19* deletsiooni kandjate diplotüübid;
3. Leida *CYP2C19* geeni deletsioonidega aheldunud SNP-id.

2.2. Materjalid ja meetodika

2.2.1. Andmestik

Antud töös kasutasin TÜ EGV-s olemasolevaid 2244 geenidoonori kogu genoomi sekveneerimisandmeid (WGS) ja 36 029 geenidoonori genotüübi andmeid, mis on saadud Omni (n = 5929) ja GSA (n = 30 100) mikrokiipe kasutades. Hiljem sain veel juurde 104 920 geenidoonori genotüübi andmed, mis olid saadud samuti GSA mikrokiipi kasutades. WGS andmete saamiseks kasutati Illumina HiSeq X Ten (Illumina, San Diego, CA, USA) sekvenaatorit 150 bp paarislugemi pikkuse ja 30x katvusega. Genotüübi andmed saadi kasutades Illumina InfiniumOmniExpress (Omni) või Infinium Global Screening Array (GSA) mikrokiipe. Omni kiip sisaldab 714 238 SNP markerit, keskmise mediaanvahemikuga 2,23 kb ja GSA kiip sisaldab umbes 654 027 SNP markerit, keskmise mediaanvahemikuga 2,3 kb.

Proovide töötlemine ja analüüsi ettevalmistamine tehti Tartu Ülikooli genoomika instituudis ja Broad Instituudis (USA) vastavalt tootjapoolsetele protokollidele. WGS andmeid analüüsiti Genome STRiP (*Genome structure in populations*) tarkvaraga (Handsaker jt., 2015) ning leiti 37 indiviidil (1,6%) *CYP2C19* geeni deletsioon. Mikrokiipidelt genotüüpide määramiseks kasutati PennCNV tarkvara, mille puhul määratud CNV-de keskmine pikkus on 13 kb (Wang jt., 2007). Selle tarkvaraga CNV-de määramine on väga tundlik mikrokiibilt tuvastatud signaali intensiivsuste suhtes, mis oleneb sellest kui kvaliteetselt on kiibikatse õnnestunud. Seetõttu tuli saadud tulemusi veel täiendavalt filtreerida signaali intensiivsuse parameetri järgi. Lisaks eemaldati edasisest analüüsist ka proovid, millel esines rohkem kui 200 CNV-d, mis viitab katsemürale. Geenidoonorite genotüübi andmetele tehti ka kvaliteedikontroll programmiga Plink 1.9 (Purcell jt., 2007).

Omni ja GSA kiibi andmetest tuvastati PennCNV tarkvaraga *CYP2C19* geeni CNV-sid ning leiti 116 indiviidi (1,9%) Omni andmehulgast ning 459 indiviidi (1,5%) esimesest GSA andmehulgast, kellel esines *CYP2C19* geeni deletsioon ning hilisemast GSA andmehulgast (n = 104 920) leiti kokku 2217 deletsiooni CNV juhtu (2,1%). Mõnede deletsiooniga indiviidide puhul esines hilisemates GSA andmetes mitu tuvastatud CNV-d samas geenis. Kui nende CNV-de koordinaadid olid samad, filtreerisin ühe neist CNV-dest enda andmetest välja, et mitte lugeda sama indiviidi mitu korda enda andmete hulka. Kui ühe indiviidi kohta käiva mitme CNV koordinaadid olid erinevad, liitsin need koordinaadid omavahel kokku, lugedes siis kaks lühemat CNV-d üheks pikemaks kokku. Nüüd kui andmetes esines üks CNV ühe indiviidi kohta, selgus et unikaalseid deletsiooniga indiviide oli viimases GSA andmekogus kokku 1677 (1,6%). Varasemalt ainult WGS platvormi ja Omni kiipi kasutades analüüsitud indiviide genotüpiseeriti hiljem uuesti, kasutades GSA kiipi, et kogu TÜ EGV andmestik oleks analüüsitud sama kiibiplatvormi ja markerite arvuga.

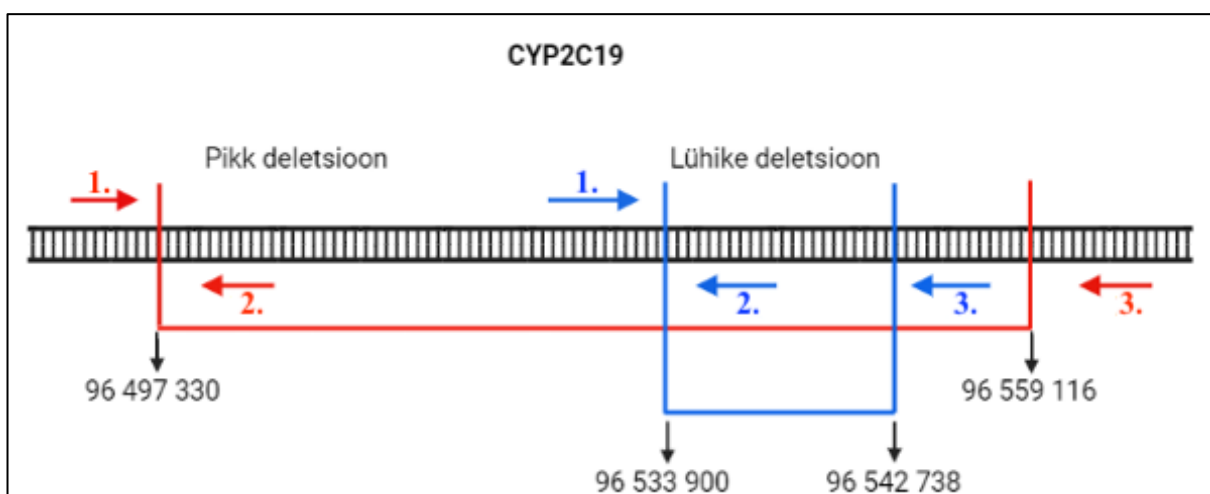
2.2.2. *CYP2C19* deletsioonide katkekohtade leidmine

WGS andmetest leitud 37 indiviidi *CYP2C19* geeni deletsioonide ennustatud algus- ja lõpppositsioonid olid (GRCh37) 10: 96 496 960 ja 10: 96 547 966 ühe deletsiooni kohta, mida edaspidi nimetan pikaks deletsiooniks, ning (GRCh37) 10: 96 530 401 ja 10: 96 542 516 teise deletsiooni kohta, mida edaspidi nimetan lühikeseks deletsiooniks. Neid koordinaate aluseks võttes, kasutasin Samtools tarkvara, et käsuga *tview* vaadata deletsiooniga indiviidide BAM faile võrrelduna referents genoomiga (GRCh37) ja seeläbi leida deletsioonide täpsed katkekohad (*breakpoints*).

Mõlema deletsiooni puhul tuvastasin visuaalse vaatluse käigus deletsioonide katkekohtade vahetust lähedusest mikrohomoloogia (6 bp pika deletsiooni ja 2 bp lühikese deletsiooni puhul). Seetõttu tegin katkekohtade kinnitamiseks pika ja lühikese deletsiooni unikaalsete 25 aluspaari (bp) pikkuste (25-mer) järjestuste otsingu WGS genoomiandmetest. Esmalt tekitasin programmiga GlistMaker WGS andmekogu indiviidide genoomi järjestustest 25-mer-ide listid. Seejärel otsisin nendest listidest deletsioonide põhiseid 25-mer-e ja sain nende esinemissageduse iga indiviidi kohta. Selleks kasutatud pika deletsiooni 25-mer oli CAGATTATAGGGTTTCTGCTGAGAG ja lühikese deletsiooni 25-mer oli AAGCCCAGATGGGATAACATTAAGA.

2.2.3. Alleelispetsiifiliste PCR praimerite disainimine

CYP2C19 geeni deletsioonide kinnitamiseks disainisin kolm praimerit kummagi deletsiooni kohta: 1) päripidine praimer; 2) äraspidine praimer deleteerunud ala sisse jääva järjestuse põhiselt, et tuvastada metsiktüüpi alleeli; 3) äraspidine praimer deleteerunud alast väljaspoole jääva järjestuse põhiselt, et tuvastada deleteerunud alleeli. Lühikese deletsiooni tuvastamiseks disainitud kõigi kolme praimerite seondumiskohad jäävad pika deletsiooni ala sisse (joonis 5).



Joonis 5. Illustreeriv joonis praimerite seondumiskohtade paiknemisest teineteise suhtes. Deletsioonide alguse ja lõpu koordinaadid on referents genoomi GRCh37 põhised. Punase joonega on märgitud pikk deletsioon ja sinise joonega lühike deletsioon. Vastavalt punaste ja siniste nooltega on märgitud deletsioonide tuvastamiseks kasutatud praimerite seondumiskohad. Joonisel on praimerid nummerdatud tabelites 2 ja 3 olevate praimerite põhiselt.

Pika deletsiooni päripidise praimerite järjestuse võtsin sama, mis on Santos jt. (2018) artiklis välja toodud, ning praimerite seondumiskoht asub deletsiooni alguspunktist umbes 300 bp eespool. Äraspidiste praimerite disainimiseks kasutasin esmalt programmi SNPmasker 1.1 (Anderson jt., 2006), referents genoomiga GRCh37/hg19, et leida soovitud piirkonnas unikaalset genoomi järjestust ja seal esinevaid SNP-e. Seejärel kasutasin programmi Primer3 (Koressaar ja Remm, 2007; Untergasser jt., 2012), et leida SNPmasker-iga saadud järjestuste seast praimerite kandidaate. Leitud praimerikandidaatide seast sobivate praimerite valikuks kasutasin NetPrimer tarkvara (NetPrimer, 2019). Eelnevalt nimetatud veebipõhise tarkvaraga kontrollisin praimeripaaride võimalike dimeeride teket, GC sisaldust ja sulamistemperatuuri (T_m), et valida enda töös kasutatavad praimerid (tabel 2). Pika deletsiooni äraspidiste praimerite seondumiskohad jäid vastavalt 150 bp pärast pika deletsiooni alguspunkti

deletsioonita alleeli tuvastamiseks, ning umbes 900 bp pärast deletsiooni lõpp-punkti deletsiooniga alleeli tuvastamiseks.

Tabel 2. PCR praimerid *CYP2C19* geeni pika deletsiooni kinnitamiseks.

	Nimetus	Järjestus (5' -> 3')	T _m (°C)
1.	wt_F_2C19	ATTAGCAATGTTGCCYGAAG	55
2.	wt_R_2C19	GTTACATCTGTTACTGGATTGTAGA	61
3.	mut_R_EGV_2C19	GGTTATACCCAGGCAAAAAGG	59

Lühikese deletsiooni praimerite (tabel 3) disainimiseks kasutasin sama meetodikat. Päripidise praimerite seondumiskoht asub deletsiooni alguspunktist umbes 900 bp eespool. Äraspidiste praimerite seondumiskohad jäid vastavalt 119 bp pärast lühikese deletsiooni alguspunkti, deletsioonita alleeli tuvastamiseks ning umbes 500 bp pärast deletsiooni lõpp-punkti, deletsiooniga alleeli tuvastamiseks.

Tabel 3. PCR praimerid *CYP2C19* geeni lühikese deletsiooni kinnitamiseks.

	Nimetus	Järjestus (5' -> 3')	T _m (°C)
1.	wt_F_s2C19	AGTGAAAGGGATCCARTGGT	57
2.	wt_R_s2C19	CCCTGAAGTTTTACCGGTGC	60
3.	mut_R_s2C19	TGGTGCAATGACAAAAGAGGT	57

Pika deletsiooni päripidine praimer sisaldab ühte olulist SNP-i (rs535814835), mille MAF on $T = 0.0042$ (21/5008, 1000G). Seetõttu on antud positsioonis nukleotiidi C asemel märgitud IUPAC koodi kohaselt Y, mida kasutatakse tähistamiseks kindlas positsioonis olevat C või T nukleotiidi. Lühikese deletsiooni päripidine praimer sisaldab samuti ühte olulist SNP-i (rs61249521), mille MAF on $A = 0.0721$ (323/4480, Eesti populatsioonist leituna). Seetõttu on antud positsioonis nukleotiidi G asemel märgitud IUPAC koodi kohaselt R, mida kasutatakse tähistamiseks kindlas positsioonis olevat G või A nukleotiidi.

2.2.4. PCR ja geelelektroforees

CYP2C19 deletsioonide katseliseks kinnitamiseks kasutasin PCR-i meetodit ja geelelektroforeesi. 20 µl PCR reaktsioonisegu sisaldas 13 µl H₂O-d, 4 µl 5x HOT FIREPol® puhvrit (Solis BioDyne), 0,5 µl kumbagi praimerit ja 2 µl DNA-d (5 ng/µl). Kasutatud PCR programm on loodud Solis BioDyne-i protokollide juhiste alusel ning on toodud välja tabelis 4.

Tabel 4. Töös kasutatud PCR programm.

Temperatuur	Aeg	Tsükli arv
95°C	15 minutit	1
95°C	30 sekundit	35 tsükli
55°C	30 sekundit	
72°C	1,5 minutit	
72°C	10 minutit	1
4°C	hoia	

PCR produktide visualiseerimiseks kasutasin 1,5% agarosgeeli (millele oli lisatud etiidiumbromiidi) ning 6x TriTrack DNA laadimisvärvi (ThermoFisher Scientific). Suurusmarkerina kasutasin Marker GeneRuler™ 100 bp Plus DNA markerit (ThermoFisher Scientific), mille fragmentide suurus on vahemikus 100 bp – 3000 bp. Saadavate PCR produktide pikkused on toodud välja tabelis 5. PCR produktide pikkuste erinevuste põhjal sain tuvastada geelelektroforeesiga deletsioonita ja deletsiooniga alleelide esinemist.

Tabel 5. Kasutatud PCR praimerid ja saadavad PCR produktide pikkused.

Praimerid	Produkti pikkus (bp)
wt_F_2C19 + wt_R_2C19	445
wt_F_2C19 + mut_R_EGV_2C19	1263
wt_F_s2C19 + wt_R_s2C19	1105
wt_F_s2C19 + mut_R_s2C19	1533

2.2.5. Deletsiooniga indiviidide diplotüüpide määramine

WGS ühenukleotiidsete variatsioonide (*single nucleotide variation*, SNV) andmeid kasutades määrasin *CYP2C19* geeni deletsiooniga indiviidide diplotüüpe. Selleks võtsin PharmVar ja LOVD (*Leiden Open Variation Database*) andmebaasidest *CYP2C19* geeni kohta käivad star-alleelid ja nendega seotud SNP-id. Ma vaatasin iga SNP-i esinemissagedust eurooplaste seas ning selle põhjal jätsin välja SNP-id, mille esinemissagedus oli 0. Samuti jätsin välja mõned star-alleelid, mille puhul SNP-idega seotus polnud lõplikult tõendatud. Diplotüüpide määramiseks kasutatud SNP-id ja nende vastavad star-alleelid on välja toodud tabelis 6.

Diplotüüpide määramisel on kõigil deletsiooniga indiviididel (n = 37) üks alleelidest *37, mis tähistab osalist geeni deletsiooni. Teise alleeli määramiseks vaatasin iga indiviidi puhul, mis

nukleotiid asus neil kindlas star-alleeli SNP positsioonis ning järeldasin nende haplotüüpe selle põhjal. Osa kasutatud SNP-idest jäid deletsioonide alast välja, mistõttu kasutasin ma Tartu Ülikooli Genoomika instituudist saadud WGS faasitud andmeid, et kindlaks määrata, kas need SNP-id asuvad üksteisega samal alleelil. Faasimine oli tehtud kasutades Eagle v2.3 tarkvara.

Tabel 6. Diplo tüüpide määramiseks kasutatud SNP-id, nende positsioonid 10. kromosoomil (GRCh37), referents ja alternatiivne nukleotiid, MAF Eesti populatsioonis (n = 4488), seostatud star-alleelid ning antud star-alleelist tulenevalt ensüümi kliiniline funktsionaalsus (PharmVar: CYP2C19, 2021; LOVD: CYP2C19, 2021).

SNP ID	Positsioon genoomis	Referents nukleotiid	Alternatiivne nukleotiid	MAF	Star-alleel	Kliiniline funktsionaalsus
rs11188072	96 519 061	C	T	0,261	*17	Suurenenud
rs3814637	96 521 045	C	T	0,072	*1; *15	Normaalne; normaalne
rs11568732	96 521 574	T	G	0,071	*1; *3	Normaalne; puudub
rs12248560	96 521 657	C	T	0,260	*17	Suurenenud
rs4986894	96 522 365	T	C	0,137	*2	Puudub
rs28399504	96 522 463	A	G	0,002	*4	Puudub
rs17885098	96 522 561	T	C	0,072	*1	Normaalne
rs17878459	96 534 922	G	C	0,027	*2	Puudub
rs12769205	96 535 124	A	G	0,136	*2; *35	Puudub; puudub
rs41291556	96 535 173	T	C	0,001	*8	Puudub
rs58973490	96 535 264	G	A	0,002	*11	Normaalne
rs181297724	96 535 296	G	C	0,010	*2	Puudub
rs4986893	96 540 410	G	A	0,001	*3	Puudub
rs7088784	96 541 373	A	G	0,072	*3	Puudub
rs4244285	96 541 616	G	A	0,137	*2	Puudub
rs3758580	96 602 622	C	T	0,136	*2	Puudub
rs3758581	96 602 623	G	A	0,070	*1; *3	Normaalne; puudub
rs17882744	96 602 691	C	T	0,003	*1	Normaalne
rs17886522	96 609 775	A	C	0,001	*1	Normaalne

2.2.6. Deletsioonidega aheldunud SNP-ide leidmine

Ma kasutasin WGS SNV andmeid, et leida deletsioonidega aheldunud SNP-e, mida oleks võimalik edaspidi kasutada indiviidide genotüüpiseerimiseks. Selleks tegin 10. kromosoomi SNV andmetest otsingu alleeli esinemise arvu järgi, mis võrdus deletsiooniga indiviidide arvuga. Seejärel sorteerisin välja SNP-id, mis ei esinenud kõigil deletsiooniga indiviididel

heterosügootselt, kuna leitud deletsioone esines WGS andmekogu indiviididel ainult heterosügootselt. Tulemuseks olid SNP-id, mis esinesid heterosügootselt ainult deletsiooniga indiviididel eraldi pika ja lühikese deletsiooni kohta. Iga leitud SNP-i puhul vaatasin ka, kus nad genoomis asuvad Ensembl Genome Browser-iga GRCh37 referents genoomi põhised, et näha, kas nad paiknevad genoomi mittekodeerivas alas, või mõne geeni või reguleeriva elemendi alas (Ensembl, 2021).

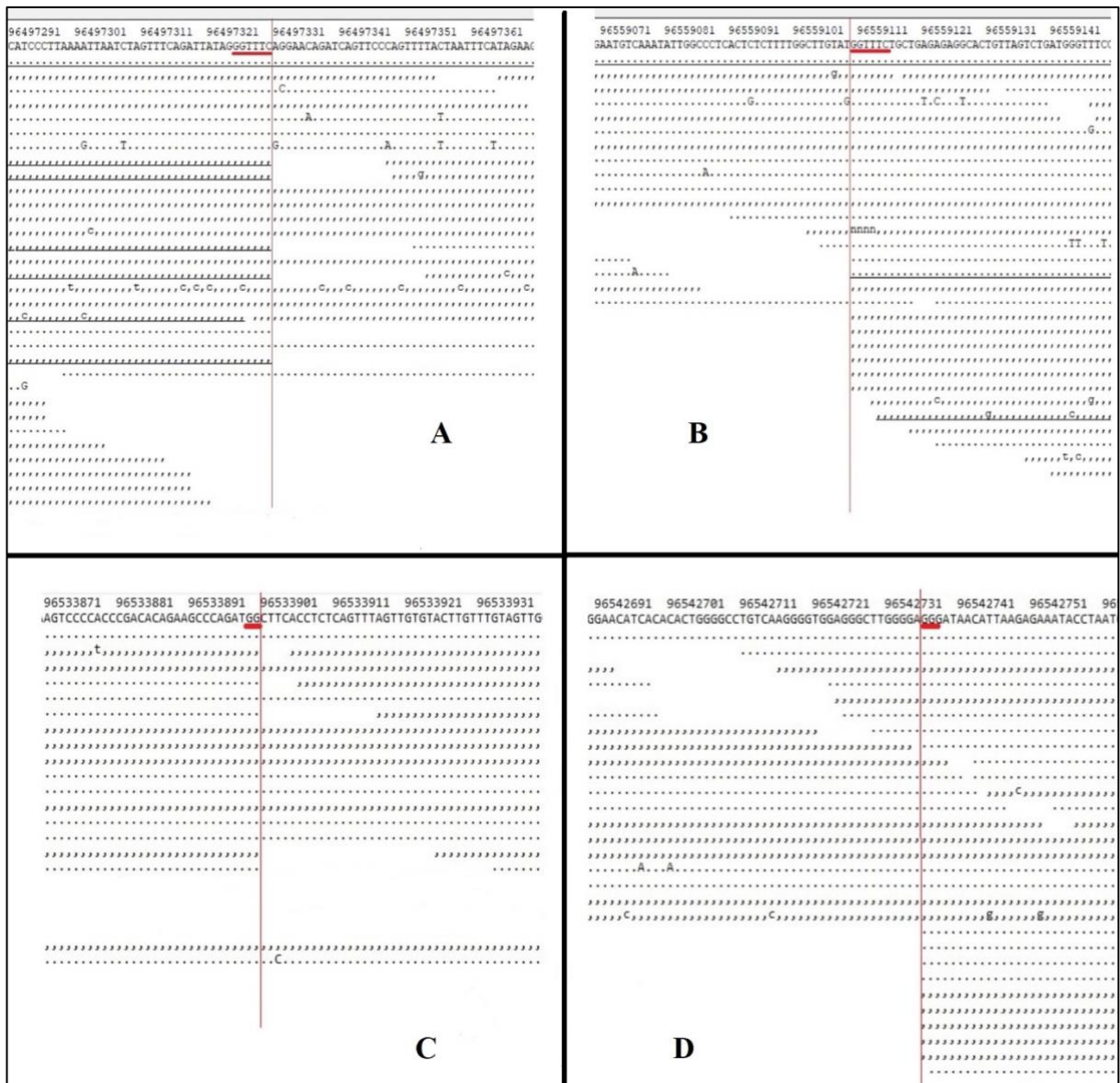
Leitud potentsiaalselt deletsiooniga aheldunud SNP-e kasutasin tuvastamiseks deletsioonide kandjate viimase ühise esivanema (*most recent common ancestor*, MRCA) vanust põlvkondades. Selleks kasutasin Human Genome Dating veebilehekülje tööriista GEVA geenivariantide vanuse genealoogiliseks hindamiseks (Human Genome Dating, 2021).

2.3. Tulemused

2.3.1. *CYP2C19* deletsioonide katkekohad

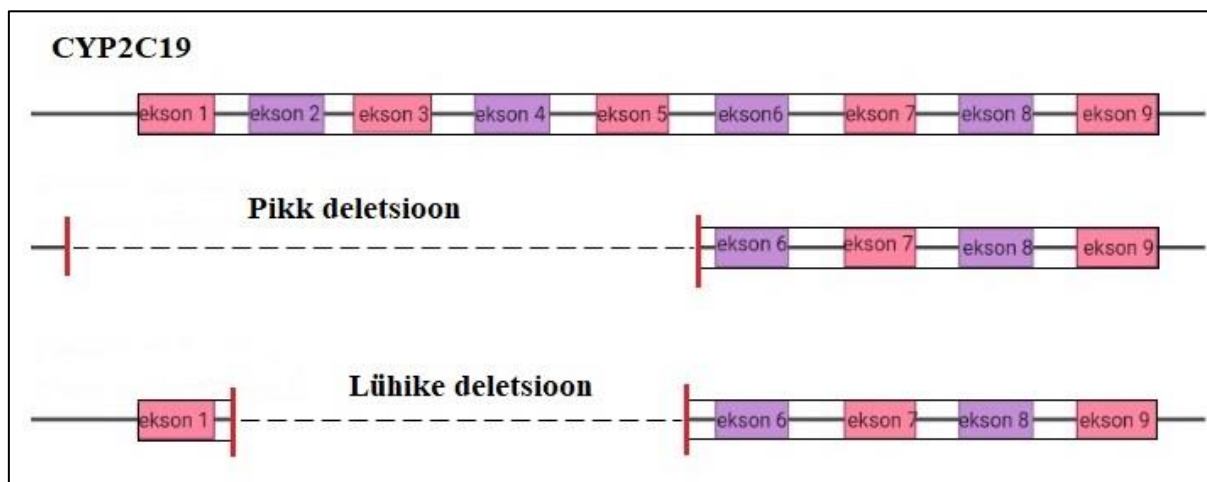
WGS andmetest leitud *CYP2C19* geeni deletsiooniga indiviidide andmete põhjal visuaalselt tuvastatud ja deletsioonide unikaalsete 25-mer-idega kinnitatud deletsioonide katkekohad on välja toodud joonisel 6. Sellelt jooniselt (joonis 6. A ja B) on näha, et pika deletsiooni katkekohtades esineb 6-nukleotiidne mikrohomoloogia järjestusega GGTTC. Lugesin pika deletsiooni alguspunktiks C nukleotiidi, mis paikneb vahetult enne deletsiooni positsioonis (GRCh37) 10: 96 497 330. Lõpp-punktiks lugesin korduvale 6-nukleotiidsele järjestusele järgnevat T nukleotiidi positsioonis (GRCh37) 10: 96 559 116. See teeb pika deletsiooni pikkuseks 61 787 bp ehk 61,7 kbp. GRCh38 referents genoomi järgselt asub pikk deletsioon koordinaatidel 10: 94 737 573 – 94 799 359.

Lugesin lühikese deletsiooni alguspunktiks G nukleotiidi (joonis 6. C ja D), mis paikneb vahetult enne deletsiooni positsioonis (GRCh37) 10: 96 533 900. Lõpp-punktiks lugesin korduvale 2 nukleotiidse GG järjestusele järgnevat G nukleotiidi positsioonis (GRCh37) 10: 96 542 738. See teeb lühikese deletsiooni pikkuseks 8837 bp ehk 8,8 kbp. GRCh38 referents genoomi järgselt asub lühike deletsioon koordinaatidel 10: 94 774 143 – 94 782 981.



Joonis 6. Pika (A ja B) ja lühikese (C ja D) deletsiooni katkekohtade leidmine visuaalselt lugemite ära langemise järgi ja võrrelduna referents genoomi järjestusega (GRCh37). Joonisel tähistab punkt (.) ühe lugemi kindlas positsioonis olevat nukleotiidi, mis on sama, mis referents genoomis peaahelel selles samas positsioonis olev nukleotiid. Koma (,) tähistab pöördahelal olevat nukleotiidi, mis on sama, mis referents genoomi pöördahelal olev nukleotiid selles positsioonis. Suurte tähtedega (A, T, G, C) on tähistatud lugemil leitud referents genoomi peaahele nukleotiidist erinev nukleotiid ning väikeste tähtedega (a, t, g, c) on tähistatud referents genoomi pöördahela nukleotiidist erinev nukleotiid. Allajoonitud lugemi alad paigutuvad referents genoomi kohaselt ka mujale genoomi piirkonda.. **A.** Visuaalselt leitud tõenäoline deletsiooni alguse katkekoht on märgitud punase vertikaalse joonega. **B.** Visuaalselt leitud tõenäoline deletsiooni lõpu katkekoht on märgitud punase vertikaalse joonega. Pika deletsiooni katkekohtades tuvastatud mikrohomologia, 6 bp pikkune järjestus GGTTTC, on punase kriipsuga alla joonitud (A ja B). **C.** Visuaalselt leitud tõenäoline deletsiooni alguse katkekoht on märgitud punase vertikaalse joonega. **D.** Visuaalselt leitud tõenäoline deletsiooni lõpu katkekoht on märgitud punase vertikaalse joonega. Lühikese deletsiooni katkekohtades tuvastatud korduv 2 bp pikkune järjestus GG, on punase kriipsuga alla joonitud (C ja D).

Pikk deletsioon algab enne *CYP2C19* geeni esimest eksonit ja lõpeb intronis 5, mis tähendab, et selle deletsiooniga on kadunud eksonid 1 – 5. Lühike deletsioon algab *CYP2C19* geeni intronis 1 ja lõpeb intronis 5, mis tähendab, et selle deletsiooniga on kadunud eksonid 2 – 5. Deletsioonide paiknemine *CYP2C19* geenil on illustreeritud joonisel 7.



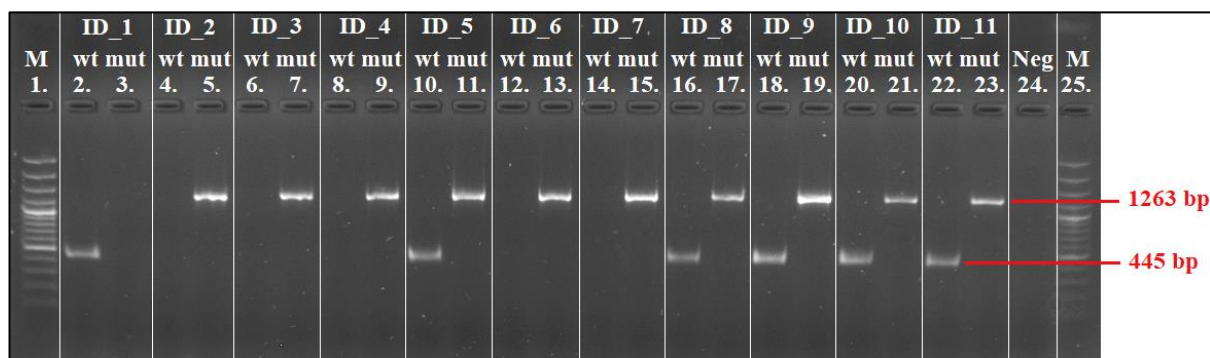
Joonis 7. Illustratsioon *CYP2C19* geenist, märgitud eksonitega. Pikk deletsioon algab enne geeni esimest eksonit ning lõpeb viiendas intronis. Lühike deletsioon algab esimeses eksonis ja lõpeb viiendas intronis. Deletsioonide katkekohad on märgitud punase vertikaalse joonega.

2.3.2. Leitud *CYP2C19* deletsioonide katseline valideerimine

WGS (n= 2244), GSA (n = 30 100) ja Omni (n = 5929) andmete põhjal oli mulle teada kõigi leitud *CYP2C19* deletsioonide ennustatud algus- ja lõppkoordinaadid. Jaotasin invidiidid gruppidesse vastavalt nendele koordinaatidele, eeldades, et samasse gruppi kuuluvatel invidiididel esineb sama deletsioon. Hilisematest GSA (n = 104 920) andmetest leitud deletsiooniga invidiidid grupeerisin eelmiste andmetega kokku ning võtsin igast grupist vähemalt kolm invidiidi katseliseks deletsiooni valideerimiseks.

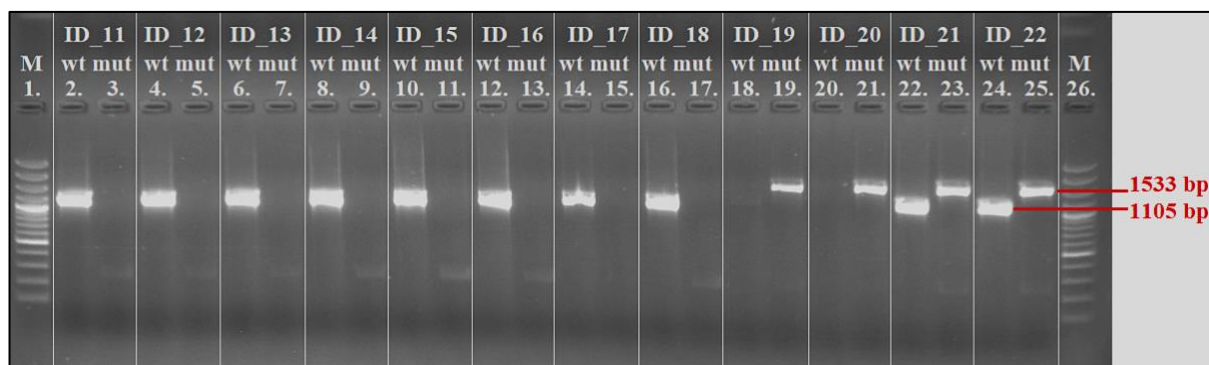
Katseliselt kontrollisin kokku 147 invidiidi, sealhulgas kontrollisin kõiki invidiide, kelle määratud geeni koopiaarv oli 0. Iga invidiidi kohta tegin mõlema deletsiooni kontrollimise puhul kaks eraldi PCR reaktsiooni ning need on geelipildidel vastavalt märgitud wt – metsiktüüpi alleeli kinnitamiseks ja mut – deletsiooniga alleeli kinnitamiseks (joonis 8 ja 9). Joonistel on invidiidid märgitud vastavalt koodidega ID_nr.

Pika deletsiooni kontrollimiseks tehtud PCR katsete tulemuste näitena on joonisel 8 välja toodud 11 indiviidi PCR tulemused. Metsiktüüpi alleeli kinnitav PCR produkti suurus on 445 bp ja pika deletsiooni alleeli kinnitav PCR produkt on suurusega 1263 bp. Jooniselt 8 on näha, et indiviidil ID_1 kinnitus ainult metsiktüüpi alleel, indiviididel ID_2, ID_3, ID_4, ID_6 ja ID_7 kinnitus ainult pika deletsiooniga alleel ning indiviididel ID_5, ID_8, ID_9, ID_10 ja ID_11 kinnitusid mõlemad alleelid.



Joonis 8. Geelipilt pika deletsiooni kinnitamiseks tehtud PCR tulemustest. Indiviidid on tähistatud koodidega ID_1 – ID_11. Iga indiviidi kohta tein kaks PCR reaktsiooni metsiktüüpi ja pika deletsiooniga alleelide olemasolu kinnitamiseks, märgitud joonisel geeliradade üleval vastavalt wt ja mut. Metsiktüüpi alleeli olemasolu kinnitab PCR produkt suurusega 445 bp ning pika deletsiooni alleeli olemasolu kinnitab PCR produkt suurusega 1263 bp, mõlemad välja toodud joonise servas punasega.

Lühikese deletsiooni kinnitamiseks tehtud PCR katsete tulemuste näitena on joonisel 9 välja toodud 12 indiviidi PCR tulemused. Metsiktüüpi alleeli kinnitav PCR produkt on suurusega 1105 bp ning lühikest deletsiooni kinnitav PCR produkt on suurusega 1533 bp. Jooniselt 9 on näha, et indiviididel ID_11, ID_12, ID_13, ID_14, ID_15, ID_16, ID_17 ja ID_18 kinnitus ainult metsiktüüpi alleel, indiviididel ID_19 ja ID_20 kinnitus ainult lühikese deletsiooniga alleel ning indiviididel ID_21 ja ID_22 kinnitusid mõlemad alleelid.



Joonis 9. Geelipilt lühikese deletsiooni kinnitamiseks tehtud PCR tulemustest. Indiviidid on tähistatud koodidega ID_11 – ID_22. Iga indiviidi kohta tein kaks PCR reaktsiooni metsiktüüpi ja lühikese deletsiooniga alleelide olemasolu kinnitamiseks, märgitud joonisel geeliradade üleval vastavalt wt ja mut. Metsiktüüpi alleeli olemasolu kinnitab PCR produkt suurusega 1105 bp ning lühikese deletsiooni alleeli olemasolu kinnitab PCR produkt suurusega 1533 bp, mõlemad välja toodud joonise servas punasega.

PCR praimerid on disainitud nii, et lühikese deletsiooni tuvastamiseks kasutatud praimerite seondumiskohad jäid pika deletsiooni ala sisse. Seega kui indiviidil esines ainult pikk deletsioon, ei olnud geelil näha lühikese deletsiooni praimeritega testimisel metsiktüüpi ega deletsiooni alleeli vööti. Kui indiviidil esines ainult lühike deletsioon, oli geelil näha pika deletsiooni praimeritega testimisel metsiktüüpi alleeli vööti. Kui indiviidil esinesid mõlemad deletsioonid, oli pika deletsiooni praimeritega testimisel näha nii metsiktüüpi kui ka deletsiooni alleeli vööti ning lühikese deletsiooni praimeritega testimisel oli näha ainult deletsiooni alleeli vööti.

Katseliselt kontrollitud 147 indiviidi puhul leidsin kokku 45, kellel minu uuritavad deletsioonid ei kinnitunud, neist 15 olid duplikatsioonidega ($cn = 3$) ning ülejäänud 30 olid heterosügootsed deletsiooni kandjad ($cn = 1$). Kõigi mittekinnitunud indiviidide puhul oli geelilt näha ainult metsiktüüpi alleeli olemasolu kinnitav PCR produkt. 102 indiviidist, kellel deletsioon kinnitus, olid 88 koopiaarvuga $cn = 1$ ehk heterosügootid ja 14 koopiaarvuga $cn = 0$ ehk homosügootid või mõlema deletsiooni kandjad. 88 heterosügootist olid 71 pika deletsiooni ja 17 lühikese deletsiooni kandjad. Ülejäänud 14 indiviidist kinnitus üheteistkümmel pika deletsiooni homosügootsus ning kolmel kinnitusid mõlemad deletsioonid.

Nagu eelnevalt mainitud, siis grupeerisin indiviidid nende deletsioonide ennustatud algus- ja lõppkoordinaatide põhjal erinevatesse gruppidesse. Samasse gruppi kuulunud indiviidide deletsioonide genotüübid kinnitusid kõigil samadena. Kindlast grupist võetud katseliselt valideeritud indiviididel kinnitunud genotüübid määrasin kõigi antud gruppi kuulunud

indiviidide genotüübiks. Duplikatsiooniga indiviidid jätsin edasisest andmeanalüüsist välja. Kõigi gruppide peale kokku leidsin 35 deletsiooniga indiviidi, kellel minu uuritavad deletsioonid ei kinnitunud ning 2254 indiviidi, kellel kinnitus. Neist 2254 indiviidist olid 2163 pika deletsiooni heterosügootid, 77 lühikese deletsiooni heterosügootid, 11 pika deletsiooni homosügootid ning kolmel indiviidil kinnitusid mõlemad deletsioonid ehk siis neil esineb pikk deletsioon ühel alleelil ning lühike deletsioon teisel. Oluline on silmas pidada, et osasid indiviide, keda analüüsiti WGS ja Omni platvormidega, analüüsiti uuesti hiljem ka GSA kiibiga, seega tõin kinnitunud deletsioonide genotüübid välja tabelis 7 kõigi kasutatud platvormide kohta eraldi. Erinevate platvormidega samade indiviidide analüüsimisel saadud tulemused langesid omavahel täielikult kokku.

Tabel 7. Katseliselt valideeritud deletsioonide genotüübid ülekantuna kõigi indiviidide genotüübi andmetele. Eraldi on välja toodud erinevate andmekogude indiviidide arv, kellel esineb kindel genotüüp. GSA_1 on esimesena analüüsitud GSA platvormi andmekogu (n = 30 100) ja GSA_2 on hiljem analüüsitud GSA platvormi andmekogu (n = 104 920).

Valideeritud deletsiooni genotüüp	WGS (n = 2244)	GSA_1 (n = 30 100)	Omni (n = 5929)	GSA_2 (n = 104 920)
Pika deletsiooni heterosügoot	35	443	108	1577
Lühikese deletsiooni heterosügoot	2	10	4	61
Pika deletsiooni homosügoot	-	1	-	10
Mõlema deletsiooniga	-	1	2	-
Ei kinnitunud	-	4	2	29
Deletsiooniga indiviide kokku	37 (1,6%)	459 (1,5%)	116 (1,9%)	1677 (1,6%)

2.3.3. Leitud *CYP2C19* deletsioonide esinemissagedused Eesti populatsioonis

Kuna andmehulkade vaheliselt esineb korduvaid indiviide, siis arvutan deletsiooniga indiviidide esinemissagedused eri platvormide andmehulkade kohta eraldi. WGS andmetest (n = 2244) leitud *CYP2C19* geeni deletsiooniga indiviidide (n = 37) esinemissagedus Eesti populatsioonis on 1,65%. Omni kiibi andmetest (n = 5929) leitud *CYP2C19* geeni deletsiooniga indiviidide (n = 116) esinemissagedus Eesti populatsioonis on 1,96%. GSA kiibi andmetest (n = 30 100) leitud *CYP2C19* geeni deletsiooniga indiviidide (n = 459) esinemissagedus Eesti populatsioonis on 1,52%. Hilisematest lisa GSA kiibi andmetest (n = 104 920) leitud *CYP2C19* geeni deletsiooniga indiviidide (n = 1677) esinemissagedus Eesti populatsioonis on 1,6%. Kõigi andmekogude peale kokku on deletsiooniga indiviidide esinemissageduste keskmine 1,68%.

Genotüüpide ja alleelisageduste arvutamisel jätsin välja WGS ja Omni platvormidega saadud andmed, et mitte üle hinnata deletsiooni kandjate esinemist Eesti populatsioonis. Kokku analüüsi GSA kiibiga 135 020 indiviidi, kelle seast 2020-1 esines pikk deletsioon heterosügootsel kujul. Lühikese deletsiooni heterosügootsus esines 71 indiviidil, pika deletsiooni homosügootsus esines 11 indiviidil ning mõlema deletsiooni kandjaid oli ainult 1. Deletsiooniga indiviidide esinemissagedused Eesti populatsioonis on välja toodud tabelis 8. Eelnevate andmete põhjal saan öelda, et pika deletsiooni alleelisagedus on 0,756% ja lühikese deletsiooni alleelisagedus on 0,026% Eesti populatsioonis.

Tabel 8. Deletsiooniga indiviidide esinemissagedused Eesti populatsioonis (n = 135 020).

CYP2C19 deletsiooni genotüüp	Kandjate esinemissagedus populatsioonis
Pika deletsiooni heterosügoot	1,49%
Lühikese deletsiooni heterosügoot	0,052%
Pika deletsiooni homosügoot	0,008%
Mõlema deletsiooni suhtes heterosügoot	0,0007%

2.3.4. *CYP2C19* diplotüüpide määramine WGS andmetest leitud deletsiooniga indiviididel

WGS SNV andmeid kasutades määrasin *CYP2C19* geeni deletsiooniga indiviidide diplotüüpe, kasutades selleks teadaolevaid star-alleelidega seostatud SNP-e. Minu tuvastatud ja katseliselt kontrollitud deletsioonid langevad star-alleeli *CYP2C19**37 alla, kuna mõlema deletsiooni puhul on tegemist osalise *CYP2C19* geeni deletsiooniga, mis on star-alleeli *37 definitsioon. Seega iga töös uuritud deletsiooniga indiviidi üks haplotüüpidest on star-alleel *37. Teise haplotüübi määramiseks vaatasin, mis star-alleelide SNP-id esinesid igal deletsiooniga indiviidil, ning leidsin, et pika deletsiooniga indiviididel (n = 35) esines kokku kolm erinevat diplotüüpi (*1/*37, *2/*37 ja *17/*37) ning lühikese deletsiooniga indiviididel (n = 2) esines kummalgi erinev diplotüüp (*1/*37 ja *2/*37). Kui palju indiviide omas kindlat diplotüüpi on toodud välja tabelis 9.

Tabel 9. *CYP2C19* deletsiooniga indiviidide diplotüübid vastavalt star-alleelide nomenklatuurile. Esimeses tulbas on välja toodud kõik määratud diplotüübid. Teises ja kolmandas tulbas on vastavalt arvud, mitmel indiviidil leiti kindlat diplotüüpi.

Deletsiooni haplotüübid	Pika deletsiooniga indiviidide arv	Lühikese deletsiooniga indiviidide arv
*1/*37	24	1
*2/*37	4	1
*17/*37	7	0

2.3.5. Deletsioonidega potentsiaalselt aheldunud SNP-id

Tuvastasin 10. kromosoomist *CYP2C19* geeni pika ja lühikese deletsiooniga potentsiaalselt aheldunud SNP-e WGS andmetest. Kuna WGS andmetest leitud deletsioonide kandjad on kõik deletsioonide suhtes heterosügootid, siis esineb leitud aheldunud SNP-e kogu WGS andmekogu peale ainult heterosügootselt deletsiooniga indiviididel, misjuures deletsioonita indiviididel neid SNP-i ei leidunud.

Pika deletsiooniga potentsiaalselt aheldunud SNP-e leidsin kokku kolm, mis esinesid ainult kõigil pika deletsiooniga indiviididel (n = 35) heterosügootselt. Nendest kandidaat SNP-idest üks asus eespool ning kaks tagapool pika deletsiooni ala. Esimesel leitud SNP-il puudub rs number, see leiti positsioonist (GRCh37) 10: 96 151 859 ning see asub umbes 345,5 kbp eespool pika deletsiooni alguspunktist. See SNP on A nukleotiidi asendus G nukleotiidiga ning genoomis paikneb see geeni TBC1 domeeni perekonna 12 liikme transkriptsioonifaktori seondumispiirkonnas. Teisena leitud SNP on rs190738878, see asub positsioonis (GRCh37) 10: 96 635 159, mis on umbes 76 kbp tagapool pika deletsiooni lõpp-punktist. Tegemist on T nukleotiidi asendusega C nukleotiidi poolt ning genoomis paikneb see mittekodeerival alal. Kolmandal leitud SNP-l puudub rs number ning see asub positsioonis (GRCh37) 10: 96 659 041, mis on umbes 99,9 kbp tagapool pika deletsiooni lõpp-punktist. Antud juhul on C nukleotiid asendunud T nukleotiidiga ning genoomis paikneb see mittekodeerivas alas.

WGS andmetes esineb ainult kaks indiviidi, kellel on *CYP2C19* geeni lühike deletsioon, seega nende põhjal leitud 19 potentsiaalselt lühikese deletsiooniga aheldunud SNP-i ei pruugi kõik sobida genotüübi andmetest deletsiooniga indiviidide määramiseks. Kõik lühikese deletsiooniga aheldunud kandidaat SNP-id on toodud välja tabelis 10, misjuures ainult kahel neist on olemas rs number, ülejäänud nimetasin järjest SNP1, SNP2 ja nii edasi.

Tabel 10. Lühikese deletsiooniga potentsiaalselt aheldunud SNP-id, nende positsioonid (Pos.) 10. kromosoomil (GRCh37), referents (Ref nt) ja alternatiivne nukleotiid (Alt nt.), kaugus lühikese deletsiooni alguspunktist ettepoole või deletsiooni lõpp-punktist tahapoole ning paiknemine genoomis. CTCF on CCCTC-seondumise faktor (tsink sõrme valk).

SNP-i nimi (rs number)	Pos. (GRCh37)	Ref nt.	Alt nt.	Kaugus lühikesest deletsioonist (bp)	Paiknemine genoomis
SNP1	90 375 792	G	C	6,15 Mbp eespool	Mittekodeerivas alas
SNP2	91 313 677	G	A	5,22 Mbp eespool	SLC16A12 geeni intronis
SNP3	94 656 711	C	A	1,87 Mbp eespool	EXOC6 geeni intronis
SNP4	95 952 352	G	C	581,5 kbp eespool	PLCE1 geeni intronis
rs182908376	96 099 729	G	A	434,1 kbp eespool	NOC3L geeni intronis
SNP5	96 125 854	G	A	408 kbp eespool	Mittekodeerivas alas
SNP6	96 480 758	A	G	53,1 kbp eespool	CYP2C18 geeni intronis
SNP7	96 621 007	A	C	78,2 kbp tagapool	Mittekodeerivas alas
SNP8	96 708 600	A	G	165,8 kbp tagapool	CYP2C9 intronis
SNP9	96 877 496	G	C	334,7 kbp tagapool	Mittekodeerivas alas
SNP10	96 880 972	C	T	338,2 kbp tagapool	Mittekodeerivas alas
SNP11	96 883 248	A	C	340,5 kbp tagapool	Mittekodeerivas alas
SNP12	97 128 384	G	C	585,6 kbp tagapool	SORBS1 geeni intronis
SNP13	97 192 446	C	T	649,7 kbp tagapool	SORBS1 geeni intronis
SNP14	97 292 067	C	T	749,3 kbp tagapool	SORBS1 geeni intronis
SNP15	97 360 141	G	C	817,4 kbp tagapool	CTCF seondumispiirkonnas
rs144543830	97 442 435	C	T	899,6 kbp tagapool	TCTN3 geeni eksonis 12
SNP16	97 482 730	A	G	939,9 kbp tagapool	ENTPD1 geeni intronis
SNP17	97 833 359	G	T	1,29 Mbp tagapool	ENTPD1 geeni intronis

Lühikese deletsiooni võimalike aheldunud SNP-ide esimene leitud rs numbriga kandidaat SNP oli nimetusega rs182908376, mis on NOC3L geeni intronis asuv variant ja see paikneb 434,1 kbp eespool lühikese deletsiooni alguspunktist. Teiseks rs numbriga kandidaat SNP-ks oli rs144543830, mis on TCTN3 geeni sünonüümne variant ja see paikneb 899,6 kbp lühikese deletsiooni lõpp-punktist tagapool. Nende kahe leitud SNP-i kohta tegin otsingu NCBI SNP-ide andmebaasist (dbSNP, 2021) ning leidsin, et nende SNP-ide raporteeritud esinemissagedused olid samade andmekogude kohta erinevad. Näiteks leidis TOPMed geenivariatsiooni andmetes (n = 264 690) rs182908376 SNP-i 114 indiviidil, kuid samast andmekogust leidis rs144543830 SNP-i 130 indiviidil. Leitud aheldunud kandidaat SNP-ide põhised on pika deletsiooni haplotüübi ulatus vähemalt 0,5 Mbp ning lühikese deletsiooni haplotüübi ulatus vähemalt 7,4 Mbp.

Ma kasutasin neid potentsiaalselt *CYP2C19* geeni deletsioonidega aheldunud SNP-e deletsioonide kandjate viimase ühise esivanema vanuse leidmiseks, Human Genome Dating veebilehekülge kasutades. Pika deletsiooniga aheldunud SNP-i (rs190738878) kandjate viimase ühise esivanema vanus on vahemikus 105 – 185 põlvkonda ning kui võtta generatsiooniajaks 29 aastat, teeb see vanuseks ligikaudu 3000 – 5300 aastat. Lühikese deletsiooniga potentsiaalselt aheldunud SNP-i rs144543830 kohta vastet ei leidnud. SNP-i rs182908376 kandjate viimase ühise esivanema vanus on vahemikus 76 – 641 põlvkonda ehk umbes 2100 – 18 592 aastat.

2.4. Arutelu

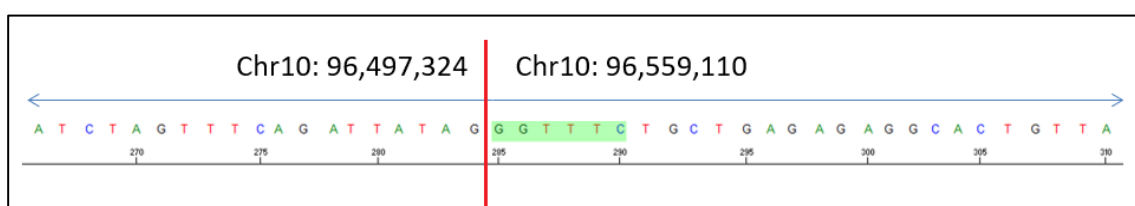
Käesoleva töö esimeseks eesmärgiks oli määrata *CYP2C19* geeni deletsioonide esinemissagedused Eesti populatsioonis TÜ EGV andmete põhjal. Selle jaoks pidin esmalt määrama ja valideerima *CYP2C19* geeni deletsioonide katkekohad ning seejärel leidma valimist ja katseliselt kinnitama *CYP2C19* geeni deletsioonide kandjad.

Töös kasutasin TÜ EGV geenidonorite andmeid, mis olid saadud järgmise põlvkonna sekveneerimise tehnoloogiaga (WGS) ning kahte erinevat mikrokiipi (Omni ja GSA) kasutades. WGS andmetest (n = 2244) leitud 37 deletsiooniga indiviidide sekveneeritud genomijärjestusi kasutasin tuvastamiseks deletsioonide katkekohti, et määrata nende deletsioonide algus- ja lõpp-punkte ning deletsioonide ulatust. Visuaalse vaatluse tulemusena leidsin pika deletsiooni katkekohtade vahetust lähedusest mikrohomoloogia (järjestus GGTTC), mis võib olla selle deletsiooni üheks tekkepõhjuseks. Tuvastatud mikrohomoloogiat arvesse võttes koostas deletsioonide üleminekujärjestuste põhised 25-mer-id, et kindlaks määrata deletsioonide täpsed katkekohad. Loodud pika deletsiooni 25-mer oli CAGATTATAGGGTTTCTGCTGAGAG ja lühikese deletsiooni 25-mer oli AAGCCAGATGGGATAACATTAAGA. Töö käigus leidis kinnitust, et loodud 25-mer-id on õiged deletsiooni üleminekujärjestused ning neid on võimalik edaspidi kasutada töös leitud deletsioonide kandjate otsimiseks teistest andmestikest.

Selgus, et WGS-i abil sekveneeritud indiviidide hulgas esines kaks erinevat *CYP2C19* geeni deletsiooni, mida töös nimetasin pikaks ja lühikeseks deletsiooniks. Pika deletsiooni kandjatel on deleteerunud referents genoomi GRCh37 piirkond 10: 96 497 330 – 96 559 116 ja lühikese deletsiooni kandjatel 10: 96 533 900 – 96 542 738, mis teeb deletsioonide pikkuseks vastavalt 61,7 kbp ja 8,8 kbp. Oluline on vanema ja uuema referents genoomi versioonide (vastavalt GRCh37 ja GRCh38) kasutamisel silmas pidada, et kui GRCh37 kohaselt asub *CYP2C19* geen kromosoomil 10 koordinaatidel 96 447 911 – 96 613 017, mis teeb geeni pikkuseks ligikaudu 165 kbp, siis GRCh38 kohaselt asub *CYP2C19* geen koordinaatidel 10: 94 762 681 – 94 855 547, mis teeb geeni pikkuseks ligikaudu 92 kbp. Geeni pikkuse erinevus tuleneb sellest, et vanemas versioonis on *CYP2C19* geenil ligikaudu 48 kbp ülekattuvus *CYP2C18* geeniga, mis on üksteisest eraldatud uuemas referents genoomi versioonis; seega *CYP2C19* uurimisel on õigem kasutada referentsina GRCh38-t. Kahjuks enda töös pidin kasutama mõlemat versiooni, mis tulenes sellest, kumma referents genoomi põhiselt andmeid, millele mulle ligipääs anti, eelnevalt analüüsitud oli. GRCh38 referents genoomi kohaselt asub pikk deletsioon

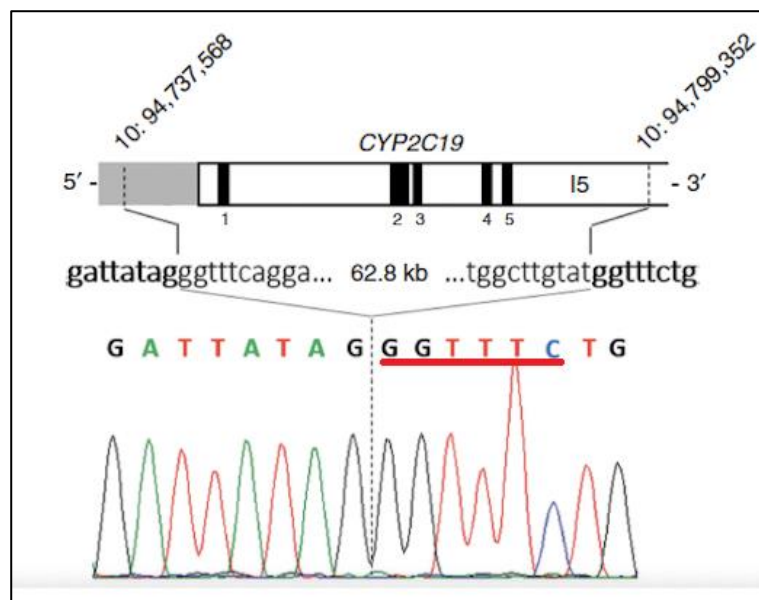
koordinaatidel 10: 94 737 573 – 94 799 359 ning lühike deletsioon koordinaatidel 10: 94 774 143 – 94 782 981.

Tervasmäki jt. (2014) artiklis räägitakse *CYP2C19* geeni deletsioonist positsioonis 10: 96 497 324 – 96 559 110 (GRCh37), mida leiti kokku 48 indiviidil heterosügootselt põhja-soomlaste populatsioonis (n = 1339). Nende koordinaatide põhjal arvasin, et nende leitud deletsioon on sama, mis minu poolt kinnitatud pikk deletsioon. Kirjeldatud deletsiooni üleminekujärjestus (joonis 10) langeb tõepoolest kokku minu poolt leitud pika deletsiooni üleminekujärjestusega (CAGATTATAGGGTTTCTGCTGAGAG), kuid deletsioonide algus- ja lõpp-punktid on erinevad piirkonnas esineva mikrohomoloogia tõttu.



Joonis 10. Tervasmäki jt. leitud *CYP2C19* geeni deletsiooni üleminekujärjestus. Deletsiooni katkekoht on märgitud punase vertikaalse joonega. Pika deletsiooni katkekohtadest leitud mikrohomoloogia järjestus on esiletõstetud rohelisega. (Kohandatud, Tervasmäki jt., 2014)

Santos jt. (2018) kirjeldavad samuti soomlaste populatsioonis (n = 463) leitud uusi *CYP2C19* geeni deletsioonide. Minu poolt leitud pika deletsiooni koordinaatidele kõige lähedasem on nende leitud deletsioon positsioonis 10: 94 737 568 – 94 799 352 (GRCh38). Selle deletsiooni üleminekujärjestus (joonis 11) langeb kokku minu leitud pika deletsiooni üleminekujärjestusega, kuid deletsioonide algus- ja lõpp-punktid on meil jälle määratud erinevalt. Tegemist on kindlasti sama *CYP2C19* deletsiooniga ehk minu poolt kinnitatud pika deletsiooniga ning edaspidise segaduse vältimiseks on oluline kokku leppida ametlikud deletsiooni koordinaadid. Santos jt. poolt leitud deletsiooni esines neljal indiviidil heterosügootselt. Artiklis raporteeritud alleelisagedus soomlaste populatsioonis (n = 463) oli MAF = 0,43%.



Joonis 11. Genoomse referents järjestuse (GRCh38) ja *CYP2C19* geeni deletsiooni lineaarne esitus. Pika deletsiooni katkekohtadest leitud mikrohomologia järjestus on alla joonitud punasega. (Kohandatud, Santos jt., 2014).

Minu poolt leitud lühikese deletsiooni koordinaatidele kõige lähedasem on Santos jt. artikli autorite poolt leitud deletsioon positsioonis 10: 94 775 058 – 94 781 997 (GRCh38), kuid antud andmete kohaselt ei saa kindlalt öelda, et tegu on sama deletsiooniga (Santos jt., 2018). Kummagi deletsiooni puhul pole *CYP2C19* ekson 1 deleteerunud, seega PharmVar andmebaasis välja toodud lause, et iga *CYP2C19* geeni deletsiooni puhul on puudu vähemalt ekson 1 (PharmVar: Structural..., 2019), on tegelikult vale.

Omni (n = 5929) ja GSA (n = 135 020) mikrokiipide genotüpiseerimise andmetest leiti kokku 2252 *CYP2C19* deletsiooniga indiviidi. Kuna mikrokiipide marker SNP-id ei kata kogu genoomi ühtlaselt ning kuna kiibil puuduvad SNP-id, mis oleksid leitud deletsioonide algus- ja lõpp-punktide vahetust lähedusest, olid sel viisil leitud deletsioonide ulatused ebatäpsed ning kohati määras PennCNV tarkvara ühe pikema CNV hoopis kaheks lühemaks CNV-ks. Mikrokiibilt deletsiooni täpse ulatuse tuvastamiseks on üheks võimaluseks luua ja lisada kiibile deletsiooni spetsiifiline järjestus (näiteks töös loodud 25-mer-id), mis deletsiooni olemasolul annaks signaali, deletsiooni puudumisel aga mitte. Teostasin ka deletsiooni spetsiifiliste 25 aluspaari järjestuste ning vastavate referents järjestuste otsingu kogu genoomist referents genoomi GRCh37 põhised Perli skriptiga. Neid järjestusi ei leidnud mujal genoomis, mis tähendab, et need järjestused on unikaalsed ja sobivad kiibilt deletsiooni tuvastamiseks.

Jaotasin leitud deletsioonidega invidiidid nende CNV-de ennustatud algus- ja lõppkoordinaatide põhjal gruppidesse. Eeldasin, et samade ennustatud CNV koordinaatidega invidiididel esineb sama deletsioon ning võtsin igast grupist vähemalt kolm invidiidi katseliseks deletsiooni valideerimiseks. Kui grupp koosnes kolmest või vähemast invidiidist, siis võtsin kõik. Gruppidesse jaotamine ja ainult osade invidiidide katseline valideerimine igast grupist oli vajalik, kuna deletsiooniga invidiide oli liiga palju, et kõiki katseliselt valideerida. Selleks, et 100% kindlusega öelda, et igale invidiidile määratud deletsiooni genotüüp on korrektne, tuleks kõiki invidiide katseliselt kontrollida, mis aga pole ressurside ja ajakulu tõttu mõistlik. Minu töö suuremaks eesmärgiks oli teha *CYP2C19* geeni uuritud deletsioonide tuvastamine tulevikus lihtsamaks.

Deletsioonide katkekohtade positsioonidest lähtuvalt disainisin alleelispetsiifilised PCR praimerid, et katseliselt valideerida deletsioonide olemasolu. Praimerite disainimisel pidasin silmas ka tekkivate PCR produktide pikkuseid, mis pidid olema üksteisest piisavalt erinevad, et nende järgi hiljem geelipildilt metsiktüüpi ja deletsiooniga alleelide kinnitamine oleks visuaalselt teostatav. Lühikese deletsiooni valideerimiseks disainitud kõigi kolme praimeri seondumiskohad jäävad pika deletsiooni ala sisse. Seega pika deletsiooni homosügootidel ei ole geelipildil näha lühikese deletsiooni praaimeritega saadavaid PCR produkte. Kui invidiidil esinesid mõlemad deletsioonid, oli pika deletsiooni praaimeritega testimisel näha nii metsiktüüpi kui ka deletsiooni alleeli vööti ning lühikese deletsiooni praaimeritega testimisel oli näha ainult deletsiooni alleeli vööti. Seda oli oluline meeles pidada invidiidide genotüüpide kinnitamisel.

Katseliselt kontrollisin kõigi kasutatud platvormide (WGS, Omni ja GSA) gruppide peale 147 invidiidi järjestusi, sealhulgas kontrollisin kõiki invidiide, kelle puhul määratud geeni koopiaarv oli 0. Invidiidide seas, kellel kumbki minu uuritud deletsioon ei kinnitunud, oli 15 invidiidi duplikatsioonidega ($cn = 3$) ning ülejäänud 30 olid heterosügootsed deletsiooni kandjad ($cn = 1$). Kõigi mitte kinnitunud invidiidide puhul oli geelilt näha ainult metsiktüüpi alleeli olemasolu kinnitav PCR produkt (joonis 8). Deletsiooniga invidiidide puhul, kellel kumbki uuritud deletsioonidest ei kinnitunud, jäid nende ennustatud deletsioonide koordinaadid täielikult või osaliselt minu töös uuritud deletsioonide piirkondadest välja. Nende invidiidide puhul võib tegu olla mingi teise *CYP2C19* geeni deletsiooniga. 102-st invidiidist, kellel deletsioon kinnitus, olid 71 pika deletsiooni heterosügootid, 17 lühikese deletsiooni heterosügootid, 11 pika deletsiooni homosügootid ja 3 olid mõlemad deletsiooni kandjad. Samasse gruppi kuulunud invidiidide deletsioonide genotüübid kinnitusid kõigil samadena, seega määrasin ka samasse gruppi kuulunud katseliselt mitte kontrollitud invidiidide

genotüübiks sama, mis sain PCR-iga valideerimise tulemustest. Ühegi indiviidi puhul ei jäänud nende genotüüp määramata.

Peale katseliselt valideeritud *CYP2C19* geeni deletsioonide genotüüpide üle kandmist kogu valimile, jätsin edasisest andmeanalüüsist duplikatsioonidega indiviidid välja. Terve valimi ($n = 143\ 193$) peale kokku leidis 35 indiviidi, kellel minu uuritavad deletsioonid ei kinnitunud ning 2254 (1,57%), kellel kinnitus. Kokku leidis 2163 indiviidi pika deletsiooni heterosügootsete kandjatena, 77 indiviidi lühikese deletsiooni heterosügootsete kandjatena, 11 indiviidi, kes olid pika deletsiooni suhtes homosügootid ning 3 indiviidi, kellel esineb pikk deletsioon ühel alleelil ning lühike deletsioon teisel. Osasid indiviide, keda analüüsiti WGS ja Omni platvormidega, analüüsiti hiljem uuesti ka GSA kiibiga. Kõikide kasutatud platvormide peale kokku esineb deletsiooniga indiviide tegelikkuses seega vähem kui 2254. Indiviidide uuesti analüüsimine GSA kiibiga oli vajalik, et kogu TÜ EGV andmestik oleks analüüsitud sama kiibiplatvormiga ja markerite arvuga ning seega lihtsamini omavahel võrreldav. Erinevate platvormidega samade indiviidide analüüsimisel saadud tulemused langesid omavahel täielikult kokku. *CYP2C19* deletsioonide katselise valideerimise kokkuvõtteks võib öelda, et kiibi andmetest leitud deletsioonide algus- ja lõppkoordinaatide järgi on võimalik usaldusväärselt kinnitada uuritud deletsioonide olemasolu kontrollitud gruppide põhjal.

Olles leidnud ja katseliselt kinnitanud *CYP2C19* geeni deletsioonide kandjad TÜ EGV valimis, sain arvutada esinemissagedusi. Kuna kohati analüüsiti samu indiviide erinevate platvormidega mitmekordselt, arvutasin deletsiooniga indiviidide esinemissagedused eri platvormide andmehulkade kohta eraldi (tabel 7). Siinkohal pean kõige õigemaks lähtuda alleelisageduste võrdlemisel ainult GSA kiibi andmete põhisel saadud numbritest, kuna see oli kõige suurem andmehulk, mida ma kasutasin. GSA kiibiga analüüsiti kokku 135 020 indiviidi, kelle hulgast leiti 2020 pika deletsiooni heterosügooti, 71 lühikese deletsiooni heterosügooti, 11 pika deletsiooni homosügooti ja 1 mõlema deletsiooni kandja. Nii Tervasmäki jt., kui ka Santos jt. leidsid soomlaste populatsioonist ainult pika deletsiooni heterosügootte, seega saan eestlaste ja soomlaste populatsioonide vaheliselt võrrelda ainult pika deletsiooni esinemissagedusi. Pika deletsiooni heterosügootide esinemissagedus Eesti populatsioonis on 1,49% ning pika deletsiooni alleelisagedus on $MAF = 0,756\%$. Santos jt. raporteerisid pika deletsiooni alleelisageduseks $MAF = 0,43\%$ (4/463), mis teeb deletsiooniga indiviidide esinemissageduseks 0,86%. Tervasmäki jt. artiklis toodud andmete põhjal on nende leitud pika deletsiooni kandjate esinemissageduseks 3,58%, mis teeb alleelisageduseks $MAF = 1,79\%$ (48/2678). Santos jt. uurisid soomlaste populatsiooni üldisemalt ja nad pole välja toonud

uuritud indiviidide geograafilist pärinemist. Tervasmäki jt. tõid välja, et uuritud indiviidid on Põhja-Soomest ning et kõik uuritud indiviidid olid naised, kuna nad uurisid rinnavähiga patsiente. Ilmselgelt on nendes kahes artiklis leitud pika deletsiooni alleelisagedused väga erinevad, mis tõstatab küsimuse, kas Tervasmäki jt. valimi puhul on tegemist kallutatud andmetega, kuna valimisse on kaasatud lähisugulased või on muud moodi tegu mittejuhusliku valimiga. Võimalik on ka, et antud deletsiooni esineb põhja-soomlastes rohkem kui mujal Soomes ja Eestis ning tegu on geograafilise iseärasusega. Või et antud deletsiooni esineb rohkem naiste seas, kuna Tervasmäki jt. uurisid ainult naisi ja teise kahe valimi puhul on mehi ka kaasatud, mis deletsiooni esinemissagedust sel juhul lahjendaks. Igal juhul on võrreldud soomlaste valimid oluliselt väiksemad kui TÜ EGV valim ning selgema pildi saamiseks peaks rohkemate soomlaste andmeid analüüsima.

Töö teiseks eesmärgiks oli määrata WGS andmetest leitud *CYP2C19* geeni deletsiooni kandjate diplotüüpe. Selleks kasutasin PharmVar ja LOVD andmebaasides olevaid *CYP2C19* geeni star-alleelide põhiseid SNP-e. Järgneva analüüsi puhul jätsin välja SNP-id, mille esinemissagedus oli eurooplaste populatsioonis 0 ning SNP-id, mille seotus kindla star-alleeliga polnud lõplikult tõendatud. Kõigil WGS andmete põhjal leitud *CYP2C19* deletsiooniga indiviididel on üks alleelidest star-alleel *37, mis tähistab osalist *2C19* geeni deletsiooni. Pika deletsiooni alguspunkt on umbes 30 kbp eespool *CYP2C19* geeni algusest, mis tähendab, et deleteerunud on ka geeni ees paiknevad regulatoorsed elemendid ning valguprodukti sellelt alleelilt ei toodeta. Lühikese deletsiooni alguspunkt asub *CYP2C19* geeni esimeses intronis, mis tähendab, et valguprodukti sellelt alleelilt võidakse toota, kuid funktsionaalne valk see ei ole. Mitte deleteerunud alleeli haplotüübi määramiseks vaatasin, mis nukleotiid asus igal deletsiooniga indiviidil kindlas star-alleeli SNP positsioonis. WGS andmetest leitud deletsiooniga indiviididel esines kokku kolm erinevat diplotüüpi, milleks olid *1/*37, *2/*37 ja *17/*37 (tabel 9). Määratud diplotüüpide põhjal saan hinnata neid kandvate indiviidide eeldatavat ensüümi funktsionaalsust. Diplotüübiga *1/*37 indiviididel esineb deleteerunud geenialleeli kõrval üks normaalse funktsionaalsusega geeni alleel (*1). Diplotüübiga *2/*37 indiviidide puhul esineb deleteerunud alleeli kõrval üks alanenud ensüümi aktiivsusega valguprodukti kodeeriv alleel (*2). Diplotüübiga *17/*37 indiviididel esineb deletsiooniga alleeli kõrval üks suurenenud ensüümi aktiivsusega valguprodukti kodeeriv alleel (*17). Kui indiviid on selle geeni suhtes heterosügoot, omades üht null alleeli ning üht normaalse funktsionaalsusega alleeli (*CYP2C19**1/*37), siis selle indiviidi summaarne valgu aktiivsus on arvatavasti madalam kui homosügootselt normaalse funktsionaalsusega alleeli kandjal (*CYP2C19**1/*1). *2 null alleel toodab valguprodukti, millel on madalam katalüütiline aktiivsus kui metsiktüüpi valkudel,

seega diplotüübiga *CYP2C19**2/*37 indiviididel on 2C19 ensüümi aktiivsus väga madal. Kui aga indiviid omab null alleeli kõrval suurenenud funktsionaalsusega alleeli (*CYP2C19**17/*37) võiks arvata, et summaarselt on indiviidil normaalne ensüümi aktiivsus. Rudberg jt. (2008) leidsid, et kui indiviidil on üks *CYP2C19* null alleel (nagu näiteks *37) ja teine suurenenud funktsionaalsusega alleel (*17), siis summaarselt valgu aktiivsus on sarnane metsiktüüpi indiviidi valgu aktiivsusele või isegi natukene madalam. Vähenenud ensüümi aktiivsus tähendab, et indiviididel on vähenenud võime lagundada 2C19 ensüümi substraatideks olevaid ravimeid (nagu näiteks mefenütoin ja omeprasool). Kiibi andmete põhjal leitud pika deletsiooni homosügootidel ja mõlema deletsiooni kandjatel puudub 2C19 ensüümi aktiivsus täielikult.

Referents alleeli (*1) määratakse indiviidile genotüpiseerimisel sõltuvalt sellest, millised SNV-d mikrokiibil on ehk milliseid geenivariante analüüsitakse ja sellest tulenevalt ka välistatakse, sest geenivariante, mida ei testita, ei saa ka määrata. Näiteks *1 indiviidil võib siiski esineda geenivariante, mis muudaks tema star-alleeli määratlust (diplotüüpi), kuid mida ei testitud kindlat mikrokiip platvormi kasutades. Samuti on võimalik, et mingi SNV ja star-alleeli vahelisi seoseid ei ole veel leitud, seega keegi ei teagi testida neid SNV-sid suurema populatsiooni genotüpiseerimisel. Teisisõnu, kui indiviidile on kiibiga genotüpiseerimisel määratud *1 haplotüüp, ei välista see võimalust, et tal võib tegelikult esineda mõni teine, seni tuvastamata või antud platvormiga testimata star-alleel, mis põhjustab metsiktüübist erinevat fenotüüpi. Tõenäosus, et *1 ehk referents alleeli vaikimisi määramine kiibi andmete põhjal on korrektne, suureneb, mida rohkem erinevaid teadaolevaid geenivariatsioone üleüldse testitakse.

Paljusid ravimeid, mida lagundab P450 2C19, võivad ka teised P450 ensüümid lagundada, mis tõstatab küsimuse, kui organism 2C19 ensüümi ei tooda või kui tegemist on keskmise või aeglase metaboliseerija fenotüübiga, kas siis tegelikult teised P450 ensüümid on võimelised asendama puudulikku 2C19 katalüütilist võimekust. Teoreetiliselt suudavad teised CYP2C alam perekonna ensüümid lagundada samu ravimeid kui 2C19, kuid siinkohal on erinevate CYP2C ensüümide efektiivsused erinevad (Niwa ja Yamazaki, 2012). Seega üks CYP2C alam perekonna liige võib teist asendada, kuid kindla ravimi lagundamise kiirus on neil erinev, kuna neil on üldiselt erinevad peamised substraadid ning seetõttu on *CYP2C19* defektse alleeliga indiviididel ikkagi väiksem ravimi lagundamise võimekus kui on metsiktüüpi indiviidil. See võib aktiivravimite puhul põhjustada ravimi kuhjumist kehas ja soovimatuid kõrvaltoimeid ning eelravimite puhul võib ravimi aktiivse metaboliidi hulk kehas olla mittepiisav, et saavutada soovitud terapeutilist vastust.

Töö kolmandaks eesmärgiks oli leida *CYP2C19* geeni deletsioonidega potentsiaalselt aheldunud SNP-e WGS andmete põhjal, mida saaks edaspidi kasutada deletsiooniga indiviidide genotüpiseerimiseks GSA mikrokiip platvormi kasutades ning ka deletsiooniga indiviidide leidmiseks teiste populatsioonide genotüübi andmestikest. Pika deletsiooniga potentsiaalselt aheldunud SNP-e leidsin kokku kolm ja ainult ühel neist oli rs number olemas (rs190738878), mida sain kasutada järgnevate otsingute tegemiseks. Lühikese deletsiooniga potentsiaalselt aheldunud kandidaat SNP-e leidsin kokku 19, kuid kuna valimis oli ainult 2 indiviidi, on suur tõenäosus, et osa nendest leitud SNP-idest on neil kahel indiviidil juhuslikult samad ning pole tegelikkuses aheldunud *CYP2C19* lühikese deletsiooniga. Leitud 19 kandidaat SNP-i seas oli ainult kahel rs number olemas (rs182908376 ja rs144543830), mida sain kasutada järgnevate otsingute tegemiseks. Leitud potentsiaalselt aheldunud SNP-ide põhjal järeldasin pika deletsiooni haplotüübi ulatuseks vähemalt 0,5 Mbp ja lühikese deletsiooni haplotüübi ulatuseks vähemalt 7,4 Mbp.

Tegin leitud lühikese deletsiooni kandidaat märgend-SNP-ide (*tag SNP*) kohta, millel oli rs number olemas, otsingu NCBI SNP andmebaasist ning leidsin, et nende SNP-ide raporteeritud esinemissagedused sama andmekogu kohta olid erinevad. Kui mõlemad SNP-id oleksid *CYP2C19* geeni lühikese deletsiooniga aheldunud, tähendaks see, et nad on ka omavahel aheldunud. See omakorda võiks tähendada, et neid mõlemaid SNP-e esineb samadel indiviididel ühes uuritud valimis, mistõttu võiks eeldada, et aheldunud SNP-ide esinemissagedus selles valimis on sama. Kuid kuna nende kahe SNP-i (rs182908376 ja rs144543830) raporteeritud sagedused sama valimi hulgas olid erinevad, siis võib öelda, et üks või mõlemad neist SNP-idest ei ole tegelikult aheldunud lühikese deletsiooniga ja seega ei saa neid reaalselt kasutada lühikese deletsiooni usaldusväärseks genotüübi määramiseks kiibi andmete põhjal. See seab ka kahtluse alla kindlasti need kandidaat SNP-id, mis asuvad veel kaugemal deletsiooni alast, kuna eeldan lookuste aheldatuse tasakaalutuse (*linkage disequilibrium*, LD) teooria kohaselt, et deletsioonile lähemal asuvad SNP-id on suurema tõenäosusega deletsiooniga aheldunud. Nagu eelpool mainitud, oli lühikese deletsiooniga aheldunud märgend-SNP-ide leidmiseks valimis ainult 2 indiviidi, kuid korrektsete lühikese deletsiooni spetsiifiliste märgend-SNP-ide leidmiseks on vaja rohkemate indiviidide andmeid ja seega leitud 19 SNP-i seast pole kindlasti kõik tegelikult lühikese deletsiooniga aheldunud. Seega ka lühikese deletsiooni haplotüübi ulatus pole ilmselt 7,4 Mbp, vaid vähem.

Leitud aheldunud kandidaat SNP-e kasutasin tuvastamiseks deletsioonide kandjate viimase ühise esivanema vanust põlvkondades. Pika deletsiooniga potentsiaalselt aheldunud SNP-i

(rs190738878) kandjate viimase ühise esivanema vanus on vahemikus 105 – 185 põlvkonda (3000 – 5300 aastat). Kuna see deletsioon on väheesinev ja leitud kindlalt seni ainult Soome ja Eesti populatsioonidest, on tõenäoline, et selle SNP-i suhteliselt noor vanus võib olla ka deletsiooni vanuseks, mis võiks tähendada seda, et see SNP on deletsiooniga samal ajal genoomis tekkinud ja on sobilik pika deletsiooni märgend-SNP-ina. Lühikese deletsiooniga potentsiaalselt aheldunud SNP-i (rs182908376) kandjate viimase ühise esivanema vanus oli ennustatud vahemikku 76 – 641 põlvkonda (2100 – 18 592 aastat). Selle põhjal võib öelda, et antud SNP pole sobilik lühikese deletsiooni märgend-SNP-iks.

Ühtegi *CYP2C19* deletsioonidega aheldunud kandidaat SNP-i ei esinenud töös kasutatud GSA kiibi SNP-ide hulgas, mistõttu ei saa neid SNP-e kasutada eelnevalt analüüsitud GSA andmete hulgast deletsiooniga indiviidide kinnitamiseks. Võib arvata, et kuna töös uuritud deletsioonide esinemissagedused on väga madalad, on nendega aheldunud SNP-ide esinemissagedused sama madalad ning kuna enamikul leitud SNP-idel puudus rs number, ei ole neid SNP-e tõenäoliselt eelnevalt analüüsitud teistes genotüübi andmestikes. Enamus mikrokiibi tootjatest võimaldavad tellija soovide kohaselt lisada SNP-e enda toodetavatele mikrokiipidele, mille puhul oleks võimalik leitud sobilikke märgend-SNP-e kasutada deletsioonide tuvastamiseks edaspidi.

Deletsioonidega aheldunud SNP-e, millel on rs number olemas, saab kasutada erinevate populatsioonide genotüübi andmestikest deletsiooniga indiviidide leidmiseks, juhul kui on eelnevalt kindlaks tehtud, et kindel SNP on tõepoolest deletsiooniga aheldunud, ning et need SNP-id on populatsioonide genotüübi andmetes olemas. Pika deletsiooniga potentsiaalselt aheldunud SNP-i rs190738878 võiks näiteks kasutada sellisteks otsinguteks, kuid tuleb silmas pidada, et ei ole võimalik vahet teha deletsiooniga haplotüübi ja tema vahetu eellashaplotüübi vahel. Seega võib andmestikest leida eellashaplotüübiga indiviide, kellel esineb uuritav SNP, kuid deletsiooni neil ei esine. Antud töös leitud lühikese deletsiooniga potentsiaalselt aheldunud SNP-ide suhtes võib öelda, et hetkel ei sobi ükski neist teistest populatsioonidest deletsioonide tuvastamiseks, kuna puudub piisavalt andmeid aheldatuse tõendamiseks. Seetõttu oleks kõige kindlam viis töös leitud *CYP2C19* geeni deletsioonide tuvastamiseks nende deletsioonide unikaalsete 25-meride kasutamine, kas siis kiibile lisatuna, et määrata deletsiooni olemasolu, või olemasolevatest sekveneerimisandmetest otsingute tegemiseks.

KOKKUVÕTE

P450 2C19 valku leidub endoplasmaatilises retiikulumis ning see on kliiniliselt oluline ensüüm, mis lagundab paljusid erinevaid ravimeid, sealhulgas krambivastast ravimit mefenütoiini, haavandivastaseid ravimeid nagu omeprasool, teatud antidepressante (estsitalopraam) ja malaariavastast ravimit proguaaniil (Blaisdell jt., 2002). P450 2C19 valku kodeeriva geeni *CYP2C19* deletsioonide puhul antud ensüümi funktsionaalsus indiviidil puudub (PM), kui tal on mõlemad geeni alleelid deletsiooniga, või esineb indiviidil vähenenud ensüümi aktiivsus (IM) kui tal on ainult üks deletsiooniga alleel. Aeglase (PM) ja keskmise (IM) metaboliseerija fenotüübiga indiviidide puhul on ravimi organismist elimineerimise kiirus vähenenud ning ravimi kontsentratsioon veres on suurenenud, mis võib kaasa tuua soovimatuid kõrvaltoimeid. Eelravimite puhul on PM ja IM fenotüübiga indiviididel organismis vähem ravimi aktiivset metaboliiti, mis võib põhjustada ravi ebaõnnestumist. Oluline on lihtsalt ja kindlalt suuta määrata selliseid valgu funktsionaalsust mõjutavaid geneetilisi polümorfisme.

Antud tööl oli kolm eesmärki:

1. Määrata *CYP2C19* geeni deletsioonide esinemissagedused Eesti populatsioonis;
2. Määrata WGS andmete põhjal *CYP2C19* deletsiooni kandjate diplotüübid;
3. Leida *CYP2C19* geeni deletsioonidega aheldunud SNP-id.

Töö esimese eesmärgi põhiselt uurisin *CYP2C19* geeni deletsioone TÜ EGV geenidoonorite andmete põhjal, et leida Eesti populatsioonis leiduvaid *CYP2C19* geeni deletsioone, ning et määrata nende deletsioonide esinemissagedused. WGS sekveneerimisandmete ning Omni ja GSA kiibi andmete ($n = 143\ 193$) analüüsimisel selgus, et *CYP2C19* geeni deletsioonidega indiviidide esinemissagedus Eesti populatsioonis on ligikaudu 1,68%. Seejuures GSA kiibi andmete ($n = 135\ 020$) põhiselt on pika deletsiooni alleeli esinemise sagedus $MAF = 0,756\%$ ja lühikese deletsiooni alleeli esinemise sagedus on $MAF = 0,026\%$. Töö teise eesmärgi põhiselt määrasin WGS andmete põhjal *CYP2C19* deletsiooniga indiviidide diplotüübid. WGS andmetes ($n = 2244$) leidis 37 indiviidil (1,6%) *2C19* geeni deletsioon ning nende indiviidide hulgas esineb ainult kolm erinevat star-alleelide põhist diplotüüpi, milleks on *CYP2C19*1/*37*, *CYP2C19*2/*37* ja *CYP2C19*17/*37*. Töö kolmanda eesmärgi põhiselt otsisin WGS andmete põhjal *CYP2C19* deletsioonidega potentsiaalselt aheldunud SNP-e. Leidsin pika deletsiooni kohta 3 ja lühikese deletsiooni kohta 19 võimalikku deletsiooniga aheldunud SNP-i.

Tulevikus kiiremaks ja usaldusväärsemaks *CYP2C19* pika ja lühikese deletsiooni määramiseks GSA kiibilt pakuti välja kaks peamist lahendust. Esiteks on võimalik GSA kiibile lisada deletsioonidega aheldunud SNP-e, mille esinemist indiviididel saab seejärel kontrollida. Sel juhul peab töös leitud potentsiaalseid aheldunud SNP-e eelnevalt täiendavalt uurima, et täiesti kindlalt saaks öelda, et need SNP-id on tõepoolest aheldunud töös uuritud deletsioonidega. Tuleb meeles pidada, et ei ole võimalik vahet teha deletsiooniga haplotüübi ja tema vahetu eellashaplotüübi vahel. Näiteks määratakse eellashaplotüübiga indiviid kiibi pealt tuvastatud märgend-SNP-i põhjal deletsiooni kandjaks, kuid tegelikult sel indiviidil pole *CYP2C19* deletsiooni. Seda probleemi ei esineks näiteks kui leitud deletsiooniga aheldunud SNP tekkis deletsiooniga samaaegselt genoomi, mispuhul poleks eellashaplotüübiga indiviidil seda SNP-i. Deletsioonide ja leitud potentsiaalsete aheldunud SNP-ide vaheliste kindlate seoste väljaselgitamine vajab edasist uurimist.

Teiseks võimaluseks on lisada GSA kiibile deletsioonide spetsiifilised järjestused (näiteks töös loodud 25-meri-id) ning vastavad referents genoomi põhised järjestused. Kui kiibil oleva oligonukleotiidsondi külge seondub sellega komplementaarset järjestust omav DNA fragment, võimaldab see polümeraasil pikendada sondi küljes olevat järjestust ühe märgistatud nukleotiidi võrra. Seondunud märgistatud nukleotiid võimaldab sondi oligonukleotiidjärjestuse pealt tuvastada laseriga ergastamise järgselt signaali. Kui indiviidil esineb genoomis deletsiooni spetsiifilist järjestust ja/või vastavat referents järjestust, on seda võimalik kindlaks teha kiibilt saadud signaalide põhjal. Näiteks tuvastatakse mõlema järjestuse pealt signaali, kui tegemist on deletsiooni suhtes heterosügootse indiviidiga.

Töös saadud tulemuste põhjal võib järeldada, et kiibi andmetest leitud *CYP2C19* deletsioonide algus- ja lõppkoordinaatide järgi on võimalik usaldusväärsest kinnitada uuritud deletsioonide olemasolu vastavalt antud töös kontrollitud gruppide põhjal. Kuid edaspidiseks deletsioonide määramiseks uutes indiviidides on soovitatav analüüsiks kasutatavale kiibile lisada deletsiooni põhiseid järjestusi või märgend-SNP-e kui nende seotus deletsiooniga on piisavalt tõendatud. Kiibilt saadava signaali põhised deletsiooni määramine on palju kiirem ja efektiivsem kui hilisem indiviidide gruppidesse jaotamine ennustatud CNV-de koordinaatide põhjal, mis tähendaks väiksemat töömahtu tulevikus nende deletsioonide määramisel.

SUMMARY

Cytochrome P450 2C19 deletions in the Estonian population

Kadri Maal

Cytochrome P450s are one of the major enzymes in humans involved in drug degradation. Individuals with genetic polymorphisms in the cytochrome P450 genes are at risk of adverse drug reactions or therapeutic failure at standard drug doses (Meyer, 2004). The goal of personalised medicine is to find an individualised prevention or treatment plan for each person by analysing their genetic data in combination with the environment, health behaviour and health data of the person (Personaalmehitsus, 2019).

Cytochrome P450 2C is one of the most important subfamilies of P450s, because its enzymes (CYP2C8, CYP2C9, CYP2C18, CYP2C19) degrade more than 20% of all therapeutic drugs (Rendic, 2002). P450 2C19 protein is found in the endoplasmic reticulum and it is a clinically important enzyme that degrades many different drugs, including the anticonvulsant drug mephenytoin, antiulcer drugs such as omeprazole, certain antidepressants (escitalopram), and the antimalarial drug proguanil (Blaisdell et al., 2002). If an individual has deletions on both of their *CYP2C19* alleles, they do not produce functional P450 2C19 protein at all and are thus deemed as a poor metaboliser (PM) for the drugs that are metabolised by P450 2C19. Individuals who have a deletion on only one of their *CYP2C19* alleles have reduced P450 2C19 enzyme activity and are deemed as intermediate metabolisers (IM). Individuals with the PM and IM phenotypes have a lower rate of elimination and higher plasma drug levels for drugs metabolised by P450 2C19, which can lead to adverse drug reactions. In the case of prodrugs, individuals with the PM and IM phenotypes produce fewer active drug metabolites, which can lead to potential loss of drug efficacy and therapeutic failure. It is important to be able to easily and reliably identify such genetic polymorphisms that affect drug metabolising enzyme functionality.

The aim of this study was to find and experimentally confirm deletions of the *CYP2C19* gene based on the data from the University of Tartu's Estonian Biobank to determine the frequencies of these deletions in the Estonian population. Based on WGS data I found two *CYP2C19* deletions in the Estonian population and based on the length of those deletions, I named them "long *CYP2C19* deletion" and "short *CYP2C19* deletion". The analysis of WGS data, Omni and GSA microarray data (n = 143 193) revealed that the frequency of individuals with the

CYP2C19 gene deletion in the Estonian population is approximately 1,68%. Based on the GSA microarray data (n = 135 020) the allele frequency of the long deletion is MAF = 0,756% and the allele frequency of the short deletion is MAF = 0,026%.

Based on WGS data (n = 2244) I determined diplotypes of individuals with these deletions. There were 37 individuals (1,6%) with a partial *CYP2C19* gene deletion (*CYP2C19**37) in that dataset with three different diplotype combinations based on PharmVar star allele database nomenclature. Those diplotypes were *CYP2C19**1/*37, *CYP2C19**2/*37 and *CYP2C19**17/*37. The *CYP2C19**1 represents a wild-type allele with normal P450 2C19 enzymatic activity, individuals with the *CYP2C19**2 and *CYP2C19**37 variant have markedly decreased enzyme activity, whereas individuals with the *CYP2C19**17 variant have increased enzymatic activity. From the WGS data I was also able to find SNPs that are potentially in linkage disequilibrium (LD) with the deletions I found previously, which could be used as tag SNPs for the deletions. For the long deletion I found 3 potential tag SNPs and for the short deletion I found 19 potential tag SNPs. Further investigation is needed to determine whether these SNPs can actually be used as tag SNPs for these deletions.

I propose two solutions for a faster and more reliable detection of *CYP2C19* deletions using the GSA microarray in the future. The first possible solution is adding deletion tag SNPs to the GSA microarray for detecting deletion based alleles in individuals. In this case the potential tag SNPs found in this work need to be further investigated in order to be able to say with absolute certainty that these SNPs are indeed linked to the deletions investigated in this study. Another problem with using tag SNPs is that based on just the SNP it is not possible to distinguish between the deletion haplotype and its immediate ancestral haplotype. For example, an individual with the precursor haplotype would be designated as a deletion carrier based on the tag SNP detected from the array, but in fact that individual does not have a *CYP2C19* deletion. This problem of distinguishing between the deletion haplotype and the precursor haplotype would not occur if the tag SNP originated in the genome at the same time as the deletion, which would mean that the precursor haplotype would not have this specific SNP.

The second possible solution could be to add deletion allele-specific oligos and corresponding reference sequence-based oligos to the GSA microarray. For the deletion-specific oligos it would be possible to use the 25-mer sequences used in this work. In this case it would be possible to determine if an individual has a specific deletion or not based on the signals you get from the specific oligos on the array beads. For example, a signal is detected from both the

deletion allele-specific oligo and the wild type allele-specific oligo if an individual is heterozygous for the deletion. This method is very deletion-specific and would not need any further analysis to determine the specific long and short deletion genotypes for any given individual.

Considering all the findings, I believe that the CNV start and end positions called with PennCNV software from the GSA microarray data are sufficient to determine if an individual has the *CYP2C19* long or short deletion when compared to the experimentally controlled groups in this study. However, for future detection of the long and short deletions using the GSA array, I would recommend adding the deletion allele-specific oligos to the microarray.

TÄNUSÕNAD

Soovin tänada enda juhendajaid Lili Milanit ja Tarmo Puuranda nende poolse abi, toetuse ja tagasiside eest. Samuti soovin tänada Liis Karo-Astoveri abi ja põhjaliku tagasiside eest.

KASUTATUD KIRJANDUSE LOETELU

- 1) Abe, I. (2007). Enzymatic synthesis of cyclic triterpenes. *Natural product report*, 24 (6): 1311 – 1331.
- 2) Aguiar, M., Masse, R. ja Gibbs, B. F. (2005). Regulation of cytochrome P450 by posttranslational modification. *Drug metabolism reviews*, 37 (2): 379 – 404.
- 3) Aithal, G. P., Day, C. P., Kesteven, P. J. ja Daly, A. K. (1999). Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *The Lancet*, 353 (9154): 717 – 719.
- 4) Albert, A. (1958). Chemical aspects of selective toxicity. *Nature*, 182 (4633): 421 – 423.
- 5) Andreson, R., Puurand, T., ja Remm, M. (2006). SNPmasker: automatic masking of SNPs and repeats across eukaryotic genomes. *Nucleic acids research*, 34: W651 – W655.
- 6) Balant, L. P., Gundert-Remy, U., Boobis, A. R. ja Von Bahr, C. H. (1989). Relevance of genetic polymorphism in drug metabolism in the development of new drugs. *European journal of clinical pharmacology*, 36 (6): 551 – 554.
- 7) Baldwin, R. M., Ohlsson, S., Pedersen, R. S., Mwinyi, J., Ingelman-Sundberg, M., Eliasson, E. ja Bertilsson, L. (2008). Increased omeprazole metabolism in carriers of the CYP2C19* 17 allele; a pharmacokinetic study in healthy volunteers. *British journal of clinical pharmacology*, 65 (5): 767 – 774.
- 8) Behrendorff, J. B., Huang, W. ja Gillam, E. M. (2015). Directed evolution of cytochrome P450 enzymes for biocatalysis: exploiting the catalytic versatility of enzymes with relaxed substrate specificity. *Biochemical Journal*, 467 (1): 1 – 15.
- 9) Blaisdell, J., Mohrenweiser, H., Jackson, J., Ferguson, S., Coulter, S., Chanas, B., Xi, T., Ghanayem, B. ja Goldstein, J. A. (2002). Identification and functional characterization of new potentially defective alleles of human CYP2C19. *Pharmacogenetics and Genomics*, 12 (9): 703 – 711.
- 10) Burchard, E. G., Ziv, E., Coyle, N., ... Risch, N. (2003). The importance of race and ethnic background in biomedical research and clinical practice. *New England Journal of Medicine*, 348 (12): 1170 – 1175.
- 11) Canfield, D. E. (2005). The Early History of Atmospheric Oxygen: Homage to Robert M. Garrels. *Annual Review of Earth and Planetary Sciences*, 33: 1 – 36.
- 12) Chiba, K., Kobayashi, K. A., Manabe, K. Y., Tani, M. A., Kamataki, T. E. ja Ishizaki, T. A. (1993). Oxidative metabolism of omeprazole in human liver

- microsomes: cosegregation with S-mephenytoin 4'-hydroxylation. *Journal of Pharmacology and Experimental Therapeutics*, 266 (1): 52 – 59.
- 13) Chang, C. C., Ng, C. C., Too, C. L., Choon, S. E., Lee, C. K., Chung, W. H., Hussein, S. H., Lim, K. S. ja Murad, S. (2017). Association of HLA-B* 15: 13 and HLA-B* 15: 02 with phenytoin-induced severe cutaneous adverse reactions in a Malay population. *The pharmacogenomics journal*, 17 (2): 170 – 173.
 - 14) Clancy, S. ja Shaw, K. (2008) DNA deletion and duplication and the associated genetic disorders. *Nature Education*, 1 (1): 23.
 - 15) Conney, A. H. (1967). Pharmacological implications of microsomal enzyme induction. *Pharmacological reviews*, 19 (3): 317 – 366.
 - 16) Correia, M. A., Sinclair, P. R. ja De Matteis, F. (2011). Cytochrome P450 regulation: the interplay between its heme and apoprotein moieties in synthesis, assembly, repair, and disposal. *Drug metabolism reviews*, 43 (1): 1 – 26.
 - 17) Daly, A. K. (2017). Pharmacogenetics: a general review on progress to date. *British Medical Bulletin*, 124: 65 – 79.
 - 18) Daxinger, L. ja Whitelaw, E. (2010). Transgenerational epigenetic inheritance: more questions than answers. *Genome research*, 20 (12): 1623 – 1628.
 - 19) Debeljak, N., Fink, M. ja Rozman, D. (2003). Many facets of mammalian lanosterol 14 α -demethylase from the evolutionarily conserved cytochrome P450 family CYP51. *Archives of Biochemistry and Biophysics*, 409 (1): 159 – 171.
 - 20) Debose-Boyd, R.A. (2007). A helping hand for cytochrome p450 enzymes. *Cell Metabolism*, 5 (2): 81 – 3.
 - 21) DeGorter, M. K., Xia, C. Q., Yang, J. J. ja Kim, R. B. (2012). Drug transporters in drug efficacy and toxicity. *Annual review of pharmacology and toxicology*, 52: 249 – 273.
 - 22) De Morais, S. M., Wilkinson, G. R., Blaisdell, J., Nakamura, K., Meyer, U. A. ja Goldstein, J. A. (1994). The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. *Journal of Biological Chemistry*, 269 (22): 15419 – 15422.
 - 23) Desta, Z., Zhao, X., Shin, J. G. ja Flockhart, D. A. (2002). Clinical significance of the cytochrome P450 2C19 genetic polymorphism. *Clinical pharmacokinetics*, 41 (12): 913 – 958.
 - 24) Edwards, I. R. ja Aronson, J. K. (2000). Adverse drug reactions: definitions, diagnosis, and management. *The Lancet*, 356 (9237): 1255 – 1259.

- 25) Evans, D. A. P. (1993). Genetic factors in drug therapy: clinical and molecular pharmacogenetics. Cambridge University Press, Cambridge.
- 26) Evans, W. E. 2006. General Principles of Clinical Pharmacokinetics, lk. 3 – 7. M. E. Burton, L. M. Shaw, J. J. Schentag ja W. E. Evans (toim.), Applied Pharmacokinetics & Pharmacodynamics: principles of therapeutic drug monitoring. Lippincott Williams & Wilkins, USA.
- 27) Evans, W. E. ja Relling, M. V. (2004). Moving towards individualized medicine with pharmacogenomics. *Nature*, 429 (6990): 464 – 468.
- 28) Ferguson, R. J., De Morais, S. M., Benhamou, S., ... Goldstein, J. A. (1998). A new genetic defect in human CYP2C19: mutation of the initiation codon is responsible for poor metabolism of S-mephenytoin. *Journal of Pharmacology and Experimental Therapeutics*, 284 (1): 356 – 361.
- 29) Funck-Brentano, C., Becquemont, L., Leneuve, A., Roux, A., Jaillon, P. ja Beaune, P. (1997). Inhibition by Omeprazole of Proguanil Metabolism: Mechanism of the Interaction In Vitro and Prediction of In Vivo Results from the In Vitro Experiments. *Journal of Pharmacology and Experimental Therapeutics*, 280 (2): 730 – 738.
- 30) Ganten, D. ja Ruckpaul, K. 2006. Encyclopedic reference of genomics and proteomics in molecular medicine. Springer, Berliin, Saksamaa.
- 31) Gilardi, G. 2018. Cytochrome P450 redox activity, lk. 90 – 109. K. Wandelt (toim.), Encyclopedia of interfacial chemistry. Elsevier, Oxford, UK.
- 32) Goldstein, J. A. (2001). Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. *British journal of clinical pharmacology*, 52 (4): 349 – 355.
- 33) Goldstein, J. A. ja Blaisdell, J. (1996). Genetic tests which identify the principal defects in CYP2C19 responsible for the polymorphism in mephenytoin metabolism, *Methods in Enzymology*, 272: 210 – 218.
- 34) Gonzalez, F. J. ja Gelboin, H. V. (1992). Human cytochrome P450: evolution and cDNA-directed expression. *Environmental Health Perspectives*, 98: 81 – 85.
- 35) Gonzalez, F. J. ja Nebert, D. W. (1990). Evolution of the P450 gene superfamily: animal-plant „warfare“, molecular drive and human genetic differences in drug oxidation. *Trends in Genetics*, 6: 182 – 186.
- 36) Guengerich, F. P. (2002). Rate-limiting steps in cytochrome P450 catalysis. *Biological Chemistry*, 383 (10): 1553 – 1564.

- 37) Guengerich, F. P. 2015. Human Cytochrome P450 Enzymes, lk. 523 – 786. P. R. Ortiz de Montellano (toim.), Cytochrome P450. Structure, Mechanism, and Biochemistry, 4th ed. Springer, California, USA.
- 38) Guengerich, F. P. 2019. Fundamental aspects of drug metabolism, lk. 1 – 178. P. G. Pearson ja L. C Wienkers (toim.), Handbook of drug metabolism, vol 186. Informa Healthcare, New York, USA.
- 39) Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A. ja Deisenhofer, J. (1995). Structure and function of cytochrome-P450: a comparative analysis of 3 crystal structures. *Structure*, 3 (1): 41 – 62.
- 40) Hein, D. W., Boukouvala, S., Grant, D. M., Minchin, R. F. ja Sim, E. (2008). Changes in consensus arylamine N-acetyltransferase (NAT) gene nomenclature. *Pharmacogenetics and genomics*, 18 (4): 367.
- 41) Holden, C. (2003). Race and medicine. *Science*, 302 (5645): 594 – 596.
- 42) Holford, N. H. ja Sheiner, L. B. (1981). Pharmacokinetic and pharmacodynamic modeling in vivo. *Critical reviews in bioengineering*, 5 (4): 273 – 322.
- 43) Housworth, E. A. ja Postlethwait, J. (2002). Measures of synteny conservation between species pairs. *Genetics*, 162 (1): 441 – 448.
- 44) Ibeanu, G. C., Blaisdell, J., Ferguson, R. J., ... Goldstein, J. A. (1999). A Novel Transversion in the Intron 5 Donor Splice Junction of CYP2C19 and a Sequence Polymorphism in Exon 3 Contribute to the Poor Metabolizer Phenotype for the Anticonvulsant Drug S-Mephenytoin. *Journal of Pharmacology and Experimental Therapeutics*, 290 (2): 635 – 640.
- 45) Ibeanu, G. C., Goldstein, J. A., Meyer, U. R. S., Benhamou, S., Bouchardy, C., Dayer, P., Ghanayem, B. I., ja Blaisdell, J. (1998). Identification of new human CYP2C19 alleles (CYP2C19* 6 and CYP2C19* 2B) in a Caucasian poor metabolizer of mephenytoin. *Journal of Pharmacology and Experimental Therapeutics*, 286 (3): 1490 – 1495.
- 46) Inaba, T., Nebert, D. W., Burchell, B., Watkins, P. B., Goldstein, J. A., Bertilsson, L. ja Tucker, G. T. (1995). Pharmacogenetics in clinical pharmacology and toxicology. *Canadian Journal of Physiology and Pharmacology*, 73: 331 – 338.
- 47) Jackson, C. J., Lamb, D. C., Marczylo, T. H., Warrilow, A. G., Manning, N. J., Lowe, D.J., Kelly, D. E. ja Kelly, S. L. (2002). A novel sterol 14 α -demethylase/ferredoxin fusion protein (MCCYP51FX) from *Methylococcus capsulatus* represents a new class of the cytochrome P450 superfamily. *Journal of Biological Chemistry*, 277: 46959 – 46965.

- 48) Janha, R. E., Worwui, A., Linton, K. J., Shaheen, S. O., Sisay-Joof, F. ja Walton, R. T. (2014). Inactive alleles of cytochrome P450 2C19 may be positively selected in human evolution. *BMC evolutionary biology*, 14 (1): 1 – 10.
- 49) Kahn, R. A. ja Durst, F. 2000. *Function and Evolution of plant cytochrome P450*, lk. 151 – 177. J. T. Romeo, R. Ibrahim, L. Varin ja V. De Luca (toim.), *Evolution of Metabolic Pathways*. Elsevier Science Ltd, Oxford, UK.
- 50) Kalman, L. V., Agúndez, J. A. G., ... Zanger, U. M. (2016). Pharmacogenetic allele nomenclature: international workgroup recommendations for test result reporting. *Clinical Pharmacology & Therapeutics*, 99 (2): 172 – 185.
- 51) Kalow, W. 1962. *Pharmacogenetics. Heredity and the response to drugs*. Saunders Company, London, UK.
- 52) Kalow, W. (2006). Pharmacogenetics and pharmacogenomics: origin, status, and the hope for personalized medicine. *The pharmacogenomics journal*, 6 (3): 162 – 165.
- 53) Kazui, M., Nishiya, Y., Ishizuka, T., Hagihara, K., Farid, N. A., Okazaki, O., Ikeda, T. ja Kurihara, A. (2010). Identification of the human cytochrome P450 enzymes involved in the two oxidative steps in the bioactivation of clopidogrel to its pharmacologically active metabolite. *Drug Metabolism and Disposition*, 38 (1): 92 – 99.
- 54) Karam, W. G., Goldstein, J. A., Lasker, J. M. ja Ghanayem, B. I. (1996). Human CYP2C19 is a major omeprazole 5-hydroxylase, as demonstrated with recombinant cytochrome P450 enzymes. *Drug Metabolism and Disposition*, 24 (10): 1081 – 1087.
- 55) Koressaar, T. ja Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23 (10): 1289 – 1291.
- 56) Küpfer, A. ja Preisig, R. (1984). Pharmacogenetics of mephenytoin: a new drug hydroxylation polymorphism in man. *European journal of clinical pharmacology*, 26 (6): 753 – 759.
- 57) Lackie, J. M. 2007. *The dictionary of cell and molecular biology*, 4th ed. Elsevier, London, UK.
- 58) Lamb, D. C., Hargrove, T. Y., Zhao, B., Wawrzak, Z., Goldstone, J. V., Nes, W. D., Kelly, S. L., Waterman, M. R., Stegeman, J. J. ja Lipesheva, G. I. (2020). Concerning P450 evolution: Structural analyses support bacterial origin of sterol 14 α -demethylases. *Molecular Biology and Evolution*, 38 (3): 952 – 967.

- 59) Lauschke, V. M., Milani, L. ja Ingelman-Sundberg, M. (2018). Pharmacogenomic biomarkers for improved drug therapy — recent progress and future developments. *The AAPS journal*, 20 (1): 1 – 16.
- 60) Lewis, D. F. V., Watson, E. ja Lake, B. G. (1998). Evolution of the cytochrome P450 superfamily: sequence alignments and pharmacogenetics. *Mutation Research/Reviews in Mutation Research*, 410 (3): 245 – 270.
- 61) Licinio J. ja Wong M-L. 2002. *Pharmacogenomics: The search for individualized therapies*. Wiley-VCH Verlag GmbH: Weinheim, Saksamaa.
- 62) Marshall A. (1997). Genset-Abbott deal heralds pharmacogenomics era. *Nature Biotechnology*, 15: 829 – 830.
- 63) Mason, H. S. 1968. Homology among oxidases, lk. 69 – 91. N. V. Thoai ja J. Roche (toim.), *Homologous enzymes and biochemical evolution*. Gordon & Breach, New York.
- 64) McCormack, M., Alfirevic, A., Bourgeois, S., ... Pirmohamed, M. (2011). HLA-A* 3101 and carbamazepine-induced hypersensitivity reactions in Europeans. *New England Journal of Medicine*, 364 (12): 1134 – 1143.
- 65) Mega, J. L., Close, S. L., Wiviott, S. D., ... Sabatine, M. S. (2009). Cytochrome p-450 polymorphisms and response to clopidogrel. *New England journal of medicine*, 360 (4): 354 – 362.
- 66) Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature reviews genetics*, 11 (1): 31 – 46.
- 67) Meyer, U. (2004). Pharmacogenetics – five decades of therapeutic lessons from genetic diversity. *Nature Reviews Genetics*, 5: 669 – 676.
- 68) Meyer, U. A. ja Zanger, U. M. (1997). Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annual review of pharmacology and toxicology*, 37 (1): 269 – 296.
- 69) Mizutani, T. (2003). PM frequencies of major CYPs in Asians and Caucasians. *Drug metabolism reviews*, 35 (2 – 3): 99 – 106.
- 70) Morey, M., Fernández-Marmiesse, A., Castiñeiras, D., Fraga, J. M., Couce, M. L. ja Cocho, J. A. (2013). A glimpse into past, present, and future DNA sequencing. *Molecular genetics and metabolism*, 110 (1 – 2): 3 – 24.
- 71) Musumba, C. O., Jorgensen, A., Sutton, L., Van Eker, D., Zhang, E., O'Hara, N., Carr, D. F., Pritchard, D. M. ja Pirmohamed, M. (2013). CYP2C19* 17 gain-of-function polymorphism is associated with peptic ulcer disease. *Clinical Pharmacology & Therapeutics*, 93 (2): 195 – 203.

- 72) Nakamura, K., Goto, F., Ray, W. A., McAllister, C. B., Jacqz, E., Wilkinson, G. R. ja Branch, R. A. (1985). Interethnic differences in genetic polymorphism of debrisoquin and mephenytoin hydroxylation between Japanese and Caucasian populations. *Clinical Pharmacology & Therapeutics*, 38 (4): 402 – 408.
- 73) Nebert, D. W. (1997). Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? *American Journal of Human Genetics*, 60: 265 – 271.
- 74) Nebert, D. W. ja Feyereisen, R. 1994. Evolutionary argument for a connection between drug metabolism and signal transduction, lk. 3 – 13. M. C. Lechner (toim.), *Cytochrome P450. Biochemistry, Biophysics and molecular biology*. John Libbey Eurotext, Paris.
- 75) Nebert, D. W., Nelson, D. R. ja Feyereisen, R. (1989). Evolution of the cytochrome P450 genes. *Xenobiotica*, 19 (10): 1149 – 1160.
- 76) Nebert, D. W. Ja Vesell, E. S. 2013. Pharmacogenetics and pharmacogenomics, ptk. 19. D. Rimoin, R. Pyeritz ja B. Korf (toim.), *Emery and Rimoin's Principles and Practice of Medical Genetics*, 6th ed. Elsevier, Oxford, UK.
- 77) Nebert, D. W., Wikvall, K. ja Miller, W. L. (2013). Human cytochromes P450 in health and disease. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368 (1612): 20120431.
- 78) Nelson, D. R. (1999). Cytochrome P450 and the individuality of species. *Archives of Biochemistry and Biophysics*, 369: 1 – 10.
- 79) Nelson, D. R., Goldstone, J. V. ja Stegeman, J. J. (2013). The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368 (1612): 20120474.
- 80) Nelson, D. R., Kamataki, T., Waxman, D. J., ... Nebert, D. W. (1993). The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes and nomenclature. *DNA and Cell Biology*, 12: 1 – 51.
- 81) Nelson, D. R., Koymans, L., Kamataki, T., ... Nebert, D. W. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, 6: 1 – 42.
- 82) Niwa, T. ja Yamazaki, H. (2012). Comparison of cytochrome P450 2C subfamily members in terms of drug oxidation rates and substrate inhibition. *Current drug metabolism*, 13 (8): 1145 – 1159.

- 83) Obayashi, E., Takahashi, S. ja Shirp, Y. (1998). Electronic-structure of reaction intermediate of cytochrome P450Nor in its nitric-oxide reduction. *Journal of American Chemical Society*, 120: 12964 – 12965.
- 84) Ohno, S. 1970. *Evolution by gene duplication*. Springer, Berlin.
- 85) Ortiz de Montellano, P. R. (2013). Cytochrome P450-activated prodrugs. *Future medicinal chemistry*, 5 (2): 213 – 228.
- 86) Rapoport, B. L., Troncone, G., Schmitt, F. ja Nayler, S. J. 2020. *Comprehensive Genomic Profiling*. Karger, Oxford, UK.
- 87) Paquette, S. M., Bak, S. ja Feyereisen, R. (2000). Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA and Cell Biology*, 19 (5): 307 – 317.
- 88) Pirmohamed, M. (2014). Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. *Annual review of genomics and human genetics*, 15: 349 – 370.
- 89) Postlethwait, J., Amores, A., Cresko, W., Singer, A. ja Yan, Y. L. (2004). Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends in Genetics*, 20 (10): 481 – 490.
- 90) Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C. ja Kraut, J. (1985). The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450. *Journal of Biological Chemistry*, 260 (30): 16122 – 16130.
- 91) Pratt, V. M., Everts, R. E., Aggarwal, P., ... Kalman, L. V. (2016). Characterization of 137 genomic DNA reference materials for 28 pharmacogenetic genes: a GeT-RM collaborative project. *The Journal of molecular diagnostics*, 18 (1): 109 – 123.
- 92) Pratt, V. M., Zehnbauer, B., Wilson, J. A., ... Kalman, L. (2010). Characterization of 107 genomic DNA reference materials for CYP2D6, CYP2C19, CYP2C9, VKORC1, and UGT1A1: a GeT-RM and Association for Molecular Pathology collaborative project. *The Journal of molecular diagnostics*, 12 (6): 835 – 846.
- 93) Purcell, S., Neale, B., Todd-Brown, K., ... Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3): 559 – 575.
- 94) Putnam, N. H., Butts, T., Ferrier, D. E., ... Rokhsar, D. S. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature*, 453 (7198): 1064 – 1071.
- 95) Raucy, J. L., Mueller, L., Duan, K., Allen, S. W., Strom, S. ja Lasker, J. M. (2002). Expression and induction of CYP2C P450 enzymes in primary cultures of human

- hepatocytes. *Journal of Pharmacology and Experimental Therapeutics*, 302 (2): 475 – 482.
- 96) Relling, M. V. ja Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature*, 526 (7573): 343 – 350.
- 97) Rendic, S. (2002). Summary of information on human CYP enzymes: human P450 metabolism data. *Drug metabolism reviews*, 34 (1-2): 83 – 448.
- 98) Režen, T., Debeljak, N., Kordis, D. ja Rozman, D. (2004). New aspects on lanosterol 14 α -demethylase and cytochrome P450 evolution: lanosterol/cycloartenol diversification and lateral transfer. *Journal of Molecular Evolution*, 59: 51 – 58.
- 99) Reynald, R. L., Sansen, S., Stout, C. D. ja Johnson, E. F. (2012). Structural Characterization of Human Cytochrome P450 2C19: active site differences between P450s 2C8, 2C9, and 2C19. *Journal of Biological Chemistry*, 287 (53): 44581 – 44591.
- 100) Rhodes, J. S. ja Crabbe, J. C. (2005). Gene expression induced by drugs of abuse. *Current opinion in pharmacology*, 5 (1): 26 – 33.
- 101) Richardson, T. H. ja Johnson, E. F. 1996. The CYP2C Subfamily, lk. 161 – 175. C. Ioannides (toim.), *Cytochromes P450: metabolic and toxicological aspects*. CRC Press, USA.
- 102) Ridderström, M., Zamora, I., Fjellström, O. ja Andersson, T. B. (2001). Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA. *Journal of medicinal chemistry*, 44 (24): 4072 – 4081.
- 103) Rodrigues, A. D., Fernandez, D., Nosarzewski, M. A., Pierce Jr, W. M. ja Prough, R. A. (1991). Inhibition of hepatic microsomal cytochrome P-450 dependent monooxygenation activity by the antioxidant 3-tert-butyl-4-hydroxyanisole. *Chemical research in toxicology*, 4 (3): 281 – 289.
- 104) Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A. ja Feldman, M. W. (2002). Genetic structure of human populations. *Science*, 298 (5602): 2381 – 2385.
- 105) Rosenborg, S. O., Mwinyi, J., Andersson, M., Baldwin, R. M., Pedersen, R. S., Sim, S. C., Bertilsson, L., Ingelman-Sundberg, M. ja Eliasson, E. (2008). Kinetics of omeprazole and escitalopram in relation to the CYP2C19* 17 allele in healthy subjects. *European journal of clinical pharmacology*, 64 (12): 1175 – 1179.
- 106) Rozman, D., Strömstedt, M. ja Waterman, M. R. (1996). The three human cytochrome P450 lanosterol 14 α -demethylase (CYP51) genes reside on

- chromosomes 3, 7, and 13: Structure of the two retrotransposed pseudogenes, association with a line-1 element, and evolution of the human CYP51 family. *Archives of Biochemistry and Biophysics*, 333 (2): 466 – 474.
- 107) Rudberg, I., Hendset, M., Uthus, L. H., Molden, E. ja Refsum, H. (2006). Heterozygous mutation in CYP2C19 significantly increases the concentration/dose ratio of racemic citalopram and escitalopram (S-citalopram). *Therapeutic drug monitoring*, 28 (1): 102 – 105.
- 108) Rudberg, I., Mohebi, B., Hermann, M., Refsum, H. ja Molden, E. (2008). Impact of the ultrarapid CYP2C19* 17 allele on serum concentration of escitalopram in psychiatric patients. *Clinical Pharmacology & Therapeutics*, 83 (2): 322 – 327.
- 109) Sanford, J. C., Guo, Y., Sadee, W. ja Wang, D. (2013). Regulatory polymorphisms in CYP2C19 affecting hepatic expression. *Drug metabolism and drug interactions*, 28 (1): 23 – 30.
- 110) Santos, M., Niemi, M., Hiratsuka, M., Kumondai, M., Ingelman-Sundberg, M., Lauschke, V. M. ja Rodríguez-Antona, C. (2018). Novel copy-number variations in pharmacogenetics contribute to interindividual differences in drug pharmacokinetics. *Genetics in Medicine*, 20 (6): 622 – 629.
- 111) Scott, A. ja Scott, G. N. (2013). Mechanisms of Drug Interactions. *Pharmacy Tech Topics™*, 18 (3): 1 – 19.
- 112) Scott, S., Sangkuhl, K., Gardner, E. E., Stein, C. M., Hulot, J. S., Johnson, J. A., Roden, D. M., Klein, T. E. ja Shuldiner, A. R. (2011). Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450–2C19 (CYP2C19) genotype and clopidogrel therapy. *Clinical Pharmacology & Therapeutics*, 90 (2): 328 – 332.
- 113) Shah, R. R., Oates, N. S., Idle, J. R., Smith, R. L. ja Lockhart, J. D. (1982). Impaired oxidation of debrisoquine in patients with perhexiline neuropathy. *British Medical Journal (Clinical Research Edition)*, 284 (6312): 295 – 299.
- 114) Shen, R., Fan, J. B., Campbell, D., ... Oliphant, A. (2005). High-throughput SNP genotyping on universal bead arrays. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 573 (1-2): 70 – 82.
- 115) Shimada, T., Yamazaki, H., Mimura, M., Inui, Y. ja Guengerich, F. P. (1994). Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *Journal of Pharmacology and Experimental Therapeutics*, 270 (1): 414 – 423.

- 116) Shuldiner, A. R., O'Connell, J. R., Bliden, K. P., ... Gurbel, P. A. (2009). Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA*, 302 (8): 849 – 857.
- 117) Sibbing, D., Koch, W., Gebhard, D., ... Kastrati, A. (2010). Cytochrome 2C19*17 allelic variant, platelet aggregation, bleeding events, and stent thrombosis in clopidogrel-treated patients with coronary stent placement. *Circulation*, 121 (4): 512 – 518.
- 118) Sim, S. C., Risinger, C., Dahl, M. L., Aklillu, E., Christensen, M., Bertilsson, L. ja Ingelman-Sundberg, M. (2006). A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clinical Pharmacology & Therapeutics*, 79 (1): 103 – 113.
- 119) Simoons, F. J. 1998. Favism and the Origin of the Pythagorean Ban on Fava Beans, lk. 216 – 247. F. J. Simoons (toim.), *Plants of Life, Plants of Death*. The University of Wisconsin Press, London.
- 120) Song, W. C. ja Brash, A. R. (1991). Purification of an allene oxide synthase and identification of the enzyme as a cytochrome-P450. *Science*, 253: 781 – 784.
- 121) Stephens, J. C., Schneider, J. A., Tanguay, D. A., ... Vovis, G. F. (2001). Haplotype variation and linkage disequilibrium in 313 human genes. *Science*, 293 (5529): 489 – 493.
- 122) Steward, D. J., Haining, R. L., Henne, K. R., Davis, G., Rushmore, T. H., Trager, W. F. ja Rettie, A. E. (1997). Genetic association between sensitivity to warfarin and expression of CYP2C9*3. *Pharmacogenetics*, 7 (5): 361 – 367.
- 123) Zanger, U. M. ja Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & therapeutics*, 138 (1): 103 – 141.
- 124) Zhao, Y., White, M. A., Muralidhara, B. K., Sun, L., Halpert, J. R. ja Stout, C. D. (2006). Structure of microsomal cytochrome P450 2B4 complexed with the antifungal drug bifonazole: insight into P450 conformational plasticity and membrane interaction. *Journal of Biological Chemistry*, 281 (9): 5973 – 5981.
- 125) Zhukov, A. ja Ingelman-Sundberg, M. (1999). Relationship between cytochrome P450 catalytic cycling and stability: fast degradation of ethanol-inducible cytochrome P450 2E1 (CYP2E1) in hepatoma cells is abolished by inactivation of its electron donor NADPH-cytochrome P450 reductase. *Biochemical Journal*, 340 (2): 453 – 458.

- 126) Taylor, C., Crosby, I., Yip, V., Maguire, P., Pirmohamed, M. ja Turner, R. M. (2020). A Review of the Important Role of CYP2D6 in Pharmacogenomics. *Genes*, 11 (11): 1295.
- 127) Tervasmäki, A., Winqvist, R., Jukkola-Vuorinen, A. ja Pylkäs, K. (2014). Recurrent CYP2C19 deletion allele is associated with triple-negative breast cancer. *BMC cancer*, 14 (1): 1 – 7.
- 128) Tiroch, K. A., Sibbing, D., Koch, W., Roosen-Runge, T., Mehilli, J., Schömig, A. ja Kastrati, A. (2010). Protective effect of the CYP2C19* 17 polymorphism with increased activation of clopidogrel on cardiovascular events. *American heart journal*, 160 (3): 506 – 512.
- 129) Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. ja Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic acids research*, 40 (15): e115.
- 130) Vesell, E. S. (1968). Genetic control of drug levels in man: antipyrine. *Science*, 161: 72 – 73.
- 131) Vogel F. (1959). Moderne probleme der humangenetik. *Ergebnisse der Inneren Medizin und Kinderheilkunde*, 12: 52 – 62.
- 132) Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., Hakonarson, H. ja Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research*, 17 (11): 1665 – 1674.
- 133) Weber, W. W. 1997. *Pharmacogenetics*. Oxford University Press, Oxford, UK.
- 134) Weinshilboum, R. (2003). Inheritance and drug response. *New England Journal of Medicine*, 348 (6): 529 – 537.
- 135) Wilkinson, G. R. 2001. *Pharmacokinetics: the dynamics of drug absorption, distribution, and elimination*, lk. 3 – 29. J. G. Hardman, L. E. Limbird ja A. G. Gilman (toim.), *Goodman & Gilman's the pharmacological basis of therapeutics*, 10th ed. McGraw-Hill, New York.
- 136) Wolfe, K. (2000). Robustness – it's not where you think it is. *Nature Genetics*, 25 (1): 3 – 4.
- 137) Wöll, S., Kim, S. H., Greten, H. J. ja Efferth, T. (2013). Animal plant warfare and secondary metabolite evolution. *Natural products and bioprospecting*, 3 (1): 1 – 7.
- 138) Xiao, Z. S., Goldstein, J. A., Xie, H. G., ... Zhou, H. H. (1997). Differences in the incidence of the CYP2C19 Polymorphism affecting the S-mephenytoin

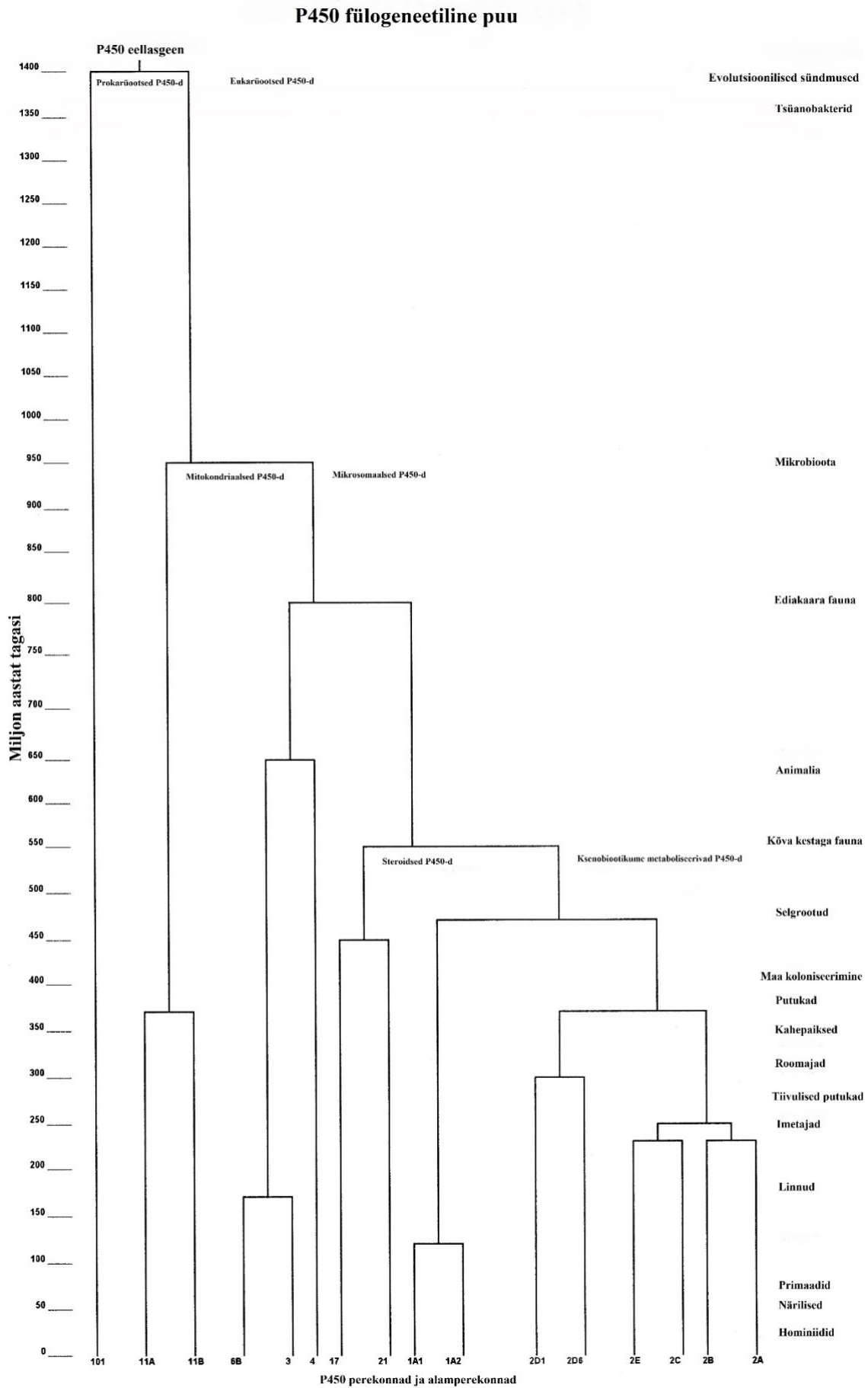
phenotype in Chinese Han and Bai populations and identification of a new rare CYP2C19 mutant allele. *Journal of Pharmacology and Experimental Therapeutics*, 281 (1): 604 – 609.

- 139) Yamazaki, H., Nakamura, M., Komatsu, T., ... Yokoi, T. (2002). Roles of NADPH-P450 reductase and apo- and holo-cytochrome b5 on xenobiotic oxidations catalyzed by 12 recombinant human cytochrome P450s expressed in membranes of *Escherichia coli*. *Protein expression and purification*, 24 (3): 329 – 337.
- 140) Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D. ja Johnson, E. F. (2004). The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution. *Journal of Biological Chemistry*, 279 (37): 38091 – 38094.
- 141) Yoshida, Y., Aoyama, Y., Noshiro, M. ja Gotoh, O. (2000). Sterol 14-demethylase P450 (CYP51) provides a breakthrough for the discussion on the evolution of cytochrome P450 gene superfamily. *Biochemical and Biophysical Research Communications*, 273: 799 – 804.
- 142) Yoshida, Y., Noshiro, M., Aoyama, Y., Kawamoto, T., Horiuchi, T. ja Gotoh, O. (1997). Structural and evolutionary studies on sterol 14-demethylase P450 (CYP51), the most conserved P450 monooxygenase: II. Evolutionary analysis of protein and gene structures. *Journal of Biochemistry*, 122: 1122 – 1128.

Kasutatud veebiaadressid

- 1) Arslan, J. (2015). Pythagoras: The fava of pharmacogenomics. Kasutatud 16.11.2019, <http://www.lateralmag.com/columns/gene-dosage/pythagoras-the-fava-of-pharmacogenomics>
- 2) CPIC. (2019). Kasutatud 19.11.2019, <https://cpicpgx.org/>
- 3) dbSNP. (2021). Kasutatud 28.07.2021, <https://www.ncbi.nlm.nih.gov/snp/>
- 4) Ensembl. (2021). Kasutatud 28.07.2021, <http://www.ensembl.org/index.html>
- 5) HGP Europe: Genomes, Diseases, Drugs and Diagnostics. (1997). Kasutatud 16.11.2019, <http://www.bio.net/bionet/mm/gen-link/1997-January/001193.html>
- 6) HUGO Gene Nomenclature Committee: CYP2C19. (2019). Kasutatud 05.01.2020, https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/HGNC:2621
- 7) Human Genome Dating. (2021). Kasutatud 27.07.2021, <https://human.genome.dating/>
- 8) Illumina: Introduction to DNA sequencing. (2021). Kasutatud 15.07.2021, <https://www.illumina.com/techniques/sequencing/dna-sequencing.html>

- 9) Illumina: Microarray Technology. (2021). Kasutatud 16.07.2021,
<https://emea.illumina.com/science/technology/microarray.html>
- 10) LOVD: CYP2C19. (2021). Kasutatud 28.07.2021,
https://databases.lovd.nl/shared/variants/CYP2C19#object_id=VariantOnTranscript%20CVariantOnGenome&id=CYP2C19&order=VariantOnTranscript%20FHaplotype%20CASC&search_transcriptid=06002&search_VariantOnGenome/Remarks=refe&page_size=100&page=1
- 11) NCBI: CYP2C19. (2019). Kasutatud 05.01.2020,
<http://view.ncbi.nlm.nih.gov/gene/1557>
- 12) NetPrimer. (2019). Kasutatud 15.10.2019, <https://www.premierbiosoft.com/netprimer/>
- 13) Personaalmehitsiin. (2019). Kasutatud 26.04.2021,
<https://www.sm.ee/et/personaalmehitsiin-1>
- 14) PharmVar. (2021). Kasutatud 06.07.2021, <https://www.pharmvar.org/>
- 15) PharmVar: CYP2C19. (2021). Kasutatud 06.07.21,
<https://www.pharmvar.org/gene/CYP2C19>
- 16) PharmVar: Structural Variation *CYP2C19*. (2019). Kasutatud 12.07.2021,
https://a.storyblok.com/f/70677/x/9225fa2a03/variation_cyp2c19.pdf
- 17) Ravimiseadus. (2004). Kasutatud 28.05.2021,
<https://www.riigiteataja.ee/akt/104052016004>
- 18) Ravimite kõrvaltoimetest teatamise juhend tervishoiutöötajale. (2020). Kasutatud 28.05.2021, <https://www.ravimiamet.ee/en/node/1461>
- 19) SuperCYP. (2021). Kasutatud 23.07.2021,
<https://bioinformatics.charite.de/supercyp/index.php?site=home>



Joonis on võetud Lewis jt. (1998) artiklist (tõlgitud ja kohandatud).

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Kadri Maal,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose “Tsütokroom P450 2C19 deletsioonide esinemine Eesti populatsioonis”,

mille juhendajad on prof. PhD Lili Milani ja MSc Tarmo Puurand,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Kadri Maal

08.08.2021