

Tartu Ülikool
Arvutiteaduse Instituut
Informaatika õppekava

Ainika Adamson

Meetod tõenäosusliku ravimisoovituse andmiseks osalise geeniinfo põhjal

Bakalaureusetöö (9 EAP)

Juhendaja: Sulev Reisberg, MSc

Tartu 2018

Meetod tõenäosusliku ravimisoovituse andmiseks osalise geeniinfo põhjal

Lühikokkuvõte: Personaalmehitsiinis on pead tõstmas valdkond, kus ravimidoos määratakse inimese geneetika põhjal, sest ravimite ainevahetuse kiirus sõltub suuresti just geneetikast. Geeniinfo põhjal ravimisoovituste tegemine on keeruline; suurimaks probleemiks on puudulik informatsioon, sest enimlevinud genotüpiseerimise võimalused tuvastavad pärilikkusainest ainult osalise info. Probleemi leevendamiseks luuakse käesolevas töös meetod, mis oskab puuduva info osas teha tõenäosuslikke ennustusi. Meetod kasutab puuduva info asemel varem publitseeritud Euroopa alleelisagedusi, mille abil ennustatakse erinevate metabolismikiiruste esinemistõenäosused. Loodud algoritmi rakendati mitme elutähtsa ravimi ainevahetust reguleeriva geeni CYP2C19 andmetel. Tulemusena selgus, et puuduva geeniinfo korral on võimalik ennustada fenotüüpide tõenäosusi.

Võtmesõnad: farmakogeneetika, andmeanalüüs, PharmGKB, ravimisoovitused

CERCS: B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika; B220 Geneetika, tsütogeneetika; P175 Informaatika, süsteemiteooria

Method for predicting dosing guidelines with missing genetic data

Abstract: The aim of this current thesis is to construct a method that helps to predict pharmaceutical dosing guidelines with missing genetic data. Making dosing guidelines for different drugs is difficult because most commonly used genotyping tools are not sufficient enough to collect all needed data. This means that if we do not get all needed genetic data then we cannot propose any guidelines. My goal was to alleviate that problem and construct method that is based on European allele frequencies and makes predictions for every phenotype.

Keywords: pharmacogenetics, data analysis, PharmGKB, dosing guidelines

CERCS: B110 Bioinformatics, medical informatics, biomathematics, biometrics, B220 Genetics, cytogenetics, P175 Informatics, systems theory

Sisukord

Sissejuhatus	4
Mõisted ja lühendid	5
1 Kirjanduse ülevaade	6
1.1 Farmakogeneetika	6
1.2 Geneetika.....	6
1.3 Olulised geenid farmakogeneetikas	7
1.4 Ravimisoovitused	7
2 Metoodika.....	9
2.1 Andmed ja geenispetsiifilised tabelid	9
2.1.1 Alleelide definitsiooni tabel	9
2.1.2 Alleelide sagedustabel.....	10
2.1.3 Diplotüübi – fenotüübi tabel	11
2.1.4 Konstrueeritud diplotüübi sagedustabel	12
2.2 Geeniinfo.....	13
2.3 Algoritmi selgitus ja skeem.....	14
3 Tulemused	19
3.1 Alleelisagedused.....	19
3.2 Fenotüüpide sagedus populatsioonis.....	20
3.3 Fenotüübis ennustuste detailsed tõenäosused.	21
3.4 Algoritmitulemuste ja Euroopa fenotüübisageduste erinevused	22
4 Arutelu.....	23
5 Kokkuvõte	24
Kasutatud allikad.....	25
Lisa 1 – CYP2C19 diplotüüpide sagedustabel	27
Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.....	29

Sissejuhatus

Tänapäeva meditsiinis kasutatakse järjest rohkem personaalmeditsiini meetodikaid. Üha enam tahetakse ravipraktikaid isikustada konkreetse inimese jaoks, kasutades tema enda geneetilist infot. Kahjuks on pärilikkusainest sageli raske üheselt mõistetavat teavet tuvastada. Kõnealune probleem on üks suuremaid põhjuseid, miks geneetikat tarvitatakse meditsiinilistes ravivõtetes vähe [1].

Ravimite soovituslikud annused sõltuvad suuresti inimese geneetikast, sest mutatsioonid määravad ravimite ainevahetuse kiiruse. Metabolismi kiirusest sõltub omakorda aeg, kui kaua püsivad ravimid organismis. Mida lühemat aega saab toimeaine mõjuda, seda suurem peaks olema annus ning mida kauem ravim organismis püsib, seda väiksemat doosi vajab inimene. Seetõttu oleks tarvis ka osalise bioloogilise andmestiku korral ravimidoose määrata. Käesoleva töö eesmärgiks ongi inimeste geeniinfo põhjal anda võimalikult täpset tagasisidet soovituslike ravimiannuste osas ka puuduliku geeniinformatsiooni korral.

Kogu vajaminevat teavet oleks võimalik välja lugeda ainult genoomi täieliku sekveneerimisega, kuid kahjuks on see meetod võrdlemisi kallis ja seetõttu vähelevinud. Käesolevas töös on kasutatud *Illumina Global Screening Array* kiibiga genotüpiseeritud andmeid. GSA on odav ja populaarne kiip, mis mõõdab ligikaudu miljonit asukohta ehk lookust geenis. [7]. Näiteks uues Geenivaramu EV100 projektis kasutatakse just GSA kiipi geenivariantide tuvastamiseks. Ometigi ei mõõda mainitud kiip alati kõiki olulisi lookusi, mistõttu katab genereeritav andmestik sageli ainult osaliselt vajamineva info.

Tänaasel päeval on ravimiannuste korrigeerimiseks kasutusel mitmeid spetsiaalseid geeniteste, mis mõõdavad iga geeni puhul üksikuid lookusi, kuid nendes eeldatakse, et kõigis ülejäänud lookustes ka kindlasti pole mutatsioone. Selline lähenemine ei ole bioloogilisest vaatepunktist täiesti korrektne, sest välistab automaatselt nende ülejäänud asukohtadega seotud geenivariandid. Lisaks tuleks kindlasti eraldi arvesse võtta mutatsioone, mida kiip ei mõõda.

Kõnealuse probleemi leevendamine ongi käesoleva töö eesmärk. Lõputöö siht on disainida meetod, millega saaks anda geenidoonoritele pädevat tagasisidet ka puuduliku geeninfo korral ning hinnata meetodi rakendatavust.

Mõisted ja lühendid

Alleel – lõik geenist

Diplo tüüp – mõlemad alleelid geeni ühes asukohas

Farmakogeneetika – geneetika valdkond, mis uurib geneetika mõju ravimitele ja nende metabolismile

Fenotüüp – avalduvate geneetiliste tunnuste kogum

Geen – funktsionaalne lõik pärilikkusainest

GSA - *Illumina Global Screening Array*, genotüüpiseerimiseks kasutatav mikrokiip

Haplotüüp – käesolevas töös kasutatakse sünonüümina alleelile

Lookus – asukoht geenis

PharmGKB – farmakogeneetika avalik andmebaas

1 Kirjanduse ülevaade

Käesolevas peatükis kirjeldatakse töös kasutatud bioloogilist terminoloogiat. Alljärgnevalt tutvustatakse farmakogeneetika valdkonda ja kirjeldatakse ravimisoovituste sisu.

1.1 Farmakogeneetika

Farmakogeneetika on teadusharu, kus uuritakse inimfüsioloogia reaktsioone erinevatele ravimitele. Üks olulisemaid valdkondi on mutatsioonide mõju ravimite metabolismikiirusele. Farmakogeneetika eesmärgiks on disainida uusi ja reguleerida olemasolevate ravimite doosierimist sõltuvalt patsiendi geneetilisest eripärast. Ravimid ei mõju kõigile inimestele ühtviisi ja seetõttu on keeruline geneetilist infot omamata ennustada, kuidas konkreetne patsient toimeainetele reageerib. [1,2]

Inimkehas määravad sageli erinevad maksaensüümid, kuidas keha ravimitele reageerib. Ensüümide sünteesi reguleerivad enamasti geenid – seega on inimese geneetilise koodi abil võimalik ennustada inimkeha reaktsioone teatud medikamentidele. Oluline on teada, kui kaua osalevad ravimid inimese ainevahetusprotsessis ning kui suur on tõenäosus allergiliseks reaktsiooniks. Mida kauem osaleb ravim inimese metabolismis, seda suurem on ravimi mõju ning mida vähem ta ainevahetusringis osaleb, seda väiksem on toimeaine mõju. [2,3]

1.2 Geneetika

Geenide ehk pärilikkusaine funktsionaalsete lõikude erinevates asukohtades paiknevad muutused määravad ravimite metabolismikiiruse. Ainevahetus sõltub omakorda nii emalt kui isalt päritud geeniinfost. Mõlema vanema kombineeritud geneetika määrabki isesuguste medikamentide vajamineva doosi, kuid geenivariantide tuvastamiseks peab esmalt mõistma mutatsioonide sisu.

Punktmutatsioonid on muutused geeni lõigus, kus üks nukleotiid on asendunud teisega. Nukleotiidid on orgaanilised molekulid, mille abil sünteesitaksegi nukleiinhappeid ehk siinkohal DNAd. Nukleotiidid koosnevad lämmastikalusest, suhkrust ja vähemalt ühest fosfaatrühmast. Bioloogias ja käesolevas töös tähistatakse ja klassifitseeritakse nukleotiide sageli lämmastikaluste järgi. Tabelis 1 kirjeldatakse DNA lämmastikaluseid, mida on neli. Pürimidiinid tsütosiin ja tümiin on väiksemad ning puriinid adeniin ja guaniin suuremad molekulid. Tabelis on märgitud ka igale lämmastikalusele vastav tähis. Punktmutatsioonides on kõige sagedamini vahetunud pürimidiinalused omavahel ehk tsütosiin tümiiniga ja vastupidi

ning puriinalustest adeniin ja guaniin vahetuvad omakeskis. Harvem vahetuvad puriinalused pürimidiiniga ja vastupidi.

Tabel 1. DNA nukleotiidide lämmastikalused

Pürimidiin		Puriin	
Tsütosiin	Tümiin	Adeniin	Guaniin
C	T	A	G

Muutused nukleotiidides on antud töö kontekstis olulised, sest just nukleotiidide abil sünteesitaksegi kõik keha valgud, sealhulgas ensüümid. Ensüümide regulatsioonist sõltub jällegi ainevahetuse kiirus ja niiviisi ongi väikesed mutatsioonid metabolismi iseärasuste aluseks.

Eelnevalt kirjeldatud mutatsioonide tuvastamisega saab määrata inimese geenivariandi, kuid selleks tuleb ühes asukohas määrata mõlemad alleelid ehk lõigud geenist. Alleelide abil määratakse diplotüüp, mis koosneb kahest alleelist. Käesoleva töö raames nimetatakse haplotüübiks ehk alleeliks mutatsioonide kogumit samas geenis. Haplotüübi tähistamiseks on numbri ees tärn, näiteks *1.

Diplo tüübi abil tuvastatakse fenotüüp, mis on juba tunnuste kogum, mis määrabki metabolismi kiiruse. Kui diplotüüp tähistab geneetilist informatsiooni, siis fenotüüp on avaldunud tunnus, näiteks silmade värvus. Käesoleva töö (ravimi metabolismi) kontekstis on kasutusel kuus erinevat fenotüüpi: ülikiire (*ultrarapid*), kiire (*rapid*), normaalne (*normal*), keskmine (*intermediate*), aeglasele kalduv (*likely intermediate*) ja aeglane fenotüüp (*poor*). [4]

1.3 Olulised geenid farmakogeneetikas

Tsütokroom P450 on ensüümiklass, mille ensüümid vastutavad ligikaudu 40% ravimite metabolismi kiiruse eest. Nende ensüümide kiirus sõltub peamiselt neljast geenist: CYP2D6, CYP2C9, CYP2C19 ja CYP3A5. Seetõttu on kõnealused geenid ravimisoovituste tegemisel eriliselt olulised. Käesolevas töös on algoritmi näide toodud geeni CYP2C19 näitel. [5]

1.4 Ravimisoovitused

Geen CYP2C19 reguleerib maksaensüümi tööd, mis omakorda määrab ligikaudu 10% ravimite ainevahetuse kiiruse. Sealhulgas vastutab CYP2C19 laialdaselt kasutusel olevate mao ülihappelisust neutraliseerivate ravimite omeprasooli ja pantoprasooli metabolismikiiruste eest.

Lisaks reguleerib kõnealune geen veresoonekonna haiguste ravis kasutatava klopidogreeli metabolismikiirust. [3]

Ravimisoovitused põhinevad punktmutatsioonide olemasolul või puudumisel konkreetsetes geeni asukohtades, mida tähistavad rs-koodid (näiteks rs12248560 tähistab 10. kromosoomil asukohas 96521657 mutatsiooni, kus C asemel on T). Punktmutatsioonide olemasolu konkreetsetes lookustes mõõdavad mikrokiibid. Käesolevas töös on kasutatud kiipi GSA, mis mõõdab ligikaudu miljonit asukohta pärilikkusaines [7]. Igas lookuses on tavaliselt normaalne lämmastikualus, mida nimetatakse wild type'ks. Kui inimesel tuvastatakse normaalne lämmastikalus, siis toimub ravimi metabolism normaalse kiirusega. Kui aga mainitud asukohas on mõni muu lämmastikalus, siis on ravimite ainevahetuse kiirus tõenäoliselt kas kiirem või aeglasem ja inimene vajab annuse korrigeerimist või asendusravimit. [1, 4]

2 Metoodika

2.1 Andmed ja geenispetsiifilised tabelid

Käesolevas töös kasutatakse geenianalüüsina Eesti Geenivaramu poolt genotüüpiseeritud andmeid GSA kiibilt ja järgmisi abitabeleid [10]:

- 1) CYP2C19 alleelide definitsiooni tabel (PharmGKB andmebaasist [8])
- 2) CYP2C19 alleelide sagedustabel (Human Genome Diversity projektist [9])
- 3) CYP2C19 diplotüübi – fenotüübi tabel (PharmGKB andmebaasist [8])

PharmGKB on avalik andmebaas, mis sisaldab infot geneetika ja ravimite metabolismi vaheliste seoste kohta.

Alljärgnevalt selgitatakse nende andmete ja tabelite sisu ning eesmärki.

2.1.1 Alleelide definitsiooni tabel

Geeni alleelide definitsioonide tabelit on meetodis vaja punktmutatsioonide alusel haplotüübi määramiseks.

Tabelis 2 on toodud geeni CYP2C19 alleelide definitsioonide lihtsustatud tabel (kogu tabel oleks visualiseerimiseks liiga mahukas). Tabeli seis pärineb 12. märtsist 2018. Esimeses veerus on loetletud haplotüübid ja nende järel iga haplotüübi jaoks vajalik lämmastikalus (mutatsioon). Mutatsiooni asukohta geenis tähistatakse rs-koodiga (ülemine rida). Teises reas *1 järel vastab igale asukohale mutatsioonita lämmastikalus.

Alljärgnevast näeme, et kui inimesel on näiteks asukohas, mida tähistab kood rs72558186 lämmastikualus adeniin (A), siis tal on haplotüüp ehk alleel *7. Kui tal oleks samas asukohas hoopis tümiin (T), siis oleks tegemist 'wild type'ga ehk normaalse alleeliga. Samas haplotüübi *17 määramiseks peaks inimesel olema rs12248560 asukohas tümiin (T), kuid kui seal on hoopis tsütosiin (C), siis on jällegi tegemist normaalse haplotüübiga.

Tabel 2. CYP2C19 modifitseeritud alleelide definitsiooni tabel [10]

Alleel	rs12248560	rs55752064	rs41291156	rs17884712	rs4986893	rs6413438	rs4244285	rs72558186	rs1789685	rs56337013
*1	C	T	T	G	G	C	G	T	C	C
*2							A			
*3					A					
*4										
*5										T
*6										
*7								A		
*8			C							
*9				A						
*10						T				
*11										
*12										
*13									T	
*14		C								
*15										
*16										
*17	T									

Alleeli tuvastamisel tekib probleem, kui kiip ei mõõda mõnda asukohta. Olukorras, kus kiip ei määra lämmastikalust kõigis positsioonides ja ei tuvasta ülejäänud asukohtade seas ühtegi mutatsiooni, ei saa täielikult kindel olla, et inimesel mutatsioone üldse pole. Seega ei saa üheselt haplotüüpi määrata. Näiteks kui rs17879685 ja rs56337013 koode kiip ei määra ja me mujal ühtegi muutust geenis ei tuvasta, siis võivad inimesel olla haplotüübid *1, *5 või *13.

2.1.2 Alleelide sagedustabel

Geeni alleelide sagedustabelit on algoritmis tarvis diplotüübi sageduste arvutamiseks ja mitteesinevate alleelide välistamiseks.

Geeni CYP2C19 alleelide sagedustabel Euroopas ja Põhja-Ameerikas on kujutatud tabelis 3. PharmGKB andmebaasis on olemas alleelisagedused ka teiste etniliste inimrühmade jaoks, kuid meie meetodis kasutame ainult eurooplaste sagedustabelit, kuna kasutatud geenandmed pärinevad kõik Euroopa populatsioonist. Tabeli andmed pärinevad 12. märtsist 2018.

Tabel 3. Geeni CYP2C19 Euroopa ja Põhja-Ameerika alleelide sagedustabel [10]

CYP2C19 alleel	Euroopa+Põhja-Ameerika alleelisagedus
*1	0,624
*2	0,146
*3	0,00619
*4	0,00278
*5	0,0000253
*6	0,0019
*7	0
*8	0,00256
*9	0,00025
*10	0
*12	0
*13	0,001
*14	0,00000333
*15	0,002
*16	0
*17	0,213
*22	0
*35	0

Esimeses veerus on geeni alleel, millele vastab teises veerus vastava alleeli sagedus. Näiteks alleeli *3 sagedus on 0,00619 ehk 0,6% populatsioonist esineb see alleel. Ühtlasi näeme, et näiteks alleeli *7 ei esine eurooplaste seas üldse. Alljärgnevalt selgub, et kõige sagedasem alleel on *1 sagedusega 0,624, mis tähistab mutatsioonideta alleeli.

2.1.3 Diplotüübi – fenotüübi tabel

Diplotüübi-fenotüübi tabelit kasutatakse diplotüübi põhjal fenotüübi määramiseks.

Geeni CYP2C19 diplotüübi – fenotüübi tabel, mis sisaldab alleeli *1 on kuvatud tabelis 4. Joonise esimeses veerus on toodud haplotüüpidest koosnevad diplotüübid, millele vastab paremal konkreetne fenotüüp. Joonisel on kuvatud ruumi kokkuhoiu mõttes ainult alleeli *1 sisaldavate diplotüüpide fenotüübid. Töö raames tuli kasutada kõiki Euroopas esinevatele haplotüüpidele vastavaid tabeleid. Tabeli seis pärineb kuupäevast 12. märts 2018. Näiteks diplotüübiga *1/*9 inimene on selle andmeil keskmise metabolismiga.

Tabel 4. Geeni CYP2C19 alleeli *1 sisaldav diplotüübi-fenotüübi tabel [10]

CYP2C19 Diplotüüp	Fenotüüp
*1/*1	CYP2C19 Normaalne metaboliseerija
*1/*2	CYP2C19 Keskmise metaboliseerija
*1/*3	CYP2C19 Keskmise metaboliseerija
*1/*4	CYP2C19 Keskmise metaboliseerija
*1/*5	CYP2C19 Keskmise metaboliseerija
*1/*6	CYP2C19 Keskmise metaboliseerija
*1/*8	CYP2C19 Keskmise metaboliseerija
*1/*9	CYP2C19 Keskmise metaboliseerija
*1/*13	CYP2C19 Normaalne metaboliseerija
*1/*14	Teadmata
*1/*15	CYP2C19 Normaalne metaboliseerija
*1/*17	CYP2C19 Ülikiire metaboliseerija

2.1.4 Konstrueeritud diplotüübi sagedustabel

Diplo tüübi sagedustabelit (tabel 3) on algoritmis tarvis, et grupeerida võimalike diplotüüpide sagedused vastavate fenotüüpide (tabel 4) järgi.

Diplo tüübi esinemissageduste tabel on konstrueeritud alleelide sagedustabeli (tabel 3) põhjal, kasutades valemit (1):

$$D_{m n} = H_m \cdot H_n,$$

kus D_{mn} tähistab diplotüübi sagedust ning diplotüüp koosneb haplotüüpidest m ja n . H_m on vastavalt haplotüübi m sagedus ja H_n haplotüübi n sagedus. Tabelis 5 on kuvatud (osaline) diplotüübi sagedustabel, kus on esindatud diplotüübid, mis sisaldavad alleeli *1, täies mahus tabel on leitav peatükist Lisa 1.

Tabel 5. *1 alleeli sisaldavate diplotüüpide esinemissagedus (kasutatud valemit 1)

Diplotüüp	Esinemissagedus
*1/*1	0,38937600
*1/*2	0,09110400
*1/*3	0,00386256
*1/*4	0,00173472
*1/*5	0,00001579
*1/*6	0,00118560
*1/*8	0,00159744
*1/*9	0,00015600
*1/*13	0,00062400
*1/*14	0,00000208
*1/*15	0,00124800
*1/*17	0,13291200
...	...
*17/*17	0,045369

Näiteks diplotüübi *1/*3 esinemissagedus Euroopas on 0,00386, mis tähendab, et 0,39 protsendil inimestest on mainitud diplotüüp.

2.2 Geeniinfo

Käesolevas töös kasutatakse Eesti Geenivaramus GSA kiibiga genotüpiseeritud inimeste andmeid. Tabelis 6 on toodud geeniinfo näide. Esimeses veerus on rs-kood, mis märgib vastavat mutatsiooni, teises veerus on kromosoomi number, millele järgneb mutatsiooni asukoht. Neljandas veerus asetseb doonori identifikaator ja kahes viimases on konkreetsed tuvastatud alleelid – üks ühelt ja teine teiselt vanemalt. Näiteks inimesel identifikaatoriga 4 olid 19ndas kromosoomis positsioonil 15990431 lämmastikualused tümiin (T) ja tsütosiin (C).

Tabel 6. GSA geeniinfo näide

RS-kood	Kromosoom	Positsioon	Geenidoonor	Alleel 1	Alleel 2
rs2108622	19	15990431	1	T	C
rs2108622	19	15990431	2	C	C
rs2108622	19	15990431	3	C	C
rs2108622	19	15990431	4	T	C
rs2108622	19	15990431	5	T	T

Kokku sisaldab geeniandmestik 19 341 inimese andmeid. Andmestik sisaldab CYP2C19 kohta 4 273 392 rida.

Nimetatud kiibiga on võimalik mõõta ligikaudu üks miljon punktmutatsiooni, kuid see moodustab vaid murdosa inimese kogu genoomist [7]. Näiteks geeni CYP2C19 täielike ravimisoovituste jaoks oleks tarvis tuvastada 38 mutatsiooni, kuid GSA mõõdab vaid 10 positsiooni. Seega on ilmne, et täielikke ja üheseid ravimisoovitusi saab anda vaid väga vähestele inimestele, sest kui GSA ei mõõda 28 asukohta, siis me ei tea, kas neis on mutatsioone või mitte. Selleks, et inimesele üheselt fenotüüpi määrata, peaksid tema mutatsioonid paiknema mõnes neis kümnes asukohas, mida GSA mõõdab. Järelikult kui meil puudub sisuline info 28 olulise asukoha kohta, siis võib juhtuda, et mõlemat haplotüüpi ei saa doonori kohta üheselt määrata.

2.3 Algoritmi selgitus ja skeem

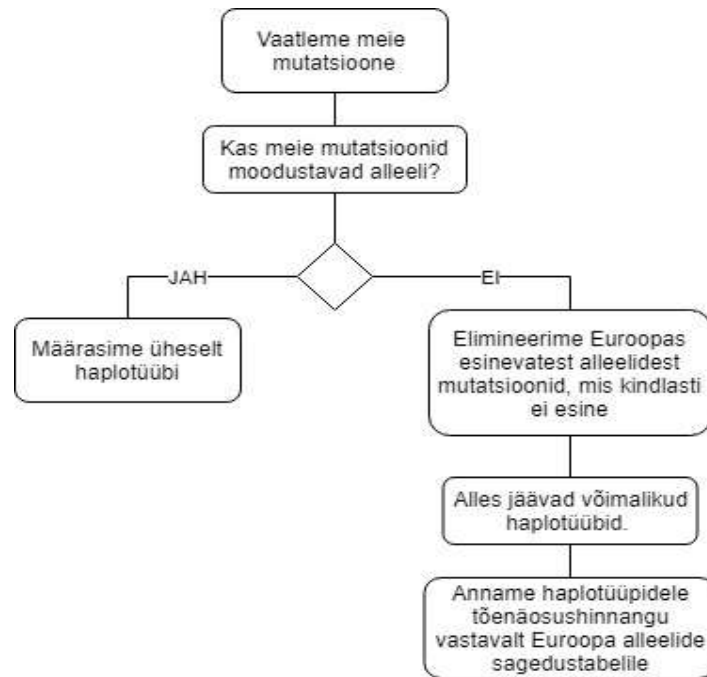
Alljärgnevalt kirjeldame algoritmi, mis suudaks siiski teha doonori kohta fenotüübi ennustuse ka osalise geeniinfo põhjal.

Algoritmi toimemehhanismi võib jagada seitsmeks etapiks. Järgnevalt kirjeldame kõiki etappe lähemalt. Lihtsuse mõttes toome näite kujuteldava geeniga X.

Etapid:

1. Filtreerime inimese geeniinfost konkreetse geeni jaoks olulised mutatsioonid (kasutades tabelit 2)
2. Püüame mutatsioonidega määrata haplotüüpi (tabeli 2 abil)
3. Leiame kõik potentsiaalsed haplotüübid
4. Konstrueerime haplotüüpidest kõikvõimalikud diplotüübid
5. Leiame diplotüüpide sagedused (tabeli 3 abil)
6. Grupeerime diplotüüpide sagedused fenotüüpide järgi (tabeli 4 abil)
7. Skaleerime sagedused 100% peale ja arvutame fenotüüpide esinemistõenäosused.

Algoritmi esimeses etapis leiame inimese geeniinfost mutatsioonid, mis on geeni ravimisoovituste tegemiseks vajalikud. Seejärel püüame muutuste põhjal tuvastada haplotüüpi. Kuna haplotüüpe on kaks, siis tuleb mõlemad määrata.



Joonis 1. Haplotüübi leidmine

Joonis 1 illustreerib haplotüübi määramist. Selleks on kaks võimalust (etapp 2): me kas suudame mutatsioonide abil üheselt määrata haplotüübi või on info puudulik, mistõttu peame välistama alleelid, mis kindlasti ei esine ja allesjäänutele andma esinemistõenäosuse PharmGKB alleelide sagedustabeli järgi.

Joonisel 2 on selgitatud, kuidas määratakse haplotüüpi kujuteldava geeni X puhul. Jooniselt on näha, et geeni kõikide haplotüüpide määramiseks oleks tarvis mõõta 10 mutatsiooni (kõik vertikaalsed tulbad), kuid realselt mõõtsime ainult neli muutust (joonisel vertikaalsed rohelise taustaga rs-koodid) Nende nelja mutatsiooniga suudaksime üheselt määrata ainult neli haplotüüpi: *3, *7, *8 ja *10. Seetõttu on selge, et ravimisoovituste tagasisidet on mõistlik anda tõenäosuslikul viisil.

Haplotüüp	rs12248560	rs55752064	rs41291556	rs17884712	rs4986893	rs6413438	rs4244285	rs72558186	rs17879685	rs56337013
*1	C	T	T	G	G	C	G	T	C	C
*2							A			
*3					A					
*4		G								
*5										T
*6	A									
*7								A		
*8			C							
*9				A						
*10						T				
*11									A	

Joonis 2. Mutatsioonide põhjal haplotüübi määramine

Järgmine tähtis aspekt ravimisoovituste juures on asjaolu, et juhul kui kiip ei tuvastanud inimesel ühtegi mutatsiooni, siis teame, et tal võivad olla (etapp 3) potentsiaalselt ainult haplotüübid *1, *2, *4, *5, *6, *9, *11, sest ülejäänute jaoks oleksime pidanud leidma mutatsioone.

Edasi tuleb neljandas etapis emalt ja isalt pärinevad võimalikud haplotüübid omavahel kõikvõimalikult kombineerida ja luua sedasi potentsiaalselt esinevad diplotüübid. Näiteks geeni X puhul oleksid konstrueeritud diplotüübid: *1/*1, *1/*2, *1/*4, *1/*5, *1/*6, *1/*9, *1/*11, *2/*1, *2/*2, *2/*4, ..., *11/*11.

Etapis 5 tuleb diplotüübi-fenotüübi tabelist (tabel 4) leida igale konstrueeritud diplotüübile vastav fenotüüp ning valida konstrueeritud diplotüübi sagedustabelist (tabel 5) diplotüübile vastav sagedus. Edasi tuleb sagedused grupeerida (etapp 6) fenotüübi järgi ja tõenäosused etapis 7 skaleerida 100% peale, mille alusel saabki inimesele anda fenotüüpide esinemistõenäosuse.

Diplootüüpide sageduste grupeerimist ja tõenäosuste skaleerimist iseloomustab järgnev valem (2):

$$F = \frac{\sum D_F}{\sum D},$$

kus F on konkreetse fenotüübi esinemistõenäosus, D_F märgib fenotüübiga F diplotüüpide sagedusi ja D märgib kõikide konstrueeritud diplotüüpide sagedusi.

Haplotüüp	*1	*2	*3	*4	*5	*6	*7	*8	*9	*10	*11
*1	0,3894	0,091104		0,00173	0	0					0,000808
*2	0,0911	0,021316		0,00041	0	0					0,000189
*3											
*4	0,0017	0,000406		7,7E-06	0	0					3,6E-06
*5	0	0		0	0	0					0
*6	0	0		0	0	0					0
*7											
*8											
*9											
*10											
*11	0,0008	0,000189		3,6E-06	0	0					1,68E-06

Joonis 3. Potentsiaalsed diplotüübid ja fenotüübid konkreetsel inimesel geeniga X

Joonisel 3 on kujutatud näitena ühe inimese potentsiaalsete diplotüüpide sagedusi, mille loomiseks läheb tarvis konstrueeritud diplotüüpide sagedustabelit (tabel 5 geeni X kohta). Näites toodud inimesel me ei suutnud üheselt haplotüüpe määrata, sest me ei tuvastanud üheski meile teadaolevas asukohas mutatsioone (st. alleele *3, *7, *8 ja *10 ei tuvastatud, sest vastavates asukohtades polnud mutatsioone). Ühtlasi kui geeni X alleelide sagedustabelis on *5 ja *6 väärtusteks 0, siis võime tõenäosushinnangutest ka need välja jätta.

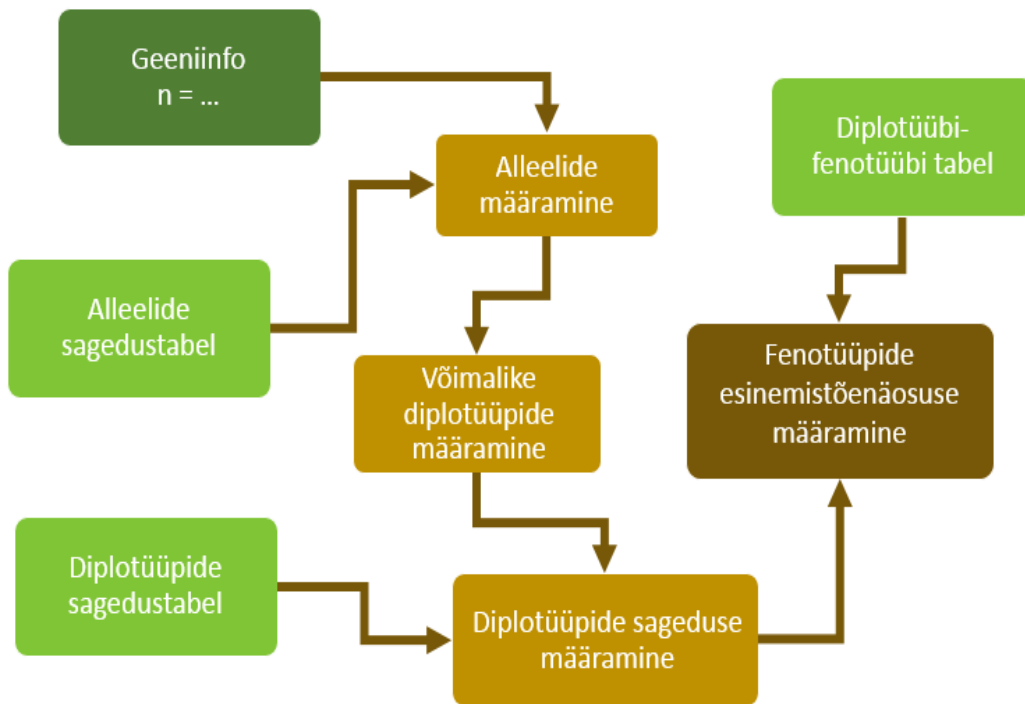
Seega summaarselt jätame geeni X puhul tõenäosushinnangust välja alleelid, mida Euroopas ei esine (*5 ja *6) ja alleelid, mis kindlasti meie populatsioonis ei esine (*3, *7, *8, *10).

Alles jäävad neli haplotüüpi *1, *2, *4 ja *11, moodustasime tabelis nende kombinatsioonid ja diplotüüpide sagedustabeli abil leiame tabelisse kirjed. Seejärel grupeerime diplotüübid fenotüüpide järgi (Geeni X diplotüübi-fenotüübi tabeli abil). Joonisel tähistab roheline taustavärv normaalset, kollane aeglast ja lilla kiiret metabolismi. Näiteks diplotüüpidel *1/*1 ja *11/*11 oleks normaalne metabolism.

Nüüd skaleerime kõik sagedused fenotüüpide järgi 100% peale ja leiame nende esinemistõenäosuse. Selleks kasutame valemit 2. Esmalt tuleb grupeerida normaalse (rohelse taustaga) fenotüübiga diplotüüpide sagedused. Seejärel tuleb liita *1/*1 ja *11/*11 sagedused ja jagada kõikide diplotüüpide sageduste summaga. Sarnaselt tuleb käituda aeglasele ja kiirele fenotüübile vastavate kirjetega. Niiviisi selgub, et näitena toodud (joonis 3) konkreetne geenidoonor on 65% tõenäosusega normaalne metaboliseerija, 34,7% tõenäosusega aeglane ja ainult 0,3% tõenäosusega kiire metaboliseerija. Sellise tagasisidega inimene teaks, et pigem on

tal kalduvus aeglasele ravimite metabolismile kui kiirele, mis omakorda tähendaks pigem keskmisest suuremat kui väiksemat ravimiannust.

Kogu kirjeldatud meetodit illustreerib joonis 4. Helerohelise taustaga plokid kujutavad PharmGKB tabeleid, tumerohelise taustaga Geenivaramu geeniandmestikku. Helepruunid kastikesed märgivad andmeanalüüsi vahepunkte ning tumepruun taust kirjeldab lõpptulemust.



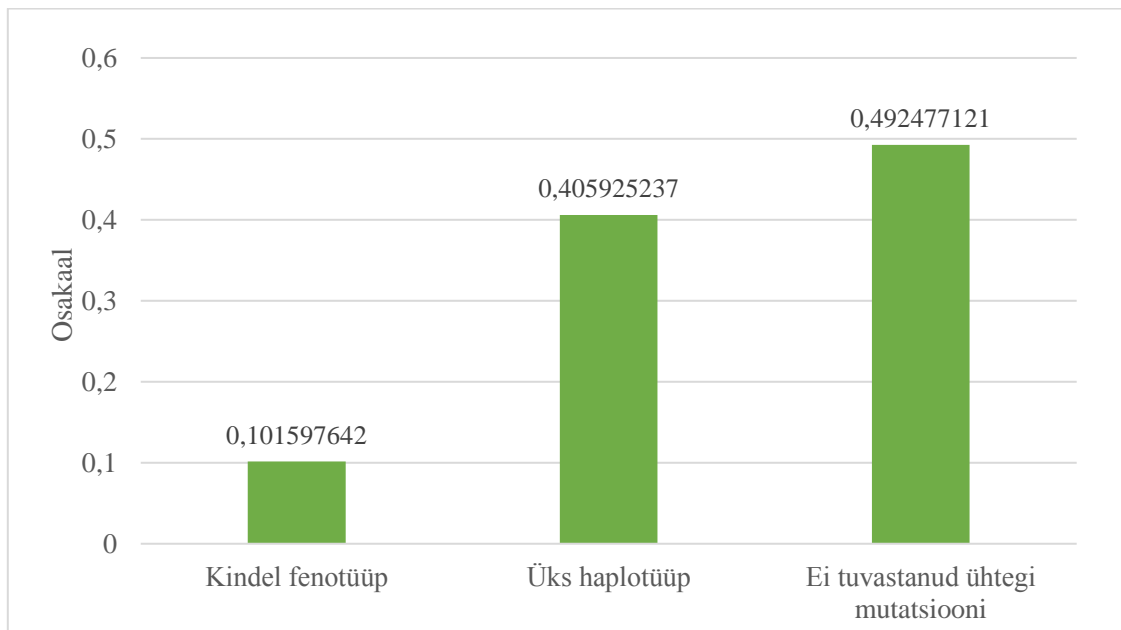
Joonis 4. Meetodi ülevaatlisk skeem

Järgnevalt rakendasime kirjeldatud meetodit Eesti Geenivaramu andmestikul ja arvutasime geeni CYP2C19 fenotüüpide esinemistõenäosused. Tulemused esitan järgmises peatükis. Meetodi testimiseks loodi spetsiaalne Python skript. Skript on kättesaadav Githubis: <https://github.com/ainika/Farmakogeneetika>; Graafikud loodi Microsoft Exceliga.

3 Tulemused

Järgnevas peatükis kirjeldan tulemusi.

GSA kiibiga saime ravimisoovitusi geeni CYP2C19 raames teha 19341 inimesele. Neist 1965 inimesele õnnestus üheselt määrata fenotüüp. 7851 indiviidile suutsime määrata ainult ühe haplotüübi ja 9525 inimesel ei tuvastatud ühtegi punktmutatsiooni. Vastavat informatsiooni kajastab joonis 5.



Joonis 5. Geeni CYP2C19 inimgrupid fenotüübi määramisel

Jooniselt näeme, et inimesi, kellel suutsime määrata mõlemad alleelid ja seega fenotüübi üheselt oli kõige vähem ehk 10%. Kõige rohkem oli inimesi, kellel me ei tuvastanud ühtegi mutatsiooni. Selliseid inimesi oli 49% ja neile saame tagasisidena öelda tõenäosused mitme fenotüübi kohta. Ligikaudu 41 protsendil inimestest saime määrata ainult ühe alleeli, mistõttu ka sellel inimgrupil ei saanud üheselt fenotüüpi määrata, vaid anda erinevatele fenotüüpidele esinemistõenäosusi.

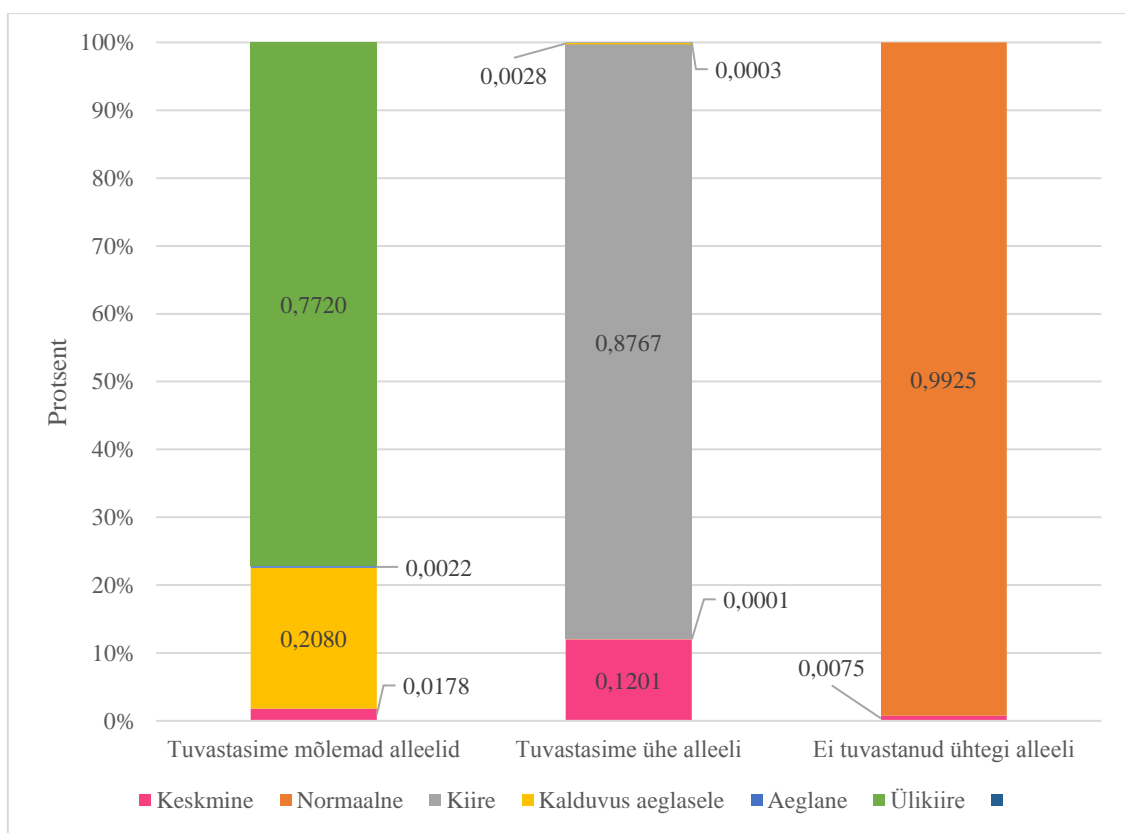
3.1 Alleelisagedused

Meie populatsioonis esines summaarselt enim alleele *2 ja *17, vastavalt 14 ja 84 protsenti. Veel leidis populatsioonis *3, *5, *8, *9, *10, *13 ja *14 alleele, kuid neid oli kokku ligikaudu 2% kogu valimist.

3.2 Fenotüüpide sagedus populatsioonis

Fenotüüpide sagedusi eelnevalt toodud (joonis 5) inimgruppide lõikes iseloomustab joonis 6 ja tabel 7.

Vasakpoolsest tulbast näeme, et suutes üheselt määrata mõlemad alleelid, saime tuvastada, et ülikiire metabolismiga inimesi oli selles grupist 77%, aeglasele metabolismile kalduvaid inimesi oli 20 % ja keskmise ning aeglase metabolismiga summaarselt vähem kui kolm protsenti populatsioonist.



Joonis 6. CYP2C19 Fenotüübi sagedus populatsioonis

Teisest tulbast näeme, et neile, kellel me ei tuvastanud ühtegi mutatsiooni, saime ennustada, et 99% sellistest inimestest on normaalse metabolismiga ja ainult üks protsent tasase metabolismiga. Ühtlasi teame seeläbi, et inimestel, kellel me ei suuda ühtegi mutatsiooni tuvastada, pole kindlasti ülikiiret ega aeglast metabolismi.

Kolmandast tulbast näeme, et inimestel, kellel tuvastasime üheselt ainult ühe mutatsiooni, on 87 protsenti tõenäosusega kiire metabolism ja 12 protsendil keskmine metabolism. Kolmanda

grupi inimeste ravimisoovitused ei ole siiski kõigile samad, vaid sõltuvad sellest, missugune haplotüüp neil esiti tuvastati. Seda jaotust on näidatud järgmises peatükis.

3.3 Fenotüübis ennustuste detailsed tõenäosused.

Lähtudes eelpool mainitust, siis määrates ainult ühe haplotüübi, sõltub ennustus konkreetsest alleelist. Tabelist 7 on näha, et iga alleeli puhul on fenotüübi tõenäosused erinevad. Näiteks *13 haplotüübiga on inimene kindlasti normaalse metabolismiga, samas tuvastades ainult *17 alleeli, teame, et metabolism on 99% tõenäosusega kiire

Tabel 7. Ennustuste detailsed tõenäosused

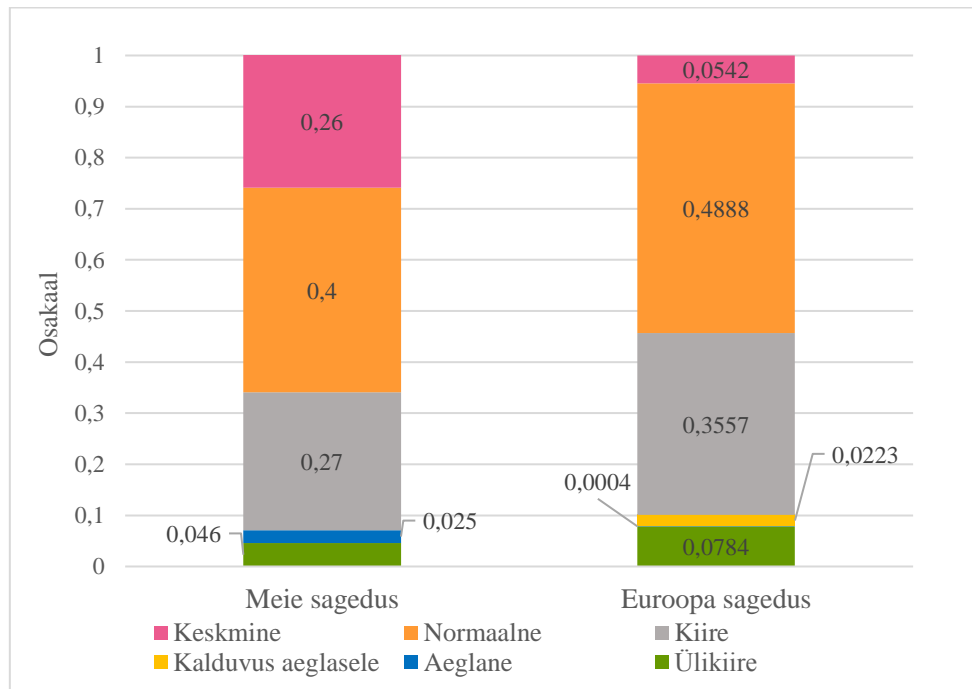
		Aeglane	Kalduvus aeglaseks	Keskmine	Normaalne	Kiire	Ülikiire
Kindel fenotüüp		0,00216	0,207995678	0,0178282			0,77202
Üks haplotüüp							
	*3	0,00303		0,996974			
	*17		0,00302596	0,0031852		0,99379	
	*5	0,00303		0,996974			
	*2	0,00303		0,996974			
	*9		0,996805112	0,0031949			
	*13				1		
	*8			1			
	*10						
	*14						
Pole ühtegi				0,0074543	0,9925457		

Ühtlasi näeme tabelist, et kui me ei suuda inimesel ühtegi mutatsiooni tuvastada, siis on ta 99 protsendi tõenäosusega normaalne metaboliseerija kalduvusega tasasele metabolismile. Oluline on siinkohal tähele panna, et sellisel juhul me teame, et inimene pole kindlasti kiirenenud metabolismiga ning see tähendab, et tal pole tarvis ravimiannust suurendada. Ühtlasi on kasulik teada, et geenidoonor ei kuulu äärmusesse (ülikiire), sest liiga kiire metabolismiga võivad kaasneda muud terviseprobleemid [11].

Kindla fenotüübi määramisel pole tegemist enam ennustustega, vaid sagedustega. Jooniselt näeme, et 77% sellest inimgrupist moodustasid ülikiire metabolismiga inimesed ja 20,7% kaldusid aeglasele metabolismile.

3.4 Algoritmitulemuste ja Euroopa fenotüübisageduste erinevused

Joonis 7 võrdleb fenotüüpide jaotust varem avaldatud kirjanduses ja meie populatsioonis. Jaotused on enamvähem sarnased, kuid võrreldes meie populatsiooniga on Euroopa normaalse fenotüübi sagedus 8 protsendi võrra väiksem ja seega moodustab 40% kogu populatsioonist. Kiiret metabolismi on Euroopas 27%, kuid keskmist on võrreldes meie populatsiooniga ligikaudu viis korda rohkem ja seega 26 protsenti. Samas pole Euroopa sagedustes märgitud aeglasele kalduvat metabolismi, kuid meil on väga vähe aeglast metabolismi.



Joonis 7. Meie ja Euroopa populatsioonide fenotüüpide summaarne jaotus

Meie populatsiooni ja Euroopa sageduste erinevuste põhjuseks on tõenäoliselt asjaolu, et meie ei tuvastanud ühegi inimese seas *4A ega *15 alleele, mis PharmGKB andmetel peaks olema Euroopas esindatud. Samas kõnealused alleelid on olulised keskmise metabolismi määramiseks.

4 Arutelu

Praegu kasutusel olevad geenitestid on ülemäära jäigad ega arvesta puuduliku geeniinformatsiooniga. Käesoleva töö raames loodud meetod on dünaamilisem ja kohandatud geeni spetsiifika ning haplotüüpide iseärasusega.

Tulemustest selgus, et kuna algoritmi tulemuste jaotus ja Euroopa fenotüüpide sagedused on sarnased, siis järelikult on algoritm meie populatsioonile rakendatav. Euroopa sagedustabelis ei olnud märgitud fenotüübina aeglasele kalduvat metabolismi, kuid umbes sama suur fenotüübijaotus oli meil aeglase metabolismiga. Seega esineb võimalus, et PharmGKB tabelis liigitati tulemus lihtsalt aeglase fenotüübi alla.

Ühtlasi on ilmne, et algoritm on suuresti sõltuv algandmete õigsusest. Kui PharmGKB andmetabelid ei ole korrektsed, siis on ka ravimisoovitused ebatäpsed. Seepärast on oluline enne meetodi rakendamist võrrelda populatsiooni alleelisagedusi PharmGKB tabelites toodud alleelisagedustega, sest enne tuleb veenduda, et meie populatsiooni alleelisagedused ja Euroopa omad (tabel 3) on sarnased. Teoreetiliselt esineb võimalus, et uuritav populatsioon on nii eripärane, et Euroopa alleelisagedusi ei tohiks meetodis kasutada.

Samas on selge, et dünaamilisema algoritmiga saab ka täpsemat tagasisidet anda. Hulgal inimestel suudame määrata ainult ühe alleeli, kuid tänasel päeval ei saaks nemad konkreetse geeni osas üldse ravimisoovitusi. Tulemustest oli näha, et tegelikult saab paljudele geenidonoritele ka sellisel puhul infot jagada. Tähtis aspekt tõenäosuslike ravimisoovituste puhul on asjaolu, et kuigi me ei saa konkreetselt öelda, et inimesel on näiteks kiire ainevahetus, küll aga pruugime mainida, et tal kindlasti ei ole aeglane metabolism. Selline tagasiside omab juba suurt sisulist tähendust, sest teadmine, et inimesel pole aeglane metabolism tähendab omakorda, et väiksemaid ravimidoose ei ole mõistlik tarbida.

Kõnealune algoritm ja valdkond vajab kindlasti jätkuvalt uurimist, sest lahendust tuleks kindlasti üldistada ka teiste geenide tarvis.

5 Kokkuvõte

Geeniinfo põhjal ravimisoovituste tegemine on muutumas kliinilises praktikas küll üha tavapärasemaks, kuid tehniliste nüansside tõttu ei ole siiski kasutusel nii laialdaselt, kui võiks. Suurimaks probleemiks on siin asjaolu, et inimese geeniinfo eraldamine kuluefektiivsel mikrokiibimeetodil ei mõõda jätkuvalt paljusid ravimisoovituste tegemiseks vajalikke geenimutatsioone. Selleks, et ravimisoovituste andmine oleks siiski võimalik, kirjeldatakse käesolevas töös osalist geeniinfot kasutaval tõenäosustel põhinevat meetodit, mis kasutab patsiendi kohta olemasolevat geeniinfot ja geenivariantide varem publitseeritud esinemissagedusi. Näitena kasutatakse geeni CYP2C19 ja populaarse mikrokiibiga Illumina Global Screening Array mõõdetud Eesti geenidonorite andmeid, millel saadud tulemused langevad kokku varem raporteeritud ravimisoovituste sagedustega. Põhjalikumaks analüüsiks vajaks teema kindlasti edasist uurimist.

Kasutatud allikad

1. Scott, A-S. Personalizing medicine with clinical pharmacogenetics, 2011, külastatud 12.02.2018,
https://www.researchgate.net/publication/51807462_Personalizing_medicine_with_clinical_pharmacogenetics
2. Drew, L. The right drug for you, 2016, külastatud 12.02.2018,
<https://www.nature.com/articles/537S60a>
3. Relling, MV. Gardner EE, Sandborn, W,J. Evans WE, Clinical Pharmacogenetics Implementation Consortium Guidelines for Thiopurine Methyltransferase Genotype and Thiopurine Dosing, 2011, külastatud 12.02.2018,
<https://www.ncbi.nlm.nih.gov/pubmed/21270794>
4. Goldstein, D-B. Tate, S-K. Sisodiya S-M. Pharmacogenetics goes genomic. 2003, 12.02.2018. <https://www.ncbi.nlm.nih.gov/pubmed/14631354>
5. Dong, Y. Xiao, Huasheng. Wang, Qi. Zhang Chunxiu. Analysis of genetic variations in CYP2C9, CYP2C19, CYP2D6 and CYP3A5 genes using oligonucleotide microarray, 2015. külastatud 24.04.2018,
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4694416/>
6. Ingelman-Sundberg, M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. 2004. Külastatud 15.02.2018, <https://www.nature.com/articles/6500285>
7. GSA tehniline info, külastatud 2.05.2018,
http://www.glimdna.org/assets/gsa_datasheet_2016.pdf
8. PharmGKB tabelid, külastatud 16.03.2018,
<https://www.pharmgkb.org/page/pgxGeneRef>

9. Human Genome Diversity Project, külastatud 14.03.2018,
<http://www.hagsc.org/hgdp/>

10. CYP2C19 geenitabelid, külastatud 12.02.2018
<https://www.pharmgkb.org/page/cyp2c19RefMaterials>

11. Fast metabolism, külastatud 8.05.2018, <https://www.livescience.com/32362-what-does-fast-metabolism-mean.html>

Lisa 1 – CYP2C19 diplotüüpide sagedustabel

Diplo tüüp	Esinemissagedus
*1/*1	0,38937600
*1/*2	0,09110400
*1/*3	0,00386256
*1/*4	0,00173472
*1/*5	0,00001579
*1/*6	0,00118560
*1/*8	0,00159744
*1/*9	0,00015600
*1/*13	0,00062400
*1/*14	0,00000208
*1/*15	0,00124800
*1/*17	0,13291200
*2/*2	0,02131600
*2/*3	0,00090374
*2/*4	0,00040588
*2/*5	0,00000369
*2/*6	0,00027740
*2/*8	0,00037376
*2/*9	0,00003650
*2/*13	0,00014600
*2/*14	0,00000049
*2/*15	0,00029200
*2/*17	0,03109800
*3/*3	0,00003832
*3/*4	0,00001721
*3/*5	0,00000016
*3/*6	0,00001176
*3/*8	0,00001585
*3/*9	0,00000155
*3/*13	0,00000619
*3/*14	0,00000002
*3/*15	0,00001238
*3/*17	0,00131847
*4/*4	0,00000773
*4/*5	0,00000007
*4/*6	0,00000528
*4/*8	0,00000712
*4/*9	0,00000070
*4/*13	0,00000278
*4/*14	0,00000001
*4/*15	0,00000556
*4/*17	0,00059214
*5/*5	0,00000000
*5/*6	0,00000005
*5/*8	0,00000006
*5/*9	0,00000001
*5/*13	0,00000003
*5/*14	0,00000000

*5/*15	0,00000005
*5/*17	0,00000539
*6/*6	0,00000361
*6/*8	0,00000486
*6/*9	0,00000048
*6/*13	0,00000190
*6/*14	0,00000001
*6/*15	0,00000380
*6/*17	0,00040470
*8/*8	0,00000655
*8/*9	0,00000064
*8/*13	0,00000256
*8/*14	0,00000001
*8/*15	0,00000512
*8/*17	0,00054528
*9/*9	0,00000006
*9/*13	0,00000025
*9/*14	0,00000000
*9/*15	0,00000050
*9/*17	0,00005325
*13/*13	0,00000100
*13/*14	0,00000000
*13/*15	0,00000200
*13/*17	0,00021300
*14/*14	0,00000000
*14/*15	0,00000001
*14/*17	0,00000071
*15/*15	0,00000400
*15/*17	0,00042600
*17/*17	0,04536900

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, **Ainika Adamson**,

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

„**Meetod tõenäosusliku ravimisoovituse andmiseks osalise geeniinfo põhjal**“,
mille juhendaja on Sulev Reisberg,

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

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

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

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

Tartus, **14.05.2018**