

UNIVERZITET U BEOGRADU
MEDICINSKI FAKULTET

Ivan R. Skadrić

Karakteristike distribucije funkcionalnih varijanti
gena iz superfamilije citohroma P450 u
stanovništvu Srbije

doktorska disertacija

Beograd, 2020

UNIVERSITY OF BELGRADE
FACULTY OF MEDICINE

Ivan R. Skadrić

Cytochrome P450 gene superfamily functional
variants distribution characteristics in Serbian
population

Doctoral Dissertation

Belgrade, 2020

Mentor:

dr Oliver Stojković, redovni profesor, Medicinski fakultet, Univerzitet u Beogradu

Članovi komisije:

dr Suzana Cvjetićanin, redovni profesor, Medicinski fakultet, Univerzitet u Beogradu

dr Branislava Medić, docent, Medicinski fakultet, Univerzitet u Beogradu

dr Viktorija Dragojević-Simić, redovni profesor, Medicinski fakultet, Univerzitet odbrane u Beogradu

Najiskrenije i sa najdubljim poštovanjem, želim da irazim svoju zahvalnost radnicima u Republici Srbiji, koji stvaraju preduslove za svako istraživanje u našoj zemlji i kojima dugujem svoje znanje.

Karakteristike distribucije funkcionalnih varijanti gena iz superfamilije citohroma P450 u stanovništvu Srbije

Sažetak

Citohrom P450 genska superfamilija se u ljudskom genomu sastoji od 57 gena, koji se na osnovu aminokiselinske sličnosti razvrstavaju u 44 genske podfamilije i 18 familija. Geni koji pripadaju ovoj familiji se prevode u enzime sa raznovrsnim funkcijama, koji se nazivaju CYP enzimi i eksprimiraju se u različitim tkivima u ljudskom organizmu. Među najznačajnijim funkcijama CYP enzima su sposobnost katalize oksidacije, redukcije i hidrosilacije, zbog čega se ovi enzimi nazivaju i monooksigenazama i poznato je da učestvuju u fazi I metabolizma ksenobiotika. U širokom spektru ksenobiotika koji se metabolišu CYP enzimima se nalazi i veliki broj lekova koji su u kliničkoj upotrebi, što omogućava proučavanje uticaja genetičke varijabilnosti na farmakodinamiju i farmakokinetiku lekova. Uticaj genetičkih promena u *CYP* genima se prvenstveno ogleda u izmeni aminokiselinske sekvence, a posledično i u smanjenoj funkciji CYP enzima. Pored toga, izmene u DNK sekvenci koja reguliše ekspresiju *CYP* gena, mogu dovesti do umanjena ili povećanja kapaciteta CYP enzima da metabolišu lek. Izmene u *CYP* genima, koje najčešće dovode do promena u funkciji CYP enzima, se nazivaju funkcionalne varijante i mogu biti tačkaste izmene ili inserciono-delecione varijante u DNK sekvenci. Na osnovu genotipizacije funkcionalnih varijanti *CYP* gena, moguće je odrediti i funkcionalnost enzima, kao i metabolički kapacitet čovekovog organizma prema leku za koji je poznato da zavisi od funkcionalnosti tog enzima. Cilj je precizniji odabir vrste i doze leka, kao i izbegavanje neželjene reakcije na lek. Ovakav pristup se naziva farmakogenetikom i predstavlja osnovu na kojoj se razvija personalizovana medicina i molekularna autopsija. Od svih *CYP* genskih familija koje su bitne za metabolizam lekova, familije *CYP1*, *CYP2* i *CYP3* imaju najvećeg uticaja i najbolje su istražene, a unutar njih postoje genetičke varijante za koje je dokumentovano da dovode do izmena u metabolizmu lekova. Međutim, ciljana primena molekularno bioloških metoda za genotipizaciju zahteva prethodno znanje o informativnosti *CYP* lokusa, koja može u velikoj meri zavistiti od učestalosti genetičkih varijanti u nekoj populaciji, što nameće potrebu da se istraživanje *CYP* varijanti sprovede na populaciono genetičkom nivou. Kako je u ovom trenutku informacija o karakteristikama raspodele učestalosti *CYP* funkcionalnih varijanti u stanovništvu Srbije oskudna i nesistematična, ova studija je imala za cilj da u uzorku koji će verodostojno predstaviti opštu populaciju Srbije, odredi učestalosti alela i genotipova 11 varijanti u *CYP* genima: *CYP1A1* (rs4646903, rs1048943), *CYP2C9* (rs1057910, rs1799853), *CYP2C19* (rs12248560, rs4244285), *CYP2D6* (rs3892097, rs1065852, rs28371706, rs28371725), *CYP3A4* (rs2740574). Pored određivanja učestalosti funkcionalnih varijanti, jedan od ciljeva je bio i da se analizira genetička struktura stanovništva pomoću formiranja uzorka od osoba koje pripadaju potencijalno najdiferenciranijoj subpopulaciji i upotrebom selektivno neutralnih mikrosatelitskih lokusa, kao i većeg broja algoritama za detekciju genetičke struktuiranosti. Radi povećanja reprezentativnosti uzorka opšte populacije, odabran je metod stratifikovanog uzorkovanja DNK izolata nesrodnih osoba prema mestu prebivališta iz pet regiona Srbije: Severna Srbija (Vojvodina), okrug Beograda, Zapadna Srbija, Centralna Srbija, Južna i Istočna Srbija. Ukupan uzorak je iznosio 550 izolata, dok je u svojstu uzorka iz potencijalno najdiferenciranije subpopulacije, odabrano 60 DNK izolata od osoba romske nacionalne pripadnosti. Svi uzorci su genotipizirani primenom lančane reakcije polimeraze u realnom vremenu (rs1057910, rs1799853,rs12248560, rs4244285, rs3892097, rs1065852, rs28371706, rs28371725, rs2740574) i analizom polimorfizama dužine restrikcioniha fragmenata (rs4646903, rs1048943). Na podacima uspešno genotipiziranih ispitanika sprovedena je bioinformatička analiza upotrebom algoritma koji se zasnivaju na Bejzovom pristupu statističkog zaključivanja, redukciji dimenzionalnosti podataka, hijerarhijskom klasterovanju, kao i primeni teorije mreža u analizi podataka. Algoritmi koji su korišćeni za detekciju sturktuiranosti populacije su BAPS,

STRUCTURE, PCA, tSNE, DBSCAN i IGRAPH. Prema rezultatima ove studije, učestalosti funkcionalnih varijanti u CYP genima su sledeće: (rs1057910) 7.4%, (rs1799853) 6.2%, (rs12248560) 17.3 % ,(rs4244285) 21.1%, (rs3892097) 8.8%, (rs1065852) 0%, (rs28371706) 2.2%, (rs28371725) 11.7%, (rs2740574) 8%, (rs4646903) 16.3%, (rs1048943) 22.2%. Prema rezultatima bioinformatičke analize genetičke struktuiranosti populacije, nije bilo većih odstupanja od uniformne raspodele alelskih i genotipskih učestalosti između stanovnika pet testiranih regiona, niti su učestalosti CYP funkcionalnih varijanti značajno različite između romske subpopulacije i ostatka opšte populacije. Od algoritama upotrebljenih za detekciju genetičke struktuiranosti, BAPS i tSNE su pokazali zadovoljavajuće performanse i jednostavnost za implementaciju na genetičkim podacima. Imajući u vidu učestalosti u opštoj populaciji Srbije, kao i uticaj na metabolizam lekova, za populaciono specifični panel koji bi se koristio za farmakogenetičko testiranje u personalizovanoj medicini ili molekularnoj autopsiji, predlažemo set od sedam lokusa, rs3892097, rs1065852, rs28371725, rs1799853, rs1057910, rs4244285, rs12248560 u genima CYP2C9, CYP2C19 i CYP2D6.

Ključne reči: alelske učestalosti, citohrom 450, farmakogenetika, populaciona genetika, genetička struktuiranost, molekularna autopsija.

Naučna oblast: Medicina

Uža naučna oblast: Molekularna medicina

Cytochrome P450 gene superfamily functional variants distribution characteristics in Serbian population

Abstract

The cytochrome P450 gene superfamily in the human genome consists of 57 genes, which are classified into 44 gene subfamilies and 18 families based on amino acid similarity. Genes belonging to this family are translated into enzymes with various functions, called CYP enzymes and are expressed in various tissues in the human body. Among the most important functions of CYP enzymes is the ability to catalyze oxidation, reduction and hydroxylation reactions, which is why these enzymes are also called monooxygenases and are known to participate in phase I metabolism of xenobiotics. The wide range of xenobiotics that are metabolized by CYP enzymes also includes a large number of drugs that are in clinical use, which enables the study of the influence of genetic variability on the pharmacodynamics and pharmacokinetics of drugs. The influence of genetic changes in *CYP* genes is primarily reflected in the change of the amino acid sequence, and consequently in the reduced function of CYP enzymes. Additionally, changes in the DNA sequence, that regulate *CYP* gene expression, may lead to a decrease or increase in the capacity of CYP enzymes to metabolize a drug. Changes in *CYP* genes, which most often lead to changes in the function of the CYP enzymes, are called functional variants and can be point changes or insertion-deletion variants in the DNA sequence. Based on the genotyping of functional variants of a CYP gene, it is possible to determine the functionality of the enzyme, as well as the metabolic capacity of the human body for a drug known to depend on the functionality of the enzyme. The goal is to accurately choose the type and dose of the drug, as well as to avoid an adverse reaction to it. This approach is called pharmacogenetics and is the basis for development of personalized medicine and molecular autopsy. Of all the CYP gene families important for drug metabolism, the *CYP1*, *CYP2*, and *CYP3* families have the greatest impact, are most researched and encompass genetic variants that have been documented to lead to changes in drug metabolism. However, the targeted application of molecular biology methods for genotyping requires prior knowledge of the informativeness of the *CYP* locus, which may largely depend on the frequency of genetic variants in a population, necessitating research on *CYP* variants at the population genetic level. As information on the characteristics of the frequency distribution of CYP functional variants in the population of Serbia is scarce and unsystematic, this study aimed to determine the frequencies of alleles and genotypes of 11 variants in CYP genes in a sample that will credibly represent the general population of Serbia: *CYP1A1* (rs4646903, rs1048943), *CYP2C9* (rs1057910, rs1799853), *CYP2C19* (rs12248560, rs4244285), *CYP2D6* (rs3892097, rs1065852, rs28371706, rs28371725), *CYP3A4* (rs2740574). In addition to determining the frequency of functional variants, one of the goals was to analyze the genetic structure of the population by forming a sample of people belonging to the potentially most differentiated subpopulation and using selectively neutral microsatellite loci, as well as a number of algorithms for detecting genetic structure. In order to increase the representativeness of the general population sample, the chosen method was the one of stratified sampling of DNA isolates of unrelated persons according to the place of residence from five regions of Serbia: Northern Serbia (Vojvodina), Belgrade district, Western Serbia, Central Serbia, Southern and Eastern Serbia. The total sample was 550 isolates, while 60 DNA isolates of persons of Roma ethnicity were selected as a sample from the potentially most differentiated subpopulation. All samples were genotyped using real-time polymerase chain reaction (rs1057910, rs1799853, rs12248560, rs4244285, rs3892097, rs1065852, rs28371706, rs28371725, rs2740574) and analysis of restriction fragment length polymorphism (rs4646903, rs1048943). Bioinformatics analysis was performed on the data of successfully genotyped subjects using algorithms based on Bayes' approach to statistical inference, dimensionality reduction, hierarchical clustering, as well as network theory. The algorithms used to detect population

structure are BAPS, STRUCTURE, PCA, tSNE, DBSCAN and IGRAPH. According to the results of this study, the frequencies of functional variants in *CYP* genes are as follows: (rs1057910) 7.4%, (rs1799853) 6.2%, (rs12248560) 17.3%, (rs4244285) 21.1%, (rs3892097) 8.8%, (rs1065852) 0%, (rs28371706) 2.2%, (rs28371725) 11.7%, (rs2740574) 8%, (rs4646903) 16.3%, (rs1048943) 22.2%. According to the results of bioinformatics analysis of genetic structure of the population, there were no major deviations from the uniform distribution of allelic and genotypic frequencies among the inhabitants of the five tested regions, nor were the frequencies of *CYP* functional variants significantly different between the Roma subpopulation and the rest of the general population. Of the algorithms used to detect genetic structuring, BAPS and tSNE have shown satisfactory performance and ease of implementation on genetic data. Having in mind the frequency in the general population of Serbia, as well as the impact on drug metabolism, for a population-specific panel that would be used for pharmacogenetic testing in personalized medicine or molecular autopsy, we propose a set of seven loci rs3892097, rs1065852, rs28371725, rs1799853, rs1057910, rs4244285, rs12248560 in *CYP2C9*, *CYP2C19* and *CYP2D6* genes.

Keywords: allele frequency, cytochrome P450, pharmacogenetics, population genetics, genetic structure, molecular autopsy.

Scientific field: Medicine

Scientific subfield: Molecular Medicine

Sadržaj

1.	UVOD	1
1.1.	Otkriće citohroma P 450.....	1
1.1.1.	Otkriće uloge citohroma P450 u metabolizmu lekova	3
1.1.2.	Otkriće diverziteta citohroma P450	4
1.2.	Lokalizacija CYP450 enzima.....	6
1.3.	Funkcija CYP450 enzima	6
1.3.1.	Monoooksigenazna funkcija.....	7
1.3.2.	Hidroksilazna i druge funkcije.....	7
1.4.	Struktura i diverzitet citohroma P450	7
1.5.	Genetička struktura i nomenklatura <i>CYP</i> gena	9
1.5.1.	Variranje u broju kopija gena (CNV)	10
1.5.2.	CYP450 genska superfamilija u ljudskom genomu.....	10
1.5.2.1.	CYP1 familija	12
1.5.2.2.	CYP2 familija	12
1.5.2.3.	CYP3 familija	14
1.6.	Citohrom P450 u personalizovanoj medicini.....	15
1.6.1.	CYP1A1.....	17
1.6.2.	CYP2C9.....	17
1.6.3.	CYP2C19.....	18
1.6.4.	CYP2D6.....	19
1.6.5.	CYP3A4.....	20
1.7.	Citohrom P450 u sudsko-medicinskoj primeni.....	21
1.8.	Teorija i analiza genetičke strukture populacije	24
1.8.1.	Ocena genetičke strukturiranosti populacije na osnovu matematičkih modela.....	27
1.8.1.1.	Osnovne definicije u verovatnoći i statistici.....	28
1.8.1.2.	Bejzova statistika	28
1.1.1.1.	Algoritmi za detekciju genetičke struktuiranosti	46
2.	Ciljevi istraživanja	59
3.	Materijal i metode	60
3.1.	Mesto i period istraživanja	60
3.2.	Ispitanici	60
3.3.	Uzorci biološkog materijala	60
3.4.	Klasifikacija i veličina uzorka.....	61
3.5.	Instrumenti merenja.....	63
3.6.	Statističke i bioinformatičke analize	69

3.6.1.	Statistička testiranja i analiza molekulske varijanse	69
4.	Rezultati	72
4.1.	Rezultati genotipizacije	72
4.2.	Rezultati analize molekulske varijanse	76
4.3.	Rezultati algoritama za detekciju genetičke struktuiranosti.....	78
5.	Diskusija.....	89
6.	Zaključci.....	104
7.	LITERATURA.....	105

1. UVOD

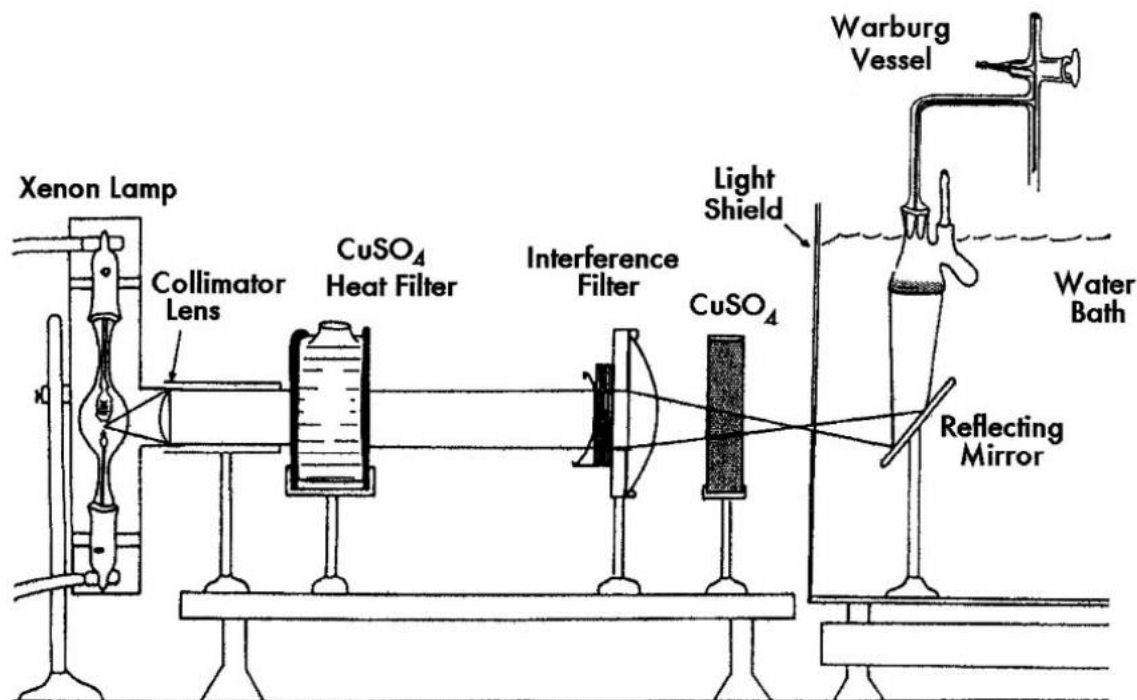
1.1. Otkriće citohroma P 450

Iako je šezdesetih godina dvadesetog veka skovan naziv citohrom P450, proučavanje ove grupe enzima korespondira sa deset godina ranije započetim istraživanjima mikrozomalnih ostataka lizata tkiva jetre i metabolizma aminazo boje (N-Ndimethyl-4-aminoazobenzene. U Mekardl (eng McArdl) laboratoriji Univerziteta u Vinskonsinu, Mjular i Miler (Mueller and Miller) opažaju da jedna komponenta lizata hepatičnog tkiva sastavljena od nepoznatih partikula, poseduje neobično svojstvo da metaboliše aminazo boje (Estabrook 2003; MUELLER and MILLER 1949). Nedugo zatim, Akslrod (eng Axelrod) pokazuje da je metabolizam psihoaktivnih supstanci u mikrozomu zavisano od redukcije trifosfopiridinskog nukleotida (NADP) i prisustva kiseonika (AXELROD 1955; Estabrook 2003). Nezavisno od ostalih, Rajn i Engl (eng Ryan i Engel) ukazuju da je za hidroksilaciju progesterona na C21 ugljeniku u mikrozomima neophodan trifosfopiridinski nukleotid i atmosferski kiseonik (Estabrook 2003; RYAN and ENGEL 1957). Ovi pionirski eksperimenti su bili osnova, odnosno preduslov za niz daljih istraživanja koja su iznedrila novu klasu enzima i promenila shvatanje metabolizma kako stranih supstanci u organizmu (ksenobiotika), tako i jedinjenja koje sam organizam sintetise (endobiotika).

Tokom šeste i sedme decenije 20. veka, razvoj i poboljšanje tehnologije masene spektrometrije, spektrofotometrije i polarografije, omogućili su karakterizaciju metaloenzima. Zahvaljujući tom tehnološkom napretku, Britn Čens i Ron Vilijams (eng Brittan Chance, Ron Williams), u nameri da bolje razumeju oksidativnu fosforilaciju i transportni lanac elektrona, izolovali su mikrozomalnu frakciju iz jetre pacova i spektrofotometrijski okarakterisali hemoprotein citohrom *m*, kasnije poznat kao citohrom *b₅* (CHANCE and WILLIAMS 1954). U njihovim eksperimentima je, nakon dodavanja ugljen monoksida u spektrofotometarsku kivetu sa mikrozomima i redukujućim agensom NADH, opažen maksimum apsorbance na talasnoj dužini od 450 nm. Kako to nije bio cilj istraživanja Čansa i Vilijamsa, samo otkriće i nije bilo interesanto van konteksta ćelijskog disanja i kinetike elektrona u mitohondrijama. Međutim, ono je bilo interesantno mladom nemačkom fizikohemičaru Martinu Klingenbergu koji je tek stigao na postdoktorske studije u Čansovu laboratoriju. Klinenberg tada pažljivo reprodukuje Čansove eksperimente i opisuje ugljen monoksid vezujući pigment (kako je tada nazvan) sa maksimumom apsorbance na talasnoj dužini od 450 nm, što je kasnije izvedeno u skraćeni naziv P450 (Klingenberg 1958, 2003). Iako su rezultati Čansa, Vilijamsa i Klingenberga ukazivali da je koncentracija novootkrivenog P 450 duplo veća od citohroma *b₅*, nije se znalo da li je uopšte u pitanju metaloprotein, već da ovaj pigment vezuje ugljen monoksid i da je najverovatnije primalac/akceptor elektrona. Još jedna neobičnost vezana za ovaj pigment, je bila pojava crvenog pomeraja prilikom vezivanja CO za redukovani hemoprotein. Taj fenomen nije bio viđen do tada i nije se uklapao u uvraženo mišljenje da hemoproteini imaju plavi pomeraj prilikom vezivanja CO. Zbog svega toga, pigment 450 nije bilo moguće klastifikovati u neku od poznatih klasa enzima¹(Estabrook 2003).

¹ Crveni i plavi pomeraj su promene talasne dužine elektromagnetnom zračenja. Crveni pomeraj ili pomak je emitovanje svetlosti u delu spektra povećane talasne dužine, dok se plavi pomeraj odnosi na pomak u smeru manje talasne dužine.

Ranih šesdesetih godina intezivira se saradnja između biohemičara, fizikohemičara i kliničkih lekara pod okriljem udruženja Džon Morgan i to sa ciljem objedinjavanja fundamentalnih i kliničkih istraživanja. Tako započinje saradnja hirurga Dejvida Kupera i biohemičara Ronalda Estabruka (eng David Cooper, Ronald Estabrook), koji su uprkos nedostatku opreme, a samo zahvaljujući sopstvenoj domišljatosti i entuzijizmu, napravili spektar fotohemijske reakcije, a potom opisali i stehiometriju P 450 (slika 1)(Estabrook 2003; Klingenberg 2003). Tada je otkriveno da hemijska reakcija ovog enzima zahteva ekvivalentne količine kiseonika i NADPH, ali, što je još bitnije, otkrivena je nova klasa enzima koji učestvuju u oksidativnom metabolizmu hepatocita-oksigenaze (Cooper et al. 1965; COOPER, ESTABROOK, and ROSENTHAL 1963).



Slika 1 - Shematski prikaz opreme kojom su Kuper i Estabruk pokazali fotohemijsku aktivnost citohroma P 450. Vodeno kupatilo je napravljeno od akvarijuma za ribe, dok je ksenonska lampa snage 2.5 KW, uslovno rečeno, pozajmljena sa reflektora obližnjeg sportskog terena. Slika preuzeta iz A Passion for P450s, Estabrook, 2003.

Uloga pigmenta P450 kao monooksigenaze opisana je u istraživanjima u vezi sa sintezom hormona nadbubrežne žlezde. Naime, 1964. godine Harding sa saradnicima opisuje hidrosilaciju steroida u mikrozomima za koju je bio potreban kiseonik i P450 (Harding, Wong, and Nelson 1964).

Za to vreme, dve grupe naučnika u nezavisnim eksperimentima istražuju enzime koji sadrže gvožđe u aktivnom centru. Prva grupa u SAD, koju su predvodili Hašimoto, Jamata i Mejson, opisuje "microsomal Fex" protein za koji će se kasnije ispostaviti da je P450 (HASHIMOTO, YAMANO, and MASON 1962). U Japanu, druga grupa koju su predvodili Sato i Omura, reprodukuje eksperiment Klinenberga i uvodi manje modifikacije koje će pokazati hemoproteinsku prirodu P450. Upravo tada se prvi put pojavljuje naziv koji postoji i danas-citohrom P450, nastao kombinacijom grčkih reči (κύτος-kontejner, sadržalac i χρωμα-boja), slova P (pigment) i broja 450 (talasna dužina maksimuma apsorbance). Naziv citohrom je upotrebljavan i ranije, ali ne u kontekstu citohroma P450, već kao citohrom c reduktaza i citohrom b₅ (HOGEBROOM 1949; T. OMURA and SATO 1962; Tsuneo OMURA 2011)

Tabela 1 : hronološki prikaz otkrića CYP enzima u različitim taksonima.				
Godina	Organizam	Tkivo	Organela	Referenca
1958	Pacov, Zec	jetra	Mikrozom	(Klingenberg 1958)
1964	Pacov	nadbubrežna žlezda	Mitohondrija	(Harding, Wong, and Nelson 1964)
1964	Kvasac, gljiva	vanćelijski ekstrakt		(Lindenmayer and Smith 1964)
1967	<i>Rhizobium</i> , bakterija			(Appleby 1967)
1968	<i>Pseudomonas putida</i> , bakterija			(Katagiri, Ganguli, and Gunsalus 1968)
1969	<i>Echinocystis macrocarpa</i> , biljka			(Murphy and West 1969)
1973	<i>Cunninghamella bainieri</i> , gljiva	micela		(Ferris et al. 1973)
1996	<i>Arhea</i>			(R. L. Wright et al. 1996)

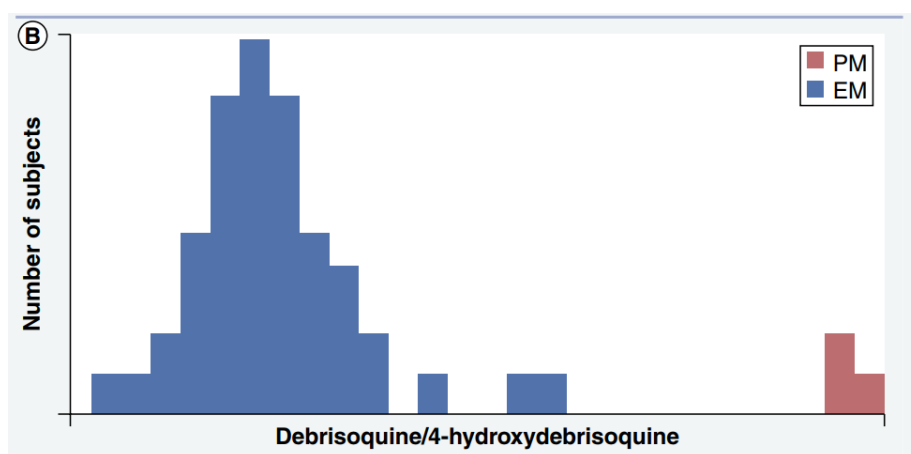
1.1.1. Otkriće uloge citohroma P450 u metabolizmu lekova

Nedugo nakon otkrića Estabruka i Kupera, nemački farmakolog Herbert Remer (Herbert Remmer) nakon niza eksperimenata opaža da koncentracija P 450 u mikrozomima jetre raste nakon tretmana lekovima, odnosno da postoji indukcija sinteze P 450 (Estabrook 2003; Remmer and Merker 1963, 1965). Ovo je bio jasan znak asocijacije citohroma P450 sa metabolizmom lekova. Međutim, bilo je potrebno upoznati mehanizam i razjasniti čitavu biohemijsku pozadinu učešća ovog enzima u metabolizmu lekova.

Nesvesni identiteta faktora koji utiče na metabolizam lekova, Abrams (William Abrams) sa još nekoliko kliničkih lekara ispituje novi antihipertenziv debrizokin na grupi pacijenata koja je pokazala veliku varijabilnosti u doziranju leka (ABRAMS et al. 1964). Isti lek je ispitivan dve godine kasnije i tom prilikom su opažena varijabilnost u doziranju i neželjena dejstva kod 3.5% ispitanika (Athassiadis et al. 1966).

Kako će se kasnije ispostaviti, najjasnija potvrda o ulozi CYP450 u metabolizmu lekova dolazi 1973. godine na St Mary's Hospital Medical School u Londonu. Grupa vrlo iskusnih istraživača iz oblasti metabolizma lekova, koristi radioaktivno obeleženi debrizokin C¹⁴ na animalnim modelima prateći koncentraciju nemetabolisanog leka u urinu i promene krvnog pritiska. Rezultati su ukazivali na dve očigledne pojave: varijabilnost u koncentraciji nemetabolisanog leka i pada krvnog pritiska, kao i njihovu međusobnu povezanost (Angelo et al. 1975; Baumann 2014). Ohrabreni prvim rezultatima na animalnim modelima, četiri istraživača ponavljaju eksperiment, ali ne na animalnim modelima, već, dobrovoljno uzimajući radioaktivni debrizokin, eksperiment ponavljaju na sebi. Slučaj je hteo da se u tako maloj grupi volontera zadese dvoje koji bi u današnjoj terminologiji bili klasifikovani kao 'loš metabolizer' i 'ekstenzivni metabolizer'. Naime, Smit (Smith), rukovodilac istraživačkog tima je i sam bio volonter u eksperimentu, kao loš metabolizer je imao najvišu koncentraciju nemetabolisanog debrizokina i to u meri koja je statistički značajno odstupala od slučajnog variranja (Angelo et al. 1976). Više nije bilo sumnje u postojanje faktora koji utiču na metabolizam lekova, no trebalo je spoznati njihov identitet u fizičkom svetu.

Svestan genetičkih osnova enzimologije, Smit proverava svoju porodičnu istoriju, ali i sprovodi dalja merenja među novim volonterima i identifikuje još osoba čiji je metabolizam debrizokina u velikoj meri bio sličan njegovom. Proučavanjem porodične istorije novih volontera, uočava genetičku pozadinu metabolizma lekova i recesivno nasleđivanje osobine kod loših metabolizera (Angelo et al. 1976; Baumann 2014; Mahgoub et al. 1977). Nakon svih otkrića koje ova grupa beleži, dolazi i ono krucijalno koje je rezultat naučne slobode i izuzetnog razumevanja statističke analize. Smit sa saradnicima histogramski projektuje odnos 4-hidroksidebrizokina i nemetabolisanog debrizokina. Na histogramu je uočena trimodalna raspodela (slika 2), a kako je u statistici poznato da modovi nastaju delovanjem faktora na raspodelu podataka, zaključak je bio da postoje tri faktora, odnosno tri vrste metabolizera koje se jasno ističu (Baumann 2014).



Slika 2 - Histogram MR. Smith i saradnici. Izvor: Baumann, 2014.

1.1.2. Otkriće diverziteta citohroma P450

Krajem šezdesetih godina, jedno otkriće u molekularnoj biologiji prokariota je prošlo nezapaženo od strane kliničkih i drugih istraživača koji su se prvenstveno bavili metabolizmom lekova u višim taksonima. U gram negativnim bakterijama *Rhizobium* je otkriven enzim sa maksimumom apsorbance na 450nm, što će se ispostaviti kao prvi ikada opisani CYP450 u prokariotima (Appleby 1967). Samo godinu dana kasnije dolazi još jedna potvrda o postojanju CYP450 enzima u prokariotima, u *Pseudomonas putida* bakterijama je pronađen enzim CYP450cam².

Prvu kompletnu sekvencu gena za CYP450 iz jetre pacova, publikovala je grupa japanskih autora 1983. godine (Mizukami et al. 1983), da bi tri godine kasnije bilo poznato čak 67 sekvenci pretežno iz eukariotskih organizama (D. Nebert 1987).

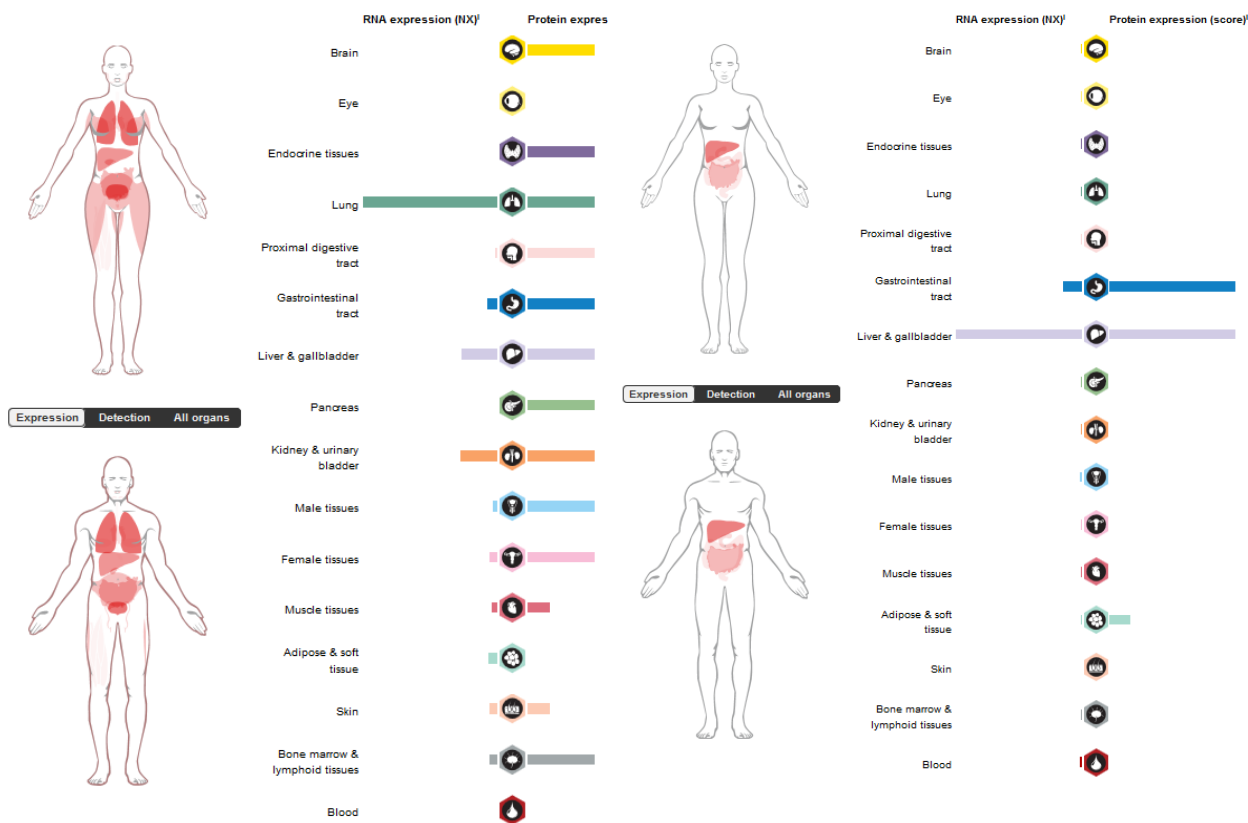
Otkriće citohrom P450 enzima u prokariotima i velike razlike u aminokiselinskoj sekvenci nametnuli su potrebu za sveobuhvatnim shvatanjem raznovrsnosti, ali i standardizacijom klasifikacije ovih enzima. Frenk Gonzales 1987. rezimira do tada publikovane radove i zaključuje da postoji bar 10 genskih familija CYP450 i da njih moraju činiti evolutivno stari geni. On zatim predlaže i sistem klasifikacije po kojem CYP450 enzimi sa manje od 36% sličnosti pripadaju istoj familiji, dok sekvence sa više od 68% sličnosti unutar familije pripadaju jednoj podfamiliji (D. Nebert 1987).

² sufiks *cam* je izveden od kamfora, smolastog sastojka kamforog drveta koji se nalazio u medijumu iz kojeg je izolovana bakterija *Pseudomonas putida*.

Decidne parametre, odnosno procenete za klasifikaciju *CYP* gena, i sam Gonzales je označio kao donekle arbitrarne, naglašavajući da ne moraju nužno odlikavati biološke entitete (D. Nebert 1987).

1.2. Lokalizacija CYP450 enzima

Unutar eukariotske ćelije, citohrom P450 se pretežno nalazi u vidu enzima vezanih za endoplazmin retikulum, Golidžijev aparat i mitohondrijsku membranu, dok je u prokariotskim ćelijama prisutan u vidu solubilnih enzima (Ortiz de Montellano 2015).



Slika 3 - Primer tkivno-specifične ekspresije CYP1A1 i CYP2D6: Izvor: <https://www.proteinatlas.org/ENSG00000140465-CYP1A1/tissue>

U ljudskom organizmu, CYP enzimi se najviše ekspimiraju u tkivu jetre i tankih creva, međutim postoje i forme koje su karakteristične za pluća, korteks nadbubrežne žlezde, mozak, ždrelo itd. Farmakološki najinteresantnijih CYP enzimi (CYP1A2, CYP2A6, CYP2B6, CYP2C9, CYP2C19, CYP2D6 i CYP3A4) se ekspimiraju u jetri i povezani su sa njenom funkcijom u metabolizmu ksenobiotika, ali i sintezom žučnih soli (Fanni et al. 2014; Guengerich, Waterman, and Egli 2016). Ekspresija u korteksu nadbubrežne žlezde, jajnicima i testisima je u vezi sa sintezom i razgradnjom steroidnih hormona (Hasler et al. 1999). Posebno je neobična ekspresija CYP enzima u mozgu, gdje se upravo zahvaljujući ovim enzimima održava homeostaza holesterola i eliminacija retinoida, što ih dovodi u vezu, kako sa embrionalnim razvojem centralnog nervnog sistema, tako i sa tretmanima retioničnom kiselinom (Catharine Ross and Zolfaghari 2011). Ekspresija *CYP450* gena je specifična i u vezi sa funkcijom tkiva, odnosno organa u kojima CYP enzimi obavljaju različite uloge.

1.3. Funkcija CYP450 enzima

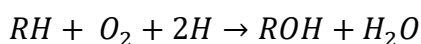
Prema svojoj funkciji, CYP enzimi mogu imati različite kategorije. Jedna od posljednjih klasifikacija je na višefunkcionalne ('moonlighting') i jednofunkcionalne ('non-moonlighting') proteine. 'Moonlighting' proteini mogu imati više funkcija ukoliko se nađu na različitim ćelijskim lokalizacijama, nekada i nakon formiranja kompleksnih struktura sa drugim proteinima, ili

menjanjem funkcije zbog različite ekspresije u ćelijskim odeljcima, na kraju mogu imati i više aktivnih mesta sa različitim funkcijama (Zhao and Waterman 2011). Treba naglasiti da CYP enzimi koji učestvuju u metabolizmu ksenobiotika spadaju u promiskuitetne, ali ne i 'moonlighting' proteine, dok za CYP enzime koji učestvuju u biosintezi endogenih jedinjenja važi upravo obrnuto (Zhao and Waterman 2011).

Fizičko-hemijski posmatrano, CYP enzimi su katalizatori oksidaciono-redukcionih reakcija u koje spadaju oksigenacija, hidroksilacija, epoksidacija, dealkilacija itd. (Isin and Guengerich 2007).

1.3.1. Monooksigenazna funkcija

Metabolizam ksenobiotika se može podeliti u dve faze, faza I i faza II. U fazi I ksenobiotici se oksiduju, redukuju i hidrolizuju, dok u fazi II dolazi do sinteze i konjugacije produkata iz faze I. CYP enzimi učestvuju u fazi I i to najčešće kao monooksigenazni enzimi koji preko hem-gvožđa u aktivnom centru katališu ugradnju jednog atoma kiseonika u supstrat, dok se drugi atom kiseonika redukuje stvaranjem molekula vode. Za ovu reakciju, energija se obezbeđuje oksidacijom, odnosno transferom elektrona sa NADPH (Barnaba et al. 2017).



U ljudskom organizmu, CYP enzimi kao monooksigenaze učestvuju u metabolizmu endogenih jedinjenja kao što su faktori rasta, hormoni, masne kiseline, retionična kiselina, prostaglandini, žučne kiseline, vitamin D. Sa druge strane, CYP enzimi razgrađuju egzogene supstance poput hemijskih jedinjenja, kancerogena i lekova (Fanni et al. 2014).

1.3.2. Hidroksilazna i druge funkcije

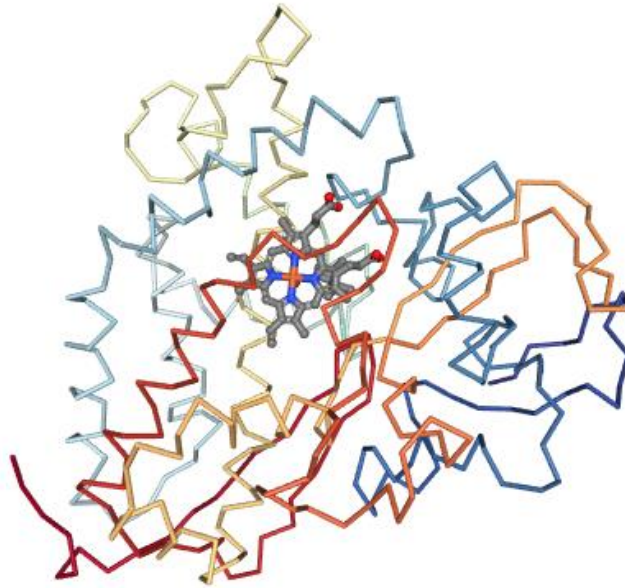
Dobar primer za višestruku funkciju CYP enzima jeste sinteza kortizola u kori nadbubrežne žlezde, koja zahteva hidroksilaznu aktivnost CYP17 enzima. Isti enzim učestvuje u sintezi androgena u gonadama svojom lijaznom aktivnošću (Zhao and Waterman 2011). Postoji i primer epoksigenazne funkcije, koju CYP2J2 ostvaruje u tkivu aorte tokom vazodilatacije ovog krvnog suda (Nayeem et al. 2011).

Ono što je zajedničko za oksidaciju, redukciju i hidrolizu jeste da su to glavne hemijske reakcije u fazi I metabolizma ksenobiotika. Može se reći da CYP enzimi zbog svojih fizičko-hemijskih svojstava učestvuju isključivo u ovom koraku metabolizam, međutim, procena je da je učešće u fazi I 80%, dok je u eliminaciji metabolita oko 70% (Bertz and Granneman 1997; Gibson and Skett 1986; P.B. Danielson 2002).

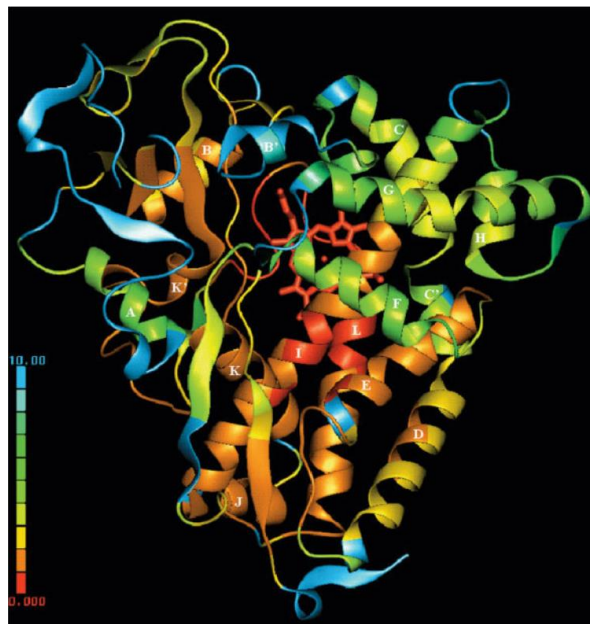
1.4. Struktura i diverzitet citohroma P450

Citohorm P450 enzimi su mali hemproteini veličine oko 200 KDa, dužine između 400 i 500 aminokiselina sa hem prostetičkom grupom u aktivnom centru enzima (Narhi and Fulco 1986). Najkonzervativniji deo enzima se sastoji od 12 alfa zavojnica označenih slovima alfabetom od A do L i manjeg broja beta ploča (Poulos, Finzel, and Howard 1987). Najbitniji delovi strukture CYP 450 enzima su I zavojnica sa treoninom koji konformaciono potpomaže katalitičku funkciju, kao i F i G zavojnice koje prave 'poklopac' iznad katalitičkog domena (Otyepka et al. 2007; Pochapsky, Kazanis, and Dang 2010). Unutar enzima se nalazi protopirinski (hem b) kofaktor, koji je vezan za katalitički

domen preko tiolatne grupe visokokonzerviranog cisteina (Mestres 2004; Šrejber et al. 2018). Unutar katalitičkog domena se nalazi atom gvožđa, stabilizovan sa šest aksijalno ligandnih veza sa molekulima vode, preko kojeg se vrše oksidaciono-redukcione reakcije (slika 4) (Guengerich 2018).



Slika 4 - Grafički prikaz kristalne strukture CYP cam enzima sa hemom u sredini. Izvor: <https://www.rcsb.org/3d-view/3CPP/1>



Slika 5 - Tipična struktura CYP450 (CYP101) sa zavojnicama. Heliksi obeleženi crvenom bojom imaju manju varijabilnost od heliksa obeleženih plavom bojom. Izvor: Sezutsu, Le Goff and Fezereisen 2013.

Iako su mnogi proteinski motivi pominju u kontekstu konzerviranih domena, kao što je WXXXR u C heliksu, treonin u heliksu I, EXXR u K heliksu, cistein ligand koji stoji aksijalno prema ravni hema je najčešće obeležje CYP enzima (Pochapsky, Kazanis, and Dang 2010; Sezutsu, Le Goff, and Feyereisen 2013). Ipak, ono što je zasigurno karakteristično za CYP enzime jeste konzervativni način zadobijanja tercijarne strukture koja je bitna za enzimsku funkciju, bez obzira na vrstu konzervativnog domena (Sezutsu, Le Goff, and Feyereisen 2013).

Citohrom P450 enzimi su prisutni u arhea i bakterija i eukariota, a otkrićem u virusima, diverzitet P450 je zadobio novu dimenziju koja ovaj enzim čini univerzalno rasprostranjenim kroz sve domene živog sveta (Lamb et al. 2019). Prema poslednjem pregledu literature, samo u prokariotima postoji preko 62 000 formi citohroma P450, dok u eukariotima taj broj premašuje 114 hiljada (Nelson 2018). Smatra se da su mutacije, a naročito duplikacije, glavni izvor diverzifikacije CYP enzima pored rekombinacija i genetičkog drifta (Lynch 2007). Takođe se smatra da je primarna funkcija CYP enzima, pre nego što su evoluirali u monooksigenaze, bila reduktazna, odnosno da je predački gen kodirao za anaerobnu reduktazu koja je omogućila veću adaptaciju u uslovima sredine pre akumulacije kiseonika u atmosferi (Sezutsu, Le Goff, and Feyereisen 2013).

1.5. Genetička struktura i nomenklatura *CYP* gena

Na nivou genoma, *CYP* geni se obično blisko pozicioniraju i sačinjavaju genski klaster, što je još jedno specifično obeležje ovih gena (Hoffman, Nelson, and Keeney 2001). Na nivou pojedinačnih enzima, varijabilnost u aminokiselinskoj sekvenci je prouzrokovana genskim varijantama u dezoksiribonukleinskoj (DNK) sekvenci *CYP* gena i to su: pojedinačne nukleotidne varijante, kratke insercije i delecije, inverzije i duplikacije kompletne genske sekvence (Polimanti et al. 2012). Genske varijante mogu ispoljiti tri vida uticaja na proteinske produkte gena u kojima se nalaze. Ukoliko je genska varijanta u kodirajućem delu sekvence i to na poziciji koja dovodi do nesinonimne mutacije, posledica je promena u aminokiselinskom sastavu što može promeniti funkcionalni fenotip enzima. Promene u nukleotidnom sastavu koje su sinonimne mutacije, nemaju uticaja na fenotip enzima. Ukoliko su pak genetičke varijante u nekodirajućem delu sekvence *CYP* gena, one mogu uticati na ekspresiju tog gena ako se nađu u promotorskoj sekvenci, ili na iskrajanje egzona kada su prisutne u intronskim sekvencama, ili mogu imati uticaja na druge vidove regulacije genske ekspresije.

U klasifikaciji *CYP* gena postoji neobičnost da definisanje alela, pa i gena, polazi od aminokiselinske sekvence, a ne od DNK sekvence. Kao posledica tog pristupa, alel se prevodi u kombinaciju genetičkih varijanti koje obrazuju/određuju haplotip. Dodatni nivo u kompleksnosti nomenklature dolaze od promene u broju kopija gena (CNV-Copy Number Variation). Ukoliko je neki haplotip u potpunosti odsutan (delecija), takvo stanje se definiše kao zaseban alel. Isto važi i za duplikacije haplotipova; naime, veći broj kopija istog haplotipa na jednom hromozomu (genetičke varijante u fazi) se smatra zasebnim alelom (Nelson 2006).

Ukoliko aminokiselinske sekvence *CYP* gena imaju najmanje 40% identičnosti, smatra se pripadaju istoj familiji i obeležavaju se dodatnim brojem npr. CYP1. Kada je ta identičnost veća od 55%, svrstavaju se u istu podfamiliju i obeležavaju dodatnim slovom npr. CYP1A, dok aleli istog gena imaju bar 95% identičnosti i predstavljaju različite izoforme jednog CYP enzima, te se obeležavaju dodatnim brojem npr. CYP1A1 (Arinç, Schenkman, and Hodgson 1995; Nelson 2006; Rose and Hodgson n.d.). Konkretno izoforme CYP enzima i haplotipovi kojima su kodirani u DNK

sekvenci se obeležavaju zvezdicom i brojem, npr. CYP1A1*1, ili brojem i slovom, npr. CYP1A1*1B. Ova klasifikacija se zasniva na kvantitativnoj skali, odnosno procentu aminokiselinske identičnosti, kojom se geni hijerarhijski grupišu u gore pomenute kategorije i kao takva ne odražava neko kvalitativno svojstvo (npr. funkciju) niti ima jasnu biološku zasnovanost, već je iz praktičnih razloga uvedena sa mogućnošću promene svojih kriterijuma na osnovu promene filogenije gena nakon otkrića novih sekvenci *CYP* gena (Nelson 2006). Postoje i predlozi da se iznad nivoa familije uvede nivo klanova, a naročito je problematčno definisanje istih gena u različitim taksonima, zbog čega je osnovan konzorcijum za dodeljivanje naziva *CYP450* genima, tako da treba očekivati dodatne izmene u nomenklaturi, koja je svakako dinamičnog karaktera (Gaedigk et al. 2018; Nelson 1999; Sim and Ingelman-Sundberg 2010).

1.5.1. Variranje u broju kopija gena (CNV)

Velika raznovrsnost *CYP* gena je najverovatnije posledica duplikacija, koje su očigledno bile učestalije od mutacija koje su se vremenom akumulirale unutar gena. Ovo se objašnjava time da sporo evoluirajući geni prilikom duplikacije imaju veću verovatnoću fiksacije, nego prilikom promena u sastavu sekvence, jer dok jedna kopija može obavljati svoju osnovnu funkciju, druga kopija može evoluirati u pravcu zadobijanja novih funkcija (Davis and Petrov 2004). Postoji i hipoteza prema kojoj u slučaju haploinsuficijencije, povećanje broja kopija funkcionalnog alela može nadoknaditi gubitak funkcije drugog alela istog gena (Kondrashov and Koonin 2004).

Prva manifestacija u promeni broja kopija gena jeste količina DNK matrice za ekspresiju tog gena. Najbolji primer za favorizaciju broja umnožaka haploinsuficijentnih gena je *CYP2D6* u genomu čoveka (Sezutsu, Le Goff, and Feyereisen 2013). *CYP2D6* gen ima dosta nefunkcionalnih alela i u heterozigotnom stanju sa nefunkcionalnim, funkcionalni alel može proizvesti čitav spektar metaboličkih fenotipova u zavisnosti od broja ponovaka. Ipak, smatra se da ovi ekstremi mogu pokazati veću adaptivnu vrednost samo u ekstremnim uslovima sredine, kakva je mogla biti pre 10 do 50 hiljada godina zbog upliva veće količine toksičnih alkaloida u ljudsku ishranu (Yasukochi and Satta 2011). Kod primata, *CYP2D6* i *CYP2D8* su evolutivno stari i bili su prisutni i pre razdvajanja prostih majmuna i čovekolikih majmuna, dok su se duplikacije 2D6 kao i nastanak 2D7P (duplikacija 2D6) odigrale pre razdvajanja čoveka od čovekolikih majmuna (Yasukochi and Satta 2011).

Variranje u broju kopija je od posebnog značaja za određivanje metaboličkih fenotipova jer može dovesti do povećanog kapaciteta za metabolizam supstrata, a kada su u pitanju ksenobiotici poput lekova, broj kopija gena za funkcionalne enzime može povećati šansu za neželjenom reakcijom ili fatalnim ishodom (Hosono et al. 2009).

1.5.2. *CYP450* genska superfamilija u ljudskom genomu

Superfamiliju *CYP450* u ljudskom genomu sačinjava 57 gena. Na osnovu aminokiselinske sličnosti, ovi geni su svrstani u 44 podfamilije i 18 familija (Ian, Elizabeth, and Nelson 2006). Pored ovih 57 gena, postoji 58 pseudogenskih sekvenci. *CYP* pseudogeni su najčešće krnje kopije funkcionalnih gena koje se ne transkribuju ili nakon transkripcije postaju nestabilne, zbog čega se ne transliraju u funkcionalne enzime.

U ljudskoj populaciji, diverzitet *CYP450* superfamilije je pod značajnim uticajem adaptacije na uslove sredine, dok se raspodela varijabilnosti asocira sa etničkom i geografskom strukturiranošću

koje primarno nastaju usled genetičkog drifta i inbridinga usled smanjenog protoka gena (Polimanti et al. 2012).

Najveći udeo u genetičkoj varijabilnosti čine pojedinačne nukleotidne varijante (SNV). Od oko 450 SNV rasutih širom 57 *CYP* gena, za 1.3% se pretpostavlja da ima visok, dok 14% ima umereni uticaj na funkciju enzima (Polimanti et al. 2012).

Na osnovu celokupne varijabilnosti u superfamiliji *CYP450*, globalna populacija (metapopulacija) se najverodostojnije grupiše u 5 populacija: Afričku, Euroazijsku, Istočno Azijsku, Američku i Okeaniju (Polimanti et al. 2012).

Raspodela genetičke varijanse u metapopulaciji je takva da se 8.17% nalazi između populacija, 4.83% između subpopulacija, 0.45% između individualnih genoma, dok je najveći deo (86.5%) unutar individualnih genoma (Polimanti et al. 2012). Za 26 SNVa, na nivou populacija i subpopulacija, indeks fiksacije je ispod statističkog praga značajnosti, ipak, učestalosti alela pojedinačnih *CYP* gena ne pokazuju toliko izraženu strukturiranost (Polimanti et al. 2012).

Tabela 2 – Prikaz diverziteta *CYP450* supergenske familije. Izvor:
<https://themedicalbiochemistrypage.org/cytochrome450.php>
<https://www.genenames.org/>

Podfamilija	gen	gen	gen	gen	gen	gen	hromozom
CYP1A	1	2					15
CYP1B	1						2
CYP2A	6	7	13				19
CYP2B	6						19
CYP2C	8	9	18	19			10
CYP2D	6						22
CYP2E	1						10
CYP2F	1						19
CYP2J	2						1
CYP2R	1						11
CYP2S	1						19
CYP2U	1						4
CYP2W	1						7
CYP3A	4	5	7	43			7
CYP4A	11	22					1
CYP4B	1						1
CYP4F	2	3	8	11	12	22	19
CYP4V	2						4
CYP4X	1						1
CYP4Z	1						1
CYP5A	1						7
CYP7A	1						8
CYP7B	1						8
CYP8A	1						20
CYP8B	1						3
CYP11A	1						15
CYP11B	1	2					8
CYP17A	1						10
CYP19A	1						15

CYP20A	1						2
CYP21A	2						6
CYP24A	1						20
CYP26A	1						10
CYP26B	1						2
CYP26C	1						10
CYP27A	1						2
CYP27B	1						12
CYP27C	1						2
CYP39A	1						6
CYP46A	1						14
CYP51A	1						7

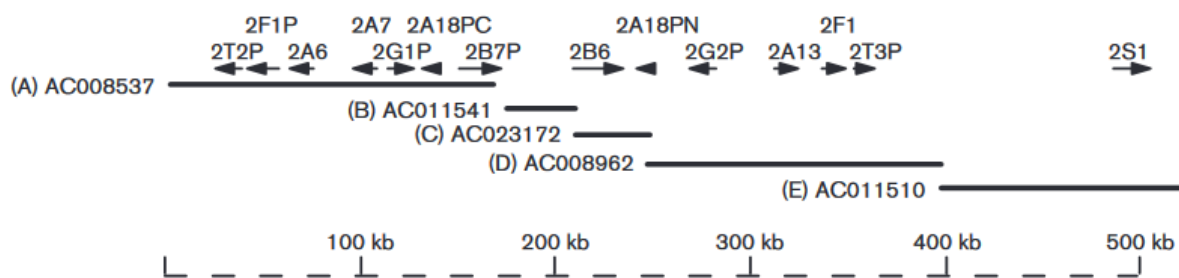
1.5.2.1. CYP1 familija

Članovi CYP1 familije se eksprimiraju kako u jetri, tako i van jetre, uglavnom u pankreasu, prostati, uterusu, plućima, gastrointestinalnom traktu, gde učestvuju u metabolizmu ksenobiotika i endogenih supstanci (Go, Hwang, and Choi 2015; Kawakami et al. 2011; Sankhwar and Sankhwar 2014). Transkripcija *CYP1* gena je regulisana preko ligand aktivirajućeg transkripcionog faktora AhR (Aryl hydrocarbon receptor) (D. W. Nebert et al. 2004). Do sada su identifikovana samo dva člana ove familije, *CYP1A1* i *CYP1A2*, koji su veoma konzervirani na interspecijskom nivou (Martignoni, Groothuis, and de Kanter 2006). Oba gena se nalaze na 15q24.1 lokusu u bidirekionalnoj poziciji jedan naspram drugome i dele 5'-bočni region, tako da imaju zajedničku regulaciju transkripcije (Ueda et al. 2006). U Pharmvar bazi podataka je opisano 27 SNV i 14 alela u *CYP1A1* genu, dok je u *CYP1A2* identifikovano 48 SNV i 40 alela (Gaedigk et al. 2018). Ipak, samo su dve SNV široko studirane (rs4646903 i rs1048943) i to uglavnom u asocijaciji sa malignim oboljenjima (Lu et al. 2020).

U ljudskoj metapopulaciji postoji značajna razlika u raspodeli alela *CYP1A1*, na način da se manje učestali aleli u Afričkoj, pojavljuju i od deset puta češće u Azijskoj populaciji (Inoue, Asao, and Shimada 2000; Rajeevan et al. 2017). Razlike su vidljive i u admiksičnim populacijama, gde je i efekat osnivača izražen kada se uporede alelske učestalosti *CYP1A1* u etničkim grupama sa učestalostima u izvornim populacijama (Coura et al. 2007).

1.5.2.2. CYP2 familija

CYP2 familija objedinjava preko 20 podfamilija, od kojih je za sedam poznata funkcija u metabolizmu kseno- i endobiotika (*CYP2A*, *CYP2B*, *CYP2C*, *CYP2D*, *CYP2E*, *CYP2S* i *CYP2J*). Geni koji kodiraju za proteine familije 2 su rasuti na različitim hromozomima. Međutim, prisustvo klastera od 6 gena na hromozomu 19, koji svi pripadaju različitim subfamilijama, unikatno je obeležje ove u odnosu na druge familije CYP enzima (Hoffman, Nelson, and Keeney 2001).



Slika 6 - CYP2 genski klaster na hromozmu 19. Izvor: Hoffman, Nelson and Keeney 2001.

Zbog mnoštva dokaza o učešću u metabolizmu ksenobiotika, naročito lekova, varijabilnost u podfamilijama 2C i 2D je najviše izučavana i dokumentovana (Manikandan and Nagini 2018).

CYP2C podfamiliju čine četiri gena (*CYP2C8*, *CYP2C9*, *CYP2C18*, *CYP2C19*) na hromozomu 10, među kojima je najbolje poznata uloga 2C9 i 2C19, dok su preostala dva još uvek nedovoljno istražena (2C8) ili od manjeg značaja za metabolizam zbog insuficijencije u translaciji (Stanley 2017). Prema broju varijabilnih mesta u sekvenci gena, 2C19 ima najveći broj do sada opisanih alela (34), ali i najveći broj SNV koji mogu uticati na promenu funkcionalnosti enzima (Bush et al. 2016). Obrazac formiranja metaboličkih fenotipova je specifičan za supstrat, ali je najčešće trimodalni karaktera, tako da homozigotna stanja produkuju ekstreme (loše ili ekstenzivne metaboličke fenotipove), dok heterozigoti obrazuju umerene tipove metabolizera (Niinuma et al. 2014; Schwarz 2003). Ekspresija podfamilije 2C je karakteristična za tkivo jetre i distalne delove gastrointestinalnog trakta (Lapple et al. 2003).

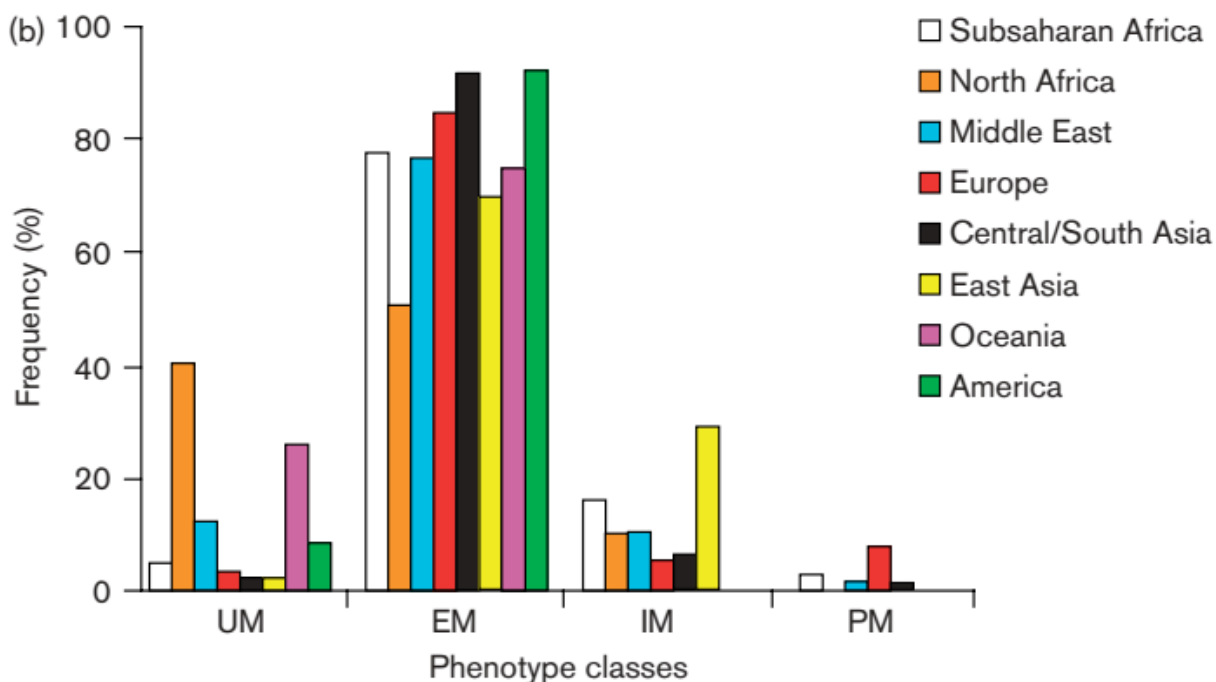
U podfamiliji 2D, *CYP2D6* ima najveću varijabilnost od svih ostalih CYP gena sa čak 418 genetičkih varijanti u egzonomima, intronima, 3' i 5' netranslirajućim regionima, kao i u promotorskoj sekvenci (Fujikura, Ingelman-Sundberg, and Lauschke 2015). Smatra se da je nekoliko alela ovog gena nastalo fuzijom sa pseudogenom *CYP2D7P*, što je doprinelo akumulaciji ovolikog broja genetičkih varijanti (Sezutsu, Le Goff, and Feyereisen 2013). U determinaciju alela je uključeno svega 97 genetičkih varijanti zbog njihovog značajnog uticaja na alteracije u aminokiselinskoj sekvenci ili regulaciji ekspresije gena (F. R. Wendt et al. 2017). Diploidne kombinacije alela 2D6 gena, posledično i izoenzima za koje ti geni kodiraju, određuju metaboličku aktivnost 2D6 enzima prema nekom supstratu. To se postiže dodeljivanjem pojedinačnih funkcionalnih vrednosti (0, 0.5, 1, 1.5, 2) svakoj izoformi proteina, odnosno alelu, odnosno haplotipu koji tu izoformu determiniše u sekvenci DNK. Metaoblički fenotip se određuje u zavisnosti od zbira pojedinačnih vrednosti metaboličke aktivnosti u diploidnom genomu. Tako se osobe koje imaju zbir 0 (dva nefunkcionalna alela) smatraju lošim metabolizerima. Kada je zbir 0.5, poput onog u kombinaciji nefunkcionalnog i polufunkcionalnog alela, fenotip se naziva umerenim. Za opseg vrednosti od 1 do 2, metabolički fenotip pripada klasi ekstenzivnih metabolizera. Pored navedene tri kategorije, postoji i četvrta-ultrabrzi metabolizeri. U ultrabrzih metabolizera mora postojati i umnožak kompletnih i funkcionalnih haplotipova/alela, tako da zbir metaboličke aktivnosti prelazi vrednost od 2 (Crews et al. 2012).

Na nivou metaboličkog fenotipa, CYP2D6 enzimi imaju karakterističnu kvadrimodalnu distribuciju koja nastaje variranjem unutar četiri klase gore pomenutih fenotipova: loših, umerenih, ekstenzivnih, ultrabrzih metabolizera.

Ovaj način klasifikacije metaboličkih fenotipova je specifičan za supstrat, ali je uz manje korekcije primenjiv u velikom broju slučajeva (za veći broj supstrata). Za tu svrhu je oformljen konzorcijum

koji objavljuje preporuke u određivanju metaboličkog fenotipa na osnovu genotipizacije *CYP2D6* gena u zavisnosti od supstrata (Relling and Klein 2011).

Neki od *CYP2D6* alela imaju vrlo uočljivu raspodelu u metapopulaciji, tako da postoje aleli koji su visoko učestali u nekim populacijama, dok u drugim populacijama ne prelaze standardni prag od 5%. Dobar primer je učestalost *4 alela u Evropskoj populaciji, koji je zbog svoje važnosti dobro opisan u skoro svim Evropskim subpopulacijama i u kojima se kreće u tipičnom opsegu od 9-20%, dok je u Azijskoj populaciji uglavnom ispod 5% učestalosti (Rajeevan et al. 2017).



Slika 7 - Učestalosti metaboličkih fenotipova u metapopulaciji. Izvor: Sistonen, 2007.

Slično je i sa metaboličkim fenotipovima, koji u najvećoj meri zavise od variranja u broju kopija gena. Ipak, treba napomenuti da postoji sumnja u tačnost procena ovih učestalosti, kako zbog metodologije određivanja CNV, tako i inferencije zasnovane na haplotipovima (Naranjo et al. 2016; Frank R. Wendt et al. 2018).

1.5.2.3. CYP3 familija

CYP3A podfamilija enzima čini najzastupljeniju grupu proteina koja se eksprimira u jetri čoveka i učestvuje u metabolizmu više od 30% lekova u kliničkoj upotrebi (U. M. Zanger et al. 2005). U ljudskom genomu se nalazi samo jedna subfamilija *CYP3A* sa genskim jatam koje čine četiri gena (*A4*, *A5*, *A7* i *A43*) na hromozomu 7 (McArthur et al. 2003). Najvarijabilniji član ovog klastera je gen *CYP3A4*, sa 46 potvrđenih alela (izvor: <https://www.pharmvar.org/gene/CYP3A4>) (Gaedigk et al. 2018). Zanimljivo je da ekspresija *CYP3A4* gena nije aktivna u prenatalnom periodu i da otpočinje nakon prve nedelje nenatusa da bi posle prvog meseca dostigla bazalni nivo prisutan i kod odrasle osobe (Fanni et al. 2014). Pretpostavlja se da je zbog svoje procesivnosti u metabolizmu egzogenih jedinjenja, *CYP3A4* gen evoluirao pod pritiskom diverzifikujuće selekcije, tako da je najveća raznovrsnost ostala u afričkoj populaciji kao izvornoj, dok je u drugim globalnim populacijama došlo do fiksacije manjeg broja alela ovog gena (Bush et al. 2016; X. Chen et al. 2009; Polimanti et al. 2012). Primer koji potvrđuje ovu hipotezu je učestalost *CYP3A4*1B* na globalnom nivou. Tako je npr. u afričko-američkoj subpopulaciji učestalost ovog alela procenjena na oko 45%, dok u

stanovništvu jugoistočne Azije nije zabeležen ni jedan 3A4*1B alel (Lamba et al. 2002; Zhou, Ingelman-Sundberg, and Lauschke 2017).

1.6. Citohrom P450 u personalizovanoj medicini

Personalizovana medicina je grana medicine koja se bavi proučavanjem individualizacije medicinskog tretmana uzimajući u obzir lična fiziološka i genetička svojstva pacijenta u cilju poboljšanja ishoda lečenja (Schork 2015).

Kako je primena medikamenata jedan od vidova lečenja, personalizovana medicina se oslanja na farmakogenetiku, pristup koji prilagođava primenu farmakoloških znanja na osnovu genetičkih svojstava organizma u cilju poboljšanja odgovora na lek. Takođe je ispravno reći i da je farmakogenetika proučava uticaj genetičke varijabilnosti na ishod primenjenih medikamenata (Bishop 2018).

Uopšteno, primena terapije lekovima može imati tri kvalitativna ishoda (Sajantila and Budowle 2016):

- odgovor organizma na terapiju
- izostanak odgovora na terapiju
- neželjena reakcija na lek (NRL).

U slučaju bilo kojeg od navedenih ishoda, postoji i kvantitativni udeo koji relativizuje ili sažima navedene kategorije ishoda. Tako odgovor na terapiju ne mora nužno izostati, već može biti nedovoljan.

Prema dostupnoj literaturi, učestalosti smrtnih ishoda koji su posledica neželjenih reakcija na lek se kreću u rasponu od 0.15% do 5% među hospitalizovanim pacijentima, dok je neadekvatna doza među pacijentima na terapiji u rasponu od 8% do 40% (Pirmohamed et al. 2004)(Wester et al. 2008)(Coste and Venot 1999).

Iako se pre uvođenja lekova u upotrebu sprovode obimna klinička istraživanja u cilju otkrića neželjenih reakcija na lekove, najveći broj NRL-a se otkrije tek nakon stavljanja leka u upotrebu (Moore et al. 1998). To se jednim delom može objasniti i uticajem genetičke varijabilnosti na metabolizam lekova. Procena je da više od 80 gena utiče na metabolizam lekova u meri i na način da je njihovo genotipiziranje opravdano prilikom određivanja vrste i doze terapije (Bush et al. 2016). Ovi geni se nazivaju i farmakogenima, a među njima su najznačajniji geni iz *CYP* superfamilije (Bush et al. 2016).

U personalizovanoj medicini, *CYP* enzimi se generalno izučavaju u dva konteksta, u vezi sa bolestima i u vezi sa metabolizmom lekova (odgovor na terapiju) (Q. Chen et al. 2011). Odnos zastupljenosti ova dva konteksta je nesrazmeran u korist drugog, ali nije zanemarljiv ni prvi, koji dovodi u vezu metabolizam raznih ksenobiotika, najčešće kancerogenih svojstava. Tako je predloženo da varijante *CYP1A1* gena treba uzeti u obzir kada se procenjuje štetni uticaj radnog okruženja zagađenog policikličnim aromatičnim hidrougljenicima (PAH) (Lin et al. 2000). Ova jedinjenja se smatraju prokancerogenima koja se u plućima mogu preko visoko eksprimiranog *CYP1A1* enzima prevesti u kancerogeni oblik (Le Marchand et al. 1998)(Lin et al. 2000).

Izučavanja uloge u metabolizmu lekova je staro koliko i samo izučavanje *CYP* enzima. Spisak lekova (izvor: <https://drug-interactions.medicine.iu.edu/MainTable.aspx>) na čiji metabolizam utiče neki od, ili više, gena iz *CYP* superfamilije je pozamašan i konstantno se produžava (D. W. Nebert

and Russell 2002). Međutim, kako zbog epistatičkih interakcija (jedan supstrat može biti metabolisan od strane više CYP enzima) i plejotropnih efekata (jedan CYP enzim može imati više supstrata) *CYP* gena, ali i uticaja sredinskih faktora, kao i (pat) fiziološkog stanja osobe, nije uvek moguće jasno asocirati genetičku sa metaboličkom varijabilnosti. Zato se u farmakogenetici za metabolizam svakog pojedinačnog supstrata (leka) određuje uticaj genskih varijanti i daju preporuke koje treba uzeti u obzir zarad personalizacije terapije (Relling and Klein 2011). Iako je često prenaplašena uloga genetičkih varijanti u varijabilnosti metabolizma lekova, procena je da među onima koji imaju manje efikasnu terapiju ili ispolje neželjene reakcije na lek, 15-30% je ljudi sa genetičkom dispozicijom kao uzrokom neželjene reakcije (Eichelbaum, Ingelman-Sundberg, and Evans 2006).

Većina varijanti *CYP* gena, koje potencijalno mogu dovesti do alternacija u metabolizmu lekova, su zapravo retke, međutim, kumulativno one čine više od 50% varijanti (Fujikura, Ingelman-Sundberg, and Lauschke 2015). Od svih varijanti koje se mogu genotipizirati za farmakogenetičku ili drugu upotrebu, izbor najčeće pada na one varijante za koje postoji dovoljno dokaza da se mogu primeniti u kliničkoj praksi (tabela 3).

Tabela 3 – Farmakogenetički najznačajniji aleli iz *CYP450* superfamilije. Izvor: Zanger, 2013.

<i>Alel</i>	RS identifikacioni br.	Pozicija	Efekat na protein
<i>CYP1A1*2C</i>	rs1048943	2454A>G	I462V
<i>CYP1A2*1C</i>	rs2069514	-3860G>A	Promoter
<i>CYP1A2*1F</i>	rs762551	-163C>A	Intron 1
<i>CYP1B1*6</i>	rs10012	142C>G	R48G
<i>CYP2A6*2</i>	rs1801272	1799T>A	L160H
<i>CYP2B6*4</i>	rs2279343	18053A>G	K262R
<i>CYP2B6*6</i>	rs3745274	15631G>T	Q172H
<i>CYP2C9*2</i>	rs1799853	3608C>T	R144C
<i>CYP2C9*3</i>	rs1057910	42614A>C	I359L
<i>CYP2C19*2</i>	rs4244285	19154G>A	Splicing defect
<i>CYP2C19*3</i>	rs4986893	17948G>A	W212X
<i>CYP2C19*17</i>	rs12248560	-806C>T	Promoter
<i>CYP2D6*3</i>	rs35742686	2549delA	Frameshift
<i>CYP2D6*4</i>	rs3892097	1846G>A	Splicing defect
<i>CYP2D6*5</i>		Recombination	Deletion
<i>CYP2D6*6</i>	rs5030655	1707delT	Frameshift
<i>CYP2D6*10</i>	rs1065852	100C>T	P34S
<i>CYP2D6*17</i>	rs28371706 rs16947	1023C>T 2850C>T	T107I R296C
<i>CYP2D6*41</i>	rs28371725	Splicing defect	
<i>CYP3A4*1B</i>	rs2740574	-392A>G	Promoter
<i>CYP3A4*22</i>	rs35599367	15389 C>T	Intron 6
<i>CYP3A5*3</i>	rs776746	6986A>G	Splicing defect
<i>POR*28</i>	rs1057868	31696C>T	A503V

1.6.1. CYP1A1

Najranija istraživanja CYP enzima su brzo proširena na potencijalnu ulogu u kancerogenezi, tako je pokazano da kancerogena supstanca N,N-dimethylaminoazobenzene (DAB) svoj demetilizovani oblik zadobija preko CYP1A1 enzima (Wahlang et al. 2015). Međutim, najčešći mehanizam aktivacije prokancerogena preko CYP1A1 jeste hidroksilacija aromatičnog prstena u policikličnim jedinjenjima (PAH) koja za rezultat ima visoko reaktivni produkt potentan da uvede onkogene mutacije (Buterin et al. 2000). Kako različite varijante enzimske izoforme mogu sa različitom efikasnošću metabolisati supstrat (prokancerogen), tako i genske varijante tih enzima imaju različit nivo asociranosti sa aktivacijom prokancerogena. Lista kancerogenih supstanci koje se metabolišu CYP1A1 enzimom nije svedena samo na DAB ili PAH, već postoji veći broj supstrata koji imaju kancerogena svojstva i koji se aktiviraju preko CYP1A1 (Young et al. 2010)(Mrozikiewicz et al. 2011). Pored dokaza o aktivaciji prokancerogena, postoje i dokazi da ovaj enzim učestvuje u eliminaciji kancerogena, što čini celokupnu sliku uloge CYP1A1 u kancerogenezi još kompleksnijom čak i bez osvrta na regulaciju genske ekspresije ili epigenetički mehanizam koji bi mogao imati uticaja na ekspresiju gena za ovaj enzim (Androutsopoulos, Tsatsakis, and Spandidos 2009)(Tekpli et al. 2012).

Do sada je opisano 14 alela gena *CYP1A1*, od kojih su tek 4 dovedena u vezu sa kancerogenezom raka pluća i to isključivo kroz studije genetičke asocijacije, što je ostavilo veliki prostor za špekulisanje o molekulskim mehanizmima kancerogeneze (Persson, Johansson, and Ingelman-Sundberg 1997; Zhang et al. 1996). Upravo zbog mogućnosti da određene varijante CYP1A1 enzima (M1, M2, M3, M4) mogu ukazati na veću podložnost kancerogenim jedinjenjima, nastala je ideja preventivnog delovanja u vidu umanjenja izloženosti sredinama sa visokom koncentracijom PAH osoba sa ovim genskim varijantama (Farmer et al. 2003). Koncept preventivne genotipizacije *CYP1A1* gena bi najviše uticao na medicinu rada, ali i na individualni nivo u smislu menjanja ličnih navika ili odabira posla.

1.6.2. CYP2C9

Najverodostojniji dokazi o uticaju CYP2C9 na efikasnost terapije jesu u vezi sa metabolizmom antagonista vitamina K, varfarinom (Guruprasad P Aithal et al. 1999). Ovaj antikoagulans se široko primenjuje u slučajevima duboke venske tromboze, infarkta miokarda, fibrilacije srčanih pretkomora, retko i u slučajevima subakutne diaseminovane intravaskularne koagulacije (Boral, Williams, and Boral 2016). Po dospeću u hepatocite, varfarin se oksiduje preko više CYP enzima, ali najpotentniji, *S* enantiomer varfarina, isključivo preko CYP2C9. Nakon oksidacije i povratka u krvotok, varfarin ima umanjeni afinitet da inhibira proteinski kompleks vitamin K epoksidne reduktaze preko kojeg zaustavlja koagulaciju snižavanjem koncentracije vitamina K (Deaton and Nappe 2020; Whirl-Carrillo et al. 2012). Ipak, neki aleli gena *CYP2C9* imaju umanjeni kapacitet za oksidaciju varfarina, kao što su *CYP2C9**2 i *3, tako da osobe koje su nosioci ovih genetičkih varijanti imaju pojačan odgovor na terapiju varfarinom i veći rizik od hemoragičnih događaja (G P Aithal et al. 1999; Dean 2012b). Pored uvećanog rizika za hemoragijom prilikom inicijalne primene varfarina, CYP aleli sa umanjenom aktivnošću ne utiču značajnije na koncentraciju varfarina u krvi nakon što ona dostigne stabilnu vrednost (Takanashi et al. 2000; Tässies et al. 2002).

Genotipizacijom *CYP2C9* alela *2 i *3 i prilagođavanjem doze varfarina prema smernicama konzorcijuma za primenu kliničke farmakogenetike (CPIC), obezbeđuje se brža i efektivnija terapija, dok se za dugoročnu upotrebu moraju uzeti u obzir genetičke varijante VKORC1 gena zbog njihovog uticaja na održavanje koncentracije ovog leka u krvi (Johnson et al. 2011).

Pored neželjenih reakcija na terapiju antikoagulanasa, aleli gena *CYP2C9* sa umanjenom funkcijom su dovedeni u vezu sa gastrointestinalnim krvarenjem izazvanim nesteroidnim antiinflamatornim lekovima (Carbonell et al. 2010). Međutim, preporuke CPIC još uvek nisu dostupne za prilagođavanje upotrebe nesteroidnih antiinflamatornih lekova zasnovanoj na genotipizaciji *CYP2C9*.

Za antiepileptik fenitoin postoje jasni dokazi i preporuke za umanjenje doze u pacijenata koji su nosioci bar jednog od *2 ili *3 alela (Swen et al. 2011). Kao i u slučaju varfarina, fenitoin se inhibira preko *CYP2C9* i u slučaju smanjenje aktivnosti ovog enzima, raste koncentracija leka u krvi što za posledicu može imati veći rizik za ispoljavanje neželjenih reakcija na ovaj lek ili pak intoksikaciju sa letalnim ishodom (McCluggage, Voils, and Bullock 2009).

1.6.3. CYP2C19

Iako je gen za *CYP2C19* enzim najpolimorfiji, odnosno prožet brojnim alteracijama u sekvenci DNK koje mogu uticati na funkcionalnost enzima (Bush et al. 2016), jasni dokazi o uticaju ovog enzima na metabolizam lekova postoje kod antitrombotičkih lekova (klopidogrel), inhibitora protonске pumpe i ponekih antidepresiva (Klotz 2009; Paré et al. 2010; Seeringer and Kirchheiner 2008; Sim et al. 2010).

Klopidogrel je antitrombotični lek koji se u hepatocitima prevodi u aktivnu formu u malom procentu, svega 15% proaktivne supstance se transformiše u svoju aktivnu formu (Scott et al. 2013). I taj mali deo od ukupne supstance zahteva dva uzastopna koraka oksidacije CYP enzimima i to prevashodno, uglavnom (ja bih izbegavao termin dominantno u genetičkim tekstovima) preko *CYP2C19* (Scott et al. 2013). Pored postojanja većeg broja alela za ovaj enzim, tri su pokazala značajan uticaj na funkcionalnost *CYP2C19* u aktivaciji klopidogrela. Dva alela, *2 (c.681G>A; rs4244285) i *3 (c.636G>A; rs4986893) dovode do nefunkcionalnosti enzima u aktiviranju klopidogrela, dok alel *17 (c.-806C>T; rs12248560) ispoljava suprotan efekat, tj. dovodi do povećane aktivnosti *CYP2C19* i posledično do ubrzane aktivacije clopidogrela (K. Kim et al. 2008; LI et al. 2012).

Preporuke CPIC u vezi sa doziranjem klopidogrela na osnovu genotipiziranja *CYP2C19*, prepoznaju četiri metabolička fenotipa: ultrabrze (sa genotipovima *1/*17, *17/*17), ekstenzivne (*1/*1), intermedijarne (*1/*2, *1/*3, *2/*17), kao i slabe metabolizere (*2/*2, *2/*3, *3/*3). Samo se za grupu intermedijarnih i slabih metabolizera preporučuje prelazak na alternativne antitrombotične medikamente (npr. prasugrel, ticagrelor) i to u cilju umanjenja rizika od tromboze stenta (Scott et al. 2013).

Za razliku od uticaja na metabolizam klopidogrela, nešto je drugačija situacija kada su u pitanju inhibitori protonске pumpe (PPI) i antidepresivi. Veći broj dokaza o uticaju *CYP2C19* alela na umanjen ili uvećan metabolizam omeprazola, pantoprazole, lansoprazola i drugih inhibitora protonске pumpe, ostao je bez standardizovane preporuke farmakogenetičkog konzorcijuma. Do sada postoje samo pojedinačne studije i na njima zasnovane preporuke o prilagođavanju terapije inhibitorima protonске pumpe na osnovu *CYP2C19* genotipa (Baldwin et al. 2008; Ogawa and Echizen 2010). Generalno, veći je uticaj na određivanje vrste i doze terapija PPI od strane *17 nego nefunkcionalnih *2 i *3 alela, ali postoji i značajna razlika u uticaju genotipa od leka do leka. Izbor PPI čiji je metabolizam nezavisan od *CYP2C19* genotipa omogućava zaobilazanje farmakogenetičkog testiranja, što se sve češće ističe kao argument za umanjenu primenu farmakogenetike u kliničkoj praksi kada su u pitanju PPI (Swen et al. 2011).

Najslabija asocijacija postoji između metabolizma antidepresiva i alela *CYP2C19* (Brøsen 2004). To ipak nije bila prepreka da CPIC preporuča da se doziranje amitriptilina redukuje za 50% u kategoriji slabih metabolizera na osnovu genotipizacije *CYP2C19* jer, iako mali, uticaj ovog enzima u metabolizmu amitriptilina je vrlo jasan (Hicks et al. 2013). Novija studija, na više od 2000 pacijenata, je istakla kao značajan uticaj nefunkcionalnog *CYP2C19* (alel *2) na koncentraciju SSRI leka escitaloprama, naročito u osoba sa *2/*2 ili *2/*1 genotipom koji imaju 3.3 puta veću koncentraciju ovog leka u serumu od osoba sa *1/*1 genotipom (Jukić et al. 2018).

1.6.4. CYP2D6

Do sada je u bazi podataka Nacionalnog instituta za zdravlje (PharmGKB) SAD označeno 14 lekova za koje postoje relevantni i verodostojni podaci o uticaju CYP2D6 enzima na njihovu kliničku primenu, kao i preporuke za individualizaciju doziranja na osnovu genotipizacije *CYP2D6* lokusa (izvor: <https://www.pharmgkb.org/gene/PA128/clinicalAnnotation>). Pored 14 označenih, klinička primena još 75 lekova zavisi od genetičke osnove CYP2D6 enzima.

Ugrubo, a na osnovu brojnih eksperimentalnih i kliničkih studija, moguće je izdvojiti nekoliko lekova zbog kojih je koncept 'personalizovane medicine' prevashodno i nastao (Bell et al. 2017; Crews et al. 2012; Dean 2012a; Gex-Fabry et al. 2008; Goetz et al. 2018; S.-H. Kim et al. 2018; Steimer et al. 2004).

- Selektivni inhibitori preuzimanja serotonina (paroksetin, fluvoksamin)
- Antiemetici (ondansetron, tropisetron)
- Triciklični antidepresivi (nortriptilin, doksepin, trimipramin, klomipramin, amitriptilin)
- Antiestrogeni citostatici (tamoksifen)
- Opioidni analgetici (kodein, tramadol)
- Inhibitori preuzimanja noradrenalina (atomoksetin)

Uzimajući u obzir posledice koje mogu nastati neadekvatnom terapijom, antiestrogeni citostatik, tamoksifen je najviše izučavan u farmakogenetičkom kontekstu (Binkhorst et al. 2015; Ingelman-Sundberg et al. 2007; Lim et al. 2005). U svojoj osnovi, tamoksifen je selektivni modulator receptora za estrogen, za koji je pokazano da za oko 30% smanjuje smrtnost žena obolelih od estrogen pozitivnog raka dojke (Abe, O., Abe, R., Enomoto, K., Kikuchi, K., Koyama, H., Masuda, H., Nomura, Y., Ohashi, Y., Sakai, K., Sugimachi, K., Toi, M., Tominaga, T., Uchino, J., Yoshida, M., Haybittle, J.L., Leonard, C.F., Calais, G., Geraud, P., Collett, V., Davies, C., Delmestr 2011). Međutim, u svom prvobitnom hemijskom obliku, tamoksifen je slab antiestrogen, što nije slučaj sa njegovim metabolitima. Oko 90% tamoksifena se demetiluje preko CYP3A4, a zatim, preko CYP2D6 enzima, oksiduje u endoksifen (Desta et al. 2004). Za razliku od tamoksifena, endoksifen ima 100 puta veći afinitet za blokiranje estrogenskog receptora i zato se smatra najbitnijim metabolitom tamoksifena (Lim et al. 2005). Problem u dostizanju adekvatne koncentracije endoksifena imaju osobe u kojih je CYP2D6 enzim neefikasan ili neefektivan u koraku oksidacije demetilovanog tamoksifena. Kada je u pitanju umanjena efikasnost, obično se podešava doza tamoksifena kako bi koncentracija endoksifena dostigla optimalni nivo (Schroth et al. 2017). Sa druge strane, potpuni izostanak oksidacije demetilovanog tamoksifena postoji kada nema CYP2D6 enzima koji može ikako oksidovati demetilovani tamoksifen (loši metabolizeri). U tom slučaju se pribegava zameni terapije sa aromataznim inhibitorima (Baatjes et al. 2018).

Farmakogenetičke preporuke za primenu tamoksifena u lečenju estrogen receptor pozitivnog raka dojke su formulisane CPIC uputstvom i predviđaju četiri kategorije metabolizera: loše, umerene, ekstenzivne i ultrarapidne (Goetz et al. 2018). Način određivanja kategorije metabolizera se zasniva na genotipizaciji većeg broja SNV i INDEL lokusa u *CYP2D6* genu na osnovu kojih se definišu aleli (haplotip) koji dobijaju funkcionalne skorove (0, 0.5 i 1) prema svojoj sposobnosti da oksiduju demetilovani tamoksifen. Pored određivanja haplotipa, određuje se i broj ponovaka haplotipa, koji može diskretno varirati od nijednog do više kopija gena (Schroth et al. 2017). Nakon određivanja skorova za pojedinačne alele, sabira se skor u diploidnom genomu i na osnovu njega određuje kategorija metabolizera: loši metabolizeri imaju skor 0; umereni metabolizer 0.5; ekstenzivni metabolizeri 1-1.5; ultrarapidni preko 2 (Goetz et al. 2018).

Metabolizam opioidnih analgetika u velikoj meri zavisi od *CYP2D6* enzima. Tako tramadol, koji je se sintetiše kao racemska smeša (1*S*1*S*)-tramadol i (1*R*1*R*)-tramadol, ima široku upotrebu kao analgetik u tretmanu umerenih i jakih bolova. Ovaj lek se u jetri prevodi u M1 metabolit koji poseduje najizraženija analgetička svojstva jer se za μ -opioidni receptor vezuje sa 200 puta većim afinitetom od ostalih metabolita, a da bi M1 nastao neophodna je O-demetilacija od strane *CYP2D6* enzima (Grond and Sablotzki 2004). Veća ekspresija *CYP2D6* funkcionalnog enzima kakva postoji u ultrabrzih metabolizera dovodi do naglog povećanja koncentracije M1 metabolita i respiratorne depresije koja može imati smrtni ishod (Elkalioubie et al. 2011; Orliaguet et al. 2015).

Za adekvatnu primenu terapije tramadolom, postoji zvanična preporuka CPIC prema kojoj genotip pacijenata može biti svrstan u tri kategorije: ultrabrzih, umerenih i loših metabolizera (Dean 2012a).

Kada su u pitanju lekovi sa svojstvom inhibitora preuzimanja serotonina i noradrenalina, često primenjivani u psihijatriji, odnosno tretmanu depresije, predikcija metabolizma na osnovu genotipa *CYP2D6* nije precizna. Određivanjem metaboličkog fenotipa na osnovu genotipa se može otkriti tek oko 16% osoba koje su loši metabolizeri ovih antidepresiva, što nije dovoljno da bi se farmakogenetička testiranja uvela u kliničku praksu (Preskorn et al. 2013).

1.6.5. *CYP3A4*

CYP3A4 enzim učestvuje u metabolizmu oko 30% svih lekova, s tim da oko 50% lekova u kliničkoj upotrebi imaju inhibitorno dejstvo na ovaj enzim (McArthur et al. 2003). Broj genetičkih varijanti koje utiču na ekspresiju ili funkciju ovog enzima nije tako veliki, ako se uporedi sa drugim *CYP* enzimima (2*C9*, 2*C19*, 2*D6*) (Werk and Cascorbi 2014; U. M. Zanger and Schwab 2013). Zbog inducibilne prirode *CYP3A4* enzima, može se očekivati da se genetičke varijante koje uvode alteracije u metabolizam supstrata 3A4 nalaze u nekodirajućem regionu sekvence DNK gena za ovaj enzim, odnosno u promotorskoj sekvenci i drugim kontrolnim sekvencama. Dosadašnje studije asocijacije su ukazale na rs35599367 i rs2740574 varijante u nekodirajućem delu sekvence (prva u intronu 6, a druga na oko 2 kb ispred start kodona), dok je u kodirajućem regionu retka varijanta rs55785340 asocirana sa funkcionalnom promenom *CYP3A4* (García-Martín 2002; Miyazaki et al. 2008; Z. Wang et al. 2013).

Od kliničkog značaja u slučajevima transplatacije bubrega jeste varijanta rs35599367 (alel *22) za koju je pokazano da negativno utiče na odstranjivanje imunosupresora iz organizma (takrolimus i ciklosporin A) (Elens et al. 2013). Nizvodna varijanta na poziciji -392A>G je asocirana sa rakom prostate i dojke, ali označena i kao varijanta koja potencijalno može uticati na sve lekove koji se metabolišu preko *CYP3A4* (Kadlubar et al. 2003; Rebbeck et al. 1998). Ipak, dosadašnje znanje o farmakogenetičkoj primeni *CYP3A4* varijanti u personalizovanoj medicini nije iznedrilo zvanične preporuke niti jasne klase metabolizera.

1.7. Citohrom P450 u sudsko-medicinskoj primeni

U nekim sudsko-medicinski slučajevima je teško objasniti način smrti, na primer kod trovanja, kada nije jasno da li je u pitanju samoubistvo ili zadesna smrt (Rockett et al. 2015). Pored toga postoje i slučajevi u kojima se uzrok smrti ne može u potpunosti objasniti (bela smrt), kao što je npr. iznenadna srčana smrt (Semsarian, Ingles, and Wilde 2015).

Otkrićem uticaja CYP enzima na metabolizam lekova, toksičnih supstanci kao i prokancerogena, ali i mogućnosti određivanja metabolizma na osnovu genotipizacije varijanti unutar *CYP* gena, nastala je realna osnova za novi pristup u objašnjavanju uzroka i porekla (manira) smrti ili drugih fizioloških stanja koja mogu biti predmet sudsko-medicinske analize kao što su hemijske i biološke povrede (Musshoff, Stamer, and Madea 2010).

Određivanjem genetičkih varijanti koje ukazuju na alteracije u metabolizmu lekova i toksičnih supstanci, ali i molekularno fiziološke alteracije (kanalopatije i miokardiopatije, venske tromboze) moguće je dobiti dodatna pojašnjenja u vezi sa uzrokom i načinom smrti (Ackerman, Tester, and Driscoll 2001; Musshoff, Stamer, and Madea 2010; Sajantila and Budowle 2016). Takav pristup u postmortem analizi je nazvan molekularnom autopsijom, a kao ideja postoji još od pedesetih godina prošlog veka kada je postalo jasno da je upliv molekularne biologije u sudsku medicinu neminovna (Davidson 1965). Ipak, zbog semantičkih i istorijskih razloga treba imati u vidu da je ovo uži pojam i da ne postoji standardizovana definicija termina 'molekularna autopsija', kao i da će dalji razvoj tehnologije dovesti do promena u značenju ovog termina (Sajantila and Budowle 2016).

Postoji izvestan broj lekova koji mogu dovesti do smrtnih ishoda (npr. kodein, etilmorfin) ili narušavanja zdravlja u značajnoj meri (Koren et al. 2006)(Haufroid and Hantson 2015). Pored toga, moguće je dodatno pojasniti način smrti u slučajevima gde je primena lekova ili izloženost nekoj supstanci, koja se metaboliše CYP enzimima, dovela do okolnosti koje su prouzrokovale smrt (depresija usled neadekvatne doze terapije, umanjena sposobnost učestvovanja u saobraćaju, neadekvatna medicinska briga). Sudsko-medicinski slučajevi u kojima je bitan toksikološki aspekt supstanci u čijem metabolizmu učestvuju CYP enzimi, mogu biti dodatno pojašnjeni određivanjem metaboličkog fenotipa za svaku supstancu od interesa ukoliko postoje jasni dokazi o vezi između genetičke i fenotipske varijabilnosti. Spisak supstanci koje zadovoljavaju ove uslove nije toliko dugačak ukoliko se ima u vidu broj supstanci koje se metabolišu preko CYP enzima. Ipak, na osnovu zabeleženih slučajeva intoksikacije supstancama za koje se dokazano da se metabolišu CYP enzimima, te na osnovu naučnih radova koji su uputili na genetičku osnovu u alteraciji metabolizma i studija koje su dokazale funkcionalne promene CYP enzima prema tim supstancama, moguće je izdvojiti one najvrednije pažnje u objašnjenju uzroka smrti (Musshoff, Stamer, and Madea 2010):

1. Kodein/Morfin

Kodein je opioidni analgetik koji se manjim delom prevodi u morfin O-demetilacijom pomoću CYP2D6 enzima (Crews et al. 2012). Osobe koje poseduju veći broj kopija gena za funkcionalni *CYP2D6* brzo povećavaju koncentraciju morfina u krvi iznad predviđene za prosečne osobe. Dostizanje visoke koncentracije morfina može ispoljiti sve neželjene reakcije, pa čak i respiratornu depresiju što je dokumentovano većim brojem pojedinačnih slučajeva (Ciszkowski et al. 2009; Hermanns-Clausen et al. 2009; Koren et al. 2006).

2. Dihidrokodein

Hemijski sličan kodeinu, dihidrokodein se oksidativnom redukcijom preko CYP2D6 prevodi u *N*- i *O*-demetilovane metabolite, nordihidrokodein i dihidromorfin. Odnos ova dva metabolita je takav

da se dihidromorfin produkuje u manjoj meri nego nordihidrokodein. Kada CYP2D6 nije funkcionalan, *O*-demetilacija izostaje i sa njom količina aktivnog metabolita u krvi. Suprotno je u slučaju ultrarapidnih metabolizera, kada koncentracija dihidroksimorfina može dostići visok nivo koji je deset puta viši u odnosu na loše metabolizere, zbog čega se mogu ispoljiti toksična dejstva slična morfinu (Kirkwood, Nation, and Somogyi 2003; Skopp et al. 1998). Inhibitori CYP2D6 enzima, kao što su kinidin sulfat i kinin, mogu umanjiti nivo dihidroksimorfina, ali i drugi komorbiditeti, kao što je renalna disfunkcija, pored toga, inhibitori mogu umanjiti brzinu eliminacije dihidroksimorfina i time dovesti do povećane koncentracije ovog opioida (Conway et al. 2006; Kirkwood, Nation, and Somogyi 2003).

3. Oksikodon

Kao i kodein i dihidroksikodein, oksikodon se metaboliše u jetri i prolazi procese *N*- i *O*-demetilacije koje katalizuje CYP2D6. Aktivni metaboliti su noroksikodon i noroksimorfin, a osobe koje su ultrabrzni metabolizeri brzo postižu visoke koncentracije ovih metabolita u krvi (Kaiko et al. 1996). Za razliku od kodeina i dihidroksikodeina, u eliminaciji oksikodona, uloga CYP3A4 je veoma izražena (C. Samer et al. 2010). Ukoliko se prilikom upotrebe oksikodona simultano upotrebi i agens koji inhibira CYP3A4, npr. ketokonazol, rizik za ispoljavanjem toksičnog dejstva se drastično povećava (Overholser and Foster 2011). Najrizičnija situacija je u slučaju CYP2D6 ultrabrzih metabolizera koji imaju i inhibiran CYP3A4 i renalnu insuficijenciju (C. F. Samer et al. 2013).

4. Metadon

Racemsku smešu metadona čine *R*- i *S*-metadon, od kojih *R*-enantiomer aktivira μ -opioidni receptor, dok *S*-enantiomer može blokirati hERG kalijumov kanal (Sanguinetti and Mitcheson 2005). Ventrikularne aritmije su direktno zavisne od hERG kalijumovog kanala, čija opstrukcija obično dovodi do produženog QT intervala, a za posledicu može imati iznenadnu smrt usled zastoja rada srca (Abriel et al. 2004; Eap et al. 2007; Sanguinetti and Mitcheson 2005). Pokazano je da CYP2B6*6 homozigotne osobe imaju veću koncentraciju *S*-metadona u plazmi i produženiji QT interval (Eap et al. 2007). Pored CYP2B6, pokazano je da CYP3A4 i 2C9 imaju značajan uticaj na metabolizam metadona. Dok se CYP3A4 dovodi u vezu sa ispoljavanjem neželjenih reakcija tokom lečenja bolesti zavisnosti, CYP2C9 je asociran sa promenama koncentracije metadona u urinu (C.-H. Chen et al. 2011; Tsai et al. 2014).

5. Fentanil

Sintetički opioidni analgetik, fentanil, je 50 do 100 puta potentniji u odnosu na morfin i može se aplicirati kako intravenski, tako i intramuskularno i transdermalno. *N*-dealkilacijom preko CYP3A4 ili oksidacijom preko CYP3A5, fentanil prelazi u norfentalin koji čini 99% metabolita (Lötsch et al. 2013). Značajane promene u koncentraciji norfentalina koje mogu ispoljiti neželjena dejstva su opažene kada u kombinaciji homozigota za CYP3A5*3 i CYP3A4*1B (Jin et al. 2005). Ipak, udeo CYP3A4 i 3A5 enzima u metabolizmu fentanila nije tako veliki, već ima dodatni efekat na dostizanje toksične koncentracije i primena genetičkog testiranja je još uvek u povoju (Ruan, Chiravuri, and Kaye 2016).

6. MDMA

Poznatija kao glavna supstanca 'ekstazija', MDMA (3,4-metilen dioksimetamfetamin) može izazvati različite štetne efekte u organizmu. Oni se kreću od hroničnih, koje odlikuje manja degeneracija serotonergičkih neurona, preko akutno umerenih, za koje je karakterističan spazam mišića i gubitak svesti, do teških akutnih stanja u koja spadaju intrakranijalna krvarenja, akutna kardiovaskularna stanja i oštećenja jetre (Beitia et al. 2000; Lange-Brock et al. 2002; Liechti and

Vollenweider 2001; Scheffel et al. 1998; Turillazzi et al. 2010). CYP2D6 učestvuje u metabolizmu oko 30% MDMA, međutim, produkti koji nastaju jesu katehol-metaboliti čija dalja oksidacija stvara reaktivne oblike kiseonika i azota potentnih da izazovu oštećenja lipoida pa i DNK (Barreto et al. 2009; Carvalho et al. 2002; Kovacic and Cooksy 2005). U studiji na fibroblastima pluća funkcionalne varijante CYP2D6 enzima su produkovale značajnu količinu *N*-metil- α -metildopamina, koji je 100 puta citotoksičniji od samog MDMA (Carmo et al. 2006). Iako ne postoji jasno podudaranje između slučajeva intoksikacije MDMA i CYP2D6 fenotipa, niti jasna dozna zavisnost, eksperimentalni podaci govore u prilog tezi da CYP2D6 ima doprinosa u štetnim dejstvima MDMA, ali ne kao jedini faktor u svim, već kao mogući faktor u idiosinkratičnim slučajevima.

7. Antidepresivi

Toksičnost izazvana antidepresivom koja je dovedena u vezu sa funkcionalnošću CYP450, odnosi se na pojedinačni slučaj upotrebe fluoksetina. Studija slučaja trovanja devetogodišnjeg dečaka pod standardnom dozom fluoksetina je otkrila defektni lokus gena *CYP2D6* (SALLEE, DeVANE, and FERRELL 2000). Međutim, nakon tog slučaja, toksičnost ovog leka nije više dovedena u vezu sa CYP450 zavisnim metabolizmom (Roberts et al. 2004). Za razliku od uzroka smrti intoksikacijom, poreklo smrti se objašnjava sagledavanjem širih okolnosti koje su mogle indirekto dovesti do smrti. Dobar primer je samoubistvo kod pacijenata sa kliničkom dijagnozom depresije, a koji su pod terapijom antidepresivima. Naime, ukoliko je jedan udeo tih pacijenata u nemogućnosti da postigne optimalnu koncentraciju aktivnog metabolita zbog uvećanog kapaciteta CYP enzima, terapija antidepresivima neće dostići optimalni nivo i za posledicu može imati suicid pacijenta. Iako deficijencija u metabolizmu antidepresiva nije direktan uzrok smrti, može se smatrati za okolnost koja je indirektno dovela do suicida (Zackrisson, Lindblom, and Ahlner 2010).

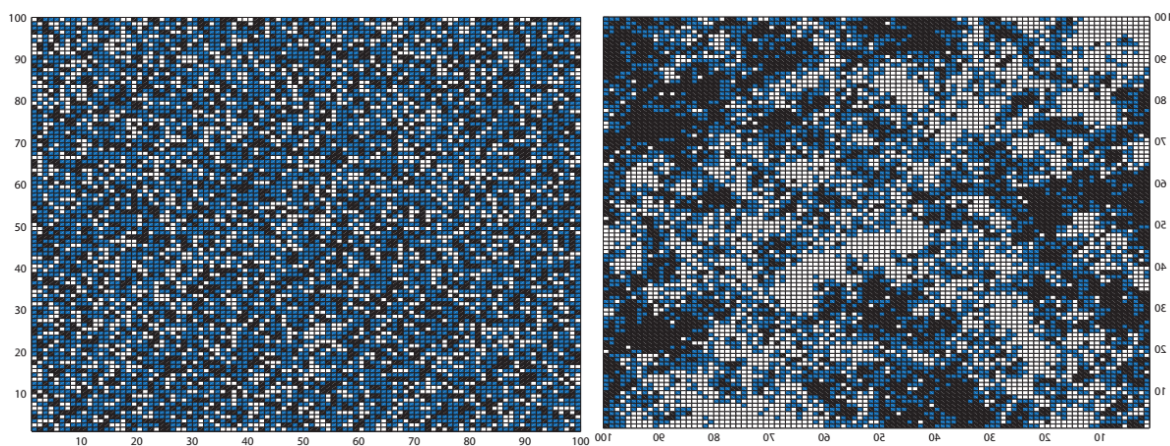
Još jedna hipoteza o koristi odmolekularne autopsije se odnosi na slučajna trovanja lekovima. U graničnim slučajevima, kada je koncentracija toksične supstance prešla prag tolerancije organizma i dovela do trovanja, teško je objasniti način smrti kod osoba koje su tu supstancu mogle uzeti kao remedium, ali i kao sredstvo za suicid. Kako se brzina metabolizma kod mrtvih ljudi ne može odrediti na eksperimentalni način, preostaje predikcija na osnovu genotipa. Na taj način se može razdvojiti zadesno trovanje kod npr. ultrabrzih metabolizera od samoubilačkog trovanja lekovima kod loših ili ekstenzivnih metabolizera (Sajantila and Budowle 2016).

I za personalizovanu medicinu i za postmortem analize (molekularnu autopsiju) je bitan odabir genetičkih varijanti koje su predmet farmakogenetičkog testiranja. Potrebno je da imaju bar tri stvojsva: da su snazno asocirane sa promenama u metabolizmu lekova, da je njihova učestalost u populaciji dovoljno visoka (populaciono specifični) i da ne postoji visok stepen genetičke struktuiranosti populacije (Bush et al. 2016).

1.8. Teorija i analiza genetičke strukture populacije

Najopštija i verovatno najtačnija definicija genetičke struktuiranosti jeste da je ona odstupanje od slučajne raspodele genetičkih varijanti u nekoj populaciji. To bi značilo da je opšta populacija sastavljena od subpopulacija u kojima se alelske učestalosti značajno razlikuju (Chakraborty 1993).

Genetička struktuiranost može biti posledica admiksije dve populacije, ili više divergentnih populacija između kojih nije bilo značajnijeg protoka gena i u kojima su alelske učestalosti različite (Halder and Shriver 2003; J. Z. Li et al. 2008; Matthew B. Hamilton 2012; Sousa et al. 2012). Sa druge strane i bez admiksije genetička struktuiranost u određenoj meri nastaje kao posledica fizičke udaljenosti između jedinki koje sačinjavaju populaciju (slika 8). Naprosto, fizički udaljene jedinke imaju manju verovatnoću ukrštanja zbog čega se stvaraju lokalne strukture u kojima pod dejstvom genetičkog drifta i inbridinga dolazi do promena u alelskim učestalostima (Matthew B. Hamilton 2012; Wijsman and Cavalli-Sforza 1984). Posledice genetičke struktuiranosti se ogledaju u smanjenju efektivne veličine N_e unutar dema i povećanju šanse da se neka genetička varijanta fiksira ili izgubi (Matthew B. Hamilton 2012).



Slika 8 - Simulacija nastanka genetičke struktuiranosti nakon 200 generacija slučajnog ukrštanja sa početnim alelskim učestalostima $q=p=0.5$. Plava polja označavaju heterozigote, dok crna i bela polja označavaju homozigote. U grafičkom prikazu levo orikazana je pocetna situacija, sa nasumicno rasporedjenim genotipovima, dok je na slici desno prikazana raspodela nakon 200 generacija slučajnog ukrštanja između susednih polja (3x3). Izvor: Hamilton et al. 2012

Ostaje pitanje šta je uopšte subpopulacija i koliko izdvojena ona mora biti u odnosu na opštu populaciju! Losan (Lawson) navodi da populaciju ili subpopulaciju čini grupa jedinki koje se nasumično razmnožavaju i čine geografski specifične deme. Međutim, isti autor navodi da postoje razlike između statističke definicije populacije i subpopulacije u odnosu na biološku (Daniel John Lawson 2015). Nasumično ukrštanje je pretpostavka koja je uvedena zarad uklapanja teorijskih modela u biološku realnost, jer se u realnosti mogu konstatovati samo različiti nivoi asortativnog ukrštanja (Jorde 1980). Možda je najtačnije reći da je subpopulacija zasebna grupa jedinki sa većim koeficijentom inbridinga u odnosu na opštu populaciju

Kada je u pitanju ljudska populacija, definisanje subpopulacije je najčešće zasnovano na geografskom poreklu, lingvističkim, kulturološkim i fizičkim svojstvima jedinki od kojih je sačinjena, ali takva subpopulacija ne mora biti i genetički izdvojeni entitet (Daniel John Lawson 2015; Pritchard, Stephens, and Donnelly 2000). I u ljudskoj populaciji panmiksija je teorijski koncept koji odstupa od realnosti, jer je moguće da naselja i gradovi mogu biti delimično izolovana ostrva,

odnosno subpopulacije (deme) koje se značajno razlikuju u odnosu na opštu populaciju (Matthew B. Hamilton 2012). Ovo viđenje se uklapa u koncept ostrvskog modela i izolacije udaljenošću (distancom) i još bolje u koncept kontinualnog modela odnosno kontinuirane izolacije udaljenošću (eng. *continuous isolation by distance*) (Hardy and Vekemans 1999; S Wright 1950). Čak i kada bi jedinke u populaciji bile ravnomerno geografski pozicionirane tako da ne postoje ostrva u vidu naselja, logično je za očekivati da bi verovatnoća pronalaska partnera za ostavljanje potomstva zavisila od geografske udaljenosti, jer je ukrštanje uvek ograničeno u nekoj meri, što bi opet rezultiralo u generisanju novih subpopulacija. Sa druge strane, prilikom nastajanja novih genskih varijanti, isti fenomen bi bio opažen, opet bi zbog opadanja verovatnoće ukrštanja između jedinki sa povećanjem fizičke udaljenosti bio opažen gradijent učestalosti novonastale varijante, od epicentara nastanka ka distalnim područjima u populaciji.

U ljudskoj populaciji, pored geografske udaljenosti, postoji društveni faktor koji može dovesti do genetičkog strukturiranja populacije. Odabir partnera za ostavljanje potomstva ima kulturnološku i socio-ekonomsku dimenziju, tako da i društveni koncept, kakava je npr. nacionalna pripadnost ili kastinski sistem, putem povećanog inbridinga, utiče na promene u alelskim učestalostima između subpopulacija i ne mora nužno biti unapred definisan (Bigham 2014).

Pored geografskih i društvenih uzroka, moguće je i postojanje latentne genetičke strukture (skrivena, kriptovana eng. *cryptic structure*) koja se ni na koji način ne može anticipirati osim populaciono genetičkim testiranjem. To stvara poseban problem u studijama genetičke asocijacije i forenzičkoj genetici (Ewens and Spielman 1995; Foreman, Smith, and Evett 1997; Pritchard, Stephens, and Donnelly 2000). Naime, latentna strukturiranost opšte populacije može za rezultat imati povećanu genetičku sličnost unutar jedne od poređenih grupa (testirane i kontrolne) praćenu povećanjem učestalosti neke genetičke varijante, što dovodi do lažno pozitivnih asocijacija bolesti sa tom genetičkom varijantom (Aistle and Balding 2009). U forenzičkoj genetici je poseban problem izračunavanja podudarnosti STR (eng. *Short Tandem Repeats*) profila i načina korigovanja rezultata u populaciji sa izraženom genetičkom strukturom, jer je uniformnost alelske učestalosti pretpostavka za primenu pravila množenja prilikom određivanja učestalosti genotipa (Balding and Nichols 1995).

Problemi koji mogu nastati kao rezultat genetičke strukturiranosti se prevazilaze njenim detektovanjem, što je zapravo primena matematičkih metoda u analizi bioloških podataka kojom se biološka realnost modelira. To znači da je ljudsko razumevanje bioloških sistema, kakva je i populacija, u velikoj meri svedeno na matematičko uprošćavanje i redukciju. Zbog takvog pristupa, koji se i u najboljim slučajevima holističkog pristupa samo približava realnosti i po pravilu ostavlja prostora za pogrešku, postoji potreba za konstantim iznalaženjem novih, tačnijih i preciznijih, algoritama za modelovanje bioloških sistema.

Klasičan pristup u detekciji genetičke strukturiranosti zahteva prethodno određivanje subpopulacija i analizu varijanse (F -statistika) (Fisher 1919; Marcel 1948; Weir and Cockerham 1984; S Wright 1950). F -statistika se zasniva na nekoliko parametara koji su izvedeni iz indeksa fiksacije F , odnosno, na smanjenju efektivne veličine populacije N_e i posledičnom padu heterozigotnosti (Valundov efekat) (WAHLUND 2010; S Wright 1950)

$$F = \frac{1}{2N_e} = \frac{H_e - H_o}{H_e}$$

gde je H_e očekivana, a H_o opažena heterozigotnost. Ti parametri su F_{IT} , F_{IS} i F_{ST} . Oznake I , S i T se odnose na nivoe u genetičkoj strukturi populacije: individue, subpopulacije (subdivizije) i opšte (totalne, ukupne) populacije (Sewall Wright 1965). Matematički se definišu kao:

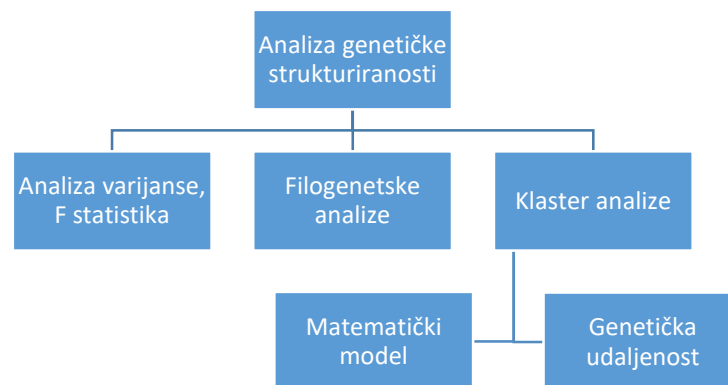
$$F_{IS}(f) = \frac{H_S - H_I}{H_S}$$

$$F_{IT}(F) = \frac{H_T - H_I}{H_T}$$

$$F_{ST}(\theta) = \frac{H_T - H_S}{H_T}$$

gde je H_I opažena heterozigotnost u celoj populaciji, H_S očekivana heterozigotnost u subpopulaciji, dok je H_T očekivana heterozigotnost u opštoj populaciji. Tri parametra (F_{IS} , F_{IT} i F_{ST}) opisuju uslovnu verovatnoću sparivanja alela u diploidnom organizmu na različitim nivoima populacione hijerarhije. F_{IS} parametrom se opisuje odnos dva alela u nasumično odabranom genotipu iz subpopulacije, na taj način što poredi prosečnu opaženu heterozigotnost u populacijama sa opštom očekivanom heterozigotnošću. F_{IS} može ukazati na odstupanje od nasumičnog ukrštanja unutar subpopulacije. F_{IT} opisuje odnos između alela unutar nasumično izabranog genotipa iz subpopulacije u kojoj mogu postojati veća odstupanja od nasumičnog ukrštanja i koja može imati različite alelske učestalosti u odnosu na druge subpopulacije. F_{ST} je uslovna verovatnoća da su dva nasumice izabrana alela identična po poreklu unutar subpopulacije, ukoliko su po poreklu identična i na višem hijerarhijskom nivou, odnosno u opštoj populaciji. Drugim rečima, ukoliko se dva ista alela nađu u jednom genotipu, F_{ST} procenjuje verovatnoću da oba alela dolaze iz iste subpopulacije. F_{ST} parameter dobro opisuje razlike u alelskim učestalostima između subpopulacija (Matthew B. Hamilton 2012).

Pored analize genetičke struktuiranosti F statistikom, moguće je primeniti i filogenetske analize za istu svrhu, ali prvenstveno za haplotipske podatke (Volz et al. 2020). Međutim, nije uvek moguće prethodno definisati subpopulacije, niti je čest slučaj da su genetički podaci haplotipski. Stoga se prvenstveno upotrebljavaju metode klasterovanja za detekciju subpopulacija u uzorku iz opšte populacije i nakon toga kvantifikuje genetička struktuiranost F statistikom (Teo 2008).



Slika 9 – Grafički prikaz odnosa metoda za analizu genetičke struktuiranosti

Grupisanje (klasterisanje) jeste raspodela jedinica posmatranja u klase, a na osnovu neke mere sličnosti. Postoji više vrsta klaster analiza kao što su: grafičko, hijerarhijsko, centroidno, zasnovano na gustini, zasnovano na raspodeli (Tan, Steinbach, and Kumar 2005). U svrhu otkrivanja genetičke strukture populacije se koriste dva analitička pristupa. Prvi se zasniva na određivanju genetičke udaljenosti, dok drugi teži primeni statističkih modela.

Analiza genetičke struktuiranosti na osnovu genetičke udaljenosti je u suštini multidimenzionalno skaliranje na osnovu matrica genetičke udaljenosti (Greenbaum, Templeton, and Bar-David 2016). Ovakav vid analize sledi sve principe vizuelizacije višedimenzionalnih podataka. Prvenstveno se teži očuvanju što veće varijabilnosti matrica genetičke udaljenosti, a zatim grafičkom predstavljanju na nižedimenzionalnoj skali, što omogućava vizuelno uočavanje klastera (grafičko klasterovanje).

Osnovni problemi sa metodama zasnovanim na genetičkoj udaljenosti, koji su po svojoj prirodi kontinuirani (neprekidni, nediskretni) modeli, jesu njihova zavisnost od mere genetičke udaljenosti u grafičkom predstavljanju podataka i nemogućnosti inkorporacije geografskih podataka (Pritchard, Stephens, and Donnelly 2000). Naime, nekada je genetička razlika između subpopulacija kvantitativna i mala, ali jasna, što metodi zasnovani na genetičkoj udaljenosti mogu zanemariti zbog malog doprinosa ukupnoj varijabilnosti.

1.8.1. Ocena genetičke strukturiranosti populacije na osnovu matematičkih modela

'Svi modeli su pogrešni' je poznata rečenica Džordža Boksa, koja je kasnije proširena u 'Svi modeli su pogrešni, ali neki su korisni'. Ova rečenica je često citirana zbog rasprostranjene prakse da se rezultati matematičkog modelovanja prihvataju 'zdravo za gotovo' ili pak opovrgavaju zbog proizvoljnih pretpostavki. Nažalost, još jedan navod Boksa nije tako ekstenzivno citiran i glasi: 'kako su svi modeli pogrešni, naučnici moraju biti upozoreni na ono što je vrlo pogrešno. Pogrešno je biti zabrinut zbog miševa kada su tigrovi u okolini/na sve strane' (Box 1976).

Smatra se da postoji objektivna realnost kojoj se ljudski um može približiti, može je modelovati, ali je nikada ne može u potpunosti opaziti (Putnam 1980). Takođe se smatra i da objektivna realnost jednostavno ne postoji i kao primer u prirodnim naukama se navode kvantni eksperimenti (Proietti et al. 2019). Ova dva pravca, realizam i nominalizam su prisutni u svakom području nauke. Bez temeljitijeg razmatranja ovog filozofskog i epistemološkog problema, moramo se osloniti na navode Boksa, racionalno je koristiti modele ukoliko su oni praktični bez obzira na pretpostavke u postulatima tih modela (Box 1976; Giere 2004). U krajnjem slučaju, logika nalaže da je svo znanje model ili koncept kojim ljudski um interaguje sa realnošću, ali to ne znači nužno da materijalni svet ne postoji već da naučnici koriste koncepte kojima jednostavnije razumeju zakonitosti materijalnog sveta (Dumontier and Hoehndorf 2010).

Matematički modeli u genetici korespondiraju sa nastankom genetike i omogućili su razvoj statističkih metoda koji su vremenom pratili razvoj molekularne biologije i evoluirali u alat za analizu molekularno-genetičkih podataka (Neuhauser 2004). U populacionoj genetici matematički modeli su osnova svake analize i kao dobar primer se može navesti Hardi-Vajnbergov princip razvnoteže, koji je zapravo populaciono genetička primena matematičkog modela očekivanih učestalosti za binomne eksperimente, formulisan još u 17. veku od strane braće Bernuli (Beerenwinkel and Siebourg 2019). Možemo čak i tvrditi da bez matematičkih modela, populaciona genetika ne bi postojala, jer je teško, gotovo nemoguće, proučavati učestalosti alela i varijabilnost bez primene matematike, odnosno, modeliranja.

Jedan segment populacione genetike, koji se bavi genetičkom strukturom populacije, je prožet posebnim pristupom u matematičkom modeliranju. Treba naglasiti da je veliki broj filozofskih kontroverzi i u vezi sa pretpostavkama koje modelovanje genetičke strukture populacije nosi sa sobom. To su pretpostavke na kojima modeli počivaju, i odnose se na stabilnost genotipova u

populaciji (Hardi Vajnbergova ravnoteža), diskretnost populacije (izolovanost), gametsku ravnotežu, nasumičnost razmnožavanja, upliv subjektivnih kriterijuma za odabir mesta uzorkovanja itd.

1.8.1.1. Osnovne definicije u verovatnoći i statistici

Razumevanje matematičkih modela u populacionoj genetici je uslovljeno razumevanjem osnovnih pojmova u verovatnoći i statistici, jer na populacionom nivou nije moguće sprovesti strogo kontrolisane eksperimente, niti izbeći generalizacije. Nasuprot determinističkim eksperimentima kada se ishodi ponavljaju pri konstantnim uslovima izvođenja eksperimenta, kod slučajnih ili stohastičkih eksperimenata, ishod događaja nije moguće predvideti već samo odrediti matematičku funkciju po kojoj se ostvarivanje događaja prevodi u verovatnoću (P). To ne znači da su u slučajnim eksperimentima svi ishodi mogući, naprotiv, bez obzira da li je prebrojiv, postoji skup svih mogućih ishoda i on se naziva prostorom slučajnih događaja (Ω), a svaki njegov element slučajni ishod odnosno elementarni slučajni događaj (ω). Osnovna karakteristika elementarnih događaja jeste da se pri slučajnom eksperimentu može ostvariti jedan i samo jedan elementaran događaj bez obzira što on nekada može biti skup više elemenata (Mališić 1997).

Tako npr. genotipizacijom bialelskog lokusa L u diploidnom organizmu imamo četiri ishoda A_1A_1 , A_1A_2 , A_2A_2 , A_2A_1 od kojih svaki ima jednaku verovatnoću 0.25. Ukoliko su ishodni događaji homozigot A_1A_1 , homozigot A_2A_2 i heterozigot, imamo tri ishodna događaja gde je jedan (heterozigot) unija dva elementarna događaja. Ovde vidimo da je definisanje ishoda ili elementarnog događaja unapred poznato, pre samog opita.

Skup svih događaja (\mathcal{F}), gde je svaki događaj skup koji može zauzeti ni jedan ili više ishoda.

Prostor verovatnoće je matematički konstrukt, tj. model koji reprezentuje skup slučajnih događaja. Ovaj konstrukt se upravo sastoji od tri komponente, skupa događaja, uzoračkog prostora i verovatnoće ostvarivanja nekog događaja (Ω, \mathcal{F}, P).

Prema klasičnoj teoriji verovatnoće i posmatrano u prostoru verovatnoće, podskup koji je presek dva skupa kojima su definisani događaj A i događaj B u redukovanom univerzumu/prostoru elementarnih događaja, jednak je verovatnoći ostvarivanja oba događaja ($P(A \cap B) = P(AB)$). Međutim, u slučajevima kada se izražava ostvarenje jednog događaja (npr. A) pod uslovom ostvarivanja drugog (npr. B), uzima se u obzir presek skupova A i B samo ukoliko je prethodno ostvaren događaj B i matematički formuliše kao $P(A|B)$ (Gentle 2013).

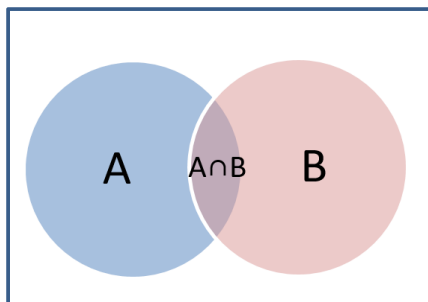
1.8.1.2. Bejzova statistika

Za razliku od klasičnog (frekvencionističkog) pristupa verovatnoći, Bejzov pristup se oslanja na subjektivnu teoriju verovatnoće u kojoj intuicija i stepen uverenja o mogućnosti realizacije nekog događaja imaju primat. Ovaj koncept je razvijan od 18. veka i pronalaskom rada Tomasa Bejza nakon čega se uvodi koncept uticaja 'prethodnog uverenja' o verovatnoći nekog događaja u određivanju posteriorne verovatnoće da se taj događaj ostvario. Neretko se ovaj pristup definiše i kao inverzna verovatnoća s obzirom da se vrednosti ulaznih parametara procenjuju na osnovu izlaznih vrednosti u vidu dobijenih podataka (Gentle 2013). Da bismo razumeli ovaj koncept, potrebno je definisati pojmove od kojih je sačinjen.

Uslovna verovatnoća (eng Likelihood)

Neka su A i B događaji iz istog prostora verovatnoća (Ω, F, P) i neka je $P(B) > 0$. Tada je uslovna verovatnoća događaja A , ako se ostvario događaj B , jednaka odnosu između preseka dva događaja $(A \cap B)$ i samog događaja (B) :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

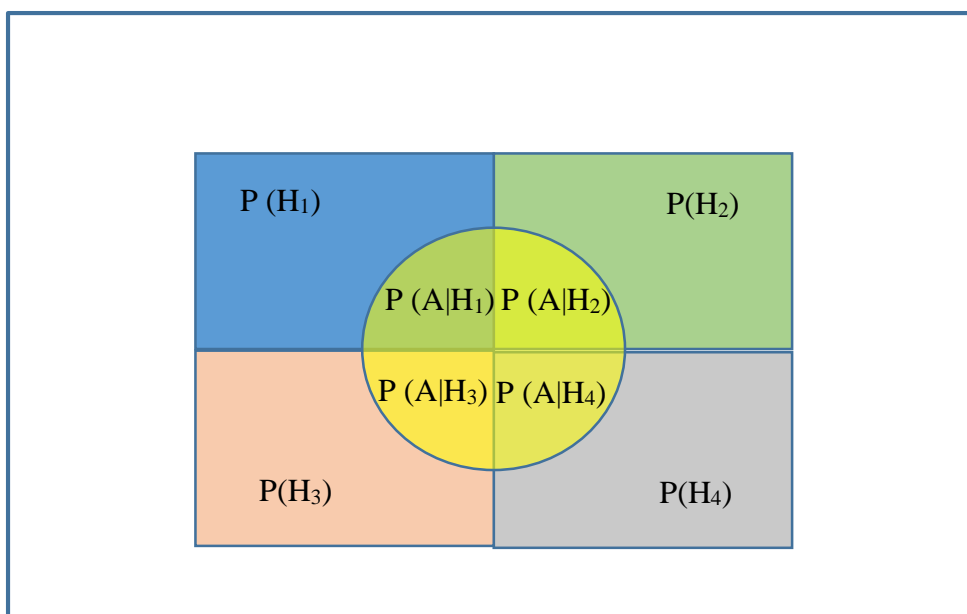


Slika 10 – presek skupa dva događaja

Potpuna verovatnoća

Neka su H_1, H_2, \dots, H_n međusobno isključivi slučajni događaji koji čine potpun sistem događaja i neka je A neki događaj iz istog prostora elementarnih ishoda. Verovatnoća događaja A se može izračunati po formuli:

$$P(A) = \sum_{j=1}^n P(H_j)P(A|H_j)$$



Slika 11 – Grafički prikaz odnosa četiri skupa događaja

, gde su $P(H_j)$ apriorne, odnosno, prethodne verovatnoće, $P(A|H_j)$ uslovne verovatnoće ako je hipoteza (H_j) tačna.

Posteriorna (ukupna) verovatnoća, odnosno mera uverenja u tačnost hipoteze na osnovu dobijenih podataka A , definiše se kao $P(H_k|A)$ ili jednostavno opaženi rezultati nekog eksperimenta.

Bejzova teorema

Kako su uslovne verovatnoće $P(A|B) = \frac{P(A \cap B)}{P(B)}$ i $P(B|A) = \frac{P(A \cap B)}{P(A)}$, vidimo da se $P(A \cap B)$ može izraziti na dva načina: $P(A \cap B) = P(A|B)P(B)$ ili $P(A \cap B) = P(B|A)P(A)$.

Ovo dalje implicira jednakost $P(A|B)P(B) = P(B|A)P(A)$, kao i da se uslovna verovatnoća događaja A može izraziti i kao:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

, kada to primenimo na veći broj međusobno isključivih događaja, odnosno hipoteza H_1, H_2, \dots, H_n pod kojima se događaj A može ostvariti, imenilac je zapravo suma verovatnoća svih H_j događaja:

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum_{i=1}^n P(H_i)P(A|H_i)}$$

, a kako se H_1, H_2, \dots, H_n ne mogu dogoditi istovremeno, suma svih uslovnih verovatnoća događaja A , odnosno $\sum_{i=1}^n P(H_i)P(A|H_i)$ se može skratiti u $P(A)$, koja se naziva marginalnom verovatnoćom i ne zavisi od tačnosti hipoteze. Tako da sada jednačina dobija formu:

$$P(H|A) = \frac{P(H)P(A|H)}{P(A)}$$

Može se reći da je ovo izraz uslovne verovatnoće da se događaj (hipoteza) H ostvario ako znamo da je ostvaren događaj A (opservirani podaci). U brojiocu se nalaze dva elementa, prethodna verovatnoća $P(H)$, odnosno prethodno uverenje o ostvarivanju događaja H (hipoteze) i verodostojnost opažanja podataka pod uslovom važenja H (hipoteze), odnosno funkcija verodostojnosti podataka pod uslovom važenja hipoteze.

$$P(\text{Hipoteza}|\text{Podaci}) = \frac{P(\text{Hipoteza})P(\text{Podaci}|\text{Hipoteza})}{P(\text{Podaci})} = \frac{\text{prethodna verovatnoća} \times \text{verodostojnost}}{\text{marginalna verovatnoća}}$$

U deliocu, marginalna verovatnoća događaja A predstavlja dokaz hipoteze i ne utiče na promene $P(H|A)$ drugačije nego kao konstanta normalizacije tako da $P(H|A)$ ne može biti veće od 1. Zahvaljujući tome, Bejzovu teoremu možemo redukovati na proporcionalnost:

$$P(\text{Hipoteza}|\text{Podaci}) \propto P(\text{Hipoteza})P(\text{Podaci}|\text{Hipoteza}).$$

Likelihood funkcija (funkcija verodostojnosti)

Ukoliko je skup promenljivih $X=(X_1, X_2, \dots, X_n)$ odraz funkcije $f(x|\theta)$, gde je θ parametar koji se procenjuje, a x opažena vrednost X , onda je likelihood funkcija određena kao $L(\theta|X)=f(x|\theta)$.

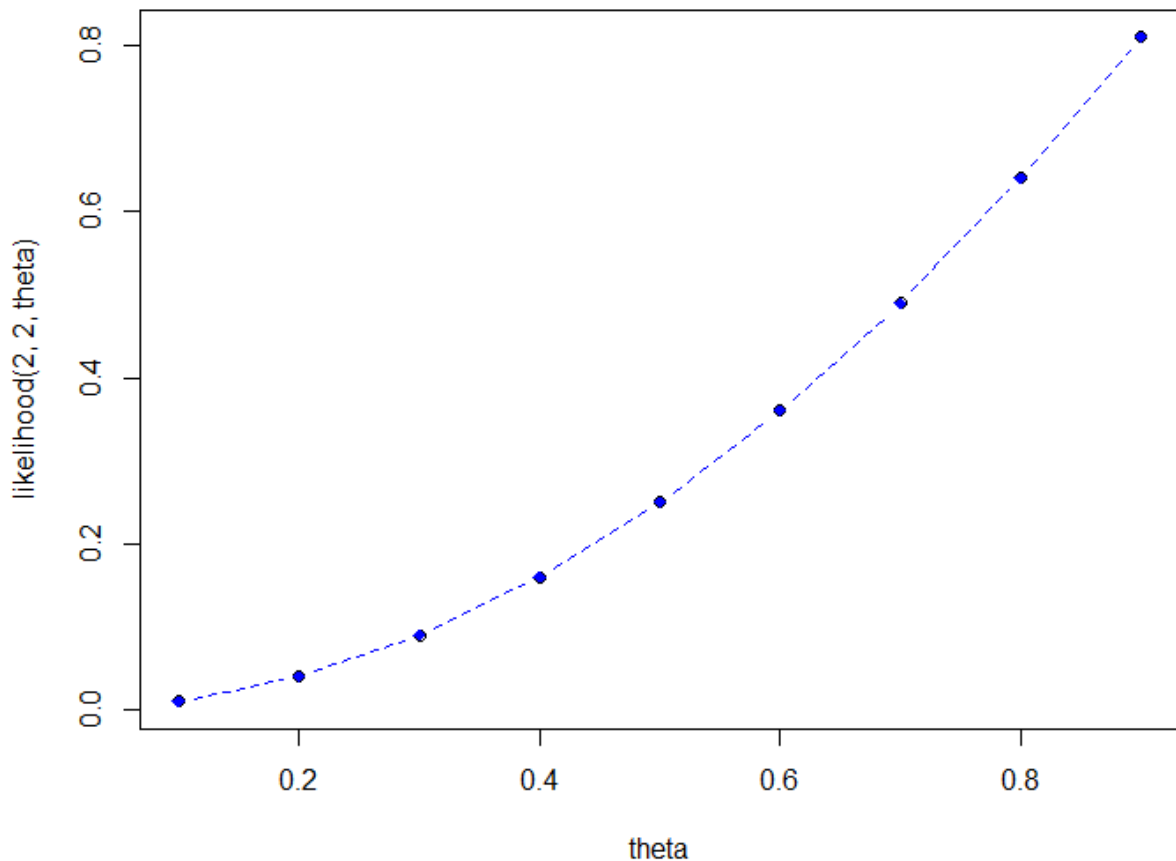
Za razliku od klasičnog pristupa u verovatnoći i statistici, kada se parametar θ prihvata kao nepromenljiva (npr. centralna tendencija), a ishodna promenljiva kao variranje oko vrednosti tog parametra, Bejzov pristup je reverzan (obrnut) i ishodne vrednosti uzima kao nepromenljive, dok

parametar procenjuje kao promenljivu na osnovu funkcije verodostojnosti (Edward 2017). Drugim rečima, prilikom procene nekog parametra θ na osnovu ishodnih promenljivih, postoji neizvesnost u toj proceni, te parametar θ može zauzeti veći broj vrednosti sa različitim verovatnoćama od kojih su neke veće, a neke manje: $L(\theta_j | x) > L(\theta_k | x)$.

Ilustracija simulacijom u R programskom jeziku:

Ukoliko je na uzorku od dva hromozoma na bialelskom lokusu i opažen samo alel A, funkcija verodostojnosti će izgledati kao na grafiku:

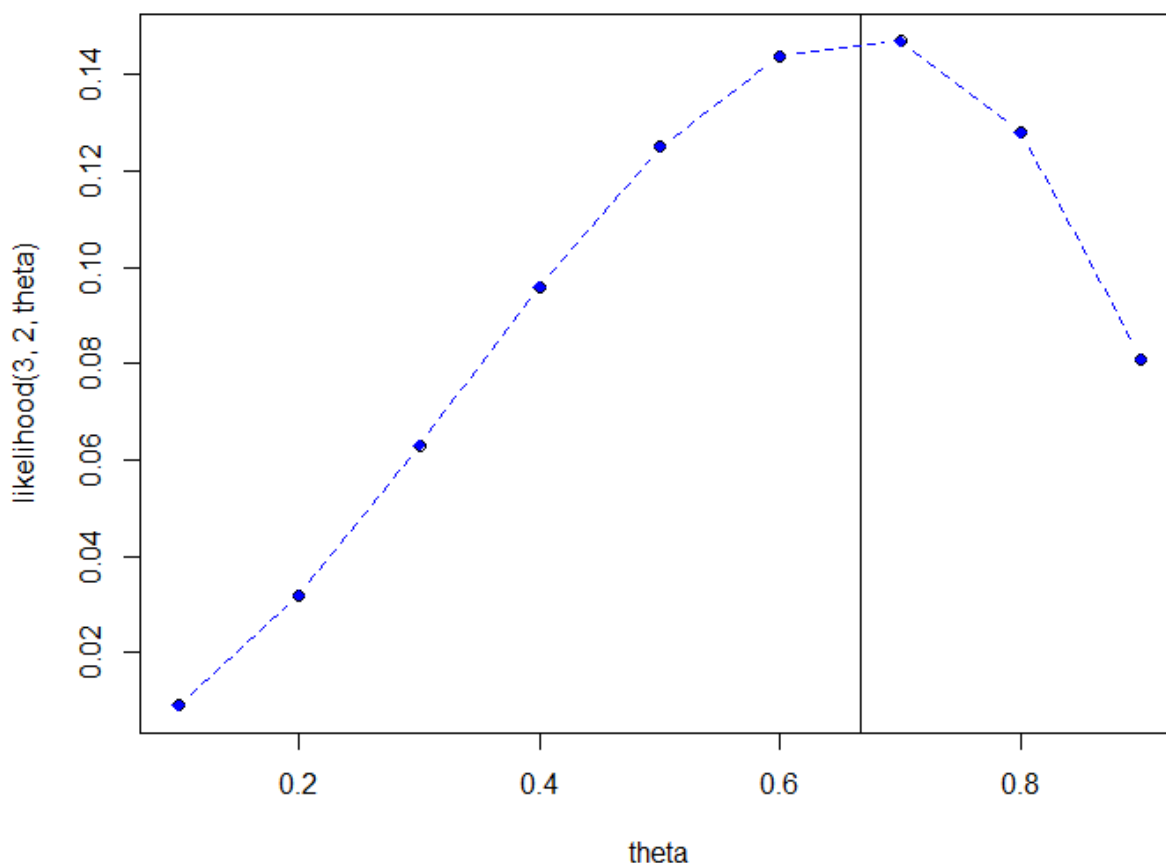
```
> likelihood=function(n,y,theta){return(theta^y*(1-theta)^(n-y))}
> theta=seq(from=0.1, to=0.9, by=0.1)
> plot(theta, likelihood(2,2, theta))
> lines(theta, likelihood(2,2,theta),pch=18, col="blue", type="b", lty=2)
```



Slika 12 – Grafički prikaz zavisnosti verodostojnosti i θ parametra.

Ukoliko je na lokusu i nakon uzorka opažen i alel T pored dva alela A, funkcija verodostojnosti menja svoj oblika kao:

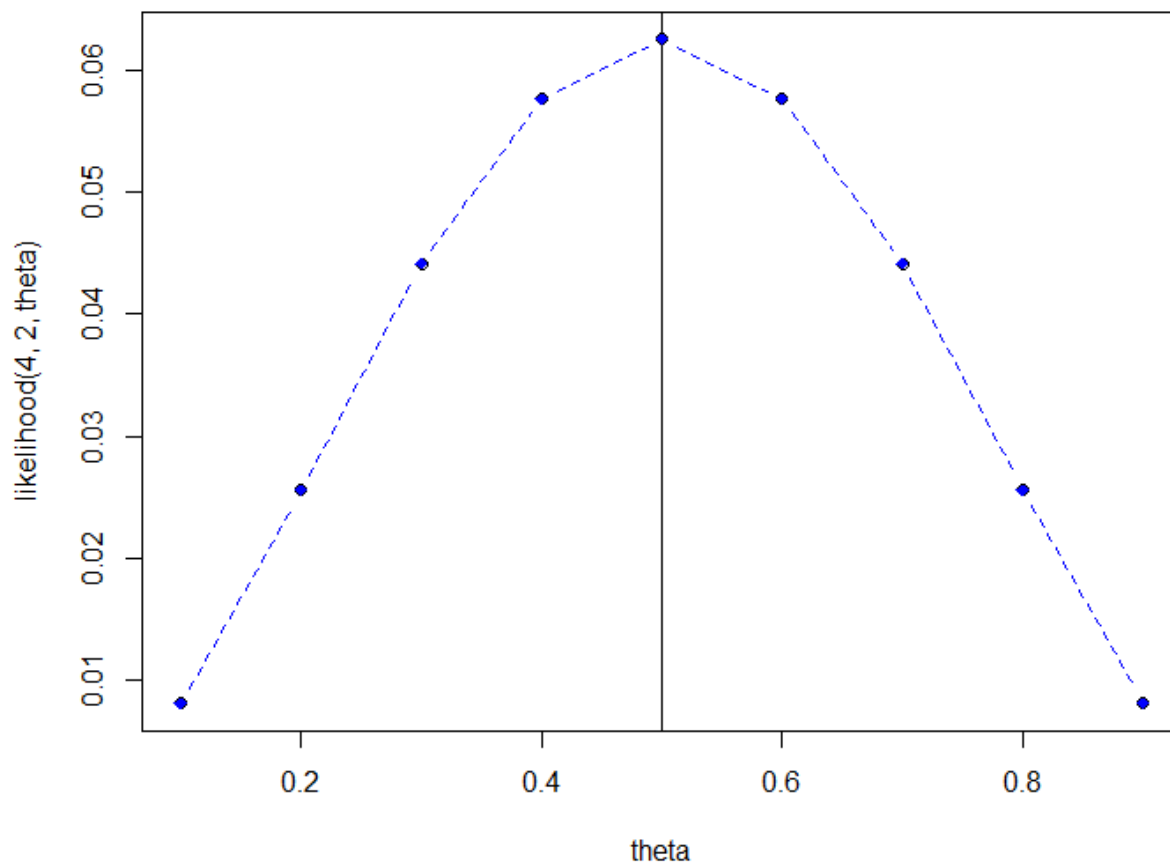
```
> likelihood=function(n,y,theta){return(theta^y*(1-theta)^(n-y))}
> theta=seq(from=0.1, to=0.9, by=0.1)
> plot(theta, likelihood(3,2, theta))
> lines(theta, likelihood(3,2,theta),pch=18, col="blue", type="b", lty=2)
> abline(v=2/3)
```



Slika 13 - Grafički prikaz zavisnosti verodostojnosti i θ parametra.

Ukoliko je nakon uzorkovanja četvrtog hromozoma na lokusu i opažen još jedan alel T, tako da ceo uzorak ima skup alela (AATT), funkcija verodostojnosti ima najvišu vrednost za $\theta=0.5$:

```
> likelihood=function(n,y,theta){return(theta^y*(1-theta)^(n-y))}
> theta=seq(from=0.1, to=0.9, by=0.1)
> plot(theta, likelihood(4,2, theta))
> lines(theta, likelihood(4,2,theta),pch=18, col="blue", type="b", lty=2)
> abline(v=0.5)
```



Slika 14 - Grafički prikaz zavisnosti verodostojnosti i θ parametra.

Marginal likelihood-verovatnoća opažanja podataka (B, raspodele učestalosti alela) ukoliko je hipoteza tačna ili nije tačna (A, broj klastera)(Brewer 2017).

Beta raspodela $Beta(\alpha, \beta)$

Familija *Beta* raspodela pripada neprekidnim raspodelama verovatnoće, sa specifičnošću da se uvek nalazi u intervalu između 0 i 1 ($0 \leq x \leq 1$). Ova raspodela ima samo dva parametra, α i β , koji su veći od nule i određuju njen oblik (William M. Bolstad 2016). Zahvaljujući tome što se nalazi u intervalu od 0 do 1, *beta* raspodela omogućava modelovanje različitih raspodela podataka u raspodelu gustine verovatnoće, npr. binomnu raspodelu učestalosti bialelskih varijanti. Za ovu raspodelu se može reći da je ona verovatnoća verovatnoće, jer se i na apscisi i na ordinati opisuje verovatnoća. Opšti izgled *beta* raspodele je predstavljen kao:

$$g(x; \alpha, \beta) = \begin{cases} k \times x^{\alpha-1}(1-x)^{\beta-1} & \text{za interval } 0 \leq x \leq 1 \\ 0 & \text{za } x \notin [0,1] \end{cases}$$

, gde je k samo konstanta koja određuje površinu ispod krive, odnosno normalizuje raspodelu:

$$k = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

, a *gama* funkcija $\Gamma(c)$, za c prirodan broj, generalizacija faktorijalne funkcije $\Gamma(c) = (c-1) \times \Gamma(c-1)$, tako da kompletan izraz *beta* raspodele jeste:

$$g(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

, dok se srednja vrednost određuje integralom u intervalu $[0,1]$ bez obzira na α i β :

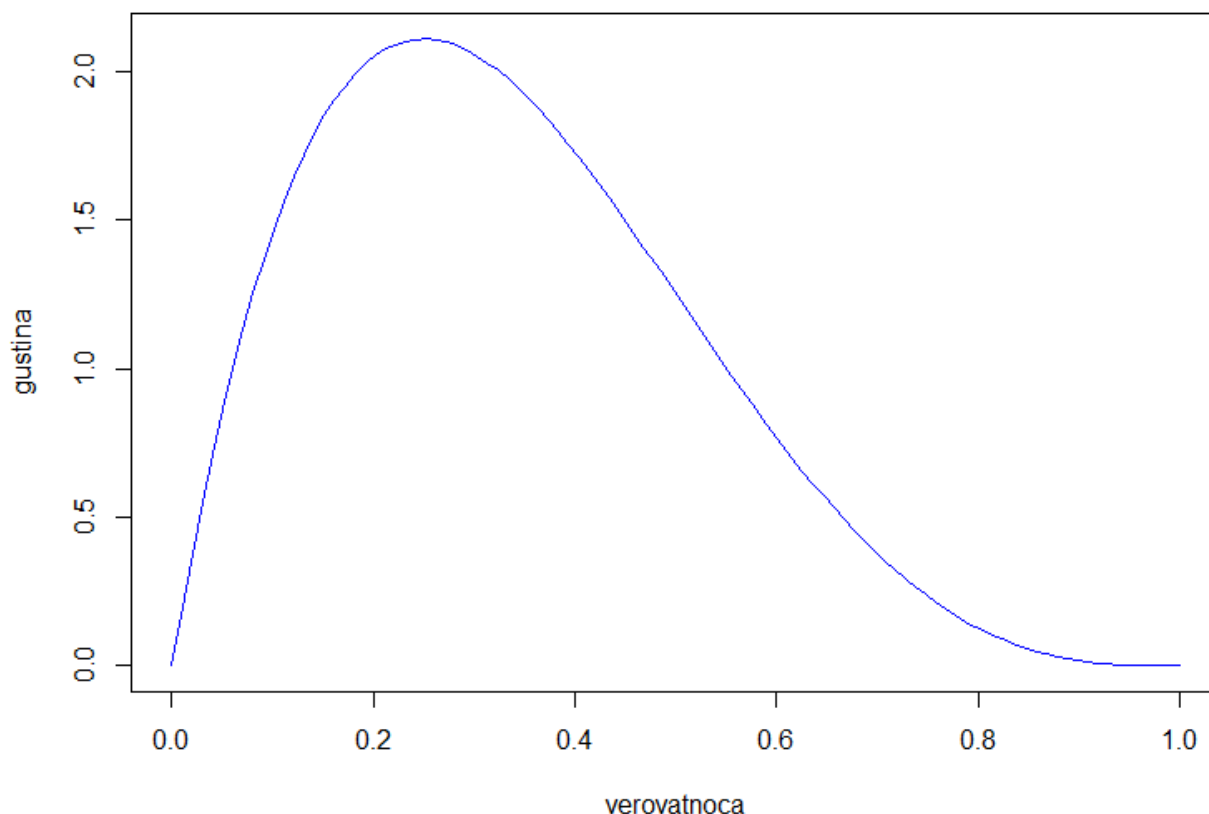
$$E[X] = \int_0^1 x \times g(x; \alpha, \beta) dx = \int_0^1 x \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx$$

, koji je jednak 1 i može se izvesti u jednostavnu jednačinu:

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

Varijansa *beta* funkcije ima oblik:

$$Var[X] = \frac{\alpha \times \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$



Slika 15 - Izgled beta funkcije za $\alpha=2$ i $\beta=4$

Ono što je karakteristično za *beta* raspodelu jeste da ona omogućava procenu većeg broja parametara na osnovu vrednosti α i β , jer se iz njih mogu izračunati srednja vrednost, mod, varijansa itd. Pored toga, glavna karakteristika koja je čini podobnom za Bejzovu statistiku jeste mogućnost ažuriranja prethodnih uverenja na osnovu dobijenih eksperimentalnih podataka. Ako ponovo pogledamo jednačinu za procenu srednje vrednosti:

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

, vidimo da se svakom promenom (svakim ažuriranjem) parametara α i ili β , menja i parameter \hat{X} , dok se rasipanje oko centralne tendencije smanjuje, jer prema jednačini:

$$Var[X] = \frac{\alpha \times \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

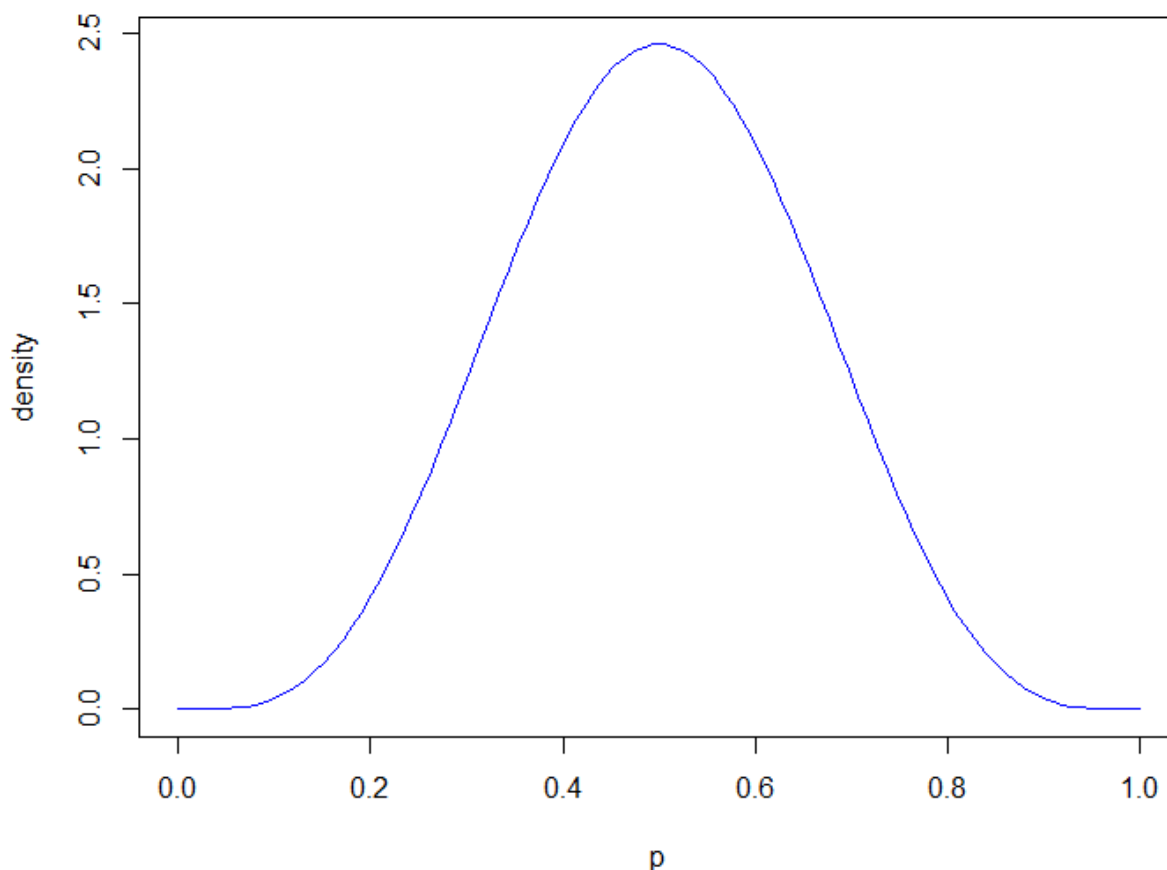
, imenilac ima komponentu koja je zavisna od kvadratne funkcije. Drugim rečima, pored stalnog ažuriranja posteriorne raspodele odnosno povećanja njene tačnosti, istovremeno se povećava i njena preciznost.

Primer u genetici-učestalost alela u populaciji

Ukoliko bialeleskom lokusu i gde su moguće varijante A i T , uzorkujemo 10 hromozoma iz testirane populacije i genotipizacijom dobijemo 10 A i ni jednu T varijantu, imaćemo monotonu raspodelu podataka sa učestalostima $A=1$ i $T=0$.

Međutim, ukoliko znamo da je u okolnim populacijama učestalost manje učestale varijante (MAF)=0.5, zbog geografske bliskosti je legitimno da i taj podatak uključiti kao prethodno uverenje o očekivanoj vrednosti $A=0.5$ i $T=0.5$ u testiranoj populaciji. U tom slučaju, $beta$ raspodela će dobiti oblik kao na grafiku ispod (slika16):

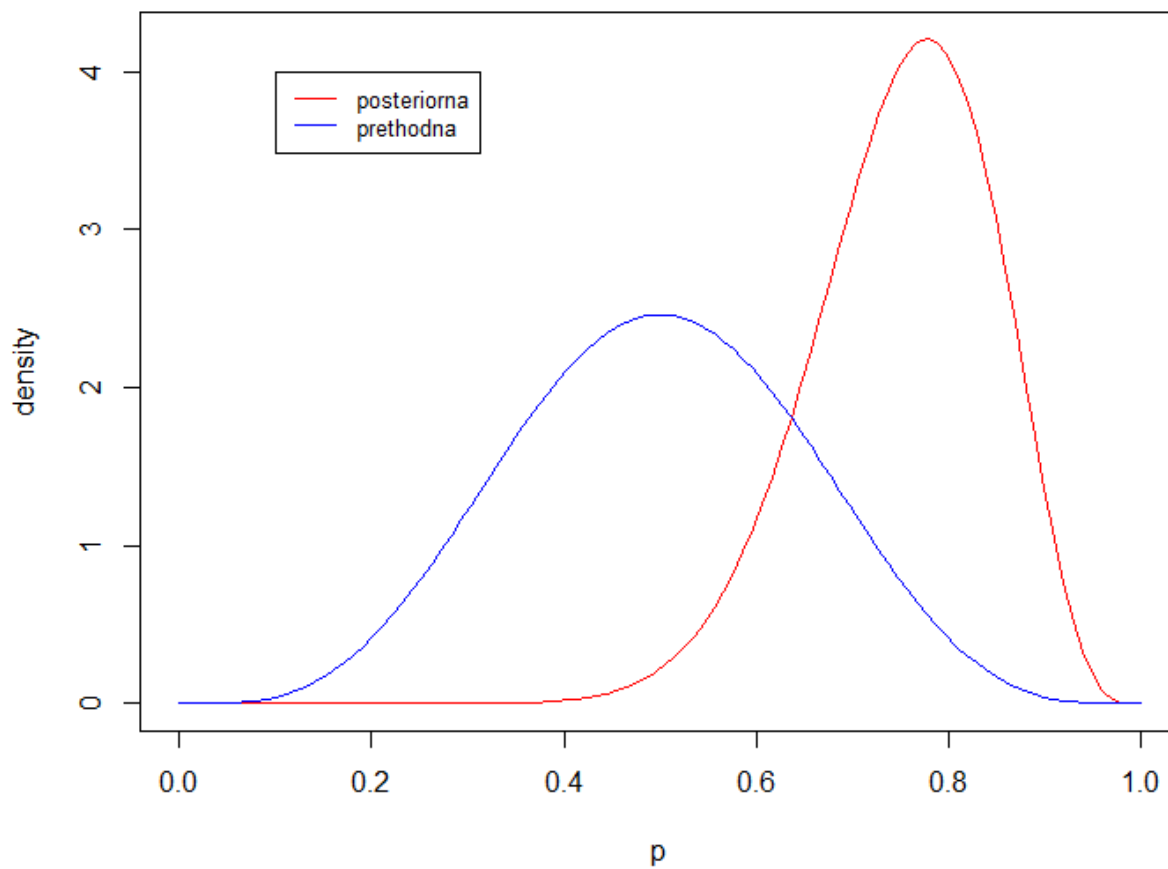
```
> p = seq(0,1, length=100)
> plot(p, dbeta(p, 5, 5), ylab="density", type="l", col=4)
```



Slika 16 - Beta raspodela za MAF=0.5

, koja se zatim ažurira dobijenim podacima eksperimenta ($A=10$ i $T=0$), tako da posteriorna $beta$ raspodela uzima parametre $\alpha=\alpha+10$ i $\beta=\beta+0$ (slika 17).

```
> plot(p, dbeta(p, 15, 5), ylab="density", type="l", col=2)
> lines(p, dbeta(p, 5, 5), col=4)
> legend(0.1, 4, legend=c("posteriorna", "prethodna"), col=c("red", "blue"), lty=1:1, cex=0.8)
```



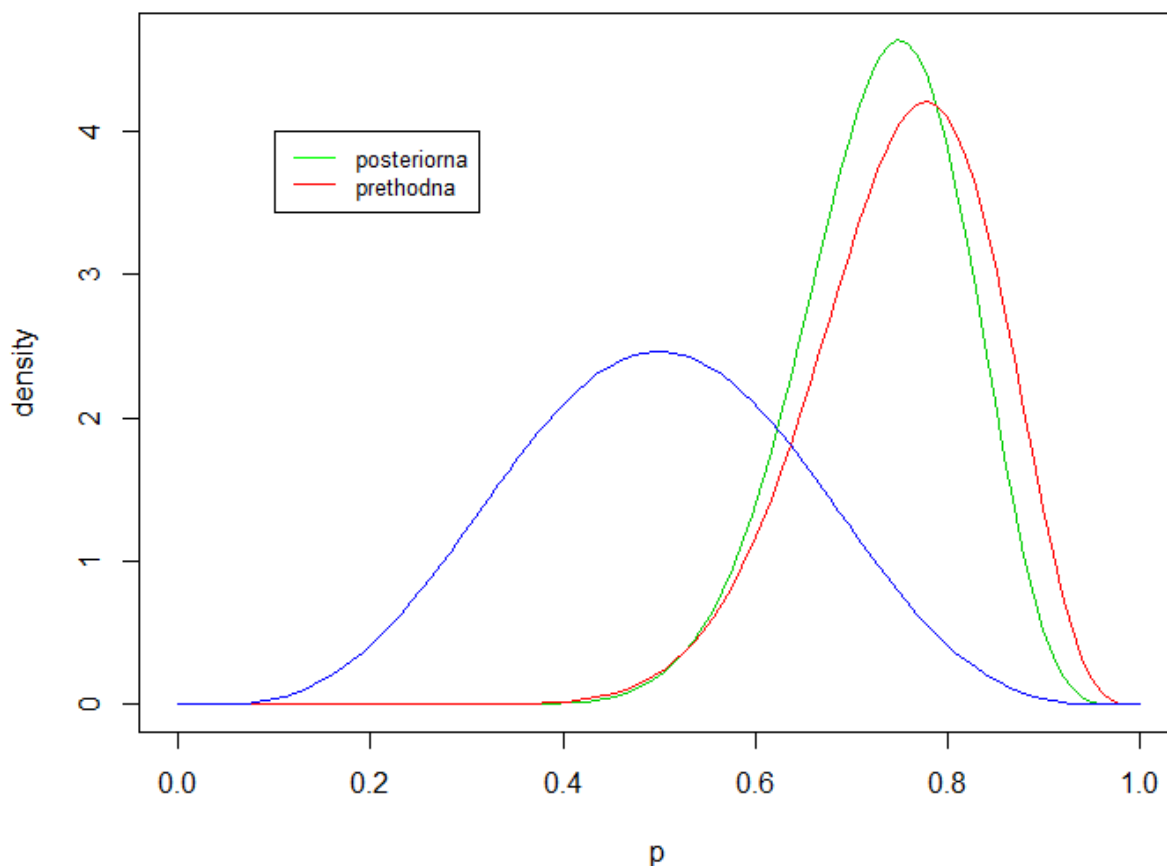
Slika 17 – Odnos posteriorne i prethodne verovatnoće.

Daljim eksperimentom pored 10 *A* varijanti, genotipiziramo još 3 osobe i uočimo 4 *A* i 2 *T* varijante. Na osnovu novih podataka, ažuriramo *beta* raspodelu sa parametrima $\alpha=\alpha+4$ i $\beta=\beta+2$ i dobijamo novu raspodelu (slika 18).

```

> plot(p, dbeta(p, 19, 7), ylab="density", type="l", col=3)
> lines(p, dbeta(p, 15, 5), col=2)
> lines(p, dbeta(p, 5, 5), col=4)
> legend(0.1, 4, legend=c("posteriorna", "prethodna"), col=c("green", "red"), lty=1:1, cex=0.8)

```



Slika 18 – Odnos posteriorne i prethodne verovatnoće.

Možemo primetiti da je procena posteriorne verovatnoće na osnovu prethodnih uverenja realnija nego što bi bila binomna raspodela $B(14, 0.125)$ za vrednosti $A=14$, $T=2$. Takođe, možemo primetiti da se *beta* funkcija ponaša tako da sa povećanjem broja eksperimenata varijansa oko centralne tendencije opada u skladu sa jednačinom:

$$Var[X] = \frac{\alpha \times \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

, a sa njom i kredibilni interval (ekvivalent intervalu poverenja u klasičnoj statistici).

Dirihleova raspodela

Opšti izraz *beta* raspodele je pogodan za modelovanje binomnih ishoda, odnosno kada su prisutna dva faktora koji utiču na raspodelu verovatnoće ili podataka. Kada je u pitanju primena *beta* raspodele u modelovanju multivarijantnih podataka, kao što su npr. multialelski lokusi, primenjuje se posebana generalizacija ove raspodele poznata kao Dirihleova raspodela gustine verovatnoće. Odnos između *beta* i Dirihleove raspodele je analogan odnosu binomne i multinominalne raspodele.

Razlog podobnosti Dirihleove distribucije jeste taj što ona procenjuje raspodelu verovatnoće parametara $\theta_1, \theta_2, \dots, \theta_n$ na osnovu opaženih podataka, tako da je suma svih vrednosti svih mogućih parametra 1.

$$\sum_{i=1}^n \theta_i = 1$$

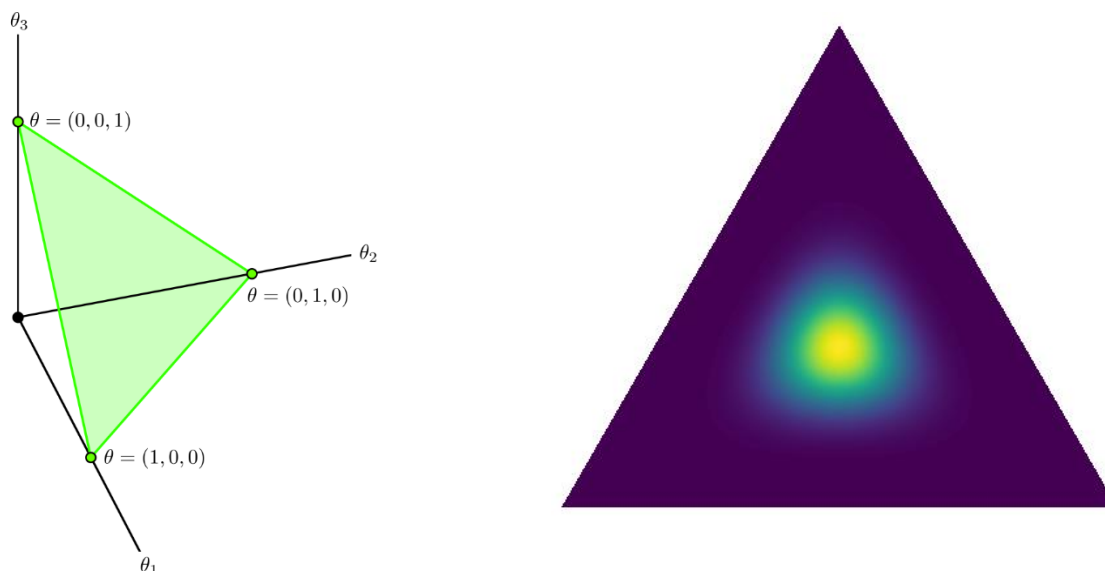
Funkcija K-dimenzionalne gustine Dirihleove raspodele jeste:

$$f(x) = \frac{\prod_{k=1}^K \Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^K \alpha_i)} \prod_{k=1}^K x_k^{\alpha_k-1}$$

, gde je Γ gama funkcija.

Ulazna vrednost Dirihletove raspodele je K-dimenzionalni vektor α sa vrednostima α_k koje su pozitivni i realni brojevi. U slučajevima kada vektor α ima dve dimenzije, Dirihletova raspodela je jednaka *Beta* raspodeli za dva parametra. Npr. ukoliko modelujemo bialelski polimorfizam, imamo samo dva parametra (učestalosti alela 1 i alela 2) koji se mogu svesti na jedan, jer je stepen slobode $2-1=1$ (kako učestalosti u zbiru imaju vrednost 1, određivanjem učestalosti jednog alela, određena je i učestalost drugog alela). Ukoliko pak modelujemo multialelske polimorfizme, imamo multinominalnu raspodelu učestalosti alela na jednom lokusu i to možemo posmatrati kao parametar $\alpha=[\text{učestalost alela}_1, \text{učestalost alela}_2, \dots, \text{učestalost alela}_n]$. Kako to nije vrednost parametara koji se procenjuju (verovatnoća), Dirihletova raspodela se normalizuje u raspodelu gustine verovatnoće.

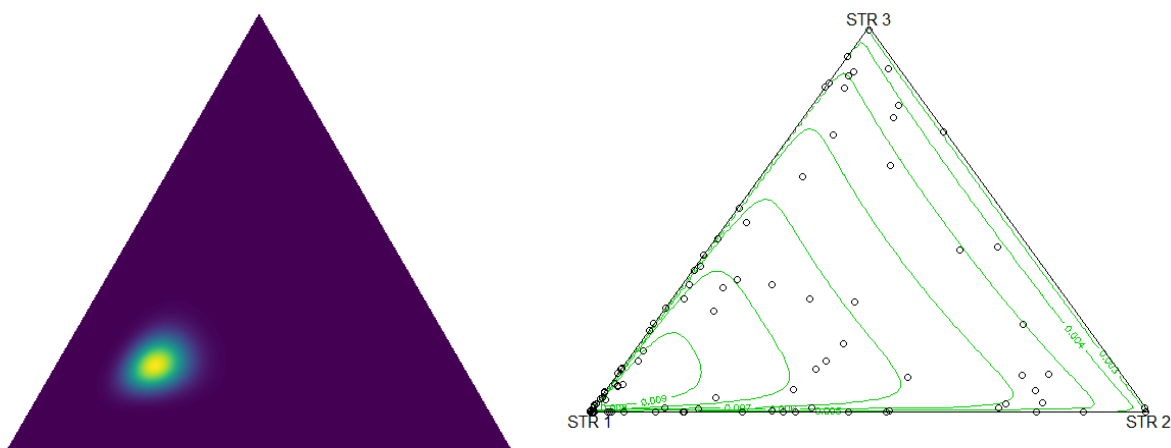
Dirihletova raspodela se može grafički predstaviti kao k-simpleks verovatnoće. Na slici 19 je se nalazi pojednostavljeni grafički prikaz raspodele za parametar $\alpha=[1,1,1]$, odnosno trodimenzionalni vektor kada su verovatnoće ostvarivanja događaja jednake.



Slika 59 – trodimenzionalni prikaz simpleksa (grafičkog prikaza za tri vrednosti) verovatnoće Dirihletove raspodele za tročlani vektor Θ

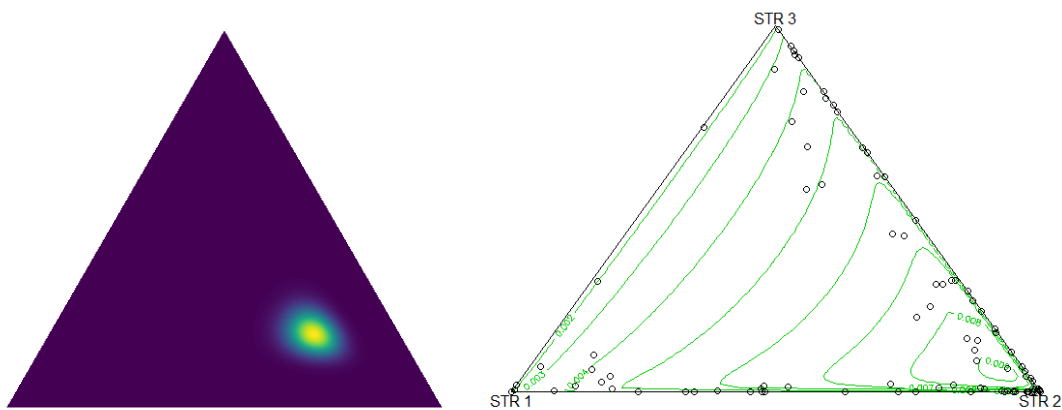
Primer u genetici-Dirihleova raspodela učestalosti mikrosatelita

Na mikrosatelitskom lokusu i gde su moguće tri varijante uzastopnih tandemskih ponovaka $STR 1$, $STR 2$ i $STR 3$, uzorkujemo 50 osoba (100 hromozoma) iz testirane populacije, i genotipizacijom dobijemo odnose 60:20:20, Dirihleova raspodela bi izgledala kao na slici 20:



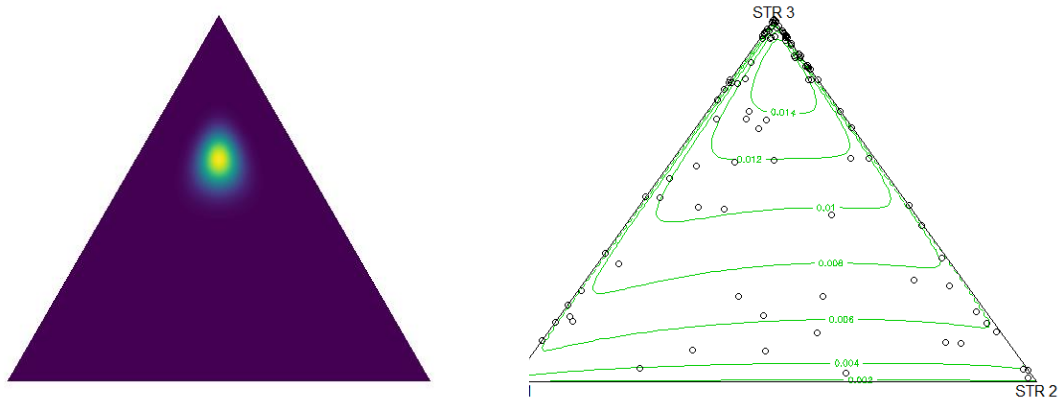
Slika 20 – Grafički prikaz Dirihleove raspodele.

, ukoliko bi taj odnos bio 20:60:20, grafički prikaz raspodele bi bio kao na slici 21:



Slika 21 – Grafički prikaz Dirihleove raspodele.

, dok bi za odnos 20:20:60 raspodela izgledala kao na slici 22:



Slika 22 – Grafički prikaz Dirihleove raspodele.

Konjugacija raspodela verovatnoće

Pod terminom konjugacija raspodela verovatnoće u Bejzovom modelovanju se podrazumeva odnos između prethodne raspodele (npr. *beta*), funkcije verodostojnosti, konstante normalizacije i posteriorne raspodele verovatnoće (takođe *beta*), tako da obe raspodele ostaju u istoj familiji distribucija verovatnoće. Familija raspodela verovatnoće je skup definisan kao:

$$P = \{P_\theta: \theta \in \Theta \subseteq R^s\}$$

u prostoru elementarnih događaja (Ω), gde je P_θ funkcija raspodele parametra θ , a R skup realnih brojeva.

S obzirom da u Bejzovoj statistici prethodna i posteriorna raspodela pripadaju *beta* familiji, konjugacija se ostvaruje preko funkcije verodostojnosti i konstante normalizacije.

U slučaju podataka gde je bilo više od dva faktora koja utiču na raspodelu podataka (npr. multialelski lokus), konjugacija raspodela je između multinomne i Dirihleove raspodele. Ovaj način modelovanja koji se primenjuje u Bejzovom pristupu, između multinomne i Dirihleove raspodele, se naziva i Dirihle-kategorijalni model (kategorijalni jer se multinomna raspodela odnosi na n kategorijalne podatke). Ako npr. imamo lokus i sa tri moguća alela koji imaju jednake učestalosti $\frac{1}{3}$, možemo posmatrati kao vektor $\Theta = [\Theta_1, \Theta_2, \Theta_3]$, onda bismo očekivali da u populaciji iz koje na slučajan način uzorkujemo i zatim genotipiziramo lokus i , dobijemo raspodelu obzerviranih učestalosti alela $X = [x_1, x_2, x_3]$. Na osnovu multinomne raspodele očekivali bismo da verovatnoću tačnog ishoda možemo izračunati na osnovu funkcije:

$$f(x_1, x_2, \dots, x_k | n, q = (q_1, q_2, \dots, q_k)) = \frac{n!}{x_1! x_2! \dots x_k!} \prod_{i=1}^k q_i^{x_i}$$

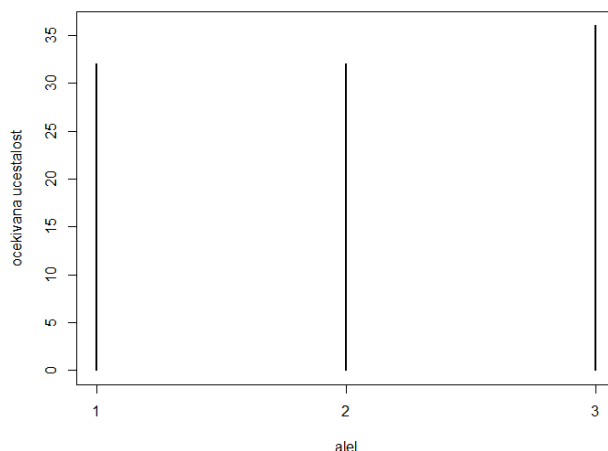
i da dobijene vrednosti x_1, x_2, x_3 ne budu drastično različite od $\frac{1}{3}$ (ilustracija u R):

```
> str.lokus <- function(){
+   lokus <- sample(1:3, size = 1, replace = TRUE)
+   return(lokus)
+ }
> for (i in 1:4){
+   simulacija <- replicate(100, str.lokus())
+   table(simulacija)
```

```

+ table(simulacija)/length(simulacija)
+ plot(table(simulacija), xlab = 'alel', ylab = 'očekivana učestalost')
+ }

```



Slika 23 – Stubičasti prikaz očekivanih učestalosti tri promenjive.

Međutim, ukoliko ne znamo vrednosti $\Theta=[\Theta_1, \Theta_2, \Theta_3]$, Bejzovim pristupom možemo proceniti verovatnoću parametra θ na osnovu opaženih rezultata i prethodnog uverenja o raspodeli Θ .

$$P(\theta|\alpha) \propto \prod_{i=1}^n \theta_i^{\alpha_i-1}$$

, pod uslovom da je $\sum_{i=1}^n \theta_i = 1$ i da je $\theta_i \geq 0$. Dirihleova raspodela se konjuguje kao prethodna raspodela za multinomnu, preko funkcije verodostojnosti koja je produkt uslovnih verovatnoća opažanja podataka D pod θ :

$$P(D|\theta) = \prod_{i=1}^n P(X_i = x_i | \theta)$$

, što se može prevesti u produkt uslovnih verovatnoća sa eksponentom b =vector sa brojem podataka:

$$P(D|\theta) = \prod_{i=1}^n \theta_i^{b_i}$$

Primenom Bejzovog pravila, odnosno, da je posteriorna verovatnoća produkt prethodne i funkcije verodostojnosti:

$$P(\theta|D) = P(D|\theta) \times P(\theta)$$

, dobijamo produkt:

$$\prod_{i=1}^n \theta_i^{b_i+\alpha_i-1} \propto Dir(\theta|b + \alpha)$$

, gde je α vektor parametara raspodele prethodne verovatnoće. Ovo se može uopštiti u tvrdnju da produkt kategorijalne i Dirihleove distribucije se inkorporira u Dirihleovu:

$$(\prod_{i=1}^n Kat(x_i(\theta)))Dir(\theta|\alpha) = Dir(\theta|b + \alpha)$$

U Bejzovoj statistici ocena dokaza, odnosno testiranje hipoteze je isto što i procena parametara, što se drastično razlikuje od frekvencionističkog pristupa u statistici (Brewer 2017). Prema klasičnom (frekvencionističkom) pristupu testiranja hipoteze, statistikom testa se izračunava verovatnoća (p vrednost) opažanja dobijenih rezultata, ili još manje verovatnih, pod pretpostavkom važenja nulte hipoteze H_0 :

P vrednost = $P(\bar{x} \leq X|H_0)$, gde je \bar{x} prosečna vrednost, a ishod testiranja može biti:

-nulta hipoteza H_0 je tačna i sva odstupanja su proizvod slučajnosti

-nulta hipoteza H_0 je netačna i odstupanja su posledica faktora koji prouzrokuje drugačiju raspodelu podataka (William M. Bolstad 2016).

Bejzov pristup testira odnos posteriornih verovatnoća hipoteza od kojih je jedna nulta $H_0 (P(H_0|x))$, a druga alternativna $H_1 (P(H_1|x))$. S obzirom da su sve hipoteze međusobno isključive, onda je $P(H_1|x) = 1 - P(H_0|x)$, tako da je odnos ovih verovatnoća indeks verodostojnosti (LR, eng Likelihood ratio) i govori o tome koliko je puta jedna hipoteza verovatnija u odnosu na drugu (Brewer 2017).

U suštini, princip je vrlo jednostavan: na osnovu opaženih podataka, procenjuju se verovatnoće parametara i zatim procenjuje odnos parametra koji se smatrao da je najverodostojniji pod nultom hipotezom i parametra koji je najverodostojniji na osnovu opaženih podataka.

Ono što jeste glavna karakteristika Bejzovog principa je mogućnost ažuriranja prethodne verovatnoće nakon svakog eksperimenta (pod eksperimentom se podrazumeva svako nezavisno uzorkovanje dokaza). Ako pogledamo jednačinu:

$$P(\text{Posteriorna verovatnoća}) = \frac{\text{prethodna verovatnoća} \times \text{verodostojnost}}{\text{marginalna verovatnoća}}$$

, jednostavno, nakon svakog pojedinačnog eksperimenta, posteriorna verovatnoća postaje prethodna verovatnoća:

$$P(\text{posteriorna verovatnoća}) = \frac{\text{prethodna verovatnoća} \times \text{verodostojnost}}{\text{marginalna verovatnoća}}$$

Metropolis-Hestingov (Markovljev lanac Monte Karlo) metod (eng Metropolis-Hastings Markov chain Monte Carlo)

Markovljev lanac Monte Karlo (MCMC) je klasa metoda za uzorkovanje iz empirijske raspodele u cilju aproksimacije raspodele verovatnoće parametara iz kojih je empirijska raspodela nastala. (Hastings 1970)

Za kompleksne podatke, normalizacija u Bejzovoj teoremi je takođe kompleksna, nekada je i nemoguće izračunati konstantu $P(A)$, odnosno, marginalnu verovatnoću dokaza, jer je potrebno integraliti prostor ispod višedimenzionalne raspodele verovatnoće. Kako je u Bejzovom modelovanju osnovni cilj odrediti raspodelu verovatnoće parametara, posebna forma MCMC metoda, Metropolis-Hestingov algoritam, koristeći nasumično kretanje u prostoru verovatnoće, može opisati raspodelu posteriorne verovatnoće na osnovu opaženih podataka, prethodne verovatnoće i funkcije verodostojnosti. Nasumično kretanje je komponenta iz Monte Karlo principa, a ideja je da iterativno

uzorkovanje iz predložene raspodele $\pi(x)$ generiše raspodelu koja je proporcionalna predloženoj raspodeli $\pi(x) \propto \tilde{\pi}(x)$.

Druga komponenta je Markovljev lanac, koji omogućava da početak kretanja može biti bilo koja vrednost iz raspodele verovatnoće (Hastings 1970). Markovljev lanac je niz brojeva (x_1, x_2, \dots, x_n) , gde x_n zavisi samo od x_{n-1} što se uvodi kao princip u uzorkovanju tako da svaki novi parametar raspodele služi kao predložena osnova za uzorkovanje sledećeg parametra. Na taj način se prilikom svake iteracije (nasumičnog koraka) predložena raspodela za uzorkovanje, zapravo menja.

Ipak, Monte Karlo algoritam ne može generisati posteriornu raspodelu, jer ne proračunava koje su vrednosti parametra θ više verovatne. Zato se uvodi Metropolis-Hastings komponenta u odlučivanju koju predloženu vrednost parametra θ treba prihvatiti ili odbaciti, kada oblik posteriorne raspodele nije poznat (Huber 2016).

MHMCMC opisuje relativne odnose između verovatnoća vrednosti parametra zahvaljujući proporcionalnosti u imeniocu Bejzove teoreme:

$$P(\text{Parametar}|\text{Podaci}) \propto P(\text{Parametar})P(\text{Podaci}|\text{Parametar})$$

Kako se za svaku vrednost parametara u statističkom modelu može odrediti verodostojnosti kao $P(\theta|X)$, gde $P(X)$ ima svoju prethodnu verovatnoću, tako se i za svaki predloženi parametar θ_i može odrediti verodostojnost množenjem sa prethodnom verovatnoćom $L = P(\theta_i|X_i) \times P(X_i)$. Na taj način se određuje jedna konkretna predložena vrednost parametra u prostoru verovatnoće vrednosti parametara.

Opisivanje raspodele verovatnoće parametra od interesa, MCMC algoritam počinje sa bilo kojom vrednošću parametra θ_i i izračunava njegovu verovatnoću $P(X|\theta_i) \times P(X)$, gde su opaženi podaci označeni sa X , a zatim uzima novu vrednost θ_j iz predložene raspodele $q(X|Y)$ (može biti bilo koja raspodela, npr. za bialelske polimorfizme je to binomna, dok je za multialelske multinomna) i izračunava $P(\theta_j|X) \times P(X)$. U sledećem koraku se odmerava odnos r verovatnoća predložene vrednost θ_j i postojeće θ_i :

$$r(\theta_{new}, \theta_{t-1}) = \frac{\text{posterior}(\theta_{new})}{\text{posterior}(\theta_{t-1})} = \frac{\text{Beta}(a, b, \theta_{new}) \times \text{Bin}(n, p, \theta_{new})}{\text{Beta}(a, b, \theta_{t-1}) \times \text{Bin}(n, p, \theta_{t-1})}$$

Ukoliko je taj odnos veći ili jednak 1, predložena vrednost verovatnoće parametra θ_j se prihvata:

$$\alpha(\theta_{new}, \theta_{t-1}) = \min\{r(\theta_{new}, \theta_{t-1}), 1\}.$$

Ukoliko je $r < 1$, odnosno, verovatnoća predloženog parametra manje verovatna od postojećeg, odlučuje se o njenom prihvatanju tako što se iz uniformne raspodele uzorkuje neka vrednost i ako je ona niža od vrednosti r , odbacuje se vrednost predloženog parametra θ_j :

$$u \sim U(0,1), \text{ ako je } u < r, \theta_i = \theta_i; \text{ ako je } u > r, \theta_i = \theta_j.$$

Ponavljanjem ovog procesa, predlaganjem nove vrednosti parametra, ocenjivanjem verodostojnosti predloga, odbacivanjem ili prihvatanjem, opisuje se posteriorna raspodela parametra od interesa, što generiše uzorak iz posteriorne raspodele parametra θ sa vrednostima θ koje su više verovatne.

Populacije u svom prirodnom habitusu obično imaju ugnježdene substrukture sa hijerarhijskom prostornom raspodelom učestalosti genskih varijanti (Corander, Waldmann, and Sillanpää 2003). Te latentne substrukture su zapravo konglomerati subpopulacija čiji broj nije poznat. Procena broja subpopulacija se može ostvariti statističkim modelovanjem na osnovu opaženih podataka, odnosno genotipova, ali i drugim informacijama koje su postojale pre genotipiziranja, kao što je npr. informacija o geografskom poreklu. Ideja iza upotrebe modela zasnovanih na Bejzovom pristupu je jednostavna, na osnovu verovatnoća različitih događaja, kao što su individualni genotipovi, ali i geografsko poreklo, može se izraziti uslovna verovatnoća predloženog broja subpopulacija (Corander, Waldmann, and Sillanpää 2003).

U jednoj varijanti raspoređivanja pojedinačnih genotipova u subpopulacije, uslovna verovatnoća se smatra uniformnom, odnosno za svaku jedinku je podjednako verovatno da se pripíše nekoj od subpopulacija ako postoje dokazi o postojanju strukturiranosti. Druga varijanta uvodi i prethodno uverenje o geografskoj tj. prostornoj pripadnosti što utiče na verovatnoću raspoređivanja u subpopulacije, dok se genotipski podaci koriste za procenu verodostojnosti postojanja predloženih subpopulacija.

Redosled po kojem se inferira broj subpopulacija jeste pravljenje statističkog modela koji uključuje i pretpostavljeni broj subpopulacija, opciono može uključiti informaciju o geografskom poreklu, zatim se genotipski podaci koriste za određivanje verodostojnosti između modela sa različitim brojem pretpostavljenih subpopulacija i to MCMC metodom određivanja posteriorne raspodele (Corander, Sirén, and Arjas 2008).

Ovde moramo naglasiti da su svi Bejzovski pristupi u modelovanju genetičke strukturiranosti populacije u biti specijalni slučajevi maksimizacije verodostojnosti (Maximum Likelihood Estimation, MLE), odnosno maksimizacije posteriorne raspodele parametara (Maximum A Posteriori, MAP). Intencija iza upotrebe maksimizacije verodostojnosti jeste pronalazak vrednosti parametra koja je najverovatnija, a time i odabir modela koji se najbolje uklapa u opservirane podatke.

Pregledom literature, dva najčešće korišćena algoritma zasnovana na Bejzovom modelovanju jesu predloženi od Corandera i Pritcharda. Prema Corander-u i saradnicima, predloženi model ima za cilj otkrivanje prostorne strukturiranosti na osnovu informacija o geografskom poreklu i alelskim učestalostima, ali i mogućnost analize admiksičnih populacija (Corander, Waldmann, and Sillanpää 2003). Pod pretpostavkom gametske ravnoteže među lokusima (eng Linkage equilibrium), kao i Hardy-Vajnbergovog ekvilibrijuma u populaciji, Corander i saradnici postavljaju model po kojem je N_L broj nezavisnih lokusa, a na svakom lokusu j se može naći k različitih alela (genskih varijanti) u intervalu $[1, \dots, N_{A(j)}]$. Broj mogućih subpopulacija v_P (broj particija populacije) se nalazi u intervalu $[1, N_P]$, gde je N_P maksimalni broj subpopulacija. Prema ovome, verovatnoća p_{ijk} da se u nekoj populaciji i na nekom lokusu j pronađe alel A_{jk} zavisi od $[i=1, \dots, v_P; j=1, \dots, N_L; k=1, \dots, N_{A(j)}]$.

Kombinacija alelskih učestalosti u nekoj populaciji i je izražena kao Θ_i , dok n_{ijk} označava ukupan i apsolutan broj alela.

Celokupna populacija se particioniše na osnovu matrice sličnosti S , dimenzija $N_P \times N_P$:

$$S_{mr} = \begin{cases} 1, & \text{ako } \theta_m = \theta_r \\ 0, & \text{za drugačije} \end{cases} \text{ gde su } m \text{ i } r \text{ u intervalu } [1, N_P]$$

Primer:

Ako su u populaciji i genotipizirana četiri lokusa i pretpostavljeno je da postoje tri subpopulacije, matrice sličnosti će imati oblik:

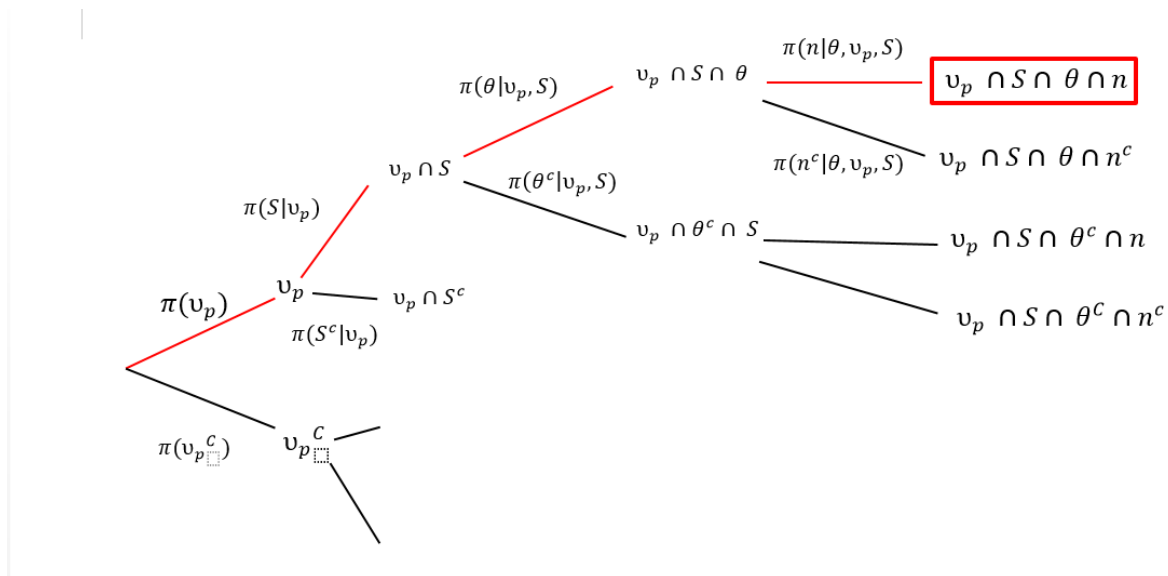
$$\begin{matrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{matrix}$$

i ukoliko se učestalosti podudare $\theta_m = \theta_r$, ćelija matrice će zauzeti vrednost 1, dok je u suprotnome 0.

S obzirom da je apsolutni broj alela n poznat nakon eksperimenta (opažanja podataka), uslovna verovatnoća ostvarivanja parametara θ , v_p , S , n na osnovu pravila množenja verovatnoća jeste:

$$\pi(\theta, v_p, S, n) = \pi(v_p)\pi(S|v_p)\pi(\theta|v_p, S) \pi(n|\theta, v_p, S)$$

, gde je $\pi(\theta|v_p, S)$ prethodna gustina θ , a $\pi(v_p)\pi(S|v_p)$ prethodna uslovna verovatnoća parametara particionisanja (strukturiranosti) populacije. Ovo se može i shematski prikazati kao jedan od uslovnih ishoda:



Slika 24 – Shematski prikaz jednog od logičkih ishoda.

1.1.1.1. Algoritmi za detekciju genetičke struktuiranosti

STRUCTURE

STRUCTURE je softverska implementacija istoimenog Bejzovog iterativnog algoritma koji grupiše genotipove na osnovu sličnosti u obrascu genetičkog variranja (Porrás-Hurtado et al. 2013a; Pritchard, Stephens, and Donnelly 2000). Ovaj algoritam pretpostavlja neutralnost genetičkih markera, odnosno gametsku ravnotežu, kao i Hardy Vajnbergov ekvilibrijum u populaciji i unutar subpopulacija. Glavna karakteristika ovog modela je da otkrivanje broja subpopulacija dolazi kao

posledica dodeljivanja genotipova u K subpopulacija i dozvoljava admiksiju, odnosno, da neki genotip može poticati iz više od jedne subpopulacije.

U osnov modela se nalaze tri komponente, tačnije tri višedimenzionalna vektora:

Z -populacija porekla jedinke

X -svi genotipovi uzorkovanih jedinki

P -alelske učestalosti u metapopulaciji

, koji se u skladu sa Bejzovom teoremom mogu staviti u međusoban odnos tako da se opiše posteriorna raspodela Z i P :

$$\Pr(Z, P|X) \propto \Pr(Z) \Pr(P) \Pr(X|Z, P)$$

, ukoliko su opservirani genotipovi X polazna osnova, a prethodne verovatnoće $\Pr(Z)$ i $\Pr(P)$ aproksimirane MCMC metodom.

U uzorku od N jedinki gde se genotipizira L broj neutralnih lokusa koji su gametskoj ravnoteži, elementi vektora Z , X i P jesu:

$(X_l^{(i,1)}, X_l^{(i,2)})$ = genotipovi jedinke i na lokusu l , gde je $i = 1, 2, 3, \dots, N$ i $l = 1, 2, \dots, L$;

$Z^{(i)}$ = populacija porekla jedinke i ;

P_{klj} = učestalost alela j na lokusu l u populaciji k , gde je $k = 1, 2, \dots, K$ i $j = 1, 2, \dots, J_l$,

tako da je J_l broj različitih alela opserviranih na lokusu l . Ovo ukazuje da je opservacija konkretnog genotipa na nekom lokusu zapravo uslovna verovatnoća izvlačenja alela iz populacije porekla i učestalosti tog alela u metapopulaciji:

$\Pr(X_l^{(i,a)} = j | Z, P) = p_{z^{(i)}l_j}$, gde je $p_{z^{(i)}l_j}$ učestalost alela j na lokusu l u populaciji porekla jedinke i . Raspodela alelskih učestalosti na svakom multialelskom lokusu unutar populacije je modelovana Dirihleovom raspodelom $p_{kl} = \text{Dir}(\lambda_1, \lambda_2, \dots, \lambda_{J_l})$. Uzorci različitih subpopulacija moraju biti slične veličine, jer model podrazumeva da je verovatnoća svakog genotipa da potiče iz neke od subpopulacija ista (Pritchard, Stephens, and Donnelly 2000):

$$\Pr(Z^i = k) = \frac{1}{K}$$

i ovo je bitna pretpostavka ovog modela, jer veći disbalans između subpopulacija može narušiti integritet celog modela.

Unutar STRUCTURE algoritma se nalaze dva podmodela za poreklo populacije: jedan se odnosi na admiksičnu populaciju, dok drugi pretpostavlja neadmiksičnu populaciju koja je sastavljena od diskretnih subpopulacija. Kada ne postoji prethodno znanje o populaciji u pogledu porekla, lingvističkih, etničkih, geografskih razlika, opravdana je upotreba neadmiksičnog modela, u suprotnom, treba pretpostaviti admiksiju i proceni broja subpopulacija priključiti i podatke o prethodnom znanju u vezi sa sastavom populacije (Porrás-Hurtado et al. 2013a).

Ocena parametara modela na osnovu maksimizacije verodostojnosti

Modeli zasnovani na Bejzovom pristupu se oslanjaju na evaluaciji verodostojnosti parametara od interesa na način da se verodostojnosti različitih vrednosti predloženih parametara međusobno

porede. Kako bi do poređenja došlo, potrebno je odrediti pojedinačne verodostojnosti i opisati njihovu raspodelu. Iz razloga upotrebljivosti, iz raspodele verodostojnosti se bira vrednost parametra koja ima najveću verodostojnost, ipak ne treba izgubiti iz vida činjenicu da ta vrednost ne mora biti prava, već je najverovatnija na osnovu opserviranih podataka. Pristup određivanja najverodostojnije vrednosti parametra se naziva metodom maksimalne verodostojnosti.

Polazeći od simetričnosti između verodostojnosti modela i parametara na osnovu opserviranih podataka sa verovatnoćom opserviranja podataka na osnovu modela i parametara:

$$L(\theta, M|D) = P(D|\theta, M)$$

, možemo rasčlanjivanjem $P(D|\theta, M)$ na zajedničku verovatnoću svakog opserviranog podatka n , verodostojnost predstaviti i kao produkt:

$$L(\theta, M|D) = p_{D_1}, p_{D_2}, \dots, p_{D_n}(D_1, D_2, \dots, D_n|\theta, M) = \prod_i p_{D_i}(D_i|\theta, M)$$

, gde je bitno istaći da se svi eksperimenti prilikom opserviranja podataka smatraju nezavisnim slučajnim događajima koji imaju istovetne raspodele verovatnoće (i.i.d., eng independent and identically distributed). Ipak, ovakav vid izračunavanja je sa informatičkog aspekta nepraktičan jer zahteva maksimizaciju produkta, što značajno opterećuje memoriju računara. Koristeći jednakost između produkta verovatnoća i sabiranja logaritma tih verovatnoća:

$$P(A \cap B) = P(A) * P(B) = e^{\log(P_{AB})} = e^{\log(P_A)+\log(P_B)}$$

```
Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32) [MSC v.1900 64 bit (AMD64)]
on win32
```

```
In[2]: import math
```

```
In[3]: p1=0.5; p2=0.5; p1p2=p1*p2; logp1p2=math.log(p1)+math.log(p2)
```

```
In[4]: p1p2==math.exp(logp1p2)
```

```
Out[4]: True
```

, ovaj problem se lako prevazilazi upotrebom sume izvoda funkcije verodostojnosti parametara, odnosno akumulacijom derivata svakog učitanoog podatka ponaosob:

$$\operatorname{argmax}_{\theta} = \prod_i p_{D_i}(D_i|\theta, M) = \operatorname{argmax}_{\theta} \sum_i \log(p(D_i|\theta, M))$$

, a kako je potrebno odrediti θ na globalnom maximum, odnosno θ čija je vrednosti izvoda jednaka 0, rešavanjem diferencijalne jednačine dobija se najverodostojniji parameter θ :

$$\operatorname{argmax}_{\theta} P(\theta) = \frac{dP(D|\theta)}{d\theta} = 0$$

Treba napomenuti da je maksimizacija logaritma verodostojnosti parametara komplementarna minimizaciji negativnog logaritma verodostojnosti i da se najčešće koristi u oceni parametara statističkih modela.

S obzirom da u Bejzovom pristupu raspodelu opserviranih podataka nazivamo posteriornom raspodelom, sama procena maksimalne verodostojnosti je nešto drugačija i naziva se aposteriorni maksimum (eng *Maximum A Posteriori*, MAP).

$$\operatorname{argmax}_{\theta} P(\theta|D) = \operatorname{argmax}_{\theta} P(\theta|D)P(\theta)$$

Da bi to bilo moguće, raspodela verovatnoće se konjuguje sa raspodelom verodostojnosti, najčešće prebacivanjem obe raspodele u *Beta* ili *Gama* raspodelu. Još jednom treba napomenuti da je *Gama* raspodela sa dva hiperparametra isto što i *Beta* funkcija. Izvodom diferencijalne jednačine za vrednost nagiba funkcije 0, dobija se maksimalna verodostojnost posteriorne raspodele parametra θ :

$$\frac{dP(\theta|D)P(\theta)}{d\theta} = 0$$

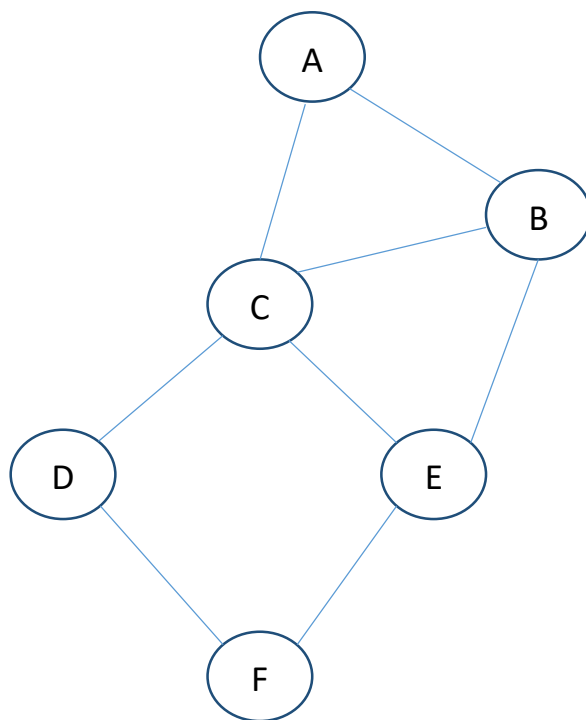
Odnos između metode maksimizacije verodostojnosti MLE i aposteriornog maksimuma MAP je takav da je MLE poseban slučaj MAP u kojoj je prethodna verovatnoća uniformna. Određivanjem MLE za parametar K (broj populacija), STRUCTURE procenjuje najoptimalniji broj populacija koje bi odgovarale opaženim genetičkim podacima.

Girvan-Njumanov algoritam

Poslednji u nizu algoritama koji se mogu primeniti za inferenciju genetičke struktiranosti populacije jeste heuristička implementacija teorije mreža kroz Girvan-Njumanov algoritam (GN) (Greenbaum, Templeton, and Bar-David 2016). Prema teoriji koju su definisali Girvan i Njuman, "mreža je skup diskretnih entiteta, čvorišta, koja su međusobno povezana linijama/granama, koje mogu imati kvantitativne vrednosti, odnosno težinu" (Newman and Girvan 2004). Same mreže se mogu definisati matricom povezanosti ili grafički, usmerenim grafikom, što se primenjuje i na populaciju koja se na osnovu genetičkih podataka pretvara u matricu genetičke udaljenosti, a potom i u grafik.

Prema Girvanu i Njumanu, grupa čvorišta, unutar neke mreže, koja je sa topološkog aspekta međusobno gušće i jače povezana u odnosu na ostatak mreže, jeste zajednica (Newman and Girvan 2004). Analog zajednice u populaciji može biti i subpopulacija ili jednostavno grupa jedinki između kojih se genetički materijal češće razmenjivao (Greenbaum, Templeton, and Bar-David 2016).

Za razliku od grupišućih algoritama, Girvan-Newmanov algoritam se zasniva na eliminaciji linija/grana (eng 'edges') između najudaljenijih zajednica, odnosno hijerarhijskoj dekompoziciji. Drugim rečima, on funkcioniše na suprotan način u odnosu na ostale grupišuće i hijerarhijske algoritme: umesto dna hijerarhije kao polazne osnove, algoritam kreće sa vrha i uklanja najopterećenije grane (eng *highest betweenness edge*) unutar grupe podataka dok ne ostanu sami klasteri koji se nazivaju zajednice. Najopterećenije grane su one kroz koje prolazi najveći broj najkraćih putanja između čvorova. Karakteristično je za GN algoritam da koristi dužine l umesto ponderisane vrednosti grana w , iako se dužine grana mogu izraziti kao recipročne vrednosti pondera $l=1/w$ (Aggarwal 2015).



Slika 25 – Shematski prikaz čvorišta (slovo) u mreži i grana/linija koje ih spajaju. Najkraći put AF je zbir najkraćih puteva AD i AC.

Modularnost je izuzeto bitno svojstvo mreže prilikom procene broja mogućih grupa i definisana je kao:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i C_j)$$

, gde su:

Q-mera modularnosti,

m-broj grana,

A-graf matrice povezanosti,

$k_i k_j$ -stepen parova čvorišta,

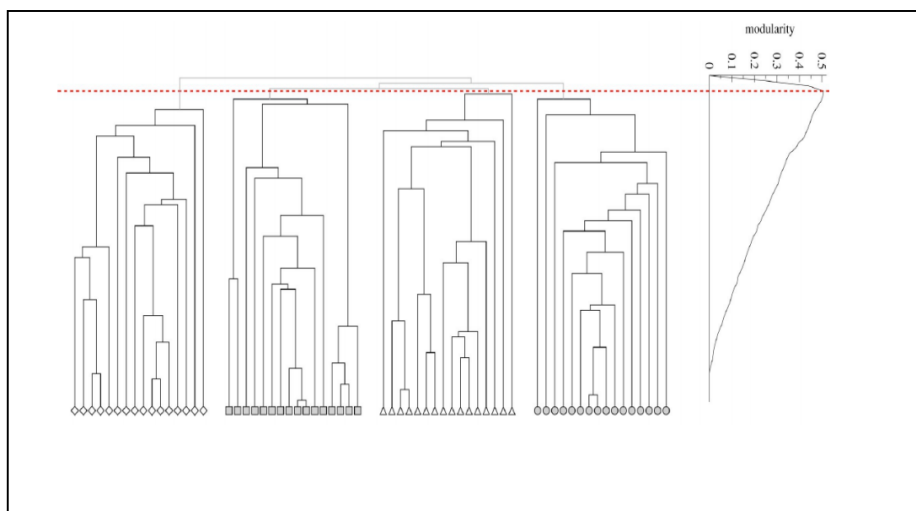
δ -pokazatelj pripadnosti grupi.

S obzirom na to da se GN algoritam bazira pronalazanju zajednica unutar mreže koja je predstavljena grafom, prvi korak u primeni ovog algoritma jeste konstrukcija grafa na osnovu matrice

genetičke udaljenosti čvorišta u mreži (genotipova). Genetička udaljenost, često nazvana i sličnost, se matematički određuje na oko 20 različitih načina (Dyer 2017), a jedan od predloženih za kvantifikaciju genetičke udaljenosti individue i i j na lokusu l gde su moguće, jeste:

$$S_{ij,l} = \frac{1}{4} ((fa - l_{ac})(fb + l_{ad}) + (fc - l_{bc})(fd + l_{bd}))$$

, gde su f učestalosti, dok je l_{ac} jednak 0 ukoliko su aleli a i c identični, isto važi i za l_{ad} , l_{bc} i l_{bd} . Ukoliko su različiti onda je njihov skor 1, a genetička udaljenost genotipova određuje kao suma distanci između svih poređenih lokusa. Imajući u vidu da GN algoritam raščlanjuje mrežu do diskretnih činilaca, potrebno je odrediti nivo u rasčlanjivanju na kojem je najverovatnije uočiti zajednice. Ukoliko izuzmemo vizuelni i krajnje subjektivni način određivanja nivoa razlaganja, određivanje najveće modularnosti preostaje kao najobjektivniji pristup.

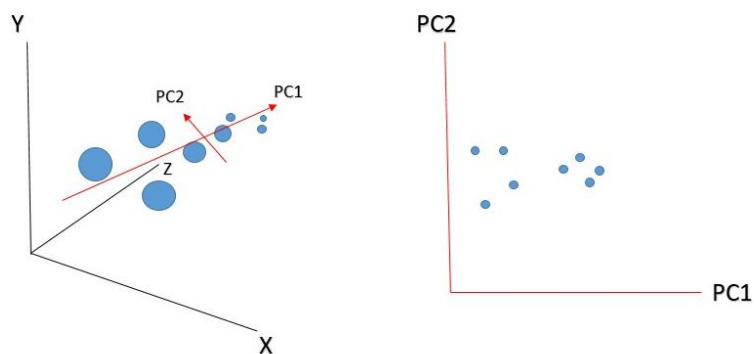


Slika 26 – Prikaz hijerarhije unutar klastera i promene modularnosti.
Izvor: Newman M E, Girvan M., 2004.

Krajnji rezultat GN algoritma je ‘ogoljeni’ grafik gde su zajednice uočljive u vidu grupacija sa međusobno povezanim čvorištima. Otkrivanje zajednice unutar mreže je analogno otkrivanju subpopulacije unutar opšte populacije (Greenbaum, Templeton, and Bar-David 2016)

Analiza glavnih komponenti (Principal Component Analysis)

PCA je linearna metoda redukcije dimenzionalnosti koja se prvenstveno zasniva na reprezentaciji nižedimenzionalnih podataka tako da se najsljedniji podaci blisko poziciraju. Sama teorija o smanjenju većeg broja međusobno zavisnih varijabli u glavne komponente postoji bezimalo 90 godina (Hotelling 1933), a po nekim tumačenjima i duže (Jolliffe and Cadima 2016; Pearson 1901). Analiza glavnih komponenti omogućava da se ukupna varijabilnost nekog uzorka bolje prikaže putem manjeg broja izvedenih varijabli (glavnih komponenti), što implicira da glavne komponente mogu sadržati više varijabilnosti od svake ulazne varijable ponaosob (slika 27).



Slika 27 - Ilustracija redukcije dimenzionalnosti analizom glavnih komponenti varijanse.

Glavne komponente su poznatije kao ajgenvektori matrice kovarijansi podataka. Svaka ćelija matrice kovarijansi se izračunava po jednačini:

$$Cov(X, Y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

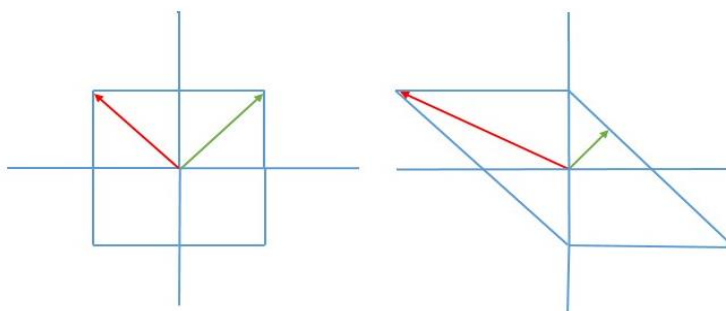
, dok matrica ima dimenzije $v \times v$, gde je v broj varijabli. Karakteristika matrice kovarijanse je da je simetrična u odnosu na dijagonalu i da se upravo u dijagonali nalaze vrednosti varijanse za svaku od varijabli:

$$\begin{matrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{matrix}$$

Ajgenvektor je posebna vrsta vektora koja nakon transformacije matrice ne menja svoj pravac, već može promeniti samo dužinu ili rotirati u istom pravcu oko tačke simetrije (zeleni vektor na slici 28).

Koeficijent uvećanja ili umanjnja ajgenvektora nakon transformacije se zove ajgenvrednost.

Matematički izraženo, postoji jednakost između produkta transformacione matrice i vektora i produkta ajgenvrednosti i vektora:



Slika 28 - Transformacija matrice. Crvenom bojom je označen običan vektor, dok je zelenom bojom označen ajgenvektor.

$$A\vec{v} = \lambda\vec{v}$$

, što znači da je i zbir varijansi pojedinačnih varijabli u matrici kovarijanse jednak zbiru ajgenvrednosti:

$$\sum v_{ii} = \sum \lambda_i$$

, a da bi se izračunale λ_i , varijanse svih v_{ii} u matrici kovarijanse se kodiraju sa 1 i takva matrica naziva matrica identičnosti I:

$$\begin{matrix} 1 & v_{21} & v_{31} \\ v_{12} & 1 & v_{32} \\ v_{13} & v_{23} & 1 \end{matrix}$$

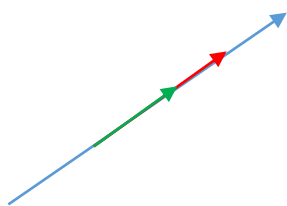
, dalje se rešavanjem jednačine:

$$A\vec{v} = (\lambda I)\vec{v}$$

, odnosno:

$$A\vec{v} - (\lambda I)\vec{v} = \vec{0}$$

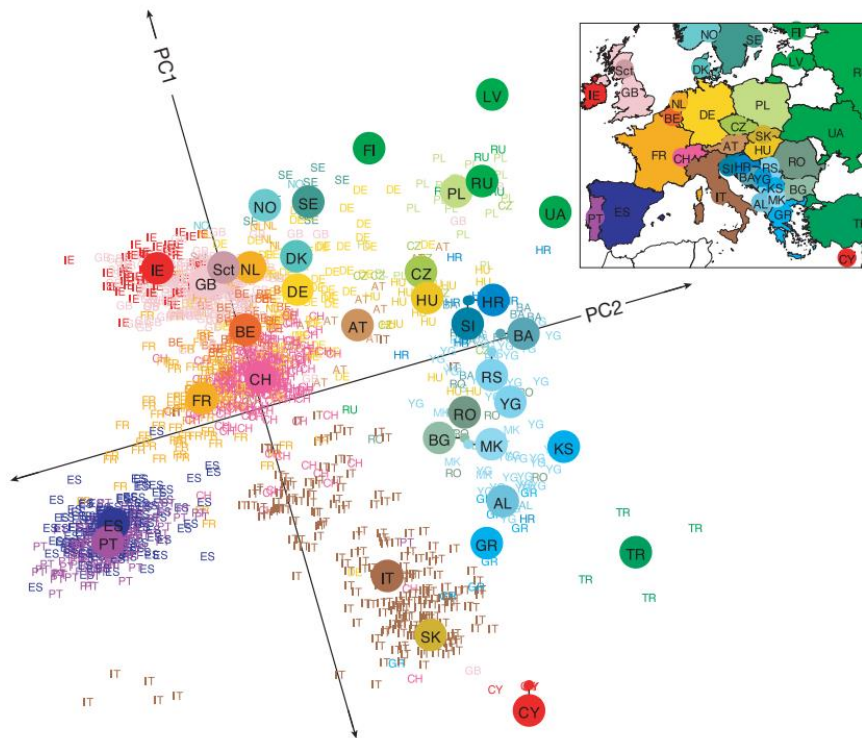
, što zahteva da se produkt matrice sa vrednošću različitom od 0 transformiše u 0. To je jedino moguće redukcijom dimenzionalnosti.



Slika 29 - Grafički prikaz redukcije dimenzionalnosti

Karakteristično je za analizu glavnih komponenti da su ulazne varijable međusobno asocirane i da ajgenvektori koji se izračunavaju u matricama kovarijanse nisu međusobno asocirani. Ovim se postiže da se veliki deo od ukupne varijabilnosti može opisati sa dva ili više ajgenvektora. Međutim, ukoliko su ulazne varijable nezavisne, raspodela udela ajgenvektora u ukupnoj varijansi ima prilično uniformnu raspodelu, što čini PCA neinformativnom.

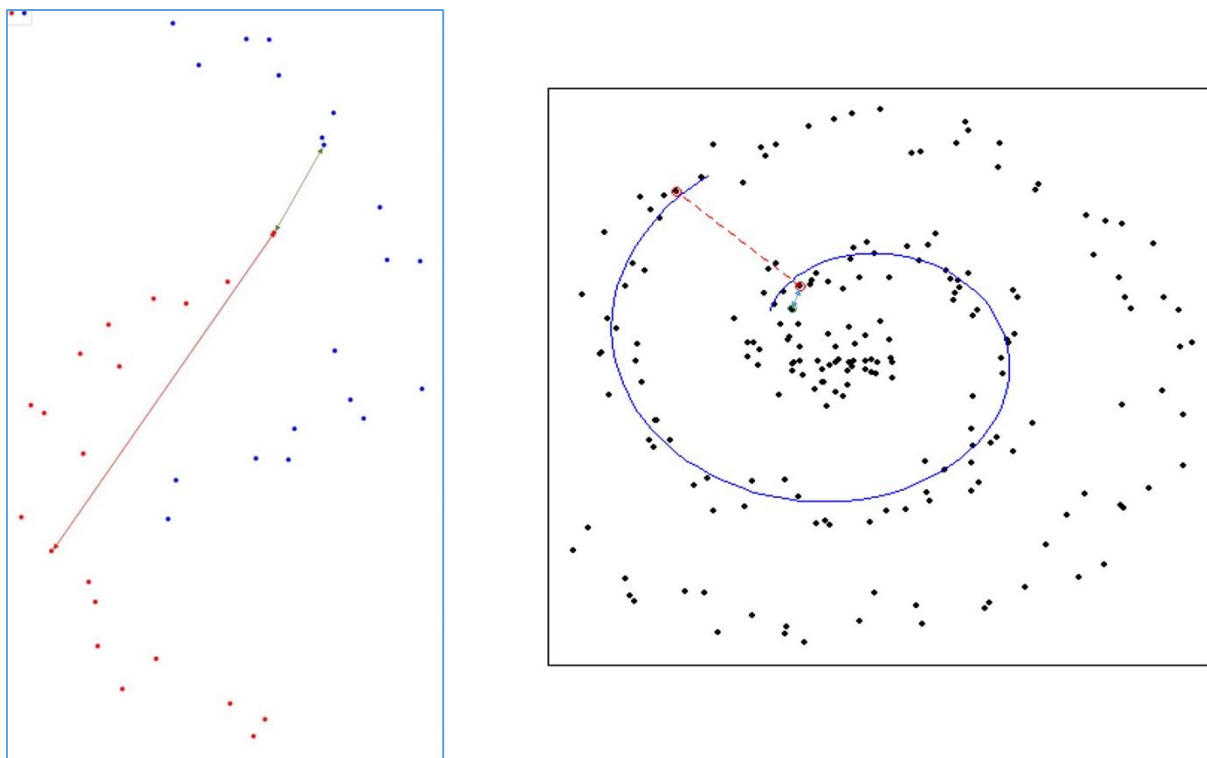
U populacionoj genetici se pokazalo da je PCA moguće primeniti na multilokusnim podacima, npr. SNV, kako bi se ilustrovala genetička struktura populacije (Novembre et al. 2008). Ovaj vid provere genetičke strukture je sastavni deo GWAS studija jer omogućava eradikaciju podataka koji zbog struktuiranosti uzorka mogu iznedriti lažno pozitivne asocijacije (Sul, Martin, and Eskin 2018)



Slika 30 - Prikaz analize glavnih komponenti na osnovu SNV iz uzorka Evropske populacije. Izvor: Novembre et al. 2008.

Stohastičko umetanje susednih tačaka na osnovu t raspodele (eng. t-distributed stochastic neighbor embedding-tSNE)

Relativno nedavno je u analizi genetičke strukture populacije primenjen i tSNE algoritam. Dve su glavne prednosti tSNE u odnosu na druge algoritme za vizuelizaciju višedimenzionalnih podataka su robusnost u odnosu na ekstremne vrednosti i mogućnost detaljnijeg prikaza lokalnih odnosa u subpopulacijama (W. Li et al. 2017). Za razliku od PCA metoda koji se fokusira na maksimizaciju varijanse, održanju razdaljine između tačaka koje su udaljene u višedimenzionalnom prostoru, tSNE je algoritam za vizuelizaciju koji ima mogućnost očuvanja lokalnih odnosa višedimenzionalnih podataka nakon redukcije dimenzionalnosti, a da pri tome u znatnoj meri sačuva i globalne odnose tako da se ne narušava opšte grupisanje (klasteri podataka) (Visualizing .Data using t-SNE, 2008). Glavna prednost tSNE u odnosu na PCA je što ne uzima u obzir samo Euklidsko rastojanje, već celokupnu sličnost tačaka u višedimenzionalnom prostoru. To znači da će dve tačke x_i i x_j koje su međusobno udaljenije u Euklidskom prostoru (plava linija na slici 31 a) biti klasifikovane u isti klaster na osnovu sličnosti sa susednim tačkama koje vode od x_i do x_j , ne samo na osnovu apsolutne udaljenosti (crvena linija na slikama 31 a i b).



Slika 31 a i b - Skup nelinearnih podataka u dvodimenzionalnom prostoru, relativne i apsolutne udaljenosti.

Prvi korak u tSNE algoritmu je izračunavanje sličnosti, a zapravo se radi o uslovnoj verovatnoći na osnovu Euklidske razdaljine između podataka u višedimenzionalnom prostoru. Ovde je uslovna verovatnoća $p_{j|i}$ da je x_i susedna tačka x_j na osnovu normalne raspodele centrirane nad x_i , drugim rečima, verovatnoća da će dve tačke biti određene kao susedne opada sa gustinom normalne raspodele od pozicije tačaka:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_k - x_i\|^2 / 2\sigma_i^2)}$$

, nakon čega se uvodi simetričnost između tačaka sa pretpostavkom da gustina raspodele verovatnoće između para x_i, x_j ne može biti različita u zavisnosti iz kojeg pravca se odnos dve tačke posmatra.

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}$$

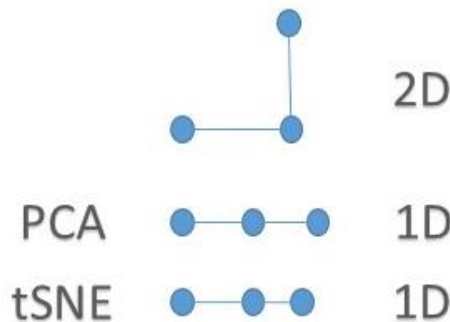
Isti par tačaka (x_i, x_j) ima korenspodirajući nižedimenzionalni par (y_i, y_j) za koji je uslovna verovatnoća $q_{j|i}$ izražena na osnovu t raspodele sa jednim stepenom slobode:

$$q_{j|i} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}}$$

, a ako tačke y_i, y_j verodostojno predstavljaju sličnost između x_i, x_j , onda će uslovne verovatnoće $p_{j|i}$ i $q_{j|i}$ biti jednake. Primena t raspodele u nižedimenzionalnom mapiranju omogućava da se ekstremne vrednosti bolje puferišu i manje utiču na globalno mapiranje. Da bi algoritam procenio podudarnost ove dve verovatnoće, primenjuje se Kulbak-Lajberov divergenca tako da ne dođe do gubitka informacija prilikom konverzije $p_{j|i}$ u $q_{j|i}$. Slikovito rečeno, ukoliko su dve tačke x_i, x_j u višedimenzionalnom prostoru udaljene tako da je gustina verovatnoće $p_{j|i}$ mala između njih, svaka predložena reprezentacija u nižedimenzionalnom prostoru čija gustina verovatnoće $q_{j|i}$ nije podudarna sa $p_{j|i}$, neće imati zadovoljavajuću vrednost na osnovu Kulbak-Lajberove divergencije po formuli:

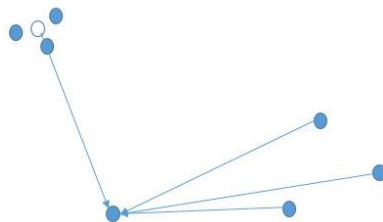
$$KL(P||Q) = \sum_i \sum_{j \neq i} P_{ij} \log \frac{p_{ij}}{q_{ij}}$$

, ali samo za slučajeva kada je mala vrednost $p_{j|i}$ modelovana velikim vrednostima $q_{j|i}$. To za posledicu ima očuvanje lokalnih sličnosti nad globalnim.



Slika 32 - Pojednostavljeni prikaz redukcije dimenzionalnosti iz 2D u 1D.

Prilikom gradijentne redukcije dimenzionalnosti, tSNE za svaku tačku uzima u obzir repulzivni uticaj ostalih tačaka, s tim što tačke koje su blisko pozicionirane imaju prosečan uticaj repulzije (Barnes-Hut aproksimacija), tako da je konačno pozicioniranje u nižoj dimenziji rezultat uticaja svih ostalih tačaka i klastera iz više dimenzije (slika 33). Matematička formulacija sume repulzivnih uticaja je definisana kao:



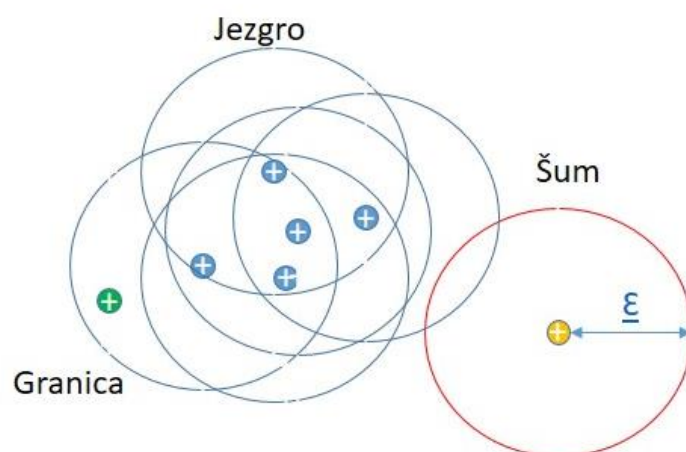
Slika 33 - Odnos drugih tačaka u višedimenzionalnom prostoru. Beli kružić označava zbir vektora repulzije unutar klastera.

$$\frac{\partial C}{\partial y_i} = 4 \sum_{j \neq i} (p_{ij} - q_{ij}) \left(1 + \|y_i - y_j\|^2 \right)^{-1} (y_i - y_j)$$

Prostorno grupisanje zasnovano na gustini raspodele aplikacija sa šumom (*Density-Based Spatial Clustering of Applications with Noise-DBSCAN*)

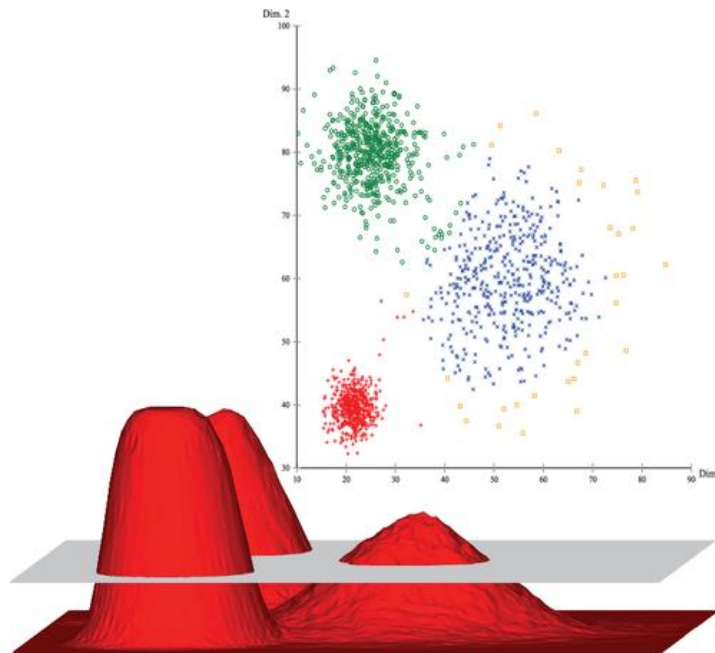
U osnovi ovog algoritma leži pretpostavka o postojanju funkcija gustine verovatnoće (eng *Probability Density Function*-PDF) na osnovu koje su podaci raspoređeni u prostoru, odnosno na grafiku raspršenosti (eng *scatter plot*). Drugim rečima, za svaku tačku na grafiku, postoji određena verovatnoća pronalaska objekta/podatka. Takođe, ovaj algoritam zahteva implementaciju funkcije udaljenosti (distance) na osnovu koje se podaci porede i grupišu. Klase podataka koje nastaju primenom ovog algoritma su centralna, granična i šum. Dobra karakteristika DBSCAN je što on može pronaći grupacije proizvoljnih oblika. Algoritam počinje nasumično odabrane q tačke da pretražuje susedne tačke p u razdaljini ϵ :

$$N_\epsilon = \{q \in D \mid \text{razdaljina}(p, q) < \epsilon\}$$



Slika 34 - Grafički prikaz funkcionisanja DBSCAN algoritma. Plavi krugovi su deo jezgra, zelena tačka je granična, crveni grup je površina koju zahvata tačka van klastera (šum).

, ukoliko je q tačka u sferi \mathcal{E} pronađe unapred definisani minimalni broj tačaka p (MinP), formira jezgro i nastavlja da traži susedne tačke za svaku p tačku iz jezgra. Nakon pripajanja svih p tačaka koje su imale kontakt sa jezgrom i u svom \mathcal{E} okruženju minimalni broj tačaka MinP, nasumice se odabira tačka koja nije pripisana jezgru i proces iznova započinje. One tačke koje u svom susedstvu nisu imale MinP tačaka, smatraju se šumom. Sve tačke koje imaju jezgro, ali i sve granične tačke, su deo klastera. Unutar svakog klastera mogu postojati prostorni kompartimenti sa većom kao i oni sa manjom gustinom tačaka. Zahvaljujući različitoj gustini, moguće je međusobno hijerarhijsko ustrojavanje klastera (slika 35)(Campello et al. 2020; Ester et al. 1996).



Slika 35 – Trodimenzionalni prikaz gustine klastera.
Izvor: Campello et al. 2020.

2. Ciljevi istraživanja

Istraživanje raspodele učestalosti funkcionalnih varijanti CYP450 supergenske familije u populaciji Srbije je imalo sledeće ciljeve:

1. Odrediti učestalosti polimorfizama u genima CYP1A1 (rs4646903, rs1048943), CYP2C9 (rs1057910, rs1799853), CYP2C19 (rs12248560, rs4244285), CYP2D6 (rs3892097, rs1065852, rs28371706, rs28371725), CYP3A4 (rs2740574) i njihove distribucije u populaciji Srbije.
2. Proceniti genetičku strukturiranost populacije stanovništva Republike Srbije primenom algoritama zasnovanih na genetičkoj udaljenosti i algoritama zasnovanih na statističkom modelu.
3. Utvrditi eventualno postojanje razlika u distribuciji učestalosti alela i gentipova za analizirane polimorfizme u odnosu na geografske-prostorne i/ili etničke faktore unutar analizirane populacije, kao i u odnosu na druge susedne i nesusedne populacije.
4. Definisati farmakološki informativne panele genskih testova za analizirane CYP enzime.

3. Materijal i metode

3.1. Mesto i period istraživanja

Celokupno istraživanje je sprovedeno u DNK laboratoriji Instituta za sudsku medicinu, Medicinskog fakulteta u Beogradu u periodu od 2016. do 2019. godine.

3.2. Ispitanici

Ispitanici su osobe koje su deponovale biološki material tokom testiranja srodničkih odnosa u periodu od 2011. do 2019. godine u DNK laboratoriji Instituta za sudsku medicinu, Medicinskog fakulteta u Beogradu. Kriterijumi na osnovu kojih su ispitanici uključeni u istraživanje su sledeći:

- Postojanje saglasnosti za upotrebu biološkog materijala za genetička istraživanja.
- Poznato mesto prebivališta na teritoriji Republike Srbije, osim za ispitanike Romske nacionalne pripadnosti.
- Da nisu u srodstvu sa drugim ispitanicima uključenim u istraživanje.
- Postojanje kompletnog STR profila na osnovu rezultata testiranja srodničkih odnosa.

3.3. Uzorci biološkog materijala

Iz biblioteke bioloških izolata DNK laboratorije Instituta za sudsku medicinu u kojoj se nalazi između 8000 i 9000 uzoraka, izabrani su svi izolati osoba koje su ispunile kriterijum za učešće u istraživanju.

3.4. Klasifikacija i veličina uzorka

Za potrebe istraživanja su formirane dve grupe uzoraka sa ciljem da reprezentuju opštu populaciju i subpopulaciju Roma u Srbiji.

U grupi koja reprezentuju opštu populaciju Srbije su izabrani uzorci osoba sa poznatim mestom prebivališta. Uzorci su zatim stratifikovani u pet kategorija prema Nomenklaturi statističkih teritorijalnih jedinica (Službeni glasnik RS, broj 104/09), bez regiona Kosova i Metohije. Za potrebe istraživanja koje se odnose na primenu linearnih metoda za redukciju dimenzionalnosti podataka, region Šumadije i Zapadne Srbije je razdvojen u dva zasebna regiona, Centralne i Zapadne Srbije, za koje je pretpostavljeno da bi iz društveno-istorijskih razloga mogli imati veću genetičku varijabilnost (Halpern 1999; Kanitz 1868; Sundhaussen 2007; Vukovic et al. 2011) (slika 36). Svi uzorci su klasifikovani u jedan od pet regiona:

1. Severna Srbija
2. Okrug Beograd
3. Zapadna Srbija
4. Centralna Srbija
5. Južna i Istočna Srbija



Slika 36 - Mapa Republike Srbije prema Nomenklaturi statističkih teritorijalnih jedinica i sa izdvojenim regionom Centralne Srbije.

, nakon čega je odabrano ukupno 550 DNK izolata prema broju stanovnika koji nastanjuju region. Pri odabiru uzoraka koji će prezentovati stanovnike regiona, upotrebljen je generator slučajnih brojeva (Haahr 2012) kako bi se uvela randomizacija unutar stratuma i izbegao uticaj pridružene varijable.

Broj ispitanika po regionima je iznosio:

- Severna Srbija – 132 uzorka (24%).
- Region Beograda – 124 uzorka (22.5%).
- Zapadna Srbija – 75 uzoraka (13.6%).
- Centralna Srbija – 92 uzoraka (16.7%)
- Južna i Istočna Srbija – 127 uzoraka (23%)

Za grupu uzoraka koji predstavljaju romsku subpopulaciju je izabrano 59 DNK izolata, od kojih se 32 izolata od osoba sa poznatim mestom prebivališta na teritoriji Republike Srbije zajednički sa grupom koja predstavlja opštu populaciju Srbije.

3.5. Instrumenti merenja

Nakon formiranja grupa od DNK izolata ispitanika, proverena je koncentracija DNK i urađena genotipizacija na 11 polimorfnih lokusa u 5 *CYP* gena (*CYP1A1*, *CYP2C9*, *CYP2C19*, *CYP2D6* i *CYP3A4*).

3.5.1. Određivanje koncentracije DNK pomoću lančane reakcije polimeraze u realnom vremenu (Real Time PCR)

U svim izolatima DNK je određena koncentracija metodom apsolutne kvantifikacije pomoću TERT Taqman Assay seta reagenasa (Applied Biosystems, Fostercity, SAD) i Universal PCR Master Mix na Real Time PCR aparatu Viiia7.

Za pripremu PCR reakcije je korišćen rastvor koji se dobija dodavanjem:

1. 5 μ L DNK izolata
2. 10 μ L Universal PCR Master Mix
3. 4 μ L dejonizovane vode
4. 1 μ L TERT Taqman Assay

PCR reakcija je izvedena pod sledećim temperaturnim uslovima:

- Početna denaturacija u trajanju od 10 minuta na temperaturi od 95°C.
- 50 ciklusa amplifikacije u dva koraka:
 - denaturacija u trajanju od 15 sekundi na temperaturi od 95°C.
 - hibridizacija i elongacija u trajanju od 1 minuta na temperaturi od 60°C.

Rezultati kvantifikacije su očitani QuantStudio™ softverom (Applied Biosystems, Foster City, SAD).

Nakon određivanja početne koncentracije DNK u uzorcima, izvršena je normalizacija razblaživanjem izolata dejonizovanom vodom tako da se postigne finalna koncentracija u opsegu od 0.5 do 5 ng/ μ L pre genotipizacije *CYP* lokusa.

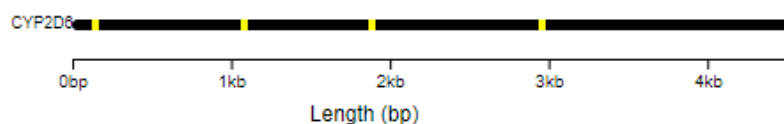
3.5.2. Genotipizacija *CYP* SNV

Za određivanje varijanti na 11 lokusa u *CYP* genima *CYP1A1*, *CYP2C9*, *CYP2C19*, *CYP2D6* i *CYP3A4*, je upotrebljen TaqMan Drug Metabolism Assay set reagenasa i Universal PCR Master Mix (Applied Biosystems, Foster City, SAD).

Genotipizacija je izvedena na Real-time PCR aparatu Vii7, (Applied Biosystems, Foster City, SAD), za sledeće genetičke varijante (varjanta, pozicija, identifikacioni broj TaqMan Drug Metabolism Assay):

CYP2D6:

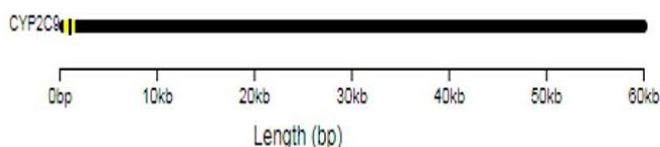
- rs3892097, (c.1847G>A), C__27102431_D0
- rs1065852, (c.100C>T), C__11484460_40
- rs28371725, (g.2988G>A), C__34816116_20
- rs28371706, (g.1023C>T), C__2222771_A0



Slika 37 - pozicije rs1065852, rs28371706, rs3892097 i rs28371725 u *CYP2D6* transkriptu. Vizuelizacija pomoću 'chromoMap' R softverskog paketa.

CYP2C9:

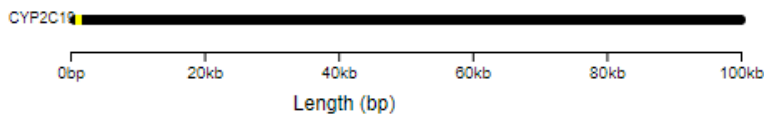
- rs1799853, (c.430C>T), C__25625805_10
- rs1057910, (c.1075A>C), C__27104892_10



Slika 38 - pozicije rs1799853 i rs1057910 u *CYP2C9* genu. Vizuelizacija pomoću 'chromoMap' R softverskog paketa.

CYP2C19:

- rs4244285, (c.681G>A), C__25986767_70
- rs12248560, (g.-806C>T), C____469857_10



Slika 39 – pozicija rs4244285 u *CYP2C19* genu. Vizuelizacija pomoću 'chromoMap' R softverskog paketa.

CYP3A4:

- rs2740574, (c.-392A>G), C__1837671_50

Za pripremu PCR reakcije je korišćen rastvor koji se dobija dodavanjem:

1. 5 µL DNK izolata
2. 10 µL Universal PCR Master Mix
3. 4 µL dejonizovane vode
4. 1 µL TaqMan Drug Metabolism Assay

Temperaturni uslovi RT-PCR genotipizacije Taqman Drug Metabolism Assay:

1. Početna denaturacija u trajanju od 10 min na 95°C
2. 50 ciklusa u dva koraka:
 - Denaturacija 15s na 92°C
 - Hibridizacija i elongacija 90s na 60°C

Rezultati genotipizacije su očitani QuantStudio™ softverom (Applied Biosystems, Foster City, SAD).

3.5.3. Analiza polimorfizma dužine restrikcionih fragmenata (RFLP)

Pored genotipizacije TaqMan assay setom reagenasa, svi uzorci su genotipizirani i metodom analize polimorfizma dužine restrikcionih fragmenata (RFLP) prethodno umnoženih Type-it Mutation Detect PCR Kit (Qiagen, Hilden, Nemačka) na PCR aparatu ABI2720 (Applied Biosystems, Foster City, SAD) i to za sledeće genetičke varijante, koristeći sledeće prajmere:

CYP11A1:

- lokus: rs4646903, pozicija: (g.1759A>G)

Prajmeri:

F-ACTCACCTGAACCCCATTC

R-GGCCCAACTACTCAGAGGCT

- lokus: rs1048943, pozicija: (g.3103T>C)

Prajmeri:

F-CTGTCTCCCTCTGGTTACAGGAAGC

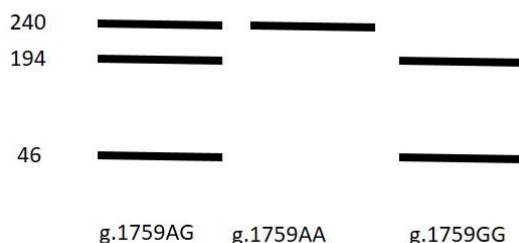
R-TTCCACCCGTTGCAGCAGGATAGCC

Temperaturni profil PCR reakcije za RFLP:

1. Početna denaturacija u trajanju od 5 min na 95°C
2. 30 ciklusa u tri koraka:
 - Denaturacija 30s na 95°C
 - Hibridizacija pramera 90s na 55°C
 - Ekstenzija 30s na 72°C
3. Finalna ekstenzija 30 min na 60°C

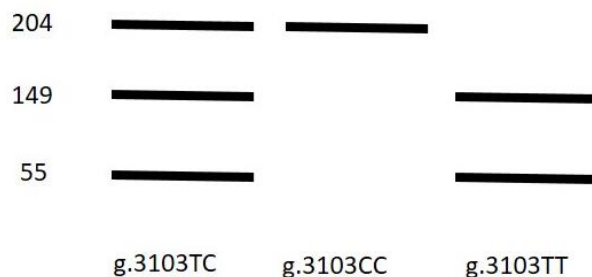
Za restrikciju DNK amplikona su upotrebljeni sledeći enzimi:

- Za varijabilno mesto na lokusu rs4646903 (g.1759A>G) je upotrebljen restrikcioni enzim MspI (New England Biolabs) kojim se dvolančani DNK amplikon dužine od 240 bp (g.1759T) fragmentiše na oligonukleotide dužina 194bp, 46bp kada je na poziciji g.1759 prisutna varjanta G.



Slika 40 – Genotipizacija rs4646903 RFLP metodom.

- Za varijabilno mesto na lokusu rs1048943 (g.3103T>C) je upotrebljen restrikcioni enzim BsrDI (New England Biolabs) kojim se dvolančani DNK amplikon dužine od 204 bp (g.3103C) fragmentiše na oligonukleotide dužina 149 , 55bp kada je na poziciji g.3103 prisutna varijanta T.



Slika 41 – Genotipizacija rs1048943 RFLP metodom.

Fragmenti nastali nakon restrikcije enzimima MspI i BsrDI su razdvojeni horizontalnom elektroforezom u 1xTBE puferu na 1.8% agaroznom gelu i vizuelizovani bojenjem etidijum bromidom.

3.5.4. Analiza mikrosatelitskih lokusa

Zahvaljujući tome što su svi testirani uzorci prethodno bili predmet analize srodničkih odnosa, postojala je informacija o genotipskom profilu kratkih uzastopnih ponovaka (eng Short Tandem Repeats, STR). Ovi profili su nastali genotipizacijom pomoću AmpFLSTR™ Identifiler™ Plus (Applied Biosystems, Foster City, SAD) seta reagenasa za forenzičko-genetičku identifikaciju ljudi. Ovaj set reagenasa se koristi za formiranje STR genotipa u vidu elektroferograma koji uključuje 15 mikrosatelitskih lokusa i to: D21S11, CSF1PO, vWA, D8S1179, TH01, D18S51, D5S818, D16S539, D3S1358, D2S1338, TPOX, FGA, D7S820, D13S317, D19S433.

Rezultati kapilarne elektroforeze (elektroferogrami) su očitani pomoću DataCollection V3 softvera (Applied Biosystems, Foster City, SAD), dok je za određivanje alela upotrebljen GeneMapper ID V3 softver (Applied Biosystems, Foster City, SAD).

3.6. Statističke i bioinformatičke analize

3.6.1. Statistička testiranja i analiza molekulske varijanse

Za proveru odstupanja od pretpostavljenih učestalosti na osnovu Hardy-Vajnbergove ravnoteže, kao i za poređenja učestalosti kategorijalnih podataka, upotrebljen je Fišerov test tačne verovatnoće (Fisher 1922, 1992), dok je prag značajnosti statističkog testa za višestruka poređenja prilagođen na osnovu Dan-Šidakove korekcije (Šidák 1967).

Parametri F statistike, su izračunati na nivou pojedninačnih genetičkih vrjanti gena *CYP450* prema matematičkim izrazima:

$$F_{IS} = \frac{H_S - H_I}{H_S}$$
$$F_{IT} = \frac{H_T - H_S}{H_T}$$
$$F_{St} = \frac{H_T - H_I}{H_T}$$

, gde su H_S i H_T očekivane heterozigotnosti u subpopulaciji i opštoj populaciji, dok je H_I opažena heterozigotnost među individualnim genotipovima (Sewall Wright 1965).

Haplotipovi gena *CYP450* (*CYP1A1*, *CYP2C9*, *CYP2C19*, *CYP2D6* i *CYP3A4*) su generisani upotrebnom softvera Haploview (Barrett et al. 2005), nakon čega su podvrgnuti analizi molekulske varijanse pomoću Arlequin softvera (Laurent Excoffier and Lischer 2010).

3.6.2. Primena algoritama za detekciju genetičke struktuiranosti

Za otkrivanje klastera unutar dobijenih genotipskih podataka iz uzorka opšte populacije sa regionalnom stratifikacijom i uzorka sa etničkim stratumom, upotrebljeni su algoritmi koji se zasnivaju na Bejzovom pristupu statističkog zaključivanja, redukciji dimenzionalnosti, hijerarhijskom klasterovanju i primeni teorije mreža.

Softveri bazirani na algoritmima Bejzovog statističkog zaključivanja:

- STRUCTURE sa pretpostavljenim modelom admiksije kao konzervativnijom opcijom, 100 000 'burn in' periodom i 150 000 iteracija MCMC u tri replike. Broj mogućih subpopulacija je iznosio od 2 do 7. Izabrani su sledeći modeli: model alelske učestalosti-korelisane alelske učestalosti koji omogućava i primenu lokusa koji nisu u potpunoj gametskoj ravnoteži i pretpostavlja da je svaka od mogućih subpopulacija prošla kroz genetički drift nakon razdva iz zajedničke ancestralne populacije zbog različite efektivne veličine populacije (Porras-Hurtado et al. 2013b); admiksični ancestralni model koji pretpostavlja svaka osoba u jednoj subpopulaciji može naslediti deo genetičkog materijala od pretka iz druge subpopulacije (Porras-Hurtado et al. 2013b; Pritchard, Stephens, and Donnelly 2000). Za određivanje najverodostojnije vrednosti broja subpopulacije je upotrebljen Delta K metod u platformi Harvester, koji predlaže vrednost broja subpopulacija na osnovu promene drugog reda funkcije verodostojnosti (EVANNO, REGNAUT, and GOUDET 2005). Pomoću STRUCTURE su analizirani uzorci STR i CYP genotipova između pet testiranih regiona, kao i uzorci STR i CYP genotipova Roma i ostalih ispitanika iz opšte populacije. Posebno su testirani i CYP haplotipski podaci opšte populacije u cilju provere podudarnosti sa rezultatima koji su dobijeni na osnovu CYP genotipskih podataka.

- BAPS V6 softver u kojem je broj testiranih subpopulacija (potencijalnih) iznosio od 2 do 7 (Corander, Waldmann, and Sillanpää 2003). Kao konzervativnija opcija, odabrana je analiza admiksije sa klasterovanjem individualnih genotipova što je onemogućilo algoritmu da uzme u obzir i prethodno uverenje o subpopulacionom geografskom poreklu testiranih uzoraka i na taj način uvede mogući artefakt zbog negenetičkih informacija (CORANDER and MARTTINEN 2006). BAPS algoritmom su analizirani STR i CYP genotipski podaci raspoređeni u regionalne klase, kao podaci raspoređeni na osnovu etničke kasifikacije, odnosno, subpopulacije Roma naspram ostatka opšte populacije.

Algoritmi zasnovani na redukciji dimenzionalnosti podataka:

- PCA. Za pripremu podataka, formiranje matrica genetičke udaljenosti i analizu glavnih komponenti (PCA) su korišćeni R software biblioteke 'adegenet', 'poppr' i 'hierfstat' (GOUDET 2005; Jombart 2008; Kamvar, Tabima, and Grünwald 2014; M Nei 1978). Genetička matrica je kao i u drugim algoritmima generisana na osnovu distance prema Neiju (M Nei 1978). PCA metodom su analizirani STR genotipski podaci zasnovani na regionalnoj klasifikaciji, kao i etničkoj, odnosno romske subpopulacije naspram opšte populacije. Haplotipski podaci i genotipski podaci CYP varijanti nisu posebno proveravani, s obzirom na to da su u pitanju bialelski polimorfizmi i da je broj lokusa mali za prekodiranje alela u diskretno promenjive, dok je sama PCA analiza zasnovana na neprekidno promenljivoj.

- tSNE. Algoritam tSNE je implementiran na matrici genetičke udaljenosti formiranom prema Nei 1978. pomoću R software biblioteke 'Rtsne' (Krijthe 2015; M Nei 1978). Najbitnijem parametru perplexity, kojim se određuju uticaj lokalne i globalne strukture prilikom redukcije dimenzionalnosti, je dodeljena vrednost 30. Za vizuelizaciju dvodimenzionalnih rezultata je upotrebljena osnovna funkcija R sotrvera plot(). tSNE algoritmom su analizirani STR genotipski podaci na osnovu regionalne i etničke klasifikacije.

Algoritmi zasnovani na hijerarhijskom klasterovanju:

- DBSCAN. Za identifikaciju klastera unutar uzoraka na osnovu geografske i etničke raspodele, upotrebljene su R biblioteke 'poppr' i 'dbscan' (Hahsler, Piekenbrock, and Doran 2019; Kamvar, Tabima, and Grünwald 2014). Prvo je konstruisana matrica genetičke udaljenosti, zatim upotrebljena funkcija kNNdistplot sa MinPt=5 za otkrivanje optimalne vrednosti ϵ . Funkcijom dbscan::dbscan je primenjen istoimeni algoritam na osnovu matrice genetičke udaljenosti i optimalne $\epsilon=7$. Za vizuelizaciju dobijenih rezultata je upotrebljena R biblioteka 'factorextra' (Kassambara 2017). DBSCAN algoritmom su analizirani STR genotipski podaci na osnovu regionalne klasifikacije i Roma naspram ostatka opšte populacije.

Algoritmi zasnovani na informatičkoj primeni teorije mreža:

- IGRAPH. Matrica genetičke udaljenosti prema Nei je konstruisana pomoću R biblioteka 'poppr' i 'adegenet'. Za pravljenje grafika i detekciju mreža uklanjem manje opterećenih grana je upotrebljena R biblioteka 'igraph' i funkcija cluster_edge_betweenness() (Csardi and Nepusz 2006). Na ovaj način su analizirani STR genotipski podaci na osnovu regionalne klasifikacije i Roma naspram ostatka opšte populacije.

Uzorci koji nisu uspešno genotipizirani na tri od jedanaest testiranih *CYP450* lokusa nisu bili predmet bioinformatičkih analiza i upotrebljeni su samo za određivanje učestalosti na pojedinačnim lokusima.

4. Rezultati

4.1. Rezultati genotipizacije

Od 550 DNK izolata poreklom od ispitanika obuhvaćenih ovom studijom, uspešno su genotipizirana 492 na *CYP450* lokusima koji su bili predmet istraživanja, sa varijacijama u broju uspešnih genotipizacija na svakom pojedinačnom lokusu (tabele 5 i 6). Uzorci osoba sa prebivalištem u regionu Severne Srbije su imali najniži udeo uspešne genotipizacije, svega 64 od 132 testirana uzorka. Iz okruga Beograda je uspešno genotipizirano 113 uzoraka, dok je taj broj iznosio 70 i 82 za regione Zapadne i Centralne Srbije. Iz regiona Južne i Istočne Srbije je uspešno genotipizirano 120 uzoraka (tabela 4). Uzorci izolata poreklom od osoba romske nacionalnosti su genotipizirani sa visokom uspešnošću, tako da je genetički profil koji obuhvata i *CYP450* i STR testirane lokuse, dobijen u 59 od 60 analiziranih uzoraka. Na ovaj način su dobijeni podaci koji dobro reprezentuju subpopulaciju Roma, kao i stanovništvo u pet testiranih regiona, osim u Severnoj Srbiji gde je uspešnost genotipizacije bila niža od 50%.

Severna Srbija	64
Okrug Beograd	113
Zapadna Srbija	70
Centralna Srbija	82
Južna i Istočna Srbija	120

Genotipizacijom 11 lokusa u *CYP450* genu u uzorku opšte populacije iz pet testiranih regiona, dobijene su alelske i genotipske učestalosti prikazane u tabeli 5. Broj uspešno genotipiziranih lokusa je varirao među uzorcima, tako da je najniža uspešnost genotipizacije bila na lokusu rs12248560 u genu *CYP2C19*, sa određenih 301 genotipova, a najviša na lokusu rs4646903 (M1) u *CYP1A1* genu, sa 486 određenih genotipova. Samo je jedan lokus pokazao monomorfnost (rs28371706, *CYP2D6*), dok je najveća polimorfnost opažena na rs12248560 u genu *CYP2C19*. Prema rezultatima statističkog testa Hardy-Vajnborgove ravnoteže, samo je na lokusu rs1065852 u *CYP2D6* genu došlo do odstupanja od očekivanog broja genotipova. Međutim, na još tri lokusa (rs3892097, rs4244285, rs12248560) se može primetiti da su vrednosti statistike testa blizu korigovanog praga značajnosti od 0.00465.

Tabela 5. Učestalosti *CYP450* varijanti u opštoj populaciji Srbije (regionalni uzorci). A1A1- homozigot učestalije varijante, A1A2 – heterozigot, A2A2 – homozigot manje učestale varijante P vrednost HW testa prema statistici Fišerovog testa tačne verovatnoće Prag značajnosti je korigovan prema Dunn-Sidak, 1967. GMAF prema ALFRED bazi podataka, Rajeevan, 2017.

Gen	CYP1A1		CYP2D6				CYP3A4	CYP2C9		CYP2C19	
pozicija	M1	M2	1846G>A	100C>T	2988G>A	1023C>T	- 392A>G	430C>T	1075A>C	681G>A	806C>T
SNV	rs4646903	rs1048943	rs3892097	rs1065852	rs28371725	rs28371706	rs2740574	rs1799853	rs1057910	rs4244285	rs12248560
GMAF	0.2433	0.1198	0.1	0.2557	0.05556	0.048	0.2011	0.06841	0.04224	0.1983	0.1524
alel 1	900	838	800	760	824	958	942	854	890	795	189
alel 2	72	52	168	204	80	0	22	114	78	156	90
alel 1 učestalost	0.925925926	0.941573034	0.826446	0.788382	0.911504	monomorfan	0.97717842	0.882231	0.919421	0.835084	0.775748
alel 2 učestalost	0.074074074	0.062052506	0.173554	0.211618	0.088496	monomorfan	0.02282158	0.117769	0.080579	0.163866	0.222591
A1A1	418	398	339	311	375	479	461	375	411	340	189
A1A2	64	42	122	138	74	0	20	104	68	116	90
A2A2	4	5	23	33	3	0	1	5	5	20	22
A1A1 učestalost	0.860	0.894	0.700	0.645	0.830	1	0.956	0.775	0.849	0.714	0.628
A1A2 učestalost	0.132	0.094	0.252	0.286	0.164	0	0.041	0.215	0.140	0.244	0.299
A2A2 učestalost	0.008	0.011	0.048	0.068	0.007	0	0.002	0.010	0.010	0.042	0.073
P vrednost HW testa	0.322	0.012	0.011	0.003	0.88	NA	0.22	0.66	0.22	0.019	0.02

Korigovani prag značajnosti	0.0465	0.00465	0.00465	0.00465	0.00465	NA	0.00465	0.00465	0.00465	0.00465	0.00465	0.00465
------------------------------------	--------	---------	---------	---------	---------	----	---------	---------	---------	---------	---------	---------

Genotipizacijom 11 lokusa u *CYP450* genu u uzorku romske subpopulacije, dobijene su alelske i genotipske učestalosti prikazane u tabeli 6. Najveća polimorfnost je opažena na lokusu rs1065852 u genu *CYP2D6*, dok je kao i u opštoj populaciji, na lokusu rs28371706 bio prisutan samo jedan alel. Za razliku od opšte populacije, učestalost M1 varijante (alel 1) na lokusu rs4646903 u genu *CYP1A1*, je procenjena na 0.152, koja je na granici statističke značajnosti (P vrednost= 0.0069) kada se uporedi sa učestalošću u opštoj populaciji, ali i statistički značajano viša u poređenju sa ostatkom opšte populacije (P vrednost=0.002). Ipak, ni na jednom od testiranih lokusa nije došlo do odstupanja učestalosti genotipova u odnosu na očekivane prema Hardi-Vajnbergovom principu.

Tabela 6. Broj i učestalosti CYP SNV u romskoj subpopulaciji											
A1A1- homozigot učestalije varijante, A1A2 – heterozigot, A2A2 – homozigot manje učestale varijante											
P vrednost HW testa prema statistici Fišerovog testa tačne verovatnoće											
Prag značajnosti je korigovan prema Dunn-Sidak, 1967.											
Gen	CYP1A1	CYP1A1	CYP2D6	CYP2D6	CYP2D6	CYP2D6	CYP3A4	CYP2C9	CYP2C9	CYP2C19	CYP2C19
SNV	rs4646903	rs1048943	rs3892097	rs1065852	rs28371725	rs28371706	rs2740574	rs1799853	rs1057910	rs4244285	rs12248560
pozicija	M1	M2	1846G>A	100C>T	2988G>A	1023C>T	-392A>G	430C>T	1075A>C	681G>A	806C>T
A1	100	111	97	91	93	114	117	106	103	87	62
A2	18	3	19	25	9	0	1	12	15	21	14
A1 učestalost	0.847	0.973	0.836	0.784	0.911	monomorf an	0.991	0.898	0.872	0.805	0.815
A2 učestalost	0.152	0.027	0.163	0.215	0.088	monomorf an	0.008	0.101	0.127	0.194	0.184
A1A1	41	55	40	37	42	57	58	47	45	34	26
A1A2	18	1	17	17	9	0	1	12	13	19	10
A2A2	0	1	1	4	0	0	0	0	1	1	2
A1A1 učestalost	0.694	0.964	0.689	0.637	0.823	1	0.983	0.796	0.576	0.629	0.684
A1A2 učestalost	0.305	0.017	0.293	0.293	0.176	0	0.016	0.203	0.22	0.351	0.263
A2A2 učestalost	0	0.017	0.017	0.068	0	0	0	0	0.016	0.018	0.052
P vrednost HW testa	0.222	0.013	0.816	0.349	0.659	NA	0.5	0.731	0.789	0.528	0.447
Korigovani prag značajnosti	0.00465	0.00465	0.00465	0.00465	0.00465	NA	0.00465	0.00465	0.00465	0.00465	0.00465

Genotipizacijom 11 lokusa u *CYP450* genu u uzorku opšte populacije bez uzoraka Roma, dobijene su alelske i genotipske učestalosti prikazane u tabeli 7. Kao i na nivou opšte populacije, najveća polimorfnost je opažena na lokusu rs12248560 u genu *CYP2C19*, dok je na lokusu rs28371706 opažena samo varijanta c.1023C. Na poziciji -392 u genu *CYP3A4* je varijanta G imala učestalost od svega 0.024. Dva lokusa (rs3892097 i rs1065852) u genu *CYP2D6*, su pokazala odstupanje od formiranja genotipova u skladu sa očekivanjima prema Hardi-Vajnbergovom principu i nakon korekcije praga značajnosti za višestruka poređenja. Za razliku od uzoraka Roma, u ostatku opšte populacije je na dva lokusa (rs3892097, rs1065852) u *CYP2D6* genu, opaženo odstupanje broja genotipova prema principima Hardi-Vajnbergove ravnoteže.

Tabela 7. Učestalosti CYP SNV u ostatku opšte populacije											
A1A1- homozigot učestalije varijante, A1A2 – heterozigot, A2A2 – homozigot manje učestale varijante											
P vrednost HW testa prema statistici Fišerovog testa tačne verovatnoće											
Prag značajnosti je korigovan prema Dunn-Sidak, 1967.											
Gen	CYP1A1	CYP1A1	CYP2D6	CYP2D6	CYP2D6	CYP2D6	CYP3A41	CYP2C9	CYP2C9	CYP2C19	CYP2C19
SNV	rs4646903	rs1048943	rs3892097	rs1065852	rs28371725	rs28371706	rs2740574	rs1799853	rs1057910	rs4244285	rs12248560
pozicija	M1	M2	1846G>A	100C>T	2988G>A	1023C>T	-392A>G	430C>T	1075A>C	681G>A	806C>T
A1	800	727	703	669	731	844	825	748	787	709	405
A2	54	49	149	179	71	0	21	102	63	135	121
A1	0.936	0.936	0.825	0.788	0.911	monomorf an	0.975	0.88	0.925	0.84	0.769
A2	0.063	0.067	0.174	0.211	0.088	monomorf an	0.024	0.12	0.074	0.159	0.23
A1A1	377	344	299	275	333	421	403	328	366	306	163
A1A2	46	41	105	119	65	0	19	92	55	97	79
A2A2	4	4	22	30	3	0	1	5	4	19	21
A1A1	0.882	0.884	0.701	0.705	0.83	NA	0.952	0.771	0.861	0.725	0.619
A1A2	0.107	0.105	0.246	0.28	0.162	NA	0.044	0.216	0.129	0.229	0.3
A2A2	0.009	0.01	0.051	0.07	0.007	NA	0.002	0.011	0.009	0.045	0.079
P vrednost HW testa	0.048	0.033	0.003	0.0015	0.877	NA	0.1236	0.732	0.212	0.0044	0.011
Korigovani prag značajnosti	0.00465	0.00465	0.00465	0.00465	0.00465	0.00465	0.00465	0.00465	0.00465	0.00465	0.00465

4.2. Rezultati analize molekulske varijanse

Analizom molekulske varijanse, odnosno primenom F statistike na uzorcima između testiranih regiona, dobijena je prosečna vrednost F_{ST} parametra od 0.004. Najvišu vrednost F_{ST} parametar je dostigao na lokusu rs4646903 u genu *CYP11A1* i ona je iznosila 0.011 (tabela 8). Na lokusima rs2740574 u *CYP3A4* i rs4244285 u *CYP2C19*, je vrednost F_{ST} parametra bila najniža, ispod treće decimale. Ovakvi rezultati ukazuju na mali deo genetičke varijanse koji postoji unutar testiranih regiona.

Tabela 8 – Rezultati F statistike za uzorak na osnovu regionalne klasifikacije, pojedinačni lokusi CYP gena.

Region	Pozicija >	M1	M2	1846G >A	100C> T	2988G> A	1023C> T	392A> G	430C> T	1075A >C	681G> A	806C>T
	Heterozigotnost	rs4646903	rs1048943	rs3892097	rs1065852	rs28371725	rs28371706	rs2740574	rs1799853	rs1057910	rs4244285	rs12248560
Vojvodina	Opažena	0.203	0.089	0.266	0.254	0.161	NA	0.032	0.313	0.095	0.306	0.263
	Očekivana	0.182	0.085	0.333	0.346	0.162	NA	0.031	0.264	0.091	0.328	0.266
Beograd	Opažena	0.151	0.107	0.204	0.250	0.167	NA	0.044	0.124	0.186	0.188	0.256
	Očekivana	0.155	0.151	0.237	0.316	0.153	NA	0.043	0.147	0.210	0.251	0.381
Zapadna Srbija	Opažena	0.057	0.081	0.261	0.286	0.145	NA	0.059	0.235	0.114	0.250	0.286
	Očekivana	0.055	0.107	0.306	0.368	0.162	NA	0.057	0.230	0.108	0.318	0.278
Centralna Srbija	Opažena	0.049	0.111	0.280	0.329	0.147	NA	0.049	0.220	0.134	0.317	0.333
	Očekivana	0.070	0.127	0.291	0.321	0.136	NA	0.048	0.195	0.125	0.299	0.366
Južna i istočna Srbija	Opažena	0.125	0.103	0.263	0.297	0.205	NA	0.034	0.252	0.144	0.178	0.343
	Očekivana	0.146	0.098	0.298	0.319	0.198	NA	0.050	0.245	0.148	0.189	0.371
	F_{ST}	0.011	0.003	0.004	0.002	0.002	1.000	0.000	0.007	0.009	0.000	0.009

Analizom molekulske varijanse između uzoraka Roma i ostalih uzoraka iz opšte populacije, prosečan F_{ST} je iznosio 0.0023, dok je navišu vrednost imao na lokusu rs4646903 u genu *CYP11A1*, gde se može uočiti i najveća razlika između opažene i očekivane heterozigotnosti u romskoj populaciji. Prema niskoj vrednosti F_{ST} parametra na lokusima *CYP450* gena se može zaključiti da je razlika između učestalosti genotipova u romskoj populaciji, izuzetno mala u odnosu na ostatak opšte populacije (tabela 9).

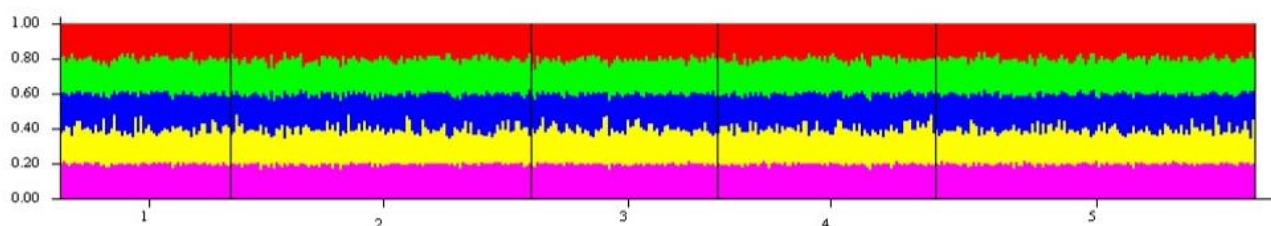
Tabela 9 - Rezultati F statistike za uzorak na osnovu entičke klasifikacije, pojedinačni lokusi. H – heterozigotnost (opažena i očekivana)												
Pozicija		m1	m2	1846G>A	100C>T	2988G>A	1023C>T	392A>G	430C>T	1075A>C	681G>A	806C>T
SNV		rs4646903	rs1048943	rs3892097	rs1065852	rs28371725	rs28371706	rs2740574	rs1799853	rs1057910	rs4244285	rs12248560
Romi	Očekivana H	0.259	0.051	0.274	0.338	0.161	0.000	0.017	0.183	0.222	0.313	0.301
	Opažena H	0.305	0.018	0.293	0.293	0.176	0.000	0.017	0.203	0.220	0.352	0.263
Ostali	Očekivana H	0.118	0.118	0.289	0.333	0.161	0.000	0.048	0.211	0.137	0.269	0.354
	Opažena H	0.108	0.106	0.246	0.281	0.162	0.000	0.045	0.216	0.129	0.230	0.300
F_{ST}		0.012	0.003	0.000	0.000	0.000		0.001	0.000	0.004	0.001	0.001

Analizom varijanse haplotipskih podataka, generisanih na osnovu testiranih SNV u *CYP450* i pomoću Haploview softvera (Barrett et al. 2005), dobijene su F_{ST} vrednosti prikazane u tabeli 10. Prosečna vrednost F_{ST} parametra je iznosila 0.0021. Pojedinačne vrednosti F_{ST} između regiona su ispod 0.01, a najviša između regiona Severne Srbije (Vojvodine) i opšte populacije. Kao i rezultati F statistike bazirane na pojedinačnim SNV lokusima u *CYP450* genima, rezultati bazirani na haplotipovima nisu ukazali na značajna odstupanja između heterozigotnosti u regionima, u poređenju sa heterozigotnošću u opštoj populaciji. Ovi rezultati ukazuju i na mala odstupanja između F_{ST} parametra procenjenih na pojedinačnim lokusima i F_{ST} parametara procenjenih na osnovu inferiranih haplotipova.

Tabela 10 - Rezultati analize molekulske varijanse haplotipskih podataka CYP450 u opštoj populaciji Srbije.		
Region	prosečni F_{ST}	P vrednost
Severna Srbija	0.00410	0.34416
Beograd	0.00057	0.74743
Zapadna Srbija	0.00176	0.61396
Centralna Srbija	0	0.65614
Južna i istočna Srbija	0	1

4.3. Rezultati algoritama za detekciju genetičke struktuiranosti

Primenom STRUCTURE algoritma za detekciju strukture populacije na *CYP* haplotipskim podacima iz uzoraka testiranih regiona, generisan je grafički prikaz (bar plot na slici 42) gde se može učitati da je svaki genotip imao sličan obrazac porekla iz pretpostavljenih subpopulacija. Broj pretpostavljenih populacija se kretao u intervalu od 2 do 7, a za svaku vrednost je određena verodostojnost prikazana na slici 43. prema Delta K metodu (EVANNO, REGNAUT, and GOUDET 2005). Najniža vrednost negativnog logaritma funkcije verodostojnosti, kao i mera raspršenosti oko vrednosti ispod moda funkcije, može se primetiti za 1 i 3 pretpostavljene subpopulacije. Međutim, Delta K parametar ukazuje da su 6 subpopulacija najoptimalniji izbor (slika 43).

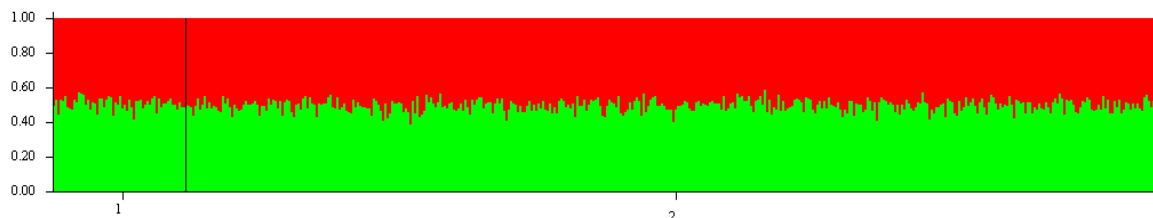


Slika 42 - Barplot prikaz nakon analize *CYP450* haplotipova za regionalnu klasifikaciju. Brojevi označavaju regione (1-Severna Srbija, 2- Beograd, 3 – Zapadna Srbija, 4 – Centralna Srbija, 5 – Južna i Istočna Srbija).

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
1	4	-25716.050000	0.129099	—	—	—
2	4	-26023.100000	161.313670	-307.050000	45.125000	0.279735
3	4	-26285.025000	150.746705	-261.925000	185.000000	1.227224
4	4	-26731.950000	216.867094	-446.925000	140.700000	0.648784
5	4	-27038.175000	198.043924	-306.225000	299.066667	1.510103
6	3	-27045.333333	283.313631	-7.158333	539.641667	1.904750
7	3	-27592.133333	374.526065	-546.800000	—	—

Slika 43 - Najoptimalniji broj klastera prema Delta K metodu na osnovu haplotipova *CYP* lokusa u genima *CYP1A1*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*.

Kada su u pitanju rezultati STRUCTURE na osnovu *CYP450* haplotipskih podataka između uzoraka Roma i uzoraka iz ostatka opšte populacije, nisu pronađene razlike koje bi ukazivale na različito poreklo subpopulacija. Na slici 44 se može videti grafički prikaz (bar plot) prema kojem svi genotipovi imaju iste proporcije pretpostavljene populacije porekla. Prema Delta K metodu, najveća verodostojnost postoji za 4 pretpostavljene populacije, ali su vrednosti negativnog logaritma funkcije verodostojnosti vrlo bliske, tako da se ne mogu uzeti kao jasna determinanta broja mogućih subpopulacije (slika 45).



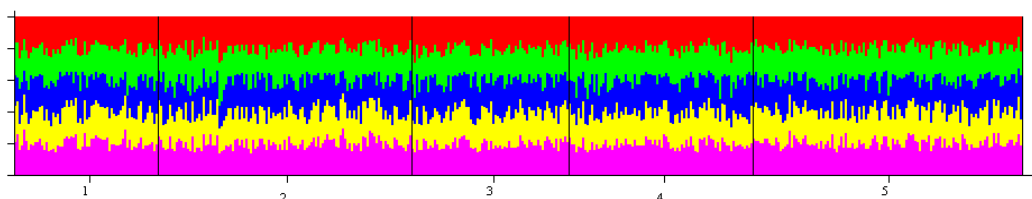
Slika 44 – Najoptimalniji broj klastera na osnovu etničke raspodele *CYP450* haplotipskih podataka. Brojem 1 su obeleženi genotipovi iz romske subpopulacije, brojem 2 iz ostatka opšte populacije.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
2	3	-2680.400000	0.953939	—	—	—
3	3	-2621.400000	4.326662	59.000000	120.600000	27.873685
4	3	-2683.000000	2.645751	-61.600000	145.933333	55.157615
5	3	-2890.533333	35.235541	-207.533333	3.000000	0.085141
6	3	-3101.066667	241.163437	-210.533333	247.966667	1.028210
7	3	-3063.633333	41.672333	37.433333	—	—

Slika 45 - Najoptimalniji broj klastera na osnovu etničke klasifikacije. Rezultati delta K metoda u Harvesteru. Evanno 2011.

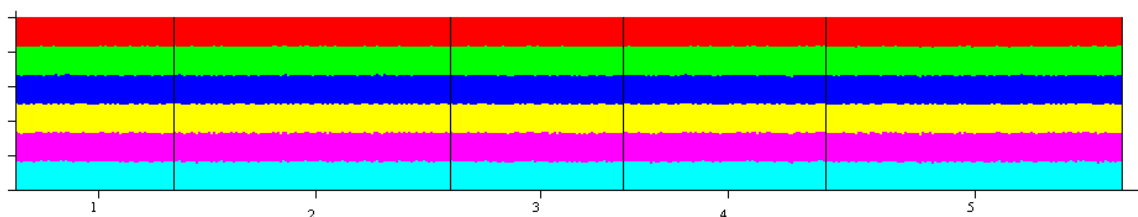
Prema rezultatima STRUCTURE prikazanim na slici 46 se može videti grafički prikaz particionisanja opšte populacije Srbije na osnovu genotipskih podataka 11 testiranih lokusa *CYP450* gena. Iako je obrazac sličan i na osnovu *CYP450* genotipskih (slika 46), kao i haplotipskih podataka

(slika 42). Moguće je uočiti nešto veće oscilacije između pojedinačnih genotipova, međutim, vizuelno se ne mogu uočiti pravilnosti u odstupanjima koje bi ukazale na grupisanje unutar nekih od regiona.



Slika 46 - Barplot prikaz analize *CYP* genotipskih podataka za regionalnu klasifikaciju. Brojevi označavaju regione (1-Severna Srbija, 2- Beograd, 3 – Zapadna Srbija, 4 – Centralna Srbija, 5 – Južna i Istočna Srbija).

Na osnovu STR podataka testiranih regiona, STRUCTURE nije ukazao na postojanje klastera (slika 47). Iako se prema rezultatima Delta K metoda može uočiti da se na nivou 3 subpopulacije našla najoptimalnija vrednost negativnog logaritma verodostojnosti, STRUCTURE nije testirao mogućnost da svi genotipovi potiču iz jedne populacije, već je poredio promene funkcije verodostojnosti i mere raspršenosti ispod modova za svaki broj pretpostavljenih subpopulacija od 2 do 10 (slika 48).



Slika 47 – Barplot prikaz analize STR podataka za regionalnu klasifikaciju. Brojevi označavaju regione (1-Severna Srbija, 2- Beograd, 3 – Zapadna Srbija, 4 – Centralna Srbija, 5 – Južna i Istočna Srbija).

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
2	3	-23869.466667	18.032563	—	—	—
3	3	-24298.733333	79.844182	-429.266667	337.500000	4.226983
4	3	-24390.500000	17.841244	-91.766667	66.333333	3.717977
5	3	-24548.600000	177.248498	-158.100000	146.733333	0.827840
6	3	-24559.966667	320.665470	-11.366667	297.200000	0.926823
7	3	-24868.533333	550.695291	-308.566667	577.833333	1.049280
8	3	-24599.266667	543.530913	269.266667	213.400000	0.392618
9	3	-24543.400000	291.772189	55.866667	516.733333	1.771016
10	3	-25004.266667	399.247409	-460.866667	—	—

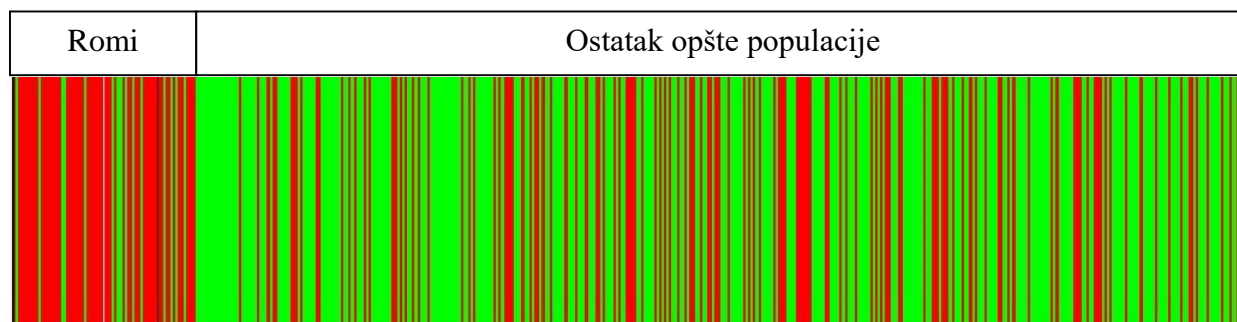
Slika 48 - Najoptimalniji broj klastera na osnovu regionalne klasifikacije i STR podataka. Rezultati delta K metoda u Harvesteru. Evanno 2011.

U BAPS softveru su analizirani STR genotipski podaci na osnovu uzoraka iz pet testiranih regiona i na osnovu etničke pripadnosti (Romi i ostatak opšte populacije). BAPS nije izdvojio posebne klasterne u opštoj populaciji, već jednu grupaciju i šum na osnovu STR podataka (slika 49). Međutim, između uzoraka Roma i drugih iz opšte populacije, obzervirana su dva klastra (slika 50). U klasteru 1 se našlo 46 roma, dok je u klasteru 2 bilo svega 11. Klaster 1 je veličine 187, dok je klaster 2 veličine 297. Verovatnoća ovakvog ishoda grupisanja na slučajan način je prema Fišerovom testu tačne verovatnoće manja od 0.00001.

Severna Srbija	Beograd	Zapadna Srbija	Centralna Srbija	Južna i istočna Srbija
----------------	---------	----------------	------------------	------------------------



Slika 49 - Barplot prikaz analize BAPS softverom na osnovu STR podataka iz pet regiona.



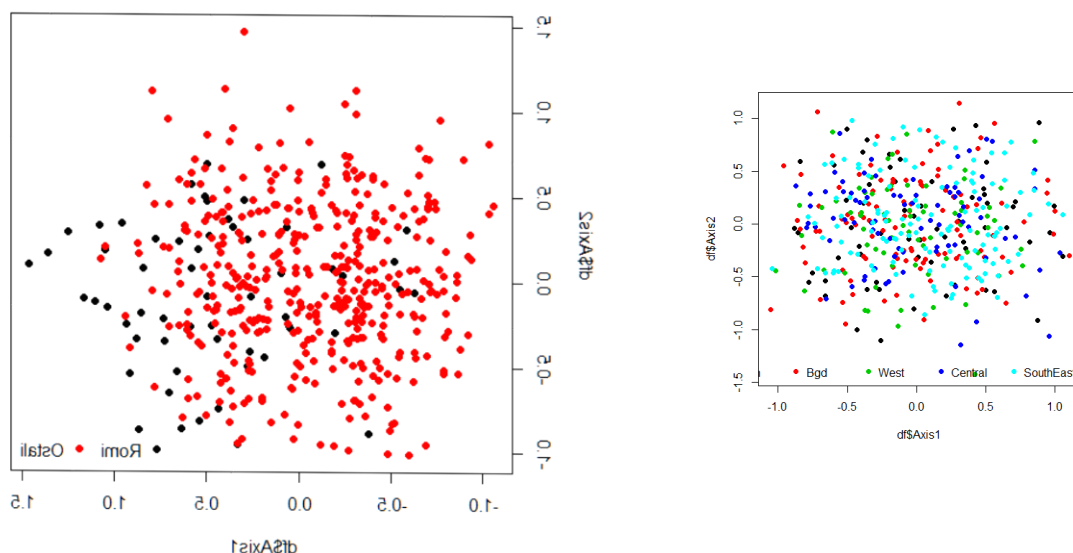
Slika 50 - Barplot generisan u BAPS na osnovu analize STR podataka. Romi i ostatak opšte populacije

Diferencijalnim poređenjem alela STR lokusa u klasterima koji su dobijeni BAPS algoritmom, izdvojene su aleli koji se nalaze samo u klasteru K1 ili K2 (tabela 11). Aleli dužina 16.2 na lokusu D5S818, 9 na lokusu D2S1338, 14 na lokusu D13S317 i 21 na lokusu D18S51 su se mogli pronaći samo u K2 klasteru gde je bio manji broj STR genotipova Roma, što sugerise da bi upravo ovi aleli mogli biti najveći doprinosioci grupisanju u različite klasterne.

Tabela 11 – Diferencijalni prikaz STR alela u klasterima K1 i K2

	K1	K2
TH01	9.3; 11; 15	
D3S1358	25	
vWA	20	
D21S11	33.2; 31; 31.2; 28.2; 32	
TPOX	10; 9; 12	
D7S820	13	
D19S433	17.2	
D5S818		16.2
D2S1338		9
D16S539	14	
CSF1PO	9	
D13S317		14
FGA	18; 20.2; 22.2; 23.2; 26	
D18S51	10; 11; 19; 20	21

Analizom glavnih komponenti (PCA) podataka dobijenih genotipizacijom STR lokusa, nisu uočeni klasteri koji se posebno izdvajaju u poređenjima između testiranih regiona (slika 51, desno), niti između Roma i ostalih ispitanika iz opšte populacije (slika 51, levo). Međutim, može se primetiti da uzorci Roma imaju više vrednosti na apscisi koja odgovara prvoj glavnoj komponenti nakon redukcije dimenzionalnosti. Imajući u vidu proporcionalnost uzoraka prema broju stanovnika i broju Roma koji žive na prostoru Republike Srbije, uticaj varijanse lokalnih struktura ne bi trebalo da je doprineo devijaciji globalne varijanse u PCA analizi.



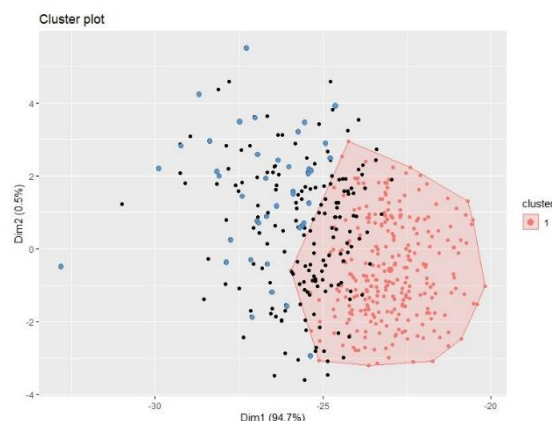
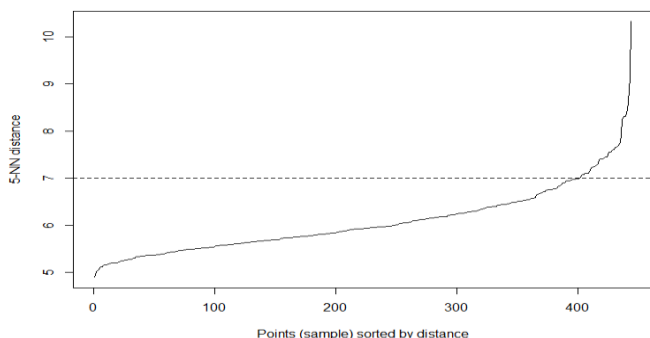
Slika 51 – Rezultati PCA na osnovu etničke klasifikacije (levo) i na osnovu regionalne klasifikacije (desno). Na slici levo su različitim bojama obeleženi Romi (teget) i ispitanici iz ostatka opšte populacije (crveno). Na slici desno su različitim bojama obeleženi ispitanici iz različitih regiona: Severna Srbija- crno, Beograd – crveno, Zapadna Srbija – zeleno, Centralna Srbija – plavo, Južna i Istočna Srbija – svetlo plavo.

Primenom DBSCAN algoritma sa prethodno određenom vrednošću $\epsilon=7$ (slika 52), na STR podacima testiranih regiona nije uočeno izdvajanje dva ili više klastera (slika 54). Ipak, može se konstatovati da je algoritam na osnovu podataka koje je učitao izdvojio jedan klaster koji čini većina genotipiziranih osoba (327), dok se manji udeo našao u šumu (155) (tabela 12).

Kada su u pitanju rezultati dobijeni DBSCAN algoritmom na osnovu genotipskih podataka Roma i ostalih ispitanika iz opšte populacije, takođe je opažen samo jedan klaster veličine 274 ispitanika od kojih je 18 bilo od osoba romskog etniciteta (slika 53). Van klastera, odnosno u šumu se našao 41 ispitanik iz uzorka Roma, što je prema Fišerovom testu tačne verovatnoće očekivano sa izuzetno malom verovatnoćom ($2.999e-05$).

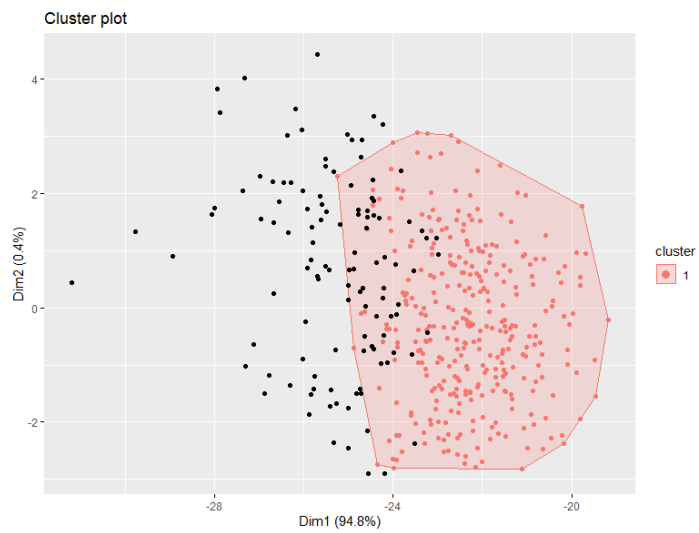
Tabela 12 - Raspodela STR genotipova geografskog/regionalnog uzorka između klastera i šuma u rezultatima DBSCAN algoritma.

	Severna Srbija	Beograd	Zapadna Srbija	Centralna Srbija	Južna i Istočna Srbija
Klaster	43	90	54	60	80
Šum	21	23	16	22	40



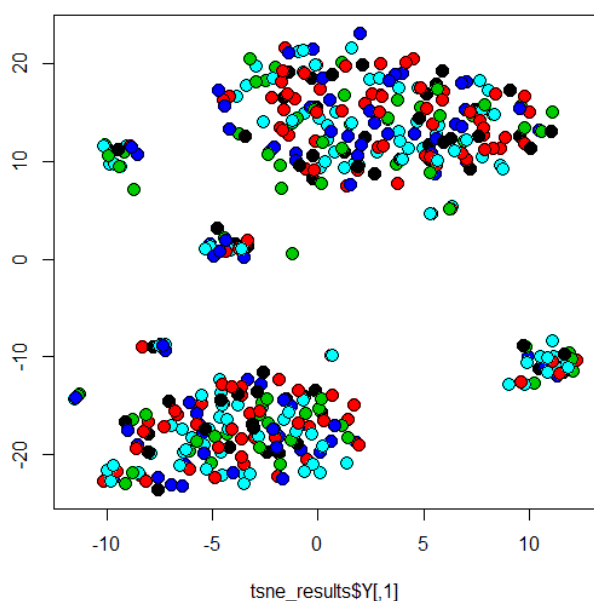
Slika 52 - Optimalna vrednost ϵ parametra za DBSCAN algoritam

Slika 53 - Grafikon raspršenosti na osnovu DBSCAN algoritma i STR podataka prema etničkoj klasifikaciji. Romi van klastera su obelježeni plavim tačkama. Podaci su vizuelizovani pomoću R softverskog paketa 'factoextra'.

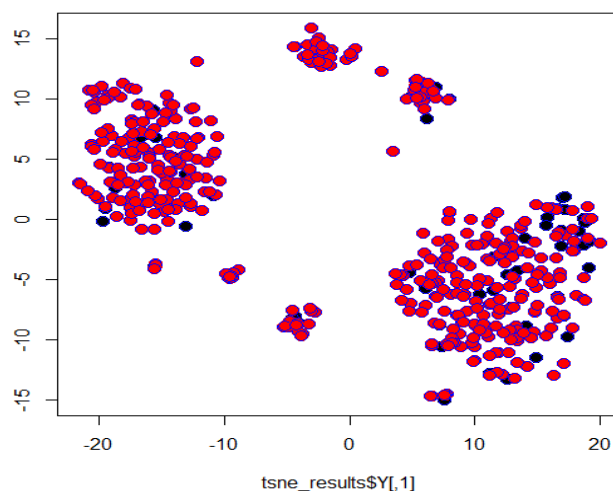


Slika 54 - Grafik raspšenosti na osnovu DBSCAN algoritma i STR podataka prema regionalnoj klasifikaciji.

tSNE algoritmom su vizuelizovani STR podaci regionalnih uzoraka i Roma naspram ostalih ispitanika iz opšte populacije nakon redukcije dimenzionalnosti (slike 55 i 56). Na nivou opšte populacije, uzorci iz testiranih regiona se grupišu u dva veća i diskretna klastera i tri manje grupacije (slika 55). Nije uočen obrazac prema kojem se u nekom od klastera nalaze dominantno genotipovi poreklom iz nekog od testiranih regiona, što potvrđuje i statistika Fišerovog testa tačne verovatnoće (P vrednost = 0.21). Rezultati su drugačiji kada su u pitanju STR podaci Roma i ostalih ispitanika (slika 56). Iako postoje dva veća klastera i u ovim podacima, broj Roma u prvom klasteru je 12, dok je u drugom 35. Prema statistici Fišerovog testa tačne verovatnoće, ovakav ishod se može očekivati u 2.6% slučajeva kada je razdvajanje u dva klastera nasumično.

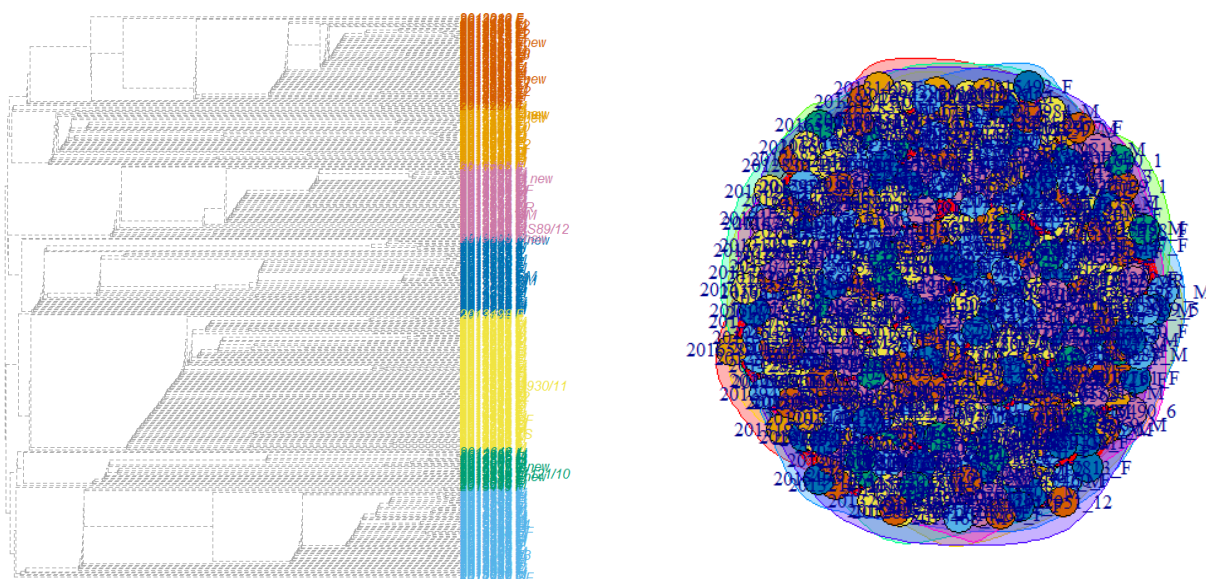


Slika 55 – Klasteri generisani tSNE algoritmom na osnovu regionalne klasifikacije



Slika 56 – Klasteri generisani tSNE algoritmom na osnovu etničke klasifikacije. Romi su obeleženi teget bojom, dok je ostatak populacije obeležen roze bojom.

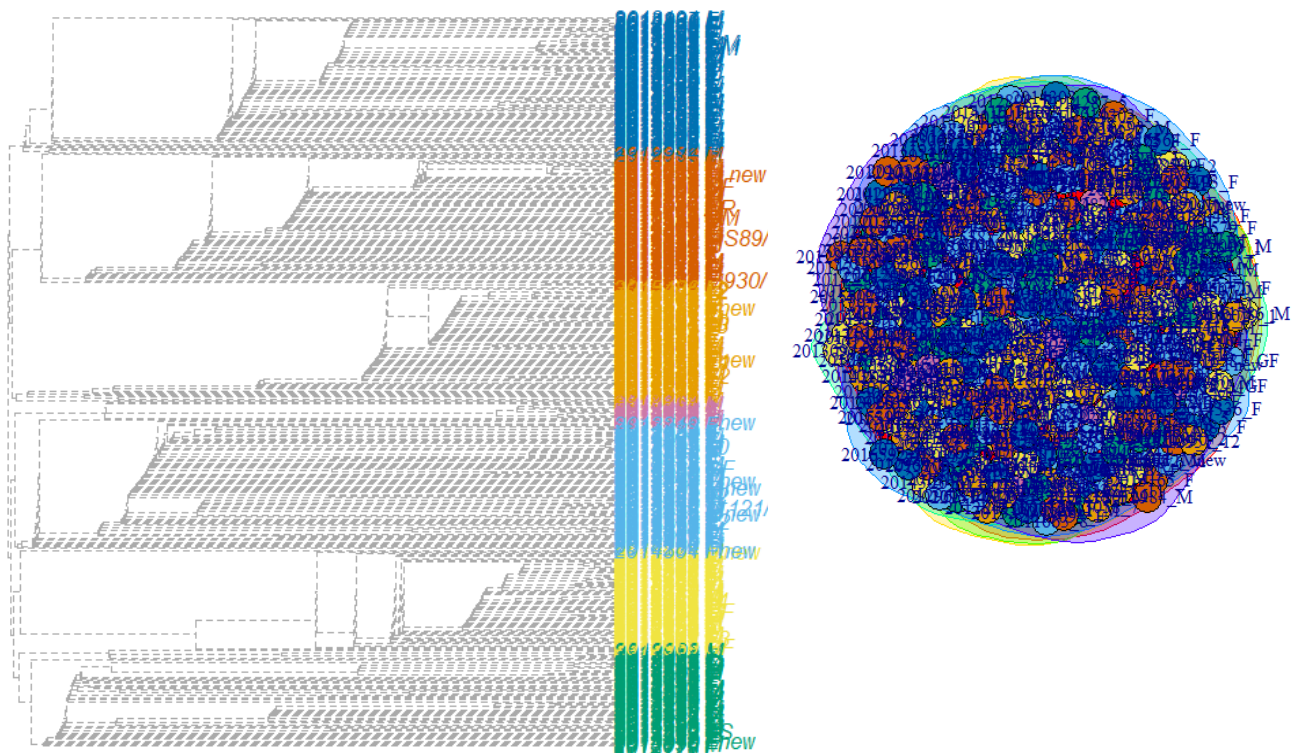
Girvan Newman algoritam je ukazao na 7 mogućih klastera u oba skupa STR podataka, na osnovu regionalne klasifikacije i na osnovu etničke klasifikacije. Iz oba skupa podataka, pre pokretanja algoritama je isključeno 48 genotipova iz neromskog uzorka opšte populacije zbog pravljenja objekta grafika koji zahteva precizno određivanje 'težine' grana između čvorišta. Vizuelnim pregledom se može uočiti da predloženi klasteri ne odslikavaju jasan obrazac grupisanja (slika 57 i 58), dok sama raspodela u predložene klasterne nema odstupanja od one koja bi se mogla očekivati na osnovu slučajnosti (tabele 13 i 14), što je prema Fišerovom testu tačne verovatnoće procenjeno kao ishod sa verovatnoćom od 0.549. Iako je uočljiva loša preglednost grafika koje generiše 'IGRAPH' softverska biblioteka za vizuelno zaključivanje o potencijalnim klasterima unutar većeg uzorka, dendogramski prikaz klasterisanja podataka daje dobar pregled, međutim, ni na njemu ne postoji jasno razgraničenje na osnovu modularnosti podataka kojom se određuju klasteri (slike 57 i 58).



Slika 57 – Grafički prikaz dendograma i grafika generisanih GN algoritmom.

Tabela 13 – presek klastera generisanih Girvan Newman algoritmom i geografska raspodele.

	K1-48	K2-72	K3-28	K4-109	K5-60	K6-70	K7-57
Sever	11	12	6	17	2	7	8
Beograd	9	20	8	31	15	17	11
Zapad	4	11	6	18	10	9	11
Centralna	10	11	4	16	16	15	10
Jugoistok	14	18	4	27	17	22	17
p-value = 0.5497							



Slika 58 - Dendrogram i grafik generisani GN algoritmom na osnovu etničke klasifikacije.

Tabela 14 – presek klastera generisanih Girvan Newman algoritmom i etničke raspodele							
klaster	K1	K2	K3	K4	K5	K6	K7
veličina	72	79	59	61	83	79	11
broj roma	5	2	3	4	2	5	0
p-value=0.7325							

5. Diskusija

Proučavanje genetičke varijabilnosti je osnova za razumevanje genetike, rekonstrukciju evolucionih događaja u nekoj populaciji, kao i osnova za primenu genetike u drugim naučnim disciplinama, posebno u medicini (Nevo 2001). Napretkom molekularne biologije stvorena je osnova za razvoj molekularne genetike i izučavanje naslednih osobina na nivou molekula DNK (Kassem, Girolami, and Sanoudou 2012), što je otvorilo čitav niz novih naučnih polja koje integrišu osnove genetike i znanja iz drugih naučnih disciplina. Takva jedna naučna disciplina je i farmakogenetika koja baštini jedinstveni pristup u genetičkom razmevanju varijabilnosti metabolizma supstanci koje se koriste u lečenju ljudi (Bishop 2018). U užem smislu, farmakogenetika se bavi proučavanjem odnosa između genetičke i fenotipske varijabilnosti, dovodeći u vezu genske varijante sa promenama u metabolizmu lekova. U širem kontekstu, farmakogenetika se može posmatrati i sa populaciono genetičkog nivoa, proučavanjem raspodele učestalosti genskih varijanti i raspodele učestalosti fenotipova prema definisanoj metaboličkoj klasifikaciji (Suarez-Kurtz 2005). Značaj proučavanja genetičke osnove varijabilnosti metabolizma lekova na populaciono genetičkom nivou ima više aspekata. S jedne strane, određivanje učestalosti genskih varijanti koje utiču na metabolizam lekova, doprinosi usmeravanju pažnje naučne javnosti ka onim varijantama koje imaju visoke učestalosti i kao takve najviše doprinose promenama u fenotipskoj raspodeli unutar opšte populacije. Sa druge strane, populaciono genetičkim sagledvanjem karakteristika raspodele učestalosti genskih varijanti, mogu se izvesti zaključci o uniformnosti raspodele učestalosti genskih varijanti između subpopulacija koje čine opštu populaciju. To praktično znači da zbog razlike u učestalostima koje postoje između subpopulacija, uticaj neke farmakogenetičke varijante ne mora biti jednak u svim subpopulacijama. Otkrivanjem subpopulacija i određivanjem njima specifičnih učestalosti genskih varijanti, stvara se mogućnost preciznijeg preventivnog delovanja u smislu odabira najinformativnijih genskih varijanti za ciljano farmakogenetičko testiranje. Odabir genskih varijanti čije učestalosti mogu biti od interesa za testiranje u populaciji, zavisi direktno od jačine dokaza o njihovom uticaju na metabolizam lekova. Do sada se kao najbolji kandidati za ovakvu vrstu testiranja pojavljuju geni iz supergenske familije *CYP450*, za koje je pokazano da dovode do promena u fazi I metabolizma većeg broja lekova koji su u kliničkoj upotrebi (Bush et al. 2016; Cook et al. 2016).

Genotipizacijom *CYP450* funkcionalnih varijanti u reprezentativnom uzorku iz opšte populacije sa neke teritorije definišu se raspodele učestalosti alela i genotipova, koje ne moraju biti jednake raspodelama prisutnim u drugim susednim i nesusednim populacijama. Pored toga, grupisanjem genetičkih podataka u skladu sa pretpostavkama o postojanju zasebnih celina unutar populacije, između kojih postoji smanjeni protok gena, može se otkriti obrazac raspodele unutar jedne populacije, koji ne mora biti uniforman i u populacionoj genetici je poznat pod terminom genetička struktura populacije. Polazna osnova za otkrivanje genetičke strukture je utvrđivanje subpopulacija u kojima alelske učestalosti mogu odstupati od onih u opštoj populaciji. Međutim, definisanje subpopulacije nije obavezno jednostavno niti precizno. Najčešći kriterijumi za svrstavanje osoba u subpopulaciju jesu obeležja koja dovode do veće verovatnoće intra nego intersubpopulacionog ukrštanja, a takva obeležja su mesto stanovanja i etnička pripadnost. Pored *CYP450* funkcionalnih varijanti, na genetičku strukturu, a time i postojanje subpopulacija, mogu ukazati i drugi genetički markeri, kao što su mikrosatelitski (STR) lokusi koji se, zbog selektivne neutralnosti i visoke polimorfnosti, koriste za identifikaciju ljudi (Rosenberg et al. 2005). U ovom radu je, na osnovu reprezentativnog uzorka, kako iz opšte populacije podeljene na stratum prema regionima koji sačinjavaju teritoriju Republike Srbije, tako i romske subpopulacije, genotipizirano 11 funkcionalnih varijanti u genima *CYP1A1*, *CYP2C9*, *CYP2C19*, *CYP2D6* i *CYP3A4* koje su bitne za metabolizam velikog broja supstrata, kao i 15 mikrosatelitskih lokusa koji se rutinski koriste u identifikaciji ljudi i analizi srodničkih odnosa. Na

osnovu dobijenih rezultata genotipizacije, sprovedena je informatička analiza koja je za cilj imala detekciju potencijalnih subpopulacija, kao i otkrivanje razlika između opšte populacije i romske subpopulacije za koju se očekivalo da je genetički najdiferenciranija usled znatne socijalne izolacije. Korist od utvrđivanja alelskih učestalosti *CYP450* gena u opštoj populaciji Srbije se odnosi prvenstveno na razvoj i primenu individualizacije lečenja, poznatijim pod nazivom personalizovana medicina, ali i razvoju panela genetičkih analiza na obdukcionom materijalu, koje se često nazivaju i 'molekularnom autopsijom'.

Individualizacija terapije svakako nije nepoznat koncept i uvek je prisutna u lečenju, kako svaki pacijent dobija dozu leka na osnovu fizičkih svojstava, kao što su telesna masa, fiziološki i biohemijski parametri, starosne godine, druge pridružene bolesti i stanja. Uvođenjem novog svojstva, genetičke osnove metabolizma lekova, samo je nastala nova dimenzija u prilagođavanju terapije pacijentu. Odgovor na pitanje zašto je farmakogenetika toliko popularizovana u naučnim istraživanjima u poslednje dve decenije treba tražiti u napretku tehnologije i boljem razumevanju uticaja genetičke varijabilnosti na primenu terapije. Ipak, publicitet koji je personalizovana medicina zadobila uveliko prevazilazi realnu i racionalnu procenu doprinosa lečenju (Janssens and Deverka 2014). Po sredi je problem u pojednostavljenom pripisivanju fenotipskih odlika genskim varijantama, sa zanemarivanjem kompleksnosti u određivanju fenotipa na osnovu genotipa, koja je više pravilo nego izuzetak. Tako neretko dolazi do nesklada između pretpostavljenog odgovora na osnovu farmakogenetičke analize i opaženog na osnovu merenja metabolita ili kliničkog odgovora na terapiju (Osnabrugge et al. 2015; Shah and Smith 2015). CYP enzimi su dobar primer za razumevanje kompleksnosti farmakogenetičke predikcije iz razloga što jedan supstrat (lek) može biti metabolisan od strane više CYP enzima, ali i zato što su mnogi CYP enzimi inducibilni ili pak inhibirani drugim lekovima (Manikandan and Nagini 2018). Jasno je da je tačnija procena metaboličke aktivnosti organizma na osnovu farmakokinetičkih merenja metabolita nego na osnovu farmakogenetičke predikcije. Ipak, nije uvek moguće svakog pacijenta podvrgnuti farmakokinetičkom testiranju, niti je tehnički jednostavno i uvek izvodljivo. Upravo u tehničkoj jednostavnosti i dostupnosti leži ključni razlog za uvođenje farmakogenetičkih testiranja u kliničku praksu. Kada je genotipizacija značajno ekonomski isplativija i kada postoji jaka asocijacija genskih varijanti sa alteracijama u odgovoru na terapiju, postoji razumna opravdanost za primenom farmakogenetičkog pristupa u personalizaciji lečenja. Između ostalog, opšti trend je da sve veći broj lekova koji se upotrebljavaju na području Evropske unije zahteva i analizu genetičkih varijanti koje imaju uticaja na metabolizam tog leka (Ehmann et al. 2015). Postoje dva, tehnički različita, načina da se utvrde genetičke varijante za koje je prethodno dokazano da uvode izmene u metabolizmu lekova, tzv. funkcionalne varijante. Prvi i obuhvatniji način utvrđivanja većeg broja funkcionalnih varijanti podrazumeva upotrebu druge ili treće generacije sekvenciranja, što sa sobom nosi velike materijalne troškove, a samim tim i manju dostupnost. Drugi način se odnosi na ciljano utvrđivanje funkcionalnih varijanti (genotipizacija poznatih lokusa) za koje je unapred pokazano da imaju visoku učestalost u nekoj populaciji. Najdostupnija metoda za ciljanu genotipizaciju jeste upotreba TaqMan eseja na RT-PCR, zbog svoje jednostavne izvedbe, široke primene RT-PCR aparata i kratkog perioda za dobijanje upotrebljivih rezultata. Međutim, kao što je već naglašeno, ciljana genotipizacija zahteva prethodno utvrđivanje učestalosti funkcionalnih varijanti u nekoj populaciji, kako bi se izbeglo nepotrebno genotipiziranje neinformativnih varijanti, koje odlikuje niska učestalost, te posledično uvećavanje troškova farmakogenetičkog testiranja.

Od 11 testiranih u ovom radu, prva je M1 varijanta tačkastog polimorfizma rs4646903, koja uvodi izmenu T6235C u 3' netranslatirajućem regionu informacione RNK gena *CYP1A1*. Ova varijanta je uglavnom statistički asociirana sa raznim oblicima neoplastičnih transformacija ćelija, a najverovatnije da svoj mehanizam delovanja na karcinogenezu ostvaruje preko regulacije ekspresije

koja je zavisna od mikro RNK i dovodi do povećanja količine CYP1A1 enzima koji aktivira PAH jedinjenja u formu koja oštećuje DNK (Jorge-Nebert et al. 2010; Ramamoorthy and C. Skaar 2011; Rotunno et al. 2009). Međutim, ta asocijacija je dovedena u pitanje zbog nepostojanja jasnog mehanizma kancerogeneze, ili pak kontradiktornih podataka koji potenciraju čak i protektivnu ulogu CYP1A1 enzima u metabolizmu kancerogenih supstanci, tako na primer u miševima koji nemaju *CYP1A1* gen, dolazi do ispoljavanja toksičnog efekta benzo[a]pirena (PAH jedinjenje) na aorte (Uno et al. 2004). Da asociranost sa karcinogenezom, odnosno aktivacijom PAH jedinjenja, treba shvatiti vrlo uslovno, govori i primer da je 3 puta viša učestalost M1 varijante u subsaharskim populacijama i oko 10 puta niža učestalost malignih oboljenja u čijoj kancerogenezi učestvuju PAH jedinjenja, što je nasuprot tezi da bi viša učestalost M1 varijante u nekoj populaciji doprinela povećanju broja oboljenja izazvanih PAH jedinjenjima (Fathimah, Adhiyanto, and Hendarto 2017). Međutim, kako je kancerogeneza poligena bolest, teško je oceniti uticaj učestalosti jedne varijante *CYP1A1* bez poznavanja i drugih genetičkih varijanti koje učestvuju u kancerogenezi PAH jedinjenjima, a koje nisu bile predmet ove studije sprovedene na uzorku opšte populacije u Srbiji i usmerene ka funkcionalnim varijantama samo u *CYP450* genima (Lee et al. 2005).

Učestalost M1 varijante na globalnom novou je 0.24, dok je u populacijama evropskog porekla blizu 0.09 (Rajeevan et al. 2017). Procenjena učestalost M1 varijante u opštoj populaciji Srbije iznosi 0.063, što jeste nešto niža vrednost u odnosu na prosek od 0.09 na nivou svih evropskih populacija. Poređenjem sa susednim populacijama na osnovu dostupnih podataka, može se uočiti neobično odstupanje u učestalosti M1 varijante. Tako je učestalost u Hrvatskoj populaciji 0.14, dok je u Mađarskoj čak 0.2 (Pašalić and Marinković 2017; Szanyi et al. 2012). Iako statistički veoma značajne (P vrednosti <0.0001), ove razlike treba prihvatiti uslovno, s obzirom na to da postoji značajna razlika u metodologiji uzorkovanja, jer u studijama koje dokumentuju raspodelu M1 varijante u Hrvatskoj i Mađarskoj, odabir učesnika istraživanja nije bio stratifikovan niti oslobođen mogućih pridruženih faktora s obzirom na to da su uzorkovanja vršena pretežno nad hospitalizovanim pacijentima (Göring, Terwilliger, and Blangero 2001). Između pet regionalnih stratumata u Srbiji se primećuje izostanak većih varijacija u učestalostima M1 varijante, što se može objasniti višim stepenom migracija između tih regiona, koje su doprinele unifikaciji populacije (Tabela 5). Ovaj podatak govori u prilog tezi da bi povećana incidenca raka pluća ili drugih pulmoloških oboljenja u Srbiji, bila više uslovljena samim sredinskim faktorima, nego interakcijom genetičkih varijanti sa sredinskim faktorima (Sun, Schiller, and Gazdar 2007).

S druge strane, razlika u učestalostima M1 varijante između Roma (0.152) i ostatka opšte populacije u Srbiji (0.063), ostaje statistički značajna i nakon korekcije praga značajnosti za višestruka poređenja (P vrednost $=0.002$). Upravo se na lokusu rs4646903 i opaža najviša vrednost F_{ST} parametra (Tabela 9), koja se u prvi mah može objasniti izolovanošću romske subpopulacije u kojoj su genetički drift i viši stepen inbridinga mogli dodatno uticati na promenu učestalosti M1 varijante, međutim tu postoje manje nedoslednosti koje ne dozvoljavaju jednostavan zaključak. Iako u apsolutnim merama vrednost F_{ST} od 0.012 nije naročito visoka (Frankham et al. 2002), ona se može dovesti u vezu sa neočekivano većom heterozigotnošću u romskoj subpopulaciji (0.305 u odnosu na 0.259 očekivane heterozigotnosti). S obzirom na to da se obično očekuje povećana homozigotnost kao posledica veće stope inbridinga u subpopulaciji, povećana heterozigotnost se može sagledati u svetlu nepostojanja M1 homozigota u ispitivanom uzorku i špekulisati o mogućnoj vezi sa prirodnom selekcijom. Na osnovu alelske učestalosti M1 u romskoj subpopulaciji, koja je iznosila 0.153, očekivani broj homozigota M1M1 u uzorku od 59 genotipova, bi iznosio 1 do 2, dok u testiranom uzorku nije opažen ni jedan, što i nije veliko odstupanje. Imajući u vidu veličinu uzorka, nije moguće donositi jednostavne zaključke o razlozima smanjena broja M1 homozigota usled dejstva selekcije, iz razloga što bi selekcija morala dovesti do manje stope preživljavanja na prenatalnom nivou ili u

ranij životnoj dobi, kako bi broj M1 homozigota bio manje učestao u uzorku subpopulacije, naročito među osobama starije životne dobi. Međutim, dizajn naše studije nam ne omogućava donošenje preciznijih zaključaka o efektu prirodne selekcije na promenu učestalosti alela. Naime, uzorak u našoj studiji se sastoji od osoba koje su bile predmet ispitivanja srodničkih odnosa, odnosno dokazivanja očinstva, tako da najtipičniji slučaj za takvo testiranje podrazumeva da su izuzeti uzorci od jednog deteta i dvoje roditelja, a kako su zbog dizajna istraživanja isključena deca kao srodnici, zastupljenije su osobe srednje i starije životne dobi. Imajući to u vidu, nije nemoguće da su M1 homozigoti više zastupljeni u mlađem delu romske subpopulacije i da se zbog veće stope smrtnosti manje nalaze u starijem delu subpopulacije. S druge strane, treba imati u vidu i kulturološko-socijalnu specifičnost romske subpopulacije koja se odnosi na visok natalitet i stupanje u reproduktivne odnose u ranijoj životnoj dobi (Radovanović, S., Knežević 2014), tako da su i među ispitanicima koji su podvrgnuti srodničkim analizama, uglavnom mlađe osobe. Daljim i obimnijim populaciono genetičkim istraživanjem među mlađim osobama romske nacionalnosti bi se dobilo više informacija o broju M1 homozigota i jasniji odgovor na pitanje da li se ovaj genotip nalazi pod blagim dejstvom prirodne selekcije. Verovatnija je situacija da je M1 varijanta češća usled genetičkog drifta, koji nije uticao na promenu u odnosima genotipova i da se ovaj alel ne nalazi pod pritiskom prirodne selekcije.

Ne treba zapostaviti ni činjenicu da je asocijacija M1 varijante sa onkološkim oboljenjima i dalje kontroverzna, a sociološki položaj Romske subpopulacije u pogledu zdravstvene zaštite veoma specifičan, što dodatno otežava izvođenje zaključaka o genetičkoj dispoziciji ove subpopulacije prema malignitetima i zahteva dodatno istraživanje (J. Wang et al. 2017). Dodatni razlog za neko buduće izučavanje ove varijante, može se naći i u revalidaciji uticaja epigenetičkih modifikacija na ekspresiju *CYP1A1*, što bi omogućilo preciznije određivanje uticaja same M1 varijante i M1 varijante sa promenom metilacionog statusa (Anttila et al. 2003). No i bez dovoljno razjašnjene uloge M1 varijante u aktivaciji prokancerogenih PAH jedinjenja, produblivanjem naučno-istraživačkog aspekta u subpopulaciji sa visokom učestalošću ove varijante, mogao bi se otvoriti prostor za preventivno delovanje i poboljšanje medicine rada. Ostaje u dobroj veri da će buduća istraživanja napraviti poseban osvrt i na visoku učestalost *CYP1A1* M1 u subpopulaciji Roma u Srbiji.

Druga ispitivana varijanta u genu *CYP1A1*, je M2 na lokusu rs1048943 koja uvodi nesinonimnu izmenu u sedmom egzonu i utiče na katalitičku aktivnost CYP1A1 enzima (Mao et al. 2019). Kao i M1 varijanta, M2 se asocira sa kancerogenezom ali samo na nivou statističke značajnosti, jer je promena u katalitičkoj aktivnosti enzima koja nastaju kao posledica M2 varijante, vrlo male da bi objasnile uzročno-posledičnu vezu preko aktivacije PAH jedinjenja (Zhang et al. 1996). Učestalost M2 varijante u opštoj populaciji Srbije iznosi 0.062 i na osnovu F_{ST} parametra čija je vrednost 0.003 na lokusu rs1048943, može se reći da ne varira značajno između pet testiranih regiona. Kada se učestalost M2 varijante u Srbiji uporedi sa podacima dostupnim za susedne populacije, 0.06 u Hrvatskoj, 0.094 u Mađarskoj i 0.09 u Rumuniji, ne mogu se uočiti veća odstupanja koja se mogu očekivati na osnovu greške u uzorkovanju. U poređenju učestalosti M2 varijante između romske subpopulacije (0.027), sa ostatkom opšte populacije u Srbiji (0.063), ne postoji statistički značajna razlika (P vrednost=0.191). Nemogućnost da se preciznije odredi razlika između učestalosti M2 varijante u romskoj i opštoj populaciji u Srbiji se može pripisati veličini uzorka Roma, što je dovelo do umanjavanja statističke snage testa da ukaže i na manje vrednosti efekat veličine, tako da ostaje nezvesnot da li je u romskoj subpopulaciji učestalost M2 varijante zaista ispod 5% i da li se kao takva može svrstati u retke varijante.

Raspodele učestalosti alela obe genske pozicije u okviru gena *CYP2C9*, rs1799853 (c.430) i rs1057910 (c.1075), između testiranih regiona ili Roma (0.101 i 0.127) i ostatka opšte populacije (0.117 i 0.08) u Srbiji, nisu imale značajnih međusobnih razlika. Minorne razlike u učestalostima

c.430 C varijante između Roma (0.89) i ostatka opšte populacije (0.88) se mogu pripisati slučajnom variranju usled uzorkovanja. Nešto veće razlike u učestalostima varijante A na poziciji c.1075, između Roma (0.87) i ostatka populacije (0.92), takođe nisu statistički značajno odstupile od raspodele koja se očekuje u slučaju važenja nulte hipoteze, jer prema statistici Fišerovog testa tačne verovatnoće, P vrednost iznosi 0.0683. Tu svakako ne treba zaboraviti da je prag značajnosti drastično ispod 0.05 i da 0.0683 ne predstavlja vrednost statistike testa koja je blizu praga statističke značajnosti. Pored očiglednog podudaranja u alelskim učestalostima, vrednosti F_{ST} parametra (niže od 0.005 za oba testirana lokusa) dodatno potvrđuju da je glavnicu varijabilnosti unutar celokupne opšte populacije, a ne između Roma i ostatka opšte populacije. Isto se može reći i za raspodelu alelskih učestalosti između regiona, niske vrednosti F_{ST} parametra (niže od 0.008, Tabela 5) ukazuju na mala odstupanja za koja se neočekuje da se u istom obrazcu održe prilikom replikacije studije (Thomas and Witte 2002). Ipak, nepostojanje genetičke struktuiranosti na lokusima rs1799853 i rs1057910, relativno visoke učestalost c.430 T i c.1075 T varijanti (0.12 i 0.074), kao i poznati uticaj na funkcionalnost metabolizma lekova, opravdavaju genotipizaciju ovih lokusa u svrhu farmakogenetičke primene ili post mortem analize u opštoj populaciji Srbije.

Ovo se naročito odnosi na kliničku primenu varfarina (Johnson et al. 2011), ali pored varfarina, toksikološki interesantan lek, fenitoin je takođe supstrat za CYP2C9, čiji metabolizam zavisi od alela *2 i *3 i zahteva optimizaciju doze prema metaboličkom fenotipu. Za slabe metabolizere, fenitoin se redukuje za 50, odnosno 25% za umerene metabolizere (Caudle et al. 2014). U aktima Agencije za lekove medicinska sredstva Republike Srbije (ALIMS), postoji napomena o uticaju CYP2C9 na metabolizam varfarina, ali ne i uputstvo o doziranju na osnovu genotipa (ALIMS 2012). Takođe, prema podacima ALIMS o neželjenim reakcijama na lek u poslednjih 10 godina, grupa B ATC (anatomsko-terapijsko-hemijske) klasifikacije, koja se odnosi na lekove za krv i krvotvorne organe, je zastupljena sa 6-11% u svim prijavljenim slučajevima neželjenih reakcija na lek (ALIMS 2012). Ostaje otvoreno pitanje koliki procenat ovih slučajeva bi bio umanjeno optimizacijom doze varfarina na osnovu genotipiziranja lokusa CYP2C9. Isto pitanje se može postaviti i kada je u pitanju udeo neželjenih reakcija na fenitoin koji u Srbiji iznosi 6.82%.

Visoke učestalosti CYP2C19 varijanti rs4244285 (c.681G>A, *2) 0.163 i rs12248560 (g.-806C>T, *17) 0.222 u Srbiji poslednično tvore učestalosti homozigotnih stanja koja su u skladu sa očekivanjem na osnovu Hardi-Vajnbergove ravnoteže (*2*2 0.045 i *17*17 0.079), ali i visoke učestalosti heterozigotnih stanja (*2*X × 0.229 i *17*X × 0.3) za koja se smatra da imaju umanjenu funkcionalnost u pogledu metabolizma pojedinih supstanci (klopidogrel, diazepam, inhibitori protonске pumpe) (Li-Wan-Po et al. 2010; Sibbing et al. 2010). Detaljnijim pregledom rezultata genotipizacije u opštoj populaciji (Tabela 2), opažaju se samo tri genotipa gde je homozigotno stanje bilo prisutno na jednom od dva testirana lokusa, što može potvrditi postojanje c.681G>A i g.-806C>T u fazi, odnosno na istom hromozomu. S obzirom da nema publikacija koje potvrđuju postojanje ova dva SNV u haplotipu, ali i visokoj neuspešnosti genotipizacije na lokusu g.-806, najverodostojnije objašnjenje jeste da je u pitanju artefakt. Međutim i u slučaju da su učestalosti oba alela precenjene, a s obzirom na to da do sada nije opisan haplotip sa genetičkim varijantama određenim sa c.681G>A i g.-806C>T, verovatnoća da će nasumično genotipizirani lokus gena za CYP2C19 biti *2 ili *17 haplotip, je jednaka zbiru ove dve verovatnoće, odnosno preko 50%. Ovaj podatak može dovesti u pitanje adekvatnost terapija lekovima za čiji metabolizam je pokazano da zavisi od CYP2C19*2 i *17 u Srbiji, s obzirom da genotipizacija CYP2C19 nije široko rasprostranjena u praksi.

Najveći uticaj na metabolizam lekova i drugih ksenobiotika u opštoj populaciji republike Srbije, svakako ima SNV rs3892097, a sa njim i nefunkcionalni *4 haplotip u genu za CYP2D6 enzim. Razlog za to je učestalost homozigota *4*4 koja iznosi 4.7 %, dok je čak četvrtina ispitanika (25.2%) imala

heterozigotno stanje. Implikacije ove varijante se odnose na široki spektar lekova koji su u kliničkoj primeni, tako se npr. može očekivati da bi bar svaka dvadeseta pacijentkinja kojoj je dijagnostikovano estrogen zavisni rak dojke, neuspešno odgovorila na terapiju tamoksifenom. Taj broj je sasvim izvesno i veći, jer pored *4*4 homozigota i drugi nefunkcionalni aleli u heterozigotnom stanju sa *4 mogu dovesti do izostanka ili značajnog umanjenja metabolizma tamoksifena (Goetz et al. 2018; Schroth et al. 2017).

U poređenju sa susednim populacijama, učestalost *4 alela se ne razlikuje od one koja se očekuje slučajnim variranjem između uzorkovanja (Kuzmanovska et al. 2015; Mizzi et al. 2016; Weber et al. 2015). Ista je situacija kada u poređenju Romske subpopulacije i ostatka opšte populacije. Jedino je evidentan manji broj homozigota *4*4 (1) u Romskoj subpopulaciji u odnosu na ostatak opšte populacije (22), što je opaženo i na lokusu rs4646903 u *CYP1A1* genu (M1 varijanta). Za razliku od rs4646903, genotipizacija svih *CYP2D6* lokusa bi trebalo da dovede i do većeg broja *4*4 homozigota zbog nemogućnosti da se heterozigoti *4*5 razdvoje kao genotipovi. Ove informacije dodatno podupiru mogućnost da je do umanjenja homozigota sa *CYP450* nefunkcionalnim varijantama došlo usled interakcije sa uslovima životne sredine, odnosno selekcionog pritiska preko enzima koji su nosioci metabolizma ksenobiotika. Međutim, veličina uzorka nije dovoljna da bi se uočio mali efekat veličine kojim bi *CYP450* varijante doprinele umanjenju homozigota *4*4. Potrebno je uzeti u obzir da su i ostale učestalosti varijanti na *CYP2D6* lokusu pod velikim uticajem *5 alela, odnosno delecije kompletnog lokusa. Ovo nije biološki fenomen, već isključivo tehničko ograničenje genotipizacije pri kojem se svako heterozigotno stanje, gde je jedan lokus deletiran, opaža kao homozigot za polimorfizam koji je bio prisutan na drugom lokusu (Bijl et al. 2008). Kako su metode za određivanje variranja broja kopija na osnovu RT PCR nepouzdanе i opterećene efikasnošću amplifikacije (Ramakers et al. 2003), teško je proceniti koliki je efekat ovog tehničkog maskiranja *5 alela, ali trebalo bi očekivati da bude jednak proizvodu učestalosti *5 i učestalosti drugog SNV ili alela koji se procenjuje (Naranjo et al. 2016). Ukoliko se uzmu u obzir učestalosti *5 u Evropskim populacijama, koje se kreću do 5-6%, ostale genetičke varijante na lokusu *CYP2D6* ne bi trebalo značajnije da odstupaju od procenjenih u Republici Srbiji (Rajeevan et al. 2017).

Bez obzira na to što nije moguće odrediti tačan broj metaboličkih klasa sa 4 SNV u *CYP2D6* genu, jasno je da je minimalna učestalost loših metabolizera blizu učestalosti *4*4 homozigota, što implicira da je realan broj svakako veći kada se pridruže heterozigotna stanja sa *10 i *41 haplotipovima. Ipak, treba imati u vidu da se ne spoznavanjem učestalosti deletiranog gena *CYP2D6*, pridodaje viši procenat učestalosti ostalih alela, ali ne i fenotipske klase, jer *4*5 genotip ima *4 dominantno ispoljavanje koje se ogleda u značajno smanjenoj mogućnosti metabolizma *CYP2D6* supstrata.

Poslednja u nizu ispitivanih varijanti jeste rs2740574 u *CYP3A4*, koja se nalazi 392 bp uzvodno od mesta početka transkripcije gena. Niska učestalost u opštoj populaciji Srbije (0.022) stavlja ovaj SNV van panela koji bi zadovoljio kriterijume za informativno ciljano testiranje u farmakogenetici ili molekularnoj autopsiji, koji podrazumeva da se zbog ograničenog broja eseja koji se mogu primeniti RT PCR metodom uključiti varijante sa učestalostima većim od 0.05 (Bomba, Walter, and Soranzo 2017). Slične učestalosti su opisane i u susednim populacijama u Bugarskoj i Hrvatskoj (Ganoci et al. 2017; Saraeva et al. 2007), a nisu opažene ni razlike između romske subpopulacije i ostatka opšte populacije. Nažalost, ne postoji ni puno drugih varijanti u ovom genu čiji je uticaj na funkcionalnost *CYP3A4* enzima evidentiran i dokazan (Gaedigk et al. 2018). Da li je nedostatak funkcionalnih varijanti u *CYP3A4* odraz važne funkcije koju ovaj enzim obavlja u metabolizmu ksenobiotika, pa je svaka genetička varijanta koja bi dovela do permanentnog izostanka ili umanjenja funkcionalnosti nevišabilna, pitanje je koje zahteva istraživanja na populaciono genetičkom i

fiziološkom planu. Poznato je da postoje izuzetno retke varijante koje bi bile dobri kandidati za testiranje umesto rs2740574 i za koje se tvrdi da uvode potpunu nefunkcionalnost ovog enzima, ipak, te varijante su u do sada pronalazene u heterozigotnom stanju sa funkcionalnim varijantama, tako da još uvek nije opisan slučaj potpunog izostanka CYP3A4 iz organizma, iako se sugerise da bi takvo stanje bilo neizbežno (Werk et al. 2014; WESTLINDJOHNSON et al. 2006). S obzirom na to da CYP3A4 enzim ima veliki broj supstrata među kojima je i znatan broj lekova koji su u kliničkoj upotrebi, bilo bi neoprezno da, zbog nepostojanja ili nepoznavanja genetičkih varijanti u promotorskoj ili genskoj sekvenci, bude zanemaren i uticaj drugih načina regulacije genske ekspresije na koncentracije aktivne supstance i da se celokupna fenotipska varijabilnost enzima pripise sredinskom uticaju. Postoji još nekoliko nivoa regulacije genske ekspresije koji bi imali sličan efekat koji se pripisuje -392G varijanti. Potrebno je sagledati širu sliku i istražiti metilacioni status promotorske sekvence, postojanje regulatornih elemenata, regulaciju malim interferirajućim RNK, regulaciju na nivou informacione RNK, kao i regulaciju na nivou stabilnosti proteina, jer svaki od prethodno navedenih nivoa regulacije ekspresije CYP3A4 gena može uticati na količinu CYP3A4 enzima. Do sada, neveliki broj studija koje su za predmet istraživanja imale pomenute nivoe regulacije genske ekspresije nisu iznedrile jasne zaključke o pojedinačnom uticaju svakog od nivoa regulacije (Hirota et al. 2004; Kacevska et al. 2012; Vyhldal et al. 2016).

Pored očigledne koristi od genotipizacije CYP450 funkcionalnih varijanti u utvrđivanju uzroka i porekla smrti kada su u pitanju intoksikacije supstancama čija koncentracija u organizmu direktno zavisi od metabolizma zavisnim od CYP enzima, postoji mogućnost i da se određivanjem metabolizma antidepresiva ili drugih lekova koji mogu umanjiti kognitivne sposobnosti osobe pod terapijom, bliže pojašni poreklo smrti. Pored toga, moguće je i preventivno delovati u odnosu na negativne posledice (suicid, saobraćajne nezgode, povrede na radu itd.), imajući u vidu da se radi o naslednim varijantama i da bi bliski srodnici osobe kod koje je identifikovana varijanta asocirana sa potencijalnom negativnom posledicom, mogli biti na isti način zdravstveno ugroženi.

Molekularna autopsija je posebno korisna u slučajevima kada nije moguće odrediti količinu toksične supstance u krvi ili urinu zbog poodmaklih truležnih promena tela koje je predmet obdukcije (van Driel, Michels, and van der Velden 2018). Kako se genetička informacija u DNK održava dugo nakon smrti, moguće je rekonstruisati metabolička svojstva preminule osobe na osnovu genotipizacije farmakogenetički bitnih lokusa. To bi bilo korisno u slučaju inkonkluzivnih nalaza obdukcije, naročito toksikološkog nalaza koji se ne uklapa u širu sliku porekla i uzroka smrti. Za tu svrhu, odabir funkcionalnih varijanti *CYP* gena bi se mogo primeniti u skladu sa drugim indicijama i saznanjima do kojih se može doći pre i nakon obdukcije. Tako bi, s obzirom na učestalosti u Srbiji, genotipizacija *CYP2C9* varijanti *2 i *3 bila korisna u rasvetljavanju kliničkih obdukcija kada je upokojeni bio podvrgnut lečenju varfarinom, ukoliko je toksikološki nalaz ukazao na toksične doze ovog leka (Dean 2012b).

Nešto je drugačija situacija sa rekonstruisanjem metabolizma zasnovanim na genotipizaciji lokusa kojima se određuju *CYP2C19* *2 i *17 aleli, jer i pored značajno visokih učestalosti u opštoj populaciji Srbije (0.163 i 0.222), najjasniji uticaj ovih izoenzima jeste na antitrombocit klopidogrel (K. Kim et al. 2008). Međutim, klopidogrel ima više uticaja na trombotičke događaje i komplikacije u organizmu u kojem već postoji narušena hemostaza krvotoka, nego što može ispoljiti toksične efekte kao što je slučaj sa varfarinom. Prema evidentiranim slučajevima predoziranja klopidogrelom, samoubistvo ovim lekom nije moguće ni dozom od 4200 mg i kao posledicu može imati samo manje devijacije u koagulaciji krvi (Kocabay et al. 2006; Mehta et al. 2010; Yilmaz et al. 2013). S toga, post

mortem analiza koja uključuje genotipizaciju varijanti *CYP2C19* bi više podataka pružila u cilju rasvetljavanja okolnosti pod kojim je smrt nastupila, nego samog uzroka smrti.

Kada su u pitanju funkcionalne varijante *CYP2D6*, opravdanost genotipizacije u svrhu post mortem analize nije jednaka opravdanosti primene u personalizovanoj medicini. Bez obzira na to što se *CYP2D6* varijanta *4 (rs3892097), koju karakteriše izostanak svake funkcije ovog enzima, često može naći u homozigotnom (0.047) ili heterozigotnom stanju (0.225) u opštoj populaciji Srbije, najveći efekat na promene u metabolizmu potencijalno smrtonosnih supstanci, imaju umnožene kopije funkcionalnih alela koje nisu bile predmet naše studije. Kada je u pitanju poreklo smrti, neobični slučajevi intoksikacije kodeinom, dihidroksikodeinom i oksikodonom bi bili dobri kandidati za redovnu proveru metaboličkog statusa upokojenih na osnovu broja kopija funkcionalnih *CYP2D6* alela (Dean 2012a; Haufroid and Hantson 2015; Jannetto et al. 2002). Ipak, treba istaći da je umanjena sposobnost za metabolizam oksikodona usled postojanja nefunkcionalnih formi *CYP2D6* enzima bila predložena kao pomoćna informacija u toksikološkoj interpretaciji idiosinkratničnih slučajeva intoksikacije oksikodonom i da bi širi pogled na epistatičko dejstvo *CYP2D6* alela sa varijantama drugih gena koji su involvirani u metabolizam ovog analgetika, mogao naći svoju primenu u 'molekularnoj autopsiji' (Jannetto et al. 2002). Uprkos predlozima za detaljnije ispitivanje metaboličkog statusa zasnovanog na *CYP2D6* genotipizaciji, kada se oceni doprinos, ali i ekonomska strana farmakogenetičke primene u postmortem analizama, nije isplativo ovaj lokus analizirati ciljanom genotipizacijom, osim u slučajevima koji zahteva izuzetno detaljno upoznavanje porekla smrti (Ehmann et al. 2015).

Pored testiranja samih učestalosti alela i genotipova *CYP450* gena, poseban osvrt u ovoj studiji je napravljen ka analizi genetičke struktuiranosti opšte populacije u Srbiji. Uverenje o stepenu izraženosti genetičke struktuiranosti, može se svesti na ocenu statističkih dokaza o odstupanjima od uniformnosti raspodele alela sa 15 mikrosatelitskih i 11 bialelskih lokusa, koji su upotrebljeni za genotipizaciju ispitanika u ovoj studiji. Izbor ispitanika koji nisu u srodstvu i za koje ne postoji posebno biološko stanje koje bi uvelo sistematsku grešku u uzorkovanju, kao što je npr. zdravstveni problem ili lokalna dostupnost uzoraka, omogućava reprezentativnost uzorka u odnosu na opštu populaciju. Jedna od bitnih karakteristika uzorkovanja jeste da je uzet u obzir uticaj geografske raspodele gustine populacije uvođenjem stratuma, odnosno randomizovanog uzorkovanja na osnovu regionalne klasifikacije i proporcionalno broju stanovnika unutar regiona. Takav način uzorkovanja je omogućio da se genetička varijabilnost opšte populacije ne nađe pod uticajem varijabilnosti lokalnih struktura. Uvođenjem uzorka iz romske subpopulacije, omogućena je provera metoda za detekciju diferencijacije subpopulacija, uz pretpostavku da bi upravo Romi mogli činiti genetički najdiferenciraniji deo opšte populacije zbog socijalne izolacije i demografskog porekla (Mendizabal et al. 2012). No, pre sagledavanja rezultata analize genetičke strukture, treba razmotriti i ograničenje koje se odnosi na razlike između testiranih lokusa. Razlog za to je taj što nisu svih 11 testiranih bialelskih lokusa međusobno u stanju gametske ravnoteže, već su varijante, koje se nalaze unutar haplotipova *CYP* gena *1A1*, *2C9*, *2C19*, *2D6* i *3A4*, vrlo blisko pozicionirane i vezano se nasleđuju. Imajući u vidu da je odnos *CYP450* i STR lokusa je takav da se na metapopulacionom nivou opaža podudarnosti između raspodela genetičke varijabilnosti superfamilije *CYP450* i većeg broja mikrosatelitskih lokusa (Rosenberg et al. 2005), treba razmotriti i mogućnost podudaranja varijabilnosti STR i *CYP* lokusa koji su bili predmet naše studije. No, u situacijama kada se koristi manji broj genetičkih markera koji su gametskoj ravnoteži sa genima koji su odgovorni za opaženi fenotip ili fiziološko stanje, postojanje genetičke strukture neće obavezno uticati na promenu učestalosti genskih varijanti. Za našu studiju bi to značilo da postojanje klastera definisanog na

osnovu STR podataka ne mora nužno za posledicu imati postojanje klastera na osnovu CYP varijanti, što se podudara sa opaženim rezultatima u slučajevima kada su klasteri identifikovani. Većina modernih studija koristi veliki broj markera koji pokrivaju sve delove genoma i mogu postići bolju rezoluciju kako bi otkrili i suptilnije promene, odnosno slabo izraženu genetičku struktuiranost (Welter et al. 2014). U našoj studiji je za detekciju genetičke struktuiranosti upotrebljen set od svega 15 mikrosatelitski lokusa, koji može ukazati na veće fluktuacije u alelskim učestalostima, ali ne i na finije obrazce odstupanja od uniformne raspodele alelskih učestalosti. Jedna od prednosti STR markera u analizama srodničkih odnosa jeste njihova neutralnost i relativno slična učestalost alela između populacija (Wilkening et al. 2006). Istovremeno, takva prednost ujedno je i mana kada je u pitanju detekcija genetičke struktuiranosti populacije, jer nizak nivo gametske neravnoteže između STR markera ukazuje na njihovu fizičku udaljenost u genomu, što ostavlja mogućnost da se druge varijante u međuprostoru nalaze u gametskoj neravnoteži, odnosno, potpuno nezavisnog nasleđivanja u odnosu na testirane STR lokuse. Zato je odabir genetičkih markera za testiranje strukturiranosti populacije od ključne važnosti. Postoje predlozi da se za analizu genetičke strukture pomoću STR lokusa prethodno proceni mera informativnosti lokusa I_n , koja poredi divergenciju po lokusu po populaciji, nakon čega treba odabrati visoko diferencirane genetičke varijante za analizu struktuiranosti populacije (Rosenberg et al. 2003). Međutim, kako je u našoj studiji upotrebljen već definisani set mikrosatelitskih lokusa za forenzičku upotrebu, prethodna procena I_n nije bila moguća, već je više pažnje posvećeno primeni robusnih algoritama za analizu strukture podataka (Corander, Waldmann, and Sillanpää 2003; Jones and Wang 2012; W. Li et al. 2017; Porrás-Hurtado et al. 2013a; Pritchard, Stephens, and Donnelly 2000; Salas, Phillips, and Carracedo 2006).

Prvi korak u testiranju struktuiranosti populacije se odnosi na analizu molekulske varijanse i izračunavanju F_{ST} parametra između unapred definisanih subpopulacija (S Wright 1950) i prema pretpostavci da bi geografska udaljenost mogla dovesti do promena u učestalostima alela i genotipova. Ipak, podesnost F_{ST} parametra za opisivanje struktuiranosti populacije se dovodi u pitanje zbog velikog uticaja diverziteta unutar subpopulacije (H_S). Ovo se naročito odnosi na mikrosatelitske lokuse, gde hipervarijabilni lokusi imaju manje stepeni slobode zbog čega umanjuju vrednost F_{ST} i prikrivaju razlike između subpopulacija (JOST 2008; MEIRMANS and HEDRICK 2011). Kako vrednost F_{ST} parametra ne može premašiti prosečnu homozigotnost subpopulacija

$$F_{ST} = 1 - \frac{H_S}{H_T} < 1 - H_S$$

, postavlja se pitanje njegove primene u odmeravanju diferenciranosti subpopulacija u situaciji kada se porede lokusi sa različitom varijabilnošću (Charlesworth 1998; Hedrick 1999, 2005; Jakobsson, Edge, and Rosenberg 2013). Taj problem ne bi trebalo da postoji kada su u pitanju bialelski SNV i više se tiče STR lokusa koji imaju veći broj alela. S obzirom da su varijante u *CYP450* genima bialelske, za očekivati je da F_{ST} parametar nije pod uticajem visoke varijabilnosti lokusa, što procenjenju vrednost čini pouzdanijom. Rezultati F statistike zasnovani na raspodeli varijabilnosti testiranih *CYP450* lokusa pokazuju da F_{ST} parameter ima srednju vrednosti od 0.0047 kada se analizira varijabilnost između geografskih regiona (Tabela 8). Dobijene vrednosti F_{ST} na osnovu geografske raspodele, takođe ukazuju da je najveća razlika u heterozigotnosti između testiranih regiona, na rs4646903 lokusu, *CYP11A1* gena (M1 varijanta). Zanimljivo je da geografski region sa najvećim odstupanjem od očekivane heterozigotnosti jeste područje Beograda, ali i da je i F_S na lokusu rs12248560 g.-806C>T imao najvišu vrednost (32%) upravo za subpopulaciju ovog regiona. Ovakav rezultat bi se mogao objasniti većom admiksijom uzrokovanom skorijim

migracijama iz drugih regiona ka području Beograda, pri čemu je došlo do očuvanja homozigotnih genotipova čiji bi broj u budućnosti trebalo da se umanjuje usled protoka gena (Campbell et al. 2009).

Vrednosti F_{ST} parametra za dve potencijalne subpopulacije u Srbiji na lokusima *CYP450* gena, koje bi se mogle zasnivati na etničkoj pripadnosti, ne ukazuju na postojanje značajne genetičke diferencijacije (Tabela 9). Na svega jednom lokusu (rs4646903, M1 varijanta u *CYP11A1* genu) od 11 testiranih se može opaziti blago povišena vrednost F_{ST} od 0.012, dok je na svim ostalim lokusima ta razlika zanemarljiva tako da prosečna vrednost F_{ST} iznosi 0.0023.

Drugačiji je pristup u analizi genetičke struktuiranosti na osnovu bialelskih varijanti, jer je bilo neophodno rekonstruisati haplotipove koji se podudaraju sa alelima CYP gena, s obzirom na to da su varijante u gametskoj neravnoteži i da se zanemarivanjem ove činjenice može uvesti greška koja za posledicu ima veći broj sličnih F_{ST} vrednosti za svaki od međusobno vezanih lokusa. To je učinjeno pomoću Haploview i F_{ST} parametara procenjenog analizom varijanse u Arlequin softveru (L Excoffier, Smouse, and Quattro 1992). Ipak, dobijeni rezultati ovog pristupa se nisu bitnije razlikovali od rezultata F statistike bazirane na pojedinačnim SNV (prosečni F_{ST} =0.0021, Tabela 10), što je i očekivano s obzirom na to da je gametska neravnoteža između SNV u jednom genu izuzetno visoka zbog bliske pozicije i nepostojanja rekombinativnog mesta između tih lokusa. No, to ne predstavlja jasnu potvrdu nulte hipoteze, odnosno nepostojanja faktora koji utiče na raspodelu alelskih učestalosti, kao što bi to mogla biti genetička struktura, već jedan dokaz više u odbacivanju alternativne hipoteze. Samo uvođenje haplotipova ima posledice na statističke modele koji su primenjeni u ocenjivanju genetičke strukture. Glavna posledica jeste umanjeno statističko snage testova zbog povećanja varijabilnosti nakon prekodiranja bialelskih u multialelske polimorfizme, te umanjeno mogućnosti da se otkrije odstupanje od uniformne raspodele alelskih učestalosti testiranih *CYP* gena (Porrás-Hurtado et al. 2013a; W. Shi et al. 2010).

U našoj studiji je napravljen i poseban osvrt na latentnu genetičku struktuiranost opšte populacije u Srbiji. Za razliku od klasičnog pristupa u kojem se pomoću analize varijanse (F statistika) poredi heterozigotnost između unapred definisanih grupa, dodatno je isproban nekoliko različitih pristupa u detekciji neuniformnosti raspodele alelskih učestalosti. Na osnovu podataka obrađenih različitim algoritmima, uverenje o postojanju ili naznakama genetičke struktuiranosti u Srbiji se može načelno steći pregledom prednosti i mana svakog od primenjenih algoritama.

Najupečatljiviji rezultati klaster analiza su generisani BAPS algoritmom (slika 50). Dok se na osnovu mikrosatelitskih podataka u opštoj populaciji nisu mogli uočiti izdvojeni klasteri, u etnički kategorisanim podacima (Romi i ostatak opšte populacije) su pronađena dva klastera. Iako je jedan od klastera manji po broju genotipova koji su se našli u njemu (187), broj genotipova Roma je iznosio 46. U drugom klasteru, u kojem se našla većina genotipova (297), udeo genotipova Roma je iznosio 11. Kao izdvojeni iz oba klastera, našla su se dva genotipa Roma. Ovaj rezultat je neobičan zbog verovatnoće da se grupa genotipova od 57 osoba, na slučajnan način raspodeli u dva klastera veličina 187 i 297 članova. Binarnom formulom se može izračunati verovatnoća za takav ishod i ona je manja od 0.00001. Ovi rezultati mogu ukazati na heterogenost opšte populacije u Srbiji i mali uticaj geografske udaljenosti na genetičku strukturu, ali i većem uticaju kulturoloških faktora, koji jednu subpopulaciju dovode u stanje veće izolovanosti. odnosno smanjenog protoka gena i višeg stepena inbridinga.

Međutim, potvrda klaster analiza pomoću BAPS algoritma se nije mogla naći u rezultatima najprominentnijeg algoritma za detekciju genetičke struktuiranosti, STRUCTURE (Pritchard, Stephens, and Donnelly 2000). Zbog heruističke prirode ovih algoritama, nije moguće opredeliti se za veću upotrebnu vrednost jednog od njih, već samo ukazati na bitnost premisa na kojima počivanju. Pored problema neuporedivosti algoritama, važno je istaći i premise u samoj osnovi istraživanja genetičke strukture u opštoj populaciji Srbije. Naime, polazna osnova za izdvajanje Roma kao posebne subpopulacije ima populaciono genetičkog utemeljenja ukoliko se uzme u obzir demografsko poreklo i socijalna izolovanost, ali ima i elemente pretpostavke bez prethodnih dokaza da su Romi genetički diferencirani od ostatka opšte populacije. Ovakva premissa može za posledicu imati artefakte, naročito ako se u obzir uzme broj algoritama koji su upotrebljeni da bi se uočila razlika, što sa sobom nosi problem višestrukih statističkih testiranja. Ipak, vrednost verovatnoće da do slučajnog razdvajanja 46 od 59 Roma u klaster veličine 187 članova i svega 11 u klaster veličine 297 članova, je veoma značajna i drastično ispod korigovanog praga značajnosti za višestruka poređenja (P vrednost < 0.001).

Proces u inferenciji broja subpopulacija u STRUCTURE algoritmu se mogu podeliti na dva podprocesa. Prvi podproces u inferenciji jeste 'burn-in' period, koji je u ovoj studiji iznosio 100 000 iteracija, dovoljan za konvergenciju prema pouzdanim procenama alelskih učestalosti unutar subpopulacija. Drugi deo je određivanje porekla svakog genotipa preko funkcije verodostojnosti, za koju je predloženo da iznosi 200 000 iteracija u podrazumevanom (eng *Default*) režimu (Porrás-Hurtado et al. 2013a). Konfiguracija podmodela (admiksični model, međusobno vezani aleli, bez uključenog *LOCPRIOR* paramtera) koja je primenjena u ovoj studiji bi se mogla smatrati optimalnom i konzervativnom u smislu ne uključivanja malo verovatnih pretpostavki ili varijabli koje utiču na izračunavanje najverodostojnijeg modela (Daniel J. Lawson, van Dorp, and Falush 2018). Tako na primer, za vrednosti predloženog broja populacija, preporuka je da se uzima najniža vrednost sa maksimumom funkcije verodostojnosti (Kalinowski 2011). Imajući u vidu da STRUCTURE dozvoljava i opciju *LOCPRIOR*, odnosno unošenje prethodno poznatih podataka o poreklu genotipova koje čine uzorak, a takvi podaci se najčešće odnose na fenotipske karakteristike (npr. visina), lingvističke (dijalekt), socijalno-ekonomske kategorije (obrazovanje, socijalni status) itd., treba biti obazriv prema rezultatima koji bi na taj način bili generisani. Ovi podaci mogu pomoći kada postoji slaba genetička struktuiranost, ali otvaraju pitanje nastajanja artefakata. Naime, kako ne postoji zlatni standard naspram kojeg se može porediti učinkovitost nekog algoritma i modela, veoma je lako formirati model koji se ne podudara sa realnošću i to baš pridruživanjem drugih varijabli kao što su gorenabrojane (Kalinowski 2011; Daniel J. Lawson, van Dorp, and Falush 2018).

S druge strane, treba imati u vidu i to da je STRUCTURE algoritam deo naivne Bejzove klasifikacije, i kao takav pretpostavlja međusobnu nezavisnost nekih od ulaznih varijabli kao što su selekciona neutralnost genetičkih markera, niska stopa mutacije, ali i gametska ravnoteža (Pritchard, Stephens, and Donnelly 2000), što sve čini ovaj algoritam u mnogome zavisim od HW ekvilibrijuma (Kalinowski 2011). Interesanto je i to da se u STRUCTURE uvodi matrica genetičke udaljenosti po Nei, što je u ovom radu bila polazna osnova za DBSCAN (kao i za PCA i tSNE), te na toj instanci ne treba tražiti objašnjenje za razlike u dobijenim rezultatima.

U STRUCTURE se, kao i u BAPS, algoritmu, broj predloženih subpopulacija određuje poređenjem verodostojnosti, odnosno pronalaženjem najniže vrednosti negativnog logaritma funkcije verodostojnosti. Takav metod se naziva delta K i prema rezultatima na osnovu genotipiziranih 15 STR lokusa u uzorku zasnovanom na regionalnoj klasifikaciji, najoptimalniji broj subpopulacija bi iznosio 3 (Slika 45) (EVANNO, REGNAUT, and GOUDET 2005). Međutim, treba imati u vidu da su razlike između vrednosti funkcija verodostojnosti za svaki od testiranih K, zapravo veoma male i

da ne mogu sa većim stepenom sigurnosti ukazati na struktuiranost i postojanje četiri subpopulacije. Ako se pažnja posveti grafičkom prikazu rezultata STRUCTURE algoritma (slika 42), koji daje intuitivniji opis populacije prikazivanjem svakog pojedinačnog genotipa i njegovog porekla iz pretpostavljenih populacija, može se uočiti da je gotovo svaki genotip, odnosno, svaka osoba iz predefinisanih regiona imala sličan obrazac porekla.

Ostaje pitanje zašto se rezultati BAPS algoritma razlikuju u odnosu na one dobijene pomoću STRUCTURE, a odgovor najverovatnije treba tražiti u razlikama ova dva algoritma. Dok se u osnovi iteracija STRUCTURE softvera nalazi MCMC pristup, u osnovi BAPS je nereverzibilni MCMC algoritam koji znatno ubrzava process izračunavanja alelskih učestalosti tokom 'burn-in' perioda, jer ne pravi korake unazad tokom 'hodanja' po raspodeli. Pored toga, BAPS uvodi ciljano 'stohastičko' dodeljivanje jedinki u prostorne klastere, što svakako umanjuje vreme izračunavanja, ali i dovodi do drugačijeg rasporeda u klastere (Corander, Sirén, and Arjas 2008). To su, naravno, tehničke razlike. Suštinska razlika se odnosi na pretpostavku o broju populacija. Dok STRUCTURE polazi od zadate vrednosti, BAPS tretira broj populacija kao nepoznati parametar koji se može inferirati iz analiziranih podataka (Rodríguez-Ramilo, Toro, and Fernández 2009). To bi značilo da STRUCTURE uklapa posteriornu raspodelu na osnovu broja zadatih populacija, dok BAPS procenjuje broj populacija, a zatim i verodostojnost tog parametra na osnovu analize genetičkih podataka. Još jedna interesantna činjenica je da BAPS uopšte ne uzima u obzir 'stepwise' model mutacije mikrosatelita za koji bi se očekivalo da bolje aproksimira verovatnoće promene alelskih učestalosti pod uticajem mutacija na mikrosatelitskim lokusima (Kimura and Ohta 1978). No, efekat neuključivanja 'stepwise' modela u Bejzov pristup statističke inferencije nije poznat, niti dokumentovan u literaturi. Interesantno je i to da, iako i BAPS i STRUCTURE koriste beta i gama raspodelu za modelovanje bialelskih i multialelskih genetičkih podataka u maksimizaciji verodostojnosti, ostaje nepoznato koliki bi uticaj imala implementacija beta-gama raspodele, kao vrlo prilagodljive na različite tipove podataka i već upotrebljene na biološkim podacima (uključuje beta-eksponencijalnu, gama i eksponencijalnu raspodelu) (Kong, Lee, and Sepanski 2007).

Primenom najčešće korišćenog metoda u detekciji genetičke strukturiranosti populacije, analize glavnih komponenti, rezultati nisu ukazali na postojanje klastera u podacima genotipizacije mikrosatelita na osnovu regionalne, niti etničke raspodele. Ovaj podatak se može tumačiti na dva načina, da struktuiranost populacije ne postoji, ili da je PCA nepodoban za analizu nelinearnih podataka kakvi su genetički podaci prekodirani u diskretne brožčane vrednosti. Čak i u slučajevima kada postoji jasna genetička struktura u nekoj populaciji, PCA je izuzetno osetljiv na način uzorkovanja, jer nesrazmernim obuhvatanjem varijabilnosti u lokalnim strukturama (npr. veća zastupljenost uzoraka sa jednog područja), narušavaju se ajgenvrednosti koje određuju glavne komponente analize. U ovoj studiji je pristupom uzorkovanja proporcionalnog broju stanovnika u diskretnim regionima, uticaj prezastupljenosti umanjen u značajnoj meri, ali ne i u potpunosti, s obzirom na to da diskretne populacije na prostoru Republike Srbije ne postoje. Najidealniji uzorak za PCA bi bio onaj koji prati gustinu stanovništva u prostoru, kao neprekidno promenjivu, no kako takav uzorak nije dostupan u biblioteci bioloških uzoraka, već se mora prikupljati na terenu, to čini upotrebu PCA za otkrivanje klastera (subpopulacija), neadekvatnom metodom. Ipak, kada se bolje analiziraju grafici prve dve komponente PCA podataka na osnovu etničke raspodele (Slika 38), može se uočiti drugačiji obrazac raspršenosti genotipova Roma i ostatka opšte populacije. Epicentar tačaka koje predstavljaju genotipove Roma je pomeren ka višim vrednostima PC1 u odnosu na epicentar tačaka ostatka opšte populacije. Može se primetiti i to da su ekstremi sa najvišim vrednostima PC1, takođe genotipovi osoba romskog etniciteta.

Kako je to već pomenuto, jedan od poznatih problema u primeni PCA jeste uticaj varijabilnosti lokalnih struktura, jer se prezastupljenošću uzorka genotipovima jedne grupe, pravi distorzija u ukupnoj varijabilnosti uzorka, što za posledicu ima umanjeње varijabilnosti u ostatku uzorka (Novembre et al. 2008; L. Shi 1997). Upravo je ovaj problem prevaziđen tSNE algoritmom jer je u njegovoj osnovi ideja da se zadržavanjem varijabilnosti lokalnih struktura, ne narušava varijabilnost globalne strukture podataka (u ovom slučaju, genetičkih)(Van Der Maaten and Hinton 2008).

Rezultati naše studije pokazuju da postoji podudarnost između klastera u rezultatima BAPS i tSNE: 127 (52%) osoba iz K1 BAPS se nalazilo i u K2 tSNE. Međutim, došlo je do preraspodele Roma unutar klastera u kojem se nalaze u većem broju (tabela 15). Broj Roma u K1 BAPS (46) i K2 tSNE (35) ima 30 identičnih genoma dok je ostalih 5 Roma iz K2 tSNE bilo svrstano u K2 BAPS. Takođe je zanimljivo to da je K2 tSNE manje gustine u odnosu na K1 tSNE, što bi moglo ukazati na veću varijabilnost unutar ovog klastera, ali i lokalne odnose genoma koji sačinjavaju ovaj klaster. Ova observacije jeste u skladu sa osnovnim prednostima tSNE algoritma, očuvanju lokalnih i ublažavanju uticaja globalnih odnosa (W. Li et al. 2017).

Tabela 15 – Prikaz preseka skupova klastera generisanih BAPS i tSNE algoritmima

BAPS		tSNE		presek BAPS tSNE			
K1	K2	K1	K2	K1 K1	K1 K2	K2 K1	K2 K2
187	297	161	230	61	96	101	135
Romi	Romi	Romi	Romi	Romi	Romi	Romi	Romi
46	12	12	35	7	30	1	7
Ostali	Ostali	Ostali	Ostali	Ostali	Ostali	Ostali	Ostali
				54	66	100	128

Upotrebom algoritma za hijerarhijsko klasterovanje zasnovano na gustini podataka, DBSCAN (Ester et al. 1996), na genetičkim podacima, prvo se može primetiti izostanak većeg broj parametara koje je potrebno modifikovati, što ukazuje na jednostavnost u primeni. Rezultati ove studije ukazuju na postojanje jednog klastera, kako u uzorku na osnovu regionalne klasifikacije, tako i na osnovu etničkog kriterijuma (Romi naspram ostatka opšte populacije). Na osnovu geografske klasifikacije, nije jasan geografski uticaj na formiranje klastera, odnosno ne postoji podudarnost regiona porekla sa pripadnošću klasteru ili šumu. Međutim, kada je u pitanju etnički kriterijum (Slika 41), neobična je raspodela Roma između klastera i tačaka koje se smatraju šumom. Mala verovatnoća (P vrednost < 0.001) da do ovakve raspodele dođe na slučajan način (18 Roma u klasteru od 274 tačke i 40 Roma u šumu od 207 tačaka), budi podozrenje da bi DBSCAN mogao ukazati na jednu centralnu tendenciju unutar populacije, odnosno jednu grupu sličnijih genotipova.

Broj podudarnih genotipova Roma iz šuma DBSCAN sa Romima iz BAPS K1 (presek šuma i K1) je 34. Odnosno, 34 identičnih osoba romske etničke pripadnosti se našla izdvojena i u analizi BAPS i DBSCAN algoritmom. Ovo je iznenađujuće s obzirom da ukupan presek između šuma u rezultatima DBSCAN i K1 u BAPS-u, iznosi 95 osoba. Dakle, od 207 osoba u šumu DBSCAN, 90 se podudarilo sa 187 osoba u K1 BAPS algoritma, oba na osnovu etničke pripadnosti. To ukazuje da DBSCAN nije sve genome svrstao u šum koje je BAPS svrstao u K1, ali genome Roma jeste preferencijalno svrstavao u istu kategoriju.

Implementacija teorije mreža kao osnove za otkrivanje genetičke strukturiranosti je relativno nova ideja, te se u literaturi može pronaći tek jedan rad i NetStruct softver za tu svrhu (Greenbaum, Templeton, and Bar-David 2016). Uprkos činjenici da je pomenuta implementacija dostupna, postoji značajna manjkavost u matematičkom proračunu genetičke udaljenosti između genoma prilikom

izračunavanja matrice udaljenosti. Izračunavanje genetičke udaljenosti na jednom lokusu je formulisano izrazom:

$$S_{ij,1} = \frac{1}{4}[(1 - f_a)(I_{ac} + I_{ad}) + (1 - f_b)(I_{bc} + I_{bd})]$$

, gde su na lokusu l aleli a i b od genoma i , sa učestalostima f_a i f_b , dok genom j na istom lokusu ima alele c i d . Parametar I_{ac} ima vrednost 1 kada su a i c identični, u suprotnom zadobija vrednost 0. Genetička sličnost između genoma se izračunava kao aritmetička sredina sume sličnosti za sve lokuse:

$$A_{ij} = \frac{1}{L} \sum_{l=1}^L S_{ij,1}$$

Problem nastaje prilikom unosa vrednosti učestalosti alela, jer krajnji rezultat genetičke sličnosti menja vrednosti u zavisnosti od redosleda unosa učestalosti alela. Ukoliko je prvo uneta učestalost alela a , zatim alela b , izračunata sličnost neće biti jednaka onoj kada je redosled bio b , pa zatim a .

Zbog gore navedenog problema, u ovoj studiji je odabrana Neijeva distanca za izračunavanje genetičke sličnosti i konstrukciju matrice međusobne udaljenosti, koja se vrlo robustna u odnosu na tip polimorfizma, dok je R biblioteka 'igraph' upotrebljena za primenu GN algoritma (Csardi and Nepusz 2006; Masatoshi Nei 1974). Treba istaći da vreme koje je potrebno za izračunavanje broja klastera drastično nadmašuje sve ostale algoritme, tako da je dvojezrgarnom procesoru Intel Core 2 duo E7500 2,93GHz bilo potrebno 20.43h za uzorak od 448 genoma sa podacima sa 15 lokusa. Rezultati analize su ukazali na postojanje 7 klastera u obe grupe podataka, geografskoj i etničkoj (Slike 45 i 46). Međutim, tih 7 lokusa po svome sastavu se nisu podudarili ni sa jednim od klastera koje su izdvojeni pomoću drugih testiranih algoritama. Takođe, rezultati zasnovani na regionalnoj klasifikaciji se nisu podudarili sa poreklom uzorkovanja (P vrednost=0.5497), dok rezultati zasnovani na etničkoj pripadnosti nisu svrstali Rome na način koji statistički značajno razlikuje od slučajnog raspoređivanja u 7 kategorija (P vrednost=0.7325). Potrebno je istaći i da obrada podataka koja zahteva upotrebu većeg broja programskih biblioteka, istovremeno zahteva i više vremena za upoznavanje sintakse i razumevanja primene programskog koda, što za posledicu može imati veću verovatnoću nastanka greške pri podešavanju parametara.

Biološki smisao primene GN algoritma u analizi genetičkih podataka se oslanja na analogiju između mreže u kojoj su pojedina čvorišta opterećenija od drugih sa populacijom u kojoj neke jedinke dele više genetičke sličnosti sa drugim jedinkama (Greenbaum, Templeton, and Bar-David 2016). Ipak, sa aspekta populacione genetike je bitno i kvantitativno izraziti odnose unutar populacije, tako da bi opterećenje grana između čvorišta unutar mreže trebalo prevesti u verovatnoću deljenja ancestralnog alela. Trenutna implementacija GN algoritma ne pruža takvu mogućnost, ali osim nepostojanja matematičkog modela koji bi to omogućio, ne postoje drugi razlozi za izostanak te mogućnosti. U ovoj studiji je GN algoritam mogao samo ukazati na postojanje strukture ali bez mere značajnosti strukture, što nije slučaj sa algoritmima zasnovanim na Bejzovoj statistici. Uprkos tome, GN algoritam se može se koristiti kao polazna osnova za definisanje klastera na kojima bi se potom primenila analiza varijanse (F statistika).

U osvrtu na rezultate algoritama za detekciju genetičke struktuiranosti, treba razdvojiti dva izvora nastanka ove pojave u nekoj populaciji. Jedan je admiksijom bar dve populacije sa različitim alelskim učestalostima, drugi je izolovanošću lokalnih subpopulacija kada usled inbridinga i genetičkog drifta dolazi do porasta ili smanjenja alelskih učestalosti. Dok se prvi uzrok može posmatrati i sistematskom greškom u uzorkovanju (pridruženom varijablom), drugi uzrok je mnogo suptilniji i teži za detekciju (Devlin, Roeder, and Bacanu 2001). Prema rezultatima algoritama za otkrivanje genetičke

struktuiranosti, u Srbiji se može opaziti izvesna i mala razlika između subpopulacije Roma i ostatka opšte populacije, što odgovara admiksiji bar dve populacije i prethodnom uverenju o demografskom poreklu Roma (Mendizabal et al. 2012). Ovu razliku je potrebno kritički sagledati, jer mali broj STR lokusa, odnosno njihova informativnost, može dosta uticati na rezultate BAPS, DBSCAN i tSNE algoritama, koji su jedini ukazali na odstupanja od slučajne raspodele. Drugi faktor koji je mogao dovesti do odstupanja od slučajne raspodele romskih genotipova između klastera, jeste broj Roma koji su bili uključeni u testiranje etničkih razlika, ali i izostanak informacije o geografskoj raspodeli alelskih učestalosti unutar romske subpopulacije, odnosno nemogućnosti da se svi testirani uzorci Roma priključe nekome od 5 testiranih regiona, zbog nepostojanja informacije o mestu prebivališta ispitanika. Ono što jeste očito, kada je reč o postojanju lokalnih subpopulacija pod dejstvom izolovanosti, odnosno većeg stepena inbridinga, ne postoji dovoljno dokaza koji bi ukazali na postojanje jasno izražene strukture u opštoj populaciji kada su u pitanju STR lokusi, dok se u romskoj populaciji alelske učestalosti *CYP450 2C9*, *2C19*, *2D6* i *3A4*, sa izuzetkom *CYP1A1 M1*, ne razlikuju u značajnoj meri u odnosu na učestalosti u opštoj populaciji.

Mogućnost postojanja neuniformne raspodele alelskih učestalosti *CYP450* se prvenstveno tiče studija genetičke asocijacije i odabira onih lokusa koji neće drastično varirati između subpopulacije, kako god one bile definisane (geografski, etnički itd.), za farmakogenetičku upotrebu. Poznati primeri lažnih asocijacija genetičkih varijanti sa bolestima, ili fenotipom uopšte, su dokumentovani u dosta slučajeva, pa i onih koji se tiču asocijacije varijanti *CYP* gena sa malignitetima (Kittles et al. 2002; Thomas and Witte 2002). Na sličan način je moguće lažno asociirati i genetičke varijante *CYP* gena sa metaboličkim fenotipom. Ukoliko bi genetička struktuiranost ostala latentna, pojava kao što je neželjena reakcija na lek se može asociirati sa nekom od varijanti *CYP* gena. Naravno, greška bi mogla biti napravljena i u suprotnom smeru, moguće je uprosečenu učestalost neželjene reakcije na lek imputirati subpopulaciji u kojoj su neželjene reakcije na lek učestalije.

Međutim, moguće je zamisliti subpopulaciju koja ima jedinstven polimorfizam i na osnovu kojeg se razlikuje od ostalih subpopulacija, ali ni u tom slučaju, postojanje mogućnosti da se subpopulacije genetički klasifikuju, neće uticati na varijante *CYP450*, ukoliko jedna od tih varijanti nije determinanta za klasifikaciju subpopulacija. S druge strane, postojanje neuniformne raspodele alelskih učestalosti na većem broju STR lokusa može biti signal da struktuiranost postoji i da je potrebno obazrivo analizirati i *CYP450* lokuse. Jedno od pitanja na koje bi ova studija mogla odgovoriti, jeste da li postoji podudarnost raspodele 15 mikrosatelitskih lokusa i 11 lokusa u četiri *CYP450* gena, a odgovor na osnovu rezultata svih primenjenih algoritama bi bio da podudarnosti nema.

6. Zaključci

1. Raspodela učestalosti funkcionalnih varijanti gena *CYP1A1*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, u opštoj populaciji Srbije je jedinstvena, ali slična raspodelama u drugim evropskim populacijama. Najveći doprinosilac izostanku ili umanjenju metabolizma ksenobiotika, među kojima su i često upotrebljavani lekovi, jeste *CYP2D6* *4 alel i genetička varijanta rs3892097 koja determiniše ovaj alel.

2. Unutar opšte populacije u Srbiji, testirani stratumi (Vojvodina, Zapadna, Centralna, Istočna i Južna Srbija, kao i područje Beograda) nemaju značajno različite raspodele učestalosti 11 SNV, osim blagog izdvajanja Istočne i Južne Srbije u odnosu na ostatak opšte populacije metodom klasterovanja. Razlike između Romske subpopulacije i ostatka opšte populacije se najviše opažaju na lokusu *CYP1A1*, varijanti M1, dok su alelske učestalosti na ostalim lokusima vrlo slične.

3. Od testiranih SNV, za uspešno testiranje u opštoj populaciji, najinformativniji su: SNV rs3892097, rs1065852, rs28371725, rs1799853, rs1057910, rs4244285, rs12248560, dok SNV rs28371706 i rs2740574, zbog niske učestalosti, ne pružaju dovoljno informacija.

4. U populaciono genetičkom smislu, postoji osnova za primenu geotipiziranja funkcionalnih varijanti *CYP450* u post mortem analizama, dok je izbor varijanti *CYP450* uslovljen uticajem na konkretni supstrat od interesa.

5. Testirane varijante rs28371706 u genu *CYP2D6* i rs2740574 u *CYP3A4*, bi se mogle izostaviti iz budućih testiranja i zameniti rs1135840 i rs35599367, za čiju funkcionalnost u pogledu metabolizma ksenobiotika se pojavljuju novi dokazi.

6. Algoritmi za testiranje genetičke struktuiranosti populacije se mogu koristiti kao polazna osnova za definisanje grupa koje se podvrgavaju klasičnom pristupu opisivanja genetičke udaljenosti, odnosno genetičkoj primeni F statistike. Postoji uočljiva razlika u metodama klasterovanja u zavisnosti od premisa na kojima se te metode zasnivaju. Posebno je istaknuta klasa algoritama (STRUCTURE, BAPS) koji se zasnivaju na Bejzovoj teoriji određivanja verovatnoće, za koju postoje interaktivne implementacije. Metodi klasterovanja zasnovani na redukciji dimenzionalnosti (PCA, tSNE) i hijerarhijskom klasterovanju (DBSCAN) su pokazali visoke performanse, ali njihova implementacija zahteva ekstenzivno bioinformatičko predznanje, kao i osnove programiranja u softverskom okruženjima R i Python. Metodi zasnovani na teoriji mreža (Girvan Newman) su takođe primenjivi u detekciji genetičke struktuiranosti, ali kompjutaciono zahtevni i bez jasne biološke osnove.

7. LITERATURA

- Abe, O., Abe, R., Enomoto, K., Kikuchi, K., Koyama, H., Masuda, H., Nomura, Y., Ohashi, Y., Sakai, K., Sugimachi, K., Toi, M., Tominaga, T., Uchino, J., Yoshida, M., Haybittle, J.L., Leonard, C.F., Calais, G., Geraud, P., Collett, V., Davies, C., Delmestr, H. 2011. "Relevance of Breast Cancer Hormone Receptors and Other Factors to the Efficacy of Adjuvant Tamoxifen: Patient-Level Meta-Analysis of Randomised Trials." *The Lancet* 378(9793): 771–84. <https://linkinghub.elsevier.com/retrieve/pii/S0140673611609938>.
- ABRAMS, WILLIAM B. et al. 1964. "Clinical Pharmacological Studies With Debrisoquin Sulfate, a New Antihypertensive Agent." *The Journal of New Drugs*.
- Abriel, Hugues et al. 2004. "Molecular and Clinical Determinants of Drug-Induced Long QT Syndrome: An Iatrogenic Channelopathy." *Swiss medical weekly* 134(47–48): 685–94. <http://www.ncbi.nlm.nih.gov/pubmed/15616901>.
- Ackerman, M J, D J Tester, and D J Driscoll. 2001. "Molecular Autopsy of Sudden Unexplained Death in the Young." *The American journal of forensic medicine and pathology* 22(2): 105–11. <http://www.ncbi.nlm.nih.gov/pubmed/11394742>.
- Aggarwal, Charu C. 2015. *Data Mining*. Cham: Springer International Publishing. <http://link.springer.com/10.1007/978-3-319-14142-8>.
- Aithal, G P, C P Day, P J Kesteven, and A K Daly. 1999. "Association of Polymorphisms in the Cytochrome P450 CYP2C9 with Warfarin Dose Requirement and Risk of Bleeding Complications." *Lancet (London, England)* 353(9154): 717–19. <http://www.ncbi.nlm.nih.gov/pubmed/10073515>.
- Aithal, Guruprasad P, Christopher P Day, Patrick JL Kesteven, and Ann K Daly. 1999. "Association of Polymorphisms in the Cytochrome P450 CYP2C9 with Warfarin Dose Requirement and Risk of Bleeding Complications." *The Lancet* 353(9154): 717–19. <https://linkinghub.elsevier.com/retrieve/pii/S0140673698044742>.
- ALIMS, 2012. 2012. "Farin, ALIMS." <https://www.alims.gov.rs/ciril/files/lekovi/smpc/515-01-7681-11-001.pdf>.
- Androutopoulos, Vasilis P, Aristidis M Tsatsakis, and Demetrios A Spandidos. 2009. "Cytochrome P450 CYP1A1: Wider Roles in Cancer Progression and Prevention." *BMC Cancer* 9(1): 187. <https://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-9-187>.
- Angelo, M. et al. 1975. "Proceedings: A Correlation between the Response to Debrisoquine and the Amount of Unchanged Drug Excreted in the Urine." *British journal of pharmacology* 55(2): 264P. <http://www.ncbi.nlm.nih.gov/pubmed/1201404>.
- Angelo, M, L G Dring, R Lancaster, and R L Smith. 1976. "The Metabolism of Debrisoquine in Rat and Man." *Biochemical Society transactions* 4(4): 704–6. <http://www.ncbi.nlm.nih.gov/pubmed/1001745>.
- Anttila, Sisko et al. 2003. "Methylation of Cytochrome P4501A1 Promoter in the Lung Is Associated with Tobacco Smoking." *Cancer research* 63(24): 8623–28. <http://www.ncbi.nlm.nih.gov/pubmed/14695173>.
- Appleby, C.A. 1967. "A Soluble Haemoprotein P 450 from Nitrogen-Fixing Rhizobium Bacteroids." *Biochimica et Biophysica Acta (BBA) - Protein Structure* 147(2): 399–402. <https://linkinghub.elsevier.com/retrieve/pii/0005279567904230>.
- Arinç, Emel, John B. Schenkman, and Ernest Hodgson, eds. 1995. *Molecular Aspects of Oxidative*

Drug Metabolizing Enzymes. Berlin, Heidelberg: Springer Berlin Heidelberg.
<http://link.springer.com/10.1007/978-3-642-79528-2>.

- Astle, William, and David J. Balding. 2009. "Population Structure and Cryptic Relatedness in Genetic Association Studies." *Statistical Science* 24(4): 451–71.
<http://projecteuclid.org/euclid.ss/1271770342>.
- Athanassiadis, D., W. I. Cranston, B. E. Juel-Jensen, and D. O. Oliver. 1966. "Clinical Observations on the Effects of Debrisoquine Sulphate in Patients with High Blood-Pressure." *BMJ* 2(5516): 732–35. <http://www.bmj.com/cgi/doi/10.1136/bmj.2.5516.732>.
- AXELROD, J. 1955. "The Enzymatic Demethylation of Ephedrine." *The Journal of pharmacology and experimental therapeutics* 114(4): 430–38.
<http://www.ncbi.nlm.nih.gov/pubmed/13252560>.
- Baatjes, K.J., M. Conradie, J.P. Apffelstaedt, and M.J. Kotze. 2018. "Pharmacogenetics of Aromatase Inhibitors in Endocrine Responsive Breast Cancer: Lessons Learnt from Tamoxifen and CYP2D6 Genotyping." *Anti-Cancer Agents in Medicinal Chemistry* 17(13).
<http://www.eurekaselect.com/151553/article>.
- Balding, DavidJ, and RichardA Nichols. 1995. "A Method for Quantifying Differentiation between Populations at Multi-Allelic Loci and Its Implications for Investigating Identity and Paternity." *Genetica* 96(1–2): 3–12. <http://dx.doi.org/10.1007/BF01441146>.
- Baldwin, R. Michael et al. 2008. "Increased Omeprazole Metabolism in Carriers of the CYP2C19*17 Allele; a Pharmacokinetic Study in Healthy Volunteers." *British Journal of Clinical Pharmacology* 65(5): 767–74. <http://doi.wiley.com/10.1111/j.1365-2125.2008.03104.x>.
- Barnaba, Carlo, Katherine Gentry, Nirupama Sumangala, and Ayyalusamy Ramamoorthy. 2017. "The Catalytic Function of Cytochrome P450 Is Entwined with Its Membrane-Bound Nature." *F1000Research* 6: 662. <http://www.ncbi.nlm.nih.gov/pubmed/28529725>.
- Barreto, George et al. 2009. "The Role of Catechols and Free Radicals in Benzene Toxicity: An Oxidative DNA Damage Pathway." *Environmental and Molecular Mutagenesis* 50(9): 771–80. <http://doi.wiley.com/10.1002/em.20500>.
- Barrett, J C, B Fry, J Maller, and M J Daly. 2005. "Haploview: Analysis and Visualization of LD and Haplotype Maps." *Bioinformatics (Oxford, England)* 21(2): 263–65.
<http://www.ncbi.nlm.nih.gov/pubmed/15297300>.
- Baumann, Pierre. 2014. CYP2D6: Genetics, Pharmacology and Clinical Relevance *CYP2D6: Genetics, Pharmacology and Clinical Relevance*. ed. Pierre Baumann. Unitec House, 2 Albert Place, London N3 1QB, UK: Future Medicine Ltd. <http://dx.doi.org/10.2217/9781780844626>.
- Beerenwinkel, Niko, and Juliane Siebourg. 2019. "Probability, Statistics, and Computational Science." In , 33–70. http://link.springer.com/10.1007/978-1-4939-9074-0_2.
- Beitia, G., A. Cobreros, L. Sainz, and E. Cenarruzabeitia. 2000. "Ecstasy-Induced Toxicity in Rat Liver." *Liver International* 20(1): 8–15. <http://doi.wiley.com/10.1034/j.1600-0676.2000.020001008.x>.
- Bell, GC et al. 2017. "Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 Genotype and Use of Ondansetron and Tropisetron." *Clinical Pharmacology & Therapeutics* 102(2): 213–18. <http://doi.wiley.com/10.1002/cpt.598>.
- Bertz, R J, and G R Granneman. 1997. "Use of in Vitro and in Vivo Data to Estimate the Likelihood of Metabolic Pharmacokinetic Interactions." *Clin Pharmacokinet* 32.

<http://dx.doi.org/10.2165/00003088-199732030-00004>.

- Bigham, Abigail W. 2014. "Human Evolutionary Genetics (2 Nd Edition). By MarkJobling, EdwardHollox, MatthewHurles, ToomasKivisild, and ChrisTyler-Smith. 650 Pp. New York: Garland Science. 2013. \$130.00 (Paper)." *American Journal of Human Biology* 26(4): 574–75. <http://doi.wiley.com/10.1002/ajhb.22564>.
- Bijl, Monique J. et al. 2008. "Influence of the CYP2D6*4 Polymorphism on Dose, Switching and Discontinuation of Antidepressants." *British Journal of Clinical Pharmacology* 65(4): 558–64. <http://doi.wiley.com/10.1111/j.1365-2125.2007.03052.x>.
- Binkhorst, Lisette, Ron H.J. Mathijssen, Agnes Jager, and Teun van Gelder. 2015. "Individualization of Tamoxifen Therapy: Much More than Just CYP2D6 Genotyping." *Cancer Treatment Reviews*.
- Bishop, Jeffrey R. 2018. "Pharmacogenetics." In , 59–73. <https://linkinghub.elsevier.com/retrieve/pii/B9780444632333000063>.
- Bomba, Lorenzo, Klaudia Walter, and Nicole Soranzo. 2017. "The Impact of Rare and Low-Frequency Genetic Variants in Common Disease." *Genome Biology* 18(1): 77. <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1212-4>.
- Boral, Benjamin M., Dennis J. Williams, and Leonard I. Boral. 2016. "Disseminated Intravascular Coagulation." *American Journal of Clinical Pathology* 146(6): 670–80. <https://academic.oup.com/ajcp/article-lookup/doi/10.1093/ajcp/aqw195>.
- Box, George E. P. 1976. "Science and Statistics." *Journal of the American Statistical Association* 71(356): 791–99. <http://www.tandfonline.com/doi/abs/10.1080/01621459.1976.10480949>.
- Brewer, Brendon J. 2017. "Introduction to Bayesian Statistics." <https://www.stat.auckland.ac.nz/~brewer/stats331.pdf>.
- Brøsen, Kim. 2004. "Some Aspects of Genetic Polymorphism in the Biotransformation of Antidepressants." *Thérapie* 59(1): 5–12.
- Bush, W. S. et al. 2016. "Genetic Variation among 82 Pharmacogenes: The PGRNseq Data from the EMERGE Network." *Clinical Pharmacology and Therapeutics* 100(2): 160–63.
- Buterin, T et al. 2000. "Unrepaired Fjord Region Polycyclic Aromatic Hydrocarbon-DNA Adducts in Ras Codon 61 Mutational Hot Spots." *Cancer research* 60(7): 1849–56. <http://www.ncbi.nlm.nih.gov/pubmed/10766171>.
- Campbell, Harry, Igor Rudan, Alan H Bittles, and Alan F Wright. 2009. "Human Population Structure, Genome Autozygosity and Human Health." *Genome medicine* 1(9): 91. <http://www.ncbi.nlm.nih.gov/pubmed/19804611>.
- Campello, Ricardo J. G. B., Peer Kröger, Jörg Sander, and Arthur Zimek. 2020. "Density-based Clustering." *WIREs Data Mining and Knowledge Discovery* 10(2). <https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1343>.
- Carbonell, N et al. 2010. "CYP2C9*3 Loss-of-Function Allele Is Associated With Acute Upper Gastrointestinal Bleeding Related to the Use of NSAIDs Other Than Aspirin." *Clinical Pharmacology & Therapeutics* 87(6): 693–98. <http://doi.wiley.com/10.1038/clpt.2010.33>.
- Carmo, Helena et al. 2006. "Influence of CYP2D6 Polymorphism on 3,4-Methylenedioxymethamphetamine ('Ecstasy') Cytotoxicity." *Pharmacogenetics and Genomics* 16(11): 789–99. <https://insights.ovid.com/crossref?an=01213011-200611000-00004>.
- Carvalho, Márcia et al. 2002. "Role of Metabolites in MDMA (Ecstasy)-Induced

- Nephrotoxicity: An in Vitro Study Using Rat and Human Renal Proximal Tubular Cells.” *Archives of Toxicology* 76(10): 581–88. <http://link.springer.com/10.1007/s00204-002-0381-3>.
- Catharine Ross, A., and Reza Zolfaghari. 2011. “Cytochrome P450s in the Regulation of Cellular Retinoic Acid Metabolism.” *Annual Review of Nutrition* 31(1): 65–87. <http://www.annualreviews.org/doi/10.1146/annurev-nutr-072610-145127>.
- Caudle, K. E. et al. 2014. “Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and HLA-B Genotypes and Phenytoin Dosing.” *Clinical Pharmacology and Therapeutics* 96(5): 542–48.
- Chakraborty, Ranajit. 1993. “Analysis of Genetic Structure of Populations: Meaning, Methods, and Implications.” In *Human Population Genetics*, Boston, MA: Springer US, 189–206. http://link.springer.com/10.1007/978-1-4615-2970-5_14.
- CHANCE, B., and G. R. WILLIAMS. 1954. “Kinetics of Cytochrome B5 in Rat Liver Microsomes.” *The Journal of biological chemistry* 209(2): 945–51. <http://www.ncbi.nlm.nih.gov/pubmed/13192148>.
- Charlesworth, B. 1998. “Measures of Divergence between Populations and the Effect of Forces That Reduce Variability.” *Molecular Biology and Evolution* 15(5): 538–43. <https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a025953>.
- Chen, Chia-Hui et al. 2011. “Genetic Polymorphisms in CYP3A4 Are Associated with Withdrawal Symptoms and Adverse Reactions in Methadone Maintenance Patients.” *Pharmacogenomics* 12(10): 1397–1406. <https://www.futuremedicine.com/doi/10.2217/pgs.11.103>.
- Chen, Qi, Tao Zhang, Jing-Fang Wang, and Dong-Qing Wei. 2011. “Advances in Human Cytochrome P450 and Personalized Medicine.” *Current drug metabolism* 12(5): 436–44. <http://www.ncbi.nlm.nih.gov/pubmed/21453272>.
- Chen, Xiaoping et al. 2009. “Molecular Population Genetics of Human CYP3A Locus: Signatures of Positive Selection and Implications for Evolutionary Environmental Medicine.” *Environmental Health Perspectives* 117(10): 1541–48. <https://ehp.niehs.nih.gov/doi/10.1289/ehp.0800528>.
- Ciszkowski, Catherine et al. 2009. “Codeine, Ultrarapid-Metabolism Genotype, and Postoperative Death.” *New England Journal of Medicine* 361(8): 827–28. <http://www.nejm.org/doi/abs/10.1056/NEJMc0904266>.
- Conway, B R, D G Fogarty, W E Nelson, and C C Doherty. 2006. “Opiate Toxicity in Patients with Renal Failure.” *BMJ* 332(7537): 345–46. <http://www.bmj.com/lookup/doi/10.1136/bmj.332.7537.345>.
- Cook, D.J. et al. 2016. “Cytochromes P450.” In *Advances in Protein Chemistry and Structural Biology*, , 105–26. <https://linkinghub.elsevier.com/retrieve/pii/S1876162316300323>.
- COOPER, D. Y., R. W. ESTABROOK, and O. ROSENTHAL. 1963. “The Stoichiometry of C21 Hydroxylation of Steroids by Adrenocortical Microsomes.” *The Journal of biological chemistry* 238: 1320–23. <http://www.ncbi.nlm.nih.gov/pubmed/14022832>.
- Cooper, David Y. et al. 1965. “Photochemical Action Spectrum of the Terminal Oxidase of Mixed Function Oxidase Systems.” *Science* 147(3656): 400–402. <http://www.sciencemag.org/cgi/doi/10.1126/science.147.3656.400>.
- CORANDER, JUKKA, and PEKKA MARTTINEN. 2006. “Bayesian Identification of Admixture Events Using Multilocus Molecular Markers.” *Molecular Ecology* 15(10): 2833–43. <http://doi.wiley.com/10.1111/j.1365-294X.2006.02994.x>.

- Corander, Jukka, Jukka Sirén, and Elja Arjas. 2008. "Bayesian Spatial Modeling of Genetic Population Structure." *Computational Statistics* 23(1): 111–29.
<http://link.springer.com/10.1007/s00180-007-0072-x>.
- Corander, Jukka, Patrik Waldmann, and Mikko J Sillanpää. 2003. "Bayesian Analysis of Genetic Differentiation Between Populations." *Genetics* 163(1): 367–74.
<http://www.genetics.org/content/163/1/367.abstract>.
- Coste, Joel, and Alain Venot. 1999. "An Epidemiologic Approach to Drug Prescribing Quality Assessment." *Medical Care* 37(12): 1294–1307. <http://www.jstor.org/stable/3766945>.
- Coura, Renata dos Santos et al. 2007. "CYP1A1 and CYP2E1 Polymorphism Frequencies in a Large Brazilian Population." *Genetics and Molecular Biology* 30(1): 1–5.
http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572007000100001&lng=en&tlng=en.
- Crews, K R et al. 2012. "Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for Codeine Therapy in the Context of Cytochrome P450 2D6 (CYP2D6) Genotype." *Clinical Pharmacology & Therapeutics* 91(2): 321–26. <http://doi.wiley.com/10.1038/clpt.2011.287>.
- Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal Complex Systems*.
- Davidson, Charles S. 1965. "The Autopsy in the Age of Molecular Biology." *JAMA: The Journal of the American Medical Association* 193(10): 813.
<http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1965.03090100059016>.
- Davis, Jerel C., and Dmitri A. Petrov. 2004. "Preferential Duplication of Conserved Proteins in Eukaryotic Genomes" ed. Ken H. Wolfe. *PLoS Biology* 2(3): e55.
<http://dx.plos.org/10.1371/journal.pbio.0020055>.
- Dean, Laura. 2012a. Medical Genetics Summaries *Tramadol Therapy and CYP2D6 Genotype*.
 . 2012b. Medical Genetics Summaries *Warfarin Therapy and VKORC1 and CYP Genotype*.
<http://www.ncbi.nlm.nih.gov/pubmed/28520347>.
- Deaton, Justin G., and Thomas M. Nappe. 2020. StatPearls *Warfarin Toxicity*.
<http://www.ncbi.nlm.nih.gov/pubmed/28613764>.
- Desta, Zeruesenay, Bryan A. Ward, Nadia V. Soukhova, and David A. Flockhart. 2004. "Comprehensive Evaluation of Tamoxifen Sequential Biotransformation by the Human Cytochrome P450 System in Vitro: Prominent Roles for CYP3A and CYP2D6." *Journal of Pharmacology and Experimental Therapeutics* 310(3): 1062–75.
<http://jpet.aspetjournals.org/lookup/doi/10.1124/jpet.104.065607>.
- Devlin, B., Kathryn Roeder, and Silviu-Alin Bacanu. 2001. "Unbiased Methods for Population-Based Association Studies." *Genetic Epidemiology* 21(4): 273–84.
<http://doi.wiley.com/10.1002/gepi.1034>.
- van Driel, B., M. Michels, and J. van der Velden. 2018. "Molecular Autopsy." *Netherlands Heart Journal* 26(10): 471–72. https://doi.org/10.1142/9781786340788_0016.
- Dumontier, Michel, and Robert Hoehndorf. 2010. 209 Proc. FOIS *Realism for Scientific Ontologies*.
- Dyer, Rodney J. 2017. *Applied Population Genetics*.
http://dyerlab.github.io/applied_population_genetics/.
- Eap, C B et al. 2007. "Stereoselective Block of HERG Channel by (S)-Methadone and QT Interval Prolongation in CYP2B6 Slow Metabolizers." *Clinical Pharmacology & Therapeutics* 81(5):

719–28. <http://doi.wiley.com/10.1038/sj.clpt.6100120>.

- Edward, Robinson L. 2017. *Data Analysis for Scientists and Engineers*. New Jersey: Princeton University Press. <https://press.princeton.edu/books/hardcover/9780691169927/data-analysis-for-scientists-and-engineers>.
- Ehmann, F et al. 2015. “Pharmacogenomic Information in Drug Labels: European Medicines Agency Perspective.” *The Pharmacogenomics Journal* 15(3): 201–10. <http://www.nature.com/articles/tpj201486>.
- Eichelbaum, Michel, Magnus Ingelman-Sundberg, and William E. Evans. 2006. “Pharmacogenomics and Individualized Drug Therapy.” *Annual Review of Medicine* 57(1): 119–37. <http://www.annualreviews.org/doi/10.1146/annurev.med.56.082103.104724>.
- Elens, Laure et al. 2013. “Impact of CYP3A4*22 Allele on Tacrolimus Pharmacokinetics in Early Period After Renal Transplantation.” *Therapeutic Drug Monitoring*: 1. <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00007691-900000000-99499>.
- Elkalioubie, Ahmed et al. 2011. “Near-Fatal Tramadol Cardiotoxicity in a CYP2D6 Ultrarapid Metabolizer.” *European Journal of Clinical Pharmacology* 67(8): 855–58. <http://link.springer.com/10.1007/s00228-011-1080-x>.
- Estabrook, Ronald W. 2003. “A PASSION FOR P450s (REMEMBRANCES OF THE EARLY HISTORY OF RESEARCH ON CYTOCHROME P450).” *Drug Metabolism and Disposition* 31(12): 1461–73. <http://dmd.aspetjournals.org/cgi/doi/10.1124/dmd.31.12.1461>.
- Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*.
- EVANNO, G., S. REGNAUT, and J. GOUDET. 2005. “Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study.” *Molecular Ecology* 14(8): 2611–20. <http://doi.wiley.com/10.1111/j.1365-294X.2005.02553.x>.
- Ewens, W J, and R S Spielman. 1995. “The Transmission/Disequilibrium Test: History, Subdivision, and Admixture.” *American journal of human genetics* 57(2): 455–64. <http://www.ncbi.nlm.nih.gov/pubmed/7668272>.
- Excoffier, L, P E Smouse, and J M Quattro. 1992. “Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data.” *Genetics* 131(2): 479–91. <http://www.ncbi.nlm.nih.gov/pubmed/1644282>.
- Excoffier, Laurent, and Heidi E L Lischer. 2010. “Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows.” *Molecular Ecology Resources*.
- Fanni, D. et al. 2014. “Cytochrome P450 Genetic Polymorphism in Neonatal Drug Metabolism: Role and Practical Consequences towards a New Drug Culture in Neonatology.” *International Journal of Immunopathology and Pharmacology* 27(1): 5–13. <http://journals.sagepub.com/doi/10.1177/039463201402700102>.
- Farmer, Peter B. et al. 2003. “Molecular Epidemiology Studies of Carcinogenic Environmental Pollutants.” *Mutation Research/Reviews in Mutation Research* 544(2–3): 397–402. <https://linkinghub.elsevier.com/retrieve/pii/S1383574203000863>.
- Fathimah, Nurul, Chris Adhiyanto, and Hari Hendarto. 2017. “The Description of Polymorphism of CYP1A1*2A Rs4646903 (TgC) Gene as Colorectal Cancer Risk Factor In Medical Study

Programs and Doctor Profession 2012-2014 UIN Syarif Hidayatullah Jakarta.” In *Proceedings of the 1st International Integrative Conference on Health, Life and Social Sciences (ICHLaS 2017)*, Paris, France: Atlantis Press. <http://www.atlantis-press.com/php/paper-details.php?id=25887004>.

- Ferris, J.P. et al. 1973. “Monooxygenase Activity in *Cunninghamella Bainieri*: Evidence for a Fungal System Similar to Liver Microsomes.” *Archives of Biochemistry and Biophysics* 156(1): 97–103. <https://linkinghub.elsevier.com/retrieve/pii/0003986173903457>.
- Fisher, R. A. 1919. “XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance.” *Transactions of the Royal Society of Edinburgh* 52(2): 399–433. https://www.cambridge.org/core/product/identifier/S0080456800012163/type/journal_article.
- . 1922. “On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P.” *Journal of the Royal Statistical Society* 85(1): 87. <https://www.jstor.org/stable/2340521?origin=crossref>.
- . 1992. “Statistical Methods for Research Workers.” In , 66–70. http://link.springer.com/10.1007/978-1-4612-4380-9_6.
- Foreman, L. A., A. F. M. Smith, and I. W. Evett. 1997. “A Bayesian Approach to Validating STR Multiplex Databases for Use in Forensic Casework.” *International Journal of Legal Medicine* 110(5): 244–50. <http://link.springer.com/10.1007/s004140050079>.
- Frankham, Richard, Jonathan D. Ballou, David A. Briscoe, and Karina H. McInnes. 2002. *Introduction to Conservation Genetics*. Cambridge University Press. <https://www.cambridge.org/core/product/identifier/9780511808999/type/book>.
- Fujikura, Kohei, Magnus Ingelman-Sundberg, and Volker M. Lauschke. 2015. “Genetic Variation in the Human Cytochrome P450 Supergene Family.” *Pharmacogenetics and Genomics* 25(12): 584–94. <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=01213011-201512000-00002>.
- Gaedigk, Andrea et al. 2018. “The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database.” *Clinical Pharmacology & Therapeutics* 103(3): 399–401. <http://doi.wiley.com/10.1002/cpt.910>.
- Ganoci, Lana et al. 2017. “Genetic Polymorphisms of Cytochrome P450 Enzymes: CYP2C9, CYP2C19, CYP2D6, CYP3A4, and CYP3A5 in the Croatian Population.” *Drug Metabolism and Personalized Therapy* 32(1). <http://www.degruyter.com/view/j/dmdi.2017.32.issue-1/dmpt-2016-0024/dmpt-2016-0024.xml>.
- García-Martín, Elena. 2002. “CYP3A4 Variant Alleles in White Individuals with Low CYP3A4 Enzyme Activity.” *Clinical Pharmacology & Therapeutics* 71(3): 196–204. <http://doi.wiley.com/10.1067/mcp.2002.121371>.
- Gentle, James R. 2013. *Theory of Statistics*. <http://www.math.ttu.edu/~atrindad/stat5381/MathStat-Gentle-2014.pdf>.
- Gex-Fabry, Marianne et al. 2008. “CYP2D6 and ABCB1 Genetic Variability: Influence on Paroxetine Plasma Level and Therapeutic Response.” *Therapeutic Drug Monitoring* PAP. <http://journals.lww.com/00007691-900000000-99983>.
- Gibson, G. Gordon, and Paul Skett. 1986. *Introduction to Drug Metabolism*. Boston, MA: Springer US. <http://link.springer.com/10.1007/978-1-4899-3188-7>.
- Giere, Ronald N. 2004. “How Models Are Used to Represent Reality.” *Philosophy of Science* 71(5):

- 742–52. <https://www.journals.uchicago.edu/doi/10.1086/425063>.
- Go, Ryeo-Eun, Kyung-A Hwang, and Kyung-Chul Choi. 2015. “Cytochrome P450 1 Family and Cancers.” *The Journal of Steroid Biochemistry and Molecular Biology* 147: 24–30. <https://linkinghub.elsevier.com/retrieve/pii/S0960076014002581>.
- Goetz, Matthew P. et al. 2018. “Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and Tamoxifen Therapy.” *Clinical Pharmacology & Therapeutics* 103(5): 770–77. <http://doi.wiley.com/10.1002/cpt.1007>.
- Göring, Harald H.H., Joseph D. Terwilliger, and John Blangero. 2001. “Large Upward Bias in Estimation of Locus-Specific Effects from Genomewide Scans.” *The American Journal of Human Genetics* 69(6): 1357–69. <https://linkinghub.elsevier.com/retrieve/pii/S0002929707612646>.
- GOUDET, JEROME. 2005. “Hierfstat, a Package for r to Compute and Test Hierarchical F-Statistics.” *Molecular Ecology Notes* 5(1): 184–86. <http://doi.wiley.com/10.1111/j.1471-8286.2004.00828.x>.
- Greenbaum, Gili, Alan R. Templeton, and Shirli Bar-David. 2016. “Inference and Analysis of Population Structure Using Genetic Data and Network Theory.” *Genetics* 202(4): 1299–1312. <http://www.genetics.org/lookup/doi/10.1534/genetics.115.182626>.
- Grond, Stefan, and Armin Sablotzki. 2004. “Clinical Pharmacology of Tramadol.” *Clinical Pharmacokinetics* 43(13): 879–923. <http://link.springer.com/10.2165/00003088-200443130-00004>.
- Guengerich, F. Peter. 2018. “Mechanisms of Cytochrome P450-Catalyzed Oxidations.” *ACS Catalysis* 8(12): 10964–76. <https://pubs.acs.org/doi/10.1021/acscatal.8b03401>.
- Guengerich, F. Peter, Michael R. Waterman, and Martin Egli. 2016. “Recent Structural Insights into Cytochrome P450 Function.” *Trends in Pharmacological Sciences* 37(8): 625–40. <https://linkinghub.elsevier.com/retrieve/pii/S0165614716300402>.
- Haahr, Mads. 2012. “True Random Number Service.” www.Random.org. www.Random.org.
- Hahsler, Michael, Matthew Piekenbrock, and Derek Doran. 2019. “DbSCAN : Fast Density-Based Clustering with R.” *Journal of Statistical Software* 91(1). <http://www.jstatsoft.org/v91/i01/>.
- Halder, Indrani, and Mark D Shriver. 2003. “Measuring and Using Admixture to Study the Genetics of Complex Diseases.” *Human Genomics* 1(1): 52. <http://humgenomics.biomedcentral.com/articles/10.1186/1479-7364-1-1-52>.
- Halpern, Joel. 1999. “The Ecological Transformation of a Resettled Area, Pig Herders to Settled Farmers in Central Serbia (Sumadija, Yugoslavia) during the 19th and 20th Centuries.” In *TRANSHUMANT PASTORALISM IN SOUTHERN EUROPE Recent Perspectives from Archaeology, History and Ethnology*, Budapest: Archaeolingua, 245. https://scholarworks.umass.edu/anthro_faculty_pubs/64.
- Harding, Boyd W., Siu Ha Wong, and Don H. Nelson. 1964. “Carbon Monoxide-Combining Substances in Rat Adrenal.” *Biochimica et Biophysica Acta (BBA) - Specialized Section on Enzymological Subjects* 92(2): 415–17. <https://linkinghub.elsevier.com/retrieve/pii/0926656964902068>.
- Hardy, Olivier J, and Xavier Vekemans. 1999. “Isolation by Distance in a Continuous Population: Reconciliation between Spatial Autocorrelation Analysis and Population Genetics Models.” *Heredity* 83(2): 145–54. <http://link.springer.com/10.1046/j.1365-2540.1999.00558.x>.

- HASHIMOTO, Y., T. YAMANO, and H. S. MASON. 1962. "An Electron Spin Resonance Study of Microsomal Electron Transport." *The Journal of biological chemistry* 237: 3843–44. <http://www.ncbi.nlm.nih.gov/pubmed/13960870>.
- Hasler, Julia A et al. 1999. "Human Cytochromes P450." *Molecular Aspects of Medicine* 20(1–2): 1–137. <https://linkinghub.elsevier.com/retrieve/pii/S0098299799000059>.
- Hastings, W. K. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57(1): 97–109. <https://academic.oup.com/biomet/article/57/1/97/284580>.
- Haufroid, Vincent, and Philippe Hantson. 2015. "CYP2D6 Genetic Polymorphisms and Their Relevance for Poisoning Due to Amfetamines, Opioid Analgesics and Antidepressants." *Clinical Toxicology* 53(6): 501–10. <http://www.tandfonline.com/doi/full/10.3109/15563650.2015.1049355>.
- Hedrick, Philip W. 1999. "PERSPECTIVE: HIGHLY VARIABLE LOCI AND THEIR INTERPRETATION IN EVOLUTION AND CONSERVATION." *Evolution* 53(2): 313–18. <http://doi.wiley.com/10.1111/j.1558-5646.1999.tb03767.x>.
- . 2005. "A STANDARDIZED GENETIC DIFFERENTIATION MEASURE." *Evolution* 59(8): 1633–38. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0014-3820.2005.tb01814.x>.
- Hermanns-Clausen, Maren et al. 2009. "Drug Dosing Error with Drops—Severe Clinical Course of Codeine Intoxication in Twins." *European Journal of Pediatrics* 168(7): 819–24. <http://link.springer.com/10.1007/s00431-008-0842-7>.
- Hicks, J K et al. 2013. "Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Tricyclic Antidepressants." *Clinical Pharmacology & Therapeutics* 93(5): 402–8. <http://doi.wiley.com/10.1038/clpt.2013.2>.
- Hirota, Takeshi et al. 2004. "Allelic Expression Imbalance of the Human CYP3A4 Gene and Individual Phenotypic Status." *Human Molecular Genetics* 13(23): 2959–69. <http://academic.oup.com/hmg/article/13/23/2959/625563/Allelic-expression-imbalance-of-the-human-CYP3A4>.
- Hoffman, Susan M G, David R. Nelson, and Diane S. Keeney. 2001. "Organization, Structure and Evolution of the CYP2 Gene Cluster on Human Chromosome 19." *Pharmacogenetics* 11(8): 687–98. <https://insights.ovid.com/crossref?an=00008571-200111000-00007>.
- HOGEBOM, G H. 1949. "Cytochemical Studies of Mammalian Tissues; the Distribution of Diphosphopyridine Nucleotide-Cytochrome c Reductase in Rat Liver Fractions." *The Journal of biological chemistry* 177(2): 847–58. <http://www.ncbi.nlm.nih.gov/pubmed/18110461>.
- Hosono, Naoya et al. 2009. "CYP2D6 Genotyping for Functional-Gene Dosage Analysis by Allele Copy Number Detection." *Clinical Chemistry* 55(8): 1546–54.
- Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24(6): 417–41. <http://content.apa.org/journals/edu/24/6/417>.
- Huber, Chuck. 2016. "Introduction to Bayesian Statistics, Part 2: MCMC and the Metropolis–Hastings Algorithm." <https://blog.stata.com/2016/11/15/introduction-to-bayesian-statistics-part-2-mcmc-and-the-metropolis-hastings-algorithm/>.
- Ian, Phillips R., Shephard A. Elizabeth, and David R. Nelson. 2006. "Cytochrome P450 Nomenclature, 2004." In *Cytochrome P450 Protocols*.

- Ingelman-Sundberg, Magnus, Sarah C Sim, Alvin Gomez, and Cristina Rodriguez-Antona. 2007. "Influence of Cytochrome P450 Polymorphisms on Drug Therapies: Pharmacogenetic, Pharmacoeconomic and Clinical Aspects." *Pharmacology & therapeutics* 116(3): 496–526.
- Inoue, K., T. Asao, and T. Shimada. 2000. "Ethnic-Related Differences in the Frequency Distribution of Genetic Polymorphisms in the CYP1A1 and CYP1B1 Genes in Japanese and Caucasian Populations." *Xenobiotica* 30(3): 285–95.
<http://www.tandfonline.com/doi/full/10.1080/004982500237677>.
- Isin, Emre M., and F. Peter Guengerich. 2007. "Complex Reactions Catalyzed by Cytochrome P450 Enzymes." *Biochimica et Biophysica Acta (BBA) - General Subjects* 1770(3): 314–29.
<https://linkinghub.elsevier.com/retrieve/pii/S0304416506001991>.
- Jakobsson, Mattias, Michael D. Edge, and Noah A. Rosenberg. 2013. "The Relationship Between F ST and the Frequency of the Most Frequent Allele." *Genetics* 193(2): 515–28.
<http://www.genetics.org/lookup/doi/10.1534/genetics.112.144758>.
- Jannetto, Paul J. et al. 2002. "Pharmacogenomics as Molecular Autopsy for Postmortem Forensic Toxicology: Genotyping Cytochrome P450 2D6 for Oxycodone Cases." *Journal of Analytical Toxicology*.
- Janssens, A C J W, and P A Deverka. 2014. "Useless Until Proven Effective: The Clinical Utility of Preemptive Pharmacogenetic Testing." *Clinical Pharmacology & Therapeutics* 96(6): 652–54.
<http://doi.wiley.com/10.1038/clpt.2014.186>.
- Jin, Ming et al. 2005. "Pharmacogenomics as Molecular Autopsy for Forensic Toxicology: Genotyping Cytochrome P450 3A4*1B and 3A5*3 for 25 Fentanyl Cases." *Journal of Analytical Toxicology* 29(7): 590–598.
- Johnson, J A et al. 2011. "Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 Genotypes and Warfarin Dosing." *Clinical Pharmacology & Therapeutics* 90(4): 625–29. <http://doi.wiley.com/10.1038/clpt.2011.185>.
- Jolliffe, Ian T., and Jorge Cadima. 2016. "Principal Component Analysis: A Review and Recent Developments." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2065): 20150202.
<https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202>.
- Jombart, Thibaut. 2008. "Adegenet: A R Package for the Multivariate Analysis of Genetic Markers." *Bioinformatics* 24(11): 1403–5. <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btn129>.
- Jones, Owen R., and Jinliang Wang. 2012. "A Comparison of Four Methods for Detecting Weak Genetic Structure from Marker Data." *Ecology and Evolution* 2(5): 1048–55.
<http://doi.wiley.com/10.1002/ece3.237>.
- Jorde, Lynn B. 1980. "The Genetic Structure of Subdivided Human Populations." In *Current Developments in Anthropological Genetics*, Boston, MA: Springer US, 135–208.
http://link.springer.com/10.1007/978-1-4613-3084-4_7.
- Jorge-Nebert, Lucia F. et al. 2010. "Analysis of Human CYP1A1 and CYP1A2 Genes and Their Shared Bidirectional Promoter in Eight World Populations." *Human Mutation* 31(1): 27–40.
<http://doi.wiley.com/10.1002/humu.21132>.
- JOST, LOU. 2008. "G ST and Its Relatives Do Not Measure Differentiation." *Molecular Ecology* 17(18): 4015–26. <http://doi.wiley.com/10.1111/j.1365-294X.2008.03887.x>.
- Jukić, Marin M., Tore Haslemo, Espen Molden, and Magnus Ingelman-Sundberg. 2018. "Impact of

- CYP2C19 Genotype on Escitalopram Exposure and Therapeutic Failure: A Retrospective Study Based on 2,087 Patients.” *American Journal of Psychiatry* 175(5): 463–70.
<http://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.2017.17050550>.
- Kacevska, Marina et al. 2012. “DNA Methylation Dynamics in the Hepatic CYP3A4 Gene Promoter.” *Biochimie* 94(11): 2338–44.
<https://linkinghub.elsevier.com/retrieve/pii/S0300908412002957>.
- Kadlubar, Fred F et al. 2003. “The CYP3A4*1B Variant Is Related to the Onset of Puberty, a Known Risk Factor for the Development of Breast Cancer.” *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 12(4): 327–31.
<http://www.ncbi.nlm.nih.gov/pubmed/12692107>.
- Kaiko, Robert F. et al. 1996. “Pharmacokinetic-Pharmacodynamic Relationships of Controlled-Release Oxycodone*.” *Clinical Pharmacology & Therapeutics* 59(1): 52–61.
[http://doi.wiley.com/10.1016/S0009-9236\(96\)90024-7](http://doi.wiley.com/10.1016/S0009-9236(96)90024-7).
- Kalinowski, S T. 2011. “The Computer Program STRUCTURE Does Not Reliably Identify the Main Genetic Clusters within Species: Simulations and Implications for Human Population Structure.” *Heredity* 106(4): 625–32. <http://www.nature.com/articles/hdy201095>.
- Kamvar, Zhian N., Javier F. Tabima, and Niklaus J. Grünwald. 2014. “Poppr : An R Package for Genetic Analysis of Populations with Clonal, Partially Clonal, and/or Sexual Reproduction.” *PeerJ* 2: e281. <https://peerj.com/articles/281>.
- Kanitz, F P. 1868. *Serbien: Historisch-Ethnographische Reisestudien Aus Den Jahren, 1859-1868*. Fries. <https://books.google.rs/books?id=Jh4UAAAAQAAJ>.
- Kassambara, A. 2017. “Extract and Visualize the Results of Multivariate Data Analysis.” <https://cran.r-project.org/web/packages/factoextra/index.html>.
- Kassem, Heba Sh., Francesca Girolami, and Despina Sanoudou. 2012. “Molecular Genetics Made Simple.” *Global Cardiology Science and Practice* 2012(1): 6.
<http://www.qscience.com/doi/abs/10.5339/gcsp.2012.6>.
- Katagiri, M., B. N. Ganguli, and I. C. Gunsalus. 1968. “A Soluble Cytochrome P-450 Functional in Methylene Hydroxylation.” *The Journal of biological chemistry* 243(12): 3543–46.
<http://www.ncbi.nlm.nih.gov/pubmed/4297783>.
- Kawakami, Hirotaka et al. 2011. “Simultaneous Absolute Quantification of 11 Cytochrome P450 Isoforms in Human Liver Microsomes by Liquid Chromatography Tandem Mass Spectrometry with In Silico Target Peptide Selection.” *Journal of Pharmaceutical Sciences* 100(1): 341–52.
<https://linkinghub.elsevier.com/retrieve/pii/S0022354915323352>.
- Kim, KA, PW Park, SJ Hong, and J-Y Park. 2008. “The Effect of CYP2C19 Polymorphism on the Pharmacokinetics and Pharmacodynamics of Clopidogrel: A Possible Mechanism for Clopidogrel Resistance.” *Clinical Pharmacology & Therapeutics* 84(2): 236–42.
<http://doi.wiley.com/10.1038/clpt.2008.20>.
- Kim, Se-Hyung et al. 2018. “Physiologically Based Pharmacokinetic Modelling of Atomoxetine with Regard to CYP2D6 Genotypes.” *Scientific Reports* 8(1): 12405.
<http://www.nature.com/articles/s41598-018-30841-8>.
- Kimura, M., and T. Ohta. 1978. “Stepwise Mutation Model and Distribution of Allelic Frequencies in a Finite Population.” *Proceedings of the National Academy of Sciences* 75(6): 2868–72.
<http://www.pnas.org/cgi/doi/10.1073/pnas.75.6.2868>.

- Kirkwood, L. C., R. L. Nation, and A. A. Somogyi. 2003. "Characterization of the Human Cytochrome P450 Enzymes Involved in the Metabolism of Dihydrocodeine." *British Journal of Clinical Pharmacology* 44(6): 549–55. <http://doi.wiley.com/10.1046/j.1365-2125.1997.t01-1-00626.x>.
- Kittles, Rick A. et al. 2002. "CYP3A4-V and Prostate Cancer in African Americans: Causal or Confounding Association Because of Population Stratification?" *Human Genetics* 110(6): 553–60. <http://link.springer.com/10.1007/s00439-002-0731-5>.
- Klingenberg, Martin. 1958. "Pigments of Rat Liver Microsomes." *Archives of Biochemistry and Biophysics* 75(2): 376–86. <https://linkinghub.elsevier.com/retrieve/pii/0003986158904363>.
- . 2003. "The Dragging Emergence of the P450 Cytochrome." *Archives of Biochemistry and Biophysics* 412(1): 2. <https://linkinghub.elsevier.com/retrieve/pii/S0003986103000602>.
- Klotz, Ulrich. 2009. "Impact of CYP2C19 Polymorphisms on the Clinical Action of Proton Pump Inhibitors (PPIs)." *European Journal of Clinical Pharmacology* 65(1): 1–2. <http://link.springer.com/10.1007/s00228-008-0571-x>.
- Kocabay, G, I Okçular, V Akkaya, and Kerim Güler. 2006. "Suicide Attempt with Clopidogrel." *Human & Experimental Toxicology* 25(12): 731–34. <http://journals.sagepub.com/doi/10.1177/0960327106073794>.
- Kondrashov, Fyodor A., and Eugene V. Koonin. 2004. "A Common Framework for Understanding the Origin of Genetic Dominance and Evolutionary Fates of Gene Duplications." *Trends in Genetics* 20(7): 287–90. <https://linkinghub.elsevier.com/retrieve/pii/S0168952504001258>.
- Kong, Lingji, Carl Lee, and J.H. Sepanski. 2007. "On the Properties of Beta-Gamma Distribution." *Journal of Modern Applied Statistical Methods* 6(1): 187–211. <http://digitalcommons.wayne.edu/jmasm/vol6/iss1/18>.
- Koren, Gideon et al. 2006. "Pharmacogenetics of Morphine Poisoning in a Breastfed Neonate of a Codeine-Prescribed Mother." *Lancet*.
- Kovacic, Peter, and Andrew L. Cooksy. 2005. "Unifying Mechanism for Toxicity and Addiction by Abused Drugs: Electron Transfer and Reactive Oxygen Species." *Medical Hypotheses* 64(2): 357–66. <https://linkinghub.elsevier.com/retrieve/pii/S0306987704004530>.
- Krijthe, Jesse H. 2015. "Rtsne: T-Distributed Stochastic Neighbor Embedding Using Barnes-Hut Implementation."
- Kuzmanovska, M et al. 2015. "CYP2D6 Allele Distribution in Macedonians, Albanians and Romanians in the Republic of Macedonia." *Balkan Journal of Medical Genetics* 18(2): 49–58. <http://content.sciendo.com/view/journals/bjmg/18/2/article-p49.xml>.
- Lamb, David C. et al. 2019. "On the Occurrence of Cytochrome P450 in Viruses." *Proceedings of the National Academy of Sciences* 116(25): 12343–52. <http://www.pnas.org/lookup/doi/10.1073/pnas.1901080116>.
- Lamba, Jatinder K, Yvonne S Lin, Erin G Schuetz, and Kenneth E Thummel. 2002. "Genetic Contribution to Variable Human CYP3A-Mediated Metabolism." *Advanced drug delivery reviews* 54(10): 1271–94. <http://www.ncbi.nlm.nih.gov/pubmed/12406645>.
- Lange-Brock, N et al. 2002. "Akutes Leberversagen Nach Dem Konsum von Ecstasy (MDMA)." *Zeitschrift für Gastroenterologie* 40(8): 581–86. <http://www.thieme-connect.de/DOI/DOI?10.1055/s-2002-33416>.
- Lapple, Florian et al. 2003. "Differential Expression and Function of CYP2C Isoforms in Human

Intestine and Liver.” *Pharmacogenetics* 13(9): 565–75.

- Lawson, Daniel J., Lucy van Dorp, and Daniel Falush. 2018. “A Tutorial on How Not to Over-Interpret STRUCTURE and ADMIXTURE Bar Plots.” *Nature Communications* 9(1): 3258. <http://www.nature.com/articles/s41467-018-05257-7>.
- Lawson, Daniel John. 2015. “Populations in Statistical Genetic Modelling and Inference.” In *Population in the Human Sciences: Concepts, Models, Evidence*, Oxford: Oxford University Press, 108–31. <http://http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199688203.001.0001/acprof-9780199688203-chapter-4>.
- Lee, Chul-Ho et al. 2005. “[Effects of Oxidative DNA Damage Induced by Polycyclic Aromatic Hydrocarbons and Genetic Polymorphism of the Paraoxonase-1 (PON1) Gene on Lung Cancer].” *Journal of preventive medicine and public health = Yebang Uihakhoe chi* 38(3): 345–50. <http://www.ncbi.nlm.nih.gov/pubmed/16323636>.
- Li-Wan-Po, Alain et al. 2010. “Pharmacogenetics of CYP2C19: Functional and Clinical Implications of a New Variant CYP2C19*17.” *British Journal of Clinical Pharmacology* 69(3): 222–30. <https://bpspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2125.2009.03578.x>.
- Li, J. Z. et al. 2008. “Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation.” *Science* 319(5866): 1100–1104. <http://www.sciencemag.org/cgi/doi/10.1126/science.1153717>.
- Li, Wentian, Jane E. Cerise, Yaning Yang, and Henry Han. 2017. “Application of T-SNE to Human Genetic Data.” *Journal of Bioinformatics and Computational Biology* 15(04): 1750017. <https://www.worldscientific.com/doi/abs/10.1142/S0219720017500172>.
- LI, Y., H.-L. TANG, Y.-F. HU, and H.-G. XIE. 2012. “The Gain-of-Function Variant Allele CYP2C19*17: A Double-Edged Sword between Thrombosis and Bleeding in Clopidogrel-Treated Patients.” *Journal of Thrombosis and Haemostasis* 10(2): 199–206. <http://doi.wiley.com/10.1111/j.1538-7836.2011.04570.x>.
- Liechti, Matthias E., and Franz X. Vollenweider. 2001. “Which Neuroreceptors Mediate the Subjective Effects of MDMA in Humans? A Summary of Mechanistic Studies.” *Human Psychopharmacology: Clinical and Experimental* 16(8): 589–98. <http://doi.wiley.com/10.1002/hup.348>.
- Lim, Young Chai, Zeruesenay Desta, David A. Flockhart, and Todd C. Skaar. 2005. “Endoxifen (4-Hydroxy-N-Desmethyl-Tamoxifen) Has Anti-Estrogenic Effects in Breast Cancer Cells with Potency Similar to 4-Hydroxy-Tamoxifen.” *Cancer Chemotherapy and Pharmacology* 55(5): 471–78. <http://link.springer.com/10.1007/s00280-004-0926-7>.
- Lin, P et al. 2000. “Association of CYP1A1 and Microsomal Epoxide Hydrolase Polymorphisms with Lung Squamous Cell Carcinoma.” *British Journal of Cancer* 82(4): 852–57. <http://www.nature.com/doi/10.1054/bjoc.1999.1011>.
- Lindenmayer, Aristid, and Lucile Smith. 1964. “Cytochromes and Other Pigments of Baker’s Yeast Grown Aerobically and Anaerobically.” *Biochimica et Biophysica Acta (BBA) - General Subjects* 93(3): 445–61. <https://linkinghub.elsevier.com/retrieve/pii/0304416564903290>.
- Lötsch, Jörn et al. 2013. “Pharmacokinetics of Non-Intravenous Formulations of Fentanyl.” *Clinical Pharmacokinetics* 52(1): 23–36. <http://link.springer.com/10.1007/s40262-012-0016-7>.
- Lu, Jian et al. 2020. “New Insights of CYP1A in Endogenous Metabolism: A Focus on Single

- Nucleotide Polymorphisms and Diseases.” *Acta Pharmaceutica Sinica B* 10(1): 91–104.
<https://linkinghub.elsevier.com/retrieve/pii/S2211383519307701>.
- Lynch, Michael. 2007. In the Light of Evolution *In the Light of Evolution*. Washington, D.C.: National Academies Press. <http://www.nap.edu/catalog/11790>.
- Van Der Maaten, Laurens, and Geoffrey Hinton. 2008. “Visualizing Data Using T-SNE.” *Journal of Machine Learning Research*.
- Mahgoub, A. et al. 1977. “POLYMORPHIC HYDROXYLATION OF DEBRISOQUINE IN MAN.” *The Lancet* 310(8038): 584–86.
<https://linkinghub.elsevier.com/retrieve/pii/S0140673677914301>.
- Mališić, J. 1997. *Verovatnoća i Statistika: Zbirka Rešenih Zadataka Za IV Razred Gimnazija, Tehničkih Škola i Matematičku Gimnaziju*. Krug.
<https://books.google.rs/books?id=o60SAAAACAAJ>.
- Manikandan, Palrasu, and Siddavaram Nagini. 2018. “Cytochrome P450 Structure, Function and Clinical Significance: A Review.” *Current Drug Targets* 19(1).
- Mao, Yaping et al. 2019. “Interaction between CYP1A1/CYP17A1 Polymorphisms and Parental Risk Factors in the Risk of Hypospadias in a Chinese Population.” *Scientific Reports* 9(1): 4123. <http://www.nature.com/articles/s41598-019-40755-8>.
- Marcel, J. 1948. “G. Malecot. — *Les Mathématiques de l’Hérédité*. 66 p., 9 Fig., Masson, Paris 1948.” *Publications de la Société Linnéenne de Lyon*: 203.
https://www.persee.fr/doc/linly_0366-1326_1948_num_17_10_8510_t1_0203_0000_2.
- Le Marchand, L et al. 1998. “Associations of CYP1A1, GSTM1, and CYP2E1 Polymorphisms with Lung Cancer Suggest Cell Type Specificities to Tobacco Carcinogens.” *Cancer research* 58(21): 4858–63. <http://www.ncbi.nlm.nih.gov/pubmed/9809991>.
- Martignoni, Marcella, Geny M M Groothuis, and Ruben de Kanter. 2006. “Species Differences between Mouse, Rat, Dog, Monkey and Human CYP-Mediated Drug Metabolism, Inhibition and Induction.” *Expert Opinion on Drug Metabolism & Toxicology* 2(6): 875–94.
<http://www.tandfonline.com/doi/full/10.1517/17425255.2.6.875>.
- Matthew B. Hamilton. 2012. *Population Genetics*. 1st ed. Chichester: Wiley-Blackwell.
<http://www.wiley.com/go/hamiltongenetics>.
- McArthur, Andrew G. et al. 2003. “Phylogenetic Analysis of the Cytochrome P450 3 (CYP3) Gene Family.” *Journal of Molecular Evolution* 57(2): 200–211.
<http://link.springer.com/10.1007/s00239-003-2466-x>.
- McCluggage, Lauren K., Stacy A. Voils, and Malcolm Ross Bullock. 2009. “Phenytoin Toxicity Due to Genetic Polymorphism.” *Neurocritical Care* 10(2): 222–24.
<http://link.springer.com/10.1007/s12028-008-9165-9>.
- Mehta, Shamir R et al. 2010. “Double-Dose versus Standard-Dose Clopidogrel and High-Dose versus Low-Dose Aspirin in Individuals Undergoing Percutaneous Coronary Intervention for Acute Coronary Syndromes (CURRENT-OASIS 7): A Randomised Factorial Trial.” *The Lancet* 376(9748): 1233–43. <https://linkinghub.elsevier.com/retrieve/pii/S0140673610610884>.
- MEIRMANS, PATRICK G., and PHILIP W. HEDRICK. 2011. “Assessing Population Structure: FST and Related Measures.” *Molecular Ecology Resources* 11(1): 5–18.
<http://doi.wiley.com/10.1111/j.1755-0998.2010.02927.x>.
- Mendizabal, Isabel et al. 2012. “Reconstructing the Population History of European Romani from

- Genome-Wide Data.” *Current Biology* 22(24): 2342–49.
<https://linkinghub.elsevier.com/retrieve/pii/S0960982212012602>.
- Mestres, Jordi. 2004. “Structure Conservation in Cytochromes P450.” *Proteins: Structure, Function, and Bioinformatics* 58(3): 596–609. <http://doi.wiley.com/10.1002/prot.20354>.
- Miyazaki, Mitsue et al. 2008. “Defective Activity of Recombinant Cytochromes P450 3A4.2 and 3A4.16 in Oxidation of Midazolam, Nifedipine, and Testosterone.” *Drug Metabolism and Disposition* 36(11): 2287–91.
<http://dmd.aspetjournals.org/lookup/doi/10.1124/dmd.108.021816>.
- Mizukami, Y. et al. 1983. “Gene Structure of a Phenobarbital-Inducible Cytochrome P-450 in Rat Liver.” *Proceedings of the National Academy of Sciences* 80(13): 3958–62.
<http://www.pnas.org/cgi/doi/10.1073/pnas.80.13.3958>.
- Mizzi, Clint et al. 2016. “A European Spectrum of Pharmacogenomic Biomarkers: Implications for Clinical Pharmacogenomics” ed. Marie-Pierre Dubé. *PLOS ONE* 11(9): e0162866.
<https://dx.plos.org/10.1371/journal.pone.0162866>.
- Moore, Nicholas, Dominique Lecointre, Catherine Noblet, and Michel Mabile. 1998. “Frequency and Cost of Serious Adverse Drug Reactions in a Department of General Medicine.” *British Journal of Clinical Pharmacology* 45(3): 301–8.
<https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2125.1998.00667.x>.
- Mrozikiewicz, Przemysław M et al. 2011. “Importance of CYP1A1 Polymorphism and Its Transcriptional Regulation in Ovarian and Endometrial Cancer.” *Ginekologia polska* 82(12): 925–32. <http://www.ncbi.nlm.nih.gov/pubmed/22384629>.
- MUELLER, G C, and J A MILLER. 1949. “The Reductive Cleavage of 4-Dimethylaminoazobenzene by Rat Liver; the Intracellular Distribution of the Enzyme System and Its Requirement for Triphosphopyridine Nucleotide.” *The Journal of biological chemistry* 180(3): 1125–36. <http://www.ncbi.nlm.nih.gov/pubmed/18139207>.
- Murphy, Patrick J., and Charles A. West. 1969. “The Role of Mixed Function Oxidases in Kaurene Metabolism in *Echinocystis Macrocarpa* Greene Endosperm.” *Archives of Biochemistry and Biophysics* 133(2): 395–407. <https://linkinghub.elsevier.com/retrieve/pii/0003986169904688>.
- Musshoff, Frank, Ulrike M. Stamer, and Burkhard Madea. 2010. “Pharmacogenetics and Forensic Toxicology.” *Forensic Science International* 203(1–3): 53–62.
<https://linkinghub.elsevier.com/retrieve/pii/S037907381000349X>.
- Naranjo, M E G et al. 2016. “High Frequency of CYP2D6 Ultrarapid Metabolizers in Spain: Controversy about Their Misclassification in Worldwide Population Studies.” *The Pharmacogenomics Journal* 16(5): 485–90. <http://www.nature.com/articles/tpj201647>.
- Narhi, L O, and A J Fulco. 1986. “Characterization of a Catalytically Self-Sufficient 119,000-Dalton Cytochrome P-450 Monooxygenase Induced by Barbiturates in *Bacillus Megaterium*.” *The Journal of biological chemistry* 261(16): 7160–69.
<http://www.ncbi.nlm.nih.gov/pubmed/3086309>.
- Nayeem, Mohammed A., Darryl C. Zeldin, Matthew A. Boegehold, and John R. Falck. 2011. “Salt Modulates Vascular Response through Adenosine A2A Receptor in ENOS-Null Mice: Role of CYP450 Epoxygenase and Soluble Epoxide Hydrolase.” *Molecular and Cellular Biochemistry* 350(1–2): 101–11. <http://link.springer.com/10.1007/s11010-010-0686-0>.
- Nebert, D. 1987. “P450 Genes: Structure, Evolution, And Regulation.” *Annual Review of Biochemistry* 56(1): 945–93.

<http://biochem.annualreviews.org/cgi/doi/10.1146/annurev.biochem.56.1.945>.

- Nebert, Daniel W., Timothy P. Dalton, Allan B. Okey, and Frank J. Gonzalez. 2004. "Role of Aryl Hydrocarbon Receptor-Mediated Induction of the CYP1 Enzymes in Environmental Toxicity and Cancer." *Journal of Biological Chemistry* 279(23): 23847–50. <http://www.jbc.org/lookup/doi/10.1074/jbc.R400004200>.
- Nebert, Daniel W, and David W Russell. 2002. "Clinical Importance of the Cytochromes P450." *Lancet (London, England)* 360(9340): 1155–62. <http://www.sciencedirect.com/science/article/pii/S0140673602112037> (April 27, 2016).
- Nei, M. 1978. "Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals." *Genetics* 89(3): 583–90. <http://www.ncbi.nlm.nih.gov/pubmed/17248844>.
- Nei, Masatoshi. 1974. "A New Measure of Genetic Distance." In *Genetic Distance*, Boston, MA: Springer US, 63–76. http://link.springer.com/10.1007/978-1-4684-2139-2_6.
- Nelson, David R. 1999. "Cytochrome P450 and the Individuality of Species." *Archives of Biochemistry and Biophysics*.
- . 2018. "Cytochrome P450 Diversity in the Tree of Life." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1866(1): 141–54. <https://linkinghub.elsevier.com/retrieve/pii/S1570963917300857>.
- Nelson, David R. 2006. "Cytochrome P450 Nomenclature, 2004." *Methods in molecular biology (Clifton, N.J.)* 320: 1–10. <http://www.ncbi.nlm.nih.gov/pubmed/16719369>.
- Neuhauser, C. 2004. "Mathematical Models in Population Genetics." In *Handbook of Statistical Genetics*, Chichester: John Wiley & Sons, Ltd. <http://doi.wiley.com/10.1002/0470022620.bbc20>.
- Nevo, Eviatar. 2001. "Genetic Diversity." In *Encyclopedia of Biodiversity*, Elsevier, 662–77. <https://linkinghub.elsevier.com/retrieve/pii/B9780123847195000654>.
- Niinuma, Y. et al. 2014. "Functional Characterization of 32 CYP2C9 Allelic Variants." *Pharmacogenomics Journal* 14(2): 107–14.
- Novembre, John et al. 2008. "Genes Mirror Geography within Europe." *Nature* 456(7218): 98–101. <http://www.nature.com/articles/nature07331>.
- Ogawa, Ryuichi, and Hirotochi Echizen. 2010. "Drug-Drug Interaction Profiles of Proton Pump Inhibitors." *Clinical Pharmacokinetics* 49(8): 509–33. <http://link.springer.com/10.2165/11531320-000000000-00000>.
- OMURA, T., and R. SATO. 1962. "A New Cytochrome in Liver Microsomes." *The Journal of biological chemistry* 237: 1375–76. <http://www.ncbi.nlm.nih.gov/pubmed/14482007>.
- OMURA, Tsuneo. 2011. "Recollection of the Early Years of the Research on Cytochrome P450." *Proceedings of the Japan Academy, Series B* 87(10): 617–40. https://www.jstage.jst.go.jp/article/pjab/87/10/87_10_617/_article.
- Orliaguet, G. et al. 2015. "A Case of Respiratory Depression in a Child With Ultrarapid CYP2D6 Metabolism After Tramadol." *PEDIATRICS* 135(3): e753–55. <http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2014-2673>.
- Ortiz de Montellano, Paul R., ed. 2015. "Cytochrome P450." In *Cytochrome P450*, Cham: Springer International Publishing, 3–32. <http://link.springer.com/10.1007/978-3-319-12108-6>.
- Osnabrugge, Ruben L. et al. 2015. "A Systematic Review and Critical Assessment of 11 Discordant

Meta-Analyses on Reduced-Function CYP2C19 Genotype and Risk of Adverse Clinical Outcomes in Clopidogrel Users.” *Genetics in Medicine* 17(1): 3–11.
<http://www.nature.com/articles/gim201476>.

Otyepka, Michal, Josef Skopalík, Eva Anzenbacherová, and Pavel Anzenbacher. 2007. “What Common Structural Features and Variations of Mammalian P450s Are Known to Date?” *Biochimica et Biophysica Acta (BBA) - General Subjects* 1770(3): 376–89.
<https://linkinghub.elsevier.com/retrieve/pii/S0304416506002881>.

Overholser, Brian R, and David R Foster. 2011. “Opioid Pharmacokinetic Drug-Drug Interactions.” *The American journal of managed care* 17 Suppl 1: S276-87.
<http://www.ncbi.nlm.nih.gov/pubmed/21999760>.

P.B. Danielson, Bentham Science Publisher. 2002. “The Cytochrome P450 Superfamily: Biochemistry, Evolution and Drug Metabolism in Humans.” *Current Drug Metabolism* 3(6): 561–97. <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-2002&volume=3&issue=6&spage=561>.

Paré, Guillaume et al. 2010. “Effects of CYP2C19 Genotype on Outcomes of Clopidogrel Treatment.” *New England Journal of Medicine* 363(18): 1704–14.

Pašalić, Daria, and Natalija Marinković. 2017. “Genetic Polymorphisms of the CYP1A1, GSTM1, and GSTT1 Enzymes and Their Influence on Cardiovascular Risk and Lipid Profile in People Who Live near a Natural Gas Plant.” *Archives of Industrial Hygiene and Toxicology* 68(1): 46–52. <http://content.sciendo.com/view/journals/aiht/68/1/article-p46.xml>.

Pearson, Karl. 1901. “LIII. On Lines and Planes of Closest Fit to Systems of Points in Space.” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2(11): 559–72. <https://www.tandfonline.com/doi/full/10.1080/14786440109462720>.

Persson, I., I. Johansson, and M. Ingelman-Sundberg. 1997. “In Vitro Kinetics of Two Human CYP1A1 Variant Enzymes Suggested to Be Associated with Interindividual Differences in Cancer Susceptibility.” *Biochemical and Biophysical Research Communications* 231(1): 227–30. <https://linkinghub.elsevier.com/retrieve/pii/S0006291X97960514>.

Pirmohamed, Munir et al. 2004. “Adverse Drug Reactions as Cause of Admission to Hospital: Prospective Analysis of 18 820 Patients.” *BMJ* 329(7456): 15–19.
<http://www.bmj.com/lookup/doi/10.1136/bmj.329.7456.15>.

Pochapsky, Thomas C., Sophia Kazanis, and Marina Dang. 2010. “Conformational Plasticity and Structure/Function Relationships in Cytochromes P450.” *Antioxidants & Redox Signaling* 13(8): 1273–96. <http://www.liebertpub.com/doi/10.1089/ars.2010.3109>.

Polimanti, Renato, Sara Piacentini, Dario Manfellotto, and Maria Fuciarelli. 2012. “Human Genetic Variation of CYP450 Superfamily: Analysis of Functional Diversity in Worldwide Populations.” *Pharmacogenomics* 13(16): 1951–60.
<https://www.futuremedicine.com/doi/10.2217/pgs.12.163>.

Porrás-Hurtado, Liliana et al. 2013a. “An Overview of STRUCTURE: Applications, Parameter Settings, and Supporting Software.” *Frontiers in Genetics* 4(1): 13–22.
<http://journal.frontiersin.org/article/10.3389/fgene.2013.00098/abstract>.

. 2013b. “An Overview of STRUCTURE: Applications, Parameter Settings, and Supporting Software.” *Frontiers in Genetics* 4.
<http://journal.frontiersin.org/article/10.3389/fgene.2013.00098/abstract>.

Poulos, Thomas L., Barry C. Finzel, and Andrew J. Howard. 1987. “High-Resolution Crystal

- Structure of Cytochrome P450cam.” *Journal of Molecular Biology* 195(3): 687–700.
<https://linkinghub.elsevier.com/retrieve/pii/0022283687901902>.
- Preskorn, Sheldon H. et al. 2013. “Cytochrome P450 2D6 Phenoconversion Is Common in Patients Being Treated for Depression.” *The Journal of Clinical Psychiatry* 74(06): 614–21.
<http://article.psychiatrist.com/?ContentType=START&ID=10008198>.
- Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. 2000. “Inference of Population Structure Using Multilocus Genotype Data.” *Genetics* 155(2): 945–59.
<http://www.ncbi.nlm.nih.gov/pubmed/10835412>.
- Proietti, Massimiliano et al. 2019. “Experimental Test of Local Observer Independence.” *Science Advances* 5(9): eaaw9832.
<https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aaw9832>.
- Putnam, Hilary. 1980. “Models and Reality.” *Journal of Symbolic Logic* 45(3): 464–82.
https://www.cambridge.org/core/product/identifier/S0022481200046466/type/journal_article.
- Radovanović, S., Knežević, A. 2014. *Romi u Srbiji*. Belgrade: Republički zavod za statistiku.
<https://pod2.stat.gov.rs/ObjavljenePublikacije/Popis2011/Romi.pdf>.
- Rajeevan, Haseena et al. 2017. “ALFRED: An Allele Frequency Database for Microevolutionary Studies.” *Evolutionary Bioinformatics* 1(20): 1–10.
- Ramakers, Christian, Jan M Ruijter, Ronald H.Lekanne Deprez, and Antoon F.M Moorman. 2003. “Assumption-Free Analysis of Quantitative Real-Time Polymerase Chain Reaction (PCR) Data.” *Neuroscience Letters* 339(1): 62–66.
<https://linkinghub.elsevier.com/retrieve/pii/S0304394002014234>.
- Ramamoorthy, Anuradha, and Todd C. Skaar. 2011. “In Silico Identification of MicroRNAs Predicted to Regulate the Drug Metabolizing Cytochrome P450 Genes.” *Drug Metabolism Letters* 5(2): 126–31.
<http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1872-3128&volume=5&issue=2&spage=126>.
- Rebbeck, Timothy R. et al. 1998. “Modification of Clinical Presentation of Prostate Tumors by a Novel Genetic Variant in CYP3A4.” *JNCI: Journal of the National Cancer Institute* 90(16): 1225–29. <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/90.16.1225>.
- Relling, M V, and T E Klein. 2011. “CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network.” *Clinical Pharmacology & Therapeutics* 89(3): 464–67. <http://doi.wiley.com/10.1038/clpt.2010.279>.
- Remmer, H., and H. J. Merker. 1963. “Drug-Induced Changes in the Liver Endoplasmic Reticulum: Association with Drug-Metabolizing Enzymes.” *Science* 142(3600): 1657–58.
<http://www.sciencemag.org/cgi/doi/10.1126/science.142.3600.1657>.
- . 1965. “EFFECT OF DRUGS ON THE FORMATION OF SMOOTH ENDOPLASMIC RETICULUM AND DRUG-METABOLIZING ENZYMES.” *Annals of the New York Academy of Sciences* 123(1): 79–97. <http://doi.wiley.com/10.1111/j.1749-6632.1965.tb12247.x>.
- Roberts, Rebecca L. et al. 2004. “No Evidence of Increased Adverse Drug Reactions in Cytochrome P450CYP2D6 Poor Metabolizers Treated with Fluoxetine or Nortriptyline.” *Human Psychopharmacology: Clinical and Experimental* 19(1): 17–23.
<http://doi.wiley.com/10.1002/hup.539>.
- Rockett, Ian R. H. et al. 2015. “Variable Classification of Drug-Intoxication Suicides across US

- States: A Partial Artifact of Forensics?" ed. Wenzhe Ho. *PLOS ONE* 10(8): e0135296. <https://dx.plos.org/10.1371/journal.pone.0135296>.
- Rodríguez-Ramilo, Silvia T, Miguel A Toro, and Jesús Fernández. 2009. "Assessing Population Genetic Structure via the Maximisation of Genetic Distance." *Genetics Selection Evolution* 41(1): 49. <https://gsejournal.biomedcentral.com/articles/10.1186/1297-9686-41-49>.
- Rose, Randy L., and Ernest Hodgson. "Metabolism of Toxicants." In *A Textbook of Modern Toxicology*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 111–48. <http://doi.wiley.com/10.1002/0471646776.ch7>.
- Rosenberg, Noah A. et al. 2005. "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure." *PLoS Genetics*.
- Rosenberg, Noah A., Lei M. Li, Ryk Ward, and Jonathan K. Pritchard. 2003. "Informativeness of Genetic Markers for Inference of Ancestry*." *The American Journal of Human Genetics* 73(6): 1402–22. <https://linkinghub.elsevier.com/retrieve/pii/S0002929707639901>.
- Rotunno, Melissa et al. 2009. "Phase I Metabolic Genes and Risk of Lung Cancer: Multiple Polymorphisms and mRNA Expression" ed. Ulrich Zanger. *PLoS ONE* 4(5): e5652. <https://dx.plos.org/10.1371/journal.pone.0005652>.
- Ruan, Xiulu, Srinivas Chiravuri, and Alan D. Kaye. 2016. "Toxicological Testing When Evaluating Fatal Cases Suspected of Acute Fentanyl Toxicity." *Forensic Science, Medicine, and Pathology* 12(3): 363–65. <http://link.springer.com/10.1007/s12024-016-9789-6>.
- RYAN, K. J., and L. L. ENGEL. 1957. "Hydroxylation of Steroids at Carbon 21." *The Journal of biological chemistry* 225(1): 103–14. <http://www.ncbi.nlm.nih.gov/pubmed/13416221>.
- Sajantila, Antti, and Bruce Budowle. 2016. "Molecular Autopsy." In *Handbook of Forensic Genetics, Security Science and Technology*, WORLD SCIENTIFIC (EUROPE), 377–413. https://doi.org/10.1142/9781786340788_0016.
- Salas, A., C. Phillips, and A. Carracedo. 2006. "Ancestry vs Physical Traits: The Search for Ancestry Informative Markers (AIMs)." *International Journal of Legal Medicine* 120(3): 188–89. <http://link.springer.com/10.1007/s00414-005-0032-3>.
- SALLEE, FLOYD R., C. LINDSAY DeVANE, and ROBERT E. FERRELL. 2000. "Fluoxetine-Related Death in a Child with Cytochrome P-450 2D6 Genetic Deficiency." *Journal of Child and Adolescent Psychopharmacology* 10(1): 27–34. <http://www.liebertpub.com/doi/10.1089/cap.2000.10.27>.
- Samer, C. F. et al. 2013. "Applications of CYP450 Testing in the Clinical Setting." *Molecular Diagnosis & Therapy* 17(3): 165–84. <http://link.springer.com/10.1007/s40291-013-0028-5>.
- Samer, CF et al. 2010. "Genetic Polymorphisms and Drug Interactions Modulating CYP2D6 and CYP3A Activities Have a Major Effect on Oxycodone Analgesic Efficacy and Safety." *British Journal of Pharmacology* 160(4): 919–30. <http://doi.wiley.com/10.1111/j.1476-5381.2010.00709.x>.
- Sanguinetti, Michael C., and John S. Mitcheson. 2005. "Predicting Drug–HERG Channel Interactions That Cause Acquired Long QT Syndrome." *Trends in Pharmacological Sciences* 26(3): 119–24. <https://linkinghub.elsevier.com/retrieve/pii/S0165614705000040>.
- Sankhwar, Monica, and S.N. Sankhwar. 2014. "Variations in CYP Isoforms and Bladder Cancer: A Superfamily Paradigm." *Urologic Oncology: Seminars and Original Investigations* 32(1): 28.e33-28.e40. <https://linkinghub.elsevier.com/retrieve/pii/S1078143912003511>.

- Saraeva, R. B. et al. 2007. "Pharmacogenetics of Acenocoumarol: CYP2C9, CYP2C19, CYP1A2, CYP3A4, CYP3A5 and ABCB1 Gene Polymorphisms and Dose Requirements." *Journal of Clinical Pharmacy and Therapeutics* 32(6): 641–49. <http://doi.wiley.com/10.1111/j.1365-2710.2007.00870.x>.
- Scheffel, U et al. 1998. "In Vivo Detection of Short- and Long-Term MDMA Neurotoxicity--a Positron Emission Tomography Study in the Living Baboon Brain." *Synapse (New York, N.Y.)* 29(2): 183–92. <http://www.ncbi.nlm.nih.gov/pubmed/9593108>.
- Schork, Nicholas J. 2015. "Personalized Medicine: Time for One-Person Trials." *Nature* 520(7549): 609–11. <http://www.nature.com/articles/520609a>.
- Schroth, Werner et al. 2017. "Improved Prediction of Endoxifen Metabolism by CYP2D6 Genotype in Breast Cancer Patients Treated with Tamoxifen." *Frontiers in Pharmacology* 8. <http://journal.frontiersin.org/article/10.3389/fphar.2017.00582/full>.
- Schwarz, U. I. 2003. "Clinical Relevance of Genetic Polymorphisms in the Human CYP2C9 Gene." *European Journal of Clinical Investigation* 33(2): 23–30.
- Scott, S A et al. 2013. "Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C19 Genotype and Clopidogrel Therapy: 2013 Update." *Clinical Pharmacology & Therapeutics* 94(3): 317–23. <http://doi.wiley.com/10.1038/clpt.2013.105>.
- Seeringer, Angela, and Julia Kirchheiner. 2008. "Pharmacogenetics-Guided Dose Modifications of Antidepressants." *Clinics in Laboratory Medicine* 28(4): 619–26. <https://linkinghub.elsevier.com/retrieve/pii/S0272271208000279>.
- Semsarian, C., J. Ingles, and A. A. M. Wilde. 2015. "Sudden Cardiac Death in the Young: The Molecular Autopsy and a Practical Approach to Surviving Relatives." *European Heart Journal* 36(21): 1290–96. <https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehv063>.
- Sezutsu, Hideki, Gaëlle Le Goff, and René Feyereisen. 2013. "Origins of P450 Diversity." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1612): 20120428. <https://royalsocietypublishing.org/doi/10.1098/rstb.2012.0428>.
- Shah, Rashmi R., and Robert L. Smith. 2015. "Addressing Phenoconversion: The Achilles' Heel of Personalized Medicine." *British Journal of Clinical Pharmacology* 79(2): 222–40. <http://doi.wiley.com/10.1111/bcp.12441>.
- Shi, L. 1997. "Local Influence in Principal Components Analysis." *Biometrika* 84(1): 175–86. <https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/84.1.175>.
- Shi, W. et al. 2010. "A Worldwide Survey of Human Male Demographic History Based on Y-SNP and Y-STR Data from the HGDP-CEPH Populations." *Molecular Biology and Evolution* 27(2): 385–93. <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msp243>.
- Sibbing, Dirk et al. 2010. "Isolated and Interactive Impact of Common CYP2C19 Genetic Variants on the Antiplatelet Effect of Chronic Clopidogrel Therapy." *Journal of Thrombosis and Haemostasis* 8(8): 1685–93.
- Šidák, Zbyněk. 1967. "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions." *Journal of the American Statistical Association* 62(318): 626–33. <https://doi.org/10.1080/01621459.1967.10482935>.
- Sim, Sarah C. et al. 2010. "Association between CYP2C19 Polymorphism and Depressive Symptoms." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 9999B: n/a-n/a. <http://doi.wiley.com/10.1002/ajmg.b.31081>.

- Sim, Sarah C, and Magnus Ingelman-Sundberg. 2010. "The Human Cytochrome P450 (CYP) Allele Nomenclature Website: A Peer-Reviewed Database of CYP Variants and Their Associated Effects." *Human Genomics* 4(4): 1–4. <http://dx.doi.org/10.1186/1479-7364-4-4-278>.
- Skopp, Gisela et al. 1998. "Postmortem Distribution of Dihydrocodeine and Metabolites in a Fatal Case of Dihydrocodeine Intoxication." *Forensic Science International* 95(2): 99–107. <https://linkinghub.elsevier.com/retrieve/pii/S0379073898000917>.
- Sousa, V C et al. 2012. "Population Divergence with or without Admixture: Selecting Models Using an ABC Approach." *Heredity* 108(5): 521–30. <http://www.nature.com/articles/hdy2011116>.
- Šrejber, Martin et al. 2018. "Membrane-Attached Mammalian Cytochromes P450: An Overview of the Membrane's Effects on Structure, Drug Binding, and Interactions with Redox Partners." *Journal of Inorganic Biochemistry* 183: 117–36. <https://linkinghub.elsevier.com/retrieve/pii/S0162013417308413>.
- Stanley, L.A. 2017. "Drug Metabolism." In *Pharmacognosy*, Elsevier, 527–45.
- Steimer, Werner et al. 2004. "Allele-Specific Change of Concentration and Functional Gene Dose for the Prediction of Steady-State Serum Concentrations of Amitriptyline and Nortriptyline in CYP2C19 and CYP2D6 Extensive and Intermediate Metabolizers." *Clinical Chemistry* 50(9): 1623–33. <https://academic.oup.com/clinchem/article/50/9/1623/5640158>.
- Suarez-Kurtz, Guilherme. 2005. "Pharmacogenetics, Pharmacogenomics and Population Admixture: Implications for Drug Development and Prescription." *Nature Reviews Genetics* 6(10): 721–721. <http://www.nature.com/articles/nrg1559-c1>.
- Sul, Jae Hoon, Lana S. Martin, and Eleazar Eskin. 2018. "Population Structure in Genetic Studies: Confounding Factors and Mixed Models" ed. Gregory S. Barsh. *PLOS Genetics* 14(12): e1007309. <https://dx.plos.org/10.1371/journal.pgen.1007309>.
- Sun, Sophie, Joan H. Schiller, and Adi F. Gazdar. 2007. "Lung Cancer in Never Smokers — a Different Disease." *Nature Reviews Cancer* 7(10): 778–90. <http://www.nature.com/articles/nrc2190>.
- Sundhaussen, Holm. 2007. *Geschichte Serbiens Geschichte Serbiens*. Wien: Böhlau Verlag. <https://www.vr-elibrary.de/doi/book/10.7767/boehlau.9783205117650>.
- Swen, J J et al. 2011. "Pharmacogenetics: From Bench to Byte— An Update of Guidelines." *Clinical Pharmacology & Therapeutics* 89(5): 662–73. <http://doi.wiley.com/10.1038/clpt.2011.34>.
- Szanyi, István et al. 2012. "Effects of Cytochrome P450 1A1 and Uridine-Diphosphate-Glucuronosyltransferase 1A1 Allelic Polymorphisms on the Risk of Development and the Prognosis of Head and Neck Cancers." *European Journal of Cancer Prevention* 21(6): 560–68. <https://insights.ovid.com/crossref?an=00008469-201211000-00009>.
- Takanashi, K et al. 2000. "CYP2C9 Ile359 and Leu359 Variants: Enzyme Kinetic Study with Seven Substrates." *Pharmacogenetics* 10(2): 95–104. <http://www.ncbi.nlm.nih.gov/pubmed/10761997>.
- Tan, P N, Michael Steinbach, and Vipin Kumar. 2005. "Cluster Analysis: Basic Concepts and Algorithms." *Introduction to Data Mining*: 487–568.
- Tàssies, Dolores et al. 2002. "Pharmacogenetics of Acenocoumarol: Cytochrome P450 CYP2C9 Polymorphisms Influence Dose Requirements and Stability of Anticoagulation." *Haematologica* 87(11): 1185–91. <http://www.ncbi.nlm.nih.gov/pubmed/12414349>.

- Tekpli, Xavier et al. 2012. "DNA Methylation of the CYP1A1 Enhancer Is Associated with Smoking-Induced Genetic Alterations in Human Lung." *International Journal of Cancer* 131(7): 1509–16. <http://doi.wiley.com/10.1002/ijc.27421>.
- Teo, Yik Y. 2008. "Common Statistical Issues in Genome-Wide Association Studies: A Review on Power, Data Quality Control, Genotype Calling and Population Structure." *Current Opinion in Lipidology* 19(2): 133–43. <http://journals.lww.com/00041433-200804000-00007>.
- Thomas, Duncan C, and John S Witte. 2002. "Point: Population Stratification: A Problem for Case-Control Studies of Candidate-Gene Associations?" *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 11(6): 505–12. <http://www.ncbi.nlm.nih.gov/pubmed/12050090>.
- Tsai, Hui-Ju et al. 2014. "Assessment of CYP450 Genetic Variability Effect on Methadone Dose and Tolerance." *Pharmacogenomics* 15(7): 977–86. <https://www.futuremedicine.com/doi/10.2217/pgs.14.19>.
- Turillazzi, Emanuela et al. 2010. "MDMA Toxicity and Pathological Consequences: A Review About Experimental Data and Autopsy Findings." *Current Pharmaceutical Biotechnology* 11(5): 500–509. <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-2010&volume=11&issue=5&spage=500>.
- Ueda, Rika et al. 2006. "A Common Regulatory Region Functions Bidirectionally in Transcriptional Activation of the Human CYP1A1 and CYP1A2 Genes." *Molecular Pharmacology* 69(6): 1924–30. <http://molpharm.aspetjournals.org/lookup/doi/10.1124/mol.105.021220>.
- Uno, Shigeyuki et al. 2004. "Oral Exposure to Benzo[a]Pyrene in the Mouse: Detoxication by Inducible Cytochrome P450 Is More Important Than Metabolic Activation." *Molecular Pharmacology* 65(5): 1225–37. <http://molpharm.aspetjournals.org/lookup/doi/10.1124/mol.65.5.1225>.
- Volz, Erik M et al. 2020. "Identification of Hidden Population Structure in Time-Scaled Phylogenies." *Systematic Biology*. <https://academic.oup.com/sysbio/advance-article/doi/10.1093/sysbio/syaa009/5734655>.
- Vukovic, Darko, Ana Jovanovic, Nebojsa Zakic, and Svetlana Vukotic. 2011. "Nomenclature of Statistical Territorial Units: Possibilities of Application in Serbia." *Journal of the Geographical Institute Jovan Cviji?, SASA* 61(2): 11–24. <http://www.doiserbia.nb.rs/Article.aspx?ID=0350-75991102011V>.
- Vyhlidal, C. A., C. Bi, S. Q. Ye, and J. S. Leeder. 2016. "Dynamics of Cytosine Methylation in the Proximal Promoters of CYP3A4 and CYP3A7 in Pediatric and Prenatal Livers." *Drug Metabolism and Disposition* 44(7): 1020–26. <http://dmd.aspetjournals.org/cgi/doi/10.1124/dmd.115.068726>.
- Wahlang, B., K. Cameron Falkner, Matt C. Cave, and Russell A. Prough. 2015. "Role of Cytochrome P450 Monooxygenase in Carcinogen and Chemotherapeutic Drug Metabolism." In , 1–33. <https://linkinghub.elsevier.com/retrieve/pii/S1054358915000356>.
- WAHLUND, STEN. 2010. "ZUSAMMENSETZUNG VON POPULATIONEN UND KORRELATIONSERSCHEINUNGEN VOM STANDPUNKT DER VERERBUNGSLEHRE AUS BETRACHTET." *Hereditas* 11(1): 65–106. <http://doi.wiley.com/10.1111/j.1601-5223.1928.tb02483.x>.
- Wang, Junjun et al. 2017. "Genetic Predisposition to Lung Cancer: Comprehensive Literature

- Integration, Meta-Analysis, and Multiple Evidence Assessment of Candidate-Gene Association Studies.” *Scientific Reports* 7(1): 8371. <http://www.nature.com/articles/s41598-017-07737-0>.
- Wang, Zhican et al. 2013. “Enhancement of Hepatic 4-Hydroxylation of 25-Hydroxyvitamin D 3 through CYP3A4 Induction in Vitro and in Vivo: Implications for Drug-Induced Osteomalacia.” *Journal of Bone and Mineral Research* 28(5): 1101–16. <http://doi.wiley.com/10.1002/jbmr.1839>.
- Weber, Agnes et al. 2015. “Increased Prevalence of Functional Minor Allele Variants of Drug Metabolizing CYP2B6 and CYP2D6 Genes in Roma Population Samples.” *Pharmacological Reports* 67(3): 460–64. <https://linkinghub.elsevier.com/retrieve/pii/S1734114014003405>.
- Weir, B. S., and C. Clark Cockerham. 1984. “ESTIMATING F -STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE.” *Evolution* 38(6): 1358–70. <http://doi.wiley.com/10.1111/j.1558-5646.1984.tb05657.x>.
- Welter, Danielle et al. 2014. “The NHGRI GWAS Catalog, a Curated Resource of SNP-Trait Associations.” *Nucleic Acids Research* 42(D1): D1001–6. <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt1229>.
- Wendt, F. R. et al. 2017. “Global Genetic Variation of Select Opiate Metabolism Genes in Self-Reported Healthy Individuals.” *Pharmacogenomics Journal*.
- Wendt, Frank R. et al. 2018. “Full-Gene Haplotypes Refine CYP2D6 Metabolizer Phenotype Inferences.” *International Journal of Legal Medicine* 132(4): 1007–24. <http://link.springer.com/10.1007/s00414-017-1709-0>.
- Werk, A. N. et al. 2014. “Identification and Characterization of a Defective CYP3A4 Genotype in a Kidney Transplant Patient with Severely Diminished Tacrolimus Clearance.” *Clinical Pharmacology and Therapeutics*.
- Werk, A. N., and I. Cascorbi. 2014. “Functional Gene Variants of CYP3A4.” *Clinical Pharmacology & Therapeutics* 96(3): 340–48. <http://doi.wiley.com/10.1038/clpt.2014.129>.
- Wester, Karin et al. 2008. “Incidence of Fatal Adverse Drug Reactions: A Population Based Study.” *British Journal of Clinical Pharmacology* 65(4): 573–79. <http://doi.wiley.com/10.1111/j.1365-2125.2007.03064.x>.
- WESTLINDJOHANSSON, A et al. 2006. “Identification and Characterization of CYP3A4*20, a Novel Rare CYP3A4 Allele without Functional Activity.” *Clinical Pharmacology & Therapeutics* 79(4): 339–49. <http://doi.wiley.com/10.1016/j.clpt.2005.11.015>.
- Whirl-Carrillo, M et al. 2012. “Pharmacogenomics Knowledge for Personalized Medicine.” *Clinical pharmacology and therapeutics* 92(4): 414–17. <http://www.ncbi.nlm.nih.gov/pubmed/22992668>.
- Wijsman, E M, and L L Cavalli-Sforza. 1984. “Migration and Genetic Population Structure with Special Reference to Humans.” *Annual Review of Ecology and Systematics*.
- Wilkening, Stefan, Bowang Chen, Kari Hemminki, and Asta Forsti. 2006. “STR Markers for Kinship Analysis.” *Human Biology* 78(1): 1–8. http://muse.jhu.edu/content/crossref/journals/human_biology/v078/78.1wilkening.html.
- William M. Bolstad, James M. Curran. 2016. *Introduction to Bayesian Statistics*. 3th ed. New Jersey: Wiley.
- Wright, Rhonda L. et al. 1996. “Cloning of a Potential Cytochrome P450 from the Archaeon *Sulfolobus Solfataricus*.” *FEBS Letters* 384(3): 235–39. <http://doi.wiley.com/10.1016/0014->

5793%2896%2900322-5.

- Wright, S. 1950. "Genetical Structure of Populations." *Nature* 166(4215): 247–49.
<http://www.ncbi.nlm.nih.gov/pubmed/15439261>.
- Wright, Sewall. 1965. "THE INTERPRETATION OF POPULATION STRUCTURE BY F-STATISTICS WITH SPECIAL REGARD TO SYSTEMS OF MATING." *Evolution* 19(3): 395–420. <http://doi.wiley.com/10.1111/j.1558-5646.1965.tb01731.x>.
- Yasukochi, Yoshiki, and Yoko Satta. 2011. "Evolution of the CYP2D Gene Cluster in Humans and Four Non-Human Primates." *Genes & Genetic Systems* 86(2): 109–16.
<http://joi.jlc.jst.go.jp/JST.JSTAGE/ggs/86.109?from=CrossRef>.
- Yilmaz, Fevzi et al. 2013. "Massive Ingestion of Clopidogrel in a Suicidal Attempt: A Case Report and Review of Literature." *Journal of Academic Emergency Medicine*.
- Young, Chi-Hsien et al. 2010. "CYP1A1 Gene Polymorphisms as a Risk Factor for Pterygium." *Molecular vision* 16: 1054–58. <http://www.ncbi.nlm.nih.gov/pubmed/20596254>.
- Zackrisson, A L, Bertil Lindblom, and Johan Ahlner. 2010. "High Frequency of Occurrence of CYP2D6 Gene Duplication/Multiduplication Indicating Ultrarapid Metabolism Among Suicide Cases." *Clinical Pharmacology & Therapeutics* 88(3): 354–59.
<http://doi.wiley.com/10.1038/clpt.2009.216>.
- Zanger, Ulrich M. et al. 2005. "Impact of Genetic Polymorphism in Relation to Other Factors on Expression and Function of Human Drug-Metabolizing P450s." *Toxicology Mechanisms and Methods* 15(2): 121–24. <http://www.tandfonline.com/doi/full/10.1080/15376520590918847>.
- Zanger, Ulrich M, and Matthias Schwab. 2013. "Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation." *Pharmacology & Therapeutics* 138(1): 103–41.
<http://www.sciencedirect.com/science/article/pii/S0163725813000065>.
- Zhang, Z Y et al. 1996. "Characterization of Purified Human Recombinant Cytochrome P4501A1-Ile462 and -Val462: Assessment of a Role for the Rare Allele in Carcinogenesis." *Cancer research* 56(17): 3926–33. <http://www.ncbi.nlm.nih.gov/pubmed/8752159>.
- Zhao, Bin, and Michael R. Waterman. 2011. "Moonlighting Cytochrome P450 Monooxygenases." *IUBMB Life* 63(7): 473–77. <http://doi.wiley.com/10.1002/iub.501>.
- Zhou, Y., M. Ingelman-Sundberg, and V. M. Lauschke. 2017. "Worldwide Distribution of Cytochrome P450 Alleles: A Meta-Analysis of Population-Scale Sequencing Projects." *Clinical Pharmacology and Therapeutics* 102(4): 688–700.

LISTA SKRAĆENICA

DNK	Dezoksiribonukleinska kiselina
PCR	Lančana reakcija polimeraze
RT-PCR	Lančana reakcija polimeraze u realnom vremenu
NSAID	Nesteroidni antiinflamatorni lekovi
STR	kratki uzastopni ponovci
NRL	Neželjena reakcija na lek

Biografija

Ivan Skadrić je rođen 15. februara 1985. godine u Beogradu gde je završio osnovnu i srednju školu. Biološki fakultet Univerziteta u Beogradu je upisao 2007. godine, na kojem je 2012. godine stekao diplomu osnovnih akademskih studija na smeru molekularne biologije i fiziologije. Diplomom master studija smera humane molekularne biologije je stekao na Biološkom fakultetu Univerziteta u Beogradu 2013. godine, odbranom master rada pod nazivom "Populaciono genetička studija inserciono-delecionih polimorfizama u lokalnim zajednicama ljudi u Srbiji i regionu". Doktorske studije smera Molekularna medicina je upisao 2014. godine na Medicinskom fakultetu Univerziteta u Beogradu. Od 2014. godine učestvuje u naučno-istraživačkom projektu "Analiza genskih polimorfizama CIP izoenzima u populaciji Srbije", finansiranom od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije, koji se realizuje u DNK laboratoriji Instituta za sudsku medicinu, Medicinskog fakulteta u Beogradu. 2016. godine je odlukom Naučnog veća Medicinskog fakulteta u Beogradu, izabran u zvanje istraživača pripravnika. Autor je dve naučne publikacije, "Prediction of autosomal STR typing success in ancient and Second World War bone samples" i "Defining screening panel of functional variants of CYP1A1, CYP2C9, CYP2C19, CYP2D6, and CYP3A4 genes in Serbian population", objavljenih u međunarodnim časopisima M21 kategorije. Rezultate svojih istraživanja je oralno prezentovao na tri naučna kongresa, 2015. godine pod naslovom "Genotyping Data of 30 Insertion-deletion Polymorphisms in Serbian Population" na Balkanskom kongresu humane genetike, 2017. godine pod naslovom "Analysis of Cytochrome P450 2D6 Functional Variants in Serbian Human Population" na I Kongresu molekularnih biologa Srbije, 2019. godine pod naslovom "CYP450 Functional Variants Panel for Postmortem Genetic Screening in Serbian Population" na VI Kongresu Društva genetičara Srbije.

образац изјаве о ауторству

Изјава о ауторству

Име и презиме аутора: Иван Скадрић
Број индекса: ММ08/14

Изјављујем

да је докторска дисертација под насловом :“Карактеристике дистрибуције функционалних варијанти гена из суперфамилије цитохрома П450 у становништву Србије”

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 06.07.2020. године

образац изјаве о истоветности штампане и електронске верзије докторског рада

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора: Иван Скадрић

Број индекса: ММ08/14

Студијски програм: молекуларна медицина

Наслов рада :“Карактеристике дистрибуције функционалних варијанти гена из суперфамилије цитохрома П450 у становништву Србије”

Ментор: проф. др Оливер Стојковић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, 06.07.2020. године

образац изјаве о коришћењу

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

:“Карактеристике дистрибуције функционалних варијанти гена из суперфамилије цитохрома П450 у становништву Србије” која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.

Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 06.07.2020. године

образац изјаве о ауторству

Изјава о ауторству

Име и презиме аутора: Иван Скадрић
Број индекса: ММ08/14

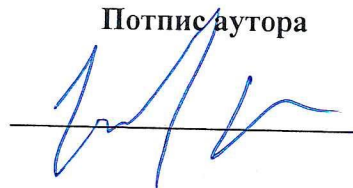
Изјављујем

да је докторска дисертација под насловом :“Карактеристике дистрибуције функционалних варијанти гена из суперфамилије цитохрома П450 у становништву Србије”

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

У Београду, 06.07.2020. године

Потпис аутора



образац изјаве о истоветности штампане и електронске верзије докторског рада

**Изјава о истоветности штампане и електронске верзије докторског
рада**

Име и презиме аутора: Иван Скадрић

Број индекса: ММ08/14

Студијски програм: молекуларна медицина

Наслов рада :“Карактеристике дистрибуције функционалних варијанти гена из суперфамилије цитохрома П450 у становништву Србије”

Ментор: проф. др Оливер Стојковић

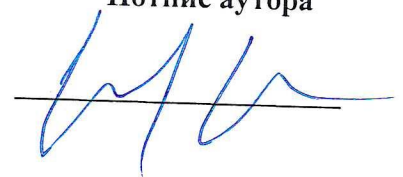
Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

У Београду, 06.07.2020. године

Потпис аутора

A handwritten signature in blue ink, appearing to be 'Ivan Skadrić', written over a horizontal line.

образац изјаве о коришћењу

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

：“Карактеристике дистрибуције функционалних варијанти гена из суперфамилије цитохрома П450 у становништву Србије”

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање. Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)
 2. Ауторство – некомерцијално (CC BY-NC)
 3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
 4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
 5. Ауторство – без прерада (CC BY-ND)
 6. Ауторство – делити под истим условима (CC BY-SA)
- (Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

У Београду, 06.07.2020. године

Потпис аутора

