

Izvirni znanstveni članek ■

Kontrola kvalitete DNA mikromrež nizke gostote

Quality Control of Low-Density DNA Microarrays

Institucije avtorjev: Medicinska fakulteta (RD, PJ, TR, DR),
Fakulteta za računalništvo in informatiko (PJ), Univerza v
Ljubljani.

Kontaktna oseba: Peter Juvan, Center za funkcijsko genomiko
in bio-čipe, Inštitut za biokemijo, Medicinska fakulteta,
Univerza v Ljubljani, Zaloška cesta 4, SI-1000 Ljubljana.
email: peter.juvan@fri.uni-lj.si.

**Rok Devjak, Tadeja Režen,
Damjana Rozman, Peter Juvan**

Izvleček. Tehnologija DNA mikromrež, ki veliko obeta na področju klinične diagnostike in prognostike, se sooča s problemom slabo izoblikovane in neenotne sheme za kontrolo kvalitete meritev. Razvili smo splošen pristop za kontrolo kvalitete DNA mikromrež nizke gostote, ki je osnovan na analizi kovariance kontrolnih sond. Pristop smo aplicirali na mikromreži Steroltalk, razviti z namenom preučevanja izražanja genov, povezanih s homeostazo holesterola. Na modelu primarnih človeških hepatocitov smo analizirali spremembe v izražanju genov pri aplikaciji rifampicina in rosuvastatina in pristop ovrednotili s primerjavo rezultatov analize z navedbami iz literature. Predlagan pristop je neodvisen od vrste uporabljene DNA mikromreže in kot tak lahko pomembno prispeva k hitrejšemu prenosu te tehnologije v klinično prakso.

Abstract. DNA microarray technology, showing great perspective in clinical diagnostics and prognostics, is facing a problem of loosely defined and non-unified data quality control. We have developed a general approach for controlling the quality of low density DNA microarrays, which is based on the analysis of covariance of control probes. We have applied the approach to the Steroltalk microarray, which was developed to study genes in connection with cholesterol homeostasis. We analyzed changes in genes expression in human primary hepatocytes treated with rifampicin and rosuvastatin and evaluated the approach by comparing our results to these from the literature. The proposed approach is independent of the type of DNA microarray and therefore may contribute to more rapid transfer of this technology to the clinical practice.

■ **Infor Med Slov:** 2007; 12(2): 1-13

Uvod

Tehnologija DNA mikromrež je v zadnjem času močno posegla na področje medicine in diagnostike. Ena izmed možnih oblik uporabe te tehnologije so mikromreže nizke gostote za ekspresijsko profiliranje. Te vsebujejo specifičen izbor genov iz celotnega genoma in so kot take namenjene proučevanju posameznih bioloških procesov. Z namenom proučevanja izražanja genov, povezanih s homeostazo holesterola in presnovo zdravil, smo v okviru evropskega projekta 6. okvirnega programa projekta Steroltalk razvili cDNA mikromrežo nizke gostote, ki vsebuje okoli 300 sond za človeške gene in njim sorodne gene pri miši. Mikromreža poleg homeostaze holesterola omogoča tudi študije homeostaze glukoze, maščobnih kislin, cirkadičnega ritma in vnetnih procesov, ter študije razvoja ateroskleroze pri miši in človeku.^{1,2}

Z uporabo tehnologije DNA mikromrež v enem samem poskusu opravimo veliko število meritev, zaradi česar je kontrola kvalitete izvedbe poskusa in pridobljenih podatkov velikega pomena. Kljub temu, da je ponovljivost poskusa z mikromrežami relativno visoka,³ v trenutnem stanju razvoja tehnologija še ni dovolj robustna, da bi jo lahko rutinsko uporabljali v klinični praksi. Pogosto namreč prihaja do primera, da na podlagi določenih meritev pridemo do različnih zaključkov,⁴ kar v klinični praksi ni dopustno. Enega večjih problemov predstavlja slabo izoblikovana in neenotna shema kontrole kvalitete podatkov. Določitev enotne in od platforme neodvisne kontrole kvalitete podatkov je nujnega pomena za razvoj diagnostičnih čipov in posledičnega prenosa tehnologije v klinično prakso, saj se bodo lahko le na ta način oblikovali diagnostični testi z dovolj visoko občutljivostjo in specifičnostjo za njihovo praktično uporabo.

Skupina External RNA Controls Consortium (ERCC) in projekt MicroArray Quality Control (MAQC) sta trenutno najbolj aktivna na področju raziskovanja kontrole kvalitete DNA mikromrež s pomočjo visoko kalibrirane referenčne RNA.⁵

Njun namen je vzpostaviti osnovne standarde kontrole kvalitete s pomočjo kontrolnih sond, kar bi pripomoglo k boljši primerljivosti podatkov, tako med različnimi poskusi kot tudi med različnimi platformami mikromrež, kar bi posledično prispevalo k hitrejši implementaciji tehnologije v klinični praksi.

Kvaliteto izvedbe poskusa z mikromrežo kontroliramo na različnih ravneh: pri izvedbi poskusa z merjenjem kvalitete RNA in uspešnosti integracije barvil, pri odčitavanju z laserskim čitalcem (saturacija, intenziteta ozadja), in pri analizi podatkov (intenziteta pozitivnih in negativnih kontrolnih sond, razmerje med intenziteto signala in ozadja, intenziteta vzdrževalnih (angl. housekeeping) genov in razmerja intenzitet različnih sond za isto tarčo). Pri analizi podatkov na podlagi določenih meril za kontrolo kvalitete izločimo tiste meritve, ki bi lahko privedle do napačnih zaključkov. Merila so pogosto specifična za uporabljeno platformo in med platformami medsebojno neprimerljiva.

Kontroli kvalitete sledi normalizacija podatkov, s katero zagotovimo medsebojno primerljivost opravljenih meritev. Pri tem se v splošnem uporablja dve strategiji. Pri mikromrežah visoke gostote, kjer lahko predpostavimo, da je skupna količina mRNA v vseh vzorcih enaka (ne glede na tretma), za izračun normalizacijske konstante (oz. krivulje) uporabimo izmerjene intenzitete vseh genov na mikromreži. Te predpostavke pri mikromrežah nizke gostote ne moremo narediti; zaradi tega za izračun normalizacijske vrednosti uporabimo normalizacijske kontrolne sonde, na katere hibridizira tuja RNA, ki jo v naprej določeni količini dodamo k RNA preiskovanih vzorcev.

Normalizacijske sonde lahko uporabimo tudi za kontrolo kvalitete podatkov DNA mikromrež. V prispevku predstavljamo inovativen pristop, ki je osnovan na analizi kovariance intenzitete normalizacijskih sond. Razvili smo ga z namenom kontrole kvalitete mikromrež Steroltalk in aplicirali na podatkih iz študije učinka rifampicina

in rosuvastatina, kjer smo na ta način ovrednotili razlike v izražanju genov.

Metode

Hibridizacija in odčitavanje mikromrež Steroltalk sta bili opravljena v Centru za funkcijsko genomiko in bio-čipe Medicinske fakultete Univerze v Ljubljani. Za normalizacijo podatkov in izračun diferencialne izraženosti smo uporabili programski paket Orange⁶ (Fakulteta za računalništvo in informatiko, Univerza v Ljubljani) Predlagan pristop za kontrolo kvalitete smo izvedli v programu SPSS (SPSS Inc., Chicago, IL, ZDA).

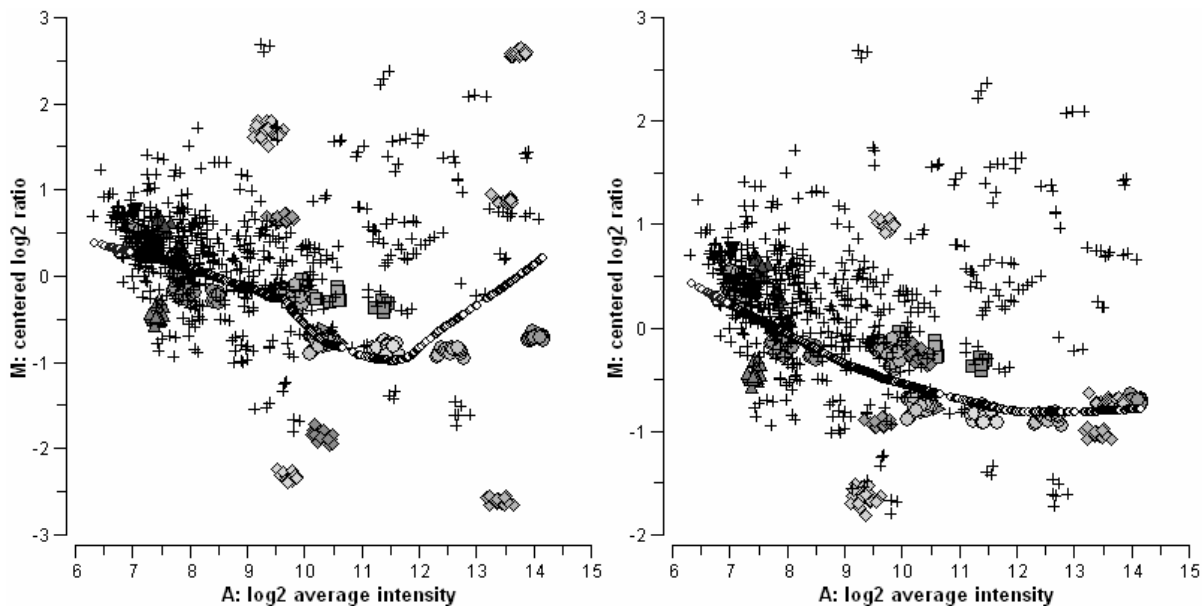
Mikromreža Steroltalk

Mikromreža Steroltalk v2 vsebuje cDNA sonde za 288 človeških genov, 18 vrst normalizacijskih kontrolnih sond kompleta Lucidea Universal ScoreCard (Amersham Biosciences), 7 vrst pozitivnih in 5 vrst negativnih kontrolnih sond. Geni so na mikromreži naneseni v treh in kontrolne sonde v šestih ponovitvah. Mikromreža omogoča hibridizacijo dveh vzorcev hkrati, ki ju predhodno označimo z različnima fluorescentnima barviloma, npr. cianin-3 (Cy3) in cianin-5 (Cy5). Negativne in pozitivne kontrolne sonde uporabljamo za kontrolo izvedbe poskusa, normalizacijske pa za določitev izhodiščnega razmerja med signaloma primerjanih vzorcev.

Normalizacija mikromrež nizke gostote temelji na normalizacijskih sondah, ki so na mikromreži nanesene skupaj s sondami genov. Na normalizacijske sonde hibridizira tuja RNA, ki jo dodamo k iz vzorcev izolirani RNA pred začetkom označevanja. Normalizacijska RNA sestoji iz dveh v naprej pripravljenih mešanic RNA prepisov, ki

se razlikujeta v koncentracijah posameznih transkriptov. Na ta način je pričakovano razmerje med signaloma posamezne normalizacijske sonde pogojeno z razmerjem med koncentracijama RNA, ki se hibridizira na to sondo. V procesu normalizacije izmerjeno razmerje med signaloma normalizacijskih sond popravimo na pričakovano razmerje, za enak faktor pa popravimo tudi razmerja signalov genov. Najpreprostejši način normalizacije je, da za vse gene uporabimo enak normalizacijski faktor. Pogosteje se uporablja normalizacija v odvisnosti od logaritma povprečne intenzitete signalov $A_i = \log_2 \sqrt{R_i * G_i}$, kjer R_i in G_i predstavljata intenziteti signalov sonde i . Normalizacijsko funkcijo v odvisnosti od A ponavadi določimo neparametrično z uporabo metode lokalno utežene regresije (LOWESS)⁷ skozi normalizacijske sonde.

Slika 1 prikazuje primer MA grafa mikromreže Steroltalk in LOWESS krivuljo skozi normalizacijske sonde; na ordinatni osi je prikazan logaritem razmerja intenzitet signalov posameznih sond ($M_i = \log_2(R_i/G_i)$), na abscisni osi pa logaritem njihove povprečne intenzitete (A). Mikromreža Steroltalk vsebuje 10 vrst normalizacijskih sond, pri katerih je pričakovano razmerje med signaloma 1:1 (Lcal1-Lcal10, označene s krogi), ter 8 vrst normalizacijskih sond s pričakovanimi razmerji med signaloma 1:3, 3:1, 1:10 in 10:1 (Lratio1-Lratio8, označene z rombi). Slednje lahko centriramo glede na pričakovano razmerje med signaloma in na ta način dosežemo boljše prileganje normalizacijske krivulje kontrolam. Slika 1 levo prikazuje normalizacijske sonde in krivuljo pred centriranjem, desno pa po centriranju. S centriranjem dosežemo, da je normalizacijska krivulja gladka tudi pri izbiri ožjega okna (širina okna je bila v obeh primerih enaka) in s tem bolj natančno določa izhodiščno razmerje med signaloma primerjanih vzorcev.



Slika 1. Primer MA grafa iz podatkov mikromreže Steroltalk in LOWESS krivulja nad normalizacijskimi sondami. S križci (+) so označeni geni, s krožci (o) normalizacijske kontrole z razmerjem koncentracij 1:1, z rombi (◊) pa normalizacijske kontrole z razmerjem koncentracij 1:3, 3:1, 1:10 in 10:1. Levo so prikazane normalizacijske sonde pred centriranjem, desno pa po centriranju.

Študija aplikacije rifampicina in rosuvastatina

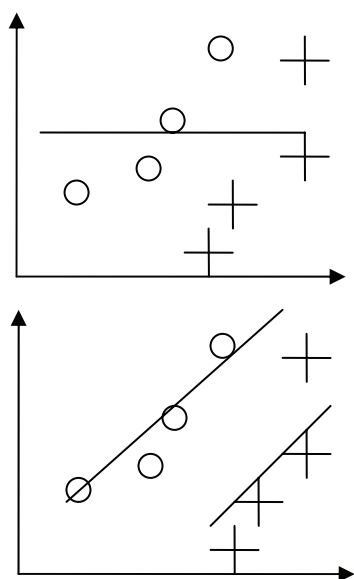
V okviru EU FP6 projekta Steroltalk so bili v laboratorijih K. Monostory, Madžarska akademija znanosti, Madžarska, in P. Maurel, INSERM, Montpellier, Francija, pripravljene človeški primarni hepatociti iz jeter sedmih donorjev. Vzorci so bili tretirani s 5 μ M rifampicinom ali z 10 μ M rosuvastatinom (oba v 0,1% DMSO). Po 48 urah je bila iz celic izolirana celokupna RNA z uporabo reagenta TRIZOL po navodilih proizvajalca (Invitrogen). Za analize transkriptoma smo uporabili 20 mikrogramov celokupne RNA ter jo označili in hibridizirali po protokolih, razvitih za mikromreže Steroltalk.¹ Pred označevanjem smo RNA dodali tuje oz. normalizacijske RNA.

Pristop za kontrolo kvalitete podatkov mikromrež na osnovi normalizacijskih sond

Normalizacijske sonde lahko poleg za izračun normalizacijske krivulje uporabimo tudi za kontrolo kvalitete meritev DNA mikromrež. Z njihovo pomočjo lahko odkrijemo mikromreže

nižje kvalitete in pripadajoče podatke izločimo iz nadaljnje analize. Kontrola kvalitete temelji na analizi kovariance (ANCOVA) normalizacijskih sond, ki jih predhodno centriramo z razmerjem koncentracij, v katerih smo dodali tujo RNA k primerjanima vzorcema.

Model ANCOVA je nadgrajen model analize variance (ANOVA) z dodano spremenljivko, ki kovariira z osnovno spremenljivko. Je mešanica modela ANOVA in linearne regresije. Moč testa ANCOVA je povečana tako, da odstrani tisti del variance, ki ga povzroči kovariiranje osnovne in dodane spremenljivke.⁸ Variabilnosti spremenljivke pri modelu ANOVA izrazimo z vsoto kvadriranih odklonov za tretmaje, pri modelu ANCOVA pa kot vsoto kvadriranih odklonov od regresijske krivulje (Slika 2). Tako izračunana vsota kvadratov je manjša od tiste pri modelu ANOVA.



Slika 2. ANOVA (zgoraj) – ena regresijska krivulja v sredini, ANCOVA (spodaj) – dve regresijski krivulji. Različni tretmaji so ponazorjeni s križci (+) in krožci (○).

Predlagani pristop je osnovan na testiranju ničelne hipoteze (H_0), da med razmerji intenzitet normalizacijskih kontrol, ki jih predhodno centriramo, ne obstajajo statistično značilne razlike, pri čemer upoštevamo tudi vrednost njihove povprečne intenzitete (A). Vrsta kontrole predstavlja fiksni faktor, logaritem razmerja intenzitet signalov (M) neodvisno spremenljivko, logaritem povprečne intenzitete (A) pa sospremenljivko (angl. covariate). Test opravimo po parih; pri 18 vrstah normalizacijskih kontrol, ki so prisotne na mikromreži Steroltalk, torej opravimo 153 testov. Število statistično značilnih rezultatov uporabimo kot mero za kvaliteto podatkov posamezne mikromreže (nižja vrednost pomeni višjo kvaliteto).

S testom primerjamo razlike med istovrstnimi normalizacijskimi sondami (na mikromreži Steroltalk so vse nanese v šestih ponovitvah), ki so posledica tehničnih omejitev pri izdelavi mikromreže in meritvene napake laserskega čitalca (torej faktorjev, na katere tekom poskusa ne moremo vplivati), z razlikami med različnimi vrstami normalizacijskih sond, do katerih lahko

pride v tekom izvedbe poskusa in katere lahko nadzorujemo: uspešnost integracije barvila, natančnost pipetiranja, uspešnost spiranja... Na ta način kontroliramo kvaliteto izvedbe poskusa in posledično kvaliteto izvedenih podatkov. S predlaganim pristopom tudi zagotovimo, da se normalizacijska krivulja dobro prilega podatkom normalizacijskih sond, in hkrati opozorimo na potencialna odstopanja posameznih vrst normalizacijskih sond, katere lahko tekom normalizacije izločimo iz izračuna normalizacijske krivulje.

Evaluacija predlaganega pristopa kontrole kvalitete podatkov mikromrež

Predlagan pristop kontrole kvalitete podatkov DNA mikromrež smo aplicirali na študijah učinka rifampicina in rosuvastatina z uporabo mikromreže Steroltalk v2, katerih namen je bil ovrednotiti razlike v izražanju genov, ki jih povzroči administracija omenjenih dveh učinkovin. Primerjali smo razlike v izražanju genov, do katerih pridemo brez in z uporabo predlaganega pristopa, in jih ovrednotili s stališča poznavanja učinka omenjenih učinkovin.

Mikromreže, ki so bile hibridizirane v omenjenih študijah, smo razvrstili glede na število signifikantnih rezultatov testov ANCOVA pri $\alpha=0,05$ od najslabše ocenjene (najvišje število signifikantnih testov) do najboljše. Postopno smo odstranjevali podatke najslabše ocenjenih mikromrež in iz vsakega nabora podatkov ovrednotili diferencialno izražene gene z uporabo dvostranskega t-testa za neodvisne vzorce pri statistični značilnosti $\alpha=0,05$. Primerjali smo sezname diferencialno izraženih genov iz šestih naborov podatkov: osnovnega nabora, ki je vseboval vse podatke, in petih naborov, pri katerih smo odstranili podatke od ene do pet najslabše ocenjenih mikromrež. Gene smo razvrstili v skupine glede na njihov vzorec pojavljanja v seznamih diferencialne izraženosti in jih ovrednotili na podlagi poznavanja njihovih bioloških funkcij in s pomočjo literature.

Rezultati

Predlagan pristop kontrole kvalitete podatkov DNA mikromrež smo ovrednotili s stališča razlik v izražanju genov pri študiji učinka rifampicina in rosuvastatina. Za vsako študijo smo primerjali šest seznamov diferencialno izraženih genov: nabor (a) je vseboval vse podatke, pri naborih (b)-(f) pa smo odstranili podatke od ene do pet najslabše ocenjenih mikromrež. Glede na vzorec pojavljanja genov v seznamih diferencialne izraženosti smo določili sledeče skupine genov:

1. geni, ki z odstranitvijo najslabše ocenjenih podatkov izginejo s seznama diferencialne izraženosti;
2. geni, ki se z odstranitvijo najslabše ocenjenih podatkov pojavijo na seznamu diferencialne izraženosti;
3. geni, pri katerih se z odstranjevanjem najslabše ocenjenih podatkov diferencialna izraženost spreminja;
4. diferencialno izraženi geni, na katere predlagan pristop ne vpliva.

Gene, pri katerih se z odstranjevanjem najslabše ocenjenih podatkov diferencialna izraženost spreminja (skupina 3), smo na podlagi manj stroge statistične značilnosti $\alpha=0,1$ poskusili prerazporediti v eno izmed drugih skupin. Pri preostalih (neprerazporejenih) genih pa smo primerjali njihove p-vrednosti po posameznih obravnavah in poskusili sklepati o njihovi diferencialni izraženosti.

Poskus na študiji rifampicina

Tabela 1 prikazuje 5 najslabše ocenjenih in 2 najbolj ocenjeni mikromreži pri poskusu z rifampicinom. Postopno smo odstranjevali podatke najslabše ocenjenih mikromrež in ovrednotili diferencialno izražene gene. Tabela 2 prikazuje primerjavo diferencialno izraženih genov, ovrednotenih iz vseh podatkov (a), po odstranitvi podatkov mikromreže Rif48h269 (b), po

odstranitvi podatkov mikromrež Rif48h269 in Rif48h271 (c), itd. Gene smo glede na vzorec njihove diferencialne izraženosti preko posameznih obravnav razvrstili v zgoraj opisane štiri skupine.

Tabela 1. Število statistično značilnih rezultatov analize kovariance med pari normalizacijskih sond pri petih najslabše (zgoraj) in dveh najboljše (spodaj) ocenjenih mikromrežah, hibridiziranih pri študiji učinka rifampicina.

Oznaka mikromreže	Št. stat. značilnih rezultatov ($\alpha=0,05$)
Rif48h269	146
Rif48h271	138
Rif48h270	137
UnRif270	136
Rif48h89	135
...	...
UnRif129	114
UnRif89	102

Tabela 2. Diferencialno izraženi geni pri študiji vpliva rifampicina glede na različne obravnave podatkov (t-test, $\alpha=0,05$). (a) vsi podatki, (b)-(f) odstranjenih 1-5 meritev najnižje kvalitete. S # je označena skupina genov glede na vzorec njihovega pojavljanja v seznamih diferencialne izraženosti. + (-) označujeta statistično značilno povišano (znižano) izražanje genov s seznama.

a b c d e f	Seznam genov	#		
+	CYP39A1			
++	CYP4F8,CYP4F11	1		
++++	GFPT2			
	+	HMGCL,LDHA	2	
	----	CYP8B1		
	--	NR1I2		
	+	APOC1,CEBPG,INSIG2,PPARA		
	++	NFKBIE		
	++	+	DHCR24	3
	++	++	SLCO1B1	
	+++	+	NR1D2	
	+++	+	CYP2A6,CYP2A7,CYP2A13	
	+++	++	NOS1	
	+++++	ALAS1,CEBPA,CYP2C8,CYP2C9, CYP2C18,CYP2C19,CYP3A4,CYP3A5, 4 CYP3A7,CYP3A43,DBP,NR5A2		

Tabela 3 prikazuje p-vrednosti genov, pri katerih se z odstranjevanjem najslabše ocenjenih mikromrež diferencialna izraženost spreminja

(skupina 3). Gene smo z uporabo manj stroge statistične značilnosti $\alpha=0,1$ poskusili prerazporediti v eno izmed drugih skupin.

Tabela 3. Diferencialno izraženi geni pri študiji vpliva rifampicina iz skupine 3 in pripadajoče p-vrednosti po posameznih obravnavah. S # je označena skupina genov, določena glede na vzorec diferencialne izraženosti preko posameznih obravnav pri statistični značilnosti $\alpha=0,1$.

a	b	c	d	e	f	Gen	#
0,250	0,305	0,048	0,028	0,051	0,112	<i>NR1I2</i>	3
0,051	0,022	0,052	0,076	0,061	0,085	<i>APOC1</i>	4
0,101	0,033	0,056	0,111	0,132	0,132	<i>CEBPG</i>	3
0,062	0,030	0,064	0,140	0,155	0,155	<i>INSIG2</i>	1
0,124	0,047	0,102	0,130	0,182	0,182	<i>PPARA</i>	3
0,065	0,018	0,045	0,070	0,116	0,116	<i>NFKBIE</i>	1
0,021	0,036	0,057	0,149	0,109	0,019	<i>DHCR24</i>	3
0,011	0,025	0,062	0,191	0,011	0,011	<i>SLCO1B1</i>	3
0,011	0,022	0,045	0,111	0,097	0,015	<i>NR1D2</i>	3
0,013	0,016	0,035	0,090	0,035	0,078	<i>CYP2A6</i>	4
0,013	0,016	0,036	0,093	0,034	0,085	<i>CYP2A7</i>	4
0,015	0,017	0,037	0,094	0,037	0,081	<i>CYP2A13</i>	4
0,017	0,014	0,036	0,055	0,048	0,048	<i>NOS1</i>	4

Poskus na študiji rosuvastatina

Tabela 4 prikazuje 5 najslabše ocenjenih in 2 najboljše ocenjeni mikromreži pri poskusu z rosuvastatinom. Na enak način kot pri študiji rifampicina smo postopno odstranjevali podatke najslabše ocenjenih mikromrež in ovrednotili diferencialno izražene gene, ki so prikazani v Tabeli 5. Gene smo glede na vzorec njihove diferencialne izraženosti preko posameznih obravnav razvrstili v zgoraj opisane štiri skupine.

Tabela 6 prikazuje p-vrednosti genov, pri katerih se z odstranjevanjem najslabše ocenjenih mikromrež diferencialna izraženost spreminja (skupina 3). Gene smo z uporabo manj stroge statistične značilnosti $\alpha=0,1$ poskusili prerazporediti v eno izmed drugih skupin.

Tabela 4. Število statistično značilnih rezultatov analize kovariance med pari normalizacijskih sond pri petih najslabše (zgoraj) in dveh najboljše (spodaj) ocenjenih mikromrežah, hibridiziranih pri študiji učinka rosuvastatina.

Oznaka mikromreže	Št. stat. značilnih rezultatov ($\alpha=0,05$)
UnRo271	150
Ro48h114	147
UnRo270	147
Ro48h89	142
Ro48h270	141
...	...
UnRo129	96
Ro48h271	90

Tabela 5. Diferencialno izraženi geni pri študiji vpliva rosuvastatina glede na različne obravnave podatkov (t-test, $\alpha=0,05$). (a) vsi podatki, (b)-(f) odstranjenih 1-5 meritev najnižje kvalitete. S # je označena skupina genov glede na vzorec njihovega pojavljanja v seznamih diferencialne izraženosti. + (-) označujeta statistično značilno povešano (znižano) izražanje genov s seznama.

a b c d e f	Seznam genov	#
+++	<i>INSIG1</i>	1
---	<i>SAA2</i>	2
----	<i>ORM1</i>	
+	<i>ACAT2,FDFT1</i>	3
----	<i>C9</i>	
--	<i>APCS</i>	
+++++	<i>ACAS2,FDPS,HMGCS1,ID11,SQLE</i>	4
-----	<i>ABCG5</i>	

Tabela 6. Diferencialno izraženi geni pri študiji vpliva rosuvastatina iz skupine 3 in pripadajoče p-vrednosti po posameznih obravnavah. S # je označena skupina genov, določena glede na vzorec diferencialne izraženosti preko posameznih obravnav pri statistični značilnosti $\alpha=0,1$.

a	b	c	d	e	f	Gen	#
0,063	0,039	0,076	0,149	0,185	0,105	<i>ACAT2</i>	1
0,066	0,044	0,078	0,078	0,130	0,062	<i>FDFT1</i>	3
0,103	0,037	0,048	0,013	0,024	0,051	<i>C9</i>	2
0,039	0,066	0,034	0,030	0,061	0,139	<i>APCS</i>	1

Diskusija

Predlagani pristop postopnega odstranjevanja podatkov mikromrež nižje kvalitete smo s stališča izvedbe ovrednotili na podlagi primerjave skupin genov, ki smo jih določili glede na vzorec njihove diferencialne izraženosti preko posameznih obravnav. Geni, ki se z odstranitvijo najslabše ocenjenih podatkov na novo pojavijo na seznamu diferencialne izraženosti (skupina 2), kažejo na uporabnost predlaganega pristopa, s katerim smo uspeli iz analize diferencialne izraženosti odstraniti tiste meritve, ki so največ prispevali k šumu v podatkih. Taki geni so trije od skupno 32 pri študiji vpliva rifampicina (rif) in 2 od skupno 13 pri študiji vpliva rosuvastatina (ros). Geni, ki z odstranitvijo najslabše ocenjenih podatkov izginejo iz seznama diferencialne izraženosti (skupina 1), kažejo na problem zmanjševanja moči statističnega testa za ugotavljanje diferencialne izraženosti, do katerega pride z odstranjevanjem podatkov posameznih meritev. Taki geni so štirje pri rif in eden pri ros.

Največjo skupino predstavljajo geni, na katere predlagan pristop ne vpliva. Sem spadajo diferencialno izraženi geni iz skupine 4 (12 genov pri rif in 6 pri ros), in hkrati tudi vsi geni, za katere diferencialna izraženost ni bila potrjena v nobeni izmed obravnav.

Geni, pri katerih se z odstranjevanjem najslabše ocenjenih podatkov diferencialna izraženost spreminja (13 genov pri študiji vpliva rifampicina in 4 pri študiji vpliva rosuvastatina), izkazujejo tako uporabnost pristopa kot tudi problem zmanjševanja moči statističnega testa, do katerega pride z odstranjevanjem podatkov posameznih meritev. Za te smo s primerjavo p-vrednosti preko posameznih obravnav pokazali, da jih lahko v večini primerov z uporabno manj stroge statistične značilnosti $\alpha=0,1$ prerazporedimo v eno izmed prej omenjenih skupin. Na ta način smo uspeli prerazporediti 7 od 13 genov pri rif (Tabela 3) in 3 od 4 genov pri ros (Tabela 6). Pri preostalih neprerazporejenih genih smo opazili, da se njihove p-vrednosti preko vseh obravnav gibljejo blizu

mejne vrednosti izbrane statistične značilnosti, zaradi česar je njihove diferencialne izraženosti ne moremo zanesljivo napovedati.

Primerjava p-vrednosti genov iz skupine 3 (Tabeli 3 in 6) pokaže, da z odstranjevanjem podatkov najslabše ocenjenih mikromrež dosežemo tako pozitiven učinek odstranjevanja šuma v podatkih kot tudi negativen učinek zmanjševanja moči statističnega testa za ugotavljanje diferencialne izraženosti (t-test). Pozitiven učinek odstranjevanja šuma je najbolj opazen pri genu *NR112* (Tabela 3) po odstranitvi podatkov dveh najslabše ocenjenih mikromrež (obravnav (c) in pri genu *C9* (Tabela 6) po odstranitvi podatkov najslabše ocenjene mikromreže (obravnav (b)).

Smotrnost uporabe predlaganega pristopa smo ovrednotili tudi s stališča poznanih bioloških funkcij genov, za katere smo v različnih obravnavah lahko potrdili statistično značilno diferencialno izraženost, in s pomočjo rezultatov iz obstoječe literature. Gene, vključene v mikromrežo Steroltalk, smo na podlagi poznavanja njihovih bioloških funkcij razdelili v funkcijske razrede; za obravnavane gene so prikazani v Tabelah 7 in 8 v Prilogi. S pomočjo razredov smo ovrednotili celovitost in smiselnost diferencialne izraženosti posameznih genov. Pri tem smo izhajali iz predpostavke, da geni znotraj posameznih razredov delujejo v skupnih genskih poteh, torej je njihovo izražanje v medsebojni soodvisnosti. Diferencialna izraženost večjega števila genov, ki pripadajo istemu funkcijskemu razredu, na ta način potrjuje dejansko spremembo v genski poti, ki uravnava določene biološke funkcije. Pristop evaluacije je konceptualno podoben pristopu GSEA (angl. Gene Set Enrichment Analysis)⁹ za avtomatsko analizo podatkov o izraženosti večjega števila genov na podlagi njihove razdelitve v razrede. Pri primerjavi z literaturo smo se osredotočili na poskuse, ki so bili biološko podobni našim v smislu opazovanega biološkega sistema (primarni človeški hepatociti) in učinkovin (statini oz. antibiotiki).

Poskus z rifampicinom

Rifampicin je antibiotik, ki se uporablja za zdravljenje tuberkuloze in nekaterih drugih bakterijskih okužb. V znanosti pa se uporablja kot klasičen aktivator človeškega jedrnega receptorja PXR (pregnanski X receptor) in preko njega aktivira metabolizem ksenobiotikov. Pri poskusu z rifampicinom smo na vzorcu primarnih človeških hepatocitov lahko potrdili naslednje diferencialno izražene gene (Tabela 2):

- Encimi faze I presnove ksenobiotikov:

Geni **CYP2A6**, **CYP2A7** in **CYP2A13** imajo na mikromreži Steroltalk skupno sondo, tako da se mRNA enega hibridizira z sondami vseh treh genov, kar pomeni, da bi se morali naštetih geni obnašati enako v vseh obravnavah.

Oscarson in soavtorji¹⁰ navajajo, da se gen **CYP2A6** inducira po aplikaciji ksenobiotika karbamazepina, ki tako kot rifampicin tudi inducira mnoge družine citokromov P450 preko PXR. V obravnavah (d) in (f) diferencialne izraženosti teh treh genov ne moremo potrditi, kar lahko pripišemo zmanjšani moči t-testa, do katerega pride z odstranjevanjem podatkov mikromrež nižje kvalitete.

Geni družine **CYP2C** imajo skupno sondo. Oscarson in soavtorji¹⁰ navajajo, da se morajo geni **CYP2C8**, **CYP2C9**, **CYP2C19** po aplikaciji rifampicina inducirati. Geni so diferencialno izraženi v vseh obravnavah.

Oscarson in soavtorji¹⁰ navajajo, da se morajo geni **CYP3A4** in **CYP3A7** in **CYP3A43** po aplikaciji rifampicina inducirati. Geni **CYP3A4**, **CYP3A5** in **CYP3A7** imajo na mikromreži Steroltalk skupno sondo. Naštetih geni so diferencialno izraženi v vseh obravnavah; na te gene podobno kot na gene družine **CYP2C** odstranjevanje podatkov ne vpliva.

Gena **CYP4F8** in **CYP4F11** se v obravnavah (a) in (b) prikažeta kot diferencialno izražena

zaradi nespecifične hibridizacije gena **CYP4F3**, kar je v skladu z navedbo Oscarsona in soavtorjev.¹⁰ Geni **CYP4F2**, **CYP4F8** in **CYP4F11** imajo na mikromreži Steroltalk skupno sondo. Kljub temu diferencialne izraženosti gena **CYP4F2** ne moremo potrditi v nobeni od obravnav.

Znižano izražanje gena **CYP8B1**, kar je v skladu z navedbo Bhalla in soavtorjev,¹¹ smo potrdili šele po odstranitvi podatkov dveh najslabše ocenjenih mikromrež, torej v obravnavah (c)-(f).

- Jedrni receptorji: gen **NR1I2** (PXR) se po administraciji karbamazepina inducira.¹⁰ V obravnavah (c) in (d), torej po odstranitvi podatkov dveh oziroma treh najslabše ocenjenih mikromrež, smo potrdili njegovo diferencialno izražanje. V nasprotju z navedbo Oscarsona in soavtorjev¹⁰ smo pokazali negativno spremembo v njegovi izraženosti, vendar pri relativno nizki vrednosti M: -0,220 v obravnavi (b) in -0,356 v obravnavi (c).
- Skupina Hem: Gen **ALAS1** se inducira po aplikaciji ksenobiotika,¹² kar lahko potrdimo v vseh obravnavah.
- Transport in presnova žolčnih kislin in ksenobiotikov: Gen **SLCOB1** se po administraciji karbamazepina inducira,¹⁰ kar lahko potrdimo v vseh obravnavah z izjemo obravnav (c) in (d).

Pri poskusu z rifampicinom smo s predlaganim pristopom potrdili diferencialno izraženo devetih genov (**HMGCL**, **LDHA**, **CYP8B1**, **NR1I2**, **APOC1**, **CEBPG**, **INSIG2**, **PPARA** in **NFKBIE**), ki jih v primeru obravnave podatkov vseh mikromrež ne moremo potrditi. Zmožnost potrditve diferencialne izraženosti trinajstih genov iz skupine 3 (**NR1I2**, **APOC1**, **CEBPG**, **INSIG2**, **PPARA**, **NFKBIE**, **DHCR24**, **SLCO1B1**, **NR1D2**, **CYP2A6**, **CYP2A7**, **CYP2A13** in **NOS1**) je odvisna od izbrane obravnave. Zmožnost potrditve diferencialne izraženosti dvanajstih genov iz skupine 4 (**ALAS1**, **CEBPA**, **CYP2C8**, **CYP2C9**,

CYP2C18, CYP2C19, CYP3A4, CYP3A5, CYP3A7, CYP3A43, DBP in *NR5A2*) pa se z aplikacijo predlaganega pristopa ne spreminja.

Na podlagi dobljenih rezultatov pri poskusu z rifampicinom lahko v obravnavah (a) in (b) potrdimo diferencialno izraženost največjega števila genov (16), ki so bili predhodno opisani v literaturi.

Poskus z rosuvastatinom

Rosuvastatin spada med statine, ki se uporabljajo za zdravljenje hiperholestemije in preprečevanje bolezni srca in ožilja. Celostna študija presnove tega statina v človeških jetrih oz. na modelu primarnih človeških hepatocitov predhodno še ni bila opisana. Pri poskusu z rosuvastatinom smo lahko potrdili naslednje diferencialno izražene gene (Tabela 5):

- Biosinteza holesterola in njeno uravnavanje: Iz te skupine smo v vseh obravnavah potrdili diferencialno izraženost genov **ACAS2**, **FDPS**, **HMGCS1**, **IDI1** in **SQLE**, v obravnavi (b) gena **ACAT2** in **FDFT1** ter v obravnavah (a)-(c) gen **INSIG1**. Administracija lovastatina v primarne človeške hepatocite poveča količino aktivne oblike transkripcijskega dejavnika SREBP2 (angl. sterol response element binding protein 2), le-ta pa poveča izražanje vseh zgoraj omenjenih genov.¹³
- Transport holesterola: Gen **ABCG5** se po administraciji statinov poveča izražanje v jetrih miši in podgan ter črevesju človeka.¹⁴⁻¹⁶ V nasprotju z omenjenimi navedbami smo pokazali negativno spremembo v njegovi izraženosti, vendar pri relativno nizki vrednosti M (-0,467 v obravnavi (a)). Slednje lahko pojasnimo z dejstvom, da je bil v našem primeru uporabljen drug statin in drugo tkivo.
- Geni **APCS**, **C9**, **ORM1**, **SAA2** sodelujejo v vnetnem odgovoru, za katerega je bila pokazana inhibicija s statini.¹⁷

Pri poskusu z rosuvastatinom smo s predlaganim pristopom potrdili diferencialno izraženost petih genov (**SAA2**, **ORM1**, **ACAT2**, **FDFT1** in **C9**), ki jih v primeru obravnave podatkov vseh mikromrež ne moremo potrditi. Zmožnost potrditve diferencialne izraženosti štirih genov iz skupine 3 (**ACAT2**, **FDFT1**, **C9** in **APCS**) je odvisna od izbrane obravnave. Zmožnost potrditve diferencialne izraženosti šestih genov iz skupine 4 (**ACAS2**, **FDPS**, **HMGCS1**, **IDI1**, **SQLE** in **ABCG5**) pa se z aplikacijo predlaganega pristopa ne spreminja.

Na podlagi dobljenih rezultatov pri poskusu z rosuvastatinom lahko v obravnavah (b)-(d) potrdimo diferencialno izraženost največjega števila genov (10), ki so bili predhodno opisani v literaturi.

Zaključek

Predlagan pristop je uporaben predvsem pri DNA mikromrežah nizke gostote, pri katerih normalizacija ni osnovana na predpostavki o enakosti količine mRNA v vzorcih, pač pa na večjem številu kontrolnih sond, kamor hibridiziramo visoko-kalibrirano tujo RNA. Pristop je neodvisen od uporabljene vrste DNA mikromreže in kot tak lahko pomembno prispeva k izoblikovanju enotne sheme kontrole kvalitete DNA mikromrež in posledično k hitrejšemu prenosu te tehnologije v klinično prakso.

Pristop izkorišča normalizacijske sonde za oceno variance v meritvah, na katero tekom poskusa ne moremo vplivati, in variance, ki jo tekom poskusa lahko nadzorujemo. S primerjavo obeh komponent varianc lahko zaznamo potencialna odstopanja posameznih vrst normalizacijskih sond, ki privedejo do slabšega prileganja normalizacijske krivulje in posledično manj natančne normalizacije podatkov. S pristopom lahko izpostavimo meritve nižje kvalitete in jih izločimo iz nadaljnje analize.

Pri študijah učinka rifampicina in rosuvastatina na modelu primarnih človeških hepatocitov se je predlagan pristop izkazal kot učinkovit in v smislu fiziologije tudi koristen. Z odstranitvijo podatkov mikromrež nižje kvalitete smo natančneje analizirali spremembe v izražanju genov, kar smo potrdili z navedbami iz literature. Pokazali smo, da je pristop postopnega odstranjevanja podatkov pri analizi diferencialne izraženosti uporaben do meje, kjer izguba na moči statističnega testa preseže prednosti odstranjevanja šuma, ki ga dosežemo z izločitvijo podatkov mikromrež nižje kvalitete iz analize.

Zahvala

Zahvalili bi se dr. Katalin Monostory iz Madžarske akademije znanosti v Budimpešti (HAS) in dr. Jean-Marc Pascussiju z INSERM (Institut national de la santé et de la recherche médicale), Montpellier v Franciji, ki sta pripravila in tretirala človeške primarne hepatocite. Zahvalili bi se tudi dr. Juan A. Contrerasu, ki je opravil del hibridizacij mikromrež Steroltalk. Študiji aplikacije rifampicina in rosuvastatina sta bili financirani iz EU FP6 projekta Steroltalk.

Literatura

1. Režen T: *S ksenobiotiki posredovano uravnavanje lanosterol 14alfa-demetilaze (CYP51) in drugih genov homeostaze holesterola pri miši in človeku*. Ph.D. thesis. Ljubljana, Slovenija 2007: Medicinska fakulteta, Univerza v Ljubljani.
2. Režen T, Contreras JA, Rozman D: Functional Genomics Approaches to Studies of the Cytochrome P450 Superfamily. *Drug Metabolism Reviews* 2007; 39(2): 389-399.
3. Canales RD, Luo Y, Willey JC, et al.: Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* 2006; 24(9): 1115-1122.
4. Tibshirani R: Immune signatures in follicular lymphoma. *N Engl J Med* 2005; 352(14): 1496-1497; author reply 1496-1497.
5. Ji H, Davis RW: Data quality in genomics and microarrays. *Nat Biotechnol* 2006; 24(9): 1112-1113.
6. Demšar J, Zupan B, Grega L: Orange: From Experimental Machine Learning to Interactive Data Mining. White paper. Ljubljana, Slovenia 2004: Faculty of Computer and Information Science, University of Ljubljana.
7. Cleveland WS: Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* 1979; 74(368): 829-836.
8. Gaddis ML: Statistical methodology: IV. Analysis of variance, analysis of covariance, and multivariate analysis of variance. *Acad Emerg Med* 1998; 5(3): 258-265.
9. Subramanian A, Tamayo P, Mootha VK, et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; 102(43): 15545-15550.
10. Oscarson M, Zanger UM, Rifki OF, et al.: Transcriptional profiling of genes induced in the livers of patients treated with carbamazepine. *Clin Pharmacol Ther* 2006; 80(5): 440-456.
11. Bhalla S, Ozalp C, Fang S, et al.: Ligand-activated pregnane X receptor interferes with HNF-4 signaling by targeting a common coactivator PGC-1alpha. Functional implications in hepatic cholesterol and glucose metabolism. *J Biol Chem* 2004; 279(43): 45139-45147.
12. Maglich JM, Stoltz CM, Goodwin B, et al.: Nuclear pregnane x receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification. *Mol Pharmacol* 2002; 62(3): 638-646.
13. Horton JD, Shah NA, Warrington JA, et al.: Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. *Proc Natl Acad Sci U S A* 2003; 100(21): 12027-12032.
14. Kamisako T, Ogawa H: Effects of pravastatin and bezafibrate on biliary lipid excretion and hepatic expression of Abcg5 and Abcg8 in the rat. *J Gastroenterol Hepatol* 2004; 19(8): 879-883.
15. Lally S, Tan CY, Owens D, et al.: Messenger RNA levels of genes involved in dysregulation of postprandial lipoproteins in type 2 diabetes: the role of Niemann-Pick C1-like 1, ATP-binding cassette, transporters G5 and G8, and of microsomal triglyceride transfer protein. *Diabetologia* 2006; 49(5): 1008-1016.
16. Tang W, Ma Y, Yu L: Plasma cholesterol is hyperresponsive to statin in ABCG5/ABCG8

transgenic mice. *Hepatology* 2006; 44(5): 1259-1266.

hypercholesterolaemia: effect of atorvastatin therapy. *J Intern Med* 2005; 257(6): 523-530.

17. Sampietro T, Bigazzi F, Rossi G, et al.:
Upregulation of the immune system in primary

Priloga

Tabela 7. Seznam potencialno diferencialno izraženih genov pri administraciji rifampicina, njihova identifikacijska številka (angl. Accession number), slovensko ime in pripadajoči funkcijski razred.

Gen	Accession	Slovensko ime	Funkcijski razred
ALAS1	NM_003920	Aminolevulinat-sintaza 1	Hem
NFKBIE	BC063609	Jedrni faktor inhibitor ojačevalca gena	Adipocitokin
NOS1	BC051765	Dušikov oksid-sintaza 1, nevronska	Aterogeneza
DHCR24	NM_004820	24-dehidroholesterol-reduktaza	Biosinteza holesterola
HMGCL	BC010570	HMG-CoA-liaza	Biosinteza holesterola
DBP	AF237982	Vezavni protein D mesta promotorja albumina	Cirkadični ritem
NR1D2	NM_006446	Jedrni receptor, 1D2	Cirkadični ritem
GFPT2	NM_000771	Glutamin:fruktoza-6-fosfat transaminaza 2	Glukoza
LDHA	NM_001645	Laktat-dehidrogenaza A	Glukoza
CYP4F11	NM_022820	Citokrom P450, 4F11	Maščobne kisline
CYP4F8	NM_000777	Citokrom P450, 4F8	Maščobne kisline
PPARA	NM_005036	Jedrni receptor peroksisomske proliferacije alfa	Maščobne kisline
CEBPA	BC021116	CCAAT/ojačevalec vezavni protein, alfa	Maščobne kisline, glukoza
CYP2A13	BC001491	Citokrom P450, 2A13	Metabolizem in transport ksenobiotikov
CYP2A6	NM_021187	Citokrom P450, 2A6	Metabolizem in transport ksenobiotikov
CYP2C18	BC005807	Citokrom P450, 2C18	Metabolizem in transport ksenobiotikov
CYP3A4	NM_000104	Citokrom P450, 3A4	Metabolizem in transport ksenobiotikov
CYP3A43	NM_000766	Citokrom P450, 3A43	Metabolizem in transport ksenobiotikov
CYP3A5	AF182275	Citokrom P450, 3A5	Metabolizem in transport ksenobiotikov
CYP3A7	NM_000772	Citokrom P450, 3A7	Metabolizem in transport ksenobiotikov
NR1I2	NM_003889	Pregnan X receptor	Metabolizem in transport ksenobiotikov
CYP39A1	NM_000392	Citokrom P450, 39A1	Metabolizem in transport žolčnih kislin
CYP8B1	NM_004391	Sterol 12 α -hidroksilaza	Metabolizem in transport žolčnih kislin
SLCO1B1	NM_002107	Nosilec organskih anionov, 1B1	Metabolizem in transport žolčnih kislin in ksenobiotikov
NR5A2	X03225	Jedrni receptor, 5A2, jetrni receptor homolog 1	Metabolizem in transport žolčnih kislin, steroidnih hormonov in holesterola
CYP2C19	NM_000778	Citokrom P450, 2C19	Metabolizem ksenobiotikov in eikozanoidov
CYP2C8	NM_004364	Citokrom P450, 2C8	Metabolizem ksenobiotikov in eikozanoidov
CYP2C9	NM_004827	Citokrom P450, 2C9	Metabolizem ksenobiotikov in eikozanoidov
CEBPG	NM_005566	CCAAT/ojačevalec vezavni protein, gama	Neznana funkcija
CYP2A7	NM_007253	Citokrom P450, 2A7	Neznana funkcija
INSIG2	AA197454	Inzulin inducirajoči gen 2	SREBP signalna pot
APOC1	NM_001486	Apolipoprotein C1	Transport holesterola

Tabela 8. Seznam potencialno diferencialno izraženih genov pri administraciji rosuvastatina, njihova identifikacijska številka (angl. Accession number), slovensko ime in pripadajoči funkcijski razred.

Gen	Accession	Slovensko ime	Funkcijski razred
ACAS2	NM_018677	Acil-CoA-sintetaza 2	Biosinteza holesterola
ACAT2	NM_005891	Acetil-CoA-acetiltransferaza 2	Biosinteza holesterola
FDFT1	BC009251	Farnezil-pirofosfat-farneziltransferaza 1	Biosinteza holesterola
FDPS	BC010004	Farnezil-pirofosfat-sintaza	Biosinteza holesterola
HMGCS1	NM_002130	HMG-CoA-sintaza 1	Biosinteza holesterola
IDI1	BC005247	Izopentenil-pirofosfat-delta-izomeraza	Biosinteza holesterola
SQLE	BC017033	Skvalen-epoksidaza	Biosinteza holesterola
APCS	BC007039	P-element serumskega amiloida	Serumski proteini
C9	BC020721	Komplement 9	Serumski proteini
ORM1	BC026238	Orozomukoid 1	Serumski proteini
SAA2	BC020795.1	Serumski amiloid A2	Serumski proteini
INSIG1	BC001880	Z inzulinom inducirani gen 1	SREBP signalna pot
ABCG5	NM_022436	ABC prenašalec, G5	Transport holesterola