



UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA
ŠTUDIJ MOLEKULSKE BIOLOGIJE

Špela DERŽIČ

**IDENTIFIKACIJA ZELO OHRANJENIH REGIJ
GENOMA (UCR) TER NJIHOVIH FUNKCIJSKIH
POLIMORFIZMOV PRI MODELNIH ORGANIZMIH**

MAGISTRSKO DELO

Magistrski študij - 2. stopnja Molekulska biologija

Ljubljana, 2013

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA
ŠTUDIJ MOLEKULSKE BIOLOGIJE

Špela DERŽIČ

**IDENTIFIKACIJA ZELO OHRANJENIH REGIJ GENOMA (UCR)
TER NJIHOVIH FUNKCIJSKIH POLIMORFIZMOV PRI
MODELNIH ORGANIZMIH**

MAGISTRSKO DELO
Magistrski študij – 2. stopnja Molekulska biologija

**IDENTIFICATION OF ULTRACONSERVED REGIONS (UCRs) AND
THEIR FUNCTIONAL POLYMORPHISMS IN MODEL ORGANISMS**

M. SC. THESIS
Master Study Programmes – Molecular biology

Ljubljana, 2013

Magistrsko delo je zaključek univerzitetnega študija – 2. stopnja Molekulska biologija.

Študijska komisija Študija biologije je za mentorico magistrskega dela imenovala prof. dr. Tanjo Kunej in za somentorja prof. dr. Simona Horvata.

Komisija za oceno in predstavitev:

Predsednica: doc. dr. Blagajana HERZOG VELIKONJA
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo

Članica: prof. dr. Tanja KUNEJ
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Član: prof. dr. Simon HORVAT
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Član: prof. dr. Gregor ANDERLUH
Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo

Datum predavitve:

Magistrsko delo je rezultat lastnega dela. Podpisana se strinjam z objavo svojega magistrskega dela v polnem tekstu na spletni strani Digitalne knjižnice Biotehniške fakultete. Izjavljam, da je naloga, ki sem jo oddala v elektronski obliki, identična tiskani verziji.

Špela Deržič

KLJUČNA DOKUMENTACIJSKA INFORMACIJA (KDI)

ŠD	Du2
DK	
KG	bioinformatika/zelo ohranjeni elementi/genetski polimorfizmi/nekodirajoča RNA/rak
KK	
AV	DERŽIČ, Špela
SA	KUNEJ, Tanja (mentorica)/HORVAT, Simon (somentor)
KZ	SI-1000, Večna pot 111
ZA	Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za biologijo
LI	2013
IN	IDENTIFIKACIJA ZELO OHRANJENIH REGIJ GENOMA (UCR) TER NJIHOVIH FUNKCIJSKIH POLIMORFIZMOV PRI MODELNIH ORGANIZMIH
TD	Magistrsko delo (univerzitetni študij)
OP	IX, 59 str., 8 pregl., 24 sl., 1 pril., 42 vir.
IJ	sl
JJ	sl/en
AI	Ozadje problema: V predhodnih raziskavah so v genomu določili 481 zelo ohranjenih regij (angl. <i>ultraconserved regions</i>; UCR), ki so daljša od 200 bp in identična med človekom, mišjo in podgano. Do sedaj so znotraj UCR-jev pri človeku odkrili 30 polimorfizmov, medtem ko pri glodalskih vrstah UCR-SNP-jev še niso identificirali. Ker se podatki o genetski variabilnosti v podatkovnih zbirkah nenehno dopolnjujejo, je bil namen magistrske naloge analizirati, če se pri človeku znotraj UCR-jev nahajajo še dodatni genetski polimorfizmi. Z analizo podatkovnih zbirk smo želeli odkriti nove UCR-SNP-je pri miši in podgani. Želeli smo preveriti, če so UCR-SNP-ji pri miši povezani z lastnostmi in/ali boleznimi ter odkriti nove UCR-je pri govedu. Metode dela: Bioinformacijsko orodje BLAT smo pri človeku, miši in podgani uporabili za določitev novih genomskega lokacij 481 UCR-jev glede na trenutne verzije podatkovnih zbirk in za identifikacijo UCR-jev pri govedu. Genetske polimorfizme znotraj UCR-jev smo iskali s pomočjo orodja Biomart in podatkovne zbirke Ensembl. Za iskanje povezav med UCR-SNP-ji in fenotipom pri miši smo uporabili podatkovni zbirki Mouse Phenome Database (MPD) in Mouse Genome Informatics (MGI). Rezultati: Pri človeku smo znotraj UCR-jev odkrili 1180 genetskih polimorfizmov oz. en genetski polimorfizem na 106,8 bp. Od 1180 genetskih polimorfizmov (1172 iz podatkovnih zbirk in 8 iz literature) je 1105 SNP-jev, 14 DNP-jev (angl. <i>dinucleotide polymorphism</i>), za 42 UCR-SNP-jev pa je predpostavljen škodljivi učinek (angl. <i>deleterious</i>) na funkcijo proteina. Pri miši smo znotraj UCR-jev odkrili 52, pri podgani pa 11 genetskih polimorfizmov. Medvrstna analiza mest polimorfizmov je pokazala ohranjeno mesto polimorfizma, in sicer se pri človeku in miši na istem mestu znotraj uc.322 nahaja UCR-SNP, ki je ohranjen med obema vrstama. Pri človeku je ta UCR-SNP rs185936189 (A>C), pri miši pa rs51296464 (A>G). Ugotovili smo, da SNP-ji znotraj UCR-jev pri miši vplivajo na različne lastnosti in bolezni. Nesinonimni UCR-SNP rs37005101 znotraj gena <i>Ccar1</i> vpliva na sestavo in maso telesa, rast, vzorce obnašanja in koncentracijo metabolitov v krvi. Sinonimni UCR-SNP rs28296350 znotraj gena <i>Hat1</i> vpliva na imunski in kardiovaskularni sistem. Sinonimni UCR-SNP rs46759302 znotraj gena <i>Ext1</i> vpliva na reprodukcijo, koncentracijo lipidov v krvi, sluh in razvoj tumorjev. Pri govedu smo na novo identificirali UCR-je in ugotovili, da jih je približno polovica 100-odstotno ohranjena z genomom človeka, miši in podgane. Zaključek: Rezultati naloge bodo služili za identifikacijo in nadaljnje eksperimentalno potrjevanje pomembnih regulatornih regij, ki se pogosto nahajajo v ohranjenih območjih. Prav tako bodo prispevali k razvijanju novih biooznačevalcev (angl. <i>biomarkers</i>) za bolezni in fenotipske lastnosti, novih tarč za zdravljenje in terapevtske posege pri človeku ter za razvoj selekcijskih označevalcev za bolezni in selekcijo pri domačih živalih.

KEY WORDS DOCUMENTATION (KWD)

ND Du2
DC
CX bioinformatica/ultraconserved regions/genetic polymorphisms/non-coding RNA/cancer
CC
AU DERŽIČ, Špela
AA KUNEJ, Tanja (supervisor)/HORVAT, Simon (co-advisor)
PP SI-1000, Večna pot 111
PB University of Ljubljana, Biotechnical Faculty, Department of Biology
PY 2013
TI IDENTIFICATION OF ULTRACONSERVED REGIONS (UCRs) AND THEIR FUNCTIONAL POLYMORPHISMS IN MODEL ORGANISMS
DT M. Sc. Thesis (University studies)
NO IX, 59 p., 8 tab., 24 fig., 1 ann., 42 ref.
LA sl
AL sl/en
AB **Background:** Previous studies identified 481 ultraconserved regions (UCRs) which are segments that are longer than 200 bp and are absolutely conserved between human, mouse and rat genomes. Till now 30 single nucleotide polymorphisms (SNPs) inside UCRs have been identified in the human genome but none of them yet in mouse and rat. Because the information about genetic polymorphisms in genome databases is continuously updated the aim of our study was to determine novel genetic polymorphisms inside UCRs in human genome. Additionally, we aimed to identify new genetic polymorphisms in UCRs of the mouse, rat and bovine genome. Genetic polymorphisms in UCRs in mouse have been analyzed and their potential role in various phenotypic traits have been determined. **Methods:** A bioinformatic tool BLAT was used to identify genomic locations of UCRs in human, mouse and rat of latest genome assembly release and to determine new UCRs in bovine. Genetic polymorphisms inside UCRs have been searched by using bioinformatic tool Biomart and Ensembl database. Associations between the mouse polymorphisms in UCRs and phenotypic traits were analyzed by bioinformatic tools within the Mouse Phenome Database (MPD) and Mouse Genome Informatics (MGI). **Results:** 1180 genetic polymorphisms were identified in the human genome or one polymorphism per 106,8 bp. Out of 1180 (1172 from databases and 8 from the literature) genetic polymorphisms 1105 are SNPs, 14 dinucleotide polymorphisms (DNPs) and 42 deleterious SNPs. We also uncovered 52 polymorphisms in UCRs in the mouse and 11 SNPs inside UCRs in rat genome. Alignment between locations of genetic polymorphisms inside UCRs has revealed a conserved SNP in uc.322 between human (rs185936189; A>C) and mouse (rs51296464; A>G) genome. We detected correlations between mouse SNPs in UCRs and phenotypic traits and diseases. Non-synonymous SNP rs37005101 within the *Ccar1* gene shows an effect on body weight and composition, growth, behaviour traits and concentration of metabolites in blood. Synonymous SNP rs28296350 within the *Hat1* gene shows an effect on immune and cardiovascular system. Synonymous SNP rs46759302 within the *Ext1* gene has an effect on reproduction, concentration of blood lipids, hearing and tumors development. We have identified UCRs in bovine and found out that about half of the UCRs in bovine are 100 % identical with the human, mouse and rat genomes. **Conclusion:** The bioinformatic analyses and collated data in the present study can serve researchers as a starting point for experimental validation and identification of important regulatory regions and functional polymorphisms. These, in turn, can become new targets in therapeutic interventions and serve as diagnostic and selective markers for diseases and selection in breeding animals.

KAZALO VSEBINE

KLJUČNA DOKUMENTACIJSKA INFORMACIJA (KDI)	III
KEY WORDS DOCUMENTATION (KWD).....	IV
KAZALO VSEBINE	V
KAZALO PREGLEDNIC	VII
KAZALO SLIK	VII
SLOVARČEK.....	IX
1 UVOD	1
2 PREGLED OBJAV	3
2.1 ZELO OHRANJENE REGIJE GENOMA (UCR).....	3
2.2 UCR IN BOLEZNI PRI ČLOVEKU	7
2.2.1 Kronična limfocitna levkemija	7
2.2.2 Rak dojke	8
2.2.3 Rak debelega črevesa.....	9
2.2.4 Rak jeter	9
2.2.5 Nevroblastom	9
2.2.6 Multipla skleroza	10
2.3 UCR IN NKRNA	10
2.4 GOSTITELJSKI GENI ZA UCR-JE.....	12
2.5 EPIGENETSKO URAVNAVANJE IZRAŽANJA GENOV ZA NEKODIRAJOČE RNA.....	13
3 MATERIAL IN METODE	14
3.1 FIZIČNA GENOMSKA KARAKTERIZACIJA UCR-JEV	14
3.2 ISKANJE SNP-JEV ZNOTRAJ UCR-JEV IN ZBIRANJE PODATKOV O SNP-JIH.....	16
3.3 ISKANJE POVEZAV MED UCR-SNP-JI IN ZNAČILNOSTMI LINIJ MIŠI	19
4 REZULTATI.....	20
4.1 FIZIČNA GENOMSKA KARAKTERIZACIJA OBMOČIJ UCR	21
4.1.1 Identifikacija UCR-jev pri različnih modelnih organizmih	21
4.1.2 Prekrivanje UCR-jev z drugimi genetskimi elementi	26
4.1.3 Iskanje genetske variabilnosti območij UCR.....	33
4.1.3.1 Iskanje genetske variabilnosti območij UCR pri človeku	33
4.1.3.2 Iskanje genetske variabilnosti območij UCR pri miši	37
4.1.3.3 Iskanje genetske variabilnosti območij UCR pri podgani	39

4.1.3.4 Iskanje genetske variabilnosti območij UCR pri govedu	39
4. 2 FUNKCIONALNA GENOMSKA KARAKTERIZACIJA OBMOČIJ UCR	39
4.2.1 Integracija podatkov s predhodnimi objavami v znanstveni literaturi.....	40
4.2.2 Iskanje povezav med UCR-SNP-ji in značilnostmi linij miši	40
5 RAZPRAVA.....	46
5.1 FIZIČNA GENOMSKA KARAKTERIZACIJA OBMOČIJ UCR	46
5.1.1 Identifikacija UCR-jev pri različnih modelnih organizmih	46
5.1.2 Prekrivanje UCR-jev z drugimi genetskimi elementi	46
5.1.3 Iskanje genetske variabilnosti območij UCR.....	47
5.1.3.1 Iskanje genetske variabilnosti območij UCR pri človeku	48
5.1.3.2 Iskanje genetske variabilnosti območij UCR pri miši	48
5.2 FUNKCIONALNA GENOMSKA KARAKTERIZACIJA OBMOČIJ UCR	48
5.2.1 Integracija podatkov s predhodnimi objavami v znanstveni literaturi.....	48
5.2.2 Iskanje povezav med UCR-SNP-ji in značilnostmi linij miši	50
6 SKLEPI	52
7.1 POVZETEK	53
7.2 SUMMARY	54
8 VIRI	56

KAZALO PREGLEDNIC

Preglednica 1: Pregled identificiranih genetskih polimorfizmov znotraj UCR-jev pri človeku.	6
Preglednica 2: Pregled UCR-jev glede na dolžino v bp.	22
Preglednica 3: Razvrstitev gostiteljskih genov za UCR-je v posamezne skupine pri človeku, miši, podgani in govedu.	26
Preglednica 4: Pregled UCR-jev, ki se prekrivajo z geni za druge tipe nkRNA (lincRNA in miRNA) pri človeku in miši.	26
Preglednica 5: Pregled UCR-jev, ki se prekrivajo z več gostiteljskimi geni pri človeku. ..	27
Preglednica 6: Pregled UCR-jev, ki se prekrivajo z več gostiteljskimi geni pri miši.	29
Preglednica 7: Pregled UCR-SNP-jev s predvidenim škodljivim učinkom na funkcijo proteina.	35
Preglednica 8: Pregled 48 UCR-SNP-jev s frekvencami alelov in linijami miši.	41

KAZALO SLIK

Slika 1: Razporeditev 481 zaporedij UCR po kromosomih pri človeku	3
Slika 2: Razdelitev molekul RNA.. ..	4
Slika 3: Vhodna stran orodja BLAT	14
Slika 4: Izpis rezultata, ki ga vrne orodje BLAT, s podatki o genomski lokaciji UCR-ja, dolžini UCR-ja, odstotku identičnosti glede na vnešeno zaporedje, kromosomu, na katerem se nahaja UCR, in orientiranosti verige.....	15
Slika 5: Orodje Biomart z vnešenimi podatki za človeka.. ..	16
Slika 6: Prikaz vnosa podatkov v orodje Biomart.. ..	17
Slika 7: Izpis rezultata, ki ga vrne orodje Biomart.. ..	17
Slika 8: Izpis rezultata, ki ga vrne podatkovna zbirka dbSNP.. ..	18
Slika 9: Prikaz poteka raziskave.....	20
Slika 10: Prikaz števila UCR-jev glede na dolžino.	21
Slika 11: Pregled UCR-jev po kromosomih pri človeku	24

Slika 12: Prikaz UCR-jev na kromosomu X pri človeku	24
Slika 13: Pregled UCR-jev po kromosomih pri miši.....	25
Slika 14: Pregled UCR-jev po kromosomih pri podgani.....	25
Slika 15: Število gostiteljskih genov za UCR-je pri vseh štirih vrstah (svetlo siva barva)..	31
Slika 16: Razdelitev gostiteljskih genov za UCR-je na posamezne skupine pri človeku..	31
Slika 17: Razdelitev gostiteljskih genov za UCR-je na posamezne skupine pri miši..	32
Slika 18: Število ohranjenih gostiteljskih genov za UCR-je med človekom, mišjo, podgano in govedom.	32
Slika 19: Dinukleotidni polimorfizmi znotraj UCR-jev pri človeku.....	34
Slika 20: Razdelitev 751 UCR-SNP-jev, ki ležijo znotraj protein-kodirajočih genov, na intronske, eksonske in tiste, ki ležijo znotraj UTR.....	35
Slika 21: Rezultati iskanja genetskih polimorfizmov znotraj UCR-jev pri človeku.	37
Slika 22: Poravnava genomov 13 vrst sesalcev v območju uc.332 z označenimi mesti UCR-SNP-jev.....	38
Slika 23: Razvrščanje UCR-SNP-jev pri miši po prioriteti.	39
Slika 24: Grafični prikaz frekvenc alelov za 48 UCR-SNP-jev pri miši.....	45

SLOVARČEK

CAGR	z rakom povezane genomske regije (angl. <i>cancer-associated genomic regions</i>)
CLL	kronična limfocitna levkemija (angl. <i>chronic lymphocytic leukemia</i>)
CRC	rak debelega črevesa (angl. <i>colorectal cancer</i>)
DNP	dinukleotidni polimorfizem (angl. <i>dinucleotide polymorphism</i>)
HCC	rak jeter (angl. <i>hepatocellular cancer</i>)
indel	insercije-delecije (angl. <i>insertion-deletion</i>)
lincRNA	lincRNA (angl. <i>long intergenic non-coding RNA</i>)
microRNA (miRNA)	mikro RNA, eden izmed razredov nekodirajočih RNA
ncRNA; nkRNA	nekodirajoča RNA; RNA, ki ne kodira proteina (angl. <i>non-coding RNA</i>)
NB	nevroblastom (angl. <i>neuroblastoma</i>)
pre-miRNA	prekurzorska miRNA
SNP	polimorfizem posameznega nukleotida (angl. <i>single nucleotide polymorphism</i>)
T-UCR; UCG	prepisana zelo ohranjena regija (angl. <i>transcribed ultraconserved region</i>); kodirajoča zelo ohranjena regija (angl. <i>ultraconserved gene</i>)
ucRNA	zelo ohranjena RNA (angl. <i>ultraconserved RNA</i>)
UCE; UCR	zelo ohranjen element (angl. <i>ultraconserved element</i>); zelo ohranjena regija (angl. <i>ultraconserved region</i>)
UCR-SNP	SNP, ki leži znotraj UCR-ja
UTR	neprevedeno območje (angl. <i>untranslated region</i>)

1 UVOD

Kljub temu da je poznavanje genomike vretenčarjev močno napredovalo, še vedno ni natančno znano, kolikšen del človeškega genoma in genoma ostalih vretenčarjev je funkcionalen v smislu kodiranja proteinov ali RNA, uravnavanja prepisovanja (transkripcije) in prevajanja (translacije), spreminjanja strukture kromatina ali opravljanja drugih pomembnih celičnih funkcij. Funkcionalne vloge zaporedij so bolje raziskane pri modelnih evkariontskih organizmih, kot so *Drosophila melanogaster*, *Caenorhabditis elegans* in *Saccharomyces cerevisiae*, vendar tudi pri teh modelnih organizmih ostaja še vedno velik del genoma nepoznan. Učinkovite metode za iskanje funkcionalnih zaporedij so zato ključnega pomena, še posebej v velikih genomih, kjer funkcionalni elementi predstavljajo le majhen del celotnega zaporedja. Pri iskanju pomembnih regij v genomu so nam v pomoč zelo ohranjeni elementi (angl. *ultraconserved elements*; UCE), ki so ohranjeni med številnimi vrstami, ne le sesalskimi, kar kaže na njihovo evolucijsko pomembnost (Siepel in sod., 2005).

Zelo ohranjene regije genoma so nov razred genskih regulatornih elementov, vendar je njihova vloga še vedno v veliki meri neznana. Nedavne študije kažejo na vpletenost UCR-jev v različne bolezni pri človeku, npr. pri raku. Nadaljnje raziskave UCR-jev bodo prispevale k boljšemu razumevanju genskega uravnavanja v genomu človeka, patogeneze bolezni pri človeku in razvoju novih metod zdravljenja. Zaradi pomembne vloge mehanizmov nastanka in uravnavanja UCR-jev pri rakavem obolenju lahko terapevtsko ciljanje na te mehanizme v bližnji prihodnosti veliko pripomore k terapiji raka. Kljub temu bo odkrivanje funkcije UCR-jev predstavljalo velik izziv. Večina dosedanjega znanja o UCR-jih je namreč pridobljena s pomočjo bioinformacijskih analiz, eksperimentalne študije pa so še vedno precej omejene. Veliko število UCR-jev v genomih seslancev bo možno določili le s povratnimi informacijami iz eksperimentalnih ugotovitev (Baira in sod., 2008).

Do sedaj so Bejarano in sod. (2004), Chen in sod. (2007), Wojcik in sod. (2010) ter Shen in sod. (2011) pri človeku odkrili 30 genetskih polimorfizmov znotraj UCR-jev. Zanimalo nas je, če poleg 30 že opisanih v teh regijah obstajajo še kateri drugi polimorfizmi. Pri miših in podganah SNP-ji znotraj že identificiranih UCR-jev še niso znani. V magistrski nalogi smo zato analizirali, če se znotraj UCR-jev nahajajo SNP-ji. Za identificirane SNP-je smo pri miših analizirali njihov vpliv na fenotipske lastnosti. Prav tako tudi ni znano, v kolikšni meri so UCR-ji ohranjeni pri drugih modelnih organizmih, zato smo v magistrski nalogi identificirali njihovo ohranjenost pri govedu.

Na podlagi poiskanih SNP-jev v UCR-jih različnih modelnih organizmov bo možno identificirati pomembne regulatorne regije in funkcijske polimorfizme, ki so lahko v prihodnosti diagnostični in selekcijski označevalci za bolezni in selekcijo pri domačih živalih ter nove tarče za terapevtske intervencije pri človeku.

Cilji naloge:

1. Ugotoviti, če pri človeku poleg poznanih 30 SNP-jev v UCR-jih obstajajo še dodatni polimorfizmi.
2. Z analizo podatkovnih zbirk ugotoviti, če se znotraj UCR-jev pri miših in podganah nahajajo SNP-ji.
3. Ugotoviti, če so polimorfizmi znotraj UCR-jev pri miših povezani s fenotipi.
3. Odkriti nove UCR-je pri različnih modelnih organizmih.

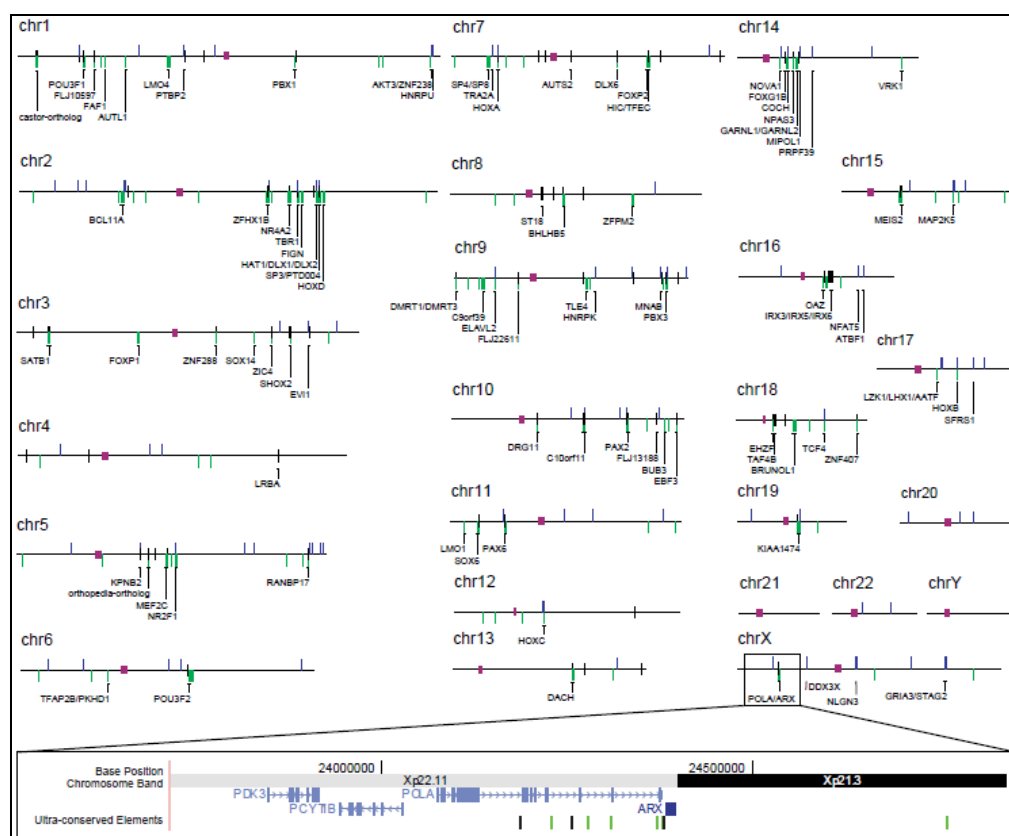
Delovne hipoteze:

1. Znotraj že identificiranih UCR-jev pri miših obstajajo SNP-ji.
2. SNP-ji znotraj UCR-jev pri miših vplivajo na različne lastnosti in/ali bolezni.

2 PREGLED OBJAV

2.1 ZELO OHRANJENE REGIJE GENOMA (UCR)

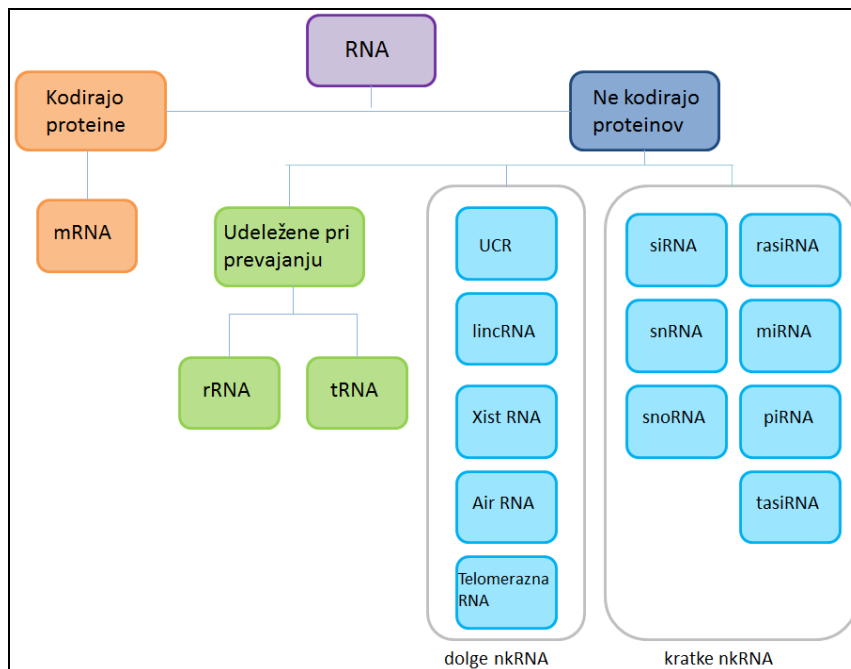
V genomih sesalcev je veliko ohranjenih zaporedij in mnoga med njimi so regulatorna. Zelo ohranjene genomske regije (angl. UCR; *ultraconserved regions*) so definirane kot genomski odseki, ki so daljši od 200 bp in imajo 100-odstotno homologijo z genomi človeka, miši in podgane. Do sedaj so jih odkrili 481. Veliko teh regij kaže tudi od 95- do 99-odstotno stopnjo identičnosti z genomi kokoši in psa, od teh pa je veliko regij močno ohranjenih tudi pri ribah. V genomu človeka UCR-ji pogosto prekrivajo eksone genov, ki so vključeni v procesiranje RNA ali ležijo v intronih ali blizu genov, ki so vpleteni v uravnavanje prepisovanja in razvoja. Primarno se UCR-ji nahajajo v bližini razvojnih genov, tisti v nekodirajočih regijah sesalskih genomov pa lahko uravnavajo izražanje sosednjih razvojnih genov. Znano je, da so UCR-ji razporejeni po celotnem genomu človeka, zato jih lahko najdemo na vseh kromosomih, razen na kromosomih 21 in Y. Nekateri izmed UCR-jev se nahajajo celo v gručinah (angl. *clusters*). Glede na njihovo lokacijo UCR-je razdelimo v delno-eksonske (angl. *partly exonic*), neeksonske (angl. *non-exonic*) in potencialno-eksonske (angl. *possibly exonic*) elemente (Bejerano in sod., 2004). Slika 1 prikazuje pregled 481 zaporedij UCR po kromosomih pri človeku.



Slika 1: Razporeditev 481 zaporedij UCR po kromosomih pri človeku (Bejerano in sod., 2004: 7)

Delno eksonski elementi UCR so označeni z modro črto, neeksonski elementi z zeleno črto in eksonski elementi s črno črto. Vijolični kvadrati prikazujejo centromere. Elementi, ki so med sabo oddaljeni manj kot 675 kb, so združeni v gruče. Takih gručin je 89 in vsebujejo dva ali več UCR-jev. Vsaka gručina je na sliki prikazana kot kvadrat in poimenovana z imenom glavnega gena ali genske družine, ortolognim genom pri organizmu *Drosophila melanogaster* ali oznako mRNA, če ime gena ni znano.

Zaporedja UCR spadajo med dolge nekodirajoče RNA (Ferdin in sod., 2010). Razdelitev nkRNA (angl. *non-coding RNA*) prikazuje slika 2. Med kratke nkRNA uvrščamo tiste nkRNA, ki obsegajo do 200 bp, med dolge pa tiste nad 200 bp.



Slika 2: Razdelitev molekul RNA (prirejeno po Kunej in sod., 2012)

Air RNA: antisense IGF2R RNA; lincRNA: long non-coding RNA; mRNA: informacijska RNA; miRNA: mikro RNA; piRNA: PIWI-interacting RNA; rRNA: ribosomska RNA; rasiRNA: repeat-associated siRNA; scnRNA: small-scanner RNA; siRNA: mala interferenčna RNA; snRNA: mala jedrna RNA; snoRNA: mala nukleolarna RNA; tRNA: prenašalna RNA; tasiRNA: transdelujoča RNA; UCR: visoko ohranjene regije; Xist: X-inactive specific transcript (non-protein coding).

Ker so Ahituv in sod. leta 2007 želeli ugotoviti, kakšen je pomen zaporedij UCR pri miših *in vivo*, so jim iz genoma odstranili štiri UCR-je (uc.248, uc.329, uc.467 in uc.482). Zaporedja UCR, ki so jih izbrali, delujejo kot ojačevalci in se nahajajo blizu genov, katerih spremenjeno izražanje ali utišanje vodi v izrazite spremembe fenotipov miši. Pričakovali so, da bo delecija UCR-jev iz genoma miši pomembno vplivala na spremembo njihovega fenotipa. Izkazalo se je, da so vse miši z odstranjenimi UCR-ji preživele, bile so plodne in niso kazale nobenih anomalij v rasti, življenjski dobi, patologiji in metabolizmu. Ti rezultati kažejo na to, da visoka ohranjenost zaporedij še ni nujno ključna za preživetje organizma (Ahituv in sod., 2007).

Znotraj 481 zaporedij UCR so Bejerano in sod. (2004) v genomu človeka odkrili šest polimorfizmov posameznih nukleotidov (angl. SNP; *single nucleotide polymorphism*) (rs1538101, rs1861100, rs2056116, rs7092999, rs7143938 in rs9572903), kar kaže na to, da so bili ti elementi pod visokim evolucijskim pritiskom in so namenjeni za ključne biološke naloge (Bejerano in sod., 2004). Kasneje so Chen in sod. (2007) odkrili še deset, Shen in sod. (2011) pa tri genetske polimorfizme znotraj UCR-jev pri človeku. Wojcik in sod. (2010) so našli 12 dodatnih genetskih polimorfizmov znotraj UCR-jev pri normalni populaciji in bolnikih z rakom. Zanimalo jih je, če se pri bolnikih s kronično limfocitno levkemijo (CLL) in rakom debelega črevesa (angl. *colorectal cancer*; CRC) v UCR-jih

nahajajo mutacije oz. genetski polimorfizmi. Pred tem so preverili, kakšna je genetska variabilnost znotraj UCR-jev pri zdravi populaciji. Naključno so izbrali in določili zaporedja 28 UCR-jem pri 95 belcih. Znotraj izbranih 28 UCR-jev so našli šest SNP-jev oziroma enega na 1572 zelo ohranjenih nukleotidov. To je približno šestkrat manj kot je pogostnost pojavljanja SNP-jev preko desetih regij človeškega haplotipa, ki znaša en SNP na 279 bp. Ti rezultati potrjujejo dejstvo, da so UCR-ji mesta z zelo nizko stopnjo variabilnosti pri zdravi populaciji. Istih 28 UCR-jev, ki so jih uporabili pri normalni populaciji, so pregledali tudi pri vzorcih raka. Določili so zaporedje tumorske genomske DNA. Znotraj 11 UCR-jev tumorske DNA so odkrili osem genetskih polimorfizmov. Šest genetskih polimorfizmov so našli samo pri 74 rakavih bolnikih (9,5 %), dva SNP-ja pri bolnikih s CLL in štiri SNP-je pri bolnikih s CRC. Nobenega od teh šestih polimorfizmov niso našli pri kontrolnih vzorcih. Pogostnost pojavljanja genetskih polimorfizmov v UCR-jih tumorske DNA je bila ena mutacija znotraj UCR-jev na vsakih 90 kb tumorske DNA. Odkrili so tudi nekatere redke polimorfizme v UCR-jih. Znotraj uc.159 so na mestu 103 našli delecijo dveh nukleotidov (del TT), ki je približno trikrat pogostejša pri bolnikih z rakom kot pri normalni populaciji. Znotraj UCR-jev so odkrili tudi take genetske polimorfizme, ki so se pojavili samo pri normalni populaciji, pri rakavih bolnikih pa ne. Za izbranih 28 UCR-jev so ugotovili, da se somatske mutacije trikrat pogosteje pojavljajo v vzorcih raka kot pri zdravih vzorcih (Wojcik in sod., 2010). V preglednici 1 so prikazani do sedaj identificirani genetski polimorfizmi znotraj UCR-jev pri človeku.

Wojcik in sod. (2010) so v svoji študiji preučili tudi možen učinek genetskih polimorfizmov znotraj UCR-jev, predvsem njihov vpliv na interakcijo med miRNA in UCG-ji. Odkrili so, da naj bi se zaporedje uc.276, ki vključuje zamenjavo G>A na mestu 335 pri normalni populaciji, povezalo z *miR-125a*, ko se prepiše v smerni (angl. *sense*) orientaciji, in povezalo z *miR-638*, ko se prepiše v protismerni (angl. *anti-sense*) orientaciji. Zamenjava A v G na mestu 90 znotraj uc.276, ki se pojavlja samo pri bolnikih s CRC, zmanjša apoptozo celic HeLa in moti interakcijo uc.276 z *miR-887* ter *miR-214*, ki se prekomerno izraža v trdnih tumorjih (angl. *solid tumors*).

Preglednica 1: Pregled identificiranih genetskih polimorfizmov znotraj UCR-jev pri človeku

UCR	SNP	GENOMSKA LOKACIJA UCR-JA (kot navedeno v raziskavi)	POVEZAVA SNP-JA Z BOLEZNIJO	REFERENCA (pri kateri je bil SNP prvič opisan)
uc.21	190 (T>C)	neeksonski (intronski ali medgenski)	CRC	Wojcik in sod., 2010
uc.51	rs17049105	medgenski	ni povezave z rakom dojke pri kitajski populaciji (Shen in sod., 2011)	Chen in sod., 2007
uc.53	rs1861100	medgenski	ni povezave z multiplo sklerozo (Ban in sod., 2005)	Bejerano in sod., 2004
uc.67	rs10496382	medgenski	/	Chen in sod., 2007
uc.72	380 (G>A) = rs141047480	neeksonski (intronski ali medgenski)	CLL	Wojcik in sod., 2010
uc.82	rs13020355	medgenski	ni povezave z rakom dojke pri kitajski populaciji (Shen in sod., 2011)	Chen in sod., 2007
uc.133	rs2682406	v intronu gena <i>RSRC1</i>	ni povezave z rakom dojke pri kitajski populaciji (Shen in sod., 2011)	Shen in sod., 2011
uc.140	rs2056116	medgenski	ni povezave z multiplo sklerozo (Ban in sod., 2005); povezava z rakom dojke pri nemški populaciji (Yang in sod., 2008); ni povezave z rakom dojke pri kitajski populaciji (Shen in sod., 2011)	Bejerano in sod., 2004
uc.140	rs2056117	medgenski	/	Chen in sod., 2007
uc.159	103 (DEL TT)	neeksonski (intronski ali medgenski)	CLL in CRC	Wojcik in sod., 2010
uc.189	83 (G>A)	eksonski	/	Wojcik in sod., 2010
uc.206	324 (G>A)	neeksonski (intronski ali medgenski)	CLL	Wojcik in sod., 2010
uc.211	rs17291131	v intronu gena <i>SKAP2</i>	/	Chen in sod., 2007
uc.243	146 (G>A) = rs73237566	potencialno eksonski	CLL	Wojcik in sod., 2010
uc.252	rs1538101	medgenski	ni povezave z multiplo sklerozo (Ban in sod., 2005)	Bejerano in sod., 2004
uc.268	rs3902936	na 3'UTR koncu gena <i>RC3H2</i>	/	Chen in sod., 2007
uc.268	rs3902937	na 3'UTR koncu gena <i>RC3H2</i>	/	Chen in sod., 2007
uc.268	rs12981	na 3'UTR koncu gena <i>RC3H2</i>	/	Chen in sod., 2007
uc.269	16 (T>C)	neeksonski (intronski ali medgenski)	/	Wojcik in sod., 2010

se nadaljuje

nadaljevanje preglednice 1

UCR	SNP	GENOMSKA LOKACIJA UCR-JA (kot navedeno v raziskavi)	POVEZAVA SNP-JA Z BOLEZNIJO	REFERENCA (pri kateri je bil SNP prvič opisan)
uc.276	335 (G>A) = rs191425730, 90 (A>G)	potencialno eksonski	CRC	Wojcik in sod., 2010
uc.295	rs7092999	medgenski	ni povezave z multiplo sklerozo (Ban in sod., 2005)	Bejerano in sod., 2004
uc.302	rs11190870	medgenski	ni povezave z rakom dojke pri kitajski populaciji (Shen in sod., 2011)	Shen in sod., 2011
uc.328	179 (G>T) = rs41275156	potencialno eksonski	CRC	Wojcik in sod., 2010
uc.341	291 (C>A)	eksonski	/	Wojcik in sod., 2010
uc.353	rs9572903	medgenski	ni povezave z multiplo sklerozo (Ban in sod., 2005); mejna povezanost z rakom dojke pri nemški populaciji (Yang in sod., 2008); ni povezave z rakom dojke pri kitajski populaciji (Shen in sod., 2011)	Bejerano in sod., 2004
uc.368	rs8004379	v intronu gena <i>NPAS3</i>	ni povezave z rakom dojke pri kitajski populaciji (Shen in sod., 2011)	Shen in sod., 2011
uc.374	rs7143938	v intronu gena <i>MIPOL1</i>	ni povezave z multiplo sklerozo (Ban in sod., 2005)	Bejerano in sod., 2004
uc.433	rs4300725	medgenski	/	Chen in sod., 2007
uc.461	rs11573440	v intronu gena <i>POLA1</i>	/	Chen in sod., 2007
uc.483	166 (A>T)	potencialno eksonski	CRC	Wojcik in sod., 2010

190 (T>C) – znotraj UCR-ja je na poziciji 190 nukleotidna zamenjava T>C; CLL – kronična limfocitna levkemija; CRC – rak debelega črevesa; / – ni povezave z boleznijo

2.2 UCR IN BOLEZNI PRI ČLOVEKU

Visoka evolucijska ohranjenost UCR-jev med različnimi vrstami sesalcev kaže na velik pomen tovrstnih regij v različnih procesih. Specifične spremembe v izražanju UCR-jev zato lahko vodijo do različnih bolezni, kot so rak dojke (Yang in sod., 2008), kronična limfocitna levkemija (CLL), rak debelega črevesa (CRC), rak jeter (Braconi in sod., 2011; Calin in sod., 2007), in nevroblastom (Mestdagh in sod., 2010; Scaruffi in sod., 2009).

2.2.1 Kronična limfocitna levkemija

Kronična limfocitna levkemija je rak kostnega mozga, pri katerem v kostnem mozgu nastajajo in se kopičijo nenormalne in nezrele bele krvničke. Raziskave genomov so pokazale, da imajo UCR-ji pomembno vlogo pri nastanku levkemij in karcinomov. Rakave celice imajo edinstven nabor izraženih UCR-jev v primerjavi z normalnimi celicami, kar kaže na to, da so značilne spremembe v izražanju UCR-jev vključene v maligne procese.

Zelo ohranjena zaporedja se pogosto nahajajo na fragilnih mestih v genomu in na genomskih mestih, ki so povezana z rakom. Pri nastanku CLL sodelujejo tudi miRNA. Calin in sod. (2007) so odkrili določene UCR-je, katerih izražanje je bilo uravnavano z miRNA, ki so bile nenormalno izražene pri CLL. Dokazali so tudi, da inhibicija prekomerno izraženih UCR-jev vodi v apoptozo rakavih celic debelega črevesa. Raziskave Calina in sod. (2007) kažejo na to, da so ncRNA in interakcije med nekodirajočimi geni vključene v razvoj tumorjev v veliko večji meri, kot je bilo znano do sedaj. Kot primer, aktivni molekuli miR-16-1/miR-15a, ki sta popolnoma ohranjeni med človekom, mišjo in podgano ter visoko ohranjeni med devetimi od desetih vrst primatov, imata pomembno vlogo pri razvoju CLL. V rakavih celicah sta pri razvoju CLL molekuli miR-15a in miR-16-1, na kromosomu 13q14.3 na 30 kb dolgi regiji, pogosto odsotni zaradi delecije (Calin in sod., 2007).

Številni UCR-ji nimajo vloge genov, ampak imajo regulatorno funkcijo kot ojačevalci, medtem ko drugi UCR-ji predstavljajo eksone protein-kodirajočih genov z znano ali neznano povezanostjo z rakom. Še posebej zanimiva regija je lokus *DACH1*, ki vsebuje sedem UCR-jev znotraj 700 kb (Bejerano in sod., 2004). Trije od UCR-jev iz tega zaporedja se različno izražajo pri različnih vrstah raka, dva od njih sodelujeta pri razvoju CLL. Večina ohranjenih regij lokusa *DACH1* pri miši so ojačevalci, vključno z uc.351, ki ni bil izražen pri nobenem od analiziranih rakavih tkiv. Zanimivo je, da sta samo dve regiji, ki nista imeli vloge ojačevalcev, uc.348 in uc.352, obe nekodirajoči in različno izraženi v rakavem tkivu. Povezali so ju s CLL, pri čemer nobeden od znanih protein-kodirajočih genov ni bil mutiran (Calin in sod., 2007).

Wojcik in sod. (2010) so znotraj štirih UCR-jev (uc.72, uc.159, uc.206 in uc.243) odkrili genetske polimorfizme, ki so se pogosteje pojavljali pri bolnikih s CLL. Genetski polimorfizem v uc.159 je delecija TT in se pojavlja tudi pri bolnikih s CRC.

2.2.2 Rak dojke

Rak dojke je takoj za pljučnim rakom drugi najpogostejši rak na svetu in najpogostejša oblika raka, ki se konča s smrtjo. Yang in sod. (2008) so preučevali vpliv šestih UCR-SNP-jev na pojavnost raka dojke. Od šestih analiziranih UCR-SNP-jev štirje niso kazali povezanosti z rakom dojke, UCR-SNP rs9572903 je kazal samo mejno povezanost z njegovim razvojem, prisotnost UCR-SNP-ja rs2056116 pa je bila višja pri bolnikih z rakom kot pri kontrolah. To kaže na povezavo med spremembo v UCR-ju in razvojem raka. Oba UCR-SNP-ja, rs9572903 in rs2056116, sta medgenska in se nahajata v elementih uc.353 in uc.140 ter delujeta kot ojačevalca. Taki regulatorni elementi imajo sposobnost uravnavanja izražanja genov tudi na zelo dolge razdalje (Yang in sod., 2008).

Najbližja gena UCR-SNP-ju rs9572903 sta *DACH1* in *FLJ22624*. Vloga gena *FLJ22624* še ni znana, gen *DACH1* pa je udeležen v razvojnih procesih. Ni izključeno, da ima UCR-SNP rs9572903 vpliv na izražanje *DACH1*, kar posledično predstavlja tveganje za razvoj raka. Najbližji gen navzdol od rs2056116 je *RAB28* (angl. *RAS-associated protein 28*), ki je od UCR-SNP-ja oddaljen 360 kb. Gen *RAB28* je majhna GTP-aza v super-družini Ras, ki uravnava membranski transport, vpletena pa je tudi v signalizacijske kaskade genskega izražanja, mitozo in apoptozo. Zaradi pomembnih funkcij proteinov Rab spremembe

izražanja ali mutacije v njih povezujemo z razvojem raka. Najbližji gen navzgor od rs2056116 je *HS3ST1* (angl. *heparan sulphate D-glucosaminyl 3-O-sulfotransferase 1*). Razdalja med njima je 1580 kb. Gen *HS3ST1* kodira heparan sulfat biosintetski encim, ki igra pomembno vlogo pri nastanku heparan sulfat glikozaminoglikanskih struktur. Heparan sulfat glikozaminoglikani se nahajajo v zunajceličnem matriksu, sodelujejo pri celičnem signaliziranju in se povezujejo z različnimi proteini, kot so rastni dejavniki in morfogeni, s čimer lahko prispevajo k razvoju raka (Yang in sod., 2008).

2.2.3 Rak debelega črevesa

Rak debelega črevesa predstavlja 13 % vseh oblik raka in je drugi najpogostejši vzrok smrti zaradi raka v zahodnem svetu (Rossi in sod., 2009). Calin in sod. (2007) so dokazali vpletenost UCR-jev pri CRC. Odkrili so, da je uc.73 eden izmed T-UCR-jev, ki se najbolj spremenjeno izražajo pri CRC. To kaže na njegovo regulatorno vlogo pri izražanju kodirajočih genov. Element uc.73 je protismerni prepis UCE 73, ki se nahaja na kromosomu 2 in ima onkogeno vlogo pri CRC. Rakave celice debelega črevesa prekomerno izražajo uc.73. Njegovo izražanje se lahko zniža z malo interferenčno RNA (angl. *small interfering RNA*; siRNA), kar vodi v programirano celično smrt (apoptozo) in ima antiproliferativni učinek na rakave celice (Calin in sod., 2007).

2.2.4 Rak jeter

V karcinogenezo raka jeter (angl. *hepatocellular cancer*; HCC) je vključeno veliko število genov, ki se kompleksno povezujejo med sabo in tako spreminjajo ključne poti, ki so odgovorne za rast tumorskih celic. S pomočjo molekularnih tehnik so opisali transkriptom in številne gene, ki se spremenjeno izražajo pri HCC. Zaradi pomembne funkcionalne vloge ohranjenih zaporedij so se osredotočili na ucRNA (angl. *ultra conserved RNA*) v HCC. Izražanje celičnih ucRNA je bilo spremenjeno s siRNA. V primerjavi z nemalignimi hepatociti je bilo v celicah HepG2 nenormalno izraženih 56 ucRNA. Izmed vseh ucRNA je bila največja sprememba v izražanju opažena pri uc.338, katerega izražanje je bilo močno povišano pri HCC v primerjavi z nekancerogenimi sosednjimi tkivi. Čeprav uc.338 delno prekriva gen *PCBP2*, je njegovo izražanje neodvisno od *PCBP2* (Braconi in sod., 2011).

2.2.5 Nevroblastom

Nevroblastom (angl. *neuroblastoma*; NB) je agresivni otroški tumor simpatičnega živčnega sistema, ki je odgovoren za 15 % vseh smrti otrok zaradi raka. Klinični potek bolezni se med bolniki močno razlikuje. Za agresivne tumorje NB je značilna kombinacija genetskih nepravilnosti, vključno z delecijami kromosomov 1p ali 11q, pridobitvijo 17q in pomnožitvami gena *MYCN* (Mestdagh in sod., 2010).

Mestdagh in sod. (2010) so dokazali, da se T-UCR-ji izražajo v tumorjih NB in da je njihovo izražanje povezano s kliničnimi in genetskimi parametri pri NB. Izražanje T-UCR-jev naj bi bilo povezano tudi s pomnoževanjem gena *MYCN*. S pomočjo celičnih modelov so ugotovili, da so T-UCR-ji povezani s številnimi z rakom povezanimi celičnimi procesi, kot sta apoptoza in diferenciacija. Element uc.73 so povezali s signalno potjo tumorskega proteina p53 (Mestdagh in sod., 2010).

Scaruffi in sod. (2009) so odkrili, da ima neustrezno uravnavanje miRNA/T-UCR pomembno vlogo pri razvoju NB. Večina T-UCR-jev, ki so jih Scaruffi in sod. (2009) našli v vzorcih NB, je neeksonskih (53 %), medtem ko se eksonski in potencialno-eksonski T-UCR-ji prepisujejo redkeje (23,3 % in 23,6 %). Omembe vrednih je pet UCR-jev (uc.322, uc.323, uc.421, uc.423 in uc.452) znotraj treh genov (*SOX6*, *ZNF521*, *TSHZ3*), ki se povišano izražajo pri bolnikih z NB, ki so preživeli dlje (angl. *long-survivors*). Scaruffi in sod. (2009) za nekatere T-UCR-je pri NB predpostavljajo, da imajo pozitivno vlogo ojačevalcev na transkripcijo. Znižano izražanje T-UCR-jev pri bolnikih z NB, ki so preživeli krajše obdobje (angl. *short-survivors*), je lahko posledica domnevno povišanega izražanja njihovih lastnih komplementarnih miRNA. Izražanje T-UCR-jev se razlikuje med bolniki, ki so preživeli dlje, in tistimi, ki so preživeli krajše obdobje, tudi zaradi epigenetskega uravnavanja izražanja genov T-UCR-jev. To potrjuje dejstvo, da se metilirani otoki CpG pojavljajo predvsem v NB z zelo slabo prognozo, prav tako epigenetski mehanizmi ne uravnavajo le izražanja kodirajočih genov, ampak tudi nekodirajočih genov, kot so miRNA. Kar 29 od 37 T-UCR-jev je povezanih z otoki CpG v promotorskih regijah njihovih lastnih gostiteljskih genov. Prav tako kot hipermetilacija otokov CpG vodi v utišanje tumor-supresorskih miRNA, kar vodi v razvoj raka, vpliva tudi na gostiteljske gene T-UCR-jev in tako utiša T-UCR-je, ki imajo potencialno vlogo onkogenov v metastatskih tumorjih NB (Scaruffi in sod., 2009).

2.2.6 Multipla skleroza

Multipla skleroza je kronična avtoimunska vnetna bolezen, ki prizadane osrednje živčevje. Ob procesu zaradi poškodb propadajo predvsem ovojnice živčnih vlaken, pride tudi do izgube mielina. Značilno so pri istem bolniku prizadeta različna mesta v osrednjem živčevju in ob različnih časih. Gre za poslabšanja in izboljšanja funkcij, ki jih opravlja osrednje živčevje, pri čemer je izguba funkcije odvisna od mesta, ki je prizadeto. Živčni impulzi se zaradi poškodb mielina ne morejo več normalno prenašati po živčnem vlaknu, kar povzroči motnje v vidu in zaznavanju ter šibkost pripadajočega dela telesa (Stoppard, 2007). Zaradi velikega števila UCR-jev, ki se izražajo v centralnem in perifrenem živčnem sistemu, so mutacije v UCR-jih poskušali povezati tudi z multiplo sklerozo (Ovcharenko, 2008). Na pojavnost multiple skleroze so analizirali vpliv šestih SNP-jev (rs1538101, rs1861100, rs2056116, rs7092999, rs7143938, rs9572903). Polimorfizem rs7143938 se nahaja na drugem intronu gena *MIPOL1* (angl. *Mirror-image Polidactyly Gene 1*) na kromosomu 14q13. Gen *MIPOL1* je razvojni gen, njegov produkt pa lahko vpliva na imunski in živčni sistem. Povezave tega SNP-ja z multiplo sklerozo niso uspeli dokazati. Prav tako niso z multiplo sklerozo povezali preostalih SNP-jev znotraj UCR-jev (Ban in sod., 2005).

2.3 UCR IN nkRNA

Nekodirajoče RNA so molekule RNA, ki ne kodirajo proteinov. Osrednje načelo molekularne biologije pravi, da RNA igrajo vlogo vmesnega prenašalca med genom in končnim proteinom, ki ga kodira ta gen. Geni, ki kodirajo nkRNA, se prepišejo v prepise, ki delujejo kot strukturne, katalitične ali regulatorne molekule RNA. Nekaj skupin nkRNA je poznanih že dolgo časa in imajo dobro znano vlogo v celici. Nekatere od njih so

vkjučene v translacijo RNA (ribosomska RNA in prenašalna RNA) ali izrezovanje intronov iz prekurzorske mRNA (majhna jedrna RNA), druge pa imajo sposobnost spreminjanja ostalih skupin nkRNA (Ferdin in sod., 2010).

Enega izmed razredov nkRNA predstavlja tudi skupina približno 3500 dolgih medgenskih nkRNA (angl. *long intergenic RNA*; *lincRNA*). Geni, ki kodirajo lincRNA, imajo zanimive lastnosti, vključno z evolucijsko ohranjenostjo, njihovi vzorci izražanja so povezani z različnimi celičnimi procesi in vezavo ključnih transkripcijskih dejavnikov na njihove promotorje, prav tako pa se lincRNA same fizično povezujejo s kromatinskimi regulatornimi proteini. Še vedno ostaja nejasno, če imajo prepisi lincRNA kakšno biološko funkcijo. Ugibajo, da imajo geni lincRNA vlogo ojačevalcev in je njihov prepis samo naključen stranski produkt, da prepisi lincRNA delujejo v cis in aktivirajo transkripcijo oz. da lahko prepisi lincRNA delujejo tudi v trans in tako zavirajo transkripcijo (Guttman in sod., 2011).

Do sedaj so najbolj raziskan razred nkRNA miRNA (Ambros, 2008). Mikro RNA so prepisi 19–25 nukleotidov, ki se izrežejo iz prekurzorske miRNA (pre-miRNA), ki ima obliko lasne zanke in je dolga 60–110 nukleotidov. Molekule miRNA so do sedaj odkrili pri nevretenčarjih, vretenčarjih in rastlinah. Mnogo miRNA je ohranjenih tudi med sorodstveno bolj oddaljenimi organizmi, kar kaže na to, da miRNA sodelujejo v pomembnih procesih. Vloga molekul miRNA je uravnavanje izražanja genov med razvojem in diferenciacijo na transkripcijski, post-transkripcijski in/ali translacijski ravni, mnogo funkcij miRNA pa je še vedno neznanih (Rossi in sod., 2009).

Med nkRNA spada tudi nedavno odkrita skupina zelo ohranjenih genov (angl. *ultraconserved genes*; UCG), ki se prepišejo iz UCR-jev in imajo tkivno specifični vzorec izražanja. Neustrezno uravnavanje izražanja UCR-jev je povezano z razvojem raka. Podobno kot miRNA tudi UCG-ji lahko delujejo kot onkogeni ali tumor-supresorski geni, prav tako je izražanje UCG-jev uravnavano z miRNA (Ferdin in sod., 2010).

Calin in sod. (2007) so z uporabo različnih tehnik, kot so prenos po northernu (angl. *Northern blot hybridization*), uporaba mikromrež (angl. *microarray profiling*), obratna transkripcija in kvantitativna verižna reakcija s polimerazo (angl. *reverse transcription qPCR*; RT-qPCR), dokazali, da se UCR-ji pogosto prepisujejo. Ugotovili so, da imajo maligne celice v primerjavi z zdravimi edinstven vzorec izražanja UCR-jev, kar kaže na velike razlike v izražanju T-UCR-jev. Spremenjeno izražanje T-UCR-jev na ravni genoma pri visokem odstotku analiziranih levkemij in karcinomov podpira dejstvo, da so tako kodirajoči kot nekodirajoči geni vpleteni v razvoj tumorjev. Poleg tega povezave med izraženim UCR-jem in miRNA predstavljajo kompleksno funkcionalno pot uravnavanja, v kateri se dva ali več tipov nkRNA poveže in vpliva na končen fenotip. Calin in sod. (2007) so odkrili, da pet T-UCR-jev (uc.269, uc.160, uc.215, uc.346A in uc.348) kaže spremembe v svojem izražanju, kar so povezali z izražanjem miRNA pri bolnikih s CLL. Mikro RNA se lahko neposredno vežejo na UCR-je in uravnavajo njihovo izražanje. To kaže na kompleksen regulatorni mehanizem med miRNA in T-UCR-ji. S pomočjo poravnav zaporedij so ugotovili, da so trije od petih UCR-jev komplementarni s petimi od trinajstih miRNA, kar da šest možnih parov: uc.160::miR-24, uc.160::miR-155, uc.160::miR-223, uc.160::miR-146a, uc.346A::miR-155 in uc.348::miR-29b. Nekodirajoči T-UCR-ji tako

predstavljajo potencialne tarče za miRNA, njihove interakcije pa imajo lahko biološki in prognostični pomen za bolnike z rakom (Calin in sod., 2007).

2.4 GOSTITELJSKI GENI ZA UCR-JE

Znano je, da se približno polovica vretenčarskih miRNA izreže iz intronov protein-kodirajočih in nekodirajočih transkriptov. V določenih primerih se miRNA lahko nahajajo tako v eksonu kot v intronu, kar je odvisno od alternativnega izrezovanja iz gostiteljskega transkripta. En gostiteljski transkript lahko vsebuje več prekrivajočih miRNA, ki tvorijo gručo in se lahko izrežejo iz istega policistronskega primarnega transkripta. Ni še znano, če dolžina gostiteljskega gena vpliva na število miRNA v tem gostiteljskem genu. Ugotovili so, da se miRNA pogosteje nahajajo znotraj kratkih gostiteljskih genov, kar naj bi izviralo iz ugodne evolucije in povezave s pre-miRNA mehanizmom izrezovanja. Evolucijska ohranjenost je bila prav tako odločilni dejavnik pri ko-izražanju miRNA in gostiteljskega gena; miRNA, ki so bile evolucijsko ohranjene, so se običajno izražale skupaj s svojimi gostiteljskimi geni, medtem ko so se miRNA, ki niso bile ohranjene, le redko izražale skupaj z gostiteljskimi geni (He in sod., 2012).

Od 481 UCR-jev jih 111 prekriva mRNA znanih protein-kodirajočih genov (vključno z regijami UTR), za 256 ni nobenega dokaza o transkripciji iz EST ali mRNA pri katerikoli vrsti, za ostalih 114 so podatki o transkripciji nezadostni. Zelo ohranjene elemente delimo na delno-eksonske oz. eksonske, neeksonske in potencialno-eksonske. Sto neeksonskih elementov se nahaja v intronih znanih genov, ostali so medgenski. Neeksonski UCR-ji, ki ležijo v intronih, so pogosto povezani z razvojnimi geni. Tak je uc.328, ki leži v četrtem intronu gena *PAX6* in ima vlogo specifičnega ojačevalca. Tako intronski kot medgenski neeksonski elementi se običajno združujejo v gručo blizu transkripcijskih dejavnikov in razvojnih genov, medtem ko so eksonski in potencialno-eksonski elementi bolj naključno razporejeni po kromosomih. Znanih je 93 genov, ki prekrivajo eksonske UCR-je; tem genom pravimo geni tipa I. Med gene tipa I uvrščamo številne gene, ki kodirajo dobro znane RNA-vezavne proteine, kot so *HNRPK*, *HNRPH1*, *HNRPU*, *HNRPDL*, *HNRPM*, *SFRS1*, *SFRS3*, *SFRS6*, *SFRS7*, *SFRS10*, *SFRS11*, *TRA2A*, *PCBP2* in *PTBP2*. Geni, ki se nahajajo v bližini neeksonskih elementov, so geni tipa II in jih je 255. Bioinformacijske analize so pokazale, da so geni tipa I povezani z vezavo RNA in uravnavanjem izrezovanja ter imajo veliko RNA-prepoznavnih motivov (angl. *RNA recognition motif*; RRM), gene tipa II pa povezujemo z uravnavanjem prepisovanja in vezavo DNA, zato imajo veliko DNA-vezavnih motivov, zlasti Homeobox. To kaže na to, da naj bi bili eksonski UCR-ji povezani s procesiranjem RNA, neeksonski pa z uravnavanjem transkripcije na nivoju DNA (Bejerano in sod., 2004).

Zelo ohranjena zaporedja se nahajajo tudi v sesalskih genih *HOX*, zato naj bi imeli pomembno vlogo pri razvoju zarodka. Predhodne študije so pokazale, da se veliko število genov *HOX*, ki vsebujejo UCR-je, izraža v placenti, čeprav je njihova funkcija še neznana. Med preučevanjem zaporedij DNA genov *HOX* med različnimi sesalskimi vrstami so Lin in sod. (2008) opazili, da se mnogo UCR-jev nahaja v protein-kodirajočih regijah izven motivov homeobox. Geni *HOX* kodirajo skupino transkripcijskih dejavnikov, ki pri živalskih zarodkih uravnava razvoj segmentacijskih struktur v smeri od glave do repa. Ti proteini vsebujejo homeodomeno, ki jo kodira homeobox motiv. Aminokislinska struktura

homeodomene je močno ohranjena, tudi med sesalci in insekti. Na drugi strani so zaporedja izven homeodomene v genih *HOX* precej divergentna in ne vsebujejo ohranjenih domen, razen nekaterih manjših ohranjenih motivov, kot sta motiv MXSXFE na N-koncu in motiv YPWM blizu homeodomene. Raziskave kažejo na to, da protein-kodirajoča zaporedja izven homeodomene vplivajo na razvoj in morfologijo zarodka. Mnogo UCR-jev se nahaja prav izven homeodomene genov *HOX*, iz česar sklepajo, da imajo UCR-ji velik pomen pri funkcijah genov *HOX* (Lin in sod., 2008).

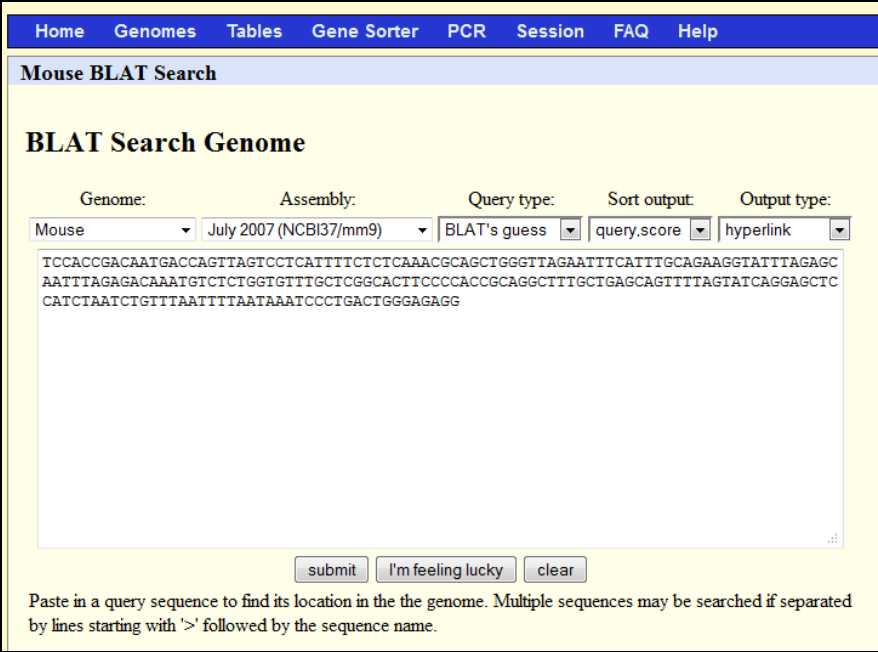
2.5 EPIGENETSKO URAVNAVANJE IZRAŽANJA GENOV ZA NEKODIRAJOČE RNA

Nedavne raziskave so pokazale, da se vzorci izražanja (angl. *expression profiles*) miRNA razlikujejo med normalnimi tkivi in tumorskimi ter med različnimi tipi tumorjev. Molekule miRNA imajo lahko vlogo onkogenov ali tumor-supresorskih genov in sodelujejo pri razvoju tumorjev. Funkcija miRNA je lahko okvarjena zaradi različnih vzrokov: zaradi napak pri post-transkripcijskem uravnavanju miRNA, zaviranja transkripcije z onkogenimi dejavniki, genetskih sprememb, ki vodijo v izgubo funkcije genov za procesiranje miRNA, in zaradi utišanja transkripcije zaradi hipermetilacije otokov CpG v promotorjih genov. Podobno kot pri miRNA lahko tudi pri ostalih nkRNA, ki so vključene v razvoj tumorjev, pride do genetskih in epigenetskih okvar. Znižano izražanje protein-kodirajočih genov, še posebej mnogih tumor-supresorskih genov in miRNA z rastno-zaviralnimi funkcijami, je močno povezano s hipermetilacijo otokov CpG v promotorjih. Lujambio in sod. (2010) so raziskovali, če je enak mehanizem odgovoren tudi za znižano izražanje T-UCR-jev v tumorjih. Odkrili so, da v otokih CpG T-UCR-jev uc.160, uc.283 in uc.346 pride do hipermetilacije in tako utišanja transkripcije pri najpogostejših vrstah raka. Zanimivo je, da je skoraj polovica otokov CpG T-UCR-jev nemetilirana v vseh tkivih, druga polovica pa kaže tkivno specifično metilacijo otokov CpG. Metilirane otoke CpG najdemo v promotorjih kodirajočih genov. To potrjuje dejstvo, da spremembe v epigenetskem uravnavanju izražanja genov vplivajo na tumorigenezo (Lujambio in sod., 2010).

3 MATERIAL IN METODE

3.1 FIZIČNA GENOMSKA KARAKTERIZACIJA UCR-JEV

Pri človeku, miši in podgani smo že znanim 481 UCR-jem določili nove genomske lokacije glede na trenutne verzije podatkovnih zbirk. Za identifikacijo novih genomskih lokacij UCR-jev smo uporabili orodje BLAT (angl. *BLAST-Like Alignment Tool*) (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>) iskalnika UCSC Genome Browser (<http://genome.ucsc.edu/>). BLAT je program, ki omogoča iskanje podobnosti med zaporedji DNA in proteinskimi zaporedji. Hitro najde najmanj 95-odstotno podobnost med zaporedji dolžine 25 bp in več, pri proteinih pa najmanj 80-odstotno podobnost med zaporedji dolžine 20 AA in več. BLAT je veliko hitrejši od starejšega orodja BLAST (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>). Za določitev novih genomskih lokacij UCR-jev pri človeku, miši in podgani smo v orodju BLAT izbrali genom ustreznega organizma; človek: Feb. 2009 (GRCh37/hg19), miš: July 2007 (NCBI37/mm9) ali podgana: Nov. 2004 (Baylor 3.4/rn4). Na sliki 3 je prikazana vhodna stran orodja BLAT z vnešenim zaporedjem UCR.



Slika 3: Vhodna stran orodja BLAT z vnešenim zaporedjem uc.1 in izbrano verzijo podatkovne zbirke miši: July 2007 (NCBI37/mm9)

Pri govedu smo na novo identificirali UCR-je in določili njihov odstotek identičnosti z zaporedji UCR pri človeku, miši in podgani. Za to smo prav tako uporabili orodje BLAT in izbrali verzijo podatkovne zbirke: Nov. 2009 (Bos_taurus_UMD_3.1/bosTau6). Po vnosu zaporedja UCR v program BLAT, ta izpiše rezultat, kot je prikazan na sliki 4.

Home Genomes Tables Gene Sorter PCR Session FAQ Help											
Mouse BLAT Results											
BLAT Search Results											
ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	207	1	207	207	100.0%	4	-	148414326	148414532	207
browser details	YourSeq	20	170	189	207	100.0%	2	-	116597547	116597566	20
browser details	YourSeq	20	131	150	207	100.0%	13	-	51423181	51423200	20
browser details	YourSeq	20	94	113	207	100.0%	11	-	99235940	99235959	20
browser details	YourSeq	20	84	103	207	100.0%	12	+	62434099	62434118	20

Slika 4: Izpis rezultata, ki ga vrne orodje BLAT, s podatki o genomski lokaciji UCR-ja, dolžini UCR-ja, odstotku identičnosti glede na vnešeno zaporedje, kromosomu, na katerem se nahaja UCR, in orientiranosti verige. Na sliki je prikazan izpis rezultata za uc.1 pri miši. Dolžina uc.1 je 207 bp, nahaja se na kromosomu 4 od 148414326 bp do 148414532 bp, orientiranost verige je negativna.

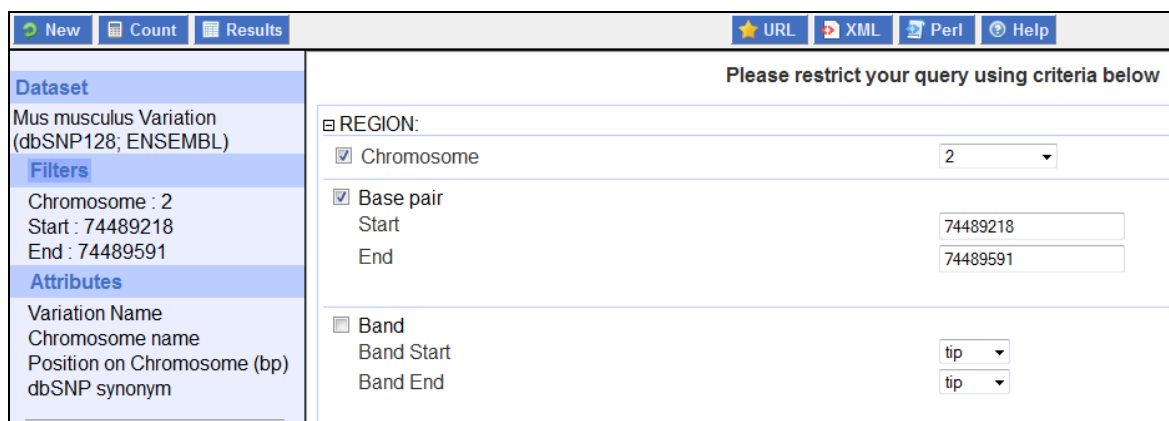
Za poenostavitev ročnega določanja genomskih lokacij UCR-jev pri človeku, miši, podgani in govedu smo uporabili avtomatiziran program, ki prebere excelovo datoteko, v kateri je shranjenih vseh 481 zaporedij UCR in jih preko protokola HTTP POST »pošlje« v program BLAT. Poleg zaporedij UCR program pošlje še ostale podatke (vrsto genoma, verzijo podatkovne zbirke in v kakšni obliki želimo izpis rezultata), ki so potrebni za določitev pravih genomskih lokacij UCR-jev. Rezultat, ki ga vrne BLAT, program obdela in izpiše v excelovo datoteko. Z avtomatiziranim programom smo tako pospešili ročni postopek določanja genomskih lokacij UCR-jev pri vseh štirih vrstah.

Za iskanje genetskih elementov (gostiteljski geni, nkRNA ...), ki se prekrivajo z UCR-ji, smo uporabili orodje Biomart (<http://www.ensembl.org/index.html/>). Izbrali smo podatkovno zbirko Ensembl Genes 67 in ustrezni organizem: *Homo sapiens* genes (GRCh37.p7), *Mus musculus* genes (NCBIM37), *Rattus norvegicus* genes (RGSC3.4) ali *Bos taurus* genes (UMD3.1). Namesto da bi vnašali podatke o genomski lokaciji za vsak UCR posebej, smo izbrali možnost, kjer smo lahko vnesli več genomskih lokacij naenkrat (angl. *Multiple Chromosomal Regions*). Da je bil postopek še hitrejši, smo excelovo datoteko s podatki o genomskih lokacijah UCR-jev pretvorili v tekstovno datoteko (angl. *text file*) in jo vnesli v orodje Biomart. Tako nam podatkov ni bilo treba vnašati ročno. Na sliki 5 je prikazano orodje Biomart z vnešenimi podatki za človeka.

Slika 5: Orodje Biomart z vnešenimi podatki za človeka
Izbrana je podatkovna zbirka *Homo sapiens* genes (GRCh37.p7) in označena možnost Multiple Chromosomal Regions za vnos več genomskih lokacij naenkrat v obliki tekstovne datoteke.

3.2 ISKANJE SNP-JEV ZNOTRAJ UCR-JEV IN ZBIRANJE PODATKOV O SNP-JIH

Za iskanje SNP-jev znotraj UCR-jev smo uporabljali ročni način, in sicer bioinformacijsko orodje Biomart, in avtomatiziran način. V orodju Biomart smo izbrali podatkovno zbirko Ensembl Variation 67 in ustrezni organizem: *Homo sapiens* Variation (dbSNP 135; ENSEMBL), *Mus musculus* Variation (dbSNP 128; ENSEMBL), *Rattus norvegicus* Variation (dbSNP 130) ali *Bos taurus* Variation (dbSNP 133). V orodje Biomart smo vnesli še podatke o genomski lokaciji UCR-ja: kromosom, na katerem se nahaja UCR, ter začetni in končni bazni par zaporedja UCR. Vnos podatkov v orodje Biomart prikazuje slika 6.



Dataset
Mus musculus Variation (dbSNP128; ENSEMBL)

Filters
Chromosome : 2
Start : 74489218
End : 74489591

Attributes
Variation Name
Chromosome name
Position on Chromosome (bp)
dbSNP synonym

Please restrict your query using criteria below

REGION:

☒ Chromosome 2

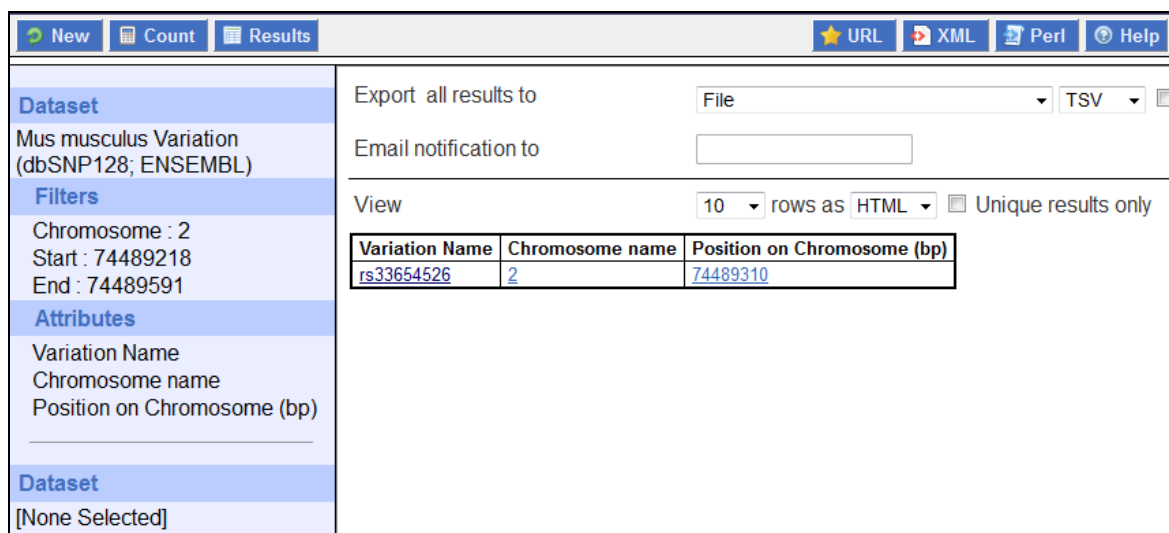
☒ Base pair
Start 74489218
End 74489591

☐ Band
Band Start tip
Band End tip

Slika 6: Prikaz vnosa podatkov v orodje Biomart

Izbrana je podatkovna zbirka *Mus musculus* Variation (dbSNP 128; ENSEMBL). Na sliki so predstavljeni podatki za uc.108, ki se nahaja na kromosomu 2 od 74489218 bp do 74489591 bp.

Po vnosu podatkov o genomski lokaciji UCR-ja nam je orodje Biomart izpisalo genetske polimorfizme, ki se nahajajo znotraj izbranega UCR-ja. Prikazana je bila identifikacijska številka polimorfizma in njegova lokacija na kromosomu. Izpis rezultata, ki ga vrne orodje Biomart, prikazuje slika 7. Da smo si olajšali delo in preverili rezultate ročnega načina iskanja polimorfizmov znotraj UCR-jev, smo polimorfizme v UCR-jih poiskali še na avtomatiziran način.



Dataset
Mus musculus Variation (dbSNP128; ENSEMBL)

Filters
Chromosome : 2
Start : 74489218
End : 74489591

Attributes
Variation Name
Chromosome name
Position on Chromosome (bp)

Export all results to File TSV

Email notification to

View 10 rows as HTML Unique results only

Variation Name	Chromosome name	Position on Chromosome (bp)
rs33654526	2	74489310

Dataset
[None Selected]

Slika 7: Izpis rezultata, ki ga vrne orodje Biomart

Na sliki je prikazan rezultat za miš za uc.108, znotraj katerega se nahaja en SNP: rs33654526, ki leži na kromosomu 2 na lokaciji 74489310 bp.

Podatke o genetskih polimorfizmih v UCR-jih pri človeku, miši, podgani in govedu smo zbirali s pomočjo podatkovnih zbirk Ensembl, dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>) zbirke NCBI (<http://www.ncbi.nlm.nih.gov/>), MGI (<http://www.informatics.jax.org/>) in MPD (<http://phenome.jax.org/>). Tako smo dobili informacije o vrsti genetske variabilnosti (nukleotidna zamenjava, indel), genomski lokaciji variabilnosti na kromosomu, lokaciji glede na gostiteljski gen, statusu validacije in objavi v znanstvenih člankih. Slika 8

prikazuje izpis rezultata, ki ga vrne podatkovna zbirka dbSNP za enega od UCR-SNP-jev pri človeku.

RefSNP
Organism: human ([Homo sapiens](#))
Molecule Type: Genomic
Created/Updated in build: 135/135
Map to Genome Build: [37.3](#)
[Validation Status:](#)

Allele
[Variation Class:](#) SNV: single nucleotide variation
RefSNP Alleles: A/G
Allele Origin:
Ancestral Allele: G
Clinical Source: unknown
Clinical Significance: NA
MAF/MinorAlleleCount: NA
MAF Source:

HGVS Names
[NC_000001.10:g.10597738G>A](#)
[NG_008340.1:g.67736G>A](#)
[NM_004565.2:c.169+1384G>A](#)

SNP Details are organized in the following sections:
[GeneView](#) | [Map](#) | [Submission](#) | [Fasta](#) | [Resource](#) | [Diversity](#) | [Validation](#)

Integrated Maps (Hint: click on 'Chr Pos' or 'Contig Pos' column value to see variation in NCBI sequence viewer)

Assembly	Genome Build	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Neighbor SNP	Map Method
GRCh37.p5	37.3	1	10597738	NT_021937.19	6602470	+	G	+	view	remap

GeneView

GeneView via analysis of contig annotation: [PEX14](#) peroxisomal biogenesis factor 14
View more variation on this gene (click to hide).
☒ Clinical Source: ☐ in gene region ☒ cSNP ☐ has frequency ☐ double hit [Go](#)

Assembly	SNP to Chr	Chr	Primary Assembly Mapping	Contig	Contig position	Allele
GRCh37.p5	+	1	10597738	NT_021937.19	6602470	G

RefSeqGene	Gene (ID)	RefSeqGene Mapping	SNP to RefSeqGene	Position	Allele
NG_008340.1	PEX14 (5195)		+	67736	G

Function class:
[rs190053770](#) is located in the intron region of [NM_004565.2](#).

NC_000001.10: 11M..11M (3.0Kb+) | Find on Sequence: | | Tools | Configure | ?

10,595,500 | 10,596 K | 10,596,500 | 10,597 K | 10,597,500 | [rs190053770](#) | 10,598 K

SNP

Genes

[PEX14/NM_004565.2/NF_004556.1](#)

PEX14
total range: NC_000001.10 (10,535,003..10,690,815)
total length: 155,813
strand: plus
MIM: 601791
Links & Tools
[View GeneID: 5195 \(PEX14\)](#)
[View HGNC: 8856](#)
[View HPRD: 03477](#)
[View MIM: 601791](#)

GeneView via direct blast against RefSeq sequences (used when no gene model is available): N/A

Slika 8: Izpis rezultata, ki ga vrne podatkovna zbirka dbSNP

Na sliki je prikazan rezultat za SNP rs190053770, ki je nukleotidna zamenjava (A>G) in se nahaja znotraj uc.1 na kromosomu 1 in lokaciji 10597738 bp. Leži v intronu gena *PEX14* in ima znan status validacije; določen je bil v Projektu »1000 genomov«.

S pomočjo podatkovne zbirke Ensembl smo dobili podatke o vplivu nesinonimnega UCR-SNP-ja na funkcijo proteina. V podatkovni zbirki Ensembl so informacije o vplivu nesinonimnega SNP-ja na funkcijo proteina pridobili z dvema orodjema: PolyPhen-2 (Polymorphism Phenotyping v2) (<http://genetics.bwh.harvard.edu/pph2/index.shtml>) (Maathuis in sod., 2010; Ramensky in sod., 2002) in SIFT (Sort intolerant from tolerant) (<http://sift.jcvi.org/>) (Ng in Henikoff, 2001). Obe orodji predvidita vpliv zamenjave aminokisline na funkcijo proteina na podlagi evolucije oz. homolognosti zaporedij med vrstami. Orodje PolyPhen se od orodja SIFT razlikuje po tem, da škodljivost zamenjave aminokisline na funkcijo proteina poda z različnimi stopnjami. Orodje PolyPhen lahko s pomočjo podatkovnih zbirk o strukturi proteinov, kot so PDB (Protein Data Bank), PQS

(Protein Quarternary Structure), DSSP (Dictionary of Secondary Structure in Proteins), in podatkovnih zbirk s tridimenzijskimi strukturami napove vpliv zamenjave aminokisline na proteinsko sekundarno strukturo, stike med proteinskimi verigami ter funkcionalna in vezavna mesta proteina. Orodje SIFT napove vpliv zamenjave na funkcijo proteina s pomočjo tolerančnega indeksa, ki ima vrednosti od 0 do 1. Če je tolerančni indeks enak ali manjši od 0,05, ima zamenjava škodljivi učinek (angl. *deleterious*) na funkcijo proteina. Če so vrednosti med 0,05 in 1, pa zamenjava nima vpliva (angl. *tolerated*) na funkcijo proteina. Orodje PolyPhen učinke nesinonimne zamenjave aminokisline na funkcijo proteina razdeli v tri skupine: verjetno škodljiv (angl. *probably damaging*) – velika verjetnost je, da nesinonimna zamenjava vpliva na funkcijo in/ali strukturo proteina; mogoče škodljiv (angl. *possibly damaging*) – zamenjava lahko vpliva na funkcijo in/ali strukturo proteina; in neškodljiv (angl. *benign*) – zamenjava nima učinka na funkcijo proteina.

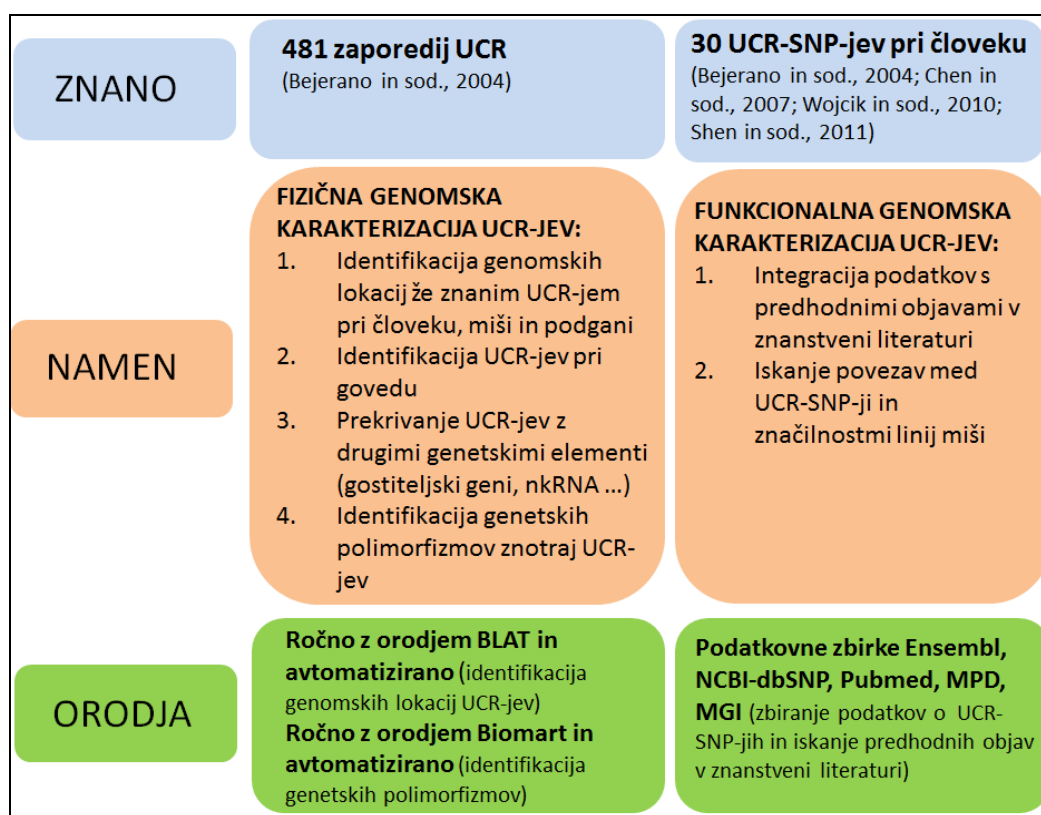
Za iskanje vloge gostiteljskih genov in bioloških poti, v katere so vključeni, smo uporabili različna orodja: KEGG (<http://www.genome.jp/kegg/>), Reactome (<http://www.reactome.org/ReactomeGWT/entrypoint.html>), Panther (<http://www.pantherdb.org/pathway/>), Biocarta (www.biocarta.com/). Podatke o povezanosti gostiteljskih genov z določenimi boleznimi smo dobili iz zbirke OMIM (<http://omim.org/>) in znanstvenih člankov iz zbirke Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed/>).

3.3 ISKANJE POVEZAV MED UCR-SNP-JI IN ZNAČILNOSTMI LINIJ MIŠI

Za genetske polimorfizme znotraj UCR-jev pri miših smo iz podatkovne zbirke za fenom miši (angl. MPD; *Mouse Phenome Database*; <http://phenome.jax.org/>) pridobili podatke o linijah miši in frekvenci alelov za posamezen genetski polimorfizem. Poiskali smo povezave med aleli, prisotnimi pri določenih linijah miši, in značilnostmi teh linij.

4 REZULTATI

Raziskava je potekala po več korakih, ki so prikazani na sliki 9. Potek dela smo razdelili na fizično in funkcionalno genomsko karakterizacijo UCR-jev. Izhajali smo iz 481 zaporedij UCR, ki so jih odkrili Bejerano in sod. (2004) (<http://users.soe.ucsc.edu/~jill/ultra.html>). Zaradi novih verzij podatkovnih zbirk smo pri človeku, miši in podgani že znanim UCR-jem najprej določili genomske lokacije UCR-jev glede na trenutne verzije podatkovnih zbirk. To smo naredili na dva načina: ročno s pomočjo orodja BLAT in avtomatizirano. Dodatno smo identificirali UCR-je pri govedu in določili njihovo ohranjenost glede na zaporedja UCR pri človeku, miši in podgani. Do sedaj so znotraj UCR-jev pri človeku odkrili 30 genetskih polimorfizmov. Nas je zanimalo, če poleg 30 identificiranih SNP-jev v UCR-jih pri človeku obstajajo še dodatni polimorfizmi. To smo preverili ročno z orodjem Biomart in avtomatizirano. Odkrili smo še dodatne SNP-je znotraj UCR-jev pri človeku in nove SNP-je znotraj UCR-jev pri miši, podgani in govedu. Za SNP-je, ki se nahajajo znotraj zaporedij UCR, smo uvedli nov izraz UCR-SNP. S pomočjo podatkovnih zbirk Ensembl, NCBI – dbSNP in PubMed smo zbrali podatke o UCR-SNP-jih: o njihovem statusu validacije, lokaciji glede na gostiteljski gen, prekrivanju z nkRNA in njihovem morebitnem opisu v objavljeni literaturi. Pri človeku in miši smo UCR-SNP-je razvrstili po prioriteti glede na njihovo lokacijo v gostiteljskem genu (intron, ekson, UTR). Eksonske UCR-SNP-je smo razvrstili na sinonimne in nesinonimne ter jim pripisali potencialni učinek na fenotip. Pri miši smo za tri izbrane eksonske UCR-SNP-je poiskali povezavo z značilnostmi linij miši.



Slika 9: Prikaz poteka raziskave

Izvorna preglednica 481 zaporedij UCR (Bejerano in sod., 2004) je dostopna na spletni strani:
<http://users.soe.ucsc.edu/~jill/ultra.html>.

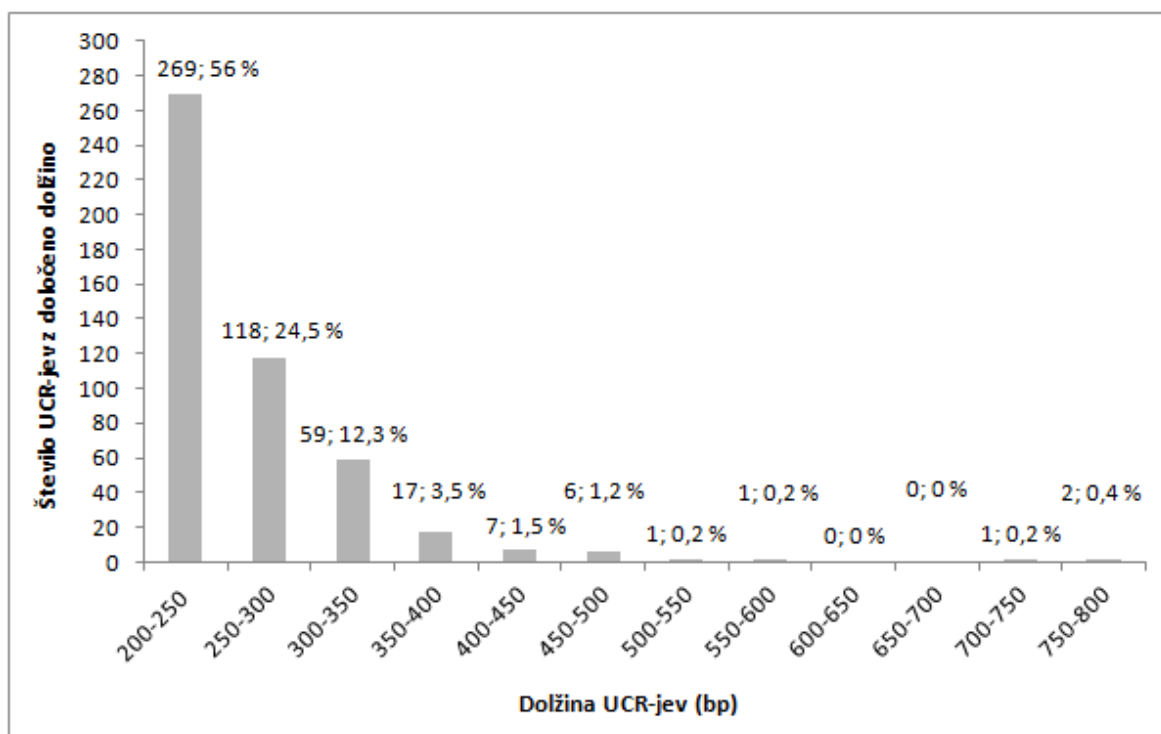
4.1 FIZIČNA GENOMSKA KARAKTERIZACIJA OBMOČIJ UCR

Fizična genomska karakterizacija območij UCR je obsegala identifikacijo genomskih lokacij že znanim UCR-jem pri človeku, miši in podgani ter identifikacijo UCR-jev pri govedu. Pregledali smo prekrivanje UCR-jev z drugimi genetskimi elementi (gostiteljski geni, nkrRNA itd.), identificirali gostiteljske gene z največ UCR-ji in poiskali genetske polimorfizme znotraj UCR-jev.

4.1.1 Identifikacija UCR-jev pri različnih modelnih organizmih

Pri človeku, miši in podgani smo že znanim UCR-jem določili nove genomske lokacije glede na trenutne verzije podatkovnih zbirk. Pri govedu smo na novo identificirali UCR-je in določili njihov odstotek identičnosti z zaporedji UCR pri človeku, miši in podgani.

Po definiciji obsegajo UCR-ji najmanj 200 bp in analiza je pokazala, da je UCR-jev s tako dolžino pet (uc.94, uc.190, uc.323, uc.352 in uc.382). Najdaljši UCR je uc.462 in je dolg 779 bp. Največ UCR-jev je dolgih od 200 do 250 bp, in sicer več kot polovica (56 %). Število UCR-jev z naraščanjem dolžine pada. Le nekaj UCR-jev (1 %) je daljših od 500 bp. Število UCR-jev z določeno dolžino prikazuje slika 10. V preglednici 2 so izpisani UCR-ji z določeno dolžino.



Slika 10: Prikaz števila UCR-jev glede na dolžino

Preglednica 2: Pregled UCR-jev glede na dolžino v bp

Dolžina UCR-jev (bp)	UCR-ji
200–250	uc.94, uc.190, uc.323, uc.352, uc.382, uc.12, uc.73, uc.135, uc.152, uc.167, uc.194, uc.213, uc.222, uc.225, uc.258, uc.9, uc.37, uc.56, uc.204, uc.241, uc.314, uc.346, uc.366, uc.394, uc.472, uc.480, uc.45, uc.58, uc.154, uc.164, uc.203, uc.267, uc.349, uc.386, uc.149, uc.169, uc.321, uc.452, uc.458, uc.481, uc.35, uc.144, uc.212, uc.226, uc.390, uc.90, uc.106, uc.116, uc.240, uc.256, uc.290, uc.400, uc.403, uc.433, uc.1, uc.2, uc.49, uc.51, uc.91, uc.100, uc.131, uc.155, uc.157, uc.263, uc.282, uc.330, uc.364, uc.415, uc.34, uc.132, uc.171, uc.191, uc.300, uc.396, uc.454, uc.54, uc.84, uc.250, uc.284, uc.295, uc.347, uc.385, uc.477, uc.82, uc.219, uc.299, uc.405, uc.436, uc.474, uc.16, uc.81, uc.134, uc.271, uc.406, uc.432, uc.450, uc.457, uc.26, uc.65, uc.129, uc.187, uc.14, uc.156, uc.251, uc.272, uc.369, uc.430, uc.5, uc.146, uc.151, uc.335, uc.399, uc.122, uc.188, uc.424, uc.427, uc.437, uc.8, uc.41, uc.104, uc.243, uc.46, uc.60, uc.67, uc.265, uc.269, uc.292, uc.377, uc.418, uc.143, uc.162, uc.172, uc.208, uc.317, uc.331, uc.337, uc.29, uc.115, uc.118, uc.179, uc.311, uc.410, uc.59, uc.168, uc.280, uc.196, uc.217, uc.50, uc.248, uc.253, uc.334, uc.417, uc.469, uc.473, uc.105, uc.140, uc.288, uc.322, uc.338, uc.423, uc.38, uc.109, uc.130, uc.158, uc.165, uc.197, uc.231, uc.306, uc.374, uc.3, uc.180, uc.324, uc.451, uc.358, uc.422, uc.47, uc.235, uc.342, uc.435, uc.355, uc.368, uc.23, uc.310, uc.411, uc.44, uc.184, uc.202, uc.207, uc.431, uc.227, uc.252, uc.260, uc.291, uc.313, uc.328, uc.53, uc.255, uc.285, uc.307, uc.380, uc.448, uc.15, uc.103, uc.420, uc.62, uc.21, uc.25, uc.315, uc.325, uc.354, uc.404, uc.75, uc.183, uc.13, uc.17, uc.70, uc.107, uc.278, uc.18, uc.98, uc.230, uc.281, uc.381, uc.387, uc.429, uc.476, uc.182, uc.362, uc.428, uc.443, uc.471, uc.55, uc.148, uc.153, uc.201, uc.316, uc.348, uc.350, uc.438, uc.357, uc.30, uc.110, uc.214, uc.266, uc.293, uc.192, uc.409, uc.64, uc.402, uc.455, uc.57, uc.176, uc.414, uc.40, uc.66, uc.113, uc.232, uc.71, uc.78, uc.85, uc.145, uc.441, uc.178, uc.395, uc.434, uc.442, uc.175, uc.209, uc.401
250–300	uc.95, uc.117, uc.268, uc.336, uc.356, uc.378, uc.205, uc.286, uc.339, uc.379, uc.408, uc.478, uc.31, uc.101, uc.200, uc.289, uc.344, uc.68, uc.262, uc.275, uc.351, uc.459, uc.7, uc.19, uc.199, uc.392, uc.43, uc.177, uc.210, uc.220, uc.287, uc.142, uc.340, uc.174, uc.249, uc.96, uc.150, uc.215, uc.426, uc.93, uc.439, uc.36, uc.257, uc.125, uc.228, uc.242, uc.363, uc.42, uc.233, uc.384, uc.236, uc.264, uc.361, uc.223, uc.309, uc.327, uc.412, uc.20, uc.383, uc.120, uc.333, uc.126, uc.389, uc.127, uc.234, uc.303, uc.304, uc.413, uc.446, uc.195, uc.447, uc.52, uc.10, uc.393, uc.460, uc.463, uc.173, uc.277, uc.133, uc.283, uc.308, uc.372, uc.63, uc.161, uc.181, uc.270, uc.365, uc.254, uc.246, uc.301, uc.218, uc.416, uc.124, uc.360, uc.419, uc.27, uc.87, uc.376, uc.449, uc.211, uc.121, uc.80, uc.114, uc.79, uc.141, uc.224, uc.482, uc.77, uc.83, uc.111, uc.229, uc.371, uc.48, uc.367, uc.388, uc.128, uc.239, uc.375

se nadaljuje

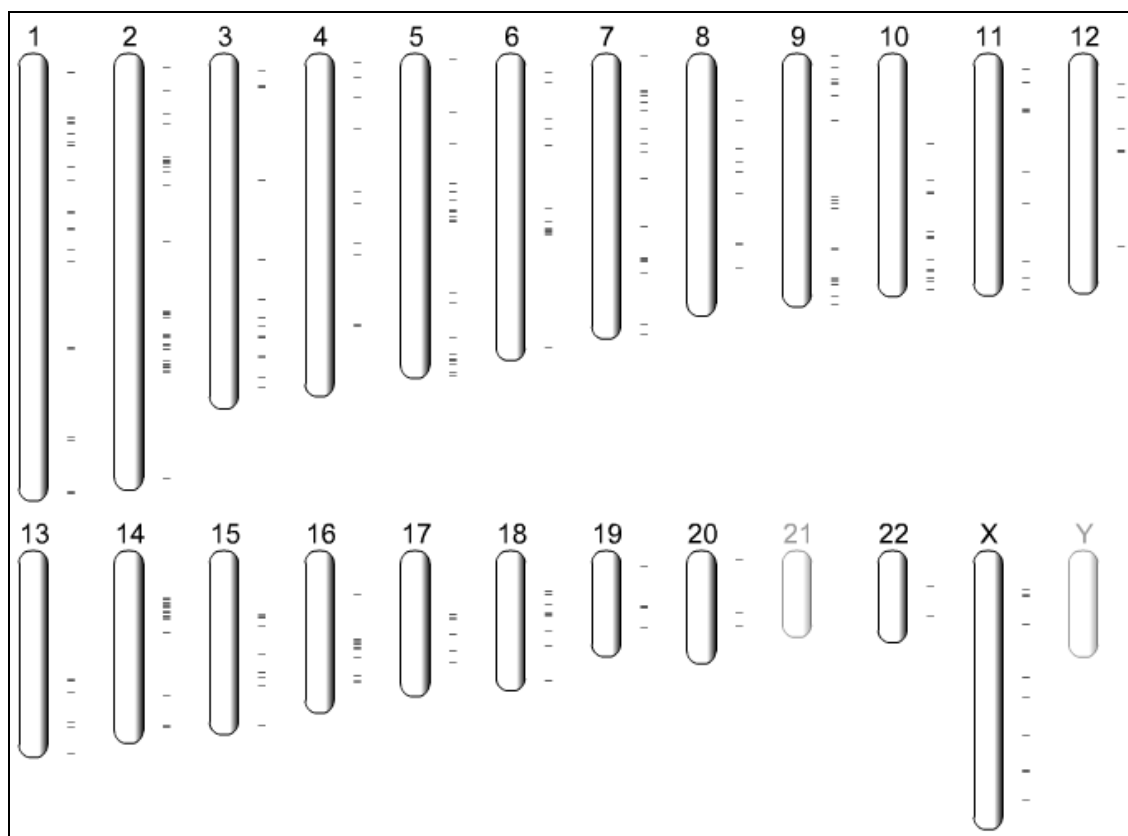
nadaljevanje preglednice 2

Dolžina UCR-jev (bp)	UCR-ji
300–350	uc.6, uc.69, uc.119, uc.345, uc.479, uc.186, uc.305, uc.89, uc.198, uc.329, uc.22, uc.147, uc.259, uc.92, uc.166, uc.170, uc.445, uc.465, uc.261, uc.391, uc.397, uc.33, uc.88, uc.216, uc.341, uc.326, uc.319, uc.193, uc.456, uc.244, uc.273, uc.318, uc.160, uc.312, uc.398, uc.353, uc.359, uc.425, uc.453, uc.61, uc.407, uc.274, uc.440, uc.76, uc.320, uc.24, uc.279, uc.332, uc.102, uc.139, uc.245, uc.86, uc.302, uc.470, uc.421, uc.112, uc.136, uc.221, uc.466
350–400	uc.28, uc.39, uc.238, uc.4, uc.298, uc.247, uc.297, uc.108, uc.163, uc.137, uc.343, uc.444, uc.370, uc.373, uc.461, uc.475, uc.99
400–450	uc.483, uc.72, uc.185, uc.138, uc.276, uc.97, uc.294
450–500	uc.296, uc.237, uc.159, uc.468, uc.123, uc.206
500–550	uc.74
550–600	uc.189
600–650	/
650–700	/
700–750	uc.467
750–800	uc.464, uc.462

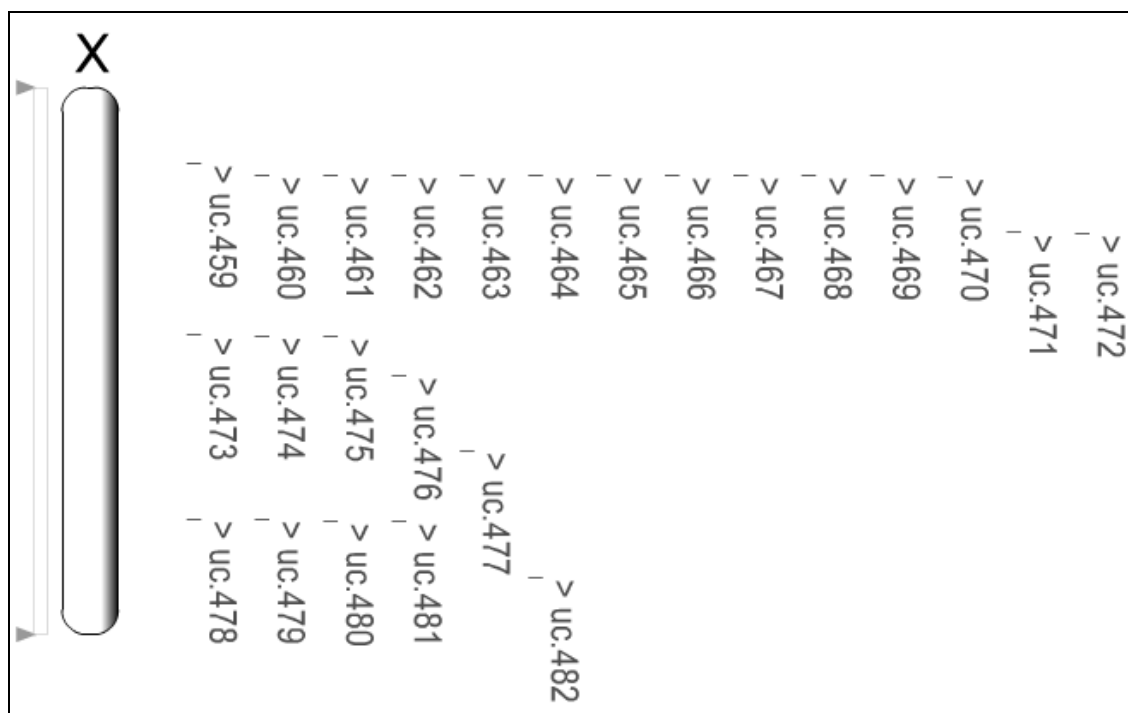
/ – ni UCR-jev s tako dolžino

Ker je znano, da so zaporedja UCR 100-odstotno identična med človekom, mišjo in podgano, nas je zanimalo, kakšna je ohranjenost zaporedij UCR pri govedu glede na UCR-je pri človeku, miši in podgani. Ugotovili smo, da pri govedu ni vseh 481 UCR-jev 100-odstotno ohranjenih glede na genome človeka, miši in podgane. Zaporedij UCR, ki so 100-odstotno identična med človekom, mišjo, podgano in govedom, je približno polovica, in sicer 241. Zaporedje UCR, ki je najmanj ohranjeno z UCR-ji pri človeku, miši in podgani, je uc.457; odstotek identičnosti je pri tem zaporedju 91,9 %. Ostali UCR-ji pri govedu imajo več kot 95-odstotno identičnost z UCR-ji v genomih človeka, miši in podgane.

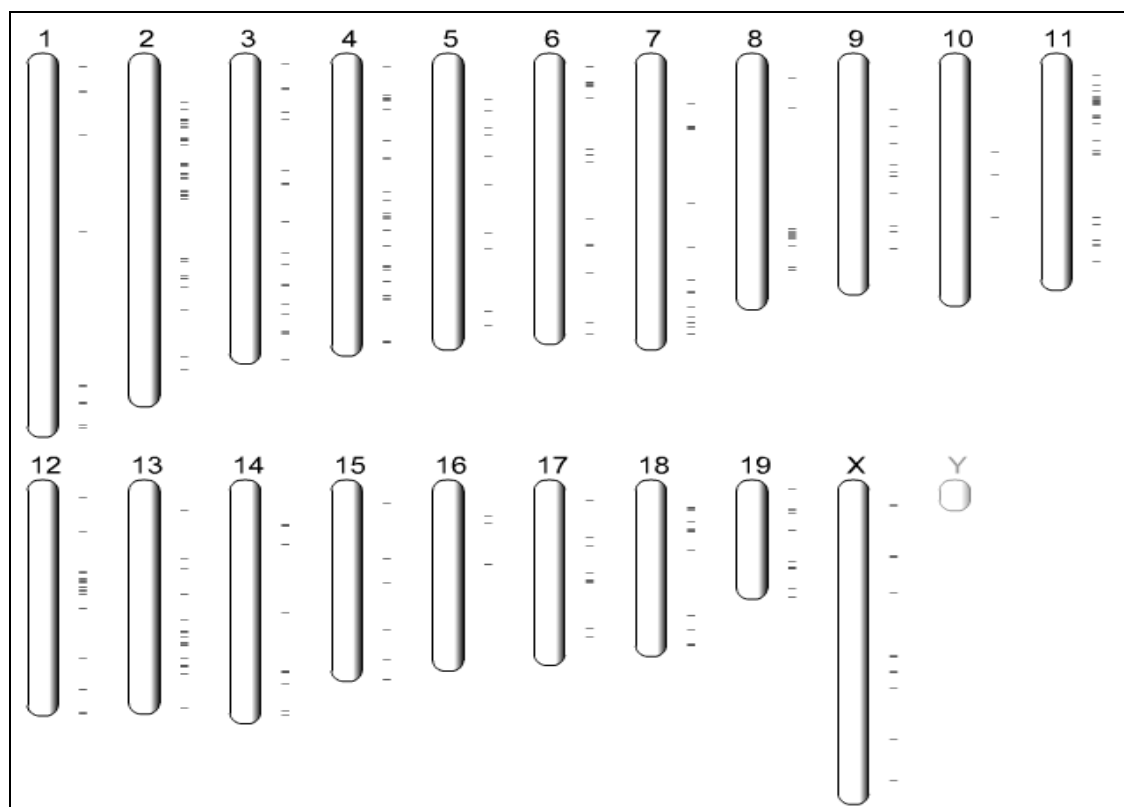
Zaporedja UCR ležijo pri različnih vrstah na različnih kromosomih, ne najdemo pa jih na kromosomu Y. Pri človeku UCR-jev ni tudi na kromosomu 21. Pregled zaporedij UCR po kromosomih pri človeku je prikazan na sliki 11. Na podlagi identificiranih UCR-jev pri človeku, miši, podgani in govedu smo ugotovili, da na kromosomu X pri vseh štirih vrstah ležijo enaki UCR-ji. To so UCR-ji od uc.459 do uc.482. Na sliki 12 so prikazani UCR-ji na kromosomu X pri človeku. Razporeditev UCR-jev po kromosomih pri miši in podgani je prikazana na slikah 13 in 14.



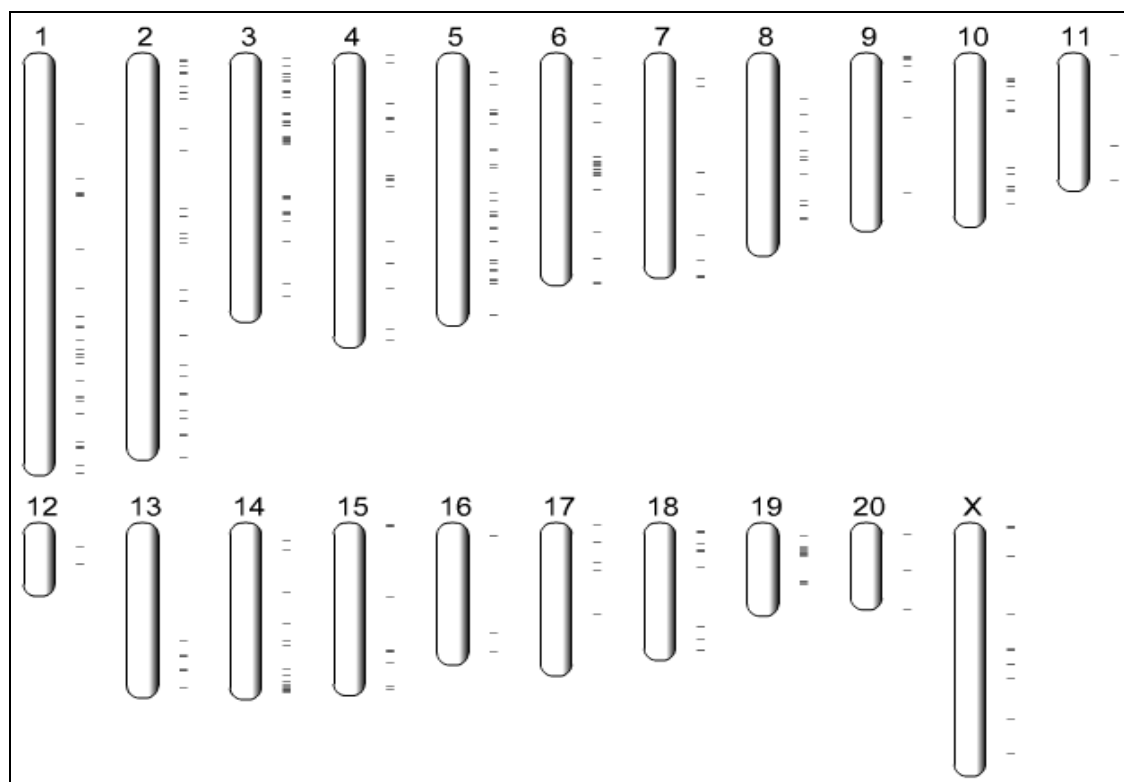
Slika 11: Pregled UCR-jev po kromosomih pri človeku (Jevšinek Skok in sod., v pripravi.
<http://www.integratomics-time.com/>)



Slika 12: Prikaz UCR-jev na kromosomu X pri človeku (Jevšinek Skok in sod., v pripravi.
<http://www.integratomics-time.com/>)



Slika 13: Pregled UCR-jev po kromosomih pri miši (Jevšinek Skok in sod., v pripravi.
<http://www.integratomics-time.com/>)



Slika 14: Pregled UCR-jev po kromosomih pri podgani (Jevšinek Skok in sod., v pripravi.
<http://www.integratomics-time.com/>)

4.1.2 Prekrivanje UCR-jev z drugimi genetskimi elementi

S pomočjo orodja Biomart smo identificirali genetske elemente (protein-kodirajoči gostiteljski geni, geni za nkRNA ...), ki se prekrivajo z UCR-ji. Zanimalo nas je, kateri gostiteljski gen vsebuje največje število UCR-jev. Izkazalo se je, da sta taka gostiteljska gena dva, *PBX3* in *POLA1*, in sicer se prekrivata z osmimi UCR-ji. Gen *PBX3* prekriva UCR-je: uc.273, uc.274, uc.275, uc.276, uc.277, uc.278, uc.279 in uc.280, ki ležijo na kromosomu 9. Dolžina osmih UCR-jev, ki se prekrivajo z genom *PBX3*, je 2404 bp. Gen *POLA1* prekriva UCR-je: uc.460, uc.461, uc.462, uc.463, uc.464, uc.465, uc.466 in uc.467, ki ležijo na kromosomu X. Skupna dolžina UCR-jev, ki se prekrivajo z genom *POLA1*, znaša 3886 bp. Med UCR-ji, ki se prekrivajo z genom *POLA1*, sta tudi dva najdaljša UCR-ja uc.462 in uc.464. Gen *CASZ1* prekriva sedem UCR-jev: uc.2, uc.3, uc.4, uc.5, uc.6, uc.7 in uc.8, ki ležijo na kromosomu 1, njihova skupna dolžina pa je 1778 bp.

Ker smo želeli ugotoviti, kaj kodirajo gostiteljski geni za UCR-je, smo jih razvrstili v posamezne skupine glede na to, ali kodirajo proteine ali druge tipe nkRNA. Razdelitev gostiteljskih genov za UCR-je v posamezne skupine prikazuje preglednica 3. Celotna preglednica UCR-SNP-jev in gostiteljskih genov za UCR-je z imeni genov za človeka, miš in podgano se nahaja v prilogi A.

Preglednica 3: Razvrstitev gostiteljskih genov za UCR-je v posamezne skupine pri človeku, miši, podgani in govedu

Razdelitev gostiteljskih genov v skupine	Človek	Miš	Podgana	Govedo
protein-kodirajoči geni	207	199	174	157
geni za druge tipe nkRNA (lincRNA, miRNA)	23	13	0	1
pseudogeni, protismerni, procesirani prepisi	27	9	2	0
geni, ki prekrivajo več UCR-jev	70	59	33	29
UCR-ji, ki se prekrivajo z več geni	44	21	7	3

Od vseh 257 gostiteljskih genov za UCR-je pri človeku jih približno 9 % predstavljajo geni za druge tipe nkRNA, pri miši pa je gostiteljskih genov za druge tipe nkRNA približno 6 %. V preglednici 4 so z imeni prikazane lincRNA in miRNA, ki se prekrivajo z UCR-ji pri človeku in miši.

Preglednica 4: Pregled UCR-jev, ki se prekrivajo z geni za druge tipe nkRNA (lincRNA in miRNA) pri človeku in miši

Človek			Miš		
uc.29	lincRNA	<i>RP4-544H6.2</i>	uc.57, uc.58, uc.59	lincRNA	<i>Gm12068</i>
uc.52, uc.53, uc.54	lincRNA	<i>AC007092.1</i>	uc.59	lincRNA	<i>4933430M04Rik</i>
uc.93, uc.94	lincRNA	<i>AC092684.1</i>	uc.162	lincRNA	<i>A830009L08Rik</i>
uc.109	lincRNA	<i>AC017048.4</i>	uc.166	lincRNA	<i>C130071C03Rik</i>
uc.114	lincRNA	<i>AC105750.1</i>	uc.169	lincRNA	<i>A830082K12Rik</i>
uc.130	lincRNA	<i>RP11-290K4.2</i>	uc.173	lincRNA	<i>Gm16953</i>
uc.150	lincRNA	<i>P11-121L11.1</i>	uc.198	lincRNA	<i>Gm11884</i>
uc.162	lincRNA	<i>CTD-2249K22.1</i>	uc.266	lincRNA	<i>Gm12478</i>
uc.165	lincRNA	<i>CTC-358I24.1</i>	uc.389	lincRNA	<i>Gm16759</i>
uc.166	lincRNA	<i>LINC00461</i>	uc.403	lincRNA	<i>D230002A01Rik</i>

se nadaljuje

nadaljevanje preglednice 4

Človek			Miš		
uc.167	lincRNA	<i>CTC-454M9.1</i>	uc.410	lincRNA	<i>Gm11431</i>
uc.169	lincRNA	<i>RP11-65F13.2</i>	uc.458	lincRNA	<i>1700109K24Rik</i>
uc.230	lincRNA	<i>RP11-458K10.2</i>	uc.420	miRNA	<i>Mir3064</i>
uc.240	lincRNA	<i>RP11-21C4.1</i>			
uc.250	lincRNA	<i>C9orf146</i>			
uc.266	lincRNA	<i>RP11-308N19.4</i>			
uc.304	lincRNA	<i>RP11-107I14.5</i>			
uc.321	lincRNA	<i>RP11-531H8.2</i>			
uc.358	lincRNA	<i>RP11-314P15.2</i>			
uc.362	lincRNA	<i>CTD-2384A14.1</i>			
uc.363	lincRNA	<i>RP11-260G13.1</i>			
uc.364, uc.365	lincRNA	<i>CTD-2251F13.1</i>			
uc.382	lincRNA	<i>RP11-702M1.1</i>			

Izmed 481 UCR-jev v genomu človeka je tudi 44 takih, ki se prekrivajo z več različnimi gostiteljskimi geni. Pri miši je UCR-jev, ki prekrivajo več gostiteljskih genov 21. V preglednicah 5 in 6 so prikazani UCR-ji, ki se prekrivajo z več geni pri človeku in miši.

Preglednica 5: Pregled UCR-jev, ki se prekrivajo z več gostiteljskimi geni pri človeku

UCR	GEN	IME GENA
uc.43	<i>AKT3</i>	v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)
	<i>RP11-370K11.1</i> (procesiran prepis)	/
uc.45, uc.46	<i>HNRNPU-AS1</i>	HNRNPU antisense RNA 1 (non-protein coding)
	<i>HNRNPU</i>	heterogeneous nuclear ribonucleoprotein U
uc.69	<i>ARHGAP15</i>	Rho GTPase activating protein 15
	<i>AC096558.1</i> (protismerni prepis)	/
	<i>RP11-570L15.2</i> (protismerni prepis)	/
uc.70	<i>ARHGAP15</i>	Rho GTPase activating protein 15
	<i>RP11-434H14.1</i> (protismerni prepis)	/
uc.71, uc.72	<i>GTDC1</i>	glycosyltransferase-like domain containing 1
	<i>AC016910.1</i> (protismerni prepis)	/
uc.96, uc.97	<i>HAT1</i>	histone acetyltransferase 1
	<i>SLC25A12</i>	solute carrier family 25 (aspartate/glutamate carrier), member 12
uc.100	<i>MLK7-AS1</i> (protismerni prepis)	MLK7 antisense RNA 1 (non-protein coding)
	<i>AC013461.1</i> (protein kodirajoči gen)	/
uc.136	<i>MECOM</i>	MDS1 and EVI1 complex locus
	<i>RP11-641D5.2</i> (protismerni prepis)	/

se nadaljuje

nadaljevanje preglednice 5

UCR	GEN	IME GENA
uc.142	<i>PHOX2B</i>	paired-like homeobox 2b
	<i>RP11-227F19.1</i> (protismerni prepis)	/
uc.167	<i>MEF2C</i>	myocyte enhancer factor 2C
	<i>CTC-454M9.1</i> (lincRNA)	/
uc.169	<i>NR2F1</i>	nuclear receptor subfamily 2, group F, member 1
	<i>RP11-65F13.2</i> (lincRNA)	/
uc.189	<i>PII6</i>	peptidase inhibitor 16
	<i>SRSF3</i>	serine/arginine-rich splicing factor 3
uc.213	<i>HOXA-AS3</i>	HOXA cluster antisense RNA 3 (non-protein coding)
	<i>HOXA3</i>	homeobox A3
	<i>HOXA5</i>	homeobox A5
uc.217	<i>VSTM2A</i>	V-set and transmembrane domain containing 2A
	<i>GS1-18A18.1</i> (procesiran prepis)	/
uc.222, uc.223	<i>FOXP2</i>	forkhead box P2
	<i>AC073626.2</i> (protismerni prepis)	/
uc.232	<i>FEZF1</i>	FEZ family zinc finger 1
	<i>RP11-560I19.4</i> (protismerni prepis)	/
uc.245	<i>ZFPM2</i>	zinc finger protein, multitype 2
	<i>RP11-127H5.1</i> (protein kodirajoči gen)	/
uc.263, uc.264	<i>HNRNPK</i>	heterogeneous nuclear ribonucleoprotein K
	<i>RP11-575L7.8</i> (protismerni prepis)	/
uc.290	<i>C10orf11</i>	chromosome 10 open reading frame 11
	<i>RP11-369F10.3</i> (protismerni prepis)	/
uc.325, uc.326, uc. 327	<i>ELP4</i>	elongation protein 4 homolog (<i>S. cerevisiae</i>)
	<i>Z83001.1</i> (procesiran prepis)	/
uc.330	<i>RBM14</i>	RNA binding motif protein 14
	<i>RBM14-RBM4</i>	RBM14-RBM4 readthrough
uc.335	<i>MGST1</i>	microsomal glutathione S-transferase 1
	<i>LMO3</i>	LIM domain only 3 (rhombotin-like 2)
uc.338	<i>PCBP2</i>	poly(rC) binding protein 2
	<i>RP11-793H13.8</i> (procesiran prepis)	/
uc.341	<i>HOXC10</i>	homeobox C10
	<i>HOXC5</i>	homeobox C5
uc.342, uc.343	<i>HOXC5</i>	homeobox C5
	<i>HOXC6</i>	homeobox C6
	<i>HOXC4</i>	homeobox C4
uc.344, uc.345	<i>HOXC5</i>	homeobox C5
	<i>HOXC4</i>	homeobox C4

se nadaljuje

nadaljevanje preglednice 5

UCR	GEN	IME GENA
uc.346	<i>RFX4</i>	regulatory factor X, 4 (influences HLA class II expression)
	<i>RP11-144F15.1</i> (protein kodirajoči gen)	/
uc.408	<i>ZFH3</i>	zinc finger homeobox 3
	<i>RP5-991G20.1</i> (procesiran prepis)	/
uc.409	<i>ZFH3</i>	zinc finger homeobox 3
	<i>RP11-346C20.1</i> (sense overlapping)	/
uc.414	<i>THRA</i>	thyroid hormone receptor, alpha
	<i>NR1D1</i>	nuclear receptor subfamily 1, group D, member 1
uc.415	<i>HOXB3</i>	homeobox B3
	<i>HOXB-AS3</i>	HOXB cluster antisense RNA 3 (non-protein coding)
uc.416	<i>HOXB3</i>	homeobox B3
	<i>HOXB-AS3</i>	HOXB cluster antisense RNA 3 (non-protein coding)
	<i>HOXB5</i>	homeobox B5
uc.417	<i>HOXB-AS3</i>	HOXB cluster antisense RNA 3 (non-protein coding)
	<i>HOXB6</i>	homeobox B6
	<i>AC103702.1</i> (protein kodirajoči gen)	/
uc.457	<i>HIRA</i>	HIR histone cell cycle regulation defective homolog A (<i>S. cerevisiae</i>)
	<i>C22orf39</i>	chromosome 22 open reading frame 39
uc.472	<i>CASK-AS1</i>	CASK antisense RNA 1 (non-protein coding)
	<i>CASK</i>	calcium/calmodulin-dependent serine protein kinase (MAGUK family)

/ - ni imena gena

Preglednica 6: Pregled UCR-jev, ki se prekrivajo z več gostiteljskimi geni pri miši

UCR	GEN	IME GENA
uc.58, uc.59	<i>Gm12068</i> (lincRNA)	/
	<i>4933430M04Rik</i> (lincRNA)	/
uc.61	<i>Bcl11a</i>	B cell CLL/lymphoma 11A (zinc finger protein)
	<i>4930538E20Rik</i> (protismerni prepis)	/
uc.143	<i>G3bp2</i>	GTPase activating protein (SH3 domain) binding protein 2
	<i>Gm15710</i> (pseudogen)	/
uc.169	<i>Nr2f1</i>	nuclear receptor subfamily 2, group F, member 1
	<i>A830082K12Rik</i> (lincRNA)	/
uc.173	<i>Ube2b</i>	ubiquitin-conjugating enzyme E2B
	<i>Gm16953</i> (lincRNA)	/

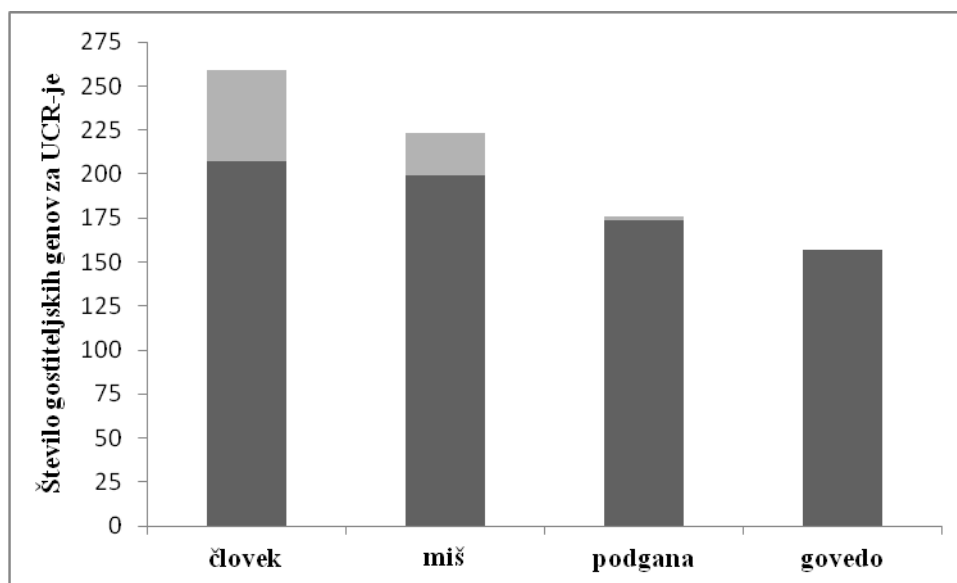
se nadaljuje

nadaljevanje preglednice 6

UCR	GEN	IME GENA
uc.193	<i>Syncrip</i>	synaptotagmin binding, cytoplasmic RNA interacting protein
	<i>Gm20537</i> (protein kodirajoči gen)	/
uc.212	<i>Hoxa2</i>	homeobox A2
	<i>RP23-103L13.8</i> (protein kodirajoči gen)	/
uc.213	<i>Hoxa3</i>	homeobox A3;
	<i>2700086A05Rik</i> (protismerni prepis)	/
	<i>RP23-103L13.8</i> (protein kodirajoči gen)	/
	<i>Hoxa5</i>	homeobox A5
uc.235	<i>Ebf2</i>	early B-cell factor 2
	<i>4930438E09Rik</i> (protismerni prepis)	/
uc.276	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3
	<i>C79798</i> (protismerni prepis)	/
uc.341	<i>Hoxc10</i>	homeobox C10
	<i>Gm20398</i> (nekodirajoči gen)	/
uc.342, uc.343	<i>Hoxc6</i>	homeobox C6
	<i>Gm20398</i> (nekodirajoči gen)	/
uc.344	<i>Hoxc5</i>	homeobox C5
	<i>Gm20398</i> (nekodirajoči gen)	/
uc.389	<i>Iqch</i>	IQ motif containing H
	<i>Gm16759</i> (lincRNA)	/
uc.414	<i>Thra</i>	thyroid hormone receptor, alpha
	<i>Nr1d1</i>	nuclear receptor subfamily 1, group D, member 1
uc.416	<i>Hoxb5</i>	homeobox B5
	<i>0610040B09Rik</i> (protismerni prepis)	/
uc.417	<i>Hoxb6</i>	homeobox B6
	<i>0610040B09Rik</i> (protismerni prepis)	/
uc.420	<i>Ddx5</i>	DEAD (Asp-Glu-Ala-Asp) box helicase 5
	<i>Mir3064</i> (miRNA)	/
uc.458	<i>Rbfox2</i>	RNA binding protein, fox-1 homolog (<i>C. elegans</i>) 2
	<i>1700109K24Rik</i> (lincRNA)	/

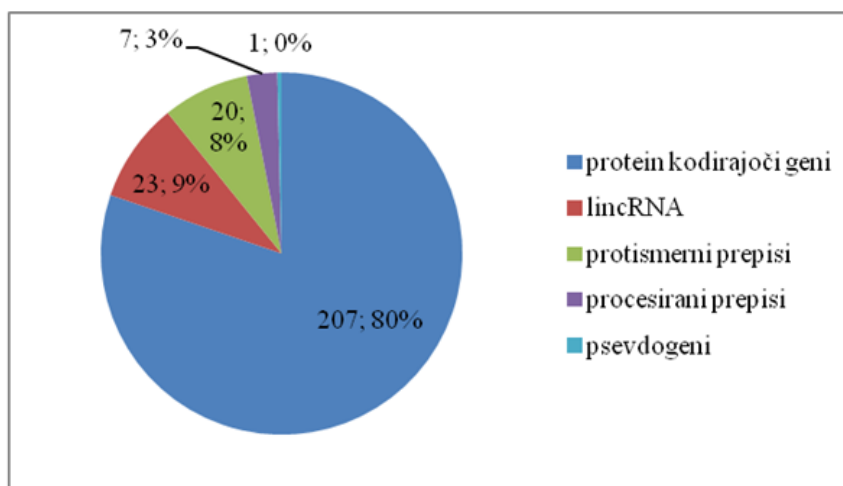
/ - ni imena gena

Zanimalo nas je, kakšno je število gostiteljskih genov za UCR-je pri posamezni vrsti: človeku, miši, podgani in govedu. Največ gostiteljskih genov za UCR-je smo našli pri človeku, in sicer 259. Pri miši je vseh gostiteljskih genov 223, pri podgani 176, pri govedu pa najmanj, 157. Večina gostiteljskih genov kodira proteine. Pri človeku je protein-kodirajočih gostiteljskih genov 207, pri miši 199, pri podgani 174, pri govedu pa 157 (Preglednica 3). Število gostiteljskih genov za UCR-je pri vseh štirih vrstah prikazuje slika 15.

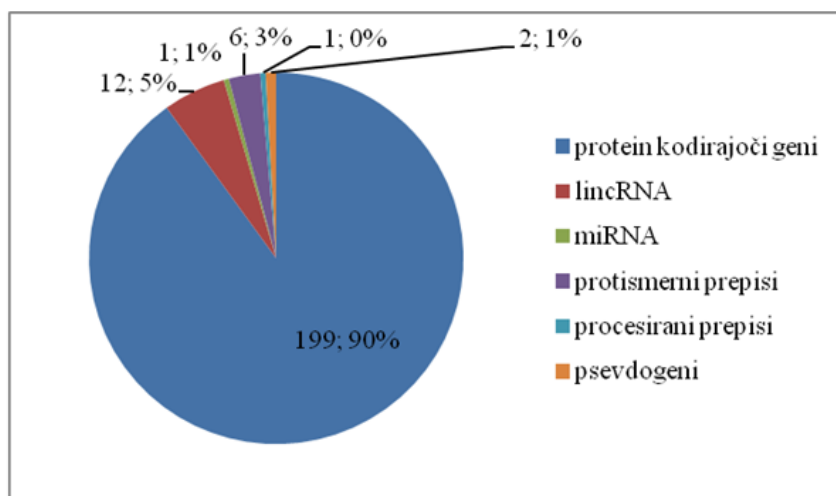


Slika 15: Število gostiteljskih genov za UCR-je pri vseh štirih vrstah (svetlo siva barva)
S temnejšo barvo je prikazano število protein-kodirajočih gostiteljskih genov.

Želeli smo ugotoviti, koliko gostiteljskih genov za UCR-je kodira druge tipe nkRNA, kamor uvrščamo lincRNA in miRNA. Pri človeku je gostiteljskih genov za druge tipe nkRNA 23, od tega so vse lincRNA, noben UCR pa se ne prekriva z miRNA. Pri miši je gostiteljskih genov za druge tipe nkRNA 13, od katerih je 12 lincRNA in ena miRNA. Pri podgani so vsi gostiteljski geni za UCR-je protein-kodirajoči geni, pri govedu pa gostiteljski gen za uc.420 kodira miRNA. Pri človeku je 20 protismernih prepisov, sedem procesiranih prepisov in en psevdogen, pri miši pa šest protismernih prepisov, dva psevdogena in en procesiran prepis. Pri podgani sta dva psevdogena. Razdelitev gostiteljskih genov na posamezne skupine pri človeku in miši prikazujeta sliki 16 in 17.

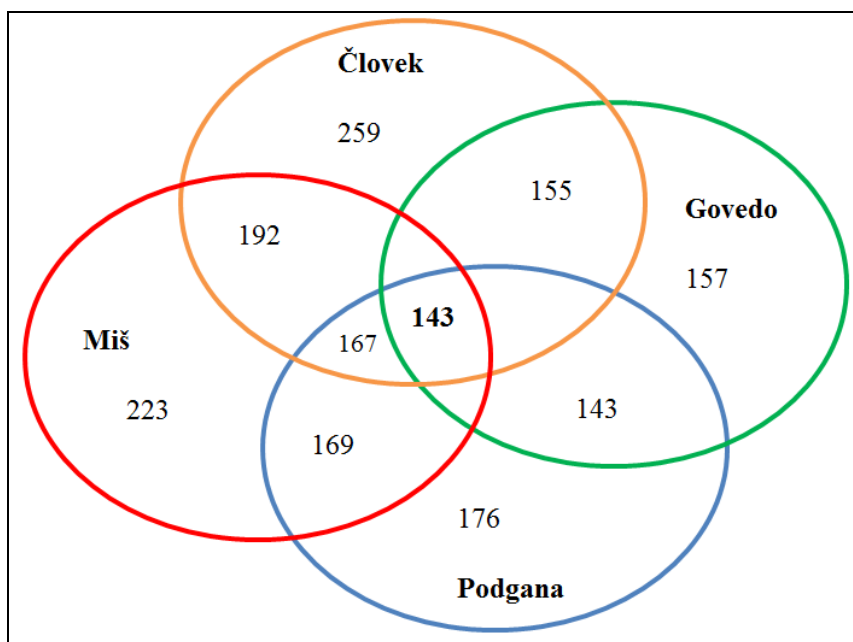


Slika 16: Razdelitev gostiteljskih genov za UCR-je na posamezne skupine pri človeku
Prikazano je število genov v posamezni skupini in njihov odstotni delež glede na skupno število gostiteljskih genov za UCR-je.



Slika 17: Razdelitev gostiteljskih genov za UCR-je na posamezne skupine pri miši. Prikazano je število genov v posamezni skupini in njihov odstotni delež glede na skupno število gostiteljskih genov za UCR-je.

Pri vseh štirih vrstah, človeku, miši, podgani in govedu, je kar nekaj gostiteljskih genov, ki se prekrivajo z več UCR-ji. Takih je pri človeku 70, pri miši 59, pri podgani 33 in pri govedu 29. Prav tako pri vseh štirih vrstah obstajajo UCR-ji, ki se prekrivajo z več različnimi gostiteljskimi geni. Pri človeku je takih UCR-jev 44, pri miši 21, pri podgani sedem in pri govedu trije. Med vsemi štirimi vrstami je ohranjenih 143 gostiteljskih genov. Med človekom, mišjo in podgano je ohranjenih nekoliko več gostiteljskih genov, 167, med človekom in mišjo pa je ohranjenih največ gostiteljskih genov, 192. Ohranjenost gostiteljskih genov za UCR-je med človekom, mišjo, podgano in govedom prikazuje slika 18.



Slika 18: Število ohranjenih gostiteljskih genov za UCR-je med človekom, mišjo, podgano in govedom.

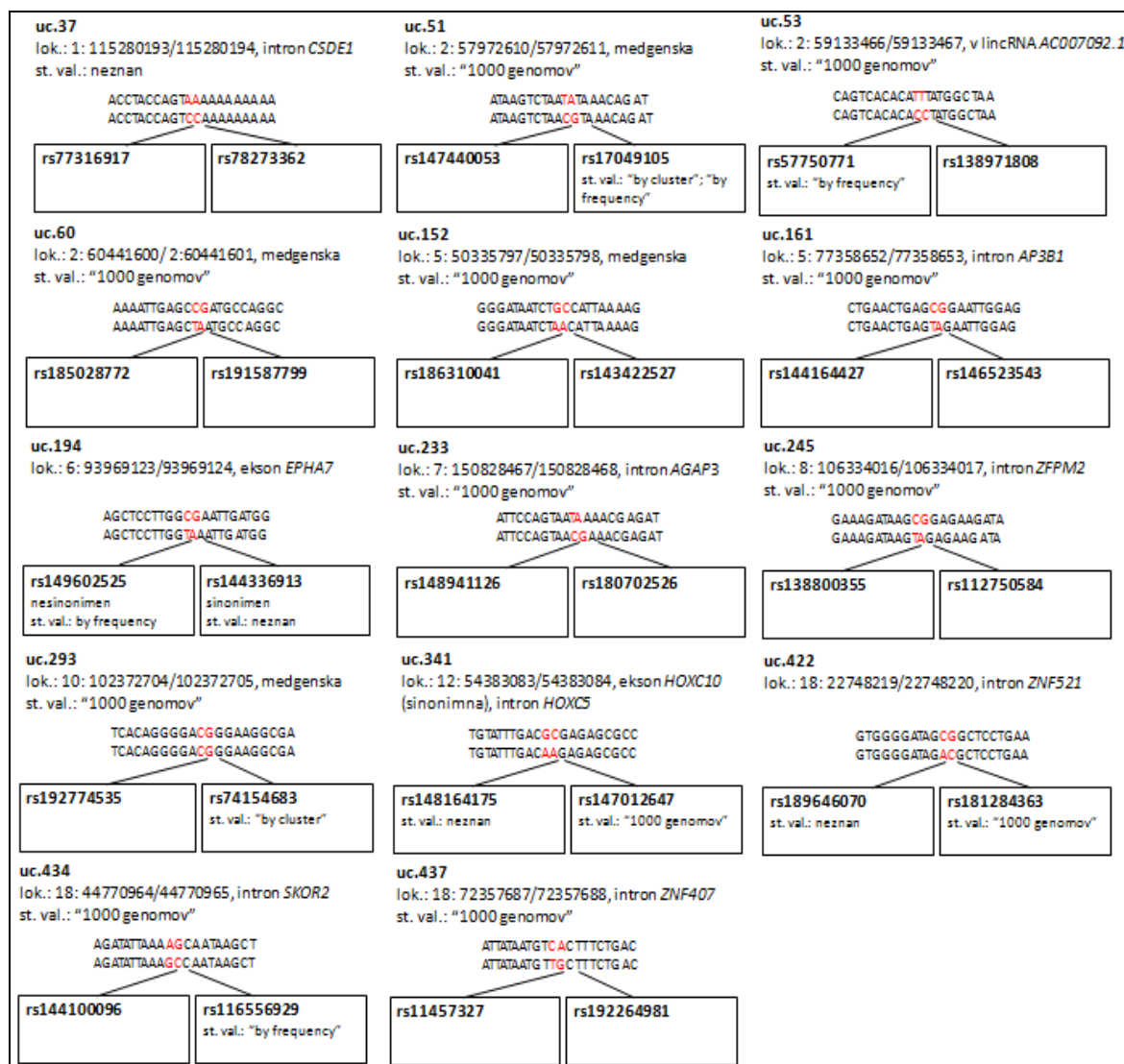
4.1.3 Iskanje genetske variabilnosti območij UCR

Ker smo želeli ugotoviti, kakšna je genetska variabilnost območij UCR, smo si pomagali z orodjem Biomart in poiskali polimorfizme znotraj UCR-jev pri človeku, miši, podgani in govedu. Da smo si olajšali delo in preverili rezultate ročnega načina iskanja polimorfizmov znotraj UCR-jev, smo polimorfizme v UCR-jih poiskali še na avtomatiziran način. Pri človeku in govedu je prišlo do razlik z ročnim preverjanjem, medtem ko jih pri miši in podgani ni bilo. Pri človeku je bila razlika med ročnim in avtomatiziranim načinom v dveh UCR-SNP-jih oz. je znašala manj kot 0,2 %. Orodje Biomart je našlo SNP-ja rs113975978 znotraj uc.166 in rs5789573 znotraj uc.320, naše avtomatizirano orodje pa ne. Pri govedu je bilo razlik več. Z ročnim načinom smo našli 179 UCR-SNP-jev, z avtomatiziranim pa 93. Razlik je bilo v 88 UCR-SNP-jih, kar predstavlja 49 %. Z orodjem Biomart smo našli 87 SNP-jev znotraj UCR-jev: uc.18, uc.20, uc.28, uc.70, uc.113, uc.122, uc.140, uc.141, uc.142, uc.166, uc.170, uc.206, uc.220, uc.233, uc.250, uc.252, uc.253, uc.270, uc.298, uc.301, uc.316, uc.319, uc.366, uc.420, uc.424, uc.444, uc.450, uc.468 in uc.470. Teh 87 UCR-SNP-jev avtomatizirano orodje ni našlo. Obratno smo z avtomatiziranim načinom našli en SNP znotraj uc.47, ki pa ga ročno nismo.

Genetske polimorfizme v UCR-jih smo opremili z različnimi informacijami. Za zbiranje podatkov o genetski variabilnosti območij UCR-jev pri človeku, miši, podgani in govedu smo uporabili podatkovne zbirke, opisane v materialih in metodah.

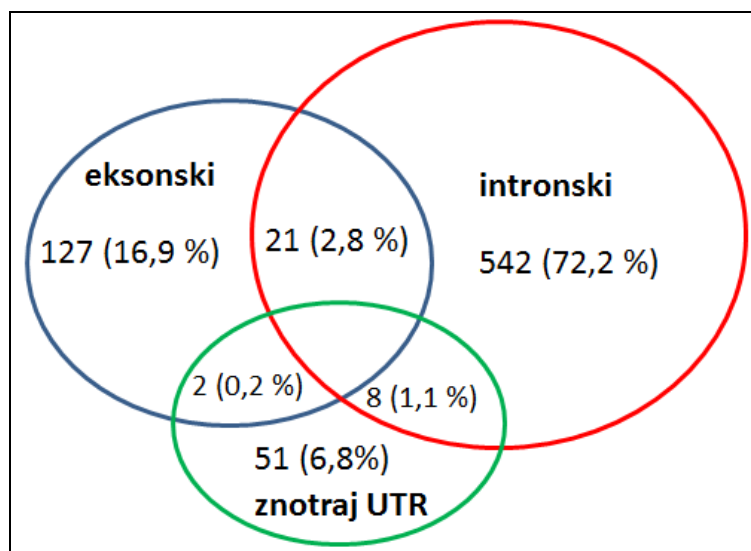
4.1.3.1 Iskanje genetske variabilnosti območij UCR pri človeku

Pri človeku smo znotraj 126007 bp UCR-jev našli 1180 genetskih polimorfizmov (1172 iz podatkovnih zbirk in 8 iz literature) oz. en genetski polimorfizem na 106,8 bp. Od 1180 genetskih polimorfizmov je bilo 1105 nukleotidnih zamenjav, 79 je bilo indel-ov, med katerimi je bilo največ delecij timina. Od odkritih 1180 genetskih polimorfizmov znotraj UCR-jev je bilo tudi 14 dinukleotidnih polimorfizmov (angl. DNP; *dinucleotide polymorphism*), ki so prikazani na sliki 19, in en večalelni SNP rs13171589 (A>C>G) znotraj uc.158.



Slika 19: Dinukleotidni polimorfizmi znotraj UCR-jev pri človeku
Lok. = lokacija, st. val. = status validacije, »1000 genomov« = projekt »1000 genomov«.

Večina genetskih polimorfizmov znotraj UCR-jev že ima znan status validacije, največ jih je bilo določenih v projektu »1000 genomov«. Podatke o statusu validacije smo pridobili iz baze dbSNP. Znotraj 259 različnih gostiteljskih genov se nahaja 816 UCR-SNP-jev, kar predstavlja skoraj 70 %. Najbolj polimorfen UCR je uc.478, znotraj katerega se nahaja 12 SNP-jev. Medgenskih UCR-SNP-jev je 353. Znotraj protein-kodirajočih gostiteljskih genov se nahaja 751 UCR-SNP-jev, ki smo jih razdelili na intronske, eksonske in tiste, ki ležijo znotraj UTR. Intronskih UCR-SNP-jev je 571, od tega je 29 UCR-SNP-jev takih, ki ležijo v intronu enega gena in hkrati v eksonu ali UTR drugega gena. Eksonskih UCR-SNP-jev je 150, od tega 23 takih, ki hkrati ležijo še v intronu ali UTR drugega gena. Znotraj UTR leži 61 UCR-SNP-jev, od tega deset takih, ki hkrati ležijo še v intronu ali eksonu drugega gena. Od 61 UCR-SNP-jev znotraj UTR jih 49 leži znotraj 3'UTR in 12 znotraj 5'UTR. Podatke o genetskih polimorfizmih znotraj UCR-jev smo dobili iz podatkovne zbirke Ensembl. Razdelitev UCR-SNP-jev, ki ležijo znotraj protein-kodirajočih gostiteljskih genov, prikazuje slika 20.



Slika 20: Razdelitev 751 UCR-SNP-jev, ki ležijo znotraj protein-kodirajočih genov, na intronske, eksonske in tiste, ki ležijo znotraj UTR. Slika prikazuje število in pripadajoči odstotek UCR-SNP-jev znotraj posamezne skupine.

Od 150 eksonskih UCR-SNP-jev je 92 nesinonimnih, 48 sinonimnih, sedem jih vpliva na premik bralnega okvira (angl. *frameshift coding*), trije pa kodirajo stop kodone (angl. *stop-gained coding*). Za nesinonimne UCR-SNP-je je predviden njihov učinek na funkcijo proteina. V podatkovni zbirki Ensembl so podatke o vplivu nesinonimnega UCR-SNP-ja na funkcijo proteina pridobili s pomočjo orodij PolyPhen-2 in SIFT. Orodje SIFT je napovedalo, da ima 40 nesinonimnih UCR-SNP-jev pri človeku škodljiv učinek na funkcijo proteina (angl. *deleterious*). Orodje PolyPhen pa je napovedalo, da ima 24 nesinonimnih UCR-SNP-jev pri človeku verjetno škodljiv učinek (angl. *probably damaging*), 13 pa mogoče škodljiv (angl. *possibly damaging*) učinek na funkcijo proteina. Od 13 nesinonimnih UCR-SNP-jev, za katere je orodje PolyPhen napovedalo, da so mogoče škodljivi, je šest takih, za katere je orodje SIFT napovedalo, da imajo škodljivi učinek na funkcijo proteina. Za ostalih sedem UCR-SNP-jev je orodje SIFT napovedalo, da zamenjava nima vpliva na funkcijo proteina (angl. *tolerated*). Preglednica 7 prikazuje UCR-SNP-je s predvidenim škodljivim učinkom na funkcijo proteina.

Preglednica 7: Pregled UCR-SNP-jev s predvidenim škodljivim učinkom na funkcijo proteina

UCR	UCR-SNP	Predviden učinek na funkcijo proteina	Orodje
uc.48	rs62123447 rs143781732	<i>deleterious</i> , <i>probably damaging</i>	SIFT PolyPhen
uc.49	rs140113540 rs62150472 rs111866703		
uc.61	rs61742690		
uc.135	rs112080036 rs190002250		

se nadaljuje

nadaljevanje preglednice 7

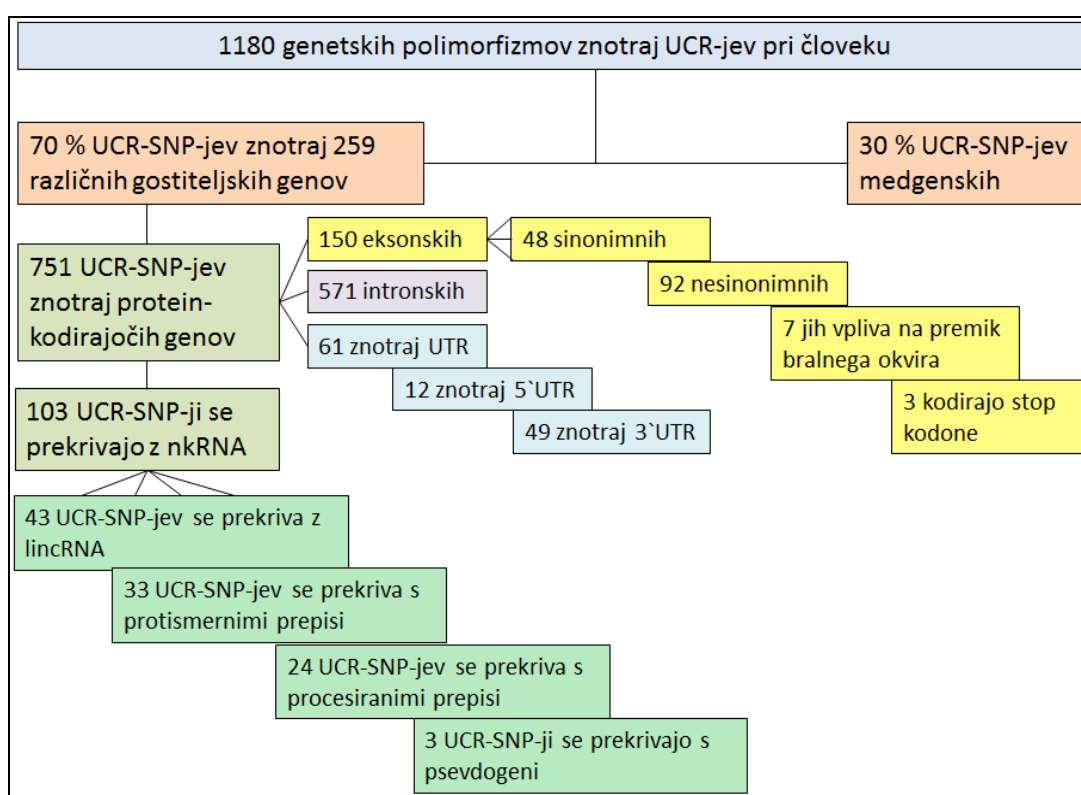
UCR	UCR-SNP	Predviden učinek na funkcijo proteina	Orodje
uc.174	rs148980401 rs190049582	<i>deleterious, probably damaging</i>	SIFT PolyPhen
uc.212	rs187162003		
uc.246	rs142281281		
uc.299	rs141576824		
uc.324	rs11556749 rs143559000		
uc.333	rs141864382		
uc.343	rs144307645 rs80157375		
uc.366	rs145027811		
uc.413	rs113351963 rs1139822		
uc.454	rs148423778		
uc.135	rs138869473	<i>deleterious, possibly damaging</i>	SIFT PolyPhen
uc.174	rs56286276		
uc.333	rs139994257		
uc.390	rs145063072		
uc.416	rs144236768		
uc.477	rs132630295		
uc.280	rs145687528 rs2149989	<i>deleterious</i>	SIFT
uc.285	rs137961156 rs142459122		
uc.331	rs138255175		
uc.343	rs141865901		
uc.344	rs35606176	<i>deleterious</i>	SIFT
uc.366	rs139630378		
uc.390	rs71400372		
uc.413	rs1139823		
uc.414	rs137853164		
uc.477	rs132630287		
uc.101	rs185888392	<i>possibly damaging</i>	PolyPhen
uc.183	rs149359755		
uc.213	rs145906139		
uc.194	rs149602525		
uc.338	rs192943135		

se nadaljuje

nadaljevanje preglednice 7

UCR	UCR-SNP	Predviden učinek na funkcijo proteina	Orodje
uc.393	rs142670255 rs151022979	<i>possibly damaging</i>	PolyPhen
uc.135	rs145677287	<i>probably damaging</i>	PolyPhen
uc.290	rs56241934		

Od 1180 UCR-SNP-jev jih 43 prekrivajo lincRNA, 33 protismerni prepisi, 24 procesirani prepisi, tri UCR-SNP-je pa prekrivajo psevdogeni. Slika 21 povzema rezultate iskanja genetskih polimorfizmov znotraj UCR-jev pri človeku.



Slika 21: Rezultati iskanja genetskih polimorfizmov znotraj UCR-jev pri človeku

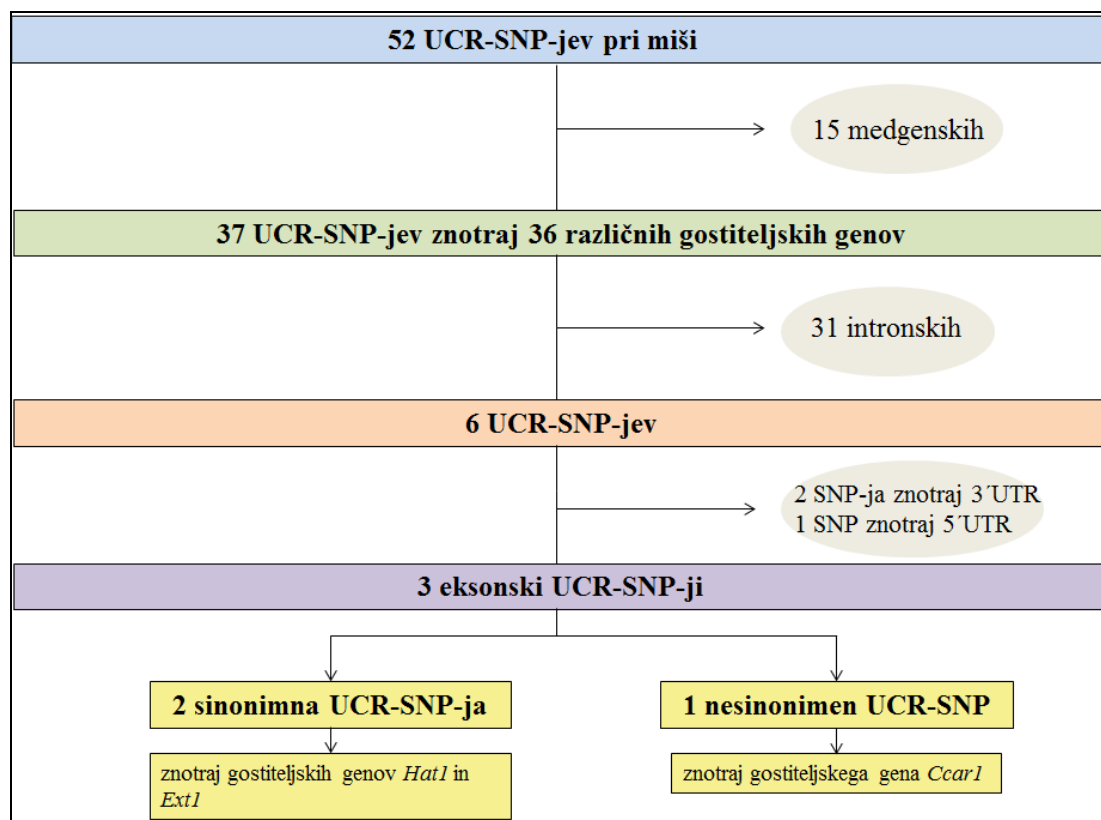
4.1.3.2 Iskanje genetske variabilnosti območij UCR pri miši

Pri miši smo znotraj UCR-jev našli 52 genetskih polimorfizmov, ki so vsi zamenjave posameznih nukleotidov. Eden izmed njih je večalelni SNP (T>C>A). Znotraj 36 različnih gostiteljskih genov se nahaja 37 UCR-SNP-jev. Najbolj polimorfen UCR pri miših je uc.207, znotraj katerega ležita dva SNP-ja. Oba sta intronska in ležita znotraj gena *Dnahc11*. En UCR-SNP, ki leži znotraj uc.332, je ohranjen med človekom in mišjo, kar pomeni, da leži na istem mestu pri obeh vrstah. Ta UCR-SNP je pri miši rs51296464 (A>G) in leži na kromosomu 9. Pri človeku se na istem mestu znotraj uc.332 nahaja rs185936189 (A>C) in leži na kromosomu 11. Poravnava genomov 13 vrst sesalcev z označenima UCR-SNP-jema pri človeku in miši je prikazana na sliki 22.

Homo sapiens	RTATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Pan troglodytes	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Gorilla gorilla	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Pongo abelii	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Macaca mulatta	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Callithrix jacchus	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Mus musculus	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Rattus norvegicus	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Oryctolagus cuniculus	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Bos taurus	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Sus scrofa	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Canis lupus familiaris	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT
Equus caballus	ATATTGCCAAGGCGGCTCTAATACTGCGACATTAAATTTTGCCTGAAAGATT

Slika 22: Poravnava genomov 13 vrst sesalcev v območju uc.332 z označenimi mesti UCR-SNP-jev Polimorfizma rs51296464 (Y) in rs185936189 (M) pri miši in človeku ležita na istem mestu v genomu. Levi R pri človeku predstavlja SNP rs180885255, desni pa rs73590085.

Ker smo želeli poiskati povezavo med identificiranimi UCR-SNP-ji in fenotipskimi lastnostmi različnih linij miši, smo UCR-SNP-je pri miši razvrstili po prioriteti. Razvrščanje UCR-SNP-jev po prioriteti smo izvedli na podlagi podatkov, ki smo jih pridobili o UCR-SNP-jih in njihovih gostiteljskih genih iz različnih podatkovnih zbirk. Vloge gostiteljskih genov in njihovo vključenost v različne biokemijske, signalne, metabolne in regulatorne poti smo pridobili s pomočjo orodij: KEGG (angl. *Kyoto Encyclopedia of Genes and Genomes*), Reactome, Panther, Biocarta. S pomočjo baze OMIM (angl. *Online Mendelian Inheritance in Man*) in znanstvenih člankov smo ugotovili, da je 18 gostiteljskih genov povezanih z različnimi boleznimi, 15 od njih z rakom. Polimorfizme posameznih nukleotidov miši smo preverili v podatkovnih zbirkah MGI (angl. *Mouse Genome Informatics*) in MPD (angl. *Mouse Phenome Database*). Tu smo pridobili podatke o statusu validacije nekaterih UCR-SNP-jev, za katere je bil status validacije v podatkovni zbirki dbSNP neznan. Prav tako smo tu dobili podatke o frekvenci alelov in genomski lokaciji UCR-SNP-ja (navzgor oz. navzdol od določenih genov, promotorskih regij, regulatornih elementov ...). Podatkovna zbirka MPD vsebuje tudi podatke iz orodja Vista Enhancer Browser (<http://enhancer.lbl.gov/>), ki napove, če je iskano zaporedje ojačevalcev. Preverili smo, če so ojačevalci medgenski UCR-SNP-ji pri miši, in ugotovili, da je od skupno 15 medgenskih UCR-SNP-jev osem UCR-SNP-jev ojačevalcev. Za ostalih sedem medgenskih UCR-SNP-jev njihova vloga ojačevalcev še ni dokončno potrjena. Prav tako nas je zanimala lokacija UCR-SNP-ja znotraj gostiteljskega gena. Na podlagi dobljenih podatkov smo UCR-SNP-je razvrstili na intronske, eksonske in tiste, ki se nahajajo znotraj regij UTR. Intronskih UCR-SNP-jev je 31, od tega štirje nimajo statusa validacije. Znotraj regij UTR ležijo trije SNP-ji. Eksonski UCR-SNP-ji so trije, in sicer dva sinonimna (rs28296350 in rs46759302) in en nesinonimen (rs37005101). Razvrščanje UCR-SNP-jev po prioriteti pri miši prikazuje slika 23.



Slika 23: Razvrščanje UCR-SNP-jev pri miši po prioriteti

4.1.3.3 Iskanje genetske variabilnosti območij UCR pri podgani

Pri podgani smo našli 11 UCR-SNP-jev, od katerih so vsi zamenjave nukleotidov. Status validacije je pri vseh neznan. Štirje UCR-SNP-ji so medgenski. Znotraj petih različnih gostiteljskih genov leži šest SNP-jev, od katerih sta dva intronska in pet eksonskih. Trije eksonski UCR-SNP-ji so nesinonimni, en je sinonimen, en pa kodira stop kodon. Najbolj polimorfen je uc.174, znotraj katerega ležijo trije UCR-SNP-ji.

4.1.3.4 Iskanje genetske variabilnosti območij UCR pri govedu

Pri govedu smo našli 179 UCR-SNP-jev, od tega je 164 nukleotidnih zamenjav in 15 indel-ov. Izmed 164 nukleotidnih zamenjav je tudi šest DNP-jev. Status validacije je pri vseh UCR-SNP-jih neznan. V uc.170 se nahaja največ SNP-jev, in sicer 40. Medgenskih UCR-SNP-jev je 111. Znotraj 39 različnih gostiteljskih genov se nahaja 68 UCR-SNP-jev. Intronskih UCR-SNP-jev je 44, eksonskih pa 22. Od tega je deset UCR-SNP-jev sinonimnih in 12 nesinonimnih. En UCR-SNP leži znotraj 5'UTR, en pa v 3'UTR. En UCR-SNP, ki leži znotraj uc.420, prekriva zrelo miRNA *ENSBTAG00000045885*.

4. 2 FUNKCIONALNA GENOMSKA KARAKTERIZACIJA OBMOČIJ UCR

Funkcionalna karakterizacija območij UCR obsega pregled predhodnih objav o UCR-jih in genetskih polimorfizmih znotraj UCR-jev v znanstveni literaturi. Funkcionalna analiza UCR-jev obsega tudi iskanje povezav med UCR-SNP-ji in značilnostmi linij miši.

4.2.1 Integracija podatkov s predhodnimi objavami v znanstveni literaturi

Za iskanje predhodnih objav o UCR-jih in genetskih polimorfizmih znotraj UCR-jev pri človeku, miši, podgani in govedu v znanstveni literaturi smo uporabili podatkovni zbirki Pubmed in OMIM. Objavo o UCR-jih smo našli pri miši, in sicer so Ahituv in sod. (2007) naredili transgene miši, tako da so jim iz genoma odstranili štiri UCR-je, in nato preučevali vpliv delecije UCR-jev na fenotip miši. Objave o UCR-SNP-jih smo našli še pri človeku, in sicer smo našli tako objave o SNP-jih, za katere ni bilo predhodne povezave z UCR-ji, kot objave o SNP-jih s predhodno povezavo z UCR-ji.

4.2.2 Iskanje povezav med UCR-SNP-ji in značilnostmi linij miši

Za UCR-SNP-je smo iz podatkovne zbirke za fenom miši (angl. MPD; *Mouse Phenome Database*) pridobili podatke o linijah miši in frekvenci alelov za posamezen genetski polimorfizem. Pregled UCR-SNP-jev s frekvencami alelov in linijami miši je prikazan v preglednici 8. Grafični prikaz frekvenc alelov za 48 UCR-SNP-jev prikazuje slika 24. Podatke smo iz podatkovne zbirke MPD dobili za 48 UCR-SNP-jev, za štiri UCR-SNP-je pa podatkov o alelih ni bilo. Najbolj polimorfen UCR-SNP je intronski UCR-SNP rs33008407 (G>C), pri čemer se alel G pogosteje pojavlja v populaciji miši. Alel G se pojavlja pri 15 linijah miši, alel C pa pri 11. Frekvenca alelov za rs33008407 je 0,577 : 0,423. Najmanj polimorfna sta dva intronska UCR-SNP-ja: rs30279798 in rs31128319, ki imata frekvenco alelov 0,962 : 0,038. Alel A se na mestu UCR-SNP-ja rs30279798 (G>A) pojavlja samo pri eni liniji miši – NOD/ShiLtJ. Prav tako se alel T na mestu UCR-SNP-ja rs31128319 (G>T) pojavi samo pri liniji miši WSB/EiJ.

Od 52 UCR-SNP-jev pri miši se trije UCR-SNP-ji nahajajo v eksonih gostiteljskih genov. Izmed treh eksonskih UCR-SNP-jev je rs37005101 nesinonimen, rs28296350 in rs46759302 pa sta sinonimna. Nesinonimni UCR-SNP rs37005101 (C>T) leži znotraj uc.285 na kromosomu 10 in ima znan status validacije. Nahaja se v eksonu gena *Ccar1*. Za izbrani nesinonimni UCR-SNP smo iz podatkovne zbirke MPD dobili podatke o alelih za 25 linij miši. Pri 23 linijah miši je na mestu UCR-SNP-ja rs37005101 alel T, pri dveh linijah miši, PWD/PhJ in PWK/PhJ, pa je na tem mestu alel C. Frekvenca alelov T : C je 0,92 : 0,08.

Sinonimni UCR-SNP rs28296350 (C>T) leži znotraj uc.97 v eksonu gena *Hat*. Podatke o alelih smo za rs28296350 dobili pri 26 linijah miši. Frekvenca alelov C : T za UCR-SNP rs28296350 je 0,65 : 0,35. Alel C se pojavlja pri 17 linijah miši, alel T pa pri devetih. Sinonimni UCR-SNP rs46759302 (A>G) leži znotraj uc.246 v eksonu gena *Ext*. Podatke o alelih smo za rs46759302 dobili pri 23 linijah miši. Pri 18 linijah miši se pojavlja alel G, pri petih pa alel A. Frekvenca alelov G : A za rs46759302 je 0,78 : 0,22.

Preglednica 8: Pregled 48 UCR-SNP-jev s frekvencami alelov in linijami miši

linije miši\SNP-ji	rs29685768	rs33241053	rs27938959	rs28296350	rs33654526	rs48914854	rs36894992	rs29813723	rs29900110	rs13469695	rs51309354	rs31283097	rs48819977	rs30279798	rs36727308	rs37023430	rs27769969	rs27749209	rs47295343	rs50387466	rs29806600	rs29968241	rs37768738	rs36307162
	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10	SNP11	SNP12	SNP13	SNP14	SNP15	SNP16	SNP17	SNP18	SNP19	SNP20	SNP21	SNP22	SNP23	SNP24
129P2/OlaHsd	T	/	C	C	T	/	/	/	G	G	C	T	C	G	T	/	T	C	A	A	T	C	A	C
129S1/SvImJ	T	/	C	C	T	T	G	C	G	G	C	T	C	G	T	A	T	C	A	A	T	C	A	C
129S5SvEvBrd	T	/	C	C	T	/	/	/	G	G	C	T	C	G	T	/	T	C	A	A	T	C	A	C
129X1/SvJ	/	/	/	C	T	/	/	/	/	/	/	T	/	G	/	/	/	/	/	/	T	C	/	/
A/J	T	G	C	T	T	T	G	G	G	G	C	C	C	G	T	A	T	C	A	A	T	C	A	C
AKR/J	T	G	C	T	T	/	G	/	G	G	C	T	C	G	T	/	T	C	A	A	C	T	A	C
BALB/cByJ	T	G	C	C	T	T	T	/	G	/	C	T	C	G	T	A	T	C	A	A	T	C	A	C
BALB/cJ	T	/	C	C	T	/	/	/	G	G	C	T	C	G	T	/	T	C	A	A	T	C	A	C
BTBR T+ tf/	/	/	C	C	T	T	G	/	G	/	C	T	C	G	T	/	T	C	A	A	T	T	A	/
C3H/HeJ	T	G	C	T	C	/	G	G	G	G	C	C	C	G	T	A	T	C	A	A	T	C	A	C
C57BL/6J	T	G	C	C	C	T	G	G	G	G	C	T	C	G	T	A	T	C	A	A	T	T	A	C
C57BL/6NJ	T	/	C	C	C	/	/	/	G	G	C	T	C	G	T	/	T	C	A	A	T	T	A	C
CAST/EiJ	C	G	C	C	C	/	G	G	G	G	A	T	C	G	C	/	G	C	T	T	T	T	G	C
CBA/J	T	/	C	C	C	/	/	/	G	G	C	C	C	G	T	/	T	C	A	A	T	C	A	C
DBA/2J	T	/	C	C	C	T	G	G	G	G	C	T	C	G	T	A	T	C	A	A	T	T	A	C
FVB/NJ	T	/	C	T	T	G	G	G	G	/	/	T	/	G	T	A	T	C	A	A	T	C	A	C
KK/H1J	/	/	C	C	T	T	G	/	G	/	C	T	C	G	T	A	T	C	A	A	C	T	A	/
LP/J	T	/	C	C	T	/	/	/	G	G	C	T	C	G	T	/	T	C	A	A	T	T	A	C
MOLF/EiJ	T	A	C	T	/	/	/	G	G	/	C	/	C	G	T	A	T	C	A	A	T	/	A	C

se nadaljuje

nadaljevanje preglednice 8

linije miši\SNP-ji	rs29685768	rs33241053	rs27938959	rs28296350	rs33654526	rs48914854	rs36894992	rs29813723	rs29900110	rs13469695	rs51309354	rs31283097	rs48819977	rs30279798	rs36727308	rs37023430	rs27769969	rs27749209	rs47295343	rs50387466	rs29806600	rs29968241	rs37768738	rs36307162
	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10	SNP11	SNP12	SNP13	SNP14	SNP15	SNP16	SNP17	SNP18	SNP19	SNP20	SNP21	SNP22	SNP23	SNP24
NOD/ShiLtJ	T	/	C	T	T	/	G	/	G	G	C	T	C	A	T	A	T	C	A	A	T	T	A	C
NZO/HILtJ	T	/	C	C	C	/	/	/	G	G	C	T	C	G	T	/	T	C	A	A	T	T	A	C
NZW/LacJ	/	/	C	T	T	T	G	G	G	/	/	T	C	G	T	T	T	C	A	A	T	T	A	C
PWD/PhJ	/	G	C	T	/	/	G	C	A	/	C	/	T	G	T	A	T	/	A	A	T	/	A	C
PWK/PhJ	T	/	C	T	C	/	/	/	A	A	C	T	T	G	T	/	T	C	A	A	T	T	A	C
SPRET/EiJ	T	/	C	C	C	/	/	/	G	G	C	T	C	G	A	/	T	C	A	A	T	T	A	C
WSB/EiJ	T	G	T	C	C	T	G	G	G	G	C	T	C	G	T	A	T	G	A	A	C	T	A	A
alel 1	T	G	C	C	C	T	G	G	G	G	C	T	C	G	T	A	T	C	A	A	T	T	A	C
n (alel 1)	13	5	16	9	9	6	10	6	14	10	16	15	14	16	16	9	18	14	18	18	15	15	18	15
frekvenca alela 1	0,93	0,71	0,84	0,47	0,56	0,86	0,91	0,67	0,74	0,77	0,94	0,94	0,74	0,84	0,94	0,75	0,95	0,82	0,95	0,95	0,79	0,94	0,95	0,83
alel 2	C	A	T	T	T	G	T	C	A	A	A	C	T	A	C	T	G	G	T	T	C	C	G	A
n (alel 2)	1	2	3	10	7	1	1	3	5	3	1	1	5	3	1	3	1	3	1	1	4	1	1	3
frekvenca alela 2	0,07	0,29	0,16	0,53	0,44	0,14	0,09	0,33	0,26	0,23	0,06	0,06	0,26	0,16	0,06	0,25	0,05	0,18	0,05	0,05	0,21	0,06	0,05	0,17
n (alel)	14	7	19	19	16	7	11	9	19	13	17	16	19	19	17	12	19	17	19	19	19	16	19	18

se nadaljuje

nadaljevanje preglednice 8

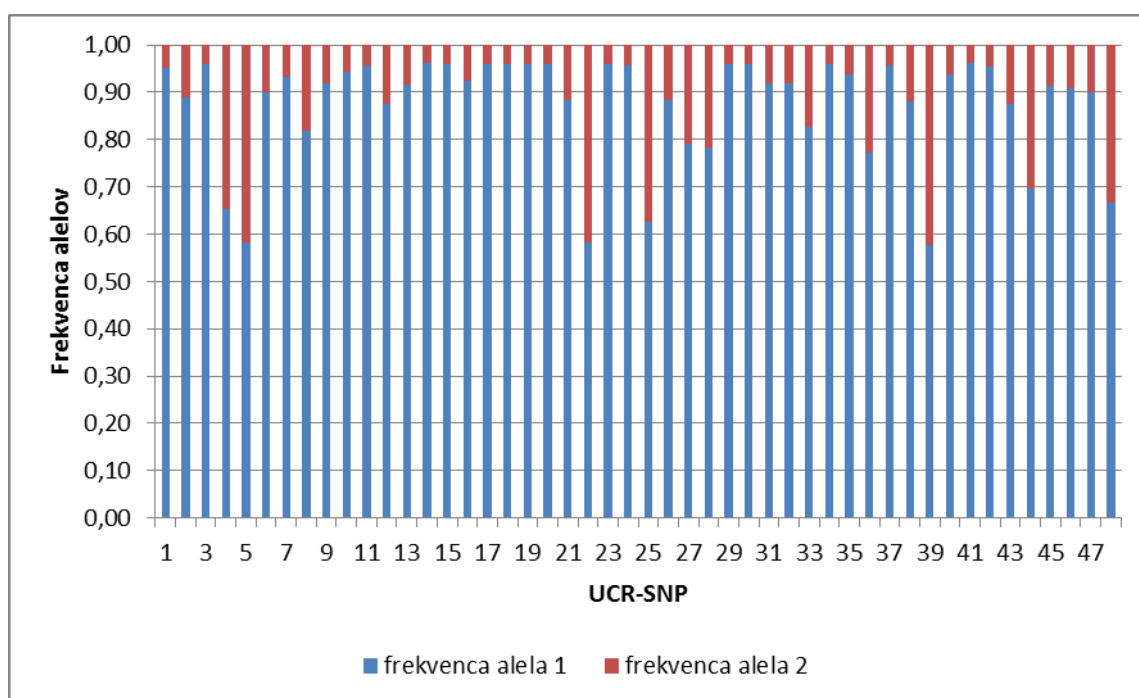
linije miši\SNP-ji	rs32131119	rs31140534	rs3707261	rs46759302	rs50733610	rs27159946	rs27192054	rs37005101	rs47973723	rs48287107	rs51296464	rs30192203	rs27456340	rs37191576	rs33008407	rs32209879	rs31128319	rs31424513	rs48294219	rs31856242	rs32568344	rs29071983	rs29939974	rs31309753
	SNP25	SNP26	SNP27	SNP28	SNP29	SNP30	SNP31	SNP32	SNP33	SNP34	SNP35	SNP36	SNP37	SNP38	SNP39	SNP40	SNP41	SNP42	SNP43	SNP44	SNP45	SNP46	SNP47	SNP48
129P2/OlaHsd	C	G	T	A	G	G	T	T	C	C	/	A	A	A	G	/	G	/	C	G	G	/	/	/
129S1/SvImJ	C	G	T	A	G	G	T	T	C	C	G	A	A	A	G	C	G	C	C	G	G	/	C	/
129S5SvEvBrd	C	G	T	A	G	G	T	T	C	C	/	A	A	A	G	/	G	/	C	G	G	/	/	/
129X1/SvJ	C	G	T	A	G	/	/	/	C	/	/	A	/	/	G	/	G	C	C	G	/	/	/	/
A/J	T	A	C	G	G	G	T	T	A	T	G	G	A	A	C	C	G	C	C	A	G	T	C	G
AKR/J	C	G	C	G	G	G	T	T	C	C	G	G	A	A	G	C	G	C	C	A	G	/	C	/
BALB/cByJ	C	G	C	G	G	G	T	T	A	C	G	G	/	A	G	C	G	C	/	A	G	T	/	/
BALB/cJ	C	G	C	G	A	G	T	T	A	C	/	G	A	A	G	/	G	C	C	A	G	/	/	/
BTBR T+ tf/	C	G	C	G	G	G	T	T	C	C	G	G	A	A	C	C	G	C	C	G	G	/	C	/
C3H/HeJ	C	A	C	G	G	G	T	T	C	C	G	G	A	A	C	C	G	C	A	G	G	/	/	/
C57BL/6J	T	G	C	G	G	G	T	T	C	C	G	G	A	A	C	C	G	C	C	G	G	T	C	A
C57BL/6NJ	T	G	C	G	G	G	T	T	C	C	/	G	A	A	C	/	G	/	C	G	G	/	/	/
CAST/EiJ	T	G	C	G	G	A	T	T	C	C	G	G	G	A	C	C	G	C	C	G	G	T	C	/
CBA/J	C	A	C	G	G	G	T	T	C	C	/	G	A	A	G	/	G	C	A	G	G	/	/	/
DBA/2J	C	G	C	G	G	G	T	T	A	C	G	G	A	A	G	C	G	C	A	G	G	T	/	G
FVB/NJ	C	G	C	G	G	G	T	T	/	C	G	/	/	A	G	C	G	C	C	A	G	T	C	/
KK/H1J	/	G	C	/	G	G	T	T	C	C	G	G	A	A	C	C	G	C	C	G	G	C	C	/
LP/J	C	G	T	A	G	G	T	T	C	C	/	A	A	A	G	/	G	C	C	A	G	/	/	/
MOLF/EiJ	C	G	/	/	/	G	T	T	/	C	A	/	A	C	C	C	G	T	C	/	/	T	T	/

se nadaljuje

nadaljevanje preglednice 8

linije miši\SNP-ji	rs32131119	rs31140534	rs3707261	rs46759302	rs50733610	rs27159946	rs27192054	rs37005101	rs47973723	rs48287107	rs51296464	rs30192203	rs27456340	rs37191576	rs33008407	rs32209879	rs31128319	rs31424513	rs48294219	rs31856242	rs32568344	rs29071983	rs29939974	rs31309753
	SNP25	SNP26	SNP27	SNP28	SNP29	SNP30	SNP31	SNP32	SNP33	SNP34	SNP35	SNP36	SNP37	SNP38	SNP39	SNP40	SNP41	SNP42	SNP43	SNP44	SNP45	SNP46	SNP47	SNP48
NOD/ShiLtJ	C	G	C	G	G	G	T	T	C	C	G	G	A	A	G	C	G	C	C	A	G	T	/	/
NZO/HILtJ	T	G	C	G	G	G	T	T	C	C	/	G	A	A	G	/	G	C	C	G	G	/	/	/
NZW/LacJ	T	G	C	G	G	G	T	T	C	C	G	/	A	A	G	C	G	C	/	/	/	T	/	/
PWD/PhJ	/	G	/	/	/	G	C	C	/	C	G	/	A	C	C	T	G	C	C	/	T	T	/	/
PWK/PhJ	T	G	C	G	G	G	C	C	C	C	/	G	A	C	C	/	G	C	C	G	T	/	/	/
SPRET/EiJ	T	G	C	G	G	G	T	T	C	C	/	G	A	A	C	/	G	/	C	G	G	/	/	/
WSB/EiJ	T	G	C	G	G	G	T	T	C	C	G	G	A	A	G	C	T	C	C	G	G	/	C	/
alel 1	T	G	C	G	G	G	T	T	C	C	G	G	A	A	G	C	G	C	C	G	G	T	C	G
n (alel 1)	9	23	19	18	23	24	23	23	19	24	15	17	22	22	15	15	25	21	21	16	21	10	9	2
frekvenca alela 1	0,38	0,88	0,79	0,78	0,96	0,96	0,92	0,92	0,83	0,96	0,94	0,77	0,96	0,88	0,58	0,94	0,96	0,95	0,88	0,70	0,91	0,91	0,90	0,67
alel 2	C	A	T	A	A	A	C	C	A	T	A	A	G	C	C	T	T	T	A	A	T	C	T	A
n (alel 2)	15	3	5	5	1	1	2	2	4	1	1	5	1	3	11	1	1	1	3	7	2	1	1	1
frekvenca alela 2	0,63	0,12	0,21	0,22	0,04	0,04	0,08	0,08	0,17	0,04	0,06	0,23	0,04	0,12	0,42	0,06	0,04	0,05	0,13	0,30	0,09	0,09	0,10	0,33
n (alel)	24	26	24	23	24	25	25	25	23	25	16	22	23	25	26	16	26	22	24	23	23	11	10	3

n – število; / – ni podatka; MPD – Mouse Phenome Database



Slika 24: Grafični prikaz frekvenc alelov za 48 UCR-SNP-jev pri miši

5 RAZPRAVA

5.1 FIZIČNA GENOMSKA KARAKTERIZACIJA OBMOČIJ UCR

5.1.1 Identifikacija UCR-jev pri različnih modelnih organizmih

Pri človeku, miši in podgani smo že znanim 481 UCR-jem določili nove genomske lokacije glede na trenutne verzije podatkovnih zbirk. Pri govedu UCR-ji še niso bili poznani, zato smo UCR-je na novo identificirali in jim določili odstotek identičnosti z zaporedji UCR pri človeku, miši in podgani. Vseh 481 UCR-jev smo razvrstili glede na njihovo dolžino. Več kot polovica UCR-jev je dolgih od 200 do 250 bp. Kot smo pričakovali, število UCR-jev z naraščanjem dolžine pada. Od 481 UCR-jev je najdaljši uc.462 z dolžino 779 bp. Pri vseh štirih vrstah (človek, miš, podgana in govedo) uc.462 leži na kromosomu X znotraj gena *POLA1*. Pri človeku se znotraj uc.462 nahaja pet intronskih UCR-SNP-jev, pri miši, podgani in govedu pa genetske variabilnosti znotraj najdaljšega UCR-ja nismo našli.

Pri govedu vsa zaporedja UCR niso 100-odstotno ohranjena z UCR-ji pri človeku, miši in podgani. Kljub temu je kar polovica UCR-jev 100-odstotno identična med človekom, mišjo, podgano in govedom. Zaporedje UCR, ki je najmanj ohranjeno z UCR-ji pri človeku, miši in podgani, je uc.457. Odstotek identičnosti je pri tem zaporedju še vedno zelo visok, in sicer kar 91,9 %. Zanimivo je, da je najmanj ohranjeno zaporedje tisto, ki je dolgo samo 211 bp. Zaporedje uc.457 pri človeku, miši in govedu leži znotraj gena *HIRA*. Genetske polimorfizme smo znotraj uc.457 našli samo pri človeku, in sicer en SNP. Vsi ostali UCR-ji pri govedu imajo več kot 95-odstotno identičnost z UCR-ji v genomih človeka, miši in podgane, kar še potrjuje njihovo visoko evolucijsko ohranjenost med različnimi sesalskimi vrstami in pomembno vlogo v različnih procesih.

Zaporedja UCR ležijo pri različnih vrstah na različnih kromosomih. Zanimivo je, da pri vseh štirih vrstah, človeku, miši, podgani in govedu, na kromosomu X ležijo enaki UCR-ji. To so UCR-ji od uc.459 do uc.482.

5.1.2 Prekrivanje UCR-jev z drugimi genetskimi elementi

Pri vseh štirih vrstah, človeku, miši, podgani in govedu, smo za UCR-je identificirali gostiteljske gene. Znotraj gostiteljskih genov *PBX3* in *POLA1* leži največ UCR-jev, in sicer osem. Zanimivo je, da se oba najdaljša UCR-ja, uc.462 z dolžino 779 bp in uc.464 z dolžino 770 bp, prekrivata z genom *POLA1*. Gen *POLA1* kodira katalitično podenoto DNA-polimeraze, ki skupaj z regulatorno in dvema podenotama primaze tvori DNA-polimerazni kompleks alfa. Katalitična podenota DNA-polimeraze ima bistveno vlogo pri začetku podvojevanja DNA (Dantzer in sod., 1998). Ravno zaradi pomembne vloge gena *POLA1* v celici je ta tako visoko ohranjen v genomu. Tudi gen *PBX3* prekriva osem UCR-jev. Gen *PBX3* kodira pre-levkemični prepišovalni dejavnik 3 in je vključen v uravnavanje izražanja razvojnih genov ter diferenciacijo urogenitalnih organov. Gen *PBX3* se visoko izraža v centralnem živčnem sistemu in v skorji adrenalnih žlez, kjer igra ključno vlogo v uravnavanju steroidogeneze. Ugotovili so, da je gen *PBX3* tarča miRNA *let-7d*, ki se znižano izraža v malignem tkivu prostate, kar pa se kaže v povišanem izražanju gena *PBX3*.

pri raku prostate. Gen *PBX3* je tarča še za eno miRNA *let-7c*, ki se znižano izraža v primarnem tkivu CRC, kar pa vodi v povišano izražanje gena *PBX3* pri CRC (Monica in sod., 1991; Ramberg in sod., 2011). V nadaljnjih raziskavah bi bilo zanimivo pogledati, če obstaja medsebojna odvisnost med dolžino gena ali cDNA in številom UCR-jev. Ali se v daljših genih nahaja več UCR-jev? Če se, bi to potrdilo »fizični moment«. Če povezave med dolžino gena in številom UCR-jev znotraj gena ni, očitno prevlada funkcijski element.

Največ gostiteljskih genov za UCR-je pri vseh vrstah kodira proteine. Ostali gostiteljski geni za UCR-je so gostiteljski geni, ki kodirajo psevdogene, protismerne in procesirane prepise, ter gostiteljski geni za druge tipe nkRNA, lincRNA in miRNA. Protein-kodirajočih gostiteljskih genov za UCR-je je pri človeku 207, gostiteljskih genov za druge tipe nkRNA pa 23. Iz razmerja med protein-kodirajočimi gostiteljskimi geni in geni za druge tipe nkRNA (207 : 23) lahko ugotovimo, da se znotraj protein-kodirajočih gostiteljskih genov nahaja devetkrat več UCR-jev kot znotraj genov za druge tipe nkRNA. Glede na to lahko sklepamo, da razporeditev UCR-jev v genomu ni naključna, saj se znotraj protein-kodirajočih gostiteljskih genov nahaja devetkrat več UCR-jev od pričakovanega. Pri miši je razmerje med protein-kodirajočimi gostiteljskimi geni za UCR-je in geni za druge tipe nkRNA še večje (199 : 13), in sicer se znotraj protein-kodirajočih gostiteljskih genov nahaja 15-krat več UCR-jev kot znotraj genov za druge tipe nkRNA. Kljub temu je treba upoštevati dejstvo, da najverjetneje še niso anotirani vsi gostiteljski geni za nkRNA. Največ gostiteljskih genov za UCR-je smo odkrili pri človeku, kar je pričakovano, saj je človeški genom najboljše anotiran. Nekaj manj gostiteljskih genov za UCR-je smo odkrili pri miši, še manj pri podgani, najmanj pa pri govedu. Gostiteljski geni, ki smo jih identificirali pri podgani, vsi kodirajo proteine. Pri govedu en gostiteljski gen za UCR-je kodira miRNA, ostali so protein-kodirajoči. Pri govedu je doslej poznanih in anotiranih najmanj genov, kar pa se bo v prihodnosti še spreminjalo.

Veliko gostiteljskih genov je ohranjenih med vsemi štirimi vrstami, človekom, mišjo, podgano in govedom, in sicer kar 143. Ta številka se bo v prihodnosti še spreminjala, saj sta genoma podgane in goveda še relativno slabo anotirana. Največ gostiteljskih genov, 192, je ohranjenih med človekom in mišjo, kar je pričakovano, saj je v človeškem in mišjem genomu poznanih največ genov.

Pri miši smo pogledali, v kakšne biološke poti so vključeni gostiteljski geni za UCR-SNP-je. Zelo veliko gostiteljskih genov za UCR-SNP-je je vključenih v uravnavanje izražanja genov, procesiranje in izrezovanje RNA, signalne poti, diferenciacijo in razvoj ter v celični cikel in apoptozo. Kar 15 od 36 gostiteljskih genov kaže povezanost z rakom.

5.1.3 Iskanje genetske variabilnosti območij UCR

Polimorfizme posameznih nukleotidov znotraj UCR-jev smo iskali ročno in na avtomatiziran način. Pri miši in podgani ni bilo razlik med obema načinoma iskanja genetskih polimorfizmov, pri človeku in govedu pa je prišlo do nekaterih odstopanj. Pri človeku je bila razlika med ročnim in avtomatiziranim načinom zelo majhna, saj je prišlo do razlik samo v dveh UCR-SNP-jih od skupno 1172. To znaša manj kot 0,2 %. Pri govedu je bilo razlik nekoliko več. Ročno smo našli več UCR-SNP-jev kot z avtomatiziranim načinom. Razlika je bila v 88 UCR-SNP-jih, kar predstavlja kar 49 %.

Vzrok za razlike med rezultati ročnega in avtomatiziranega načina bi bile lahko različne podatkovne zbirke, iz katerih avtomatiziran program črpa podatke, glede na podatkovno zbirko Biomart, ki smo jo uporabili pri ročnem načinu. Podatki in informacije med podatkovnimi zbirkami se namreč razlikujejo. Ostale vzroke za različne rezultate med ročnim in avtomatiziranim načinom je težko ugotoviti, saj nisem sodelovala pri izdelavi avtomatiziranega programa in zato ne poznam algoritmov, ki jih uporablja.

5.1.3.1 Iskanje genetske variabilnosti območij UCR pri človeku

Pri človeku smo znotraj UCR-jev našli zelo veliko genetskih polimorfizmov, in sicer en genetski polimorfizem na 106,8 bp. Veliko od teh genetskih polimorfizmov bi bilo potrebno še potrditi v laboratoriju, saj jih ima slabih 25 % UCR-SNP-jev status validacije neznan. Skoraj 70 % UCR-SNP-jev se nahaja znotraj gostiteljskih genov.

Najbolj polimorfen UCR je uc.478, znotraj katerega se nahaja 12 SNP-jev. Leži na kromosomu X, prekriva pa ga gostiteljski gen *GRIA3*. En UCR-SNP leži v intronu gena *GRIA3*, ostalih 11 UCR-SNP-jev pa v njegovih eksonih. Od eksonskih UCR-SNP-jev sta dva nesinonimna in devet sinonimnih. Gen *GRIA3* kodira ionotropni glutamatni receptor AMPA3, ki sodeluje pri transmembranskem transportu ionov in ob vezavi glutamata aktivira receptorje AMPA in NMDA. Nepravilnosti v delovanju ionotropnega glutamatnega receptorja AMPA3 so povezali z mentalno zaostalostjo, migrenami, samomorilskimi nagnjenji in celo z rakom (Géczi in sod., 1999).

5.1.3.2 Iskanje genetske variabilnosti območij UCR pri miši

Pri miši smo znotraj UCR-jev identificirali 52 genetskih polimorfizmov, od katerih so le trije eksonski, in sicer dva sinonimna in en nesinonimen. Sinonimni SNP rs28296350 leži znotraj gena *Hat1*, ki je vključen v modifikacije histonov (modifikacije kromatina, acetilacija histonov, utišanje kromatina) in je povezan z rakom (Verreault in sod., 1997). Drugi sinonimni SNP rs46759302 prekriva regijo, ki ima s promotorjem povezane regulatorne lastnosti (angl. *promoter associated regulation feature*) in leži znotraj gena *Ext1*. Ta je povezan z multiplo eksostozo tipa 1 in hondrosarkomom (Simmons in sod., 1999). Nesinonimni SNP rs37005101 leži 1100 bp navzgor od *Snord98* (angl. *small nucleolar RNA*) in znotraj gena *Ccar1*.

Velika težava pri izdelavi magistrske naloge so bile podatkovne zbirke podatkov, ki se stalno posodablja in spreminjajo. Posledično sem morala posodobiti vse svoje podatke in rezultate, kar mi je vzelo veliko časa. Informacije v podatkovnih zbirkah se med sabo tudi razlikujejo, kar oteži interpretacijo rezultatov, saj je težko vedeti, katera podatkovna zbirka vsebuje »pravo« informacijo.

5.2 FUNKCIONALNA GENOMSKA KARAKTERIZACIJA OBMOČIJ UCR

5.2.1 Integracija podatkov s predhodnimi objavami v znanstveni literaturi

Predhodno objavo o UCR-jih v znanstveni literaturi smo našli pri miši. Zaradi visoke ohranjenosti UCR-jev v genomih človeka, miši in podgane ter pomembnih funkcij UCR-

jev so Ahituv in sod. (2007) želeli ugotoviti, kakšen je pomen teh regij pri miših *in vivo*. Iz genoma miši so odstranili štiri UCR-je (uc.248, uc.329, uc.467 in uc.482), ki so dolgi od 222 do 731 bp. Da bi še povečali učinek delecije UCR-jev na fenotip, so izbrali UCR-je, ki delujejo kot ojačevalci in ležijo blizu genov, katerih spremenjeno izražanje ali utišanje vodi v izrazite spremembe fenotipov miši. Presenetljivo so vse miši z odstranjenimi UCR-ji iz genoma preživele, bile so plodne in niso kazale nobenih anomalij v rasti, življenjski dobi, patologiji in metabolizmu. Ti rezultati kažejo na to, da visoka ohranjenost zaporedij še ni nujno ključna za preživetje organizma (Ahituv in sod., 2007).

Pri človeku smo našli predhodne objave o genetskih polimorfozmih znotraj UCR-jev. Za večino SNP-jev, ki so bili analizirani v znanstvenih člankih, so raziskovalci vedeli, da se nahajajo znotraj zaporedij UCR. Analizirali so jih zato, da bi ugotovili, če imajo kakšno pomembno vlogo znotraj UCR-jev oz. če jih lahko povežejo z določenim fenotipskim znakom.

Shen in sod. (2011) so skušali povezati sedem SNP-jev (rs17049105, rs13020355, rs2682406, rs2056116, rs11190870, rs9572903 in rs8004379) znotraj uc.51, uc.82, uc.133, uc.140, uc.302, uc.353 in uc.368 z rakom dojke na kitajski populaciji, vendar povezave niso našli (Shen in sod., 2011). Chen in sod. (2007) so analizirali 16 SNP-jev (rs17049105, rs1861100, rs10496382, rs13020355, rs2056116, rs2056117, rs17291131, rs12981, rs3902936, rs3902937, rs7092999, rs2111796, rs9572903, rs7143938, rs4300725, rs11573440) znotraj uc.51, uc.53, uc.67, uc.82, uc.140, uc.211, uc.268, uc.295, uc.353, uc.374, uc.433 in uc.461 v referenčnem genomu šimpanza. V nasprotju s pričakovanji so odkrili, da niso vsi UCR-ji 100-odstotno ohranjeni v genomu šimpanza, kar je presenetljivo, saj je evolucijska razdalja med ljudmi in šimpanzi veliko krajša kot med ljudmi in glodalci (Chen in sod., 2007). Yang in sod. (2008) so preučevali vpliv šestih UCR-SNP-jev (rs1538101, rs1861100, rs2056116, rs7092999, rs7143938 in rs9572903) na pojavnost raka dojke. Ugotovili so, da sta z rakom dojke povezana dva UCR-SNP-ja; UCR-SNP rs9572903 je kazal samo mejno povezanost z razvojem raka dojke, prisotnost UCR-SNP-ja rs2056116 pa je bila višja pri bolnikih z rakom kot pri kontrolah. Ban in sod. (2005) so analizirali vpliv šestih SNP-jev (rs1538101, rs1861100, rs2056116, rs7092999, rs7143938, rs9572903) na pojavnost multiple skleroze, vendar z njenim razvojem niso povezali nobenega od šestih UCR-SNP-jev. Povezave z multiplo sklerozo niso uspeli dokazati niti za polimorfizem rs7143938, ki se nahaja v intronu razvojnega gena *MIPOL1*, katerega produkt lahko vpliva na imunski in živčni sistem.

V nasprotju z omenjenimi raziskavami so nekateri UCR-SNP-ji že bili omenjeni v znanstvenih člankih, vendar raziskovalci niso navedli, da se nahajajo znotraj UCR-jev. Takih UCR-SNP-jev je pet in se nahajajo v dveh UCR-jih: uc.299 in uc.477. Znotraj uc.299 se nahajata dva SNP-ja (rs77453353 in rs75462234), ki ležita v genu *PAX2*. Sta tudi klinično pomembna, saj so ju povezali s papilorenalnim sindromom (angl. *papillorrenal syndrome*) (Amiel in sod., 2000; Ford in sod., 2001; Sanyanusin in sod., 1995; Schimmenti in sod., 1997; Schimmenti in sod., 1999; Tellier in sod., 2000; Weaver in sod., 1988). Znotraj uc.477 se nahajajo trije SNP-ji (rs132630295, rs132630287 in rs132630292), ki ležijo v genu *PLP1*. Dva od teh SNP-jev (rs132630295 in rs132630287) so povezali s spastično paraplegijo 2 (angl. *spastic paraplegia 2*), enega (rs132630292) pa

z boleznijo Pelizaeus-Merzbacher (angl. *Pelizaeus-Merzbacher disease*) (Gorman in sod., 2007; Hodes in sod., 1997; Saugier-Veber in sod., 1994).

5.2.2 Iskanje povezav med UCR-SNP-ji in značilnostmi linij miši

Ker smo pri miši želeli poiskati povezavo med UCR-SNP-ji in značilnostmi linij miši, smo za 48 UCR-SNP-jev iz podatkovne zbirke MPD dobili podatke o alelih pri posameznih linijah miši. Za šest UCR-SNP-jev smo dobili podatke o alelih pri 26 linijah miši, pri ostalih UCR-SNP-jih pa pri manjšem številu linij. Frekvence alelov 48 UCR-SNP-jev kažejo na to, da SNP-ji znotraj UCR-jev pri miši niso zelo polimorfni, saj se pri 27 UCR-SNP-jih eden od alelov pojavi samo pri eni liniji miši.

Bolj podrobno smo povezave polimorfizmov z značilnosti linij miši iskali pri treh eksonskih UCR-SNP-jih. Nesinonimni UCR-SNP rs37005101 leži v eksonu gena *Ccar1*, ki ima pomembno vlogo v celičnem ciklu, uravnavanju apoptoze in izražanju genov. Protein *Ccar1* se poveže s koaktivatorskim kompleksom Mediator in p160, ki se kot odgovor na ustrezne hormone veže na jedrne receptorje tarčnih genov. Kompleks Mediator omogoči vezavo RNA-polimeraze II na promotorje protein-kodirajočih genov. Pomanjkanje endogenega proteina *Ccar1* zavre s hormoni posredovano izražanje jedrnih receptorjev tarčnih genov, s hormoni posredovano vezavo kompleksa Mediator in RNA-polimeraze II na promotorje tarčnih genov in od estrogena odvisno rast rakavih celic dojke. Gen *Ccar1* uravnava izražanje ključnih genov za proliferacijo in deluje kot koaktivator proteina p53. Kaže, da ima protein *Ccar1* dvojno vlogo; vpliva na proliferacijo celic, odvisno od estrogena, prav tako pa spodbuja apoptozo ob poškodbi DNA in drugih celičnih okvarah (Kim in sod., 2008).

Za nesinonimni UCR-SNP rs37005101 smo iz podatkovne zbirke MPD dobili podatke o alelih za 25 linij miši. Pri 23 linijah miši je na mestu UCR-SNP-ja rs37005101 alel T, pri dveh linijah miši, PWD/PhJ in PWK/PhJ, pa je na tem mestu alel C. Frekvenca alelov T : C je 0,92 : 0,08 oz. 92 % : 8 %. Liniji miši PWD/PhJ in PWK/PhJ sta visoko inbridirani liniji, ki izhajata iz podvrste divje miši *Mus musculus musculus*. Samci v F1 hibridni generaciji potomcev, ki nastanejo pri križanju linije miši PWD/PhJ z večino laboratorijskih linij miši, so sterilni. Pregled 37 biokemijskih označevalcev pri PWD/PhJ je razkril številne alelne razlike s standardnimi laboratorijskimi sevi miši. V primerjavi z laboratorijskimi linijami miši kažeta liniji PWD/PhJ in PWK/PhJ množico sprememb v mnogih kompleksnih fenotipih, kot so telesna masa, vzorec razporeditve maščobe v telesu, serumske koncentracije intermediarnih metabolitov, dovzetnost na diabetes tipa I in različni vzorci obnašanja. Nesinonimni UCR-SNP rs37005101 zato lahko povežemo s sestavo in težo telesa, rastjo, obnašanjem in koncentracijo metabolitov v krvi.

Sinonimni UCR-SNP rs28296350 leži v eksonu gena *Hat1*. Gen *Hat1* kodira histonsko aminotransferazo 1, ki acetilira novo sintetizirane histone, sodeluje pri modifikacijah kromatina, utišanju kromatina na predelu telomer, vključena pa naj bi bila tudi v signalno pot proteina EGFR1 in razvoj raka (Verreault in sod., 1997). Za sinonimni UCR-SNP rs28296350 smo dobili podatke o alelih za 26 linij miši. Frekvenca alelov C : T za rs28296350 je 0,65 : 0,35. Alel T se pojavlja pri sedmih linijah miši: A/J, AKR/J, C3H/HeJ, FVB/NJ, MOLF/EiJ, NOD/ShiLtJ, NZW/LacJ, PWD/PhJ in PWK/PhJ. Za linije

A/J, AKR/J, C3H/HeJ je značilna višja verjetnost nastanka tumorjev, kardiovaskularnih bolezni (ateroskleroza) in imunske pomanjkljivosti. Z imunskimi in vnetnimi boleznimi so povezane tudi linije miši FVB/NJ, NOD/ShiLtJ in NZW/LacJ. Za linije AKR/J, NOD/ShiLtJ, PWD/PhJ in PWK/PhJ je značilen razvoj diabetesa. Sklepamo, da sinonimni UCR-SNP rs28296350 vpliva na imunski in kardiovaskularni sistem.

Sinonimni UCR-SNP rs46759302 leži v eksonu gena *Ext1*. Gen *Ext1* kodira transmembransko glikoziltransferazo, ki je vključena v biosintezo heparan sulfata. Mutacije v genu *Ext1* lahko vodijo v nastanek multiple eksostoze tipa I in hondrosarkoma (Simmons in sod., 1999). Za sinonimni UCR-SNP rs46759302 smo dobili podatke za 23 linij miši. Frekvenca alelov A : G za rs46759302 je 0,22 : 0,78. Alel A se pojavlja pri petih linijah miši: 129P2/OlaHsd, 129S1/SvImJ, 129S5SvEvBrd, 129X1/SvJ in LP/J. Kar štiri od teh linij izhajajo iz družine 129, za katero so značilne dolga življenjska doba, visoka koncentracija holesterola v plazmi pri 12 in 24 tednih starosti, visoko razmerje med težo možganov in telesa, preferenca do slanega v primerjavi s sladkim okusom in slaba reprodukcija. Za linijo miši LP/J je značilna večja verjetnost nastanka tumorjev, epilepsija in okvare sluha. Sklepamo, da ima sinonimni UCR-SNP rs46759302 vpliv na reprodukcijo, koncentracijo lipidov v krvi, sluh in na razvoj tumorjev.

V prihodnosti bi bilo potrebno identificirane genetske polimorfizme znotraj UCR-jev pri vseh štirih vrstah (človeku, miši, podgani in govedu) potrditi v laboratoriju, saj ima veliko polimorfizmov še neznan status validacije. Pri miših bi bilo zanimivo primerjati, če so genetski polimorfizmi v UCR-jih razporejeni podobno kot ostali polimorfizmi med linijami miši glede na njihovo poreklo in genetsko razdaljo. S pomočjo rezultatov naloge bo v nadaljnje mogoče identificirati pomembne regulatorne regije, ki se pogosto nahajajo v ohranjenih območjih. Rezultati naloge bodo prav tako prispevali k razvijanju novih biooznačevalcev za bolezni in fenotipske lastnosti, novih tarč za zdravljenje in terapevtske posege pri človeku ter za razvoj selekcijskih označevalcev za bolezni in selekcijo pri domačih živalih.

6 SKLEPI

Ugotovili smo, da pri človeku poleg do sedaj identificiranih 30 SNP-jev znotraj UCR-jev obstajajo še dodatni UCR-SNP-ji. Znotraj UCR-jev smo pri človeku našli 1180 genetskih polimorfizmov (1172 iz podatkovnih zbirk in 8 iz literature) oz. en genetski polimorfizem na 106,8 bp. Od 1180 UCR-SNP-jev jih skoraj 70 % leži znotraj 259 gostiteljskih genov. Najbolj polimorfen UCR je uc.478, znotraj katerega se nahaja 12 SNP-jev. Od 1180 UCR-SNP-jev je tudi 14 DNP-jev, 40 UCR-SNP-jev pa ima škodljiv učinek (angl. *deleterious*) na funkcijo proteina.

S pomočjo bioinformatičnih orodij in analizo podatkovnih zbirk smo potrdili hipotezo, da se znotraj UCR-jev pri miši in podgani nahajajo genetski polimorfizmi. Pri miši smo znotraj UCR-jev odkrili 52, pri podgani pa 11 genetskih polimorfizmov. Tako kot pri človeku tudi pri miši približno 70 % UCR-SNP-jev leži znotraj gostiteljskih genov. Znotraj uc.332 se pri miši in človeku nahaja UCR-SNP, ki je ohranjen med obema vrstama, kar pomeni, da leži na istem mestu tako pri človeku kot pri miši. Pri miši je ta UCR-SNP rs51296464 (A>G) in leži na kromosomu 9. Pri človeku se na istem mestu znotraj uc.332 nahaja rs185936189 (A>C) in leži na kromosomu 11. Izmed 52 UCR-SNP-jev pri miši so trije eksonski, in sicer dva sinonimna (rs28296350 in rs46759302) in en nesinonimen (rs37005101).

Potrdili smo hipotezo, da SNP-ji znotraj UCR-jev pri miši vplivajo na različne lastnosti in bolezni. Na podlagi frekvenc alelov za UCR-SNP-je in lastnosti linij miši smo ugotovili, da nesinonimni UCR-SNP rs37005101 vpliva na sestavo in težo telesa, rast, vzorce obnašanja in koncentracijo metabolitov v krvi. Sinonimni UCR-SNP rs28296350 vpliva na imunski in kardiovaskularni sistem. Sinonimni UCR-SNP rs46759302 vpliva na reprodukcijo, koncentracijo lipidov v krvi, sluh in razvoj tumorjev.

Pri govedu smo na novo identificirali UCR-je in določili njihovo ohranjenost glede na UCR-je pri človeku, miši in podgani. Približno polovica zaporedij UCR je 100-odstotno identična med človekom, mišjo, podgano in govedom. Najmanj ohranjeno zaporedje UCR med govedom, človekom, mišjo in podgano je uc.457; odstotek identičnosti je pri tem zaporedju 91,9 %. Ostala zaporedja UCR pri govedu imajo več kot 95-odstotno identičnost z UCR-ji v genomih človeka, miši in podgane. Znotraj UCR-jev smo pri govedu našli 179 UCR-SNP-jev. Najbolj polimorfen UCR je pri govedu uc.170, znotraj katerega se nahaja 40 SNP-jev.

7.1 POVZETEK

V genomu so določili 481 UCR-jev, ki so identični med človekom, mišjo in podgano ter so daljši od 200 bp. Ti elementi so celo bolj ohranjeni kot regije, ki kodirajo proteine, kar kaže na njihovo pomembnost in vpletenost v ključne biološke vloge. Do sedaj so znotraj UCR-jev pri človeku odkrili 30 genetskih polimorfizmov. Spremenjeno izražanje UCR-jev lahko vodi v različne vrste raka. Zaporedja UCR se pogosto nahajajo na krhkih mestih genoma (angl. *fragile sites*) in na regijah, povezanih z rakom (angl. *cancer-associated genomic regions*; CAGR). Do sedaj so spremembe v UCR-jih že povezali z rakom debelega črevesa in dva SNP-ja z rakom dojke. Identificirali so tudi UCR-je, katerih izražanje je uravnano z miRNA in ki so nenormalno izraženi pri kronični limfocitni levkemiji. Ker je veliko UCR-jev izraženih v centralnem in perifernem živčnem sistemu, so poskušali povezati mutacije v UCR-jih tudi z multiplo sklerozo, vendar zaenkrat še neuspešno. Zaporedja UCR so nov razred genskih regulatornih elementov, njihova vloga pa še vedno ni dokončno znana. Zaradi vpletenosti UCR-jev v boleznih pri človeku, predvsem pri raku, bodo nadaljne študije prispevale k boljšemu razumevanju patogeneze bolezni in razvoju novih metod zdravljenja. Ker se podatki o genetski variabilnosti v podatkovnih zbirkah nenehno dopolnjujejo, je bil namen magistrske naloge analizirati, če se pri človeku znotraj UCR-jev nahajajo še dodatni genetski polimorfizmi. Z analizo podatkovnih zbirk smo želeli odkriti nove UCR-SNP-je pri miši in podgani. Želeli smo preveriti, če so UCR-SNP-ji pri miši povezani z lastnostmi in/ali boleznimi ter odkriti nove UCR-je pri govedu. Pri človeku smo znotraj UCR-jev odkrili 1180 genetskih polimorfizmov oz. en genetski polimorfizem na 106,8 bp. Od 1180 genetskih polimorfizmov (1172 iz podatkovnih zbirk in 8 iz literature) je 1105 SNP-jev, 14 DNP-jev, za 42 UCR-SNP-jev pa je predpostavljen škodljivi učinek na funkcijo proteina. Približno 70 % UCR-SNP-jev pri človeku prekriva gostiteljske gene, od katerih je največ protein-kodirajočih. Okrog 9 % gostiteljskih genov za UCR-je pri človeku predstavljajo geni za druge tipe nkRNA – lincRNA. Od UCR-SNP-jev, ki prekrivajo gostiteljske gene, je 74 % intronskih, 19 % eksonskih, 7 % pa jih leži znotraj UTR. Pri miši smo znotraj UCR-jev odkrili 52, pri podgani pa 11 genetskih polimorfizmov. Znotraj 36 gostiteljskih genov pri miši leži približno 70 % UCR-SNP-jev. Pri miši je med gostiteljskimi geni za UCR-je, ki kodirajo druge tipe nkRNA, poleg lincRNA tudi ena miRNA *Mir3064*, ki prekriva uc.420. Pri človeku in miši se na istem mestu znotraj uc.322 nahaja UCR-SNP, ki je ohranjen med obema vrstama. Pri človeku je ta UCR-SNP rs185936189 (A>C), pri miši pa rs51296464 (A>G). Izmed 52 UCR-SNP-jev pri miši je 31 intronskih, trije ležijo znotraj UTR, trije pa so eksonski, in sicer dva sinonimna (rs28296350 in rs46759302) in en nesinonimen (rs37005101). Ugotovili smo, da SNP-ji znotraj UCR-jev pri miši vplivajo na različne lastnosti in bolezni. Nesinonimni UCR-SNP rs37005101 vpliva na sestavo in težo telesa, rast, vzorce obnašanja in koncentracijo metabolitov v krvi. Sinonimni UCR-SNP rs28296350 vpliva na imunski in kardiovaskularni sistem. Sinonimni UCR-SNP rs46759302 vpliva na reprodukcijo, koncentracijo lipidov v krvi, sluh in razvoj tumorjev. Pri govedu smo na novo identificirali UCR-je in ugotovili, da jih je približno polovica 100-odstotno ohranjena z genomom človeka, miši in podgane. Rezultati naloge bodo služili za identifikacijo pomembnih regulatornih regij, ki se pogosto nahajajo v ohranjenih območjih. Prav tako bodo prispevali k razvijanju novih biooznačevalcev za bolezni in fenotipske lastnosti, novih tarč za zdravljenje in terapevtske posege pri človeku ter za razvoj selekcijskih označevalcev za bolezni in selekcijo pri domačih živalih.

7.2 SUMMARY

Previous studies identified 481 UCRs which are segments that are longer than 200 bp and are absolutely conserved between human, mouse and rat genomes. Overall, these genetic elements are even more highly conserved between species than proteins which strongly suggests that UCRs might play a critical regulatory function in the genome. Till now 30 genetic polymorphisms inside UCRs have been identified in the human genome. Recent studies show that the expression of UCRs is altered in tumours, suggesting UCRs may also be involved in human disease such as cancer. It is reported that UCRs are frequently located at fragile sites and genomic regions involved in cancers (cancer-associated genomic regions; CAGR). It is known that a deregulation of UCRs expression is involved in colorectal cancer. Two SNPs also show an association with familial breast cancer. Certain UCRs have been identified whose expression may be regulated by miRNA abnormally expressed in human chronic lymphocytic leukemia. Because many of UCRs are expressed in central and peripheral nervous system, some of the UCR-SNPs were tested for evidence of association with multiple sclerosis but unsuccessfully yet. It has been accepted that UCRs are a new class of genetic regulatory elements, although their function is still largely unknown. Therefore, further studies on UCRs will contribute to a better understanding not only of gene regulation in the human genome but also in the pathogenesis of human disease and development of novel treatment methods. Because the information about genetic polymorphisms in genome databases is continuously updated the aim of our study was to determine novel genetic polymorphisms inside UCRs in human genome. Additionally, we aimed to identify new genetic polymorphisms in UCRs of the mouse, rat and bovine genome. 1180 genetic polymorphisms were identified in the human genome or one polymorphism per 106,8 bp. Out of 1180 (1172 from databases and 8 from the literature) genetic polymorphisms 1105 are SNPs, 14 dinucleotide polymorphisms (DNPs) and 42 deleterious SNPs. In the human genome approximately 70 % of UCR-SNPs overlap the host genes, most of them are protein-coding. About 9 % of the host genes for UCRs are non-coding genes – lincRNA. In the human genome 74 % of UCR-SNPs are intronic, 19 % are exonic and 7 % of UCR-SNPs are located inside UTRs. We also uncovered 52 polymorphisms in UCRs in the mouse and 11 SNPs inside UCRs in rat genome. In the mouse genome about 70 % of UCR-SNPs are located inside 36 host genes. Beside non-coding host genes for UCRs (lincRNA) there is also a miRNA *Mir3064* overlapping uc.420 in mouse genome. Alignment between locations of genetic polymorphisms inside UCRs has revealed a conserved SNP in uc.322 between human (rs185936189; A>C) and mouse (rs51296464; A>G) genome. Out of 52 UCR-SNPs in mouse genome there are 31 intronic SNPs, three SNPs are located inside regions UTR and three SNPs are exonic, two of which are synonymous (rs28296350 and rs46759302) and one non-synonymous (rs37005101). We detected correlations between mouse SNPs in UCRs and phenotypic traits and diseases. Non-synonymous SNP rs37005101 within the *Ccar1* gene shows an effect on body weight and composition, growth, behaviour traits and concentration of metabolites in blood. Synonymous SNP rs28296350 within the *Hat1* gene shows an effect on immune and cardiovascular system. Synonymous SNP rs46759302 within the *Ext1* gene has an effect on reproduction, concentration of blood lipids, hearing and tumors development. We have identified UCRs in bovine and found out that about half of the UCRs in bovine are 100 % identical with the human, mouse and rat genomes. The bioinformatic analyses and collated data in the present study can serve researchers as a

starting point for experimental validation and identification of important regulatory regions and functional polymorphisms. These, in turn, can become new targets in therapeutic interventions and serve as diagnostic and selective markers for diseases and selection in breeding animals.

8 VIRI

- Ahituv N., Zhu Y., Visel A., Holt A., Afzal V., Pennacchio L. A., Rubin E. M. 2007. Deletion of ultraconserved elements yields viable mice. *Plos Biology*, 5, 9: 1906–1911.
- Ambros V. 2008. The evolution of our thinking about microRNAs. *Nature Medicine*, 14: 1036–1040.
- Amiel J., Audollent S., Joly D., Dureau P., Salomon R., Tellier A. L., Auge J., Bouissou F., Antignac C., Gubler M. C., Eccles M. R., Munnich A., Vekemans M., Lyonnet S., Attie-Bitach T. 2000. PAX2 mutations in renal-coloboma syndrome: mutational hotspot and germline mosaicism. *European Journal of Human Genetics*, 8: 820–826.
- Ban M., Maranian M., Yeo T. W., Gray J., Compston A., Sawcer S. 2005. Ultraconserved regions in multiple sclerosis. *European Journal of Human Genetics*, 13: 998–999.
- Bejerano G., Pheasant M., Makunin I., Stephen S., Kent W. J., Mattick J. S., Haussler D. 2004. Ultraconserved Elements in the Human Genome. *Science Express*, 304: 1321–1325.
- Braconi C., Valeri N., Kogure T., Gasparini P., Huang N., Nuovo G. J., Terracciano L., Croce C. M., Patel T. 2011. Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2: 786–791.
- Calin G. A., Liu C.-g., Ferracin M., Hyslop T., Spizzo R., Sevignani C., Fabbri M., Cimmino A., Lee E. J., Wojcik S. E., Shimizu M., Tili E., Rossi S., Taccioli C., Pichiorri F., Liu X., Zupo S., Herlea V., Gramantieri L., Lanza G., Alder H., Rassenti L., Volinia S., Schmittgen T. D., Kipps T. J., Negrini M., Croce C. M. 2007. Ultraconserved Regions Encoding ncRNAs Are Altered in Human Leukemias and Carcinomas. *Cancer Cell*, 12: 215–229.
- Chen C. T. L., Wang J. C., Cohen B. A. 2007. The Strength of Selection on Ultraconserved Elements in the Human Genome. *The American Journal of Human Genetics*, 80: 692–704.
- Dantzer F., Nasheuer H.-P., Vonesch J.-L., Murcia G., Ménissier-de Murcia J. 1998. Functional association of poly(ADP-ribose) polymerase with DNA polymerase α -primase complex: a link between DNA strand break detection and DNA replication. *Nucleic Acids Research*, 26, 8: 1891–1898.
- Ferdin J., Kunej T., Calin G. A. 2010. Non-coding RNAs: Identification of Cancer-Associated microRNAs by Gene Profiling. *Technology in Cancer Research and Treatment*, 9, 2: 123–138.
- Ford B., Rupps R., Lirenman D., Van Allen M. I., Farquharson D., Lyons C., Friedman J. M. 2001. Renal-coloboma syndrome: prenatal detection and clinical spectrum in a large family. *American Journal of Medical Genetics*, 99, 2: 137–141.
- Gécz J., Barnett S., Liu J., Hollway G., Donnelly A., Eyre H., Eshkevari H. S., Baltazarc R., Grunnc A., Nagarajae R., Gilliamc C., Peltonenf L., Sutherlanda G. R., Baronc M., Mulley J. C. 1999. Characterization of the Human Glutamate Receptor Subunit 3 Gene (GRIA3), a Candidate for Bipolar Disorder and Nonspecific X-Linked Mental Retardation. *Genomics*, 62, 3: 356–368.

- Gorman M. P., Golomb M. R., Walsh L. E., Hobson G. M., Garbern J. Y., Kinkel R. P., Darras B. T., Urion D. K., Eksioglu Y. Z. 2007. Steroid-responsive neurologic relapses in a child with a proteolipid protein-1 mutation. *Neurology*, 68, 16: 1305–1307.
- Guttman M., Donaghey J., Carey B. W., Garber M., Grenier J. G., Munson G., Young G., Bergstrom Lucas A., Ach R., Bruhn L., Yang X., Amit I., Meissner A., Regev A., Rinn J. L., Root D. E., Lander E. S. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, 477: 295–300.
- He C., Li Z., Chen P., Huang H., Hurst L. D., Chen J. 2012. Young intragenic miRNAs are less coexpressed with host genes than old ones: implications of miRNA–host gene coevolution. *Nucleic Acids Research*, 40, 9: 4002–4012.
- Hodes M. E., Blank C. A., Pratt V. M., Morales J., Napier J., Dlouhy S. R. 1997. Nonsense mutation in exon 3 of the proteolipid protein gene (PLP) in a family with an unusual form of Pelizaeus-Merzbacher disease. *American Journal of Medical Genetics*, 69, 2: 121–125.
- Jevšinek Skok D., Zorc M., Deržic S., Horvat S., Calin G. A., Kunej T. v pripravi.
- Kim J. H., Yang C. K., Heo K., Roeder R. G., An W., Stallcup M. R. 2008. CCAR1, a key regulator of Mediator complex recruitment to nuclear receptor transcription complexes. *Molecular Cell*, 31, 4: 510–519.
- Kunej T., Čeh E., Jevšinek Skok D., Ogorevc J., Ferdin J., Godnič I., Zorc M., Dovč P. 2012. Podatkovne zbirke in bioinformacijska orodja za raziskave mikro RNA. Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko: 66 str.
- Lin Z., Ma H., Nei M. 2008. Ultraconserved coding regions outside the homeobox of mammalian Hox genes. *BMC Evolutionary Biology*, 8, 260: 1–13.
- Lujambio A., Portela A., Liz J., Melo S. A., Rossi S., Spizzo R., Croce C. M., Calin G. A., Esteller M. 2010. CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. *Oncogene*, 29: 6390–6401.
- Maathuis M. H., Colombo D., Kalisch M., Bühlmann P. 2010. A method and server for predicting damaging missense mutations. *Nature Methods*, 7, 4: 248–249.
- Mestdagh P., Fredlund E., Pattyn F., Rihani A., Van Maerken T., Vermeulen J., Kumps C., Menten B., De Preter K., Schramm A., Schulte J., Noguera R., Schleiermacher G., Janoueix-Lerosey I., Laureys G., Powel R., Nittner D., Marine J. C., Ringner M., Speleman F., Vandesompele J. 2010. An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours. *Oncogene*, 29: 3583–3592.
- Monica K., Galili N., Nourse J., Saltman D., Cleary M. L. 1991. PBX2 and PBX3, New Homeobox Genes with Extensive Homology to the Human Proto-Oncogene PBXJ. *Molecular and Cellular Biology*, 11, 12: 6149–6157.
- Ng P. C., Henikoff S. 2001. Predicting Deleterious Amino Acid Substitutions. *Genome Research*, 11: 863–874.
- Ovcharenko I. 2008. Widespread Ultraconservation Divergence in Primates. *Molecular Biology and Evolution*, 25, 8: 1668–1676.
- Ramberg H., Alshbib A., Berge V., Svindland A., Taskén K. A. 2011. Regulation of PBX3 expression by androgen and Let-7d in prostate cancer. *Molecular Cancer*, 10, 50: 1–10.
- Ramensky V., Bork P., Sunyaev S. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Research*, 30, 17: 3894–390.

- Rossi S., Kopetz S., Davuluri R., Hamilton S. R., Calin G. A. 2009. MicroRNAs, ultraconserved genes and colorectal cancers. *The International Journal of Biochemistry and Cell Biology*, 42: 1291–1297.
- Sanyanusin P., McNoe L. A., Sullivan M. J., Weaver R. G., Eccles M. R. 1995. Mutation of PAX2 in two siblings with renal-coloboma syndrome. *Human Molecular Genetics*, 4, 11: 2183–2184.
- Saugier-Verber P., Munnich A., Bonneau D., Rozet J. M., Le Merrer M., Gil R., Boespflug-Tanguy O. 1994. X-linked spastic paraplegia and Pelizaeus-Merzbacher disease are allelic disorders at the proteolipid protein locus. *Nature Genetics*, 6, 3: 257–262.
- Scaruffi P., Stigliani S., Moretti S., Coco S., De Vecchi C., Valdora F., Garaventa A., Bonassi S., Tonini G. P. 2009. Transcribed-ultra conserved region expression is associated with outcome in high-risk neuroblastoma. *BMC Cancer*, 9: 441–450.
- Schimmenti L. A., Cunliffe H. E., McNoe L. A., Ward T. A., French M. C., Shim H. H., Zhang Y. H., Proesmans W., Leys A., Byerly K. A., Braddock S. R., Masuno M., Imaizumi K., Devriendt K., Eccles M. R. 1997. Further delineation of renal-coloboma syndrome in patients with extreme variability of phenotype and identical PAX2 mutations. *American Journal of Medical Genetics*, 60, 4: 869–878.
- Schimmenti L. A., Shim H. H., Wirtschafter J. D., Panzarino V. A., Kashtan C. E., Kirkpatrick S. J., Wargowski D. S., France T. D., Michel E., Dobyns W. B. 1999. Homonucleotide expansion and contraction mutations of PAX2 and inclusion of Chiari 1 malformation as part of renal-coloboma syndrome. *Human Mutation*, 14, 5: 369–376.
- Shen H., Lu C., Jiang Y., Tang J., Chen W., Zhang H., Zhang Q., Wang J. C., Liang J., Hu Z., Shen H. 2011. Genetic variants in ultraconserved elements and risk of breast cancer in Chinese population. *Breast Cancer Research and Treatment*, 128, 3: 855–861.
- Siepel A., Bejerano G., Pedersen J. S., Hinrichs A. S., Hou M., Rosenbloom K., Clawson H., Spieth J., Hillier L. W., Richards S., Weinstock G. M., Wilson R. K., Gibbs R. A., Kent W. J., Miller W., Haussler D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15: 1034–1050.
- Simmons A. D., Musy M. M., Lopes C. S., Hwang L.-Y., Yang Y.-P., Lovett M. 1999. A direct interaction between EXT proteins and glycosyltransferases is defective in hereditary multiple exostoses. *Human Molecular Genetics*, 8, 12: 2155–2164.
- Tellier A. L., Amiel J., Delezoide A. L., Audollent S., Auge J., Esnault D., Encha-Razavi F., Munnich A., Lyonnet S., Vekemans M., Attie-Bitach T. 2000. Expression of the PAX2 gene in human embryos and exclusion in the CHARGE syndrome. *American Journal of Medical Genetics*, 93, 2: 85–88.
- Verreault A., Kaufman P. D., Kobayashi R., Stillman B. 1997. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Current Biology*, 8: 96–108.
- Weaver R. G., Cashwell L. F., Lorentz W., Whiteman D., Geisinger K. R., Ball M. 1988. Optic nerve coloboma associated with renal disease. *American Journal of Medical Genetics*, 29, 3: 597–605.
- Wojcik S. E., Rossi S., Shimizu M., Nicoloso M. S., Cimmino A., Alder H., Herlea V., Rassenti L. Z., Rai K. R., Kipps T. J., Keating M. J., Croce C. M., Calin G. A. 2010. Non-codingRNA sequence variations in human chronic lymphocytic leukemia and colorectal cancer. *Carcinogenesis*, 31, 2: 208–215.

Yang R., Frank B., Hemminki K., Bartram C. R., Wappenschmidt B., Sutter C., Kiechle M., Bugert P., Schmutzler R. K., Arnold N., Weber B. H. F., Niederacher D., Meindl A., Burwinkel B. 2008. SNPs in ultraconserved elements and familial breast cancer risk. *Carcinogenesis*, 29, 2: 351–355.

Stoppard M. 2007. Zdravstveni vodnik za vso družino. Ljubljana, Vizita.si.
<http://vizita.si/clanek/leksikon/multipla-skleroza.html> (avgust, 2010).

ZAHVALA

Posebna zahvala gre prof. dr. Tanji Kunej za mentorstvo, strokovno usmerjanje in svetovanje pri pisanju magistrske naloge. Iskreno se ji zahvaljujem za njeno prijaznost, dobro voljo, potrpežljivost in spodbudo med nastajanjem naloge.

Daši Jevšinek Skok in Minji Zorc se zahvaljujem za vso podporo, potrpežljivost, koristne nasvete in pomoč pri izdelavi programov, ki jih sama ne bi znala narediti. S tem sta mi zelo olajšali delo in brez njiju naloga ne bi bila takšna, kot je.

Zahvaljujem se prof. dr. Simonu Horvatu za somentorstvo, koristne nasvete, predloge in usmeritve pri pisanju magistrske naloge.

Iskrena hvala fantu Maticu, ki mi je med študijem in ob vseh težkih trenutkih stal ob strani, me spodbujal in verjel vame. Hvala tudi za vse računalniške nasvete in pomoč.

Zahvaljujem se tudi svoji družini, ki me je podpirala, spodbujala in mi stala ob strani med celotnim študijem, še posebej mami Majdi, ki mi je nalogo zlektorirala.

PRILOGE

Priloga A: Pregled gostiteljskih genov za UCR-je pri človeku, miši in podgani ter SNP-jev znotraj posameznega UCR-ja

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.1	rs190053770, rs181426894	<i>PEX14</i>	/	<i>Pex14</i>	/	<i>Kif1b, Pex14</i>	peroxisomal biogenesis factor 14; kinesin family member 1B
uc.2	/	<i>CASZ1</i>	/	<i>Cas21</i>	/	<i>Kif1b</i>	kinesin family member 1B
uc.3	rs147621668, rs141631312, rs191263455, rs147072317, rs182214723	<i>CASZ1</i>	/	<i>Cas21</i>	/	<i>Kif1b</i>	castor zinc finger 1; kinesin family member 1B
uc.4	rs149004687, rs115141721, rs191491960, rs184287221, rs190398256, rs182389787, rs112980687	<i>CASZ1</i>	/	<i>Cas21</i>	/	<i>Kif1b</i>	castor zinc finger 1; kinesin family member 1B
uc.5	rs56121564, rs185631376	<i>CASZ1</i>	/	<i>Cas21</i>	/	/	castor zinc finger 1
uc.6	rs150506810, rs58824710, rs187762845	<i>CASZ1</i>	/	<i>Cas21</i>	/	/	castor zinc finger 1
uc.7	rs191035764, rs183976709, rs141418716	<i>CASZ1</i>	/	<i>Cas21</i>	/	/	castor zinc finger 1
uc.8	rs181670343, rs148793361, rs143358158	<i>CASZ1</i>	/	<i>Cas21</i>	/	/	castor zinc finger 1
uc.9	rs60400758	/	/	/	/	/	/
uc.10	rs114012169	/	/	/	/	/	/
uc.12	rs116497941	<i>SFPQ</i>	/	<i>Sfpq</i>	/	<i>Sfpq</i>	splicing factor proline/glutamine-rich
uc.13	/	<i>EIF2C1</i>	/	<i>Eif2c1</i>	/	<i>Eif2c1</i>	eukaryotic translation initiation factor 2C, 1
uc.14	rs186811602, rs115042546, rs191270944, rs141195286, rs145072960	/	/	/	/	/	/
uc.15	rs148358540, rs140543490, rs186031080, rs191352463, rs183064560	<i>RP5-884C9.2</i> (antisense)	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.16	rs185926611, rs188560967, rs181190474, rs186024081	/	/	/	/	/	/
uc.17	rs145399457	/	/	/	/	/	/
uc.18	rs79475765, rs192520588, rs148753309, rs116280768	<i>ERI3</i>	/	<i>Eri3</i>	/	/	exoribonuclease family member 3
uc.19	rs185314667, rs189737458, rs150809666	<i>RNF220</i>	/	<i>Rnf220</i>	/	/	ring finger protein 220
uc.20	rs186641479, rs35519518, rs71587055	<i>RNF220</i>	/	<i>Rnf220</i>	/	/	ring finger protein 220
uc.21	/	<i>AGBL4</i>	/	<i>Agbl4</i>	/	<i>F1LUW2_RAT</i>	ATP/GTP binding protein-like 4
uc.22	rs192560022, rs146520494, rs112195677, rs140961504	<i>FAF1</i>	/	<i>Faf1</i>	/	<i>Faf1</i>	Fas (TNFRSF6) associated factor 1
uc.23	rs183893736, rs57468024, rs189064368, rs138688315	<i>FAF1</i>	/	<i>Faf1</i>	/	<i>Faf1</i>	Fas (TNFRSF6) associated factor 1
uc.24	rs112586674, rs61781023	<i>FAF1</i>	/	<i>Faf1</i>	/	<i>Faf1</i>	Fas (TNFRSF6) associated factor 1
uc.25	/	<i>FAF1</i>	/	<i>Faf1</i>	/	<i>Faf1</i>	Fas (TNFRSF6) associated factor 1
uc.26	rs142176691	/	/	/	/	/	/
uc.27	rs41296151, rs147789427	/	/	/	/	/	/
uc.28	rs188168633	<i>SRSF11</i>	/	<i>Srsf11</i>	/	<i>F1MAM3_RA T</i>	serine/arginine-rich splicing factor 11
uc.29	rs34945188, rs12076829	<i>RP4- 544H6.2 (lincRNA)</i>	/	/	/	/	/
uc.30	/	/	/	/	/	/	/
uc.31	/	/	/	/	/	/	/
uc.33	rs151232160, rs34796807, rs12125403	<i>PTBP2</i>	/	<i>Ptbp2</i>	/	<i>Ptbp2</i>	polypyrimidine tract binding protein 2
uc.34	rs180870026	<i>PTBP2</i>	/	<i>Ptbp2</i>	/	/	polypyrimidine tract binding protein 2
uc.35	rs191664292, rs183168383, rs146839023	<i>DPYD</i>	/	<i>Dpyd</i>	/	<i>F1M891_RAT</i>	dihydropyrimidine dehydrogenase
uc.36	rs189944613, rs41301275, rs34438190, rs34664000	<i>PRPF38B</i>	rs29685768	<i>Prpf38b</i>	/	<i>Prpf38b</i>	PRP38 pre-mRNA processing factor 38 (yeast) domain containing B

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.37	rs190940416, rs142641678, rs77316917, rs78273362	<i>CSDE1</i>	/	<i>Csde1</i>	/	<i>CSDE1_RAT</i>	cold shock domain containing E1, RNA-binding
uc.38	rs114776633, rs12090689, rs116263201, rs66875870	/	/	/	/	/	/
uc.39	rs72688618, rs183153996, rs186399650	/	/	/	/	/	/
uc.40	rs7556157, rs117349984, rs111434935, rs41313884	<i>PBX1</i>	/	<i>Pbx1</i>	/	<i>Pbx1</i>	pre-B-cell leukemia homeobox 1
uc.41	/	/	/	/	/	/	/
uc.42	rs1857227, rs117489616, rs79728292, rs150308668, rs12132018	<i>USH2A</i>	/	<i>Ush2a</i>	/	/	Usher syndrome 2A (autosomal recessive, mild)
uc.43	rs12029205, rs144897294, rs56158846	<i>RP11-370K11.1</i> (processed transcript), <i>AKT3</i>	/	<i>Akt3</i>	/	/	v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)
uc.44	rs11588791, rs150420107, rs141329979, rs149523952, rs143195681	<i>ZNF238</i>	/	<i>Zfp238</i>	rs105454451, rs105778700	<i>Zfp238</i>	zinc finger protein 238
uc.45	rs15320	<i>HNRNPU-AS1</i> (antisense RNA), <i>HNRNPU</i>	/	<i>Hnrnpu</i>	/	/	HNRNPU antisense RNA 1 (non-protein coding); heterogeneous nuclear ribonucleoprotein U
uc.46	rs1802387, rs11709	<i>HNRNPU-AS1</i> (antisense RNA), <i>HNRNPU</i>	/	<i>Hnrnpu</i>	/	<i>Hnrnpu</i>	HNRNPU antisense RNA 1 (non-protein coding); heterogeneous nuclear ribonucleoprotein U
uc.47	rs145952615, rs139800258, rs142991839	/	/	/	/	/	/
uc.48	rs74393229, rs35439873, rs77758128, rs62123447, rs143781732	<i>PUM2</i>	/	<i>Pum2</i>	/	<i>Pum2</i>	pumilio homolog 2 (Drosophila)

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.49	rs62150472, rs140113540, rs181897533, rs111866703	<i>FAM98A</i>	/	<i>Fam98a</i>	/	<i>Fam98a</i>	family with sequence similarity 98, member A
uc.50	rs11547147	<i>SRSF7</i>	/	<i>Srsf7</i>	/	<i>Sfrs7</i>	serine/arginine-rich splicing factor 7
uc.51	rs147440053, rs17049105, rs188749292, rs112000154, rs142225496, rs192133539	/	/	/	/	/	/
uc.52	rs73942621, rs191945469	<i>AC007092.1</i> (lincRNA)	/	<i>5730522E02</i> <i>Rik</i> (protein coding)	/	/	/
uc.53	rs57750771, rs138971808	<i>AC007092.1</i> (lincRNA)	/	<i>5730522E02</i> <i>Rik</i> (protein coding)	/	/	/
uc.54	/	<i>AC007092.1</i> (lincRNA)	/	<i>5730522E02</i> <i>Rik</i> (protein coding)	/	/	/
uc.55	rs184360703, rs57908461, rs142264171	<i>AC007131.2</i> (processed transcript)	/	/	/	/	/
uc.56	/	<i>AC007131.2</i> (processed transcript)	/	/	/	/	/
uc.57	rs150384741, rs191820994	/	/	<i>Gm12068</i> (lincRNA)	/	/	/
uc.58	rs186141849, rs189107023	/	/	<i>Gm12068</i> (lincRNA), <i>4933430M0</i> <i>4Rik</i> (lincRNA)	/	/	/
uc.59	rs181500647	/	/	<i>Gm12068</i> (lincRNA), <i>4933430M0</i> <i>4Rik</i> (lincRNA)	/	/	/
uc.60	rs10186617, rs185028772, rs191587799, rs182987489	/	/	/	/	/	/
uc.61	rs61742690	<i>BCL11A</i>	/	<i>Bcl11a</i> , <i>4930538E20</i> <i>Rik</i> (antisense)	/	<i>D3ZSY3_RAT</i>	B cell CLL/lymphoma 11A (zinc finger protein)
uc.62	rs111946551, rs34642248	<i>BCL11A</i>	/	<i>Bcl11a</i>	/	<i>ENSRNOG000</i> <i>00042733</i> (pseudogen), <i>D3ZSY3_RAT</i>	B cell CLL/lymphoma 11A (zinc finger protein)
uc.63	rs143158075, rs180886367, rs186298975	<i>XPO</i>	/	<i>Xpo1</i>	/	<i>Xpo1</i>	exportin 1 (CRM1 homolog, yeast)

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.64	/	<i>EHBP1</i>	/	<i>Ehbp1</i>	/	<i>F1LXY4_RAT</i>	EH domain binding protein 1
uc.65	rs183905107, rs188687892, rs148174240	/	/	/	/	/	/
uc.66	rs149451031, rs183314214	<i>SFXN5</i>	/	<i>Sfxn5</i>	/	<i>D4A441_RAT</i>	sideroflexin 5
uc.67	rs10496382, rs116527523, rs10189742, rs35926232, rs189344245, rs181992322, rs151007589	/	rs33241053	/	/	/	/
uc.68	rs187926931	<i>ARHGAP15</i>	/	<i>Arhgap15</i>	/	<i>Arhgap15</i>	Rho GTPase activating protein 15
uc.69	rs141419384, rs150812052	<i>ARHGAP15</i> , <i>AC096558.1</i> (antisense), <i>RP11-570L15.2</i> (antisense)	/	<i>Arhgap15</i>	/	<i>Arhgap15</i>	Rho GTPase activating protein 15
uc.70	/	<i>ARHGAP15</i> , <i>RP11-434H14.1</i> (antisense)	/	<i>Arhgap15</i>	/	<i>Arhgap15</i>	Rho GTPase activating protein 15
uc.71	/	<i>AC016910.1</i> (antisense), <i>GTDC1</i>	/	<i>Gtdc1</i>	/	<i>Gtdc1</i>	glycosyltransferase-like domain containing 1
uc.72	rs112900080, rs146617980, rs141047480	<i>AC016910.1</i> (antisense), <i>GTDC1</i>	/	<i>Gtdc1</i>	/	<i>Gtdc1</i>	glycosyltransferase-like domain containing 1
uc.73	rs143711808, rs184697675, rs189288425, rs144162550	<i>GTDC1</i>	/	<i>Gtdc1</i>	/	<i>Gtdc1</i>	glycosyltransferase-like domain containing 1
uc.74	rs184858358, rs139610047, rs189120539, rs115830889, rs181447685	<i>GTDC1</i>	/	<i>Gtdc1</i>	/	<i>Gtdc1</i>	glycosyltransferase-like domain containing 1
uc.75	/	<i>ZEB2</i>	/	<i>Zeb2</i>	/	<i>ENSRNOG00000039025</i> (pseudogen)	zinc finger E-box binding homeobox 2
uc.76	rs187729797	<i>ZEB2</i>	/	<i>Zeb2</i>	/	<i>Zeb2</i>	zinc finger E-box binding homeobox 2
uc.77	rs142542447	<i>ZEB2</i>	/	<i>Zeb2</i>	/	<i>Zeb2</i>	zinc finger E-box binding homeobox 2
uc.78	rs149168368, rs35186800	<i>ZEB2</i>	/	<i>Zeb2</i>	/	<i>Zeb2</i>	zinc finger E-box binding homeobox 2
uc.79	rs145147754	<i>ZEB2</i>	/	<i>Zeb2</i>	/	<i>Zeb2</i>	zinc finger E-box binding homeobox 2
uc.80	rs183516368	<i>ZEB2</i>	/	<i>Zeb2</i>	/	<i>Zeb2</i>	zinc finger E-box binding homeobox 2

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.81	rs186006615, rs73009092, rs189206134, rs112670531	/	/	/	/	/	/
uc.82	rs34497887, rs13020355, rs140837085	/	/	/	/	/	/
uc.83	rs182759756, rs11896224	/	/	/	/	/	/
uc.84	/	NR4A2	/	Nr4a2	/	/	nuclear receptor subfamily 4, group A, member 2
uc.85	rs138818909	/	/	/	/	/	/
uc.86	rs79609723, rs145513622, rs139275163, rs144089793	/	rs27938959	/	/	/	/
uc.87	rs78458911, rs149769110	/	/	/	/	/	/
uc.88	rs111271017	AC009299.2 (antisense)	/	/	/	/	/
uc.89	/	PSMD14	/	Psmid14	/	Psmid14	proteasome (prosome, macropain) 26S subunit, non-ATPase, 14
uc.90	rs140860352	TBR1	/	Tbr1	/	Tbr1	T-box, brain, 1
uc.91	rs79076496, rs79982499	FAP	/	Fap	/	F1LMH7_RAT	fibroblast activation protein, alpha
uc.92	rs184607231, rs188187360, rs112773514	FIGN	/	/	/	Fign	fidgetin
uc.93	rs137869979, rs35009800	AC092684.1 (lincRNA)	/	/	/	/	/
uc.94	rs150970010	AC092684.1 (lincRNA)	/	/	/	/	/
uc.95	rs148257590, rs183919483, rs77203228, rs150109661, rs186745331	SP5	/	Sp5	/	Sp5	Sp5 transcription factor
uc.96	rs34912502	HAT1, SLC25A12	/	Hat1	/	Hat1	histone acetyltransferase 1; solute carrier family 25 (aspartate/glutamate carrier), member 12
uc.97	rs147062013, rs138153000, rs142923920, rs79541997	HAT1, SLC25A12	rs28296350	Hat1	/	Hat1	histone acetyltransferase 1; solute carrier family 25 (aspartate/glutamate carrier), member 12

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.98	/	/	/	<i>Dlx1as</i> (antisense RNA)	/	/	distal-less homeobox 1, antisense
uc.99	/	/	/	/	/	/	/
uc.100	rs146675902	<i>AC013461.1</i> (protein coding), <i>MLK7-AS1</i> (antisense RNA)	/	<i>B230120H2</i> <i>3Rik</i> (protein coding)	/	<i>Zak</i>	MLK7 antisense RNA 1 (non-protein coding); sterile alpha motif and leucine zipper containing kinase AZK
uc.101	rs185888392, rs13392217	<i>SP3</i>	/	<i>Sp3</i>	/	<i>Sp3</i>	Sp3 transcription factor
uc.102	rs113895258, rs138590546, rs111563353, rs145357368, rs150580441	<i>OLA1</i>	/	<i>Ola1</i>	/	<i>Ola1</i>	Obg-like ATPase 1
uc.103	/	<i>OLA1</i>	/	<i>Ola1</i>	/	<i>Ola1</i> , <i>Q7TPI6_RAT</i>	Obg-like ATPase 1
uc.104	/	<i>OLA1</i>	/	<i>Ola1</i>	/	<i>Ola1</i>	Obg-like ATPase 1
uc.105	/	<i>OLA1</i>	/	<i>Ola1</i>	/	<i>Ola1</i>	Obg-like ATPase 1
uc.106	/	<i>OLA1</i>	/	<i>Ola1</i>	/	<i>Ola1</i>	Obg-like ATPase 1
uc.107	rs144623604	/	/	<i>Sp9</i>	/	/	Sp9 transcription factor
uc.108	rs79981404, rs187871630, rs193240116, rs184519785, rs189455360, rs113546130, rs143743862, rs180694033, rs75304143	/	rs33654526	/	/	/	/
uc.109	rs34384113, rs192802675	<i>AC017048.4</i> (lincRNA)	/	/	/	/	/
uc.110	rs138763955, rs77187601	/	/	/	/	/	/
uc.111	rs35713478, rs34168198	<i>SETD5</i>	/	<i>Setd5</i>	/	<i>Setd5</i>	SET domain containing 5
uc.112	rs145785706, rs145439355, rs58040687	<i>TBC1D5</i>	/	/	/	/	TBC1 domain family, member 5
uc.113	rs183317163	/	/	/	/	/	/
uc.114	rs9816478, rs139911597	<i>AC105750.1</i> (lincRNA)	/	/	/	/	/
uc.115	rs4566562	/	/	/	/	/	/
uc.116	rs60930884, rs149763066, rs12486347	/	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.117	rs190659225, rs148980720, rs182823990, rs143780526, rs114719041	/	/	/	/	/	/
uc.118	rs138530814, rs75979941	/	/	/	/	/	/
uc.119	rs190023265, rs76579951, rs182250856, rs113152439	<i>ZBTB20</i>	/	<i>Zbtb20</i>	/	<i>Zbtb20</i>	zinc finger and BTB domain containing 20
uc.120	rs35243278, rs142361553	<i>ZBTB20</i>	/	<i>Zbtb20</i>	/	<i>Zbtb20</i>	zinc finger and BTB domain containing 20
uc.121	rs140756956	<i>ZBTB20</i>	rs48914854	<i>Zbtb20</i>	/	<i>Zbtb20</i>	zinc finger and BTB domain containing 20
uc.122	rs117910155	<i>ZBTB20</i>	/	<i>Zbtb20</i>	/	<i>Zbtb20</i>	zinc finger and BTB domain containing 20
uc.123	rs141996143, rs190025083, rs182252351, rs186625571	/	/	/	/	/	/
uc.124	rs188001721, rs80201959, rs77494121, rs73219093	/	/	/	/	/	/
uc.125	rs151145471, rs80194376, rs192713607	/	/	/	/	/	/
uc.126	rs150509502, rs186823000, rs139486070	/	/	/	/	/	/
uc.127	rs74512168, rs115633876, rs186427351, rs190970039, rs116165701	<i>RP11- 649A16.1 (antisense)</i>	/	/	/	/	/
uc.128	rs4490318, rs190318257	<i>RP11- 649A16.1 (antisense)</i>	/	/	/	/	/
uc.129	rs142636884, rs76277365	<i>MBNL1</i>	/	<i>Mbnl1</i>	/	<i>F1LPR3_RAT</i>	muscleblind-like splicing regulator 1
uc.130	/	<i>RP11- 290K4.2 (lincRNA)</i>	/	/	/	/	/
uc.131	rs139351639	<i>RSRC1</i>	/	<i>Rsrc1</i>	/	/	arginine/serine-rich coiled-coil 1
uc.132	rs113630780, rs77260280, rs117247632	<i>RSRC1</i>	rs36894992	<i>Rsrc1</i>	/	/	arginine/serine-rich coiled-coil 1
uc.133	rs2682406, rs181312926, rs151282528	<i>RSRC1</i>	/	<i>Rsrc1</i>	/	/	arginine/serine-rich coiled-coil 1

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.134	rs74767115, rs150733808	<i>RSRC1</i>	/	<i>Rsrc1</i>	/	/	arginine/serine-rich coiled-coil 1
uc.135	rs148137162, rs145677287, rs138337432, rs138869473, rs35594969, rs140534613, rs112080036, rs190002250	<i>MECOM</i>	/	<i>Mecom</i>	/	<i>Mecom</i>	MDS1 and EVI1 complex locus
uc.136	rs149268725, rs144470760, rs73174306, rs115730333	<i>RP11-641D5.2</i> (antisense RNA), <i>MECOM</i>	rs29813723	<i>Mecom</i>	/	/	MDS1 and EVI1 complex locus
uc.137	rs117034744, rs142217459, rs34700547	<i>CCDC39</i>	rs29900110	/	/	/	coiled-coil domain containing 39
uc.138	rs6784818, rs186122722, rs144330881	<i>TRA2B</i>	/	<i>Tra2b</i>	/	<i>Tra2b</i>	transformer 2 beta homolog (Drosophila)
uc.139	rs55884441	<i>STX18</i>	/	<i>Stx18</i>	/	<i>Stx18</i>	syntaxin 18
uc.140	rs149589194, rs114647501, rs2056116, rs143169624, rs2056117	/	/	/	/	/	/
uc.141	rs1801730, rs71854036, rs71730324, rs193193226, rs1801731	<i>DHX15</i>	rs13469695	<i>Dhx15</i>	/	<i>Dhx15</i>	DEAH (Asp-Glu-Ala-His) box polypeptide 15
uc.142	/	<i>RP11-227F19.1</i> (antisense), <i>PHOX2B</i>	/	<i>Phox2b</i>	/	<i>Phox2b</i>	paired-like homeobox 2b
uc.143	rs150020326, rs145238804	<i>G3BP2</i>	/	<i>G3bp2</i> , <i>Gm15710</i> (pseudogene)	/	<i>G3bp2</i>	GTPase activating protein (SH3 domain) binding protein 2
uc.144	/	<i>HNRPDL</i>	/	<i>Hnrpdl</i>	/	<i>D3ZN89_RAT</i>	heterogeneous nuclear ribonucleoprotein D-like
uc.145	rs150494696, rs138375935	/	rs51309354	/	/	/	/
uc.146	rs192397634, rs183913503	/	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.147	rs74459799, rs113756616, rs76927516, rs112241943, rs113296355, rs111282500, rs78877620	<i>LRBA</i>	/	<i>Lrba</i>	/	<i>F1LZG6_RAT</i>	LPS-responsive vesicle trafficking, beach and anchor containing
uc.148	/	<i>LRBA</i>	/	<i>Lrba</i>	/	<i>F1LZG6_RAT</i>	LPS-responsive vesicle trafficking, beach and anchor containing
uc.149	rs141918562	<i>LRBA</i>	rs31283097	<i>Lrba</i>	/	<i>F1LZG6_RAT</i>	LPS-responsive vesicle trafficking, beach and anchor containing
uc.150	rs192188964, rs146716769, rs144319752	<i>P11- 121L11.1</i> (lincRNA)	/	/	/	/	/
uc.151	/	<i>ZFR</i>	/	<i>Zfr</i>	/	<i>Zfr</i>	zinc finger RNA binding protein
uc.152	rs72758286, rs180730681, rs148827233, rs78806477, rs115999115, rs186310041, rs143422527	/	/	/	/	/	/
uc.153	rs112888451	<i>TNPO1</i>	/	<i>Tnpol</i>	/	<i>Tnpol</i>	transportin 1
uc.154	rs7879	<i>TNPO1</i>	/	<i>Tnpol</i>	/	/	transportin 1
uc.155	/	<i>OTP</i>	/	<i>Otp</i>	/	<i>Otp</i>	orthopedia homeobox
uc.156	rs139310356, rs143340412	<i>OTP</i>	/	/	/	/	orthopedia homeobox
uc.157	rs34752877	/	/	/	/	/	/
uc.158	rs13157302, rs13171589, rs13171597	<i>TBCA</i>	/	/	/	/	tubulin folding cofactor A
uc.159	rs35429261, rs139020094, rs78393253	<i>TBCA</i>	/	/	/	/	tubulin folding cofactor A
uc.160	rs150825065	/	/	/	/	/	/
uc.161	rs189837266, rs144164427, rs146523543, rs149023862, rs182409685	<i>AP3B1</i>	/	<i>Ap3b1</i>	/	<i>Ap3b1</i>	adaptor-related protein complex 3, beta 1 subunit
uc.162	rs182472668, rs188040102	<i>CTD- 2249K22.1</i> (lincRNA)	/	<i>A830009L08 Rik</i> (lincRNA)	/	/	/
uc.163	rs114651573, rs115220017, rs118112261, rs150488838	/	rs48819977	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.164	rs190343306, rs141207010, rs181546074	/	/	/	/	/	/
uc.165	rs11742553, rs78678674	<i>CTC-358I24.1</i> (lincRNA)	/	/	/	/	/
uc.166	rs189518044, rs150687433, rs113701909, rs140023994, rs113975978	<i>LINC00461</i> (lincRNA)	/	<i>C130071C03Rik</i> (lincRNA)	/	/	/
uc.167	rs148180366	<i>CTC-454M9.1</i> (lincRNA), <i>MEF2C</i>	/	<i>Mef2c</i>	/	/	myocyte enhancer factor 2C
uc.168	rs185824534	/	/	/	/	/	
uc.169	/	<i>RP11-65F13.2</i> (lincRNA), <i>NR2F1</i>	/	<i>A830082K12Rik</i> (lincRNA), <i>Nr2f1</i>	/	<i>Nr2f1</i>	nuclear receptor subfamily 2, group F, member 1
uc.170	/	<i>FAM172A</i>	/	<i>Fam172a</i>	/	<i>Fam172a</i>	family with sequence similarity 172, member A
uc.171	rs183181049	<i>KIAA0825</i>	rs30279798	<i>2210408I21Rik</i> (protein coding)	/	/	KIAA0825
uc.172	rs182167722, rs35849927	<i>KIAA0825</i>	/	<i>2210408I21Rik</i> (protein coding)	rs105802692	/	KIAA0825
uc.173	rs66724509, rs184830298	<i>UBE2B</i>	/	<i>Ube2b</i> , <i>Gm16953</i> (lincRNA)	/	<i>Ube2b</i>	ubiquitin-conjugating enzyme E2B
uc.174	rs148980401, rs190049582, rs56286276, rs192888077	<i>MATR3</i>	/	<i>Matr3</i>	rs105172356, rs8155490, rs8155487	<i>Matr3</i>	matrin 3
uc.175	rs187078138, rs191890759	<i>EBF1</i>	/	<i>Ebf1</i>	/	<i>Ebf1</i>	early B-cell factor 1
uc.176	rs182885583	<i>ODZ2</i>	/	<i>Odz2</i>	/	<i>Odz2</i>	odz, odd Oz/ten-m homolog 2 (Drosophila)
uc.177	rs142052059, rs77002625	<i>RANBP17</i>	/	<i>Ranbp17</i>	/	<i>Ranbp17</i>	RAN binding protein 17
uc.178	rs79540181, rs141109364	<i>RANBP17</i>	/	<i>Ranbp17</i>	/	<i>Ranbp17</i>	RAN binding protein 17
uc.179	rs183865013, rs77347996	<i>RANBP17</i>	/	<i>Ranbp17</i>	/	<i>Ranbp17</i>	RAN binding protein 17
uc.180	rs188527425	<i>RANBP17</i>	/	<i>Ranbp17</i>	/	<i>Ranbp17</i>	RAN binding protein 17
uc.181	rs35053326, rs147152482, rs140287869	<i>RANBP17</i>	/	<i>Ranbp17</i>	/	<i>Ranbp17</i>	RAN binding protein 17
uc.182	rs76623517	<i>RANBP17</i>	/	<i>Ranbp17</i>	/	<i>Ranbp17</i>	RAN binding protein 17

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.183	rs149359755	<i>FBXW11</i>	/	<i>Fbxw11</i>	/	<i>Fbxw11</i>	F-box and WD repeat domain containing 11
uc.184	/	<i>CPEB4</i>	/	<i>Cpeb4</i>	/	/	cytoplasmic polyadenylation element binding protein 4
uc.185	rs112874977, rs142915100, rs183667057	<i>CLK4</i>	/	<i>Clk4</i>	/	<i>Clk4</i>	CDC-like kinase 4
uc.186	rs181355904, rs3615, rs186428663	<i>HNRNPH1</i>	rs13461682	<i>Hnrnph1</i>	/	<i>Hnrph1</i>	heterogeneous nuclear ribonucleoprotein H1 (H)
uc.187	rs192318874, rs6923442	<i>TFAP2A</i>	/	/	/	/	transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)
uc.188	rs150423599, rs3834854	<i>ATXN1</i>	/	/	/	/	ataxin 1
uc.189	rs3208960, rs3208961, rs3208962	<i>PI16, SRSF3</i>	/	<i>Srsf3</i>	/	<i>Sfrs3</i>	peptidase inhibitor 16; serine/arginine-rich splicing factor 3
uc.190	rs151226504, rs139738238, rs140295784, rs145061523	<i>FOXP4</i>	/	<i>Foxp4</i>	/	<i>Foxp4</i>	forkhead box P4
uc.191	rs183865565	/	rs36727308	/	/	/	/
uc.192	rs150746229	/	rs37023430	/	/	/	/
uc.193	rs10223397, rs150956081, rs79750012	<i>SYNCRIP</i>	/	<i>Gm20537</i> (protein coding), <i>Syncrip</i>	rs107522357, rs105480159	/	synaptotagmin binding, cytoplasmic RNA interacting protein
uc.194	rs191903622, rs139211112, rs149602525, rs144336913	<i>EPHA7</i>	/	<i>Epha7</i>	/	<i>Epha7</i>	EPH receptor A7
uc.195	rs142332088, rs79771805	<i>MMS22L</i>	rs27769969	<i>Mms22l</i>	/	<i>RGD1304693</i> (protein coding)	MMS22-like, DNA repair protein
uc.196	rs184573900	/	/	/	/	/	/
uc.197	rs140569660, rs111236979, rs150477403, rs115515185	/	/	/	/	/	/
uc.198	rs182617815	/	/	<i>Gm11884</i> (lincRNA)	/	/	/
uc.199	rs187568687, rs189726566	/	rs27749209	/	/	/	/
uc.200	rs193098405	/	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.201	rs74770783, rs11433406, rs35218387, rs35695007, rs142391108	/	/	/	/	/	/
uc.202	rs183353839, rs185817301	ASCC3	/	Ascc3	/	F1LPQ2_RAT	activating signal cointegrator 1 complex subunit 3
uc.203	rs71569412, rs146558778, rs116587937	QKI	/	Qk	/	/	QKI, KH domain containing, RNA binding; quaking
uc.204	rs146674497	/	/	/	/	/	/
uc.205	rs184385875, rs41304918	/	/	/	/	/	/
uc.206	rs193271985, rs143686971, rs146797660, rs148923742	AC006481.1 (processed transcript)	/	/	/	/	/
uc.207	rs115980058, rs144620701, rs34915629	DNAH11	rs47295343, rs50387466	Dnahc11	/	E9PU24_RAT	dynein, axonemal, heavy chain 11
uc.208	rs180816601, rs140545554	TRA2A	/	Tra2a	/	Tra2a	transformer 2 alpha homolog (Drosophila)
uc.209	/	TRA2A	/	Tra2a	/	Tra2a	transformer 2 alpha homolog (Drosophila)
uc.210	rs186137141, rs75444887	/	/	/	/	/	/
uc.211	rs147209602, rs114690711, rs148617300, rs17291131	SKAP2	rs29806600	Skap2	/	Skap2	src kinase associated phosphoprotein 2
uc.212	rs187162003, rs146562288	HOXA2	/	Hoxa2, RP23- 103L13.8 (protein coding)	/	Hoxa2	homeobox A2
uc.213	rs145906139, rs149964084, rs143040629, rs146076178	HOXA-AS3 (antisense RNA), HOXA3, HOXA5	/	2700086A05 Rik (antisense), RP23- 103L13.8 (protein coding), Hoxa3, Hoxa5	/	F1M020_RAT, HXA5_RAT	HOXA cluster antisense RNA 3 (non-protein coding); homeobox A3; homeobox A5
uc.214	rs181600633, rs115780071, rs185718546, rs190428534, rs181279802	/	/	/	/	/	/
uc.215	rs138751100	GLI3	/	Gli3	/	Gli3	GLI family zinc finger 3

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.216	rs189757940	<i>IKZF1</i>	/	<i>Ikzf1</i>	/	<i>Ikzf1</i>	IKAROS family zinc finger 1
uc.217	/	<i>GS1-18A18.1</i> (processed transcript), <i>VSTM2A</i>	/	<i>Vstm2a</i>	/	<i>D3ZM70_RAT</i>	V-set and transmembrane domain containing 2A
uc.218	rs73173536	<i>AUTS2</i>	/	<i>Auts2</i>	/	/	autism susceptibility candidate 2
uc.219	rs57767351, rs189003572	<i>AUTS2</i>	/	<i>Auts2</i>	/	/	autism susceptibility candidate 2
uc.220	rs148816760, rs117486161, rs75288581, rs182247768, rs80292422	<i>DLX6-AS1</i> (antisense eRNA)	rs29968241	<i>Dlx6as1</i> (antisense RNA)	/	/	DLX6 antisense RNA 1 (non-protein coding)
uc.221	rs56175227	<i>DLX6-AS1</i> (antisense RNA)	/	<i>Dlx6as1</i> (antisense RNA)	/	/	/
uc.222	rs191014628, rs180893093	<i>AC073626.2</i> (antisense), <i>FOXP2</i>	/	<i>Foxp2</i>	/	/	forkhead box P2
uc.223	rs143840670, rs191732829, rs183250600	<i>AC073626.2</i> (antisense), <i>FOXP2</i>	/	<i>Foxp2</i>	/	/	forkhead box P2
uc.224	rs186584820, rs191436170, rs184032779	<i>FOXP2</i>	/	<i>Foxp2</i>	/	/	forkhead box P2
uc.225	rs141904641, rs78572788, rs139343994, rs144080515, rs191234894, rs180762891	<i>FOXP2</i>	/	<i>Foxp2</i>	/	/	forkhead box P2
uc.226	rs184157203, rs41311772, rs145907860	<i>FOXP2</i>	/	<i>Foxp2</i>	/	/	forkhead box P2
uc.227	rs188992761	<i>FOXP2</i>	/	<i>Foxp2</i>	/	<i>Foxp2</i>	forkhead box P2
uc.228	rs112954659, rs181695726	/	/	/	/	/	/
uc.229	/	/	/	/	/	/	/
uc.230	rs185896350, rs138127783, rs117536479	<i>RP11-458K10.2</i> (lincRNA)	/	/	/	/	/
uc.231	rs138110588, rs143565973, rs41303399	<i>TFEC</i>	/	<i>Tfec</i>	/	<i>Tfec</i>	transcription factor EC
uc.232	rs35005407, rs67446039, rs72148128, rs138407547	<i>RP11-560I19.4</i> (antisense), <i>FEZF1</i>	/	<i>Fezf1</i>	/	<i>Fezf1</i>	FEZ family zinc finger 1

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.233	rs2303946, rs148941126, rs180702526, rs186054083, rs143648199, rs189983164	AGAP3	/	<i>Agap3</i>	/	<i>Agap3</i>	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3
uc.234	rs57284979	/	/	/	/	/	/
uc.235	rs36098746, rs180971500, rs184330239	<i>EBF2</i>	/	4930438E09 <i>Rik</i> (processed transcript), <i>Ebf2</i>	/	<i>Ebf2</i>	early B-cell factor 2
uc.236	rs189955721, rs151048543, rs140978373	/	rs37768738	/	/	/	/
uc.237	rs185523527, rs112434932, rs189186014, rs78357199	<i>ST18</i>	rs36307162	<i>St18</i>	/	<i>St18</i>	suppression of tumorigenicity 18 (breast carcinoma) (zinc finger protein)
uc.238	rs142495076, rs150242598, rs189659310, rs145829026, rs149055472	<i>ST18</i>	/	<i>St18</i>	/	/	suppression of tumorigenicity 18 (breast carcinoma) (zinc finger protein)
uc.239	rs142283636, rs187783253	<i>TOX</i>	rs32131119	<i>Tox</i>	/	<i>Tox</i>	thymocyte selection- associated high mobility group box
uc.240	rs189871454, rs182034675	<i>RP11- 21C4.1</i> (lincRNA)	/	/	/	/	/
uc.241	rs145165836	/	/	/	/	/	/
uc.242	rs117813405	/	rs31140534	/	/	/	/
uc.243	rs191727525, rs73237566, rs146458568, rs183894209	<i>ZFHx4</i>	/	<i>Zfhx4</i>	/	<i>Zfhx4</i>	zinc finger homeobox 4
uc.244	rs142768056, rs190555563, rs151067592, rs182249696, rs186926591, rs141009416	<i>RP11- 127H5.1</i> (protein coding)	/	/	/	/	/
uc.245	rs80245989, rs190865307, rs183886735, rs138800355, rs112750584, rs34047814, rs189242141	<i>RP11- 127H5.1</i> (protein coding), <i>ZFPM2</i>	rs3707261	<i>Zfpm2</i>	/	<i>Zfpm2</i> , <i>F1LV48_RAT</i>	zinc finger protein, multitype 2

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.246	rs142281281, rs139785421, rs61757381, rs191550540	<i>EXT1</i>	rs46759302	<i>Ext1</i>	/	<i>Ext1</i>	exostosin 1
uc.247	rs139708661, rs142924484, rs190792292, rs146115020	/	rs50733610	/	/	/	/
uc.248	rs41314215	/	/	/	/	/	/
uc.249	rs138102451, rs186813768, rs76042111	/	/	/	/	/	/
uc.250	rs77916072, rs72715812, rs184161180	<i>C9orf146</i> (lincRNA)	/	/	/	/	long intergenic non-protein coding RNA 583
uc.251	rs112496573, rs10962189	<i>C9orf93/CCDC171</i>	/	<i>4930473A06 Rik</i> (protein coding)	/	<i>RGD1309931</i> (protein coding)	coiled-coil domain containing 171
uc.252	rs1538101, rs183192107, rs12551811	<i>BNC2</i>	/	<i>Bnc2</i>	/	<i>Bnc2</i>	basonuclin 2
uc.253	rs191292228	<i>CNTLN</i>	/	<i>Cntln</i>	/	<i>F1LVK7_RAT</i>	centlein, centrosomal protein
uc.254	rs45624833, rs191574834	/	/	/	/	/	/
uc.255	rs183269647	<i>ELAVL2</i>	/	<i>Elavl2</i>	/	<i>Elavl2</i>	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)
uc.256	rs7847710	<i>ELAVL2</i>	/	<i>Elavl2</i>	/	<i>Elavl2</i>	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)
uc.257	rs185395451	<i>ZCCHC7</i>	/	<i>Zcchc7</i>	/	<i>Zcchc7</i>	zinc finger, CCHC domain containing 7
uc.258	rs148850137, rs143468386, rs192848882	<i>ZCCHC7</i>	/	<i>Zcchc7</i>	/	<i>Zcchc7</i>	zinc finger, CCHC domain containing 7
uc.259	rs144530044	/	/	/	/	/	/
uc.260	rs139430020, rs186163238, rs188276506, rs180800303	/	/	/	/	/	/
uc.261	rs140009626, rs41312178, rs144892289, rs190612894	/	/	/	/	/	/
uc.262	rs150451931, rs191042505, rs183400905, rs187526508, rs192804180	/	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.263	/	<i>RP11-575L7.8</i> (antisense), <i>HNRNPK</i>	/	<i>Hnrnpk</i>	/	<i>Hnrnpk</i>	heterogeneous nuclear ribonucleoprotein K
uc.264	/	<i>RP11-575L7.8</i> (antisense), <i>HNRNPK</i>	/	<i>Hnrnpk</i>	/	<i>Hnrnpk</i>	heterogeneous nuclear ribonucleoprotein K
uc.265	rs142281369, rs41306484, rs140434459, rs148680326, rs113045741	<i>SLC44A1</i>	/	<i>Slc44a1</i>	/	<i>F1LYW4_RAT</i> , <i>Slc44a1</i>	solute carrier family 44, member 1
uc.266	/	<i>RP11-308N19.4</i> (lincRNA)	/	<i>Gm12478</i> (lincRNA)	/	/	/
uc.267	rs145685186, rs57065013	<i>MRRF</i>	rs27159946	<i>Mrrf</i>	/	<i>Mrrf</i>	mitochondrial ribosome recycling factor
uc.268	rs188387883, rs1131758, rs193025380, rs3902936, rs3902937, rs56528822, rs12981, rs185603861	<i>RC3H2</i>	/	<i>Rc3h2</i>	rs8146984	/	ring finger and CCCH-type domains 2
uc.269	rs1752168, rs7871805	<i>DENND1A</i>	/	<i>Dennd1a</i>	/	/	DENN/MADD domain containing 1A
uc.270	rs148710527, rs141972813, rs146382639, rs139760593, rs114529341, rs181063974, rs113788992	<i>MAPKAP1</i>	rs27192054	<i>Mapkap1</i>	/	<i>Mapkap1</i>	mitogen-activated protein kinase associated protein 1
uc.271	rs78715270, rs185617540	<i>MAPKAP1</i>	/	<i>Mapkap1</i>	/	<i>Mapkap1</i>	mitogen-activated protein kinase associated protein 1
uc.272	rs149050332, rs112751374	<i>MAPKAP1</i>	/	<i>Mapkap1</i>	/	<i>Mapkap1</i>	mitogen-activated protein kinase associated protein 1
uc.273	rs10986880, rs190931273	<i>PBX3</i>	/	<i>Pbx3</i>	/	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3
uc.274	rs181261148, rs184571756, rs189892914	<i>PBX3</i>	/	<i>Pbx3</i>	/	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3
uc.275	rs149345595, rs111973387	<i>PBX3</i>	rs10820374 7	<i>Pbx3</i>	/	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.276	rs12683881, rs145082420, rs192264399, rs184097004, rs189076431, rs191425730, rs138967507	<i>PBX3</i>	/	<i>C79798</i> (antisense), <i>Pbx3</i>	/	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3
uc.277	rs186798673, rs41307493, rs193249964	<i>PBX3</i>	/	<i>Pbx3</i>	/	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3
uc.278	/	<i>PBX3</i>	/	<i>Pbx3</i>	/	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3
uc.279	rs151214466, rs189401522	<i>PBX3</i>	/	<i>Pbx3</i>	/	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3
uc.280	rs145687528, rs2149989, rs114889309, rs114411468	<i>PBX3</i>	/	<i>Pbx3</i>	/	<i>Pbx3</i>	pre-B-cell leukemia homeobox 3
uc.281	rs184714154, rs146932898	<i>DDX31</i>	/	<i>Ddx31</i>	/	<i>Ddx31</i>	DEAD (Asp-Glu- Ala-Asp) box polypeptide 31
uc.282	rs75345761	<i>GRIN1</i>	/	<i>Grin1</i>	/	<i>Grin1</i>	glutamate receptor, ionotropic, N-methyl D-aspartate 1
uc.283	/	/	/	/	/	/	/
uc.284	rs146385759, rs138417306, rs78741647	/	/	/	/	/	/
uc.285	rs150319796, rs137961156, rs142459122, rs1782338	<i>CCAR1</i>	rs37005101	<i>Ccar1</i>	/	<i>Ccar1</i>	cell division cycle and apoptosis regulator 1
uc.286	rs147233105, rs36014122, rs73287017	<i>ZNF503</i>	/	/	/	/	zinc finger protein 503
uc.287	rs146902011, rs138741615, rs112437071, rs183429660	<i>C10orf11</i>	/	<i>1700112E06</i> <i>Rik</i> (protein coding)	/	<i>F1LU89_RAT</i>	/
uc.288	/	<i>C10orf11</i>	/	<i>1700112E06</i> <i>Rik</i> (protein coding)	/	<i>F1LU89_RAT</i>	/
uc.289	rs41314990, rs185448585	<i>C10orf11</i>	/	<i>1700112E06</i> <i>Rik</i> (protein coding)	/	<i>F1LU89_RAT</i>	/
uc.290	rs77676688, rs7087327	<i>C10orf11</i> , <i>RP11-</i> <i>369F10.3</i> (antisense)	/	<i>1700112E06</i> <i>Rik</i> (protein coding)	/	<i>F1LU89_RAT</i>	/
uc.291	rs113326412, rs145390075, rs41305709	<i>C10orf11</i>	/	<i>1700112E06</i> <i>Rik</i> (protein coding)	/	<i>F1LU89_RAT</i>	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.292	rs114110248, rs147037110	<i>LCOR</i>	/	<i>Lcor</i>	/	<i>D3ZAA6_RAT</i>	ligand dependent nuclear receptor corepressor
uc.293	rs145364394, rs192774535, rs74154683, rs184339673, rs79242821	/	/	/	/	/	/
uc.294	rs145948132, rs185891328, rs77311125, rs188837027, rs151206697, rs140307208, rs150311263, rs181255944, rs137990523	/	/	/	/	/	/
uc.295	rs192296826, rs150842766, rs7092999, rs139326250	/	/	/	/	/	/
uc.296	rs150803427, rs184469753, rs41308667, rs139128864, rs112602804, rs73343996	/	/	/	/	/	/
uc.297	rs186155229	/	/	/	/	/	/
uc.298	rs187515892, rs12220795, rs150956939, rs140799032, rs4394768	/	/	/	/	/	
uc.299	rs116664440, rs74550754, rs76803183, rs75999420, rs77453353, rs75462234, rs141576824, rs146392954	<i>PAX2</i>	/	<i>Pax2</i>	/	<i>Pax2</i>	paired box 2
uc.300	rs35967178, rs143705908, rs34603018, rs147227087	<i>PAX2</i>	/	<i>Pax2</i>	/	<i>Pax2</i>	paired box 2
uc.301	rs182632814, rs187761904, rs7097681, rs80328765	<i>PAX2</i>	/	<i>Pax2</i>	/	<i>Pax2</i>	paired box 2

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.302	rs77380815, rs185438446, rs72842101	/	/	/	/	/	/
uc.303	rs191078441	/	/	/	/	/	/
uc.304	rs115119994, rs191278887	<i>RP11- 107I14.5</i> (lincRNA)	/	/	/	/	/
uc.305	rs41300239, rs140831102	<i>BTRC</i>	/	<i>Btrc</i>	/	<i>Btrc</i>	beta-transducin repeat containing E3 ubiquitin protein ligase
uc.306	rs192034325, rs183912927, rs116608882	<i>BTRC</i>	/	<i>Btrc</i>	/	<i>Btrc</i>	beta-transducin repeat containing E3 ubiquitin protein ligase
uc.307	rs141693919	<i>BTRC</i>	/	<i>Btrc</i>	/	<i>Btrc</i>	beta-transducin repeat containing E3 ubiquitin protein ligase
uc.308	rs41303831, rs147681042	<i>BTRC</i>	rs47973723	<i>Btrc</i>	/	<i>Btrc</i>	beta-transducin repeat containing E3 ubiquitin protein ligase
uc.309	rs187514322, rs117326913, rs151147415	<i>BTRC</i>	/	<i>Btrc</i>	/	<i>Btrc</i>	beta-transducin repeat containing E3 ubiquitin protein ligase
uc.310	rs41310286	<i>VTI1A</i>	/	<i>Vti1a</i>	/	<i>Vti1a</i>	vesicle transport through interaction with t-SNAREs homolog 1A (yeast)
uc.311	rs183157527, rs115364952	<i>FAM204A</i>	rs48287107	<i>D19Ertd737 e</i>	/	/	family with sequence similarity 204, member A
uc.312	rs146453316, rs41308665, rs187810963	<i>FAM204A</i>	/	<i>D19Ertd737 e</i>	/	/	family with sequence similarity 204, member A
uc.313	rs36057393	<i>TIAL1</i>	/	<i>Tial1</i>	/	<i>Tial1</i>	TIA1 cytotoxic granule-associated RNA binding protein-like 1
uc.314	rs78020880, rs141259466, rs55963282	/	/	<i>Fgfr2</i>	/	/	fibroblast growth factor receptor 2
uc.315	rs190620214, rs139141243, rs182041872, rs117494402	/	/	<i>Fgfr2</i>	/	/	fibroblast growth factor receptor 2
uc.316	rs140668720	/	/	/	/	/	/
uc.317	rs144253327, rs36116890	<i>MGMT</i>	/	<i>Mgmt</i>	/	<i>Mgmt</i>	O-6-methylguanine- DNA methyltransferase
uc.318	rs182326832	<i>EBF3</i>	/	<i>Ebf3</i>	/	<i>Ebf3</i>	early B-cell factor 3
uc.319	rs191950901	/	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.320	rs34634217, rs150437595, rs72517885, rs3028953, rs72051865, rs181281095, rs185537203, rs61879734, rs5789573	/	/	/	/	/	/
uc.321	rs182282427, rs142441369	<i>RP11- 531H8.2</i> (lincRNA)	/	/	/	/	/
uc.322	/	<i>SOX6</i>	/	<i>Sox6</i>	/	<i>Sox6</i>	SRY (sex determining region Y)-box 6
uc.323	rs140310050	<i>SOX6</i>	/	<i>Sox6</i>	/	<i>Sox6</i>	SRY (sex determining region Y)-box 6
uc.324	rs11556749, rs140184093, rs143559000	<i>MPPED2</i>	/	<i>Mpped2</i>	/	<i>Mpped2</i>	metallophosphoestera se domain containing 2
uc.325	/	<i>Z83001.1</i> (processed transcript), <i>ELP4</i>	/	<i>Elp4</i>	/	<i>F1M086_RAT</i>	elongation protein 4 homolog (S. cerevisiae)
uc.326	rs189673923, rs182848170	<i>Z83001.1</i> (processed transcript), <i>ELP4</i>	/	<i>Elp4</i>	/	<i>F1M086_RAT</i>	elongation protein 4 homolog (S. cerevisiae)
uc.327	rs191695186, rs183750416, rs3026413, rs3026412	<i>Z83001.1</i> (processed transcript), <i>ELP4</i>	/	<i>Elp4</i>	/	<i>F1M086_RAT</i>	elongation protein 4 homolog (S. cerevisiae)
uc.328	rs41275156	<i>PAX6</i>	/	<i>Pax6</i>	/	<i>Pax6</i>	paired box 6
uc.329	rs150303127, rs117009917, rs144930329	<i>RP1-65P5.1</i> (pseudogene)	/	<i>0610012H0 3Rik</i> (protein coding)	/	<i>LOC691083</i> (protein coding)	/
uc.330	/	<i>RBM14, RBM14- RBM4</i>	/	<i>Rbm14</i>	/	<i>Rbm4</i>	RNA binding motif protein 14; RBM14- RBM4 readthrough
uc.331	rs149617923, rs138255175	<i>DLG2</i>	/	<i>Dlg2</i>	/	<i>Dlg2</i>	discs, large homolog 2 (Drosophila)
uc.332	rs184766181, rs189677597, rs180885255, rs185936189, rs73590085, rs190970313, rs181804313	/	rs51296464	/	/	/	/
uc.333	rs139994257, rs141864382	<i>MSANTD2</i>	/	<i>BC024479</i> (protein coding)	/	<i>F1LR51_RAT</i>	Myb/SANT-like DNA-binding domain containing 2

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.334	rs191004784, rs142817091, rs756629	<i>NTM</i>	/	<i>Ntm</i>	/	<i>Ntm</i>	neurotrimin
uc.335	rs141344958	<i>MGST1</i> , <i>LMO3</i>	/	<i>Lmo3</i>	/	<i>RGD1561357</i> (protein coding)	microsomal glutathione S-transferase 1; LIM domain only 3 (rhombotin-like 2)
uc.336	rs182917092, rs137977193, rs187746138, rs150583867	/	/	<i>Sox5</i>	/	/	SRY (sex determining region Y)-box 5
uc.337	/	<i>PDZRN4</i>	/	<i>Pdzrn4</i>	/	/	PDZ domain containing ring finger 4
uc.338	rs138735670, rs11556936, rs11556932, rs11556937, rs192943135	<i>RP11-793H13.8</i> (processed transcript), <i>PCBP2</i>	/	<i>Pcbp2</i>	/	<i>D4ACH3_RAT</i>	poly(rC) binding protein 2
uc.339	rs77902138, rs191688792, rs184281575, rs146886097	<i>ATP5G2</i>	/	/	/	/	ATP synthase, H ⁺ transporting, mitochondrial Fo complex, subunit C2 (subunit 9)
uc.340	rs139414711, rs188425949, rs74091505	/	/	/	/	/	/
uc.341	rs142781030, rs143836884, rs148164175, rs147012647, rs138069369, rs141068564	<i>HOXC10</i> , <i>HOXC5</i>	/	<i>Hoxc10</i> , <i>Gm20398</i> (non coding)	/	<i>Hoxc10</i>	homeobox C10; homeobox C5
uc.342	/	<i>HOXC5</i> , <i>HOXC6</i> , <i>HOXC4</i>	/	<i>Hoxc6</i> , <i>Gm20398</i> (non coding)	/	/	homeobox C5; homeobox C6; homeobox C4
uc.343	rs145731630, rs144307645, rs141865901, rs111518456, rs80157375, rs148984909	<i>HOXC5</i> , <i>HOXC6</i> , <i>HOXC4</i>	/	<i>Hoxc6</i> , <i>Gm20398</i> (non coding)	/	<i>Q63991_RAT</i>	homeobox C5; homeobox C6; homeobox C4
uc.344	rs35606176, rs151147734, rs139064806	<i>HOXC5</i> , <i>HOXC4</i>	/	<i>Hoxc5</i> , <i>Gm20398</i> (non coding)	/	<i>Hoxc5</i>	homeobox C5; homeobox C4
uc.345	/	<i>HOXC5</i> , <i>HOXC4</i>	/	<i>Hoxc4</i>	/	<i>Hoxc4</i>	homeobox C5; homeobox C4
uc.346	rs142182915, rs183166157	<i>RP11-144F15.1</i> (protein coding), <i>RFX4</i>	/	<i>Rfx4</i>	/	/	regulatory factor X, 4 (influences HLA class II expression)

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.347	rs190189110, rs114698059, rs182575420	/	/	/	/	/	/
uc.348	rs41300584	<i>DACH1</i>	/	<i>Dach1</i>	/	<i>LOC306096</i> (protein coding)	dachshund homolog 1 (Drosophila)
uc.349	rs2325421, rs192245837, rs74095972	<i>DACH1</i>	/	<i>Dach1</i>	/	<i>LOC306096</i> (protein coding)	dachshund homolog 1 (Drosophila)
uc.350	/	<i>DACH1</i>	/	<i>Dach1</i>	/	/	dachshund homolog 1 (Drosophila)
uc.351	rs139796764, rs183399294	/	/	/	/	/	/
uc.352	rs76567188, rs144797660, rs186759703	/	/	/	/	/	/
uc.353	rs2111796, rs77703734, rs185507119, rs115915196, rs9572903, rs143235579	/	/	/	/	/	/
uc.354	/	/	rs30192203	/	/	/	/
uc.355	rs142018602, rs188204006, rs192960730	/	/	/	/	/	/
uc.356	/	<i>MBNL2</i>	/	<i>Mbnl2</i>	/	<i>Mbnl2</i>	muscleblind-like splicing regulator 2
uc.357	rs138542649, rs75262363	/	/	<i>Gm5607</i> (protein coding)	/	/	/
uc.358	rs28712045	<i>RP11- 314P15.2</i> (lincRNA)	/	/	/	/	/
uc.359	rs190617392, rs1062462	<i>NOVA1</i>	/	<i>Nova1</i>	/	/	neuro-oncological ventral antigen 1
uc.360	rs138137204, rs182846549, rs187694970, rs191362743, rs183069455	<i>NOVA1</i>	/	<i>Nova1</i>	/	/	neuro-oncological ventral antigen 1
uc.361	rs149602255	<i>RP11- 966I7.1</i> (antisense)	/	/	/	/	/
uc.362	rs146415190, rs140726039, rs187817350	<i>CTD- 2384A14.1</i> (lincRNA)	/	/	/	/	/
uc.363	rs45482994	<i>RP11- 260G13.1</i> (lincRNA)	/	/	/	/	/
uc.364	rs35961721	<i>CTD- 2251F13.1</i> (lincRNA)	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.365	rs148287038	<i>CTD-2251F13.1</i> (lincRNA)	/	/	/	/	/
uc.366	rs145027811, rs147210893, rs139630378	<i>STRN3</i>	/	<i>Strn3</i>	/	<i>Strn3</i>	striatin, calmodulin binding protein 3
uc.367	rs1315144, rs45538937	<i>NPAS3</i>	/	<i>Npas3</i>	/	/	neuronal PAS domain protein 3
uc.368	rs145531514, rs112970390, rs8004379, rs188837158	<i>NPAS3</i>	/	<i>Npas3</i>	/	/	neuronal PAS domain protein 3
uc.369	rs181222326, rs140190401	<i>NPAS3</i>	/	<i>Npas3</i>	/	/	neuronal PAS domain protein 3
uc.370	rs181338647, rs147974716, rs141640672, rs185975204	<i>NPAS3</i>	/	<i>Npas3</i>	/	/	neuronal PAS domain protein 3
uc.371	rs80045939	<i>RALGAPA1</i>	/	<i>Ralgapa1</i>	/	<i>RGPA1_RAT</i>	Ral GTPase activating protein, alpha subunit 1 (catalytic)
uc.372	rs145330992	<i>RALGAPA1</i>	/	<i>Ralgapa1</i>	/	<i>RGPA1_RAT</i>	Ral GTPase activating protein, alpha subunit 1 (catalytic)
uc.373	/	<i>RP11-259K15.3</i> (non-coding)	/	/	/	/	/
uc.374	rs148850480, rs7143938, rs180822220, rs143449041	<i>MIPOL1</i>	/	<i>Mipol1</i>	/	/	mirror-image polydactyly 1
uc.375	rs113046805	<i>MIPOL1</i>	/	<i>Mipol1</i>	/	<i>F1LUS3_RAT</i>	mirror-image polydactyly 1
uc.376	/	<i>PRPF39</i>	/	<i>Prpf39</i>	/	<i>D4A0B1_RAT</i>	PRP39 pre-mRNA processing factor 39 homolog (S. cerevisiae)
uc.377	/	<i>PRPF39</i>	/	<i>Prpf39</i>	/	<i>D4A0B1_RAT</i>	PRP39 pre-mRNA processing factor 39 homolog (S. cerevisiae)
uc.378	/	<i>NRXN3</i>	/	<i>Nrxn3</i>	rs8160446	<i>NRX3A_RAT</i>	neurexin 3
uc.379	rs150851724	/	/	/	/	/	/
uc.380	rs142252570, rs150349399	/	/	/	/	/	/
uc.381	rs145759927, rs181533379, rs148957475, rs5810816	/	/	/	/	/	/
uc.382	rs189456481	<i>RP11-702M1.1</i> (lincRNA)	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.383	rs186652931, rs191607593, rs182954301	/	/	/	/	/	/
uc.384	rs76549073, rs181012113, rs34551805	<i>C15orf41</i>	rs27456340	<i>BC052040</i> (protein coding)	/	<i>RGD1563680</i> (protein coding)	/
uc.385	/	<i>MEIS2</i>	/	<i>Meis2</i>	/	<i>Meis2</i>	Meis homeobox 2
uc.386	rs34048156, rs58195779, rs138079777	/	/	/	/	/	/
uc.387	rs61757233, rs190653892	<i>MGA</i>	/	<i>Mga</i>	/	<i>Mga</i>	MAX gene associated
uc.388	rs186962893, rs76813819, rs146567044	<i>TCF12</i>	rs37191576	<i>Tcf12</i>	/	<i>Tcf12</i>	transcription factor 12
uc.389	rs141989899, rs190887636, rs111922182, rs111761274, rs34983272, rs76024542	<i>IQCH</i>	/	<i>Gm16759</i> (lincRNA), <i>Iqch</i>	/	<i>Iqch</i>	IQ motif containing H
uc.390	rs145063072, rs71400372, rs56241934	<i>MAP2K5</i>	/	<i>Map2k5</i>	/	<i>MP2K5_RAT</i>	mitogen-activated protein kinase kinase 5
uc.391	rs146841009	<i>MAP2K5</i>	rs51604954	<i>Map2k5</i>	/	<i>MP2K5_RAT</i>	mitogen-activated protein kinase kinase 5
uc.392	rs604684, rs193080686	/	/	/	/	/	/
uc.393	rs142670255, rs151022979	<i>CLK3</i>	/	<i>Clk3</i>	/	<i>Clk3</i>	CDC-like kinase 3
uc.394	rs189970172, rs149796142	/	/	/	/	/	/
uc.395	/	<i>RBBP6</i>	/	<i>Rbbp6</i>	/	<i>Rbbp6</i>	retinoblastoma binding protein 6
uc.396	rs142185441	/	/	/	/	/	/
uc.397	rs137915109, rs35742977	<i>ZNF423</i>	/	<i>Zfp423</i>	/	/	zinc finger protein 423
uc.398	rs113918901	<i>ZNF423</i>	/	/	/	/	zinc finger protein 423
uc.399	rs141510339, rs147042234	/	/	/	/	/	/
uc.400	rs117819096, rs191817552	/	/	/	/	/	/
uc.401	rs189603470, rs73583140	<i>TOX3</i>	rs33008407	<i>Tox3</i>	/	<i>Tox3</i>	TOX high mobility group box family member 3
uc.402	rs114606911, rs180909866, rs186335160, rs13339138	<i>RPGRIP1L</i>	/	<i>Rpgrip1l</i>	/	<i>Rpgrip1l</i>	RPGRIP1-like

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.403	rs74022614, rs147559652	/	/	<i>D230002A01Rik</i> (lincRNA)	/	/	/
uc.404	rs41313543, rs149863224, rs145793371	/	/	/	/	/	/
uc.405	rs141226786, rs41301801, rs183669227, rs189145878, rs138800533	/	rs32209879	/	/	/	/
uc.406	rs8053982, rs181783058, rs34134352	<i>NFAT5</i>	/	<i>Nfat5</i>	/	<i>Nfat5</i>	nuclear factor of activated T-cells 5, tonicity-responsive
uc.407	rs56393495, rs138012931, rs143392508	<i>NFAT5</i>	/	<i>Nfat5</i>	/	<i>Nfat5</i>	nuclear factor of activated T-cells 5, tonicity-responsive
uc.408	rs139068178, rs149908041, rs147553973, rs141983735, rs145799976, rs147952239, rs149017053, rs143760732, rs148140195, rs141871061	<i>RP5-991G20.1</i> (processed transcript), <i>ZFHX3</i>	/	<i>Zfhx3</i>	/	<i>Zfhx3</i>	zinc finger homeobox 3
uc.409	rs146875021, rs12934016, rs12933631	<i>RP11-346C20.1</i> (sense overlapping) , <i>ZFHX3</i>	/	/	/	/	zinc finger homeobox 3
uc.410	rs140569420, rs191133374, rs150039766	/	/	<i>Gm11431</i> (lincRNA)	/	/	/
uc.411	rs150444413, rs188083352, rs114765326	<i>AATF</i>	/	<i>Aatf</i>	/	<i>Aatf</i>	apoptosis antagonizing transcription factor
uc.412	rs137929653, rs59799585, rs149527844	<i>AATF</i>	/	<i>Aatf</i>	/	<i>Aatf</i>	apoptosis antagonizing transcription factor
uc.413	rs1139824, rs113351963, rs1139823, rs1139822	<i>MED1</i>	/	<i>Med1</i>	/	<i>Med1</i>	mediator complex subunit 1
uc.414	rs137853164	<i>THRA</i> , <i>NR1D1</i>	/	<i>Thra</i> , <i>Nr1d1</i>	/	<i>Thra</i> , <i>Nr1d1</i>	thyroid hormone receptor, alpha; nuclear receptor subfamily 1, group D, member 1
uc.415	/	<i>HOXB3</i> , <i>HOXB-AS3</i> (antisense RNA)	/	/	/	/	homeobox B3; HOXB cluster antisense RNA 3 (non-protein coding)

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.416	rs144236768, rs148136297, rs145148229	<i>HOXB3</i> , <i>HOXB-AS3</i> (antisense RNA), <i>HOXB5</i>	/	<i>Hoxb5</i> , <i>0610040B09</i> <i>Rik</i> (antisense)	/	<i>Hoxb5</i>	homeobox B3; HOXB cluster antisense RNA 3 (non-protein coding); homeobox B5
uc.417	rs56805315, rs138810538, rs78222746	<i>HOXB-AS3</i> (antisense RNA), <i>HOXB6</i> , <i>AC103702.1</i> (protein coding)	/	<i>Hoxb6</i> , <i>0610040B09</i> <i>Rik</i> (antisense)	/	/	HOXB cluster antisense RNA 3 (non-protein coding); homeobox B6
uc.418	/	<i>SRSF1</i>	/	<i>Srsf1</i>	/	<i>D4A9L2_RAT</i>	serine/arginine-rich splicing factor 1
uc.419	rs2233912, rs147579081, rs142072157, rs150523813	<i>SRSF1</i>	/	<i>Srsf1</i>	/	<i>D4A9L2_RAT</i>	serine/arginine-rich splicing factor 1
uc.420	/	<i>DDX5</i>	/	<i>Ddx5</i> , <i>Mir3064</i> (miRNA)	/	<i>Ddx5</i>	DEAD (Asp-Glu- Ala-Asp) box helicase 5
uc.421	rs143316155, rs12456225	<i>ZNF521</i>	/	<i>Zfp521</i>	/	/	zinc finger protein 521
uc.422	rs189646070, rs181284363, rs141915119, rs78092980, rs150830564, rs45455094	<i>ZNF521</i>	/	<i>Zfp521</i>	/	/	zinc finger protein 521
uc.423	rs149551397, rs41303041, rs116260226	<i>ZNF521</i>	/	<i>Zfp521</i>	/	/	zinc finger protein 521
uc.424	/	<i>ZNF521</i>	/	<i>Zfp521</i>	/	/	zinc finger protein 521
uc.425	rs138709311, rs188206171, rs192439076, rs62083937	<i>ZNF521</i>	/	<i>Zfp521</i>	/	<i>Znf521</i>	zinc finger protein 521
uc.426	rs140667406	<i>TAF4B</i>	rs31128319	<i>Taf4b</i>	/	<i>RGD1562997</i> (protein coding)	TAF4b RNA polymerase II, TATA box binding protein (TBP)-associated factor, 105kDa
uc.427	rs189600095, rs140610287, rs180829291	/	/	<i>Kctd1</i>	/	<i>F1LVD9_RAT</i>	potassium channel tetramerisation domain containing 1
uc.428	rs188242868, rs181733584	<i>AC012123.1</i>	/	<i>4930426D0</i> <i>5Rik</i> (pseudogene)	/	<i>D3ZUE9_RAT</i>	/
uc.429	rs183446318	<i>KIAA1328</i>	rs31424513	<i>AW554918</i> (protein coding)	/	/	KIAA1328

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.430	rs41370646, rs41534145, rs41356048	<i>CELF4</i>	/	<i>Celf4</i>	/	<i>F1M6I9_RAT</i>	CUGBP, Elav-like family member 4
uc.431	rs111834469, rs111228431, rs186927934, rs190745607	/	/	/	/	/	/
uc.432	rs12962352, rs41442949, rs41419050, rs76054315, rs111594923	/	/	/	/	/	/
uc.433	rs140523539, rs182787929, rs188725597, rs75128538	/	/	/	/	/	/
uc.434	rs144100096, rs116556929	<i>SKOR2</i>	/	<i>Skor2</i>	/	<i>Skor2</i>	SKI family transcriptional corepressor 2
uc.435	rs186334179, rs190709022, rs181478739, rs112001726	<i>TCF4</i>	/	<i>Tcf4</i>	/	<i>Tcf4</i>	transcription factor 4
uc.436	rs34968341, rs34555808, rs74676655	<i>TCF4</i>	rs47564759	<i>Tcf4</i>	/	<i>Tcf4</i>	transcription factor 4
uc.437	rs148062178, rs114573277, rs192264981, rs17055306	<i>ZNF407</i>	/	<i>Zfp407</i>	/	<i>D4A2J7_RAT</i>	zinc finger protein 407
uc.438	/	<i>ZNF407</i>	/	<i>Zfp407</i>	/	<i>D4A2J7_RAT</i>	zinc finger protein 407
uc.439	/	<i>ZNF407</i>	/	<i>Zfp407</i>	/	<i>D4A2J7_RAT</i>	zinc finger protein 407
uc.440	rs17818580, rs114924640, rs189222459, rs192512285	<i>ZNF407</i>	rs48294219	<i>Zfp407</i>	/	<i>D4A2J7_RAT</i>	zinc finger protein 407
uc.441	rs192469983, rs183429872	<i>ZNF407</i>	/	<i>Zfp407</i>	/	<i>D4A2J7_RAT</i>	zinc finger protein 407
uc.442	rs74627831, rs143914079, rs147244007	<i>ZNF407</i>	/	<i>Zfp407</i>	/	<i>D4A2J7_RAT</i>	zinc finger protein 407
uc.443	/	<i>HNRNPM</i>	/	<i>Hnrnmp</i>	/	<i>Hnrpm</i>	heterogeneous nuclear ribonucleoprotein M
uc.444	rs77055544, rs144540397, rs186010732, rs191480016, rs41301953, rs115791645	<i>URI1</i>	rs31856242	<i>C80913</i> (protein coding)	/	<i>RGD1310358</i> (protein coding)	URI1, prefoldin-like chaperone
uc.445	/	/	/	/	/	/	/

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.446	rs181979067, rs142730227, rs150746281	/	/	<i>Zfp536</i>	/	/	zinc finger protein 536
uc.447	rs34711524	/	/	<i>Zfp536</i>	/	/	zinc finger protein 536
uc.448	rs115650222	/	/	<i>Zfp536</i>	/	/	zinc finger protein 536
uc.449	rs114798778, rs148199453, rs186612542, rs193103927	<i>ZNF536</i>	/	<i>Zfp536</i>	/	<i>Znf536</i>	zinc finger protein 536
uc.450	rs140587014, rs183074880, rs188161379	/	/	/	/	/	/
uc.451	rs41309439, rs76368616	<i>TSHZ3</i>	/	<i>Tshz3</i>	/	<i>Tshz3</i>	teashirt zinc finger homeobox 3
uc.452	/	<i>TSHZ3</i>	rs32568344	<i>Tshz3</i>	/	<i>Tshz3</i>	teashirt zinc finger homeobox 3
uc.453	rs188453644, rs41303845, rs191492837, rs113382461	/	/	/	/	/	/
uc.454	rs148423778, rs117118794, rs112779445, rs35861107, rs139616286	<i>SLC23A2</i>	/	<i>Slc23a2</i>	/	<i>Slc23a2</i>	solute carrier family 23 (nucleobase transporters), member 2
uc.455	rs180983278	<i>RBM39</i>	/	<i>Rbm39</i>	/	<i>Rbm39</i>	RNA binding motif protein 39
uc.456	rs112197164	<i>SRSF6</i>	/	<i>Srsf6</i>	/	<i>Sfrs6</i>	serine/arginine-rich splicing factor 6
uc.457	rs142062346	<i>HIRA</i> , <i>C22orf39</i>	/	<i>Hira</i>	/	/	HIR histone cell cycle regulation defective homolog A (<i>S. cerevisiae</i>)
uc.458	rs117874679	<i>RBFOX2</i>	/	<i>Rbfox2</i> , <i>1700109K24</i> <i>Rik</i> (lincRNA)	/	<i>Rbm9</i>	RNA binding protein, fox-1 homolog (<i>C.</i> <i>elegans</i>) 2
uc.459	rs142081737	<i>CNKS2</i>	/	<i>Cnksr2</i>	/	<i>Cnksr2</i>	connector enhancer of kinase suppressor of Ras 2
uc.460	rs190103302, rs141728753	<i>POLA1</i>	/	<i>Pola1</i>	/	<i>DPOLA_RAT</i>	polymerase (DNA directed), alpha 1, catalytic subunit
uc.461	rs11573440	<i>POLA1</i>	/	<i>Pola1</i>	/	<i>DPOLA_RAT</i>	polymerase (DNA directed), alpha 1, catalytic subunit
uc.462	rs113072356, rs190100081, rs139820873, rs181170257, rs5944694	<i>POLA1</i>	/	<i>Pola1</i>	/	<i>DPOLA_RAT</i>	polymerase (DNA directed), alpha 1, catalytic subunit

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.463	rs145853984, rs138819937, rs189030471	<i>POLA1</i>	/	<i>Pola1</i>	/	<i>DPOLA_RAT</i>	polymerase (DNA directed), alpha 1, catalytic subunit
uc.464	rs11573463, rs191316710	<i>POLA1</i>	/	<i>Pola1</i>	/	<i>DPOLA_RAT</i>	polymerase (DNA directed), alpha 1, catalytic subunit
uc.465	rs181991476, rs185089697	<i>POLA1</i>	/	<i>Pola1</i>	/	<i>DPOLA_RAT</i>	polymerase (DNA directed), alpha 1, catalytic subunit
uc.466	rs113774923, rs138658126	<i>POLA1</i>	rs29071983	<i>Pola1</i>	/	<i>DPOLA_RAT</i>	polymerase (DNA directed), alpha 1, catalytic subunit
uc.467	rs144880670, rs186842380, rs6629933, rs76067703, rs191626524	<i>POLA1</i>	/	<i>Pola1</i>	/	<i>DPOLA_RAT</i>	polymerase (DNA directed), alpha 1, catalytic subunit
uc.468	rs112911817, rs140518939	/	/	/	/	/	/
uc.469	/	/	/	/	/	/	/
uc.470	rs190114603	/	/	/	/	/	/
uc.471	rs55668011, rs191556194	<i>DDX3X</i>	/	<i>Ddx3x</i>	/	<i>D4ADE8_RAT</i>	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked
uc.472	/	<i>CASK-AS1</i> (antisense RNA), <i>CASK</i>	/	<i>Cask</i>	/	<i>Cask</i>	<i>CASK</i> antisense RNA 1 (non-protein coding); calcium/calmodulin-dependent serine protein kinase (MAGUK family)
uc.473	rs7062937, rs188141012, rs180681274	<i>NLGN3</i>	/	<i>Nlgn3</i>	/	<i>Nlgn3</i>	neuroligin 3
uc.474	/	<i>ZMYM3</i>	/	<i>Zmym3</i>	/	<i>Zmym3</i>	zinc finger, MYM-type 3
uc.475	/	<i>OGT</i>	/	<i>Ogt</i>	rs8168858	<i>Ogt</i>	O-linked N-acetylglucosamine (GlcNAc) transferase
uc.476	rs189938519, rs181242483	/	/	/	/	/	/
uc.477	rs132630295, rs132630287, rs132630292, rs113345335, rs112667596	<i>PLP1</i>	/	<i>Plp1</i>	/	<i>Plp1</i>	proteolipid protein 1
uc.478	rs3752325, rs1132301, rs1132302, rs1132303, rs1052539, rs1052540, rs1052541, rs1052542, rs1052543, rs1052544, rs1052545, rs1052546	<i>GRIA3</i>	/	<i>Gria3</i>	/	/	glutamate receptor, ionotropic, AMPA 3

UCR	SNP-ji znotraj UCR-jev pri človeku	Gostiteljski geni (človek)	SNP-ji znotraj UCR-jev pri miši	Gostiteljski geni (miš)	SNP-ji znotraj UCR-jev pri podgani	Gostiteljski geni (podgana)	Ime gena
uc.479	rs78034276, rs76068519, rs77856488, rs150138519	<i>GRIA3</i>	rs29939974	<i>Gria3</i>	/	<i>Gria3</i>	glutamate receptor, ionotropic, AMPA 3
uc.480	rs183029659, rs1044020	<i>STAG2</i>	/	<i>Stag2</i>	/	/	stromal antigen 2
uc.481	rs187706335	<i>STAG2</i>	rs31309753	<i>Stag2</i>	/	/	stromal antigen 2
uc.482	rs186364973, rs145902649	/	/	<i>Atp11c</i>	/	/	ATPase, class VI, type 11C
uc.483	rs112917029, rs148758482, rs142410570, rs114256065	TBC1D5	/	<i>Tbc1d5</i>	/	<i>Tbc1d5</i>	TBC1 domain family, member 5

/ - ni podatka