UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA

Minja ZORC

RAZVOJ BIOOZNAČEVALCEV ZA NALAGANJE MAŠČOBE Z INTEGRACIJO GENOMSKIH PODATKOV IN BIOINFORMACIJSKO ANALIZO

DOKTORSKA DISERTACIJA

Ljubljana, 2013

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA

Minja ZORC

RAZVOJ BIOOZNAČEVALCEV ZA NALAGANJE MAŠČOBE Z INTEGRACIJO GENOMSKIH PODATKOV IN BIOINFORMACIJSKO ANALIZO

DOKTORSKA DISERTACIJA

DEVELOPMENT OF BIOMARKERS FOR FAT DEPOSITION USING INTEGRATION OF GENOMIC DATA AND BIOINFORMATICS ANALYSIS

DOCTORAL DISSERTATION

Ljubljana, 2013

Na podlagi Statuta Univerze v Ljubljani ter po sklepu Senata Biotehniške fakultete in sklepa Senata Univerze z dne 15. septembra 2009 je bilo potrjeno, da kandidatka izpolnjuje pogoje za neposreden prehod na doktorski Univerzitetni podiplomski študij Biomedicine ter opravljanje doktorata znanosti s področja genetike. Za mentorico je bila imenovana prof. dr. Tanja Kunej.

Komisija za oceno in zagovor:

Predsednik:	prof. dr. Simon HORVAT Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko, Kemijski inštitut Ljubljana, Laboratorij za biotehnologijo
Član:	prof. dr. Tanja KUNEJ Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko
Član:	prof. dr. Blaž ZUPAN Univerza v Ljubljani, Fakulteta za računalništvo in informatiko

Datum zagovora: 11. 6. 2013

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

Minja Zorc

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD Dd
DK UDK 575(043.3)=163.3
KG bioinformatika/biooznačevalci/debelost/genomika/nalaganje maščobe/polimorfizmi
KK AGRIS /
AV ZORC, Minja, univ. dipl. inž. rač. in inf.
SA KUNEJ, Tanja (mentorica)
KZ SI-1000 Ljubljana, Jamnikarjeva 101
ZA Univerza v Ljubljani, Biotehniška fakulteta, Univerzitetni podiplomski študij Biomedicine, področje genetike
LI 2013
IN RAZVOJ BIOOZNAČEVALCEV ZA NALAGANJE MAŠČOBE Z INTEGRACIJO GENOMSKIH PODATKOV IN BIOINFORMACIJSKO ANALIZO
TD Dokt. disertacija
OP XII, 89 str., 12 pregl., 18 sl., 7 pril., 239 vir.
IJ sl
JI sl/en
AI Debelost je večgenska bolezen, ki predstavlja pereč zdravstveni problem. Za debelostjo zbolevajo ljudje vseh starosti, pojavlja pa se tudi pri domačih živalih, kjer pogosteje govorimo o zamaščenosti. Odkritje genetskih osnov nalaganja maščobe lahko pripomore k razumevanju pojava in terapiji debelosti. Količina razpoložljivih genomskih podatkov narašča, hkrati pa se povečujejo zahteve po metodah za njihovo analizo. Vse večjo vlogo pri raziskavah kompleksnih fenotipov pridobivajo sistemski pristopi. Izdelali smo genomski atlas, ki predstavlja centralno spletno mesto genetskih vzrokov za nalaganje maščobe. Uporabili smo primerjalni in integrativni pristop zbiranja lokusov za nalaganje maščobe pri človeku, miši, podgani in govedu ter z vizualizacijo integriranih podatkov dosegli enostaven vpogled v do sedaj poznane lokuse, povezane z nalaganjem maščobe. Izdelali smo prikaze genomske razporeditve lokusov, identificirali kandidatne biološke poti in genske mreže za nalaganje maščobe, ki so bile osnova za razvrščanje kandidatnih genov po prioritetah. Razvili smo bioinformacijski orodji za analizo nekodirajočih kandidatnih genov (miRNA SNiPer in miRNA Viewer). Iz nabora kandidatnih lokusov smo izbrali potencialne biooznačevalce *(Akt1, Ubc, Grb2, Mir599)* in z analizo povezanosti genotipa s fenotipom preverili njihov vpliv na lastnosti nalaganja maščobe pri miši. Razvili smo strategijo za raziskave genetskih vzrokov nalaganja maščobe, ki jo je možno uporabiti tudi za druge kompleksne fenotipe.

KEY WORDS DOCUMENTATION

DN Dd
DC UDC 575(043.3)=163.3
CX bioinformatics/biomarkers/fat deposition/genomics/obesity/polymorphisms
CC AGRIS /
AU ZORC, Minja
AA KUNEJ, Tanja (supervisor)
PP SI-1000 Ljubljana, Jamnikarjeva 101
PB University of Ljubljana, Biotechnical Faculty, Postgraduate Study Programme in Biomedicine, Field: Genetics
PY 2013
TI DEVELOPMENT OF BIOMARKERS FOR FAT DEPOSITION USING INTEGRATION OF GENOMIC DATA AND BIOINFORMATICS ANALYSIS
DT Doctoral dissertation
NO XII, 89 p.,12 tab., 18 fig., 7 ann., 239 ref.
LA sl
AL sl/en
AB Obesity is polygenic disease which presents a major health issue. It affects people of all ages as well as domestic animals. The unravelling of genetic bases of fat deposition might help to develop therapeutics and understand the process of fat deposition. The amount of available genomic data and the need for genomic data analysis methods grow. Systemic approaches are becoming important in complex phenotypes research. We created the genomic atlas, which presents the central web resource of genetic causes for fat deposition. The comparative and integrative approach to collect the loci associated with fat deposition in human, mouse, rat and cattle was used. By visualization of the integrated data the insight into known fat deposition loci was enabled. We created genomic views of loci, identified candidate biological pathways and determined genetic networks for fat deposition, which were basis for candidate genes prioritisation. Two bioinformatics tools for analysis of noncoding candidate genes were developed (miRNA SNiPer and miRNA Viewer). From the set of candidate loci we selected potential biomarkers (*Akt1*, *Ubc*, *Grb2*, *Mir599)* and tested their effect on fat deposition traits in mice using analysis of association between genotype and phenotype. We developed a strategy for research of genetic causes for fat deposition. The same approach can be used for analysis of other complex phenotypes.

KAZALO VSEBINE

		str.
	Ključna dokumentacijska informacija	III
	Key Words Documentation	IV
	Kazalo preglednic	VIII
	Kazalo slik	IX
	Kazalo prilog	XI
	Okrajšave in simboli	XII
1	**UVOD**	1
1.1	CILJI RAZISKAVE IN HIPOTEZE	2
1.2	PRISPEVKI K ZNANOSTI	2
2	**PREGLED OBJAV**	3
2.1	NALAGANJE MAŠČOBE	3
2.1.1	**Maščobno tkivo**	3
2.1.2	**Lokusi, povezani z nalaganjem maščobe**	4
2.1.3	**Biološke poti in biološki procesi, povezani z nalaganjem maščobe**	5
2.1.4	**Zbirke lokusov, povezanih z debelostjo**	6
2.1.5	**Genske terapije za zdravljenje debelosti**	6
2.2	RAZISKOVANJE KOMPLEKSNIH BOLEZNI	6
2.2.1	**Sistemska biologija**	7
2.2.2	**Strategije raziskovanja kompleksnih bolezni**	7
2.2.3	**Integracija »omskih« podatkov ali integratomika**	8
2.3	ZBIRANJE LOKUSOV, POVEZANIH S FENOTIPOM	9
2.3.1	**Specializirane zbirke lokusov, povezanih s fenotipom**	9
2.3.2	**Zbiranje lokusov iz bibliografskih zbirk**	12
2.3.3	**Ontologije fenotipov in bolezni**	13
2.3.4	**Orodja za rudarjenje v podatkih**	13
2.4	BIOLOŠKE POTI	13
2.5	BIOLOŠKE MREŽE	14
2.5.1	**Mreže proteinskih interakcij**	15
2.5.2	**Mreže uravnavanja izražanja genov (DNA-protein, RNA-RNA)**	15

2.5.3 Metabolne mreže 16
2.5.4 Signalne mreže 16
2.5.5 Uporaba pristopov teorije grafov za analizo bioloških mrež 16
2.5.6 Orodja za analizo bioloških mrež 17
2.6 RAZVRŠČANJE KANDIDATNIH GENOV PO PRIORITETI 18
2.7 BIOOZNAČEVALCI 18
2.8 ASOCIACIJSKE ŠTUDIJE 19
3 MATERIALI IN METODE 21
3.1 ZBIRANJE LOKUSOV IN IZDELAVA GENOMSKEGA ATLASA ZA NALAGANJE MAŠČOBE 21
3.1.1 Zbiranje lokusov, povezanih z nalaganjem maščobe 21
3.1.2 Izdelava kataloga lokusov, povezanih z nalaganjem maščobe 21
3.1.3 Spletno centralno mesto za lokuse, povezane z nabiranjem maščobe 22
3.2 INTEGRACIJA RAZNOVRSTNIH GENOMSKIH PODATKOV IN BIOINFORMACIJSKA ANALIZA 22
3.2.1 Prikaz genomske razporeditve lokusov, povezanih z nalaganjem maščobe 22
3.2.2 Analiza obogatenosti bioloških poti 23
3.2.3 Integracija in analiza bioloških mrež 23
3.2.4 Bioinformacijska analiza in razvrščanje kandidatnih genov po prioriteti 24
3.3 VPLIV KANDIDATNIH GENOV NA LATNOSTI NALAGANJA MAŠČOBE 24
4 REZULTATI 27
4.1 GENOMSKI ATLAS ZA RAZISKAVE LOKUSOV, POVEZANIH Z NALAGANJEM MAŠČOBE 28
4.2 INTEGRACIJA RAZNOVRSTNIH GENOMSKIH PODATKOV V POVEZAVI Z NALAGANJEM MAŠČOBE 30
4.2.1 Genomski prikaz razporeditve lokusov 30
4.2.2 Analiza obogatenosti bioloških poti 31
4.2.3 Analiza mreže proteinskih interakcij in mreže uravnavanja izražanja genov 39
4.2.4 Razvrščanje protein-kodirajočih kandidatnih genov po prioriteti 43

4.2.5 Razvrščanje kandidatnih miRNA genov po prioriteti z orodjem miRNA SNiPer 46

4.2.6 Razvrščanje kandidatnih miRNA genov po prioriteti s pomočjo orodja miRNA Viewer 47

4.3 ANALIZA POVEZANOSTI GENOTIPA S FENOTIPOM 49

5 RAZPRAVA 55

5.1 GENOMSKI ATLAS ZA NALAGANJE MAŠČOBE 55

5.2 INTEGRACIJA GENOMSKIH PODATKOV 56

5.2.1 Genomski prikazi razporeditve lokusov 56

5.2.2 Analiza obogatenosti bioloških poti 57

5.2.3 Genska mreža proteinskih interakcij 57

5.2.4 Integracija mreže uravnavanja izražanja genov z molekulami miRNA 58

5.2.5 Postopek integracije genomskih podatkov 58

5.3 RAZVRŠČANJE KANDIDATNIH GENOV PO PRIORITETI 58

5.3.1 Razvrščanje protein kodirajočih kandidatnih genov po prioriteti 58

5.3.2 Razvrščanje kandidatnih zapisov za nekodirajoče RNA po prioriteti 59

5.4 ANALIZA POVEZAVE GENOTIPA S FENOTIPOM 60

5.5 BIOOZNAČEVALCI ZA NALAGANJE MAŠČOBE 61

6 SKLEPI 63

7 POVZETEK (SUMMARY) 64

7.1 POVZETEK 64

7.2 SUMMARY 65

8 VIRI 66

ZAHVALA

PRILOGE

KAZALO PREGLEDNIC

str.

Preglednica 1: Specializirane podatkovne zbirke lokusov, povezanih z boleznimi pri človeku 10
Preglednica 2: Zbirke bioloških poti, ki jih zajame orodje DAVID za analizo obogatenosti množice genov 23
Preglednica 3: Genotipski podatki iz zbirke MPD za polimorfizme rs261613149, rs37362582, rs13461180, rs13471888 in rs37362582 pri 24 inbridiranih linijah miši 26
Preglednica 4: Biološke poti, v katerih so udeleženi kandidatni geni za nalaganje maščobe 31
Preglednica 5: Biološke poti, ki imajo skupne kandidatne gene iz Genomskega atlasa za nalaganje maščobe 35
Preglednica 6: Vrednosti koeficientov grupiranja in vmesnostne centralnosti za 20 vozlišč z najvišjo mero vmesnosti v mreži proteinskih interakcij 41
Preglednica 7: Kandidatni miRNA geni za nalaganje maščobe 43
Preglednica 8: Seznam razvrščenih kandidatnih genov za nalaganje maščobe po prioriteti 44
Preglednica 9: Število kandidatnih miRNA genov za nalaganje maščobe pri miši 46
Preglednica 10: Geni za miRNA pri človeku in miši, ki imajo SNP v regiji seed, ki je odgovorna za vezavo na tarče 47
Preglednica 11: Rezultat analize povezanosti genotipa s fenotipom pri miši. Število fenotipskih lastnosti pri inbridiranih linijah miši, na katere imajo polimorfizmi rs37362582, rs32568344, rs261613149, rs13461180 in rs13471888 učinek. 50
Preglednica 12: Fenotipski podatki iz zbirke MPD: telesne mase samcev pri miši ob koncu testa 53

KAZALO SLIK

str.

Slika 1: Vhodna stran projekta Integratomics TIME (http://integratomics-time.com) za raziskave kompleksnih fenotipov 9

Slika 2: Shema poteka dela 27

Slika 3: Obrazec za vnos podatkov o lokusu za nalaganje maščobe v zbirko MySQL s pomočjo aplikacije Xataface 28

Slika 4: Spletno mesto Obesity Genomic Atlas 29

Slika 5: Genomski prikaz A) Razporeditev genomskih lokusov za nalaganje maščobe pri človeku B) Kromosom 8 pri človeku z lokusi za nalaganje maščobe. Prikazana je lokacija gena *DEPTOR*, ki se prekriva z dvema QTL-oma. 30

Slika 6: Razporeditev genomskih lokusov za nalaganje maščobe pri miši 31

Slika 7: Vennov diagram, ki prikazuje število kandidatnih genov iz Genomskega atlasa za nalaganje maščobe, ki so udeleženi v bioloških poteh *Adipocytokine signaling pathway*, *Insulin signaling pathway*, *Endothelin signaling pathway*, *PPAR signaling pathway* in *Neuroactive ligand-receptor interaction* 34

Slika 8: Biološka pot »Adipocytokine signaling pathway« iz zbirke KEGG 37

Slika 9: Biološka pot »Insulin signaling pathway« iz zbirke KEGG 38

Slika 10: Biološka pot »Endothelin signaling pathway« iz zbirke PANTHER 39

Slika 11: Vrednosti koeficientov grupiranja in vmesnostne centralnosti za vozlišča v mreži proteinskih interakcij 40

Slika 12: Shema razvrščanja lokusov po prioriteti 44

Slika 13: Rezultat orodja miRNA SNiPer: polimorfni regiji seed pri genih za miRNA pri miši *mmu-mir-717* in *mmu-mir-599* 47

Slika 14: Genomska razporeditev genov miRNA in QTL-ov, povezanih z nalaganjem maščobe ter gostiteljskih genov za miRNA pri človeku 48

Slika 15: Genomska razporeditev genov miRNA in QTL-ov, povezanih z nalaganjem maščobe ter gostiteljskih genov za miRNA pri miši 48

Slika 16: Prekrivanje genomskih lokusov, povezanih z nalaganjem maščobe: SNP, gen za miRNA, protein-kodirajoči gen in QTL 48
Slika 17: Ocena razlik med aleloma po analiziranih SNP-jih 51
Slika 18: Meritve telesne mase miši po alelih SNP-ja rs37362582 52

KAZALO PRILOG

Priloga A: Catalog of microRNA seed polymorphisms in vertebrates (Zorc in sod., 2012)

Priloga B: Genetic variability of microRNA genes in farm animals (Jevšinek Skok in sod., 2012)

Priloga C: Obesity gene atlas in mammals (Kunej in sod., 2012)

Priloga D: Pivotal role of the muscle-contraction pathway in cryptorchidism and evidence for genomic connections with cardiomyopathy pathways in RASopathies (Cannistraci in sod., 2013)

Priloga E: The microRNA decalog of cancer involvement (Kunej in sod, v tisku)

Priloga F: Cross talk between microRNA and coding cancer genes (Kunej in sod., 2012)

Priloga G: Genome-wide *in silico* screening (GWISS) for microRNA genetic variability in livestock species (Jevšinek Skok in sod., 2013)

Priloga H: Genome-wide and species-wide *in silico* screening for intragenic microRNAs in human, mouse and chicken (Godnič in sod., 2013)

OKRAJŠAVE IN SIMBOLI

BAT	Rjavo maščobno tkivo (angl. *brown adipose tissue*)
DNP	Polimorfizem dveh zaporednih nukleotidov (angl. *dinucleotide polymorphism*)
GO	Ontologija genov (angl. *gene ontology*)
ChIP	Kromatinska imunoprecipitacija (angl. *chromatin immunoprecipitation*)
GRN	Mreža uravnavanja izražanja genov (angl. *gene regulatory networ*k)
GWAS	Asociacijska analiza na ravni celotnega genoma (angl. *genome wide association study*)
GWISS	Analiza *in silico* na ravni celotnega genoma (angl. *genome-wide in silico screening*)
KO	Poskus z izbitim genom (angl. *gene knockout*)
MNP	Polimorfizem več zaporednih nukleotidov (angl. *multiple nucleotide polymorphism*)
MPD	Podatkovna zbirka fenoma miši (angl. *Mouse Phenome Database*)
mRNA	Informacijska RNA (angl. *messenger RNA*)
miRNA	Mikro RNA (angl. *microRNA*)
PPI	Proteinske interakcije (angl. *protein-protein interactions*)
PIN	Mreža proteinskih interakcij (angl. *protein-protein interaction network)*
QTL	Kvantitativni lokus (angl. *quantitative trait locus*)
SNP	Polimorfizem posameznega nukleotida (angl. s*ingle nucleotide polymorphism*)
TF	Transkripcijski dejavnik (angl. *transcription factor*)
WAT	Belo maščobno tkivo (angl. *white adipose tissue*)
Y2H	Dvohibridni sistem kvasovke (angl. *yeast two hybrid system*)

1 UVOD

Debelost je večgenska bolezen, ki predstavlja pereč zdravstveni problem v razvitem svetu, v zadnjem času pa tudi v deželah v razvoju. Za debelostjo zbolevajo ljudje vseh starosti, pojavlja pa se tudi pri domačih živalih, kjer pogosteje govorimo o zamaščenosti. Odkritje genetskih osnov nalaganja maščobe lahko pripomore k razumevanju pojava in terapiji debelosti.

V medicini je bilo do nedavnega preučevanje posameznih kandidatnih genov običajen pristop k odkrivanju genetskih vzrokov bolezni. Ta vrsta raziskav je pripomogla k razumevanju, diagnostiki in terapiji širokega spektra bolezni. Kljub uspehom ima redukcionističen pristop mnogo omejitev, še posebej pri preučevanju večgenskih bolezni in lastnosti, kjer je za fenotipski izid odgovorno večje število različnih genov. Spremembe v posameznih genih prispevajo k skupnemu učinku, ki je vsota prispevkov vseh vpletenih genov. Preučevanje kompleksnih lastnosti in bolezni je zapleteno, saj je izražanje genov in delovanje njihovih proteinskih produktov soodvisno, na njihov nastanek pa vplivajo tudi dejavniki iz okolja. Vse večjo vlogo pridobivajo sistemski pristopi, pri katerih si pomagamo z raziskavo celotnega genoma.

Hiter napredek tehnik določanja zaporedja DNA, mikromrež (angl. *microarrays*) in drugih novih tehnologij ponuja nove strategije za identifikacijo polimorfizmov, povezanih z boleznijo. Poznavanje zaporedij DNA ter boljše razumevanje genoma in izražanja genov omogoča prepoznavanje ključnih regulatornih elementov na globalni ravni genoma. Genomske raziskave proizvedejo ogromne količine podatkov. Dostopni molekularni podatki na ravni celotnega genoma eksponentno rastejo. Hitro kopičenje podatkov v podatkovnih zbirkah in literaturi zahteva vedno bolj sistematično zbiranje in organizacijo informacij. Danes, ko je analitskih podatkov na pretek, manjkajo učinkovite metode za identifikacijo in validacijo vzročnih dejavnikov, njihovih funkcij ter interakcij.

Dostopnost raznovrstnih genomskih podatkov, skupaj z bioinformacijskimi metodami, omogoča raziskave mehanizmov kompleksnih bolezni in razvoj biooznačevalcev (angl. *biomarkers*), ki lahko pripomorejo k natančnejšim diagnozam ter razvoju terapij. V razvoj biooznačevalcev je usmerjenih veliko raziskav, saj nekatere kompleksne bolezni, kot je debelost, dosegajo epidemiološke razsežnosti.

1.1 CILJI RAZISKAVE IN HIPOTEZE

Cilji raziskave so 1) postaviti centralno spletno mesto za zbiranje raznovrstnih genetskih lokusov, poveznih z nalaganjem maščobe pri ljudeh in živalih, 2) razviti postopek za analizo kandidatnih genov in izbor najobetavnejših med njimi, ki bodo podlaga za nadaljnje eksperimentalno potrjevanje, 3) s pomočjo razvitega pristopa med vsemi zbranimi kandidatnimi geni za nalaganje maščobe poiskati najobetavnejše ter razviti potencialne biooznačevalce.

Postavili smo hipotezi:

- Z integracijo do sedaj poznanih genetskih vzrokov za nalaganje maščobe ter bioinformacijsko analizo je možno identificirati genske mreže in biološke poti, ki bi jih bilo možno razviti v biooznačevalce.

- Kandidatne gene za nalaganje maščobe je možno razvrstiti po prioriteti glede na njihovo lokacijo na genski mapi, njihovo število povezav v genski mreži in glede na njihovo vlogo v biološki poti.

1.2 PRISPEVKI K ZNANOSTI

Rezultat doktorske naloge je prosto dostopna zbirka kandidatnih lokusov za nalaganje maščobe, nov pristop za razvoj biooznačevalcev večgenskih bolezni z integracijo genomskih podatkov in bioinformacijsko analizo ter seznam potencialnih biooznačevalcev za nalaganje maščobe, preverjenih z analizo povezave genotipa s fenotipom na modelu miši. Zbirka kandidatnih genov za nalaganje maščobe je centralno spletno mesto za raziskovalce s tega področja. Pristop za razvoj biooznačevalcev je možno uporabiti tudi za druge kompleksne fenotipe, zato predstavlja pomemben prispevek k razvoju biooznačevalcev.

Gre za interdisciplinaren projekt, ki je močno vpet v mednarodni prostor (KAUST, WSU, MDACC) in zahteva tesno sodelovanje bioinformatikov in raziskovalcev s področja bioloških znanosti. Večina laboratorijev je še vedno bolj usmerjena v generiranje velike količine informacij, kot pa v integriranje obstoječih. Zaradi nezmožnosti velikih vlaganj v genomske raziskave in omejenih bioloških resursov bi konkurenčnost malim državam omogočila usmeritev v razvoj dobro organiziranih zbirk, na podlagi katerih bi z bioinformacijskimi metodami razvijali kandidatne biooznačevalce, in jih nato eksperimentalno preverjali. Raziskovanja na področju genomike lahko pripomorejo k napredku v zdravstvenem varstvu in znižanju stroškov s pomočjo natančnejših diagnoz, k napredku individualiziranega zdravljenja ter učinkovitejšim razvojnim potem do novih zdravil, terapij in drugih produktov novih tehnologij.

2 PREGLED OBJAV

2.1 NALAGANJE MAŠČOBE

Prekomerna telesna masa in debelost predstavljata problem svetovnih razsežnosti in sta največja dejavnika tveganja za razvoj sladkorne bolezen tipa 2, bolezni srca in ožilja, povečanega krvnega tlaka, dislipidemije, bolezni ledvic, respiratornih in mišičnih obolenj, zmanjšanja plodnosti, psiholoških težav in nekaterih vrst raka (Kopelman, 2007). Pojavnost debelosti je začela naraščati pred stotimi leti (Helmchen in Henderson, 2004) in se je iz nepojasnjenih razlogov hitro povečala v drugi polovici prejšnjega stoletja (Klimentidis in sod., 2011) ter začenja izpodrivati svetovne zdravstvene probleme, kot so podhranjenost in infekcijske bolezni (World Health Organ, 2000). Svetovna zdravstvena organizacija (angl. *World Health Organization*; WHO) je leta 1997 prvič poročala o epidemiološki razsežnosti pojavnosti debelosti (World Health Organ, 2000).

Hitro širjenje pojava debelosti je povezano s sodobnim življenjskim slogom. Zaradi preobilja visoko kalorične hrane in telesne neaktivnosti prihaja do kroničnega presežka energije, ki se kopiči v obliki maščobe. Debelost je dedna, pogojena je z interakcijo mnogih genov, okolijskih ter vedenjskih dejavnikov. Raziskave genetskih dejavnikov debelosti so zelo pomembne, saj je debelost eden največjih izzivov današnjega javnozdravstvenega sistema (Yang in sod., 2007). Poznavanje molekularnega mehanizma nalaganja maščobe je nujno za razvoj diagnostičnih biooznačevalcev. To znanje je pomembno tudi pri reji domačih živali, saj prekomerno nalaganje maščevja vpliva na gospodarnost reje in kakovost mesa.

2.1.1 Maščobno tkivo

Maščobno tkivo je kompleksen endokrini organ, ki poleg maščobnih celic (adipocitov), vsebuje tudi živčno tkivo, stromalno vaskualarne in imunske celice (Kershaw in Flier, 2004). Pri sesalcih ločimo dve vrsti maščobnega (adipoznega) tkiva, belo (angl. *white adipose tissue*; WAT) in rjavo (angl. *brown adipose tissue*; BAT). Belo maščobno tkivo predstavlja vsaj 10% telesne mase odraslega človeka in hrani kemično energijo v obliki trigliceridov. Razvoj belega maščobnega tkiva predstavlja evolucijsko prilagoditev v obdobjih pomanjkanja hrane. Celice belega maščobnega tkiva so specializirane za sintezo trigliceridov iz glukoze, kot tudi za uvažanje maščobnih kislin iz krvi. Bele maščobne celice izločajo peptidne in steroidne hormone, ki uravnavajo energijsko ravnovesje, metabolizem glukoze in lipidov, vaskularno homeostazo, imunski odziv in reprodukcijo (Guerre-Millo, 2002). Pri debelosti je ravnotežje izločanja hormonov maščobnega tkiva porušeno (Rajala in Scherer, 2003). Rjavo maščobno tkivo sodeluje pri vzdrževanju telesne temperature in porabi presežka energije. Pri mladičih predstavlja rjavo maščevje znaten delež v telesu in je prisotno na hrbtnem delu ter ob ramenih. Dolgo je veljalo prepričanje,

da pri odraslem človeku rjavega maščobnega tkiva ni. Morfološke študije so pokazale, da je rjavo maščobno tkivo prisotno in aktivno tudi pri odraslih. V majhnih količinah se nahaja okoli nadledvičnih žlez in velikih krvnih žil, v zgornjem delu prsnega koša in vratnem predelu. Maščobne celice rjavega maščobnega tkiva imajo v citoplazmi več manjših maščobnih kapljic, medtem ko imajo adipociti belega maščobnega tkiva samo eno večjo kapljico. Rjavo barvo jim dajejo številni mitohondriji, ki vsebujejo železo. Glavna funkcija rjavega maščevja je generacija toplote preko netresave termogeneze v primeru padca temperature pod vrednost, ki je določena v hipotalamusu. Raziskave kažejo, da rjavo tkivo sodeluje pri uravnavanju porabe energije, zaradi česar je pomemben dejavnik pri nastanku debelosti (Nedergaard in sod., 2007).

Leptin je hormon, ki ga izloča maščobno tkivo in ima ključno vlogo pri uravnavanju energijske homeostaze (ravnovesje med vnosom in porabo energije) (Considine in sod., 1996). Na izražanje leptina vplivajo inzulin, dejavnik tumorske nekroze-α, estrogeni proste maščobne kisline in rastni hormon (Saladin in sod., 1995). Leptin se prenaša s krvjo v hipotalamus in deluje kot signalna molekula, ki sporoča stanje energijskih zalog v telesu (Cheung in sod., 1997). Adiponektin je hormon, ki ga izločajo maščobne celice in kroži v krvni plazmi. Pri debelosti, odpornosti na inzulin, metabolnem sindromu in sladkorni bolezni tipa 2 je sinteza adiponektina zmanjšana (Hajer in sod., 2007).

Poznamo dva osnovna tipa debelosti, hiperplastično (povečano število maščobnih celic) in hipertrofično (povečanje volumna maščobnih celic). Hipertrofija adipocitov je prisotna pri vseh prekomerno prehranjenih in debelih odraslih, medtem ko je hiperplazija adipocitov značilna le za resne oblike debelosti (Hirsch in Batchelor, 1976). Število maščobnih celic je določeno do zgodnje odraslosti, do sprememb mase maščobe pri odraslih pa prihaja samo preko spreminjanja volumna maščobnih celic. Pri odraslem človeku se letno obnovi približno 10 odstotkov maščobnega tkiva. Raziskave so pokazale, da se stopnji odmiranja in nastajanja novih maščobnih celic (adipogeneza) pri odraslih ne spreminjata niti v stanju debelosti, kar nakazuje na to, da je tudi takrat število adipocitov natančno uravnavano (Spalding in sod., 2008).

2.1.2 Lokusi, povezani z nalaganjem maščobe

Študije dvojčkov, posvojencev in družin so pokazale visoko stopnjo dednosti debelosti (Stunkard in sod., 1986a; Stunkard, in sod., 1986b; Rice in sod., 1999). Ocena stopnje dednosti je po podatkih študije ocenjena na 50% do 70% (Allison in sod., 1996). Monogeno debelost povzročijo mutacije v enem genu. Te oblike debelosti so zelo redke, resne, običajno se začnejo v otroštvu (Farooqi in O'Rahilly, 2004). Do leta 2007 je bilo znanih 11 različnih genov pri človeku, ki povzročijo monogeno debelost (*CRHR1*, *CRHR2*, *GPR24*, *LEP*, *LEPR*, *MC3R*, *MC4R*, *NTRK2*, *POMC*, *PCSK1* in *SIM1*) (Rankinen in sod., 2006). Sindromna debelost se pojavi pri vsaj 20 redkih sindromih, ki jih povzročajo genetske mutacije ali kromosomske nepravilnosti. Večino sindromnih debelosti spremlja

tudi prizadetost, zato je treba pri iskanju genetskih dejavnikov upoštevati tudi okoljske. Najpogostejša oblika sindromne oblike debelosti (1 na 25000 rojstev) je sindrom Prader–Willi (PWS) (Bell in sod., 2005). Večgenska ali navadna debelost se pojavi zaradi interakcij med številnimi geni in okoljskimi dejavniki. Leta 2012 je zabeleženih 1736 lokusov, ki so povezani z debelostjo (Kunej in sod., 2012). Nekateri izmed genov, ki so odgovorni za nastanek monogene debelosti, so vpleteni tudi v večgensko debelost (*MC4R*, *PCSK1*, *POMC* in *BDNF*) (Hirschhorn, 2009). Polimorfizmi v genu *FTO* imajo močan vpliv na fenotip pri večgenski obliki debelosti (Speliotes in sod., 2010). Vloga FTO pri uravnavanju energijskega ravnotežja je očitna. Izguba funkcije gena *FTO* pri miših pripelje do zmanjšanja maščobnega tkiva in nizke telesne mase (Fischer in sod., 2009). O močnem prispevku pri večgenski obliki debelosti so poročali tudi za gene, kot so *TMEM18*, *KCTD15*, *GNPDA2*, *SH2B1*, *MTCH2*, *NEGR1* (Willer in sod., 2009) in *DEPTOR* (Laplante in sod., 2012). Z razvojem maščobnih celic so študije povezale transkripcijske dejavnike (*PPARγ*, *C/EBP* in *ADD1/SREBP1c*) (Rosen in sod., 2000). Pomembno regulatorno vlogo v mnogih bioloških procesih, povezanih z debelostjo (diferenciacija maščobnih celic, inzulin, metabolizem maščobe), imajo mikro RNA (angl. *microRNA*; miRNA) molekule. Zadnje študije so pokazale, da so miRNA v maščobnem tkivu pri debelosti čezmerno ali premalo izražene (Kunej in sod., 2010; McGregor in Choi, 2011). Hitro povečanje pojavnosti debelosti poudarja vlogo dejavnikov iz okolja. Vzroke za epidemijo debelosti iščejo tudi v epigenetskih mehanizmih. Na vzorce metilacije DNA genov, ki povečajo možnost pojava debelosti, vpliva sodoben način življenja (Gluckman in Hanson, 2008; Haemer in sod., 2009; Newnham in sod., 2009; Herrera in sod., 2011). Raziskave debelosti potekajo tudi na živalskih modelih, kot so miši, podgane in nečloveški primati (angl. *non-human primates*) (Speakman in sod., 2008). Kot model za raziskave debelosti pri človeku se vedno večkrat pojavlja tudi prašič (Houpt in sod., 1979).

2.1.3 Biološke poti in biološki procesi, povezani z nalaganjem maščobe

Pri obolelih za debelostjo obstaja povečano tveganje za razvoj zdravstvenih težav, kot so inzulinska rezistenca in diabetes tipa 2, hipertenzija, dislipidemija, srčno-žilne bolezni, možganska kap, apneja v spanju (angl. *sleep apnea*), bolezen žolčnika, hiperurikemia, putika in osteoartritis. Tudi nekatere vrste raka so povezane z debelostjo, na primer rak debelega črevesa, prostate, dojke, žolčnika in endometrija. Genski mehanizem razvoja debelosti in sočasnih bolezni (angl. *co-morbidity*) še ni raziskan (Khaodhiar in sod., 1999). V povezavi z debelostjo je bilo opravljenih veliko asociacijskih študij na ravni celotnega genoma (angl. *genome-wide assciation study*; GWAS), ki se običajno osredotočajo na polimorfizme posameznih nukleotidov (angl. *single-nucleotide polymorphism*; SNP) brez upoštevanja bioloških interakcij med raziskovanimi geni. V eni izmed asociacijskih študij na ravni celotnega genoma, ki je temeljila na bioloških poteh, so z debelostjo povezali biološko pot »*vasoactive intestinal peptide*« (VIP) (Liu in sod., 2010).

2.1.4 Zbirke lokusov, povezanih z debelostjo

Lokuse, povezane z debelostjo, so deset let zbirali v zbirki *Obesity Gene Map* (Bouchard in Pérusse, 1996). Posodobitve so objavljali enkrat letno v reviji *Obesity,* leta 2000 pa je prvič izšla tudi spletna različica. Iz več virov (PubMed, OMIM) so zbirali lokuse (QTL, gene, kromosomske regije, označevalce), ki so jih z debelostjo povezali v različnih študijah. Leta 2006 so zaradi preobsežnosti količine podatkov in omejenih finančnih sredstev zbirko prenehali posodabljati (Rankinen in sod., 2006). Zadnja objavljena zbirka *Obesity Gene Map* vsebuje okoli 430 lokusov, ki so povezani z debelostjo. V prosto dostopnih zbirkah *AnimalQTLdb* (Hu in sod., 2007) in *Rat Genome Database* (De la Cruz in sod., 2005) so zbrani kvantitativni lokusi (angl. *Quantitative Trait Loci*; QTL) za lastnosti pri domačih živalih, med katerimi so tudi tisti, povezani z nalaganjem maščobe. Ker se je kazala potreba po centralni podatkovni zbirki, smo leta 2012 izdelali prosto dostopno spletno zbirko *Obesity Gene Atlas in Mammals*, ki vsebuje 1736 lokusov, povezanih z nalaganjem maščobe (Kunej in sod., 2012). V zbirki *Obesity Gene Atlas in Mammals* so zbrani lokusi iz vseh zgoraj naštetih virov.

2.1.5 Genske terapije za zdravljenje debelosti

Današnje zdravljenje debelosti temelji na omejevanju dnevnega vnosa kalorij in povečevanju telesne aktivnosti, farmakoterapiji in kirurgiji (Melnikova in Wages, 2006). Omejevanje vnosa kalorij in povečevanje telesne aktivnosti je pogosto neučinkovito (zmanjšanje telesne mase za približno 7-10%), zdravila pa imajo mnoge stranske učinke. Zaradi vedno večjega števila obolelih je nujen razvoj novih terapij za zdravljenje debelosti (Ahima, 2002). Možnosti za zdravljenje debelosti se kažejo v uporabi receptorskih ligandov (antagonistov ali agonistov), ali inhibitorjev znotrajceličnih signalnih mehanizmov, ki so povezani z biološkimi potmi za nalaganje maščobe. V razvoju so tudi tehnike, kjer z vnosom virusnih vektorjev, ki kodirajo določen gen (npr. *LEP* ali *POMC*), nadomestijo izgubo funkcije tega gena (Li in sod., 2003).

2.2 RAZISKOVANJE KOMPLEKSNIH BOLEZNI

Genske bolezni povzročijo spremembe v enem ali več genih. Do sedaj poznamo več kot 1800 bolezni, ki jih povzročijo mutacije v enem genu (anemija srpastih celic, Marfanov sindrom, Huntingtonova bolezen) (O'Connor in Crystal, 2006). Takšnim boleznim pravimo monogene. Večina monogenih bolezni je redkih. Mnogo več bolezni, med njimi tudi zelo pogoste (rak, diabetes, bolezni srca in ožilja), povzročijo mutacije več genov (Gibson, 2009) in interakcije med njihovimi produkti in okoljem (Motulsky, 2006). Odkrivanje vzročnih genov za določen fenotip je počasno in zahtevno (Hardy in Singleton 2009). Za razumevanje genetskih mehanizmov bolezni ni dovolj poznati seznam vzročnih genov. Potrebno je razvozlati povezave med celičnimi enotami, ki jih uravnavajo vzročni

geni ter njihovi produkti. Zapletena mreža povezav med celičnimi enotami in vzročnimi geni odpre vprašanje o medsebojni povezanosti med boleznimi (Barabási, 2007).

2.2.1 Sistemska biologija

Biomolekule, udeležene v biološke procese, delujejo usklajeno, so organizirane v bioloških mrežah in poteh. Molekularna biologija je bila usmerjena v raziskave posamičnih interakcij med biomolekulami. Takšni pristopi so pripomogli k razumevanju omejenega števila signalnih poti in funkcij proteinov v celicah. Napredek visokozmogljivih metod omogoča raziskave na ravni celotnega biološkega sistema. Zanimanje za raziskave na sistemski ravni izhaja iz prepričanja, da imajo celotni sistemi funkcije, ki jih nobena od posameznih podenot nima, in, da je celota več kot vsota njenih delov (Bader in sod., 2008). Prehod od molekularne na sistemsko raven v biologiji obljublja revolucijo v razumevanju kompleksnih bioloških sistemov in zagotavlja nove možnosti praktične uporabe tega znanja (Bader in sod., 2008). Izraz sistemska biologija (angl. *systems biology*) se je pojavil pred dobrim desetletjem (Ideker in sod., 2001). V tem času se je sistemska biologija razvila, uporablja vse več eksperimentalnih in računalniških metod in vedno več je v uporabi (Chuang in sod., 2010). Sistemska biologija je pravzaprav logično nadaljevanje funkcijske genomike (Aggarwal in Lee, 2003; Auffray in sod., 2003). Razumevanje biološkega sistema na sistemski ravni pomeni razumevanje strukture in dinamike sistema ter metod za uravnavanje in načrtovanje sistema. Razumevanje strukture sistema zajema poznavanje mreže genskih interakcij in bioloških poti ter poznavanje mehanizmov, s katerimi interakcije uravnavajo znotrajcelične in zunajcelične strukture. Razumevanje dinamike sistema zajema poznavanje obnašanja sistema v različnih pogojih. Z namenom, da bi bili zmožni zmanjšati okvare, poiskati terapevtske tarče, spreminjati in konstruirati biološke sisteme, pa je potrebno poznati metode uravnavanja in načrtovanja bioloških sistemov (Kitano, 2002). Sistemski pristop zahteva raziskave iz različnih vidikov sistema, zato je treba povezati genomske informacije na različnih ravneh (genomika v ožjem smislu; raven DNA, transkriptomika; raven RNA, proteomika, metabolomika).

2.2.2 Strategije raziskovanja kompleksnih bolezni

Odkrivanje genov, ki povzročajo bolezni, razumevanje interakcij med njimi in ključnih bioloških poti, je glavni izziv današnjih raziskav v genetiki. Pogosto uporabljene pristope k odkrivanju vzročnih genov lahko v grobem razdelimo v dve skupini, to so študije kandidatnih genov in analize povezanosti na celotnem genomu (angl. *genome-wide linkage studies*) (Bell in sod., 2005). Pri študijah kandidatnih genov genotipiziramo vzorce za polimorfni označevalec znotraj kandidatnega gena. Kandidatni geni so lahko funkcijski ali pozicijski. Funkcijski kandidatni geni so vključeni v patogenezo bolezni. Odkrivanje funkcijskih kandidatnih genov je odvisno od trenutnega znanja o fenotipu. Pozicijski kandidatni geni so tisti, ki se nahajajo znotraj kromosomskih regij, ki so povezane s fenotipom. Pozicijski kandidatni geni se lahko nahajajo znotraj kromosomskih mutacij, ali

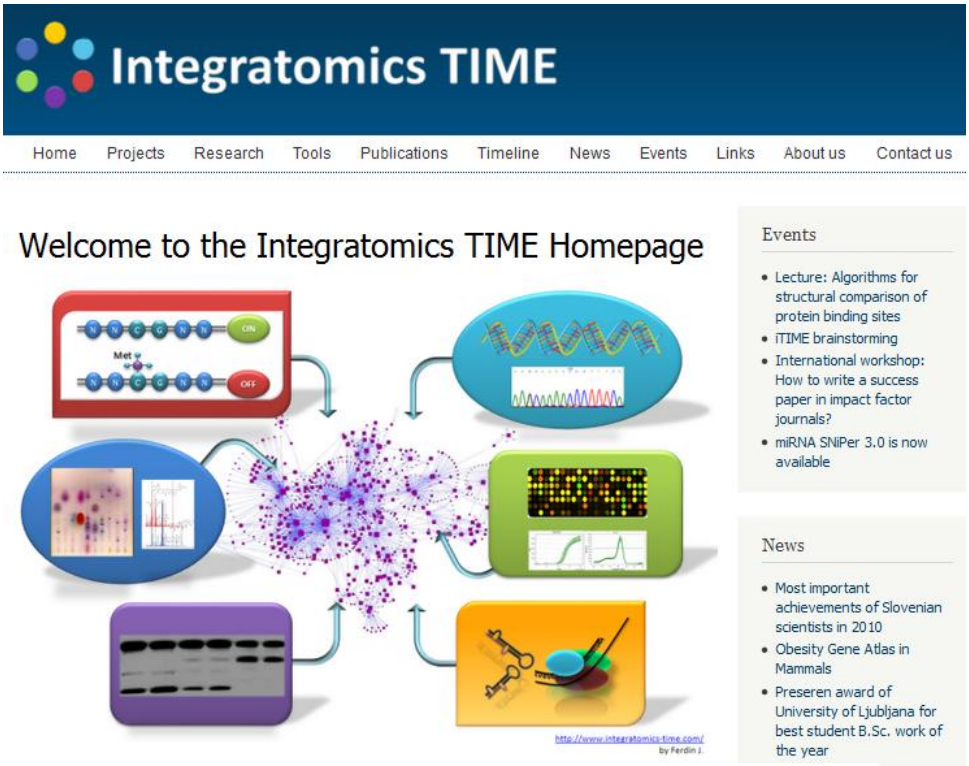
pa so jih s fenotipom povezali s pomočjo asociacijskih študij (angl. *association studies*) ali študij genske povezanosti (angl. *linkage studies*). Pri analizah povezanosti na celotnem genomu genotipiziramo vzorce za polimorfne označevalce vzdolž celotnega genoma in ocenimo stopnjo povezanosti označevalcev s fenotipom. V tem primeru odkrivanje genov ni odvisno od predhodnega znanja o fenotipu. Pri študijah genske povezanosti na ravni celega genoma (angl. *genome-wide linkage studies*) iščemo polimorfne označevalce vzdolž genomov znotraj družine in računamo stopnjo povezanosti označevalca z boleznijo ali lastnostjo. Na ta način zaznamo regije na genomu, ki so povezane z opazovanim fenotipom. Kadar gre za kvantitativne lastnosti, lahko govorimo o kvantitativnih lokusih. Z namenom, da bi razvili še bolj učinkovite metode za iskanje vzročnih genov, mnoge raziskovalne skupine združujejo različne pristope. Orodje BioMercator omogoča iskanje kandidatnih genov na podlagi analize kvantitativnih lokusov več neodvisnih študij (Arcade in sod., 2004).

2.2.3 Integracija »omskih« podatkov ali integratomika

Pristop GWAS se je izkazal kot zelo učinkovit pri odkrivanju genetskih lokusov, povezanih z mnogimi fenotipi (Hindorff in sod., 2009), vendar ne omogočajo etioloških vpogledov v bolezni, učinkovitejšega napovedovanja in zdravljenja bolezni (Loos in Schadt, 2012). Možnosti za boljši vpogled v molekularne mehanizme bolezni se odpirajo z integriranjem rezultatov, pridobljenih z dosedanjimi študijami, zato interes za zbiranje in integriranje raznovrstnih fenotipskih in genotipskih podatkov narašča (Loos in sod., 2012).

Kompleksnost in raznolikost bioloških sistemov zahtevata integracijo raznovrstnih podatkov s pomočjo analitskih in računalniških tehnologij. Integracija podatkov je nujna tako znotraj posameznih ravni genomskih raziskav (razpršeni podatki, pridobljeni z različnimi pristopi) kot tudi med različnimi ravnmi genomskih raziskav (genomika, proteomika, transkiptomika, epigenomika). Prav tako vsi poskusi niso možni pri vseh vrstah (inbridirane linije in poskusi z izbijanjem genov (angl. *knockout*; KO)), zato je pomembna integracija podatkov na ravni različnih vrst (primerjalna genomika).

Primeri študij, kjer so raziskovalci uporabili integrativni pristop, so raziskave genov za odpornost proti mastitisu (Ogorevc in sod., 2008), razvoj mlečne žleze in prireje mleka (Ogorevc in sod., 2009), shizofrenije (Jia in sod., 2010), reprodukcije pri moškem (Ogorevc in sod., 2011), raka dojke (Mosca in sod., 2010) in kriptorhizma (retencije testisov) (Cannistraci in sod., 2013). V naštetih študijah so zbrali lokuse, povezane z raziskovanim fenotipom, ter z različnimi bioinformacijskimi pristopi (analiza bioloških poti, analiza genskih mrež) analizirali molekularne mehanizme kompleksnih fenotipov. V okviru projekta Integratomics TIME (Slika 1) so zbrane študije nekaterih kompleksnih bolezni in predstavljene strategije za raziskavo kompleksnih fenotipov s pristopom integracije genomskih podatkov (Kunej in sod., 2012; Cannistraci in sod., 2013).



Slika 1: Vhodna stran projekta Integratomics TIME (http://integratomics-time.com) za raziskave kompleksnih fenotipov

Figure 1: The entry page of the Integratomics TIME (http://integratomics-time.com) project for the research of complex phenotypes

2.3 ZBIRANJE LOKUSOV, POVEZANIH S FENOTIPOM

Kompleksne fenotipe oblikujejo interakcije med številnimi geni in okoljem. Osnova za raziskave kompleksnih fenotipov je poznavanje vzročnih genov, zato je nujno zbiranje lokusov, povezanih z določenimi fenotipi.

2.3.1 Specializirane zbirke lokusov, povezanih s fenotipom

Primeri prosto dostopnih zbirk, ki vsebujejo lokuse, povezane z določenim fenotipom, so T2D-Db z lokusi, povezanimi z diabetesom tipa 2 (Agrawal in sod., 2008), T1DBase z lokusi, povezanimi z diabetesom tipa 1 (Burren in sod., 2011), Obesity Gene Atlas z lokusi, povezanimi z nalaganjem maščobe (Kunej in sod., 2012) in ostale (Preglednica 1). Na spletni strani HUGO Mutation Database Initiative (MDI) (Cotton in sod., 1998) je seznam nekaterih izmed specializiranih zbirk lokusov, povezanih z boleznimi pri človeku (http://www.hgvs.org/dblist/disease.html).

Preglednica 1: Specializirane podatkovne zbirke lokusov, povezanih z boleznimi pri človeku

Table 1: Specialized databases of loci associated with human diseases

Bolezen/ Fenotip	**Ime zbirke**	**Povezava**	**Dostopnost, zadnja posodobitev**	**Referenca**
Albinizem	Albinism database	http://albinismdb.med.umn.edu/	Dostopna, 2009	
Astma	Asthma and Allergy Database	http://cooke.gsf.de	Ni dostopna	(Immervoll in Wjst 1999)
	Asthma Local Gene Database	https://research.cchmc.org/mershalab/AsthmaGeneBrowser/Home.html	Dostopna	
Alzheimerjeva bolezen	AlzGene	http://www.alzgene.org/	Dostopna	(Bertram in sod., 2007)
Debelost	Obesity Gene Map		Ni dostopna	(Rankinen in sod., 2006)
	Obesity Genomic Atlas	http://www.integratomics-time.com/fat_deposition	Dostopna	(Kunej in sod., 2012)
Diabetes 1	T1DBase	http://www.t1dbase.org/page/Welcome/display	Dostopna	(Burren in sod., 2011)
Diabetes 2	T2D-Db	http://t2ddb.ibab.ac.in/home.shtml	Dostopna	(Agrawal in sod., 2008)
Kriptorhizem	Cryptorchidism Genomic Atlas	http://www.integratomics-time.com/cryptorchidism	Dostopna, 2012	(Cannistraci in sod., 2013)
Možganska kap	SigCS base	http://sysbio.kribb.re.kr/sigcs/pageHome.php?m=h	Dostopna	(Park in sod., 2011)

se nadaljuje

nadaljevanje

Bolezen/ Fenotip	**Ime zbirke**	**Povezava**	**Dostopnost, zadnja posodobitev**	**Referenca**
Izguba sluha	Hereditary Hearing Loss Homepage	http://hereditaryhearingloss.org/	Dostopna	
Reprodukcija pri moškem spolu sesalcev	Genomic atlas of male reproduction in mammals	http://www.integratomics-time.com/male_infertility	Dostopna	(Ogorevc in sod., 2011)
Periferne nevropatije	Mutation Database of Inherited Peripheral Neuropathies (IPNMDB)	http://www.molgen.ua.ac.be/CMTMutations/Home/IPN.cfm	Dostopna	
Parkinsonova bolezen	Parkinson's disease	http://www2.cancer.ucl.ac.uk/Parkinson_Db2/index.php	Dostopna, december 2010	
Rak	Roche Cancer Genome Database (RCGDB)	http://rcgdb.bioinf.uni-sb.de/MutomeWeb/	Dostopna	(Küntzer in sod., 2011)
Rak dojke	Genes-to-Systems Breast Cancer Database (G2SBC)	http://www.itb.cnr.it/breastcancer/	Dostopna	(Mosca in sod., 2010)
Samovnetne bolezni	Infevers	http://fmf.igh.cnrs.fr/ISSAID/infevers/	Dostopna	(Touitou in sod., 2004; Milhavet in sod., 2008)
Shizofrenija	Schizophrenia Gene Resource (SZGR)	http://bioinfo.mc.vanderbilt.edu/SZGR/index.jsp	Dostopna	(Jia in sod., 2010)
Staranje	GenAge Database of Ageing-Related Gene	http://genomics.senescence.info/genes/	Dostopna	(Tacutu in sod., 2013)
Vnetna bolezen črevesa	IBDsite	http://www.itb.cnr.it/ibd/	Dostopna	(Merelli in sod., 2012)

V zbirki OMIM (*Mendelian Inheritance in Man*) so podatki o genih, ki so povezani z boleznimi. Zbirka je ročno pregledana in vsebuje podatke o več kot 4500 fenotipih s poznano molekularno osnovo ter opise skoraj 3000 genov z mutacijami, ki vplivajo na fenotip. Pomanjkljivost zbirke OMIM pa je, da jo zelo počasi posodabljajo in vsebuje samo lokuse z zelo močnim vplivom na fenotip. Pri raziskavah kompleksnih fenotipov pa so pomembni tudi lokusi z manjšim vplivom na fenotip. Najdemo jih lahko v zbirki GAD (*Genetic Association Database*) (Masseroli in sod., 2005), ki hrani podatke asociacijskih študij. V zbirki CTD (*Comparative Toxicogenomics Database*) so zbrani podatki o interakcijah med kemičnimi spojinami iz okolja in proteini v povezavi z boleznimi pri vretenčarjih in nevretenčarjih (Davis in sod., 2013). V zbirki DisGeNET (Bauer-Mehren in sod., 2011) so podatki o vzročnih genih za več kot 6000 bolezni pri človeku. Zbirka vsebuje ročno pregledane podatke iz literature, vključuje pa tudi podatke iz zgoraj naštetih zbirk (OMIM, GAD, CTD).

2.3.2 Zbiranje lokusov iz bibliografskih zbirk

V zadnjih dveh desetletjih smo priča hitremu porastu količine genomskih podatkov in objavljenih publikacij na področju biomedicine. Čeprav so genomski podatki in publikacije osnova za raziskave, sistematična integracija genomskih podatkov z literaturo zaostaja. Podatki o nukleotidnih zaporedjih genomov so zbrani in prosto dostopni preko spletnih portalov Ensembl (Flicek in sod., 2013), UCSC (Meyer in sod., 2013), FlyBase (Drysdale in FlyBase Consortium 2008), *Saccharomyces Genome Database* (Dwight in sod., 2004), WormBase (Yook in sod., 2012) in miRBase (Kozomara in Griffiths-Jones 2011). Naštete zbirke so integrirane z viri genomskih podatkov, kot so zbirka podatkov o izražanju genov ArrayExpress (Parkinson in sod., 2011) ter zbirka proteinov UniProt (Consortium, 2013). Literatura s področja biomedicine je dostopna preko spletnega portala PubMed (Lu, 2011). Informacije v besedilih niso strukturirane tako, da bi omogočale učinkovito računalniško analizo. Povezati je treba genomske lokuse s fenotipi, v publikacijah pa se pojavljajo različna poimenovanja tako za lokuse (Chen in sod., 2005; Tamames in Valencia, 2006) kot za fenotipe. Ob vnosu podatkov o novih publikacijah v PubMed informacija o genomski regiji, na katero se besedilo nanaša, ni vnesena. Za nekatere modelne organizme strokovnjaki redno pregledujejo objave in jih povezujejo z geni in genomskimi regijami (Hirschman in sod., 2010), kljub temu pa ostaja na tisoče publikacij, nepovezanih z regijami na genomih (Kersey in Apweiler, 2006). Ensembl2pubmed (Baran in sod., 2011) je razširitev orodja BioMart. Ensembl2pubmed omogoča iskanje publikacij, ki so povezane z vneseno ključno besedo, ter sezname genov, na katere se publikacije nanašajo.

2.3.3 Ontologije fenotipov in bolezni

Cilj konzorcija za ontologijo genov (angl. *Gene Ontology*; GO) je razvoj ontologije, ki omogoča pripisovanje molekularnih funkcij in bioloških procesov genom pri različnih vrstah. Iz zbirke genov s pripisanimi funkcijami iz ontologije GO lahko dobimo sezname vseh genov z določeno molekularno funkcijo ali sezname tistih genov, ki sodelujejo pri določenih bioloških procesih. Poznamo tudi projekte ontologije bolezni in fenotipov, kot so HPO (*Human Phenotype Ontology*) (Robinson in sod., 2008), DO (*Disease Ontology*) (Osborne in sod., 2009), MP (*Mammalian Phenotype Ontology*) (Smith in sod., 2005) in MPATH (*Mammalian Phenotype Ontology*) (Schofield in sod., 2010). Genom človeka je anotiran z biološkimi funkcijami iz ontologije GO, s fenotipi in boleznimi pa zelo malo. Anotacija genov z ontologijami fenotipov in boleznimi bi olajšala iskanje genov, povezanih z boleznimi in lastnostmi.

2.3.4 Orodja za rudarjenje v podatkih

GeneRIF (*Gene Reference Into Function*) (Mitchell in sod., 2003) omogoča vnos funkcijskih anotacij za gene v zbirki Entrez Gene. Vsak GeneRIF vnos vsebuje anotacijo z največ 255 črk, povezavo do publikacije, v kateri je opisan gen, in elektronski naslov vnašalca. Anotacije lahko vnaša kdorkoli, večinoma pa jih vnašajo strokovnjaki z NCBI. MILANO (Microarray Literature-based Annotation) (Rubinstein in Simon 2005) je orodje, s pomočjo lahko poiščemo publikacije, v katerih se hkrati pojavljajo geni in ključne besede, ki jih vnese uporabnik. Orodje MILANO omogoča preiskovanje po zbirki publikacij PubMed in v GeneRIF vnosih. Orodje je uporabno za anotacijo rezultatov mikromrež. Podobni orodji sta še PubMatrix (Becker in sod., 2003) in B.E.A.R. GeneInfo (Zhou in sod., 2004).

2.4 BIOLOŠKE POTI

Biološke poti (metabolične, signalne, regulatorne) so množice proteinov in drugih biomakromolekul, ki predstavljajo prostorsko predstavljene kaskade interakcij, ki so odgovorne za določen fenotipski izid. Biološke poti so idealizirani modeli. Od zelo zapletene mreže interakcij genov in proteinov so ločeni glede na celično ali fiziološko funkcijo. Hitro naraščanje zanimanja za analizo bioloških poti je sprožila dostopnost visoko zmogljivih metod in obsežnih študij izražanja genov. Z integracijo raznovrstnih informacij, kot so funkcije genov in proteinov, mreže molekularnih interakcij in biološke poti, lahko proučujemo biološki sistem na sistemski ravni.

Večina informacij o bioloških poteh je razbrana iz znanstvene literature. Eksperti preberejo in povzamejo na tisoče objav na vnaprej določen način. V izdelavo zbirk bioloških poti je treba vložiti veliko truda (časa, finančnih vlaganj), zato so nekatere zbirke plačljive, kot na primer: Ingenuity Pathway Analysis (IPA), Linnea Pathways, MetaCore, PathArt in

ResNet. Prosto dostopne zbirke bioloških poti so: BioCarta, BioCyc, GenMAPP, Kyoto Encyclopedia of Genesand Genomes (KEGG), MIPS CYGD (Comprehensive Yeast Genome Database), PANTHER, Protein Lounge, Reactome, Science Database of Cell Signaling in Saccharomyces Genome Database (SGD) (Michael Cherry in sod., 2012). V metapodatkovni zbirki Pathguide (Bader in sod., 2006) je zbranih več kot 190 povezav do zbirk bioloških poti in mrež.

Pristope k analizi bioloških poti lahko razdelimo v dve skupini, na analizo kandidatnih bioloških poti in na analizo bioloških poti na ravni celotnega genoma. Pri analizi kandidatnih poti testiramo povezanost fenotipa s kandidatnimi potmi, ki jih izberemo na podlagi predhodnega znanja. Pri analizi bioloških poti na ravni celotnega genoma ugotavljamo povezanost med fenotipom in biološkimi potmi na podlagi genomskih podatkov (Ramanan in sod., 2012). Z orodji za analizo bioloških poti, kot so Ingenuity Pathway Analysis (IPA), MetaCore, DAVID (Huang da in sod., 2007) in PathJam (Glez-Peña in sod., 2010), je možno integrirati podatke o bioloških poteh pri človeku iz različnih virov.

2.5 BIOLOŠKE MREŽE

Biološke genske mreže so robustne in odporne proti mutacijam in spremembam okolja (Leclerc, 2008). Robustnost in kompleksnost celičnih sistemov sta tesno povezani in zahtevni za razumevanje (Carlson in Doyle 2002). Nekatere izmed značilnosti genskih mrež, ki omogočajo robustnost celičnih sistemov, so modularnost, povratne zanke in redundanca (Queitsch in sod., 2012). Hkrati pa so biološki sistemi zelo občutljivi na nekatere spremembe (Stelling in sod., 2004). Medtem ko mnoge mutacije ne vplivajo na določen fenotip, lahko posamezne mutacije oziroma kombinacije mutacij sprožijo vrsto sprememb v genski mreži in posledično razvoj bolezni (Zhu in sod., 2007). Raziskave genskega mehanizma kompleksnih bolezni so možne s pomočjo analize topologije in dinamike bioloških mrež, značilnih za bolezensko stanje (Del Sol in sod., 2010). Glede na to, da temelji večina bioloških procesov v organizmu na interakcijah med proteini, so mreže proteinskih interakcij osnova za raziskave bioloških sistemov. Čeprav so mreže proteinskih reakcij predstavljene kot statične mreže proteinov in interakcij med njimi, je treba upoštevati dinamično naravo bioloških sistemov. Funkcijsko stanje mreže je odvisno od stopnje izražanja proteinov (Han in sod., 2004) , ki jo uravnava vrsta regulatornih mehanizmov. Proteini povezovalniki (angl. *hubs*), ki tvorijo veliko interakcij z ostalimi proteini, imajo najpomembnejšo vlogo pri ohranjanju funkcionalnosti mreže proteinskih interakcij (Barabási in Oltvai, 2004). V proteomu pri človeku ločimo dve vrsti povezovalnikov, intramodularne in intermodularne (Taylor in sod., 2009; Dong in sod., 2011). Intramodularni povezovalniki imajo nizek, intermodularni povezovalniki pa visok koeficient grupiranja, njihova vloga je uravnavanje ostalih modulov. Mutacije v intermodualrnih povezovalnikih so bile močneje povezane s fenotipi raka od mutacij v intramodularnih povezovalnikih (Wang in Marcotte, 2010). Izkazalo se je, da je

mehanizem uravnavanja izražanja genov z molekulami miRNA bolj kompleksen in pomembnejši pri intermodularnih kot pri intramodularnih povezovalnikih (Liang in Li, 2007).

2.5.1 Mreže proteinskih interakcij

Proteinske interakcije (angl. *protein-protein interactions*; PPIs) igrajo ključno vlogo v celičnih procesih (Gavin in sod., 2006); spremembe v interakcijah med proteini povzročajo razvoj bolezni. Sistem proteinskih interakcij lahko predstavimo z mrežo proteinskih interakcij (angl. *protein interaction network;* PIN), kjer so proteini vozlišča, interakcije med njimi pa povezave. Za odkrivanje in karakterizacijo proteinskih interakcij poznamo eksperimentalne, za napovedovanje proteinskih reakcij pa računalniške tehnike. Eksperimentalne tehnike, kot so rentgenska kristalografija (angl. *X-ray christalography*), fluorescenčni prenos resonančne energije (angl. *fluorescence resonance energy transfer*), plazmonska resonanca (angl. *surface plasmon resonance*), atomska mikroskopija (angl. *atomic force microskopy*; AFM), elektronska mikroskopija (angl. *electron microscopy*), karatkerizirajo posamične proteinske interakcije, za identifikacijo proteinskih interakcij na ravni celega genoma pa poznamo tehnike, kot so dvohibridni sistem kvasovke (angl. *yeast two-hybrid*; Y2H), masna spektroskopija (angl. *mass spectroscopy;* MS), DNA in proteinske mikromreže, sintetična letalnost (angl. *synthetic lethality*) ter predstavitev na fagu (phage display). Podatki o proteinskih interakcijah (pridobljeni eksperimentalno in *in silico*) so zbrani v podatkovnih zbirkah (Mathivanan in sod., 2006). Nekatere izmed zbirk z eksperimentalnimi podatki o proteinskih interakcijah so DIP (Database of Interacting Proteins) (Xenarios in sod., 2002), BIND (Biomolecular Interaction Network Database) (Isserlin in sod., 2011), MINT (The Molecular INTeraction database) (Licata in sod., 2012), IntAct (Kerrien in sod., 2012), BioGRID (The Biological General Repository for Interaction Datasets) (Chatr-Aryamontri in sod., 2013), HPRD (Human Protein Reference Database) (Keshava Prasad in sod., 2009). Zbirki proteinskih interakcij, ki so zbrane iz literature, sta Stitch (Kuhn in sod., 2012) in STRING (Franceschini in sod., 2013). V zbirki STRING so poleg eksperimentalnih podatkov tudi računalniške napovedi proteinskih interakcij.

2.5.2 Mreže uravnavanja izražanja genov (DNA-protein, RNA-RNA)

Mreže uravnavanja izražanja genov (angl. *gene regulatory network;* GRN) vsebujejo informacije o uravnavanju izražanja genov. Proces uravnavanja izražanja genov je pod vplivom več dejavnikov. To so transkripcijski dejavniki (angl. *transcription factor*; TF) (Carninci in sod., 2005), potranslacijske spremembe (angl. *post-translational modifications*) in povezovanje z drugimi biomolekulami (Linding in sod., 2008). Transkripcijski dejavniki so proteini, ki se vežejo na DNA. Večina transkripcijskih dejavnikov se veže na več vezavnih mest v genomu in tvori kompleksno mrežo uravnavanja genov (Ravasi in sod., 2010). Človeški genom kodira 1400 transkripcijskih

dejavnikov, ki uravnavajo izražanje več kot 1400 genov (Vaquerizas in sod., 2009). Tehnologije za raziskavo mrež genskega uravnavanja so ChIP-chip (angl. *immunoprecipitation* (*"ChIP"*) *with microarray technology* (*"chip"*)), ki je kombinacija kromatinske imunoprecipitacije in mikromrež, ChIP-seq (angl. *ChIP-sequencing*), ki je kombinacija kromatinske imunoprecipitacije in sekveniranja, in CliP-seq (angl. *cross-linking immunoprecipitation-high-throughput sequencing*). Informacije o interakcijah med proteini in DNA so zbrane v zbirkah, kot so na primer JASPAR (Portales-Casamar in sod., 2010), TRANSFAC (Wingender in sod., 1996), Human B-cell interactome (HBCI) (Lefebvre in sod., 2010), ENCyclopedia Of DNA Elements (ENCODE) (Consortium, 2011) in Transcription Factor Encyclopedia (TFe) (Yusuf in sod., 2012). Informacije o interakcijah med molekulami miRNA in njihovimi tarčami so zbrane v MicroCosm Targets (Griffiths-Jones in sod., 2006), miRecords (Xiao in sod., 2009) in miRTarBase (Hsu in sod., 2011). Informacije o potranslacijskih spremembah so zbrane v zbirkah Phospho.ELM (Dinkel in sod., 2011), NetPhorest (Miller in sod., 2008) in PHOSIDA (Gnad in sod., 2011). V zbirki DrugBank so informacije o zdravilih in njihovih tarčah (Knox in sod., 2011).

2.5.3 Metabolne mreže

Kemične spojine v celici so med seboj povezane z biokemijskimi reakcijami, ki pretvorijo eno spojino v drugo. Reakcije katalizirajo encimi. Tako so vse spojine v celici del zapletene biokemijske mreže reakcij, ki jo imenujemo metabolna mreža (angl. *metabolic network*). V metabolnih mrežah so predstavljene informacije o proteinih in metabolitih. Informacije o metabolnih mrežah se nahajajo v podatkovnih zbirkah Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa in Goto, 2000), EcoCyc (Keseler in sod., 2011) in metaTIGER (Whitaker in sod., 2009).

2.5.4 Signalne mreže

Znotraj celic in med celicami se prenašajo signali, ki tvorijo kompleksno signalno mrežo. MAPK/ERK pot se prenaša s površja celice v jedro z zaporedjem proteinskih interakcij, fosforilacijskih reakcij in drugimi procesi. Signalne mreže integrirajo mreže proteinskih interakcij, mreže regulacije genov in metabolne mreže. Podatki o signalnih mrežah so zbrani v podatkovnih zbirkah, kot sta MiST (Ulrich in Zhulin, 2007) in TRANSPATH (Krull in sod., 2003).

2.5.5 Uporaba pristopov teorije grafov za analizo bioloških mrež

Današnje visoko zmogljive genomske tehnologije ustvarijo veliko količino podatkov. Razvoju eksperimentalnih tehnik sledi napredek računalniških pristopov za analizo genomskih podatkov. S kopičenjem razpoložljivih genomskih podatkov se povečuje kompleksnost bioloških mrež (Miller in sod., 2004). S kompleksnimi mrežami,

apliciranimi v biologiji in na drugih področjih, se ukvarja matematična disciplina teorija grafov. Tako kot na drugih področjih se je tudi v biologiji uveljavil pristop modeliranja pojavov z matematičnim konceptom, ki mu pravimo mreža (graf). Analiza kompleksnih mrež nam omogoča določati lastnosti, karakterizacijo in klasifikacijo kompleksnih struktur. Mreže so definirane kot grafi, neusmerjeni graf *G* (angl. *undirected graph*) je urejen par (*V, E*), ki ga označimo z *G* = (*V, E*). *V* je množica vozlišč (angl. *vertices*), *E* pa množica dvosmernih povezav (angl. *edges*). Če obstaja med vozliščema v grafu povezava, pravimo, da sta vozlišči soseda (angl. *neighbours*). Vozlišča v grafih je možno razvrščati glede na njihove značilnosti. Izbira metode razvrščanja vozlišč je odvisna od vprašanja na katerega poskušamo odgovoriti. V bioloških mrežah pogosto iščemo vozlišča, ki vplivajo na topologijo. Primer biološkega vprašanja pri raziskavi mreže je, katere molekule v biološki poti, ki niso nujno centralne, imajo ključno vlogo v mreži proteinskih interakcij (Pavlopoulos in sod., 2011). Za razvrščanje vozlišč glede na pomembnost njihove vloge v biološki mreži uporabljamo centralnostne indekse. Zaradi različnih interpretacij pomembnosti obstaja več definicij centralnosti. Stopnja centralnosti vozlišča (angl. *degree centrality*) je definirana kot število povezav z drugimi vozlišči. Vozlišča z visoko stopnjo centralnosti so povezovalniki (angl. *hubs*), ker so povezani z veliko sosedi. Odstranitev povezovalnika ima velik vpliv na topologijo grafa. Biološke mreže so odporne na naključne spremembe in občutljive na odpovedi povezovalnikov (Zotenko in sod., 2008; Levy in Siegal 2008). Vmesnostna centralnost (angl. *betweenes centrality*) predstavlja število najkrajših poti med katerimkoli parom vozlišč, ki potekajo skozi dano vozlišče. Joy in sod., ter Hahn in Kern so povezali esencialnost genov z visoko stopnjo vmesnosti (Joy in sod., 2005); (Hahn in Kern, 2005). Bližinska centralnost (angl. *closeness centrality*) predstavlja vsoto najkrajših poti od vozlišča do prav vseh ostalih vozlišč v grafu. S pomočjo mere bližinske centralnosti so identificirali najpomembnejše metabolite v metabolnih mrežah (Ma in Zeng, 2003). Eigenvektorska centralnost (angl. *eigenvector centrality*) temelji na predpostavki, da so pomembna vozlišča povezana s pomembnimi sosedi. S pomočjo mere za eigenvektorsko centralnost so identificirali vzročne gene za bolezni (Ozgür in sod., 2008) in povezovalnike v bioloških mrežah (Zotenko in sod., 2008).

2.5.6 Orodja za analizo bioloških mrež

Analiza bioloških mrež običajno poteka v treh osnovnih korakih (priprava podatkov, analiza in vizualizacija mreže), ki jih je možno izvesti s pomočjo bioinformacijskega orodja Cytoscape. Poleg Cytoscape poznamo tudi orodja NextBio, Ingenuity Pathway Analysis (IPA), GeneGo, Ondex, Osprey, visANT, NAViGaTOR. Aplikacija Cytoscape je prosto dostopna in je zasnovana tako, da omogoča dodajanje programskih vtičnikov, ki jih razvijajo raziskovalne skupine, in na ta način dodajajo osnovni aplikaciji vrsto novih funkcionalnosti. Aprila 2012 je bilo 152 registriranih vtičnikov za Cytoscape (Saito in sod., 2012). Vtičnik BisoGenet (Martin in sod., 2010) omogoča uvoz in integracijo podatkov iz več zbirk (DIP, BIND, HPRD, BioGRID, MINT in Intact) v Cytoscape. Vtičnik

CyTargetLinker omogoča integracijo obstoječe mreže z mrežami uravnavanja genov (DrugBank, ENCODE, MicroCosm, TFe, miRecords, miRTarBase). Cyto-Hubba (Lin in sod., 2008) je vtičnik za Cytoscape, ki omogoča iskanje pomembnih vozlišč v bioloških mrežah s pomočjo mer centralnosti, kot so stopnja vozlišča, ozko grlo, mera vmesnosti in koeficient grupiranja.

2.6 RAZVRŠČANJE KANDIDATNIH GENOV PO PRIORITETI

Razvrščanje kandidatnih genov po prioriteti (angl. *gene prioritisation*) je postopek ocenjevanja povezanosti kandidatnih genov z biološkim procesom, z namenom, da bi opravili nadaljnje analize z najbolje ocenjenimi geni. Z obema osnovnima pristopoma k iskanju vzročnih genov za bolezni, tako na ravni celotnega genoma kot na podlagi kandidatnih genov, dobimo množice genov ali kromosomske regije z možnim vplivom na fenotip. Proučevanje vseh kandidatnih genov je lahko zelo zamudno, zato so se pojavila različna bioinformacijska orodja za razvrščanje kandidatnih genov po prioriteti. Veliko izmed teh orodij temelji na do sedaj znanih podatkih o kandidatnih genih in povezovanju genov z biološkimi procesi; uporabljajo koncept krivde zaradi povezanosti (angl.*guilty by association*) (Tranchevent in sod., 2011).

Na spletnem portalu Gene Prioritization Portal (Tranchevent in sod., 2011) so zbrana orodja za razvrščanje genov po prioriteti. Nekatera izmed teh orodij so SUSPECTS (Adie in sod., 2006), Endeavour (S. Aerts in sod., 2009), CANDID (Hutz in sod., 2008), G2D (Gefen in sod., 2010), ToppGene Suite (Chen in sod., 2009), PosMed (Yoshida in sod., 2009), PolySearch (Cheng in sod., 2008), GeneProspector (Yu in sod., 2008), GenTrepid (George in sod., 2006), GeneWanderer (Kohler in sod., 2008), PhenoPred (Radivojac in sod., 2008).

2.7 BIOOZNAČEVALCI

Biooznačevalci so biološke značilnosti, ki jih je možno objektivno izmeriti in oceniti kot pokazatelje normalnih bioloških procesov, patogenih procesov ali farmakoloških odzivov na terapevtske intervencije (Naylor, 2003). Povezani so z vzrokom bolezni ali z določenimi lastnostmi. Uporabni so za diagnozo, prognozo bolezni in terapijo ter omogočajo boljši vpogled v patogenezo bolezni, iskanje novih terapevtskih tarč in strategij. Na kompleksne bolezni vplivajo spremembe več med seboj odvisnih genov ter vplivi okolja. Za identifikacijo biooznačevalcev je treba pravilno zaznati množico genov ali proteinov, ki spreminjajo biološki sistem in povzročajo bolezni. Iskanje biooznačevalcev kompleksnih bolezni je usmerjeno v integracijo podatkov različnih virov in pridobivanje kandidatnih biooznačevalcev na sistemski ravni (Azuaje, 2010).

Poznamo dva osnovna tipa biooznačevalcev; biooznačevalce ekspozicije, ki so uporabni za napovedovanje tveganja za obolevnost, ter biooznačevalce bolezni, ki so uporabni za diagnozo in spremljanje razvoja bolezni (Mayeux, 2004). Biooznačevalci so lahko

enostavne molekule (metaboliti, steroidi, lipidi), peptidi in proteini (inzulin, hemoglobin) ali celice (T celice) (Jain, 2010). Genetske biooznačevalce lahko razvijamo na ravni DNA, RNA, proteinov in peptidov ali epigenetike. Pri biooznačevalcih na ravni DNA iščemo variacije v zaporedju DNA, kot so delecije, insercije in polimorfizmi posameznih nukleotidov. Pri biooznačevalcih na ravni RNA iščemo spremembe vzorcev izražanja (angl. *transcriptional alterations*). Pri biooznačevalcih na ravni proteinov in peptidov lahko opazujemo količino ali funkcionalnost oziroma nefunkcionalnost proteinov. Pri epigenetskih biooznačevalcih opazujemo prisotnost ali izgubo metilne skupine na citozinih CpG dinukleotidov.

Bioinformatika ima ključno vlogo v razvoju biooznačevalcev. Biooznačevalce lahko iščemo na podlagi proteinskih interakcij. Chuang in sod. so za ločevanje metastatičnega od nemetastatičnega raka dojke primerjali stopnjo izražanja genov v celotnih podmrežah proteinskih interakcij in dosegali boljšo klasifikacijsko točnost kot s primerjavo stopnje izražanja posameznih genov (Chuang in sod., 2007). Študije so pokazale, da so biooznačevalci, ki temeljijo na bioloških poteh, zanesljivejši od tistih, ki temeljijo na posameznih genih (Kim in sod., 2012).

2.8 ASOCIACIJSKE ŠTUDIJE

Cilj genetskih asociacijskih študij je s testiranjem povezave med genotipom in fenotipom identificirati kandidatne gene ali genomske regije, ki vplivajo na določen fenotip (Lewis in Knight, 2012), in oceniti njihov prispevek k razvoju tega fenotipa (Ambrosius in sod., 2004). Asociacijske študije, ki temeljijo na razlikovanju genotipa glede na polimorfizem posameznega nukleotida, so zelo pogoste v raziskavah genetskih osnov kompleksnih bolezni (Collins in sod., 1997). Poznamo asociacijske študije na ravni celotnega genoma, kjer uporabimo SNP-je vzdolž celotnega genoma, ter asociacijske študije kandidatnih genov (angl. *candidate-gene association studies;* CGAS) vzdolž ali v bližini kandidatnih genov. Značilna genetska povezanost polimorfizma s fenotipom lahko pomeni 1) direktno povezanost, kjer je SNP vzrok za razliko v fenotipu, 2) posredno povezanost, kjer je SNP v vezavnem neravnotežju (angl. *linkage disequilibrium*; LD) z vzročno variacijo za fenotip in 3) lažno pozitiven rezultat (Lewis in Knight, 2012).

S pomočjo asociacijskih študij na celotnem genomu (GWAS) so v zadnjih letih identificirali na tisoče polimorfizmov, povezanih s kompleksnimi fenotipi. Pristop GWAS pomeni velik napredek na področju raziskav kompleksnih fenotipov, saj je razkril mnogo do sedaj nepoznanih kandidatnih genov. S pomočjo študij GWAS so identificirali preko 50 novih kandidatnih lokusov, povezanih z diabetesom tipa 2 (Mohlke in sod., 2008), preko 20 novih kandidatnih lokusov za nekatere vrste raka (Easton in Eeles, 2008), lokuse, povezane z avtoimunsko boleznijo (Lettre in Rioux, 2008) in druge.

Študije kandidatnih genov se začnejo z zbiranjem kandidatnih genov za proučevani fenotip, ki mu sledi izbor polimorfizmov z možnim vplivom na uravnavanje izražanja kandidatnih genov ali na njihove produkte (Collins in sod., 1997). Temu sledi testiranje povezave med polimorfizmom in fenotipom z opazovanjem prisotnosti različic polimorfizma pri različnih fenotipih. Pri asociacijskih študijah kandidatnih genov iščemo polimorfizme v kandidatnem genu in njegovi okolici. Povezanost fenotipa lahko testiramo z vsemi polimorfizmi ali samo z nekaterimi izmed njih (Wu in Cui, 2013). Polimorfizme za testiranje zbiramo lahko glede na to, kje se nahajajo v genu, ali glede na njihov tip (sinonimni, nesinonimni). Asociacijske študije kandidatnih genov so v primerjavi z asociacijskimi študijami na celotnem genomu bolj primerne, kadar so frekvence polimorfizmov nizke (Wilkening in sod., 2009), in kadar so na voljo podatki za majhno populacijo.

3 MATERIALI IN METODE

3.1 ZBIRANJE LOKUSOV IN IZDELAVA GENOMSKEGA ATLASA ZA NALAGANJE MAŠČOBE

3.1.1 Zbiranje lokusov, povezanih z nalaganjem maščobe

Lokuse za nalaganje maščobe smo iskali pri štirih vrstah (človek, govedo, podgana in miš) iz literature ter podatkovnih zbirk. Za iskanje literature, povezane z nalaganjem maščobe pri zgoraj omenjenih vrstah, smo uporabili bibliografsko zbirko PubMed. Ključne besede, ki smo jih uporabili pri iskanju, so: *gene*, *genetics*, *epigenetics*, *non-coding RNA*, *microRNA*, *obesity*, *adipose tissue*, *marbling*, *fat deposotion*, *adipogenesis*, *human*, *cattle*, *rat* in *mouse*. Za prenos povzetkov publikacij smo uporabili vmesnik Entrez Programming Utilities (E-utilities). Poizvedba, ki smo jo uporabili za zbiranje literature, je:

http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pubmed&mindate=2011/02&term=(obesity+OR+adipose+tissue+OR+(marbling+AND+meat)+OR+fat+deposotion+OR+adipogenesis)+AND+(gene+OR+genetics+OR+epigenetics+OR+non-coding+RNA+OR+microRNA)+AND+(human+OR+cattle+OR+rat+OR+mouse+OR+pig)&reldate=60&datetype=edat&retmax=1000000&usehistory=n

Podatkovne zbirke, v katerih smo iskali gene za nalaganje maščobe, so Obesity Gene Map, GeneCards, MGI (Mouse Genome Informatics) in RGD (Rat Genome Database). V RGD in AnimalQTLdb (Animal QTL Database) smo iskali QTL-e za lastnosti, povezane z nalaganjem maščobe. Uporabili smo ključne besede: *body fat*, *body weight*, *adiposity*, *obesity* in *diabetes*. V obeh zbirkah so ročno pregledani QTL-i iz literature in tisti, ki še niso bili objavljeni, vendar so jih skupinam RGD in AnimalQTLdb poslali raziskovalci. Poleg imena, opisa in lokacije QTL-a na genomu so na voljo še nekateri drugi podatki: vrednost LOD (angl. *logarithm of odds ratio*), vrednost p, varianca, kandidatni geni znotraj QTL-a in drugi (De la Cruz in sod., 2005; Hu in sod., 2007). Iskanje QTL-ov smo omejili na ime in opis lastnosti, ki se je morala ujemati z naštetimi ključnim besedami, povezanimi z nalaganjem maščobe.

3.1.2 Izdelava kataloga lokusov, povezanih z nalaganjem maščobe

Vso literaturo, povezano z nalaganjem maščobe pri človeku, govedu, podgani in miši, smo ročno preverili (angl. *manual curation*) in lokuse združili s tistimi, ki smo jih našli v podatkovnih zbirkah. Poimenovanje genov v katalogu lokusov, povezanih z nalaganjem maščobe, smo poenotili v skladu s priporočili skupine za nomenklaturo genov pri človeku HUGO Gene Nomenclature Committee (HGNC) (Gray in sod., 2013). Zbirka HGNC je prosto dostopna in vsebuje enolična imena ter simbole genov pri človeku. Nastala je z namenom poenotiti nomenklaturo genov. Vsi HGNC vnosi so ročno pregledani, imena in

simboli HGNC pa predstavljajo standard za poimenovanje genov v publikacijah in podatkovnih zbirkah. Nomenklaturo imen genov pri miših smo uskladili s terminologijo v zbirki MGI (http://www.informatics.jax.org/mgihome/nomen/). Za imena genov miRNA smo uporabili nomenklaturo miRBase. Za gene pri govedu, podgani in miši smo poiskali ortologne gene pri človeku s pomočjo zbirk MGI in Ensembl. Do podatkov smo dostopali s pomočjo orodja BioMart (http://www.ensembl.org/biomart/martview/). Lokuse za nalaganje maščobe smo vnesli v relacijsko podatkovno zbirko MySQL s pomočjo spletne aplikacije Xataface.

3.1.3 Spletno centralno mesto za lokuse, povezane z nabiranjem maščobe

Izdelali smo spletno centralno mesto za raziskave genskih mehanizmov nalaganja maščobe na spletnem naslovu http://integratomics-time.com/fat_deposition (Priloga C). Centralno mesto za nalaganje maščobe se nahaja v okviru projekta Integratomics TIME (http://integratomics-time.com/), namenjenega raziskavam kompleksnih bolezni s pomočjo integracije raznovrstnih genomskih podatkov in sistemske biologije. Za implementacijo spletne strani smo uporabili spletno tehnologijo PHP in relacijsko podatkovno zbirko MySQL.

3.2 INTEGRACIJA RAZNOVRSTNIH GENOMSKIH PODATKOV IN BIOINFORMACIJSKA ANALIZA

V genomskem atlasu sta dve vrsti lokusov (geni in QTL-i), ki smo jih zbrali iz različnih virov (podatkovne zbirke, literatura). Lokusi so bili povezani z nalaganjem maščobe v neodvisnih študijah in različnih eksperimentalnih pristopih pri štirih vrstah (človek, govedo, podgana in miš). Prikazali smo genomsko razporeditev QTL-ov in genov. Za gene pri govedu, podgani in miši smo poiskali ortologne gene pri človeku. S pomočjo prikaza genomske razporeditve lokusov smo integrirali genomske podatke iz različnih virov podatkov, iz različnih študij, pri različnih vrstah in identificirali prekrivanja med njimi. Na podlagi zbranih kandidatnih genov smo poiskali kandidatne biološke poti za nalaganje maščobe. Kandidatne gene, ki kodirajo proteine, smo povezali s prvimi sosedi v mreži proteinskih interakcij. Identificirali smo tudi mrežo uravnavanja izražanja genov za nalaganje maščobe z molekulami miRNA.

3.2.1 Prikaz genomske razporeditve lokusov, povezanih z nalaganjem maščobe

Za genomski prikaz porazdelitve lokusov na genomu smo uporabili orodje Flash GViewer (http://gmod.org/wiki/Flash_GViewer). Orodje omogoča vnos podatkov o številu in dolžini kromosomov ter lokacijah prog, zato je primerno za prikaz genomskih lokacij pri katerikoli vrsti. Flash GViewer omogoča animiran prikaz dveh vrst lokusov na genomu: 1) lokusi označeni s puščico, primerni za prikaz genov ali SNP-jev in 2) lokusi označeni z vzdolžnimi črtami, primerni za prikaz genomskih regij, kot so QTL-i ali kromosomske

mutacije. Lokuse je možno različno pobarvati, jih označiti z imeni ter jim dodati spletne naslove do podatkovnih zbirk.

3.2.2 Analiza obogatenosti bioloških poti

Analizo obogatenosti bioloških poti smo izvedli s pomočjo orodja DAVID Bioinformatics Resources 6.7 (http://david.abcc.ncifcrf.gov/) (Huang da in sod., 2007) (podobno kot v prilogi D). Orodje DAVID zajema podatke bioloških poti iz naslednjih zbirk: BBID (The biological biochemical image database) (Becker in sod., 2000), BioCarta (http://www.biocarta.com/), ExPASy Biochemical Pathways (Gasteiger in sod., 2003), KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa in sod., 2000), PANTHER (Mi in sod., 2013) in Reactome (Croft in sod., 2011) (Preglednica 2). Pri iskanju obogatenih bioloških poteh smo se omejili na tiste, ki imajo Bonferronijevo *p*-vrednost manjšo od 0,01. Za risanje Vennovih diagramov smo uporabili spletno aplikacijo Venn, ki je prosto dostopna na naslovu http://bioinformatics.psb.ugent.be/webtools/Venn/.

Preglednica 2: Zbirke bioloških poti, ki jih zajame orodje DAVID za analizo obogatenosti množice genov

Table 2: Biological pathways databases included in the tool DAVID for gene-set enrichment analysis

Zbirka bioloških poti	**Število genov v bioloških poteh, ki so zajeti v analizo orodja DAVID**
BBID	140
BioCarta	752
ExPASy	4307
KEGG	9013
PANTHER	3278
Reactome	613

3.2.3 Integracija in analiza bioloških mrež

Za analizo bioloških mrež smo uporabili orodje Cytoscape (Smoot in sod., 2011) in vtičnike BisoGenet (Martin in sod., 2010), CyTargetLinker in Cyto-Hubba (Lin in sod., 2008). Mrežo proteinskih interakcij genov, povezanih z nalaganjem maščobe, smo izdelali s pomočjo orodja BisoGenet, ki omogoča uvoz podatkov iz zbirk proteinskih interakcij DIP (Xenarios in sod., 2002), BioGRID (Chatr-Aryamontri in sod., 2013), HPRD (Goel in sod., 2012), BIND (Isserlin in sod., 2011), MINT (Licata in sod., 2012) in IntAct (Kerrien in sod., 2012). S seznama genov, ki so povezani z nalaganjem maščobe, smo uporabili samo tiste, ki kodirajo proteine. Povezali smo jih z njihovimi prvimi sosedi (proteini, ki so z njimi v interakcijah) in dobili mrežo proteinskih interakcij z 8046 vozlišči (proteini) in 92638 povezavami (interakcije). Proteine v mreži smo poimenovali z imeni genov, ki jih kodirajo. Za analizo topoloških značilnosti mreže proteinskih interakcij smo uporabili

vtičnik za Cytoscape Cyto-Hubba, s pomočjo katerega smo izračunali mere centralnosti za vozlišča. Za analizo mreže uravnavanja genov smo uporabili orodje CyTargetLinker. Gene, povezane z nalaganjem maščobe, smo integrirali s podatki o regulatornih interakcijah. Interakcije smo pridobili iz zbirke miRecords. Geni in povezave med njimi predstavljajo mrežo uravnavanja genov, povezanih z nalaganjem maščobe.

3.2.4 Bioinformacijska analiza in razvrščanje kandidatnih genov po prioriteti

Razvili smo orodje miRNA SNiPer za iskanje polimorfizmov znotraj genov miRNA (Zorc in sod., 2012) (Priloga A, Priloga B, Priloga G). Spletna aplikacija miRNA SNiPer je dostopna na naslovu http://www.integratomics-time.com/miRNA-SNiPer. Uporabniški vmesnik omogoča vnos seznama genov miRNA in vrača tabelo polimorfizmov znotraj vnesenih genov. Polimorfizmi so razvrščeni glede na različne regije znotraj genov za miRNA: 1) pre-miRNA, 2.) zrela miRNA, 3) regija *seed*, ki je odgovorna za vezavo na tarčo. Orodje vsebuje podatke iz različnih virov: 1) nukleotidna zaporedja genov za miRNA, genomske lokacije in nomenklaturo iz zbirke miRBase (različica 19) (http://www.mirbase.org/) (Kozomara in sod., 2011), 2) lokacije regij *seed* v genih za miRNA iz zbirke TargetScan (različica 6.2) (http://www.targetscan.org/) (Lewis in sod., 2005) in 3) lokacije polimorfizmov iz zbirke Ensembl Variation database (http://www.ensembl.org/) (Flicek in sod., 2013). Podatke iz zbirk miRBase, TargetScan in Ensembl Variation database smo prenesli v relacijsko podatkovno zbirko. S pomočjo skripte, napisane v skriptnem jeziku Perl, smo izvedli poizvedbe po relacijski zbirki in poiskali polimorfizme znotraj različnih regij genov za miRNA.

Za prikaz genomske porazdelitve genov za miRNA, njihovih gostiteljskih genov (angl. *host genes*), QTL-ov in polimorfizmov pri človeku in miši smo razvili orodje miRNA Viewer, ki je dostopno na naslovu http://www.integratomics-time.com/miRNA-genomic-viewer/. Orodje integrira podatke iz zbirk miRBase, Ensembl, RGD (Rat Genome Database) in OMIM.

3.3 VPLIV KANDIDATNIH GENOV NA LATNOSTI NALAGANJA MAŠČOBE

Preverili smo vpliv polimorfizmov v nekaterih izmed kandidatnih genov (*Akt1*, *Ubc*, *Grb2*, *Mir599*) na nalaganje maščobe pri miši (podobno kot v prilogi A). Podatke za analizo smo dobili v podatkovni zbirki Mouse Phenome Database (MPD; http://phenome.jax.org/) (Maddatu in sod., 2012). Fenotipski podatki iz zbirke MPD zajemajo meritve za 3379 lastnosti pri inbridiranih linijah miši. Lastnosti so razdeljene v 32 skupin (izgled in barva kožuha, obnašanje, klinična kemija, hematologija, lipidi, ksenobiotiki, sestava telesa, telesne maščobne blazinice, telesna teža, velikost in rast, kosti, možgani, rak, kardiovaskularni sistem, poškodbe tkiv in celic, razvoj, uho, endokrini sistem, oko, žolčnik, imunski sistem, prednostno zauživanje, ledvica, jetra, dolgoživost, metabolizem, mišice, živčni sistem, nevrosenzorika, reprodukcija, dihalni sistem, spalni vzorci in

vranica). Analizo smo opravili za lastnosti, ki so povezane z nalaganjem maščobe (sestava telesa, telesne maščobne blazinice in telesna teža). Število linij, vključenih v analizo, je odvisno od dostopnosti genotipskih podatkov (Preglednica 3). Za statistično analizo smo uporabili linearni model (1), kjer y_{ijkl} predstavlja opazovano lastnost, μ povprečje, G_i sistematski vpliv SNP-ja, L_{ij} ugnezden vpliv linije znotraj SNP-ja, e_{ijkl} pa ostanek. Za statistično analizo smo uporabili programski paket SAS/STAT.

$$y_{ijk} = \mu + G_i + L_{ij} + e_{ijk} \quad (1)$$

Preglednica 3: Genotipski podatki iz zbirke MPD za polimorfizme rs261613149, rs37362582, rs13461180, rs13471888 in rs37362582 pri 24 inbridiranih linijah miši

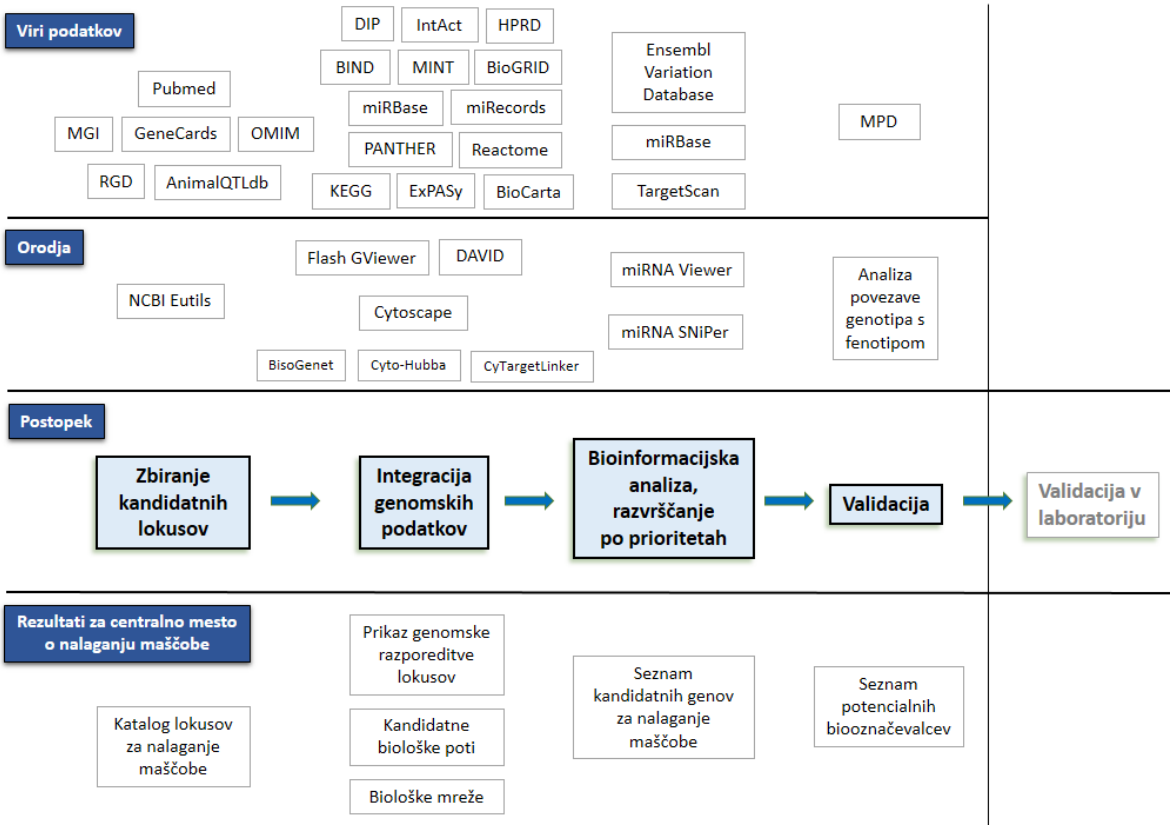
Table 3: Genotipic data from the MPD database for polymorphisms rs261613149, rs37362582, rs13461180, rs13471888 and rs37362582 in 24 inbred strains of mice

Gen	***Akt1***	***Mir599***	***Ubc***	***Grb2***	***Mapkap1***
SNP	rs261613149	rs37362582	rs13461180	rs13471888	rs37362582
Zamenjava	**A>C**	**A>T**	**C>T**	**A>G**	**A>T**
C57BL/6J	C	A	G	A	A
129S1/SvImJ	C	A	A	G	A
129X1/SvJ	C	-	A	G	-
A/J	A	A	A	G	A
AKR/J	A	A	A	G	A
BALB/cByJ	C	A	A	A	A
BALB/cJ	C	A	A	A	A
BTBR_T+_tf/J	C	A	G	A	A
C3H/HeJ	A	A	A	G	A
C57BL/6NJ	C	A	G	A	A
CAST/EiJ	C	A	A	G	A
CBA/J	A	A	A	G	A
DBA/2J	C	A	A	A	A
FVB/NJ	A	A	G	G	A
KK/HlJ	C	A	-	G	A
LP/J	C	A	A	G	A
MOLF/EiJ	-	A	-	-	-
NOD/ShiLtJ	C	A	A	G	A
NZO/HlLtJ	C	A	A	G	A
NZW/LacJ	C	A	G	G	-
PWD/PhJ	-	T	-	-	T
PWK/PhJ	C	T	A	G	T
SPRET/EiJ	A	A	A	G	A
WSB/EiJ	A	A	G	G	A

(-) ni podatka

4 REZULTATI

Za izdelavo kataloga smo uporabili primerjalni in integrativni pristop zbiranja do sedaj objavljene literature o nalaganju maščobe. S pomočjo bioinformacijskih orodij smo integrirali znane podatke o nalaganju maščobe ter izdelali zbirko genetskih lokusov. Izdelali smo primerjalne genomske prikaze in genske mreže kandidatnih genov ter identificirali kandidatne biološke poti. Iz nabora kandidatnih genov smo izbrali najbolj obetavne, ki smo jih določili na podlagi prekrivanj neodvisnih študij, kandidatnih bioloških poti, genskih mrež in z uporabo novih bioinformacijskih orodij miRNA SNiPer ter miRNA Viewer. Razvili smo strategijo raziskave kompleksne lastnosti (Slika 2), ki jo je možno uporabiti tudi za ostale kompleksne fenotipe.

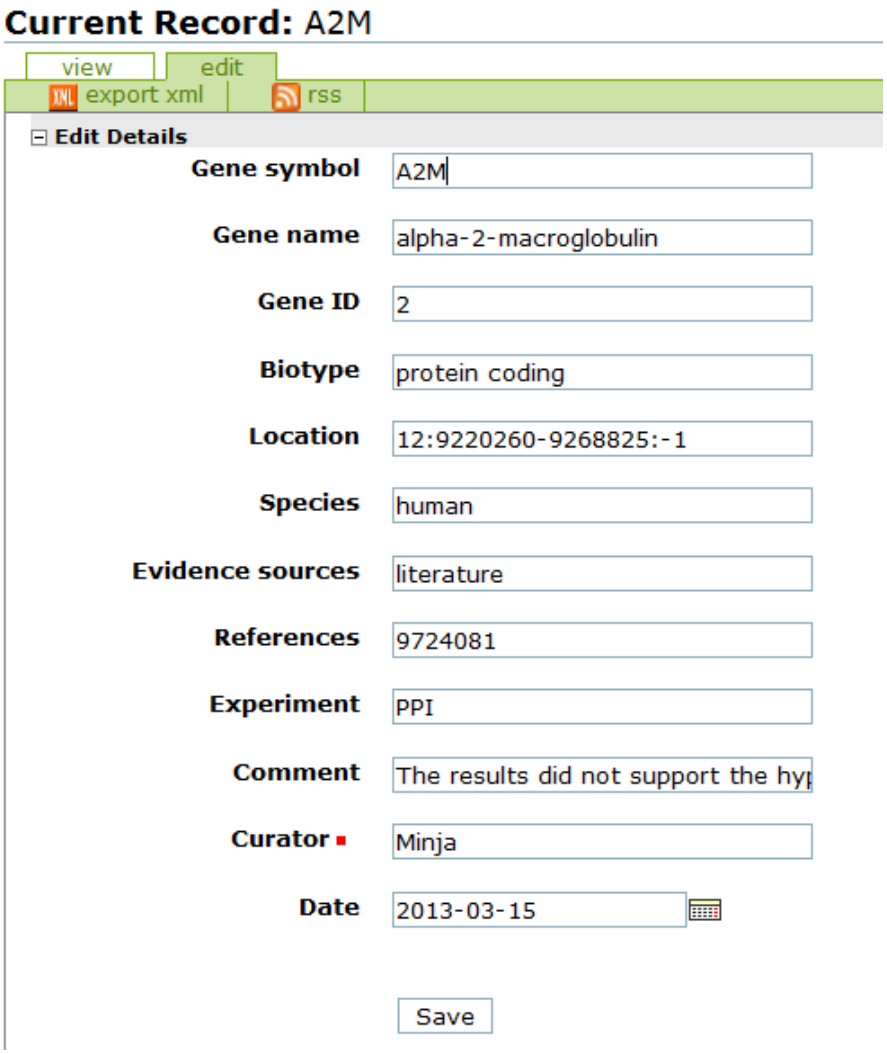


Slika 2: Shema poteka dela

Figure 2: Workflow of the study

4.1 GENOMSKI ATLAS ZA RAZISKAVE LOKUSOV, POVEZANIH Z NALAGANJEM MAŠČOBE

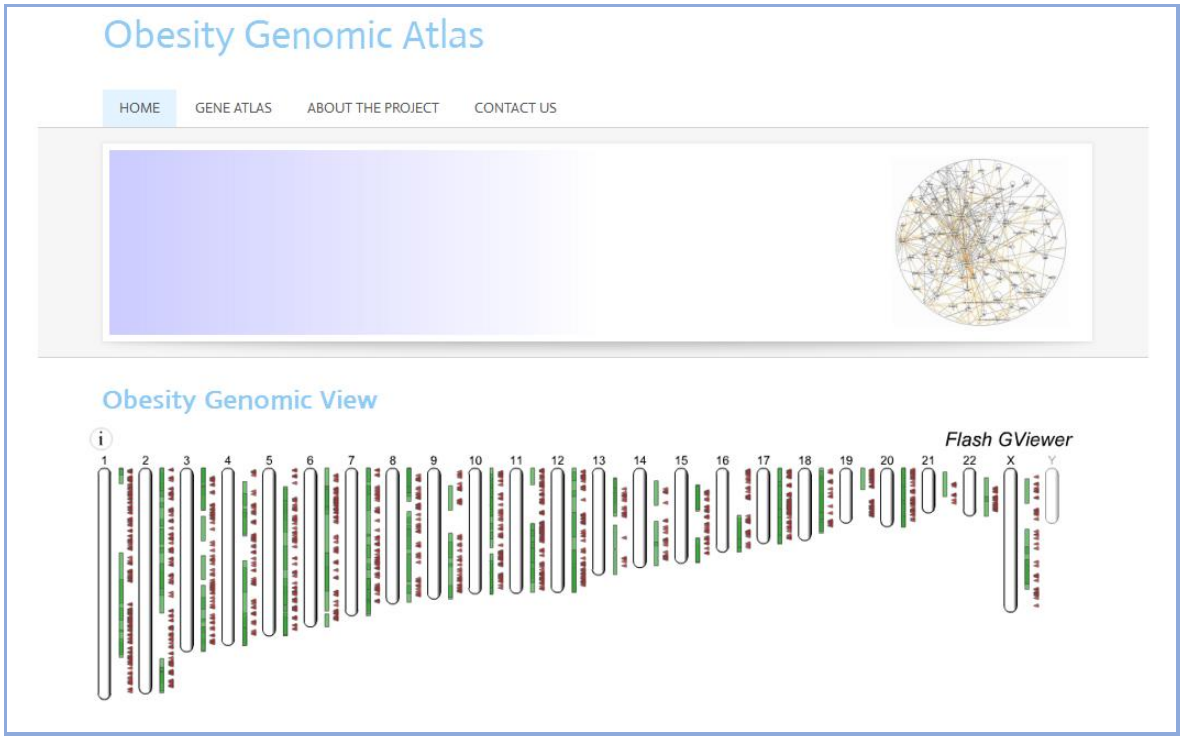
Zbrali smo 1774 genov, povezanih z nalaganjem maščobe, od tega je 1553 kodirajočih genov in 221 genov za miRNA. Poiskali smo tudi 222 QTL-ov za nalaganje maščobe pri človeku in 38 pri miši. Podatke o lokusih smo shranjevali v relacijsko podatkovno zbirko MySQL s pomočjo aplikacije Xataface (Slika 3).



Slika 3: Obrazec za vnos podatkov o lokusu za nalaganje maščobe v zbirko MySQL s pomočjo aplikacije Xataface

Figure 3: Data entry form for the fat deposition related loci in MySQL database using Xataface application

Zbirko lokusov smo objavili na spletni strani Obesity Genomic Atlas, ki se nahaja na spletnem naslovu http://www.integratomics-time.com/fat_deposition (Slika 4). Stran predstavlja centralno spletno mesto za raziskave, povezane z nalaganjem maščobe (Priloga C).



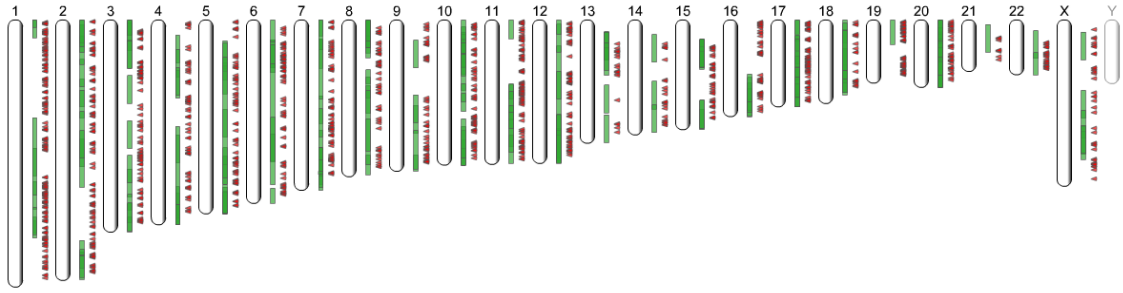
Slika 4: Spletno mesto Obesity Genomic Atlas

Figure 4: The website for Obesity Genomic Atlas

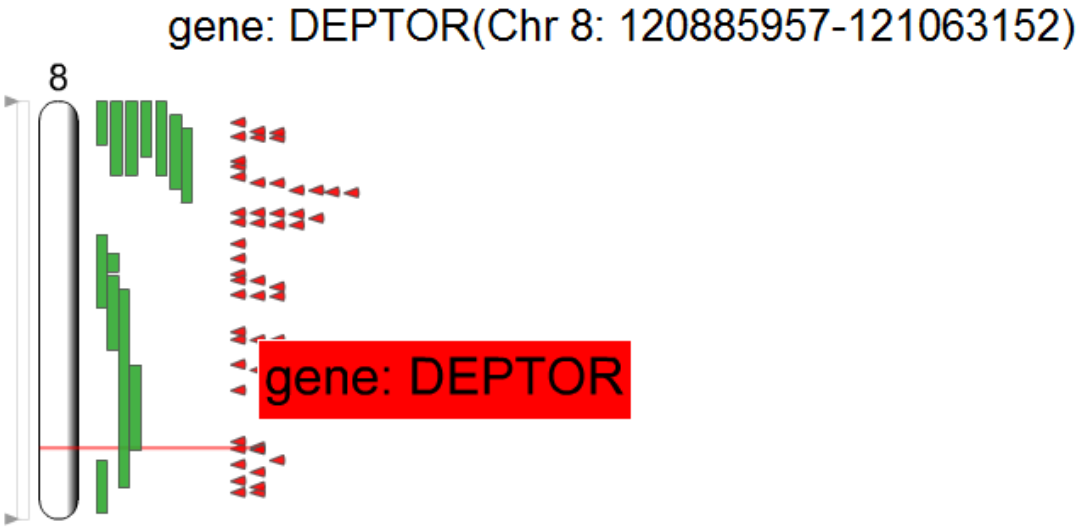
4.2 INTEGRACIJA RAZNOVRSTNIH GENOMSKIH PODATKOV V POVEZAVI Z NALAGANJEM MAŠČOBE

4.2.1 Genomski prikaz razporeditve lokusov

Izdelali smo interaktivne prikaze genomskih razporeditev lokusov, povezanih z nalaganjem maščobe pri človeku (Slika 5), miši (Slika 6) in govedu. Na ta način smo integrirali rezultate različnih študij, zbranih iz več virov (literature in podatkovnih zbirk) in jih predstavili na enem mestu ter omogočili prikaz prekrivanj lokusov.



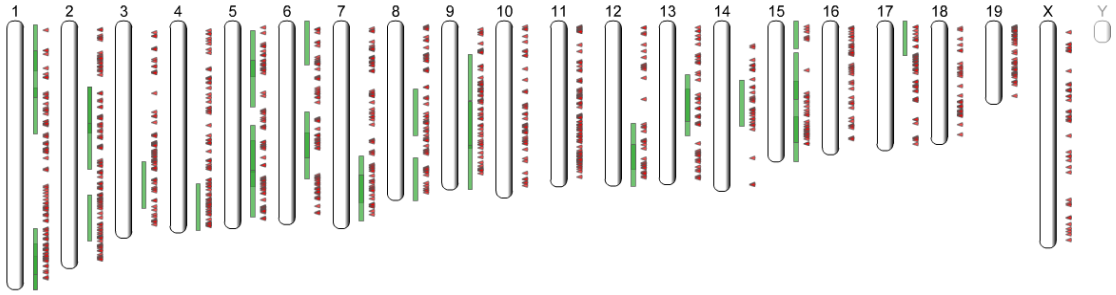
A



B

Slika 5: Genomski prikaz A) Razporeditev genomskih lokusov za nalaganje maščobe pri človeku. B) Kromosom 8 pri človeku z lokusi za nalaganje maščobe. Prikazana je lokacija gena *DEPTOR*, ki se prekriva z dvema QTL-oma.

Figure 5: Genomic view A) Genomic locations of fat deposition associated loci in human. B) Chromosome 8 in human with fat deposition associated loci. The location of gene *DEPTOR*, which overlaps with two QTL is shown.



Slika 6: Razporeditev genomskih lokusov za nalaganje maščobe pri miši

Figure 6: Genomic locations of loci, associated with fat deposition in mouse

4.2.2 Analiza obogatenosti bioloških poti

Analizo obogatenosti bioloških poti smo izvedli z orodjem DAVID (Priloga D). Identificirali smo 139 bioloških poti, v katerih so udeleženi geni, ki so povezani z nalaganjem maščobe (Preglednica 4).

Preglednica 4: Biološke poti, v katerih so udeleženi kandidatni geni za nalaganje maščobe

Table 4: Biological pathways, containing fat deposition candidate genes

Zbirka bioloških poti	**Biološka pot**	**Bonferroni vrednost p**	**Kandidatni geni, vpleteni v biološko pot**
KEGG	*Adipocytokine signaling pathway*	$9.6*10^{-29}$	*CD36, JAK2, TRAF2, TRADD, ACACB, ACSL1, ADIPOR1, ADIPOR2, ADIPOQ, AGRP, CAMKK2, CPT1A, CHKB, CPT1B, CHUK, G6PC, G6PC2, IKBKB, IKBKG, IRS1, IRS2, LEP, LEPR, MTOR, MAPK8, NPY, NFKB1, NFKBIA, NFKBIB, PPARA, PPARGC1A, PCK1, PCK2, POMC, PRKCQ, PRKAA1, PRKAA2, PRKAB1, PRKAG1, PRKAG2, PRKAG3, PTPN11, RXRA, RXRG, STK11, STAT3, SLC2A1, SLC2A4, SOCS3, TNF, TNFRSF1A, TNFRSF1B, AKT1, AKT2, AKT3, RELA*
KEGG	*Insulin signaling pathway*	$2.8*10^{-21}$	*PDPK1, BAD, CBL, SHC1, ACACA, ACACB, EIF4EBP1, FASN, FOXO1, FBP2, GCK, G6PC, G6PC2, GYS1, GSK3B, GRB2, HK1, HK2, MAPK3, IKBKB, INPP5K, INPP5D, INSR, IRS1, IRS2, INS, IGF2, LIPE, MTOR, MAPK1, MAPK8, MAP2K1, PPARGC1A, PDE3A, PDE3B, PCK1, PCK2, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIK3R2, PIK3R3, PIK3R5, PRKCZ, PRKAA1, PRKAA2, PRKAB1, PRKAG1, PRKAG2, PRKAG3, PRKACA, PRKACB, PRKACG, PRKAR1A, PRKAR1B, PRKAR2A, PRKAR2B, PPP1R3A, PPP1R3C, PTPN1, PTPRF, PKLR, RPS6KB1, RPS6, SLC2A4, SORBS1, SREBF1, SOCS1, SOCS3, AKT1, AKT2, AKT3*

se nadaljuje

nadaljevanje

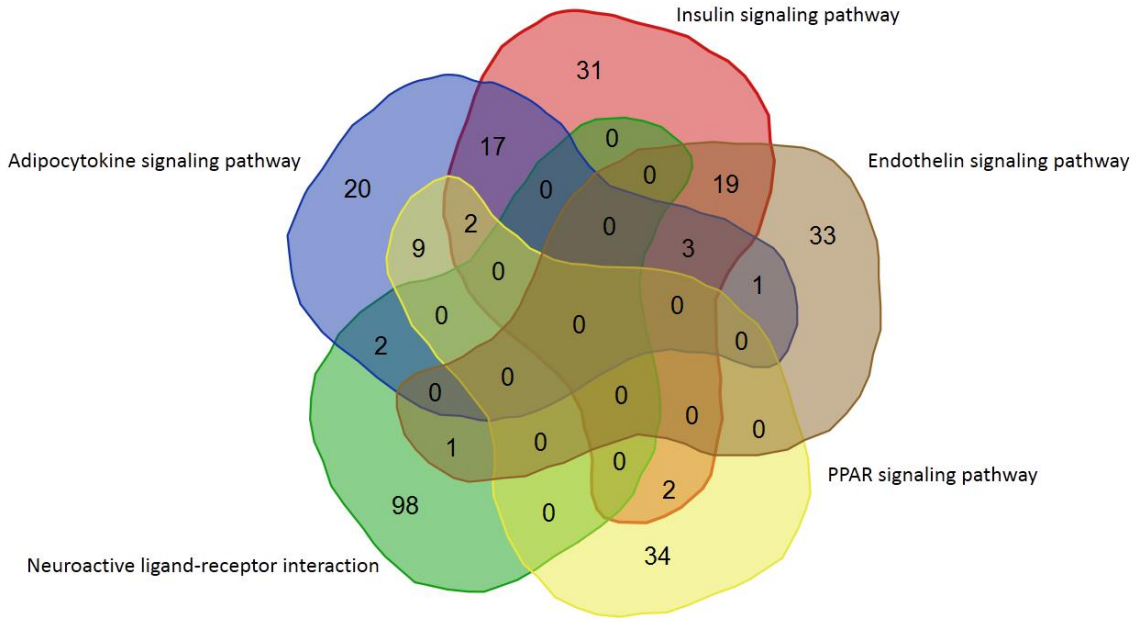
Zbirka bioloških poti	**Biološka pot**	**Bonferroni vrednost p**	**Kandidatni geni, vpleteni v biološko pot**
PANTHER	*Endothelin signaling pathway*	$5.1*10^{-19}$	*GNAS, ADCY1, ADCY10, ADCY2, ADCY3, ADCY4, ADCY5, ADCY6, ADCY7, ADCY8, ADCY9, EDN1, EDNRA, GNAQ, MAPK3, ITPR3, MAPK1, MAP2K1, NPR1, NOS1, NOS2, NOS3, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3C2A, PIK3C2B, PIK3C2G, PIK3C3, PIK3R1, PIK3R2, PIK3R3, PIK3R5, PLA2G4A, PLCB1, PLCB2, PLCB3, PLCB4, PTGS2, PRKCA, PRKCB, PRKCD, PRKCE, PRKCQ, PRKCZ, PRKACA, PRKACB, PRKACG, PRKAR1A, PRKAR1B, PRKAR2A, PRKAR2B, PRKG1, AKT1, AKT2, AKT3*
KEGG	*PPAR signaling pathway*	$6.8*10^{-18}$	*PDPK1, CD36, ACSL1, ACADM, ADIPOQ, ANGPTL4, APOA1, APOA2, APOA5, APOC3, AQP7, CPT1A, CPT2, CHKB, CPT1B, CYP27A1, CYP7A1, CYP8B1, FABP1, FABP2, FABP3, FABP4, FABP5, FABP6, FADS2, GK, LPL, NR1H3, OLR1, PLIN1, PPARA, PPARD, PPARG, PCK1, PCK2, PLTP, RXRA, RXRG, SLC27A1, SLC27A2, SLC27A4, SLC27A6, SORBS1, SCD, SCD5, SCP2, UCP1*
KEGG	*Neuroactive ligand-receptor interaction*	$2.7*10^{-17}$	*HTR1A, HTR1B, HTR2A, HTR2C, HTR5A, HTR6, GPR35, GPR50, ADORA1, ADRA1B, ADRA2A, ADRA2B, ADRB1, ADRB2, ADRB3, AGTR1, AGTR2, APLNR, BRS3, BDKRB1, BDKRB2, CNR1, CNR2, CCKAR, CCKBR, CHRM3, F2, F2R, CRHR1, CRHR2, CYSLTR2, DRD2, DRD3, DRD4, EDNRA, FSHR, GALR1, GALR2, GABRA2, GABRA6, GABRG3, GABBR1, GIPR, GRPR, GCGR, GLP1R, GRIN1, GRM1, GRM5, GRM8, GLRA1, CGA, GH1, GHR, GHRHR, GHSR, HRH1, HRH3, HCRTR1, HCRTR2, LEP, LEPR, LHB, LPAR1, LPAR3, MCHR1, MCHR2, MC2R, MC3R, MC4R, MC5R, MTNR1B, NMBR, NMUR1, NMUR2, NPFFR2, NPY1R, NPY2R, NPY5R, NPBWR1, NTSR1, NR3C1, OPRD1, OPRM1, PPYR1, PRL, PRLHR, PTGIR, PRSS3, SSTR2, SSTR5, S1PR1, S1PR5, TBXA2R, THRA, THRB, TSHR, TRPV1, MC1R, UTS2R, VIPR2*

se nadaljuje

nadaljevanje

Zbirka bioloških poti	**Biološka pot**	**Bonferroni vrednost p**	**Kandidatni geni, vpleteni v biološko pot**
Reactome	*Metabolism of lipids and lipoproteins*	$3.9*10^{-17}$	*AGPAT2, DECR1, DHCR24, HMGCR, ACLY, ABCA1, ABCB11, ABCC3, ABCG1, ABCG5, ABCG8, AMACR, NPC1L1, ABHD5, ACACA, ACACB, ACADS, ACADM, ACADVL, ALB, APOA1, APOA2, APOA4, APOB, APOC2, APOC3, CEL, CRAT, CPT2, CETP, CHKB, CPT1B, CLPS, CYP11A1, CYP11B2, CYP19A1, CYP21A2, CYP27A1, CYP7A1, CYP8B1, DGAT1, FDFT1, FABP4, FABP6, FASN, GK, GPAM, GPD1, GPD2, HSD3B1, HADH, HADHA, HSD11B1, APOE, IDH1, LBR, LCAT, LIPE, LPL, LDLR, MTTP, MGLL, PNLIP, PLIN1, PRKACA, PRKACB, PRKACG, SCARB1, SLC10A2, SLC27A2, SQLE, STAR, SCP2, SDC1*
KEGG	*Cytokine-cytokine receptor interaction*	$4.9*10^{-16}$	*CD40LG, FAS, KITLG, TNFSF12, AMH, BMP2, BMP7, BMPR1A, BMPR2, FIGF, CTF1, CLCF1, XCL1, CCL11, CCL15, CCL2, CCL20, CCL25, CCL28, CCL3, CCL4, CCL5, CCR1, CCR2, CCR3, CCR5, CXCL10, CXCL14, CXCL2, CXCL5, CX3CR1, CNTFR, CSF1, CSF2, CSF3, EGFR, EPO, FLT1, FLT3, FLT3LG, GH1, GHR, HGF, INHBA, INHBB, IFNA1, IFNG, IL1R1, IL1A, IL1B, IL10, IL10RA, IL11, IL13, IL15, IL15RA, IL17A, IL17B, IL18, IL18R1, IL2, IL2RA, IL2RB, IL2RG, IL20, IL22, IL23A, IL3, IL4, IL5, IL6, IL6R, IL6ST, IL7, IL8, CXCR1, KDR, LEP, LEPR, LIF, LIFR, LTA, LTB, OSM, PF4, PRL, TSLP, TGFB1, TNF, TNFSF11, TNFSF14, TNFRSF11A, TNFRSF11B, TNFRSF12A, TNFRSF14, TNFRSF1A, TNFRSF1B, VEGFA, VEGFC, CNTF*
KEGG	*Chemokine signaling pathway*	$4.1*10^{-14}$	*JAK2, SHC1, ADCY1, ADCY2, ADCY3, ADCY4, ADCY5, ADCY6, ADCY7, ADCY8, ADCY9, ADRBK1, XCL1, CCL11, CCL15, CCL2, CCL20, CCL25, CCL28, CCL3, CCL4, CCL5, CCR1, CCR2, CCR3, CCR5, CXCL10, CXCL14, CXCL2, CXCL5, CX3CR1, CHUK, GSK3A, GSK3B, GRB2, GNAI1, GNB3, GNG3, MAPK3, IKBKB, IKBKG, IL8, CXCR1, MAPK1, MAP2K1, NFKB1, NFKBIA, NFKBIB, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIK3R2, PIK3R3, PIK3R5, PLCB1, PLCB2, PLCB3, PLCB4, PF4, PRKCB, PRKCD, PRKCZ, PRKACA, PRKACB, PRKACG, RHOA, RAC1, RAC2, STAT1, STAT3, STAT5B, AKT1, AKT2, AKT3, RELA*

Preverili smo, kateri geni iz kataloga kandidatnih lokusov za nalaganje maščobe se nahajajo v najbolj obogatenih bioloških poteh in kateri izmed njih se nahajajo v več bioloških poteh hkrati (Preglednica 5). S pomočjo Vennovega diagrama smo prikazali število skupnih kandidatnih genov v petih najbolj obogatenih kandidatnih poteh (Slika 7).



Slika 7: Vennov diagram, ki prikazuje število kandidatnih genov iz Genomskega atlasa za nalaganje maščobe, ki so udeleženi v bioloških poteh *Adipocytokine signaling pathway*, *Insulin signaling pathway*, *Endothelin signaling pathway*, *PPAR signaling pathway* in *Neuroactive ligand-receptor interaction*

Figure 7: Venn diagram showing the numbers of fat deposition candidate genes from Obesity Genomic Atlas, involved in biological pathways *Adipocytokine signaling pathway*, *Insulin signaling pathway*, *Endothelin signaling pathway*, *PPAR signaling pathway* and *Neuroactive ligand-receptor interaction*

Preglednica 5: Biološke poti, ki imajo skupne kandidatne gene iz Genomskega atlasa za nalaganje maščobe

Table 5: Biological pathways that share common candidate genes from Obesity Genomic Atlas

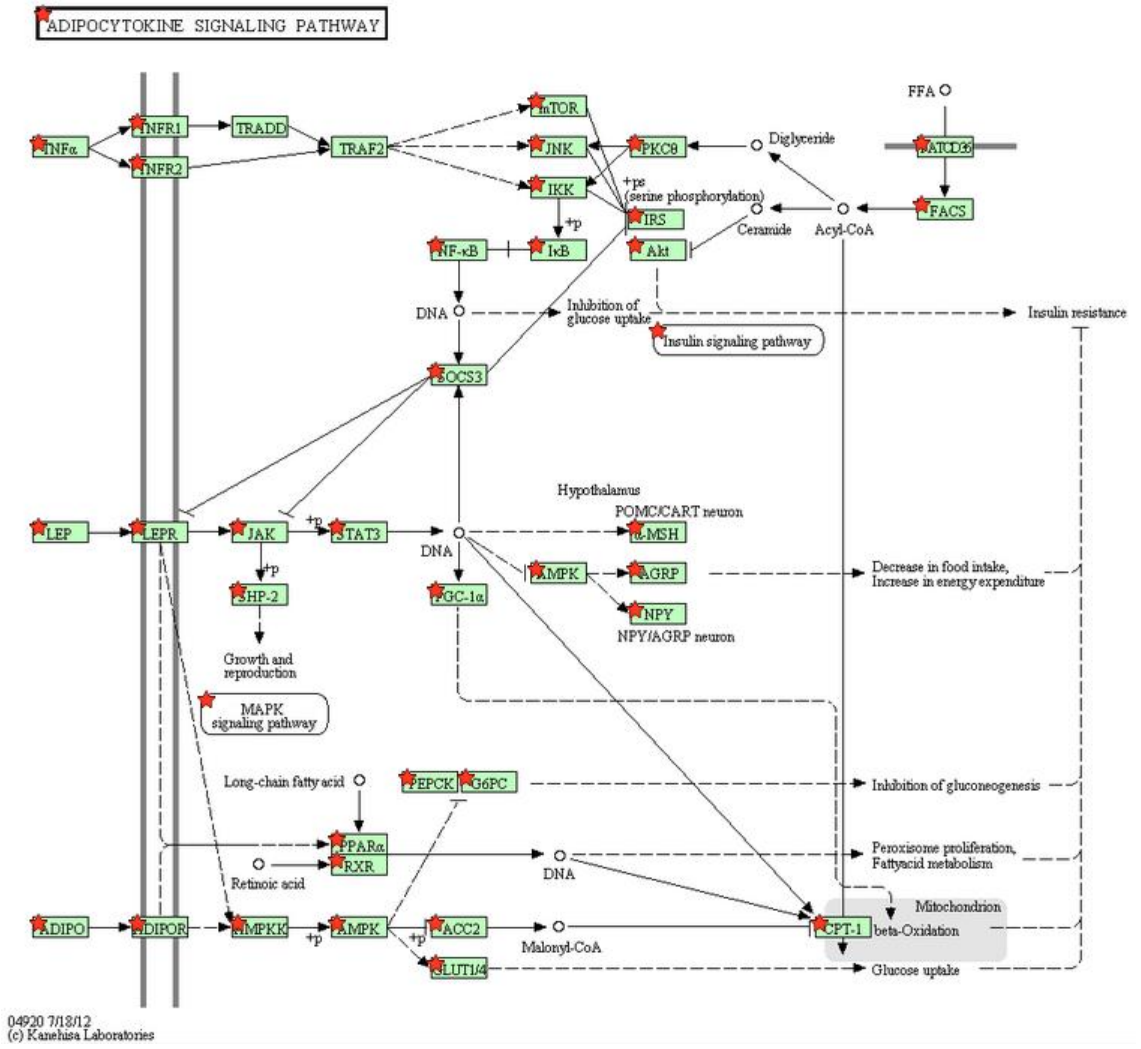
Biološke poti	**Število skupnih genov**	**Imena skupnih genov**	**Ortologni geni pri miši**
Adipocytokine signaling pathway, Insulin signaling pathway, Endothelin signaling pathway	3	*AKT1, AKT2, AKT3*	*Akt1, Akt2, Akt3*
Adipocytokine signaling pathway, Insulin signaling pathway, PPAR signaling pathway	2	*PCK1, PCK2*	*Pck1, Pck2*
Adipocytokine signaling pathway, Insulin signaling pathway	17	*MTOR, PPARGC1A, PRKAA2, G6PC2, ACACB, PRKAB1, IKBKB, G6PC, IRS2, PRKAG3, MAPK8, IRS1, PRKAG2, SOCS3, SLC2A4, PRKAA1, PRKAG1*	*Mtor, Ppargc1a, Prkaa2, G6pc2, Acacb, Prkab1, Ikbkb, G6pc, Irs2, Prkag3, Mapk8, Irs1, Prkag2, Socs3, Slc2a4, Prkaa1, Prkag1*
Adipocytokine signaling pathway, Endothelin signaling pathway	1	*PRKCQ*	*Prkcq*
Adipocytokine signaling pathway PPAR signaling pathway	9	*RXRG, PPARA, RXRA, CPT1B, CPT1A, CD36, ACSL1, CHKB, ADIPOQ*	*Rxrg, Ppara, Rxra, Cpt1b, Cpt1a, Cd36, Acsl1, Chkb, Adipoq*
Adipocytokine signaling pathway, Neuroactive ligand-receptor interaction	2	*LEPR, LEP*	*Lepr, Lep*

se nadaljuje

nadaljevanje

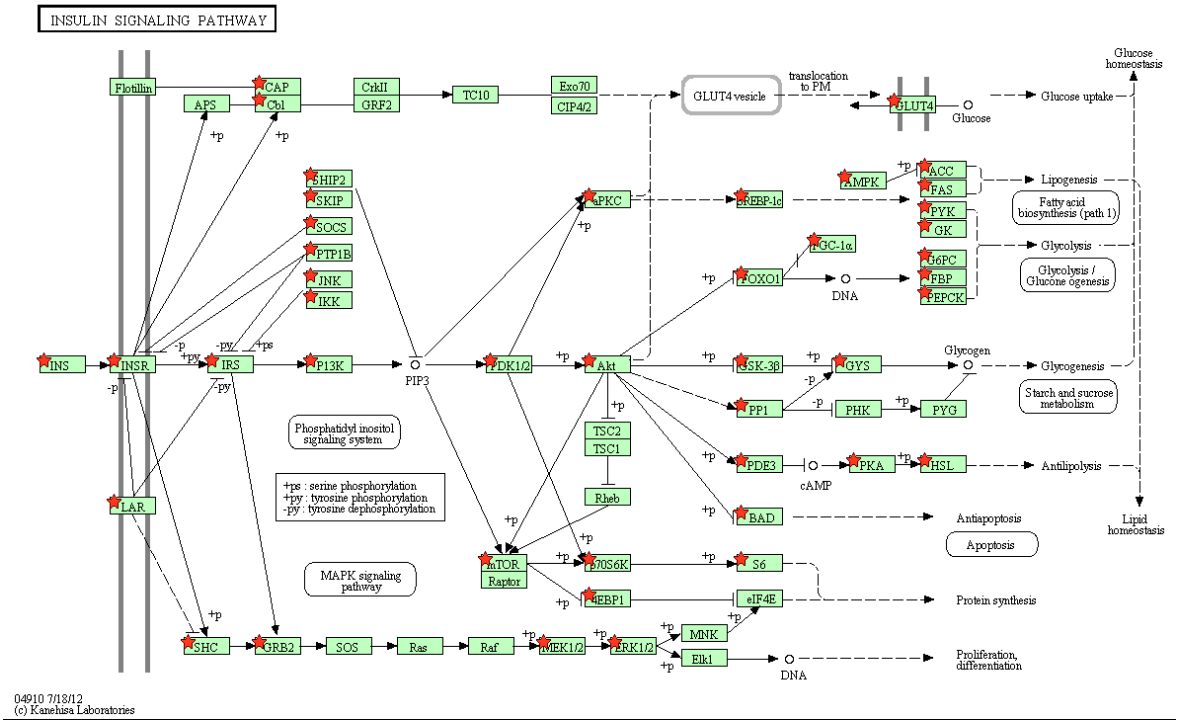
Biološke poti	**Število skupnih genov**	**Imena skupnih genov**	**Ortologni geni pri miši**
Insulin signaling pathway, Endothelin signaling pathway	19	*PRKAR1B, PRKAR2B, PRKCZ, PRKAR1A, PIK3R1, PIK3R5, MAPK1, PRKACA, PIK3CG, MAP2K1, PRKAR2A, PIK3CA, PRKACB, PIK3CB, PIK3CD, PIK3R2, MAPK3, PRKACG, PIK3R3*	*Prkar1b, Prkar2b, Prkcz, Prkar1a, Pik3r1, Pik3r5, Mapk1, Prkaca, Pik3cg, Map2k1, Prkar2a, Pik3ca, Prkacb, Pik3cb, Pik3cd, Pik3r2, Mapk3, Pik3r3*
Insulin signaling pathway, PPAR signaling pathway	2	*PDPK1, SORBS1*	*Pdpk1, Sorbs1*
Endothelin signaling pathway, Neuroactive ligand-receptor interaction	1	*EDNRA*	*Ednra*

V bioloških poteh *Adipocytokine signaling pathway* (Slika 8), *Insulin signaling pathway* (Slika 9) in *Endothelin signaling pathway* (Slika 10) so udeleženi geni, ki so povezani z nalaganjem maščobe. Z zvezdicami so označeni kandidatni geni iz Genomskega atlasa za nalaganje maščobe.



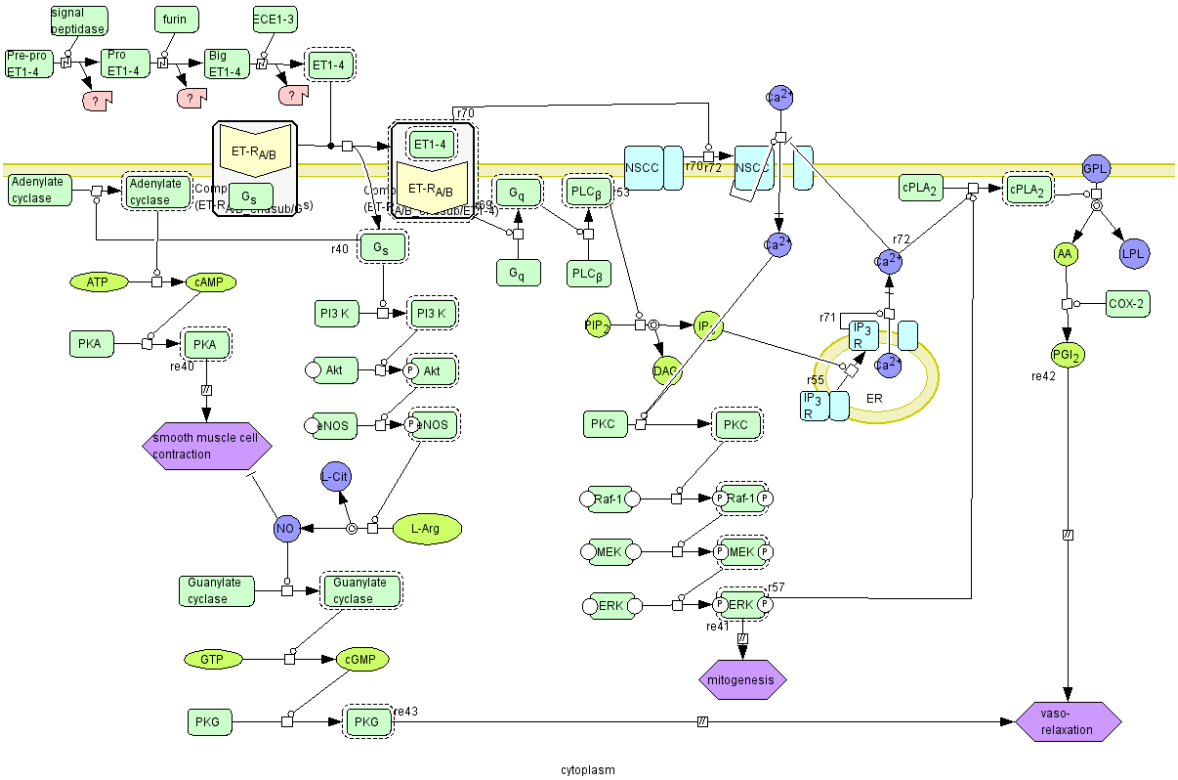
Slika 8: Biološka pot »*Adipocytokine signaling pathway*« iz zbirke KEGG. Z zvezdicami so označeni kandidatni geni iz Genomskega atlasa za nalaganje maščobe.

Figure 8: Biological pathway »*Adipocytokine signaling pathway*« from the database KEGG. Candidate genes from Obesity Genomic Atlas are marked with stars.



Slika 9: Biološka pot »*Insulin signaling pathway*« iz zbirke KEGG. Z zvezdicami so označeni kandidatni geni iz Genomskega atlasa za nalaganje maščobe.

Figure 9: Biological pathway »*Insulin signaling pathway*« from the database KEGG. Candidate genes from Obesity Genomic Atlas are marked with stars.

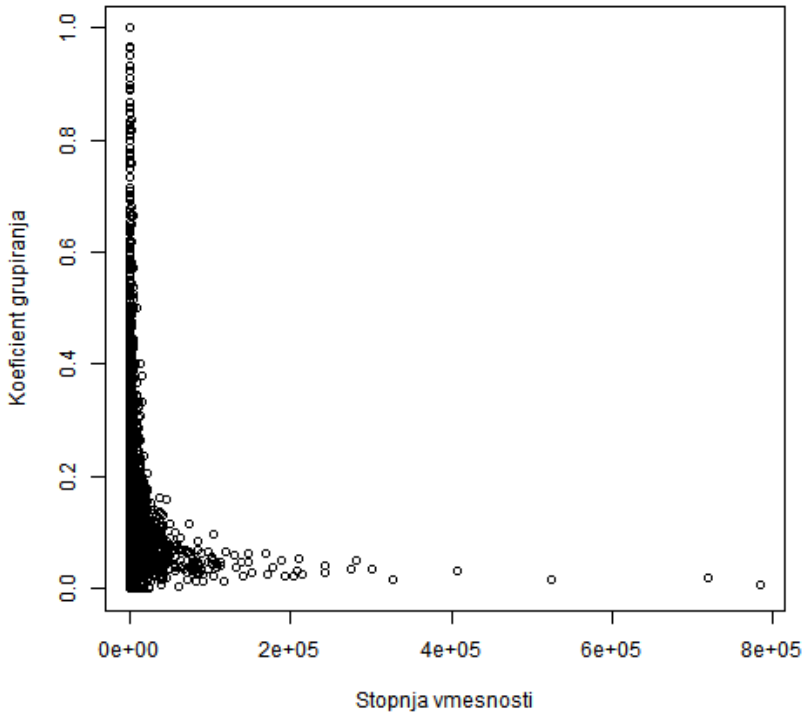


Slika 10: Biološka pot »*Endothelin signaling pathway*« iz zbirke PANTHER

Figure 10: Biological pathway »*Endothelin signaling pathway*« from the PANTHER database

4.2.3 Analiza mreže proteinskih interakcij in mreže uravnavanja izražanja genov

Mrežo proteinskih interakcij genov, povezanih z nalaganjem maščobe, smo izdelali s pomočjo orodja BisoGenet (Martin in sod., 2010) za Cytoscape. Poleg 1553 proteinov iz kataloga genov za nalaganje maščobe so v mreži proteinskih interakcij tudi njihovi prvi sosedi, ki smo jih poiskali v zbirkah proteinskih interakcij DIP, BioGRID, BIND, MINT in IntAct. V mreži proteinskih interakcij je 8046 vozlišč (proteinov) in 92638 povezav (interakcij). S pomočjo vtičnika CytoHubba smo za vozlišča v mreži izračunali vmesnostno centralnost (angl. *betweenness centrality*) in koeficient grupiranja (angl. *clustering coefficient*). Izkazalo se je, da imajo vozlišča z najvišjo vmesnostno centralnostjo nizek koeficient grupiranja (Slika 11). Vozlišča smo razporedili glede na vmesnostno centralnost (od najvišje do najnižje) (Preglednica 5).



Slika 11: Vrednosti koeficientov grupiranja in vmesnostne centralnosti za vozlišča v mreži proteinskih interakcij

Figure 11: Values of clustering coefficient and betweenness centrality for nodes in the protein interaction network

Preglednica 6: Vrednosti koeficientov grupiranja in vmesnostne centralnosti za 20 vozlišč z najvišjo mero vmesnosti v mreži proteinskih interakcij

Table 6: Values of clustering coefficient and betweenness centrality for the 20 genes with the highest betweenness centrality in the protein interaction network

Gen	**Ortolog pri miši**	**Ime gena**	**Vmesnostna centralnost**	**Koeficient grupiranja**
UBC	*Ubc*	ubiquitin C	$2{,}10*10^{12}$	0,00495
*GRB2**	*Grb2*	growth factor receptor-bound protein 2	709077	0,02024
SUMO2	*Sumo2*	SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae)	494299	0,01792
CUL3	*Cul3*	cullin 3	390324	0,01904
*TP53**	*Trp53*	tumor protein p53	390098	0,03234
*SRC**	*Src*	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)	293855	0,03631
KIAA0101	*2810417H13Rik*	*KIAA0101*	284686	0,01712
EP300	*Ep300*	E1A binding protein p300	273381	0,05062
*EGFR**	*Egfr*	epidermal growth factor receptor	260452	0,03411
YWHAZ	*Ywhaz*	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	250431	0,03546
*IKBKG**	*Ikbkg*	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma	250153	0,03118
SUMO1	*Sumo1*	SMT3 suppressor of mif two 3 homolog 1 (S. cerevisiae)	243752	0,03025
*SMAD3**	*Smad3*	SMAD family member 3	224847	0,04277

se nadaljuje

nadaljevanje

Gen	Ortolog pri miši	Ime gena	Vmesnostna centralnost	Koeficient grupiranja
*TRAF6**	*Traf6*	TNF receptor-associated factor 6, E3 ubiquitin protein ligase	214340	0,02265
*MYC**	*Myc*	v-myc myelocytomatosis viral oncogene homolog (avian)	209230	0,02474
*EWSR1**	*Ewsr1*	Ewing sarcoma breakpoint region 1	202317	0,03236
COPS5	*Cops5*	COP9 constitutive photomorphogenic homolog subunit 5 (Arabidopsis)	199250	0,02970
*ESR1**	*Esr1*	estrogen receptor 1	195516	0,05656
*PIK3R1**	*Pik3r1*	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	190760	0,05108
*GSK3B**	*Gsk3b*	glycogen synthase kinase 3 beta	174475	0,02789

Z zvezdico so označeni kandidatni geni iz Genomskega atlasa za nalaganje maščobe.

S pomočjo orodja CyTargetLinker za Cytoscape smo ustvarili mrežo interakcij genov, povezanih z nalaganjem maščobe, z molekulami miRNA pri miši. S kandidatnimi geni iz kataloga lokusov za nalaganje maščobe smo povezali gene za miRNA, če je kandidatni gen 1) eksperimentalno potrjena tarča molekule miRNA iz zbirke miRecords, 2) gostiteljski gen miRNA (Priloga H). Tako smo dobili mrežo s 533 vozlišči in 351 povezavami. Zbrali smo vse molekule miRNA, ki so povezane z geni za nalaganje maščobe (Preglednica 7).

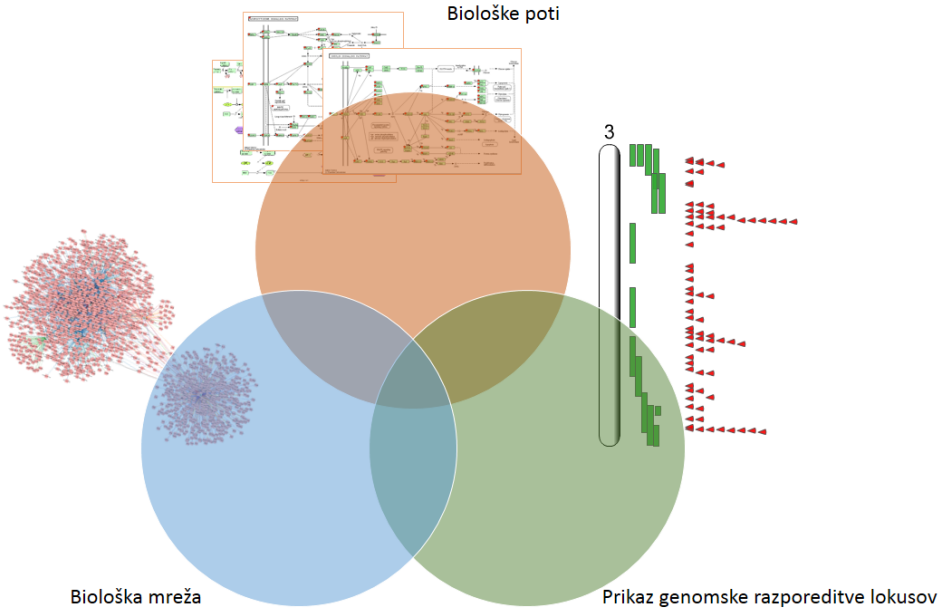
Preglednica 7: Kandidatni miRNA geni za nalaganje maščobe

Table 7: Candidate miRNA genes for fat deposition

Vrsta povezave	**Geni za miRNA**
Kandidatni gen za nalaganje maščobe je tarča miRNA	*let-7b, let-7c-1, let-7c-2, let-7e, mir-101b, mir-106a, mir-10b, mir-122, mir-124-1, mir-124-2, mir-124-3, mir-125b-1, mir-125b-2, mir-126, mir-128-1, mir-128-2, mir-130a, mir-133a-1, mir-133a-2, mir-134, mir-135a-1, mir-135a-2, mir-138-1, mir-138-2, mir-139, mir-140, mir-141, mir-142, mir-146a, mir-150, mir-155, mir-15a, mir-16-1, mir-16-2, mir-17, mir-181a-1, mir-181a-2, mir-181b-1, mir-181b-2, mir-196a-1, mir-196a-2, mir-199a-1, mir-199a-2, mir-19b-1, mir-19b-2, mir-1a-1, mir-1a-2, mir-200a, mir-200b, mir-205, mir-206, mir-208a, mir-20a, mir-20b, mir-210, mir-214, mir-216a, mir-21a, mir-223, mir-23a, mir-23b, mir-24-1, mir-24-2, mir-26a-1, mir-26a-2, mir-27a, mir-27b, mir-2861, mir-290, mir-291a, mir-292, mir-293, mir-294, mir-295, mir-298, mir-29a, mir-29b-1, mir-29b-2, mir-29c, mir-31, mir-328, mir-346, mir-34a, mir-375, mir-449a, mir-466l, mir-470, mir-495, mir-497, mir-706, mir-7b, mir-9-1, mir-9-2, mir-92a-1, mir-92a-2, mir-93, mir-9-3, mir-96*
Kandidatni gen za nalaganje maščobe je gostiteljski gen za miRNA	*mir-1264, mir-1298, mir-149, mir-1904, mir-1912, mir-1938, mir-1954, mir-1957a, mir-1958, mir-218-2, mir-3076, mir-3095, mir-33, mir-335, mir-343, mir-361, mir-378a, mir-3967, mir-448, mir-483, mir-5120, mir-5122, mir-5123, mir-5623, mir-582, mir-592, mir-599, mir-6353, mir-6384, mir-6391, mir-6392, mir-6400, mir-6418, mir-717, mir-721, mir-763, mir-764*

4.2.4 Razvrščanje protein-kodirajočih kandidatnih genov po prioriteti

Pri razvrščanju genov po prioriteti smo za protein-kodirajoče gene in gene za miRNA uporabili različna pristopa. Med kandidatne gene, ki kodirajo proteine, so prišli v poštev vsi geni iz mreže proteinskih interakcij. Razvrstili smo jih glede na vmesnostno centralnost v mreži proteinskih interakcij. Začeli smo z genom z najvišjo vmesnostno centralnostjo. Preverili smo, v katerih izmed prvih šestih najbolj obogatenih bioloških poteh se nahaja gen in s koliko QTL-i se prekriva. Postopek smo ponovili za vse gene iz mreže proteinskih interakcij. Izbrali smo samo tiste gene, ki so zadostili obema pogojema: 1) prisotnost v eni izmed prvih šestih najbolj obogatenih poteh in 2) prekrivanje z vsaj enim QTL-om (Slika 12), (Preglednica 8).



Slika 12: Shema razvrščanja lokusov po prioriteti

Figure 12: Scheme of candidate gene prioritisation

Preglednica 8: Seznam razvrščenih kandidatnih genov za nalaganje maščobe po prioriteti

Table 8: List of prioritised candidate genes for fat deposition

Gen	**Ime gena**	**QTL**	**Biološke poti**
UBC	ubiquitin C	BW104_H, BW201_H, BW218_H, BW476_H, BW501_H, BW502_H, BW198_H, BW235_H, BW278_H, BW58_H	*PPAR signaling pathway*
GRB2	growth factor receptor-bound protein 2	BW116_H, BW332_H, BW487_H	*Insulin signaling pathway*
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	BW407_H, BW61_H	*Insulin signaling pathway*, *Endothelin signaling pathway*
GSK3B	glycogen synthase kinase 3 beta	BW337_H	*Insulin signaling pathway*
PRKCA	protein kinase C, alpha	BW108_H, BW332_H, BW444_H, BW487_H, BW92_H	*Endothelin signaling pathway*

se nadaljuje

nadaljevanje

Gen	**Ime gena**	**QTL**	**Biološke poti**
PRKACA	protein kinase, cAMP-dependent, catalytic, alpha	BW138_H	*Insulin signaling pathway, Endothelin signaling pathway*
TRAF2	TNF receptor-associated factor 2	BW231_H	*Adipocytokine signaling pathway*
STAT3	signal transducer and activator of transcription 3	BW310_H, BW363_H, BW271_H, BW499_H	*Adipocytokine signaling pathway*
CALM1	calmodulin 1 (phosphorylase kinase, delta)	BW266_H	*Insulin signaling pathway*
CALM2	calmodulin 2 (phosphorylase kinase, delta)	BW483_H	*Insulin signaling pathway*
RELA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)	BW329_H, BW452_H	*Adipocytokine signaling pathway*
MAPK1	mitogen-activated protein kinase 1	BW313_H	*Insulin signaling pathway, Endothelin signaling pathway*
PTPN11	protein tyrosine phosphatase, non-receptor type 11	BW104_H, BW201_H, BW37_H, BW195_H, BW198_H, BW235_H, BW278_H, BW58_H	*Adipocytokine signaling pathway*
GNAQ	guanine nucleotide binding protein (G protein), q polypeptide	BW102_H, BW44_H	*Endothelin signaling pathway,*
CBL	Cbl proto-oncogene, E3 ubiquitin protein ligase	BW474_H, BW298_H	*Insulin signaling pathway*
PPP1CA	protein phosphatase 1, catalytic subunit, alpha isozyme	BW329_H, BW452_H, BW274_H	*Insulin signaling pathway*
CAV1	caveolin 1, caveolae protein, 22kDa	BW18_H,	*Metabolism of lipids and lipoproteins*

se nadaljuje

nadaljevanje

Gen	Ime gena	QTL	Biološke poti
NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)	BW380_H, BW381_H, BW61_H	*Neuroactive ligand-receptor interaction*
CRK	v-crk sarcoma virus CT10 oncogene homolog (avian)	BW145_H, BW163_H	*Insulin signaling pathway*
NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	BW441_H	*Adipocytokine signaling pathway*

4.2.5 Razvrščanje kandidatnih miRNA genov po prioriteti z orodjem miRNA SNiPer

S postopkom integracije raznovrstnih podatkov smo zbrali 135 kandidatnih miRNA genov, povezanih z nalaganjem maščobe pri miši. Od tega jih je 98 zbranih, ker so njihove tarče (eksperimentalno dokazane in napovedane) geni, povezani z nalaganjem maščobe, 37 pa jih je znotraj gostiteljskih genov, ki so povezani z nalaganjem maščobe. V katalogu lokusov za nalaganje maščobe je 87 kandidatnih miRNA genov (Preglednica 9).

Preglednica 9: Število kandidatnih miRNA genov za nalaganje maščobe pri miši

Table 9: Number of miRNA candidate genes for fat deposition in mouse

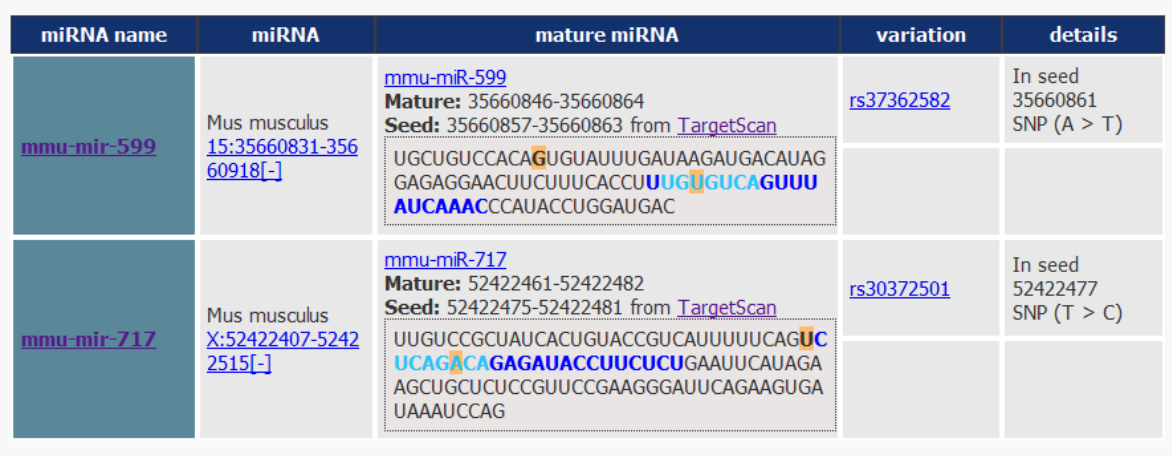
Kandidatni geni miRNA	Število kandidatnih miRNA genov
gen iz kataloga za nalaganje maščobe je tarča gena miRNA pri miši	98
miRNA znotraj gostiteljskega gena iz kataloga kandidatnih lokusov za nalaganje maščobe pri miši	37
miRNA iz kataloga kandidatnih lokusov za nalaganje maščobe pri miši	87

Za razvrščanje po prioriteti na osnovi genetskih polimorfizmov smo razvili orodje miRNA SNiPer, s pomočjo katerega lahko poiščemo SNP-je znotraj različnih regij genov miRNA (pre-miRNA, zrela miRNA, regija *seed*). V orodje smo vnesli seznam vseh kandidatnih miRNA genov in poiskali tiste, ki imajo polimorfno regijo *seed*. Pri miši sta imela polimorfno regijo *seed* gena *mmu-mir-717* in *mmu-mir-599* (Slika 13), miRNA geni s polimorfnimi regijami *seed* pri človeku so navedeni v Preglednici 10.

Preglednica 10: Geni za miRNA pri človeku in miši, ki imajo SNP v regiji *seed*, ki je odgovorna za vezavo na tarče

Table 10: Human and mouse miRNA genes with SNP within seed region, responsible for target binding

miRNA	***SNP***
hsa-mir-122	*rs41292412*
hsa-mir-125a	*rs12975333*
hsa-mir-145	*rs190323149*
hsa-mir-15b	*rs192595529*
hsa-mir-221	*rs113054794*
mmu-mir-717	*rs30372501*
mmu-mir-599	*rs37362582*



miRNA name	miRNA	mature miRNA	variation	details
mmu-mir-599	Mus musculus 15:35660831-35660918[-]	mmu-miR-599 **Mature:** 35660846-35660864 **Seed:** 35660857-35660863 from TargetScan UGCUGUCCACAGUGUAUUUGAUAAGAUGACAUAG GAGAGGAACUUCUUUCACCUUUGUGUCAGUUU AUCAAACCCAUACCUGGAUGAC	rs37362582	In seed 35660861 SNP (A > T)
mmu-mir-717	Mus musculus X:52422407-52422515[-]	mmu-miR-717 **Mature:** 52422461-52422482 **Seed:** 52422475-52422481 from TargetScan UUGUCCGCUAUCACUGUACCGUCAUUUUUCAGUC UCAGACAGAGAUACCUUCUCUGAAUUCAUAGA AGCUGCUCUCCGUUCCGAAGGGAUUCAGAAGUGA UAAAUCCAG	rs30372501	In seed 52422477 SNP (T > C)

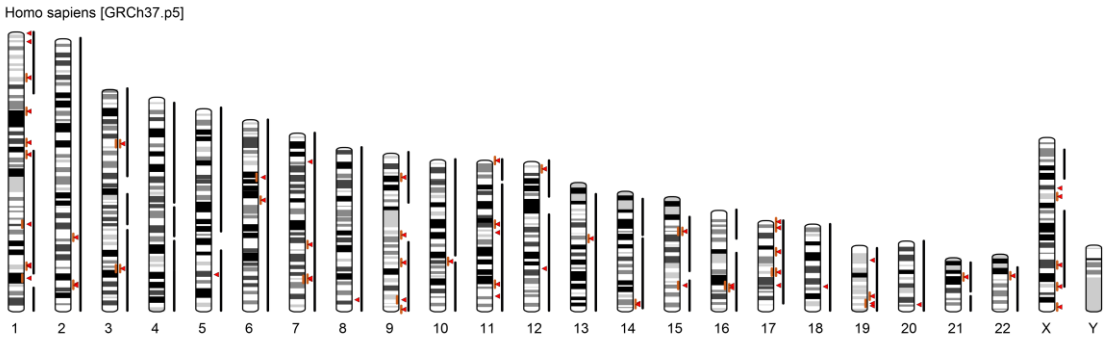
Slika 13: Rezultat orodja miRNA SNiPer: polimorfni regiji *seed* pri genih za miRNA pri miši *mmu-mir-717* in *mmu-mir-599*

Figure 13: Result table of the miRNA SNiPer tool: polymorphic seed regions within miRNA genes in mouse *mmu-mir-717* and *mmu-mir-599*

4.2.6 Razvrščanje kandidatnih miRNA genov po prioriteti s pomočjo orodja miRNA Viewer

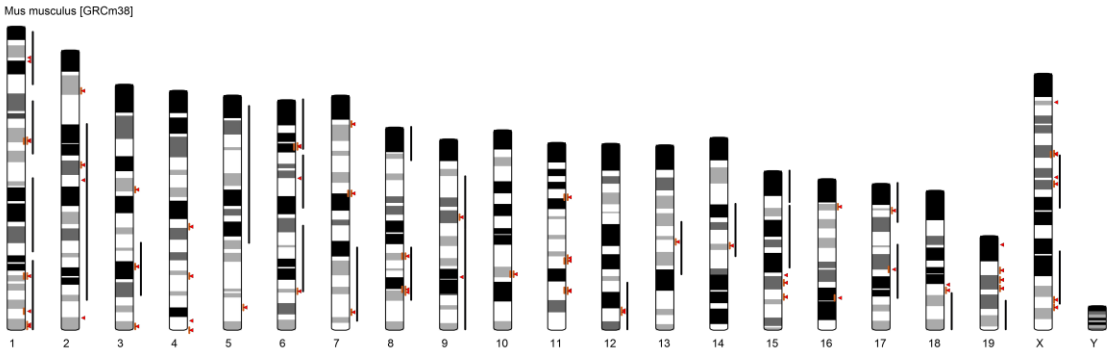
Za razvrščanje kandidatnih genov miRNA po prioriteti smo razvili tudi orodje miRNA Viewer, ki omogoča prikaz genomske razporeditve genov miRNA in njihovih gostiteljskih genov. Orodje zajema podatke o QTL-ih pri človeku in miši iz zbirk OMIM in RGD. S pomočjo miRNA Viewer-ja smo prikazali genomsko razporeditev kandidatnih miRNA genov za nalaganje maščobe na genomu človeka (87 ortologov) (Slika 14) in na genomu miši (87 ortologov) (Slika 15). Prikazali smo tudi QTL-e za nalaganje maščobe in odkrili prekrivanje 47 genov za miRNA s 53 gostiteljskimi geni in 92 QTL-i, povezanimi z nalaganjem maščobe pri človeku. V 64 pre-miRNA smo našli SNP-je, od teh jih je pet znotraj regij *seed*. Odkrili smo primer genomskega prekrivanje lokusov, povezanih z

nalaganjem maščobe, na štirih ravneh (Slika 16): 1) gen za miRNA (hsa-mir-15b), 2) QTL (BW348_H; body weight QTL), 3) protein-kodirajoči gostiteljski gen za miRNA *SMC4* v smerni (angl. *sense*) orientaciji, RP11 v protismerni (angl. *antisense*) orientaciji in 4) SNP v regiji *seed* gena za miRNA, rs192595529.



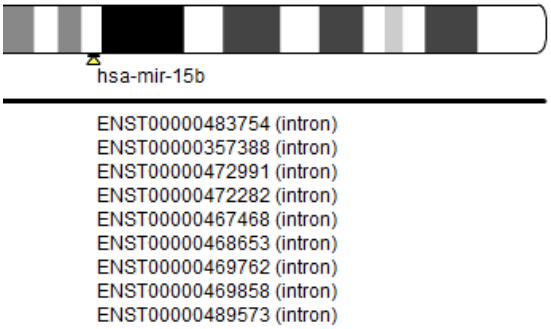
Slika 14: Genomska razporeditev genov miRNA in QTL-ov, povezanih z nalaganjem maščobe ter gostiteljskih genov za miRNA pri človeku

Figure 14: Genomic view of miRNA genes and QTL associated with obesity and miRNA host genes in human



Slika 15: Genomska razporeditev genov miRNA in QTL-ov, povezanih z nalaganjem maščobe ter gostiteljskih genov za miRNA pri miši

Figure 15: Genomic view of miRNA genes and QTL associated with obesity and miRNA host genes in mouse



Slika 16: Prekrivanje genomskih lokusov, povezanih z nalaganjem maščobe: SNP, gen za miRNA, protein-kodirajoči gen in QTL

Figure 16: Genomic overlap of obesity related loci: SNP, miRNA gene, protein-coding host gene and QTL

4.3 ANALIZA POVEZANOSTI GENOTIPA S FENOTIPOM

Z analizo povezanosti genotipa s fenotipom smo preverili vpliv polimorfizmov v kandidatnih genih na nalaganje maščobe pri miših. Genetsko variabilnost genov (Preglednica 3) in meritve za lastnosti nalaganja maščobe ter telesne sestave smo pridobili v podatkovni zbirki MPD. Analizirali smo vpliv polimorfizmov štirih kandidatnih genov na fenotip: v genu *Akt1*, ki se nahaja v treh kandidatnih poteh za nalaganje maščobe, v genih *Ubc* in *Grb2*, ki imata visoko vmesnostno centralnost v mreži proteinskih interakcij, se nahajata znotraj QTL-ov in sta v kandidatnih bioloških poteh za nalaganje maščobe, ter v genu *Mir599*, ki ima polimorfizem v območju miRNA, odgovornem za vezavo na mRNA, in se nahaja znotraj gostiteljskega gena *VPS13B*, ki je povezan z nalaganjem maščobe. Preverili smo tudi vpliv polimorfizma znotraj gena *Mapkap1*, ki ima nizko vmesnostno centralnost, se ne nahaja v kandidatnih bioloških poteh, prekriva pa se z dvema QTL-oma za nalaganje maščobe.

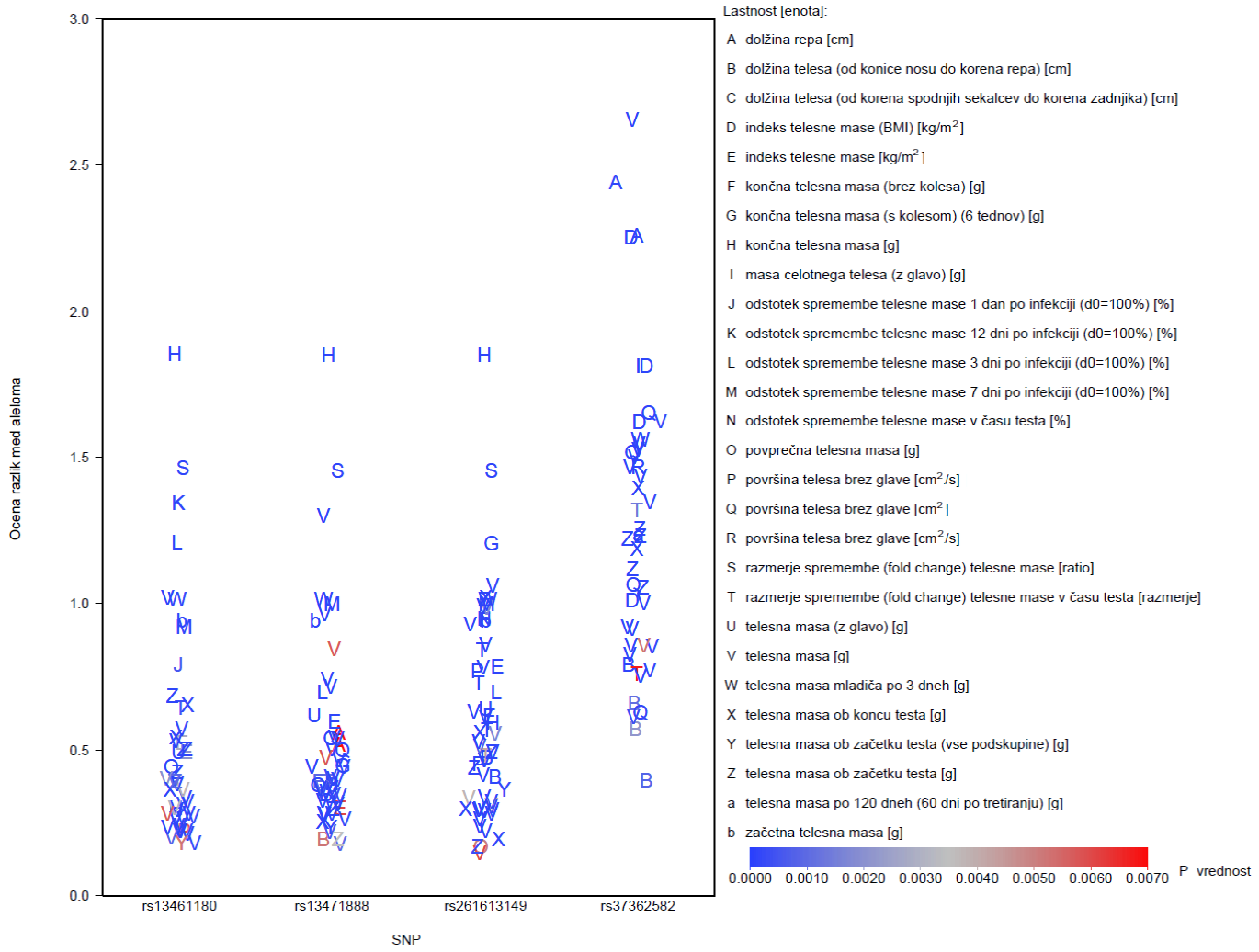
V asociacijsko analizo smo vključili 268 lastnosti. Število lastnosti, povezanih z analiziranimi kandidatnimi geni, je v preglednici 11. Od 76 lastnosti, ki opisujejo sestavo telesa, jih je 32 povezanih s SNP-jem gena *Akt1*, 34 s SNP-jem gena *Ubc*, 30 s SNP-jem gena *Grb2*, 30 s SNP-jem gena *Mir599* in dva SNP-ja gena *Mapkap1*. Lastnosti, ki so povezane s telesnimi maščobnimi blazinicami, je 26, od tega so 4 povezane s SNP-jem v genu *Akt1*, 2 v genu *Ubc* in 2 v genu *Grb2*. Polimorfizma v genih *Mir599* in *Mapkap1* nista povezana s skupino lastnosti telesne maščobne blazinice. Telesno težo opisuje 166 lastnosti, od tega jih je 59 povezanih z SNP-jem v genu *Akt1*, 50 v genu *Ubc*, 57 v genu *Grb2*, 45 v genu *Mir599* in 26 v genu *Mapkap1*.

Preglednica 11: Rezultat analize povezanosti genotipa s fenotipom pri miši. Število fenotipskih lastnosti pri inbridiranih linijah miši, na katere imajo polimorfizmi rs37362582, rs32568344, rs261613149, rs13461180 in rs13471888 učinek.

Table 11: The results of analysis of association between genotype and phenotype. The number of phenotypic traits in inbred strains of mice, on which polymorphisms rs37362582, rs32568344, rs261613149, rs13461180 and rs13471888 have effect.

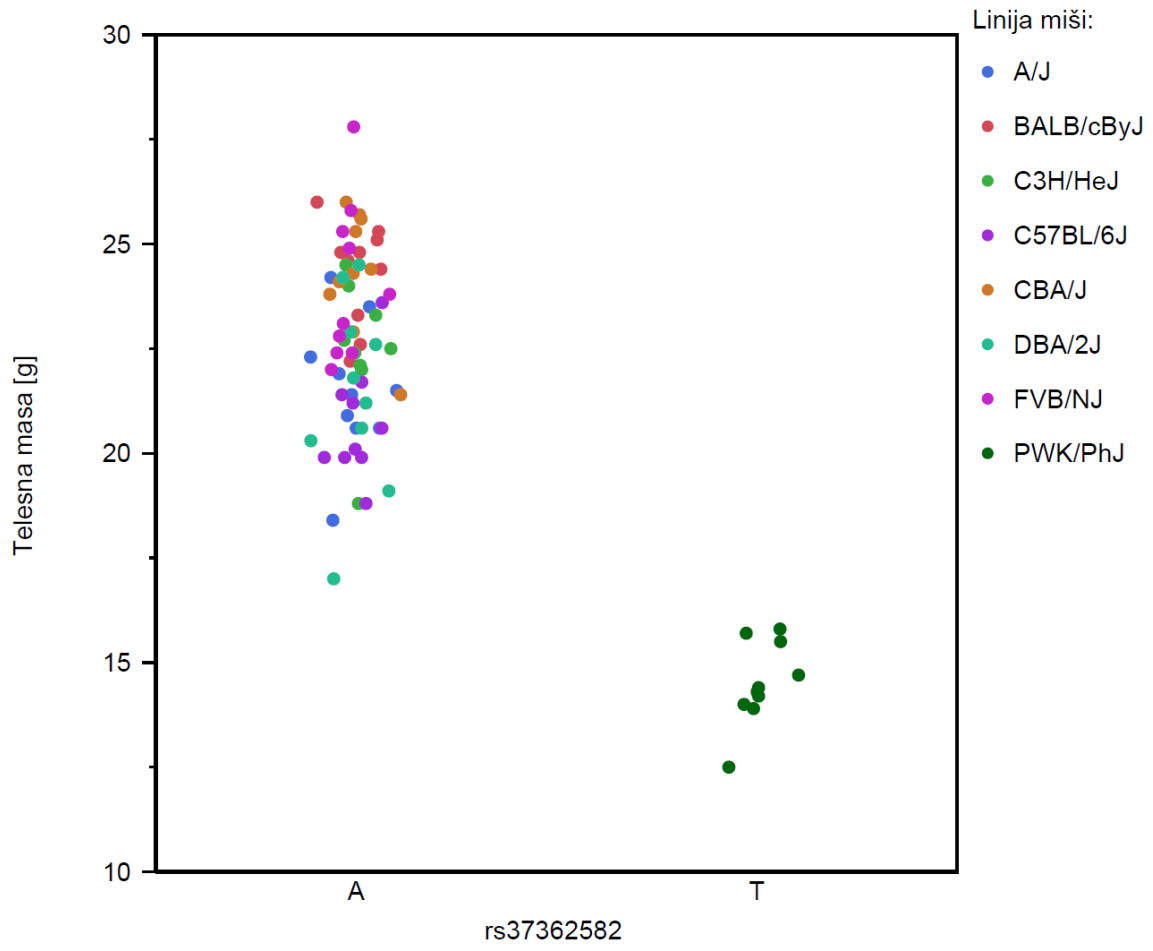
skupina opazovane lastnosti	*Mir599* rs37362582	*Mapkap1* rs32568344	*Akt1* rs261613149	*Ubc* rs13461180	*Grb2* rs13471888
sestava telesa (76 lastnosti)	30	2	32	34	30
telesne maščobne blazinice (26 lastnosti)	0	0	4	2	2
telesna teža, velikost in rast (166 lastnosti)	45	26	59	50	57
skupno (268 lastnosti)	75	28	95	86	89

Rezultate analize povezanosti genotipa s fenotipom lahko razberemo tudi s slike, ki prikazuje ocene razlik med aleloma v analiziranih SNP-jih (Slika 17). Za lažjo primerjavo smo vrednosti standardizirali. Črke ponazarjajo statistično značilne povezave med analiziranimi lastnostmi in SNP-ji ($p < 0,01$). Največje razlike opazimo pri povezavi med SNP-jem rs37362582 in lastnostmi: telesna masa (na sliki označena s črko V), dolžina repa (A) in indeks telesne mase (D). Pri ostalih treh SNP-jih je največja razlika med aleloma pri lastnosti telesna masa (H), sledi razmerje spremembe telesne mase (S). Ocena vpliva genotipa je največja (največja razlika med izmerjeno težo za različna alela) za lastnost telesna masa v SNP-ju rs37362582 gena *Mir599*. Meritve telesne mase pri osmih linijah miši, od katerih je pri sedmih linijah (A/J, BALB/cByJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J, FVB/NJ) na SNP-ju rs37362582 alel A, pri eni liniji (PWK/PhJ) pa alel T, so prikazane v preglednici (Preglednica 12) in na grafu (Slika 18).



Slika 17: Ocena razlik med aleloma po analiziranih SNP-jih

Figure 17: Estimated differences in alleles by analysed SNPs



Slika 18: Meritve telesne mase miši po alelih SNP-ja rs37362582

Figure 18: Measurements of body weight in mice by alleles SNP rs37362582

Preglednica 12: Fenotipski podatki iz zbirke MPD: telesne mase samcev pri miši ob koncu testa

Table 12: Phenotypic data from MPD database: final body weight of male mice

Oznaka živali	**Linija**	**Telesna masa [g]**
1201	A/J	20,6
1202	A/J	21,9
1203	A/J	21,5
1204	A/J	21,4
1205	A/J	23,5
1206	A/J	22,3
1207	A/J	20,9
1208	A/J	24,2
1209	A/J	20,6
1210	A/J	18,4
1001	BALB/cByJ	23,3
1002	BALB/cByJ	22,2
1003	BALB/cByJ	25,3
1004	BALB/cByJ	22,6
1005	BALB/cByJ	26,0
1006	BALB/cByJ	24,8
1007	BALB/cByJ	24,4
1008	BALB/cByJ	25,1
1009	BALB/cByJ	24,8
1010	BALB/cByJ	24,6
301	C3H/HeJ	24,5
302	C3H/HeJ	22,0
303	C3H/HeJ	22,4
304	C3H/HeJ	22,8
305	C3H/HeJ	23,3
306	C3H/HeJ	22,7
307	C3H/HeJ	22,5
308	C3H/HeJ	24,0
309	C3H/HeJ	22,1
310	C3H/HeJ	18,8
1301	C57BL/6J	21,7
1302	C57BL/6J	21,4
1303	C57BL/6J	21,2
1304	C57BL/6J	20,6
1305	C57BL/6J	19,9
1306	C57BL/6J	19,9
1307	C57BL/6J	19,9
1308	C57BL/6J	20,1
1309	C57BL/6J	18,8
1310	C57BL/6J	23,6
401	CBA/J	25,3
402	CBA/J	25,6
403	CBA/J	21,4
404	CBA/J	22,9

se nadaljuje

nadaljevanje

Oznaka živali	**Linija**	**Telesna masa [g]**
405	CBA/J	24,4
406	CBA/J	23,8
407	CBA/J	25,7
408	CBA/J	24,3
409	CBA/J	24,1
410	CBA/J	26,0
501	DBA/2J	24,2
502	DBA/2J	22,9
503	DBA/2J	21,8
504	DBA/2J	21,2
505	DBA/2J	22,6
506	DBA/2J	19,1
507	DBA/2J	20,3
508	DBA/2J	24,5
509	DBA/2J	17,0
510	DBA/2J	20,6
601	FVB/NJ	22,4
602	FVB/NJ	27,8
603	FVB/NJ	23,1
604	FVB/NJ	23,8
605	FVB/NJ	25,3
606	FVB/NJ	25,8
607	FVB/NJ	22,8
608	FVB/NJ	22,4
609	FVB/NJ	22,0
610	FVB/NJ	24,9
801	PWK/PhJ	14,0
802	PWK/PhJ	15,7
803	PWK/PhJ	12,5
804	PWK/PhJ	14,2
805	PWK/PhJ	13,9
806	PWK/PhJ	14,7
807	PWK/PhJ	14,4
808	PWK/PhJ	15,5
809	PWK/PhJ	15,8
810	PWK/PhJ	14,3

5 RAZPRAVA

5.1 GENOMSKI ATLAS ZA NALAGANJE MAŠČOBE

Razumevanje molekularnih mehanizmov kompleksnih fenotipov je danes največji izziv raziskav v biomedicini, zato je veliko študij usmerjenih v odkrivanje vzročnih genov za večgenske bolezni (Botstein in Risch, 2003). Pri razvoju večine bolezni pri človeku ni enega vzročnega gena, temveč gre za zapleteno mrežo soodvisnosti več genetskih in okoljskih dejavnikov (Hirschhorn in Daly, 2005). Raziskovalne skupine objavljajo prosto dostopne zbirke vzročnih genov za bolezni (OMIM, CTD, GAD). Nalaganje maščobe je kompleksen fenotip, prekomerno nalaganje maščobe pa je vzrok za debelost, ki dobiva epidemiološke razsežnosti. Z debelostjo so povezane številne bolezni, kot so artritis, bolezen žolčnika, diabetes tipa 2, bolezni srca in ožilja, številne vrste raka, neplodnost in druge. Prekomerno nalaganje maščobe je nezaželeno tudi pri prireji mesa in mleka v živinoreji. V povezavi z debelostjo so do sedaj odkrili več kot tisoč posameznih lokusov, molekularni mehanizem prekomernega nalaganja maščobe pa še vedno ni pojasnjen.

Izdelali smo Genomski atlas lokusov za nalaganje maščobe. Poiskali smo zbirke lokusov, ki so povezane z nalaganjem maščobe in združili podatke iz zbirk GeneCards, OMIM, RGD, MGI, Obesity Gene Map in literature. Ročno smo pregledali dosedanje objave o genetskih vzrokih za nalaganje maščobe pri človeku, miši, podgani in govedu ter zbrali 1774 genov, povezanih z nalaganjem maščobe, od tega je 1553 protein-kodirajočih genov in 221 genov za miRNA. Katalog smo dopolnili s QTL-i za nalaganje maščobe iz zbirk RGD in AnimalQTLdb. Zbrali smo 222 QTL-ov pri človeku in 38 pri miši.

Podatki o vzročnih genih za nalaganje maščobe so razpršeni in nepopolni, poimenovanja genov in bolezni v različnih virih niso poenotena. V zbirki OMIM so samo lokusi z močnim vplivom na fenotip, v zbirki GAD so rezultati asociacijskih študij (Zhang in sod., 2010), zbirka CTD (Davis in sod., 2013) vsebuje le 125 lokusov, ki so povezani z debelostjo pri človeku, specializirana zbirka kandidatnih genov za nalaganje maščobe Obesity Gene Map (Rankinen in sod., 2006) pa je bila nazadnje posodobljena leta 2006.

Genomski atlas lokusov za nalaganje maščobe predstavlja centralno mesto za raziskave na tem področju in je trenutno najobsežnejša zbirka z ročno preverjenimi lokusi, povezanimi z nalaganjem maščobe, ter predstavlja pomemben vir informacij za nadaljnje raziskave. Ročno pregledovanje literature in zbiranje kandidatnih genov je časovno zahtevno (Wei in sod., 2012), zato bi bilo treba v prihodnje postopek zbiranja kandidatnih genov podpreti z računalniškimi programi, kar bi olajšalo sprotno posodabljanje zbirke. Pri analizi zbranih podatkov se je izkazalo, da bi bili potrebni za nadaljnje raziskave še nekateri podatki o kandidatnih lokusih, kot so na primer vrednost LOD pri QTL-ih, stopnja procesa razvoja debelosti, pri kateri je kandidatni gen vpleten, ter eksperimentalna metoda, s katero so povezali lokus z nalaganjem maščobe. Na primer, kandidatni gen *TRAF6* je vnetni gen, ki

prispeva k pospeševanju debelosti, ni pa primarni vzrok za debelost. Podobno je s kandidatnim genom *TP53*, ki je tumor-supresorski gen in se verjetno vključi, ko že prihaja do močne nekroze in apoptoze tkiva. Za razumevanje molekularnih osnov mehanizma nalaganja maščobe je potrebno zajeti kandidatne gene na vseh stopnjah procesa razvoja debelosti, vendar je treba pri analizi upoštevati, ali gre za vzročne ali sekundarne dejavnike. Do sedaj samo zbirali molekularno genetske lokuse in QTL-e, v prihodnje pa bi bilo treba dodati še več vrst genomskih podatkov, kot so molekularno-citogenetski (CGH, aCGH) in epigenetski mehanizmi.

5.2 INTEGRACIJA GENOMSKIH PODATKOV

V zadnjih letih je na razpolago vedno več genomskih podatkov (Hamid in sod., 2009). Primeri genomskih in proteomskih podatkov so podatki o izražanju genov, SNP-jih, različicah v številu kopij (angl. *copy number variation;* CNV) in proteinskih interakcijah, pridobljenih z visokozmogljivimi metodami, kot so mikromreže (Schena in sod., 1995), aCGH (Oostlander in sod., 2004) in masna spektrometrija (Aebersold in Mann, 2003). Vsaka izmed vrst genomskih podatkov prispeva k popolnejšemu vpogledu v celoten genom. Integracija raznovrstnih genomskih podatkov je postala osnovno orodje v genomskih raziskavah. Načinov za integracijo genomskih podatkov je vedno več, zato se pojavlja potreba po izboljšavah in standardizaciji metod za integracijo (Reif in sod., 2004).

5.2.1 Genomski prikazi razporeditve lokusov

Z združevanjem raznovrstnih genomskih podatkov smo poskušali celostno povezati različne ravni dosedanjih ugotovitev o nalaganju maščobe pri več vrstah. Za gene pri govedu, podgani in miši iz Genomskega atlasa lokusov za nalaganje maščobe smo poiskali ortologne gene pri človeku in izdelali genomski prikaz vseh kandidatnih lokusov za nalaganje maščobe ter ga objavili na spletni strani (Kunej in sod., 2012). Identificirali smo prekrivanja med lokusi, ki so jih z nalaganjem maščobe povezali v neodvisnih študijah z različnimi eksperimentalnimi pristopi.

Genomski prikazi kandidatnih lokusov so v zbirki RGD (Laulederkind in sod., 2011) za prikaz kandidatnih genov in QTL-ov za nekatere bolezni (rak, bolezen srca in ožilja, diabetes, debelost, nevrološke bolezni) pri človeku, miši in podgani ter v zbirki AnimalQTLdb (Hu in sod., 2007) za prikaz QTL-ov za širok spekter kvantitativnih lastnosti pri prašiču, kravi, kokoši, ovci in postrvi. Prikaze prekrivanj kandidatnih lokusov so uporabili tudi za prikaz kandidatnih genov iz zbirke za nalaganje maščobe, ki je več ne posodabljajo (Rankinen in sod., 2006), ter za prikaz kandidatnih genov za razvoj mlečne žleze in prireje mleka (Ogorevc in sod., 2009). Prikaz prekrivanj med QTL-i omogočata orodji BioMercator (Sosnowski in sod., 2012) in Flash GViewer.

Za raziskavo genetskih mehanizmov kompleksnega tipa nalaganja maščobe smo uporabili pristop sistemske biologije. Združili smo genomske podatke iz več virov (podatkovne

zbirke, literatura), ki zajemajo rezultate različnih študij pri štirih vrstah. Genomski prikaz kandidatnih lokusov Genomskega atlasa za nalaganje maščobe predstavlja najobsežnejši pregled do sedaj znanih genetskih vzrokov za nalaganje maščobe. Tako kot v zbirki RGD smo za genomske prikaze lokusov uporabili orodje Flash GViewer, ki ga je možno uporabiti za interaktivni genomski prikaz dveh vrst lokusov pri izbranem organizmu in ga vključiti v spletne strani. Orodje ima nekatere omejitve: prikaz postane nepregleden, kadar prikazujemo večje število lokusov, orodje ne omogoča omejevanja prikaza glede na vrsto lokusov in ne vključuje iskalnika lokusov, zato bi ga bilo treba v prihodnje izpopolniti.

5.2.2 Analiza obogatenosti bioloških poti

Biološke poti združujejo ročno pregledane podatke iz različnih virov in predstavljajo pomemben vir znanja o bioloških procesih na molekularni ravni (Cerami in sod., 2011). S pomočjo zbirke znanih lokusov za nalaganje maščobe smo izvedli analizo obogatenosti bioloških poti in identificirali kandidatne biološke poti, vpletene v mehanizem nalaganja maščobe. Nekatere izmed kandidatnih bioloških poti za nalaganje maščobe so signalna pot adipocitokinov (angl. *adipocytokine signaling pathway*), signalna pot inzulina (angl. *insulin signaling pathway*), signalna pot endotelina (angl. *endothelin signaling pathway*) in signalna pot PPAR (angl. *PPAR signaling pathway*).

Na osnovi kandidatnih genov so bile do sedaj z analizo obogatenosti bioloških poti že identificirane kandidatne biološke poti za nekatere fenotipe: biološke poti za uravnavanje zasvojenosti z drogami (Li in sod., 2008) z orodjem KOBAS (Wu in sod., 2006), biološke poti za depresijo (Jia in sod., 2011) z orodjem IPA (Ingenuity Pathway Analysis system) in biološke poti za kriptorhizem (Cannistraci in sod., 2013) z orodjem DAVID (Huang in sod., 2007). Množico genov, ki so vpleteni v kandidatne biološke poti, v nekaterih študijah uvrstijo med kandidatne gene (Ochagavía in sod., 2011).

5.2.3 Genska mreža proteinskih interakcij

Interakcije med proteini imajo v večini celičnih procesov centralno vlogo. Analize mrež proteinskih interakcij lahko pripomorejo k razumevanju celičnih funkcij in bioloških procesov. Protein-kodirajoče kandidatne gene smo povezali v mrežo proteinskih interakcij z 8046 vozlišči in 92638 povezavami. Mrežo smo sestavili s pomočjo orodij, ki omogočajo integracijo več podatkovnih zbirk eksperimentalno potrjenih proteinskih interakcij.

Količina dostopnih podatkov o interakcijah med proteini narašča, analiza mrež proteinskih interakcij pa je vedno bolj pogost pristop k razumevanju mehanizmov bolezni (Kann, 2007). Množico znanih kandidatnih genov (v literaturi tudi »*seed genes*«) povežemo z geni, ki so z njimi v interakcijah. Prvi sosedi v mreži proteinskih interakcij so lahko uvrščeni med potencialne kandidatne gene (Oti in sod., 2006; Ochagavía in sod., 2011) med katerimi so lahko tudi novi kandidatni geni. Med prvimi sosedi kandidatnih genov iz Genskega atlasa za nalaganje maščobe je na primer gen *NFKB1*, ki je mediator vnetnega in

imunskega odziva in ni primarni vzrok za debelost (gen *NFKB1* podobno kot gen *TRAF6* prispeva k pospeševanju debelosti).

5.2.4 Integracija mreže uravnavanja izražanja genov z molekulami miRNA

Tiste protein-kodirajoče kandidatne gene za nalaganje maščobe, ki so tarče ali gostiteljski geni za miRNA, smo v mreži povezali z geni za miRNA. Dobili smo mrežo s 533 vozlišči in 351 povezavami. Geni za miRNA uravnavajo protein-kodirajoče gene na potranskripcijski ravni z vezavo na 3'-UTR (Guo in sod., 2010). V mreži smo zajeli tako eksperimentalno potrjene kot napovedane tarče miRNA genov. V nekaterih primerih se miRNA geni nahajajo znotraj gostiteljskih genov.

5.2.5 Postopek integracije genomskih podatkov

Veliko raziskav je usmerjenih v integracijo genomskih podatkov. Količina genomskih podatkov narašča. Z združevanjem raznovrstnih podatkov je možno razkriti, česar z upoštevanjem samo enega vira ali ravni ne bi bilo možno. Z integracijo raznovrstnih podatkov lahko zmanjšamo šum in možnost napak. V nadaljnjih študijah bi lahko uporabili še več vrst genomskih podatkov, kot so na primer podatki o izražanju kandidatnih genov in epigenetski podatki. Postopek integracije genomskih podatkov, ki smo ga izvedli v tej študiji, je lahko osnova za bolj robusten protokol, ki bi ga lahko uporabili za raziskavo drugih kompleksnih fenotipov. Vse korake integracije bi bilo treba povezati in avtomatizirati, kar zahteva razvoj računalniških aplikacij.

5.3 RAZVRŠČANJE KANDIDATNIH GENOV PO PRIORITETI

Pri postopku razvrščanja kandidatnih genov po prioriteti gre za ocenjevanje pomembnosti vloge kandidatnih genov v biološkem procesu, z namenom izbrati najobetavnejše gene za nadaljnje raziskave. Biologi so kandidatne gene razvrščali ročno, kar je postalo zaradi velike količine genomskih podatkov v zadnjem času neobvladljivo. Raziskovalci so začeli razvijati bioinformacijska orodja za razvrščanje kandidatnih genov po prioriteti, ki večinoma temeljijo na iskanju podobnosti med kandidatnimi geni in geni, za katere je že ugotovljena določena vloga pri boleznih ali bioloških procesih (Masoudi-Nejad in sod., 2012).

5.3.1 Razvrščanje protein kodirajočih kandidatnih genov po prioriteti

Poiskali smo skupne gene, ki se nahajajo v najbolj obogatenih bioloških poteh (kandidatnih bioloških poteh za nalaganje maščobe). V prvih treh najbolj obogatenih poteh hkrati se nahajajo geni *AKT1*, *AKT2* in *AKT3*. Vozlišča v mreži proteinskih interakcij smo razvrstili glede na vrednost vmesnostne centralnosti. Med geni *AKT1*, *AKT2* in *AKT3* ima najvišjo vrednost vmesnostne centralnosi gen *AKT1*, zato smo ga uvrstili med potencialne biooznačevalce. Potem smo za gene iz mreže proteinskih interakcij, razvrščene glede na

vrednost vmesnostne centralnosti, preverili, ali so udeleženi v vsaj eni od prvih šestih kandidatnih bioloških poti in ali se hkrati nahaja znotraj vsaj enega QTL-a za nalaganje maščobe. Začeli smo z genom z najvišjo vmesnostno centralnostjo. Kot potencialna biooznačevalca smo zbrali prva dva gena, ki sta ustrezala naštetima pogojema, to sta gena *UBC* in *GRB2*. Vozlišča z visoko vmesnostno centralnostjo v mreži proteinskih interakcij so *UBC*, *GRB2*, *SUMO2*, *CUL3*, *TP53*, *SRC*, *KIAA0101*, *EP300*, *EGFR*, *YWHAZ*, *IKBKG*, *SUMO1*, *SMAD3*, *TRAF6*, *MYC*, *EWSR1*, *COPS5*, *ESR1*, *PIK3R1* in *GSK3B.*

Kompleksne bolezni vplivajo na nepravilno delovanje več organov oziroma tkiv (Gao in sod., 2012). Kompleksne fenotipe, kot je nalaganje maščobe, uravnava več bioloških poti hkrati. Nastanek debelosti je posledica motenj ravnovesja med vnosom hrane in porabo energije, ki je nadzorovano preko kompleksnega homeostatskega sistema, ki ga sestavljajo signalne poti, več tkiv in organov (Xu in sod., 2012). Nekateri geni (oziroma celotni funkcijski moduli) so udeleženi v več bioloških poteh hkrati, kar nakazuje na možnost, da so ti geni vpleteni v mehanizem sopojavnosti bolezni. Vzroke za sopojavnost motnje avtističnega spektra (angl. *autism spectrum disorder;* ASD) in bipolarne motnje so iskali v bioloških poteh, ki so kandidatne poti za oba fenotipa (Ragunath in sod., 2011). Kandidatni gen *AKT1* smo označili kot močnejši kandidatni gen za nalaganje maščobe, ker je vpleten v tri kandidatne poti za nalaganje maščobe hkrati. Gen, ki se nahaja na preseku več bioloških poti, ima lahko centralno vlogo pri oblikovanju fenotipa.

S pomočjo mreže proteinskih interakcij so Lim in sod. iskali kandidatne gene za marmoriranost mišic pri govedu (Lim in sod., 2011). V študiji so zbrali 121 kandidatnih genov, ki so jih povezali v mrežo proteinskih interakcij. Razširjena mreža s sosednimi proteini je vsebovala 1090 vozlišč in 1517 povezav. Podobno kot Lim in sod. smo tudi mi opravili topološko analizo mreže proteinskih interakcij in vozlišča razvrstili glede na vmesnostno centralnost. Visoka mera vmesnosti nakazuje na globalno centralno vlogo vozlišča. Če iz mreže odstranimo vozlišče z visoko mero vmesnosti, močno vplivamo na topologijo mreže. Močnejši vpliv na fenotip ima lahko izguba funkcije proteinov, ki imajo v mreži proteinskih interakcij visoko vmesnost.

Pri postopku razvrščanju kandidatnih genov po prioriteti smo integrirali več ravni genomskih podatkov in zajeli rezultate različnih neodvisnih študij in eksperimentalnih tehnik. Na primer gen *UBC* ima v mreži proteinskih interakcij zelo visoko vmesnostno centralnost, hkrati pa se nahaja znotraj 10 QTL-ov za nalaganje maščobe ter je vpleten v kandidatno biološko pot za nalaganje maščobe (*PPAR signaling pathway).*

5.3.2 Razvrščanje kandidatnih zapisov za nekodirajoče RNA po prioriteti

Z namenom, da bi v razvrščanje kandidatnih genov po prioritetah zajeli še nkRNA, ki so jih mnoge raziskave že povezale z nalaganjem maščobe (McGregor in Choi, 2011), smo zajeli vse gene miRNA iz Genomskega atlasa za nalaganje maščobe in iz mreže

uravnavanja izražanja genov z molekulami miRNA. S pomočjo orodja miRNA SNiPer, ki smo ga razvili, smo izmed teh genov za miRNA poiskali tiste, ki imajo polimorfno regijo *seed*, ki je odgovorna za vezavo na tarče; to so geni *mmu-mir-717* in *mmu-mir-599* pri miši ter *hsa-mir-15b* pri človeku.

Pri razvrščanju kandidatnih genov za miRNA po prioriteti smo si pomagali tudi s prikazom genomske razporeditve miRNA genov. Geni za miRNA so pogosto nakopičeni v gručah (angl. *clustered*) v regijah genomske nestabilnosti in so hkrati izraženi z gostiteljskimi geni, kar nakazuje na to, da imajo skupen mehanizem prepisovanja (Priloga E, Priloga F, Priloga H). Pri genu *hsa-mir-15b* pri človeku smo ugotovili, da se gen nahaja znotraj QTL-a za nalaganje maščobe (BW348_H; body weight QTL), ima protein-kodirajoči gostiteljski gen *SMC4* v smerni orientaciji in *RP11* v protismerni orientaciji ter SNP v regiji *seed* (rs192595529). Kot potencialni biooznačevalec smo izbrali gen *mmu-mir-599* (*Mir599*) pri miši, ker smo želeli vpliv tega gena na nalaganje maščobe preveriti z analizo povezanosti genotipa s fenotipom, za katero imamo na razpolago samo podatke pri miši.

5.4 ANALIZA POVEZAVE GENOTIPA S FENOTIPOM

Z genetskimi asociacijskimi študijami testiramo povezave med genetskimi variacijami in fenotipi, da bi poiskali gene ali odseke na genomu, ki prispevajo k oblikovanju določenega fenotipa. Višja frekvenca alela SNP-ja pri obolelih osebkih lahko pomeni, da testirani SNP poveča tveganje za obolevanjem (Lunetta, 2008). V asociacijskih študijah so najbolj pogosto testirani označevalci SNP-ji.

Analizirali smo povezavo SNP-jev znotraj štirih izbranih kandidatnih genov z lastnostmi nalaganja maščobe. Uporabili smo genotipske in fenotipske podatke pri inbridiranih linijah miši iz podatkovne zbirke MPD. Izkazalo se je, da so vsi štirje geni, ki smo jih izbrali za potencialne biooznačevalce, povezani z nalaganjem maščobe. Analizirali smo tudi povezavo SNP-ja znotraj kandidatnega gena *Mapkap1*, ki ga v postopku razvrščanja kandidatnih genov nismo uvrstili k močnejšim kandidatom. S polimorfizmom v genu *Akt1* je povezanih 95 lastnosti, s polimorfizmom v genu *Ubc* 86 lastnosti, s polimorfizmom v genu *Grb2* 89 lastnosti, s polimorfizmom v *Mir599* 75 lastnosti, s polimorfizmom v genu *Mapkap1* pa 28 lastnosti za nalaganje maščobe.

Razlike v frekvenci alelov v genetskih asociacijskih študijah pri človeku lahko privedejo do lažno pozitivnih povezav med genotipom in fenotipom (Freedman in sod, 2004; Marchini in sod., 2004). Prav tako so omejene genetske asociacijske študije z inbridiranimi linijami miši zaradi strukture populacije in sorodnosti med osebki (Flint in sod., 2005; Peirce in sod., 2008). Pri izboru SNP-jev znotraj kandidatnih genov smo upoštevali frekvence alelov, izbrali smo čim bolj izenačene. Pri kandidatnem genu *Mir599* smo testirali vpliv SNP-ja znotraj regije *seed*, kjer frekvenca alelov ni bila izenačena. V genu *Mir599* sta še dva SNP-ja (rs222270936 in rs37482983), vendar je pri obeh frekvenca

alelov enaka kot pri SNP-ju rs37362582 v regiji *seed*. V prihodnjih raziskavah bi bilo treba opraviti analizo povezanosti genotipa s fenotipom za več SNP-jev znotraj kandidatnih genov.

Za fenotipske podatke smo zajeli 268 lastnosti, ki so povezane z nalaganjem maščobe, to je 76 lastnosti za sestavo telesa, 26 za telesne maščobne blazinice, 166 za telesno težo. Upoštevati bi bilo treba, da na vse analizirane lastnosti ne vplivajo samo geni za debelost, na telesno maso lahko vplivajo tudi geni, ki povišujejo mišično in kostno maso.

Miš je dober modelni organizem za sesalce zaradi poznanega celotnega genoma, inbridiranih linij, enostavne vzreje in vzdrževanja v laboratoriju ter razvitih standardnih postopkov za rokovanje z njim (Cox in Church, 2011). Genom miši je približno 14% manjši od genoma pri človeku, za 80% genov pri miši obstaja ortolog pri človeku in za samo 1% genov pri miši do sedaj še ni identificiranih homolognih genov pri človeku (Waterston in sod., 2002). Miš je zelo pogosto uporabljen kot model za raziskavo biologije in bolezni pri človeku (Blake in sod., 2011), med drugim tudi debelosti in diabetesa tipa 2. Uravnavanje ravni glukoze v krvi je pri obeh organizmih v številnih pogledih podobno, pri obeh kažeta motena toleranca za glukozo in diabetes podobne simptome. Razlika pri obeh organizmih se na primer kaže v ravni lipoproteinov z visoko gostoto (angl. *high density lipoprotein;* HDL), ki je pri miših višja, kar je možen vzrok za odpornost miši proti arteriosklerozi. Gen *ABCA1* je zadolžen za transport holesterola in za produkcijo HDL-ja. Gena *miR-33a* in *miR-33b* ciljata tarče, ki so med drugimi tudi mRNA gena *ABCA1*. Pri miši gena *miR-33b* ni, kar je možen vzrok za višjo raven HDL-ja pri miših (Najafi-Shoushtari in sod., 2010). Kljub razlikam med organizmoma (nekatere so celo pripomogle k razumevanju mehanizmov razvoja bolezni), velja miš za dober model za raziskave debelosti in diabetesa tipa 2 pri človeku (Cox in Church, 2011).

5.5 BIOOZNAČEVALCI ZA NALAGANJE MAŠČOBE

Biooznačevalci so molekularni dejavniki, s katerimi lahko prepoznavamo ali spremljamo normalne ali patološke biološke procese. Uporabni so za ugotavljanje tveganja, prisotnosti ali suma na bolezen, tudi kot vodilo za diagnostične in terapevtske intervencije ter za prilagajanje poteka zdravljenja posamezniku. Biooznačevalce lahko predstavljajo meritve količine proteinov ali peptidov, polimorfizmi DNA, količina molekul mRNA v celicah (Margulies, 2009).

Kot potencialne biooznačevalce za nalaganje maščobe smo predlagali gene *Ubc*, *Akt1*, *Grb2* in *Mir599*, ki smo jih s postopkom integracije genomskih podatkov in z razvrščanjem po prioriteti ocenili kot močnejše kandidatne gene za nalaganje maščobe. Z analizo povezanosti genotipa s fenotipom smo potrdili vpliv izbranih potencialnih biooznačevalcev na lastnosti nalaganja maščobe pri miši.

Gen *Ubc* kodira majhen, zelo ohranjen regulatorni protein ubikvitin, ki je izražen v skoraj vseh tkivih pri evkariontih. Poleg gena *Ubc*, ubikvitin kodirajo še geni *Uba52*, *Rps27s* in *Ubb*. Ubikvitin se veže na proteine in jih označi za razgradnjo. Ubikvitinacija predstavlja potranslacijsko spremembo, ki ima pomembno vlogo pri apoptozi, diferenciaciji, celičnem ciklu, biogenezi ribosomov in drugih pomembnih celičnih funkcijah pri evkariontih (Kimura in Tanaka, 2010). Ubikvitin so v dosedanjih študijah že povezali z nalaganjem maščobe. Raven ubikvitina v plazmi pri človeku je povezana z indeksom telesne mase (BMI) (Chang in sod., 2009). Eksogeno dodajanje ubikvitina v plazmo povzroči zmanjšanje ravni dejavnika tumorske nekroze TNF-*α* (Patel in sod, 2006), ki pa se je izkazal kot biooznačevalec kroničnega vnetja maščobnega tkiva (Moller, 2000; Kern in sod, 2001). Potencialni biooznačevalec za nalaganje maščobe je tudi gen *Akt1*, ki ga uravnava ubikvitin (Yang in sod., 2010). Miši z utišanim genom *Akt1* so odporne na visokokalorično dieto (Min Wan in sod., 2012). Potencialni biooznačevalec *Grb2* kodira protein, ki je izražen v vseh tkivih. Gen *Grb2* vsebuje domeno SH2, ki lahko veže fosforilirane tirozinske ostanke proteinov. Pri podhranjenih miših se je izražanje *Grb2* v skeletnih mišicah znižalo (Liu in sod., 2009). Potencialni biooznačevalec, *Mir599* do sedaj še ni bil povezan z nalaganjem maščobe, nahaja pa se znotraj gostiteljskega gena *VPS13B*, ki je vzročni gen za sindrom Cohen, pri katerem je prisotna tudi debelost (Balikova in sod., 2009).

Pri razvrščanju kandidatnih genov smo zajeli gene na vseh stopnjah procesa razvoja debelosti in nekateri izmed njih so bili zelo visoko ocenjeni kandidatni geni. Do tega je prišlo, ker smo med kandidatne gene za razvrščanje prišteli tudi gene, ki so prvi sosedi v proteinskih interakcijah znanih kandidatnih genov. Na ta način smo omogočili iskanje genov za nalaganje maščobe, ki do sedaj še niso bili povezani s tem fenotipom. V raziskavah v prihodnje bi bilo treba pri že znanih kandidatnih genih zabeležiti tudi to, v kateri stopnji razvoja debelosti so udeleženi in to upoštevati pri analizi.

6 SKLEPI

V delovnih hipotezah smo predvideli, 1) da je z integracijo do sedaj poznanih genetskih vzrokov za nalaganje maščobe ter bioinformacijsko analizo možno identificirati genske mreže in biološke poti, ki jih je možno razviti v biooznačevalce, in, 2) da je kandidatne gene za nalaganje maščobe možno razvrstiti po prioriteti glede na njihovo lokacijo na genomskem prikazu, njihovem številu povezav v genski mreži in glede na njihovo udeleženost v biološki poti. Potrdili smo obe hipotezi.

- Zbrali smo do sedaj poznane kandidatne lokuse za nalaganje maščobe in s prikazom genomske razporeditve lokusov odkrili prekrivanja med njimi. S pomočjo zbranih kandidatnih lokusov smo identificirali tudi kandidatne biološke poti in genske mreže, ki smo jih razvili v potencialne biooznačevalce za nalaganje maščobe.

- Kandidatne gene za nalaganje maščobe smo razvrstili po prioriteti na osnovi genomskih razporeditev, kandidatnih bioloških poti in genskih mrež ter med njimi izbrali najobetavnejše, potencialne biooznačevalce za nalaganje maščobe. Z analizo povezave med genotipom in fenotipom smo potrdili vpliv izbranih kandidatnih genov na lastnosti nalaganja maščobe pri miši.

Rezultati doktorske naloge so:

- prosto dostopna zbirka kandidatnih lokusov za nalaganje maščobe,

- nov pristop za razvoj biooznačevalcev večgenskih bolezni z integracijo genomskih podatkov in bioinformacijsko analizo, ki temelji na identifikaciji prekrivanj lokusov neodvisnih študij, kandidatnih bioloških poteh in genskih mrežah,

- seznam potencialnih biooznačevalcev za nalaganje maščobe, preverjenih z analizo povezave genotipa s fenotipom na modelu miši.

Na podlagi rezultatov raziskovalnega dela bodo možne nadaljnje raziskave genskih osnov nalaganja maščobe. Pristop za razvoj biooznačevalcev bo možno uporabiti tudi za druge kompleksne fenotipe, zato predstavlja pomemben prispevek k znanosti.

7 POVZETEK (SUMMARY)

7.1 POVZETEK

Debelost je večgenska bolezen, ki predstavlja pereč zdravstveni problem. Za debelostjo zbolevajo ljudje vseh starosti, pojavlja pa se tudi pri domačih živalih, kjer pogosteje govorimo o zamaščenosti. Odkritje genetskih osnov nalaganja maščobe lahko pripomore k razumevanju pojava in terapiji debelosti. Količina razpoložljivih genomskih podatkov narašča, hkrati pa se povečujejo zahteve po metodah za njihovo analizo. Vse večjo vlogo pri raziskavah kompleksnih fenotipov pridobivajo sistemski pristopi. Izdelali smo genomski atlas, ki predstavlja centralno spletno mesto genetskih vzrokov za nalaganje maščobe. Uporabili smo primerjalni in integrativni pristop zbiranja lokusov za nalaganje maščobe pri človeku, miši, podgani in govedu ter z vizualizacijo integriranih podatkov dosegli enostaven vpogled v do sedaj poznane lokuse, povezane z nalaganjem maščobe. Izdelali smo prikaze genomske razporeditve lokusov, identificirali kandidatne biološke poti in genske mreže za nalaganje maščobe, ki so bile osnova za razvrščanje kandidatnih genov po prioritetah. Razvili smo bioinformacijski orodji za analizo nekodirajočih kandidatnih genov (miRNA SNiPer in miRNA Viewer). Iz nabora kandidatnih lokusov smo izbrali potencialne biooznačevalce *(Akt1, Ubc, Grb2, Mir599)* in z analizo povezanosti genotipa s fenotipom preverili njihov vpliv na lastnosti nalaganja maščobe pri miši. Razvili smo strategijo za raziskave genetskih vzrokov nalaganja maščobe, ki jo je možno uporabiti tudi za druge kompleksne fenotipe.

7.2 SUMMARY

Obesity is polygenic disease which presents a major health issue. It affects people of all ages as well as domestic animals. The unravelling of genetic bases of fat deposition might help to develop therapeutics and understand the process of fat deposition. The amount of available genomic data and the need for genomic data analysis methods grow. Systemic approaches are becoming important in complex phenotypes research. We created the genomic atlas, which presents the central web resource of genetic causes for fat deposition. The comparative and integrative approach to collect the loci associated with fat deposition in human, mouse, rat and cattle was used. By visualization of the integrated data the insight into known fat deposition loci was enabled. We created genomic views of loci, identified candidate biological pathways and determined genetic networks for fat deposition, which were basis for candidate genes prioritisation. Two bioinformatics tools for analysis of noncoding candidate genes were developed (miRNA SNiPer and miRNA Viewer). From the set of candidate loci we selected potential biomarkers (*Akt1*, *Ubc*, *Grb2*, *Mir599)* and tested their effect on fat deposition traits in mice using analysis of association between genotype and phenotype. We developed a strategy for research of genetic causes for fat deposition. The same approach can be used for analysis of other complex phenotypes.

8 VIRI

Adie E.A., Adams R.R., Evans K.L., Porteous D.J., Pickard B.S. 2006. SUSPECTS: enabling fast and effective prioritization of positional candidates. Bioinformatics, 22: 773–774

Aebersold R., Mann M. 2003. Mass spectrometry-based proteomics. Nature, 422: 198–207

Aerts S., Vilain S., Hu S., Tranchevent L.C., Barriot R., Yan J., Moreau Y., Hassan B.A., Quan X.J. 2009. Integrating computational biology and forward genetics in Drosophila. PLoS Genetics, 5: e1000351

Aggarwal K., Lee K.H. 2003. Functional genomics and proteomics as a foundation for systems biology. Briefings in Functional Genomics & Proteomics, 2: 175–184

Agrawal S., Dimitrova N., Nathan P., Udayakumar K., Lakshmi S.S., Sriram S., Manjusha N., Sengupta U. 2008. T2D-Db: an integrated platform to study the molecular basis of Type 2 diabetes. BMC Genomics, 9: 320

Ahima R.S. 2002. Obesity gene therapy: Slimming immature rats. Gene Therapy, 10: 196–197

Allison D.B., Kaprio J., Korkeila M., Koskenvuo M., Neale M.C., Hayakawa K. 1996. The heritability of body mass index among an international sample of monozygotic twins reared apart. International Journal of Obesity and Related Metabolic Disorders : Journal of The International Association for the Study of Obesity, 20: 501–506

Ambrosius W.T., Lange E.M., Langefeld C.D. 2004. Power for genetic association studies with random allele frequencies and genotype distributions. American Journal of Human Genetics, 74: 683–693

Arcade A., Labourdette A., Falque M., Mangin B., Chardon F., Charcosset A., Joets J. 2004. BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. Bioinformatics, 20: 2324–2326

Auffray C., Imbeaud S., Roux-Rouquié M., Hood L. 2003. From functional genomics to systems biology: concepts and practices. Comptes Rendus Biologies, 326: 879–892

Azuaje F. 2010. Bioinformatics and Biomarker Discovery: "Omic" Data Analysis for Personalized Medicine. Hoboken, N.J, John Wiley and Sons: 248 str.

Bader G.D., Cary M.P., Sander C. 2006. Pathguide: a pathway resource list. Nucleic Acids Research, 34: D504– D506

Bader S., Kühner S., Gavin A.C. 2008. Interaction networks for systems biology. FEBS Letters, 582: 1220–1224

Balikova I., Lehesjoki A.-E., De Ravel T.J.L., Thienpont B., Chandler K.E., Clayton-Smith J., Träskelin A.-L., Fryns J.-P., Vermeesch J.R. 2009. Deletions in the VPS13B (COH1) gene as a cause of Cohen syndrome. Human Mutation, 30: E845– E854

Barabási A.-L. 2007. Network Medicine — From Obesity to the "Diseasome". The New England Journal of Medicine, 357: 404–407

Barabási A.L., Oltvai Z.N. 2004. Network biology: understanding the cell's functional organization. Nature Reviews. Genetics, 5: 101–113

Baran J., Gerner M., Haeussler M., Nenadic G., Bergman C.M. 2011. pubmed2ensembl: a resource for mining the biological literature on genes. PLoS One, 6: e24716

Bauer-Mehren A., Bundschus M., Rautschka M., Mayer M.A., Sanz F., Furlong L.I. 2011. Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. PLoS One, 6: e20284

Becker K.G., Hosack D.A., Dennis G., Lempicki R.A., Bright T.J., Cheadle C., Engel J. 2003. PubMatrix: a tool for multiplex literature mining. BMC Bioinformatics, 4: 61

Becker K.G., White S.L., Muller J., Engel J. 2000. BBID: the biological biochemical image database. Bioinformatics, 16: 745–746

Bell C.G., Walley A.J., Froguel P. 2005. The genetics of human obesity. Nature Reviews. Genetics, 6: 221–234

Bertram L., McQueen M.B., Mullin K., Blacker D., Tanzi R.E. 2007. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nature Genetics, 39: 17–23

Blake J.A., Bult C.J., Kadin J.A., Richardson J.E., Eppig J.T. 2011. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. Nucleic Acids Research, 39: D842– D848

Botstein D., Risch N. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genetics, 33, Suppl: 228–237

Bouchard C., Pérusse L. 1996. Current status of the human obesity gene map. Obesity Research, 4: 81–90

Burren O.S., Adlem E.C., Achuthan P., Christensen M., Coulson R.M., Todd J.A. 2011. T1DBase: update 2011, organization and presentation of large-scale data sets for type 1 diabetes research. Nucleic Acids Research, 39: D997–D1001

Cannistraci C. V, Ogorevc J., Zorc M., Ravasi T., Dovč P., Kunej T. 2013. Pivotal role of the muscle-contraction pathway in cryptorchidism and evidence for genomic connections with cardiomyopathy pathways in RASopathies. BMC Medical Genomics, 6: 5

Carlson J.M., Doyle J. 2002. Complexity and robustness. Proceedings of the National Academy of Sciences of the United States of America, 99, Suppl 1: 2538–2545

Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M., Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impiombato A., Apweiler R., Aturaliya R.N., Bailey T.L., Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M., Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R., Crowe M.L., Dalla E., Dalrymple B.P., De Bono B., Della Gatta G., Di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G., Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M., Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E., Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N., Hill D., Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H., Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I. V, Lareau L.F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K., Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G., Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M., Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Schönbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y., Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B., Sperling S., Stupka E., Sugiura K., Sultana R., Takenaka Y., Taki K., Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A., Ueda H.R., Van Nimwegen E., Verardo R., Wei C.L., Yagi K., Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C., Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J., Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T., Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N., Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S., Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y., Consortium F., Group) R.G.E.R.G. and G.S.G. (Genome N.P.C.) 2005. The transcriptional landscape of the mammalian genome. Science, 309: 1559–1563

Cerami E.G., Gross B.E., Demir E., Rodchenkov I., Babur O., Anwar N., Schultz N., Bader G.D., Sander C. 2011. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Research, 39: D685– D690

Chang T.-L., Chang C.-J., Lee W.-Y., Lin M.-N., Huang Y.-W., Fan K. 2009. The roles of ubiquitin and 26S proteasome in human obesity. Metabolism: Clinical and Experimental, 58: 1643–1648

Chatr-Aryamontri A., Breitkreutz B.J., Heinicke S., Boucher L., Winter A., Stark C., Nixon J., Ramage L., Kolas N., O'Donnell L., Reguly T., Breitkreutz A., Sellam A., Chen D., Chang C., Rust J., Livstone M., Oughtred R., Dolinski K., Tyers M. 2013. The BioGRID interaction database: 2013 update. Nucleic Acids Research, 41: D816–D823

Chen J., Bardes E.E., Aronow B.J., Jegga A.G. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Research, 37: W305– W311

Chen L., Liu H., Friedman C. 2005. Gene name ambiguity of eukaryotic nomenclatures. Bioinformatics, 21: 248–256

Cheng D., Knox C., Young N., Stothard P., Damaraju S., Wishart D.S. 2008. PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. Nucleic Acids Research, 36: W399–W405

Cherry J.M., Hong E.L., Amundsen C., Balakrishnan R., Binkley G., Chan E.T., Christie K.R., Costanzo M.C., Dwight S.S., Engel S.R., Fisk D.G., Hirschman J.E., Hitz B.C., Karra K., Krieger C.J., Miyasato S.R., Nash R.S., Park J., Skrzypek M.S., Simison M., Weng S., Wong E.D. 2012. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Research, 40: D700– D705

Cheung C.C., Clifton D.K., Steiner R.A. 1997. Proopiomelanocortin neurons are direct targets for leptin in the hypothalamus. Endocrinology, 138: 4489–4492

Chuang H.Y., Hofree M., Ideker T. 2010. A decade of systems biology. Annual Review of Cell and Developmental Biology, 26: 721–744

Chuang H.Y., Lee E., Liu Y.T., Lee D., Ideker T. 2007. Network-based classification of breast cancer metastasis. Molecular Systems Biology, 3: 140

Collins F.S., Guyer M.S., Charkravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. Science, 278: 1580–1581

Considine R. V, Sinha M.K., Heiman M.L., Kriauciunas A., Stephens T.W., Nyce M.R., Ohannesian J.P., Marco C.C., McKee L.J., Bauer T.L. 1996. Serum immunoreactive-leptin concentrations in normal-weight and obese humans. The New England Journal of Medicine, 334: 292–295

Consortium E.P. 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biology, 9: e1001046

Consortium U. 2013. Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Research, 41: D43– D47

Cotton R.G., McKusick V., Scriver C.R. 1998. The HUGO Mutation Database Initiative. Science, 279: 10–11

Cox R.D., Church C.D. 2011. Mouse models and the interpretation of human GWAS in type 2 diabetes and obesity. Disease Models & Mechanisms, 4: 155–164

Croft D., O'Kelly G., Wu G., Haw R., Gillespie M., Matthews L., Caudy M., Garapati P., Gopinath G., Jassal B., Jupe S., Kalatskaya I., Mahajan S., May B., Ndegwa N., Schmidt E., Shamovsky V., Yung C., Birney E., Hermjakob H., D'Eustachio P., Stein L. 2011. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Research, 39: D691– D697

Davis A.P., Murphy C.G., Johnson R., Lay J.M., Lennon-Hopkins K., Saraceni-Richards C., Sciaky D., King B.L., Rosenstein M.C., Wiegers T.C., Mattingly C.J. 2013. The Comparative Toxicogenomics Database: update 2013. Nucleic Acids Research, 41: D1104– D1114

Dinkel H., Chica C., Via A., Gould C.M., Jensen L.J., Gibson T.J., Diella F. 2011. Phospho.ELM: a database of phosphorylation sites--update 2011. Nucleic Acids Research, 39: D261– D267

Dong Y.-Y., Wang Z.-H., Liu Q.-J., Wang Y.-X. 2011. Understanding centrality-lethality rule in yeast Protein-Protein Interaction network by topological property. V: Information Technology and Artificial Intelligence Conference (ITAIC), 6th IEEE Joint International, Chongqing, 20-22. avg. 2011. 116–118

Drysdale R., Consortium F. 2008. FlyBase : a database for the Drosophila research community. Methods in Molecular Biology, 420: 45–59

Dwight S.S., Balakrishnan R., Christie K.R., Costanzo M.C., Dolinski K., Engel S.R., Feierbach B., Fisk D.G., Hirschman J., Hong E.L., Issel-Tarver L., Nash R.S., Sethuraman A., Starr B., Theesfeld C.L., Andrada R., Binkley G., Dong Q., Lane C., Schroeder M., Weng S., Botstein D., Cherry J.M. 2004. Saccharomyces genome database: underlying principles and organisation. Briefings in Bioinformatics, 5: 9–22

Easton D.F., Eeles R.A. 2008. Genome-wide association studies in cancer. Human Molecular Genetics, 17: R109– R115

Farooqi I.S., O'Rahilly S. 2004. Monogenic human obesity syndromes. Recent Progress in Hormone Research, 59: 409–424

Fischer J., Koch L., Emmerling C., Vierkotten J., Peters T., Brüning J.C., Rüther U. 2009. Inactivation of the Fto gene protects from obesity. Nature, 458: 894–898

Flicek P., Ahmed I., Amode M.R., Barrell D., Beal K., Brent S., Carvalho-Silva D., Clapham P., Coates G., Fairley S., Fitzgerald S., Gil L., García-Girón C., Gordon L., Hourlier T., Hunt S., Juettemann T., Kähäri A.K., Keenan S., Komorowska M., Kulesha E., Longden I., Maurel T., McLaren W.M., Muffato M., Nag R., Overduin B., Pignatelli M., Pritchard B., Pritchard E., Riat H.S., Ritchie G.R., Ruffier M., Schuster M., Sheppard D., Sobral D., Taylor K., Thormann A., Trevanion S., White S., Wilder S.P., Aken B.L., Birney E., Cunningham F., Dunham I., Harrow J., Herrero J., Hubbard T.J., Johnson N., Kinsella R., Parker A., Spudich G., Yates A., Zadissa A., Searle S.M. 2013. Ensembl 2013. Nucleic Acids Research, 41: D48–D55

Flint J., Valdar W., Shifman S., Mott R. 2005. Strategies for mapping and cloning quantitative trait genes in rodents. Nature Reviews. Genetics, 6: 271–286

Franceschini A., Szklarczyk D., Frankild S., Kuhn M., Simonovic M., Roth A., Lin J., Minguez P., Bork P., Von Mering C., Jensen L.J. 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research, 41: D808– D815

Freedman M.L., Reich D., Penney K.L., McDonald G.J., Mignault A.A., Patterson N., Gabriel S.B., Topol E.J., Smoller J.W., Pato C.N., Pato M.T., Petryshen T.L., Kolonel L.N., Lander E.S., Sklar P., Henderson B., Hirschhorn J.N., Altshuler D. 2004. Assessing the impact of population stratification on genetic association studies. Nature Genetics, 36: 388–393

Gao S., Roberts H.K., Wang X. 2012. Cross tissue trait-pathway network reveals the importance of oxidative stress and inflammation pathways in obesity-induced diabetes in mouse. PloS One, 7: e44544

Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research, 31: 3784–3788

Gavin A.C., Aloy P., Grandi P., Krause R., Boesche M., Marzioch M., Rau C., Jensen L.J., Bastuck S., Dümpelfeld B., Edelmann A., Heurtier M.A., Hoffman V., Hoefert C., Klein K., Hudak M., Michon A.M., Schelder M., Schirle M., Remor M., Rudi T., Hooper S., Bauer A., Bouwmeester T., Casari G., Drewes G., Neubauer G., Rick J.M., Kuster B., Bork P., Russell R.B., Superti-Furga G. 2006. Proteome survey reveals modularity of the yeast cell machinery. Nature, 440: 631–636

Gefen A., Cohen R., Birk O.S. 2010. Syndrome to gene (S2G): in-silico identification of candidate genes for human diseases. Human Mutation, 31: 229–236

George R.A., Liu J.Y., Feng L.L., Bryson-Richardson R.J., Fatkin D., Wouters M.A. 2006. Analysis of protein sequence and interaction data for candidate disease gene prediction. Nucleic Acids Research, 34: e130

Gibson G. 2009. Decanalization and the origin of complex disease. Nature reviews. Genetics, 10: 134–140

Glez-Peña D., Reboiro-Jato M., Domínguez R., Gómez-López G., Pisano D.G., Fdez-Riverola F. 2010. PathJam: a new service for integrating biological pathway information. Journal of Integrative Bioinformatics, 7: 147

Gluckman P.D., Hanson M.A. 2008. Developmental and epigenetic pathways to obesity: an evolutionary-developmental perspective. International journal of obesity: journal of the International Association for the Study of Obesity, 32, Suppl 7: S62–S71

Gnad F., Gunawardena J., Mann M. 2011. PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Research, 39: D253– D260

Goel R., Harsha H.C., Pandey A., Prasad T.S. 2012. Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis. Molecular BioSystems, 8: 453–463

Gray K.A., Daugherty L.C., Gordon S.M., Seal R.L., Wright M.W., Bruford E.A. 2013. Genenames.org: the HGNC resources in 2013. Nucleic Acids Research, 41: D545–D552

Griffiths-Jones S., Grocock R.J., Van Dongen S., Bateman A., Enright A.J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Research, 34: D140– D144

Guerre-Millo M. 2002. Adipose tissue hormones. Journal of Endocrinological Investigation, 25: 855–861

Guo H., Ingolia N.T., Weissman J.S., Bartel D.P. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature, 466: 835–840

Haemer M.A., Huang T.T., Daniels S.R. 2009. The effect of neurohormonal factors, epigenetic factors, and gut microbiota on risk of obesity. Preventing Chronic Disease, 6: A96

Hahn M.W., Kern A.D. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Molecular Biology and Evolution, 22: 803–806

Hajer G.R., Van der Graaf Y., Olijhoek J.K., Edlinger M., Visseren F.L.J. 2007. Low plasma levels of adiponectin are associated with low risk for future cardiovascular events in patients with clinical evident vascular disease. American Heart Journal, 154: 750.e1–7

Hamid J.S., Hu P., Roslin N.M., Ling V., Greenwood C.M.T., Beyene J. 2009. Data integration in genetics and genomics: methods and challenges. Human Genomics and Proteomics, HGP: 2009

Han J.D., Bertin N., Hao T., Goldberg D.S., Berriz G.F., Zhang L. V, Dupuy D., Walhout A.J., Cusick M.E., Roth F.P., Vidal M. 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature, 430: 88–93

Hardy J., Singleton A. 2009. Genomewide association studies and human disease. The New England Journal of Medicine, 360: 1759–1768

Helmchen L.A., Henderson R.M. 2004. Changes in the distribution of body mass index of white US men, 1890-2000. Annals of Human Biology, 31: 174–181

Herrera B.M., Keildson S., Lindgren C.M. 2011. Genetics and epigenetics of obesity. Maturitas, 69: 41–49

Hindorff L.A., Sethupathy P., Junkins H.A., Ramos E.M., Mehta J.P., Collins F.S., Manolio T.A. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of America, 106: 9362–9367

Hirsch J., Batchelor B. 1976. Adipose tissue cellularity in human obesity. Clinics in Endocrinology and Metabolism, 5: 299–311

Hirschhorn J.N. 2009. Genomewide association studies--illuminating biologic pathways. The New England Journal of Medicine, 360: 1699–1701

Hirschhorn J.N., Daly M.J. 2005. Genome-wide association studies for common diseases and complex traits. Nature Reviews. Genetics, 6: 95–108

Hirschman J., Berardini T.Z., Drabkin H.J., Howe D. 2010. A MOD(ern) perspective on literature curation. Molecular Genetics and Genomics : MGG, 283: 415–425

Houpt K.A., Houpt T.R., Pond W.G. 1979. The pig as a model for the study of obesity and of control of food intake: a review. The Yale Journal of Biology and Medicine, 52: 307–329

Hsu S.D., Lin F.M., Wu W.Y., Liang C., Huang W.C., Chan W.L., Tsai W.T., Chen G.Z., Lee C.J., Chiu C.M., Chien C.H., Wu M.C., Huang C.Y., Tsou A.P., Huang H.D. 2011. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Research, 39: D163– D169

Hu Z.L., Fritz E.R., Reecy J.M. 2007. AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. Nucleic Acids Research, 35: D604– D609

Huang da W., Sherman B.T., Tan Q., Kir J., Liu D., Bryant D., Guo Y., Stephens R., Baseler M.W., Lane H.C., Lempicki R.A. 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Research, 35: W169– W175

Hutz J.E., Kraja A.T., McLeod H.L., Province M.A. 2008. CANDID: a flexible method for prioritizing candidate genes for complex human traits. Genetic Epidemiology, 32: 779–790

Ideker T., Galitski T., Hood L. 2001. A new approach to decoding life: systems biology. Annual Review of Genomics and Human Genetics, 2: 343–372

Immervoll T., Wjst M. 1999. Current status of the Asthma and Allergy Database. Nucleic Acids Research, 27: 213–214

Isserlin R., El-Badrawi R.A., Bader G.D. 2011. The Biomolecular Interaction Network Database in PSI-MI 2.5. Database: The Journal of Biological Databases and Curation, 2011: baq037

Jain K.K. 2010. The Handbook of Biomarkers. New Jersey, Humana Press/Springer: 488 str.

Jia P., Kao C.-F., Kuo P.-H., Zhao Z. 2011. A comprehensive network and pathway analysis of candidate genes in major depressive disorder. BMC Systems Biology, 5, Suppl 3: S12

Jia P., Sun J., Guo A.Y., Zhao Z. 2010. SZGR: a comprehensive schizophrenia gene resource. Molecular Psychiatry, 15: 453–462

Joy M.P., Brock A., Ingber D.E., Huang S. 2005. High-betweenness proteins in the yeast protein interaction network. Journal of Biomedicine & Biotechnology, 96–103

Kanehisa M., Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research, 28: 27–30

Kann M.G. 2007. Protein interactions and disease: computational approaches to uncover the etiology of diseases. Briefings in Bioinformatics, 8: 333–346

Kern P.A., Ranganathan S., Li C., Wood L., Ranganathan G. 2001. Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance. American journal of physiology. Endocrinology and Metabolism, 280: E745– E751

Kerrien S., Aranda B., Breuza L., Bridge A., Broackes-Carter F., Chen C., Duesbury M., Dumousseau M., Feuermann M., Hinz U., Jandrasits C., Jimenez R.C., Khadake J., Mahadevan U., Masson P., Pedruzzi I., Pfeiffenberger E., Porras P., Raghunath A., Roechert B., Orchard S., Hermjakob H. 2012. The IntAct molecular interaction database in 2012. Nucleic Acids Research, 40: D841– D846

Kersey P., Apweiler R. 2006. Linking publication, gene and protein data. Nature Cell Biology, 8: 1183–1189

Kershaw E.E., Flier J.S. 2004. Adipose tissue as an endocrine organ. The Journal of Clinical Endocrinology and Metabolism, 89: 2548–2556

Keseler I.M., Collado-Vides J., Santos-Zavaleta A., Peralta-Gil M., Gama-Castro S., Muñiz-Rascado L., Bonavides-Martinez C., Paley S., Krummenacker M., Altman T., Kaipa P., Spaulding A., Pacheco J., Latendresse M., Fulcher C., Sarker M., Shearer A.G., Mackie A., Paulsen I., Gunsalus R.P., Karp P.D. 2011. EcoCyc: a comprehensive database of Escherichia coli biology. Nucleic Acids Research, 39: D583– D590

Keshava Prasad T.S., Goel R., Kandasamy K., Keerthikumar S., Kumar S., Mathivanan S., Telikicherla D., Raju R., Shafreen B., Venugopal A., Balakrishnan L., Marimuthu A., Banerjee S., Somanathan D.S., Sebastian A., Rani S., Ray S., Harrys Kishore C.J., Kanth S., Ahmed M., Kashyap M.K., Mohmood R., Ramachandra Y.L., Krishna V., Rahiman B.A., Mohan S., Ranganathan P., Ramabadran S., Chaerkady R., Pandey A. 2009. Human Protein Reference Database--2009 update. Nucleic Acids Research, 37: D767– D772

Khaodhiar L., McCowen K.C., Blackburn G.L. 1999. Obesity and its comorbid conditions. Clinical Cornerstone, 2: 17–31

Kim S., Kon M., DeLisi C. 2012. Pathway-based classification of cancer subtypes. Biology Direct, 7: 21

Kimura Y., Tanaka K. 2010. Regulatory mechanisms involved in the control of ubiquitin homeostasis. Journal of Biochemistry, 147: 793–798

Kitano H. 2002. Systems biology: a brief overview. Science, 295: 1662–1664

Klimentidis Y.C., Beasley T.M., Lin H.Y., Murati G., Glass G.E., Guyton M., Newton W., Jorgensen M., Heymsfield S.B., Kemnitz J., Fairbanks L., Allison D.B. 2011. Canaries in the coal mine: a cross-species analysis of the plurality of obesity epidemics. Proceedings. Biological Sciences / The Royal Society, 278: 1626–1632

Knox C., Law V., Jewison T., Liu P., Ly S., Frolkis A., Pon A., Banco K., Mak C., Neveu V., Djoumbou Y., Eisner R., Guo A.C., Wishart D.S. 2011. DrugBank 3.0: a comprehensive resource for “omics” research on drugs. Nucleic Acids Research, 39: D1035– D1041

Kohler S., Bauer S., Horn D., Robinson P.N. 2008. Walking the interactome for prioritization of candidate disease genes. American Journal of Human Genetics, 82: 949–958

Kopelman P. 2007. Health risks associated with overweight and obesity. Obesity reviews: an official journal of the International Association for the Study of Obesity, Suppl 1: 13-17

Kozomara A., Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research, 39: D152– D157

Krull M., Voss N., Choi C., Pistor S., Potapov A., Wingender E. 2003. TRANSPATH: an integrated database on signal transduction and a tool for array analysis. Nucleic Acids Research, 31: 97–100

Kuhn M., Szklarczyk D., Franceschini A., Von Mering C., Jensen L.J., Bork P. 2012. STITCH 3: zooming in on protein-chemical interactions. Nucleic Acids Research, 40: D876– D880

Kunej T., Jevšinek Skok D., Horvat S., Dovč P., Jiang Z. 2010. The glypican 3-hosted murine mir717 gene: sequence conservation, seed region polymorphisms and putative targets. International Journal of Biological Sciences, 6: 769–772

Kunej T., Jevšinek Skok D., Zorc M., Ogrinc A., Michal J.J., Kovač M., Jiang Z. 2012. Obesity Gene Atlas in Mammals. Journal of Genomics, 1: 45–55

Küntzer J., Maisel D., Lenhof H.P., Klostermann S., Burtscher H. 2011. The Roche Cancer Genome Database 2.0. BMC medical genomics, 4: 43

De la Cruz N., Bromberg S., Pasko D., Shimoyama M., Twigger S., Chen J., Chen C.F., Fan C., Foote C., Gopinath G.R., Harris G., Hughes A., Ji Y., Jin W., Li D., Mathis J., Nenasheva N., Nie J., Nigam R., Petri V., Reilly D., Wang W., Wu W., Zuniga-Meyer A., Zhao L., Kwitek A., Tonellato P., Jacob H. 2005. The Rat Genome Database (RGD): developments towards a phenome database. Nucleic Acids Research, 33: D485– D491

Laplante M., Horvat S., Festuccia W.T., Birsoy K., Prevorsek Z., Efeyan A., Sabatini D.M. 2012. DEPTOR cell-autonomously promotes adipogenesis, and its expression is associated with obesity. Cell Metabolism, 16: 202–212

Laulederkind S.J.F., Shimoyama M., Hayman G.T., Lowry T.F., Nigam R., Petri V., Smith J.R., Wang S.-J., De Pons J., Kowalski G., Liu W., Rood W., Munzenmaier D.H., Dwinell M.R., Twigger S.N., Jacob H.J. 2011. The Rat Genome Database curation tool suite: a set of optimized software tools enabling efficient acquisition, organization, and presentation of biological data. Database: The Journal of Biological Databases and Curation, 2011: bar002

Leclerc R.D. 2008. Survival of the sparsest: robust gene networks are parsimonious. Molecular Systems Biology, 4: 213

Lefebvre C., Rajbhandari P., Alvarez M.J., Bandaru P., Lim W.K., Sato M., Wang K., Sumazin P., Kustagi M., Bisikirska B.C., Basso K., Beltrao P., Krogan N., Gautier J., Dalla-Favera R., Califano A. 2010. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Molecular Systems Biology, 6: 377

Lettre G., Rioux J.D. 2008. Autoimmune diseases: insights from genome-wide association studies. Human Molecular Genetics, 17: R116– R121

Levy S.F., Siegal M.L. 2008. Network hubs buffer environmental variation in Saccharomyces cerevisiae. PLoS Biology, 6: e264

Lewis B.P., Burge C.B., Bartel D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell, 120: 15–20

Lewis C.M., Knight J. 2012. Introduction to genetic association studies. Cold Spring Harbor Protocols, 2012: 297–306

Li C.-Y., Mao X., Wei L. 2008. Genes and (common) pathways underlying drug addiction. PLoS Computational Biology, 4: e2

Li G., Mobbs C. V, Scarpace P.J. 2003. Central pro-opiomelanocortin gene delivery results in hypophagia, reduced visceral adiposity, and improved insulin sensitivity in genetically obese Zucker rats. Diabetes, 52: 1951–1957

Liang H., Li W.H. 2007. MicroRNA regulation of human protein protein interaction network. RNA, 13: 1402–1408

Licata L., Briganti L., Peluso D., Perfetto L., Iannuccelli M., Galeota E., Sacco F., Palma A., Nardozza A.P., Santonico E., Castagnoli L., Cesareni G. 2012. MINT, the molecular interaction database: 2012 update. Nucleic Acids Research, 40: D857–D861

Lim D., Kim N.K., Park H.S., Lee S.H., Cho Y.M., Oh S.J., Kim T.H., Kim H. 2011. Identification of candidate genes related to bovine marbling using protein-protein interaction networks. International Journal of Biological Sciences, 7: 992–1002

Lin C.Y., Chin C.H., Wu H.H., Chen S.H., Ho C.W., Ko M.T. 2008. Hubba: hub objects analyzer--a framework of interactome hubs identification for network biology. Nucleic Acids Research, 36: W438– W443

Linding R., Jensen L.J., Pasculescu A., Olhovsky M., Colwill K., Bork P., Yaffe M.B., Pawson T. 2008. NetworKIN: a resource for exploring cellular phosphorylation networks. Nucleic Acids Research, 36: D695– D699

Liu X., Liu M., Zhang J., Bai X., Ramos F., Van Remmen H., Richardson A., Liu F.-Y., Dong L.Q., Liu F. 2009. Downregulation of Grb2 contributes to the insulin-sensitizing effect of calorie restriction. American journal of physiology. Endocrinology and Metabolism, 296: E1067– E1075

Liu Y.J., Guo Y.F., Zhang L.S., Pei Y.F., Yu N., Yu P., Papasian C.J., Deng H.W. 2010. Biological pathway-based genome-wide association analysis identified the vasoactive intestinal peptide (VIP) pathway important for obesity. Obesity, 18: 2339–2346

Loos R.J., Schadt E.E. 2012. This I believe: gaining new insights through integrating "old" data. Frontiers in Genetics, 3: 137

Lu Z. 2011. PubMed and beyond: a survey of web tools for searching biomedical literature. Database: The Journal of Biological Databases and Curation, baq036

Lunetta K.L. 2008. Genetic association studies. Circulation, 118: 96–101

Ma H.W., Zeng A.P. 2003. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics, 19: 1423–1430

Maddatu T.P., Grubb S.C., Bult C.J., Bogue M.A. 2012. Mouse Phenome Database (MPD). Nucleic Acids Research, 40: D887– D894

Marchini J., Cardon L.R., Phillips M.S., Donnelly P. 2004. The effects of human population structure on large genetic association studies. Nature Genetics, 36: 512–517

Margulies K.B. 2009. MicroRNAs as novel myocardial biomarkers. Clinical Chemistry, 55: 1897–1899

Martin A., Ochagavia M.E., Rabasa L.C., Miranda J., Fernandez-de-Cossio J., Bringas R. 2010. BisoGenet: a new tool for gene network building, visualization and analysis. BMC Bioinformatics, 11: 91

Masoudi-Nejad A., Meshkin A., Haji-Eghrari B., Bidkhori G. 2012. Candidate gene prioritization. Molecular Genetics and Genomics : MGG, 287: 679–698

Masseroli M., Galati O., Manzotti M., Gibert K., Pinciroli F. 2005. Inherited disorder phenotypes: controlled annotation and statistical analysis for knowledge mining from gene lists. BMC Bioinformatics, 6, Suppl 4: S18

Mathivanan S., Periaswamy B., Gandhi T.K., Kandasamy K., Suresh S., Mohmood R., Ramachandra Y.L., Pandey A. 2006. An evaluation of human protein-protein interaction data in the public domain. BMC Bioinformatics, 7, Suppl 5: S19

Mayeux R. 2004. Biomarkers: potential uses and limitations. NeuroRx: the journal of the American Society for Experimental NeuroTherapeutics, 1: 182–188

McGregor R.A., Choi M.S. 2011. microRNAs in the regulation of adipogenesis and obesity. Current Molecular Medicine, 11: 304–316

Melnikova I., Wages D. 2006. Anti-obesity therapies. Nature reviews. Drug Discovery, 5: 369–370

Merelli I., Viti F., Milanesi L. 2012. IBDsite: a Galaxy-interacting, integrative database for supporting inflammatory bowel disease high throughput data analysis. BMC Bioinformatics, 13, Suppl 1: S5

Meyer L.R., Zweig A.S., Hinrichs A.S., Karolchik D., Kuhn R.M., Wong M., Sloan C.A., Rosenbloom K.R., Roe G., Rhead B., Raney B.J., Pohl A., Malladi V.S., Li C.H., Lee B.T., Learned K., Kirkup V., Hsu F., Heitner S., Harte R.A., Haeussler M., Guruvadoo L., Goldman M., Giardine B.M., Fujita P.A., Dreszer T.R., Diekhans M., Cline M.S., Clawson H., Barber G.P., Haussler D., Kent W.J. 2013. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Research, 41: D64–D69

Mi H., Muruganujan A., Thomas P.D. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Research, 41: D377–D386

Milhavet F., Cuisset L., Hoffman H.M., Slim R., El-Shanti H., Aksentijevich I., Lesage S., Waterham H., Wise C., Sarrauste de Menthiere C., Touitou I. 2008. The infevers autoinflammatory mutation online registry: update with new genes and functions. Human Mutation, 29: 803–808

Miller K., O'Neill A.J., Chopra I. 2004. Escherichia coli mutators present an enhanced risk for emergence of antibiotic resistance during urinary tract infections. Antimicrobial Agents and Chemotherapy, 48: 23–29

Miller M.L., Jensen L.J., Diella F., Jørgensen C., Tinti M., Li L., Hsiung M., Parker S.A., Bordeaux J., Sicheritz-Ponten T., Olhovsky M., Pasculescu A., Alexander J., Knapp S., Blom N., Bork P., Li S., Cesareni G., Pawson T., Turk B.E., Yaffe M.B., Brunak S., Linding R. 2008. Linear motif atlas for phosphorylation-dependent signaling. Science Signaling, 1: ra2

Mitchell J.A., Aronson A.R., Mork J.G., Folk L.C., Humphrey S.M., Ward J.M. 2003. Gene indexing: characterization and analysis of NLM's GeneRIFs. AMIA Annual Symposium Proceedings, 460–464

Mohlke K.L., Boehnke M., Abecasis G.R. 2008. Metabolic and cardiovascular traits: an abundance of recently identified common genetic variants. Human Molecular Genetics, 17: R102– R108

Moller D.E. 2000. Potential Role of TNF-α in the Pathogenesis of Insulin Resistance and Type 2 Diabetes. Trends in Endocrinology and Metabolism: TEM, 11: 212–217

Mosca E., Alfieri R., Merelli I., Viti F., Calabria A., Milanesi L. 2010. A multilevel data integration resource for breast cancer study. BMC Systems Biology, 4: 76

Motulsky A.G. 2006. Genetics of complex diseases. Journal of Zhejiang University. Science. B, 7: 167–168

Najafi-Shoushtari S.H., Kristo F., Li Y., Shioda T., Cohen D.E., Gerszten R.E., Näär A.M. 2010. MicroRNA-33 and the SREBP host genes cooperate to control cholesterol homeostasis. Science, 328: 1566–1569

Naylor S. 2003. Biomarkers: current perspectives and future prospects. Expert Review of Molecular Diagnostics, 3: 525–529

Nedergaard J., Bengtsson T., Cannon B. 2007. Unexpected evidence for active brown adipose tissue in adult humans. American journal of physiology. Endocrinology and Metabolism, 293: E444– E452

Newnham J.P., Pennell C.E., Lye S.J., Rampono J., Challis J.R. 2009. Early life origins of obesity. Obstetrics and Gynecology Clinics of North America, 36: 227–244

O'Connor T.P., Crystal R.G. 2006. Genetic medicines: treatment strategies for hereditary disorders. Nature Reviews. Genetics, 7: 261–276

Ochagavía M.E., Miranda J., Nazábal M., Martin A., Novoa L.I., Bringas R., Fernández-DE-Cossío J., Camacho H. 2011. A methodology based on molecular interactions and pathways to find candidate genes associated to diseases: its application to schizophrenia and Alzheimer's disease. Journal of Bioinformatics and Computational Biology, 9: 541–557

Ogorevc J., Dovč P., Kunej T. 2011. Comparative genomics approach to identify candidate genetic loci for male fertility. Reproduction in Domestic Animals = Zuchthygiene, 46: 229–239

Ogorevc J., Kunej T., Dovč P. 2008. An integrated map of cattle candidate genes for mastitis. Acta Agriculturae Slovenica, suplement, 2: 85–91

Ogorevc J., Kunej T., Razpet A., Dovč P. 2009. Database of cattle candidate genes and genetic markers for milk production and mastitis. Animal Genetics, 40: 832–851

Oostlander A.E., Meijer G.A., Ylstra B. 2004. Microarray-based comparative genomic hybridization and its applications in human genetics. Clinical Genetics, 66: 488–495

Osborne J.D., Flatow J., Holko M., Lin S.M., Kibbe W.A., Zhu L.J., Danila M.I., Feng G., Chisholm R.L. 2009. Annotating the human genome with Disease Ontology. BMC Genomics, 10, Suppl, 1: S6

Oti M., Snel B., Huynen M.A., Brunner H.G. 2006. Predicting disease genes using protein-protein interactions. Journal of Medical Genetics, 43: 691–698

Ozgür A., Vu T., Erkan G., Radev D.R. 2008. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. Bioinformatics, 24: i277–i285

Park Y.K., Bang O.S., Cha M.H., Kim J., Cole J.W., Lee D., Kim Y.J. 2011. SigCS base: an integrated genetic information resource for human cerebral stroke. BMC Systems Biology, 5, Suppl 2: S10

Parkinson H., Sarkans U., Kolesnikov N., Abeygunawardena N., Burdett T., Dylag M., Emam I., Farne A., Hastings E., Holloway E., Kurbatova N., Lukk M., Malone J., Mani R., Pilicheva E., Rustici G., Sharma A., Williams E., Adamusiak T., Brandizi M., Sklyar N., Brazma A. 2011. ArrayExpress update--an archive of microarray and high-throughput sequencing-based functional genomics experiments. Nucleic Acids Research, 39: D1002–D1004

Patel M.B., Proctor K.G., Majetschak M. 2006. Extracellular ubiquitin increases in packed red blood cell units during storage. The Journal of Surgical Research, 135: 226–232

Pavlopoulos G.A., Secrier M., Moschopoulos C.N., Soldatos T.G., Kossida S., Aerts J., Schneider R., Bagos P.G. 2011. Using graph theory to analyze biological networks. BioData Mining, 4: 10

Peirce J.L., Broman K.W., Lu L., Chesler E.J., Zhou G., Airey D.C., Birmingham A.E., Williams R.W. 2008. Genome Reshuffling for Advanced Intercross Permutation (GRAIP): simulation and permutation for advanced intercross population analysis. PloS One, 3: e1977

Portales-Casamar E., Thongjuea S., Kwon A.T., Arenillas D., Zhao X., Valen E., Yusuf D., Lenhard B., Wasserman W.W., Sandelin A. 2010. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Research, 38: D105–D110

Queitsch C., Carlson K.D., Girirajan S. 2012. Lessons from model organisms: phenotypic robustness and missing heritability in complex disease. PLoS Genetics, 8: e1003041

Radivojac P., Peng K., Clark W.T., Peters B.J., Mohan A., Boyle S.M., Mooney S.D. 2008. An integrated approach to inferring gene-disease associations in humans. Proteins, 72: 1030–1037

Ragunath P., Chitra R., Mohammad S., Abhinand P. 2011. A systems biological study on the comorbidity of autism spectrum disorders and bipolar disorder. Bioinformation, 7: 102–106

Rajala M.W., Scherer P.E. 2003. Minireview: The adipocyte--at the crossroads of energy homeostasis, inflammation, and atherosclerosis. Endocrinology, 144: 3765–3773

Ramanan V.K., Shen L., Moore J.H., Saykin A.J. 2012. Pathway analysis of genomic data: concepts, methods, and prospects for future development. Trends in Genetics: TIG, 28: 323–332

Rankinen T., Zuberi A., Chagnon Y.C., Weisnagel S.J., Argyropoulos G., Walts B., Pérusse L., Bouchard C. 2006. The human obesity gene map: the 2005 update. Obesity, 14: 529–644

Ravasi T., Suzuki H., Cannistraci C. V, Katayama S., Bajic V.B., Tan K., Akalin A., Schmeier S., Kanamori-Katayama M., Bertin N., Carninci P., Daub C.O., Forrest A.R., Gough J., Grimmond S., Han J.H., Hashimoto T., Hide W., Hofmann O., Kamburov A., Kaur M., Kawaji H., Kubosaki A., Lassmann T., Van Nimwegen E., MacPherson C.R., Ogawa C., Radovanovic A., Schwartz A., Teasdale R.D., Tegner J., Lenhard B., Teichmann S.A., Arakawa T., Ninomiya N., Murakami K., Tagami M., Fukuda S., Imamura K., Kai C., Ishihara R., Kitazume Y., Kawai J., Hume D.A., Ideker T., Hayashizaki Y. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. Cell, 140(5): 744–752

Reif D.M., White B.C., Moore J.H. 2004. Integrated analysis of genetic, genomic and proteomic data. Expert Review of Proteomics, 1: 67–75

Rice T., Perusse L., Bouchard C., Rao D.C. 1999. Familial aggregation of body mass index and subcutaneous fat measures in the longitudinal Quebec family study. Genetic Epidemiology. 16: 316–334

Robinson P.N., Köhler S., Bauer S., Seelow D., Horn D., Mundlos S. 2008. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. American Journal of Human Genetics, 83: 610–615

Rosen E.D., Walkey C.J., Puigserver P., Spiegelman B.M. 2000. Transcriptional regulation of adipogenesis. Genes & Development, 14: 1293–1307

Rubinstein R., Simon I. 2005. MILANO--custom annotation of microarray results using automatic literature searches. BMC Bioinformatics, 6: 12

Saito R., Smoot M.E., Ono K., Ruscheinski J., Wang P.L., Lotia S., Pico A.R., Bader G.D., Ideker T. 2012. A travel guide to Cytoscape plugins. Nature Methods, 9: 1069–1076

Saladin R., De Vos P., Guerre-Millo M., Leturque A., Girard J., Staels B., Auwerx J. 1995. Transient increase in obese gene expression after food intake or insulin administration. Nature, 377: 527–529

Schena M., Shalon D., Davis R.W., Brown P.O. 1995. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. Science, 270: 467–470

Schofield P.N., Gruenberger M., Sundberg J.P. 2010. Pathbase and the MPATH ontology. Community resources for mouse histopathology. Veterinary Pathology, 47: 1016–1020

Smith C.L., Goldsmith C.A., Eppig J.T. 2005. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome Biology, 6: R7

Smoot M.E., Ono K., Ruscheinski J., Wang P.L., Ideker T. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics, 27: 431–432

Del Sol A., Balling R., Hood L., Galas D. 2010. Diseases as network perturbations. Current Opinion in Biotechnology, 21: 566–571

Sosnowski O., Charcosset A., Joets J. 2012. BioMercator V3: an upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. Bioinformatics, 28: 2082–2083

Spalding K.L., Arner E., Westermark P.O., Bernard S., Buchholz B.A., Bergmann O., Blomqvist L., Hoffstedt J., Näslund E., Britton T., Concha H., Hassan M., Rydén M., Frisén J., Arner P. 2008. Dynamics of fat cell turnover in humans. Nature, 453: 783–787

Speakman J., Hambly C., Mitchell S., Król E. 2008. The contribution of animal models to the study of obesity. Laboratory Animals, 42: 413–432

Speliotes E.K., Willer C.J., Berndt S.I., Monda K.L., Thorleifsson G., Jackson A.U., Lango Allen H., Lindgren C.M., Luan J., Mägi R., Randall J.C., Vedantam S., Winkler T.W., Qi L., Workalemahu T., Heid I.M., Steinthorsdottir V., Stringham H.M., Weedon M.N., Wheeler E., Wood A.R., Ferreira T., Weyant R.J., Segrè A. V, Estrada K., Liang L., Nemesh J., Park J.H., Gustafsson S., Kilpeläinen T.O., Yang J., Bouatia-Naji N., Esko T., Feitosa M.F., Kutalik Z., Mangino M., Raychaudhuri S., Scherag A., Smith A. V, Welch R., Zhao J.H., Aben K.K., Absher D.M., Amin N., Dixon A.L., Fisher E., Glazer N.L., Goddard M.E., Heard-Costa N.L., Hoesel V., Hottenga J.J., Johansson A., Johnson T., Ketkar S., Lamina C., Li S., Moffatt M.F., Myers R.H., Narisu N., Perry J.R., Peters M.J., Preuss M., Ripatti S., Rivadeneira F., Sandholt C., Scott L.J., Timpson N.J., Tyrer J.P., Van Wingerden S., Watanabe R.M., White C.C., Wiklund F., Barlassina C., Chasman D.I., Cooper M.N., Jansson J.O., Lawrence R.W., Pellikka N., Prokopenko I., Shi J., Thiering E., Alavere H., Alibrandi M.T., Almgren P., Arnold A.M., Aspelund T., Atwood L.D., Balkau B., Balmforth A.J., Bennett A.J., Ben-Shlomo Y., Bergman R.N., Bergmann S., Biebermann H., Blakemore A.I., Boes T., Bonnycastle L.L., Bornstein S.R., Brown M.J., Buchanan T.A., Busonero F., Campbell H., Cappuccio F.P., Cavalcanti-Proença C., Chen Y.D., Chen C.M., Chines P.S., Clarke R., Coin L., Connell J., Day I.N., Den Heijer M., Duan J., Ebrahim S., Elliott P., Elosua R., Eiriksdottir G., Erdos M.R., Eriksson J.G., Facheris M.F., Felix S.B., Fischer-Posovszky P., Folsom A.R., Friedrich N., Freimer N.B., Fu M., Gaget S., Gejman P. V, Geus E.J., Gieger C., Gjesing A.P., Goel A.,

Goyette P., Grallert H., Grässler J., Greenawalt D.M., Groves C.J., Gudnason V., Guiducci C., Hartikainen A.L., Hassanali N., Hall A.S., Havulinna A.S., Hayward C., Heath A.C., Hengstenberg C., Hicks A.A., Hinney A., Hofman A., Homuth G., Hui J., Igl W., Iribarren C., Isomaa B., Jacobs K.B., Jarick I., Jewell E., John U., Jørgensen T., Jousilahti P., Jula A., Kaakinen M., Kajantie E., Kaplan L.M., Kathiresan S., Kettunen J., Kinnunen L., Knowles J.W., Kolcic I., König I.R., Koskinen S., Kovacs P., Kuusisto J., Kraft P., Kvaløy K., Laitinen J., Lantieri O., Lanzani C., Launer L.J., Lecoeur C., Lehtimäki T., Lettre G., Liu J., Lokki M.L., Lorentzon M., Luben R.N., Ludwig B., Manunta P., Marek D., Marre M., Martin N.G., McArdle W.L., McCarthy A., McKnight B., Meitinger T., Melander O., Meyre D., Midthjell K., Montgomery G.W., Morken M.A., Morris A.P., Mulic R., Ngwa J.S., Nelis M., Neville M.J., Nyholt D.R., O'Donnell C.J., O'Rahilly S., Ong K.K., Oostra B., Paré G., Parker A.N., Perola M., Pichler I., Pietiläinen K.H., Platou C.G., Polasek O., Pouta A., Rafelt S., Raitakari O., Rayner N.W., Ridderstråle M., Rief W., Ruokonen A., Robertson N.R., Rzehak P., Salomaa V., Sanders A.R., Sandhu M.S., Sanna S., Saramies J., Savolainen M.J., Scherag S., Schipf S., Schreiber S., Schunkert H., Silander K., Sinisalo J., Siscovick D.S., Smit J.H., Soranzo N., Sovio U., Stephens J., Surakka I., Swift A.J., Tammesoo M.L., Tardif J.C., Teder-Laving M., Teslovich T.M., Thompson J.R., Thomson B., Tönjes A., Tuomi T., Van Meurs J.B., Van Ommen G.J., Vatin V., Viikari J., Visvikis-Siest S., Vitart V., Vogel C.I., Voight B.F., Waite L.L., Wallaschofski H., Walters G.B., Widen E., Wiegand S., Wild S.H., Willemsen G., Witte D.R., Witteman J.C., Xu J., Zhang Q., Zgaga L., Ziegler A., Zitting P., Beilby J.P., Farooqi I.S., Hebebrand J., Huikuri H. V, James A.L., Kähönen M., Levinson D.F., Macciardi F., Nieminen M.S., Ohlsson C., Palmer L.J., Ridker P.M., Stumvoll M., Beckmann J.S., Boeing H., Boerwinkle E., Boomsma D.I., Caulfield M.J., Chanock S.J., Collins F.S., Cupples L.A., Smith G.D., Erdmann J., Froguel P., Grönberg H., Gyllensten U., Hall P., Hansen T., Harris T.B., Hattersley A.T., Hayes R.B., Heinrich J., Hu F.B., Hveem K., Illig T., Jarvelin M.R., Kaprio J., Karpe F., Khaw K.T., Kiemeney L.A., Krude H., Laakso M., Lawlor D.A., Metspalu A., Munroe P.B., Ouwehand W.H., Pedersen O., Penninx B.W., Peters A., Pramstaller P.P., Quertermous T., Reinehr T., Rissanen A., Rudan I., Samani N.J., Schwarz P.E., Shuldiner A.R., Spector T.D., Tuomilehto J., Uda M., Uitterlinden A., Valle T.T., Wabitsch M., Waeber G., Wareham N.J., Watkins H., Wilson J.F., Wright A.F., Zillikens M.C., Chatterjee N., McCarroll S.A., Purcell S., Schadt E.E., Visscher P.M., Assimes T.L., Borecki I.B., Deloukas P., Fox C.S., Groop L.C., Haritunians T., Hunter D.J., Kaplan R.C., Mohlke K.L., O'Connell J.R., Peltonen L., Schlessinger D., Strachan D.P., Van Duijn C.M., Wichmann H.E., Frayling T.M., Thorsteinsdottir U., Abecasis G.R., Barroso I., Boehnke M., Stefansson K., North K.E., McCarthy M.I., Hirschhorn J.N., Ingelsson E., Loos R.J., MAGIC, Consortium P. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genetics, 42: 937–948

Stelling J., Sauer U., Szallasi Z., Doyle F.J., Doyle J. 2004. Robustness of cellular functions. Cell, 118: 675–685

Stunkard A.J., Foch T.T., Hrubec Z. 1986a. A twin study of human obesity. JAMA: The Journal of the American Medical Association, 256: 51–54

Stunkard A.J., Sorensen T.I., Hanis C., Teasdale T.W., Chakraborty R., Schull W.J., Schulsinger F. 1986b. An adoption study of human obesity. The New England Journal of Medicine, 314: 193–198

Tacutu R., Craig T., Budovsky A., Wuttke D., Lehmann G., Taranukha D., Costa J., Fraifeld V.E., De Magalhães J.P. 2013. Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Research, 41: D1027– D1033

Tamames J., Valencia A. 2006. The success (or not) of HUGO nomenclature. Genome Biology, 7: 402

Taylor I.W., Linding R., Warde-Farley D., Liu Y., Pesquita C., Faria D., Bull S., Pawson T., Morris Q., Wrana J.L. 2009. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature Biotechnology, 27: 199–204

Touitou I., Lesage S., McDermott M., Cuisset L., Hoffman H., Dode C., Shoham N., Aganna E., Hugot J.P., Wise C., Waterham H., Pugnere D., Demaille J., Sarrauste de Menthiere C. 2004. Infevers: an evolving mutation database for auto-inflammatory syndromes. Human Mutation, 24: 194–198

Tranchevent L.C., Capdevila F.B., Nitsch D., De Moor B., De Causmaecker P., Moreau Y. 2011. A guide to web tools to prioritize candidate genes. Briefings in Bioinformatics, 12: 22–32.

Ulrich L.E., Zhulin I.B. 2007. MiST: a microbial signal transduction database. Nucleic Acids Research, 35: D386–390

Vaquerizas J.M., Kummerfeld S.K., Teichmann S.A., Luscombe N.M. 2009. A census of human transcription factors: function, expression and evolution. Nature Reviews. Genetics, 10: 252–263

Wan M., Easton R.M., Gleason C.E., Monks B.R., Ueki K., Kahn C.R., Birnbaum M.J. 2012. Loss of Akt1 in mice increases energy expenditure and protects against diet-induced obesity. Molecular and Cellular Biology, 32: 96–106

Wang P.I., Marcotte E.M. 2010. It's the machine that matters: Predicting gene function and phenotype from protein networks. Journal of Proteomics, 73: 2277–2289

Waterston R.H., Lindblad-Toh K., Birney E., Rogers J., Abril J.F., Agarwal P., Agarwala R., Ainscough R., Alexandersson M., An P., Antonarakis S.E., Attwood J., Baertsch R., Bailey J., Barlow K., Beck S., Berry E., Birren B., Bloom T., Bork P., Botcherby M., Bray N., Brent M.R., Brown D.G., Brown S.D., Bult C., Burton J., Butler J., Campbell R.D., Carninci P., Cawley S., Chiaromonte F., Chinwalla A.T., Church D.M., Clamp M., Clee C., Collins F.S., Cook L.L., Copley R.R., Coulson A., Couronne O., Cuff J., Curwen V., Cutts T., Daly M., David R., Davies J., Delehaunty K.D., Deri J., Dermitzakis E.T., Dewey C., Dickens N.J., Diekhans M., Dodge S.,

Dubchak I., Dunn D.M., Eddy S.R., Elnitski L., Emes R.D., Eswara P., Eyras E., Felsenfeld A., Fewell G.A., Flicek P., Foley K., Frankel W.N., Fulton L.A., Fulton R.S., Furey T.S., Gage D., Gibbs R.A., Glusman G., Gnerre S., Goldman N., Goodstadt L., Grafham D., Graves T.A., Green E.D., Gregory S., Guigó R., Guyer M., Hardison R.C., Haussler D., Hayashizaki Y., Hillier L.W., Hinrichs A., Hlavina W., Holzer T., Hsu F., Hua A., Hubbard T., Hunt A., Jackson I., Jaffe D.B., Johnson L.S., Jones M., Jones T.A., Joy A., Kamal M., Karlsson E.K., Karolchik D., Kasprzyk A., Kawai J., Keibler E., Kells C., Kent W.J., Kirby A., Kolbe D.L., Korf I., Kucherlapati R.S., Kulbokas E.J., Kulp D., Landers T., Leger J.P., Leonard S., Letunic I., Levine R., Li J., Li M., Lloyd C., Lucas S., Ma B., Maglott D.R., Mardis E.R., Matthews L., Mauceli E., Mayer J.H., McCarthy M., McCombie W.R., McLaren S., McLay K., McPherson J.D., Meldrim J., Meredith B., Mesirov J.P., Miller W., Miner T.L., Mongin E., Montgomery K.T., Morgan M., Mott R., Mullikin J.C., Muzny D.M., Nash W.E., Nelson J.O., Nhan M.N., Nicol R., Ning Z., Nusbaum C., O'Connor M.J., Okazaki Y., Oliver K., Overton-Larty E., Pachter L., Parra G., Pepin K.H., Peterson J., Pevzner P., Plumb R., Pohl C.S., Poliakov A., Ponce T.C., Ponting C.P., Potter S., Quail M., Reymond A., Roe B.A., Roskin K.M., Rubin E.M., Rust A.G., Santos R., Sapojnikov V., Schultz B., Schultz J., Schwartz M.S., Schwartz S., Scott C., Seaman S., Searle S., Sharpe T., Sheridan A., Shownkeen R., Sims S., Singer J.B., Slater G., Smit A., Smith D.R., Spencer B., Stabenau A., Stange-Thomann N., Sugnet C., Suyama M., Tesler G., Thompson J., Torrents D., Trevaskis E., Tromp J., Ucla C., Ureta-Vidal A., Vinson J.P., Von Niederhausern A.C., Wade C.M., Wall M., Weber R.J., Weiss R.B., Wendl M.C., West A.P., Wetterstrand K., Wheeler R., Whelan S., Wierzbowski J., Willey D., Williams S., Wilson R.K., Winter E., Worley K.C., Wyman D., Yang S., Yang S.-P., Zdobnov E.M., Zody M.C., Lander E.S. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature, 420: 520–562

Wei C.-H., Harris B.R., Li D., Berardini T.Z., Huala E., Kao H.-Y., Lu Z. 2012. Accelerating literature curation with text-mining tools: a case study of using PubTator to curate genes in PubMed abstracts. Database: The Journal of Biological Databases and Curation, 2012: bas041

Whitaker J.W., Letunic I., McConkey G.A., Westhead D.R. 2009. metaTIGER: a metabolic evolution resource. Nucleic Acids Research, 37: D531– D538

Wilkening S., Chen B., Bermejo J.L., Canzian F. 2009. Is there still a need for candidate gene approaches in the era of genome-wide association studies? Genomics, 93: 415–419

Willer C.J., Speliotes E.K., Loos R.J., Li S., Lindgren C.M., Heid I.M., Berndt S.I., Elliott A.L., Jackson A.U., Lamina C., Lettre G., Lim N., Lyon H.N., McCarroll S.A., Papadakis K., Qi L., Randall J.C., Roccasecca R.M., Sanna S., Scheet P., Weedon M.N., Wheeler E., Zhao J.H., Jacobs L.C., Prokopenko I., Soranzo N., Tanaka T., Timpson N.J., Almgren P., Bennett A., Bergman R.N., Bingham S.A., Bonnycastle L.L., Brown M., Burtt N.P., Chines P., Coin L., Collins F.S., Connell J.M., Cooper C., Smith G.D., Dennison E.M., Deodhar P., Elliott P., Erdos M.R., Estrada K., Evans

D.M., Gianniny L., Gieger C., Gillson C.J., Guiducci C., Hackett R., Hadley D., Hall A.S., Havulinna A.S., Hebebrand J., Hofman A., Isomaa B., Jacobs K.B., Johnson T., Jousilahti P., Jovanovic Z., Khaw K.T., Kraft P., Kuokkanen M., Kuusisto J., Laitinen J., Lakatta E.G., Luan J., Luben R.N., Mangino M., McArdle W.L., Meitinger T., Mulas A., Munroe P.B., Narisu N., Ness A.R., Northstone K., O'Rahilly S., Purmann C., Rees M.G., Ridderstråle M., Ring S.M., Rivadeneira F., Ruokonen A., Sandhu M.S., Saramies J., Scott L.J., Scuteri A., Silander K., Sims M.A., Song K., Stephens J., Stevens S., Stringham H.M., Tung Y.C., Valle T.T., Van Duijn C.M., Vimaleswaran K.S., Vollenweider P., Waeber G., Wallace C., Watanabe R.M., Waterworth D.M., Watkins N., Witteman J.C., Zeggini E., Zhai G., Zillikens M.C., Altshuler D., Caulfield M.J., Chanock S.J., Farooqi I.S., Ferrucci L., Guralnik J.M., Hattersley A.T., Hu F.B., Jarvelin M.R., Laakso M., Mooser V., Ong K.K., Ouwehand W.H., Salomaa V., Samani N.J., Spector T.D., Tuomi T., Tuomilehto J., Uda M., Uitterlinden A.G., Wareham N.J., Deloukas P., Frayling T.M., Groop L.C., Hayes R.B., Hunter D.J., Mohlke K.L., Peltonen L., Schlessinger D., Strachan D.P., Wichmann H.E., McCarthy M.I., Boehnke M., Barroso I., Abecasis G.R., Hirschhorn J.N., Consortium W.T.C.C., Consortium G.I. of An.T. 2009. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nature Genetics, 41: 25–34

Wingender E., Dietze P., Karas H., Knüppel R. 1996. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Research, 24: 238–241

World Health Organ. 2000. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organization Technical Report Series, 894: i–xii, 1–253

Wu C., Cui Y. 2013. Boosting signals in gene-based association studies via efficient SNP selection. Briefings in Bioinformatics.

Wu J., Mao X., Cai T., Luo J., Wei L. 2006. KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Research, 34: W720– W724

Xenarios I., Salwínski L., Duan X.J., Higney P., Kim S.M., Eisenberg D. 2002. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Research, 30: 303–305

Xiao F., Zuo Z., Cai G., Kang S., Gao X., Li T. 2009. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Research, 37: D105– D110

Xu J., Ji J., Yan X.-H. 2012. Cross-talk between AMPK and mTOR in regulating energy balance. Critical Reviews in Food Science and Nutrition, 52: 373–381

Yang W., Kelly T., He J. 2007. Genetic epidemiology of obesity. Epidemiologic Reviews, 29: 49–61

Yang W.-L., Wu C.-Y., Wu J., Lin H.-K. 2010. Regulation of Akt signaling activation by ubiquitination. Cell Cycle, 9: 487–497

Yook K., Harris T.W., Bieri T., Cabunoc A., Chan J., Chen W.J., Davis P., De la Cruz N., Duong A., Fang R., Ganesan U., Grove C., Howe K., Kadam S., Kishore R., Lee R., Li Y., Muller H.M., Nakamura C., Nash B., Ozersky P., Paulini M., Raciti D., Rangarajan A., Schindelman G., Shi X., Schwarz E.M., Ann Tuli M., Van Auken K., Wang D., Wang X., Williams G., Hodgkin J., Berriman M., Durbin R., Kersey P., Spieth J., Stein L., Sternberg P.W. 2012. WormBase 2012: more genomes, more data, new website. Nucleic Acids Research, 40: D735–D741

Yoshida Y., Makita Y., Heida N., Asano S., Matsushima A., Ishii M., Mochizuki Y., Masuya H., Wakana S., Kobayashi N., Toyoda T. 2009. PosMed (Positional Medline): prioritizing genes with an artificial neural network comprising medical documents to accelerate positional cloning. Nucleic Acids Research, 37: W147–W152

Yu W., Wulf A., Liu T., Khoury M.J., Gwinn M. 2008. Gene Prospector: an evidence gateway for evaluating potential susceptibility genes and interacting risk factors for human diseases. BMC Bioinformatics, 9: 528

Yusuf D., Butland S.L., Swanson M.I., Bolotin E., Ticoll A., Cheung W.A., Zhang X.Y., Dickman C.T., Fulton D.L., Lim J.S., Schnabl J.M., Ramos O.H., Vasseur-Cognet M., De Leeuw C.N., Simpson E.M., Ryffel G.U., Lam E.W., Kist R., Wilson M.S., Marco-Ferreres R., Brosens J.J., Beccari L.L., Bovolenta P., Benayoun B.A., Monteiro L.J., Schwenen H.D., Grontved L., Wederell E., Mandrup S., Veitia R.A., Chakravarthy H., Hoodless P.A., Mancarelli M.M., Torbett B.E., Banham A.H., Reddy S.P., Cullum R.L., Liedtke M., Tschan M.P., Vaz M., Rizzino A., Zannini M., Frietze S., Farnham P.J., Eijkelenboom A., Brown P.J., Laperrière D., Leprince D., De Cristofaro T., Prince K.L., Putker M., Del Peso L., Camenisch G., Wenger R.H., Mikula M., Rozendaal M., Mader S., Ostrowski J., Rhodes S.J., Van Rechem C., Boulay G., Olechnowicz S.W., Breslin M.B., Lan M.S., Nanan K.K., Wegner M., Hou J., Mullen R.D., Colvin S.C., Noy P.J., Webb C.F., Witek M.E., Ferrell S., Daniel J.M., Park J., Waldman S.A., Peet D.J., Taggart M., Jayaraman P.S., Karrich J.J., Blom B., Vesuna F., O'Geen H., Sun Y., Gronostajski R.M., Woodcroft M.W., Hough M.R., Chen E., Europe-Finner G.N., Karolczak-Bayatti M., Bailey J., Hankinson O., Raman V., LeBrun D.P., Biswal S., Harvey C.J., DeBruyne J.P., Hogenesch J.B., Hevner R.F., Héligon C., Luo X.M., Blank M.C., Millen K.J., Sharlin D.S., Forrest D., Dahlman-Wright K., Zhao C., Mishima Y., Sinha S., Chakrabarti R., Portales-Casamar E., Sladek F.M., Bradley P.H., Wasserman W.W. 2012. The transcription factor encyclopedia. Genome Biology, 13: R24

Zhang Y., De S., Garner J.R., Smith K., Wang S.A., Becker K.G. 2010. Systematic analysis, comparison, and integration of disease based human genetic association data and mouse genetic phenotypic information. BMC Medical Genomics, 3: 1

Zhou G., Wen X., Liu H., Schlicht M.J., Hessner M.J., Tonellato P.J., Datta M.W. 2004. B.E.A.R. GeneInfo: a tool for identifying gene-related biomedical publications through user modifiable queries. BMC Bioinformatics, 5: 46

Zhu X., Gerstein M., Snyder M. 2007. Getting connected: analysis and principles of biological networks. Genes & Development, 21: 1010–1024

Zorc M., Jevšinek Skok D., Godnič I., Calin G.A., Horvat S., Jiang Z., Dovč P., Kunej T. 2012. Catalog of MicroRNA Seed Polymorphisms in Vertebrates. PLoS One, 7: e30737

Zotenko E., Mestre J., O'Leary D.P., Przytycka T.M. 2008. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. PLoS Computational Biology, 4: e100014

ZAHVALA

Zahvaljujem se mentorici prof. dr. Tanji Kunej za odlično vodenje in vsestransko podporo na podiplomskem študiju.

Raziskovalcem z Oddelka za zootehniko Biotehniške fakultete, še posebej Daši Jevšinek Skok, Ireni Godnič, Evi Čeh, Jani Ferdin in Jerneju Ogorevcu, se zahvaljujem za sodelovanje, nasvete, pripombe in namige, ki so pripomogli k nastanku tega dela.

Zahvaljujem se družini, ki mi je vedno stala ob strani.

PRILOGE

Priloga A

Catalog of microRNA seed polymorphisms in vertebrates (Zorc in sod., 2012)

OPEN ACCESS Freely available online

PLoS one

Catalog of MicroRNA Seed Polymorphisms in Vertebrates

Minja Zorc[1,2☯], Dasa Jevsinek Skok[1☯], Irena Godnic[1☯], George Adrian Calin[3], Simon Horvat[1,4], Zhihua Jiang[5], Peter Dovc[1], Tanja Kunej[1]*

1 Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Domzale, Slovenia, **2** Centre for Mathematical and Computational Biology, Rothamsted Research, Harpenden, United Kingdom, **3** Department of Experimental Therapeutics and The Center for RNA Interference and Non-Coding RNAs, The University of Texas, M.D. Anderson Cancer Center, Houston, Texas, United States of America, **4** National Institute of Chemistry, Ljubljana, Slovenia, **5** Department of Animal Sciences, Washington State University, Pullman, Washington, United States of America

Abstract

MicroRNAs (miRNAs) are a class of non-coding RNA that plays an important role in posttranscriptional regulation of mRNA. Evidence has shown that miRNA gene variability might interfere with its function resulting in phenotypic variation and disease susceptibility. A major role in miRNA target recognition is ascribed to complementarity with the miRNA seed region that can be affected by polymorphisms. In the present study, we developed an online tool for the detection of miRNA polymorphisms (miRNA SNiPer) in vertebrates (http://www.integratomics-time.com/miRNA-SNiPer) and generated a catalog of miRNA seed region polymorphisms (miR-seed-SNPs) consisting of 149 SNPs in six species. Although a majority of detected polymorphisms were due to point mutations, two consecutive nucleotide substitutions (double nucleotide polymorphisms, DNPs) were also identified in nine miRNAs. We determined that miR-SNPs are frequently located within the quantitative trait loci (QTL), chromosome fragile sites, and cancer susceptibility loci, indicating their potential role in the genetic control of various complex traits. To test this further, we performed an association analysis between the *mmu-miR-717* seed SNP rs30372501, which is polymorphic in a large number of standard inbred strains, and all phenotypic traits in these strains deposited in the Mouse Phenome Database. Analysis showed a significant association between the *mmu-miR-717* seed SNP and a diverse array of traits including behavior, blood-clinical chemistry, body weight size and growth, and immune system suggesting that seed SNPs can indeed have major pleiotropic effects. The bioinformatics analyses, data and tools developed in the present study can serve researchers as a starting point in testing more targeted hypotheses and designing experiments using optimal species or strains for further mechanistic studies.

Citation: Zorc M, Jevsinek Skok D, Godnic I, Calin GA, Horvat S, et al. (2012) Catalog of MicroRNA Seed Polymorphisms in Vertebrates. PLoS ONE 7(1): e30737. doi:10.1371/journal.pone.0030737

Editor: Alfons Navarro, University of Barcelona, Spain

Received June 15, 2011; **Accepted** December 23, 2011; **Published** January 27, 2012

Funding: This work was supported by the Slovenian Research Agency (ARRS) through the Research programme Comparative genomics and genome biodiversity [grant number P4-0220]. G.A.C. is supported as a Fellow at The University of Texas MD Anderson Research Trust, as a University of Texas System Regents Research Scholar and by the CLL Global Research Foundation. Work in Dr. Calin's laboratory is supported in part by the NIH/NCI (1R01CA135444), a Department of Defense Breast Cancer Idea Award, Developmental Research Awards in Breast Cancer, Ovarian Cancer, Brain Cancer, Multiple Myeloma and Leukemia SPOREs, and the Laura and John Arnold Foundation and RGK Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tanja.kunej@bf.uni-lj.si

☯ These authors contributed equally to this work.

Introduction

MicroRNAs (miRNA) are non-coding RNA molecules with approximately 21 nucleotides in length, which play an important role in posttranscriptional regulation of mRNA. By binding to the target gene's complementary sequence of the 3′ untranslated region (3′UTR) they repress translation [1]. Changes in miRNA expression profiles have been identified in diseases, including several cancer types (reviewed in [2,3]). Additionally, single nucleotide polymorphisms (SNPs) of miRNA precursors, their target sites, and silencing machinery were reported to interfere with miRNA function and they are likely to affect phenotypic variation, including disease susceptibility [4]. For example, genetic variants affecting the miRNA pathways were involved in diseases such as cancer, neurological disorders, muscular hypertrophy, gastric mucosal atrophy, cardiovascular disease, and type 2 diabetes [5–7]. The term miR-SNP refers to the variation that occurs in the miRNA gene sequence, while the miR-TS-SNP to the SNP that occurs in the miRNA target site (TS) or binding site [8]. Because one miRNA can have multiple targets, miR-SNPs would be expected to exhibit more profound and broader biological effects than miR-TS-SNPs [8]. SNPs in miRNA genes may alter their sequences and therefore enhance, diminish or even generate or cancel out their ability to bind to target sites [9]. Therefore, miR-SNPs could have an impact on the catalogue of miRNA targets, not only by disrupting the interaction of the mutant miRNA with its target genes, but also by creating illegitimate targets that are not targeted by the wild type miRNA [10].

The key binding location for translational suppression resides in the mature miRNA sequence, more accurately nucleotides 2–7 or 2–8 from the 5′ end of the miRNA, also called the seed region [8]. The minimal pairing requirement is a 6-nt match of the seed region (2–7 nucleotides), which can be extended to a 7-nt match (2–8 nucleotides) due to a highly conserved nucleotide position immediately upstream [11]. The causal effect of the miR-SNPs in

the seed region (miR-seed-SNPs) on phenotypic variation has been shown recently; two groups discovered that a miR-seed-SNP in miR-96 was responsible for hearing loss in human and mouse [10,12]. A genomic overlap of four layers was also identified consisting of growth associated quantitative trait locus (QTL), body mass associated *Gpc3* gene, miRNA gene (*mmu-miR-717*), and miR-seed-SNP identified in the lean mouse strain 129/Sv [13].

Polymorphisms within miRNA genes have been reported to be rare, with only approximately 10% (65/474 reported miRNAs at that time) of human pre-miRNAs having documented SNPs, and <1% (3/474) of miRNAs having SNPs in the functional seed region [9]. Similarly, a survey on 173 human miRNA genes revealed 10 SNPs in the pre-miRNAs but none in the seed region [14]. As such, the information about miR-seed-SNPs has received much less attention, while the data remains limited mostly to human and mouse, fragmented, and dispersed among various databases and publications: Patrocles [15], PolymiRTS [16], miRvar [17], and miRNASNP [18]. In contrast, miR-TS-SNPs that influence disease susceptibility, especially cancer, have been the subject of intense research in the last few years [19,20]. Additionally, catalog of SNPs residing in miRNA binding regions has already been compiled [21]. Therefore, the aim of the present study was to develop the tool for detection of miRNA polymorphisms in vertebrates and assemble the information associated with SNPs residing within the miRNA seed region into a single catalog. This web-based public resource should enable faster and more targeted studies on miR-seed-SNP biology precluding a need for preliminary bioinformatics, mouse model and phenotype screens.

Materials and Methods

Development of the online tool for the detection of genetic variations within miRNA genes

A web based tool named "miRNA SNiPer" was developed for the detection of polymorphisms residing within miRNA genes in vertebrates. It accepts a list of miRNA genes and returns a table of variations within different regions of miRNA genes: pre-miRNA, mature, and seed region. The mature sequences are designated as "miR" and the precursor hairpins as "mir" [22]. The tool retrieves data from multiple sources: 1) miRNA gene sequences, genomic coordinates, and nomenclature from miRBase, release 18 (http://www.mirbase.org/) [23], 2) locations of miRNA seed regions from TargetScan, release 5.2 (http://www.targetscan.org/) [11], and 3) locations of genetic polymorphisms from Ensembl Variation database, release 64 (http://www.ensembl.org/) [24]. The assemblies from miRBase, TargetScan, and Ensembl Variation database were downloaded and locally inserted into a MySQL database. The tool is implemented as a CGI (Common Gateway Interface) script written in Perl. Script triggers SQL commands to the MySQL database to perform the searches of variations within miRNA genes. The tool miRNA SNiPer is available at http://integratomics-time.com/miRNA-SNiPer/.

Catalog of the miR-seed polymorphisms

The developed tool miRNA SNiPer was used to generate an assembly of miR-seed polymorphisms in vertebrates (http://www.integratomics-time.com/miR-seed-SNPs/catalog/). The assembled list was supplemented with information relevant for further analysis from the literature (PubMed: http://www.ncbi.nlm.nih.gov/pubmed; Web of Science: http://apps.webofknowledge.com/) and from other sources. Validation status and allele frequencies were retrieved from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). Overlaps with host genes were retrieved from miRBase, release 18 [23]. Validated miRNA targets were extracted from TarBase v5.0 (http://diana.cslab.ece.ntua.gr/tarbase/) [25] and miRecords (http://mirecords.biolead.org) databases [26]. The list of assembled miR-seed polymorphisms was additionally verified with other online databases and tools.

Physical characterization of the miRNA polymorphisms

Genomic distribution of miR-seed polymorphisms was presented on a genomic view (http://www.integratomics-time.com/miR-seed-SNPs/genomic_view/) using Flash GViewer web tool (http://gmod.org/wiki/Flashgviewer/). Overlap analysis of miRNAs comprising seed-SNPs with genomic fragile sites was performed using data retrieved from Ensembl *via* BioMart. Human and mouse QTL were retrieved from the Rat Genome Database (RGD) (http://rgd.mcw.edu/) [27] and chicken QTL were retrieved from Animal QTL Database, release 15 (http://www.animalgenome.org/cgi-bin/QTLdb/index/) [28].

Functional characterization of the miR-seed polymorphisms

TargetScan Custom feature (http://www.targetscan.org/) was used to analyze whether the miR-seed-SNP cause the formation of seed regions annotated to different miRNA [11]. The information regarding the association between miRNAs with polymorphic seed regions and diseases was retrieved from miR2Disease (http://watson.compbio.iupui.edu:8080/miR2Disease/) [29] and published literature (PubMed).

Seed SNP genotype to phenotype association analysis

Association between the mouse seed SNP in *mmu-miR-717* (rs30372501) and phenotypes was analyzed. Data for the genotype-phenotype association analysis was downloaded from the Mouse Phenome Database (MPD; http://phenome.jax.org/) [30]. A test was carried out on all phenotypic data from MPD consisting of 2586 traits in 35 groups (appearance and coat color, behavior, blood-clinical chemistry, blood-hematology, blood-lipids, blood-miscellaneous, body composition, body weight size and growth, bone, brain, breathing pattern, cancer, cardiovascular, cell and tissue damage, development, ear, endocrine, eye, gallbladder, immune system, ingestive preference, kidney, liver, local experiment parameter, longevity, metabolism, metastatic progression, mouse procurement, muscle, nervous system, neurosensory, reproduction, respiratory, sensory gating, and spleen). Seed SNP in *mmu-miR-717* (rs30372501 genotypic and phenotypic data was available for 14 inbred mouse strains (129S1/SvImJ, A/J, AKR/J, BALB/cByJ, C3H/HeJ, C57BL/6J, DBA/2J, FVB/NJ, KK/HlJ, MOLF/EiJ, NOD/ShiLtJ, NZW/LacJ, PWK/PhJ, and WSB/EiJ) consisting of a various number of measurements, ranging from one to 311 for each strain.

Statistical package SAS/STAT [31] was used for statistical analyses. The following linear model was used in the analysis:

$$y_{ijkl} = \mu + G_i + L_{ij} + S_k + e_{ijkl} \quad (1)$$

where y_{ijkl} represents the observation for the traits, μ trait average, G_i fixed effect of genotype *mmu-miR-717* seed SNP rs30372501 (i = CC, TT), L_{ij} nested effect of strain ($j = 1-14$), S_k fixed effect of sex (k = f, m), and e_{ijkl} random error.

Results and Discussion

We developed a web-based tool for the detection of genetic variations in miRNA genes in vertebrates and generated an

open access catalog of polymorphisms that overlap with miRNA seed regions (**Fig. 1A**). This catalog was supplemented with information relevant for further functional analysis. Genotype-phenotype analysis of the murine miR-seed-SNP located in *mmu-miR-717* showed association with a diverse array of traits.

Development of the online tool for the detection miRNA gene polymorphisms

The online tool miRNA SNiPer for the detection of genetic polymorphisms residing within miRNA genes in vertebrates (http://www.integratomics-time.com/miRNA-SNiPer/) was developed using data assembled from miRBase, TargetScan and Ensembl Variation database (**Fig. 2**). The search for miR-seed SNPs was performed in thirteen vertebrate species: chicken, chimpanzee, dog, horse, human, mouse, opossum, pig, platypus, pufferfish, rat, zebra finch, and zebrafish.

Display settings enable the miR-SNPs to be arranged according to their location in pre-miRNAs, mature or seed regions. In six vertebrate species (human, mouse, chicken, chimpanzee, rat, and zebra finch) 149 polymorphisms overlapped with miRNA seed regions (**Fig. 1B, Table S1**). These polymorphisms included SNPs, double nucleotide polymorphisms (DNPs), and insertion/deletions (indels). An example of miR-SNPs located within the pre-miRNA, mature and seed regions of miRNA gene is demonstrated in **Figure 2**.

Data from species in which the latest releases of Ensembl Variation Database and miRBase assemblies are currently not compatible were not included in the catalog (cat, cattle, rabbit, etc.). The miRNA SNiPer tool is going to be updated with each new release of compatible databases.

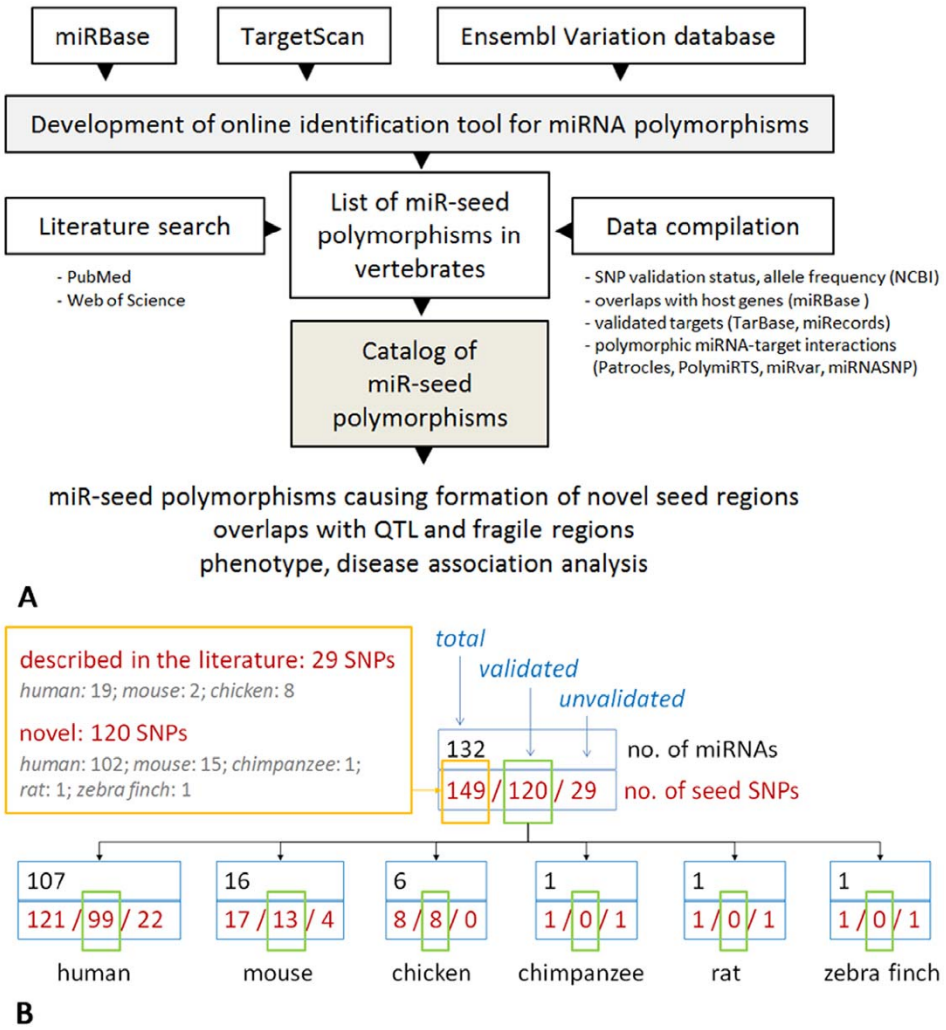
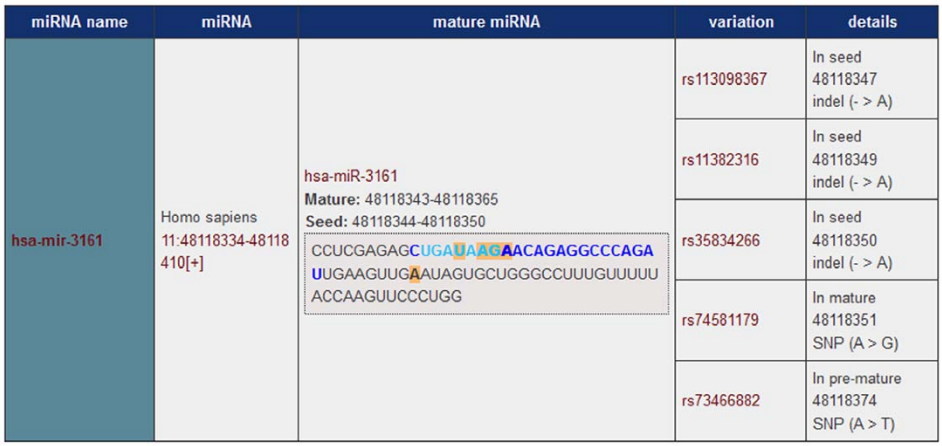


Figure 1. Workflow diagram of the study and diagram of assembled polymorphic miRNAs. (**A**) Workflow diagram of the study: approaches applied for search of known and novel seed miRNA variations and further bioinformatic analysis performed on the database of miR-seed polymorphisms. (**B**) Diagram of assembled miRNAs comprising miR-seed-SNPs according to source, validation status and species.
doi:10.1371/journal.pone.0030737.g001



miRNA name	miRNA	mature miRNA	variation	details
hsa-mir-3161	Homo sapiens 11:48118334-48118410[+]	hsa-miR-3161 **Mature:** 48118343-48118365 **Seed:** 48118344-48118350 CCUCGAGAGCUGAUAAGAACAGAGGCCCAGA UUGAAGUUGAAUAGUGCUGGGCCUUUGUUUUU ACCAAGUUCCCUGG	rs113098367	In seed 48118347 indel (- > A)
			rs11382316	In seed 48118349 indel (- > A)
			rs35834266	In seed 48118350 indel (- > A)
			rs74581179	In mature 48118351 SNP (A > G)
			rs73466882	In pre-mature 48118374 SNP (A > T)

Figure 2. Output of developed miRNA SNiPer tool. An example of miR-SNPs located in pre-miRNA, mature, or seed region of the human miRNA *hsa-miR-3161*. Mature miRNA sequence are highlighted in dark blue, seed regions in light blue, and polymorphisms in orange.
doi:10.1371/journal.pone.0030737.g002

Catalog and genomic view of miR-seed polymorphisms in vertebrates

The list of miR-seed polymorphisms in six species (human, mouse, chicken, chimpanzee, rat, and zebra finch) was supplemented with data retrieved from literature and databases, and presented as an open access online catalog (http://www.integratomics-time.com/miR-seed-SNPs/catalog/). From the total of 149 identified miR-seed polymorphisms only 29 have previously been described in the miRNA context [8–10,12,13,32–45] (**Fig. 1B**). Among them four studies discussed the miR-seed-SNPs as located in the mature region of the miRNA without referring to their miRNA seed location [36,39,44,45]. The remaining 120 miR-seed polymorphisms from the catalog have not been described in publications previously. Data integration revealed that 120 of 149 miR-seed polymorphisms had been previously validated or genotyped. The catalog was additionally supplemented with information of the host gene location and orientation, validation status of miRNA target genes and SNPs (**Table S1**). Genomic distribution of the assembled miR-seed-SNPs was presented on the genomic view (http://www.integratomics-time.com/miR-seed-SNPs/genomic_view) (**Fig. 3**).

The frequency estimations of miRNAs having polymorphic seed regions at this point should be treated with caution because all miRNAs have not yet been systematically sequenced and screened for polymorphisms, and some SNPs in the databases still have unvalidated status. Our preliminary data showed that in human 88 of 1527 (5.7%) currently annotated miRNAs had polymorphic seed regions (99 validated seed-SNPs in total). Similarly, in mouse 13 of 741 (1.7%) miRNAs had miR-seed-SNPs (all previously validated), whereas in chicken six of 499 (1.2%) miRNAs had miR-seed-SNPs (eight validated). Nevertheless, our analysis of currently available data revealed a much higher frequency (5.9%) of miR-seed polymorphisms in human than the frequency of >1% reported by Didiano and Hobert [46]. Similarly, Muiños-Gimeno *et al.* [47] also revealed much lower frequencies than our study mapping 24 SNPs within 325 mature miRNAs and detecting only two miR-seed-SNPs. They also estimated the density of SNPs in miRNAs to be 4.5-times lower that in the rest of the genome [47]. In contrast to this, we found that some miRNAs had highly polymorphic seed regions. For instance, pre-miRNA *gga-mir-1658* had one SNP within the mature sequence (*gga-miR-1658*) and two SNPs within the minor miR sequence (*gga-miR-1658**). Our study also revealed four human miRNAs (*hsa-miR-96*, *-518e-5p*, *-1304-5p*, and *-3939*) that had two validated consecutive nucleotides altered, so called DNPs, which have been found to occur with a frequency of ~1% of the total number of SNPs in the genome [48]. It has been suggested that DNPs have a greater propensity to be involved in disease causing mutations in protein coding regions as they effect two positions in a codon, resulting in a more likely non-synonymous mutation [48]. We can speculate that miR-seed-DNPs identified in our study (four validated and five unvalidated) also have a potential to cause more profound effects on the regulation of target genes and phenotypes that single miR-seed-SNPs, but this is to be evaluated experimentally in future studies.

Allele frequency was available for 90 SNPs in human, 12 in mouse, and eight SNPs in chicken (NCBI). Population-based differences were observed for 41 human SNPs; among them rs12975333 described as polymorphic in three studies [9,32,35], but monomorphic in the Spanish [47] and Scandinavian populations [49]. As expected, transitions (purine ↔ purine or pyrimidine ↔ pyrimidine substitutions) were twice as more frequent than transversions (purine ↔ pyrimidine). Because transversions induce greater genetic alternations that are more likely to cause functional consequences [50] we have examined their prevalence in miR-seed-SNPs and found 31 transversions; 27 in human, three in mouse, and one in chicken. Interestingly, transversions were observed in *hsa-miR-96* and *mmu-miR-96* which have been previously linked with hearing loss in both human and mouse [10,12].

Data analysis

The data from the catalog were further analyzed to prioritize promising SNPs for further functional analysis. We examined each miR-seed SNP for their potential to generate novel seed regions that also match to a different miRNA, for their genomic distribution, overlaps with host genes, QTL, and fragile regions, as well as their association with diseases and phenotypes.

miR-seed polymorphisms causing the formation of a seed region annotated to a different miRNA. As shown in Mencia *et al.*, [10] a miR-seed-SNP+13G>A in *hsa-miR-96* caused a

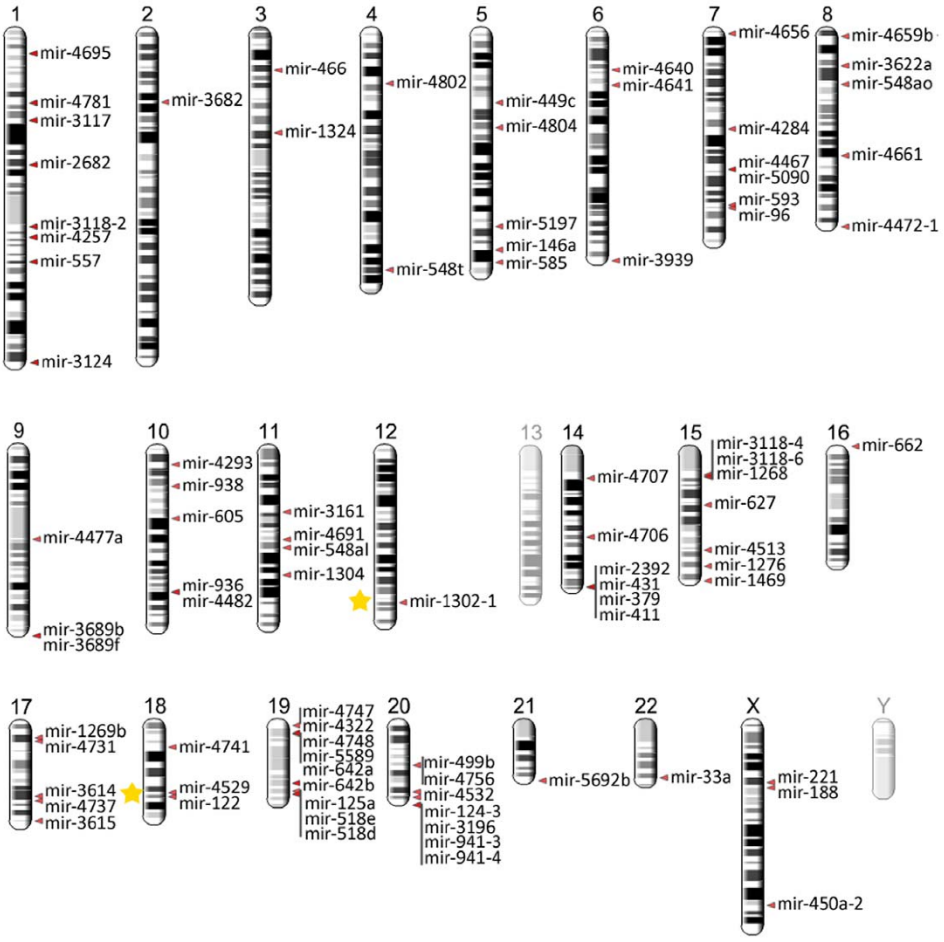


Figure 3. Genomic location of miRNAs with polymorphic seed regions in human. miRNAs comprising validated seed region polymorphisms mapping to two overlapping fragile sites are marked with yellow stars.
doi:10.1371/journal.pone.0030737.g003

change in the seed region to perfectly match another annotated miRNA *hsa-miR-514*, implying a possibility of targets shared by both miRNAs. To determine whether miR-seed-SNPs cause the formation of seed regions annotated to different miRNAs we screened the catalog using TargetScan Custom. SNPs in *hsa-miR-3117-3p* and *-4467* matched two different seed regions of *hsa-miR-499-5p* and *-885-3p*, respectively (**Figure S1**). A change of annotated seed regions may lead to altered recognition and selection of miRNA targets, which could possibly be a part of a different biological pathway.

Genomic distribution of miR-seed-SNPs and their overlaps with host genes, QTL, and fragile regions. Genomic locations of miRNAs comprising seed polymorphisms are shown in **Figure 3**, **Figures S2** and **S3** for human, mouse, and chicken, respectively. The highest number of miR-seed polymorphisms was present on human chromosomes 1, 15, 19, and 20. Several miRNAs with polymorphic seed regions overlapped with host genes, QTL, fragile sites, or cancer susceptibility sites.

We observed 75 miRNAs with polymorphic seed regions residing within protein coding host genes; 16 in antisense, 55 in sense orientation and four miRNAs that overlapped with host genes both orientations (**Table S1**). MiRNA genes and their host genes in antisense orientation have been shown to have independent transcription mechanisms [51], whereas sense transcriptional orientation suggests that miRNAs and host genes can be transcribed from shared promoters [1]. Sense oriented miRNA genes from our catalog were either exonic (five in human and three in mouse), intronic (34 in human, three in mouse, and three in chicken), or overlapped both exonic and intronic transcripts (seven in human) (**Table S1**). Intronic miRNAs have previously been found to be co-expressed and regulated by co-activation of both miRNA and its host gene [52,53]. Several studies have also shown that host genes are functionally linked with their resident miRNAs [53–55].

By comparing locations of polymorphic seed regions with QTL and fragile sites, several overlaps were found. MiRNAs with validated seed SNPs overlapped with 830 QTL in human, 118 in mouse and 20 in chicken. Highest number of overlapped QTL in human was observed for *hsa-miR-4737* and *hsa-miR-4756* each overlapping 43 QTL. In mouse mmu-miR-628 overlapped with

23 QTL and in chicken gga-miR-1658 with seven QTL (data not shown). MiRNA genes have also been observed to be frequently located near the mouse cancer susceptibility loci, which is in concordance with a previous study of Sevignani *et al.* [56]. These results support previous observations that miRNAs are an important player in generating genetic variability and important genomic sites in the trait's genetic architecture.

Hsa-miR-1302-1 overlapped with a common fragile site located at 12q24.1 and *hsa-miR-122-3p* overlapped with a fragile site located at 18q21.3 (**Fig. 3**). *Hsa-miR-513a-5p* overlapped with a fragile site located at cytogenetic band Xq27; however, this SNP is yet to be validated. This observation is in concordance with a previous study showing that miRNAs are frequently located at fragile sites, as well as minimal regions of loss of heterozygosity, minimal regions of amplification, or common breakpoint regions [57].

MiR-seed-SNP association with diseases and phenotypes. We reviewed published associations between miRNAs with polymorphic seed regions and diseases/phenotypes. Additionally, we performed a statistical genotype-phenotype association analysis using the data from the Mouse Phenome Database. Four human and two mouse miRNAs comprising seed-SNPs have already been associated with diseases and phenotypes (**Table 1**) [10,12, 13,32–34,37–39,41,42,44,45,58]. Hsa-miR-146a-3p and -499-3p were associated with the largest variety of pathologies affecting all organ systems, especially the reproductive and digestive system. Two separate studies have linked the SNP in the seed region of miR-96 to the same clinical pathology, hearing loss, in mouse [12] and human [10], which also represents the first case implicating a miRNA in a Mendelian inherited disorder. In a recent study, Kunej *et al.* [13] analyzed a murine SNP (rs30372501) in *mmu-miR-717* seed region which was found to be associated with leanness. Evidence also exists that miR-717 is involved in osmoregularity control and is regulated rapidly in response to high salt exposure in mice [59]. To verify if these associations hold true for some other standard inbred mouse strains we searched for association between the *mmu-miR-717* SNP (rs30372501) genotypes and all 2586 phenotypes within the Mouse Phenome Database (see Material and Methods). The SNP rs30372501 showed a significant effect ($p<0.01$) on 363 measurements-parameters that are grouped by MPD into 25 trait-groups: 23 in behavior, 27 in blood-clinical chemistry, 40 in blood-hematology, 22 in blood-lipids, one in blood-miscellaneous, 19 in body composition, 38 in body weight size and growth, 50 in bone, one in brain, two in breathing pattern, one in cancer, five in cardiovascular, three in cell and tissue damage, five in ear, four in endocrine, two in gallbladder, 18 in immune system, three in ingestive preference, 11 in kidney, four in liver, seven in local experiment parameter, nine in muscle, 23 in respiratory, one in sensory gating, two in spleen, and 42 in ungrouped (**Table S2**). A result that seed SNP rs30372501 was significantly associated with 363 measurements-parameters should be closely examined and interpreted further. As shown in **Table S2**, many measurements within a group are highly correlated (e.g. body weights at various ages within a group "body weight size and growth") and also groups of traits can be highly correlated (e.g., body weight and body composition traits, fat depot weight etc.). Therefore, one should not interpret these associations as causal but rather as a list of potential groups of traits that a particular miR-SNP could affect. In this sense a visual presentation (**Figure S4**) of associations shown in **Table S2** can be informative as one can observe immediately which groups have the highest number of significant associations and hence string support and which trait groups can be further joined into related "super" groups (e.g. body weight, body composition, blood lipids etc.). Such examination can help researchers to prioritize further causation experiments by providing them only a small number of different traits likely to be controlled by a miR-SNP.

Associations between SNP rs30372501 and obesity traits reported by Kunej *et al.* [13] were confirmed also in the present statistical analysis. **Figure 4** shows significant differences in Fat weight (g) between the lean mouse strains carrying a seed SNP rs30372501 allele C (e.g., 129S1/SvImJ, NOD/ShiLtJ) and allele T-carrying high fat strains (e.g., A/J, DBA/2J) in both sexes. Such miR-SNP-genotype to phenotype association analyses can help researchers select an optimal strain and phenotype for further experiments as well as identifying traits and pathologies likely to be affected by miR-SNP variability.

Future perspectives

The following open questions could be addressed in future studies: 1) To examine the effects of SNPs that cause a formation of seed regions already annotated to different miRNAs. 2) To experimentally validate DNPs found in seed regions for their effect

Table 1. Diseases and phenotypes associated with miRNA gene polymorphisms within the seed region in human and mouse.

miRNA	miRNA genomic location	Associated diseases and phenotypes
HUMAN		
hsa-miR-96	chr 7: 129414532–129414609	nonsyndromic progressive hearing loss [10]
hsa-miR-125a-5p	chr 19: 52196507–52196592	breast cancer [32]
hsa-miR-146a-3p	chr 5: 159912359–159912457	breast and ovarian cancer [33]; hepatocellular carcinoma [34]; thyroid cancer [37]; esophageal squamous cell carcinoma [38]; gastric cancer [39]; dilated cardiomyopathy [58]; prostate cancer [41]; cervical squamous cell carcinoma [45]
hsa-miR-499a-3p	chr 20: 33578179–33578300	breast cancer [36]; gastric cancer [39]; prostate cancer [42]; ulcerative colitis [44]; cervical squamous cell carcinoma [45]
MOUSE		
mmu-miR-96	chr 6: 30119446–30119551	hearing loss [12]
mmu-miR-717	chr X: 49775584–49775692	leanness [13]; behavior, blood-clinical chemistry, blood-hematology, blood-lipids, blood-miscellaneous, body composition, body weight size and growth, bone, brain, breathing pattern, cancer, cardiovascular, cell and tissue damage, ear, endocrine, gallbladder, immune system, ingestive preference, kidney, liver, local experiment parameter, muscle, respiratory, sensory gating, spleen (present study)

doi:10.1371/journal.pone.0030737.t001

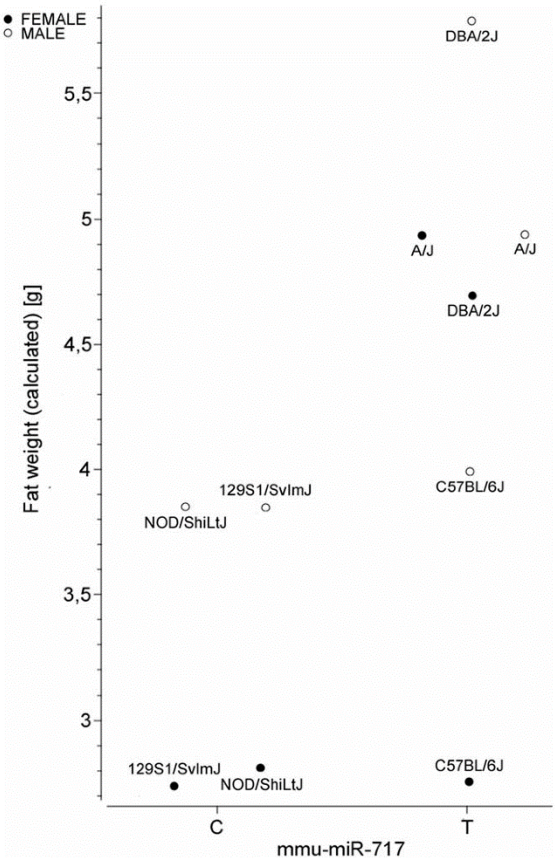


Figure 4. MiR-seed-SNP within *mmu-miR-717* is associated with fat weight in mice. Association analysis between *mmu-miR-717* SNP genotypes in different inbred mouse strains and phenotypes within the Mouse phenome database revealed that *mmu-miR-717* significantly affects several different traits. An example of significant difference between lean and high fat strains that differ for *mmu-miR-717* SNP genotype (C>T) and fat weight, for both female and male is shown.
doi:10.1371/journal.pone.0030737.g004

on miRNA target selection. 3) To study effects of miR-seed-SNPs identified herein on shared transcriptional regulation, expression and function of polymorphic miRNAs and their host genes. 4) Our statistical association analysis of seed-SNP with mouse phenotypes showed a diverse array for associated phenotypes. Further studies should be designed to examine the molecular mechanism for such differential miR-SNP pleiotropic effects.

In conclusion, miR-seed polymorphisms may have a profound effect on a wide range of phenotypes. Using the database integration we assembled all known and identified novel miR-seed-SNPs, and performed a first systematic case-study in this field. The project is ongoing, as novel miRNAs and SNPs are yet to be discovered in human as well as in other animal species. However, results and tools developed in this study can be immediately used by interested scientific community to help retrieve valuable information and design efficient experimental plans in the field of miR-SNP research.

Supporting Information

Figure S1 miR-seed-SNP causing formation of novel seed regions. Three examples of miRNAs (green) with seed-SNPs which cause a formation of a seed region annotated to another miRNA are indicated (red).
(TIF)

Figure S2 Genomic location of miRNAs with polymorphic seed regions in mouse.
(TIF)

Figure S3 Genomic location of miRNAs comprising seed polymorphisms in chicken.
(TIF)

Figure S4 Graphical representation of Table S2 showing association between *mmu-miR-717* seed SNP rs30372501 and 363 traits clustered into 25 groups: behavior, blood-clinical chemistry, blood-hematology, blood-lipids, blood-miscellaneous, body composition, body weight size and growth, bone, brain, breathing pattern, cancer, cardiovascular, cell and tissue damage, ear, endocrine, gallbladder, immune system, ingestive preference, kidney, liver, local experiment parameter, muscle, respiratory, sensory gating and spleen.
(TIF)

Table S1 Catalog of miRNAs with polymorphic seed regions in human, mouse, chicken, chimpanzee, rat, and zebra finch: genomic location, host gene orientation, nucleotide substitution and validation status of the SNP.
(DOC)

Table S2 Estimated differences between miR-seed- SNP (rs30372501) alleles (C>T), associated standard errors and P_values for 363 traits.
(DOC)

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the submitted paper.

Author Contributions

Conceived and designed the experiments: TK SH GAC. Performed the experiments: TK IG MZ SH DJS. Analyzed the data: IG MZ SH DJS. Contributed reagents/materials/analysis tools: MZ SH. Wrote the paper: IG TK SH MZ DJS GAC. Final editing of the text: TK SH GAC ZJ PD.

References

1. Bartel DP (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell 116: 281–297.
2. Ferdin J, Kunej T, Calin G (2010) Non-coding RNAs: Identification of Cancer-Associated microRNAs by Gene Profiling. Technology in Cancer Research & Treatment. pp 123 138.
3. Kunej T, Godnic I, Ferdin J, Horvat S, Dovc P, et al. Epigenetic regulation of microRNAs in cancer: An integrated review of literature. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis In Press, Corrected Proof.
4. Georges M, Coppieters W, Charlier C (2007) Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. Current Opinion in Genetics & Development 17: 166–176.
5. Fabbri M, Valeri N, Calin GA (2009) MicroRNAs and genomic variations: from Proteus tricks to Prometheus gift. Carcinogenesis 30: 912 917.
6. Mishra PJ, Bertino JR (2009) MicroRNA polymorphisms: the future of pharmacogenomics, molecular epidemiology and individualized medicine. Pharmacogenomics 10: 399–416.

7. Hoffman AE, Zheng T, Yi C, Leaderer D, Weidhaas J, et al. (2009) microRNA miR-196a-2 and Breast Cancer: A Genetic and Epigenetic Association Study and Functional Analysis. Cancer Research 69: 5970–5977.
8. Sun G, Yan J, Noltner K, Feng J, Li H, et al. (2009) SNPs in human miRNA genes affect biogenesis and function. RNA 15: 1640–1651.
9. Saunders MA, Liang H, Li W-H (2007) Human polymorphism at microRNAs and microRNA target sites. Proceedings of the National Academy of Sciences 104: 3300–3305.
10. Mencia A, Modamio-Hoybjor S, Redshaw N, Morin M, Mayo-Merino F, et al. (2009) Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet 41: 609–613.
11. Lewis BP, Burge CB, Bartel DP (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell 120: 15–20.
12. Lewis MA, Quint E, Glazier AM, Fuchs H, De Angelis MH, et al. (2009) An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. Nat Genet 41: 614–618.
13. Kunej T, Skok D, Horvat S, Dovc P, Jiang Z (2010) The Glypican 3-Hosted Murine Mir717 Gene: Sequence Conservation, Seed Region Polymorphisms and Putative Targets. International Journal of Biological Sciences. pp 769–772.
14. Iwai N, Naraba H (2005) Polymorphisms in human pre-miRNAs. Biochemical and Biophysical Research Communications 331: 1439–1444.
15. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D (2010) Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Research 38: D640–D651.
16. Ziebarth JD, Bhattacharya A, Chen A, Cui Y (2011) PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. Nucleic Acids Res.
17. Bhartiya D, Laddha SV, Mukhopadhyay A, Scaria V (2011) miRvar: A comprehensive database for genomic variations in microRNAs. Hum Mutat 32: E2226–2245.
18. Gong J, Tong Y, Zhang HM, Wang K, Hu T, et al. (2011) Genome-wide identification of SNPs in MicroRNA genes and the SNP effects on MicroRNA target binding and biogenesis. Hum Mutat.
19. Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, et al. (2010) Single-Nucleotide Polymorphisms Inside MicroRNA Target Sites Influence Tumor Susceptibility. Cancer Research 70: 2789–2798.
20. Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, et al. (2008) A SNP in a let-7 microRNA Complementary Site in the KRAS 3′ Untranslated Region Increases Non–Small Cell Lung Cancer Risk. Cancer Research 68: 8535–8540.
21. Landi D, Gemignani F, Barale R, Landi S (2007) A Catalog of Polymorphisms Falling in MicroRNA-Binding Regions of Cancer Genes. DNA and Cell Biology 27: 35–43.
22. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34: D140–144.
23. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research 39: D152–D157.
24. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26: 2069–2070.
25. Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG (2009) The database of experimentally supported targets: a functional update of TarBase. Nucleic Acids Research 37: D155–D158.
26. Xiao F, Zuo Z, Cai G, Kang S, Gao X, et al. (2009) miRecords: an integrated resource for microRNA–target interactions. Nucleic Acids Research 37: D105–D110.
27. Twigger SN, Shimoyama M, Bromberg S, Kwitek AE, Jacob HJ, et al. (2007) The Rat Genome Database, update 2007—Easing the path from disease to data and back again. Nucleic Acids Research 35: D658–D662.
28. Hu Z-L, Reecy J (2007) Animal QTLdb: beyond a repository. Mammalian Genome 18: 1–4-4.
29. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, et al. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Research 37: D98–D104.
30. Grubb SC, Maddatu TP, Bult CJ, Bogue MA (2009) Mouse Phenome Database. Nucleic Acids Research 37: D720–D730.
31. Institute S (2002) The SAS System for Windows, Release 9.1. CaryNC, .
32. Duan R, Pak C, Jin P (2007) Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. Human Molecular Genetics 16: 1124–1131.
33. Shen J, Ambrosone CB, DiCioccio RA, Odunsi K, Lele SB, et al. (2008) A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. Carcinogenesis 29: 1963–1966.
34. Xu T, Zhu Y, Wei Q-K, Yuan Y, Zhou F, et al. (2008) A functional polymorphism in the miR-146a gene is associated with the risk for hepatocellular carcinoma. Carcinogenesis 29: 2126–2131.
35. Duan S, Mi S, Zhang W, Dolan M (2009) Comprehensive analysis of the impact of SNPs and CNVs on human microRNAs and their regulatory genes. Rna Biology. pp 412–425.
36. Hu Z, Liang J, Wang Z, Tian T, Zhou X, et al. (2009) Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. Human Mutation 30: 79–84.
37. Jazdzewski K, Liyanarachchi S, Swierniak M, Pachucki J, Ringel MD, et al. (2009) Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer. Proceedings of the National Academy of Sciences.
38. Guo H, Wang K, Xiong G, Hu H, Wang D, et al. (2010) A functional varient in microRNA-146a is associated with risk of esophageal squamous cell carcinoma in Chinese Han. Familial Cancer 9: 599–603-603.
39. Okubo M, Tahara T, Shibata T, Yamashita H, Nakamura M, et al. (2010) Association Between Common Genetic Variants in Pre-microRNAs and Gastric Cancer Risk in Japanese Population. Helicobacter 15: 524–531.
40. Zhang C-s, Geng L-y, Zhang J, Zhu W-j, Du L-x (2010) Chicken Polymorphism at Pre-MicroRNAs Inferred from SNP Data. Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on. Chengdu. pp 1–4.
41. Xu B, Feng N-H, Li P-C, Tao J, Wu D, et al. (2010) A functional polymorphism in Pre-miR-146a gene is associated with prostate cancer risk and mature miR-146a expression in vivo. The Prostate 70: 467–472.
42. George G, Gangwar R, Mandal R, Sankhwar S, Mittal R (2011) Genetic variation in microRNA genes and prostate cancer risk in North Indian population. Molecular Biology Reports 38: 1609–1615-1615.
43. Mittal RD, Gangwar R, George GP, Mittal T, Kapoor R (2011) Investigative Role of Pre-MicroRNAs in Bladder Cancer Patients: A Case–Control Study in North India. DNA and Cell Biology.
44. Okubo M, Tahara T, Shibata T, Yamashita H, Nakamura M, et al. (2011) Association Study of Common Genetic Variants in Pre-microRNAs in Patients with Ulcerative Colitis. Journal of Clinical Immunology 31: 69–73-73.
45. Zhou B, Wang K, Wang Y, Xi M, Zhang Z, et al. (2011) Common genetic polymorphisms in pre-microRNAs and risk of cervical squamous cell carcinoma. Molecular Carcinogenesis. pp n/a–n/a.
46. Didiano D, Hobert O (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. Nat Struct Mol Biol 13: 849–851.
47. Muinos-Gimeno M, Montfort M, Bayes M, Estivill X, Espinosa-Parrilla Y (2009) Design and evaluation of a panel of single-nucleotide polymorphisms in microRNA genomic regions for association studies in human disease. Eur J Hum Genet 18: 218–226.
48. Rosenfeld JA, Malhotra AK, Lencz T (2010) Novel multi-nucleotide polymorphisms in the human genome characterized by whole genome and exome sequencing. Nucleic Acids Research.
49. Hansen T, Olsen L, Lindow M, Jakobsen KD, Ullum H, et al. (2007) Brain expressed microRNAs implicated in schizophrenia etiology. PloS one 2: e873.
50. Freudenberg-Hua Y, Freudenberg J, Kluck N, Cichon S, Propping P, et al. (2003) Single Nucleotide Variation Analysis in 65 Candidate Genes for CNS Disorders in a Representative Sample of the European Population. Genome Research 13: 2271–2276.
51. Li S-C, Tang P, Lin W-C (2007) Intronic MicroRNA: Discovery and Biological Implications. DNA and Cell Biology 26: 195–207.
52. Saito Y, Friedman JM, Chihara Y, Egger G, Chuang JC, et al. (2009) Epigenetic therapy upregulates the tumor suppressor microRNA-126 and its host gene EGFL7 in human cancer cells. Biochemical and Biophysical Research Communications 379: 726–731.
53. Baskerville S, Bartel D (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. Rna-a Publication of the Rna Society. pp 241–247.
54. Baker SJ, Sumerson R, Reddy CD, Berrebi AS, Flynn DC, et al. (2001) Characterization of an alternatively spliced AATYK mRNA: expression pattern of AATYK in the brain and neuronal cells. Oncogene 20: 1015–1021.
55. Fitch MJ, Campagnolo L, Kuhnert F, Stuhlmann H (2004) Egfl7, a novel epidermal growth factor-domain gene expressed in endothelial cells. Dev Dyn 230: 316–324.
56. Sevignani C, Calin GA, Nnadi SC, Shimizu M, Davuluri RV, et al. (2007) MicroRNA genes are frequently located near mouse cancer susceptibility loci. Proceedings of the National Academy of Sciences 104: 8017–8022.
57. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, et al. (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proceedings of the National Academy of Sciences of the United States of America 101: 2999–3004.
58. Zhou B, Rao L, Peng Y, Wang Y, Chen Y, et al. (2010) Common genetic polymorphisms in pre-microRNAs were associated with increased risk of dilated cardiomyopathy. Clinica Chimica Acta 411: 1287–1290.
59. Huang W, Liu H, Wang T, Zhang T, Kuang J, et al. (2011) Tonicity-responsive microRNAs contribute to the maximal induction of osmoregulatory transcription factor OREBP in response to high-NaCl hypertonicity. Nucleic Acids Research 39: 475–485.

Priloga B

Genetic variability of microRNA genes in farm animals (Jevšinek Skok in sod., 2012)

COBISS: 1.08
Agris category code: L10

GENETIC VARIABILITY OF microRNA GENES IN FARM ANIMALS

Daša JEVŠINEK SKOK [1], Minja ZORC [1, 2], Simon HORVAT [1, 3], Peter DOVČ [1], Milena KOVAČ [1], Tanja KUNEJ [1]

ABSTRACT

MicroRNAs (miRNAs) are a class of non-coding RNAs that plays an important role in posttranscriptional regulation of target genes. Regulation requires complementarity between the target mRNA and the miRNA seed region, which is responsible for their recognition and binding. Previous studies in human and mouse have shown that variability of miRNA gene (miR-SNPs) might interfere with its function resulting in phenotypic variation. Polymorphisms within miRNA genes could represent biomarkers for phenotypic traits important in farm animals. The aim of this study was to: 1) update previously developed web-based tool for identification of polymorphisms within miRNA genes (miRNA SNiPer), 2) systematically collect polymorphisms of miRNA genes in pig, cattle, chicken, and horse, and 3) experimentally validate SNPs within miRNA seed regions (miR-seed-SNPs) in cattle. Using miRNA SNiPer tool, polymorphisms within 32 mature miRNA regions, including 12 miR-seed-SNPs, were identified in pig, cattle, and chicken. Bovine miR-seed SNPs were chosen for experimental validation. The bta-mir-2313 locus was shown to be very polymorphic, therefore we validated one SNP with previously unknown validation status within the mature seed region in population of Slovenian Simmental cattle. Additionally, two SNPs in corresponding pri-miRNA were identified. Results of this study can serve researchers for follow up hypothesis-driven experimental studies to evaluate the phenotypic effect of identified miRNA genetic variability in vertebrates.

Keywords: farm animals / biomarkers / microRNA / genetic variability

1 INTRODUCTION

MicroRNAs (miRNAs) are non-coding RNA molecules with approximately 21 nucleotides in length that play an important role in posttranscriptional regulation of mRNA. By binding to the different target gene regions, *i.e.,* 3' untranslated region (3'UTR), 5'UTR, promoter, or coding sequences, they repress or activate translation (reviewed in Kunej *et al.*, 2012). MicroRNA biogenesis begins in the nucleus with the primary transcript (pri-miRNA) of several hundreds or thousands base pairs in length that is cleaved by the action of endonuclease DROSHA to 60 to 70 nucleotides long precursor miRNA (pre-miRNA). Pre-miRNA, with its characteristic stem-loop structure (Fig. 1) is then transported to the cytoplasm where endonuclease DICER cleaves both duplex chains to form mature miRNA (Lee *et al.*, 2002; Bartel, 2004). Products of DICER action also include complementary sequences of mature miRNAs, which are referred to as miRNA* (Lau *et al.*, 2001), and are usually transcribed in lower percentage as mature miRNAs (Lim *et al.*, 2003). The key binding location for translational suppression, also called the seed region, resides in the mature miRNA sequence, more accurately situated at position 2–7 or 2–8 nucleotides from the 5' end of the miRNA (Sun *et al.*, 2009).

Changes in the miRNA expression profile were linked with several diseases (reviewed in Ferdin *et al.*, 2011). Moreover, single nucleotide polymorphisms (SNPs) within 1) miRNA genes, 2) miRNA targets or in

1 Univ. of Ljubljana, Biotechnical Fac., Dept. of Animal Science, Groblje 3, SI-1230 Domžale, Slovenia
2 Centre for Mathematical and Computational Biology, Rothamsted Research, Harpenden, UK
3 Dept. of Biotechnology, National Institute of Chemistry, Ljubljana 1000, Slovenia

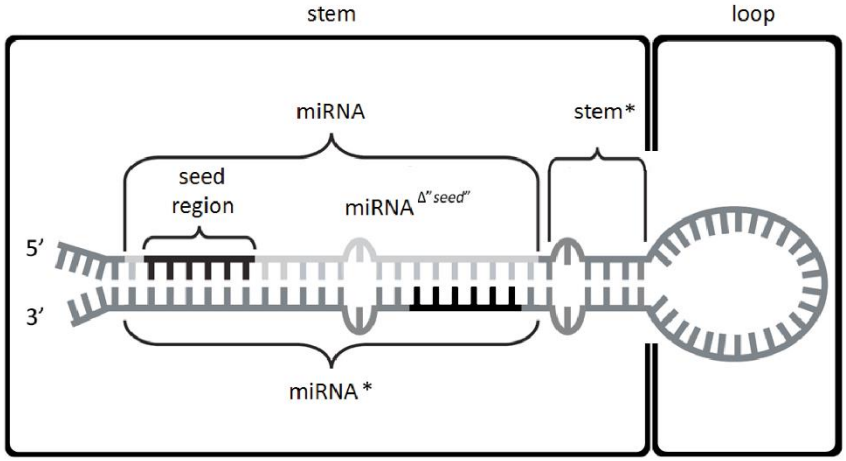


Figure 1: *Secondary structure of miRNA and an example of miRNA stem-loop (adapted from Saunders et al., 2007)*

3) protein-coding genes involved in miRNA biogenesis result in phenotypic differences and therefore affects (associated with) production traits and susceptibility to diseases (Georges *et al.*, 2007). Previously, we developed a web based tool miRNA SNiPer for the identification of polymorphisms residing within miRNA genes (Zorc *et al.*, 2012). Since some farm animals (pig, cattle, chicken, sheep, etc.) represent both – a source of food and disease models, the biomarkers based on miRNA polymorphisms could contribute to the study of phenotypic properties in farm and in medicine. In this study we systematically collected polymorphic miRNA genes from four farm animal species (pig, cattle, chicken, and horse) and further experimentally validated miRNA SNPs in cattle.

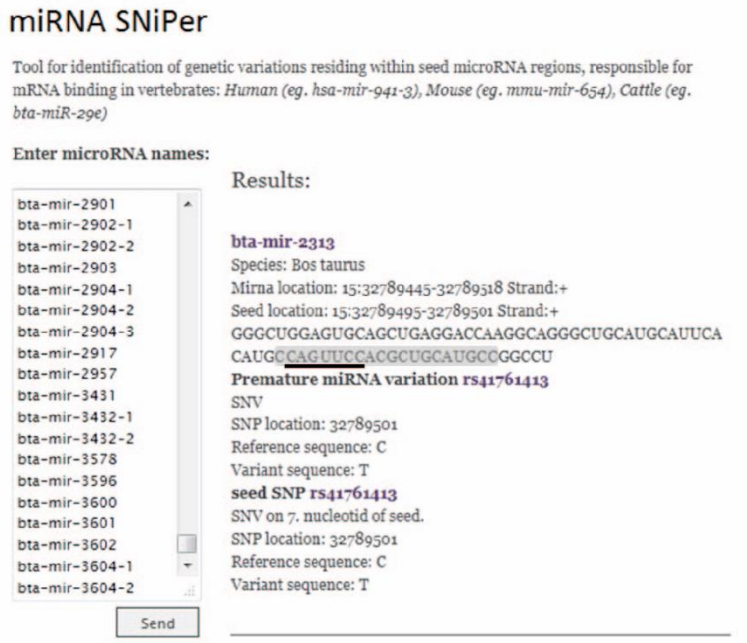


Figure 2: *An example of predicted genetic variability within miRNA in cattle (bta-mir-2313), using a web-based tool miRNA SNiPer. Gray: mature miRNA, underlined region: seed region*

2 MATERIALS AND METHODS

2.1 UPDATE OF THE ONLINE TOOL FOR THE DETECTION OF GENETIC VARIATIONS WITHIN MIRNA GENES.

Previously developed tool miRNA SNiPer was updated with the latest versions of miRBase (release 18; http://www.mirbase.org/) (Kozomara and Griffiths-Jones, 2011), TargetScan (release 5.2; http://www.targetscan.org/) (Lewis *et al.*, 2005), and Ensembl Variation database for pig (Sscrof10.2), horse (EquCab2), cattle (Btau4.0), and chicken (WASHUC2).

2.2 SAMPLES, SEQUENCING OF A DNA PANEL OF SIRES AND VALIDATION OF MIR-SNPS IN CATTLE

Animals were selected from the National progeny test for Slovenian Simmental cattle. DNA samples were extracted from frozen semen of sires using DNeasy Blood & Tissue DNA extraction kit (QiaGen, Düseldorf, Germany). Three bovine miR-seed-SNPs were chosen for experimental validation. Primers were selected using online tool Primer3 (Rozen and Skaletsky, 2000). Due to presence of the *repetitive sequence,* primers in region of SNP rs42658514 within bta-mir-2450c couldn't be designed. For experimental validation, the following primers were used: for bta-mir-2313 forward primer (F) 5'-GCACAGACTCTCAGCCACTG-3' and reverse primer (R) 5'-CTGACTGAGGCTCTCGCTCT-3', and for bta-miR-29e (F) 5'-TGTAGGGACTGGTTGTGGAA-3' and (R) 5'-TCTACTGAACACAGCCCCATC-3'. PCR products were purified using *Exo*I (Exonuclease I) and SAP (shrimp alkaline phosphatase (both Fermentas, Vilnius, Lithuania) and following sequencing reaction for capillary electrophoresis on ABI3130xl.

3 RESULTS AND DISCUSSION

Previously developed web-based tool miRNA SNiPer, designed for the detection of genetic variations within miRNA genes in vertebrates, was updated. The tool accepts a list of miRNA genes and returns a table of variations within different regions of miRNA genes: pre-miRNA, mature, and seed region (Fig. 2).

The number of total known miRNAs and putative SNPs, as well as the list of miR-SNPs for four farm animal species (pig, horse, cattle, and chicken) is presented in table 1. The highest number of SNPs in farm animals is currently known for chicken and cattle, while the numbers for other farm animals are significantly lower. Thirty-two polymorphisms overlapping mature miRNAs, including 12 within seed region (miR-seed-SNPs), were identified. One miR-seed-SNP was identified in pig, three in cattle, and seven in chicken. Bovine miR-seed-SNPs were selected for experimental validation in eight Slovenian Simmental cattle sires.

All collected miR-seed-SNPs have an unknown

Table 1: *Number of known miRNAs and SNPs within four farm species (pig, cattle, chicken, and horse) and list of SNPs within miRNA mature and seed region*

Species	Total number of known miRNAs	Total number of SNPs	No. of SNPs within seed/mature miRNA region	miRNA comprising seed SNP	miR-seed-SNP ID
Pig	228	545.950	1/8	ssc-mir-4335	rs80984906
Cattle	662	2.201.071	3/9	bta-mir-29e	rs41825418
				bta-mir-2313	rs41761413
				bta-mir-2450c	rs42658514
Chicken	499	3.292.991	8/14	gga-miR-1568	rs14511527
				gga-miR-1614*	rs15172520
				gga-miR-1644	rs14076349
				gga-miR-1648*	rs14281065
				gga-miR-1657	rs14934924
				gga-miR-1658	rs16681031
					rs16681032
					rs16681033
Horse	341	1.163.258	0/1	/	/

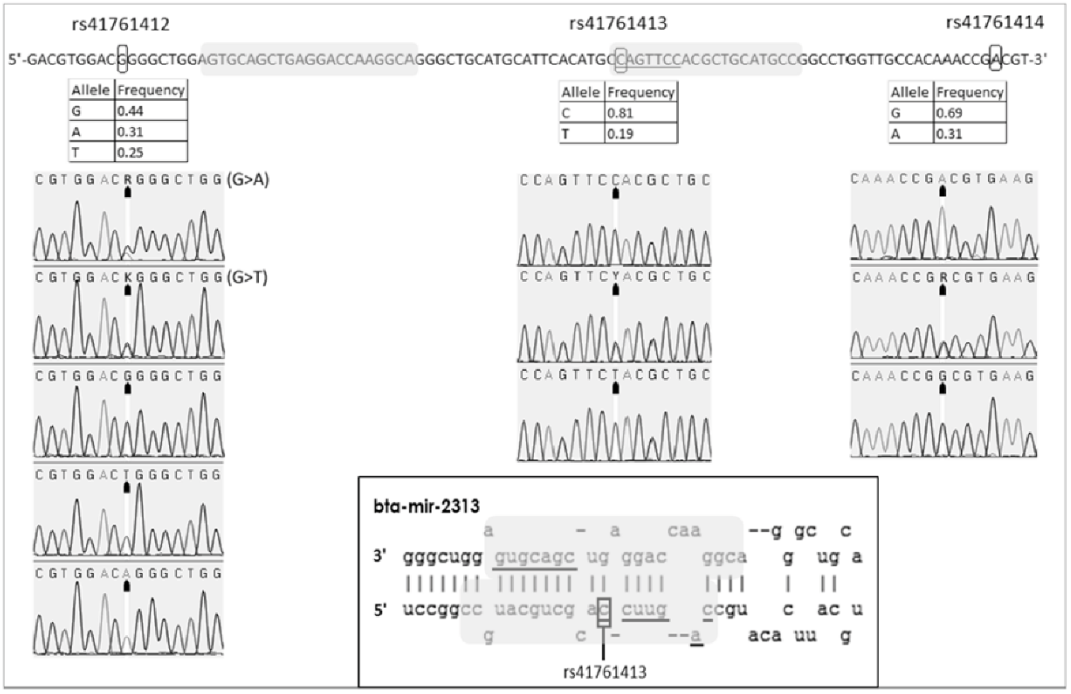


Figure 3: *Validation of three pri-miRNA SNPs located within bta-mir-2313 in cattle. Gray: mature miRNA, underlined region: seed region*

validation status according to the NCBI database and have not been genotyped yet. Experimental validation revealed that polymorphism rs41825418 within bta-mir-29e was monomorphic, while miR-seed-SNP rs41761413 within bta-mir-2313 was polymorphic in analyzed population of Slovenian Simmental cattle (Fig. 3). Additional two SNPs in corresponding pri-miRNA were identified: rs41761412 with poly-allelic substitution (G> T> A) and rs41761414 with substitution (A> G).

4 CONCLUSION

Since miRNA polymorphisms may have a profound effect on a wide range of phenotypes, genetic variability of miRNA genes in farm animals was examined. Bioinformatics tool miRNA SNiPer was updated and used for assembling a list of all known miRNA SNPs in four farm species (pig, horse, cattle, and chicken). Our results show, that most of the miR-SNPs still need to be validated. The project is ongoing, as novel miRNAs and SNPs are yet to be discovered in farm animals, as well as in other animal species. Collated data can be used by interested scientific community to help retrieve valuable information and design efficient experimental plans in the field of miR-SNP research. Association study between validated SNPs located within bta-mir-2313 and carcass traits in paternal half-sib families of the Slovenian Simmental cattle is still under way and current results will be presented. This project may yield new findings useful for development of molecular markers in selection programs allowing more effective, marker assisted selection in farm animals.

5 REFERCES

Bartel D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116: 281–297

Ferdin J., Kunej T., Calin G.A. 2011. MicroRNA: Genomic Association with Cancer Predisposition. Journal of the Association of Genetic Technologists, 37: 11–19

Georges M., Coppieters W., Charlier C. 2007. Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. Current Opinion in Genetics and Development, 17: 166–176

Kozomara A., Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research, 39: D152–D157

Kunej T., Godnic I., Horvat S., Zorc M., Calin G.A. 2012. Cross Talk Between MicroRNA and Coding Cancer Genes. The Cancer Journal, 18: 223–231

Lau N.C., Lim L.P., Weinstein E.G., Bartel D.P. 2001. An abun-

dant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science, 294: 858–862

Lee Y., Jeon K., Lee J.-T., Kim S., Kim V.N. 2002. MicroRNA maturation: stepwise processing and subcellular localization. The EMBO Journal, 21: 4663–4670

Lewis B.P., Burge C.B., Bartel D.P. 2005. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell, 120: 15–20

Lim L.P., Lau N.C., Weinstein E.G., Abdelhakim A., Yekta S., Rhoades M.W., Burge C.B., Bartel D.P. 2003. The microRNAs of Caenorhabditis elegans. Genes and Development, 17: 991–1008

Rozen S., Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology, 132: 365–386

Saunders M.A., Liang H., Li W.H. 2007. Human polymorphism at microRNAs and microRNA target sites. Proceedings of the National Academy of Sciences of the United States of America, 104: 3300–3305

Sun G., Yan J., Noltner K., Feng J., Li H., Sarkis D.A., Sommer S.S., Rossi J.J. 2009. SNPs in human miRNA genes affect biogenesis and function. RNA, 15: 1640–1651

Zorc M., Jevsinek Skok D., Godnic I., Calin G.A., Horvat S., Jiang Z., Dovc P., Kunej T. 2012. Catalog of MicroRNA Seed Polymorphisms in Vertebrates. PloS one, 7: e30737

Priloga C

Obesity gene atlas in mammals (Kunej in sod., 2012)

IVYSPRING INTERNATIONAL PUBLISHER

Journal of Genomics

2012; 1:45-55. doi: 10.7150/jgen.3996

Research Paper

Obesity Gene Atlas in Mammals

Tanja Kunej[1*]✉, Dasa Jevsinek Skok[1*], Minja Zorc[1,2*], Ana Ogrinc[1], Jennifer J. Michal[3], Milena Kovac[1], and Zhihua Jiang[3]✉

1. Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Domzale, Slovenia;
2. Centre for Mathematical and Computational Biology, Rothamsted Research, Harpenden, Herts, AL5, 2JQ, UK;
3. Department of Animal Sciences, Washington State University, Pullman, Washington 99164-6351, USA.

* Authors contributed equally to the work.

✉ Corresponding author: Dr. T. Kunej, Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Domzale, Slovenia, Tel: +386 13203890; FAX: +386 17241005; E-mail address: tanja.kunej@bf.uni-lj.si and Prof. Zhihua Jiang, Department of Animal Sciences, Washington State University, Pullman, Washington 99164-6351, E-mail address: jiangz@wsu.edu.

Published: 2012.02.10

Abstract

Obesity in humans has increased at an alarming rate over the past two decades and has become one of the leading public health problems worldwide. Studies have revealed a large number of genes/markers that are associated with obesity and/or obesity-related phenotypes, indicating an urgent need to develop a central database for helping the community understand the genetic complexity of obesity. In the present study, we collected a total of 1,736 obesity associated loci and created a freely available obesity database, including 1,515 protein-coding genes and 221 microRNAs (miRNAs) collected from four mammalian species: human, cattle, rat, and mouse. These loci were integrated as orthologs on comparative genomic views in human, cattle, and mouse. The database and genomic views are freely available online at: http://www.integratomics-time.com/fat_deposition. Bioinformatics analyses of the collected data revealed some potential novel obesity related molecular markers which represent focal points for testing more targeted hypotheses and designing experiments for further studies. We believe that this centralized database on obesity and adipogenesis will facilitate development of comparative systems biology approaches to address this important health issue in human and their potential applications in animals.

Key words: adipogenesis, fat deposition, integratomics, mammals, microRNA (miRNA), obesity.

Introduction

Obesity is a result of excess body fat accumulation, which often causes adverse effects on human health, such as cardiovascular diseases, type 2 diabetes mellitus, and cancer. Depending on suspected etiology, obesity is commonly classified into three subgroups: monogenic, syndromic, and polygenic or common obesity [1]. Incidence of obesity has dramatically increased over the past few decades and is a global health problem worldwide. In particular, polygenic obesity depends on an individual's genetic makeup that is susceptible to an environment that promotes energy consumption over energy expenditure. As such, search for genes associated with polygenic obesity has focused on multiple interacting alleles contributing to common diseases. Major approaches used to identify novel gene variants associated with polygenic obesity, include candidate gene, genome-wide linkage and genome-wide association studies [2]. Genes that define polygenic obesity have been implicated in a wide variety of biological

functions, such as the regulation of food intake, energy expenditure, lipid and glucose metabolism, and adipose tissue development [3]. There is growing evidence for a relationship between obesity-associated insulin resistance and mitochondrial dysfunction [4].

Characterization of genes responsible for increased fat deposition is needed to develop methods and tools that can identify people at high risk for obesity. This information can also be used to design appropriate targeted therapies that will improve the quality of care of obesity-linked conditions [5]. The identification of obesity candidate loci is also attractive for animal breeding, due to growing consumer demand for products with lower fat content. On the other hand, selection of animals with superior marbling or intramuscular fat (IMF) phenotypes is an important objective for livestock producers since meat with a high marbling content tends to be juicier and is more tender and flavorful than very lean meat [6].

The rapid development of molecular genetic marker technology in recent years has led to the identification of genes that contribute to the genetic variation (so called quantitative trait loci, QTL) of marbling or IMF content in livestock species, and hence to molecular farming by marker-assisted selection [7, 8]. Therefore, the same orthologous genes may have conserved functions in biological or biochemical pathways, and thus explain the same or similar variations of the concordant QTL among different species [9]. Comparative approach may reveal novel candidate genes and functional insights into obesity in human [10]. In a previous study, we collected over 2,000 reports on genes/markers affecting fat phenotypes in several species [5], assigned them to the human orthologous regions and subsequently used the markers for identification of genetic networks associated with various fat and fat-related phenotypes in a Wagyu x Limousin cattle population [11].

Recent discoveries also link the development of obesity to microRNAs (miRNAs) [12]. MiRNAs regulate expression of most genes and play critical roles in different biological processes, such as cell differentiation, proliferation, death, metabolism, and energy homeostasis [13]. Some miRNAs have also been implicated in regulation of adipogenic differentiation [14-27]. These data strongly suggest that miRNAs represent a new class of adipogenic inhibitors that may play a role in the pathological development of obesity [19].

The human obesity gene map was updated yearly from 1996 to 2006 and published in the journal *Obesity* (previously, *Obesity Research*). This map displayed gene/marker information associated with obesity and obesity-related phenotypes using data collected from a variety of sources, such as PubMed using a combination of key words, authors, and journals; continuous reviews of obesity and genetics journals; personal collection of reprints; and papers made available to the authors by colleagues around the world [8]. Although the yearly review was a valuable tool for researchers involved in this field, it was discontinued after 2006. However, an increasing number of studies have been published on obesity and adipogenesis in human as well as in other animals.

The aim of the present study was, therefore, to assemble all available information associated with fat deposition into a publicly available online obesity database. We collected data with obesity associated genomic loci from different sources and species (human, cattle, rat, and mice) in an obesity gene atlas, which includes a database and genomic views. We believe that this database will serve as a valuable source for fine mapping and narrowing QTL regions to a few genes and identification of major pathways involved in obesity. In order to integrate data from different sources, a holistic (map-driven) approach was used and an interactive genomic view of collected obesity candidate loci was developed. The map-based approach reveals positional overlaps between candidate loci, thus allowing complementation of different pieces of information from various species for cloning of positional candidate genes.

Material and Methods

Obesity associated candidate loci identified in human, cattle, rat, and mice were collected from literature and publicly available databases. All relevant publications were identified after searching PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Web of Science (http://apps.webofknowledge.com/), with key phrases, such as gene, genetics, epigenetics, non-coding RNA, microRNA, obesity, adipose tissue, marbling, fat deposition, adipogenesis, human, cattle, rat, and mouse. Obesity-associated loci were also extracted from the following databases: Human obesity gene map (OGM) [8], GeneCards, version 3.07 (http://www.genecards.org), Mouse Genome Informatics, release 4.42 (MGI: http://www.informatics.jax.org). To analyze genomic overlaps of obesity associated loci, QTL were downloaded from Rat Genome Database, release date: October 2011 (RGD: http://rgd.mcw.edu/) [28], and Animal QTL Database, release 15 (http://www.animalgenome.org/cgi-bin/QTLdb/index/). QTL associated with obese phenotypes were selected (body fat, body weight, adiposity, obesity, and diabetes) for cattle,

mouse, and human. Gene nomenclature was based on the HUGO Gene Nomenclature Guidelines (http://www.genenames.org). MiRNA nomenclature was adjusted according to the miRBase, release 18 (http://microrna.sanger.ac.uk/). Ingenuity Pathway Analysis (IPA: Ingenuity Systems® www.ingenuity.com) was used to interpret the obesity candidate genes in the context of biological processes and pathways. The Ingenuity Knowledge Base contains information from databases including: Gene Ontology, Entrez Gene, RefSeq, OMIM, KEGG metabolic pathway information etc. (www.ingenuity.com). IPA analysis was performed twice, once with protein coding genes and miRNAs and the second time only with protein coding genes. Genomic views (graphical overview of the chromosomal locations) of obesity associated loci were constructed using the web-based interactive visualization tool Flash GViewer (http://gmod.org/wiki/Flashgviewer/) developed by the GMOD project.

Results and Discussion

Collecting obesity related data and creating the database

Previous studies have confirmed a large number of genetic loci associated with obesity and adipogenesis in mammals. We retrieved data from the literature and publicly available databases (PubMed, Web of Science, OGM, GeneCards, MGI, and RGD) for 1,515 protein-coding genes and 221 miRNAs in four different species (human, cattle, rat, and mouse). The workflow of data collection is presented in **Figure 1**.

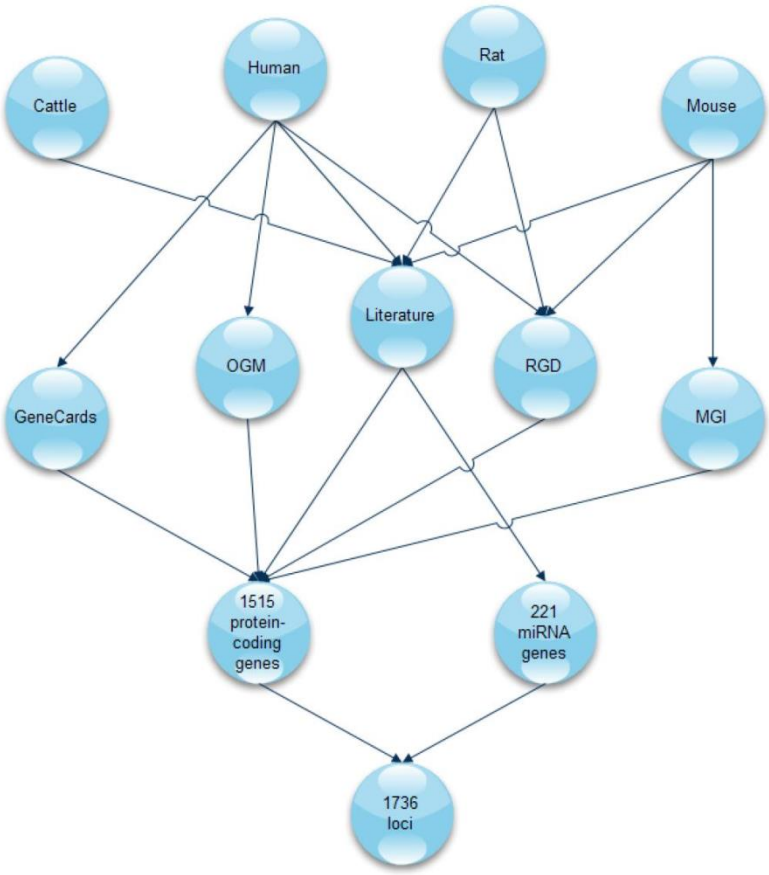


Figure 1: The workflow of data collection. A total of 1,736 loci were retrieved from five databases and current literature for four mammalian species (Rat, Human, Cattle, and Mouse). Genes and miRNAs associated with adipogenesis were retrieved from literature, OGM, Gene Cards, MGI, and RGD. OGM: Obesity gene map, RGD: Rat genome database, MGI: Mouse genome informatics database.

As early as in 2004, miRNAs were identified to play regulatory roles in adipocyte differentiation [14]. Since then, 221 miRNAs have been found to be expressed or disregulated in adipose tissue or adipocytes in human, cattle, rat, and mouse. Those miRNAs were collected in studies using 3T3-L1 pre-adipocytes, adipocytes, adipogenesis, adipose tissue, back fat, diabetes, omental fat and subcutaneous fat, and obesity (**Figure 2**) [14-27]. As shown on a Venn diagram (**Figure 3**), some miRNAs have been found to be associated with fat deposition in more than one species. For example, 14 miRNAs (let-7a, let-7b, let-7c, let-7e, let-7f, mir-103, mir-10b, mir-125a, mir-125b, mir-143, mir-23a, mir-23b, mir-26a, and mir-99b) have been reported to affect fat deposition in all four species. Interestingly, Wang et al. [29] also described involvement of mir-143 in regulation of porcine adipocyte lipid metabolism, which indicates its strong impact on adipogenesis.

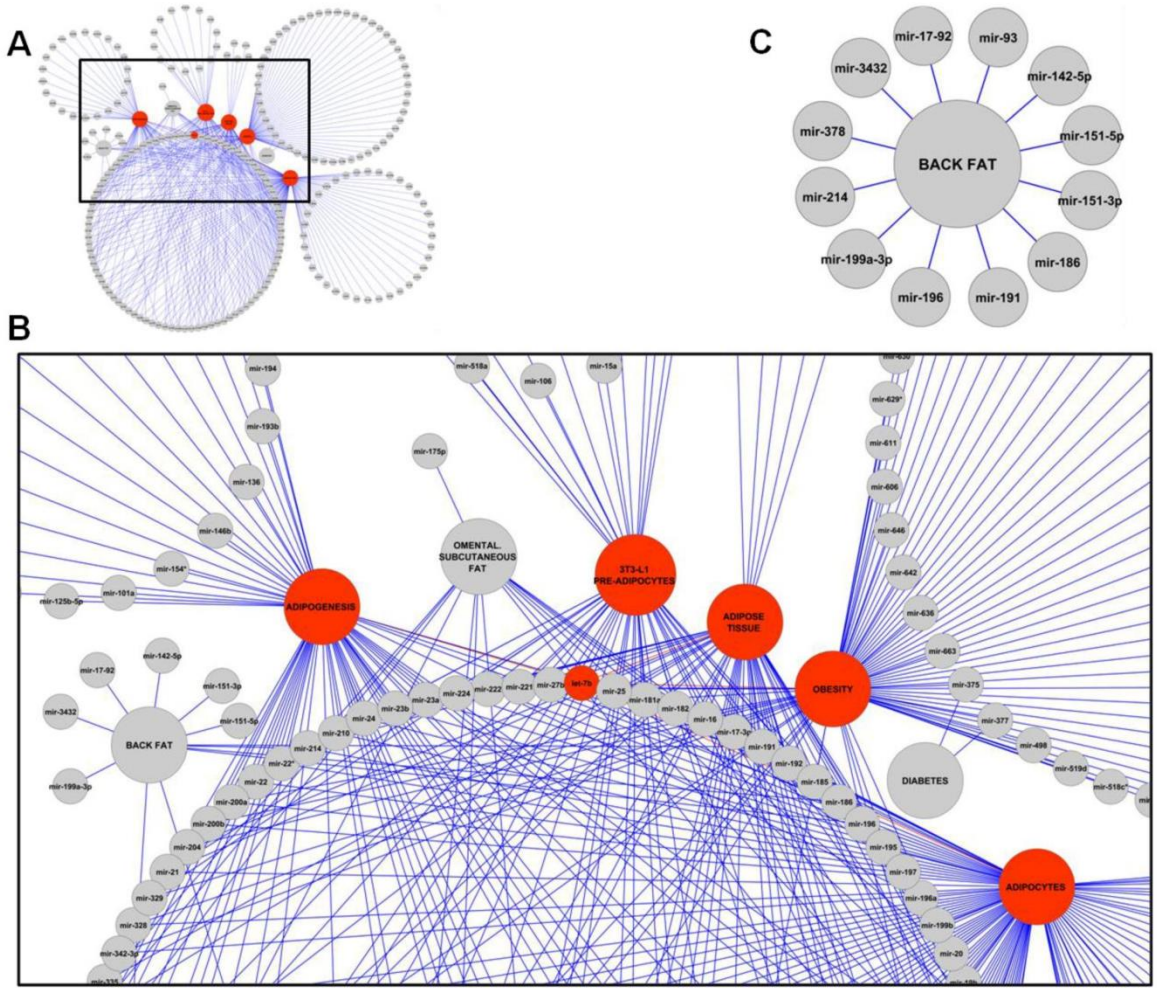


Figure 2: (A) MiRNAs associated with 3T3-L1 pre-adipocytes, adipocytes, adipogenesis, adipose tissue, back fat, diabetes, omental fat and subcutaneous fat, and obesity retrieved from current literature. An example for miRNA let-7b associated with five obesity-associated traits is shown in red. (B) Close up of miRNAs with regulatory roles in obesity-associated traits. (C) MiRNAs associated with back fat.

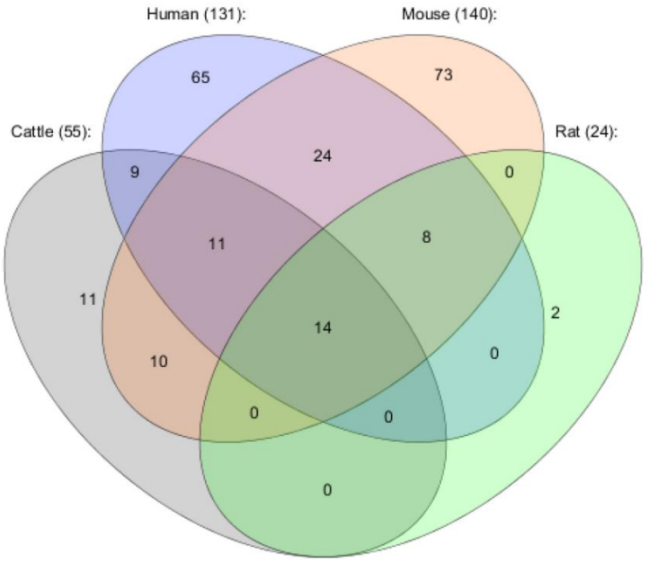


Figure 3: Venn diagram presenting common obesity miRNAs for four species: cattle, human, mouse, and rat.

Based on the data collected from four mammalian species: human, mouse, rat, and cattle, we created an online database to share the information with the community on genetics and genomics of fat deposition/obesity. The database consists of 1,736 obesity-associated loci, including 1,515 protein-coding genes and 221 miRNAs. The database is freely available on the Web site at http://www.integratomics-time.com/fat_deposition. In order to further strengthen the community's involvement in the database expansion and utilization, we also generated an online data entry interface, which will enable users to update or submit new obesity-associated candidate genes and markers. The newly submitted information will be reviewed by curators before releasing to the public (http://www.integratomics-time.com/fat_deposition/add_genes).

The centralized database on genes/markers associated with obesity and adipogenesis will provide fundamental information on what genetic loci are involved in obesity-related traits, how their effects are similar or different among different species and how they interact under different conditions. The knowledge generated from the genetics/genomics studies will provide insight into the molecular basis of obesity and allow for prospective identification of people who have a genetic predisposition to become overweight and/or obese. We believe that the newly constructed database will advance us to the forefront of mapping of QTL for complex traits and understanding their biological pathways by combining a traditional quantitative genetics approach with a modern molecular genetics approach. More importantly, the information obtained in the present study will offer potential and exciting possibilities for future development of successful therapies and new treatments for obesity in humans. On the other hand, due to growing consumer demand for products with lower fat content, an important objective in animal breeding is to improve meat quality traits. Therefore, addition of data related to livestock animal models in our database will have implications in agriculture and biomedicine.

Comparative genomic distribution of fat deposition associated loci

An integrative, comparative-genomics approach allowed us to join obesity-associated information for various species regardless of the study approach, and to present collected loci from human and animals in the form of a single species (human, mouse, and cattle) genomic view available at http://www.integratomics-time.com/fat_deposition/genomic_view/. Genomic view of obesity-associated loci is visible through Flash GViewer provided by the GMOD consortium. Human obesity associated orthologous protein-coding genes, miRNA genes and overlapping QTL are shown in **Figure 4**.

Additionally, a large number of obesity QTL have been identified in domestic and model animal species. To analyze genomic overlaps of obesity associated loci, QTL for cattle, mouse, and human were

downloaded from AnimalQTLdb and RGD. QTL associated with obese phenotype were selected: 469 in human, 355 in cattle and 274 in mouse. Genomic overlap analysis was performed comparing the location of obesity QTL, protein-coding and miRNA genes. Obesity associated orthologs overlapped with 222 human, 176 cattle, and 38 mouse QTL, respectively.

Some candidate genes overlapped with a high number of QTL and therefore could potentially be stronger molecular markers. Three human obesity candidate genes: *CEBPB* (CCAAT/enhancer binding protein (C/EBP), beta), *PTPN1* (protein tyrosine phosphatase, non-receptor type 1), and *SLPI* (secretory leukocyte peptidase inhibitor) each overlapped with 11 different QTL. Gene *ALB* (albumin) overlapped with 16 QTL in cattle. Similarly, using integratomics approach, Kunej *et al.* [25] identified a molecular marker residing within the mmu-mir-717 gene, growth rate associated gene *Gpc3*, and growth associated QTL.

Several studies have also shown that protein-coding host genes are functionally linked with their resident miRNAs [23, 30]. MiRNA genes and their sense oriented host genes can be transcribed from shared promoters [31], whereas antisense orientation suggested that miRNAs and host genes have independent transcription mechanisms [32]. Our analysis showed that among 221 obesity-associated miRNAs, 54 resided within human and 57 within mouse host genes, including some of non-protein-coding genes; large intergenic non-coding RNAs (lincRNAs) (**Supplementary Material: Table S1**). Sense oriented obesity associated miRNA genes from our obesity database overlapped with introns (39 in human and mouse), exons (eight in human and nine in mouse), or 3'-UTRs (two in mouse) of their host genes. Interestingly, three of the host genes have been previously associated with obesity: mir-335 resided within host gene *MEST* (mesoderm specific transcript homolog (mouse)) in human and mouse, mir-378 within *Ppargc1b* (peroxisome proliferator-activated receptor gamma, coactivator 1 beta), and mir-33 within *Srebf2* (sterol regulatory element binding transcription factor 2). In addition to these three miRNA-host gene pairs, other miRNA-host gene pairs warrant further experimental analyses to explore their potential functional link.

Comparative genomics allowed exploitation of animal models for elucidation of obesity phenotype in human. However, extrapolating the gained knowledge from one species to another is often difficult, due to different anatomical and physiological characteristics. Similarly, integratomic/genomic overlap approach was already successfully used for identification of candidate loci for mammary gland associated phenotypes [33] and male infertility [34].

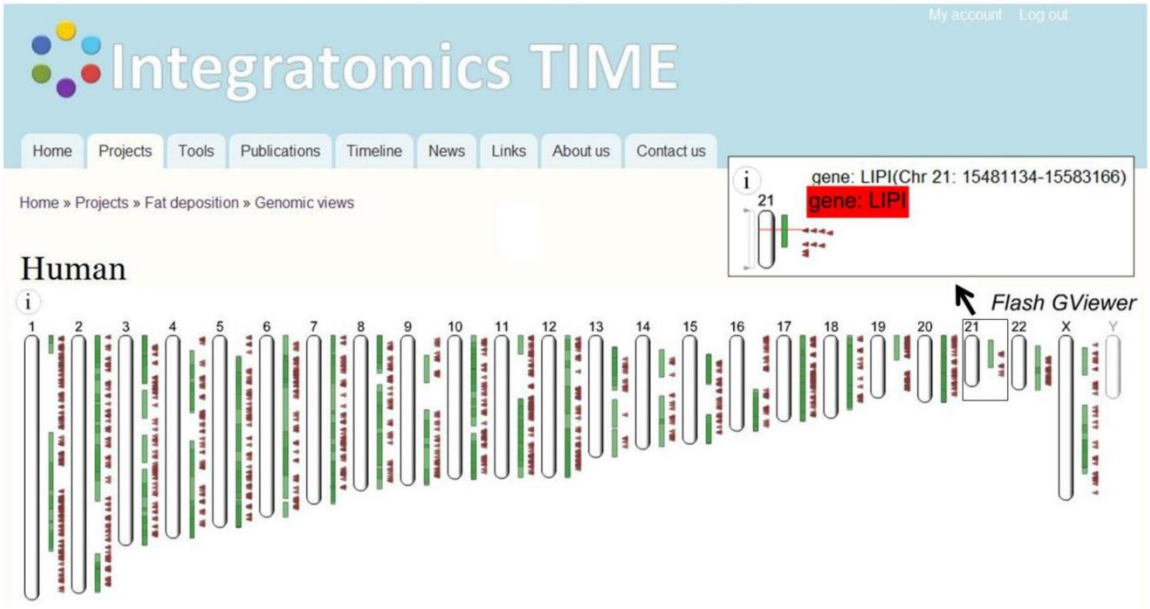


Figure 4: Human chromosomes with obesity-associated loci. Genomic view of the obesity-associated candidate loci presented as human orthologs. Enlargement of the chromosome 21 is showing an overlap between QTL for body weight (BW276_H) and seven obesity related loci *ADAMTS1*, *APP*, *GABPA*, *HSPA13*, *LIPI*, *NRIP1*, and hsa-mir-99a.

Network analysis of obesity associated genes

Human orthologs of obesity candidate genes were integrated into interactome networks using Ingenuity Pathway Analysis (IPA) (**Figure 5**). Input data set consisted of 1,676 genes (protein coding and miRNA genes) and can be accessed on the web page http://www.integratomics-time.com/fat_deposition/genomic_view/. Five of the highest scored networks, diseases and disorders, molecular and cellular functions, as well as canonical pathways are shown in **Figure 5a**. The analysis revealed that obesity candidate genes were associated with nutritional disease, genetic, gastrointestinal, developmental disorder, and cancer. From the total of 25 top networks four of them were found to be associated with lipid metabolism (data not shown). A merged diagram of these four networks includes 140 obesity candidate genes and 175 molecules (**Figure 5b, Supplementary Material: Table S2**). Peroxisome proliferator-activated receptor alpha (PPARα) and retinoid X receptor alpha (RXRα) were identified as central nodes, also called hub molecules. Peroxisome proliferator-activated receptors (PPARs) are nuclear hormone receptors that regulate genes in energy metabolism. Activity of PPARs is dependent on RXR biological activity; PPARs require cognate lipid ligands, heterodimerization with retinoic X receptors (RXR) and coactivation by coactivators [35]. The PPAR-RXR transcriptional complex plays a critical role in energy balance, including triglyceride metabolism, fatty acid handling and storage, and glucose homeostasis: processes whose dysregulation characterize obesity, diabetes, and atherosclerosis [36]. Interestingly, our network analysis of obesity associated genes identifies both of them as central nodes. Moreover, the *PPARα/RXRα* activation pathway was one of the top five canonical pathways identified using IPA analysis.

A

Top Networks

ID		Associated Network Functions	Score
1	View	Behavior, Digestive System Development and Function, Cell Signaling	44
2	View	Lipid Metabolism, Small Molecule Biochemistry, Vitamin and Mineral Metabolism	39
3	View	Cell-To-Cell Signaling and Interaction, Cellular Growth and Proliferation, Hematological System Development and Function	36
4	View	Nucleic Acid Metabolism, Small Molecule Biochemistry, Genetic Disorder	35
5	View	DNA Replication, Recombination, and Repair, Cellular Growth and Proliferation, Reproductive System Disease	35

Top Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Nutritional Disease	1,10E-243 - 1,92E-45	378
Genetic Disorder	2,41E-191 - 8,40E-36	1211
Developmental Disorder	7,35E-149 - 8,38E-40	540
Cancer	8,64E-146 - 8,18E-37	841
Gastrointestinal Disease	2,84E-118 - 2,15E-35	862

Molecular and Cellular Functions

Name	p-value	# Molecules
Lipid Metabolism	1,90E-211 - 1,61E-35	632
Molecular Transport	1,90E-211 - 2,60E-35	778
Small Molecule Biochemistry	1,90E-211 - 3,36E-35	815
Cellular Growth and Proliferation	1,29E-164 - 5,10E-35	780
Cellular Movement	8,21E-156 - 8,90E-36	558

Top Canonical Pathways

Name	p-value	Ratio
Leptin Signaling in Obesity	5,96E-70	71/84 (0,845)
G-Protein Coupled Receptor Signaling	1,82E-58	174/530 (0,328)
PPARα/RXRα Activation	1,3E-57	97/186 (0,522)
Hepatic Cholestasis	3,63E-52	84/176 (0,477)
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	2,63E-50	122/333 (0,366)

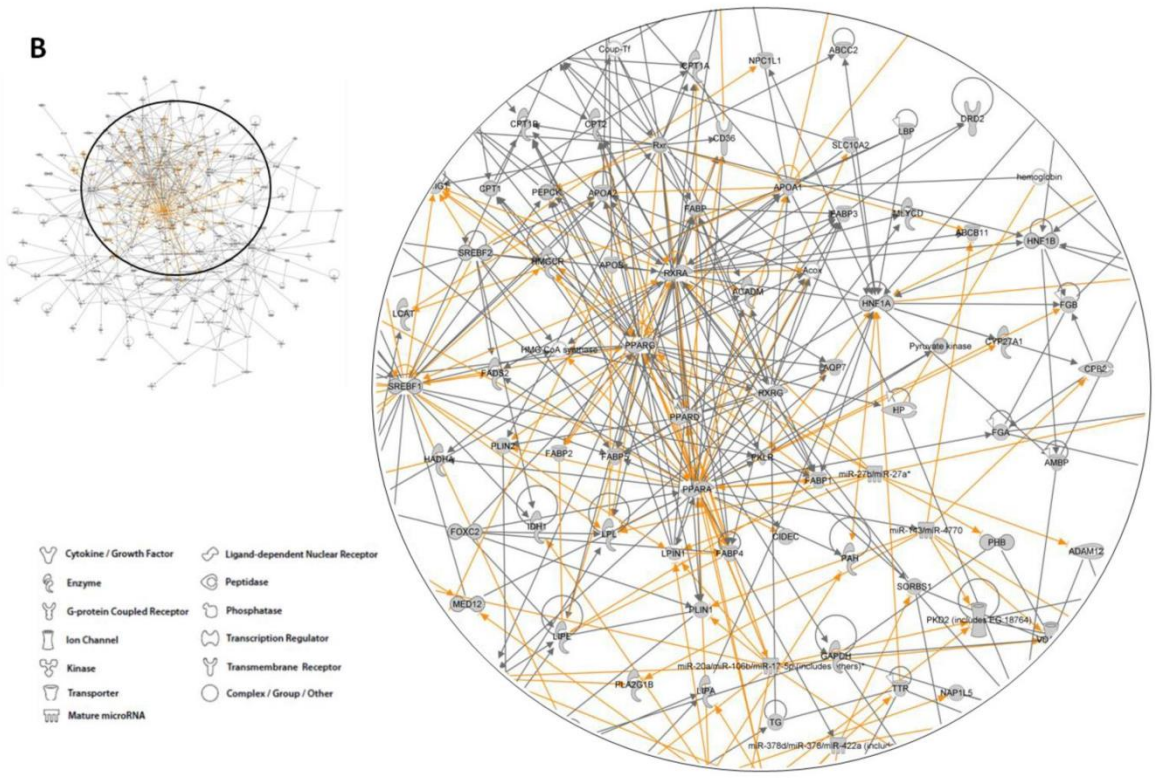


Figure 5: Association of human orthologs of obesity genes with molecular networks, biological functions and canonical pathways according to Ingenuity Pathway Analysis software. The information contained in Canonical Pathways comes from specific journal articles, review articles, text books, and KEGG Ligand (www.ingenuity.com). (A) Five top networks, biological functions and canonical pathways. (B) Pathway analysis of obesity associated loci, shown as diagram of a molecular network associated with lipid metabolism. A network close-up including central nodes PPARα and RXRα is shown on the right.

Canonical pathway *PPARa/RXRa* activation is representative at the cellular level (**Figure 6**) and depicts the involved genes/proteins, their interactions, and the cellular and metabolic reactions in which the pathway is involved. Grey colored molecules represent genes/proteins that appear in the dataset. To avoid miRNA genes that have a bias towards adipogenesis pathway as many of these miRNAs were isolated in studies using 3T3-L1 pre-adipocytes and adipocytes, IPA analysis was repeated with miRNA genes excluded. Interestingly, the same two central nodes were confirmed (data not shown).

Future directions

The centralized obesity database presented in this research will enable processing of collected data in various ways that will lead to faster development of obesity-associated molecular markers for functional studies in human and animals. Our study provides a basis for uncovering miRNA regulatory networks and identification of miRNA targets for combating obesity. For example, analysis of the obesity gene catalog with the miRNA SNiPer tool [37] revealed four human obesity-associated miRNAs (hsa-miR-96, hsa-miR-122-3p, hsa-miR-125a-5p, and hsa-miR-221-5p) that had polymorphisms within the mature miRNA seed region, which is responsible for target binding. Additionally, the central nodes of the molecular networks associated with lipid metabolism, such as PPARα and RXRα as well as genes and miRNAs that overlap with adipogenesis-associated QTL are strong candidates for further experimental validation. Potential obesity markers developed in the present study can help researchers in choosing the most promising biomarkers for further experiments.

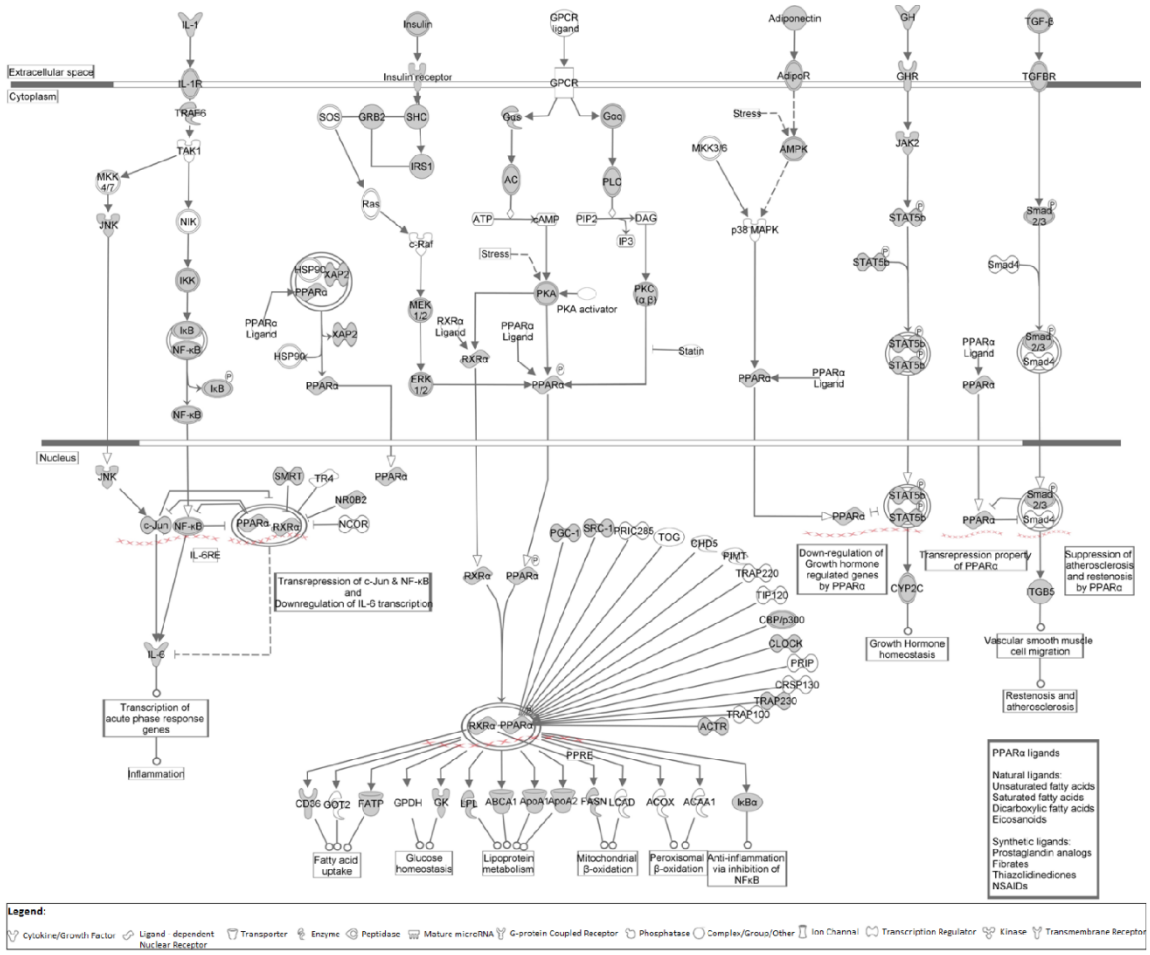


Figure 6: Canonical pathway *PPARα/RXRα* activation shown at the cellular level. Grey colored molecules represent genes/proteins from our dataset.

It is expected that the number of obesity-associated loci will increase; therefore there is also a need to develop new bioinformatics tools for collecting and presenting a large amount of obesity-associated information. An option for updating our central obesity database with new loci associated with fat deposition, including miRNAs and other non-coding RNA genes, and their regulatory mechanisms will allow development of novel biomarkers and will lead to better understanding and consequently more effective treatment and control of such obesity-related disorders both in humans and in animals.

Conclusions

This study presents an integrated resource for obesity candidate genes and miRNAs potentially involved in obese phenotypes and currently includes 1,736 loci associated to obesity in four mammalian species. The centralized online obesity database collects dispersed data in a central location and aims to be an entry point for human and animal obesity research allowing users to retrieve and submit information, which is evaluated by curators. Systems biology approach will contribute to understanding of genetic causes for obesity and also presents a novel approach to study genetic background of complex traits.

Supplementary Material

Table S1: Obesity associated miRNAs (54 human and 57 murine) with corresponding host genes. **Table S2:**

List of 175 molecules included in lipid metabolism network. http://www.jgenomics.com/v01p0045s1.pdf

Acknowledgements

This work was supported by the Slovenian Research Agency (ARRS) through the Research programme Comparative genomics and genome biodiversity (P4-0220).

Conflict of Interest

The authors have declared that no conflict of interest exists.

References

[1] Herrera BM, Lindgren CM. The Genetics of Obesity. Current Diabetes Reports, 2010;10: 498-505.
[2] Choquet H, Meyre D. Molecular basis of obesity: current status and future prospects. Curr Genomics. 2011; 12(3): 154-68.
[3] Mutch DM, Clément K. Unraveling the Genetics of Human Obesity. PLoS Genet. 2006; 2(12): e188.
[4] Kunej T, Wang Z, Michal JJ, Daniels TF, Magnuson NS, Jiang Z. Functional UQCRC1 Polymorphisms Affect Promoter Activity and Body Lipid Accumulation. Obesity. 2007; 15(12): 2896-901.
[5] Jiang Z, Pappu SS, Rothschild MF. Hitting the jackpot twice: identifying and patenting gene tests related to muscle lipid accumulation for meat quality in animals and type 2 diabetes/obesity in humans. Recent Pat DNA Gene Seq. 2007; 1(2): 100-11.
[6] Elías Calles JA, Gaskins CT, Busboom JR, Duckett SK, Cronrath JD, Reeves JJ, Wright R.W. Differences among Wagyu sires for USDA carcass traits and palatability attributes of cooked ribeye steaks. J Anim Sci. 2000; 78(7): 1710-5.
[7] Jiang Z, Kunej T, Michal JJ, Gaskins CT, Reeves JJ, Busboom JR, Dovc P, Wright RW. Significant associations of the mitochondrial transcription factor A promoter polymorphisms with marbling and subcutaneous fat depth in Wagyu x Limousin F2 crosses. Biochem Biophys Res Commun. 2005; 334(2): 516-23.
[8] Rankinen T, Zuberi A, Chagnon YC, Weisnagel SJ, Argyropoulos G, Walts B, Pérusse L, Bouchard C. The human obesity gene map: the 2005 update. Obesity (Silver Spring). 2006; 14(4): 529-644.
[9] Rothschild MF, Hu ZL, Jiang Z. Advances in QTL mapping in pigs. Int J Biol Sci. 2007; 3(3): 192-7.
[10] Jiang Z, Michal JJ, Williams GA, Daniels TF, Kunej T. Cross species association examination of UCN3 and CRHR2 as potential pharmacological targets for antiobesity drugs. PLoS One. 2006; 1:e80.
[11] Jiang Z, Michal JJ, Chen J, Daniels TF, Kunej T, Garcia MD, Gaskins CT, Busboom JR, Alexander LJ, Wright RW, Macneil MD. Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. Int J Biol Sci. 2009; 5(6): 528-42.
[12] Romao JM, Jin W, Dodson MV, Hausman GJ, Moore SS, Guan IL. MicroRNA regulation in mammalian adipogenesis. Exp Biol Med (Maywood). 2011; 236(9): 997-1004.
[13] Bao B, Wang Z, Li Y, Kong D, Ali S, Banerjee S, Ahmad A, Sarkar FH.. The complexities of obesity and diabetes with the development and progression of pancreatic cancer. Biochim Biophys Acta. 2011; 1815(2): 135-46.
[14] Esau C, Kang X, Peralta E, Hanson E, Marcusson EG, Ravichandran LV, Sun Y, Koo S, Perera RJ, Jain R, Freier SM, Bennett CF, Lollo B, Griffey R. MicroRNA-143 regulates adipocyte differentiation. J Biol Chem. 2004; 279(50): 52361-5.
[15] Georges M, Coppieters W, Charlier C. Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. Curr Opin Genet Dev. 2007 Jun;17(3):166-76.
[16] Kajimoto K, Naraba H, Iwai N. MicroRNA and 3T3-L1 pre-adipocyte differentiation. RNA. 2006;12(9): 1626-32.
[17] Klöting N, Berthold S, Kovacs P, Schön MR, Fasshauer M, Ruschke K, Stumvoll M. Blüher M. MicroRNA expression in human omental and subcutaneous adipose tissue. PLoS One. 2009; 4(3): e4699.
[18] Mack GS. MicroRNA gets down to business. Nat Biotechnol. 2007; 25(6): 631-8.
[19] Lin Q, Gao Z, Alarcon RM, Ye J, Yun Z. A role of miR-27 in the regulation of adipogenesis. FEBS J. 2009; 276(8): 2348-58.
[20] Nakanishi N, Nakagawa Y, Tokushige N, Aoki N, Matsuzaka T, Ishii K, Yahagi N, Kobayashi K, Yatoh S, Takahashi A, Suzuki H, Urayama O, Yamada N, Shimano H. The up-regulation of microRNA-335 is associated with lipid metabolism in liver and white adipose tissue of genetically obese mice. Biochem Biophys Res Commun. 2009; 385(4): 492-6.
[21] Ortega FJ, Moreno-Navarrete JM, Pardo G, Sabater M, Hummel M, Ferrer A, Rodriguez-Hermosa JI, Ruiz B, Ricart W, Peral B, Fernández-Real JM. MiRNA expression profile of human subcutaneous adipose and during adipocyte differentiation. PLoS One. 2010; 5(2): e9022.
[22] Takanabe R, Ono K, Abe Y, Takaya T, Horie T, Wada H, Kita T, Satoh N., Shimatsu A, Hasegawa K. Up-regulated expression of microRNA-143 in association with obesity in adipose tissue of mice fed high-fat diet. Biochem Biophys Res Commun. 2008; 376(4): 728-32.
[23] Wang Q, Wang Y, Minto AW, Wang J, Shi Q, Li X, Quigg RJ. MicroRNA-377 is up-regulated and can lead to increased fibronectin production in diabetic nephropathy. FASEB J. 2008; 22(12): 4126-35.
[24] Zhao E, Keller MP, Rabaglia ME, Oler AT, Stapleton DS, Schueler KL, Neto EC, Moon JY, Wang P, Wang IM, Lum PY, Ivanovska I, Cleary M, Greenawalt D, Tsang J, Choi YJ, Kleinhanz R, Shang J, Zhou YP, Howard AD, Zhang BB, Kendziorski C, Thornberry NA, Yandell BS, Schadt EE, Attie AD. Obesity and genetics regulate microRNAs in islets, liver, and adipose of diabetic mice. Mamm Genome. 2009; 20(8): 476-85.
[25] Kunej T, Jevsinek Skok D, Horvat S, Dovc P, Jiang Z. The glypican 3-hosted murine mir717 gene: sequence conservation, seed region polymorphisms and putative targets. International journal of biological sciences. 2010; 6(7): 769-72.
[26] Jin W, Dodson MV, Moore SS, Basarab JA, Guan LL. Characterization of microRNA expression in bovine adipose tissues: a potential regulatory mechanism of subcutaneous adipose tissue development. BMC Mol Biol. 2010; 11: 29.
[27] Martinelli R, Nardelli C, Pilone V, Buonomo T, Liguori R, Castano I, Buono P, Masone S, Persico G, Forestieri P, Pastore L, Sacchetti L. miR-519d Overexpression Is Associated With Human Obesity. Obesity. 2010 Nov;18(11):2170-6.
[28] Twigger SN, Shimoyama M, Bromberg S, Kwitek AE, Jacob HJ, the RGD Team. The Rat Genome Database, update 2007—Easing the path from disease to data and back again. Nucleic Acids Research. 2007; 35(suppl 1): D658-D62.
[29] Wang T, Li M, Guan J, Li P, Wang H, Guo Y, Shuai S, Li X. MicroRNAs miR-27a and miR-143 Regulate Porcine Adipocyte Lipid Metabolism. International Journal of Molecular Sciences. 2011; 12(11): 7950-9.
[30] Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA. 2005; 11(3): 241-7.

[31] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116(2): 281-97.
[32] Li SC, Tang P, Lin WC. Intronic microRNA: discovery and biological implications. DNA Cell Biol. 2007; 26(4): 195-207.
[33] Ogorevc J, Kunej T, Razpet A, Dovc P. Database of cattle candidate genes and genetic markers for milk production and mastitis. Anim Genet. 2009; 40(6): 832-51.
[34] Ogorevc J, Dovc P, Kunej T. Comparative genomics approach to identify candidate genetic loci for male fertility. Reprod Domest Anim. 2011; 46(2): 229-39.
[35] Haemmerle G, Moustafa T, Woelkart G, Büttner S, Schmidt A, van de Weijer T, Hesselink M, Jaeger D, Kienesberger PC, Zierler K, Schreiber R, Eichmann T, Kolb D, Kotzbeck P, Schweiger M, Kumari M, Eder S, Schoiswohl G, Wongsiriroj N, Pollak NM, Radner FP, Preiss-Landl K, Kolbe T, Rülicke T, Pieske B, Trauner M, Lass A, Zimmermann R, Hoefler G, Cinti S, Kershaw EE, Schrauwen P, Madeo F, Mayer B, Zechner R. ATGL-mediated fat catabolism regulates cardiac mitochondrial function via PPAR-α and PGC-1. Nat Med. 2011; 17(9): 1076-85.
[36] Plutzky J. The PPAR-RXR Transcriptional Complex in the Vasculature. Circulation Research. 2011; 108(8): 1002-16.
[37] Zorc M, Jevsinek Skok D, Godnic I, Calin GA, Horvat S, Jiang Z, Dovc P, Kunej T. Catalog of microRNA seed polymorphisms in vertebrates. PloS One. 2012; 7(1): e30737.

Author Biography

Dr. Tanja Kunej is an assistant professor at the Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Slovenia. Dr. Kunej is co-author of over 30 publications and has received two US patent awards. Her main research interest is integratomics in agriculture and biomedicine.

Dasa Jevsinek Skok is a PhD student of Bioscience - Bioinformatics on Biotechnical Faculty, University of Ljubljana, Slovenia. She graduated from Animal Science at the same University. She is employed as a researcher at the Department of Animal Science, Unit for pig production. Her main research interests are bioinformatics and genetics.

Minja Zorc is a PhD student of Biomedicine, University of Ljubljana, Slovenia. She graduated from Computer and information science at the same University. Her main research interest is integratomics. She is employed as a bioinformatics scientist at Rothamsted Research (BBSRC) Harpenden, UK.

Ana Ogrinc is an MSc student of Biotechnology, University of Ljubljana, Slovenia. She graduated from Biotechnology at the same University.

Jennifer J. Michal is an Associate in Research in the Department of Animal Sciences at Washington State University. She received her M.S. in Animal Sciences from the same University. She is co-author of over 40 publications.

Dr. Milena Kovac is a Professor at the Department of Animal Science, Unit for pig production, Biotechnical Faculty, University of Ljubljana, Slovenia. Dr. Kovac's main research interests are statistical methods and software development for genetic evaluation in livestock (PEST, PeRun, VCE), applications of genetic evaluation in livestock, pig breeding and husbandry.

Dr. Zhihua Jiang is an Associate Professor in the Animal Sciences Department at Washington State University. Since the early 1990's, Dr. Jiang's research has focused on comparative genome biology with aims at understanding how orthologous genes have evolved and shuffled during evolution. The overall objective of his research program is to develop comparative genomic tools and reagents for determining gene sequence, location, expression and function, thus advancing genome sciences and their applications in agriculture and biomedicine. Dr. Jiang has published more than 110 publications that were well cited in the field. He was editor of the book "Reproductive Genomics in Domestic Animals" Published by Wiley-Blackwell in 2010. Dr. Jiang has received eight US patent awards.

Priloga D

Pivotal role of the muscle-contraction pathway in cryptorchidism and evidence for genomic connections with cardiomyopathy pathways in RASopathies (Cannistraci in sod., 2013)

Cannistraci et al. BMC Medical Genomics 2013, 6:5
http://www.biomedcentral.com/1755-8794/6/5

BMC Medical Genomics

RESEARCH ARTICLE Open Access

Pivotal role of the muscle-contraction pathway in cryptorchidism and evidence for genomic connections with cardiomyopathy pathways in RASopathies

Carlo V Cannistraci[1,2,3*†], Jernej Ogorevc[4†], Minja Zorc[4†], Timothy Ravasi[1], Peter Dovc[4] and Tanja Kunej[4*]

Abstract

Background: Cryptorchidism is the most frequent congenital disorder in male children; however the genetic causes of cryptorchidism remain poorly investigated. Comparative integratomics combined with systems biology approach was employed to elucidate genetic factors and molecular pathways underlying testis descent.

Methods: Literature mining was performed to collect genomic loci associated with cryptorchidism in seven mammalian species. Information regarding the collected candidate genes was stored in MySQL relational database. Genomic view of the loci was presented using Flash GViewer web tool (http://gmod.org/wiki/Flashgviewer/). DAVID Bioinformatics Resources 6.7 was used for pathway enrichment analysis. Cytoscape plug-in PiNGO 1.11 was employed for protein-network-based prediction of novel candidate genes. Relevant protein-protein interactions were confirmed and visualized using the STRING database (version 9.0).

Results: The developed cryptorchidism gene atlas includes 217 candidate loci (genes, regions involved in chromosomal mutations, and copy number variations) identified at the genomic, transcriptomic, and proteomic level. Human orthologs of the collected candidate loci were presented using a genomic map viewer. The cryptorchidism gene atlas is freely available online: http://www.integratomics-time.com/cryptorchidism/. Pathway analysis suggested the presence of twelve enriched pathways associated with the list of 179 literature-derived candidate genes. Additionally, a list of 43 network-predicted novel candidate genes was significantly associated with four enriched pathways. Joint pathway analysis of the collected and predicted candidate genes revealed the pivotal importance of the muscle-contraction pathway in cryptorchidism and evidence for genomic associations with cardiomyopathy pathways in RASopathies.

(Continued on next page)

* Correspondence: kalokagathos.agon@gmail.com; tanja.kunej@bf.uni-lj.si
†Equal contributors
[1]Integrative Systems Biology Laboratory, Biological and Environmental Sciences and Engineering Division, Computer, Electrical and Mathematical Sciences and Engineering Division, Computational Bioscience Research Center, King Abdullah University for Science and Technology (KAUST), Thuwal, Saudi Arabia
[4]Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Domzale, Slovenia
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: The developed gene atlas represents an important resource for the scientific community researching genetics of cryptorchidism. The collected data will further facilitate development of novel genetic markers and could be of interest for functional studies in animals and human. The proposed network-based systems biology approach elucidates molecular mechanisms underlying co-presence of cryptorchidism and cardiomyopathy in RASopathies. Such approach could also aid in molecular explanation of co-presence of diverse and apparently unrelated clinical manifestations in other syndromes.

Keywords: Cryptorchidism, Muscle-contraction pathway, Cardiomyopathy, Comparative integratomics, Protein-protein interactions, Systems biology, Undescended testes, RASopathy

Background

Cryptorchidism (CO) is the most frequent congenital disorder in male children (2-4% of full-term male births) and is defined as incomplete descent of one (unilateral) or both (bilateral) testes and associated structures. Cryptorchidism has a potential effect on health; defects in testes descent usually cause impaired spermatogenesis, resulting in reduced fertility and increased rates of testicular neoplasia, and testicular torsion (reviewed in [1]). Cryptorchidism is common in human, pigs, and companion animals (2–12%) but relatively rare in cattle, and sheep (≤ 1%) [2].

Testicular descent is a complex series of events which requires concerted action of hormones, constitutive mechanisms, and the nervous system. In most species, including human, the complete descent of testes usually occurs prenatally, while in some (*e.g.* dogs), postnatally. Beside environmental factors like endocrine disruptors, CO is at least in part determined by genetic causes (chromosome or gene mutations), and is often a common feature of different syndromes. For example, Klinefelter syndrome and mutations in *INSL3* gene have already been recognized as a cause of CO in some cases [3].

The comparative knowledge attained through study of animal models has been of great importance in understanding complex disease etiology, suggesting several candidate genes involved also in the pathogenesis of human diseases [4]. Therefore, the use of comparative genomics approach, integrating and cross-filtering the available knowledge from different species seems highly justified. Different animal models for CO exist; for example natural mutants or transgenic mice, rat, rabbit, dog, pig and rhesus monkeys are used to elucidate the role of different factors involved in CO [5]. Based on mouse knock-out models from Mouse Genome Informatics (MGI) database, several genes appear as possible candidates (*AR*, *HOX* genes, *INSL3*, *RXFP2*, and *WT1*). Additionally, the technological progress in the last years enabled the use of high-throughput omics-information, at coding (DNA), expression (RNA), and proteomic level. This technological revolution creates a vast amount of data, which increases the need for application of bioinformatics tools that are able to connect omics data with phenotype and enable search for overlapping pathogenetic mechanisms in different genetic diseases [6]. However, this existing technology hasn't been significantly employed in human CO research on a genome and transcriptome-wide scale; to date only one genome-wide expression study has been performed in rat [7].

Integratomics represents a novel trend in the omics-research and is based on the integration of diverse omics-data (genomic, transcriptomic, proteomic, *etc.*), regardless of the study approach or species [8-10]. High genetic homology between mammals and the availability of well annotated genomes from different species allows the assembled data to be presented in a form of a comparative genomic view, displaying candidate genes as a single species orthologs.

Information extracted from diverse and methodologically focused studies are often fragmented and controversial. To overcome this problem we integrated the collected data, using a holistic (map-driven) approach, and developed freely available interactive genomic visualization tool. Such map-based approach allows identification and prioritization of candidate genes [11] based on a number of literature sources (references), genomic position, and pathway analyses, employing all currently available knowledge in different species. However, extrapolating the gained knowledge from one species to another is often difficult due to different anatomical and physiological characteristics, which should be considered when comparing pathology of the disease in different species.

To identify genetic factors potentially involved in CO pathogenesis in human we 1) applied comparative integratomics approach and assembled the database of all CO-associated genomic loci reported in the literature, regardless of the study approach and species, 2) presented the loci on a genomic map as human orthologs, and 3) prioritized the collected data using systems biology approach. The collected candidate genes were classified in corresponding biological pathways and the most significant CO-enriched pathways were proposed. Such classification of candidate genes allowed us to prioritize biological pathways (characterized by genes involved in the

pathogenesis of CO), which revealed importance of several pathways (for example muscle contraction mechanisms) that may also play a role in the pathogenesis of other clinical features distinctive for different syndromes often concurrent with CO. In order to additionally illuminate the CO-associated pathways we performed a network-based protein-protein interaction analysis, which resulted in prediction of 43 additional CO candidate genes.

Methods

In search for CO associated candidate loci seven different research approaches were considered: (i) chromosomal abnormalities associated with CO, (ii) copy number variations, (iii) clinical syndromes with known genetic mutations that feature CO, (iv) transgenes and knock-outs that result in CO associated phenotypes, (v) association studies/mutation screening that show association between sequence variation/mutation screening and CO, (vi) expression patterns associated with CO, and (vii) candidates associated with CO at proteomic level.

Data mining

We reviewed the literature published up to 9/2012 searching for the relevant publications through PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and Web of Science (http://isiknowledge.com) using key phrases: genetics, gene candidates, cryptorchidism, testicular descent, undescended testes, male infertility, QTL, microarray, association, microRNA, non-coding RNA, epigenetic, reproduction, and assisted reproduction. CO-associated candidate genes from different sources and species were retrieved from the literature search. Human clinical syndromes that may cause or feature CO were retrieved from Online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim) and Disease database (http://www.diseasesdatabase.com/). The data for CO-related experiments on mouse models were retrieved from the Mouse Genome Informatics (MGI) database (http://www.informatics.jax.org/). Human orthologs for the CO associated genes were extracted from the MGI database, which contains information about mammalian ortholog genes for different species. Overlap analysis of the CO candidate genes with genomic regions involved in chromosome mutations was performed using data retrieved from Ensembl via BioMart data mining tool.

Database implementation

CO-associated candidate genes database is a web resource, which provides integrated and curated information on molecular components involved in the pathogenesis of CO. Information regarding collected CO-associated candidate genes has been stored in relational MySQL database, which is publicly available for search, data entry and update at http://www.integratomics-time.com/cryptorchidism/. Search interface enables users to find specific CO-associated candidate genes based on the number of criteria. Online data entry interface enables users to update or submit new CO-associated candidate genes.

Genomic view of the CO associated loci

Overview of the chromosomal locations of CO associated loci is graphically represented in genomic view, as previously described [12]. It is possible to visualize the literature-collected and network-predicted CO genes on the same genomic view or separately. Genomic view is visible through the web-based interactive visualization tool Flash GViewer (http://gmod.org/wiki/Flashgviewer/), which was developed by the GMOD project.

Pathway and network analysis

In the first pathway analysis we considered human orthologs of the literature-collected candidate genes (179 genes). DAVID Bioinformatics Resources 6.7 [13] was employed for the enrichment (overrepresentation) analysis. The background for the analysis was defined using the 179 candidate genes plus their first neighbours (5018 proteins) selected in the human protein-protein interaction network (PPIN). The result of the enrichment analysis was obtained using Bonferroni multiple test correction and a p-value significant threshold of 0.01. The human PPIN was obtained by fusion of the following human networks: IRefIndex [14], Chuang *et al.* article [15], Ravasi *et al.* article [16], ConsensusPathDB [17].

A new cohort of 43 candidate genes was predicted using PiNGO 1.11 [18]. PiNGO is a tool designed to find candidate genes in biological networks and it is freely provided as a plug-in for Cytoscape 2.8 [19], which is an open source software platform for visualizing and integrating molecular interaction networks. PiNGO predicts the categorization of a candidate gene based on the annotations of its neighbors, using enrichment statistics. In our analysis we quested which first-neighbour-genes significantly interact with the original cohort of 179 literature-collected genes in the human PPIN. We adopted: hypergeometric statistical test, Bonferroni multiple testing correction and p-value significant threshold of 0.01. The cohort of 43 network-predicted genes resulted strongly significant (Bonferroni p-value < 0.0095) for being new candidate genes.

In order to evaluate the importance of this new cohort of 43 candidate genes we performed the pathway analysis according to the procedure already described for the 179 literature-collected candidate genes.

Finally, in order to investigate the biological relations between the 179 literature-collected and 43 network-predicted genes, we repeated the pathway analysis in DAVID (using the same procedure previously described) considering the 222 (179 + 43) candidate genes. The background for the analysis was defined using the 222 candidate genes plus their first neighbours in the human PPIN. In addition, we visualized the protein-protein interactions occurring between the genes present in at least two pathways using the STRING database (version 9.0) [20] and selecting only interactions with high confidence score.

Genetic variability of candidate genes

Genetic variability for the most promising CO candidate genes was extracted from the Ensembl database (http://www.ensembl.org/). Probably damaging genetic variations were predicted by PolyPhen-2, version 2.1.0, provided by Ensembl database. Putative polymorphic miRNA target sites in candidate genes were obtained from Patrocles database (http://www.patrocles.org/) [21].

Results and discussion

Extensive literature mining was performed resulting in 217 collected candidate loci (chromosome regions and genes) reported to be involved in CO in human or/and animals. The generated database served as the foundation for the development of freely available interactive genomics viewer designed to integrate multi-species data from various research approaches. Enriched biological pathways and 43 additional CO candidate genes were suggested, based on protein-protein interaction network (PPIN) analysis. The workflow of the study is presented in the Figure 1.

Collection of the cryptorchidism associated loci from the literature

The collected data incorporates genomic loci associated with cryptorchidism by seven different types of research approaches (chromosomal mutations, copy number variations, clinical syndromes, transgenes and knock-outs, association studies/mutation screening, transcriptomic/expression studies, and proteomic studies). The collected

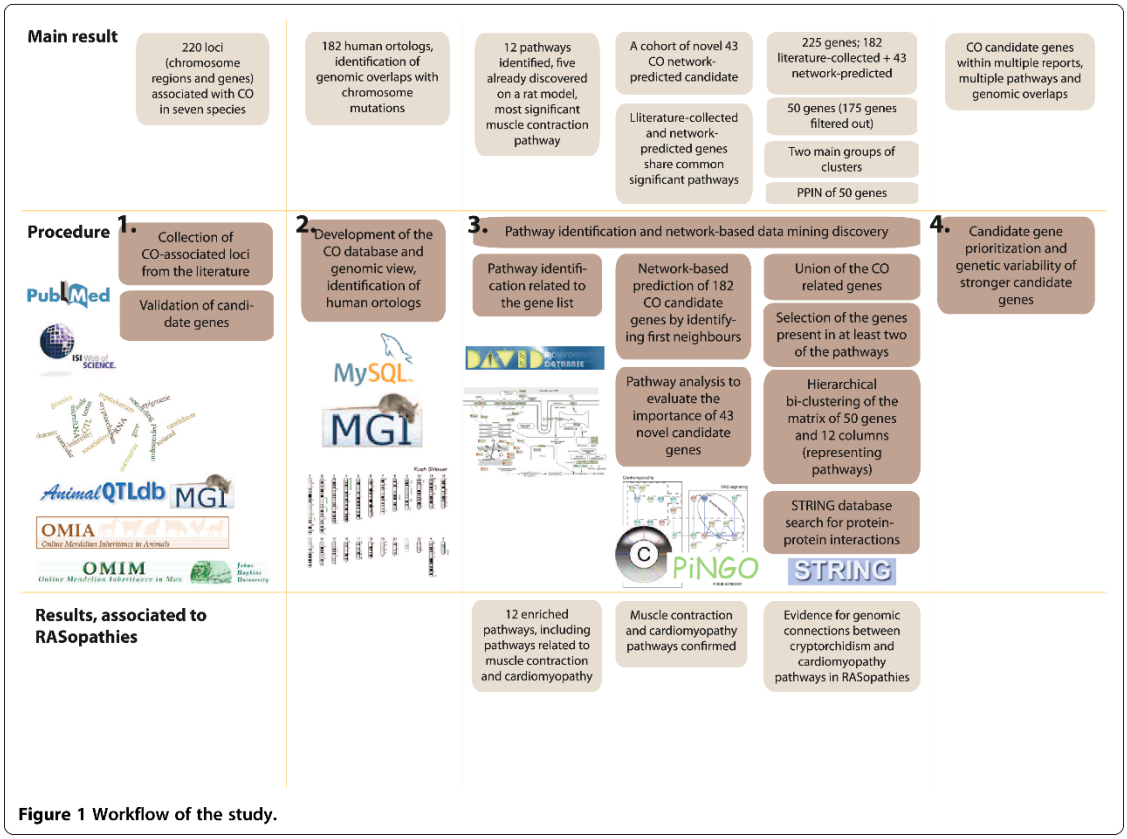


Figure 1 Workflow of the study.

data originates from seven different species (human, cattle, horse, sheep, dog, rat, and mouse) (Table 1). The collected CO data is available in Additional file 1: Table S1, Additional file 2: Table S2, Additional file 3: Table S3, Additional file 4: Table S4 and Additional file 5: Table S5 and include physical locations of the candidate loci in human and species of origin.

Chromosomal aberrations and copy number variations

We reviewed studies reporting 32 different chromosomal mutations including numerical and structural aberrations associated with cryptorchidism [22-33]. Additionally, two *de novo* copy number variations (CNVs) - microduplications were found to be associated with CO using array-based comparative genomic hybridization (aCGH) [34]. The collected data is available in Additional file 1: Table S1.

Clinical syndromes

Studies of complex disease traits can be facilitated by analysis of the molecular pathways represented by genes responsible for monogenic syndromes that also exhibit these traits [7,35]. There are over 200 different human syndromes with known molecular basis in OMIM database that feature "cryptorchidism" or "undescended testis" as a possible feature in their clinical synopsis. Since cryptorchidism phenotype prevalence is low in some syndromes, and could only occur coincidentally, it is difficult to justify association of syndrome causative genes with a particular phenotype.

To collect CO candidate genes (Additional file 2: Table S2) we obtained list of syndromes from the literature [4,36,37], OMIM and Diseases database ("may be caused or feature") and then further examined phenotype-gene relationships and clinical features for each of the syndromes. Only syndromes where cryptorchidism is present as a regular feature, described in multiple clinical cases, and where gene(s) causing the syndrome is/are known were included.

Table 1 The summary of CO associated candidate loci

Locus type / study approach	Number of loci
DNA level	
Chromosomal aberrations	32
Copy number variants (CNVs)	2
Clinical syndromes	42
Knock-out and transgenic experiments	40
Association studies	12
RNA level	
Expression study	112
Protein level	
Injection of exogenous protein	1
Total	217*

* Unique loci (individual locus/gene reported by multiple studies was counted only once).

Transgenes and knock-outs

From the Mouse Genome Informatics (MGI) database and the literature [38-42] we retrieved 39 mouse and one rat KO and transgenic experiments that result in phenotypes associated with CO (Additional file 3: Table S3).

Association studies/mutation screening

Nine genes (*AR*, *BMP7*, *ESR1*, *HOXA10*, *INSL3*, *KISS1R*, *NR5A1*, *RXFP2*, and *TGFBR3*) in human [43-59], *INSL3* in sheep [60] and dog [61], and *COL2A1* in dog [62] showed positive association between sequence polymorphisms/mutations and CO susceptibility (Additional file 4: Table S4). In the case of androgen receptor (*AR*) gene Ferlin *et al.* [45] found no difference between the numbers of CAG and GGC repeats, resulting in variable lengths of PolyGln/PolyGly in the *AR* gene and cryptorchidism; however, it has been proposed that a particular combination of the PolyGln/PolyGly polymorphisms may be linked to CO. In some cases opposing results have been found; for example, no association between the sequence polymorphisms and CO have been found for the genes *ESR1* [63-65], *INSL3* [66-68], *HOXA10* [69], and *RXFP2* [70]. The *LHCGR* has been excluded as a CO candidate gene in an association study in men [71], although KO of this gene in mice showed cryptorchid phenotype (MGI) and is causative gene of Leydig cell hypoplasia-a syndrome that features CO as one of the clinical signs (OMIM). In addition, Y chromosome microdeletions have been found to be present in patients with CO, but are not likely to be a common etiological cause of CO [72-74].

Expression patterns

There are several studies comparing expression profiles in testes between cryptorchid and normal males investigating the resulting effects of but not causes for development of CO (*e.g.* [75,76]). To our knowledge, there is only one microarray study that analyzed transcript profiles in gubernaculum during normal and abnormal testicular descent and reported 3589 differentially expressed genes between inherited cryptorchydism orl rats and a control group [7]. We included a subset of 112 promising candidate genes to our candidate gene list that were selected by the authors of the study based on expression levels, inclusion in specific pathways of interest and/or previous reports showing association with cryptorchidism (Additional file 5: Table S5).

Protein level

Hutson *et al.* (1998) [77] investigated the effect of exogenous calcitonin gene-related peptide (CGRP) in

neonatal pigs. They found that exogenous CGRP, in pigs also known as calcitonin gene-related peptide B (CALCB), stimulated migration of inguinal testes that had been arrested in the line of descent, while ectopic testes did not respond. The results support the role for this protein in testicular descent. However, mutation screening performed by Zuccarello *et al.* (2004) [78] failed to confirm *CGRP* (in human also known as *CALCA*) pathway genes as a major players in human sporadic CO.

Development of the CO database and genomic viewer

The CO-associated loci, obtained by comparative integratomics approach, were assembled into a freely accessible database available at http://www.integratomics-time.com/cryptorchidism/. The curated database is open for public data entry. Researchers are invited to submit new cryptorchidism candidate genes from their research or other publications by filling data entry form on our web site. The collected loci from human and animal species were presented as genomic view for human orthologs (in a form of a human genomic view) (Figure 2).

Some candidate genes have been associated with CO by multiple independent literature reports in multiple species. For example, twenty genes (*AMH, AMHR2, AR, ARID5B, BMP7, EPHA4, ESR1, FGFR2, HOXA10, HRAS, INSL3, LHCGR, MAP2K1, MSX1, NR5A1, RXFP2, SOS1, TNNI2, TNNT3*, and *WT1*) have been associated with

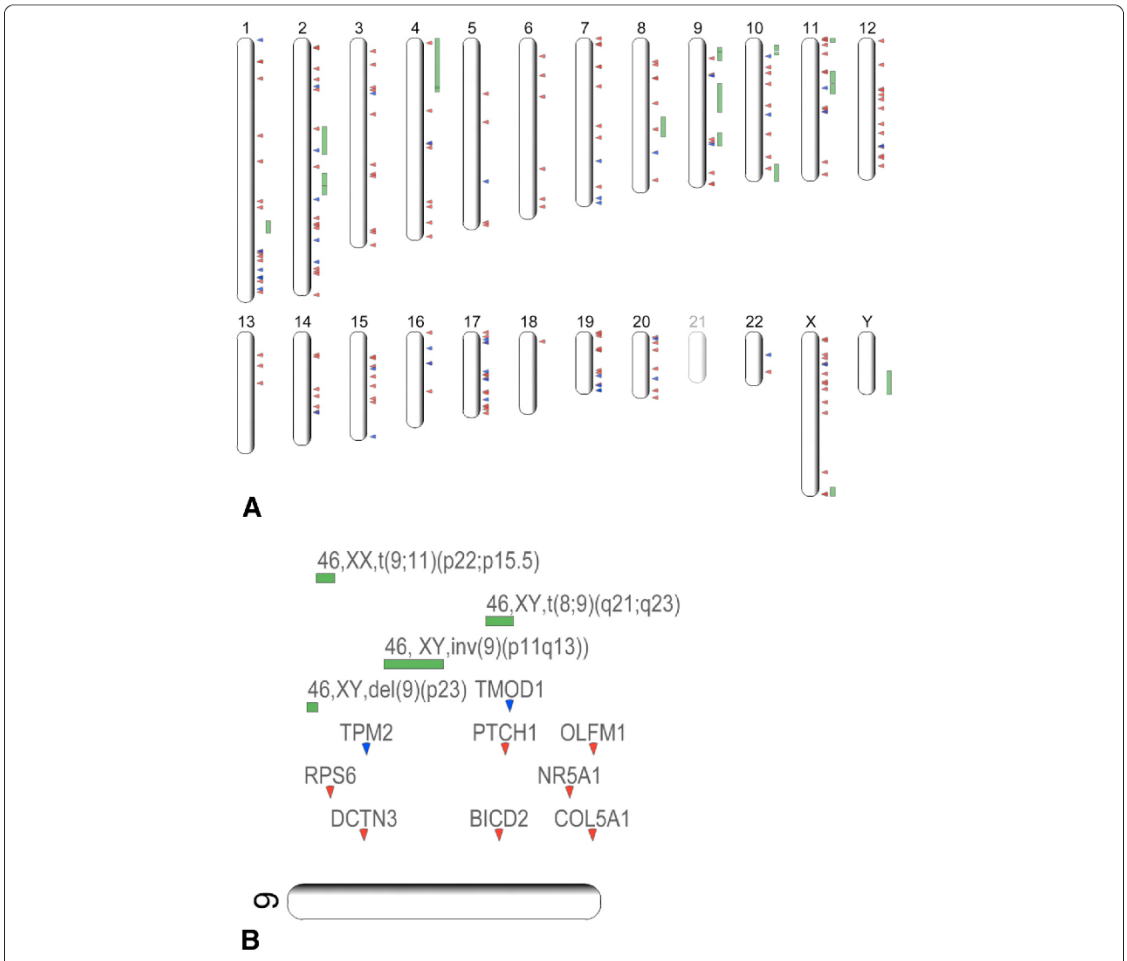


Figure 2 Genomic view of the cryptorchidism candidate genes. A. Genomic view of the literature-collected (red) and network-predicted (blue) CO associated candidate loci presented as human orthologs. The view includes syndromes with known genetic mutations that feature CO, mouse transgenic and knock-out experiments, chromosomal abnormalities, genes tested for association with CO, genes with expression patterns associated with CO, and genes associated with CO on proteomic level. Loci are placed at approximate positions on chromosome map. **B**. Enlargement of the chromosome 9.

CO in at least two independent studies using different study approaches (Table 2). These genes are denoted in bold in the online database (http://www.integratomics-time.com/cryptorchidism/candidate_genes/).

The CO associated loci mapped to all human chromosomes, except HSA21. Genomic distribution of the selected loci revealed several overlapping areas between the candidate loci. Overlaps between structural chromosomal mutations and candidate genes can be observed in Figure 2 or by using interactive genomic view available on the website (http://www.integratomics-time.com/cryptorchidism/genomic_view/). Genomic regions involved in chromosome mutations on chromosomes 2, 4, 8, 9, 11, and X [22-33] overlapped with 13 literature-collected candidate genes: *CAPG*, *MSX1*, *E2F5*, *PTCH1*, *BICD2*, *RPS6*, *FGFR2*, *HRAS*, *PAX6*, *WT1*, *TNNI2*, *TNNT3*, *FLNA*, and *MECP2*. For instance, a breakpoint on 11p15.5 overlapped with three CO candidate genes: *HRAS*, *TNNI2*, and *TNNT3*. Additionally, in some cases two regions involved in chromosome mutations overlapped; duplication on position 4p overlapped with *MSX1* gene and the region involved in chromosomal translocation on position 4p12. Interestingly, three network-predicted candidate genes, *FHL2*, *TMOD1* and *MYBPC3*, also overlapped with genomic regions involved in the chromosome mutations.

Pathway identification and network-based data mining discovery

Pathway analysis of the cryptorchidism associated candidate genes

We performed pathway analysis of the 179 literature-collected CO-candidate genes (refer to Methods). This pathway enrichment analysis, conducted by applying very stringent criteria (Bonferroni multiple test correction and p-value significant threshold of 0.01), yielded the presence of twelve significant pathways associated with the list of our CO candidate genes in human (Table 3). The literature-collected candidate genes involved in multiple (at least four) pathways are presented in Additional file 6: Table S6 and marked with an asterisk in the online database.

The presence of pathways related to "cytoskeleton", "muscle development", "muscle contraction", "focal adhesion", and "insulin signaling" was previously reported in rat [7]. In addition to these pathways, our analysis showed new pathways: "cardiomyopathy" (hypertrophic and dilated),"RAS signaling", "signaling by PDGF", "signaling by EGFR", "role of MAL in Rho-mediated activation of SRF", "IGF-1 signaling pathway", and "integrin signaling". The results represent a valid example of pathway-based data mining discovery.

Table 2 Literature-collected candidate genes associated with CO in at least two independent literature reports

Gene	Clinical syndromes	KOs and transgenes	Overlapping chromosome mutations	Association studies (number of studies)	Expression experiments
AMH	human	mouse			
AMHR2	human	mouse			
AR	human	mouse		human (2)	
ARID5B		mouse			rat
BMP7				human	rat
EPHA4		mouse			rat
ESR1		mouse		human (2)	
FGFR2	human		human		rat
HOXA10		mouse		human	rat
HRAS	human		human		rat
INSL3		mouse		human (5), sheep, dog	
LHCGR	human	mouse			
MAP2K1	human				rat
MSX1			human		rat
NR5A1		mouse		human	
RXFP2		mouse		human (4)	
SOS1	human				rat
TNNI2	human		human		
TNNT3	human		human		
WT1	human	mouse			

Table 3 Pathway analysis of the literature-collected and network-predicted candidate genes, respectively

Database	Pathway	Bonferonni p-value	Candidate genes involved in the pathway
Literature-collected candidate genes			
KEGG	Regulation of actin cytoskeleton	4.70E-06	*ACTB, BRAF, CDC42, CFL1, CHRM3, EZR, FGD1, FGF9, FGFR1, FGFR2, HRAS, ITGB1, KRAS, MAP2K1, MAP2K2, MYL2, MYL9, PDGFA, PFN1, PPP1CA, PPP1CB, PXN, RAC1, RAF1, RHOA, RRAS, SOS1*
REACTOME	Muscle contraction	5.39E-06	*DES, MYH3, MYL2, MYL3, TNNI2, TNNT2, TNNT3, TPM1, TPM3, TPM4, TTN*
KEGG	Focal adhesion	1.05E-05	*ACTB, BRAF, CCND1, CDC42, COL1A2, COL2A1, COL5A1, FLNA, GRB2, GSK3B, HRAS, IGF1, ILK, ITGB1, MAP2K1, MYL2, MYL9, PDGFA, PPP1CA, PPP1CB, PXN, RAC1, RAF1, RHOA, SOS1, THBS4*
REACTOME	Signaling by PDGF	3.61E-5	*COL1A2, COL2A1, COL5A1, GRB2, HRAS, KRAS, MAP2K1, MAP2K2, PDGFA PTPN11, RAF1, SOS1, STAT3, THBS4*
REACTOME	Signaling by insulin receptor	7.02E-05	*EIF4E, EIF4EBP1, GRB2, HRAS, KRAS, MAP2K1, MAP2K2, RAF1, RPS6, RPS6KB1, SOS1*
REACTOME	Signaling by EGFR	0.0022	*CDC42, GRB2, HRAS, KRAS, MAP2K1, MAP2K2, PTPN11, PXN, RAF1, SOS1*
PANTHER	RAS pathway	0.0025	*BRAF, CDC42, GRB2, GSK3B, HRAS, KRAS, MAP2K1, MAP2K2, RAC1, RAF1, RHOA, RRAS, SOS1, STAT3*
KEGG	Hypertrophic cardiomyopathy (HCM)	0.0035	*ACTB, DES, IGF1, ITGB1, MYH7, MYL2, MYL3, TNNT2, TPM1, TPM3, TPM4, TTN*
BIOCARTA	Role of MAL in Rho-mediated activation of SRF	0.044	*ACTA1, CDC42, MAP2K1, MAP2K2, RAC1, RAF1, RHOA*
BIOCARTA	IGF-1 signaling	0.0067	*FOS, GRB2, HRAS, IGF1, MAP2K1, PTPN11, RAF1, SOS1*
PANTHER	Integrin signaling	0.0070	*BRAF, CDC42, COL1A2, COL2A1, COL5A1, FLNA, GRB2, HRAS, ILK, ITGB1, KRAS, MAP2K1, MAP2K2, PXN, RAC1, RAF1, RHOA, RND2, RRAS, SOS1*
KEGG	Dilated cardiomyopathy	0.0099	*ACTB, DES, IGF1, ITGB1, MYH7, MYL2, MYL3, TNNT2, TPM1, TPM3, TPM4, TTN*
Network-predicted candidate genes			
REACTOME	Muscle contraction	2.71E-24	*ACTN2, DMD, MYBPC1, MYBPC2, MYBPC3, MYH8, MYL1, MYL4, NEB, TCAP, TMOD1, TNNC1, TNNC2, TNNI1, TNNI3, TNNT1, TPM2, VIM*
KEGG	Hypertrophic cardiomyopathy (HCM)	9.68E-6	*ACTC1, DMD, MYBPC3, TGFB1, TGFB2, TGFB3, TNNC1, TNNI3, TPM2,*
PANTHER	TGF-beta signaling pathway	1.07E-5	*BMP2, LEFTY1, LEFTY2, LOC100271831, MAPK1, MAPK3, MSTN, NODAL, TGFB1, TGFB2, TGFB3*
KEGG	Dilated cardiomyopathy	2.35E-5	*ACTC1, DMD, MYBPC3, TGFB1, TGFB2, TGFB3, TNNC1, TNNI3, TPM2*

As an additional validation analysis, we excluded the 112 candidate genes proposed by Barthold *et al.* (2008) [7] from the overall candidate genes list (consisting of 179 unique human genes) and repeated the pathway analysis. Nine genes from Barthold *et al.* (2008) [7] were reported as CO candidate genes also in other studies, therefore we retained them in the analysis, so that the new list of candidate genes consisted of 79 genes. The pathway analysis of these remaining 79 candidate genes returned similar results as were obtained when using the overall 179 candidate gene list. In fact, 10 of the 12 enriched pathways were the same after excluding the discussed data from the candidate gene list. In particular, the five pathways reported by Barthold *et al.* (2008) [7] in rat ("cytoskeleton", "muscle development", "muscle contraction", "focal adhesion", and "insulin signaling") were all confirmed in this independent validation analysis. The main effect of the gene removal were higher, but still significant, p-values in the pathway analysis. According to these results we can infer that inclusion of the candidate genes from Barthold *et al.* (2008) [7] is not the reason for the substantial overlap of the five pathways identified in both studies. On the contrary, the findings proposed here are a further confirmation of the validity of the conclusions made by Barthold *et al.* (2008) [7].

Surprisingly, when we searched the medical literature for articles that describe pathologies where CO, cardiomyopathy, and RAS signaling are common features, we found a perfect matching with Noonan, Cardiofaciocutaneous, LEOPARD, and Costello syndrome that all belong to the class of RASopathies [79,80]. Features of all four syndromes are different physical anomalies including concomitant presence of cardiomyopathy due to heart defects and, in males, cryptorchidism [79]. Noonan syndrome (NS) is the most common single gene cause of congenital heart disease, and NS subjects also present other features as leukemia predisposition [81]. In particular, five different mutations in *RAF1* were identified in individuals with NS; four mutations causing changes in the CR2 domain of RAF1 were associated with hypertrophic cardiomyopathy (HCM), whereas mutations in

the CR3 domain were not [82]. Additionally, *PTPN11*, *RAF1*, and *SOS1* mutants were identified as a major cause of Noonan syndrome, *BRAF* of Cardiofaciocutaneous, *PTPN11* of LEOPARD, and *HRAS* of Costello syndrome, providing new insights into RAS regulation [80,81]. These genes have also been found to be mutated in patients with RASopathies having cryptorchidism in a clinical picture. In NS patients having CO in their clinical picture 11/14 had mutated *PTPN11*, 4/5 had mutated *SOS1*, and 1/2 had mutated *RAF1*. *BRAF* has been found to be mutated in 2/3 patients with Cardiofaciocutaneous syndrome having CO, *PTPN11* in 1/4 patients with LEOPARD having CO, and *HRAS* in 2/4 patients with Costello syndrome and CO [80,81]. However, the genes responsible for the remainder are unknown, and the gene pathway relations responsible for potential connections between unrelated features such as cryptorchidism and HCM in RASopathies are not clear. Therefore, we performed a network-based prediction (see next paragraph) of CO candidate genes by identifying the most significant first neighbors (in the human protein-protein interaction network; PPIN) of the 179 literature-collected candidates.

Pathway analysis of the network-predicted candidate genes

A new cohort of 43 candidate genes (Additional file 7: Table S7) was predicted by PiNGO 1.11 [18], which is a Cytoscape plug-in (see Methods) [19]. The question we tried to address was which first-neighbor genes significantly interact with the original cohort of 179 literature-collected genes in the human PPIN. We adopted hypergeometric statistical test and Bonferroni multiple testing correction. The cohort of 43 network-predicted genes was strongly significant (Bonferroni p-value < 0.0095); therefore, we consider them as additional CO candidate genes.

In order to evaluate the importance of these new candidate genes we performed the pathway analysis (Table 3), according to the same procedure already used in the previous paragraph (and described in the methods). The most intriguing evidence is the presence of significant pathways related to cardiomyopathy and muscle contraction in both sets of candidate genes (*i.e.* literature-collected and network-predicted). Pathways common to both sets of candidate genes represent a confirmation of the validity and robustness of the results obtained in the first pathway analysis and regarding the hypothesis of connection between CO and cardiomyopathy, in NS. Yet, it is also a quality proof of the procedure adopted for network prediction of new candidate genes.

Pathway analysis of the overall CO candidate gene list (179 literature-collected and 43 network-predicted genes)

The first cohort of 179 literature-collected genes and the second one containing 43 network-predicted genes were condensed in a list of 222 unique genes - the overall candidate gene list. We repeated the pathway analysis on this list applying the same very stringent criteria used above (Bonferroni multiple test correction and p-value significant threshold of 0.01). The analysis suggested the presence of 12 significant pathways associated with the overall list of candidate genes in human (Table 4).

The "muscle contraction" pathway was the most significant (in absolute) with Bonferroni corrected p-value of 4.55E-33 (Figure 3A); while the "hypertrophic cardiomyopathy" was the second most significant pathway with Bonferroni corrected p-value of 1.21E-09 (Figure 3B). These results are crucial for our study because they suggest the presence of a strong genomic connection among diverse pathways associated with clinical features that seemed unrelated. To address relationship among these mechanisms we created a matrix merging the information related to the gene participation in several identified pathways. Of the 222 (179 + 43) candidate genes, 172 were filtered out because they were not present in at least two of the 12 significant pathways. The resulting matrix consists of 50 candidate genes in the rows and 12 enriched pathways in the columns (Additional file 8: Table S8). The matrix values are binary: 0 indicates that the gene is not present in a pathway, whereas 1 indicates that the gene is present.

Hierarchical bi-clustering of the matrix [83], both in the rows and in the columns, was performed to detect clusters of genes which participated in common pathways, and clusters of pathways which share the same genes, respectively. The result of this analysis is provided in the Figure 4. The presence of two main groups of clusters is evident. The first group is constituted of "cardiomyopathy" (hypertrophic and dilated), "muscle contraction" and "cardiac muscle contraction" pathways. The second group is constituted of "focal adhesion", "regulation of actin cytoskeleton, "integrin signaling", "vascular smooth muscle contraction", "signaling by insulin receptor", "signaling by PDGF", "RAS pathway", and "TGF-beta signaling".

In order to further investigate the relation between the genes involved in the "cardiomyopathy" (hypertrophic and dilated), the "muscle contraction" and the "RAS pathway" and to interpret their role in creating connections between the diverse pathway modules, we searched the STRING database [20] for protein-protein interactions, selecting only the interactions with high confidence score. The outcome of this analysis is represented in Figure 5. All the 50 genes presented at least one interaction in the PPI network produced by the STRING database. This network is provided as a supplementary material (Additional file 9: Table S9).

The principal pathways involved in both, CO and RASopathies, are displayed on the PPIN (Figure 5), and

Table 4 Pathway analysis of the overall candidate gene list (literature-collected and network-predicted)

Database	Pathway	Bonnferoni p-value	Candidate genes involved in the pathway
REACTOME	Muscle contraction	4.55E-33	*ACTN2, DES, DMD, MYBPC1, MYBPC2, MYBPC3, MYH3, MYH8, MYL1, MYL2, MYL3, MYL4, NEB, TCAP, TMOD1, TNNC1, TNNC2, TNNI1, TNNI2, TNNI3, TNNT1, TNNT2, TNNT3, TPM1, TPM2, TPM3, TPM4, TTN, VIM*
KEGG	Hypertrophic cardiomyopathy (HCM)	1.21E-9	*ACTB, ACTC1, DES, DMD,IGF1, ITGB1, MYBPC3, MYH7, MYL2, MYL3, TGFB1, TGFB2, TGFB3, TNNC1, TNNI3, TNNT2, TPM1, TPM2, TPM3, TPM4, TTN*
KEGG	Dilated cardiomyopathy	1.21E-8	*ACTB, ACTC1, DES, DMD, IGF1, ITGB1, MYBPC3, MYH7, MYL2, MYL3, TGFB1, TGFB2, TGFB3, TNNC1, TNNI3, TNNT2, TPM1, TPM2, TPM3, TPM4, TTN*
KEGG	Focal adhesion	4.68E-7	*ACTB, ACTN2, BRAF, CAV1, CCND1, CDC42, COL1A2, COL2A1, COL5A1, FLNA, GRB2, GSK3B, HRAS, IGF1, IGF1R, ILK, ITGB1, MAP2K1, MAPK1, MAPK3, MYL2, MYL9, PDGFA, PPP1CA, PPP1CB, PRKCA, PXN, RAC1, RAF1, RHOA, SOS1, THBS4*
REACTOME	Signaling by insulin receptor	3.83E-6	*EIF4E, EIF4EBP1, GRB2, HRAS, KRAS, MAP2K1, MAP2K2, MAPK1, MAPK3, RAF1, RHEB, RPS6, RPS6KB1, SOS1*
KEGG	Regulation of actin cytoskeleton	4.29E-6	*ACTB, ACTN2, BRAF, CDC42, CFL1, CHRM3, EZR, FGD1, FGF3, FGF9, FGFR1, FGFR2, HRAS, ITGB1, KRAS, MAP2K1, MAP2K2, MAPK1, MAPK3, MYL2, MYL9, PDGFA, PFN1, PPP1CA, PPP1CB, PXN, RAC1, RAF1, RHOA, RRAS, SOS1*
PANTHER	TGF-beta signaling pathway	7.01E-6	*AMH, AMHR2, BMP2, BMP4, BMP5, BMP7, CDC42, FOS, FOXO1, FOXP3, HRAS, KRAS, LEFTY1, LEFTY2, MAPK1, MAPK3, MSTN, NODAL, RHEB, RRAS, TGFB1, TGFB2, TGFB3*
BIOCARTA	Integrin signaling pathway	1.35E-5	*ACTA1, ACTN2, CAV1, GRB2, HRAS, ITGB1, MAP2K1, MAP2K2, MAPK1, MAPK3, PXN, RAF1, RHOA, SOS1*
REACTOME	Signaling by PDGF	7.27E-5	*COL2A1, COL1A2, COL5A1, GRB2, HRAS, KRAS, MAP2K1, MAP2K2, MAPK1, MAPK3, PDGFA, PTPN11, RAF1, SOS1, STAT3, THBS4*
KEGG	Cardiac muscle contraction	0.0024	*ACTC1, MYH7, MYL2, MYL3, TNNC1, TNNI3, TNNT2, TPM1, TPM2, TPM3, TPM4*
PANTHER	RAS pathway	0.0030	*BRAF, CDC42, GRB2, GSK3B, HRAS, KRAS, MAP2K1, MAP2K2, MAPK1, MAPK3, RAC1, RAF1, RHOA, RRAS, SOS1, STAT3*
KEGG	Vascular smooth muscle contraction	0.0058	*ACTA2, BRAF, MAP2K1, MAP2K2, MAPK1, MAPK3, MYH11, MYL9, PPP1CA, PPP1CB, PRKCA, PRKCE, RAF1, RHOA*

also marked in the Figure 4 to facilitate the comparison. This figure addresses the question of the relation between the common genetic mechanisms underlying CO and RASopathies. Figure 5 provides a clear visualization of the overlapping pathways and of the integrated network of relations existing on proteomic level. At the best of our knowledge, this is the first time that such relation is presented, and it might help in understanding the relation between co-presence of CO and cardiomyopathy as clinical and apparently unrelated features in RASopathies. This fact is clarified by the layout offered in Figure 5 that reveals how the "cardiomyopathy" and the "RAS signaling" pathways are connected by a plethora of interactions with high confidence score in the STRING database. To investigate the precise type of intra- and inter- pathway interactions we suggest to mine the network that we provide in the supplementary material (Additional file 9: Table S9). Figure 5 further emphasizes how the "focal adhesion" and the "TGF-beta signaling" pathways overlap the "cardiomyopathy", the "muscle contraction" and the "RAS signaling" pathways by connecting proteins at different metabolic levels. The relevance of the "focal adhesion" pathway, as well as the importance of "cytoskeleton", "muscle development", "muscle contraction", and "insulin signaling" pathways in cryptorchidism were widely discussed [7]. However, the referred study was conducted on a rat model and all of the pathways were considered and treated separately. Here, for the first time, we proceed to an integratomic investigation of the genetic factors linked to CO in human. Meanwhile, we offer the holistic perspective that points out how clinical features apparently unrelated with CO might be generated by genetic mutation(s) which propagate at different pathway levels of the network. This propagation on different pathway-modules can justify the onset of multiple unrelated clinical features in complex diseases, such as RASopathies. The selection of 43 network-based predicted genes considered together with these last disease-related evidences are another proof that confirms the power of PPIN for association of genes with diseases [21,84].

Our results are in concordance with previous observations that alignment of human interactome with human phenome enables identification of causative genes (and networks) underlying disease families. Phenotypic overlap implies genetic overlap and human phenome can be viewed as a landscape of interrelated diseases, which reflects overlapping molecular causation [6,85-88]. In addition, it has been already shown that causative genes from syndromes that are phenotypically

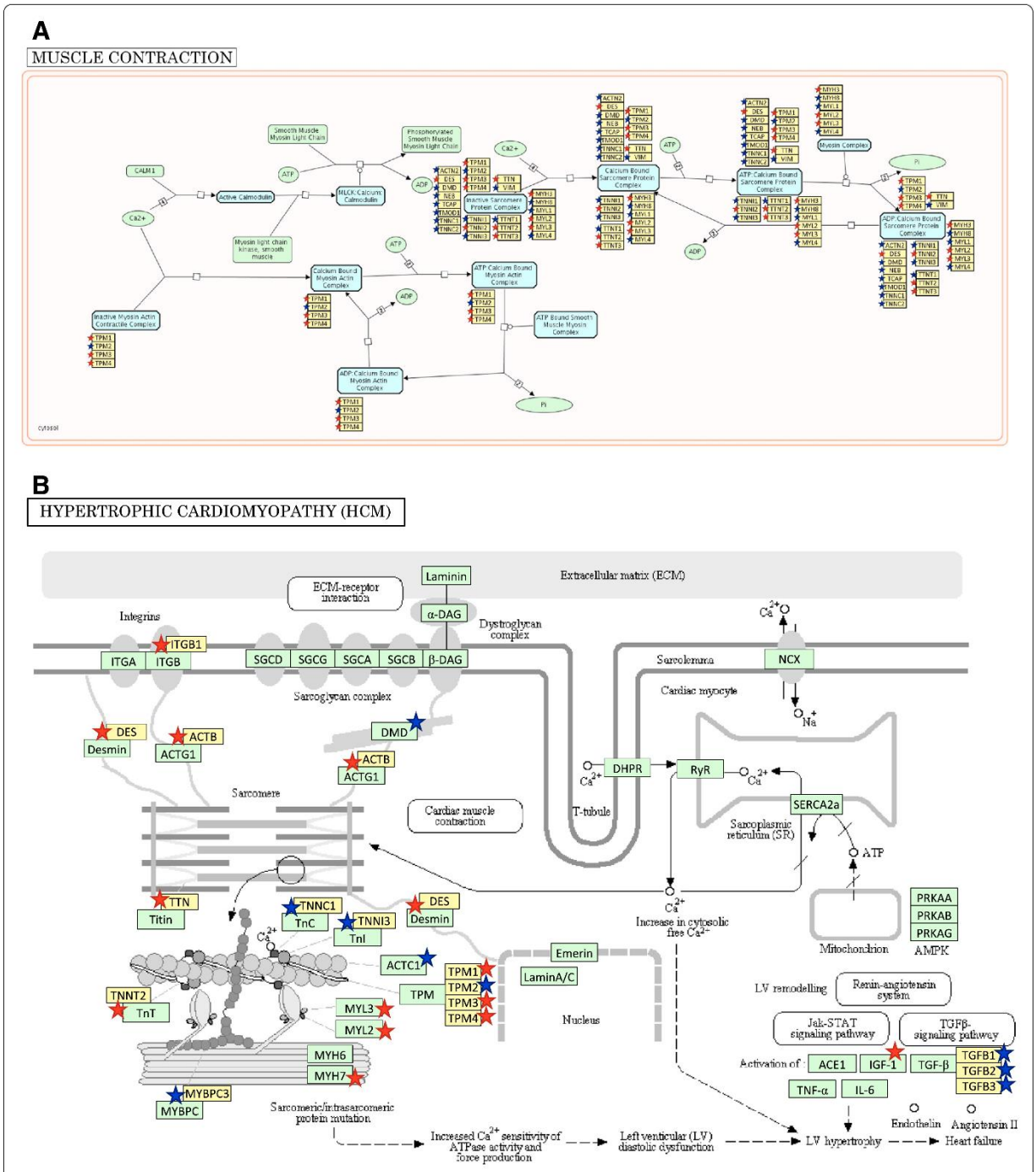


Figure 3 Pathways, containing literature-collected and network-predicted CO candidate genes. CO genes are marked with stars: red stars stand for literature-collected genes, blue stars stand for network-predicted genes. Gene names are added in yellow boxes to the original pathway images. **A**. Reactome pathway: Muscle contraction – human. **B**. KEGG pathway: Hypertrophic cardiomyopathy (HCM) – human.

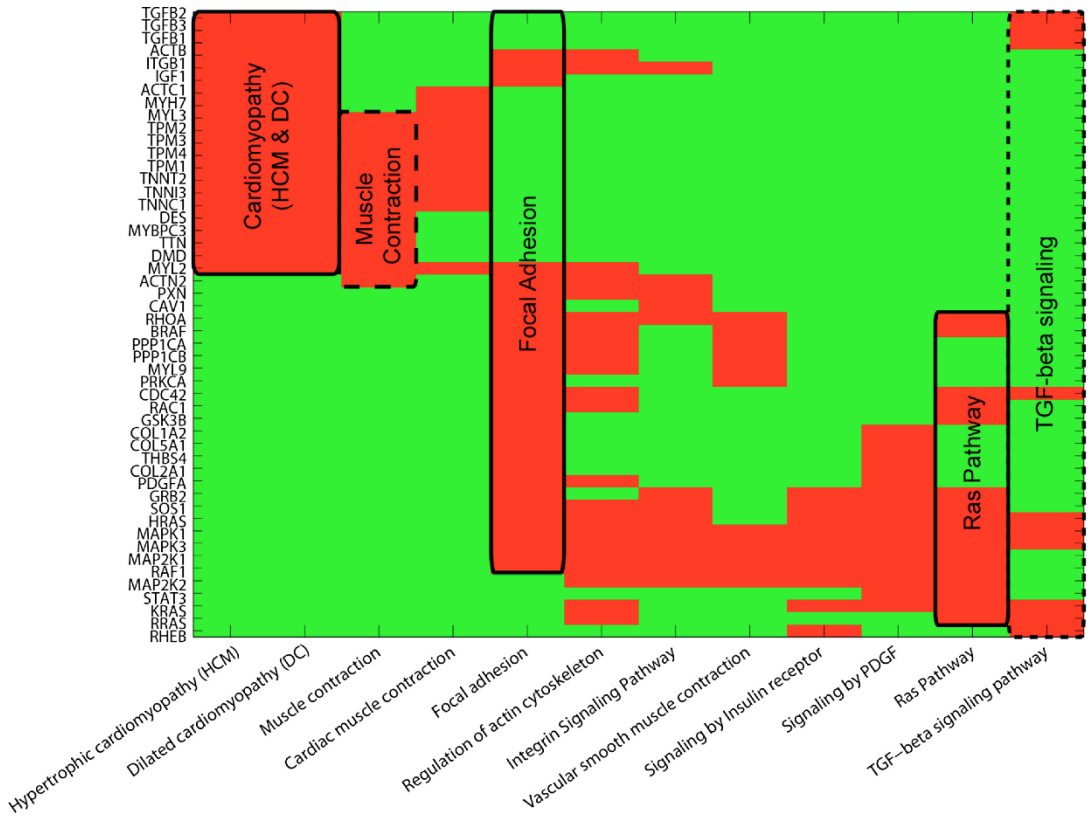


Figure 4 Hierarchical bi-clustering of the CO candidate genes. Hierarchical bi-clustering of the matrix of the CO candidate genes present in at least two pathways. The matrix consists of 50 candidate genes (rows) and 12 enriched pathways (columns). The black lines (full and dashed) are used to indicate the modules corresponding to clusters of interacting proteins in the respective pathways.

similar to a genetically uncharacterized syndrome can be used to query the gene network for functionally related candidate genes [89].

Candidate gene prioritization

Prioritization of candidate genes underlying complex traits remains one of the main challenges in molecular biology [11]. In this study we used three criteria for selecting the most promising candidate genes: 1) number of independent literature reports connecting the candidate gene with CO (Table 2), 2) involvement of candidate genes in enriched pathways (Table 3), and 3) position of candidate genes on the genomic map (genes positioned in regions where multiple CO associated data overlap were considered positional candidates) (Figure 2).

Twenty genes have been suggested as a genetic cause for CO in at least two independent studies (criterion 1) using different study approaches (*AMH*, *AMHR2*, *AR*, *ARID5B*, *BMP7*, *EPHA4*, *ESR1*, *FGFR2*, *HOXA10*, *HRAS*, *INSL3*, *LHCGR*, *MAP2K1*, *MSX1*, *NR5A1*, *RXFP2*, *SOS1*, *TNNI2*, *TNNT3*, and *WT1*). Among them, *INSL3* has been associated with CO in eight, *RXFP2* in five, and *AR* in four independent studies. However, this approach should be treated with some caution because of the possible bias towards research interest into more "popular" genes. The approach will be more reliable after significant amount of unbiased genome-wide studies is available.

Considering involvement in enriched pathways (criterion 2), the most promising candidates would be *HRAS*, *MAP2K1*, *MAP2K2*, *GRB2*, *RAF1* and *SOS1*, which are all involved in seven or more enriched pathways. For the literature-collected candidate genes involved in multiple (four or more) CO-enriched pathways we assembled genetic information relevant for further functional analyses: assignment to corresponding biological pathways, genetic variability, and putative presence of polymorphic microRNA (miRNA) target sites (Additional file 6: Table S6). The importance of small non-coding RNAs (ncRNAs) in gene regulation and pathogenesis of the

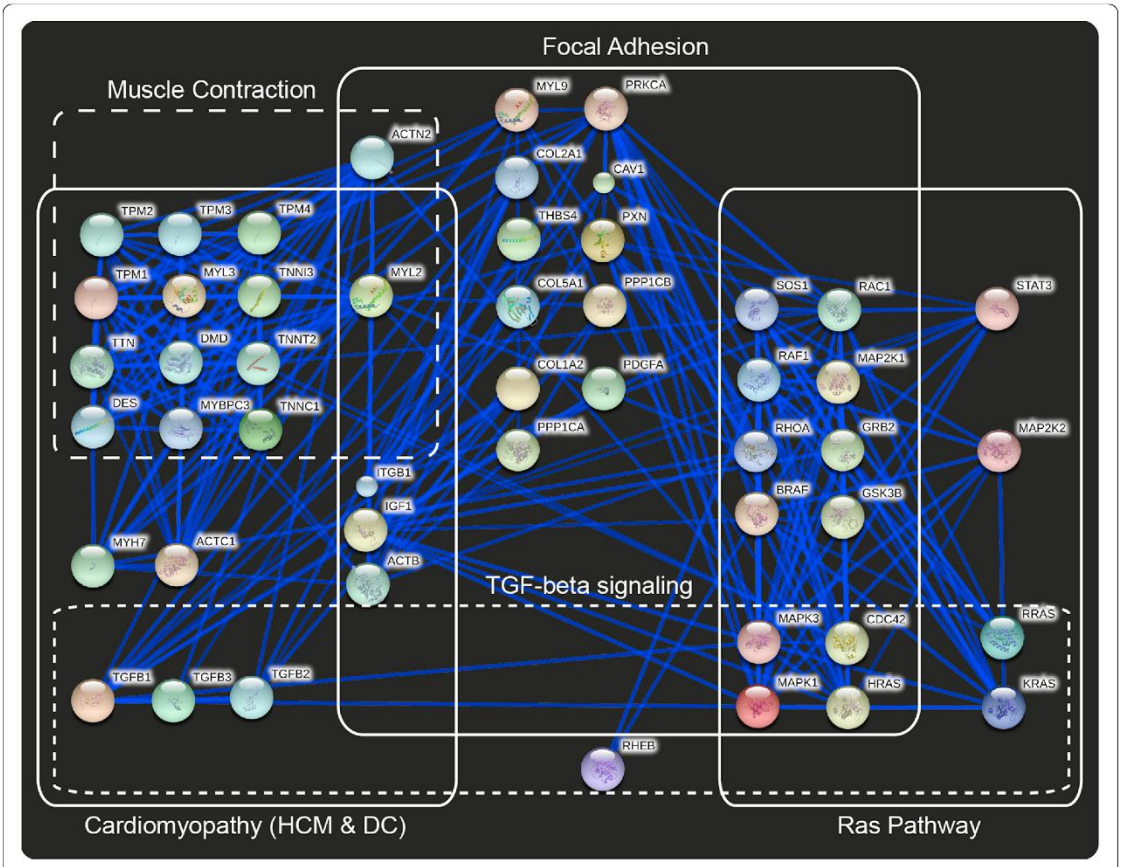


Figure 5 String protein-protein interaction network (PPIN) of the CO candidate genes. PPIN is obtained from the 50 CO candidate genes interacting in the STRING database. The lines (full and dashed) delimit the presence of the diverse overlapped protein-pathway-modules. The same line is used in the Figure 4 to indicate the modules corresponding to clusters of interacting proteins in the respective pathways.

diseases, including reduced fertility, is today evident [90]. However, to our knowledge, there are no literature reports associating ncRNAs or epigenetic factors with CO.

The most promising candidates meeting both suggested criteria (1 and 2) are *FGFR2* (reported in two CO-associated studies/ involved in one CO-associated pathway), *HRAS* (3/8), *MAP2K1* (2/9), and *SOS1* (2/5). Additionally, *TNNI2* and *TNNT3* are reported in the literature (once each), involved in one CO enriched pathway (*i.e.* "muscle contraction"), and positioned in a region overlapping chromosomal mutation.

Genomic regions involved in the chromosome mutations on chromosomes 2, 4, 8, 9, 11, and X overlapped with 14 candidate genes suggested as positional candidates (criterion 3): *CAPG*, *MSX1*, *E2F5*, *PTCH1*, *BICD2*, *RPS6*, *FGFR2*, *HRAS*, *PAX6*, *WT1*, *TNNI2*, *TNN3*, *FLNA*, and *MECP2*. Additionally, three network-predicted candidate genes, *FHL2*, *TMOD1* and *MYBPC3* overlapped with chromosomal mutations. Considering suggested prioritization criteria, *HRAS* gene meets all of them.

Reliability of such methodologically different approaches is not always comparable (for example, data from genome-wide expression experiments is much less validated than syndromic or transgenic data); therefore, ranking candidate genes based only on a number of different reports/approaches is not always feasible. However, less validated data may also be of high biological relevance and should not be discarded for hypothesis-driven approaches. To increase reliability of the collected heterogeneous data we tested *in silico* how candidate genes interact at the proteomic level. Although integratomic approaches are only partially established yet and have several drawbacks, including already mentioned heterogeneity of input data, we believe that such approaches are a reasonable and at the moment among the most promising

ways for hypothesis generation, which should be further experimentally validated in animal and/or human populations. Similar integratomics approach was already used for identification of candidate loci for mammary gland associated phenotypes [8], male infertility [9], and obesity [10,91], and could be adapted to any other complex trait.

Conclusions

In this study we present an overview of CO associated candidate regions/genes and suggest pathways potentially involved in the pathogenesis of the disease. The integrative, comparative-genomics approach, and *in silico* analyses of the collected data aim to help solving the problem of fragmented and often contradictory data extracted from different methodologically focused studies. The protein-protein interactions analysis revealed the most relevant pathways associated with CO candidate gene list and enabled us to suggest additional candidate genes based on network prediction. Described systems biology approach will contribute to a better understanding of genetic causes for cryptorchidism and provides possible example how integration and linking of complex traits related data can be used for hypothesis generation. Publicly available online CO gene atlas and data entry option will allow researcher to enter, browse, and visualize CO associated data. The proposed network-based approach elucidates co-presence of similar pathogenetic mechanisms underlying diverse clinical syndromes/defects and could be of a great importance in research in the field of molecular syndromology. This approach has also a potential to be used for future development of diagnostic, prognostic, and therapeutic markers. The developed integratomics approach can be extrapolated to study genetic background of any other complex traits/diseases and to generate hypothesis for downstream experimental validation.

Additional files

Additional file 1: Table S1. Chromosomal abnormalities and CNVs associated with cryptorchidism.

Additional file 2: Table S2. Selected clinical syndromes that feature CO in their clinical picture.

Additional file 3: Table S3. Transgenic and knock-out murine models that display cryptorchid phenotype.

Additional file 4: Table S4. Genes tested for association with CO.

Additional file 5: Table S5. Genes with expression patterns associated with CO. Genes with expression patterns associated with CO in rat (adapted from [7]) and genomic location of their human orthologs.

Additional file 6: Table S6. The literature collected candidate genes involved in multiple (four or more) CO-associated pathways.

Additional file 7: Table S7. Forty-three network-predicted CO-associated candidate genes.

Additional file 8: Table S8. Protein network information matrix for candidate genes. Protein network information matrix for candidate genes involved in at least two pathways significant for literature-collected and network-predicted candidate genes. The matrix consists of 50 rows (each row corresponds to a different gene involved in at least two pathways) and 12 columns (each column corresponds to a different pathway), where 0 indicates that the gene is not present in a pathway and 1 indicates that the gene is present in it.

Additional file 9: Table S9. STRING network data: list of protein network interactions present in STRING with high confidence score.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CVC performed pathway and PPIN analyses. JO, TK, and MZ performed the data mining and established the database. TK, JO, and CVC interpreted the results and drafted the manuscript. MZ developed web-based interactive visualization tool. PD and TR provided feedback on the initial draft and contributed to the final editing of the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

This work was supported by the Slovenian Research Agency (ARRS) through the Research programme Comparative genomics and genome biodiversity (P4-0220). C.V.C. received financial support from the Italian Inter-polytechnic School of Doctorate (SIPD) and from the King Abdullah University of Science and Technology.

Author details

[1]Integrative Systems Biology Laboratory, Biological and Environmental Sciences and Engineering Division, Computer, Electrical and Mathematical Sciences and Engineering Division, Computational Bioscience Research Center, King Abdullah University for Science and Technology (KAUST), Thuwal, Saudi Arabia. [2]Department of Mechanics, Politecnico di Torino, Turin, Italy. [3]Proteome Biochemistry Unit, San Raffaele Scientific Institute, Milan, Italy. [4]Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Domzale, Slovenia.

Received: 14 February 2012 Accepted: 6 February 2013
Published: 14 February 2013

References

1. Dovc P, Kunej T, Williams GA: **Genetics and genomics of reproductive disorders.** In *Reproductive Genomics of Domestic Animals.* 1st edition. Edited by Jiang Z, Ott TL. Oxford, UK: Wiley-Blackwell; 2010:67–97.
2. Amann RP, Veeramachaneni DNR: **Cryptorchidism in common eutherian mammals.** *Reproduction* 2007, **133**(3):541–561.
3. Foresta C, Zuccarello D, Garolla A, Ferlin A: **Role of hormones, genes, and environment in human cryptorchidism.** *Endocr Rev* 2008, **29**(5):560–580.
4. Barthold JS: **Undescended testis: current theories of etiology.** *Curr Opin Urol* 2008, **18**(4):395–400.
5. Mortell A, Montedonico S, Puri P: **Animal models in pediatric surgery.** *Pediatr Surg Int* 2006, **22**(2):111–128.
6. Oti M, Huynen MA, Brunner HG: **Phenome connections.** *Trends Genet* 2008, **24**(3):103–106.
7. Barthold JS, McCahan SM, Singh AV, Knudsen TB, Si X, Campion L, Akins RE: **Altered expression of muscle- and cytoskeleton-related genes in a rat strain with inherited cryptorchidism.** *J Androl* 2008, **29**(3):352–366.
8. Ogorevc J, Kunej T, Razpet A, Dovc P: **Database of cattle candidate genes and genetic markers for milk production and mastitis.** *Anim Genet* 2009, **40**(6):832–851.
9. Ogorevc J, Dovc P, Kunej T: **Comparative genomics approach to identify candidate genetic loci for male fertility.** *Reprod Domest Anim* 2011, **46**(2):229–239.
10. Kunej T, Jevsinek Skok D, Zorc M, Ogrinc A, Michal JJ, Kovac M, Jiang Z: **Obesity gene atlas in mammals.** *J Genomics* 2012, **1**:45–55.
11. Moreau Y, Tranchevent LC: **Computational tools for prioritizing candidate genes: boosting disease gene discovery.** *Nat Rev Genet* 2012, **13**(8):523–536.

12. Zorc M, Jevsinek Skok D, Godnic I, Calin GA, Horvat S, Jiang Z, Dovc P, Kunej T: **Catalog of MicroRNA Seed Polymorphisms in Vertebrates.** *PLoS One* 2012, **7**(1):e30737.
13. da Huang W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, *et al*: **DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists.** *Nucleic Acids Res* 2007, **35**(Web Server issue): W169–W175.
14. Razick S, Magklaras G, Donaldson IM: **iRefIndex: a consolidated protein interaction database with provenance.** *BMC Bioinforma* 2008, **9**:405.
15. Chuang HY, Lee E, Liu YT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol* 2007, **3**:140.
16. Ravasi T, Cannistraci CV, Suzuki H, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, *et al*: **An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man.** *Cell* 2010, **140**(5):744–752.
17. Kamburov A, Wierling C, Lehrach H, Herwig R: **ConsensusPathDB–a database for integrating human functional interaction networks.** *Nucleic Acids Res* 2009, **37**(Database issue):D623–D628.
18. Smoot M, Ono K, Ideker T, Maere S: **PiNGO: a Cytoscape plugin to find candidate genes in biological networks.** *Bioinformatics* 2011, **27**(7):1030–1031.
19. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T: **Cytoscape 2.8: new features for data integration and network visualization.** *Bioinformatics* 2011, **27**(3):431–432.
20. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, *et al*: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.** *Nucleic Acids Res* 2011, **39**(Database issue):D561–D568.
21. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D: **Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates.** *Nucleic Acids Res* 2010, **38**(Database issue):D640–D651.
22. Miyake Y, Kaneda Y: **A new type of Robertsonian translocation (1/26) in a bull with unilateral cryptorchidism, probably occurring de novo.** *Nihon juigaku zasshi J Vet Sci* 1987, **49**(6):1015–1019.
23. Sasagawa I, Nakada T, Ishigooka M, Sawamura T, Adachi Y, Hashimoto T: **Chromosomal anomalies in cryptorchidism.** *Int Urol Nephrol* 1996, **28**(1):99–102.
24. Ogata T, Muroya K, Matsuo N, Hata J, Fukushima Y, Suzuki Y: **Impaired male sex development in an infant with molecularly defined partial 9p monosomy: implication for a testis forming gene(s) on 9p.** *J Med Genet* 1997, **34**(4):331–334.
25. Suzuki Y, Sasagawa I, Nakada T, Onmura Y: **Bilateral cryptorchidism associated with terminal deletion of 10q.** *Urol Int* 1998, **61**(3):186–187.
26. Goldschmidt B, El-Jaick KB, Souza LM, Carvalho ECQ, Moura VLS, Benevides Filho IM: **Cryptorchidism associated with 78, XY/79, XXY mosaicism in dog.** *Israel J Vet Med* 2001, **56**:56–58.
27. Moreno-Garcia M, Miranda EB: **Chromosomal anomalies in cryptorchidism and hypospadias.** *J Urol* 2002, **168**(5):2170–2172. discussion 2172.
28. Prabhakara K, Angalena R, Ramadevi AR: **Familial (9;11)(p22;p15.5)pat translocation and XX sex reversal in a phenotypic boy with cryptorchidism and delayed development.** *Genetic counseling (Geneva, Switzerland)* 2004, **15**(1):37–41.
29. Brito L, Sertich PL, Durkin K, Chowdhary BP, Turner RM, Greene LM, McDonnell S: **Autosomic 27 Trisomy in a Standardbred Colt.** *J Equine Vet Sci* 2008, **28**(7):431–436.
30. Tartaglia N, Davis S, Hench A, Nimishakavi S, Beauregard R, Reynolds A, Fenton L, Albrecht L, Ross J, Visootsak J, *et al*: **A new look at XXYY syndrome: medical and psychological features.** *Am J Med Genet A* 2008, **146A**(12):1509–1522.
31. van der Veken LT, Dieleman MM, Douben H, van de Brug JC, van de Graaf R, Hoogeboom AJ, Poddighe PJ, de Klein A: **Low grade mosaic for a complex supernumerary ring chromosome 18 in an adult patient with multiple congenital anomalies.** *Mol Cytogenet* 2010, **3**:13.
32. Niyazov DM, Nawaz Z, Justice AN, Toriello HV, Martin CL, Adam MP: **Genotype/phenotype correlations in two patients with 12q subtelomere deletions.** *Am J Med Genet A* 2007, **143A**(22):2700–2705.
33. Melis D, Genesio R, Boemio P, Del Giudice E, Cappuccio G, Mormile A, Ronga V, Conti A, Imperati F, Nitsch L, *et al*: **Clinical description of a patient carrying the smallest reported deletion involving 10p14 region.** *Am J Med Genet A* 2012, **158A**(4):832–835.
34. Tannour-Louet M, Han S, Corbett ST, Louet JF, Yatsenko S, Meyers L, Shaw CA, Kang SH, Cheung SW, Lamb DJ: **Identification of de novo copy number variants associated with human disorders of sexual development.** *PLoS One* 2010, **5**(10):e15392.
35. Peltonen L, Perola M, Naukkarinen J, Palotie A: **Lessons from studying monogenic disease for common disease.** *Hum Mol Genet* 2006, **15**:R67–R74. Spec No 1.
36. Gianotten J, van der Veen F, Alders M, Leschot NJ, Tanck MW, Land JA, Kremer JA, Hoefsloot LH, Mannens MM, Lombardi MP, *et al*: **Chromosomal region 11p15 is associated with male factor subfertility.** *Mol Hum Reprod* 2003, **9**(10):587–592.
37. Klonisch T, Fowler PA, Hombach-Klonisch S: **Molecular and genetic regulation of testis descent and external genitalia development.** *Dev Biol* 2004, **270**(1):1–18.
38. Kreidberg JA, Sariola H, Loring JM, Maeda M, Pelletier J, Housman D, Jaenisch R: **WT-1 IS REQUIRED FOR EARLY KIDNEY DEVELOPMENT.** *Cell* 1993, **74**(4):679–691.
39. Ikadai H, Ajisawa C, Taya K, Imamichi T: **Suprainguinal ectopic scrota of TS inbred rats.** *J Reprod Fertil* 1988, **84**(2):701–707.
40. Li X, Nokkala E, Yan W, Streng T, Saarinen N, Warri A, Huhtaniemi I, Santti R, Makela S, Poutanen M: **Altered structure and function of reproductive organs in transgenic male mice overexpressing human aromatase.** *Endocrinology* 2001, **142**(6):2435–2442.
41. Caron P, Imbeaud S, Bennet A, Plantavid M, Camerino G, Rochiccioli P: **Combined hypothalamic-pituitary-gonadal defect in a hypogonadic man with a novel mutation in the DAX-1 gene.** *J Clin Endocrinol Metab* 1999, **84**(10):3563–3569.
42. Donaldson KM, Tong SY, Washburn T, Lubahn DB, Eddy EM, Hutson JM, Korach KS: **Morphometric study of the gubernaculum in male estrogen receptor mutant mice.** *J Androl* 1996, **17**(2):91–95.
43. Gorlov IP, Kamat A, Bogatcheva NV, Jones E, Lamb DJ, Truong A, Bishop CE, McElreavey K, Agoulnik AI: **Mutations of the GREAT gene cause cryptorchidism.** *Hum Mol Genet* 2002, **11**(19):2309–2318.
44. Canto P, Escudero I, Soderlund D, Nishimura E, Carranza-Lira S, Gutierrez J, Nava A, Mendez JP: **A novel mutation of the insulin-like 3 gene in patients with cryptorchidism.** *J Hum Genet* 2003, **48**(2):86–90.
45. Ferlin A, Garolla A, Bettella A, Bartoloni L, Vinanzi C, Roverato A, Foresta C: **Androgen receptor gene CAG and GGC repeat lengths in cryptorchidism.** *Eur J Endocrinol* 2005, **152**(3):419–425.
46. Yoshida R, Fukami M, Sasagawa I, Hasegawa T, Kamatani N, Ogata T: **Association of cryptorchidism with a specific haplotype of the estrogen receptor alpha gene: implication for the susceptibility to estrogenic environmental endocrine disruptors.** *J Clin Endocrinol Metab* 2005, **90**(8):4716–4721.
47. Ferlin A, Bogatcheva NV, Gianesello L, Pepe A, Vinanzi C, Agoulnik AI, Foresta C: **Insulin-like factor 3 gene mutations in testicular dysgenesis syndrome: clinical and functional characterization.** *Mol Hum Reprod* 2006, **12**(6):401–406.
48. Silva-Ramos M, Oliveira JM, Cabeda JM, Reis A, Soares J, Pimenta A: **The CAG repeat within the androgen receptor gene and its relationship to cryptorchidism.** *Int Braz J Urol* 2006, **32**(3):330–334. discussion 335.
49. Wada Y, Okada M, Fukami M, Sasagawa I, Ogata T: **Association of cryptorchidism with Gly146Ala polymorphism in the gene for steroidogenic factor-1.** *Fertil Steril* 2006, **85**(3):787–790.
50. Bogatcheva NV, Ferlin A, Feng S, Truong A, Gianesello L, Foresta C, Agoulnik AI: **T222P mutation of the insulin-like 3 hormone receptor LGR8 is associated with testicular maldescent and hinders receptor expression on the cell surface membrane.** *Am J Physiol Endocrinol Metab* 2007, **292**(1):E138–E144.
51. El Houate B, Rouba H, Sibai H, Barakat A, Chafik A, Chadli el B, Imken L, Bogatcheva NV, Feng S, Agoulnik AI, *et al*: **Novel mutations involving the INSL3 gene associated with cryptorchidism.** *J Urol* 2007, **177**(5):1947–1951.
52. Yamazawa K, Wada Y, Sasagawa I, Aoki K, Ueoka K, Ogata T: **Mutation and polymorphism analyses of INSL3 and LGR8/GREAT in 62 Japanese patients with cryptorchidism.** *Horm Res* 2007, **67**(2):73–76.
53. Wang Y, Barthold J, Figueroa E, Gonzalez R, Noh PH, Wang M, Manson J: **Analysis of five single nucleotide polymorphisms in the ESR1 gene in cryptorchidism.** *Birth Defects Res A Clin Mol Teratol* 2008, **82**(6):482–485.
54. Harris RM, Finlayson C, Weiss J, Fisher L, Hurley L, Barrett T, Emge D, Bathgate RA, Agoulnik AI, Jameson JL: **A missense mutation in LRR8 of RXFP2 is associated with cryptorchidism.** *Mamm Genome* 2010, **21**(9–10):442–449.
55. Tang KF, Zheng JZ, Xing JP: **Molecular analysis of SNP12 in estrogen receptor alpha gene in hypospadiac or cryptorchid patients from Northwestern China.** *Urol Int* 2011, **87**(3):359–362.
56. Feng S, Ferlin A, Truong A, Bathgate R, Wade JD, Corbett S, Han S, Tannour-Louet M, Lamb DJ, Foresta C, *et al*: **INSL3/RXFP2 signaling in testicular descent.** *Ann N Y Acad Sci* 2009, **1160**:197–204.

57. Dalgaard MD, Weinhold N, Edsgard D, Silver JD, Pers TH, Nielsen JE, Jorgensen N, Juul A, Gerds TA, Giwercman A, *et al*: **A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation.** *J Med Genet* 2012, **49**(1):58–65.
58. Kolon TF, Wiener JS, Lewitton M, Roth DR, Gonzales ET Jr, Lamb DJ: **Analysis of homeobox gene HOXA10 mutations in cryptorchidism.** *J Urol* 1999, **161**(1):275–280.
59. Teles MG, Trarbach EB, Noel SD, Guerra-Junior G, Jorge A, Beneduzzi D, Bianco SD, Mukherjee A, Baptista MT, Costa EM, *et al*: **A novel homozygous splice acceptor site mutation of KISS1R in two siblings with normosmic isolated hypogonadotropic hypogonadism.** *Eur J Endocrinol* 2010, **163**(1):29–34.
60. Williams GA, Ott TL, Michal JJ, Gaskins CT, Wright RW Jr, Daniels TF, Jiang Z: **Development of a model for mapping cryptorchidism in sheep and initial evidence for association of INSL3 with the defect.** *Anim Genet* 2007, **38**(2):189–191.
61. Cassata R, Iannuzzi A, Parma P, De Lorenzi L, Peretti V, Perucatti A, Iannuzzi L, Di Meo GP: **Clinical, cytogenetic and molecular evaluation in a dog with bilateral cryptorchidism and hypospadias.** *Cytogenet Genome Res* 2008, **120**(1–2):140–143.
62. Zhao X, Du ZQ, Rothschild MF: **An association study of 20 candidate genes with cryptorchidism in Siberian Husky dogs.** *J Anim Breed Genet* 2010, **127**(4):327–331.
63. Galan JJ, Guarducci E, Nuti F, Gonzalez A, Ruiz M, Ruiz A, Krausz C: **Molecular analysis of estrogen receptor alpha gene AGATA haplotype and SNP12 in European populations: potential protective effect for cryptorchidism and lack of association with male infertility.** *Hum Reprod* 2007, **22**(2):444–449.
64. Pathirana IN, Tanaka K, Kawate N, Tsuji M, Kida K, Hatoya S, Inaba T, Tamada H: **Analysis of single nucleotide polymorphisms in the 3′ region of the estrogen receptor 1 gene in normal and cryptorchid Miniature Dachshunds and Chihuahuas.** *J Reprod Dev* 2010, **56**(4):405–410.
65. Lo Giacco D, Ars E, Bassas L, Galan JJ, Rajmil O, Ruiz P, Caffaratti J, Guarducci E, Ruiz-Castane E, Krausz C: **ESR1 promoter polymorphism is not associated with nonsyndromic cryptorchidism.** *Fertil Steril* 2011, **95**(1):369–371. 371 e361-362.
66. Krausz C, Quintana-Murci L, Fellous M, Siffroi JP, McElreavey K: **Absence of mutations involving the INSL3 gene in human idiopathic cryptorchidism.** *Mol Hum Reprod* 2000, **6**(4):298–302.
67. Takahashi I, Takahashi T, Komatsu M, Matsuda J, Takada G: **Ala/Thr60 variant of the Leydig insulin-like hormone is not associated with cryptorchidism in the Japanese population.** *Pediatr Int* 2001, **43**(3):256–258.
68. Baker LA, Nef S, Nguyen MT, Stapleton R, Nordenskjold A, Pohl H, Parada LF: **The insulin-3 gene: lack of a genetic basis for human cryptorchidism.** *J Urol* 2002, **167**(6):2534–2537.
69. Bertini V, Bertelloni S, Valetto A, Lala R, Foresta C, Simi P: **Homeobox HOXA10 gene analysis in cryptorchidism.** *J Pediatr Endocrinol Metab* 2004, **17**(1):41–45.
70. Nuti F, Marinari E, Erdei E, El-Hamshari M, Echavarria MG, Ars E, Balercia G, Merksz M, Giachini C, Shaeer KZ, *et al*: **The leucine-rich repeat-containing G protein-coupled receptor 8 gene T222P mutation does not cause cryptorchidism.** *J Clin Endocrinol Metab* 2008, **93**(3):1072–1076.
71. Simoni M, Tuttelmann F, Michel C, Bockenfeld Y, Nieschlag E, Gromoll J: **Polymorphisms of the luteinizing hormone/chorionic gonadotropin receptor gene: association with maldescended testes and male infertility.** *Pharmacogenet Genomics* 2008, **18**(3):193–200.
72. Kunej T, Zorn B, Peterlin B: **Y chromosome microdeletions in infertile men with cryptorchidism.** *Fertil Steril* 2003, **79**(Suppl 3):1559–1565.
73. Bor P, Hindkjaer J, Kolvraa S, Rossen P, von der Maase H, Jorgensen TM, Sorensen VT, Eiberg H, Ingerslev HJ: **Screening for Y microdeletions in men with testicular cancer and undescended testis.** *J Assist Reprod Genet* 2006, **23**(1):41–45.
74. Gurbuz N, Ozbay B, Aras B, Tasci AI: **Do microdeletions in the AZF region of the Y chromosome accompany cryptorchidism in Turkish children?** *Int Urol Nephrol* 2008, **40**(3):577–581.
75. Hejmej A, Gorazd M, Kosiniak-Kamysz K, Wiszniewska B, Sadowska J, Bilinska B: **Expression of aromatase and oestrogen receptors in reproductive tissues of the stallion and a single cryptorchid visualised by means of immunohistochemistry.** *Domest Anim Endocrinol* 2005, **29**(3):534–547.
76. Nguyen MT, Delaney DP, Kolon TF: **Gene expression alterations in cryptorchid males using spermatozoal microarray analysis.** *Fertil Steril* 2009, **92**(1):182–187.
77. Hutson JM, Watts LM, Farmer PJ: **Congenital undescended testes in neonatal pigs and the effect of exogenous calcitonin gene-related peptide.** *J Urol* 1998, **159**(3):1025–1028.
78. Zuccarello D, Morini E, Douzgou S, Ferlin A, Pizzuti A, Salpietro DC, Foresta C, Dallapiccola B: **Preliminary data suggest that mutations in the CgRP pathway are not involved in human sporadic cryptorchidism.** *J Endocrinol Invest* 2004, **27**(8):760–764.
79. Bertola DR, Pereira AC, Passetti F, de Oliveira PS, Messiaen L, Gelb BD, Kim CA, Krieger JE: **Neurofibromatosis-Noonan syndrome: molecular evidence of the concurrence of both disorders in a patient.** *Am J Med Genet A* 2005, **136**(3):242–245.
80. Digilio MC, Lepri F, Baban A, Dentici ML, Versacci P, Capolino R, Ferese R, De Luca A, Tartaglia M, Marino B, *et al*: **RASopathies: Clinical Diagnosis in the First Year of Life.** *Molecular syndromology* 2011, **1**(6):282–289.
81. Roberts AE, Araki T, Swanson KD, Montgomery KT, Schiripo TA, Joshi VA, Li L, Yassin Y, Tamburino AM, Neel BG, *et al*: **Germline gain-of-function mutations in SOS1 cause Noonan syndrome.** *Nat Genet* 2007, **39**(1):70–74.
82. Razzaque MA, Nishizawa T, Komoike Y, Yagi H, Furutani M, Amo R, Kamisago M, Momma K, Katayama H, Nakagawa M, *et al*: **Germline gain-of-function mutations in RAF1 cause Noonan syndrome.** *Nat Genet* 2007, **39**(8):1013–1017.
83. Cannistraci CV, Ravasi T, Montevecchi FM, Ideker T, Alessio M: **Nonlinear dimension reduction and clustering by Minimum Curvilinearity unfold neuropathic pain and tissue embryological classes.** *Bioinformatics* 2010, **26**(18):i531–i539.
84. Navlakha S, Kingsford C: **The power of protein interaction networks for associating genes with diseases.** *Bioinformatics (Oxford, England)* 2010, **26**:1057–1063. England.
85. Rzhetsky A, Wajngurt D, Park N, Zheng T: **Probing genetic overlap among complex human phenotypes.** *P Natl Acad Sci USA* 2007, **104**(28):11694–11699.
86. Wu X, Jiang R, Zhang MQ, Li S: **Network-based global inference of human disease genes.** *Mol Syst Biol* 2008, **4**:189.
87. Wu X, Liu Q, Jiang R: **Align human interactome with phenome to identify causative genes and networks underlying disease families.** *Bioinformatics* 2009, **25**(1):98–104.
88. Loscalzo J, Barabasi AL: **Systems biology and the future of medicine.** *Wiley Interdiscip Rev Syst Biol Med* 2011, **3**(6):619–627.
89. Oti M, Brunner HG: **The modular nature of genetic diseases.** *Clin Genet* 2007, **71**(1):1–11.
90. Ogorevc J, Dovc P, Kunej T: **Polymorphisms in microRNA targets: a source of new molecular markers for male reproduction.** *Asian J Androl* 2011, **13**(3):505–508.
91. Kunej T, Skok DJ, Horvat S, Dovc P, Jiang Z: **The glypican 3-hosted murine mir717 gene: sequence conservation, seed region polymorphisms and putative targets.** *Int J Biol Sci* 2010, **6**(7):769–772.

doi:10.1186/1755-8794-6-5
Cite this article as: Cannistraci *et al.*: Pivotal role of the muscle-contraction pathway in cryptorchidism and evidence for genomic connections with cardiomyopathy pathways in RASopathies. *BMC Medical Genomics* 2013 **6**:5.

Priloga E

The microRNA decalog of cancer involvement (Kunej in sod, v tisku)

THE microRNA DECALOG OF CANCER INVOLVEMENT

Tanja Kunej[1], Irena Godnic[1], Minja Zorc[1,2], Simon Horvat[1,3], George A Calin[4]

[1] Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Groblje 3, 1230, Domzale, Slovenia
[2] Department of Computational and Systems Biology, Rothamsted Research, Harpenden, Herts, AL5, 2JQ, UK
[3] Department of Biotechnology, National Institute of Chemistry, Ljubljana 1000, Slovenia
[4] Experimental Therapeutics Department and Center for RNA Interference and Non-Coding RNA, The University of Texas MD Anderson Cancer Center, Houston, TX 77030.

Abstract

MicroRNAs (miRNAs) are short (19-24 nucleotides in length) non-coding RNAs (ncRNAs) that post-transcriptionally regulate gene expression. In this chapter, the fundamentals for understanding miRNA involvement in human cancers are summarized as the decalog of miRNA principles: 1. MicroRNAs can up- or down-regulate gene expression by targeting various genic regions. 2. Each miRNA can regulate the expression of numerous target genes and likewise can multiple miRNAs target the same gene. Such miRNA-target interplay is involved in regulation of various physiological processes and pathophysiology of several diseases, including cancer. 3. About a half of mammalian miRNAs are intragenic, predominantly intronic, and can be coordinately expressed and functionally linked with their host genes. 4. MicroRNAs are frequently located within cancer-associated genomic regions (CAGR) and act as tumor-suppressors or oncogenes. 5. Genetic variations within miRNA genes, their targets, and genes encoding components of processing machinery can affect phenotypic variation and disease susceptibility. 6. MicroRNA-mediated gene silencing is one of the classes of epigenetic mechanisms and together with DNA methylation and histone modifications function in an interacting regulatory circuit. 7. Aberrant miRNA gene expression signatures characterize cancer cells and are as such valuable biomarkers applied in diagnosis, prognosis, and treatment in cancer. 8. Circulating miRNAs have been shown as non-invasive biomarkers in cancer. 9. RNA inhibition using miRNAs is a potential treatment method which can be applied for specific types of cancer. 10. Interplay between miRNAs, other ncRNAs, and protein-coding genes forms a complex network of interactions in normal and disease tissues.
The knowledge of overall functional roles of miRNAs in cancer would therefore be important not only for scientists, but could be applied by clinicians and oncologists. It may have a major impact on biomedical research by providing miRNA-based therapeutics and diagnostics opportunities.

Keywords: microRNA (miRNA), non-coding RNA (ncRNA), cancer, miRNA profiling, biomarker, miRNA-related therapy

I INTRODUCTION

II THE DECALOG OF PRINCIPLES OF miRNA INVOLVEMENT IN HUMAN CANCERS

1. MicroRNAs are a class of ncRNAs that are estimated to regulate expression of one third of the mammalian genome by binding to promoter, coding and untranslated regions, proteins or other ncRNAs.
2. Each miRNA can regulate the expression of numerous target genes, likewise can multiple miRNAs regulate the same target gene, and this interplay is involved in regulation of various physiological processes and pathophysiology of several diseases, including all analyzed types of human cancers.
3. About a half of mammalian miRNAs are intragenic, predominantly intronic, and on the same strand as their host genes, and can be coordinately expressed and functionally linked with them.
4. MicroRNAs are frequently located within cancer associated genomic regions (CAGR) and can act as tumor suppressors or oncogenes.
5. Genetic variations in miRNA genes and their precursors, target sites, and genes encoding components of processing machinery can affect phenotypic variation and disease susceptibility.
6. Three types of epigenetic concepts have been associated with miRNAs: A) miRNA-mediated gene silencing is one of the epigenetic mechanisms, B) miRNAs can also directly control the epigenetic machinery with a subclass of miRNAs (epi-miRNAs), and C) miRNA expression can be downregulated via promoter hypermethylation.
7. Aberrant miRNA gene expression signatures characterize cancer cells and miRNA profiling can be applied in diagnosis, prognosis, and treatment in cancer patients.
8. Circulating miRNAs are potential noninvasive biomarkers in cancer.
9. RNA inhibition using miRNAs is a potential treatment method for specific types of cancer.
10. Interplay between miRNAs, other ncRNAs, and protein-coding genes forms a complex network of interactions in normal and disease tissues.

III CONCLUSIONS

IV LITERATURE CITED

Abbreviations

ceRNA	competitive endogenous RNA
DNMT	DNA methyltransferase
HDAC	histone deacetylase
lincRNA	long intergenic non-coding RNA
lncRNA	long non-coding RNA
miRNA	microRNA
MRE	miRNA response elements
ncRNA	non-coding RNA
pre-miRNA	miRNA precursor
pri-miRNA	miRNA primary transcript
RISC	RNA-induced silencing complex
UTR	untranslated region

I. INTRODUCTION

MicroRNAs (miRNAs) are noncoding RNAs (ncRNAs) with gene-regulatory functions, involved in a variety of molecular functions and biological processes in multicellular organisms (reviewed in (Bartel 2004; Fabian, Sonenberg, and Filipowicz 2010)). Initially transcribed by RNA polymerase II as long, capped and polyadenylated primary transcripts (pri-miRNAs), they are processed by the Microprocessor protein complex, which contains Drosha, an RNase III endonuclease, and *DGCR8* (DiGeorge syndrome critical region gene 8) (also known as Pasha) (Gregory et al. 2004; Denli et al. 2004). First, Drosha in conjunction with its binding partner DGCR8 processes pri-miRNAs into hairpin RNAs of 70-100 nucleotides (nt) in length known as precursor miRNAs (pre-miRNAs). Translocated from the nucleus to the cytoplasm by Exportin 5, pre-miRNAs are processed by a RNase III endonuclease Dicer and TRBP (TAR RNA-binding protein), in an 19-24 nt long duplex. Finally, the duplex interacts with the RNA-induced silencing complex (RISC), which includes proteins of the Argonaute family (Ago1-4 in humans) (Hammond et al. 2001). One strand of the miRNA duplex remains stably associated with RISC and becomes the mature miRNA, which guides the RISC complex to target mRNAs. Recent evidences show that miRNAs are able to up- or down-regulate target gene expression by binding to its different regions (reviewed in (Kunej, Godnic, et al. 2012)).

II. THE DECALOG OF PRINCIPLES OF miRNA INVOLVEMENT IN HUMAN CANCERS

The development of high throughput methods to detect miRNA expression in human samples has provided invaluable tools to investigate the role of miRNAs both in physiological and pathological conditions. Exponentially accumulated data in the last years clearly shows that perturbations of miRNA genes at a DNA, RNA or expression level play a critical role in cancer initiation and progression. Determining the miRNome, the whole set of miRNAs, specific for cancer would further contribute to better diagnosis and prognosis of human diseases. Such findings are important not only for scientists in general, but also for clinicians and oncologists, as the field of ncRNA touches every aspect of human oncology. The main principles for understanding miRNA involvement in human cancers could be summarized in the following way:

1. **MicroRNAs are a class of ncRNAs that are estimated to regulate expression of one third of the mammalian genome by binding to promoter, coding and untranslated regions, proteins or other ncRNAs.**

The initial discovery that miRNAs target the 3'-untranslated regions (UTRs) of mRNAs and downregulate the expression of protein-coding genes in cytoplasm has been expanded with the following additional observations: (**A**) miRNAs can be localized in the nucleus (Hwang, Wentzel, and Mendell 2007), (**B**) in addition to 3'-UTR miRNAs target other genic regions at a DNA or RNA level (5'-UTR, promoter regions, coding regions) (Lytle, Yario, and Steitz 2007; Place et al. 2008; Tay et al. 2008; Forman, Legesse-Miller, and Coller 2008) and proteins (Eiring et al. 2010), (**C**) miRNAs can upregulate translation (Vasudevan, Tong, and Steitz 2007; Place et al. 2008), and (**D**) miRNAs interact with other ncRNAs and various types of RNA transcripts in a "competing endogenous RNA" (ceRNA) hypothesis (Salmena et al. 2011) (**Fig. 1**).

(A) MicroRNAs have initially been considered to be located only in cytoplasm. However, recent studies have reported that most miRNAs found in the cytoplasm might also localize to or function in the nucleus (Liao et al. 2010). Hwang et al. (2007) showed that human *MIR29B* was significantly enriched in the nucleus; the hexanucleotide terminal motif of *MIR29B* was found to be required for nuclear localization (Hwang, Wentzel, and Mendell 2007). (**B**) It has been shown that miRNAs can also at a DNA level affect transcription by direct binding to promoters. For example, human *MIR373* binds to the E-cadherin (*CDH1*) promoter which induces gene expression (Place et al. 2008). MicroRNA-dependent mRNA repression also occurs through binding sites located in mRNA coding sequences (CDS), as shown for miRNAs targeting *Nanog* (Nanog homeobox), *Pou5f1* (POU class 5 homeobox 1), *Sox2* (SRY (sex determining region Y)-box 2) (Tay et al. 2008) and *DICER* (Forman, Legesse-Miller, and Coller 2008). In addition to miRNA-mediated gene silencing through base pairing with DNA or mRNA target sequences, miRNAs also interfere with function of

regulatory proteins (decoy activity). In particular, *MIR328* binds to poly-C binding protein 2 (PCBP2), alternatively known as heterogeneous ribonucleoprotein (hnRNP) E2. This binding does not involve the miRNA's seed region and prevents its interaction with the target mRNA. In chronic myelogenous leukemia downregulation of *MIR328* enables inhibition of myeloid differentiation with PCBP2, which, as a result, leads to tumor progression (Eiring et al. 2010). (**C**) It has also been shown that miRNAs upregulate translation through binding to AU-rich elements (AREs) in cell cycle arrested cells. For example, human *MIR369-3* targets AREs in the *TNF* (tumor necrosis factor) mRNA; during the cell cycle, while the cell was proliferating, the miRNA down-regulated translational activity, but upon cell-cycle arrest, it reversed its effect (Vasudevan, Tong, and Steitz 2007). MicroRNAs can therefore both repress and activate gene expression at different points in the cell cycle. (**D**) Other ncRNAs, such as non-coding ultraconserved genes (UCGs), have been found to be consistently altered at the genomic level in a high percentage of leukemias and carcinomas and may interact with miRNAs (Calin et al. 2007). Recently, the role of coding and non-coding RNAs has been emphasized and grouped in a unifying theory of ceRNAs that can regulate one another through their ability to compete for miRNA binding. The ceRNA hypothesis suggests that long non-coding RNAs (lncRNAs), transcribed pseudogenes, and mRNAs communicate with each other using miRNA response elements (MREs), which are sequences with partial complementarity on target mRNA transcripts (Salmena et al. 2011). Competing endogenous RNAs may therefore be active partners in miRNAs regulation exerting effects on their expression levels, which may have important implications in pathological conditions, such as cancer.

2. **Each miRNA can regulate the expression of numerous target genes, likewise can multiple miRNAs regulate the same target gene, and this interplay is involved in regulation of various physiological processes and pathophysiology of several diseases, including all analyzed types of human cancers.**

It was estimated that about one-third of human mRNAs are considered as miRNA targets (Lewis, Burge, and Bartel 2005). Vertebrate miRNAs target about 200 transcripts each and more than one miRNA might coordinately regulate a single target, thereby providing a basis for complex networks (Krek et al. 2005). By specifically cleaving the homologous mRNAs, or by inhibiting protein synthesis, miRNAs are likely to target separate or multiple effectors of pathways involved in cell differentiation, proliferation, and survival. Several computational tools for miRNA target prediction have been developed (e.g. TargetScan, PicTar, miRanda), however, only a limited number of predicted targets were experimentally confirmed and are being collected in curated databases (e.g. miRecords, miRTarBase).

For several miRNAs, the involvement in essential biological processes has been demonstrated, such as B-cell lineage fate (*MIR181*) (Chen et al. 2004), B-cell survival

(*MIR15A* and *MIR16-1*) (Calin et al. 2002), cell proliferation control (*MIR125B* and *MIRLET7*) (Lee et al. 2005; Takamizawa et al. 2004), brain patterning (*MIR430*) (Giraldez et al. 2005), pancreatic cell insulin secretion (*MIR375*) (Poy et al. 2004), and adipocyte development (*MIR143*) (reviewed in (Kunej, Skok, et al. 2012)). MicroRNAs were found to be involved in the pathophysiology of all analyzed types of human tumors, including benign and malignant tumors. MicroRNAs differentially expressed between tumors and normal tissues have been identified in lymphoma, breast cancer, lung cancer, papillary thyroid carcinoma, glioblastoma, hepatocellular carcinoma (HCC), pancreatic tumors, pituitary adenomas, cervical cancer, brain tumors, prostate cancer, kidney and bladder cancers, and colorectal cancers. Furthermore, miRNA alterations at the genic or expression levels were identified in other human diseases, including schizophrenia, autoimmune or cardiac disorders (reviewed in (Ha 2011; Stahlhut Espinosa and Slack 2006)).

3. About a half of mammalian miRNAs are intragenic, predominantly intronic, and on the same strand as their host genes, and can be coordinately expressed and functionally linked with them.

Depending on the genomic position, miRNAs can be classified as intragenic and intergenic (**Fig. 2**). It has been estimated that approximately half of vertebrate miRNAs are processed from introns of protein-coding and non-protein coding transcripts (Rodriguez et al. 2004; Erdmann et al. 2000). A single host gene transcript can comprise multiple and overlapping resident miRNAs, called a cluster, which is processed from the same polycistronic primary transcript (Rodriguez et al. 2004; Ambros 2004). Intergenic miRNAs have their own transcriptional mechanisms, whereas intragenic, more specifically intronic miRNAs, are co-transcribed with their host genes (Baskerville and Bartel 2005; Rodriguez et al. 2004). Because expression profiles of intronic miRNAs in many cases coincided with the transcription profiles of their host genes this raised a question as to how these miRNAs were processed and coordinately regulated (Baskerville and Bartel 2005). Intronic miRNAs, like most ncRNAs, are released from the excised host introns in the post-splicing process (Kim and Kim 2007; Rearick et al. 2011) but it was later indicated that intronic miRNAs might also be processed from un-spliced intronic regions prior to splicing catalysis (Kim and Kim 2007). It was observed, that E2F1 (E2F transcription factor 1) regulates both an intronic miRNA cluster (*MIR106B-25*) and its host gene *MCM7* (minichromosome maintenance complex component 7) and induces their accumulation in gastric primary tumors (Petrocca et al. 2008). Another study reported coordinated expression of *MIR218* and its host gene *SLIT2* (slit homolog 2 (Drosophila)), and between *MIR224* and its host gene *GABRE* (gamma-aminobutyric acid (GABA) A receptor, epsilon), in the clear cell renal cell carcinoma (White et al. 2011). MicroRNA genes located in closely linked clusters exhibit highly correlated expression patterns (Baskerville and Bartel 2005; Sempere et al. 2004). Co-transcription and correlated expression pattern of host genes and their resident miRNAs strongly support their transcriptional co-

regulation. In addition to co-expression and proposed co-regulation of miRNA and host genes, several studies have described a functional link between them in cells. Different roles have been attributed to intronic miRNAs, from providing a negative feedback regulatory mechanism with their host genes (Li, Tang, and Lin 2007) to targeting genes that are functionally antagonistic to their host genes (Barik 2008). For example, *MIR126* was found to regulate expression of its host gene *EGFL7* (EGF-like-domain, multiple 7) in a negative feed-back loop (Fish et al. 2008). The association between introns and resident ncRNAs was also considered to have a synergistic effect with important implications for fine-tuning gene expression patterns in the genome (Rearick et al. 2011). Both *MIR126* and its host gene *EGFL7* are associated with vascular abnormalities (Wang et al. 2008) and are also epigenetically regulated in human cancer cells (Saito et al. 2009). Therefore, genomic location, especially of the intragenic miRNAs, strongly influences their co-expression regulation and often also functionally links miRNAs with their host genes.

4. MicroRNAs are frequently located within cancer associated genomic regions (CAGR) and can act as tumor suppressors or oncogenes.

The role of miRNAs in cancer was proposed early in the history of miRNA research by three important observations: miRNA genes are not randomly distributed in the genome, but are frequently located at fragile sites and cancer-associated genomic regions (CAGRs) (Calin et al. 2004), miRNAs are involved in cell proliferation and apoptosis (Brennecke et al. 2003; Lee, Feinbaum, and Ambros 1993), and miRNA expression is deregulated in malignant tumors and tumor cell lines in comparison with normal tissues (Gaur et al. 2007; Lu, Tej, et al. 2005; Calin and Croce 2006).

MicroRNAs can function as oncogenes, by activating malignant potential, or as tumor suppressors, by blocking the cell's malignant potential, and are therefore referred to as oncomiRs. Components required for miRNA biogenesis have also been associated with various cancers (reviewed in (Esquela-Kerscher and Slack 2006)). First direct evidence of miRNA oncogenic activity was reported when *MIR15A* and *MIR16-1* were found deleted or down-regulated in most chronic lymphocytic leukemia (CLL) patients (Calin et al. 2002). Follow-up studies reported that miRNAs can act by various mechanisms as oncogenes such as *MIR21* (Medina, Nolde, and Slack 2010) or *MIR155* (Costinean et al. 2006), for which the transgenic mice models developed acute B cell leukemias, or as tumor suppressors such as *MIR15A/16-1* cluster, for which the knockout (KO) mice produced and developed CLL (reviewed in (Croce 2009)). In some instances the same miRNA can act as an oncogene in one type of cells and as a suppressor in another due to different targets and mechanisms of action. For example, *MIR222* is overexpressed in liver cancers where it targets suppressor *PTEN* (phosphatase and tensin homolog) (Garofalo et al. 2009), while the same miRNA is downregulated in erythroblastic leukemias where it targets oncogene *KIT* (v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog) (Felli et al. 2005). Therefore, miRNAs can function both as oncogenes or tumor suppressors but their mode of

action cannot be predicted in cancer as they function differentially depending on cell or tissue type.

Modulation of miRNA biogenesis pathway can promote tumorigenesis through increased repression of tumor suppressors and/or through incomplete repression of oncogenes. It was indicated that the dysregulation of a single oncogenic miRNA can lead to the development of a malignant tumor. Such alterations can be a result of various mechanisms, such as deletions, amplifications or mutations involving miRNA loci, and also epigenetic modifications and dysregulation of transcription factors (TFs) which target specific miRNAs (reviewed in (Croce 2009)) (**Fig. 2**). In several cases miRNAs have been shown to affect all hallmarks of malignant cells: 1) self-sufficiency in growth signals (*MIRLET7* family), 2) insensitivity to anti-growth signals (*MIR17-92* cluster), 3) evasion from apoptosis (*MIR34A*), 4) limitless replicative potential (*MIR372/373* cluster), 5) angiogenesis (*MIR210*), and 6) invasion and metastases (*MIR10B*) (reviewed in (Santarpia, Nicoloso, and Calin 2010)). Based on this, it is possible to propose that miRNAs are master regulators of tumor biology features and their deregulation can therefore contribute to oncogenesis in various ways.

5. Genetic variations in miRNA genes and their precursors, target sites, and genes encoding components of processing machinery can affect phenotypic variation and disease susceptibility.

Genetic variations have been shown to affect miRNA-mediated gene regulation. Genetic variations such as single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) within pri-miRNAs and genes encoding silencing components contribute to phenotypic variations, including disease susceptibility (reviewed in (Georges, Coppieters, and Charlier 2007)). Germline mutations or SNPs can occur in miRNA precursors, their target sites, and miRNA processing machinery (**Fig. 2**) and have been found to affect miRNA-mediated regulatory functions, which can lead to phenotypic effects including activating cancerogenesis.

Germline and somatic mutations in active mature, precursor or primary miRNA molecules contribute to cancer predisposition and initiation, as was observed in *MIR15A/16* cluster mutations in rare families with CLL and breast cancer (Calin et al. 2005). Even though these mutations are rare, a similar mutation was observed in the New Zealand Black (NZB) strain of mice that were susceptible to the development of CLL late in life. Researchers identified a point mutation 6 nt downstream from the identical miRNA, murine *MIR16*, whose levels of expression were decreased in NZB lymphoid tissue. When delivered to an NZB malignant B-1 cell line exogenous *MIR16* resulted in cell cycle alterations and increased apoptosis (Raveche et al. 2007). Taken together, these two studies (one of human CLL and the other of a murine model of human indolent CLL) indicate that *MIR16* is the

first miRNA proven to be involved in CLL predisposition and cancer predisposition in general.

Polymorphisms occurring in miRNA genes are referred to as “miR-SNPs”, and the term “miR-TS-SNP” is used for SNPs located within miRNA target binding sites (Sun et al. 2009). Even though many miRNA sequence variations observed in cancer have altered the secondary structure with no demonstrated effects on miRNA processing (Diederichs and Haber 2006), several recent reports show that miR-SNPs can be associated with cancer susceptibility (Shen et al. 2008; Tian et al. 2009; Hu et al. 2009). It was observed that miR-SNPs affect function by modulating the miRNA precursor transcription, processing and maturation (Zeng and Cullen 2005), or miRNA-mRNA interaction (Johnson et al. 2005). Sequence variations in the mature miRNA, especially in the seed region, may have an effect on miRNA target recognition (Sun et al. 2009) and can have an effect on a diverse array of traits (Zorc et al. 2012). A catalog of genetic variations residing within miRNA seed region (miR-seed-SNPs) has been generated and will serve researchers as a starting point in testing more targeted hypothesis (Zorc et al. 2012). Because of the miRNA-target interaction, the miR-SNPs (including miR-seed-SNPs) and miR-TS-SNPs function in the same manner to create or destroy miRNA binding sites.

Furthermore, polymorphisms in protein-coding mRNAs which are targeted by miRNAs can also influence cancer risk. For example, the *MIRLET7* complementary SNP site in the *KRAS* 3'-UTR was found significantly associated with an increased risk for non-small cell lung carcinoma among moderate smokers (Chin et al. 2008). A catalog of SNPs residing within miRNA binding regions of cancer genes has also been compiled (Landi et al. 2007). Even though miR-TS-SNPs were shown to influence susceptibility to tumorigenesis, additional association studies and follow-up functional experiments should still be applied to provide a clearer view on the interplay of these variations in disease development (Nicoloso et al. 2010).

Another class of polymorphisms with a potentially profound effect on the phenotype are SNPs in the miRNA-processing machinery. They can have deleterious effects on the miRNome and global repression of miRNA maturation, shown to lead to tumorigenesis (Kumar et al. 2007). Several studies reported that such genetic polymorphisms affect cancer susceptibility: SNPs in *GEMIN4* (gem (nuclear organelle) associated protein 4) were significantly associated with altered renal cell carcinoma (Horikawa et al. 2008) and bladder cancer risk (Yang et al. 2008), whereas a SNP in the 3′-UTR of *DICER1* was associated with an increased risk of premalignant oral lesions in individuals with leukoplakia and/or erythroplakia (Clague et al. 2010). In a recent study, Sung *et al.* found SNPs located within *EIF2C2* (eukaryotic translation initiation factor 2C, 2), *DICER1*, *PIWIL1* (piwi-like 1 (Drosophila)), *DGCR8*, *DROSHA*, and *GEMIN4* that were associated with breast cancer survival (Sung et al. 2012).

The essential part in miRNA variation studies is identification of SNPs, which can be aided with bioinformatic tools by intercalating and cross-referencing data from dbSNP. MicroRNA variation identification has improved the understanding of the disease complexity, provided SNPs as genetic markers of increased cancer susceptibility as well as biomarkers of cancer type, outcome and response to therapy (reviewed in (Pelletier and Weidhaas 2010)). Additionally, large scale and high throughput new generation sequencing (*i.e.* RNA-seq) will allow identification of miRNA somatic mutations and expression alterations in a large numbers of patients and cancer types. This will eventually sort out which miRNA mutations (SNPs, insertions, deletions, amplifications) are most frequent in a particular cancer type or stage of cancerogenesis, and how that affects global miRNA and target genes expression profiles. Such genome-wide understanding of miRNA somatic mutations and expression changes they cause are likely to identify novel and previously unappreciated mechanisms in oncogenesis.

6. **Three types of epigenetic concepts have been associated with miRNAs: A) miRNA-mediated gene silencing is one of the epigenetic mechanisms, B) miRNAs can also directly control the epigenetic machinery with a subclass of miRNAs (epi-miRNAs), and C) miRNA expression can be downregulated via promoter hypermethylation.**

MicroRNAs have been found associated with several epigenetic mechanisms (**Fig. 2**):

(**A**) MicroRNAs have the ability to regulate gene expression on a post-transcriptional level and are therefore themselves regarded as one of the classes of epigenetic mechanisms (reviewed in (Sharma, Kelly, and Jones 2010)). Together with other epigenetic mechanisms like promoter DNA methylation and histone modifications miRNAs are involved in an interacting network of epigenetic regulation of gene expression (reviewed in (Chuang and Jones 2007; Esteller 2008)).

(**B**) A subclass of miRNAs (epi-miRNAs) directly controls the epigenetic machinery through a regulatory loop by targeting its regulating enzymes. The first epi-miRNAs identified where the *MIR29* family (*MIR29A*, *MIR29B*, and *MIR29C*) shown to directly target and downregulate *de novo* DNA methyltransferases (*DNMT3A* and *DNMT3B*) and indirectly target *DNMT1* in lung cancer (Fabbri et al. 2007) and acute myeloid leukemia (Garzon et al. 2009). This led to demethylation of CpG islands in promoter regions of tumor-suppressor genes, allowing their reactivation and a loss of the cell's tumorigenicity (Fabbri et al. 2007). It was also reported that *MIR449A* targets histone deacetylase 1 (*HDAC1*), which is frequently overexpressed in many types of cancer, and, for example, induces growth arrest in prostate cancer (Noonan et al. 2009). *MIR101* was shown to directly modulate the expression of enhancer of zeste homolog 2 (EZH2), a catalytic subunit of the polycomb repressive complex 2 (PRC2), which mediates epigenetic silencing of tumor-suppressor genes in cancer (Friedman et al. 2009).

(**C**) Expression of miRNA genes has also been found silenced in human tumors by epigenetic mechanisms, such as aberrant hypermethylation of CpG islands encompassing or in proximity of miRNA genes, and/or by histone acetylation (Weber et al. 2007). The first evidence of deregulated miRNA expression in cancer due to an altered methylation status was reported for *MIR127*, embedded within a CpG island promoter, which was silenced in several cancer cells, but strongly upregulated after treatment with a hypomethylating agent (DNMT inhibitor) (Saito et al. 2006). A similar scenario was observed with *MIR124A* whose function can be restored by erasing DNA methylation, and has functional consequences on cyclin D kinase 6 (CDK6) activity (Lujambio et al. 2007). On the other hand, Brueckner *et al.* (Brueckner et al. 2007) observed that hypomethylation facilitated reactivation of *MIRLET7A3* and elevated its expression in human lung cancer cell lines, which resulted in enhanced tumor phenotypes. Compared to protein-coding genes, human oncomiRs were found to have an order of magnitude higher methylation frequency (Weber et al. 2007; Kunej et al. 2011). Epigenetically regulated miRNAs have been found to be associated with various types of cancer and also present cancer-specific biomarker potential (Kunej et al. 2011). Future studies of epigenetic regulation of miRNA expression together with downstream signaling pathways are likely to lead to development of novel drug targets in cancer therapy.

7. Aberrant miRNA gene expression signatures characterize cancer cells and miRNA profiling can be applied in diagnosis, prognosis, and treatment in cancer patients.

Profiling of miRNA transcriptome is used to document typical expression signatures of a particular cell or tissue and to identify variability in disease states, such as cancer. The main mechanism of miRNome alterations in cancer cells is represented by aberrant gene expression, characterized by abnormal levels of expression for mature and/or precursor miRNA sequences in comparison with the corresponding normal tissues. It was shown that miRNome signatures and aberrations from a wild type signature provide a more accurate diagnostic tool for cancer classification than the transcriptome of protein-coding genes (Lu, Getz, et al. 2005) (reviewed in (Calin and Croce 2006)). Deciphering the miRNome expression in normal and diseased states will be useful for the identification of miRNA targets, and alterations in miRNA expression patterns may disclose new pathogenic pathways in human tumorigenesis (Liu et al. 2004). Changes in miRNA expression pattern can be a consequence of mechanisms that can act independently or in combination, such as the location of miRNAs at CAGR, epigenetic regulation of miRNA expression, and abnormalities in miRNA processing genes and proteins, including mutations in *DICER1*, *TRBP* or *XPO5* (exportin 5). In cancer, the loss of tumor-suppressor miRNAs enhances the expression of target oncogenes, whereas increased expression of oncogenic miRNAs represses target tumor-suppressor genes. Paired expression profiles of miRNAs and mRNAs can be used to identify functional miRNA-target relationships with high precision (Huang et al. 2007). The aberrant expression of miRNAs in cancer is

characterized by abnormal levels of expression for mature and/or precursor miRNA transcripts in comparison to those in the corresponding normal tissue. Lu *et al*. (Lu, Getz, et al. 2005) observed a general down-regulation of miRNAs in tumor samples compared to normal tissue samples. It was also found that miRNA expression profiles could be used to differentiate human cancers according to their developmental origin, with cancers of epithelial and hematopoietic origin having distinct miRNA profiles (Lu, Getz, et al. 2005). Therefore, determining miRNA transcriptome profile in cancer cells or tissues and its deviations from expression signatures of normal cells or tissues are informative in identifying cancer sub-types and possibly even causal variability.

MicroRNAs profiling achieved by various methods has allowed the identification of signatures associated with diagnosis, staging, progression, prognosis, and response to treatment of human tumors. For example, the miRNA-based classifier is much better in establishing the correct diagnosis of metastatic cancer of unknown primary site than the classifier based on messenger RNAs of coding genes. As miRNA expression levels and tissue distribution pattern changes with differentiation, the poorly differentiated tumors have lower global expression levels of miRNAs compared with well-differentiated tumors from control groups (Lu, Getz, et al. 2005). Because reduced expression levels of miRNAs present a hallmark in poorly differentiated tumors, miRNA profiling can therefore present an effective tool in the diagnosis of cancer of unknown primary site.

Profiling of miRNA expression correlates well with clinical and biological characteristics of tumors and has enabled the identification of signatures associated with diagnosis, staging, progression, prognosis, and response to treatment of human tumors (reviewed in (Barbarotto, Schmittgen, and Calin 2008)). Profiling miRNA transcriptome of cancer cells or tissues therefore provides new insights of basic research interest as well as a novel fingerprinting tool to aid clinical oncology diagnostics.

8. Circulating miRNAs are potential noninvasive biomarkers in cancer.

MicroRNAs have been found to function not only within cells but can also act at neighboring cells and more distant sites within the body. The measurement of miRNAs in body fluids, including plasma and serum, has been used to distinguish cancer patients from healthy subjects (Mitchell et al. 2008). Since deregulated miRNA expression is an early indicator in tumorigenesis, circulating miRNAs can be used for cancer detection and therefore represent a gold mine for noninvasive biomarkers in cancer (reviewed in (Cortez et al. 2011)). Biomarker potential of serum miRNAs relies mainly on their high stability and resistance to storage handling; they remain stable after being subjected to severe conditions that would normally degrade most RNAs, such as boiling, very low or high pH levels, extended storage, and 10 freeze-thaw cycles (Chen et al. 2008). Chen et al. (Chen et al. 2008) identified expression patterns of serum miRNAs that were specific for lung and colorectal cancer, and diabetes, which provides evidence that serum miRNAs contain

fingerprints for various diseases (Chen et al. 2008). Correlation between circulating miRNA levels and response to a given anticancer agent was also observed and may be useful in predicting patterns of resistance and sensitivity to drugs used in cancer treatment. This was shown in the case of serum *MIR21* levels that were higher in hormone-refractory prostate cancer patients, whose disease was resistant to docetaxel-based chemotherapy, when compared to those with chemo-sensitive disease (Zhang et al. 2011). Additional and more detailed investigations of types and levels of circulating miRNAs, comparisons between cancer stages, types of cancer and treatments may provide novel and more precise clinical laboratory cancer biomarkers with established reference intervals to allow appropriate clinical interpretation and therapy.

9. RNA inhibition using miRNAs is a potential treatment method for specific types of cancer.

RNA inhibition by using miRNAs, although not the "universal panacea" for any type of cancer, could represent valid options for the treatment of specific patients in the near future. These patients should have a concordant expression between a specific miRNA and the experimentally proven targets. RNA inhibition can be used to treat cancer patients in two ways: a) by using RNA or DNA molecules as therapeutic drugs against messenger RNA of genes involved in the pathogenesis of cancers, and b) by directly targeting ncRNAs that participate in cancer pathogenesis (reviewed in (Spizzo et al. 2009)). The advantage by using miRNAs is double: first, miRNAs are naturally occurring in human cells (by difference to chemotherapies or antisense oligonucleotides), and second, miRNAs target multiple genes from the same pathway and therefore the action occurs at multiple levels in the same pathway (for example *MIR16* targets both anti-apoptotic genes *BCL2* (B-cell CLL/lymphoma 2) (Cimmino et al. 2005) and *MCL1* (myeloid cell leukemia sequence 1 (BCL2-related)) (Calin et al. 2008). On the other hand, approaches for sequence-specific inhibition with miRNAs in tumors address several difficulties, such as target specificity and delivery efficiency. MiRNA-mediated therapy may lead to unwanted gene silencing (off-target effect). Delivery of the therapeutic miRNAs to the target tissue without compromising the integrity of the miRNA remains challenging (reviewed in (Akhtar and Benter 2007; Castanotto and Rossi 2009; Whitehead, Langer, and Anderson 2009)). Current strategies for miRNA-based delivery use antisense oligonucleotides such as antagomirs, locked nucleic acid (LNA) anti-miR constructs, miRNA sponges, miR-masks, to block the oncogenic miRNAs, and synthetic miRNA mimics to restore miRNA expression (reviewed in (Garzon, Marcucci, and Croce 2010)). Appropriate target gene selection and therapeutic molecule design are crucial for efficient therapeutic design.

Virus-mediated delivery of *MIR26A, that is* normally expressed at high levels in diverse tissues but reduced in HCC cells, for liver cancer treatment in the mouse model resulted in inhibition of cancer cell proliferation, induction of tumor-specific apoptosis, and protection

from disease progression (Kota et al. 2009). Exogenous delivery of synthetic *MIRLET7* miRNA to established tumors significantly reduced tumor growth in mouse models of lung cancer (Trang et al. 2010). Therapeutic delivery of *MIR34A* mimic strongly inhibited cancer cell growth in mouse models for prostate (Liu et al. 2011) and lung cancer (Wiggins et al. 2010). Virus-mediated delivery of *MIR145* combined with 5-fluorourasil (5-FU) showed significant inhibition of tumor growth in breast tumor bearing mice (Kim et al. 2011). Intravenous administration of antagomirs against *MIR16*, *MIR122*, *MIR192* and *MIR194* effectively inhibited corresponding miRNA levels in liver, lung, kidney, heart, intestine, fat, skin, bone marrow, muscle, ovaries and adrenals (Krützfeldt et al. 2005). *MIR10B* silencing did not inhibit the growth of the primary tumor, but drastically decreased the number of pulmonary metastases (Ma et al. 2010). The results of miRNA inhibition studies suggest great potential for miRNAs as a powerful tool for gene regulation research and therapeutic intervention.

10. Interplay between miRNAs, other ncRNAs, and protein-coding genes forms a complex network of interactions in normal and disease tissues.

Despite the leading role of miRNAs as cancer-related ncRNAs in published research, recently new categories of not-translated RNAs have emerged. Other ncRNAs such as lncRNAs, including long intergenic non-coding RNAs (lincRNAs) (Gupta et al. 2010) and UCGs (Calin et al. 2007) were found to be abnormally expressed in cancer and involved in tumorigenic mechanisms. As the spectrum of ncRNAs is much larger than that of miRNAs (the estimates are as high as 1,000,000 ncRNA transcripts versus as many as 10,000 potential miRNAs), it will have a profound impact on any aspect of basic and translational cancer research.

Recent analyses of miRNA genes, their targets and genes encoding for processing machinery, genetic polymorphisms, and epigenetic modifications revealed that miRNA-mediated regulation in gene regulatory networks involves a far more complex system than initially expected. MicroRNA genes are linked with TFs in complex regulatory networks where they reciprocally regulate one another (Yu et al. 2008). It has been shown that tumor-associated transcribed-UCRs (T-UCRs) in leukemias are negatively regulated by direct interaction with miRNAs (Calin et al. 2007). The competing endogenous RNA (ceRNA) hypothesis asserts that RNA transcripts can indirectly regulate each other by competing for binding to miRNAs (Salmena et al. 2011). An example for ceRNA activity was described for tumor suppressor phosphatase and tensin homolog (*PTEN*) and its highly homologous transcript from phosphatase and tensin homolog pseudogene 1 (*PTENP1*) (Poliseno et al. 2010). The pseudogene transcript can compete with *PTEN* mRNA for miRNA binding and thereby modulate the expression of *PTEN*. Additionally, recent evidences implicate the inter-connection of miRNAs and epigenetics. A subclass of miRNAs (epi-miRNAs) directly controls the epigenetic machinery through a regulatory loop by targeting key enzymes involved in establishing epigenetic memory (reviewed in

(Chuang and Jones 2007)). Finally, the miRNA decoy functions have an effect on therapeutic approaches in human diseases, which include specific ways to overcome resistance to drug therapy and design of miRNA-based clinical trials in the future (reviewed in (Almeida, Reis, and Calin 2012)). The broken interactions within the complex network of miRNA regulation may lead to great disruption in the cell, possibly leading to *cancerous phenotypes.*

III. CONCLUSIONS

There is no more doubt that miRNAs are involved in the regulation of tumorigenic pathways involved in tumor initiation, development, progression and dissemination. The question of whether miRNAs represent the "dark side" of cancer predisposition started to be answered by studies in large populations of cancer patients. MicroRNAs are identified as significant new diagnostic and prognostic tools for cancer patients, and the miRNA-based cancer therapy should represent a future option for medical oncologists. The progressively increasing understanding of the implications of ncRNAs for the malignant phenotype represents the essential background to achieve the goal for earlier detection and more effective treatment of cancer patients.

IV. LITERATURE CITED

Akhtar, S., and I. F. Benter. 2007. Nonviral delivery of synthetic siRNAs in vivo. *J Clin Invest* 117 (12):3623-32.

Almeida, M. I., R. M. Reis, and G. A. Calin. 2012. Decoy activity through microRNAs: the therapeutic implications. *Expert Opin Biol Ther*.

Ambros, V. 2004. The functions of animal microRNAs. *Nature* 431 (7006):350-5.

Barbarotto, E., T. D. Schmittgen, and G. A. Calin. 2008. MicroRNAs and cancer: profile, profile, profile. *Int J Cancer* 122 (5):969-77.

Barik, S. 2008. An intronic microRNA silences genes that are functionally antagonistic to its host gene. *Nucleic Acids Res* 36 (16):5232-41.

Bartel, David P. 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* 116 (2):281-297.

Baskerville, S, and DP Bartel. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *Rna-a Publication of the Rna Society*:241-247.

Brennecke, J., D. R. Hipfner, A. Stark, R. B. Russell, and S. M. Cohen. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* 113 (1):25-36.

Brueckner, Bodo, Carlo Stresemann, Ruprecht Kuner, Cora Mund, Tanja Musch, Michael Meister, Holger Sültmann, and Frank Lyko. 2007. The Human let-7a-3 Locus Contains an Epigenetically Regulated MicroRNA Gene with Oncogenic Function. *Cancer Research* 67 (4):1419-1423.

Calin, G. A., A. Cimmino, M. Fabbri, M. Ferracin, S. E. Wojcik, M. Shimizu, C. Taccioli, N. Zanesi, R. Garzon, R. I. Aqeilan, H. Alder, S. Volinia, L. Rassenti, X. Liu, C. G. Liu, T. J. Kipps, M. Negrini, and C. M. Croce. 2008. MiR-15a and miR-16-1 cluster functions in human leukemia. *Proc Natl Acad Sci U S A* 105 (13):5166-71.

Calin, G. A., and C. M. Croce. 2006. MicroRNA signatures in human cancers. *Nat Rev Cancer* 6 (11):857-66.

Calin, G. A., C. D. Dumitru, M. Shimizu, R. Bichi, S. Zupo, E. Noch, H. Aldler, S. Rattan, M. Keating, K. Rai, L. Rassenti, T. Kipps, M. Negrini, F. Bullrich, and C. M. Croce. 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 99 (24):15524-9.

Calin, G. A., M. Ferracin, A. Cimmino, G. Di Leva, M. Shimizu, S. E. Wojcik, M. V. Iorio, R. Visone, N. I. Sever, M. Fabbri, R. Iuliano, T. Palumbo, F. Pichiorri, C. Roldo, R. Garzon, C. Sevignani, L. Rassenti, H. Alder, S. Volinia, C. G. Liu, T. J. Kipps, M. Negrini, and C. M. Croce. 2005. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* 353 (17):1793-801.

Calin, G. A., C. G. Liu, M. Ferracin, T. Hyslop, R. Spizzo, C. Sevignani, M. Fabbri, A. Cimmino, E. J. Lee, S. E. Wojcik, M. Shimizu, E. Tili, S. Rossi, C. Taccioli, F. Pichiorri, X. Liu, S. Zupo, V. Herlea, L. Gramantieri, G. Lanza, H. Alder, L. Rassenti, S. Volinia, T. D. Schmittgen, T. J. Kipps, M. Negrini, and C. M. Croce. 2007. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 12 (3):215-29.

Calin, George Adrian, Cinzia Sevignani, Calin Dan Dumitru, Terry Hyslop, Evan Noch, Sai Yendamuri, Masayoshi Shimizu, Sashi Rattan, Florencia Bullrich, Massimo Negrini, and Carlo M. Croce. 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America* 101 (9):2999-3004.

Castanotto, D., and J. J. Rossi. 2009. The promises and pitfalls of RNA-interference-based therapeutics. *Nature* 457 (7228):426-33.

Chen, C. Z., L. Li, H. F. Lodish, and D. P. Bartel. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303 (5654):83-6.

Chen, X., Y. Ba, L. Ma, X. Cai, Y. Yin, K. Wang, J. Guo, Y. Zhang, J. Chen, X. Guo, Q. Li, X. Li, W. Wang, J. Wang, X. Jiang, Y. Xiang, C. Xu, P. Zheng, J. Zhang, R. Li, H. Zhang, X. Shang, T. Gong, G. Ning, K. Zen, and C. Y. Zhang. 2008. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res* 18 (10):997-1006.

Chin, L. J., E. Ratner, S. Leng, R. Zhai, S. Nallur, I. Babar, R. U. Muller, E. Straka, L. Su, E. A. Burki, R. E. Crowell, R. Patel, T. Kulkarni, R. Homer, D. Zelterman, K. K. Kidd, Y. Zhu, D. C. Christiani, S. A. Belinsky, F. J. Slack, and J. B. Weidhaas. 2008. A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res* 68 (20):8535-40.

Chuang, J. C., and P. A. Jones. 2007. Epigenetics and microRNAs. *Pediatr Res* 61 (5 Pt 2):24R-29R.

Cimmino, A., G. A. Calin, M. Fabbri, M. V. Iorio, M. Ferracin, M. Shimizu, S. E. Wojcik, R. I. Aqeilan, S. Zupo, M. Dono, L. Rassenti, H. Alder, S. Volinia, C. G. Liu, T. J. Kipps, M. Negrini, and C. M. Croce. 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* 102 (39):13944-9.

Clague, J., S. M. Lippman, H. Yang, M. A. Hildebrandt, Y. Ye, J. J. Lee, and X. Wu. 2010. Genetic variation in MicroRNA genes and risk of oral premalignant lesions. *Mol Carcinog* 49 (2):183-9.

Cortez, M. A., C. Bueso-Ramos, J. Ferdin, G. Lopez-Berestein, A. K. Sood, and G. A. Calin. 2011. MicroRNAs in body fluids--the mix of hormones and biomarkers. *Nat Rev Clin Oncol* 8 (8):467-77.

Costinean, S., N. Zanesi, Y. Pekarsky, E. Tili, S. Volinia, N. Heerema, and C. M. Croce. 2006. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A* 103 (18):7024-9.

Croce, C. M. 2009. Causes and consequences of microRNA dysregulation in cancer. *Nature Reviews Genetics* 10 (10):704-714.

Denli, A. M., B. B. Tops, R. H. Plasterk, R. F. Ketting, and G. J. Hannon. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* 432 (7014):231-5.

Diederichs, S., and D. A. Haber. 2006. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. *Cancer Res* 66 (12):6097-104.

Eiring, A. M., J. G. Harb, P. Neviani, C. Garton, J. J. Oaks, R. Spizzo, S. Liu, S. Schwind, R. Santhanam, C. J. Hickey, H. Becker, J. C. Chandler, R. Andino, J. Cortes, P. Hokland, C. S. Huettner, R. Bhatia, D. C. Roy, S. A. Liebhaber, M. A. Caligiuri, G. Marcucci, R. Garzon, C. M. Croce, G. A. Calin, and D. Perrotti. 2010. miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. *Cell* 140 (5):652-65.

Erdmann, V. A., M. Szymanski, A. Hochberg, N. Groot, and J. Barciszewski. 2000. Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res* 28 (1):197-200.

Esquela-Kerscher, Aurora, and Frank J. Slack. 2006. Oncomirs [mdash] microRNAs with a role in cancer. *Nat Rev Cancer* 6 (4):259-269.

Esteller, M. 2008. Epigenetics in cancer. *N Engl J Med* 358 (11):1148-59.

Fabbri, M., R. Garzon, A. Cimmino, Z. Liu, N. Zanesi, E. Callegari, S. Liu, H. Alder, S. Costinean, C. Fernandez-Cymering, S. Volinia, G. Guler, C. D. Morrison, K. K. Chan, G. Marcucci, G. A. Calin, K. Huebner, and C. M. Croce. 2007. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc Natl Acad Sci U S A* 104 (40):15805-10.

Fabian, M. R., N. Sonenberg, and W. Filipowicz. 2010. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 79:351-79.

Felli, N., L. Fontana, E. Pelosi, R. Botta, D. Bonci, F. Facchiano, F. Liuzzi, V. Lulli, O. Morsilli, S. Santoro, M. Valtieri, G. A. Calin, C. G. Liu, A. Sorrentino, C. M. Croce, and C. Peschle. 2005. MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. *Proc Natl Acad Sci U S A* 102 (50):18081-6.

Fish, J. E., M. M. Santoro, S. U. Morton, S. Yu, R. F. Yeh, J. D. Wythe, K. N. Ivey, B. G. Bruneau, D. Y. Stainier, and D. Srivastava. 2008. miR-126 regulates angiogenic signaling and vascular integrity. *Dev Cell* 15 (2):272-84.

Forman, J. J., A. Legesse-Miller, and H. A. Coller. 2008. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A* 105 (39):14879-84.

Friedman, J. M., G. Liang, C. C. Liu, E. M. Wolff, Y. C. Tsai, W. Ye, X. Zhou, and P. A. Jones. 2009. The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. *Cancer Res* 69 (6):2623-9.

Garofalo, M., G. Di Leva, G. Romano, G. Nuovo, S. S. Suh, A. Ngankeu, C. Taccioli, F. Pichiorri, H. Alder, P. Secchiero, P. Gasparini, A. Gonelli, S. Costinean, M. Acunzo, G. Condorelli, and C. M. Croce. 2009. miR-221&222 regulate TRAIL resistance and enhance tumorigenicity through PTEN and TIMP3 downregulation. *Cancer Cell* 16 (6):498-509.

Garzon, R., S. J. Liu, M. Fabbri, Z. F. Liu, C. E. A. Heaphy, E. Callegari, S. Schwind, J. X. Pang, J. H. Yu, N. Muthusamy, V. Havelange, S. Volinia, W. Blum, L. J. Rush, D. Perrotti, M. Andreeff, C. D. Bloomfield, J. C. Byrd, K. Chan, L. C. Wu, C. M. Croce, and G. Marcucci. 2009. MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. *Blood* 113 (25):6411-6418.

Garzon, R., G. Marcucci, and C. M. Croce. 2010. Targeting microRNAs in cancer: rationale, strategies and challenges. *Nat Rev Drug Discov* 9 (10):775-89.

Gaur, A., D. A. Jewell, Y. Liang, D. Ridzon, J. H. Moore, C. Chen, V. R. Ambros, and M. A. Israel. 2007. Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Res* 67 (6):2456-68.

Georges, M., W. Coppieters, and C. Charlier. 2007. Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. *Curr Opin Genet Dev* 17 (3):166-76.

Giraldez, A. J., R. M. Cinalli, M. E. Glasner, A. J. Enright, J. M. Thomson, S. Baskerville, S. M. Hammond, D. P. Bartel, and A. F. Schier. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* 308 (5723):833-8.

Gregory, R. I., K. P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432 (7014):235-40.

Gupta, R. A., N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M. C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar, and H. Y. Chang. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464 (7291):1071-6.

Ha, T. Y. 2011. MicroRNAs in Human Diseases: From Cancer to Cardiovascular Disease. *Immune Netw* 11 (3):135-54.

Hammond, S. M., S. Boettcher, A. A. Caudy, R. Kobayashi, and G. J. Hannon. 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* 293 (5532):1146-50.

Horikawa, Y., C. G. Wood, H. Yang, H. Zhao, Y. Ye, J. Gu, J. Lin, T. Habuchi, and X. Wu. 2008. Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma. *Clin Cancer Res* 14 (23):7956-62.

Hu, Z., J. Liang, Z. Wang, T. Tian, X. Zhou, J. Chen, R. Miao, Y. Wang, X. Wang, and H. Shen. 2009. Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. *Hum Mutat* 30 (1):79-84.

Huang, J. C., T. Babak, T. W. Corson, G. Chua, S. Khan, B. L. Gallie, T. R. Hughes, B. J. Blencowe, B. J. Frey, and Q. D. Morris. 2007. Using expression profiling data to identify human microRNA targets. *Nat Methods* 4 (12):1045-9.

Hwang, H. W., E. A. Wentzel, and J. T. Mendell. 2007. A hexanucleotide element directs microRNA nuclear import. *Science* 315 (5808):97-100.

Johnson, S. M., H. Grosshans, J. Shingara, M. Byrom, R. Jarvis, A. Cheng, E. Labourier, K. L. Reinert, D. Brown, and F. J. Slack. 2005. RAS is regulated by the let-7 microRNA family. *Cell* 120 (5):635-47.

Kim, S. J., J. S. Oh, J. Y. Shin, K. D. Lee, K. W. Sung, S. J. Nam, and K. H. Chun. 2011. Development of microRNA-145 for therapeutic application in breast cancer. *J Control Release* 155 (3):427-34.

Kim, Y. K., and V. N. Kim. 2007. Processing of intronic microRNAs. *EMBO J* 26 (3):775-83.

Kota, J., R. R. Chivukula, K. A. O'Donnell, E. A. Wentzel, C. L. Montgomery, H. W. Hwang, T. C. Chang, P. Vivekanandan, M. Torbenson, K. R. Clark, J. R. Mendell, and J. T. Mendell. 2009. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell* 137 (6):1005-17.

Krek, A., D. Grün, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus, M. Stoffel, and N. Rajewsky. 2005. Combinatorial microRNA target predictions. *Nat Genet* 37 (5):495-500.

Krützfeldt, J., N. Rajewsky, R. Braich, K. G. Rajeev, T. Tuschl, M. Manoharan, and M. Stoffel. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 438 (7068):685-9.

Kumar, M. S., J. Lu, K. L. Mercer, T. R. Golub, and T. Jacks. 2007. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet* 39 (5):673-7.

Kunej, T., I. Godnic, J. Ferdin, S. Horvat, P. Dovc, and G. A. Calin. 2011. Epigenetic regulation of microRNAs in cancer: an integrated review of literature. *Mutat Res* 717 (1-2):77-84.

Kunej, T., I. Godnic, S. Horvat, M. Zorc, and G. A. Calin. 2012. Cross Talk Between MicroRNA and Coding Cancer Genes. *Cancer J* 18 (3):223-31.

Kunej, Tanja, Dasa Jevsinek Skok, Minja Zorc, Ana Ogrinc, Jennifer J. Michal, Milena Kovac, and Zhihua Jiang. 2012. Obesity Gene Atlas in Mammals. *Journal of Genomics*.

Landi, Debora, Federica Gemignani, Roberto Barale, and Stefano Landi. 2007. A Catalog of Polymorphisms Falling in MicroRNA-Binding Regions of Cancer Genes. *DNA and Cell Biology* 27 (1):35-43.

Lee, R. C., R. L. Feinbaum, and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75 (5):843-54.

Lee, Y. S., H. K. Kim, S. Chung, K. S. Kim, and A. Dutta. 2005. Depletion of human micro-RNA miR-125b reveals that it is critical for the proliferation of differentiated cells but not for the down-regulation of putative targets during differentiation. *J Biol Chem* 280 (17):16635-41.

Lewis, Benjamin P., Christopher B. Burge, and David P. Bartel. 2005. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* 120 (1):15-20.

Li, S. C., P. Tang, and W. C. Lin. 2007. Intronic microRNA: discovery and biological implications. *DNA Cell Biol* 26 (4):195-207.

Liao, J. Y., L. M. Ma, Y. H. Guo, Y. C. Zhang, H. Zhou, P. Shao, Y. Q. Chen, and L. H. Qu. 2010. Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. *PLoS One* 5 (5):e10563.

Liu, C. G., G. A. Calin, B. Meloon, N. Gamliel, C. Sevignani, M. Ferracin, C. D. Dumitru, M. Shimizu, S. Zupo, M. Dono, H. Alder, F. Bullrich, M. Negrini, and C. M. Croce. 2004. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci U S A* 101 (26):9740-4.

Liu, C., K. Kelnar, B. Liu, X. Chen, T. Calhoun-Davis, H. Li, L. Patrawala, H. Yan, C. Jeter, S. Honorio, J. F. Wiggins, A. G. Bader, R. Fagin, D. Brown, and D. G. Tang. 2011. The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. *Nat Med* 17 (2):211-5.

Lu, C., S. S. Tej, S. Luo, C. D. Haudenschild, B. C. Meyers, and P. J. Green. 2005. Elucidation of the small RNA component of the transcriptome. *Science* 309 (5740):1567-9.

Lu, Jun, Gad Getz, Eric A. Miska, Ezequiel Alvarez-Saavedra, Justin Lamb, David Peck, Alejandro Sweet-Cordero, Benjamin L. Ebert, Raymond H. Mak, Adolfo A. Ferrando, James R. Downing, Tyler Jacks, H. Robert Horvitz, and Todd R. Golub. 2005. MicroRNA expression profiles classify human cancers. *Nature* 435 (7043):834-838.

Lujambio, Amaia, Santiago Ropero, Esteban Ballestar, Mario F. Fraga, Celia Cerrato, Fernando Setién, Sara Casado, Ana Suarez-Gauthier, Montserrat Sanchez-

Cespedes, Anna Gitt, Inmaculada Spiteri, Partha P. Das, Carlos Caldas, Eric Miska, and Manel Esteller. 2007. Genetic Unmasking of an Epigenetically Silenced microRNA in Human Cancer Cells. *Cancer Research* 67 (4):1424-1429.

Lytle, J. R., T. A. Yario, and J. A. Steitz. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* 104 (23):9667-72.

Ma, L., F. Reinhardt, E. Pan, J. Soutschek, B. Bhat, E. G. Marcusson, J. Teruya-Feldstein, G. W. Bell, and R. A. Weinberg. 2010. Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. *Nat Biotechnol* 28 (4):341-7.

Medina, P. P., M. Nolde, and F. J. Slack. 2010. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature* 467 (7311):86-90.

Mitchell, P. S., R. K. Parkin, E. M. Kroh, B. R. Fritz, S. K. Wyman, E. L. Pogosova-Agadjanyan, A. Peterson, J. Noteboom, K. C. O'Briant, A. Allen, D. W. Lin, N. Urban, C. W. Drescher, B. S. Knudsen, D. L. Stirewalt, R. Gentleman, R. L. Vessella, P. S. Nelson, D. B. Martin, and M. Tewari. 2008. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 105 (30):10513-8.

Nicoloso, M. S., H. Sun, R. Spizzo, H. Kim, P. Wickramasinghe, M. Shimizu, S. E. Wojcik, J. Ferdin, T. Kunej, L. Xiao, S. Manoukian, G. Secreto, F. Ravagnani, X. Wang, P. Radice, C. M. Croce, R. V. Davuluri, and G. A. Calin. 2010. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res* 70 (7):2789-98.

Noonan, E. J., R. F. Place, D. Pookot, S. Basak, J. M. Whitson, H. Hirata, C. Giardina, and R. Dahiya. 2009. miR-449a targets HDAC-1 and induces growth arrest in prostate cancer. *Oncogene* 28 (14):1714-24.

Pelletier, C., and J. B. Weidhaas. 2010. MicroRNA binding site polymorphisms as biomarkers of cancer risk. *Expert Rev Mol Diagn* 10 (6):817-29.

Petrocca, F., R. Visone, M. R. Onelli, M. H. Shah, M. S. Nicoloso, I. de Martino, D. Iliopoulos, E. Pilozzi, C. G. Liu, M. Negrini, L. Cavazzini, S. Volinia, H. Alder, L. P. Ruco, G. Baldassarre, C. M. Croce, and A. Vecchione. 2008. E2F1-regulated microRNAs impair TGFbeta-dependent cell-cycle arrest and apoptosis in gastric cancer. *Cancer Cell* 13 (3):272-86.

Place, R. F., L. C. Li, D. Pookot, E. J. Noonan, and R. Dahiya. 2008. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* 105 (5):1608-13.

Poliseno, L., L. Salmena, J. Zhang, B. Carver, W. J. Haveman, and P. P. Pandolfi. 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465 (7301):1033-8.

Poy, M. N., L. Eliasson, J. Krutzfeldt, S. Kuwajima, X. Ma, P. E. Macdonald, S. Pfeffer, T. Tuschl, N. Rajewsky, P. Rorsman, and M. Stoffel. 2004. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 432 (7014):226-30.

Raveche, E. S., E. Salerno, B. J. Scaglione, V. Manohar, F. Abbasi, Y. C. Lin, T. Fredrickson, P. Landgraf, S. Ramachandra, K. Huppi, J. R. Toro, V. E. Zenger, R. A. Metcalf, and G. E. Marti. 2007. Abnormal microRNA-16 locus with synteny to human 13q14 linked to CLL in NZB mice. *Blood* 109 (12):5079-86.

Rearick, D., A. Prakash, A. McSweeny, S. S. Shepard, L. Fedorova, and A. Fedorov. 2011. Critical association of ncRNA with introns. *Nucleic Acids Res* 39 (6):2357-66.

Rodriguez, Antony, Sam Griffiths-Jones, Jennifer L. Ashurst, and Allan Bradley. 2004. Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Research* 14 (10a):1902-1910.

Saito, Yoshimasa, Jeffrey M. Friedman, Yoshitomo Chihara, Gerda Egger, Jody C. Chuang, and Gangning Liang. 2009. Epigenetic therapy upregulates the tumor suppressor microRNA-126 and its host gene EGFL7 in human cancer cells. *Biochemical and Biophysical Research Communications* 379 (3):726-731.

Saito, Yoshimasa, Gangning Liang, Gerda Egger, Jeffrey M. Friedman, Jody C. Chuang, Gerhard A. Coetzee, and Peter A. Jones. 2006. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* 9 (6):435-443.

Salmena, L., L. Poliseno, Y. Tay, L. Kats, and P. P. Pandolfi. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146 (3):353-8.

Santarpia, L., M. Nicoloso, and G. A. Calin. 2010. MicroRNAs: a complex regulatory network drives the acquisition of malignant cell phenotype. *Endocr Relat Cancer* 17 (1):F51-75.

Sempere, L. F., S. Freemantle, I. Pitha-Rowe, E. Moss, E. Dmitrovsky, and V. Ambros. 2004. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol* 5 (3):R13.

Sharma, Shikhar, Theresa K. Kelly, and Peter A. Jones. 2010. Epigenetics in cancer. *Carcinogenesis* 31 (1):27-36.

Shen, Jie, Christine B. Ambrosone, Richard A. DiCioccio, Kunle Odunsi, Shashikant B. Lele, and Hua Zhao. 2008. A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. *Carcinogenesis* 29 (10):1963-1966.

Spizzo, R., D. Rushworth, M. Guerrero, and G. A. Calin. 2009. RNA inhibition, microRNAs, and new therapeutic agents for cancer treatment. *Clin Lymphoma Myeloma* 9 Suppl 3:S313-8.

Stahlhut Espinosa, C. E., and F. J. Slack. 2006. The role of microRNAs in cancer. *Yale J Biol Med* 79 (3-4):131-40.

Sun, Guihua, Jin Yan, Katie Noltner, Jinong Feng, Haitang Li, Daniel A. Sarkis, Steve S. Sommer, and John J. Rossi. 2009. SNPs in human miRNA genes affect biogenesis and function. *RNA* 15 (9):1640-1651.

Sung, H., S. Jeon, K. M. Lee, S. Han, M. Song, J. Y. Choi, S. K. Park, K. Y. Yoo, D. Y. Noh, S. H. Ahn, and D. Kang. 2012. Common genetic polymorphisms of microRNA biogenesis pathway genes and breast cancer survival. *BMC Cancer* 12 (1):195.

Takamizawa, J., H. Konishi, K. Yanagisawa, S. Tomida, H. Osada, H. Endoh, T. Harano, Y. Yatabe, M. Nagino, Y. Nimura, T. Mitsudomi, and T. Takahashi. 2004. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* 64 (11):3753-6.

Tay, Y., J. Zhang, A. M. Thomson, B. Lim, and I. Rigoutsos. 2008. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455 (7216):1124-8.

Tian, T., Y. Shu, J. Chen, Z. Hu, L. Xu, G. Jin, J. Liang, P. Liu, X. Zhou, R. Miao, H. Ma, Y. Chen, and H. Shen. 2009. A functional genetic variant in microRNA-196a2 is

associated with increased susceptibility of lung cancer in Chinese. *Cancer Epidemiol Biomarkers Prev* 18 (4):1183-7.

Trang, P., P. P. Medina, J. F. Wiggins, L. Ruffino, K. Kelnar, M. Omotola, R. Homer, D. Brown, A. G. Bader, J. B. Weidhaas, and F. J. Slack. 2010. Regression of murine lung tumors by the let-7 microRNA. *Oncogene* 29 (11):1580-7.

Vasudevan, S., Y. Tong, and J. A. Steitz. 2007. Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318 (5858):1931-4.

Wang, S., A. B. Aurora, B. A. Johnson, X. Qi, J. McAnally, J. A. Hill, J. A. Richardson, R. Bassel-Duby, and E. N. Olson. 2008. The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis. *Dev Cell* 15 (2):261-71.

Weber, Barbara, Carlo Stresemann, Bodo Brueckner, and Frank Lyko. 2007. Methylation of Human MicroRNA Genes in Normal and Neoplastic Cells. *Cell Cycle* 6 (9):1001-1005.

White, N. M., T. T. Bao, J. Grigull, Y. M. Youssef, A. Girgis, M. Diamandis, E. Fatoohi, M. Metias, R. J. Honey, R. Stewart, K. T. Pace, G. A. Bjarnason, and G. M. Yousef. 2011. miRNA profiling for clear cell renal cell carcinoma: biomarker discovery and identification of potential controls and consequences of miRNA dysregulation. *J Urol* 186 (3):1077-83.

Whitehead, K. A., R. Langer, and D. G. Anderson. 2009. Knocking down barriers: advances in siRNA delivery. *Nat Rev Drug Discov* 8 (2):129-38.

Wiggins, J. F., L. Ruffino, K. Kelnar, M. Omotola, L. Patrawala, D. Brown, and A. G. Bader. 2010. Development of a lung cancer therapeutic based on the tumor suppressor microRNA-34. *Cancer Res* 70 (14):5923-30.

Yang, H., C. P. Dinney, Y. Ye, Y. Zhu, H. B. Grossman, and X. Wu. 2008. Evaluation of genetic variants in microRNA-related genes and risk of bladder cancer. *Cancer Res* 68 (7):2530-7.

Yu, X., J. Lin, D. J. Zack, J. T. Mendell, and J. Qian. 2008. Analysis of regulatory network topology reveals functionally distinct classes of microRNAs. *Nucleic Acids Res* 36 (20):6494-503.

Zeng, Y., and B. R. Cullen. 2005. Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* 280 (30):27595-603.

Zhang, H. L., L. F. Yang, Y. Zhu, X. D. Yao, S. L. Zhang, B. Dai, Y. P. Zhu, Y. J. Shen, G. H. Shi, and D. W. Ye. 2011. Serum miRNA-21: elevated levels in patients with metastatic hormone-refractory prostate cancer and potential predictive factor for the efficacy of docetaxel-based chemotherapy. *Prostate* 71 (3):326-31.

Zorc, M., D. Jevsinek Skok, I. Godnic, G. A. Calin, S. Horvat, Z. Jiang, P. Dovc, and T. Kunej. 2012. Catalog of MicroRNA Seed Polymorphisms in Vertebrates. *PLoS One* 7 (1):e30737.

FIGURES:

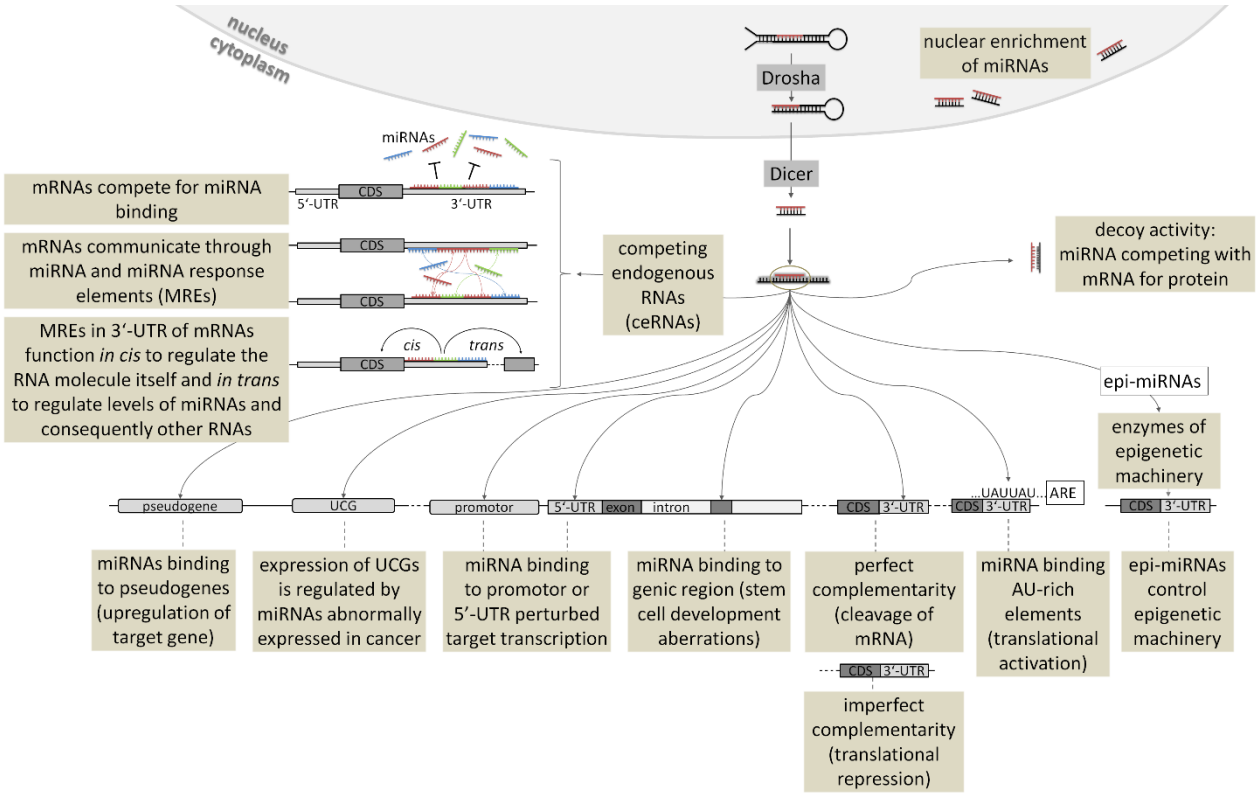


Figure 1: MicroRNAs can target and regulate different genic regions: promoters, 3'- and 5'-untranslated regions (UTRs), coding sequences (CDS), AU-rich elements (AREs), other ncRNAs, including ultraconserved genes (UCGs), and proteins. Competing endogenous RNA (ceRNA) in this case represent an interacting activity between mRNAs and miRNAs.

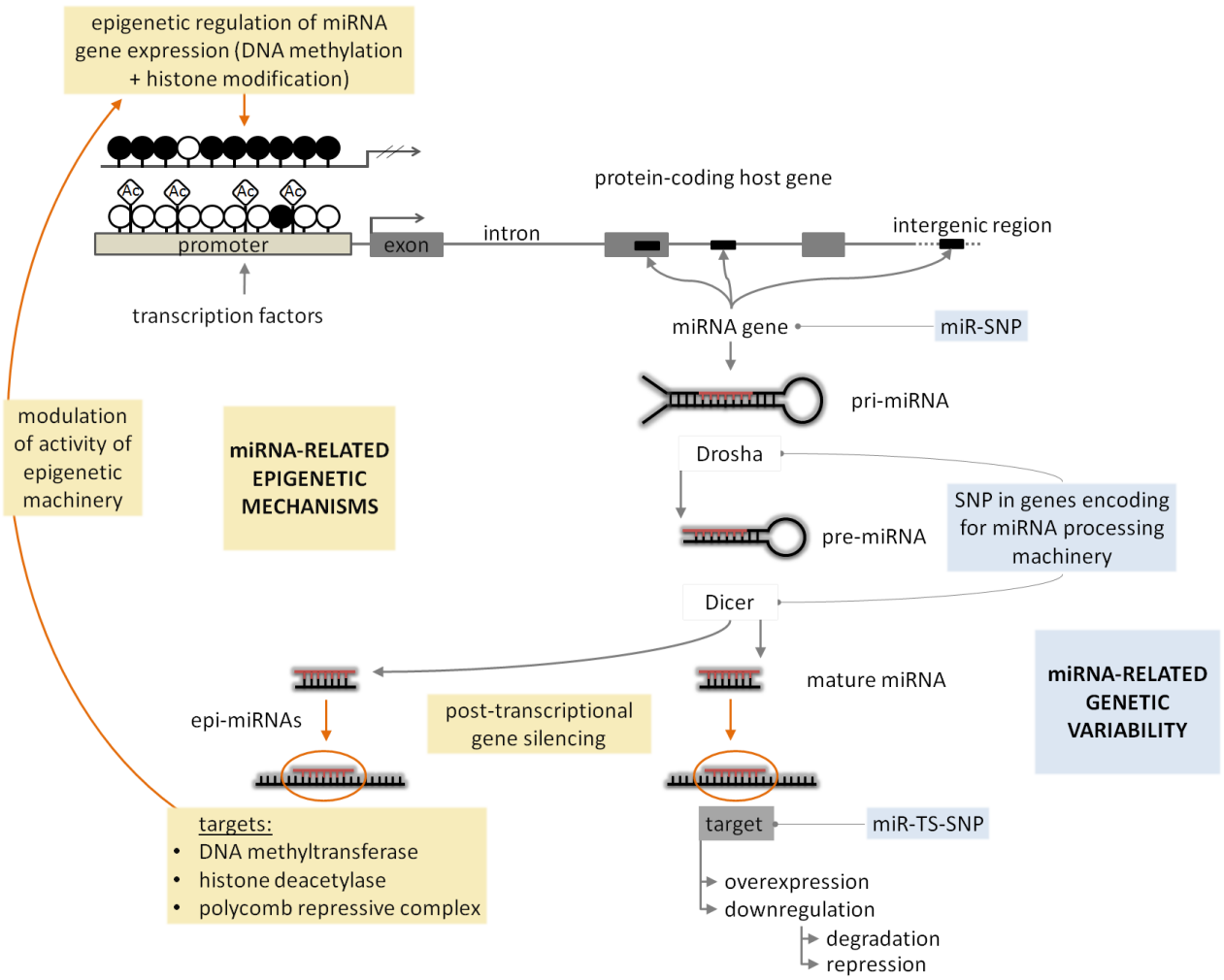


Figure 2: Interplay of mechanisms affecting miRNA biogenesis and function: genomic location of miRNA genes, polymorphisms, transcription factors, and epigenetic mechanisms. Ac – acetyl groups; *empty circles* – unmethylated CpG sites; *filled circles* – methylated CpG circles; SNP – single nucleotide polymorphism; miR-SNP – SNP located within the miRNA gene; miR-TS-SNP – SNP located within miRNA target site.

Priloga F

Cross talk between microRNA and coding cancer genes (Kunej in sod., 2012)

Cross Talk Between MicroRNA and Coding Cancer Genes

Tanja Kunej, PhD, Irena Godnic, DVM,* Simon Horvat, PhD,*† Minja Zorc, BSc,*‡ and George A. Calin, MD, PhD§*

Abstract: MicroRNAs (miRNAs) are a class of noncoding RNAs (ncRNAs) and posttranscriptional gene regulators shown to be involved in pathogenesis of all types of human cancers. Their aberrant expression as tumor suppressors can lead to cancerogenesis by inhibiting malignant potential, or when acting as oncogenes, by activating malignant potential. Differential expression of miRNA genes in tumorous tissues can occur owing to several factors including positional effects when mapping to cancer-associated genomic regions, epigenetic mechanisms, and malfunctioning of the miRNA processing machinery, all of which can contribute to a complex miRNA-mediated gene network misregulation. They may increase or decrease expression of protein-coding genes, can target 3′-UTR or other genic regions (5′-UTR, promoter, coding sequences), and can function in various subcellular compartments, developmental, and metabolic processes. Because expanding research on miRNA-cancer associations has already produced large amounts of data, our main objective here was to summarize main findings and critically examine the intricate network connecting the miRNAs and coding genes in regulatory mechanisms and their function and phenotypic consequences for cancer. By examining such interactions, we aimed to gain insights for the development of new diagnostic markers as well as identification of potential venues for more selective tumor therapy. To enable efficient examination of the main past and current miRNA discoveries, we developed a Web-based miRNA timeline tool that will be regularly updated (http://www.integratomics-time.com/miRNA_timeline). Further development of this tool will be directed at providing additional analyses to clarify complex network interactions between miRNAs, other classes of ncRNAs, and protein-coding genes and their involvement in development of diseases including cancer. This tool therefore provides curated relevant information about the miRNA basic research and therapeutic application all at hand on one site to help researchers and clinicians in making informed decision about their miRNA cancer–related research or clinical practice.

Key Words: microRNA, cancer, oncogene, tumor suppressor, epigenetics, genetic variation, transcriptional regulation

(*Cancer J* 2012;18: 223–231)

Cancer develops through a complex multistep process involving structural and expression abnormalities of genes, including those encoding microRNAs (miRNAs).[1,2] MicroRNAs are a class of non–protein-coding RNAs that post-transcriptionally regulate the expression of target mRNAs. MicroRNAs that have been associated with cancer are referred to as oncomiRs.[3,4] The role of miRNAs in cancer was hinted early in the history of miRNA research by 3 important observations[5]: (1) miRNAs are involved in cell proliferation and apoptosis,[6,7] (2) miRNA genes are frequently located at fragile sites and cancer-associated genomic regions (CAGRs),[8] and (3) miRNA expression is deregulated in malignant tumors and tumor cell lines in comparison with normal tissues.[1,9,10] However, the first direct evidence of their oncogenic activity was reported when *MIR15A* and *MIR16-1* were found to be deleted or down-regulated in most chronic lymphocytic leukemia (CLL) patients.[11] Recent analyses of miRNA genes, their targets, processing machinery, genetic polymorphisms, and epigenetic modifications revealed that miRNA-mediated regulation in gene regulatory networks involves a far more complex system than initially expected. We therefore aimed here to present the main miRNA-related discoveries in an online timeline format that will be updated regularly with new discoveries (http://www.integratomics-time.com/miRNA_timeline). A summary of the timeline is presented in Table 1. We also presented an online list of the most extensive review articles sorted according to the topic of miRNA research in cancer; for example, (1) general, (2) polymorphisms, (3) miRNA host genes, (4) transcription factors (TFs), and (5) epigenetics (http://www.integratomics-time.com/miRNA_timeline/reviews). This review and Web-based tool developed should enable efficient examination of past and current miRNA-cancer publications and enable critical exploration of interaction networks involved. Curated and regularly updated information at one site will be useful for researchers and clinicians in guiding their miRNA basic research and therapeutic applications in clinical practice.

INTERPLAY BETWEEN MIRNA AND CANCER GENES

A predominant reason for cancerogenic cell transformation is a combined interaction of both tumor suppressors and oncogenes. MicroRNAs can function as oncogenes, by activating malignant potential, or as tumor suppressors, by blocking the cell's malignant potential.[3,27] Modulation of miRNA biogenesis pathway can also promote tumorigenesis through increased repression of tumor suppressors and/or through incomplete repression of oncogenes (Fig. 1). MicroRNAs can affect all major hallmarks of malignant cells: self-sufficiency in growth signals, evasion of apoptosis, insensitivity to antigrowth signals, sustained angiogenesis, limitless replicative potential, and tissue invasion and metastasis.[28,29] MicroRNA deregulation can therefore contribute to oncogenesis in various ways. Two studies described a more direct relationship between a miRNA cluster *MIR17HG*, *MYC* and cancer oncogenic pathway, revealing a complex genetic circuit that regulates cell proliferation, growth, and apoptosis.[13,14] Overexpression of the *MIR17HG* cluster was found to act with c-*Myc* expression to accelerate tumorigenesis in mice; therefore, this cluster was suggested to be a potential noncoding oncogene.[13] On the other hand, the MIRLET7 family showed tumor suppressor activity by regulating the expression of

From the *Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Domzale; †Department of Biotechnology, National Institute of Chemistry, Ljubljana, Slovenia; ‡Department of Computational and Systems Biology, Rothamsted Research, Harpenden, Herts, UK; and §Department of Experimental Therapeutics and The Center for RNA Interference and Non-Coding RNAs, The University of Texas, MD Anderson Cancer Center, Houston, TX.

The authors have disclosed that they have no significant relationships with, or financial interest in, any commercial companies pertaining to this article.

Reprints: George A. Calin, MD, PhD, Department of Experimental Therapeutics, University of Texas, MD Anderson Cancer Center, So Campus Research Bldg 3, 1881 E Rd, Unit 1950, Houston, TX 77030. E-mail: gcalin@mdanderson.org; or Tanja Kunej, PhD, Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Groblje 3, 1230, Domzale, Slovenia. E-mail: tanja.kunej@bf.uni-lj.si.

ISSN: 1528-9117

TABLE 1. Timeline of Main miRNA Discoveries

Year	Discovery	Species	Associated With	References
1960s	Non–protein-coding transcripts (activator RNAs) regulate gene activity	Eukaryotes		Britten and Davidson (1969)
1993	First miRNA (*lin-4*) discovered	*Caenorhabditis elegans*		Lee et al[7]
2000	Discovered microRNA (*MIRLET7*)	*Caenorhabditis elegans*		Reinhart et al (2000)
2000	RNAi "unit": 21–23 nt	*Drosophila*		Zamore et al (2000)
2001	A large class of small RNAs (named miRNAs) are coexpressed in clusters and have potential regulatory roles	*Caenorhabditis elegans*, invertebrates, vertebrates		Lau et al (2001), Lagos-Quintana et al (2001), and Lee and Ambros (2001)
2001	Dicer in miRNA biogenesis pathway	*Caenorhabditis elegans*, *Drosophila*		Grishok et al (2001) and Hutvágner et al (2001)
2002	miRNAs discovered in plants	Plants		Reinhart et al (2002) and Rhoades et al (2002)
2002	miRNA alterations found in cancer cells (*MIR15A* and *MIR16-1* deleted or down-regulated in most CLLs)	Human	Cancer	Calin et al[11]
2004	More than 50% of miRNA genes are located in CAGRs or fragile sites	Human	Cancer	Calin et al[8]
2004	miRNA as diagnostic/prognostic biomarker	Human	Cancer	Takamizawa et al (2004)
2004	Coexpression of miRNAs and their host genes	Mouse		Rodriguez et al (2004)
2005	miRNA-target interaction relevant to cancer	*Caenorhabditis elegans*, human	Cancer	Johnson et al[12]
2005	Altered expression of miRNAs affects tumor formation/growth *in vivo*	Human	Cancer	He et al[13]
2005	Connection between miRNAs and the MYC oncogene	Human, rat	Cancer	O'Donnell et al[14]
2005	Inhibition of miRNA by antagomirs in mammals	Mouse		Krützfeldt et al (2005)
2006	Molecule of the year (*MIR155* and *MIRLET7A2*)	Human	Cancer	Yanaihara et al (2006)
2006	Epigenetic regulation (DNA methylation and histone deacetylase inhibition) of miRNAs	Human	Cancer	Saito et al[15]
2007	MiRNA target sites can also occur in 5′-UTR	*Caenorhabditis elegans*		Lytle et al[16]
2007	miRNAs deregulation in cancer metastasis	Human	Cancer	Ma et al (2007)
2007	miRNAs can up-regulate mRNA expression and initiate the translation of proteins	Human		Vasudevan et al[17]
2007	miRNAs can affect epigenetic changes and cause the reactivation of silenced tumor suppressor genes	Human	Cancer	Fabbri et al[18]
2007	miRNAs can regulate ncRNAs from the category of long ultraconserved genes (UCGs)	Human	Cancer	Calin et al[19]
2007	miRNAs carrying hexanucleotide terminal motifs are enriched in the nucleus			Hwang et al[20]
2008	miRNA (MIR373) targets promoter sequences and induces gene expression	Human		Place et al[21]
2008	miRNAs can transcriptionally silence gene expression	Human		Kim et al (2008)
2008	Functional SNP in the miRNA seed region	Human	Cancer	Shen et al[22]
2008	miRNA binding sites located within mRNA-coding sequence			Tay et al[23]
2009	Proof of concept of miRNA delivery as cancer therapy	Human, mouse	Cancer	Kota et al[24]
2010	miRNA as molecular decoys	Human, mouse	Cancer	Eiring et al[25]
2010	miRNAs predominantly cause mRNA destabilization	Human, mouse		Guo et al (2010)
2010	Pseudogene PTEN saturates miRNA binding sites	Human	Cancer	Poliseno et al (2010)
2010	Overexpression of a single miRNA is sufficient to cause cancer	Mouse	Cancer	Medina et al (2010)
2011	ceRNA communicate with and regulate other RNA transcripts by competing for shared miRNAs	Human		Salmena et al[26]

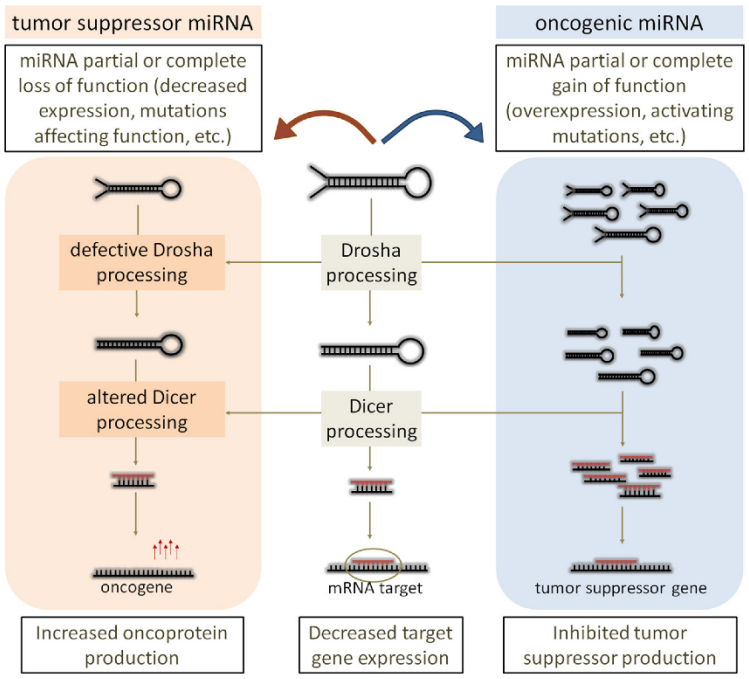


FIGURE 1. Tumorigenesis promoted by modulation of miRNA biogenesis pathway.

a proto-oncogene, the RAS protein, which is a membrane-associated signaling protein that regulates cell growth.[12] Several studies therefore support the view that miRNAs are a class of noncoding nucleic acids that can function as oncogenes or as tumor suppressors contributing to oncogenesis.

It was estimated that about one third of human mRNAs are considered as miRNA targets.[30] Vertebrate miRNAs target about 200 transcripts each, and more than 1 miRNA might coordinately regulate a single target,[31] thereby providing a basis for complex networks. Predicted targets for the differentially expressed miRNAs in human solid tumors have been shown to be significantly enriched for protein-coding tumor suppressors and oncogenes.[32] The standard "dogma," that is, miRNAs target the 3′-UTR of genes and down-regulate the expression of protein-coding genes in cytoplasm, has been expanded with the following additional observations: (1) miRNAs can be localized in the nucleus,[20] (2) miRNAs target other genic regions in addition to 3′-UTR (5′-UTR, promoter regions, coding regions),[16,21,23,33] and (3) miRNA up-regulate translation[17,21] (Fig. 2). It has been shown that human miRNA *MIR369-3* targets AU-rich elements in the target gene *TNF* to activate translation of proteins whose expression they normally repress during cell proliferation.[17] It has also been shown that miRNAs can affect transcription at the promoter level: human *MIR373* binds to the E-cadherin (*CDH1*) promoter and induces transcription.[21] MicroRNA-dependent mRNA repression also occurs through binding sites located in mRNA coding sequences (CDS), as shown for miRNAs with targets in developmentally regulated genes.[23,33] In addition to miRNA-mediated gene silencing through base pairing with mRNA targets, miRNAs also interfere with the function of regulatory proteins (decoy activity).[25] In particular, *MIR328* binds to poly-C binding protein 2 (*PCBP2*), alternatively known as heterogeneous ribonucleoprotein (*hnRNP*) E2, independently of the miRNA's seed region and prevents its interaction with the target mRNA.[25] Down-regulation of *MIR328* in chronic myelogenous leukemia allowed *PCBP2* inhibition of myeloid differentiation and, as a result, led to tumor progression.[25A] Other noncoding RNAs (ncRNAs), such as noncoding ultraconserved genes, have been found to be consistently altered at the genomic level in a high percentage of leukemias and carcinomas and may interact with miRNAs in leukemias.[19] The findings provide support for a model in which both coding and noncoding genes are involved and cooperate in human tumorigenesis. Recently, the role of coding and ncRNAs has been emphasized and grouped in a unifying theory of competing endogenous RNAs (ceRNAs) that can regulate one another through their ability to compete for miRNA binding.[26] The ceRNA hypothesis suggests that long ncRNAs may elicit their biological activity through the ability to act as endogenous decoys for miRNAs and that such activity would in turn affect the distribution of miRNAs on their targets.[26] Competing endogenous RNAs may therefore be active partners in miRNAs regulation, exerting effects on their expression levels, which may have important implications in pathological conditions, such as cancer.

The causes of dysregulated expression can be explained by analyzing the many layers of gene network regulation, including the miRNA gene location in CAGRs, epigenetic mechanisms, and alterations in the miRNA processing machinery[34,35] (Fig. 3). Because each miRNA has numerous targets, inherited minor variations in miRNA expression may have important consequences for the expression of various protein-coding oncogenes and tumor suppressors involved in malignant transformation. Accumulation of additional somatic abnormalities in protein-coding genes or ncRNAs, including miRNAs, is necessary for the full development of the malignant phenotype.[36] The expanding field of miRNA and cancer research therefore requires the

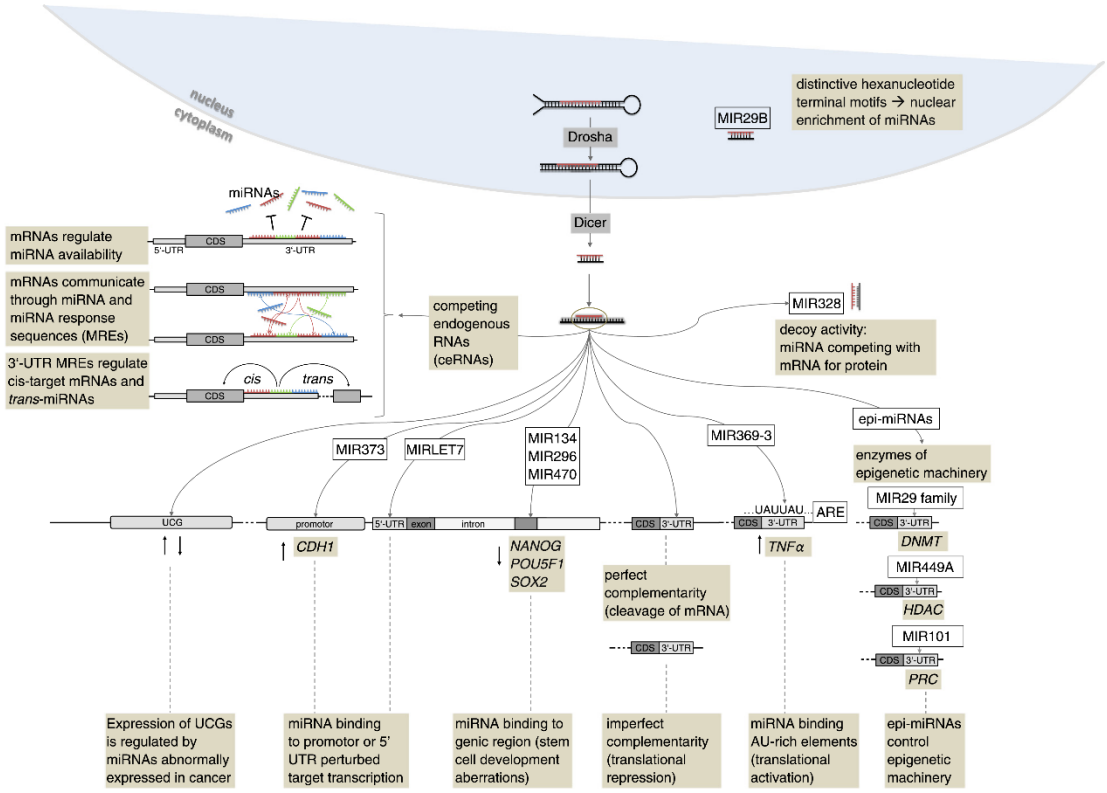


FIGURE 2. Cross talk between miRNAs and targets. ARE indicates AU-rich elements; *CDH1*, E-cadherin; CDS, coding sequence; *DNMT*, DNA methyltransferase; *HDAC*, histone deacetylase; *NANOG*, Nanog homeobox; *POU5F1*, POU class 5 homeobox 1 (synonym: *OCT4*); *PRC*, polycomb repressive complex; *SOX2*, sex-determining region Y box 2; TNFα, tumor necrosis factor α; UCG, ultraconserved gene.

consideration of interplay that connects the regulatory mechanisms and their function into an intricate network.

SINGLE NUCLEOTIDE POLYMORPHISMS IN MIRNA GENES, THEIR TARGETS, AND PROCESSING MACHINERY

Single nucleotide polymorphisms (SNPs) of miRNA precursors, their target sites, and miRNA processing machinery were reported to affect miRNA function and lead to phenotypic effects.[37] When referring to SNPs occurring in miRNA genes, the term miR-SNPs is used, and miR-TS-SNP is used for SNPs located within miRNA target binding sites[38,39] (Fig. 3). Although many miRNA sequence variations observed in cancer alter the secondary structure with no demonstrated effects on miRNA processing,[40] several recent reports show that SNPs located in miRNA genes are associated with cancer susceptibility.[22,39,41] It was observed that miR-SNPs can affect function by modulating the miRNA precursor transcription, processing and maturation,[42] or miRNA-mRNA interaction.[12] Sequence variations in the mature miRNA, especially in the seed region (miR-seed-SNP), may affect miRNA target recognition.[43] Human miRNAs comprising miR-seed-SNPs have been shown to be frequently located within quantitative trait loci, chromosome fragile sites, and cancer susceptibility loci.[43] Because of the miRNA-target interaction, the miR-SNPs (and/or miR-seed-SNPs) and miR-TS-SNPs function in the same manner to create or destroy miRNA binding sites. Chin et al[44] demonstrated a SNP that modified the *MIRLET7* binding site in the v-Ki-*ras* 2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*) and was significantly associated with increased risk for non–small cell lung cancer. Although miR-TS-SNPs were shown to influence susceptibility to tumorigenesis,[45] additional association studies and follow-up functional experiments should still be applied to provide a clearer view on the interplay of these variations in disease development. In order for the miR-TS-SNP to be functional, it must have a proven association with cancer, both miRNA and its predicted target must be expressed in the tissue, and allelic changes must result in different binding affinity of miRNA and affect expression of the target gene.[46] Nevertheless, we can conclude from available published studies that "miR-SNPs" provide an additional layer of functional variability of miRNAs in cancerogenesis and that ongoing studies using next-generation sequencing technology and systems biology analyses are likely to provide additional evidence for miR-SNPs–cancer associations in the near future.

Single nucleotide polymorphisms in miRNA-processing machinery may also have profound effects on the phenotype. Single nucleotide polymorphisms that affect the proteins involved in miRNA biogenesis may have deleterious effects on the miRNAome, and global repression of miRNA maturation was shown to lead to tumorigenesis.[47] Several studies reported

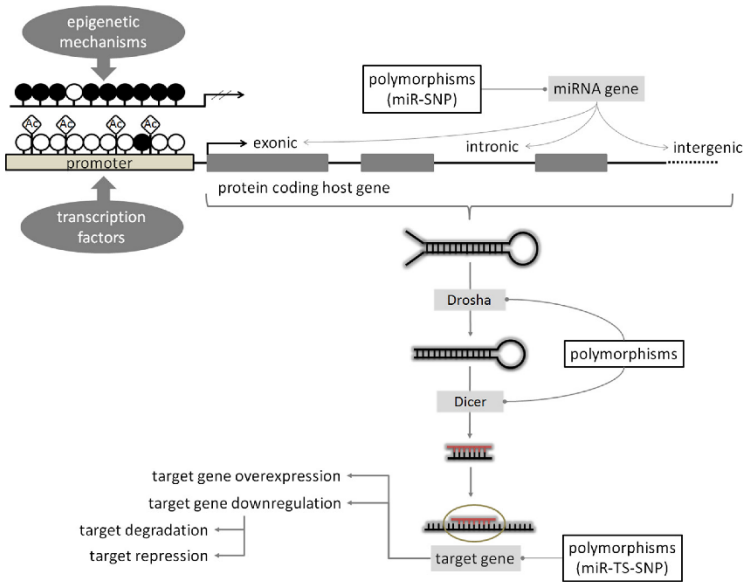


FIGURE 3. MicroRNA biogenesis and mechanisms of regulation. MicroRNA expression and regulation can be affected by transcriptional deregulation, epigenetic modifications (DNA methylation and/or histone acetylation), polymorphisms (SNPs) present in miRNA genes, and their processing machinery and targets. Empty circles indicate unmethylated CpG sites; filled circles, methylated CpG sites. Ac indicates acetyl groups; miR-SNP, SNPs in the miRNA gene; miR-TS-SNP, SNPs in the miRNA target sites.

that genetic polymorphisms of the proteins involved in miRNA machinery affect cancer susceptibility.[48–50] Single nucleotide polymorphisms in the *GEMIN4* gene were significantly associated with altered renal cell carcinoma[48] and bladder cancer risk,[49] and SNP in the 3′ UTR of *DICER1* was associated with an increased risk of premalignant oral lesions in individuals with leukoplakia and/or erythroplakia.[50] The essential part in miRNA variation studies is identification of SNPs located within miRNA genes, their processing machinery, and targets, with bioinformatics tools such as Patrocles,[51] miRNA SNiPer,[43] and PolymiRTS.[52] These tools intercalate and cross-reference the data from dbSNP and, as such, aid in the search for miRNA-related polymorphisms. Although these tools provide useful information on existence of miR-SNPs and their possible effects on target regulation, we still need more experimental data to gain insights on which miR-SNP locations and types of nucleotide substitutions have the most profound effects on cancerogenesis. Such knowledge would have diagnostic power in predicting a person's risk for cancer development based on his/her high-risk miR-SNP genotype. In addition, in people carrying low-risk miR-SNP alleles that have developed primary tumors, genotyping for miR-SNPs of tumor biopsies may reveal somatic mutations that generated high-risk miR-SNPs potentially aiding in the correct diagnosis of the cancer type and in therapeutic decisions.

TRANSCRIPTION FACTOR–MIRNA REGULATORY NETWORK

MicroRNAs are transcriptional and posttranscriptional gene regulators and, like protein-coding genes, are also regulated by TFs, another class of gene regulators that act at the transcriptional level. MicroRNA genes are also linked with TFs in complex regulatory networks where they reciprocally regulate one another[53] (Fig. 4A). It is estimated that up to 43% of human genes are under combined regulation at transcriptional and posttranscriptional level.[54] Owing to the TF's and/or miRNA's involvement in cancer, the disruption of their coregulation may be associated with oncogenesis. O'Donnell et al[14] found that the *MIR17HG* cluster is transactivated by *MYC*, an oncogene frequently dysregulated in cancer. Similarly, *TP53*, a tumor suppressor gene whose pathway mutations have been discovered in many cancer types, was found to regulate expression of *MIR34A*.[55] On the other hand, overexpression of *MIR125B* has

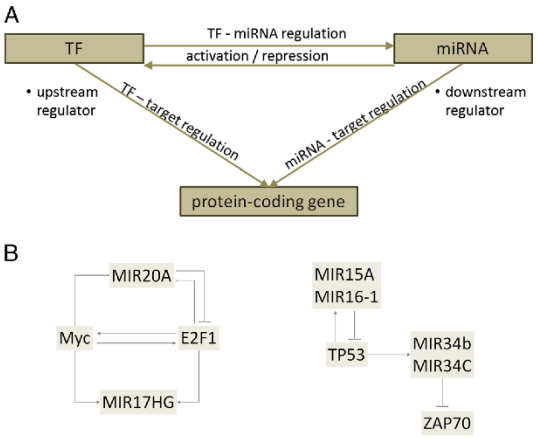


FIGURE 4. Transcription factor–miRNA regulatory network. A, Basic network motif in transcriptional and posttranscriptional gene regulation. B, Schematic representation of examples of regulatory circuit, including MIR17-92 - MYC - E2F1 pathway (left) (adapted from Drakaki and Iliopoulos 2009) and the MIR15A/16-1 cluster - TP53 - MIR34B/34C cluster - ZAP70 pathway (right) (adapted from Fabbri et al. 2011).

been shown to reduce levels of TP53 protein and suppress apoptosis in human cancer cells, whereas knockdown of *MIR125B* elevates the level of TP53.[56]

Coordinated miRNA/TF regulation engage in a wider diversity of biological processes that can have a higher specificity than regulation within only 1 layer of regulation.[57] MicroRNAs and TFs have been found to cooperate in tuning gene expression, by which miRNAs were found to preferentially target genes with transcriptional regulation complexity.[58,59] In regulatory networks, miRNAs and TFs can reciprocally regulate one another and form feedback loops or form feed forward loops in which both TFs and miRNAs regulate their target genes (Fig. 4B). An example of a complex network interaction in cancer was described by O'Donnell et al[28] who found a circuit comprising c-Myc, E2F transcription factor 1 (E2F1), and two miRNAs from the MIR17-92 cluster (MIR20A and MIR17-5p) (Figure 4b). Both miRNAs modulate the translation of E2F1 mRNA, which also induces and is induced by c-Myc. [28] Moreover, *E2F1* binds to the promoter of the *MIR17HG* cluster, activating its transcription.[84] A miRNA/protein feedback circuitry (miRNA/TP53) has been found to be associated with pathogenesis and prognosis of CLL.[60] In 13q deleted CLLs, the *MIR15A/16-1* cluster directly targets TP53 and its downstream effectors. In leukemic cell lines and primary B-CLL cells, TP53 stimulates the transcription of both *MIR15A/16-1* and *MIR34B/34C* clusters, and the MIR34B/34C cluster directly targeted the ZAP70 kinase. This mechanism provides a novel pathogenetic model for the association of 13q deletions with the indolent form of CLL that involves miRNAs, *TP53*, and *ZAP70*.[60] Complex patterns in miRNA-TF interplay have also been computationally analyzed, and the generated databases (TransmiR and dPORE-miRNA) present valuable regulatory framework for future experimental analyses.[61,62] Transcription factor–miRNA regulation, therefore, is not only confined to the "classic" action of TFs on miRNA promoters and their regulation but also extended to reciprocal and mutual interactions forming a complex regulatory network.

MIRNA AND EPIGENETICS IN CANCER

Several aspects of epigenetic regulation have been found to be associated with miRNAs: (1) They regulate target gene expression; regulation of gene expression mediated by miRNAs is frequently reported in cancer and presents one component of an interacting network of epigenetic mechanisms. (2) A subclass of miRNAs (epi-miRNAs) directly controls the epigenetic machinery through a regulatory loop by targeting its regulating enzymes. (3) MiRNA expression could be affected by CpG island hypermethylation-associated silencing in the promoter region or CpG demethylation-associated activation of miRNA promoters (Fig. 5).

Several epi-miRNAs have been shown to be directly connected to the epigenetic machinery by regulating the expression of its regulatory enzymes.[18,63,64] The first epi-miRNAs were identified in lung cancer in a study where *MIR29* family (*MIR29A*, *MIR29B*, and *MIR29C*) was shown to target and downregulate *de novo* DNA methyltransferases (DNMT3A and DNMT3B).[18] In addition, *MIR29B* has been shown to induce global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly *DNMT3A* and *3B* and indirectly the DNMT (*DNMT1*).[65] This led to demethylation of CpG islands in the promoter regions of tumor suppressor genes, allowing their reactivation and a loss of the cell's tumorgenicity.[18] It was also reported that *MIR449A* targets histone deacetylase 1 (*HDAC1*), a gene that is frequently

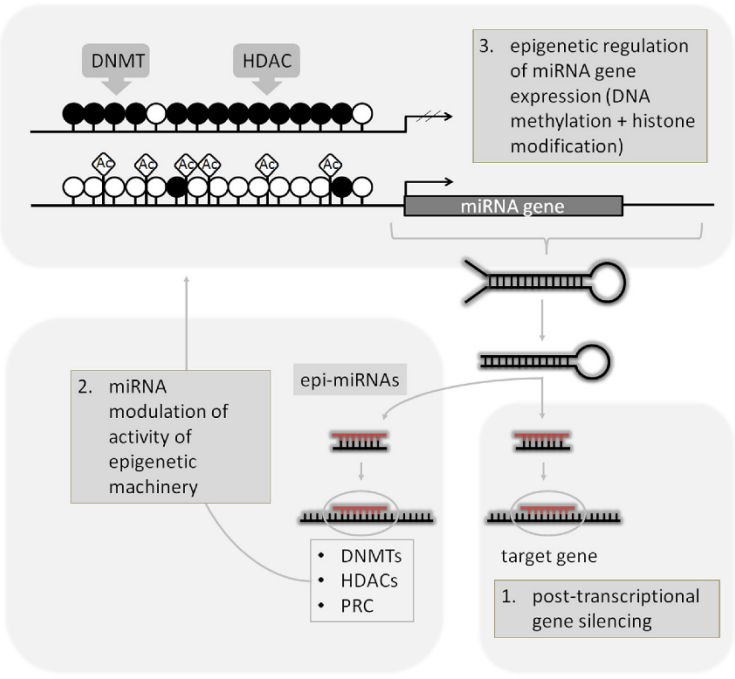


FIGURE 5. Epigenetic concepts of miRNA regulatory network. Empty circles indicate unmethylated CpG sites; filled circles, methylated CpG sites. Ac indicates acetyl groups; DNMTs, DNA methyltransferases; HDACs, histone deacetylases; PRC, polycomb repressive complex.

overexpressed in many types of cancer, and consecutively induces growth arrest in prostate cancer.[63] Also, *MIR101* was shown to directly modulate the expression of enhancer of zeste homolog 2 (EZH2), a catalytic subunit of the polycomb repressive complex 2 (PRC2), which mediates epigenetic silencing of tumor suppressor genes in cancer.[64]

The expression of miRNA genes has also been found to be silenced in human tumors by epigenetic mechanisms, such as aberrant hypermethylation of CpG islands encompassing and/or in proximity of miRNA genes, or by histone acetylation.[66] The first evidence that altered methylation status can deregulate the expression of a miRNA in cancer was reported by Saito et al.[15] *MIR127*, embedded within a CpG island promoter, was silenced in several cancer cells but strongly up-regulated after treatment with a hypomethylating agent (DNMT inhibitor). A similar scenario was observed with *MIR124A* whose function can be restored by erasing DNA methylation and has functional consequences on cyclin D kinase 6 (CDK6) activity.[67] On the other hand, Brueckner et al[68] observed that hypomethylation of *MIRLET7A3* facilitates reactivation of the gene and elevates expression of *MIRLET7A3* in human lung cancer cell lines, which resulted in enhanced tumor phenotypes. Compared with protein-coding genes, human oncomiRs were found to have an order of magnitude higher methylation frequency.[66,69] Future studies of epigenetic regulation of miRNA expression coupled to downstream signaling pathways are likely to lead to the development of novel drug targets in cancer therapy.[69]

MIRNA EXPRESSION PROFILES/SIGNATURES

Profiling of miRNAs is used to document their expression variability, and it was shown to be more accurate for cancer classification than by using sets of known protein-coding genes.[36,70] In cancer, the loss of tumor suppressor miRNAs enhances the expression of target oncogenes, whereas increased expression of oncogenic miRNAs can repress target tumor suppressor genes. Paired expression profiles of miRNAs and mRNAs can be used to identify functional miRNA-target relationships with high precision.[71] The aberrant expression of miRNAs in cancer is characterized by abnormal levels of expression for mature and/or precursor miRNA transcripts in comparison to those in the corresponding normal tissue. Lu et al[70] observed a general down-regulation of miRNAs in tumor samples compared with normal tissue samples. It was also found that miRNA expression profiles could differentiate human cancers according to their developmental origin, with cancers of epithelial and hematopoietic origin having distinct miRNA profiles.[70] The first evidence that miRNA expression could be altered in cancer came from the observation by Calin et al[11] that *MIR15A/MIR16-1* gene cluster is located in a genomic region frequently deleted in CLL and that their expression is frequently down-regulated or deleted in CLL. Afterward, numerous studies examined aberrant miRNA expression signatures in cancer. A review analyzing 58 studies[72] revealed 70 differentially expressed miRNAs in cancers, which were reported in at least 5 studies. The causes of widespread alterations of miRNA expression in cancer cells include different factors such as the location of miRNAs at CAGR,[8] epigenetic regulation,[73] and abnormalities in genes and proteins of the miRNA processing machinery.[74] Because deregulated miRNA expression is an early event in tumorigenesis, measuring the levels of circulating miRNAs may also be useful for early cancer detection. Fluid-expressed miRNAs have been discussed as reliable cancer biomarkers and treatment response predictors as well as potential new patient selection criteria for clinical trials.[75] Profiling of miRNA expression correlates with clinical and biological characteristics of tumors and has enabled the identification of signatures associated with diagnosis, staging, progression, prognosis, and response to treatment of human tumors.[76] MicroRNA fingerprinting therefore represents an additional tool in the clinical oncology.

MIRNAS AS DIAGNOSTIC, PROGNOSTIC, AND THERAPEUTIC TARGETS IN CANCER

On the basis of their regulatory function, miRNAs are important players in the oncogenic signaling pathway, which is why they should be considered in cancer diagnosis and prognosis. As already mentioned, miRNA expression profiles contain much information that could explain developmental processes in cancer and are disrupted by different mechanisms mentioned in the previous section on miRNA expression profiles/signatures.[36] A general down-regulation of miRNAs was observed in tumor samples compared with normal tissue samples. A unique miRNA signature is associated with prognostic factors and disease progression in CLL. Mutations in miRNA transcripts are common and may have functional importance.[77]

Several studies have indicated various strategies for therapeutic usage of miRNAs in cancer. Four different strategies for potential therapies were proposed[78]: (1) Anti-miRNA oligonucleotides—inhibitory molecules that block the interactions between miRNA and its target mRNA by competition.[79] (2) MicroRNA sponges—synthetic miRNAs that contain multiple binding sites for an endogenous miRNA and prevent the interaction between miRNA and its endogenous targets.[80] Ebert et al[80] have also designed these sponges with complementary seed regions, which effectively repress an entire miRNA seed family. (3) MiRNA masking—a sequence with perfect complementarity to the binding site for an endogenous miRNA in the target gene, which can in turn form a duplex with the target mRNA with higher affinity and block the access of the miRNA.[81] (4) Small molecule inhibitors against specific miRNAs—chemicals or reagents able to specifically inhibit miRNA synthesis. One such example is azobenzene, a specific and efficient inhibitor of *MIR21* biogenesis.[82] Apart from the 4 aforementioned therapeutic approaches proposed by Li et al,[78] others have proposed and demonstrated alternative strategies such as restoring activity of tumor-suppressive miRNAs to rescue its anti-tumor function.[83–85] Some studies also suggest that, by selecting miRNAs that are highly expressed in normal tissues but lost in cancer cells, they can be used in the general strategy for restoring tumor suppressor miRNAs as therapy.[24,86,87] Another approach to use miRNA in cancer therapy is in sensitizing tumors to chemotherapy.[88] Owing to the ability of miRNAs to target signaling pathways that are frequently misregulated in cancers, studies have examined the potential of miRNAs or antagomirs to sensitize resistant cells to already known and successful cancer therapies (eg, tamoxine, gefitinib treatments). Several promising *in vitro* and mouse model studies have already shown efficacy of this approach guiding now the clinical development of miRNA-based therapies for sensitization to chemotherapy.

However, despite the advances made in the miRNA-mediated therapy, 2 major hurdles still remain: the first is to maintain target specificity, which is especially challenging and its effect needs to be evaluated on a proteome-wide scale to prevent unwanted gene alterations, owing to partial complementary binding between miRNAs and protein-coding transcripts.[78] The second hurdle is to achieve high therapeutic efficiency, which is linked with delivery efficiency. In addition to lipid- and polymer-based nanoparticles for systemic delivery, viral vectors may be also

used; these approaches were found suitable for certain types of tumors, and further investigations are needed for the evaluation of these approaches in various tumor types.[78]

CONCLUSIONS AND FUTURE DIRECTIONS

The causes of differential expression of miRNA genes in tumorous tissues can be better understood by including multiple layers, such as their location in CAGRs, epigenetic mechanisms, and alterations in the miRNA processing machinery into a coordinately regulated network. The expanding field of miRNA and cancer research therefore requires the consideration of such an interplay that connects the regulatory mechanisms and their function into an intricate network. Future studies will further clarify complex interactions of short and long ncRNAs with protein coding genes and their involvement in shaping phenotypes and cancer development.

REFERENCES

1. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;6:857–866.
2. Kanellopoulou C, Monticelli S. A role for microRNAs in the development of the immune system and in the pathogenesis of cancer. *Semin Cancer Biol*. 2008;18:79–88.
3. Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. *Nat Rev Cancer*. 2006;6:259–269.
4. Liang Y, Ridzon D, Wong L, et al. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*. 2007;8:166.
5. Sassen S, Miska EA, Caldas C. MicroRNA: implications for cancer. *Virchows Arch*. 2008;452:1–10.
6. Brennecke J, Hipfner DR, Stark A, et al. *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell*. 2003;113:25–36.
7. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. 1993;75:843–854.
8. Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A*. 2004;101:2999–3004.
9. Gaur A, Jewell DA, Liang Y, et al. Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Res*. 2007;67:2456–2468.
10. Lu C, Tej SS, Luo S, et al. Elucidation of the small RNA component of the transcriptome. *Science*. 2005;309:1567–1569.
11. Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro- RNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2002;99:15524–15529.
12. Johnson SM, Grosshans H, Shingara J, et al. RAS is regulated by the let-7 microRNA family. *Cell*. 2005;120:635–647.
13. He L, Thomson JM, Hemann MT, et al. A microRNA polycistron as a potential human oncogene. *Nature*. 2005;435:828–833.
14. O'Donnell KA, Wentzel EA, Zeller KI, et al. c-Myc–regulated microRNAs modulate E2F1 expression. *Nature*. 2005;435:839–843.
15. Saito Y, Liang G, Egger G, et al. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell*. 2006;9:435–443.
16. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc Natl Acad Sci U S A*. 2007;104:9667–9672.
17. Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. *Science*. 2007;318:1931–1934.
18. Fabbri M, Garzon R, Cimmino A, et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc Natl Acad Sci U S A*. 2007;104:15805–15810.
19. Calin GA, Liu CG, Ferracin M, et al. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell*. 2007;12:215–229.
20. Hwang HW, Wentzel EA, Mendell JT. A hexanucleotide element directs microRNA nuclear import. *Science*. 2007;315:97–100.
21. Place RF, Li LC, Pookot D, et al. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A*. 2008;105:1608–1613.
22. Shen J, Ambrosone CB, DiCioccio RA, et al. A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. *Carcinogenesis*. 2008;29:1963–1966.
23. Tay Y, Zhang J, Thomson AM, et al. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*. 2008;455:1124–1128.
24. Kota J, Chivukula RR, O'Donnell KA, et al. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell*. 2009;137:1005–1017.
25. Eiring AM., Harb JG, Neviani P, et al. miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. *Cell*. 2010;140:652–665.

25A. Lopez-Serra P, Esteller M. DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. *Oncogene*. 2012;31(13):1609–1622.

26. Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011;146:353–358.
27. Zhang B, Pan X, Cobb GP, et al. microRNAs as oncogenes and tumor suppressors. *Dev Biol*. 2007;302:1–12.
28. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100:57–70.
29. Negrini M, Nicoloso MS, Calin GA. MicroRNAs and cancer—new paradigms in molecular oncology. *Curr Opin Cell Biol*. 2009;21:470–479.
30. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120:15–20.
31. Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet*. 2005;37:495–500.
32. Volinia S, Calin GA, Liu CG, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A*. 2006;103:2257–2261.
33. Forman JJ, Legesse-Miller A, Coller HA. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A*. 2008;105:14879–14884.
34. Calin GA, Croce CM. Chromosomal rearrangements and microRNAs: a new cancer link with clinical implications. *J Clin Invest*. 2007;117:2059–2066.
35. Calin GA, Croce CM. Chronic lymphocytic leukemia: interplay between noncoding RNAs and protein-coding genes. *Blood*. 2009;114:4761–4770.
36. Calin GA, Croce CM. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res*. 2006;66:7390–7394.
37. Georges M, Coppieters W, Charlier C. Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. *Curr Opin Genet Dev*. 2007;17:166–176.
38. Sun G, Yan J, Noltner K, et al. SNPs in human miRNA genes affect biogenesis and function. *RNA*. 2009;15:1640–1651.
39. Tian T, Shu Y, Chen J, et al. A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese. *Cancer Epidemiol Biomarkers Prev*. 2009;18:1183–1187.
40. Diederichs S, Haber DA. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. *Cancer Res*. 2006;66:6097–6104.
41. Hu Z, Liang J, Wang Z, et al. Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. *Hum Mutat*. 2009;30:79–84.
42. Zeng Y, Cullen BR. Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem*. 2005;280:27595–27603.
43. Zorc M, Jevsinek Skok D, Godnic I, et al. Catalog of microRNA seed polymorphisms in vertebrates. *PLoS One*. 2012;7:e30737.

44. Chin LJ, Ratner E, Leng S, et al. A SNP in a let-7 microRNA complementary site in the KRAS 3′ untranslated region increases non–small cell lung cancer risk. *Cancer Res*. 2008;68:8535–8540.
45. Nicoloso MS, Sun H, Spizzo R, et al. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res*. 2010;70:2789–2798.
46. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer*. 2010;10:389–402.
47. Kumar MS, Lu J, Mercer KL, et al. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet*. 2007;39:673–677.
48. Horikawa Y, Wood CG, Yang H, et al. Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma. *Clin Cancer Res*. 2008;14:7956–7962.
49. Yang H, Dinney CP, Ye Y, et al. Evaluation of genetic variants in microRNA-related genes and risk of bladder cancer. *Cancer Res*. 2008;68:2530–2537.
50. Clague J, Lippman SM, Yang H, et al. Genetic variation in MicroRNA genes and risk of oral premalignant lesions. *Mol Carcinog*. 2010;49:183–189.
51. Hiard S, Charlier C, Coppieters W, et al. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Res*. 2010;38[suppl 1]:D640–D651.
52. Ziebarth JD, Bhattacharya A, Chen A, et al. PolymiRTS Database 2.0: linking polymorphisms in microRNA target sites with human diseases and complex traits. *Nucleic Acids Res*. 2011;40:216–221.
53. Yu XP, Lin J, Zack DJ, et al. Analysis of regulatory network topology reveals functionally distinct classes of microRNAs. *Nucleic Acids Res*. 2008;36:6494–6503.
54. Shalgi R, Lieber D, Oren M, et al. Global and local architecture of the mammalian microRNA–transcription factor regulatory network. *PLoS Comput Biol*. 2007;3:1291–1304.
55. Tarasov V, Jung P, Verdoodt B, et al. Differential regulation of microRNAs by p53 revealed by massively parallel sequencing—miR-34a is a p53 target that induces apoptosis and G(1)-arrest. *Cell Cycle*. 2007;6:1586–1593.
56. Le MTN, The C, Shyh-Chang N, et al. MicroRNA-125b is a novel negative regulator of p53. *Genes Dev*. 2009;23:862–876.
57. Chen CY, Chen ST, Fuh CS, et al. Coregulation of transcription factors and microRNAs in human transcriptional regulatory network. *BMC Bioinformatics*. 2011;12[suppl 1]:S41.
58. Enright AJ, John B, Gaul U, et al. MicroRNA targets in *Drosophila*. *Genome Biol*. 2003;5:R1.
59. Cui Q, Yu Z, Pan Y, et al. MicroRNAs preferentially target the genes with high transcriptional regulation complexity. *Biochem Biophys Res Commun*. 2007;352:733–738.
60. Fabbri M, Bottoni A, Shimizu M, et al. Association of a microRNA/TP53 feedback circuitry with pathogenesis and outcome of B-cell chronic lymphocytic leukemia. *JAMA*. 2011;305:59–67.
61. Wang J, Lu M, Qiu C, et al. TransmiR: a transcription factor–microRNA regulation database. *Nucleic Acids Res*. 2010;38[database issue]:D119–D122.
62. Schmeier S, Schaefer U, MacPherson CR, et al. dPORE-miRNA: polymorphic regulation of microRNA genes. *PLoS One*. 2011;6:e16657.
63. Noonan EJ, Place RF, Pookot D, et al. miR-449a targets HDAC-1 and induces growth arrest in prostate cancer. *Oncogene*. 2009;28:1714–1724.
64. Friedman JM, Liang G, Liu CC, et al. The putative tumor suppressor microRNA-101 modulates the cancer epigenome by repressing the polycomb group protein EZH2. *Cancer Res*. 2009;69:2623–2629.
65. Garzon R, Liu SJ, Fabbri M, et al. MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. *Blood*. 2009;113:6411–6418.
66. Weber B, Stresemann C, Brueckner B, et al. Methylation of human microRNA genes in normal and neoplastic cells. *Cell Cycle*. 2007;6:1001–1005.
67. Lujambio A, Ropero S, Ballestar E, et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res*. 2007;67:1424–1429.
68. Brueckner B, Stresemann C, Kuner R, et al. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res*. 2007;67:1419–1423.
69. Kunej T, Godnic I, Ferdin J, et al. Epigenetic regulation of microRNAs in cancer: an integrated review of literature. *Mutat Res*. 2011;717:77–84.
70. Lu J, Getz G, Miska EA, et al. Using expression profiling data to identify human microRNA targets. *Nat Methods*. 2007;4:1045–1049.
71. Huang JC, Babak T, Corson TW, et al. Using expression profiling data to identify human microRNA targets. *Nat Methods*. 2007;4:1045–1049.
72. Ferdin J, Kunej T, Calin GA. Non-coding RNAs: identification of cancer-associated microRNAs by gene profiling. *Technol Cancer Res Treat*. 2010;9:123–138.
73. Esteller M. Epigenetics in cancer. *N Engl J Med*. 2008;358:1148–1159.
74. Melo SA, Esteller M. A precursor microRNA in a cancer cell nucleus: get me out of here! *Cell Cycle*. 2011;10:922–925.
75. Cortez MA, Bueso-Ramos C, Ferdin J, et al. MicroRNAs in body fluids—the mix of hormones and biomarkers. *Nat Rev Clin Oncol*. 2011;8:467–477.
76. Barbarotto E, Schmittgen TD, Calin GA. MicroRNAs and cancer: profile, profile, profile. *Int J Cancer*. 2008;122:969–977.
77. Calin GA, Ferracin M, Cimmino A, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med*. 2005;353:1793–1801.
78. Li C, Feng Y, Coukos G, et al. Therapeutic microRNA strategies in human cancer. *AAPS J* 2009;11:747–757.
79. Weiler J, Hunziker J, Hall J. Anti-miRNA oligonucleotides (AMOs): ammunition to target miRNAs implicated in human disease? *Gene Ther*. 2006;13:496–502.
80. Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods*. 2007;4:721–726.
81. Xiao J, Yang B, Lin H, et al. Novel approaches for gene-specific interference via manipulating actions of microRNAs: examination on the pacemaker channel genes *HCN2* and *HCN4*. *J Cell Physiol*. 2007;212:285–292.
82. Gumireddy K, Young DD, Xiong X, et al. Small-molecule inhibitors of microrna miR-21 function. *Angew Chem Int Ed Engl*. 2008;47:7482–7484.
83. Kumar MS, Erkeland SJ, Pester RE, et al. Suppression of non–small cell lung tumor development by the let-7 microRNA family. *Proc Natl Acad Sci U S A*. 2008;105:3903–3908.
84. Johnson CD, Esquela-Kerscher A, Stefani G, et al. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res*. 2007;67:7713–7722.

84A. Drakaki A, Iliopoulos D. MicroRNA Gene Networks in Oncogenesis. *Curr Genomics*. 2009;10:35–41.

85. Esquela-Kerscher A, Trang P, Wiggins JF, et al. The let-7 microRNA reduces tumor growth in mouse models of lung cancer. *Cell Cycle*. 2008:7:759–764.
86. Cimmino A, Calin GA, Fabbri M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A*. 2005;102:13944–13949.
87. Bonci D, Coppola V, Musumeci M, et al. The miR-15a–miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. *Nat Med*. 2008;14:1271–1277.
88. Kasinski AL, Slack FJ. Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat Rev Cancer*. 2011;11:849–864.

Priloga G

Genome-wide *in silico* screening (GWISS) for microRNA genetic variability in livestock species (Jevšinek Skok in sod., 2013)

Genome-wide *in silico* screening (GWISS) for microRNA genetic variability in livestock species

D Jevsinek Skok[1*], I Godnic[1*], M Zorc[1], S Horvat[1,2], P Dovc[1], M Kovac[1], and T Kunej[1]

[1]Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Groblje 3, 1230, Domzale, Slovenia

[2]National Institute of Chemistry, Hajdrihova 19, SI-1001 Ljubljana, Slovenia

* Authors contributed equally to the work.

E-mail addresses:
Dasa Jevsinek Skok: dasa.jevsinek.skok@bf.uni-lj.si
Irena Godnic: irena.godnic@gmail.com
Minja Zorc: minja.zorc@gmail.com
Simon Horvat: simon.horvat@bf.uni-lj.si
Peter Dovc: peter.dovc@bf.uni-lj.si
Milena Kovac: milena.kovac@bf.uni-lj.si
Tanja Kunej: tanja.kunej@bf.uni-lj.si

Corresponding author:
Tanja Kunej, PhD
Department of Animal Science
Biotechnical Faculty
University of Ljubljana
Groblje 3, Slovenia
Phone: +3861-320-3890
Fax: +3861 7217 888
Email: tanja.kunej@bf.uni-lj.si

Keywords: microRNA (miRNA), single nucleotide polymorphism (SNP), double nucleotide polymorphism (DNP), seed region, livestock species, cattle, chicken, prioritization

Abstract

MicroRNAs are a class of non-coding RNAs that post-transcriptionally regulate target gene expression. Previous studies have shown that microRNA gene variability can interfere with its function resulting in phenotypic variation. Polymorphisms within microRNA genes present a source of novel biomarkers for phenotypic traits in animal breeding. However, little is known about microRNA genetic variability in livestock species, which is also due to incomplete data in genomic resource databases. Therefore, the aim of this study was to perform a genome-wide *in silico* screening of genomic sources and determine the genetic variability of microRNA genes in livestock species using miRNA SNiPer 3.0 (http://www.integratomics-time.com/miRNA-SNiPer/), a new version of our previously developed tool. By examining Ensembl and miRBase genome builds it was possible to design a tool-based generated search of 16 genomes including four livestock species: pig, horse, cattle, and chicken. The analysis revealed 65 polymorphisms located within mature microRNA regions in these four species, including 28 present within the seed region in cattle and chicken. Polymorphic microRNA genes in cattle and chicken were further examined for mapping to quantitative trait loci regions associated with production and health traits. The developed bioinformatics tool enables analysis of polymorphic microRNA genes and prioritization of potential regulatory polymorphisms, and therefore contributes to the development of microRNA-based biomarkers in livestock species. The assembled catalog and the developed tool can serve animal science community to efficiently select miRNA SNPs for further quantitative and molecular genetic evaluations of their phenotypic effects and causal associations with livestock production traits.

Introduction

MicroRNAs (miRNAs) are non-coding RNA (ncRNA) molecules, approximately 21 nucleotides in length, and important post-transcriptional regulators of target mRNAs. By binding to the different target gene regions, *i.e.*, 3'-untranslated region (3'-UTR), 5'-UTR, promoter, or coding sequences, they repress or activate translation (reviewed in (Kunej *et al.* 2012)). The biogenesis of miRNAs begins in the nucleus with the primary transcript (pri-miRNA), which are encoded either in the intergenic regions or within overlapping genes (protein-coding or non-coding) (Godnic *et al.* 2013) . Several hundreds or thousands base pairs in length, the pri-miRNA is cleaved by endonuclease Drosha to 60 to 70 nucleotides long precursor miRNA (pre-miRNA). Precursor miRNA, with its characteristic stem-loop structure (**Figure 1**) is transported to the cytoplasm where endonuclease Dicer cleaves both duplex chains to form mature miRNAs, either in the 5' arm (*e.g.* miR-1-5p) or on the 3' arm (*e.g.* miR-1-3p) (Lee *et al.* 2002; Bartel 2004). The key binding location for translational suppression is called the seed region and resides within the mature miRNA sequence. It was shown that the seed region is flexible within the miRNA, allowing the mRNA to be regulated following base pairing of nucleotides positioned either on 2-7, 2-8, or 3-9 from the 5' end of the miRNA (Lewis *et al.* 2005; Nahvi *et al.* 2009; Sun *et al.* 2009). Additionally, it was shown that nearly perfect complementarity between the first nine miRNA nucleotides is needed in order for the protein-coding genes to be regulated (Kiriakidou *et al.* 2004).

About half of miRNA genes are located within host genes and miRNA-host gene pairs have frequently been reported to be co-expressed in human and mouse (Godnic *et al.* 2013). Aberrant miRNA expression signatures have previously been linked with several diseases in human (reviewed in (Ferdin *et al.* 2010; Kunej *et al.* 2011)). Similarly, transcriptome profiling studies in livestock revealed that expression of many miRNAs are species- and tissue-specific and suggested their potential important roles in tissue and organ development, immune response, and metabolism (reviewed in (Liu *et al.* 2010)). Moreover, single nucleotide polymorphisms (SNPs) located within miRNA genes (miR-SNPs), miRNA target sites (miR-TS-SNPs) or in protein-coding genes involved in miRNA biogenesis result in phenotypic differences and therefore affect production traits and susceptibility to diseases (Georges *et al.* 2007). However, because a miRNA is estimated to target 200 transcripts on average (Krek *et al.* 2005), miR-SNPs are expected to have more profound biological impact than those residing within miRNA target sites (Sun *et al.* 2009). A miR-TS-SNP can be prone to either disrupt or create new interacting sites for miRNAs. For example, it has been shown, that a mutation creating a target site for mir-1 and mir-206 within the myostatin gene (*MSTN*; previous symbol *GDF8*) affects muscular hypertrophy in Texel sheep (Clop *et al.* 2006). Cargill et al. (Cargill *et al.* 2008) reported that polymorphism residing within the 3'-UTR of synaptojanin 1 gene (*SYNJ1*) was found to disrupt target sites of let-7a and miR-98, suggesting an effect on polledness (absence of

horns) in cattle. On the other hand, a SNP occurring within a miRNA gene, especially in the seed region can alter its secondary structure, affect miRNA processing, and have an influence on the phenotype. For example, a SNP residing within the seed region (miR-seed-SNP) of murine *mmu-mir-717*, a body mass and growth-associated protein-coding host gene glypican (*Gpc3*) and quantitative trait locus (QTL), affected fat deposition (Kunej *et al.* 2010). Moreover, a polymorphism residing within the miRNA seed region (miR-seed-SNP) (rs14934924) located within *gga-mir-1657* was found to be associated with growth and meat quality traits in chicken (Li *et al.* 2012). Twelve polymorphisms within the porcine miR-206/miR-133b cluster have also been found to affect muscle and meat quality traits (Lee *et al.* 2013).

Over 20 bioinformatics tools have been developed for the search of polymorphic miRNA targets using different algorithms, however only four enabling identification of polymorphisms residing within miRNA genes (**Table S1)**. Patrocles was the first tool for the search of polymorphisms within miRNA genes, as well as their targets, and genes encoding the components of the silencing machinery (Hiard *et al.* 2010). BioMart is a data mining tool which also enables the search of polymorphisms within mature miRNA regions (Kinsella *et al.* 2011), whereas miRNASNP searches within miRNA seed and precursor regions (Gong *et al.* 2012). Because of the need for a regularly updated bioinformatics tool for identification of SNPs within all miRNA gene regions (pre-miRNA, mature, and seed) and species with available genomic data (annotated miRNAs and genetic variability data), we developed miRNA SNiPer tool (Zorc *et al.* 2012). To assure data accuracy, we manually inspected the tool-based catalog of polymorphic miRNA genes, supplementing it with information regarding overlapping genes and QTL, as well as validation status of polymorphisms within miRNA genes. Even though several analyses investigating the phenotypic effect of polymorphisms have been performed in human and model organisms, little is known about genetic variability of miRNA genes in livestock species, which is also due to still ongoing sequencing projects in genomic resource databases.

The focus of our previous study Zorc *et al*. (Zorc *et al.* 2012) was genetic variability of miRNA seed regions in 13 vertebrate species, whereas the aim in the present study was to perform a genome-wide *in silico* screening (GWISS) in livestock species to generate a catalog of polymorphisms located within miRNA mature, including its seed region. We therefore updated miRNA SNiPer, a bioinformatics web-based tool for identification of polymorphisms residing within miRNA genes, with the latest matching releases of source databases. Based on the current genomic database releases, the tool for automated search of miRNA polymorphisms enables search for 16 species, including four livestock species. Additionally, a QTL overlap analysis was performed and three miRNA genes were selected for experimental validation of their miR-seed-SNPs, including a double nucleotide polymorphism (DNP).

Materials and methods

Update of the miRNA SNiPer tool.

The online tool miRNA SNiPer 1.0, previously developed for the search of polymorphisms within the precursor, mature, and seed miRNA regions in 13 vertebrate species was updated to release miRNA SNiPer 3.0, using the latest matching versions of source databases (**Table S2**). The differences between the tool versions are presented online http://www.integratomics-time.com/miRNA-SNiPer/about.php. 1) Ensembl Variation database was used to retrieve data for polymorphisms from releases 66 and 68 (http://www.ensembl.org/index.html). 2) miRBase, releases 18 and 19 (http://www.mirbase.org/) (Kozomara & Griffiths-Jones 2011), was used to retrieve the location of miRNA genes. The seed region was defined according to the 7mer used in TargetScan (release 6.2; http://www.targetscan.org/) (Lewis *et al.* 2005), the area of 2-8 nucleotides from the 5' end in the mature miRNA region.

Physical and functional characterization of miRNA SNPs in cattle and chicken

Genomic distribution of miR-seed polymorphisms was presented on a genomic view using Flash GViewer web tool (http://gmod.org/wiki/Flashgviewer/). To perform a genomic overlap analysis QTL in chicken and cattle were downloaded from Animal QTL Database, release 19 (http://www.animalgenome.org/cgi-bin/QTLdb/index/) (Hu *et al.* 2013). TargetScan Custom feature (http://www.targetscan.org/) was used to analyze whether the miR-seed-SNP cause the formation of seed regions annotated to different miRNA (Lewis *et al.* 2005). The data of miRNA gene variability was extracted from the following DNA-microarrays: Illumina 60K SNP chip for chicken, PorcineSNP60, and 770K Illumina HD SNPs for cattle.

Samples and sequencing of a DNA panel comprising seed polymorphisms in cattle and chicken

The bovine samples were obtained from Holstein, Slovenian Simmental and Slovenian autochthonous cattle breed Cika. DNA samples were extracted using DNeasy Blood & Tissue DNA extraction kit (Qiagen, Düseldorf, Germany). Chicken DNA samples of three local breeds (brown, barred and Styrian hen) and German commercial hybrid LB (Lohmann brown classic), were extracted using a standard phenol-chloroform extraction protocol.

Six polymorphisms residing within the miRNA seed region of bovine *bta-mir-2489* and *bta-mir-2313*, and chicken *gga-mir-1658* were selected for experimental validation. The polymerase chain reaction (PCR) primers were selected using Primer 3 tool (http://frodo.wi.mit.edu/):

bta-mir-2489-F: 5’-TGCTTCTGGCCAATAAATCC-3’,
bta-mir-2489-R: 5’-CCTCAGAGGATGCCAGTGAT-3’,
bta-mir-2313-F: 5'-GCACAGACTCTCAGCCACTG-3',
bta-mir-2313-R: 5'-CTGACTGAGGCTCTCGCTCT-3',
gga-mir-1658-F, 5’-GGCCTCACAGCAGGATTTAC-3’, and
gga-mir-1658-R: 5’- GCATGCAATTTCCAAGTGTG-3’.

Conditions for the PCR were: 94°C for 10 min, 30 cycles of 94°C for 1 min, 63°C for 1 min, 72°C for 1 min followed by a further 10 min extension at 72°C. The PCR products were purified using Exonuclease I (*Exo*I) and shrimp alkaline phosphatase (SAP) (both Fermentas, Vilnius, Lithuania), followed by the sequencing reaction for capillary electrophoresis on ABI3130xl (Applied Biosystems, USA).

Results and discussion

Using the upgraded version of the bioinformatics tool miRNA SNiPer 3.0 we generated a catalog of miRNA polymorphisms in four livestock species supplemented with information regarding overlaps with genes (protein-coding or non-coding) and QTL, and validation status of the polymorphisms. Six miR-seed polymorphisms located within two cattle and one chicken miRNA genes were experimentally validated and interesting hotspots for future functional analysis were revealed.

Update of bioinformatics tool to miRNA SNiPer 3.0

We released miRNA SNiPer 3.0 (http://www.integratomics-time.com/miRNA-SNiPer/), a new version of the previously developed tool for search of polymorphisms in pre-miRNA, seed and mature region, in 16 species: human, chimpanzee, orangutan, macaque, cow, horse, pig, opossum, platypus, rat, mouse, chicken, zebra finch, tetraodon, zebrafish, and fruit fly. The tool searches for polymorphisms by retrieving data from matching releases of genomic databases. Releases included in the tool have to contain matching information regarding the location of miRNA genes and polymorphisms. Compared to the previous version of the online tool miRNA SNiPer 1.0, the present version of miRNA SNiPer 3.0 is updated with the latest matching releases of its source databases: miRBase (miRNA location data) and Ensembl (genetic variability data), as shown online (http://www.integratomics-time.com/miRNA-SNiPer/about.php). Additionally, in the new version four genomes were added, including cattle.

Using the tool we performed a genome-wide *in silico* screening (GWISS) of genomic resources and generated a catalog of miRNA polymorphisms in livestock species. By examining the latest matching Ensembl and miRBase genome builds it was possible to design a tool-based generated search for four species: cattle, chicken, horse, and pig (**Table S2**). Six other animal species (goat, sheep, rabbit, duck, turkey, and honeybee) could not be included in this version of the tool due to their currently mismatching or missing genome build data. The miRNA SNiPer 3.0 accepts a list of miRNA genes and returns a table of variations located within different regions of miRNA genes: pre-miRNA, mature, and seed region (**Figure 2**).

The developed bioinformatics tool enables analysis of polymorphic miRNA genes and prioritization of potential regulatory polymorphisms, and therefore contributes to the development of miRNA-based biomarkers for production traits in livestock species. Upgrades of source databases will enable the future miRNA SNiPer releases to search for miRNA gene polymorphisms in other species.

Catalog of miRNA polymorphisms in livestock species

The catalog of miRNA polymorphisms in livestock species presents a list of genetic variations residing within the mature and seed miRNA region in four species: pig, horse, cattle, and chicken (**Table 1**). The assembled list is supplemented with updated information regarding the genomic location of miRNA genes (intergenic or within host gene), their position within host genes (intron or exon), strand orientation, and validation status of the polymorphisms. The genomic location of miRNA genes comprising miR-seed polymorphisms in cattle and chicken is presented on the online Genomic View (http://www.integratomics-time.com/miR-seed-SNPs/genomic_view), which also presents overlapping QTL.

The number of polymorphisms within pre-miRNAs in chicken and cattle (125 and 92, respectively) was much higher compared to the horse and pig genomes (one and two, respectively) (**Figure 3A**). This observation may not reflect the lack of genetic variability in the horse or pig but rather differences in the timing of genome project (*i.e.* chicken was assembled in 2006, pig in 2011) and intensity of genome variation research in a particular species. The number of assembled polymorphisms is not final and will change with time as all miRNAs have not yet been systematically sequenced and screened for polymorphisms. However, using current genome data of species examined in our study, we found 59 polymorphic miRNA genes that contained 65 polymorphisms within the mature miRNAs (**Figure 3A**, **Table 1**). These include 28 polymorphisms overlapping with the seed region: 19 in cattle and nine in chicken. Two consecutive miR-seed-SNPs forming a DNP were detected residing within the cattle *bta-mir-2489*. The distribution of polymorphisms within the miRNA mature region in four analyzed species is presented in **Figure 3B**. Polymorphisms adjacent to the seed region; located on the nucleotide positions 1 and 9 within the mature miRNA, could also be considered for their effect on target recognition and binding. Our catalog includes three polymorphisms located on nucleotide position 1 and four on position 9, in three species (**Figure 3B**). Other locations of interest are Drosha and Dicer cleavage sites, located on both ends of mature miRNAs, which were found polymorphic in cattle (four SNPs) and chicken (one SNP). According to the NCBI database all assembled polymorphisms residing within the mature region in cattle have an unknown validation status. On the other hand, 16/21 polymorphisms within the mature miRNA region in chicken have available allele frequency or genotype data. Moreover, four miR-seed polymorphisms (rs14076349, rs14281065, rs14934924, and rs16681031) were experimentally validated (Chuan-sheng *et al.* 2010; Geng *et al.* 2011; Zhang *et al.* 2011). The polymorphism rs14934924 was also genotyped and associated with chicken growth and meat traits in the Chinese Gushi-Anka F2 resource population (Li *et al.* 2012). Additionally, by examining three SNP microarrays (chicken Illumina 60K SNP chip, PorcineSNP60, and cattle 770K Illumina HD SNPs) we found one SNP overlapping miRNA gene; namely chicken SNP rs15190357 is located within mature region of *gga-*

mir-1596. However, because all miRNAs have not yet been systematically sequenced, many polymorphisms remain still unvalidated, while on the other hand some SNPs in the databases may be results of sequencing errors.

It was shown in Mencia et al. (Mencia et al. 2009) and our previous study (Zorc et al. 2012) that miR-seed-SNPs can cause the seed region from one miRNA to perfectly match with the seed region from another miRNA, thereby potentially generating targets shared by both miRNAs or novel targets. The analysis using TargetScan Custom tool, which defines miRNA targets based on miRNA seed region recognition, revealed that SNP rs41825418 (T>C) alters the seed region of *bta-mir-29e*, resulting in the seed sequence *bta-mir-29b-2* (**Figure 4**). Due to this T>C substitution 153/758 predicted targets were gained/lost.

Genomic distribution of polymorphic miRNA genes and their overlaps with host genes and QTL

As expected approximately half of miRNA genes comprising polymorphisms were found to be located within protein-coding host genes (28/57) (**Table 1**). Two bovine miRNAs *bta-mir-1291* and *-2484* additionally overlapped with a non-coding gene; small nucleolar RNA (snoRNA) *SNORA2* and *SNORD61*, respectively. Most of the intragenic polymorphic miRNA genes (26/28) had the same strand orientation as their host genes. The majority of intragenic miRNA genes resided within introns of their host genes (27/28), whereas only *bta-mir-2900* was exonic, comprising a synonymous polymorphism rs137603330. It was previously observed, that miRNA genes exhibiting a high degree of evolutionary conservation, are often coordinately regulated and/or expressed with their host genes (He *et al.* 2012). For example, a miRNA/host gene pair (*mir-338*/*AATK*) was observed expressed in the nervous system in both human and mouse (Baskerville & Bartel 2005; Bak *et al.* 2008; Barik 2008). The matching location and orientation of polymorphic miRNA genes and their host gene pairs provides a starting-point for further functional analyses of the present polymorphisms.

Our previous study revealed several overlaps comprising polymorphic miRNA genes, QTL and fragile sites (Zorc et al. 2012). In this study we found that bovine and chicken miRNA genes comprising seed-polymorphisms overlapped with several QTL associated with production traits (http://www.integratomics-time.com/miR-seed-SNPs/genomic_view). Bovine miRNA genes were found to overlap with 101 QTL. The analysis revealed an interesting hotspot, namely *bta-mir-2489* comprising a DNP within the seed region overlaps with the largest number of QTL (n=25), associated with reproduction, behavior, meat and milk quality. Chicken miRNA genes comprising miR-seed polymorphisms overlap with 37 QTL, associated with several production (*e.g.* body weight, drumstick, thigh and wing weight, egg weight) and health traits (*e.g.* Marek's disease-related traits). These miRNA SNPs can now be used in re-analyses of existing crosses or mapping

populations to test for associations with indicated traits or to evaluate if these SNPs affect expression levels of target genes in selected tissues. The performed miRNA/QTL overlap will enable more effective phenotype-related analyses, which would provide the researchers with an additional source of biomarkers for traits in animal production and breeding.

Experimental validation of selected miRNA genes comprising miR-seed polymorphisms

Several polymorphisms deposited in genomic databases result from sequencing errors and therefore we experimentally validated six miR-seed-polymorphisms; three in chicken and three in cattle, two of which formed a DNP (**Figure S1A-C**). The DNP comprising rs110544069 and rs109834057 located within *bta-mir-2489* and gene *sterile alpha motif domain containing 12* (*SAMD12*) was validated in Cika cattle breed (**Figure S1A**). Because DNPs consist of two consecutive nucleotide polymorphisms and hence have a greater potential effect on target recognition and functional consequences, we performed a TargetScan Custom analysis, which showed a change in the number of targets. Due to the CA>TG substitution 278/147 predicted targets were gained/lost, out of which only eight targets were shared by both miRNA seed region variants. Interestingly, several DNPs were previously found to be present within miRNA seed regions in human and mouse (Zorc *et al.* 2012). Additionally, miR-seed-SNP rs41761413 residing within *bta-mir-2313* and protein-coding host gene *GRAM domain containing 1B* (*GRAMD1B*) was found to be polymorphic in the analyzed population of Slovenian Simmental cattle (**Figure S1B**). All three chicken miR-seed-SNPs (rs16681031, rs16681032, and rs16681033) residing within *gga-mir-1658* and host gene *guanine monophosphate synthetase* (*GMPS*) were polymorphic in four breeds (Brown, Barred, Styrian hen, and Lohmann brown classic) (**Figure S1C**). These miR-seed-SNPs could be further applied in research investigating their association with production traits in cattle and chicken.

Conclusion

MicroRNA polymorphisms may have a profound effect on a wide range of phenotypes; therefore genetic variability of miRNA genes in livestock animals provides a source of novel biomarkers for phenotypic traits in animal breeding. The upgraded bioinformatics tool miRNA SNiPer 3.0 was used for assembling a list of all known miRNA SNPs in four livestock species (pig, horse, cattle, and chicken), which overlapped with several QTL. The miRNA SNiPer tool will be periodically updated with novel miRNAs and SNPs, in livestock and other species. This tool has an added value in comparative genomics aspect of miRNA-SNP research enabling retrieval and prioritization of valuable information across several mammalian species and hence helps researchers to improve experimental designs. This project may yield new findings useful for development of molecular markers in selection programs allowing more effective, marker assisted selection in livestock species.

Acknowledgments

The authors thank Antonija Holcman, PhD and Dusan Tercic, PhD for providing the chicken samples.
The authors declare that there are no conflicts of interest.

Funding

This work was supported by the Slovenian Research Agency (ARRS) through the Research programme Comparative genomics and genome biodiversity [grant number P4-0220] and I.G. is supported through the young researcher scheme of ARRS.

FIGURES AND TABLES:

Figure 1: Secondary structure of miRNA.

Figure 2: MiRNA SNiPer 3.0 output showing genetic variability *of bta-mir-2489* in cattle. Double nucleotide polymorphism consisting of two consecutive SNPs, rs110544069 and rs109834057 residing within seed region, and SNP rs36326300 within pre-miRNA region.

Figure 3: Distribution of polymorphisms within miRNA genes in cattle, chicken, horse, and pig. **A**) Distribution of polymorphisms among miRNA regions: pre-miRNA, mature and seed region. **B**) Number of polymorphisms per nucleotide position within mature miRNA.

Figure 4: MiR-seed-SNP causing formation of a novel seed region. MiR-seed-SNP rs41825418 located within miRNA *bta-mir-29e* causes a formation of a seed region annotated to miRNA *bta-mir-29b-2*. Seed region is indicated with a rectangle.

Table 1: MicroRNA genes with polymorphisms located within their mature and seed regions in four livestock species (cattle, chicken, horse, and pig).

SUPPORTING INFORMATION:

Figure S1: Experimental validation of polymorphisms located within the seed region in three miRNA genes. **A**) Validation of a DNP located within *bta-mir-2489.* **B**) Validation of rs41761413 located within *bta-mir-2313* in cattle. **C**) Validation of all three miR-seed-SNPs located within *gga-mir-1658* in chicken.

Table S1: Main characteristics of bioinformatics tools for analysis of miRNA genetic variability.

Table S2: Genome assemblies for ten livestock species applied by source genomic databases.

Table 1: MicroRNA genes with polymorphisms located within their mature and seed regions in four livestock species: cattle, chicken, horse, and pig.

• Species [assembly] • no. of miRNAs • no. of polymorphisms	miRNA host gene (orientation)	orientation/ genic location	miRNA genes comprising polymorphisms within mature region	ID of polymorphisms located within the miRNA mature region (polymorphisms in miRNA seed region in bold)	identification / validation / genotyping
• Pig [Sscrofa9] • 271 miRNAs • 545,950 polymorphisms	*F1SR76_PIG*	as / intron 10	*ssc-miR-4334-3p*	rs10719650	/
			ssc-miR-4334-5p	rs10719651	/
• Horse [EquCab2] • 341 miRNAs • 1,163,258 polymorphisms	intergenic	-	*eca-miR-514*	rs69481895	/
• Cattle [UMD3.1] • 766 miRNAs • 2,201,071 polymorphisms	*MAML2*	-	*bta-miR-1260b*	**rs134960452**	/
	CL041_BOVIN SNORA2	s / intron 3 s	*bta-miR-1291*	**rs110817643**	/
	intergenic	-	*bta-miR-132*	rs133988074	/
	MIB1	as / intron 12	*bta-miR-133a-2*	**rs137070651**	/
	TAGL2	s / intron 1	*bta-miR-1584-3p*	rs110121441	/
	F1MBP1_BOVIN	s / intron 1	*bta-miR-1777b*	rs133532692	/
	intergenic	-	*bta-miR-2285m-4*	**rs109856770**	/
	intergenic	-	*bta-miR-2286*	rs132885360 rs135512047	/
	RIN2	s / intron 4	*bta-miR-2305*	rs135483577	/
	GRAMD1B	s / intron 2	*bta-miR-2313-3p*	**rs41761413**	this study
	MADD	s / intron 14	*bta-miR-2318*	**rs134638324**	/
	intergenic	-	*bta-miR-2321*	rs137731966	/
	SF3B3_BOVIN	s / intron 20	*bta-miR-2324*	rs134817229	/
	PTPRG	s / intron 1	*bta-miR-2369*	**rs136780194 rs134753593**	/
	intergenic	-	*bta-miR-2402*	**rs137803906**	/
	intergenic	-	*bta-miR-2403*	**rs135523674**	/
	LAMB1	s / intron 11	*bta-miR-2418*	rs110555814	/
	intergenic	-	*bta-miR-2419-3p*	rs43400521	/
	NCKAP5L	s / intron 1	*bta-miR-2425-3p*	**rs134756375**	/
	intergenic	-	*bta-miR-2436-3p*	rs135821414	/
	HTT	s / intron 20	*bta-miR-2450c*	**rs42658514**	/
	intergenic	-	*bta-miR-2455*	rs137716256	/
	intergenic	-	*bta-miR-2467-*	rs109063650	/

• Species [assembly] • no. of miRNAs • no. of polymorphisms	miRNA host gene (orientation)	orientation/ genic location	miRNA genes comprising polymorphisms within mature region	ID of polymorphisms located within the miRNA mature region (polymorphisms in miRNA seed region in bold)	identification / validation / genotyping
			5p		
	D3JUI8_BOVIN SNORD61	s / intron 2 s	*bta-miR-2484*	rs135930835	/
	SAMD12	s / intron 4	*bta-miR-2489*	**rs110544069*** **rs109834057***	this study this study
	intergenic	-	*bta-miR-2892*	**rs109795153**	**/**
	intergenic	-	*bta-miR-2899*	**rs109462250**	**/**
	PEG3	s / exon 8	*bta-miR-2900*	rs137603330	/
	intergenic	-	*bta-miR-2903*	rs134036560	/
	intergenic	-	*bta-miR-29b-2*	**rs133392085**	**/**
	intergenic	-	*bta-miR-29d*	rs137831203	/
	intergenic	-	*bta-miR-29e*	**rs41825418**	**/**
	F1MIB7_BOVIN	s / intron 6	*bta-miR-338*	rs133207384	/
	intergenic	-	*bta-miR-375*	rs134683062	/
	intergenic	-	*bta-miR-411c-5p*	rs132740734	/
	MEGF6	s / intron 3	*bta-miR-551a*	**rs136014887**	**/**
	KAT6B	s / intron 2	*bta-miR-584-3*	rs135786138	/
	intergenic	-	*bta-miR-940*	rs110936400	/
• Chicken [WASHUC2] • 499 miRNAs • 3,292,991 polymorphisms	*PTPRG*	s / intron 2	*gga-miR-1550-5p*	rs14981477	NCBI
	intergenic	-	*gga-miR-1568*	**rs14511527** rs14511526	NCBI NCBI
	CELF2	s / intron 1	*gga-miR-1596-3p*	rs15190357	NCBI, Illumina 60K SNP chip
	intergenic	-	*gga-miR-1605*	ss538056207	/
	intergenic	-	*gga-miR-1614-3p*	**rs15172520**	NCBI
	ENSGALG00000005676	s / intron 12	*gga-miR-1625-3p*	ss538087094	/
	XYLT1	s / intron 5	*gga-miR-1644*	**rs14076349**	(Chuan-sheng *et al.* 2010)
	intergenic	-	*gga-miR-1648-3p*	**rs14281065** rs14281066	(Zhang *et al.* 2011)
	E1C752_CHICK	s / intron 2	*gga-miR-1657*	**rs14934924**	(Li *et al.* 2012)
	GMPS	s / intron 13	*gga-miR-1658-3p*	**rs16681031** **rs16681032**	(Geng *et al.* 2011), this study this study
			gga-miR-1658-5p	**rs16681033**	this study
	intergenic	-	*gga-miR-1675*	rs16643637	NCBI
	TRAIP	s / intron 4	*gga-miR-1678*	**ss537949696**	**/**

• Species [assembly] • no. of miRNAs • no. of polymorphisms	miRNA host gene (orientation)	orientation/ genic location	miRNA genes comprising polymorphisms within mature region	ID of polymorphisms located within the miRNA mature region (polymorphisms in miRNA seed region in bold)	identification / validation / genotyping
	intergenic	-	*gga-miR-1703*	rs16718265	NCBI
	TTK	s / intron 4	*gga-miR-1712-5p*	rs15403608	NCBI
	intergenic	-	*gga-miR-1747*	ss538180149	/
	COL27A1	s / intron 17	*gga-miR-455-5p*	rs15031616	NCBI
	intergenic	-	*gga-miR-460b-5p*	ss538066607	/

*two consecutive SNPs forming a DNP

s: sense orientation; as: antisense orientation

-: data not applicable; /: data not available

host gene names: *CELF2* - CUGBP, Elav-like family member 2; *CL041_BOVIN* (*KANSL2*) - KAT8 regulatory NSL complex subunit 2; *COL27A1* - collagen, type XXVII, alpha 1; *E1C752_CHICK* (*RAB38*) - RAB38, member RAS oncogene family; *ENSGALG00000005676* (*TTC28*) - tetratricopeptide repeat domain 28; *D3JUI8_BOVIN* (***RBMX***) - RNA binding motif protein, X-linked; *F1MBP1_BOVIN* (***PDGFRB***) - beta-type platelet-derived growth factor receptor precursor; *F1MIB7_BOVIN (AATK)* - apoptosis-associated tyrosine kinase; *F1SR76_PIG* (*DNPEP*) - aspartyl aminopeptidase; *GMPS* - guanine monphosphate synthetase; *GRAMD1B* - GRAM domain containing 1B; *HTT* - huntingtin; *KAT6B* - K(lysine) acetyltransferase 6B; *LAMB1* - laminin, beta 1; *MADD* - MAP-kinase activating death domain; *MAML2* - mastermind-like 2 (Drosophila); *MEGF6* - multiple EGF-like-domains 6; *MIB1* - mindbomb E3 ubiquitin protein ligase 1; *NCKAP5L* - NCK-associated protein 5-like; *PEG3* - paternally expressed 3; *PTPRG* - protein tyrosine phosphatase, receptor type, G; *RIN2* - Ras and Rab interactor 2; *SAMD12* - sterile alpha motif domain containing 12; *TAGL2* - Transgelin-2; *TRAIP* - TRAF interacting protein; *TTK* - TTK protein kinase; *XYLT1* - xylosyltransferase I.

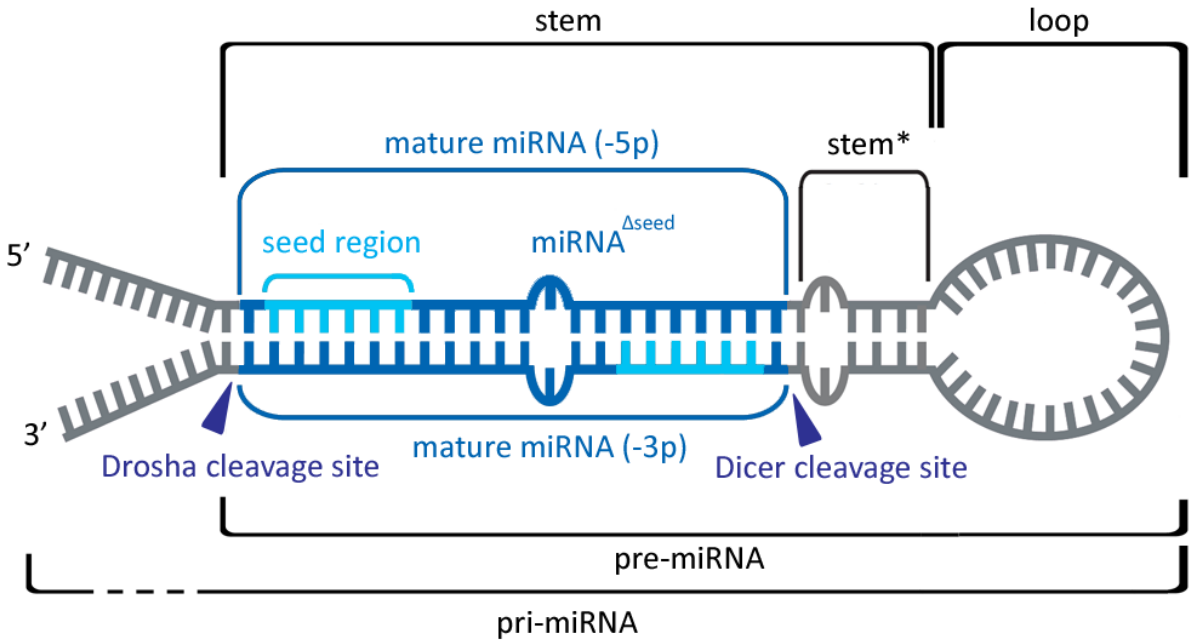


Figure 1: Secondary structure of miRNA.

miRNA name	miRNA	mature miRNA	variation	details
bta-mir-2489	Bos taurus 14:48094677-48094754[+]	bta-miR-2489 **Mature:** 48094688-48094709 **Seed:** 48094689-48094695 from TargetScan UGUUGUGUUUCAAAUGACAGGGGACAUGAGU UUUUUAAUCUGGCAGUUUAAACUUUUGCCCUCC UUCAUUACAUAACG	rs110544069	In seed 48094694 SNP (C > T)
			rs109834057	In seed 48094695 SNP (A > G)
			rs136326300	In pre-mature 48094753 SNP (C > T)

Figure 2: MiRNA SNiPer 3.0 output showing genetic variability *of bta-mir-2489* in cattle. Double nucleotide polymorphism consisting of two consecutive SNPs, rs110544069 and rs109834057 residing within seed region, and SNP rs36326300 within pre-miRNA region.

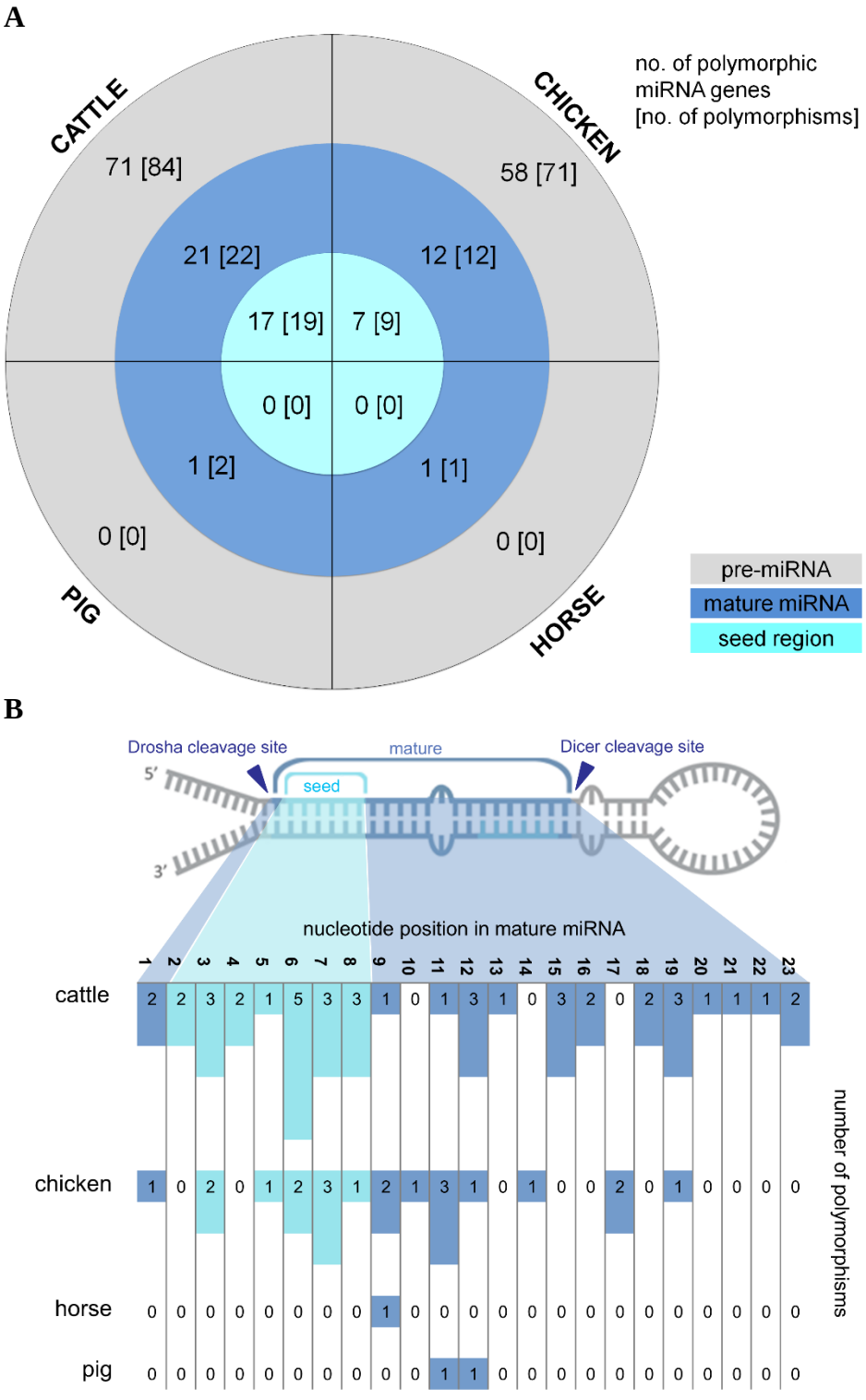


Figure 3: Distribution of polymorphisms within miRNA genes in cattle, chicken, horse, and pig. **A**) Distribution of polymorphisms among miRNA regions: pre-miRNA, mature and seed region. **B**) Number of polymorphisms per nucleotide position within mature miRNA.

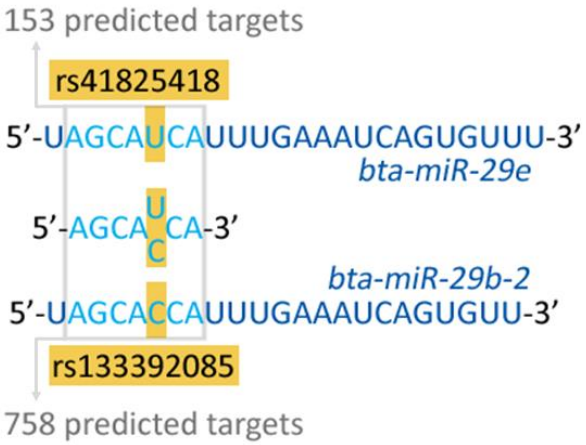


Figure 4: MiR-seed-SNP causing formation of a novel seed region. MiR-seed-SNP rs41825418 located within miRNA *bta-mir-29e* causes a formation of a seed region annotated to miRNA *bta-mir-29b-2*. Seed region is indicated with a rectangle.

References

Bak M., Silahtaroglu A., Møller M., Christensen M., Rath M.F., Skryabin B., Tommerup N. & Kauppinen S. (2008) MicroRNA expression in the adult mouse central nervous system. *RNA* **14**, 432-44.

Barik S. (2008) An intronic microRNA silences genes that are functionally antagonistic to its host gene. *Nucleic Acids Res* **36**, 5232-41.

Bartel D.P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281-97.

Baskerville S. & Bartel D. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *Rna-a Publication of the Rna Society*, 241-7.

Cargill E.J., Nissing N.J. & Grosz M.D. (2008) Single nucleotide polymorphisms concordant with the horned/polled trait in Holsteins. *BMC Res Notes* **1**, 128.

Chuan-sheng Z., Li-ying G., Zheng-zhu L., Zhi-xin F., Shen P. & Li-xin D. (2010) Frequency distribution of T > C polymorphisms in the seed region of the chicken mir-1644 gene. In: *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*, pp. 2570-3.

Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X., Bibé B., Bouix J., Caiment F., Elsen J.M., Eychenne F., Larzul C., Laville E., Meish F., Milenkovic D., Tobin J., Charlier C. & Georges M. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* **38**, 813-8.

Ferdin J., Kunej T. & Calin G. (2010) Non-coding RNAs: Identification of Cancer-Associated microRNAs by Gene Profiling. *Technology in Cancer Research & Treatment* **9**, 123-38.

Geng L.-Y., Zhang C.-S., Li Y.-Y., Li Y.-Y., Qiu-Yuewang, Sun H.-N. & Li X.-S. (2011) The Chicken *GGA-Mir-1658** Gene: Seed Region Polymorphisms, Frequency Distribution and Putative Targets. *Journal of Animal and Veterinary Advances* **10**, 1187-93.

Georges M., Coppieters W. & Charlier C. (2007) Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. *Curr Opin Genet Dev* **17**, 166-76.

Godnic I., Zorc M., Jevsinek Skok D., Calin G.A., Horvat S., Dovc P., Kovac M. & Kunej T. (2013) Genome-wide and species-wide *in silico* screening for intragenic microRNAs in human, mouse and chicken. PLoS ONE - in press.

Gong J., Tong Y., Zhang H.M., Wang K., Hu T., Shan G., Sun J. & Guo A.Y. (2012) Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum Mutat* **33**, 254-63.

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 Res*.

Hiard S., Charlier C., Coppieters W., Georges M. & Baurain D. (2010) Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Research* **38**, D640-D51.

Hu Z.L., Park C.A., Wu X.L. & Reecy J.M. (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res* **41**, D871-9.

Kinsella R.J., Kähäri A., Haider S., Zamora J., Proctor G., Spudich G., Almeida-King J., Staines D., Derwent P., Kerhornou A., Kersey P. & Flicek P. (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**, bar030.

Kiriakidou M., Nelson P.T., Kouranov A., Fitziev P., Bouyioukos C., Mourelatos Z. & Hatzigeorgiou A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* **18**, 1165-78.

Kozomara A. & Griffiths-Jones S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* **39**, D152-D7.

Krek A., Grün D., Poy M.N., Wolf R., Rosenberg L., Epstein E.J., MacMenamin P., da Piedade I., Gunsalus K.C., Stoffel M. & Rajewsky N. (2005) Combinatorial microRNA target predictions. *Nat Genet* **37**, 495-500.

Kunej T., Godnic I., Ferdin J., Horvat S., Dovc P. & Calin G.A. (2011) Epigenetic regulation of microRNAs in cancer: an integrated review of literature. *Mutat Res* **717**, 77-84.

Kunej T., Godnic I., Horvat S., Zorc M. & Calin G.A. (2012) Cross Talk Between MicroRNA and Coding Cancer Genes. *Cancer J* **18**, 223-31.

Kunej T., Jevsinek Skok D., Horvat S., Dovc P. & Jiang Z. (2010) The Glypican 3-Hosted Murine Mir717 Gene: Sequence Conservation, Seed Region Polymorphisms and Putative Targets. *International Journal of Biological Sciences*, 769-72.

Lee J.S., Kim J.M., Lim K.S., Hong J.S., Hong K.C. & Lee Y.S. (2013) Effects of polymorphisms in the porcine microRNA MIR206 / MIR133B cluster on muscle fiber and meat quality traits. *Anim Genet* **44**, 101-6.

Lee Y., Jeon K., Lee J.-T., Kim S. & Kim V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663-70.

Lewis B.P., Burge C.B. & Bartel D.P. (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* **120**, 15-20.

Li H., Sun G.R., Lv S.J., Wei Y., Han R.L., Tian Y.D. & KANG X.T. (2012) Association study of polymorphisms inside the miR-1657 seed region with chicken growth and meat traits. *British Poultry Science*.

Liu H.C., Hicks J.A., Trakooljul N. & Zhao S.H. (2010) Current knowledge of microRNA characterization in agricultural animals. *Anim Genet* **41**, 225-31.

Mencia A., Modamio-Hoybjor S., Redshaw N., Morin M., Mayo-Merino F., Olavarrieta L., Aguirre L.A., del Castillo I., Steel K.P., Dalmay T., Moreno F. & Moreno-Pelayo M.A. (2009) Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet* **41**, 609-13.

Nahvi A., Shoemaker C.J. & Green R. (2009) An expanded seed sequence definition accounts for full regulation of the hid 3' UTR by bantam miRNA. *RNA* **15**, 814-22.

Sun G., Yan J., Noltner K., Feng J., Li H., Sarkis D.A., Sommer S.S. & Rossi J.J. (2009) SNPs in human miRNA genes affect biogenesis and function. *RNA* **15**, 1640-51.

Zhang C.-S., Geng L.-Y., Yang-Qing, Zhu W.-J., Wang Q.-Y., Zhang X.-Z., Sun H.-N., Li X.-S., Liu R.-Z. & Liu X.-H. (2011) The RAB38-Hosted Chicken *Mir-1657* Gene: Sequence Conservation, Seed Region Polymorphisms and Putative Targets. *Journal of Animal and Veterinary Advances* **10**, 2343-8.

Zorc M., Jevsinek Skok D., Godnic I., Calin G.A., Horvat S., Jiang Z., Dovc P. & Kunej T. (2012) Catalog of MicroRNA Seed Polymorphisms in Vertebrates. *PLoS One* **7**, e30737.

Priloga H

Genome-wide and species-wide *in silico* screening for intragenic microRNAs in human, mouse and chicken (Godnič in sod., 2013)

OPEN ACCESS Freely available online

PLOS ONE

Genome-Wide and Species-Wide *In Silico* Screening for Intragenic MicroRNAs in Human, Mouse and Chicken

Irena Godnic[1], Minja Zorc[1], Dasa Jevsinek Skok[1], George Adrian Calin[2], Simon Horvat[1,3], Peter Dovc[1], Milena Kovac[1], Tanja Kunej[1]*

1 Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Domzale, Slovenia, **2** Department of Experimental Therapeutics and The Center for RNA Interference and Non-Coding RNAs, The University of Texas, M. D. Anderson Cancer Center, Houston, Texas, United States of America, **3** National Institute of Chemistry, Ljubljana, Slovenia

Abstract

MicroRNAs (miRNAs) are non-coding RNAs (ncRNAs) involved in regulation of gene expression. Intragenic miRNAs, especially those exhibiting a high degree of evolutionary conservation, have been shown to be coordinately regulated and/or expressed with their host genes, either with synergistic or antagonistic correlation patterns. However, the degree of cross-species conservation of miRNA/host gene co-location is not known and co-expression information is incomplete and fragmented among several studies. Using the genomic resources (miRBase and Ensembl) we performed a genome-wide *in silico* screening (GWISS) for miRNA/host gene pairs in three well-annotated vertebrate species: human, mouse, and chicken. Approximately half of currently annotated miRNA genes resided within host genes: 53.0% (849/1,600) in human, 48.8% (418/855) in mouse, and 42.0% (210/499) in chicken, which we present in a central publicly available Catalog of intragenic miRNAs (http://www.integratomics-time.com/miR-host/catalog). The miRNA genes resided within either protein-coding or ncRNA genes, which include long intergenic ncRNAs (lincRNAs) and small nucleolar RNAs (snoRNAs). Twenty-seven miRNA genes were found to be located within the same host genes in all three species and the data integration from literature and databases showed that most (26/27) have been found to be co-expressed. Particularly interesting are miRNA genes located within genes encoding for miRNA silencing machinery (*DGCR8*, *DICER1*, and *SND1* in human and *Cnot3*, *Gdcr8*, *Eif4e*, *Tnrc6b*, and *Xpo5* in mouse). We furthermore discuss a potential for phenotype misattribution of miRNA host gene polymorphism or gene modification studies due to possible collateral effects on miRNAs hosted within them. In conclusion, the catalog of intragenic miRNAs and identified 27 miRNA/host gene pairs with cross-species conserved co-location, co-expression, and potential co-regulation, provide excellent candidates for further functional annotation of intragenic miRNAs in health and disease.

Citation: Godnic I, Zorc M, Jevsinek Skok D, Calin GA, Horvat S, et al. (2013) Genome-Wide and Species-Wide *In Silico* Screening for Intragenic MicroRNAs in Human, Mouse and Chicken. PLoS ONE 8(6): e65165. doi:10.1371/journal.pone.0065165

Editor: Rolf Müller, Philipps University, Germany

Received January 8, 2013; **Accepted** April 22, 2013; **Published** June 6, 2013

Funding: This work was supported by the Slovenian Research Agency (ARRS) through the Research programme Comparative genomics and genome biodiversity [grant number P4–0220]. GAC is supported as a Fellow at The University of Texas M. D. Anderson Research Trust, as a Fellow of The University of Texas System Regents Research Scholar, and by the CLL Global Research Foundation. Work in Dr. Calin's laboratory is supported in part by National Institutes of Health, by Department of Defense, by Developmental Research Awards in Breast Cancer, Ovarian Cancer and Leukemia SPOREs, and by 2009 Seena Magowitz - Pancreatic Cancer Action Network - AACR Pilot Grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tanja.kunej@bf.uni-lj.si

Introduction

MicroRNAs (miRNAs) are non-coding RNAs (ncRNAs) that post-transcriptionally regulate gene expression. The standard dogma states that expression of protein-coding genes is repressed by binding the target gene's complementary sequence in the 3′ untranslated region (3′-UTR) with the miRNA's seed region: 2–7 or 2–8 consecutive nucleotides from the 5′-end of the miRNA, which are crucial for target recognition [1,2]. This earlier postulated dogma has now been extended with new discoveries. MicroRNAs have also been shown to increase or decrease expression of protein-coding genes by targeting different genomic regions (3′-UTR, 5′-UTR, promoter, and coding sequences) and interact with proteins. Additionally, they have been shown to function in various subcellular compartments, and developmental and metabolic processes [3]. Several components of the miRNA processing machinery are included in miRNA biogenesis, which first take place in the nucleus. Primary miRNA transcripts (pri-miRNAs) are processed by the complex Drosha-DGCR8 (DiGeorge syndrome critical region gene-8), a component of the miRNA processing machinery [4,5]. Thereafter precursor miRNAs (pre-miRNAs) are transported to the cytoplasm where they are further cleaved by RNase III Dicer, a key enzyme in miRNA maturation, to form functional mature miRNAs [6]. They are incorporated into the RNA-induced silencing complex (RISC) composed of many associated proteins [7]. Disruption of the miRNA processing machinery core components, miRNA genes and their targets affects overall efficiency of silencing [8]. Indeed, polymorphisms as well as aberrant miRNA expression patterns have previously been shown to be involved in disease development, including several cancer types [9–12].

Approximately half of vertebrate miRNAs are processed from introns of protein-coding genes or genes encoding for other

ncRNA classes (*e.g.* snoRNAs, miRNAs, lincRNAs) [13], whereas miRNA genes can also be encoded in intergenic regions of DNA, therefore referred to as intergenic miRNAs. In some cases, a miRNA gene can have a "mixed" location, *i.e.* can be located either in an exon or an intron of the same or different host gene transcripts which depends on their alternative splicing [13].

A single host gene can comprise multiple and overlapping resident miRNA genes, called a cluster, which are processed from the same polycistronic primary transcript [13,14]. It has been observed that miRNA genes which are located in a polycistron and co-expressed in the clusters are pivotal in coordinately regulating multiple processes, including embryonic development, cell cycle and cell differentiation [15]. It was also observed that miRNA genes are more frequently hosted within the short genes than expected by chance, which was hypothesized as a favorable evolutionary feature due to the gene's interaction with the pre-miRNA splicing mechanism [16].

Host genes and resident ncRNAs have been considered to have a synergistic effect with important implications for fine-tuning gene expression patterns in the genome [17,18]. Expression profiles of intronic miRNAs were in many cases found to coincide with the transcription of their host genes, which raised a question as to how these miRNAs were processed [19]. Intronic miRNAs, like most ncRNAs, are released from the excised host introns in the post-splicing process [17,20]. However, it was later indicated that intronic miRNAs might also be processed from unspliced intronic regions prior to splicing catalysis [20]. A class of miRNA precursors, named mirtrons, are processed in an alternative miRNA biogenesis pathway where certain debranched introns mimic the structural features of pre-miRNAs and enter the miRNA-processing pathway, however without the Drosha-mediated cleavage [21].

Highly correlated expression patterns have been found in closely clustered miRNA genes (50 kb of each other), which coincides with the idea of a polycistronic primary transcript [19,22]. He et al. [23] additionally showed that evolutionary conserved miRNA genes tend to be co-expressed with their host genes: even though the non-conserved miRNAs dominate in the human genome, the majority of intragenic miRNAs exhibiting co-expression with their host gene are phylogenetically old. A high conservation between orthologous intronic miRNAs has been demonstrated in several species [24,25]. In addition to co-expression and proposed co-regulation of miRNA and host genes, several studies have described a functional link between them [19,26,27]. Interestingly, genes highly correlated in expression with a resident miRNA gene were found to be more likely predicted as miRNA targets [28]. The expression of miRNA/host genes and that of predicted miRNA targets tend to be positively or negatively correlated, suggesting that the coordinated transcriptional regulation of a miRNA and its target is an abundant motif in gene networks [28].

The proportion of miRNA genes located within the same host genes among different species remains unknown, whether their coordinated expression is conserved, and to what degree. The miRNA/host gene co-expression has been analyzed in several studies, yet the data remains fragmented and incomplete. However, based on the report by He *et al.* [23] that evolutionary conserved ("old") miRNA genes tend to be co-expressed with their host genes, but, in contrast, non-conserved ("young") ones rarely do so, it might be reasonable to predict the same co-expression patterns of miRNA/host gene pairs with conserved cross-species co-location. The conserved pairs would present candidate genes whose matching expression profiles would be of assistance for further annotation and functional analysis.

The aim of this study was to create a central Catalog of intragenic miRNAs in three well-annotated vertebrate species (human, mouse, and chicken) serving as a framework for researchers working in the field of intragenic miRNAs. The supplemented information regarding the miRNA/host gene pair's conserved cross-species co-location, expression data, and disease associations provides a list of high priority intragenic miRNAs for further functional analyses. These include identification and annotation of genes based on cross-species conservation, functional analyses and studies to re-examine potential misattribution of phenotype previously ascribed to host genes or hosted miRNA genes only.

Materials and Methods

Datasets of miRNA/host gene pairs were downloaded from genomic resources: the coordinates of miRNA genes and their host genes in human, mouse, and chicken were downloaded from miRBase, release 19 (http://www.mirbase.org/) [29] and Ensembl, release 69 (http://www.ensembl.org/index.html), using the latest matching assemblies: GRCh37 for human, GRCm38 for mouse, and WASHUC2 for chicken. The catalog is accessible through a web application written in PHP language, which allows retrieving miRNA/host gene pairs (http://www.integratomics-time.com/miR-host/catalog). The nomenclature of miRNA and host genes was unified according to The HUGO Gene Nomenclature Committee (HGNC) (http://www.genenames.org/) and Mouse Genome Informatics (MGI) (http://www.informatics.jax.org/). The list of miRNA host genes was manually inspected; cases with doubtful gene nomenclature after automatic annotation (*e.g.* overwriting of a miRNA record with an overlapping snoRNA and lincRNA record) were reported to the source database (Ensembl) and solved case by case. Genomic distribution of miRNA/host gene pairs in human, mouse, and chicken was presented in a genomic view format using Flash GViewer web tool (http://gmod.org/wiki/Flashgviewer/). MicroRNA and host gene expression profiles, their functional links and diseases associated with dysregulated expression were retrieved from: 1) literature using PubMed (http://www.ncbi.nlm.nih.gov/pubmed), Web of Science (http://apps.webofknowledge.com/), and 2) databases Gene Expression Atlas (GEA), release 2.0.11.1 (http://www.ebi.ac.uk/gxa/). Small RNA expression data was obtained from University of California Santa Cruz (UCSC) Genome Bioinformatics (http://genome.ucsc.edu/) based on the ENCODE project [30]. Genetic variability of miRNA genes residing within host genes (protein-coding and non-coding) was determined using miRNA SNiPer tool 3.0 (http://www.integratomics-time.com/miRNA-SNiPer) [31]. Predicted and experimentally validated miRNA targets were obtained using TargetScan (http://www.targetscan.org/), miRecords (http://mirecords.biolead.org/), and miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/). The list of components of the miRNA silencing machinery was obtained from Patrocles database (http://www.patrocles.org) [32]. Pathway enrichment analysis for miRNA host genes was performed using the Ingenuity Pathway Analysis (IPA), release 8.8 (Ingenuity® Systems, http://www.ingenuity.com/) [33]. Multispecies sequence alignments were performed using Ensembl, option Comparative genomics - Alignments (text).

Results and Discussion

We developed a central Catalog of intragenic miRNAs in three well-annotated vertebrate genomes (human, mouse, and chicken) by performing a genome-wide *in silico* screening (GWISS) of

genomic resource databases (**Figures 1** and **2**). The miRNAs were hosted by protein-coding genes or genes encoding for other ncRNA classes. Further species-wide *in silico* screening (SWISS) revealed 27 miRNA/host gene pairs with conserved co-location in all three species, most of which have been found to be co-expressed. Coordinately expressed miRNA/host gene pairs with cross-species conserved co-location are considered prioritized candidate genes for future functional analysis.

1. Genome-wide *in silico* Screening (GWISS) for Sense-oriented miRNA/host Gene Pairs in Human, Mouse and Chicken

Intragenic miRNAs (**Figure 3**) have become a topic of increasing research interest. We performed a genome-wide *in silico* screening (GWISS) of the latest genome assemblies of three well-annotated vertebrate genomes (human, mouse, and chicken) to define how many miRNA genes are located within host genes. The Catalog of intragenic miRNAs is available through a web application (http://www.integratomics-time.com/miR-host/ catalog), which allows users to retrieve single or multiple miRNA/host gene pairs, based on 1) selection of species, biotype of host genes, and genomic position of resident miRNAs (exon, intron, 3′ and 5′-UTR), or 2) by querying individual miRNA or their host genes. In all three species approximately half of currently annotated miRNAs are intragenic, residing within protein-coding and/or ncRNA genes: 53.0% (849/1,600) in human, 48.8% (418/855) in mouse, and 42.0% (210/499) in chicken (**Figure 2**). This percentage however should be considered as an estimate that will change with time as both miRNA and host genes (protein-coding and ncRNA genes) are still being annotated and added to database upgrades. Manual inspection of host genes revealed examples with doubtful annotation in regions with two or three overlapping genes, for which we contacted the source database (Ensembl) and solved ambiguous annotations case by case. Namely, it was observed that in cases where two ncRNA genes (miRNA and snoRNA) overlapped in the same region, the automatic annotation pipeline favored the longer RNA; for example, the record of snoRNA gene *SNORA36B* overwrote the record of the

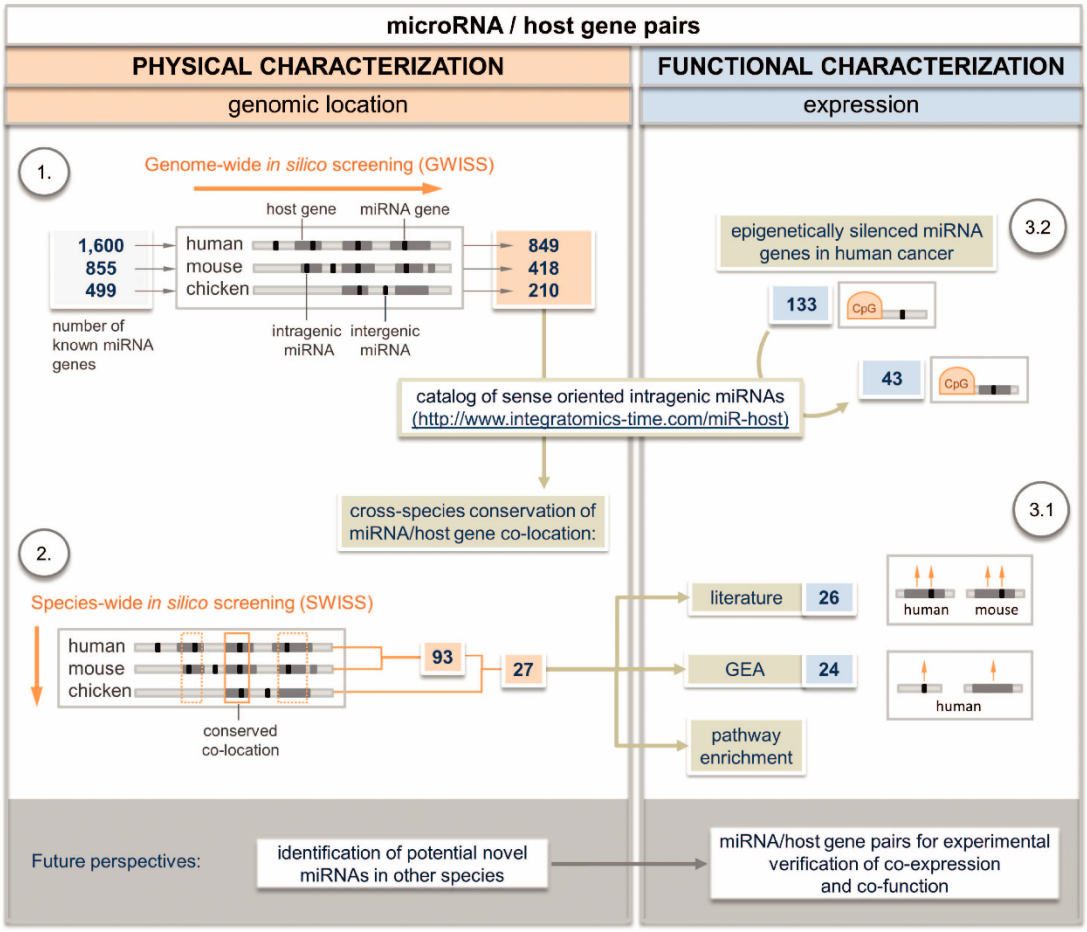


Figure 1. Workflow of the study. GEA – Gene Expression Atlas.
doi:10.1371/journal.pone.0065165.g001

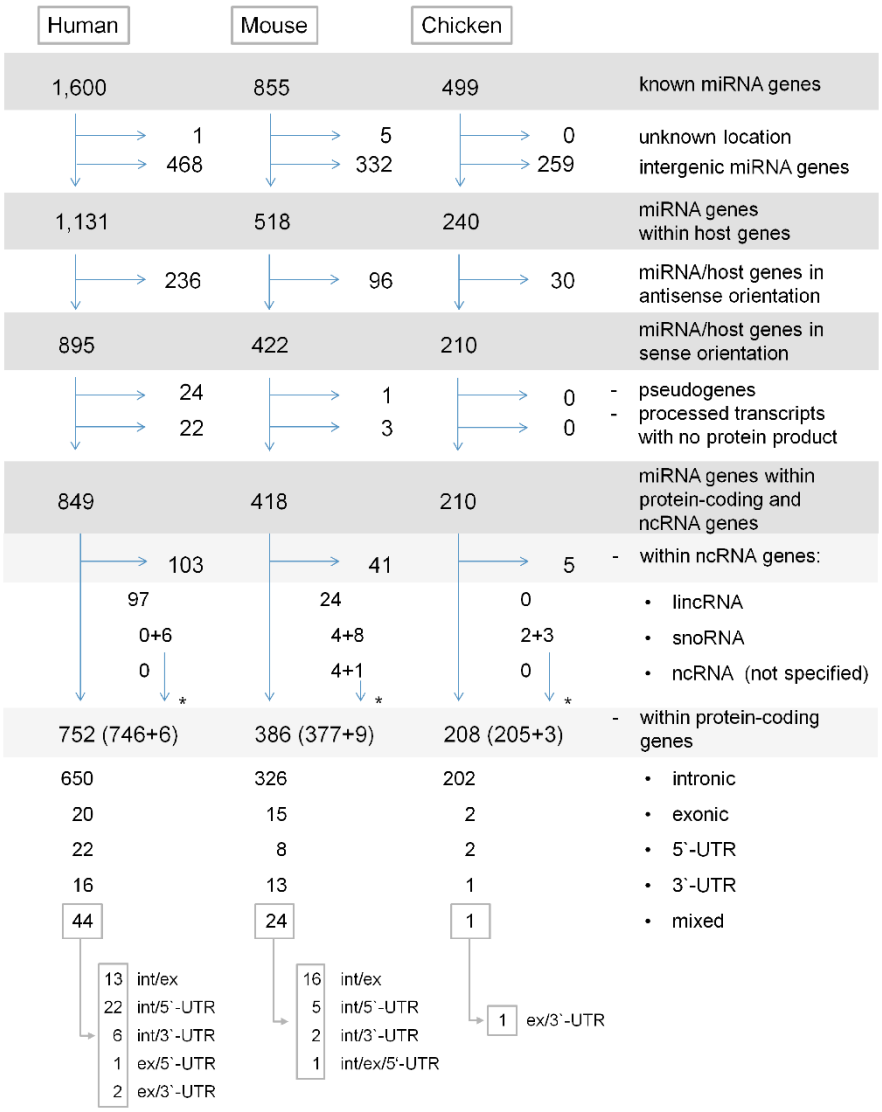


Figure 2. Diagram of genomic distribution of miRNA genes in human, mouse, and chicken. * - microRNA genes overlapping protein-coding and ncRNA genes; mixed - microRNA genes overlapping intron, exon or UTR, depending on overlapping host gene transcripts. For details see online table: http://www.integratomics-time.com/miR-host/catalog.
doi:10.1371/journal.pone.0065165.g002

overlapping miRNA gene *hsa-mir-664a*. One of the reasons for annotation error may also be the use of non-official and inconsistent nomenclature of genes. For example, a miRNA host transcript with a lincRNA biotype (ENSG00000253522) was merged between the Ensembl automatic pipeline and the Havana manual curation and was found to be given two names, *CTC-231O11.1* or *hsa-mir-146a*. Any updates of the catalog of miRNA/host gene pairs should therefore take into consideration the importance of nomenclature when searching for single or overlapping miRNA genes.

MicroRNA genes that do not share the same strand orientation as their host genes (*i.e.* are antisense-oriented) have been shown to have independent transcription mechanisms [34], whereas sense transcriptional orientation suggests that miRNA and host genes can be transcribed from shared promoters [1]. Additionally, it was found that a majority of predicted promoter regions of intronic miRNA genes (94.2%; 49/52) overlapped with their host gene promoters [35]. In addition to protein-coding host genes, ncRNA genes comprised snoRNAs, lincRNAs, and other unspecified ncRNAs (**Figure 2**). Long ncRNAs were found to also host

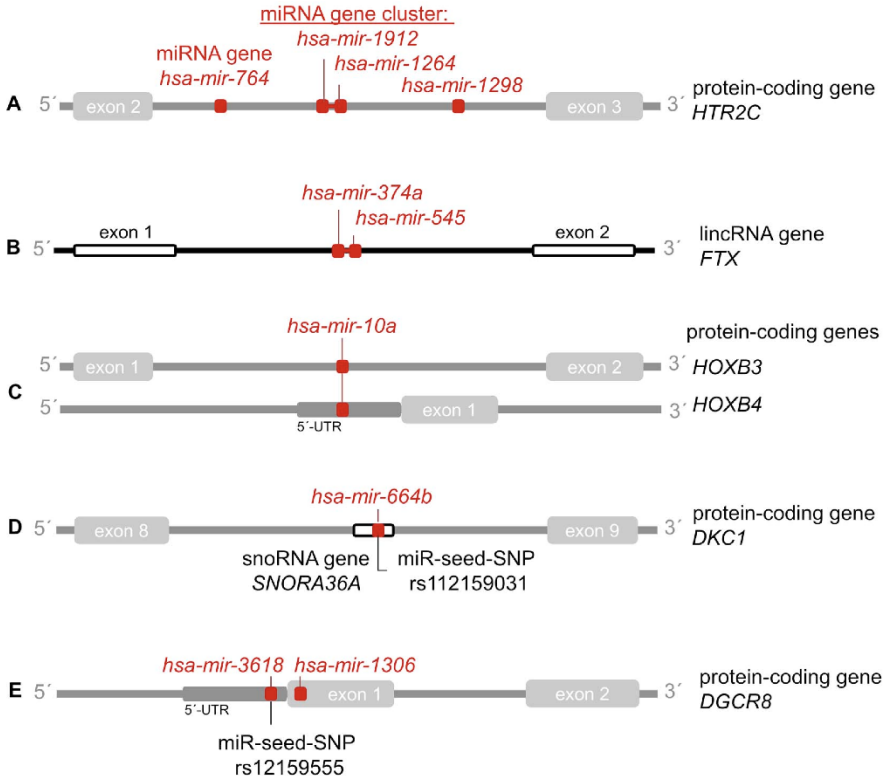


Figure 3. Examples of co-location of miRNA genes with protein-coding and ncRNA genes. A) Protein-coding gene *HTR2C* with four resident miRNA genes, two of which form a cluster. **B**) A miRNA gene cluster located within lincRNA gene *FTX*. **C**) MicroRNA gene *hsa-mir-10a* located within two overlapping protein-coding genes. **D**) Overlapping miRNA gene (*hsa-mir-664b*) comprising a miR-seed-SNP, and snoRNA gene (*SNORA36A*) residing within protein-coding *DKC1*. **E**) Gene *DGCR8*, associated with miRNA biogenesis, hosts two miRNA genes, one of which comprises a miR-seed-SNP.
doi:10.1371/journal.pone.0065165.g003

clusters of miRNA genes and therefore encode polycistronic primary transcripts that can yield several miRNAs; for example lincRNA *FTX* (FTX transcript, XIST regulator (non-protein coding)) comprises two miRNA genes: *hsa-mir-374a* and *hsa-mir-545* (**Figure 3B**). Because miRNA clusters can also overlap with a single protein-coding host gene (**Figure 3A**), the total number of host genes is lower than the number of intragenic miRNAs: we identified 687 protein-coding host genes in human (with 752 resident miRNA genes), 288 in mouse (with 386 miRNA genes), and 192 in chicken (with 208 miRNA genes). In all three species intragenic miRNA clusters most frequently comprise two miRNAs per host gene, as shown in the online table: http://www.integratomics-time.com/miR-host/catalog. The mouse host gene *Sfmbt2* (Scm-like with four mbt domains 2), located on MMU2, was found to comprise the largest number of resident miRNA genes (n = 70) belonging to the *mir-297*, *mir-466*, and *mir-467* gene families. Our study revealed that around one tenth of miRNA genes formed clusters in protein-coding host genes: 8.8% (141/1,600) in human, 14.5% (124/855) in mouse, and 8.2% (41/499) in chicken. It was also proposed that human miRNAs that share a host gene or are organized in clusters might also, due to clustering propensity, share a significant biological role [36,37]. Accordingly, miRNA genes that formed clusters were also found to be coordinately expressed with their host genes, which will be described in section 3.

For all three species (human, mouse, and chicken) we presented online genomic-views of intragenic miRNAs genes, connected to miRBase and host genes connected to Ensembl, with an outgoing link (http://www.integratomics-time.com/miR-host/GViews). The human genomic-view is presented in **Figure S1**. Intragenic miRNAs were found distributed among all chromosomes, however some, *e.g.* HSA14, HSA19, and HSAX, were found to comprise less intragenic miRNA genes compared to other chromosomes (**Figure S2**). In most cases miRNA genes resided within a single host gene. For example, human *hsa-mir-1307* gene overlaps with a single host gene *USMG5* (up-regulated during skeletal muscle growth 5 homolog (mouse)) gene. On the other hand, ten miRNA genes were found to overlap with two protein-coding host genes in human (http://www.integratomics-time.com/miR-host/human_coding). For example *hsa-mir-10a* overlapped with both, *HOXB3* (homeobox B3) and *HOXB4* (homeobox B4) (**Figure 3C**). Regarding the location of miRNA genes, we found that in accordance with previous publications [13,20,38] a majority of intragenic miRNA genes were located within introns of their protein-coding host genes: 86.4% (650/752) in human, 84.4% (326/386) in mouse, and 97.1% (202/208) in chicken (**Figure 2**).

Intronic miRNAs were also most frequently found to be coordinately expressed with their host genes among species, which will be further discussed in results section 2 and 3.

1.1. Co-location of miRNA with other ncRNA genes. Besides the half of miRNAs located within protein-coding genes, we found that around 4% were positioned within genes encoding for other ncRNA classes. These include lincRNAs, snoRNAs, or other ncRNAs: 6.4% (103/1,600) in human, 4.8% (41/855) in mouse, and 1% (5/499) in chicken, which can be accessed at http://www.integratomics-time.com/miR-host/catalog. Nomenclature conflicts of miRNA and ncRNA names may occur due to annotation difficulties: information merged from the Ensembl automatic pipeline and the Havana manual curation, which assign gene names according to miRBase and the HUGO Gene Nomenclature Committee. Six human miRNA genes were found located in both, protein-coding and ncRNA genes: *hsa-mir-600*, *-664a*, *-664b*, *-1248*, *-1291*, and *-3651* (**online table** http://www.integratomics-time.com/miR-host/human_table). MicroRNA gene *hsa-mir-664b*, its overlapping protein-coding host gene *DKC1* (dyskeratosis congenita 1, dyskerin) and snoRNA *SNORA36A* gene are shown in **Figure 3D**. Some miRNA genes were found to form clusters within hosting ncRNA genes: for example the miRNA gene cluster, comprising *hsa-mir-374a* and *hsa-mir-545*, is located within lincRNA gene *FTX* (**Figure 3B**). Additionally, lincRNAs have also been found to be the most frequent type of ncRNA host genes (97/103) as shown in the online table: http://www.integratomics-time.com/miR-host/human_table. In some cases the designated lincRNAs have been found to be the primary transcripts and not actual lincRNA genes, for example *MIR155HG* (also known as *BIC*) and *DLEU2* (deleted in lymphocytic leukemia 2 (non-protein coding), previously known as *LEU2*, are primary transcripts of their resident miRNA genes *hsa-mir-155* and *hsa-mir-15a/16-1*, respectively. Besides miRNAs themselves being regulators of gene expression participating in a wide regulatory network [1,3], their long ncRNA genes have likewise been found associated with human diseases. For example, lincRNA *H19* (H19, imprinted maternally expressed transcript (non-protein coding)), which hosts *hsa-mir-675*, was implicated in human tumor growth [39] in esophageal [40] and breast cancer [41], and different carcinomas and hepatic metastases [42]. Another study demonstrated that *H19* and *hsa-mir-675* were upregulated in human colon cancer cell lines and primary colorectal cancer tissues [43]. Long intergenic ncRNA *MEG3* (maternally expressed gene 3) could act as a tumor suppressor [44], while both the miRNA gene *hsa-mir-155* and *BIC* RNA (*MIR155HG*) from which it is processed, were overexpressed in human B-cell lymphomas [45]. Similarly, it was shown that the deletion of the 13q14 region, which encodes both, lincRNA *DLEU2* and its resident miRNA cluster *hsa-mir-15a/16-1*, led to chronic lymphocytic leukemia in both human [46] and mouse [47].

1.2. Genetic variability of intragenic miRNA genes. The intragenic miRNAs were also analyzed for genetic variability within the miRNA seed region (miR-seed-SNPs). By analyzing variation databases we found that 14.2% of intragenic miRNAs had polymorphic seed regions in human (121/849), 2.1% in mouse (9/418), and 1.4% in chicken (3/210) (**Table S1**). According to the NCBI database 18 out of 121 miRNA genes in human and two murine miRNA genes have not yet had validated miRNA seed polymorphisms. The actual proportion of polymorphic miRNA genes cannot yet be determined because miRNAs and polymorphisms, most of which are experimentally unvalidated, are still being discovered and added to the databases. That is why the results from previous studies tend to differ: Saunders *et al.* [48] found that less than 1% (3/474) of human miRNA genes miR-seed-SNPs, whereas in our previous study, Zorc *et al.* [31], we reported that 5.9% of miRNA genes comprised miR-seed-SNPs. Polymorphic miRNA genes are an interesting feature to include in the host gene analysis because they have previously been found to have functional associations. For example, we found a link between two independent studies: human *MYH7B* gene (myosin, heavy chain 7B, cardiac muscle, beta) hosts *hsa-mir-499a*, a miRNA upregulated in human and murine cardiac hypertrophy and cardiomyopathy [49], which comprises miR-seed-SNP rs3746444 linked with increased risk of dilated cardiomyopathy [50]. A similar overlap was demonstrated previously comprising a mouse miRNA gene *mmu-mir-717*, a miR-seed-SNP identified in the lean mouse strain 129/Sv, a body mass associated host gene *Gpc3* (glypican 3), as well as a growth associated quantitative trait locus (QTL) [51]. Our catalog provides the basis for a more targeted selection of SNPs and functional connections with the miRNA and host genes.

1.3. MicroRNA/host gene pairs in miRNA biogenesis and regulation. By considering the host gene's function our study revealed an interesting observation that miRNAs are also located within genes encoding for components of the miRNA processing machinery. There were four miRNAs in human located within genes encoding for components of miRNA biogenesis: *DGCR8*, *DICER1*, and *SND1* (**Figure 4**). Similarly, five miRNA genes in mouse were located within *Cnot3*, *Dgcr8*, *Eif4e*, *Tnrc6b*, and *Xpo5* (**Figure S3**). Two miRNA genes (*hsa-mir-1306* and *hsa-mir-3618*) reside within gene *DGCR8*, whose protein product is essential for miRNA biogenesis (**Figure 3E**). Human miRNA gene *hsa-mir-3173*, was found located within an intron of host gene *DICER1*, encoding a protein that functions as a ribonuclease required to produce active RNAs. MicroRNA gene *hsa-mir-593* resided within an intron of *SND1* (staphylococcal nuclease and tudor domain containing 1), a component of RISC. By performing a target gene analysis we found that each of the residing miRNAs was predicted to target genes which also host other miRNA genes (**Figure 4**). According to previous experimental studies, *DICER1* was found targeted by nine miRNAs: *hsa-let-7a*, *-7b*, *-7c*, and *-7d*, *hsa-mir-18a*, *-103*, *-107*, *-374a*, and *-519a* [52 55]. Additionally, *hsa-mir-3618* and *hsa-mir-593* were found to comprise a miR-seed-SNPs (rs12159555 and rs73721294, respectively), however both SNPs still need to be validated. Where miRNA molecule targets a gene from a miRNA processing machinery this could indicate a negative regulatory loop and a multi-layer regulatory cross-point, possibly associated with the disrupted processing of miRNAs. Also, alterations in gene regulation could have pathologic implications, as all three miRNA silencing machinery genes have previously been linked to certain diseases: *DICER1* with cancer [11,56], *DGCR8* with DiGeorge syndrome [57], and *SND1* was found frequently up-regulated in human and mouse cancers, as well as in aberrant crypt foci [58]. To summarize, this miRNA-related genomic cross-points consists of: 1) intragenic miRNAs, 2) miRNA gene polymorphisms, 3) miRNA host genes encoding for proteins involved in miRNA biogenesis and silencing, 4) miRNA target sites within miRNA host genes, and 5) their resident miRNAs targeting other host genes. Polymorphisms and aberrations in this miRNA-related and disease-associated genomic cross-point could therefore have a significant effect on phenotypic variation, including disease susceptibility and deserve further analysis.

2. Cross-species Conservation of miRNA/host Gene Co-location

In order to determine how many intragenic miRNAs are located within the same host genes in human, mouse, and chicken,

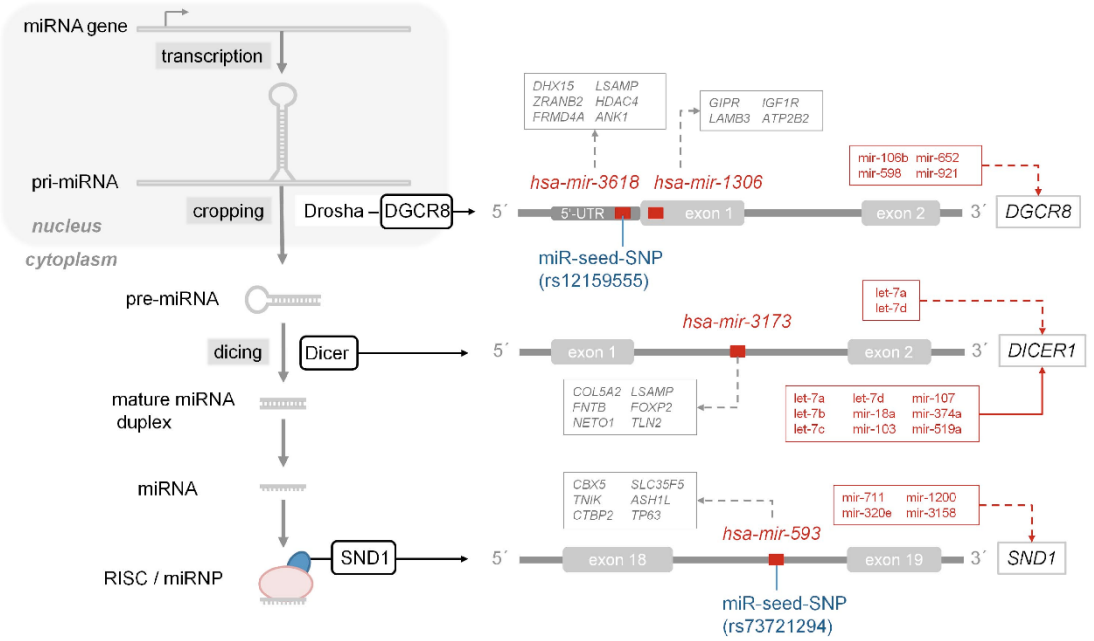


Figure 4. Cross talk of miRNA-related genomic elements. Overlapping miRNA genes (*hsa-mir-3618* and *mir-1306*, *mir-3173*, and *mir-593*), miRNA polymorphisms (miR-seed-SNPs (rs12159555 and rs73721294), host genes encoding for miRNA processing machinery components (*DGCR8*, *DICER1*, and *SND1*), miRNA target sites within host genes, and miRNAs targeting other host genes. Arrow with solid line: experimentally validated miRNA targets; arrow with dashed line: predicted miRNA targets.
doi:10.1371/journal.pone.0065165.g004

we performed a species-wide *in silico* screening (SWISS) of their co-location. We found that 27 miRNA genes had conserved co-location within the same 23 host genes in all three species (**Table 1**, **Figure S4**). In some cases the host genes (*NFYC*, *SMC4*, and *C9orf3*) encompassed more than one resident miRNA, explaining the co-location of the 27 miRNAs within 23 host genes. Moreover, additional 93 miRNA/host gene pairs were found to have conserved co-location in human and mouse (**online table:** http://www.integratomics-time.com/miR-host/species_cons). Most of the intragenic miRNAs were found to reside within introns of their host genes (25/27) (**Table 1**). MicroRNA/host gene pairs with conserved co-location offer a foundation for structural annotation of novel miRNA genes in other species. Using this approach, we proposed a novel miRNA gene in chicken (*mir-3064*) based on its pre-miRNA region that was found conserved in human and mouse (**Figure S5**). Similarly, 15 potential miRNA genes in human have been suggested by comparing the annotated murine miRNA genes with the human genome. Sequences of potential human miRNAs were examined for small RNA expression data using the UCSC database. Four of the human sequences (complementary to mouse *mmu-mir-677*, *-1839*, *-1897*, and *-1949*) had available expression data (**Figure S5**), which further confirms that these sequences encode miRNAs. The proposed novel miRNA genes present candidates for further experimental validation, annotation and expression analysis. In this manuscript the proposed miRNAs (one in chicken and 15 in human) have been given temporary names and will be submitted to the miRBase upon acceptance of this manuscript by the peer review process.

3. Coordinated Expression and Functional Association of miRNA/host Gene Pairs

To find out whether miRNA/host gene pairs with conserved cross-species co-location are also co-expressed, we integrated experimental data from two different sources: published studies that experimentally confirmed miRNA/host gene co-expression and databases providing gene expression data for miRNA and host genes separately.

3.1. Co-expression of miRNA/host gene pairs with conserved cross-species co-location. For the first step in determining if the 27 miRNA/host gene pairs with conserved cross-species co-location (in human, mouse, and chicken) (**Table 1**) are also co-expressed, we analyzed data from 28 studies that experimentally confirmed their coordinated expression [19,26,27,59–71]. The data integration revealed that most miRNA/host gene pairs (26/27) have previously been found to be coordinately expressed (either both up- or down-regulated) in human and/or mouse (**online table**: http://www.integratomics-time.com/miR-host/co-exp). Co-expression of only one miRNA/host gene pair, *mir-1306/DGCR8*, has not yet been experimentally demonstrated. We also found opposing results regarding the expression of two miRNA/host gene pairs, murine *mmu-mir-103/Pank3* and *mmu-mir-107/Pank1* – these have previously been demonstrated to have coordinate [71] as well as anti-correlative (or discordant) expression patterns [72]. Out of the 26 miRNA/host gene pairs with coordinated expression, 11 have been found to be coordinately expressed in both, human and mouse [19,27,59,61–64,67–69,71,73–79]: *mir-103/PANK3*, *mir-107/PANK1*, *mir-126/EGFL7*, *mir-128-1/R3HDM1*, *mir-140/WWP2*,

Table 1. Twenty-seven miRNA/host gene pairs with conserved co-location in human, mouse, and chicken.

Human			Mouse			Chicken		
miRNA gene	**location within host gene**	**host gene (synonym)**	**miRNA gene**	**location within host gene**	**host gene (synonym)**	**miRNA gene**	**location within host gene**	**host gene (synonym)**
let-7g	intron 2, 3	*WDR82*	*let-7g*	intron 2	*Wdr82*	*let-7g*	intron 2	*WDR82_CHICK*
mir-101-2	intron 4, 7, 8	*RCL1*	*mir-101b*	intron 8	*Rcl1*	*mir-101*	intron 8	*RCL1*
mir-103a-1	intron 2, 5	*PANK3*	*mir-103-1*	intron 5	*Pank3*	*mir-103-1*	intron 4, 5	*PANK3*
mir-107	intron 4, 5	*PANK1*	*mir-107*	intron 5	*Pank1*	*mir-107*	intron 5	*PANK1*
mir-126	intron 5–7	*EGFL7*	*mir-126*	intron 1, 6, 7, exon 4	*Egfl7*	*mir-126*	intron 7	*EGFL7*
mir-128-1	intron 7, 15, 18	*R3HDM1*	*mir-128-1*	intron 19	*R3hdm1*	*mir-128-1*	intron 18	*R3HDM1*
mir-128-2	intron 3, 6, 17, 18	*ARPP21*	*mir-128-2*	intron 14–17, 19	*Arpp21*	*mir-128-2*	intron 16	*ARPP21*
mir-1306	5′-UTR	*DGCR8*	*mir-1306*	exon 1–3	*Dgcr8*	*mir-1306*	5′-UTR	*DGCR8*
mir-140	intron 3, 6, 7, 9, 14, 16	*WWP2*	*mir-140*	intron 16	*Wwp2*	*mir-140*	intron 16, 17	*WWP2*
mir-15b	intron 1–5	*SMC4*	*mir-15b*	intron 5	*Smc4*	*mir-15b*	intron 4	*SMC4*
mir-16-2	intron 1–5	*SMC4*	*mir-16-2*	intron 5	*Smc4*	*mir-16-2*	intron 4	*SMC4*
mir-190a	intron 22, 27, 51, 53	*TLN2*	*mir-190a*	intron 53	*Tln2*	*mir-190*	intron 54	*TLN2*
mir-211	intron 4–7	*TRPM1*	*mir-211*	intron 1, 2, 4–6	*Trpm1*	*mir-204-2* (orthologue)	intron 5	*TRPM1*
mir-218-1	intron 14–16	*SLIT2*	*mir-218-1*	intron 1, 8, 14–16	*Slit2*	*mir-218-1*	intron 13, 15	*Q90XG3_CHICK* (orthologue)
mir-218-2	intron 4, 14	*SLIT3*	*mir-218-2*	intron 14	*Slit3*	*mir-218-2*	intron 1, 9	*Q90Z43_CHICK* (orthologue)
mir-23b	intron 4–6, 14, 15	*C9orf3*	*mir-23b*	intron 3, 15	*2010111I01Rik* (orthologue)	*mir-23b*	intron 15	*C9orf3*
mir-24-1	intron 4–6, 14,15, exon 7	*C9orf3*	*mir-24-1*	intron 3, 15	*2010111I01Rik* (orthologue)	*mir-24*	3′-UTR	*C9orf3*
mir-26a-1	intron 2, 4, 5	*CTDSPL*	*mir-26a-1*	intron 4, 5	*Ctdspl*	*mir-26a*	intron 5	*CTDSL_CHICK*
mir-27b	intron 4–6, 14, 15	*C9orf3*	*mir-27b*	intron 3, 15	*2010111I01Rik* (orthologue)	*mir-27b*	intron 15	*C9orf3*
mir-301a	intron 1	*SKA2*	*mir-301a*	intron 1	*SKA2 (Fam33a)*	*mir-301b* (orthologue)	intron 1	*SKA2*
mir-30c-1	intron 1–6, 10	*NFYC*	*mir-30c-1*	intron 3–5	*Nfyc*	*mir-30c-1*	intron 4	*NFYC*
mir-30e	intron 1–6, 10	*NFYC*	*mir-30e*	intron 3–5	*Nfyc*	*mir-30e*	intron 4	*NFYC*
mir-32	intron 8, 12, 14	*TMEM245 (C9orf5)*	*mir-32*	intron 8, 12, 14	*TMEM245*	*mir-32*	intron 23	*Tmem245 (C9orf5)*
mir-33a	intron 1, 2, 9, 10, 16, 18, 19	*SREBF2*	*mir-33*	intron 16	*Srebf2*	*mir-33 (mir-33-1)*	intron 13	*SREBF2*
mir-455	intron 5–7, 10	*COL27A1*	*mir-455*	intron 7, 10	*Col27a1*	*mir-455*	intron 18	*COL27A1*
mir-499a	intron 20	*MYH7B*	*mir-499*	intron 19	*Myh7b*	*mir-499*	intron 18, 19	*MYH7B*
mir-7-1	intron 1, 3, 15, 16	*HNRNPK*	*mir-7a-1*	intron 2, 3, 5, 7, 11, 14, 15, 17	*Hnrnpk*	*mir-7-1*	intron 15	*HNRNPK*

Host gene names: *ARPP21*: cAMP-regulated phosphoprotein, 21kDa; *COL27A1*: collagen, type XXVII, alpha 1; *CTDSPL*: CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like; *C9orf3*: chromosome 9 open reading frame 3; *C9orf5*: chromosome 9 open reading frame 5; *EGFL7*: EGF-like-domain, multiple 7; *Fam33a*: spindle and kinetochore associated complex subunit 2; *DGCR8*: DiGeorge syndrome critical region gene 8; *HNRNPK*: heterogeneous nuclear ribonucleoprotein K; *MYH7B*: myosin, heavy chain 7B, cardiac muscle, beta; *NFYC*: nuclear transcription factor Y, gamma; *PANK1*: pantothenate kinase 1; *PANK3*: pantothenate kinase 3; *R3HDM1*: R3H domain containing 1; *RCL1*: RNA terminal phosphate cyclase-like 1; *SKA2*: spindle and kinetochore associated complex subunit 2; *SLIT2*: slit homolog 2 (Drosophila); *SLIT3*: slit homolog 3 (Drosophila); *SMC4*: structural maintenance of chromosomes 4; *SREBF2*: sterol regulatory element binding transcription factor 2; *TLN2*: talin 2; *TMEM245*: transmembrane protein 245; *TRPM1*: transient receptor potential cation channel, subfamily M, member 1; *WDR82*: WD repeat domain 82; *WWP2*: WW domain containing E3 ubiquitin protein ligase 2.

doi:10.1371/journal.pone.0065165.t001

mir-211/TRPM1, *mir-218-1/SLIT2*, *mir-218-2/SLIT3*, *mir-27b/C9orf3*, *mir-33/SREBF2*, and *mir-499/MYH7B*. Moreover, two miRNA/host gene pairs have been found to have expression patterns associated with the same phenotype in both species: *mir-499/MYH7B* with heart development [79] and *mir-33/SREBF2* with cholesterol homeostasis [74,75,77]. Several independent studies in chicken have similarly indicated that *gga-mir-33* and its host gene *SREBF2* are highly expressed in the liver, suggesting involvement in expression upregulation of genes related to cholesterol biosynthesis [80,81].

To further test the hypothesis that miRNA/host gene pairs with cross-species conserved co-location are coordinately expressed, we integrated expression data for 27 miRNA and their host genes using the GEA database. By comparing the gene expression data, we found that 24 miRNAs and their host genes had matching expression patterns in at least one disease (either over- or under-expression) (**Table S2**). Because of the same expression patterns and similar functions, the miRNA/host gene pairs are likely to be controlled by the same regulatory mechanisms. The miRNA/host gene pairs with conserved cross-species co-location, co-expression, and potential co-regulation provide a starting point for researchers investigating the involvement of intragenic miRNAs with disease development or control of production traits.

To better determine the role of the miRNA host genes from the pairs with conserved cross-species co-location, we performed a pathway enrichment analysis, using the IPA software [33]. Pathway analysis performed on the 23 host genes (**Table 1**) revealed networks associated with cancer, dermatological diseases and conditions, and hematological diseases (**Figure S6A**). Most significant biological functions included cancer, in addition to reproductive system diseases and infectious diseases. A molecular network diagram was constructed involving 14 miRNA host genes (*CTDSPL*, *C9orf3*, *COL27A1*, *EGFL7*, *HNRNPK*, *NFYC*, *PANK1*, *SLIT2*, *SLIT3*, *SMC4*, *SREBF2*, *TLN2*, *TRPM1*, and *WWP2*) which were found related to cancer, dermatologic and hematological diseases (**Figure S6B**). Within this network, several hubs were found encoding transcription factors, the largest two of which were *MYC* (v-myc myelocytomatosis viral oncogene homolog (avian)) and *TP53* (tumor protein p53), previously also linked with regulation of miRNA gene expression [82,83].

3.2. Epigenetically silenced miRNA genes located within host genes. Silenced expression of co-located miRNA and host genes might also be a subject of epigenetic regulation [27]. Namely, the proximal CpG islands located within their promoter or 5′UTR regions could epigenetically silence gene expression through DNA hypermethylation. In a recent study, 81.2% of protein-coding genes harboring miRNA genes in their 5′-end have been found located 500 bp downstream of CpG islands [84]. By performing a cross-section of 133 miRNA genes that have previously been found to be epigenetically regulated in cancer [85], we found that 30 are located within protein-coding, and 13 within ncRNA host genes, *i.e.* genes encoding for lincRNAs (**Figure 1**, **Table 2**). However, in order to determine the exact proportion of epigenetically regulated miRNA/host gene pairs a systematic genome-wide epigenetic analysis should be performed. Previous studies revealed that five miRNA genes as well as their host genes (*hsa-mir-10a/HOXB4*, *hsa-mir-126/EGFL7*, *hsa-mir-152/COPZ2*, *hsa-mir-191/DALRD3*, and *hsa-mir-342/EVL*) were found to be epigenetically downregulated, either by histone modification and/or CpG island hypermethylation in the promoter region in cancer cells [27,86–89] (**Table 2**). Additionally, several host genes have, independently of miRNA studies, been found to be silenced through DNA hypermethylation: *DALRD3* [88], *HOXA9* [90–92], *HOXB4* [93], *HOXB7* [94], *HOXC4* [95], *HOXD3* [96], *HTR2C* [97], and *IGF2* [98]. The identified epigenetically regulated intragenic miRNA genes can now be analyzed together with their host genes in order to study their potential epigenetic co-regulation. We found that around half (20/43) of the epigenetically silenced miRNA genes were located within the 5′-UTR or in the first intron or exon of their host genes, suggesting the possibility of shared promoter regions that comprise CpG islands. Further studies on epigenetic regulation of miRNA/genes may reveal novel approaches for prevention or treatment of human cancer.

4. MicroRNA/host Gene Pairs – Potential for Misattribution of Phenotype?

In our study we demonstrated that a very large proportion of miRNAs are located within the host genes (**Figure 2**) in human (1,131/1,600), mouse (518/855), and chicken (240/499) and that miRNA/host gene pairs have important conservation and co-expression issues. Our study can be used as a platform for researchers to re-examine questions related to earlier or planned studies correlating genetic variation or modification of the miRNA/host gene pairs with diseases or trait control. Namely, it is prudent to ask if some of the gene variation-phenotype association studies targeted at the miRNA host genes, spontaneous, radiation or chemically induced mutations, knockout and overexpression models need reinterpretation to take into account collateral effects on miRNAs. MicroRNA genes harbored within another host gene, as shown by many examples in our study, may have several target genes and functions unrelated to their host genes. The host gene mutations or modifications may also collaterally affect the level, time or tissue specificity of miRNA expression thereby leading to several pleiotropic effects in the phenotype that could not be causally ascribed to the host gene only. Many types of spontaneous and induced mutations within the host gene locus (*e.g.* promoter, splicing mutations, or mRNA stability mutations) may affect the transcript quantity, temporal and/or spatial expression pattern of hosted miRNA.

In addition to aforementioned effects, transgenic overexpression and knockout host gene models may alter hosted miRNA function through exogenous sequences left in the locus such as selection marker genes (*e.g.* neomycin resistance, *NeoR*), plasmid vector and other sequences (*e.g.* strong phosphoglycerate kinase (*pgk*) gene promotor). We note that among the knockout mice of relevance in **Tables 1A and B**, most models retained the *NeoR* marker and also other exogenous sequences that can potentially affect expression and function of hosted miRNA gene in addition to the target host gene itself. Many targeting constructs are designed to delete large portions of the target gene in order to ensure loss of function of the host locus. The weakness of this strategy is that some of the deleted sequence may contain miRNAs or regulatory sequences affecting neighboring genes. Significantly for this discussion, inadvertent deletion of *mmu-mir-126* has led to the misattribution of phenotype - angiogenesis defects previously reported in a knockout of the *Egfl7* locus were subsequently shown to have arisen due to deletion of the *mmu-mir-126* [99].

A degree of common sense can be applied to assessing the level of confidence attributed to specific phenotypes of the miRNA/host gene pairs. Where the phenotype is consistent with what was expected from knowledge of gene expression and biochemistry for the host gene and hosted miRNA gene, one can be reasonably comfortable in attributing a phenotype to the host target gene function. However, where the phenotype is unexpected, or where multiple genotype-phenotype or multiple gene modification models show disparate effects, then one is justified in being more cautious and to proceed by further experimentation to differentiate the host gene from hosted miRNA gene phenotypic effects. In the

Table 2. Host genes for epigenetically silenced miRNA genes in cancer.

miRNA gene	location within host gene	host gene	study describing epigenetic regulation of host gene
Protein-coding host genes			
hsa-let-7a-3	exon 5	*RP4-695O20__B.10*	/
hsa-mir-107	intron 4–6	*PANK1*	/
hsa-mir-10a	intron 1, 5′-UTR	*HOXB3*	/
	5′-UTR	*HOXB4*	Zhend et al., 2009 [93], Shen et al., 2012 [89]*
hsa-mir-10b	intron 1	*HOXD3*	Kron et al., 2010 [96]
hsa-mir-1-1	intron 1, 2	*C20orf166*	/
hsa-mir-126	intron 5–7	*EGFL7*	Saito et al., 2009 [27]*
hsa-mir-139	intron 1–3	*PDE2A*	/
hsa-mir-140	intron 3, 6, 7, 9, 13, 14, 16	*WWP2*	/
hsa-mir-148b	intron 1, 2	*COPZ1*	/
hsa-mir-152	intron 1, 2	*COPZ2*	Tsuruta et al., 2011 [87]*
hsa-mir-188	intron 3	*CLCN5*	/
hsa-mir-191	intron 1	*DALRD3*	He et al., 2011 [88]*
hsa-mir-196a-1	intron 1	*HOXB7*	Bennett et al., 2009 [94]
hsa-mir-196b	exon 1–3	*HOXA9*	Bandyopadhyay et al., 2012 [92], Hwang et al., 2011 [90], Wu et al., 2007 [91]
	intron 1	*RP1-170O19.20*	/
hsa-mir-198	3′-UTR	*FSTL1*	/
hsa-mir-204	intron 3–7	*TRPM3*	/
hsa-mir-23b	intron 4–6, 14, 15	*C9orf3*	/
hsa-mir-24-1	intron 4, 5, 14, 15, exon 7	*C9orf3*	/
hsa-mir-25	intron 2, 4, 8, 12, 13	*MCM7*	/
hsa-mir-27b	intron 4–6, 14, 15	*C9orf3*	/
hsa-mir-342	intron 2–4	*EVL*	Grady et al., 2008 [86]*
hsa-mir-425	intron 1	*DALRD3*	He et al., 2011 [88]
hsa-mir-448	intron 4, 5	*HTR2C*	Anderton et al., 2008 [97]
hsa-mir-483	intron 2, 3, 5	*IGF2*	Dejeux et al., 2009 [98]
	intron 5	*INS-IGF2*	/
hsa-mir-548c-1	intron 14–16	*ATAD2*	/
hsa-mir-570	intron 3	*MUC20*	/
hsa-mir-582	intron 1–3	*PDE4D*	/
hsa-mir-615	intron 1	*HOXC5*	/
	intron 1	*HOXC4*	Issa, 2009 [95]
hsa-mir-744	intron 1–5	*MAP2K4*	/
hsa-mir-9-1	intron 1, 2	*C1orf61*	/
Non-coding RNA genes *(gene type according to Ensembl)*			
hsa-mir-124-1	exon 1, 3, 4	*LINC00599* (lincRNA)	/
hsa-mir-124-2	intron 1	*RP11-32K4.2* (lincRNA)	/
hsa-mir-137	exon 3	*MIR137HG* (lincRNA)	/
hsa-mir-17	exon 3, intron 3	*MIR17HG* (lincRNA)	/
hsa-mir-193b	intron 1	*RP11-65J21.3* (lincRNA)	/
hsa-mir-205	exon 2, 4, intron 2, 3	*MIR205HG* (lincRNA)	/
hsa-mir-20a	exon 3, intron 3	*MIR17HG* (lincRNA)	/
hsa-mir-30a	intron 3	*LINC00472* (lincRNA)	/
hsa-mir-31	intron 1	*MIR31HG* (lincRNA)	/
hsa-mir-370	intron 5	*MEG8* (lincRNA)	/
hsa-mir-9-2	exon 3, 4, intron 2, 3	*LINC00461* (lincRNA)	/
hsa-mir-9-3	intron 1	*CTD-2335A18.1* (lincRNA)	/

Table 2. Cont.

miRNA gene	location within host gene	host gene	study describing epigenetic regulation of host gene
hsa-mir-99a	intron 1, 3, 5, 6	*LINC00478* (lincRNA)	/

/− host gene not found to be regulated by DNA methylation in references.
*- studies describing epigenetically regulated host gene and resident miRNA gene.
Host gene names: *ATAD2*: ATPase family, AAA domain containing 2; *C1orf61*: chromosome 1 open reading frame 61; *C20orf166*: chromosome 20 open reading frame 166; *CLCN5*: chloride channel, voltage-sensitive 5; *COPZ1*: coatomer protein complex, subunit zeta 1; *COPZ2*: coatomer protein complex, subunit zeta 2; *DALRD3*: DALR anticodon binding domain containing 3; *EVL*: Enah/Vasp-like; *FSTL1*: follistatin-like 1; *HOXA9*: homeobox A9; *HOXB4*: homeobox B4; *HOXB7*: homeobox B7; *HOXC4*: homeobox C4; *HOXC5*: homeobox C5; *HOXD3*: homeobox D3; *HTR2C*: 5-hydroxytryptamine (serotonin) receptor 2C, G protein-coupled; *IGF2*: insulin-like growth factor 2 (somatomedin A); *INS-IGF2*: INS-IGF2 readthrough; *LINC00461*: long intergenic non-protein coding RNA 461; *LINC00478*: long intergenic non-protein coding RNA 478; *LINC00472*: long intergenic non-protein coding RNA 472; *MAP2K4*: mitogen-activated protein kinase kinase 4; *MCM7*: minichromosome maintenance complex component 7; *MIR137HG*: mir-137 host gene (non-protein coding); *MEG8*: maternally expressed 8 (non-protein coding); *MIR17HG*: mir-17–92 cluster host gene (non-protein coding); *MIR205HG*: mir-205 host gene (non-protein coding); *MIR31HG*: mir-31 host gene (non-protein coding); *MUC20*: mucin 20, cell surface associated; *PDE2A*: phosphodiesterase 2A, cGMP-stimulated; *PDE4D*: phosphodiesterase 4D, cAMP-specific; *TRPM3*: transient receptor potential cation channel, subfamily M, member 3.
doi:10.1371/journal.pone.0065165.t002

future gene modification experiments many concerns raised above can be minimized by using recent technology of Zinc finger [100] and Tal nucleases [101]. These methods generate minimal targeted modifications (*i.e.* point mutation generating premature stop codon) and do not leave exogenous sequence in the genome thereby providing excellent transgenic *in vitro* and *in vivo* models for miRNA/host gene pairs studies.

Our web site (http://www.integratomics-time.com/miR-host/) provides an efficient tool to check which host genes contain miRNAs while other tables list important functional and literature information to aid researchers in re-examining potential misattribution of phenotype previously ascribed to host genes or hosted miRNA genes only.

5. Future Perspective

Our assembled and supplemented catalog of miRNA/host gene pairs available via the web application will provide researchers with a data mining tool for investigating miRNA/host gene pair involvement of their coordinated expression, shared regulation, and function in diseases: 1) structural annotation - miRNA/host gene pairs with conserved cross-species co-location in the examined species present candidate genes for future annotation in other species. 2) Functional annotation - miRNA/host gene pairs with matching expression patterns integrated from databases are high priority candidates for experimental validation of their potential co-expression and co-regulation. 3) MicroRNAs overlapping with protein-coding and other ncRNA host genes (lincRNA and snoRNA) present candidates for evaluating molecular mechanisms underlying previously shown functional links. 4) MicroRNAs residing within genes encoding for miRNA silencing machinery present important miRNA-related regulatory cross-talk needing additional mechanistic experimentation to elucidate targeting interplay in which miRNAs target genes for miRNA processing components and, in a feedback loop, influences the production of miRNAs. 5) Identification and validation of polymorphisms located within miRNA genes, their host genes, and genes encoding for and processing machinery components may also reveal whether they contribute to phenotypic variation, including disease susceptibility. 6) Epigenetic silencing of both, miRNA and their host genes, offers insights into their shared regulation and their re-expression may be used to contribute to the effects of epigenetic therapy. The assembled epigenetically regulated intragenic miRNAs represent candidate genes for the study of miRNA/host gene pair epigenetic co-regulation. 7) Our web site also provides an efficient tool to identify certain miRNA/host gene pairs where previous studies show inconsistencies of the effects of natural or induced mutations on the phenotype. We point to examples where such phenotype misinterpretations could arise due to attribution collateral effects of such mutations on hosted miRNAs. Our catalog can therefore direct researchers to critically examine designs and interpretation of such miRNA/host gene cases.

Conclusion

In conclusion, the assembled catalog is, to our knowledge, the most comprehensive integrated assembly of intragenic miRNAs and their host genes in human, mouse, and chicken. The systematically integrated physical (genomic location and cross-species conserved co-location) and functional characterization (co-expression data) of miRNA/host gene pairs provides a starting point for researchers investigating involvement of intragenic miRNAs with human and animal health, and animal production traits. Using this approach we found that miRNA/host gene pairs with cross-species conserved co-location are very likely to be co-expressed. The expanding field of miRNA research requires a consideration of interplay of interconnecting regulatory mechanisms and their function into an intricate network, in which miRNA genes and their co-expressed host genes also play a role.

Supporting Information

Figure S1 Print-screen of genomic view of intragenic miRNAs in human. Enlarged chromosome 22 showing *hsa-mir-1306* and its host gene *DGCR8* with databases linked through outgoing links.
(TIF)

Figure S2 Distribution of intragenic miRNA genes according to chromosome in A) human, B) mouse, and C) chicken.
(PDF)

Figure S3 MicroRNA genes located within genes encoding for the miRNA processing machinery in mouse.
(TIF)

Figure S4 Venn diagram of the number of miRNA/host gene pairs with cross-species conserved co-location.
(TIF)

Figure S5 Alignment of orthologous miRNA genes. A) Human (*hsa-mir-3064*) and mouse (*mmu-mir-3064*) miRNA genes matching the sequence in chicken. Mature miRNA regions are marked with a square. **B**) Murine miRNA genes (*mmu-mir-677*, *-686*, *-717*, *-763*, *-1839*, *-1893*, *-1896*, *-1897*, *-1898*, *-1902*, *-1907*, *-1949*, *-2139*, *-3059*, and *-5125*) aligned with human sequences. **C**) Fifteen potential human miRNA genes acquired based on alignment with 15 murine miRNA genes. **D**) Small RNA expression data for sequences matching the four potential new miRNA genes in human (*hsa-mir-677*, *-1839*, *-1897*, and *-1949*).
(DOC)

Figure S6 Network analysis of host genes from 27 conserved miRNA/host gene pairs, in human and mouse. A) Top network and biological functions associated miRNA host genes. **B**) Diagram of a top molecular network showing 14 miRNA host genes (gray-filled shapes) associated with cancer, dermatological diseases and conditions, and hematological diseases. White-filled shapes indicate connecting elements in between host genes in the network.
(DOCX)

Table S1 Intragenic miRNAs with polymorphic seed regions in human, mouse, and chicken.
(DOC)

Table S2 Dysregulation of expression in diseases associated with 27 human miRNA/host gene pairs with cross-species conserved co-location.
(DOCX)

Author Contributions

Conceived and designed the experiments: TK. Analyzed the data: IG MZ DJS GAC SH TK. Contributed reagents/materials/analysis tools: MZ GAC. Wrote the paper: IG SH TK. Reviewed and evaluated the article: GAC SH PD MK TK. Designed the software: MZ.

References

1. Bartel DP (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell 116: 281–297.
2. Sun G, Yan J, Noltner K, Feng J, Li H, et al. (2009) SNPs in human miRNA genes affect biogenesis and function. RNA 15: 1640–1651.
3. Kunej T, Godnic I, Horvat S, Zorc M, Calin GA (2012) Cross Talk Between MicroRNA and Coding Cancer Genes. Cancer J 18: 223–231.
4. Lee Y, Ahn C, Han J, Choi H, Kim J, et al. (2003) The nuclear RNase III Drosha initiates microRNA processing. Nature 425: 415–419.
5. Han J, Lee Y, Yeom KH, Kim YK, Jin H, et al. (2004) The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev 18: 3016–3027.
6. Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409: 363–366.
7. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell 123: 631–640.
8. Georges M, Coppieters W, Charlier C (2007) Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. Current Opinion in Genetics & Development 17: 166–176.
9. Ferdin J, Kunej T, Calin G (2010) Non-coding RNAs: Identification of Cancer-Associated microRNAs by Gene Profiling. Technology in Cancer Research & Treatment: 123–138.
10. Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, et al. (2010) Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. Cancer Res 70: 2789–2798.
11. Sand M, Gambichler T, Skrygan M, Sand D, Scola N, et al. (2010) Expression levels of the microRNA processing enzymes Drosha and dicer in epithelial skin cancer. Cancer Invest 28: 649–653.
12. Sand M, Skrygan M, Georgas D, Arenz C, Gambichler T, et al. (2011) Expression levels of the microRNA maturing microprocessor complex component DGCR8 and the RNA-induced silencing complex (RISC) components Argonaute-1, Argonaute-2, PACT, TARBP1, and TARBP2 in epithelial skin cancer. Mol Carcinog.
13. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A (2004) Identification of Mammalian microRNA Host Genes and Transcription Units. Genome Research 14: 1902–1910.
14. Ambros V (2004) The functions of animal microRNAs. Nature 431: 350–355.
15. Zhang Y, Zhang R, Su B (2009) Diversity and evolution of MicroRNA gene clusters. Sci China C Life Sci 52: 261–266.
16. Golan D, Levy C, Friedman B, Shomron N (2010) Biased hosting of intronic microRNA genes. Bioinformatics 26: 992–995.
17. Rearick D, Prakash A, McSweeny A, Shepard SS, Fedorova L, et al. (2011) Critical association of ncRNA with introns. Nucleic Acids Res 39: 2357–2366.
18. Lutter D, Marr C, Krumsiek J, Lang EW, Theis FJ (2010) Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. BMC Genomics 11: 224.
19. Baskerville S, Bartel D (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. Rna-a Publication of the Rna Society: 241–247.
20. Kim YK, Kim VN (2007) Processing of intronic microRNAs. EMBO J 26: 775–783.
21. Ruby JG, Jan CH, Bartel DP (2007) Intronic microRNA precursors that bypass Drosha processing. Nature 448: 83–86.
22. Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, et al. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol 5: R13.
23. He C, Li Z, Chen P, Huang H, Hurst LD, et al. (2012) Young intragenic miRNAs are less coexpressed with host genes than old ones: implications of miRNA-host gene coevolution. Nucleic Acids Res.
24. Saini HK, Enright AJ, Griffiths-Jones S (2008) Annotation of mammalian primary microRNAs. BMC Genomics 9: 564.
25. Ying SY, Lin SL (2005) Intronic microRNAs. Biochem Biophys Res Commun 326: 515–520.
26. Wang S, Aurora AB, Johnson BA, Qi X, McAnally J, et al. (2008) The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis. Dev Cell 15: 261–271.
27. Saito Y, Friedman JM, Chihara Y, Egger G, Chuang JC, et al. (2009) Epigenetic therapy upregulates the tumor suppressor microRNA-126 and its host gene EGFL7 in human cancer cells. Biochemical and Biophysical Research Communications 379: 726–731.
28. Tsang J, Zhu J, van Oudenaarden A (2007) MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. Mol Cell 26: 753–767.
29. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research 39: D152–D157.
30. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, et al. (2013) ENCODE Data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res 41: D56–63.
31. Zorc M, Jevsinek Skok D, Godnic I, Calin GA, Horvat S, et al. (2012) Catalog of MicroRNA Seed Polymorphisms in Vertebrates. PLoS One 7: e30737.
32. Hiard S, Charlier C, Coppieters W, Georges M, Baurain D (2010) Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. Nucleic Acids Research 38: D640–D651.
33. Ingenuity Pathway Analysis system.
34. Li S-C, Tang P, Lin W-C (2007) Intronic MicroRNA: Discovery and Biological Implications. DNA and Cell Biology 26: 195–207.
35. Wang G, Wang Y, Shen C, Huang YW, Huang K, et al. (2010) RNA polymerase II binding patterns reveal genomic regions involved in microRNA gene regulation. PLoS One 5: e13798.
36. Sikand K, Slane SD, Shukla GC (2009) Intrinsic expression of host genes and intronic miRNAs in prostate carcinoma cells. Cancer Cell Int 9: 21.
37. Zhang Y, Zhang R, Su B (2009) Diversity and evolution of MicroRNA gene clusters. Sci China C Life Sci 52: 261–266.
38. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34: D140–144.
39. Matouk IJ, DeGroot N, Mezan S, Ayesh S, Abu-lail R, et al. (2007) The H19 non-coding RNA is essential for human tumor growth. PLoS One 2: e845.
40. Hibi K, Nakamura H, Hirai A, Fujikake Y, Kasai Y, et al. (1996) Loss of H19 imprinting in esophageal cancer. Cancer Res 56: 480–482.
41. Berteaux N, Lottin S, Monté D, Pinte S, Quatannens B, et al. (2005) H19 mRNA-like noncoding RNA promotes breast cancer cell proliferation through positive control by E2F1. J Biol Chem 280: 29625–29636.
42. Fellig Y, Ariel I, Ohana P, Schachter P, Sinelnikov I, et al. (2005) H19 expression in hepatic metastases from a range of human carcinomas. J Clin Pathol 58: 1064–1068.
43. Tsang WP, Ng EK, Ng SS, Jin H, Yu J, et al. (2010) Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer. Carcinogenesis 31: 350–358.
44. Zhang X, Zhou Y, Mehta KR, Danila DC, Scolavino S, et al. (2003) A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. J Clin Endocrinol Metab 88: 5119–5126.

45. Eis PS, Tam W, Sun L, Chadburn A, Li Z, et al. (2005) Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci U S A 102: 3627–3632.
46. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, et al. (2002) Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 99: 15524–15529.
47. Klein U, Lia M, Crespo M, Siegel R, Shen Q, et al. (2010) The DLEU2/miR-15a/16–1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. Cancer Cell 17: 28–40.
48. Saunders MA, Liang H, Li W-H (2007) Human polymorphism at microRNAs and microRNA target sites. Proceedings of the National Academy of Sciences 104: 3300–3305.
49. Matkovich SJ, Hu Y, Eschenbacher WH, Dorn LE, Dorn GW (2012) Direct and indirect involvement of microRNA-499 in clinical and experimental cardiomyopathy. Circ Res 111: 521–531.
50. Zhou B, Rao L, Peng Y, Wang Y, Chen Y, et al. (2010) Common genetic polymorphisms in pre-microRNAs were associated with increased risk of dilated cardiomyopathy. Clin Chim Acta 411: 1287–1290.
51. Kunej T, Jevsinek Skok D, Horvat S, Dovc P, Jiang Z (2010) The Glypican 3-Hosted Murine Mir717 Gene: Sequence Conservation, Seed Region Polymorphisms and Putative Targets. International Journal of Biological Sciences: 769–772.
52. Tokumaru S, Suzuki M, Yamada H, Nagino M, Takahashi T (2008) let-7 regulates Dicer expression and constitutes a negative feedback loop. Carcinogenesis 29: 2073–2077.
53. Forman JJ, Legesse-Miller A, Coller HA (2008) A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. Proc Natl Acad Sci U S A 105: 14879–14884.
54. Tao J, Wu D, Li P, Xu B, Lu Q, et al. (2012) microRNA-18a, a member of the oncogenic miR-17–92 cluster, targets Dicer and suppresses cell proliferation in bladder cancer T24 cells. Mol Med Report 5: 167–172.
55. Huang Y, Chuang A, Hao H, Talbot C, Sen T, et al. (2011) Phospho-ΔNp63α is a key regulator of the cisplatin-induced microRNAome in cancer cells. Cell Death Differ 18: 1220–1230.
56. Yan M, Huang HY, Wang T, Wan Y, Cui SD, et al. (2011) Dysregulated Expression of Dicer and Drosha in Breast Cancer. Pathol Oncol Res.
57. Shiohama A, Sasaki T, Noda S, Minoshima S, Shimizu N (2003) Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. Biochem Biophys Res Commun 304: 184–190.
58. Tsuchiya N, Ochiai M, Nakashima K, Ubagai T, Sugimura T, et al. (2007) SND1, a component of RNA-induced silencing complex, is up-regulated in human colon cancers and implicated in early stage colon carcinogenesis. Cancer Res 67: 9568–9576.
59. Liang Y, Ridzon D, Wong L, Chen C (2007) Characterization of microRNA expression profiles in normal human tissues. BMC Genomics 8: 166.
60. Böhlig L, Friedrich M, Engeland K (2011) p53 activates the PANK1/miRNA-107 gene leading to downregulation of CDK6 and p130 cell cycle proteins. Nucleic Acids Res 39: 440–453.
61. Musiyenko A, Bitko V, Barik S (2008) Ectopic expression of miR-126*, an intronic product of the vascular endothelial EGF-like 7 gene, regulates prostein translation and invasiveness of prostate cancer LNCaP cells. J Mol Med (Berl) 86: 313–322.
62. Watanabe K, Emoto N, Hamano E, Sunohara M, Kawakami M, et al. (2012) Genome structure-based screening identified epigenetically silenced microRNA associated with invasiveness in non-small-cell lung cancer. Int J Cancer 130: 2580–2590.
63. Wang YP, Li KB (2009) Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data. BMC Genomics 10: 218.
64. Lages E, Guttin A, El Atifi M, Ramus C, Ipas H, et al. (2011) MicroRNA and target protein patterns reveal physiopathological features of glioma subtypes. PLoS One 6: e20600.
65. Donzelli S, Fontemaggi G, Fazi F, Di Agostino S, Padula F, et al. (2011) MicroRNA-128–2 targets the transcriptional repressor E2F5 enhancing mutant p53 gain of function. Cell Death Differ.
66. Beezhold K, Liu J, Kan H, Meighan T, Castranova V, et al. (2011) miR-190-mediated downregulation of PHLPP contributes to arsenic-induced Akt activation and carcinogenesis. Toxicol Sci 123: 411–420.
67. Tie J, Pan Y, Zhao L, Wu K, Liu J, et al. (2010) MiR-218 inhibits invasion and metastasis of gastric cancer by targeting the Robo1 receptor. PLoS Genet 6: e1000879.
68. Bak M, Silahtaroglu A, Møller M, Christensen M, Rath MF, et al. (2008) MicroRNA expression in the adult mouse central nervous system. RNA 14: 432–444.
69. Hackler L, Wan J, Swaroop A, Qian J, Zack DJ (2010) MicroRNA profile of the developing mouse retina. Invest Ophthalmol Vis Sci 51: 1823–1831.
70. Careccia S, Mainardi S, Pelosi A, Gurtner A, Diverio D, et al. (2009) A restricted signature of miRNAs distinguishes APL blasts from normal promyelocytes. Oncogene 28: 4034–4040.
71. Xie H, Lim B, Lodish HF (2009) MicroRNAs induced during adipogenesis that accelerate fat cell development are downregulated in obesity. Diabetes 58: 1050–1057.
72. Polster BJ, Westaway SK, Nguyen TM, Yoon MY, Hayflick SJ (2010) Discordant expression of miR-103/7 and pantothenate kinase host genes in mouse. Mol Genet Metab 101: 292–295.
73. Yang J, Qin S, Yi C, Ma G, Zhu H, et al. (2011) MiR-140 is co-expressed with Wwp2-C transcript and activated by Sox9 to target Sp1 in maintaining the chondrocyte proliferation. FEBS Lett 585: 2992–2997.
74. Najafi-Shoushtari SH, Kristo F, Li Y, Shioda T, Cohen DE, et al. (2010) MicroRNA-33 and the SREBP host genes cooperate to control cholesterol homeostasis. Science 328: 1566–1569.
75. Horie T, Ono K, Horiguchi M, Nishi H, Nakamura T, et al. (2010) MicroRNA-33 encoded by an intron of sterol regulatory element-binding protein 2 (Srebp2) regulates HDL in vivo. Proc Natl Acad Sci U S A 107: 17321–17326.
76. Rayner KJ, Suárez Y, Dávalos A, Parathath S, Fitzgerald ML, et al. (2010) MiR-33 contributes to the regulation of cholesterol homeostasis. Science 328: 1570–1573.
77. Marquart TJ, Allen RM, Ory DS, Baldán A (2010) miR-33 links SREBP-2 induction to repression of sterol transporters. Proc Natl Acad Sci U S A 107: 12228–12232.
78. Dávalos A, Goedeke L, Smibert P, Ramírez CM, Warrier NP, et al. (2011) miR-33a/b contribute to the regulation of fatty acid metabolism and insulin signaling. Proc Natl Acad Sci U S A 108: 9232–9237.
79. van Rooij E, Quiat D, Johnson BA, Sutherland LB, Qi X, et al. (2009) A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. Dev Cell 17: 662–673.
80. Hicks JA, Trakooljul N, Liu HC (2010) Discovery of chicken microRNAs associated with lipogenesis and cell proliferation. Physiol Genomics 41: 185–193.
81. Sakakura Y, Shimano H, Sone H, Takahashi A, Inoue N, et al. (2001) Sterol regulatory element-binding proteins induce an entire pathway of cholesterol synthesis. Biochem Biophys Res Commun 286: 176–183.
82. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT (2005) c-Myc-regulated microRNAs modulate E2F1 expression. Nature 435: 839–843.
83. Tarasov V, Jung P, Verdoodt B, Lodygin D, Epanchintsev A, et al. (2007) Differential regulation of microRNAs by p53 revealed by massively parallel Sequencing - miR-34a is a p53 target that induces apoptosis and G(1)-arrest. Cell Cycle 6: 1586–1593.
84. Kozaki K, Inazawa J (2012) Tumor-suppressive microRNA silenced by tumor-specific DNA hypermethylation in cancer cells. Cancer Sci 103: 837–845.
85. Kunej T, Godnic I, Ferdin J, Horvat S, Dovc P, et al. (2011) Epigenetic regulation of microRNAs in cancer: an integrated review of literature. Mutat Res 717: 77–84.
86. Grady WM, Parkin RK, Mitchell PS, Lee JH, Kim YH, et al. (2008) Epigenetic silencing of the intronic microRNA hsa-miR-342 and its host gene EVL in colorectal cancer. Oncogene 27: 3880–3888.
87. Tsuruta T, Kozaki K, Uesugi A, Furuta M, Hirasawa A, et al. (2011) miR-152 is a tumor suppressor microRNA that is silenced by DNA hypermethylation in endometrial cancer. Cancer Res 71: 6450–6462.
88. He Y, Cui Y, Wang W, Gu J, Guo S, et al. (2011) Hypomethylation of the hsa-miR-191 locus causes high expression of hsa-mir-191 and promotes the epithelial-to-mesenchymal transition in hepatocellular carcinoma. Neoplasia 13: 841–853.
89. Shen J, Wang S, Zhang YJ, Kappil MA, Chen Wu H, et al. (2012) Genome-wide aberrant DNA methylation of microRNA host genes in hepatocellular carcinoma. Epigenetics 7.
90. Hwang SH, Kim KU, Kim JE, Kim HH, Lee MK, et al. (2011) Detection of HOXA9 gene methylation in tumor tissues and induced sputum samples from primary lung cancer patients. Clin Chem Lab Med 49: 699–704.
91. Wu Q, Lothe RA, Ahlquist T, Silins I, Tropé CG, et al. (2007) DNA methylation profiling of ovarian carcinomas and their in vitro models identifies HOXA9, HOXB5, SCGB3A1, and CRABP1 as novel targets. Mol Cancer 6: 45.
92. Bandyopadhyay S, Harris DP, Adams GN, Lause GE, McHugh A, et al. (2012) HOXA9 methylation by PRMT5 is essential for endothelial cell expression of leukocyte adhesion molecules. Mol Cell Biol 32: 1202–1213.
93. Zheng CL, Guo ZX, Han ZC, Zhou YL, Lu SH, et al. (2009) [Analysis on promoter CpG methylation and expression of HOXB4 gene in cord blood CD34(+) cells and peripheral blood mononuclear cells]. Zhongguo Shi Yan Xue Ye Xue Za Zhi 17: 674–678.
94. Bennett LB, Schnabel JL, Kelchen JM, Taylor KH, Guo J, et al. (2009) DNA hypermethylation accompanied by transcriptional repression in follicular lymphoma. Genes Chromosomes Cancer 48: 828–841.
95. Issa J-P (2009) DNA methylation as an epigenetic factor in the development and progression of polycythemia vera. University of Texas M.D. Anderson Cancer Center Houston, TX 77030.
96. Kron KJ, Liu L, Pethe VV, Demetrashvili N, Nesbitt ME, et al. (2010) DNA methylation of HOXD3 as a marker of prostate cancer progression. Lab Invest 90: 1060–1067.
97. Anderton JA, Lindsey JC, Lusher ME, Gilbertson RJ, Bailey S, et al. (2008) Global analysis of the medulloblastoma epigenome identifies disease-subgroup-specific inactivation of COL1A2. Neuro Oncol 10: 981–994.
98. Dejeux E, Olaso R, Dousset B, Audebourg A, Gut IG, et al. (2009) Hypermethylation of the IGF2 differentially methylated region 2 is a specific

event in insulinomas leading to loss-of-imprinting and overexpression. Endocr Relat Cancer 16: 939–952.
99. Kuhnert F, Mancuso MR, Hampton J, Stankunas K, Asano T, et al. (2008) Attribution of vascular phenotypes of the murine Egfl7 locus to the microRNA miR-126. Development 135: 3989–3993.
100. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11: 636–646.
101. Bogdanove AJ, Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. Science 333: 1843–1846.