

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA

Uroš JAMNIKAR

**TRANSKRIPTOMSKA ANALIZA STABILNOSTI
PROIZVODNIH KLONOV CELIC OVARIJEV
KITAJSKEGA HRČKA (CHO)**

DOKTORSKA DISERTACIJA

Ljubljana, 2016

UNIVERZA V LJUBLJANI
BIOTEHNIŠKA FAKULTETA

Uroš JAMNIKAR

**TRANSKRIPTOMSKA ANALIZA STABILNOSTI PROIZVODNIH
KLONOV CELIC OVARIJEV KITAJSKEGA HRČKA (CHO)**

DOKTORSKA DISERTACIJA

**TRANSCRIPTOMIC ANALYSIS OF STABILITY IN PRODUCTION
CLONES OF CHINESE HAMSTER OVARY CELLS**

DOCTORAL DISSERTATION

Ljubljana, 2016

Na podlagi Statuta Univerze v Ljubljani ter po sklepu Senata Biotehniške fakultete in sklepa Komisije za doktorski študij Univerze v Ljubljani z dne 12.11.2009 je bilo potrjeno, da kandidat izpolnjuje pogoje za opravljanje doktorata znanosti na Interdisciplinarnem doktorskem študijskem programu Biomedicine, področje genetika. Za mentorja je bila imenovana prof. dr Kristina Gruden.

Doktorska disertacija pomeni sklep podiplomskega študija Biomedicine, področje genetika. Delo je bilo opravljenju v LEK-u d.d., Kolodvorska 27, Mengeš.

Komisija za oceno in zagovor:

Predsednica: Prof.dr. Mojca Narat

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Član: Doc. dr. Bojan Doljak

Univerza v Ljubljani, Fakulteta za farmacijo, Oddelek za farmacevtsko biologijo

Član: Prof.dr. Peter Dovč

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Datum zagovora:

Podpisani izjavljam, da je disertacija rezultat lastnega raziskovalnega dela. Izjavljam, da je elektronski izvod identičen tiskanemu. Na univerzo neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravico shranitve avtorskega dela v elektronski obliki in reproduciranja, ter pravico omogočanja javnega dostopa do avtorskega dela na svetovnem spletu preko Digitalne knjižnice Biotehniške fakultete.

Uroš Jamnikar

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

- ŠD Dd
- DK 575:615.32(043.2)=163.6
- KG CHO/rekombinantna učinkovina/biološka zdravila/izražanje genov
- AV JAMNIKAR Uros, univerzitetni diplomiran mikrobiolog
- SA GRUDEN Kristina (mentorica)
- KZ SI-1000 Ljubljana, Jamnikarjeva 101
- ZA Univerza v Ljubljani, Biotehniška fakulteta, Interdisciplinarni doktorski študij Biomedicine, področje Genetike.
- LI 2016
- IN TRANSKRIPTOMSKA ANALIZA STABILNOSTI PROIZVODNIH KLONOV CELIC OVARIJEV KITAJSKEGA HRČKA (CHO)
- TD Doktorska disertacija
- OP XI, 94 str., 12 pregl., 24 sl., 54 vir.
- IJ sl
- JI sl / en
- AI Zaradi pravilnega zvijanja in potranslacijskih modifikacij proteinov so ovarijske celice kitajskega hrčka (celice CHO) postale glavni gostiteljski organizem za proizvodnjo rekombinantnih proteinov. Najbolj uporaben ekspresijski sistem za proizvodnjo rekombinantnih proteinov v celicah CHO je CHO-*Dhfr* ekspresijski sistem. Celice CHO so znane po nestabilnem kariotipu, še posebej kadar so celice CHO tretirane z metotreksatom (MTX). Trenutno uporabljeni metoda za testiranje stabilnosti produkcije rekombinantnega proteina skozi daljše časovno obdobje temelji na eksperimentalnih podatkih, ki se zbirajo 10 tednov. Z namenom izpopolnitve tega sistema, smo z DNA-mikromrežami naredili profil transkriptoma na izbranih klonih celic CHO. Identificirali smo gene, ki so bili statistično različno izraženi med kloni s stabilno/nestabilno produkcijo rekombinantnega proteina in med kloni z visoko/nizko produkcijo rekombinantnega proteina. Izbrane gene smo dodatno potrdili z RT-qPCR. V nadaljevanju smo z algoritmom k-najbližjega soseda pokazali, da lahko uporabimo kombinacijo treh pomembnejših genov za napoved stabilne produkcije rekombinantnega proteina, že v začetku razvoja celične linije.

KEY WORDS DOCUMENTATION

DN Dd
DC 575:615.32(043.2)=163.6
CX CHO/recombinant product /biological drug/gene expression/genes
AU JAMNIKAR Uros
AA GRUDEN Kristina (mentor)
PP SI-1000 Ljubljana, Jamnikarjeva 101
PB UnIversity of Ljubljana, Biotechnical Faculty, Interdisciplinary Doctoral
Programme in Biomedicine, scientific field Genetics.
PY 2016
TI TRANSCRIPTOMIC ANALYSIS OF STABILITY IN PRODUCTION
CLONES OF CHINESE HAMSTER OVARY CELLS
DT Doctoral dissertation
NO XI, 94 p., 12 tab., 24 fig., 54 ref.
LA sl
AL sl / en
AB Chinese hamster ovary (CHO) cells have become the host of choice for the production of recombinant proteins, due to their capacity for correct protein folding, assembly, and posttranslational modifications. The most widely used system for recombinant proteins is the gene amplification procedure, which uses CHO-*Dhfr* expression system. However, CHO cells are known to have a very unstable karyotype. This is due to chromosome rearrangements that can arise from translocations and homologous recombinations, especially when the cells with CHO-*Dhfr* expression system are treated with methotrexate hydrate (MTX). The present method used in the industry for testing clones for their long-term stability and high production of recombinant protein production is empirical, and it involves the cultivation over 10 weeks prior to the selection of the most suitable clone for further bioprocess development. The transcriptome profiles of CHO clones were investigated using DNA microarrays. We identified marker genes that were differently expressed between clones with stable/unstable and high/low recombinant protein production. Their expression was validated by reverse-transcription quantitative real-time PCR (RT-qPCR). Furthermore, k-nearest neighbour algorithm approach showed that the combination of the gene expression patterns of only three genes is sufficient to predict stable recombinant protein production in clones already in the early phases of cell line development.

KAZALO VSEBINE

KLJUČNA DOKUMENTACIJSKA INFORMACIJA	III
KEY WORDS DOCUMENTATION	IV
KAZALO VSEBINE	V
KAZALO PREGLEDNIC	VII
KAZALO SLIK	VIII
OKRAJŠAVE IN SIMBOLI	IX
1 UVOD	1
2 PREGLED OBJAV	2
2.1 ZGODOVINSKI PREGLED NASTANKA BIOLOŠKIH ZDRAVIL	2
2.2 RAZVOJ PRODUKCIJE BIOLOŠKO PODOBNIH PROTEINOV	4
2.3 VARNOST BIOLOŠKIH IN BIOLOŠKO PODOBNIH PROTEINOV	5
2.4 CHO-DHFR EKSPRESIJSKI SISTEM ZA PRODUKCIJO REKOMBinantnega proteina	6
2.5 CILJI DOKTORSKE NALOGE	7
2.6 DELOVNE HIPOTEZE	8
3 MATERIALI IN METODE	9
3.1 MATERIAL	9
3.2 UPORABLJENE METODE	9
3.2.1 Transfekcija	9
3.2.2 Kloniranje transficiranih celic CHO	11
3.2.3 Gojenje pripravljenih rekombinantnih klonov	12
3.2.4 Producija rekombinantnega proteina	14
3.3 IZOLACIJA DNA IN RNA	14
3.3.1 Izolacija genomske DNA (gDNA)	14
3.3.2 Izolacija celokupne RNA (totRNA) in analiza kvalitete izolacije	14
3.3.3 Odstranitev genomske DNA	15
3.3.4 Obratno prepisovanje	16
3.4 ANALIZA TRANSKRIPTOMA Z DNA-MIKROMREŽAMI	17
3.4.1 Priprava in hibridizacija vzorca za DNA-mikromreže	17
3.4.2 Obdelava in analiza podatkov iz DNA-mikromrež	18
3.5 PCR V REALNEM ČASU	18
3.5.1 Analiza izražanja genov	18
3.5.2 Izbor genov in načrtovanje njihovih oligonukleotidnih začetnikov/sond ..	19
3.5.3 Izvedba reakcij PCR v realnem času	20
3.5.4 Določitev optimalnih parametrov delovanja testov qPCR	22
3.5.5 Statistične analize podatkov	25
3.5.5.1 Test t	25
3.5.5.2 Razvrščanje in analiza PCA	25
3.5.6 Analiza števila kopij rekombinantnega gena	26
4 REZULTATI	30
4.1 PRODUKCIJA REKOMBINANTE UČINKOVINE IN VPLIV MTX	31
4.2 PRODUKCIJA REKOMBINANTNEGA PROTEINA IN ŠTEVILO KOPIJ REKOMBinantnega gena	33

4.3	OBDELAVA IN ANALIZA PODATKOV IZ DNA-MIKROMREŽ.....	36
4.3.1	Stabilnost produkcije rekombinante učinkovine skozi čas (stabilnost).....	40
4.3.2	Učinkovitost produkcije rekombinantnega proteina (produktivnost)	41
4.3.3	Vpliv MTX na produkcijo rekombinantnega proteina (tretiranje z/brez MTX).....	42
4.4	POTRJEVANJE IZRAŽANJA IZBRANIH GENOV Z METODO RT-qPCR..	44
4.4.1	Stabilnost produkcije rekombinantnega proteina skozi čas (stabilnost).....	47
4.4.1.1	Test t (stabilnost)	50
4.4.1.2	Diskriminatorna analiza.....	55
4.4.1.3	Analiza PCA	58
4.4.2	Učinkovitost produkcije rekombinantnega proteina (produktivnost)	60
4.4.2.1	Test – t (produktivnost)	63
4.4.3	Vpliv MTX na produkcijo rekombinantnega proteina (tretiranje z/brez MTX).....	66
4.4.3.1	Test t (tretiranje z / brez MTX)	69
5	RAZPRAVA	72
5.1	ŠTEVILO KOPIJ REKOMBINANTNEGA GENA IN UČINKOVITOST PRODUKCIJE REKOMBINANTNEGA PROTEINA NISTA POVEZANI	72
5.2	ANALIZA TRANSKRIPTOMA OMOGOČA IDENTIFIKACIJO GENETSKIH MARKERJEV.....	74
5.3	OD ANALIZE TRANSKRIPTOMA DO IZBORA GENOV POMEMBNIH ZA STABILNOST PRODUKCIJE REKOMBINANTE UČINKOVINE.....	76
5.4	GENI POVEZANI Z UČINKOVITOSTJO PRODUKCIJE REKOMBINANTNEGA PROTEINA	79
5.5	VPLIV MTX NA PRODUKCIJO REKOMBINANTNEGA PROTEINA	80
5.6	MESTO INTEGRACIJE REKOMBINANTNEGA GENA V GOSTITELJSKI GENOM IMA POMEMBNO VLOGO PRI STABILNEM IZRAŽANJU REKOMBINANTNEGA GENA	81
6	SKLEPI	85
7	POVZETEK (SUMMARY)	87
7.1	POVZETEK	87
7.2	SUMMARY	88
8	VIRI.....	90
	ZAHVALA	

KAZALO PREGLEDNIC

Pregl. 1:	Priprava reakcijskih mešanic za obdelavo z DNazo.....	16
Pregl. 2:	Potek inkubacij pri obdelavi RNA z DNazo.	16
Pregl. 3:	Priprava reakcijskih mešanic za prepis v cDNA.	17
Pregl. 4:	Potek inkubacij za prepis v cDNA.	17
Pregl. 5:	Reakcijska mešanica za reakcijo qPCR – določanje števila kopij rekombinantnega gena.	20
Pregl. 6:	Začetni oligonukleotidi in sonde genov uporabljenih za določanje relativnega izražanja in števila kopij genov s qPCR.	27
Pregl. 7:	Priprava reakcijskih mešanic za qPCR – izražanje genov.....	28
Pregl. 8:	Produktivnost klonov z in brez MTX v gojišču skozi 10 tednov dolgo študijo.	32
Pregl. 9:	Glavni parametri, ki opisujejo značilnosti pomnoževanja amplikonov qRT-PCR.....	44
Pregl. 10:	Opis funkcije genov, potencialnih markerjev za stabilno produkcijo rekombinantnega proteina.....	49
Pregl. 11:	Opis funkcije genov iz analize učinkovitosti produkcije rekombinantnega proteina.	62
Pregl. 12:	Opis funkcije genov iz sklopa vpliva MTX na produkcijo rekombinante učinkovine. ...	68

KAZALO SLIK

Sl. 1: Encim DHFR deluje kot katalizator pri pretvorbi dihidrofolne kisline v tetrafolno kislino (Voet in Voet, 1995: 356).....	6
Sl. 2: Prikaz razvoja mešane populacije celic.....	10
Sl. 3: Princip kloniranja transficiranih celic CHO.....	12
Sl. 4: Shematski prikaz vzorčenja.....	13
Sl. 5: Shematski prikaz poteka poskusa od nukleotidnega zaporedja za rekombinant gen do izbora statistično značilno različno izraženih genov.....	31
Sl. 6: Odvisnost med produktivnostjo in številom kopij rekombinantnega gena.....	34
Sl. 7: Prikaz korelacije med številom kopij rekombinantnega gena v gostiteljski celici in med produktivnostjo.....	35
Sl. 8: Porazdelitev vrednosti izražanja genov vzorcev iz DNA-mikromrež pred normalizacijo z RMA algoritmom (GeneSpring).....	36
Sl. 9: Porazdelitev vrednosti izražanja genov pri različnih vzorcih iz DNA-mikromrež po normalizaciji z RMA algoritmom (GeneSpring).....	37
Sl. 10: Normalizirane (\log_2) vrednosti signala hibridizacijskih kontrolnih sond na DNA-mikromrežah za namen preverjanja učinkovitosti hibridizacije vzorcev na DNA-mikromreže.....	38
Sl. 11: Razlika v intenziteti izražanja genov na DNA-mikromrežah med vzorci s stabilno in nestabilno produkcijo rekombinantnega proteina.....	40
Sl. 12: Toplotni diagram prikazuje razliko v izražanju genov med vzorci, ki imajo visoko produkcijo in med vzorci z nizko produkcijo rekombinantnega proteina ob prisotnosti MTX v gojišču.....	41
Sl. 13: Venn-ov diagram prikazuje presek statistično različno izraženih genov med dvema analizama izražanja genov.....	43
Sl. 14: Prikaz korelacije izmerjenega izražanja 36 genov med DNA-mikromrežami in RT-qPCR.	46
Sl. 15: Primerjava izražanja genov značilnih za stabilnost produkcije rekombinantnega proteina med metodo DNA-mikromrež in metodo RT-qPCR.	48
Sl. 16: Grafikon kvantilov prikazuje izražanje izbranih genov pri stabilnih in nestabilnih vzorcih.	52
Sl. 17: Grafikon kvantilov prikazuje izražanje izbranih genov med stabilnimi vzorci na začetku in koncu študije.....	54
Sl. 18: Diskriminatorska analiza prikazuje ločitev na skupino vzorcev s stabilno in na skupino vzorcev z nestabilno produktivnostjo že na začetku (teden 1 in 2) študije, glede na izražanje treh izbranih genov.....	56
Sl. 19: Diskriminatorska analiza prikazuje ločitev vzorcev na skupino stabilnih in nestabilnih vzorcev na koncu študije (teden 9 in 10), glede na izražanje izbranih genov.	57
Sl. 20: Tri-dimenzionalna predstavitev, kjer se stabilni in nestabilni vzorci ločijo v 2 skupini, glede na lastnosti 3 komponent (P1, P2, P3) PCA analize.	59
Sl. 21: Primerjava izražanja genov značilnih za učinkovitost produkcije rekombinantnega proteina med metodo DNA-mikromrež in metodo RT-qPCR.	61
Sl. 22: Grafikon kvantilov prikazuje relativno izražanje izbranih genov med kloni z visoko in nizko produkcijo rekombinantne učinkovine.	65
Sl. 23: Primerjava izražanja genov z in brez dodanega MTX v gojišču med metodo DNA-mikromrež in metodo RT-qPCR.	67
Sl. 24: Grafikon kvantilov prikazuje relativno izražanje izbranih genov med vzorci tretiranih z MTX ali brez tretiranja z MTX.	71

OKRAJŠAVE IN SIMBOLI

Ac115880.11	protein LEO1 v povezavi s polimerazo RNA, <i>angl.</i> RNA polymerase-associated protein LEO1
Ac158560.4	ni podatka
AcGa	normalizator, izračunan iz geometrijskega povprečja vzdrževalnih genov Actb in Gapdh
Actb	vzdrževalni gen - beta aktin
Ankrd26	protein 26 s ponavljajočo domeno ankiran, <i>angl.</i> ankyrin repeat domain-containing protein 26
B2bf95_rat	TCP vezavni protein 2, <i>angl.</i> TCP binding protein 2
BHK	ledvične celice mladiča hrčka
Bx842664.2 / Hist1h3c	histon H3.2, <i>angl.</i> histone H3.2
C1qtnf1	protein 1 soroden komplementu C1q tumor nekrotizirajočemu faktorju, <i>angl.</i> complement C1q tumor necrosis factor-related protein 1
Cd68	makrosialin, <i>angl.</i> macrosialin
Cdh1	kadherin-1, <i>angl.</i> cadherin-1
cDNA	komplementarna DNA, <i>angl.</i> complementary DNA
Ces2g	karboksilesteraza 2G, <i>angl.</i> carboxylase 2G
CHO	celice iz ovarijskega kitajskega hrčka
Cntn1	kontaktin-1, <i>angl.</i> contactin-1
Cq	cikel kvantifikacije, <i>angl.</i> quantification cycle
Cspg4	hondroitin sulfat proteoglikan 4, <i>angl.</i> chondroitin sulfate proteoglycan 4
Cu459186.17	ni podatka
ΔCq	razlika povprečnih vrednosti Cq med višjo in nižjo redčino produkta
ΔΔCq	razlika ΔCq vzorca tarče in ΔCq vzorca normalizatorja
ddH ₂ O	destilirana voda
DHFR	dihidrofolat reduktaza
DNA	deoksiribonukleinska kislina
ΔRn	razlika emisij fluorescence Rn ⁺ in Rn ⁻
E	učinkovitost pomnoževanja tarčne DNA
E130203b14Rik	ni podatka
Egr1	protein 1 zgodnjega odziva rasti, <i>angl.</i> early growth response protein 1
ECACC	Evropska zbirka celičnih kultur, <i>angl.</i> European Collection of Cell Culture
EMA	Evropska agencija za zdravila, <i>angl.</i> European Medicines Agency
FAM	fluorescenčno barvilo na qPCR sondi
FDA	Agencija za hrano in zdravila, <i>angl.</i> Food and Drug Administration
Fgfr2	receptor 2 za rastni faktor fibroblastov, <i>angl.</i> fibroblast growth factor receptor 2
FISH	fluorescenčna hibridizacija <i>in situ</i>

Flt1	receptor 1 za vaskularno endotelni rastni faktor, <i>angl.</i> vascular endothelial growth factor receptor 1
Foxp2	forkhead box protein P2, <i>angl.</i> forkhead box protein P2
Gapdh	vzdrževalni gen - gliceraldehid 3-fosfat dehidrogenaza
gDNA	genomska DNA
Gluc	glukagen
GMP	dobra proizvodna praksa, <i>angl.</i> good manufacturing principles
Gpc6	glipican-6, <i>angl.</i> glycan-6
GS	glutamin-sintetaza
HEK-293	človeške embrionalne celice
Hmgn2	ne-histonski kromosomalni protein HMG-17, <i>angl.</i> non-histone chromosomal protein HMG-17
Hist1h2bc	histon H2B tipa 1-c/e/g, <i>angl.</i> histone H2B type 1-C/E/G
k	naklon krivulje
mAb	monoklonalno protitelo
MGB-probe	vezava sonde na manjšo brazdo DNA, <i>angl.</i> minor groove binding
Mmp10	matrična metalopeptidaza 10, <i>angl.</i> matrix metallopeptidase 10
»mpc«	mešana populacija celic
MTX	metotreksat
Myrip	efektor Rab MyRIP, <i>angl.</i> rab effector MyRIP
NFW	voda brez vsebnosti nukleaz, <i>angl.</i> Nuclease free water
NS0	mišje mielomske celice
NTC	negativna kontrola
qPCR	kvantitativni PCR v realnem času
PBS	fosfatni pufer, <i>angl.</i> phosphate-buffered saline
PCA	metoda glavnih komponent, <i>angl.</i> principal component analysis
PCR	verižna reakcija s polimerazo
PER-C6	celice iz človeške retine
Plac1	protein 1 specifičen za placento, <i>angl.</i> placenta-specific protein 1
Plekhb1	protein 1 z domeno, homologno plekstrinu iz družine B, <i>angl.</i> pleckstrin homology domain-containing family B member 1
Psma8	protein podoben podenoti proteasoma alfa tipa 7, <i>angl.</i> proteasome subunit alpha type-7-like
Ptpre	receptor tipa tirozin-protein fosfataza ipsilon, <i>angl.</i> receptor-type tyrosine-protein phosphatase epsilon
R	relativno izražanje gena
R ²	linearni koreacijski koeficient
RIN	integriteta RNA, <i>angl.</i> RNA integrity number
RMA	robustno povprečje več mikromrež, <i>angl.</i> robust multi-array average
Rn ⁺	emisija fluorescence produkta
Rn ⁻	emisija fluorescence pasivnega referenčnega barvila

RNA	ribonukleinska kislina
ROX	pasivno referenčno barvilo
Rps6ka2	rinaza alfa-2 ribosomalnega proteina S6, <i>angl.</i> ribosomal protein S6 kinase alpha-2
RT-qPCR	qPCR s predhodnjim obratnim prepisovanjem
Spn1	snurportin-1, <i>angl.</i> snurportin-1
Sprr2h	majhen protein 2H bogat s prolinom, <i>angl.</i> small proline-rich protein 2H
Sprr2k	majhen protein 2K bogat s prolinom, <i>angl.</i> small proline-rich protein 2K
Thsd7a	protein 7A z domeno trombospondin tipa 1, <i>angl.</i> thrombospondin type-1 domain-containing protein 7A
Tinagl1	protein podoben antigenu tubulointerstitialni nefritis, <i>angl.</i> tubulointerstitial nephritis antigen-like
totRNA	celokupna RNA, <i>angl.</i> total RNA
Vsnl1	vizininu podoben protein 1, <i>angl.</i> visinin-like protein 1
Zdhhc6	palmitoiltransferaza ZDHHC6, <i>angl.</i> palmitoyltransferase ZDHHC6

1 UVOD

S pomočjo rekombinantne tehnologije farmacevtske družbe izdelujejo učinkovine bioloških zdravil. Preden se biološko zdravilo dostopno na trgu, je treba preizkusiti njegovo kakovost, varnost in učinkovitost. Treba je opraviti vse farmacevtske, kemijske, biološke in mikrobiološke analize, farmakološko-toksikološke preizkuse na živalih, ki dokazujejo njegovo varnost, ter seveda klinične preizkuse na zdravih prostovoljcih in bolnikih, s čimer se dokaže njegova varnost in učinkovitost pri ljudeh. Biološka zdravila pridobivajo iz celic, ki jih gojijo v celičnih kulturah. Kakršnekoli spremembe v procesu gojenja lahko vodijo do sprememb v zgradbi učinkovine. Zaradi razlik v procesih pri izdelavi bioloških zdravil, pa tudi zaradi posledičnih možnih sprememb v zgradbi med istimi biološkimi zdravili različnih proizvajalcev in zaradi izjemno strogih zahtev glede varnosti in z njo povezane potencialne imunogenosti, so regulatorne ustanove v razvitem svetu, kot sta Agencija za hrano in zdravila (FDA) in Evropska agencija za zdravila (EMA), uvedle novo skupino zdravil z imenom biološkim podobna zdravila (*angl. biosimilars*). Po poteku patentne zaščite inovativnega biološkega zdravila se lahko na trg uvrsti biološkemu podobno zdravilo, ki vsebuje enako učinkovino in se po jakosti, kakovosti, varnosti in učinkovitosti ne razlikuje od inovativnega. Kakovost, varnost in učinkovitost biološkemu podobnega zdravila dokazujemo z lastnimi kliničnimi preiskavami, ki po številu bolnikov niso tako obsežne kot v primeru inovativnega zdravila (Štrukelj in Kos, 2007). Poleg kakovosti in učinkovitosti biološko podobnega zdravila je med samim razvojem biološko podobnega zdravila pomembna visoka produktivnost in predvsem stabilna produkcija učinkovine uporabljene v biološkem zdravilu. Med delom smo se osredotočili na iskanje pomembnejših genov za ugotavljanje stabilne in visoke produkcije rekombinantnega proteina, ki se uporablja v bioloških zdravilih.

2 PREGLED OBJAV

2.1 ZGODOVINSKI PREGLED NASTANKA BIOLOŠKIH ZDRAVIL

Leta 1982 je bil za uporabo v medicinske namene odobren rekombinantni inzulin, ki je bil prvo biološko zdravilo na trgu. Zaradi enostavne zgradbe molekule je njegovo pridobivanje mogoče v bakteriji *Escherichia coli*. Pri proizvodnji učinkovin v mikrobnih ekspresijskih sistemih pa ti niso glikozilirani, ali pa se glikozilacija razlikuje od tiste v humanih celicah. Zato se je zaradi potrebe po proizvodnji kompleksnejših rekombinantnih učinkovin začel razvoj sesalskih produkcijskih celičnih linij, ki omogočajo post-translacijske modifikacije, značilne za humane celice. Začetki njihove uporabe segajo v petdeseta leta prejšnjega stoletja. Uporabljati so jih začeli za proizvodnjo cepiv, izredno široko uporabo pa so doživele v sedemdesetih in osemdesetih letih z razvojem tehnologije za proizvodnjo rekombinantnih učinkovin. Največkrat se uporablajo celice ovarija kitajskega hrčka (CHO), celice mieloma miši (NS0), celice ledvic mladiča hrčka (BHK), linija HEK-293 iz embrionalnih celic človeka ter linija PER-C6 iz retine človeka (Butler, 2005).

Delež rekombinantnih učinkovin, proizvedenih v celicah CHO, je okoli 30 %. Pri produkciji terapevtskih rekombinantnih monoklonskih protiteles, ki so izredno velika skupina med biološkimi zdravili, je ta delež kar 50 % (Omasa in sod., 2010). Celice CHO so tako postale standarden sistem za proizvodnjo rekombinantnih učinkovin (Durocher in Butler, 2009). Prvič so bile izolirane leta 1958 (Landauer in sod., 2014), od takrat je bilo razvitih veliko različnih derivatov (Werner in sod., 2007). Leta 1986 je bila na trgu dovoljena uporaba prvega rekombinantnega proteina, proizvedenega v celicah CHO. To je bil tkivni plazminogen aktivator (Omasa in sod., 2010). Celice CHO so najpomembnejša celična linija za proizvodnjo rekombinantnih učinkovin tudi zato, ker omogočajo gojenje po t. i. standardih dobre proizvodne prakse (GMP). Posledično to pomeni, da regulatorne ustanove hitreje izdajo dovoljenje za prodajo biološkega zdravila.

Pri proizvodnji rekombinantnih učinkovin je aminokislinsko zaporedje proteina kodirano z nukleotidnim zaporedjem gena, ki ga vnesemo v celico. Glikozilacija je posledica post-translacijskih modifikacij in se lahko razlikuje med različnimi celičnimi linijami,

različnimi kloni enake celične linije, različnimi gojišči, različnimi pogoji gojenja (vsebnost hranil, pH, temperatura), različno specifično produktivnostjo in fiziološkim stanjem celice (Werner in sod., 2007). Poleg glikozilacije je pomembno tudi pravilno zvijanje in sestavljanje proteinov in njihovo izločanje iz celice. Vse to lahko vpliva na biološko aktivnost proteinov, hkrati pa lahko neustrezna glikozilacija sproži imunski odziv po vnosu proteina v krvni obtok pacienta (Štrukelj in Kos, 2007). Omogočajo tudi izločanje ustrezeno zvitih in ustrezeno glikoziliranih proteinov v gojišče, ki so kompatibilni in bioaktivni v ljudeh (Werner in sod., 2007).

Vse omenjene lastnosti so ključne za varno delovanje rekombinantnega proteina pri pacientih. Poleg tega so celice CHO, tudi zelo dobro karakterizirane, kar je posledično pripomoglo, da so regulatorni organi izdali že več dovoljenj za uporabo učinkovin za zdravljenje pacientov (Chu in Robinson, 2001).

Slabost celic CHO je njihova nizka produkcija rekombinantnih učinkovin. Visoko produktivnost celic CHO dosežemo šele ob uporabi enega od sistemov za amplifikacijo vstavljenih genov. Poznamo več takšnih sistemov, eden je dihidrofolat reduktaza (DHFR), ki je najpogosteje uporabljen, drug zelo pogosto uporabljen pa je glutamin-sintetaza (GS). Pri obeh sistemih izvajamo seleksijski pritisk, v primeru DHFR sistema izvajamo seleksijski pritisk na DHFR encim z dodajanjem metotreksata (MTX) v gojišče, ki inhibira delovanje omenjenega encima. V primeru GS sistema, se seleksijski pritisk izvaja na celice s pomočjo gojišča, ki ne vsebuje esencialne aminokisline L-glutamin. V tem primeru preživijo samo celice, ki so sprejele ekspresijski vektor z zapisom za sintezo L-glutamina. V obeh primerih izvajanja seleksijskega pritiska na celice, celice pomnožujejo število genov, v prvem primeru *Dhfr* gen in v drugem primeru gen, ki kodira zapis za glutamin sintetazo. Zraven omenjenih genov se pomnožijo tudi geni, ki nosijo zapis za rekombinantno učinkovino. Geni z zapisom za rekombinantno učinkovino se pomnožujejo, ker so na ekspresijskem vektorju, ki se vstavi v celico, locirani zelo blizu genov na katere se izvaja seleksijski pritisk (Werner in sod., 2007). Na takšen način je mogoče zelo povisati produkcijo rekombinantnih učinkovin.

Celice CHO so v preteklosti gojili v gojišču, ki je bil deloma sestavljen iz živalskih komponent. V zadnjih letih so se na trgu pojavila gojišča brez živalskih komponent, ki so kemijsko karakterizirana in omogočajo rast celic CHO do velikih koncentracij. Zaradi razvoja gojišč, brez živalskih komponent, se lahko celice CHO gojijo v suspenziji, kar ni bilo mogoče ob prisotnosti živalskih komponent v gojišču (Butler, 2005). Poleg tega jih lahko gojimo do visoke gostote celic v dohranjevalnih bioreaktorjih velikega volumena nad 10.000 litrov.

2.2 RAZVOJ PRODUKCIJE BIOLOŠKO PODOBNIH PROTEINOV

Pri razvoju biološko podobnih proteinov je največji izviv doseganje ustrezone biološke podobnosti. V primeru razvoja bioloških proteinov izbira ustrezne učinkovine ni pomembna, saj je nukleotidno zaporedje gena, ki ga vstavljamo v celico, kodira enako zaporedje kot pri originalnem biološkem zdravilu. Pomembno je uravnavanje post-translacijskih modifikacij med razvojem celične linije in sam proces gojenja celic (Jostock in Knopf, 2012). Pri razvoju procesa produkcije biološko podobnih proteinov je zaradi ekonomike procesa pomembno tudi, da se izbere celično linijo, ki bo proizvajala proteine z visoko učinkovitostjo.

V začetnih fazah razvoja biološko podobnega proteina testirajo večje število različnih celičnih linij, v katere vnesejo ekspresijski vektor z genom za želeni protein. Na ta način pridobijo mešane populacije celic, t. i. »mpc« (skupek celic z različnimi lastnostmi), pri katerih podobnost proizvedenih proteinov primerjajo s karakteristikami proteinov originatorja. Karakterizacijo biološko podobnih proteinov izvajajo z različnimi analitskimi tehnikami in tako izberejo najprimernejše »mpc« za nadaljnje delo. S postopkom kloniranje iz »mpc« osamijo posamezne klone, pri katerih ponovno preverijo kakovost in podobnost produkta z originatorskim produktom. Na podlagi teh rezulatov izberejo manjše število klonov, katerih produkt je najbolj podoben inovativnemu biološkemu proteinu. Z optimizacijo gojišča in z razvojem procesa gojenja celic v bioreaktorju (razvoj bioprosesa), se še bolj približajo biološki podobnosti originalnemu zdravilu v okvire, ki jih dopušča regulativa (Štrukelj in Kos, 2007).

Sodoben pristop pri proizvodnji biološko podobnih zdravil je tako imenovana kakovost z načrtovanjem, pri kateri vnaprej ugotovijo, katere strukture želijo spremeniti in v kakšnem obsegu, nato pa glede na poglobljeno znanje o produkcijskih celičnih linijah in procesih, vnaprej načrtujejo celotno pot razvoja biološko podobnega proteina. V primeru, da določene strukture kljub uporabi ustrezne celične linije in modifikacije procesa kultivacije ni možno spremeniti v zadostni meri, se lahko razvije novo genetsko modificirano celično linijo, ki bo proizvajala ustrezen produkt (Štrukelj in Kos, 2007).

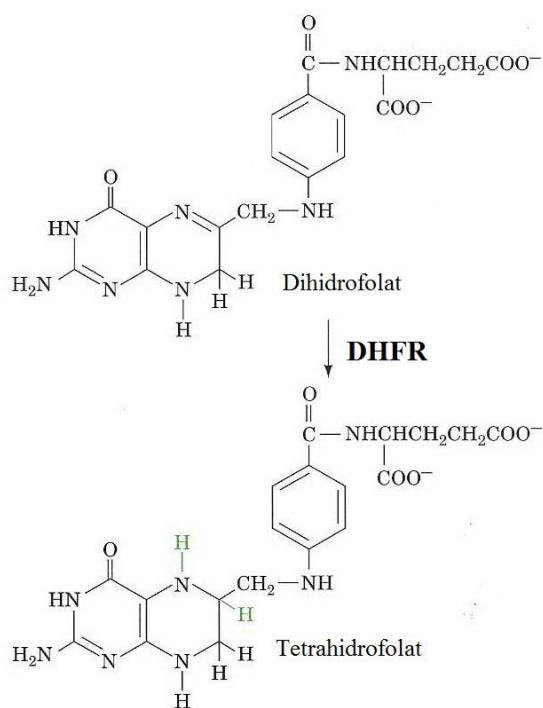
2.3 VARNOST BIOLOŠKIH IN BIOLOŠKO PODOBNIH PROTEINOV

Pri proizvodnji bioloških in biološkim podobnih proteinov je poleg doseganja visoke kakovosti in učinkovitosti ključnega pomena zagotavljanje njihove varnosti. Potrebna sta zagotavljanje specifičnosti delovanja in preprečevanje imunogenosti.

Poleg tega je potrebno zagotoviti odsotnost patogenov v produktu, ki bi se lahko vnesli v pacienta hkrati z vnosom biološkega zdravila. Pri večini bioloških zdravil se učinkovina namreč injicira neposredno v krvni obtok. Zaradi takšnega načina vnosa je treba zagotoviti, da biološki kontaminanti, imenovani tudi adventicijski agensi, niso prisotni. Vir okužbe so lahko parentalne celične linije ali pa sestavine gojišča. V gojišča se lahko dodajajo komponente živalskega izvora, zaradi česar bi bil potencialno mogoč prenos virusov. Varnost bioloških zdravil zagotavljajo z uporabo ustreznih surovin za pripravo gojišč, kar se zagotavlja s sledljivostjo njihovega izvora in dodatnimi kemijskimi analizami s katerimi preverjajo ustreznost kemijske sestave. Osnova za proizvodnjo so dobro karakterizirane celične banke. Poleg tega se preverja prisotnost okužbe tudi med gojenjem rekombinantnih celičnih kultur, v procesu čiščenja produkta pa vse nečistoče odstranjujejo z zaporednimi čiščenji z uporabo različnih kromatografskih tehnik (Štrukelj in Kos, 2007).

2.4 CHO-DHFR EKSPRESIJSKI SISTEM ZA PRODUKCIJO REKOMBINANTNEGA PROTEINA

Za zagotavljanje visoke produkcije rekombinantnega proteina, se najpogosteje uporablja CHO-*Dhfr* ekspresijski sistem. CHO-*Dhfr* ekspresijski sistem je sestavljen iz gena za rekombinantni protein in iz rekombinantnega gena za dihidrofolat reduktazo (*Dhfr*). Encim DHFR reducira dihidrofolno kislino v tetrafolno kislino, ki je pomembna pri sintezi purinov, torej prekurzorjev za sintezo DNA (Sl. 1). Zato je encim DHFR bistven za celično delitev in rast. MTX inhibira delovanje encima DHFR v celicah in posledično povzroča pomanjkanje prekurzorjev za DNA. Zaradi tega celica pomnoži gen *dhfr*, kar vključuje tudi rekombinantni gen *dhfr*, ki se skupaj z genom za rekombinantni protein nahaja na istem odseku vnešene DNA (ekspresijski vektor). Posledično se poveča tudi število kopij gena za rekombinantni protein, kar pomeni večjo produkcijo rekombinantnega proteina (Reff in sod., 1993).



Slika 1: Encim DHFR deluje kot katalizator pri pretvorbi dihidrofolne kisline v tetrafolno kislino (Voet in Voet, 1995: 356).

Figure 1: DHFR enzyme catalyses conversion of tetrahydropholic acid to dihydropholic acid (Voet in Voet, 1995: 356).

Celice CHO so znane po zelo nestabilnem kariotipu zaradi prerazporeditve kromosomov, ki so posledica translokacij in homolognih rekombinacij, še posebej če celice gojimo v prisotnosti MTX (Flintoff in sod., 1984; Bacsi in sod., 1986; Rolig in sod., 1997; Yoshikawa in sod., 2000). V povezavi s tem je pogosto nestabilna tudi produkcija rekombinantnih proteinov. Nestabilna produkcija je bila dokazana pri 8 % do 63 % vseh rekombinantnih celic CHO, kjer je bil uporabljen CHO-*Dhfr* ekspresijski sistem, ne glede na to ali je bil seleksijski pritisk (MTX) prisoten ali ne (Fann in sod., 2000; Heller-Harrison in sod., 2009; Kim in sod., 1998; Kim in sod., 1999).

2.5 CILJI DOKTORSKE NALOGE

Predhodne študije transkriptoma celic CHO s pomočjo DNA-mikromrež, so bile osredotočene na celične mehanizme povezane s produkcijo rekombinante učinkovine. Poleg tega je bilo več študij osredotočenih na vpliv pogojev gojenja (sprememba temperature gojišča, število obratov bioreaktorjev) v povezavi s produkcijo rekombinantnega proteina (Doolan in sod., 2008; Kantardjieff in sod. 2010; Schaub in sod., 2010; Trummer in sod., 2008). Objavljene ni bilo nobene študije, ki bi spremljala transkriptom celic CHO skozi daljše časovno obdobje, kjer bi se osredotočili na iskanje pomembnejših genov, ki so specifični za celice CHO s stabilno produkcijo rekombinantnega proteina.

Cilj doktorske naloge je razumevanje razlik v celičnih procesih pri visoko in nizko producirajočih klonih s proučevanjem transkriptoma celic CHO. Predvsem nas zanima ali obstajajo geni pri celicah CHO, ki lahko napovedo stabilno produkcijo rekombinantnega proteina, skozi daljše časovno obdobje. Stabilna produkcija skozi daljše časovno obdobje, je eden izmed ključnih dejavnikov pri razvoju procesa priprave biološkega zdravila. Gen za rekombinantno učinkovino, ki nosi nukleotidni zapis za učinkovino mora biti enakomerno izražen, skozi celotno obdobje razvoja in proizvodnjo rekombinantnega proteina, kar je poleg ustrezne transkripcije in translacije predpogoj za stabilno produkcijo rekombinante učinkovine. V primeru, da najdemo zanesljive genetske markerje za stabilno produkcijo rekombinantnega proteina, lahko z vedenjem o njihoven načinu izražanja

prihranimo veliko časa pri razvoju rekombinantnega proteina. Za dosego zastavljenih ciljev, smo se odločili za pristop analize transkriptoma transgenih celic CHO s pomočjo DNA-mikromrež in PCR v realnem času s predhodnjim korakom obratnega prepisovanja (RT-qPCR).

2.6 DELOVNE HIPOTEZE

- S pomočjo DNA-mikromrež bo možno netarčno proučevati razlike v spremembah izražanja na nivoju mRNA med visoko in nizko producirajočimi kloni.
- Z DNA-mikromrežami bo možno identificirati gene, ki so ključni za stabilno izražanje rekombinantnega gena (transgena).
- S pomočjo orodij za genetsko karakterizacijo bomo v zgodnji fazi celičnega razvoja lahko predvideli kateri kloni so visoko produktivni in predvsem produkcijsko stabilni skozi daljše časovno obdobje.
- Mesto integracije rekombinantnega gena je pomembno za stabilno izražanje rekombinantnega gena (transgena).
- Amplificirani rekombinantni geni (transgeni) locirani blizu telomernih regij kromosomov so bolj stabilni.

3 MATERIALI IN METODE

3.1 MATERIAL

Vse poskuse smo izvedli na suspenzijskih kulturah celic CHO. CHO-*Der* celična linija je bila kupljena pri Evropski zbirki celičnih kultur (ECACC).

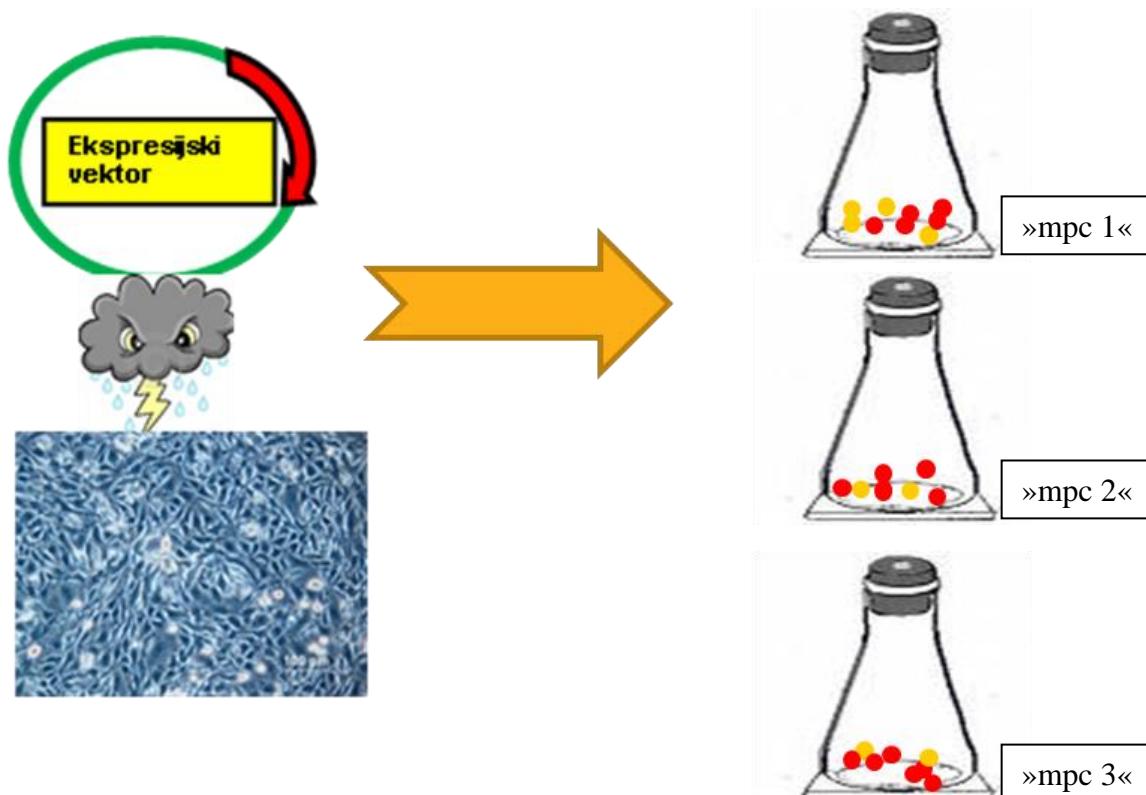
V študiji uporabljena CHO-*Der2* gostiteljska celična linija je bila razvita v LEK d.d., Mengeš iz CHO-*Der* celične linije brez genetskih modifikacij. CHO-*Der2* celično linijo smo prilagodili na brez-serumsko gojišče. V nadaljevanju smo v gostiteljsko celično linijo CHO-*Der2* s transfekcijo vnesli ekspresijski vektor, ki je vseboval rezistenco za antibiotik (geneticin), modificiran gen *Dhfr* in nukleotidni zapis za rekombinantno učinkovino. Iz tako pridobljenih »mpc« smo glede na njihovo produktivnost in kvaliteto rekombinantnega proteina, izbrali najbolj primerno »mpc« za nadaljni razvoj.

3.2 UPORABLJENE METODE

3.2.1 Transfekcija

Za vnos nukleinskih kislin v celice smo uporabili metodo transfekcije z elektroporacijo membran, kar ob uporabi ustreznih reagentov omogoča vnos nukleinske kisline v jedro. Pred transfekcijo smo količino in viabilnost celic določili z napravo ViCell (Beckman Coulter) po navodilih proizvajalca. ViCell (Beckman Coulter) omogoča avtomatizirano določanje viabilnosti celic s pomočjo barvila tripan modro, ki prodira samo v mrtve celice in jih takoobarva. Za transfekcijo je bila potrebna več kot 90 % viabilnost celic. V centrifugirko smo prenesli volumen celične kulture, ki vsebuje 5×10^6 celic. Kulturo smo centrifugirali ($180 \times g$, 5 min pri sobni temperaturi), odstranili gojišče in celičnim peletom dodali po $100 \mu\text{l}$ raztopine za transfekcijo in $3 \mu\text{g}$ plazmidne DNA, ki smo jo pred transfekcijo linearizirali z uporabo restrikcijske encima *SwaI*. Za transfekcijo z elektroporacijo membran smo uporabili Cell Line Nucleofector Kit (Lonza), po navodilih proizvajalca. Po transfekciji z elektroporacijo membran, smo z uporabo sterilne pipete

celice prenesli v 125 ml stresalno steklenico, (Sl. 2), napolnjeno s 15 ml gojišča, in jih gojili v inkubatorju pri 37 °C, v atmosferi z 10 % CO₂ in s stresanjem pri 110 obr. / min.



Slika 2: Prikaz razvoja mešane populacije celic.

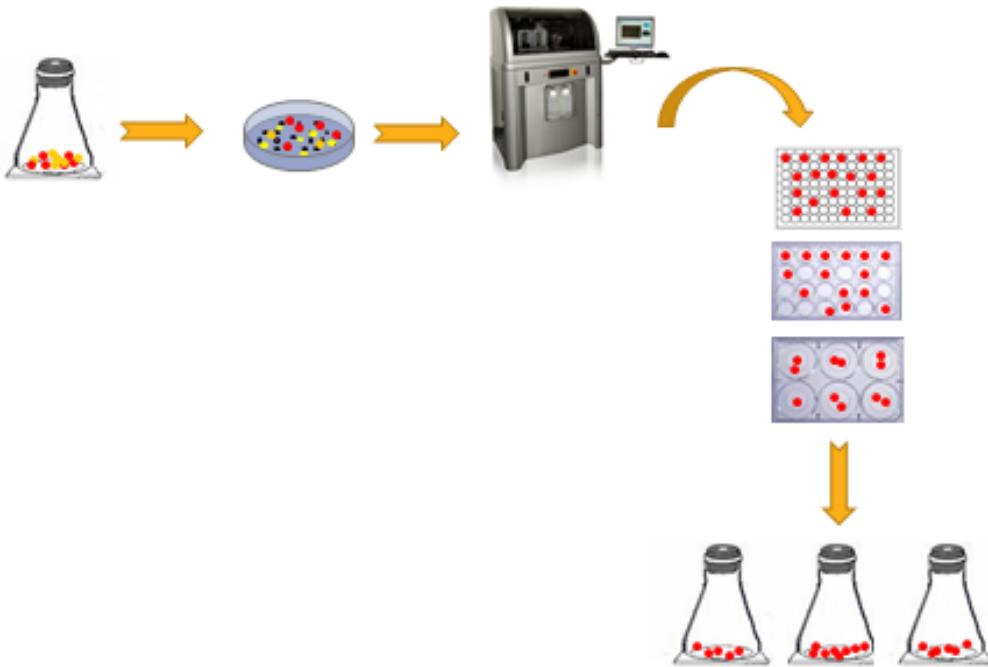
Ekspresijski vektor, ki nosi nukleotidni zapis za rekombinanten protein smo s pomočjo transfekcije vnesli v genom gostiteljske celice CHO. Zaradi različnega mesta integracije ekspresijskega vektorja v genom, je izražanje in produkcija rekombinantnega proteina različna, ne samo med različnimi »mpc-i«, ampak zaradi heterogene populacije tudi znotraj samega »mpc-a«.

Figure 2: The pool development principle.

The expression vector with nucleotide sequence for recombinant protein was transfected in CHO genome. Because of the random integration site of the expression vector in the CHO genome, the expression of the recombinant protein varies between pools.

3.2.2 Kloniranje transficiranih celic CHO

V nadaljevanju smo izbrani »mpc« redčili tako, da smo vzgojili klone iz posameznih transficiranih celic »mpc«. Pri pripravi klonov smo izbran »mpc« gojili od 4 do 5 dni, nato pa prenesli po 50 celic / ml v 100 ml metil-celuloznega poltrdega gojišča z dodanim rastnim gojiščem (40 ml poltrdega gojišča + 60 ml rastnega gojišča) in rahlo premešali. 2ml gojišča s celicami smo prenesli v posamezno jamico mikrotitrske ploščice s šestimi jamicami in celice gojili deset dni v inkubatorju pri 37 °C in v atmosferi z 10 % CO₂. Plošče smo pustili nedotaknjene 7 dni, nato smo kolonije pregledali pod svetlobnim mikroskopom. Po 10 dneh smo posamezne kolonije iz poltrdega gojišča z robotom ClonePix FL (Genetix) prenesli v mikrotitrske ploščice s 96 jamicami. Te smo prej napolnili s po 100 µl gojišča. Pri izbiri kolonij za prenos, smo bili pozorni na velikosti in obliko kolonij, ter na razdaljo med posameznimi kolonijami. S tem smo zagotovili monoklonalnost pripravljenih linij. Pred prenosom kolonij, smo v polrdo gojišče dodali fluorescenčno označena protitelesa, ki so se vezala na ciljni protein. Več kot je posamezna kolonija proizvedla proteina, več flourescence smo zaznali. Fluoresenco smo zaznali s pomočjo kamere, ki je del robotskega sistema ClonePix FL (Genetix). Tarčno smo prenesli samo kolonije, ki so uspešno proizvajale naš protein. Plošče s prenesenimi kolonijami smo gojili v inkubatorju pri 37 °C in v atmosferi z 10 % CO₂ brez stresanja. Po treh dneh smo v vsako jamico dodali po 50 µl svežega gojišča. Preraščenost površine smo spremljali z aparaturom Clone Select Imager (Genetix). Ko so posamezni kloni dosegli 20 % preraščenost površine, smo jih prenesli v mikrotitrske ploščice s 24 jamicami s svežim gojiščem. Po 20 % preraščenosti površine v mikrotitrski ploščici s 24 jamicami, smo celice prenesli v ploščico s 6 jamicami, ki je vsebovala sveže gojišče. Nazadnje smo celice prenesli v 125 ml stresalne plastenke (Sl. 3).



Slika 3: Princip kloniranja transficiranih celic CHO.

Celice izbranega »mpc-a« smo redčili do posameznih celice in jih nacepili v poltrdi medij. Iz poltrdega medija smo z robotom v ploščico s 96 jamicami prenesli samo tiste klone, ki so producirali rekombinantni protein in jih zaporedno prenašali v ploščice z vedno večjimi jamicami, ter nato vsak klon prenesli v ločeno stresalno plastenko.

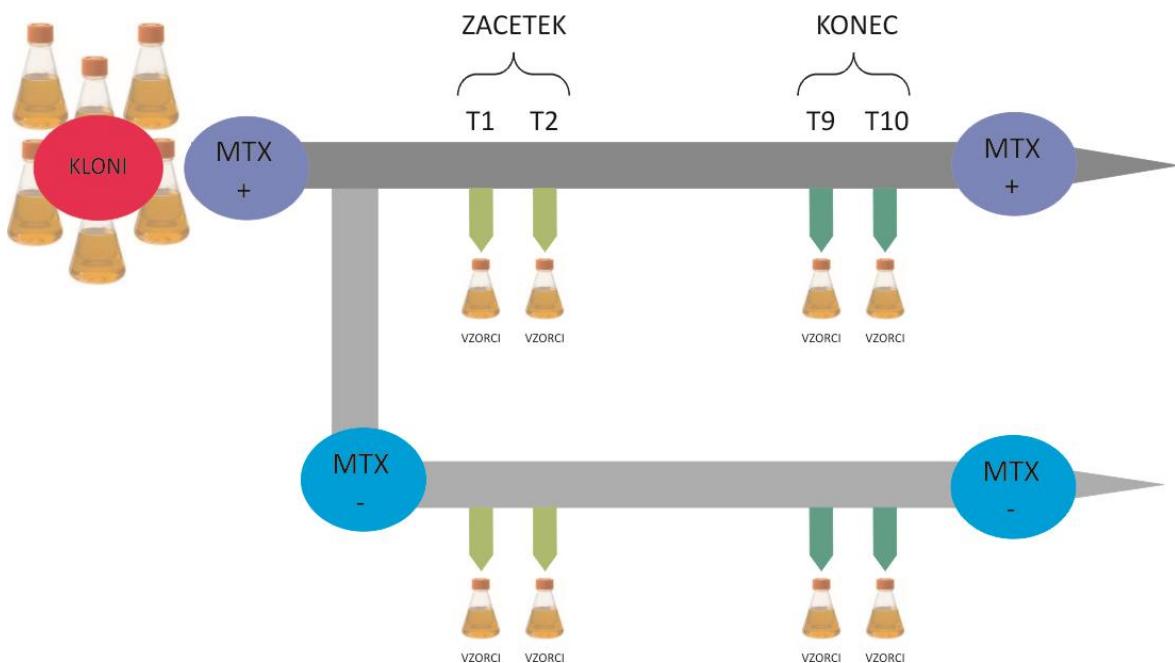
Figure 3: The cloning of transfected CHO cells.

The cells from selected pool were diluted and seeded in semi-solid medium. Only producing colonies were transferred by robot from semi-solid medium to 96 well plates. The individual clones were transferred to larger wells and finally to shake flasks.

3.2.3 Gojenje pripravljenih rekombinantnih klonov

Za namen proučevanja izražanja rekombinantnega gena skozi daljše časovno obdobje smo izbrali 6 klonov, ki smo jih 10 tednov gojili v stresalnih plastenkah. Na začetku 10 tednov dolgega poskusa smo vseh 6 klonov razdelili v 2 paralelki, eni paralelki smo v gojišče dodali MTX (Sigma), drugo paralelko smo gojili pod enakimi pogoji ($37\text{ }^{\circ}\text{C}$, 110 obr. / min, 10 % CO_2) brez MTX v gojišču. Dvakrat tedensko smo celični kulturi nadomestili staro gojišče s svežim gojiščem, tako da je nacepitvena gostota celic znašala $2.0 - 3.0 \times 10^5$ celic / ml. Pred menjavo gojišča smo najprej odvzeli vzorec kulture (1 ml), ki smo ga uporabili za štetje celic z ViCell (Beckman Coulter) aparatom. Glede na izmerjeno število celic v vzorcu, smo odstranili toliko izrabljenega gojišča (skupaj s celično kulturo), da je ostalo dovolj kulture za nacepitveno gostoto $2.0 - 3.0 \times 10^5$ celic / ml. Odvzeto gojišče smo

nadomestili s svežim gojiščem. Ob menjavi gojišča smo v tednih 1, 2, 9 in 10 (v začetku eksponentne faze rasti) odvzeli celice za namen izolacije gDNA in celokupne RNA (Sl. 4). Vzorce celične kulture za merjenje produktivnosti vzorcev smo vzeli v tednih 1, 2 in 9, 10 (natančnejši opis v poglavju 3.2.4)



Slika 4: Shematski prikaz vzorčenja.

Izbrane klone smo 10 tednov gojili v 2 paralelkah (z in brez MTX v gojišču), kjer smo ob menjavi gojišča iz vsakega vzorca v tednih 1, 2, 9 in 10 odvzeli vzorec celic za izolacijo gDNA in celokupne RNA, ter za merjenje produkcijskega potenciala.

Figure 4: Experimental setup.

Selected clones were grown 10-weeks in parallel with/without MTX in the media. The samples for gDNA / total RNA isolation and productivity were taken from each sample on weeks 1, 2, 9 and 10.

Glede na rezultate rasti in produkcijo rekombinantnega proteina med 10 tednov dolgo študijo, smo na osnovi rastnih lastnosti in produktivnosti izbrali 6 klonov za analizo transkriptoma z DNA-mikromrežami. Ker smo želeli transkriptom celic analizirati pri klonih z različnimi rastnimi in produkcijskimi karakteristikami, smo izbrali klone, ki so imeli visoko / nizko in stabilno / nestabilno produkcijo rekombinantnega proteina skozi celotno 10-tedensko študijo. Za analizo z DNA-mikromrežami smo pri vsakem klonu odvzeli 4 vzorce RNA v tednu 1, 2, 9 in 10 iz paralele z MTX v gojišču in 4 vzorce na klon brez MTX v gojišču. Transkriptom celic smo z DNA-mikromrežami tako analizirali na 48 vzorcih.

3.2.4 Producija rekombinantnega proteina

Za merjenje proizvodnje rekombinantnega proteina smo celično kulturo klonov v tednih 1, 2, 9 in 10 nacepili v sveže gojišče (nacepitvena gostota celic $2,0 \times 10^5$ celic / ml) in kulturo brez menjavanja gojišča gojili 10 dni (37°C , 10 % CO_2). Po 10 dnevih smo iz kulture odvzeli vzorec kulture (1 ml) in izmerili proizvodnjo rekombinantnega proteina na aparaturi Octet (Forte Bio). Aparatura Octet kvantificira rekombinantno učinkovino kot je monoklonalno protitelo (mAb) s pomočjo bio-senzorja obdanega s proteinom-A.

3.3 IZOLACIJA DNA IN RNA

3.3.1 Izolacija genomske DNA (gDNA)

Genomsko DNA (gDNA) smo izolirali z uporabo kompleta regentov Blood & Cell Culture DNA Mini Kit (Qiagen) iz 5×10^6 celic, ki smo jim odstranili gojišče (centrifugiranje 10 min pri 4°C in $1500 \times g$) in sprali s PBS pufrom. Izolacijo smo izvedli, kot je opisano v navodilih proizvajalca. Po izolaciji smo izmerili koncentracijo na spektrofotometru NanoDrop 1000 (NanoDrop Technologies). Razmerji 280 nm / 260 nm in 230 nm / 260 nm sta nam služili za oceno čistosti DNA. DNA smo po izolaciji razredčili na koncentracijo 100 ng / μl z vodo brez nukleaz, in jo shranili na -20°C . Tako pripravljeno gDNA smo uporabili za analizo s kvantitativno verižno reakcijo PCR v realnem času (qPCR).

3.3.2 Izolacija celokupne RNA (totRNA) in analiza kvalitete izolacije

Celokupno RNA (totRNA) smo izolirali z uporabo kompleta regentov RNeasy Mini Kit (Qiagen). Za izolacijo smo uporabili 1×10^7 celic, ki smo jim odstranili gojišče (centrifugiranje 10 min pri 4°C in $1500 \times g$) in sprali s PBS pufrom. Celične pelete smo najprej inkubirali v lizirnem pufru, sledila pa je mehanska obdelava v pufru s kolono QIAshredder. Po mehanski obdelavi smo izolacijo nadaljevali, kot je opisano v navodilih proizvajalca. RNA smo eluirali iz kolone v vodi brez nukleaz. Po izolaciji smo izmerili koncentracijo na spektrofotometru NanoDrop 1000 (NanoDrop Technologies). Razmerji 280 nm / 260 nm in 230 nm / 260 nm sta nam služili za oceno čistosti RNA.

Integriteto in koncentracijo RNA smo vzporedno z NanoDrop preverili tudi s kapilarno elektroforezo na Bioanalizator 2100 (Agilent Technologies). Uporabili smo komplet regentov Agilent RNA 6000 Pico Kit (Agilent Technologies) tako, da smo RNA ustrezno redčili, analizo pa smo izvedli, kot je opisano v navodilih proizvajalca. Določanje integritete temelji na detekciji vrhov za ribosomski RNA 18S in 28S. Bioanalizator z algoritmom izračuna vrednost RIN (*angl.* RNA integrity number), ki je numerično merilo za določitev integritete RNA z razponom lestvice od 1 do 10, pri kateri vrednost 10 ustreza intaktni RNA in 1 popolnoma razgrajeni RNA. Za nadaljnje analize smo uporabili RNA z $RIN \geq 9,0$.

Izolirano RNA smo shranili pri -80°C . Tako pripravljeno celokupno RNA smo nato obdelali z DNazo in uporabili za analizo z DNA-mikromrežami in analizo s qPCR.

3.3.3 Odstranitev genomske DNA

Pri izolaciji RNA se vedno hkrati izolira še nekaj genomske DNA, ki jo je pred nadaljnji analizami izražanja genov treba odstraniti. Če DNA zaznamo v območju, kjer kvantificiramo proučevani gen, je določitev ravni izražanja izbranega gena napačna. Prisotnost genomske DNA v vzorcu določimo pred reakcijo reverzne transkripcije z analizo qPCR z uporabo za gen specifičnih začetnih oligonukleotidov. V tem primeru detektiramo le prisotno DNA, saj DNA polimeraza ne deluje z RNA kot osnovno matrico.

Genomsko DNA smo iz vzorca izolirane RNA odstranili z DNazo (Ambion). Priprava reakcijskih mešanic in potek inkubacije sta opisana v preglednici 1 in preglednici 2.

Preglednica 1: Priprava reakcijskih mešanic za obdelavo z DNazo.

Table 1: Reaction mix for treating with Dna-ase.

Komponenta	Volumen [µl]
totRNA (5 µg)	X
10x DNase pufer	5,0
rDNase	2,5
NFW	42,5 - X
Skupni volumen	50,0

Preglednica 2: Potek inkubacij pri obdelavi RNA z DNazo.

Table 2: Incubation settings for DNase.

Temperatura [°C]	Čas [min]
37	38
75	15
4	neskončno

3.3.4 Obratno prepisovanje

Pri postopku obratnega prepisovanja RNA v cDNA smo izolirano celokupno RNA, po obdelavi z DNazo, prepisali v cDNA (komplementarna DNA), ki smo jo uporabili kot matrico za kvantifikacijo s qPCR. Stopnja reverzne transkripcije je najbolj kritična za občutljivo in zanesljivo kvantifikacijo (Ståhlberg in sod., 2004). Za prepis smo uporabili komplet reagentov SuperScript VILO (Thermo Scientific) z naključnimi začetnimi oligonukleotidi, ki omogočajo prepisovanje RNA celotnega nukleotidnega zaporedja matrične RNA. Priprava reakcijskih mešanic za prepis v cDNA je opisana v preglednici 3, potek inkubacije pa v preglednici 4. Verižna reakcija s polimerazo je potekala v aparaturi Eppendorf Mastercycler (Eppendorf). Po končani inkubaciji smo cDNA shranili pri -20°C za nadaljnjo kvantifikacijo s qPCR.

Preglednica 3: Priprava reakcijskih mešanic za prepis v cDNA.

Table 3: Reagent mix for cDNA.

Komponenta	Volumen [μ l]
totRNA obdelana z DNazo	5,0
5 × VILO reaction mix	4,0
10 × SuperScript Enzyme Mix	2,0
NFW voda	9,0
Skupni volumen	20,0

Preglednica 4: Potek inkubacij za prepis v cDNA.

Table 4: Incubation settings for cDNA.

Temperatura [°C]	Čas [min]
25	10
42	90
85	5
4	neskončno

3.4 ANALIZA TRANSKRIPTOMA Z DNA-MIKROMREŽAMI

3.4.1 Priprava in hibridizacija vzorca za DNA-mikromreže

Pri študiju transkriptoma transgenih CHO celičnih linij smo uporabili DNA-mikromreže (Affymetrix), ki so sestavljene iz 61,223 oligonukleotidnih sond, ki predstavljajo približno 26,227 genov kitajskega hrčka (*Cricetulus griseus*) in 14,657 genov miške (*Mus musculus*). Za DNA-mikromreže smo uporabili celokupno RNA, ki smo jo uporabili tudi za RT-qPCR analizo.

Pred prepisom celopukne RNA v cDNA smo celokupno RNA redčili na enako koncentracijo (5 ng / ul). Z biotinom označena cRNA je bila sintetizirana po navodilih proizvajalca (Affymetrix). Hibridizacija je potekala v GeneChip hibridizacijski pečici

(GeneChip Hybridisation Oven 640, Affymetrix). Nadaljna obdelava DNA-mikromrež je potekala v GeneChip Fluidics postaji po navodilih proizvajalca (Affymetrix).

3.4.2 Obdelava in analiza podatkov iz DNA-mikromrež

DNA-mikromreže smo slikali na GeneArray (Affymetrix). Po slikanju DNA-mikromrež smo originalne slikovne dokumente izvozili v GeneSpring GX (Agilent Technologies) programski paket za obdelavo podatkov. Uvožene podatke smo v GeneSpring Gx programu normalizirali s pomočjo algoritma za izračun povprečja med različnimi DNA-mikromrežami (Bolstad in sod., 2003; Irizarry in sod., 2003). Vse nadaljnje statistične analize smo naredili v programskem okolju R (www.bioconductor.org) in uporabili programski paket *limma* (Smyth, 2005).

Z namenom zmanjšanja lažno pozitivnih statistično značilnih rezultatov smo iz podatkov najprej odstranili neizražene gene (geni pri katerih se izraženost pri 80 % vzorcev ni razlikovala od ozadja). Za iskanje statistično značilno različno izraženih genov med različnimi vzorci smo uporabili empirični Bayesov model (Smyth, 2005).

3.5 PCR V REALNEM ČASU

3.5.1 Analiza izražanja genov

Za namen potrditve rezultatov DNA-mikromrež smo na vseh vzorcih (6 klonov v 4 časovnih točkah, z in brez MTX v gojišču) skupno preverili izražanje 36 genov z uporabo PCR v realnem času (qPCR). PCR v realnem času je nadgradnja PCR, saj omogoča detekcijo nastanka produktov PCR med potekom reakcije. Detekcija poteka na podlagi merjenja naraščanja fluorescence, nastajanje produktov pa sledimo s fluorescenco, kjer je količina produkta PCR sorazmerna izmerjeni fluorescenci. Fluorescenco merimo ves čas poteka reakcije, zato lahko izrišemo krivuljo, ki podaja odvisnost jakosti fluorescence od števila ciklov. Končni rezultat je podan kot časovna točka oz. število ciklov, pri katerem

prvič zaznamo fluorescenco (Bustin in sod., 2005; Ginzinger in sod., 2002; Kubista in sod., 2006; Wilhelm in sod., 2003). Takšen pristop omogoča natančno kvantitativno analizo nukleinskih kislin. Poleg tega so prednosti qPCR še široko dinamično območje detekcije ter povečana občutljivost in natančnost. Zaradi načina detekcije fluorescence je mogoče pomnoževanje kratkih odsekov DNA (od 70 do 150 baznih parov), kar zagotavlja večjo učinkovitost pomnoževanja. Sonda, uporabljeni v reakciji PCR, omogoča večjo specifičnost reakcije PCR (Ginzinger in sod., 2002; Wilhelm in sod., 2003; Wiseman in sod., 2002). Specifičnost in občutljivost metode qPCR sta močno odvisni od tipa kemije, na kateri temelji metoda qPCR (Nikolić in sod., 2010).

3.5.2 Izbor genov in načrtovanje njihovih oligonukleotidnih začetnikov/sond

Za načrtovanje amplikonov za proučevanja izražanja genov, smo uporabili identične nukleotidne sekvence genov, kot so bile uporabljene za izdelavo sond na DNA-mikromrežah. Nukleotidne sekvence smo preverili *in silico* s primerjavo znanih zaporedij v bazah podatkov (GenBank) z uporabo programa Basic Local Alignment Tool (BLAST), (NCBI, 2007). Oblikovanje začetnih oligonukleotidov je temeljilo na osnovi znanega nukleotidnega zaporedja uporabljenega pri zasnovi sond za DNA-mikromreže. Sintezo začetnih oligonukleotidov in sond smo naročili pri podjetju Thermo Scientific s pomočjo programa FileBuilder (Thermo Scientific). Za doseganje večje specifičnosti smo za analize qPCR uporabljali kemijo TaqMan. Uporabili smo TaqMan MGB-sonde, kar pomeni, da je bila sonda na koncu 5' označena z barvilom FAM, na koncu 3' pa z nefluorescenčnim dušilcem MGB (minor groove binder). Sonde DNA z ligandi MGB tvorijo z enoverižno molekulo DNA stabilne duplekse in tako povečujejo temperaturo taljenja oligonukleotidov v primerjavi z nemodificiranimi sondami DNA (Afonina in sod., 1997). Posledično so lahko sonde MGB krajše od navadnih sond TaqMan, s čimer omogočajo izdelavo krajših amplikonov, ki so lahko še bolj specifični.

Ko smo od proizvajalca (Thermo Scientific) dobili zaporedja začetnih oligonukleotidov / sond, smo tudi te preverili *in silico* z BLASTn algoritmom, prilagojenim za krajša zaporedja (Altschul in sod., 1990).

3.5.3 Izvedba reakcij PCR v realnem času

Analize so potekale na aparaturah ABI Prism 7900HT Sequence Detection System in 7900HT Fast Real-Time PCR System (Thermo Scientific). Vse reakcije smo izvedli pri univerzalnih pogojih pomnoževanja (*angl.* universal cycling conditions): 2 min pri 50 °C in 10 min pri 95 °C (aktivacija polimeraze) ter nato 45 ciklov pri pogojih: 10 s pri 95 °C in 1 min pri 60 °C, v optičnih ploščicah formata 384 (384 Well Clear Optical Reaction Plates, Thermo Scientific), ki smo jih prekrili z optičnimi adhezivnimi folijami (Optical Adhesive Covers, Thermo Scientific).

Volumen reakcij je bil 10 µL, 3 µL DNA in 7 µL reakcijske mešanice. Poleg cDNA je reakcija vsebovala 900 nM vsakega izmed oligonukleotidnih začetnikov (smiselnega in protismiselnega), 250 nM sonde (20-kratna mešanica sonde in začetnih oligonukleotidov) in še eno polovico volumna master miks reagenta (Pregl. 5). Uporabili smo dvakratni Universal PCR Master Mix (Thermo Scientific). Reagent vsebuje ROX™ pasivno referenčno barvilo, katerega funkcija je normalizacija fluorescence sonde po celi ploščici.

Preglednica 5: Reakcijska mešanica za reakcijo qPCR – določanje števila kopij rekombinantnega gena.

Table 5: Reaction mix for the qPCR- recombinant gene copy number.

Reagent	Končna koncentracija	Volumen [µL]
2-kratni TaqMan Universal Master Mix (Applied Biosystems)	1-kratni	5
20-kratni miks sonde in oligonukleotidnih začetnikov	1-kratni	0,5
ddH ₂ O		1,5
DNA		3
Skupaj		10

V vsako jamico, v optičnih ploščicah formata 384 (384 Well Clear Optical Reaction Plates, Thermo Scientific), smo nanesli 7 µL reakcijske mešanice in dodali 3 µL ustrezno redčene cDNA. Za vsak vzorec smo pripravili tri ponovitve in dve zaporedni 10-kratni redčini. Na 384-ploščici smo za vsak amplikon pripravili še reakcije brez cDNA, ki so služile kot

negativna kontrola (NTC – *angl. non template control*). V te tri jamice smo namesto cDNA nanesli vodo brez nukleaz. Za natančnejše spremljanje značilnosti amplikona (izračun meje detekcije, linearnega razpona meritev in učinkovitosti pomnoževanja) smo na vsako ploščico za vsak amplikon nanesli še en združen vzorec v treh ponovitvah in sedmih zaporednih trikratnih redčenjih (10-kratni do 7290-kratni).

V vsakem vzorcu smo preverili izražanje:

- a) tarčnih genov iz treh različnih sklopov (stabilnost produkcije rekombinantnega proteina skozi čas (stabilnost), obseg produkcije rekombinantnega proteina (produktivnost) in vpliv MTX na produkcijo rekombinante učinkovine (tretiranje z / brez MTX) in
- b) dva kontrolna gena (*Actb* in *Gapdh*), ki smo jih uporabili za preverjanje učinkovitosti izolacije DNA / RNA, preverjanje uspešnosti prepisovanja RNA v cDNA. V procesu kvantifikacije smo jih uporabili kot normalizatorja. Vsak vzorec smo tako testirali s 16, 14 oz. 12 različnimi qPCR reakcijami.

Zaradi velikega števila reakcij qPCR smo vzorce na 384-ploščice nanašali z robotom CAS-1200 (Corbett). Začetno analizo podatkov smo opravili v programu SDS 2.1 (Thermo Scientific). Fluorescence reporterskega barvila program normalizira na barvilo ROX, ki ga vsebuje reakcijska mešanica, na bazno linijo (šum metode) in na fluorescentno ozadje. Normalizirana fluorescencija je podana kot ΔRn : kjer je Rn^+ emisija fluorescence produkta v določenem času in Rn^- emisija pasivnega referenčnega barvila ROX.

$$\Delta Rn = Rn^+ - Rn^- \quad \dots(1)$$

Program za vse jamice 384-ploščice izriše graf naraščanja ΔRn skozi cikle reakcije (krivulje pomnoževanja). Za nastavitev bazne linije (vrednost cikla pred dvigom fluorescence nad vrednost ozadja) smo izbrali avtomatsko opcijo. Nato smo na eksponentnem delu krivulje pomnoževanja določili linijo fluorescenčnega praga. Ta predstavlja intenzitetu fluorescence (ΔRn), ki je značilno višja od ozadja. Točko, pri kateri fluorescencija vzorca (ΔRn) preseže fluorescentni prag, imenujemo cikel kvantifikacije (Cq)

– *angl. quantification cycle*). Po obdelavi podatkov v programu SDS 2.1 smo te izvozili v tekstovno datoteko in to prenesli v program Microsoft Excel in nadaljevali z obdelavo.

3.5.4 Določitev optimalnih parametrov delovanja testov qPCR

Z orientacijskimi testiranjji smo določili lastnosti testov PCR v realnem času za vsak amplikon posebej:

- a) določitev optimalne količine cDNA v reakciji;
- b) določitev linearne razpona meritev;
- c) določitev učinkovitosti pomnoževanja.

Sonde TaqMan MGB in oligonukleotidni začetniki, naročeni pri Thermo Scientific so prišli v obliki enotne raztopine z znanimi koncentracijami (sonde – 250 μM in oligonukleotidni začetniki – 900 μM).

Primerno količino cDNA v reakciji smo določili tako, da smo izbrali 3 vzorce, ki so imeli srednjo izmerjeno količino RNA. Pripravili smo zaporedne 3-kratne redčine z razponom redčitvenih faktorjev od 10 x do 7290 x in vsako testirali z vsakim amplikonom v treh ponovitvah in tako dobili umeritvene krivulje. Izris umeritvene krivulje (vrednosti Cq v odvisnosti od \log_{10} začetne količine tarčne DNA) ne potrebuje absolutne začetne količine tarčne DNA, ampak si postavimo umetne količine, pomembno je le, da se začetna količina DNA med 3-kratnimi redčinami zmanjšuje s faktorjem 3.

Učinkovitost reakcije qPCR smo izračunali po formuli (Pfaffl, 2001): kjer je (k) naklon linearne regresijske premice med logaritmiranimi vrednostmi količine tarče in vrednostmi Cq. Definirana je kot odstotek pomnoževanja tarčne molekule cDNA (0 % – pomnoževanja ni bilo oz. 100 % – vse molekule tarčne molekule cDNA se v enem ciklu reakcije PCR podvojijo). Stoodstotna učinkovitost ustreza naklonu $k = -3.33$.

$$E = (10^{[-1/k]} - 1) \times 100 \quad \dots(2)$$

Določili smo tudi linearni razpon meritev in koreacijski koeficient (R^2) samo v linearinem razponu meritev, to je razpon koncentracij oz. faktorjev redčitve, pri katerih so vrednosti Cq v linearinem odnosu z \log_{10} koncentracijami tarčne cDNA. Za vsak amplikon smo izračunali naslednje parametre: naklon krivulje – k (iz diagrama odvisnosti vrednosti Cq od \log_{10} začetne količine tarčne cDNA), učinkovitost pomnoževanja – E in linearni regresijski koeficient – R^2 .

Vsakemu amplikonu smo z umeritveno krivuljo referenčnega vzorca, sestavljenega iz vsaj deset naključno izbranih vzorcev, izračunali učinkovitost pomnoževanja in tako kontrolirali uspešnost reakcije PCR. Vsakemu vzorcu smo nato izračunali njegovo relativno količino izraženega gena (npr. *Spn1*). Izbrali smo t. i. kalibratorski vzorec, na katerega smo primerjali vse preostale vzorce in normalizirali ekspresijo proučevanega gena na vzdrževalni gen, katerega izražanje se v vzorcih ne spreminja (npr. *Actb*, *Gapdh*). Kalibratorski vzorec smo izbirali na podlagi meje, ki določa kontrolo kvalitete izračunov relativnega izražanja. Gre za merilo podobnosti učinkovitosti pomnoževanj med proučevanimi geni (tarčni geni) na eni strani in normalizatorju na drugi strani. Pri vsakem vzorcu smo izračunali razliko povprečij vrednosti Cq med višjo in nižjo redčino vseh genov (ΔCq). Nadalje smo izračunali vrednost $\Delta\Delta Cq$ za vsak vzorec pri vseh tarčnih genih po formuli 5. Če vrednost $\Delta\Delta Cq$ ni presegala meje $\Delta\Delta Cq < 0,5$ pri vseh tarčnih genih, je vzorec lahko kandidiral za t.i. kalibratorski vzorec.

Izražanje genov smo normalizirali glede na geometrijsko povprečje dveh vzdrževalnih genov *Actb* in *Gapdh* ter tako dobili izračunani normalizator AcGa. Geometrijsko povprečje več vzdrževalnih genov naj bi bilo dobro orodje za normalizacijo (Vandesompele in sod., 2002). Po analizi podatkov v programu SDS 2.1 in njihovem prenosu v program Microsoft Excel smo za vsak gen vsakega vzorca imeli po šest vrednosti Cq , pri čemer so tri pripadli eni redčini, tri pa za faktor 10 višji redčini. Izračunali smo povprečne vrednosti Cq vseh treh ponovitev za obe redčini. Nato smo izračunali učinkovitost pomnoževanja po enačbi 3 in za tem še relativno ekspresijo gena po enačbi 4, povzeti po Pfaffl in sod. (2002).

$$E = 10^{(-1/\Delta Cq)} \quad \dots(3)$$

Stoddotna učinkovitost pomnoževanja se v tem primeru nanaša na vrednost E=2. Izračunali smo jo za vsak vzorec posebej (razlika povprečij vrednosti Cq obeh redčin – ΔCq), in ne iz umeritvene krivulje enega, združenega vzorca, kot je to opisano v Pfaffl (2001) oz. Pfaffl in sod. (2002). Tako smo se izognili predvidevanju, da je E za vsak vzorec enaka.

$$R = \frac{E_{\text{tarča}}^{[\text{povprCq(tarča kalibrator)} - \text{povprCq(tarča vzorec)}]} }{E_{\text{normalizator}}^{[\text{povprCq(naormalizator kalibrator)} - \text{povprCq(normalizator vzorec)}]}} \dots (4)$$

Izračun relativne kvantifikacije za vsak gen v vsakem vzorcu posebej temelji na primerjavi učinkovitosti pomnoževanja tarčnega gena (npr. *Spn1*) v natančno določenem vzorcu glede na kalibratorski vzorec z vzdrževalnim genom (AcGa) istega vzorca, glede na kalibratorski vzorec. Tarča se nanaša na proučevani gen (npr. *Spn1*), ki ga želimo kvantificirati, normalizator pa na geometrijsko povprečje dveh vzdrževalnih genov, Actb in Gapdh. Na koncu dobimo za vse vzorce podatke o relativni količini izraženega gena, ki jih lahko med seboj primerjamo. Dobljene vrednosti R smo \log_2 transformirali.

Dodatno smo pri obdelavi podatkov upoštevali tudi kontrolo kakovosti relativnega izražanja. Uvedli smo merilo podobnosti učinkovitosti pomnoževanj tarčnega gena in normalizatorja: za vsak vzorec smo izračunali razliko med ΔCq tarčnih genov in normalizatorja po formuli 6 in določili mejo merila $\Delta\Delta Cq < 0,5$. Če je vrednost $\Delta\Delta Cq$ za posamezni gen določenega vzorca presegala mejno 0,5 smo vzorec izločili iz nadaljnjih statističnih analiz.

$$\Delta\Delta Cq_{\text{tarča}} = \Delta Cq_{\text{vzorec tarča}} - \Delta Cq_{\text{vzorec normalizator}} \dots (5)$$

Izračun merila podobnosti učinkovitosti pomnoževanj med tarčnimi geni in geni normalizatorja. Vse obdelave smo opravili v odprtokodno programskem okolju R (R Development Core Team, 2005).

3.5.5 Statistične analize podatkov

3.5.5.1 Test t

V programskem okolju R (R Development Core Team, 2005) smo z Welchevim dvo-temim testom t znotraj posameznega sklopa analizirali po dve skupini neodvisnih vzorcev:

- a) vzorce s stabilno in nestabilno produkcijo rekombinantnega proteina skozi čas (stabilnost);
- b) vzorce z visoko in nizko produkcijo rekombinantnega proteina (produktivnost);
- c) vzorce z in brez tretiranja z MTX.

Za vsak sklop smo izračunali, ali so bile razlike v povprečjih relativnega izražanja posameznega gena med neodvisnimi skupinami vzorcev statistično značilne (popravljena P-vrednost < 0,05).

3.5.5.2 Razvrščanje in analiza PCA

Na osnovi rezultatov analize izražanja genov med kloni s stabilno in nestabilno produkcijo rekombinantnega proteina, smo izbrali tri gene za tri-dimenzionalno vizualizacijo. Vsaka izmed treh osi grafa, zajema izražanje enega izmed treh genov med vsemi vzorci in tako omogoča vizualizacijo v tri-dimenzionalnem prostoru. Za sistematično oceno ločevanja med kloni s stabilno in nestabilno produkcijo rekombinantnega proteina, smo uporabili algoritem najbližjega soseda (Jackson, 2003). Cilj uporabe algoritma je ločevanje podatkov v definirano število razredov brez predhodne delitve vzorcev v posamezne razrede. Algoritem minimalizira vsoto kvadratov razdalj med vzorci in centralnimi točkami razredov. Na osnovi tega se oceni najverjetnejša centralna točka med razredom in vzorcem.

Podatke relativne kvantifikacije, iz sklopa stabilnosti produkcije rekombinantnega proteina, smo analizirali tudi z metodo glavnih komponent (PCA). Namen te metode je opisati dani sistem z manjšim številom spremenljivk, ki so linearne kombinacije izvornih spremeljivk. Tako znižamo dimenzionalnost podatkov, hkrati pa ohranimo čim večjo količino informacij iz izhodnih podatkov. Prva glavna komponenta – PC1 vsebuje največ

variabilnosti, druga glavna komponenta – PC2, ki je pravokotna na PC1, nekoliko manj, tretja in naslednje komponente še manj. Tako vse podatke reduciramo v manj dimenzionalni prostor in pri tem izgubimo le malo informacij. Analizo PCA smo uporabili na petih (*E130203B14Rik*, *BX842664.2/Hist1h3c*, *Ptpre Cspg4*, in *Fgfr2*) najbolj različno izraženih genih. Analiza za tri-dimenzionalni prikaz in analiza PCA sta bili narejeni s pomočjo MATLAB2014 (The Mathworks Inc.) programa.

3.5.6 Analiza števila kopij rekombinantnega gena

Za določitev števila kopij tarčnega gena smo uporabili absolutno kvantifikacijo. Absolutno kvantifikacijo uporabljamo, kadar nas zanima točno število kopij tarčnega nukleotidnega zaporedja. Pri tem načinu analize odčitamo število kopij nukleotidnega zaporedja v vzorcu iz standardne krivulje, ki jo pridobimo z zaporednim redčenjem standarda znane koncentracije. Standardna krivulja kaže linearno razmerje med vrednostmi Cq in logaritemskimi vrednostmi števila kopij standardne DNA. V programu SDS 2.1 smo posameznim točкам standardnih krivulj pripisali pripadajoče vrednosti, kar je omogočilo poznejše odčitanje števila kopij tarčnega nukleotidnega zaporedja v vzorcih. Po obdelavi podatkov v programu SDS 2.1 smo te izvozili v tekstovno datoteko in jo prenesli v program Microsoft Excel in obdelavo nadaljevali.

Standardne krivulje smo pripravili iz genomske DNA, izolirane iz celic CHO *Der2*. Pripravili smo jih v petih razredčitvenih točkah (koncentracij 10 ng / μL , 2 ng / μL , 1 ng / μL , 0,2 ng / μL in 0,1 ng / μL), vsako točko v treh paralelkah. Standardne krivulje smo pripravili za endogeni gen za glukagon (*Gluc*), ki je v genomu celice CHO v eni kopiji, in za tarčne gene. Število kopij genov v reakciji za standardno krivuljo smo izračunali po formuli 6 z upoštevanjem velikosti haploidnega genoma celice CHO.

Izračun števila kopij genomske DNA celic CHO velikosti $2,45 \times 10^9$ bp v 100 ng genomske DNA:

$$\begin{aligned} M \text{ genoma} &= (2,45 \times 10^9 \text{ bp}) \times (618 \text{ g/mol/bp}) = 1,5141 \times 10^{12} \text{ g/mol} \\ m \text{ 1 kopije genoma} &= (1,5141 \times 10^{12} \text{ g/mol}) \times (1 \text{ mol}/6,022 \times 10^{23}) = \\ &= 2,51386 \times 10^{-12} \text{ g} = \\ &= 2,51386 \text{ pg DNA} \end{aligned}$$

$$\text{št. kopij v 100 ng DNA} = 100 * 10^3 \text{ pg}/2,51386 \text{ pg} = 39779 \text{ kopij.}$$

Masa ene kopije genomske DNA celic CHO velikosti $2,45 \times 10^9$ bp je 2,51386 pg.
 100 ng genomske DNA vsebuje 39779 kopij genomske DNA celic CHO.

... (6)

Vzorce smo analizirali v dveh razredčinah (koncentracije 10 ng / μL in 1 ng / μL), vsako v treh paralelkah. Iz standardne krivulje smo odčitali število kopij za endogeni gen *Gluc* in tarčnim genom v vzorcu, ter izračunali razmerje med njima.

Uporabljeni začetni oligonukleotidi in sonde za določitev števila kopij gena so opisani v preglednici 6, priprava reakcij qPCR pa v preglednici 7.

Preglednica 6: Začetni oligonukleotidi in sonde genov uporabljenih za določanje relativnega izražanja in števila kopij genov s qPCR.

Gen	Začetni oligonukleotid [5'-3']	Končni oligonukleotid [5'-3']	Sonda [5'-3']
<i>Rekombinanten gen*</i>	/	/	
<i>Glucagen (Gluc)**</i>	attgc当地aaacggccacgat	ccaagcaatgaattcctttgc	ctgaaggggacccttacca
<i>Fgf2</i>	gcctgagttcacatccatcaca	gatgtgaaggctctgaagctgtta	ttggcctcacatctcc
<i>Bx842664.2/Hist1h3c</i>	ggccc当地agacatggacact	ccatgaggcactggacttt	aagegccccatcagc
<i>Egr1</i>	getcacctctggc当地taagg	cattctggagaaccaaagct	cagetc当地gecccttc
<i>Cspg4</i>	gccatgtgc当地ctgctcat	aaacaggtgagaatagaggactttgg	caagctt当地gaattcc
<i>Hist1h2bc</i>	acgaggaggtagacctgatgatgt	gtatcacctattccatgtctcaattgc	cagtgctggacgttgtt
<i>E130203b14Rik</i>	ccagtgggtacatcacatgagaga	cccgagttgggactgtact	aaactgtgccaactc
<i>Ptpre</i>	ccctccagttcttggctaatg	gcaaaacttgactctgtgtctttagg	ccacaaccaaattcag
<i>Cu459186.17</i>	gggaggccggttttg	ttagcaacacccagagactac	ttgctgccc当地tatcc
<i>Mmp10</i>	caggaaatcgaggccacaaaattgt	tcaaactgtgatgatccatggaaagaa	aatgc当地ctgcaacaccg
<i>Vsnl</i>	acccttaagecatatgtcttggaaatttga	tccgaaatgacaacaatcgctgtt	tcatec当地gccccctcc
<i>Foxp2</i>	gggc当地ttacggcttatactctatgt	cccagttgtgatattctatcaagtacttt	acgggtgcatgatcc
<i>C1qtnf1</i>	cattccacagacactggatgga	gccaaagaagccaggactga	ctgaccccatc当地atcc
<i>hDhfr</i>	ccatctctgagaagaatcgacctt	ggcatcatctaaactttggcaagaa	tctgtgggtggtcttg
<i>Ac115880.11</i>	c当地gacatcaccagtagagatgtt	ttgacacatacagctccaatt-cca	acgggcttc当地agtcttc

»se nadaljuje«

»nadaljevanje preglednice 6. Začetni oligonukleotidi in sonde genov uporabljenih za določanje relativnega izražanja in števila kopij genov s qPCR.«

*Opomba: Sekvence oligonukleotidov in sond za rekombinanti gen so poslovna skrivnost (za izračun števila kopij gena).

*Remark: The sequences for primers and probes for recombinant gene are confidential (for gene copy number calculation).

****Opomba:** Sekvenca oligonukleotidov in sonda uporabljena za izračun števila kopij gena.

**Remark: The sequences for primers and probe used for gene copy number calculation.

Preglednica 7: Priprava reakcijskih mešanic za qPCR – izražanje genov.

Table 7: Reaction mix for qPCR-gene expression.

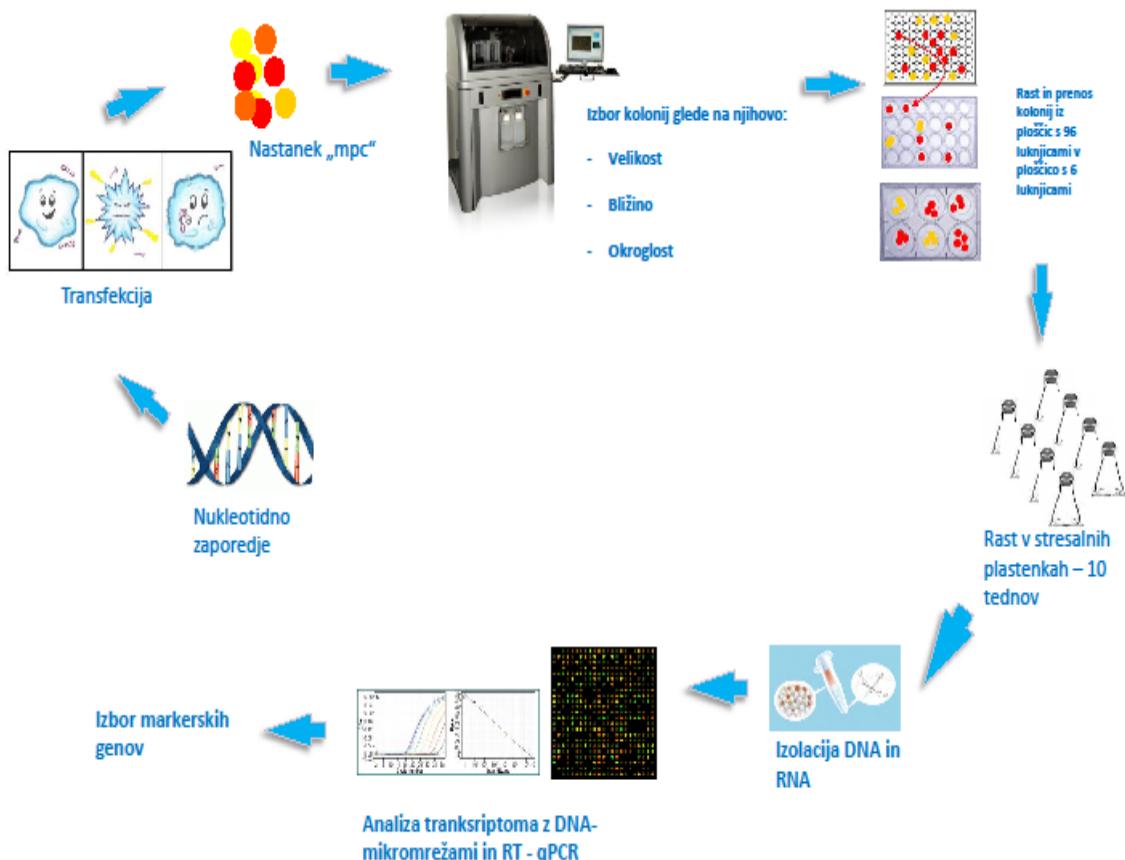
Komponenta	Volumen [µl]
5 x TaqMan Master Mix	5,0
začetni oligonukleotid 3' µM)	(10 0,9
začetni oligonukleotid 5' µM)	(10 0,9
sonda (10 µM)	0,2
matrična DNA	3,0
Skupni volumen	10,0

Za določitev ustreznosti standardnih krivulj smo iz naklona standardne krivulje (k) izračunali tudi učinkovitost pomnoževanja (E) po formuli 2. Določili smo tudi koreacijski koeficient (R^2).

$$E = 10^{[1/k]} - 1 \quad \dots(7)$$

4 REZULTATI

Takoj ko je znana aminokislinska sekvenca rekombinantega proteina, se na osnovi aminokislinske sekvence enoznačno določi nukleotidna sekvenca. Nukleotidna sekvenca za rekombinantno učinkovino se s kloniranjem vstavi v ekspresijski vektor, ki se nato s transfekcijo vnese v gostiteljsko celično linijo. Za potrebe poskusa smo s transfekcijo vnesli DNA, z zapisom za rekombinantno učinkovino, v gostiteljske celice CHO – in tako naredili »mpc«. Iz »mpc« smo izolirali klone z dobrimi produkciskimi in rastnimi lastnostmi. Nastavili smo poskus za spremljanje stabilnosti produkcije na 6 klonih v paraleli z in brez tretiranja z MTX, ki je trajal 10 tednov. Pri vseh klonih smo v tednih 1, 2, 9 in 10 odvzeli vzorec za merjenje produkcije rekombinantnega proteina, izolirali gDNA in celokupno RNA. Iz gDNA smo izračunali število kopij gena za rekombinanto učinkovino, totRNA smo uporabili za analizo transkriptoma z DNA-mikromrežami in potrjevanje izražanja izbranih genov z metodo RT-qPCR (Sl. 5).



Slika 5: Shematski prikaz poteka poskusa od nukleotidnega zaporedja za rekombinanen gen do izbora statistično značilno različno izraženih genov.

Nukleotidni zapis za ciljni protein se s transfekcijo vnese v genom gostiteljske celice. Po selekciji in amplifikaciji smo klone gojili v stresalnih plastenkah, izolirali DNA in RNA za analizo transkriptoma z DNA-mikromrežami in RT-qPCR.

Picture 5: Schematic overview of an experiment – from nucleotide sequence of recombinant gene to selection of marker genes.

Nucleotidic sequence for target protein were transferred in the host cell. After the selection and amplification the clones were grown in shake flasks. DNA and RNA were isolated for the transcriptom analysis by DNA microarrays and RT-qPCR.

4.1 PRODUKCIJA REKOMBINANTE UČINKOVINE IN VPLIV MTX

Producijo rekombinantnega proteina smo izmerili v 4 točkah (teden 1, 2, 9 in 10) med 10-tednov dolgo študijo. Vseh 6 klonov je bilo gojenih z in brez prisotnega MTX v gojišču. Iz rezultatov v preglednici 8 je razviden vpliv MTX na produktivnost in predvsem na stabilnost produkcije rekombinante učinkovine skozi čas. Kadar je MTX prisoten v gojišču, je produktivnost stabilna in se ne zmanjšuje bistveno skozi čas. Skozi celotno

študijo se je produkcija rekombinante učinkovine v primeru stabilnih klonov znižala za 9 % (iz 7 na 6,3 enot). Izmed 6 klonov, smo na začetku študije imeli 4 visoko producirajoče in 2 nizko producirajoča klena (K-1 in K-10). Pri vseh 4 visoko producirajočih klonih (K-50, K-90, K-102, K-115), kjer MTX ni bil prisoten v gojišču, se je produkcija rekombinantnega proteina zmanjšala za 67 % (iz 7 na 2,3 enoti), kar pomeni, da produkcija rekombinantnega proteina ni bila stabilna med 10 tednov dolgo kultivacijo. Pri ostalih 2 klonih, brez prisotnega MTX v gojišču, upada produkcije ni bilo opaziti (Pregl. 8).

Preglednica 8: Produktivnost klonov z in brez MTX v gojišču skozi 10 tednov dolgo študijo.

Table 8: Clone productivity with and without MTX in the media in 10-week long cultivation.

Oznaka klena	K-1		K-10		K-50		K-90		K-102		K-115	
MTX	+	-	+	-	+	-	+	-	+	-	+	-
Vzorčenje	Produktivnost											
Teden 1	nizka	nizka	nizka	nizka	visoka	visoka	visoka	visoka	visoka	visoka	visoka	visoka
Teden 2						nizka						
Teden 9							nizka					
Teden 10								nizka				
Stabilnost	Stabilnost produkcije rekombinannte učinkovine											
	stabilno	ne-stabilno	stabilno	ne-stabilno	stabilno	ne-stabilno	stabilno	ne-stabilno	stabilno	ne-stabilno	stabilno	ne-stabilno

Glede na stabilnost produkcije rekombinantnega proteina skozi časovno obdobje 10 tednov, smo klone razdelili po naslednjih kriterijih:

a) stabilni kloni;

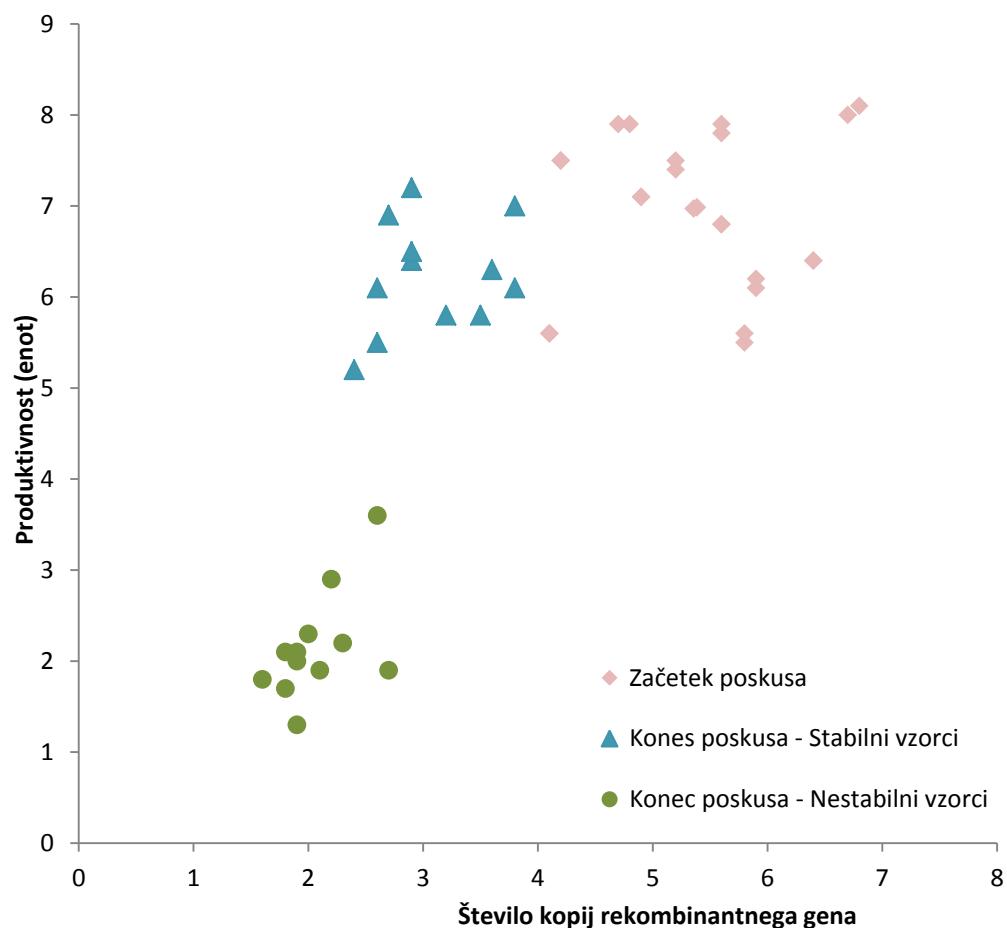
Producija rekombinantnega proteina se ni spremenila za več kot 30 % med povprečjem razlike produktivnosti v tednu 9 in 10 in med povprečjem produktivnostjo v tednu 1 in 2. Ob prisotnosti MTX v gojišču, so imeli vsi kloni stabilno produkcijo rekombinantnega proteina. Če MTX ni bil prisoten v gojišču, sta imela samo 2 klona (K-1 in K-10) stabilno produkcijo rekombinantnega proteina. Le-ta sta imela nizko produktivnost skozi celotno študijo. Klona K-1 in K-10 sta torej nizka producenta, vendar glede na to, da njuna produktivnost rekombinantnega proteina skozi čas ni padala, smo ju uvrstili v skupino vzorcev s stabilno produkcijo rekombinantnega proteina.

b) nestabilni kloni;

Producija rekombinantnega proteina se je spremenila za več kot 30 % med povprečjem razlike produktivnosti v tednu 9 in 10 in med povprečjem produktivnostjo v tednu 1 in 2. Štirje izmed 6 klonov so imeli nestabilno produkcijo rekombinantnega proteina, kadar MTX ni bil prisoten v gojišču.

4.2 PRODUKCIJA REKOMBINANTNEGA PROTEINA IN ŠTEVILO KOPIJ REKOMBINANTNEGA GENA

Za namen opisa lastnosti posameznega klena, smo poleg produkcije rekombinante učinkovine izmerili tudi število kopij vstavljenega gena za rekombinantno učinkovino v 4 časovnih točkah 10 tedenske študije. Producijo in število kopij rekombinantnega gena smo izračunali iz povprečja dobljenih vrednosti na začetku študije (teden 1 in 2) in iz povprečja dobljenih vrednosti na koncu študije (teden 9 in 10). Pri stabilnih klonih se je število kopij rekombinantnega gena v povprečju znižalo za 45 % v 10-tedenskem obdobju študije, v primerjavi z 61 % upadom števila kopij pri nestabilnih klonih (Sl. 6).



Slika 6: Odvisnost med produktivnostjo in številom kopij rekombinantnega gena.

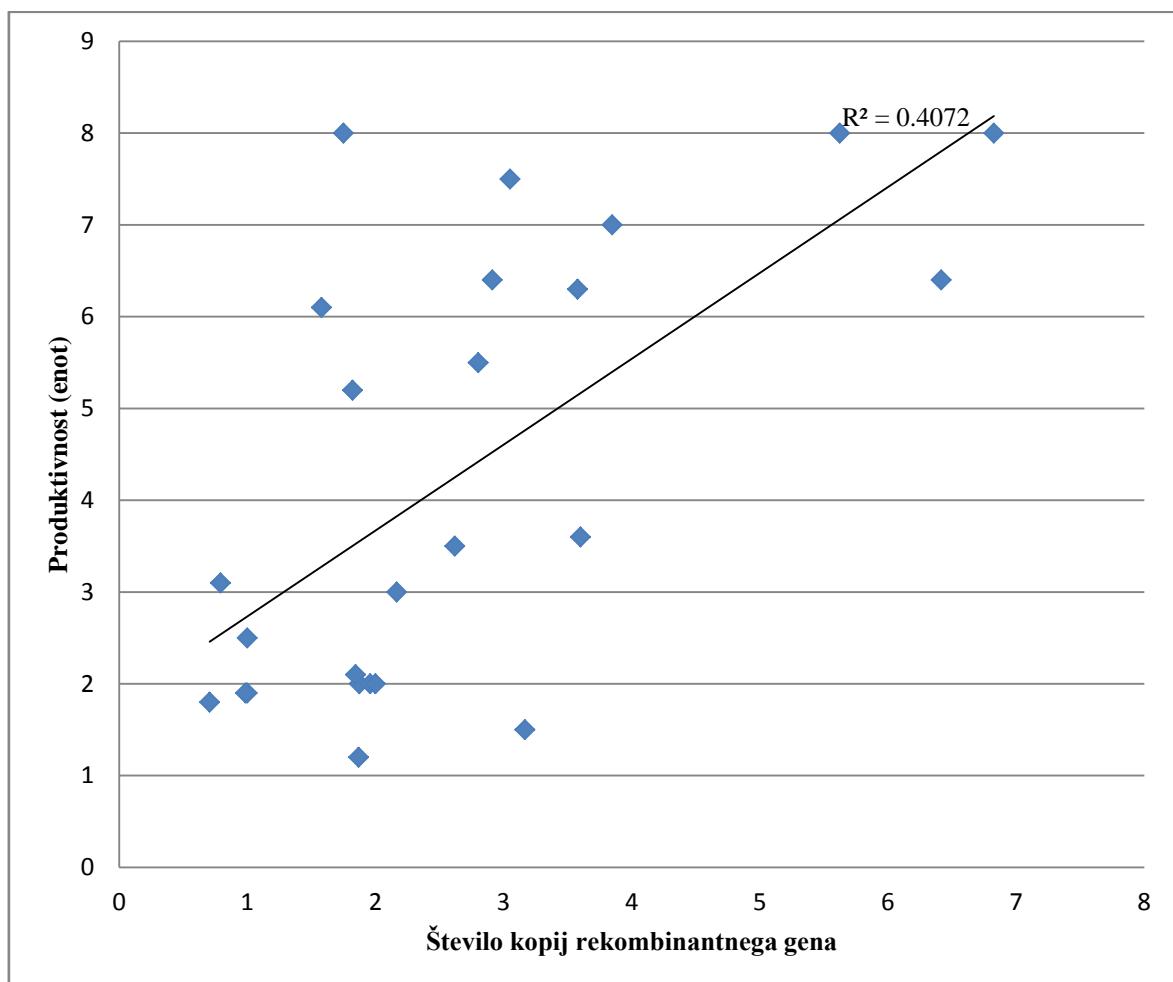
Produktivnost in število kopij rekombinantnega gena so prikazani na začetku študije (teden 1 in 2) in koncu (teden 9 in 10) 10-tednov dolge študije. Na začetku študije smo 24 vzorcev razdelili v 2 skupini po 12 vzorcev, kjer smo eni skupini vzorcev (stabilni vzorci) dodali MTX v gojišče, drugo skupino vzorcev (nestabilni vzorci) smo gojili brez prisotnosti MTX v gojišču. Skupina vzorcev brez dodanega MTX v gojišču je med študijo izgubila veliko število kopij rekombinantnega gena, znižala se je tudi produktivnost.

Figure 6: Relation between the productivity and recombinant gene copy number in clones.

The productivity and recombinant gene copy numbers of the clones were measured at the beginning (weeks 1 and 2) and the end (weeks 9 and 10) of the 10-week study, as indicated. At the beginning of the study 24 samples were separated in two groups, each containing 12 samples. One group was grown with and the other group without MTX in the medium. The group grown without MTX in the medium lost recombinant gene copies and also their productivity was reduced.

V nadaljevanju smo primerjali tudi višino produkcije rekombinante učinkovine v povezavi s številom kopij rekombinantnega gena vključenega v genomu gostiteljske celice. Načeloma je večja verjetnost za višjo produkcijo rekombinantnega proteina kadar je v gostiteljskem genomu više število kopij rekombinantnega gena, vendar v našem primeru med njima ni korelacije ($R^2 = 0,4072$). Kot je razvidno iz slike 7, je produktivnost

neodvisna od števila kopij rekombinantnega gena. Pri nekaterih vzorcih z 2-3 kopijami rekombinantnega gena v genomu gostiteljske celice je produktivnost enako visoka kot pri vzorcih kjer je 5-6 kopij rekombinantnega gena v genomu gostiteljske celice.



Slika 7: Prikaz korelacije med številom kopij rekombinantnega gena v gostiteljski celici in med produktivnostjo.

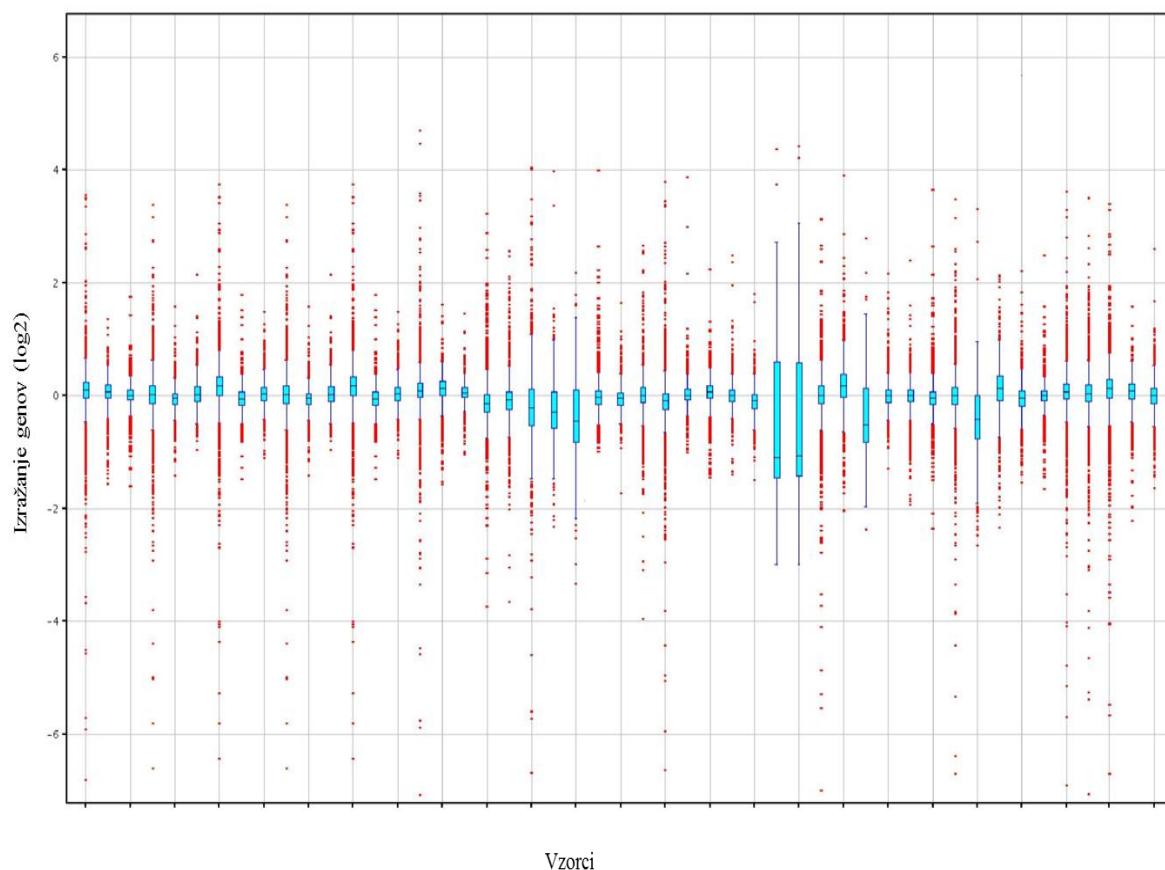
Na osnovi meritev števila kopij rekombinantnega gena v gostiteljski celici in njihovo produkcijo rekombinantnega proteina smo izračunali koeficient korelacije, ki je pokazal, da med številom kopij rekombinantnega gena in med produkcijo rekombinantnega proteina ni povezave.

Figure 7: Correlation between recombinant gene copy number and productivity.

The coefficient of correlation was calculated based on the recombinant gene copy number calculations and measurement on recombinant protein production. Based on the coefficient of correlation there is no correlation between recombinant gene copy numbers and the productivity of recombinant protein.

4.3 OBDELAVA IN ANALIZA PODATKOV IZ DNA-MIKROMREŽ

Na DNA-mikromreže smo hibridizirali 48 vzorcev, in s tem analizirali globalni transkriptom 6 klonov vzorčenih v 4 točkah (teden 1, 2, 9, 10) gojenih v pogojih, kjer je MTX bil prisoten v gojišču in vzporedno, kjer MTX ni bil prisoten. V začetku analize smo izražanje genov na vseh proučevanih vzorcih določili na surovih podatkih pred normalizacijo podatkov (Sl. 8).



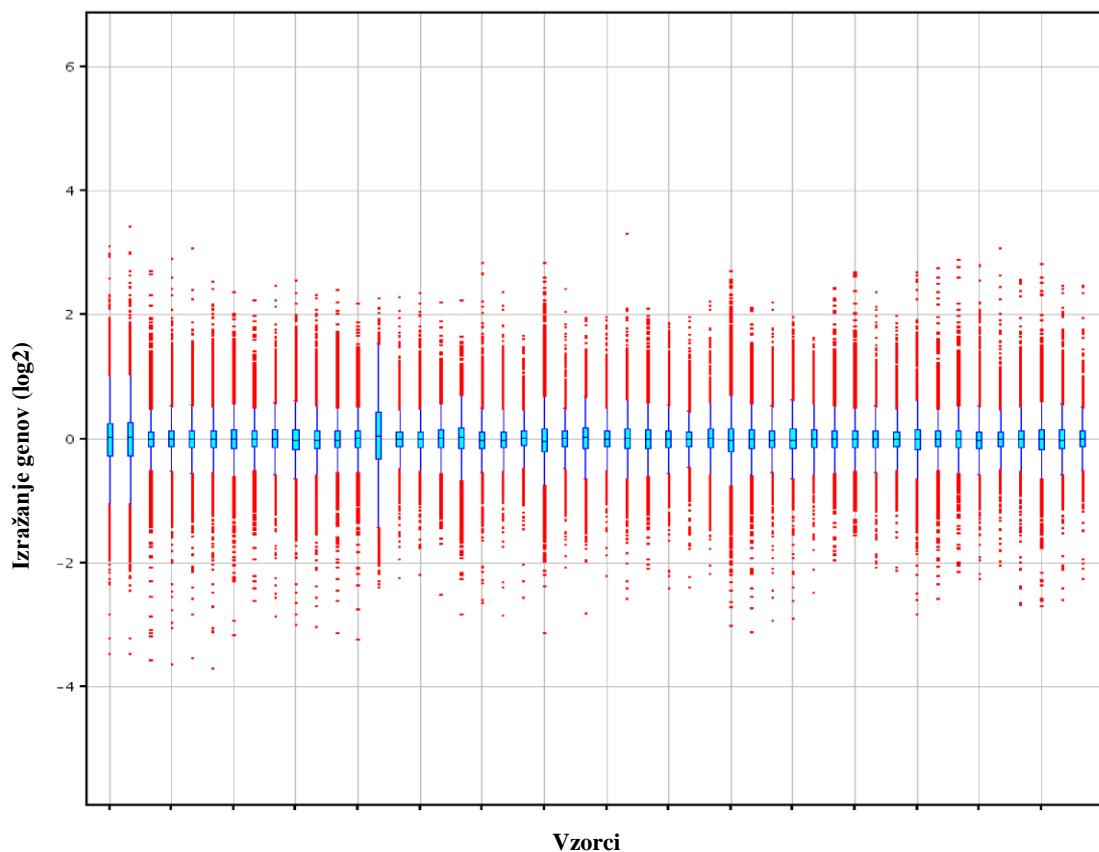
Slika 8: Porazdelitev vrednosti izražanja genov vzorcev iz DNA-mikromrež pred normalizacijo z RMA algoritmom (GeneSpring).

Surove podatke izražanja vseh genov na posameznih vzorcih pridobljenih iz DNA-mikromrež smo prikazali na skupnem grafu. Izražanje genov iz posameznih vzorcev je na sliki prikazano v obliki grafikona kvantilov. Zaradi razlik med izražanjem genov med posameznimi DNA-mikromrežami direktna primerjava izraženih genov ni možna brez normalizacije surovih podatkov.

Figure 8: Distribution of gene expression values obtained by DNA microarrays before RMA normalisation (GeneSpring).

Raw gene expression data obtained by DNA microarrays from all samples are represented on the graph. Gene expression values are represented as a Box Whisker plots for each sample. Due to the differences between gene expression values between DNA microarrays a direct comparison is not possible without a normalisation of the raw data.

Da smo lahko rezultate med vzorci medsebojno primerjali, smo jih normalizirali in s tem izničili razlike v začetnih količinah RNA, v učinkovitosti sinteze cDNA in hibridizaciji (Sl. 9).

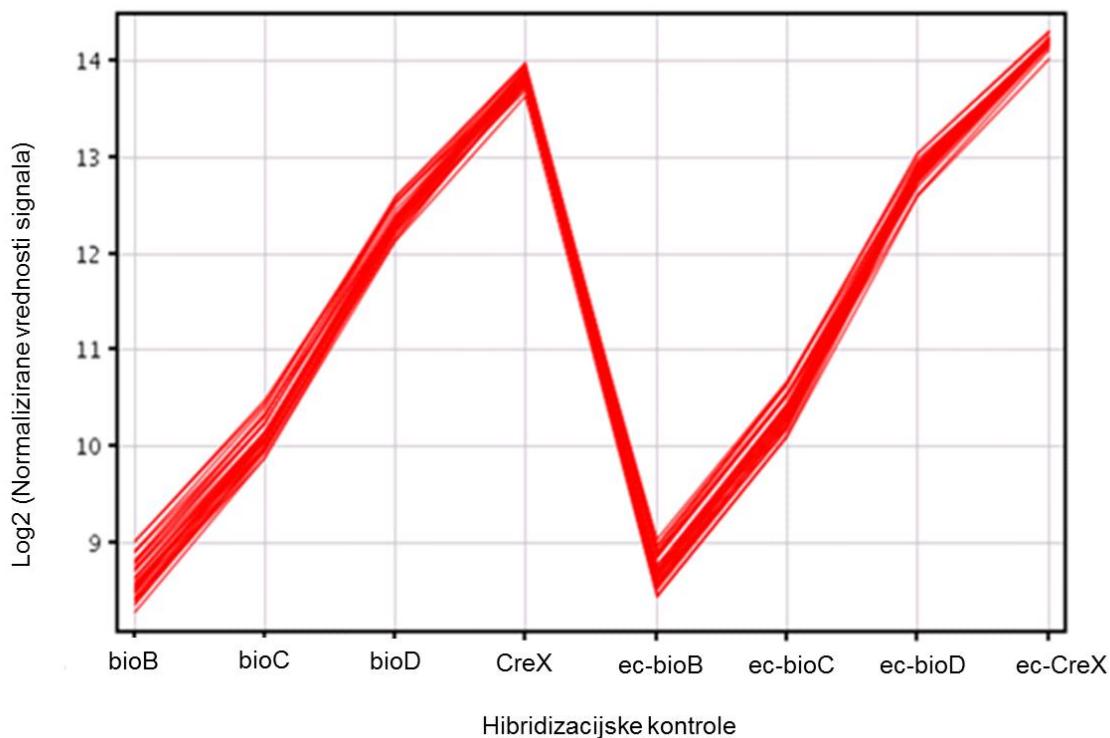


Slika 9: Porazdelitev vrednosti izražanja genov pri različnih vzorcih iz DNA-mikromrež po normalizaciji z RMA algoritmom (GeneSpring).
Izražanje genov na posameznih vzorcih je prikazano na skupnem grafu v obliki grafikona kvantilov, kjer vsak grafikon kvantilov prikazuje izražanje genov na posamezni DNA-mikromreži oz. vzorcu. Izražanje genov je normalizirano s pomočjo RMA algoritma.

Figure 9: Distribution of gene expression values obtained by DNA microarrays after the RMA normalisation (GeneSpring).

Gene expression is shown as Box Whisker plot where each Box Whisker plot represents gene expression values obtained from one DNA-microarray. Gene expression is normalized by RMA algorithm.

Nato smo preverili kvaliteto hibridizacije naših vzorcev na DNA-mikromreže. Ker smo uporabili komercialno dostopne DNA-mikromreže proizvajalca Affymetrix, so bile sonde namenjene preverjanju kvalitete hibridizacije prisotne že na sami DNA-mikromreži (Sl. 10).



Slika 10: Normalizirane (\log_2) vrednosti signala hibridizacijskih kontrolnih sond na DNA-mikromrežah za namen preverjanja učinkovitosti hibridizacije vzorcev na DNA-mikromreže.

Affymetrix DNA-mikromreže vsebujejo hibridizacijske kontrole (bioB, bioC, bioD, CreX, ec-bioB, ec-bioC, ec-bioD, ec-CreX) v različnih koncentracijah. Hibridizacijske kontrole (bioB, bioC, bioD, ec-bioB, ec-bioC) izvirajo iz različnih delov nukleotidnega zaporedja genov udeleženih v sintezi biotina v organizmu *Escherichia coli*, CreX in ec-CreX hibridizacijski kontroli izvirata iz različnih delov nukleotidne sekvene gena *cre* iz bakteriofaga P1. Nobena izmed hibridizacijskih kontrol ne hibridizira z nukleotidno sekvenco iz celic CHO. Hibridizacijske kontrole so prisotne na vseh DNA-mikromrežah in morajo imeti enako normalizirano vrednost signala, da je hibridizacija uspešna.

Figure 10: Normalized (\log_2) signal values of DNA microarrays probes as a quality control of the hybridisation on DNA microarrays.

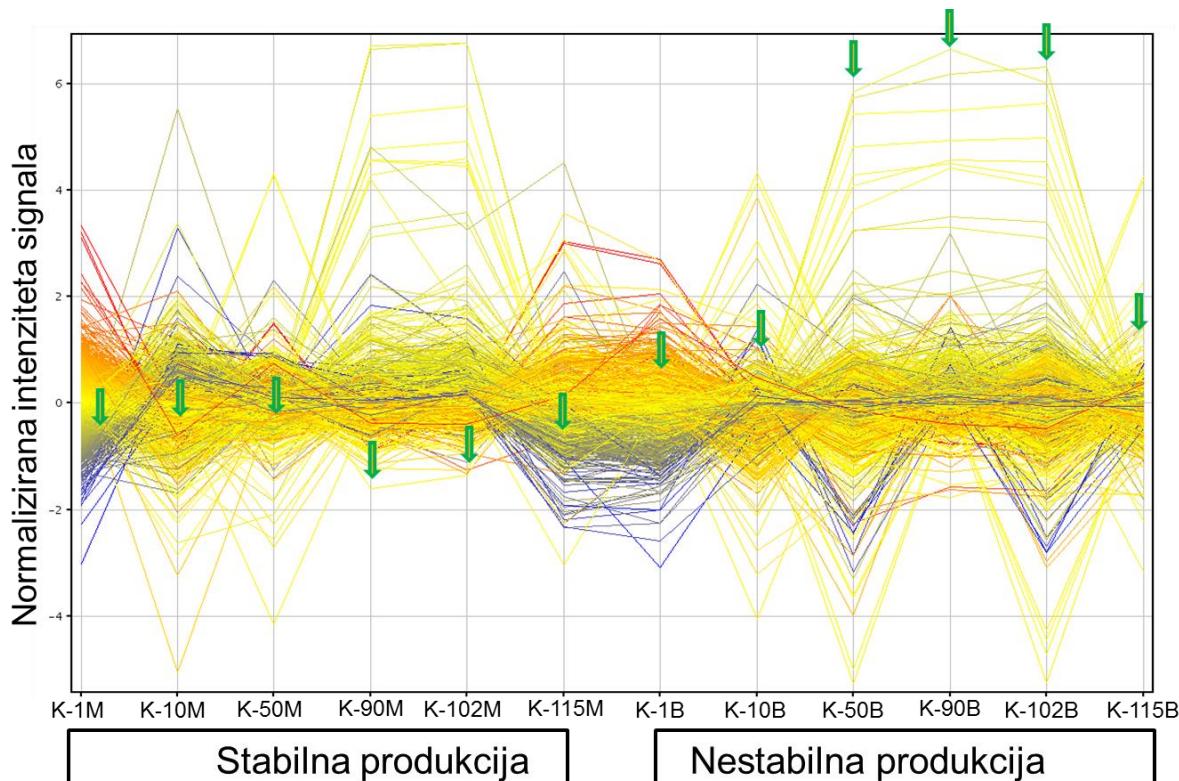
Affymetrix DNA microarrays include hybridization controls (bioB, bioC, bioD, CreX, ec-bioB, ec-bioC, ec-bioD, ec-CreX) in different concentrations. Hybridisation controls (bioB, bioC, bioD, ec-bioB, ec-bioC) originates from different parts of nucleotidic sequence of the genes in the biotin synthesis pathway of *Escherichia coli*, CreX and ec-CreX controls originates from the different parts of nucleotidic sequence of the cre gene in P1 bacteriophage. Non of the hybridisation controls hybridised with the CHO cells. Hybridisation controls are present on all DNA-microarrays and the normalized signal intensity has to be the same among DNA-micorarrays to be hybridisation successfull.

S statistično analizo smo primerjali izražanje genov pri skupinah vzorcev s stabilno/nestabilno produkcijo rekombinantne učinkovine, visoko/nizko produktivne vzorce, ter vzorce tretirane z in brez MTX v gojišču. Vzorce smo pred analizo transkriptoma razdelili v skupine po naslednjih kriterijih:

- a) stabilnost produkcije rekombinantne učinkovine skozi čas (stabilnost);
- b) učinkovitost produkcije rekombinantnega proteina (produktivnost);
- c) vpliv MTX na produkcijo rekombinantnega proteina (tretiranje z / brez MTX).

4.3.1 Stabilnost produkcije rekombinante učinkovine skozi čas (stabilnost)

Med primerjavo transkriptomskega profila med stabilnimi in nestabilnimi vzorci smo identificirali 295 izraženih genov (popravljena P-vrednost < 0,05). Na sliki 11 je prikazano izražanje genov med vzorci s stabilno in med vzorci z nestabilno produkcijo.



Slika 11: Razlika v intenziteti izražanja genov na DNA-mikromrežah med vzorci s stabilno in nestabilno produkcijo rekombinantnega proteina.

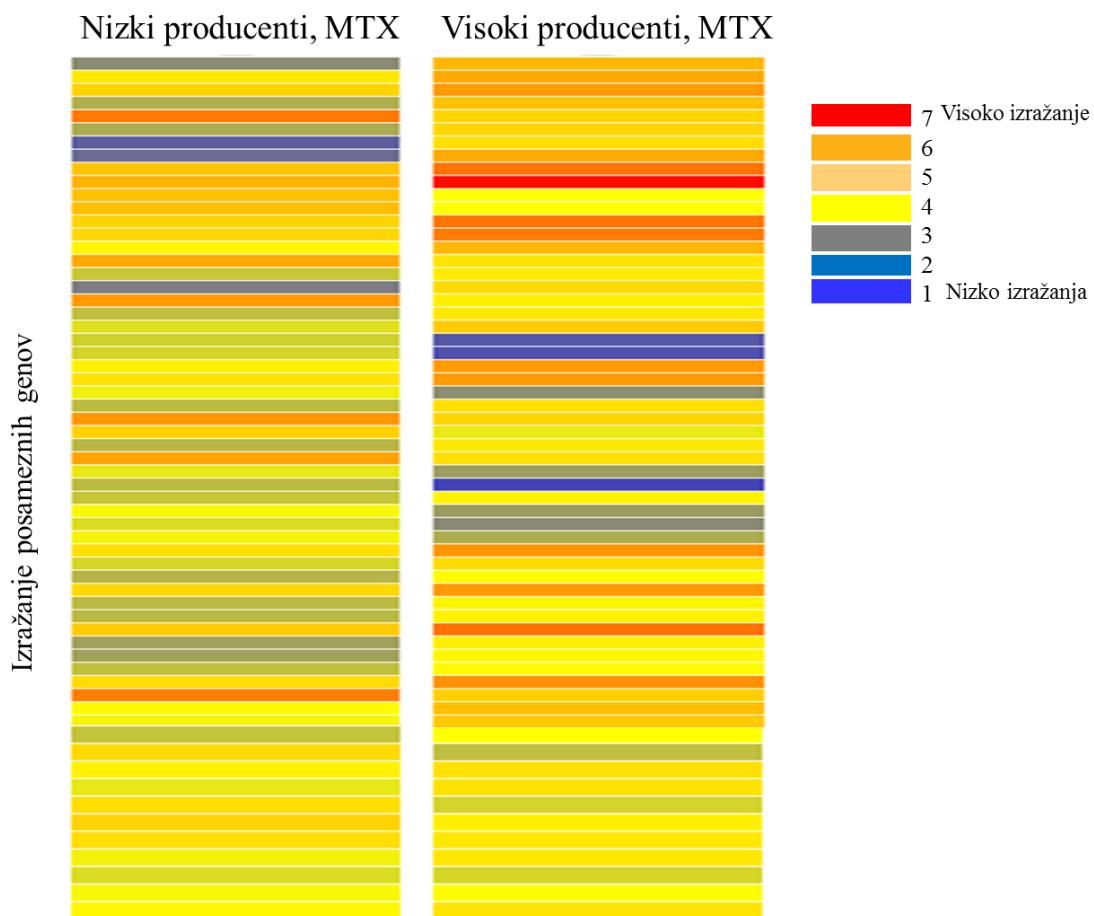
Intenziteta izražanja genov pri klonih s stabilno produkcijo rekombinantnega proteina, kjer je bil MTX ves čas prisoten v gojišču v primerjavi z izražanjem genov pri klonih, kjer MTX ni bil prisoten v gojišču (nestabilna produkcija). Puščice označujejo intenziteto izražanja gena med posameznimi kloni (K-xM – okrajšava za klon in številko klena, kjer je bil MTX prisoten v gojišču, K-xB – okrajšava za klon in številko klena, kjer MTX ni bil prisoten v gojišču). Trije kloni z nestabilno produkcijo rekombinantnega proteina imajo zelo povisano intenziteto signala pri označenem genu v primerjavi s kloni s stabilno produkcijo rekombinantnega proteina (označeno s puščico).

Figure 11: Difference in gene expression intensity on DNA microarrays between samples with stable and unstable production of recombinant protein.

Gene expression intensity comparison between clones with stable recombinant protein production (with MTX in the medium) and clones without MTX in the medium (unstable production). Arrays represent gene expression intensity among different clones (K-xM – abbreviation for clone and clone number with MTX present in the medium, K-xB – abbreviation for clone and clone number without MTX present in the medium). Three clones with unstable recombinant protein production have very high intensity values in marked gene in comparison to clones with stable recombinant protein production (marked with arrays).

4.3.2 Učinkovitost produkcije rekombinantnega proteina (produktivnost)

Med primerjavo transkriptomskega profila vzorcev z visoko in nizko produkcijo rekombinantnega proteina smo identificirali 62 statistično različno izraženih genov (popravljena P-vrednost < 0,05). Slika 12 prikazuje toplotni diagram, kjer je prikazano izražanje posameznih genov med vzorci z visoko in nizko produkcijo rekombinantnega proteina, ki v gojišču vsebujejo MTX. Izražanje genov je barvno kodirano, kjer rdeča barva pomeni visoko izražanje in modra barva nizko izražanje genov.



Slika 12: Toplotni diagram prikazuje razliko v izražanju genov med vzorci, ki imajo visoko produkcijo in med vzorci z nizko produkcijo rekombinantnega proteina ob prisotnosti MTX v gojišču.
Vsaka barvna alineja predstavlja izražanje posameznega gena. Alineje so barvno kodirane tako, da je možna hitra primerjava v izražanju gena med nizkimi in visokimi producenti.

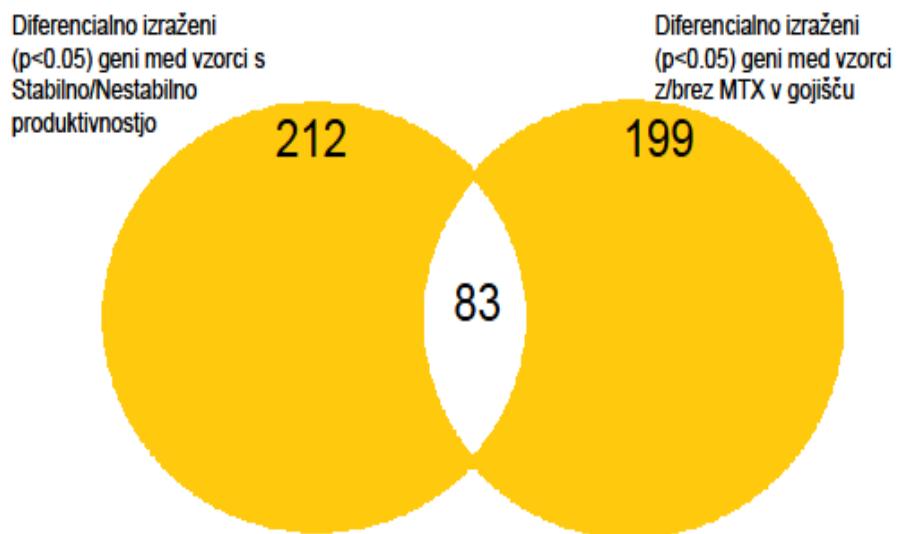
Figure 12: Heat map representing difference in gene expression between high and low producing samples where MTX was present in the media. Each color coded line represents gene expression of a gene. Lines are color coded for gene expression comparison between high and low producing samples.

Iz slike 12 je razvidno, da so nekateri geni višje izraženi pri vzorcih z visoko produkcijo rekombinantne učinkovine, nekateri geni pa nižje izraženi v primerjavi z vzorci z nizko produkcijo rekombinantne učinkovine. Z nadaljnjim filtriranjem ($\log FC_{abs} > 0,8$) rezultatov, smo izbrali 12 genov, za markerje najbolj primernih genov, katerih izražanje smo v nadaljevanju potrdili z metodo RT-qPCR.

4.3.3 Vpliv MTX na produkcijo rekombinantnega proteina (tretiranje z/brez MTX)

V nadaljevanju smo primerjali transkriptomski profil vzorcev z visoko produkcijo rekombinantne učinkovine z in brez vsebnosti MTX v gojišču. Med analizo transkriptomskega profila vzorcev z in brez MTX v gojišču, smo identificirali 199 značilno izraženih genov (popravljena P-vrednost $< 0,05$). Z nadaljnjim filtriranjem ($\log FC_{abs} > 0,8$) rezultatov smo identificirali 10 genov za nadaljne potrjevanje izražanja genov z metodo RT-qPCR.

Ker so vsi nestabilni vzorci izvirali iz gojišča, kjer ni bilo prisotnega selekcijskega pritiska (brez MTX v gojišču) in vsi stabilni kloni iz okolja, kjer je selekcijski pritisk bil prisoten (MTX v gojišču), je očiten vpliv MTX na stabilno produkcijo rekombinantnega proteina v vzorcih. Zaradi tega smo preverili, kateri geni so različno izraženi med stabilnimi in nestabilnimi kloni z in brez prisotnosti MTX v gojišču. Izmed 295 genov je bilo 199 genov statistično značilno različno izraženih med vzorci, kjer je MTX bil prisoten v gojišču, v primerjavi z vzorci kjer MTX ni bil prisoten (popravljena P-vrednost $< 0,05$). Med primerjavo statistično različno izraženih genov med stabilnimi in nestabilnimi vzorci in med vzorci z in brez prisotnosti MTX v gojišču, smo ugotovili, da je 83 genov skupnih obema skupinama vzorcev. Preostalih 212 genov je bilo statistično različno izraženih samo med stabilnimi in nestabilnimi kloni in zaradi tega so bili zanimivi za nadaljne študije (Sl. 13).



Slika 13: Venn-ov diagram prikazuje presek statistično različno izraženih genov med dvema analizama izražanja genov.

Venn-ov diagram smo uporabili za prikaz genov, ki so različno izraženi med vzorci s stabilno in nestabilno produkcijo rekombinantega proteina, niso pa različno izraženi pri vzorcih z in brez MTX v gojišču (212 genov). Statistično različno izraženi geni med vzorci, kjer je bil MTX prisoten v gojišču in med vzorci kjer MTX ni bil prisoten v gojišču. Druga analiza predstavlja različno izražene gene med vzorci s stabilno produkcijo in med vzorci z nestabilno produkcijo rekombinantega proteina.

Figure 13: Venn diagram represents differentially expressed genes in two analysis: comparing stable/unstable samples and samples with or without MTX present in the media.

Venn diagram was used to represent specific genes which are differently expressed between samples with stable and unstable recombinant protein production and are not differently expressed between samples with and without MTX present in the media. Differently expressed genes between samples with and without MTX present in the media. The second analysis represent differently expressed genes between samples with and without stable recombinant protein production.

Izmed vseh 212 statistično različno izraženih genov med stabilnimi in nestabilnimi vzorci, ne glede na vpliv MTX, smo z nadaljnjim filtriranjem rezultatov ($\log FC_{abs} > 0,8$) izbrali 14 genov. Specifičen vzorec izražanja 14 genov smo v nadaljevanju potrdili z metodo RT-qPCR.

4.4 POTRJEVANJE IZRAŽANJA IZBRANIH GENOV Z METODO RT-QPCR

V preglednici 9 so prikazani glavni parametri, ki opisujejo značilnosti pomnoževanja amplikonov vseh 36 genov, katerih izražanje smo preverili z metodo RT-qPCR. Podane so tudi povprečne vrednosti treh vzorcev s standardnimi deviacijami. Podatki za umetni amplikon AcGa so bili izračunani iz podatkov za *Actb* gen in *Gapdh* gen. Prav tako sta podana še razpon vrednosti Cq, ki smo jih uporabili za relativno kvantifikacijo, ter razpon redčitvenih faktorjev za vzorce cDNA, ki smo jih uporabili pri posameznem amplikonu.

Preglednica 9: Glavni parametri, ki opisujejo značilnosti pomnoževanja amplikonov qRT-PCR.
Podane so povprečne vrednosti meritev vseh vzorcev s standardnimi deviacijami. Podatki za umetni amplikon AcGa so bili izračunani iz podatkov za *Actb* in *Gapdh*. V tabeli sta podana še razpon vrednosti Cq, ki smo jih uporabili za izračun relativne kvantifikacije in razpon redčitvenih faktorjev za vzorce cDNA, ki smo jih uporabili pri posameznem amplikonu (k – naklon premice linearne regresije, R² – korelacijski koeficient, E – učinkovitost pomnoževanja).

Table 9: Performance characteristics of all qRT-PCR amplicons.

Mean values of all individual measurements are represented with standard deviations. Data for the artificial amplicon AcGa have been calculated from the *Actb* and *Gapdh* data. For each amplicon the Cq range is shown from which the relative quantification was calculated and the dilution factor range of the cDNA samples, which was used in for each amplicon (k – slope of linear regression line, R² – correlation coefficient, E – efficiency of amplification)

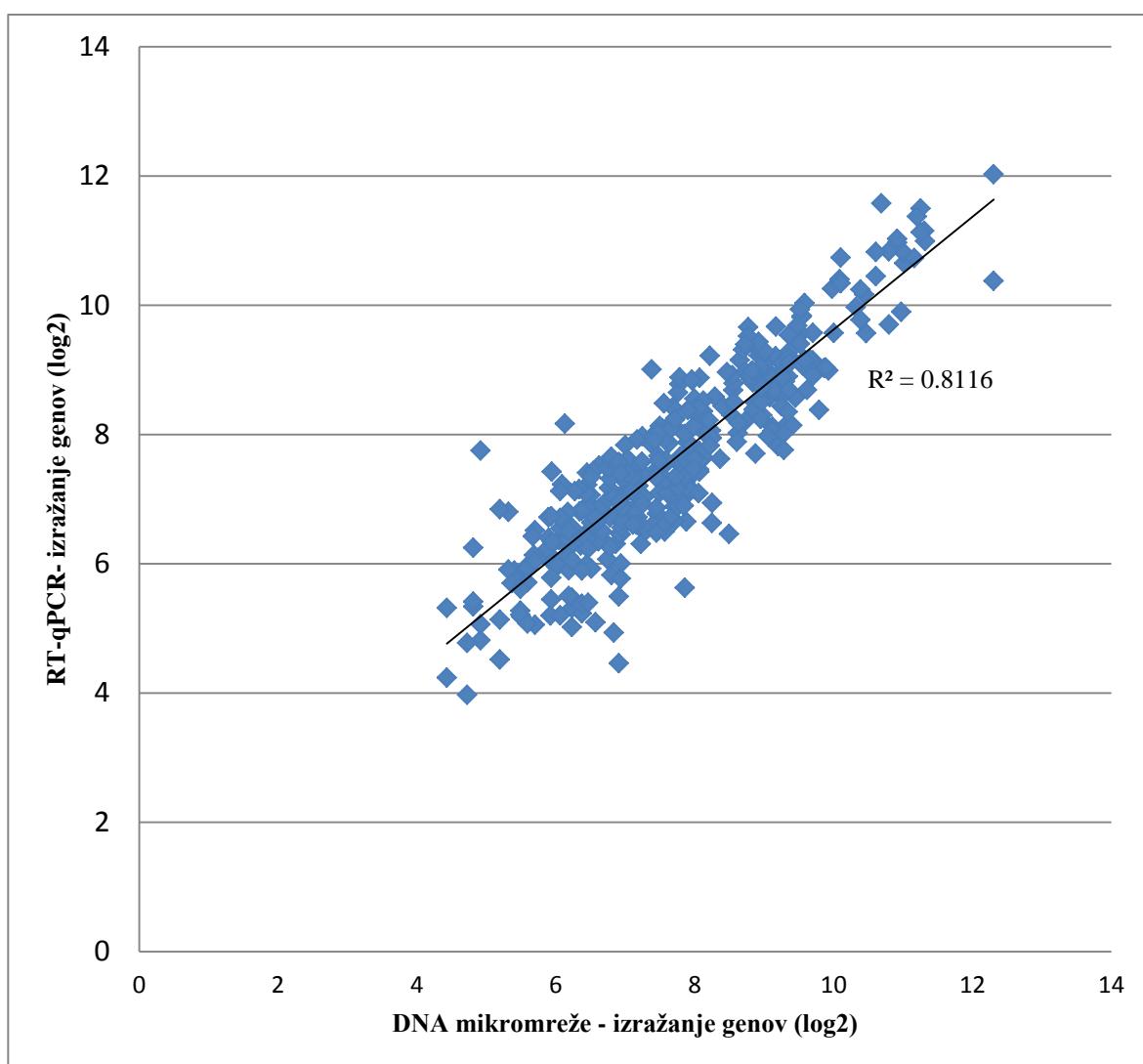
Test	Razpon redčitvenih faktorjev	Linearna regresija			Razpon vrednosti Cq
		naklon (k)	R ²	E	
<i>Ac115880.11</i>	30 - 300	-3.362±0.157	0.997±0.002	0.986±0.064	26.8 - 35.8
<i>Ac158560.4</i>	30 - 300	-3.468±0.053	0.991±0.011	0.943±0.02	25.2 - 34.9
<i>Ankrd26</i>	30 - 300	-3.607±0.148	0.995±0.006	0.895±0.05	25.9 - 35.2
<i>B2bf95_rat</i>	9 - 90	-3.476±0.06	0.991±0.01	0.94±0.022	26.8 - 35.6
<i>Bx842664.2/Hist1h3c</i>	30 - 300	-3.518±0.138	0.99±0.001	0.926±0.05	23.5 - 35.9
<i>C1qtnf1</i>	30 - 300	-3.753±0.01	0.998±0.001	0.847±0.003	23.4 - 31.3
<i>Cd68</i>	30 - 300	-3.491±0.045	0.999±0.001	0.935±0.017	23.8 - 33.7
<i>Cdh1</i>	30 - 300	-3.566±0.048	0.996±0.005	0.908±0.017	26.7 - 35.4
<i>Ces2g</i>	30 - 300	-3.326±0.133	0.986±0.001	1±0.056	25.8 - 35.2
<i>Cntn1</i>	9 - 90	-3.526±0.008	0.993±0.003	0.922±0.003	25.4 - 33.6
<i>Cspg4</i>	30 - 300	-3.389±0.04	0.985±0.005	0.973±0.016	23.5 - 32.0
<i>Cu459186.17</i>	30 - 300	-3.465±0.104	0.989±0.007	0.945±0.039	24.5 - 31.9
<i>E130203b14Rik</i>	30 - 300	-3.424±0.079	0.99±0.004	0.96±0.031	23.9 - 31.6
<i>Egr1</i>	30 - 300	-3.757±0.017	0.984±0.021	0.846±0.005	24.6 - 35.9
<i>Fgfr2</i>	30 - 300	-3.566±0.129	0.998±0.001	0.909±0.045	26.7 - 35.9
<i>Flt1</i>	9 - 90	-3.408±0.13	0.99±0.001	0.967±0.051	27.2 - 36.7

»se nadaljuje«

»nadaljevanje preglednice 9. Glavni parametri, ki opisujejo značilnosti pomnoževanja amplikonov qRT-PCR.«

Test	Razpon redčitvenih faktorjev	Linearna regresija			Razpon vrednosti Cq
		naklon (k)	R ²	E	
<i>Foxp2</i>	30 - 300	-3.69±0.006	0.993±0.004	0.867±0.002	27.0 - 37.4
<i>Gpc6</i>	30 - 300	-3.529±0.006	0.994±0.003	0.921±0.002	24.5 - 33.6
<i>hDhfr</i>	30 - 300	-3.474±0.003	0.995±0.001	0.941±0.002	16.7 - 26.2
<i>Hist1h2bc</i>	30 - 300	-3.393±0.004	0.993±0.001	0.972±0.002	21.9 - 29.9
<i>Hmgn2</i>	30 - 300	-3.553±0.125	0.989±0.001	0.914±0.044	25.7 - 33.4
<i>Mmp10</i>	9 - 90	-3.608±0.058	0.984±0.003	0.894±0.02	26.9 - 36.8
<i>mouse Dhfr</i>	30 - 300	-3.433±0.087	0.998±0.001	0.957±0.033	16.5 - 27.4
<i>Myrip</i>	9 - 90	-3.588±0.014	0.988±0.001	0.9±0.005	27.8 - 37.9
<i>Plac1</i>	9 - 90	-3.684±0.018	0.997±0.002	0.869±0.006	26.1 - 34.6
<i>Plekhb1</i>	30 - 300	-3.363±0.061	0.995±0.001	0.984±0.025	26.0 - 34.8
<i>Psma8</i>	9 - 90	-3.621±0.033	0.989±0.002	0.89±0.011	29.8 - 38.7
<i>Ptpre</i>	30 - 300	-3.499±0.02	0.975±0.002	0.932±0.008	26.3 - 35.7
<i>Rps6ka2</i>	30 - 300	-3.682±0.061	0.999±0.003	0.87±0.02	25.1 - 34.2
<i>Spn1</i>	30 - 300	-3.7±0.04	0.977±0.031	0.864±0.013	23.7 - 34.1
<i>Sprr2h</i>	30 - 300	-3.697±0.031	0.997±0.003	0.865±0.01	23.6 - 32.5
<i>Sprr2k</i>	30 - 300	-3.557±0.046	0.997±0.002	0.911±0.016	22.5 - 31.8
<i>Thsd7a</i>	30 - 300	-3.618±0.034	0.993±0.008	0.891±0.012	26.2 - 36.7
<i>Tinagl1</i>	30 - 300	-3.673±0.024	0.992±0.005	0.872±0.008	20.3 - 29.0
<i>Vsnl</i>	9 - 90	-3.529±0.02	0.985±0.004	0.921±0.007	31.6 - 37.8
<i>Zdhhc6</i>	30 - 300	-3.498±0.034	0.996±0.001	0.932±0.013	18.7 - 27.1
<i>Actb</i>	900 - 9000	-3.417±0.049	0.998±0.001	0.963±0.019	20.3 - 28.0
<i>Gapdh</i>	900 - 9000	-3.5±0.023	1±0.001	0.931±0.009	21.6 - 28.4
<i>AcGa</i>	/	-3.473±0.1	0.999±0.001	0.942±0.037	21.4 - 28.1

V celotnem poskusu smo z RT-qPCR analizirali izražanje 36 pomembnejših genov, ki smo jih izbrali na osnovi analize podatkov iz DNA-mikromrež. Korelacija med izražanjem vseh 36 genov analiziranih z DNA-mikromrežami in RT-qPCR je prikazana na sliki 14. Izračunali smo tudi koeficient korelacije, ki je znašal $R^2 = 0.8116$.

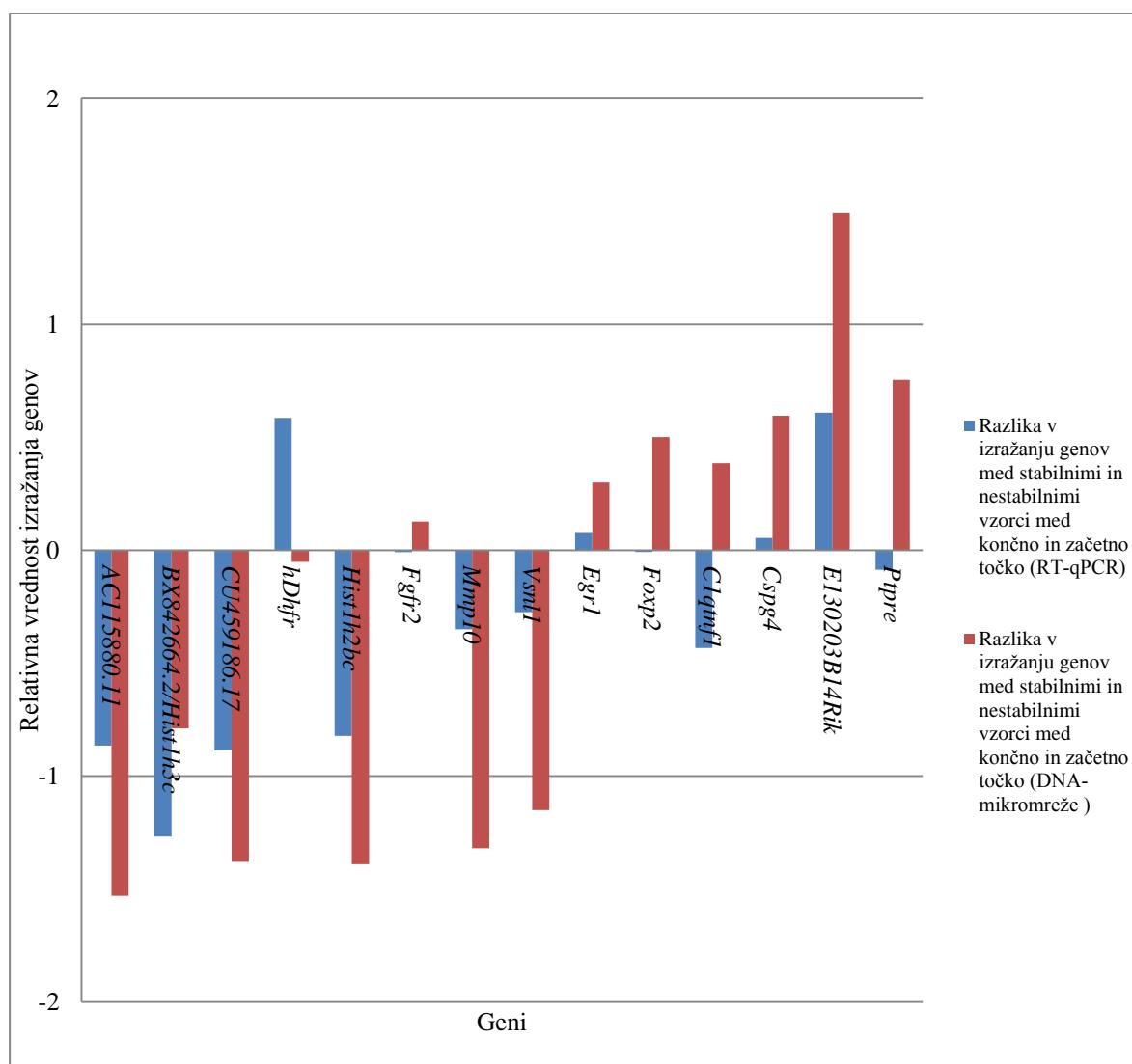


Slika 14: Prikaz korelacije izmerjenega izražanja 36 genov med DNA-mikromrežami in RT-qPCR.
Vseh 36 genov, katerih izražanje smo potrjevali z metodo RT-qPCR, smo primerjali z izražanjem na DNA-mikromrežah in izračunali korelacijo med obema metodama ($R^2 = 0,8116$). Večina izraženih genov je enako izražena pri obeh metodah.

Figure 14: Correlation between the gene expression values of 36 genes measured by DNA-microarrays and RT-qPCR. For all 36 genes where gene expression was confirmed by RT-qPCR method the coefficient of correlation was calculated ($R^2 = 0,8116$). Most of the genes have the same gene expression ratio in both methods.

4.4.1 Stabilnost produkcije rekombinantnega proteina skozi čas (stabilnost)

Izražanje 14 izbranih genov, povezanih s stabilno produkcijo rekombinantnega proteina smo preverili z metodo RT-qPCR in njihovo izražanje primerjali z rezultati pridobljenimi z DNA-mikromrežami (Sl. 15). Zaradi večje preglednosti primerjave smo rezultate časovno združili. Za vsak gen smo izračunali povprečje vrednosti izražanja genov vseh vzorcev s stabilno produkcijo rekombinantnega proteina iz konca poskusa (teden 9 in 10) in vrednost odšteli od povprečja izražanja genov vseh vzorcev iz začetka poskusa (teden 1 in 2). Enako smo naredili za vse vzorce z nestabilno produkcijo rekombinantnega proteina. Za prikaz na spodnji sliki smo še dodatno, za vsak gen, izračunali razliko v izražanju gena med vzorci s stabilno in nestabilno produkcijo rekombinantnega proteina v končni in začetni točki. Tako smo za vsak gen dobili eno vrednost za DNA-mikromreže in eno vrednost za RT-qPCR, kar je prikazano na sliki 15.



Slika 15: Primerjava izražanja genov značilnih za stabilnost produkije rekombinantega proteina med metodo DNA-mikromrež in metodo RT-qPCR.

Prikazano je relativno izražanje genov kjer smo za namen prikaza rezultate časovno združili (izračunali razliko med povprečjem vrednosti izražanja genov iz konca poskusa in povprečjem vrednosti iz začetka poskusa) in dodatno izračunali razliko v izražanju genov med vzorci s stabilno in nestabilno produkcijo rekombinantnega proteina. Enak izračun smo naredili za rezultate obeh metod (DNA-mikromreže in RT-qPCR). Iz vsake metode smo tako dobili en rezultat in ga prikazali na sliki. Pri genih *Fgfr2* in *Foxp2* vrednost izražanja pri metodi RT-qPCR ni vidna, ker je relativno vrednost izražanja pri metodi RT-qPCR zelo majhna (*Fgfr2* 0,01 in *Foxp2* 0,03).

Figure 15: Comparison of measured gene expression between DNA microarrays and RT-qPCR for individual genes imported for stable recombinant protein production.

The relative expression is shown where the results were combined on time bases (the difference between an average of results of gene expression from the end of the study and an average of the gene expression at the beginning of the study). Additionally the difference in gene expression between samples with stable and unstable recombinant protein production was calculated. The same calculation was used for the results of gene expression for both methods (DNA microarray and RT-qPCR). Finally, one result from each method was obtained and it is shown in a graph. In some genes (*Fgfr2* and *Foxp2*) the calculated values based on RT-qPCR method are not visible, due to low values (*Fgfr2* 0,01 and *Foxp2* 0,03).

Iz slike 15 je razvidno, da je večina genov (11 od 14) reguliranih v isto smer (znižana ali povišana raven izražanja), tako pri DNA-mikromrežah kot pri metodi RT-qPCR. Pri genih *Fgfr2* in *Foxp2* vrednost izražanja pri metodi RT-qPCR na sliki 15 ni vidna, ker je relativno vrednost izražanja pri metodi RT-qPCR zelo majhna (*Fgfr2* 0,01 in *Foxp2* 0,03). Pri preostalih 3 genih (*hDhfr*, *Ptpre* in *C1tnf1*) smo opazili razliko pri regulaciji izražanja genov med obema metodama.

Biološka vloga 14 izbranih genov (*Fgfr2*, *BX842664.2/Hist1h3c*, *AC115880.11*, *E130203b14Rik*, *hDhfr*, *Hist1h2bc*, *Cspg4*, *C1qtnq*, *Foxp2*, *Mmp10*, *Vsnl1*, *CU459186.17*, *Egr1*, in *Ptpre*) iz sklopa stabilnosti produkcije rekombinantnega proteina skozi čas (stabilnost), je prikazana v preglednici 10.

Preglednica 10: Opis funkcije genov, potencialnih markerjev za stabilno produkcijo rekombinantnega proteina.

Table 10: Gene description of potential marker genes for stable production of recombinant protein.

Okrajšava gena	Gen kodira	Gen kodira (angl.)	Najpomembnejše biološke vloge
<i>Fgfr2</i>	Receptor 2 za rastni faktor fibroblastov	Fibroblast growth factor receptor 2	regulacija celične proliferacije, diferenciacije, migracije in apoptoze; regulacija embrionalnega razvoja
<i>Bx842664.2 / Hist1h3c</i>	Histon H3.2	Histone H3.2	regulacija transkripcije (utišanja genov), popravila DNA, podvojevanja DNA in kromosomalne stabilnosti
<i>Egr1</i>	Protein 1 zgodnjega odziva rasti	Early growth response protein 1	regulacija transkripcije (predvsem genov vključenih v mitogenezo in diferenciacijo)
<i>Cspg4</i>	Hondroitin sulfat proteoglikan 4	Chondroitin sulfate proteoglycan 4	regulacija celične proliferacije in migracije
<i>Hist1h2bc</i>	Histon H2B tipa 1-c/e/g	Histone H2B type 1-C/E/G	regulacija transkripcije (utišanja genov), popravila DNA, podvojevanja DNA in kromosomalne stabilnosti
<i>E130203b14Rik</i>	Ni podatka	Ni podatka	Ni podatka
<i>Ptpre</i>	Receptor tipa tirozin-protein fosfataza epsilon	Receptor-type tyrosine-protein phosphatase epsilon	udeležen v metabolni poti glukoze, inducirani od inzulina; udeležen v signalno transdukcijo sinteze citokinov
<i>Cu459186.17</i>	Ni podatka	Ni podatka	Ni podatka
<i>Mmp10</i>	Matrična metalopeptidaza 10	Matrix metallopeptidase 10	udeležen v metabolni poti razgradnje fibronektina, kolagena; regulacija celične migracije
<i>Vsnl1</i>	Vizinin podoben protein 1	Visinin-like protein 1	udeležen v signalnih potekh odvisnih od kalcija
<i>Foxp2</i>	Forkhead box protein P2	Forkhead box protein P2	regulacija transkripcije; regulacija proliferacije epitelnih in mezenhimalnih celic

»se nadaljuje«

»nadaljevanje predlednice 10. Opis funkcije genov, potencialnih markerjev za stabilno produkcijo rekombinantnega proteina.«

Okrajšava gena	Gen kodira	Gen kodira (angl.)	Najpomembnejše biološke vloge
<i>C1qtnf1</i>	Protein 1 soroden komplementu C1q tumor nekrotizirajočemu faktorju	Complement C1q tumor necrosis factor-related protein 1	regulacija izražanja genov; regulacija (platelets)
<i>hDhfr</i>	Dihidrofolat reduktaza	Dihydrofolate reductase	ključni enzim v metabolizmu folatov
<i>Ac115880.11</i>	Protein LEO1 v povezavi s polimerazo RNA	RNA polymerase-associated protein LEO1	regulacija transkripcije; regulacija signalne poti Wnt

Z metodo RT-qPCR smo analizirali izražanje 14 genov (*Fgfr2*, *BX842664.2/Hist1h3c*, *AC115880.11*, *E130203b14Rik*, *hDhfr*, *Hist1h2bc*, *Cspg4*, *C1qtnq*, *Foxp2*, *Mmp10*, *Vsnl1*, *CU459186.17*, *Egr1*, in *Ptpre*) na skupno 46 vzorcih; od tega je bilo 32 vzorcev s stabilno produkcijo rekombinantnega proteina in 14 vzorcev z nestabilno produkcijo rekombinante učinkovine. Dva izmed 48 vzorcev, zaradi tehnične napake med izvajanjem analize, nista bila primerna za nadaljno obdelavo rezultatov.

4.4.1.1 Test t (stabilnost)

Za oceno statistične relevantnosti zaznanih razlik v izražanju smo izvedli Welchov dvo-repi test t med skupino vzorcev s stabilno produkcijo in skupino vzorcev z nestabilno produkcijo rekombinantnega proteina. Pokazal je, da se izmed 14 analiziranih genov statistično značilno različno izražajo vsi geni, razen gena *Vsnl 1* ($p = 0.22$); pri njem razlike niso bile statistično značilne. Sedem izmed 14 genov je bilo višje izraženih pri stabilnih klonih v primerjavi z nestabilnimi kloni (*Fgfr2*, *BX842664.2/Hist1h3c*, *AC115880.11*, *hDhfr*, *Hist1h2bc*, *Mmp10*, *CU459186.17*) in šest genov (*E130203b14Rik*, *Cspg4*, *C1qtnq*, *Foxp2*, *Egr1*, *Ptpre*) je bilo nižje izraženih pri stabilnih klonih v primerjavi z nestabilnimi kloni (S1.17). Največja razlika v izražanju genov med stabilnimi in nestabilnimi kloni je bila pri naslednjih štirih genih:

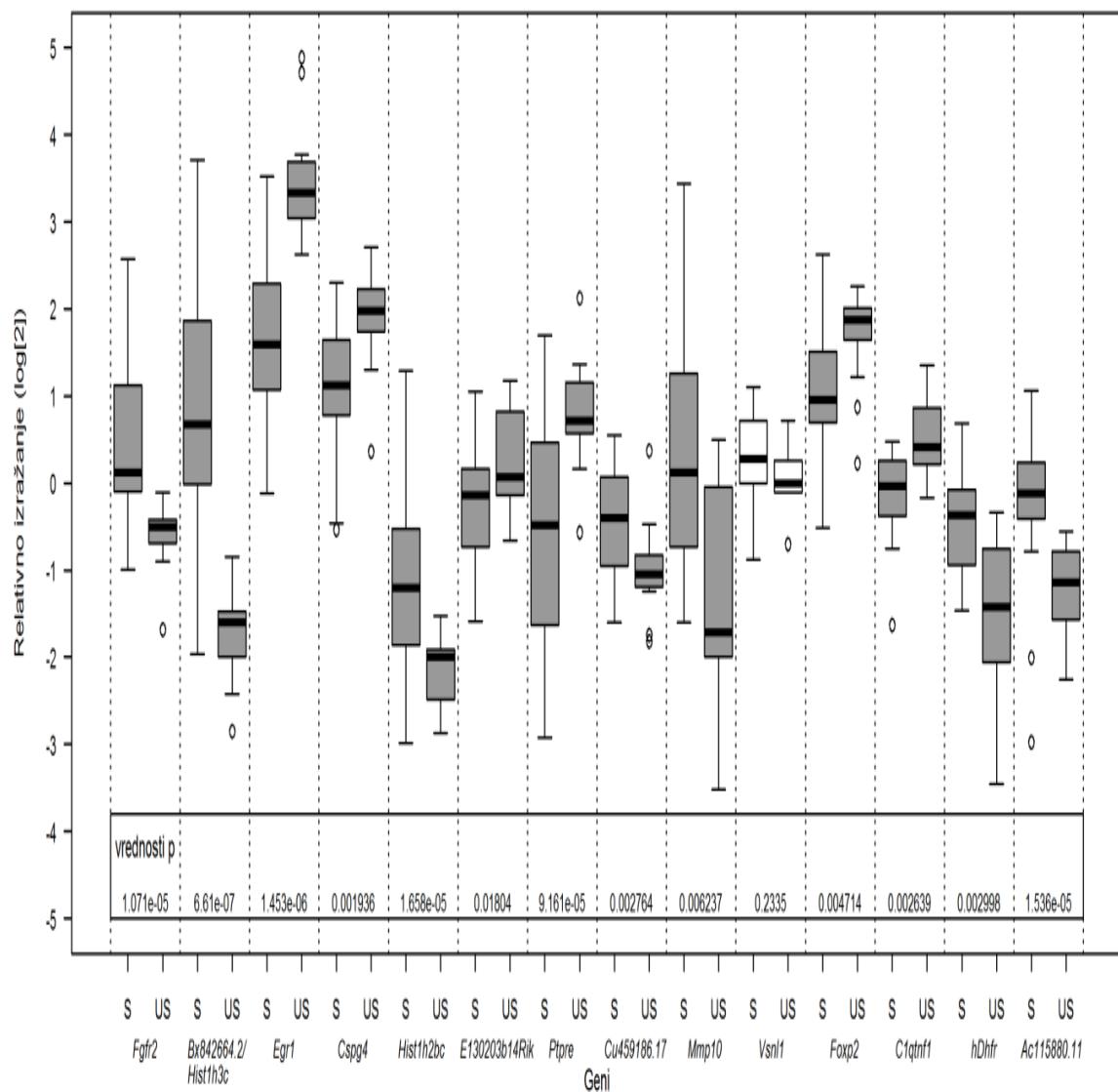
- *BX842664.2/Hist1h3c* ($\log_2\text{FC} = 2,0$);
- *Egr1* ($\log_2\text{FC} = 2,0$);

- *hDhfr*, ($\log_2\text{FC} = 1,0$);
- *Hist1h2bc* ($\log_2\text{FC} = 0,7$).

Trije (*BX842664.2/Hist1h3c,hDhfr,Hist1h2bc*) izmed štirih genov so bili višje izraženi pri vzorcih s stabilno produkcijo rekombinantnega proteina v primerjavi z vzorci z nestabilno produkcijo rekombinantnega proteina. Gena *Hist1h2bc* in *BX842664.2/Hist1h3c* kodirata histonske proteine. Njuna biološka funkcija je povezana z regulacijo transkripcije, popravilom in podvojevanjem DNA, ter stabilnostjo kromosomov. Gen *hDhfr* kodira encim DHFR, ki je ključen v metabolizmu folatov. Udeležen je pri redukciji dihidrofolične kisline v tetrafolično kislino, ki je pomembna pri sintezi purinov, ki so bistveni za celično delitev in rast. DHFR encim je ključnega pomena pri sintezi prekurzorjev za DNA. Nativni gen *Dhfr* je v gostiteljskih celicah CHO okvarjen, vendar se še vedno izraža.

Samo gen *Egr1* je bil pri vzorcih s stabilno produkcijo nižje izražen kot pri vzorcih s stabilno produkcijo. Gen *Egr1* kodira protein zgodnjega odziva rasti in je udeležen pri regulaciji transkripcije, predvsem v regulaciji genov vključenih v diferenciacijo.

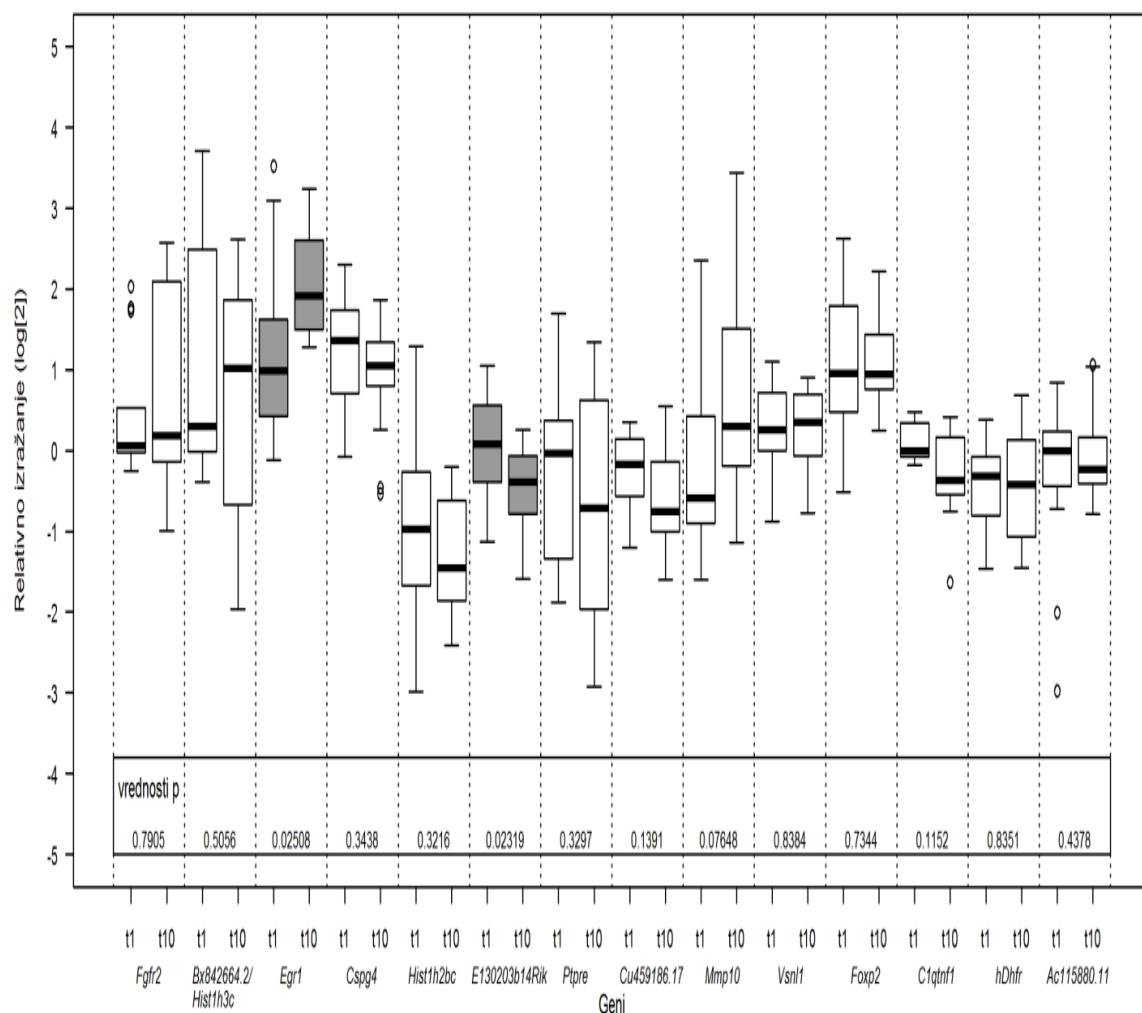
Slika 16 prikazuje relativno izražanje posameznega gena analiziranega z metodo RT-qPCR. Na sliki so ločeno prikazani grafikoni kvantilov izražanja genov pri vzorcih s stabilno produkcijo rekombinante učinkovine in pri vzorcih z nestabilno produkcijo rekombinante učinkovine.



Slika 16: Grafikon kvantilov prikazuje izražanje izbranih genov pri stabilnih in nestabilnih vzorcih. Grafikoni kvantilov predstavljajo izražanje 14 genov: *Fgfr2*, *Bx842664.2/Hist1h3c*, *Hist1h2bc*, *Cu459186.17*, *Mmp10*, *hDhfr*, *Ac115880.11*, *Egr1*, *Cspg4*, *E130203b14Rik*, *Ptpre*, *Foxp2*, *C1qtnf1* in *Vsnl1* transformiranih z \log_2 in normaliziranih z *AcGa* (geometrijsko povprečje vrednosti Cq za *Actb* in *Gapdh*) pri stabilnih in nestabilnih vzorcih. Grafikoni kvantilov so označeni temno sivo pri genih, kjer je popravljena P vrednost $< 0,05$; kjer je popravljena P-vrednost $> 0,05$, so grafikoni kvantilov označeni belo. Stabilni vzorci (S) in nestabilni vzorci (US). Osamelci so označeni s krogci. Pod vsakim parom grafikonov za vsak gen so pripisane še vrednosti p.

Figure 16: Boxplots showing relative expression ratios in stable and unstable samples. Box plot of the relative gene expression ratios ($\log_2\text{FC}$) of genes (*Fgfr2*, *BX842664.2/Hist1h3c*, *AC115880.11*, *hDhfr*, *Hist1h2bc*, *Mmp10*, *CU459186.17*, *E130203b14Rik*, *Cspg4*, *C1qtnq*, *Foxp2*, *Egr1*, *Ptpre*) for the stable (S) and unstable (US) clones. The relative expression values were normalized using geometric mean of reference genes *Actb/Gapdh*. Threshold for statistical significance between relative expression ratios of stable and unstable clones is $P=0.05$ (Welch two tailed t-test). Box plots for genes where such differences were observed are represented with dark gray (P corrected < 0.05) colors in contrast to white (P corrected > 0.05) boxplots for non-significant genes.

V nadaljevanju smo na istih podatkih izražanja 14 genov izvedli še Welchov dvo-repi test t za iskanje razlik med stabilnimi vzorci iz začetka študije (začetni dve točki - teden 1 in 2 - smo z izračunom povprečja združili v eno – t1) in med stabilnimi vzorci iz konca študije (končni dve točki - teden 9 in 10 - smo z izračunom povprečja združili v eno točko t10). Skupno smo analizirali 32 samo stabilnih vzorcev, 16 vzorcev iz začetne točke (t1) in 16 iz končne točke (t10). Želeli smo ugotoviti ali izražanje genov skozi čas ostaja enako. Welchov Test t je pokazal, da se izražanje genov pri večini genov po času ne spreminja. Izjemi sta le gena *Egr1* in *E130203b14Rik*. Pri izražanju genov *Egr1* in *E130203b14Rik* je bila statistično značilna (popravljena P-vrednost $> 0,05$) razlika med izražanjem obeh genov ob primerjavi izražanja genov v istih vzorcih odvzetih na začetku in koncu študije (Sl. 17).



Slika 17: Grafikon kvantilov prikazuje izražanje izbranih genov med stabilnimi vzorci na začetku in koncu študije.

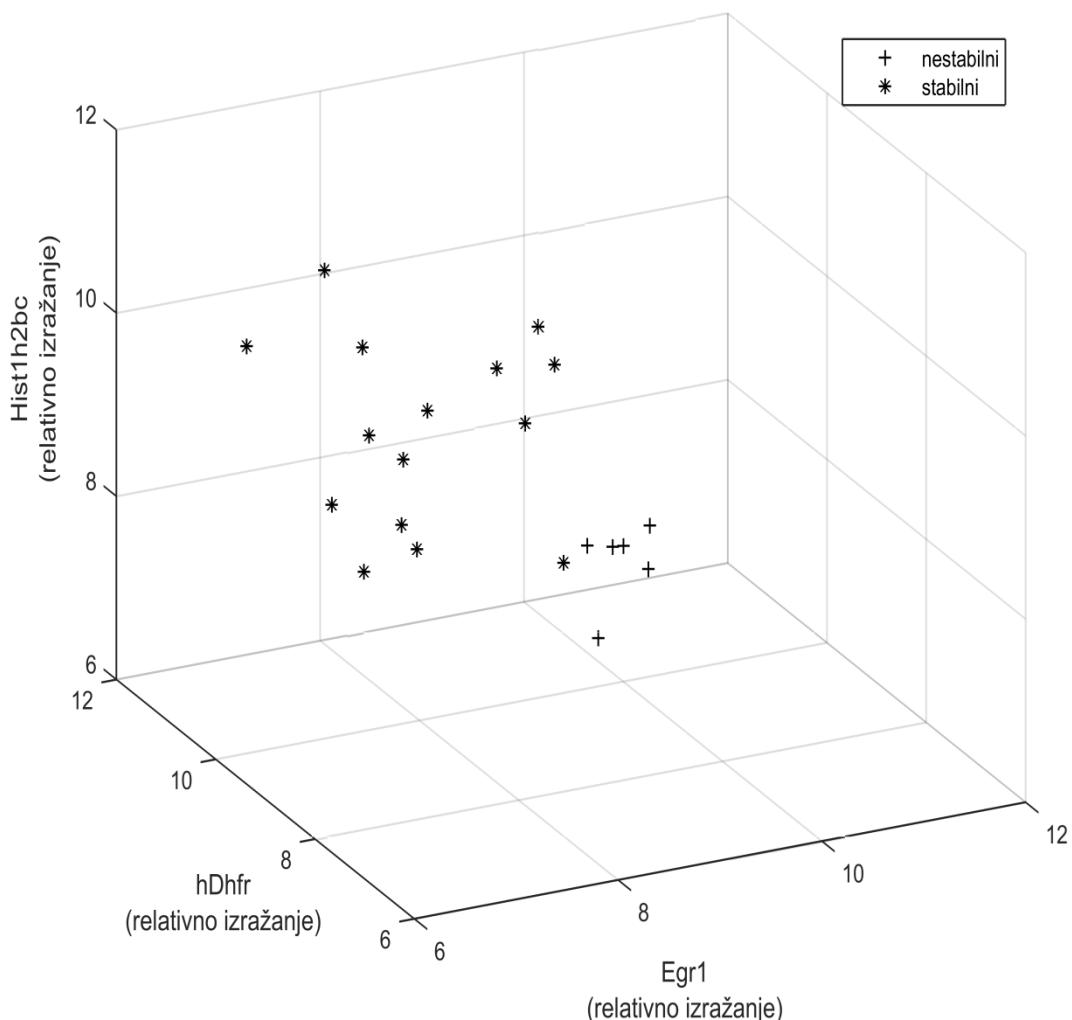
Grafikoni kvantilov predstavljajo izražanje 14 genov: *Fgfr2*, *Bx842664.2/Hist1h3c*, *Hist1h2bc*, *Cu459186.17*, *Mmp10*, *hDhfr* *Ac115880.11*, *Egr1*, *Cspg4*, *E130203b14Rik*, *Ptpre*, *Foxp2*, *Clqtnf1* in *Vsnl1* transformiranih z \log_2 in normaliziranih z *AcGa* (geometrijsko povprečje vrednosti Cq za *Actb* in *Gapdh*) pri stabilnih vzorcev na začetku in koncu študije. Prag statistično značilnega izražanja genov med izbranima skupinama vzorcev je popravljena P-vrednost $< 0,05$. Grafikoni kvantilov so označeni temno sivo pri genih, kjer je popravljena P-vrednost $< 0,05$; kjer je popravljena P-vrednost $> 0,05$, so grafikoni kvantilov označeni z belo barvo. S testom t smo analizirali 32 stabilnih vzorcev; od tega 16 iz začetne točke (t1) in 16 iz končne točke (t10). Osamelci so označeni s krogci. Pod vsakim parom grafikonov za vsak gen so pripisane še popravljene P-vrednosti.

Figure 17: Boxplots showing relative expression ratios in stable samples at the beginning and at the end of the study.

The graphicons represent the gene expression of 14 genes: *Fgfr2*, *Bx842664.2/Hist1h3c*, *Hist1h2bc*, *Cu459186.17*, *Mmp10*, *hDhfr* *Ac115880.11*, *Egr1*, *Cspg4*, *E130203b14Rik*, *Ptpre*, *Foxp2*, *Clqtnf1* and *Vsnl* which are \log_2 transformed and *AcGa* normalised (geometric mean Cq for *Actb* and *Gapdh* genes) in stable samples in the beginning and end of the study. Corrected P value of $p < 0,05$ is used for statistical significance. The boxplots are colour coded - grey for corrected $P < 0,05$ and white for corrected $P > 0,05$. 32 samples were analysed by t-test: 16 samples from the beginning of the study (t1) and 16 samples from the end point (t10).

4.4.1.2 Diskriminatorna analiza

Z nadalnjim delom smo želeli ugotoviti, če lahko vzorce ločimo na skupino stabilnih in nestabilnih le na osnovi izražanja treh genov, v našem primeru smo izbrali gene *hDhfr*, *Egr1* in *Hist1h2bc*. Diskriminatorna analiza je pokazala, da lahko na osnovi naših izbranih markerjev ločimo vzorce s stabilno in nestabilne produkcijo rekombinantnega proteina. Samo 1 vzorec (od 21) se napačno razvrsti v skupino nestabilnih vzorcev, čeprav ima stabilno produkcijo rekombinantnega proteina (Sl.18). Tudi z uporabo algoritma knajbližjega soseda smo potrdili, uspešno ločitev vzorcev v 2 skupini (vzorci s stabilno in nestabilno produkcijo rekombinantnega proteina). Algoritem je z 85 % zanesljivostjo razdelil vzorce v skupino stabilnih ali nestabilnih vzorcev že na začetku (teden 1) študije (brez slike) in s tem potrdil napovedno moč izbranih markerjev za določanje stabilnosti produkcije rekombinantnega proteina pri določenem klonu.

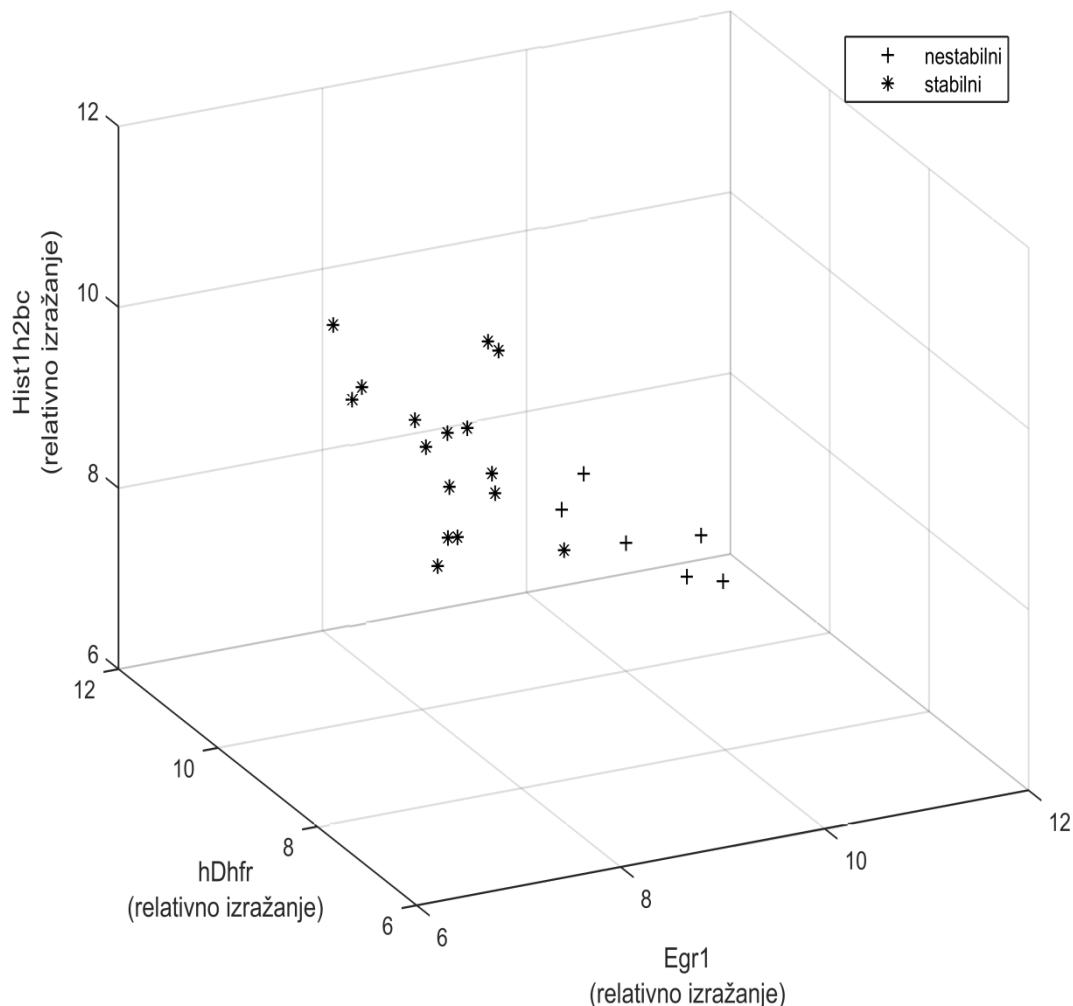


Slika 18: Diskriminatorna analiza prikazuje ločitev na skupino vzorcev s stabilno in na skupino vzorcev z nestabilno produktivnostjo že na začetku (teden 1 in 2) študije, glede na izražanje treh izbranih genov. Tri-dimenzionalna predstavitev temelji na diskriminatorni analizi, kjer so lastnosti vzorcev opisane z izražanjem 3 genov iz začetka študije (*hDhfr*, *Egr1* in *Hist1h2bc*), kjer vsaka izmed treh osi predstavlja izražanje enega gena.

Figure 18: Three-dimensional representation based on discrimination analysis where stable and unstable samples are separated in the beginning (week 1 and 2) based on the expression values of 3 potential marker genes.

The properties of the clones (originating from the beginning of the study – w1 and w2) are described by the expression of the three selected genes *hDhfr*, *Egr1* and *Hist1h2bc*, as indicated on the three-dimensional representation.

Prav tako smo uporabili diskriminatorno analizo in tri-dimenzionalno predstavitev na vzorcih odvzetih iz konca študije (teden 9 in 10). Tudi v tej časovni točki se stabilni vzorci ločijo od nestabilnih vzorcev (Sl. 19).



Slika 19: Diskriminatorna analiza prikazuje ločitev vzorcev na skupino stabilnih in nestabilnih vzorcev na koncu študije (teden 9 in 10), glede na izražanje 3 izbranih genov.

Tri-dimenzionalna predstavitev temelji na diskriminatorični analizi, kjer so lastnosti vzorcev opisane z gensko ekspresijo 3 genov iz konca študije (*hDhfr*, *Egr1* in *Hist1h2bc*), kjer vsaka izmed treh osi predstavlja ekspresijo enega gena.

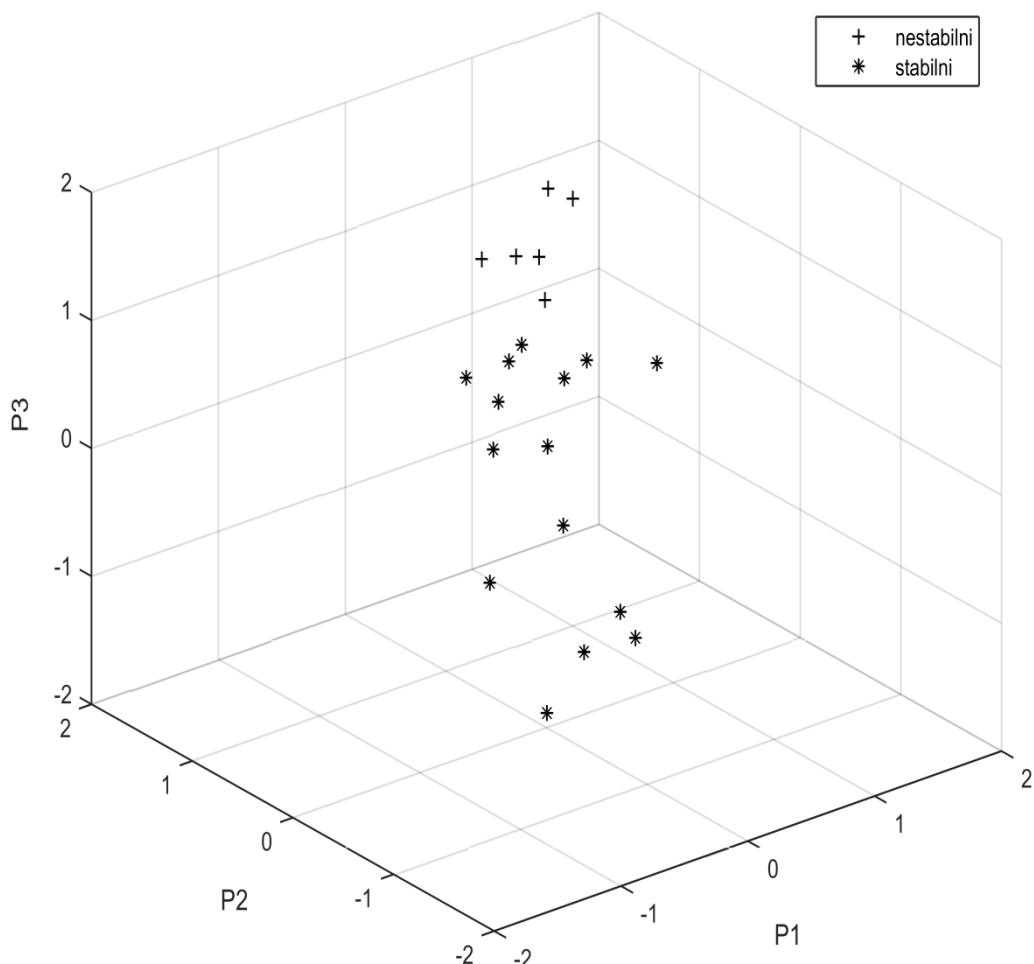
Figure 19: Three-dimensional representation based on discrimination analysis, where stable and unstable samples are distinguished at the end (week 9 and 10) based on the expression values of 3 potential marker genes.

The properties of the clones (originating from the end of the study) are described by the expression of the three selected genes *hDhfr*, *Egr1* and *Hist1h2bc*, as indicated on the three axes. In the three-dimension representation, there is clear separation of the stable and unstable clones, with only one false negative.

4.4.1.3 Analiza PCA

Izmed 14 genov je 5 genov (*Fgfr2*, *BX842664.2/Hist1h3c*, *E130203b14Rik*, *Cspg4*, in *Ptpre*) statistično različno izraženih samo med stabilnimi in nestabilnimi vzorci in njihovo izražanje ni pod vplivom prisotnosti MTX v gojišču (brez slike).

Približno 92 % variabilnosti izražanja 5 genov, lahko z analizo PCA opišemo z uporabo 3 komponent (P1, P2, P3). S pomočjo tri-dimenzionalne predstavitev 3 komponent analize PCA, lahko vidimo, da se stabilni in nestabilni vzorci ločijo v dve neodvisni skupini vzorcev (Sl. 20). Razdelitev vzorcev v dve neodvisni skupini vzorcev smo tudi pokazali z algoritmom k-najbljižjega soseda (brez slike). Algoritem je pravilno klasificiral vse vzorce iz začetka študije v skupino stabilnih ali v skupino nestabilnih vzorcev.



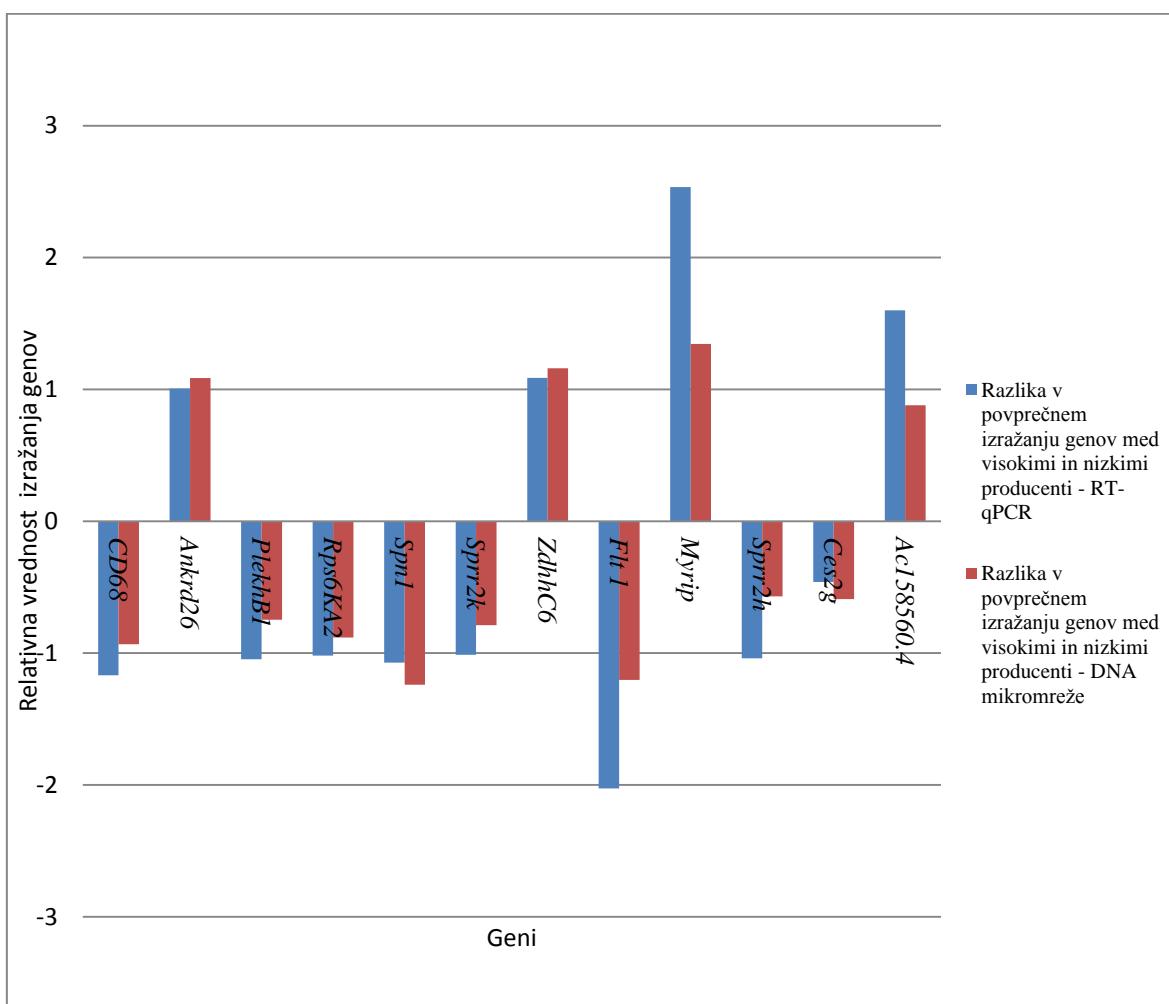
Slika 20: Tri-dimenzionalna predstavitev, kjer se stabilni in nestabilni vzorci ločijo v 2 skupini, glede na lastnosti 3 komponent (P1, P2, P3) PCA analize.
Vsaka os (komponenta) temelji na izražanju 5 genov. Lastnosti klonov iz začetka študije so opisani na treh komponentah grafa glede na izražanje genov *Fgfr2*, *BX842664.2/Hist1h3c*, *E130203b14Rik*, *Cspg4* in *Ptpre*.

Figure 20: Three-dimensional representation where stable and unstable samples are separated in 2 groups based on properties of 3 components in PCA analysis.

The properties of the clones (originating from the beginning of the study) are described by the first three principal components (P1, P2, and P3) of the expression of the genes *Fgfr2*, *BX842664.2/Hist1h3c*, *E130203b14Rik*, *Cspg4* and *Ptpre*.

4.4.2 Učinkovitost produkcije rekombinantnega proteina (produktivnost)

Za namen primerjanja izražanja genov analiziranih na dveh različnih metodah (DNA-mikromreže in RT-qPCR) smo naredili primerjalni graf (Sl. 21) relativnega izražanja posameznega gena. Izražanje 12 izbranih genov, povezanih z učinkovitostjo produkcije rekombinantnega proteina, smo preverili z metodo RT-qPCR in njihovo izražanje primerjali z rezultati pridobljenimi z DNA-mikromrežami. Zaradi večje preglednosti primerjave obeh metod, smo rezultate za prikaz na spodnji sliki združili. V osnovi smo za vsak gen izračunali povprečje vrednosti vseh vzorcev z visoko in povprečje vseh vzorcev z nizko produkcijo rekombinantnega proteina. Za prikaz na sliki 21 smo podatke v nadaljevanju pripravili enako kot za sliko 15.



Slika 21: Primerjava izražanja genov značilnih za učinkovitost produkcije rekombinantnega proteina med metodo DNA-mikromrež in metodo RT-qPCR.

Prikazano je relativno izražanje genov kjer smo za namen prikaza rezultate časovno združili (izračunali razliko med povprečjem vrednosti izražanja genov iz konca poskusa in povprečjem vrednosti iz začetka poskusa) in dodatno izračunali razliko v izražanju genov med vzorci z visoko in nizko produkcijo rekombinantnega proteina. Enak izračun smo naredili za rezultate obih metod (DNA-mikromreže in RT-qPCR). Iz vsake metode smo tako dobili en rezultat in ga prikazali na sliki.

Figure 21: Comparison of measured gene expression between DNA microarrays and RT-qPCR for individual genes imported for productivity of recombinant protein production. The relative expression is shown where the results were combined on time bases (the difference between an average of results of gene expression from the end of the study and an average of the gene expression at the beginning of the study). Additionally the difference in gene expression between samples with high and low recombinant protein production was calculated. The same calculation was used for the results of gene expression for both methods (DNA microarray and RT-qPCR). Finally, one result from each method was obtained and it is shown in a graph.

Iz slike 21 je razvidno, da so vsi geni regulirani v isto smer (znižana ali povišana raven izražanja), tako pri DNA-mikromrežah kot pri metodi RT-qPCR.

Biološka vloga 12 izbranih genov (*Ac158560.4*, *Ankrd26*, *Cd68*, *Ces2g*, *Sprr2h*, *Flt1*, *Myrip*, *Plekhb1*, *Rps6ka2*, *Spn1*, *Sprr2k* in *Zdhhc6*) iz analize transkriptoma med skupino vzorcev, ki imajo visoko produkcijo rekombinante učinkovine v primerjavi z vzorci z nizko produkcijo rekombinante učinkovine, je prikazana v preglednici 11.

Preglednica 11: Opis funkcije genov iz analize učinkovitosti produkcije rekombinantnega proteina.

Table 11: Description of genes involved in recombinant protein productivity.

Okrajšava gena	Gen kodira	Gen kodira (angl.)	Najpomembnejše biološke vloge
<i>Ces2g</i>	Karboksilesteraza 2G	Carboxylase 2G	udeležen pri signalni modulaciji
<i>Ac158560.4</i>	NA	NA	NA
<i>Sprr2h</i>	Majhen protein 2H bogat s prolinom	Small proline-rich protein 2H	udeležen v razvoju epidermisa in diferenciaciji keratinocitov
<i>Ankrd26</i>	Protein 26 s ponavljajočo domeno ankirin	Ankyrin repeat domain-containing protein 26	regulacija homeostaze glukoze; regulacija metabolizma maščobnih kislin, lipidov
<i>Cd68</i>	Makrosialin	Macrosialin	udeležen v fagocitotskih aktivnostih tkivnih makrofagov; udeležen v metabolizmu lisosomov in interakcijah celica-celica, celica-patogen
<i>Flt1</i>	Receptor 1 za vaskularno endotelni rastni faktor	Vascular endothelial growth factor receptor 1	regulacija angiogeneze in celične diferenciacije, udeležen pri embrionalni morfogenezi, udeležen pri celičnih odzivih na stimuluse rastnih faktorjev
<i>Myrip</i>	Efektor Rab MyRIP	Rab effector MyRIP	udeležen pri intracelularnem transportu proteinov in vezikularnem transportu vzdolž aktinskih filamentov
<i>Plekhb1</i>	Protein 1 z domeno, homologno plekstrinu iz družine B	Pleckstrin homology domain-containing family B member 1	regulacija celične diferenciacije (razvojni protein)
<i>Rps6ka2</i>	Kinaza alfa-2 ribosomalnega proteina S6	Ribosomal protein S6 kinase alpha-2	regulacija translacije, posrednik celične proliferacije, preživetja in diferenciacije; udeležen pri celičnem odzivu na stimulus karbohidratov
<i>Spn1</i>	Snurportin-1	Snurportin-1	regulacija transkripcije, udeležen pri importu proteinov in snRNA v jedro
<i>Sprr2k</i>	Majhen protein 2K bogat s prolinom	Small proline-rich protein 2K	udeležen pri razvoju epidermisa in diferenciaciji keratinocitov
<i>Zdhhc6</i>	Palmitoiltransferaza ZDHHC6	Palmitoyltransferase ZDHHC6	udeležen pri palmitoilaciji proteinov - metabolizem proteinov (pravilno zvijanje proteinov)

4.4.2.1 Test – t (produktivnost)

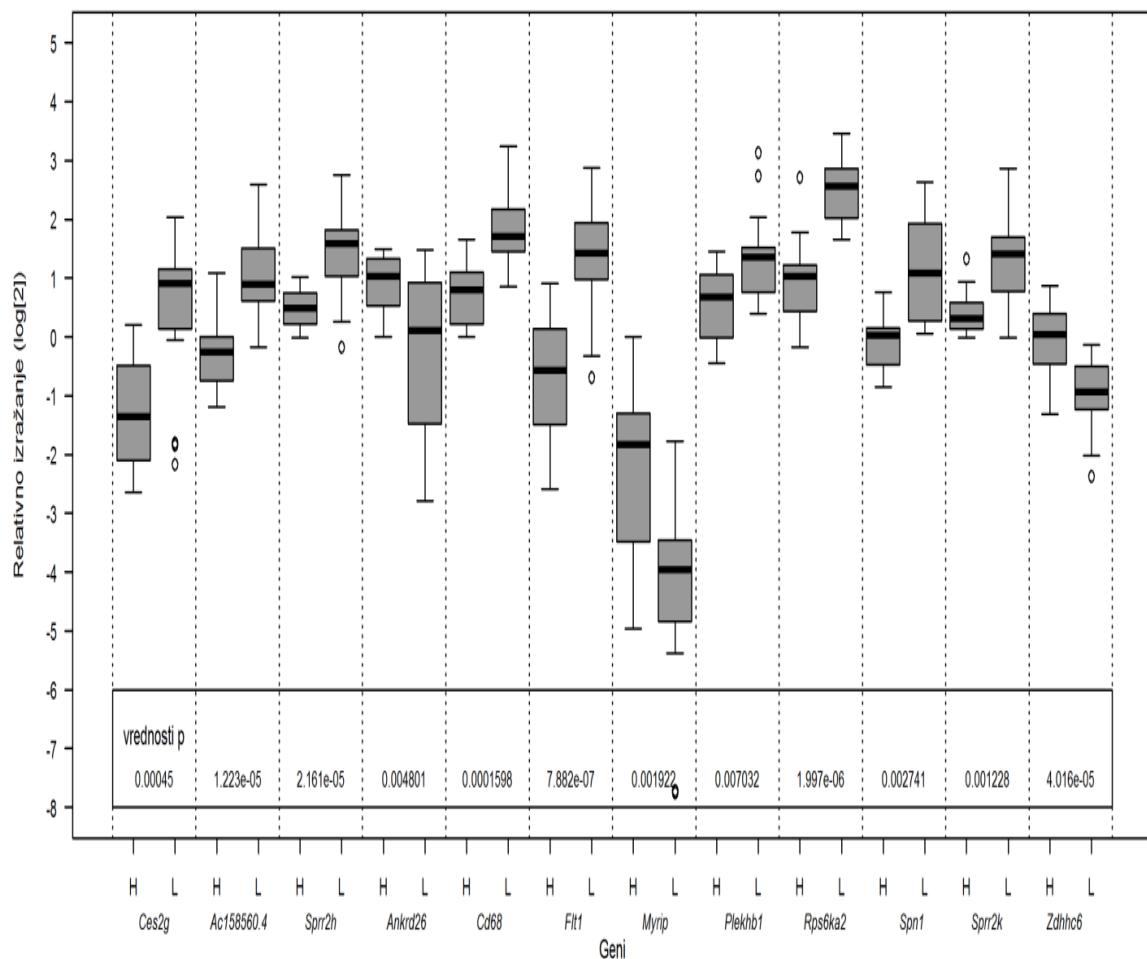
Za oceno statistične relevantnosti zaznanih razlik v izražanju smo izvedli Welchev dvo-repi test t med skupino vzorcev z visoko produkcijo in skupino vzorcev z nizko produkcijo rekombinantnega proteina. Pokazal je, da se vseh 12 genov statistično značilno različno izraža (popravljena P-vrednost < 0,05; $\log_{2}FC_{abs} > 0,8$). Izražanje 12 genov (*Ces2g*, *Ac15850.4*, *Sprr2h*, *Ankrd26*, *Cd68*, *Flt1*, *Myrip*, *Plekhb1*, *Rpska6*, *Spn1*, *Sprr2k* in *Zdhc6*), ki so bili statistično značilno drugače izraženi med vzorci z visoko in med vzorci z nizko produktivnostjo rekombinante učinkovine, smo preverili z metodo RT-qPCR (Sl. 22).

Z RT-qPCR metodo smo analizirali izražanje genov na skupno 39 vzorcih; od tega 18 visokih producentov in 21 nizkih producentov. Devet vzorcev, izmed 48, zaradi tehnične napake med izvedbo analize, niso bili primerni za nadaljno analizo rezultatov. Na osnovi rezultatov pridobljenih z metodo RT-qPCR, smo izvedli Welchev dvo-repi test t med skupino vzorcev z visoko produkcijo in skupino vzorcev z nizko produkcijo rekombinantnega proteina. Pokazal je, da se med 12 analiziranimi geni, statistično značilno (popravljena P-vrednost < 0,05) različno izražajo vsi geni. Trije izmed 12 genov so bili višje izraženi pri vzorcih z visoko produkcijo v primerjavi z vzorci z nizko produkcijo rekombinantnega proteina (*Ankrd6*, *Myrip* in *Zdhc6*). Ostali geni (9) so bili visoko izraženi pri klonih z nizko produkcijo rekombinantnega proteina (*Ces2g*, *Ac15850.4*, *Sprr2h*, *Cd68*, *Flt1*, *Plekhb1*, *Rpska6*, *Spn1* in *Sprr2k*). Največjo razliko v izražanju genov med vzorci z visoko oz. nizko produkcijo rekombinantnega proteina smo opazili pri naslednjih štirih genih:

- *Ces2g* ($\log_{2}FC=2,5$);
- *Flt1* ($\log_{2}FC=2,0$);
- *Myrip* ($\log_{2}FC=2,3$);
- *Rps6ka2* ($\log_{2}FC = 1,7$).

Izmed 4 genov, z največjo razliko v izražanju je samo gen *Myrip* bolj izražen pri vzorcih z visoko produkcijo rekombinantnega proteina v primerjavi z vzorci, ki so imeli nizko produkcijo rekombinantne učinkovine (Sl. 22). Protein, ki ga kodira gen *Myrip* je udeležen pri intracelularnem transportu proteinov in vezikularnem transportu vzdolž aktinskih filamentov. Ostali 3 geni, (*Ces2g*, *Flt1*, *Rps6ka2*) z največjo razliko v izražanju med vzorci z visoko in nizko produkcijo rekombinantnega proteina, so bili bolj izraženi pri vzorcih z nizko produkcijo rekombinantnega proteina, kot pri vzorcih z visoko produkcijo rekombinantnega proteina. Gen *Ces2g* kodira encim karboksilesterazo, ki je udeležena pri signalni modulaciji, medtem ko je produkt gena *Flt1* udeležen pri regulaciji celične diferenciacije in angiogeneze. Gen *Rps6ka2* pa je udeležen pri regulaciji translacije, posrednik pri celični proliferaciji in diferenciaciji. Glede na znane biološke funkcije, noben izmed produktov genov (*Ces2g*, *Flt1*, *Rps6ka2*), ki so visoko izraženi pri vzorcih z nizko produkcijo rekombinantnega proteina, nima vpliva na višino produkcije rekombinantnega proteina.

Grafikoni kvantilov prikazujejo relativno izražanje posameznega gena, analiziranega z metodo RT-qPCR, kjer so ločeno prikazani grafikoni kvantilov za vzorce z visoko produkcijo rekombinantne učinkovine in vzorci z nizko produkcijo rekombinantne učinkovine (Sl. 22). Relativno izražanje posameznega gena smo izračunali tako, da smo izražanje gena iz posamezne skupine vzorcev (visoko in nizko producirajoči vzorci) transformirali z \log_2 in normalizirali z AcGa, ki predstavlja geometrijsko povprečje vrednosti Cq za dva referenčna gena *Actb* in *Gapdh*. Za vsak gen smo izračunali statistično značilnost različno izražanja gena med skupino vzorcev z visoko in nizko produkcijo rekombinantne učinkovine. Izmed 12 genov, kjer smo testirali statistično značilno različno izražanje med dvema skupinama vzorcev (visoko in nizki producirajoči vzorci) smo ugotovili, da se vsi geni izražajo statistično različno (popravljena P-vrednost < 0,05). Za statistično različno izražanje genov med dvema skupinama vzorcev (visoko in nizki producirajoči vzorci), smo uporabili Welchev dvo-repi test t (Sl. 22).



Slika 22: Grafikon kvantilov prikazuje relativno izražanje izbranih genov med kloni z visoko in nizko produkcijo rekombinantne učinkovine.

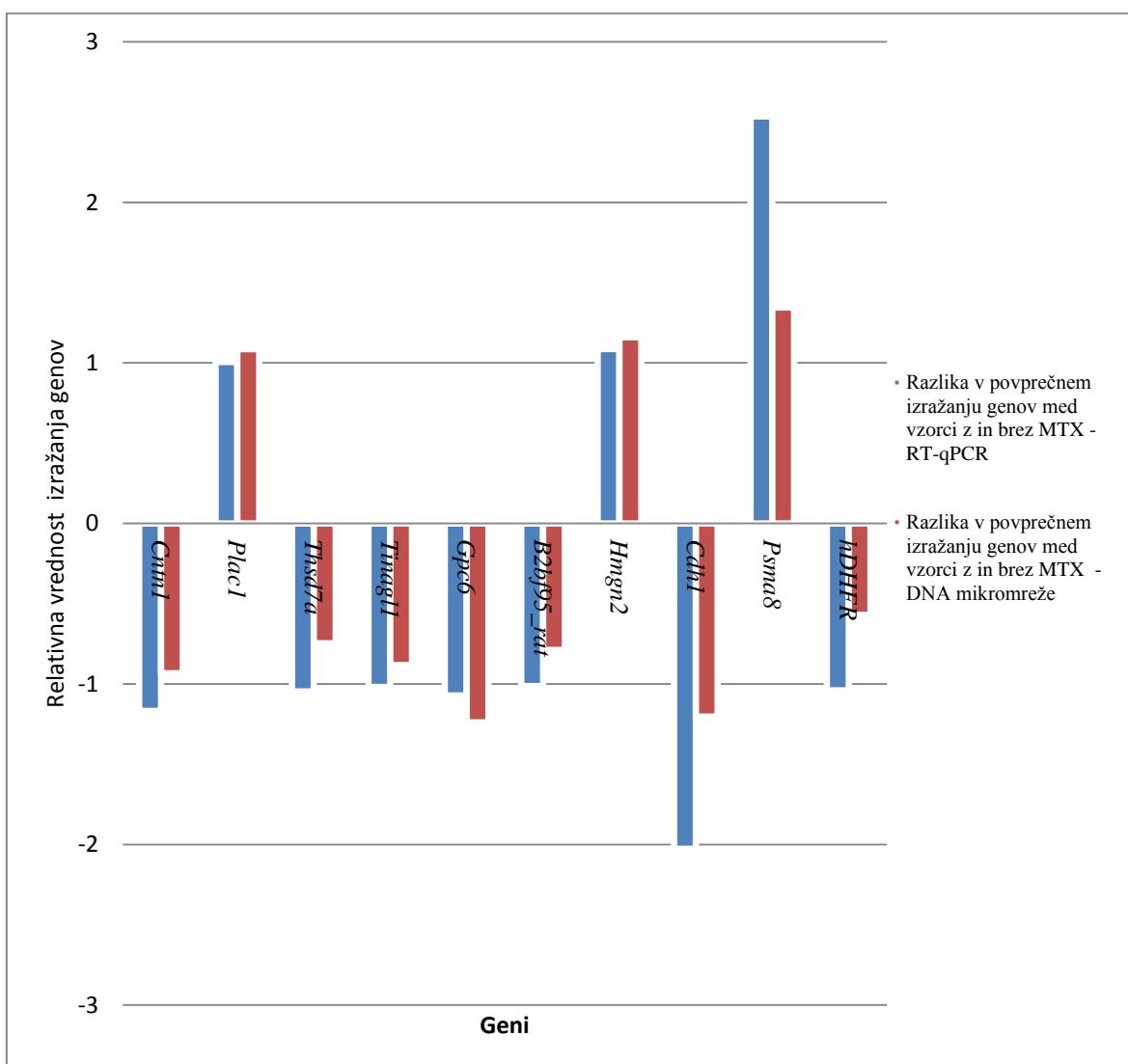
Grafikoni kvantilov predstavljajo relativno izražanje 12 genov: *Ces2g*, *Ac15850.4*, *Sprr2h*, *Ankrd26*, *Cd68*, *Flt1*, *Myrip*, *Plekhb1*, *Rpska6*, *Spn1*, *Sprr2k* in *Zdhhc6* transformiranih z \log_2 in normaliziranih z *AcGa* (geometrijsko povprečje vrednosti Cq za *Actb* in *Gapdh*) pri vzorcih z visoko oz. nizko produkcijo rekombinantnega proteina. Prag statistično značilnega izražanja med visokimi producenti in nizkimi producenti je popravljenja P-vrednost $< 0,05$. Grafikoni kvantilov so označeni temno sivo pri genih, kjer je popravljenja P-vrednost $< 0,05$; kjer je popravljenja P-vrednost $> 0,05$, so grafikoni kvantilov označeni z belo. S testom t smo analizirali 39 vzorcev; od tega 18 visokih producentov (H) in 21 nizkih producentov (L). Osamelci so označeni s krogci. Pod vsakim parom grafikonov za vsak gen so pripisane še vrednosti p.

Figure 22: Boxplots showing relative gene expression of 12 genes (*Ces2g*, *Ac15850.4*, *Sprr2h*, *Ankrd26*, *Cd68*, *Flt1*, *Myrip*, *Plekhb1*, *Rpska6*, *Spn1*, *Sprr2k*, *Zdhhc6*) in high producing clones compared to low producing clones.

The graphicons represent the relative gene expression of 12 genes *Ces2g*, *Ac15850.4*, *Sprr2h*, *Ankrd26*, *Cd68*, *Flt1*, *Myrip*, *Plekhb1*, *Rpska6*, *Spn1*, *Sprr2k* in *Zdhhc6* which are \log_2 transformed and *AcGa* normalised (geometric mean Cq for *Actb* and *Gapdh* genes) in low and high producing samples. P-corrected $< 0,05$ is used for statistical significance. The boxplots are colour coded - grey for P-corrected $< 0,05$ and white for P-corrected $> 0,05$. 39 samples were analysed by t-test: 18 high producing samples and 21 low producing samples.

4.4.3 Vpliv MTX na produkcijo rekombinantnega proteina (tretiranje z/brez MTX)

Za namen primerjanja izražanja genov analiziranih na dveh različnih metodah (DNA-mikromreže in RT-qPCR) smo naredili primerjalni graf (Sl. 23) relativnega izražanja posameznega gena. Izražanje 10 izbranih genov, povezanih z vplivom MTX na produkcijo rekombinantnega proteina, smo preverili z metodo RT-qPCR in njihovo izražanje primerjali z rezultati pridobljenimi z DNA-mikromrežami. Za prikaz na sliki 23 smo podatke v nadaljevanju pripravili enako kot za sliko 21.



Slika 23: Primerjava izražanja genov z in brez dodanega MTX v gojišču med metodo DNA-mikromrež in metodo RT-qPCR.

Prikazano je relativno izražanje genov kjer smo za namen prikaza rezultate časovno združili (izračunali razliko med povprečjem vrednosti izražanja genov iz konca poskusa in povprečjem vrednosti iz začetka poskusa) in dodatno izračunali razliko v izražanju genov med vzorci z dodanim MTX v gojišče in vzorci brez dodanega MTX v gojišču. Enak izračun smo naredili za rezultate obeh metod (DNA-mikromreže in RT-qPCR). Iz vsake metode smo tako dobili en rezultat in ga prikazali na sliki.

Figure 23: Comparison of gene expression between DNA microarrays and RT-qPCR for individual genes important for productivity of recombinant protein production.

The relative expression is shown where the results were combined on time bases (the difference between an average of results of gene expression from the end of the study and an average of the gene expression at the beginning of the study). Additionally the difference in gene expression between samples with and without MTX present in the media was calculated. The same calculation was used for the results of gene expression for both methods (DNA microarray and RT-qPCR). Finally, one result from each method was obtained and it is shown in a graph.

Iz slike 23 je razvidno, da so vsi geni regulirani v isto smer (znižana ali povišana raven izražanja), tako pri DNA-mikromrežah kot pri metodi RT-qPCR.

Biološka vloga 10 izbranih genov (*Cntn1*, *Plac1*, *Thsd7a*, *Tinagl1*, *Gpc6*, *B2bf95_rat*, *Hmgn2*, *rDhfr*, *Cu459186.17(Cdh1)* in *Psma8*) iz analize transkriptoma med skupino vzorcev, ki so bili ves čas študije tretirani z MTX v primerjavi z vzorci, ki med študijo genetske stabilnosti niso bili tretirani z MTX, so prikazane v preglednici 12.

Preglednica 12: Opis funkcije genov iz sklopa vpliva MTX na produkcijo rekombinante učinkovine.

Table 12: Description of genes influenced by MTX on productivity.

Okrajšava gena	Gen kodira	Gen kodira (angl.)	Najpomembnejše biološke vloge
<i>Cntn1</i>	Kontaktin-1	Contactin-1	udeležen pri adheziji celic, razvoju živčnega sistema; regulacija transkripcije
<i>Plac1</i>	Protein 1 specifičen za placento	Placenta-specific protein 1	udeležen pri razvoju placente
<i>Thsd7a</i>	Protein 7A z domeno trombospondin tipa 1	Thrombospondin type-1 domain-containing protein 7A	udeležen pri angiogenezi, celični diferenciaciji
<i>Tinagl1</i>	Protein podoben antiguenu tubulointerstitialni nefritis	Tubulointerstitial nephritis antigen-like	udeležen pri imunskemu odzivu
<i>Gpc6</i>	Glipican-6	Glypican-6	udeležen pri migraciji celic
<i>B2bf95_rat</i>	TCP vezavni protein 2	TCP binding protein 2	Ni podatka
<i>Hmgn2</i>	Ne-histonski kromosomalni protein HMG-17	Non-histone chromosomal protein HMG-17	regulacija transkripcije (DNA vezavni protein)
<i>Cdh1</i>	Kadherin-1	Cadherin-1	udeležen pri povezavah med celicami odvisnih od Ca; regulacija celične adhezije
<i>Psma8</i>	Protein podoben podenotni proteasoma alfa tipa 7	Proteasome subunit alpha type-7-like	udeležen pri metabolizmu razgradnje proteinov
<i>hDhfr</i>	Dihidrofolat reduktaza	Dihydrofolate reductase	ključni encim v metabolizmu folatov

4.4.3.1 Test t (tretiranje z / brez MTX)

Izražanje 10 izbranih genov iz sklopa vpliva MTX na produkcijo rekombinantnega proteina smo v nadaljevanju preverili z metodo RT-qPCR. Z metodo RT-qPCR smo analizirali izražanje genov na skupno 46 vzorcih; od tega 22 vzorcev, kjer je bil MTX prisoten v gojišču ves čas študije in 24 vzorcev, kjer smo pred začetkom študije odstranili MTX iz gojišča. Dva izmed 48 vzorcev, zaradi tehnične napake med izvajanjem analize, nista bila primerna za nadaljno obdelavo rezultatov.

Na osnovi rezultatov pridobljenih z metodo RT-qPCR smo izvedli Welchev dvo-repi test t med skupino vzorcev tretiranih z MTX in skupino vzorcev brez tretiranja z MTX. Pokazal je, da se med 10 analiziranimi geni statistično značilno različno izraža sedem genov: *Cntn1*, *Plac1*, *Thsd7a*, *Tinagl1*, *Hmng2*, *Cdh1* in *Psma8* (Sl. 24). Od tega je bila pri štirih genih (*Cntn1*, *Plac1*, *Thsd7a*, *Tinagl1*) izražanje nižje v vzorcih tretiranih z MTX; pri treh genih (*Hmng2*, *Cdh1* in *Psma8*) pa je bilo izražanje v vzorcih, ki so bili tretirani z MTX višje. Pri genih *Gpc6*, *B2bf95.rat* in *hDhfr* izražanje ni bilo statistično značilno različno (Sl. 24).

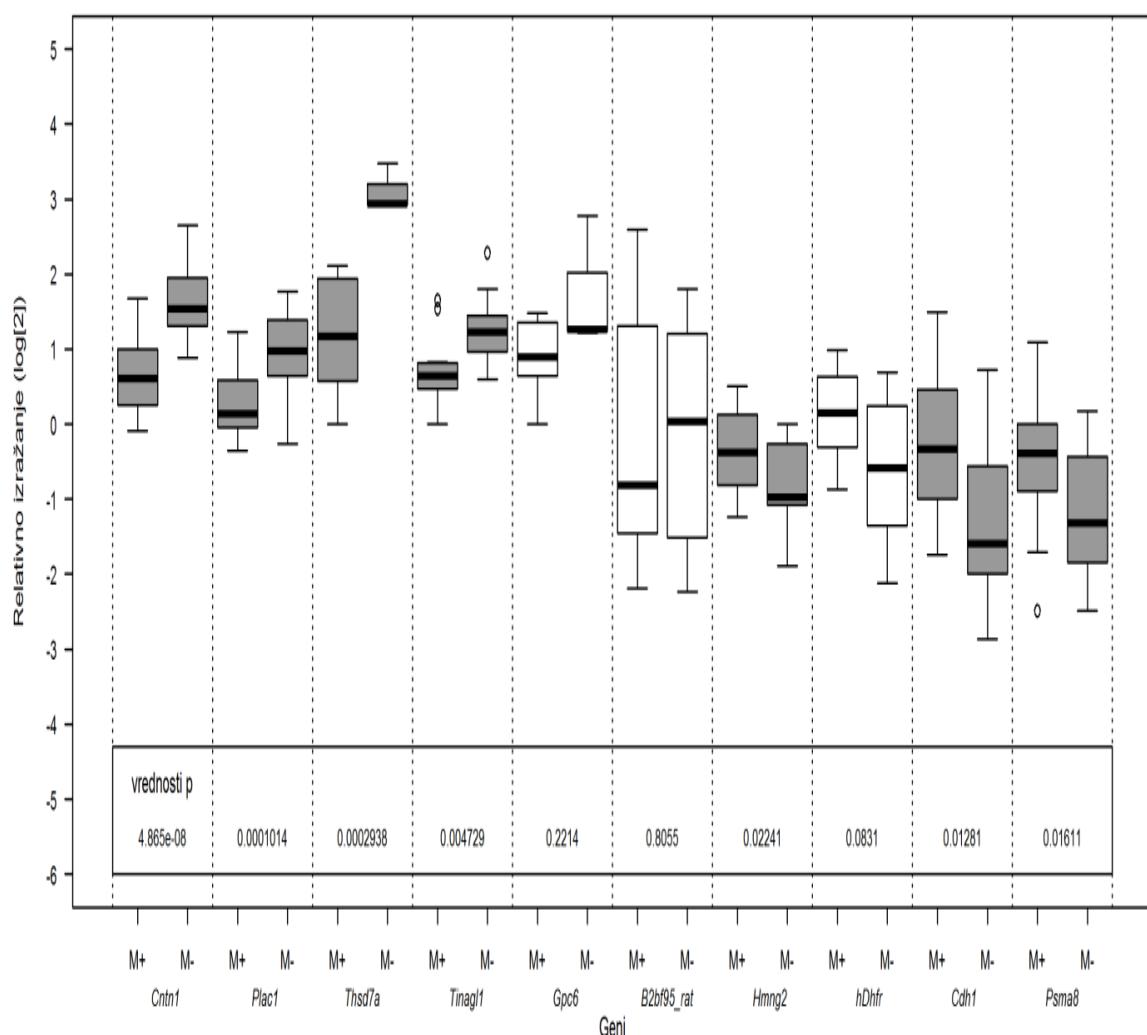
Največjo razliko v izražanju genov med vzorci, tretiranimi z oz. brez MTX smo opazili pri naslednjih dveh genih:

- *Thsd7a* ($\log_2\text{FC}=2$);
- *Cdh1* ($\log_2\text{FC}=1,7$).

Protein 7A z domeno trombospondin, ki ga kodira gen *Thsd7a* je udeležen pri celični diferenciaciji in angiogenezi, medtem ko je protein kadherin-1, ki ga kodira gen *Cdh1*, udeležen pri povezavah med celicami, ki so odvisne od kalcija in pri regulaciji celične adhezije.

Grafikoni kvantilov prikazuje relativno izražanje posameznega gena, analiziranega z metodo RT-qPCR, kjer so ločeno prikazani grafikoni kvantilov za vzorce tretirane z MTX in vzorce, ki niso bili tretirani z MTX (Sl.24). Relativno izražanje posameznega gena smo

izračunali tako, da smo izražanje gena iz posamezne skupine vzorcev (tretirani z in brez MTX) transformirali z \log_2 in normalizirali z AcGa, ki predstavlja geometrijsko povprečje vrednosti Cq za dva referenčna gena *Actb* in *Gapdh*. Za vsak gen smo izračunali statistično značilnost različno izražanja gena med skupino vzorcev. Izmed 10 genov, kjer smo testirali statistično značilno različno izražanje med dvema skupinama vzorcev, smo ugotovili, da se 3 geni (*Gpc6*, *B2bf95_rat in hDhfr*) od 10 izražajo statistično različno (popravljena P-vrednost < 0,05). Za statistično različno izražanje genov med dvema skupinama vzorcev, smo uporabili Welchev dvo-repi test t (Sl. 24).



Slika 24: Grafikon kvantilov prikazuje relativno izražanje izbranih genov med vzorci tretiranih z MTX ali brez tretiranja z MTX.

Grafikoni kvantilov predstavljajo izražanja 10 genov: *Cntn1*, *Plac1*, *Thsd7a*, *Tinagl1*, *Gpc6*, *Hmng2*, *B2bf95_rat*, *hDhfr*, *Cdh1* in *Psma8* transformiranih z \log_2 in normaliziranih z AcGa (geometrijsko povprečje vrednosti Cq za *Actb* in *Gapdh*) pri vzorcih tretiranih z MTX oz. brez MTX. Prag statistično značilne izražanja med vzorci tretiranih z MTX in tistih brez MTX je popravljena P-vrednost $< 0,05$. Grafikoni kvantilov so označeni temno sivo pri genih, kjer je popravljena P-vrednost $< 0,05$; kjer je popravljena P-vrednost $> 0,05$, so grafikoni kvantilov označeni z belo. S testom *t* smo analizirali 46 vzorcev; od tega 22 z MTX (M+) in 24 brez MTX (M-). Osamelci so označeni s krogci. Pod vsakim parom grafikonov za vsak gen so pripisane še vrednosti popravljene P-vrednosti.

Figure 24: Boxplots showing relative expression ratios in clones treated with or without MTX.
 The graphicons represent the gene expression of 10 genes *Ces2g*, *Ac15850.4*, *Spr2h*, *Ankrd26*, *Cd68*, *Flt1*, *Myrip*, *Plekhb1*, *Rpska6*, *Spn1*, *Spr2k* and *Zdhhc6* which are \log_2 transformed and AcGa normalised (geometric mean Cq for *Actb* and *Gapdh* genes) in samples with and without MTX. P-corrected $< 0,05$ is used for statistical significance. The boxplots are colour coded - grey for p- corrected $< 0,05$ and white for P-corrected $> 0,05$. 46 samples were analysed by t-test: 22 samples with MTX (M+) and 24 samples without MTX (M-).

5 RAZPRAVA

Celice CHO so v zadnjih dveh desetletjih postale najpomembnejša celična linija za proizvodnjo rekombinantnih učinkovin. Omogočajo post-translacijske modifikacije rekombinantnih proteinov, ki so kompatibilni in bioaktivni pri ljudeh (Kim in sod., 2012). Zaradi tega so celice CHO postale najpogosteje uporabljen gostiteljski organizem za proizvodnjo rekombinantnega proteina (Andersen in sod., 2001; Chu in sod., 2002).

5.1 ŠTEVilo KOPIJ REKOMBINANTNEGA GENA IN UČINKOVITOST PRODUKCIJE REKOMBINANTNEGA PROTEINA NISTA POVEZANI

Najbolj razširjen sistem za produkcijo rekombinantnega proteina v gostiteljski celici CHO je ekspresijski sistem CHO-Dhfr, ki omogoča visoko produkcijo rekombinantnega proteina (Reff in sod., 1993). Znano je, da imajo celice CHO zelo nestabilen kariotip zaradi kromosomskih reorganizacij, ki so posledica homolognih rekombinacij in translokacij. Kariotip celic CHO je še posebej nestabilen ob dodatku MTX v gojišče, ki je nujen za visoko in stabilno produkcijo rekombinantnega proteina (Bacsı in sod., 1986; Flintoff in sod., 1984; Rolig in sod., 1997; Yoshikawa in sod., 2000). Nujnost prisotnosti MTX v gojišču za učinkovito in stabilno produkcijo rekombinantnega proteina, je razvidna tudi iz naših rezultatov, kjer smo iste vzorce gojili vzporedno z in brez dodanega MTX v gojišču. Pri klonih, kjer smo iz gojišča odvzeli MTX, se je produkcija rekombinantnega proteina znižala, v primerjavi z istimi kloni kjer je bil MTX prisoten v gojišču. MTX ima vpliv na produkcijo in na dolgotrajno stabilnost produkcije rekombinantnega proteina.

V primeru, ko smo iz gojišča odstranili MTX, ki predstavlja selekcijski pritisk, se je število kopij rekombinantnega gena znižalo za 61 % in tudi produkcija rekombinantnega proteina se je znižala za 67% (Sl. 6). Tudi ta rezultat je primerljiv z rezultati ostalih študij, kjer so celice prav tako daljši čas gojili brez prisotnosti MTX v gojišču (Kaufman in sod., 1985; Weidle in sod., 1988). V omenjenih študijah so prav tako poročali o padcu števila kopij rekombinantne učinkovine in o zmanjšanju produkcije rekombinantnega proteina. Iz naših rezultatov in literature, je viden vpliv MTX na učinkovitost in stabilnost produkcije. S

pomočjo MTX-a se zviša produkcija rekombinatne učinkovine in ob stalni prisotnosti MTX v gojišču, je produkcija rekombinatne učinkovine stabilna.

Nestabilna produkcija rekombinantnega proteina v rekombinantnih celicah CHO brez prisotnosti MTX je značilna lastnost CHO-Dhfr ekspresijskega sistema (Fann in sod., 2000). V idealnem primeru bi v proizvodnji rekombinante učinkovine uporabili MTX samo na začetku razvoja produkta z namenom dvigniti produkcijo, kasneje v proizvodnji pa MTX ne bi bilo potrebno uporabljati. To bi bilo možno, če bi bila produkcija rekombinantnega proteina stabilna tudi brez prisotnosti MTX v gojišču. Neprestana prisotnost MTX v gojišču povečuje verjetnost nezaželenih sprememb v genom vstavljenem ekspresijskem vektorju, ki nosi nukleotidni zapis za rekombinant gen.

V literaturi objavljenih študijah, kjer so z DNA-mikromrežami proučevali transkriptom rekombinantnih celic CHO, so se osredotočili predvsem na razlike v izražanju genov med visoko in nizko produktivnimi vzorci in na mehanizme, ki vplivajo na visoko produkcijo rekombinantnega proteina (Doolan in sod., 2008; Kantardjief in sod., 2010; Schaub in sod., 2010; Trummer in sod., 2008). V nekaterih drugih študijah so hkrati iskali gene, ki so povezani z visoko produkcijo rekombinantnega proteina v bioreaktorjih in stresnimi dejavniki (npr. strižne sile zaradi mešanja kulture) za celice, ki so gojene v bioreaktorjih (Carlage in sod., 2009; Meleady in sod., 2008; Nissam in sod., 2006). V naši študiji, smo imeli enake pogoje gojenja celic skozi celotno študijo in smo se osredotočili samo na razliko v izražanju genov tekom gojenja.

Ker je dolgotrajno stabilna učinkovitost proizvodnje rekombinantnega proteina ključna za uspešen razvoj in proizvodnjo vseh bioloških zdravil, smo se v nalogi osredotočili za podrobnejšo analizo transkriptoma vzorcev celic CHO skozi daljše časovno obdobje. Za ekonomično proizvodnjo rekombinantnega proteina je predvsem pomembna visoka produkcija rekombinatne učinkovine. Med primerjavo števila kopij gena z zapisom za rekombinatno učinkovino in produktivnostjo smo ugotovili, da se število kopij gena z zapisom za rekombinatno učinkovino in produktivnost s časom zmanjšujeta. V naši študiji se je število kopij rekombinatne učinkovine zmanjšalo iz začetne 5,4 na 3 kopije na celico v povprečju pri vzorcih, ki so imeli stabilno produkcijo rekombinatne učinkovine v

obdobju 10-tednov, kolikor je trajala študija. Pri vzorcih z nestabilno produkcijo rekombinatne učinkovine je bil upad števila kopij v istem časovnem obdobju višji (iz 5,4 na 2,1 kopije na celico v povprečju) (Sl. 6).

Število kopij rekombinatne učinkovine v celici ni direktno povezano s produktivnostjo. To smo pokazali z našimi rezultati primerjave med učinkovitostjo produkcije in številom kopij rekombinantnega gena. Nekateri vzorci so imeli visoko produkcijo (nad 6 enot), kljub temu, da so imeli različno število kopij rekombinantnega gena (od 2 kopiji rekombinantnega gena do 7 kopij rekombinantnega gena) (Sl.7). Izračunali smo tudi koeficient korelacije ($R^2 = 0,4072$), ki potrjuje, da ne obstaja direktna povezava med številom kopij rekombinantnega gena in produktivnostjo.

To lahko razložimo z različno intenziteto izražanja posameznih kopij rekombinantnega gena. Ker je integracija gena za rekombinatno učinkovino v celični genom naključna, se lahko zgodi, da je integriran gen za rekombinatno učinkovino bolj ali manj izražen, lahko pa se zgodi, da se gen sploh ne izraža. Iz stališča proizvodnje je pomembna predvsem visoka produktivnost, ne glede na število kopij rekombinatne učinkovine, je pa seveda zaželjeno, da je število kopij čim manjše in produkcija čim bolj stabilna. Tako smo pri stabilnih vzorcih ugotovili, da se, kljub upadu povprečnega števila kopij rekombinatne učinkovine (5,4 na 3 kopije na celico), produktivnost ni zmanjšala za več kot 9 %, v primerjavi z nestabilnimi vzorci, kjer je bil padec produkcije rekombinatne učinkovine 67 % (Sl. 6). Na visoko in stabilno produkcijo torej vpliva več znanih in neznanih faktorjev.

5.2 ANALIZA TRANSKRIPTOMA OMOGOČA IDENTIFIKACIJO STATISTIČNO ZNAČILNO RAZLIČNO IZRAŽENIH GENOV

Glavna prednost DNA-mikromrež je možnost hkratnega netarčnega proučevanja izražanja velikega števila genov, glavni pomankljivosti sta, v primerjavi z drugimi metodami za analizo izražanja genov, nizka občutljivost metode in dinamični razpon. Predvsem je namen DNA-mikromrež in kasnejše analize rezultatov identifikacija statistično značilno različno izraženih genov iz velikega nabora genov.

Tehnologija DNA-mikromrež, ki smo jo uporabili v našem poskusu, omogoča vpogled v veliko število genov (približno 26,227) hkrati. Večje kot je število hkrati analiziranih genov, bolj celostno sliko dobimo o proučevanem vzorcu. Ker genom za gostiteljske celice, ki smo jih uporabili v poskusu, ni bil na na voljo, je proizvajalec DNA-mikromrež uporabil genom kitajskega hrčka (*Cricetulus Griseus*), iz katerega so bile celice CHO izolirane. Ker je na genetskem nivoju razlika med celicami CHO, izoliranimi iz ovarijskega kitajskega hrčka in kitajskim hrčkom, smo ob obdelavi podatkov naleteli na veliko genov, ki pri celičnih kulturah nimajo relevantne vloge.

Tako sta gena *Sprr2h* in *Sprr2k* udeležena v razvoju epidermisa in diferenciaciji keratinocitov, gen *Cd68* je udeležen v fagocitotskih aktivnostih tkivnih makrofagov; v metabolizmu lizosomov in interakcijah celica-patogen, *Flt1* regulira angiogenezo in celično diferenciacijo in je udeležen pri embrionalni morfogenezi, *Plac1* je udeležen pri razvoju placente, *Tinagl1* je udeležen pri imunskemu odzivu (Pregl. 11 in Pregl. 12).

Za potrditev rezultatov diferencialno izraženih genov smo uporabili metodo RT-qPCR in analizirali izražanje 36 izbranih genov, vsak gen na vseh 48 vzorcih. Korelacija med izražanjem 36 genov na vseh 48 vzorcih med metodo RT-qPCR in DNA-mikromrežami je pokazala visoko korelacijo ($R^2 = 0,8116$) rezultatov obeh metod (Sl. 14).

Z analizo transkriptoma vzorcev smo se osredotočili predvsem na 3 sklope, kjer smo iskali statistično značilno izražene gene, ki bi jih lahko uporabili v razvoju rekombinantnih proteinov:

- a) stabilnost produkcije rekombinantnega proteina skozi čas (stabilnost);
- b) učinkovitost produkcije rekombinantnega proteina (produktivnost);
- c) vpliv MTX na produkcijo rekombinantnega proteina (tretiranje z / brez MTX).

5.3 OD ANALIZE TRANSKRIPTOMA DO IZBORA GENOV, POMEMBNIH ZA STABILNOST PRODUKCIJE REKOMBINANTE UČINKOVINE

Z namenom razumevanja razlik med vzorci s stabilno in med vzorci z nestabilno produkcijo rekombinatne učinkovine, smo z analizo transkriptoma primerjali vzorce s stabilno in nestabilno produkcijo rekombinatnega proteina. S podrobno statistično analizo podatkov iz DNA-mikromrež smo zožili število genov na 14 genov (*Fgfr2*, *BX842664.2/Hist1h3c*, *AC115880.11*, *E130203b14Rik*, *hDhfr*, *Hist1h2bc*, *Cspg4*, *C1qtnq*, *Foxp2*, *Mmp10*, *Vsnl1*, *CU459186.17*, *Egr1* in *Ptpre*) uporabnih za razlikovanje med vzorci s stabilno in nestabilno produkcijo rekombinantnega proteina. Metoda RT-qPCR je, v primerjavi z DNA-mikromrežami, veliko bolj natančna, vendar se običajno uporablja le za analizo izražanja manjšega števila genov.

Relativno izražanje 14 statistično značilno različno izraženih genov iz primerjave vzorcev s stabilno in nestabilno produkcijo rekombinante učinkovine, smo primerjali med metodama RT-qPCR in DNA-mikromrežami. Ugotovili smo, da obstajajo razlike med izražanjem genov analiziranih z obema metodama (Sl. 15). Razlike v izražanju genov med obema metodama so bile pričakovane. Pri tej primerjavi je pomembno, da je izražanje gena v obeh primerih regulirano v isto smer, kar pri večini naših rezultatov ni bil problem. Samo 3 geni (*hDhfr*, *Ptpre* in *C1tnf1*) pri proučevanju stabilnosti produkcije rekombinantne učinkovine skozi čas, niso bili regulirani v isto smer pri obeh metodah. Izražanje preostalih 11 genov je bilo pri obeh metodah reguliranih v isto smer (Sl. 15).

Med primerjavo vzorcev s stabilno in nestabilno produkcijo rekombinantnega proteina smo ugotovili, da se pri vzorcih s stabilno produkcijo rekombinantnega proteina 7 genov višje izraža v primerjavi z vzorci, ki imajo nestabilno produkcijo rekombinantnega proteina (Sl. 16). Šest izmed 7 genov je udeleženih v regulaciji na različnih nivojih, od diferencijacije, migracije, signalnih poti, transkripcije in podvojevanja DNA. Gen *hDHFR* kodira encim DHFR, ki je ključen v metabolizmu folatov. Dva (*E130203b14Rik* in *Cu459186.17*) izmed 14 genov nimata znane funkcije (Pregl. 10). Gen *E130203b14Rik* je pri vzorcih s stabilno produkcijo nižje, gen *Cu459186.17* pa višje izražen v primerjavi z vzorci z nestabilno produkcijo rekombinante učinkovine (Sl. 16).

V naslednjem koraku smo primerjali izraženost istih 14 genov samo med vzorci s stabilno produkcijo v različnih časovnih točkah. Njihovo izražanje smo primerjali na skupno 48 vzorcih na začetku študije (teden 1 in 2) in na koncu študije (teden 9 in 10). Pri večini genov (12 od 14) se izraženost genov med začetkom in koncem študije pri stabilnih vzorcih ne spreminja, kar pomeni, da lahko stabilnost produkcije napovemo že v začetnih fazah razvoja celične linije in so zato dobri kandidati za gene, ki so povezani s stabilnostjo produkcije rekombinantnega proteina (Sl. 17).

Napovedovanje stabilnosti produkcije z izraženostjo določenih genov do sedaj ni bila opisana v literaturi. V študiji (Dorai in sod., 2012) so opisali poskus znižanja stroškov pri razvoju rekombinantnega proteina v celicah CHO z odstranitvijo vzorcev, ki so vsebovali markerje za apoptozo, kot sta visoka izraženost aneksina 5 in kaspaze 3. Pomankljivost opisane metode je predvsem v neselektivnosti, ker ne omogoča odstranitve samo tistih celic, ki imajo izražen apoptotični marker, ampak je potrebno zavreči celoten vzorec celic, kjer so prisotni apoptotični markerji. Dodatno je potrebno, da je celična kultura relativno starja (pozna eksponentna/stacionarna faza rasti), da se apoptotični markerji dovolj izrazijo in jih je možno zaznati. V zgodnji fazi razvoja celične linije, se gojišče v celični kulturi menjava vsak drugi ali tretji dan, kar pomeni, da so celice v kulturi vseskozi v zgodnji eksponentni fazi rasti z visoko celično viabilnostjo ($> 95\%$), kjer je prisotnih zelo malo apoptotičnih celic.

Prav tako so (Barnes in sod., 2003) razvili Northern blot metodo, kjer so spremljali količino mRNA rekombinantnega proteina v vzorcih in iz razvoja odstranili vzorce z nizko količino mRNA rekombinantnega proteina. Pri našem pristopu nismo povezovali rezultatov izražanja statistično značilno različno izraženih genov s količino mRNA, ampak s produkcijo rekombinante učinkovine. Med količino mRNA in višino produkcije rekombinante učinkovine je veliko regulacijskih stopenj (transkripcija, translacija, post-translacijske modifikacije), ki pomembno vplivajo na produkcijo in kakovost končnega produkta, še posebej, če je produkt kompleksno sestavljen molekula kot v našem primeru.

Opisane metode zahtevajo veliko dela in so neprimerne za veliko količino vzorcev, ki nastaja pri razvoju biološko podobnih zdravil, predvsem pa niso tako natančne kot metoda qPCR. Zraven tega pa bi lahko zgoraj opisan pristop z Northern blot metodo prenesli na metodo RT-qPCR, kar bi poenostavilo izvedbo analize in posledično omogočilo analizo večjega števila vzorcev.

Poleg tega uporaba DNA-mikromrež omogoča pregled hkratnega izražanja velikega števila genov in izbor genov, ki so statistično značilno drugače izraženi med vzorci, brez predhodnih informacij o samem genu. Pri uporabi DNA-mikromrež se ni potrebno omejiti samo na določene gene, katerih izražanje proučujemo, ampak dobimo veliko širšo sliko izražanja velikega števila genov. In prav takšen pristop nam omogoča, da lahko najdemo morebitne pomembnejše gene za specifičen namen. V našem primeru smo se omejili na iskanje specifičnega vzorca izražanja genov, katerih izražanje lahko uporabimo za napoved stabilne produkcije rekombinantnega proteina že v začetku razvoja celične linije. Na takšen način lahko skrajšamo testiranje stabilnosti produkcije na vzorcih za 80 %. Kar pomeni ogromen prihranek časa in denarja pri razvoju rekombinantnih učinkovin.

Z analizo qPCR rezultatov smo pokazali, da lahko s kombinacijo izražanja 3 identificiranih genov (*Hist1h2bc*, *Egr1*, *hDhfr*) ločimo vzorce v skupino stabilnih in skupino nestabilnih vzorcev že v začetku (v prvem ali drugem tednu) razvoja rekombinantnega proteina (Sl. 18). Razdelitev vzorcev na osnovi izraženosti 3 genov na vzorce s stabilno in nestabilno produkcije rekombinantnega proteina, je ostala nespremenjena do konca 10-tednov dolge študije (Sl. 19). Skupina stabilnih vzorcev se na začetku študije razdeli v svojo skupino, katera se je ohranila ločena od skupine nestabilnih vzorcev od začetka do konca študije. Že v začetku študije smo videli, da se je v skupino nestabilnih vzorcev uvrstil tudi en stabilen vzorec (Sl. 18). Kar je glede na večje število (15) stabilnih vzorcev zelo majhna napaka. Vsekakor je veliko boljše, da se zaradi napake uvrsti en stabilen vzorec v skupino nestabilnih, kot pa obratno. V tem primeru pomeni, da bi zavrgli manj kot 7 % stabilnih vzorcev, ker bi jih zaradi napake uvrstili v skupino vzorcev z nestabilno produkcijo (Sl. 18 in Sl. 19).

Prav tako smo razdelitev vzorcev v skupino stabilnih in skupino nestabilnih vzorcev potrdili z metodo k-najbližjega soseda, kjer je algoritem s 85 % zanesljivostjo razdelitev vzorcev v eno izmed dveh skupin (brez slike).

Natančnost razvrščanja smo poskušali popraviti z vključevanjem dodatnih genov in z alternativno metodo analize PCA. Analizirali smo 5 genov (*Fgfr2*, *BX842664.2/Hist1h3c*, *E130203b14Rik*, *Cspg4* in *PtPre*), statistično značilno izraženih samo med stabilnimi in nestabilnimi vzorci in katerih izražanje ni odvisno od prisotnosti MTX v gojišču. Na ta način smo z večjo gotovostjo (92 % zanesljivost) razdelili vzorcev v skupino s stabilno in nestabilno produkcijo rekombinantne učinkovine (Sl.20).

5.4 GENI POVEZANI Z UČINKOVITOSTJO PRODUKCIJE REKOMBINANTNEGA PROTEINA

Relativno izražanje 12 statistično značilno različno izraženih genov iz primerjave vzorcev z visoko in nizko produkcijo rekombinantne učinkovine, smo primerjali med metodama RT-qPCR in DNA-mikromrežami. Ugotovili smo, da obstajajo razlike med izražanjem genov analiziranih z metodo RT-qPCR in DNA-mikromrežami (Sl. 21). Razlike v izražanju genov med obema metodama so bile pričakovane. Razločno se vidi trend izražanja genov v isto smer, tako pri DNA-mikromrežah kot pri metodi RT-qPCR. V nekaterih primerih (*Ankrd26* in *Zdhc6*) se izražanje genov zelo pokriva (manj kot 10 % odstopanje) med analizo izražanja z metodo RT-qPCR in DNA-mikromreže (Sl. 21).

Prav tako smo analizirali razlike v izražanju genov med vzorci z visoko in nizko produkcijo rekombinantne učinkovine, skozi 10-tednov dolgo študijo. Izmed 12 kandidatnih genov (*Ces2g*, *Ac15850.4*, *Spr2h*, *Ankrd26*, *Cd68*, *Flt1*, *Myrip*, *Plekhb1*, *Rpska6*, *Spn1*, *Spr2k* in *Zdhc6*), kjer smo z metodo RT-qPCR potrjevali izražanje genov, smo pokazali, da so vsi geni statistično značilno različno izraženi med vzorci z visoko in nizko produktivnostjo, ne glede na čas vzorčenja (Sl. 22). Trije (*Ankrd26*, *Myrip*, *Zdhc6*) izmed 12 genov so bili višje izraženi pri vzorcih z visoko produkcijo v primerjavi z vzorci z nizko produkcijo rekombinantnega proteina. Gena *Ankrd26* in *Zdhc6* sta udeležena pri

regulaciji metabolizma maščobnih kislin in pri pravilnem zvijanju proteinov. Gen *Myrip* je udeležen pri intracelularnem transportu proteinov in vezikularnem transportu vzdolž aktinskih filamentov (Pregl. 11). Glede na vlogo genov, ki so višje izraženi pri visokih producentih v primerjavi z nizkimi producenti, je pričakovati, da so omenjeni geni višje izraženi pri visokih producentih. Povišano izražanje genov udeleženih pri pravilnem zvijanju proteinov in pri transportu proteinov pomeni, da se omenjeni procesi v celici odvijajo hitreje, kar pomeni višjo produkциjo rekombinantne učinkovine. Višje izražanje gena *Ankrd26* pomeni tudi višjo regulacijo metabolizma maščobnih kislin, kar posledično pomeni, da ima celica na voljo več energije za celično rast.

5.5 VPLIV MTX NA PRODUKCIJO REKOMBINANTNEGA PROTEINA

Pri 10 genih smo z RT-qPCR preverili spremenjeno izražanje genov med vzorci tretiranimi z in brez MTX v gojišču. Razločno se vidi enak trend izražanja genov, tako pri DNA-mikromrežah, kot pri metodi RT-qPCR (Sl. 23). V nadaljevanju smo analizirali razlike v izražanju genov med vzorci tretiranimi z in brez MTX v gojišču. Izmed 10 kandidatnih genov (*Cntn1*, *Plac1*, *Thsd7a*, *Tinagl1*, *Hmng2*, *Cdh1*, *Psma8*, *Gpc6*, *B2bf95_rat*, *hDHFR*), kjer smo z metodo RT-qPCR potrjevali izražanje genov, smo pokazali, da je 7 genov (*Cntn1*, *Plac1*, *Thsd7a*, *Tinagl1*, *Hmng2*, *Cdh1*, *Psma8*) statistično značilno izraženih med vzorci tretiranimi z in brez MTX v gojišču (Sl. 24). Pri tem je zanimivo, da se gen *hDHFR* ni statistično značilno izražal med primerjavo izražanja vzorcev tretiranih z in brez MTX, čeprav se je statistično značilno izražal med vzorci s stabilno in nestabilno produkциjo rekombinantne učinkovine (Sl. 16 in Sl. 24).

Izmed 7 statistično značilno različno izraženih genov, so 3 geni (*Hmng2*, *Cdh1*, *Psma8*) višje izraženi v vzorcih tretiranih z MTX v primerjavi z vzorci, ki niso bili tretirani z MTX (Sl. 24). Gen *Hmgn2* (Pregl. 12) je udeležen pri regulaciji transkripcije (DNA vezavni proteini), tako kot geni *BX842664.2/Hist1h3c*, *Hist1h2bc* in *AC115880.11* iz sklopa stabilne produkcijske rekombinantne učinkovine skozi čas (Pregl. 10). Glede na znane biološke funkcije in naše rezultate, so omenjeni geni najverjetne medsebojno testno povezani. Gen *Psma8* je udeležen pri metabolizmu razgradnje proteinov in je po biološki

funkciji (Pregl. 12) zelo podoben genu *Mmp10* (Pregl. 10), ki je višje izražen pri vzorcih s stabilno produkcijo rekombinante učinkovine (Sl. 16). Gen *Cdh1* je po biološki funkciji udeležen pri povezavah med celicami in pri regulaciji celične adhezije (Pregl. 12). Noben drug proučevan gen v naši študiji nima podobne biološke funkcije (Pregl. 10, 11 in 12). Kot smo pokazali v naših rezultatih prisotnost MTX v gojišču posredno vpliva na produkcijo in stabilnost produkcije rekombinante učinkovine (Sl. 6). Glede na to, da smo naredili poskus na istih klonih, kjer je MTX bil prisoten v gojišču in smo ga kasneje pri eni paraleli vzorcev odvzeli iz gojišča, samo integracijsko mesto nima vpliva na razlike v produkciji in stabilnosti rekombinantnega proteina. V naših rezultatih (Sl. 6) smo opazili, da kloni brez stalne prisotnosti MTX-a v gojišču, sčasoma izgubljajo kopije ekspresijskega vektorja. Iz tega vidika so geni, ki so višje izraženi pri vzorcih tretiranih z MTX najverjetneje neposredno udeleženi pri učinkovitosti in stabilnosti produkcije rekombinante učinkovine.

5.6 MESTO INTEGRACIJE REKOMBINANTNEGA GENA V GOSTITELJSKI GENOM IMA POMEMBNO VLOGO PRI STABILNEM IZRAŽANJU REKOMBINANTNEGA GENA

Raziskave so pokazale, da so pomnoženi rekombinanti geni locirani blizu telomernih regij bolj stabilni.

V študiji je Dickinson (2009) z MTX pomnožil rekombinantni gen in z metodo fluorescenčne hibridizacije *in situ* (FISH) določil lokacijo le-tega. Pri nekaterih vzorcih je bila lokacija rekombinantnih genov na telomerni regiji kromosomov, pri nekaterih pa na ostalih delih kromosomov. Vsi kloni celic z visoko produktivnostjo so imeli rekombinantni gen lociran na telomerni regiji kromosoma in so imeli 6 krat višjo produktivnost v primerjavi s kloni, kjer je bila lokacija rekombinantnega gena na drugih delih kromosoma. Glede na rezultate študije kjer so z metodo FISH locirali mesto integracije rekombinantnega gena, je lokacija rekombinantnega gena na kromosому, v tesni povezavi z visoko produktivnostjo.

V isti študiji so ugotovili, da je lokacija vstavljenega rekombinantnega gena odvisna tudi od koncentracije MTX v gojišču. Razlike so opazili ob postopnem zviševanju koncentracije MTX v gojišču. V tem primeru se je pri 80% vzorcev rekombinanti gen vstavil v telomerno regijo kromosomov. Po drugi strani pa se je rekombinanti gen vstavil v telomerno regijo samo pri 40 % vzorcev, katerim so koncentracijo MTX v gojišču dvignili brez predhodnih nižjih koncentracij.

V nadaljnjih korakih študije so proučevali še razlike med vzorci v številu kopij rekombinantnega gena in mestom integracije rekombinantnega gena na kromosому. Rekombinanti gen vstavljen v telomerno regijo je bil prisoten v več kopijah v primerjavi z rekombinantnim genom vstavljenim v ostale regije kromosomov. Večje število kopij rekombinantnega gena v telomerni regiji bi naj nastalo z dodatnim pomnoževanjem le-tega zaradi selekcijskega pritiska, ki ga predstavlja MTX. To pa je potrebno, da se rekombinanti gen, skupaj z rekombinantnim genom *dhfr* ohranja v velikem številu kopij.

Hipotezo so skušali potrditi s primerjavo števila kopij nativnega gena *dhfr* v ne-transfenciranih celicah in v vzorcih celic, kjer se je rekombinantni gen *dhfr* vgradil v telomerno regijo kromosomov. Nativni gen *dhfr* je lociran v "sredini" kromosoma, kjer je manjša možnost utišanja gena v primerjavi s telomerno regijo. Ugotovili so, da je aktivnost gena *dhfr* lociranega na "sredini" kromosoma in v telomerni regiji enaka. Tako so pokazali, da so rekombinantni geni locirani na telomerni regiji kromosov stabilno integrirani in niso podvrženi utišanju. Z omenjenim poskusom so pokazali, da za ohranitev gena *dhfr* na telomernih regijah za preživetje celice ob prisotnosti MTX v gojišču pomnoževanje ni nujno potrebno. Najverjetneje so za preživetje celice ob prisotnosti MTX v gojišču zraven gena *dhfr* vključeni tudi neznani mehanizmi, kot je npr. sprememba v propustnosti celičnih membran kot posledica mutacij.

V nadaljevanju so dodatno še ugotovili, da se število celic z rekombinantnim genom, vstavljenim v telomerni regiji, gojene 50 dni brez prisotnosti MTX v gojišču, ni zmanjšalo. Med tem, ko se je število celic z rekombinantnim genom, vstavljenem na drugih regijah, postopoma zmanjševalo.

Rezultati opisane študije dokazujejo, da so rekombinantni geni vstavljeni v telomerno regijo kromosoma stabilni, tudi brez prisotnosti selekcijskega pritiska (MTX) in prispevajo k višji produktivnosti v primerjavi z vzorci z rekombinantnim genom lociranim na ostalih delih kromosoma (Dickinson, 2009).

V študiji so Karpinski in sod. (2009) opisali poskuse na celicah CHO, kjer so proučevali stabilnost vstavljenih genov v vzorcih z in brez vsebnosti MTX v gojišču. Ugotovili so, da so z MTX pomnoženi rekombinantni geni bolj stabilni, ob stalni prisotnosti MTX v gojišču. Z namenom pridobitve vzorcev s pomnoženimi rekombinantnimi geni, so v poskusu gojili celice ob prisotnosti MTX v gojišču, ki inhibira encim DHFR. S povečevanjem koncentracije MTX v gojišču, se je število viabilnih celic zelo zmanjšalo, preživele so samo celice, kjer se je pomnoževal tudi vstavljen rekombinant gen *dhfr*. Med dolgotrajnim gojenjem istih vzorcev brez vsebnosti MTX v gojišču, so ugotovili, da se je rekombinantni gen izločil iz celic gostiteljske DNA. Prav tako pa se je znižala produkcija rekombinantnega proteina. Z metodo FISH so analizirali lokacijo vstavljenih genov v gostiteljsko kromosomske DNA med gojenjem vzorcev z in brez prisotnosti MTX v gojišču. Lokacija vstavljenih genov je bila odvisna od naraščajoče koncentracije MTX v gojišču. Kadar so uporabili večstopenjsko naraščanje koncentracije MTX v gojišču, je bilo 82 % vstavljenih in z MTX pomnoženih rekombinantnih genov na telomernih regijah. Ko pa so uporabili visoko koncentracijo MTX v gojišču, brez predhodnjih postopnih zviševanj koncentracije, se je število pomnoženih, rekombinantnih genov na telomernih regijah znižalo.

V nadaljevanju so v študiji raziskali razmerje med stabilno integracijo pomnoženih rekombinantnih genov in njihovo lokacijo na kromosому. Uporabili so vzorce, kjer je bil rekombinantni gen vstavljen na telomerah. Vzorce so gojili 50 dni z in brez MTX v gojišču. Med študijo, kjer je bil MTX prisoten ves čas, je bila produktivnost nespremenjena skozi celoten čas študije. V tem primeru se tudi ni spremenila lokacija rekombinantnega gena *dhfr* na kromosomu. Pri vzorcih, kjer MTX ni bil prisoten ves čas študije, se je produktivnost po 46 dneh kultivacije zmanjšala za 30 %. Zmanjšala se je tudi prisotnost rekombinantnih genov na telomernih regijah, kar so ugotovili z metodo FISH.

Rezultati študije so pokazali, da so med dolgotrajno kultivacijo, pomnoženi rekombinantnimi geni locirani blizu telomerne regije bolj stabilni ob stalni prisotnosti MTX v gojišču. Prav tako so rezultati študije pokazali, da je lokacija rekombinantnih genov na kromosomu povezana z njihovim izražanjem (Karpinski in sod., 2009).

V naslednji študiji so Kim in sod. (2012) proučevali vpliv prisotnosti MTX v gojišču na lokacijo rekombinantnega gena na različnih lokacijah kromosomov. Ob dodatku MTX v gojišče se je povečala raznolikost mest na kromosomih, kamor se je vstavil rekombinantni gen. Več kot 60 % vzorcev je imelo drugačno integracijsko mesto rekombinantnega gena, kot pred dodatkom MTX v gojišče. Glede na rezultate metode FISH, se je število integracijskih mest povečalo znotraj enega kromosoma. Prav tako so se integracijska mesta razširila na več kromosomov. Ob odstranitvi MTX iz gojišča se je zmanjšalo število integracijskih mest rekombinantnega gena. Tudi integracijska mesta so postala bolj homogena. Približno 99 % vzorcev je imelo nizko število integracijskih mest rekombinantnega gena, večinoma locirana na telomerni regiji kromosoma. Omenjeni vzorci so imeli nizko število integracijskih mest, vendar so bili rekombinanti geni stabilno integrirani v telomere in so ohranili stabilnost izražanja rekombinantnega gena tudi brez prisotnosti MTX v gojišču (Kim in sod., 2012).

Opisane študije potrjujejo pomembnost lokacije, kamor se rekombinantni gen integrira v kromosom gostiteljske celice. Ob postopnem zviševanju koncentracije MTX v gojišču, je večja verjetnost, da se bo rekombinanti gen stabilno integriral v telomerno regijo kromosoma. V naši študiji smo uporabili samo visoko koncentracijo MTX, brez postopnega zviševanja koncentracije MTX v gojišču. Ob tem smo ugotovili, da ni povezave med številom kopij rekombinantnega gena in produkcijo rekombinantnega proteina (Sl. 7). Prav tako smo ugotovili, da imajo naši vzorci nestabilno produkcijo rekombinantnega proteina kadar MTX ni stalno prisoten v gojišču, kar glede na opisane študije (Karpinski in sod., 2009, Dickinson, 2009 in Kim in sod., 2012) najverjetneje pomeni, da rekombinanti gen ni integriran v telomerno regijo kromosoma.

6 SKLEPI

V sklopu doktorskega dela, smo analizirali transkriptom klonov gostiteljsih celic CHO, ki proizvajajo rekombinantno učinkovino. Ob tem smo identificirali statistično značilno izražene gene, ki nam bodo v prihodnosti olajšale izbor klonov za nadaljni razvoj celičnih linij, potrebnih za proizvodnjo biološko podobnih zdravil. Na osnovi vseh opravljenih poskusov in analiz med visoko/nizko producirajočimi vzorci in med vzorci s stabilno/nestabilno produkcijo rekombinantnega proteina, smo prišli do zaključka, da smo našli statistično značilno različno izražene gene, ki so koristni pri napovedovanju stabilne in visoke produkcije rekombinantnega proteina.

Identifikacija statistično značilno različno izraženih genov za razlikovanje vzorcev z visoko in stabilno produkcijo ima veliko dodatno vrednost pri razvoju rekombinantnih učinkovin. Glede na to, da je razvoj bioloških zdravil zelo dolgotrajen in finančno zahteven proces, je vsako skrajšanje procesa zelo dobrodošlo. Ena izmed daljših študij, ki potekajo v začetnem razvoju rekombinantnega proteina je testiranje stabilnosti produkcije, ki traja več tednov. Producijski klon mora namreč imeti stabilno produkcijo rekombinante učinkovine, od začetka do konca proizvodnje rekombinantnega proteina. Proizvodnja rekombinante učinkovine traja od nekaj tednov do nekaj mesecev, kar je odvisno od rastnih lastnosti celic in od samega postopka proizvodnje. Z uporabo statistično značilno različno izraženih genov, lahko že v zelo zgodnjo fazo razvoja celičnih linij, predvidimo kateri kloni so visoki producenti s stabilno produkcijo rekombinantnega proteina skozi časovno obdobje 10-tednov. Toliko časa je potrebnega od odmrznitve viale, ki vsebuje dovolj celic za namnožitev več tisoč litrskega bioreaktorja, kjer poteka proizvodnja rekombinantnega proteina na producijski skali, do konca produkcije. Na osnovi rezultatov naše študije lahko prihranimo 80% časa, potrebnega za potrjevanje stabilne produkcije rekombinante učinkovine.

Kljub temu, da natančen mehanizem, ki omogoča stabilno in visoko produkcijo rekombinantnega proteina v gostiteljskem sistemu celice CHO ni poznan, smo identificirali statistično značilno izražene gene, katerih izražanje lahko uporabimo za analizo pri izboru najprimernejšega klena za nadaljni razvoj rekombinantnega proteina.

Potrdili smo hipotezo, da bomo z opisanim pristopom identificirali statistično značilno izražene gene za napovedovanje stabilne produkcije rekombinantnega proteina. Prav tako smo potrdili hipotezo, da bomo zaznali razlike v izražanju genov med vzorci z visoko in nizko produkcijo rekombinante učinkovine. Pokazali smo tudi, da lahko razlike med vzorci z različno učinkovitostjo in stabilnostjo produkcije analiziramo z metodo RT-qPCR.

Kljud temu, da smo izvedli poskus z relativno malo število klonov (6) skozi daljše časovno obdobje, smo prišli do iskanih statistično značilno različno izraženih genov, ki jih lahko uporabimo za skrajšanje razvoja rekombinantnega proteina v prihodnosti. Seveda, pa je v prihodnje potrebno testirati uporabnost identificiranih genov na večjem številu vzorcev, kar zaradi časovnih in prostorkih omejitev v tej študiji ni bilo mogoče.

Ne nazadnje, se rezultati in uporabnost našega dela odraža tudi ob prijavi dveh patentov (PAT056379 in PAT056361) na evropsko patentno organizacijo.

7 POVZETEK (SUMMARY)

7.1 POVZETEK

Pri proizvodnji proteinov v mikrobnih ekspresijskih sistemih proteini niso glikozilirani, ali pa se glikozilacija razlikuje od tiste v humanih celicah. Zato se je zaradi potrebe po proizvodnji kompleksnejših rekombinantnih proteinov začel razvoj sesalskih celičnih linij, ki omogočajo post-translacijske modifikacije, značilne za humane celice. Največkrat se uporablja celice CHO, ki so najpomembnejša celična linija za proizvodnjo rekombinantnih proteinov, saj mogočajo gojenje po t. i. standardih dobre proizvodne prakse.

Proizvodnja rekombinantnih proteinov v sesalčih celicah je danes zelo razširjen način proizvodnje rekombinantnih proteinov (učinkovin). Celice CHO so najbolj razširjen gostiteljski organizem za rekombinantno proizvodnjo proteinov. Imajo sposobnost pravilnega sestavljanja proteinov, predvsem pa so pomembne potranslacijske spremembe, kot je pravilno zvijanje proteinov. Vse omenjene lastnosti so ključne za varno delovanje rekombinantnega proteina pri pacientih. Zraven tega so celice CHO, tudi zelo dobro karakterizirane, kar je pripomoglo, da so regulatorni organi izdali številna dovoljenja za uporabo učinkovin za zdravljenje pacientov.

Na eni strani je slabost celic CHO nizka produktivnost, vendar le to lahko povečamo z uporabo sistemov za amplifikacijo genov; kot npr. selekcijski pritisk na encim DHFR. Po drugi strani jih lahko preprosto prilagodimo za rast v suspenzijskih gojiščih brez prisotnih živalskih komponent, poznamo preproste načine za vnos genov (transfekcijo), ter za namnoževanje in selekcijo klonov, ki proizvajajo velike količine rekombinantne učinkovin. Celice CHO so znane po zelo nestabilnem kariotipu zaradi prerazporeditve kromosomov, ki so posledica translokacij in rekombinantnih rekombinacij, še posebej, kadar celice tretiramo z MTX.

V sklopu doktorskega dela, smo analizirali transkriptom klonov gostiteljsih celic CHO, ki proizvajajo rekombinantno učinkovino. Ob tem smo identificirali statistično značilno

različno izražene gene, ki nam bodo v prihodnosti olajšale izbor klonov za nadaljni razvoj celičnih linij, potrebnih za proizvodnjo biološko podobnih zdravil. Na osnovi vseh opravljenih poskusov in analiz med visoko/nizko producirajočimi vzorci, med vzorci s stabilno/nestabilno produkcijo rekombinantnega proteina in med vzorci tretiranimi z in brez MTX, smo prišli do zaključka, da smo identificirali gene, ki so koristni pri napovedovanju stabilne in visoke produkcije rekombinantnega proteina. Statistično značilno različno izražene gene smo dodatno potrdili z RT-qPCR. V nadaljevanju smo z algoritmom k-najbližjega soseda pokazali, da lahko uporabimo kombinacijo petih genov za napoved stabilnega izražanja rekombinantnega proteina, že v začetku razvoja celične linije.

7.2 SUMMARY

In the protein production in microbial expression systems the proteins are not glycosylated or the glycosylation is not comparable to the human one. Because microbial expression systems are not suitable for production of more complex recombinant proteins, the production of more complex proteins is performed in mammalian cell lines. For that purpose the CHO cells, originating from Chinese hamster ovary (*Cricetus griseus*) is mainly used. The CHO cells are the most important cell line for the production of recombinant proteins, because they can be grown based on GMP level.

The production of recombinant proteins in mammalian cells is very popular way of the production of recombinant proteins. CHO cells are capable of correct folding of proteins and posttranslational modifications. All that is critical for safe product when it is consumed by patients. Beside that the CHO cells are very good characterised which was important for approving the product by regulatory organs.

The disadvantage of CHO cells is low productivity which can be enhanced by gene amplification – like introducing selection pressure on DHFR enzyme. In the other hand they can easily grow in media without any origin of animal components and they can be easily transfected with expression vectors. CHO cells have very unstable karyotype because of chromosome rearrangements, specially when treated with MTX.

In doctoral dissertation we have analysed the transcriptome of CHO clones producing the recombinant protein. We have identified (DNA microarrays) genes for easier clone selection in the cell line development which can be used for further improvement in the production of biosimilars. Based on performed analyzes between high/low productive samples, samples with stable/unstable and MTX treated/untreated samples the significantly differently expressed genes were found. The genes can be used for predicting the high and stable production samples already early in the cell line development.

The significantly differently expressed genes identified by DNA microarrays were confirmed by RT-qPCR method. Furthermore we have showed, by k-nearest neighbour algorithm, that combined expression of 5 genes can be used for predicting stable producing samples already at the beginning of cell line development.

8 VIRI

- Afonina I., Zivarts M., Kutyavin I., Lukhtanov E., Gamper H., Meyer R. B. 1997. Efficient priming of PCR with short oligonucleotides conjugated to a minor groove binder. *Nucleic Acid Research*, 25, 2: 2657-2660
- Altschul S. F., Gish W., Miller W., Mayers E. W., Lipman D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215, 4: 403-410
- Andersen D.C., Krummen L. 2002. Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology*, 13, 2: 117-123
- Bacsi S. G., Wejksnora P. J. 1986. Effect of increase in ploidy on the activation of nucleolar organizer regions in Chinese hamster ovary (CHO) cells. *Experimental Cell Research*, 165, 3: 283-289
- Barnes L. M., Bentley C.M., Dickson A.J. 2003. Stability of protein production from recombinant mammalian cells. *Biotechnology and Bioengineering*, 81, 6: 631-639
- Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society series - Statistical Methodology*, 57, 1: 289-300
- Bolstad B.M., Irizarry R.A., Astrand M., Speed T.P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19, 2: 185-193
- Bustin S. A. 2005. Real-time, fluorescence-based quantitative PCR: a snapshot of current procedures and preferences. *Expert Review of Molecular Diagnostics*, 5: 493-498
- Butler M. 2005. Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. *Applied Microbiology and Biotechnology*, 68, 4: 283-291
- Carlage T., Hincapie M., Zang L., Lyubarskaya Y., Madden H., Mhatre R., Hancock W.S. 2009. Proteomic profiling of a high-producing Chinese hamster ovary cell culture. *Analytical Chemistry*, 81, 17: 7357-7362
- Chu L., Robinson D. K. 2001. Industrial choices for protein production by large-scale cell culture. *Current Opinion in Biotechnology*, 12, 2: 180-187
- Dickinson A. 2009. Importance of genetic environment for recombinant gene expression V: Cell Line Development. Al-Rubeai M. (ed.). Dordrecht, Springer: 83-96

- Dorai H., Corisdeo S., Ellis D., Kinney C., Chomo M., Hawley-Nelson P., Moore G., Betenbaugh M.J., Ganguly S. 2012. Early prediction of instability of Chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. *Biotechnology and Bioengineering*, 109, 4: 1016-1030
- Doolan P., Melville M., Gammell P., Sinacore M., Meleady P., McCarthy K., Francullo L., Leonard M., Charlebois T., Clynes M. 2008. Transcriptional profiling of gene expression changes in a PACE-transfected CHO DUKX cell line secreting high levels of rhBMP-2. *Molecular Biotechnol*, 39, 5: 187-199
- Durocher Y., Butler M. 2009. Expression systems for therapeutic glycoprotein production. *Current Opinion in Biotechnology*, 20, 1: 700-707
- Ginzinger D. G. 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Experimental Hematology*, 30, 3: 503-512
- Fann C. H., Guirgis F., Chen G., Lao M. S., Piret J. M. 2000. Limitations to the amplification and stability of human tissue-type plasminogen activator expression by Chinese hamster ovary cells. *Biotechnology and Bioengineering*, 69, 5: 204-212
- Flintoff W. F., Livingston E., Duff C., Worton R. G. 1984. Moderate-level gene amplification in methotrexate-resistant Chinese hamster ovary cells is accompanied by chromosomal translocations at or near the site of the amplified DHFR gene. *Molecular Cell Biology*, 4, 1: 69-76
- Heller-Harrison R., Crowe K., Cooley C., Hone M., McCarthy K., Leonard M. 2009. Managing cell line instability and its impact during cell line development. New Jersey, BioPharm International: 5 str.
<http://www.biopharminternational.com/biopharm/article/articleDetail.jsp?id=601147&sk=&date=&pageID=7> (oktober 2014)
- Irizarry R.A., Bolstad B.M., Collin F., Cope L.M., Hobbs B., Speed T.P. 2003. Summaries of affymetrix GeneChip probe level data. *Nucleic Acids Research*, 31, 4: e15
- Jackson E.J. 2003. A user's guide to principal components. New Jersey, John Wiley & Sons: 569 str.
- Jostock T., Knopf H.P. 2012. Mammalian stable expression of biotherapeutics. *Methods in Molecular Biology*, 899: 227-238
- Karpinski J., Yang Y., Yeo J. H., Toh P. C., Asvaldi P., Wong N. S. 2009. A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? *Biotechnology and Bioengineering*, 102, 4: 1189-1196

- Kantardjieff A., Jacob N. M., Yee J. C., Epstein E., Kok Y. J., Philp R., Betenbaugh M., Hu W. S. 2010. Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment. *Journal of Biotechnology*, 145, 2: 143-159
- Kaufman R. J., Wasley L. C., Spiliotes A. J., Gossels S. D., Latt S. A., Larsen G. R., Kay R. M. 1985. Coamplification and coexpression of human tissue-type plasminogen activator and murine dihydrofolate reductase sequences in Chinese hamster ovary cells. *Molecular and Cellular Biology*, 5, 7: 1750-1759
- Kim N. S., Kim S. J., Lee G. M. 1998. Clonal variability within dihydrofolate reductase-mediated gene amplified Chinese hamster ovary cells: Stability in the absence of selective pressure. *Biotechnology and Bioengineering*, 60, 6: 679-688
- Kim S. J., Lee G. M. 1999. Cytogenetic analysis of chimeric antibody-producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnology and Bioengineering*, 64, 6: 741-749
- Kim J.Y., Kim Y.G., Lee G.M. 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied Microbiology and Biotechnology*, 93, 3: 917-930
- Kubista M., Andrade J. M., Bengtsson M., Forootan A., Jonak J., Lind K., Sindelka R., Sjoback R., Sjogreen B., Strombom L., Stahlberg A., Zoric N. 2006. The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, 27: 95-125
- Landauer K. 2014. Designing media for animal cell culture: CHO cells, the industrial standard. *Methods in Molecular Biology*, 1104: 89-103
- Meleady P., Henry M., Gammell P., Doolan P., Sinacore M., Melville M., Francullo L., Leonard M., Charlebois T., Clynes M. 2008. Proteomic profiling of CHO cells with enhanced rhBMP-2 productivity following co-expression of PACEsol. *Proteomics*, 8, 13: 2611-2624
- NCBI. 2007. Basic local alignment search tool. Bethesda, National Center for Biotechnology Information., U. S. National Library of Medicine.
<http://www.ncbi.nlm.nih.gov/> (oktober 2014)
- Nikolić P., Mehle N., Gruden K., Ravnikar M., Dermastia M. 2010. A panel of real-time PCR assays for specific detection of three phytoplasmas from the apple proliferation group. *Molecular and Cell Probes*, 24, 5: 303-309
- Nissom P.M., Sanny A., Kok Y.J., Hiang Y.T., Chuah S.H., Shing T.K., Lee Y.Y., Wong K.T., Hu W.S., Sim M.Y. 2006. Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. *Molecular Biotechnology* 34, 2: 125-140

- Omasa T., Onitsuka M., Kim W. D. 2010. Cell engineering and cultivation of chinese hamster ovary (CHO) cells. Current Pharmaceutical Biotechnology, 11, 3: 233-240
- Pfaffl M. W. 2001. A new mathematical model for relative quantification in real-time (RT PCR). Nucleic Acids Research, 29, 9: e45
- Pfaffl M. W., Horgan G. W., Dempfle L. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research, 30, 9: e36
- R Development Core Team. 2005. R: a language and environment for statistical computing (manual). Vienna, R Foundation for Statistical Computing: 1369 str.
- Reff M. E. 1993. High-level production of recombinant immunoglobulins in mammalian cells. Current Opinion in Biotechnology, 4, 5: 573-576
- Rolig R. L., Layher S. K., Santi B., Adair G. M., Gu F., Rainbow A. J., Nairn R. S. 1997. Survival, mutagenesis, and host cell reactivation in a Chinese hamster ovary cell ERCC1 knock-out mutant. Mutagenesis, 12, 4: 277-283
- R version 2.14.2. 2014. Vienna, R Foundation for Statistical Computing.
<http://www.r-project.org/> (september 2014)
- Schaub J., Clemens C., Schorn P., Hildebrandt T., Rust W., Mennerich D., Kaufmann H., Schulz T. W. 2010. CHO gene expression profiling in biopharmaceutical process analysis and design. Biotechnology and Bioengineering, 105, 2: 431-438
- Smyth G.K. 2005. Limma: linear models for microarray data. V: Bioinformatics and computational biology solutions using R and bioconductor. Gentleman R., Carey V., Huber W., Irizarry R., Dudoit S. (eds.). New York, Springer: 397-420
- Stahlberg A., Hakansson J., Xian X., Semb H., Kubista M. 2004. Properties of the reverse transcription reaction in mRNA quantification. Clinical Chemistry, 50, 3: 509-515
- Štrukelj B., Kos J. 2007. Biološka zdravila. Od gena do učinkovine. Ljubljana, Slovensko farmacevtsko društvo: 673 str.
- Trummer E., Ernst W., Hesse F., Schriebl K., Lattenmayer C., Kunert R., Vorauer-Uhl K., Katinger H., Muller D. 2008. Transcriptional profiling of phenotypically different Epo-Fc expressing CHO clones by cross-species microarray analysis. Biotechnology Journal, 3, 7: 924-937
- Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology, 3, 7: 624-629

Voet D. in Voet J. G., 1995. Biochemistry 2nd edition. Rose N. (ed.) New York, Chichester, Brisbane, Toronto, Singapore. John Wiley & Sons, inc: 1361 str.

Weidle U.H., Buckel P., Wienberg J. 1988. Amplified expression constructs for human tissue-type plasminogen activator in Chinese hamster ovary cells: instability in the absence of selective pressure. *Gene*, 66, 2: 193-203

Werner R. G., Kopp K., Schlueter M. 2007. Glycosylation of therapeutic proteins in different production systems. *Acta Paediatrica*, 96: 17-22

Wilhelm J., Pingoud A. 2003. Real-time polymerase chain reaction. *ChemBioChem*, 4: 1120-1128

Wiseman G. 2002. State of the art and limitations of quantitative polymerase chain reaction. *Journal of AOAC International*, 85: 792-796

Yoshikawa T., Nakanishi F., Ogura Y., Oi D., Omasa T., Katakura Y., Kishimoto M., Suga K. 2000. Amplified gene location in chromosomal DNA affected recombinant protein production and stability of amplified genes. *Biotechnology Progress*, 16, 5: 710-715

ZAHVALA

Kljub temu, da je na platnicah zapisano samo moje ime, je pri doktorski disertaciji na takšen ali drugačen način sodelovalo veliko ljudi, nekateri s svojim strokovnim znanjem in spet drugi z moralno oporo skozi celotno pot, in prav vam bi se želel zahvaliti v naslednjih vrsticah.

V prvi vrsti bi se želel zahvaliti svoji mentorici prof. dr. Kristini Gruden. Hvala za usmerjanje skozi čudovit svet transkriptomike, konstruktivne komentarje pri pisanju članka in doktorske disertacije. Andreju Franckyu in Dominiku Gaserju hvala za vso podporo ter omogočanje izvedbe celotne zgodbe od ideje do realizacije.

Zahvalil bi se rad tudi članom komisije prof. dr. Mojci Narat, prof.dr. Petru Dovču in doc. dr. Bojanu Doljaku za koristne nasvete in kritičen pregled doktorske naloge, ki je naloga dodatno oplemenitil.

Dodatno bi se rad zahvalil Vesni Ješe-Janežič za vso iskreno pomoč od začetka prijave teme do zagovora doktorske disertacije.

Hvala tudi mojim prijateljem, Vesni, Miši, Tomažu, Alešu in vsem ostalim za urice, ki smo jih preživeli skupaj in razumevanje za tiste, ki jih zaradi doktorske naloge nismo uspeli. Hvala za oporo in razumevanje tekom mnogih težav, na katere sem naletel med izdelavo naloge.

Zahvalil bi se rad še svojima staršema, ki sta me usmerjala na moji življenjski poti in me vzpodbujala kadar sem vzpodbudo najbolj potreboval.

Predvsem bi se rad zahvalil vama Ljudmila in Yogesh, ki sta me vzpodbujala in napolnjevala s pozitivno energijo, ter optimizmom.

Nenazadnje pa bi se zahvalil tebi, Petra, ki si mi stala ob strani in verjela vame od prvega do zadnjega dne izdelave naloge. Samo ti veš koliko odpovedovanja je bilo potrebnega na tej poti. Prav tako se zahvaljujem najini Flori, ki mi je pokazala, da so otroci najboljši motivatorji.

Verjetno sem še koga izpustil, ki bi ga prav tako želel navesti tukaj, vendar vem, da se čisto vsem ne bom uspel nikoli zahvaliti, zato se iskreno zahvaljujem tudi vsem Vam, ki ste mi kakorkoli pomagali, vendar Vas nisem posebej omenil.

Hvala vam.