

**Univerzita Karlova v Praze**  
**Přírodovědecká fakulta**

Katedra genetiky a mikrobiologie

Studijní program: Biologie

Studijní obor: Biologie



**Kristýna Maksimová**

Housekeeping geny

Housekeeping genes

Bakalářská práce

Školitelka: RNDr. Michaela Schierová, Ph.D.

Praha, 2016

**Prohlášení:**

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 12. 05. 2016

Podpis

Ráda bych touto cestou poděkovala RNDr. Michaele Schierové, Ph.D. za nesmírnou ochotu, trpělivost a cenné rady při psaní této práce. Dále děkuji svým rodičům za podporu během celého studia na vysoké škole a svému příteli za starost a pochopení během psaní práce.

# Abstrakt:

Housekeeping geny jsou geny exprimované ve většině tkání organismu za určitých podmínek daného prostředí. Jejich exprese je pro buňku zpravidla esenciální, částečně neměnná a může probíhat i na nízké úrovni. Rozvoj genomových a transkriptomových analýz přispěl k zpřesnění znalostí o housekeeping genech, jejich struktuře, počtu, intenzitě i stabilitě genové exprese. Housekeeping geny se využívají jako referenční geny pro kalibraci v semi-kvantitativních a kvantitativních studiích v oblasti molekulární biologie. V práci zmíním několik kandidátů, které byly mnoha studiemi určeny jako nejvhodnější referenční geny. Zaměřím se převážně na housekeeping geny u člověka a nejvíce prozkoumaných modelových organismů, *Saccharomyces cerevisiae* a *Arabidopsis thaliana*

**Klíčová slova:** Housekeeping gen, referenční gen, tkáňově specifický gen, RT-PCR, genová exprese

# Abstract:

Housekeeping genes are genes expressed in most tissues of the organism under certain conditions of the environment. Their expression in the cell is essential and relatively stable and may proceed at a low level. The development of genomic and transcriptomic analyses contributed to improving of our knowledge about housekeeping genes, about their structure, number, as well as the intensity and stability of their expression. Housekeeping genes are used as the reference genes for a calibration in semi-quantitative and quantitative studies in molecular biology. In my thesis, I will mention several candidates that have been selected as the most suitable reference genes by many studies. I will focus mainly on human housekeeping genes and the most studied model organisms like *Saccharomyces cerevisiae* or *Arabidopsis thaliana*.

**Keywords:** Housekeeping gene, reference gene, tissue specific gene, RT-PCR, gene expression

## Seznam použitých zkratek

<b>ATP</b>	adenosine triphosphate	adenotrifosfát
<b>Kbp, bp</b>	kilobase pair, base pair	kilobáze, párů bází
<b>CSC</b>	cancer stem cell	rakovinná kmenová buňka
<b>Ct</b>	cycle of treshold	prahový cyklus
<b>CV</b>	coefficient of variation	variační koeficient
<b>d<sub>SM</sub></b>	shared motif divergence	divergence sdílených motivů
<b>ENC</b>	effective number of codons	efektivní počet kodonů
<b>HKG</b>	housekeeping gene	housekeeping gen
<b>Ka</b>	rate of nonsynonymous substitutions	frekvence nesynonymních substitucí
<b>Ks</b>	rate of synonymous substitutions	Frekvence synonymních substitucí
<b>mRNA</b>	messenger RNA	mediátorová ribonukleová kyselina
<b>RT-PCR</b>	real time polymerase chain reaction	kvantitativní polymerázová řetězová reakce
<b>TBP</b>	TATA binding protein	TATA vazebný protein
<b>TSG</b>	tissue specific gene	tkáňově specifický gen
<b>UTR</b>	untranslated region	nepřekládaná oblast

# Obsah

<b>1.</b>	<b>Úvod</b>	<b>1</b>
<b>2.</b>	<b>Vývoj definice pojmu “housekeeping gen“</b>	<b>2</b>
<b>3.</b>	<b>Jak housekeeping geny vypadají?</b>	<b>3</b>
	<i>3.1. Délka exonů a intronů</i>	<i>3</i>
	<i>3.2 Tandemové repetice a GC oblasti</i>	<i>7</i>
	<i>3.3 Evoluční pohled</i>	<i>8</i>
	<i>3.3 Promotory</i>	<i>10</i>
<b>4.</b>	<b>Housekeeping geny při studiu genové exprese</b>	<b>11</b>
	<i>4.1 Výběr vhodných referenčních genů pro studie u kvasinek</i>	<i>13</i>
	<i>3.2 Výběr vhodných referenčních genů pro studie u Arabidopsis thaliana</i>	<i>19</i>
	<i>4.3 Výběr vhodných referenčních genů pro studie genové exprese u lidských nádorů</i>	<i>22</i>
	<i>4.4 Rakovinné kmenové buňky a housekeeping geny</i>	<i>22</i>
<b>5.</b>	<b>Funkce housekeeping genů</b>	<b>23</b>
<b>6.</b>	<b>Závěr</b>	<b>26</b>
<b>7.</b>	<b>Seznam použité literatury</b>	<b>27</b>

# 1. Úvod

Toto téma jsem si vybrala po přečtení článku „Do Housekeeping Genes Exist?“ od Zhanga se sp. (2015), který mě pobídl k tomu, zjistit si o těchto záhadných genech více. I přes to, že pojem „housekeeping gen“ můžeme najít v řadě učebnic, neexistuje dodnes přesná definice, na které by se vědci dokázali shodnout.

Existuje několik různých metod, které se používají při hledání housekeeping genů (dále jen HKG). Za nejúčinnější techniku se v současné době považuje RNA-seq, jelikož má nejvyšší výkon, dynamický rozsah a spolehlivou přesnost. To všechno umožňuje účinnější studii HKG. Zároveň s pomocí této techniky vznikla současná definice HKG, kterou zmíním v další kapitole.

Rozvoj metod zaměřených na analýzu genové exprese, např. microarrayové analýzy, RNA-sekvenování, Expressed Sequence Tag (EST) analýzy nebo masívně paralelního sekvenování (MPSS, Massive Parallel Signature Sequence) umožnil rozkvět databází expresních dat. Zatímco rozbořením dílčích dat lze získat odpovědi na jednoduché otázky, např. jak se projeví vliv určité chemické látky u daného typu tkáně, rozbořením velkých souborů expresních dat lze zjistit, jaký podíl genů v genomu představují HKG a jaké jsou jejich typické vlastnosti. Posun v metodických možnostech, zejména v citlivosti analytických metod i v počtu analyzovaných tkání, pletiv, vývojových stádií a experimentálních podmínek má podstatný vliv na vývoj našeho názoru o vlastnostech HKG i jejich počtu.

Cílem této práce je shrnout poznatky o vlastnostech HKG a o jejich využití jako referenčních genů v molekulární biologii. V práci se zaměřím na jejich strukturu, funkci a evoluci. Soustředím se na význam využití referenčních genů v mnoha studiích, především v RT-PCR (*real time polymerase chain reaction*, kvantitativní polymerázová řetězová reakce) zejména z hlediska volby vhodných genů. Přestože je pojem „housekeeping“ gen hojně používán, definice pojmu se stále vyvíjí. Dílčím cílem této práce je představit i některé sporné poznatky o těchto genech.

## 2. Vývoj definice pojmu “housekeeping gen“

Pojem “housekeeping gen“ se začal v literatuře objevovat před téměř 50 lety. V minulosti byly HKG definovány jednoduše jako geny, které jsou exprimované vždy a v každém buněčném typu (Watson se sp., 1965). Speiser (1974) předpokládal existenci HKG, které se musí nacházet ve všech buňkách a jsou neustále transkribovány (Hall se sp., 1999). Podle novější definice jsou HKG exprimované ve všech typech tkání a buněk a jsou potřebné pro údržbu základních buněčných funkcí (Warrington se sp., 2000). Rozdělování genů na HKG a TSG (*tissue specific gene*, tkáňově specifický gen) je základem pro studii genové exprese a buněčné diferenciaci (Zhu se sp., 2008 a).

Warrington se sp. (2000) měřili expresi 7000 genů z 11 různých lidských tkání. (srdce, mozek, plíce, ledvina, slinivka břišní, děloha a játra). Bylo použito 7 tkání dospělého člověka a 4 tkáně plodu. Autoři našli 535 genů, které byly exprimovány ve všech těchto tkáních u plodu i u dospělého člověka. Tyto geny se aktivovaly v rané fázi vývoje zárodku a byly nadále funkční v průběhu života. Pouze 47 genů se nacházelo přibližně ve stejném množství ve všech těchto tkáních. U plodu zjistili expresi 400 genů, které nebyly exprimovány u tkání z dospělých jedinců. Ve všech čtyřech tkáních plodu bylo nalezeno 767 genů, z toho 397 na stejné úrovni. U tkání z dospělých jedinců bylo zjištěno 695 genů, které se exprimovaly ve všech sedmi tkáních, z toho 241 na stejné úrovni (Warrington se sp., 2000).

Hsiao se sp. (2001) analyzovali expresi 7070 genů v 19 různých tkáních dospělých lidí. Vzorky tkání byly odebrány 24 mužům a 25 ženám ve věku kolem 55 let. Pomocí microarrayové analýzy (Affymetrix GeneChip 3.1) zjišťovali přítomnost nebo absenci příslušné mRNA. Do souboru HKG zařadili všechny geny (celkem 451 genů), jejichž mRNA byla přítomna alespoň v jednom vzorku od každé tkáně. Studie Warrington se sp. (2000) a Hsiao se sp. (2001) se ve výsledku shodovaly ve 358 HKG. Podle Hsiao se sp. (2001) HKG nejsou exprimované nutně na stejné úrovni ve všech tkáňových typech. Naopak se zdá, že každá tkáň má svou vlastní specifickou úroveň exprese

Levanon se sp. (2003) tento počet zvýšili na 575 lidských HKG. Vybrané geny otestovali pomocí veřejně dostupné databáze z microarray analýz, Affymetrix U95A chip, obsahující 12600 genů ze 47 odlišných lidských tkání (Su se sp., 2002).



I když všechny tři výše uvedené studie ve výsledku našly přibližně 500 HKG, překryvů mezi nimi bylo velmi málo. Shodovaly se pouze ve 155 genech a dvě studie se shodovaly ve 340 genech z důvodu použití stejné technologie. (Zhu se sp., 2008 a)

Podle Zhanga se sp. (2015) počet identifikovaných HKG klesá s počtem tkání, které byly zahrnuty do analýzy. U člověka bylo do roku 2015 provedeno 15 studií zaměřených na odhalení HKG, které zahrnuly 180 tkání (90 % celkového počtu). Spojením souborů HKG identifikovaných v těchto analýzách vznikl soubor 12500 HKG, avšak pouze 1 se vyskytl ve všech analýzách (peroxiredoxin 1, *PRDX1*) a 17 genů ve 14 z nich. Po odstranění dvou nejvíce vychýlených analýz, bylo 20 genů společných pro 13 zbývajících analýz. Největší vzájemná shoda byla u analýz provedených metodou sekvenování RNA.

Ve studii též popsal současnou definici HKG, která rozšiřuje tu předešlou od Warringtona se sp. (2000) a zdůrazňuje, že slabě exprimované geny, například transkripční faktory, mohou být též zahrnuty do HKG (Zhang se sp., 2015).

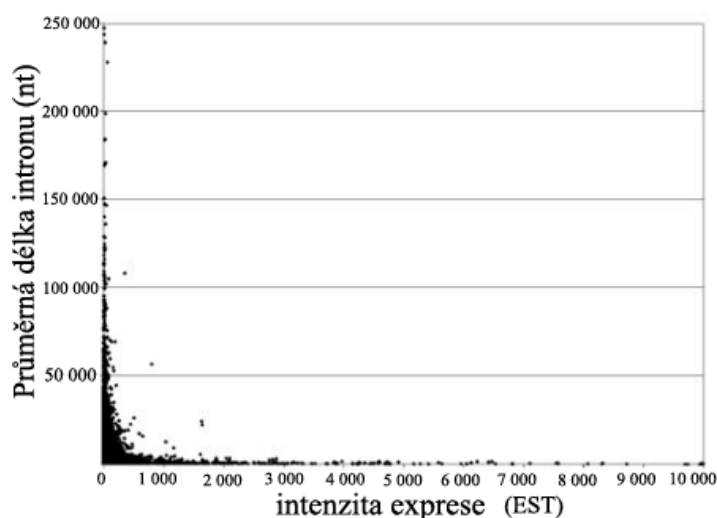
### **3. Jak housekeeping geny vypadají?**

#### **3.1. Délka exonů a intronů**

O tom, jak HKG vypadají, se dodnes vedou spory a můžeme najít mnoho protikladných názorů o jejich struktuře. Studie struktury, regulace i evoluce HKG je velmi důležitá pro poznání jejich funkce a pro jejich lepší využití při analýzách genové exprese.

Je známo, že transkripce i translace jsou pro buňku energeticky náročné, a čím více kopií mRNA buňka potřebuje, tím je to pro ni náročnější. U eukaryotických organismů je 20 nukleotidů transkribováno přibližně za 1 sec se spotřebou nejméně dvou molekul ATP na jeden nukleotid. Vyplývá z toho, že geny, které mají nižší intenzitu transkripce, mají introny zpravidla delší, zatímco geny s vyšší intenzitou transkripce mají v průměru introny krátké (Graf 1). To se týká převážně HKG. S tímto faktem souvisí ekonomická hypotéza, která předpokládá, že selekce vede u genů s intenzivní expresí k jejich zmenšování (zvyšuje se jejich kompaktnost), zatímco u genů, které se exprimují méně, selekční tlak na redukci jejich délky nepůsobí (Castillo-Davis se sp., 2002). Li se sp. (2007) však hypotézu o kompaktnosti

HKG vyvracejí. Pokud srovnávali HKG a TSG s podobnou intenzitou genové exprese, zjistili, že HKG mají delší introny než odpovídající TSG.



**Graf 1.**

*Korelace délky intronů a intenzity genové exprese u člověka. (Převzato z Castillo-Davis se sp., 2002)*

Přestože jsou HKG esenciální pro funkci buňky, je jejich genová struktura rozdílná u různých organismů. Data jsou však k dispozici jen u malého počtu modelových organismů. Při srovnání struktury lidských HKG ( $n = 532$ ) s tkáňově specifickými ( $n = 5404$ ) geny (v tabulce číslo 1 jsou označeny jako non-HK) Levanon se sp. (2003) zjistili, že nejen introny, ale i všechny ostatní části HKG jsou v průměru kratší (Tabulka 1), u intronů jsou tyto rozdíly ale mnohem znatelnější (Levanon se sp., 2003).

	HK genes ( $n = 532$ )	Non-HK ( $n = 5404$ )	P-value
Average intron length	$2573 \pm 145$ ( $n = 4353$ )	$5025 \pm 71$ ( $n = 57447$ )	$4 \times 10^{-130}$
	672	1365	
Total intron length	$21050 \pm 1781$	$53418 \pm 1425$	$7 \times 10^{-28}$
	9293	20804	
Average exon length	$212 \pm 5$ ( $n = 4885$ )	$240 \pm 2$ ( $n = 62851$ )	$9 \times 10^{-5}$
	128	132	
5' UTR length	$135 \pm 8$	$173 \pm 3$	$4 \times 10^{-7}$
	79	106	
3' UTR length	$599 \pm 30$	$846 \pm 13$	$3 \times 10^{-13}$
	333	552	
Coding sequence length	$1211 \pm 44$	$1770 \pm 26$	$3 \times 10^{-26}$
	928	1322	
Number of exons	$8.2 \pm 0.3$	$10.6 \pm 0.2$	$6 \times 10^{-7}$
	6	8	
Intron bps per coding bp	$20 \pm 2$	$31.8 \pm 0.8$	$2 \times 10^{-11}$
	9.9	15.6	

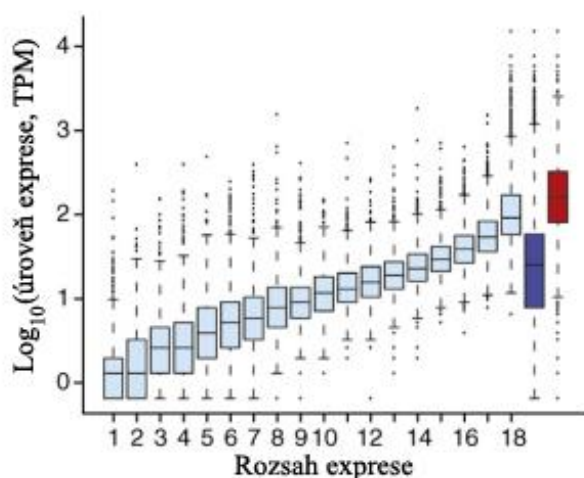
**Tabulka 1:** Porovnání struktury housekeeping genů (HK genes) a ostatních genů (Non-HK). První řádek udává průměrnou hodnotu se směrodatnou odchylkou, druhý řádek udává medián. Průměrná délka byla určena ze všech genů daného souboru. N je počet (Převzato z Levanon se sp., 2003).

V následující studii (Huang a Niu, 2008) srovnávali strukturu genů člověka a myši s využitím databází genové exprese (Su se sp., 2004) zahrnujících 69 vzorků z tkání a orgánů dospělého člověka a 55 vzorků z tkání a orgánů z dospělé myši. Analyzované geny byly rozděleny na základě úrovně exprese. Geny v horním kvantilu 30 % byly označeny jako vysoce exprimované a ve spodním kvantilu 30 % jako slabě exprimované. Jak je uvedeno v tabulce (Tabulka 3), introny a UTRs (*untranslated region*, nepřekládaná oblast) u vysoce exprimovaných genů jsou podstatně kratší, zatímco u počtu intronů nebo délky CDS není významný rozdíl.

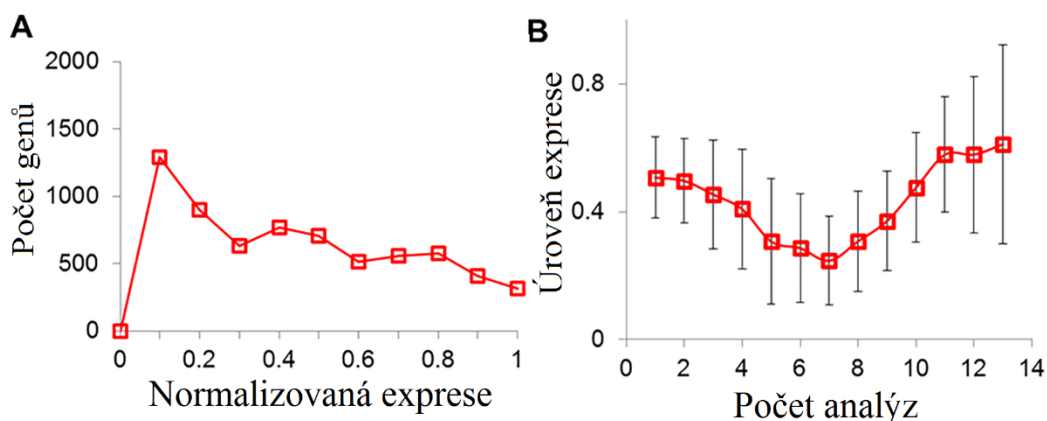
	Average intron length	Total intron length	Intron number	CDS length	UTR length	Expression level
<b>Human genes</b>						
Top 30% quantile	2768 ± 608	28117 ± 7347	8 ± 1	1313 ± 90	775 ± 107	5369 ± 770
Versus						
Bottom 30% quantile	10448 ± 4237	901046 ± 33210	9 ± 1	1764 ± 232	1478 ± 244	267 ± 14
	P = 0.001	P = 0.019	P = 0.844	P = 0.273	0.019	
<b>Mouse genes</b>						
Top 30% quantile	2631 ± 290	16190 ± 1828	7 ± 1	1214 ± 65	779 ± 136	6219 ± 794
versus						
bottom 30% quantile	8032 ± 2706	37391 ± 4615	8 ± 1	1450 ± 128	1496 ± 190	365 ± 16
	P = 0.001	P = 0.001	P = 0.444	P = 0.589	P = 0.001	

**Tabulka 2:** Srovnání genů člověka a myši v závislosti na intenzitě jejich exprese. (Převzato z Huang a Niu, 2008)

Na základě výsledků dále zmíněné studie (Zhu se sp., 2008 b), by bylo možné usoudit, že charakteristiky zjištěné pro vysoce exprimované geny lze extrapolovat pro HKG (Graf 2). Geny exprimované ve všech 18 analyzovaných typech tkání mají v průměru vyšší expresi než průměr všech exprimovaných genů. S jeho výsledky nesouhlasí studie Zhanga se sp. (2015), která ukazuje, že většina genů, označených v různých analýzách jako HKG, má nižší úroveň genové exprese. Znamená to tedy, že charakteristiky uváděné v grafech 3 pro vysoce exprimované geny platí pouze pro HKG s vyšší intenzitou exprese.



**Graf 2:** Vztah mezi úrovní a rozsahem exprese. TPM je četnost konkrétního transkriptu v milionu všech typů transkriptů detekovaných při RNA sekvencování. Rozsah exprese určuje počet tkání, u kterých byl daný gen exprimován. Světle modré bloky představují rozptyl hodnot mezi 25. a 75. percentilem. Pomocí tmavě modrého a červeného boxu je vyjádřena průměrná exprese všech genů, resp. genů identifikovaných jako HKG pomocí microarrayových analýz (Převzato z Zhu se sp., 2008; b).



**Graf 3:** Vliv intenzity genové exprese na zařazení genu do souboru HKG. Normalizovaná exprese, zobrazená v grafu A, je odrazem intenzity genové exprese. Na ose y je počet HKG s dosaženou úrovní genové exprese. Většina genů označených jako HKG má nižší intenzitu genové exprese. Graf B ukazuje, že počet analýz, ve kterých byl gen zahrnut do seznamu HKG, souvisí s intenzitou genové exprese. Intenzivně exprimované geny měly vyšší pravděpodobnost, že budou detekovány. Výsledky uvedené v grafu odrážejí stav analýzy po odstranění dvou nejméně kompletních studií (Převzato z Zhang se sp., 2015).

Zhu se sp. (2008 b) zkoumali expresi 17288 lidských genů z 18 různých tkání. Data pocházela z microarrayové analýzy a také z analýzy EST (expressed sequence tag). Zjistili, že délka genů pozitivně koreluje s intenzitou exprese a že HKG mají větší počet exonů (medián 11) než TSG (medián 4). HKG byly zpravidla delší a více exprimované než TSG (v průměru 70x). Zjistili, že některé geny nebyly na podkladě microarrayových dat dříve definovány jako HKG, což způsobilo chybný odhad délky intronů u této skupiny.

*Gallus Gallus (kur domácí)* z čeledi *Phasianidae* je důležitý modelový organismus, který vyplňuje evoluční mezeru mezi savci a ostatními strunatci. Jeho genom je významně menší než genom myši ( $2.8 \times 10^9$  bp) nebo člověka ( $3.3 \times 10^9$  bp) (Hillier se sp., 2004). V následující studii byl použit jako modelový organismus pro objasnění interakcí mezi vlastnostmi genu a expresním vzorcem při selekčním tlaku. Analýzou 10289 genů zjistili, že celková délka genu, délka CDS i délka intronů negativně korelují s úrovní a rozsahem exprese. Více než 30 % variability v kompaktnosti genů může být vysvětleno úrovní genové exprese. Porovnáním HKG a TSG s podobnou úrovní exprese zjistili, že podobně jako u člověka jsou HKG méně kompaktní, než TSG s podobnou úrovní exprese (Rao se sp., 2010), což potvrzuje výsledky studie Li se sp. (2007).

Vinogradov (2004) analyzoval expresi 7708 lidských genů ve 32 tkáních a zjišťoval vztah mezi délkou intronů (6874 intronů) nebo intergenových sekvencí (5104 IGS) a intenzitou a rozsahem exprese genu. Zjistil, že délka intronů i kódujících sekvencí je u genů exprimovaných ve většině tkání menší než u genů s vyšší tkáňovou specifitou (exprese v 1 až několika málo tkáních). Kódující sekvence u HKG byly 1,5 krát kratší než u TSG. Větší délku u TSG vysvětluje větším počtem domén i větší průměrnou délkou domény u TSG. Zajímavá zmínka se týká striktně tkáňově specifických genů, u kterých „ekonomické pravidlo“ neplatí. Vinogradov spekuluje, že nekódující oblasti, včetně intergenových, mohou obsahovat důležité regulační sekvence, které se podílejí na komplexní regulaci TSG. Pro porovnání analyzoval genom *Caenorhabditis elegans* (6250 genů) a *Drosophila melanogaster* (11250 genů). V obou případech zahrnul transkriptomy z embryonálních vývojových stádií i z dospělých jedinců. Zjistil, že u *C. elegans* jsou kódující sekvence u HKG delší, zatímco introny a především mezigenové oblasti jsou krátké. Naopak u *D. melanogaster* jsou ve většině HKG kódující sekvence, introny i oblasti mezi geny kratší, a proto je délka genu nepřímo úměrná intenzitě exprese.

Li se sp. zjistili, že u *A. thaliana* není negativní korelace mezi rozsahem genové exprese a délkou intronů. Tvrdili, že HKG s podobnou úrovní exprese nejsou u této rostliny více kompaktní než TSG (Li se sp., 2007).

### 3.2 Tandemové repetice a GC oblasti

Tandemové repetice jsou opakující se sekvence DNA dlouhé 1 – 6 bp, lokalizované na chromosomu v řadě za sebou. Vyskytují se u prokaryotických i eukaryotických organismů, v kódujících i nekódujících oblastech. Tyto repetice se využívají jako genetické markery. Lawson a Zhang (2008) se zaměřili na výskyt repetitivních sekvencí v 5'-UTR oblasti u myši a člověka. Zjistili, že u HKG je počet repetic přibližně 1,7 krát vyšší než u TSG. Tri-SSR (trinucleotide simple sequence repeats) jsou u HKG výrazně bohatší na GC páry než u TSG. Dominantním typem repetice u HKG je triplet CGG, který tvoří většinu tri-SSR. Vyskytuje se přibližně u 74 % myších HKG a u 79 % HKG u člověka, zatímco u TSG tvoří pouze 42 % u myši a 57 % u člověka. CG bohaté tri-SSR oblasti jsou lokalizovány nejčastěji v první polovině 5'UTR (70 % repetice), a proto mohou mít vliv na rozdíly v genové expresi mezi HKG a TSG (Lawson a Zhang, 2008).

Genom teplokrevných obratlovců obsahuje úseky bohatší na GC páry (v kódující i nekódující oblasti) nazývané izochory (Vinogradov, 2003). Těžké řetězce izochor (GC bohaté oblasti) představující časně replikovaný a méně kondenzovaný chromatin, obsahují HKG, zatímco lehké řetězce izochor (GC chudé oblasti) skrývají TSG (Gonçalves se sp., 2000).

Zatímco i podle Vinogradova (2003) jsou HKG člověka a myši bohatší na sekvence s GC páry než TSG, jiná studie naopak tvrdí, že množství GC párů není v korelaci s mírou tkáňové specifity (D'Onofrio, 2002).

U teplokrevných obratlovců se GC bohaté sekvence vyskytují častěji v euchromatinových oblastech, blíže ke středu jádra. GC chudé oblasti jsou naopak lokalizované v periferním, více kondenzovaném chromatinu (Vinogradov, 2004). Shlukování skupin společně exprimovaných genů, nejčastěji na operonech, je u prokaryot obvyklé. Geny kódující proteiny jedné metabolické dráhy mají tendenci zůstat propojené. U eukaryot se naopak předpokládá, že jsou geny náhodně rozmístěné, avšak některé studie to vyvracejí. Například Lercher se sp. (2002) ukazují, že geny s vysokou expresí mají tendenci se v lidském genomu shlukovat. Provedli analýzu transkriptů 5112 jednokopiových genů ze 14 tkání v lidském těle. Geny rozdělili na tkáňově specifické (exprese maximálně ve 2 tkáních), geny se střední specifitou (exprese v méně než 9 tkáních) a housekeeping geny (exprese ve více než 9 tkáních). Pro každou třídu byla vypočítána genová disperze, jako průměr rozdílu genové density v oblastech 300 kbp. Disperze HKG byla významně nižší, než se očekávala. Tento výsledek potvrdili pomocí nezávislé analýzy exprese 6298 genů z 60 tkání. HKG opět vykazovaly nízkou disperzi na rozdíl od TSG (Lercher se sp., 2002).

### 3.3 Evoluční pohled

Pro ověření, zdali jsou HKG v průměru starší než TSG, byla provedena srovnávací analýza ortologních genů z pěti organismů: *H. sapiens sapiens*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans* a *Saccharomyces cerevisiae*. Bylo klasifikováno 15501 (89,7 % z celkového počtu) lidských genů do pěti hlavních kategorií v závislosti na jejich fylogenetické distribuci. Podle míry podobnosti s ortology z výše zmíněných druhů rozdělili 15501 lidských genů (89,7 % z celkového počtu) do pěti hlavních kategorií: A-Geny specifické pro savce, B-geny přítomné jen u obratlovců, C-geny přítomné u obratlovců

i členovců, D-geny specifické jen pro Metazoa, E-geny přítomné u všech eukaryot. Toto členění odpovídá přibližně i stáří genů, od mladších ke starším. 54% HKG genů bylo přítomno ve skupinách C, D a E, kde naopak bylo jen 5,4 % TSG. Většina TSG (70,8 %) byla ve skupině A, která obsahovala jen 11,6 % ze všech HKG. Výsledky analýzy tedy původní hypotézu potvrdily. 54.5% HKG a 5.4% TSG vzniklo před oddělením linie obratlovců, zatímco pro savce je unikátních 11.6 % HKG a 70.8 % TSG. HKG jsou ve srovnání s TSGs v průměru starší (Zhu, se sp., 2008 b).

Z několika evolučních studií vyplynulo, že intenzivně exprimované geny mají tendenci podléhat evoluci pomaleji než méně exprimované geny. (Duret a Mouchiroud 2000; Hastings 1996; Hughes a Hughes 1995).

Studie (Zhang a Li, 2004) prováděná u myši a člověka ukázala, že mezi HKG a TSG jsou významné rozdíly v poměru  $Ka/Ks$  ( $Ka$  je rychlost nesynonymních substitucí, zatímco  $Ks$  je rychlost synonymních substitucí, resp. počet substitucí na dané pozici nukleotidové sekvence za  $10^9$  let). Potřebná data byla stažena z genové databáze HuGE Index a tvořila 1980 genů, z čehož 451 HKG a 1529 TSG exprimovaných ve svalech, mozku, prostatě, plicích, pochvě, játrech a ledvinách. Zjistili, že kódující sekvence HKG mají v průměru menší počet nesynonymních mutací ( $Ka = 0.046$ ) než TSG ( $Ka = 0.083$ ). Rozdíly v rychlosti synonymních substitucí jsou méně výrazné (HKG je  $Ks = 0.447$ , zatímco u TSG je  $Ks = 0.492$ ). Shrnutí výsledků analýzy Zhang a Li (2004) je uvedeno v tabulce 3. Z tabulky je zřejmé, že u genů exprimovaných v játrech nebo v plicích je frekvence nesynonymních substitucí větší než u genů exprimovaných v ostatních tkáních, lze předpokládat, že substituce v genech druhé skupiny TSG jsou striktněji odstraňovány jako selekčně nevýhodné.

Tyto výsledky se shodují s výsledky studie Hastings (1996), který porovnával sekvenční variabilitu mezi alelami 15 genových rodin u obratlovců. Potvrdil, že alely s širší expresí podléhají evoluci pomaleji, než jejich tkáňově specifické protějšky.

Tissue	Ka	Ks	Ka/Ks	ENC <sup>a</sup>
Brain	0.059 (0.062)	0.465 (0.136)	0.121 (0.125)	48.7 (6.5)
Kidney	0.089 (0.071)	0.517 (0.125)	0.166 (0.116)	48.3 (5.9)
Lung	0.150 (0.120)	0.540 (0.139)	0.259 (0.172)	47.9 (6.0)
Liver	0.131 (0.082)	0.546 (0.127)	0.233 (0.132)	50.6 (5.3)
Muscle	0.055 (0.059)	0.458 (0.150)	0.116 (0.112)	48.7 (6.3)
Prostate	0.066 (0.095)	0.540 (0.181)	0.108 (0.125)	48.6 (8.2)
Vulva	0.109 (0.085)	0.519 (0.152)	0.201 (0.139)	45.6 (7.5)
House-keeping	0.046 (0.063)	0.447 (0.146)	0.093 (0.114)	49.1 (6.6)

**Tabulka 3:** Srovnání substituční rychlosti u HKG a TSG. Efektivní počet kodonů (ENC, effective number of codons) používaný u genů dané skupiny. Každému genu se přiřadí hodnota ENC mezi 20 a 61: hodnota 20 znamená, že každá aminokyselina je kódována pouze jedním kodónem a hodnota 61 znamená, že se používají všechny kodony. Ka je rychlost nesynonymních substitucí v dané pozici genomu za 10<sup>9</sup> let, Ks je frekvence synonymních substitucí (Převzato z Zhang a Li, 2004).

### 3.3 Promotory

Studie Farré se sp. (2007) je první prací, která se důsledně věnuje mezidruhovému variabilitě promotorových sekvencí u HKG a TSG u savců. Na rozdíl od kódující sekvence, promotory nemají přesně definované hranice, avšak oblast o velikosti přibližně 100 bp upstream od počátečního místa transkripce, nazývaná basální promotor, má významnou roli při skládání transkripčního iniciačního komplexu (Farré se sp., 2007).

Podle starší studie se u oblastí ve vzdálenosti delší než 2 kbp od transkripčního počátku podobnost mezi promotory ortologních genů drasticky snižuje, což by potvrdilo, že většina funkčních elementů se vyskytuje v prvních 2 kbp. Jako příklad uvádějí, že 85 % myších transkripčních regulačních motivů se vyskytuje v blízkosti do 2 kbp (<2 kbp) od transkripčního počátku proti směru transkripce, 10 % v intronech a pouze 5 % ve větší vzdálenosti (>2 kbp) od transkripčního počátku (Waterston se sp., 2002).

Farré se sp. (2007) shromáždili DNA sekvence promotorů lidských a myších ortologních genů (6698 párů, 2 kbp od počátku transkripce) a zaměřili se na vyčíslení odlišnosti promotorových sekvencí, zhodnocení účinnosti selekce a určení významných vztahů mezi sekvencí variabilitou promotoru a kódujících sekvencí. Vzájemným porovnáním promotorů ortologních genů (pairwise sequence alignment) (Castillo-Davis se sp., 2004) určili hodnotu divergence promotorové sekvence  $d_{SM}$  (*shared motif divergence*, divergence sdílených motivů). Hodnota  $d_{SM}$  vyjadřuje podíl promotorové sekvence, která u



ortologních genů nelze přiložit. Další charakteristikou je hodnota  $K_P$ , která charakterizuje rychlost substitučních mutací v oblasti promotoru.

Divergence promotorových sekvencí  $d_{SM}$  ani hodnota  $K_P$  v promotorové oblasti v rozsahu 2000 bp od počátku transkripce (poloha -2000, +1) byla u HKG neprůkazně zvýšená, na rozdíl od výrazného rozdílu poměru  $Ka/Ks$  u kódujících sekvencí (viz předchozí kapitola). Avšak, vezme-li se v úvahu oblast promotoru (-2000, -500), kde je míra divergence větší, jsou rozdíly v hodnotě  $d_{SM}$  mezi HKG a TSG výraznější. Trend vyšší sekvenční diverzity promotorů u HKG se uplatňuje bez ohledu na anotovanou funkci těchto genů (Farré se sp., 2007).

Promotory HKG jsou bohaté na CpG ostrovy (Gardiner-Garden a Frommer, 1987). To může mít vliv na úroveň konzervativnosti promotorových sekvencí. Z tohoto důvodu rozdělili soubor analyzovaných genů podle přítomnosti ostrovů v oblasti -100 až 100 bp a analyzovali tyto dvě skupiny odděleně. U myši, 65 % genů bylo klasifikováno jako CpG+, v rámci HKG je podíl CpG+ 88 %. Délka CpG ostrovů nebyla významně odlišná u HKG a TSG. Autoři zjistili, že v souboru CpG+ genů jsou větší rozdíly v hodnotě  $d_{SM}$  mezi HKG a TSG než v nesetříděném souboru nebo v souboru genů CpG-. Nejvyšší variabilita promotorů byla u HKG CpG+, s hodnotou  $d_{SM} = 0,739$ , zatímco u TSG CpG+ byla hodnota  $d_{SM} = 0,679$ .

Tato studie ukázala, že promotory, které neobsahují CpG ostrovy, jsou více konzervované (Farré se sp., 2007).

## 4. Housekeeping geny při studiu genové exprese

V současnosti slouží HKG při kalibraci v mnoha teoretických i aplikovaných studiích, včetně klinické praxe. Jsou to tzv. referenční geny – geny, které se exprimují stabilně za všech podmínek experimentu. HKG jsou všeobecně využívány v semi- kvantitativních a kvantitativních RNA analýzách, Northern blot analýzách, RT-PCR a DNA microarrayové analýze. Správná volba referenčního genu je velmi důležitá v kvantitativní RNA analýze. Vlastnímu experimentu by měla předcházet pilotní studie, která při stejných experimentálních podmínkách otestuje stabilitu exprese několika možných referenčních genů (Hellemans se sp., 2007).

Metodu real-time PCR (RT-PCR) objevil v roce 1992 Higuchi se sp. (1992). Dnes je považována za nejpřesnější a nejspolehlivější způsob ověření údajů získaných jinými metodami, např. microarrayovou analýzou. Navíc, úroveň exprese některých genů je často natolik malá, že RT-PCR je jedinou možnou technikou, která je schopna tuto RNA detekovat. Výhodami RT-PCR jsou bezpochyby velká citlivost, detekce dynamického vývoje hladiny RNA, rychlost analýzy a přesnost. Pro RT-PCR je potřeba vybrat referenční gen s podobným  $C_t$  (*cycle of threshold*, prahový cyklus) jako u cílového genu. S pomocí referenčního genu lze určit míru variability výsledků plynoucí z chyb v technologii a při přípravě vzorků. K tomu se používají převážně HKG. Zároveň se doporučuje použít minimálně dva HKG, jelikož využití pouze jednoho často způsobuje chyby při vyhodnocení výsledků měření. Při větším počtu HKG jsou výsledky přesnější (Kozera a Rapacz, 2013).

Existuje několik algoritmů, které analýzou výsledků expresní analýzy vyhodnotí ze seznamu použitých referenčních genů ty nejvhodnější (s minimálními změnami hladiny RNA). Pomocí algoritmu GeNorm lze pro každý referenční gen určit hodnotu jeho stability  $M$  a rozptyl jeho  $C_t$  hodnot  $V$ , resp. jejich variační koeficient  $CV$  (*coefficient of variation*) za všech experimentálních podmínek. Optimální hodnoty  $M$  a  $V$  jsou 0,5–1,0 pro  $M$  a 0,25–0,5 pro  $CV$ . S rostoucím počtem analyzovaných podmínek a genů, hodnoty rostou (Vandesompele se sp., 2002). Pro analýzu většího počtu referenčních genů byl na bázi GeNorm vyvinut software qBase. Tento program se skládá ze dvou modulů. Prvním je "qBase browser" pro správu a archivaci dat a druhým "qBase Analyzer" pro zpracování prvotních dat do biologicky smysluplných výsledků (Hellemans se sp., 2007).

Tabulkový procesor v Excelu, BestKeeper, byl také vyvinut pro nalezení nejstabilnějších HKG ze seznamu. Z deseti kandidátních genů dokáže určit ty nejvhodnější a zařadit je do indexu. Tento index může být dále srovnáván s dalšími deseti geny, aby se rozhodlo, zdali je dostatečně stabilní (Pfaffl se sp., 2004). Podobný princip používá i NormFinder (Andersen se sp., 2004).

V únoru 2009 byly sepsány základní pokyny pro experimenty s qPCR. Zahrnují detailní seznam postupů, které by měly být dodržovány. Týkají se zejména přípravy vzorků, manipulace s nimi, syntézy cDNA a analýzy dat. Hlavním doporučením v těchto pokynech bylo zvolit si více než jeden referenční gen. Je zároveň nezbytné ověřit, zda vybrané geny vykazují stabilní expresi v cílových tkáních (Bustin se sp., 2009).

Výběrem referenčních genů a reálnou stabilitou jejich genové exprese se zabývala studie Chapmana a Waldenströma (2015). Z výsledků vyplynulo, že průměrný počet použitých HKG ve studiích je 1,23 (počet studií = 128) a tedy pouze 13 % (17/128) studií použilo více jak jeden HKG pro normalizaci. Z analýzy nebylo evidentní, že by se počet použitých HKG s časem zvyšoval. Jenom 19 studií dokázalo, že panel referenčních genů byl testován, aby byly nalezeny ty, které mají nejstabilnější expresi. Studie, které HKG testovaly, nakonec použily více HKG ve svých výzkumech než ty které test neprováděly. Nejčastěji se jako HKG používaly ACTB (38 %), GAPDH (37 %) a méně často používaný byl gen pro 18S rRNA (12 %). Tyto tři geny bývají zvoleny s mnohem menší pravděpodobností, pokud je testováno více referenčních genů. Zároveň bylo zjištěno, že mnohem více HKG bylo testováno na stabilitu exprese u modelových organismů než u nemodelových. Většina výzkumů byla provedena na člověku (38 %), na hlodavcích (22 %) a hospodářských zvířatech (19 %). Průměrný počet HKG testovaných z hlediska stability jejich exprese byl 9,53.

Potvrdilo se, že myšlenka univerzálního referenčního genu je nesmysl. I přes to, že ve více než 72 % studií (93 / 128) byly GAPDH, ACTB a 18S rRNA použity jako jediný referenční gen, nejsou tyto geny vhodné pro všechny experimentální podmínky a typy experimentů. (Chapman and Waldenström, 2015) Navíc, využití genů GAPDH a ACTB je zatíženo i výskytem pseudogenů (zjištěno u člověka a myši). Pseudogeny představují další potenciální zdroj chyb během qPCR studií, vedoucích k nepřesným výsledkům (Sun se sp., 2012).

#### **4.1 Výběr vhodných referenčních genů pro studie u kvasinek**

Cílem studie Teste se sp. (2009) bylo vybrat vhodnou sadu referenčních genů pro *Saccharomyces cerevisiae*, použitelnou pro analýzu genové exprese při dlouhodobé kultivaci. Navázali tak na podobné studie provedené u *Candida albicans* a *Aspergillus nidulans*. Z databáze microarrayových analýz (DeRisi se sp., 1997) a (Gasch se sp., 2000) vybrali 8 genů s nejstabilnější expresí při kultivaci na glukóze. Tento soubor doplnili o 5 genů běžně používaných při hodnocení genové exprese u *S. cerevisiae* (*ACT1*, *PDA1*, *RDN18*, *IPP1* a *TDH3*). Funkce genů zmiňovaných v této studii i studii následující je uvedena v tabulce 4. Sledovali expresi vybraných 13 genů při dlouhodobé kultivaci kmene

s delecí *TPS1* (defekt v syntéze trehalózy) v galaktózovém médiu a také dvou nemutovaných kmenů při kultivaci v glukózovém nebo galaktózovém médiu. Hladinu vybraných RNA měřili v klíčových fázích vývoje kvasinkové kultury, od počátku exponenciální fáze až do dosažení stacionární fáze růstu (exponenciální fáze (respiro-fermentační fáze), diauxický shift, postdiauxický shift (čistá respirace), stacionární fáze -hladovění). U vybraných referenčních genů sledovali hodnotu  $C_t$  a její rozptyl mezi kultivacemi i v průběhu jedné kultivace (Teste se sp., 2009).

Stabilita exprese genů byla rovněž hodnocena pomocí *geNORM* (Vandesompele se sp., 2002). Z analýzy vyplynulo, že geny *ALG9*, *TAF10*, *TFC1*, *UBC6* a hlavně *KRE11* patří mezi nejstabilnější geny v testované kultuře, zatímco *ACT1*, *IPPI* a *TDH3* patří mezi nejméně vhodné kandidáty na referenční geny. Vybrané geny vykazovaly  $C_t$  hodnoty v rozpětí 20 – 30 cyklů, zatímco  $C_t$  hodnoty *RDN18* a *TDH3* byly podstatně nižší (*RDN18* 8 a *TDH3* 17 cyklů). Stabilita exprese vybraných referenčních genů byla dále ověřena rozborem dat z dostupných microarrayových analýz. Autoři zjistili, že geny *ALG9*, *TAF10*, *TFC1*, *UBC6* vykazovaly mnohem častěji hodnoty  $\log_2$  poměru exprese za dvou různých podmínek blízké nule než gen *ACT1*, tj. jejich exprese byla vyrovnanější. Tím potvrdili, že vybrané geny jsou kvalitní referenční geny. Dále úspěšně využili sestavu 3 nejstabilnějších genů *UBC6*, *TAF10* a *ALG9* při kvantifikaci exprese genů metabolismu glykogenu *GSY2* (Glycogen Synthase), *GPH1* (glycogen phosphorylase) a *SGA1* (Sporulation-specific GlycoAmylase) v závislosti na fázi kultivace (Teste se sp., 2009).

Zajímavý přístup k identifikaci vhodné sestavy referenčních genů zvolili Cankorur-Cetinkaya se sp. (2012). Analyzovali data 31 microarrayových analýz, zahrnujících vliv doby kultivace na genovou expresi. Nejdříve pro každý experiment vybrali 100 nejstabilnějších genů. Je překvapivé, že ani jediný gen se neobjevil mezi 100 geny u všech analyzovaných souborů. *TPI1* a *ACT1* byly ve stovce nejstabilnějších v 81% případů. Autoři vybrali 10 genů, které se vyskytly jako nejstabilnější ve více než 55% případů (*TPI1*, *ACT1*, *TDH3*, *FBA1*, *CCW12*, *CDC19*, *ADH1*, *PGK1*, *GCN4* a *PDC1*). Pouze 2 geny z této skupiny (*ACT1* a *TDH3*) patřily mezi běžně používané referenční geny.

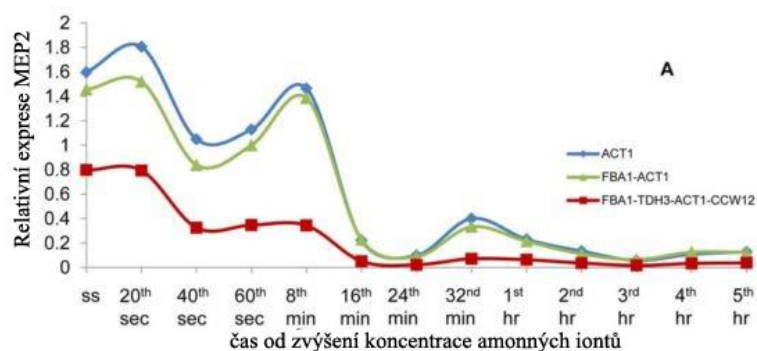
Jako alternativní přístup při identifikaci referenčních genů byla hodnocena stabilita exprese genů v náhodných souborech, sestavených kombinací arrayových dat použitých v

předchozí analýze. Rozsah tohoto výzkumu byl obrovský: zahrnoval analýzu 5423 transkriptů v 888 různých časových bodech (resp. kombinace času a konkrétních podmínek experimentu). Soubor 100 nejstabilnějších genů z této analýzy měl  $C_v$  hodnotu (poměr směrodatné odchylky a průměru expresních hodnot) v intervalu 0,05 až 0,11. Zpřísněním podmínek pro  $C_v < 0,07$  bylo ze souboru vybráno 5 genů: *TDH3*, *RPS26A*, *TPII*, *CDC19* a *ARF1*. Srovnání obou top-referenčních souborů je opět zarážející: překryv mezi nimi představují jen geny *TDH3*, *CDC19* a *TPII*. Podle autorů studie lze za kvalitní referenční geny považovat všech 12 genů, vybraných jedním z obou postupů.

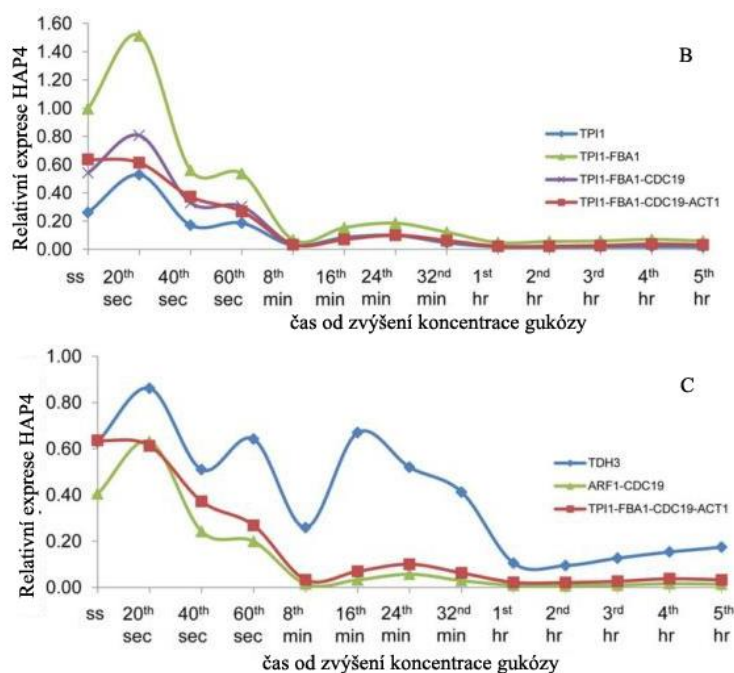
Cílem třetí analýzy bylo zjistit, které referenční geny jsou nejvhodnější pro dané podmínky experimentu. Stabilita genové exprese všech 12 genů byla hodnocena ve dvou (reálných) experimentech pomocí qPCR. Ze souboru byl vyřazen gen *PGKI* kvůli nízké účinnosti qPCR (Cankorur-Cetinkaya se sp., 2012).

Následně, v prvním experimentu byla sledována odpověď na náhlé zvýšení koncentrace amonných iontů ke kultuře s limitující koncentrací zdroje dusíku. Ve druhém experimentu byla ke kultuře s nedostatečným zdrojem uhlíku náhle přidána glukóza. V obou případech byla známa očekávaná odpověď. Reakcí na doplnění amonných iontů je snížená exprese genu *MEP2*, kódujícího vysokoafinitní přenašeč pro amonné ionty. Podobně, jako odpověď na přítomnost glukózy se sníží exprese genu *HAP4*, který kóduje transkripční faktor aktivující respirační geny. Expresní hodnoty kandidátních referenčních genů byly po dvojicích hodnoceny korelačním koeficientem. Přítomnost dvou genů s korelovanou genovou expresí v souboru referenčních genů může zkreslit data o genové expresi testovaných genů (Vandesompele se sp., 2002). Ukázalo se, že soubor kandidátních referenčních genů geny s korelovanou expresí obsahuje. V prvním pokusu korelovala exprese *TDH3* a *PDC1*, ve druhém pokusu korelujícími geny bylo víc: exprese *ARF1* korelovala s geny *TPII* a *CDC19* a exprese *FBA1* korelovala s *CCW12*. Z toho je zřejmé, že i korelace genové exprese závisí na konkrétních podmínkách experimentu.

V dalším kroku byl optimalizován počet referenčních genů, které nejlépe přispějí k dosažení očekávané expresní odpovědi u genů *MEP2* a *HAP4*. Výsledky nejlépe vystihují grafy 4 a 5. Pro první experiment byly jako optimální sestava vybrány geny *FBA1*, *TDH3*, *CCW12* a *ACT1*, ve druhém experimentu geny *TPH1*, *FBA1*, *CDC19* a *ACT1* (Cankorur-Cetinkaya se sp., 2012).



**Graf 5:** Vývoj exprese *MEP2* v reakci na zvýšení koncentrace amonných iontů; osa x: doba kultivace po přidání amonných iontů, osa y: relativní exprese *MEP2*. Z grafu je zřejmé, že tradičně používaný referenční gen *ACT1* nelze použít pro normalizaci expresních dat za daných podmínek experimentu (Převzato z Cetinkaya se sp., 2012).



**Graf 6:** Vývoj exprese HAP4 v reakci na zvýšení koncentrace glukózy; osa x: doba kultivace po přidání glukózy, osa y: relativní exprese HAP4. Z horního grafu je zřejmý pozitivní vliv zvyšujícího se počtu referenčních genů: očekávaná odpověď exprese HAP4 byla nalezena po normalizaci dat pomocí kombinace 4 genů: TPI1-FBA1-CDC19 a ACT1. Graf C ukazuje srovnání normalizované exprese HAP4 s jinými variantami referenčních genů. THD3 je běžně používaný referenční gen, dvojice ARF1-CDC19 měla nejstabilnější expresi v souboru vybraných referenčních genů, avšak tyto dva geny mají koncentrovanou expresi (Převzato z Cetinkaya se sp., 2012).

Z toho vyplývá, že volba vhodných referenčních genů i jejich počtu představuje kritický bod expresních analýz. Zároveň je jasné, že i po předchozím, velmi pečlivém výběru optimálních referenčních genů je nezbytné ověřit, zda je exprese těchto genů stabilní i za daných podmínek experimentu.

Název	Molekulární funkce	Biologická funkce
ARF1	GTPase activity	Transport, macroautophagy
IPP1	Inorganic diphosphatase activity	cell growth
PDA1	Pyruvate dehydrogenase activity	Pyruvate metabolism
RDN18	Structural constituent of ribosome	Translation
RPS26A	Structural constituent of ribosome	Cytoplasmic translation, rRNA export from nucleus
<b>ALG9</b>	Mannosyltransferase activity	Protein amino acid glycosylation
<b>FPR2</b>	Membrane-bound peptidyl-prolyl cis-trans isomerase activity	Unknown
<b>HEM2</b>	Porphobilinogen synthase activity	Heme biosynthesis
<b>KRE11</b>	Unknown	ER to Golgi vesicle-mediated transport
<b>RPN2</b>	Protein binding, bridging	Ubiquitin-dependent protein catabolic process
<b>TAF10</b>	RNA Pol II transcription factor activity	Transcription initiation and chromatin modification
<b>TFC1</b>	RNA Pol III transcription factor activity	Transcription initiation on Pol III promoter
<b>UBC6</b>	Ubiquitin-protein ligase activity	ER-associated protein catabolic process
ACT1 <sub>c</sub>	Actin, Structural constituent of cytoskeleton	cell polarity, cell division
ADH1 <sub>c</sub>	Alcohol dehydrogenase (NAD) activity	Ethanol biosynthetic process involved in glucose fermentation to ethanol
CCW12 <sub>c</sub>	Structural constituent of cell wall	Agglutination involved in conjugation with cellular fusion
CDC19 <sub>c</sub>	Pyruvate kinase activity	Glycolytic process, pyruvate metabolic process
FBA1 <sub>c</sub>	Fructose-bisphosphate aldolase activity	Gluconeogenesis, glycolytic process
GCN4 <sub>c</sub>	Chromatin binding, sequence-specific DNA binding	Negative regulation of transcription from RNA polymerase II promoter
PDC1 <sub>c</sub>	Pyruvate decarboxylase activity	Pyruvate metabolic process, tryptophan catabolic process
PGK1 <sub>c</sub>	Phosphoglycerate kinase activity	Gluconeogenesis, glycolytic process
TDH3 <sub>c</sub>	Glyceraldehyde-3-phosphate dehydrogenase (NAD <sup>+</sup> ) activity, RNA binding	Glycolysis & Gluconeogenesis
TPI1 <sub>c</sub>	Triose-phosphate isomerase activity	Glycolytic process

**Tabulka 4:** Funkce kandidátních referenčních genů. Údaje byly získány z databáze genomu *S. cerevisiae* (SGD). Geny zvýrazněné tučným písmem byly vybrány analýzou Teste se sp. (2009). Geny vybrané studií Cetinkaya se sp. (2012) mají index c za názvem genu. Ostatní geny jsou běžně používané jako kontrolní geny pro studie u kvasinek.



Huisinga a Pugh (2004) ukázali, že genom kvasinek může být regulován dvěma způsoby. Expres malé části genů (10 %) je ovlivněna různými stresory a je vysoce regulována chromatinem, TBP (*TATA binding protein*, TATA vazebný protein) a RNA polymeráza II regulátory. Druhý typ genů lze zařadit do HKG (90 %). Tyto geny jsou celkově méně regulované, jejich exprese je stabilní a jejich produkty mají podobnou funkci jako u ostatních organismů (syntéza proteinů, růst buňky a základní funkce metabolismu).

### 3.2 Výběr vhodných referenčních genů pro studie u *Arabidopsis thaliana*

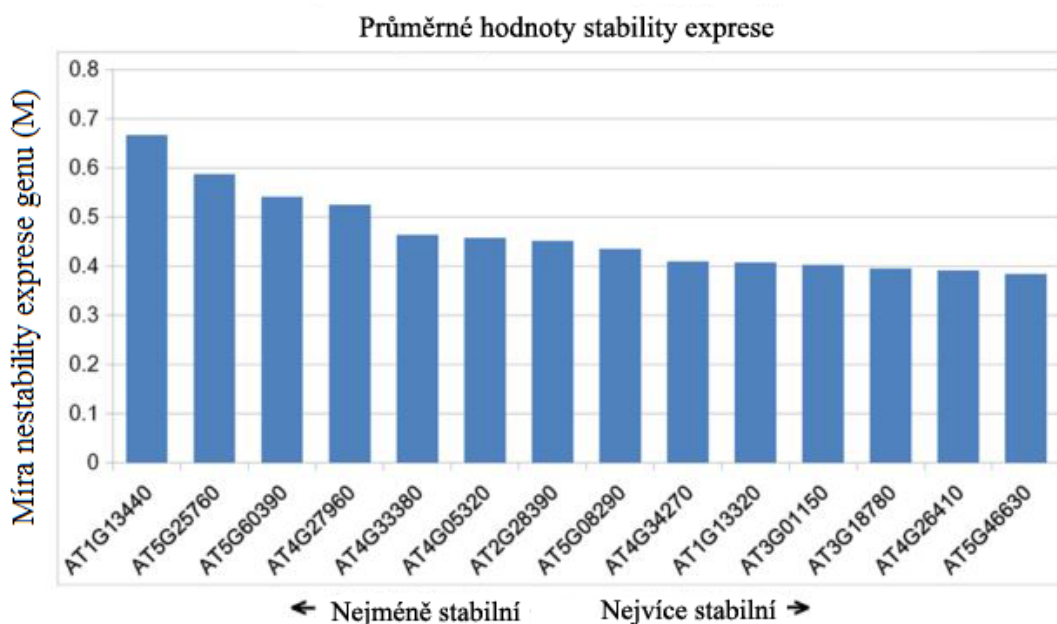
Huseniček (*Arabidopsis thaliana*) je oblíbená modelová rostlina z čeledi brukvovitých. Předpoklady pro určení nejvhodnějších referenčních genů k normalizaci byly splněny v podobě databáze získané pomocí Affymetrix ATH1 GeneChip s 23500 geny (Redman se sp., 2004). Shrnuje data o úrovni transkripce genů za různých vývojových a environmentálních podmínek, u 79 pletiv, orgánů, různých ontogenetických stádií i různých genotypů, při působení různých abiotických stresorů (chlad, osmotický tlak, salinita, oxidativní stres a teplotní výkyvy). Analýzy zahrnují i působení biotických stresorů, především fytopatogenních bakterií a hub.

Czechowski se sp. (2005) srovnáním expresních dat z databází, doplněných o vlastní analýzy vybrali 18 nových genů vhodných pro normalizaci expresních dat u *A. thaliana*. Expresi sledovali v závislosti na vývojovém stádiu či orgánu, abiotických stresech, dostupnosti živin, osvětlení atd. Zjistili, že rozdíly v expresi genů způsobují především různá vývojová stadia, průměru se nejvíce vymykají pylová zrna a semena. Pro různé podmínky experimentu vybrali sestavu 100 genů s nejmenšími výkyvy v genové expresi. Ve studii dále zkoumali rozdíly v normalizované genové expresi při 101 různých podmínkách experimentu u genů, které se tradičně používají, a genů, které vybral Czechowski se sp. Expresi vybraných genů byla 10 – 20 % ve srovnání s tradičně používanými geny, jak ukazovaly velké rozdíly v hodnotách Ct (8-12 u tradičních genů, 12-24 u nově vybraných genů). U genů vybraných na základě hodnocení expresních hodnot v databázích bylo provedeno jejich ověření u 20 náhodných vzorků cDNA. Příkladem z 18 nově vybraných genů jsou geny *TIP41*-like family protein, *PP2AA3*, *PP2A-1*, *At4g33380* nebo *MON1* (Tabulka 5, Czechowski se sp., 2005).

Zkratka genu	Celý název	Molekulární funkce	Biologická funkce
TIP41-like family protein	x	x	x
PP2AA3	PROTEIN PHOSPHATASE 2A SUBUNIT A3	protein phosphatase type 2A regulator activity	regulation of phosphorylation
PP2A-1	PROTEIN PHOSPHATASE 2A-1	metal ion binding, protein serine/threonine phosphatase activity	protein dephosphorylation
At4g33380	Neznámý protein	x	x
MON1	MONENSIN SENSITIVITY1	x	vesicle-mediated transport

**Tabulka 5:** Přehled referenčních genů u *A.thaliana* vybraných ve studii Czechowski se sp. (2005), data byla stažena z databáze [www.arabidopsis.org](http://www.arabidopsis.org).

Wang se sp. (2014) porovnali funkci stejných HKG u diploidních a tetraploidních rostlin *Arabidopsis thaliana*. K analýze využili data stažená z Gene Expression Omnibus (GEO) a sadu 5 tradičně používaných a 16 nově vybraných genů, dříve určených pomocí microarrayových dat u diploidní *A. thaliana*. Stabilita těchto 21 genů byla posouzena pomocí programů GeNorm a NormFinder. Ačkoli se tyto referenční geny projevovaly jako vhodné u diploidní *A. thaliana*, v případě tetraploidních rostlin tomu tak nebylo: 7 z 21 genů (*AT1G62930*, *AT5G55840*, *AT4G38070*, *AT5G15710*, *AT3G53090*, *AT2G32170* a *AT5G1224*) (Tabulka 6) s nízkou hodnotou RPKM (počet fragmentů (readů) specifického genu na kilobázi délky jeho transkriptu na milión mapovaných fragmentů (readů)) u diploidní *A. thaliana* nebylo detekováno u všech tetraploidních rostlin, takže tyto geny nebyly pro tetraploidní rostliny vhodné pro normalizaci a byly vyřazeny. Pomocí statistické analýzy u zbývajících 14 genů zjistili, že 6 genů vykazovalo výrazný rozdíl v intenzitě transkripce mezi diploidními a tetraploidními rostlinami. Nejstabilnějším genem byl u tetraploidů gen *AT5G46630*. Výsledky analýzy dokumentuje graf 6. Ze studie vyplývá, že je nutné vytvořit samostatnou sadu HKG pro tetraploidní rostliny *A. thaliana*.



**Graf 6:** Hodnota  $M$  pro gen  $x$  je mírou nestability exprese genu  $x$ . Čím nižší je tato hodnota, tím stabilnější je jeho exprese v porovnání s ostatními geny v testovaném souboru 14 vybraných HKG u tetraploidních rostlin *A. thaliana*. (Převzato z Wang se sp., 2014).

Označení	Zkatka	Celý název	Funkce
AT1G62930	<i>RPF3</i>	RNA PROCESSING FACTOR 3	mRNA processing
AT5G55840	x	Pentatricopeptide repeat-containing protein At5g55840	x
AT4G38070	x	x	DNA binding, transcription factor activity
AT3G53090	<i>UPL7</i>	UBIQUITIN-PROTEIN LIGASE 7	ligase activity, ubiquitin-protein transferase activity, cellular protein modification process,
AT2G32170	<i>EXPRS</i>	x	methylation, methyltransferase activity
AT5G15710	x	x	x

**Tabulka 6:** Přehled referenčních genů u *A. thaliana*, které musely být ze seznamu vyřazeny, jelikož nebyly detekovány u všech tetraploidních rostlin i přes to, že u diploidní *A. thaliana* byly označeny jako vhodné referenční geny (Wang se sp., 2014). Data byla stažena z databáze [www.arabidopsis.org](http://www.arabidopsis.org).

### 4.3 Výběr vhodných referenčních genů pro studie genové exprese u lidských nádorů

Studie genové exprese jsou velmi důležité např. při stanovení diagnózy u nádorových onemocnění. I v případě těchto specificky zaměřených expresních analýz se používaly stále stejné housekeeping geny. Protože není zaveden standardní postup a sestava referenčních genů, je obtížné až nemožné získaná data porovnávat. Špatná volba referenčního genu může vést k chybnému porovnání genové exprese, jak tomu bylo například u výše zmíněné studie se *S. cerevisiae*. Studie Sharan se sp. (2015) se zabývala nalezením vhodných genů rozborem monografií zabývajících se genovou expresí u třinácti nejrozšířenějších nádorů u člověka. Autoři zjistili, že ve třinácti studiích odlišných typů lidské rakoviny bylo použito 50 různých HKG. Po rozboru dostupných analýz doporučili kombinovat gen *PPIA* (peptidylprolyl isomerase A) s 2–3 geny ze skupiny *GADPH*, *ACTB*, *HPRT*, *TBP* (Sharan se sp., 2015). Příkladem použitých studií je studie McNeill se sp., kteří hledali referenční geny pro rakovinu prsu a jako nejstabilnější geny zvolili *MRPL19* a *PPIA*. (McNeill se sp., 2007) U rakoviny ledvin byly doporučeny jako nejvhodnější HKG *PPIA* a *TBP* (Jung se sp., 2007).

### 4.4 Rakovinné kmenové buňky a housekeeping geny

Rakovinné kmenové buňky (CSC, cancer stem cell) mají zvláštní rysy. Liší se rychlostí růstu, adhezivitou i metabolismem. Lze u nich proto očekávat i rozdíly v genové expresi některých HKG, např. u *ACTB*, a proto je důležité určit pro ně nejvhodnější HKG pro normalizaci dat z qRT-PCR.

Lemma se sp. (2016) srovnávali stabilitu exprese 15 běžně používaných HKG u kmenových buněk sarkomů a karcinomů. Stabilita exprese HKG byla stanovena za pomoci softwaru NormFinder a GeNorm a porovnána s variačním koeficientem. Hodnota stability pro NormFinder a hodnota M pro GeNorm nepřímo úměrně korelují se stabilitou exprese HKGs. Všechny kandidátní HKG měli M hodnotu nižší než je prahová hodnota 1.5, což naznačovalo, že všechny mohou být s ohledem na jejich stabilitu přijatelné (Vandesompele se sp., 2002). Tři nejstabilnější geny pro CSC za použití NormFinder byly *GAPDH*, *PGK1* a *HMBS*, při použití GeNormu *YWHAZ*, *GAPDH* a *TBP*. Nejstabilnější HKG v nekmenových rakovinných buňkách stanovené pomocí NormFinder byly geny pro *18S*

*rRNA*, *TBP*, *PPIA* . Pomocí GeNormu vyhodnotili jako nejstabilnější geny pro *PPIA* a *18S rRNA*. Ve výsledku zjistili, že geny *TBP*, *YWHAZ*, *PPIA* a *HMBS* (Tabulka 7) jsou nejstabilnější HKG pro porovnání mezi CSC a rakovinnými buňkami a že pro spolehlivou analýzu je třeba použít alespoň 4 referenční geny. Zmiňují, že specifické HKG jsou vhodné jen pro specifické typy nádorů, jako jsou sarkomy a karcinomy (Lemma se sp., 2016).

Zkratka genu	Celý název	Funkce
TBP	TATA Box Binding Protein	General RNA polymerase II transcription factor.
YWHAZ	Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein, Zeta	Belongs to the 14-3-3 family of protein which mediate signal transduction
PPIA	Peptidylprolyl Isomerase A	Isomerase involved in the cis-trans isomerization of peptide bonds in oligopeptides
HMBS	Hydroxymethylbilane Synthase	Third enzyme in the heme biosynthetic pathway

**Tabulka 7:** Zvolené geny určené k porovnání mezi CSC a nekmenovými rakovinnými buňkami.

## 5. Funkce housekeeping genů

HKG jsou zodpovědné za udržování základních a životně důležitých funkcí buňky, nezávisle na tkáni, ve které se buňka nachází. To je jeden z důvodů, proč se očekává, že jsou HKG exprimované ve všech buňkách organismu nezávisle na podmínkách nebo stádiu vývoje (Eisenberg a Levanon, 2013).

Překvapivé rozdíly mezi jednotlivými studiemi, které se zaměřily na hledání konkrétních HKG, vedou ke kacířským otázkám, zda skupina HKG vůbec existuje (Zhang se sp. (2015): „Do Housekeeping Genes Exist?“). Podle Zhang se sp. (2015) HKG přeci jen existují, ale je třeba změnit jejich definici. Většina HKG nalezených a ověřených v poslední době je exprimována pouze ve většině tkání při určitých podmínkách daného prostředí, s nízkou úrovní exprese namísto toho, aby byly exprimované hojně ve všech tkáních. V souvislosti s post-transkripčními a post-translačními úpravami, které výrazně zvyšují diverzitu genových produktů, je třeba zvážit, zda není vhodnější definovat housekeeping funkce místo konkrétních genů (Zhang se sp., 2015).

Cílem studie Zhu se sp. (2008 a) bylo zjistit, jaký podíl genů z lidského genomu představují HKG. Srovnávali proto expresní databáze EST a microarrayových analýz nového typu (s 13986 geny) transkriptomu u člověka. Z jejich analýzy vyplynulo, že EST databáze zahrnovaly jen omezený počet tkání a nejsou proto kompletní. Jejich výhodou však byl vyšší záchyt genů z 1 vzorku, než je tomu u microarrayových analýz. Autoři článku se přiklánějí k názoru, že housekeeping geny nelze definovat jako geny se stabilní expresí ve všech tkáních, ale spíše podle jejich funkce. Proto na základě databází Reactome a KEGG sestavili soubor HK408 (Tabulka 8) obsahující seznam genů, jejichž produkty zaručeně zajišťují housekeeping funkce.

Do svého seznamu zařadili geny, jejichž produkty se podílejí na transkripci, translaci, jaderném transportu a proteolýze (Tabulka, Zhu se sp., 2008 a). Z tabulky je zřejmé, že pouze u genů s očekávanou vysokou úrovní exprese je detekční citlivost mikroarrayové analýzy dostatečná k jejich záchytu. To způsobilo velké rozdíly v původních a současných odhadech počtu HKG a potažmo i při popisu jejich vlastností.

Výčet spolehlivých funkcí HKG uvedený v tabulce (Tabulka 8) jistě není kompletní. Pokud se podíváme na funkce spolehlivých referenčních genů u kvasinky *Saccharomyces cerevisiae* nebo rostliny *Arabidopsis thaliana* (Tabulky 4; 5 a 6), lze k těmto funkcím přiřadit např. buněčné dělení, glykolýzu, biosyntézu hemu, regulaci fosforylace, metylaci a mnoho dalších. Sekvenováním RNA a EST analýzami se do souboru HKG zařadili i DNA sekvence, jejichž biologická funkce je zatím neznámá. To se týká převážně *A. thaliana*, zatímco referenční geny pro *S. cerevisiae* jsou v tomto ohledu až na výjimky mnohem lépe prostudované.

Zhu se sp. (2008 a) pomocí souboru HK408 testovali vlastnosti microarrayové a EST analýzy. Předpokládali, že u dobře provedené expresní analýzy by všechny geny ze souboru HK408 měly být zastoupeny v analyzovaném souboru bez ohledu na typ tkáně. U EST dat to platilo u 13 tkání z 18 (u 5 tkání byl proveden jen malý počet analýz). Při microarrayových analýzách byla nižší detekční schopnost, která vedla k absenci genů s nízkou expresí v microarrayových souborech. To se týkalo např. transkripčních faktorů. U obou typů analýz však byla většina genů ze souboru HK408 exprimována ve všech 18 analyzovaných tkáních.

Protože pomocí EST analýzy lze nalézt 93% genů ze souboru HK408 exprimovaných nejméně v 16 tkáních z 18, autoři odhadují, že počet housekeeping genů v genomu člověka

definovaných jako „exprimovaný nejméně v 16 tkáních“ odpovídá rozpětí 3140 až 6909 genů (s klesající mírou falešně negativních genů, „false negative“). U microarrayových analýz bylo 58,5 % genů z HK408 exprimovaných nejméně v 16 tkáních. Podle microarrayových analýz by počet HKG byl 1206-2403. Překvapivý výsledek poskytuje vzájemné porovnání obou souborů HKG. Soubor EST s 6909 HKG obsahuje 71 % genů nepřítomných v souboru microarray s 2403 HKG, obráceně je číslo nižší, pouze 17 %. Soubor EST obsahuje téměř všechny geny z předchozích analýz HKG, provedených pomocí microarray (Warrington se sp., 2000, Hsaio se sp., 2001). Ze srovnání tedy vyplývá, že EST analýzy přinášejí konzistentnější data než microarrayové analýzy a počet HKG je pravděpodobně mnohem vyšší, než jsme původně čekali (Zhu se sp., 2008 a).

Function	Pathway/Complex <sup>a</sup>	# Genes	MA18 <sup>b</sup>	MA16	EST18	EST16
Transcription	Transcription pre-initiation complex	40	3	13	12	28
	Basal transcription elongation factor	17	4	7	5	15
	Capping, splicing and polyadenylation	99	42	55	72	96
Transport	Nuclear pore complex	29	2	4	10	26
Translation	Basal translation factor	37	20	28	30	37
	tRNA synthetase	20	8	13	19	20
	Cytosolic ribosome	82	79	79	78	81
Proteolysis	Ubiquitin mediated proteolysis	45	8	12	22	40
	Proteasome	43	17	26	31	41
<b>Total</b>		<b>408</b>	<b>182</b>	<b>235</b>	<b>278</b>	<b>379</b>

**Tabulka 8:** *Funkční klasifikace HK408 genů.* Sloupec Genes ukazuje celkový počet HKG. Sloupec Ma18 (microarray) reprezentuje výsledný počet HKG vyskytující se ve všech 18 tkáních, Ma16 počet HK vyskytujících se v 16 z 18 tkání. To samé platí pro EST18 a EST1 (Převzato z Zhu se sp., 2008 a).

## 6. Závěr

Pojem housekeeping gen je známý již polovinu století, avšak dodnes se vedou spory o jeho přesné definici. Studie věnované HKG rozebírají několik zásadních otázek. Jaký podíl genomu představují HKG? Jsou kompaktní? Je jejich exprese za všech okolností stejná? Jsou jejich sekvence konzervativní? Tou hlavní však je - zda vůbec existují? Zhang se sp. (2015) na základě rozsáhlé analýzy dostupných expresních dat navrhuje původní definici HKG upravit. Je totiž nepravděpodobné, že lze najít geny, které budou mít ve všech 200 typech tkání u člověka stejnou úroveň exprese v průběhu celého ontogenetického vývoje a nezávisle na dalších podmínkách prostředí apod. Navrhuje zavést termín housekeeping funkce místo housekeeping gen.

Celkový počet HKG v organismech nebyl dosud určen. Výsledky starších výzkumů, zabývající se počtem lidských HKG, se pohybovaly v počtu okolo 500 HKG, avšak shodujících se genů mezi nimi bylo významně málo. Studie Zhu se sp. (2008) předpokládá existenci až 6900 genů. Ve srovnání se staršími studii tento soubor zahrnuje geny s nízkou úrovní exprese a někdy i s dosud neznámou molekulární a biologickou funkcí. Hledání HKG je velmi závislé na citlivosti použité metody, počtu genů, které lze detekovat i počtu tkání a podmínek, které jsou testovány.

Housekeeping geny s vysokou úrovní exprese jsou v souladu s tzv. ekonomickou hypotézou a lze je označit jako kompaktní (Vinogradov se sp., 2004). Z jiných studií však vyplývá, že intenzita genové exprese negativně koreluje především s délkou intronů, ale nezávisle na tom, zda se jedná o HKG nebo TSG (Rao se sp., 2010).

Nezávislé evoluční studie potvrdily, že HKG jsou vývojově starší než TSG (Zhu, se sp., 2008 b). Je rovněž potvrzeno, že frekvence nesynonymních substitucí je v této skupině až 2x nižší než u TSG (Zhang a Li, 2004).

Specifickou skupinou HKG jsou tzv. referenční geny, které se – dle doporučení nejlépe ve větším počtu - používají při normalizaci dat v kvantitativních studiích genové exprese. Stabilita genové exprese u HKG za různých podmínek je důležitým faktorem, který určuje, zda jsou vhodnými referenčními geny. K určení stability byly vyvinuty programy jako GeNorm nebo NormFinder. Volba vhodných referenčních genů zásadně ovlivňuje interpretaci výsledků experimentu i klinického vyšetření.

Studie HKG byly provedeny i některých modelových organismů, např. u myši, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* nebo *Gallus gallus*.



## 7. Seznam použité literatury

- Cankorur-Cetinkaya A., Dereli E., Eraslan S., Karabekmez E., Dikicioglu D., Kirdar B., 2012 A novel strategy for selection and validation of reference genes in dynamic multidimensional experimental design in yeast. *PLoS One* **7**.
- Castillo-Davis C. I., Mekhedov S. L., Hartl D. L., Koonin E. V, Kondrashov F. a, 2002 Selection for short introns in highly expressed genes. *Nat. Genet.* **31**: 415–418.
- Castillo-davis C. I., Hartl D. L., Achaz G., 2004 cis -Regulatory and Protein Evolution in Orthologous and Duplicate Genes. : 1530–1536.
- Claus Lindbjerg Andersen J. L. J., Ørntoft T. F., 2004 Normalization of real-time quantitative RT-PCR data: a mode-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**: 5245–5250.
- Czechowski T., Stitt M., Altmann T. identification and testing of superior reference genes for transcript normalization, Udvardi M. K., 2005 Genome-wide identification and testing of superior reference genes for transcript normalization. *Society* **139**: 5–17.
- D’Onofrio G., 2002 Expression patterns and gene distribution in the human genome. *Gene* **300**: 155–160.
- DeRisi J. L., 1997 Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* (80-. ). **278**: 680–686.
- Eisenberg E., Levanon E. Y., 2013 Human housekeeping genes, revisited. *Trends Genet.* **29**: 569–574.
- Farré D., Bellora N., Mularoni L., Messeguer X., Albà M. M., 2007 Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol.* **8**: R140.
- Gardiner-Garden M., Frommer M., 1987 CpG Islands in Vertebrate Genomes. *J.Mol.Biol.* **196**: 261–282.
- Gasch a P., Spellman P. T., Kao C. M., Carmel-Harel O., Eisen M. B., Storz G., Botstein D., Brown P. O., 2000 Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**: 4241–4257.
- Gonçalves I., Duret L., Mouchiroud D., 2000 Nature and structure of human genes that generate retropseudogenes. *Genome Res.* **10**: 672–678.
- Hellemans J., Mortier G., Paepe A. De, Speleman F., Vandesompele J., 2007 qBase relative quantification framework and software for management and automated analysis of real-time

- quantitative PCR data. *Genome Biol.* **8**: R19.
- Higuchi R., Dollinger G., Walsh P. S., Griffith R., 1992 Simultaneous amplification and detection of specific DNA sequences. *Biotechnology. (N. Y.)* **10**: 413–417.
- Hillier L. W., Miller W., Birney E., Warren W., Hardison R. C., Ponting C. P., Bork P., Burt D. W., Groenen M. A. M., Delany M. E., 2004 Sequencing and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695 – 716.
- Hsiao L.-L., Dangond F., Yoshida T., Hong R., Jensen R. V., Misra J., Dillon W., Lee K. F., Clark K. E., Haverty P., Weng Z., Mutter G. L., Frosch M. P., MacDonald M. E., Milford E. L., Crum C. P., Bueno R., Pratt R. E., Mahadevappa M., Warrington J. A., Stephanopoulos G., Stephanopoulos G., Gullans S. R., 2001 A compendium of gene expression in normal human tissues. *Physiol. Genomics* **7**: 97–104.
- Huang Y. F., Niu D. K., 2008 Evidence against the energetic cost hypothesis for the short introns in highly expressed genes. *BMC Evol. Biol.* **8**: 154.
- Huisinga K. L., Pugh B. F., 2004 A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13**: 573–585.
- Chapman J. R., Waldenström J., 2015 With reference to reference genes: A systematic review of endogenous controls in gene expression studies. *PLoS One* **10**: 1–18.
- Jeffrey C. Hall, Theodore Friedmann, Jay C. Dunlap and Francesco Giannelli. 1999. "Advances in Genetics", Vol. 39 p 18.
- Jung M., Ramankulov A., Roigas J., Johannsen M., Ringsdorf M., Kristiansen G., Jung K., 2007 In search of suitable reference genes for gene expression studies of human renal cell carcinoma by real-time PCR. *BMC Mol. Biol.* **8**: 47.
- Kozera B., Rapacz M., 2013 Reference genes in real-time PCR. *J. Appl. Genet.* **54**: 391–406.
- Lawson M. J., Zhang L., 2008 Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5'-UTR region. *Gene* **407**: 54–62.
- Lemma S., Avnet S., Salerno M., Chano T., Baldini N., 2016 Identification and Validation of Housekeeping Genes for Gene Expression Analysis of Cancer Stem Cells. *PLoS One* **11**: e0149481.
- Lercher M. J., Urrutia A. O., Hurst L. D., 2002 Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.* **31**: 180–183.
- Levanon E. E. and E. Y., 2003 Human housekeeping genes are compact. *Trends Genet.* **19**: 356–362.

- Li S. W., Feng L., Niu D. K., 2007 Selection for the miniaturization of highly expressed genes. *Biochem. Biophys. Res. Commun.* **360**: 586–592.
- McNeill R. E., Miller N., Kerin M. J., 2007 Evaluation and validation of candidate endogenous control genes for real-time quantitative PCR studies of breast cancer. *BMC Mol. Biol.* **8**: 107.
- Pfaffl M. W., Tichopad A., Prgomet C., Neuvians T. P., 2004 Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**: 509–515.
- Rao Y. S., Wang Z. F., Chai X. W., Wu G. Z., Zhou M., Nie Q. H., Zhang X. Q., 2010 Selection for the compactness of highly expressed genes in *Gallus gallus*. *Biol. Direct* **5**: 35.
- Sharan R. N., Vaiphei S. T., Nongrum S., Keppen J., Ksoo M., 2015 Consensus reference gene(s) for gene expression studies in human cancers: end of the tunnel visible? *Cell. Oncol.* **38**: 419–431.
- Stephen A. Bustin, Benes V., Garson J. A., Hellemans J., 2009 The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Appl. Environ. Microbiol.* **65**: 611–622.
- Su A. I., Cooke M. P., Ching K. A., Hakak Y., Walker J. R., Wiltshire T., Orth A. P., Vega R. G., Sapinoso L. M., Moqrich A., Patapoutian A., Hampton G. M., Schultz P. G., Hogenesch J. B., 2002 Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 4465–4470.
- Su A. I., Wiltshire T., Batalov S., Lapp H., Ching K. a, Block D., Zhang J., Soden R., Hayakawa M., Kreiman G., Cooke M. P., Walker J. R., Hogenesch J. B., 2004 A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 6062–7.
- Sun Y., Li Y., Luo D., Liao D. J., 2012 Pseudogenes as weaknesses of ACTB (Actb) and GAPDH (Gapdh) used as reference genes in reverse transcription and polymerase chain reactions. *PLoS One* **7**.
- Teste M.-A., Duquenne M., François J. M., Parrou J.-L., 2009 Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *BMC Mol. Biol.* **10**: 99.
- Vandesompele J., Preter K. De, Pattyn F., Poppe B., Roy N. Van, Paepe A. De, Speleman F., 2002 Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**.
- Vinogradov A. E., 2003 Isochores and tissue-specificity. *Nucleic Acids Res.* **31**: 5212–5220.

- Vinogradov A. E., 2004 Compactness of human housekeeping genes: Selection for economy or genomic design? *Trends Genet.* **20**: 248–253.
- Wang H., Wang J., Jiang J., Chen S., Guan Z., Liao Y., Chen F., 2014 Reference genes for normalizing transcription in diploid and tetraploid *Arabidopsis*. *Sci. Rep.* **4**: 6781.
- Warrington J. A., Nair A., Mahadevappa M., Tsyganskaya M., 2000 Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol. Genomics* **2**: 143–7.
- Watson, J. D., N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner. 1965. *Molecular biology of the gene*, vol. 1. p 704. Benjamin/Cummings, Menlo Park, Calif.
- Zhang L., Li W. H., 2004 Mammalian Housekeeping Genes Evolve More Slowly than Tissue-Specific Genes. *Mol. Biol. Evol.* **21**: 236–239.
- Zhang Y., Li D., Sun B., 2015 Do Housekeeping Genes Exist? *PLoS One* **10**: e0123691.
- Zhu J., He F., Song S., Wang J., Yu J., 2008 (A) How many human genes can be defined as housekeeping with current expression data? *BMC Genomics* **9**: 172.
- Zhu J., Fuhong He, Songnian Hu J. Y., 2008 (B) On the nature of human housekeeping genes. *Trends Genet.* **24**: 478–481.