universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Meta analysis for the finding of tissue specific reference genes in the sections of the small intestine in mice

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer.nat.)

Verfasser:	Florian Rudolf Lukas Meyer
Studienrichtung:	A 441 Diplomstudium Genetik - Mikrobiologie (Stzw) UniStG
Betreuer:	Ao.Univ.Prof. Dr. Ralf Steinborn, Univ.Prof Dr. Thomas Decker

Wien, am 17. Juni 2010

Contents

List of Figures

List of Tables

1 Abstract

Even relatively modest expression changes can have significant biological effects. To quantify such low expression changes the gold standard reverse transcription quantitative real-time PCR (RT-qPCR) has to be combined with accurate normalisation. Successful application of this technique relies on the careful selection of reference genes for normalisation. Tissue wide expression profiling necessitates reference genes similarly stably expressed in every tissue resulting in only a small proportion of genes with a moderate stable expression. This trade-off for expression stability is expected to cease when profiling is being addressed in a specific tissue.

In this work reference genes for expression profiling in a specific tissue was studied for the small intestine of mice. Candidate genes were identified by a meta analysis of microarray RNA expression data of internal and publicly available microarray data. Jejunum samples of one outbred and three inbred strains were used to validate 15 candidates by RT-qPCR. The genes *Plekha7*, *6430706D22Rik*, *EG666853* and *Zfyve19* were identified as the most suitable internal controls with least expression variance of $<$0.8-fold. They exhibited a similar expression alteration as *Oaz1*, but a lower one compared to the B1 and B2 element retrotransposons recently introduced for tissue wide normalisation. This low variance determined for the novel tissue-specific reference genes is superior to the expression stability value identified by GeNorm software previously found for novel reference genes identified by a tissue wide multi platform meta analysis. The high stably expression of the jejunal reference genes suggests that they are involved in tightly regulated pathways in this specific tissue. The functional diversity of these normalisers allows to survey target gene expressions over a wide range of biological conditions and stimuli. Additionally the meta analysis approach was also applied on the other sections of the small intestine and the small intestine as a whole.

The enhancement of normalisation accuracy achieved by expression data meta analysis in a specific tissue underlines the potential for improvement of normalisation strategies in other tissues. The novel reference genes being subject to functional constraints indicated by their highly stable expression are promising targets for future functional analyses.

2 Introduction

Even relatively modest expressionchanges can have significant biological consequences, as seen for the tumor suppressor gene APC, in which a 50% change in gene expression can lead to development of familial neoplasia [67]. Due to its large quantification range, sensitivity and specificity RT-qPCR is the method of choice when the amount of mRNA has to be determined and is used in a wide variety of experiments. However to reach optimal results data normalisation is of critical importance for RT-qPCR. To address this a lot of different methods were invented. When RT-qPCR was introduced, the first generation of reference genes consisted of reference genes like *HMBS (PBGD)*, *GAPDH*, *18S rRNA*, *28S rRNA*, *ACTB* or a number of ribosomal protein genes [12] were used for normalisation. Like every gene, these 1st generation normaliser genes are affected by a number of stimuli [75, 69] and several studies have shown that the genes commonly used for normalisation are differently expressed in different tissues, or show generally high level of variance [64, 76]. Still today reference genes are often incorrectly referred to as housekeeping genes . Later, multi-gene normalisation methods became the standard approach in normalising RT-qPCR and are still the standard approach today [79]. This approach used various mathematical algorithms like the software tool BestKeeper [58], to merge a set of reference genes into an reference index. Initially these methods were used in conjunction with first generation reference genes. However the development of software tools, like GeNorm [79], NormFinder [3], Global Pattern Recognition [2] or Equivalence tests [28], enabled scientists to find the best suited genes for normalisation among the genes used in their study. Yet, especially for GeNorm, the most widespread, because user friendly, tool, the usability is often disputed [3, 12, 22, 18], because the algorithms preference for coregulated pairs of genes. Furthermore these approaches need the incorporation of prechosen reference genes in a low or medium density RT-qPCR study, although they might proof to be of no value to the experiment *post hoc*, binding biological and financial resources.

In the last century finding valuable reference genes was more or less luck based, as genome wide information on gene expression was scarce. Publicly available microarray databases, like Gene Expression Omnibus (GEO) [5] or ArrayExpress [52] or derivatives from thereof like Genevestigator [32] or A-MADMAN [10], have changed this and made it possible to use meta analyses to gain valuable information from a large number of already existing microarrays. Commonly these meta analyses are used to search for deregulated genes. Yet in the mid of the last decade meta analyses were also used to identify genes not deregulated by the experimental conditions [15, 17, 25, 43, 68, 92], comparing thousands of microarrays at once. A comparison of microarray studies is depicted in tab. 1. Basic principles used were similar to some extent using some statistical or math-

ematical models to rank the genes. For example, the relinquishment not to weighten the used experiments for biological importance addressed, the number of samples used per experiment or for the platform used. Indications regarding the stability of the expression levels of specific genes in model organisms may be acquired from the data compiled in public microarray databases. However, this approach raises several challenges. One of those is the reliability of using indicators of the stability of genes expression levels derived by averaging values obtained from large numbers of samples collected under widely varying experimental conditions. As microarrays generate an average signal from different numbers of probes, targeting different splice variants and the possibly from cross reactions, microarrays and RT-qPCR do not correlate in 13-16% of the genes [16] and differ in their accuracy [27]. . However as for the time this work was conducted all meta analyses to find reference genes were done using a single microarray platform. The multi platform approach was first established for the finding of differentially expressed genes. Only recently different microarrays were used for the finding of reference genes [39]. This might be attributed to the difficulties of merging different microarray platforms in a meta analysis, be it for differential expression or not [31, 37, 62]. However laboratory specific influences on experiments also influence meta analyses on a single platform, although presumably to a lesser extend [72]. In recent days the development of on-line meta analysing tools, like A-MADMAN [10] and PHOENIX [8] has started.

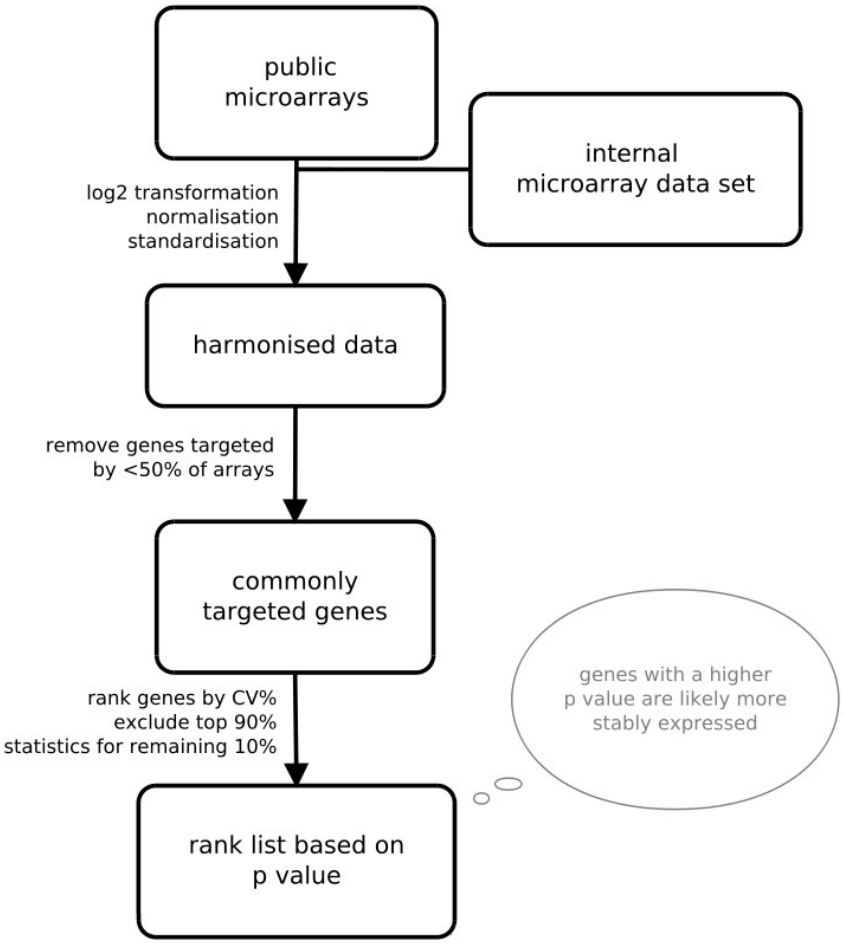

Figure 1: Flowchart of the methodology of the meta analysis to find new reference genes giving a short introduction in the method used.

As each tissue is subject to different functional constraints, they express specific sets of

Table 1: Analysis and meta analysis strategies used in previous studies for identification of reverence genes across tissues and experimental conditions.

Study	Organism	Platform	Array no.	Gene exclusion criteria	Ranking criteria	Platform no.	Weightening	Validation
Jin et al 2004		2 color cDNA microarray	384	missing spots	SD and fluorescence intesity	1	no	qPCR
Czechowski et al 2005	*Arabidopsis thaliana*	1 color 25-mer microarray	721	no exclusion	CV%	1	no	qPCR
Shulzhenko et al 2005	*Homo sapiens*	1 color oligonucleotide array	75	no exclusion	two subsets statistically compared	1	no	qPCR
de Jonge et al 2007	*Homo sapiens/Mus musculus*	1 color 25-mer microarray	13629 /2543	no exclusion	CV%, maximum fold change and expression level	2[3]	no	PCR
Maccoux et al 2007	*Canis lupus familiaris*	60-mer microarray	26	t-test	fold change and CV%	1	no	qPCR
Saviozzi et al 2007	*Homo sapiens*	1 color 25-mer microarray	82	no exclusion	integrative correlation and SD	1	no	qPCR
Waxman and Wurmbach 2007	*Homo sapiens*	1 color 25-mer microarray	72	no exclusion	SD	1	no	qPCR
Frericks et al 2008	*Mus musculus*	1 color 25-mer microarray	1968	no exclusion	CV% and SD	1	no	qPCR
Kwon et al 2009	*Homo sapiens*	EST[1], SAGE[2], 1 color 25-mer microarray	77/32 6/567	0's proportion, mean expression and CV%	z-test and clustering	3	no	qPCR
Zhou et al 2010	*Rattus norvegicus*	1 color 50-mer microarray	12	no exclusion	CV%	1	no	qPCR
This work	*Mus musculus*	1 color 25-mer to cDNA microarrays	220	CV%	Chi[2] test	9	no	qPCR

[1] Data derived from Expressed Sequence Tag Sequencing libraries

[2] Data derived from Serial Analysis of Gene Expression libraries

[3] near identical set of genes targeted

genes [59], we therefore hypothesised that reference genes should be searched and used in a tissue-specific manner. We therefore conducted a meta analysis aiming to find specific sets of reference genes for the three segments of the small intestine of mice. Because of the small number of microarrays of one platform available for the different sections of the small intestine, a multi platform approach was chosen. We also included a set of microarrays from an internal study, testing the effect of genetically modified MON810 maize on the small intestine of mice *in vivo*. This maize contains both artificially introduced resistances to the broad-spectrum herbicide RoundUp (Monsanto, St. Louis, USA) and the european corn borer larvae (*Ostrinia nubilalis*). The latter is mediated by the crystal protein Cry1Ab of *Bacillus thuringensis*. Cry1Ab was reported to bind to the mucosal surface of the mouse intestine and to induce *in situ* temporal changes in the electrophysiological properties of the mouse jejunum [81]. It can not be excluded that each sample analysed in this internal microarray data was not generated at the same time of day. Here I summarise molecular indication supporting the assumption that sampling was not in accordance with exclusion of circadian clock effects. The circadian clock system, altering a large percentage of the cellular transcripts ($\geq$8–10%), but not uniformly across tissues [73]. If not controlled for it, the circadian clock system can be a major pitfall for a study, rendering the data meaningless, as shown in this work, as the expression alterations could not be allocated to the stimulus studied. Still this internal data set was valuable for the conducted meta analysis and aided us in the finding of a new sets of reference genes for the jejunum and the small intestine, enlarging the number of reference genes derived from different pathways for these tissues.

3 Results

The effect of genetically modified maize (Mon810 x NK603), expressing the crystal toxin Cry1Ab of *Bacillus thuringensis* and the herbicide resistance inducing protein (Epsps) of the *Agrobacterium tumefaciens* strain CP4 ,on the outbred mouse strain OF1 was studied. Two experimental designs were chosen. In the multi generation study (MGS), usually at least three generations of the test organism are exposed to the substance being assessed. The other study design featured a risk assessment using continuous breeding (RACB) design, investigate midterm effects caused by repeated reproductive stress . Mon810 x NK603 mice feed (feed group G) was tested against isogenic NK603-maize (feed group K), featuring glyphosate resistance and an other conventional maize (SARASTRO) (feed group A) as control samples, in the MGS, in the RACB SARASTRO-maize was neglected. After the feeding studies were completed samples of the jejunum of each 12 male and female mice were taken, in RACB scrape samples of all tested parental female mice were also taken. For the MGS mice of the 3rd generation (F3) were sampled at the age of 7 weeks, for RACB mice of the 3rd litter of the tested pairs were sampled 7 weeks after birth(3rd F1),and F0 mice were sampled 26 weeks *post natum*.

Nine male mice samples of the MGS were used in an other study for microarray analysis [9]. Based on this microarray data micro fluidic cards (TLDA) were designed, targeting 45 genes derived mainly from protein metabolism with putative altered gene expression on the microarray ($p \leq 0.05$) and three first generation reference genes (see supplementary tab. 10). For the MGS six male mice of all three feeding groups were examined by TLDAs. For the F0 of the RACB six male mice of each feeding group and three female mice of each group as well as three scrape samples of female mice of both groups were analysed. Four sample of groups G and K were tested in the 3rd F1 of the RACB. The crossing point (C_q)of the resulting plots was determined using SDS 2.3 software. REST [57], featuring a very stringent statistic, was used to analyse the differential gene expression between the individual groups. For each experiment the usability of genes for normalisation demonstrated was by GeNorm [79] ranking and the number of normaliser genes need for best results determined. BestKeeper [58] was used to calculate an artificial normalisers gene for each experiment. For REST analysis an efficiency of 1.0 was assumed for all genes [4].

Unexpectedly, several target genes placed higher in GeNorm analysis than than the reference genes included in the TaqMan assays [9] (Fig.2A). Further GeNorm analysis of the other experiments showed similar results (Fig. 2b)

MGS samples of *18SrRNA* (0.74-fold change), *Fkbp5* (3.73-fold change), *Igtp* (0.33-fold change), *Per3* (3.29-fold change), *Pnpla3* (0.26-fold change) and *Socs1* (0.59-fold change) were differentially expressed (p=0.05), when the effects of feed G was compared to both

Table 2: Fold change of the individual genes in the different experiments. Blue, green and yellow colored boxes indicate a p value below 0.001, 0.05 or 0.055, respectively, determined by REST. 18S rRNA deregulation shows the inadequacy of first generation reference genes

	MGS				RACB(F0)			MGS and RACB(F0)	MGS, RACB(F0) and RACB 3rdF1)
	G compared to K and B	G compared to K	G compared to B	B compared to K	G compared to K				
	Jejunum	Jejunum	Jejunum	Jejunum	Jejunum		Scrape	Jejunum	Jejunum
	♂	♂	♂	♂	♀	♂	♀	♀/♂	♀/♂
18SrRNA	0.74		0.73						
Cd40			0.48					0.64	0.69
Cd68			0.65						
Csf3r							0.51		
Fkbp5	3.73	2.94	4.42					1.87	1.61
Foxq1						0.55			
Igtp	0.33	0.26	0.38						
Per3	3.29		5.83	0.29		4.34		2.36	2.08
Pnpla3	0.26	0.23	0.17					0.45	0.53
Socs1	0.59	0.54	0.61						

feeds K and A. *Fkbp5* (2.94-fold change), *Igtp* (0.26-fold change), *Pnpla3* (0.23-fold change) and *Socs1* (0.54-fold change) showed significant fold change (p=0.05) when Feeding group G was compared to isogenic feeding group K. The comparison of group G and A revealed larger differences (*18SrRNA* (0.73-fold change), *Cd40* (0.48-fold change), *Cd68* (0.65-fold change), *Fkbp5* (4.42-fold change), *Per3* (5.83-fold change), *Pnpla3* (0.17-fold change), and *Socs1* (0.61-fold change)(p=0.05) furthermore *Igtp* (0.38-fold change) was bordering to significance, showing a p-value of 0.055). It has to be noted that *18SrRNA* was initially included as a reference gene. Comparison of feed group A and K showed that *Per3* was downregulated in group A by 0.29-fold.

RACB(F0) samples of male mice showed deregulation of *Foxq1* (0.55-fold change) and *Per3* (4.43-fold change) (p=0.05). Only *Csf3r* was downregulated 0.51-times (p=0.05) in the scrape sample of female mice in the RACB2, no gene was found to be deregulated in the whole jejunum sample

None of the 47 genes under investigation was found to be changed in the 3rd F1 generation of the RACB experiment.

When data of both male and female mice from the MGS and RACB(F0) was pooled, expression of *CD40* (0.64-fold change), *Fkbp5* (1.87-fold change), *Per3* (2.36-fold change) and *Pnpla3* (0.45-fold change)(p=0.05) were altered. Addition of RACB(3rd F1) led to the same results, although the fold-changes found were slightly reduced (*CD40* (0.69-fold change), *Fkbp5* (1.61-fold change), *Per3* (2.08-fold change) and *Pnpla3* (0.53-fold change))(p=0.05).

As noted before we found that genes not commonly used for normalisation and shown for the MGS found to be deregulated by other experimental techniques were ranked

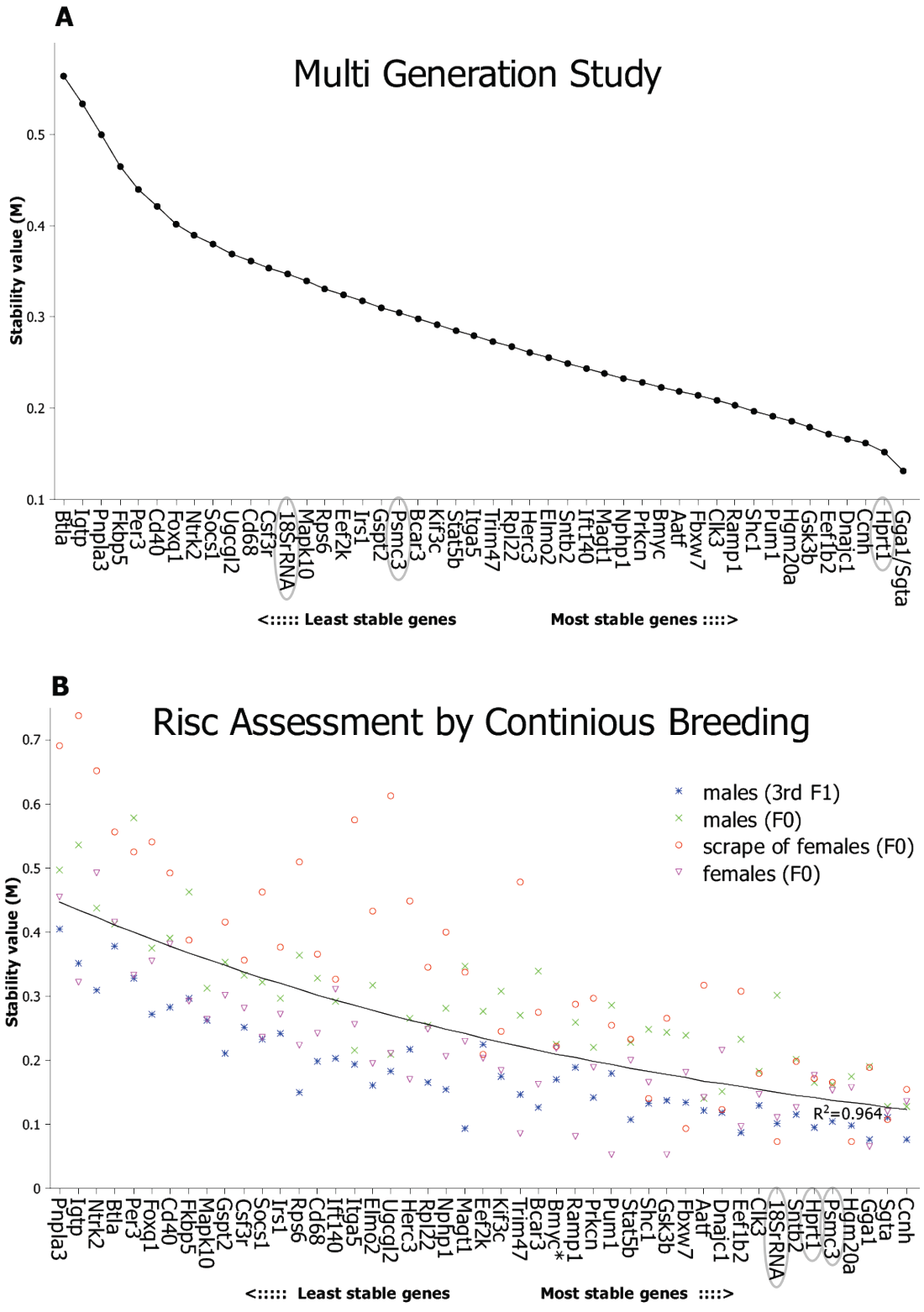

Figure 2: Distribution of stability falues (M values) determined by GeNorm, indicating new normaliser genes. (A) M values determined by the C_qs of MGS data. (B) M values calculated by the use of C_qs of the RACB study. A exponential trendline is shown in black indicating a strong correlation between the average M value of a gene and its positioning by an additive rank (R^2= 0.964).

*Probe also recognises A230005M16Rik

higher as normalisation genes by GeNorm than the first generation reference genes used on the TLDA. This indicates the potential existence of normalisation genes outperforming the classically used reference genes.

Previous studies have already revealed new reference genes by the meta analysis of microarray data across a wide set of tissues [17, 25]

Here we argued that meta analysing expression data of microarrays for a specific tissue or its substructures would identify novel candidate normaliser genes being more stably expressed. We addressed this issue exemplarily by performing a meta analysis of public and own microarray mRNA-expression data for the sections of the small intestine and small intestinal samples of unspecified origin. 22 studies comprising a total of 220 microarrays derived from samples with defined (duodenum: n=23, jejunum: n=53, ileum: n=63) and unspecified origin (n=92) were meta analysed. A survey of the microarrays used can be seen in tab.3

For this meta analysis genes and expressed sequence tags (ESTs) were selected, if detectable above a given S/N ratio on at least half of the intestinal microarrays.

The meta-analysis list's top 100

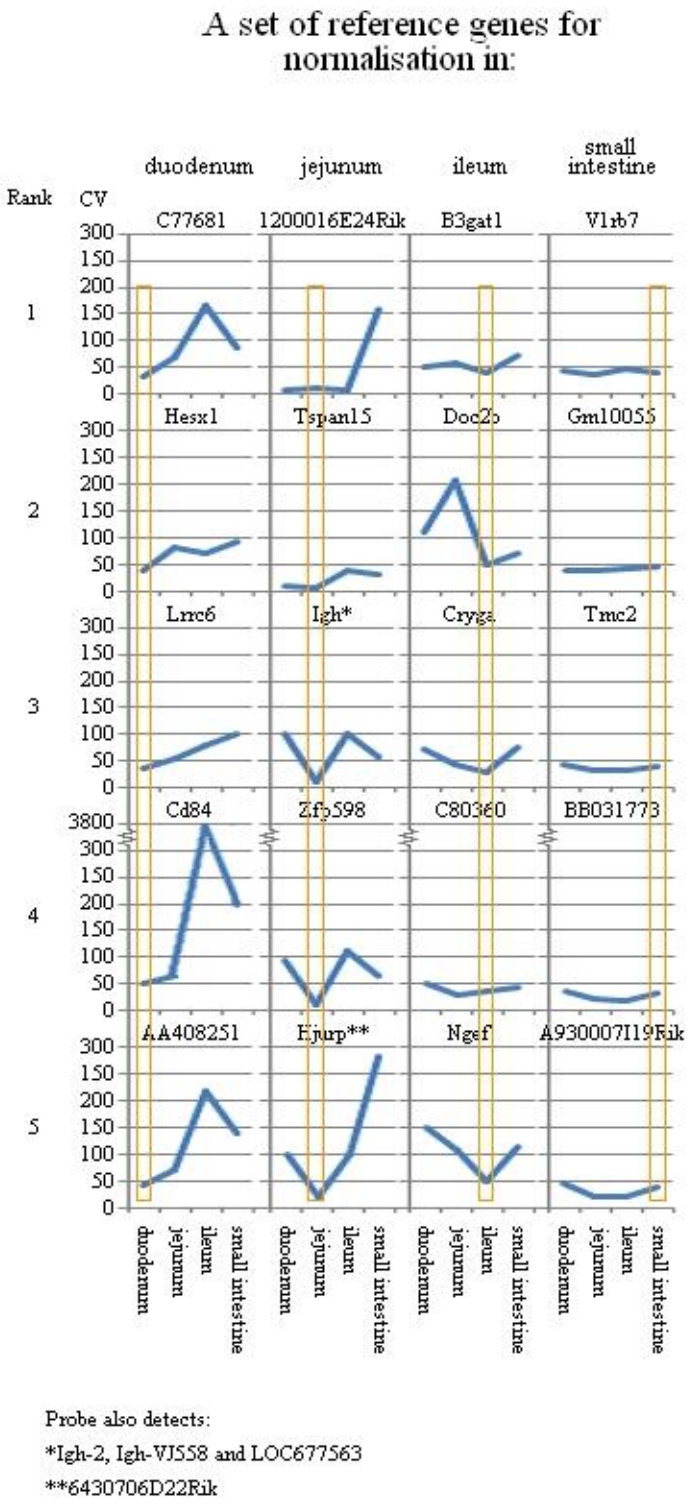

Figure 3: Comparison of the CV% of the top ranked genes in each small intestinal section in the different tissues

Table 3: Guidance characteristics of the microarray experiments used in for the metaanalyses. Gray colored boxes indicate outbred mice strains, light gray colored boxes indicate a mixed genetic background.

Study issue	Mouse strain	Microarray	No of arrays	Reference	GEO accession number
GMO diet	OF1	Gene Expression Array System (Applied Biosystems)	9	unpublished*	to be submitted
gene knock-out (GATA4)	C57BL/6J x SJL/J C57BL/6	Mouse Genome 430 2.0 Array (Affymetrix)	9	Battle et al., 2008	GSE11194
PUFA-diet	C57BL/6J		2	van Schothorst et al., 2009	GSE11936
Metagenome-effects	NMRI		18	Rawls et al. 2006	GSE5198
gene knock-out (PTEN)	not specified		5	He et al. 2007	GSE6078
gene knock-out (Hfe)	B6;D2		12	Coppin et al., 2007	GSE7357
high fat/low fat diet	C57BL/6J		18	de Wit et al., 2008	GSE8582
immunological challenge	B6		9	Peterson et al., 2007	GSE9018
gene knock-out (PLAGL2)	Swiss Webster x 129/SvJ		8	Van Dyck et al., 2007	GSE9123
RNA profiling across tissues	not specified	IncyteMouseGEM1 (non commercial)	23	unpublished	GSE1701
RNA profiling across tissues	C57BL/6J		2	Hutton et at., 2004	GSE2168
RNA profiling across tissues	C57BL/6J		12	Zhang et al., 2004	GSE2178
RNA profiling across tissues	not specified	20K Riken cDNA array; Spotted Riken cDNA array 20-40K; Spotted Riken cDNA array 40-60K (non-commercial)	2	Ravasi et al., 2006	GSE3098
high fat diet	C57BL/6J; A/J	Mouse Genome 430A Array (Affymetrix)	4	Kondo et al., 2006	GSE3433
gene knock-out (Por)	C57BL/6		24	Mutch et al., 2006	GSE4257 GSE4258
gene knock-out (Klf9)	C57BL/6J		10	Simmen et al., 2007	GSE6443
RNA profiling	BALB/c	NTU_CGM_MCF Mouse 6.1k Microarray (non-commercial)	10	Yu et al., 2008	GSE4715
immunological challenge	B6	MG-U74A,B,C (Affymetrix)	20	Lecuit et al., 2007	GSE7013
RNA profiling	Hsd:ICR(CD-1)		27	Mutch et al., 2004	GSE849
tumor study	Apc1638N	Mouse Genome 430A 2.0 Array (Affymetrix)	3	Kucherlapati et al., 2008	GSE13298
RNA profiling across tissues	C57BL/6	MG-U74Av2 (Affymetrix)	4	Freilich et al., 2005	HGMP2**

* unpublished data by Binter C., Gülly C., Velimirov A., Steinborn R., and Zentek J.

** Array Express accession number

genes in each of the four meta analyses contained exclusively novel genes, not found before by other transcriptome wide meta analyses across a wide range of tissues and experimental conditions.

Next we asked whether a section-specific, optimised set of normaliser genes can be composed. From each gene list constituted for the duodenum, jejunum, ileum or the small intestine, the top five genes were selected and their CV%s were crosswise compared in all analysed tissue sections (Fig. 3).

The minimal CV% for 13 of the 15 genes selected from the list of the duodenum, jejunum and ileum was found in the section they originated from. The exceptions were two gene/EST sequences (*1200016E24Rik* and *Doc2b*) derived from the lists of the jejunum and ileum respectively.

Most evaluated genes are expressing a wide range of CV%s in the tested tissues, indicating a high variability in gene expression in different tissue sections, for the tested genes. This gives evidence to the hypothesis that tissue section specific normaliser genes should be used.

In the meta analysed small intestine list however, in contrast to results of the tissue specific meta analyses, crosswise analysis showed that none of the five genes had its minimum CV% in the tissue it originated from. We do not think that this is in contrast to our prior findings, as a set of tissues was investigated, and only genes with low CVs in all other tissues were likely to show up as normaliser genes for the small intestine. This is supported by the generally lower differences between the CVs observed for the potential normaliser genes in the small intestine compared to the other tissues.

The top 16 genes ranked by meta-analysis of microarray expression data in the jejunum were exemplarily further analysed by RT-qPCR. Out of these Igh (Immunglobolin Heavy chain) was excluded, because of its susceptibility to immunological responses and the challenges raised by site-directed DNA rearrangement, known as VD(J) recombination, for assay design. Three genes listed top in a large meta-analysis data set of murine microarrays from multiple tissues and one B element (Huggett and Vandesompele, unpublished), a transcribed, repetitive genomic sequence, were included in addition.

mRNA profiles of one outbred and three inbred mouse strains were monitored by RT-qPCR, representing the gold standard technique for measuring RNA expression. Considering that meta-analysis as well as RT-qPCR data were derived from a wide range of nuclear genetic backgrounds the congruence the ranges of the expression changes measured with the two techniques could be demonstrated (tab. 4). A detailed list of the genes ranked by their variation in RT-qPCR is presented in Table 4.

The very low fold-changes down to 1.4 and M values down to 0.135 (tab. 4), the identification of new stably expressed genes (*Plekha7*, *6430706D22Rik*[1], *Ube2v1/EG666853*, *Zfyve19*

[1]MSOAR software [26] appoints *HJRP* as the human orthologous gene of *6430706D22Rik*

Table 4: The rank of reference genes based on their differential expression found by RT-qPCR, compared to the results obtained by the meta analysis.

		Microarray Data		qRT-PCR					
Gene Symbol	Accession Number	Log2 Median Fluorescence	Fold Deregulation	Median Ct (calculated to 100% Eff.)	Fold Deregulation	M value	Gene ontology term (GO number)	ChiSq	Rank
Plekha7	NM_172743	1.31±0.09	0.35	26(23) ± 0.13	0.44	0.240	not available	0.19	1
6430706D22 Rik	XM_001005388.1	1.62±0.16	0.66	26(22) ± 0.22	0.77	0.274	0034080 0034509 0007059	0.26	2
Ube2v1	NM_023230.2	2.12±0.013	0.56	31(24) ± 0.22	0.78	0.245	0043687 0051246	0.12	3
Zfyve19	NM_028054.2	1.52±0.13	0.65	27(24) ± 0.2	0.79	0.135	not available	0.07	4
Tspan15	NM_197996.2	2.52±0.19	0.94	23(22) ± 0.25	1.06	0.135	not available	0.32	5
Zfp598	NM_183149.1	2.08±0.14	0.67	25(22) ± 0.37	1.26	0.252	not available	0.26	6
Atp6v0d1	NM_013477.3	2.32±0.11	0.47	24(21) ± 0.35	1.40	0.202	0007420 0006810 0006811 0015986 0015992	0.07	7
Fbln1	NM_010180.2	2.33±0.38	1.59	25(17) ± 0.39	1.48	0.314	0030198	0.00	8
Tmem14c	NM_025387.3	2.7±0.49	2.36	24(20) ± 0.41	1.64	0.335	0006783	0.01	9
1200016E24R ik	XM_001472371.1	2.02±0.25	1.18	24(19) ± 0.68	2.17	0.391	not available	0.71	10
AI314976	NM_207219.3	1.14±0.13	0.53	25(19) ± 0.55	2.32	0.363	not available	0.15	11
2310046K01R ik	NM_027172.2	2.78±0.38	1.62	25(19) ± 0.72	2.66	0.483	0034605	0.03	12
B3gnt3	NM_028189.3	2.27±0.38	1.45	27(23) ± 0.93	2.97	0.576	0006486	0.00	13
LOC630729	XM_904332.3	2.63±0.29	1.02	23(20) ± 0.77	2.98	0.439	not available	0.01	14
Hadhb	NM_145558.1	3.27±0.23	1.12	23(20) ± 0.88	3.50	0.525	0008152 0006631 0006635	0.04	15

Median Ct gives an indication of the range the treshold cycle can be expected.
Only GO numbers of the processes the protein is involved in are given for detailed information see http://www.ebi.ac.uk/QuickGO
Biological pathway gives the biological pathway the gene is involved according to PANTHER (Thomas et al. 2003)
M values as determined by GeNorm software

Table 5: Ranking of the individual Meta analyses for different sections of the small intestine and their p-value calculated by a X^2 test

Rank	Jejunum		duodenum		illeum		small intestine	
	Gene detected	p-value	Gene detected	p-value	Gene detected	p-value	Gene detected	p-value
1	1200016E24Rik	0.7090	C77681	0.9972	B3gat1	0.9993	V1rb7	0.0869
2	LOC669875 /// Tspan15	0.3397	Hesx1	0.9646	Doc2b	0.5681	ENSMUSG 00000058934	0.075
3	Igh /// Igh-2 /// Igh-VJ558 /// LOC677563	0.2921	Lrrc6	0.9638	Cryga	0.562	Tmc2	0.0316
4	Zfp598	0.2898	Cd84	0.959	C80360	0.5556	BB031773	0.0144
5	6430706D22Rik /// C330011F01Rik	0.276	AA408251	0.9342	Ngef	0.4238	A930007I19Rik	0.0095
6	Plekha7	0.2234	Tectb	0.9236	Limk2	0.341	Tex16	0.0064
7	AI314976 /// LOC672876	0.1734	Dmrt1	0.9133	Gabrr1	0.2843	1700123K08Rik	0.0017
8	EG666853 /// Ube2v1	0.1429	6332401O19Rik	0.9124	Tekt2	0.1578	OTTMUSG 00000015282	0.0012
9	Zfyve19	0.1001	Cryba1	0.9016	Serpinb3a	0.154	Asb18	0.0008
10	Atp6v0d1	0.0988	Prl3c1	0.8986	Cacna1g	0.1125	Slc5a7	0.0005
11	Hadhb /// LOC623031	0.0565	Ambn	0.8891	Fcho1	0.0952	Abpb	0.0004
12	2310046K01Ri	0.0501	Gabrr1	0.8868	Col4a3	0.0735	1700074P13Rik	0.0002
13	LOC630729	0.0235	Zfp57	0.8356	Nkx2-9	0.0719	LOC100044065;Odz1	0.0002
14	Tmem14c	0.0237	Tyrp1	0.8275	Top3b	0.0718	B230217O12Rik	0.0001
15	Fbln1	0.0115	Fbxo16	0.8190	Prl6a1	0.0675	remaining genes	<.0001

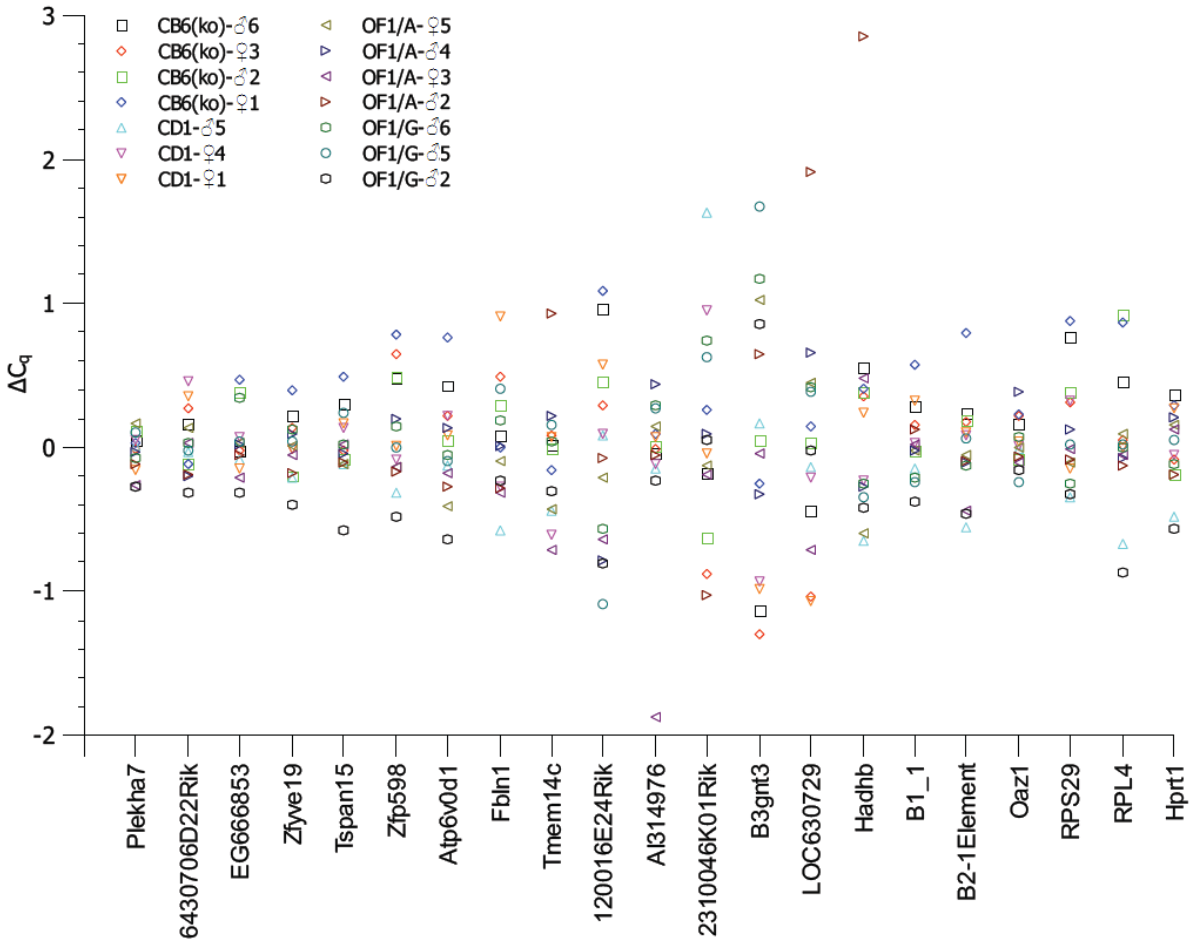

Figure 4: Relative C_q of 15 potential reference genes in comparison to one first generation reference gene, three reference genes suggested by [17] and two B-elements (Huggett and Vandesompele, unpublished), in 3 different mouse strains and 4 different feeding regimes.

and others) and their derivation from differential biological processes display the usability of our approach to find uniformly expressed genes. One gene showed very stable expression across samples, but deviated in expression in one outbred sample. We assume that this is caused rather by genotype than by a technical issue (Fig.4).

In summary, RT-qPCR has supported reasonability of our meta-analysis protocol. It demonstrated that even a transcript that is altered in its expression at a common set of microarrays analysed in this study can fall within the ranges of acceptable variation measured by RT-qPCR.

Next we used meta-analysis to select a toolbox of potential reference genes for the use in expression studies of intestinal samples beyond the jejunum. (see above), further sets of microarrays were meta-analysed to find new genes stably expressed in the duodenum, ileum and the whole small intestine.

The numbers of genes above the threshold of significance (p = 0.05) determined by meta-analysis of the ileum (n = 14) and the jejunum (n = 10) were comparable. However, neither of the 15 genes ranked top in the jejunum appeared among the top 100 of the respective ileum data. This can be attributed to the large number of target genes, the diversity of biological issues addressed by the study set and the impact of an individual array given the limited number of samples microarrayed.

In contrast, in the meta-analysis performed for the duodenum a much larger number of genes were found to be above the significance threshold (n = 123). This is not surprising considering that the duodenum is adjacent to pancreas, a tissue with an abundant set of different RNases, presumably caused by reduction of the accuracy of the measurement. Likewise, neither of the 15 genes ranked top in the duodenum listed among the top 100 of the respective jejunum or ileum data.

All four meta analyses performed identified genes from different biological processes . Thus they constitute a valuable set of new genes allowing a multigene normalisation strategy in RT-qPCR expression studies.

However the results also show that microarray-data and qPCR-data are only comparable to some extend when it is looked for small changes ($\leq$ 2-fold) in gene expression. While the microarray data is correlating with the qPCR data, the quality of genes as normaliser genes can only be determined by qPCR. Therefore verification of meta analysis-data in conjunction with normaliser genes is mandatory.

In contrast to other findings our study showed that genes from mostly different pathways can be found.

4 Discussion

This study started from an internal data set [9] of a circadian clock non-controlled microarray profiling experiment addressing a genetically modified maize diet in a MGS setup. In addition to reporting the day time of an experiment (as requested [65]), the randomisation of sampling is necessary to exclude circadian clock-based influences.This type of expression regulation affects about 8 to 10% of the genes expressed in a cell, shown to be tissue specific for a number of tissues ([50, 73]).The fact that sampling was not randomised and that the time and order of sampling was not specified does not allow to associate the deregulations found with either the diet and/or the sampling time/order.

This theory was supported by RT-qPCR low density array data obtained for a set of genes selected from the top biological processes ranked by overrepresentation analysis of the microarray data (see supplementary tab. 8).

Altogether, all RT-qPCR analyses performed for MGS, RACB or a combination of both showed a total of ten altered genes. Five of those were reported in conjunction with circadian clock-expression(*Per3*, *Socs1*, *Fkbp5*, *Foxq1*, *Pnpla3*) ([63, 49, 80, 45]). This indicates the influence of the circadian clock system on the experiments. Considering the still missing transcriptome-wide expression data for the intestine, it can not be excluded that some of the remaining 38 genes are also under a different circadian clock phase and/or amplitude.

PANTHER analysis of the microarray data showed the highest significance for the circadian clock system in two out of three comparisons done and was most obvious in the overrepresentation analysis performed between G and A groups (see supplementary tab. 8), but not significantly deregulated in the third (G compared to K) [9]. This probably indicates that feed groups G and K where sampled at a similar point in contrast to A. The higher significance in the comparison of A and K (p=6.4E-06) compared to G vs both controls (P=2.0E-0.3) supports this argumentation line. However these indications do not rule out a possible effect caused of the diet.

Although all genes assayed by RT-qPCR (except the first generation reference genes used) were shown to be deregulated on the microarray, only 10 genes were altered in RT-qPCR (Tab.2). The generally low number of significantly deregulated genes is partly attributed to the fact that the statistical test used for TaqMan array analysis (Pair Wise Fixed Reallocation Randomisation Test ([57])) is more stringent than the Welch t-test (not assuming an equal variance between samples) used for the analysis of the microarrays.

To show the effect of statistical test stringency on the significance value, I exemplarily calculated the differences in feed group G compared to the controls using a heteroscedastic student's t-test with a low stringency and comparing them to the results obtained by the more stringent Pair Wise Fixed Reallocation Randomisation Test. P values gained by the

Table 6: Statistics for deregulated genes of the TaqMan array profiling data

Gene	p-value (PWFRRT)	p-value (t-test)
18S	0.007	0.002
2610529C 04Rik	0.530	0.112
Bmyc	0.810	0.418
Aatf	0.276	0.030
Bcar3	0.320	0.019
Btla	0.495	0.276
Ccnh	0.784	0.181
Cd40	0.102	0.034
Cd68	0.212	0.095
Clk3	0.879	0.174
Csf3r	0.424	0.262
Dnajc1	0.554	0.065
Eef1b2	0.331	0.024
Eef2k	0.086	0.008
Elmo2	0.389	0.272
Fbxw7	0.330	0.250
Fkbp5	<0.000	<0.000
Foxq1	0.092	0.015
Gga1	0.843	0.188
Gsk3b	0.520	0.053
Gspt2	0.533	0.053
Herc3	0.988	0.320
Hmg20a	0.793	0.399
Hprt1	0.944	0.330
Ift140	0.345	0.039
Igtp	0.001	0.005
Irs1	0.475	0.083
Itga5	0.277	0.212
Kif3	0.511	0.358
Mapk10	0.507	0.369
Nphp1	0.573	0.401
Ntrk2	0.709	0.414
Per3	0.001	<0.000
Pnpla3	<0.000	<0.000
Prkcn	0.731	0.494
Psmc3	0.209	0.192
Pum1	0.919	0.229
Ramp1	0.449	0.065

continuation

Gene	p-value (PWFRRT)	p-value (t-test)
Rgs6	0.775	0.199
Rpl22	0.767	0.454
Sgta	0.887	0.352
Shc1	0.883	0.340
Sntb2	0.392	0.046
Socs1	0.009	0.004
Stat5b	0.750	0.495
Trim47	0.832	0.431
Ugcgl2	0.669	0.429

student's t-test were lower or equal to the ones received by the Pair Wise Fixed Reallocation Randomisation Test and more than twice as many genes were deregulated significantly (n=17 $p \leq 0.05$) (tab.6). For two thirds of the genes deregulation was found by microarray analysis but not by RT-qPCR. This could be explained by the different number of target genes assayed (33012 vs. 46+1)

Furthermore (as noted earlier) we found genes significantly altered in a microarray experiment to be better suited for normalisation than the first generation genes included in the TaqMan array. Therefore we performed a meta analysis of public and own RNA expression microarrays of the duodenum, jejunum, ileum and the entire small intestine to find new genes for normalisation of RT-qPCR and increase the number of pathways and processes (GO categories) they are assigned to. Briefly, we removed genes targeted only by less than half of the experimental microarrays, calculated the CV of the remaining ones, and listed the top ten percent genes based on a p value calculated using the parametric Wald test. Finally, the reasonability of the meta analysis approach was demonstrated by RT-qPCR quantification in the jejunum (tab. 4).

The meta analysed expression data sets for the duodenum, the jejunum, the ileum and the small intestine showed very little overlap regarding the top ranked genes. Furthermore Fig. 3 shows that the CV of most of the top five genes for a specific section of the small intestine had their lowest CV of all surveyed sections in the section they where found to be useful as normaliser genes. However in the meta analysis done for the whole small intestine this was not true. We believe that this is because of the need to find genes evenly expressed in all tissues of the small intestine. These findings demonstrates the existence of tissue- and section-specific normaliser genes in the small intestine.

Compared to the two pathways found in the set of 15 putative normaliser genes in a tissue-wide meta analysis ([17]), our meta analysis supported by RT-qPCR identified several genes belonging to alternative GO categories, the eight genes with an assigned GO process belonged to 17 individual processes, for the remaining seven genes no GO process was specified. Of the first three genes found by the latter only RPL4 was not excluded prior to calculation, the remaining genes were not among the top 10 % of CV% calculated in this study. However they were included in RT-qPCR analysis as a reference, and proved to be in the same range of variance as the genes found in this study. Still both results showed that a validation by RT-qPCR is indispensable, as the results showed that meta analysis can only be used as a reference to find suitable reference genes, but only RT-qPCR is sensitive enough to show which gene is more stably expressed. A number of reasons could be thought of to explain why microarray derived meta analysis data is not completely in line with RT-qPCR, like cross-hybridisation, splice variants, exons targeted or capping problems, however it is a subject for future studies to explain and the reproducibility of microarrays and the influence of platforms

are still a matter of debate ([1, 72, 55]).

The validation of genes showed to be of importance, as meta analysis derived data only showed genes of low variance but showed a much greater variability in RT-qPCR.

The genes found originate from a wide array of different biological pathways (tab. 4), an important aspect for their use in different profiling experiments were certain pathways could be effected by the experimental conditions. We also found that two genes, namely *Tspan15* and *AI314976*, seemed to be very uniformly expressed in inbred mice. Their different expression found in some individuals of outbred strains, might be the reflection of numerous SNPs within introns or UTRs(see supplementary tab. 7 and tab. 8). However, as for *Tspan15* and *AI314976*, we strongly believe that these genes can be used as normaliser gene in inbred strains, as it is very uniformly expressed in these strains.

We also present a meta analysis for the whole small intestine, including all arrays from the individual arrays as well as a number of arrays lacking intestinal segment information. While the general idea of this work was that of a tissue specific normaliser-gene set, a meta analysis over a small part of tissues, is useful if the exact origin of the sample is unknown, or it the sample is taken from a transition point between the tissues. This issue is commonly encountered in the small intestine as the borders are not clear cut. Furthermore a gradient study would also call for a set of reference genes stably expressed over a selection of intestinal sections.

The finding of very evenly expressed genes is of special importance, when minor changes in gene regulation are issued. Variances in the genes used for normalisation, while without effect on gene deregulation larger than 10-fold, easily could disguise deregulations smaller than two fold. Therefore it is of importance to improve the quality of normalisation procedures and the genes used for it.

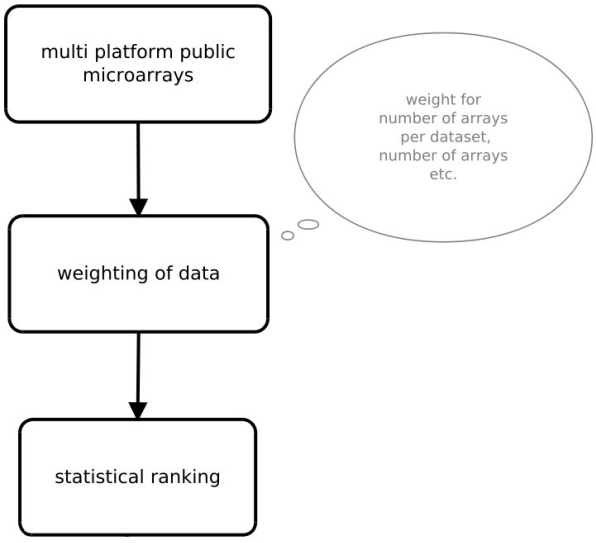

Figure 5: Potential model of a more accurate meta analysis strategy

We believe that new uprising techniques, like second generation DNA sequencing technology ([39])will provide further data for the finding of normaliser genes, and methods will be found that enable scientists to combine the information of second generation DNA sequencing data and microarrays in future meta analyses. However today these techniques do not detect expression changes as reliable as microarrays ([84]), but future methods will

overcome this problem.

At present little is known about the effect of small deregulations of gene expression, largely due to the technical barriers involved. However a lot of daily influences, like diet changes or the time of exposure to light on an organism are thought to alter gene expression only slightly ([13, 23, 29, 51]). Future experiment with the ability to detect changes smaller than two-fold, will shed light in this yet neglected field of science.

The resulting genes proved to be of good usability as reference genes.

Mostly, general direction (up- and down-regulation) and rank order of the fold-changes are similar, but the levels of the fold changes of microarray experiments differ compared to qPCR data ([83, 42, 86, 85]).

Still we think that our meta analysis approach needs amendment. Because of the typically non-gaussian distribution ([46]) of microarray data a non-parametric test should have been chosen, but was neglected because of the better usability and power of parametric tests (Claus Vogl and Heinrich Grausgruber, personal communication). Furthermore for practicability reasons we neglected the effect of the different microarray platforms on the resulting fluorescence. This could be achieved by calculating the genes best suited for normalisation in one type of array individually and combining weightedly, according to the number of positive probes involved or the experimental setup used. In future experiments this issue will be addressed. More sophisticated statistical data may be obtained with a meta-analysis using different statistical tests, like for example the Winer's combined test [35].

5 Material and Methods

5.1 Study design

5.1.1 Multi Generation Study (MGS)

Three feed groups (G, K and A) were analysed in a MGS design. For the MGS (Fig. 6A) pairs of mice were housed together in standard cages (marcolone type II, III and IV (Ehret, Emmendingen, Germany) cages together with their offspring. Pups were weaned after three weeks and male and female pups were kept separately till sexual maturation. After sexual maturation the offspring were arranged into new pairs randomly. From F1 onwards males were sampled after one week of mating. Feed was applied *ad libidum*.

Jejunum samples of male F3 mice were taken for this study.

5.1.2 Risk Assessment by Continuous Breeding (RACB)

Two repetitions of a RACB experiment were performed. In the RACB (Fig. 6B) pairs of mice were kept in standard cages (marcolone type II, III and IV) and allowed to breed. The pups were separated from the adults three weeks *post natum*. During the experimental period of 20 weeks, 4 litters/Pair were produced.

Parental mice and mice from the 3rd litter were sampled after the experiment. Two feed groups (G and K) were analysed in this study. Feed was applied *ad libidum*.

Jejunum samples and scrape samples of male and female parental mice, as well as jejunum samples of males of the 3rd litter were taken (see RNA isolation below).

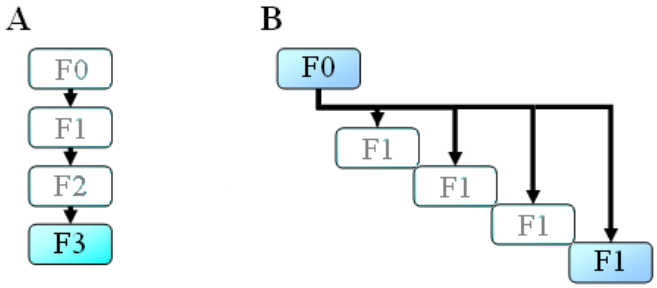

Figure 6: Breeding set up of MGS and RACB study designs. A: Set up of the MGS study design, showing the alternation of generations. B:In the RACB study one pair of mice is studied over the course of four litters.

5.2 Sample preparation

5.2.1 Mouse strains

The internal microarray set was done using OF1 mice. The outbred strains OF1 and Cd1 and the inbred strains CB6 Irf-/- and Rac2(c) stat1-/- were used for RT-qPCR experiments.

The mice were kept in marcolon cages of varying sizes adhering to the standards given by national ethics committee for animal experiments (GZ: 68.205/0042 – BrGT/2006). Mice were kept under a 12 hours light/dark regime and fed with a costume made diet containing 33% maize (Ssniff, Soest, Germany). Mon810/NK603 (feed G), the isogenic line NK603 (feed K) or the non-isogenic maize line SARASTRO (feed A) were used as maize for feed groups G, K or A respectively. The nutritional value of the feed was analysed by [9].

CD1, CB6 and Rac2(c) mice were fed a mixture of R/M-H Ered 1, M-Z, M-Z Extrudat and M-Z Ereich of varying composition (Ssniff).

5.2.2 Sample collection

Sampling was performed at day time in April and June 2008 for the MGS, in august 2008 for the parental mice of the RACB and in November and December 2006 for the 3rd litter in the RACB. Mice were fasted for 3 hours before applying vertebral dislocation.

The small intestine beginning from the pylorus to the ileocaecal junction was dissected immediately and divided into 2 fragments of similar length. In addition a 2 cm segment was dissected 1 cm proximal and 1 cm distal of the angular point representing the distal jejunum.

Tissue samples were immediately shock frozen in 2-methyl-butane (Merck, Darmstadt, Germany) cooled on liquid nitrogen. After shock freezing the samples were transferred and stored in the cryo tubes (Bertoni, Vienna, Austria) in liquid nitrogen until further analysis.

5.2.3 RNA isolation

50 mg intestinal tissue was homogenised in 700 μL Qiazol Lysis Reagent (Qiagen, Hilden, Germany) with 1.4 mm ceramic beads (MagNA Lyser Green Beats, Roche Diagnostics, Mannheim, Germany) in a MagNA Lyser (Roche Diagnostics) for 20 s. Lysate was stored at -80 °C.

QIAzol (Qiagen), using a monophasic solution of phenol and GTC (QIAzol lysis reagent), were used to separate the RNA from DNA and protein. The total RNA from murine jejunum samples was than isolated using the miRNeasy Kit (Qiagen) capable of recovering all cellular RNA including small RNA. The isolation was performed on the QIAcube (Qiagen) as described by the manufacturer.

RNA concentration was determined in a BioPhotometer (Eppendorf, Hamburg, Germany) using a Hellma TrayCell (Hellma, Müllheim, Germany) and yield ranged from 0.5 - 2 μg/μL. RNA integrity was determined by the use of the Agilent 2100 Bioanalyser (Agilent Technologies, Santa Clara, USA) using the RNA 6000 Nano Chip (Agilent Technologies) as described in the manual.
For dye-based RT-qPCR samples were treated with 1 U RQ1 DNase (Promega, Leiden, Germany) per 8 μl of RNA (125 μg/μL) to remove genomic DNA followed by spin-column based purification and concentrationing of RNA using the miRNeasy Kit (Qiagen). To check for remaining DNA contamination a minus reverse transcriptase negative control (-RT) using a equal mixture of all tested samples for each assay was analysed.

5.2.4 Integrity of RNA

To determine the integrity of RNA the RIN (RNA Integrity Number) value of each RNA sample was determined using the Agilent 2100 Bioanalyser (Agilent Technologies) in combination with RNA 6000 Nano chips (Agilent Technologies). A gel-dye mix was prepared at room temperature by adding 1 μl of well vortexed (10 s) RNA 6000 dye concentrate to 65 μl of gel, which was filtered prior using supplied spin filters. The gel-dye mix was again vortexed and centrifuged for 10 min at 13000 g. 9 μl were pumped into the chip using the supplied chip priming station. The wells were than filled with gel-dye mix or RNA 6000 Nano marker as indicated in the protocol. 1 μl of RNA was added to sampling wells and 1 μL of RNA 6000 Ladder was added to the appropriate well. After vortexing the chip for 60 s the chip was inserted into the 2100 Bioanalyser and the measurement was started using the 2100 Expert (Version B.02.03.SI307) software (Agilent Technologies).

Only RIN values of $>$8 were used for TLDAs and microarrays. For the RT-qPCR experiments a RIN $>$6 was mandatory.

5.2.5 Reverse transcription (RT)

The High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, USA) was used for cDNA synthesis from RNAs with acceptable RIN values of 7.8 to 9.6. The RNA was diluted in RNase-free water (Sigma-Aldrich, Vienna, Austria) to a concentration of 0.2 μg/μl each. The cDNA synthesis was done according to the manufacturers protocol. 2 μl 10 x RT Buffer, 25 μl dNTP mix (100 mM), 10 x RT Random Primers, 1 U Multiscribe Reverse Transciptase and 4.2 μl nuclease-free H_2O per reaction were mixed with 10 μl of the diluted RNA. The RTreaction was incubated at 25 °C for 10 min and 37 °C for 120 min. The reaction was stopped at 85 °C for 5 s. Sample cDNAs were run in duplicates. The minus RT controls were analyed in unicate. To examine the outcome of the RT a qPCR targeting

HPRT by dye based RT-qPCR.

5.3 Microarray analysis

DIG-labeled cDNA probes were reverse transcribed from 40 μg total RNA using the Chemiluminescent RT-Labelling kit (Applied Biosystems) as described by the manufacturer. Array hybridization, chemiluminescence detection, image acquisition and analysis were performed using the Chemiluminescence Detection Kit (Applied Biosystems) and the 1700 Chemiluminescence Microarray Analyzer (Applied Biosystems) following the manufacturer's instructions.

Each microarray was first pre-hybridized at 55 °C for 1 h in hybridization buffer with blocking reagent. Oligo-dT primed, DIG-labeled cDNA targets were fragmented, mixed with internal control target and then hybridized to the equilibrated microarrays in a volume of 1.5 ml at 55 °C for 16 h. After hybridisation, the arrays were washed with hybridization wash buffer and chemiluminescence rinse buffer. Enhanced chemiluminescent signals were generated by incubating arrays with alkaline phosphatase conjugated anti-digoxigenin antibody followed by incubation with chemiluminescence enhancing solution and a final addition of chemiluminescence substrate. Four images were collected for each microarray using the ABI 1700 Chemiluminescent Microarray Analyzer. Images were auto-gridded and the chemiluminescence signals were quantified, corrected for background and spot and spatially normalised.

5.4 Microarray processing

5.4.1 Meta-analysis of microarray expression data

Microarray experiments, 23 for the duodenum, 53 for the duodenum, 63 for the ileum as well as 92 unassigned small intestinal samples, containing a total of 220 single microarrays [6, 78, 61, 30, 19, 56, 77, 34, 60, 36, 48, 66, 87, 40, 47, 38, 24, 14, 88] (Tab. 3). Arrays of nine different platforms were selected and downloaded form the Gene Expression Omnibus (GEO) [21] and Array Express [52]. Either the raw data or, if the raw data was not available, MAS 5.0 scaled data (defined as the anti-log of a robust average (Turkey biweight) of the values $\log(PM_{ij} - CT_{ij})$ was taken and used for further analysis. Data had to be well defined to be included in this study to avert the influence of possible pre-normalised data on the outcome. Furthermore solely single-dye chips were meta analysed. No data derived from cell cultures were used. Only probes with a given NCBI gene symbol were considered. Quality-weighting as proposed by [33] was not considered, as the detection system generated detection p-value needed for quality-weighting was not included in all researched microarray datasets.

First each individual microarray experiment was normalised by quantile normalisation us-

ing GeneStat software (11th Edition for Windows; VSN International Ltd, Hemel Hempstead, England). During quantile normalisation each array is ranked by the fluorescent intensity level. Then the arithmetic mean of each rank across the arrays is calculated and the values of each rank on each microarray are replaced by the ranks mean. After quantile normalisation the data was log2 transformed. This was done to eliminate interassay differences caused by systematic errors. After normalisation spots lacking a gene symbol were removed and genes being detected by more than one set of probes were averaged. Subsequently the individual numbers were standardised by the formula

$$X_{\mathrm{i}\to\mathrm{j}} = \frac{X_{i\to j} - X_{avg}}{X_{SD}}$$

where X ... fluroescence intensity signal

After normalising and standardising the data the coefficient of variation in percent (CV%) of each gene across all arrays was calculated. The 10% genes with the lowest CV% were selected for further analysis. A parametric Chi^2 test was performed, using SAS 9.1 software (SAS Institute, Cary, USA). The resulting p values were used to identify reference gene candidates. The p values were not used to determine statistical significance, but rather as indicators for a uniform gene's expression across all arrays being consistent with a high p value.

5.4.2 Overrepresentation analysis

Overrepresentation studies are designed to identify common biological pathways or functions among a set of genes altered in a study. Many overrepresentation analysis platforms are available online or as downloadable tools, like GoSurfer [91], Go-Bayes [89], BiNGO [44], eGOn [7], GOrilla [20], GOEAST [90] and many more. We used the PANTHER database (Version 7 beta) [74] for this analysis. A plain text file containing the gene symbols found to be deregulated in the analysed microarray experiment, was up-loaded to the data base and analysed using the Compare Classifications of Lists tool. The Mouse AB 1700 genes list was used as a reference list. Both pathways and biological functions were analysed. No Bonferroni correction was used.

5.5 RT-qPCR

5.5.1 Low density profiling of mRNA

The custom made TaqMan low density arrays (configuration 4) (Applied Biosystems) were used according to the manual. In brief, 100 μl of a mix of cDNA and TaqMan Universal PCR Master Mix (2X) (Applied Biosystems), containing a total of 250 ng cDNA, was loaded into each fill reservoir. After centrifugation in the Heraeus Multifuge 3S-R Centrifuge (Thermo Fischer Scientific, Vienna, austria) (2 x 1 min at 311 g) the microfluic card was closed and qPCR was started immediately

on the ABI Prism HT 7900 Sequence detection system (Applied Biosystems). SDS 2.3 software (Applied Biosystems) was used for analysis.

It has to be noted that the *Gpr39* assay did not yield a amplification signal (C_q>45) in all experiments and was turned down in all following analyses.

The samples were run in duplicate or unicate (see supplementariy tab. 9).

5.5.2 Dye based assays

All RT-qPCR reactions included 3.5 mM $MgCl_2$ (Solis Biodyne, Tartu, Estonia), 200 nM of each Primer, 0.2 mM of each dNTP (Solis Biodyne) , 0.4 x EvaGreen (Biotium, Hayward, USA), 1 x Buffer BD (0.8 M Tris-HCl, 0.2 M $(NH_4)_2SO_4$) (Solis Biodyne) and 1 Unit of HotFire polymerase (Solis Biodyne). qPCR was performed in a 10 μl Volume and performed on the LightCycler 480 (Roche Diagnostics) in 384-well setup. The cycling conditions consisted of a hot start phase of 15 min at 95 °C to activate the HotFire polymerase, followed by 40 cycles of 95 °C for 15 s , 60 °C for 40 s and 72 °C for 20 s.
Intron-spanning or flanking primer sequences available in the Harvard PrimerBank [70, 71, 82] or RTPrimerDB [54, 53, 41] were used. Only primer pairs spanning introns larger than 500 bp and produced an amplicon shorter than 200 bp were used. (Tab. 7). If a gene was not covered by a publicly available primer pair, primers were designed using Primer Express 2.0 software (Applied Biosystems) on basis of the current reference sequences derived from the NCBI-homepage. All primers were designed to anneal at 60 °C ±2 °C. The primers were tested for primer dimerisation using PerlPrimer 1.1.14 software (Marshall 2004) (Tab. 7). All primers were synthesised by Sigma-Aldrich.

Primers were tested for amplification efficiency using an equal mixture of all tested samples. To determine the efficiency a dilution series was used. To achieve C_qs below 40 a dilution series of 1:1, 1:5, 1:10, 1:20 and 1:30 was used. All Primers were tested using the LightCycler 480 (Roche Diagnostics) and analysed by the LightCycler 480 software 1.5.0 using the second derivative to determine the C_q. As given by the LightCycler 480 manual an error of below 0.2 was considered necessary. However in case of the EG666853 an error of 0.241 was accepted because of the late C_q.

RT-qPCR assays for the testing of reference genes was pipeted by the liquid handling system EpMotion 5075 TMX (Eppendorf).

For RT-qPCR assays MIQE guidelines [11] were followed as close as possible I feel compelled to advert to the declared conflicts of interest in the respective publication.

Table 7: PCR primer pairs designed and/or used in this work. The error value is calculated by the LightCycler 480 software, a value below 0.2 is considered acceptable.

Symbols of Target gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Amplicon length	Efficency	Error[1]	Tm[2] (°C) (±0.3 °C)	Assay ID
1200016E24Rik	TGGAGTAGGCTGGCCACCTA	CAGCCAAGCCAAACGCTATT	131	0.786	0.0526	84.5	
2310046K01Rik	ACCAGGGACCACCTTGAACACTA	CAGCAGCACGTCAGGAAAGAC	159	0.756	0.0228	88.8	
64030706D22Rik	CGGTGCGGATACGAGCTT	CTCTGCGTTCTTAAAATACTCTTCATCT	187	0.842	0.0080	86.4	
AI314976	GGCGTCACTGACCTCTCCAT	TTGACTCAAACACCTCTTCGAGAA	97	0.766	0.1340	83.9	
Atp6v0d1	ATTTCAACGTGGACAATGGCTACTT	TCTCGCACTGCACCAGGTT	99	0.883	0.0194	87.7	
B3gnt3	CCCCTACGAGATCCTCCTCC	CACGGATAGATTAGCTCGGCA	181	0.849	0.0269	91.7	31982625a1[3]
EG666853	GGACGACGAGGACATGACACTT	GCTCTCGGATCCACTCGTG	72	0.781	0.2410	81.2	
Fbln1	ACCAGGCCGACATCATCTTC	TTCAGGACAGCGTAAAACGGG	142	0.68	0.0184	87.3	6753822a2[3]
Hadhb	GCTAGAGCTGCACTTTCGGG	GCCACATTGCTTGTTTTCACTTC	110	0.847	0.0222	79.8	21704100a2[3]
Hprt1	AGCTACTGTAATGATCAGTCAACG	AGAGGTCCTTTTCACCAGCA	198	0.816	0.0051	81.3	3583[4]
LOC630729	TGACAACATCCCTACTGTGGTCTT	GGGTAAAGGCAGTCGAGTAGATTTTC	122	0.876	0.0038	83.6	
Plekha7	GAACGCTTCCGTGCTCAAAG	GATCGACAAACAAGGCATCCT	103	0.857	0.0008	86.3	27370088a2[3]
Tmem14c	CTGTCTCAGGATCCCAGGAATGT	CAAACTGGCTCCTGCGATTAA	129	0.833	0.0465	83.8	
Tspan15	CGGCAGAAATACAAAACCCTGG	CAGAAGGCACAGGTTGTCC	132	0.948	0.0058	87.9	13278031a2[3]
Zfp598	GGTGCTCTACCAAGATGCGG	GGATGGGGATCAGGGCAAA	117	0.871	0.0055	89.4	34147169a1[3]
Zfyve19	GCTGTCAAGTTCACCCTCTTCAA	GGCTCGGCCACAATTCTTACA	57	0.897	0.0538	78.7	

[1] Mean squared error of the single data points fit to the regression line

[2] Experimental amplicon Tm

[3] Primer Bank

[4] Rtprimer DB

Acknowledgments

First and foremost I am deeply thankful to Ao. Univ. Prof. Dr. Ralf Steinborn for his excellent help, whose knowledge and supervision guided me through this work. Without his dedication for science and teaching, this work could neither be started nor finished. I further want to thank Univ. Prof. Dr. Thomas Decker for his assistance to make this work possible. I am also grateful to Univ. Prof. Dr. Jürgen Zentek for supplying samples and financing this work. I further want to thank Univ.Ass. Dr. Heinrich Grausgruber and Dr. Claus Vogl for their helpful support regarding statistical questions. Next I want to credit is Dr. Claudia Binter for her support, especially in the beginning of this diploma work. I also want to thank the whole Vetomics-team for their great help and friendship during the course of this work. Last but not least I want to thank my girlfriend Katerina Sitner and my friends for their love, friendship and support.

In loving memory of Rudolf Meyer.

6 References

References

[1] L V Abruzzo, K Y Lee, A Fuller, A Silverman, M J Keating, L J Medeiros, and K R Coombes. Validation of oligonucleotide microarray data using microfluidic low-density arrays: a new statistical method to normalize real-time rt-pcr data. *Biotechniques*, 38(5):785–92, 2005.

[2] S Akilesh, D J Shaffer, and D Roopenian. Customized molecular phenotyping by quantitative gene expression and pattern recognition analysis. *Genome Research*, 13(7):1719–27, 2003.

[3] C L Andersen, J L Jensen, and T F Orntoft. Normalization of real-time quantitative reverse transcription-pcr data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res*, 64(15):5245–50, 2004.

[4] AppliedBiosystems. Amplification efficiency of taqman gene expression assay, 2006.

[5] T Barrett, D B Troup, S E Wilhite, P Ledoux, D Rudnev, C Evangelista, I F Kim, A Soboleva, M Tomashevsky, K A Marshall, K H Phillippy, P M Sherman, R N Muertter, and R Edgar. Ncbi geo: archive for high-throughput functional genomic data. *Nucleic Acids Research*, 37(Database issue):D885–90, 2009.

[6] M A Battle, B J Bondow, M A Iverson, S J Adams, R J Jandacek, P Tso, and S A Duncan. Gata4 is essential for jejunal function in mice. *Gastroenterology*, 135(5):1676–1686.e1, 2008.

[7] V Beisvag, F K R JÃŒnge, H Bergum, L Jolsum, S Lydersen, C GÃŒnther, H Ramampiaro, M Langaas, A K Sandvik, and A Laegreid. Genetools–application for functional annotation and statistical hypothesis testing. *BMC Bioinformatics*, 7:470, 2006.

[8] F Berge, B De Hertogh, M Pierre, E Bareke, A Gaigneaux, and E Depiereux. Phoenix, a web interface for (re)analysis of microarray data. *Central European Journal of Biology*, 4:603–618, 2009.

[9] C Binter, A Velimirov, and J Zenteck. Biological effects of transgenic maize nk603 x mon 810 fed in long term reproduction studies in mice. 2008.

[10] A Bisognin, A Coppe, F Ferrari, D Risso, C Romualdi, S Bicciato, and S Bortoluzzi. A-madman: annotation-based microarray data meta-analysis tool. *BMC Bioinformatics*, 10:201, 2009.

[11] S A Bustin, V Benes, J A Garson, J Hellemans, J Huggett, M Kubista, R Mueller, T Nolan, M W Pfaffl, G L Shipley, J Vandesompele, and C T Wittwer. The miqe guidelines: minimum information for publication of quantitative real-time pcr experiments. *Clinical Chemistry*, 55(4):611–22, 2009.

[12] S A Bustin, V Benes, T Nolan, and M W Pfaffl. Quantitative real-time rt-pcr - a perspective. *J Mol Endocrinol*, 34(3):597–601, 2005.

[13] C Cailotto, J Lei, J van der Vliet, C van Heijningen, C G van Eden, A Kalsbeek, P Pevet, and R M Buijs. Effects of nocturnal light on (clock) gene expression in peripheral organs: a role for the autonomic innervation of the liver. *PLoS One*, 4(5):e5650, 2009.

[14] H Coppin, V Darnaud, L Kautz, D Meynard, M Aubry, J Mosser, M Martinez, and M Roth. Gene expression profiling of hfe-/- liver and duodenum in mouse strains with differing susceptibilities to iron loading: identification of transcriptional regulatory targets of hfe and potential hemochromatosis modifiers. *Genome Biol*, 8(10):R221, 2007.

[15] T Czechowski, M Stitt, T Altmann, M K Udvardi, and W Scheible. Genome-wide identification and testing of superior reference genes for transcript normalization in arabidopsis. *Plant Physiology*, 139(1):5–17, 2005.

[16] P B Dallas, N G Gottardo, M J Firth, A H Beesley, K Hoffmann, P A Terry, J R Freitas, J M Boag, A J Cummings, and U R Kees. Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time rt-pcr – how well do they correlate? *BMC Genomics*, 6(1):59, 2005.

[17] H J M de Jonge, R S N Fehrmann, E S J M de Bont, R M W Hofstra, F Gerbens, W A Kamps, E G E de Vries, A G J van der Zee, G J te Meerman, and A ter Elst. Evidence based selection of housekeeping genes. *PLoS ONE*, 2(9):e898, 2007.

[18] C De Santis, C Smith-Keune, and D R Jerry. Normalizing rt-qpcr data: Are we getting the right answers? an appraisal of normalization approaches and internal reference genes from a case study in the finfish lates calcarifer. *Mar Biotechnol (NY)*, 2010.

[19] N J de Wit, H Bosch-Vermeulen, P J de Groot, G J Hooiveld, M M G Bromhaar, J Jansen, M Muller, and R van der Meer. The role of the small intestine in the development of dietary fat-induced obesity and insulin resistance in c57bl/6j mice. *BMC Med Genomics*, 1:14, 2008.

[20] E Eden, R Navon, I Steinfeld, D Lipson, and Z Yakhini. Gorilla: a tool for discovery and visualization of enriched go terms in ranked gene lists. *BMC Bioinformatics*, 10:48, 2009.

[21] R Edgar, M Domrachev, and A E Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res*, 30(1):207–10, 2002.

[22] M Exposito-Rodriguez, A A Borges, A Borges-Perez, and J A Perez. Selection of internal control genes for quantitative real-time rt-pcr studies during tomato development process. *BMC Plant Biol*, 8:131, 2008.

[23] R Feil. Environmental and nutritional effects on the epigenetic regulation of genes. *Mutat Res*, 600(1-2):46–57, 2006.

[24] S Freilich, T Massingham, S Bhattacharyya, H Ponsting, P A Lyons, T C Freeman, and J M Thornton. Relationship between the tissue-specificity of mouse gene expression and the evolutionary origin and function of the proteins. *Genome Biol*, 6(7):R56, 2005.

[25] M Frericks and C Esser. A toolbox of novel murine house-keeping genes identified by meta-analysis of large scale gene expression profiles. *Biochim Biophys Acta*, 2008.

[26] Z Fu, X Chen, V Vacic, P Nan, Y Zhong, and T Jiang. Msoar: a high-throughput ortholog assignment system based on genome rearrangement. *J Comput Biol*, 14(9):1160–75, 2007.

[27] S Guenin, M Mauriat, J Pelloux, O Van Wuytswinkel, C Bellini, and L Gutierrez. Normalization of qrt-pcr data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *J Exp Bot*, 60(2):487–93, 2009.

[28] F Haller, B Kulle, S Schwager, B Gunawan, A von Heydebreck, H Sultmann, and L Fuzesi. Equivalence test in quantitative reverse transcription polymerase chain reaction: confirmation of reference genes suitable for normalization. *Anal Biochem*, 335(1):1–9, 2004.

[29] J Hannibal, F Jamen, H S Nielsen, L Journot, P Brabet, and J Fahrenkrug. Dissociation between light-induced phase shift of the circadian rhythm and clock gene expression in mice lacking the pituitary adenylate cyclase activating polypeptidadenylate cyclase activating polypeptide type 1 receptor. *J Neurosci*, 21(13):4883–90, 2001.

[30] X C He, T Yin, J C Grindley, Q Tian, T Sato, W A Tao, R Dirisina, K S Porter-Westpfahl, M Hembree, T Johnson, L M Wiedemann, T A Barrett, L Hood, H Wu, and L Li. Pten-deficient intestinal stem cells initiate intestinal polyposis. *Nat Genet*, 39(2):189–98, 2007.

[31] F Hong and R Breitling. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics*, 24(3):374–82, 2008.

[32] T Hruz, O Laule, G Szabo, F Wessendorp, S Bleuler, L Oertle, P Widmayer, W Gruissem, and P Zimmermann. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics*, 2008:420747, 2008.

[33] P Hu, C M Greenwood, and J Beyene. Statistical methods for meta-analysis of microarray data: A comparative study. *Information Systems Frontiers*, 8(1):9–20, 2006.

[34] J J Hutton, A G Jegga, S Kong, A Gupta, C Ebert, S Williams, J D Katz, and B J Aronow. Microarray and comparative genomics-based identification of genes and gene regulatory regions of the mouse immune system. *BMC Genomics*, 5(1):82, 2004.

[35] J L Johnson, S L Wolf, and U M Kabadi. Efficacy of insulin and sulfonylurea combination therapy in type ii diabetes. a meta-analysis of the randomized placebo-controlled trials. *Arch Intern Med*, 156(3):259–64, 1996.

[36] H Kondo, Y Minegishi, Y Komine, T Mori, I Matsumoto, K Abe, I Tokimitsu, T Hase, and T Murase. Differential regulation of intestinal lipid metabolism-related genes in obesity-resistant a/j vs. obesity-prone c57bl/6j mice. *Am J Physiol Endocrinol Metab*, 291(5):E1092–9, 2006.

[37] X Kong, V Mas, and K J Archer. A non-parametric meta-analysis approach for combining independent microarray datasets: application using two microarray datasets pertaining to chronic allograft nephropathy. *BMC Genomics*, 9:98, 2008.

[38] M H Kucherlapati, K Yang, K Fan, M Kuraguchi, D Sonkin, A Rosulek, M Lipkin, R T Bronson, B J Aronow, and R Kucherlapati. Loss of rb1 in the gastrointestinal tract of apc1638n mice promotes tumors of the cecum and proximal colon. *Proc Natl Acad Sci U S A*, 105(40):15493–8, 2008.

[39] M J Kwon, E Oh, S Lee, M R Roh, S E Kim, Y Lee, Y Choi, Y In, T Park, S S Koh, and Y K Shin. Identification of novel reference genes using multiplatform expression data and their validation for quantitative gene expression analysis. *PLoS One*, 4(7):e6162, 2009.

[40] M Lecuit, J L Sonnenburg, P Cossart, and J I Gordon. Functional genomic studies of the intestinal response to a foodborne enteropathogen in a humanized gnotobiotic mouse model. *J Biol Chem*, 282(20):15065–72, 2007.

[41] S Lefever, J Vandesompele, F Speleman, and F Pattyn. Rtprimerdb: the portal for real-time pcr primers and probes. *Nucleic Acids Res*, 37(Database issue):D942–5, 2009.

[42] R W Li and G C Waldbieser. Production and utilization of a high-density oligonucleotide microarray in channel catfish, ictalurus punctatus. *BMC Genomics*, 7:134, 2006.

[43] L J Maccoux, D N Clements, F Salway, and P J R Day. Identification of new reference genes for the normalisation of canine osteoarthritic joint tissue transcripts from microarray data. *BMC Mol Biol*, 8:62, 2007.

[44] S Maere, K Heymans, and M Kuiper. Bingo: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 21(16):3448–9, 2005.

[45] J J McCarthy, J L Andrews, E L McDearmon, K S Campbell, B K Barber, B H Miller, J R Walker, J B Hogenesch, J S Takahashi, and K A Esser. Identification of the circadian transcriptome in adult mouse skeletal muscle. *Physiol Genomics*, 31(1):86–95, 2007.

[46] E S Motakis, G P Nason, P Fryzlewicz, and G A Rutter. Variance stabilization and normalization for one-color microarray data using a data-driven multiscale approach. *Bioinformatics*, 22(20):2547–53, 2006.

[47] D M Mutch, P Anderle, M Fiaux, R Mansourian, K Vidal, W Wahli, G Williamson, and M Roberts. Regional variations in abc transporter expression along the mouse intestinal tract. *Physiol Genomics*, 17(1):11–20, 2004.

[48] D M Mutch, V Crespy, J Clough, C J Henderson, S Lariani, R Mansourian, J Moulin, C R Wolf, and G Williamson. Hepatic cytochrome p-450 reductase-null mice show reduced transcriptional response to quercetin and reveal physiological homeostasis between jejunum and liver. *Am J Physiol Gastrointest Liver Physiol*, 291(1):G63–72, 2006.

[49] K Oishi, N Amagai, H Shirai, K Kadota, N Ohkura, and N Ishida. Genome-wide expression analysis reveals 100 adrenal gland-dependent circadian genes in the mouse liver. *DNA Res*, 12(3):191–202, 2005.

[50] S Panda, M P Antoch, B H Miller, A I Su, A B Schook, M Straume, P G Schultz, S A Kay, J S Takahashi, and J B Hogenesch. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*, 109(3):307–20, 2002.

[51] Wenqiang Pang, Chunying Li, Yue Zhao, Shiming Wang, Wei Dong, Pengjiu Jiang, and Jianfa Zhang. The environmental light influences the circulatory levels of retinoic acid and associates with hepatic lipid metabolism. *Endocrinology*, 149(12):6336–42, Dec 2008.

[52] H Parkinson, M Kapushesky, N Kolesnikov, G Rustici, M Shojatalab, N Abeygunawardena, H Berube, M Dylag, I Emam, A Farne, E Holloway, M Lukk, J Malone, R Mani, E Pilicheva, T F Rayner, F Rezwan, A Sharma, E Williams, X Bradley, T Adamusiak, M Brandizi, T Burdett, R Coulson, M Krestyaninova, P Kurnosov, E Maguire, S G Neogi, P Rocca-Serra, S Sansone, N Sklyar, M Zhao, U Sarkans, and A Brazma. Arrayexpress update–from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res*, 37(Database issue):D868–72, 2009.

[53] F Pattyn, P Robbrecht, A De Paepe, F Speleman, and J Vandesompele. Rtprimerdb: the real-time pcr primer and probe database, major update 2006. *Nucleic Acids Res*, 34(Database issue):D684–8, 2006.

[54] F Pattyn, F Speleman, A De Paepe, and J Vandesompele. Rtprimerdb: the real-time pcr primer and probe database. *Nucleic Acids Res*, 31(1):122–3, 2003.

[55] P Pedotti, P A C 't Hoen, E Vreugdenhil, G J Schenk, R H Vossen, Y Ariyurek, M de Hollander, R Kuiper, G J B van Ommen, J T den Dunnen, J M Boer, and R X de Menezes. Can subtle changes in gene expression be consistently detected with different microarray platforms? *BMC Genomics*, 9:124, 2008.

[56] D A Peterson, N P McNulty, J L Guruge, and J I Gordon. Iga response to symbiotic bacteria as a mediator of gut homeostasis. *Cell Host Microbe*, 2(5):328–39, 2007.

[57] M W Pfaffl, G W Horgan, and L Dempfle. Relative expression software tool (rest) for group-wise comparison and statistical analysis of relative expression results in real-time pcr. *Nucleic Acids Res*, 30(9):e36, 2002.

[58] M W Pfaffl, A Tichopad, C Prgomet, and T P Neuvians. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: Bestkeeper–excel-based tool using pair-wise correlations. *Biotechnol Lett*, 26(6):509–15, 2004.

[59] D Ramskold, E T Wang, C B Burge, and R Sandberg. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol*, 5(12):e1000598, Dec 2009.

[60] T Ravasi, H Suzuki, K C Pang, S Katayama, M Furuno, R Okunishi, S Fukuda, K Ru, M C Frith, M M Gongora, S M Grimmond, D A Hume, Y Hayashizaki, and J S Mattick. Experimental validation of the regulated expression of large numbers of non-coding rnas from the mouse genome. *Genome Res*, 16(1):11–9, 2006.

[61] J F Rawls, M A Mahowald, R E Ley, and J I Gordon. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell*, 127(2):423–33, 2006.

[62] L Robb, R Stephens, and J Coleman. Application of survival and multivariate methods to gene expression data combined from two sources.

[63] A Sadki, M Bentivoglio, K Kristensson, and M Nygard. Suppressors, receptors and effects of cytokines on the aging mouse biological clock. *Neurobiol Aging*, 28(2):296–305, 2007.

[64] S Saviozzi, F Cordero, M Lo Iacono, S Novello, G V Scagliotti, and R A Calogero. Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. *BMC Cancer*, 6:200, 2006.

[65] L A Scheving and W E Russell. It's about time: clock genes unveiled in the gut. *Gastroenterology*, 133(4):1373–6, 2007.

[66] F A Simmen, R Xiao, M C Velarde, R D Nicholson, M T Bowman, Y Fujii-Kuriyama, S P Oh, and R C M Simmen. Dysregulation of intestinal crypt cell proliferation and villus cell migration in mice lacking kruppel-like factor 9. *Am J Physiol Gastrointest Liver Physiol*, 292(6):G1757–69, 2007.

[67] D A Skelly, J Ronald, and J M Akey. Inherited variation in gene expression. *Annu Rev Genomics Hum Genet*, 10:313–32, 2009.

[68] K Skovgaard, S Mortensen, K T Poulsen, O Angen, and P M H Heegaard. Validation of putative reference genes for qrt-pcr normalization in tissues and blood from pigs infected with actinobacillus pleuropneumoniae. *Vet Immunol Immunopathol*, 118(1-2):140–6, 2007.

[69] E Spanakis. Problems related to the interpretation of autoradiographic data on gene expression using common constitutive transcripts as controls. *Nucleic Acids Res*, 21(16):3809–19, 1993.

[70] A Spandidos, X Wang, H Wang, S Dragnev, T Thurber, and B Seed. A comprehensive collection of experimentally validated primers for polymerase chain reaction quantitation of murine transcript abundance. *BMC Genomics*, 9:633, 2008.

[71] A Spandidos, X Wang, H Wang, and B Seed. Primerbank: a resource of human and mouse pcr primer pairs for gene expression detection and quantification. *Nucleic Acids Res*, 38(Database issue):D792–9, 2010.

[72] P Stafford and M Brun. Three methods for optimization of cross-laboratory and cross-platform microarray expression data. *Nucleic Acids Res*, 35(10):e72, 2007.

[73] K Storch, O Lipan, I Leykin, N Viswanathan, FC Davis, W H Wong, and C J Weitz. Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417(6884):78–83, 2002.

[74] P D Thomas, A Kejariwal, M J Campbell, H Mi, K Diemer, N Guo, I Ladunga, B Ulitsky-Lazareva, A Muruganujan, S Rabkin, J A Vandergriff, and O Doremieux. Panther: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res*, 31(1):334–41, 2003.

[75] L Thorrez, K Van Deun, L Tranchevent, L Van Lommel, K Engelen, K Marchal, Y Moreau, I Van Mechelen, and F Schuit. Using ribosomal protein genes as reference: a tale of caution. *PLoS ONE*, 3(3):e1854, 2008.

[76] V Valente, S A Teixeira, L Neder, O K Okamoto, S M Oba-Shinjo, S Kn Marie, C A Scrideli, M L Paco-Larson, and C G Jr Carlotti. Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative rt-pcr. *BMC Mol Biol*, 10(1):17, 2009.

[77] F Van Dyck, C V Braem, Z Chen, J Declercq, R Deckers, B Kim, S Ito, M K Wu, D E Cohen, M Dewerchin, R Derua, E Waelkens, L Fiette, A Roebroek, F Schuit, W J M Van de Ven, and Ramesh A Shivdasani. Loss of the plagl2 transcription factor affects lacteal uptake of chylomicrons. *Cell Metab*, 6(5):406–13, 2007.

[78] E M van Schothorst, P Flachs, N L W Franssen-van Hal, O Kuda, A Bunschoten, J Molthoff, C Vink, GJ E J Hooiveld, J Kopecky, and J Keijer. Induction of lipid oxidation by polyunsaturated fatty acids of marine origin in small intestine of mice fed a high-fat diet. *BMC Genomics*, 10:110, 2009.

[79] J Vandesompele, K De Preter, F Pattyn, B Poppe, N Van Roy, A De Paepe, and F Speleman. Accurate normalization of real-time quantitative rt-pcr data by geometric averaging of multiple internal control genes. *Genome Biology*, 3(7):research0034.1–research0034.11, 2002.

[80] P Varju, K C Chang, E Hrabovszky, I Merchenthaler, and Z Liposits. Temporal profile of estrogen-dependent gene expression in lhrh-producing gt1-7 cells. *Neurochem Int*, 54(2):119–34, 2009.

[81] R I Vazquez-Padron, J Gonzales-Cabrera, C Garcia-Tovar, L Neri-Bazan, R Lopez-Revilla, M Hernandez, L Moreno-Fierro, and G A de la Riva. Cry1ac protoxin from bacillus thuringiensis sp. kurstaki hd73 binds to surface proteins in the mouse small intestine. *Biochem Biophys Res Commun*, 271(1):54–8, 2000.

[82] X Wang and B Seed. A pcr primer bank for quantitative gene expression analysis. *Nucleic Acids Res*, 31(24):e154, 2003.

[83] S Waxman and E Wurmbach. De-regulation of common housekeeping genes in hepatocellular carcinoma. *BMC Genomics*, 8(1):243, 2007.

[84] H Willenbrock, J Salomon, R Sokilde, K B Barken, T N Hansen, F C Nielsen, S Moller, and T Litman. Quantitative mirna expression analysis: comparing microarrays with next-generation sequencing. *RNA*, 15(11):2028–34, 2009.

[85] C Wu, A Budhu, S Chen, X Zhou, N C Popescu, K Valerie, and X W Wang. Effect of hepatitis c virus core protein on the molecular profiling of human b lymphocytes. *Mol Med*, 12(1-3):47–53, 2006.

[86] E Wurmbach, T Yuen, B J Ebersole, and S C Sealfon. Gonadotropin-releasing hormone receptor-coupled gene network organization. *J Biol Chem*, 276(50):47195–201, 2001.

[87] S Yu, S Singh, H Chen, H Chen, J J W Chen, W Chen, H Chen, and S Chen. Intra-abdominal adhesion formation induces anti-oxidative injury, enhances cell proliferation, and prevents complement-mediated lysis. *Wound Repair Regen*, 16(3):388–98, 2008.

[88] J Zhang, A Moseley, A G Jegga, A Gupta, D P Witte, M Sartor, M Medvedovic, S S Williams, C Ley-Ebert, L M Coolen, G Egnaczyk, M B Genter, M Lehman, J Lingrel, J Maggio, L Parysek, R Walsh, M Xu, and Bruce J Aronow. Neural system-enriched gene expression: relationship to biological pathways and neurological diseases. *Physiol Genomics*, 18(2):167–83, 2004.

[89] S Zhang, J Cao, Y M Kong, and R H Scheuermann. Go-bayes: Gene ontology-based overrepresentation analysis using a bayesian approach. *Bioinformatics*, 26(7):905–11, 2010.

[90] Q Zheng and X Wang. Goeast: a web-based software toolkit for gene ontology enrichment analysis. *Nucleic Acids Res*, 36(Web Server issue):W358–63, 2008.

[91] S Zhong, L Tian, C Li, K Storch, and W H Wong. Comparative analysis of gene sets in the gene ontology space under the multiple hypothesis testing framework. *Proc IEEE Comput Syst Bioinform Conf*, pages 425–35, 2004.

[92] L Zhou, Q Lim, G Wan, and H Too. Normalization with genes encoding ribosomal proteins but not gapdh provides an accurate quantification of gene expressions in neuronal differentiation of pc12 cells. *BMC Genomics*, 11:75, 2010.

7 Supplements

7.1 Zusammenfassung (dt.)

Selbst verhältnismäßig kleine Veränderungen der Genexpression können signifikante biologische Effekte haben. Um solche kleinen Veränderungen quantifizieren zu können, muß der Goldstandard für Expressionsquantifizierung, die Reverse Transcription Quantitative Real-Time PCR (RT-qPCR), mit einer passenden Normalisierungsstrategie verbunden werden. Der erfolgreiche Einsatz dieser Technik benötigt eine sorgfältige Auswahl von Referenzgenen für die Normalisierung. Gewebeübergreifende Expressionsanalysen benötigen Gene die in allen Geweben gleichmäßig stabil exprimiert werden, wodurch nur eine geringe Anzahl an einigermaßen stabilen Genen zur Verfügung steht. Wenn jedoch nur ein bestimmtes Gewebe untersucht wird, dürfte die Anzahl der möglichen Referenzgene jedoch steigen.

In dieser Arbeit wurde nach Referenzgenen gesucht, die speziell auf die einzelnen Abschnitte des Dünndarms von Mäusen abgestimmt sind. Kandidatengene wurden durch eine Metaanalyse ermittelt. Für die Metaanalyse wurden Daten von internen und öffentlich zugänglichen Micoarrays verwendet. Jejunum Proben eines Auszuchtstammes und drei Inzuchtstämmen wurden verwendet um 15 Kandidatengene mittles RT-qPCR zu validieren. Die Gene *Plekha7*, *6430706D22Rik*, *EG666853* und *Zfyve19* wurden als diejenigen identifiziert die als interne Kontrollen am geeignetsten sind, da ihr Expression nur um $<$0.8-fach schwankten. Ihre Expressionstabilität liegt im selben Bereich wie jene von *Oaz1*, übertrifft jedoch die der kürzlich für gewebeübergreifene Normalisierung eingeführten Retrotransposons B1 Element und B2 Element. Der Expressions Stabilitätswert, errechnet durch die Software GeNorm, übertrifft die Werte früherer gewebe- und plattformübergreifenden Metaanalysen. Die hohe Stabilität der neu gefunden Jejunum-Referenzgene legt nahe, daß sie in diesem Gewebe, in stark regulierten Stoffwechselwegen involviert sind. Die funktionelle Diversivität der neu gefundenen Referenzgene erlaubt es Zielgene unter einer Vielzahl biologischer Konditionen und Stimuli zu untersuchen. Zusätzlich wurde dieser Metaanalyseansatz noch auf die anderen Untereinheiten des murinen Dünndarms und auf den gesamten Dünndarm angewendet.

Die Verbesserung der Genauigkeit der Normalisierung durch eine Metaanalyse von Expressionsdaten eines spezifischen Gewebes unterstreicht die Möglichkeiten dieser Strategie in anderen Geweben. Durch ihre, impliziert duch ihre gleichförmige Expression, biologische Bedeutung sind die neu gefundenen Normalisierungsgene auch lohnende Ziele für weitere Untersuchungen.

7.2 Curriculum Vitae

First- and Surename:

Florian Ruldof Lukas Meyer

Date of birth :

05.04.1982 Wien, Austria

Citizenship:

Austria

Education:

1988-1992 Volksschule der Albertus Magnus Schule

1992-1997 Realgymnasium der Albertus Magnus Schule

1997-2000 Realgymnasium Marianum der Schulbrüder - Abschluss mit Matura

March 2001-Mai 2004 academic studies of human medicine at the University of Vienna

March 2001-November 2004 academic studies of biology at the University of Vienna (1.Abschnitt)

Since November 2004 academic studies of microbiology and genetics at the University of Vienna (2.Abschnitt)

Since Ferbruary 2008 Diplomathesis at Vetomics-Corefacility for research (University for Veterinary Medicine Vienna)

National service:

Juli 2000-February 2001

Languages:

German

English

7.3 Supplementary materials

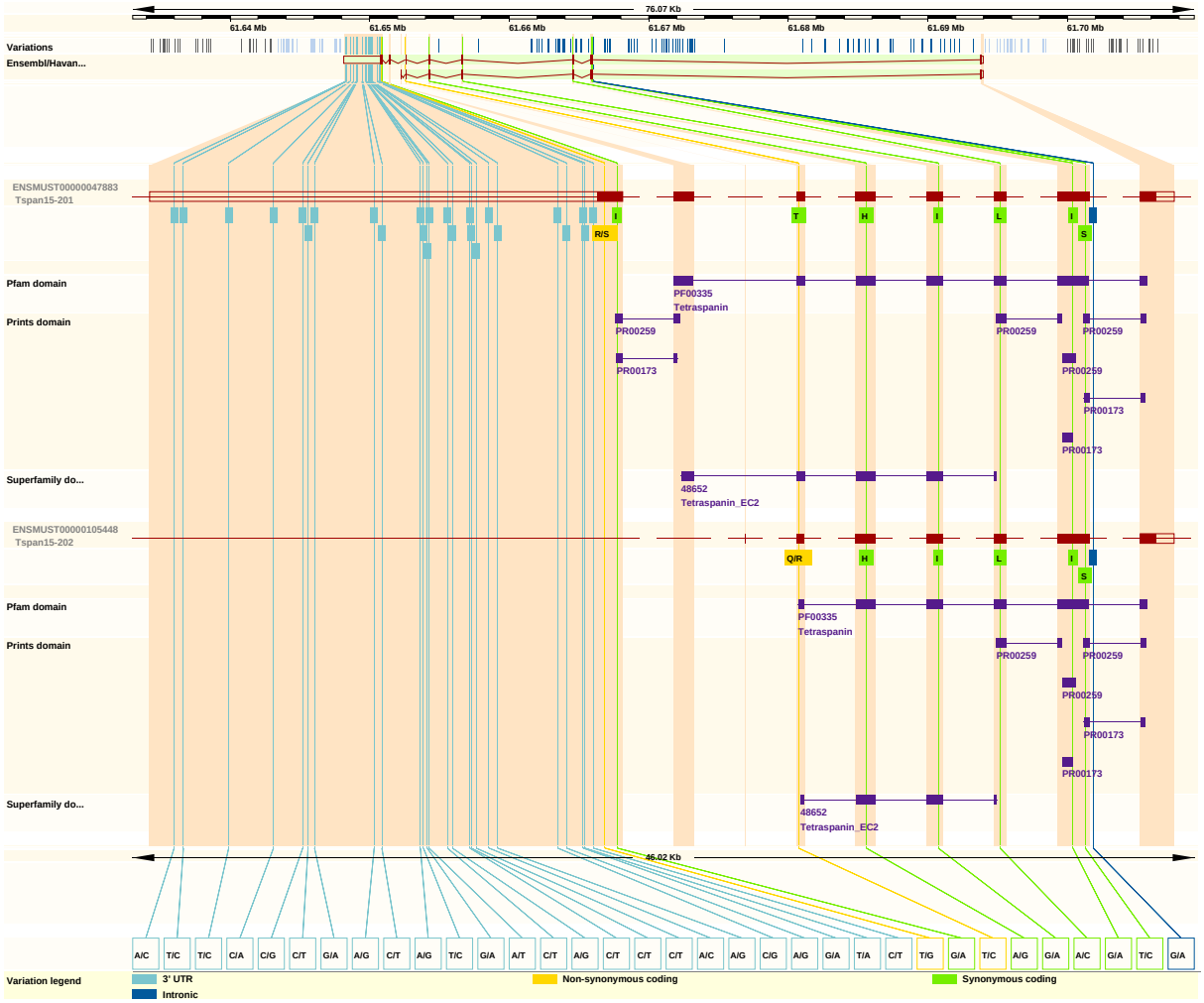

Figure 7: SNPs in the murine *Tspan15* gene, indicating possible transcriptional effects of trans-acting regulatory sequences. Graphic provided by e!ensamble.

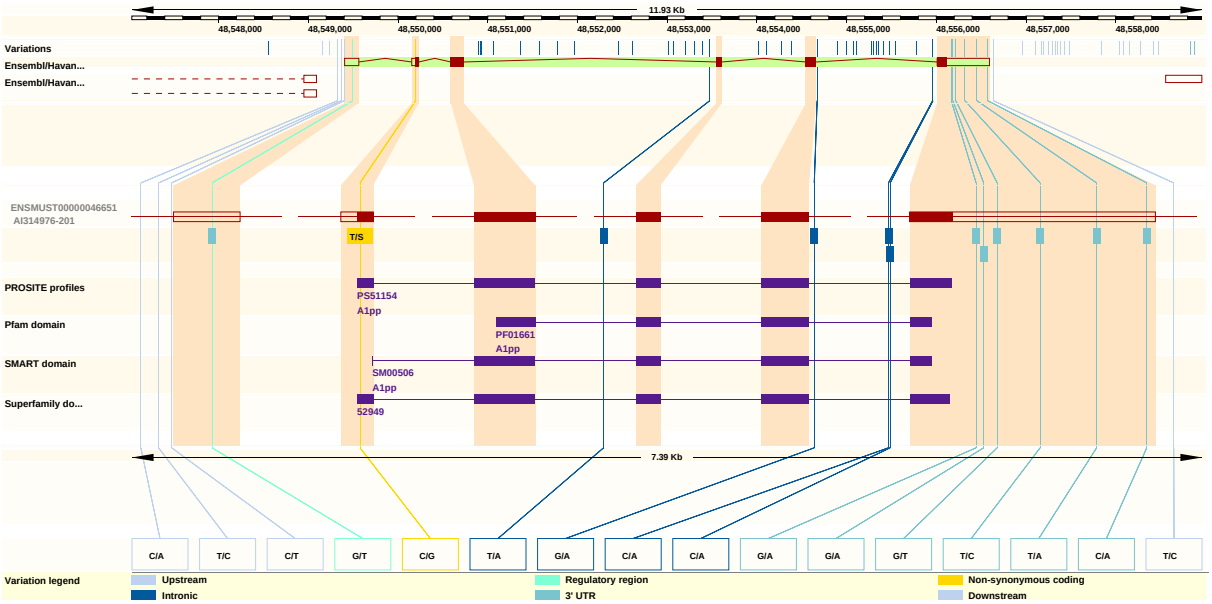

Figure 8: SNPs in the murine*AI314976* gene, indicating possible transcriptional effects of trans-acting regulatory sequences. Graphic provided by e!ensamble.

Table 8: PANTHER classification of biological processes of deregulated genes in the circadian-clock non-controlled nutritional expression study showing strong relations to certain biological processes. (sample maize: n=3, control: n=3 isogenic maize)

Biological Process	Genes on Mouse AB 1700 genes	Number of genes	Expected by chance	over/under represented	P-value
Protein biosynthesis	446	301	44	+	1.30E-148
Protein metabolism and modification	2720	552	269	+	5.30E-61
Nucleoside, nucleotide and nucleic acid metabolism	2779	397	275	+	5.30E-14

Table 9: Samples assigned to the TaqMan low density arrays

	TLDA1	TLDA2	TLDA3	TLDA4	TLDA5
Slot1	MGS K1♂ Jejunum	MGS K1♂ Jejunum	MGS K4♂ Jejunum	MGS G1♂ Duodenum	MGS G11♂ Duodenum
Slot2	MGS K3♂ Jejunum	MGS K3♂ Jejunum	MGS K6♂ Jejunum	MGS G3♂ Duodenum	MGS G8♂ Ileum
Slot3	MGS K4♂ Jejunum	MGS K8♂ Jejunum	MGS K8♂ Jejunum	MGS K7♂ Duodenum	MGS K3♂ Ileum
Slot4	MGS K6♂ Jejunum	MGS K11♂ Jejunum	MGS K11♂ Jejunum	MGS K1♂ Ileum	MGS K4♂ Ileum
Slot5	MGS G1♂ Jejunum	MGS G1♂ Jejunum	MGS G4♂ Jejunum	MGS A3♂ Duodenum	MGS A8♂ Ileum
Slot6	MGS G3♂ Jejunum	MGS G3♂ Jejunum	MGS G5♂ Jejunum	MGS A7♂ Duodenum	MGS A10♂ Ileum
Slot7	MGS G4♂ Jejunum	MGS G8♂ Jejunum	MGS G8♂ Jejunum	MGS A10♂ Duodenum	MGS A1♂ Jejunum
Slot8	MGS G5♂ Jejunum	MGS G11♂ Jejunum	MGS G11♂ Jejunum	MGS A7♂ Ileum	MGS A3♂ Jejunum

	TLDA6	TLDA7	TLDA8	TLDA9	TLDA10
Slot1	MGS G10♂ Ileum	RACB (3.F1) K1♂ Jejunum	RACB (3.F1) K3♂ Jejunum	RACB (3.F1) K5♂ Jejunum	RACB (F0) K3♂ Jejunum
Slot2	MGS G11♂ Ileum	RACB (3.F1) K2♂ Jejunum	RACB (3.F1) K4♂ Jejunum	RACB (3.F1) K6♂ Jejunum	RACB (F0) K9♂ Jejunum
Slot3	MGS K4♂ Duodenum	RACB (3.F1) G1♂ Jejunum	RACB (3.F1) G3♂ Jejunum	RACB (3.F1) G5♂ Jejunum	RACB (F0) K10♂ Jejunum
Slot4	MGS K6♂ Duodenum	RACB (3.F1) G2♂ Jejunum	RACB (3.F1) G4♂ Jejunum	RACB (3.F1) G6♂ Jejunum	RACB (F0) K12♂ Jejunum
Slot5	MGS A7♂ Jejunum	RACB (3.F1) G1♀ Scrape	RACB (3.F1) G3♀ Scrape	RACB (3.F1) G5♀ Scrape	RACB (F0) G7♂ Jejunum
Slot6	MGS A8♂ Jejunum	RACB (3.F1) G1♀ Jejunum	RACB (3.F1) G3♀ Jejunum	RACB (3.F1) G5♀ Jejunum	RACB (F0) G8♂ Jejunum
Slot7	MGS A9♂ Jejunum	RACB (3.F1) K3♀ Scrape	RACB (3.F1) K2♀ Scrape	RACB (3.F1) K8♀ Scrape	RACB (F0) G9♂ Jejunum
Slot8	MGS A10J♂ Jejunum	RACB (3.F1) K3♀ Jejunum	RACB (3.F1) K5♀ Jejunum	RACB (3.F1) K8♀ Jejunum	RACB (F0) G10♂ Jejunum

Table 10: Apportionment of genes as used on the TaqMan low density arrays

Lane 1		Lane 2		Lane 3		Lane 4		Lane 5		Lane 6		Lane 7		Lane 8	
Igtp	Irs1	Igtp	Irs1	Igtp	Irs1	Igtp	Irs1	Igtp	Irs1	Igtp	Irs1	Igtp	Irs1	Igtp	Irs1
Pnpla3	Aatf	Pnpla3	Aatf	Pnpla3	Aatf	Pnpla3	Aatf	Pnpla3	Aatf	Pnpla3	Aatf	Pnpla3	Aatf	Pnpla3	Aatf
Ugcgl2	Cd40	Ugcgl2	Cd40	Ugcgl2	Cd40	Ugcgl2	Cd40	Ugcgl2	Cd40	Ugcgl2	Cd40	Ugcgl2	Cd40	Ugcgl2	Cd40
Itga5	A230005M16Rik;Bmyc	Itga5	A230005M16Rik;Bmyc	Itga5	A230005M16Rik;Bmyc	Itga5	A230005M16Rik;Bmyc	Itga5	A230005M16Rik;Bmyc	Itga5	A230005M16Rik;Bmyc	Itga5	A230005M16Rik;Bmyc	Itga5	A230005M16Rik;Bmyc
Bcar3	Nphp1	Bcar3	Nphp1	Bcar3	Nphp1	Bcar3	Nphp1	Bcar3	Nphp1	Bcar3	Nphp1	Bcar3	Nphp1	Bcar3	Nphp1
Elmo2	Gspt2	Elmo2	Gspt2	Elmo2	Gspt2	Elmo2	Gspt2	Elmo2	Gspt2	Elmo2	Gspt2	Elmo2	Gspt2	Elmo2	Gspt2
Gpr39	Prkcn	Gpr39	Prkcn	Gpr39	Prkcn	Gpr39	Prkcn	Gpr39	Prkcn	Gpr39	Prkcn	Gpr39	Prkcn	Gpr39	Prkcn
Ntrk2	Per3	Ntrk2	Per3	Ntrk2	Per3	Ntrk2	Per3	Ntrk2	Per3	Ntrk2	Per3	Ntrk2	Per3	Ntrk2	Per3
Shc1	Stat5b	Shc1	Stat5b	Shc1	Stat5b	Shc1	Stat5b	Shc1	Stat5b	Shc1	Stat5b	Shc1	Stat5b	Shc1	Stat5b
Gga1	Mapk10	Gga1	Mapk10	Gga1	Mapk10	Gga1	Mapk10	Gga1	Mapk10	Gga1	Mapk10	Gga1	Mapk10	Gga1	Mapk10
18S	Sntb2	18S	Sntb2	18S	Sntb2	18S	Sntb2	18S	Sntb2	18S	Sntb2	18S	Sntb2	18S	Sntb2
Socs1	Dnajc1	Socs1	Dnajc1	Socs1	Dnajc1	Socs1	Dnajc1	Socs1	Dnajc1	Socs1	Dnajc1	Socs1	Dnajc1	Socs1	Dnajc1
Pum1	Fbxw7	Pum1	Fbxw7	Pum1	Fbxw7	Pum1	Fbxw7	Pum1	Fbxw7	Pum1	Fbxw7	Pum1	Fbxw7	Pum1	Fbxw7
Sgta	Cd68	Sgta	Cd68	Sgta	Cd68	Sgta	Cd68	Sgta	Cd68	Sgta	Cd68	Sgta	Cd68	Sgta	Cd68
Clk3	Hmg20a	Clk3	Hmg20a	Clk3	Hmg20a	Clk3	Hmg20a	Clk3	Hmg20a	Clk3	Hmg20a	Clk3	Hmg20a	Clk3	Hmg20a
2610529C04Rik	Ccnh	2610529C04Rik	Ccnh	2610529C04Rik	Ccnh	2610529C04Rik	Ccnh	2610529C04Rik	Ccnh	2610529C04Rik	Ccnh	2610529C04Rik	Ccnh	2610529C04Rik	Ccnh
Eef1b2	Btla	Eef1b2	Btla	Eef1b2	Btla	Eef1b2	Btla	Eef1b2	Btla	Eef1b2	Btla	Eef1b2	Btla	Eef1b2	Btla
Ramp1	Foxq1	Ramp1	Foxq1	Ramp1	Foxq1	Ramp1	Foxq1	Ramp1	Foxq1	Ramp1	Foxq1	Ramp1	Foxq1	Ramp1	Foxq1
Gsk3b	Trim47	Gsk3b	Trim47	Gsk3b	Trim47	Gsk3b	Trim47	Gsk3b	Trim47	Gsk3b	Trim47	Gsk3b	Trim47	Gsk3b	Trim47
Csf3r	Fkbp5	Csf3r	Fkbp5	Csf3r	Fkbp5	Csf3r	Fkbp5	Csf3r	Fkbp5	Csf3r	Fkbp5	Csf3r	Fkbp5	Csf3r	Fkbp5
Ift140	Kif3c	Ift140	Kif3c	Ift140	Kif3c	Ift140	Kif3c	Ift140	Kif3c	Ift140	Kif3c	Ift140	Kif3c	Ift140	Kif3c
Herc3	Rgs6	Herc3	Rgs6	Herc3	Rgs6	Herc3	Rgs6	Herc3	Rgs6	Herc3	Rgs6	Herc3	Rgs6	Herc3	Rgs6
Rpl22	Hprt1	Rpl22	Hprt1	Rpl22	Hprt1	Rpl22	Hprt1	Rpl22	Hprt1	Rpl22	Hprt1	Rpl22	Hprt1	Rpl22	Hprt1
Eef2k	Psmc3	Eef2k	Psmc3	Eef2k	Psmc3	Eef2k	Psmc3	Eef2k	Psmc3	Eef2k	Psmc3	Eef2k	Psmc3	Eef2k	Psmc3