RESEARCH ARTICLE

The Orthology Clause in the Next Generation Sequencing Era: Novel Reference Genes Identified by RNA-seq in Humans Improve Normalization of Neonatal Equine Ovary RT-qPCR Data

Dragos Scarlet[1] *, Reinhard Ertl[2], Christine Aurich[1], Ralf Steinborn[2]

1 Centre for Artificial Insemination and Embryo Transfer, University of Veterinary Medicine, Vienna, Austria,
2 Genomics Core Facility, VetCore, University of Veterinary Medicine, Vienna, Austria

* dragos.scarlet@vetmeduni.ac.at

Abstract

Background

Vertebrate evolution is accompanied by a substantial conservation of transcriptional programs with more than a third of unique orthologous genes showing constrained levels of expression. Moreover, there are genes and exons exhibiting excellent expression stability according to RNA-seq data across a panel of eighteen tissues including the ovary (Human Body Map 2.0).

Results

We hypothesized that orthologs of these exons would also be highly uniformly expressed across neonatal ovaries of the horse, which would render them appropriate reference genes (RGs) for normalization of reverse transcription quantitative PCR (RT-qPCR) data in this context. The expression stability of eleven novel RGs (*C1orf43*, *CHMP2A*, *EMC7*, *GPI*, *PSMB2*, *PSMB4*, *RAB7A*, *REEP5*, *SNRPD3*, *VCP* and *VPS29*) was assessed by RT-qPCR in ovaries of seven neonatal fillies and compared to that of the expressed repetitive element *ERE-B*, two universal (*OAZ1* and *RPS29*) and four traditional RGs (*ACTB*, *GAPDH*, *UBB* and *B2M*). Expression stability analyzed with the software tool RefFinder top ranked the normalization factor constituted of the genes *SNRPD3* and *VCP*, a gene pair that is not co-expressed according to COEXPRESdb and GeneMANIA. The traditional RGs *GAPDH*, *B2M*, *ACTB* and *UBB* were only ranked 3rd and 12th to 14th, respectively.

Conclusions

The functional diversity of the novel RGs likely facilitates expression studies over a wide range of physiological and pathological contexts related to the neonatal equine ovary. In addition, this study augments the potential for RT-qPCR-based profiling of human samples

by introducing seven new human RG assays (*C1orf43*, *CHMP2A*, *EMC7*, *GPI*, *RAB7A*, *VPS29* and *UBB*).

Introduction

In recent years the mare has intensively been used as an animal model for age-associated changes in human reproductive medicine [1]. Better approaches for the analysis of the molecular mechanisms behind ovarian development and maturation would help to prevent and treat infertility in both the human and the horse. The developing fetal equine ovary possesses several morphologic and developmental differences compared to other domestic species [2]. After the interstitial cell hypertrophy and hyperplasia that results in marked gonadal enlargement during the time spent in utero, a period of regression and ovarian cortical migration follows [2]. Recently, we have described the presence of tertiary follicles up to 4 mm in diameter in newborn fillies [3]. However, the regulatory mechanisms behind these processes have not been elucidated yet.

Quantitative real-time RT-PCR (RT-qPCR) is the most commonly used method for measuring transcript copies being especially useful when samples contain small numbers of cells [4]. The accuracy of RT-qPCR measurement is strongly dependent on appropriate normalization—usually accomplished using reference genes (RGs; [5]). RGs are mainly derived from the class of housekeeping genes required for the maintenance of basal cellular functions that are essential for the existence of a cell, regardless of its specific role in the tissue or organism [6]. RGs are the major drawback for RT-qPCR, as their validation on an individual basis for all treatment and experimental conditions is essential for accurate data interpretation, especially when normalizing for inter-sample variability [7]. Depending on the biological context, two [8,9] or three [8,10,11] of the most stably expressed RGs are sufficient to normalize the expression.

Over the past years, several mammalian genes have been intensively used as internal controls in gene expression studies, such as the traditional RGs *RN18S* (*18S rRNA*), *GAPDH*, *ACTB*, *HPRT*, *B2M* and *ALB*. Some of the latter genes (*ACTB*, *GAPDH* and *RN18S*) appear in 70% of the publications on RNA based qPCR as the solely normalizer, a fact that makes it highly unlikely that these have been validated [12]. Traditional RGs were derived by educated guesswork in the era of Northern blotting [13–15]. In the absence of any transcriptome-wide expression information, the expression uniformity of these genes was supposed to be stably expressed not only in model species, but also in their genetic relatives and hence frequently used as normalizers across plants and animals [16–19]. Some years ago, the orthology clause was suggested for a particular target context, the liver of mice, cattle and pig [20]. The clause indicated a similar expression of gene orthologs across related species. A huge amount of information, including novel aspects on human and mouse transcriptomes, was recently uncovered through RNA-seq [6,21]. This has opened a new dimension for studying mechanisms of molecular regulation. This data also indicated a substantial conservation of transcriptional programs between human and mouse by uncovering a distinct class of genes with levels of expression that have been constrained early in vertebrate evolution [22]. In detail, it was found that evolutionary constraint in gene expression levels is associated with conserved epigenetic marking and a characteristic post-transcriptional regulatory program.

Stability of RNA expression can be assessed by software tools such as GeNorm [23], NormFinder [24] or BestKeeper [25]. These major computational programs are integrated by the

tool RefFinder ([26], http://fulxie.0fees.us/?type = reference), thus compensating for their individual weaknesses.

No gene exhibits a stable pattern of expression across all tissues and conditions of an individual [20]. Commonly used reference genes such as *ACTB*, *GAPDH*, *UBC* and *RN18S* can vary considerably depending on tissue types, developmental stage, sex, pathology, and experimental conditions [27]. For example, in whole blood donated by healthy human individuals, expression of *ACTB*, *GAPDH* or *HPRT* differed by a factor of 20- to 25-fold [28]. This level of variation considerably exceeds the strong threshold criterion of $\Delta Cq \leq \pm0.5$ (twofold expression alteration) recommended for RG suitability [7]. Only few genes, termed "universal" reference genes, are relatively uniformly expressed across a wide range of conditions in men or mice [20]. For each biological context, a subset of genes exists that is more stably expressed than traditional normalizers or RGs that were selected by meta-analysis based on their expression stability across a plenitude of experimental conditions [20].

So far, there is a lack of reliable RGs for the biological context of this study—the neonatal ovary of *Equus caballus*. Here we tested the application of the orthology clause in the next generation sequencing era for this experimental setup. We targeted the equine orthologous genes of the most stably expressed human counterparts identified by RNA-seq from a wide panel of human tissues (Illumina Human Body Map 2.0) including the ovary. We argued that the extended number of transcripts identified, a resolution down to the level of individual exons and the expression consensus obtained by profiling across a large set of target tissues would facilitate a successful application of the clause, hence improve the gene repertoire for normalization of RT-qPCR data in this context.

Material and Methods

In agreement with University Guidelines for Good Scientific practice, owners' approval for scientific use of post-mortem material was obtained. In accordance with Austrian legislation on animal experimentation, collection of postmortem material from animal that had to be euthanized for veterinary medical reasons is not considered animal experimentation. Thus neither an animal experimentation license issued by the competent authority nor an internal approval by the University Animal Care and Use Committee was required.

Biological material

Fourteen ovaries were collected from euthanized newborn female foals ($n = 7$). Euthanasia was performed due to various pathological conditions not related to the female genital tract using 10 mg/kg Thiopental (Inresa, Freiburg, Germany) followed by 4–6 ml/50kg T61 (MSD Animal Health, Vienna, Austria). The average age of foals was two days (range: 0 to 8 days).

Sample preparation and RNA extraction

Fresh ovary biopsies were embedded in Tissue-Tek O.C.T. compound (Sakura Finetek, Alphen aan den Rijn, Netherlands), snap-frozen in liquid nitrogen and then stored at -80°C for less than 12 months. Frozen sections of 5 × 100 μm tissue were prepared on the cryostat CryoStar NX70 (Thermo Scientific, Waltham, USA). After removal of excess embedding medium, the sections were placed into 2 ml-screw caps filled with 1.3 mm ceramic beads (Peqlab Biotechnologie GmbH, Erlangen, Germany) and 350 μl RLT lysis buffer supplemented with 3.5 μl β-mercaptoethanol (Qiagen, Hilden, Germany). Mechanical homogenization was performed on the MagNA Lyser instrument (Roche, Rotkreuz, Switzerland) at 6500 rpm for 20 s followed by cooling on ice for 1 min. If necessary, the homogenization step was repeated. Total RNA was extracted using the RNeasy Micro Kit (Qiagen) according to Qiagen's RNeasy Micro Kit

protocol for fibrous tissues except an extended time for Proteinase K incubation (30 min). The procedure included on-column DNase I digestion for 15 min to remove residual DNA that can generate a false positive signal in some of the RT-qPCR assays. RNA concentration was determined by UV spectrophotometry on the NanoDrop 2000c (Peqlab Biotechnologie GmbH). RNA integrity was assessed by capillary electrophoresis on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA) using the RNA 6000 Nano Kit (Agilent Technologies). Only samples with RNA integrity number (RIN) values of at least seven were used for expression analysis (Fig 1). To further reduce unspecific signals from the presence of genomic DNA as concluded from a RT-qPCR assay against a traditional RG, *ACTB* (data not shown), samples were again DNase digested (TURBO DNA-free Kit; Life Technologies, Austin, USA). Exemplarily, two of the seven samples were again assessed on the Agilent Bioanalyzer to exclude that the second DNase digestion had affected the integrity of the RNAs (RIN > 7, ΔRIN of 0 and 1.0).

Selection of RG candidates and design of RT-qPCR primers

Eleven genes showing the most ubiquitous and uniform expression across the ovary and 15 other normal human tissue types [6] were selected as novel RG candidates, namely *C5H1orf43* (ortholog of human *C1orf43)*, *CHMP2A*, *EMC7*, *GPI*, *PSMB2*, *PSMB4*, *RAB7A*, *REEP5*, *SNRPD3*, *VCP*, and *VPS29*. This expression data aiming at housekeeping gene detection in the era of massive parallel sequencing and RNA-seq was part of the Human BodyMap 2.0 project (www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513). This RNA-seq data delivered expressions at the exon level. This measure takes care of normalization for the two most obvious factors affecting the raw number of reads per gene, transcript, or exon: the total number of reads produced and their length. The variation of exon expression given as reads per kilobase per million mapped reads (RPKM) is estimated by the standard deviation of $\log_2$ (RPKM) over samples, abbreviated as $SD(\log_2(RPKM))$.

DNA and cDNA sequences including the transcript isoforms available for the horse were retrieved from NCBI. Aligning them with the orthologous human sequences allowed identifying primer target sequences that were conserved across the two species and less affected by individual polymorphisms.

A biological sample is composed of different types of cells mixed at different proportions. If an RT-qPCR assay is sensitive to alternative splicing, sample heterogeneity can lead to the identification of artificially deregulated transcript isoforms and in case of a RG to an increased expression variation. Therefore, our protocol was designed to target exons that are present in multiple tissue types.

As the exon/intron architecture of genes influences the way of mRNA processing, it was argued that selecting flanking introns of sizes around 1 to 2 kb would be an appropriate compromise to reduce the chance of alternative splicing caused by exon skipping and the alternative usage of 5' or 3' splice sites (based on human data from [29]). In general, the orthologous human exons targeted by the equine RT-qPCR assays were moderately to highly and uniformly expressed across the panel of human reference tissues ($SD(\log_2(RPKM))$ of 0.32 to 0.52 [6]). In a single case (*SNRPD3*), a less stable exon ($SD(\log_2(RPKM))$ = 0.70 [6]) had to be accepted as a target region in order to achieve the proposed size (1 to 2 kb) of the flanked intron.

To avoid co-amplification of contaminating genomic DNA due to the short elongation time, a critical size limit of at least 750 bp was adopted for the central intron, *i.e.* the one flanked by the target exons in genomic DNA. The limit was met for nine of the eleven novel putative RGs (Table 1). Two of the primer pairs flanked short introns of 73 and 412 bp (*CHMP2A* and *PSMB4*, respectively). In these cases, the minus RT control routinely performed for each experimental RNA confirmed specificity of mRNA amplification.
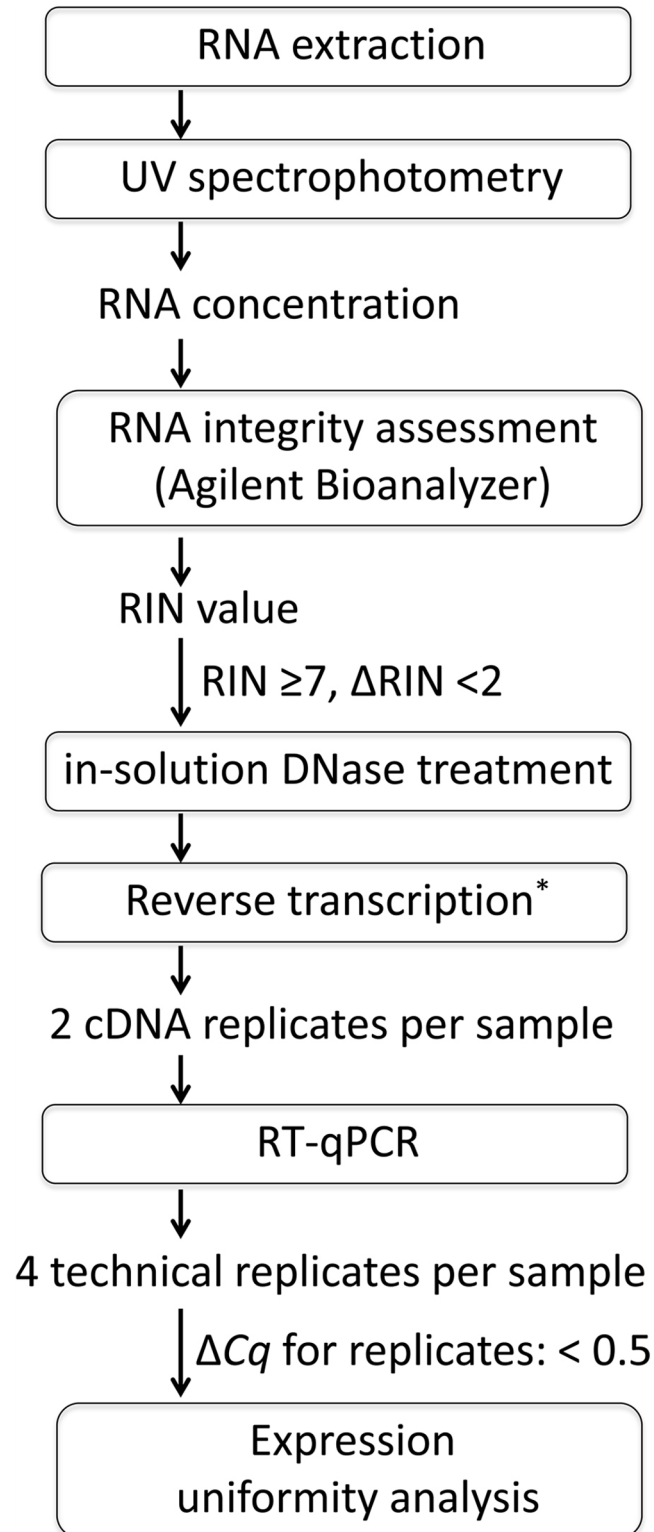
Fig 1. Processing of RNA for expression uniformity analysis by RT-qPCR. *Input RNA might contain nucleotides or short DNA fragments as a result of DNase digestion of co-purified DNA. The same amount of RNA was used for RT as concluded from a second spectrophotometric measurement of RNA mass before the RT step and based on the subsequent determination of Cq values for the rRNA genes *RN18S* and *RN28S* (Fig 2).

doi:10.1371/journal.pone.0142122.g001

Table 1. Details on qPCR assays.

Gene symbol	NCBI accession number	5'-3' sequence	E	Amplicon size(bp)	Targeted exonsF, R	Central intronsize (bp)	Upstreamintron size(bp)	Down-streamintron size (bp)	Assay ID
C5H1orf43	XM_003365015.2, XM_001496636.2, XM_001496656.2, XM_008526884.1, NW_007673359.1	F:GCTATCAGGAGGCCCTGAGT R:GGAGGTATGTCACCTGGATGG	1.01	141	6,7	2020	87	-	Hs:8778
CHMP2A	XM_005596626.1, XM_008544637.1, XM_008544638.1	F:ATGGGCACCATGAACAGACAG R:TCTCCTCTTCATCTTCCTCATCAC	0.90	151	1,2	73	-	90	Hs:8779
EMC7	XM_001503666.3, XM_008516349.1	F:GTCAGACTGCCCTATCCTCTCC R:CATGTCAGGATCACTTGTGTTGAC	1.00	177	3,4	1302	3781	2214	Hs:8784
GPI	XM_005596131.1, XM_008515114.1, XM_001490607.3	F:CCAAGTCCAGGGGCGTG R:CTTGTTGACCTCTGGCATCACA	0.90	158	3,4	1433	358	4394	Hs:8780
PSMB2	XM_001503676.3, XM_003364409.2, XM_008533360.1	F:ATATCATGTCAACCTCCTCCTGG R:AAGCTCCCACTGCCCTCTCAC	0.90	187	4,5	1475	18074	2003	
PSMB4	XM_001492317.4, XM_005610132.1, XM_008515015.1, XM_005613704.1	F:CTTGGTGTAGCCTATGAAGCCC R:CCAGAATTTCTCGCAGCAGAG	0.93	82	4,5	412	178	216	
RAB7A	XM_001488301.3, XM_008506841.1, XM_008506842.1	F:CTGGTATTTGATGTGACTGCCC R:AATCGTCTGGAACGCCTGC	1.02	252	3,4	1623	8255	4181	Hs:8782
REEP5	XM_005599722.1, XM_008535732.1	F:CCTGAAGCACGAGTCCCAG R:CCCAGTAAATTCACGGCAGC	0.92	114	3,4	5086	10372	-	
SNRPD3	XM_001489060.4, XM_008511652.1	F:ACGCACCTATGTTAAAGAGCATG R:CACGTCCCATTCCAGTC	0.99	120	3,4	1861	16047	-	
VCP	XM_005605574.1, XM_008508520.1	F:GAGTGAGATCAGGCGAGAACG R:CCTCTTCCACCTCCATGGC	0.92	56	14,15	1673	279	154	Hs:8786
VPS29	XM_001495080.4, XM_001495099.4, XM_008508375.1	F:AAACCTTTGCACCAAAGAGAGTTATG R:CAACAGTCACAACTTTCTGTTCTGG	0.90	119	2,3	2105	4766	244	
ACTB	NM_001081838.1	F:CGGGACCTGACGGACTA R:CCTTGATGTCACGCACGATT	0.91	94	3,3	-	448	85	
GAPDH	NM_001163856.1, XM_008513936.1	F:GGCAAGTTCCATGCCACAGT R:CACAACATATTCAGCACCAGCAT	0.90	129	4,5	119	82	127	Hs:8781
UBB	AF506969.1	F:TTCGTGAAGACCCTGACC R:CCTTATCCTGGATCTTGGC	0.92	91	2,2	-	NA	-	
B2M	XM_005602594.1, XM_005602595.1	F:CCTGCTCGGGCTACTCTC R:CATTCTCTGCTGGGTGACG	0.94	89	1,2	3266	-	583	
OAZ1	CD467318.1	F:CGGCTCCCTGTACATCGAGAT R:GCGGTTCTTGTGGAAGCAGAT	0.98	133	3,4	NA	NA	NA	
RPS29	XM_001494060.4, XM_008528084.1	F:CAGCTCTACTGGACGCCATC R:ACATGTTGAGGGCCGTACTTC	0.96	103	1,2	207	-	1133	

(Continued)

Table 1. (*Continued*)

Gene symbol	NCBI accession number	5'-3' sequence	E	Amplicon size(bp)	Targeted exonsF, R	Central intronsize (bp)	Upstreamintron size(bp)	Down-streamintron size (bp)	Assay ID
ERE-B	-	F:GTGGCCTAGTGGTTAAGTTCG R: ACGCCGCCACAGCATG	0.93	101–113			intron-free gene		
RN18S	*NR_046271.1*	F:CCATTCGAACGTCTGCCCTA R: TCACCCGTGGTCACCATG	0.94	68			intron-free gene		
RN28S	*NR_046309.1*	F:TAGCCAAATGCCTCGTCATCT R: AACAGTTAGGGACAGTGGGAATCT	0.96	74			intron-free gene		

F and R: forward and reverse primers; *E*: amplification efficiency; NA: information not available due to incomplete sequence data.

Hs: *Homo sapiens*.

Assay ID according to RTPrimerDB (http://medgen.ugent.be/rtprimerdb/).

doi:10.1371/journal.pone.0142122.t001
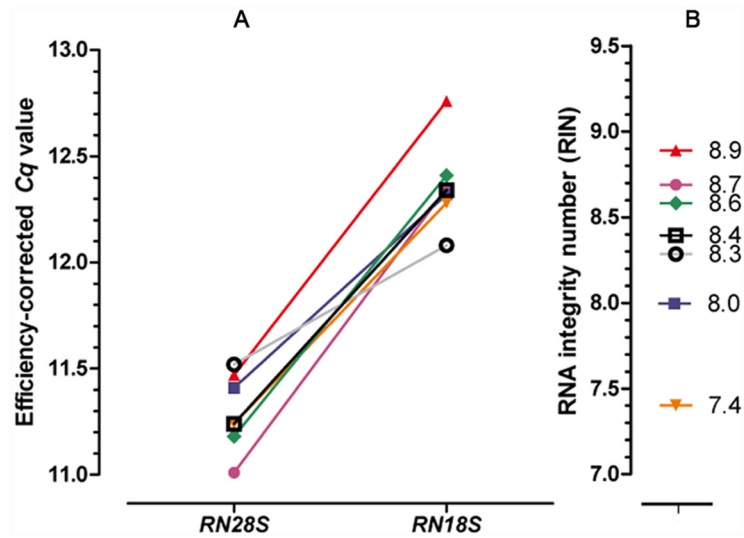
Fig 2. RT-qPCR profiling of the rRNA genes *RN18S* and *RN28S* (A) and assessment of RNA integrity (B).

Each RT-qPCR assay targeted all of the two to four transcripts available at the NCBI Gene database for *E. caballus* and *E. przewalskii* (S1 File).

In order to exclude differences in the amount of co-isolated DNA, samples were analysed for expression of *18S* und *28S* ribosomal fractions.

For comparison four traditional RGs, two universal RGs and one expressed repetitive element were selected. The traditional RGs were selected just by educated guess-work and tradition (*ACTB* and *GAPDH*, [30]) or based on stable expression across a panel of equine tissues (*UBB* and *B2M*, [31,32]). The universal RGs, *OAZ1* and *RPS29*, exhibited highly uniform expression in men and mice across a plenitude of conditions [33–35]. The assay for the expressed repetitive element targeted a consensus sequence deduced from all equine repetitive element B (*ERE-B*) sequences available at the SINEBase database ([36]; http://sines.eimb.ru/).

Primers for the novel candidates, the universal RGs and the expressed repetitive element were designed in the Primer Express 2.0 program (Life Technologies, Foster City, USA). Primer dimerization was assessed by the NetPrimer web tool (Premier Biosoft, Palo Alto, USA) freely available at www.premierbiosoft.com. Amplicon specificity was evaluated by the Primer-BLAST tool of NCBI using the non-redundant database of the taxid "horses". Amplicon folding was analyzed using the Mfold web server [37]. Primer sequences and assay details are listed in Table 1.

RT-qPCR

In order to account for variations during the RT step, two replicates of each RNA sample were converted into cDNA using the High Capacity Reverse Transcription Kit (Life Technologies) (Fig 1). For each replicate, 2 µg total RNA input were reverse-transcribed with random hexamers in a reaction volume of 40 µl following the recommended protocol. Since sensitivity towards genomic DNA contamination differs greatly between assays [38,39], controls in which no reverse transcriptase was added were run for each assay and each experimental RNA. RT-qPCR was performed in duplicates for each cDNA replicate (equals four technical replicates for each RNA sample) using the Power SYBR Green Master Mix (Life Technologies). The 20-µl reaction included 150 nM of each primer and 2 µl of five-fold diluted cDNA. All qPCR runs

were performed on a Viia7 Real-Time PCR System (Life Technologies). After an initial denaturation at 95°C for 10 min, 40 amplification cycles of 95°C for 15 s and 60°C for 1 min were performed. Amplicon dissociation was monitored over a temperature range of 60°C to 95°C. Additionally, PCR products were separated on a 2% agarose gel stained with GelGreen (Biotium, Hayward, USA) and visualized by blue light excitation (S2 File). Absence of unspecific amplification derived from the mastermix was confirmed by melting curve analysis of the no template control product. Absence of PCR inhibitors was assessed by RT-qPCR against *RN18S* using eightfold dilution of all experimental cDNAs as template.

Reverse transcription was repeated if the *Cq* values of two cDNA replicates differed by more than 0.5 cycles.

A standard curve was generated for each assay for calculating PCR efficiency (*E*) from the term $10^{-1/s}$—1, were *s* is the slope of the graph obtained by plotting the *Cq* value against the log_{10} of the cDNA input [40]. For *Cq* values generated at $E < 1$, efficiency corrected *Cq* values were obtained from the term $Cq * log_{10}(E + 1)/log_{10}(2)$ [41].

Data analysis

The efficiency-corrected *Cq* values were processed using the RefFinder tool integrating the currently available major computational programs geNorm, Normfinder, BestKeeper, and the comparative $\Delta\Delta Cq$ method to compare and rank the tested candidate reference genes ([26]; http://fulxie.0fees.us/?type = reference). Based on the rankings from each program, RefFinder assigns an appropriate weight to an individual gene or set of genes and calculates the geometric mean of their weights for the overall final ranking. A normalization factor, *i.e.* the geometric mean of several top-ranking genes, was considered suitable as soon as it was ranked ahead of all single genes.

In order to exclude the association of highly coexpressed genes which could affect the stability of the normalization factor, coexpression was identified with COXPRESdb [42] software, which calculates Pearson's correlation coefficient for each gene pair, and then transfers these values to the mutual rank value, which is the geometric average of asymmetric ranks in coexpressed gene lists.

To confirm COXPRESdb results, the predicted coexpression of the novel RGs was analysed with GeneMANIA [43], a software based on functional association data such as protein and genetic interactions, pathways, co-expression, co-localization, and protein domain similarity. The default approach, the automatically selected weighting method, was used when combining all networks into the final composite. In case of our search list with more than five genes, GeneMANIA assigns weights to maximize connectivity between all input genes using the 'assigned based on query gene' strategy. According to this strategy the weights are chosen automatically using linear regression, to make genes on the search list interact as much as possible with each other and as little as possible with genes not on this list.

As *E. caballus* species is neither reviewed by COEXPRESdb, nor by GeneMANIA, *Homo sapiens* databases were used in both cases.

The RT-qPCR data of this work comply with the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines [5].

Statistical analysis

Statistical analysis was performed using SPSS Statistics 22 (IBM-SPSS, Armonk, NY, USA). The Kolmogorov–Smirnov test was used to demonstrate a normal distribution of efficiency corrected and log transformed *Cq* values of the candidate RGs. Correlations between parameters were analyzed using the Pearson's correlation test. A $P < 0.05$ was considered significant.

Results and Discussion

This study addressed the orthologous use of exons ubiquitously and stably expressed across a wide set of human tissues for improving RT-qPCR normalization in the context of the equine ovary. If genome-wide transcript data is missing for a specific biological context in a related species, for example the neonatal ovary of the horse, this assumption could be applied to select RGs for improving RT-qPCR data normalization.

Choosing the most suitable RGs for one experiment is being complicated by the fact that the majority of mRNAs undergoes alternative polyadenylation [44]. This is more likely to occur when the target exons are located in the three prime parts of a gene. In addition, most human multi-exon genes are known to be alternatively spliced. Out of these, 85% have a minor isoform frequency of at least 15% [45]. For the human genome, alternative splicing by exon skipping is more likely to occur when exons are flanked by long introns, whereas the activation of alternative splice sites is less likely if the flanking introns are long [29,46]. Therefore, for the first time we considered these crucial principles when selecting appropriate exons for targeting by the RT-qPCR primers.

The eleven novel RGs showed high to low expression levels (efficiency-corrected Cq values < 15 to > 30, respectively; S3 File) and span a wide range of biological functions in general (S4 File). According to the GeneMANIA prediction server, they are not part of the same biochemical process ("high" false discovery rate of 5.44e-2) with exception of $PSMB2$, $PSMB4$ and VCP which belong to the proteasome complex.

Intactness of experimental RNAs was confirmed by RIN values above the required threshold of at least seven (range: 7.4 to 8.9; Fig 2). Co-purification of traces of DNA during RNA extraction is inevitable [39] with some RNA preparations yielding virtually pure RNA and others virtually pure DNA and this is independent of tissue type or operator [47]. Therefore, after determination of RIN values, samples were digested by DNase I to include also primer pairs from previous publications that were not intron-spanning or that flanked smaller introns (< 750 bp). Considering that in solution digestion did not remove the degradation products of carry-over DNA, we ruled out that the experimental cDNAs were derived from different proportions of total RNA due to differences in the amount of co-extracted genomic DNA. This was concluded from similar Cq values of RT-qPCRs targeting the main contributors to total RNA mass, $RN18S$ and $RN28S$ (Fig 2). In other words, the similarity of Cq values indicated that the selection of novel uniformly expressed genes for normalization of RT-qPCR data in the context of postnatal equine ovaries would not be affected by differences in the amount of co-isolated DNA. We argued that the approximately 50% change in the amount of input RNAs concluded from a ΔCq value of approximately 0.5 for the $RN28S$ transcript (Fig 2) did not affect the aim of this study—selection of novel RGs, since Cq values of $RN28S$ and putative RGs were not correlated ($P > 0.1$).

Absence of PCR inhibitors from the experimental cDNAs was confirmed by RT-qPCR profiling of experimental cDNAs and their eightfold dilutions with an assay against $RN18S$. The low standard deviation of Cq values determined for the eightfold dilutions of the cDNAs indicated homogeneity regarding sample purity (mean ΔCq of 2.7 ± 0.1 corresponding to $E = 1$). The slight deviation from the theoretical ΔCq of three was attributed to inter-assay variation considering that the sets of diluted and undiluted samples were run on different plates.

Expression stability rankings produced by analytic algorithms like NormFinder and Best-Keeper are often in overall agreement [27]. However, a discrepancy in stability rankings and outcomes between different analytical methods has been reported [24,25,48]. To increase the chance for deducing a "consensus" in the expression uniformity ranking of the novel candidate reference genes we used the comprehensive tool RefFinder (S5 File). The comprehensive

approach identified *SNRPD3*—a core component of the spliceosome and *VCP*—an ATPase of the endoplasmic reticulum that prevents mutations caused by DNA damage, as the most stably expressed pair of genes in the biological context studied. Traditional RGs like *GAPDH*, *B2M*, *ACTB and UBB* were less uniformly expressed as documented by their ranks 3 and 12 to 14, respectively. If normalization should be performed with more than one RG, the pair *SNRPD3* and *VCP* is recommended for calculating a normalization factor ([Fig 3](#)).

Finally, co-expression analysis of traditional and universal reference genes as well as the novel normalization candidates was performed using COEXPRESdb using all tissues available. For the best-performing gene pair, *VCP* and *SNRPD3*, absence of a significant co-expression was found (mutual rank: 716.7; Pearson's coefficient: 0.33; [S6 File](#)). The latter values were far from being biologically significant (respective thresholds: < 30 [49] and > 0.55 [50]). While the Pearson's correlation coefficient is affected by the function of the target gene and the method of gene expression database construction, the alternative use of correlation ranks helps to retrieve only genes that are functionally related [51]. Likewise, identification of co-expressed genes by GeneMANIA, containing peer-reviewed data collected mostly from the Gene Expression Omnibus (GEO), also failed to detect a co-expression for this pair of genes ([Fig 4](#)).

For seven genes (*C1orf43*, *CHMP2A*, *EMC7*, *GPI*, *RAB7A*, *VPS29* and *UBB*; [S7 File](#)) RT-qPCR primers could be designed based on a consensus sequence obtained by aligning the equine transcripts with all human orthologs available in the GenBank. The design of cross-species primers considerably reduces the likelihood of mismatches in their priming sites and hence improves quantification accuracy. For experimental validation of these seven RT-qPCR assays for human samples, RNA isolated from HeLa and Jurkat cells was used as template. The specificity of amplicons was demonstrated by agarose gel electrophoresis ([S8 File](#)).
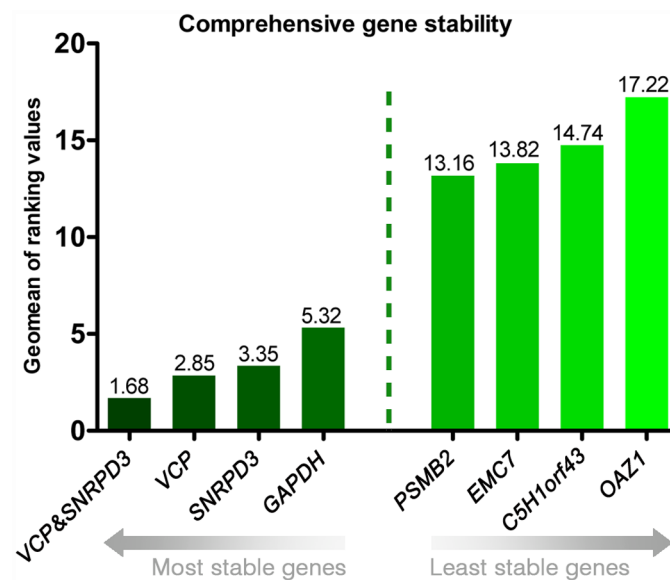

Fig 3. Expression uniformity analysis using the RefFinder software. The individual genes or the normalization factor calculated from the best-ranking, not co-expressed pair of genes are depicted. Only the "expressors" of highest and lowest ranks are presented.
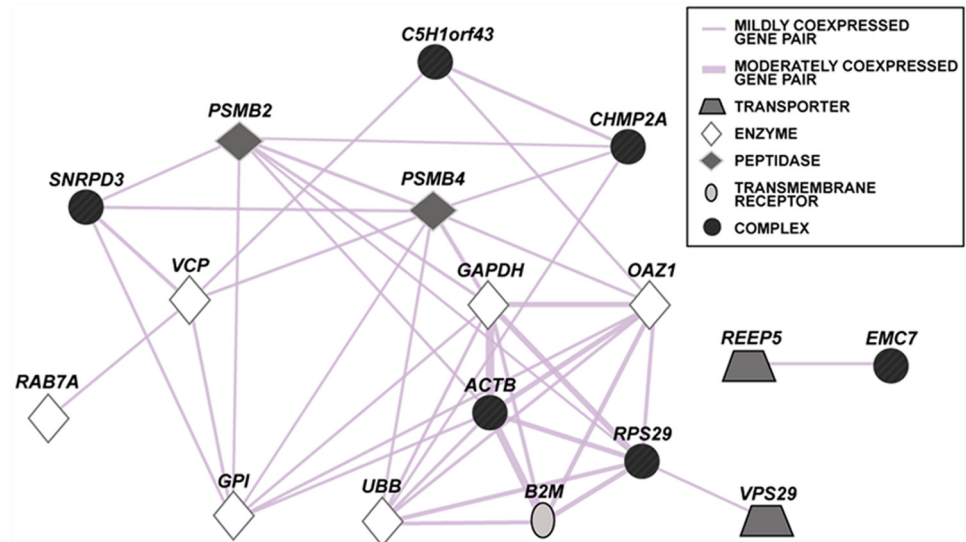
Fig 4. Co-expression analysis for all reference gene candidates according to human expression data contained in GeneMANIA. Two genes are linked if their expression levels are similar across conditions in a gene expression study, mostly derived from peer-reviewed publication data publicly accessible via the Gene Expression Omnibus (GEO). A thicker line indicates a higher co-expression. Note that *ERE-B* represents a repetitive element specific to the horse.

doi:10.1371/journal.pone.0142122.g004

Conclusion

This study assayed the expression stability of highly conserved human exon orthologs in the neonatal horse ovary. Novel housekeeping genes identified in the related species *Homo sapiens* including the best-ranking gene pair of our study, *SNRPD3* and *VCP,* have been confirmed as more stable RGs for RT-qPCR profiling compared to traditional ones. The superiority of the novel RG set is characterized by high expression uniformity and obviously no co-expression. We attributed the successful application of the orthology clause in the next generation sequencing era to the extended number of transcripts identified, a resolution down to the level of individual exons and to the expression consensus obtained by profiling across a large set of target tissues.

The functional diversity of the novel orthologous RGs likely facilitates expression studies over a wide range of physiological and pathological contexts related to the neonatal equine ovary. However, for each biological context a specific subset of RGs needs to be validated from this extended pool of genes. With regard to context-specific RG validation in equine research, the ethical aspect needs to be strongly considered, as sample collection at different stages of development would involve inducing abortion in late pregnant mares or euthanizing young healthy foals.

In the near future RNA-seq analysis putatively performed for the neonatal and the adult equine ovary will help to further extend the repertoire of RGs for RT-qPCR normalization by adding context-optimized RGs with superior expression uniformity as shown before for the context of the mouse jejunum [34].

We also note an extension of the toolkit for human molecular biology by introducing novel validated RG assays for human samples that are in the truest sense in accordance with the MIQE guidelines regarding full transparency of oligonucleotide sequences [5].

Supporting Information

S1 File. The RT-qPCR assays for the novel reference genes detect all transcript isoforms currently known.
(PDF)

S2 File. RT-PCR product size of equine ovary sample determined on a 2% agarose gel stained with GelGreen.
(EPS)

S3 File. Estimated transcript abundance of all reference gene candidates.
(XLSX)

S4 File. Names and functions of reference gene candidates.
(DOCX)

S5 File. Complete list of expression uniformity analysis by the RefFinder software.
(XLSX)

S6 File. Co-expression analysis for all pairs of gene candidates using COEXPRESdb.
(XLSX)

S7 File. Validation of RT-qPCR assays for novel human reference genes.
(XLSX)

S8 File. RT-PCR product size of HeLa and Jurkat cells on a 2% agarose gel stained with Gel-Green.
(EPS)

Acknowledgments

Author Contributions

Conceived and designed the experiments: DS RS. Performed the experiments: RE. Analyzed the data: DS RE CA RS. Wrote the paper: DS RE CA RS. Study initiation: DS CA RS. Supervised the study: CA RS.

References

1. Carnevale EM (2008) The mare model for follicular maturation and reproductive aging in the woman. Theriogenology 69 (1): 23–30. PMID: 17976712

2. Walt ML, Stabenfeldt GH, Hughes JP, Neely DP, Bradbury R (1979) Development of the equine ovary and ovulation fossa. Journal of reproduction and fertility. Supplement (27:): 471–477. PMID: 289826

3. Scarlet D, Walter I, Hlavaty J, Aurich C (2015) Expression and immunolocalisation of follicle-stimulating hormone receptors in gonads of newborn and adult female horses. Reproduction, fertility, and development.

4. Cai J, Li T, Huang B, Cheng H, Ding H, Dong W et al. (2014) The use of laser microdissection in the identification of suitable reference genes for normalization of quantitative real-time PCR in human FFPE epithelial ovarian tissue samples. PloS one 9 (4): e95974. doi: 10.1371/journal.pone.0095974 PMID: 24776823

5. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry 55 (4): 611–622. doi: 10.1373/clinchem.2008.112797 PMID: 19246619

6. Eisenberg E, Levanon EY (2013) Human housekeeping genes, revisited. Trends in genetics: TIG 29 (10): 569–574. doi: 10.1016/j.tig.2013.05.010 PMID: 23810203

7.   Ferguson BS, Nam H, Hopkins RG, Morrison RF (2010) Impact of reference gene selection for target gene normalization on experimental outcome using real-time qRT-PCR in adipocytes. PloS one 5 (12): e15208. doi: 10.1371/journal.pone.0015208 PMID: 21179435

8.   Pan H, Yang X, Siegfried BD, Zhou X (2015) A Comprehensive Selection of Reference Genes for RT-qPCR Analysis in a Predatory Lady Beetle, Hippodamia convergens (Coleoptera: Coccinellidae). PLoS ONE 10 (4): e0125868. doi: 10.1371/journal.pone.0125868 PMID: 25915640

9.   Saint-Marcoux D, Proust H, Dolan L, Langdale JA (2015) Identification of reference genes for real-time quantitative PCR experiments in the liverwort Marchantia polymorpha. PLoS ONE 10 (3): e0118678. doi: 10.1371/journal.pone.0118678 PMID: 25798897

10. Thomas KC, Zheng XF, Garces Suarez F, Raftery JM, Quinlan KG, Yang N et al. (2014) Evidence based selection of commonly used RT-qPCR reference genes for the analysis of mouse skeletal muscle. PLoS ONE 9 (2): e88653. doi: 10.1371/journal.pone.0088653 PMID: 24523926

11. Sun M, Lu M, Tang X, Du Y (2015) Exploring valid reference genes for quantitative real-time PCR analysis in Sesamia inferens (Lepidoptera: Noctuidae). PLoS ONE 10 (1): e0115979. doi: 10.1371/journal.pone.0115979 PMID: 25585250

12. Dijkstra JR, van Kempen LC, Nagtegaal ID, Bustin SA (2014) Critical appraisal of quantitative PCR results in colorectal cancer research: can we rely on published qPCR results. Mol Oncol 8 (4): 813–818. doi: 10.1016/j.molonc.2013.12.016 PMID: 24423493

13. Goldsworthy SM, Goldsworthy TL, Sprankle CS, Butterworth BE (1993) Variation in expression of genes used for normalization of Northern blots after induction of cell proliferation. Cell Prolif 26 (6): 511–518. PMID: 9116118

14. Bhatia P, Taylor WR, Greenberg AH, Wright JA (1994) Comparison of glyceraldehyde-3-phosphate dehydrogenase and 28S-ribosomal RNA gene expression as RNA loading controls for northern blot analysis of cell lines of varying malignant potential. Anal Biochem 216 (1): 223–226. PMID: 8135355

15. Savonet V, Maenhaut C, Miot F, Pirson I (1997) Pitfalls in the use of several "housekeeping" genes as standards for quantitation of mRNA: the example of thyroid cells. Anal. Biochem. 247 (1): 165–167. PMID: 9126387

16. Smits K, Goossens K, van Soom A, Govaere J, Hoogewijs M, Vanhaesebrouck E et al. (2009) Selection of reference genes for quantitative real-time PCR in equine in vivo and fresh and frozen-thawed in vitro blastocysts. BMC Res Notes 2: 246. doi: 10.1186/1756-0500-2-246 PMID: 20003356

17. Reddy DS, Bhatnagar-Mathur P, Cindhuri KS, Sharma KK (2013) Evaluation and validation of reference genes for normalization of quantitative real-time PCR based gene expression studies in peanut. PLoS ONE 8 (10): e78555. doi: 10.1371/journal.pone.0078555 PMID: 24167633

18. Hao X, Horvath DP, Chao WS, Yang Y, Wang X, Xiao B (2014) Identification and Evaluation of Reliable Reference Genes for Quantitative Real-Time PCR Analysis in Tea Plant (Camellia sinensis (L.) O. Kuntze). International Journal of Molecular Sciences 15 (12): 22155–22172. doi: 10.3390/ijms151222155 PMID: 25474086

19. Omondi BA, Latorre-Estivalis JM, Rocha Oliveira IH, Ignell R, Lorenzo MG (2015) Evaluation of reference genes for insect olfaction studies. Parasit Vectors 8: 243. doi: 10.1186/s13071-015-0862-x PMID: 25896676

20. Hruz T, Wyss M, Docquier M, Pfaffl MW, Masanetz S, Borghi L et al. (2011) RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC genomics 12: 156. doi: 10.1186/1471-2164-12-156 PMID: 21418615

21. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T et al. (2014) A comparative encyclopedia of DNA elements in the mouse genome. Nature 515 (7527): 355–364. doi: 10.1038/nature13992 PMID: 25409824

22. Pervouchine DD, Djebali S, Breschi A, Davis CA, Barja PP, Dobin A et al. (2015) Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. Nat Commun 6: 5903. doi: 10.1038/ncomms6903 PMID: 25582907

23. Vandesompele J, de Preter K, Pattyn F, Poppe B, van Roy N, De Paepe A et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3 (7): RESEARCH0034. PMID: 12184808

24. Andersen CL, Jensen JL, Orntoft TF (2004) 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): 5245–5250. PMID: 15289330

25. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pairwise correlations. Biotechnol Lett 26 (6): 509–515. PMID: 15127793

26. Chen D, Pan X, Xiao P, Farwell MA, Zhang B (2011) Evaluation and identification of reliable reference genes for pharmacogenomics, toxicogenomics, and small RNA expression analysis. Journal of cellular physiology 226 (10): 2469–2477. doi: 10.1002/jcp.22725 PMID: 21412770

27. Svingen T, Letting H, Hadrup N, Hass U, Vinggaard AM (2015) Selection of reference genes for quantitative RT-PCR (RT-qPCR) analysis of rat tissues under physiological and toxicological conditions. PeerJ 3: e855. doi: 10.7717/peerj.855 PMID: 25825680

28. Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A (2004) Validation of housekeeping genes for normalizing RNA expression in real-time PCR. BioTechniques 37 (1): 112–4, 116, 118–9. PMID: 15283208

29. Fox-Walsh KL, Dou Y, Lam BJ, Hung S, Baldi PF, Hertel KJ (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci USA 102 (45): 16176–16181. PMID: 16260721

30. Herrera-Luna CV, Budik S, Aurich C (2012) Gene expression of ACTH, glucocorticoid receptors, 11βHSD enzymes, LH-, FSH-, GH receptors and aromatase in equine epididymal and testicular tissue. Reproduction in domestic animals 47 (6): 928–935. doi: 10.1111/j.1439-0531.2012.01993.x PMID: 22335522

31. Ahn K, Bae J, Nam K, Lee C, Park K, Lee H et al. (2011) Identification of reference genes for normalization of gene expression in thoroughbred and Jeju native horse(Jeju pony) tissues. Genes Genom 33 (3): 245–250.

32. Beekman L, Tohver T, Dardari R, Léguillette R (2011) Evaluation of suitable reference genes for gene expression studies in bronchoalveolar lavage cells from horses with inflammatory airway disease. BMC molecular biology 12: 5. doi: 10.1186/1471-2199-12-5 PMID: 21272375

33. de Jonge HJ, Fehrmann RS, de Bont ES, Hofstra RM, Gerbens F, Kamps WA et al. (2007) Evidence based selection of housekeeping genes. PloS one 2 (9): e898. PMID: 17878933

34. Meyer FR, Grausgruber H, Binter C, Mair GE, Guelly C, Vogl C et al. (2013) Cross-platform microarray meta-analysis for the mouse jejunum selects novel reference genes with highly uniform levels of expression. PloS one 8 (5): e63125. doi: 10.1371/journal.pone.0063125 PMID: 23671661

35. Kwon MJ, Oh E, Lee S, Roh MR, Kim SE, Lee Y et al. (2009) Identification of novel reference genes using multiplatform expression data and their validation for quantitative gene expression analysis. PloS one 4 (7): e6162. doi: 10.1371/journal.pone.0006162 PMID: 19584937

36. Vassetzky NS, Kramerov DA (2013) SINEBase: a database and tool for SINE analysis. Nucleic acids research 41 (Database issue): D83–9. doi: 10.1093/nar/gks1263 PMID: 23203982

37. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research 31 (13): 3406–3415. PMID: 12824337

38. Laurell H, Iacovoni JS, Abot A, Svec D, Maoret J, Arnal JF et al. (2012) Correction of RT-qPCR data for genomic DNA-derived signals with ValidPrime. Nucleic Acids Res. 40 (7): e51. doi: 10.1093/nar/gkr1259 PMID: 22228834

39. De Keyser E, Desmet L, Van Bockstaele E, De Riek J (2013) How to perform RT-qPCR accurately in plant species? A case study on flower colour gene expression in an azalea (Rhododendron simsii hybrids) mapping population. BMC Mol. Biol. 14: 13. doi: 10.1186/1471-2199-14-13 PMID: 23800303

40. Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology (N Y) 11 (9): 1026–1030.

41. Kenkel CD, Aglyamova G, Alamaru A, Bhagooli R, Capper R, Cunning R et al. (2011) Development of gene expression markers of acute heat-light stress in reef-building corals of the genus Porites. PloS one 6 (10): e26914. doi: 10.1371/journal.pone.0026914 PMID: 22046408

42. Okamura Y, Aoki Y, Obayashi T, Tadaka S, Ito S, Narise T et al. (2015) COXPRESdb in 2015: coexpression database for animal species by DNA-microarray and RNAseq-based expression data with multiple quality assessment systems. Nucleic acids research 43 (Database issue): D82–6. doi: 10.1093/nar/gku1163 PMID: 25392420

43. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research 38: W214–20. doi: 10.1093/nar/gkq537 PMID: 20576703

44. Hoque M, Ji Z, Zheng D, Luo W, Li W, You B et al. (2013) Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. Nature methods 10 (2): 133–139. doi: 10.1038/nmeth.2288 PMID: 23241633

45. Hallegger M, Llorian M, Smith Christopher W J (2010) Alternative splicing: global insights. The FEBS journal 277 (4): 856–866. doi: 10.1111/j.1742-4658.2009.07521.x PMID: 20082635

46. Sterner DA, Carlo T, Berget SM (1996) Architectural limits on split genes. Proc Natl Acad Sci U S A 93 (26): 15081–15085. PMID: 8986767

47. Bustin SA (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 29 (1): 23–39. PMID: 12200227

48. Martínez-Beamonte R, Navarro MA, Larraga A, Strunk M, Barranquero C, Acin S et al. (2011) Selection of reference genes for gene expression studies in rats. J. Biotechnol. 151 (4): 325–334. doi: 10.1016/j.jbiotec.2010.12.017 PMID: 21219943

49. Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM et al. (2011) PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. The Plant cell 23 (3): 895–910. doi: 10.1105/tpc.111.083667 PMID: 21441431

50. Aoki K, Ogata Y, Shibata D (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. Plant & cell physiology 48 (3): 381–390.

51. Obayashi T, Kinoshita K (2009) Rank of Correlation Coefficient as a Comparable Measure for Biological Significance of Gene Coexpression. DNA Res 16 (5): 249–260. doi: 10.1093/dnares/dsp016 PMID: 19767600