Check for updates

RESEARCH ARTICLE

REVISED

# Identification of qPCR reference genes suitable for normalising gene expression in the developing mouse embryo [version 2; peer review: 3 approved]

John C.W. Hildyard [iD] [1], Dominic J. Wells [iD] [2], Richard J. Piercy [iD] [1]

[1]Comparative Neuromuscular Diseases Laboratory, Clinical Science and Services, Royal Veterinary College, London, NW1 0TU, UK
[2]Comparative Biomedical Sciences, Royal Veterinary College, London, London, NW1 0TU, UK

## Abstract

**Background**: Progression through mammalian embryogenesis involves many interacting cell types and multiple differentiating cell lineages. Quantitative polymerase chain reaction (qPCR) analysis of gene expression in the developing embryo is a valuable tool for deciphering these processes, but normalisation to stably-expressed reference genes is essential for such analyses. Gene expression patterns change globally and dramatically as embryonic development proceeds, rendering identification of consistently appropriate reference genes challenging.

**Methods**: We have investigated expression stability in mouse embryos from mid to late gestation (E11.5–E18.5), both at the whole-embryo level, and within the head and forelimb specifically, using 15 candidate reference genes ( *ACTB, 18S, SDHA, GAPDH, HTATSF1, CDC40, RPL13A, CSNK2A2, AP3D1, HPRT1, CYC1, EIF4A, UBC, B2M* and *PAK1IP1*), and four complementary algorithms (geNorm, Normfinder, Bestkeeper and deltaCt).

**Results**: Unexpectedly, all methods suggest that many genes within our candidate panel are acceptable references, though *AP3D1, RPL13A* and *PAK1IP1* are the strongest performing genes overall (scoring highly in whole embryos, heads or forelimbs alone, and in all samples collectively). *HPRT1* and *B2M* are conversely poor choices, and show strong developmental regulation. We further show that normalisation using our three highest-scoring references can reveal subtle patterns of developmental expression even in genes ostensibly ranked as acceptably stable ( *CDC40, HTATSF1*).

**Conclusion**: *AP3D1, RPL13A* and *PAK1IP1* represent universally suitable reference genes for expression studies in the E11.5-E18.5 mouse embryo.

## Keywords

Embryogenesis, Embryo, mouse, qPCR, reference genes, mammal, embryonic development

**Open Peer Review**

**Approval Status** ✓ ✓ ✓

|  | 1 | 2 | 3 |
|---|---|---|---|
| **version 2** (revision) 21 Sep 2022 |  | ✓ view | ✓ view |
| **version 1** 03 Aug 2021 | ✓ view | ? view | ? view |

1. **Caroline L. Smith** [iD], University of Westminster, London, UK

2. **Anush Kosakyan**, Biology Centre Czech Academy of Science, České Budějovice, Czech Republic

3. **Liang Hu** [iD], Reproductive and Genetic Hospital of CITIC-XIANGYA, Changsha, China

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** John C.W. Hildyard (jhildyard@rvc.ac.uk)

**Author roles: Hildyard JCW**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Visualization, Writing – Original Draft Preparation; **Wells DJ**: Investigation, Resources, Supervision, Writing – Review & Editing; **Piercy RJ**: Funding Acquisition, Supervision, Writing – Review & Editing

**How to cite this article:** Hildyard JCW, Wells DJ and Piercy RJ. **Identification of qPCR reference genes suitable for normalising gene expression in the developing mouse embryo [version 2; peer review: 3 approved]** Wellcome Open Research 2022, **6**:197 https://doi.org/10.12688/wellcomeopenres.16972.2

**First published:** 03 Aug 2021, **6**:197 https://doi.org/10.12688/wellcomeopenres.16972.1

**REVISED** **Amendments from Version 1**

We would like to thank all three reviewers for their excellent questions and suggestions, which we have implemented accordingly.

We have made several minor edits to the text to better place our work in context, expanded the table of reference genes to include accession numbers and amplicon sizes, made minor edits and clarifications to Figure 2, and provided additional extended data files detailing efficiency calculations for the qPCR primer pairs used.

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

The mouse (*Mus musculus*) is the principal model organism for the study of mammalian embryonic development, offering large litter sizes and a predictable gestation period (mating in this species can moreover be timed with relative ease). The stages of mouse embryonic development are thus well documented[1,2], and given the conserved nature of mammalian embryogenesis, can also be mapped across mammalian species despite gestation times differing by more than an order of magnitude (20 days in the mouse, 270 days in the human, up to 645 days in the African elephant *Loxodonta*)[3].

The earliest stages of development establish fundamental morphological patterns: blastocyst formation, implantation and establishment of polarity occur with the first five days (Embryonic days 1–5, or E1-E5)[4], with gastrulation and development of the primitive streak following (E6-E8.5)[5]. Intermediate stages (E8.5-E10.5) feature the cyclical 'clock and wave' of somitogenesis[6], the turning of the embryo, neural tube closure and the laying down of organ precursor cell lineages, while the later stages (E11.5 onward) involve maturation and development of those organs[1,2]. During this later period (from E11.5 to E18.5, see Figure 1) almost all limb development occurs, from a primitive limb bud to an essentially mature state complete with ossifying bones and fully-defined joints and digits[7–9], and similarly almost all skeletal muscle is laid down (both primary and secondary myogenesis)[10,11]. The head undergoes a broad panel of changes, with the brain enlarging and maturing[12], the palate closing and teeth forming[13], the ears emerging, and during this period the eye progresses from a simple indentation to a fully encapsulated globe (which is subsequently sheltered beneath fused eyelids)[14]. More globally, this later period also spans substantial changes in haematopoiesis (with the source of blood cells switching from the yolk sac to the liver, and then subsequently to the bone marrow)[15], maturation of the chambers of the heart[16], gradual replacement of the mesonephros with the metanephros[17], and establishment of the skin barrier[1,18].

Although the morphological and developmental changes during embryogenesis are well-characterised at the histological level, understanding at the molecular level is less comprehensive. The interactions of conflicting growth factors, signalling cascades and downstream transcriptional programs play a critical role in coordination of development, but the diversity of cell types and the often transient nature of such interactions renders these processes challenging to investigate. More focussed data can be obtained by careful isolation of specific embryonic tissues, however not all tissues are tractable to such an approach, and restricting analysis in this fashion loses wider developmental context. Recent technological advances have permitted elegant studies of embryonic gene expression down to single-cell level[19,20], suggesting that detailed study of the interactions between individual cell lineages might now be possible, however such approaches also represent a substantial undertaking both in time and resources. Furthermore, these approaches are not without limitation: the whole transcriptome nature of the method typically favours breadth over depth[21,22]: large-scale changes involving many gene components will be mapped well, while more specific, smaller-scale processes can be missed. In particular, constraints in read depth mean low abundance transcripts can be under-represented in such datasets[23], and thus more conventional methods remain vital research tools.

Measurement of gene expression via qPCR (using cDNA prepared from extracted RNA) is a comparatively inexpensive and flexible means of measuring expression of specific gene targets, including rare transcripts expressed at low levels or restricted to minority cell populations. Given the quantitative nature of this technique, normalisation of expression data is essential: RNA extraction efficiency, RNA integrity and cDNA synthesis efficiency all represent sources of variability that can obscure genuine changes in gene expression, or create the appearance of change where none exists (best summarised in the MIQE guidelines[24]). Normalisation can be conducted using reference genes: genes known to be stably expressed under the conditions studied. RNA is an inherently labile molecule, however, and cellular turnover at the transcript level is often both more dramatic, and more rapid, than at the protein level. Genes considered broadly appropriate for protein normalisation such as *GAPDH* and *ACTB* (beta-actin) are historically popular choices, but we and others have shown these often perform poorly as references[25–31]. 18S ribosomal RNA is also used widely, but this necessitates use of random priming in cDNA preparation (rRNA lacks polyA tails, thus oligo dT priming will not reverse transcribe *18S*). Synthesis and degradation of ribosomal RNA also differs from that of mRNA[32], and the sheer abundance of rRNA sequences could mask marked changes in the mRNA pool (ribosomes represent ~80–90% of total cellular RNA, while mRNA accounts for ~5%, thus mRNA levels could change by a factor of two or more without significantly altering measured rRNA content). Crucially, reference genes suitable for one comparative scenario are not guaranteed to be appropriate for another, and identification of reference genes appropriate for the conditions studied represents a key step in any qPCR-based investigation. Identification of appropriate reference genes *a priori* is challenging, and various mathematical approaches exist: the geNorm[33], Normfinder[34], Bestkeeper[35], and deltaCt[36] methods all require a representative collection of cDNA samples, and a broad panel of candidate reference genes,
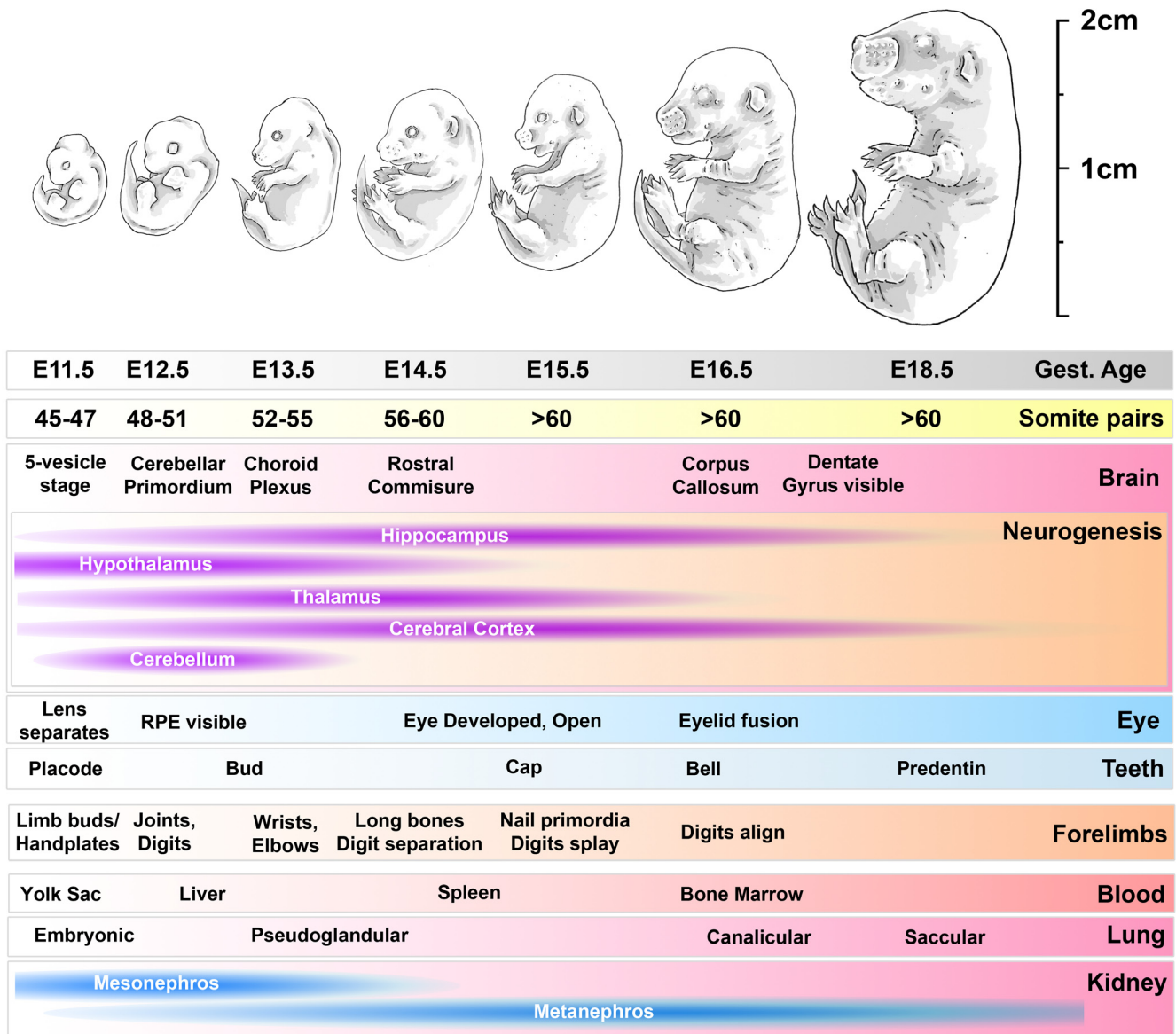
**Figure 1. developmental timeline of mouse embryos used in this study.** Progress from E11.5 to E18.5 involves marked increases in size and mass and progressive accumulation of somite pairs. Within the head, multiple defined brain structures emerge over this period, and neurogenesis peaks in different brain regions (as indicated). The eye matures, as do teeth. Within the forelimbs, development proceeds from primitive limb bud through to joint and digit formation, maturation and ossification of long bones and emergence of nail primordia. Within the body, the source of haematopoiesis shifts from the yolk sac to the liver, and thence transiently to the spleen before establishing in the bone marrow; the lungs mature through primitive pseudoglandular stages to saccular morphology; the mesonephros of the kidney degenerates and is replaced with the final metanephros. RPE: retinal pigmented epithelium.

but each assesses suitability via subtly different criteria (see *Extended data*: reference gene analysis packages[37] for a more detailed overview). Combining these complementary approaches increases the power of investigations: individual rankings might differ between methods, but truly strong candidates should consistently score highly regardless of assessment method (and discrepancies between methods can moreover highlight interesting biological information, as reported previously[26,38]).

The developing embryo represents an especially dynamic transcriptional environment, transcriptional changes are likely to be dramatic as development proceeds, making determination of reference genes especially critical; such developmental changes might be even more profound in a mammal with a short gestation like the mouse. Several studies have investigated appropriate references for earlier developmental stages, with a focus on early morula/blastocyst formation[39–41], or progression

from blastocyst (E3.5) to mid organogenesis (E11.5)[42,43], though these studies have produced conflicting results, with *ACTB* scoring highly in some, but poorly in others (some investigators have reported that embryonic reference gene suitability might even vary by mouse strain[40]). Less attention has been dedicated to later stages of embryogenesis, where cell and organ lineages are more established, and developmental changes become consequently more focussed and tissue-specific. Several studies have been conducted on specific organs (such as the developing gonads[44], heart[45] or thymus[46]), but no specific study has addressed reference genes appropriate for normalisation at the whole embryo level at these later stages.

We have thus investigated reference genes appropriate for normalising gene expression in whole mouse embryos collected at E11.5, E12.5, E13.5, E14.5, E15.5, E16.5 and E18.5, spanning development from mid to late gestation. Given the marked developmental changes that occur within the head and the forelimbs over this period (Figure 1), we have also determined reference genes appropriate for these specific tissues (at E13.5, E16.5 and E18.5). Our collection of samples is substantial (N=3–5 per time point) and we have used a broad panel of 15 candidate reference genes (*ACTB, 18S, SDHA, GAPDH, HTATSF1, CDC40, RPL13A, CSNK2A2, AP3D1, HPRT1, CYC1, EIF4A, UBC, B2M* and *PAK1IP1* -see Table 1). This panel includes candidates widely used historically (*18S, GAPDH, ACTB*)[47,48], those that we have shown be strong references in mouse skeletal muscle (*CSNK2A2, AP3D1, RPL13A*)[25,38], and also shown by others to be viable references in rodent brains

(*UBC, HPRT1, RPL13A, 18S*)[49,50]. To maximise the power of our study, we have assessed reference gene suitability using all four algorithms described above (geNorm, deltaCt, BestKeeper and Normfinder).

## Methods
### Ethics statement
We have previously demonstrated that N values of 3–5 per time point/tissue are sufficient to obtain robust assessments of reference gene suitability[26,38], and to permit nuanced comparative analysis. Given expected litter sizes of 5–9, we determined that 10 pregnant females would be required to generate sufficient samples (seven whole embryo time points, three head/forelimb time points), A total of 70 mouse embryos (strain C57BL/10) were obtained post-mortem from these 10 pregnant females: 35 embryos were used for this study (all additional embryos were collected for a separate study). All mice were bred under UK Home Office Project Licence, approved by the Royal Veterinary College Animal Welfare and Ethical Review Board (AWERB review number RVC 2018-0113N). Mice were held in individually ventilated cages in a minimal disease unit at an average 21°C in a 12 hours light/12 hours dark light cycle with food and water provided ad-lib. Mating trios (one male, 2 females) were set up and monitored each morning until all females had been mated (as shown by a vaginal plug). All efforts were made to ameliorate any suffering of animals. At the required gestational age, pregnant females were killed (schedule 1) via cervical dislocation, with all embryos then killed via hypothermia and exsanguination.

**Table 1. list of candidate genes and their full names/functions, accession numbers and PCR amplicon sizes.**

| Gene name | Full name/function | Accession number | Amplicon (bp) |
|---|---|---|---|
| ACTB | Beta-Actin: cytoskeletal component | NM_007393.3 | 94 |
| 18S | 18S ribosomal RNA: small ribosomal subunit | NR_003278.3 | 151 |
| SDHA | Succinate Dehydrogenase subunit A: mitochondrial complex II | NM_023281.1 | 181 |
| GAPDH | Glyceraldehyde 3-Phosphate Dehydrogenase: glycolysis pathway | NM_008084.2 | 180 |
| HTATSF1 | HIV-1 Tat-specific factor 1: transcriptional elongation/splicing factor | NM_028242 | 136 |
| CDC40 | Cell Division Cycle 40: cell cycle/spliceosome factor | NM_027879 | 167 |
| RPL13a | Ribosomal Protein L13: component of large ribosomal subunit | NM_009438.5 | 180 |
| CSNK2A2 | Casein Kinase 2 Alpha 2: subunit of casein kinase | NM_009974 | 177 |
| AP3D1 | Adaptor Related Protein Complex 3 Subunit Delta 1: vesicle trafficking | NM_007460 | 183 |
| HPRT1 | Hypoxanthine Phosphoribosyltransferase 1: purine salvage pathway | NM_013556 | 195 |
| CYC1 | Cytochrome C1: mitochondrial complex III | NM_025567.2 | 141 |
| EIF4A | Eukaryotic Initiation Factor 4A: helicase component of translation initiation complex | NM_013506.2 | 152 |
| UBC | Ubiquitin C | NM_019639.4 | 129 |
| B2M | Beta 2 Microglobulin: component of MHC class 1 complex | NM_009735.3 | 159 |
| PAK1IP1 | PAK1 Interacting Protein: negative regulator of PAK1 kinase | NM_026550 | 114 |

## Sample collection and study design

Matings were assumed to occur at midnight, thus first detection of a vaginal plug was designated as 0.5 days post coitum, or E0.5, with subsequent days incremented accordingly (E1.5, E2.5 etc). Seven time points were used for whole embryo samples (E11.5, E12.5, E13.5, E14.5, E15.5, E16.5 and E18.5) with E13.5, E16.5 and E18.5 also used for head and forelimb samples. A minimum of three embryos were used for each time point, though where litter sizes permitted, greater numbers were collected (Whole embryos: E13.5 N=4, E18.5 N=5; Heads: E18.5 N=5, for a total of 24 whole embryo samples, 11 head samples and nine forelimb samples – see *Extended data*: animal numbers and sample sizes[51]). Given potential variation in precise mating, ovulation and conception times, gestational ages should be considered approximate, particularly where different litters are pooled for a single time point (E13.5, E14.5, E18.5). At the appropriate gestational stage (all sample collections were performed between midday and 2pm), pregnant females were killed as described above. Uterine horns were quickly removed and placed on ice (additional organs and muscles were collected from the adults for separate studies). After death, embryos were dissected from their uterine environments (including removal from amniotic sac) and either kept intact (whole embryos), or further dissected to isolate forelimbs (each sample used both forelimbs) and heads (remaining tissue from the body was also collected). All samples were snap-frozen in liquid nitrogen and stored at -80°C until use.

## RNA isolation and qPCR

Tissues (whole embryos, heads, or forelimbs) were pulverised via dry-ice-cooled cell-crusher (Cellcrusher Ltd), and ~100-200mg frozen tissue powder (well mixed to ensure representative sampling) was used to prepare RNA. RNA was isolated using TRIzol (Invitrogen) as described previously[26,38], with inclusion of an additional chloroform extraction (1:1) after the phase separation step and inclusion of 10µg glycogen during precipitation to maximise RNA yield. RNA yield and purity were assessed via nanodrop (ND1000) and samples with 260/230 ratios below 1.7 were subjected to a second precipitation. All cDNA was prepared using the RTnanoscript2 kit (Primerdesign), using 1600ng of RNA per reaction, with oligo dT and random nonamer priming. All cDNA samples were subsequently diluted 1/20 with nuclease-free water to minimise downstream PCR inhibition. qPCR reactions were performed in duplicate or triplicate in 10µl volumes using 2µl diluted cDNA (~8ng cDNA per well assuming 1:1 conversion) in a CFX384 lightcycler using PrecisionPLUS SYBR green qPCR mastermix (Primerdesign), with a melt curve included as standard (See *Underlying data*: qPCR raw data[52]). All Cq values were determined via linear regression. Where necessary, between-plate correction was conducted using three shared calibration samples. Primers to *ACTB, SDHA, GAPDH, HTATSF1, CDC40, RPL13A, CSNK2A2, AP3D1, CYC1, EIF4A, UBC, B2M* and *PAK1IP1* were taken from the geNorm and geNorm PLUS kits (Primerdesign): all give efficiencies of 95–105% and produce single amplicons (see *Extended data*: additional reference gene validation[53]). Sequences are proprietary

but anchor nucleotides and context sequences are available[38]. Primers to *18S* were those used previously[25], and those to *HPRT1* were the pan-species set (HPSF) validated by Valadan *et al.*[54] Primers targeting the 5' terminus of dystrophin dp71 were designed using Primer3 (v 4.1.0). All sequences are provided below:

18S F 5'-GTAACCCGTTGAACCCCATT-3'

18S R 5'-CCATCCAATCGGTAGTAGCG-3'

HPSF F 5'-GGACTAATTATGGACAGGACTG-3'

HPSF R 5'-GCTCTTCAGTCTGATAAAATCTAC-3'

Dp71 F 5'-GTGAAACCCTTACAACCATGAG-3'

Dp71 R 5'-CTTCTGGAGCCTTCTGAGC-3'

## Data analysis

All data was analysed using Microsoft Excel, using four different reference gene analysis approaches: geNorm[33], deltaCt[36], BestKeeper[35] and Normfinder[34]. BestKeeper and deltaCt methods used mean Cq values, while for geNorm and Normfinder Cq values were first linearised to relative quantities (RQ). All data were analysed within a single dataset (44 samples) or as embryos (24 samples), heads (11) or forelimbs (9) alone. Normfinder analysis further allows designation of user-defined groups: accordingly, data was analysed ungrouped as for the other packages above, or grouped as follows. Whole dataset, grouped by tissue type or age; whole embryos, grouped by age; heads grouped by age; forelimbs grouped by age.

To integrate the outputs of all approaches, the per-gene geometric mean was generated from the scores of each specific comparison (whole dataset, embryos, heads, forelimbs). Bestkeeper ranks genes by coefficient of correlation (where low values represent poor correlation), while all other methods rank by stability (were low values represent high stability). Accordingly, values for Bestkeeper were inverted (1-value), with any negative correlation coefficients first set to zero (to give a subsequent score of 1). GeNorm suggests a best pair: for comparative purposes, each of the best pair were assigned the same score. Data from Normfinder was ungrouped (grouped analysis was not used in the integrated assessment).

For gene candidate validation, the three high scoring candidates were used to generate a normalisation factor (NF: geometric mean of the per-sample RQ values): this factor was used to normalise data from lower scoring candidates (all RQ values are linear, thus normalisation is a division operation). Normalised data were used to assess trends, and pre/post-normalisation values were used to calculate changes in coefficient of variation.

## Statistical analysis

Statistical analysis was performed using Graphpad Prism 8.0 for Pearson correlations and Microsoft Excel for coefficients of variation (free software alternatives such as R and JASP

could also be used). Pearson correlations used all samples in the dataset, comparing geometric mean values using RQs of all three high scoring candidates, and geometric mean values using combinations of RQs from only two genes (as indicated). Dataset coefficients of variation were determined for each time point individually (N=3-5), then time point CoVs were summed to provide the overall variation.

## Results

### Raw Cq values

Raw Cq data (all genes, all samples) serves as a simple first-pass assessment of a reference gene panel (see *Underlying data*[52]). Most genes showed highly consistent expression across all samples (particularly *UBC*), however expression of *18S* was unexpectedly varied (Figure 2A): closer examination of the data suggested that levels of this gene varied not by age or tissue, but by cDNA synthesis batch, with one batch producing *18S* Cqs ~3 cycles later (corresponding to a roughly 10-fold reduction in template). Accordingly, preparation of fresh cDNA for these samples corrected this discrepancy (Figure 2B). Given the comparatively consistent expression for other genes in our panel, we investigated further: restriction to *18S* signal only implied a defect specifically in random priming. Random priming is not required for cDNA synthesis *per se*, but is necessary for reverse transcription of ribosomal RNAs (which lack polyA tails) and more importantly, to capture 5' sequence of longer mRNAs: deficient random priming would be expected to lower apparent expression of such mRNAs. To confirm this we measured expression of the dystrophin isoform dp71: this isoform is modestly expressed in the developing embryo, can only be distinguished from other dystrophin isoforms by primers targeting the unique first exon, and with a transcript ~5kb in length, efficient capture of the 5' terminus requires random priming. As shown (Figure 2C), measured Cq values for this isoform showed a comparable pattern of batch-specific variation which was similarly corrected following preparation of fresh cDNA from these samples.

Assessment of our corrected dataset (Figure 2B) was largely in line with expected expression profiles, with known, highly abundant RNAs (*18S*, *ACTB*, *GAPDH*) showing the lowest mean Cq values. Interestingly, *EIF4A*, an RNA helicase that unwinds mRNA secondary structure for translational initiation, was the least abundant transcript in our panel. Given the fundamental role protein synthesis plays, and the high translational demands of a growing embryo, this finding was unexpected (though we note that even as the lowest expression gene in our panel, levels of this transcript were still comparatively high: mean Cq ~23.9). Studies have suggested that initiation via the eIF4 protein complex involves a high degree of recycling at the 5' UTR[56,57], with the same complex being reused multiple times in rapid succession: initiation complexes might be in lower overall demand than genes associated with metabolism or kinase signalling.
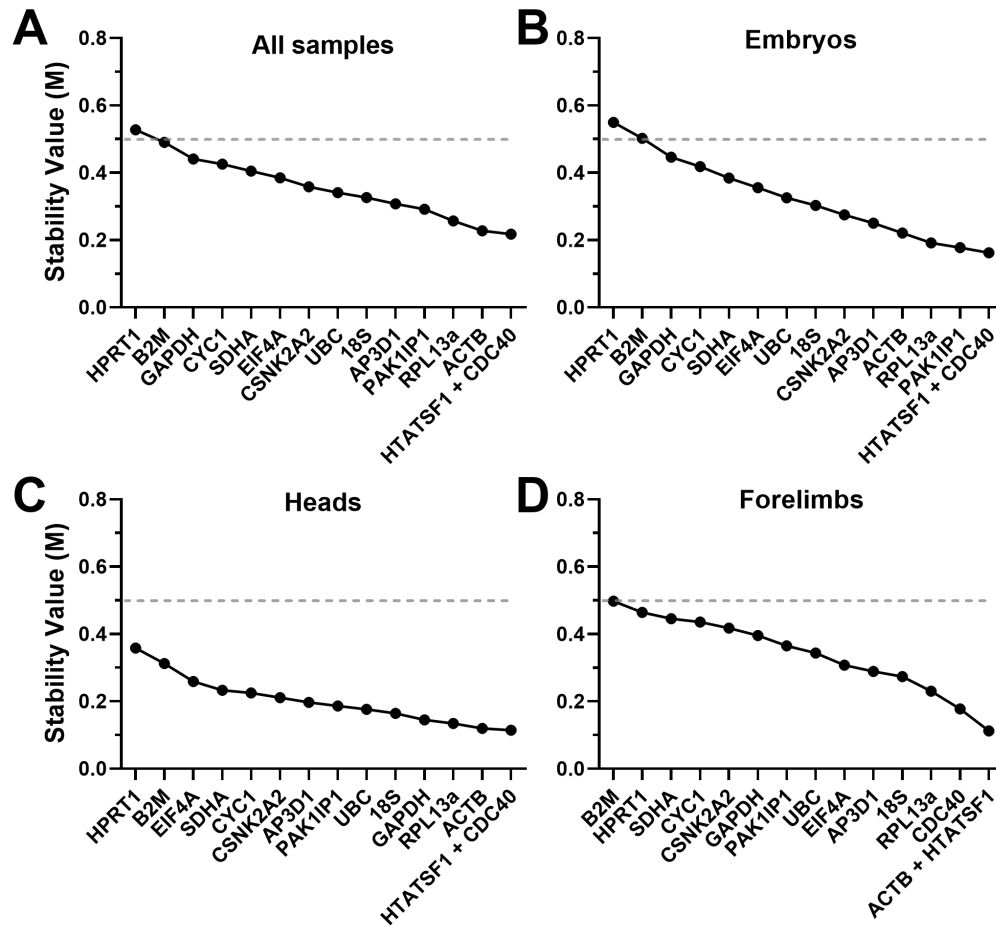
### geNorm analysis

The iterative geNorm algorithm ranks candidate genes by their mean pairwise variation M (where high M represents high variation) to determine the 'best pair', the pair of genes
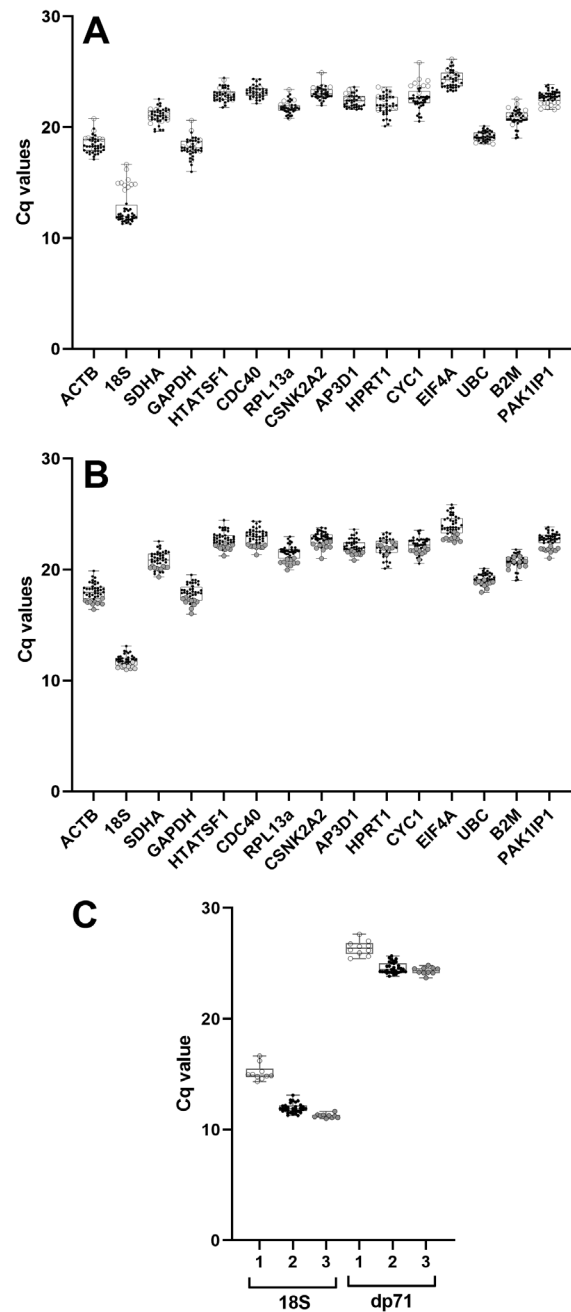


**Figure 2. Raw quantification cycle (Cq) values and validation.** Box and whisker plots of all sample Cq values for each candidate gene used in this study: each point represents the Cq value of an individual sample, for the gene indicated. (**A**) Initial raw Cq values revealed a marked variance in 18S expression (but not other genes), restricted to samples from a single cDNA synthesis batch (open circles), compared to samples from other cDNA synthesis batches (black points), indicative of impaired random priming. (**B**) Preparation of fresh cDNA for these samples corrected this discrepancy (filled circles), producing values consistent with other batches (black points). (**C**) qPCR for 18S and the dystrophin isoform dp71 using the initial poor samples (1, open circles), the other samples (2, black points) and the fresh preparation (3, filled circles) shows that impaired random priming prevents accurate quantification of long transcripts. Details of specific cDNA synthesis batches can be found in the underlying data[55].

**Figure 3. GeNorm rankings.** Rankings and scores of the fifteen candidate genes assessed by the geNorm algorithm, using the entire dataset (**A**), whole embryo samples only (**B**), heads only (**C**) or forelimbs only (**D**). High M values represent less stable genes. Genes with M values <0.5 (dashed line) are considered acceptable references.

that correlate most closely across all samples (see *Underlying data*[58]). Low M values thus represent more suitable genes, but there is no fixed threshold: conventionally, M values <0.5 are considered acceptable as reference genes, while for samples with innately variable expression patterns (such as different tissues) values <1.0 may be accepted. GeNorm analysis (Figure 3, Table 2) revealed high overall stability: assessment of our dataset as a whole (all samples, Figure 3A) revealed 14 of our 15 candidate genes had acceptable M values (M<0.5). *CDC40* and *HTATSF1* were the 'best pair', though *RPL13A*, *PAK1IP1* and *ACTB* also performed well. *CYC1*, *GAPDH*, *B2M* and *HPRT1* all performed poorly, though only the latter failed to clear the M<0.5 threshold.

We repeated the analysis using specific subsets of our data (embryos, heads or forelimbs only: Figure 3B, C and D

respectively). Embryo samples alone gave results very similar to the dataset as a whole, both in ranking and magnitude of stability: *CDC40* and *HTATSF1* were again the best pair, with *RPL13A*, *PAK1IP1* and *ACTB* also scoring highly, while *CYC1*, *GAPDH*, *B2M* and *HPRT1* scored poorly. In heads alone, *CDC40* and *HTATSF1* were also the best pair, with *RPL13A* and *ACTB* being similarly high scoring (though here *GAPDH* also performed well, while *PAK1IP1* ranked poorly), but stability overall was greater in this tissue: all M values were <0.5, with 8 of the 15 showing M values <0.2. Greater overall stability was similarly revealed by analysis of forelimbs: a different best pair was obtained here (*ACTB* replacing *CDC40*), though *CDC40* (along with *RPL13a*) remained high scoring.

Determination of a best pair does not imply that only two reference genes should be used: geNorm analysis also therefore

**Table 2. geNorm rankings.** GeNorm results for the entire dataset or tissue-specific subsets (as indicated), ranked from most stable to least stable. Genes with stability value M > 0.5 –poor scoring candidates- are indicated by italics.

| | All samples | Embryos | Heads | Forelimbs |
|---|---|---|---|---|
| **Best pair** | **HTATSF1 + CDC40** | **HTATSF1 + CDC40** | **HTATSF1 + CDC40** | **ACTB + HTATSF1** |
| Most | ACTB | PAK1IP1 | ACTB | CDC40 |
| stable | RPL13a | RPL13a | RPL13a | RPL13a |
| | PAK1IP1 | ACTB | GAPDH | 18S |
| | AP3D1 | AP3D1 | 18S | AP3D1 |
| | 18S | CSNK2A2 | UBC | EIF4A |
| | UBC | 18S | PAK1IP1 | UBC |
| | CSNK2A2 | UBC | AP3D1 | PAK1IP1 |
| | EIF4A | EIF4A | CSNK2A2 | GAPDH |
| | SDHA | SDHA | CYC1 | CSNK2A2 |
| | CYC1 | CYC1 | SDHA | CYC1 |
| | GAPDH | GAPDH | EIF4A | SDHA |
| Least | B2M | *B2M* | B2M | HPRT1 |
| stable | *HPRT1* | *HPRT1* | HPRT1 | B2M |

determines the change in pairwise variation elicited by increasing the number of reference genes. Typically, variation below 0.2 is considered acceptable, and while our data showed addition of a third reference gene lowered variation, the best pair alone was sufficient in all cases (see *Extended data*: additional reference gene validation[53], Figure E1).

## DeltaCt analysis

The DeltaCt (dCt) method ranks genes by the mean standard deviation of their pairwise Cq differences, with lower values indicating more stable references (see *Underlying data*[58]). Using this method (Figure 4, Table 3), comparable results were obtained from analysis of the dataset as a whole, or restricted to embryos only: sorting by mean deltaCt standard deviation revealed *HPRT1* and *B2M* were the worst performing candidates in both cases, and by a considerable margin. *GAPDH*, *EIF4A* and *CYC1* also scored poorly, but modestly so: indeed, *HPRT1* and *B2M* aside, increases in apparent stability across the panel were slight, suggesting most of our panel represent valid reference genes (as suggested by geNorm, above). *AP3D1*, *PAK1IP1*, *CSNK2A2*, *RPL13A*, *UBC*, *CDC40* and *HTATSF1* all scored highly: despite slight differences in ranking between the whole dataset and embryos alone, all six

genes had very similar scores (ranging from 0.42-0.5). Restricting analysis to the head alone (Figure 4C) produced similar ranking, and moreover increased apparent stability across the entire panel. With the exception of *HPRT1* and *B2M*, essentially every candidate gene represented a strong candidate for normalising gene expression in embryonic head tissues. *HPRT1* and *B2M* also ranked poorest in forelimbs alone (Figure 4D), but here *18S* and *EIF4A* were much higher scoring, while *ACTB* and *CDC40* were not, implying clear differences in expression patterns in these tissues (and a marked departure from geNorm, where *ACTB* formed part of the best pair in forelimbs).

## Bestkeeper analysis

The Bestkeeper method averages the entire presented dataset to generate a consensus profile: the 'Bestkeeper'. Individual gene profiles are then ranked by correlation with this Bestkeeper (see *Underlying data*[58]). As shown above, analysis via geNorm and dCt tended to be more effective at highlighting poor genes from our panel than at identifying consistently strong candidates. Bestkeeper analysis added to this emerging pattern (Figure 5, Table 4): most candidates performed well, with *HPRT1* and *B2M* remaining the exceptions
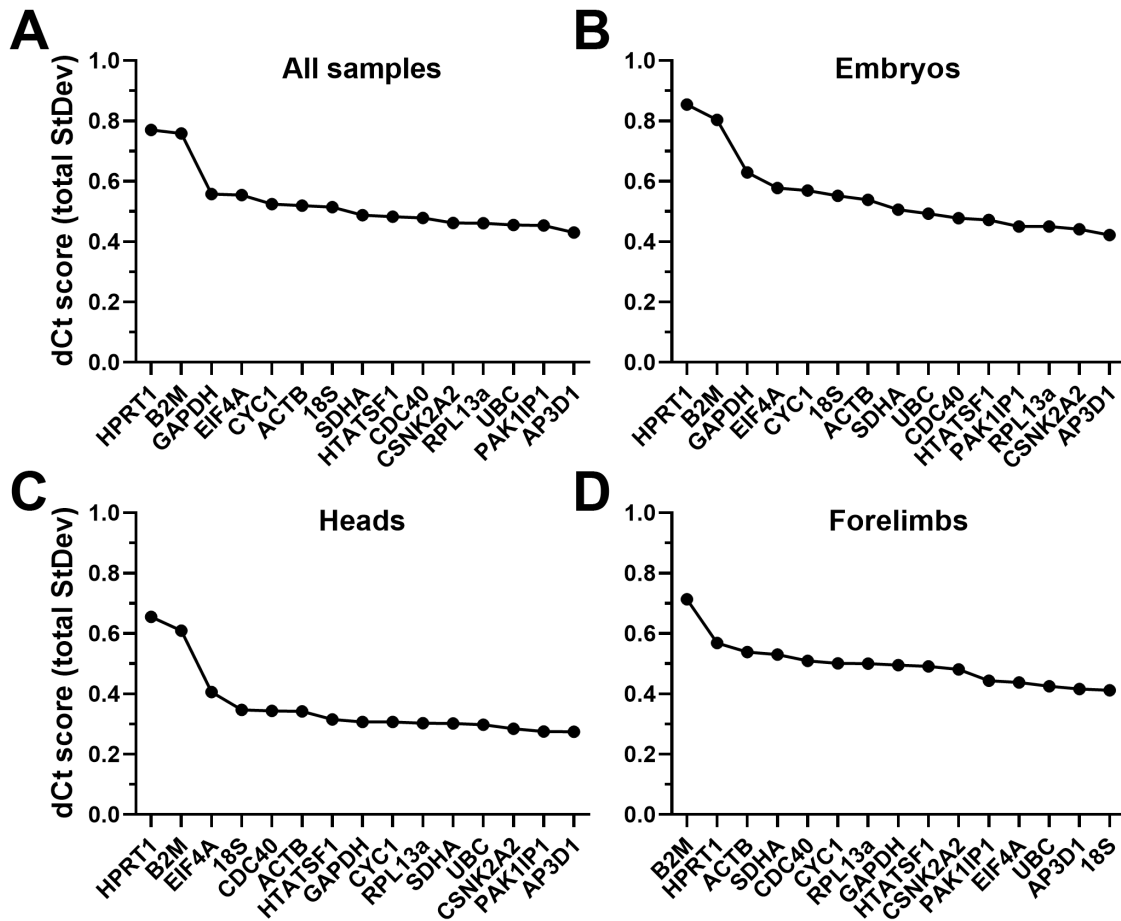
**Figure 4. deltaCt rankings.** Rankings and scores of the fifteen candidate genes assessed by the deltaCt method, using the entire dataset (**A**), whole embryo samples only (**B**), heads only (**C**) or forelimbs only (**D**). High dCt scores represent less stable genes.

(exhibiting negative correlation with the Bestkeeper in the case of heads and forelimbs). Rankings of higher scoring genes disagreed with previous analyses, however: the gene *EIF4A* -ranked lower by both geNorm and dCt- correlated well with a Bestkeeper derived from the whole dataset (or from embryos only), while *ACTB*, *HTATSF1* and *CDC40* correlated less closely. Interestingly, *EIF4A* also correlated closely with a Bestkeeper derived from forelimb data only (Figure 5D), but not one derived from head samples (Figure 5C), while *CYC1* was high ranking in this latter tissue, but not in embryos, forelimbs or the dataset as a whole. *18S* was also closely correlated with the forelimb Bestkeeper (in agreement with dCt) but performed poorly in the other dataset groupings.

Normfinder analysis
Unlike the other three algorithms, Normfinder does not make pairwise comparisons, but instead assesses each gene for individual stability across the dataset. This method also allows two different analysis approaches: grouped and ungrouped. Ungrouped analysis considers the dataset without distinction between samples, while grouped analysis allows discrete subcategories to be defined within the dataset. Accordingly, we used ungrouped analysis to assess our combined dataset and tissue-restricted subsets (Figure 6, Table 5), with further grouped analysis to introduce age and tissue-type as subcategories within those datasets where appropriate (see *Underlying data*[58]). In some respects, ungrouped analysis agreed with the above methods (particularly dCt): *B2M* and *HPRT1* were

**Table 3. DeltaCt rankings.** DeltaCt results for the entire dataset or tissue-specific subsets (as indicated), ranked from most stable to least stable. Genes with mean dCt standard deviations >0.6 –poor scoring candidates- are indicated by italics.

|  | All samples | Embryos | Heads | Forelimbs |
|---|---|---|---|---|
| Most | AP3D1 | AP3D1 | AP3D1 | 18S |
| stable | PAK1IP1 | CSNK2A2 | PAK1IP1 | AP3D1 |
|  | UBC | RPL13a | CSNK2A2 | UBC |
|  | RPL13a | PAK1IP1 | UBC | EIF4A |
|  | CSNK2A2 | HTATSF1 | SDHA | PAK1IP1 |
|  | CDC40 | CDC40 | RPL13a | CSNK2A2 |
|  | HTATSF1 | UBC | CYC1 | HTATSF1 |
|  | SDHA | SDHA | GAPDH | GAPDH |
|  | 18S | ACTB | HTATSF1 | RPL13a |
|  | ACTB | 18S | ACTB | CYC1 |
|  | CYC1 | CYC1 | CDC40 | CDC40 |
|  | EIF4A | EIF4A | 18S | SDHA |
|  | GAPDH | *GAPDH* | EIF4A | ACTB |
| Least | *B2M* | *B2M* | *B2M* | HPRT1 |
| stable | *HPRT1* | *HPRT1* | *HPRT1* | *B2M* |

consistently ranked last (by a considerable margin), while many of the remaining genes had comparable high scores. *EIF4A* ranked lower here in all tissues but forelimbs, while *CNSK2A2*, *PAK1IP1* and especially *AP3D1* all typically performed well (in agreement with dCt and geNorm, but in contrast to Bestkeeper). *18S* was again ranked highest in forelimbs alone, but *ACTB* fared very poorly (in agreement with dCt and Bestkeeper, but not geNorm)

Grouped analysis of our datasets (adding in an age or tissue component to the combined data, and an age component to tissue-specific subsets) differed from ungrouped in several respects (Figure 7, Table 6), suggesting interesting nuances of expression stability within our datasets. When the entire dataset was grouped by tissue (embryo, head, forelimb) no gene had a stability value greater than 0.2, i.e. stability across the panel was substantially greater overall than for the ungrouped dataset, or the same data grouped by age. This suggests tissue-specific differences in gene expression might be mild (and certainly milder than age-specific changes). *HPRT1*, *B2M*, *EIF4A* and *18S* were the lowest ranked genes in both tissue and age-specific grouping, and indeed *HPRT1* and *B2M*

scored poorly in all rankings, with the marked exception of embryos grouped by age (Figure 7C). Here *CYC1*, *PAK1IP1* and *18S* performed worse (exceptionally so, in the case of *18S*). Grouping embryos by age moreover revealed much lower overall stability for the panel than all other grouped analyses: these findings imply global age-related changes are of greatest magnitude, and that *CYC1*, *PAK1IP1* and *18S* in particular might exhibit consistent age-associated changes. Grouped analysis additionally generates a best pair: two genes that may vary between groups, but in opposite directions (given greater overall stability than any individual gene, when combined). In most cases, this best pair did not feature the most stable genes, but as the majority of genes performed well under the majority of comparisons, this was not unexpected (and the stability advantages offered by this best pair were in most cases only modest improvements over the highest scoring candidates alone).

## Summary and validation
Despite the differences in highest scoring candidates between algorithms, all four analysis methods tended to suggest that many of our candidate genes represented stable references. In
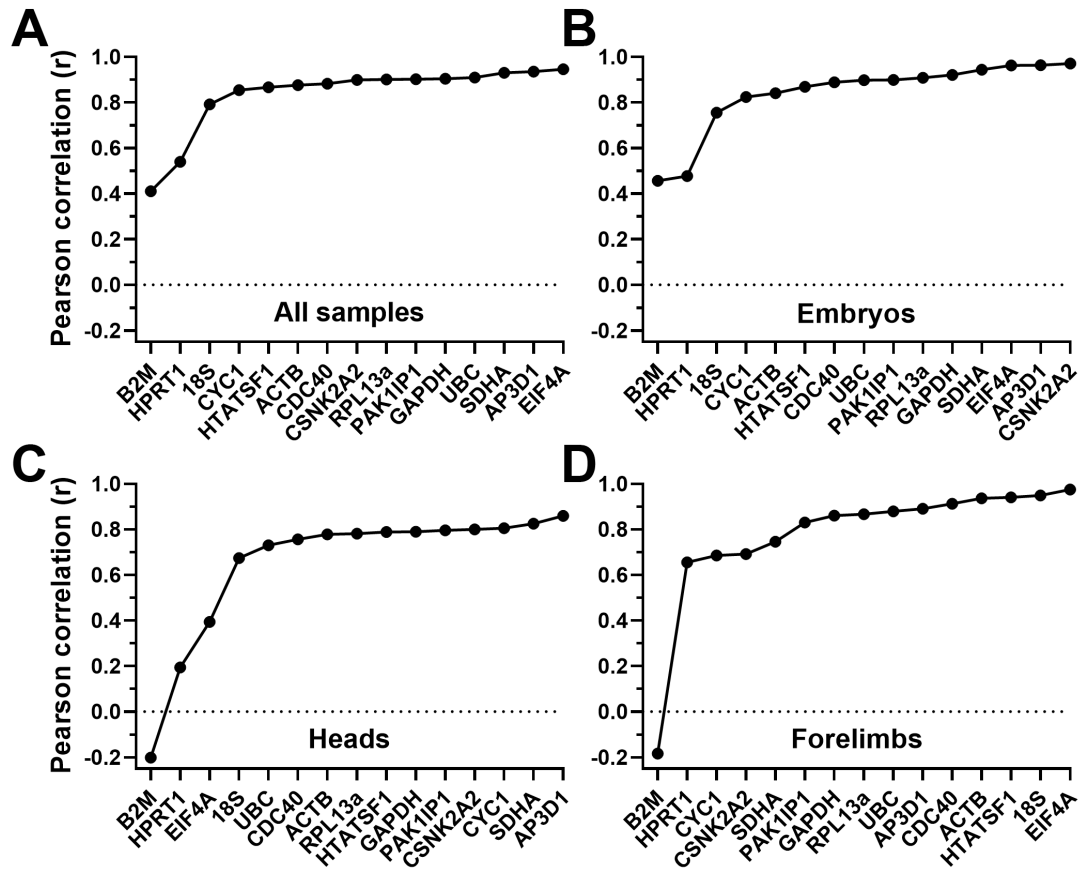
**Figure 5. Bestkeeper rankings.** Rankings and scores of the fifteen candidate genes assessed by the Bestkeeper method, using the entire dataset (**A**), whole embryo samples only (**B**), heads only (**C**) or forelimbs only (**D**). Genes are ranked by their individual Pearson correlation coefficient (r) to the bestkeeper derived from all candidate genes: low r values correspond to less stable genes. Dashed line: r value of zero.

most cases, the highest-ranking candidate was only fractionally more stable than candidates ranked seventh or eighth. Conversely, *HPRT1* and *B2M* were near-universally ranked last (and often by a substantial margin) regardless of method used or tissue assessed, implying strongly that these two genes are inappropriate choices. To illustrate this more clearly, we integrated our collected data in a manner similar to the RefFinder method developed by Xie *et al.*[59], taking the geometric mean of all algorithm scores, either for the entire dataset, or for embryos, heads and forelimbs alone (see Methods). When combined in this fashion (Figure 8, Table 7), the patterns and rankings are remarkably similar for all sets except forelimbs alone. In embryos, heads, or all samples together, *AP3D1* is the strongest candidate, while *RPL13A*, *PAK1IP1* and *CSNK2A2* are also highly ranked. In forelimbs, *EIF4A*, *18S*, *ACTB* and *HTATSF1* are the strongest candidates, however *AP3D1*, *RPL13A* and *PAK1IP1* also score highly in this tissue (while

*CSNK2A2* does not). Taken together, these data suggest that *AP3D1*, *PAK1IP1* and *RPL13A* might represent a universally suitable panel.

For validation, we employed an approach we have used previously[26,38]: using our high scoring candidates (*AP3D1*, *RPL13A*, *PAK1IP1*) to normalise our lowest scoring candidates (*HPRT1* and *B2M*). Raw data for these genes suggested modest expression with a dramatic increase at E18.5, however following normalisation both genes show a progressive increase in expression with increasing age, whether in whole embryos, heads, or limbs (Figure 9). Normalisation would also be expected to reduce the overall coefficient of variation (CoV), and as shown this was indeed the case.

This validation method can be extended: the approaches used by each algorithm are subtly different, and thus information

**Table 4. Bestkeeper rankings.** Bestkeeper results for the entire dataset or tissue-specific subsets (as indicated), ranked by Pearson correlation coefficient (r) with the Bestkeeper, from most stable to least stable. Genes with r <0.6 –poor scoring candidates- are indicated by italics.

| | All samples | Whole embryos | Heads | Forelimbs |
|---|---|---|---|---|
| Most | EIF4A | CSNK2A2 | AP3D1 | EIF4A |
| stable | AP3D1 | AP3D1 | SDHA | 18S |
| | SDHA | EIF4A | CYC1 | HTATSF1 |
| | UBC | SDHA | CSNK2A2 | ACTB |
| | GAPDH | GAPDH | PAK1IP1 | CDC40 |
| | PAK1IP1 | RPL13a | GAPDH | AP3D1 |
| | RPL13a | PAK1IP1 | HTATSF1 | UBC |
| | CSNK2A2 | UBC | RPL13a | RPL13a |
| | CDC40 | CDC40 | ACTB | GAPDH |
| | ACTB | HTATSF1 | CDC40 | PAK1IP1 |
| | HTATSF1 | ACTB | UBC | SDHA |
| | CYC1 | CYC1 | 18S | CSNK2A2 |
| | 18S | 18S | *EIF4A* | CYC1 |
| Least | *HPRT1* | *HPRT1* | *HPRT1* | HPRT1 |
| stable | *B2M* | *B2M* | *B2M* | *B2M* |

can be gleaned by examining cases where candidate rankings disagree. GeNorm, a method that ranks genes by pairwise variation, suggested *HTATSF1* and *CDC40* as the best pair near-unanimously, while these two genes performed only modestly under most other assessments. The implication is that these genes share a very similar pattern of expression, but in a manner unstable across samples. To investigate this directly, we used our high scoring candidates to normalise expression of *HTATSF1* and *CDC40*. Raw data was indeed highly variable between samples, and between tissues and ages, yet this variability was well-matched between the two genes, confirming their geNorm score (Figure 10). Following normalisation (which again substantially lowered CoV), this variability resolved into a mild, but remarkably consistent, decrease in expression with increasing age, in all tissues. Normalisation of *CYC1*, *UBC* or *EIF4A* expression (*Extended data*: additional reference gene validation[53], Figure E2) gave similar results: *CYC1*, which typically scored highly in heads but low elsewhere, was revealed to be very consistently expressed in the former tissue while exhibiting age-associated

increases in embryos and forelimbs (*Extended data*: additional reference gene validation[53], Figure E2A and B); *UBC*, which was ranked modestly overall, exhibited a similar age-associated increase in expression in all tissues except the head (*Extended data*: additional reference gene validation[53], Figure E2C and D); *EIF4A*, which scored particularly highly in forelimbs, was indeed stable in this tissue, while showing age-specific increases in heads and highly variable behaviour in whole embryos (*Extended data*: additional reference gene validation[53], Figure E2E and F). Given *18S* and *EIF4A* were the highest scoring candidates in forelimbs, we further compared normalisation using these two genes with that obtained using the three genes above. Normalised forelimb expression of *B2M*, *HPRT1*, *CDC40* and *HTATSF1* was comparable with either reference gene combination, though overall CoV values were lower using *AP3D1*, *RPL13A* and *PAK1IP1* (*Extended data*: additional reference gene validation[53], Figure E3).

Finally, use of all three reference genes might not be necessary: while the MIQE guidelines suggest use of two
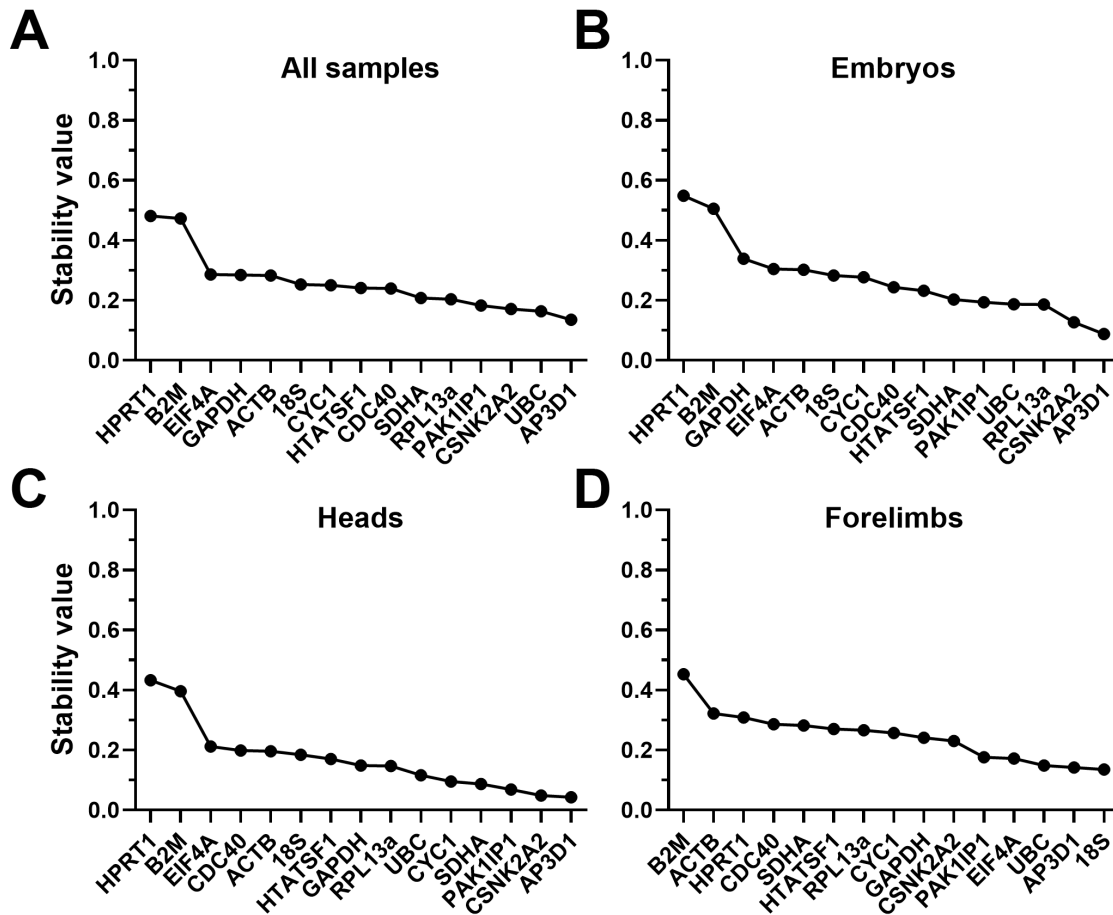
**Figure 6. Normfinder rankings (ungrouped).** Ungrouped rankings and scores of the fifteen candidate genes assessed by the Normfinder algorithm, using the entire dataset (**A**), whole embryo samples only (**B**), heads only (**C**) or forelimbs only (**D**). High stability values correspond to less stable genes.

reference genes at a minimum[24], our geNorm analysis suggested no substantial benefit in increasing number of genes from 2 to 3 (or indeed from 3 to 4). Accordingly, we compared a normalisation factor (NF) using all three genes to one using *AP3D1* with either *RPL13A* or *PAK1IP1* alone: the three gene NF was essentially identical to either 2-gene NF (gradients of 1.00 and 0.978 respectively, both with Pearson correlations of 0.99 - *Extended data*: additional reference gene validation[53], Figure E4), suggesting that two genes (*AP3D1* plus one other) are indeed sufficient.

## Discussion

The later stages of embryonic development involve many significant changes: substantial organogenesis occurs over this period, as does formation of skeletal muscle and the skeleton itself. Within the head, the brain matures from a comparatively simple tube to an intricately subdivided organ complete with discrete ventricles and regional specialisation, the teeth develop from buds to near-maturity, while the eye progresses from the simple optic cup stage to a defined globe complete with lens and iris, hidden beneath the fused eyelids. Normalisation of gene expression throughout this period might well be expected to be challenging, however our data unexpectedly suggests otherwise. For our dataset, whether assessed in its entirety, or restricted to whole embryos, heads, or forelimbs alone, all four analysis methods suggested that a substantial number of our candidate genes would serve as suitable reference genes: with the exception of *HPRT1* and *B2M*, essentially any gene from our panel represents an adequate reference. Given our panel consists of genes specifically selected for their

**Table 5. Normfinder rankings (ungrouped).** Ungrouped Normfinder results for the entire dataset or tissue-specific subsets (as indicated), ranked from most stable to least stable. Genes with stability values >0.5 –poor scoring candidates- are indicated by italics.

|  | All samples | Whole embryos | Heads | Forelimbs |
|---|---|---|---|---|
| Most | AP3D1 | AP3D1 | AP3D1 | 18S |
| stable | UBC | CSNK2A2 | CSNK2A2 | AP3D1 |
|  | CSNK2A2 | RPL13a | PAK1IP1 | UBC |
|  | PAK1IP1 | UBC | SDHA | EIF4A |
|  | RPL13a | PAK1IP1 | CYC1 | PAK1IP1 |
|  | SDHA | SDHA | UBC | CSNK2A2 |
|  | CDC40 | HTATSF1 | RPL13a | GAPDH |
|  | HTATSF1 | CDC40 | GAPDH | CYC1 |
|  | CYC1 | CYC1 | HTATSF1 | RPL13a |
|  | 18S | 18S | 18S | HTATSF1 |
|  | ACTB | ACTB | ACTB | SDHA |
|  | GAPDH | EIF4A | CDC40 | CDC40 |
|  | EIF4A | GAPDH | EIF4A | HPRT1 |
| Least | B2M | *B2M* | B2M | ACTB |
| stable | HPRT1 | *HPRT1* | HPRT1 | B2M |

reported stability, this is perhaps more reassuring than not, however we note that use of a near-identical panel of genes in skeletal muscle revealed markedly lower overall stability (for geNorm in particular, only *ACTB* and *RPL13a* gave M values <0.5)[38]. Our findings here imply that the developing embryo might be more transcriptionally consistent than different mature skeletal muscles. Even with most genes scoring highly there were differences, however, both between algorithms and between tissues, but the combination of *AP3D1*, *RPL13A* and *PAK1IP1* (or indeed *AP3D1* plus one of the others) emerged as consistently stable overall, and thus represent universal reference genes for the developing mouse embryo. We note these findings do not imply that these genes will necessarily be appropriate for more focussed embryonic contexts (such as specific organs), and suitable reference genes might need to be independently validated for such specific scenarios: indeed, as shown here *18S* and *EIF4A* apparently represent better candidates in forelimbs specifically (though *AP3D1*, *RPL13A* and *PAK1IP1* also scored highly).Studies addressing developmental changes in limbs alone might technically be better served by these two, but this would necessarily sacrifice the utility advantages offered by a universal panel (and *AP3D1*, *RPL13A* and *PAK1IP1* perform comparably

to *18S*/ *EIF4A* in forelimbs - *Extended data*: additional reference gene validation[53], Figure E3).

*AP3D1* codes for a component of the adaptor complex, which mediates non-clathrin-coated vesicle trafficking: we have previously found this gene to be stable throughout myogenic differentiation in culture[25] and in mature mouse skeletal muscle (healthy and dystrophic)[38], and our data here support the possibility that this gene might exhibit high overall stability in mouse. *RPL13A* codes for a protein component of the small ribosomal subunit: ribosomal proteins are ubiquitous, and some have proposed that such ubiquity renders them generally suitable as references[60], though this remains contentious[61]. As with *AP3D1*, we have previously found this gene to be stable in mature mouse skeletal muscle, and indeed also suitable as a reference in canine skeletal muscle (again, both healthy and dystrophic[26]). *PAK1IP1* encodes a negative regulator of PAK1 kinase, and its presence in our top scoring candidates represents something of a surprise: this gene was included in our candidate panels for both myogenic cell cultures and mature mouse muscle[25,38], and in those scenarios its performance was near-uniformly mediocre. Here this gene was ranked lower even than *HPRT1* and *B2M* in Normfinder analysis of
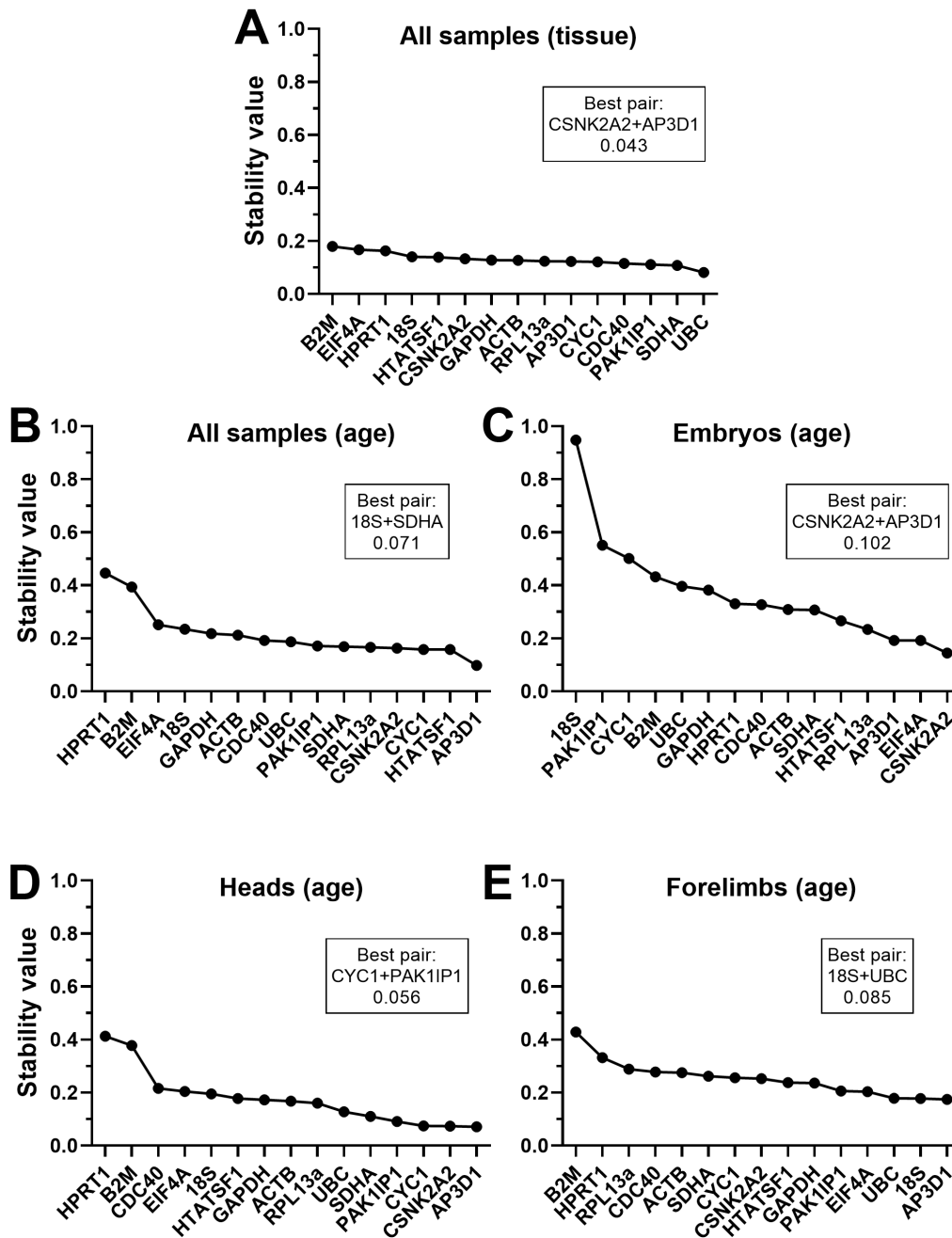
**Figure 7. Normfinder rankings (grouped).** Grouped rankings and scores of the fifteen candidate genes assessed by the Normfinder algorithm, using the entire dataset grouped by tissue (**A**) or by age (**B**), or whole embryo samples only grouped by age (**C**), heads grouped by age (**D**) or forelimbs grouped by age (**E**). High stability values correspond to less stable genes. The best pair of genes (which may not be the highest scoring individually) and the stability of that pair are indicated (boxes).

embryos grouped by age, but outside of this specific scenario the gene performed well, and as shown (*Extended data*: additional reference gene validation[53], Figure E4), *RPL13A* and *PAK1IP1* are essentially interchangeable when used in combination with *AP3D1*.

We note that all three genes are expressed at relatively high levels (Cq values of 21-24), potentially rendering them less appropriate for normalising low abundance transcripts (where minor differences in PCR efficiency can compound over multiple additional cycles). Identifying stably expressed low

**Table 6. Normfinder rankings (grouped).** Grouped Normfinder results for the entire dataset or tissue-specific subsets, grouped internally by age or tissue (as indicated), ranked from most stable to least stable. The best pair of genes are indicated separately (bold). Genes with stability values >0.5 –poor scoring candidates- are indicated by italics.

| | All samples | | Whole embryos | Heads | Forelimbs |
|---|---|---|---|---|---|
| | By age | By tissue | By age | By age | By age |
| Best pair | **18S + SDHA** | **CSNK2A2 + AP3D1** | **CSNK2A2 + AP3D1** | **CYC1 + PAK1IP1** | **18S + UBC** |
| Most | AP3D1 | UBC | CSNK2A2 | AP3D1 | AP3D1 |
| stable | HTATSF1 | SDHA | EIF4A | CSNK2A2 | 18S |
| | CYC1 | PAK1IP1 | AP3D1 | CYC1 | UBC |
| | CSNK2A2 | CDC40 | RPL13a | PAK1IP1 | EIF4A |
| | RPL13a | CYC1 | HTATSF1 | SDHA | PAK1IP1 |
| | SDHA | AP3D1 | SDHA | UBC | GAPDH |
| | PAK1IP1 | RPL13a | ACTB | RPL13a | HTATSF1 |
| | UBC | ACTB | CDC40 | ACTB | CSNK2A2 |
| | CDC40 | GAPDH | HPRT1 | GAPDH | CYC1 |
| | ACTB | CSNK2A2 | GAPDH | HTATSF1 | SDHA |
| | GAPDH | HTATSF1 | UBC | 18S | ACTB |
| | 18S | 18S | B2M | EIF4A | CDC40 |
| | EIF4A | HPRT1 | *CYC1* | CDC40 | RPL13a |
| Least | B2M | EIF4A | *PAK1IP1* | B2M | HPRT1 |
| stable | HPRT1 | B2M | *18S* | HPRT1 | B2M |

abundance genes can be challenging, however, and matching reference and GOI expression levels on a case-by-case basis is both time- and resource-intensive (and indeed impossible when investigating GOIs with dramatic changes in expression). Provided care is taken to use multiple stable, well-validated references (as per MIQE guidelines[24]), the utility of a broadly applicable high-abundance panel likely outweighs these concerns. Accordingly, Our validation process confirmed the utility of these genes, and also revealed insight into changes in gene expression during development: normalised expression of *HPRT1* and *B2M* (our lowest scoring genes) exhibited clear age-associated upregulation (Figure 9). *B2M* (beta 2 microglobulin) is a component of the MHC class I complex, and while this complex is present even at the earliest stages of development, studies in the rat have shown marked increases in expression in specific tissues such as skin, lung and inter-organ connective tissue[62]. Many of these tissues are only emerging at the earliest stages of our sample set (E11.5-E13.5) and indeed

the skin barrier is complete only toward the end of development (E16.5-E18.5)[18]. A developmentally associated upregulation in *B2M* is consistent with these changes. *HPRT1* encodes Hypoxanthine Phosphoribosyltransferase 1, a central component of the purine salvage pathway: this gene is stably expressed under many circumstances (we have reported this gene to be stable in healthy and dystrophic canine muscle[26]). Mutations in this gene however cause the neurodevelopmental condition, Lesch-Nyhan disease, and the gene itself plays a key role in glial/neuronal fate choice[63], implying that expression is developmentally regulated. Our data supports this, and indeed suggests that expression of this gene increases earlier in the head than within the embryo as a whole (Figure 9B, expression at E16.5). Conversely, both *CDC40* and *HTATSF1* displayed age-correlated downregulation following normalisation (Figure 10): this decrease was modest (insufficient to exclude them as acceptable, though not optimal, references) and moreover was sufficiently consistent between the two genes as to
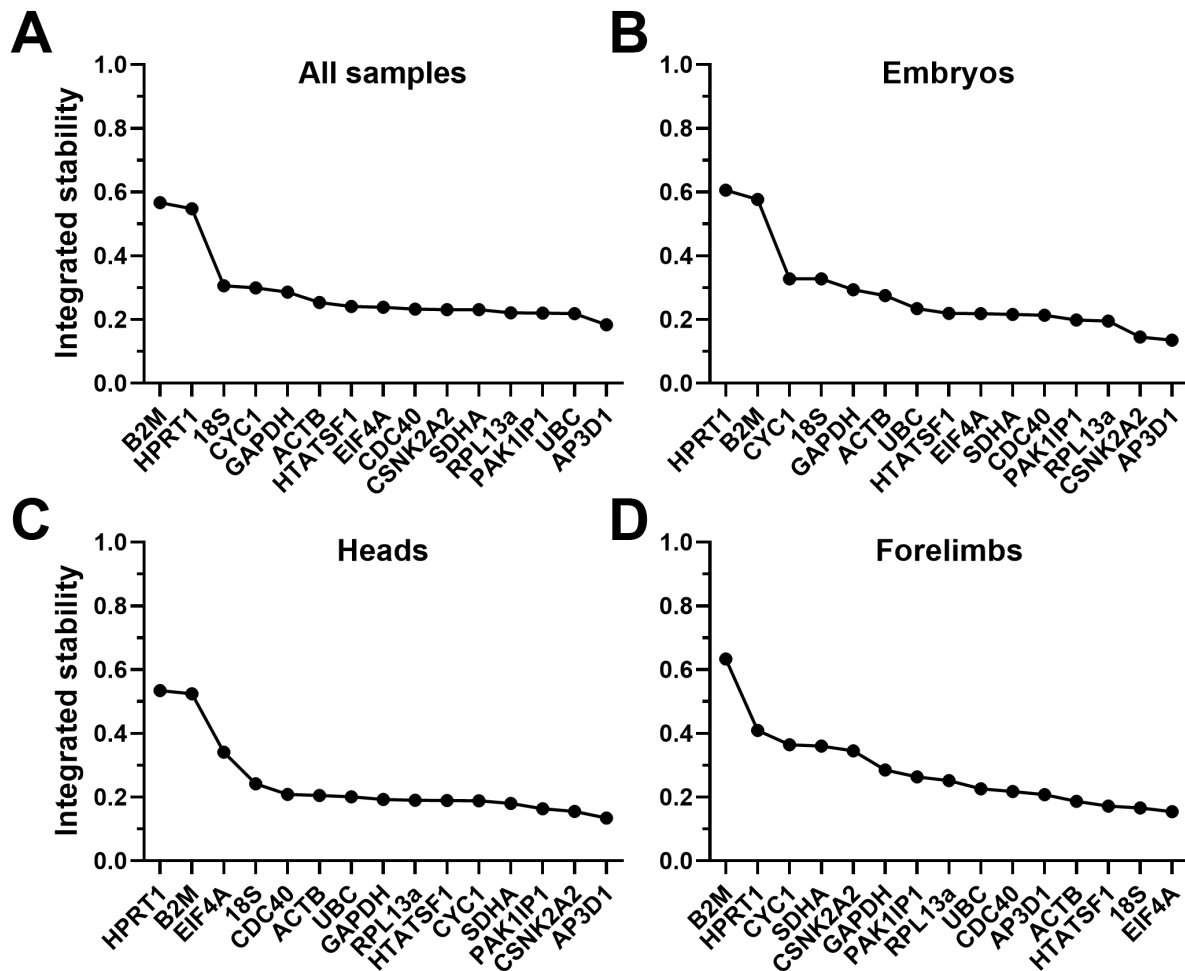
**Figure 8. Integrated rankings of the four analysis methods.** Geometric mean scores of all fifteen candidate genes from all four algorithms (see methods), assessed as either a complete dataset (**A**), whole embryos only (**B**), heads (**C**) or forelimbs (**D**). High integrated stability values correspond to less stable genes.

flag them as the best pair under geNorm analysis. This finding serves as a prominent reminder of the potential risks in relying on a single reference gene analysis method, but also indicates these genes might be co-ordinately regulated. *HTATSF1* encodes HIV-TAT specific factor 1, an RNA binding protein. Recent studies in embryonic stem cells implicate this factor in both regulation of ribosomal RNA processing and splicing of mRNA for ribosomal proteins, influencing protein synthesis and cell differentiation as a consequence[64]. *CDC40* (cell division cycle 40, also known as *PRP17*, or pre-mRNA processing factor 17) encodes a protein component of the spliceosome, and is also involved in progression through the cell cycle. Given our data, it seems likely that a shared involvement in mRNA splicing underpins the remarkably well-matched expression of these two genes, with requirements for

spliceosomal components decreasing gradually with developmental progress.

Our dataset also revealed a marked age-related increase in *CYC1* expression in all tissues except the head (*Extended data*: additional reference gene validation[53], Figure E2B). This gene encodes the mitochondrial electron transfer chain component cytochrome C, and our findings are thus best explained by developmental mitochondrial biogenesis. The most likely source of such marked increases in mitochondrial content is skeletal muscle: E11.5 represents the approximate midpoint of primary myogenesis, in which small numbers of myofibres are laid down to serve as scaffolds for subsequent musculature, while the period from E14.5-E18.5 spans secondary myogenesis, a phase of more substantial muscle synthesis[10]. This latter

**Table 7. Integrated ranking.** Reference genes ranked by geometric mean of geNorm, dCt, Bestkeeper and Normfinder scores. Genes with stability values >0.5 –poor scoring candidates- are indicated by italics.

|  | All samples | Whole embryos | Heads | Forelimbs |
|---|---|---|---|---|
| Most | AP3D1 | AP3D1 | AP3D1 | EIF4A |
| stable | UBC | CSNK2A2 | CSNK2A2 | 18S |
|  | PAK1IP1 | RPL13a | PAK1IP1 | HTATSF1 |
|  | RPL13a | PAK1IP1 | SDHA | ACTB |
|  | SDHA | CDC40 | CYC1 | AP3D1 |
|  | CSNK2A2 | SDHA | HTATSF1 | CDC40 |
|  | CDC40 | EIF4A | RPL13a | UBC |
|  | EIF4A | HTATSF1 | GAPDH | RPL13a |
|  | HTATSF1 | UBC | UBC | PAK1IP1 |
|  | ACTB | ACTB | ACTB | GAPDH |
|  | GAPDH | GAPDH | CDC40 | CSNK2A2 |
|  | CYC1 | 18S | 18S | SDHA |
|  | 18S | CYC1 | EIF4A | CYC1 |
| Least | *HPRT1* | *B2M* | *B2M* | HPRT1 |
| stable | *B2M* | *HPRT1* | *HPRT1* | *B2M* |

phase might well be associated with a dramatic increase in mitochondrial content, both in the embryo as a whole but especially in the forelimb where skeletal muscle comprises a greater fraction of the total tissue. The head is comparatively muscle-poor by contrast, and craniofacial muscles are moreover derived from a different progenitor pool which need not mature along identical timescales[65]. The brain is a highly energetic tissue, however neurodevelopmental changes over this period are predominantly structural/differentiation rather than bulk increases in tissue mass. In contrast, *EIF4A* showed the reverse trend (*Extended data*: additional reference gene validation[53], Figure E2F), increasing in heads but not forelimbs or embryos as a whole. This gene codes for an RNA helicase that unwinds 5' secondary structure for the translational preinitiation complex, and increases in expression in the head over this period might reflect the demands of the diverse transcription/translation programs necessary for brain maturation.

Finally, our experience with *18S* merits additional mention. Ribosomal RNAs comprise the vast bulk of total RNA (~80–85%), but the absence of polyA tails necessitates random priming for conversion to cDNA. Oligo dT does not contribute to *18S* cDNA, thus this ribosomal target effectively serves as an internal control for efficiency of random priming. As we show here, batch-dependent variance in *18S* specifically (i.e. other genes remain comparable) is highly indicative of a failure in this specific step, perhaps due to degradation of random primer stocks. Under many circumstances, this might be of little consequence: while the processivity of reverse transcriptase is such that even highly optimised enzymes seldom incorporate more than 1500 nucleotides in a single binding event[66], the majority of mRNAs are of modest (~3kb) length[67] and can typically be reverse-transcribed in their entirety from initial oligo dT priming alone. 5' sequence of longer transcripts (5kb+) will however be increasingly underrepresented or even absent under such conditions, and random priming is thus essential for qPCR studies investigating such transcripts (such as the early embryonic dystrophin isoform dp71). Critically, a failure in random priming will not be immediately apparent if all reference genes used can be successfully reverse-transcribed via oligo dT: our data suggests that for studies using random priming in combination with oligo dT, particularly those investigating expression of long genes, measurement of *18S* could represent a prudent quality check regardless of its efficacy as a reference.
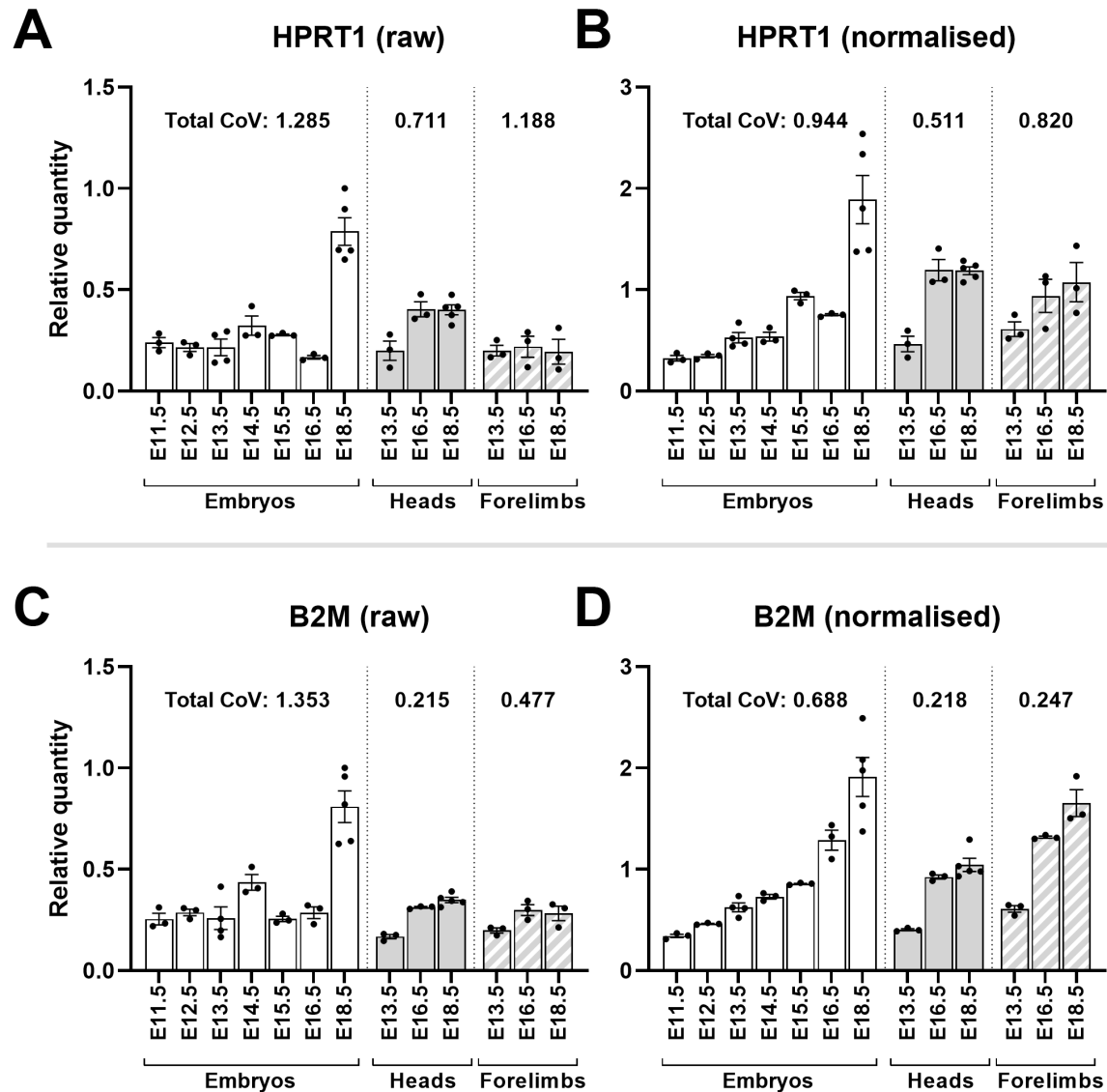
**Figure 9. Normalisation reveals HPRT1 and B2M are developmentally upregulated.** Mean raw RQ values for HPRT1 (**A**) and B2M (**C**) suggest modest expression at earlier stages with marked increases at E18.5. Normalisation with AP3D1, RPL13A and PAK1IP1 (**B** and **D**) lowers total coefficient of variation (CoV) and shows both genes are upregulated with increasing gestational age in whole embryos, heads and forelimbs (as indicated). Total CoV values were obtained by summing the individual CoVs per time-point (see methods). Points represent individual RQ values (arbitrary units).

## Conclusion

We have investigated potential reference genes for normalising quantitative PCR expression data in the developing mouse embryo from E11.5 to E18.5. We investigated expression both in whole embryos, or heads or forelimbs in isolation. Our data suggests that normalisation of expression over this period is not only possible, but that many reference genes are acceptable, though the genes *AP3D1*, *RPL13A* and *PAK1IP1* are the strongest candidates overall (using the entire sample set), within whole embryos alone, or within heads. These genes are also strong candidates for use in forelimbs (though *18S* and *EIF4A* score fractionally higher in this tissue). Our data further suggests only a pair of genes is necessary for effective normalisation: *AP3D1* and one other. We and
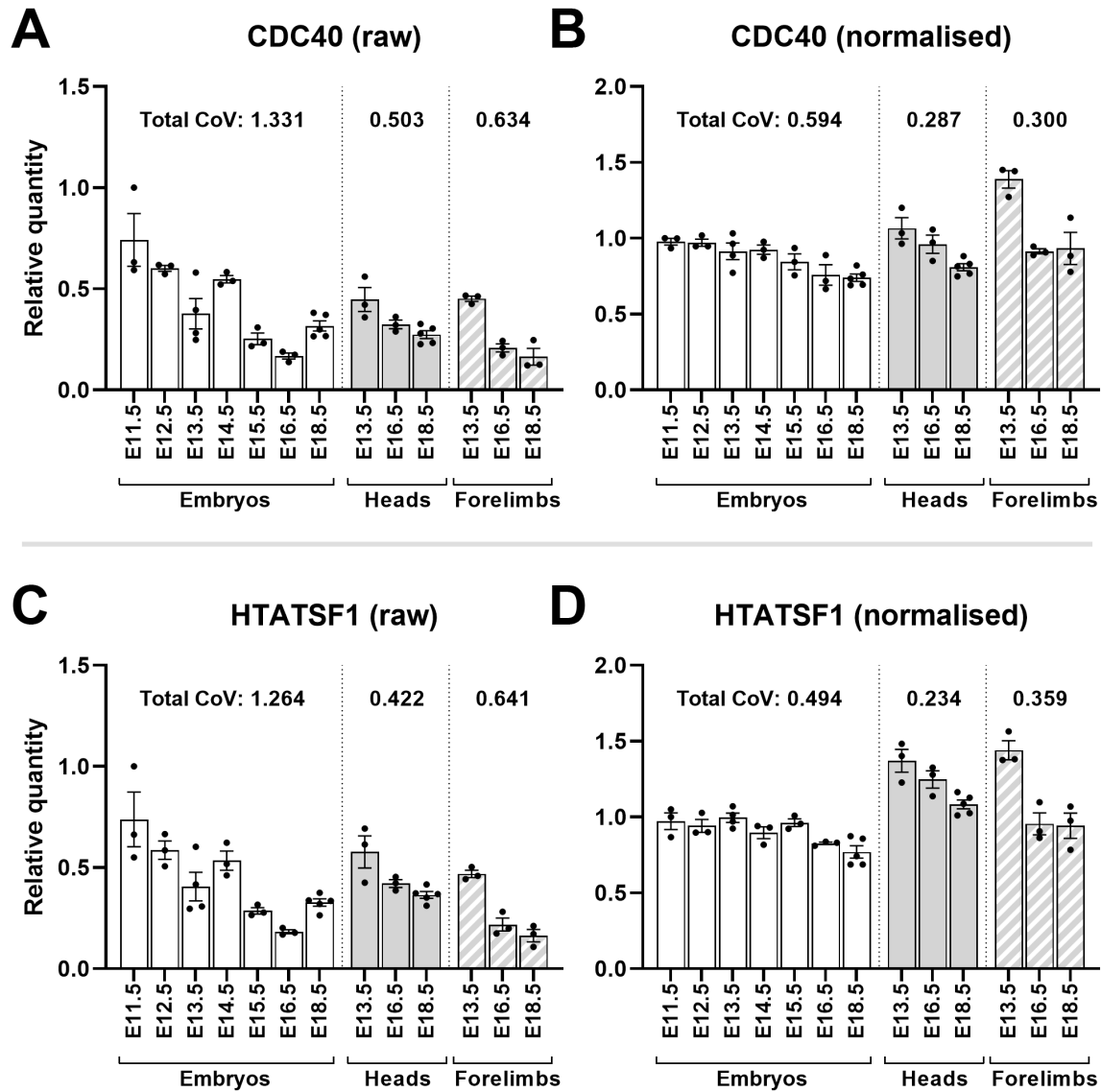
**Figure 10. Normalisation reveals CDC40 and HTATSF1 are developmentally downregulated.** Mean raw RQ values for CDC40 (**A**) and HTATSF1 (**C**) suggest closely matched but variable expression. Normalisation with AP3D1, RPL13A and PAK1IP1 (**B** and **D**) lowers total coefficient of variation (CoV) and shows both genes are modestly downregulated with increasing gestational age in whole embryos, and more markedly downregulated in heads and forelimbs (as indicated). Total CoV values were obtained by summing the individual CoVs per time-point (see methods). Points represent individual RQ values (arbitrary units).

others have reported *RPL13A* to be a strong candidate under other scenarios[26,38,58,68,69], while literature for *PAK1IP1* is more limited: *AP3D1* plus *RPL13A* thus seem the most practical pair for normalising gene expression in the developing mouse embryo.

## Data availability
### Underlying data
Figshare: Identification of qPCR reference genes suitable for normalising gene expression in the developing mouse embryo. https://doi.org/10.6084/m9.figshare.c.3136658.v1[55].

This project contains the following underlying data:
- Hildyard Wells Piercy Embryo qPCR raw data.xlsx (all raw Cq values, derived RQ values and melt curve data used in this manuscript)[52]
- Hildyard Wells Piercy Embryo qPCR RefAnalysis.xlsx (all derived reference gene rankings as determined via deltaCt, geNorm, Normfinder and Bestkeeper)[70]

## Extended data

Figshare: Extended data: Identification of qPCR reference genes suitable for normalising gene expression in the developing mouse embryo. https://doi.org/10.6084/m9.figshare.c.3136658.v1[55].

This project contains the extended data:
- Hildyard Wells Piercy Embryo qPCR Extended data Animal numbers and sample sizes.docx (detailed breakdown of animal and embryo numbers and sample assignment)[51]

- Hildyard Wells Piercy Embryo qPCR Extended data Reference gene analysis packages.docx (overview of the four reference gene algorithms used in this manuscript, with discussion of strengths and weaknesses and links to source packages where available)[37]

- Hildyard Wells Piercy Embryo qPCR Extended data file Additional reference gene validation.docx (geNorm pairwise variation analysis and additional validations of the three reference genes identified (AP3D1, RPL13A and PAK1IP1)), Extended data qPCR efficiencies.docx (methods used to calculate PCR efficiencies), Hildyard *et al* Embryo qPCR efficiencies.xlsx (qPCR primer pair efficiencies)[53]

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

## References

1. Kaufman MH: **The Atlas of Mouse Development.** 1st ed.: Academic Press; 1992.
   **Reference Source**
2. Hill M: **Embryology**.
   **Reference Source**
3. Macdonald DW: **The Encyclopedia of Mammals.** 3rd ed.: Oxford University Press; 2007.
   **Publisher Full Text**
4. Rossant J, Tam PPL: **Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse.** *Development.* 2009; **136**(5): 701–713.
   **PubMed Abstract** | **Publisher Full Text**
5. Tam PP, Behringer RR: **Mouse gastrulation: the formation of a mammalian body plan.** *Mech Dev.* 1997; **68**(1–2): 3–25.
   **PubMed Abstract** | **Publisher Full Text**
6. Maroto M, Bone RA, Dale JK: **Somitogenesis.** *Development.* 2012; **139**(14): 2453–2456.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
7. Taher L, Collette NM, Murugesh D, *et al.*: **Global gene expression analysis of murine limb development.** *PLoS One.* 2011; **6**(12): e28358.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
8. Pollard AS, McGonnell IM, Pitsillides AA: **Mechanoadaptation of developing limbs: shaking a leg.** *J Anat.* 2014; **224**(6): 615–23.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
9. Pitsillides AA: **Early effects of embryonic movement: 'a shot out of the dark'.** *J Anat.* 2006; **208**(4): 417–431.
   **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
10. Chal J, Pourquié O: **Making muscle: skeletal myogenesis *in vivo* and *in vitro*.** *Development.* 2017; **144**(12): 2104–2122.
    **PubMed Abstract** | **Publisher Full Text**
11. Buckingham M, Bajard L, Chang T, *et al.*: **The formation of skeletal muscle: from somite to limb.** *J Anat.* 2003; **202**(1): 59–68.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
12. Chen VS, Morrison JP, Southwell MF, *et al.*: **Histology Atlas of the Developing Prenatal and Postnatal Mouse Central Nervous System, with Emphasis on Prenatal Days E7.5 to E18.5.** *Toxicol Pathol.* 2017; **45**(6): 705–744.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
13. Gaete M, Lobos N, Torres-Quintana MA: **Mouse tooth development time sequence determination for the ICR/Jcl strain.** *J Oral Sci.* 2004; **46**(3): 135–41.
    **PubMed Abstract** | **Publisher Full Text**
14. Heavner W, Pevny L: **Eye development and retinogenesis.** *Cold Spring Harb Perspect Biol.* 2012; **4**(12): a008391.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
15. Sugiyama D, Inoue-Yokoo T, Fraser ST, *et al.*: **Embryonic regulation of the mouse hematopoietic niche.** *ScientificWorldJournal.* 2011; **11**: 1770–80.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
16. de Boer BA, van den Berg G, de Boer PAJ, *et al.*: **Growth of the developing mouse heart: an interactive qualitative and quantitative 3D atlas.** *Dev Biol.* 2012; **368**(2): 203–13.
    **PubMed Abstract** | **Publisher Full Text**
17. Scott RP, Maezawa Y, Kreidberg J, *et al.*: **Embryology of the Kidney.** Elsevier; 2016.
18. Schmitz A, Lazi'ć E, Koumaki D, *et al.*: **Assessing the *in vivo* epidermal barrier in mice: dye penetration assays.** *J Invest Dermatol.* 2015; **135**(2): 1–4.
    **PubMed Abstract** | **Publisher Full Text**
19. Cao J, Spielmann M, Qiu X, *et al.*: **The single-cell transcriptional landscape of mammalian organogenesis.** *Nature.* 2019; **566**(7745): 496–502.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
20. He P, Williams BA, Trout D, *et al.*: **The changing mouse embryo transcriptome at whole tissue and single-cell resolution.** *Nature.* 2020; **583**(7818): 760–767.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
21. Zhang MJ, Ntranos V, Tse D: **Determining sequencing depth in a single-cell RNA-seq experiment.** *Nat Commun.* 2020; **11**(1): 774.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
22. Baccarella A, Williams CR, Parrish JZ, *et al.*: **Empirical assessment of the impact of sample number and read depth on RNA-Seq analysis workflow performance.** *BMC Bioinformatics.* 2018; **19**(1): 423.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
23. Liu Y, Ferguson JF, Xue C, *et al.*: **Evaluating the impact of sequencing depth on transcriptome profiling in human adipose.** *PLoS One.* 2013; **8**(6): e66883.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**
24. Bustin SA, Benes V, Garson JA, *et al.*: **The MIQE guidelines: *m*inimum *i*nformation for publication of *q*uantitative real-time PCR *e*xperiments.** *Clin Chem.* 2009; **55**(4): 611–22.
    **PubMed Abstract** | **Publisher Full Text**
25. Hildyard JCW, Wells DJ: **Identification and validation of quantitative PCR reference genes suitable for normalizing expression in normal and dystrophic cell culture models of myogenesis.** *PLoS Curr.* 2014; **6**: ecurrents.md.faafdde4bea8df4aa7d06cd5553119a6.
    **PubMed Abstract** | **Free Full Text**
26. Hildyard JCW, Taylor-Brown F, Massey C, *et al.*: **Determination of qPCR Reference Genes Suitable for Normalizing Gene Expression in a Canine Model of Duchenne Muscular Dystrophy.** *J Neuromuscul Dis.* 2018; **5**(2): 177–191.
    **PubMed Abstract** | **Publisher Full Text**
27. Panina Y, Germond A, Masui S, *et al.*: **Validation of Common Housekeeping Genes as Reference for qPCR Gene Expression Analysis During iPS Reprogramming Process.** *Sci Rep.* 2018; **8**(1): 8716.
    **PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

28. Glare EM, Divjak M, Bailey MJ, *et al.*: **beta-Actin and GAPDH housekeeping gene expression in asthmatic airways is variable and not suitable for normalising mRNA levels.** *Thorax.* 2002; **57**(9): 765–770.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

29. Sikand K, Singh J, Ebron JS, *et al.*: **Housekeeping Gene Selection Advisory: Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) and β-Actin Are Targets of miR-644a.** *PLoS One.* 2012; **7**(10): e47510.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

30. Lin J, Redies C: **Histological evidence: housekeeping genes beta-actin and GAPDH are of limited value for normalization of gene expression.** *Dev Genes Evol.* 2012; **222**(6): 369–376.
**PubMed Abstract** | **Publisher Full Text**

31. Nazari F, Parham A, Maleki AF: **GAPDH, β-actin and β2-microglobulin, as three common reference genes, are not reliable for gene expression studies in equine adipose- and marrow-derived mesenchymal stem cells.** *J Anim Sci Technol.* 2015; **57**(1): 18.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

32. Lafontaine DLJ: **A 'garbage can' for ribosomes: how eukaryotes degrade their ribosomes.** *Trends Biochem Sci.* 2010; **35**(5): 267–277.
**PubMed Abstract** | **Publisher Full Text**

33. Vandesompele J, De Preter K, Pattyn F, *et al.*: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.** *Genome Biol.* 2002; **3**(7): RESEARCH0034.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

34. Andersen CL, Jensen JL, Ørntoft TF: **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.* 2004; **64**(15): 5245–5250.
**PubMed Abstract** | **Publisher Full Text**

35. Pfaffl MW, Tichopad A, Prgomet C, *et al.*: **Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations.** *Biotechnol Lett.* 2004; **26**(6): 509–515.
**PubMed Abstract** | **Publisher Full Text**

36. Silver N, Best S, Jiang J, *et al.*: **Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR.** *BMC Mol Biol.* 2006; **7**: 33.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

37. Hildyard J: **Reference gene determination packages: overview, strengths and weaknesses.** *Figshare.* Dataset. 2021.
**http://www.doi.org/10.6084/m9.figshare.14762064**

38. Hildyard JCW, Finch AM, Wells DJ: **Identification of qPCR reference genes suitable for normalizing gene expression in the *mdx* mouse model of Duchenne muscular dystrophy.** *PLoS One.* 2019; **14**(1): e0211384.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

39. Gu Y, Shen X, Zhou D, *et al.*: **Selection and expression profiles of reference genes in mouse preimplantation embryos of different ploidies at various developmental stages.** *PLoS One.* 2014; **9**(6): e98956.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

40. Jeong JK, Kang MH, Gurunathan S, *et al.*: **Evaluation of reference genes in mouse preimplantation embryos for gene expression studies using real-time quantitative RT-qPCR (RT-qPCR).** *BMC Res Notes.* 2014; **7**: 675.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

41. Mamo S, Gal AB, Bodo S, *et al.*: **Quantitative evaluation and selection of reference genes in mouse oocytes and embryos cultured *in vivo* and *in vitro*.** *BMC Dev Biol.* 2007; **7**: 14.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

42. Willems E, Mateizel I, Kemp C, *et al.*: **Selection of reference genes in mouse embryos and in differentiating human and mouse ES cells.** *Int J Dev Biol.* 2006; **50**(7): 627–35.
**PubMed Abstract** | **Publisher Full Text**

43. Barratt KS, Diamand KEM, Arkell RM: **Identification of reference genes suitable for RT-qPCR studies of murine gastrulation and patterning.** *Mamm Genome.* 2018; **29**(9–10): 656–662.
**PubMed Abstract** | **Publisher Full Text**

44. van den Bergen JA, Miles DC, Sinclair AH, *et al.*: **Normalizing Gene Expression Levels in Mouse Fetal Germ Cells.** *Biol Reprod.* 2009; **81**(2): 362–370.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

45. Ruiz-Villalba A, Mattiotti A, Gunst QD, *et al.*: **Reference genes for gene expression studies in the mouse heart.** *Sci Rep.* 2017; **7**(1): 24.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

46. Halouani A, Jmii H, Michaux H, *et al.*: **Housekeeping Gene Expression in the Fetal and Neonatal Murine Thymus Following Coxsackievirus B4 Infection.** *Genes (Basel).* 2020; **11**(3): 279.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

47. Kozera B, Rapacz M: **Reference genes in real-time PCR.** *J Appl Genet.* 2013; **54**(4): 391–406.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

48. Chapman JR, Waldenström J: **With Reference to Reference Genes: A Systematic Review of Endogenous Controls in Gene Expression Studies.** *PLoS One.* 2015; **10**(11): e0141853.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

49. Ramhøj L, Axelstad M, Svingen T: **Validation of endogenous reference genes in rat cerebral cortex for RT-qPCR analyses in developmental toxicity studies.** *PeerJ.* 2019; **7**: e7181.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

50. Kang Y, Wu Z, Cai D, *et al.*: **Evaluation of reference genes for gene expression studies in mouse and N2a cell ischemic stroke models using quantitative real-time PCR.** *BMC Neurosci.* 2018; **19**(1): 3.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

51. Hildyard J: **Embryo qPCR reference genes: animal numbers and sample sizes.** *Figshare.* Dataset. 2021.
**http://www.doi.org/10.6084/m9.figshare.14757957**

52. Hildyard J: **Embryo qPCR reference genes: raw data.** *Figshare.* Dataset. 2021.
**http://www.doi.org/10.6084/m9.figshare.14755803**

53. Hildyard J: **Embryo reference genes: additional validation.** *Figshare.* Dataset. 2021.
**http://www.doi.org/10.6084/m9.figshare.14763024**

54. Valadan R, Hedayatizadeh-Omran A, Alhosseini-Abyazani MN, *et al.*: **Data supporting the design and evaluation of a universal primer pair for pseudogene-free amplification of HPRT1 in real-time PCR.** *Data Brief.* 2015; **4**: 384–389.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

55. Hildyard J: **Identification of qPCR reference genes suitable for normalising gene expression in the developing mouse embryo.** Dataset. 2021.
**http://www.doi.org/10.6084/m9.figshare.c.3136658**

56. Pause A, Méthot N, Svitkin Y, *et al.*: **Dominant negative mutants of mammalian translation initiation factor eIF-4A define a critical role for eIF-4F in cap-dependent and cap-independent initiation of translation.** *EMBO J.* 1994; **13**(5): 1205–1215.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

57. Andreou AZ, Klostermeier D: **The DEAD-box helicase eIF4A: paradigm or the odd one out?** *RNA Biol.* 2013; **10**(1): 19–32.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

58. Gentile AM, Lhamyani S, Coín-Aragüez L, *et al.*: ***RPL13A* and *EEF1A1* Are Suitable Reference Genes for qPCR during Adipocyte Differentiation of Vascular Stromal Cells from Patients with Different BMI and HOMA-IR.** *PLoS One.* 2016; **11**(6): e0157002.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

59. Xie F, Xiao P, Chen D, *et al.*: **miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs.** *Plant Mol Biol.* 2012; 1573–5028(Electronic).
**PubMed Abstract** | **Publisher Full Text**

60. Hsiao LL, Dangond F, Yoshida T, *et al.*: **A compendium of gene expression in normal human tissues.** *Physiol Genomics.* 2001; **7**(2): 97–104.
**PubMed Abstract** | **Publisher Full Text**

61. Thorrez L, van Deun K, Tranchevent LC, *et al.*: **Using ribosomal protein genes as reference: a tale of caution.** *PLoS One.* 2008; **3**(3): e1854.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

62. Grabowska A, Lampson LA: **MHC expression in nonlymphoid tissues of the developing embryo: Strongest class I or class II expression in separate populations of potential antigen-presenting cells in the skin, lung, gut, and inter-organ connective tissue.** *Dev Comp Immunol.* 1995; **19**(5): 425–450.
**PubMed Abstract** | **Publisher Full Text**

63. Kang TH, Park Y, Bader JS, *et al.*: **The housekeeping gene hypoxanthine guanine phosphoribosyltransferase (HPRT) regulates multiple developmental and metabolic pathways of murine embryonic stem cell neuronal differentiation.** *PLoS One.* 2013; **8**(10): e74967.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

64. Corsini NS, Peer AM, Moeseneder P, *et al.*: **Coordinated Control of mRNA and rRNA Processing Controls Embryonic Stem Cell Pluripotency and Differentiation.** *Cell Stem Cell.* 2018; **22**(4): 543–558.e12.
**PubMed Abstract** | **Publisher Full Text**

65. Noden DM, Francis-West P: **The differentiation and morphogenesis of craniofacial muscles.** *Dev Dyn.* 2006; **235**(5): 1194–218.
**PubMed Abstract** | **Publisher Full Text**

66. Baranauskas A, Paliksa S, Alzbutas G, *et al.*: **Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants.** *Protein Eng Des Sel.* 2012; **25**(10): 657–668.
**PubMed Abstract** | **Publisher Full Text**

67. Piovesan A, Caracausi M, Antonaros F, *et al.*: **GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics.** *Database (Oxford).* 2016; **2016**: baw153.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

68. Bian Z, Yu Y, Quan C, *et al.*: **RPL13A as a reference gene for normalizing mRNA transcription of ovarian cancer cells with paclitaxel and 10-hydroxycamptothecin treatments.** *Mol Med Rep.* 2015; **11**(4): 3188–94.
**PubMed Abstract** | **Publisher Full Text**

69. Curtis KM, Gomez LA, Rios C, *et al.*: **EF1alpha and RPL13a represent normalization genes suitable for RT-qPCR analysis of bone marrow derived mesenchymal stem cells.** *BMC Mol Biol.* 2010; **11**: 61.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

70. Hildyard J: **Embryo qPCR reference genes: analysis.** *Figshare.* Dataset. 2021.
**http://www.doi.org/10.6084/m9.figshare.14756484**

# Open Peer Review

## Current Peer Review Status: ✓ ✓ ✓

---

**Version 2**

Reviewer Report 05 October 2022

https://doi.org/10.21956/wellcomeopenres.20390.r52483

✓    **Liang Hu** (iD)

Reproductive and Genetic Hospital of CITIC-XIANGYA, Changsha, China

No new comments.

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Genetics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 23 September 2022

https://doi.org/10.21956/wellcomeopenres.20390.r52485

✓    **Anush Kosakyan**

Institute of Parasitology, Biology Centre Czech Academy of Science, České Budějovice, Czech Republic

I am satisfied with the author's answers and corrections and do not have further comments.

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Microbial eukaryotes, taxonomy, phylogeny, transcriptomics, genomics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

**Version 1**

Reviewer Report 21 April 2022

https://doi.org/10.21956/wellcomeopenres.18735.r49745

**?**

**Liang Hu** (iD)

Reproductive and Genetic Hospital of CITIC-XIANGYA, Changsha, China

In this manuscript, the authors tried to identify the stably expressed genes as normalizing controls in E11.5-E18.5 mouse embryos for future studies. It is interesting and important for some researchers.

Several main problems are listed as follows:

1. The diversity and the complexity of the cell types in mouse embryos may affect the reliability of this study. In this study, the authors only chose head and forelimb samples, which cannot reflect other tissue and cell types. The best way is to analyze the published single-cell sequencing data and find the candidate genes, and then select the best ones by quantitative PCR.

2. I would recommend the authors verified the stability of their selected genes in this study in other published sequencing data.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Partly

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Partly

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Genetics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 12 Sep 2022
**John Hildyard**, Royal Veterinary College, London, UK

We thank the reviewer for their careful consideration of our manuscript, and their helpful comments. We address these as discussed below (original reviewer comments in italics for reference).

*Several main problems are listed as follows:*
*The diversity and the complexity of the cell types in mouse embryos may affect the reliability of this study. In this study, the authors only chose head and forelimb samples, which cannot reflect other tissue and cell types. The best way is to analyze the published single-cell sequencing data and find the candidate genes, and then select the best ones by quantitative PCR.*
*I would recommend the authors verified the stability of their selected genes in this study in other published sequencing data.*

This is an excellent point to raise: thank you. Embryos are indeed a complex, dynamic transcriptional environment, with many different cell types present. Critically, however, many of these specific nuances are lost following bulk RNA extraction, which necessarily homogenises all cells present. Our approach is intended to address the normalisation of gene expression in this latter scenario only: as it is often both more practical and more expedient to determine gene expression behaviour in bulk extracts, especially when the focus is on expression of only a few, specific targets. Moreover, study of gene expression via such approaches is markedly less expensive, placing it within the reach of more investigators. As we state within our introduction:

"Recent technological advances have permitted elegant studies of embryonic gene expression down to single-cell level [19, 20], suggesting that detailed study of the interactions between individual cell lineages might now be possible, however such approaches also represent a substantial undertaking both in time and resources. Furthermore, these approaches are not without limitation: the whole transcriptome nature of the method typically favours breadth over depth [21, 22]: large-scale changes involving many gene components will be mapped well, while more specific, smaller-scale processes can be missed. In particular, constraints in read depth mean low abundance transcripts can be under-represented in such datasets [23], and thus more conventional methods remain vital research tools."

As we show, isolating specific embryonic tissues prior to RNA extraction (heads, forelimbs) does alter the transcriptional behaviour measured (presumably as a consequence of the more restricted cellular milieu present in such tissue subsets), but as we further shown in our manuscript, the reference genes identified here as appropriate for whole embryos remain applicable even within these subset isolates. We do not assume that the genes isolated will be applicable within *all* embryonic tissue subsets (we deem this unlikely, and have edited the text to make this distinction more explicit), but repeating the work presented, over the developmental stages shown, for increasingly specific (and potentially challenging to isolate) tissue subsets seems firmly beyond the scope of this study. Our work does, however, represent a solid technical and instructive foundation for investigators who might wish to pursue such further studies.

With respect to single-cell data, we do not assume our candidate genes will necessarily apply to single-cell data, and indeed, we would strongly advise against such an approach: the depth and detail of single-cell sequencing data offers alternative, more appropriate normalisation methods that can be determined *a priori,* on a sample-by-sample basis, or even a cell-by-cell basis, if necessary (Lytal *et al*, doi: 10.3389/fgene.2020.00041). By the same token, we do not feel that single-cell data can be readily used to inform bulk cDNA normalisation: the transcriptional behaviour of individual cells and cell types is not necessarily representative of behaviour at whole tissue level.

We also note that RNAseq data (bulk or single-cell) is not strictly analogous to the cDNA approaches used here: RNAseq most commonly uses oligodT priming (and/or oligodT purification), leading to an inherent 3' bias. As our data shows, 3' bias can strongly affect the measured expression of long genes. We avoid this bias by use of random priming (and demonstrate the difference this makes), but such adjustments cannot necessarily be translated to RNAseq methodologies.

Finally, we note that others have shown that use of RNAseq datasets is not required for reference gene identification (Sampathkumar et al, doi:10.1371/journal.pcbi.1009868), and that (as the reviewer rightly notes) any genes identified as potential candidates via RNAseq would need to be validated using the methods we already employ in this manuscript. As we show, our highest scoring candidates are eminently suitable references already: we are reluctant to recreate this entire study with additional genes for (at best) a fractional gain in score.

We hope this addresses the reviewer's concerns.

***Competing Interests:*** No competing interests.

Reviewer Report 19 April 2022

https://doi.org/10.21956/wellcomeopenres.18735.r49742

**?**   **Anush Kosakyan**

Institute of Parasitology, Biology Centre Czech Academy of Science, České Budějovice, Czech Republic

This study validates potential reference genes for gene expression studies in the developing mouse embryo from E11.5 to E18.5. Authors investigated expression of 15 candidate genes in whole embryos, heads and forelimbs, and suggested *AP3D1*, *RPL13A* and *PAK1IP1* as the most suitable ones for normalising gene expression in the developing mouse embryo. The manuscript is overall well written and structured, the methods are transparent, however there are several points that must be considered before indexing:

**Abstract:**

"Unexpectedly, all methods suggest that many genes within our candidate panel are acceptable references, though *AP3D1*, *RPL13A* and *PAK1IP1* are the strongest performing genes overall. *HPRT1* and *B2M* are conversely poor choices and show strong developmental regulation. We further show that normalization using our three highest-scoring references can reveal subtle patterns of developmental expression even in genes ostensibly ranked as acceptably stable (*CDC40*, *HTATSF1*)."
  - Rather than saying here that HPRT1 and B2M are conversely poor choices (in general, we are not interested in what is the poor choice, but what is the best choice), I would say which genes performed best specifically for the overall embryo level and specifically for head and forelimbs, because some future studies might focus only on head or forelimbs.

**Introduction:**

"This panel includes candidates widely used historically (*18S*, *GAPDH*, *ACTB*), those that we have shown be strong references in mouse skeletal muscle (*CSNK2A2*, *AP3D1*, *RPL13A*), and also shown by others to be viable references in rodent brains (*UBC*, *HPRT1*, *RPL13A, 18S*). To maximise the power of our study, we have assessed reference gene suitability using all four algorithms described above (geNorm, deltaCt, BestKeeper and Normfinder)."
  - Provide references for 18S, GAPDH, ACTB. Also, it would be great if you could provide a table (or just an extension of Table 1) with a brief description of the condition/scenario in which the gene was suggested/validated as the best reference gene and the corresponding reference. That would make your introduction really valuable.

**Methods:**

"Uterine horns were quickly removed and placed on ice (additional organs and muscles were collected from the adults for separate studies)."
  - I did not understand why this information is relevant to this study.

"RNA yield and purity were assessed via nanodrop (ND1000) and samples with 260/230 ratios below 1.7 were subjected to a second precipitation."
  - = > just for your information, the optimal values for RNA purity are: 260/280 = 2, and

260/230 = 2-2.2.
"All sequences are provided below:
18S F 5'-GTAACCCGTTGAACCCCATT-3'
18S R 5'-CCATCCAATCGGTAGTAGCG-3'
HPSF F 5'-GGACTAATTATGGACAGGACTG-3'
HPSF R 5'-GCTCTTCAGTCTGATAAAATCTAC-3'
Dp71 F 5'-GTGAAACCCTTACAACCATGAG-3'
Dp71 R 5'-CTTCTGGAGCCTTCTGAGC-3'"

- ○ Make a table for all gene primer sequences and give the efficiency values for each.

**Results:**

- ○ Fig2. I am a little confused about this figure or maybe the legend is not clear enough. Are these points in each boxplot for all samples (i.e. all biological replicates for whole embryos, heads and forelimbs) per gene? Why is C shown without a box?

**Discussion:**

- ○ It can be seen from Fig. 2 that the Ct values for the proposed reference genes are between 22 and 23, indicating that the genes are comparatively highly expressed. I would add a few lines to discuss whether you see a limitation in using these genes for the normalization of low expressed genes.

- ○ I would also add a few lines emphasizing that the proposed reference genes are uniformly expressed throughout embryonic development, whereas the choice of these genes might be different if we were looking at individual parts (e.g., head, forelimbs) or scenarios.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Microbial eukaryotes, taxonomy, phylogeny, transcriptomics, genomics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have**

**significant reservations, as outlined above.**

Author Response 12 Sep 2022

**John Hildyard**, Royal Veterinary College, London, UK

We thank the reviewer for their careful and detailed assessment of our manuscript, and their insightful suggestions (which we discuss below -where appropriate, the reviewer's comments are reiterated for easy reference, in italics).

*Rather than saying here that HPRT1 and B2M are conversely poor choices (in general, we are not interested in what is the poor choice, but what is the best choice), I would say which genes performed best specifically for the overall embryo level and specifically for head and forelimbs, because some future studies might focus only on head or forelimbs.*

Thank you for this suggestion. We feel for studies of this nature it is important not only to identify strong candidates, but also to highlight those that are not simply 'less strong' candidates, but that are actively poor candidates (such as genes that exhibit strong developmental regulation: HPRT1 and B2M). Use of these genes as references in embryonic contexts runs the risk of generating data that is not necessarily poorly normalised (i.e. high variance, low statistical value) but *incorrectly* normalised (i.e. low variance, statistically tractable, but not reflective of biological reality). We feel this latter point is worth emphasising, especially when such genes might perform strongly under other contexts -we and others have shown that HPRT1 is a very strong candidate when normalising muscle gene expression, for example (Hildyard *et al*: DOI: 10.3233/JND-170267, Wang *et al*: DOI: 10.3892/mmr.2019.10102).

We have, however, edited the abstract to better illustrate the overall suitability of our strongest candidates: as we note in our discussion, the three genes identified *are* the strongest performing in whole embryos and in heads alone ("In embryos, heads, or all samples together, *AP3D1* is the strongest candidate, while *RPL13A*, *PAK1IP1* and *CSNK2A2* are also highly ranked"), and while 18S and EIF4A scored fractionally higher for use in forelimbs, forelimb data normalised using these two were highly comparable to data normalised using AP3D1, RPL13A and PAK1IP1, suggesting that these latter three do indeed represent a universal panel (we provide extended data to specifically illustrate this). The edited abstract is now as follows (edited text in bold):

"Unexpectedly, all methods suggest that many genes within our candidate panel are acceptable references, though *AP3D1*, *RPL13A* and *PAK1IP1* are the strongest performing genes overall **(scoring highly in whole embryos, heads or forelimbs alone, and in all samples collectively)**."

*Provide references for 18S, GAPDH, ACTB.*

An entirely reasonable suggestion: thank you. As we note, these three are widely used historically: so widely, in fact, that it might almost be easier to provide citations that do *not* use one or more of these three. While the observation that these three are not optimal under all circumstances (and are in fact often markedly poor choices) is gradually gaining

acceptance, use of these genes is still prevalent throughout the literature. We have thus provided citations of review articles addressing the historical use of these three (Kozera and Rapacz: doi:10.1007/s13353-013-0173-x, Chapman and Waldenstrom: doi:10.1371/journal.pone.0141853), which we feel more comprehensively addresses the pervasive use of these genes.

*Also, it would be great if you could provide a table (or just an extension of Table 1) with a brief description of the condition/scenario in which the gene was suggested/validated as the best reference gene and the corresponding reference. That would make your introduction really valuable.*

This is an excellent suggestion, but perhaps one best suited to a review article examining the use of reference genes more generally (rather than one focussed on identification of the correct references in developing mouse embryos specifically). Some of these genes (18S, ActB and GAPDH in particular, as discussed in the citations added above) have been used broadly, in scenarios ranging from mammalian cell cultures (Liu et al, doi: 10.1371/journal.pone.0117058) all the way to celery (Feng et al, doi: 10.7717/peerj.7925): including such a breadth of information, even in table form, seems challenging. Additionally, historical use of these genes might not always have been appropriate: validation of specific genes for specific scenarios is scientifically rigorous, but is also a comparatively recent approach. Such validation is typically conducted *a priori* rather than with the assumption that behaviour of a given gene in one context will translate to another. Indeed, several genes from the geNorm/geNorm PLUS collections (which comprise the bulk of those used here) were selected not due to their use in the literature, but based on their stability across multiple publicly accessible microarray datasets: the core principle being that providing a large panel of candidate genes (all of which should ostensibly be relatively stable) allows users to determine the most stable genes for their specific scenario (as in this manuscript). 18S, GAPDH and ActB are typically included in geNorm collections not because they are likely to be stable, but for the simple reason that they have such broad historical use: their inclusion thus allows analysis to be placed in context with established literature.

We already cite studies validating individual genes where appropriate (i.e. in neural or skeletal muscle contexts), and we are reluctant to further expand table 1 (which has already been enlarged to include accession numbers and amplicon lengths).

*Methods:*

*"Uterine horns were quickly removed and placed on ice (additional organs and muscles were collected from the adults for separate studies)."*
*I did not understand why this information is relevant to this study.*

We apologise for any confusion: Wellcome Open Research adheres closely to the ARRIVE guidelines, and to the principles of Replacement, Reduction and Refinement (3Rs) -we feel this is laudable, and as a Veterinary College, we too adhere closely to these principles. This information was included to clarify explicitly that tissues from animals sacrificed for this work were utilised more broadly, both to demonstrate our commitment to reductions in

animal use, and to encourage others to employ similar ethical approaches. We would prefer to retain this information for these reasons, but we are willing to remove it at the discretion of the editor.

*"RNA yield and purity were assessed via nanodrop (ND1000) and samples with 260/230 ratios below 1.7 were subjected to a second precipitation."*
*= > just for your information, the optimal values for RNA purity are: 260/280 = 2, and 260/230 = 2-2.2.*

Thank you. We are aware of these optimal values, however we are also aware that it is not always possible to achieve such high ratios (particularly for 260/230), and that repeat cleaning of perfectly useable RNA in an attempt to attain greater purity is an approach with rapidly diminishing returns. In our experience, using the reagents cited in our manuscript, samples with 260/230 ratios above 1.7 behave essentially identically to those with optimal ratios, and this more modest threshold is markedly easier to attain. We feel this information might be of use/interest to other investigators.

*Make a table for all gene primer sequences and give the efficiency values for each.*

As noted within the manuscript, sequences for most of the primers used in this work are proprietary, and thus cannot be provided. For the remaining six primers (as shown) we did not feel a table was necessary. The suggestion to show the efficiency values is excellent and appreciated, however: accordingly we have added all efficiency calculations and derived values to the extended data (see also our response to reviewer 1).

*Results:*
*Fig2. I am a little confused about this figure or maybe the legend is not clear enough. Are these points in each boxplot for all samples (i.e. all biological replicates for whole embryos, heads and forelimbs) per gene? Why is C shown without a box?*

Yes, for each gene, all sample Cq values are shown (whole embryos, heads, forelimbs). Boxplots are depicted to more clearly illustrate the outlier behaviour of the specific samples identified as aberrant. Figure 2C serves exclusively to illustrate differences in specific sample batches within a single gene (18S or dp71), rather than overall consistency within a gene, and thus boxplots would not convey any further information beyond that already depicted. We accept that this might lead to confusion, however, and have thus altered this latter plot to box and whisker for consistency.

We have also edited the figure legend to clarify these specifics (edited text in bold):

"Box and whisker plots of **all** sample Cq values for each candidate gene used in this study: **each point represents the Cq value of an individual sample, for the gene indicated**. ( **A**) Initial raw Cq values revealed a marked variance in 18S expression (but not other genes), restricted to samples from a single cDNA synthesis batch (open circles), compared to samples from other cDNA synthesis batches (black points), indicative of impaired random priming. ( **B**) Preparation of fresh cDNA for these samples corrected this discrepancy (filled circles), producing values consistent with other batches (black points). ( **C**) qPCR for 18S and

the dystrophin isoform dp71 using the initial poor samples (1, **open circles**), the other samples (2, **black points)** and the fresh preparation (3, **filled circles**) shows that impaired random priming prevents accurate quantification of long transcripts. **Details of specific cDNA synthesis batches can be found in the underlying data [67].**"

*Discussion:*
*It can be seen from Fig. 2 that the Ct values for the proposed reference genes are between 22 and 23, indicating that the genes are comparatively highly expressed. I would add a few lines to discuss whether you see a limitation in using these genes for the normalization of low expressed genes.*

An excellent suggestion: thank you! We agree that it is preferable to use reference genes of comparable expression level to a given gene of interest, wherever practical. It can be challenging to identify stably expressed genes of low abundance, however (particularly within the transcriptionally dynamic tissue of embryos): low abundance is often an indicator of tightly controlled (rather than ubiquitous) expression, and moreover even if such genes are indeed constitutively expressed, lower levels of expression are inherently more vulnerable to stochastic noise. This can be partly addressed by use of multiple references (and is a strong argument in favour of such) but identifying *multiple* stably expressed low abundance genes is particularly challenging.

Provided care is taken to ensure that efficiencies are appropriate and comparable between references and GOI, we feel that good, high abundance references are preferable to more noise-susceptible low abundance references, but we thank the reviewer for raising this excellent point, and have added words to address this within our discussion, as follows (new text in bold):

"**We note that all three genes are expressed at relatively high levels (Cq values of 21-24), potentially rendering them less appropriate for normalising low abundance transcripts (where minor differences in PCR efficiency can compound over multiple additional cycles). Identifying stably expressed low abundance genes can be challenging, however, and matching reference and GOI expression levels on a case-by-case basis is both time- and resource-intensive (and indeed impossible when investigating GOIs with dramatic changes in expression). Provided care is taken to use multiple stable, well-validated references (as per MIQE guidelines [24]), the utility of a broadly applicable high-abundance panel likely outweighs these concerns. Accordingly**, our validation process confirmed the utility of these genes, and also revealed insight into changes in gene expression during development…"

*I would also add a few lines emphasizing that the proposed reference genes are uniformly expressed throughout embryonic development, whereas the choice of these genes might be different if we were looking at individual parts (e.g., head, forelimbs) or scenarios.*

A sensible suggestion: we agree. We have endeavoured to make these distinctions clear throughout our manuscript, and to stress that the three candidates identified are in fact not only highest scoring in whole embryos, but also within heads, and within the entire dataset (including whole embryos, heads and forelimbs). As noted above (and within the

manuscript) they are not the *highest* scoring within forelimbs, but remain *high* scoring (such that data normalised using these three is comparable to data normalised using 18S/EIF4A. Given the reviewer's suggestions, we have made changes to better clarify this (and to stress that such universality should not be assumed to extrapolate to additional scenarios). We have edited the text of the discussion and conclusions accordingly:

"We note **these findings do not imply that these genes will necessarily be appropriate for more focussed embryonic contexts (such as specific organs), and suitable reference genes might need to be independently validated for such specific scenarios: indeed, as shown here** *18S* and *EIF4A* apparently represent better candidates in forelimbs specifically **(though *AP3D1*, *RPL13A* and *PAK1IP1* also scored highly)**. **S**tudies addressing developmental changes in limbs alone might technically be better served by these two, but this would necessarily sacrifice the utility advantages offered by a universal panel..."

"Our data suggests that normalisation of expression over this period is not only possible, but that many reference genes are acceptable, though the genes *AP3D1*, *RPL13A* and *PAK1IP1* are the strongest candidates overall **(using the entire sample set), within whole embryos alone, or within heads. These genes are also strong candidates for use in forelimbs (though *18S* and *EIF4A* score fractionally higher in this tissue).** Our data **further** suggests only a pair of genes is necessary for effective normalisation: *AP3D1* and one other."

***Competing Interests:*** No competing interests.

Reviewer Report 21 October 2021

✔  **Caroline L. Smith** 🆔

School of Life Sciences, University of Westminster, London, UK

This is a valuable resource for researchers which provides details of reference genes at different stages of embryonic development. As the authors comment, incorrect selection of reference genes can significantly skew qPCR data and furthermore some genes considered to be stable do change throughout development. The reports developmental regulation of about HPRT1 and B2M are valuable findings.

Although the authors state that primer efficiencies were all 95-105%, it would be extremely helpful to have these values in a table alongside details of range of dilutions of cDNA used to generate these standard curves and which ages/tissues were used to determine the efficiencies.

The primer sequences for the GeNorm are not available but it would be useful to know the amplicon sizes and accession numbers of the genes provided in Table 1.
In Figure 2 - the legend lacks detail about which ages/tissues are used in the cDNA batches.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Gene expression, cardiovascular signalling

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 12 Sep 2022

**John Hildyard**, Royal Veterinary College, London, UK

We thank the reviewer for their supportive assessment of this work, and for the excellent suggestions (repeated below in italics, for reference), which we address as follows.

*Although the authors state that primer efficiencies were all 95-105%, it would be extremely helpful to have these values in a table alongside details of range of dilutions of cDNA used to generate these standard curves and which ages/tissues were used to determine the efficiencies*

This is a pertinent and sensible suggestion: we have added two supplementary data files (extended data, Additional reference gene validation) containing all the efficiency calculations and the methodology used to generate these.

*The primer sequences for the GeNorm are not available but it would be useful to know the amplicon sizes and accession numbers of the genes provided in Table 1*

Again, an excellent suggestion: we have added these details to the table.

*In Figure 2 - the legend lacks detail about which ages/tissues are used in the cDNA batches.*

Correct, it did not: the primary purpose of figure 2 is to illustrate that batch-specific effects can be identified and then subsequently addressed (i.e. batch-specific behaviour eliminated), and thus for all subsequent analyses within the manuscript, no such batch effects should be assumed to apply. We did not consequently feel this additional information was necessarily of benefit to the reader, however we are entirely happy to implement this change as suggested. Accordingly, we have edited the legend to refer to the underlying data, which now contains the relevant batch-specific information (edits in bold).

"Box and whisker plots of **all** sample Cq values for each candidate gene used in this study: **each point represents the Cq value of an individual sample, for the gene indicated**. ( **A**) Initial raw Cq values revealed a marked variance in 18S expression (but not other genes), restricted to samples from a single cDNA synthesis batch (open circles), compared to samples from other cDNA synthesis batches (black points), indicative of impaired random priming. ( **B**) Preparation of fresh cDNA for these samples corrected this discrepancy (filled circles), producing values consistent with other batches (black points). ( **C**) qPCR for 18S and the dystrophin isoform dp71 using the initial poor samples (1, **open circles**), the other samples (2, **black points)** and the fresh preparation (3, **filled circles**) shows that impaired random priming prevents accurate quantification of long transcripts. **Details of specific cDNA synthesis batches can be found in the underlying data** [67]."

***Competing Interests:*** No competing interests.