Predicting the age of mosquitoes using transcriptional profiles

Peter E Cook[1], Leon E Hugo[1,2], Iñaki Iturbe-Ormaetxe[1], Craig R Williams[3], Stephen F Chenoweth[1], Scott A Ritchie[4,5], Peter A Ryan[2], Brian H Kay[2], Mark W Blows[1] & Scott L O'Neill[1]

[1]School of Integrative Biology, The University of Queensland, Brisbane, Queensland 4072, Australia. [2]Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia. [3]Sansom Institute, School of Pharmacy & Medical Sciences, University of South Australia, Adelaide, South Australia 5000, Australia. [4]Anton Breinl Centre for Public Health and Tropical Medicine, James Cook University, Cairns, Queensland 4870, Australia. [5]Tropical Population Health Unit, Queensland Health, Cairns, Queensland 4870, Australia. Correspondence should be addressed to S.L.O.N. (scott.oneill@uq.edu.au).

Published online 1 November 2007; doi:10.1038/nprot.2007.396

The use of transcriptional profiles for predicting mosquito age is a novel solution for the longstanding problem of determining the age of field-caught mosquitoes. Female mosquito age is of central importance to the transmission of a range of human pathogens. The transcriptional age-grading protocol we present here was developed in *Aedes aegypti*, principally as a research tool. Age predictions are made on the basis of transcriptional data collected from mosquitoes of known age. The abundance of eight candidate gene transcripts is quantified relative to a reference gene using quantitative reverse transcriptase-PCR (RT-PCR). Normalized gene expression (GE) measures are analyzed using canonical redundancy analysis to obtain a multivariate predictor of mosquito age. The relationship between the first redundancy variate and known age is used as the calibration model. Normalized GE measures are quantified for wild-caught mosquitoes, and ages are then predicted using this calibration model. Rearing of mosquitoes to specific ages for calibration data can take up to 40 d. Molecular analysis of transcript abundance, and subsequent age predictions, should take ~3–5 d for 100 individuals.

INTRODUCTION

Age is a key factor underlying the ability of female mosquitoes to transmit pathogens such as dengue viruses and malaria. These pathogens must replicate and disseminate in the mosquito's body before transmission to humans can occur. The duration of this developmental period comprises a significant proportion of the expected lifespan of the mosquito. As a consequence, only a small proportion of mosquitoes in the population live long enough to transmit pathogens. Mosquito population age structure is therefore a critical determinant of the population's capacity for pathogen transmission and a key target for vector control strategies[1].

Mosquito age-grading techniques

Despite the central importance of mosquito age to the understanding of mosquito-borne disease epidemiology, few robust and informative age-grading techniques were available until recently[2,3]. One of the earliest age-grading techniques, based on changes to female reproductive structures, was exceedingly laborious (i.e., ovariolar dissection), and the other using ovarian tracheation only differentiated young or teneral (newly eclosed) adults from those that had previously oviposited[4]. These early methods, while of epidemiological utility, only differentiated a limited number of age classes.

More recent insect age-grading methods have monitored age-related fluctuations in the abundances of pteridines and cuticular hydrocarbons (CHCs). Changes in pteridine levels have been used to predict age in several dipterans[5–7], and the initial application of this methodology with mosquitoes showed promise[8]. However, further studies revealed that it was generally unreliable for predicting mosquito age because pteridines occur in limited quantities and their concentration fluctuates with blood feeding[9]. Changes in the proportion of particular CHCs from mosquito legs have been successfully used to predict adult *Aedes aegypti* age up to 15 d posteclosion in field evaluations[3]. Physiological age is predicted using models that vary depending on the season.

We have reported the development of a transcriptional profiling technique for age-grading *A. aegypti*, the primary dengue vector, under field conditions[2]; here, we present a detailed step-by-step protocol for this method. Our initial report[2] included a comparative study where field-reared mosquito age was predicted using both transcriptional and CHC age-grading techniques. Mosquito legs were removed for CHC quantification, while transcript abundance was determined from the head and thorax. This study demonstrated that transcriptional age predictions fall within ±5 d of actual age, compared with approximately ±10 d for CHC age predictions.

Overview of transcriptional age-grading

The transcriptional age-grading method uses age-related changes in the expression of multiple mosquito genes (**Table 1**) to determine the adult female mosquito age. Age predictions are generated using a multivariate calibration approach (**Fig. 1**). This approach requires a calibration model to be constructed from transcriptional profiles (referred to as training data) collected from individual female mosquitoes of known age. Recently colonized mosquitoes (ideally G1 from field collection) are reared under field conditions and collected at specific ages to generate the training data, which are then used to build the subsequent calibration model. All age predictions are made on the basis of the data contained in the calibration model; therefore, it is important that individuals used to construct the training data are reared under conditions comparable to those experienced by field-collected individuals.

Transcript levels of eight genes of interest and a reference gene (**Table 1**) are measured in the head and thorax of individual mosquitoes using quantitative reverse transcriptase-PCR (qRT-PCR).

TABLE 1 | *Aedes aegypti* genes and their corresponding primer sequences used in the transcriptional age-grading assay. The expression of each gene is normalized to the reference gene, *Ae-RpS17*. Refer to Cook *et al.*[2] for the putative function of each gene. Table modified with permission from ref. 2. Copyright 2006 National Academy of Sciences, USA.

Target gene	Forward primer (5′→3′)	Reverse primer (5′→3′)	Amplicon size (bp)
Ae-RpS17	CACTCCCAGGTCCGTGGTAT	GGACACTTCCGGCACGTAGT	81
Ae-4274	GGACGCTTAGCGGGAAGAC	TTGGCGTTTGGGATTTACCT	81
Ae-4679	GGAGGCTATAACCATCCGAGAA	GGCCAAGGACGGTTCGAT	81
Ae-4916	ACATTCCCACTGAGCGTGAAG	GGTGAGCTCCTCCATCAGGTT	81
Ae-6639	CCACGCGAATATCGATGATG	AGGGAACCTGCGTCCCATAC	81
Ae-7471	CGAGACGATCTCCCGTTAGG	TTTCGAAGGTGATGGAGACCTT	81
Ae-8505	CTCCCTGTGGTTGGAATCCA	TAACCTACACCGCCCCAGAT	81
Ae-12750	ACCCATCGACTGGACACGAA	CCATCCGATAGAGTCGGTGAA	81
Ae-15848	TGACGGCATCCTTGAACTCTT	AGGCTCTGTGGGATGAGATCTC	84

The abdomen is removed before RNA extraction, as blood meals have been shown to induce large transcriptional changes in the midgut and reproductive tissues[10,11]. Gene expression (GE) measures are normalized to the reference gene by calculating logcontrasts (see Step 30). These eight normalized measures are entered into the canonical redundancy analysis, a statistical procedure that reduces the dimensionality of the GE measures by calculating new variates. These new redundancy variates are linear combinations of these GE measures that maximize correlation with mosquito age. The first redundancy variate provides the most informative predictor of adult female mosquito age, and the regression of this redundancy variate on mosquito age represents the calibration model (**Fig. 1**).

Mosquitoes of unknown age (referred to as test data) are then scored for the redundancy variate based on the linear relationship determined by the canonical redundancy analysis. Age predictions are calculated by inverting the regression relationship from the calibration model (**Fig. 1**). A population of individuals of known age, and which has not been used for constructing the calibration model, can be included as blinded samples in the test data set to evaluate the accuracy of age predictions. Estimates of the precision of age predictions are not possible using fiducial limits and, therefore, nonparametric bootstrapping is required to estimate 95% confidence intervals for age predictions. This bootstrapping procedure (see **Supplementary Note**) is implemented using SAS (SAS Institute). See **Box 1** for an overview of multivariate calibration and a glossary of associated terms.

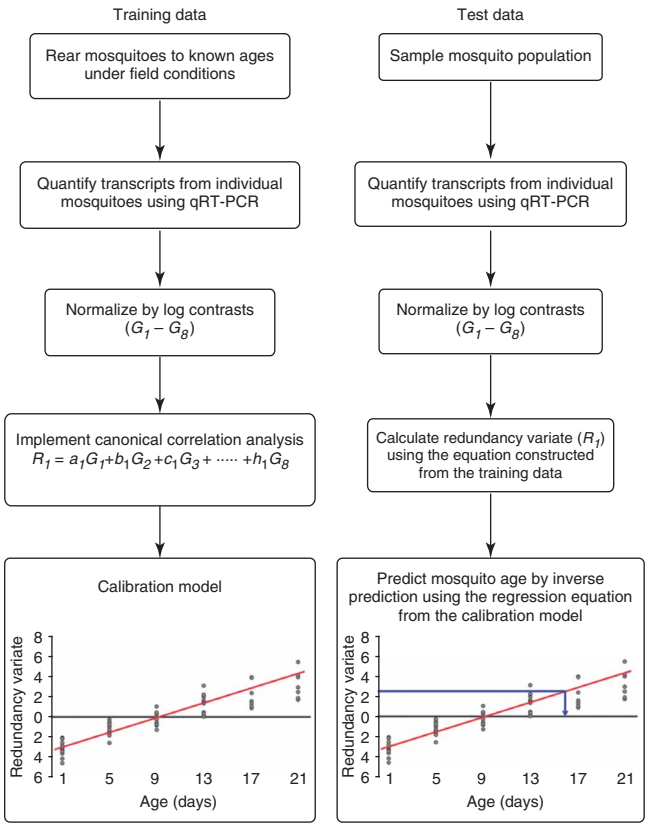

Figure 1 | Overview of the transcriptional age-grading technique. Training data refer to transcriptional profiles, specifically log contrast normalized measures of gene expression (G1–G8), generated from individual mosquitoes reared to known ages. Entering the training data into a canonical redundancy analysis generates a redundancy variate (R_1), which is a linear combination of the gene expression measures that maximizes the correlation with adult female age. The calibration data are represented as a plot of R_1 on mosquito age, and the calibration model is the linear regression of this distribution. Test data denote log contrast normalized measures of transcript abundance from the mosquitoes of unknown age. For each individual of unknown age, R_1 is calculated using the equation constructed from the training data ($R_1 = a_1G_1 + b_1G_2 + c_1G_3 + \ldots + h_1G_8$). Age predictions are calculated from the score for R_1 by inverting the regression equation from the calibration model (i.e., inverse prediction). A nonparametric bootstrapping method is required to estimate the 95% confidence intervals of predicted ages (see Step 33). qRT-PCR, quantitative reverse transcriptase-PCR.

Applications

At present, we consider this technique to be primarily a research tool and not yet an appropriate tool for groups undertaking operational mosquito control. It is hoped that, with further research, the transcriptional age-grading approach will be developed into a high-throughput technique that allows rapid and accurate estimation of mosquito age. A reliable, high-throughput method would allow comprehensive field studies to be undertaken to assess the efficacy of current vector control strategies, particularly new approaches that specifically target vector longevity[1]. Large-scale surveys of population age structure may provide reliable empirical data on the contribution of vector age to pathogen transmission dynamics. For example, a multiplex assay has been recently developed for the simplified assay (P.E.C., I.I.-O. & S.L.O., unpublished data). This assay uses dual-labeled (Taqman) probes (Biosearch Technologies) to monitor the amplification of multiple PCR products in a single tube.

BOX 1 | MULTIVARIATE CALIBRATION; OVERVIEW AND GLOSSARY

An underlying strength of the transcriptional age-grading protocol is the robust multivariate calibration procedures by which age predictions are generated. Calibration problems involve predicting the value of an immeasurable trait (mosquito age) from other indirect traits (gene expression (GE) measures). Multivariate calibration is used extensively in chemometric research, but has widespread application within biology. The body of literature discussing multivariate calibration can be challenging to read, especially when unfamiliar with statistics. To offset this we have included the following glossary of statistical terms used in the text. All statistical procedures in this protocol are implemented in SAS, and the editor syntax has been provided (see **Supplementary Note**). Readers unfamiliar with SAS or the statistical procedures used within SAS can find introductory guides and information within the SAS program help files or online http://support.sas.com/documentation/onlinedoc/sas9doc.html.

Logcontrasts—transformation applied to data that are proportional measures of a unit (compositional data).

Canonical redundancy analysis—multivariate statistical procedure that finds the linear combination of multiple response variables that best predicts a single independent variable.

Fiducial limits—standard measures of error derived from standard distributions.

Nonparametric bootstrapping—resampling procedure that allows confidence limits to be generated for parameters that have unknown distributions.

Current limitations

Cost. A current limitation to large-scale application of the transcriptional age-grading technique is that the method requires a number of genes to be analyzed, which constrains sample throughput and increases assay costs. Current assay cost is between AUD$10 and 15 per mosquito, which is more expensive per mosquito than gas chromatography analysis required for CHC-based age predictions. Assay expense is also dependent on reagent and plastic-ware costs, which vary considerably across real-time PCR platforms, suppliers and countries. Assay costs would also be expected to decrease with increasing sample throughput and multiplexing capabilities. In our initial study[2], we reported a simplified assay that predicted mosquito age using only the three most informative genes, *Ae-15848*, *Ae-8505* and *Ae-4274* (**Table 1**). Age predictions made from this simplified assay were slight over-estimations of the actual age of individuals. This bias was apparent in the age predictions across most age classes and may be tolerable, particularly given the considerable simplification of the assay. The identification of additional candidate genes that are transcriptionally active in older age classes may remove the bias in the simplified assay and improve the accuracy of age predictions for older individuals.

Need for further validation. Transcriptional age-grading is a new research tool, and researchers should be aware that aspects of the technique require additional validation and optimization. For instance, while we expect the protocol described here to be applicable to all field populations of *A. aegypti*, we are yet to assess the qRT-PCR assays in geographically distinct mosquito populations. It may be possible that sequence polymorphisms between different populations may affect the reliability of the qRT-PCR assays and further primer optimization may be required to use the technique in other geographical regions.

Ongoing studies

Additional field validation. A release–recapture study has recently been conducted to validate the method with free-roaming mosquitoes in Cairns, North Queensland, Australia (L.E.H., S.A.R. & S.L.O., personal communication). Recently colonized *A. aegypti* were released and females were collected every 2 d, up to 29 d posteclosion, using adult resting catches. Preliminary results support our initial field validation[2]; however, we are yet to evaluate the transcriptional age-grading technique in geographically distinct *A. aegypti* strains.

Seasonal evaluation. We are currently assessing seasonal fluctuations in *A. aegypti* population age structure using this assay. In addition to providing new data on demographic changes in mosquito populations, this study should provide preliminary data on the variability seen in calibration data collected across seasonal gradients.

Experimental design

Storage of mosquito samples. Samples stored in RNA*later* can be stored for extended periods of time without RNA degradation. Samples can be safely stored at 37 °C for 1 d, 4 °C for 1 month and at –20 °C indefinitely. When conducting field collections in remote areas, plan the storage and transport of samples back to your laboratory for molecular analysis. RNA*later* provides a safe, long-term storage solution when you have access to wet ice or basic refrigeration. RNA*later* is nonflammable, allowing for DNA/RNA samples to be transported by plane in a standard cooler containing wet ice.

RNA extraction. Numerous RNA extraction kits are commercially available, which provide reliable results. Generally, RNA isolation protocols using TRIzol reagent provide higher RNA yields than column-based methods[12]. Column-based methods, while more expensive, tend to yield highly pure RNA. Researchers need to be aware that phenolic compounds in TRIzol reagent can coprecipitate with the RNA and can inhibit later enzymatic reactions. We recommend using the TRIzol isolation method when isolating total RNA from small biological samples like individual mosquitoes.

RNA quantification. RNA quantity and quality can be assessed using a Nanodrop spectrophotometer, Ribogreen (Invitrogen), Agilent's Bioanalyser, or Bio-Rad's Experion. It is important that all samples being compared are quantified using the same technique[12] as all these quantification methods vary in terms of accuracy and reproducibility. Note that the DNase treatment is

completed after determining the quantity of total RNA. It is preferable to digest any contaminating DNA and then repurify the sample to remove free nucleotides and small oligonucleotides before RNA quantification. However, due to concerns of maintaining rare mRNA transcripts, the DNase treatment was done after RNA quantification.

DNase I treatment of RNA sample. Researchers should ensure that they employ an effective DNase I treatment. We have experienced variable performance with DNase I from different suppliers, and we recommend that any DNase I treatment be validated before conducting this protocol. We have previously used Roche DNase I (RNase-free) with Promega RQ1 10× reaction buffer and stop solution with good results.

First-strand cDNA synthesis. The RT reaction is considered to be a major source of variability in qRT-PCR experiments. The cDNA priming strategy (i.e., oligo(dT), random primers or gene-specific primers)[13], type of RT[14] and experimental conditions used can all influence cDNA yield[15]. When processing large numbers of RT reactions, it is important to avoid incorporating batch-to-batch variability from the RT reactions into your experimental treatments (i.e., age classes or collection periods). Therefore, we randomly distribute samples from different experimental treatments into different RT reaction batches.

Primer design. It is important to ensure that qRT-PCR primers are designed properly to ensure specific amplification and accurate quantification of GE. Poor primer design may result in low amplification efficiency, unspecific amplification or complete assay failure (see TROUBLESHOOTING). New primers may need to be designed to avoid sequence polymorphisms between mosquito species and strains. If new primers are needed, they should be designed from coding DNA sequence following the primer design resources and considerations outlined by Nolan *et al.*[12].

Design of qRT-PCR assay. Quantitative PCR (qPCR) relies on a fluorescence detection system to monitor the amplification of PCR products in real time. The appropriate detection system will depend on the number of amplicons being used for the age-grading method and the sample throughput required. Dyes that bind double-stranded DNA, such as SYBR Green I, are the most cost-effective and simplest detection systems to optimize. Reaction specificity is determined exclusively by primer sequence; however, melt curve analysis can potentially be used to identify primer–dimers or amplification of nonspecific products. Intercalating dyes are the obvious starting point for researchers developing a transcriptional age-grading technique in a new system. Once an age-grading technique is validated, dual-labeled fluorescent (Taqman) probes may allow for a high-throughput multiplex (multiple amplicons per reaction) assay to be developed.

Calculating qPCR C_ts. Numerous methods are available for analyzing qPCR amplification plots and determining C_t values. When using the Roche Lightcycler or the Corbett Research Rotor-Gene, we prefer to assign C_t values automatically using the second-derivative maximum[16,17] or comparative quantification methods (Rotor-Gene software, Corbett Research), respectively. See Nolan *et al.*[12] for the discussion on manually assigning threshold values on other real-time platforms.

MATERIALS

REAGENTS

- 10% (wt/vol) sucrose solution
- RNA*later* (Ambion, cat. no. 7021)
- TRIzol reagent (Invitrogen, cat. no. 15596-018) ! CAUTION Toxic and should be handled under a fumehood.
- Chloroform ! CAUTION Toxic and should be handled under a fumehood.
- Lithium chloride ! CAUTION Toxic.
- Water (DNase/RNase-free)
- 70% ethanol
- DNase I, RNase-free (Roche, cat. no. 10 776 785 001)
- DNase I incubation buffer and DNase I stop solution (Promega, cat. no. M6101)
- Oligo(dT)$_{15}$ (Promega, cat. no. C1101)
- dNTP mix (RNase-free)
- RNaseOUT Recombinant ribonuclease inhibitor (Invitrogen, cat. no. 10777-019)
- Superscript III reverse transcriptase (Invitrogen, cat. no. 18080-044) ▲ CRITICAL We recommend Superscript III as it is a thermostable, engineered M-MLV RT, which should minimize the problems associated with RNA secondary structure and provide high cDNA yield. Significant gene-specific variability has been reported in the efficiency of different RTs[14]. We have not tested the reproducibility of this assay using RTs other than Superscript III.
- Platinum SYBR green qPCR Supermix-UDG (Invitrogen, cat. no. 11733-046)
- Oligonucleotides (see **Table 1**)
- *Taq* DNA polymerase for standard PCR ▲ CRITICAL All reagents, plasticware and glassware must be RNase-free.

EQUIPMENT

- Standard mosquito rearing equipment (e.g., cages, larval trays)
- BG-Sentinel traps (BioGents GmbH) or another mosquito trap that does not kill collected mosquitoes
- Mini bead-beater (Biospec Products) or another device to homogenize mosquito tissues
- Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies) or another device for RNA quantification
- Real-time thermal cycler
- SAS software (version 9.1; SAS Institute)
- Microfuge tubes
- Glass beads

PROCEDURE

Rearing and collection of *A. aegypti* ● TIMING ~38 d

1| Hatch eggs collected from recently colonized mosquitoes (e.g., G1–2 from field). Rear larvae and adult mosquitoes under ambient conditions at field site. Maintain larvae at low densities (~150 larvae l^{-1}) to promote synchronous development and transfer pupae to a cage to eclose. Adults that eclose within a 24-h period are considered a single cohort. Remaining pupae can

be transferred to a new cage to establish the next cohort, and this process can be continued until sufficient numbers of adults are obtained. Supply adults with 10% (wt/vol) sucrose solution and offer a blood meal every 4 d.
▲ CRITICAL STEP If sufficient adult female mosquitoes are to be alive at 29 d of age for collection, then at least 4,000 adults would need to be hatched, assuming daily survival of ~0.85 and equal sex ratio.
▲ CRITICAL STEP Mosquitoes need to be provided with a food regime that will generate individuals of similar size to wild caught individuals. Size is easily assessed through wing length measures.

2| Collect individual female mosquitoes at specific ages (e.g., 1, 5, 9, 13, 17, 21, 25 and 29 d posteclosion) from multiple cages. Dissect the head and thorax from the mosquito and immediately store in RNA*later* following manufacturer's recommendations. Previous results suggest 10–15 individuals per age class should be adequate for constructing a robust calibration model[2]. Additional individuals can be collected at each age to establish blinded samples. Blinded samples can be included in the test data set and used to evaluate the accuracy of age predictions.
■ PAUSE POINT Samples stored in RNA*later* can be safely stored without RNA degradation at 37 °C for 1 d, 4 °C for 1 month and at −20 °C indefinitely.

3| Sample mosquitoes from the wild population using an appropriate trapping method, such as the BG-Sentinel trap or backpack aspirator[18]. Some adult mosquito sampling methods may be biased toward the collection of certain age classes, and this should be considered if the age of individual mosquitoes are used to model population age structure. Remove abdomens and store heads and thoraces as above.
■ PAUSE POINT Samples stored in RNA*later* can be safely stored without RNA degradation at 37 °C for 1 d, 4 °C for 1 month and at −20 °C indefinitely.

Total RNA isolation ● TIMING ~3–24 h depending on precipitation time
4| Remove mosquito tissues from RNA*later* and immediately transfer into a 1.5-ml microfuge tube containing 500 µl TRIzol reagent and a 2.5-mm glass bead. Place in the Bead-beater for 1–2 min to homogenize the sample. If a Bead-beater cannot be accessed, comparable results can be obtained by freezing each sample in liquid nitrogen and grind with a micropestle before adding the TRIzol reagent.

5| Proceed with the TRIzol reagent RNA extraction following the manufacturer's protocol from the phase separation step, adding an additional chloroform step to minimize carry over of DNA, proteins and phenolics during the transfer of the aqueous phase.

6| Add lithium chloride (to 2 M final concentration) and place at −30 °C for 1 h to overnight to precipitate the RNA.

7| Collect precipitated RNA by centrifugation and discard supernatant.

8| Wash RNA pellet with 75% (vol/vol) ethanol, centrifuge again, remove all ethanol using a pipette, and then air dry pellet.
▲ CRITICAL STEP Do not completely dry the RNA pellet as this may make dissolution of the pellet more difficult.

9| Dissolve RNA in 20 µl RNase-free water by incubating the sample at 55 °C for up to 10 min.
▲ CRITICAL STEP For RT-PCR applications, we recommend that RNA be dissolved in RNase-free water rather than in TE buffer. Dissolution of RNA in TE buffer provides more stable long-term storage, but the presence of EDTA may interfere with later enzymatic reactions.
■ PAUSE POINT Samples can be stored at −80 °C until further analysis, if necessary.

Quantification of RNA ● TIMING 2 min per sample
10| Initialize the Nanodrop and blank with 2 µl RNase-free water. Wipe water from sample pedestal using a Kimwipe.

11| Dispense 2 µl of undiluted RNA onto the clean sample pedestal and measure RNA quantity and purity. The Nanodrop software displays the absorbance spectrum for the sample and automatically calculates RNA concentration and two absorbance ratios. The A_{260}:A_{280} absorbance ratio indicates the purity of the RNA preparation with respect to the presence of proteins and phenol[19]. Pure RNA preparations have A_{260}:A_{280} ratios of 1.8–2.0. RNA samples resuspended in water often have acidic pH that significantly lowers the A_{260}:A_{280} ratio (by ~0.2–0.3)(ref. 20). The A_{260}:A_{230} ratio indicates purity with regard to organic compounds, such as phenolate ions and thiocyanates[19]. Pure RNA preparations should have A_{260}:A_{230} ratios between 2.0 and 2.3. High-level absorbance at 320 nm indicates the presence of suspended particulates[19].
▲ CRITICAL STEP Vortex, then collect the RNA samples by brief centrifugation before taking spectrophotometric readings to ensure the sample is homogeneously mixed.

Removal of genomic DNA in RNA sample ● TIMING 50 min

12| To remove genomic DNA, treat the RNA samples with DNase I. Add the following to a microcentrifuge tube on ice:

Component	Volume per reaction	Final concentration
1 μg total RNA		100 ng μl^{-1}
Dnase I	0.5 μl	5 U
10× DNase incubation buffer	1 μl	1×
RNase-Free water	To 10 μl	

▲ **CRITICAL STEP** RNA samples need to be free of contaminating genomic DNA to ensure reliable qRT-PCR results are obtained.

13| Incubate the mixture at 37 °C for 30 min.

14| Add 1 μl of DNase I stop solution and incubate at 65 °C for 10 min to terminate the reaction.

First-strand cDNA synthesis ● TIMING ~2 h

15| Add 1 μl (500 ng) of oligo$(dT)_{15}$, 1 μl of 10 mM dNTPs (RNase-free) to the DNase-treated RNA.

16| Incubate at 65 °C for 5 min, then transfer immediately to ice for at least 2 min.

17| Centrifuge briefly to collect the reaction. Transfer half of the reaction (6.5 μl) to a clean microcentrifuge tube and assign each tube as either +RT or –RT (no RT) reaction. The –RT reaction functions as a control to indicate the contaminating genomic DNA.

18| Add the following components to the +RT and –RT reactions.

	Volume per reaction (μl)	
Component	**+RT**	**–RT**
RNA/oligo$(dT)_{15}$ mix	6.5	6.5
5× RT buffer	4	4
0.1 M DTT	1	1
RNase OUT	1	1
Superscript III	1	—
RNase-free water	6.5	7.5
Total	20	20

19| Mix and centrifuge briefly. Incubate at 50 °C for 60 min.

20| Inactivate the reactions by heating to 70 °C for 15 min.

21| Add 80 μl water (RNase-free) to the +RT and –RT reactions to dilute out PCR inhibitors.

■ **PAUSE POINT** First-strand cDNA can be stored at –20 °C for at least 6 months.

PCR screen for contaminating genomic DNA ● TIMING ~2–3 h

22| Combine the components listed in the table below to make a PCR mastermix. This PCR screen can be done using any mosquito gene; however, we routinely use the reference gene *Ae-RpS17* (see **Table 1** for primer sequences). Make adequate mastermix to screen all +RT and –RT reactions, a positive control and a no-template control. Mix thoroughly and dispense 18 μl into PCR tubes.

Component	Volume per reaction (μl)	Final concentration
DH_2O	9.6	
10× PCR buffer	2	1×
25 mM $MgCl_2$	1.2	1.5 mM
2.5 mM dNTPs	1	125 μM
10 μM Forward primer	2	1 μM
10 μM Reverse primer	2	1 μM
Taq DNA polymerase	0.2	

23| Add 2 µl of the appropriate cDNA template from Step 21 to each reaction, vortex and centrifuge briefly. In a standard thermal cycler, run the following thermal profile:

1 cycle	95 °C	3 min
35 cycles	95 °C	15 s
	60 °C	15 s
	72 °C	30 s
1 cycle	72 °C	5 min

24| Separate PCR products on a 2% (wt/vol) agarose gel by electrophoresis. –RT reactions should be negative, that is, no bands should be seen in lanes containing –RT reactions. Discard any samples with –RT reactions that show PCR amplification as they may contain genomic DNA contamination.

■ **PAUSE POINT** DNase-treated cDNA can be stored at –20 °C for at least 6 months.

Quantitative RT-PCR ● TIMING < 2 h

25| Prepare a mastermix of the components listed below for each target gene (see **Table 1** for primer sequences). Make adequate mastermix to run triplicate reactions for each template and a no-template control. Triplicate reactions control for variability due to the amount of template added. The no-template control should not amplify, thereby verifying that no contaminating DNA or primer dimers are present. Vortex and dispense 8 µl of mastermix into the reaction tubes or wells.

Component	Volume per reaction (µl)	Final concentration
RNase-free water	2	
SYBR Supermix	5	1×
10 µM Forward primer	0.5	500 nM
10 µM Reverse primer	0.5	500 nM

▲ **CRITICAL STEP** Some real-time PCR machines require BSA or ROX reference dye to be added into the qPCR mastermix. Researchers should refer to the manufacturer's specifications for their real-time thermal cycler. If BSA or ROX are required, then decrease the volume of water in the mastermix accordingly.

26| Add 2 µl of the appropriate cDNA template from Step 21 (~10 ng) to the reaction tubes.

27| Carefully seal the reaction tube and centrifuge briefly to collect the reaction. Place samples in the real-time thermal cycler.

28| Perform the following three-step thermal profile:

1 cycle	UDG incubation[a]	50 °C	2 min	
1 cycle	*Taq* activation[b]	95 °C	2 min	
40–50 cycles	Denaturation	95 °C	5 s	
	Annealing	60 °C	5 s	
	Extension	72 °C	10 s	Acquire data
1 cycle	Melting curve analysis	68–95 °C	1 °C steps	Acquire data each step

[a]Uracil DNA glycosylase (UDG) prevents the amplification of contaminating quantitative PCR (qPCR) products, as dUTP is incorporated into amplified products. [b]Platinum *Taq* polymerase is coupled with an antibody that blocks polymerase activity at ambient temperatures. This activation step is required only for 'hot-start' *Taq* polymerases and can be omitted for standard *Taq* polymerases. Refer to manufacturer specifications when using 'hot-start' polymerases, as the activation requirements vary for chemically modified *Taq* polymerases (e.g., 95 °C for 10 min).

29| Determine the mean C_t value for each target gene (see Calculating qPCR C_ts section).

Compile training and test data files ● TIMING ~1 h

30| Compile qRT-PCR data into separate data files for both the training and test data (see **Tables 2** and **3**). Calculate the mean C_t values for each gene and determine relative measures of GE by calculating logcontrasts. Logcontrasts are calculated as[21]:

$$\text{logconstrast } X_i = \log_{10}\left(\frac{X_i/X_{total}}{X_{ref}/X_{total}}\right)$$

where X_i is the mean C_t value of a gene, X_{ref} is the mean C_t of the reference gene (*Ae-RpS17*) and X_{total} is the sum of the C_t values for all genes from an individual.

TABLE 2 | Structure of the training data file that will be used in the canonical correlation analysis.

Sample	Age	Rep	G1	G2	G3	G4	G5	G6	G7	G8
1	1	1								
2	1	2								
3	1	3								
4	1	4								
5	5	1								
6	5	2								
7	5	3								

The training data set contains the following variables: sample identifies each individual mosquito of known age; Rep denotes replicates within each age class, and G1–8 are log contrast normalized gene expression measures included in the analysis.

TABLE 3 | Structure of the test data file. *Sample* and *Rep* are identifiers of each individual mosquito of unknown age.

New_id	Sample	Rep	G1	G2	G3	G4	G5	G6	G7	G8
1	1	1								
2	3	2								
3	4	3								
4	6	4								
5	8	5								
6	12	2								
7q	12	3								

New_id is a consecutive integer that is required when implementing the bootstrap procedure. G1–8 are the log contrast normalized gene expression measures included in the analysis.

Construct calibration model ● TIMING 30 min

31| Open SAS. Copy and paste the first SAS editor syntax (see **Supplementary Note**) into the editor window. Modify the data file directory information (line 2 of syntax) to locate the training data file before running the procedure. This syntax implements a canonical redundancy analysis that generates a redundancy variate for each individual in the training data. In addition, the syntax implements a linear regression of the calculated redundancy variates on age, which will describe the calibration model.

32| Examine the SAS log to ensure that no errors have occurred. Then examine the SAS output, noting the amount of variance in normalized GE measures, which are explained by age (**Fig. 2**) and the fit of the regression (R^2 value) of the redundancy variate on mosquito age (**Fig. 3**). The redundancy variate (termed 'red') calculated for each individual in the training data is accessible in the SAS worktable 'redund_data' and can be used to graphically display the calibration data.

Generate age predictions with 95% confidence limits ● TIMING 10 min

33| Copy and paste the second SAS editor syntax (see **Supplementary Note**) into a new SAS editor window. Modify the SAS syntax to suit experimental parameters by adjusting: youngest age class, oldest age class, age class interval, number

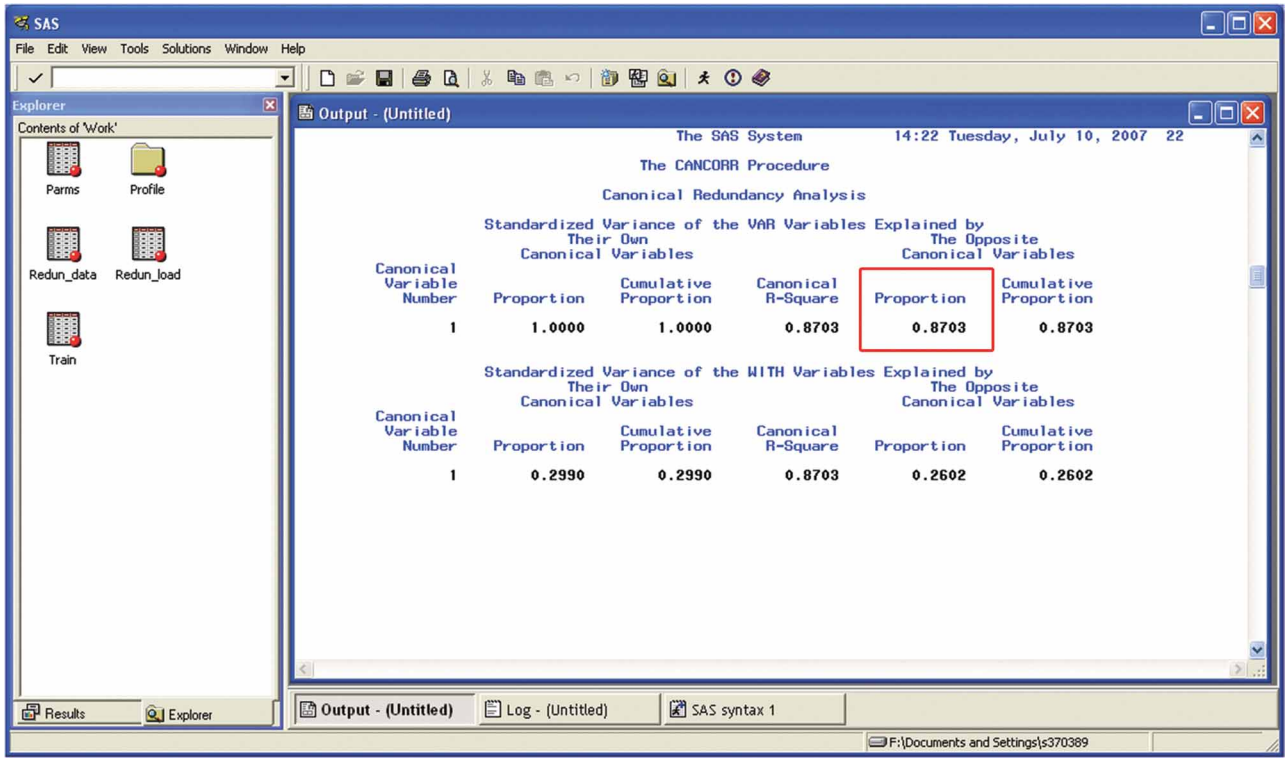

Figure 2 | SAS output file showing the percentage of variance in gene expression measures (highlighted with red box) explained by adult female mosquito age. This value indicates the degree of association between gene expression measures and mosquito age.

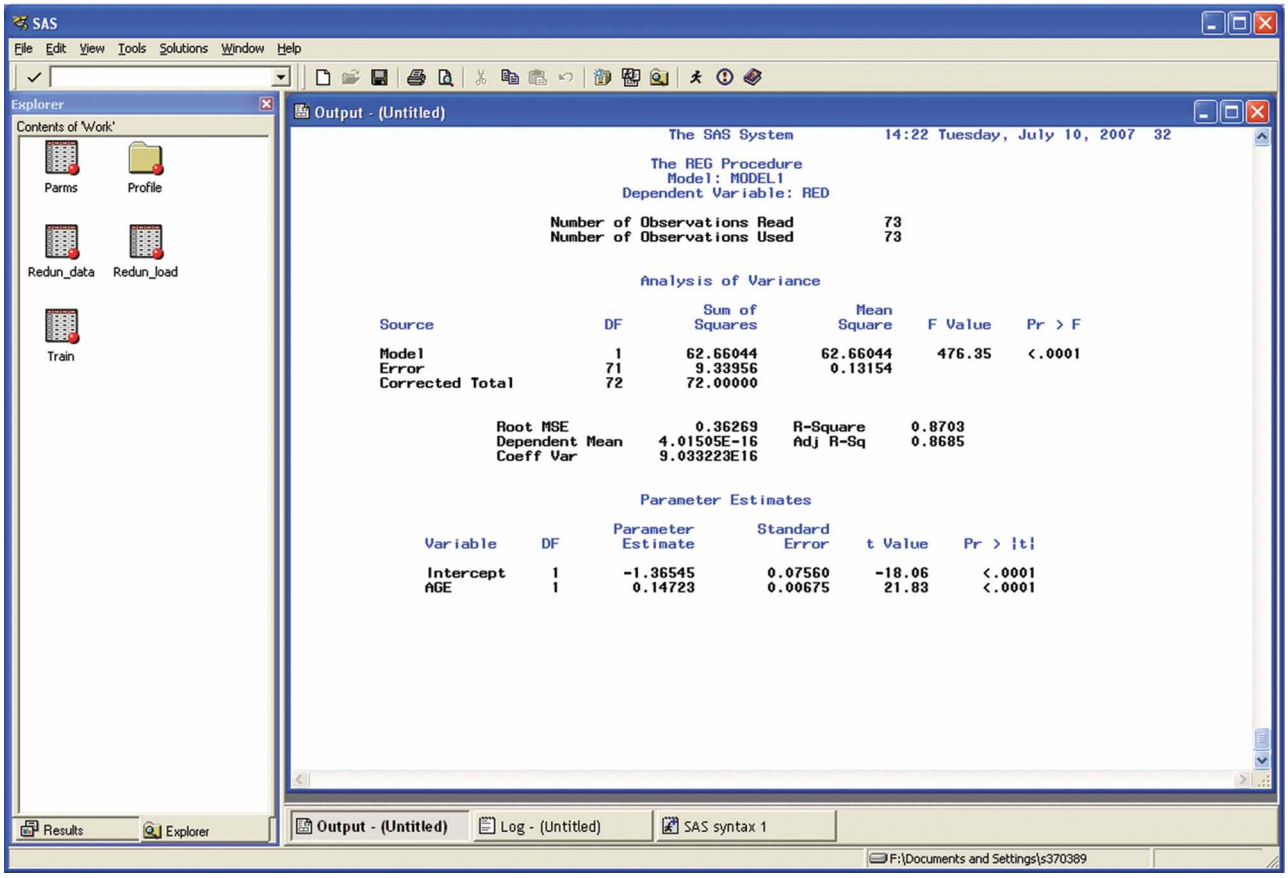

Figure 3 | SAS output showing the regression statistics for linear regression of the first redundancy variate on adult female mosquito age. Note the coefficient of correlation (R^2) to evaluate the fit of the regression.

of observations in each age class (lines 10–13 of SAS syntax) and data file directory information following all import and export statements. This syntax implements a nonparametric bootstrapping procedure to derive 95% confidence limits for age predictions. Upper and lower confidence limits for the predicted age of each individual in the test data set will be exported to the Microsoft Excel spreadsheet indicated in the SAS syntax (**Fig. 4**).

▲ **CRITICAL STEP** This SAS syntax can only be used when the training data has an equal number of replicates in each age class and age classes are at set intervals. Minor modification of the SAS syntax will be required for the studies involving unequal samples sizes and age class interval. Users will have to define each age and number of replicates in each age class in the SAS syntax to bootstrap data correctly.

? TROUBLESHOOTING

● TIMING

Steps 1–3, rearing and collection of *A. aegypti*: ~38 d
Steps 4–9, RNA extraction: up to 24 h
Steps 10–11, quantification of RNA: 2 min per sample
Steps 12–24, first-strand cDNA synthesis and –RT screen: 8 h
Steps 25–29, quantitative RT-PCR: ~2 h
Step 30, compile data: ~1 h
Steps 31 and 32, construct calibration model: 30 min
Step 33, generate age predictions: 10 mins

? TROUBLESHOOTING

Troubleshooting for transcriptional age-grading protocol mainly concerns the optimization of the molecular biology techniques. For extensive troubleshooting advice concerning RNA isolation or qRT-PCR, refer to Chomczynski and Sacchi[22] and Nolan *et al.*[12], respectively. Transcriptional age-grading has yet to be evaluated in geographically distinct *A. aegypti* strains. It may be possible that the performance of the qRT-PCR assays used for transcriptional age determination may vary between *A. aegypti* strains. The qRT-PCR may perform poorly in different *A. aegypti* strains due to sequence polymorphisms in the primer-binding sites of the target genes. If these problems occur, researchers will need to obtain strain-specific sequence data and redesign qRT-PCR primers avoiding sequence polymorphisms. The SAS log file will provide error messages related to the data analysis procedures.

ANTICIPATED RESULTS

The amount of total RNA isolated from a single mosquito head and thorax should be ~2–8 µg. The transcriptional profiles of the eight genes of interest across mosquito age vary considerably. Transcript abundance of two genes, *Ae-15848* and *Ae-8505*, decrease noticeably in older age classes. This decrease in transcription is indicated by an increase in C_t value. *Ae-15848* is expressed at the levels comparable to the reference gene, *Ae-RpS17*. Investigators can also expect the C_t values for *Ae-15848* to be lower than that of the reference gene in the youngest age class. It should also be noted that some genes display only minor changes in C_t values across age. However, these genes provide additional information to the calibration model that removes the bias present in the three gene assays. The calibration model generated from eight genes should have an observable linear relationship with age. It should be noted that the scale of the calibration model may vary to that presented by Cook *et al.*[2] due to their use of standardized data. Based on these results, we expect that female *A. aegypti* age to be predicted ± 5 d of actual age.

predicted_age

	A	B	C	D	E	F
1	sample	new_id	lower	upper		
2						
3	1	1	10.09344	13.0483		
4	2	2	12.11861	15.72818		
5	3	3	0.852686	2.919104		
6	5	4	13.49466	16.71417		
7	6	5	7.004522	9.059627		
8	8	6	13.75682	17.51183		
9	9	7	0.240535	3.459309		
10	10	8	4.153216	6.389264		
11	15	9	1.985654	4.727147		
12	16	10	16.34001	21.17434		
13	17	11	11.59242	15.33376		
14	18	12	8.949103	11.43264		
15	20	13	4.574292	7.599591		
16	21	14	4.409799	6.051764		
17	22	15	5.082459	6.968591		
18	23	16	2.239143	4.211618		
19	24	17	2.029668	3.846611		
20	25	18	5.625098	7.474483		
21	26	19	1.734209	5.014636		
22	27	20	2.975582	5.552		
23	28	21	0.854074	5.116615		
24	29	22	4.312592	7.526282		
25	30	23	7.061066	9.428554		
26	31	24	7.532393	9.687286		
27	32	25	9.368541	14.87676		
28	34	26	15.55998	19.9368		
29	35	27	10.92027	13.90951		
30	36	28	16.38667	21.8842		
31	38	29	12.2236	16.16973		
32	39	30	13.78594	18.63737		
33						
34						
35						

PRED_AGE | CI_LIMITS

Figure 4 | Microsoft Excel worksheet exported from the second SAS procedure. Upper and lower limits of the 95% confidence interval of the predicted age for each individual mosquito in the test data set.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS We are grateful to Dr. Jeremy C. Brownlie, Conor J. McMeniman and Grant Hughes for constructive reviews of earlier drafts of this manuscript. This research was funded by grants from the Australian Research Council (LP0455732), the Mosquito and Arbovirus Research Committee, and the Foundation for the National Institutes of Health through the Grand Challenges in Global Health Initiative.

Published online at http://www.natureprotocols.com
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Cook, P.E., McMeniman, C.J. & O'Neill, S.L. Modifying insect population age structure to control vector-borne disease. In *Transgenesis and the Management of Vector-borne Disease* (ed. Askoy, S.) (Landes Bioscience, Austin, TX, 2007).
2. Cook, P.E. *et al.* The use of transcriptional profiles to predict adult mosquito age under field conditions. *Proc. Natl. Acad. Sci. USA* **103**, 18060–18065 (2006).
3. Gerade, B.B. *et al.* Field validation of *Aedes aegypti* (Diptera: Culicidae) age estimation by analysis of cuticular hydrocarbons. *J. Med. Entomol.* **41**, 231–238 (2004).
4. Detinova, T.S. Age-grouping methods in Diptera of medical importance with special reference to some vectors of malaria. *Monogr. Ser. World Health Organ.* **47**, 13–191 (1962).
5. Hayes, E.J. & Wall, R. Age-grading adult insects: a review of techniques. *Physiol. Entomol.* **24**, 1 (1999).
6. Lehane, M.J. Determining the age of an insect. *Parasitol. Today* **1**, 81–85 (1985).
7. Robson, S.K., Vickers, M., Blows, M.W. & Crozier, R.H. Age determination in individual wild-caught *Drosophila serrata* using pteridine concentration. *J. Exp. Biol.* **209**, 3155–3163 (2006).
8. Wu, D. & Lehane, M.J. Pteridine fluorescence for age determination of *Anopheles* mosquitoes. *Med. Vet. Entomol.* **13**, 48–52 (1999).
9. Penilla, R.P., Rodríguez, M.H., López, A.D., Viader-Salvadó, J.M. & Sánchez, C.N. Pteridine concentrations differ between insectary-reared and field-collected *Anopheles albimanus* mosquitoes of the same physiological age. *Med. Vet. Entomol.* **16**, 225–234 (2002).
10. Holt, R.A. *et al.* The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129–149 (2002).
11. Sanders, H.R., Evans, A.M., Ross, L.S. & Gill, S.S. Blood meal induces global changes in midgut gene expression in the disease vector, *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **33**, 1105–1122 (2003).
12. Nolan, T., Hands, R.E. & Bustin, S.A. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* **1**, 1559–1582 (2006).
13. Ståhlberg, A., Håkansson, J., Xian, X., Semb, H. & Kubista, M. Properties of the reverse transcription reaction in mRNA quantification. *Clin. Chem.* **50**, 509–515 (2004).

14. Ståhlberg, A., Kubista, M. & Pfaffl, M. Comparison of reverse transcriptases in gene expression analysis. *Clin. Chem.* **50**, 1678–1680 (2004).
15. Bustin, S.A. & Nolan, T. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* **15**, 155–166 (2004).
16. Luu-The, V., Paquet, N., Calvo, E. & Cumps, J. Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction. *BioTechniques* **38**, 287–293 (2005).
17. Rasmussen, R. Quantification on the Lightcycler. In *Rapid Cycle Real-time PCR: Methods and Applications* (eds. Meuer, S.C., Wittwer, C. & Nakagawara, K.) pp. 21–34 (Springer, Heidelberg, Germany, 2001).
18. Williams, C.R., Long, S.A., Russell, R.C. & Ritchie, S.A. Field efficacy of the BG-sentinel compared with CDC backpack aspirators and CO_2-baited EVS traps for collection of adult *Aedes aegypti* in Cairns, Queensland, Australia. *J. Am. Mosq. Control Assoc.* **22**, 296–300 (2006).
19. Sambrook, J. & Russell, D.W. *Molecular Cloning: A Laboratory Manual.* 3rd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
20. Wilfinger, W.W., Mackey, K. & Chomczynski, P. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *BioTechniques* **22**, 474–476 (1997).
21. Aitchison, J. *The statistical analysis of compositional data* Reprint ed. (Blackburn Press, Caldwell, NJ, 2003).
22. Chomczynski, P. & Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat. Protoc.* **1**, 581–585 (2006).