

1 **Accurate Predictions of Postmortem Interval Using**
2 **Linear Regression Analyses of Gene Meter Expression**
3 **Data**

4
5 Authors: M. Colby Hunter¹, Alex E. Pozhitkov², and Peter A. Noble^{1,2,3}

6 **Author Affiliations:**

7 1. Ph.D. Microbiology Program, Department of Biological Sciences, Alabama State
8 University, Montgomery, Alabama, USA 36104

9 2. Department of Oral Health Sciences, University of Washington, Box 357444, Seattle,
10 WA USA 98195.

11 3. Department of Periodontics, School of Dentistry, Box 355061, University of
12 Washington, Seattle, Washington, USA 98195

13 *Correspondence to:

14 Peter A Noble

15 Email: panoble@u.washington.edu

16 Phone: 206-409-6664.

17

18 Other authors' emails:

19 MCH: cmooghww@gmail.com

20 AEP: pozhit@u.washington.edu

21 **Short title:**

22 PMI Prediction Using Gene Meter Analysis of Expression Data

23 **Keywords**

24 Postmortem transcriptome; postmortem gene expression; Gene meters; calibrated DNA
25 microarrays, thanatotranscriptome; postmortem interval, forensic science.

26 **Abstract**

27

28 In criminal and civil investigations, postmortem interval is used as evidence to help sort
29 out circumstances at the time of human death. Many biological, chemical, and physical
30 indicators can be used to determine the postmortem interval – but most are not accurate.
31 Here, we sought to validate an experimental design to accurately predict the time of death
32 by analyzing the expression of hundreds of upregulated genes in two model organisms,
33 the zebrafish and mouse. In a previous study, the death of healthy adults was conducted
34 under strictly controlled conditions to minimize the effects of confounding factors such as
35 lifestyle and temperature. A total of 74,179 microarray probes were calibrated using the
36 Gene Meter approach and the transcriptional profiles of 1,063 significantly upregulated
37 genes were assembled into a time series spanning from life to 48 or 96 h postmortem. In
38 this study, the experimental design involved splitting the gene profiles into training and
39 testing datasets, randomly selecting groups of profiles, determining the modeling
40 parameters of the genes to postmortem time using over- and/or perfectly- defined linear
41 regression analyses, and calculating the fit (R^2) and slope of predicted versus actual
42 postmortem times. This design was repeated several thousand to million times to find the
43 top predictive groups of gene transcription profiles. A group of eleven zebrafish genes
44 yielded R^2 of 1 and a slope of 0.99, while a group of seven mouse liver genes yielded a
45 R^2 of 0.98 and a slope of 0.97, and seven mouse brain genes yielded a R^2 of 0.93 and a
46 slope of 0.85. In all cases, groups of gene transcripts yielded better postmortem time
47 predictions than individual gene transcripts. The significance of this study is two-fold:
48 selected groups of upregulated genes provide accurate prediction of postmortem time,
49 and the successfully validated experimental design can now be used to accurately predict
50 postmortem time in cadavers.

51

52

53 **Introduction**

54 The postmortem interval (PMI) is the elapsed time between death of an organism and the
55 initiation of an official investigation to determine the cause of death. Its determination is
56 important to civil investigations such as those involving life insurance fraud because
57 investigators need to determine if the person was alive or not when the policy was in
58 effect [1]. The PMI is also important to criminal investigations, especially suspicious
59 death cases where there are no witnesses, because it can help determine the time
60 relationship between a potential suspect and the victim and eliminate people from a
61 suspect list, which speeds up investigations. Accurate prediction of PMI is considered
62 one of the most important and complex tasks performed by forensic investigators [2].

63 Several studies have suggested that RNA could be used to estimate PMI [3,4,5,6,7].
64 While most studies focused on the degradation of mRNA gene markers, some examined
65 gene expression. The RNA degradation studies include: a model to predict PMI based on
66 the degradation of Beta actin (*Actb*), Glyceraldehyde-3-phosphate dehydrogenase
67 (*Gapdh*), Cyclophilin A (*Ppia*) and Signal recognition particle 72 (*Srp72*) genes in
68 mouse muscle tissue samples [3], a model to predict PMI based on degradation of an
69 amplified *Actb* gene and temperature in rat brain samples [4], and a study that predicted
70 PMI based on the degradation of *Gapdh*, *Actb* and 18S rRNA genes in the spleens of rats
71 [5]. The gene expression studies include: a study that found increased expression of
72 myosin light chain 3 (*Myl3*), matrix metalloprotease 9 (*Mmp9*) and vascular endothelial
73 growth factor A (*Vegfa*) genes in human body fluids after 12 h postmortem [6], a study
74 that found increased expression of Fas Ligand (*Fasl*) and ‘phosphatase and tensin
75 homologue deleted on chromosome 10’ (*Pten*) genes with postmortem time in rats [7],
76 and a study that found individual gene transcripts did not increase using PCR-based gene
77 expression arrays of frozen human brain cadaver samples [8]. Common to these studies
78 is the requirement: (i) to amplify cDNA by polymerase chain reaction (PCR) and (ii) to
79 normalize the data with a control in order to facilitate sample comparisons. These
80 requirements introduce methodological biases that could significantly affect
81 interpretation of the data. An approach that minimizes or eliminates these biases is
82 highly desirable because it might lead to better PMI predictions.

83 Since conventional DNA microarray approaches yield noisy data [9], in 2011 we
84 developed the “Gene Meter” approach that precisely determines specific gene
85 abundances in biological samples and minimizes noise in the microarray signal [10,11].
86 The reason this approach is precise is because the behavior of every microarray probe is
87 determined by calibration – which is analogous to calibrating a pH meter with buffers.
88 Without calibration, the precision and accuracy of a meter is not known, nor can one
89 know how well the experimental data fits to the calibration (i.e., R^2). The advantages of
90 the Gene Meter approach over conventional DNA microarray approaches is that the
91 calibration takes into consideration the non-linearity of the microarray signal and
92 calibrated probes do not require normalization to compare biological samples. Moreover,
93 PCR amplification is not required. We recognize that next-generation-sequencing (NGS)
94 approaches could have been used to monitor gene expression in this study. However, the
95 same problems of normalization and reproducibility are pertinent to NGS technology
96 [12]. Hence, the Gene Meter approach is currently the most advantageous high
97 throughput methodology to study postmortem gene expression and might have utility for
98 determining the PMI.

99 The Gene Meter approach has been used to examine thousands of postmortem gene
100 transcription profiles from 44 zebrafish (*Danio rerio*) and 20 house mice (*Mus musculus*)
101 [13]. Many genes were found to be significantly upregulated (relative to live controls).
102 Given that each sampling time was replicated two or three times, we conjectured that the
103 datasets could be used to assess the feasibility for predicting PMIs from gene expression
104 data. Although many approaches are available to determine PMI (see Discussion), an
105 approach that accurately determines the time of death is highly desired and it is the goal
106 of our study to determine if specific gene transcripts or groups of gene transcripts could
107 accurately predict postmortem time. Zebrafish and mice are ideal for testing
108 experimental designs because the precise time of human deaths is often not known, and
109 other variables such as lifestyle, temperature, and health condition are also often not
110 known or sufficiently controlled in human studies. Given that these variables could have
111 confounding effects on the interpretation of gene expression data in human studies,
112 testing experimental designs under controlled conditions using model organisms is ideal.
113 In our study, the timing of death and health of the zebrafish and mice are known, which

114 enables the testing of different experimental designs to provide “proof of principle”. It is
115 our intent to use the best design to determine PMI of cadavers for future studies.

116 The objectives of our study are twofold: (1) to identify specific upregulated genes or
117 groups of upregulated genes that accurately predict the PMI in the zebrafish and mouse,
118 and (2) to design and evaluate a robust experimental approach that could later be
119 implemented to predict PMI from cadavers.

120 **Materials and Methods**

121 Although the details of zebrafish and mouse processing, the extraction of RNA, and
122 microarray calibrations are presented in a previous study [13], we have provided relevant
123 experimental protocols to aid readers in the interpretation of the results of this study.

124 **Zebrafish processing.** The 44 zebrafish were maintained under standard conditions in
125 flow-through aquaria with a water temperature of 28°C. Prior to sacrifice, the zebrafish
126 were placed into 1 L of water of the same temperature as the aquaria. At zero time, four
127 fish were extracted and snap frozen in liquid nitrogen. These live controls were then
128 placed at -80°C. To synchronize the time of death, the remaining 40 fish were put into a
129 small container with a bottom made of mesh and placed into an 8 L container of ice water
130 for 5 mins. The small container with the mesh bottom was placed into the flow-through
131 aquarium with a water temperature of 28°C for the duration of each individual’s
132 designated postmortem time. The postmortem sampling times used for the zebrafish
133 were: 0, 15 min, 30 min, 1, 4, 8, 12, 24, 48 and 96 h. At each sampling time, 4
134 individuals were taken out of the small container in the flow-through aquarium, snap
135 frozen in liquid nitrogen and then stored at -80°C. One zebrafish sample was not
136 available for use (it was accidentally flushed down the sink) however this was taken into
137 account for calculation of extraction volumes.

138 **Mouse processing.** Twenty C57BL/6JRj male mice of the same age (5 months) and
139 similar weight were used. Prior to sacrifice, the mice had *ad libitum* access to food and
140 water and were maintained at room temperature. At zero time, the mice were sacrificed
141 by cervical dislocation and each mouse was placed in a unique plastic bag with pores to
142 permit the transfer of gases. The mice were kept at room temperature for the designated
143 postmortem sampling times. The sampling times used were: “zero” time, 30 min, 1, 6,

144 12, 24 and 48 h. At each sampling time, a brain and two liver samples were obtained
145 from each of three mice, except for the 48 h sampling where only two mice were
146 sampled. The samples were immediately snap frozen in liquid nitrogen and placed at -
147 80°C.

148 **RNA Processing and Labeling.** Gene expression samples for each PMI were done in
149 duplicate for zebrafish and in triplicate for mice (except for the 48 h PMI sample that was
150 duplicated). The zebrafish samples were homogenized with a TissueLyzer (Qiagen) with
151 20 ml of Trizol. The mouse brain and liver samples (~100 mg) were homogenized in 1
152 ml of Trizol. One ml of the homogenate was placed into a centrifuge tube containing 200
153 µl of chloroform. The tube was vortexed and placed at 25°C for three min. Following
154 centrifugation for 15 min at 12000 RPM, the supernatant was placed into a new
155 centrifuge tube containing an equal volume of 70% ethanol. Purification of the RNA was
156 accomplished using the PureLink RNA Mini Kit (Life Technologies, USA). The purified
157 RNA was labeled using the One-Color Microarray-based Gene Expression Analysis
158 (Quick Amp Labeling). The labeled RNA was hybridized to the DNA microarrays using
159 the Tecan HS Pro Hybridization kit (Agilent Technologies). The zebrafish RNA was
160 hybridized to the Zebrafish (v2) Gene Expression Microarray (Design ID 019161) and
161 the mouse RNA was hybridized to the SurePrint G3 Mouse GE 8x60K Microarray
162 Design ID 028005 (Agilent Technologies) following the manufacturer's recommended
163 protocols. The microarrays were loaded with 1.65 µg of labeled cRNA for each
164 postmortem time and sample.

165 **Calibration of the DNA microarray.** Oligonucleotide probes were calibrated by
166 hybridizing pooled serial dilutions of all samples for the zebrafish and the mouse. The
167 dilution series for the Zebrafish array was created using the following concentrations of
168 labeled cRNA: 0.41, 0.83, 1.66, 1.66, 1.66, 3.29, 6.60, and 8.26 µg. The dilution series
169 for the Mouse arrays was created using the following concentrations of labeled cRNA:
170 0.17, 0.33, 0.66, 1.32, 2.64, 5.28, 7.92, and 10.40 µg. The behavior of each probe was
171 determined from these pooled dilutions as described in the previous studies [10,11]. The
172 equations of the calibrated probes were assembled into a dataset so that they could be
173 used to back-calculate gene abundances of unknown samples (Supporting Information
174 Files S1 and S2 in Ref. 13).

175 **Statistical analyses.** Gene transcription profiles were constructed from the gene
176 abundance data determined from the 74,179 calibrated profiles. Expression levels were
177 log-transformed for analysis to stabilize the variance. A one-sided Dunnett's T-statistic
178 was applied to test for upregulation at one or more postmortem times compared to live
179 control (fish) or time 0 (mouse). A bootstrap procedure with 10^9 simulations was used to
180 determine the critical value for the Dunnett statistics in order to accommodate departures
181 from parametric assumptions and to account for multiplicity of testing. The profiles for
182 each gene were centered by subtracting the mean values at each postmortem time point to
183 create "null" profiles. Bootstrap samples of the null profiles were generated to determine
184 the 95th percentile of the maximum (over all genes) of the Dunnett statistics. Significant
185 postmortem upregulated genes were selected as those having Dunnett T values larger
186 than the 95th percentile. Only significantly upregulated genes were retained for further
187 analyses. The significantly upregulated transcriptional profiles are found in the
188 Supporting Information - Compressed/ZIP File Archive. The archive contains 3 files:
189 zebrafish_calib_probe_abundance.txt, mice_liver_probe_log10_abundance.txt, and
190 mice_brain_probe_log10_abundance.txt. Each file has the following four columns:
191 Agilent Probe Identification Tag, sample time, sample number and log10 concentration.

192 The software for calculating the numerical solution of the over- and perfectly-defined
193 linear regressions was coded in C++ and has been used in previous studies [14,15]. The
194 C++ code allowed us to train and test thousands to millions of regression models. A
195 description of the analytical approach can be found in the original publication [15].
196 Briefly, the abundances of each gene transcript in a gene set was numerically solved in
197 terms of predicting the postmortem times with modeling parameters (i.e. coefficients). A
198 version of the C++ source code is available at <http://peteranoble/software> under the
199 heading: "Determine the coefficients of an equation using matrix algebra". The web page
200 includes a Readme and example files to help users implement the code. To aid readers in
201 understanding the linear matrix algebra used in the study, we have provided a primer in
202 the Supplemental Information section. The postmortem time was predicted from the sum
203 of the product of the gene abundances multiplied by the coefficients for each gene
204 transcript. Comparing the predicted to actual PMIs with the testing dataset was used to
205 assess the quality of the prediction (the fit (i.e., R^2) and slope.

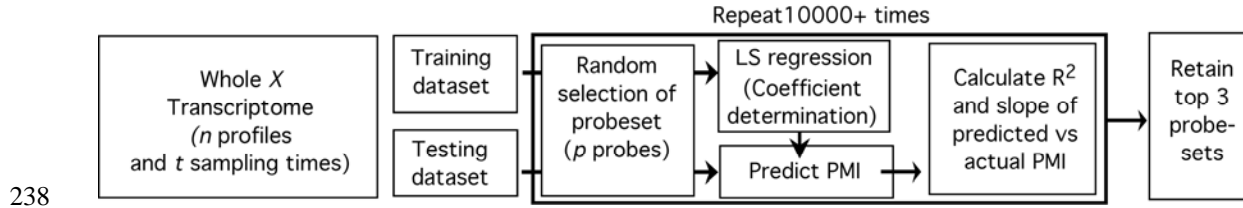
206 **Gene annotation.** The genes were annotated by performing BLASTN searches using the
207 NCBI databases. Genes that had a bit score of greater than or equal to 100 were
208 annotated.

209 **Experimental design.** Three different datasets were used in this study: the whole
210 zebrafish transcriptome, the mouse brain transcriptome, and the mouse liver
211 transcriptome. The datasets were split into training and testing data. The training data
212 was used to build the regression equations and the testing data was used to validate the
213 equations. Three different experimental designs were tested.

- 214 1. Simple linear regressions using individual genes. We examined if simple linear
215 regressions ($PMI_{\text{predict}}=m*\text{ transcript abundance} + b$) of individual gene transcripts
216 could be used to predict PMIs. The values of m and b were determined using the
217 training dataset. The performance of the regression was assessed using the R^2 of
218 the predicted versus actual PMIs with both training and testing datasets.
- 219 2. Over-defined linear regressions using top performing genes from Experimental
220 Design 1. An over-defined linear regression is used when the data consisted of
221 more rows (postmortem times) than columns (gene transcripts). The top three
222 individual gene transcripts in Experimental Design 1 were combined and trained
223 to predict PMIs using an over-defined linear regression model. The performance
224 of the model was assessed using the R^2 of the predicted versus actual PMIs of
225 both training and testing datasets.
- 226 3. Perfectly defined linear regressions using randomly selected gene transcript sets.
227 A perfectly-defined linear regression is used for data consisting of equal number
228 of rows (postmortem times) and columns (gene transcripts). A random number
229 generator was used to select sets of genes from the datasets in order to find the top
230 PMI predictors. The analysis yields a set of coefficients (i.e., m 's), one
231 coefficient for each gene transcript in a set. The coefficients were used to predict
232 the PMIs of a gene set. The R^2 and slope of the predicted versus actual PMIs
233 were determined using the training and testing data. The procedure of selecting
234 the gene transcript sets from the training set, determining the coefficients, and

235 testing the coefficients was repeated at least 50,000 or more times and the gene
236 transcript sets generating the best fit (R^2) and slopes were identified (Fig 1).

237



239 **Fig 1. Cartoon of experimental design for three different datasets. Bold box was repeated**
240 **10,000+ times. The top 3 probe datasets were determined by the R^2 between predicted**
241 **versus actual PMI and the slope closest to one using the test dataset. If X =’zebrafish’ then**
242 **$n=548, t=11, p=11$; if X =’mouse brain’ then $n=478, t=7, p=7$; if X =’mouse liver’ then $n=36,$**
243 **$t=7, p=7$.**

244

245 Results

246 The 36,811 probes of the zebrafish and 37,368 probes of the mouse were calibrated. Of
247 these, the transcriptional profiles of 548 zebrafish genes and 515 mouse genes were found
248 to be significantly upregulated. Of the 515 upregulated genes, 36 were from the liver and
249 478 genes were from the brain. It is important to note that each datum point in a
250 zebrafish transcriptional profile represents the mRNA obtained from two zebrafish and
251 each datum point in a mouse profile represents the mRNA obtained from one mouse. In
252 other words, each datum point represents a true biological replicate. Duplicate samples
253 were collected for each postmortem time for the zebrafish profiles, and triplicate samples
254 were collected for the mouse (with exception of the 48 h postmortem sample which was
255 duplicated) at each postmortem time.

256 **Predicting PMI with 1 or 3 gene transcripts.**

257 The ability of individual gene transcripts to accurately predict PMIs was assessed using
258 the simple linear regression:

$$259 \text{PMI}_{\text{predict}} = m \log_2 G + b,$$

260 with m as the slope (i.e., the coefficient), G is the individual gene transcript abundance,
261 and b is the intercept.

262 For the zebrafish, one of the duplicates (at each postmortem time) was used to determine
 263 the linear regression equation (i.e., m and b) and the other one was used to test the
 264 regression equation. For the mouse, one of the triplicates at each postmortem time was
 265 used to determine the linear equation and the remaining data (2 data points) were used to
 266 test the regression equation. The three gene transcripts of the zebrafish, mouse brain, and
 267 mouse liver with the highest fits (R^2) between predicted and actual PMIs are shown in
 268 Table 1.

269 **Table 1. Top three fits (R^2) of predicted and actual PMIs by organism/organ based on the**
 270 **training and testing datasets of individual probes (probe names were designated by Agilent)**
 271 **targeting specific transcripts. Corresponding gene names and functions are shown. Whole,**
 272 **RNA was extracted from whole organisms; Brain, RNA extracted from mouse brains;**
 273 **Liver, RNA extracted from mouse livers.**

Organism/ Organ	Oligonucleotide Probe Name	R^2	Number of data points	Gene Name and Function
Zebrafish				
Whole	A_15_P121158	0.94	11 duplicates	Non-coding
	A_15_P295031	0.84	11 duplicates	Non-coding
	A_15_P407295	0.82	11 duplicates	Non-coding
Mouse				
Brain	A_66_P130916	0.67	6 triplicates and 1 duplicate	Histocompatibility 2, O region beta locus
	A_55_P2127959	0.61	6 triplicates and 1 duplicate	Zinc finger protein 36, C3H type-like 3
	A_55_P2216536	0.60	6 triplicates and 1 duplicate	E3 ubiquitin-protein ligase
Liver	A_55_P2006861	0.94	6 triplicates and 1 duplicate	Triple functional domain protein
	A_30_P01018537	0.91	6 triplicates and 1 duplicate	Prokineticin-2 isoform 1 precursor
	A_51_P318381	0.90	6 triplicates and 1 duplicate	Placenta growth factor isoform 1 precursor

274
 275 For the zebrafish, the gene transcript targeted by probe A_15_P121158 yielded a fit
 276 (combined training and testing data) of $R^2=0.94$, while the other gene transcripts yielded
 277 moderate fits ($R^2<0.90$). The top predictors of PMIs for the mouse brain samples yielded
 278 weak R^2 -values (0.61 to 0.67), and the top predictors for the mouse liver samples yielded
 279 reasonable R^2 -values (0.90 to 0.94) (Table 1) suggesting that the liver was more suitable
 280 for predicting PMI than the brain.

281 In addition to assessing the PMI prediction of individual gene transcripts, we investigated
 282 if a combination of the top gene transcripts would improve upon PMI predictions. Using
 283 an over-defined linear regression:

284
$$PMI_{predict} = \sum_{i=1}^3 m_i \log_2 G_i$$

285 and one of the duplicate/triplicate samples from each postmortem time as the training
286 data, we determined the coefficient for each gene transcript and tested the regression
287 equation using the remaining test data. For the zebrafish, the derived coefficients for
288 genes targeted by probes A_15_P295031, A_15_P121158, and A_15_P407295 were
289 -162.97, 22.44, and 35.10, respectively. Using the gene transcript abundances for these
290 probes at 48 h postmortem (-0.33 a.u., -0.89 a.u., and -0.25 a.u., respectively) and the
291 equation above, the predicted PMI is ~25.3 h. For the mouse brain gene transcripts
292 targeted by probes A_66_P130916, A_55_P2127959, and A_55_P2216536, the derived
293 coefficients were 3.70, -3.57, and 45.25, respectively. Using the gene transcript
294 abundances for these probes at 48 h postmortem (1.21 a.u., 0.36 a.u., and 0.80 a.u.,
295 respectively) and the equation above, the predicted PMI is ~39.2 h. For the mouse liver
296 gene transcripts, the derived coefficients targeted by probes A_51_P318381,
297 A_30_P01018537, and A_55_P2006861 were -3.75, 36.21, and -13.93, respectively.
298 Using the gene transcript abundances for these probes at 48 h postmortem (1.04 a.u., 1.65
299 a.u., and 0.48 a.u., respectively) and the equation above, the predicted PMI is ~49.3 h.
300 The fits (R^2) of the predicted versus actual PMIs for the zebrafish, the mouse brain and
301 mouse liver were 0.74, 0.64, and 0.86, respectively.

302

303 While some of the individual gene transcript abundances yielded reasonable PMI
304 predictions using simple linear equations (Table 1), combining the individual gene
305 transcript abundances and using an over-defined linear regression did not significantly
306 improve upon PMI predictions based on individual genes.

307

308 These experiments showed that neither simple linear regression equations derived from
309 the individual gene transcripts, nor over-defined linear regressions derived from the top
310 three individual gene transcripts satisfactorily predicted PMIs.

311 ***Predicting PMI with many genes***

312

313 To predict PMIs using perfectly defined linear regressions, the number of gene transcripts
314 used for the regression has to equal the number of postmortem sampling times. The
315 zebrafish was sampled 11 times and the mouse was sampled 7 times, therefore 11 and 7

316 genes could be used for the regressions, respectively. The regression equation for the
317 zebrafish was:

318

$$319 \quad PMI_{predict} = \sum_{i=1}^{11} m_i \log_2 G_i$$

320

321 The regression equation for the mouse was:

322

$$323 \quad PMI_{predict} = \sum_{i=1}^7 m_i \log_2 G_i$$

324

325 The procedure to find gene transcript sets that provide the best PMI predictions included:
326 assigning randomly-selected genes to gene transcript sets, determining the coefficients of
327 the gene transcripts in the set using a defined least squared linear regression, and
328 validating the regression model with gene transcript sets in the test data. We rationalized
329 that if this process was repeated thousands to millions of times, groups of gene transcripts
330 could be identified that accurately predict PMIs with high $R^2 > 0.95$ and slopes of 0.95 to
331 1.05.

332

333 The number of upregulated genes in the zebrafish, mouse brain, and mouse liver
334 transcriptome datasets is relevant to determining the optimal gene transcript set for PMI
335 predictions because of the magnitude of possible combinations to be explored. For
336 example, there are 2.85×10^{22} combinations of 11 gene transcripts for the zebrafish
337 dataset ($n=548$), 1.08×10^{15} combinations of 7 gene transcripts for the mouse brain
338 dataset ($n=478$) and 8.35×10^6 possible combinations of 7 gene transcripts for the mouse
339 liver dataset ($n=36$). Therefore, for some transcriptome datasets (i.e., zebrafish and
340 mouse brain), the determination of the best gene transcript set to accurately predict PMI
341 was constrained by the number of possible combinations explored in reasonable
342 computer time.

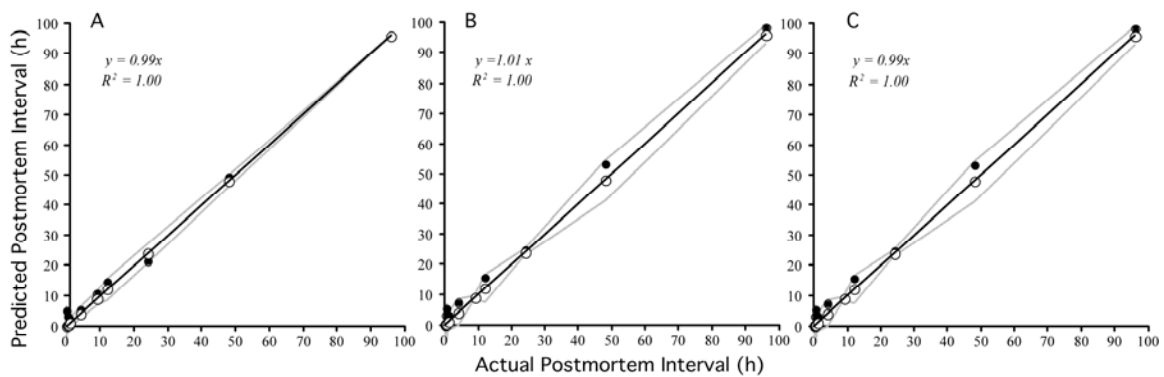
343 ***Validation of PMI prediction***

344 After training 50,000 random selections, about 95% ($n=47,582$) of the selected gene
345 transcript sets yielded R^2 and slopes of 1 with the training datasets. The remaining

346 selections ($n=2,418$) did not yield R^2 and/or slopes of 1 because the equations could not
347 be resolved, or else the fits and/or slopes were <1 . The R^2 and slopes of the predicted
348 versus actual PMIs using the testing data were used to identify the top gene sets.

349 The top three gene transcript sets with the highest R^2 and slopes closest to one are shown
350 in Fig 2. The gene transcript set used in Panel A had smaller confidence intervals than
351 those found in Panels B and C. At the 99% confidence level, the predicted PMIs for the
352 gene set in Panel A ranged from 7 to 11 h for the actual PMI of 9 h, from 8 to 16 h for the
353 actual PMI of 12 h, from 21 to 27 h for the actual PMI of 24 h, from 46 to 50 h for the
354 actual PMI of 48 h, from 96 h for the actual PMI of 96 h. These results suggest that PMIs
355 could be accurately predicted using zebrafish gene sets and the derived coefficients.

356

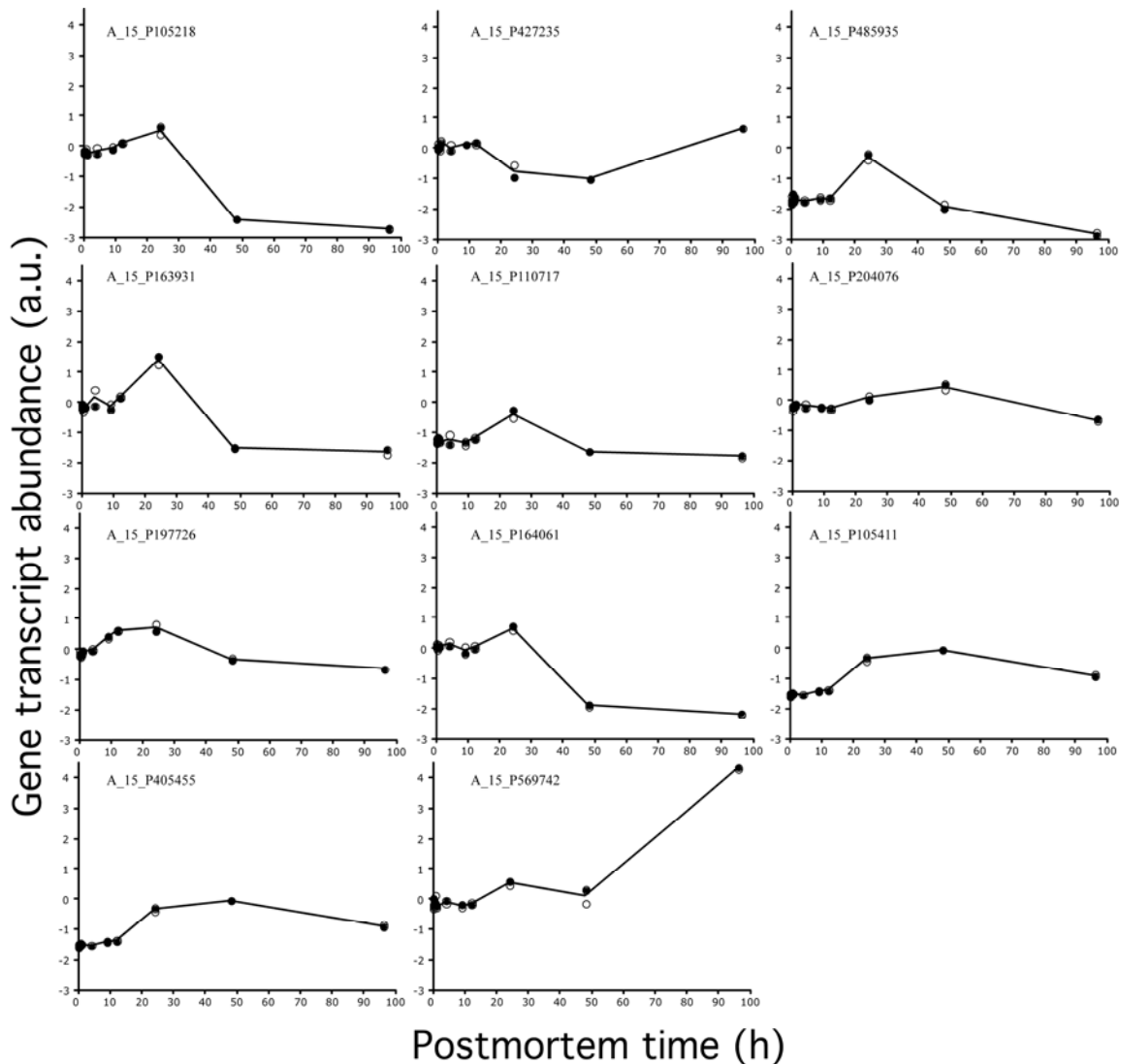


357

358 **Fig 2. Predicted versus actual PMIs determined for the zebrafish by three equations**
359 **representing different gene transcript sets. R^2 and slopes are based on both training and**
360 **testing datasets. Gray line represents 99% confidence limits of the linear regression. Open**
361 **circles, training data; closed circles, testing data. See Table 1 for information on the genes**
362 **and annotations.**

363

364 The gene transcription profiles of the 11 genes used in Fig 2, Panel A are shown in Fig 3.
365 Note that the gene transcript abundances of the duplicate samples used for training and
366 testing are similar at all sampling times. These results show the high precision of the
367 Gene Meter approach since each datum point represents different zebrafish. Note that
368 each gene has a different postmortem transcriptional profile.



369
370
371
372
373

Fig 3. Zebrafish transcriptional profiles contributing to Fig 2, Panel A. Open circles, training data; closed circles, testing data; black line, average. See Table 2 for information on the equations and probe annotations.

374 Table 2 shows the probe labels for the gene transcript sets used in Fig 2 Panels A to C
375 and their corresponding coefficients derived from the training dataset. Note that only
376 some genes could be annotated using NCBI. We assumed that genes not annotated
377 represent non-coding mRNA. The PMIs in Panel A to C could be predicted by adding
378 the products of the \log_2 abundance of each gene to its corresponding coefficient. For
379 example, the equation for Table 2 Panel A is:

380
$$\text{PMI} = -13.39P_1 - 5.72P_2 + 5.72P_3 + 12.82P_4 - 7.07P_5 - 10.26P_6 + 13.13P_7 - 16.00P_8 +$$

381
$$5.44P_9 - 4.39P_{10} + 11.94P_{11};$$

382 where P_i are the gene abundances represented by the probes A_15_P105218 (0.39 a.u.),
 383 A_15_P427235 (-0.54 a.u.), A_15_P485935 (-0.39 a.u.), A_15_P163931 (1.27 a.u.),
 384 A_15_P110717 (-0.53 a.u.), A_15_P204076 (0.16 a.u.), A_15_P197726 (0.82 a.u.),
 385 A_15_P164061 (0.58 a.u.), A_15_P105411 (-0.40 a.u.), A_15_P405455 (-1.13 a.u.),
 386 A_15_P569742 (0.46 a.u.). In this example, the predicted PMI is ~24 h. Based on
 387 Figure 2 panel A, the 99% confidence interval is between 20.9 and 27.1 h.

388

389 **Table 2. Zebrafish genes used to predict PMI by Panel. The gene annotations of the probes were**
 390 **determined using NCBI with a 100 bit minimum.**

Panel	Probe Label	Coefficient	Gene	Gene Name
A	A_15_P105218	-13.39	<i>Gpr98</i>	G-protein coupled receptor 98 precursor
	A_15_P427235	-5.72		Noncoding
	A_15_P485935	5.72		Noncoding
	A_15_P163931	12.82	<i>Moxd1</i>	DBH-like monooxygenase protein 1 homolog precursor
	A_15_P110717	-7.07	<i>Svep1</i>	Sushi von Willebrand factor type A, EGF and pentraxin
	A_15_P204076	-10.26	<i>Pde4b</i>	5'-cyclic-AMP and -GMP phosphodiesterase 11
	A_15_P197726	13.13	<i>Plek2</i>	Pleckstrin-2
	A_15_P164061	-16.00	<i>Rassf6</i>	Ras association domain-containing protein 6
	A_15_P105411	5.44	<i>Grm7</i>	Metabotropic glutamate receptor 7-like
	A_15_P405455	-4.39		Noncoding
	A_15_P569742	11.94	<i>Trim25</i>	E3 ubiquitin/ISG15 ligase TRIM25-like
B	A_15_P104895	19.66		Noncoding
	A_15_P522677	-26.62		Noncoding
	A_15_P105411	13.32	<i>Grm7</i>	Metabotropic glutamate receptor 7-like
	A_15_P119193	15.64		Noncoding
	A_15_P401770	5.89	<i>Lrrc59</i>	Leucine-rich repeat-containing protein 1
	A_15_P104490	-27.55	<i>Wif1</i>	Wnt inhibitory factor 1 precursor
	A_15_P177366	-3.88	<i>Bmpr2</i>	Bone morphogenetic protein receptor, type II a
	A_15_P586597	-7.64		Noncoding
	A_15_P168556	6.77	<i>Sema6d</i>	Semaphorin-6D isoform X1
	A_15_P105618	17.47	<i>Gpr143</i>	G-protein coupled receptor 143
	A_15_P171831	6.86	<i>Il20</i>	Interleukin-20 isoform X1
C	A_15_P569842	3.20	<i>Myo3a</i>	Myosin-IIIa
	A_15_P176341	2.03	<i>Prrt4</i>	Proline-rich transmembrane protein 4
	A_15_P107601	8.59	<i>Atf3</i>	Cyclic AMP-dependent transcription factor ATF-3
	A_15_P168526	-5.03	<i>Pglyrp1</i>	Peptidoglycan recognition protein 1
	A_15_P328806	-5.37	<i>Kdm5b</i>	Lysine-specific demethylase 5B
	A_15_P309786	2.33	<i>FimC</i>	Integumentary mucin C.1-like
	A_15_P120901	0.31	<i>Gnai1</i>	Guanine nucleotide-binding protein G(i) subunit alpha-1

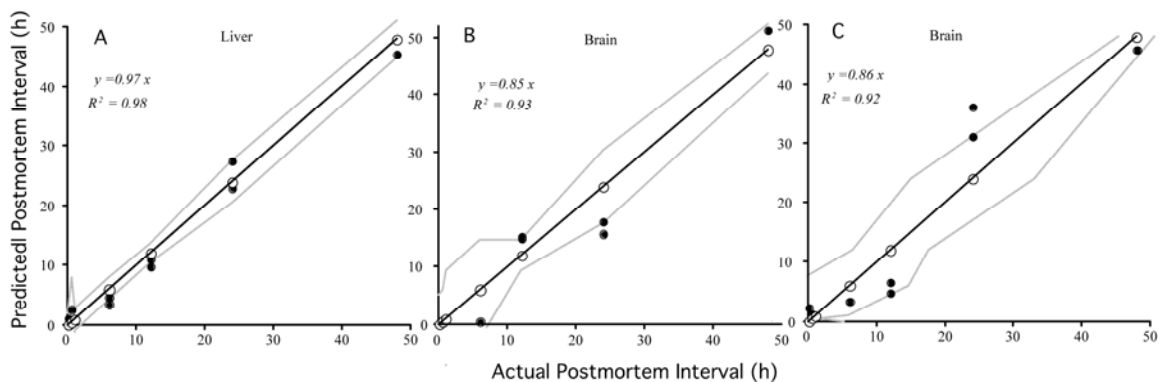
A_15_P165836	6.98	<i>C3ar1</i>	C3a anaphylatoxin chemotactic receptor-like
A_15_P576307	1.91		Noncoding
A_15_P110820	-34.27	<i>Gucy1a3</i>	Guanylate cyclase soluble subunit alpha-3
A_15_P105618	32.01	<i>Gpr143</i>	G-protein coupled receptor 143

391

392 Mouse

393 After training 50,000 random selections (each selection consisted of 7 genes), about 96%
 394 ($n=47,847$) of the selected gene sets yielded R^2 and slopes of 1. The remaining selections
 395 were not used for validation because the equations ($n=2,153$ selections) could not be
 396 resolved, or they had fits and/or slopes that were <1 ($n=25$ selections). The R^2 and slopes
 397 of predicted versus actual PMIs determined using the testing dataset identified the top
 398 performing gene sets.

399 The top selected gene transcript sets for the mouse liver and brain are shown in Fig 4. As
 400 indicated by the R^2 , slopes, and size of the 99% confidence intervals, gene transcript sets
 401 from the liver were better at predicting PMIs than those from the brain. The mouse genes
 402 used in the gene transcript sets, their coefficients, and annotations are shown in Table 3
 403 and the transcriptional gene profiles for the mouse liver samples are shown in Fig 5.
 404 Note that the high similarity in the gene transcript abundance between the data used for
 405 training and testing of the selected genes. In most cases (but not all), the duplicate
 406 samples (represented by dots) are located on top of one another.



407

408 **Fig 4. Predicted versus actual PMI determined for the mouse for three different equations**
 409 **as represented by the panels. R^2 and slopes are based on both training and testing datasets.**
 410 **Gray line represents 99% confidence limits of the linear regression. Open circles, training**
 411 **data; closed circles, testing data. See Table 3 for information on the equations and probes.**
 412

413 The poor predictability of the brain gene transcript sets (i.e., $R^2 < 0.95$) could be attributed
 414 to the low number of repeated selections of gene transcript sets and the variability in gene

415 abundances between the training and testing datasets. We repeated the analysis of the
416 brain samples an additional 1,000,000 times, which resulted in some improvement. The
417 best fit and slope for 50,000 gene transcript set selections was $R^2=0.83$ and $m=0.77$ (not
418 shown). The best fit and slope for 1,000,000 selections was $R^2=0.93$ and $m=0.85$ (Fig 4,
419 Panel B) with the second best being $R^2=0.92$ and $m=0.86$ (Fig 4, Panel C). Hence, the
420 number of combination of gene transcript sets examined is important for selecting the
421 best ones. It is important to emphasize that the computation time for running 1,000,000
422 selections was approximately 1 week using a Mac OS X 10.8.6.

423 The PMIs in Panel A to C could be predicted by adding the products of the \log_2
424 abundance of each gene to its corresponding coefficient. The predicted PMIs for mouse is
425 calculated same way as for zebrafish (shown above).

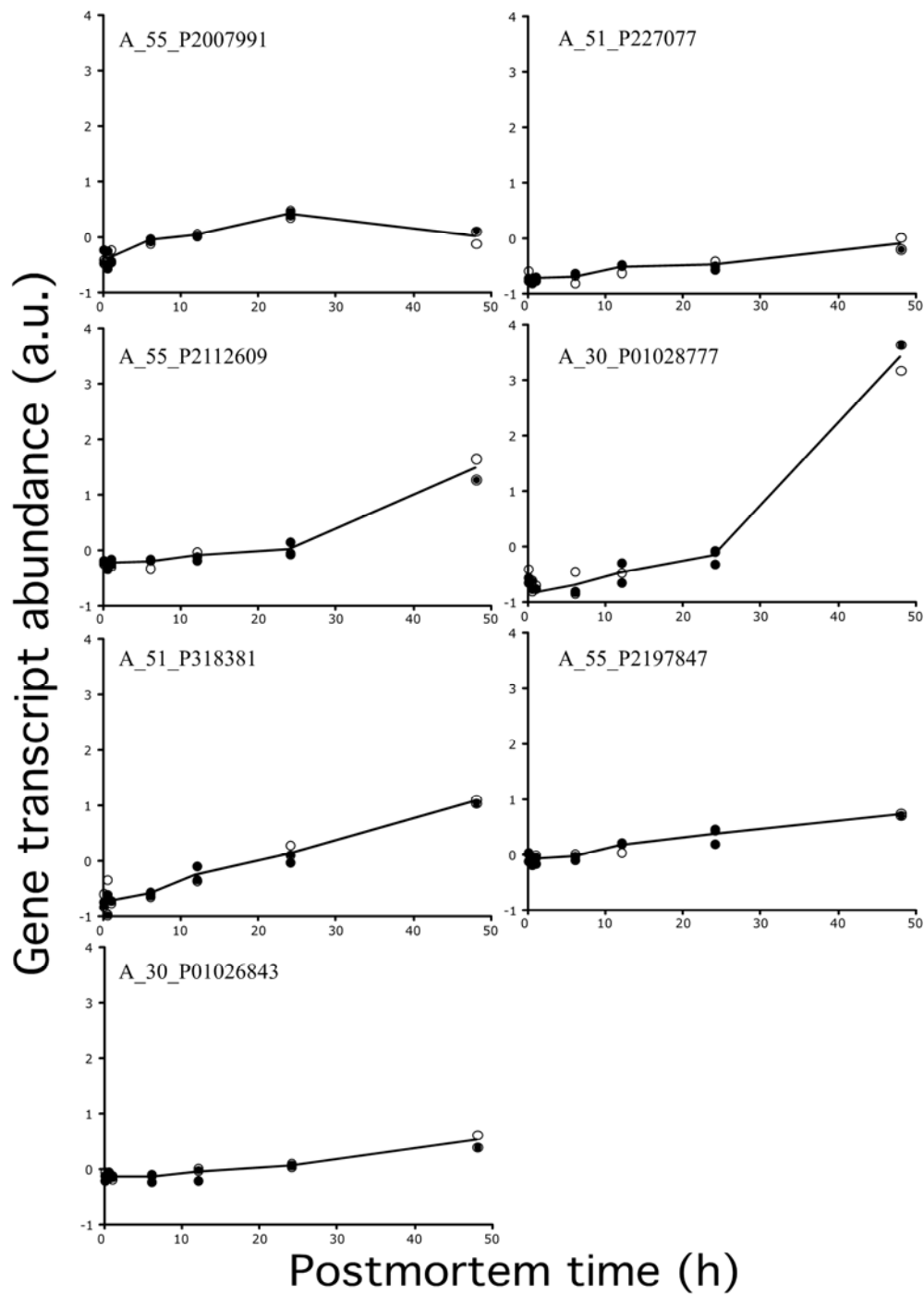
426

427 **Table 3. Mouse probes used to target gene transcripts and their coefficients used to predict**
 428 **PMIs by Panel. The annotations of probes were determined by using NCBI database with**
 429 **100 bit minimum.**

Panel	Organ	Probe	Coefficient	Gene	Gene Name
A	Liver	A_55_P2007991	12.7	<i>Tuba3b</i>	Tubulin, alpha 3B
		A_51_P227077	-21.43	<i>Mdh1b</i>	Malate dehydrogenase 1B, NAD (soluble)
		A_55_P2112609	14.97		Non-coding
		A_30_P01028777	-1.6		Non-coding
		A_51_P318381	9.91	<i>Pgf</i>	Placental growth factor
		A_55_P2197847	18.41		Non-coding
		A_30_P01026843	7.59	<i>Ifitm2</i>	Interferon induced transmembrane protein 2
B	Brain	A_52_P627085	14.2729	<i>Mrps18c</i>	28S ribosomal protein S18c, mitochondrial
		A_30_P01025266	-59.3569	<i>Klf14</i>	Krueppel-like factor 14
		A_55_P1955891	-5.57821	<i>Ppm1e</i>	Protein phosphatase 1E
		A_55_P2109107	1.92003	<i>Gfra2</i>	GDNF family receptor alpha-2 isoform 3 precursor
		A_30_P01031213	-61.0776	<i>Acs1l</i>	Long-chain-fatty-acid--CoA ligase 4 isoform 1
		A_52_P418795	-11.3343	<i>Grk4</i>	G protein-coupled receptor kinase 4
		A_66_P100268	59.9399		Non-coding
C	Brain	A_30_P01028032	9.30294	<i>Flo11l</i>	Flocculation protein FLO11-like
		A_55_P1972018	-37.7273	<i>Hist1h4a</i>	Histone H4
		A_55_P2410304	-10.5218		Non-coding
		A_52_P1082736	15.8909	<i>Sept1</i>	Septin-1
		A_66_P117204	9.66267	<i>Gpc3</i>	Glypican-3 isoform
		A_30_P01020727	6.36972	<i>Adam2</i>	Disintegrin and metalloprotease domain 4b precursor
		A_52_P236705	5.9439	<i>Ripply3</i>	Protein ripply3

430
431

432 We compared the variability in gene transcript abundances between training and testing
 433 data sets for the mouse liver and mouse brain. Transcriptional gene profiles of the gene
 434 sets used in Fig 4 Panels A and B are shown in Figs 5 and 6, respectively. While most of
 435 the mouse liver gene transcript abundances are similar for the training and testing data
 436 sets in Fig 5, many of the mouse brain gene transcript abundances are not similar in Fig
 437 6. A two-tailed T-test of the standard deviations of the gene transcript abundances in the
 438 training and testing datasets for the liver and brain samples (Fig 5 versus Fig 6) by
 439 postmortem time were significantly different ($P < 0.006$), with higher standard deviations
 440 in the brain samples than the liver. This finding indicates that variability in the gene
 441 transcript abundances affects PMI predictability.



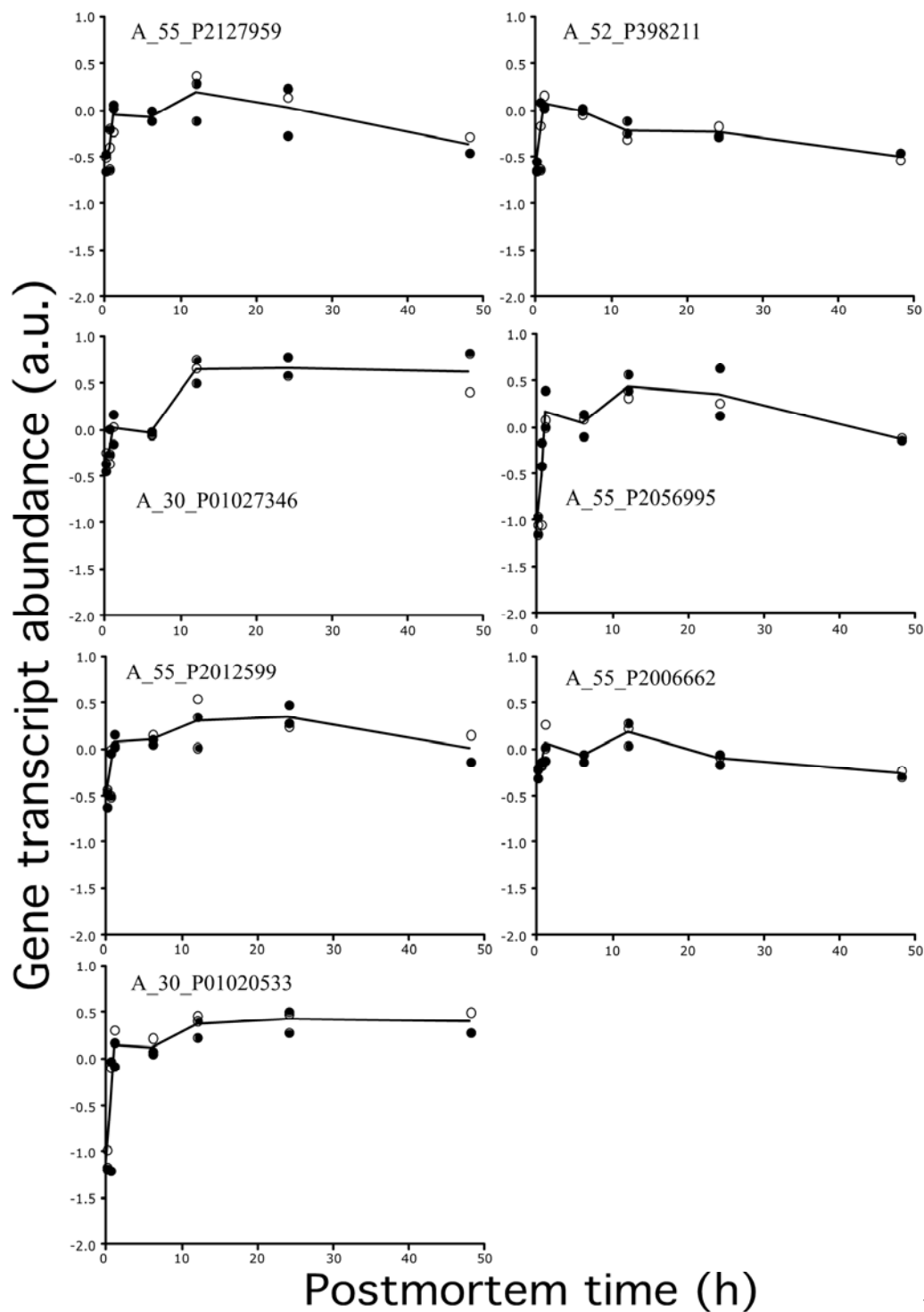
442

443

444

445

Fig 5. Mouse liver transcriptional profiles contributing to Fig 4, Panel A. Open circles, training data; closed circles, testing data; black line, average. See Table 3 for information on the equations and probe annotations.



446
447
448
449
450

Fig 6.
Mouse brain transcriptional profiles contributing to Fig 4, Panel B. Open circles, training data; closed circles, testing data; black line, average. See Table 3 for information on the equations and probe annotations.

451 To further test this phenomenon, a small amount of random noise was added to the
452 abundances of mouse liver gene transcripts (Fig 5), which originally had very low

453 standard deviations by postmortem time. When the introduction of noise approached
454 10%, the fit and slopes were drastically altered, indicating that similarity in the gene
455 transcript abundances between the training and test data sets can directly affect the fit and
456 slopes of predicted versus actual PMIs.

457 **Randomization challenge**

458 Experiments using perfectly-defined linear regressions revealed that 95% of the training
459 data for the zebrafish and mouse yielded fits (R^2) and slopes of 1 for predicted versus
460 actual PMIs. To demonstrate that the ‘perfect’ fits and slopes are functions of the linear
461 regressions, we randomized gene transcript abundances at every postmortem time for all
462 genes in the zebrafish dataset. This randomization maintained the variance in the dataset
463 at each postmortem time so that the variance of the first postmortem time in the
464 randomized dataset was the same as the variance of the gene transcript abundances at the
465 first postmortem time in the original dataset and so on for the gene transcript abundances
466 at all postmortem times.

467 As anticipated, training of the randomized zebrafish dataset using perfectly-defined linear
468 regressions (11 genes by 11 postmortem times x 50,000 repeated gene transcript
469 selections) yielded fits (R^2) and slopes of 1 for predicted versus actual PMIs for 95% of
470 the regressions. When we tested the 50,000 regressions using a test dataset, not one of the
471 gene transcript sets approached a fit (R^2) and slope of 1. In fact, most yielded slopes of
472 zero and $R^2 < 0.80$. The significance of this experiment is twofold: (i) it confirms that
473 ‘perfect’ fits and slopes using the training datasets are a function of ‘perfectly-defined’
474 linear regressions, and (ii) it confirms the need to validate the regression equations using
475 testing datasets.

476 **Discussion**

477 In addition to the different stages of body decomposition (i.e., rigor mortis, livor mortis,
478 algor mortis and putrefaction) [16,17,18], there are many biochemical, biological,
479 chemical, and physical ways to determine PMI. Biochemical indicators and
480 corresponding sample sites include: pH and spectrophotometer readings of blood and
481 serum [19], cardiac troponin-I and cadaveric blood in the heart [19,20], lactate and malate
482 dehydrogenase in the liver [21], melatonin in the brain, sera, and urine [22], DNA

483 degradation in many tissues and organs [23,24,25,26], endothelial growth factors in the
484 brain, heart, liver, and kidneys [27], insulin and glucagonin in pancreatic beta cells
485 [28,29], cells in the cerebrospinal fluid [30], apoptotic cells in skin bruises [31] and
486 histology of labial muscosa [32]. Biological indicators and sample sites include: ciliary
487 motility in the nose [33], sweat gland morphology in the arm pit [34], muscle contraction
488 [35] and pyrosequencing of the buccal cavity, rectum and GI tract samples [36],
489 entomological [37,38,39] and botanical processes occurring in and around the body
490 [40,41]. Chemical indicators and sample sites include: electrolytes in human vitreous
491 humour [42], biomarkers (e.g. amino acids, neurotransmitters) in body organs and
492 muscles [43], hypoxanthine in the vitreous humour or cerebrospinal fluid [44,45,46] and
493 potassium in the vitreous humour [47,48,49]. Physical indicators and sample sites
494 include: microwave probe to the skin [50], infrared tympanic thermography and
495 temperature of the ear [51,52], and temperature of the eye and body core [53,54,55].
496 Several formulae have also been developed to estimate PMI that are based on multiple
497 environmental and physicochemical conditions [e.g., 56]. Despite these many
498 approaches, accurate PMI prediction remains an enigma [43]. The motivation for this
499 study was to test experimental designs that could accurately predict PMI using
500 upregulated gene expression data in order to provide “proof of principle”.

501 The abundance of a gene transcript is determined by its rate of synthesis and its rate of
502 degradation [57]. In this study, the synthesis of mRNA had to far exceed its degradation
503 to be a significantly upregulated gene (at some postmortem time) in our study. As
504 demonstrated in the previous study [13] and shown in this study (Figs 3, 5, and 6), the
505 timings of the upregulation differed between genes. Some gene transcripts, such as the
506 one targeted by probe A_15_P105218, were upregulated right after organismal death and
507 reached maximize abundance at 24 h postmortem while, others, such as the one targeted
508 by probe A_15_P569742, increased substantially at 48 h postmortem (Fig 3). It is
509 presumed that differences in the transcript profiles affect the value of the coefficients in
510 the linear equations because it is not possible to generate coefficients if the gene
511 transcript abundances changed in the same way. That is, a numerical solution could not
512 be mathematically resolved.

513 It should be noted that the upregulation of postmortem genes is optimal for PMI
514 prediction because only about 1% of the total genes of an organism were upregulated in
515 organismal death – which is rare indeed. In contrast, a focus on downregulated genes
516 would not be practical because one does not know if downregulated genes are due to
517 repression, degradation of the total RNA, or exhausted resources such as those needed for
518 the transcript machinery function (e.g., dNTPs and RNA polymerase).

519 Given that gene transcripts from the liver were better at PMI predictions than those from
520 the brain suggests that mRNA transcripts from some organs are better than others. It is
521 conceivable that upregulated postmortem genes could be found in the heart, kidney,
522 spleen or muscle, which needs further exploration.

523 It is important to recognize that this study would probably not be possible using
524 conventional microarray approaches because normalizations could yield up to 20 to 30%
525 differences in the up- or down-regulation depending on the procedure selected [59-62].
526 The Gene Meter approach does not require the data to be normalized since the microarray
527 probes are calibrated. Moreover, in the processing of samples, the same amount of
528 labeled mRNA was loaded onto the DNA microarray for each sample (1.65 μ g), which
529 eliminates the need to divide the microarray output data by a denominator in order to
530 compare samples.

531 We recognize that our experimental design did not consider factors such as temperature,
532 which have been considered in other models [e.g., 4]. To do so would go beyond the
533 stated objectives of providing a “proof of principle” for the optimal experimental design
534 (i.e., perfectly-defined linear regressions based on multiple gene transcripts) using a high
535 throughput approach. Nonetheless, future studies could make our experimental design
536 more universal by integrating temperature and other factors into the regression models.

537 In addition to providing “proof of principle” of a new forensic tool for determining PMI,
538 the approach could be used as a tool for prospective studies aimed at improving organ
539 quality of human transplants.

540 **Conclusion**

541 We examined if significantly upregulated genes could be used to predict PMIs in two
542 model organisms using linear regression analyses. While PMIs could be accurately
543 predicted using selected zebrafish and mouse liver gene transcripts, predictions were poor
544 using selected mouse brain gene transcripts, presumably due to high variability of the
545 biological replicates. The experimental design of selecting groups of gene transcripts,
546 extracting the coefficients with linear regression analyses, and testing the regression
547 equations with testing data, yielded highly accurate PMI predictions. This study warrants
548 the implementation of our experimental design towards the development of an accurate
549 PMI prediction tool for cadavers and possibly a new tool for prospective studies aimed at
550 improving organ quality of human transplants.

551 **Authors' contributions**

552 PAN and AEP designed the study. MCH and PAN conducted the statistical analyses and
553 wrote the manuscript. All authors read and approved the final manuscript.

554 **Acknowledgements**

555

556 **Financial disclosure**

557 The work was supported by funds from the Max-Planck-Society.

558 **Supplemental Information**

559 ***Primer on Matrix Algebra***

560 Our PMI prediction approach using a perfectly defined system relies on matrix linear
561 algebra. The following is an explanation of matrix linear algebra and how it was used in
562 our study to predict PMI with gene expression profiles. A matrix is defined as a
563 rectangular array of related values. These values, which are called elements, usually are
564 scalars. Scalars are numbers that represent physical quantities. Elements in a horizontal
565 line are called rows and elements in a vertical line are columns and the number of rows
566 and columns describe a matrix. A matrix, \mathbf{S} , with y number of rows and z number of
567 columns is denoted as a $y \times z$ matrix. This matrix can also be notated with subscripts and
568 appears as $\mathbf{S}_{y,z}$. Mathematical operations can be performed using matrices, including

569 multiplication. Matrices can be multiplied by one another if one matrix has as many
570 columns as the other matrix has rows. We used linear equations to obtain a matrix
571 product. The number of gene transcripts used in a selected gene transcript set was limited
572 by the number of postmortem sample times.

573 **Example:**

574 If we arrange individual genes with their transcript abundances at specific postmortem
575 times in columns then, essentially, we have a matrix **A** where the columns are the
576 transcriptional profiles for individual genes. Furthermore, we can construct another
577 matrix, **B**, which defines the data in a different way. In this matrix the values are the
578 actual postmortem sampling times that are ordered in the same way as the abundances of
579 individual gene transcripts. Finally, we can define another matrix, consisting of one
580 column of coefficient values. These are the weighing factors that we will determine
581 using the rules of linear matrix algebra. To deconvolute the transcriptional profiles from
582 the mixtures and solve for the weighing factors, we set up the matrices like so: $\mathbf{A} \times \mathbf{C} =$
583 **B**. When this is done we are left with several equations. To solve for x and y we need to
584 follow the rules of linear matrix algebra. First, we must transpose **A**, which becomes \mathbf{A}^T ,
585 then multiply both sides of the equation $\mathbf{A} \times \mathbf{B} = \mathbf{C}$ with \mathbf{A}^T . Next, we are required to
586 invert the matrix product of $\mathbf{A}^T \times \mathbf{A}$, thus **A** becomes \mathbf{A}^{-1} and also multiply both sides of
587 our original equation $\mathbf{A} \times \mathbf{C} = \mathbf{B}$. Our modified equation, which looks like this $(\mathbf{A}^T * \mathbf{A})^{-1}$
588 $* \mathbf{A}^T * \mathbf{A} * \mathbf{C} = (\mathbf{A}^T * \mathbf{A})^{-1} * \mathbf{A}^T * \mathbf{B}$, is now ready to be solved for x and y in matrix **C**.
589 Solving for **C**, we get $\mathbf{C} = (\mathbf{A}^T * \mathbf{A})^{-1} * \mathbf{A}^T * \mathbf{B}$. Now the values in the matrix **C** are x and
590 y . Next, we plug our values for x and y into our matrix equations to obtain predicted
591 (calculated) PMI values. To ascertain whether the predicted PMIs from our group of
592 gene transcripts is accurate, we plotted the actual and predicted PMIs and determined the
593 slope and fit of the regression line. The R^2 and the slope of the line was observed to
594 determine how well a group of gene transcripts predicted PMI. The R^2 is the measure of
595 how much variability is accounted for by the model. For example, if the R^2 is 0.95, then
596 the model accounts for 95% of the variability. The other 5% is due to undetermined
597 phenomena.

598

599
600

Literature Cited

- 601 1. Anderson, B. Five notorious, homicidal tales of life insurance fraud Real-life clients
602 from hell. LifeHealthPro. Published October 25, 2013
603 <http://www.lifehealthpro.com/2013/10/25/5-notorious-homicidal-theses-of-life->
604 [insurance-frau](http://www.lifehealthpro.com/2013/10/25/5-notorious-homicidal-theses-of-life-)
- 605 2. Ferreira M. T. and Cunha E. Can we infer post mortem interval on the basis of
606 decomposition rate? A case from a Portuguese cemetery, *Forensic Science*
607 *International* 2013;298: e1-298.e6.
- 608 3. Sampaio-Silva F, Magalhães T, Carvalho F, Dinis-Oliveira RJ, Silvestre R. Profiling
609 of RNA degradation for estimation of post mortem interval. *PLoS One*. 2013; 8:
610 e56507.
- 611 4. Ma J, Pan H, Zeng Y, Lv Y, Zhang H, Xue A, Jiang J, Ma K, Chen L. Exploration of
612 the R code-based mathematical model for PMI estimation using profiling of RNA
613 degradation in rat brain tissue at different temperatures. *Forensic Sci Med Pathol*.
614 2015;11: 530-7.
- 615 5. Lv YH, Ma KJ, Zhang H, He M, Zhang P, Shen YW, Jiang N, Ma D, Chen L. A time
616 course study demonstrating mRNA, microRNA, 18S rRNA, and U6 snRNA changes
617 to estimate PMI in deceased rat's spleen. *J Forensic Sci*. 2014, 59: 1286-94.
- 618 6. González-Herrera L, Valenzuela A, Marchal JA, Lorente JA, Villanueva E. Studies
619 on RNA integrity and gene expression in human myocardial tissue, pericardial fluid
620 and blood, and its postmortem stability. *Forensic Sci Int*. 2013; 232: 218-28.
- 621 7. Zapico SC, Menéndez ST, Núñez P. Cell death proteins as markers of early
622 postmortem interval. *Cell Mol Life Sci*. 2014;71: 2957-62.
- 623 8. Birdsill AC, Walker DG, Lue L, Sue LI, Beach TG. Postmortem interval effect on
624 RNA and gene expression in human brain tissue. *Cell Tissue Bank*. 2011;12: 311-8.
- 625 9. Tu Y, Stolovitzky G, Klein U. Quantitative noise analysis for gene expression
626 microarray experiments. *Proc Natl Acad Sci USA* 2002;99: 14031.
- 627 10. Pozhitkov AE, Noble PA, Bryk J, Tautz D. A revised design for microarray
628 experiments to account for experimental noise and uncertainty of probe response.
629 *PLoS One*. 2014;9: e91295.
- 630 11. Harrison A., Binder, H. Buhot, A., Burden, C.J., Carlon, E., Gibas, C., Gamble, L.J.
631 Halperin, A., Hooyberghs, J., Kreil, D.P., Levicky, R., Noble, P.A., Ott, A., Pettitt,
632 B.M., Tautz, D., Pozhitkov, AE. Physico-chemical foundations underpinning
633 microarray and next-generation sequencing experiments. *Nucl Acids Res*. 2013; 41:
634 2779-96.
- 635 12. Amend AS, Seifert KA, Bruns TD. Quantifying microbial communities with 454
636 pyrosequencing: does read abundance count? *Mol Ecol* 2010;19: 5555.
- 637 13. Pozhitkov AE, Neme R, Domazet-Loo T, Leroux BG., Soni S, Tautz D and Noble
638 PA. Thanatotranscriptome: genes actively expressed after organismal death. *PlosOne*
639 2016; (PONE-6-02773).
- 640 14. Pozhitkov AE, Nies G, Kleinhenz B, Tautz D, Noble PA. Simultaneous quantification
641 of multiple nucleic acid targets in complex rRNA mixtures using high density

- 642 microarrays and nonspecific hybridization as a source of information. *J Microbiol*
643 *Methods*. 2008;75: 92-102. doi: 10.1016/j.mimet.2008.05.013.
- 644 15. Pozhitkov AE, Bailey KD, Noble PA. Development of a statistically robust
645 quantification method for microorganisms in mixtures using oligonucleotide
646 microarrays. *J Microbiol Methods*. 2007;70: 292-300.
- 647 16. Council of Europe, Strasbourg, 1999. Recommendation No. R (99) 3 on the
648 Harmonization of Medico-Legal Autopsy Rules and Its Explanatory Memorandum.
649 Strasbourg: Council of Europe; 1999.
- 650 17. Kolusayın Ö, Koç S. Ölüm, içinde Soysal Z, Çakalır C. *Forensic Medicine, Print I*,
651 *Istanbul University Cerrahpaşa School of Medicine's Publication, Istanbul*; 1999; 93-
652 152.
- 653 18. Nashelsky MB, Lawrence CH. Accuracy of cause of death determination without
654 forensic autopsy examination. *Am J Forensic Med Pathol*. 2003;24: 313-9.
- 655 19. Donaldson AE, Lamont IL. Biochemistry changes that occur after death: potential
656 markers for determining post-mortem interval. *PLoS One*. 2013;8: e82011.
- 657 20. Sabucedo AJ, Furton KG. Estimation of postmortem interval using the protein marker
658 cardiac Troponin I. *Forensic Sci Int*. 2003;134: 11–16.
- 659 21. Gos T, Raszeja S. Postmortem activity of lactate and malate dehydrogenase in human
660 liver in relation to time after death. *Int J Legal Med*. 1993;106: 25–29
- 661 22. Mikamimi H, Terazawa K, Takatori T, Tokudome S, Tsukamoto T, Haga K.
662 Estimation of time of death by quantification of melatonin in corpses. *Int J Legal*
663 *Med*. 1994;107: 42–51.
- 664 23. Di Nunno N, Costantinides F, Cina SJ, Rizzardi C, Di Nunno C, Melato M. What is
665 the best sample for determining the early postmortem period by on-the-spot flow
666 cytometry analysis? *Am J Forensic Med Pathol*. 2002, 23:173-80.
- 667 24. Di Nunno N, Costantinides F, Melato M. Determination of the time of death in a
668 homicide-suicide case using flow cytometry. *Am J Forensic Med Pathol*. 1999;20:
669 228-31.
- 670 25. Di Nunno N, Costantinides F, Bernasconi P, Bottin C, Melato M. Is flow cytometric
671 evaluation of DNA degradation a reliable method to investigate the early postmortem
672 period? *Am J Forensic Med Pathol*. 1998;19:50-3.
- 673 26. Cina SJ. Flow cytometric evaluation of DNA degradation: a predictor of postmortem
674 interval? *Am J Forensic Med Pathol*. 1994;15: 300-2.
- 675 27. Thaik-Oo M, Tanaka E, Tsuchiya T, Kominato Y, Honda K, Yamazaki K, Misawa S.
676 Estimation of postmortem interval from hypoxic inducible levels of vascular
677 endothelial growth factor. *J Forensic Sci*. 2002;47: 186-9.
- 678 28. Wehner F, Wehner HD, Schieffer MC, Subke J. Delimitation of time of death by
679 immunohistochemical detection of insulin in pancreatic beta-cells. *Forensic Sci Int*.
680 1999; 105: 161-9.
- 681 29. Wehner F, Wehner HD, Subke J. Delimitation of the time of death by
682 immunohistochemical detection of glucagon in pancreatic a-cells. *Forensic Sci Int*.
683 2002;124: 192–199.

- 684 30. Wyler D, Marty W, Bär W. Correlation between the post-mortem cell content of
685 cerebrospinal fluid and time of death. *Int J Legal Med.* 1994;106: 194–199.
- 686 31. Sawaguchi T, Jasani B, Kobayashi M, Knight B. Post-mortem analysis of apoptotic
687 changes associated with human skin bruises. *Forensic Sci Int.* 2000;108: 187-203.
- 688 32. Yadav A, Angadi PV, Hallikerimath S. Applicability of histologic post-mortem
689 changes of labial mucosa in estimation of time of death - a preliminary study. *Aust J*
690 *For Sci.* 2012;44: 343-352.
- 691 33. Romanelli MC, Gelardi M, Fiorella ML, Tattoli L, Di Vella G, Solarino B. Nasal
692 ciliary motility: a new tool in estimating the time of death. *Int J Legal Med.* 2012;
693 126: 427-33.
- 694 34. Cingolani M, Osculati A, Tombolini A, Tagliabracci A, Ghimenton C, Ferrara SD.
695 Morphology of sweat glands in determining time of death. *Int J Legal Med.*
696 1994;107: 132–140.
- 697 35. Warther S, Sehner S, Raupach T, Püschel K, Anders S. Estimation of the time since
698 death: post-mortem contractions of human skeletal muscles following mechanical
699 stimulation (idiomuscular contraction). *Int J Legal Med.* 2012;126: 399-405.
- 700 36. Hyde ER, Haarmann DP, Lynne AM, Bucheli SR, Petrosino JF. The living dead:
701 bacterial community structure of a cadaver at the onset and end of the bloat stage of
702 decomposition. *PLoS One.* 2013;8: e77733.
- 703 37. Meiklejohn KA, Wallman JF, Dowton M. DNA barcoding identifies all immature life
704 stages of a forensically important flesh fly (Diptera: Sarcophagidae). *J Forensic Sci.*
705 2013;58: 184-7.
- 706 38. Warren JA, Anderson GS. Effect of fluctuating temperatures on the development of a
707 forensically important blow fly, *Protophormia terraenovae* (Diptera: Calliphoridae).
708 *Environ Entomol.* 2013;42: 167-72.
- 709 39. Vasconcelos SD, Soares TF, Costa DL. Multiple colonization of a cadaver by insects
710 in an indoor environment: first record of *Fannia trimaculata* (Diptera: Fanniidae) and
711 *Peckia (Peckia) chrysostoma* (Sarcophagidae) as colonizers of a human corpse. *Int J*
712 *Legal Med.* 2014;128: 229-33.
- 713 40. Cardoso HF, Santos A, Dias R, Garcia C, Pinto M, Sérgio C, Magalhães T.
714 Establishing a minimum postmortem interval of human remains in an advanced state
715 of skeletonization using the growth rate of bryophytes and plant roots. *Int J Legal*
716 *Med.* 2010;124: 451-6.
- 717 41. Lancia M, Conforti F, Aleffi M, Caccianiga M, Bacci M, Rossi R. The use of
718 *Leptodictium riparium* (Hedw.) Warnst in the estimation of minimum postmortem
719 interval. *J Forensic Sci.* 2013;58 Suppl 1: S239-42.
- 720 42. Bocaz-Beneventi G, Tagliaro F, Bortolotti F, Manetto G, Havel J. Capillary zone
721 electrophoresis and artificial neural networks for estimation of the post-mortem
722 interval (PMI) using electrolytes measurements in human vitreous humour. *Int J*
723 *Legal Med.* 2002;116: 5-11.
- 724 43. Vass AA. The elusive universal post-mortem interval formula. *Forensic Sci Int.* 2011;
725 204: 34-40.

- 726 44. Madea B, Herrmann N, Henbge C. Precision of estimating the time since death by
727 vitreous potassium--comparison of two different equations. *Forensic Sci Int.* 1990;
728 46: 277-84.
- 729 45. Madea B, Käferstein H, Hermann N, Sticht G. Hypoxanthine in vitreous humor and
730 cerebrospinal fluid--a marker of postmortem interval and prolonged (vital) hypoxia?
731 Remarks also on hypoxanthine in SIDS. *Forensic Sci Int.* 1994;65: 19-31.
- 732 46. Madea B. Importance of supravitality in forensic medicine. *Forensic Sci Int.* 1994; 69:
733 221-41.
- 734 47. Lange N, Swearer S, Sturmer WQ. Human postmortem interval estimation from
735 vitreous potassium: an analysis of original data from six different studies. *Forensic*
736 *Sci Int.* 1994;66: 159-74.
- 737 48. Zhou B, Zhang L, Zhang G, Zhang X, Jiang X. The determination of potassium
738 concentration in vitreous humor by low pressure ion chromatography and its
739 application in the estimation of postmortem interval. *J Chromatogr B Analyt Technol*
740 *Biomed Life Sci.* 2007;852: 278-81.
- 741 49. Lendoiro E, Cordeiro C, Rodríguez-Calvo MS, Vieira DN, Suárez-Peñaranda JM,
742 López-Rivadulla M, Muñoz-Barús JI. Applications of Tandem Mass Spectrometry
743 (LC-MSMS) in estimating the post-mortem interval using the biochemistry of the
744 vitreous humour. *Forensic Sci Int.* 2012;223: 160-4.
- 745 50. Al-Alousi LM, Anderson RA, Land DV. A non-invasive method for postmortem
746 temperature measurements using a microwave probe. *Forensic Sci Int.* 1994;64: 35-
747 46.
- 748 51. Cattaneo C, Di Giancamillo A, Campari O, Orthmann N, Martrille L, Domeneghini
749 C, Jouineau C, Baccino E. Infrared tympanic thermography as a substitute for a probe
750 in the evaluation of ear temperature for post-mortem interval determination: a pilot
751 study. *J Forensic Leg Med.* 2009;16: 215-7.
- 752 52. Baccino E, De Saint Martin L, Schuliar Y, Guilloteau P, Le Rhun M, Morin JF,
753 Leglise D, Amice J. Outer ear temperature and time of death. *Forensic Sci Int.*
754 1996;83: 133-46.
- 755 53. Kaliszán M. Studies on time of death estimation in the early post mortem period --
756 application of a method based on eyeball temperature measurement to human bodies.
757 *Leg Med (Tokyo).* 2013;15: 278-82.
- 758 54. Nelson EL. Estimation of short-term postmortem interval utilizing core body
759 temperature: a new algorithm. *Forensic Sci Int.* 2000;109: 31-8.
- 760 55. Muggenthaler H, Sinicina I, Hubig M, Mall G. Database of post-mortem rectal
761 cooling cases under strictly controlled conditions: a useful tool in death time
762 estimation. *Int J Legal Med.* 2012;126: 79-87.
- 763 56. Megyesi MS, Nawrocki SP, Haskell NH. Using accumulated degree-days to estimate
764 the postmortem interval from decomposed human remains. *J Forensic Sci.* 2005;50:
765 618-26.
- 766 57. Neymotin, Athanasiadou R, Gresham D. Determination of *in vivo* RNA kinetics using
767 RATE-seq. *RNA.* 2014;20: 1645.

- 768 58. Bauer M, Gramlich I, Polzin S, Patzelt D. Quantification of mRNA degradation as
769 possible indicator of postmortem interval--a pilot study. *Leg Med (Tokyo)*. 2003;5:
770 220-7.
- 771 59. Barash Y, Dehan E, Krupsky M, Franklin W, Geraci M, Friedman N, Kaminski N.
772 Comparative analysis of algorithms for signal quantitation from oligonucleotide
773 microarrays. *Bioinform*. 2004, 20: 839.
- 774 60. Seo J, Bakay M, Chen YW, Hilmer S, Shneiderman B, Hoffman EP. Interactively
775 optimizing signal-to-noise ratios in expression profiling: project-specific algorithm
776 selection and detection p-value weighting in Affymetrix microarrays. *Bioinform*.
777 2004, 20: 2534.
- 778 61. Harr B, Schlötterer C, Comparison of algorithms for the analysis of Affymetrix
779 microarray data as evaluated by co-expression of genes in known operons. *Nucl*
780 *Acids Res*. 2006, 34: e8.
- 781 62. Millenaar FF, Okyere J, May ST, van Zanten M, Voeselek LA, Peeters AJ. How to
782 decide? Different methods of calculating gene expression from short oligonucleotide
783 array data will give different results. *BMC Bioinform*. 2006, 7: 137.
- 784
- 785