MOLECULAR ECOLOGY

Inferring regulatory change from gene expression: the confounding effects of tissue scaling

SCHOLARONE™
Manuscripts

1   Inferring regulatory change from gene expression: the confounding

2   effects of tissue scaling

3

4   Stephen H. Montgomery[1,2] and Judith E. Mank[1]

5

6   [1] Dept. Genetics, Evolution and Environment, University College London, London

7   WC1E 6BT, UK

8   [2] Corresponding author: Stephen.Montgomery@cantab.net

9

12

13   Running title: Tissue scaling and transcript abundance

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34   **Abstract**

35   Comparative studies of gene expression are often designed with the aim of identifying

36   regulatory changes associated with phenotypic variation. In recent years large-scale

37   transcriptome sequencing methods have increasingly been applied to non-model

38   organisms to ask important ecological or evolutionary questions. Although

39   experimental design varies, many of these studies have been based on RNA libraries

40   obtained from heterogeneous tissue samples, for example homogenised whole bodies.

41   Comparisons between groups of samples that vary in tissue composition can introduce

42   sufficient variation in RNA abundance to produce patterns of differential expression

43   that are mistakenly interpreted as evidence of regulatory differences. Here we present

44   a simple model that demonstrates this effect. The model describes the relationship

45   between transcript abundance and tissue composition in a two-tissue system, and how

46   this relationship varies under different scaling relationships. Using a range of

47   biologically realistic variables, including real biological examples, to parameterise the

48   model we highlight the potentially severe influence of tissue scaling on relative

49   transcript abundance. We use these results to identify key aspects of experimental

50   design and analysis that can help to limit the influence of tissue scaling on the

51   inference of regulatory difference from comparative studies of gene expression.

52

53

54

55

56

57

58

59

60

61

62

63

64

Introduction

A substantial amount of intra- and inter-specific diversity results from regulatory variation. Within species, a single genome can encode multiple distinct phenotypes by varying expression levels for the underlying loci. Examples of regulatory-based phenotypes include social insect castes (Toth *et al.* 2008), some instances of plastic alternative morphs such as dominant and subordinate turkeys (Pointer *et al.* 2013) or territorial, satellite and sneaker males in wrasses (Alonzo *et al.* 2000; Stiver *et al.* 2015), caring and non-caring in beetles (Parker *et al.* 2015), and a substantial proportion of differences between males and females (Moczek & Rose 2009; Khila *et al.* 2012). Similarly, across species or divergent populations, gene regulation provides an important route for the evolution of diversity (Carroll 2008; Stern & Orgogozo 2008) with many adaptive phenotypic changes linked to regulatory evolution (e.g. Shapiro *et al.* 2004; Steiner *et al.* 2007).

Given the importance of regulatory variation in shaping phenotypic diversity, transcriptome analyses based on RNA-Seq methods are increasingly used in evolutionary and ecological studies with the explicit aim of identifying genes that underlie phenotypic variation. These studies assume that differential gene expression is the result of altered transcriptional regulation which lead to phenotypic differences between groups of individuals. In many cases functional validation experiments have demonstrated causative relationships between variation in gene expression and variation in phenotypic development (e.g Abzhanov *et al.* 2006; Khila *et al.* 2012). However, functional validation is often inhibited by the polygenic nature of many traits, or a lack of functional genetics tools for the study species. For the moment at least, interpretation of the results of such studies are largely dependent on the assumption that expression differences have functional importance to the phenotypic variation observed across samples.

However, regulatory differences are not the only source of variation in gene expression in heterogeneous tissue samples. The composition of the tissue sampled for RNA extraction, and subsequent quantification of expression level, is a major source of variation that may undermine the validity of any inferred relationship between differential gene expression and phenotypic variation, but is yet to be scrutinised in any detail.

97   The design of published expression studies varies substantially. Although
98 recent studies have demonstrated the potential to study gene expression in single cells
99 (Sandberg 2014), these remain limited and most studies are based on larger samples,
100 ranging from comparisons between organs (e.g. Enard *et al.* 2002; Khaitovich *et al.*
101 2004; Ghalambor *et al.* 2007; Brawand *et al.* 2011; Chen *et al.* 2015; Harrison *et al.*
102 2015), body parts composed of many constituent tissues such as heads (e.g. Parker *et*
103 *al.* 2015; Standage *et al.* 2016), or whole body samples (e.g. Kvist *et al.* 2013;
104 Feldmeyer *et al.* 2014; Hollis *et al.* 2014; Immonen *et al.* 2014; Stuglik *et al.* 2014).
105 In all these cases, tissue samples are homogenized before mRNA extraction,
106 purification and sequencing, with the resulting expression levels forming the primary
107 data for comparison.

108   The homogenization of heterogeneous tissue samples provides one source of
109 non-regulatory variation in estimated expression levels. The composition of these
110 heterogeneous tissues depends on the nature of their constituent parts, the scaling
111 relationships between these constituent parts, and the overall size of the tissue or
112 individual. When comparing expression levels between groups of samples, for
113 example groups of biological replicates of different sexes or different phenotypic
114 morphs, the assumed connection between expression level and gene regulation is only
115 valid if we also assume subcomponents of the tissue sample scale isometrically with
116 total size, and do not vary between the groups under comparison. Numerous
117 biological examples suggest isometry between traits is not the norm (Voje 2016),
118 strongly questioning the validity of how we interpret comparative studies of gene
119 expression.

120   Under isometric scaling the relationship between two component traits is one-
121 to-one. Any individual, regardless of its total size, will have an equal percentage of its
122 mass given over to its constituent parts. Deviation from isometry means this one-to-
123 one relationship is no longer true (Figure 1, rows 1 to 3). As total size varies, an
124 allometric relationship results in the size of component parts of a tissue sample
125 varying to a greater or lesser degree and, as a result, the proportional size of each
126 tissue component can vary. For example, the effects of both scaling patterns can be
127 illustrated in fiddler crabs with asymmetric claw sizes. The smaller 'minor' claw
128 scales isometrically with body size, whereas the larger 'major' claw scales with
129 positive allometry, or hyper-allometry (Rosenberg 2002). Hence, as body size

4

130   increases the size of the minor claw as a proportion of body mass is constant, whereas

131   the size of the major claw becomes disproportionately larger.

132   When sampling heterogeneous tissue, different forms of scaling relationships

133   will affect comparative studies of gene expression in different ways. Isometry does

134   not present a problem for studies of gene expression because the proportion of the

135   RNA library attributable to a given tissue is constant (Figure 1, panels A3, A4). Any

136   robust and repeatable change in expression level is therefore likely to be attributed to

137   regulatory variation between the groups under comparison. However, under non-

138   isometric scaling this is no longer the case. If we consider the allometric equation ($y =$

139   αx^{β}), isometry assumes the scaling coefficient, β, is one (Figure 1A1, A2 and A3).

140   Under hyper-allometry, or positive allometry, β is greater than one. In this case, as

141   trait x increases in size, trait y increases in size more rapidly (Figure 1B1). As a result,

142   the size of y as a proportion of the total size increases in larger individuals (Figure

143   1B2 and B3). In contrast, under hypo-allometry, or negative allometry, β is less than

144   one and as trait x increases in size trait y increases more slowly and accounts for a

145   smaller proportion of total size in larger individuals (Figure 1C1, C2 and C3). As the

146   proportions of each sub-tissue in a sample change, expression levels of some genes in

147   RNA-Seq datasets could vary in a way that looks like regulatory variation, but is in

148   fact a sampling artefact.

149   A further confounding effect arises when groups of samples differ in their

150   scaling coefficient, β, or the scaling constant α (Figure 1D1, E1). For example,

151   variation in α results in 'grade-shifts' between groups of individuals under

152   comparison, for example the two sexes, two phenotypic morphs or two populations or

153   species (Figure 1D1). This is often observed between morphs within species, for

154   example in testis mass between male morphs (e.g. Tomkins & Simmons 2002), or

155   between species, such as in the size of testes under different reproductive ecologies

156   (Harcourt *et al.* 1981) or of different brain components (Barton & Harvey 2000;

157   Barton & Venditti 2014). Grade-shifts are also commonly observed in experimental

158   selection lines and appear to be a major axis of evolvability (e.g. Wilkinson 1993;

159   Emlen 1996; Egset *et al.* 2012; Kotrschal *et al.* 2013). Where these grade-shifts

160   occur, individuals will differ in the proportions of their constituent parts regardless of

161   total size (Figure 1D2 and D3).

162   Shifts in β are perhaps more rare in nature, possibly due to stronger

163   developmental or functional constraint (e.g. Egset *et al.* 2012), but they do occur

164   between cell or tissue types within tissues and across species (Simmons & Tomkins

165   1996; Herculano-Houzel *et al.* 2015). The main result of β differences between

166   groups is that the similarity of tissue composition between those groups will vary with

167   total size (Figure 1E1, E2 and E3). This will likely increase variance within a group as

168   well as predictably altering mean transcript abundance between groups.As a result of

169   non-isometric scaling relationships, groups of individuals - be they species, morphs,

170   castes, or sexes - can vary substantially in body or tissue composition. In the case of

171   hyper- and hypo-allometry this can occur in the absence of any functional or

172   developmental reorganization, and is a mere consequence of variation in total size.

173   The proportion, or percentage size, of different tissue components is important for

174   studies of gene expression because RNA-Seq is always a proportional rather than

175   absolute measure of expression level, regardless of sequencing depth. RNA

176   abundance within a sample is therefore directly related to the *proportion* of cells in

177   the sample expressing a gene at a certain level. As a result of this, variation among

178   samples in the proportion of different cell types will alter the proportion of mRNA

179   transcripts in the homogenized tissue pool, and therefore expression level estimates.

180   Expression levels are therefore related to variation in proportions of tissue

181   components (Figure 1, rows 3 and 4) rather than the variation around scaling

182   relationships between those tissues, i.e. 'relative' size (as indicated in Figure 1, row

183   1). As a result, comparing variation in expression level between samples of

184   homogenized, heterogeneous tissue may partly reflect differences in regulation, but

185   could also reflect differences in composition. Unfortunately, these alternatives are not

186   mutually exclusive, further complicating analysis of expression variation.

187   Differences in tissue scaling are not problematic to studies of RNA-Seq if the

188   sole aim is to simply identify expressed genes. However, if the aim is to identify loci

189   with altered regulation that underpins phenotypic variation, and then to subsequently

190   study the evolutionary characteristics of those loci, tissue scaling becomes a key

191   concern. This is perhaps more apparent in RNA-Seq analyses based on whole-body or

192   amalgamated body parts because of the obvious potential for variation in the

193   proportion of constituent tissues. However, scaling relationships between cell types

194   within organs can also deviate from isometry and can differ between groups of

195 individuals or species (e.g. Herculano-Houzel *et al.* 2015). As such, finer-scale
196 preparations may also be affected.

197 If allometric scaling contributes to large differences in gene expression, the
198 central assumption of comparative studies of gene expression, that divergence in
199 expression level reflects divergence in gene regulation, would be difficult to support.
200 However, it is not clear what magnitude of differences we might expect under
201 different scaling scenarios, or how this may vary across different expression levels.
202 Without this knowledge, it is difficult to know when a shift in gene expression is
203 more likely explained by regulatory variation than an effect of scaling, or vice versa.
204 Our goal here is to explore the ways that tissue scaling can influence RNA-Seq
205 studies using a modelling approach, and to offer some suggested guidelines that may
206 facilitate improved interpretation of RNA-Seq studies that aim to study the
207 phenotypic effects of variation in gene regulation.

208

209 **Materials and methods**

210 *A tissue-scaling model of gene expression differences*

211 To explore the effects of allometric scaling on patterns of gene expression we
212 developed a simple model. In this model, a sample is comprised of two tissues, *x* and
213 *y*, which scale with each other according to the allometric equation $y = \alpha x^{\beta}$ where β
214 is the scaling coefficient and α is the scaling constant, The total size of the sample (*S*)
215 is therefore the sum of tissue *y* and tissue *x*:

216 $$S = y + x = \alpha x^{\beta} + x \qquad \text{[eq. 1]}$$

217

218 Within each tissue, we assume the total expression level of an individual gene (*C*) is
219 constant for a given unit of size (e.g. mass or cell number). To reflect the independent
220 regulation of expression level for different genes in tissue types we allow this constant
221 to vary between tissues, and between genes. The number of transcripts for a gene in
222 tissues *x* and *y* are therefore:

223 $$Transcript\ count\ of\ gene\ a\ in\ tissue\ x = C_{a,x} \times x \qquad \text{[eq. 2]}$$
224 $$Transcript\ count\ of\ gene\ a\ in\ tissue\ y = C_{a,y} \times \alpha x^{\beta} \qquad \text{[eq. 3]}$$

225

226 In a homogenised sample, the total expression will be the sum of eq. 2 and eq 3.
227 However, with current methods, the observed value will be a proportion of the total

228 transcript count (C_{total}). This is modelled as the average expression of a gene across

229 both tissues (C_m) multiplied total sample size (S) and the number of expressed genes

230 (G):

231 $$C_{total} = C_m \times [\alpha x^\beta + x] \times G \qquad \text{[eq. 4]}$$

232

233 The relative expression of an individual gene (RE_a) will therefore equal the sum of its

234 abundance in tissues x and y (eq. 2 and eq. 3) divided by the total transcript count

235 (C_{total}; eq. 4):

236 $$RE_a = \frac{[C_{a,x} \times x] + [C_{a,y} \times \alpha x^\beta]}{[\alpha x^\beta + x] \times G \times C_m} \qquad \text{[eq. 5]}$$

237

238 RE_a is easily converted to be equivalent to commonly used measures of relative gene

239 expression such as 'counts per million' (CPM), by simple multiplication:

240 $$CPM = RE_a \times 10^6$$

241 [eq. 6]

242

243 *CPM* is used to compare the expression level of a gene between groups of samples,

244 for example between sexes, morphs, populations or species. Significant shifts in log-

245 transformed *CPM* can be identified using traditional statistics such as *t*-tests or a

246 Mann-Whitney U test. The log$_2$-fold change (*FC*) between two groups is calculated

247 as:

248 $$FC = log_2(CPM_{group\ 1}) - log_2(CPM_{group\ 2}) \qquad \text{[eq. 7]}$$

249

250 Using this model we can estimate *FC* between two samples which do not differ in the

251 expression level of gene *a* but that can vary for *x* (and therefore *y* and *S*), α or β, as

252 indicated by the subscript numbers:

253 $$FC = log_2\left(\frac{[C_{a,x} \times x_1] + [C_{a,y} \times \alpha_1 x_1^{\beta_1}]}{[\alpha_1 x_1^{\beta_1} + x_1] \times G \times C_m} \times 10^6\right) - log_2\left(\frac{[C_{a,x} \times x_2] + [C_{a,y} \times \alpha_2 x_2^{\beta_2}]}{[\alpha_2 x_2^{\beta_2} + x_2] \times G \times C_m} \times \right.$$

254 $$\left. 10^6\right)$$

255

256 This model was used to investigate the expected effect on *FC* under three scenarios: i)

257 effects of size differences under conserved allometric scaling by varying *S* between

258 two groups while α and β remain constant, ii) effects of varying the allometric

259 constant (α) between two groups while *S* and β remain constant, iii) effects of varying

260   the allometric coefficient (β) between two groups while S and α remain constant. In

261   each analysis, β was set according to the range of values (0.1-3.0) observed in over

262   3,200 datasets recently reviewed by Voje (2016). S was varied by setting different

263   values of x. Across real datasets, values of x and α will vary greatly and depend on the

264   units of measurements used. Generally, however, α is small relative to x. Unless

265   otherwise stated we therefore set x to 10 units and α to 0.1. We also examined how

266   the size of these effects varies with variable levels of tissue-biased expression

267   (measured as $\log_2(C_{a,x})$-$\log_2(C_{a,y})$). In all comparisons we fixed G and C_m, to 10,000

268   and 5,000 respectively, to reflect raw values of read counts obtained in a recent RNA-

269   Seq dataset (Harrison *et al.* 2015). $C_{a,y}$ was set to 5,000 so that results obtained reflect

270   an 'average gene'. $C_{a,x}$ varied between 0 and 50,000. It is important to note that results

271   obtained for genes limited to, or biased towards, x will be similar, but inverted relative

272   to y-biased genes with a relationship defined by the rearranging the allometric

273   equation for x.

274   To further explore the practical relevance of these effects we also used our

275   model to simulate expected results using published scaling parameters from real

276   biological data. We chose two examples to reflect the sorts of studies being conducted

277   with real data: i) scaling relationships between soma and testis tissue in different male

278   morphs from four species of insects; ii) scaling relationships between cell types in

279   mammalian brains.

280

281   **Results**

282

283   i. Model effects

284   *Effects of size differences under conserved allometric scaling*

285   We modelled the effect of allometric scaling by varying S between two

286   hypothetical groups, keeping α and β constant in order to identify the influence of

287   simple size differences on the relative proportions of sub-tissues on comparative

288   studies of gene expression (Figure 2A). Specifically, we used our model to compare

289   gene expression levels between two groups, where $x = 10$ for group one, and $0.1 < x <$

290   100 in group two, a ten-fold change in size in both directions. β was fixed at either

291   0.1, 0.5, 1.0, 1.5 and 2.0 in both groups. As expected, under isometric scaling ($\beta = 1$)

292   *FC* is consistently *zero* regardless of the magnitude of size differences between the

293   two groups, or the extent of tissue-biased expression. Turning to allometric scaling,

294   we first consider tissue-specific expression ($C_{a,y}$ = 5,000; $C_{a,x}$ = 0) as we anticipated

295   this would reflect the worst case scenario. The model predicts consistent differences

296   in CPM between groups that increase with greater size differences, or greater

297   deviation from isometry. The effects of negative and positive allometry generally

298   mirror one another, except where extreme positive allometry results in y comprising

299   nearly all of S, minimizing the influence of tissue-biased expression. The opposite

300   will occur for x-specific genes. Large fold-changes (FC $\geq$1 or <-1) are expected to

301   require relatively large size differences. For example, under strong negative allometry

302   (β = 0.1) if x = 10 for group one, group two requires x<4.5 or >22 (a S ratio of <0.45

303   or > 2.19; note in Figure 2 the $\log_{10}(S$ ratio) is plotted to compress the variance for

304   visual clarity) to produce a two-fold expression difference (FC $\geq$1 or <-1). Under

305   strong positive allometry (β = 2) this occurs only when x <3.25 for group two. When

306   the degree of tissue-bias in expression is varied ($C_{a,y}$ = 5,000; $C_{a,x}$ = 0-50,000),

307   increasing tissue-bias in either direction results in larger FC (Figure 2B, C). This

308   effect is amplified according to the degree to which β deviates from one. In summary,

309   our model predicts that where the sample differs in mean size between groups under

310   comparison any deviation from isometric scaling could produce difference in

311   transcript abundance.

312

313   *Effects of varying the allometric constant between groups*

314   We next used our model to assess the impact that differences in the allometric

315   constant between groups have on relative transcript abundance, modelling the

316   expected effects of 'grade-shifts' between groups. This was done by varying α

317   between two groups while S and β remained constant (Figure 3, panel A). With x set

318   to 10 in both groups and an α of 0.1 in group one, we varied α in group two

319   between 0.1 and 1 (a ten fold range). First considering tissue-specific genes ($C_{a,x}$ =

320   5,000; $C_{a,y}$ = 0), the model predicts absolute FC will increase linearly with the log-

321   ratio of α values. When β <1, the magnitude of the effect is largely unaffected by

322   variation in β. Where β >1, the effect is dampened as β increases because the

323   contribution of expression in tissue y quickly overwhelms that of tissue x. The

324   opposite will occur for x-specific genes. Large fold-changes ($\geq$1 or <-1) occur from

325   relatively small shifts in α. Under negative allometry (β <1), if α is 0.1 in group two,

326   the FC is ≥ 1 or <-1 when $0.05 > \alpha > 2$ in group two (an α ratio <0.05 or >2). Under

327   positive allometry the necessary magnitude of shift in α to produce this size of effect

328   increases, but the opposite will occur for *x*-specific genes. Finally, when the degree of

329   tissue bias in expression is varied ($C_{a,y} = 5,000$; $C_{a,x} = 0$-50,000), tissue-specificity is

330   again always the worst-case scenario. Increasing tissue-bias in either direction

331   produces larger *FC*, an effect amplified by increased variance in α between groups

332   (Figure 4B, C). In summary, our model predicts that differences in allometric

333   constants between groups under comparison can have a large impact on transcript

334   abundance, regardless of the similarity in total size of the tissue sampled.

335

336   *Effects of varying the allometric coefficient between groups*

337   Finally, we used our model to predict how this will affect patterns of

338   differential expression in scenarios where the total size across two groups is constant

339   but the scaling relationships between their constituent parts differ. We first varied β

340   while *S* and α remain constant setting $\beta = 0.5$ (Figure 4A) or 1.5 (Figure 4B) in

341   group one, and $0.1 < \beta < 3$ in group two. We repeated this analysis using different

342   values of *x* to explore how variation in β interacts with variation in size (Figure

343   4C,D). First, considering tissue-specific genes ($C_{a,y} = 5,000$; $C_{a,x} = 0$), the model

344   predicts *FC* will increase linearly with β until the contribution of expression in tissue

345   *y* overwhelms that of tissue *x*. The opposite will occur for *x*-specific genes. We find

346   that modest differences in β can produce large *FC* (≥ 1 or <-1). For example, when *x* =

347   10 in both groups and $\beta = 0.5$ in group one, $-1 > FC > 1$ when $0.2 > \beta > 0.9$ in group two

348   (a β ratio of <0.4 or >1.8; Figure 4A). As *x* increases the shift in β necessary to

349   produce this scale of difference decreases; when *x* = 100 it will occur when $0.3 > \beta$

350   >0.7 (a β ratio of <0.6 or >0.78), when *x* = 1,000 it will occur when $0.4 > \beta > 0.6$ (a β

351   ratio of <0.8 or >1.2). Similar results are found regardless of the value set for β in

352   group one. Again, when the degree of tissue-bias in expression is varied ($C_{a,y} = 5,000$;

353    $C_{a,x} = 0$-50,000), genes with tissue-specific expression are always most affected.

354   Increasing tissue-bias in either direction produces larger *FC*, an effect amplified by

355   increased variance in β between groups (Figure 4C,D). In summary, any deviation

356   between the scaling exponents governing the scaling relationships between tissue

357   types in two groups will again lead to predictable differences in transcript abundance.

358

Tissue scaling can produce false negatives

359

360   The previous results focus on false-positives, however it is likely that the same

361 scaling effects will obscure real patterns of group differences in gene expression. To

362 illustrate this effect we used our model to vary $C_{a,x}$ between two groups. In group one

363 $C_{a,x}$ and $C_{a,y}$ were both set to 5,000. In group two $C_{a,y}$ was again set to 5,000 but $C_{a,x}$

364 was set to either 20,000, 10,000, 5,000, 2,500 or 1,250. This simulates the gain of

365 tissue-biased expression in group two with an inter-group $\log_2$-fold change (FC) for

366 $C_{a,x}$ of 2, 1, 0, -1 and -2 respectively. We first examined the effects of varying the

367 average size of the sample (as described above with $x = 10$ for group one, and $0.1 < x$

368 < 100 in group two) whilst keeping α and β constant. We set the scaling parameters to

369 reflect moderately hyper-allometric scaling. As expected, as the size difference

370 between groups increases, the estimated FC rapidly declines (Figure 5A). Turning

371 next to inter-group differences in α, we set α to 0.1 in group one and varied α in

372 group two between 0.1 and 1, whilst keep x at 10 and β at 1.5. Again, as the

373 discrepancy between α_1 and α_2 increases, the measured FC decreases exponentially,

374 with even large FC differences in $C_{a,x}$ dropping below and FC of ± 0.5 (Figure 5B).

375 Finally, we examined the effects of varying β by keeping β at 1.5 in group 1 and

376 varying β between 0.1 and 3 in group 2. α was set to 0.1 and x was set to 10 in both

377 groups. Again an effect of reduced detected FC is found with increase inter-group

378 differences in scaling parameters. Here, the effect is sigmoidal with an accelerated

379 decline in FC as the β ratio exceeds ~2.5 (Figure 5C). Similar results are obtained

380 with alternative values for the scaling parameters. Together the model demonstrates

381 that with increasing deviation from isometry, or increasing inter-group differences in

382 scaling, the detection of true shifts in gene expression becomes increasingly

383 inaccurate potentially leading to substantial numbers of false negatives.

384

385   ii. Biological examples

386

387   *Testes size in male morphs*

388 Relative testes size can vary dramatically across species, often in association with

389 reproductive competition imposed by multiple-mating in females (Harcourt *et al.*

390 1981; Hosken & Ward 2001). In many species multiple male morphs have evolved to

391 exploit alternative reproductive strategies (Gross 1996; Sinervo & Lively 1996).

392   These morphs typically reflect trade-offs in pre and post-copulatory male-male

393   competition, and by extension, investment in sperm production and testes size. Many

394   studies of gene expression in smaller organisms, such as insects, utilise whole-body

395   samples in order to avoid laborious dissections and/or to obtain sufficient RNA for

396   sequencing. However, as whole-body samples are particularly prone to tissue scaling

397   problems, we explored how differences in testes size might affect results using

398   published scaling parameters from $\log_{10}$-$\log_{10}$ regressions between soma and testis

399   mass for two species of dung beetle, (*Onthophagus taurus* and *O. binodis*), a

400   burrowing bee (*Amegilla dawsoni*), and an earwig (*Forficula auricularia*) (Tomkins

401   & Simmons 2002). Each of these species has two male morphs, one that guards

402   females and one that adopts a 'sneaky' male strategy. We are not aware of any

403   whole-body RNA-Seq analyses based on these particular species, but rather use them

404   as an example of how the composition of the tissue sampled may affect perceived

405   levels of differential expression between groups of individuals without the need to

406   invoke morph-specific regulation of gene expression.

407   For each pair of morphs, we used the estimated morph-specific values of β and

408   α to parameterise the model (Table S1), and varied the degree of tissue-bias (here,

409   towards the testis) in expression for an average gene by setting $C_{a,y}$ to 5,000 and $C_{a,x}$

410   to range incrementally between 0 and 50,000, with S set to an realistic body mass. We

411   also extended this range to include increases in $C_{a,y}$ up to 50,000 whilst $C_{a,x}$ was set to

412   0 (i.e. soma-specific gene expression). We then plotted the estimated $\log_2$-fold change

413   in expression (*FC*) between the morphs against the degree of tissue-bias ($\log_2(C_{a,x})$-

414   $\log_2(C_{a,y})$). With the exception of *O. taurus*, the difference in gonad-soma scaling

415   between morphs was sufficient to produce *FC* ≥ 0.5 for genes modelled as testis-

416   specific, with *FC* increasing with testis-specificity in expression (Figure 6A).

417   We further explored how this effect might influence the kind of statistical

418   methods used in real analyses by simulating a modest dataset of 1,000 genes for 5

419   individuals of each morph using the scaling relationships as described above. Here,

420   $C_{a,x}$ and $C_{a,y}$ for each gene were set as equal, random numbers between 1 and 50,000

421   with 100 testis-specific genes and 100 soma-specific genes. Across individuals $C_{a,x}$

422   and $C_{a,y}$ were constrained to be within 10% of expression level of the corresponding

423   gene in the first simulated individual. Under these conditions we would not expect

424   any evidence of significant expression differences due between groups because there

425   is no contribution of regulatory variation, as such, all gene expression differences are

426 solely caused by scaling effects. When we plotted expression in both morphs against

427 one another, the correlations are significant, but show a range of *FC*. Importantly, a

428 proportion of genes is identified as 'significantly differentially expressed' between

429 morphs using standard *t*-tests with no fold-change threshold (Table 1). We next used

430 these data in two multivariate analyses, often utilised in RNA-Seq studies. First we

431 used a Principal Component Analysis (PCA) to compress the variation in the dataset

432 into PCs, we then asked if these PCs are significantly different between morphs using

433 a *t*-test. Second we used hierarchical clustering to test if the simulated data can

434 separate each morph. In three of the four cases the clustering grouped morphs by gene

435 expression and had one PC significantly associated with morph, accounting for 10-

436 16% of variance (Figure 6B-E). We note these values will depend on the permitted

437 degree of variation in expression of a gene between simulations. In each of these

438 analyses the influence of allometry directly reflects differences in the estimated ratio

439 of percentage testis volumes between morphs (Table 1).

440

441 *Cellular scaling in mammalian brains*

442 Many comparative studies have been conducted across species with the aim of

443 identifying species-specific shifts in gene expression. These may focus on specific

444 organs or tissues, but the scaling relationships among cell types could potentially

445 drive some of the observed patterns. Recently interspecific datasets on the cellular

446 composition of mammalian brain regions have revealed variation in the scaling

447 relationship between neurons and non-neuronal cells between brain regions, and for

448 individual structures across mammalian orders (Herculano-Houzel *et al.* 2015). We

449 used these data to explore how allometric relationships between cell types might

450 affect estimates of relative levels of gene expression across species. Using published

451 data we re-estimated the scaling relationship between neurons and non-neuronal cells

452 for two brain structures, the cerebral cortex and cerebellum, across two mammalian

453 orders, glires and primates, using Phylogenetic Generalised Least Square Regressions

454 (Pagel 1999) (Table S2). We used these scaling parameters to explore how variation

455 in cellular scaling might affect comparative studies of gene expression on brain tissue.

456   We first examined the effects of varying S assuming a conserved allometric

457 relationship between neuron and non-neuronal cell number within each order. By

458 setting x_1 to the minimum and x_2 maximum values of non-neuronal cell number

459 observed in each dataset we asked what size of $\log_2$-fold change (*FC*) in gene

460 expression might be observed when comparing gene expression across species within
461 each order, at varying levels of cell-bias in gene expression. The results demonstrate
462 moderate *FC* are expected, but their range varies across structures and orders (Figure
463 7). For the cerebral cortex (Figure 7, panel A), variation in *S* in primates produces
464 more modest *FC* than observed in glires with the largest *FC* (1.49) predicted for genes
465 expressed exclusively in neurons. In contrast, for the cerebellum the pattern is
466 reversed. Primates are predicted to show a greater range of *FC* as *S* varies, with the
467 largest *FC* (-2.8) predicted for genes expressed exclusively in non-neuronal cells
468 (Figure 7, panel B). This difference in pattern between cerebral cortex and cerebellum
469 is most likely related to the pattern of variation in β, which is higher in primates for
470 the cerebral cortex, and higher in glires for the cerebellum.

471 We next explored how the difference in allometric parameters would affect
472 comparisons of individuals (with constant *S*, set to the approximated midpoint in the
473 overlap in ranges of *x* between groups) in different orders (i.e. under different β and
474 α). For the cerebral cortex the model predicts modest *FC* between the two orders (-
475 0.1<*FC*<0.3) (Figure 7, panel C), whereas for the cerebellum we predict a larger
476 range in *FC*, with *FC* increasing as gene expression becomes increasingly biased
477 towards non-neuronal cells (-0.97<*FC*<0.15) (Figure 7, panel D). The analyses above
478 assume gene expression is related to cell number, independently of cell size, in the
479 Supplementary Information we explore the effects of considering cell type mass,
480 rather than number, which leads to broadly similar conclusions.

481

482 **Discussion**
483 Our results illustrate that non-isometric scaling relationships between tissue or cell
484 types within groups of samples, and heterogeneity in scaling relationships across
485 groups of samples, may influence inferred patterns of differential expression. This
486 will occur at multiple biological levels, be it organ types within whole body samples,
487 or cell type abundance when specific tissues are targeted for RNA extraction. We
488 illustrated the effects of our model using simulated expression data, which we
489 generated due to the absence of real RNA-Seq data from samples with accompanying
490 morphometric-scaling information. Although a simplification of a complex problem,
491 our model illustrates how the scaling relationships between sub-components of a
492 heterogeneous tissue sample can result in apparent differences in expression without

493   changes in the regulatory control of a gene. In particular, we highlight the following
494   conclusions:

- Scaling will *always* affect estimates of relative expression except when all
  components of a sample scale isometrically.

- Even where groups have common allometric scaling relationships, large
  differences in mean size between groups can lead to the appearance of
  differential expression. The effect increases with increasing deviation from
  isometry.

- Small differences in the allometric coefficient (β) or allometric constant (α)
  between groups can produce large fold-changes in gene expression. The effect
  is greater with increased deviation in scaling parameters between groups.

- In all cases the effect increases with tissue-bias in expression, and is most
  pronounced for genes expressed only in one tissue.

- Tissue scaling effects can produce both false positive and false negative
  detection of differential gene expression between groups.

508

509   *Recommendations on how to minimise the influence of tissue scaling when inferring*
510   *regulatory variation*

511   Differences in relative expression level between groups or across species will reflect a
512   combination of measurement error, drift, selection and variation in tissue
513   composition. We have presented a simple model that suggests variation in tissue
514   composition caused by non-isometric tissue scaling between groups may have strong
515   implications for identifying genes with altered regulation. The size of the effect is
516   dependent on the variability in tissue composition, variability in tissue size, and the
517   properties of scaling relationships between sub-components of the sampled tissue.
518   Although the effect size varies, any consistent effect between groups that is greater
519   than intra-group variation could produce signatures of significant differential gene
520   expression without any underlying regulatory variation. In real datasets the effects are
521   likely to be more complex than presented above, as variation in tissue size will
522   interact with scaling parameters across multiple classes of cell or tissue types.

523   Recent bioinformatic approaches have been developed to parse expression
524   differences from heterogeneous samples (Gong & Szustakowski 2013; Li & Xie
525   2013). These approaches can be useful if the goal is to identify heterogeneity in cell
526   type abundance across samples. However, they may have limited scope for ecological

16

527   and evolutionary studies. First, they are based on the assumption of conserved

528   regulatory architecture within similar cell types across samples, and may therefore

529   struggle to identify regulatory variation in constituent cells. Second, they require

530   information about transcriptional abundance in 'pure' samples of at least one sub-

531   tissue, and/or data on the proportions of constituent tissue types. This data is unlikely

532   to be available for the majority of ecological studies, and if it were, it would often be

533   a preferable source of the primary data for analysis. In the absence of readily

534   applicable bioinformatics tools we recommend the influence of tissue scaling should

535   be considered in the design and analysis of comparative studies of gene expression. In

536   particular we recommend the following approaches:

537

538   1) **Use fold-change thresholds:** Small but consistent effects of tissue scaling may

539   produce significant differences in gene expression when analyzed with standard

540   pairwise statistical tests. Introducing fold-change thresholds when identifying

541   differentially expressed genes will go a long way to reducing the false-positive

542   effects of tissue scaling on downstream analyses. Based on the results described

543   above, a $\log_2$-fold change of 1, as previously used in several studies (e.g. Pointer

544   *et al.* 2013; Harrison *et al.* 2015), would provide an adequate threshold in a range

545   of scenarios. We would recommend higher thresholds when comparing tissues or

546   groups/species with increasingly different phenotypic sizes or compositions. It

547   may also be necessary to consider higher thresholds for tissue-specific genes. Of

548   course, fold-change thresholds do not avoid false negatives, and to combat the

549   false positive inflation it may be necessary to accept an increase in false-negative

550   rate. However, we note that many studies of gene expression have identified genes

551   with considerably higher fold-changes between comparisons than we suggest as a

552   minimal threshold. This is true both for candidate genes (e.g. Palmer *et al.* 2016)

553   and transcriptome-wide analyses (e.g. Brawand *et al.* 2011 see Figure 3).

554   Although sometimes controversial, adopting fold-change thresholds is therefore

555   unlikely to be prohibitive to the inference of altered regulation in sufficiently well

556   powered and well-designed studies.

557

558   2) **Know your phenotype:** Many RNA-Seq experiments are conducted with the aim

559   of understanding the molecular basis of divergent phenotypes, be they specific

560   differences in the development of a trait or broad differences in individuals with

different behavioural or ecological strategies. At least a modest understanding of the phenotype in question is necessary to design informative studies of divergence in gene expression. Where possible, more precise tissue sampling will likely produce estimates of relative gene expression that more accurately reflect real variation in gene regulation. In addition to manual dissections, in 'ideal' conditions laser capture micro-dissection may provide a route to more accurate tissue sampling (Espina *et al.* 2006). In the many situations where such an approach is currently un feasible, quantifying variation in the size or composition of tissue to be analyzed may still help improve both experimental design and the interpretation of results. Estimates of scaling parameters between major tissues in the sample, either measured directly from samples for RNA-Seq, or approximated from comparable phenotypic studies, can be used to estimate the fold-change thresholds needed to minimise the effects of tissue scaling and maximise power to detect true signals of regulatory divergence. Technical difficulties in performing dissections while maintaining RNA integrity, small organism size, or simply time and expense required for additional samples, may still prevent collecting data on scaling parameters. In cases such as these, ruling out the contribution of tissue scaling is more difficult, but steps can still be taken to minimise the effect, for example by implementing more conservative fold-change thresholds.

3) **Be wary of tissue-specific genes:** Our model suggests genes with strong tissue- or cell-biased expression will be particularly prone to large changes in expression level caused by tissue scaling, and the most susceptible genes are tissue- or cell-specific. Where possible, genes identified as being differentially expressed in heterogeneous tissue samples should be examined for over-representation of tissue-specific genes in detailed expression databases, such as Flybase (Attrill *et al.* 2015) or the Mouse Atlas (Richardson *et al.* 2014). Of course, this is only possible in model species and their close relatives. It is also worth noting that tissue-biased genes may be more amenable to the action of selection, and/or may have biologically important roles in the phenotype of interest. It may therefore be reasonable to expect tissue-biased genes to be among the most differentially expressed genes in a comparative study using RNA-Seq for multiple reasons.

594   4) **Be wary of divergence along single principal components:** Multivariate

595   analyses have frequently been applied to gene expression studies to show that

596   different groups of individual samples can be distinguished based on their patterns

597   of gene expression (e.g. Brawand *et al.* 2011; Ghalambor *et al.* 2015). Our

598   analyses suggest this result can be produced solely by differences in tissue

599   composition. The variance accounted for by this effect will depend on the relative

600   balance between within group variation and the effect size of any scaling

601   differences between groups. We expect that in many cases the scaling effects will

602   primarily load on one single Principal Component (see Figure 6). To demonstrate

603   that groups of samples are genuinely distinct in their transcription patterns we

604   recommend requiring isolation across at least two dimensions in any multivariate

605   analysis. We also note that where phenotypic data can be collected, it may be

606   possible to include this in a multivariate analysis of gene expression to control for

607   major differences in tissue composition between groups.

608

609   5) **Introduce phenotypic data into neutral models of gene expression:** Although

610   we have focused on pairwise comparisons of groups, the effects of tissue scaling

611   will also affect phylogenetic analysis of gene expression. For example, an

612   Ornstein-Uhlenbeck (OU) model has been proposed as a potential model of

613   expression divergence, facilitating the identification of shifts in expression that

614   were putatively caused be selection (Brawand *et al.* 2011; Rohlfs *et al.* 2014). OU

615   models simulate adaptive optima across a phylogeny with stabilizing selection

616   constraining divergence around these optima (Martins 1994; Beaulieu *et al.* 2012).

617   The presence of multiple optima is interpreted as evidence of variation in

618   selection pressure across species. We suspect that tissue scaling could also

619   produce a pattern of divergence across species which is similar to that predicted

620   under an OU model. Where species in a phylogenetic dataset vary extensively by

621   size, or differ in their scaling relationships, patterns of expression linked to tissue

622   composition may not fit an OU model with a single optimum, giving the

623   appearance of adaptive changes in expression level. Similar effects could be

624   imagined under alternative comparative models which may prove useful for

625   studying gene expression if large enough datasets can be assembled, such as

626   incorporating heterogeneity in evolutionary rate across branches of a phylogeny

627   (Venditti *et al.* 2011). We suggest further exploration of how the effects of tissue

628   scaling may affect these methods is necessary. If found to be prohibitive, one

629   solution may be to incorporate phenotypic variation in the null model as an

630   explicit error term, as has been done in studies of intraspecific variation (Rohlfs *et*

631   *al.* 2014), or as a co-factor in the analysis.

632

633   6) **Single-cell transcriptome analysis:** Analysis of gene expression within single

634   cells is becoming an increasingly feasible option (Sandberg 2014). Single-cell

635   transcriptomics is free from the complicating effects of scaling between

636   components of a heterogeneous tissue sample making the inference of regulatory

637   change more direct. However, these analyses remain technically difficult partly

638   because they require either cell culture or dissociation of cell aggregates from

639   live-caught samples, and partly because they require many replicates of many cell

640   types to uncover the full regulatory diversity of any single organ. Due to the need

641   for increased amplification steps, single-cell analyses may also require substantial

642   replication to overcome inaccuracy in measuring all but the highest expression

643   ranges. The combination of technical difficulty, cell culture or disaggregation and

644   expense from extra replication may discourage many labs from adopting single-

645   cell analysis for evolutionary or ecological questions, particularly in non-model

646   species. However, as with all next-generation technologies, improvements may

647   soon remove some of these technical barriers leaving sample availability and

648   collection as the primary limiting step.

649

650   **Conclusion**

651   Comparative analysis of gene expression provides a potentially powerful tool in the

652   evolutionary biologist's toolkit. In an ecological or evolutionary context, most studies

653   utilizing this tool aim to understand the relationship between variation in the

654   regulation of gene expression and phenotypic variation. We have argued that our

655   ability to infer this relationship can be affected by the scaling relationships between

656   sub-tissues of the sample used to obtain RNA. In some scenarios the effect can

657   produce the appearance large fold changes in gene expression. We have presented a

658   simple model to explore whether, and under what scenarios, tissue scaling can

659   produce perceptions of large expression differences without altered gene regulation.

660   Our results suggest that under non-isometric scaling, or when comparing individuals

661   with different scaling relationships, the effects can be moderate to severe. Based on

662   these analyses, we have suggested a number of experimental and analytical
663   approaches that may go some way to minimising the effects of tissue scaling on down
664   stream analyses of genes with divergent gene expression. The absence of datasets
665   with both gene expression datasets and information on tissue scaling relationships has
666   prevented a full exploration of these effects in real data. The addition these kinds of
667   datasets, potentially derived from experimental mixing of cell cultures, would permit
668   a useful test of our results and may potential provide further improvements on how to
669   analyse expression data derived from heterogeneous tissues. However, we note many
670   of the effects we describe are observable in published work and are most notable
671   where direct comparisons can be made between whole-body and tissue-specific
672   expression datasets. For example, Perry *et al.* (2014) showed that tissue specific
673   sequencing of gonad transcriptomes produce greater numbers of sex-biased genes,
674   consistent with the effects of somatic tissue diluting this signal in whole-body RNA
675   libraries. Although we fully expect comparative studies of gene expression to
676   continue to illuminate the gene-phenotype relationship, we caution against the naïve
677   assumption that all differences in expression level are the result of altered gene
678   regulation.
679

Acknowledgments

692

Data Accessibility

694   This paper has no accompanying data.

695 References

696 Abzhanov A, Kuo WP, Hartmann C *et al.* (2006) The calmodulin pathway and
697   evolution of elongated beak morphology in Darwin's finches. *Nature*, **442**, 563–
698   567.

699 Alonzo SH, Taborsky M, Wirtz P (2000) Male alternative reproductive behaviours in
700   a mediterranean wrasse, *Symphodus ocellatus*: Evidence from otoliths for
701   multiple life-history pathways. *Evolutionary Ecology Research*, **2**, 997–1007.

702 Attrill H, Falls K, Goodman JL *et al.* (2015) FlyBase: establishing a Gene Group
703   resource for *Drosophila melanogaster. Nucleic Acids Research*, **44**, gkv1046–.

704 Barton RA, Harvey PH (2000) Mosaic evolution of brain structure in mammals.
705   *Nature*, **405**, 1055–8.

706 Barton RA, Venditti C (2014) Rapid evolution of the cerebellum in humans and other
707   great apes. *Current Biology*, **24**, 2440–2444.

708 Beaulieu JM, Jhwueng DC, Boettiger C, O'Meara BC (2012) Modeling stabilizing
709   selection: Expanding the Ornstein-Uhlenbeck model of adaptive evolution.
710   *Evolution*, **66**, 2369–2383.

711 Brawand D, Soumillon M, Necsulea A *et al.* (2011) The evolution of gene expression
712   levels in mammalian organs. *Nature*, **478**, 343–348.

713 Carroll SB (2008) Evo-Devo and an expanding evolutionary synthesis: A genetic
714   theory of morphological evolution. *Cell*, **134**, 25–36.

715 Chen Y-C, Harrison PW, Kotrschal A *et al.* (2015) Expression change in
716   *Angiopoietin-1* underlies change in relative brain size in fish. *Proceedings of the
717   Royal Society B: Biological Sciences*, **282**, 20150872.

718 Egset CK, Hansen TF, Le Rouzic a. *et al.* (2012) Artificial selection on allometry:
719   Change in elevation but not slope. *Journal of Evolutionary Biology*, **25**, 938–
720   948.

721 Emlen D (1996) Artificial selection on horn length-body size allometry in the horned
722   beetle *Onthophagus acuminatus* (Coleoptera: Scarabaeidae). *Evolution*, **50**,
723   1219–1230.

724 Enard W, Khaitovich P, Klose J *et al.* (2002) Intra- and interspecific variation in
725   primate gene expression patterns. *Science*, **296**, 340–343.

726 Espina V, Wulfkuhle JD, Calvert VS *et al.* (2006) Laser-capture microdissection.
727   *Nature Protocols*, **1**, 586–603.

728   Feldmeyer B, Elsner D, Foitzik S (2014) Gene expression patterns associated with

729   caste and reproductive status in ants: Worker-specific genes are more derived

730   than queen-specific ones. *Molecular Ecology*, **23**, 151–161.

731   Ghalambor CK, Hoke KL, Ruell EW *et al.* (2015) Non-adaptive plasticity potentiates

732   rapid adaptive evolution of gene expression in nature. *Nature*, **525**, 372–375.

733   Ghalambor C, McKay J, Carroll S, Reznick D (2007) Adaptive versus non-adaptive

734   phenotypic plasticity and the potential for contemporary adaptation in new.

735   *Functional Ecology*, **21**, 394–407.

736   Gong T, Szustakowski JD (2013) DeconRNASeq: A statistical framework for

737   deconvolution of heterogeneous tissue samples based on mRNA-Seq data.

738   Bioinformatics, 29, 1083–1085.

739   Gross MR (1996) Alternative reproductive strategies and tactics: diversity within

740   sexes. *Trends in Ecology & Evolution*, **11**, 92–98.

741   Harcourt AH, Harvey PH, Larson SG, Short R V (1981) Testis weight, body weight

742   and breeding system in primates. *Nature*, **293**, 55–57.

743   Harrison PW, Wright AE, Zimmer F *et al.* (2015) Sexual selection drives evolution

744   and rapid turnover of male gene expression. *Proceedings of the National*

745   *Academy of Sciences*, **112**, 201501339.

746   Herculano-Houzel S, Catania K, Manger PR, Kaas JH (2015) Mammalian brains are

747   made of these: A dataset of the numbers and densities of neuronal and

748   nonneuronal cells in the brain of glires, primates, scandentia, eulipotyphlans,

749   afrotherians and artiodactyls, and their relationship with body mass. *Brain,*

750   *Behavior and Evolution*, **2015**, 145–163.

751   Hollis B, Houle D, Yan Z, Kawecki TJ, Keller L (2014) Evolution under monogamy

752   feminizes gene expression in *Drosophila melanogaster*. *Nature communications*,

753   **5**, 3482.

754   Hosken DJ, Ward PI (2001) Experimental evidence for testes size evolution via sperm

755   competition. *Ecology Letters*, **4**, 10–13.

756   Immonen E, Snook RR, Ritchie MG (2014) Mating system variation drives rapid

757   evolution of the female transcriptome in *Drosophila pseudoobscura. Ecology*

758   *and Evolution*, **4**, 2186–2201.

759   Khaitovich P, Muetzel B, She X *et al.* (2004) Regional patterns of gene expression in

760   human and chimpanzee brains. *Genome Research*, **14**, 1462–1473.

761   Khila A, Abouheif E, Rowe L (2012) Function, developmental genetics, and fitness

762 consequences of a sexually antagonistic trait. *Science*, **336**, 585–589.

763 Kotrschal A, Rogell B, Bundsen A *et al.* (2013) Artificial selection on relative brain

764 size in the guppy reveals costs and benefits of evolving a larger brain. *Current*

765 *Biology*, **23**, 168–171.

766 Kvist J, Wheat CW, Kallioniemi E *et al.* (2013) Temperature treatments during larval

767 development reveal extensive heritable and plastic variation in gene expression

768 and life history traits. *Molecular Ecology*, **22**, 602–619.

769 Li Y, Xie X (2013) A mixture model for expression deconvolution from RNA-seq in

770 heterogeneous tissues. BMC bioinformatics, 14 Suppl 5, S11.

771 Martins EP (1994) Estimating the Rate of Phenotypic Evolution from Comparative

772 Data. *American Naturalist*, **144**, 193–209.

773 Moczek AP, Rose DJ (2009) Differential recruitment of limb patterning genes during

774 development and diversification of beetle horns. *Proceedings of the National*

775 *Academy of Sciences of the United States of America*, **106**, 8992–8997.

776 Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**,

777 877–884.

778 Palmer WJ, Duarte A, Schrader M *et al.* (2016) A gene associated with social

779 immunity in the burying beetle N*icrophorus vespilloides*. *Proceedings of the*

780 *Royal Society B: Biological Sciences*, **238** (1823), 0–7.

781 Parker DJ, Cunningham CB, Walling CA *et al.* (2015) Transcriptomes of parents

782 identify parenting strategies and sexual conflict in a subsocial beetle. *Nature*

783 *Communications*, **6**, 8449.

784 Perry JC, Harrison PW, Mank JE (2014) The ontogeny and evolution of sex-biased

785 gene expression in drosophila melanogaster. Molecular Biology and Evolution,

786 31, 1206–1219.

787 Pointer MA, Harrison PW, Wright AE, Mank JE (2013) Masculinization of gene

788 expression is associated with exaggeration of male sexual dimorphism. *PLoS*

789 *Genetics*, **9**.

790 Richardson L, Venkataraman S, Stevenson P *et al.* (2014) EMAGE mouse embryo

791 spatial gene expression database: 2014 update. *Nucleic Acids Research*, **42**, 1–

792 10.

793 Rohlfs R V., Harrigan P, Nielsen R (2014) Modeling gene expression evolution with

794 an extended ornstein-uhlenbeck process accounting for within-species variation.

795 *Molecular Biology and Evolution*, **31**, 201–211.

796   Rosenberg MS (2002) Fiddler crab claw shape variation: a geometric morphometric

797   analysis across the genus *Uca* (Crustacea: Brachyura: Ocypodidae). *Biological*

798   *Journal of the Linnean Society*, **75**, 147–162.

799   Sandberg R (2014) Entering the era of single-cell transcriptomics in biology and

800   medicine. *Nature Methods*, **11**, 22–24.

801   Shapiro MD, Marks ME, Peichel CL *et al.* (2004) Genetic and developmental basis of

802   evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717–723.

803   Simmons LW, Tomkins JL (1996) Sexual selection and the allometry of earwig

804   forceps. *Evolutionary Ecology*, **10**, 97–104.

805   Sinervo B, Lively CM (1996) The rock–paper–scissors game and the evolution of

806   alternative male strategies. *Nature*, **380**, 240–243.

807   Standage DS, Berens AJ, Glastad KM *et al.* (2016) Genome, transcriptome, and

808   methylome sequencing of a primitively eusocial wasp reveal a greatly reduced

809   DNA methylation system in a social insect. *Molecular Ecology*, **25**, 1769-1784.

810   Steiner CC, Weber JN, Hoekstra HE (2007) Adaptive variation in beach mice

811   produced by two interacting pigmentation genes. *PLoS Biology*, **5**, 1880–1889.

812   Stern DL, Orgogozo V (2008) The loci of evolution: How predictable is genetic

813   evolution? *Evolution*, **62**, 2155–2177.

814   Stiver KA, Harris RM, Townsend JP, Hofmann HA, Alonzo SH (2015) Neural gene

815   expression profiles and androgen levels underlie alternative reproductive tactics

816   in the ocellated wrasse, *Symphodus ocellatus*. *Ethology*, **121**, 152–167.

817   Stuglik MT, Babik W, Prokop Z, Radwan J (2014) Alternative reproductive tactics

818   and sex-biased gene expression: The study of the bulb mite transcriptome.

819   *Ecology and Evolution*, **4**, 623–632.

820   Tomkins JL, Simmons LW (2002) Measuring relative investment: a case study of

821   testes investment in species with alternative male reproductive tactics. *Animal*

822   *Behaviour*, **63**, 1009–1016.

823   Toth AL, Smith CR, Suarez A V., Robinson GE (2008) Genetic and genomic analyses

824   of the division of labour in insect societies. *Nature Reviews Genetics*, **9**, 735–

825   748.

826   Venditti C, Meade A, Pagel M (2011) Multiple routes to mammalian diversity.

827   *Nature*, **479**, 393–6.

828   Voje KL (2016) Scaling of morphological characters across trait type, sex, and

829   environment: A meta-analysis of static allometries. *The American Naturalist*,

830   **187**, 89–98.

831   Wilkinson GS (1993) Artificial sexual selection alters allometry in the stalk-eyed fly

832   *Cyrtodiopsis dalmanni* (Diptera: Diopsidae). *Genetical Research*, **62**, 213–222.

833

834

835

836

837

838

839

840

841

842

843

844

845

846   Tables

847   Table 1. Results of the simulated data sets based on scaling parameters between male morphs of multiple insects

848

Species	Pearson correlation		log$_2$-fold change (N)			differentially expressed[1]	
	r	**p**	**mean**	**minimum**	**maximum**	**p < 0.05**	**p < 0.001**
Onthophagus binodis	0.995	<0.001	-0.233	-2.239	0.007	121 (111)	106 (103)
Onthophagus taurus	0.999	<0.001	0.003	-0.001	0.026	1 (0)	0 (0)
Forficula auriculaira	0.998	<0.001	-0.104	-0.866	0.014	166 (126)	119 (96)
Amegilla dawsoni	0.999	<0.001	-0.05	-0.473	0.002	107 (101)	79 (31)

849

850   [1] numbers in parentheses are after Bonferoni correction for multiple tests.

851

852

853

854

855

856

857

858

859

860

861

862   Figure legends

863

864   Figure 1. **Types of scaling relationships and how they shape proportional size.**

865   Here we show a hypothetical comparison between two groups of individuals which

866   may differ in size and which are comprised of two tissues. In each scenario, row 1

867   shows the relationship between tissue A and total size for individuals from two groups

868   (red and blue). The scaling relationships are determined by the allometric equation $y =$

869   αx^{β}, where β is the scaling coefficient and α is the scaling constant. Row 2 shows

870   illustrative examples of individuals from each group imagining tissue A as gonad size.

871   Note, this is only an example and components tissues can be any aspect of

872   morphology. Row 3 shows an illustration of how the proportion of tissue A (coloured)

873   varies between groups as a result of the scaling relationship and differences in mean

874   size. Row 4 shows the effects these proportional differences might have on relative

875   gene expression, illustrated with box whisker plots.

876

877   Figure 2. **Effects of size differences under conserved allometric scaling.** A) Effects

878   of comparing two groups with different total sizes under alternative scaling

879   coefficients, β. The $\log_2$-fold change is plotted against the ratio of the total size of two

880   groups. In this comparison $x = 10$ in group one and varied x in group two between 0.1

881   and 100. Effects of comparing two groups with different levels of tissue-biased

882   expression B) under hyper-allometry ($\beta = 2$) and C) under hypo-allometry ($\beta =$

883   0.1). In B and C coloured lines indicate comparisons where expression of gene a is set

884   to 5,000 in component y and it's expression in component x is varied as indicated in

885   the colour key. The black dashed line indicates a comparison where expression of

886   gene a is set to 0 in component y and 5,000 in component x. Dashed grey lines

887   indicated a FC of ±1, often used as a threshold of significant difference in expression.

888

889   Figure 3. **Effects of varying the allometric constant between groups.** A) Effects of

890   comparing two groups with different scaling constants, α, across different shared

891   scaling coefficients (β), with α in group one set to 0.1 and varying α in group two

892   between 0.1 and 10. The effects of comparing two groups with different α across

893   different levels of tissue-biased expression B) under hyper-allometry ($\beta = 1.5$) and C)

894   under hypo-allometry ($\beta = 0.5$). In B and C coloured lines indicate comparisons where

895   expression of gene *a* is set to 5,000 in component *y* and it's expression in component

896   *x* is varied as indicated in the colour key. The black dashed line indicates a

897   comparison where expression of gene *a* is set to 0 in component *y* and 5,000 in

898   component *x*. The $\log_2$-fold change is plotted against the ratio of the α of each group.

899   Dashed grey lines indicated a FC of ±1.

900

901   Figure 4. **Effects of varying the allometric coefficient between groups.** Effects of

902   comparing two groups with different scaling coefficients, *β*, across different units of

903   size (*x*) with A) *β* in group one set to 0.5 and varying *β* in group two between 0.1 and

904   3, and B) *β* in group one set to 1.5 and varying *β* in group two between 0.1 and 3.

905   Effects of comparing two groups with different levels of tissue-biased expression with

906   C) *β* in group one set to 0.5 and varying *β* in group two between 0.1 and 3 and D) *β* in

907   group one set to 1.5 and varying *β* in group two between 0.1 and 3. In C and D

908   coloured lines indicate comparisons where expression of gene *a* is set to 5,000 in

909   component *y* and it's expression in component *x* is varied as indicated in the colour

910   key. The black dashed line indicates a comparison where expression of gene *a* is set

911   to 0 in component *y* and 5,000 in component *x*. The $\log_2$-fold change is plotted against

912   the ratio of the *β* of each group. Dashed grey lines indicated a FC of ±1, often used as

913   a threshold of significant difference in expression.

914

915   Figure 5. **Tissue scaling effects can mask true positives.** A) Effects of non-isometric

916   but conserved scaling on the detection of a differentially expressed gene. Two groups

917   were modelled with conserved scaling constant, *α* (0.1), and scaling coefficient, *β*

918   (1.5), values but different total sizes. The estimated $\log_2$-fold change is plotted against

919   the mass ratio, setting *x* in group one to be 10, and varying *x* in group two between 0.1

920   and 100. B) Effects of 'grade-shifts', or group differences in α, on the detection of a

921   differentially expressed gene. Two groups were modelled with conserved sizes (*x* =

922   10) and *β* (1.5) values but different α values. The estimated $\log_2$-fold change is plotted

923   against the mass ratio, setting α in group one to be 0.1, and varying *x* in group two

924   between 0.1 and 10. C) Effects of group differences in *β* on the detection of a

925   differentially expressed gene. Two groups were modelled with conserved sizes (*x* =

926   10) and α (0.1) values but different α values. The estimated $\log_2$-fold change is plotted

927   against the mass ratio, setting *β* in group one to be 1.5, and varying *β* in group two

928   between 0.1 and 3. In each case expression of gene *a* in subcomponent *y* is 5,000. In

929   group one expression of *a* in *x* is 5,000 but expression of *a* in *x* varies in group two

930   taking values of either 20,000, 10,000, 5,000, 2,500 or 1,250 (representing $\log_2$-fold

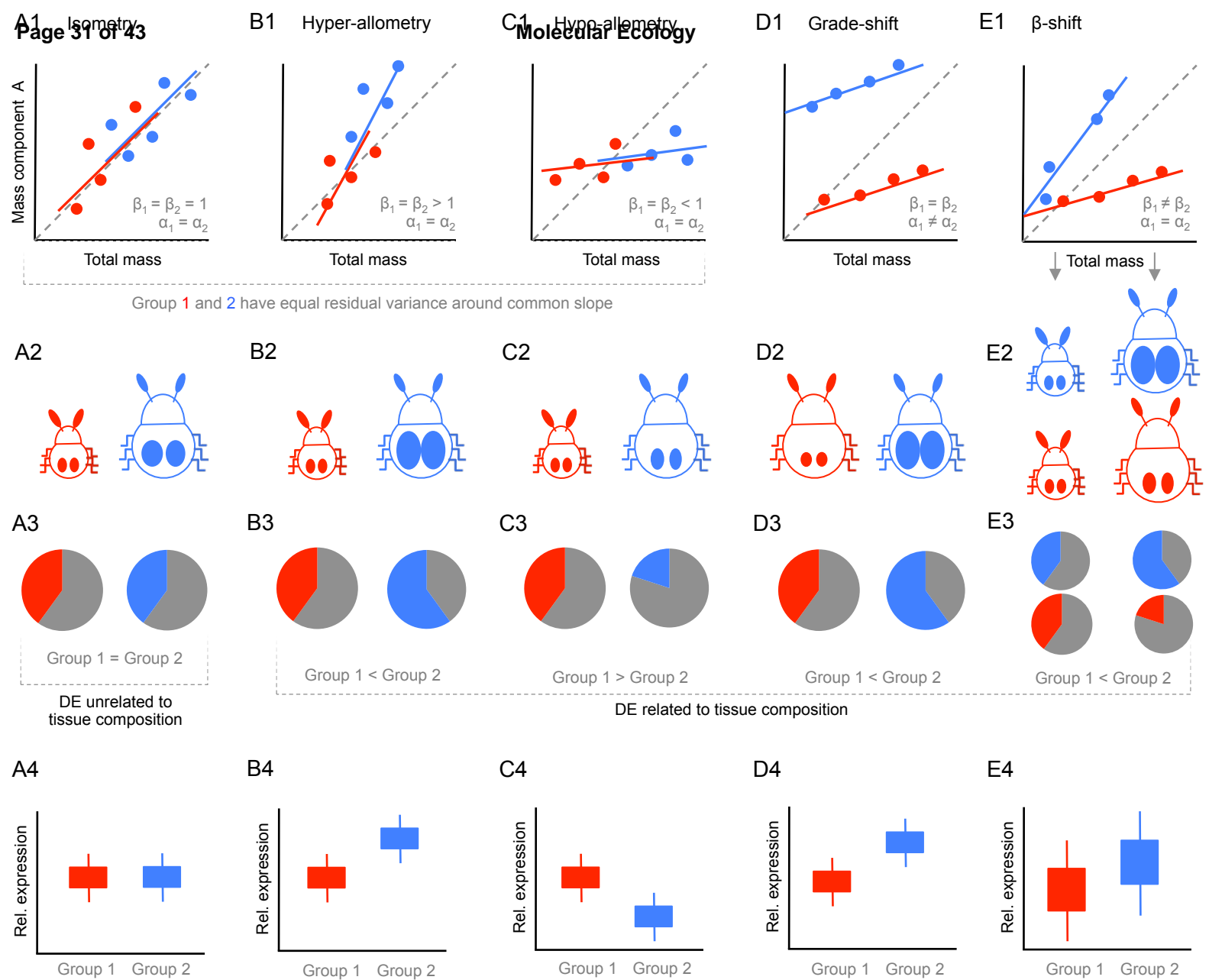931   change values of 2, 1, 0, -1 and -2 respectively).

932

933   Figure 6. **Predicted differences in relative expression level between male morphs**

934   **of multiple species of insect based on testis~soma scaling**. A) Predicted fold-change

935   in expression across different levels of tissue-biased expression ($C_{a,x}$ = gonad

936   expression, $C_{a,y}$ = soma expression). B-E) Results of Principal Component Analyses

937   (B1-E1) and hierarchical clustering (B2-E2) using simulated datasets from the model

938   paramterised using testis~soma scaling relationships for *O. taurus* (B), *A. dawsoni*

939   (C), *F. auricularia* (D) and *O. binodis* (E). In the PCAs, we plot the PC significantly

940   associated with morph type (indicated by *) against PC1. Colours indicate different

941   categories of male morph.

942

943   Figure 7. **Predicted differences in relative expression level between or within**

944   **primates and glires based on scaling relationships between neuron number and**

945   **non-neuronal cell number in the cerebral cortex and cerebellum.** A-B) Predicted

946   fold-change between two groups representing the smallest and largest individuals

947   within primates (blue) and glires (red) assuming conserved, order-specific scaling

948   relationships and varying levels of tissue-biased expression. A) Results for cerebral

949   cortex and B) results for cerebellum. C-D) Predicted differences in gene expression

950   between two group of individuals, one with glire-scaling relationships and one with

951   primate-scaling relationships, but which have an equal, constant size. Results show

952   the predicted fold-change across different levels of tissue bias for C) the cerebral

953   cortex, and D) the cerebellum.

954

A1 Isometry

$\beta_1 = \beta_2 = 1$
$\alpha_1 = \alpha_2$

B1 Hyper-allometry

$\beta_1 = \beta_2 > 1$
$\alpha_1 = \alpha_2$

C1 Hypo-allometry

$\beta_1 = \beta_2 < 1$
$\alpha_1 = \alpha_2$

D1 Grade-shift

$\beta_1 = \beta_2$
$\alpha_1 \neq \alpha_2$

E1 β-shift

$\beta_1 \neq \beta_2$
$\alpha_1 = \alpha_2$

Mass component A — Total mass

Group 1 and 2 have equal residual variance around common slope

A2 **B2** **C2** **D2** **E2**

A3 Group 1 = Group 2

DE unrelated to tissue composition

B3 Group 1 < Group 2

C3 Group 1 > Group 2

D3 Group 1 < Group 2

E3 Group 1 < Group 2

DE related to tissue composition

A4 Rel. expression — Group 1 Group 2

B4 Rel. expression — Group 1 Group 2

C4 Rel. expression — Group 1 Group 2

D4 Rel. expression — Group 1 Group 2

E4 Rel. expression — Group 1 Group 2

A

Log$_2$ fold change

α = 0.1
β = 1.5
x = variable

Log$_{10}$[Mass ratio] (S_2/S_1)

B

α = variable
β
x = 10

α ratio ($α_2/α_1$)

C

α = 0.1
β = variable
x = 10

$C_{a,x2} = 20×10^3$
$C_{a,x2} = 10×10^3$
$C_{a,x2} = 5×10^3$
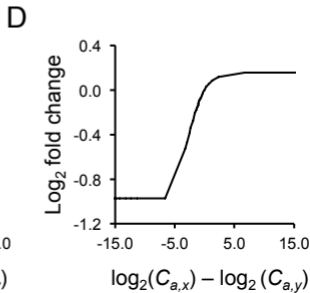$C_{a,x2} = 2.5×10^3$
$C_{a,x2} = 1.25×10^3$

β ratio ($β_2/β_1$)

Molecular Ecology
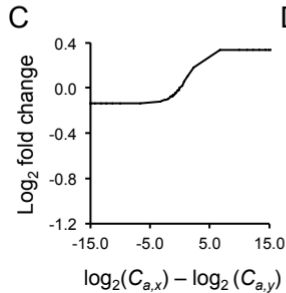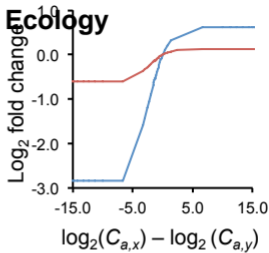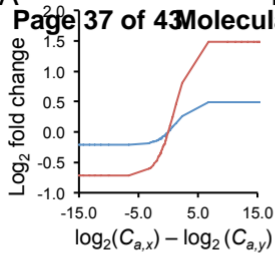

A

B

C

D

Inferring regulatory change from gene expression:

the confounding effects of tissue scaling

Stephen H. Montgomery[1,2] and Judith E. Mank[1]

[1] Dept. Genetics, Evolution and Environment, University College London, London
WC1E 6BT, UK

[2] Corresponding author: Stephen.Montgomery@cantab.net

SUPPLEMENTARY INFORMATION

1. Supplementary Results: Cellular scaling in mammalian brains

2. Supplementary Figure 1: Predicted differences in relative expression level
    between or within primates and glires based on scaling relationships between
    neuron number/mass and non-neuronal cell number/mass in the cerebral
    cortex and cerebellum.

3. Supplementary Table 1: Scaling parameters and variation in proportional
    gonad size in male morphs of multiple insects

    Supplementary Table 2: Scaling parameters and variation in number of non-
    neuronal cells (x) in mammalian brain components

4. Supplementary Reference

1. Supplementary Results

Cellular scaling in mammalian brains

The analyses in the main text assume gene expression is related to cell number, independently of cell size. Neuronal cell size can differ dramatically across mammalian orders and across brain components (Mota & Herculano-Houzel 2014) (Table 3). In contrast, non-neuronal cells are much more consistent in size (Mota & Herculano-Houzel 2014). We repeated the analyses above multiplying the estimated transcript number per cell by the average cell size estimated for rodents and primates for each brain structure, assuming rodents reflect the glire average (Table S2; Figure S1 panels A2, B2, C2 and D2). The effects of incorporating cell size vary across structures. In the cerebral cortex, where neuron size is much greater than non-neuronal size and more variable across orders, incorporating cell size shifts the range of *FC* estimated when varying S such that the effect on neuron specific genes is reduced whilst the effect on non-neuron specific genes is increased (Figure S1, panel A2). A similar pattern is found for the cerebellum, but in the opposite direction (Figure S1, panel B2). Accounting for variation in cell size between orders reduces predicted *FC*, for the cerebral cortex the range of *FC* becomes very small when *S* is constant (Figure S1, panel C2), whilst for the cerebellum *FC* is reduced primarily in the range of genes with biased expression towards non-neuronal cells.
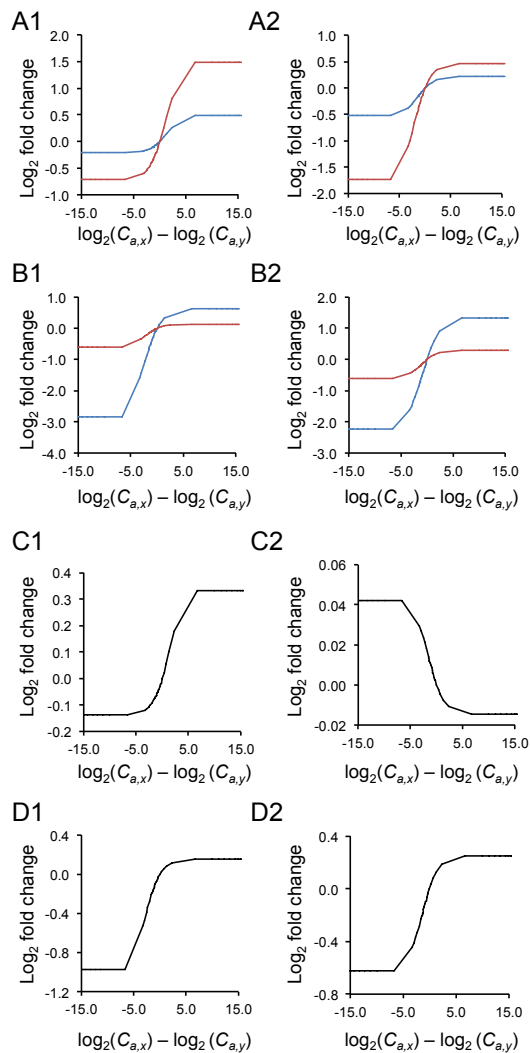
2. Supplementary Figure

Figure S1. **Predicted differences in relative expression level between or within primates and glires based on scaling relationships between neuron number and non-neuronal cell number in the cerebral cortex and cerebellum.** A-B) Predicted fold-change between two groups representing the smallest and largest individuals within primates (blue) and glides (red) assuming conserved, order-specific scaling relationships and varying levels of tissue-biased expression. A) Results for cerebral cortex and B) results for cerebellum. C-D) Predicted differences in gene expression between a group of individuals with glire-scaling relationships and a group with primate-scaling relationships which have an equal, constant size. Results show the predicted fold-change across different levels of tissue bias for C) the cerebral cortex, and D) the cerebellum. For comparison, A1-D1 show the results based on cell number as in Figure 7. A2-D2 show the results incorporating variation in average cell size.

3

3. Supplementary Tables

Table S1. Scaling parameters and variation in proportional gonad size in male morphs of multiple insects

Species[2]	Morph A (testes ~ soma)		Morph B (testes ~ soma)		x [Soma mass (mg)]	approx. % Testes (y/x)		
	β	α[1]	β	α[1]		Morph A	Morph B	Ratio
Onthophagus binodis	1.0480	0.0010	0.8160	0.0138	100	0.591	0.124	4.766
Onthophagus taurus	0.6100	0.1702	0.8040	0.0755	60	3.443	3.406	1.011
Forficula auriculaira	0.6020	0.0718	0.4190	0.0675	20	2.112	1.11	1.903
Amegilla dawsoni	0.5430	2.4491	0.8450	4.1976	500	0.467	0.336	1.390

[1] calculated from Log(α) in Tomkins and Simmons

[2] the allometric parameters for *A. dawsoni* were estimated using body mass in g. All other estimates used mg. Parameters for *F. auricularia* are based on dry mass.

Table S2. Scaling parameters and variation in number of non-neuronal cells (*x*) in mammalian brain components

Order	Brain structure	Neuron number ~ non-neuron number		Range of *x* (non-neuron number) [1]		Mean cell mass (pg) [1]	
		β	log$_{10}$(α)	Range of *x* (non-neuron number)	Log$_{10}$(Maximum)	neuron	non-neuron
Primates	Cerebral cortex	0.928	0.299	7.849	10.784	25.513	4.417
Glires	Cerebral cortex	0.717	1.962	6.924	9.267	37.697	4.527
Primates	Cerebellum	0.649	3.788	7.241	10.205	1.479	4.217
Glires	Cerebellum	0.890	1.521	6.739	8.757	1.889	4.03

[1] smallest/largest primate: *Microcebus murinus/Homo sapiens*

[1] smallest/largest glire: *Heterocephalus glaber/Hydrochaeris hydrochaeris*

[1] Mota and Herculano-Houzel 2014

5

4. Supplementary References

Mota B, Herculano-Houzel S (2014) All brains are made of this: a fundamental
building block of brain matter with matching neuronal and glial masses.
Frontiers in neuroanatomy, **8**, 127.