

1 **Title Page**
2

3 **A comparison of gene expression and DNA methylation
4 patterns across tissues and species**
5

6 Lauren E. Blake^{1*}, Julien Roux^{1,2,3*}, Irene Hernando-Herraez⁴, Nicholas E. Banovich¹, Raquel
7 Garcia Perez⁴, Chiaowen Joyce Hsiao¹, Ittai Eres¹, Claudia Cuevas¹, Tomas Marques-Bonet⁴⁻⁷,
8 and Yoav Gilad^{1,8,‡}
9

- 10 1. University of Chicago, Department of Human Genetics, Chicago, IL.
11 2. Department of Biomedicine, University of Basel, Basel, Switzerland
12 3. Swiss Institute of Bioinformatics, Lausanne, Switzerland
13 4. Universitat Pompeu Fabra, Institute of Evolutionary Biology, Barcelona, Spain
14 5. Passeig de Lluís Companys, Catalan Institution of Research and Advanced Studies,
15 Barcelona, Spain.
16 6. Barcelona Institute of Science and Technology, Centre for Genomic Regulation.
17 7. Universitat Autònoma de Barcelona, Institut Català de Paleontologia Miquel Crusafont,
18 Barcelona, Spain.
19 8. University of Chicago, Department of Medicine, Chicago, IL.

20 *Authors contributed equally to this work.
21

22 **Contact**
23

24 ‡Correspondence should be addressed to Yoav Gilad (gilad@uchicago.edu)
25 Cummings Life Sciences Center, 920 E. 58th Street, CLSC 317
26 Chicago, IL, 60637, USA
27 Telephone: (773) 702-8507
28

29 Fax: (773) 834-8470
30 **Running title:** Gene regulatory patterns across primate tissues
31 **Key words:** genomics, gene expression, primate comparative genomics
31

32 **Abstract**

33 Previously published comparative functional genomic datasets from primates using
34 frozen tissue samples, including many datasets from our own group, were often collected and
35 analyzed using non-optimal study designs and analysis approaches. In addition, when samples
36 from multiple tissues were studied in a comparative framework, individual and tissue were
37 confounded. We designed a multi-tissue comparative study of gene expression and DNA
38 methylation in primates that minimizes confounding effects, by using a balanced design with
39 respect to species, tissues, and individuals. We also developed a comparative analysis pipeline
40 that minimizes biases due to sequence divergence. Thus, we present the most comprehensive
41 catalog of similarities and differences in gene expression and DNA methylation levels between
42 livers, kidneys, hearts, and lungs, in humans, chimpanzees, and rhesus macaques. We
43 estimate that overall, inter-species and inter-tissue differences in gene expression levels can
44 only modestly be accounted for by corresponding differences in promoter DNA methylation.
45 However, the expression pattern of genes with conserved inter-tissue expression differences
46 can be explained by corresponding inter-species methylation changes more often. Finally, we
47 show that genes whose tissue-specific regulatory patterns are consistent with the action of
48 natural selection are highly connected in both gene regulatory and protein-protein interaction
49 networks.

50

51 **Introduction**

52 Gene regulatory differences between humans and other primates are hypothesized to
53 underlie human-specific traits (King and Wilson 1975). Over the past decade, dozens of
54 comparative genomic studies focused on characterizing mRNA expression level differences
55 between primates in a large number of tissues (e.g., (Enard et al. 2002; Khaitovich et al. 2005;
56 Blekhman et al. 2008; Brawand et al. 2011; Barbash and Sakmar 2017)), typically focusing on

57 differences between humans and other primates. A few studies have also characterized inter-
58 primate differences in regulatory mechanisms and phenotypes other than gene expression
59 levels, such as DNA methylation levels, chromatin modifications and accessibility, and protein
60 expression levels (Cain et al. 2011; Pai et al. 2011; Hernando-Herraez et al. 2013; Ward et al.
61 2013; Stergachis et al. 2014; Zhou et al. 2014; Hernando-Herraez et al. 2015a; Hernando-
62 Herraez et al. 2015b; Villar et al. 2015). These studies often construct catalogs of gene
63 expression levels and other mechanisms. These catalogs have been useful to better understand
64 the evolutionary processes that led to adaptations in humans (Enard et al. 2002; Caceres et al.
65 2003; Karaman et al. 2003; Khaitovich et al. 2004a; Gilad et al. 2005; Khaitovich et al. 2005;
66 Lemos et al. 2005; Blekhman et al. 2008; Nowick et al. 2009; Babbitt et al. 2010; Blekhman et
67 al. 2010; Pai et al. 2011; Shibata et al. 2012; Capra et al. 2013; Khan et al. 2013) and ancestral
68 or derived phenotypes that may be relevant to human diseases (Cooper and Shendure 2011;
69 Romero et al. 2012).

70 One caveat that is shared among practically all comparative studies in primates is
71 related to difficulty in obtaining multiple tissue samples from the same individual. To date, there
72 have been no published comparative studies in primates that have analyzed multiple tissues
73 sampled from the same individuals across multiple species in a balanced design (Romero et al.
74 2012). As a result, regulatory differences between tissues are always confounded with
75 regulatory differences between individuals (Blekhman et al. 2008; Brawand et al. 2011; Pai et al.
76 2011; Chen et al. 2019). In turn, catalogs from these studies can not be used to compare tissue-
77 specific regulatory differences between species to inter-tissue differences in regulatory variation
78 within species (see Discussion in (Pai et al. 2011)).

79 Our group and others often use previously published catalogs of comparative data from
80 primates in our different studies. While we do not expect previously observed patterns to be
81 erroneous, we are aware that data on gene-specific inter-species regulatory differences, and
82 especially data that pertain to comparisons of divergence across tissues, may be inaccurate for

83 the reasons we discussed above. We thus designed the current study to produce a new
84 comprehensive catalog of comparative gene expression and DNA methylation data from
85 humans, chimpanzees, and rhesus macaques, attempting to minimize possible confounders.

86 The goal of our study is not to challenge previous conclusions or document specific
87 differences between the current and previous data. Instead, we aim to provide a new and more
88 accurate comparative catalog of inter-tissue and inter-species differences in gene regulation
89 between humans and other primates, with substantial sample and study design documentation.
90 Overall, we believe that this catalog can be useful for many future applications and can serve as
91 a new benchmark for regulatory divergence in primates.

92

93 **Results**

94

95 **Study design and data collection**

96 To comparatively study gene expression levels and DNA methylation patterns in
97 primates, we collected primary heart, kidney, liver, and lung tissue samples from four human,
98 four chimpanzee, and four rhesus macaque individuals (Figure 1A, Supplemental Table S1A).
99 From these 48 samples, we harvested RNA and DNA in parallel (see Methods). After confirming
100 that the RNA from all samples was of acceptable quality (Supplemental Fig. S1A; Supplemental
101 Table S1B), we performed RNA-sequencing to obtain estimates of gene expression levels.
102 Additional details about the donors, tissue samples, sample processing, and sequencing
103 information can be found in the Methods and Supplemental Table S1.

104 We estimated gene expression levels using an approach designed to prevent biases
105 driven by sequence divergence across species (similar to the approach of (Blake et al. 2018b)).
106 Briefly, we mapped RNA-sequencing reads to each species' respective genome. To compare
107 gene expression levels across species, we only calculated the number of reads mapping to
108 exons that can be classified as clear orthologs across all three species (Supplemental Table

109 S1B). We excluded data from genes that were lowly expressed in over half of the samples as
110 well as data from one human heart sample that was an obvious outlier, probably due to a
111 sample swap (Supplemental Fig. S2A-B). We normalized the distribution of gene expression
112 levels to remove systematic expression differences between species (maximizing the number of
113 genes with invariant expression levels across species corresponds to our null hypothesis; see
114 Methods). Through this process, we obtained TMM- and cyclic loess-normalized \log_2 counts per
115 million (CPM) values for 12,184 orthologous genes to be used in downstream analyses
116 (Supplemental Table S2).

117 Elements of study design, including sample processing, have been shown to impact
118 gene expression data (Gilad and Mizrahi-Man 2015). Consequently, we tested the relationship
119 between a large number of technical factors recorded throughout our experiments and the
120 biological variables of interest in our study, namely tissue and species (see Methods,
121 Supplemental Materials, Supplemental Table S3A-B). We found that there were no technical
122 confounders with tissue but two technical factors were confounded with species: time post-
123 mortem until collection and RNA extraction date (Supplemental Fig. 1B-1C). Due to the
124 opportunistic nature of sample collection, these confounders are practically impossible to avoid
125 in comparative studies of primates (especially apes). We discuss possible implications of these
126 confounders throughout the paper.

127

128 **Gene expression varies more across tissues than across species**

129 We first examined broad patterns in the gene expression data. A principal component
130 analysis (PCA) and a separate clustering analysis indicated that, as expected (Brawand et al.
131 2011; Barbosa-Morais et al. 2012; Merkin et al. 2012), the primary sources of gene expression
132 variation are tissue (Figure 1B, regression of PC1 by tissue = 0.81; $P < 10^{-14}$; regression of PC2
133 by tissue = 0.70; $P < 10^{-10}$; Supplemental Tables S1A-B and S3A-B; Supplemental Fig. S3),
134 followed by species (regression of PC2 by species = 0.27; $P < 10^{-3}$; Supplemental Tables S1A-B

135 and S3A-B). We then confirmed that, globally, gene expression levels across tissues from the
136 same individual are more highly correlated than gene expression levels across tissues from
137 different individuals (Supplemental Fig. S2C). This observation supports the intuitive notion that
138 collecting and analyzing multiple tissues from the same individual is highly desirable in
139 functional genomics studies.

140 We sought further explicit evidence that incorporating balanced collection of multiple
141 tissues from the same individuals is an effective study design. To do so, we used lung and heart
142 data from the GTEx Consortium (The GTEx Consortium 2017). We first identified differentially
143 expressed (DE) genes between lung and heart; we designated these classifications, which are
144 based on hundreds of samples, as the ‘truth’ (Supplemental Materials; Supplemental Table
145 S3E). Next, we repeatedly identified DE genes between lung and heart using GTEx data from
146 randomly chosen sets of just 4 samples from each tissue. We then compared the results to DE
147 genes identified from an equivalent analysis of sets of 4 samples from each tissue, in which the
148 tissue samples originated from the same donor. Compared to the designated ‘truth’, DE
149 analyses using data from tissue samples that are matched for donors result in a higher ratio of
150 true positives to false positives than analyses using tissue samples that are unmatched for
151 donors ($P = 0.03$; Supplemental Table S3F). Given the small number of false positives in both
152 datasets, study design is unlikely to impact large-scale, highly robust trends across species.
153 However, this study design choice is particularly important if one is interested in individual
154 genes (as demonstrated by an example in Figure 1C).

155

156 **Putatively functional tissue-specific gene expression patterns**

157 To analyze the pairwise regulatory differences across tissues and species, we used the
158 framework of a linear model (see Methods). We first identified (at FDR < 1%) 3,695 to 7,027
159 (depending on the comparison we considered) differences in gene expression levels between
160 tissues, within each species (Table 1; Supplemental Table S4A-C). Overall, the patterns of inter-

161 tissue differences in gene expression levels are similar in the three species, significantly more
162 so than expected by chance alone ($P < 10^{-16}$, hypergeometric distribution; Supplemental
163 Materials; Supplemental Table S5). A range of 17 – 26% of inter-tissue DE genes have
164 conserved inter-tissue expression patterns in all three species (Supplemental Table S5).
165 Regardless of species, we found the fewest inter-tissue DE genes when we contrasted liver and
166 kidney, and the largest number of DE genes between liver and either heart or lung (Table 1;
167 Supplemental Table S4B). Unfortunately, since our data were produced from bulk RNA-
168 sequencing, we were unable to determine the impact of cell composition on the number of inter-
169 tissue DE genes.

170 We used the same framework of linear modeling to identify gene expression differences
171 between species, within each tissue (Supplemental Table S4A). Depending on the tissue and
172 species we considered, we identified between 805 to 4,098 inter-species DE genes (at FDR =
173 1%; Table 1). As expected given the known phylogeny of the three species, within each tissue,
174 we classified far fewer DE genes between humans and chimpanzees than between either of
175 these species and rhesus macaques (Supplemental Table S4B).

176 It is a common notion that genes with tissue-specific expression patterns may underlie
177 tissue-specific functions. Previous catalogs of such patterns in primates were always
178 confounded by the effect of individual variation (because each tissue was sampled from a
179 different individual). To classify tissue-specific genes using our data, we focused on genes that
180 are either up-regulated or down-regulated in a single tissue relative to the other three tissues
181 (within one or more species). We define such genes as having a ‘tissue-specific’ expression
182 pattern, acknowledging that this definition may only be relevant in the context of the four tissues
183 we considered here.

184 Using this approach and considering the human data across all tissue comparisons, we
185 identified 5,284 genes with tissue-specific gene expression patterns (FDR 1%, Figure 2A-D). By
186 performing similar analyses using the chimpanzee and rhesus macaque data, we found that the

187 degree of conservation of tissue-specific expression patterns is higher than expected by chance
188 ($P < 10^{-16}$; Figure 2A-D). This observation is robust with respect to the statistical cutoffs we used
189 (Supplemental Table S6), indicating that many of these conserved tissue-specific regulatory
190 patterns are likely of functional significance.

191 To broadly analyze the biological function of genes with conserved tissue-specific
192 expression, we performed a Gene Ontology enrichment analysis (GO, see Supplemental
193 Materials). We found these genes are indeed highly enriched with functional annotations that
194 are relevant to the corresponding tissue (Supplemental Tables S7A-D, S8). For example, genes
195 with conserved heart-specific expression patterns were enriched in GO categories related to
196 muscle filament sliding (e.g. *ACTA1*, *MYL2*) and cardiac muscle contraction (e.g. *MYBPC3*,
197 *TNNI3*).

198
199 **Functional analysis of gene regulatory differences**
200

201 We sought further evidence that the classification of genes with conserved tissue-
202 specific expression patterns is meaningful. To do so, we considered transcription co-expression
203 networks (Stuart et al. 2003; Zhang and Horvath 2005) based on GTEx data from heart and
204 lung (Pierson et al. 2015). We found that genes with conserved tissue-specific expression
205 patterns are more likely to appear as nodes in the networks than genes without tissue-specific
206 expression patterns or genes whose tissue-specific expression patterns are not conserved ($P <$
207 10^{-5}). When we only considered genes that appear as nodes in the network, we found that
208 genes with conserved tissue-specific expression patterns are more likely to be classified as
209 hubs in the networks than genes without tissue-specific expression patterns or genes whose
210 tissue-specific expression patterns are not conserved ($P < 0.007$).

211 Motivated by these findings, we focused on gene expression patterns that are consistent
212 with the action of natural selection (as described in (Blekhman et al. 2008); see Supplemental
213 Materials and Supplemental Table S7E). We found that genes whose expression patterns are

214 consistent with the action of either stabilizing or directional selection (top 10%; Supplemental
215 Table S7F) have more interactions with other genes in the network than genes whose
216 expression patterns are not consistent with the action of natural selection (bottom 10%; $P < 0.05$
217 for all comparisons; Figure 2E). This observation is fairly robust with respect to percentile cutoff
218 (Supplemental Table 7F).

219 We repeated a similar analysis using protein–protein interaction data from the Human
220 Protein Atlas (Uhlen et al. 2015; Yu et al. 2015; Lindskog 2016; Thul and Lindskog 2018) in all
221 four tissues. We again found that genes whose expression patterns are consistent with
222 selection have more annotated protein-protein interactions ($P < 0.05$ in all 8 comparisons,
223 Figure 2F; Supplemental Table 7G). These interaction results suggest that functionally
224 important genes are carefully regulated. Furthermore, this tight regulation occurs at both the
225 gene expression and protein levels in primates.

226

227 Variation in DNA methylation across tissues and species

228 We used low-coverage whole genome bisulfite sequencing (BS-seq) to study DNA
229 methylation patterns in each sample. The bisulfite conversion reaction efficiency was higher
230 than 99.4% for all samples (Supplemental Table S1C). Following sequencing, we mapped the
231 high-quality BS-seq reads to *in silico* bisulfite-converted genomes of the corresponding species.
232 We measured DNA methylation levels in 12.5M to 22.9M CpG sites per sample, with a minimum
233 coverage of two sequencing reads per site (Supplemental Table S1C).

234 We estimated local methylation levels by smoothing the data across nearby CpG sites
235 (see Supplemental Materials; Supplemental Figs. S4-S6; (Hansen et al. 2012)). To facilitate a
236 comparison of methylation levels across species, we annotated 10.5M orthologous CpGs in the
237 human and chimpanzee genomes, as well as a smaller set of 2.4M orthologous CpGs in all
238 three primate genomes (Supplemental Table S1C-E). To identify differences in DNA methylation
239 levels between tissues and species we again employed a linear model framework (see

240 Methods, Figure 1D). Focusing on DNA methylation patterns across tissues within species, we
241 identified between 7,026 to 41,280 differentially methylated regions between tissues, within
242 species (T-DMRs; Table 2; Supplemental Table S9A; (Blake et al. 2018a)). Pairwise
243 comparisons between hearts and lungs showed the lowest number of T-DMRs, regardless of
244 species (7,026 in rhesus macaques, 8,524 in chimpanzees, 14,208 in humans), while
245 comparisons involving heart and liver showed the largest number of T-DMRs (22,561 in
246 humans, 28,767 in chimpanzee and 41,280 in rhesus macaques; Table 2). We found that
247 human T-DMRs overlapped genic and regulatory features significantly more than expected by
248 chance. In particular, there is an enrichment of T-DMRs in intergenic regions, introns, 5'UTRs,
249 3'UTRs, and active enhancers (as defined by (Andersson et al. 2014); $P < 0.04$ for all tests;
250 Supplemental Table S9B).

251 We found strong evidence for T-DMR conservation across all three species ($P < 10^{-16}$
252 across all comparisons; Supplemental Table S10A). Though this level of conservation is higher
253 than expected by chance, we recognize that in each tissue comparison we performed, we had
254 incomplete power to identify T-DMRs and so the true conservation of T-DMRs is expected to be
255 even higher. To compare T-DMRs across species more effectively, we considered DNA
256 methylation data from all T-DMR orthologous regions that were classified as such in at least one
257 species. When we performed hierarchical clustering using orthologous DNA methylation data
258 from these T-DMRs, the data clustered first by tissue, then by species (Supplemental Fig. S7).
259 This trend is robust with respect to the species used to initially locate T-DMRs (Supplemental
260 Fig. S8-S9). Thus, our results suggest that in general, inter-tissue DNA methylation differences
261 within a species tend to be conserved, consistent with the observations of previous studies
262 (Martin et al. 2011; Molaro et al. 2011; Pai et al. 2011; Hernando-Herraez et al. 2013;
263 Hernando-Herraez et al. 2015b).

264 We next focused specifically on tissue-specific DMRs, as these may contribute to tissue-
265 specific function. In contrast to differences in DNA methylation between any pair of tissues, a

266 tissue-specific DMR is defined as having a similar methylation level in three of the tissues we
267 considered, but a significantly different DNA methylation level in the remaining tissue. We found
268 that there were more DMRs specific to liver (3,278 to 11,433 DMRs depending on the species)
269 than to kidney (2,300 to 3,957 DMRs), heart (1,597 to 2,969 DMRs), or lung (453 to 5,018
270 DMRs, Figure 3A-D; Supplemental Table S10B). Tissue-specific DMRs are highly conserved
271 regardless of the comparisons we made ($P < 10^{-13}$ for all comparisons, at least 25% bp overlap
272 was required to be considered shared).

273 In all four tissues, over 59% of conserved DMRs are hypo-methylated in a tissue-specific
274 manner. We evaluated the overlap between tissue-specific DMRs and genomic regions marked
275 with H3K27ac, a mark often associated with active gene expression (The ENCODE Project
276 Consortium 2012). We found that conserved hypo-methylated tissue-specific DMRs were
277 annotated with H3K27ac more frequently than tissue-specific DMRs identified only in humans (P
278 < 0.001 , difference of proportions test; Supplemental Materials; Supplemental Table S10C). We
279 then asked about the potential impact of these conserved hypo-methylated tissue-specific
280 DMRs on the expression of nearby genes. We found that genes with the closest TSSs to
281 conserved tissue-specific DMRs are highly enriched with relevant functional annotations in
282 hearts and livers (the tissues with the largest numbers of conserved hypo-methylated tissue-
283 specific DMRs; Figure 3E-F, Supplemental Table S10D) (Supek et al. 2011). For example,
284 conserved heart-specific DMRs are closest to genes in cardiovascular-related pathways,
285 including ventricular cardiac muscle cell development, canonical Wnt signaling pathway, and the
286 MAPK7 cascade. Overall, these observations suggest that conserved tissue-specific DMRs are
287 likely to underlie tissue-specific gene regulation in primates.

288

289 **Inter-species differences in gene expression and DNA methylation levels**

290 Our comparative catalog can be used to identify DNA methylation differences that could
291 potentially explain gene expression differences across species and tissues. To do so, we first

292 identified 7,725 orthologous genes with expression data and corresponding promoter DNA
293 methylation data in humans and chimpanzees, and 4,155 orthologous genes with the same
294 information for all three species. We then determined to what extent divergence in DNA
295 methylation levels could potentially underlie inter-species differences in gene expression by
296 comparing the gene expression effect size associated with 'species' before and after accounting
297 for methylation levels. To determine significant effect size differences, we applied adaptive
298 shrinkage (Stephens 2017) – a flexible empirical Bayes approach for estimating false discovery
299 rate (see Methods). We note that this mediation approach does not consider the possibility that
300 a third, unobserved event, may be causally responsible for both the DNA methylation and
301 expression patterns.

302 Considering DE genes between humans and chimpanzees (in at least one tissue), we
303 found that between 11% and 25% of genes (depending on tissue) showed a difference in the
304 effect of species on gene expression levels once average promoter methylation levels were
305 accounted for (significant difference in effect size classified at FSR 5%, represented by red in
306 Supplemental Fig. S10; Supplemental Table S11A). As a control analysis, we considered only
307 the genes that were not originally classified as DE between humans and chimpanzees, and
308 found that the difference in the effect size of species on gene expression levels was reduced in
309 less than 1% of genes once DNA methylation was accounted for (FSR 5%, Supplemental Fig.
310 S10; Supplemental Table S11A); thus, our approach is well calibrated.

311 We applied the same approach to the human and rhesus macaque data, and found that
312 the percentage of genes for which gene expression differences could potentially be explained
313 by DNA methylation differences ranges from 21% in the lung to 40% in the liver (Supplemental
314 Fig. S11; Supplemental Table S11B). This observation may reflect the more extreme gene
315 expression differences between humans and rhesus macaques than between humans and
316 chimpanzees (prior to accounting for DNA methylation levels, $P < 0.003$ in all tissues, *t*-test
317 comparing the absolute values of the effect sizes for both groups of DE genes).

318 Next, we examined the genes in which DNA methylation differences may underlie inter-
319 tissue gene expression differences (example in Figure 4A-C). Using adaptive shrinkage, we
320 found that 9% to 17% of inter-tissue gene expression differences could potentially be explained
321 by DNA methylation differences across tissues (FSR 5%). When we performed the control
322 analysis and considered only data from genes that were not DE between tissues, less than 1%
323 of effect sizes differed once we accounted for the DNA methylation data (Figure 4F;
324 Supplemental Table S11C-E).

325 Finally, we focused on regulatory patterns that are most likely to be functional; namely,
326 conserved inter-tissue gene regulatory differences. These differences were more likely to be
327 explained by variation in DNA methylation levels than non-conserved inter-tissue gene
328 expression differences (minimum difference is 7%, $P < 0.005$ for all comparisons; at FDR < 5%
329 and FSR < 5%; Figure 4D-4E, 4H; Supplemental Table S11C-E). This observation is robust with
330 respect to the FDR and FSR cutoff used (Supplemental Table S11C-E). Indeed, the correlation
331 between DNA methylation and gene expression data is higher for genes with conserved inter-
332 tissue expression patterns compared to genes whose expression patterns are not conserved
333 (Figures 1E-F).

334 One way to maintain conserved inter-tissue gene expression differences could be
335 through DNA methylation level differences. We compared the genes whose variation in inter-
336 tissue gene expression can potentially be explained by variation in DNA methylation levels
337 (assuming no independent effect of an unobserved factor) to all genes with conserved inter-
338 tissue expression differences. We found that these genes are enriched for 'essential tissue
339 functions' (Supplemental Table S11F). For example, the heart genes are enriched for cardiac
340 and smooth muscle contraction, whereas those in liver are enriched for regulation of cholesterol
341 transport and hormone secretion (Figure 4G; Supplemental Table S11F). These observations
342 suggest that DNA methylation levels may mark or even drive differences in the expression
343 levels in functionally relevant genes.

344
345
346

Discussion

347
348 We designed a comparative study of gene regulation in humans, chimpanzees, and rhesus
349 macaques that minimized confounding effects and bias. Consistent with previous studies, we
350 found a high degree of conservation in gene expression levels when we considered the same
351 tissue across species (Barlow 1993; Brawand et al. 2011; Sharp et al. 2011; Lin et al. 2014;
352 Gallego Romero et al. 2015a). We also found evidence for conservation of tissue-specific
353 DMRs. Our observations are qualitatively consistent with those of previous studies that mostly
354 used microarrays to measure DNA methylation levels (Pai et al. 2011; Hernando-Herraez et al.
355 2015b; Tsankov et al. 2015), however, the high resolution of our BS-seq data allowed us to
356 examine a much larger number of CpG sites. Thus, we were able to show that while DNA
357 methylation can potentially explain a modest proportion of expression differences between
358 tissues (Pai et al. 2011), it is more likely to impact conserved tissue-specific gene expression
359 levels.

360 We created and made available the most comprehensive, and likely most accurate
361 comparative catalog of gene expression and methylation levels in humans, chimpanzees, and
362 rhesus macaques. Comparative functional genomic studies in primates, including from our own
363 lab, often are not designed to test specific hypotheses. Rather, many of these comparative
364 genome-scale studies aim to build catalogs of similarities and differences in gene regulation
365 between humans and other primates. These catalogs have been shown to be quite useful; for
366 example, they can be used to identify inter-species regulatory changes that have likely evolved
367 under natural selection (Enard et al. 2002; Caceres et al. 2003; Karaman et al. 2003; Khaitovich
368 et al. 2004a; Gilad et al. 2005; Khaitovich et al. 2005; Lemos et al. 2005; Blekhman et al. 2008;
369 Nowick et al. 2009; Babbitt et al. 2010; Blekhman et al. 2010a; Pai et al. 2011; Shibata et al.
370 2012; Capra et al. 2013; Khan et al. 2013), and thereby help us better understand the

371 evolutionary processes that led to adaptations in humans. These catalogs are also used to
372 establish informed models of the relative importance of changes in different molecular
373 mechanisms to regulatory evolution (Khaitovich et al. 2004b; Warnefors and Eyre-Walker 2012),
374 and to inform us about ancestral or derived phenotypes that may be relevant to human diseases
375 (Cooper and Shendure 2011; Gallego Romero et al. 2015a). Ultimately, comparative catalogs of
376 gene regulatory phenotypes are used to develop and test specific hypotheses regarding the
377 connection between inter-species regulatory changes and physiological, anatomical, and
378 cognitive phenotypic differences between species.

379 In this study, we used a comparative catalog to identify species-specific and, in particular,
380 tissue-specific regulatory patterns, as these genes are often drug targets (Dezso et al. 2008)
381 and are likely important for the evolution of human traits (Blekhman et al. 2008). We showed
382 that genes with conserved tissue-specific regulatory patterns have more regulatory interactions
383 and protein-protein interactions than genes whose regulatory patterns are not conserved or are
384 not tissue-specific. These patterns became even more pronounced when we focused on genes
385 whose expression patterns are consistent with the action of natural selection. Put together,
386 these observations consistently support the inference that when genes perform an important
387 function that needs to be carefully regulated, evolution can act on multiple levels of the
388 regulatory cascade in primates.

389 Focusing on species-specific patterns of tissue-specific gene regulation, our observations
390 can help formulate specific functional hypotheses regarding human-specific adaptations. For
391 example, genes with tissue-specific gene regulation identified only in humans are enriched for
392 GO pathways that may contribute to human-specific features, including sodium ion import
393 across the plasma membrane in kidneys (e.g. *SLC9A3* and *TRPM4*), the glycogen biosynthetic
394 process in livers (e.g. *PGM1* and *AKT1*), and paraxial mesoderm morphogenesis in lungs (e.g.
395 *MST1R* and *MAP9*).

396

397 **Consideration of study design and record keeping**

398 Regardless of the model system used and the types of data that are collected, study
399 design is critical. Perhaps because comparative studies in primates typically rely on
400 opportunistic sample collection, there are not recognized standards for study design that are
401 kept and consistently reported in most existing studies (including many earlier studies from our
402 group). We thus believe that it is worthwhile to explicitly discuss a few important considerations
403 regarding study design and the recording of meta-data.

404 Without a balanced study design, it would have been impossible to independently
405 estimate the effects of individual, tissue, and species on our data. Because the sources of
406 confounding factors are difficult to predict in advance, we strongly recommend that samples are
407 collected using a balanced design with respect to as many parameters as possible. These
408 include the distribution of tissue samples per individual, the number of individuals from each
409 species, sex, age range, cause of death and collection time (in the case of post-mortem
410 tissues), or sample collection and cell culturing (in the case of iPSC-based models). All steps of
411 sample processing (RNA extraction, library preparation, etc.) should be done in batches that are
412 randomized or balanced with respect to species, tissue, and any other variables of interest.

413 Most importantly, all sample processing steps should be recorded in a sample history file
414 that includes anything that happened to any sample. We have documented many of these steps
415 in Supplemental Tables S1A-E. This documentation can help provide evidence that a
416 phenomenon is driven by biological rather than technical factors. It may also benefit future
417 studies by facilitating effective meta-analysis of multiple datasets, which would help to address
418 the problems of tissue availability and small sample sizes. We believe that, moving forward, it
419 should be a requirement that these meta-data are available with every published comparative
420 genomic dataset.

421 Methods**422 Sample description**

423 We collected heart (left ventricle), kidney (cortex), liver and lung tissues from four
424 individual donors in human (*Homo sapiens*, all of reported Caucasian ethnicity), chimpanzee
425 (*Pan troglodytes*), and Indian rhesus macaque (*Macaca mulatta*), for a total of 48 samples (3
426 species * 4 tissues * 4 individuals; Figure 1A). The choice of these tissues was guided by their
427 relative homogeneity with respect to cellular composition (e.g. (Balashova and Abdulkadyrov
428 1984)), which does not change substantially across primate species. In contrast, other tissues,
429 such as brain subparts, differ substantially in cellular composition across primates (Brodal
430 1983), which could potentially confound the analyses.

431 Human samples were obtained from the National Disease Research Interchange (IRB
432 protocol #14378B). Non-human samples were obtained from several sources, including the
433 Yerkes primate center and the Southwest Foundation for Biomedical Research, under IACUC
434 protocol 71619. When possible, samples were collected from adult individuals whose cause of
435 death was unrelated to the tissues studied.

436

437 RNA library preparation and sequencing

438 In total, we prepared 48 unstranded RNA-sequencing libraries as previously described
439 (Marioni et al. 2008; Blekhman et al. 2010). Twenty-four barcoded adapters were used to
440 multiplex different samples on two pools of libraries. RNA-sequencing libraries were sequenced
441 on 26 lanes on 4 different flow-cells on an Illumina HiSeq 2500 sequencer in either the Gilad lab
442 or at the University of Chicago Genomics Facility (50bp single end reads; Supplemental
443 Materials; Supplemental Table S1B).

444

445 Quantifying the number of RNA-seq reads from orthologous genes

446 We used FastQC (version 0.10.0; <http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>)
447 to generate read quality report and TrimGalore (version 0.2.8;
448 http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), a wrapper based on Cutadapt
449 (version 1.2.1)(Martin 2011), to trim adaptor sequences from RNA-seq reads. We trimmed using
450 a stringency of 3. To cut the low-quality ends of reads, we used a quality threshold (Phred
451 score) of 20. Reads shorter than 20 nt after trimming were eliminated before mapping
452 (Supplemental Table S1B).

453 For each sample, we used TopHat2 (version 2.0.8b) (Kim et al. 2013) to map the reads
454 to the correct species' genome: human reads to the hg19 genome, chimpanzee reads to the
455 panTro3 genome, and rhesus macaque reads to the rheMac2 genome (Supplemental
456 Materials). Expression level estimates may be biased across the species due to factors such as
457 mRNA transcript size and different genome annotation qualities. To circumvent these issues, we
458 only retained reads that mapped to a set of 30,030 Ensembl gene orthologous meta-exons
459 available for each of the 3 genomes, as described and used previously (Blekhman et al. 2010;
460 Blekhman 2012; Gallego Romero et al. 2015b). We defined the number of reads mapped to
461 orthologous genes as the sum of the reads mapped to the orthologous meta-exons of each
462 gene. We quantified gene expression levels using the program *coverageBed* from the
463 BEDTools suite and then performed TMM and cyclic loess normalization (Supplemental
464 Materials).

465 As all 3 releases of GTEx use hg19 and only a fraction of the 1000 Genomes Project
466 data are available in GRCh38 coordinates, we also used hg19. However, to demonstrate that
467 the results we report would not change much if we used the GRCh38 build, we leveraged the
468 fact that differential expression analysis compares gene expression levels from groups of
469 samples (e.g. human liver samples to human lung samples). Therefore, we compared the ranks
470 of the normalized gene expression levels in the 15 human samples mapped using hg19 to the
471 same samples mapped to GRCh38. The correlations of these ranks were extremely high

472 (median Pearson's correlation = 0.96). These strong correlations suggest that our general
473 conclusions—and indeed, many genes we identified as DE—would remain if we had used the
474 GRCh38 build.

475

476 **Analysis of technical variables**

477 To assess whether the study's biological variables of interest—tissue and species-- were
478 confounded with the study's recorded sample and technical variables, we used an approach
479 described in (Blake et al. 2018b).

480 For the 12 RNA-seq related technical variables that were the most highly correlated with
481 tissue or species, we assessed which technical variables constitute the “best set” of
482 independent variables to be included in a linear model for gene expression levels. Because of
483 the partial correlations between the variables, we applied lasso regression using the package
484 “glmnet” (Friedman et al. 2010). Before performing the analysis, we also protected our variables
485 of interest, tissue and species, in the model for each gene. We summarized each technical
486 variable’s influence across the genes by counting the number of times each technical variable
487 was included in the “best set” of the gene models. We found that none of the technical variables
488 appeared in more than 25% of the best sets (i.e. more than 25% of the gene models).
489 Therefore, we chose not to include these technical variables in our model for testing differential
490 expression.

491 Finally, during our analysis of technical factors, we discovered that RNA extraction date
492 was confounded with species. In 2012, we extracted RNA from the chimpanzee samples on
493 March 8, from the human samples on three days between March 12-29, and from the rhesus
494 samples on March 6. To test the relationship between the date of RNA extraction and gene
495 expression PCs in humans, we performed individual linear models on PCs 1-5 using RNA
496 extraction date as a predictor. None of the models were statistically significant at FDR 10%,

497 suggesting that tissue type is more highly associated with gene expression levels than RNA
498 extraction date.

499

500 **Differential expression analysis using a linear model-based framework**

501 To perform differential expression analysis, we used the same approach as in (Blake et
502 al. 2018b). We applied a linear model-based empirical Bayes method (Smyth 2004; Smyth et al.
503 2005) that accounts for the mean-variance relationship of the RNA-sequencing read counts,
504 using weights specific to both genes and samples (Law et al. 2014).

505 To be considered a “tissue-specific DE gene” under our stringent definition, the gene
506 must be in the same direction and statistically significant in all pairwise comparisons including
507 the given tissue but not significant in any comparison without that tissue. For example, for a
508 gene to be classified as having heart-specific upregulation in a given species, the gene needed
509 to be upregulated (a significant, positive effect size) in heart versus liver, heart versus lung,
510 heart versus kidney, but not significantly different between the liver versus lung, liver versus
511 kidney, or kidney versus lung, in the same species. Under the more lenient definition of tissue-
512 specific DE genes, we compared the gene expression level of one tissue to the mean of the
513 other three tissues. To do so, we grouped the three tissues together and again used the
514 limma+voom framework to identify significant differences in one tissue versus the group of the
515 other tissues.

516 To identify inter-species differences in gene expression patterns across tissues within
517 species (tissue-by-species interactions), we used the limma+voom framework and looked for
518 the significance of tissue-by-group interactions. In one analysis, the groups were great ape
519 versus rhesus macaque and in another analysis, the groups were human versus non-human
520 primates. To minimize the number of interactions, we compared one tissue relative to a group of
521 the other 3 tissues (e.g. great ape versus rhesus macaque heart versus non-heart). Significant
522 tissue-specific interactions were detected using the adaptive shrinkage method, *ashr* (Stephens

523 2017). Specifically, for each test, we input the regression estimates from *limma* to *ashr*:
524 regression coefficients, posterior standard errors, and posterior degrees of freedom. We used
525 the default settings in *ashr* to calculate the shrunken regression coefficients (called the
526 “posterior mean” in *ashr*), false discovery rate (FDR, also known as q-value), and false sign rate
527 (FSR, also known as s-value: the probability that sign of the estimated effect size is wrong in
528 either direction). We assigned directionality based on the sign of the posterior mean and
529 determined significance based on the FSR.

530

531 **The impact of matched tissue samples on DE results**

532 To determine the impact of matched tissue samples on DE results, we compared inter-
533 tissue DE analysis results when using tissues from the same or different individuals in GTEx v7
534 data (The GTEx Consortium 2017). We first subset the GTEx raw gene expression count data
535 to only individuals for which there was gene count information in the heart and lung tissues, for
536 genes included in all 3 tissues. (There were the most GTEx samples in heart and lung; we
537 decided to focus on these samples). Furthermore, to minimize the number of covariates needed
538 in the linear model, we decided to only analyze individuals of the same sex and whose samples
539 were sequenced on the same platform (sex = 1 and platform = 1 from the GTEx
540 documentation). We then normalized the data and performed DE analysis using a voom+limma
541 pipeline. In the linear model, we included tissue and 3 GTEx-provided covariates (covariates 1
542 and 2 and inferred covariate 1 from the covariate file for each tissue) as fixed effects and
543 individual as a random effect. We chose to renormalize the raw counts data rather than use the
544 normalized counts from GTEx because the voom+limma pipeline requires raw counts to assign
545 voom weights. We considered the output of the inter-tissue DE analysis for all individuals (DE
546 versus non-DE genes at FDR 5%) as the “ground truth”. To evaluate the impact of our study
547 design, we then subset the gene expression information to the individuals for which there is
548 information in all 3 tissues. We obtained gene expression level information from 4 randomly

549 selected individuals and used the voom+limma pipeline to identify inter-tissue DE genes. Next,
550 we compared the list of DE genes from this analysis to the “ground truth” list. We performed this
551 downsampling procedure for tissues from the same 4 individuals as well as different 4
552 individuals 10 times each and compared the number of true and false positives from the tests.
553 For the analysis with 8 different individuals, there were no repeated individuals, so we did not
554 use the “duplicateCorrelation” function in voom.

555

556 **BS-seq library preparation, sequencing, and mapping**

557 We prepared a total of 48 whole-genome BS-seq libraries from extracted DNA as
558 previously described (Tung et al. 2012; Banovich et al. 2014). We aligned the trimmed reads to
559 the human (hg19, February 2009), chimpanzee (panTro3, October 2010), or rhesus macaque
560 (rheMac2, January 2006) genomes, and to the lambda phage genome using the Bismark
561 aligner (version 0.8.1)(Krueger and Andrews 2011).

562 We estimated the percentage of methylation at an individual cytosine site by the ratio of
563 the number of cytosines (unconverted) found in mapped reads at this position, to the total
564 number of reads covering this position (sequenced as cytosine or thymine, i.e., converted or
565 unconverted) using the methylation extractor tool from Bismark (version 0.8.1). We additionally
566 collapsed information from both DNA strands (because CpG methylation status is highly
567 symmetrical on opposite strands (Lister et al. 2009)) to achieve better precision in methylation
568 estimates across the genome.

569 To obtain CpGs that mapped to multiple species, the chimpanzee and rhesus macaque
570 CpG sites were mapped to human coordinates (hg19) using chain files from
571 <ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/liftOver/> and the liftOver tool from the UCSC
572 Genome Browser (Karolchik et al. 2014). These files had previously been filtered for paralogous
573 regions and repeats, but we also removed positions that were not remapped to their original
574 position when we mapped from human back to their original genome. Chimpanzee and rhesus

575 macaque CpG sites were mapped to human, even if their orthologous positions were not a CpG
576 site in human.

577

578 **Identifying differentially methylated regions (DMRs)**

579 We were interested in identifying regions exhibiting consistent differences between pairs
580 of tissues or pairs of species, taking biological variation into account. To identify DMRs we used
581 the linear model-based framework in the Bioconductor package *bsseq* (version 0.10.0)(Hansen
582 et al. 2012). For a given pairwise comparison (e.g., human liver vs. human heart), the *bsseq*
583 package produces a signal-to-noise statistic for each CpG site similar to a *t*-test statistic,
584 assuming that methylation levels in each condition have equal variance. As recommended by
585 the authors of the package, we used a low-frequency mean correction to improve the marginal
586 distribution of the *t*-statistics. Similar to previous studies using this methodology, a *t*-statistic
587 cutoff of $-4.6, 4.6$ was used for significance, (Hansen et al. 2011; Hansen et al. 2014). DMRs
588 were defined as regions containing at least three consecutive significant CpGs, an average
589 methylation difference of 10% between conditions, and at least one CpG every 300 bp (Hansen
590 et al. 2012). We used BEDTools (version 2.26.0) (Quinlan and Hall 2010) to calculate the
591 number of overlapping DMRs across tissues and/or species (Blake et al. 2018a). We required
592 overlapping DMRs to have a minimum base pair overlap of at least 25%, unless otherwise
593 stated.

594 To be considered a tissue-specific DMR, the region was required to be a significant
595 tissue DMR (T-DMR) in 1 tissue compared to the other 3 tissues pairwise (in the same direction)
596 but not a significant T-DMR across any of the other 3 tissues in pairwise comparisons. We again
597 used BEDTools to ensure a minimum base pair overlap of 25%. Once the tissue-specific DMRs
598 were identified within a species, we then classified them as species-specific or conserved. To
599 be considered conserved (across humans and chimpanzees or all three species), the tissue-

600 specific DMR had to be significant in all species in the comparison and have a minimum base
601 pair overlap of the T-DMR of at least 25%.

602

603 **Calculating the average methylation levels of conserved promoters**

604 To calculate the DNA methylation levels of orthologous CpGs around the transcription
605 start site (TSS) of orthologous genes, we first had to determine the orthologous TSSs. We
606 began with the 12,184 orthologous genes in our RNA-sequencing analysis. Of these, we found
607 that 11,131 of these orthologous genes had an hg19 RefSeq TSS annotation
608 ([https://sourceforge.net/projects/seqminer/files/Reference%20coordinate/refGene_hg19_TSS.b
609 ed/download](https://sourceforge.net/projects/seqminer/files/Reference%20coordinate/refGene_hg19_TSS.bed/download)). We used liftOver to find orthologous sites in the chimpanzees and rhesus
610 macaque genomes in 9,682 of those 11,131 genes. We then determined which of the hg19
611 RefSeq TSS annotations were closest to the first hg19 orthologous exon, and repeated this
612 process with the other two species and their respective genomes. We found that 9,604 out of
613 9,682 of the closest TSS annotations in humans at the same liftOver coordinates in the other
614 two species. We then calculated the distance between the first orthologous exon to the TSS site
615 in all 3 species individually. To minimize this difference between the 3 species, we filtered all
616 genes with a maximum distance difference across the species of larger than 2,500 bp (for
617 reference, the 75th percentile of the maximum difference in distance was 2,078 bp). 7,263
618 autosomal genes remained after this filtering step. 4,155 genes had at least 2 orthologous CpGs
619 250 bp upstream and 250 bp downstream of the orthologous TSS. We chose a 250 bp window
620 around the TSS based on DNA methylation levels around the promoter in (Banovich et al. 2014)
621 and calculated the average of orthologous CpGs within this window for the 4,155 genes. Using
622 the same method but in humans and chimpanzees only, we found and calculated the average of
623 orthologous CpGs within this window for 7,725 genes.

624

625 **Joint analysis of promoter DNA methylation and gene expression levels**

626 To determine whether DNA methylation may underlie inter-species differences in gene
627 expression levels, we used a joint analysis method as described below. For each gene, we
628 analyzed the gene expression levels, along with the accompanying average methylation levels
629 250 bp upstream and downstream of the TSS (found above). For a given tissue, we first
630 determined the effect of species on gene expression levels using a linear model, with species
631 and RIN score as fixed effects (Model 1). Next, we parameterized a linear model attempting to
632 predict expression levels exclusively from methylation levels. We refer to these residuals as
633 “methylation-corrected” gene expression values. We then used these values to again determine
634 the effect of species, this time on gene expression levels “corrected” for methylation, using a
635 linear model with species and RIN score as fixed effects (Model 2). To determine the
636 contribution of DNA methylation levels to inter-species differences in gene expression, we
637 computed the difference in the species effect size between Model 1 and Model 2 for each gene,
638 as well as the standard error of the difference. Large effect size differences between Models 1
639 and 2 for a given gene suggest that methylation status may be a significant driver of DE. To
640 assess the significance of this difference, we used adaptive shrinkage (*ashr*) (Stephens 2017) to
641 compute the posterior mean of the differences in the effect sizes, using *vashr*, with the degrees
642 of freedom equal to the number of samples in the linear model minus 2. The shrunken variances
643 from *vashr* were used in the *ashr* posterior mean computation. From this procedure, we
644 obtained the number of genes where species has a significant difference in effect sizes before
645 and after regressing out methylation. We assessed significance using the s-value statistic (FSR,
646 (Stephens 2017)). Using the s-values, rather than the q-values, not only takes significance into
647 account but also has the added benefit of assessing our confidence in the direction of the effect.

648 We performed the above analysis separately for inter-species DE genes and non-DE
649 genes, and in each tissue individually. We identified inter-species DE genes in our tissue of
650 interest as those with a significant species term in the model of species and RIN score as fixed
651 effects. We assessed significance of DE genes at FDR 5%, unless otherwise noted.

652 We also applied the same analysis framework to determine whether DNA methylation
653 may underlie inter-tissue differences in gene expression levels. For the inter-tissue DE genes
654 and non-DE genes, we replaced “species” with “tissue” as a fixed effect in models 1 and 2. We
655 assessed significance with various FDR and FSR thresholds, as specified in the text.

656

657 **Data Access**

658 All raw and processed sequencing data generated in this study have been submitted to
659 the NCBI’s Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo>) using GEO
660 Series accession number GSE112356. Data and scripts used in this paper are available at
661 https://github.com/Lauren-Blake/Reg_Evo_Primates and in the Supplemental Code. The results
662 of our scripts can be viewed at https://lauren-blake.github.io/Regulatory_Evol/analysis/.

663

664 **Acknowledgments**

665 Members of the Gilad, Marques-Bonet, Robinson-Rechavi, Stephens, and Pritchard labs
666 provided helpful discussions and comments on the manuscript. In particular, Matthew Stephens
667 provided guidance on integrating the DNA methylation and gene expression data, and Michelle
668 Ward and Natalia Gonzales provided helpful comments on a draft of the manuscript. We thank
669 Kasper Hansen, Jenny Tung, and Luis Barreiro for discussions regarding multispecies
670 methylation analysis. We acknowledge Athma Pai for sharing a methylation protocol, Bryce van
671 de Geijn for help with the WASP pipeline, Charlotte Soneson, Jacob Degner, and Unjin Lee. We
672 also thank the Yerkes Primate Center and Southwest Foundation for Biomedical Research,
673 Anne Stone and Jéssica Hernández Rodríguez for providing and/or helping to ship the tissue
674 samples.

675 L.E.B. was supported by the National Science Foundation Graduate Research
676 Fellowship (DGE-1144082). Additionally, L.E.B. and I.E. were funded by the Genetics and Gene

677 Regulation Training Grant (T32 GM07197). J.R. was funded by a Swiss NSF postdoc mobility
678 fellowship (PBLAP3-134342) and the Marie Curie International Outgoing Fellowship
679 PRIMATE_REG_EVOL. This project was funded in part by the ORIP/OD P51OD011132 grant.
680 T.M.B. is supported by BFU2017-86471-P (MINECO/FEDER, UE), U01 MH106874 grant,
681 Howard Hughes International Early Career, Obra Social "La Caixa" and Secretaria
682 d'Universitats i Recerca and CERCA Programme del Departament d'Economia i Coneixement
683 de la Generalitat de Catalunya. The content presented in this article is solely the responsibility
684 of the authors and does not necessarily reflect the official views of the funders. The funders had
685 no role in study design, data collection and analysis, decision to publish, or preparation of the
686 manuscript.

687

688 **Disclosure Declaration**689
690 The authors declare no competing financial interests.

691

692

693 **Figure Legend**

694

695 **Figure 1. Surveying gene expression and DNA methylation in diverse tissues across**
696 **primates.** (A) Study design. (B) Principal components analysis (PCA) of gene expression levels
697 in 47 samples. (C) Normalized gene expression (quantile-transformed RPKMs) from 4 donors in
698 the GTEx heart collection ("heart same individuals") compared to the lungs from different ("lung
699 different individuals") and the same donors' lungs ("lung same individuals") in AC020922.1. (D)
700 PCA of average methylation levels 250 bp upstream and downstream in 47 samples. (E)
701 Density function of the correlation between gene expression and DNA methylation levels in
702 human-chimpanzee orthologous genes. (F) Density function of the correlation between gene
703 expression and DNA methylation levels in genes orthologous across humans and rhesus
704 macaques.

705

706 **Figure 2. Tissue-specific DE genes (FDR = 0.01).** The number of conserved tissue-specific
707 DE genes across all three species, in the (A) heart, (B) kidney, (C) liver, and (D) lung, is greater
708 than the number expected by chance. In each tissue, genes with tissue-specific regulatory
709 patterns that are consistent with the action of natural selection ("top") are more likely to appear
710 in (E) gene co-expression networks and (F) have an increased number of protein-protein
711 interactions those that are less consistent with the action of natural selection ("bottom"). * $P <$
712 0.05, *** $P < 0.001$. Note: The x-axes of 2E and 2F are cut off at 80 interactions for readability.
713 In both cases, <5% of the data points are beyond this cutoff.

714

715 **Figure 3. Tissue-specific DMRs (FDR = 0.01).** The number of conserved tissue-specific DMRs
716 in the (A) heart, (B) kidney, (C) liver, and (D) lung is greater than expected by chance. Genes
717 with the closest TSSs to conserved tissue-specific DMRs are enriched for relevant functional
718 annotations in (E) hearts and (F) livers.

719
720 **Figure 4. Inter-tissue DNA methylation and gene expression levels (FDR = 0.05 and FSR =**
721 **0.05). A-C.** A representative example of the *PRKACA* gene in which the variation of methylation
722 levels (A) may explain the differences in gene expression levels between human heart and
723 kidney. (C) The residuals of normalized gene expression levels after regressing out methylation
724 levels. **D-G.** Next, we compared the tissue effect sizes before and after controlling for DNA
725 methylation levels in inter-tissue DE and non-DE genes, separately. Genes in red are significant
726 at s-value < 0.05. Effect size differences in (D) conserved DE genes in human heart relative to
727 human kidney, (E) non-conserved DE genes in human heart relative to human kidney and (F)
728 non-DE genes in human heart and human kidney. (G) The conserved DE genes are enriched
729 for heart-related function. (H) Variation in DNA methylation is more likely to explain variation in
730 conserved DE genes than non-conserved DE genes (DE in the human tissues listed, but not in
731 the same tissues in chimpanzees).

732
733 **Table Legend**

734
735 **Table 1: Pairwise differentially expressed (DE) genes at FDR 1%.**

736
737 **Table 2: Pairwise differentially methylated regions (DMRs) in autosomal chromosomes**
738 **(cutoff recommended by (Hansen et al. 2012), percentage out of total pairwise methylated**
739 **regions)**

740
741
742

743 **References**

- 744
- 745 Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X,
746 Schmidl C, Suzuki T et al. 2014. An atlas of active enhancers across human cell types
747 and tissues. *Nature* **507**: 455-461.
- 748 Babbitt CC, Fedrigo O, Pfefferle AD, Boyle AP, Horvath JE, Furey TS, Wray GA. 2010. Both
749 noncoding and protein-coding RNAs contribute to gene expression evolution in the
750 primate brain. *Genome Biol Evol* **2**: 67-79.
- 751 Balashova V, Abdulkadyrov K. 1984. Cellular composition of hemopoietic tissue of the liver and
752 spleen in the human fetus. *Arkh Anat* **86**: 80.
- 753 Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, Roux J, Pritchard
754 JK, Gilad Y. 2014. Methylation QTLs Are Associated with Coordinated Changes in
755 Transcription Factor Binding, Histone Modifications, and Gene Expression Levels. *PLoS*
756 *Genet* **10**: e1004663. doi: 10.1371/journal.pgen.1004663.
- 757 Barbash S, Sakmar TP. 2017. Brain gene expression signature on primate genomic sequence
758 evolution. *Sci Rep* **7**: 17329.
- 759 Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter
760 C, Watt S, Colak R et al. 2012. The evolutionary landscape of alternative splicing in
761 vertebrate species. *Science* **338**: 1587-1593.
- 762 Barlow DP. 1993. Methylation and imprinting: from host defense to gene regulation? *Science*
763 **260**: 309-310.
- 764 Blake LE, Roux J, Gilad Y. 2018a. Pairwise DMRs for, "A comparison of gene expression and
765 DNA methylation patterns across tissues and species". doi: 10.5281/zenodo.1495829.
- 766 Blake LE, Thomas SM, Blischak JD, Hsiao CJ, Chavarria C, Myrthil M, Gilad Y, Pavlovic BJ.
767 2018b. A comparative study of endoderm differentiation in humans and chimpanzees.
768 *Genome Biol* **19**: 162.
- 769 Blekhman R. 2012. A database of orthologous exons in primates for comparative analysis of
770 RNA-seq data. *Nat Preced* doi: hdl:10101/npre.2012.7054.1.
- 771 Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. 2010. Sex-specific and lineage-
772 specific alternative splicing in primates. *Genome Res* **20**: 180-189.
- 773 Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y. 2008. Gene regulation in primates
774 evolves under tissue-specific selection pressures. *PLoS Genet* **4**: e1000271. doi:
775 10.1371/journal.pgen.1000271.
- 776 Brawand D, Soumillon M, Neacsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A,
777 Aximu-Petri A, Kircher M et al. 2011. The evolution of gene expression levels in
778 mammalian organs. *Nature* **478**: 343-348.
- 779 Brodal A. 1983. The perihypoglossal nuclei in the macaque monkey and the chimpanzee. *J
780 Comp Neurol* **218**: 257-269.
- 781 Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss
782 TM, Barlow C. 2003. Elevated gene expression levels distinguish human from non-
783 human primate brains. *Proc Natl Acad Sci U S A* **100**: 13030-13035.
- 784 Cain CE, Blekhman R, Marioni JC, Gilad Y. 2011. Gene expression differences among primates
785 are associated with changes in a histone epigenetic modification. *Genetics* **187**: 1225-
786 1234.
- 787 Capra JA, Erwin GD, McKinsey G, Rubenstein JL, Pollard KS. 2013. Many human accelerated
788 regions are developmental enhancers. *Philos Trans R Soc Lond B Biol Sci* **368**:
789 20130025.
- 790 Chen J, Swofford R, Johnson J, Cummings BB, Rogel N, Lindblad-Toh K, Haerty W, Palma FD,
791 Regev A. 2019. A quantitative framework for characterizing the evolutionary history of
792 mammalian gene expression. *Genome Res* **29**: 53-63.

- 793 Cooper GM, Shendure J. 2011. Needles in stacks of needles: finding disease-causal variants in
 794 a wealth of genomic data. *Nat Rev Genet* **12**: 628-640.
- 795 Dezso Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, Dosymbekov D, Bugrim A,
 796 Rakhmatulin E, Brennan RJ, Guryanov A et al. 2008. A comprehensive functional
 797 analysis of tissue specificity of human gene expression. *BMC Bio* **6**: 49.
- 798 Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore
 799 E, Varki A, Ravid R et al. 2002. Intra- and interspecific variation in primate gene
 800 expression patterns. *Science* **296**: 340-343.
- 801 The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the
 802 human genome. *Nature* **489**(7414): 57-74.
- 803 Friedman J, Hastie T, Tibshirani R. 2010. Regularization Paths for Generalized Linear Models
 804 via Coordinate Descent. *J Stat Softw* **33**: 1-22.
- 805 Gallego Romero I, Pavlovic BJ, Hernando-Herraez I, Zhou X, Ward MC, Banovich NE, Kagan
 806 CL, Burnett JE, Huang CH, Mitrano A et al. 2015a. A panel of induced pluripotent stem
 807 cells from chimpanzees: a resource for comparative functional genomics. *eLife* **4**:
 808 e07103. doi: 10.7554/eLife.07103.
- 809 Gallego Romero I, Pavlovic BJ, Hernando-Herraez I, Zhou X, Ward MC, Banovich NE, Kagan
 810 CL, Burnett JE, Huang CH, Mitrano A et al. 2015b. A panel of induced pluripotent stem
 811 cells from chimpanzees: a resource for comparative functional genomics. *eLife* **4**:
 812 e07103. doi: 07110.07554/eLife.07103.
- 813 The GTEx Consortium. 2017. Genetic effects on gene expression across human tissues.
 814 *Nature* **550**: 204–213.
- 815 Gilad Y, Mizrahi-Man O. 2015. A reanalysis of mouse ENCODE comparative gene expression
 816 data. *F1000 Research* **4**: doi: 10.12688/f11000research.16536.12681.
- 817 Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP. 2005. Multi-species microarrays reveal the
 818 effect of sequence divergence on gene expression profiles. *Genome Res* **15**: 674-680.
- 819 Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing
 820 reads to differentially methylated regions. *Genome Biol* **13**: R83.
- 821 Hansen KD, Sabunciyan S, Langmead B, Nagy N, Curley R, Klein G, Klein E, Salomon D,
 822 Feinberg AP. 2014. Large-scale hypomethylated blocks associated with Epstein-Barr
 823 virus-induced B-cell immortalization. *Genome Res* **24**: 177-184.
- 824 Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu
 825 Y, Diep D et al. 2011. Increased methylation variation in epigenetic domains across
 826 cancer types. *Nat Genet* **43**: 768-775.
- 827 Hernando-Herraez I, Garcia-Perez R, Sharp AJ, Marques-Bonet T. 2015a. DNA Methylation:
 828 Insights into Human Evolution. *PLoS Genet* **11**: e1005661. doi:
 829 10.1371/journal.pgen.1005661.
- 830 Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H, Prado-
 831 Martinez J, Sharp AJ, Esteller M, Marques-Bonet T. 2015b. The interplay between DNA
 832 methylation and sequence divergence in recent human evolution. *Nucleic Acids Res* **43**:
 833 8204-8214.
- 834 Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C,
 835 Navarro A, Esteller M, Sharp AJ, Marques-Bonet T. 2013. Dynamics of DNA methylation
 836 in recent human and great ape evolution. *PLoS Genet* **9**: e1003763. doi:
 837 10.1371/journal.pgen.1003763.
- 838 Karaman MW, Houck ML, Chemnick LG, Nagpal S, Chawannakul D, Sudano D, Pike BL, Ho
 839 VV, Ryder OA, Hacia JG. 2003. Comparative analysis of gene-expression patterns in
 840 human and African great ape cultured fibroblasts. *Genome Res* **13**: 1619-1630.
- 841 Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer TR, Fujita PA,
 842 Guruvadoo L, Haeussler M et al. 2014. The UCSC Genome Browser database: 2014
 843 update. *Nucleic Acids Res* **42**: D764-770.

- 844 Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M,
845 Paabo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of
846 humans and chimpanzees. *Science* **309**: 1850-1854.
- 847 Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, Dietzsch J, Steigale S, Do HH,
848 Weiss G, Enard W et al. 2004a. Regional patterns of gene expression in human and
849 chimpanzee brains. *Genome Res* **14**: 1462-1473.
- 850 Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W,
851 Paabo S. 2004b. A neutral model of transcriptome evolution. *PLoS Biol* **2**: e132. doi:
852 10.1371/journal.pbio.0020132.
- 853 Khan Z, Ford MJ, Cusanovich DA, Mitrano A, Pritchard JK, Gilad Y. 2013. Primate transcript
854 and protein expression levels evolve under compensatory selection pressures. *Science*
855 **342**: 1100-1104.
- 856 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate
857 alignment of transcriptomes in the presence of insertions, deletions and gene fusions.
858 *Genome Biol* **14**.
- 859 King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**:
860 107-116.
- 861 Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-
862 Seq applications. *Bioinformatics* **27**: 1571-1572.
- 863 Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: Precision weights unlock linear model
864 analysis tools for RNA-seq read counts. *Genome Biol* **15**: R29.
- 865 Lemos B, Meiklejohn CD, Caceres M, Hartl DL. 2005. Rates of divergence in gene expression
866 profiles of primates, mice, and flies: stabilizing selection and variability among functional
867 categories. *Evolution* **59**: 126-137.
- 868 Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, Dobin A, Zaleski C, Beer MA, Chapman
869 WC et al. 2014. Comparison of the transcriptional landscapes between human and
870 mouse tissues. *Proc Natl Acad Sci U S A* **111**: 17224-17229.
- 871 Lindskog C. 2016. The Human Protein Atlas - an important resource for basic and clinical
872 research. *Expert Rev Proteomics* **13**: 627-629.
- 873 Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z,
874 Ngo Q-M et al. 2009. Human DNA methylomes at base resolution show widespread
875 epigenomic differences. *Nature* **462**: 315-322.
- 876 Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: an assessment of
877 technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**:
878 1509-1517.
- 879 Martin DI, Singer M, Dhahbi J, Mao G, Zhang L, Schroth GP, Pachter L, Boffelli D. 2011.
880 Phyloepigenomic comparison of great apes reveals a correlation between somatic and
881 germline methylation states. *Genome Res* **21**: 2049-2057.
- 882 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
883 *EMBnet.journal* **17**. doi: 10.14806/ej.17.1.200.
- 884 Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform
885 regulation in Mammalian tissues. *Science* **338**: 1593-1599.
- 886 Molnar A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. 2011. Sperm
887 methylation profiles reveal features of epigenetic inheritance and evolution in primates.
888 *Cell* **146**: 1029-1041.
- 889 Nowick K, Gernat T, Almaas E, Stubbs L. 2009. Differences in human and chimpanzee gene
890 expression patterns define an evolving network of transcription factors in brain. *Proc Natl
891 Acad Sci U S A* **106**: 22358-22363.
- 892 Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y. 2011. A genome-wide study of DNA
893 methylation patterns and gene expression levels in multiple human and chimpanzee
894 tissues. *PLoS Genet* **7**: e1001316. doi: 10.1371/journal.pgen.1001316.

- 895 Pierson E, Consortium GT, Koller D, Battle A, Mostafavi S, Ardlie KG, Getz G, Wright FA, Kellis
 896 M, Volpi S et al. 2015. Sharing and Specificity of Co-expression Networks across 35
 897 Human Tissues. *PLoS Comput Biol* **11**: e1004220. doi: 10.1371/journal.pcbi.1004220.
- 898 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
 899 features. *Bioinformatics* **26**: 841-842.
- 900 Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the
 901 evolution of gene regulation. *Nat Rev Genet* **13**: 505-516.
- 902 Sharp AJ, Stathaki E, Migliavacca E, Brahmachary M, Montgomery SB, Dupre Y, Antonarakis
 903 SE. 2011. DNA methylation profiles of human active and inactive X chromosomes.
 904 *Genome Res* **21**: 1592-1600.
- 905 Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, Tewari AK, London D, Song L,
 906 Lee BK, Iyer VR et al. 2012. Extensive evolutionary changes in regulatory element
 907 activity during human origins are associated with altered gene expression and positive
 908 selection. *PLoS Genet* **8**: e1002789. doi: 10.1371/journal.pgen.1002789.
- 909 Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential
 910 expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.
- 911 Smyth GK, Michaud J, Scott HS. 2005. Use of within-array replicate spots for assessing
 912 differential expression in microarray experiments. *Bioinformatics* **21**: 2067-2075.
- 913 Stephens M. 2017. False discovery rates: a new deal. *Biostatistics* **18**: 275-294.
- 914 Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, Zhang M, Byron R, Canfield T,
 915 Stelhing-Sun S, Lee K et al. 2014. Conservation of trans-acting circuitry during
 916 mammalian regulatory evolution. *Nature* **515**: 365-370.
- 917 Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global discovery
 918 of conserved genetic modules. *Science* **302**: 249-255.
- 919 Supek F, Bosnjak M, Skunca N, Smuc T. 2011. REVIGO summarizes and visualizes long lists
 920 of gene ontology terms. *PLoS One* **6**: e21800. doi: 10.1371/journal.pone.0021800.
- 921 Thul PJ, Lindskog C. 2018. The human protein atlas: A spatial map of the human proteome.
 922 *Protein Sci* **27**: 233-244.
- 923 Tsankov AM, Gu H, Akopian V, Ziller MJ, Donaghey J, Amit I, Gnirke A, Meissner A. 2015.
 924 Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**:
 925 344-349.
- 926 Tung J, Barreiro LB, Johnson ZP, Hansen KD, Michopoulos V, Toufexis D, Michelini K, Wilson
 927 ME, Gilad Y. 2012. Social environment is associated with gene regulatory variation in
 928 the rhesus macaque immune system. *Proc Natl Acad U S A*
 929 doi:10.1073/pnas.1202734109.
- 930 Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A,
 931 Kampf C, Sjostedt E, Asplund A et al. 2015. Proteomics. Tissue-based map of the
 932 human proteome. *Science* **347**: 1260419.
- 933 Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen
 934 JT, Jasinska AJ et al. 2015. Enhancer evolution across 20 mammalian species. *Cell*
 935 **160**: 554-566.
- 936 Ward MC, Wilson MD, Barbosa-Morais NL, Schmidt D, Stark R, Pan Q, Schwalie PC, Menon S,
 937 Lukk M, Watt S et al. 2013. Latent regulatory potential of human-specific repetitive
 938 elements. *Mol Cell* **49**: 262-272.
- 939 Warnefors M, Eyre-Walker A. 2012. A selection index for gene expression evolution and its
 940 application to the divergence between humans and chimpanzees. *PLoS One* **7**: e34935.
 941 doi: 10.1371/journal.pone.0034935.
- 942 Yu NY, Hallstrom BM, Fagerberg L, Ponten F, Kawaji H, Carninci P, Forrest AR, Fantom C,
 943 Hayashizaki Y, Uhlen M et al. 2015. Complementing tissue characterization by
 944 integrating transcriptome profiling from the Human Protein Atlas and from the FANTOM5
 945 consortium. *Nucleic Acids Res* **43**: 6787-6798.

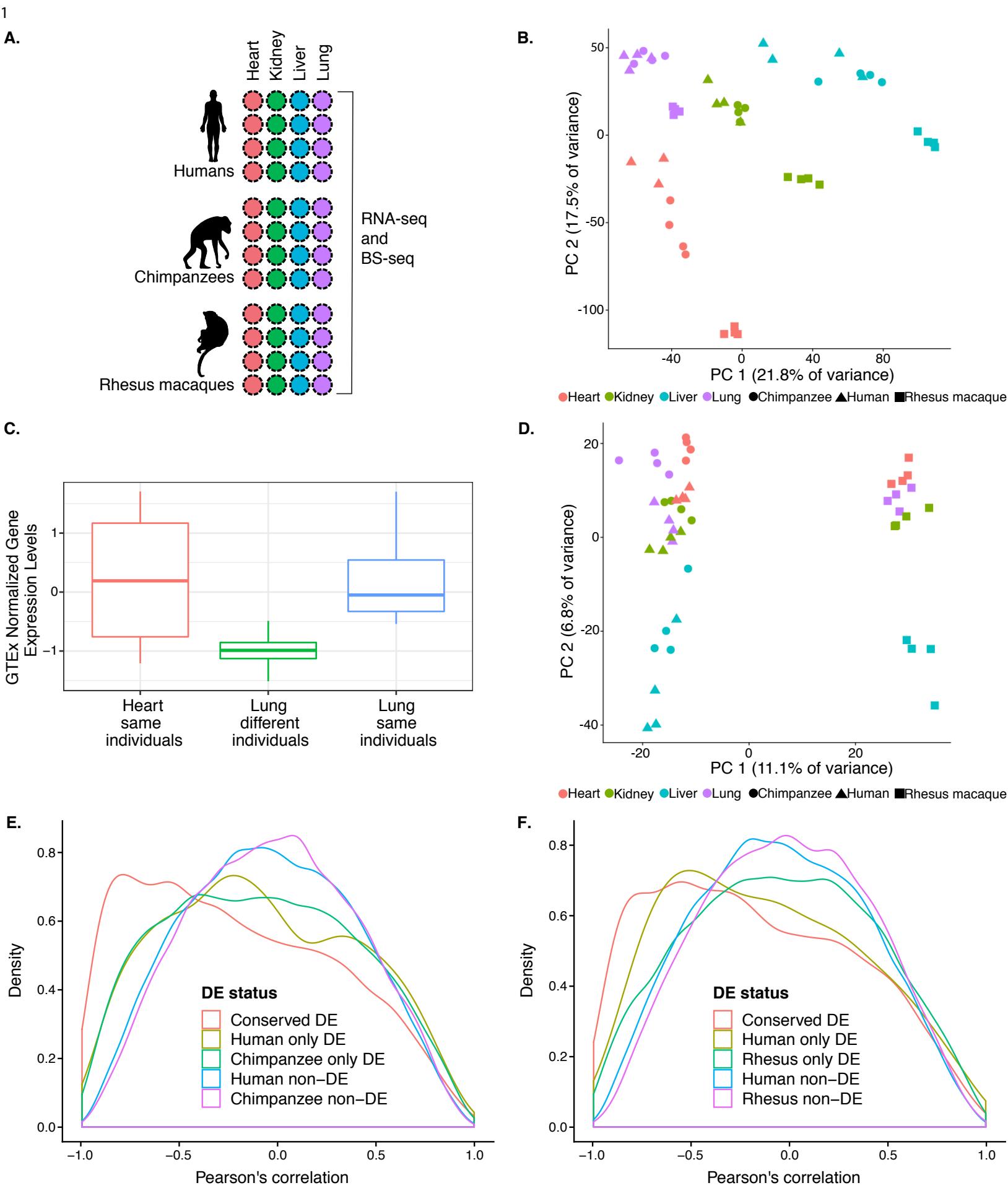
- 946 Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network
947 analysis. *Stat Appl Genet Mol Biol* **4**: Article17.
- 948 Zhou X, Cain CE, Myrthil M, Lewellen N, Michelini K, Davenport ER, Stephens M, Pritchard JK,
949 Gilad Y. 2014. Epigenetic modifications are associated with inter-species gene
950 expression variation in primates. *Genome Biol* **15**: 547.
- 951

Table 1

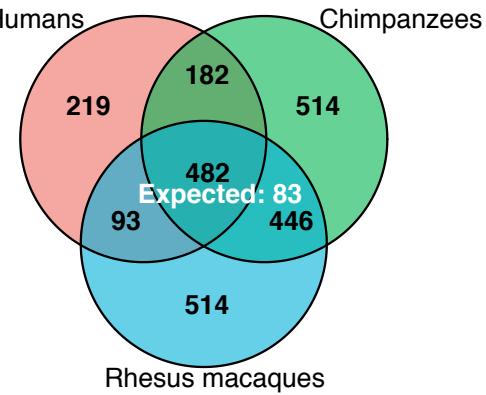
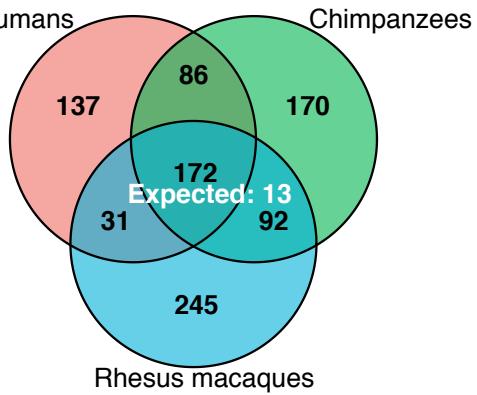
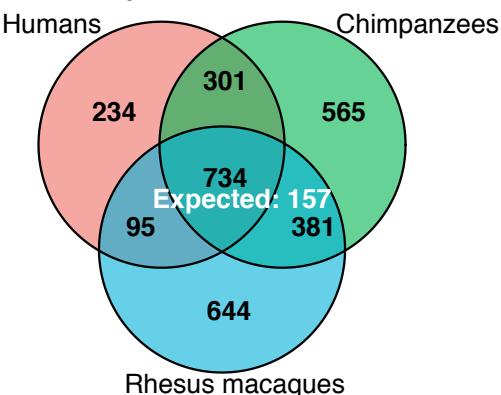
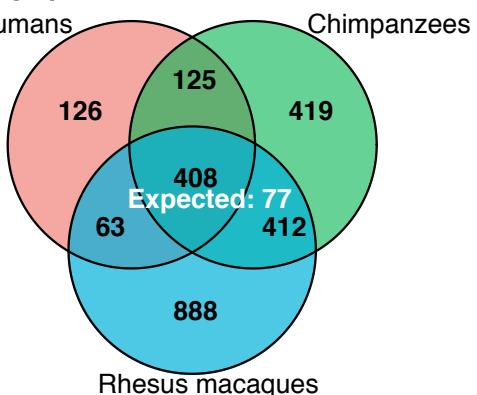
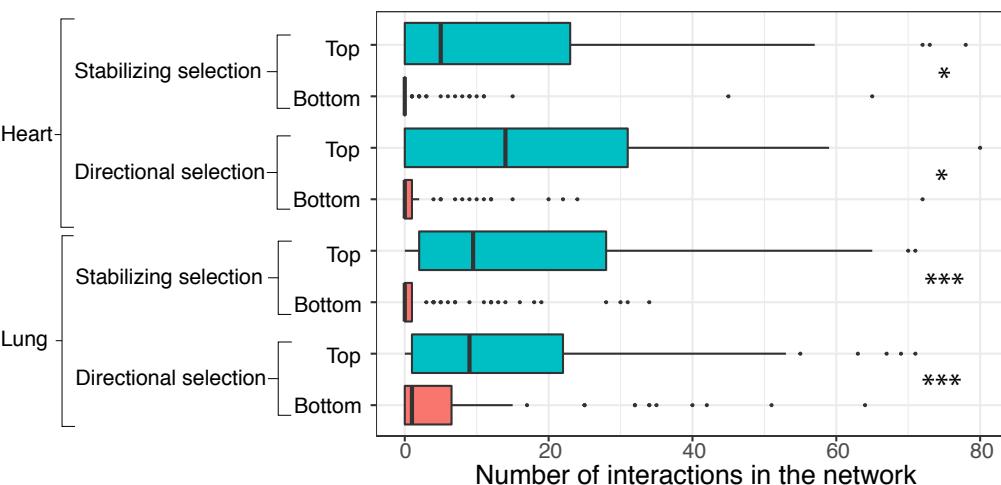
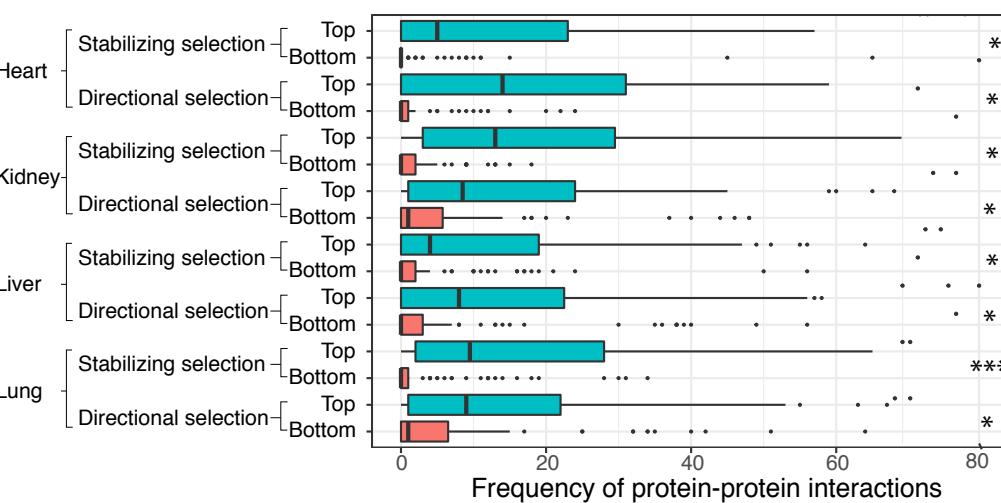
DE between tissues (within species)	Heart-Kidney	Heart-Liver	Heart-Lung	Kidney-Liver	Kidney-Lung	Liver-Lung
Human	4224	4776	4037	4248	3695	4701
Chimpanzee	4971	6295	5365	4625	4623	6247
Rhesus macaque	5980	6933	6814	5126	5721	7027
DE between species	Heart	Kidney	Liver	Lung		
Human vs. Chimpanzee	2195	799	1364	805		
Human vs. Rhesus Macaque	4098	2347	2868	2833		
Chimpanzee vs. Rhesus Macaque	2781	2286	3139	1917		

Table 2

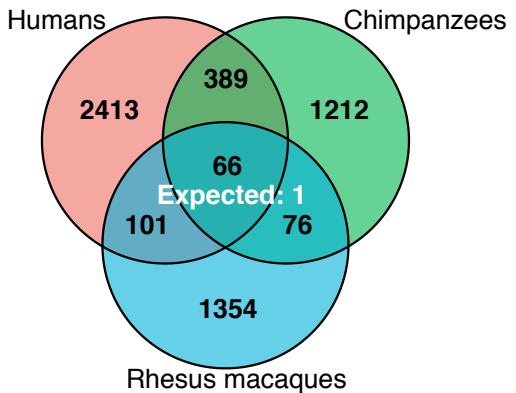
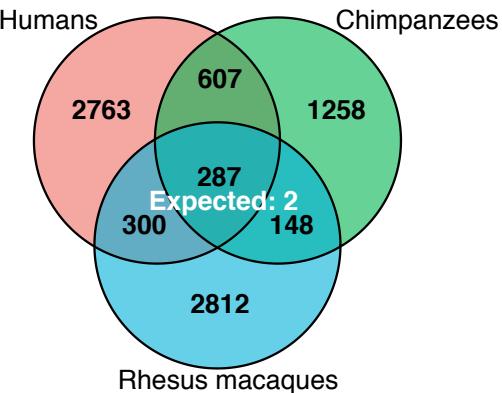
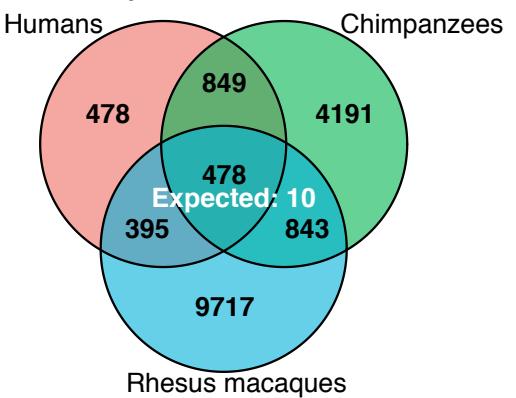
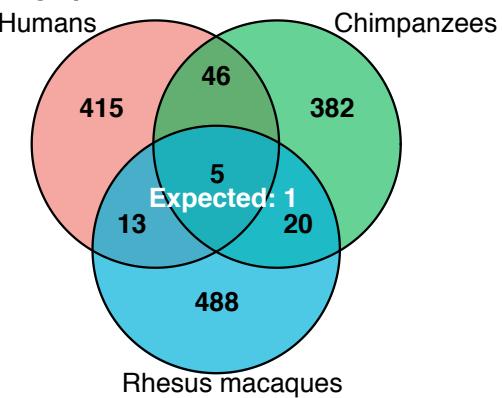
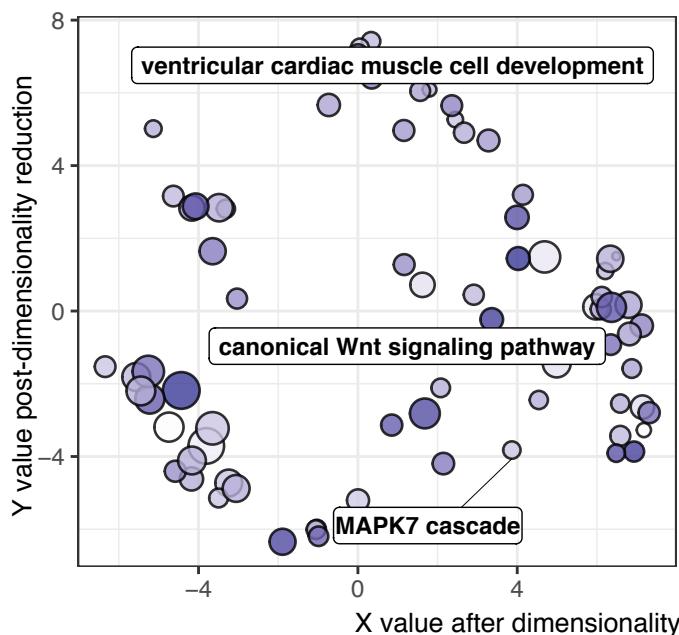
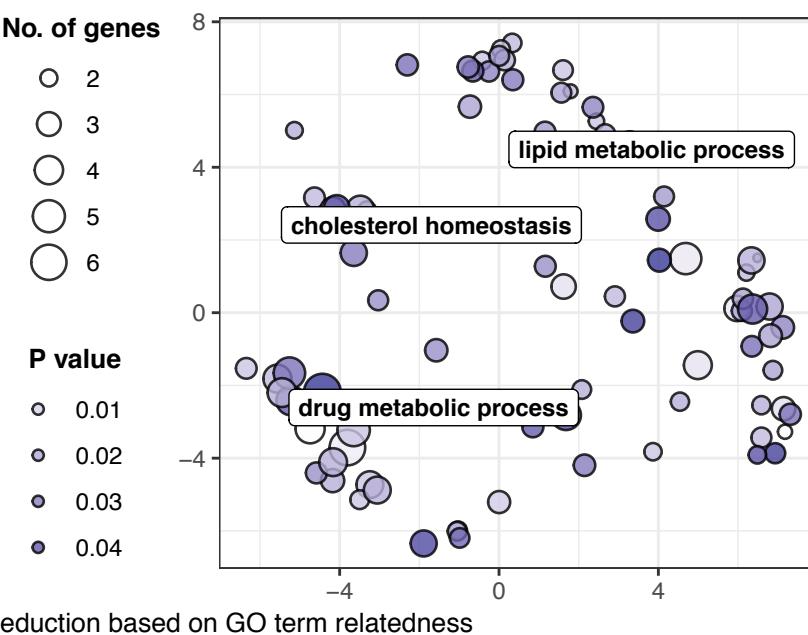
Tissue DMRs (within species)	Heart-Kidney	Heart-Liver	Heart-Lung	Kidney-Liver	Kidney-Lung	Liver-Lung
Human	30291 (1.7%)	22561 (1.3%)	14208 (0.8%)	17910 (1.1%)	16521 (1.0%)	12842 (0.8%)
Chimpanzee	17699 (1.1%)	28767 (1.7%)	8524 (0.5%)	23847 (1.4%)	12076 (0.7%)	22107 (1.3%)
Rhesus macaque	23023 (1.5%)	41280 (2.7%)	7026 (0.5%)	35910 (2.4%)	15889 (1.1%)	32636 (2.1%)
Species DMRs	Heart	Kidney	Liver	Lung		
Human vs. Chimpanzee	14504 (1.1%)	10667 (0.8%)	9476 (0.8%)	8617 (0.7%)		
Human vs. Rhesus Macaque	25539 (5.3%)	21292 (4.5%)	21639 (4.5%)	17696 (3.9%)		

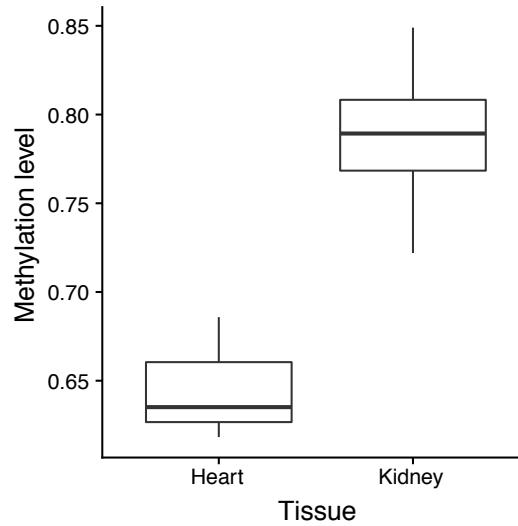
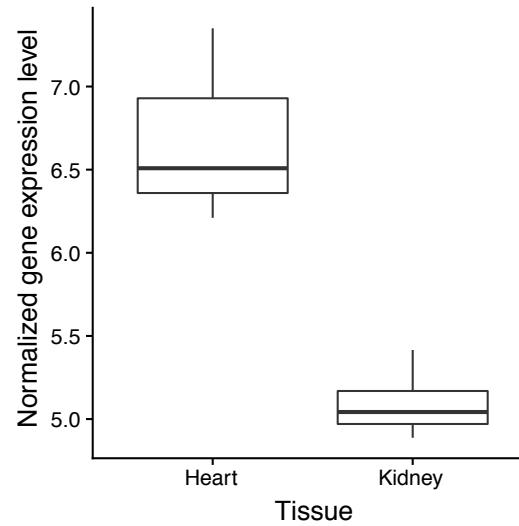
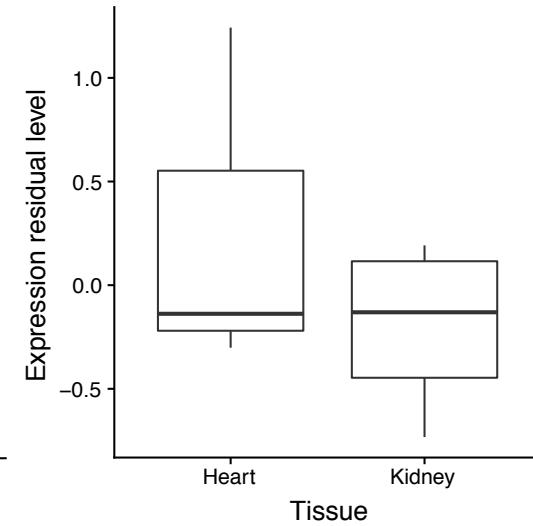
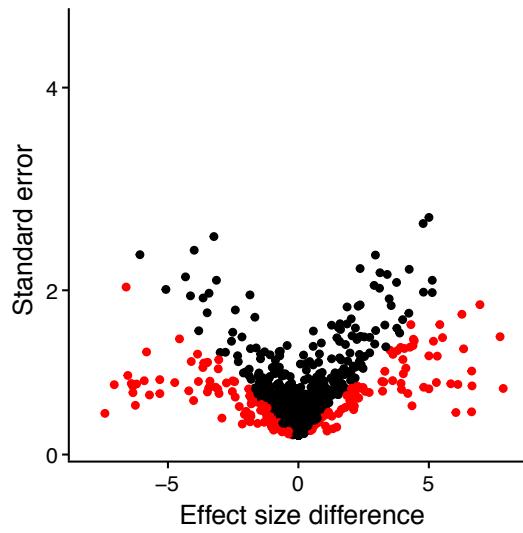
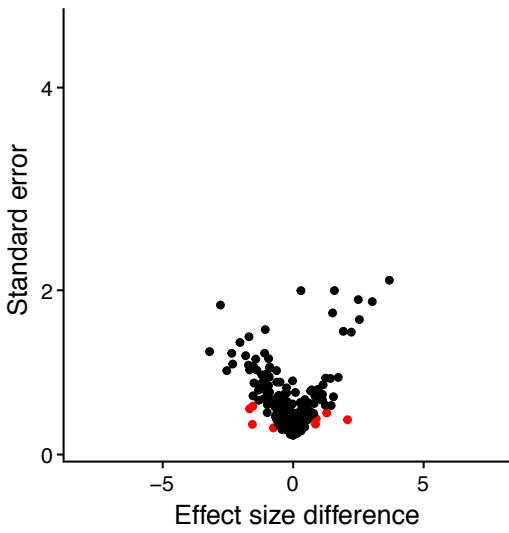
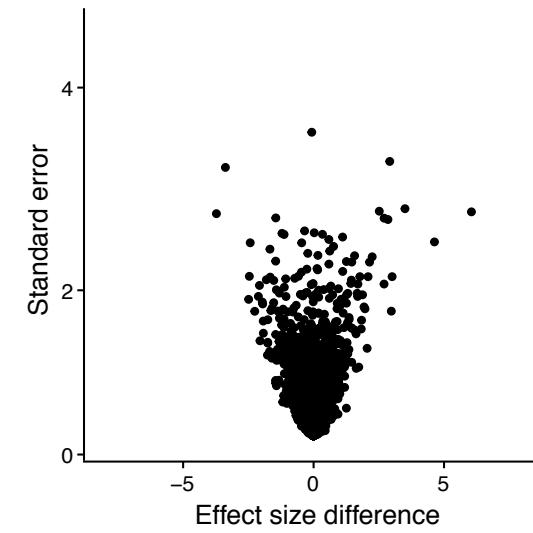
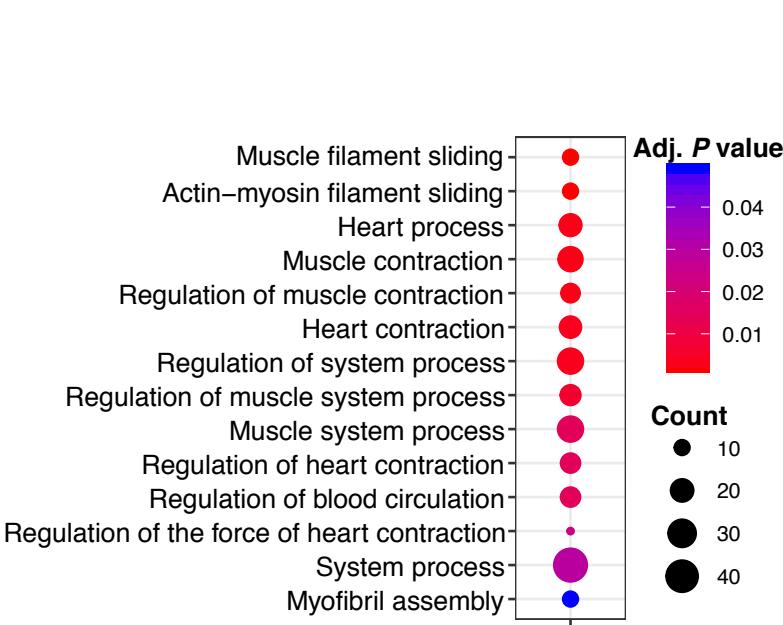
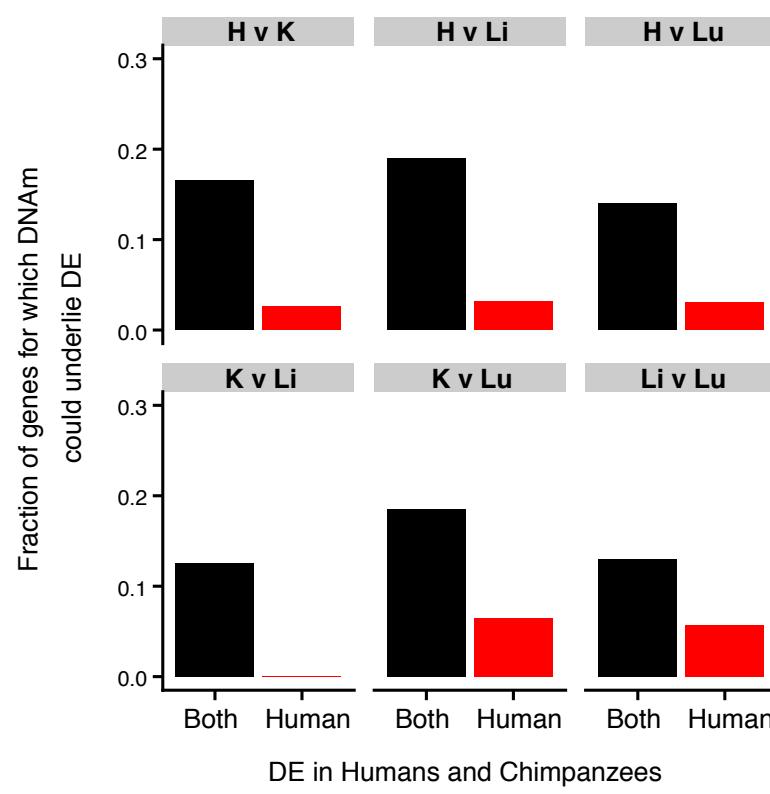


2

A. Heart-specific**B. Kidney-specific****C. Liver-specific****D. Lung-specific****E. Overlap in gene co-expression networks****F. Overlap in protein-protein interactions**

3

A. Heart-specific**B. Kidney-specific****C. Liver-specific****D. Lung-specific****E. Enrichment in hearts****F. Enrichment in livers**

A. DNA methylation**B. Gene expression****C. Expression residuals****D. Conserved DE heart v. kidney****E. Non-conserved DE****F. Non-DE heart v. kidney****G. Conserved DE enrichment****H. Additional tissues**



A comparison of gene expression and DNA methylation patterns across tissues and species

Lauren E Blake, Julien Roux, Irene Hernando-Herraez, et al.

Genome Res. published online January 17, 2020

Access the most recent version at doi:[10.1101/gr.254904.119](https://doi.org/10.1101/gr.254904.119)

P<P	Published online January 17, 2020 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Open Access	Freely available online through the <i>Genome Research</i> Open Access option.
Creative Commons License	This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at http://creativecommons.org/licenses/by-nc/4.0/ .
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .

An advertisement for a webinar. The background is dark blue with a faint grid pattern. On the left, the word "Webinar" is written in white. In the center, the text "Automation-friendly full-length scRNA-seq" is displayed in white. To the right, there is a green circular logo with the text "that's GOOD for science" and the Takara logo. Below the logo, the words "Oncotek Takara cellartis" are written in small white text.

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>
