Supplemental Materials for

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

Lauren E. Blake[1*], Julien Roux[1,2,3*], Irene Hernando-Herraez[4], Nicholas E. Banovich[1], Raquel Garcia Perez[4], Chiaowen Joyce Hsiao[1], Ittai Eres[1], Claudia Cuevas[1], Tomas Marques-Bonet[4-7], and Yoav Gilad[1,8,‡]

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

*Authors contributed equally to this work.

‡Correspondence should be addressed to Yoav Gilad (gilad@uchicago.edu)

Contents

Supplemental Methods

RNA library preparation and sequencing…….4

Quantifying the number of RNA-seq reads from orthologous genes…….4

RNA-seq data transformation and normalization…….4

SNP calling in the RNA-seq and BS-seq data…….6

Analysis of technical variables…….6

Differential expression analysis using a linear model-based framework…….7

Comparing the rank of tissue-specific DE genes in our dataset to the GTEx Project…….7

Expected overlap of genes and significance of the observed overlap…….8

The overlap between DE genes and previously defined networks…….8

BS-seq library preparation, sequencing, and mapping…….9

Methylation level estimate smoothing…….11

Identifying differentially methylated regions (DMRs)…….11

Overlap of tissue-specific DMRs with regulatory regions…….11

Supplemental Information

Assessing the impact of technical variables on gene expression levels and DNA methylation levels…….13

Tissue-specific gene expression patterns…….15

Adaptive shrinkage and false sign rate (FSR) to identify tissue-specific genes…….16

Identifying inter-species differences between tissues…….16

Promoter DNA methylation quality…….17

Identification of DMRs across species (S-DMRs)…….18

References…….19

List of Supplemental Figures

S1. Distributions of potential confounders across biological variables of interest…….25

S2. Sample QC…….26

S3. Correlation matrix of normalized log2(CPM) gene expression values from 12,184 genes…….27

S4. Density function of DNA methylation levels across all species & tissues…….28

S5. Correlation matrix of smoothed DNA methylation levels from all orthologous CpGs…….29

S6. Principal components analysis (PCA) in human and chimpanzee hearts, kidneys, and livers…….30

S7. When comparing DNA methylation levels of human T-DMRs and of orthologous regions in the same tissues, clustering is more highly correlated with tissue than species…….31

S8. When comparing DNA methylation levels of chimpanzee T-DMRs and of orthologous regions in the same tissues, clustering is more highly correlated with tissue than species…….32

S9. When comparing DNA methylation levels of rhesus macaque T-DMRs and of orthologous regions in the same tissues, clustering is more highly correlated with tissue than species…….33

S10. Inter-species DNA methylation and gene expression levels (FDR = 0.05 and FSR = 0.05) in humans and chimpanzees…….34

S11. Interspecies DNA methylation and gene expression levels (FDR = 0.05 and FSR = 0.05) in humans and rhesus macaques…….35

Supplemental Methods

RNA library preparation and sequencing

A relatively small percentage of reads could not be assigned to any sample because their adaptor sequence did not match any of the adaptors used in the study. Therefore, we ran in-house Perl scripts to recover those reads that differed at a single position. Because of a calibration issue with the Gilad lab Illumina HiSeq sequencer, the first and the third flow-cell of the study (16 lanes) yielded a low number of reads. However, this problem did not affect the quality of the reads, so we kept these lanes for the analysis.

Quantifying the number of RNA-seq reads from orthologous genes

When mapping to the species' genome, we allowed for up to two mismatches in each read and kept only reads that mapped uniquely. Tophat2 uses unmapped reads to perform gapped alignments to the genome and discover new exon-exon junction sites. For this step, we disabled the coverage-based search, and only 1 mismatch was allowed in the anchor region of the reads (>=8 nt). The minimum intron length was set to 70 nt and the maximum to 50,000 nt. This yielded from 28,544,039 (R1H) to 72,808,273 (R4Li) mapped reads across samples (mean = 46,514,692 reads). Mapping rates were between 71% and 94%.

We performed all downstream analyses in R (versions 3.1.1, 3.2.2, or 3.4.3) unless otherwise stated.

RNA-seq data transformation and normalization

We calculated the $\log_2$-transformed counts per million (CPM) from the raw gene counts of each sample using edgeR (Robinson et al. 2010). We then filtered out lowly expressed genes, keeping only genes with an expression level of $\log_2(\text{CPM}) > 1.5$ in at least 24 of the 48 samples (Robinson and Oshlack 2010). We normalized the original read counts using the weighted trimmed mean of M-values algorithm (TMM) (Robinson and Oshlack 2010). This process helped

us to account for differences in the read counts at the extremes of the distributions. We then calculated the TMM-normalized $\log_2$-transformed CPM values for each of the genes.

After performing normalization, we performed principal components analysis (PCA) using the TMM-normalized $\log_2$-transformed CPM values of all genes, 1 human heart sample (H1H) clustered with the human livers rather than the hearts. After performing SNP calling on the RNA-seq data (see section below), we found that the SNPs in sample H1H matched those from the other tissues from this individual. We removed this sample (H1H) from the list of the original gene counts. We again filtered for lowly expressed genes keeping only genes with an expression level of $\log_2(\text{CPM}) > 1.5$ in at least 24 of the 47 samples. We also wanted to allow for small differences in the distributions of gene expression across tissues. Therefore, on the 12,184 remaining genes, we performed a TMM normalization and then performed a cyclic loess normalization with the function *normalizeCyclicLoess* from the R/Bioconductor package limma (Ballman et al. 2004; Ritchie et al. 2015). To run PCA, we used the R function *prcomp*. For hierarchical clustering, we used unsupervised agglomerative clustering on the correlation matrix of the gene expression data.

In this transformation and normalization process, we were interested in the impact of sample-specific biases in GC content on the gene expression counts. Therefore, we used the WASP pipeline (van de Geijn et al. 2015) to obtain expected GC-normalized counts. Specifically, we filtered the genes with lowly expressed counts so that only genes with the $\log_2$-transformed CPM > -5.5 in at least 2 of the 4 samples in each species-tissue pair (e.g. 2/4 chimpanzee hearts) remained. For each of the 16,616 genes that remained, we summed the read depth (raw counts) of the 4 samples in each tissue-species pair. We used the WASP pipeline (van de Geijn et al. 2015) (https://github.com/bmvdgeijn/WASP/blob/master/CHT/update_total_depth.py) to obtain expected read counts, adjusted for read depth and GC content. The GC content for each of the orthologous metaexons was previously calculated as part of (Gallego Romero et al. 2015). For each tissue-species pair, the adjusted raw counts and the actual raw counts were highly

correlated (> 0.98). Therefore, we did not adjust for read depth or GC content in our RNA-sequencing data.

SNP calling in the RNA-seq and BS-seq data

We called single nucleotide variants on RNA-seq data from each tissue and sample using standard hard filtering parameters according to GATK recommendations (Van der Auwera et al. 2013). Briefly, duplicated reads were removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard). Reads were then subjected to local realignment, base-score recalibration, and candidate-variant calling using the IndelRealigner, TableRecalibration, and HaplotypeCaller tools from GATK (McKenna et al. 2010). We required a base quality score ≥20. We only considered variants that were observed in at least four of the samples.

We used the same method for the BS-seq data. Through this process, we found that 2 groups of samples had been mislabeled during sequencing: the sample labelled R3Li was actually R2Li, and R3Lu was actually R2Lu.

Analysis of technical variables

We recorded variables related to the samples (e.g. sex), variables specific to gene expression (e.g. RNA-seq flow cell number), and variables related to methylation levels (e.g. number of CpG sites covered) (Supplemental Table S1A-E). Briefly, we determined which of our recorded technical variables were significant predictors for each of the gene expression PCs 1-5 using individual linear models for each of the gene expression variables (FDR < 10% for each test). The significant technical variables were then tested against our biological variables of interest, tissue and species, again with individual linear models. For the numerical technical variables, we quantified the strength of these associations using the *P* values from analysis of variance (ANOVA), and used a Chi-squared test (using Monte Carlo simulated *P* values) for the categorical technical variables (significance at FDR < 10%). We repeated the same analysis for

159 methylation data, testing the associations between methylation PCs 1-5 and sample
160 information.
161
162 **Differential expression analysis using a linear model-based framework**
163 We implemented a linear model-based framework using the R packages *limma* and *voom*
164 (Smyth 2004; Smyth et al. 2005; Law et al. 2014). This pipeline has previously been shown to
165 perform well with at least 3 samples per condition (Rapaport et al. 2013; Soneson and Delorenzi
166 2013).
167 We hypothesized that RNA quality may be impacted by post-mortem time prior to
168 collection. According to the documentation that we received from the different sites, all of the
169 rhesus macaque samples were collected earlier than all of the chimpanzee samples. These
170 differences could impact RNA quality. Hence, we used RIN score as a proxy for RNA quality, and
171 included RIN score in the linear models.
172 In the linear models, species, tissue, RIN score, and species-by-tissue interaction terms
173 were modeled as fixed effects. Individual was modeled as a random effect. We used contrast
174 tests in limma to identify genes that were differentially expressed between tissues within each
175 species and across species in the same tissue. We corrected for multiple testing with the
176 Benjamini and Hochberg FDR (Benjamini and Hochberg 1995). Genes were considered
177 significantly DE at FDR-adjusted *P* values < 0.01, unless otherwise stated.
178
179
180 **Comparing the rank of tissue-specific DE genes in our dataset to the GTEx Project**
181 To benchmark the conserved tissue-specific DE genes, we compared the rank of each
182 gene's expression level in our data to its corresponding rank in the GTEx v6 heart, liver, lung, and
183 kidney data (The GTEx Consortium 2017). After this comparison, we looked for enrichment of
184 genes with a given rank. To do so, we used the R package *topGO* (Alexa et al. 2006), with the
185 same implementation as in (Blischak et al. 2015). This implementation included the use of Fisher's

186 Exact Test, with topGO's *weight01* algorithm (which takes into account the correlation among GO
187 categories within the graph structure of the program). We then repeated this process for the
188 tissue-specific DE genes identified in humans only.
189
190 **Expected overlap of genes and significance of the observed overlap**
191 We used the process from (Pai et al. 2011), based on the hypergeometric distribution, to
192 assess the expected overlap of the conserved DE genes and significance of the observed number
193 of conserved DE genes. This process relies on comparing a population proportion to a sample
194 proportion.
195 We first asked about the overlap of DE genes in humans and chimpanzees. To be
196 conservative, in the case of the human and chimpanzee overlap, we assigned the species with
197 the greater number of genes in the direction of interest as the population and the other species
198 as the sample. To assess the expected overlap in upregulated human and chimpanzee genes in
199 a given tissue, we used the *P* value from the hypergeometric distribution with the following
200 parameters: *m* is the total number of DE genes in the population, *n* is the total number of
201 upregulated genes in a population minus *m*, *q* is the observed overlap of upregulated DE genes
202 (between the humans and chimpanzees), and *k* is the total number of upregulated DE genes in
203 the sample, all within the given tissue. The "expected overlap" is the value at which the maximum
204 likelihood estimate for which *m, n,* and *k* occurs. We then repeated this process for the
205 upregulated and downregulated DE genes, in all four tissues separately. To obtain the same
206 statistics for the tissue-specific DE genes, we used only the tissue-specific DE genes, and
207 calculated *n* as the total number of genes upregulated in the tissue of interest compared to the
208 other three tissues in the population minus the number of tissue-specific DE genes in the
209 population.
210 To calculate these statistics for all three species, we used this framework to ask whether
211 the observed overlap between all three species was significant relative to the overlap of the

human and chimpanzee DE genes. In the same manner, we used the hypergeometric distribution to assess the expected overlap and significance of the number of conserved tDMRs (tissue differentially methylated regions) and conserved tissue-specific DMRs.

The overlap between DE genes and previously defined networks

To find gene expression patterns that are consistent with the action of natural selection, for each significant DE gene, we determined the within-species variance of all 3 species and found the average of the 3 variances. We then ranked the genes by the mean variances. For the co-transcription network analysis, we used the shared TE-TE networks for the heart and lung as well as the heart-specific network from the Supplementary Materials in (Saha et al. 2017). We downloaded the list of protein-protein interactions in the heart, kidney, liver, and lung from the Human Protein Atlas (Uhlen et al. 2015). For the interaction analyses, we counted the number of interactions for each gene in the co-transcription networks or the protein-protein interactions list, in the appropriate tissue.

BS-seq library preparation, sequencing, and mapping

To assess the efficiency of the conversion reaction (Bock 2012), we spiked the extracted DNA with unmethylated lambda phage DNA. For each sample, we prepared at least two libraries, with independent PCR amplifications to minimize PCR duplication rates. The BS-seq libraries were sequenced on 111 lanes on 17 flow-cells on an Illumina HiSeq 2500 sequencer in the Gilad lab or at the University of Chicago Genomics Facility. Reads were single-end and 49 to 59 bp. The distribution of libraries of technical replicates over the flow-cells and additional related information is described in Supplemental Tables S1C-E.

Similar to RNA-seq data, we used FastQC to generate quality reports. TrimGalore (version 0.2.8) was used to trim adapter sequences incorporated in the BS-seq reads, using a stringency

of 3, and to cut the low-quality ends of reads, using a quality threshold of 20. We eliminated reads shorter than 15 bp post-trimming.

We aligned the trimmed reads to the human (hg19, February 2009), chimpanzee (panTro3, October 2010), or rhesus macaque (rheMac2, January 2006) genomes, and to the lambda phage genome using the Bismark aligner (version 0.8.1)(Krueger and Andrews 2011). The Bismark aligner maps reads to *in-silico* converted (G to A and C to T) genome sequences using Bowtie (version 1.0.0). This aligner was shown to perform well on benchmark studies (Chatterjee et al. 2012; Tran et al. 2014; Tsuji and Weng 2016). We permitted one mismatch in the seed of the alignment, and by default Bismark reports only uniquely mapped reads. Across technical replicates, mapping rates ranged from 49% to 82% (median 76%). We applied the Bismark deduplication script to each technical replicate to remove reads mapped to the same starting genomic position, which likely arise through PCR amplification of the same DNA fragments during library preparation (Bock 2012). Across technical replicates, the duplication rates ranged from 2.8% to 44% (median 11%).

To determine the bisulfite conversion efficiency, we calculated the conversion rate at cytosines from the spiked-in lambda phage DNA (for which coverage ranged from 12× to 107×). We found this rate to be at least 99.4% across technical replicates, increasing confidence in our data.

After combining all technical replicates, the average genome-wide coverage, calculated at CpG sites shared across the three species, ranged from 1.7× to 5.7× per sample (median of 4×), corresponding to 12M to 23M CpG sites with a coverage of at least 2× (median 19M).

DNA methylation level estimate smoothing

Since DNA methylation data from BS-seq is typically lower coverage than DNA methylation array data, we first applied a smoothing procedure on raw methylation levels for each

sample. It has previously been shown that this procedure increases the precision of low-coverage BS-seq data, and yields methylation estimates that are in excellent agreement with high-coverage BS-seq data without smoothing (Hansen et al. 2012; Ziller et al. 2014). To perform the smoothing, we used the BSmooth method (as implemented in the Bioconductor package *bsseq*, version 0.10.0) (Hansen et al. 2012; Ziller et al. 2014) -- with the default parameters for smoothing -- at least 70 CpG sites with methylation data in a smoothing window of at least 1 kb.

We had 17.6M human-chimpanzee orthologous CpGs and 7.5M CpGs orthologous across all 3 species. Next, we filtered the orthologous CpGs based on coverage to increase confidence in our data. We eliminated sites with > 10x coverage, as the relative sparsity of the data suggests that these reads were likely mapped to repeated regions. For each CpG site in each species/tissue combination, we required an average of 2x coverage and that at least 2 out of 4 individuals had a coverage >= 2x. After filtering based on coverage, we had 2.4M autosomal CpGs orthologous in all three species (10.5 million in humans and chimpanzees only).

Identifying differentially methylated regions (DMRs)

For a given pairwise comparison (e.g., human liver vs. human heart), the *bsseq* package produces a signal-to-noise statistic for each CpG site similar to a *t*-test statistic, assuming that DNA methylation levels in each condition have equal variance. As recommended by the authors of the package, we used a low-frequency mean correction to improve the marginal distribution of the *t*-statistics. Similar to previous studies using this methodology, a *t*-statistic cutoff of –4.6,4.6 was used for significance (Hansen et al. 2011; Hansen et al. 2014).

Overlap of tissue-specific DMRs with regulatory regions

We extracted the coordinates of the following features from GENCODE annotation (release 19)(Harrow et al. 2012): exons, first exons, CDS, 3' and 5' UTRs, introns, intergenic regions, promoters (-2 kb to +2 kb from TSS (Eckhardt et al. 2006; Zhou et al. 2014)) and

proximal promoters (-250 nt to +250 nt from TSS (Butler and Kadonaga 2002)) of all genes, or only of protein coding genes. We downloaded coordinates of the CpG islands from the UCSC Genome Browser (Karolchik et al. 2014). CpG islands (identified as segments of the genome with %G+C > 50%, length > 200 nt, and a ratio of observed over expected number of CG dinucleotides based on the number of Gs and Cs in the segment > 0.6 (Gardiner-Garden and Frommer 1987)). CpG island shores were defined as 2 kb regions flanking CpG islands, and CpG islands shelves were defined as 2 kb regions outside of CpG islands shores (Irizarry et al. 2009). To test the overlap of t-DMRs with enhancers, we used the set of tissue-specific enhancers defined by the FANTOM consortium using CAGE-seq data on primary tissues (Andersson et al. 2014). To control for the potential effects of CpG density and region length in these analyses, we generated 100 sets of randomly located control regions matching the length and CpG densities of the DMRs in the studied set.

We calculated the overlap of tissue-specific DMRs with H3K27ac in the left ventricle of the heart, kidney, liver, and lung adult tissues from the Epigenome Roadmap (The Roadmap Epigenomics Consortium 2015). We used the consolidated broad peak data for the left ventricle, liver, and lung (available from http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/broadPeak/ucsc_compatible/). Since a consolidated version was not available for the kidneys, we used unconsolidated kidney sample numbered 153 (available from http://egg2.wustl.edu/roadmap/data/byFileType/peaks/unconsolidated/broadPeak/ucsc_compatible/). We used BEDTools (version 2.26.0) (Quinlan and Hall 2010) to calculate the number of tDMRs that overlap an H3K27ac mark.

Supplementary Text

Assessing the impact of technical variables on gene expression levels and DNA methylation levels

We tested the relationship between our technical factors and biological variables of interest, namely tissue and species (see Methods). Through this process, we discovered that RNA extraction date was confounded with species (Supplemental Fig. S1B). In subsequent analysis with only the human samples (as humans were the only species with samples processed on multiple days), we found that the RNA extraction date did not highly correlate with tissue (Supplemental Fig S1C). We thought it unlikely that differences in RNA extraction date had a larger impact on variation in gene expression levels than tissue type.

Furthermore, the time of tissue collection post-mortem is also confounded with species (Supplemental Table S1). These differences could impact RNA quality, which can be approximated by RIN score. Indeed, RIN scores were typically higher in rhesus macaques than in the other species (Supplemental Table S1B; Supplemental Fig. S1A). As a result, we included RIN score as a covariate when modeling gene expression levels.

RNA quality may have impacted the number of DE genes identified between tissues. The number of pairwise DE genes was higher in rhesus macaque than in chimpanzee and human across FDR cutoffs (Supplemental Table S5B). This finding is potentially due to the higher sample quality, and therefore lower gene expression level variance, in rhesus macaques.

We also tested for associations between technical factors and biological variables of interest in the BS-seq data. Most of the significant associations were related to DNA methylation levels (e.g. number of orthologous CpGs sites with low methylation, mean methylation level at orthologous CpGs) across species and tissues. We expect the DNA methylation level densities to vary somewhat across tissues (Pai et al. 2011) and therefore, these inter-tissue differences are likely biological rather than technical.

Overall, we found slightly higher DNA methylation levels in human, compared to chimpanzee and macaque samples (average in humans = 0.664, in chimpanzees = 0.646, in

rhesus macaques = 0.626; Supplemental Fig. S4B), which persisted in the raw, unsmoothed data (Supplemental Fig. S4A). The distribution of DNA methylation levels could potentially be biased by CpG to TpG homozygous or heterozygous SNPs, which are erroneously inferred as unmethylated CpG sites. If the rate of such SNPs was higher in the chimpanzee or macaque individuals compared to human (with respect to their respective reference genome), we could observe differences between species. Since we had performed SNP calling on our RNA-seq dataset, we retained only CpG sites located in orthologous exons, and excluded sites with C to T SNPs in any of the samples. However, we still observed differences between species (Supplemental Fig. S4C). Moreover, higher DNA methylation rates in humans compared to chimpanzees were previously reported (but rarely discussed) in a diverse set of tissues, using various technologies to measure DNA methylation (Martin et al. 2011; Molaro et al. 2011; Hernando-Herraez et al. 2013; Hernando-Herraez et al. 2015). Therefore, this result may be driven by biases when mapping to different species' genomes.

Read coverage on the lambda phage genome was statistically significant (FDR < 10^{-10}). However, further analysis showed that this trend was driven by some low values in the chimpanzee samples, rather than the rhesus macaques (Supplemental Table S1D). Indeed, there is no difference in this factor across the human and the rhesus macaque samples (P = 0.15, Student's t-test). Therefore, we do not think that coverage on the lambda phage genome can account for the differences in DNA methylation between the great apes and the rhesus macaque samples (Figure 1D).

We found evidence for a dependent relationship between species and lane number (Chi squared test, FDR= 10^{-13}). Since these lanes were spread across multiple flow-cells, we do not think that lane substantially contributed to the variance in DNA methylation levels. We also found evidence for a dependent relationship between tissue and library preparation date (Chi squared test, FDR = 0.006). Since the correlation was modest (Pearson's correlation = -0.22) and most tissues within a species had libraries made on multiple days, we chose not to correct

for this variable. Finally, we note that sample age has previously been shown to impact methylation status in a subset of genes (Day et al. 2013; Horvath 2013). DNA methylation levels were weakly positively correlated with age (age quantile relative to the species' average lifespan; Pearson correlation's = 0.18). However, in our factor analysis, this relationship was non-significant (FDR > 10%). Therefore, we did not correct for this variable.

Tissue-specific gene expression patterns

We asked whether the tissue-specific gene expression patterns we found in a sample of four individuals from each species are indeed indicative of regulatory patterns in a larger population. To examine this, we again considered human GTEx data from the same four tissues we included in our study (see Methods). Because the sample size of the GTEx data is much larger than in our study, we compared the normalized gene expression ranks in the four tissues. For example, in both our and the GTEx data, troponin T2 (*TNNT2*) shows the highest expression in the heart (rank 1) and the lowest expression in the liver (rank 4). Using this approach, we found that 428 (62%) of the 687 genes with a tissue-specific expression pattern exclusively in our human data have the same tissue-based ranked expression in the GTEx data. This observation suggests that tissue-specific expression patterns found in just four individuals are quite often not representative of the regulatory patterns in the larger population. In contrast, however, we found that 1,530 (88%) of the 1,739 genes with a conserved tissue-specific expression pattern based on our data have the same tissue-based ranked expression in the GTEx data. Thus, conserved differential expression significantly increases the confidence of classifying tissue-specific expression patterns in a larger human population ($P < 10^{-16}$, difference of proportions test).

Adaptive shrinkage and false sign rate to identify tissue-specific genes

We investigated to what extent our ability to detect tissue-specific genes could be substantially impacted by differences in effect sizes across the species (Supplemental Table

S6). Therefore, we tested the use of an adaptive shrinkage method (Stephens 2017) to identify genes with a small effect size but consistent direction of effect in each species and used the accompanying false sign rate (FSR) instead of FDR thresholds. The percentage overlap was relatively robust to threshold method (Supplemental Table S6). We found that this method increases both the total number of tissue-specific differences and the species-specific differences. However, it also increases the number of conserved tissue-specific gene expression differences in humans and chimpanzees relative to those in chimpanzees and rhesus macaques (Supplemental Table S6), more closely reflecting established phylogenetic relationships.

Identifying inter-species differences between tissues

Since our data contained multiple tissues and species, we identified genes with inter-species differences between tissues (tissue-by-species interactions). These tissue-by-species interactions are potentially informative for great ape evolution (when the contribution of species on gene expression in a given tissue is different between great apes and rhesus) and the evolution of human-specific mechanisms in tissues (when the effect of species on gene expression in a given tissue is different between humans and a group containing chimpanzees and rhesus macaques). Using a linear-model based framework, we modeled these differences with tissue-by-species interaction terms (see Methods). We found 664 total significant interactions in the great ape versus rhesus macaque comparison and 91 in the human versus chimpanzee and rhesus macaque (FDR 1%; Supplemental Table S8). Given our sample size and the small effect sizes of these interactions, we are probably underpowered to detect such interactions. To address this, we employed an adaptive shrinkage method (Stephens 2017) to identify genes with a small effect size but consistent direction of effect in each species, and used FSR instead of FDR thresholds. This method was used to identify cases where the observed sign of the effect across tissues was different between species. After applying this

method, we found 1,006 great ape-by-tissue interactions and 257 human-by-tissue interactions (FSR = 1%; Supplemental Table S8).

Potentially the most interesting class of tissue-by-species interaction is when species impacts one tissue differently than the other three tissues. Therefore, we used *ashr* to find 799 great ape-by-tissue interactions and 249 human-by-tissue interactions only present in one tissue (FSR = 1%; Supplemental Table S8). We defined tissue-by-species specific interactions as interactions with an effect size sign different from the signs of the other interactions (e.g. a positive sign when all other signs are 0 or negative). Unsurprisingly, even after accounting for small effect sizes, there were more tissue-by-species interactions for great apes versus rhesus macaques than human-specific ones.

Promoter DNA methylation quality

To check our promoter DNA methylation levels in the humans and chimpanzees, we subset the DNA methylation promoter data to the 3 human and chimpanzee tissues tested in a previous study from our lab (Pai et al. 2011). Consistent with this previous study, PC1 was more highly correlated with tissue than species and PC2 was more highly correlated with species than tissue (Supplemental Fig. S6A). Even in this subset of the data, there was more clear separation between tissues in the gene expression levels than the promoter DNA methylation data for these genes (Supplemental Fig. S6B).

Identification of DMRs across species (S-DMRs)

Using the same method to identify DMRs across tissues, we then identified thousands of DMRs across species (S-DMRs). We found the lowest number of S-DMRs on autosomal chromosomes in lungs (8,617 DMRs between human and chimpanzees, 17,696 DMRs between humans and rhesus macaques, and 15,544 between chimpanzees and rhesus macaques) and highest total number in hearts (14,504 DMRs between human and chimpanzees, 25,539 DMRs

445 between humans and rhesus macaques, and 15,544 between chimpanzees and rhesus
446 macaques, Table 2). Similar to the pairwise DE analysis across species, the number of DMRs
447 between species are consistent with known phylogenetic relationships. However, unlike in the
448 pairwise DE analysis across species, the number of S-DMRs is sometimes higher than the
449 number of pairwise T-DMRs. For example, there are more lung S-DMRs than human heart-lung
450 DMRs. This trend is somewhat unexpected given the gene expression data, but consistent with
451 clustering pattern of the DNA methylation data (Figure 1D).
452

453 **References**
454
455 Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from
456 gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**:
457 1600-1607.
458 Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y,
459 Zhao X, Schmidl C, Suzuki T et al. 2014. An atlas of active enhancers across
460 human cell types and tissues. *Nature* **507**: 455-461.
461 Ballman K, Grill D, Oberg A, Therneau T. 2004. Faster cyclic loess: normalizing RNA
462 arrays via linear models. *Bioinformatics* **20**: 2778-2786.
463 Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and
464 Powerful Approach to Multiple Testing. *J Royal Stat Soc* **57**: 289-300.
465 Blischak JD, Tailleux L, Mitrano A, Barreiro LB, Gilad Y. 2015. Mycobacterial infection
466 induces a specific human innate immune response. *Sci Rep* **5**: 16882.
467 Bock C. 2012. Analysing and interpreting DNA methylation data. *Nat Rev Genet* **13**:
468 705-719.
469 Butler JEF, Kadonaga JT. 2002. The RNA polymerase II core promoter: a key
470 component in the regulation of gene expression. *Genes Dev* **16**: 2583-2592.
471 Chatterjee A, Stockwell PA, Rodger EJ, Morison IM. 2012. Comparison of alignment
472 software for genome-wide bisulphite sequence data. *Nucleic Acids Res*
473 doi:10.1093/nar/gks150.
474 Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD, Myers RM,
475 Absher D. 2013. Differential DNA methylation with age displays both common
476 and dynamic features across human tissues that are influenced by CpG
477 landscape. *Genome Biol* **14**: R102.

478 Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV,
479 Davies R, Down TA et al. 2006. DNA methylation profiling of human
480 chromosomes 6, 20 and 22. *Nat Genet* **38**: 1378-1385.
481 Gallego Romero I, Pavlovic BJ, Hernando-Herraez I, Zhou X, Ward MC, Banovich NE,
482 Kagan CL, Burnett JE, Huang CH, Mitrano A et al. 2015. A panel of induced
483 pluripotent stem cells from chimpanzees: a resource for comparative functional
484 genomics. *Elife* **4**: e07103. doi: 10.7554/eLife.07103.
485 Gardiner-Garden M, Frommer M. 1987. CpG Islands in vertebrate genomes. *J Mol Biol*
486 **196**: 261-282.
487 The GTEx Consortium. 2017. Genetic effects on gene expression across human tissues.
488 *Nature* **550**: 204–213.
489 Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite
490 sequencing reads to differentially methylated regions. *Genome Biol* **13**: R83.
491 Hansen KD, Sabunciyan S, Langmead B, Nagy N, Curley R, Klein G, Klein E, Salamon
492 D, Feinberg AP. 2014. Large-scale hypomethylated blocks associated with
493 Epstein-Barr virus-induced B-cell immortalization. *Genome Res* **24**: 177-184.
494 Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B,
495 Wu H, Liu Y, Diep D et al. 2011. Increased methylation variation in epigenetic
496 domains across cancer types. *Nat Genet* **43**: 768-775.
497 Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL,
498 Barrell D, Zadissa A, Searle S et al. 2012. GENCODE: the reference human
499 genome annotation for The ENCODE Project. *Genome Res* **22**: 1760-1774.
500 Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H,
501 Prado-Martinez J, Sharp AJ, Esteller M, Marques-Bonet T. 2015. The interplay

between DNA methylation and sequence divergence in recent human evolution. *Nucleic Acids Res* doi:10.1093/nar/gkv693.

Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, Navarro A, Esteller M, Sharp AJ, Marques-Bonet T. 2013. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet* **9**: e1003763. doi: 10.1371/journal.pgen.1003763.

Horvath S. 2013. DNA methylation age of human tissues and cell types. *Genome Biol* **14**: R115.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M et al. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**: 178-186.

Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M et al. 2014. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* **42**: D764-770.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571-1572.

Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**: R29.

Martin D, Singer M, Dhahbi J, Mao G, Zhang L, Schroth G, Pachter L, Boffelli D. 2011. Phyloepigenomic comparison of great apes reveals a correlation between somatic and germline methylation states. *Genome Res* **21**: 2049-2057.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297-1303.

Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**: 1029-1041.

Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y. 2011. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet* **7**: e1001316. doi: 10.1371/journal.pgen.1001316.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, Zumbo P, Mason CE, Socci ND, Betel D. 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* **14**: R95.

Ritchie M, Phipson B, Wu D, Hu Y, Law C, Shi W, Smyth G. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**: e47. doi: 10.1093/nar/gkv1007.

The Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317-330.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: 11:R25.

Saha A, Kim Y, Gewirtz ADH, Jo B, Gao C, McDowell IC, Consortium GT, Engelhardt BE, Battle A. 2017. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res* **27**: 1843-1858.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.

Smyth GK, Michaud J, Scott HS. 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**: 2067-2075.

Soneson C, Delorenzi M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14**: 91.

Stephens M. 2017. False discovery rates: a new deal. *Biostatistics* **18**: 275-294.

Tran H, Porter J, Sun MA, Xie H, Zhang L. 2014. Objective and comprehensive evaluation of bisulfite short read mapping tools. *Adv bioinformatics* **2014**: 472045.

Tsuji J, Weng Z. 2016. Evaluation of preprocessing, mapping and postprocessing algorithms for analyzing whole genome bisulfite sequencing data. *Brief Bioinform* **17**: 938-952.

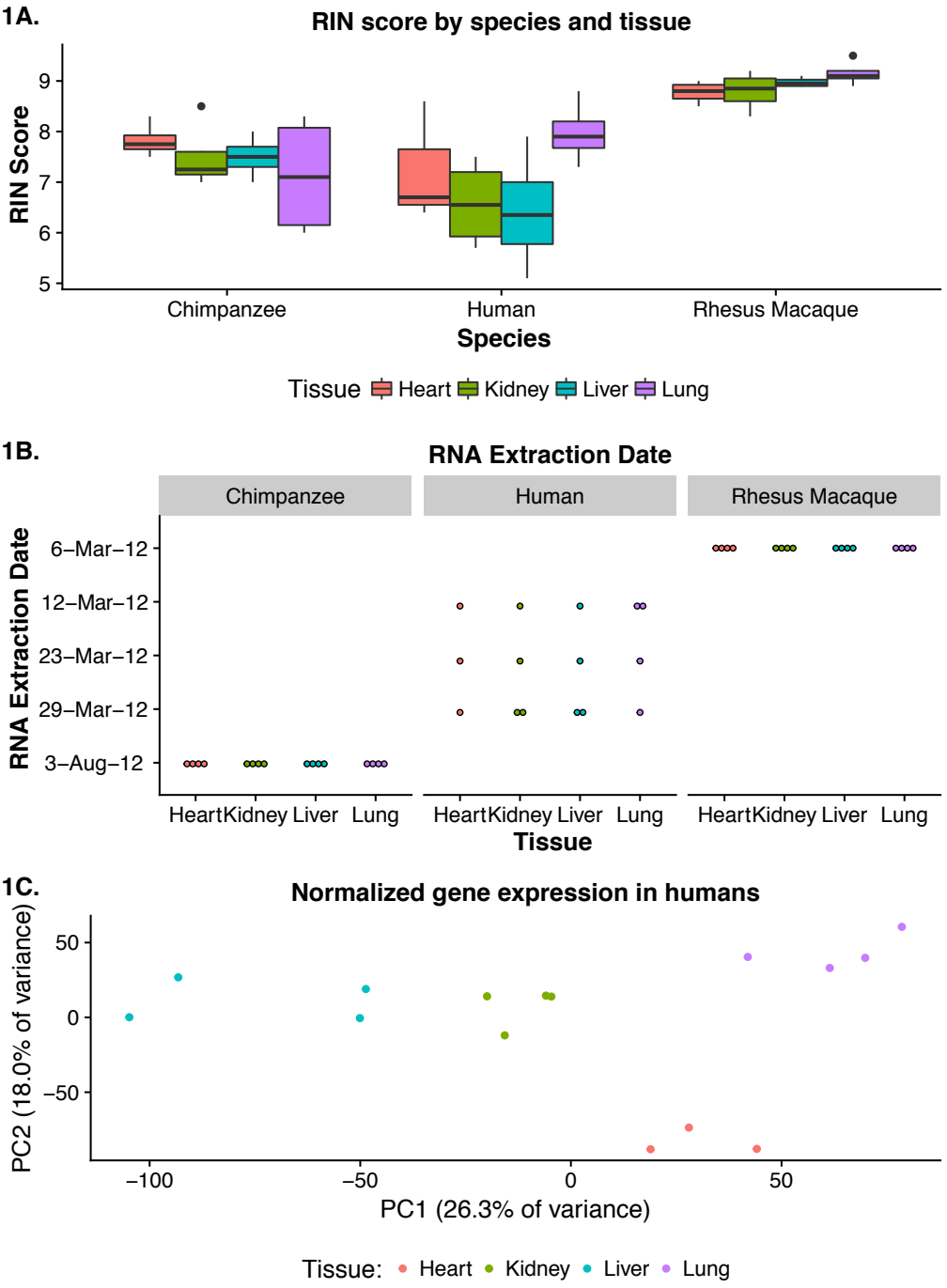
Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A et al. 2015. Proteomics. Tissue-based map of the human proteome. *Science* **347**: 1260419.

van de Geijn B, McVicker G, Gila Y, Pritchard J. 2015. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods* **12**: 1061-1063.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**: 11 10 11-33.

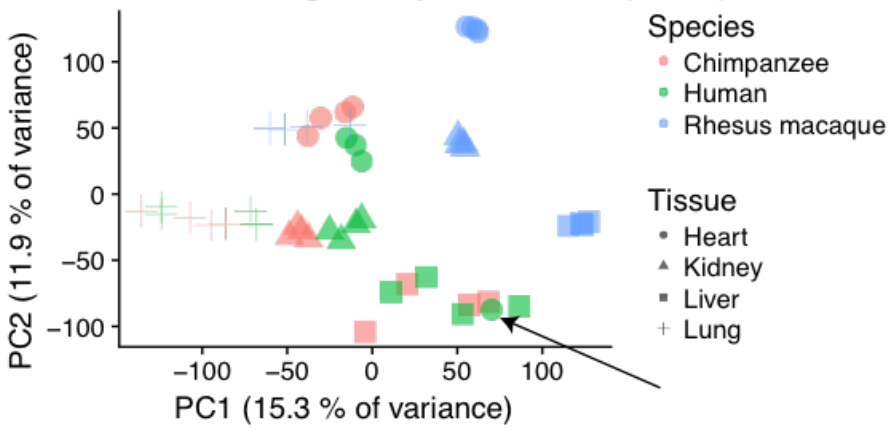
Zhou X, Cain CE, Myrthil M, Lewellen N, Michelini K, Davenport ER, Stephens M, Pritchard JK, Gilad Y. 2014. Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome Biol* **15**: 547.

Ziller MJ, Hansen KD, Meissner A, Aryee MJ. 2015. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat Meth* **12**: 230-2.

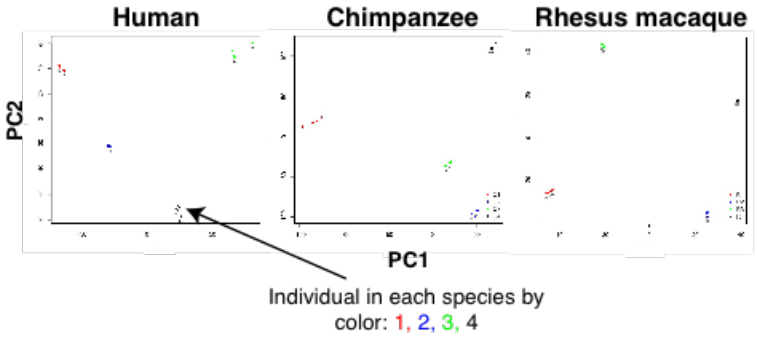


583
584 **S1. Distributions of potential confounders across biological variables of interest.** (A) RIN
585 score across the samples. (B) RNA extraction date by species. (C) PCA of RNA extraction date
586 in humans.
587

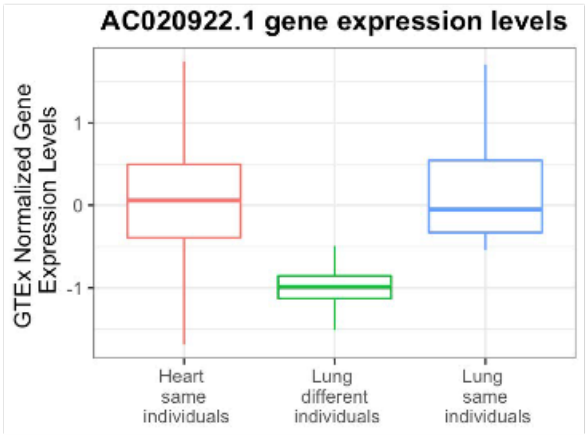
2A.



2B.

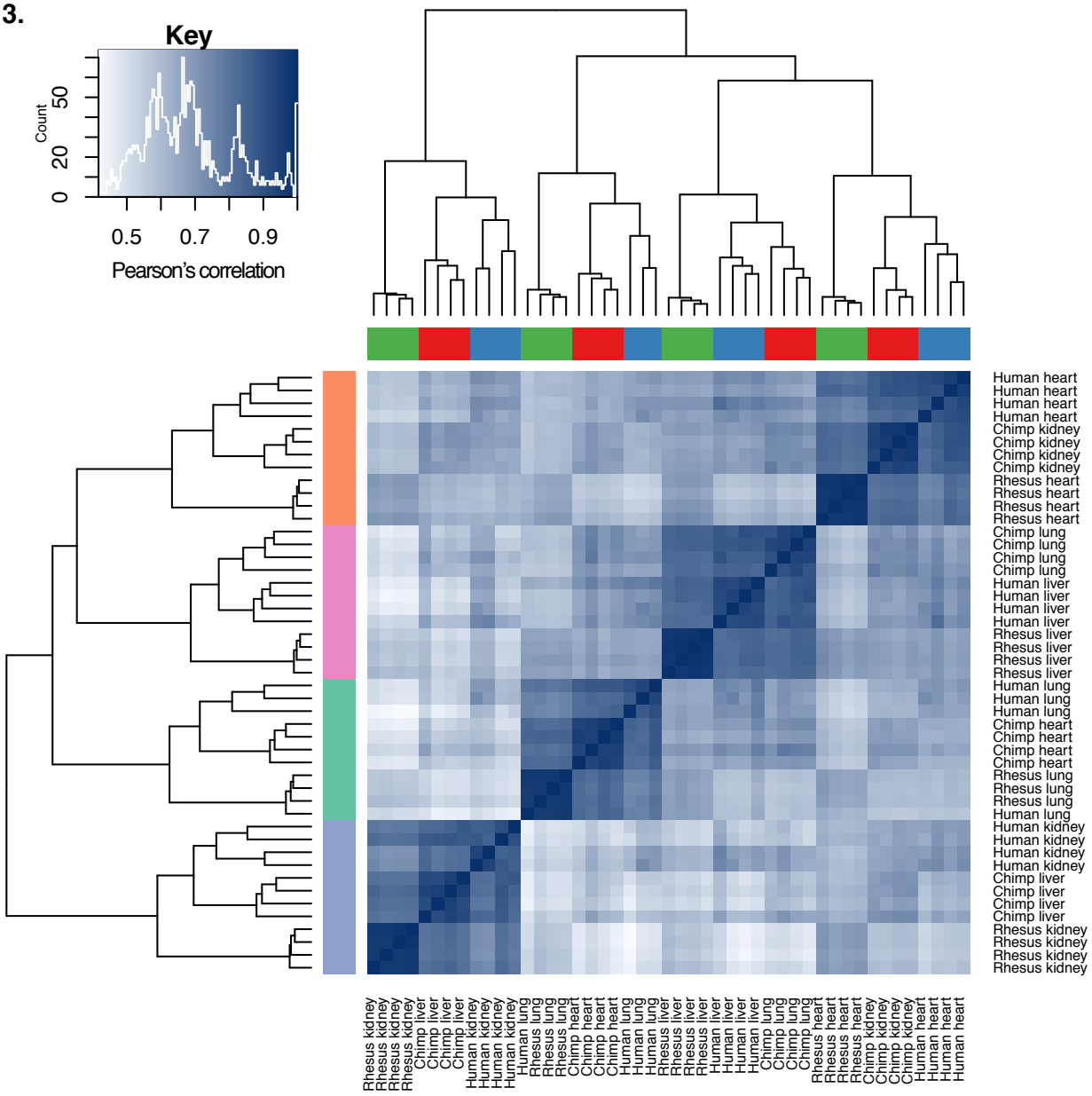


2C.

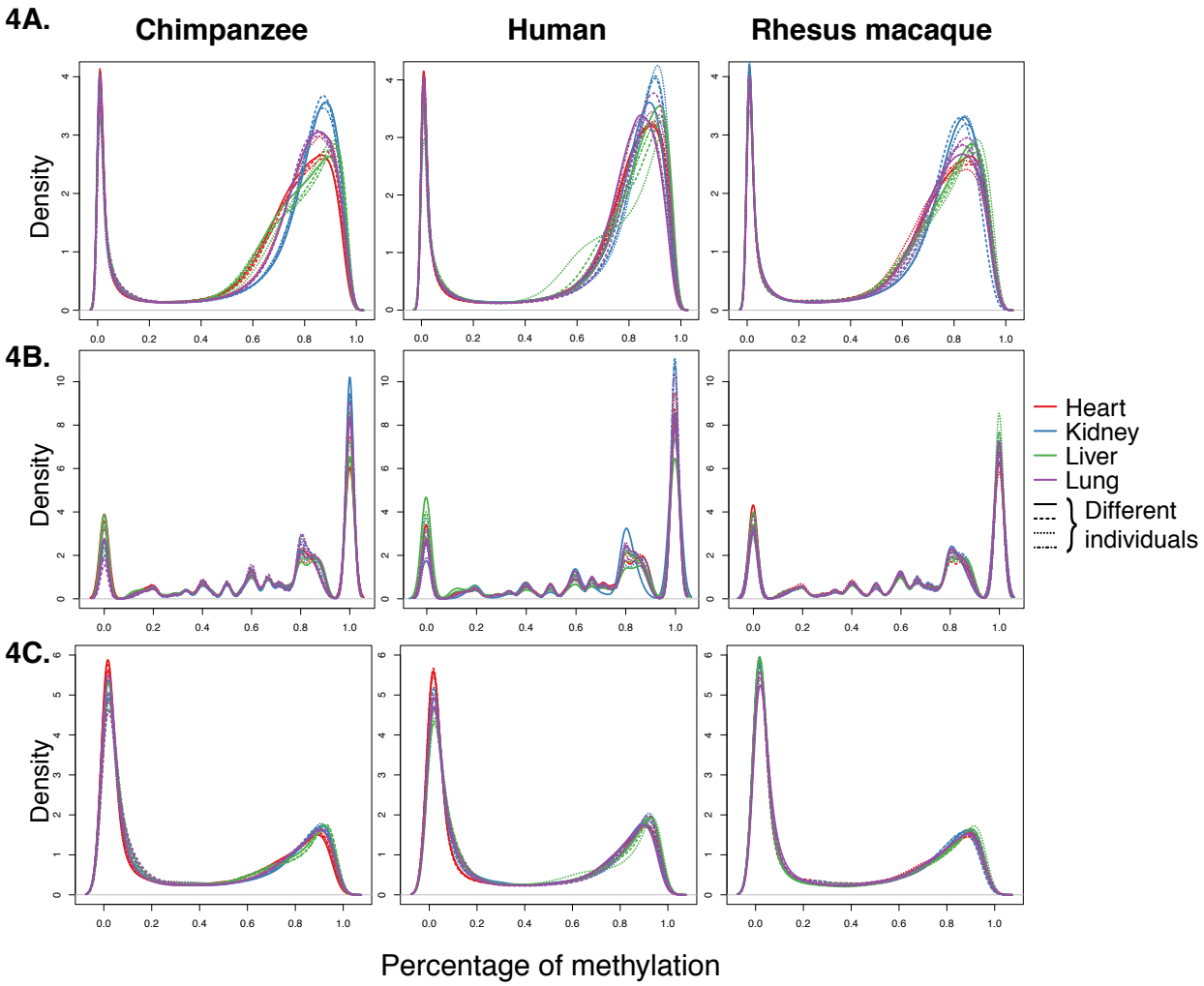


S2. Sample QC. (A) One human heart (green circle, see arrow) clusters with the human livers (green squares). The sample originally labeled as H1H ("human 1 heart") is likely a liver from the same human. (B) GATK analysis of the sample labelled H1H (see arrow) clusters with H1L, "human 1 liver". (C) Gene expression levels from different tissues in the same individual ("different tissue") are more highly correlated than gene expression levels between tissues from different individuals ("both different") and all combinations of tissues and individuals ("all"). Within a given tissue, gene expression levels are most highly correlated ("different individual").

3.

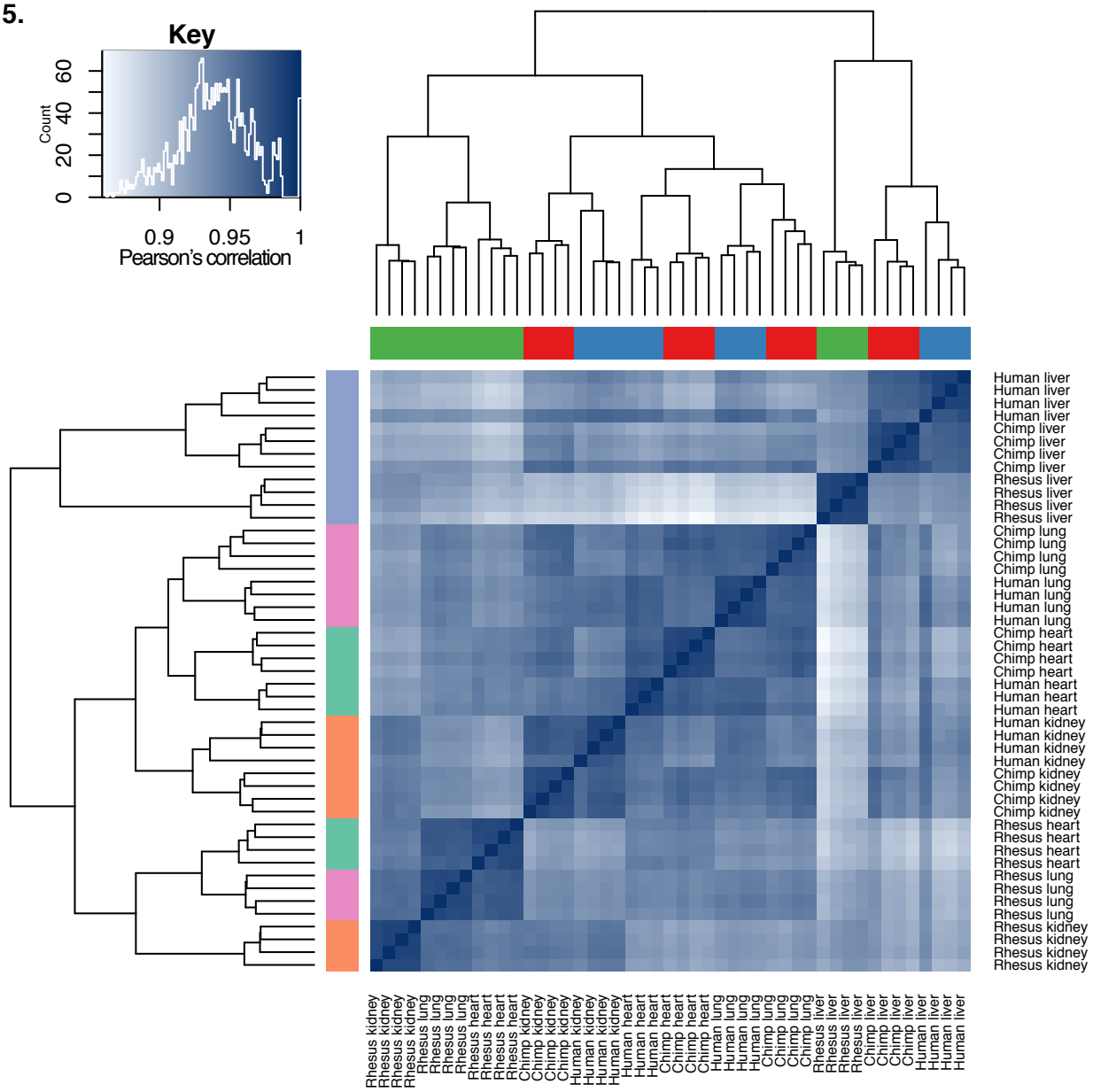


S3. Correlation matrix of normalized log2(CPM) gene expression values from 12,184 genes.

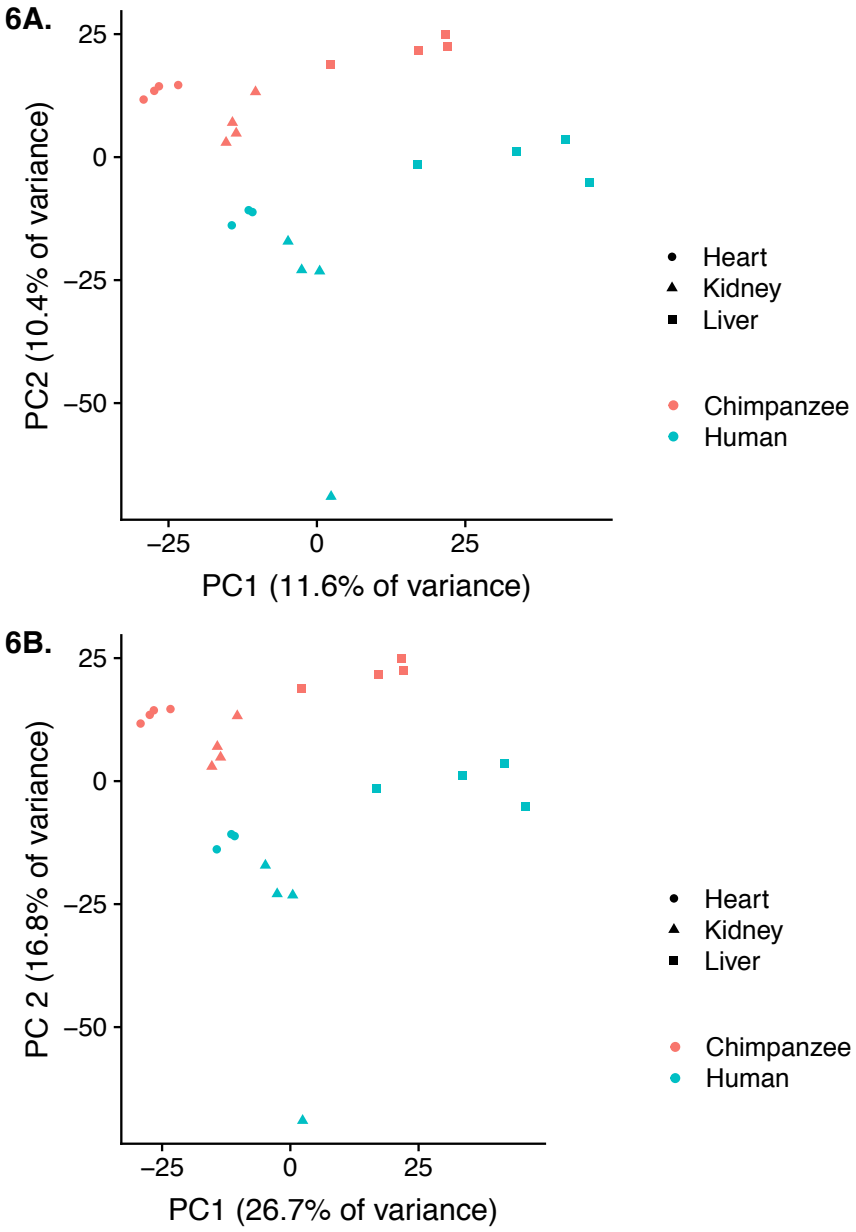


S4. Density function of DNA methylation levels across all species and tissues. (A) Using raw methylation estimates at the subset of orthologous CpG sites showing a read coverage of at least 5× and no more than 10× in each sample. (B) Using smoothed methylation estimates at all orthologous CpG sites across the three species. (C) Using orthologous CpG sites located in orthologous exons, and excluding sites with C to T SNPs in any of the samples.

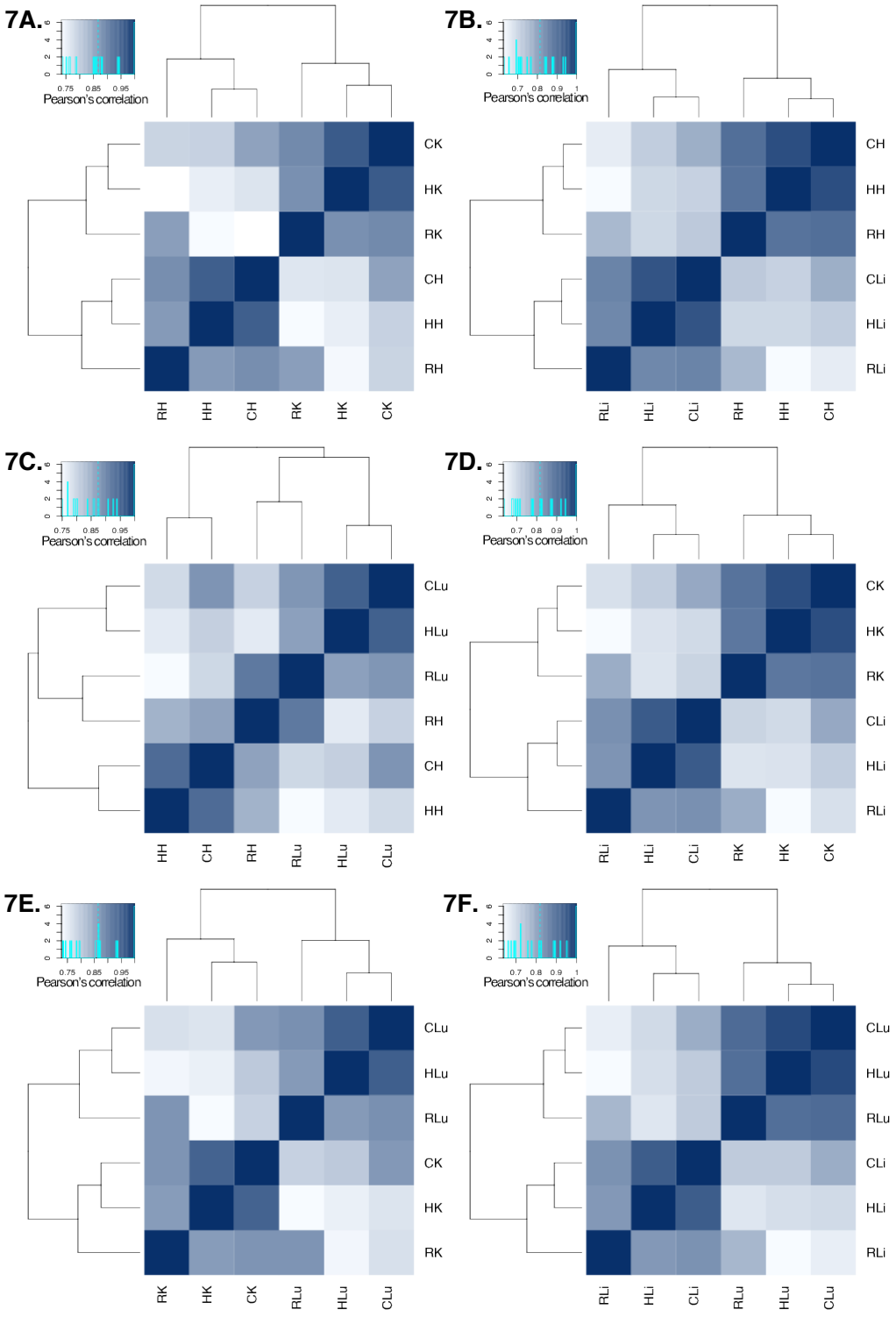
5.



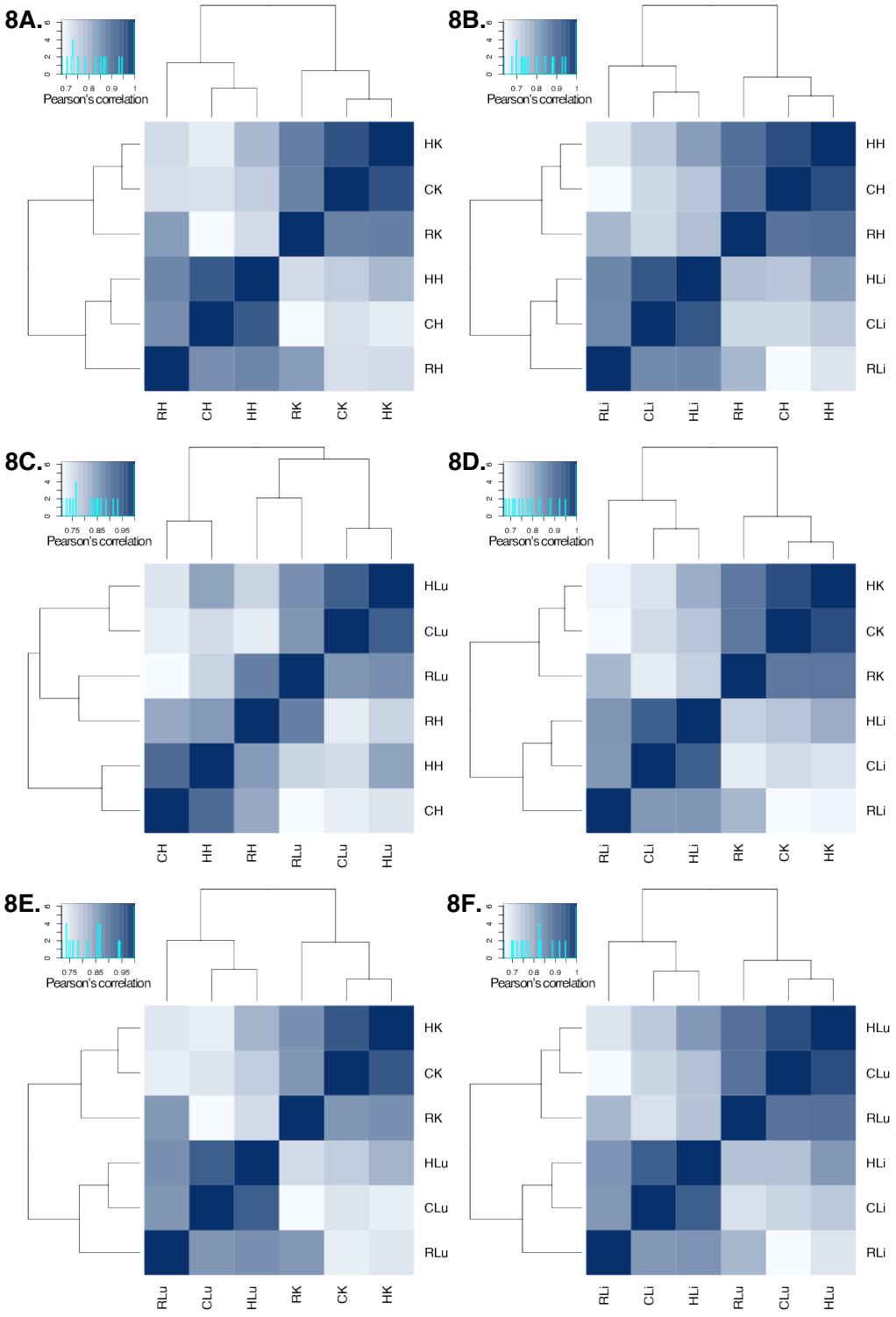
S5. Correlation matrix of smoothed DNA methylation levels from all orthologous CpGs.



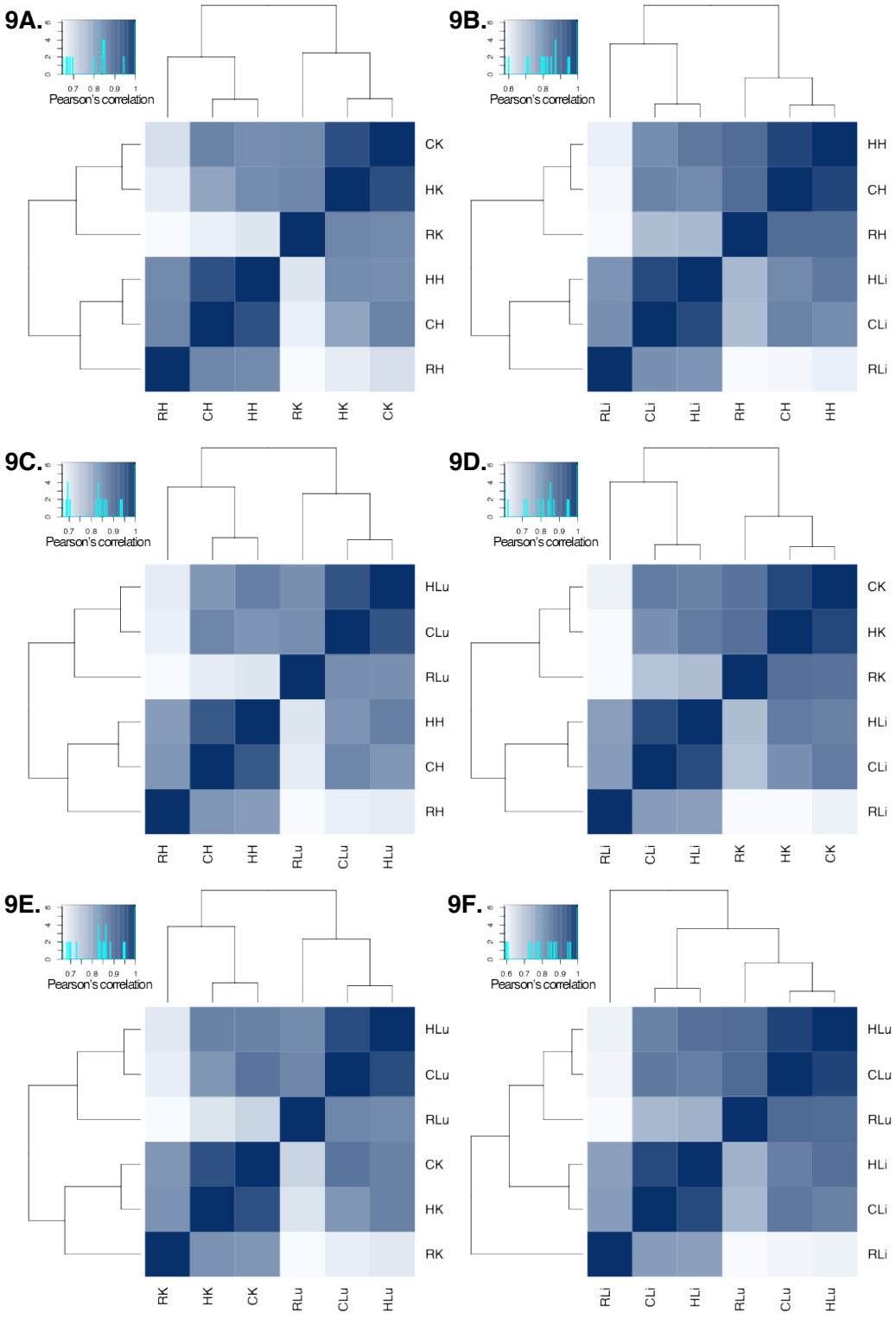
S6. Principal components analysis (PCA) in humans and chimpanzee hearts, kidneys, and livers. (A) Average promoter DNA methylation values. (B) Gene expression levels in the same genes from (A).



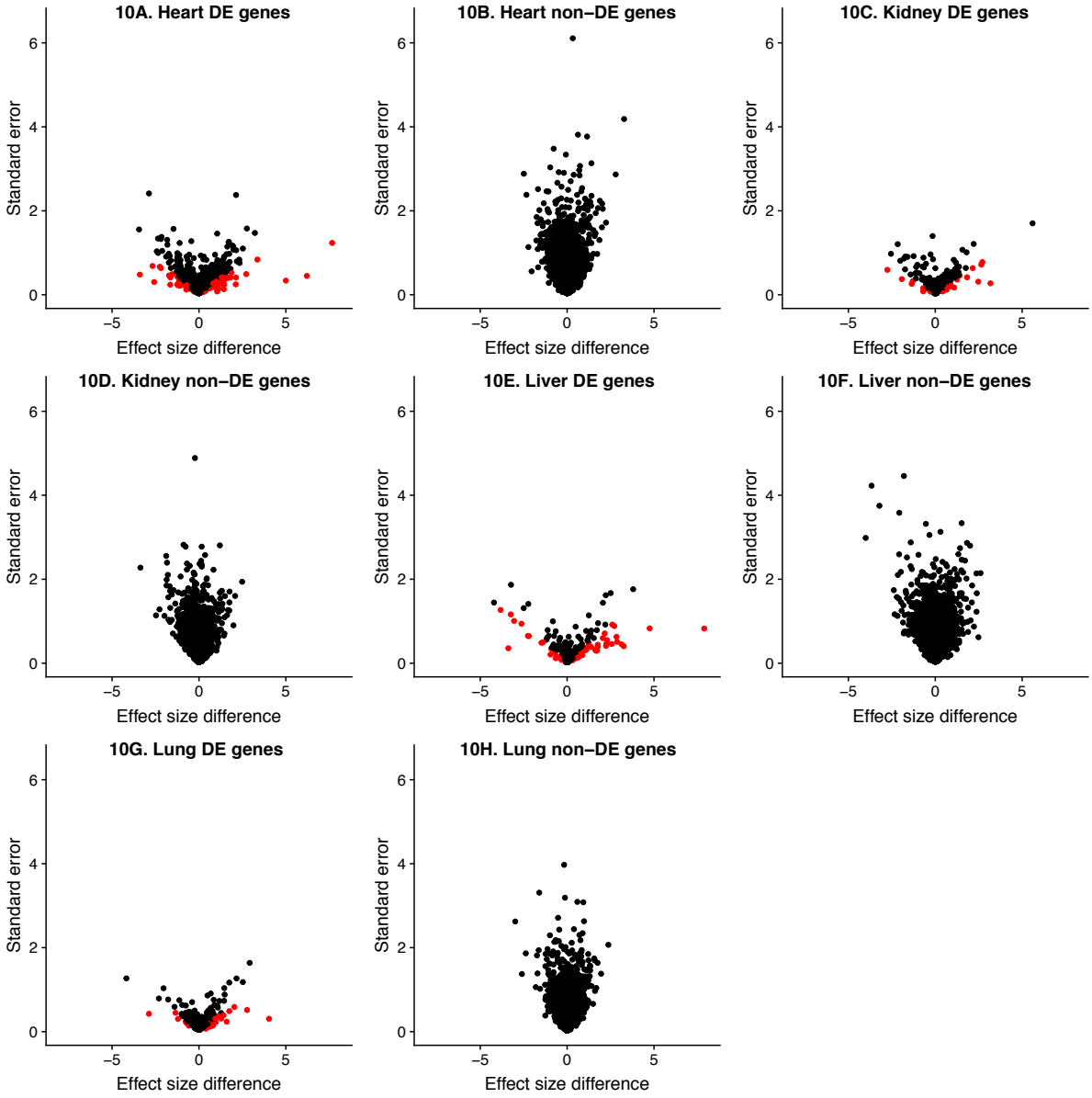
S7. When comparing DNA methylation levels of human T-DMRs and of orthologous regions in the same tissues, clustering is more highly correlated with tissue than species. Heatmaps are based on human (A) heart-kidney T-DMRs, (B) heart-liver T-DMRs, (C) heart-lung T-DMRs, (D) kidney-liver T-DMRs, (E) kidney-lung T-DMRs, and (F) liver-lung T-DMRs.

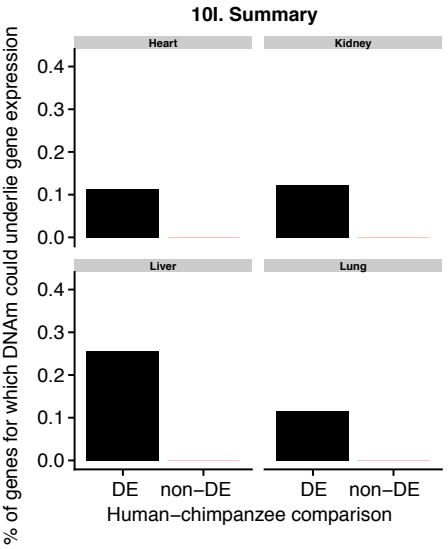


S8. When comparing DNA methylation levels of chimpanzee T-DMRs and of orthologous regions in the same tissues, clustering is more highly correlated with tissue than species. Heatmaps are based on chimpanzee (A) heart-kidney T-DMRs, (B) heart-liver T-DMRs, (C) heart-lung T-DMRs, (D) kidney-liver T-DMRs, (E) kidney-lung T-DMRs, and (F) liver-lung T-DMRs.

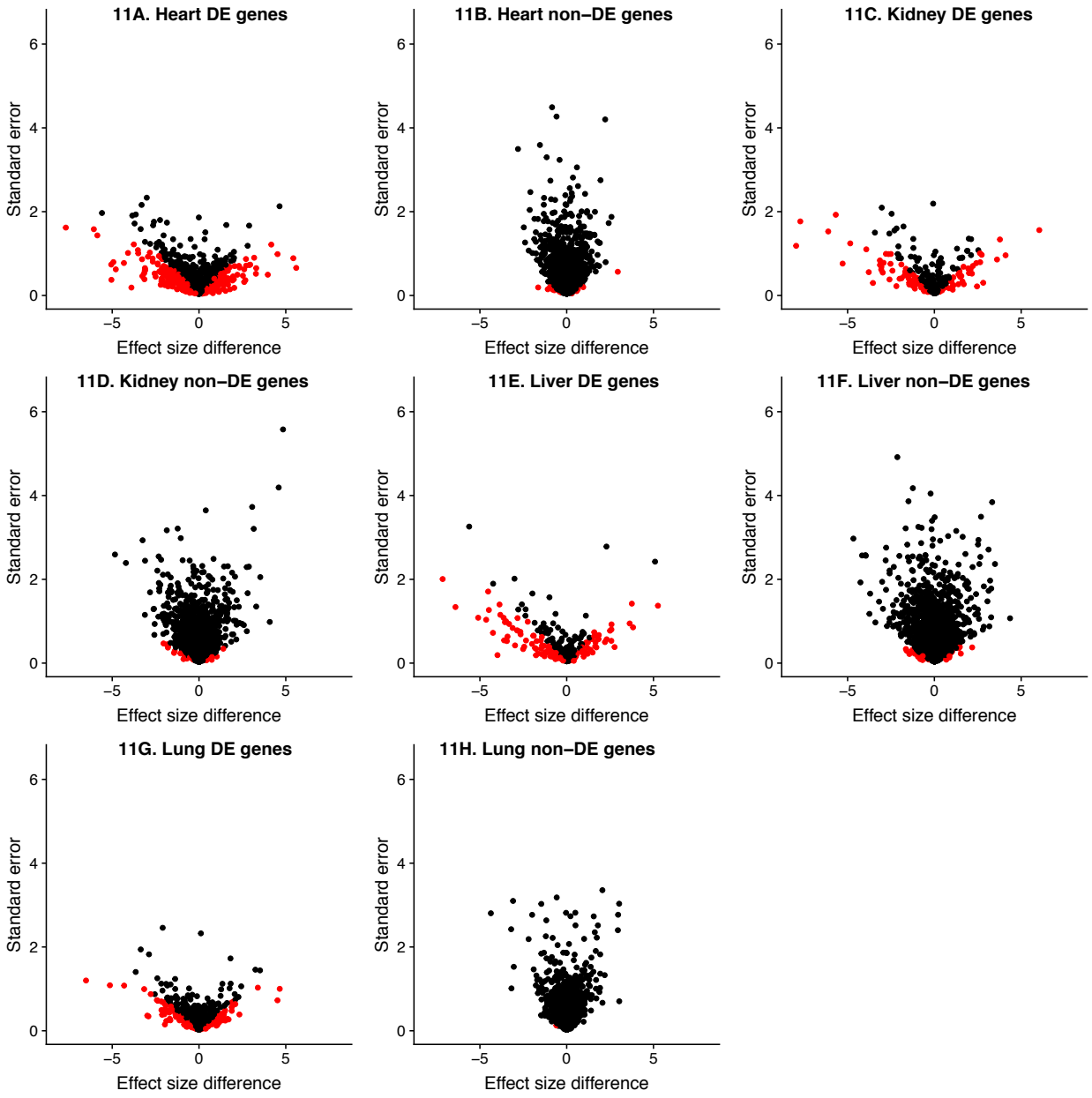


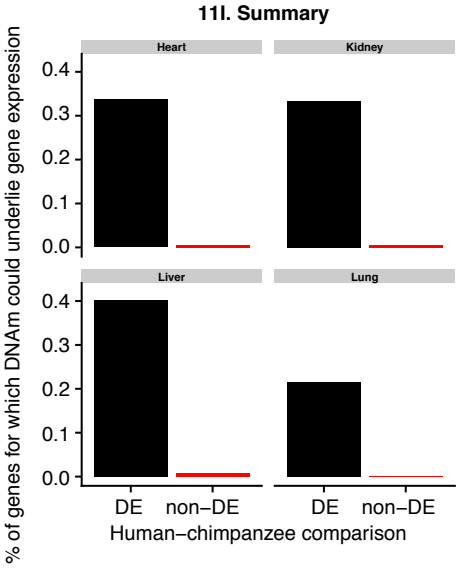
S9. When comparing DNA methylation levels of rhesus macaque T-DMRs and of orthologous regions in the same tissues, clustering is more highly correlated with tissue than species. Heatmaps are based on rhesus macaques (A) heart-kidney T-DMRs, (B) heart-liver T-DMRs, (C) heart-lung T-DMRs, (D) kidney-liver T-DMRs, (E) kidney-lung T-DMRs, and (F) liver-lung T-DMRs.





S10. Inter-species DNA methylation and gene expression levels (FDR = 0.05 and FSR = 0.05), in humans and chimpanzees. Difference in species effect size before and after accounting for DNA methylation levels in genes (A) DE and (B) non-DE in the human and chimpanzee heart, (C) DE and (D) non-DE in the human and chimpanzee kidney, (E) DE and (F) non-DE in the human and chimpanzee liver, (G) DE and (H) non-DE in the human and chimpanzee lung. (I) The percentage of genes for which the evidence for inter-species differences in gene expression levels is reduced after correcting for DNA methylation levels in humans and chimpanzees.





S11. Inter-species DNA methylation and gene expression levels (FDR = 0.05 and FSR = 0.05), in humans and rhesus macaques. Difference in species effect size before and after accounting for DNA methylation levels in genes (A) DE and (B) non-DE in the human and rhesus heart, (C) DE and (D) non-DE in the human and rhesus kidney, (E) DE and (F) non-DE in the human and rhesus liver, (G) DE and (H) non-DE in the human and rhesus lung. (I) The percentage of genes for which the evidence for inter-species differences in gene expression levels is reduced after correcting for DNA methylation levels in humans and rhesus macaques.