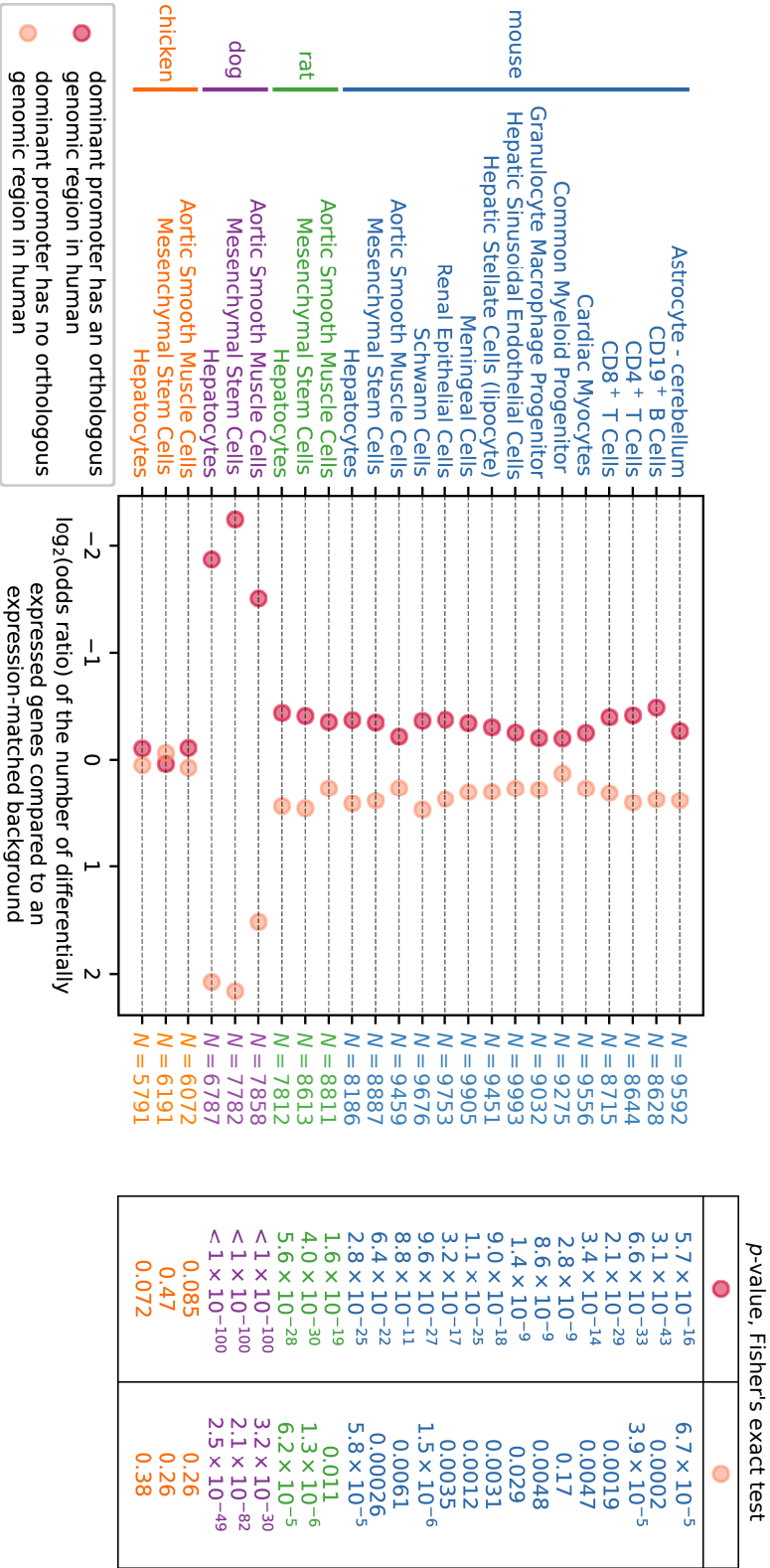


# Supplemental Figures

This file contains the following Supplemental Figures:

|                         |   |
|-------------------------|---|
| Supplemental Figure S1  | Promoter analysis of differentially expressed genes compared to an expression-matched background.   |
| Supplemental Figure S2  | Conservation analysis of differentially expressed genes compared to an expression-matched background.   |
| Supplemental Figure S3  | Integrative Correlation Analysis.   |
| Supplemental Figure S4  | Gene expression correlation of human and mouse samples.   |
| Supplemental Figure S5  | Gene expression correlation of human, mouse, rat, dog, and chicken samples.   |
| Supplemental Figure S6  | Phylogenetic analysis of expression divergence.   |
| Supplemental Figure S7  | Promoter analysis of gene expression correlations.  |
| Supplemental Figure S8  | Promoter analysis of gene expression correlations compared to an expression-matched background.   |
| Supplemental Figure S9  | Conservation analysis of gene expression correlations.  |
| Supplemental Figure S10 | Conservation analysis of gene expression correlations compared to an expression-matched background.   |
| Supplemental Figure S11 | Gene Ontology analysis of gene expression correlations.   |
| Supplemental Figure S12 | Analysis of RNA-Seq expression data for endometrial stromal fibroblast primary cells (Kin et al., 2016).  |
| Supplemental Figure S13 | Comparative analysis of RNA-Seq expression data for 12 matching tissues in human and mouse (The ENCODE Project Consortium, 2012).   |
| Supplemental Figure S14 | Evaluation of predicted transcription factor binding sites.   |
| Supplemental Figure S15 | Pearson's correlation across cell types between orthologous genes, as a function of TFBS conservation in their promoter regions.  |
| Supplemental Figure S16 | Cumulative distribution of the Pearson's correlation $r$ across cell types in motif activity between promoters and enhancers in human, mouse, rat, dog, and chicken.                          |
| Supplemental Figure S17 | Comparison of CAGE expression levels of the pri-miRNA to the mature miRNA expression levels measured by sRNA sequencing in aortic smooth muscle cells in human, mouse, rat, dog, and chicken. |
| Supplemental Figure S18 | Comparison of differential expression analysis of pri-miRNAs and mature miRNAs.   |

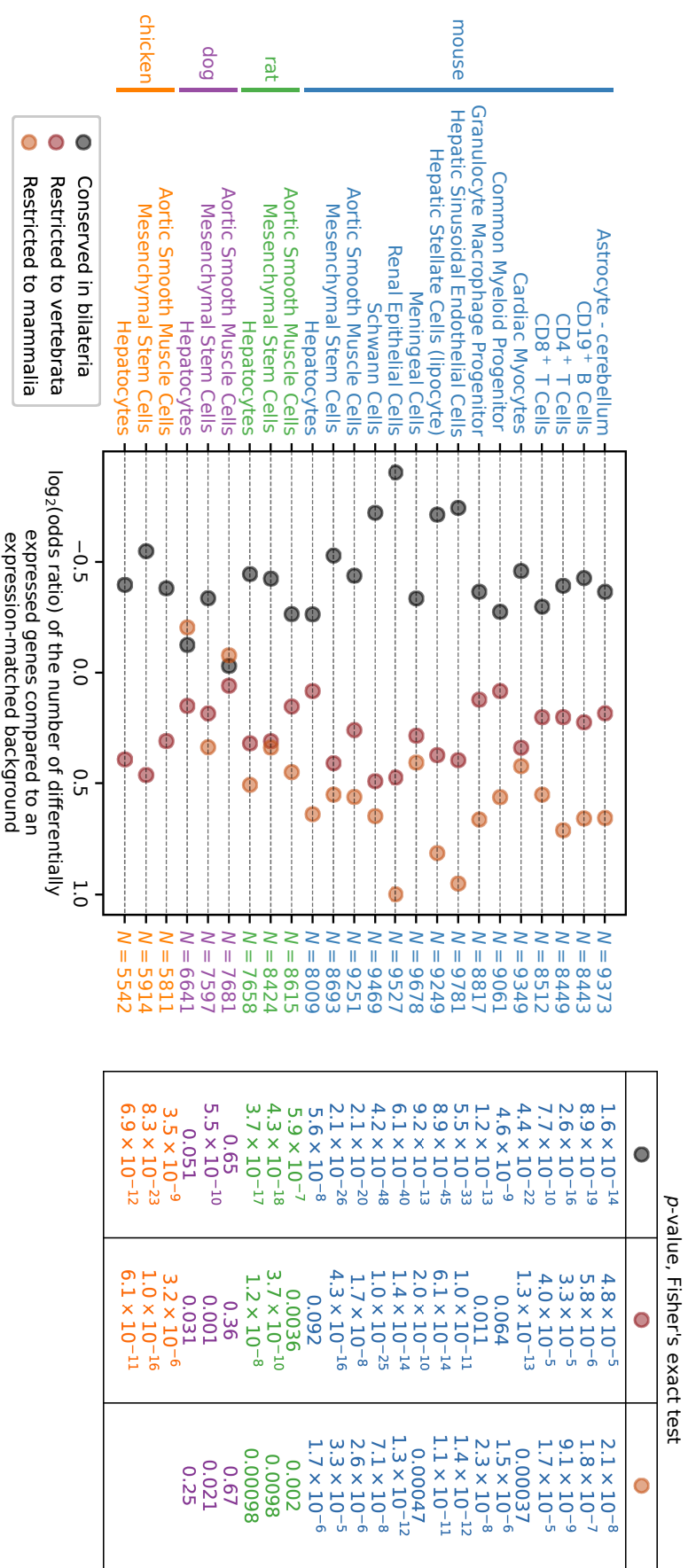




**Supplemental Figure S1.**

Promoter analysis of differentially expressed genes compared to an expression-matched background. Genes were divided into two categories depending on whether the genomic region of the dominant promoter of the gene had an orthologous genomic region in the human genome. The enrichment was calculated by comparing the number of differentially expressed genes in each of the two categories to the number of differentially expressed genes in an expression-matched background, with the log<sub>2</sub>(odds ratio) shown on the horizontal axis of the left panel and *p*-value calculated using Fisher's exact test for each comparison in the right panel. *N* is the total number of expressed genes in each cell type.

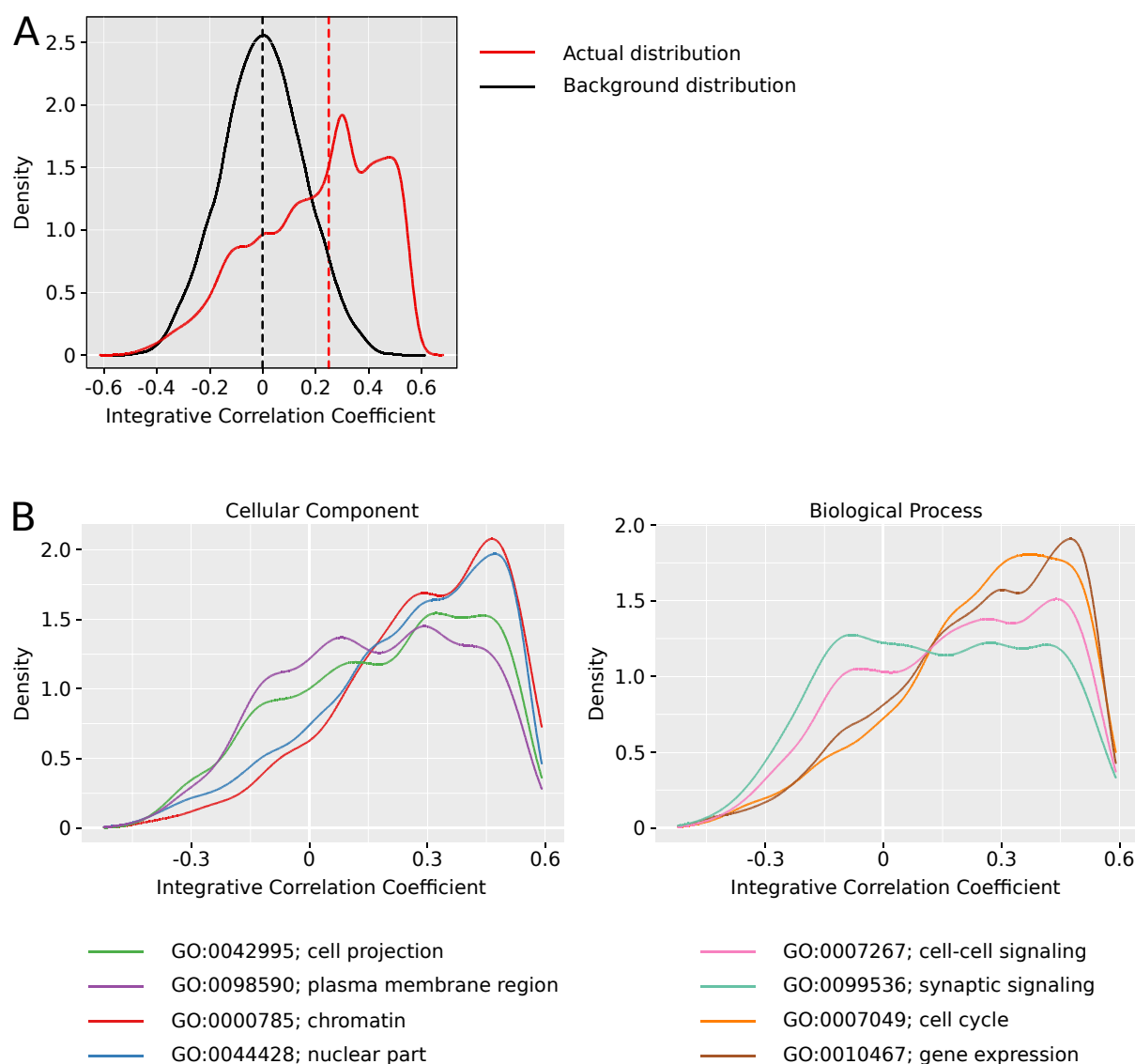




**Supplemental Figure S2.**

Conservation analysis of differentially expressed genes compared to an expression-matched background. Genes were divided into three categories based on the age of the most recent common ancestor. The enrichment was calculated by comparing the number of differentially expressed genes to the number of differentially expressed genes in an expression-matched background, with the log<sub>2</sub>(odds ratio) shown on the horizontal axis of the left panel and *p*-value calculated using Fisher's exact test for each comparison in the right panel. *N* is the total number of expressed genes in each cell type with an annotation in the NCBI HomoloGene database.





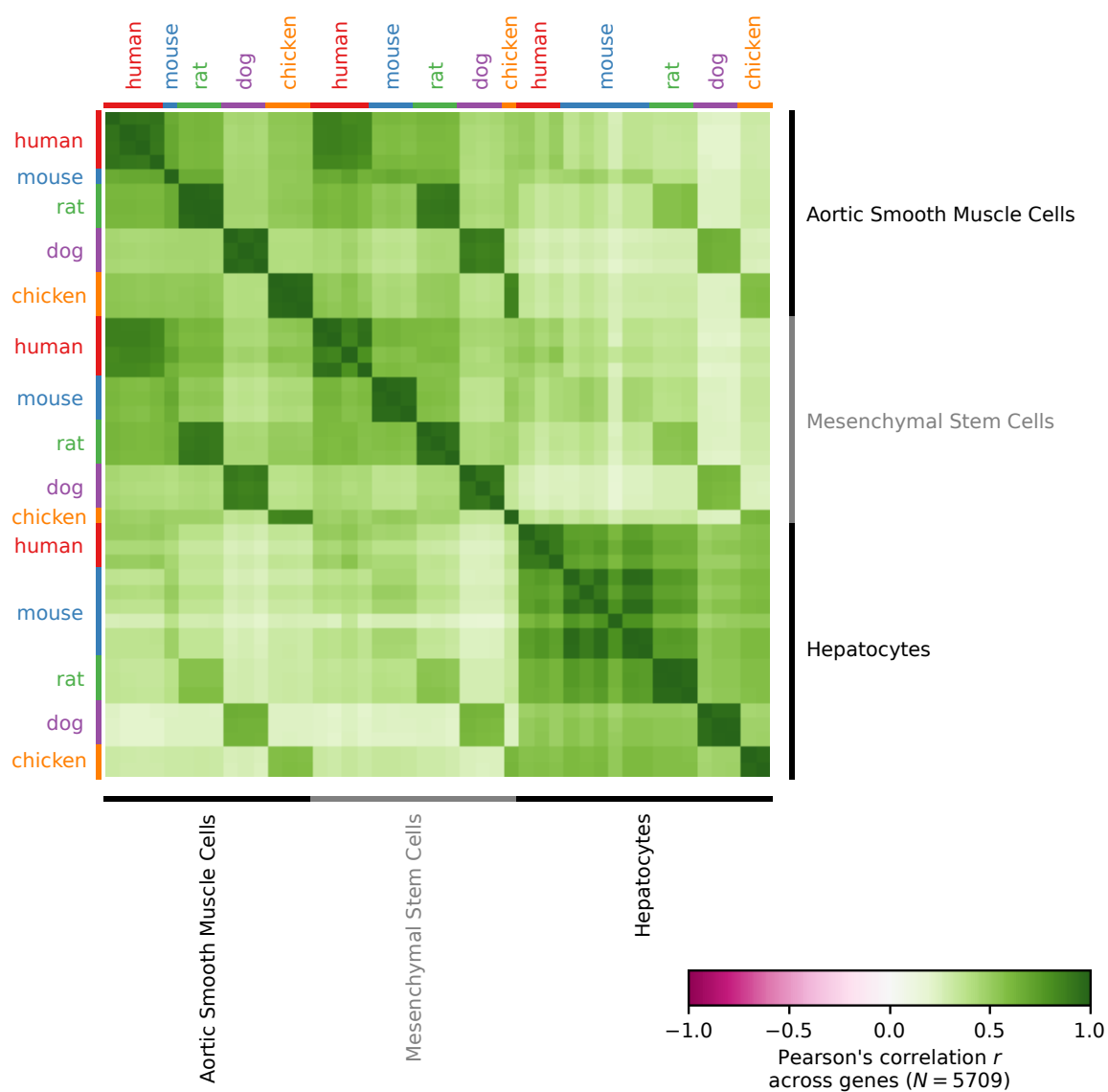
### Supplemental Figure S3.

**Integrative Correlation Analysis.** (A) Distribution of the integrative correlation coefficient computed for 15,538 expressed genes in common between human and mouse. The actual distribution is shown in red, while the null distribution obtained by permutation is depicted in black. Dashed lines indicate the median of each distribution. (B) Examples the distribution of the integrative correlation coefficient between human and mouse for genes belonging to specific Gene Ontology terms identified as significant ( $FDR < 0.05$ ) by Analysis of Functional Annotation. The distribution for Gene Ontology terms associated the key cellular processes and components (chromatin, nuclear part, cell cycle, and gene expression) are shifted to right suggesting a higher degree of conservation, while processes related to cell-to cell interaction (cell-cell and synaptic signaling, cell projection and plasma membrane region) are skewed to the left, suggesting a lesser degree of conservation across species.





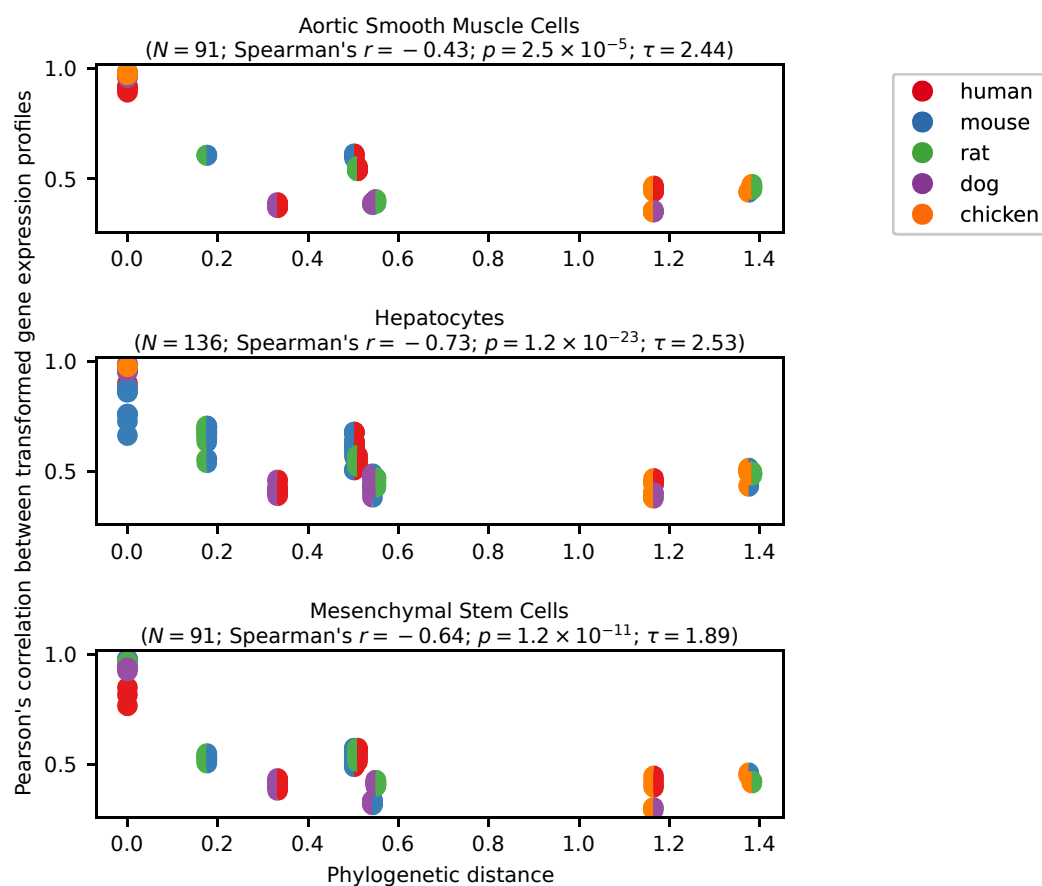




### Supplemental Figure S5.

Gene expression correlation of human, mouse, rat, dog, and chicken samples. The heatmap visualizes the value of Pearson's correlation coefficient across genes between all samples for cell types with CAGE expression data in all five species.

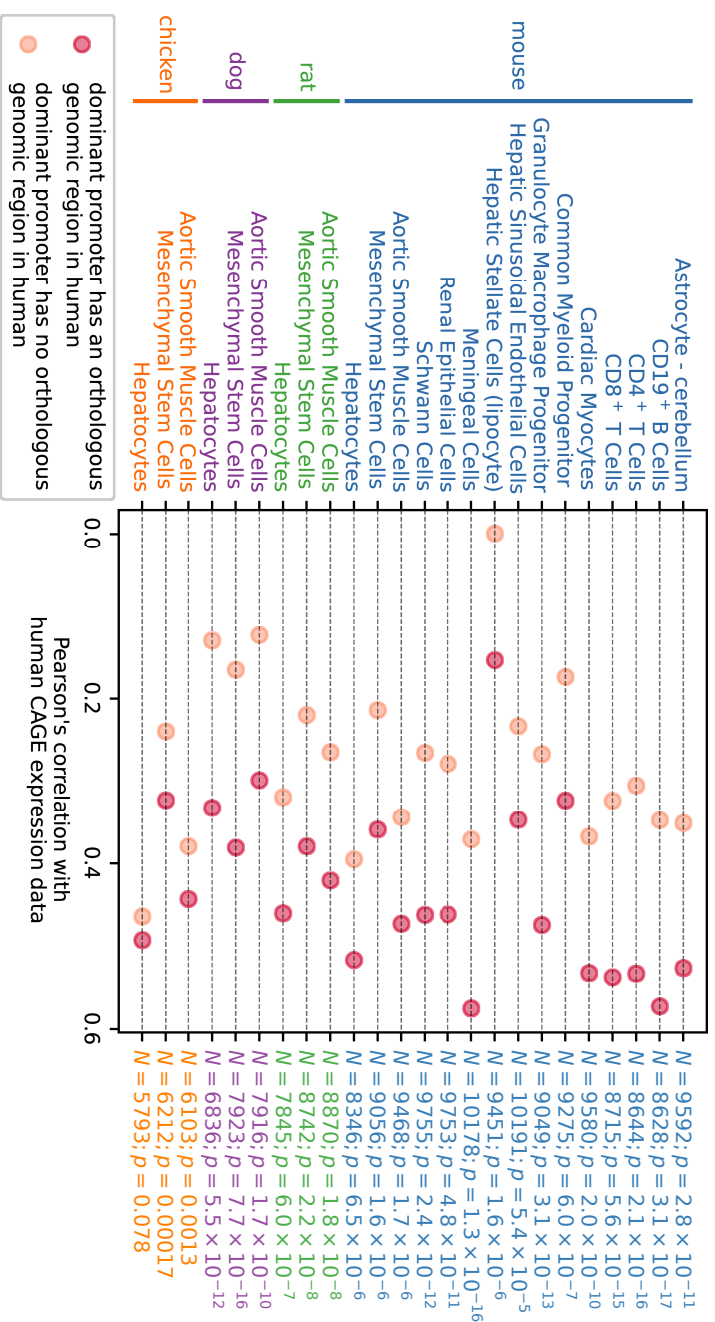




### Supplemental Figure S6.

Phylogenetic analysis of expression divergence. The value of Pearson's correlation  $r$  across genes is shown as a function of phylogenetic distance between species. Phylogenetic distances were calculated as the sum of branch lengths (which are proportional to the average number of substitutions per site in the genome) in the phylogenetic tree provided by UCSC (Miller et al. 2007). The decay time  $\tau$  was estimated by linear regression of the natural logarithm of  $r$  against the phylogenetic distance.

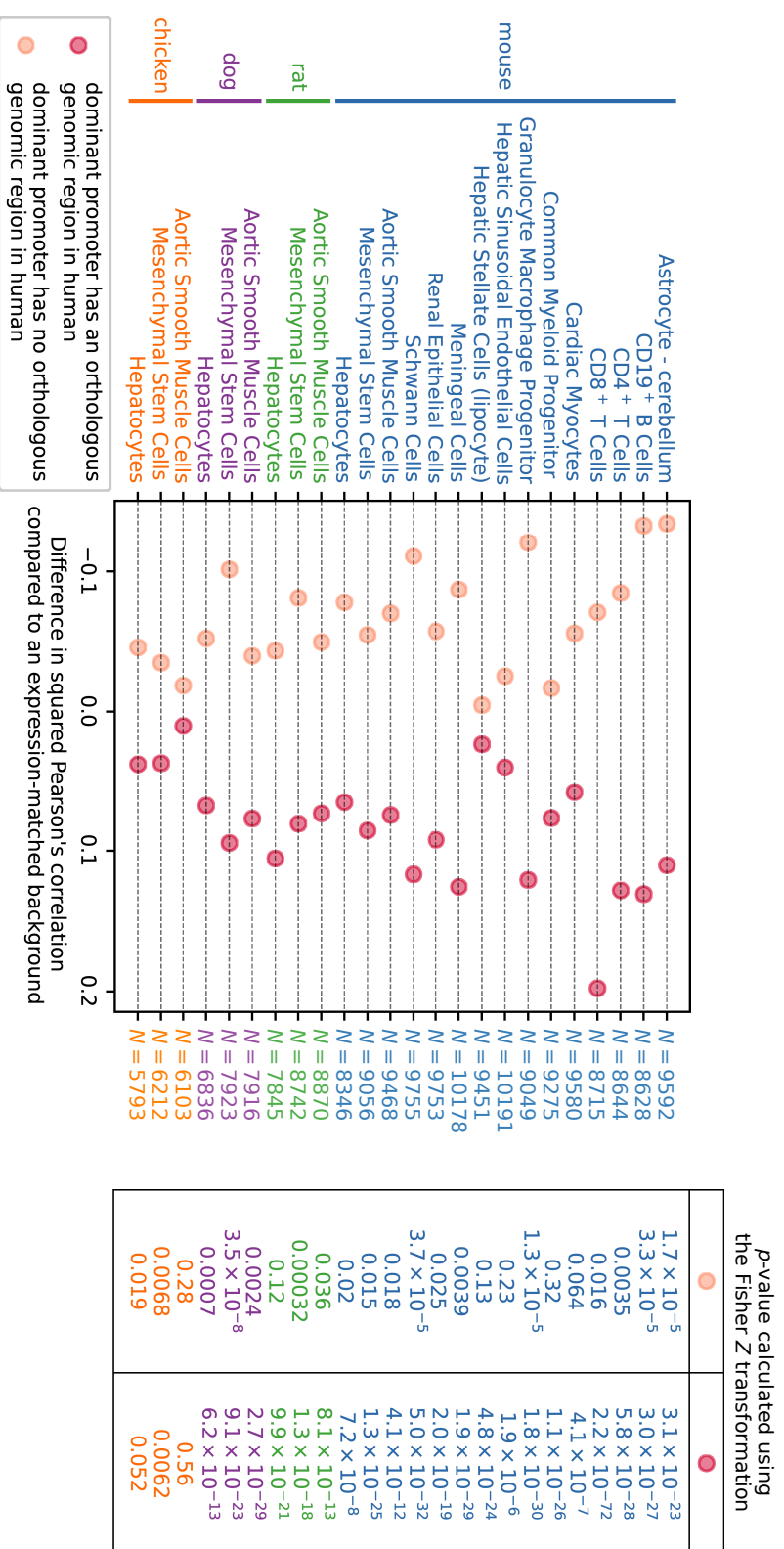




**Supplemental Figure S7.**

Promoter analysis of gene expression correlations. Pearson's correlation in each cell type and species was calculated separately for genes for which the dominant promoter had an orthologous genomic region in the human genome, and for genes for which the dominant promoter did not have an orthologous genomic region in the human genome. The one-sided  $p$ -value shown on the right was calculated by applying the Fisher Z-transformation to each correlation value and performing a paired Z-test.  $N$  is the number of expressed genes in each cell type.

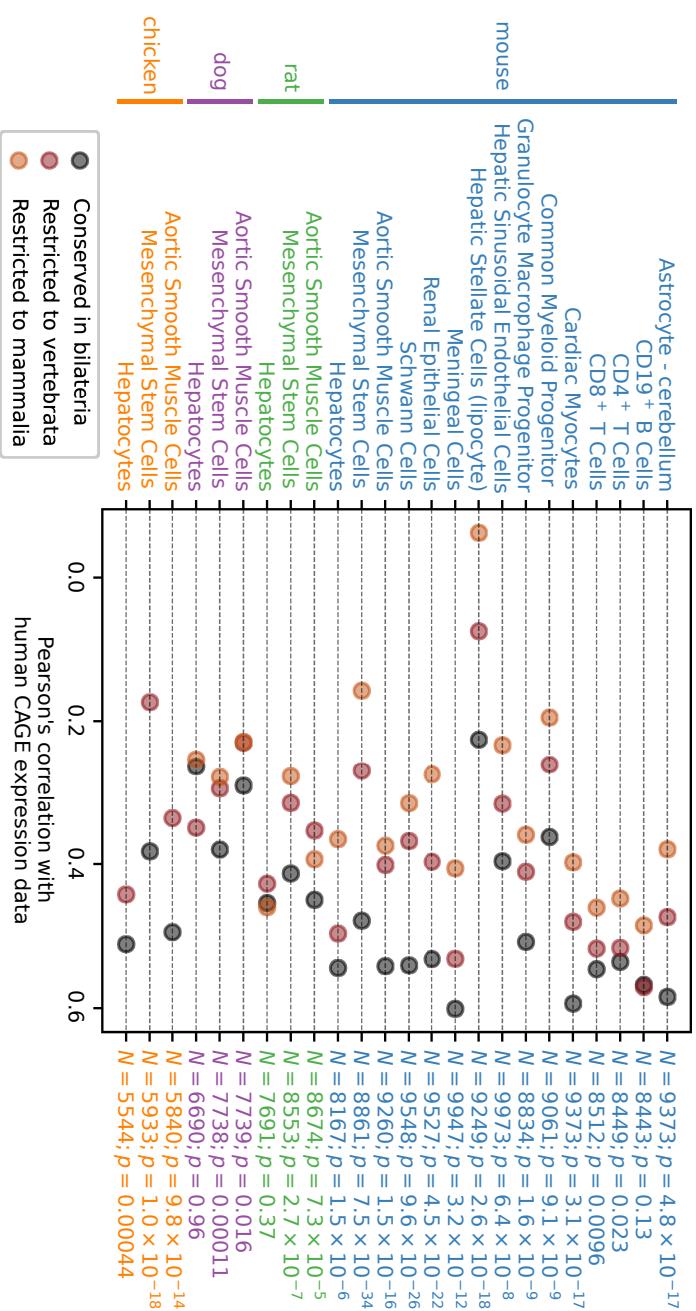




**Supplemental Figure S8.**

Promoter analysis of gene expression correlations compared to an expression-matched background. Difference in squared Pearson's correlation  $r^2$  in each cell type and species, relative to an expression-matched background, was calculated for genes for which the dominant promoter had an orthologous genomic region in the human genome, and for genes for which the dominant promoter did not have an orthologous genomic region in the human genome. The one-sided  $p$ -value shown on the right was calculated by applying the Fisher Z-transformation to each correlation value and performing a paired Z-test relative to the expression-matched background.  $N$  is the number of expressed genes in each cell type.

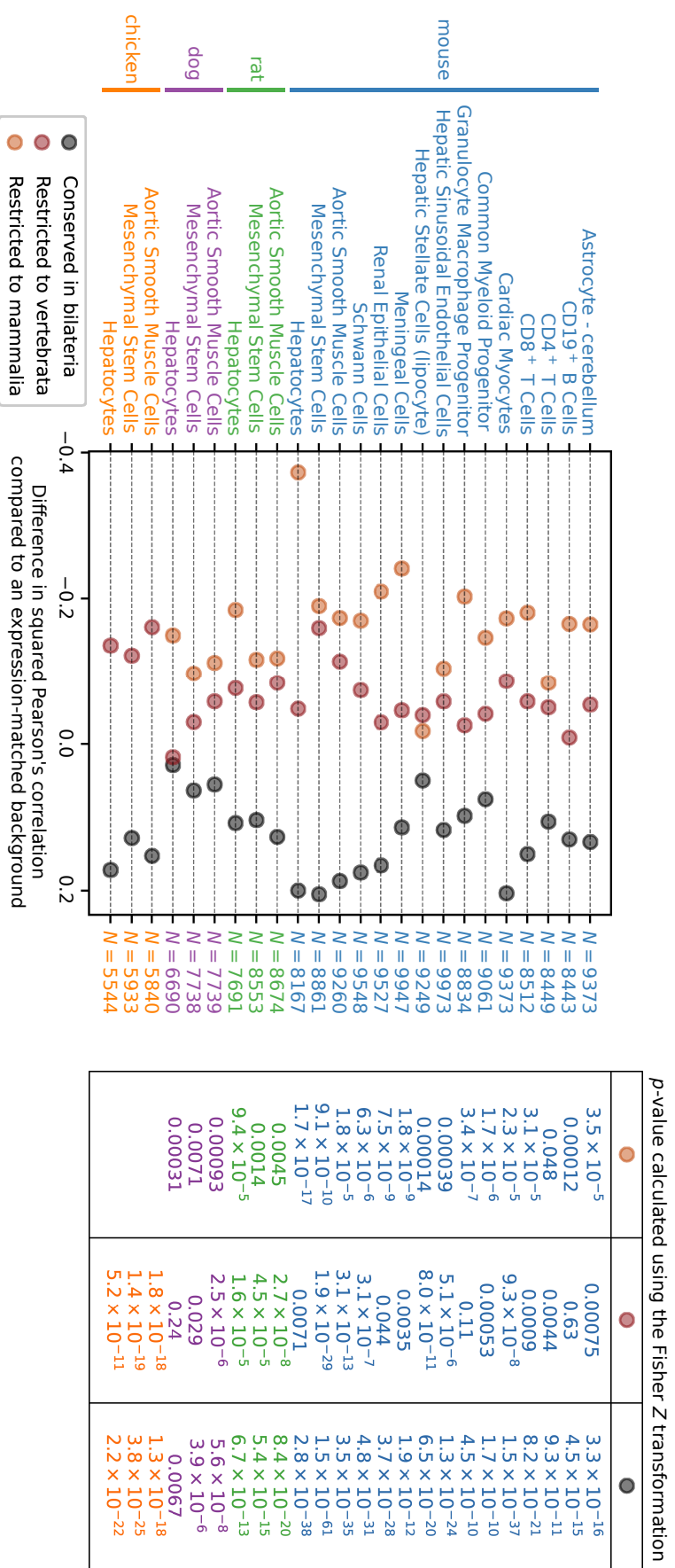




**Supplemental Figure S9.**

Conservation analysis of gene expression correlations. Pearson's correlation was calculated in each cell type and species across orthologous genes for three categories of genes defined by the age of their most recent common ancestor. The Fisher Z-transformation was applied to each correlation value to obtain Z-scores. The one-sided  $p$ -value of a linear regression model of the Z-scores against the evolutionary age category is shown on the right, together with the number  $N$  of expressed genes in each cell type with an annotation in the NCBI HomioGene database.

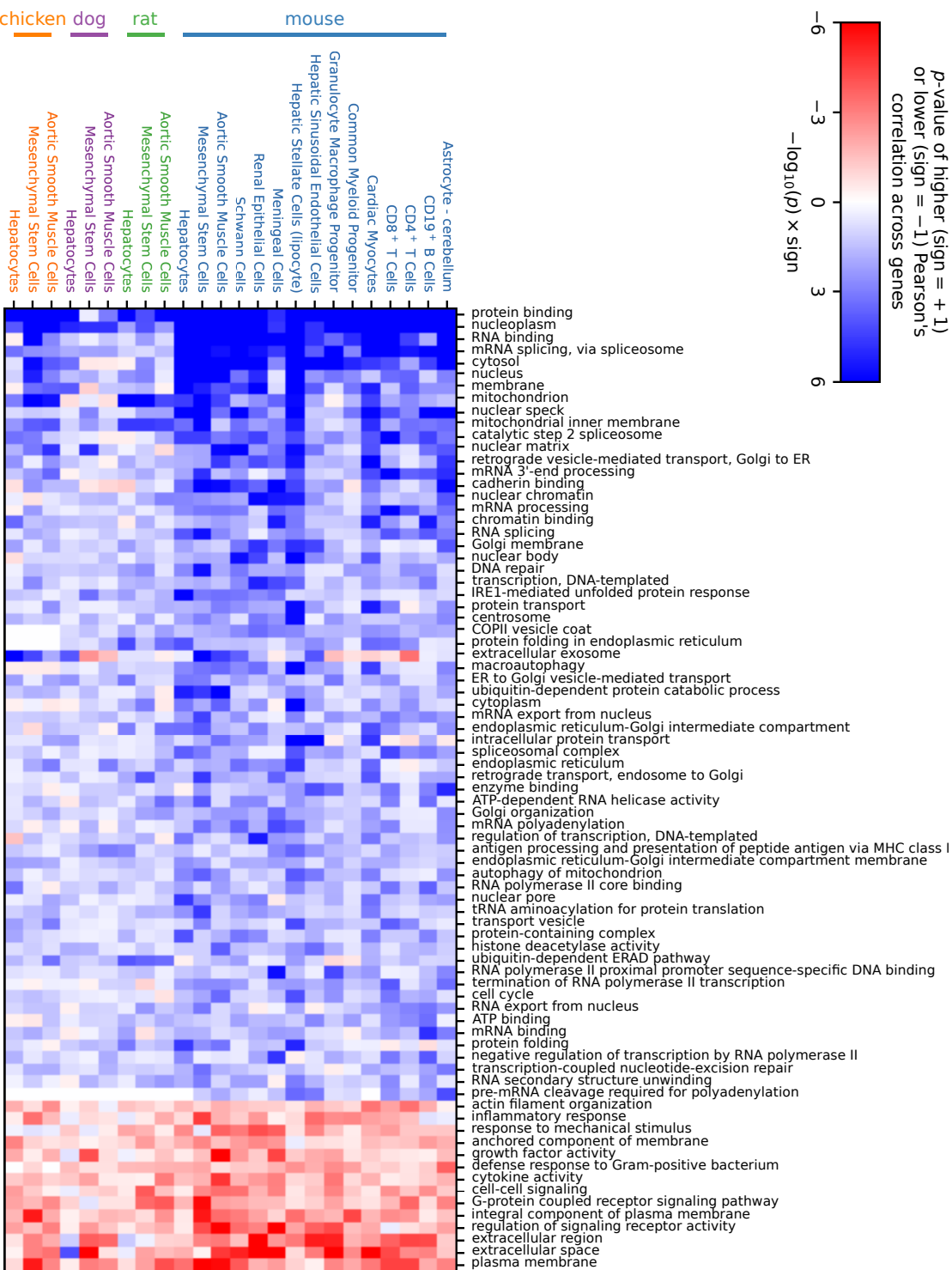




**Supplemental Figure S10.**

Conservation analysis of gene expression correlations compared to an expression-matched background. Difference in squared Pearson's correlation  $r^2$  in each cell type and species, relative to an expression-matched background, was calculated for three categories of genes defined by the age of their most recent common ancestor. The one-sided  $p$ -value shown on the right was calculated by applying the Fisher Z-transformation to each correlation value and performing a paired Z-test relative to the expression-matched background.  $N$  is the number of expressed genes in each cell type.

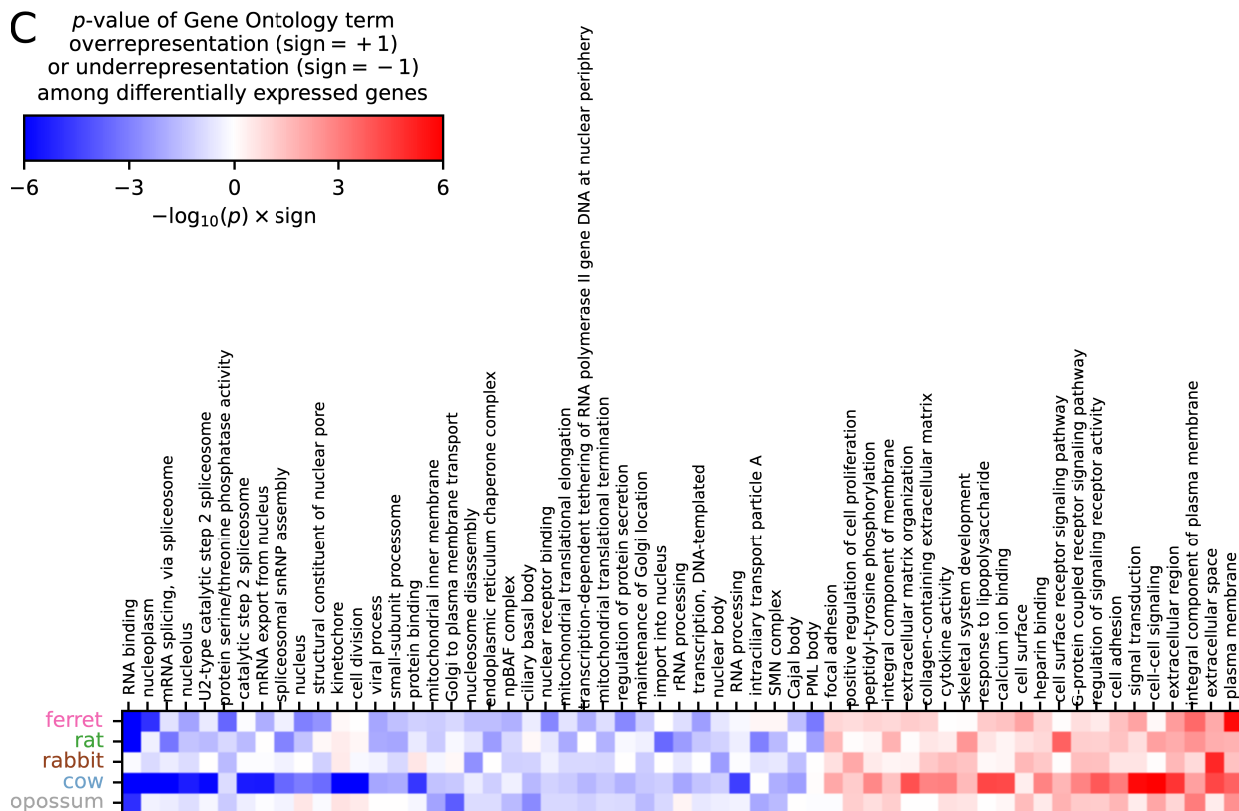
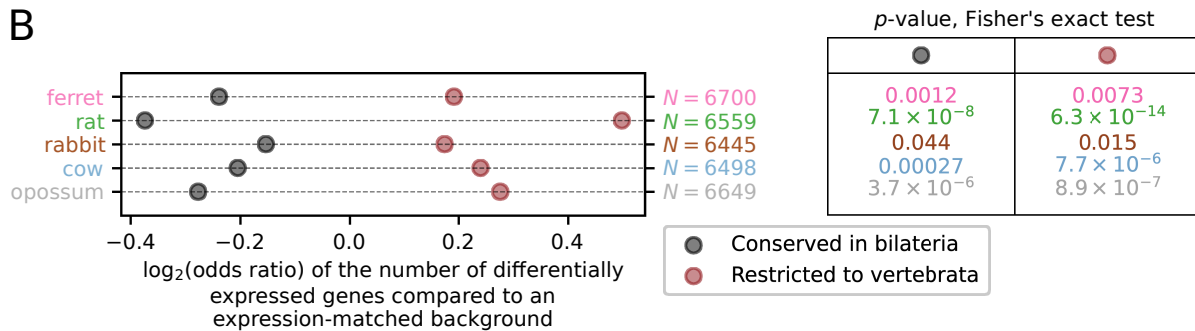
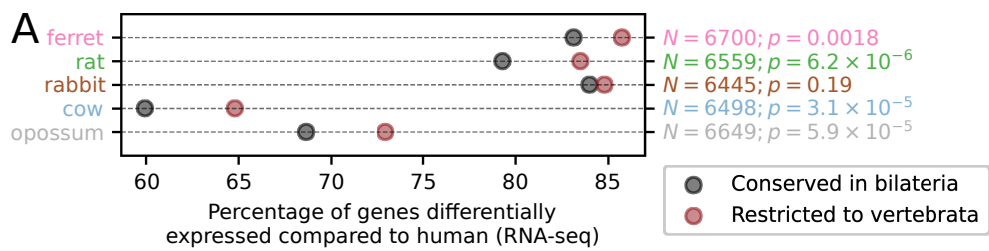




**Supplemental Figure S11.**

Gene Ontology analysis of gene expression correlations. The  $p$ -value of each Gene Ontology category in each cell type and species was calculated by applying the Fisher Z-transformation to the correlation value and performing a paired Z-test relative to the expression-matched background, with a higher or lower correlation value shown in blue and red, respectively.



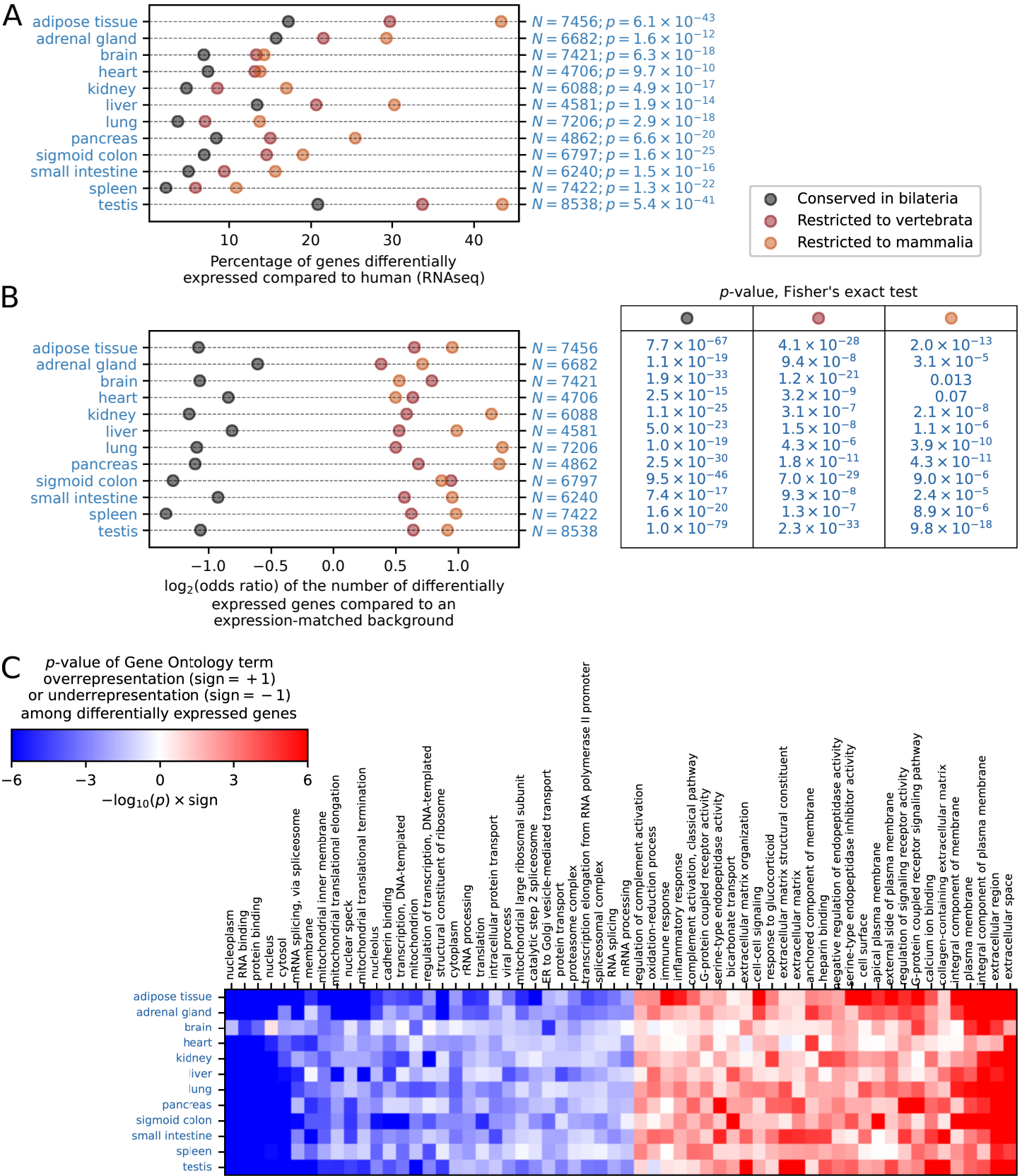




### Supplemental Figure S12 (previous page).

Analysis of RNA-Seq expression data for endometrial stromal fibroblast primary cells (Kin *et al.*, 2016). (A) Percentage of differentially expressed genes in each species, separated by the age of the most recent common ancestor for each gene. The one-sided  $p$ -value calculated using Fisher's exact test is shown on the right, together with the number  $N$  of expressed genes in each species with an annotation in the NCBI HomoloGene database. (B) Enrichment of differentially expressed genes in each species, separated by the age of the most recent common ancestor. The enrichment was calculated by comparing the number of differentially expressed genes to the number of differentially expressed genes in an expression-matched background, with the  $\log_2(\text{odds ratio})$  shown on the horizontal axis of the left panel and  $p$ -value calculated using Fisher's exact test for each comparison in the right panel.  $N$  is the total number of expressed genes in each species with an annotation in the NCBI HomoloGene database. (C) Gene Ontology analysis of differentially expressed genes. The  $p$ -value, calculated using Fisher's exact test, of overrepresentation or underrepresentation of differentially expressed genes in each Gene Ontology category compared to an expression-matched set of background genes is shown in red and blue, respectively.



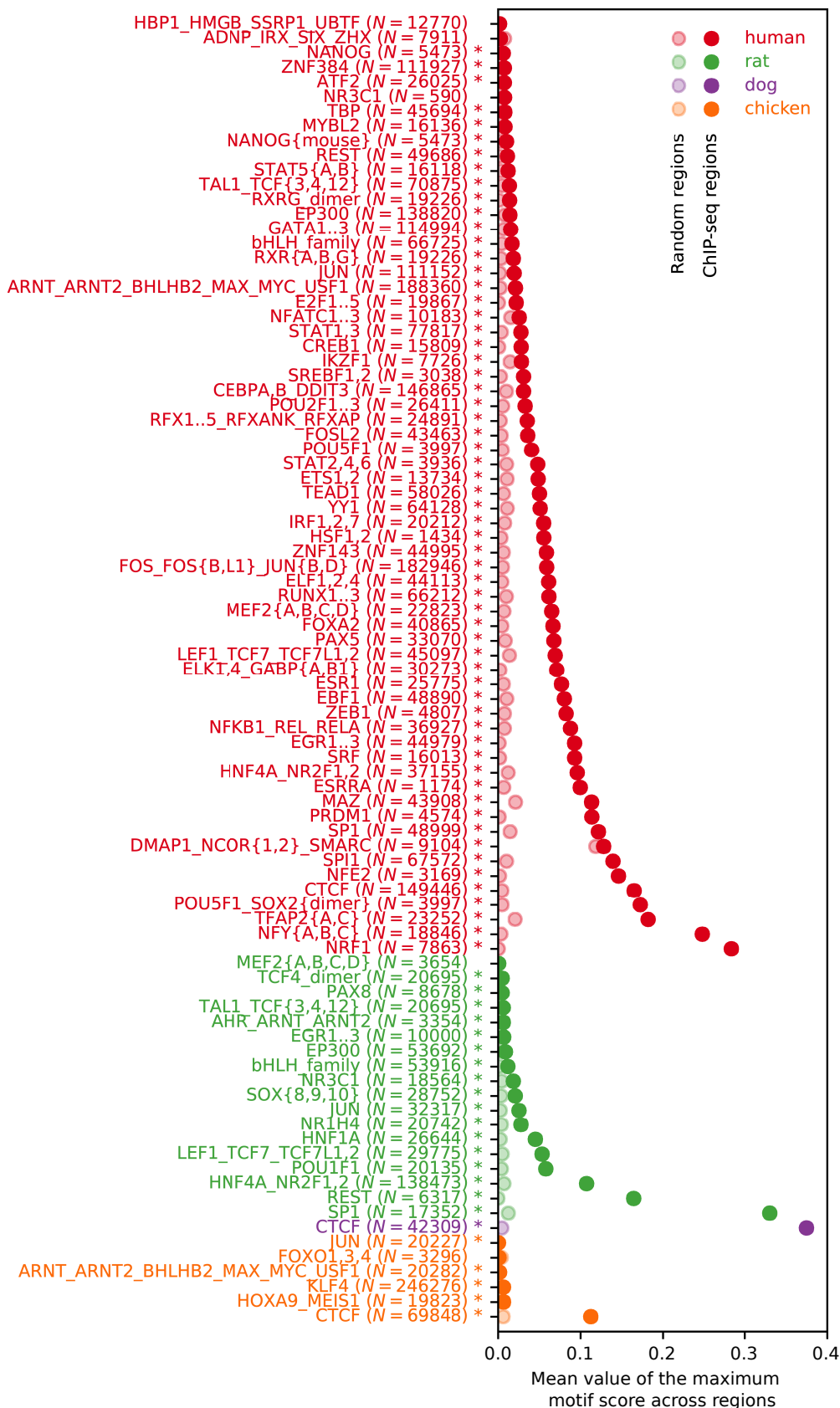




### Supplemental Figure S13 (previous page).

Comparative analysis of RNA-Seq expression data for 12 matching tissues in human and mouse (The ENCODE Project Consortium, 2012). (A) Percentage of differentially expressed genes in each tissue, separated by the age of the most recent common ancestor for each gene. The one-sided  $p$ -value calculated using Fisher's exact test is shown on the right, together with the number  $N$  of expressed genes in each tissue with an annotation in the NCBI HomoloGene database. (B) Enrichment of differentially expressed genes in each tissue, separated by the age of the most recent common ancestor. The enrichment was calculated by comparing the number of differentially expressed genes to the number of differentially expressed genes in an expression-matched background, with the  $\log_2(\text{odds ratio})$  shown on the horizontal axis of the left panel and  $p$ -value calculated using Fisher's exact test for each comparison in the right panel.  $N$  is the total number of expressed genes in each tissue with an annotation in the NCBI HomoloGene database. (C) Gene Ontology analysis of differentially expressed genes. The  $p$ -value, calculated using Fisher's exact test, of overrepresentation or underrepresentation of differentially expressed genes in each Gene Ontology category compared to an expression-matched set of background genes is shown in red and blue, respectively.

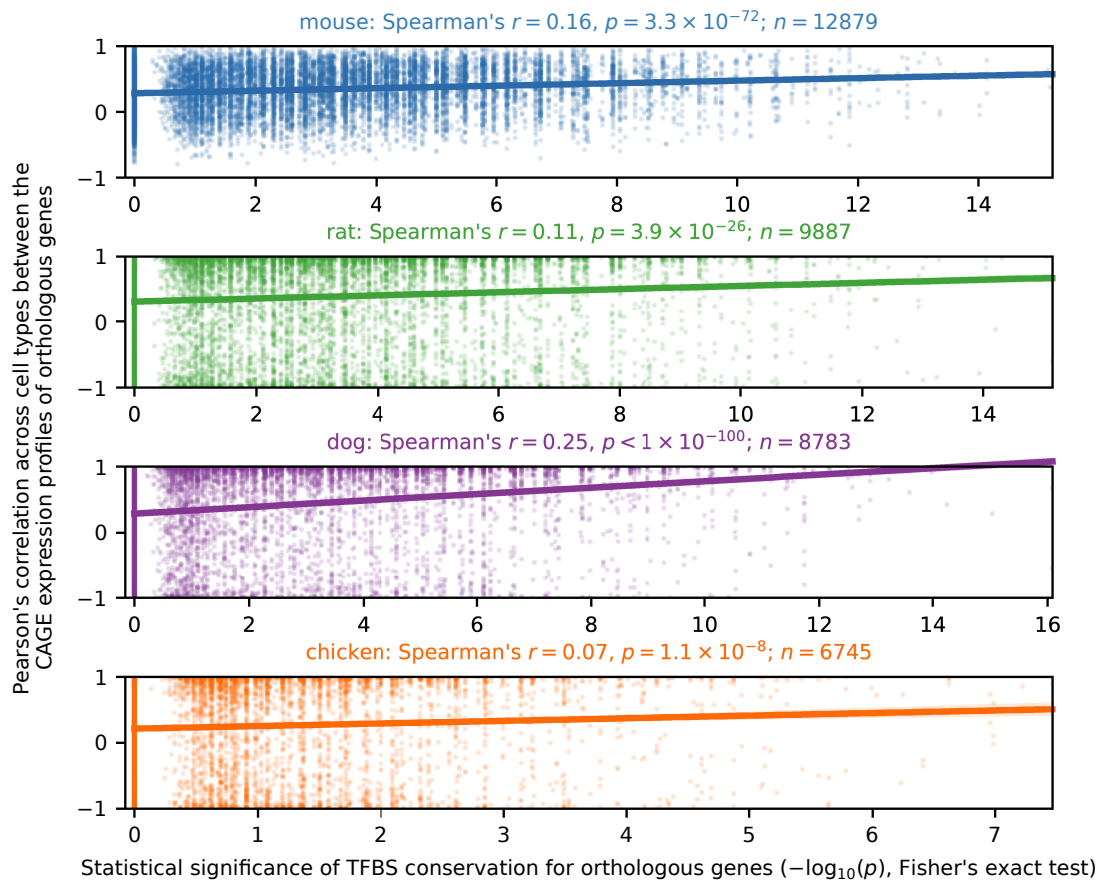




**Supplemental Figure S14.**

Evaluation of predicted transcription factor binding sites. The maximum TFBS score was calculated for each genomic region found in ChIP-seq experiments to be bound by transcription factors associated with the motif, and compared to an equal number of randomly selected genomic regions. The number  $N$  of ChIP-seq regions for each motif is shown. Motifs for which the maximum motif scores are significantly ( $p < 0.05$ , Mann-Whitney  $U$  test) higher for ChIP-seq regions compared to background regions are indicated by an asterisk.



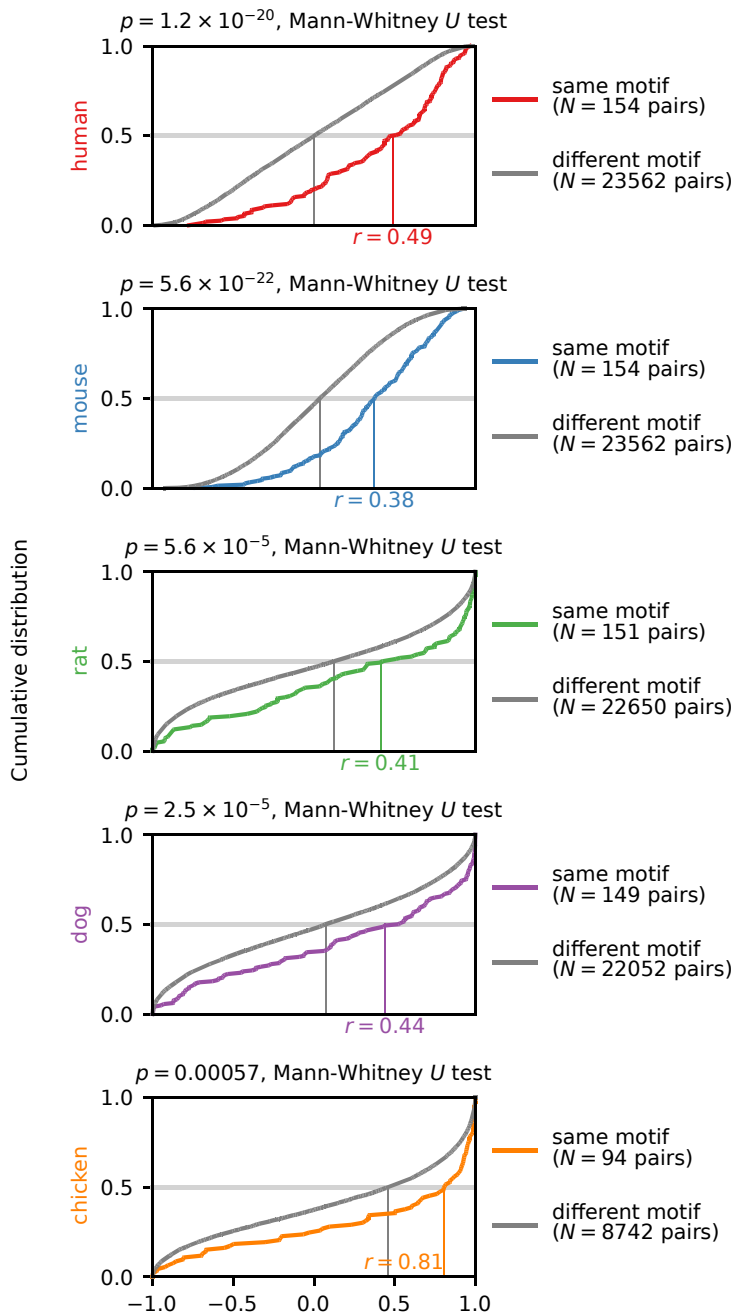


### Supplemental Figure S15.

Pearson's correlation across cell types between orthologous genes, as a function of TFBS conservation in their promoter regions. TFBS conservation was evaluated as the  $p$ -value, calculated using Fisher's exact test, of co-occurrence of binding sites, predicted by MotEvo, for the same transcription factors in the promoters of each orthologous gene pair.



# Motif activity comparison, promoters vs enhancers

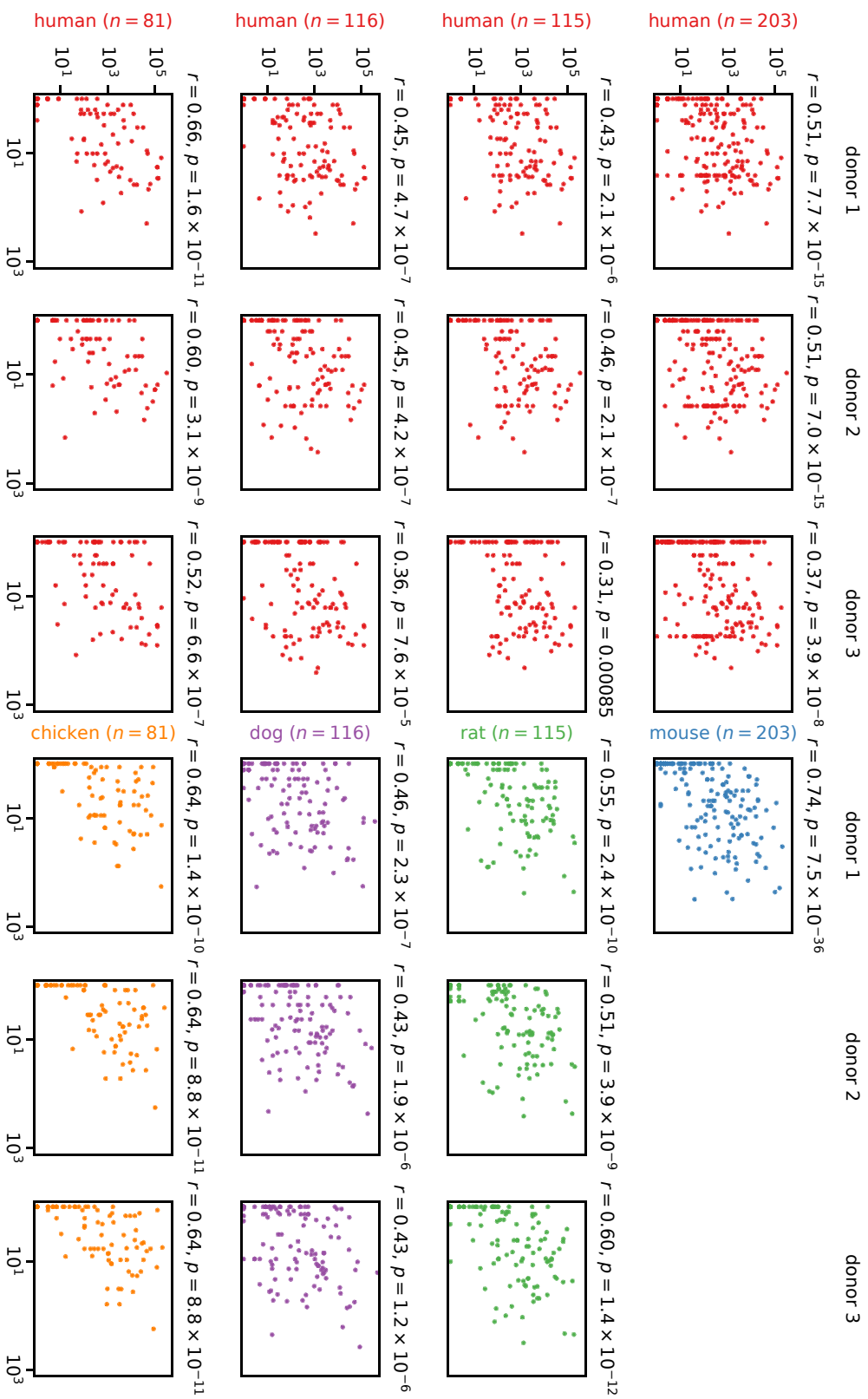


Pearson's correlation between motif activities for promoters and enhancers in matching cell types in each organism

## Supplemental Figure S16.

Cumulative distribution of the Pearson's correlation  $r$  across cell types in motif activity between promoters and enhancers in human, mouse, rat, dog, and chicken. The estimated median value of  $r$  is indicated on the horizontal axis of each graph. As a background distribution, we calculated the same correlation between pairs of different motifs in each of the five species. The Mann-Whitney  $U$  test  $p$ -value comparing the actual correlation values to the correlation values of the background distribution is shown for each species.

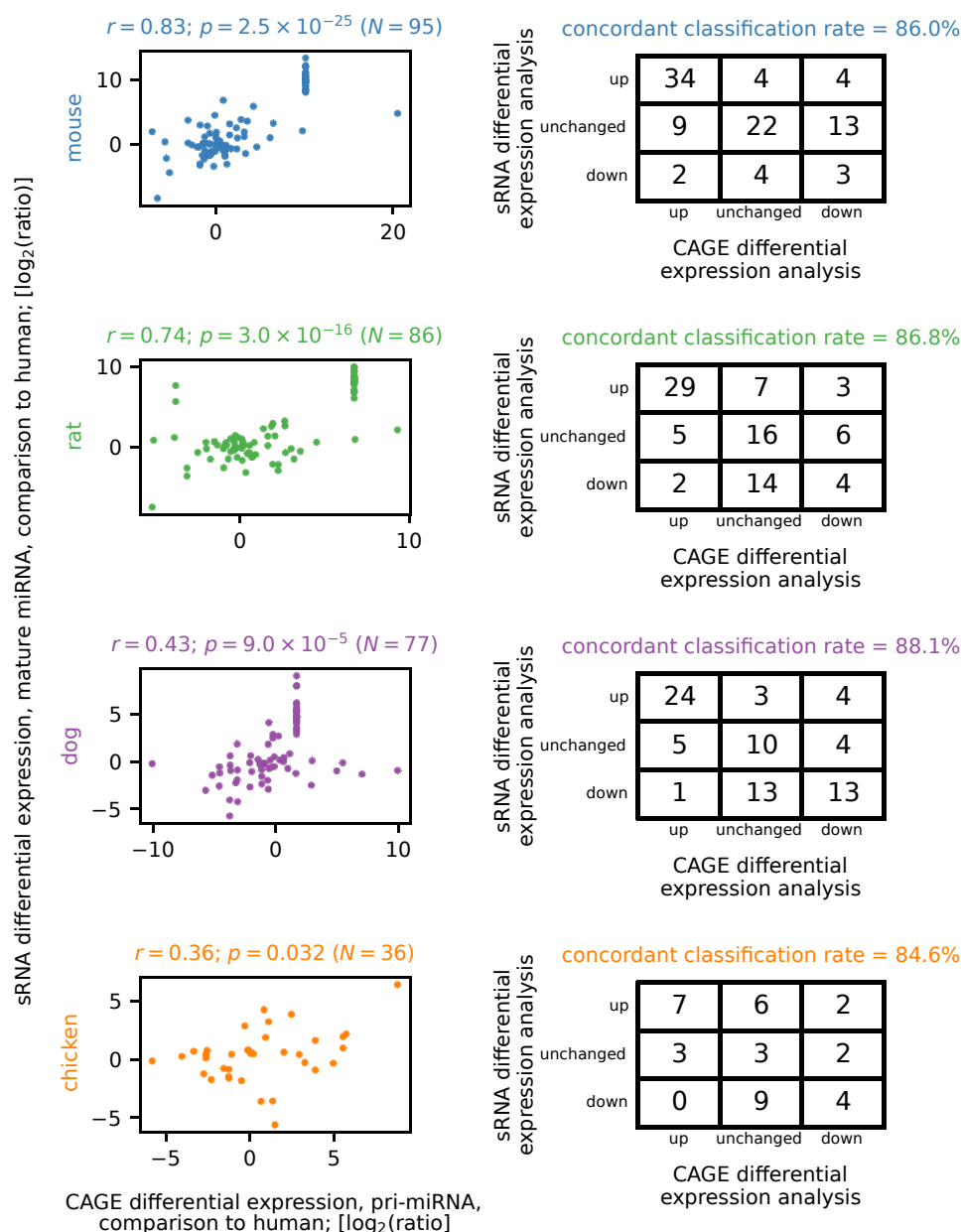




**Supplemental Figure S17.**

Comparison of CAGE expression levels of the pri-miRNA to the mature miRNA expression levels measured by sRNA sequencing in aortic smooth muscle cells in human, mouse, rat, dog, and chicken. To allow a direct comparison of the correlation values observed in human and each of the species, each row compares the CAGE and sRNA expression levels for the same set of miRNAs in human and mouse, rat, dog, or chicken.





### Supplemental Figure S18.

Comparison of differential expression analysis of pri-miRNAs and mature miRNAs. (*left*) Comparison of  $\log_2$  expression ratios obtained in differential expression analysis performed on the expression levels of pri-miRNAs (using CAGE sequencing data) and mature miRNA (using sRNA sequencing data). (*right*) Number of differentially expressed miRNAs found in differential expression analysis of pri-miRNA (CAGE) and mature miRNA (sRNA) expression data. The concordant classification rate is calculated as the number of miRNAs that are significantly up- or down-regulated compared to human both in CAGE differential expression analysis and in sRNA differential expression analysis, divided by the total number of miRNAs differentially expressed in both CAGE and sRNA differential expression analysis.