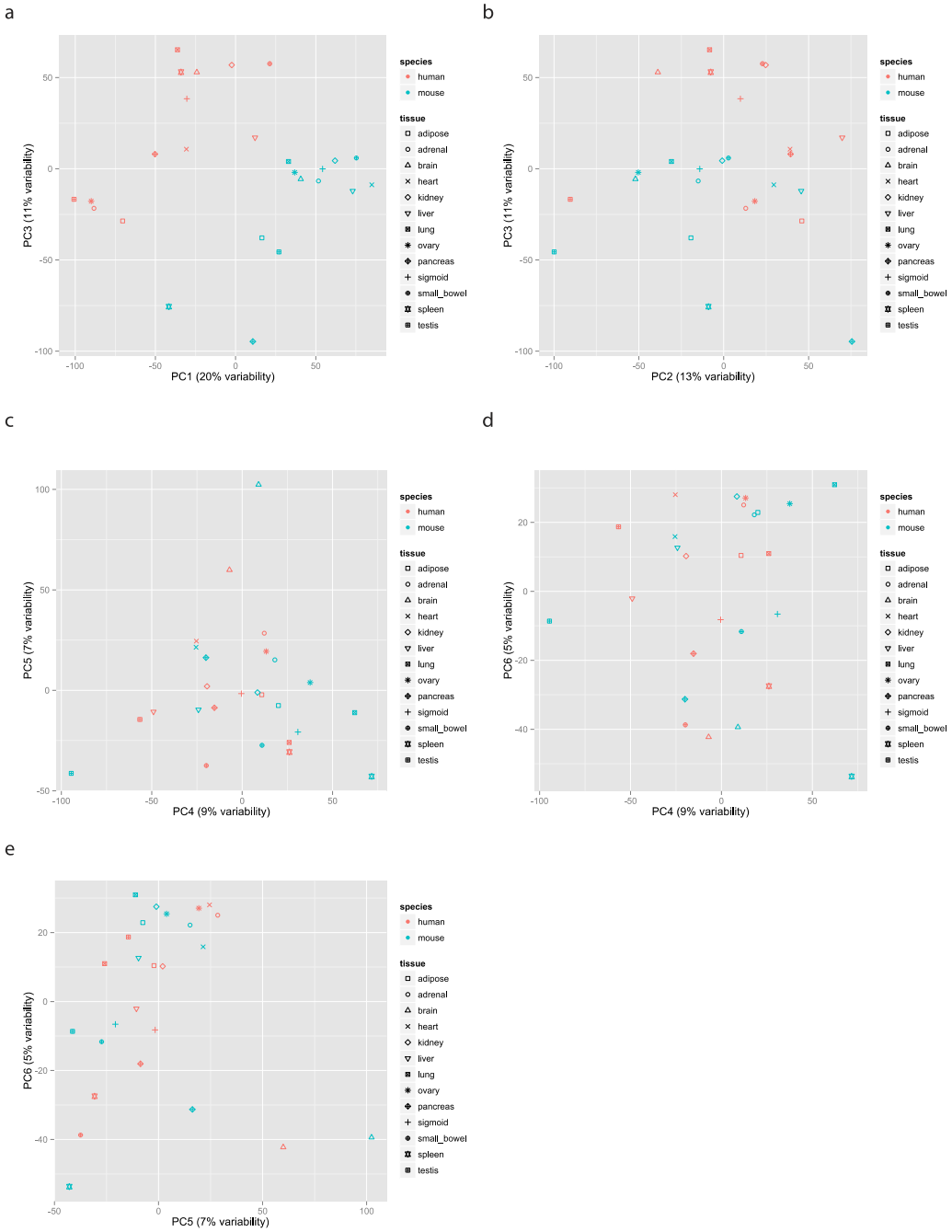

Figure S1. Two-dimensional plots of principal components resulting from PCA of the log-transformed FPKM values for the 26 samples. We performed PCA with the 'prcomp' R function (with the 'scale' and 'center' options set to TRUE) on the transposed matrix of log-transformed FPKM values (i.e. samples in rows and genes in columns) after removal of invariant columns (genes). Shown are two-dimensional plots of a. PC1 vs. PC3 b. PC2 vs. PC3 c. PC4 vs. PC5 d. PC4 vs. PC6 e. PC5 vs. PC6 (a plot of PC1 and PC2 is part of Figure 2 in the main text).

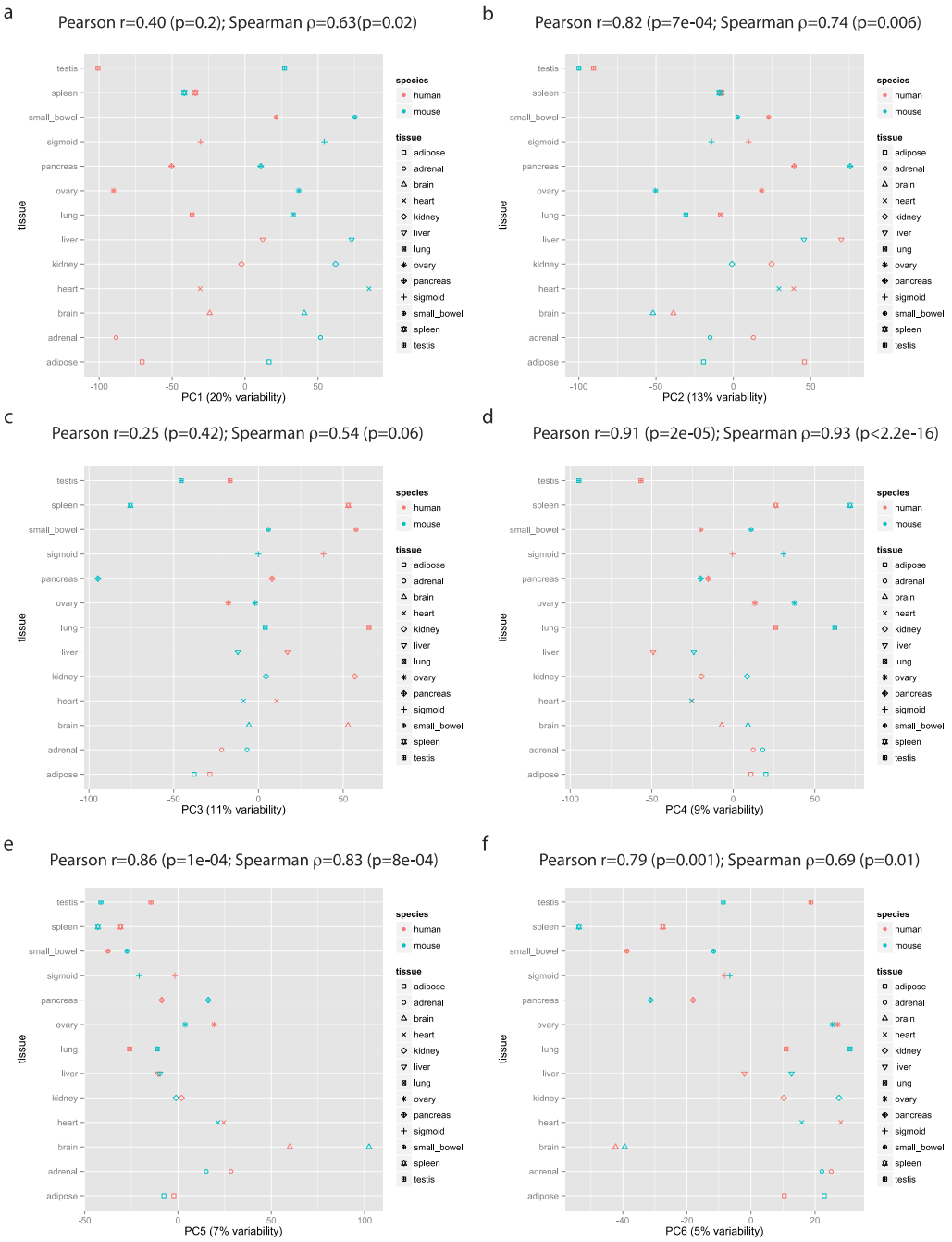

Figure S2. Plots of individual principal components resulting from PCA of the log-transformed FPKM values for the 26 samples. We performed PCA with the 'prcomp' R function (with the 'scale' and 'center' options set to TRUE) on the transposed matrix of log-transformed FPKM values (i.e. samples in rows and genes in columns) after removal of invariant columns (genes). For each principal component we calculated the Pearson and Spearman correlations between the human and mouse values ordered by tissue, along with their respective uncorrected p-values. a. PC1 b. PC2 c. PC3 d. PC4 e. PC5 f. PC6

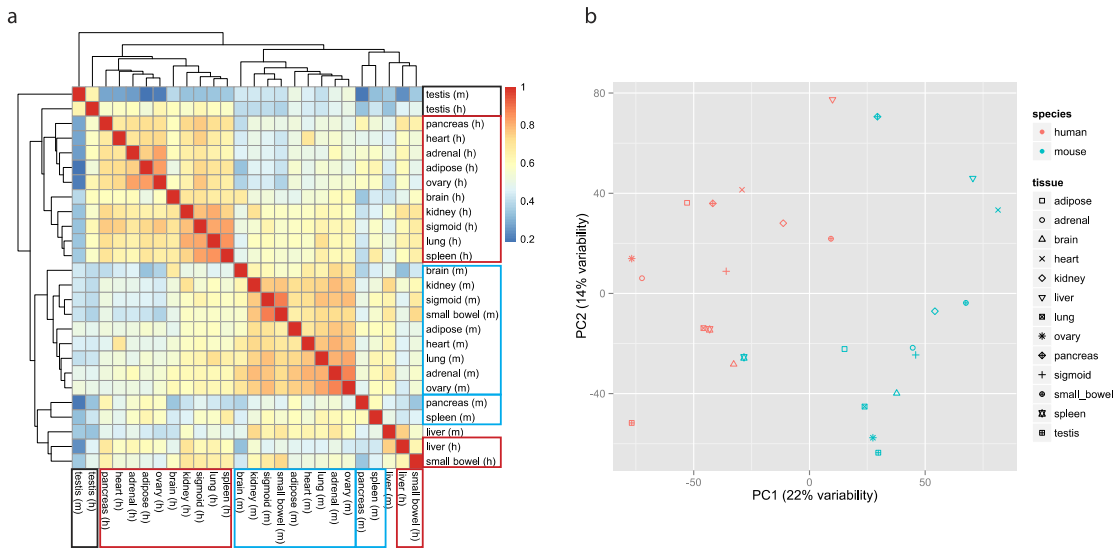

Figure S3. Filtering of lowly expressed genes is not sufficient to remove clustering of the data by species. a. Heatmap based on pairwise Pearson correlation of expression data (log-transformed FPKM values) from the same 10,309 orthologous gene pairs used in Figure 3 (resulting from exclusion of lowly-expressed genes and mitochondrial genes). Boxes indicate human clusters (red), mouse clusters (cyan) or same-tissue human-mouse pairs that are clustered together (black) b. Two-dimensional plot of the first two principal components calculated by performing PCA (using 'prcomp' with options 'scale' and 'center' set to TRUE) on the transposed matrix of log-transformed FPKM values for the 26 samples, after removal of invariant columns (genes).

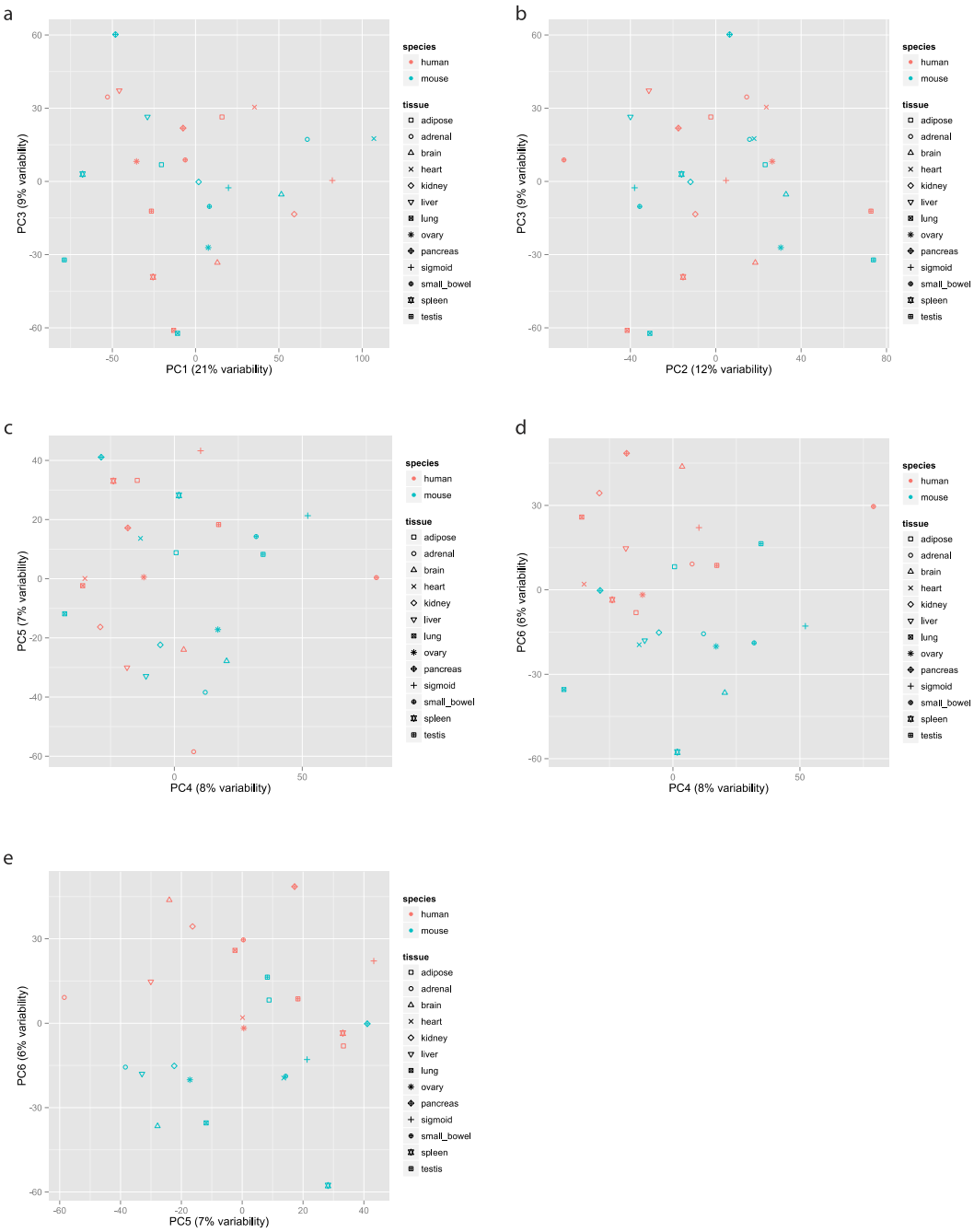

Figure S4. Two-dimensional plots of principal components resulting from PCA of the batch-cleaned log-transformed normalized fragment counts for the 26 samples. Raw fragment counts for the 26 samples were filtered, normalized, log-transformed and cleaned from batch effects (see the 'Results' for details and Supplementary Text 2 for R code). We then performed PCA with the 'prcomp' R function (with the 'scale' and 'center' options set to TRUE) on the transposed matrix of 'clean' values (i.e. samples in rows and genes in columns) after removal of invariant columns (genes). Shown are two-dimensional plots of a. PC1 vs. PC2 b. PC2 vs. PC3 c. PC4 vs. PC5 d. PC4 vs. PC6 e. PC5 vs. PC6 (a plot of PC1 and PC2 is part of Figure 3 in the main text).

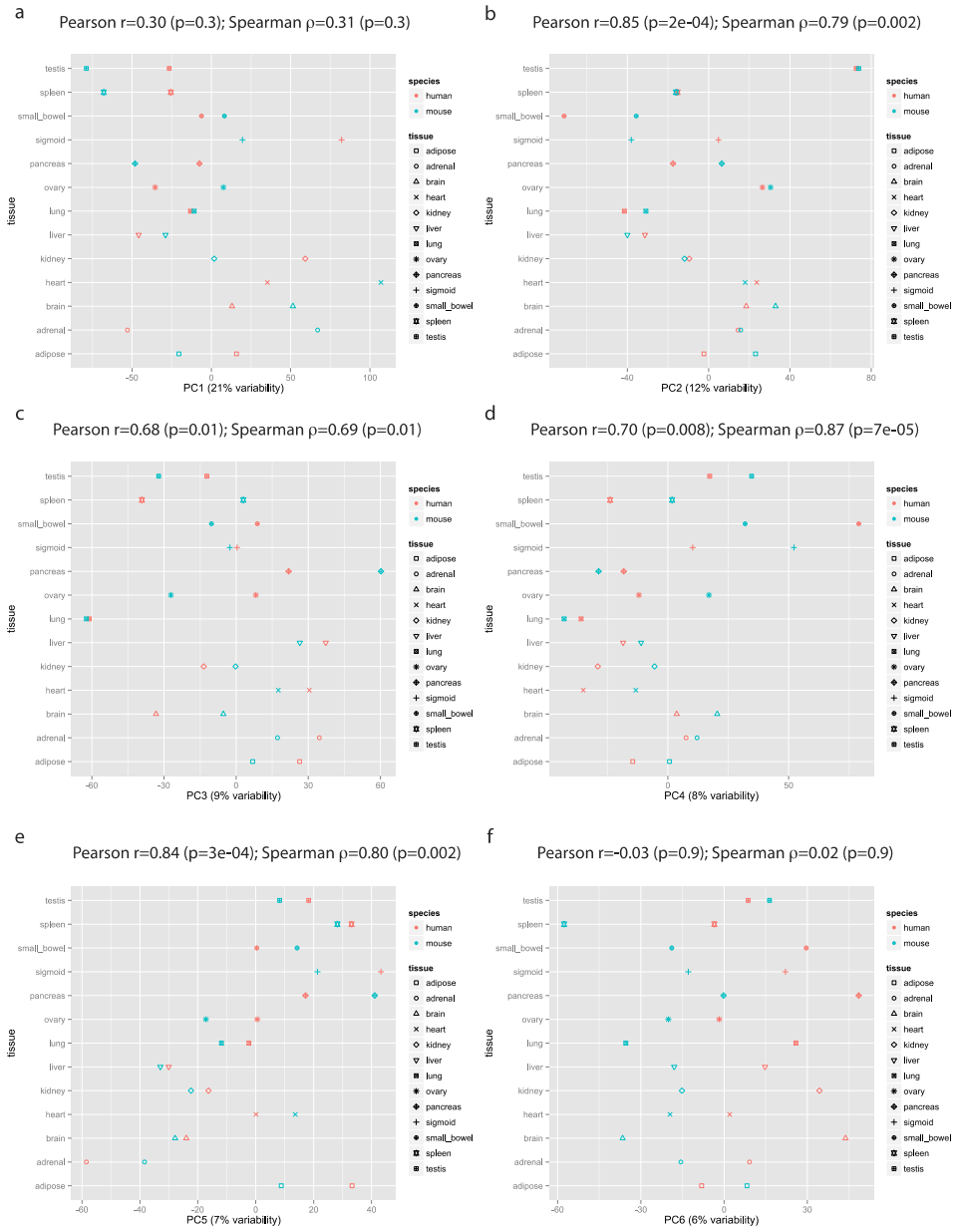

Figure S5. Plots of individual principal components resulting from PCA of the batch-cleaned log-transformed normalized fragment counts for the 26 samples. Raw fragment counts for the 26 samples were filtered, normalized, log-transformed and cleaned from batch effects (see the 'Results' for details and Supplementary Text 2 for R code). We then performed PCA with the 'prcomp' R function (with the 'scale' and 'center' options set to TRUE) on the transposed matrix of 'clean' values (i.e. samples in rows and genes in columns) after removal of invariant columns (genes). For each principal component we calculated the Pearson and Spearman correlations between the human and mouse values ordered by tissue, as well as the uncorrected p-value. a. PC1 b. PC2 c. PC3 d. PC4 e. PC5 f. PC6. Note that while no species separation is seen for the first five principal components, the sixth principal component shows a clear separation between the two species.

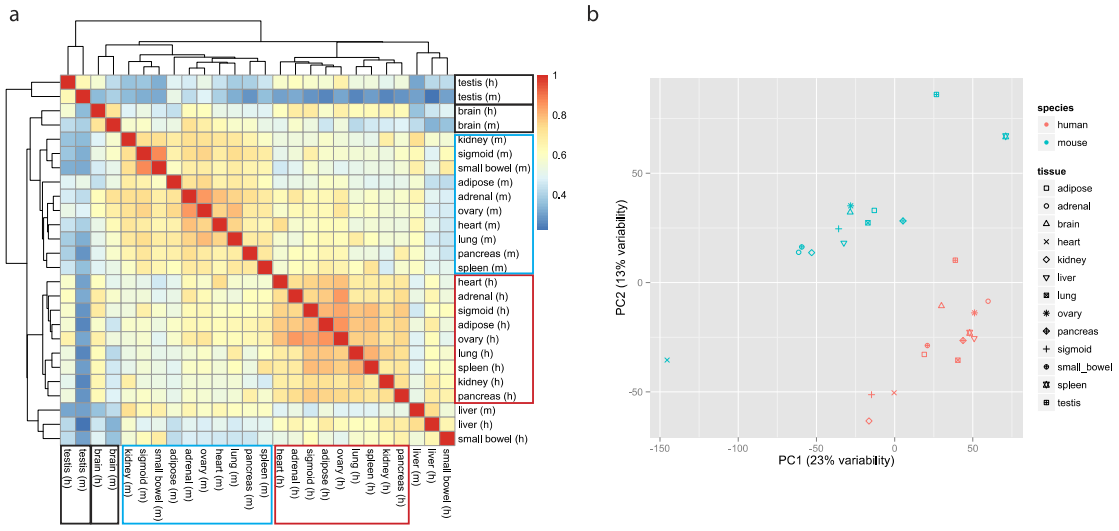

Figure S6. Filtration and normalization of raw fragment counts is not sufficient to remove clustering of the data by species. The same analysis and visualization pipeline that produced Figure 3 was applied to the raw fragment counts, with the only modification being the omission of the last step, i.e. the correction for batch effects (see 'Results' section for details of the processing of the data and Supplementary Text 2 for the R code). a. Heatmap based on pairwise Pearson correlation of expression data. Boxes indicate human clusters (red), mouse clusters (cyan) or same-tissue human-mouse pairs that are clustered together (black) b. Two-dimensional plot of the first two principal components calculated by performing PCA (using 'prcomp' with options 'scale' and 'center' set to TRUE) on the transposed matrix of log-transformed values for the 26 samples, after removal of invariant columns (genes).

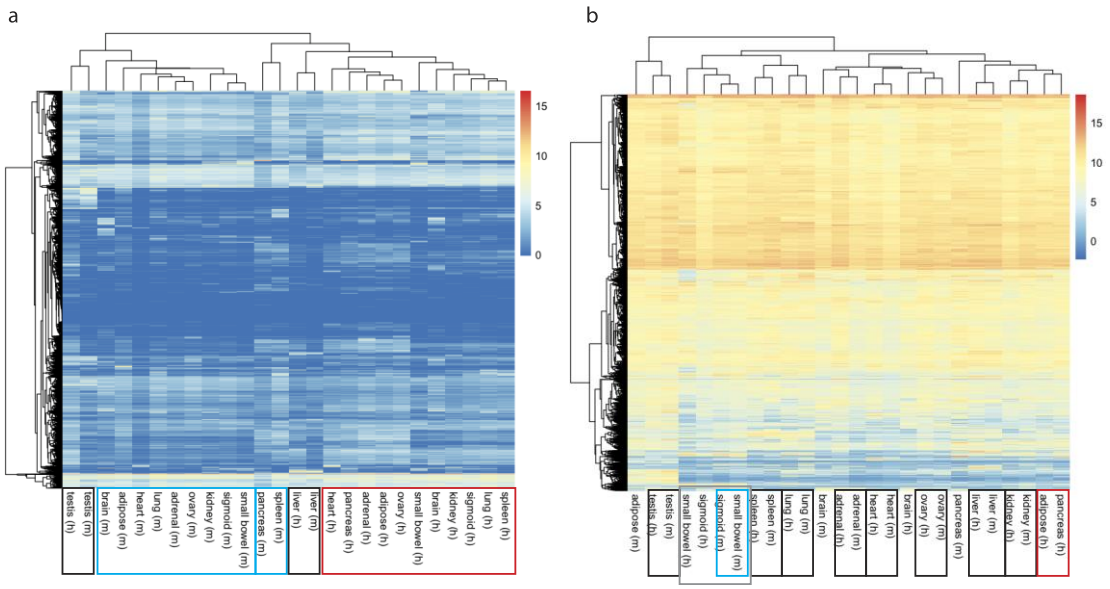

Figure S7. Clustering of vectors of expression values shows similar patterns to those obtained by clustering the correlation matrix derived from these vectors. a. Heatmap based on the complete-linkage clustering of the log-transformed FPKM values, using correlation distance. b. Heatmap based on the complete-linkage clustering of the batch-corrected log-transformed normalized fragment counts remaining after the exclusion steps described in the main text (10,309 values for each of the 26 samples), using correlation distance. Boxes indicate human clusters (red), mouse clusters (cyan) or same-tissue human-mouse pairs that are clustered together (black)

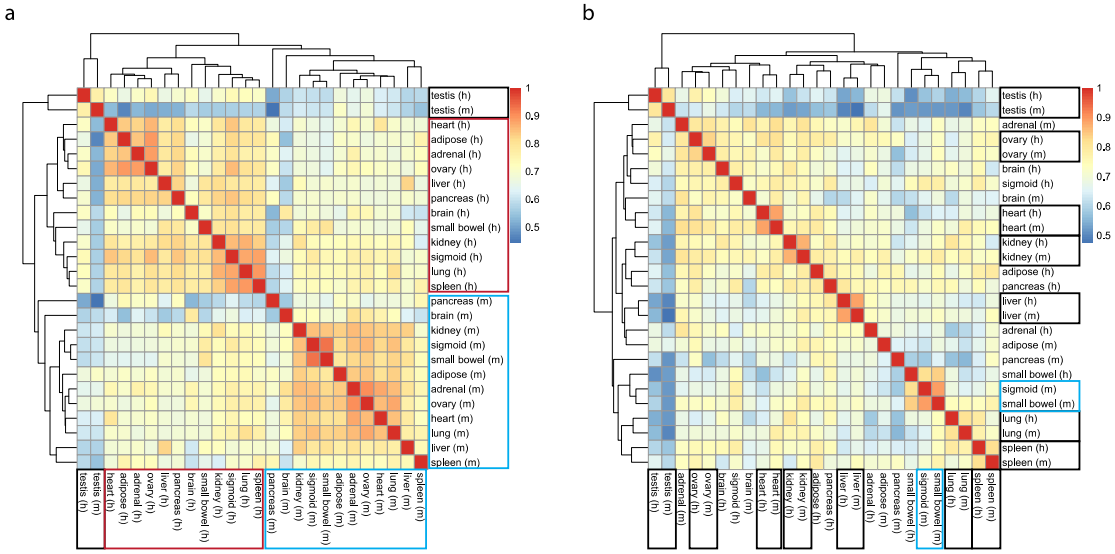

Figure S8. Heatmaps clustering Spearman correlations of expression values. a. Expression values used are log-transformed FPKM values of 14,744 orthologous genes from 26 samples b. Expression values are batch-corrected, log-transformed normalized fragment counts of 10,309 orthologous genes remaining after filtration (as detailed in the Results section) from 26 samples. In both panels we used Euclidean distance and complete linkage as distance measure and clustering method, respectively. Red boxes: all-human cluster; cyan boxes: all-mouse cluster; black boxes: same-tissue human-mouse pairs that were clustered together.

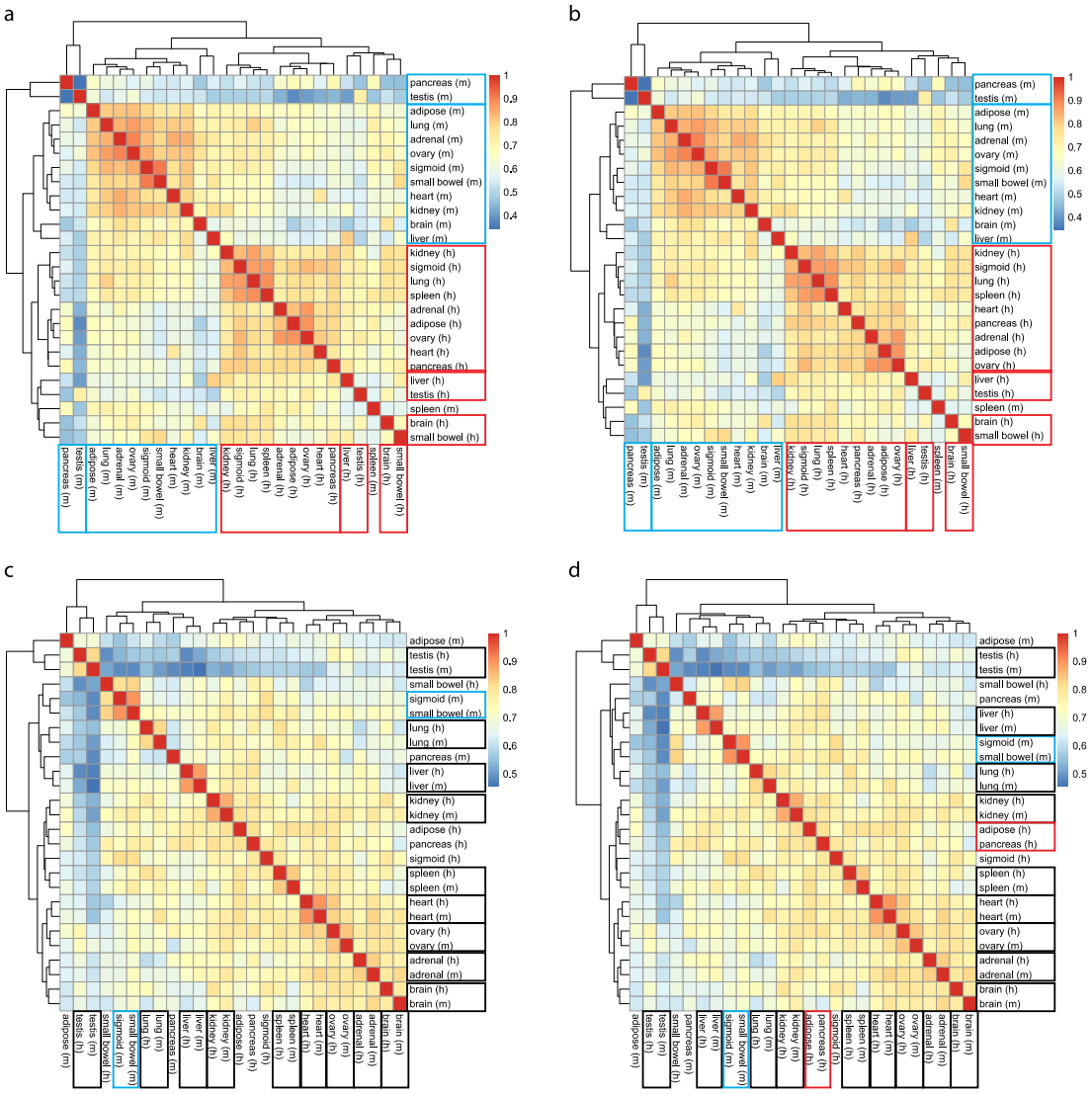

Figure S9. Clustering results are robust to choice of distance measure. In the main text figures we show heatmaps that cluster the Pearson correlation matrices using Euclidean distance and complete linkage as distance measure and clustering method, respectively. Here, we repeat the clustering using alternative distance measures: Manhattan (a and c) and Canberra (b and d) distances. a. and b. Clustering of the Pearson correlations of the log-transformed FPKM values from 14,744 orthologous gene pairs. c. and d. Clustering of Pearson correlations of the batch-corrected, log-transformed normalized fragment counts from 10,309 orthologous gene pairs that remained after the exclusion steps described in the Results section. Red boxes: all-human cluster; cyan boxes: all-mouse cluster; black boxes: same-tissue human-mouse pairs that were clustered together.

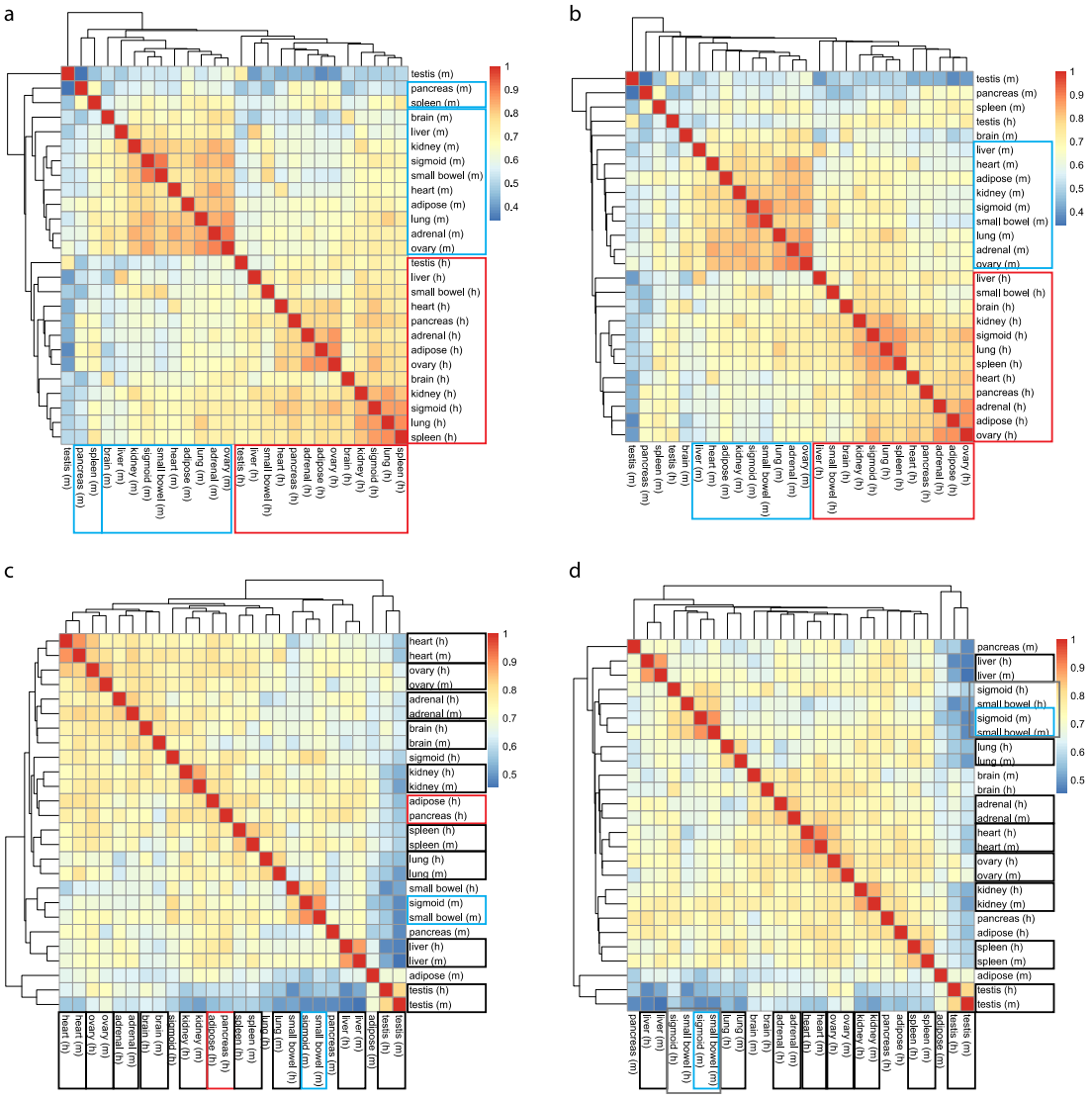

Figure S10. Clustering results are robust to choice of clustering method. In the main text figures we show heatmaps that cluster the Pearson correlation matrices using Euclidean distance and complete linkage as distance measure and clustering method, respectively. Here we repeat the clustering using alternative clustering methods: average linkage (a and c) and single linkage (b and d). a. and b. Clustering of the Pearson correlations of the log-transformed FPKM values from 14,744 orthologous gene pairs. c. and d. Clustering of Pearson correlations of the batch-corrected, log-transformed normalized fragment counts from 10,309 orthologous gene pairs that remained after the exclusion steps described in the results. Red boxes: all-human cluster; cyan boxes: all-mouse cluster; black boxes: same-tissue human-mouse pairs that were clustered together.