

1 Appendix Materials for

2

3 Transcriptomic and selection signatures of longevity in mammals

4

5 Weiqiang Liu, Pingfen Zhu, Meng Li, Zihao Li, Yang Yu, Gaoming Liu, Juan Du, Xiao Wang,
6 Jing Yang, Ran Tian, Inge Seim, Alaattin Kaya, Mingzhou Li, Ming Li, Vadim N. Gladyshev,
7 Xuming Zhou

8

9 Corresponding author: Xuming Zhou, zhouxuming@ioz.ac.cn

10

11

12 The PDF file includes:

13 Appendix results ----- Page 2- Page 3

14 Appendix Fig S1 to S19----- Page 4- Page 36

15

16

Appendix Results

Batch effect assessment

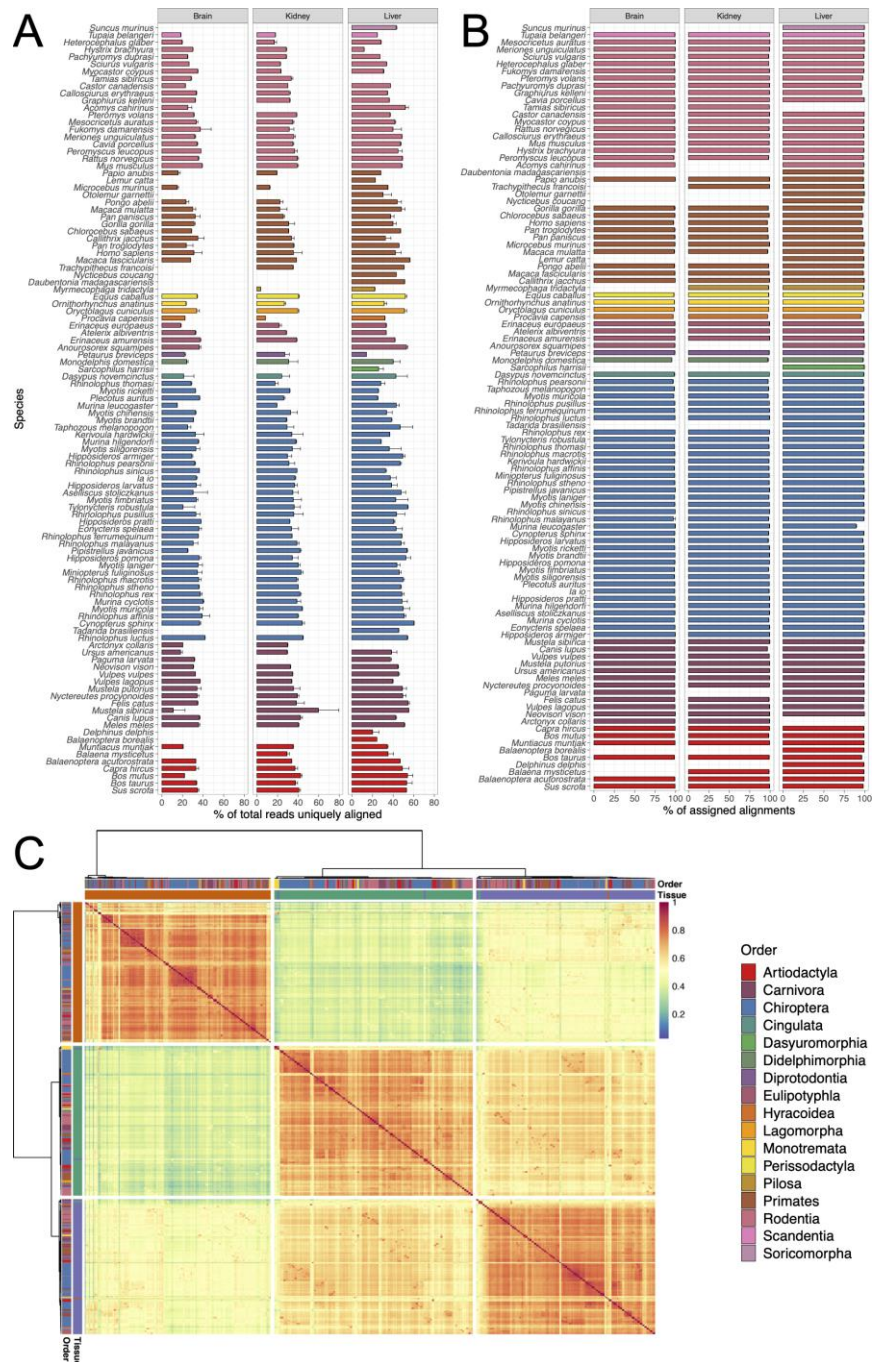
To verify the effectiveness of batch correction, we selected several subsets to compare the batch correction of BioProject, reads type (short- vs. paired-end reads), and sequencing platform. The data were mainly from three sources: new sequencing data generated in this study and data generated by Fushan *et al.* (2015) and Brawand *et al.* (2011). We performed PCA and conducted a Kruskal-Wallis rank sum (K-W) test on PC1 before and after batch correction. The results showed that before correction PC1 separates the samples by tissues but also by BioProject (Kruskal-Wallis chi-squared = 30.19, $P = 0.02$). After batch correction, PC1 separated the samples by tissue but not by BioProject (Kruskal-Wallis chi-squared = 15.30, $P = 0.50$). In addition, we examined each tissue separately and consistently found that the main components did not distinguish the samples by BioProject after removing batch effect (Liver: $P = 0.71$; Kidney: $P = 0.04$; Brain: $P = 0.05$) (**Appendix Fig S17** and **Table EV5**).

We then tested the influence of the read type using the data set generated by Brawand *et al.* (2011) (NCBI BioProject: PRJNA143627). This data set includes brain RNA-seq reads of different types (76bp single-end and 100bp paired-end) generated on the Illumina Genome Analyzer IIX platform. PC1 and PC2 separated samples by species but not read type before and after batch removal (**Appendix Fig S18**), indicating the read type has a limited influence on gene expression variation.

Finally, we used the data from the ferret (*Mustela putorius*) and sugar glider (*Petaurus breviceps*) to examine the batch effect of sequencing platforms. The ferret data was generated on the Illumina HiSeq 2000 platform, and the sugar glider data using the Illumina Genome Analyzer IIX. Our data from the same species were generated on the Illumina Novaseq 6000 instrument. PC1 and PC2 separated samples by species but not by

46 sequencing platforms (**Appendix Fig S19**), indicating that the sequencing
47 platform does not significantly affect gene expression.

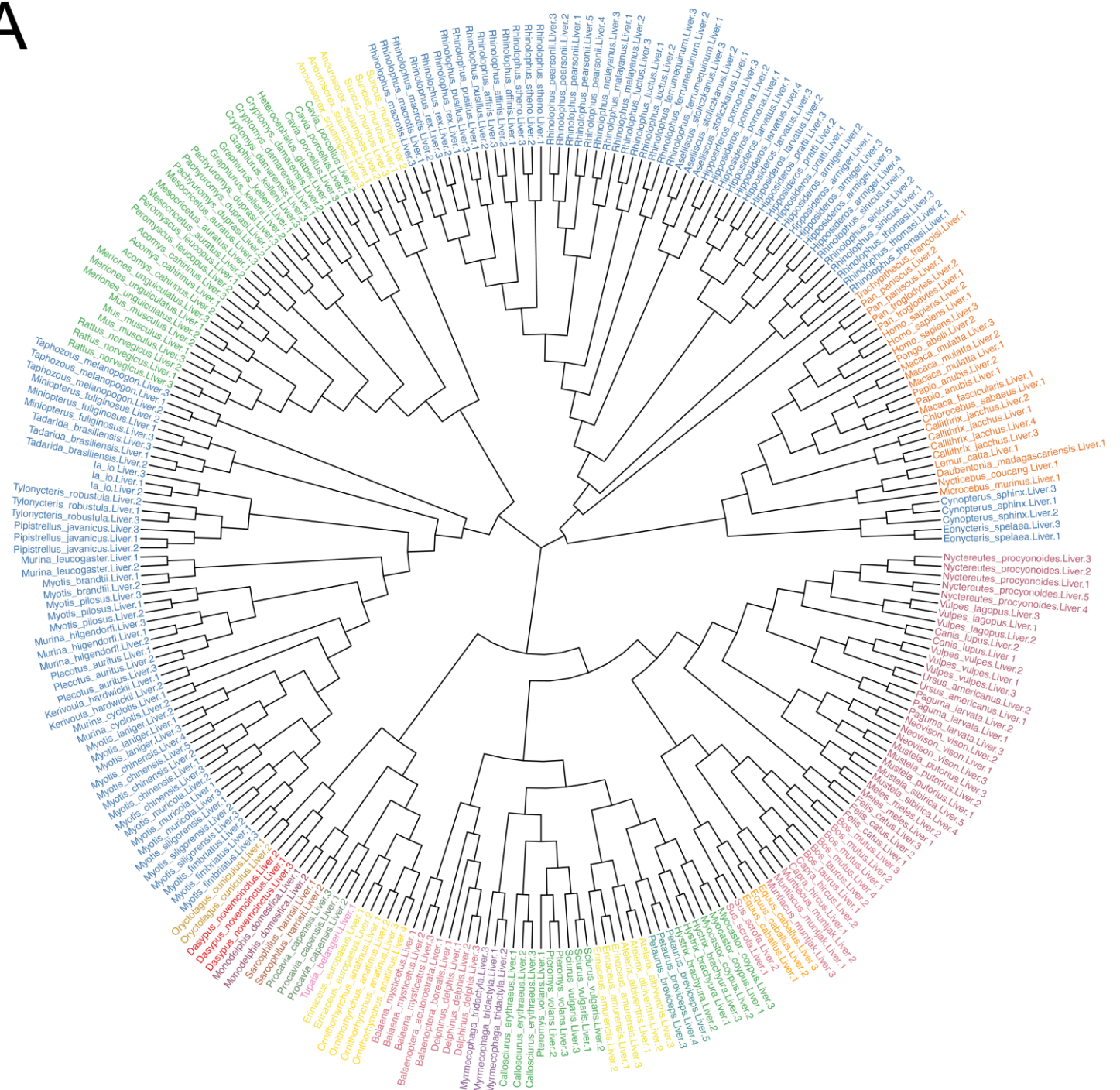
48 **Appendix Figure**



49 **Appendix Fig S1 Sample alignment statistics and inter-sample correlation. (A)**
50 RNA-seq unique mapping rates in each tissue. (B) The ratio of reads assigned to genes
51 in each tissue. (C) Heatmap of correlation coefficient between samples. All group colors
52 represent different order.

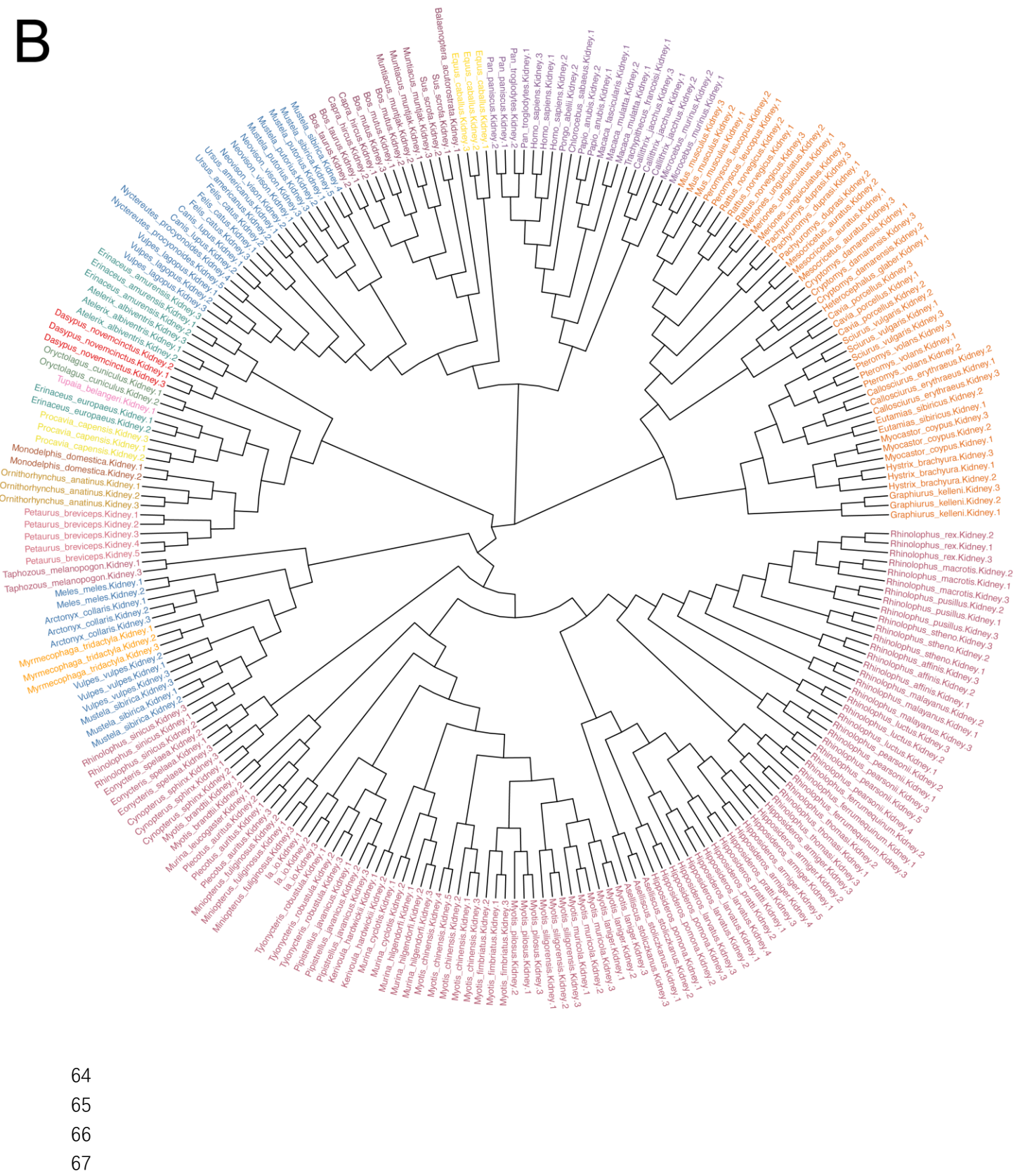
53
54
55

A



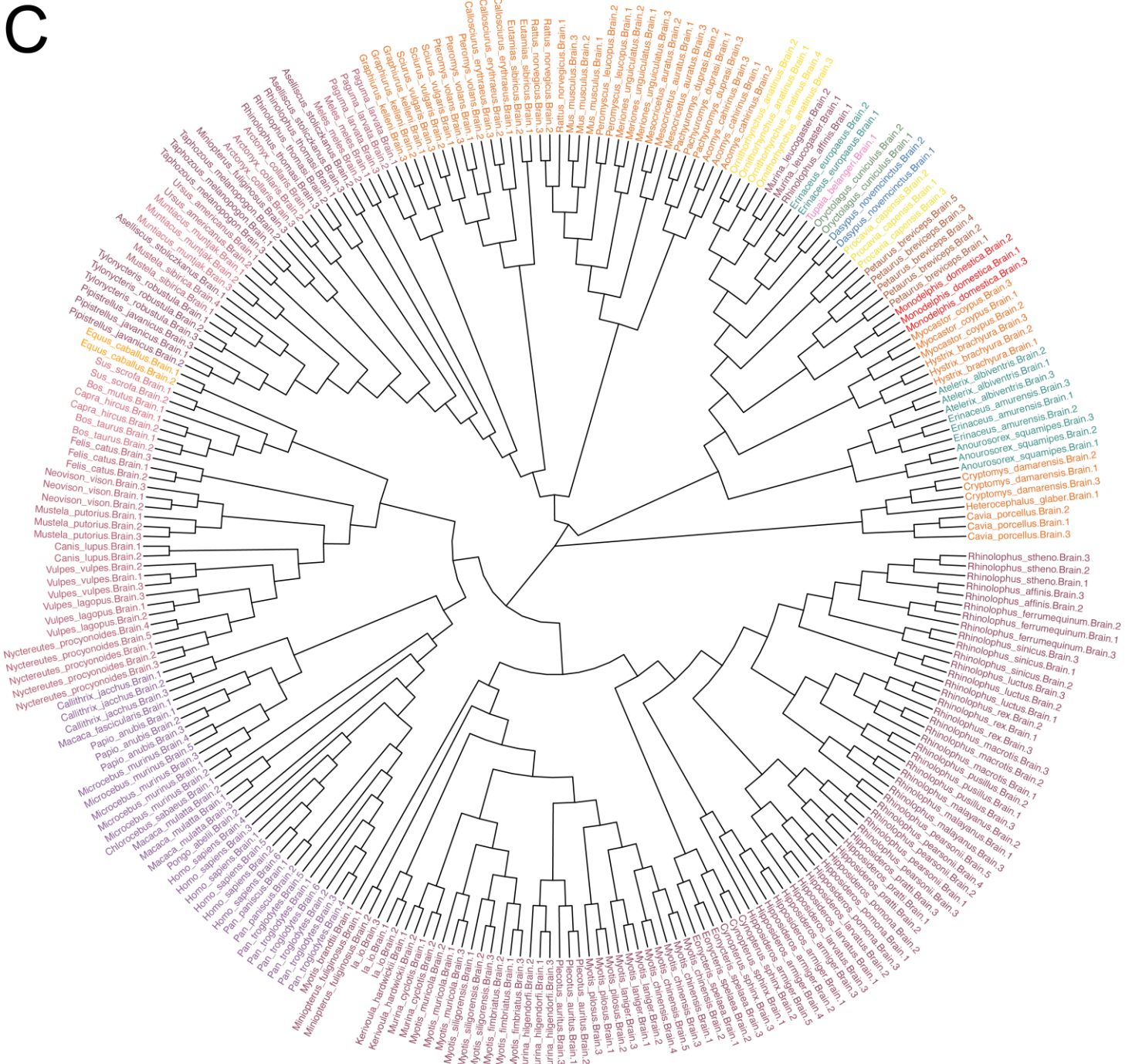
56
57
58
59

B



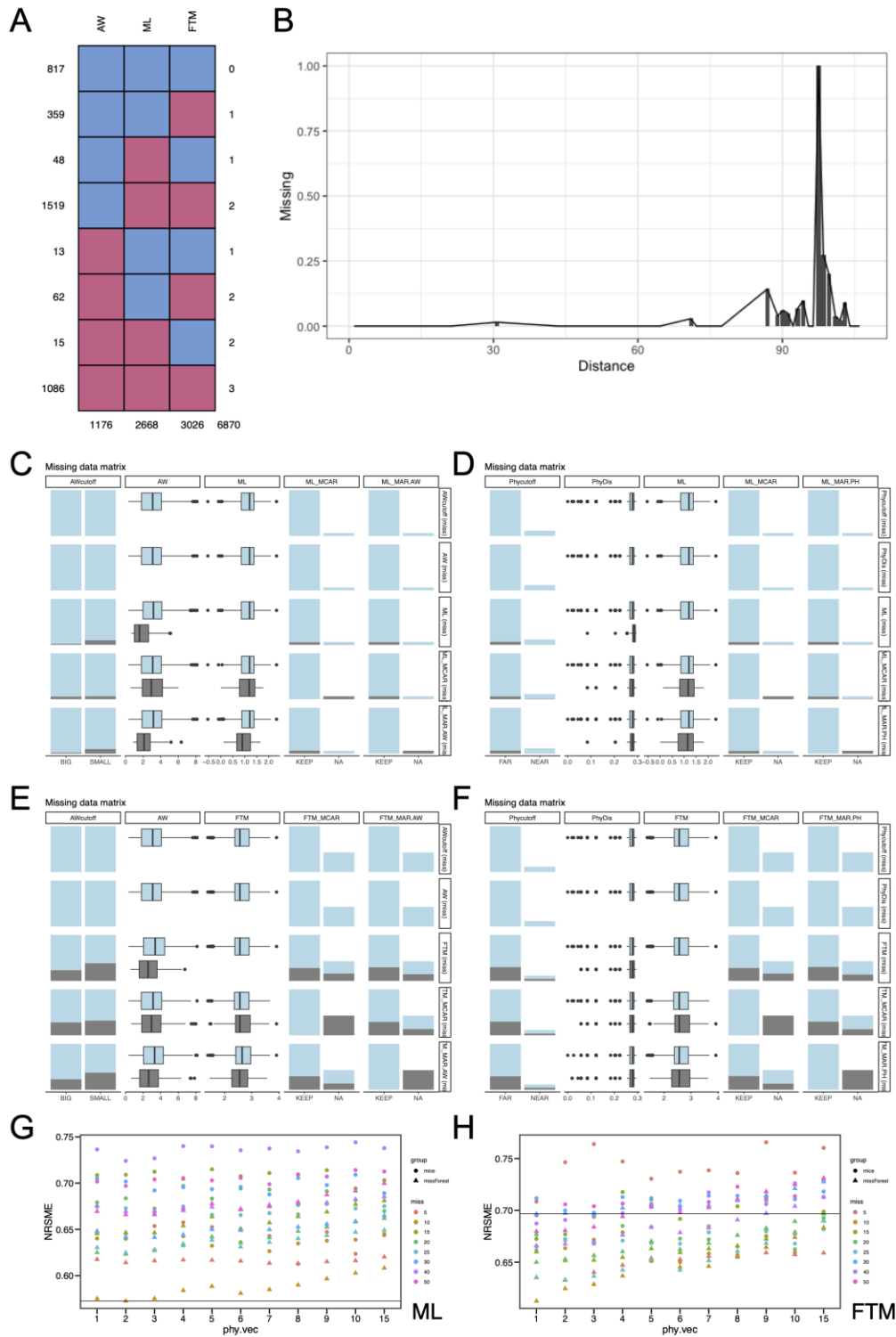
68
69
70
71

C



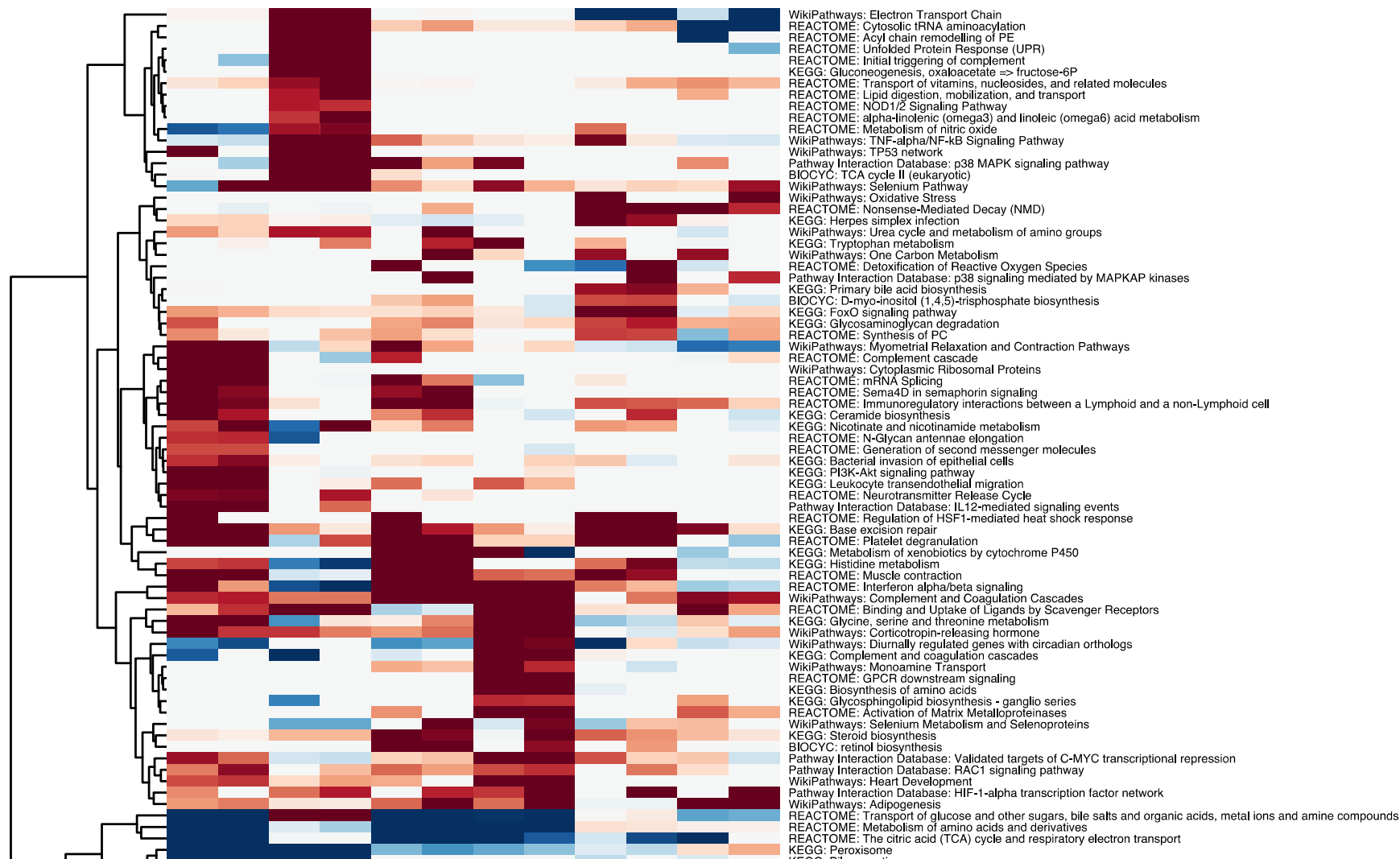
72

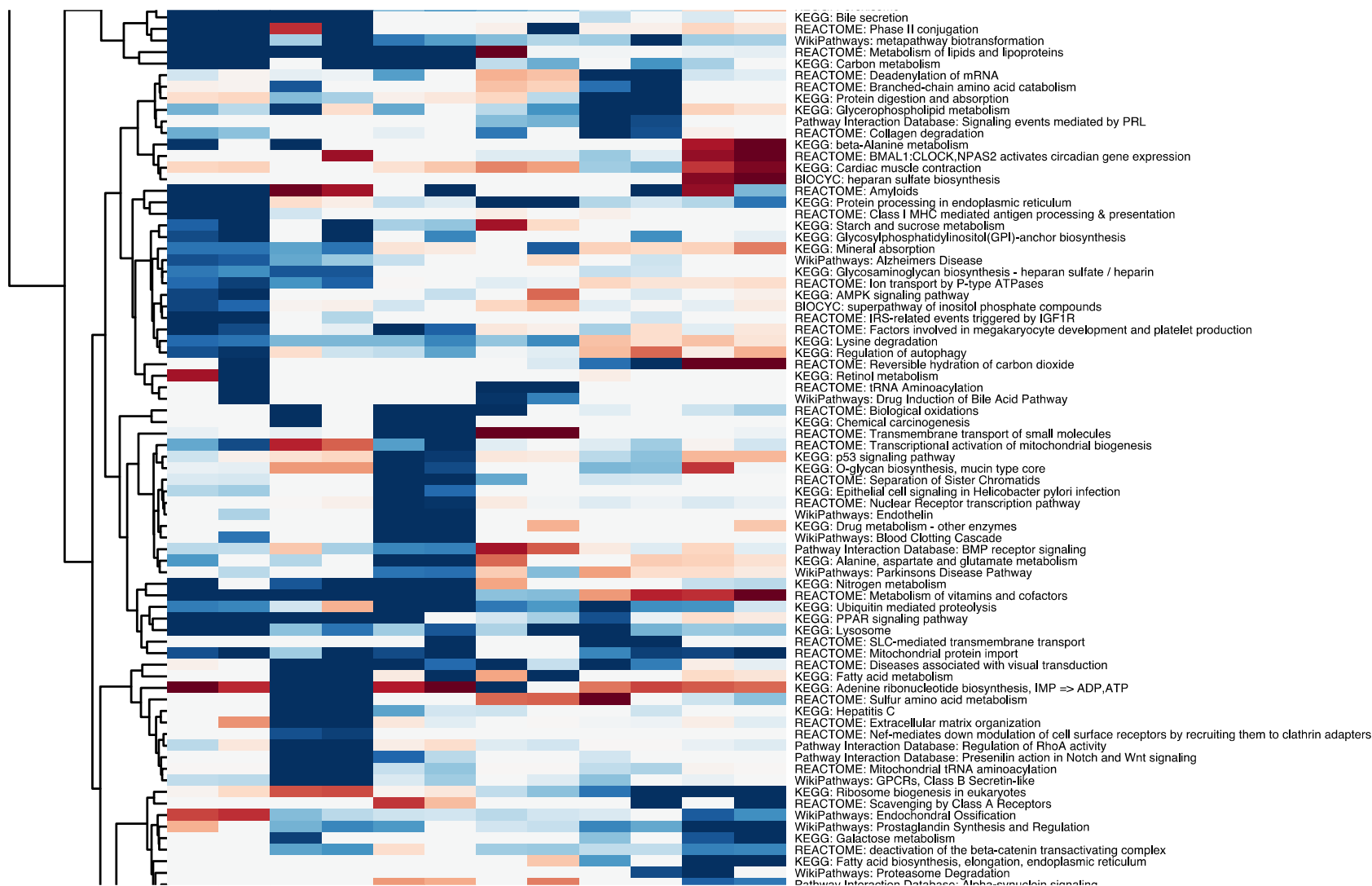
73 **Appendix Fig S2 Sample clustering after outlier removal.** The distance between
74 samples is 1-Pearson correlation coefficient, and the tree is constructed by the neighbor-
75 joining (NJ) method. (A) liver, (B) kidney and (C) brain.

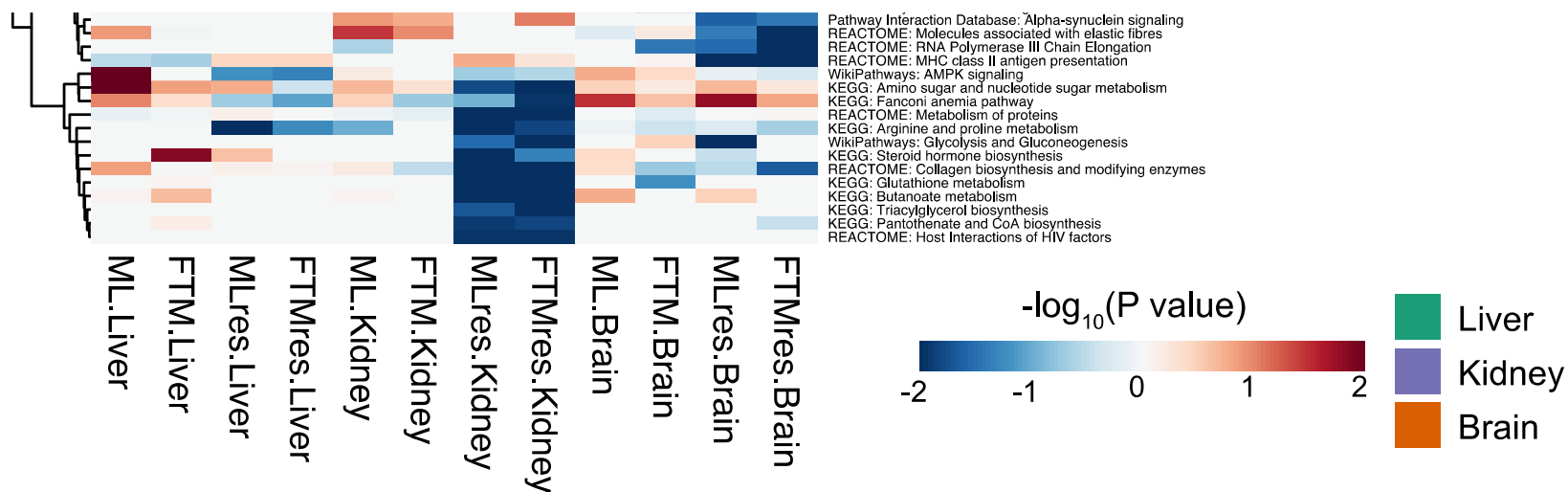


76 **Appendix Fig S3 Life history data imputation.** (A) Distribution of missing values in
 77 different life histories. (B) Distribution of missing values with increasing genetic distance
 78 from humans. (C) Correlation between missing value of maximum lifespan and adult
 79 weight. (D) Correlation between missing value of maximum life span and genetic
 80 distance from human. (E) Correlation between missing value of female time to maturity

and adult weight. (*F*) Correlation between missing value of female time to maturity and genetic distance from human. (*G*) For the maximum lifespan, the gradient estimation accuracy comparison of the number of phylogenetic vectors added in the model. The colors represent different proportions of missing values, and the shapes represent two imputation methods. (*H*) For the female time to maturity, the gradient estimation accuracy comparison of the number of phylogenetic vectors added in the model. The colors represent different proportions of missing values, and the shapes represent two imputation methods.

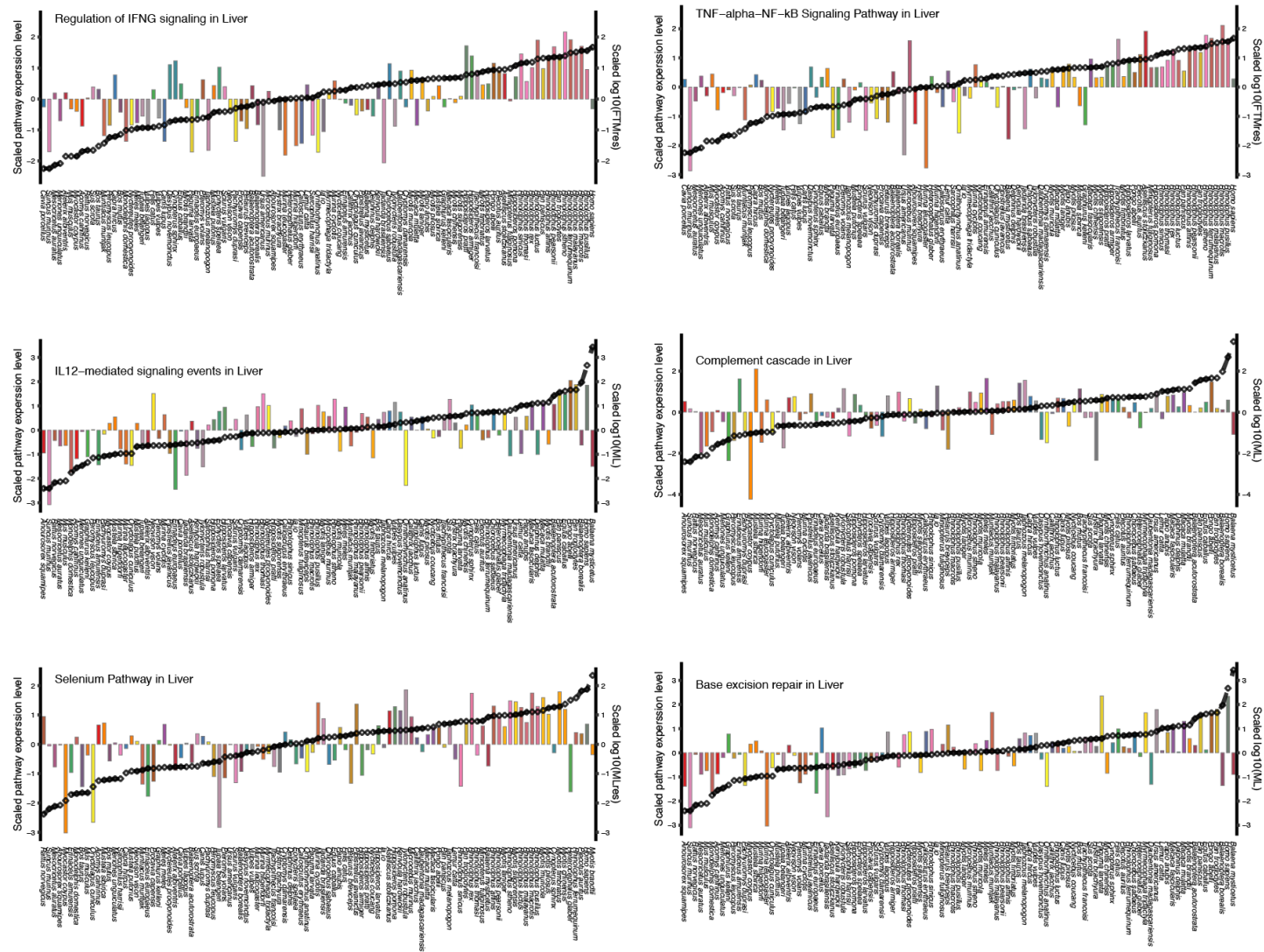


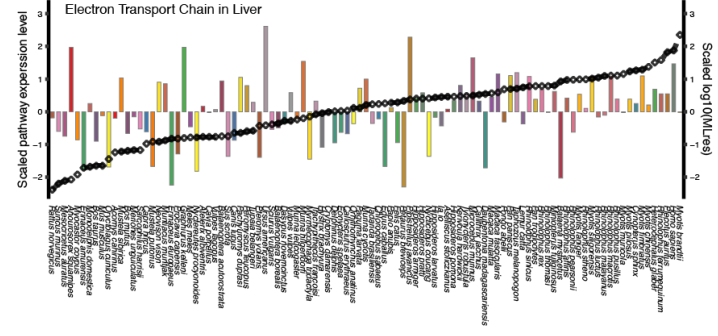
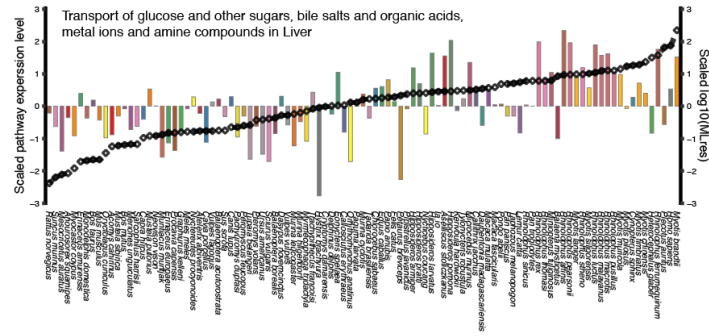
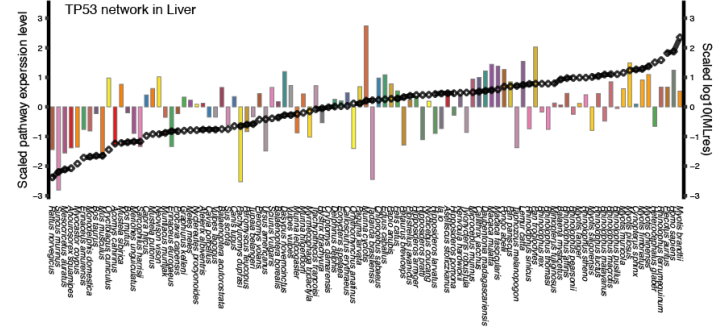
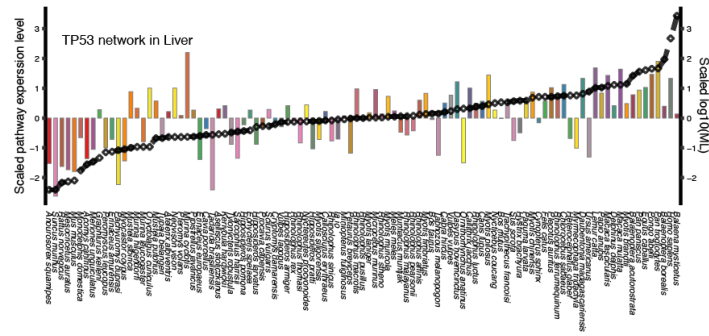




Appendix Fig S4 Gene set enrichment analysis heat map for all traits (postpruning). Represents a clustering heat map of pathways enriched by longevity-related genes and AW-related genes. The color intensity indicates the degree of significance, and the P value has undergone a $-\log_{10}$ transformation. Each row represents a different pathway, and each column represents related longevity traits (marked at the bottom). Among them, red is positive correlation, and blue is negative correlation.

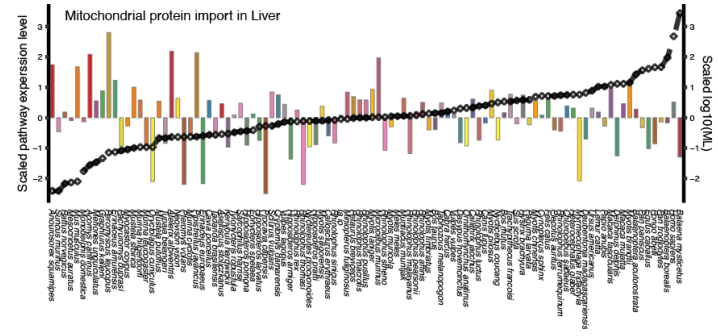
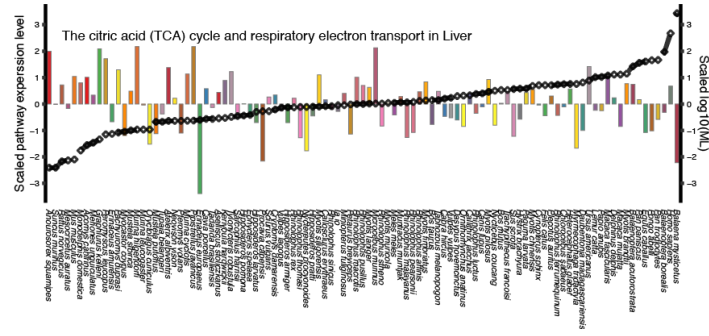
Liver

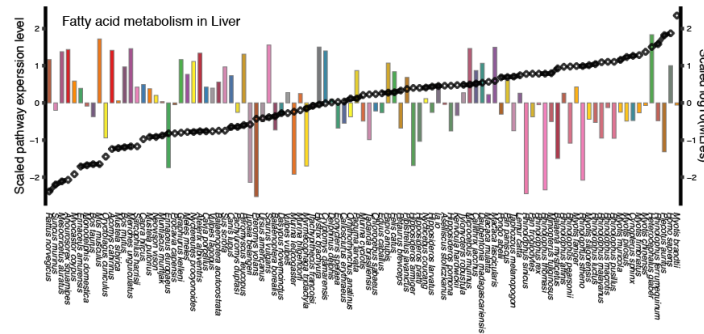




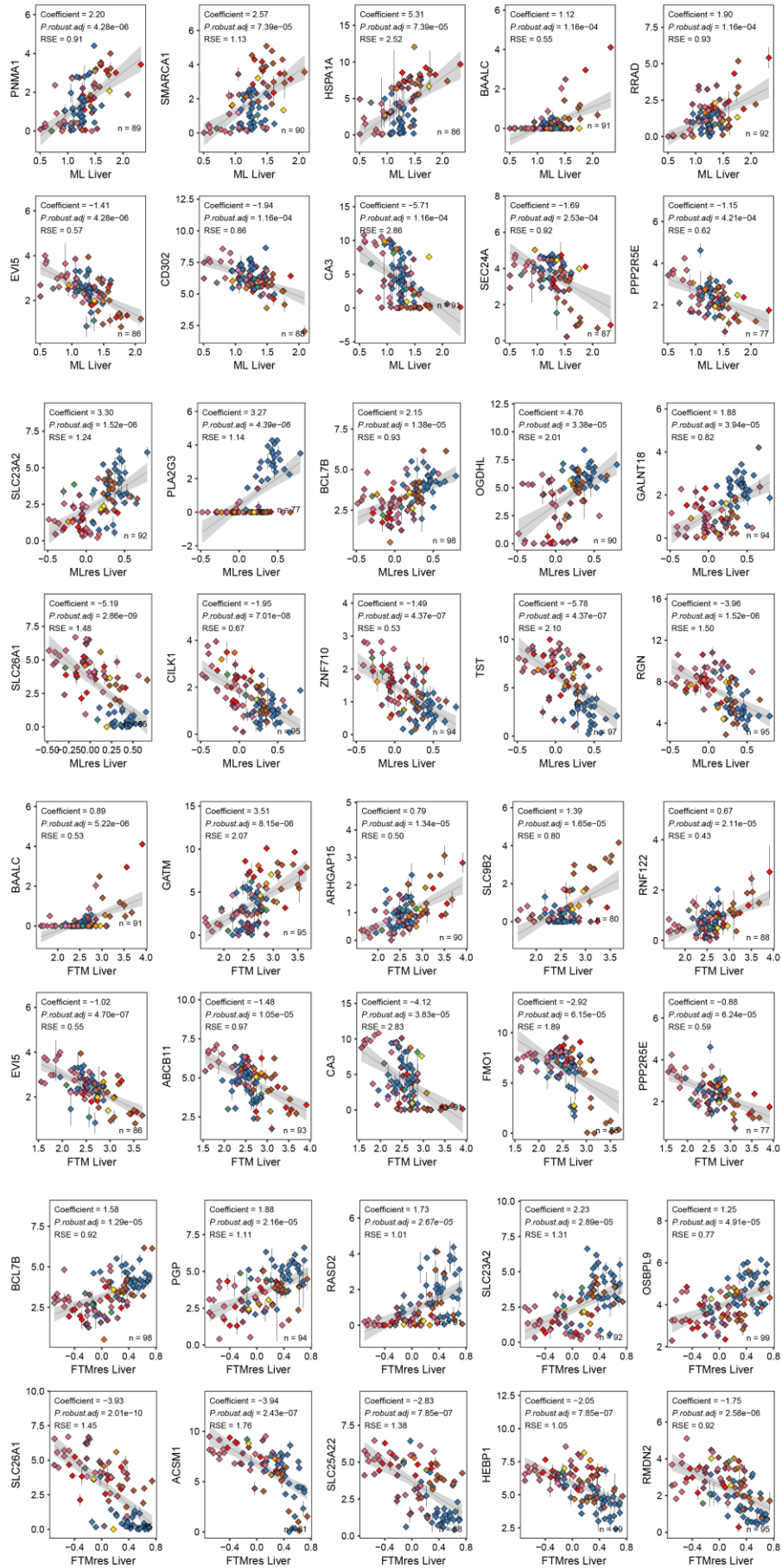
Positive

Negative



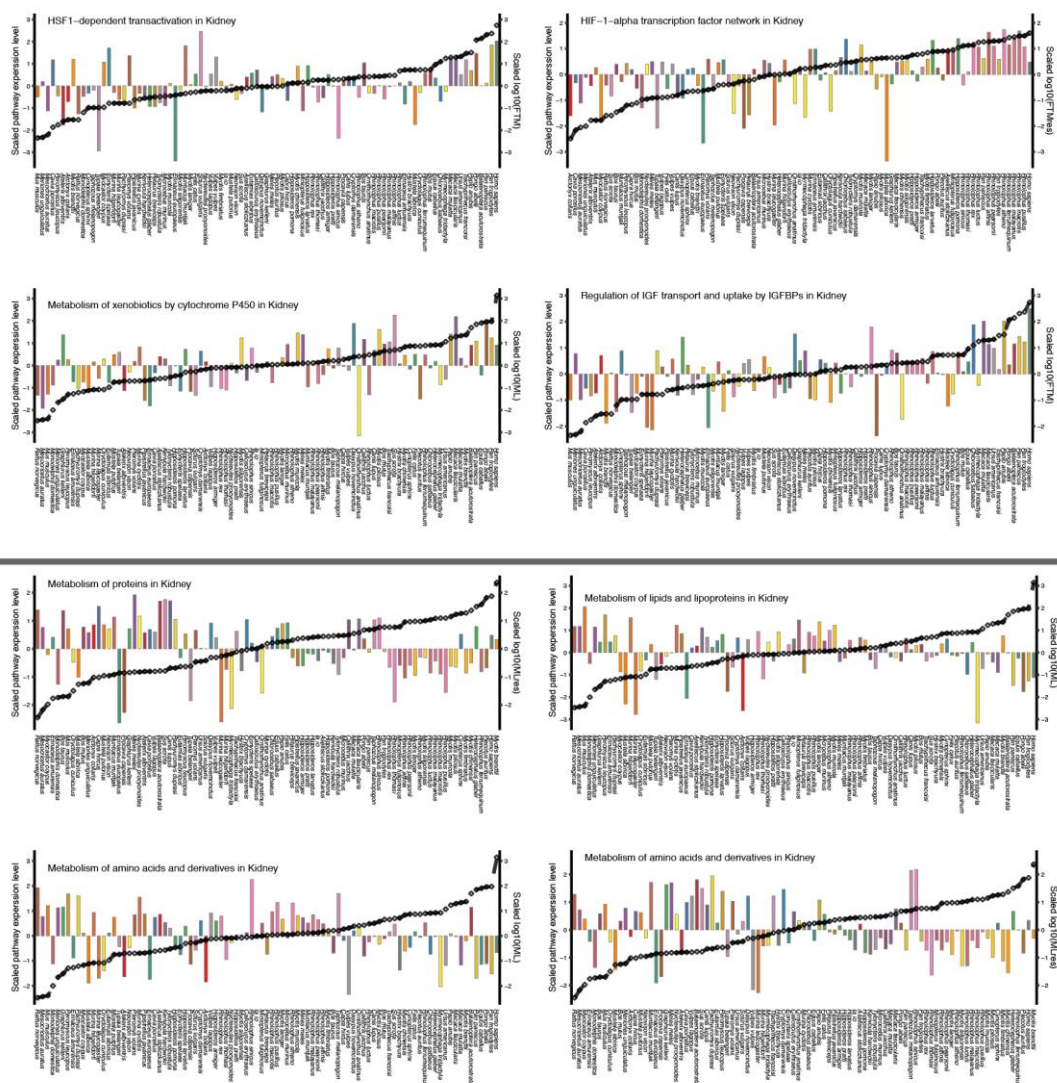


99 **Appendix Fig S5 Top pathway for gene set enrichment in liver.** The y-axis on the left represents the sum of the expression levels of all genes in
 100 the alas pathway of each species. The y-axis on the right represents the value of longevity-related traits are centered at 0 on log10 scale. The x-axis is
 101 the species name, and the ranking increases with the longevity-related traits. The upper part of the gray line is the positive correlation pathway, and
 102 the lower part is the negative correlation pathway.



Appendix Fig S6 Top five up/down-regulated genes with significant correlation to longevity in liver. In each figure, the y-axis is the scaled expression level of each gene with 0 as the center, and the x-axis is the longevity traits (ML: maximum lifespan; FTM: female time to maturity; MLres and FTMres: ML and FTM residuals adjusted for adult weight, respectively). Error bars represents SE. Potential outliers have been removed. The coefficient of phylogenetic regression, adj.P.robust and residual standard error (RSE) is included in the figure.

Kidney

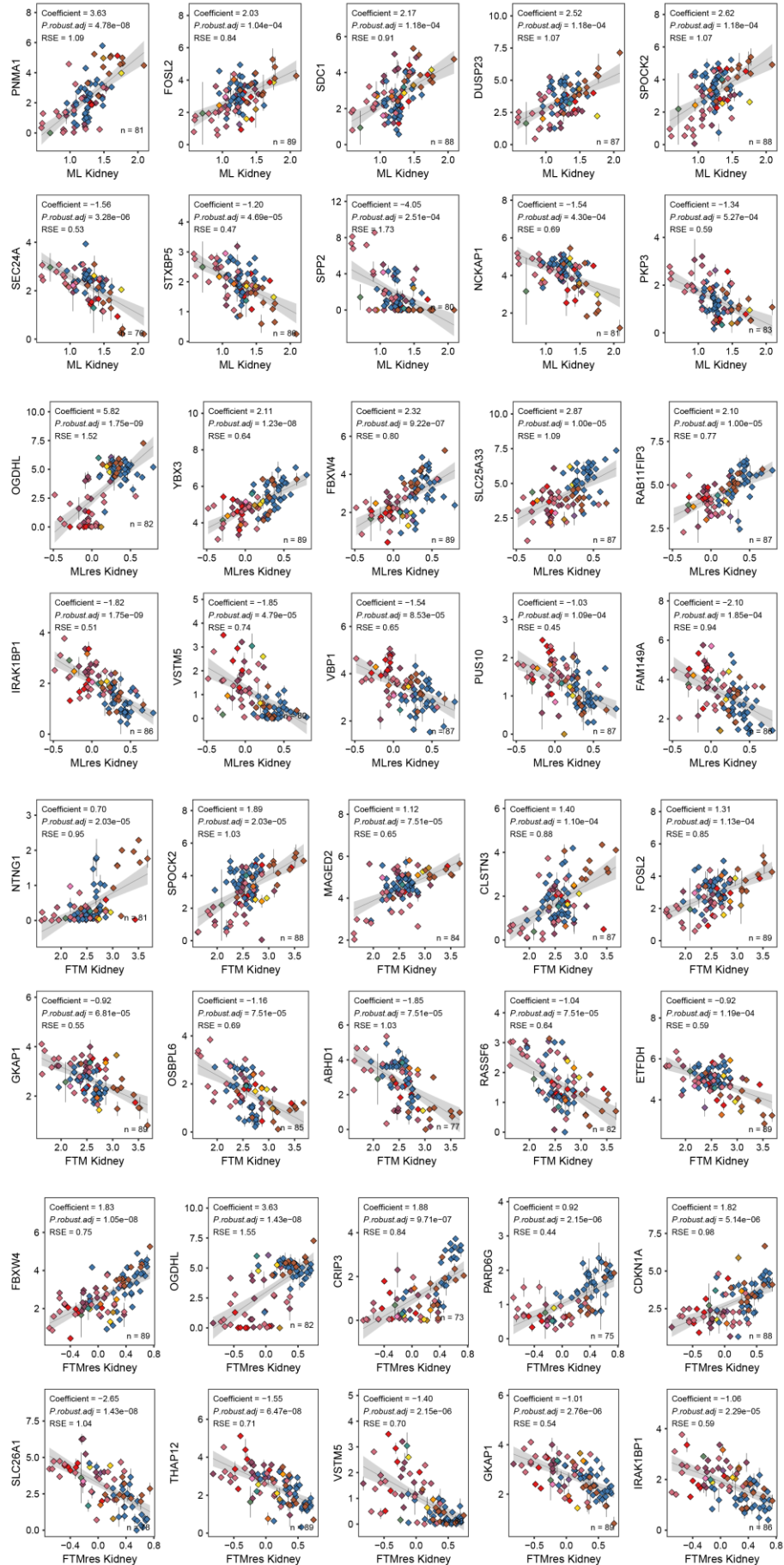


Positive

Negative

112 **Appendix Fig S7 Top pathway for gene set enrichment in kidney.** The y-axis on the
 113 left represents the sum of the expression levels of all genes in the alas pathway of each
 114 species. The y-axis on the right represents the value of longevity-related traits are
 115 centered at 0 on log10 scale. The x-axis is the species name, and the ranking increases
 116 with the longevity-related traits. The upper part of the gray line is the positive correlation
 117 pathway, and the lower part is the negative correlation pathway.

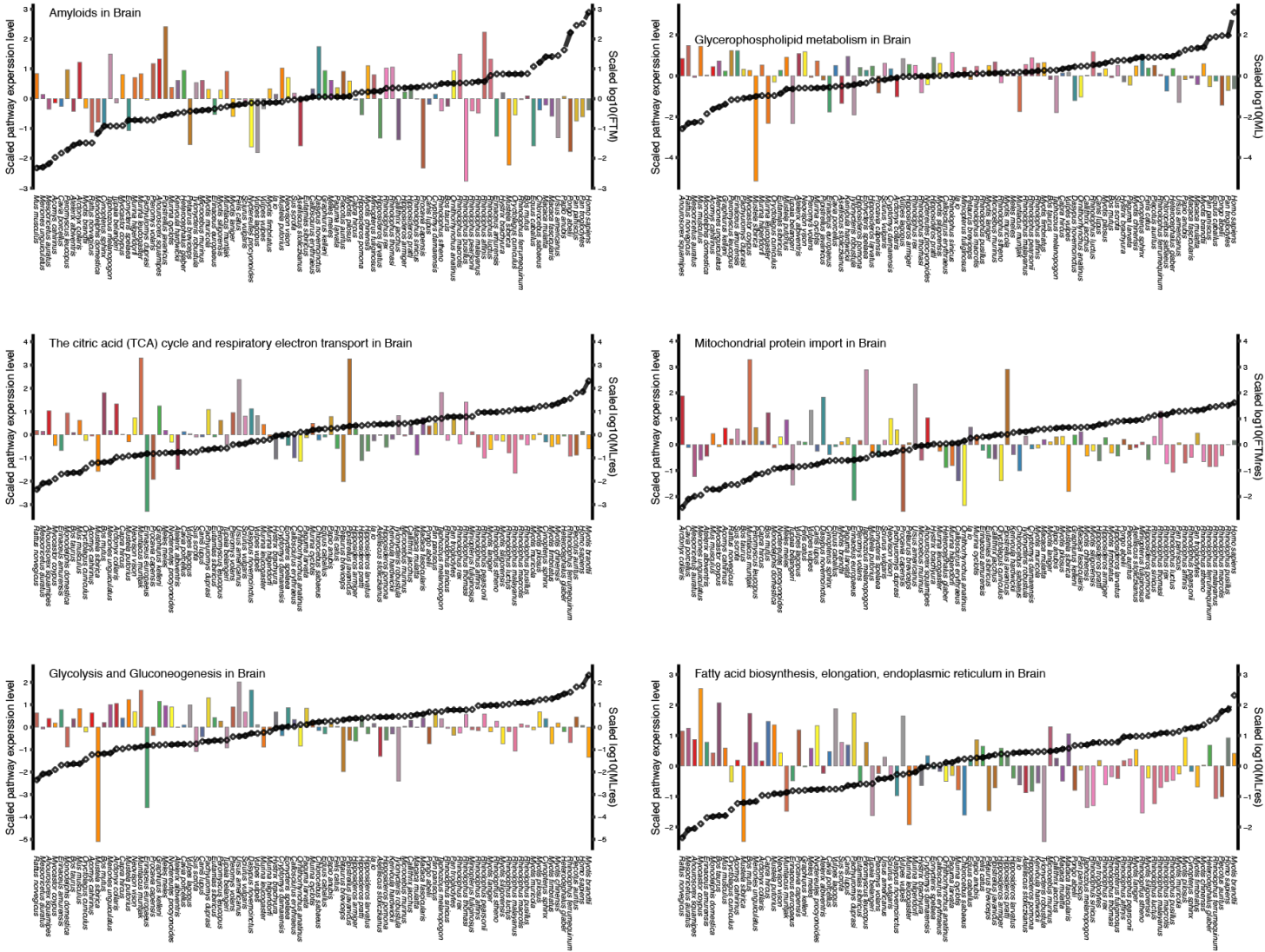
118



Appendix Fig S8 Top five up/down-regulated genes with significant correlation to longevity in kidney. In each figure, the y-axis is the scaled expression level of each gene with 0 as the center, and the x-axis is the longevity traits (ML: maximum lifespan; FTM: female time to maturity; MLres and FTMres: ML and FTM residuals adjusted for adult weight, respectively). Error bars represents SE. Potential outliers have been removed. The coefficient of phylogenetic regression, adj.P.robust and residual standard error (RSE) is included in the figure.

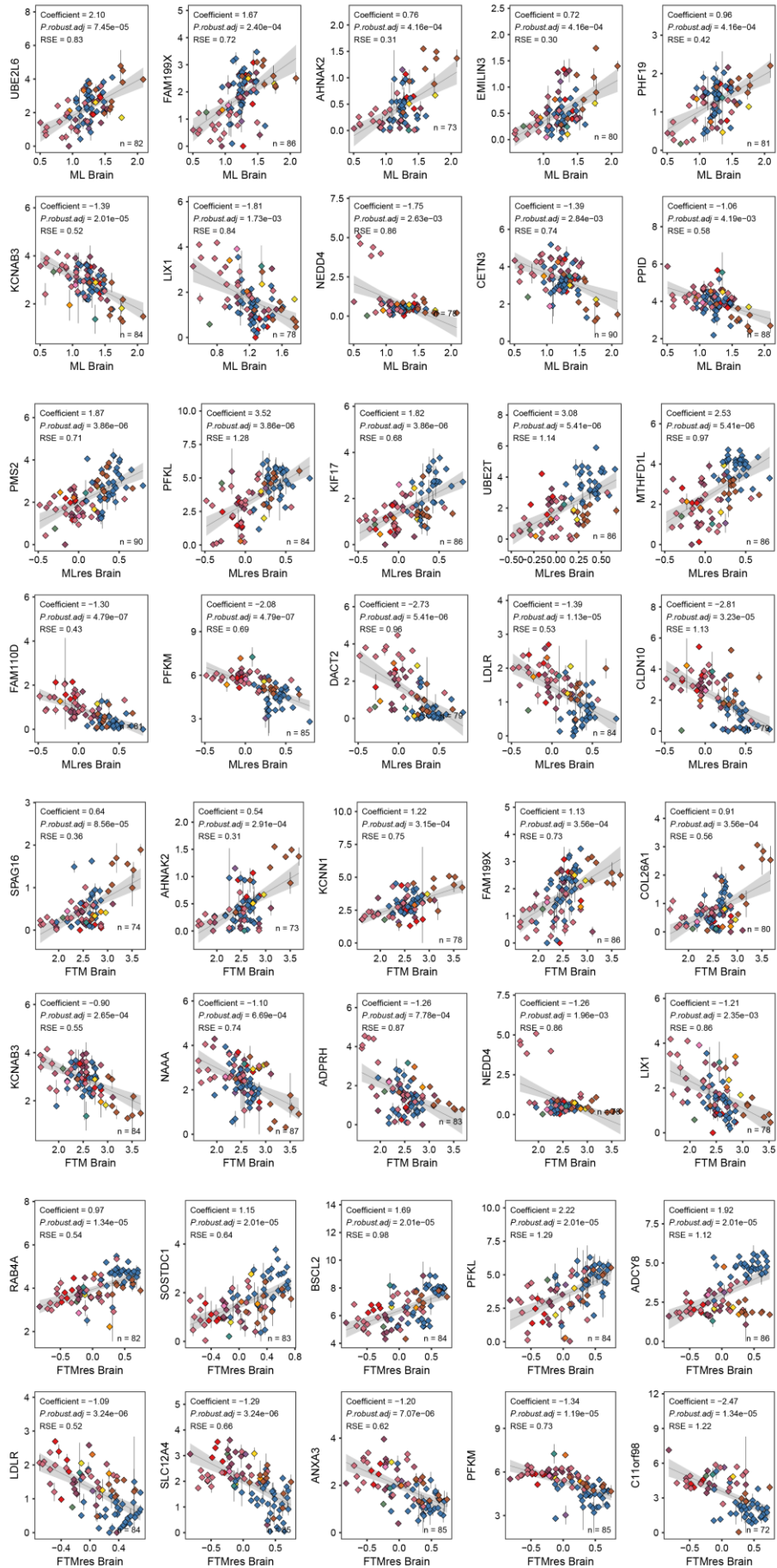


Negative

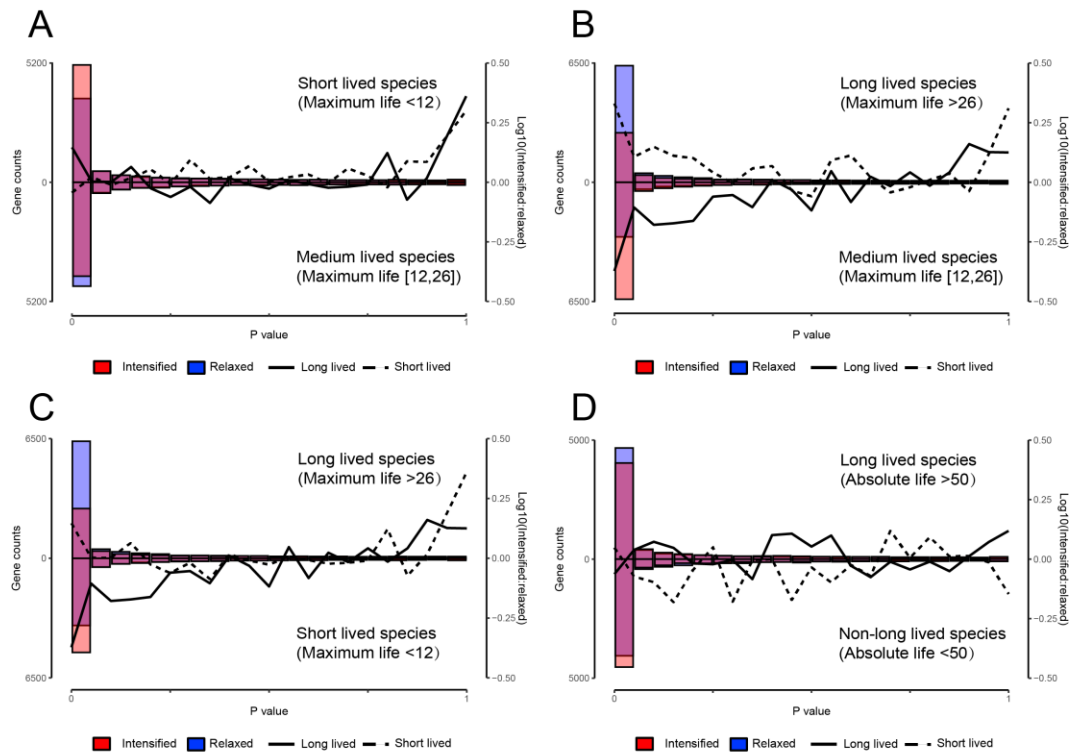


130 **Appendix Fig S9 Top pathway for gene set enrichment in brain.** The y-axis on the left represents the sum of the expression levels of all genes in
131 the alas pathway of each species. The y-axis on the right represents the value of longevity-related traits are centered at 0 on log10 scale. The x-axis is
132 the species name, and the ranking increases with the longevity-related traits. The upper part of the gray line is the positive correlation pathway, and
133 the lower part is the negative correlation pathway.

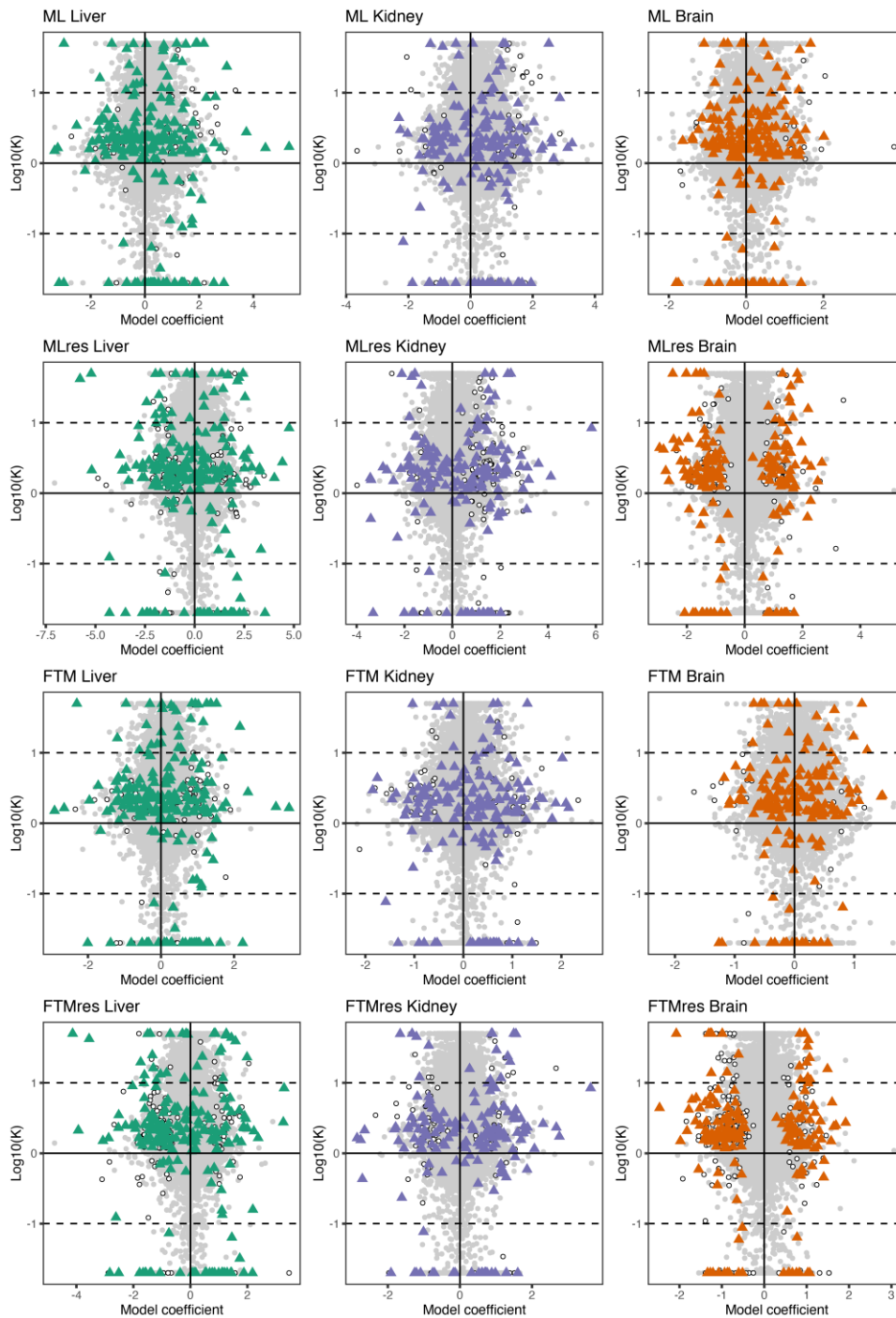
134



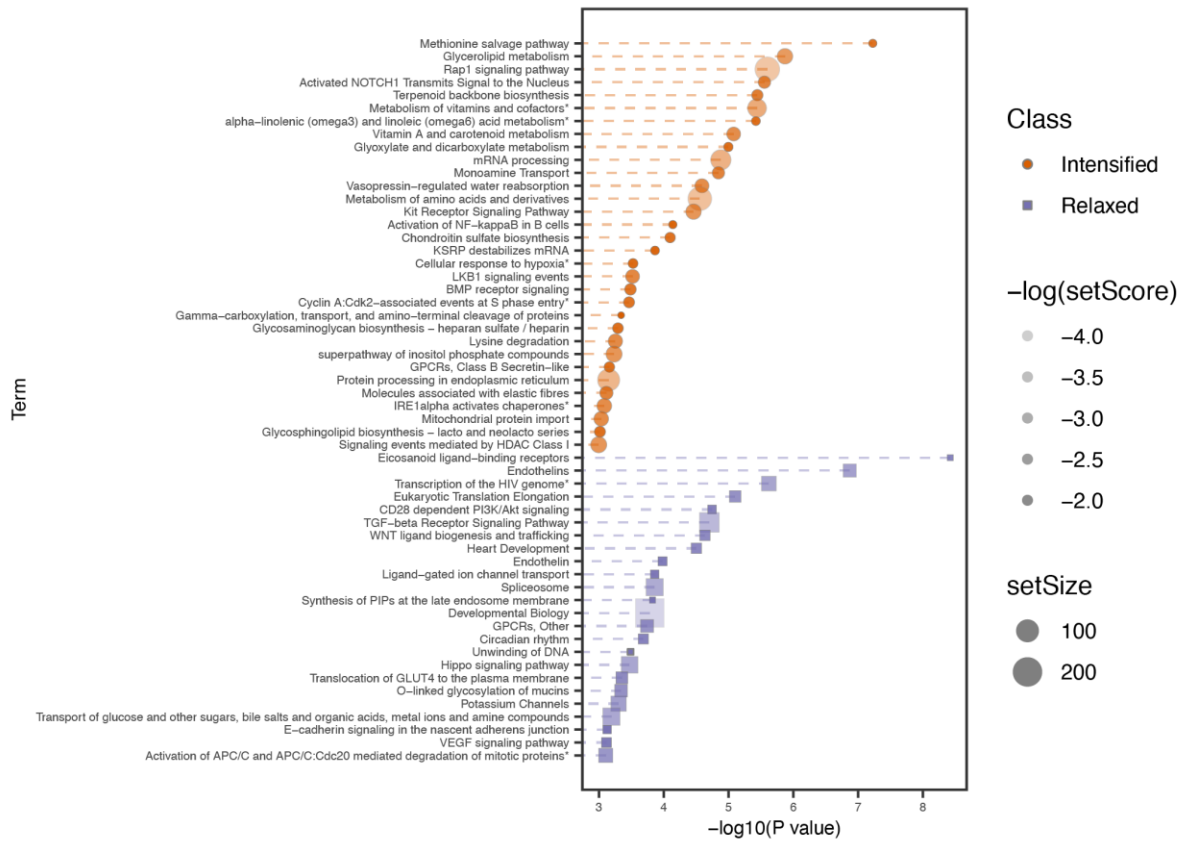
Appendix Fig S10 Top five up/down-regulated genes with significant correlation to longevity in brain. In each figure, the y-axis is the scaled expression level of each gene with 0 as the center, and the x-axis is the longevity traits (ML: maximum lifespan; FTM: female time to maturity; MLres and FTMres: ML and FTM residuals adjusted for adult weight, respectively). Error bars represents SE. Potential outliers have been removed. The coefficient of phylogenetic regression, adj.P.robust and residual standard error (RSE) is included in the figure.



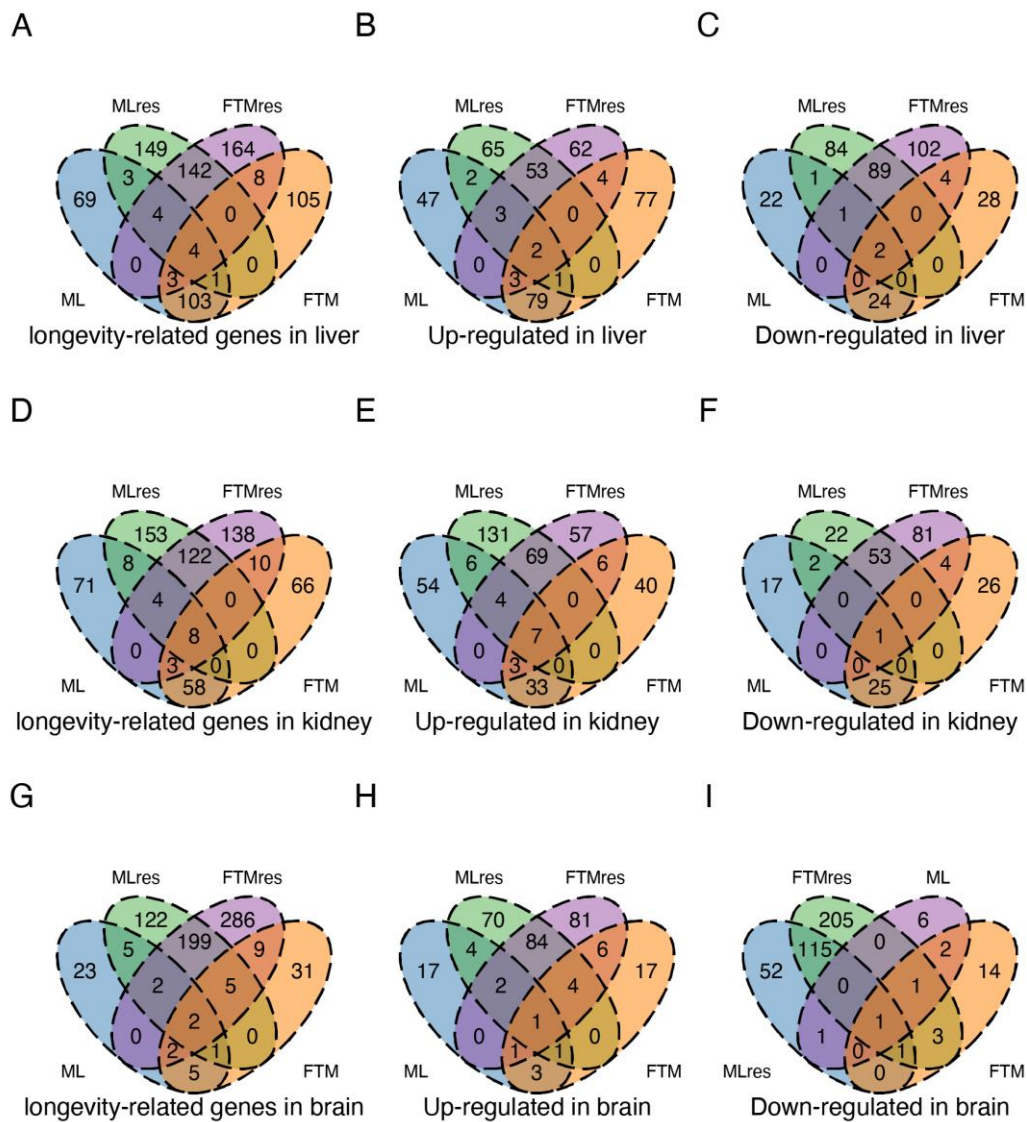
Appendix Fig S11 Relaxation of Selection in different lifespan interval
P value distributions of the RELAX tests of long-lived species (ML > 26 or ML > 50), medium-lived species (12 ≤ ML ≤ 26), short-lived species (ML < 12) and non-long-lived species (ML < 50). (A) Contrast between short-lived species and medium-lived species, (B) Contrast between long-lived species and medium-lived species, (C) Contrast between long-lived species and short-lived species, and (D) Contrast between long-lived species and non-long-lived species are shown. Blue bars, genes under relaxed selection ($k < 1$); red bars, intensified genes ($k > 1$). Dotted (short-lived) and solid (long-lived) lines are the log ratios of $k > 1$: $k < 1$ gene counts.



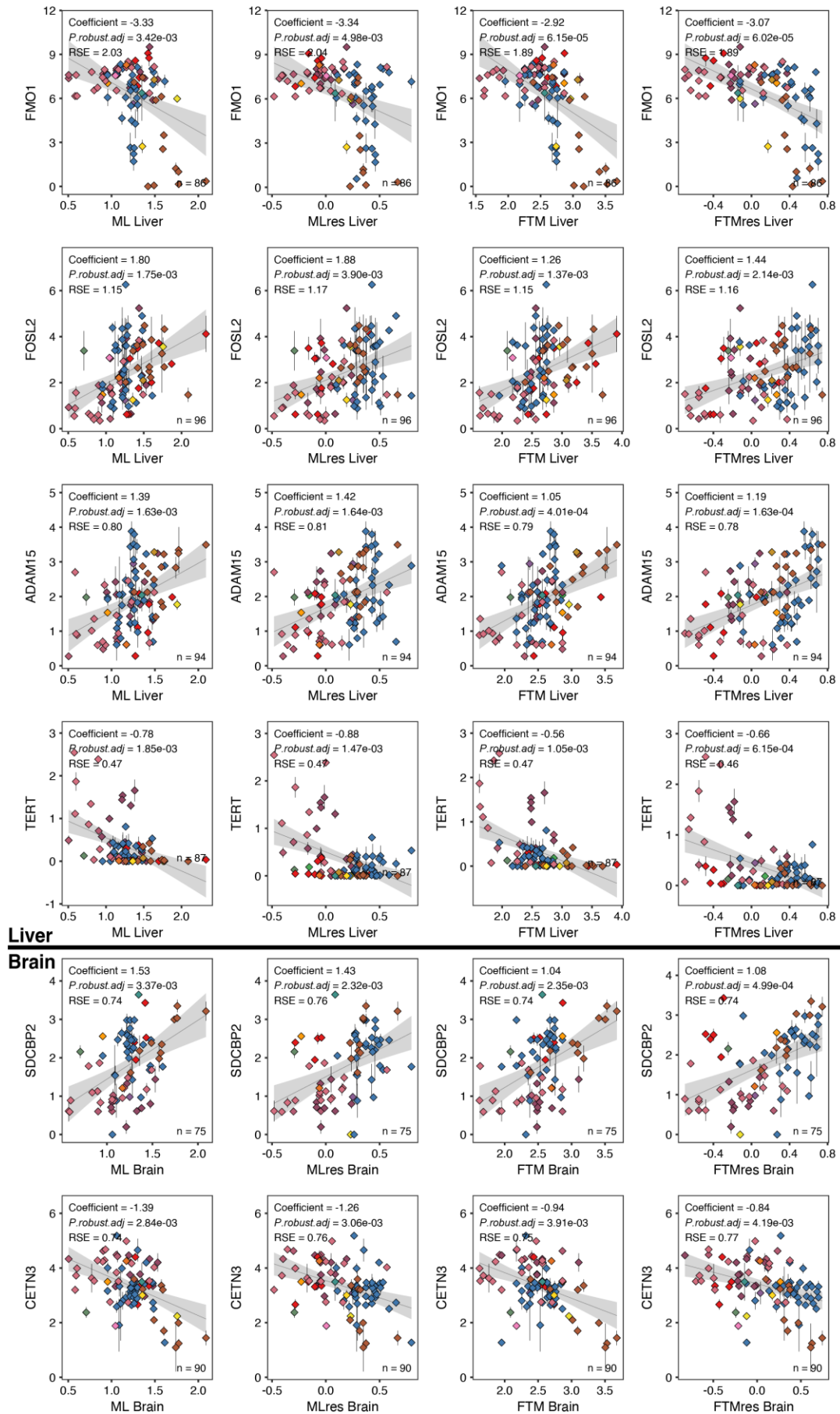
Appendix Fig S12 The relationship between the index of relaxed choice (K) and the regression coefficient. Scatter plot showing the log10-transformed relaxation parameter (k) on the y-axis, and the variation rate of gene expression along the longevity trait gradient on the x-axis. The colored points represent longevity-related genes (the color corresponds to the tissue type shown in Fig.4B), white point represent genes that are only significant related to this trait and the grey points represent genes that are only significant in RLEAX.



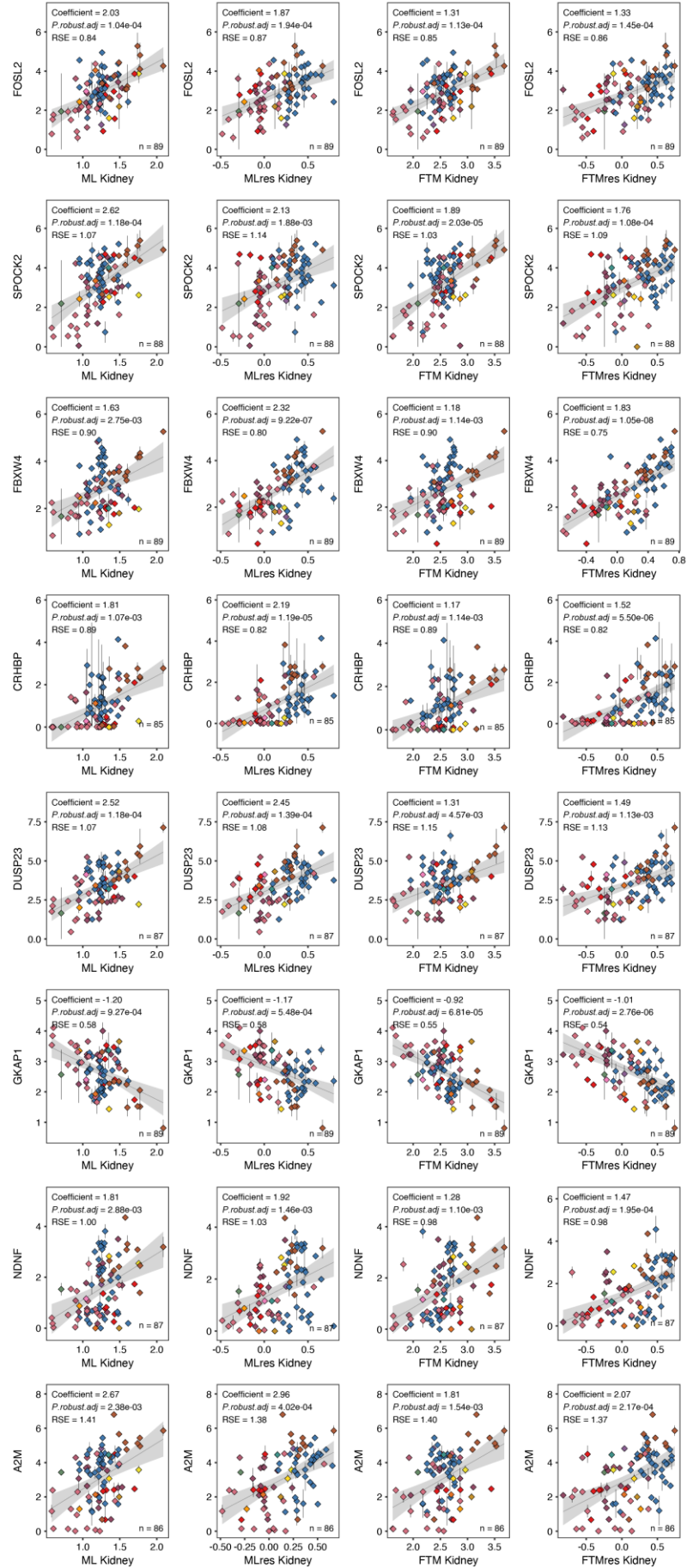
Appendix Fig S13 Gene set enrichment of genes under relaxed or strengthened selection. Polysel gene set enrichment analysis of genes under relaxed or strengthened selection. The x-axis is the P value on $-\log_{10}$ scale. Different selection uses different shapes and colors. The shade of the color represents the SUMSTAT, and the size of the dot represents the size of the gene set.



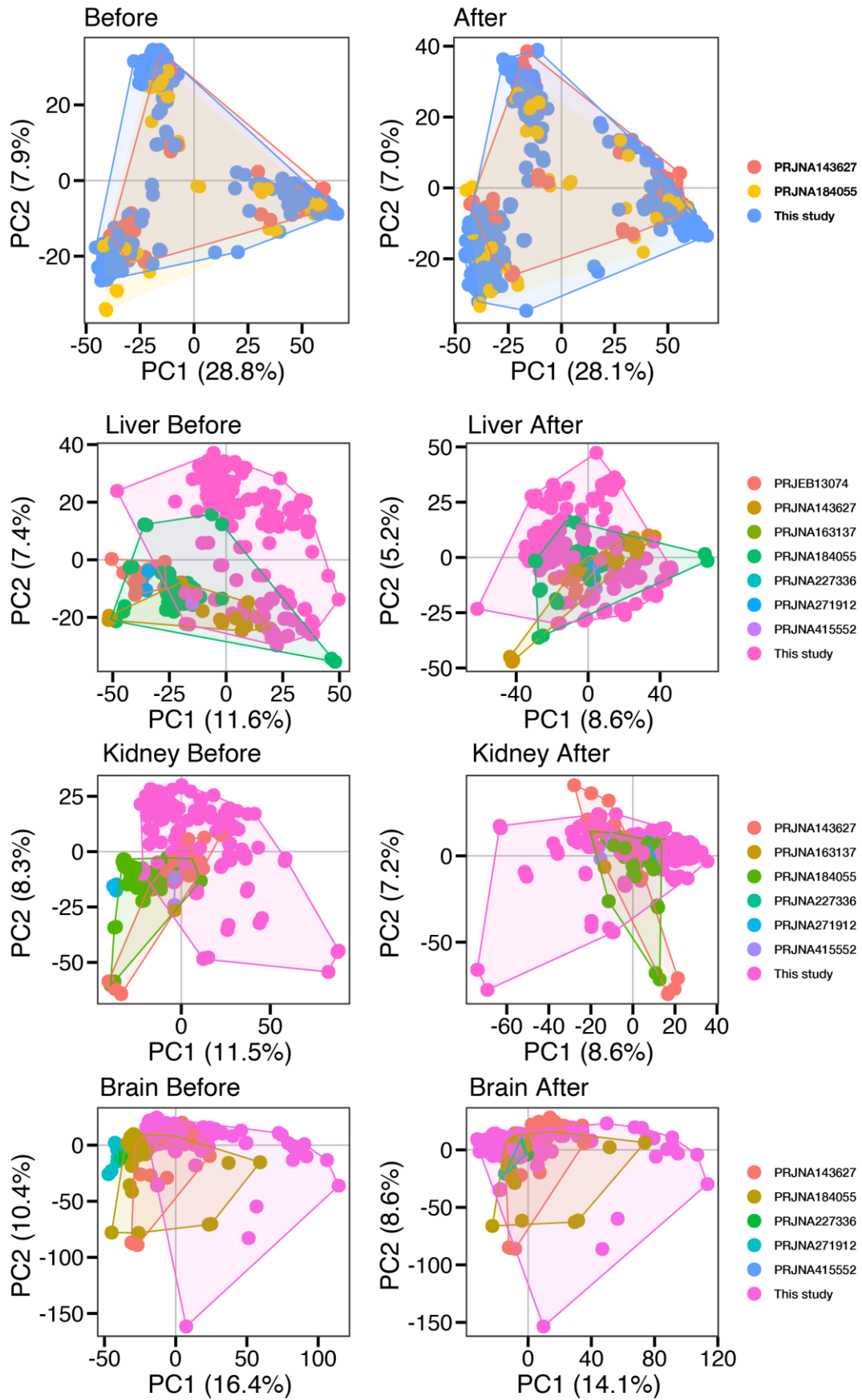
Appendix Fig S14 Overlap of longevity-related genes. (A, D, G) Venn diagrams of each tissue for longevity-related genes. (B, E, H) Venn diagrams of each tissue for positive longevity-related genes. (C, F, I) Venn diagrams of each tissue for negative longevity-related genes.



Appendix Fig S15 Overlap genes in liver and brain. In each figure, the y-axis is the scaled expression level of each gene with 0 as the center, and the x-axis is the longevity traits (ML: maximum lifespan; FTM: female time to maturity; MLres and FTMres: ML and FTM residuals adjusted for adult weight, respectively). Error bars represents SE. Potential outliers have been removed. The coefficient of phylogenetic regression, adj.P.robust and residual standard error (RSE) is included in the figure.

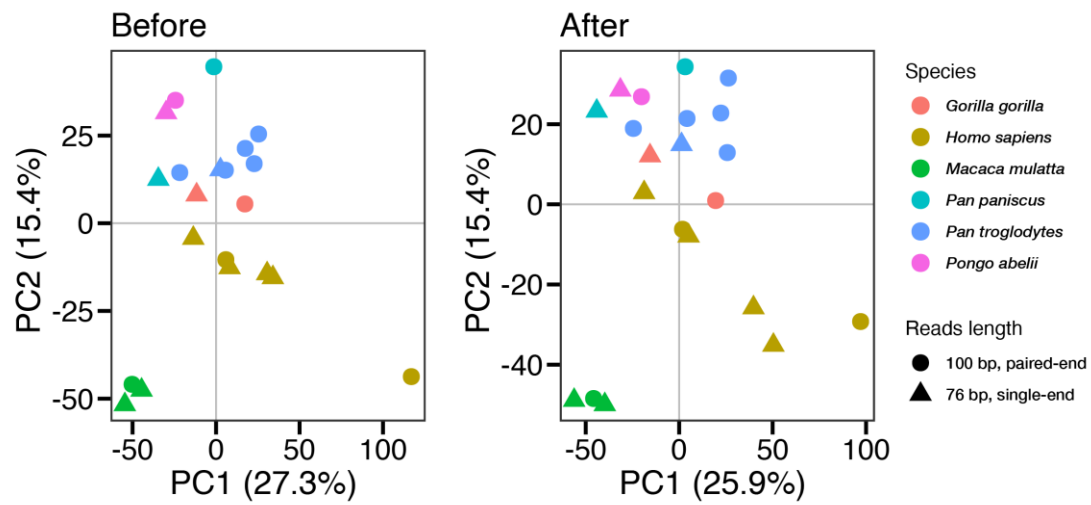


Appendix Fig S16 Overlap genes in kidney. In each figure, the y-axis is the scaled expression level of each gene with 0 as the center, and the x-axis is the longevity traits (ML: maximum lifespan; FTM: female time to maturity; MLres and FTMres: ML and FTM residuals adjusted for adult weight, respectively). Error bars represents SE. Potential outliers have been removed. The coefficient of phylogenetic regression, adj.P.robust and residual standard error (RSE) is included in the figure.



Appendix Fig S17 PCA of before and after batch correction.

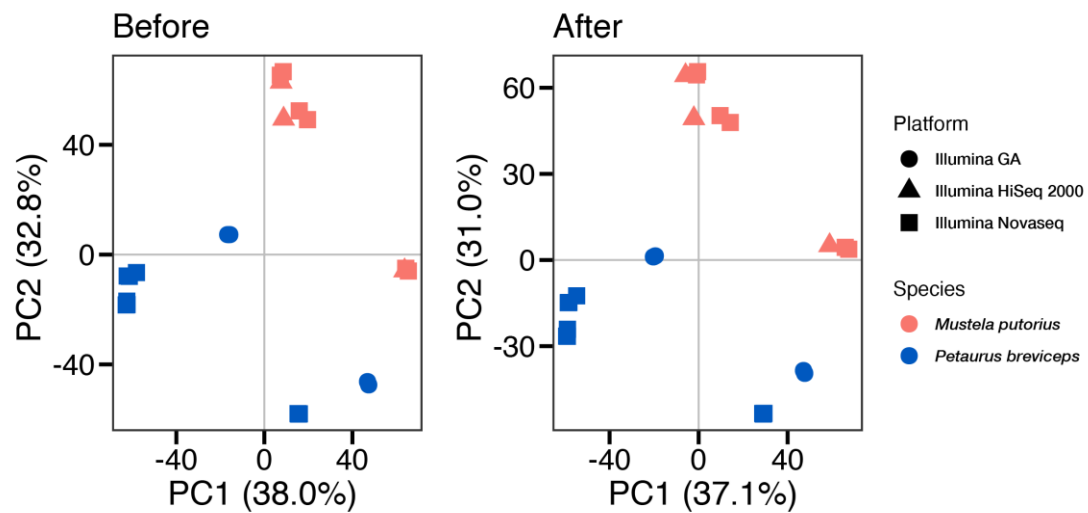
193



194

195 **Appendix Fig S18 Comparison of the impact of reads length on PCA before and**
196 **after batch correction**

197



198

199 **Appendix Fig S19 Comparison of the impact of sequencing platforms on PCA before**
200 **and after batch correction.**