Supplementary data for:

Day-night and seasonal variation of human gene expression across tissues

Valentin Wucher, Reza Sodaei, Raziel Amador, Manuel Irimia, Roderic Guigó

Manuel Irimia
Centre for Genomic Regulation
Dr. Aiguader, 88, 08003 Barcelona, Spain
e-mail: mirimia@gmail.com
Phone: +34933160212, Fax: +34933160099

Roderic Guigó
Centre for Genomic Regulation
Dr. Aiguader, 88, 08003 Barcelona, Spain
e-mail: roderic.guigo@crg.eu
Phone: +34933160168, Fax: +34933160099

Supplementary Table Legends

Table S1 - Genes with day-night expression in human tissues. Number of genes with differentially upregulated expression during the day or during the night, and log2 of the ratio. Corresponding to Fig. 2A.

Table S2 - Core clock genes.

Table S3 - Homologous tissues for human and baboon. Compared homologous tissues between the human GTEx dataset and baboon dataset from Mure *et al.*

Table S4 - Homologous tissues for human and mouse. Compared homologous tissues between the human GTEx dataset and mouse dataset from Li *et al.*

Table S5 - Day-night variation by gene. For each human gene, number of tissues with differentially up-regulated expression during the day and the night.

Table S6 - Genes associated with sleep traits. List of 254 protein-coding genes expressed in GTEx and that were previously reported to increase the risk of insomnia or to be associated with other sleep traits in humans.

Table S7 - Genes with seasonal expression in human tissues. Number of unique genes with differentially up- or down-regulated expression in at least one season respect to the rest, and log2 of the ratio.

Table S8 - Overlap between day-night and seasonal variation. For each tissue, number of genes with day-night expression, seasonal expression, overlap between the two sets and p-value for the overlap (Fisher's Exact test with and without Benjamini - Hochberg multiple testing correction).

Table S9 - Genes with by type of seasonal expression in human tissues. Number of genes with differentially up- or down-regulated expression at each season respect to the rest.

Table S10 - Seasonal variation by gene. For each human gene, number of tissues with differentially up- or down-regulated expression at each season.

Table S11 - List of human hormone genes. List of 62 genes with hormone-encoding capability based on Mirabeau *et al.*

Table S12 - Cell type specific markers. List of genes serving as cell type specific markers in the human brain from McKenzie *et al.*

Supplementary Dataset 1 - Raw data for day-night gene expression variation. For each human gene, results of the differential gene expression analysis between day and night.

Supplementary Dataset 2 - Raw data for seasonal gene expression variation. For each human gene, results of the differential gene expression analysis for each season against the rest.

Supplementary Figures

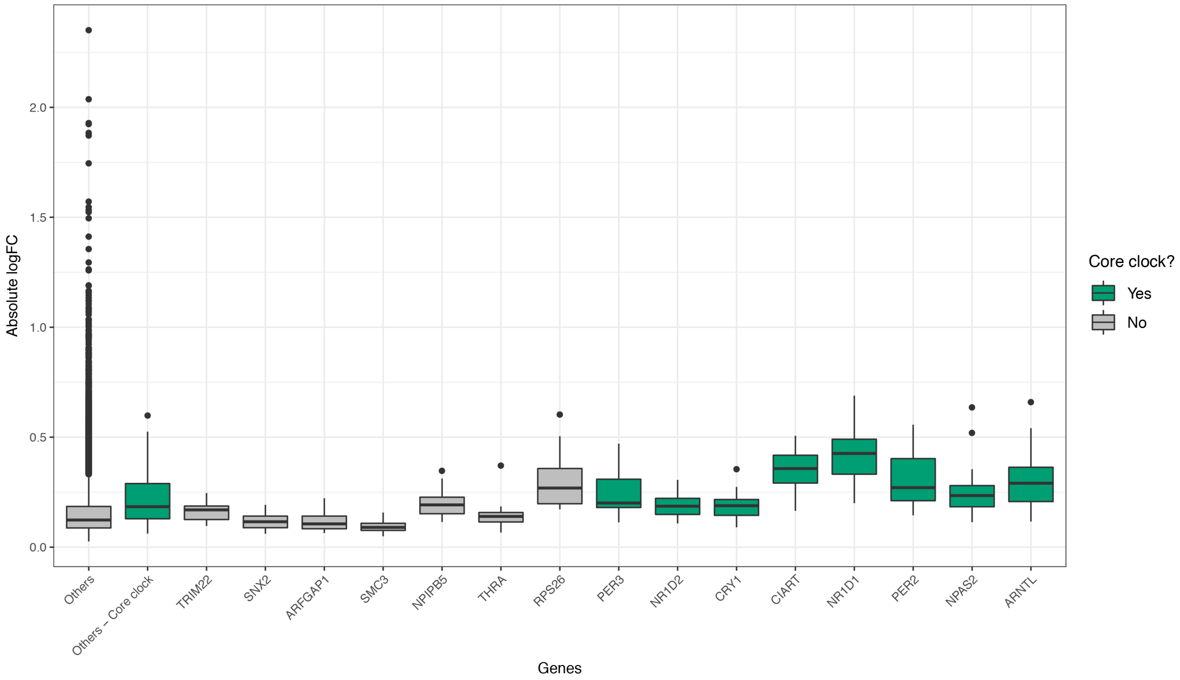

Fig. S1 - Absolute *log2* fold-change (or effect size) of day-night genes. One boxplot per gene is drawn for genes expressed in 15 tissues or more. The rest of the genes are pooled together in 'Others - Core clock' if they are in the list of the core clock genes, in 'Others' otherwise. Boxplot of core clock genes are in green.

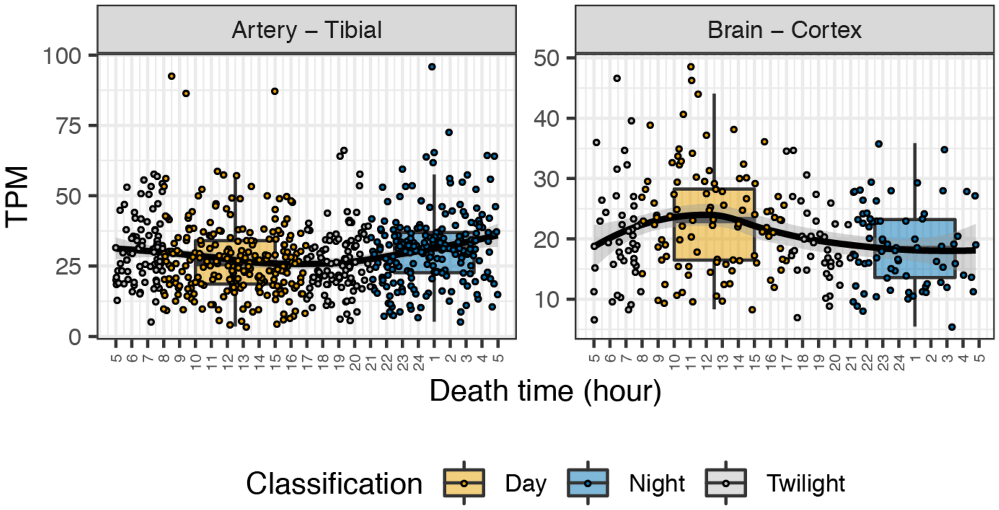

Fig. S2 - Contrasting *NR1D2* circadian behaviors in brain and other tissues. TPM values for *NR1D2* in the artery - tibial (left) and in the brain - cortex (right). The colors of the dots represent the classification of the individuals according to the time of death of the donor: during the day (yellow), during the night (blue), or in-between for twilight (grey). The samples classified as twilight have been discarded for the day-night analysis. The "circadian" curve was created using the *geom_smooth* function from *ggplot2* in R with the 'loess' method.

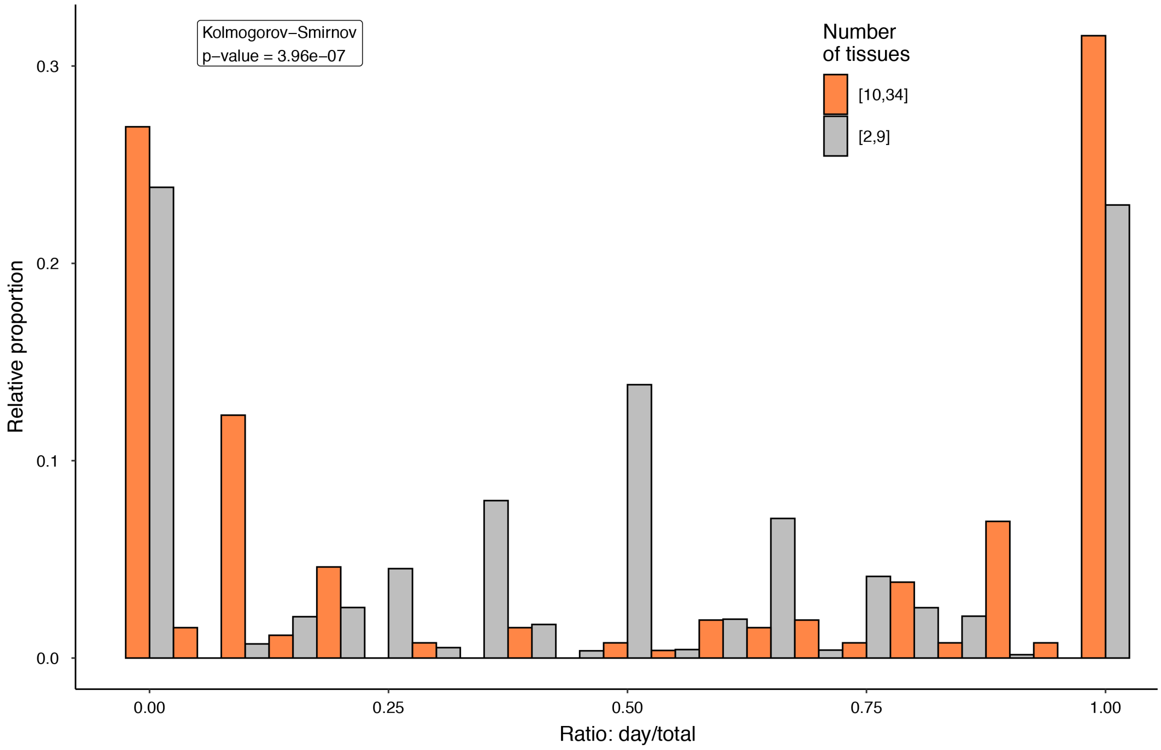

Fig. S3 - Day/night ratio consistency across tissues. Histogram of the ratio, per gene, of the number of tissues in which the gene is classified as day over the total number of tissues for which the gene is day-night. The closer the ratio was to 1, the more 'day' tissues for a particular gene and vice versa for 0 and 'night' tissues. The genes were separated according to the number of tissues in which they are day-night: between two and nine tissues (grey; 11,168 genes) and ten tissues or more (orange; 260 genes).}

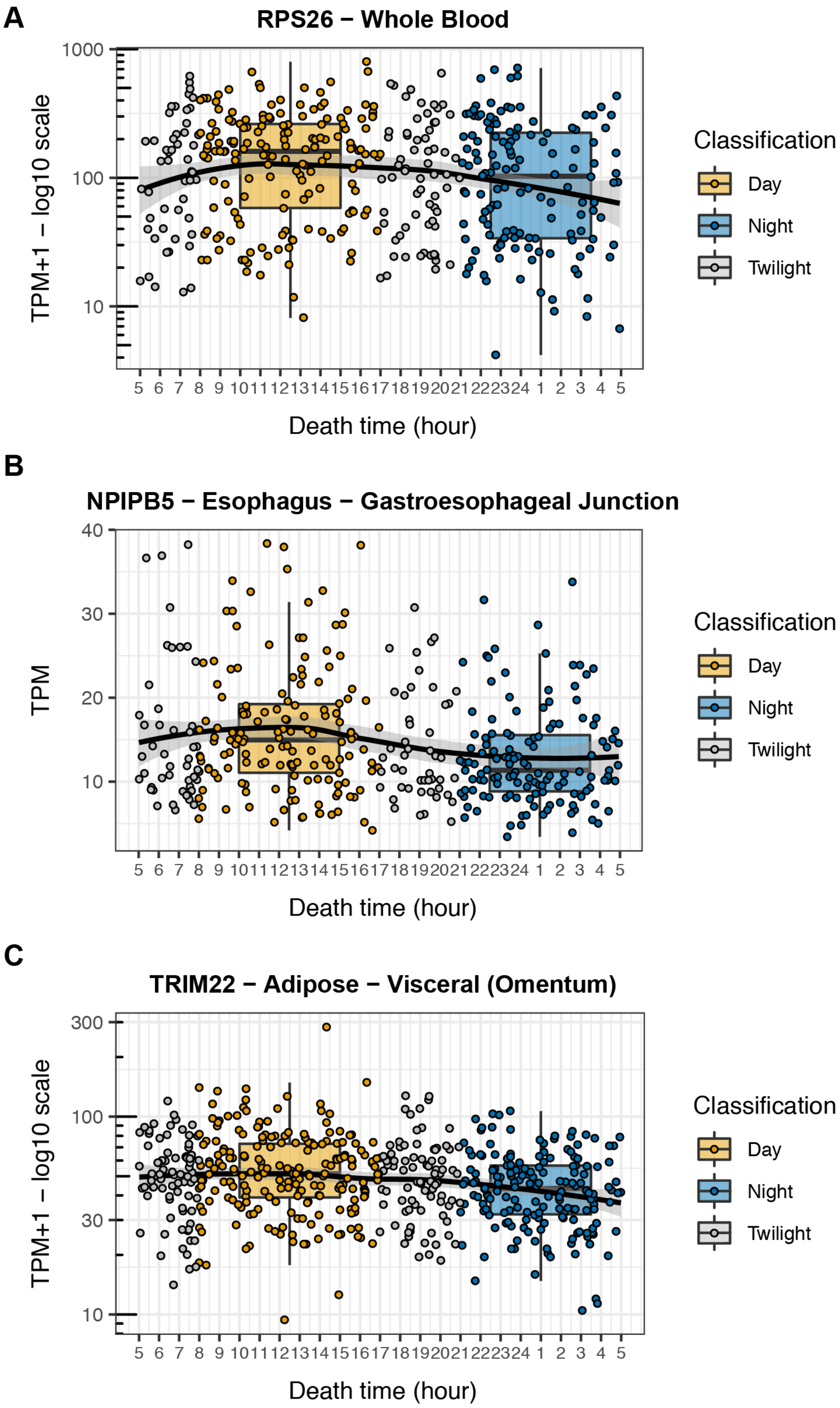

Fig. S4 - Examples of day-night genes. Expression values (TPM) values for three highly recurrent day-night genes at the time of death of the GTEx donors: *RPS26* in the whole blood (A), *NPIPB5* in the esophagus - gastroesophageal function (B), and *TRIM22* in the adipose - visceral (omentum) (C). The colors of the dots represent the classification of the individuals according to their time of death of the donor: during the day (yellow), during the night (blue), or in-between for twilight (grey). The samples classified as twilight have been discarded for the day-night analysis. The "circadian" curve was created using the *geom_smooth* function from *ggplot2* in R with the 'loess' method.

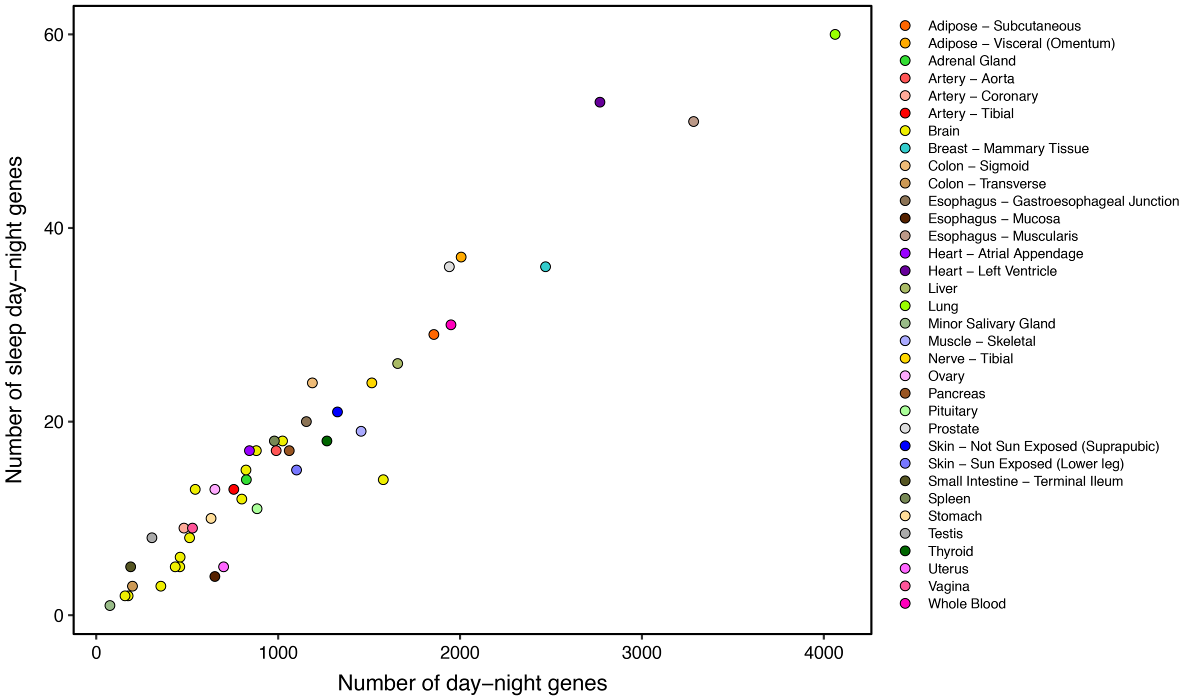

Fig. S5 - Day-night variation of sleep associated genes. Number of sleep day-night genes (y-axis) vs. total day-night genes (x-axis) per tissue.

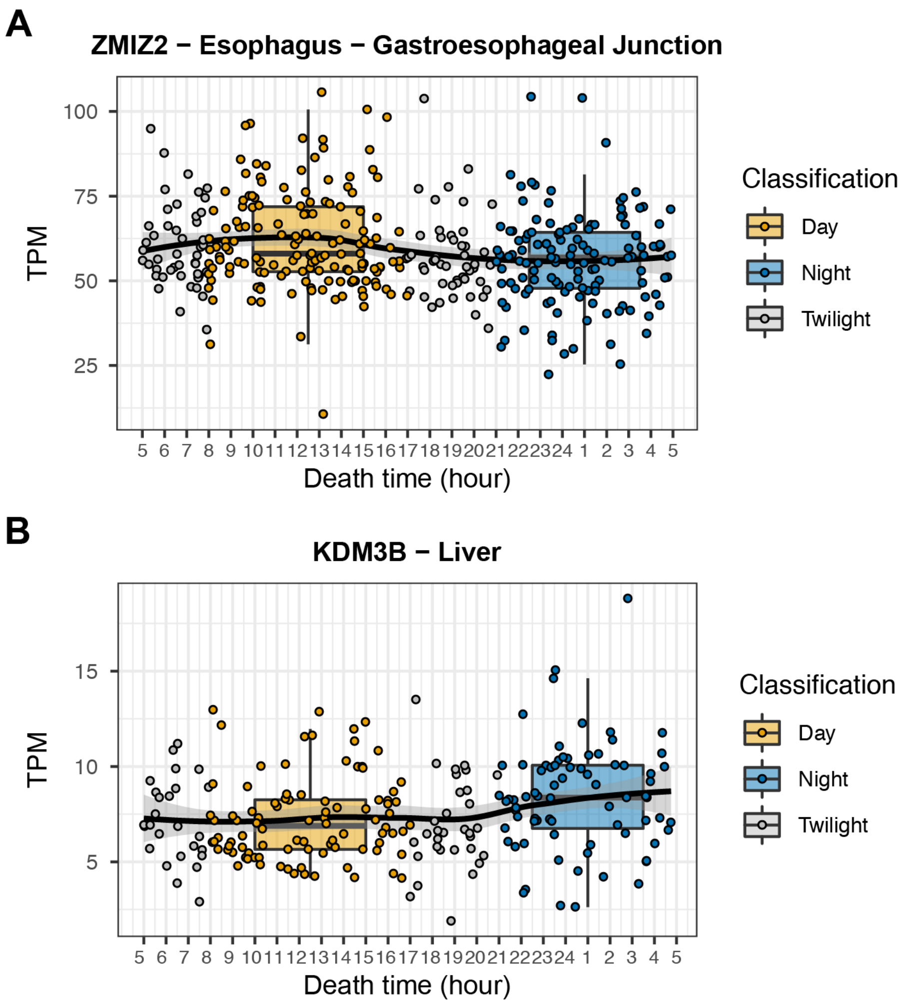

Fig. S6 - Example of day-night sleep genes. Expression values (TPMs) for two highly recurrent day-night sleep genes: *ZMIZ2* in the esophagus - gastroesophageal junction (A) and *KDM3B* in the liver (B) at the time of death of the GTEx donors. The colors of the dots represent the classification of the individuals according to the time of death of the donor: during the day (yellow), during the night (blue), or in-between for twilight (grey). The samples classified as twilight have been discarded for the day-night analysis. The "circadian" curve was created using the *geom_smooth* function from *ggplot2* in R with the 'loess' method.

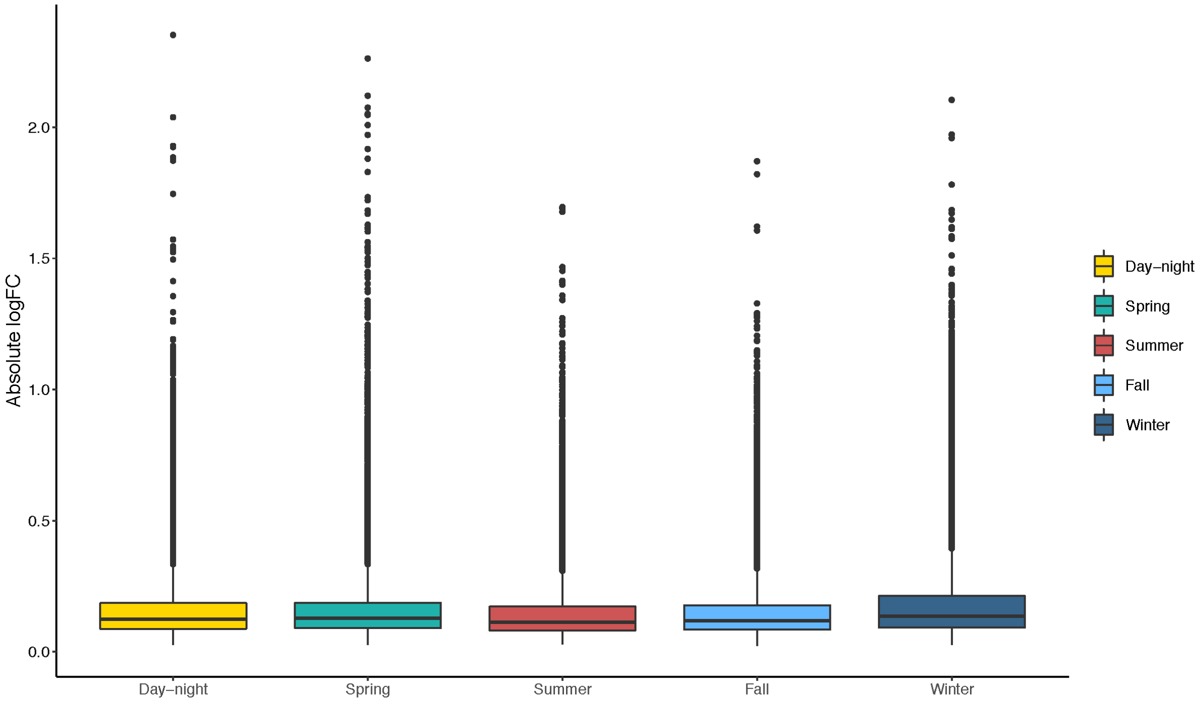

Fig. S7 - Effect size of day-night and seasonal variation in gene expression. Boxplots showing the absolute *log2* fold-change (or effect size) of day-night genes (yellow) or seasonal genes (red: summer, green: spring, light blue: fall, dark blue: winter).

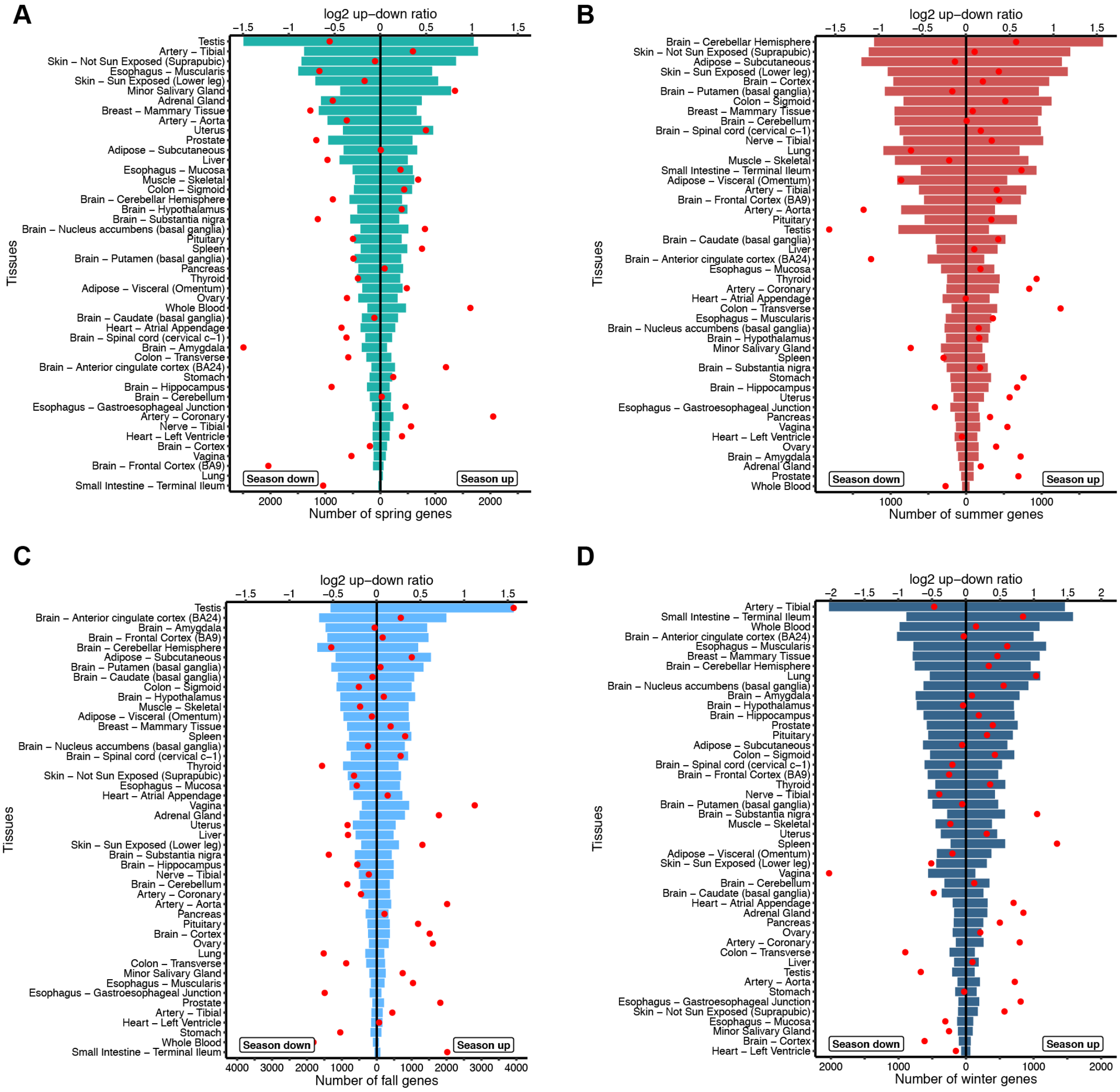

Fig. S8 - Landscape of seasonal gene expression variation by tissue. Number of over-expressed (right side) and under-expressed (left side) seasonal genes per tissue in GTEx (bottom axis) for each season, spring (A), summer (B), fall (C), and winter (D). The tissues were ordered according to the total number of seasonal genes per season. The red dot represents the *log2* ratio between the number of over and under genes (top axis).

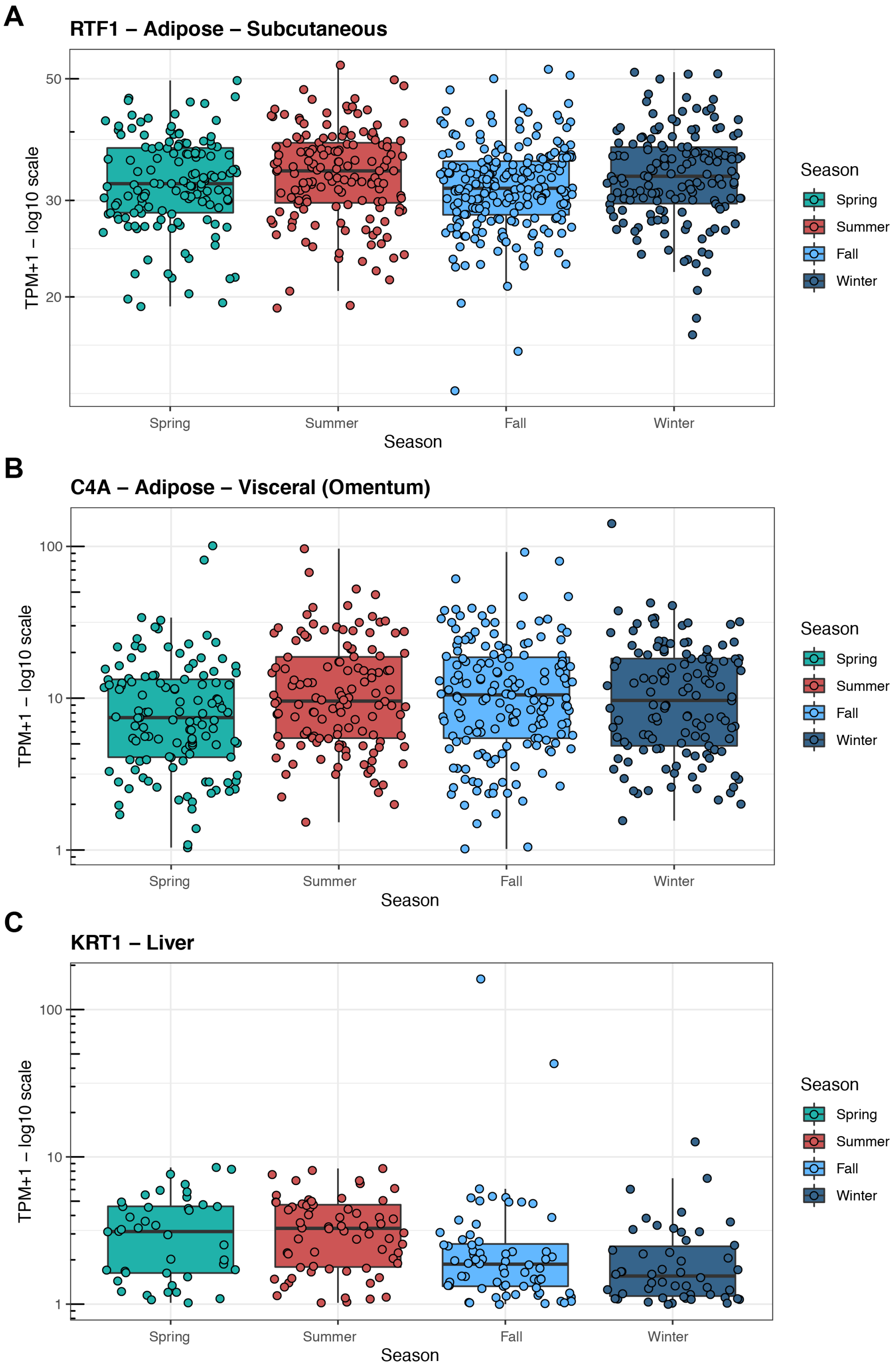

Fig. S9 - Examples of seasonal genes. Boxplot of the TPM values on a *log10* scale of three top seasonal genes in a tissue where they are differentially expressed: *RTF1*, underexpressed in fall (A), *C4A* underexpressed in spring (B), and *KRT1* underexpressed in winter (C).

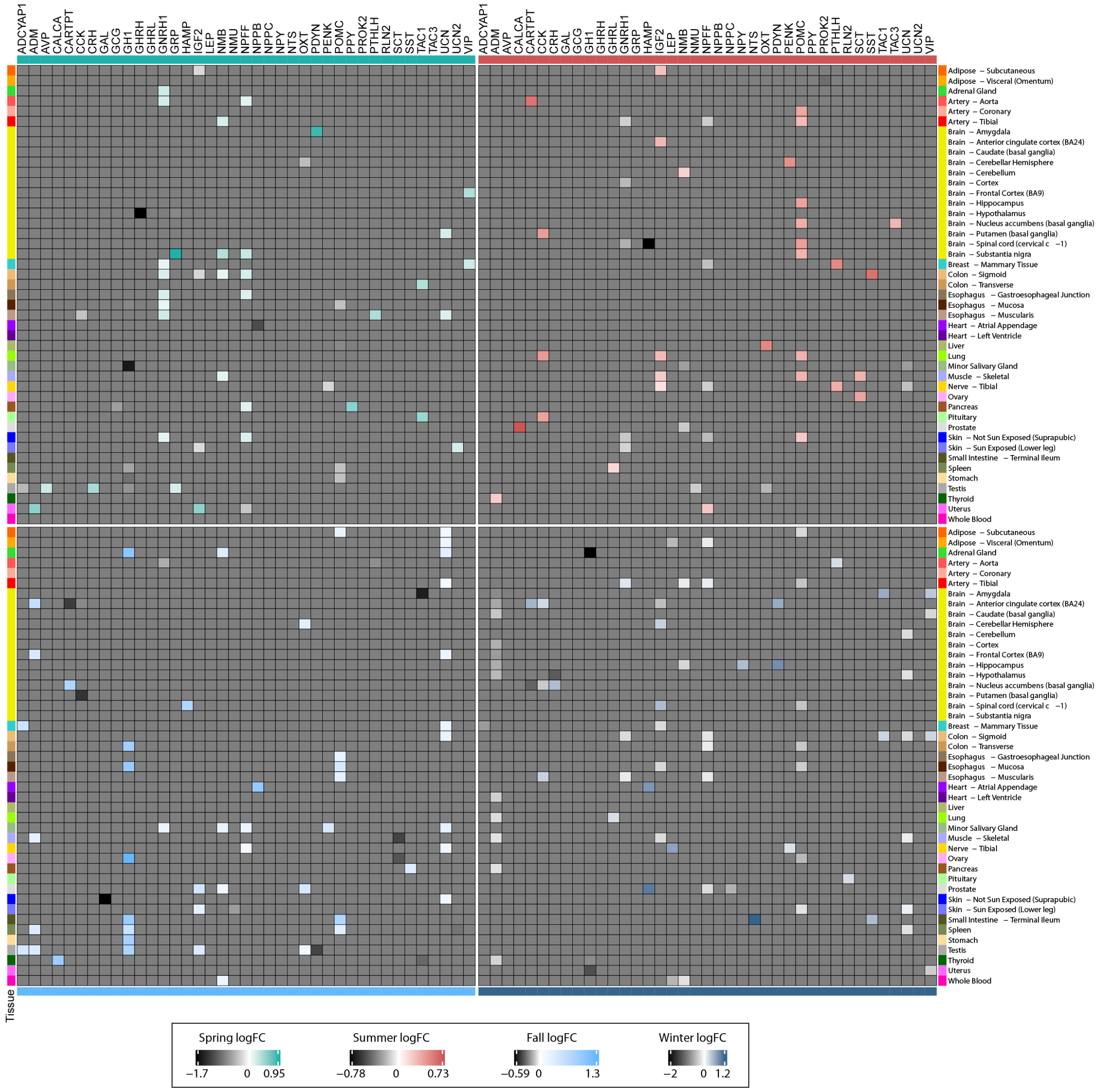

Fig. S10 - Landscape of the seasonal variation of human hormone genes. Seasonal *log2* fold-change for the 39 hormone genes in the 45 tissues (all except vagina) with at least one hormonal gene showing seasonal expression.

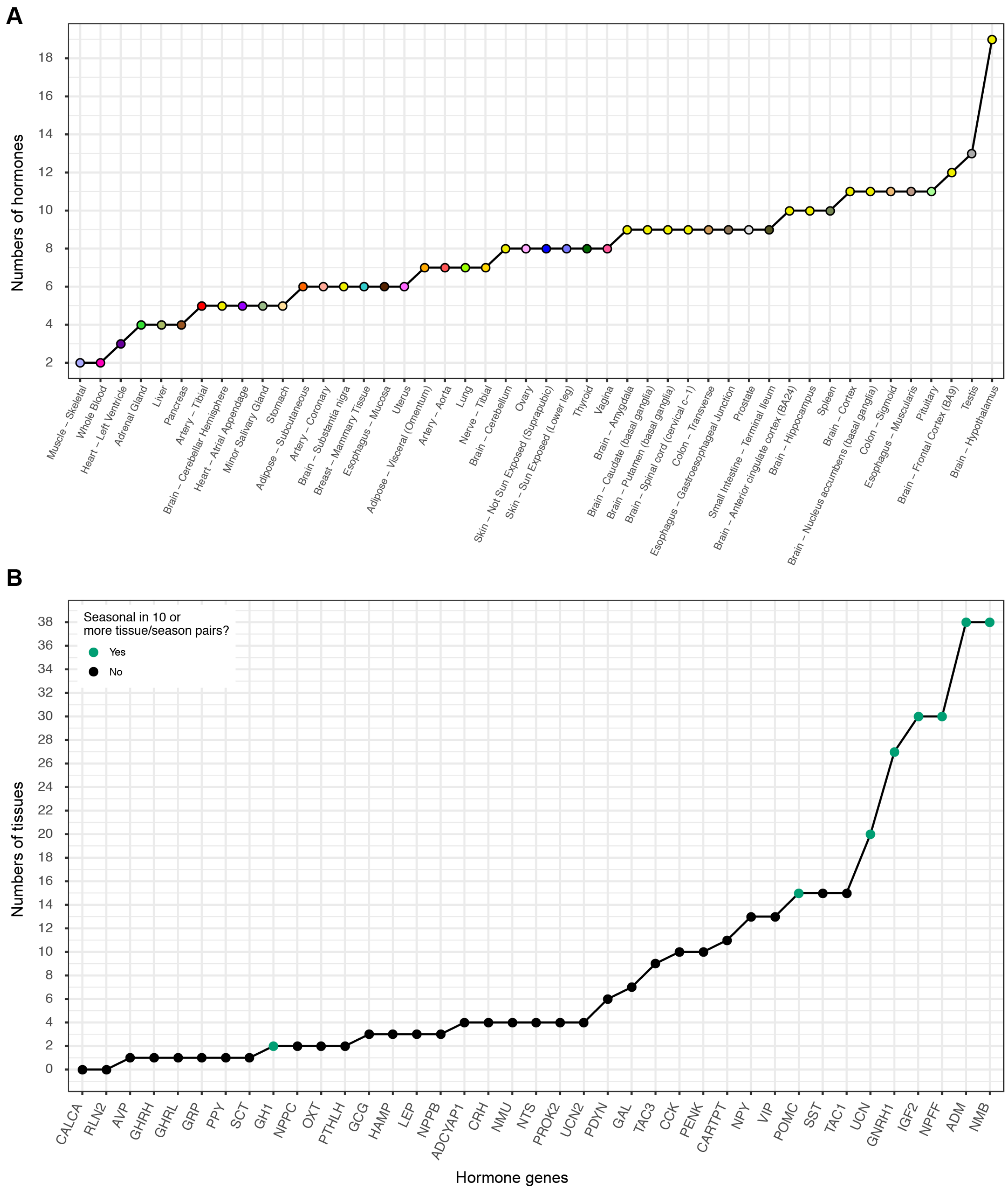

Fig. S11 - Summary of the seasonal variation of human hormone genes. (A) For each tissue, the number of seasonal hormone genes with a median expression greater or equal than 5 TPMs. (B) For each seasonal hormone gene, the number of tissues for which the gene had a median expression greater or equal than 5 TPM.

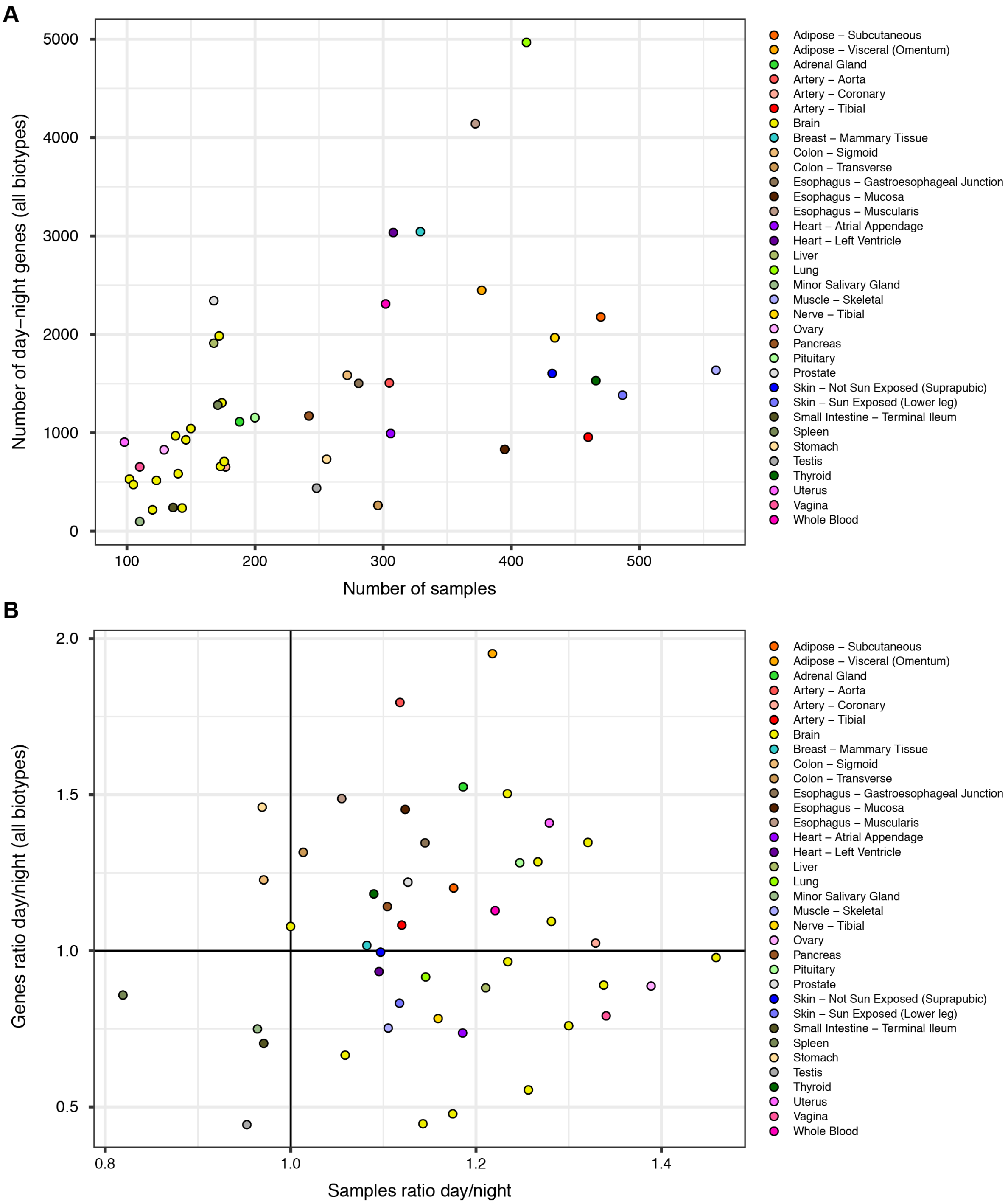

Fig. S12 - Correlations between day-night variation and sample size. (A) Number of day-night genes per number of samples. (B) Ratio of the number of day over number of night genes vs. the ratio of day over night samples.

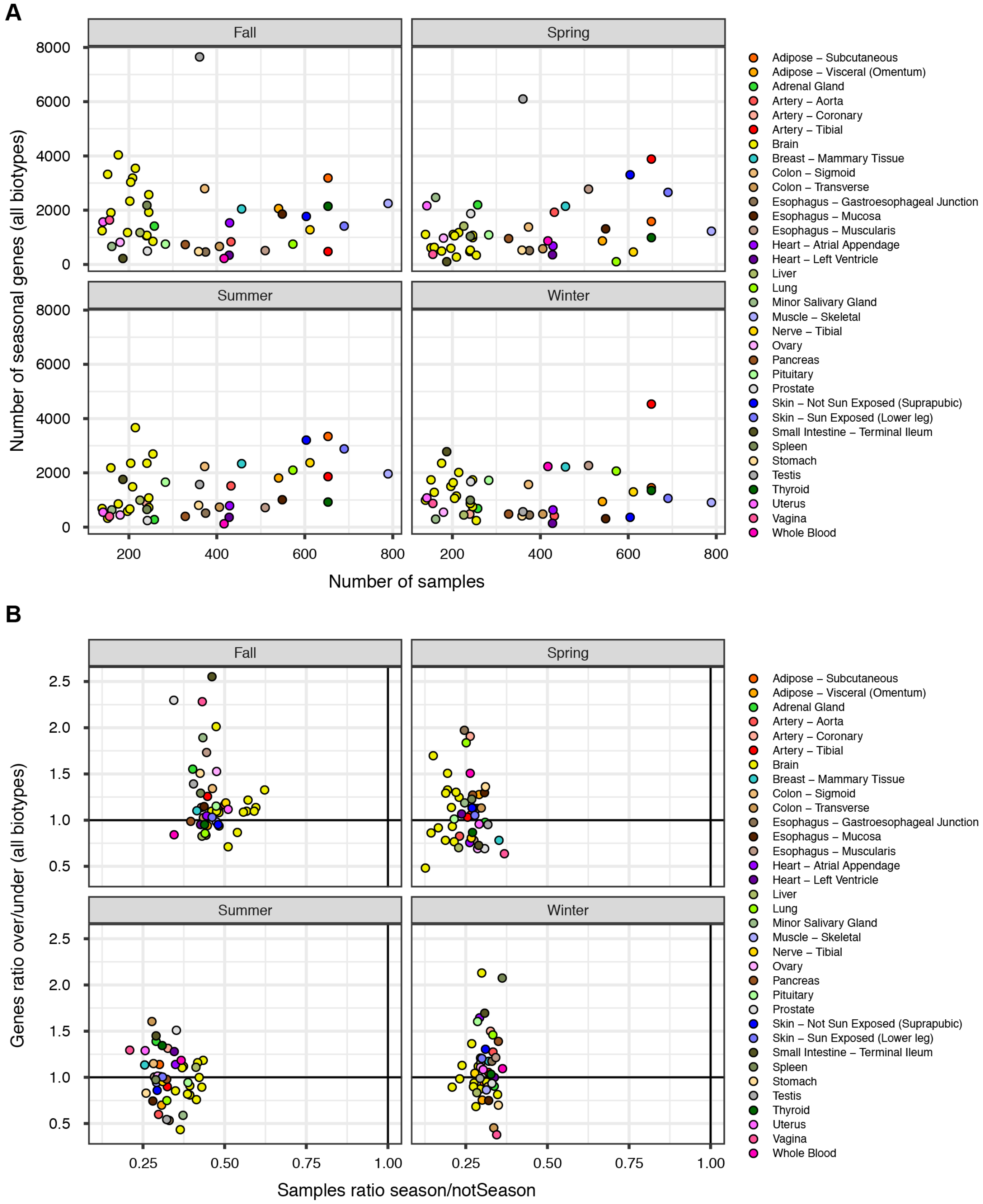

Fig. S13 - Correlations between seasonal variation and sample size. (A) Number of seasonal genes vs number of seasonal samples per season. (B) Ratio of the number of genes found over-expressed over genes found under-expressed per season vs the ratio of the samples from a specific season over all samples.