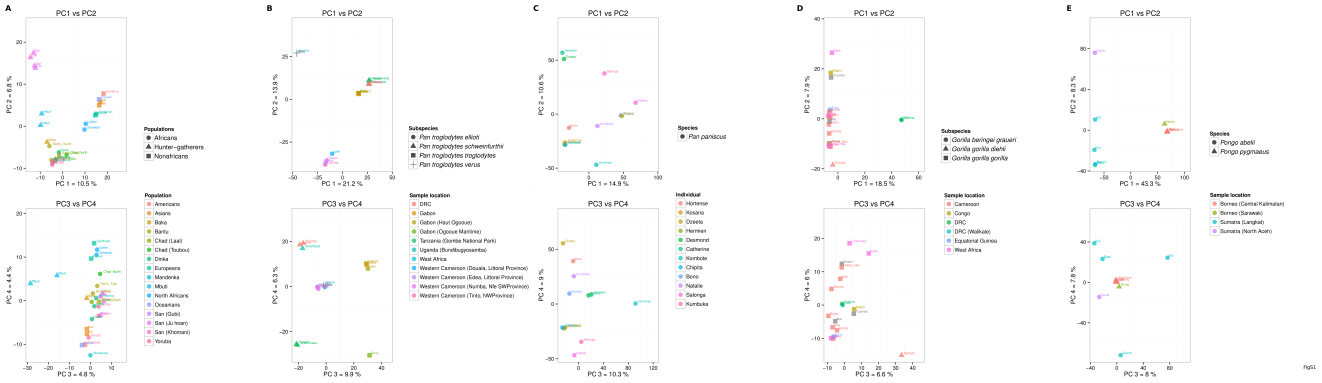
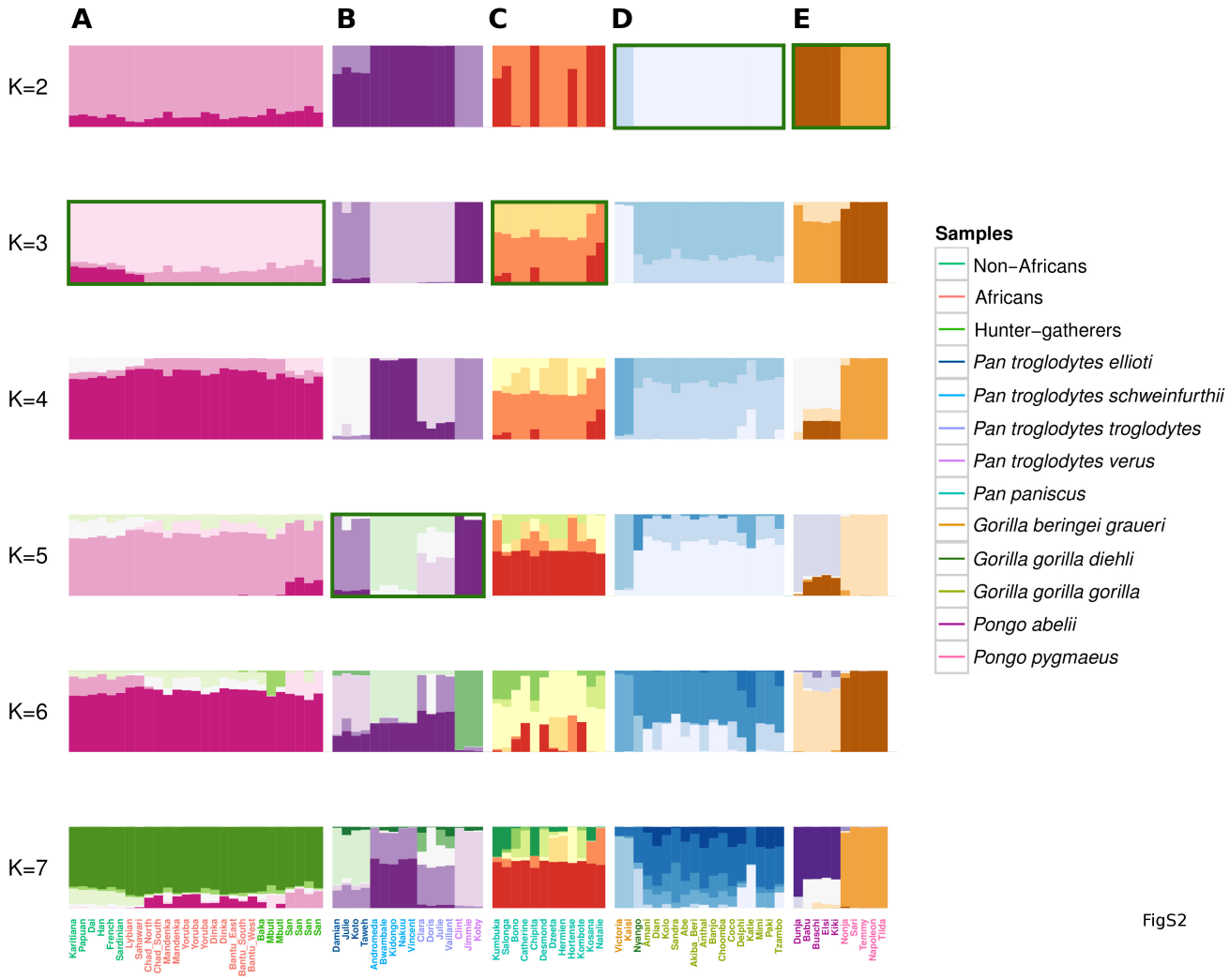
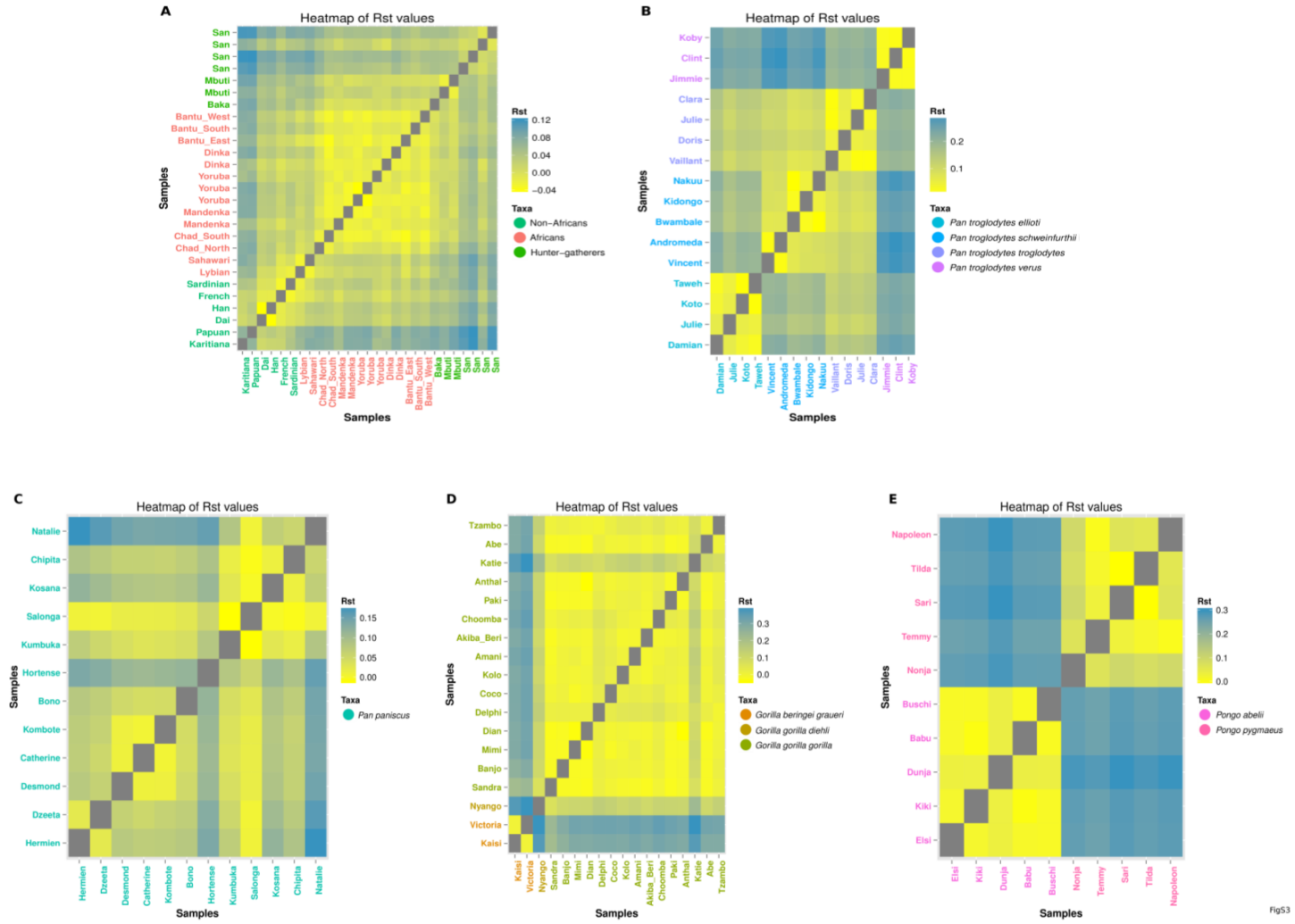


Supplementary Figure S1. Population structure analysis using principal component analysis for **(A)** humans, **(B)** chimpanzees, **(C)** bonobos, **(D)** gorillas, and **(E)** orangutans. Each axis corresponds to a principal component (PC), and shows the percentage of the total variation it explains. Each of the top plots show the two first PCs, while the plots below show both the third and the fourth PCs. Shapes correspond to different groups in humans, subspecies in gorillas and chimpanzees, and species in orangutans. Colours correspond to populations for humans, and to the sample's geographical origin for nonhuman primates (DRC stands for Democratic Republic of the Congo), except for bonobos for which they represent the different individuals since there is no information for sample origin. The principal components are able to identify structure at subspecies and species level in great apes, but also at population level and differentiate samples from different geographical locations

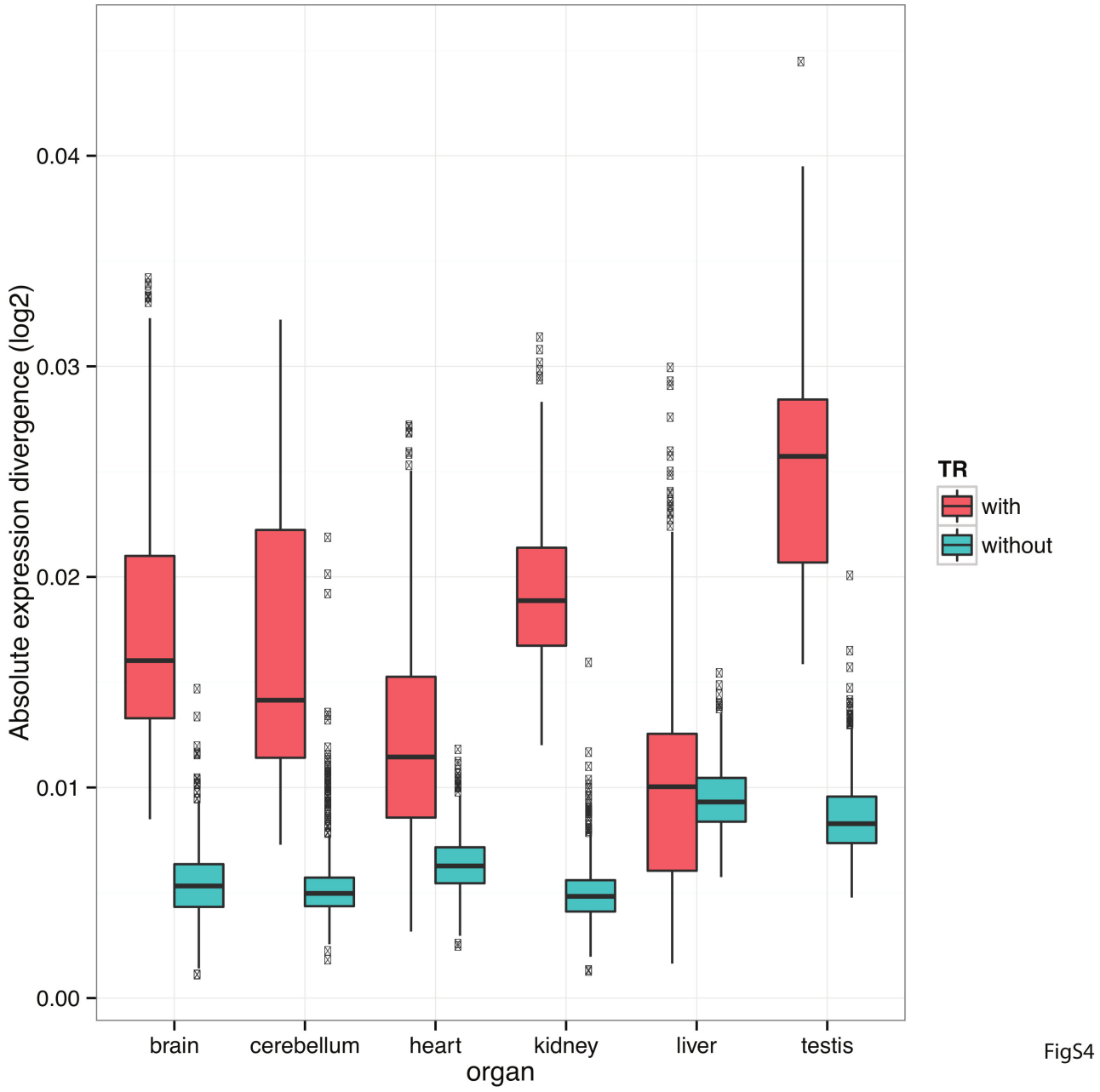


FigS2

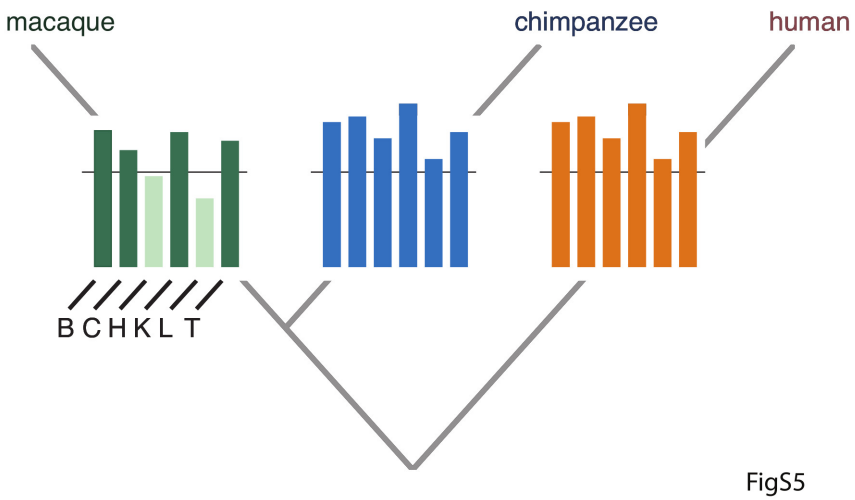
Supplementary Figure S2. Population structure analysis using STRUCTURE for **(A)** humans, **(B)** chimpanzees, **(C)** bonobos, **(D)** gorillas, and **(E)** orangutans. Each row represents a different K number of clusters into which the samples were partitioned. Within each plot, each vertical bar represents a different sample. For each taxon, barplots corresponding to the optimal K number of clusters, determined using Evanno's method, are surrounded by a green rectangle. Clear patterns can be observed corresponding to the sample's different population, subspecies or species designation.



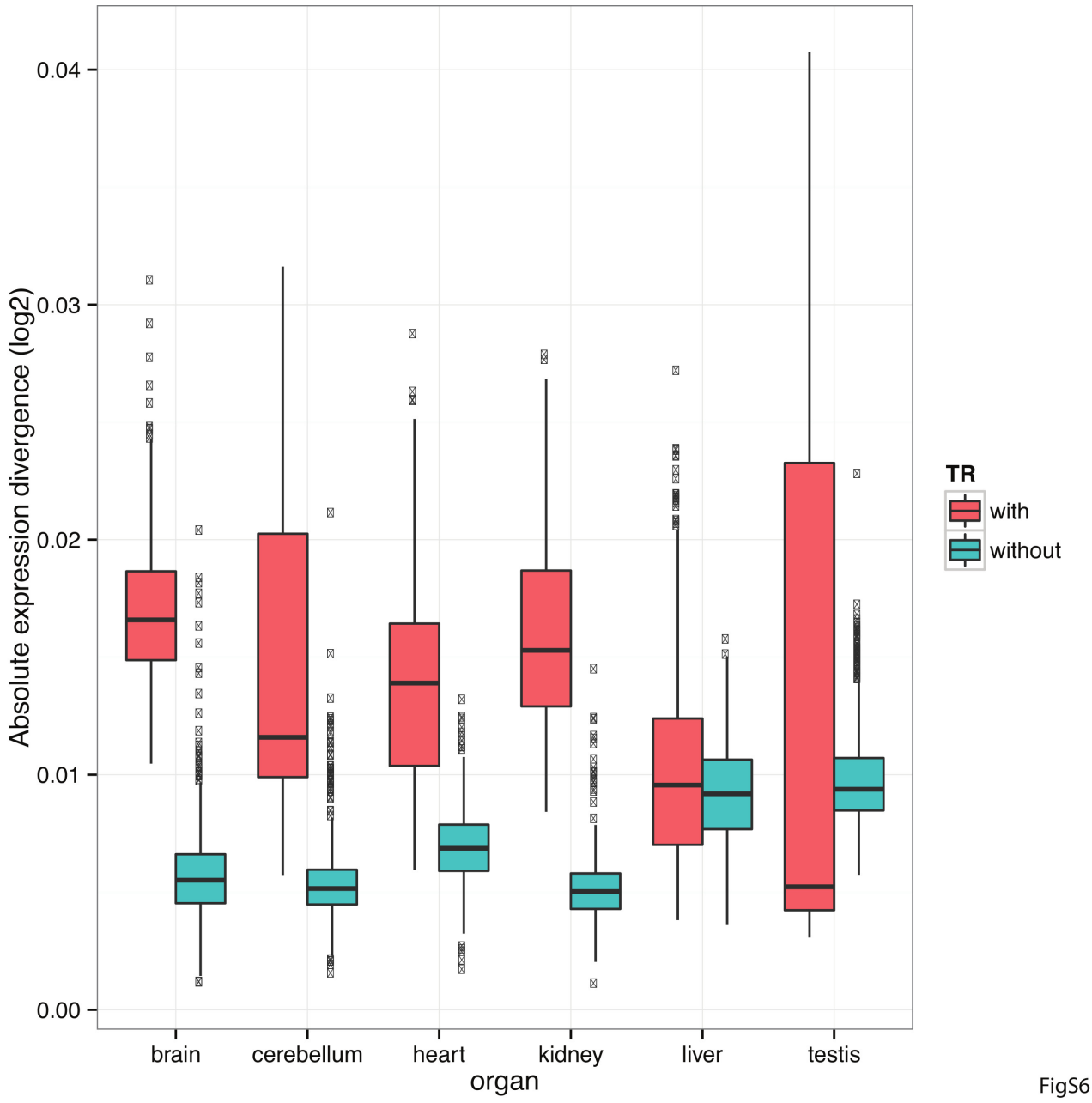
Supplementary Figure S3. Heatmap plots with Rst values for all great ape taxa. Heatmap plots of Rst for human and nonhuman primate taxa, produced by comparing samples in a pairwise fashion. Yellow indicates a low Rst value, and corresponds to high similarity between a pair of samples, while blue indicates high Rst and dissimilar samples. Clusters of higher similarity between samples corresponding to populations, subspecies, and species can be observed.



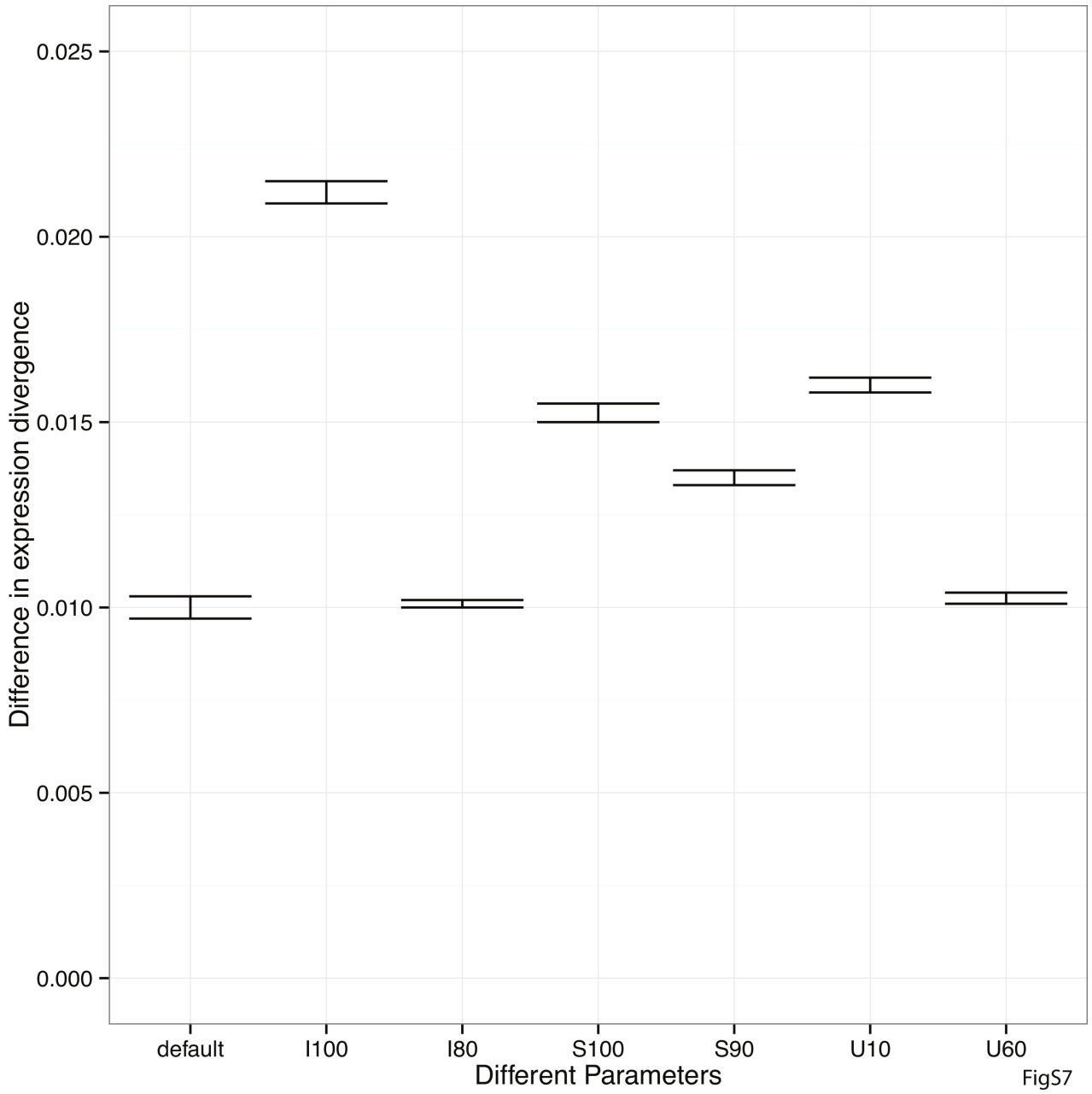
Supplementary Figure S4. Expression divergence analysis based on promoter repeats identified in the newest human reference genome. Boxplot of total tree lengths of genes with repeats (red boxes) and genes without repeats (green boxes). Horizontal lines in the middle of each box mark the median, edges of boxes correspond to the 25th and 75th percentiles, and whiskers cover 99.3% of the data points. Pair of different colored boxes corresponds to the tree lengths of gene expression trees for their below specified organ.



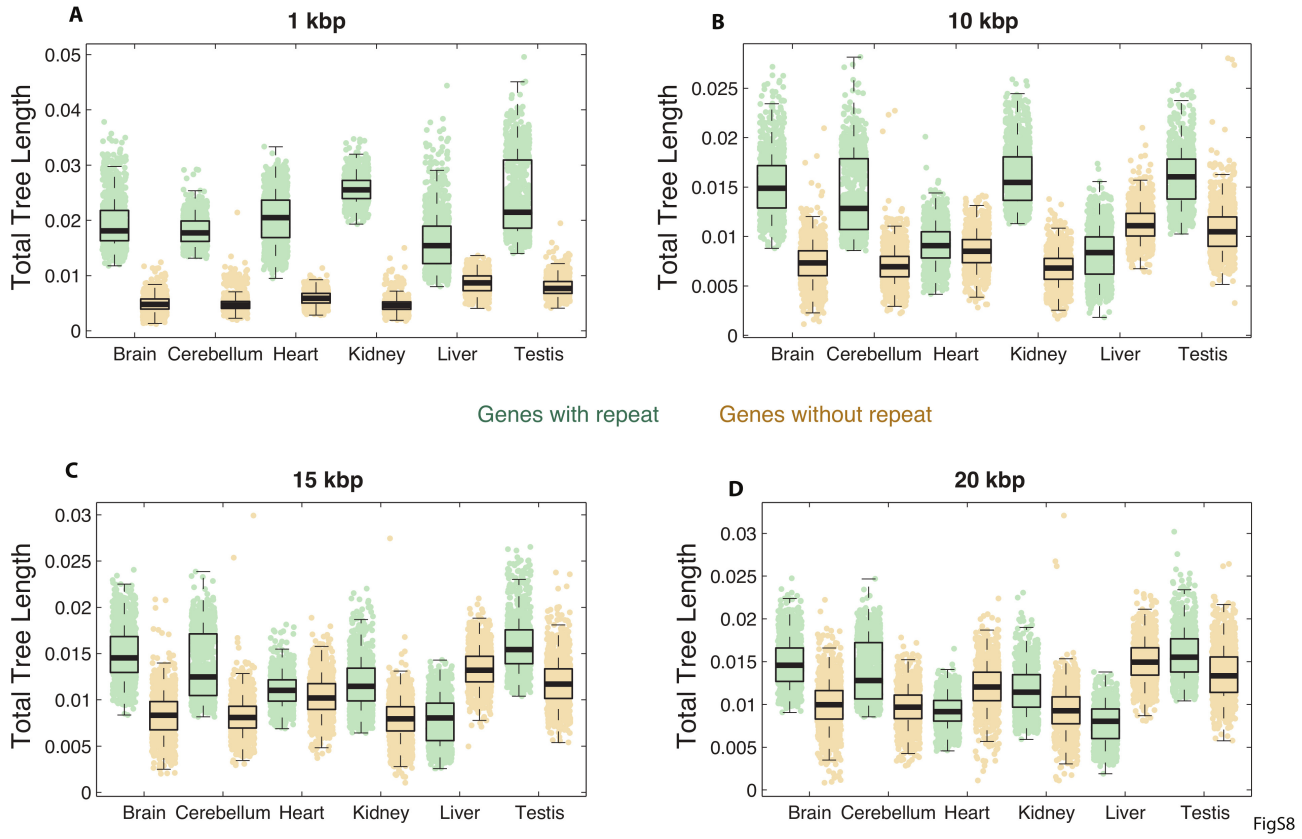
Supplementary Figure S5. Presence of tandem repeats in promoters were associated with increased expression divergence. Schematic phylogenetic tree of our three study species with a bar-chart superimposed on each branch. The length of each bar indicates the average ratio of branch lengths in 1000 sampled gene expression trees for genes with repeats relative to genes without repeats. Bars in each chart, from left to right, correspond to expression divergence in brain (B), cerebellum (C), heart (H), kidney (K), liver (L), and testis (T). Bars extending above the horizontal line indicate that genes with repeats show greater expression divergence that genes without repeats.



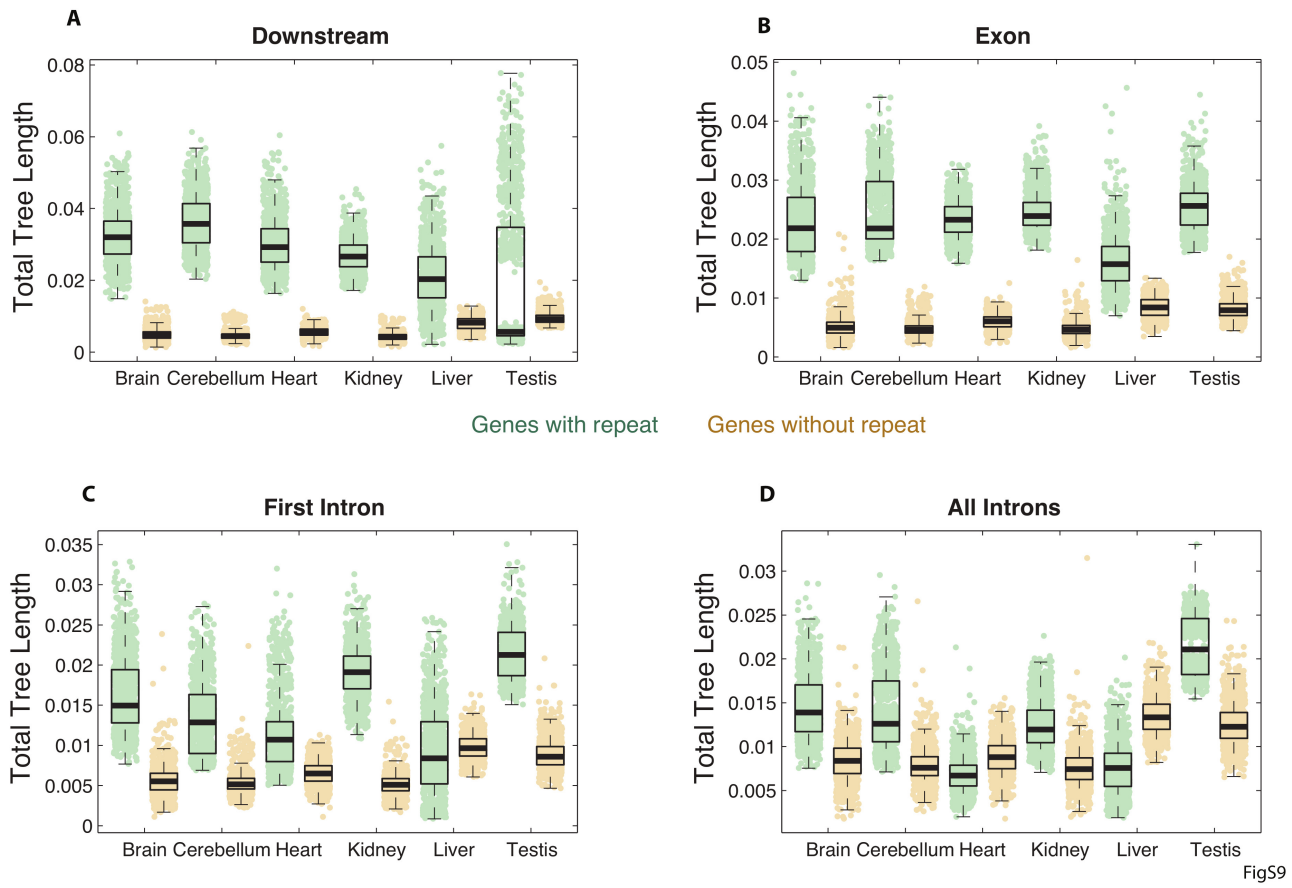
Supplementary Figure S6. Expression divergence analysis based on repeats identified by Phobos. The horizontal axis shows the organs for which we constructed organ-specific gene expression trees. The vertical axis shows the distribution of total gene expression divergence tree lengths for genes with repeats (red boxes) and genes without repeats (green boxes), as identified by Phobos. Horizontal lines in the middle of each box mark the median. The edges of the boxes correspond to the 25th and 75th percentiles. Whiskers cover 99.3% of the data points. Except for the testis, differences in expression divergences in all other organs were significant.



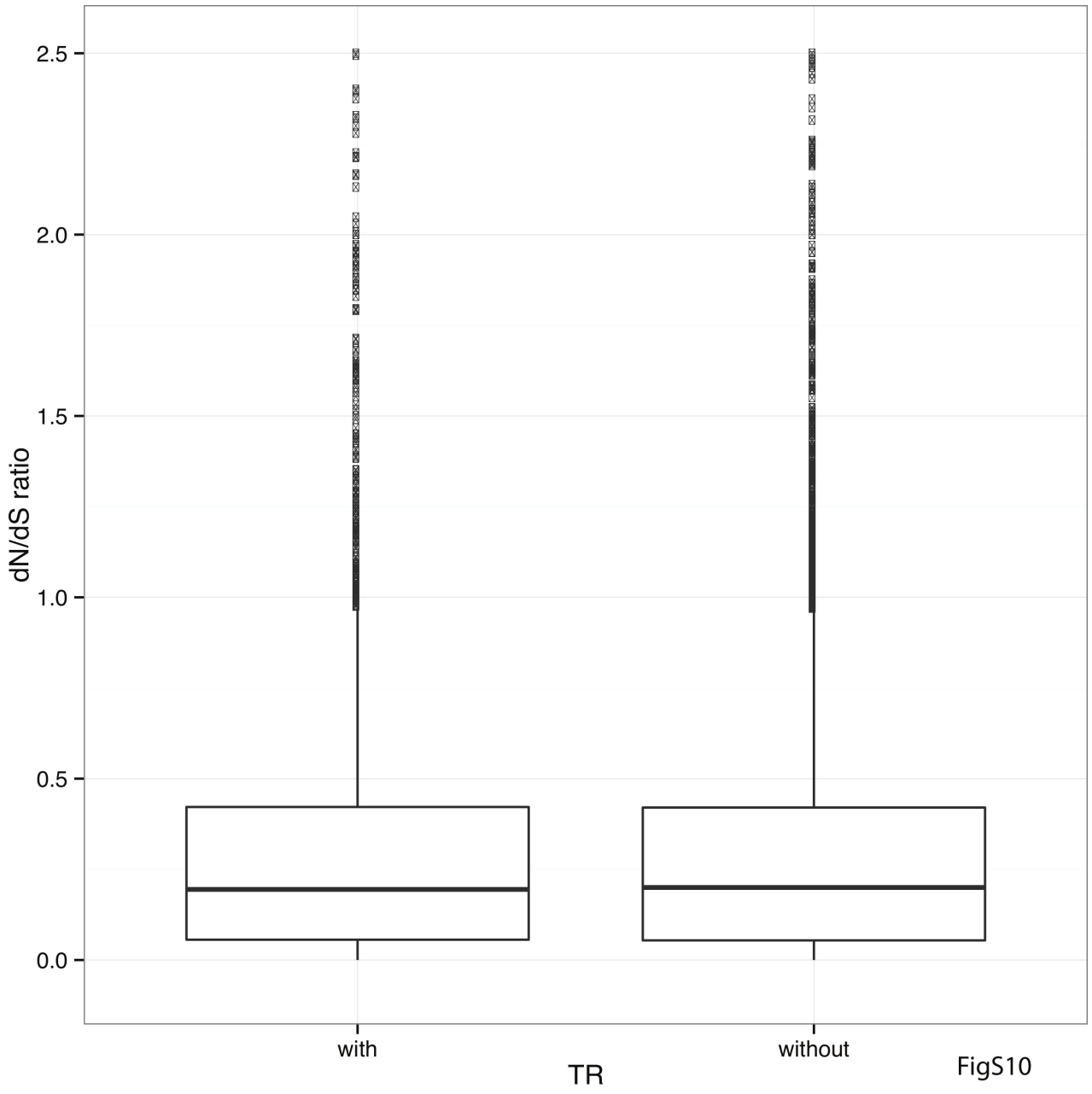
Supplementary Figure S7. Changing the parameters for TR identification does not the change the results. Y-axis shows the mean differences in expression divergence, based on pairwise expression tree length differences between repeat-containing and nonrepeat-containing genes. X-axis shows various different parameters used to identify TRs: I80 (sequence identity of 80 or more), I100 (total sequence identity between repeat units), S90 (Tandem Repeeats Finder (TRF) Score of at least 90), S100 (TRF Score of at least 100), U10 (unit size up to 10 nucleotide), U40 (unit size up to 40 nucleotides). Parameters for the default based on the given abbreviations would be I90, S80, U50. Whiskers represent 99.3% confidence intervals.



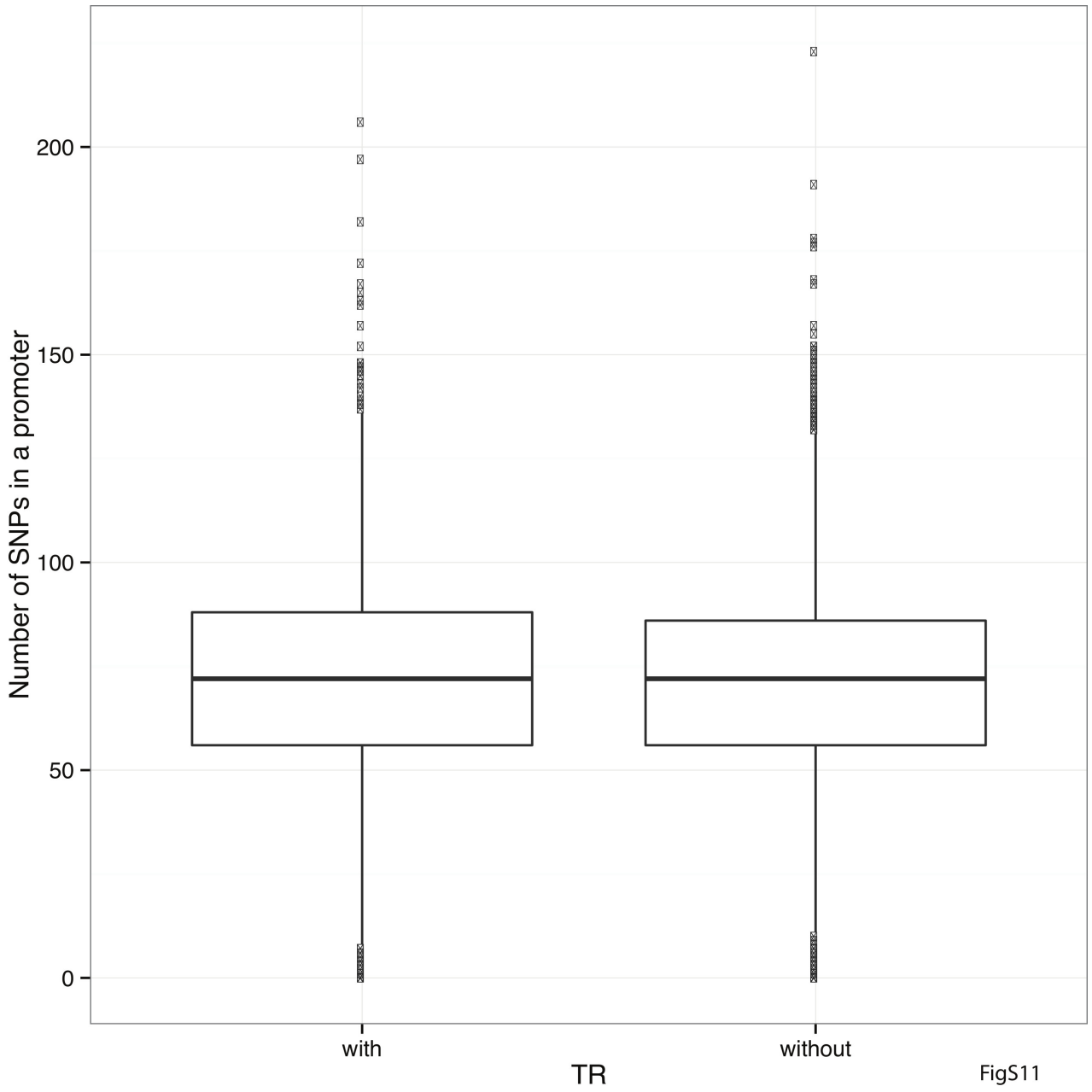
Supplementary Figure S8. Repeats closer to transcription start sites were associated more strongly with expression divergence. The horizontal axes show the organs for which we constructed organ-specific gene expression trees. The vertical axes show the distribution of total gene expression divergence tree lengths for genes with repeats (green) and without repeats (orange), where repeats could occur in upstream regions of length **(A)** 1 kbp **(B)** 10 kbp **(C)** 15 kbp **(D)** 20 kbp. Horizontal lines in the middle of each box mark the median. The edges of the boxes correspond to the 25th and 75th percentiles. Colored dots represent the distribution of tree lengths for each of the 1000 replicates (See methods). Whiskers cover 99.3% of the data points.



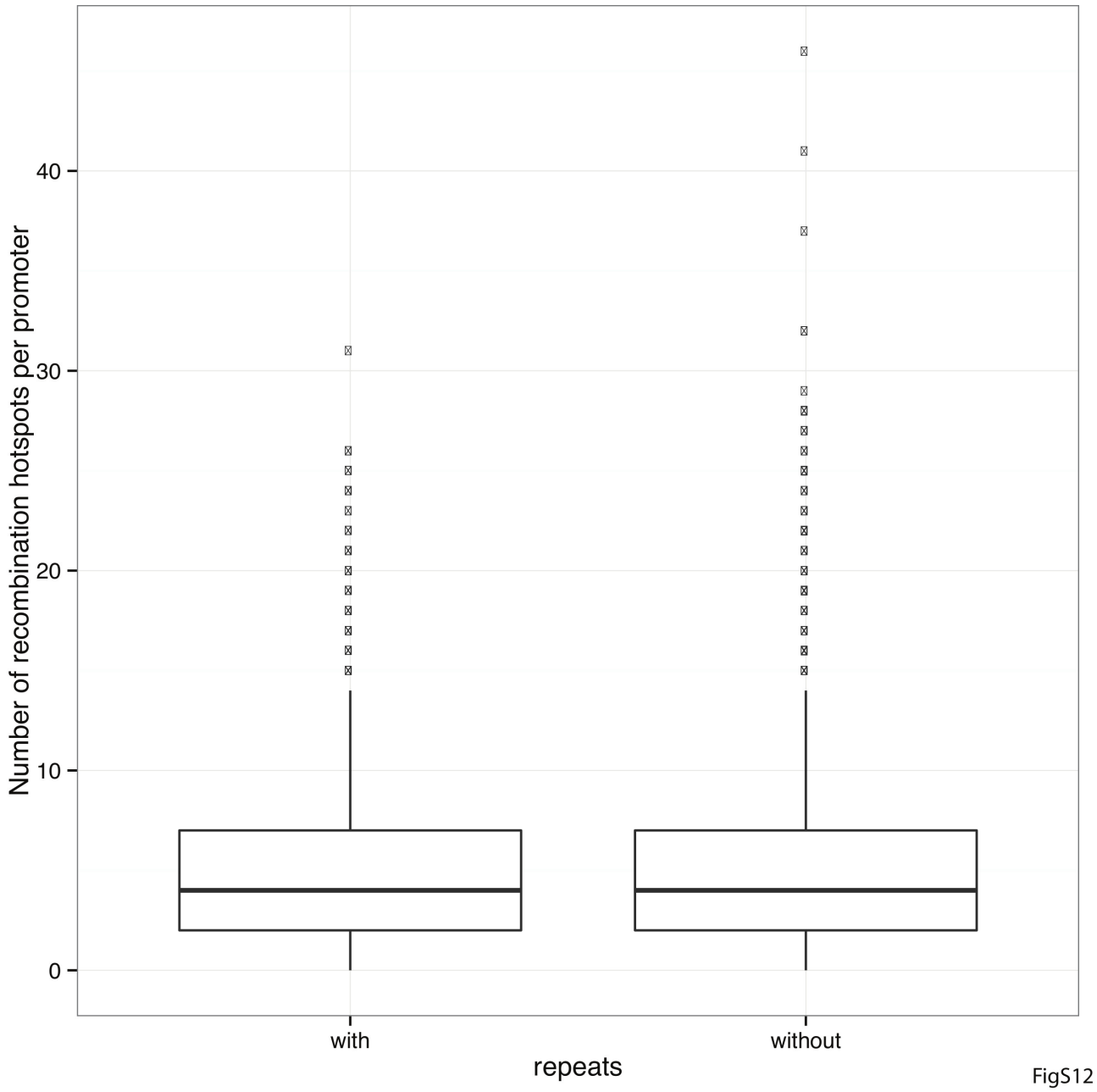
Supplementary Figure S9. Tandem repeats associated with increased expression divergence in different regulatory regions. The horizontal axes show the organs for which we constructed organ-specific gene expression trees. The vertical axes show the distribution of total gene expression divergence tree lengths for genes with repeats (green) and genes without repeats (orange), based on repeats found in **(A)** 3'UTR regions, **(B)** exons, **(C)** the first intron of a gene, **(D)** all introns. Horizontal lines in the middle of each box mark the median. The edges of the boxes correspond to the 25th and 75th percentiles. Colored dots present the distribution of tree lengths for each of 1000 replicates. Whiskers cover 99.3% of the data points.



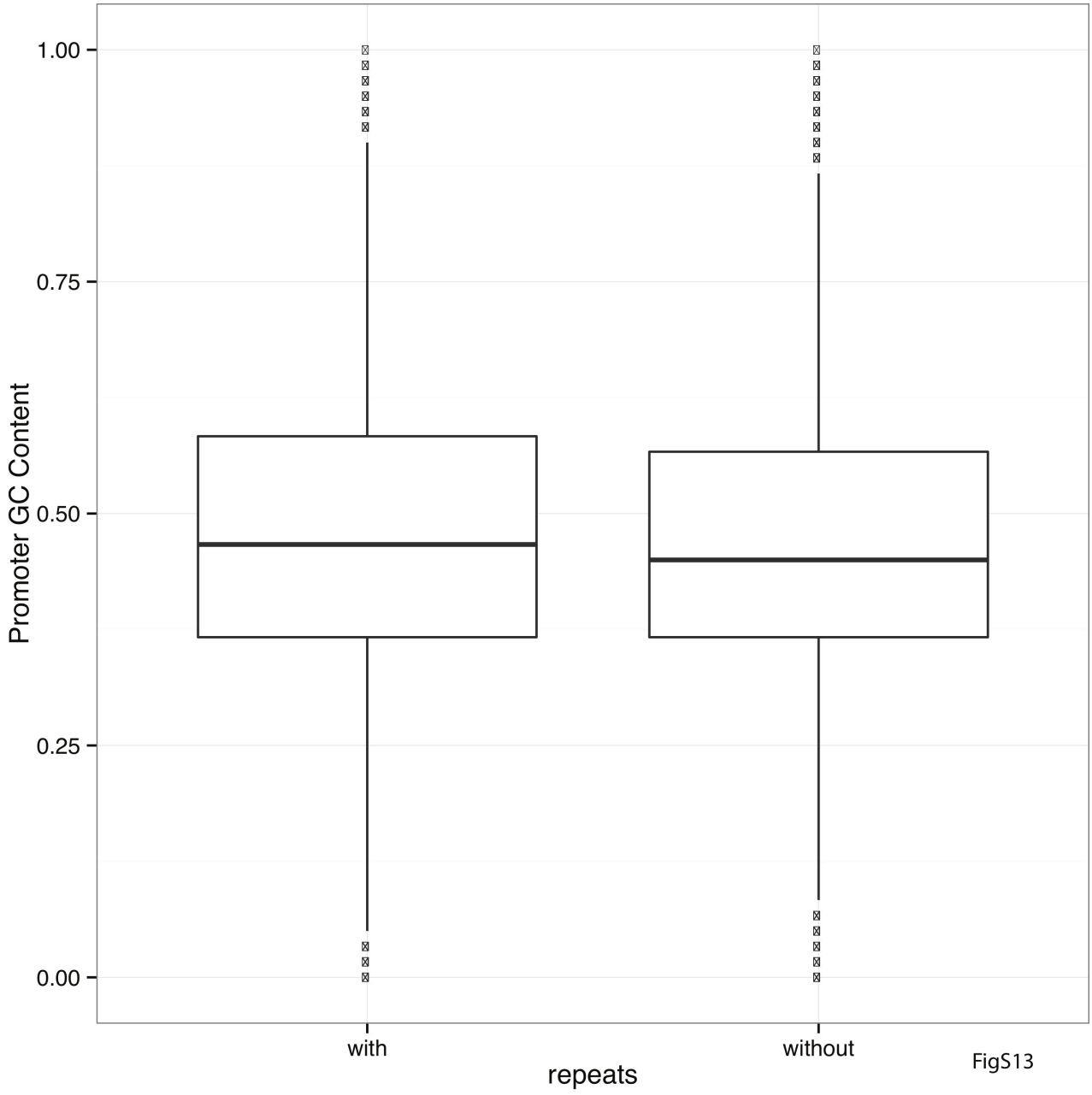
Supplementary Figure S10. No difference in d_N/d_S ratio distributions in promoters with and without large tandem repeats (repeat unit 2-50 bp). Boxplot of d_N/d_S ratio distributions in promoters with repeats and without repeats (thin lines). Horizontal lines in the middle of each box mark the median, edges of boxes correspond to the 25th and 75th percentiles, and whiskers cover 99.3% of the data points.



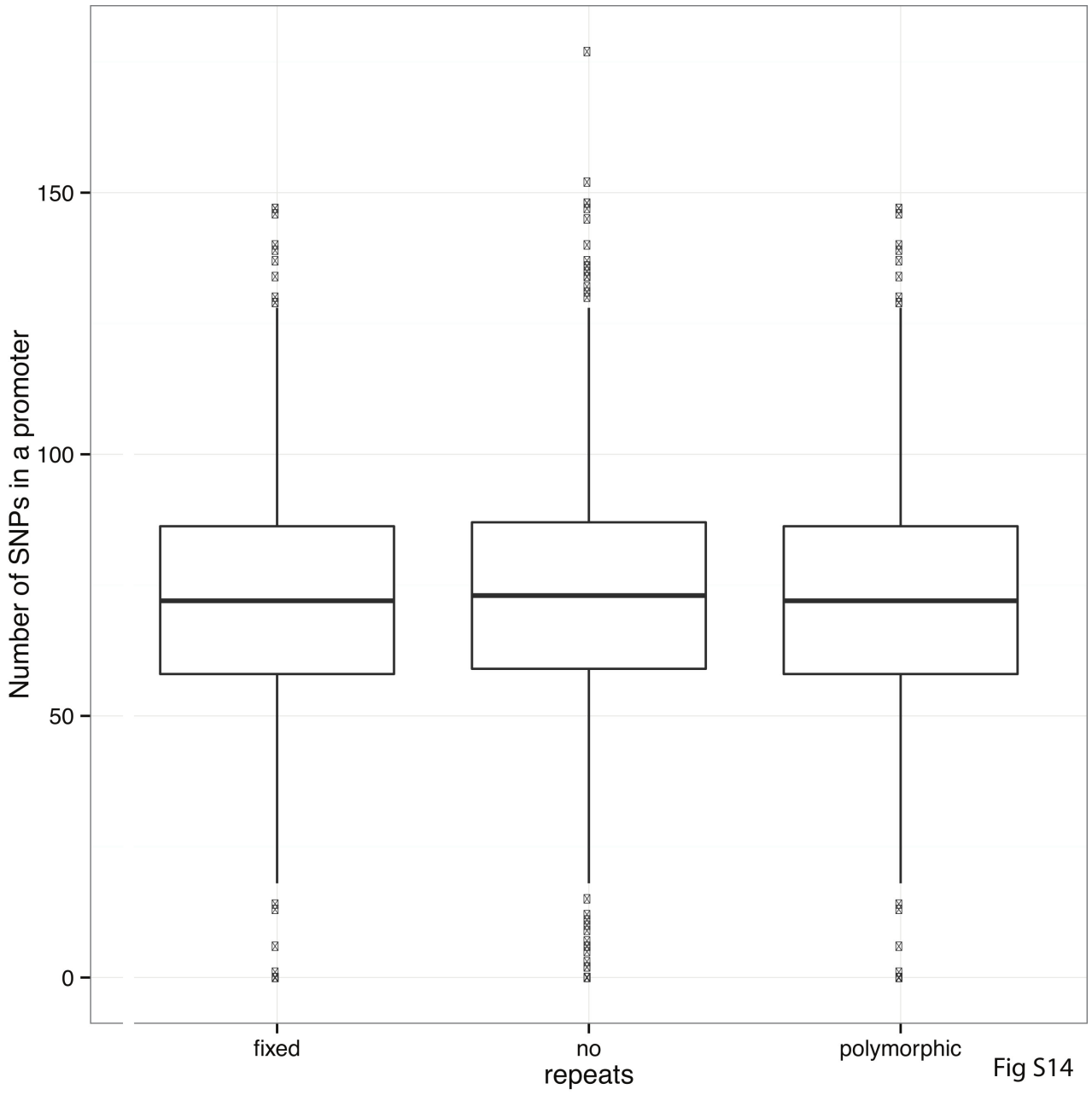
Supplementary Figure S11. No difference in number of SNPs in promoters with and without large tandem repeats (repeat unit 2-50 bp). Boxplot of number of SNPs in promoters with repeats and without repeats (thin lines). Horizontal lines in the middle of each box mark the median, edges of boxes correspond to the 25th and 75th percentiles, and whiskers cover 99.3% of the data points.



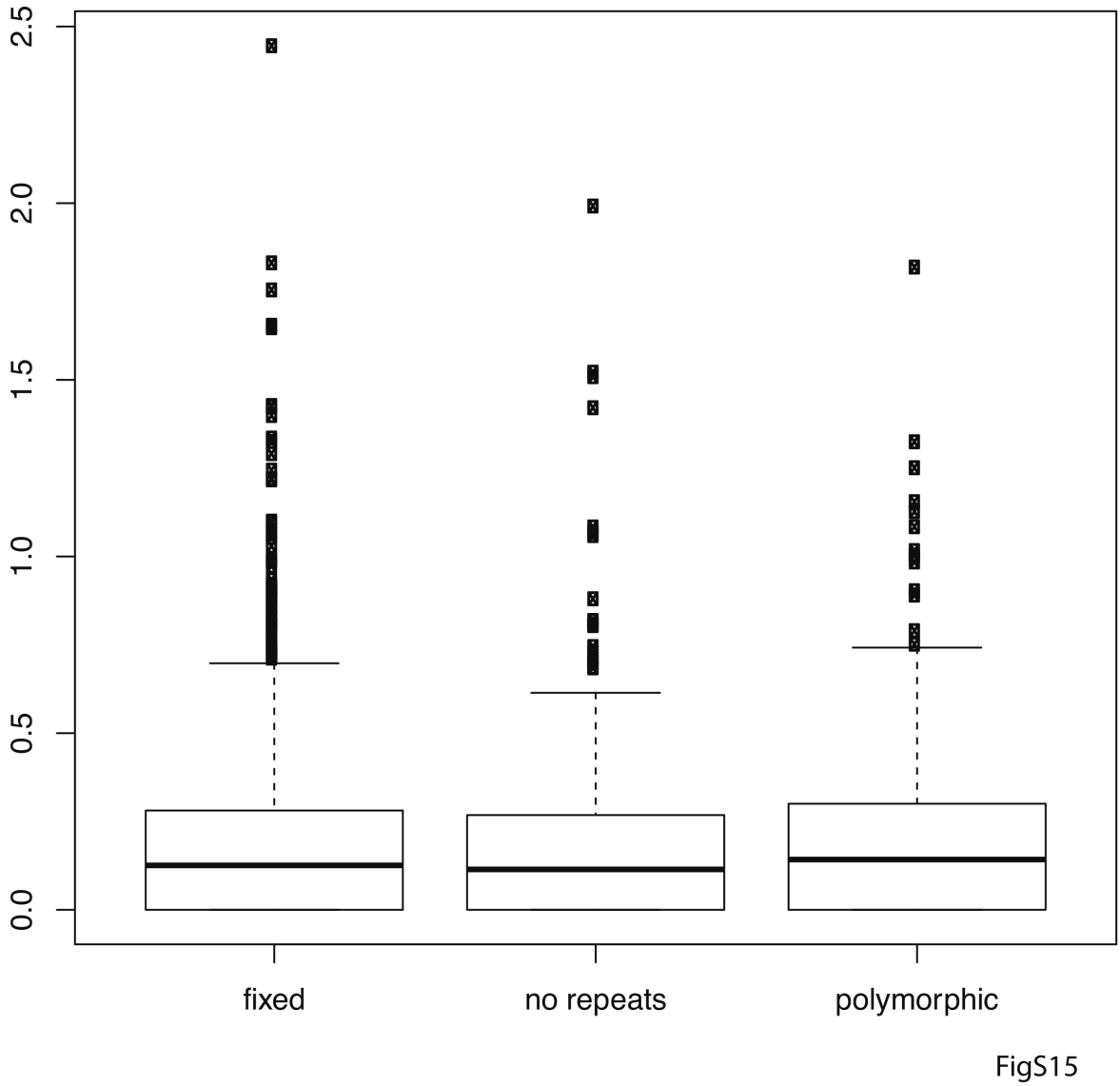
Supplementary Figure S12. No increase in number of recombination hotspots in promoters with large tandem repeats (repeat unit 2-50 bp). Boxplot of number of recombination hotspots in promoters with repeats and without repeats (thin lines). Horizontal lines in the middle of each box mark the median, edges of boxes correspond to the 25th and 75th percentiles, and whiskers cover 99.3% of the data points.



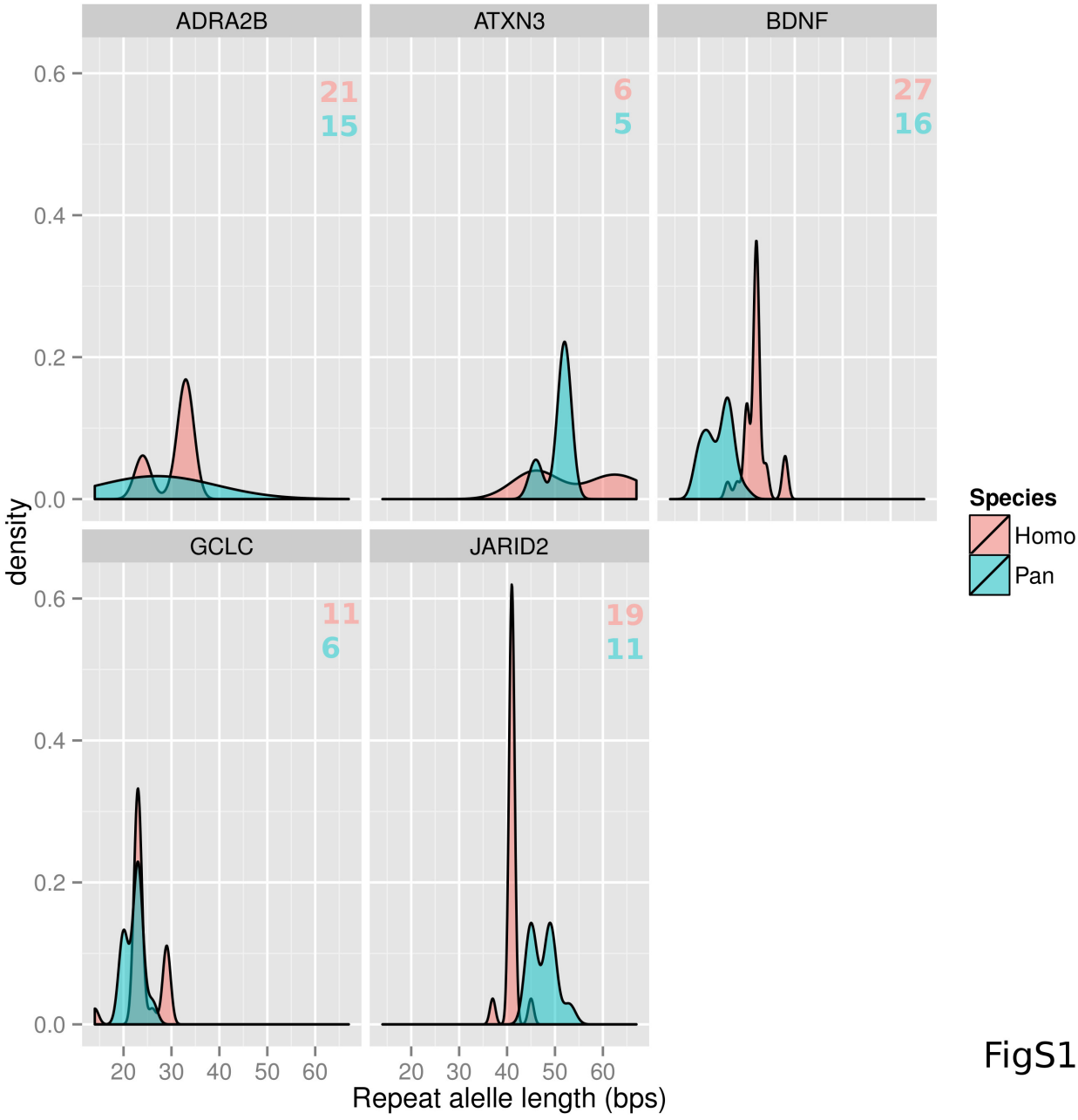
Supplementary Figure S13. Promoter GC Content in genes with and without TRs. Boxplot of GC Content in promoters with repeats and without repeats. Horizontal lines in the middle of each box mark the median, edges of boxes correspond to the 25th and 75th percentiles, and whiskers cover 99.3% of the data points.



Supplementary Figure S14. No difference in number of SNPs in promoters with no/fixed/polymorphic small repeats (repeat unit 1-5 bp and maximum length 100 bp). Boxplot of number of SNPs in promoters with fixed, no, and polymorphic repeats in human and chimpanzee populations (thin lines). Horizontal lines in the middle of each box mark the median, edges of boxes correspond to the 25th and 75th percentiles, and whiskers cover 99.3% of the data points.



Supplementary Figure S15. No difference in d_N/d_S ratio distributions of genes with no/fixed/polymorphic small repeats (repeat unit 1-5 bp and maximum length 100 bp). Boxplot of d_N/d_S ratios of promoters with fixed, no, and polymorphic repeats in human and chimpanzee populations (thin lines). Horizontal lines in the middle of each box mark the median, edges of boxes correspond to the 25th and 75th percentiles, and whiskers cover 99.3% of the data points.



FigS16

Supplementary Figure S16. Repeat allele length distribution for human (pink) and chimpanzee (blue) populations in loci described to contain disease associated repeats. Each locus corresponds to a different locus, where the X-axis corresponds to the length of the repeat alleles in number of repeat copies.