Supplemental information

Evolvability of cancer-associated genes

under APOBEC3A/B selection

Joon-Hyun Song, Liliana M. Dávalos, Thomas MacCarthy, and Mehdi Damaghi

Supplemental information

Evolvability of cancer-associated genes under APOBEC3A/B selection

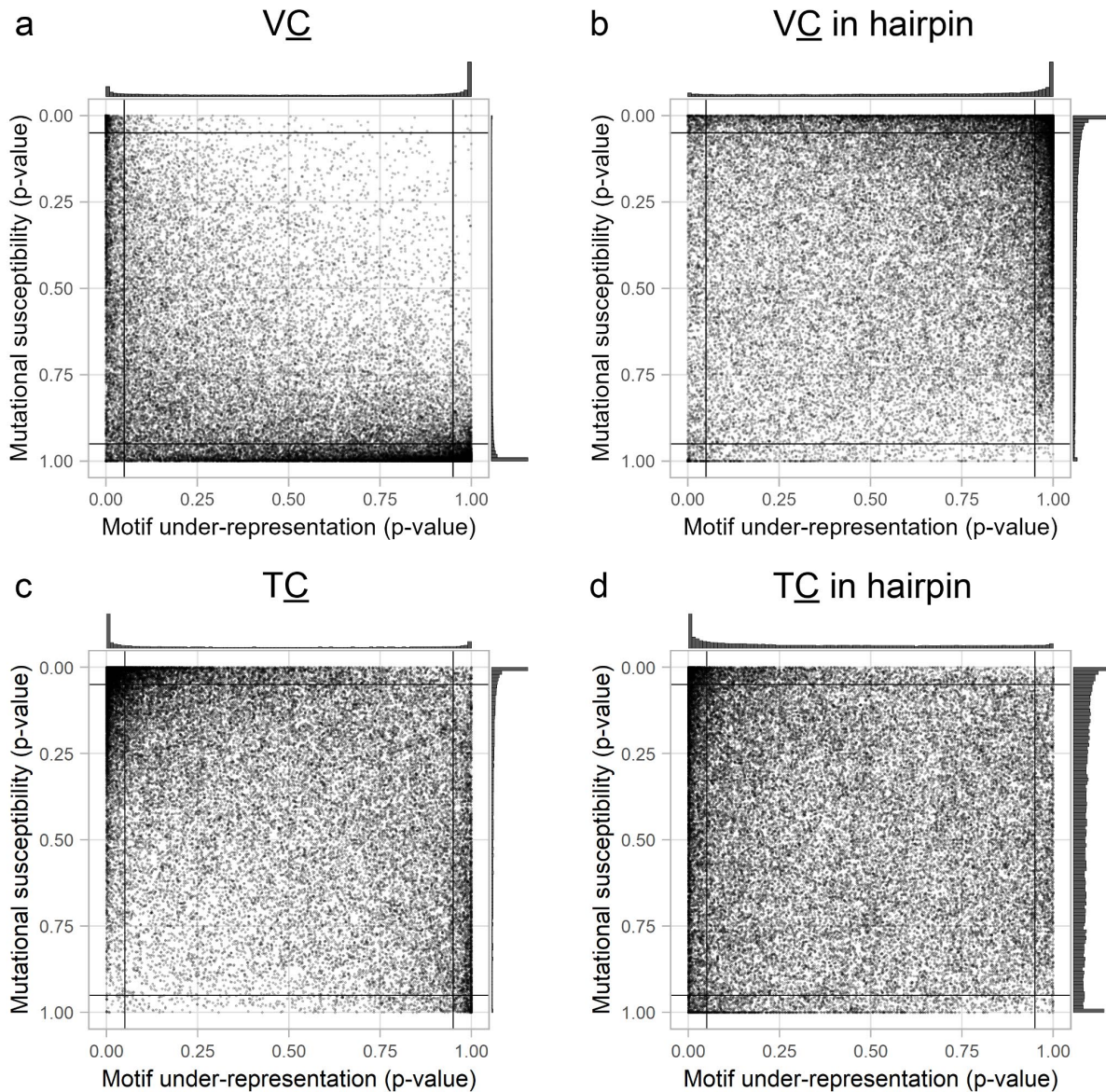Joon-Hyun Song, Liliana M. Dávalos, Thomas MacCarthy, Mehdi Damaghi

Figure S1. Comparison of CDUR statistics between APOBEC motif in hairpin-forming structure, related to Figure 1a. a, Motif under-representation and mutational susceptibility of VC motif in the human genome. b, Motif under-representation and mutational susceptibility of VC motif in hairpin-forming structures in the human genome. c, Motif under-representation and mutational susceptibility of TC motif in the human genome. d, Motif under-representation and mutational susceptibility of TC motif in hairpin-forming structures in the human genome.
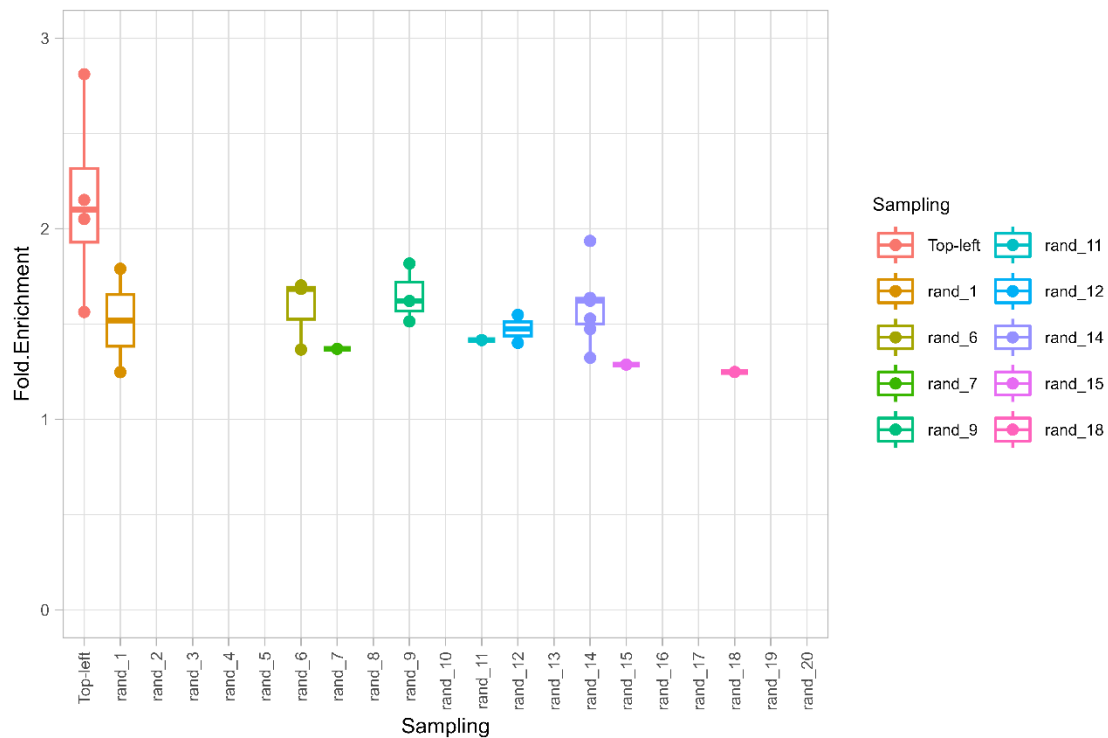
Figure S2. Human top-left corner genes KEGG pathway enrichment compared to the same number of randomly selected genes, related to Figure 1c. Cancer-associated KEGG pathway fold enrichment of randomly selected genes was significantly lower than cancer-associated KEGG pathway fold enrichment of the top-left genes in Figure 1a.
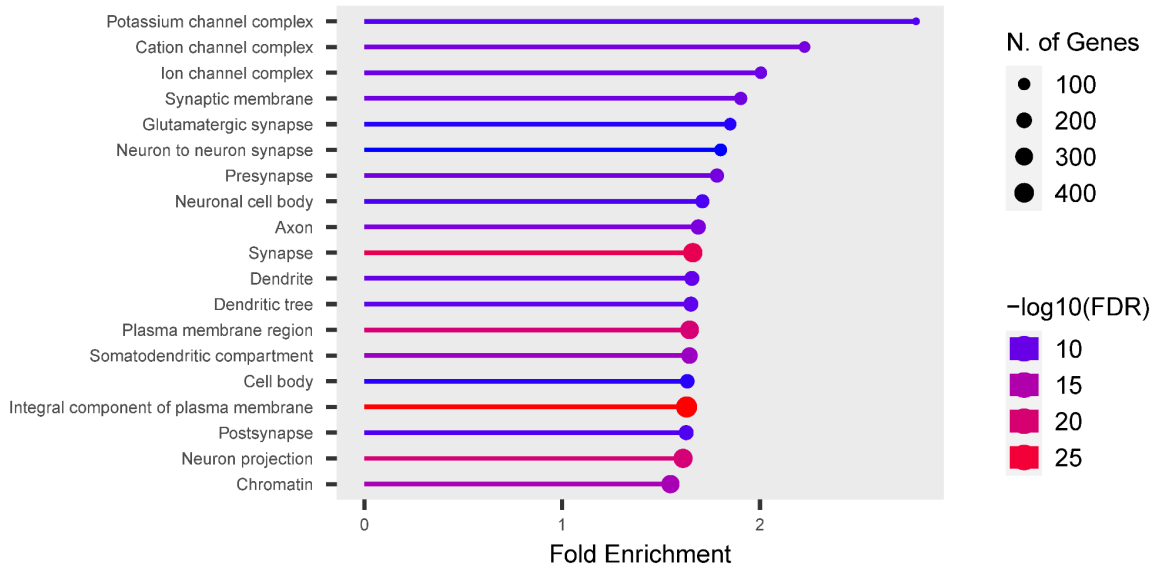
Figure S3. Gene enrichment analysis of GO Cellular Component on top-left corner genes in human APOBEC3A/B TC motif CDUR plot, related to Figure 1. Genes in the top-left partition of the CDUR plot showed significant enrichment of GO Cellular Components associated with channel complex and synapse components.
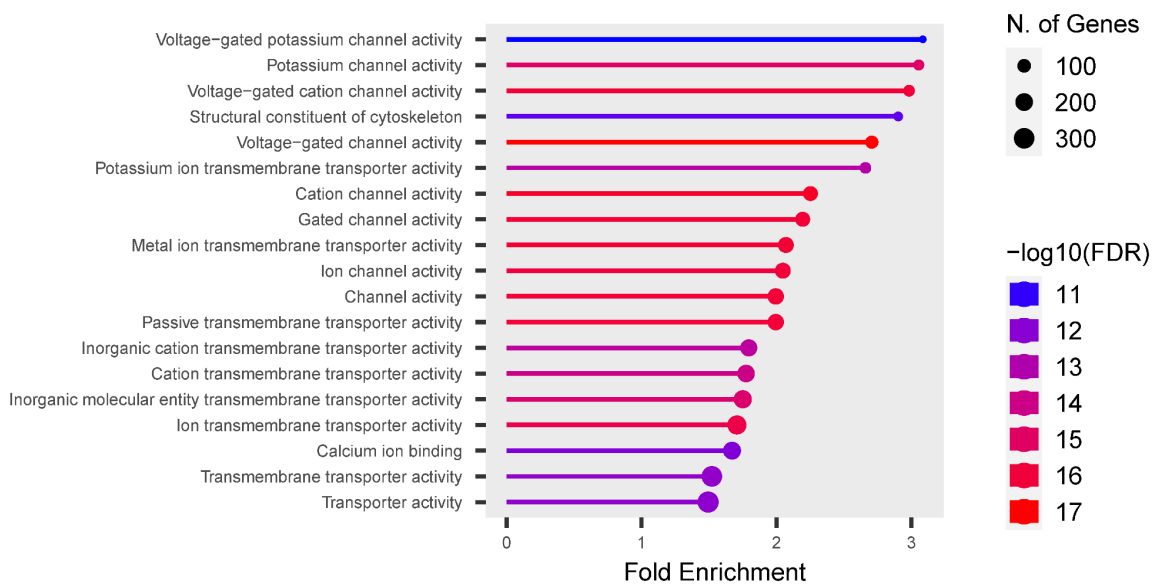
Figure S4. Gene enrichment analysis of GO Molecular Functions on top-left corner genes in human APOBEC3A/B TC motif CDUR plot, related to Figure 1. Genes in the top-left partition of the CDUR plot showed significant enrichment of GO Molecular Function associated with ion channel or transporter activity.
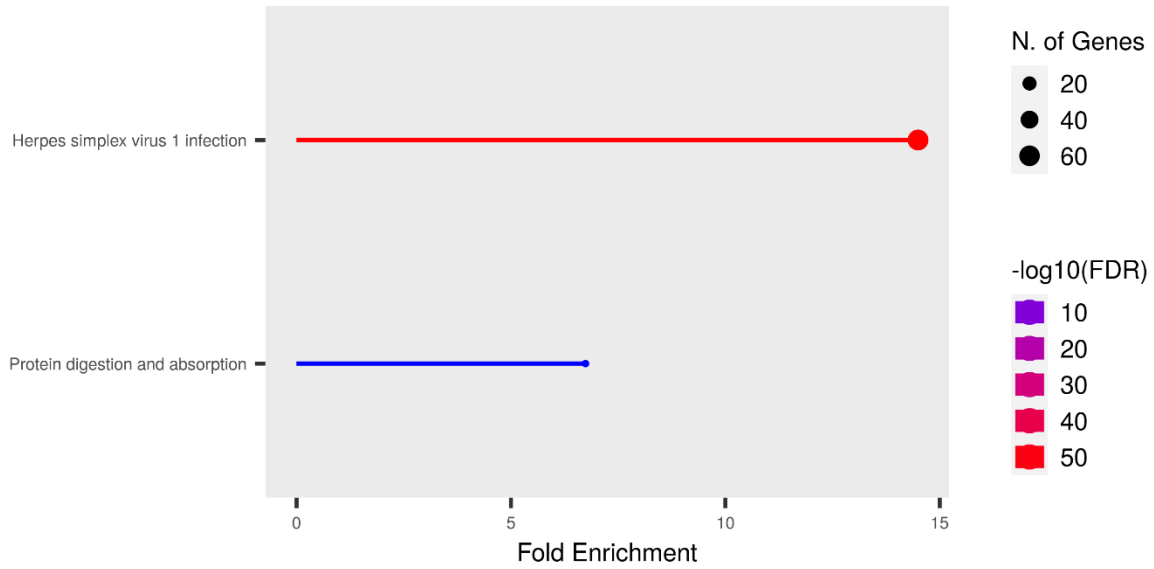
Figure S5. Gene enrichment analysis of KEGG pathways on the bottom-right corner genes in human APOBEC3A/B TC motif CDUR plot, related to Figure 1. Genes in the bottom-right partition of the CDUR plot showed significant enrichment of the KEGG Pathway associated with herpes simplex virus 1 infection.
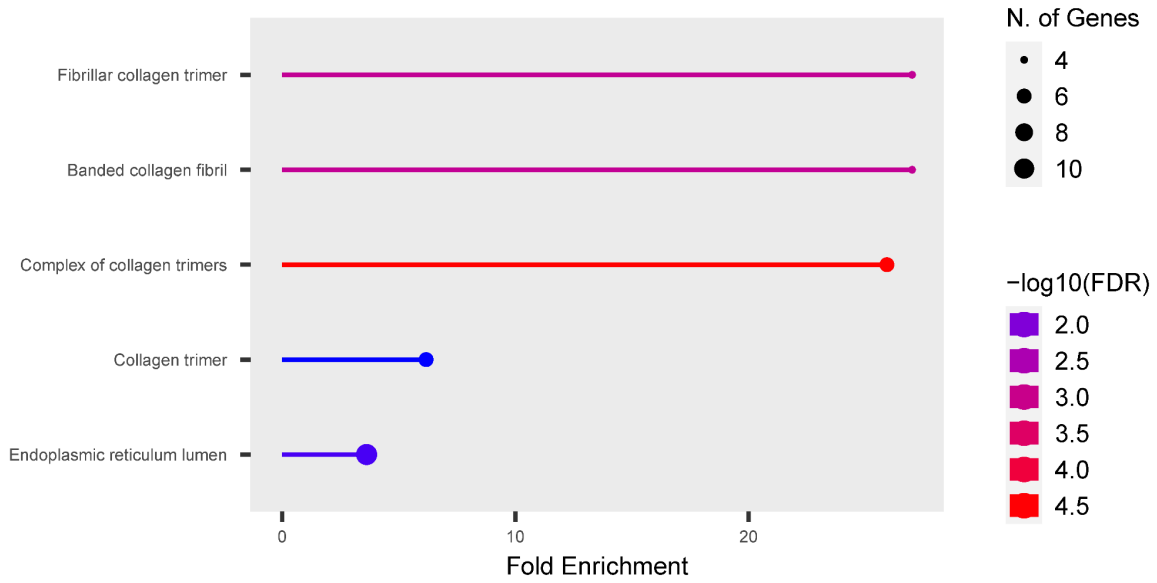
Figure S6. Gene enrichment analysis of GO Cellular Component on the bottom-right corner genes in human APOBEC3A/B TC motif CDUR plot, related to Figure 1. Genes in the bottom-right partition of the CDUR plot showed significant enrichment of GO Cellular Component associated mostly with collagen.
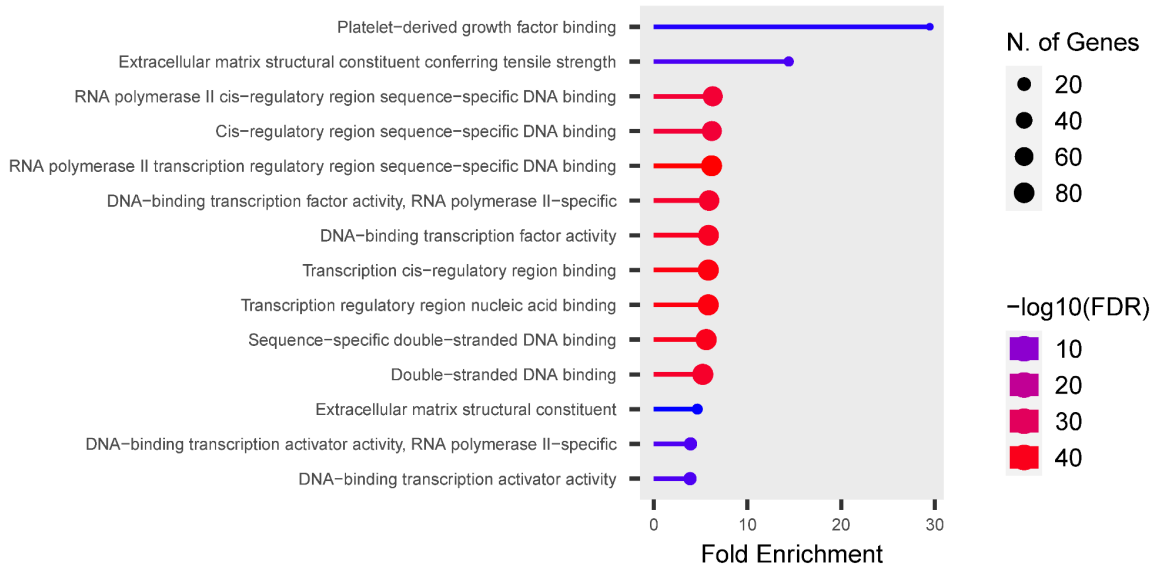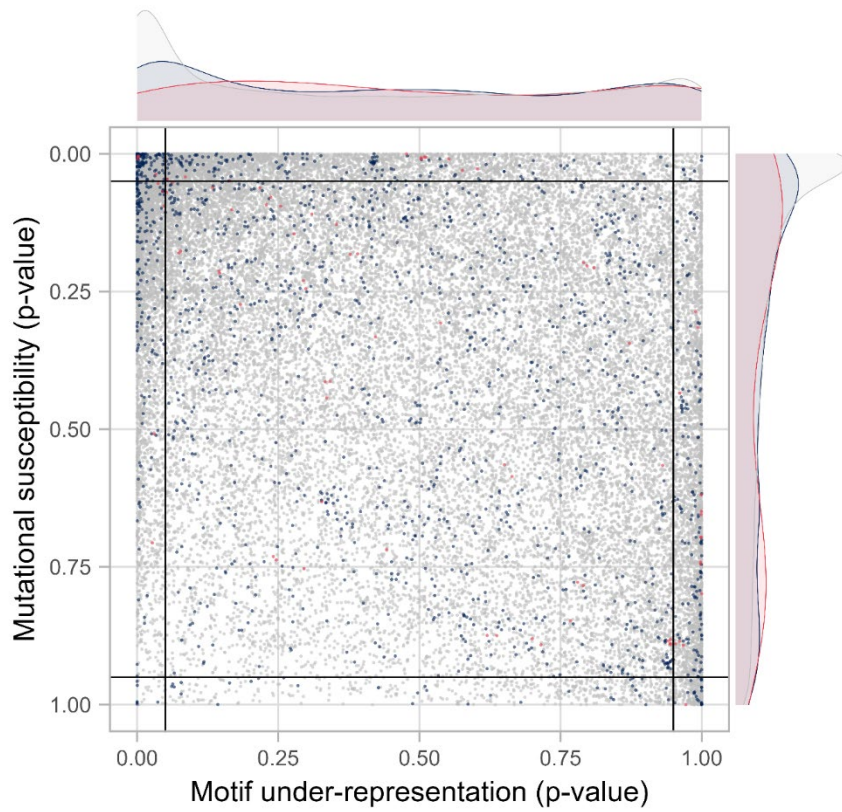
Figure S7. Gene enrichment analysis of GO Molecular Functions on the bottom-right corner genes in human APOBEC3A/B TC motif CDUR plot, related to Figure 1. Genes in the bottom-right partition of the CDUR plot showed significant enrichment of GO Molecular Function associated with DNA binding.

Figure S8. Distribution comparison between genes from two sex chromosomes and autosomal chromosomes, related to Figure 2. The number of genes compared for analysis: 1,676 genes from chromosome X and 83 from chromosome Y.
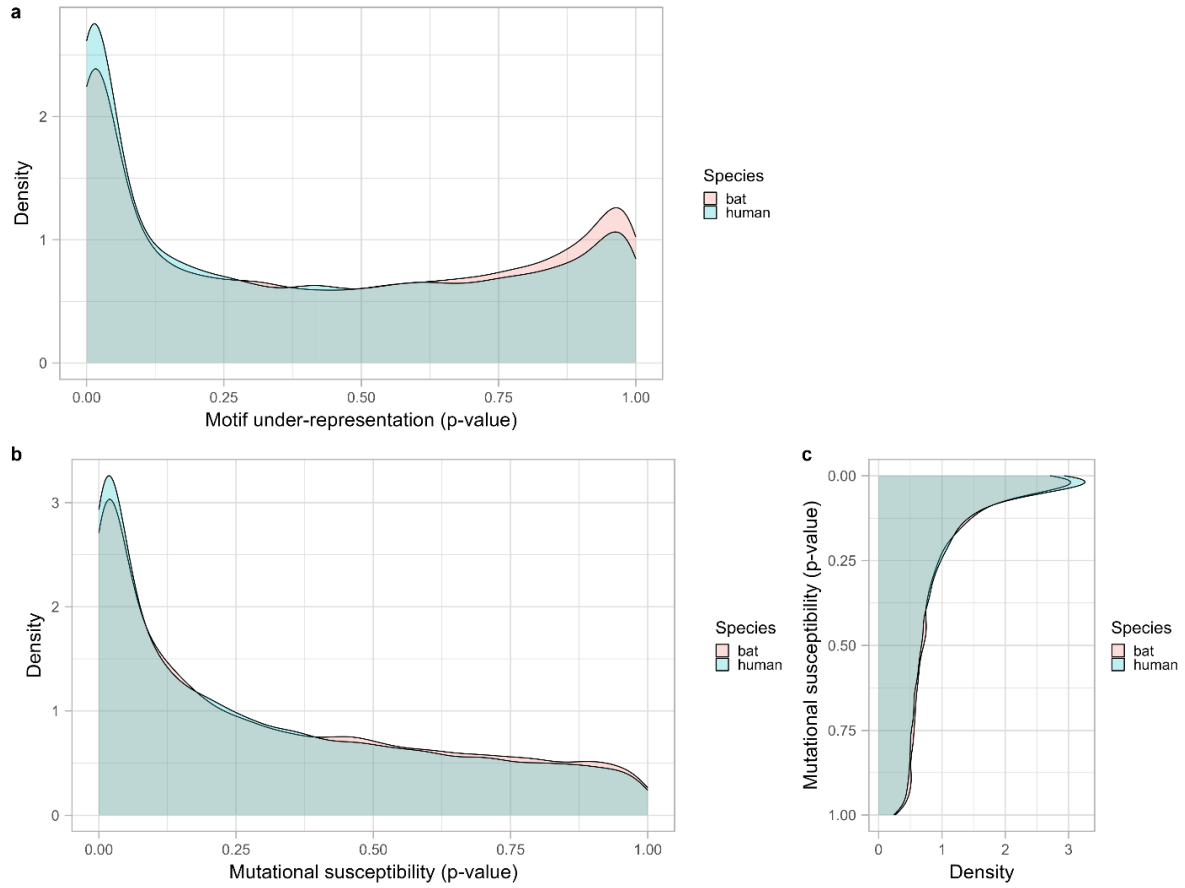
Figure S9. Comparison between human and *Pteropus alecto* genome motif under-representation and mutational susceptibility, related to Figure 3. a, Comparison of TC motif under-representations of human and *Pteropus alecto*. b, Comparison of TC motif mutational susceptibility of human and *Pteropus alecto*. c, The same comparison of Figure S9b in Y-axis representation.
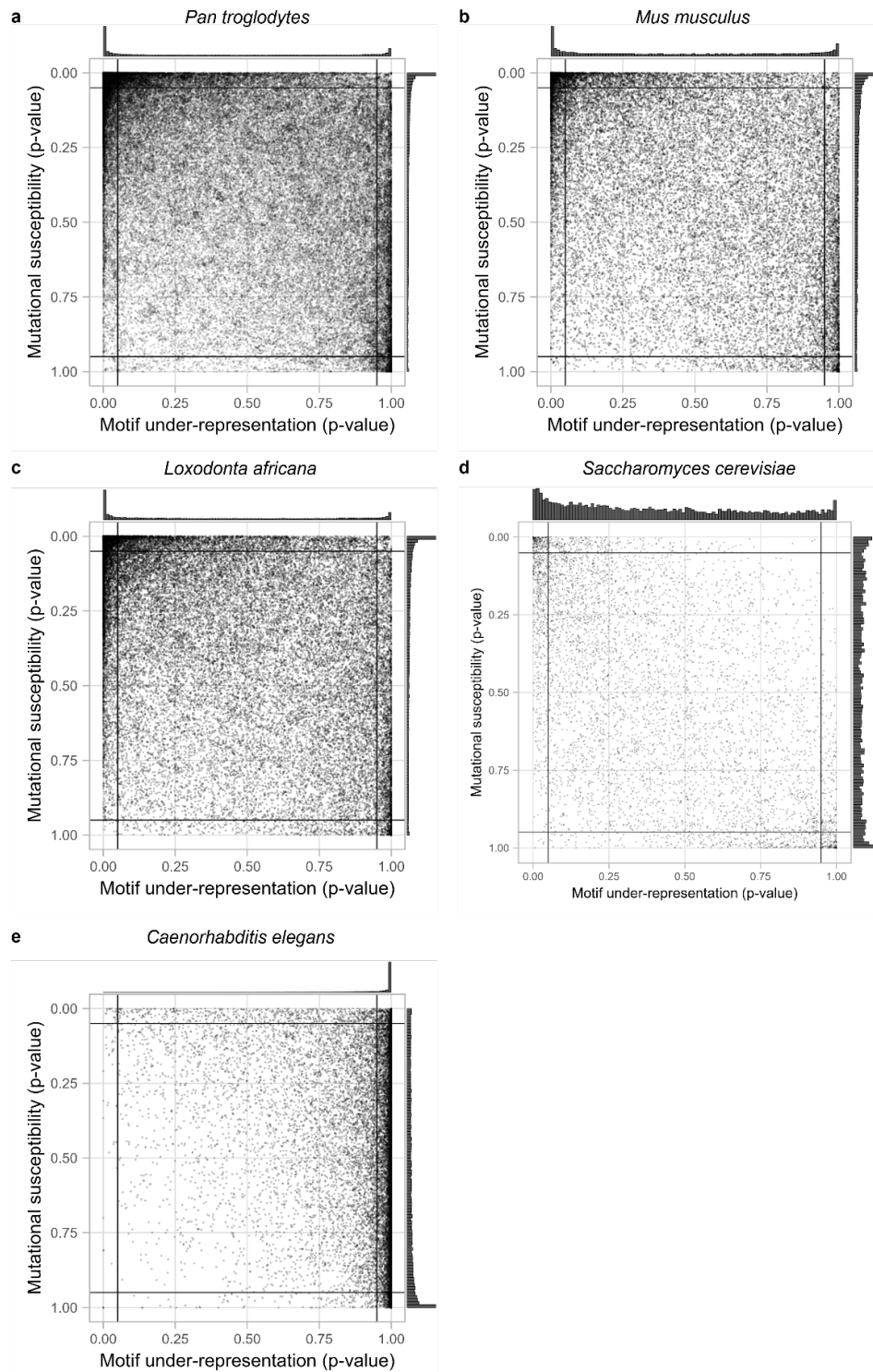
Figure S10. CDUR plots for 5 organisms: chimpanzees, mice, elephants, yeast, and *C. elegans*, related to Figure 3. CDUR statistics (motif under-representation and mutational susceptibility) of T<u>C</u> motifs of a, chimpanzee (*Pan troglodytes*), b, mice (*Mus musculus*), c, elephant (*Loxodonta africana*), d, yeast (*Saccharomyces cerevisiae*), e, nematode (*Caenorhabditis elegans*).
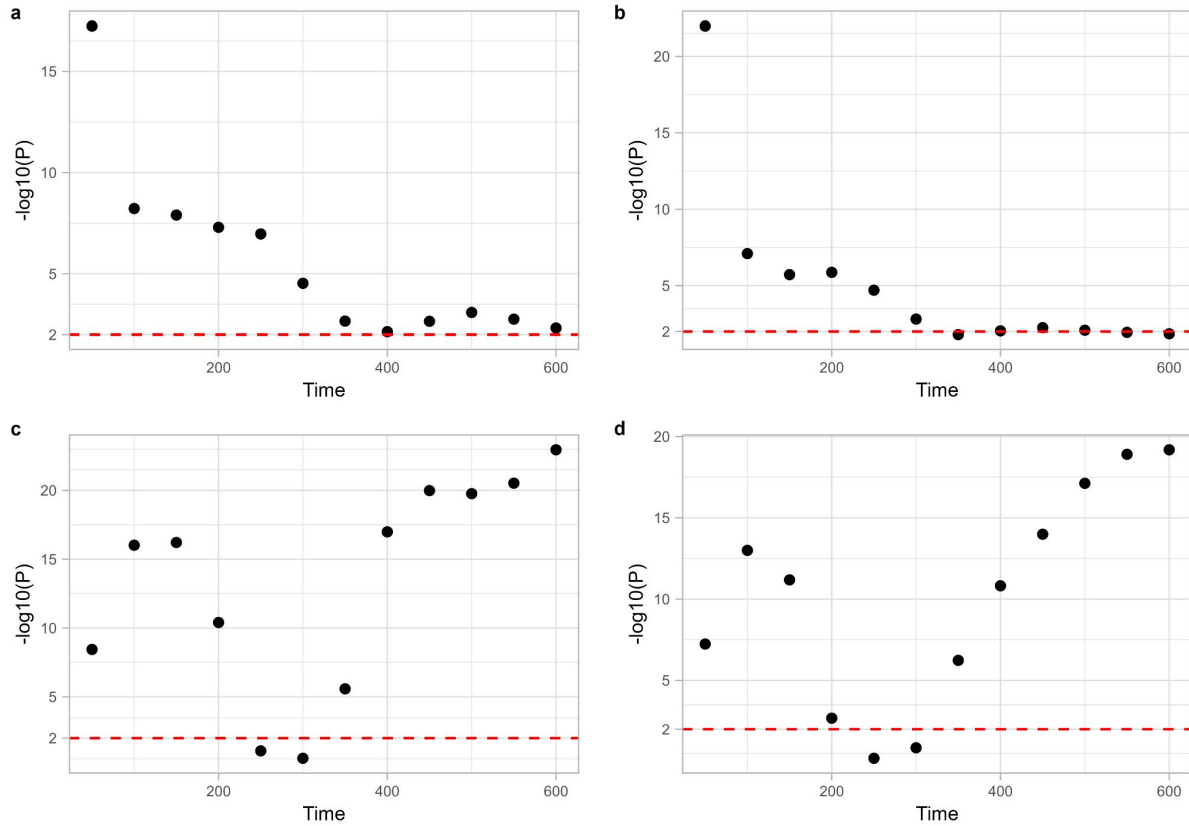
Figure S11. Statistical significance between simulations at every 50 time-points by the number of genotypes and heterogeneity, related to Figure 4. P-values were obtained by t-test on each simulation at every 50 time points. a. Statistical difference by the number of genotypes between with and without APOBEC mutations decreases over time. b. Statistical difference by heterogeneity between with and without APOBEC mutations decreases over time. c. Statistical difference by the number of genotypes between uniform and bimodal distribution increases over time. d. Statistical differences by the number of genotypes between uniform and bimodal distribution increase over time.