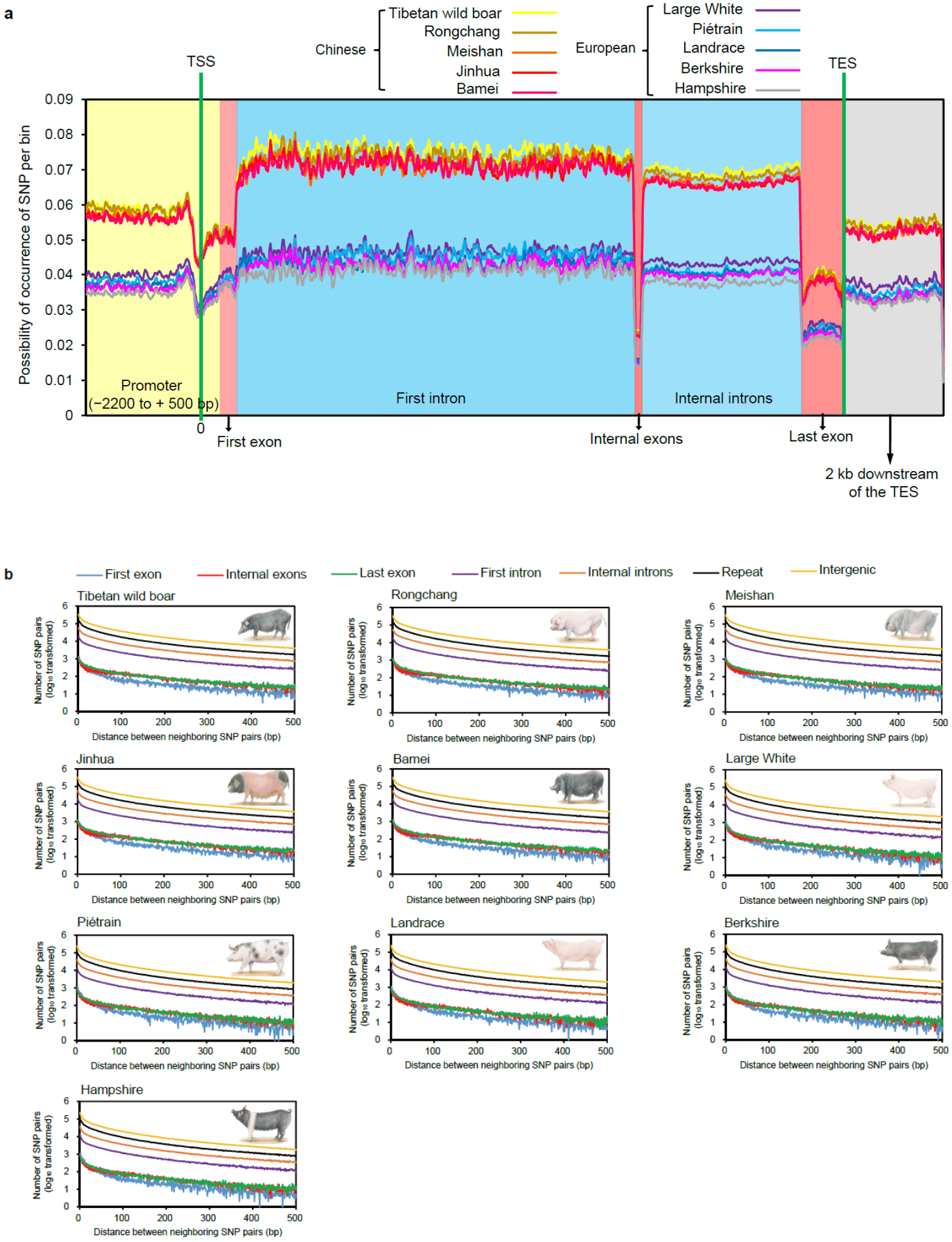

Supplemental Fig. S10. Characterization of SNPs in protein coding genes. (a) Possibility of SNP occurrence in protein coding genes. The canonical gene structure is defined by seven different features, denoted by the following on the *x* axis: 1, promoter; 2, first exon; 3, first intron; 4, internal exons; 5, internal introns; 6, last exon; and 7, 2kb downstream of the TES. *Y*-axis represents the possibility of SNP occurrence per bin. Each feature of

various length of coding genes was analyzed separately and fitted into equal numbers of bins. Each dot in the respective lines denotes the moving average of 5 bins. TSS (left green line), transcription start site. TES (right green line) transcription end site. As expected, SNP ratio is not evenly distributed across the different functional structures of genes, with the coding sequences having lower SNP rate than introns. We also observed a genome-wide trend for elevated SNP levels in first exon regions than internal and last exons, possibly due to some functional motifs overlapping between proximal region of promoters and first exons. Notably, the Chinese pigs also exhibited significantly higher SNP ratio across genes than European pigs ($P < 10^{-16}$, Mann-Whitney *U* test). **(b)** Distribution of distances between neighboring SNPs in different genomic elements. Compared with SNPs in first and internal introns, repeat and intergenic regions, SNPs in CDS region (i.e. first, internal and last exons) exhibited large fluctuations, which may be attributed to increased occurrence of coding SNPs at the 3rd position; therefore, neighboring SNPs are more likely to be separated by 3N-1 bp.