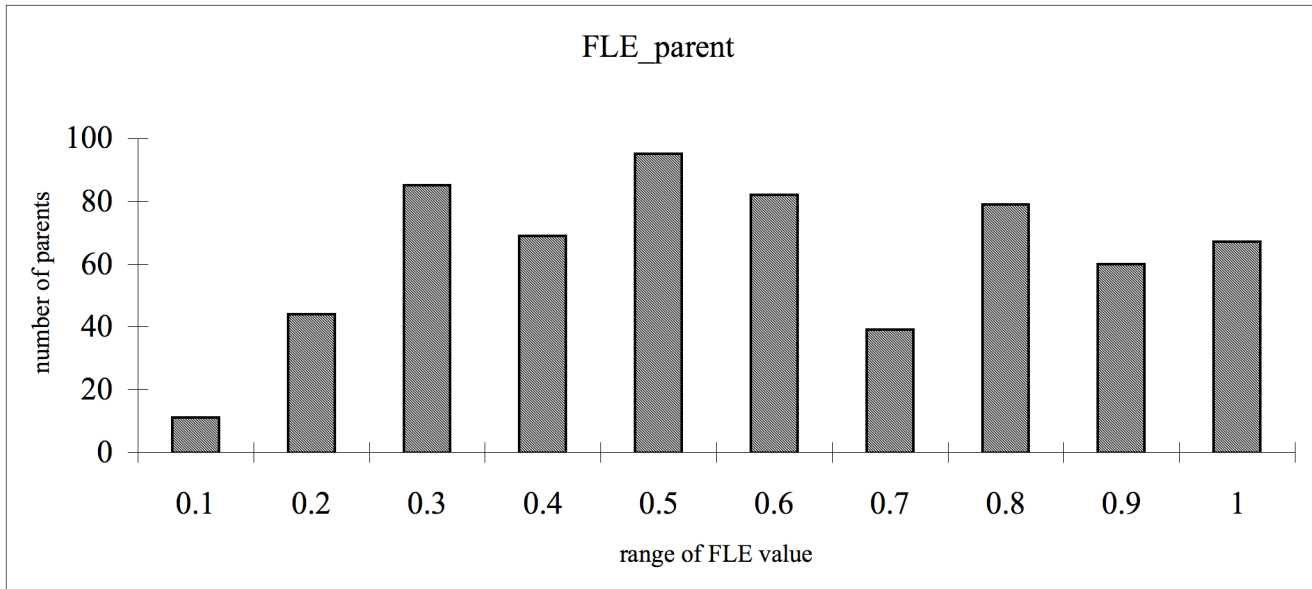
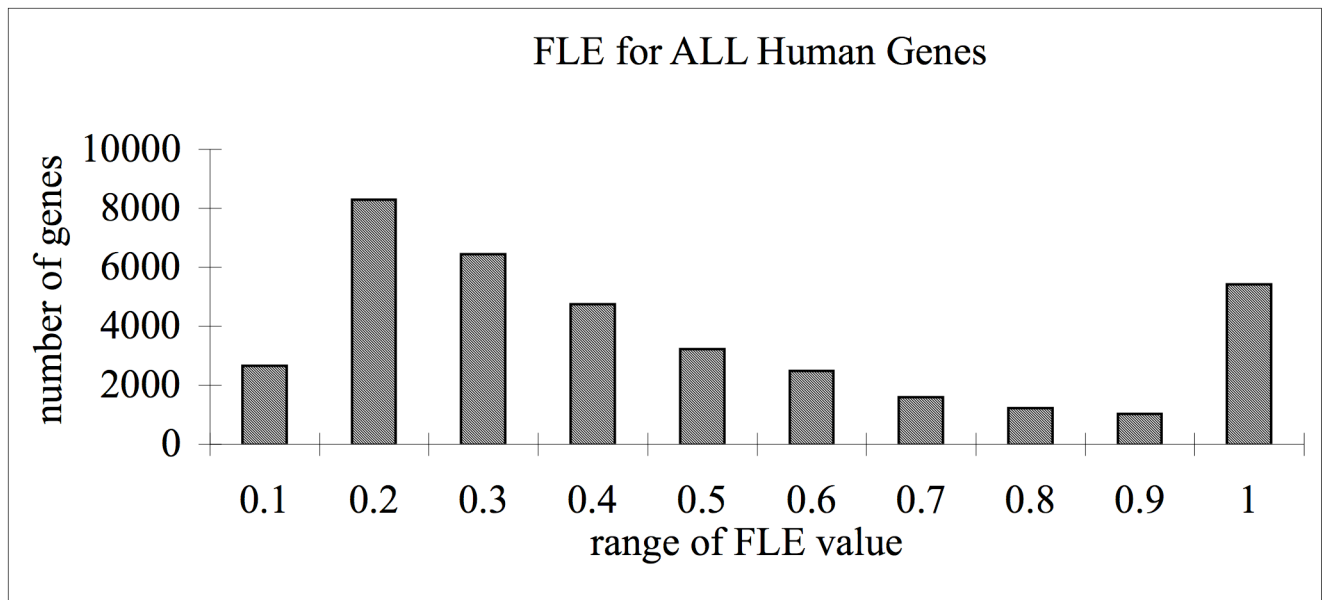


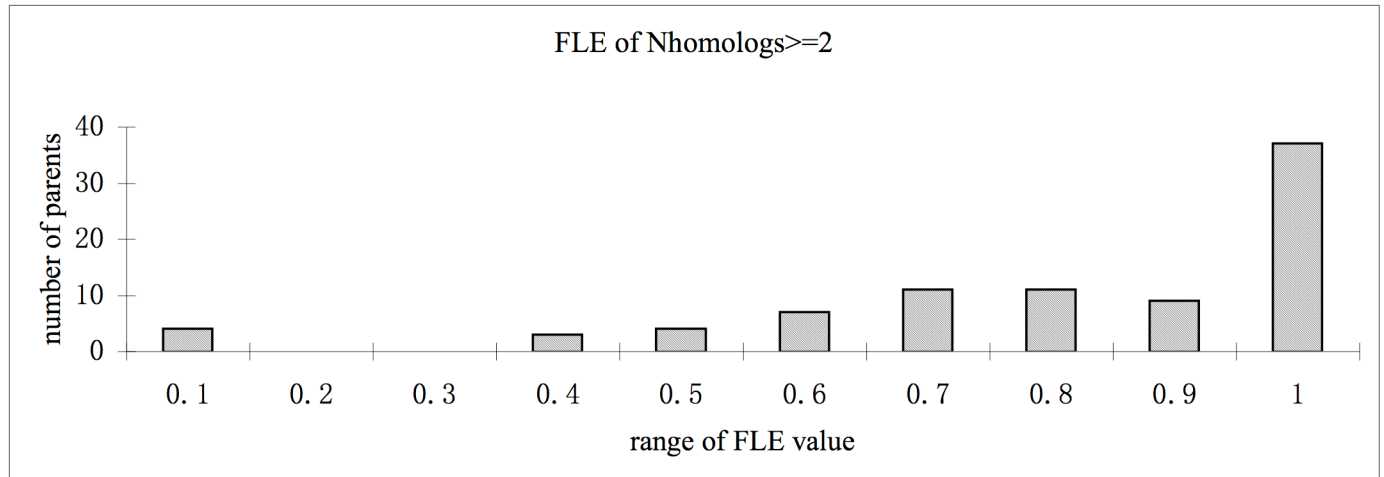
(A)



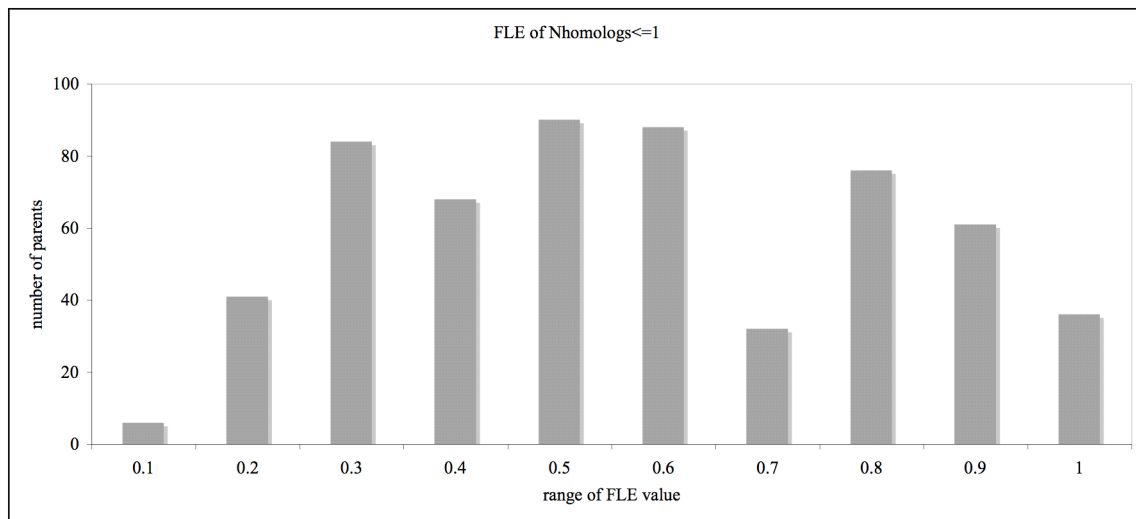
(B)



(C)



(D)



Supplementary Figure 1: Distributions of the fraction of the largest exon (FLE). FLE is the ‘fraction of the largest exon’, covered within the coding sequence of the proteins. We derived histograms for this measure for various sets of sequences. Histogram bins labelled x contain all values such that $x-0.1 <$

FLE $\leq x$. **(A)** We calculated FLE_{parent} for all parent genes for human PRs. FLE_{parent} has an approximately uniform distribution for this data set. **(B)** The overall distribution for human genes shows a peak in FLE between 0.2-0.3; the peak at 0.9-1.0 is due to single-exon genes. **(C)** The distribution of FLE_{parent} for all human PRs with $N_{\text{homologs}} \geq 2$, according to the local gene order test (see *Methods* for details). Note that the putative parents for these PRs tend to have a single large exon; more than two-thirds of these apparent PRs arise from genomic clusters of Zn-finger containing proteins. **(D)** As for (C), but with $N_{\text{homologs}} \leq 1$.