Figure S4: Distribution of variant counts per gene as observed or predicted under best-fit parameters of model A using 3 different estimation techniques. A: linear space, B: log space. Black curve: observed counts, red curve: prediction based on least-squares fit to the cumulative distribution function (see Online Methods), green curve: maximum likelihood estimate, blue curve: least squares fit to the accumulation curve of truncated genes as shown in Figure 1. The CDF method was chosen and maximum likelihood was discarded because its estimates did not fit the observations.