



Supplemental Figure 3: Histograms of null distributions for standardized global statistics for 4 GO cellular component categories in the analysis of survival among adenocarcinomas. A scaled normal density is overlaid on each histogram as the asymptotic distribution of a standardized Wilcoxon,

$$Z_W = \frac{W - \frac{g(n+1)}{2}}{\sqrt{\frac{g(n-g)(n+1)}{12}}} \quad \text{where} \quad g = \sum_{i=1}^n c_i$$

Empirical p-values were calculated for the observed statistic relative to the permuted nulls for each category, and are invariant to standardization of the Wilcoxon statistics. The upper panels show good agreement between the theoretical and empirical null distributions for (A) the most significant category and (B) a marginally significant category. The lower panels (C) and (D) display poorly standardized statistics that have greater variance than the theoretical distribution, and would thus have inflated Type I errors for pooling-based p-values ($p = 0.0133$ and $p = 0.0352$, respectively). The variability in null distributions demonstrates that pooling the permuted statistics of all categories into a single null would be inappropriate.