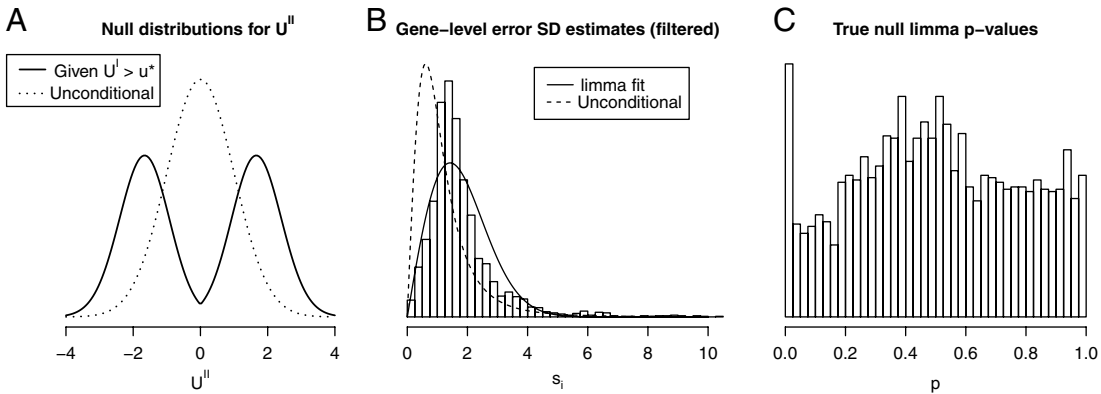

Fig. 2. (*A*) The null distribution of the test statistic is affected by filtering on the maximum of within-class averages. In this example, all genes have a known common variance, the filter statistic is the maximum of within-class means, and the test statistic is a *z*-score. The unconditional distribution of the test statistic for nondifferentially expressed genes is a standard normal. Its conditional null distribution, given that the filter statistic (U^I) exceeds a certain threshold (u^*), however, has much heavier tails. Using the unconditional null distribution to compute *p*-values after filtering would therefore be inappropriate. See *SI Text* for full details. (*B* and *C*) Overall variance filtering and the *limma* moderated *t*-statistic. Data for 5,000 nondifferentially expressed genes were generated according to the *limma* Bayesian model ($n_1 = n_2 = 2$, $d_0 = 3$, $s_0^2 = 1$). (*B*) Filtering on overall variance ($\theta = 0.5$) preferentially eliminated genes with small s_i, causing gene-level standard deviation estimates for genes passing the filter (histogram) to be shifted relative to the unconditional distribution used to generate the data (*dashed curve*). The *limma* inverse χ^2 model was unable to provide a good fit (*solid curve*) to the s_i passing the filter. (*C*) The fitting problems lead to a posterior degrees-of-freedom estimate of ∞. As a consequence, *p*-values were computed using an inappropriate null distribution, producing too many true-null *p*-values close to zero, i.e., loss of type I error rate control. An analogous analysis comparing biological replicates from the ALL study—so that real array data were used but no gene was expected to exhibit significant differential expression—yielded qualitatively similar results.