Supplement 1: Analysis of an Artificial Spike-in Data Set

We generated an artificial spike-in data set using microarray data on a group of patients with Hyperdiploid type of childhood leukemia identified through the St. Jude Children's Research Hospital Database (Yeoh et al., 2002). There were 88 subjects in this group. We selected (at random) 1255 probe sets and limited all subsequent analyses to this subset of (putative) genes. Next we designated 125 (randomly selected) genes for further modification as described below. The following resampling procedure was employed to artificially produce the effects of differential expression. Two subsamples of size $n = 10$ are drawn without replacement from a total of 88 arrays, each reporting expression levels on the pre-chosen set of 1255 genes. One sample of 10 arrays is left intact, while a constant effect size of 1 is added to log-expressions of the pre-designated 125 genes in all arrays pertaining to the second sample. The two samples thus created are used to detect "truly differentially expressed" genes by the extended Bonferroni multiple testing procedure. The resampling procedure was repeated 1000 times and the proportions of true discoveries were averaged over the subsamples.

Shown in Figure 1 is the true positive rate (power) resulting from the application of $Bonf^{\gamma}$ to the biological data set thus generated at different values of the parameter γ. The overall low power of the t-test in this case is attributable to the small sample size (n=10). Possible violations of the normality assumption can also reduce the power of this test in the presence of nuisance parameters (Wilcox, 2001). However, the gain in power is substantial when the power values for $\gamma \geq 1$ are compared to the first point on the curve obtained at $\gamma = 0.5$.

References

Yeoh, E.J., Ross, M.E., Shurtleff, S.A., Williams, W.K., Patel, D., Mahfouz, R., Behm, F.G., Raimondi, S.C., Relling, M.V., Patel, A., Cheng, C., Campana, D., Wilkins, D., Zhou, X., Li, J., Liu, H., Pui, C.H., Evans, W.E., Naeve, C., Wong, L. & Downing, J.R. (2002), 'Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling', *Cancer Cell* **1**(2), 133-143.

Wilcox, R. R. (2001), *Fundamentals of Modern Statistical Methods*, Springer, NY.

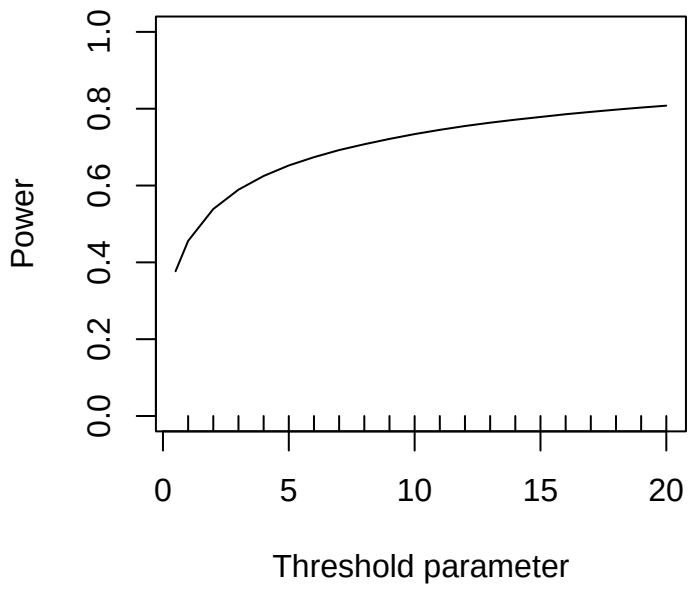

Figure 1: The power of $Bonf^{\gamma}$ when applied to the artificial spike-in data set generated from the St. Jude Children's Research Hospital Database.