gFWER(k)-controlling single-step common-cutoff augmentation procedure

Alternatively, we use an augmentation procedure for controlling tail probabilities of the number of false positives. For this, we use the generalized family-wise error rate (gFWER) controlling single-step common-cut-off T(k+1) procedure with a reference permutation distribution. The gFWER is defined as the chance of committing at least (k+1) type I errors:

$\mathrm{gFWER}(k) \equiv \Pr(V_n > k)$

In the special case where k=0, it corresponds to the usual FWER. Here, the cutoff is based on the 0.95th quantile of (k+1)st ordered T-statistics for the reference distribution, to control the gFWER(k) at level $\alpha=0.05$ (probability of having more than k type I errors is 0.05) (Dudoit, et al., 2004).

For the reference permutation distribution, T-statistics was derived using randomized class labels. A 100 permutations were used to limit the computational time for simulations.

The high cutoffs obtained using maxT for controlling the family wise error rate (FWER) were caused by a few extreme clones for each permutation. We, therefore, tested if using T(k+1)th most extreme T-statistics from the permutation distribution would give us a better cutoff. Figure S 9 shows a few examples of T-statistics smoothed by HMM. It is evident that the segmentation is successful for both cases as segments with true copy number differences (red) are mostly separated from segments with no true copy number differences (black). However, using the maxT cutoff (green, k=0), whole segments are just missing significance (left) or barely significant (right). A more stable cutoff seems to be defined by using e.g. k=5 (blue, 6th most extreme T-statistics from each permutation). Using this cutoff, most true copy number differences are identified for both methods while only picking up a few additional false positives.

Figure S 10 shows the revised results for sensitivity and specificity using k=5. The sensitivity for HMM smoothed T-statistics is now 0.85 compared to 0.48 before, while specificity has only decreased to 0.99 compared to 1.00 before. The performance for GLAD with outlier adjustment has also increased significantly and so has the sensitivity for smoothing of original $\log_2$ ratios. However, for smoothing of $\log_2$ ratios, DNAcopy still performs the best, and even though HMM smoothed T-statistics performs slightly better (Wilcoxon P-value: 0.002), the specificity is also significantly decreased (Wilcoxon P-value: 1.92e-68).

References

Dudoit, S., van der Laan, M.J. and Birkner, M.D. (2004) Multiple testing procedures for controlling tail probability error rates, *Division of Biostatistics, University of California, Berkeley*, **Technical Report 166**.