Greatly improved allele-specific tumor copy numbers with DNA microarrays when a matched normal is available

Pierre Neuvial*, Henrik Bengtsson, and Terence P. Speed
Department of Statistics, University of California, Berkeley, CA 94720, USA.

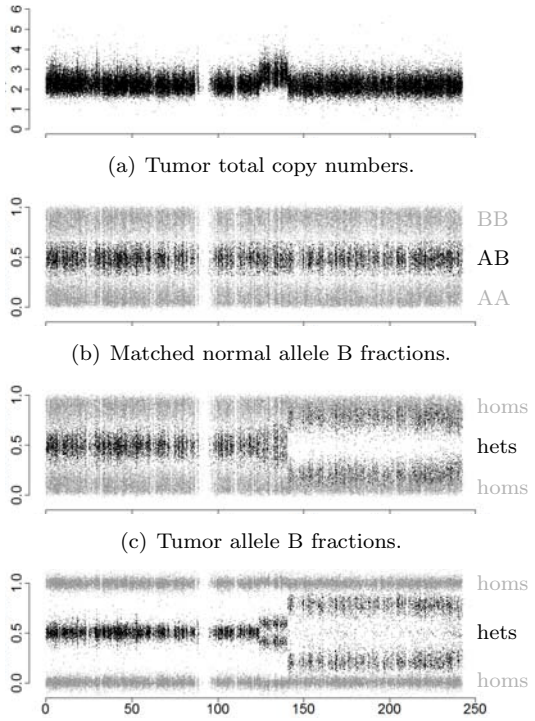
High-throughput genotyping microarrays assess both total DNA copy number (Figure 1a) and allelic composition (Figure 1c), which makes them a tool of choice for copy number studies in cancer, including total copy number and loss of heterozygosity (LOH) analyses. Even after state of the art preprocessing methods, allelic signal estimates from genotyping arrays still suffer from systematic effects, resulting in poor signal-to-noise ratios (Figure 1c) that make them difficult to use effectively for such downstream analyses.

We propose a method, TumorBoost[2], for normalizing allelic estimates of one tumor (Figure 1c) sample based on estimates from a single matched normal (Figure 1b). The method applies to any paired tumor-normal estimates from any microarray-based technology, combined with any preprocessing method. We demonstrate that it increases the signal-to-noise ratio of allelic signals (Figure 1d), making it significantly easier to detect allelic imbalances. TumorBoost increases the power to detect somatic copy-number events (including copy-neutral LOH) in the tumor from allelic signals of Affymetrix or Illumina origin.

Importantly, high-precision allelic estimates can be obtained from a single pair of tumor-normal hybridizations, if TumorBoost is combined with single-array preprocessing methods such as (allele-specific) CRMA v2 for Affymetrix or BeadStudio's (proprietary) XY-normalization method for Illumina. This makes the method suitable for both large and small studies, and practical for applied medical diagnostics, because each patient can be analyzed independently of others. Based on these results, we recommend the use of matched normal samples in cancer copy number studies.

The method is implemented in the open-source R package aroma.cn part of the Aroma Project[1].

*Corresponding author (pierre@stat.berkeley.edu).



(a) Tumor total copy numbers.

(b) Matched normal allele B fractions.

(c) Tumor allele B fractions.

(d) Tumor allele B fractions after TumorBoost.

Fig. 1. After TumorBoost it is significantly easier to identify the three regions of allelic imbalances that the tumor carry. Data is from chromosome 2 in a TCGA ovarian sample. Only two arrays were used - no other reference samples are needed neither for preprocessing nor normalization.

References

1. Henrik Bengtsson. The Aroma Project - an open-source R framework for your microarray analysis. http://www.aroma-project.org/, 2010.
2. Henrik Bengtsson, Pierre Neuvial, and Terence P. Speed. TumorBoost: Normalization of allele-specic tumor copy numbers from a single pair of tumor-normal genotyping microarrays. *BMC Bioinformatics*, 11(245), May 2010.