# Estimate tumor purity using B-allele frequency (BAF) information in SNP arrays

# We use two steps to estimate purity and determine whether a copy number alternation (CNA) was clonal or subclonal.

# Step 1: For each segment mixed by CN2 (copy neutral) and a CNA event (CN1, LOH or CN3), we used the BAF pattern to estimate , the fraction of cells carrying the CNA. First, we made a histogram of BAF and estimated the center of the two BAF bands: . Then, we can estimate as a function of according to the following table. For CN0, however, BAF pattern is similar to CN2 and thus we cannot estimate . In practice, we found it difficult to decide the absolute copy number for amplifications. Misspecification of the absolute copy number typically severely biases the estimate of for amplifications. Thus, we only estimate for CN1 deletions and LOH events. See the left panel of Figure S1B.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | Estimate |
| CN2 + CN1 |  |  |  |
| CN2 + CN3 |  |  |  |
| CN2 + LOH |  |  |  |

# Jacobs et al., Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. (2012) 44(6):651-658.

# Step 2: Estimate the number of subclones. For each sample, we derive for all CN1 deletions and LOH events. After estimating for all deletions and LOHs, we estimated the density of using a nonparametric statistical method. Each peak based on the histogram was determined as a subclone. The proportion of cells for each subclone was estimated as the center of each cluster. We assumed that the most right clone was the primary clone and its estimated represented the purity of the tumor. All CN1 deletions and LOH events with belonging to the primary clone were determined as clonal CNA. All other CN1 deletions and LOHs were determined as subclonal CNAs.