Modified PLASQ algorithm. Briefly, after normalization by invariant set method of the normal and tumor samples, the normalized CEL files are separated into normal and tumor folders. For each SNP, a model is fitted on the normal samples. The signals of probes are assumed to be normally distributed with means depending on the type of the probes. Normal CEL files are used to calculate the coefficients. These coefficients are used to calculate the raw copy numbers of tumor samples. Then these raw copy numbers are used to calculate the inferred parental specified copy numbers (PSCN) after applying the circular binary segmentation algorithm to find segmentations with roughly constant SNP raw copy numbers within each segmentation. Each SNP then has a minor PSCN and a major PSCN, reflecting the copy numbers from each of the two parents. The inferred allele specified copy numbers (ASCN) are then calculated from PSCN. LOH calls are inferred to be any loci with the minor PSCN equal to 0 and the major PSCN greater than 0. For more detail of the algorithm, see [23]. Our main modifications to the package are:

1. Extended applicable area form data of 100K set only to include data of 10K set.
2. Modified the PSCN inference algorithm to make the algorithm less conservative by considering the stroma (mixed) effect.
3. Modified the algorithm to make data reading process more efficient and running faster.

For points 2, suppose we observed the raw copy numbers of certain SNP as $X$. This $X$ will be treated as summation of two random components

$$X = pX_1 + (1 - p)X_2,$$

(1)

where $X_1 \sim Normal(2, \sigma^2)$ is the (random) raw copy numbers contributed by normal DNA contaminated in the tumor DNA and $X_2 \sim Normal(\mu, \sigma^2)$ is the random raw copy numbers contributed by the tumor sample, where $p$ is the proportion of the normal DNA contained in the whole DNA sample when running the chip. This $p$ could be different from sample to sample, but in our analysis that follows, we simply set it to be a constant 0.15. The raw copy numbers are thus defined as $E(X_2) = (X - 2p)/(1 - p)$ by taking expectation on both sides of equation 1. It equivalent to inflate the original raw copy numbers out on both sides from 2.
Figure 3: Modified PLASQ results: A) Raw copy numbers of SNPs by summing up two allele raw copy numbers. B) averaged raw copy numbers among the 16 samples. C) parental specified copy numbers by modified PLASQ. Green: minor parental copy number, red: major parental copy number. D) percents of events of true loss of heterozygosity. E) Percent of events of gains among the 16 samples. F) LOH calls.