LOH Analysis in Partek Genomics Suite

Overview

Loss of Heterozygosity (LOH) in Partek uses a Hidden Markov Model (HMM) to find regions that are most likely to be loss events based on the genotype error and the expected heterozygous frequency at each SNP. Both paired and unpaired analysis are available, however paired analysis is preferred when possible as it is more accurate in its expected genotype frequencies and does not report regions of LOH caused by common haplotype blocks within the study population.

HMM Emission Probability

The HMM will uses the expected probability of observing a given genotype call for every informative SNP. The expected probability used will depend on the type of analysis being used.

Unpaired

For unpaired analysis, the probability of observing a heterozygous SNP in a region of LOH is the genotype error rate. In a region without LOH, the probability of observing a heterozygous SNP is estimated using the observed frequency from the baseline samples.

The probability of each state emitting each observed genotype is described as follows:

P(AB | LOH) = e
P(~AB | LOH) = 1 - e
P(AB | ~LOH) = O(AB)
P(~AB | ~LOH) = 1 – O(AB)

AB represents a heterozygous genotype call. The parameter e is the expected genotype error rate specified in the LOH dialog. O(AB) is the observed frequency of heterozygous calls for each SNP. If a genotype baseline is not available to estimate O(AB) for a given SNP, the default heterozygous frequency parameter value will be used.

Paired Analysis

The paired LOH analysis also uses a similar HMM model for each pair of samples. Homozygous SNPs in the paired normal do not provide any information of LOH in the study sample, and are excluded from paired analysis.

P(AB | LOH) = e
P(~AB | LOH) = 1 – e
P(AB | ~LOH) = 1 – e
P(~AB | ~LOH) = e

HMM Transition Probability

The HMM uses the expected probability of being in a current state given the previous state to find the most likely regions of LOH.

The probability of being in a state given the previous state is calculated as

$$a = e^{-d/decay}$$
$$P(S_t = S_{t-1}) = a\,P_{max} + (1-a)\,P_{initial}$$
$$P(S_t != S_{t-1}) = 1 - P(S_t = S_{t-1})$$

Where d is the number of base pairs between neighboring observations, decay is a parameter specified in base pairs, and S is the hidden state. P_{max} is specified within the dialog and represents the maximum probability of retaining the same hidden LOH state as the previous SNP. Setting the decay parameter to 0 will disable the genomic decay using P_{max} for every transition probability. This is the recommended and default setting for LOH analysis within Partek.