White Paper: Allele-Specific Copy Number

Introduction

Allele specific copy number (AsCN) estimates provide potentially important information for the analysis, visualization, and interpretation of copy number changes. AsCN provides an estimated number of copies for each allele rather than an estimated number of total copies of each chromosome. In tumor samples, changes in copy number or loss of heterozygosity (LOH) often occur in only a proportion of the tissue sample due to the heterogeneous nature of biopsy samples, which include some degree of "normal" contamination. When these differences are in only a fraction of the tissue sample, genotyping algorithms may still make heterozygous calls, making it less likely to detect LOH using discrete genotype calls. AsCN allows these changes to be detected as an imbalance in copy number between alleles. In addition, AsCN is useful in the interpretation of total copy number analysis results. A gain of total copy number may be shown by one allele or both alleles. AsCN allows the researcher to resolve these uncertainties.

The calculation of allele specific copy number is as follows:

$AsCN_{ij} = f\left(I_{ij} / R_{ij}\right)$ if informative, missing otherwise

- $AsCN_{ij}$ is the allele specific copy number estimate for allele *i* of SNP *j*
- I_{ij} is the intensity of allele *i* for SNP *j*
- R_{ij} is the reference intensity of allele *i* for SNP *j*. This reference represents the expected intensity for one copy of the allele
- $f(x)$ is a function correcting bias in the intensity measurement for each allele

$$AsCN_{max,j} = max\left(AsCN_{A,j}, AsCN_{B,j}\right)$$
$$AsCN_{min,j} = min\left(AsCN_{A,j}, AsCN_{B,j}\right)$$

Not all SNPs are informative in a sample. When a SNP is not informative, it is treated as a missing value. A SNP is considered informative when it would be expected to be heterozygous if the tissue did not contain any copy number variations.

AsCN can be calculated using two different references depending on the experiment design:

- Paired analysis
- Unpaired analysis

These two analyses will be described below.

Paired Analysis

Paired analysis is the most ideal way to generate AsCN. For best results, we recommend the tumor and normal samples be processed and hybridized together, which reduces the chance for a difference due to batch effects. In addition, the informative SNPs will be determined based on the normal sample regardless of copy number changes in the tumor sample.

In paired analysis, Partek requires genotype calls for the normal paired sample and allele intensities for both the normal and study sample.

Using paired analysis, only the heterozygous SNPs in the reference (normal) sample are considered informative and the reference intensity R_{ij} is taken to be the intensity of allele *j* of SNP *i* in the normal sample.

Unpaired Analysis

Unpaired analysis requires genotypes and allele intensities for both the reference (normal) and study (tumor) samples. The reference intensity R_{ij} is taken as the average intensity of all reference (normal) samples that are heterozygous for SNP *i*.

Unpaired analysis considers a SNP in the tumor sample to be informative if it is a heterozygous call. In tumor samples with LOH or gains of homozygosity, the genotype algorithms will call many more homozygous SNPs than heterozygous relative to the normal samples. This limits the usefulness of unpaired analysis to be informative only to mixed tissue samples. Long stretches of homozygous calls will appear as segments with no informative SNPs (missing values). Consequently, paired analysis is recommended, when possible

Allele Imbalance

The allele imbalance procedure uses the min and max alleles to determine regions that are believed to diverge from a "normal" balance of 1 copy each. Partek defines a proportion score for each informative SNP as:

$$Proportion = \frac{\left(AsCN_{max} - AsCN_{min}\right)}{\left(AsCN_{max} + AsCN_{min}\right)}$$

In idealized data, the following table can illustrate some common scenarios and the expected proportion scores.

$AsCN_{max}$	$AsCN_{min}$	*Proportion*	*Example Description*
1	1	0	Expected balance
2	0	1	Copy neutral LOH
1.5	0.5	0.5	Copy neutral LOH in 50% of a mixed tissue sample
1	0	1	Loss of one allele
1	0.5	0.33	Loss of one allele in 50% of mixed tissue
2	1	0.33	Gain of one allele
2	2	0	Gain of both allele—allele balance does not change.

As seen from the table, the proportion score does not change with changes in total copy number, but rather the relative mixture of each allele. Total copy number analysis is also necessary to find regions of amplification or deletion of both alleles.

To determine regions of similar allele imbalance across many SNPs, Partek transforms the allele specific copy number for each SNP into its proportion score. This score is then segmented to find regions of similar proportion. The proportion reported for each detected region is the mean proportion score of all informative SNPs in the region. The mean proportion score per segment is reported in the imbalance table, which can be sorted on proportion to find segments with the largest degree of allelic imbalance within the sample.

It is very rare to have equal $AsCN_{max}$ and $AsCN_{min}$ (both alleles with identical intensity), which would be required to produce a proportion score of 0. Since the min and max are assigned after AsCN is estimated, a region's $AsCN_{min}$ will always be lower than or equal to the $AsCN_{max}$. For this reason, we recommended considering any regions of small allele proportion as normal. When analyzing good performing samples, we have found proportions less than 0.15 to be common in normal regions. This value may increase in noisier data and may be specific to each sample.

Copyright © 2009 by Partek Incorporated. All Rights Reserved. Reproduction of this material without expressed written consent from Partek Incorporated is strictly prohibited.