

Overview of vignettes for copy number estimation

Rob Scharpf

March 31, 2011

The workflow for copy number analyses in the `crmm` package includes preprocessing and genotyping of the raw intensities followed by estimation of parameters for copy number estimation using `crmmCopynumber`. Supported platforms are those for which a corresponding annotation package is available. Table 1 provides an overview of the available vignettes pertaining to copy number estimation. These vignettes are located in the `inst/scripts` subdirectory of the `crmm` package. HapMap datasets are used to illustrate the workflow and are not provided as part of the `crmm` package. Users wishing to reproduce the analysis should download the HapMap CEL files (Affymetrix) or the `idat` files (Illumina) and modify the paths to the raw data files as appropriate.

Vignette	Platform	Annotation package	Scope
Infrastructure	Affy/Illumina		The CNSet container / large data support using the <code>ff</code> package
AffymetrixPreprocessCN	Affy 5.0, 6.0	genomewidesnp5Crlmm, genomewidesnp6Crlmm	Preprocessing and genotyping
IlluminaPreprocessCN	Illumina	several [†]	Preprocessing and genotyping
copynumber	Affy/Illumina	N/A	raw copy number estimates

Table 1: Vignettes for copy number estimation. [†] Annotation packages available for the Illumina platform include `human370v1cCrlmm`, `human370quadv3cCrlmm`, `human550v3bCrlmm`, `human650v3aCrlmm`, `human610quadv1bCrlmm`, `human660quadv1aCrlmm`, `human1mduov3bCrlmm`, and `humanomni1quadv1bCrlmm`

In general, the workflow is

1. preprocess and genotype the arrays (`AffymetrixPreprocessCN` for Affymetrix and `IlluminaPreprocessCN` vignettes for Illumina)
2. copy number estimation (`copynumber` vignette)

The **Infrastructure** vignette provides additional details on the *CNSet* container used to organize the processed data as well as a brief discussion regarding large data support through the `ff` package.