Overview of vignettes for copy number estimation

Rob Scharpf

March 29, 2011

The workflow for copy number analyses in the crlmm package requires preprocessing and genotyping, followed by estimation of parameters for copy number estimation. Supported platforms are those for which a corresponding annotation package is available (see Tables 1 and 2). 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 crlmm package. HapMap datasets are used to illustrate the workflow and are not provided as part of the crlmm 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			The CNSet container / large data support
AffymetrixPreprocessCN	Affy 5.0, 6.0	genomewidesnp5Crlmm, genomewidesnp6Crlmm	Preprocessing and genotyping
IlluminaPreprocessCN	Illumina platforms	several†	Preprocessing and genotyping
copynumber	Affy/Illumina	N/A	raw copy number estimates
SmoothingRawCN	Affy/Illumina	N/A	smoothing via segmentation or hidden Markov models

Table 1: Vignettes for copy number estimation. † See table 2 for the annotation packages available for the Illumina platform.

human370v1cCrlmm
human370quadv3cCrlmm
human550v3bCrlmm
human650v3aCrlmm
human610quadv1bCrlmm
human660quadv1aCrlmm
human1mduov3bCrlmm
humanomni1quadv1bCrlmm

Table 2: Annotation packages for the Illumina platform.

In general, the workflow is

1. preprocessing and genotyping (`AffymetrixPreprocessCN` or `IlluminaPreprocessCN` vignettes)
2. copy number estimation (`copynumber` vignette)
3. inferring regions of copy number gain and loss (`SmoothingRawCN` vignette)

The `SmoothingRawCN` vignette illustrates one approach for interfacing with packages such as DNAcopy and VanillaICE for identifying regions of copy number gain or loss. 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.