Extending \textit{oligo} with \textit{SNPchip}

Robert Scharpf

April 21, 2009

Introduction

This vignette describes a pipeline for preprocessing and visualizing SNP-level summaries using the packages \textit{oligo} and \textit{SNPchip}. We use a set of unprocessed Affymetrix files (CEL files) available as experimental data packages on Bioconductor. A minimal set of commands to perform pre-processing with \textit{oligo} are provided here, though one should consult the \textit{oligo} vignette for additional information. An object of the processed data is provided with this package to reduce the time of computation – the code chunks for the preprocessing steps are not evaluated in the vignette. An example of using \textit{oligo} to process a batch of 209 Affymetrix 100k CEL files and \textit{VanillaICE} to identify regions of alterations are provided in the \textit{hapmap100k} vignette in the directory \texttt{inst/testing} of the \textit{VanillaICE} package. The \textit{hapmap100k} vignette is not reproducible as it depends on access to the 209 CEL hapmap CEL files that are not provided with the \textit{VanillaICE} package, but may be useful as a guideline when performing your own analyses. Comparable vignettes for \textit{hapmapAffy500k}, \textit{hapmapAffy5.0}, \textit{hapmapAffy6.0}, and Illumina will also be added to \textit{VanillaICE} in the near future.

An approach for estimating copy number using the package \textit{oligo} is not yet available. A simple ad-hoc approach to estimate copy number is to assume that the allele A and B summary statistics from CRLMM are proportional to copy number, but these do not generally produce very reliable estimates. A more careful treatment of copy number in \textit{oligo} is forthcoming – this vignette is largely a placeholder.

1 Creating an instance of \texttt{oligoSnpSet}

The \textit{oligo} vignette creates an instance of \texttt{SnpCallSetPlus}, \texttt{crlmmOut}, from the call to the function \texttt{crlmm}. For purposes of illustration, I subset the object to only include SNPs on chromosome 1. I also took the liberty of adding chromosome and physical position to the \texttt{featureData} slot. This object can be loaded by

\begin{verbatim}
> library(SNPchip)
> data(crlmmOut)
> class(crlmmOut)

[1] "SnpCallSetPlus"
attr(,"package")
[1] "oligoClasses"
\end{verbatim}

The elements in the \texttt{assayData} for instances of \texttt{SnpCallSetPlus} is dependent on the Affymetrix platform.

\begin{verbatim}
> annotation(crlmmOut)

[1] "pd.mapping50k.xba240"
\end{verbatim}
Estimating copy number

Copy number estimates are not currently available in CRLMM. In my experience, ad-hoc approaches for estimating copy number from the CRLMM-processed data have not been that successful.

Combining objects that use different annotation packages

Here we illustrate how one may combine two objects of class SnpCallSetPlus that use different annotation packages: e.g., pd.mapping50k.hind240 and pd.mapping50k.xba240. Following the oligo vignette, I created hind and xba instances of SnpCallSetPlus. The following code is not evaluated due to time constraints.

To combine into one object, simply

```r
callset <- combine(xba, hind)
```
3 Session Information

- R version 2.9.0 (2009-04-17), x86_64-unknown-linux-gnu
- Locale: LC_CTYPE=en_US;LC_NUMERIC=C;LC_TIME=en_US;LC_COLLATE=en_US;LC_MONETARY=C;LC_MESSAGES=en_US;
- Base packages: base, datasets, graphics, grDevices, methods, stats, tools, utils
- Other packages: Biobase 2.4.0, oligoClasses 1.6.0, SNPchip 1.8.0
- Loaded via a namespace (and not attached): Biostrings 2.12.0, IRanges 1.2.0