Extending *oligo* with *SNPchip*

Robert Scharpf

April 2, 2008

Introduction

This vignette describes a pipeline for preprocessing and visualizing SNP-level summaries using the packages *oligo* and *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 *oligo* are provided here, though one should consult the *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 *oligo* to process a batch of 209 Affymetrix 100k CEL files and *VanillaICE* to identify regions of alterations are provided in the `hapmap100k` vignette in the directory `inst/testing` of the *VanillaICE* package. The `hapmap100k` vignette is not reproducible as it depends on access to the 209 CEL hapmap CEL files that are not provided with the *VanillaICE* package, but may be useful as a guideline when performing your own analyses. Comparable vignettes for hapmapAffy500k, hapmapAffy5.0, hapmapAffy6.0, and Illumina will also be added to *VanillaICE* in the near future.
An approach for estimating copy number using the package *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 *oligo* is forthcoming – this vignette is largely a placeholder.

1 Creating an instance of `oligoSnpSet`

The *oligo* vignette creates an instance of `SnpCallSetPlus`, `crlmmOut`, from the call to the function `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 featureData slot. This object can be loaded by

```
> library(SNPchip)
> data(crlmmOut)
> class(crlmmOut)
```

```
[1] "SnpCallSetPlus"
attr(,"package")
[1] "oligoClasses"
```

The elements in the `assayData` for instances of `SnpCallSetPlus` is dependent on the Affymetrix platform.

```
> annotation(crlmmOut)
```

```
[1] "pd.mapping50k.xba240"
```

```
> ls(assayData(crlmmOut))

[1] "antisenseThetaA" "antisenseThetaB" "calls"
[4] "callsConfidence" "senseThetaA"     "senseThetaB"

> callset <- crlmmOut
```

1.1 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.

2 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.

```
> library("oligo")
> library("hapmap100kxba")
> pathCelFiles <- system.file("celFiles", package = "hapmap100kxba")
> fullFilenames <- list.celfiles(path = pathCelFiles,
+     full.names = TRUE)
> aboutSamples <- data.frame(gender = c("female",
+     "female", "male"))
> rownames(aboutSamples) <- basename(fullFilenames)
> aboutVars <- data.frame(labelDescription = "male/female")
> rownames(aboutVars) <- "gender"
> pd <- new("AnnotatedDataFrame", data = aboutSamples,
+     varMetadata = aboutVars)
> xba <- justCRLMM(fullFilenames, phenoData = pd,
+     verbose = FALSE)
> library("hapmap100khind")
> pathCelFiles <- system.file("celFiles", package = "hapmap100khind")
> fullFilenames <- list.celfiles(path = pathCelFiles,
+     full.names = TRUE)
> aboutSamples <- data.frame(gender = c("female",
+     "female", "male"))
> rownames(aboutSamples) <- basename(fullFilenames)
> aboutVars <- data.frame(labelDescription = "male/female")
> rownames(aboutVars) <- "gender"
> pd <- new("AnnotatedDataFrame", data = aboutSamples,
+     varMetadata = aboutVars)
> hind <- justCRLMM(fullFilenames, phenoData = pd,
+     verbose = FALSE)
```

To combine into one object, simply

```
> callset <- combine(xba, hind)
```

3 Session Information

- R version 2.7.0 Under development (unstable) (2008-01-28 r44219), `powerpc-apple-darwin8.11.0`
- Locale: `C`
- Base packages: base, datasets, grDevices, graphics, methods, splines, stats, tools, utils
- Other packages: AnnotationDbi 0.99.23, Biobase 1.99.4, DBI 0.2-3, RSQLite 0.6-4, SNPchip 1.3.21, affxparser 1.9.5, genefilter 1.15.10, oligo 1.3.15, oligoClasses 1.1.18, pd.mapping50k.xba240 0.3.4, preprocessCore 0.99.22, survival 2.32
- Loaded via a namespace (and not attached): annotate 1.15.6