Introduction to the gsplom Package

Andrew D. Yates

March 12, 2014

Contents

1	Background	1
2	How to produce SCAN estimates for Agilent two-color microarrays	3
3	Conclusion	4

1 Background

This vignette describes how to normalize samples with the Single-Channel Array Normalization (SCAN) and Universal exPression Codes (UPC) methods.

```
> library(GEOquery)
> tmpDir = tempdir()
> library(GEOquery)
> getGEOSuppFiles("GSM555237", makeDirectory=FALSE, baseDir=tmpDir)
> celFilePath = file.path(tmpDir, "GSM555237.CEL.gz")
```

To normalize the sample, we invoke the SCAN function. This function requires one mandatory parameter: a path specifying the location of the file to be normalized.

```
> library(SCAN.UPC)
> normalized = SCAN(celFilePath)
```

For convenience, it is also possible to download microarray samples from GEO and normalize them in a single step. To do this, substitute the file path with a GEO identifier.

> normalized = SCAN("GSM555237")

The SCAN function returns an *ExpressionSet* object containing a row for each probeset value. Detailed status information, including the number of iterations required for mathematical convergence of the mixture model, is printed to the console.

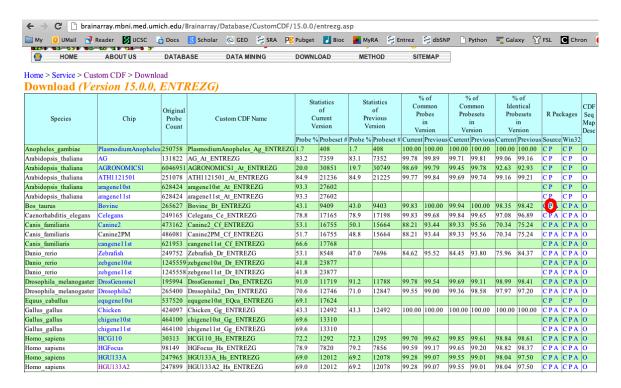
Multiple input files can be processed in one command via specifying wildcard characters (e.g., "*.CEL") or GEO experiment identifiers (e.g., "GSE22309"). In this case, the SCAN function returns an *ExpressionSet* object with a row for each probeset and a column for each input file.

Using the optional outFilePath parameter, the normalized values can be saved to a text file. The example below demonstrates this option.

> normalized = SCAN(celFilePath, outFilePath="output_file.txt")

By default, SCAN maps Affymetrix probes to "probeset" identifiers provided by the manufacturer. However, these mappings may be outdated and include problematic probes (for example, those that may cross hybridize). In addition, multiple probesets may be assigned to a single gene. As an alternative, the BrainArray resource (see http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/genomic_curated_CDF.asp) provides regularly updated mappings that exclude problematic probes and map to genes rather than probesets. When invoking SCAN, users can specify such alternative mappings via the probeSummaryPackage parameter. (Packages other than those provided by BrainArray can be used if they conform to the standards of the AnnotationDbi package.)

BrainArray packages can be downloaded manually from http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download.asp. When downloading, be sure to download the R source package for probe-level mappings (example below).



After such a package has been downloaded, it can be installed using code such as the following.

> install.packages("hgu95ahsentrezgprobe_15.0.0.tar.gz", repos=NULL, type="source")

Or instead of downloading BrainArray packages manually, it is now possible to download these packages via the InstallBrainArrayPackage function.

> pkgName = InstallBrainArrayPackage(celFilePath, "15.0.0", "hs", "entrezg")

These mappings can be applied during normalization using code such as the following.

> normalized = SCAN(celFilePath, probeSummaryPackage=pkgName)

2 How to produce SCAN estimates for Agilent twocolor microarrays

The SCAN.UPC package also supports the ability to normalize Agilent two-color microarrays. The general concept is similar to Affymetrix arrays; however, SCAN also corrects for biases that can arise due to the dyes used in each channel, as well as inter-channel correlation. (This package does not yet support normalizing Agilent one-color arrays.)

3 Conclusion

Please see the SCAN.UPC documentation for full descriptions of functions and the various options they support.