Building a PDInfo Package for an Affymetrix Mapping Chip

Seth Falcon   Benilton Carvalho

November, 2006

1   What you will need

Aside from the **pdInfoBuilder** package and its dependencies, you will need the following files in order to follow along and build a "PDInfo" package for the Affymetrix Mapping 250K NSP chip.

- `Mapping250K_Nsp.cdf`: The binary CDF file from Affymetrix containing feature-level annotation.

- `Mapping250K_Nsp_annot.csv`: Delimited text file from Affymetrix containing featureSet-level annotation.

- `Mapping250K_Nsp_probe_tab`: Delimited text file from Affymetrix containing probe sequence information for all PM probes.

2   Building a PDInfo package

Assuming the source files described above are in your current working directory, the following commands will produce a PDInfo package.

```
R> library("pdInfoBuilder")
R> cdfFile <- "Mapping250K_Nsp.cdf"
R> csvAnno <- "Mapping250K_Nsp_annot.csv"
R> csvSeq <- "Mapping250K_Nsp_probe_tab"
R> pkg <- new("AffySNPPDInfoPkgSeed", version = "0.1.5",
      author = "Seth Falcon", email = "sfalcon@fhcrc.org",
      biocViews = "AnnotationData", genomebuild = "NCBI Build 35, May 2004",
      cdfFile = cdfFile, csvAnnoFile = csvAnno,
      csvSeqFile = csvSeq)
R> makePdInfoPackage(pkg, destDir = ".")
```

We are able to complete the above step on a dual-CPU dual-core 3GHz Xeon box with 8GB RAM in about 20 minutes. The only step that requires a significant amount of RAM is generating the coded sequence matrix. I haven't yet looked carefully at it, but I think it should run on a 4GB system.