# How to use Resourcerer

Jianhua Zhang

October 3, 2007

## 1 Overview

TIGR (`http://pga.tigr.org/tigr-scripts/magic/r1.pl`) offers annotation data files for some commonly used Affymetrix and cDNA chips for various organisms. The annotation data files are available from a ftp site (`ftp://ftp.tigr.org/pub/data/tgi/Resourcerer/`. Referred to as Resourcerer hereafter). The *Resourcerer* allows users to either read an annotation file from Resourcerer or build an annotation data package as the ones that are available in the MetaData section of Bioconductor. This vignette provides instructions on how to do both.

## 2 Getting Started

### 2.1 Requirements

The package *AnnBuilder* is required to perform some of the tasks described below.

```
> require("AnnBuilder", quietly = TRUE)
> require("Resourcerer", quietly = TRUE)
```

### 2.2 Read annotation files from Resourcerer

To read an annotation file from Resourcerer, one may call the function (getResourcerer) by providing the name of the target file to be read and the name of the organism the target file concerns. Please read the help file for (getResourcerer) for more information on the function and the arguments. In our example, we will try to read a file named `Agilent_Human1_cDNA.zip`, which is the annotation data file for a cDNA chip for human located in the human subdirectory of the root ftp site of TIGR Resourcerer.

```
> agilent <- getResourcerer("Agilent_Human1_cDNA.zip", organism = "human",
+     destDir = file.path(.path.package("Resourcerer"), "temp"),
+     baseUrl = "ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Resourcerer",
+     clean = TRUE)
```

The path for destDir tells (getResourcerer) to store the temporary files in the temp subdirectory of (Resourcerer) and clean = TRUE requests that the temporary files be removed when they are no longer needed. The first 4 columns of the top five rows of the file read are:

```
> agilent[1:5, 1:4]
```

```
     Probe ID Clone Name Genbank Acc UniGene ID
[1,] "1"      "523"      "AI422141"  "Hs.532851"
[2,] "10"     "1977"     "AL137410"  "Hs.632554"
[3,] "100"    "67577"    "U92985"    "Hs.515714"
[4,] "1000"   "625374"   "AK026310"  "Hs.465413"
[5,] "10000"  "3257452"  "Y08614"    "Hs.370770"
```

The full column names for `Agilent_Human1_cDNA.zip` are:

```
> as.vector(colnames(agilent))
```

```
 [1] "Probe ID"            "Clone Name"          "Genbank Acc"
 [4] "UniGene ID"          "EntrezGene ID"       "Gene Symbol & Name"
 [7] "Gene Synonyms"       "Human TC"            "Human GC"
[10] "RefSeq Acc"          "TC PubMed Ref"       "GO"
[13] "TGI Annotation"      "Phy Map"             "Genetic Marker"
[16] "Mouse ortholog"      "Rat ortholog"        "Zebrafish ortholog"
[19] "Xenopus ortholog"    "Cattle ortholog"     "Elegans ortholog"
[22] "Yeast ortholog"      "Dog ortholog"        "Chicken ortholog"
```

## 2.3 Build BioC annotation data pacakges based on Resourcerer

To build an annotation data package that is similar to the ones available in the MetaData section of Bioconductor, one may call the function (Resourcerer2BioC). In the following example, we try to build an annotation data package for the same cDNA chip we used before. What the function does is to subtract probe ids and the matching public database ids (GenBank, UniGene, or Locuslink) from the annotation file obtained from Resourcerer and then map the probe ids to annotation data provided by other public data sources such as LocusLink, Gene ontology, KEGG, and so on. As those annotation data are normally very large, we only show the example code without running. One may try the example by coping and pasting the code. It may take up to an hour to have the sample file annotated.

```
> if (interactive()) {
+     resourcerer2BioC("Agilent_Human1_cDNA.zip", organism = "human",
+         destDir = file.path(.path.package("Resourcerer"), "temp"),
```

```
+         pkgName = "AgilentHsa1", srcUrls = getSrcUrl("all", "Homo sapiens"),
+         pkgPath = file.path(.path.package("Resourcerer"), "temp"),
+         otherSrc = NULL, baseMapType = "gb", version = "1.1.0",
+         fromWeb = TRUE, baseUrl = "ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/R
+         check = TRUE, author = list(authors = "Anonymous", maintainer = "Anonymous
+ } else {
+     print("Code is executed only when invoked interactively")
+ }
```

```
[1] "Code is executed only when invoked interactively"
```

When the above code has been executed successfully, there should be a data package named AgilentHsa1 in the temp subdirectory of *Resourcerer*. The package can be used the same way as those in the MetaData section of Bioconductor.