Alternative CDF environments for 2(or more)-genomes chips

Laurent Gautier

October 24, 2023

Introduction

Let's start by loading the package:

```
> library(altcdfenvs)
```

The *Plasmodium* / *Anopheles* is taken as an example:

```
> library(plasmodiumanophelescdf)
```

One will adapt easily the code below for other chips.

How to build a CdfEnvAffy object from the cdfenv package

The first step is to wrap the naked enviroment in the package *plasmodiumanophelescdf* in an object:

```
> planocdf <- wrapCdfEnvAffy(plasmodiumanophelescdf, 712, 712, "plasmodiumanophelescdf"
> print(planocdf)
```

```
Instance of class CdfEnvAffy:
 name     : plasmodiumanophelescdf
 chip-type: plasmodiumanophelescdf
 size     : 712 x 712
 22769 probe set(s) defined.
```

The numbers 712 and 712 correspond to the dimension of the array. If you do not know these numbers for your chip, the easiest (for the moment) is to read CEL data in an `AffyBatch` and call the function `print` on this object.

How to create a CdfEnvAffy that is a subset of the 2-genomes one

If the identifiers starting with 'Pf' correspond to plasmodium, it is an easy job to find them:

```
> ids <- geneNames(planocdf)
> ids.pf <- ids[grep("^Pf", ids)]
```

Subsetting the `CdfEnvAffy` is also an easy task:

```
> ## subset the object to only keep probe sets of interest
> plcdf <- planocdf[ids.pf]
> print(plcdf)
```

```
Instance of class CdfEnvAffy:
 name     : plasmodiumanophelescdf-subsetProbeSets
 chip-type: plasmodiumanophelescdf
 size     : 712 x 712
 4514 probe set(s) defined.
```

However, this is not that simple:**the environment created does not contain all the probe set ids from Plasmodium**. Unfortunately, one cannot rely on pattern matching on the probe set id to find all the probe set ids associated with Plasmodium. The list of plasmodium ids included in the package can let us build a Plasmodium-only CdfEnvAffy (contributed by Zhining Wang).

```
> filename <- system.file("exampleData", "Plasmodium-Probeset-IDs.txt",
+                                 package="altcdfenvs")
> ids.pf <- scan(file = filename, what = "")
> plcdf <- planocdf[ids.pf]
> print(plcdf)
```

```
Instance of class CdfEnvAffy:
 name     : plasmodiumanophelescdf-subsetProbeSets
 chip-type: plasmodiumanophelescdf
 size     : 712 x 712
 5407 probe set(s) defined.
```

Before we eventually save our environment, we may want to give it an explicit name:

```
> plcdf@envName <- "Plasmodium ids only"
> print(plcdf)
```

```
Instance of class CdfEnvAffy:
 name     : Plasmodium ids only
 chip-type: plasmodiumanophelescdf
 size     : 712 x 712
 5407 probe set(s) defined.
```

Assign the new Cdf data to an AffyBatch

Handling of AffyCdfEnv directly in within an AffyBatch, or AffyBatch-like, structure is being completed. . . in the meanwhile, the current mecanism for cdfenvs has to be used.

If your CEL files were read into an AffyBatch named `abatch`.

```
envplcdf <- as(plcdf, "environment")
abatch@cdfName <- "plcdf"
```

From now on, `abatch` will only consider Cdf information from `plcdf`. If you want to save this further use, I would recommend to do:

```
save(abatch, plcdf, envplcdf, file="where/to/save.rda")
```