

learning bioconductor

Eric Yang

learning this book

《Bioinformatics and computational biology solutions using R and Bioconductor》

2. 预处理 High-density Oligonucleotide Arrays

2.1 原理与扫盲

- Affymetrix 就是 HOA，由于此类芯片有多个 probe，因此预处理 (pre-processing) 显得格外重要；一个 gene 用 11-20 个 probe 来检测
- PM: 与目标 gene 的转录子配对，但别的 mRNA 也可能会与其配对
- MM: 目的在于配对与 PM 一样的 gene

2.2 importing probe-level data

- CEL: import the cel files using affy 包, 转化为 AffyBatch 对象

```
library(affy)
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
##
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:stats':
##
##   IQR, mad, xtabs

## The following objects are masked from 'package:base':
##
##   anyDuplicated, append, as.data.frame, cbind, colnames,
##   do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##   grepl, intersect, is.unsorted, lapply, lengths, Map, mapply,
##   match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##   Position, rank, rbind, Reduce, rownames, sapply, setdiff,
##   sort, table, tapply, union, unique, unsplit

## Loading required package: Biobase

## Welcome to Bioconductor
##
##   Vignettes contain introductory material; view with
##   'browseVignettes()'. To cite Bioconductor, see
##   'citation("Biobase")', and for packages 'citation("pkgname")'.
```

- CDF files: 包含信息->probe 属于哪一个 probesets, 以及哪些是 PM 和 MM
- AffyBatch 对象: 一个 s4 object, 包含有如下 slots: ++ cdfName: 一个 character, 包含 CDF file 的信息 ++ nrow, ncol: array 的维度 ++ assayData: 包含 raw data 的 object, 里面有 expression 矩阵 ++ phenoData: 一个 AnnotatedDataFrame object, 用来储存 phenodata 的信息

```
library(affydata)
```

```
##      Package      LibPath                               Item
## [1,] "affydata" "C:/Users/Eric/Documents/R/win-library/3.3" "Dilution"
##      Title
## [1,] "AffyBatch instance Dilution"
```

```
data(Dilution)
```

```
pm(Dilution, '1001_at')[1:3,]
```

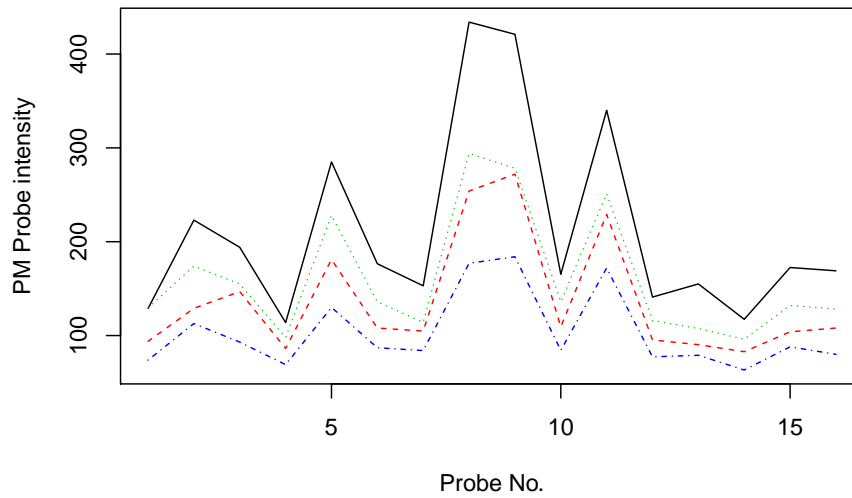
```
## Warning: replacing previous import 'AnnotationDbi::tail' by 'utils::tail'
## when loading 'hgu95av2cdf'
```

```
## Warning: replacing previous import 'AnnotationDbi::head' by 'utils::head'
## when loading 'hgu95av2cdf'
```

```
##
```

```
##           20A   20B   10A   10B
## 1001_at1 128.8  93.8 129.5  73.8
## 1001_at2 223.0 129.0 174.0 112.8
## 1001_at3 194.0 146.8 155.0  93.0
```

```
matplot(pm(Dilution, '1001_at'), type = 'l', xlab = "Probe No.", ylab = "PM Probe inten
```



```
matplot(t(pm(Dilution, '1001_at')), type = 'l', xlab = "Array No.", ylab = "PM Probe in
```

