MAQCsubsetAFX: MAQC reference subset for the Affymetrix platform

Laurent Gatto

March 12, 2008

Contents

1 The MAQC reference datasets

The MAQC (MicroArray Quality Control) project[1] provides a set of reference datasets for a set of 10 platforms (see *Summary of the MAQC Data Sets*[2] for more details). This package provides a subset of the Affymetrix MAQC dataset[3].

Regarding the Affymetrix platform (AFX prefix), a total of 120 Human Genome U133 Plus 2.0 GeneChips have been generated. Four different reference RNAs have been used: (A) 100% of Stratagene's *Universal Human Reference RNA*, (B) 100% of Ambion's Human Brain Reference RNA, (C) 75% of A and 25% of B and (D) 25% of A and 75% of B. Each reference has been repeated 5 times (noted ‿A1‿ to ‿A5‿) on six different test sites (noted ‿1‿ to ‿6‿). As an example, the .CEL result file for the first replicate of test site 2, for the reference ARN C is named `AFX_2_C1.CEL`.

These datasets are freely available and allow, for example, researchers to compare the reproducibility of their own Human Genome U133 Plus 2.0 arrays with a set of high

---

[1] `http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc`

[2] `http://edkb.fda.gov/MAQC/MainStudy/upload/Summary_MAQC_DataSets.pdf`

[3] Packages for the datasets of other platforms will follow and will all be named MAQCsubsetXXX where XXX is the three-letter code used by the MAQC consortium.

quality .CEL files. Nevertheless, using all the 30 available .CEL files (per reference RNA) is memory consuming. As such, we randomly chose 6 .CEL file for each reference RNA, one for each test site as reference to compare the user's data to. These 6 .CEL files are distributed with the *MAQCsubsetAFX* package as 1 data object par reference RNA, respectively called `refA.RData`, `refB.RData`, `refC.RData` and `refD.RData`. These subsets are used by the *yaqcaffy* package to estimate the reproducibility of Human Genome U133 Plus 2.0 with the MAQC subset.

More information concerning the MAQC initiative can be found in the September 2006 special issue of *Nature Biotechnology*.

2   Loading the reference data

Once the library has been installed and loaded, the reference datasets can be loaded using the (data()) function as shown below.

```
> library("MAQCsubsetAFX")
> data(refA)
> refA


AffyBatch (storageMode: lockedEnvironment)
assayData: 1164 features, 1164 samples
  element names: exprs
phenoData
  sampleNames: AFX_1_A2.CEL, AFX_2_A5.CEL, ..., AFX_6_A1.CEL  (6 total)
  varLabels and varMetadata description:
    sample: arbitrary numbering
    site: NA
    ref: NA
    replicate: NA
featureData
  featureNames: 1, 2, ..., 1354896  (1354896 total)
  fvarLabels and fvarMetadata description: none
experimentData: use 'experimentData(object)'
Annotation: hgu133plus2
```