MAQCsubsetILM: MAQC reference subset for the Illumina platform

Laurent Gatto

June 9, 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 Illumina MAQC dataset[3].

Regarding the Illumina platform (ILM prefix), a total of 59 Human-6 BeadChip 48K v1.0 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[4] times (noted _A1_ to _A5_)[5] on three different test sites (noted _1_ to _3_). As an example, the .CEL result file for the first replicate of test site 2, for the reference ARN C is named `ILM_2_C1.CEL`.

[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.

[4] except for site 1,reference C, where 4 replicates are available

[5] the replicates for site 2, reference D are labelled _D1_, _D2_, _D4_, _D6_ and _D7_

These datasets are freely available and allow, for example, researchers to compare the reproducibility of their own Human-6 BeadChip 48K v1.0 data with the MAQC data. *MAQCsubsetILM* offers 3 randomly chosen BeadChips for each reference RNA, one for each test site. Each reference RNA subset is accessible as an R data object, respectively called `refA`, `refB`, `refC` and `refD`.

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("MAQCsubsetILM")
> data(refA)
> refA
```

```
This is mgcv 1.3-31
LumiBatch (storageMode: lockedEnvironment)
assayData: 47293 features, 3 samples
  element names: beadNum, detection, exprs, se.exprs
phenoData
  sampleNames: ILM_1_A5, ILM_2_A1, ILM_3_A2
  varLabels and varMetadata description:
    sampleID: The unique Illumina microarray Id
    site: NA
    ref: NA
    replicate: NA
featureData
  featureNames: GI_10047089-S, GI_10047091-S, ..., trpF  (47293 total)
  fvarLabels and fvarMetadata description:
    TargetID: The Illumina microarray identifier
experimentData: use 'experimentData(object)'
Annotation:
```