

MAQCsubsetILM: MAQC reference subset for the Illumina platform

Laurent Gatto

May 7, 2020

Contents

1	The MAQC reference datasets	1
2	Loading the reference data	2

1 The MAQC reference datasets

The MAQC (MicroArray Quality Control) project¹ 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 dataset³.

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 5⁴ 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/maq>

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

³Packages 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.

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

⁵the 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
```

Summary of data information:

Major Operation History:

```
              submitted              finished
1 2008-02-29 12:24:41 2008-02-29 12:24:43
2 2008-02-29 12:24:43 2008-02-29 12:24:43

              command lumiVersion
1  lumiR("ILM_1_A1.txt", parseColumnName = FALSE)      1.5.17
2 lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)     1.5.17
...

              submitted              finished              command
72 2008-02-29 12:25:22 2008-02-29 12:25:22 combine(x = x.lumi, y = x.lumi.i)
73 2008-02-29 12:27:25 2008-02-29 12:27:25      Subsetting 3 samples.
   lumiVersion
72      1.5.17
73      1.5.17
```

Object Information:

```
LumiBatch (storageMode: lockedEnvironment)
assayData: 47293 features, 3 samples
  element names: beadNum, detection, exprs, se.exprs
```

```
protocolData: none
phenoData
  sampleNames: ILM_1_A5 ILM_2_A1 ILM_3_A2
  varLabels: sampleID site ref replicate
  varMetadata: labelDescription
featureData
  featureNames: GI_10047089-S GI_10047091-S ... trpF (47293 total)
  fvarLabels: TargetID
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:
Control Data: N/A
QC information: Please run summary(x, 'QC') for details!
```