

MAQCsubsetILM: MAQC reference subset for the Illumina platform

Laurent Gatto

September 10, 2009

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")
```

```
This is mgcv 1.5-5 . For overview type `help("mgcv-package")`.
```

```
> 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
3	2008-02-29	12:24:46	2008-02-29	12:24:46
4	2008-02-29	12:24:43	2008-02-29	12:24:45
5	2008-02-29	12:24:45	2008-02-29	12:24:45
6	2008-02-29	12:24:46	2008-02-29	12:24:46
7	2008-02-29	12:24:46	2008-02-29	12:24:46
8	2008-02-29	12:24:48	2008-02-29	12:24:48
9	2008-02-29	12:24:46	2008-02-29	12:24:48
10	2008-02-29	12:24:48	2008-02-29	12:24:48
11	2008-02-29	12:24:48	2008-02-29	12:24:49
12	2008-02-29	12:24:49	2008-02-29	12:24:49
13	2008-02-29	12:24:51	2008-02-29	12:24:51

14 2008-02-29 12:24:49 2008-02-29 12:24:51
15 2008-02-29 12:24:51 2008-02-29 12:24:51
16 2008-02-29 12:24:51 2008-02-29 12:24:51
17 2008-02-29 12:24:51 2008-02-29 12:24:51
18 2008-02-29 12:24:54 2008-02-29 12:24:54
19 2008-02-29 12:24:51 2008-02-29 12:24:54
20 2008-02-29 12:24:54 2008-02-29 12:24:54
21 2008-02-29 12:24:54 2008-02-29 12:24:54
22 2008-02-29 12:24:54 2008-02-29 12:24:54
23 2008-02-29 12:24:57 2008-02-29 12:24:57
24 2008-02-29 12:24:54 2008-02-29 12:24:56
25 2008-02-29 12:24:56 2008-02-29 12:24:56
26 2008-02-29 12:24:57 2008-02-29 12:24:57
27 2008-02-29 12:24:57 2008-02-29 12:24:57
28 2008-02-29 12:24:59 2008-02-29 12:25:00
29 2008-02-29 12:24:57 2008-02-29 12:24:59
30 2008-02-29 12:24:59 2008-02-29 12:24:59
31 2008-02-29 12:25:00 2008-02-29 12:25:00
32 2008-02-29 12:25:00 2008-02-29 12:25:00
33 2008-02-29 12:25:02 2008-02-29 12:25:02
34 2008-02-29 12:25:00 2008-02-29 12:25:02
35 2008-02-29 12:25:02 2008-02-29 12:25:02
36 2008-02-29 12:25:02 2008-02-29 12:25:03
37 2008-02-29 12:25:03 2008-02-29 12:25:03
38 2008-02-29 12:25:05 2008-02-29 12:25:05
39 2008-02-29 12:25:03 2008-02-29 12:25:04
40 2008-02-29 12:25:04 2008-02-29 12:25:05
41 2008-02-29 12:25:05 2008-02-29 12:25:05
42 2008-02-29 12:25:05 2008-02-29 12:25:05
43 2008-02-29 12:25:08 2008-02-29 12:25:08
44 2008-02-29 12:25:05 2008-02-29 12:25:07
45 2008-02-29 12:25:07 2008-02-29 12:25:07
46 2008-02-29 12:25:08 2008-02-29 12:25:08
47 2008-02-29 12:25:08 2008-02-29 12:25:08
48 2008-02-29 12:25:10 2008-02-29 12:25:10
49 2008-02-29 12:25:08 2008-02-29 12:25:10
50 2008-02-29 12:25:10 2008-02-29 12:25:10
51 2008-02-29 12:25:10 2008-02-29 12:25:11
52 2008-02-29 12:25:11 2008-02-29 12:25:11
53 2008-02-29 12:25:13 2008-02-29 12:25:13
54 2008-02-29 12:25:11 2008-02-29 12:25:13
55 2008-02-29 12:25:13 2008-02-29 12:25:13

```

56 2008-02-29 12:25:13 2008-02-29 12:25:13
57 2008-02-29 12:25:13 2008-02-29 12:25:13
58 2008-02-29 12:25:16 2008-02-29 12:25:16
59 2008-02-29 12:25:13 2008-02-29 12:25:15
60 2008-02-29 12:25:15 2008-02-29 12:25:16
61 2008-02-29 12:25:16 2008-02-29 12:25:16
62 2008-02-29 12:25:16 2008-02-29 12:25:16
63 2008-02-29 12:25:19 2008-02-29 12:25:19
64 2008-02-29 12:25:16 2008-02-29 12:25:18
65 2008-02-29 12:25:18 2008-02-29 12:25:18
66 2008-02-29 12:25:19 2008-02-29 12:25:19
67 2008-02-29 12:25:19 2008-02-29 12:25:19
68 2008-02-29 12:25:22 2008-02-29 12:25:22
69 2008-02-29 12:25:19 2008-02-29 12:25:21
70 2008-02-29 12:25:21 2008-02-29 12:25:21
71 2008-02-29 12:25:22 2008-02-29 12:25:22
72 2008-02-29 12:25:22 2008-02-29 12:25:22
73 2008-02-29 12:27:25 2008-02-29 12:27:25

```

```

                                command lumiVersion
1      lumiR("ILM_1_A1.txt", parseColumnName = FALSE)      1.5.17
2  lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)      1.5.17
3      Subsetting 47293 features.      1.5.17
4      lumiR("ILM_1_A2.txt", parseColumnName = FALSE)      1.5.17
5  lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)      1.5.17
6      Subsetting 47293 features.      1.5.17
7      combine(x = x.lumi, y = x.lumi.i)      1.5.17
8      Subsetting 47293 features.      1.5.17
9      lumiR("ILM_1_A3.txt", parseColumnName = FALSE)      1.5.17
10 lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)      1.5.17
11      Subsetting 47293 features.      1.5.17
12      combine(x = x.lumi, y = x.lumi.i)      1.5.17
13      Subsetting 47293 features.      1.5.17
14      lumiR("ILM_1_A4.txt", parseColumnName = FALSE)      1.5.17
15 lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)      1.5.17
16      Subsetting 47293 features.      1.5.17
17      combine(x = x.lumi, y = x.lumi.i)      1.5.17
18      Subsetting 47293 features.      1.5.17
19      lumiR("ILM_1_A5.txt", parseColumnName = FALSE)      1.5.17
20 lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)      1.5.17
21      Subsetting 47293 features.      1.5.17
22      combine(x = x.lumi, y = x.lumi.i)      1.5.17
23      Subsetting 47293 features.      1.5.17

```

24	lumiR("ILM_2_A1.txt", parseColumnName = FALSE)	1.5.17
25	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17
26	Subsetting 47293 features.	1.5.17
27	combine(x = x.lumi, y = x.lumi.i)	1.5.17
28	Subsetting 47293 features.	1.5.17
29	lumiR("ILM_2_A2.txt", parseColumnName = FALSE)	1.5.17
30	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17
31	Subsetting 47293 features.	1.5.17
32	combine(x = x.lumi, y = x.lumi.i)	1.5.17
33	Subsetting 47293 features.	1.5.17
34	lumiR("ILM_2_A3.txt", parseColumnName = FALSE)	1.5.17
35	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17
36	Subsetting 47293 features.	1.5.17
37	combine(x = x.lumi, y = x.lumi.i)	1.5.17
38	Subsetting 47293 features.	1.5.17
39	lumiR("ILM_2_A4.txt", parseColumnName = FALSE)	1.5.17
40	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17
41	Subsetting 47293 features.	1.5.17
42	combine(x = x.lumi, y = x.lumi.i)	1.5.17
43	Subsetting 47293 features.	1.5.17
44	lumiR("ILM_2_A5.txt", parseColumnName = FALSE)	1.5.17
45	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17
46	Subsetting 47293 features.	1.5.17
47	combine(x = x.lumi, y = x.lumi.i)	1.5.17
48	Subsetting 47293 features.	1.5.17
49	lumiR("ILM_3_A1.txt", parseColumnName = FALSE)	1.5.17
50	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17
51	Subsetting 47293 features.	1.5.17
52	combine(x = x.lumi, y = x.lumi.i)	1.5.17
53	Subsetting 47293 features.	1.5.17
54	lumiR("ILM_3_A2.txt", parseColumnName = FALSE)	1.5.17
55	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17
56	Subsetting 47293 features.	1.5.17
57	combine(x = x.lumi, y = x.lumi.i)	1.5.17
58	Subsetting 47293 features.	1.5.17
59	lumiR("ILM_3_A3.txt", parseColumnName = FALSE)	1.5.17
60	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17
61	Subsetting 47293 features.	1.5.17
62	combine(x = x.lumi, y = x.lumi.i)	1.5.17
63	Subsetting 47293 features.	1.5.17
64	lumiR("ILM_3_A4.txt", parseColumnName = FALSE)	1.5.17
65	lumiQ(x.lumi = x.lumi, detectionTh = detectionTh)	1.5.17

```

66             Subsetting 47293 features.           1.5.17
67         combine(x = x.lumi, y = x.lumi.i)       1.5.17
68             Subsetting 47293 features.           1.5.17
69     lumiR("ILM_3_A5.txt", parseColumnName = FALSE) 1.5.17
70 lumiQ(x.lumi = x.lumi, detectionTh = detectionTh) 1.5.17
71             Subsetting 47293 features.           1.5.17
72         combine(x = x.lumi, y = x.lumi.i)       1.5.17
73             Subsetting 3 samples.                 1.5.17

```

Object Information:

```

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:
Control Data: Available
QC information: Please run summary(x, 'QC') for details!

```