

The AffyRNADegradation Package

Mario Fasold

March 31, 2012

Affymetrix 3' expression arrays employ a specific experimental protocol and a specific probe design that allows assessment of RNA integrity based on probe signal data. Problems of RNA integrity are primarily governed to the degradation of the target transcripts. It can be shown that

- (i) degradation leads to a probe positional bias that needs to be corrected in order to compare expression of samples with varying degree of degradation, and
- (ii) it is possible to estimate a robust and accurate measure of RNA integrity from the probe signals that, for example, can be used to study degradation within the large number of available microarray data.

The rationale and further analysis are described in the accompanying publication by Fasold and Binder. We here show how to utilize this package for both problems.

1 Basic RNA Degradation Analysis

We here show how to use the package for the analysis of RNA degradation. Let us first load exemplar data provided by the *AmpAffyExample* package into the environment.

```
> library(AffyRNADegradation)
> library(AmpAffyExample)
> data(AmpData)
> AmpData
```

```
AffyBatch object
size of arrays=712x712 features (12 kb)
cdf=HG-U133A (22283 affyids)
number of samples=6
number of genes=22283
annotation=hgu133a
notes=
```



Figure 1: The tongs plot shows that the intensity difference between 3' and 5' probes increases with $\Sigma = \langle \log I \rangle$. $\langle \rangle$ here denotes either averaging over all probes within the probeset, or averaging over the 3' or 5' subset of probes in Σ_{subset} .

Every transcript is measured by a set of 11-16 probes. The log-average intensity difference between probes located closer to the 3' end of the target transcripts and those located further away constitutes the probe positional bias. It can be visualized using the *tongs plot*.

```
> tongs <- GetTongs(AmpData, chip.idx = 4)
> PlotTongs(tongs)
```

Figure 1 shows that the bias relates to the expression level of the transcripts. As this can vary from sample to sample, it must be considered in estimating of RNA degradation.

The function `RNAdegradation` performs the basal analysis of RNA degradation based on raw probe intensities stored in an `AffyBatch` object. The result is an `AffyDegradationBatch` object that contains the corrected probe intensities as well as several statistical parameters.

```
> rna.deg <- RNAdegradation(AmpData, location.type = "index")
```



Figure 2: Probe degradation plot. The points show the average probe intensity of expressed genes for each index $x = 1, \dots, 11$ relative to the average intensity at position $x = 1$. The lines are a fitted decay function.

We can visualize the probe positional bias using the `PlotDx` function.

```
> plotDx(rna.deg)
```

Figure 2 shows the results. Different degradation between different samples are observed.

To access the parameter d , which provides a robust, sample-wise measure for the degree of RNA degradation, one can use the function

```
> d(rna.deg)
```

1_Std_TB_75-6	2_Std_TB_75-6	5_Std_TB_75-6	13r_Amp_TB_75-6	14_Amp_TB_75-6
0.6161481	0.6339041	0.6129885	0.3804619	0.4462438
17_Amp_TB_75-6				
0.3710109				

2 Absolute Probe Locations

Instead of using the probe index within the probeset as argument of the degradation degree, one can use the actual probe locations within the transcript. We have pre-computed the distance of each probe to the 3' end of its target transcript for all Affymetrix 3' expression arrays. These probe location files are available under the URL http://www.izbi.uni-leipzig.de/downloads_links/programs/rna_integrity.php.

In order to perform the analysis and correction using absolute probe locations, one must first download the probe location file for the used chip type. You can then start the analysis using `RNADegradation`, as above, but selecting `absolute` as `location.type`. The parameter `location.file.dir` must specify the download directory of the probe location file.

3 Correction of the Bias and Integration into the Microarray Calibration Process

The correction of the probe positional bias is performed within the `AffyRNA-Degradation` function. The result is a new `AffyBatch` object with corrected probe level intensities. It can be accessed using the `afbatch` function

```
> afbatch(rna.deg)
```

It is possible to replace the original raw data with this data corrected for probe positional bias, before performing further microarray normalization and summarization (e.g. using RMA).

Alternatively, correction can be performed after probe-level normalization in a microarray calibration pipeline. The following example shows how to first apply the VSN normalization method, then correct for probe positional bias to finally get summarized expression measures

```
> library(vsn)
> affydata.vsn <- do.call(affy::normalize, c(alist(AmpData, "vsn"), NULL))
> affydata.vsn <- afbatch(RNADegradation(affydata.vsn))
> expr <- computeExprSet(affydata.vsn, summary.method="medianpolish", pmcorrect.method="pmor")
```

4 Details

This document was written using:

```
> sessionInfo()
```

```
R version 2.15.0 (2012-03-30)
Platform: i386-pc-mingw32/i386 (32-bit)
```

```
locale:
```

```
[1] LC_COLLATE=C
[2] LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.1252
```

attached base packages:

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

other attached packages:

```
[1] hgu133acdf_2.10.0      AnnotationDbi_1.18.0      AmpAffyExample_1.2.10
[4] AffyRNADegradation_1.0.0 affy_1.34.0               Biobase_2.16.0
[7] BiocGenerics_0.2.0
```

loaded via a namespace (and not attached):

```
[1] BiocInstaller_1.4.0    DBI_0.2-5                IRanges_1.14.1
[4] RSQLite_0.11.1         affyio_1.24.0            preprocessCore_1.18.0
[7] stats4_2.15.0          tools_2.15.0             zlibbioc_1.2.0
```