AffyProbePositionBias Tutorial

Mario Fasold, Hans Binder

November 30, 2011

This file gives a step-by-step tutorial of how to use the accompanying R scripts to correct a data set of Affymetrix Microarray expression data for RNA Degradation effects. The rationale and further analysis are described in the paper by Fasold and Binder.

1 Basic Correction

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(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=
```

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*.

```
> source("DegradationAnalysis.R")
> 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.

Correcting the expression data practically is a single function applied to the expression set data structure.

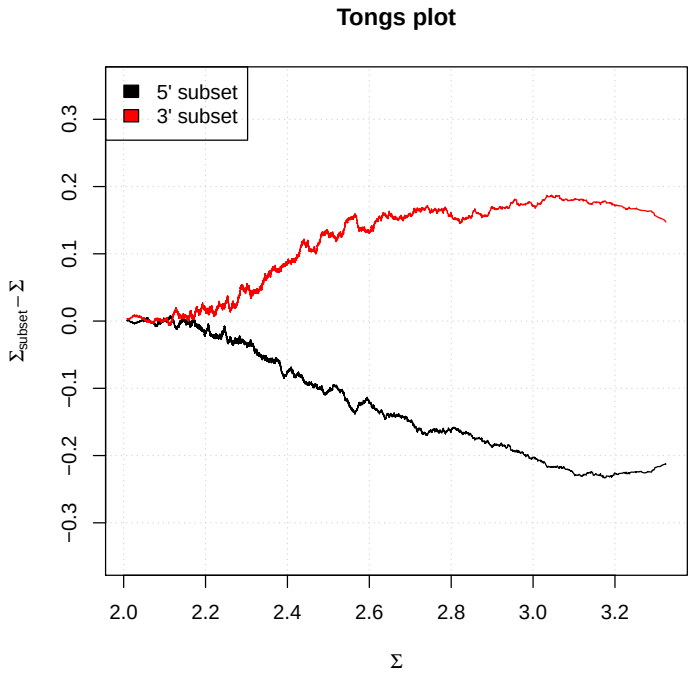

Figure 1: The tongs plot shows that the intensity difference between 3' and 5' probes increases with $\Sigma = \langle \log I \rangle$. $\langle\rangle$ here denotes either averaging over all probes within the probeset, or averaging over the 3' or 5' subset of probes in Σ_{subset}.

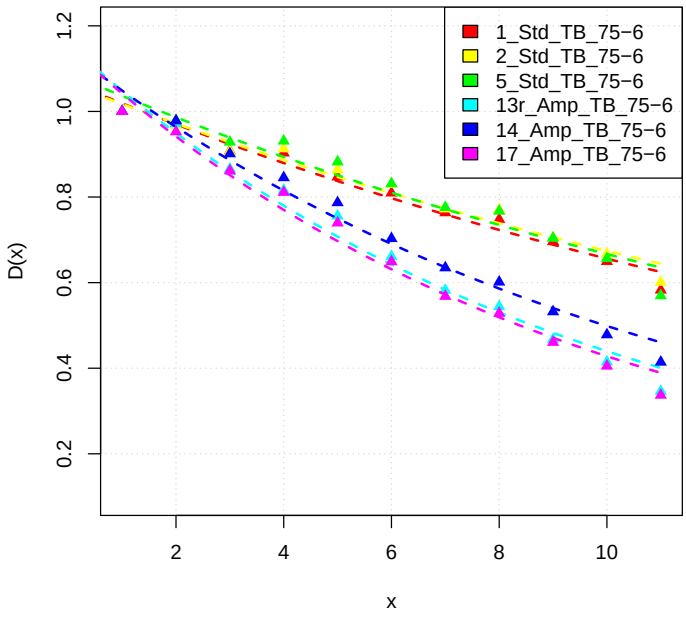

Figure 2: Probe degradation plot. The points show the average probe intensity of expressed genes for each index $x = 1, ..11$ relative to the average intensity at position $x = 1$. The lines are a fitted decay function.

```
> affydataIndexCorrect <- DegradationAnalysis(AmpData, location.type = "index")
```

The resulting data structure contains the corrected data as expression set as well as some statistics for either specific binding (*.S) or non-specific binding (*.Ns).

The matrices mean.pm and mean.mm contain the average probe intensity for each index $k = 1, ..11$ relative to the to the average intensity at position $k = 1$. The parameter decay provides a good measure for the degree of degradation of that particular chip.

You can plot the Degradation using the following command

```
> plotDegradation(affydataIndexCorrect)
```

Figure 2 shows the results.

2 Integration into the Data Calibration Process

Since degradation is a biological effect, but no technical, it should ideally be applied at the end of a microarray correction pipeline. The following example shows how to first apply the VSN normalization method, then correct for probe position bias to finally get summarized expression measures

```
> library(vsn)
> normalize.method = "vsn"; pmcorrect.method = "pmonly"; summary.method = "medianpolish"
> affydata.vsn <- do.call(affy:::normalize, c(alist(AmpData, normalize.method), NULL))
> affydata.vsn <- DegradationAnalysis(AmpData)
> em <- computeExprSet(affydata.vsn$afbatch, summary.method=summary.method, pmcorrect.method

22283 ids to be processed
|                    |
|####################|
```

Please note that many approaches such as RMA perform boisterous normalizations that cover the degradation effect. We suggest to use a single-chip correction method such as HOOK.