

# Description of affyILM package

K. Myriam Kroll, Fabrice Berger, Gerard Barkema, Enrico Carlon

April 25, 2023

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Getting started</b>	<b>1</b>
2.1	Preliminaries . . . . .	1
2.2	First Steps . . . . .	2
<b>3</b>	<b>More Examples with options</b>	<b>3</b>

## 1 Introduction

*affyILM* is a preprocessing tool which estimates gene expression levels for Affymetrix Gene Expression Chips. *affyILM* computes gene expression levels using the Langmuir model. In contrast to other measures, this method outputs the gene expression level as concentrations measured in  $pM$  (picoMolar).

*affyILM* allows the user to simultaneously read-in several CEL-files; it does *not* require raw data (CEL-files) to be specifically formatted like e.g. as *AffyBatch*.

## 2 Getting started

### 2.1 Preliminaries

To install the package:

```
R CMD INSTALL affyILM_x.y.z.tar.gz
```

*affyILM* imports several functions from other packages. Make sure to have the following installed:

*affxparser*, *affy Biobase* and *germa*. Chip-specific probe packages which are not yet installed on your system will be automatically downloaded from the bioconductor webpage if needed.

## 2.2 First Steps

For demonstration purposes we use a test CEL-file supplied by *AffymetrixDataTestFiles*.

```
> require(AffymetrixDataTestFiles)
```

Load the library

```
> library(affyILM)
```

and locate the test CEL-file

```
> path <- system.file("rawData", "FusionSDK_HG-Focus", "HG-Focus", "2.Calvin",  
+ package="AffymetrixDataTestFiles")  
> file1 <- file.path(path, "HG-Focus-1-121502.CEL")
```

Calculation of the hybridization free energies for each probe, and estimation of concentrations using the Langmuir isotherm:

```
> result <- ilm(file1);
```

Chip dimension 448 x 448

```
[1] "Checking to see if your internet connection works..."
```

Probepackage hgfocusprobe loaded

Reading intensities...[1] ...done

Now let's have a look at the output printed on the screen:

- Chip dimension
- probe package downloaded if missing

Take a look at the experimental PM's

```
> getIntens(result, "AFFX-r2-Ec-bioD-5_at")
```

	HG-Focus-1-121502.CEL	HG-Focus-1-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	644.5	0
ps.AFFX-r2-Ec-bioD-5_at.id	694.3	0
ps.AFFX-r2-Ec-bioD-5_at.id	602.5	0
ps.AFFX-r2-Ec-bioD-5_at.id	1384.0	0
ps.AFFX-r2-Ec-bioD-5_at.id	1329.8	0
ps.AFFX-r2-Ec-bioD-5_at.id	1410.8	0
ps.AFFX-r2-Ec-bioD-5_at.id	1404.0	0
ps.AFFX-r2-Ec-bioD-5_at.id	3265.8	0
ps.AFFX-r2-Ec-bioD-5_at.id	2616.3	0
ps.AFFX-r2-Ec-bioD-5_at.id	1045.5	0
ps.AFFX-r2-Ec-bioD-5_at.id	628.3	0

Plot the result:

```
> plotIntens(result, "AFFX-r2-Ec-bioD-5_at", "HG-Focus-1-121502.CEL")
```

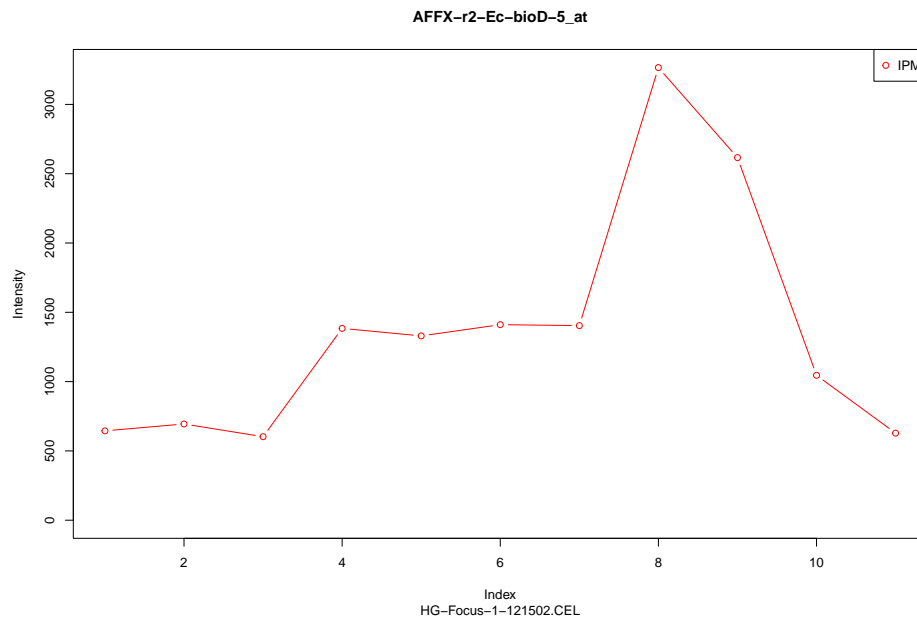


Figure 1: Probes intensities

### 3 More Examples with options

Analyze two or more CEL-files

```
> file2 <- file.path(path, "HG-Focus-2-121502.CEL")
> result2files <- ilm(c(file1, file2), satLim=12000)
```

```
Chip dimension 448 x 448
Probepackage hgfocusprobe loaded
Reading intensities...[1] ...done
```

where the saturation limit of the Langmuir isotherm is increased to 12000 (default: 10000)

Get intensity values:

```
> getIntens(result2files, "AFFX-r2-Ec-bioD-5_at")
```

	HG-Focus-1-121502.CEL	HG-Focus-2-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	644.5	692.3
ps.AFFX-r2-Ec-bioD-5_at.id	694.3	809.5
ps.AFFX-r2-Ec-bioD-5_at.id	602.5	687.3

ps.AFFX-r2-Ec-bioD-5_at.id	1384.0	1708.5
ps.AFFX-r2-Ec-bioD-5_at.id	1329.8	1513.5
ps.AFFX-r2-Ec-bioD-5_at.id	1410.8	1643.8
ps.AFFX-r2-Ec-bioD-5_at.id	1404.0	1838.0
ps.AFFX-r2-Ec-bioD-5_at.id	3265.8	3985.3
ps.AFFX-r2-Ec-bioD-5_at.id	2616.3	3331.0
ps.AFFX-r2-Ec-bioD-5_at.id	1045.5	1102.0
ps.AFFX-r2-Ec-bioD-5_at.id	628.3	746.0
	HG-Focus-1-121502.CEL	HG-Focus-2-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0

- 1st column: Probeset name
- 2nd and 3rd column: Measured PM intensities IPM of each file.
- 4th and 5th column: IO intensities of each file (default value is 0 in current release, no background estimation).

To obtain the probe concentrations (or expression levels), use

```
> getProbeConcs(result2files, "AFFX-r2-Ec-bioD-5_at")
```

	HG-Focus-1-121502.CEL	HG-Focus-2-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	265.98771	286.9227
ps.AFFX-r2-Ec-bioD-5_at.id	225.46838	265.5849
ps.AFFX-r2-Ec-bioD-5_at.id	587.82734	675.5887
ps.AFFX-r2-Ec-bioD-5_at.id	511.07926	650.8028
ps.AFFX-r2-Ec-bioD-5_at.id	255.39839	295.7714
ps.AFFX-r2-Ec-bioD-5_at.id	149.64574	178.2833
ps.AFFX-r2-Ec-bioD-5_at.id	132.74302	181.1978
ps.AFFX-r2-Ec-bioD-5_at.id	447.36533	594.9351
ps.AFFX-r2-Ec-bioD-5_at.id	131.99693	181.9098
ps.AFFX-r2-Ec-bioD-5_at.id	352.33599	373.3020
ps.AFFX-r2-Ec-bioD-5_at.id	94.30074	113.1372

Use | to subset the results on one or more probesets

```
> res_1 <- result["AFFX-r2-Ec-bioD-5_at"]
> res_1
```

	HG-Focus-1-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	644.5
ps.AFFX-r2-Ec-bioD-5_at.id	694.3
ps.AFFX-r2-Ec-bioD-5_at.id	602.5
ps.AFFX-r2-Ec-bioD-5_at.id	1384.0
ps.AFFX-r2-Ec-bioD-5_at.id	1329.8
ps.AFFX-r2-Ec-bioD-5_at.id	1410.8
ps.AFFX-r2-Ec-bioD-5_at.id	1404.0
ps.AFFX-r2-Ec-bioD-5_at.id	3265.8
ps.AFFX-r2-Ec-bioD-5_at.id	2616.3
ps.AFFX-r2-Ec-bioD-5_at.id	1045.5
ps.AFFX-r2-Ec-bioD-5_at.id	628.3

	HG-Focus-1-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
[1] 10000	

```
> res_2 <- result[c("AFFX-r2-Ec-bioD-5_at", "207218_at")]
> res_2
```

	HG-Focus-1-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	644.5
ps.AFFX-r2-Ec-bioD-5_at.id	694.3
ps.AFFX-r2-Ec-bioD-5_at.id	602.5
ps.AFFX-r2-Ec-bioD-5_at.id	1384.0
ps.AFFX-r2-Ec-bioD-5_at.id	1329.8
ps.AFFX-r2-Ec-bioD-5_at.id	1410.8
ps.AFFX-r2-Ec-bioD-5_at.id	1404.0
ps.AFFX-r2-Ec-bioD-5_at.id	3265.8

ps.AFFX-r2-Ec-bioD-5_at.id	2616.3
ps.AFFX-r2-Ec-bioD-5_at.id	1045.5
ps.AFFX-r2-Ec-bioD-5_at.id	628.3
ps.207218_at.id	63.8
ps.207218_at.id	62.5
ps.207218_at.id	83.5
ps.207218_at.id	97.0
ps.207218_at.id	83.8
ps.207218_at.id	191.5
ps.207218_at.id	60.8
ps.207218_at.id	73.5
ps.207218_at.id	133.0
ps.207218_at.id	58.8
ps.207218_at.id	82.5

HG-Focus-1-121502.CEL

ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.AFFX-r2-Ec-bioD-5_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0
ps.207218_at.id	0

[1] 10000

and/or on one or more files:

```
> res2_1 <- result2files["AFFX-r2-Ec-bioD-5_at"]
> res2_1
```

	HG-Focus-1-121502.CEL	HG-Focus-2-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	644.5	692.3
ps.AFFX-r2-Ec-bioD-5_at.id	694.3	809.5
ps.AFFX-r2-Ec-bioD-5_at.id	602.5	687.3
ps.AFFX-r2-Ec-bioD-5_at.id	1384.0	1708.5
ps.AFFX-r2-Ec-bioD-5_at.id	1329.8	1513.5
ps.AFFX-r2-Ec-bioD-5_at.id	1410.8	1643.8
ps.AFFX-r2-Ec-bioD-5_at.id	1404.0	1838.0
ps.AFFX-r2-Ec-bioD-5_at.id	3265.8	3985.3
ps.AFFX-r2-Ec-bioD-5_at.id	2616.3	3331.0
ps.AFFX-r2-Ec-bioD-5_at.id	1045.5	1102.0
ps.AFFX-r2-Ec-bioD-5_at.id	628.3	746.0

	HG-Focus-1-121502.CEL	HG-Focus-2-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0

[1] 12000

```
> res2_2 <- result2files[c("AFFX-r2-Ec-bioD-5_at", "207218_at")]
> res2_2
```

	HG-Focus-1-121502.CEL	HG-Focus-2-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	644.5	692.3
ps.AFFX-r2-Ec-bioD-5_at.id	694.3	809.5
ps.AFFX-r2-Ec-bioD-5_at.id	602.5	687.3
ps.AFFX-r2-Ec-bioD-5_at.id	1384.0	1708.5
ps.AFFX-r2-Ec-bioD-5_at.id	1329.8	1513.5
ps.AFFX-r2-Ec-bioD-5_at.id	1410.8	1643.8
ps.AFFX-r2-Ec-bioD-5_at.id	1404.0	1838.0
ps.AFFX-r2-Ec-bioD-5_at.id	3265.8	3985.3
ps.AFFX-r2-Ec-bioD-5_at.id	2616.3	3331.0
ps.AFFX-r2-Ec-bioD-5_at.id	1045.5	1102.0
ps.AFFX-r2-Ec-bioD-5_at.id	628.3	746.0
ps.207218_at.id	63.8	66.3
ps.207218_at.id	62.5	66.0

ps.207218_at.id	83.5	105.3
ps.207218_at.id	97.0	109.5
ps.207218_at.id	83.8	98.5
ps.207218_at.id	191.5	240.3
ps.207218_at.id	60.8	77.5
ps.207218_at.id	73.5	102.5
ps.207218_at.id	133.0	147.8
ps.207218_at.id	58.8	63.0
ps.207218_at.id	82.5	85.3
	HG-Focus-1-121502.CEL	HG-Focus-2-121502.CEL
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.AFFX-r2-Ec-bioD-5_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
ps.207218_at.id	0	0
[1] 12000		

The output objects are of class `ILM`.



Plot the Langmuir Isotherm :

```
> pILM<-plotILM(result2files,"AFFX-r2-Ec-bioD-5_at","HG-Focus-1-121502.CEL")
```

Median = 255

M.a.d. = 181.85

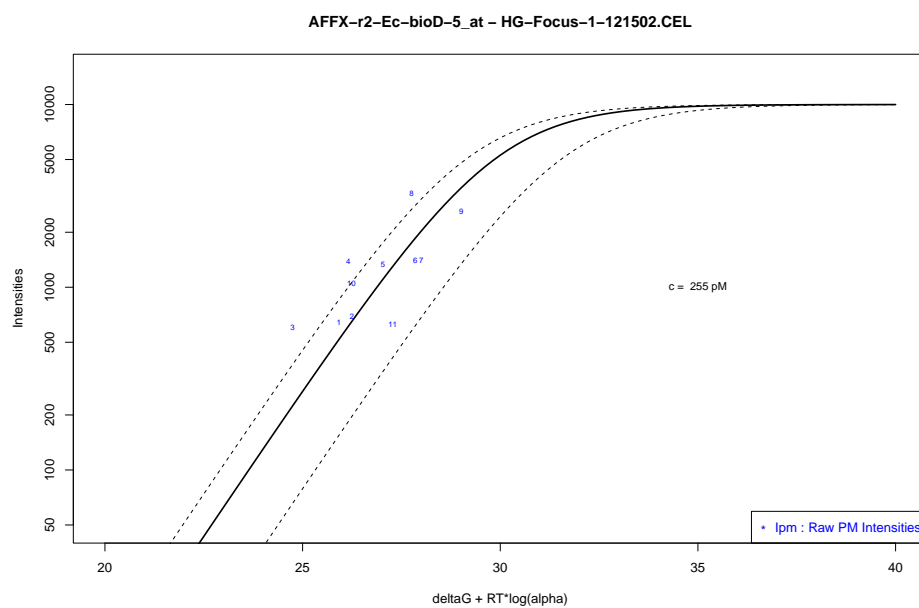


Figure 2: Illustration of the Langmuir Isotherm

This function also provides a list with computed values:

```
> print(str(pILM))
```

List of 8

```
$ Ipm          : Named num [1:11] 644 694 602 1384 1330 ...  
..- attr(*, "names")= chr [1:11] "ps.AFFX-r2-Ec-bioD-5_at.id" "ps.AFFX-r2-Ec-bioD-5_a  
$ I0pm         : Named num [1:11] 0 0 0 0 0 0 0 0 0 0 ...  
..- attr(*, "names")= chr [1:11] "ps.AFFX-r2-Ec-bioD-5_at.id" "ps.AFFX-r2-Ec-bioD-5_a  
$ ImI0         : Named num [1:11] 644 694 602 1384 1330 ...  
..- attr(*, "names")= chr [1:11] "ps.AFFX-r2-Ec-bioD-5_at.id" "ps.AFFX-r2-Ec-bioD-5_a  
$ deltaG       : Named num [1:11] 39.1 37 34.1 33.7 32.8 ...  
..- attr(*, "names")= chr [1:11] "ps.AFFX-r2-Ec-bioD-5_at.id" "ps.AFFX-r2-Ec-bioD-5_a  
$ deltaGp      : Named num [1:11] 60.7 58.1 56.6 54.6 52.7 ...  
..- attr(*, "names")= chr [1:11] "ps.AFFX-r2-Ec-bioD-5_at.id" "ps.AFFX-r2-Ec-bioD-5_a  
$ alpha        : Named num [1:11] 5.98e-05 3.60e-04 9.62e-04 3.80e-03 1.41e-02 ...  
..- attr(*, "names")= chr [1:11] "ps.AFFX-r2-Ec-bioD-5_at.id" "ps.AFFX-r2-Ec-bioD-5_a  
$ deltaGpRTlogA: Named num [1:11] 25.9 26.2 24.7 26.2 27 ...  
..- attr(*, "names")= chr [1:11] "ps.AFFX-r2-Ec-bioD-5_at.id" "ps.AFFX-r2-Ec-bioD-5_a  
$ Concs        : Named num [1:11] 266 225 588 511 255 ...  
..- attr(*, "names")= chr [1:11] "ps.AFFX-r2-Ec-bioD-5_at.id" "ps.AFFX-r2-Ec-bioD-5_a
```

NULL

## References

- K. M. Kroll, G. T. Barkema, and E. Carlon. Modeling background intensity in DNA microarrays. *Phys. Rev. E*, 77:061915, 2008.
- K. M. Kroll, G. T. Barkema, and E. Carlon. Linear model for fast background subtraction in oligonucleotide microarrays. *Algorithms for Molecular Biology*, 4:15, 2009.
- G. Mulders, G.T. Barkema, and E. Carlon. Inverse langmuir method for oligonucleotide microarray analysis. *BMC Bioinformatics*, 10:64, 2009.
- N. Sugimoto, S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, and M. Sasaki. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, 34:11211, 1995.