# oneChannelGUI Package Vignette

Raffaele A Calogero, Francesca Cordero, Remo Sanges

May 12 2011

## 1 Introduction

This package is an add-on of affylmGUI for *mouse-click* based QC, statistical analysis and data mining for one channel microarray data. It is designed for Bioconductor beginners having limited or no experience in interacting with Bioconductor line commands. OneChannelGUI is a set of functions extending the affylmGUI capabilities, rearranging and extending affylmGUI menus.

This package performs, in a graphical environment, the analysis pipe-line shown in figure 1, green box.

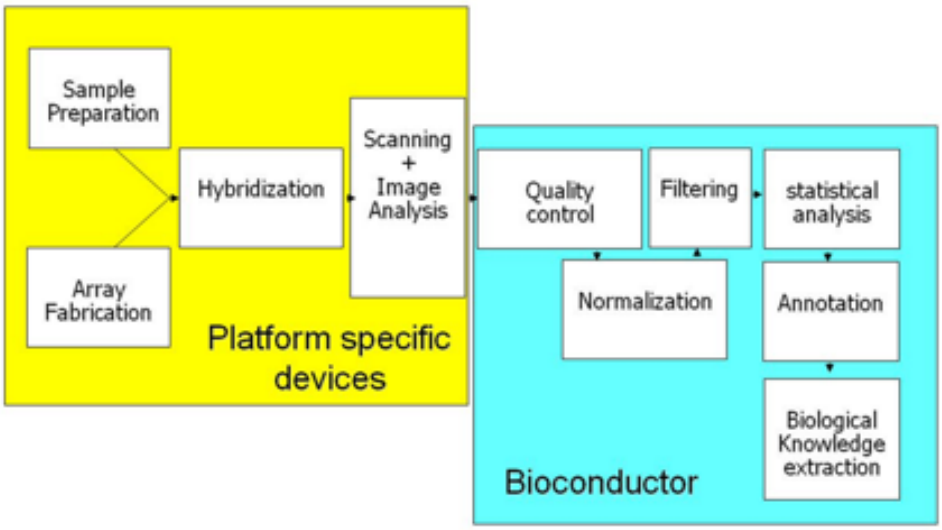



Figure 1: Microarray analysis pipe-line.

This vignette gives a general overview of the graphical interfaces available in oneChannelGUI for exn-level analysis using Affymetrix Exon 1.0 ST arrays.

*N.B:*

*All the oneChannelGUI graphical outputs are visualized in the R main window, to reduce RAM consumption, which is a critical issue when Affymetrix array data or large set of data are loaded. Furthermore, exon data generated with APT tools produce, in the working directory, a certain amount of temporary files and directories. A cleanup function is under development. At the present time, user can manually remove, from the working folder, any file starting with target, elevels, glevels, e.g. target51f81aeb, elevels3e9f6b76, and folders starting with out and outMidas, e.g. out17fb164, outMidas4a31ac4, without affecting the results stored in oneChannelGUI.*

# 2 Installation

For the complete functionality of oneChannelGUI some external softwares and data need to be installed. Please refer to the *install vignette* of oneChannelGUI package.

# 3 Main graphical window

oneChannelGUI inherits the core functionalities of affylmGUI and its main GUI. In oneChannelGUI some extra topics are available in the main affylmGUI info left frame, e.g. maSigPro results, Normalized Exon data, APT DABG, APT MiDAS, Splice Index, etc. Furthermore, four different menus are automatically exchanged depending on the type of array loaded:

1. .CEL IVT Affymetrix arrays.
2. .CEL exon 1.0 ST arrays uploaded in oneChannelGUI by Affymetrix APT tools or gene/exon data exported from Affymetrix Expression Console.
3. .CEL Gene 1.0 ST arrays uploaded in oneChannelGUI by Affymetrix APT tools.
4. GEO/flat tab delimited expression data file.
5. ILLUMINA output from BeadStudio software version 1, 2 and 3.

Each item in the menus is simply a graphical implementation of a function of a specific Bioconductor library , e.g. ssize: sample size and statistical power estimation. To get more information on those libraries please refer to their specific vignettes, accessible from the *Help menu*.

# 4 File

This menu allows the loading of .CEL IVT Affymetrix arrays as well as exon arrays, GEO Matrix Series files, tab delimited files containing only expression data and ILLUMINA data produced by BeadStudio software version 1 or 2. In this menu, fig. 2, are given the main functionalities to handle a microarray analysis project.

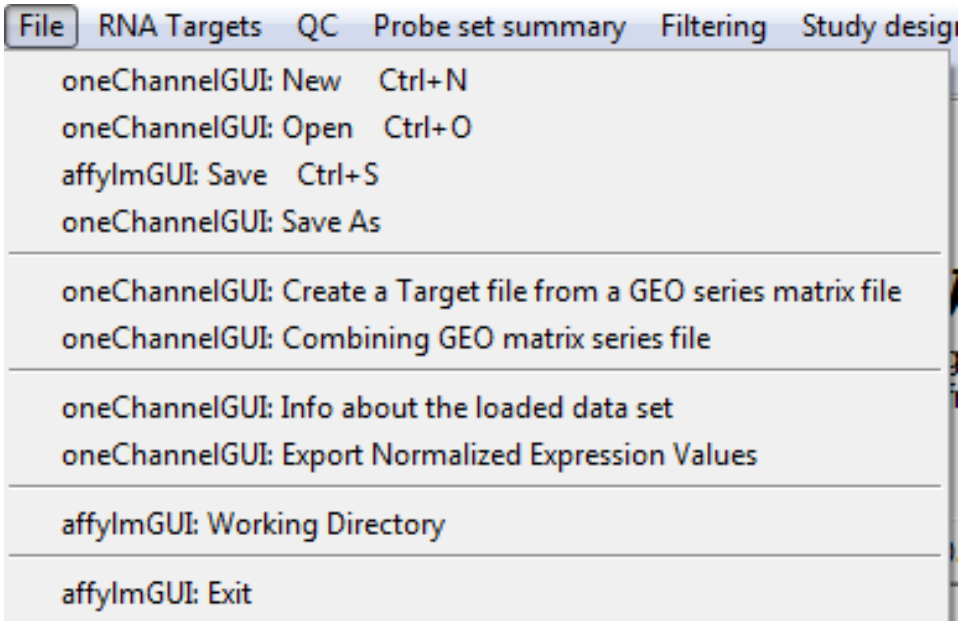



Figure 2: File menu.

## 4.1 New

The item *New*, fig. 2, allow to load various types of array data, using the sub menu shown in fig. 3,

### 4.1.1 Target file structure

To load arrays oneChannelGUI uses the information available in a file describing the experimental structure of the data set. This file is called *target file* and it is a tab delimited file with a fixed header structure also used by affylmGUI, fig. 4.

*IMPORTANT:*
*TARGET FILE MUST NOT CONTAIN CHARACTERS LIKE ;,:_-|\!?+*^()[]{}*

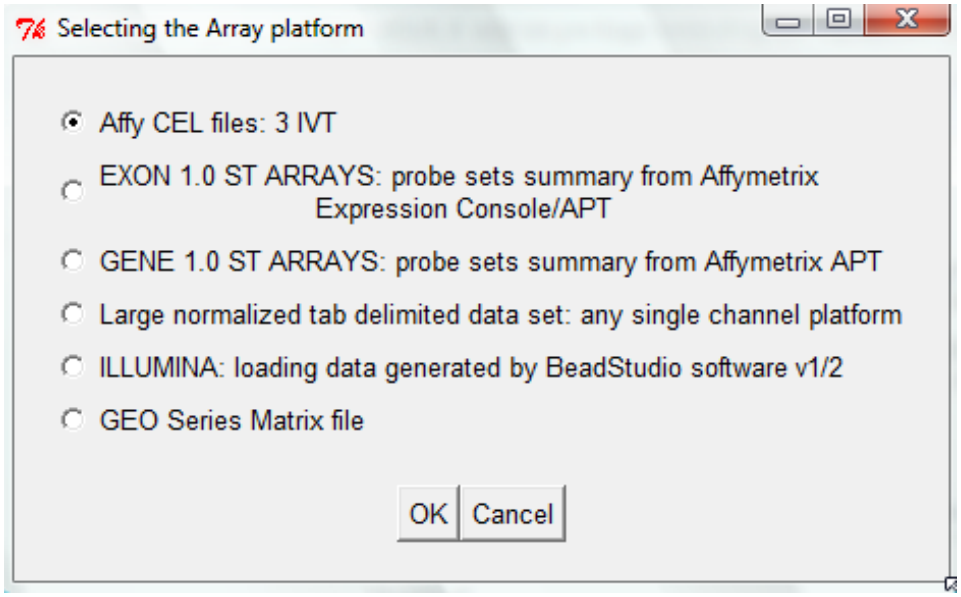
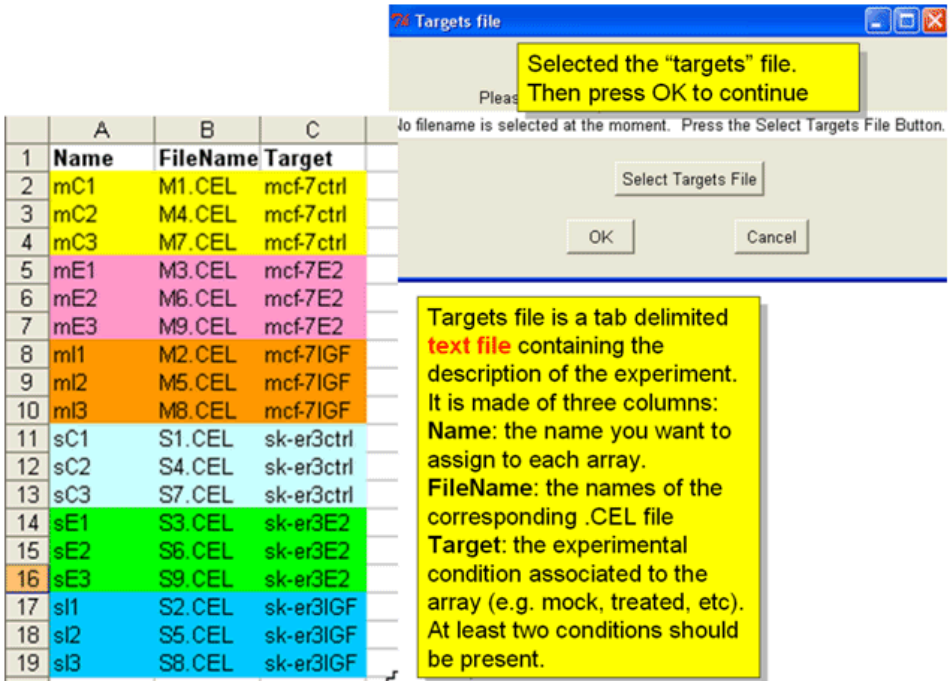



Figure 3: New: array type selection menu.




| | A | B | C |
|---|---|---|---|
| 1 | **Name** | **FileName** | **Target** |
| 2 | mC1 | M1.CEL | mcf-7ctrl |
| 3 | mC2 | M4.CEL | mcf-7ctrl |
| 4 | mC3 | M7.CEL | mcf-7ctrl |
| 5 | mE1 | M3.CEL | mcf-7E2 |
| 6 | mE2 | M6.CEL | mcf-7E2 |
| 7 | mE3 | M9.CEL | mcf-7E2 |
| 8 | mI1 | M2.CEL | mcf-7IGF |
| 9 | mI2 | M5.CEL | mcf-7IGF |
| 10 | mI3 | M8.CEL | mcf-7IGF |
| 11 | sC1 | S1.CEL | sk-er3ctrl |
| 12 | sC2 | S4.CEL | sk-er3ctrl |
| 13 | sC3 | S7.CEL | sk-er3ctrl |
| 14 | sE1 | S3.CEL | sk-er3E2 |
| 15 | sE2 | S6.CEL | sk-er3E2 |
| 16 | sE3 | S9.CEL | sk-er3E2 |
| 17 | sI1 | S2.CEL | sk-er3IGF |
| 18 | sI2 | S5.CEL | sk-er3IGF |
| 19 | sI3 | S8.CEL | sk-er3IGF |

Figure 4: Target file structure.

### 4.1.2 Loading EXON/GENE ARRAYS

This sub menu, fig. 3, allows to load exon/gene 1.0 ST arrays starting from .CEL, taking advantage of Affymetrix APT tools (`http://www.affymetrix.com/support/developer/powertools/index.affx`), or flat tab delimited files containing gene/exon level expression data exported from Affymetrix Expression Console (EC, `http://www.affymetrix.com/support/technical/software_downloads.affx`). If APT tool option is not used (it works only for Exon 1.0 ST data exported from EC), a sub-menu allows to select, for tab delimited data, the organism and the subset of exon data to be evaluated, fig. 5

*IMPORTANT:*
*TO USE APT TOOLS THE DOWNLOAD OF GENE/EXON LIBRARY FILES IS REQUIRED.*
*THIS CAN BE DONE WITH THE FUNCTION*
*oeChannelGUI: Set library folder and install Affy gene/Exon library files*
*LOCATED IN THE GENERAL TOOLS MENU*

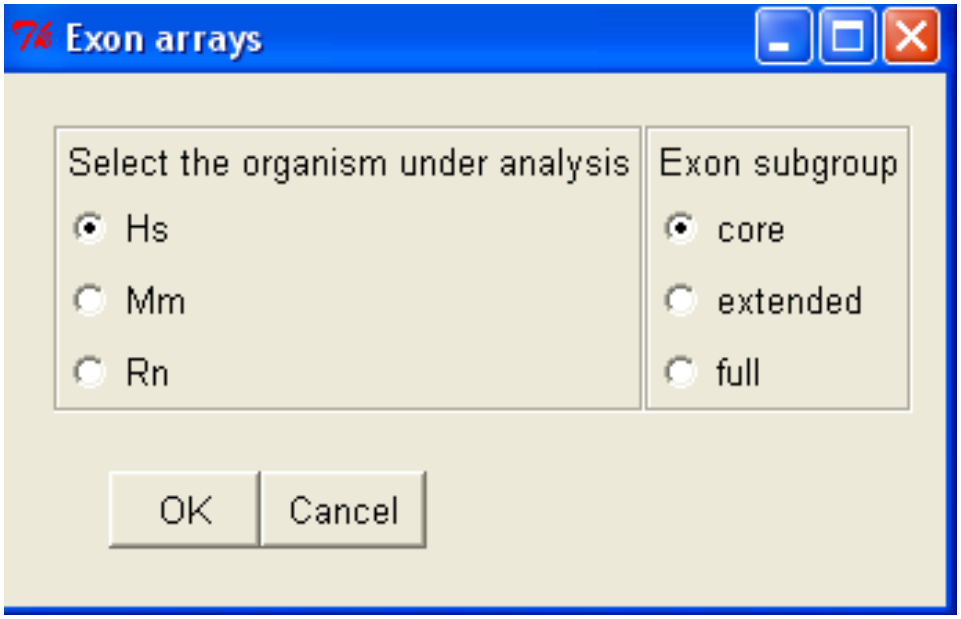



Figure 5: Sub menu to define the organism and the subset of exon data that will be loaded.

Subsequently, the user will select:

1. a working directory, a target file,
2. the flat tab delimited files containing respectively gene-level and exon-level data.

If instead, APT tool option is selected, user will select:

1. the organism and the subset of exon arrays to be evaluated, fig. 5,
2. a working directory,
3. a target file,
4. the type of probe set summary to be applied to gene/exon level data, fig. 6.

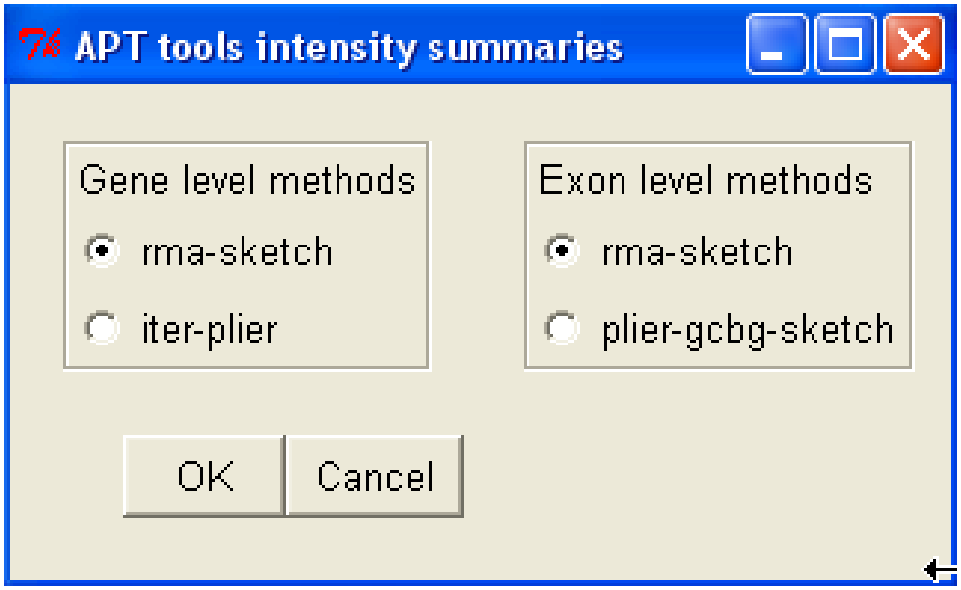



Figure 6: Sub menu to define the type of probe set summary to be applied.

Concerning probe set summary options, fig. 6, PLIER/RMA are the model-based algorithms available. Exons that are alternatively spliced in the samples, therefore exhibiting different expression patterns compared to the constitutive exons, will have down-weighted effect in overall gene-level target response values. A better estimation of gene-level signal could be obtained using IterPLIER, which is a variation of PLIER that iteratively discards features (probes) that do not correlate well with the overall gene-level signal and then recalculates the signal estimate to derive a robust estimation of the gene expression value primarily based on the expression levels of the constitutive exons. Concerning exon level expression estimation, most probe sets only have four probes, which is too limited to be useful with IterPLIER at the individual exon level, therfore il will be better to use PLIER/RMA.

Probe set summary calculation and uploading will take few minutes depending on the number of .CEL to be loaded and the PC in use. Ones probe set summary has been calculated, using APT tool, it is also possibile to calculate DABG p-values, fig. 7.

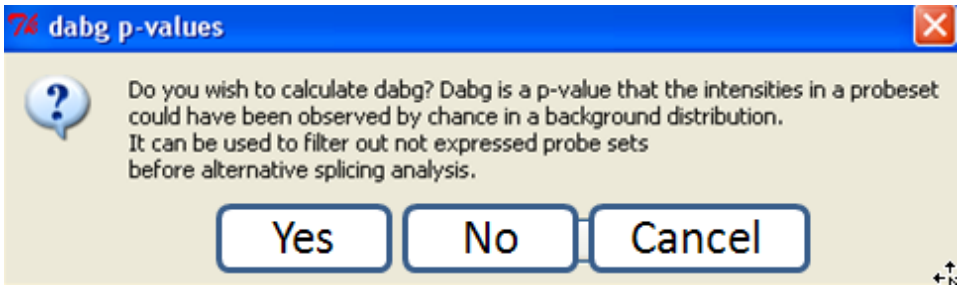



Figure 7: Selecting DABG p-value calculation.

DABG p-values represent *data above background*, it is a p-value similar to that used to derive presence/absence calls in MAS 5.0. DABG p-values could be useful to remove low intensity signals which could produce mis-leading results when alternative splicing events are evaluated using the Splice Index, where signal intensity information is not considered.

The progress of the probe set summary calculation is shown in the main R window.

```
Gene level probe sets summary started
Read 6 cel files from: target3d92750
Opening bgp file: HuEx-1_0-st-v2.r2.antigenomic.bgp
Opening clf file: HuEx-1_0-st-v2.r2.clf
Opening pgf file: HuEx-1_0-st-v2.r2.pgf
Expecting 1 iteration.
Doing iteration: 1
Opening clf file: HuEx-1_0-st-v2.r2.clf
Opening pgf file: HuEx-1_0-st-v2.r2.pgf
Loading 22011 probesets and 908532 probes.
Reading 6 cel files......Done.
Processing Probesets.....................Done.
Cleaning up.
Done.
Run took approximately: 9.56 minutes.

Gene level probe sets summary ended

Gene level probe sets summary ended

Exon level probe sets summary started

Exon level probe sets summary started
Read 6 cel files from: target3d92750
```

```
Opening bgp file: HuEx-1_0-st-v2.r2.antigenomic.bgp
Opening clf file: HuEx-1_0-st-v2.r2.clf
Opening pgf file: HuEx-1_0-st-v2.r2.pgf
Expecting 1 iteration.
Doing iteration: 1
Opening clf file: HuEx-1_0-st-v2.r2.clf
Opening pgf file: HuEx-1_0-st-v2.r2.pgf
Loading 287329 probesets and 1111849 probes.
Reading 6 cel files......Done.
Processing Probesets.....................Done.
Cleaning up.
Done.
Run took approximately: 6.41 minutes.

Exon level probe sets summary ended

Exon level probe sets summary ended

DABG calculation started
Read 6 cel files from: target3d92750
Opening bgp file: HuEx-1_0-st-v2.r2.antigenomic.bgp
Opening clf file: HuEx-1_0-st-v2.r2.clf
Opening pgf file: HuEx-1_0-st-v2.r2.pgf
Expecting 1 iteration.
Doing iteration: 1
Opening clf file: HuEx-1_0-st-v2.r2.clf
Opening pgf file: HuEx-1_0-st-v2.r2.pgf
Loading 22011 probesets and 908532 probes.
Reading 6 cel files......Done.
Processing Probesets.....................Done.
Cleaning up.
Done.
Run took approximately: 3.55 minutes.

DABG calculation ended
```

## 4.2 Open, Save, Save as

A project can be saved using the functions *Save as* fig. 2. A microarray project can also be uploaded again in oneChannelGUI with the function *open*.

## 4.3 Exporting normalized expression values

This function, fig. 2, allows to export, as tab delimited files expression data, loaded in oneChannelGUI. This function is also located in *filtering menu* and in the *exon menu*. If exon arrays are loaded in oneChannelGUI it is possible to extract not only the gene level expression data available but also exon level expression data and any other data generated during exon array analysis: Splice Index, MiDAS p-values, RP alternative splicing data.

## 4.4 Info about the loaded data set

This function, fig. 2, gives information about the set of data loaded in oneChannelGUI and on the corresponding annotation library, if available.

## 4.5 Attaching annotation lib info

If a Bioconductor library is available this is attached to the data loaded in oneChannelGUI and it will appear in the output of *Info about the loaded data set*. Using *Attaching annotation lib info* function, after loading expression data as a tab delimited file, it is possible to attach the Bioconductor annotation library associated to it. In case gene-level Affymetrix Whole transcriptome data are loaded as tab delimited file HuEx for human, MoEx for mouse and RaEx for rat annotation need to be attached. This is needed to allow the attachment of annotation information to the data set.

### 4.5.1 Probe set annotation

Concerning Exon 1.0 ST arrays, gene level annotation information are actually embedded in oneChannelGUI, and a stand alone function is provided to use gene-level annotation externaly to oneChannelGUI. Exon-level annotation is now provided by three annotation packages: HuExExonProbesetLocation, MoExExonProbesetLocation and RaExExonProbesetLocation . For exon arrays annotation is available at the gene level for the core subset of Hs/Mm/Rn. As soon as Bioconductor annotation libraries will be available for exon arrays the oneChannelGUI annotation will use them for annotation. Info about the available Affymetrix annotation release can be found in the main R window as part of the oneChannelGUI release major changes. For EXON 1.0 ST arrays, it is possible to link GeneBank accession numbers and EG to the gene-level probe sets using the function *Attaching ACC to Probe set IDs*, present in the Biological Interpretation menu. This function also allows to link EGs to glevel probe sets of a tab delimited fine that has in the first column the probe set ids. It is also possible to extract exon-level Probe Selection Region sequences associated to a specific gene-level probeset using the function *Extracting exon-level PSR sequences associated to one gene-level probeset* in the Biological Interpretation menu.

# 5 RNA target

The first item in the menu, fig. 8, is inherited from affylmGUI and allows the visualization of the experimental structure described by the target file used to load the expression data.

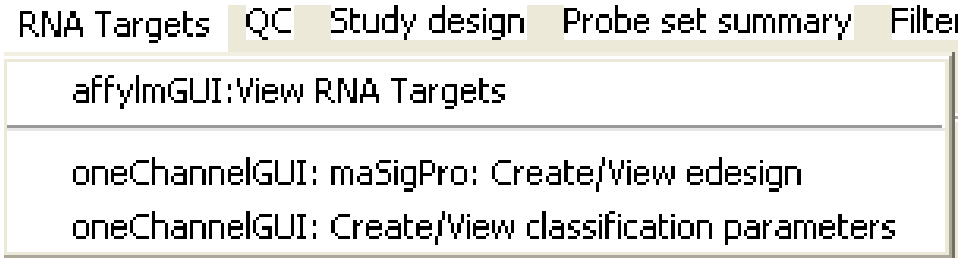



Figure 8: RNA target menu.

# 6 QC

The functions available in this menu are specific of the type of microarray data set loaded

## 6.1 QC for exon arrays

In the case of exon array the QC menu is slighly different, as shown in fig. 9

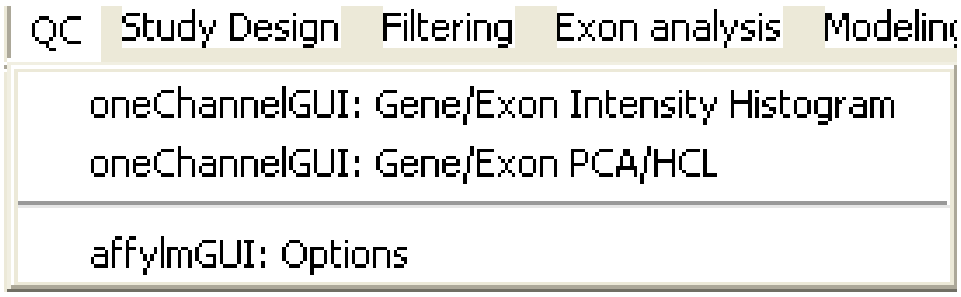



Figure 9: QC menu for exon arrays.

Two functions are available:

**Sample QC: PCA/HCL** This function will produce a PCA/HCL for both gene/exon level data.

**Gene/Exon intensity histogram** This function will produce a density histogram for gene or exon expression levels.

**Controls raw intensity histogram** This function will produce a box plot for exon, positive controls, and introns, negative controls, for housekeeping genes. Probe level data are directly extracted from CEL files using APT tools.

It useful, as quality control, to check intensities before normalization. Furthermore, the function *Gene/Exon Intensity Histogram* will show the density plot of the normalized intensities both at gene and at exon level.

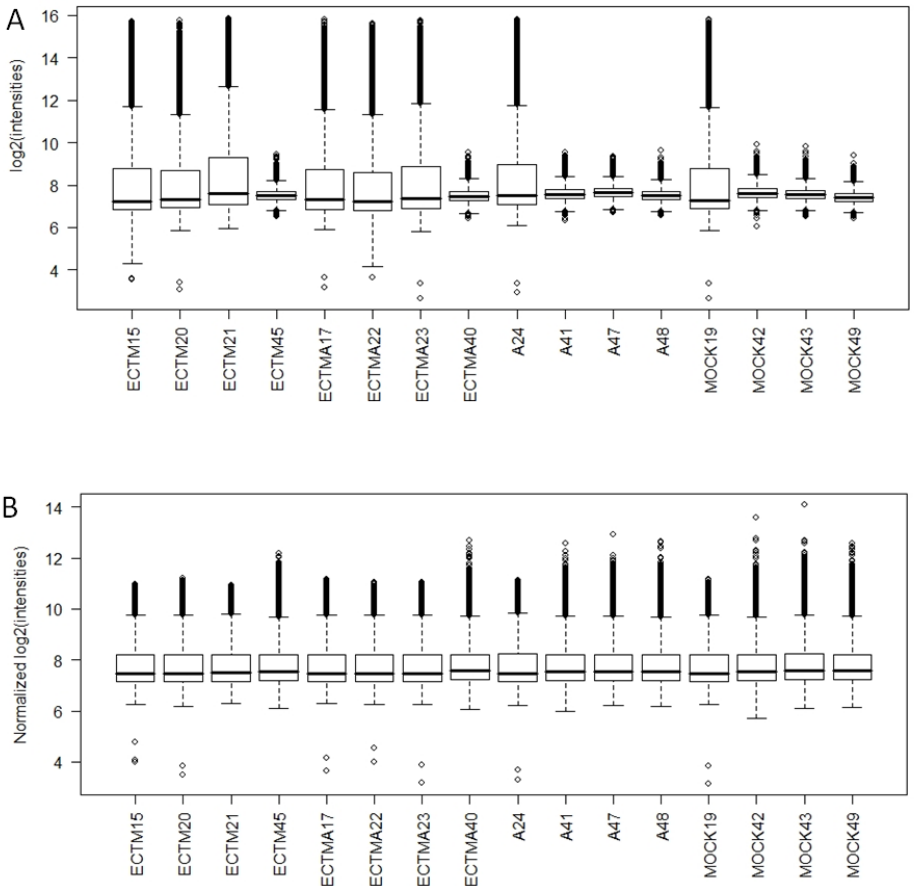



Figure 10: A set of Illumina arrays before and after data normalization.

As it can be seen in fig. 10 normalization masks the fact that a sub set of arrays, i.e. those with a very narrow boxplot 10A, had something wrong in hybridization. This problem is completely masked in the normalized data 10B. For this reason *Controls raw intensity histogram* was written for exon array data since probe sets data are directly uploaded as normalized in oneChannelGUI, via APT tools. This function produce a box plot for exon, positive controls, and introns, negative controls, for housekeeping genes. This box plot gives an idea of signals both at high and low intensity range.

# 7 Filtering

A central problem in microarray data analysis is the high dimensionality of gene expression space, which prohibits a comprehensive statistical analysis without focusing on particular aspects of the joint distribution of the gene expression levels. Possible strategies are to perform data-driven nonspecific filtering of genes (von Heydebreck, 2004) before the actual statistical analysis or to filter, making use of biologically relevant a priori knowledge. This menu allows user to apply a variety of filtering procedures, fig. 11

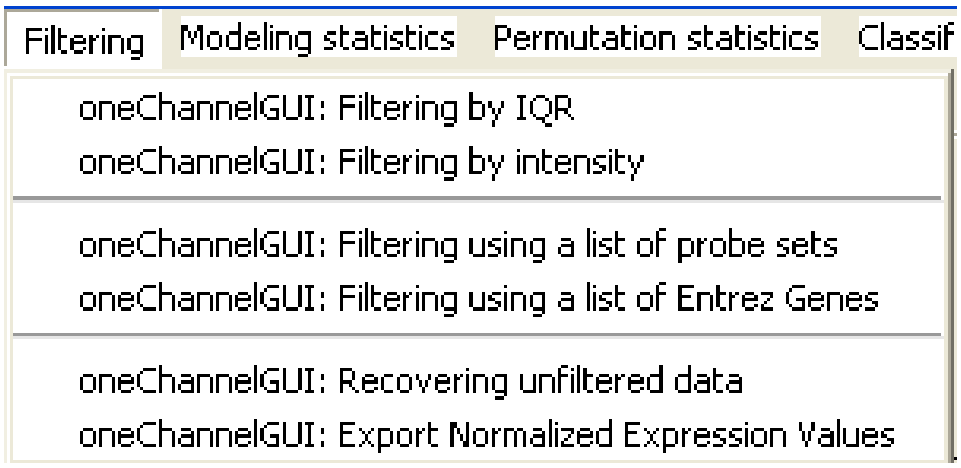



Figure 11: Filtering menu for GEO/Affy IVT arrays.

## 7.1 Filtering menu: exon data

If exon data are loaded the filtering menu appear slighly different, fig. 12.

In particular, the function *Set background threshold* collects the exon/intron expression values for a set of housekeeping genes present in the chip within chip quality controls and it offers the opportunity to set a background intensity threshold on the basis of the desidered level of intersection between the expression of exons versus introns. RMA intensity calculation is preferred, fig. 13, since, if probe set summaries are calculated with Plier or iterPlier, the differences in expression distribution between exon and introns are not enought wide, fig. 14.

Setting a background threshold using exon/intro distributions for HK genes, it is possible to apply to the full data set an intensity filtering that will remove gene and the corresponding exons on the basis of the selected threshold. The intensity filter for exon arrays works exactly as that for IVT arrays but using a fixed threshold defined as described.

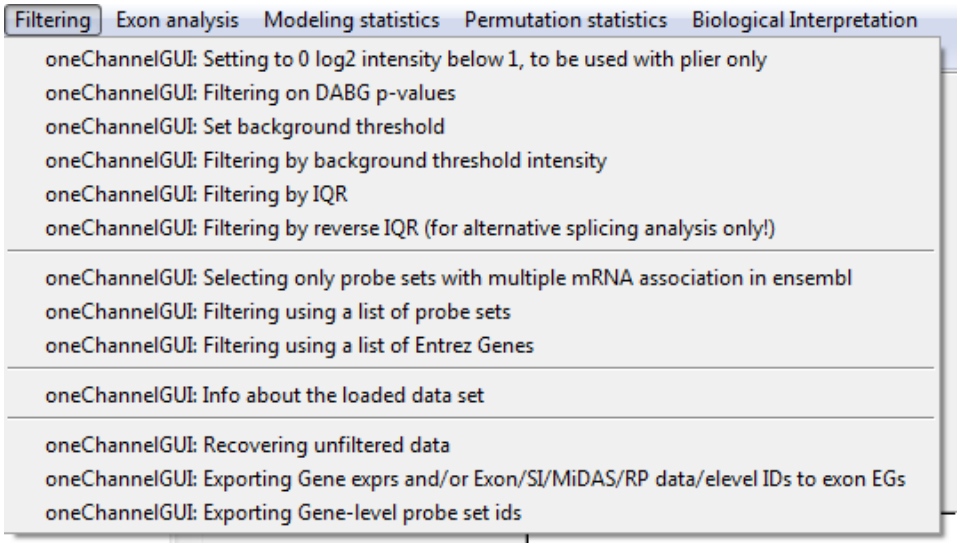



Figure 12: Filtering menu for exon data.

An other filter that allows the removal of low intensity probe sets is based on the DABG p-values. Using the function *Filtering on DABG p-values* it is possible to select the desidered level of filtering using a mask, fig. 15.

A threshold of 50% means that only probe sets where in half of the samples over the selected DABG p-value threshold will kept. As can be seen in fig. 16 this filtering also removes low intensity signals very near to zero.

*N.B. Recovering the data prior filtering is not implemented for DABG p-value filtering, yet.*

Reguarding very low intensity probe sets present if iterPlier/Plier are used, the function *Setting to 0 log2 intensity below 1, to be used with plier only* will set them zero.

The function *Selecting only probe sets with multiple mRNA association in ensembl* it is very useful when alternative splicing events are investigated, if the researcher is interested to investigate only those probe sets associated to multiple transcripts annotated on ensembl database. We strongly suggest to apply this filter at least to get an overview of the possibile known alternative splicing events that could be collected within the annotated ensembl data. This filter will reduce both the computational time to calculate splice index and type I statistical error, at the level of statistical analysis for alternative splicing detection.

Specifically, this function select at gene-level only those probe sets which are associated to multiple entries on ensembl data base. The filter uses the biomaRt package to collect this information from ensembl database.

The function oneChannelGUI: *Exporting Gene-level probe set ids* is useful to extract the list of probe set ids associated to the gene-level data set loaded on oneChannelGUI.

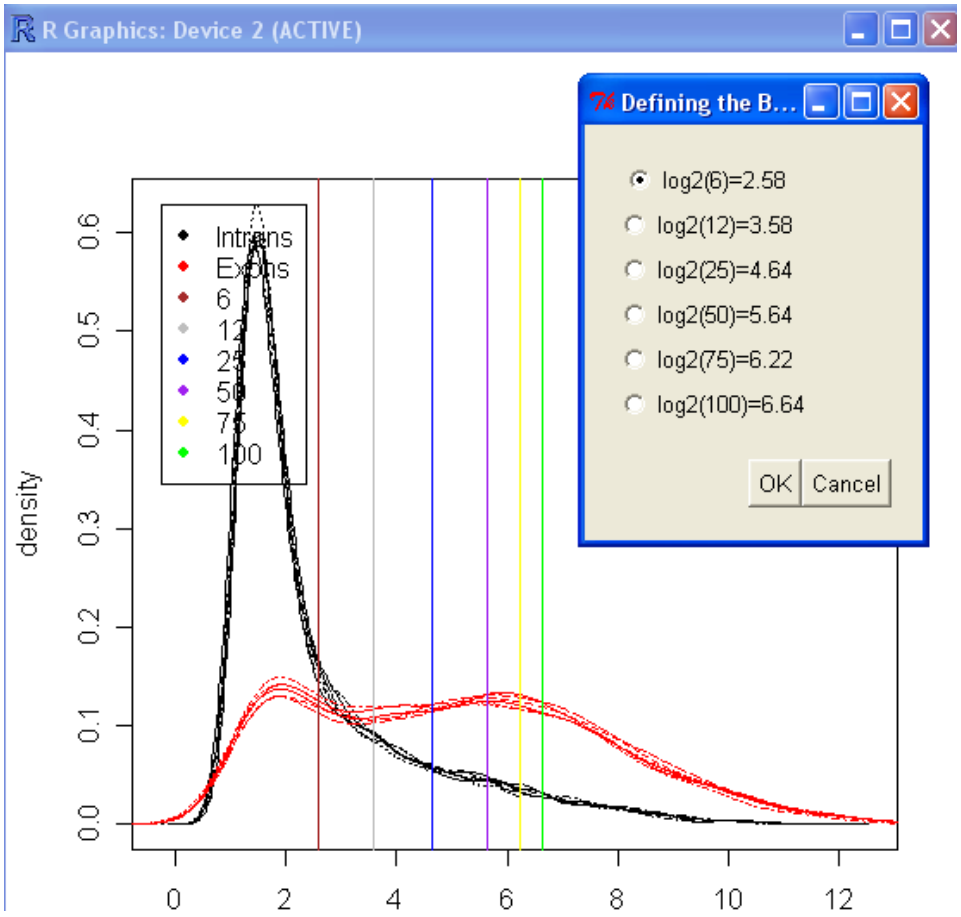



Figure 13: Human exon arrays, probe set summaries were calculated with RMA, exon/intron distribution of HK present in the chip as quality controls.

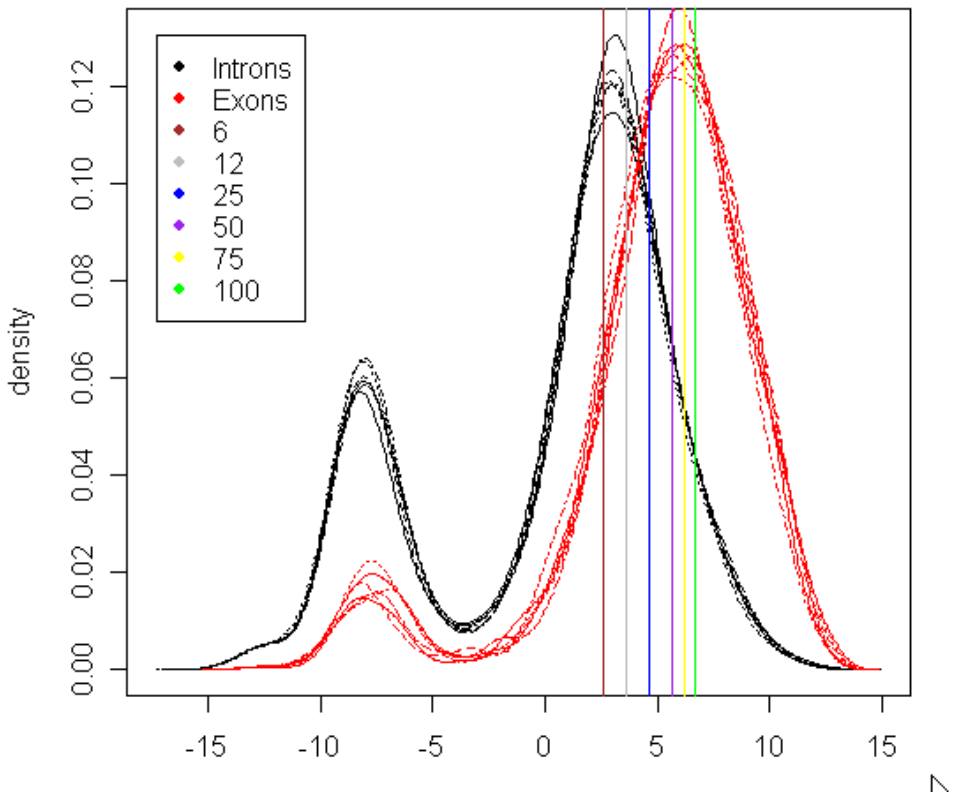



Figure 14: Human exon arrays, probe set summaries were calculated with iterPlier (gene level) and Plier (exon level), exon/intron distribution of HK present in the chip as quality controls.

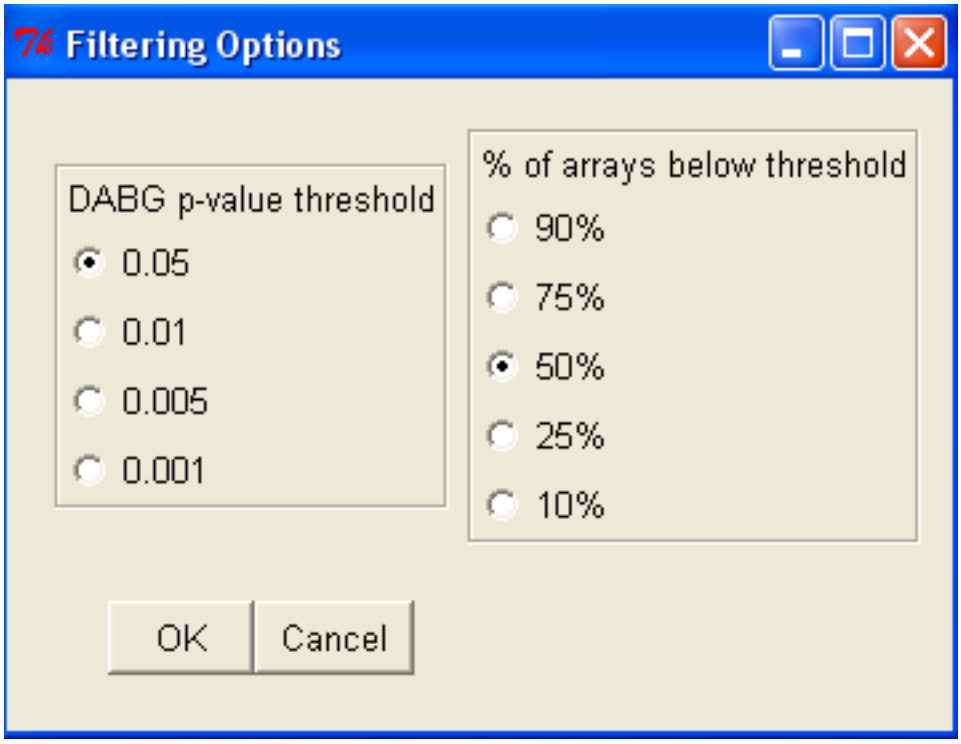



Figure 15: DABG p-value based filtering selection mask.

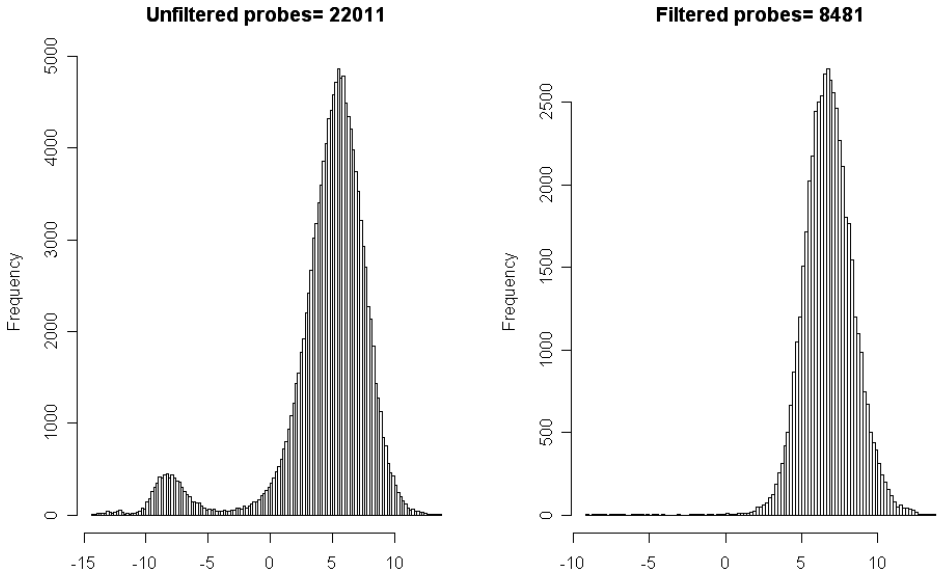



Figure 16: DABG p-value filtering results with parameters: DABG p-value threshold 0.05 and 50% of arrays over the threshold.

# 8 Exon analysis and data mining

Exon analysis menu allows a certain number of functions to identify and visualize alternative splicing events. The part related to loading gene/exon level data is described in the File menu chapter. If APT tools are used to calculate probe set intensities in oneChannelGUI will be available gene level expression data in Normalized Affy Data, exon level expression data in Normalized Exon data and, if selected, DABG p-values.

The functions actually available for exon analysis are summarised in fig. 17.

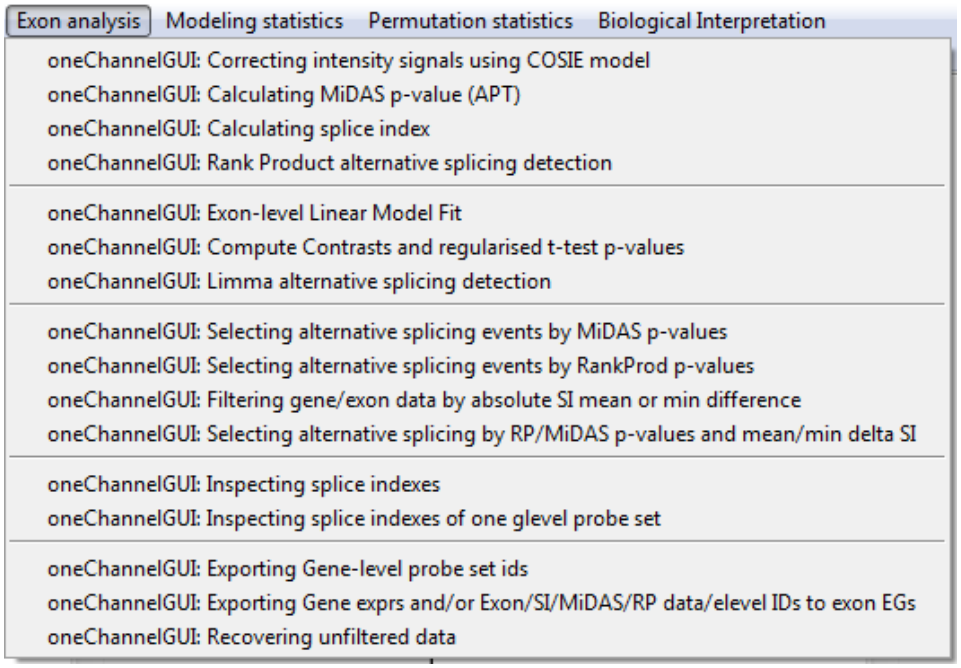



Figure 17: Exon menu.

Splice Index (SI), which represents the exon expression normalized with respect to the transcript expression, can be calculated with *oneChannelGUI: Calculating splice index*. Before SI calculation it is possible to correct gene/exon-levels intensity signals for probe sequence effect: *oneChannelGUI: Correcting intensity signals using COSIE model*. This function is a wrapper to the function developed by Gaidatzis et al. (Nucleic Acids Research, 2009). They have analyzed exon array data from many different human and mouse tissues and have uncovered a systematic relationship between transcript fold change and alternative splicing as reported by the splicing index. Evidence from dilution experiments and deep sequencing suggest that this effect is of technical rather than biological origin and that it is driven by sequence features of the probes. This effect is substantial and results in a 12-fold overestimation of alternative splicing events in genes that are differentially expressed. They have developed a R function called COSIE (Corrected Splicing Indices for Exon arrays) that for any given set of new exon array

experiments, core and full exons set for human and mouse, corrects for the observed bias and improves the detection of alternative splicing.

Starting from the work of Shah and Pallas work BMC Bioinformatics. 2009 Jan 20;10:26. Limma routines available for gene-level analysis were also implemented at exon-level to detect alternative splicing events. It is notable that it is possibile in many cases to apply BH type I error correction at exon-level. The limma analysis is performed at intensity level, therefore it will not discriminate between exon-level probe sets alternative spliced or belonging to a differentially expressed gene. Although this limitation is present we did not applyed limma analysis as Shah and Pallas to SI because there are no evidences that SI variance will be independent by the gene it belong to. The analysis steps are very similar to gene level: *oneChannelGUI: Exon-level Linear Model Fit* will fit the linear model. Then *oneChannelGUI: Compute Contrasts and regularised t-test p-values* will compute the contrasts and the regularised t-test p-values. Then *oneChannelGUI: Limma alternative splicing detection* will produce a file containing the exon-level ids of the altternative spliced exons and will filter the gene and exon-level data to retain only the genes affected by alternative splicing events.

For a two group experiment the function *oneChannelGUI: Calculating MiDAS p-value (APT)* uses APT tools to calculate MiDAS p-values for the difference between SIs in the two conditions, i.e. alternative splicing events. It is possible to subset gene/exon level data on the basis of a MiDAS p-value threshold using the function *oneChannelGUI: Selecting alternative splicing events by MiDAS p-values.* We have also applied the rank product method (RankProd package) *oneChannelGUI: Rank Product alternative splicing detection (devel)* to detect significant differences between SI or exon-level log2(intensities) in two experimental conditions, i.e. alternative splicing events. Rank Product is a non-parametric statistic that detects items that are consistently highly ranked in a number of lists. It is based on the assumption that under the null hypothesis that the order of all items is random the probability of finding a specific item among the top $r$ of $n$ items in a list is $p = \frac{r}{n}$. Multiplying these probabilities leads to the definition of the rank product $RP = \prod_i \frac{r_i}{n_i}$, where $r_i$ is the rank of the item in the $i$-th list and $n_i$ is the total number of items in the $i$-th list. The smaller the $RP$ value, the smaller the probability that the observed placement of the item at the top of the lists is due to chance. Due to performance reasons on windows based computers, the number of random permutations is fixed to 100, a menu to select the number of permutations will be implemented soon. At the end of the analysis p-values of class 1 $<$ class2 and p-values of class 1 $>$ class2 and average SI difference histograms are shown in the main R window.

*IMPORTANT All filtering functions devoted to the selection of alternative spliced events produce a flat file containing only the ids of the detected spliced exons.*

It is possible to subset gene/exon level data on the basis of rank product results using the function *oneChannelGUI: Selecting alternative splicing events by RankProd p-values.* It is also possible to filter data on the basis of the average mean or min SI difference with the function *oneChannelGUI: Filtering gene/exon data by absolute SI mean or min*

*difference*

It is also possible to filter exon data integrating midas p-values with RP p-values and average mean SI difference. This option is given by the function *oneChannelGUI: Selecting alternative splicing by RP/MiDAS p-values/average mean SI difference* Since no correction for statistical type I error is given for MiDAS we decided to use the integration of two statistical tests based on different approaches to reduce statistical type I errors. Furthermore, in this filter is integrated also the possibility to subset data on the basis of a average mean SI difference threshold. Visualization of the splicing events using one gene-level probe set at a time is possible with the function

In the Biological Interpretation menu, fig. 18, the function *oneChannelGUI: Associating alternative spliced exon-level probe set to variant exons* a set of spliced probe sets can be associated on the basis of the variant exons mapped on the UCSC genome browser.

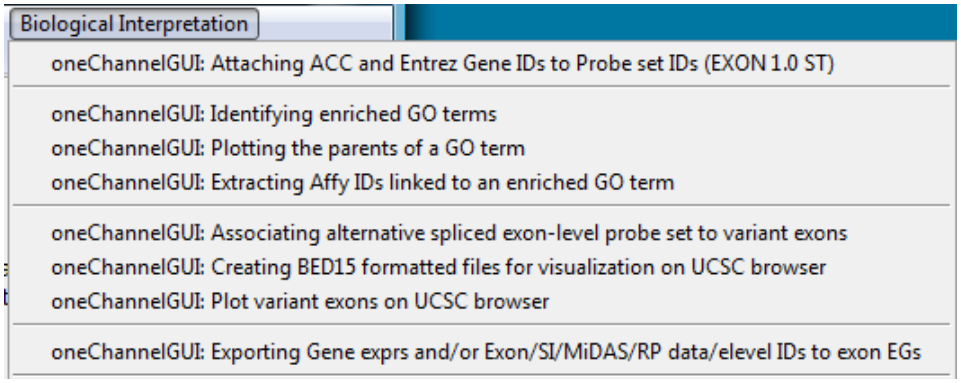



Figure 18: Biological Interpretation menu for exon arrays

Variant exons are those exons that are associated only to a subgroup of the available isoforms mapped on a gene. This association is very useful if researcher is not interested to select splicing events associated to exons conserved over all isoforms associated to a specific gene. The structure of the file is shown in fig. fig. 19

The tab delimited file produced by this filtering can be visualized on the UCSC genome browser, using two functions: *oneChannelGUI: Creating BED15 formatted files for visualization on UCSC browser* creates for each chromosome a BED15 formatted file, which allows the visualization of microarray data on the UCSC genome browser. Files need to be uploaded on the browser as shown in fig. 20,21

Exon-level feature are visualized as red colour in case of insertion and in green colour in case of skipping. Colour range is limited ranges between 3 and -3 deltaSI. Therefore colour for a deltaSI of 3 and 4 will be the same but the true deltaSI is also inserted as part of exon-level probeset name.

*oneChannelGUI: Plot variant exons on UCSC browser* takes advantage of the rtracklayer and plots directly on the UCSC genome browser from R. The output migth take

| C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|
| affyend | affywidth | affystrand | vspname | vspstart | vspend | vspwidth | vspstrand | chr | genome | exon.fc |
| 65658967 | 62 | + | NST000003710( | 65658901 | 65659011 | 111 | + | 1 | =Human&vers | 0.476112 |
| 65658967 | 62 | + | NST000003710( | 65658957 | 65659011 | 55 | + | 1 | =Human&vers | 0.476112 |
| 67222999 | 95 | + | NST000004010( | 67222842 | 67223955 | 1114 | + | 1 | =Human&vers | 0.450294 |
| 67222999 | 95 | + | NST000003576( | 67222842 | 67226302 | 3461 | + | 1 | =Human&vers | 0.450294 |
| 90265838 | 338 | + | NST000003704( | 90265500 | 90266685 | 1186 | + | 1 | =Human&vers | 0.910732 |
| 168031362 | 162 | + | NST000002860( | 168031173 | 168031748 | 576 | + | 1 | =Human&vers | 0.354134 |
| 226615373 | 126 | + | NST000002845( | 226613920 | 226615573 | 1654 | + | 1 | =Human&vers | 0.93415 |
| 15723282 | 131 | - | NST000003758( | 15723150 | 15723527 | 378 | - | 1 | =Human&vers | 0.472062 |
| 56817402 | 40 | - | NST000003712( | 56817138 | 56817845 | 708 | - | 1 | =Human&vers | 0.931792 |
| 150222342 | 36 | - | NST000003688( | 150222010 | 150222424 | 415 | - | 1 | =Human&vers | 0.400616 |
| 150222342 | 36 | - | NST000003688( | 150222017 | 150222424 | 408 | - | 1 | =Human&vers | 0.400616 |
| 154936307 | 235 | - | NST000003682( | 154936033 | 154936492 | 460 | - | 1 | =Human&vers | 1.424034 |
| 154936307 | 235 | - | NST000003682( | 154936171 | 154936492 | 322 | - | 1 | =Human&vers | 1.424034 |
| 154941905 | 27 | - | NST000003682( | 154941792 | 154941919 | 128 | - | 1 | =Human&vers | 0.858686 |
| 154941905 | 27 | - | NST000003682( | 154941792 | 154941962 | 171 | - | 1 | =Human&vers | 0.858686 |
| 154941905 | 27 | - | NST000003682( | 154941792 | 154941999 | 208 | - | 1 | =Human&vers | 0.858686 |

Figure 19: The columns present in the tab delimited file are the following: affyname: exon-level probe set, affystart: start position of PSR from Affymetrix annotation, affyend: end position of PSR from Affymetrix annotation, affywidth: width of the PSR, affystrand: stand of the PSR, vspname: ESEMBL transcript overlapping to the PSR, vspstart: start of the ESEMBL transcript, vspend: end of the ESEMBL transcript, vspwidth: width of the ESEMBL transcript, vspstrand: strand of the ESEMBL transcript, chr: chromosome location, genome: genome release used ofr the mapping, exon.fc: delta Splice Index, i.e. the fold change variation between the expression of the two experimental conditions for the signals normalized for the gene expression level.

sume time to be plotted on the default web browser.

The behaviour of exon-level probe sets, within a gene characterized by the presenc eof alternative splicing event can be inspected using two functions: *oneChannelGUI: Inspecting splice indexes of one glevel probe set* and *oneChannelGUI: Inspecting splice indexes*. The first function produces a pdf as the second one alows the analysis of gene-level probeset at a time. The structure of the output of these function is shown in fig. 22,

The function *oneChannelGUI: Inspecting splice indexes*, fig. 17. produces a pdf and txt file as output containing gene level probeset id and exon level probeset ids for all spliced exon in the following format:

```
"glevel id|exon level ids "
"3899173|3899229"
"3210737|3358127"
"3358112|3234972"
"2587961|3267416"
"3644510|3357399|3357446"
"3415109|2759224"
"3611625|3308001|3308013|3308031"
"3234760|3308001|3308013|3308031"
"3267382|3308001|3308013|3308031"
```

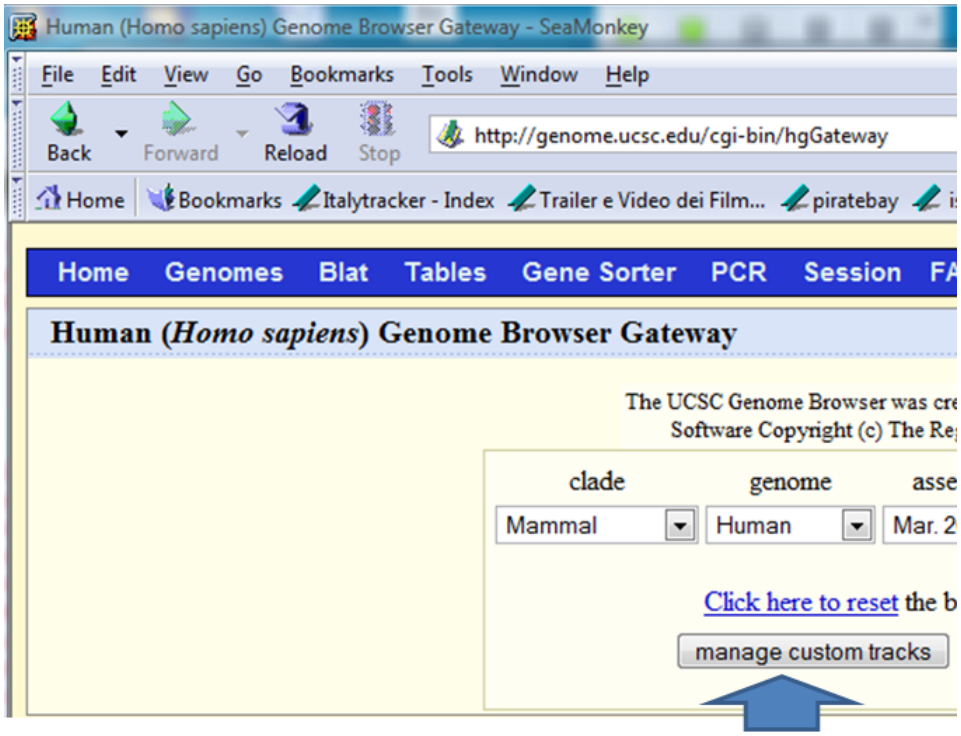



Figure 20: How to handle BED15 formatted files on UCSC genome browser, first step.

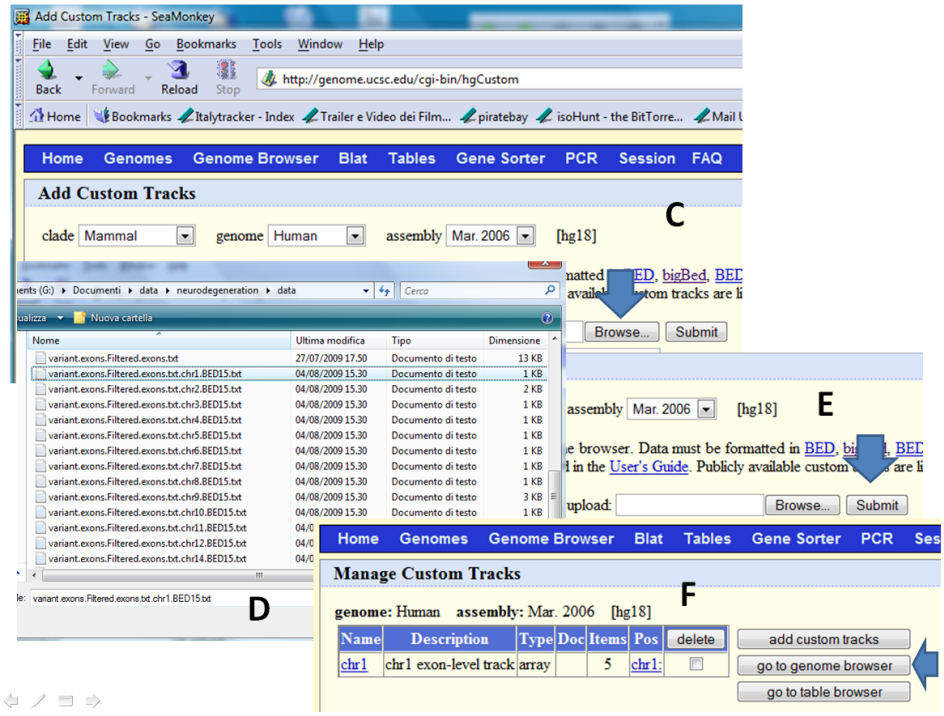



Figure 21: How to handle BED15 formatted files on UCSC genome browser, second step.

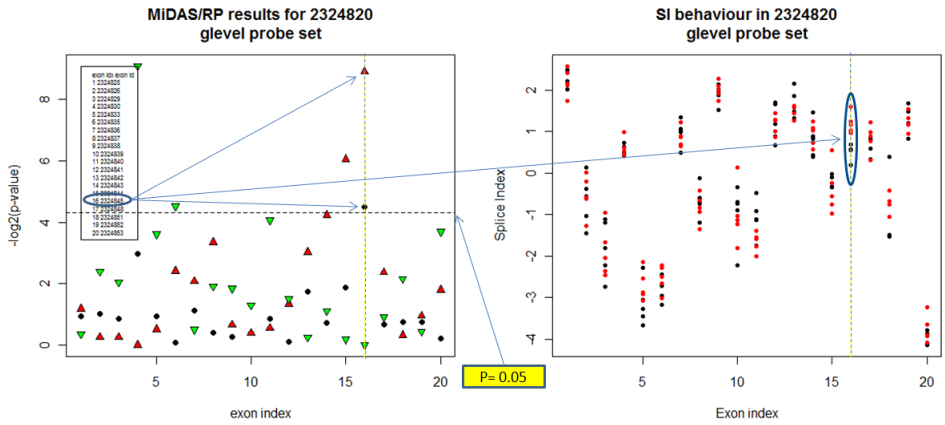



Figure 22: Example of the output of the putative alternative splicing inspection. The output is made of a tab delimited file where glevel probe sets associated to elevel probe sets and of a pdf file where each page is made of a plot of MiDAS/RP p-values with respect to exon index (black dot MiDAS, red triangle and green triangle RP). The horizontal black dashed line indicates a p-value of 0.05. The vertical yellow dashed line indicates a condition in which both MiDAS and RP p-values are below 0.05 value. In the second plot, it is shown the behaviour of splice indexes with respect to exon indexes. The vertical yellow dashed lines indicate those exon-level log2(intensity)/SI associated to MiDAS and RP p-values below 0.05 value.

*"3357397|2455983|2455993|2456013"*
*"2759205|2455983|2455993|2456013"*
*"3307939|3733603|3733609"*

Exon-level data can be saved using the function *oneChannelGUI: Exporting Gene exprs and/or Exon/SI/MiDAS/RP data/elevel IDs to exon EGs*

## 8.1 Plotting single splicing event

It is possible to plot average intensity signals over the genes and transcripts structure to identify a specific splicing event. This can be done using the function *oneChannelGUI: Plotting a splicing event* located in the exon menu. User has to pass to the function the exon-level probeset id for the spliced exon. This is used to retrieve gene-level information as well as the other exon-levels data. Using the package GenomeGraphs gene and transcript chromosomal structures are retrieved from ENSEMBL. In case the chromosome location data are associated to obsolete exon-level probeset data. These data can be provided externally using a file with the structure shown in fig. 23

| EPROBESETID | GPROBESETID | CHR | START | STOP | STRAND | ANNLEVEL | SCORE |
|---|---|---|---|---|---|---|---|
| 3804147 | 3804143 | chr18 | 33569794 | 33571808 | - | core | 1000 |
| 3804148 | 3804143 | chr18 | 33569794 | 33571808 | - | core | 1000 |
| 3804149 | 3804143 | chr18 | 33569794 | 33571808 | - | core | 1000 |
| 3804150 | 3804143 | chr18 | 33569794 | 33571808 | - | core | 1000 |
| 3804151 | 3804143 | chr18 | 33571818 | 33573263 | - | core | 1000 |
| 3804152 | 3804143 | chr18 | 33571818 | 33573263 | - | core | 1000 |
| 3804153 | 3804143 | chr18 | 33571818 | 33573263 | - | core | 1000 |
| 3804170 | 3804143 | chr18 | 33605561 | 33606838 | - | core | 1000 |
| 3804172 | 3804143 | chr18 | 33606863 | 33607038 | - | core | 1000 |
| 3804173 | 3804143 | chr18 | 33607147 | 33608143 | - | core | 1000 |
| 3804176 | 3804143 | chr18 | 33610771 | 33610857 | - | core | 1000 |
| 3804177 | 3804143 | chr18 | 33610954 | 33611060 | - | core | 1000 |
| 3804178 | 3804143 | chr18 | 33613671 | 33613800 | - | core | 1000 |
| 3804180 | 3804143 | chr18 | 33620770 | 33620812 | - | core | 1000 |
| 3804188 | 3804143 | chr18 | 33647166 | 33647442 | - | core | 1000 |
| 3804189 | 3804143 | chr18 | 33647166 | 33647442 | - | core | 1000 |

Figure 23: Example of the file structure describing chromosomal information for exon-level probesets. A file with this structure can be provided to the *oneChannelGUI: Plotting a splicing event* when user does not want to use annotation data available internally in oneChannelGUI.