

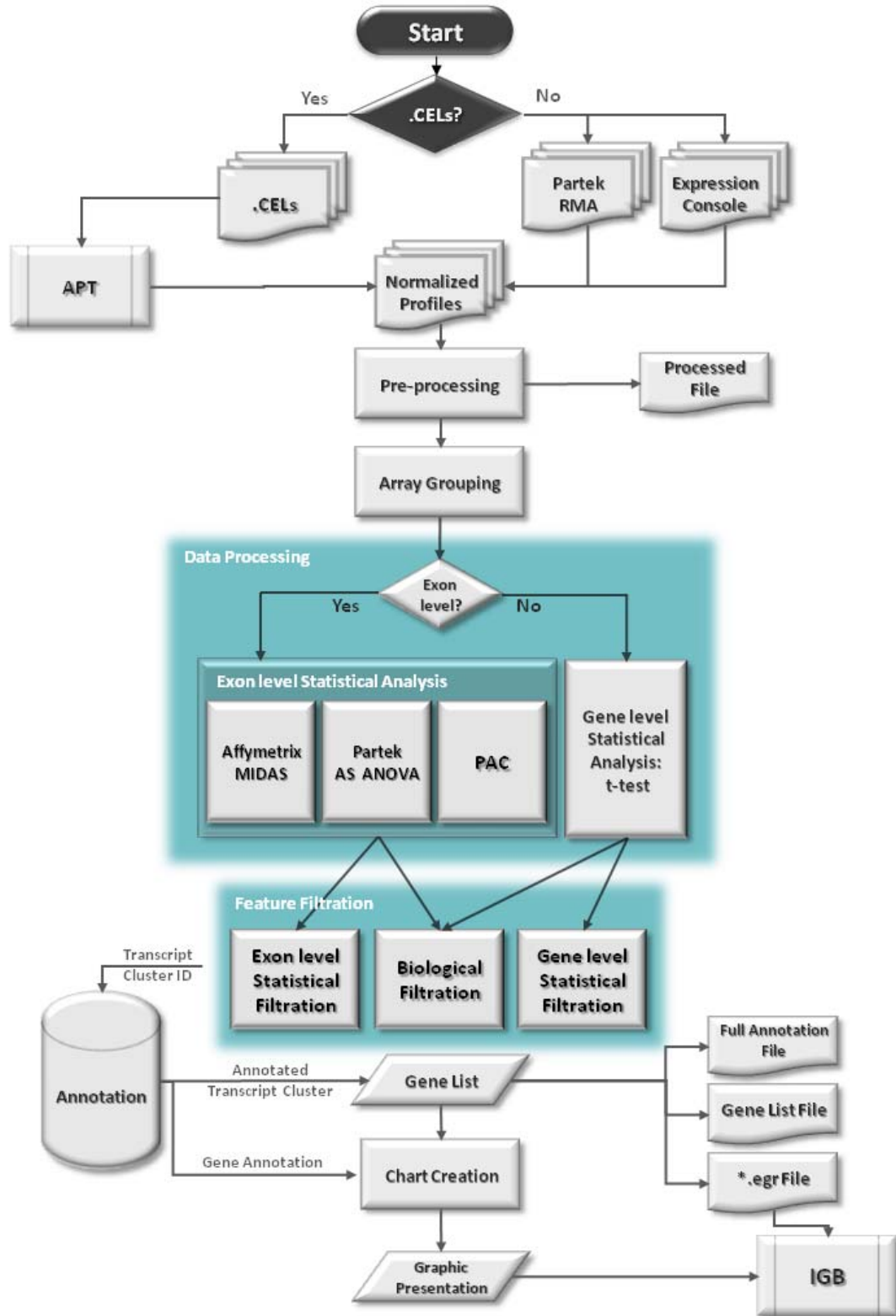
Tutorial for the easyExon

Version 1.0.3

Contents

0. Flowchart of easyExon	1
1. Alt-Splicing Study	2
Step 0: Input Data Format Selection.....	2
Step 1: Data Preparation	3
Step 1.a: Get summary files by Affymetrix Power Tools (APT)	3
Step 1.b: Load *.txt summary files	10
Step 2: Array Grouping and Data Processing	13
Step 2-1: Array Grouping.....	13
Step 2-2: Meta Probeset List Information	14
Step 2-2-1: Select Gene-level summary file.....	14
Step 2-2-2: Select meta probeset definitions	14
Step 2-3: Data Processing	16
Step 2-3.a: Exon Level Statistical Filtration	17
Step 2-3.b: Gene Level Statistical Filtration	19
Step 2-4: Export Processed Profile (Optional)	20
Step 3: Feature Filtration	21
Step 3-1.a: Statistical Filtration (Exon Level)	21
Step 3-1.b: Statistical Filtration (Gene Level)	24
Step 3-2: Biological Filtration	25
Step 4: Gene List	29
Step 4-1: User-defined Gene List	29
Step 4-2: Gene Annotation	30
Step 4-3: Export Full Annotation (Optional).....	31
Step 4-4: Export Table Information (Optional)	32
Step 4-5: Export *.egr for IGB (Optional).....	33
Step 5: Graphic Presentation	35
Step 5.a: Graphic Presentation for Exon Level.....	35
Step 5.b: Graphic Presentation for Gene Level.....	38
2. Environment Setting.....	39
2-1: Setting the Path of Meta Probeset Files	39
2-2: Setting the Path of Database	40

0. Flowchart of easyExon

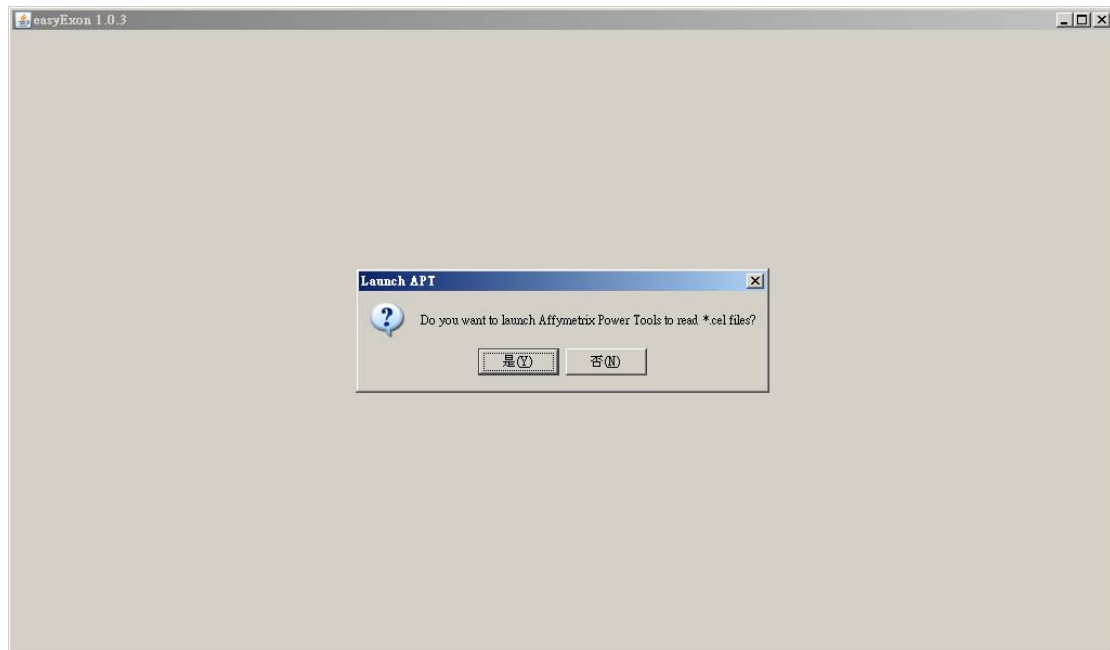


1. Alt-Splicing Study

Step 0: Input Data Format Selection

User needs to select the input data format. Two data formats are allowed:

- (i) CEL files
Choose **YES** to launch Affymetrix Power Tools (APT), and follow the [Step 1.a](#).
- (ii) *.txt summary files
Choose **NO** to load *.txt summary files, and follow the [Step 1.b](#).



Step 1: Data Preparation

Step 1.a: Get summary files by Affymetrix Power Tools (APT)

Before using easyExon to analyze exon arrays, user have to prepare Affymetrix Power Tools, corresponding library files, and *.CEL files. When these files are ready and corresponding directory paths are set, easyExon will launch APT to get the summary file that contains signals of each exon across all samples.

Data preparation:

(i) Affymetrix Power Tools (APT)

APT can be downloaded from the website:

<http://www.affymetrix.com/support/developer/powertools/index.affx>

(ii) Library files

Library files for APT have to be in the same directory. Those library files can be downloaded from the website:

<http://www.affymetrix.com/support/technical/libraryfilesmain.affx>

Here are the descriptions of library files.

Extension	Content
*.pgf	File defining probe sets E.g.: HuEx-1_0-st-v2.r2.pgf
*.clf	File defining x, y <-> probe id conversion E.g.: HuEx-1_0-st-v2.r2.clf
*.bgp	File defining probes to be used for GC background E.g.: HuEx-1_0-st-v2.r2.antigenomic.bgp
*.ps	File specifying probe sets to summarize E.g.: HuEx-1_0-st-v2.r2.dt1.hg18.core.ps
*.mps	File containing meta probeset definitions. File must contain a probeset_id column and a probeset_list column. E.g.: HuEx-1_0-st-v2.r2.dt1.hg18.extended.mps

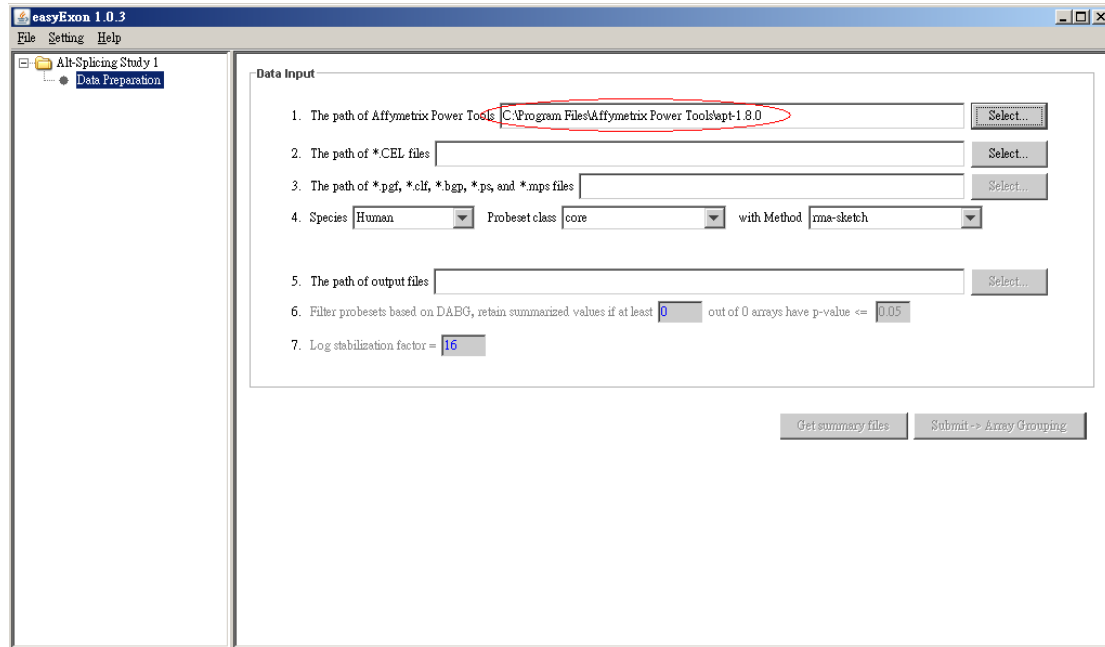
(iii) Probe results files (CEL files)

Put all CEL files in the same directory. The CEL files in the same directory will be used to get summary files.

Directory path setting:

(i) Path of the APT

Select the path of APT which contains the “bin” directory.



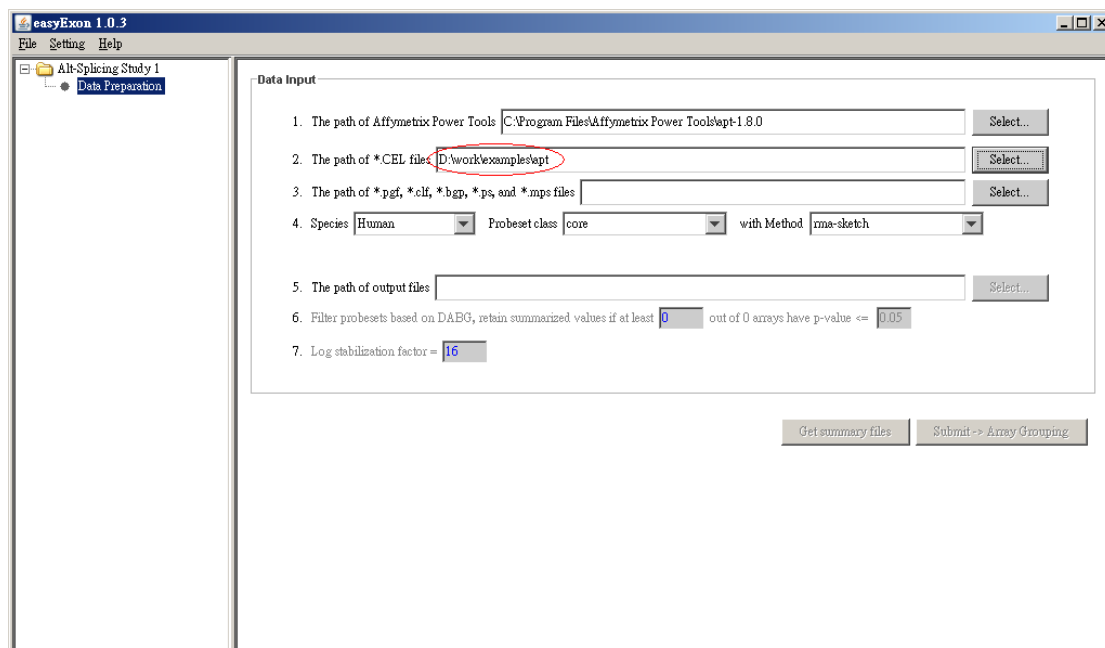
The screenshot shows the 'easyExon 1.0.3' application window. The 'Data Input' section contains the following fields and controls:

- 1. The path of Affymetrix Power Tools: (This field is circled in red)
- 2. The path of *.CEL files:
- 3. The path of *.pgf, *.clf, *.bgp, *.ps, and *.mps files:
- 4. Species: with Method:
- 5. The path of output files:
- 6. Filter probesets based on DABG, retain summarized values if at least out of 0 arrays have p-value <=
- 7. Log stabilization factor =

Buttons at the bottom:

(ii) Path of CEL files

Select the directory which includes CEL files.



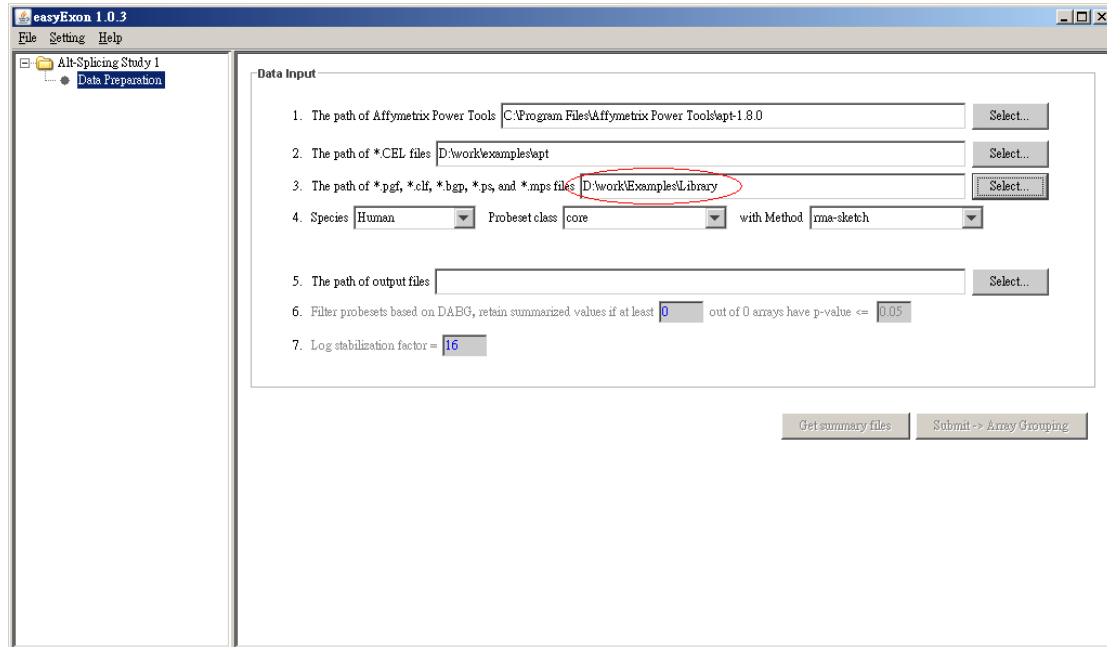
The screenshot shows the 'easyExon 1.0.3' application window. The 'Data Input' section contains the following fields and controls:

- 1. The path of Affymetrix Power Tools:
- 2. The path of *.CEL files: (This field is circled in red)
- 3. The path of *.pgf, *.clf, *.bgp, *.ps, and *.mps files:
- 4. Species: with Method:
- 5. The path of output files:
- 6. Filter probesets based on DABG, retain summarized values if at least out of 0 arrays have p-value <=
- 7. Log stabilization factor =

Buttons at the bottom:

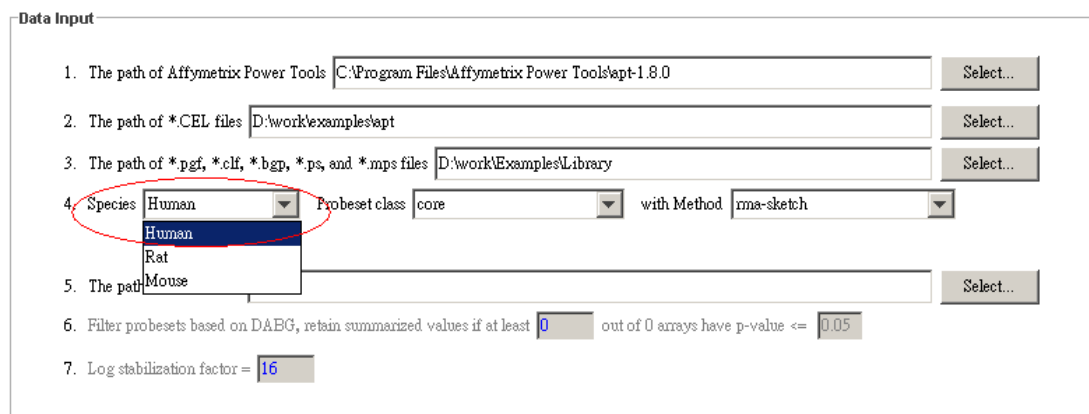
(iii) Path of library files

Select the path of library files directory which includes *.pgf files, *.clf files, *.bpg files, *.ps files, and *.mps files.



(iv) Select the species, probeset class and the method for getting summary files.

Species selection (Human, Rat, or Mouse)



Probeset class selection (core, full, extended, or defined by user)

Data Input

1. The path of Affymetrix Power Tools - 2. The path of *.CEL files - 3. The path of *.pgf, *.clf, *.bgp, *.ps, and *.mps files - 4. Species with Method
- 5. The path of output files - 6. Filter probesets based on DABG, retain summarized values if at least out of 0 arrays have p-value <=
- 7. Log stabilization factor =

When user selects “defined by user,” user needs to select the meta-probesets file location.

Meta-probesets file format:

The First row of the data contains “probeset_id”, “transcript_cluster_id”, and “probeset_list”. The first column contains probeset_ids; the second column is transcript_cluster_id. The third column is probeset_list with corresponding transcript_cluster_id.

E.g.:

```
probeset_id<Tab>transcript_cluster_id<Tab>probeset_list<Tab>...  
3096575<Tab>3096575<Tab>3096630 3096631 3096632<Tab>...  
2949118<Tab>2949118<Tab>2949119 2949120 2949121 2949130<Tab>...
```

Data Input

1. The path of Affymetrix Power Tools - 2. The path of *.CEL files - 3. The path of *.pgf, *.clf, *.bgp, *.ps, and *.mps files - 4. Species with Method
- 5. The path of output files - 6. Filter probesets based on DABG, retain summarized values if at least out of 0 arrays have p-value <=
- 7. Log stabilization factor =

Group probesets by

Method selection (rma-sketch or plier-gcbg-sketch)

Data Input

1. The path of Affymetrix Power Tools
2. The path of *.CEL files
3. The path of *.pgf, *.clf, *.bcp, *.ps, and *.mps files
4. Species with Method
5. The path of output files
6. Filter probesets based on DABG, retain summarized values if at least out of 0 arrays have p-value <=
7. Log stabilization factor =

(v) Path of output files

Select the output directory for summary files

easyExon 1.0.3
File Setting Help

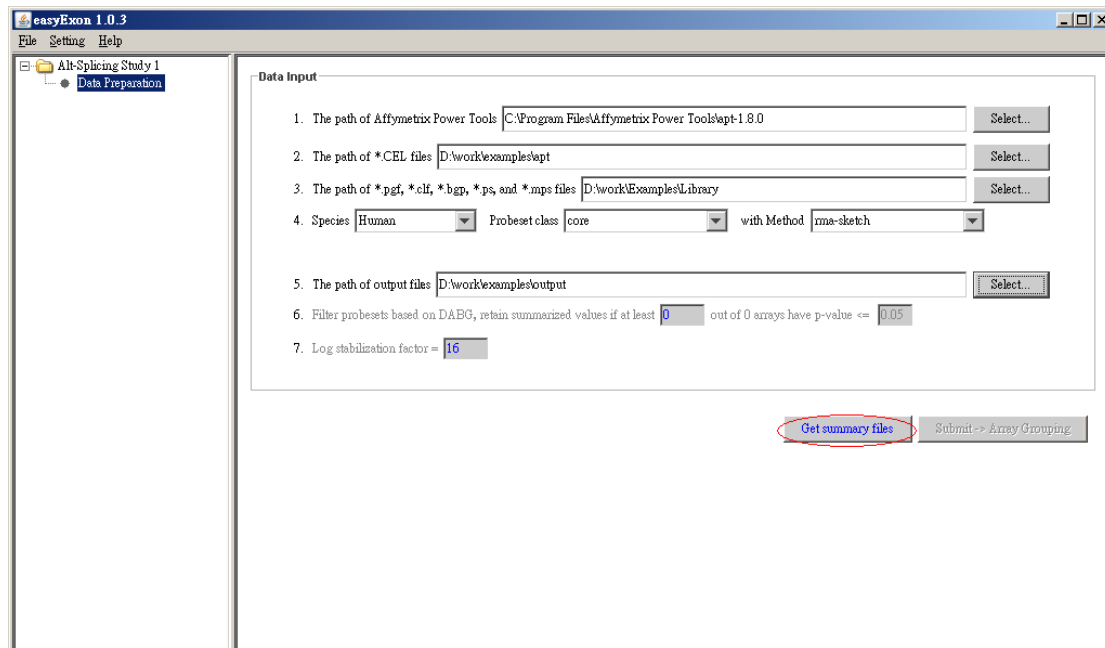
Alt-Splicing Study 1
Data Preparation

Data Input

1. The path of Affymetrix Power Tools
2. The path of *.CEL files
3. The path of *.pgf, *.clf, *.bcp, *.ps, and *.mps files
4. Species with Method
5. The path of output files
6. Filter probesets based on DABG, retain summarized values if at least out of 0 arrays have p-value <=
7. Log stabilization factor =

Get summary files!!!

Note: It may take few minutes to several hours, please have patience.



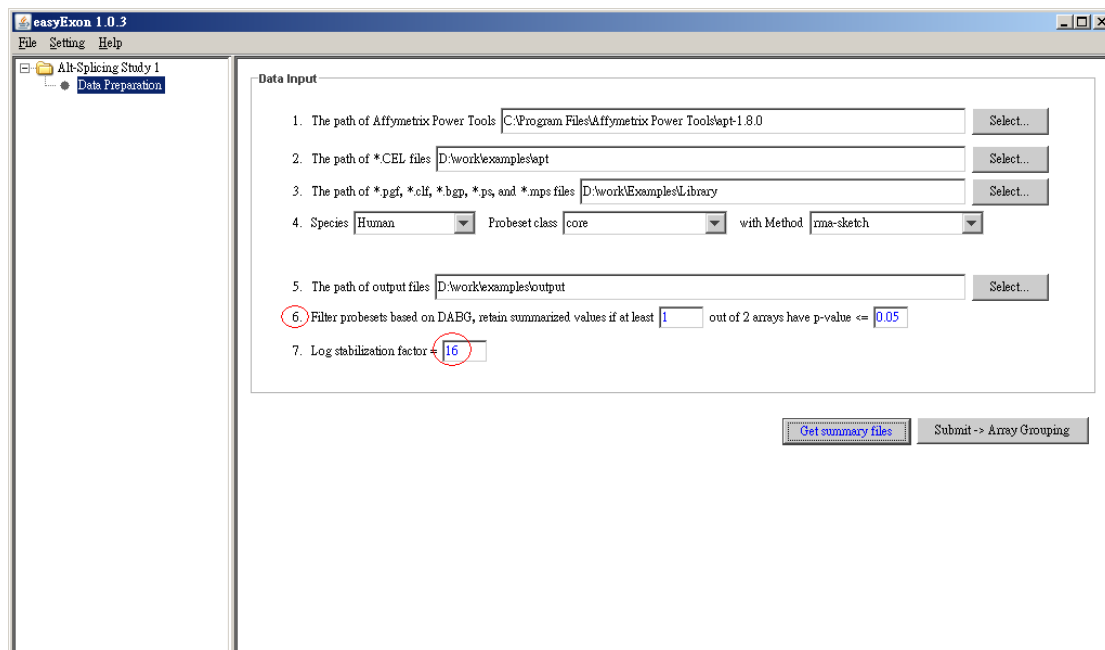
The screenshot shows the 'easyExon 1.0.3' application window. The 'Data Input' section contains the following fields and controls:

- The path of Affymetrix Power Tools:
- The path of *.CEL files:
- The path of *.pgf, *.clf, *.bgp, *.ps, and *.mps files:
- Species: with Method:
- The path of output files:
- Filter probesets based on DABG, retain summarized values if at least out of 0 arrays have p-value <=
- Log stabilization factor =

At the bottom right, there are two buttons: (circled in red) and .

Filter probesets based on DABG, and set the log stabilization factor.

When the summary file is generated by APT, user can filter probesets based on DABG p -value. User can also set the log stabilization factor value. The summarized signal will add the factor before the log transformation. The default value is 16.



The screenshot shows the 'easyExon 1.0.3' application window. The 'Data Input' section contains the following fields and controls:

- The path of Affymetrix Power Tools:
- The path of *.CEL files:
- The path of *.pgf, *.clf, *.bgp, *.ps, and *.mps files:
- Species: with Method:
- The path of output files:
- Filter probesets based on DABG, retain summarized values if at least out of 2 arrays have p-value <=
- Log stabilization factor = (circled in red)

At the bottom right, there are two buttons: and .

Submit to [Step 2!!!](#)

The screenshot shows the 'easyExon 1.0.3' application window. The title bar includes 'File', 'Setting', and 'Help' menus. On the left, a tree view shows 'Alt-Splicing Study 1' with a sub-item 'Data Preparation'. The main area is titled 'Data Input' and contains the following fields and buttons:

1. The path of Affymetrix Power Tools:
2. The path of *.CEL files:
3. The path of *.pgf, *.clf, *.bgp, *.ps, and *.mps files:
4. Species: Probeset class: with Method:
5. The path of output files:
6. Filter probesets based on DABG, retain summarized values if at least out of 2 arrays have p-value \leq
7. Log stabilization factor =

At the bottom right, there are two buttons: 'Get summary files' and 'Submit -> Array Grouping' (which is circled in red).

Step 1.b: Load *.txt summary files

Select the Exon-level summary files.

Select the data format (Expression Console or Partek) of Exon-level summary files, and set the path of file.

Data format:

(i) Expression Console:

The First row of the data contains “probeset_id” and CEL names. The first column contains probeset_ids, and remaining columns are intensity of each CEL with corresponding probeset_id.

E.g.:

```
probeset_id<Tab>01_KL1.CEL<Tab>02_KL3.CEL<Tab>03_KL5.CEL<Tab>...  
2315252<Tab>4.31<Tab>3.28<Tab>3.72<Tab>...  
2315253<Tab>2.46<Tab>1.59<Tab>2.20<Tab>...
```

(ii) Partek:

The First row of the data contains “Filename” and probeset_ids. The first column contains CEL names, and remaining columns are intensity of each probeset_id with corresponding. CEL.

E.g.:

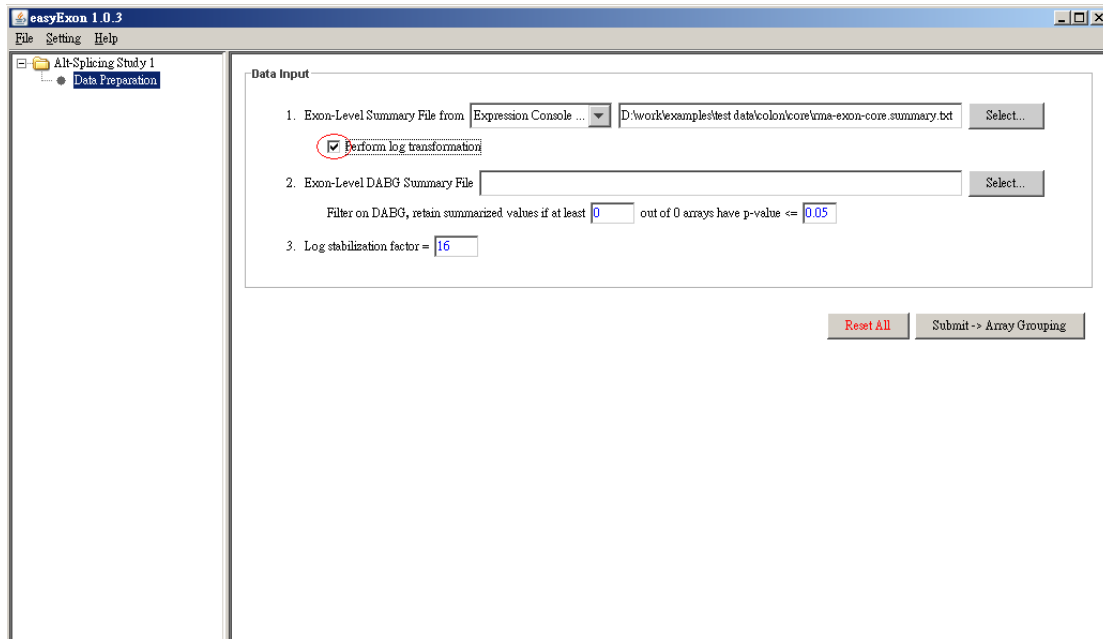
```
Filename<Tab>...<Tab>2315252<Tab>2315253<Tab>...  
01_KL1.CEL<Tab>...<Tab>4.31<Tab>2.46<Tab>...  
02_KL3.CEL<Tab>...<Tab>3.28<Tab>1.59<Tab>...  
03_KL5.CEL<Tab>...<Tab>3.72<Tab>2.20<Tab>...
```

The screenshot shows the 'easyExon 1.0.3' application window. The 'Data Input' section contains the following elements:

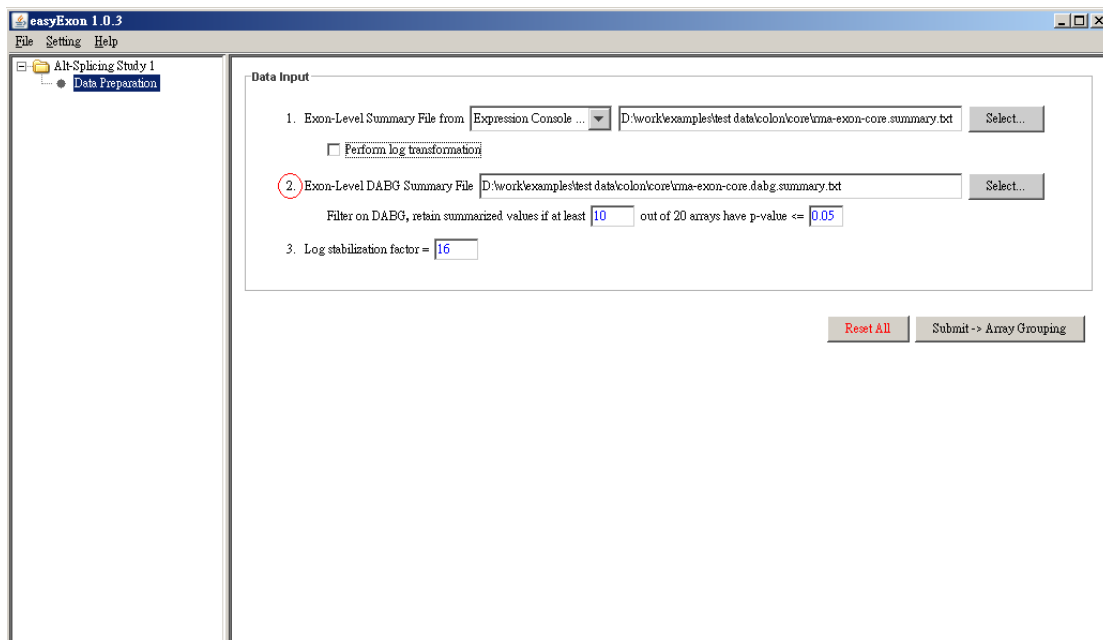
- 1. Exon-Level Summary File from: A dropdown menu is set to 'Expression Console'. The text box contains 'D:\work\example\test data\colon\core\kma-exon-core.summary.txt'. A red circle highlights the dropdown and text box area.
- Perform log transform: This checkbox is unchecked.
- 2. Exon-Level DABG Summary File: An empty text box.
- Filter on DABG, retain summarized values if at least out of 0 arrays have p-value <=
- 3. Log stabilization factor =

At the bottom right, there are two buttons: 'Reset All' and 'Submit -> Array Grouping'.

When user needs to perform log transformation on data, check the Perform log transformation.

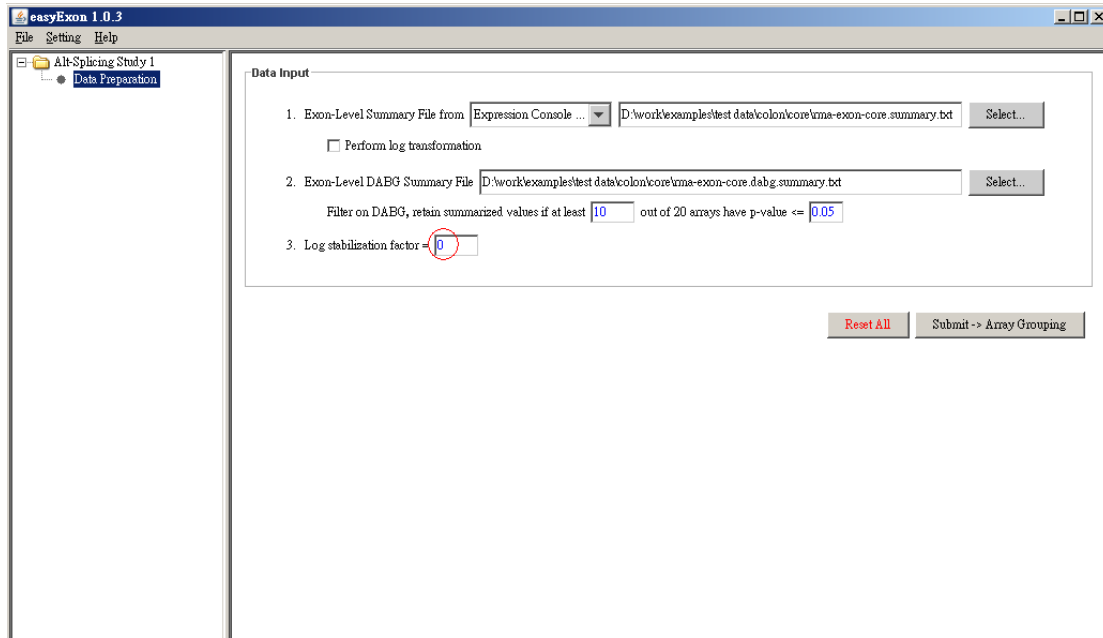


When user selects Expression Console, user needs to select the path of Exon-level DABG summary files and set probesets filter based on DABG p -value.



Set the log stabilization factor

When user selects Expression Console or Custom format, user needs to set the log stabilization factor. The summarized signal will add the factor before the log transformation. The default value is 16.

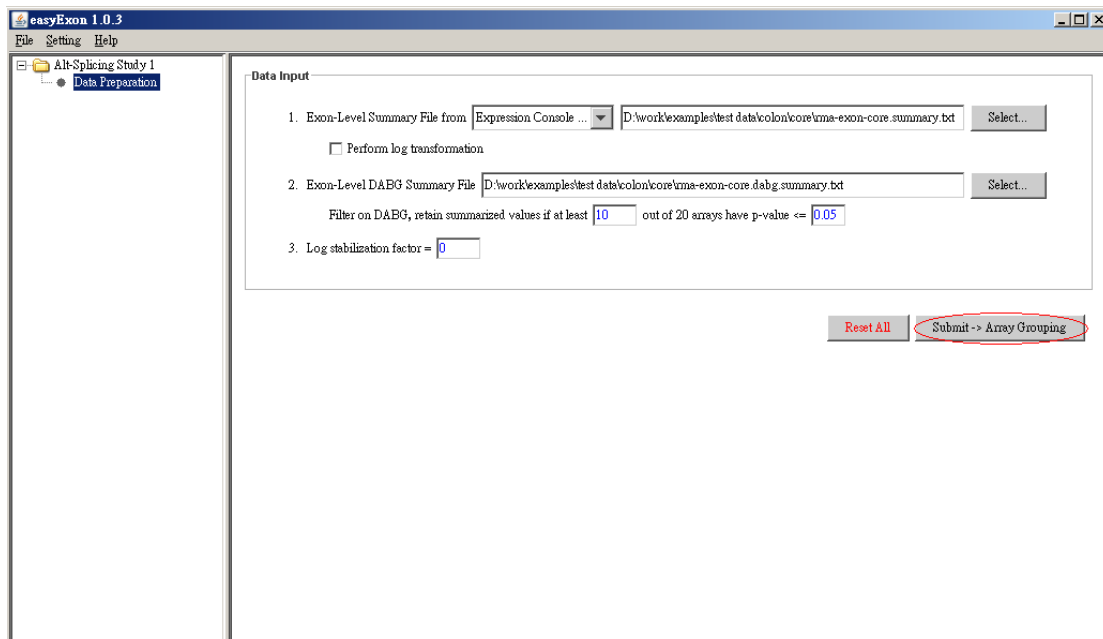


The screenshot shows the 'easyExon 1.0.3' application window. The 'Data Input' section contains the following fields and controls:

- 1. Exon-Level Summary File from: Expression Console ... (dropdown) | D:\work\example\test\data\colon\core\rma-exon-core.summary.txt (text) | Select... (button)
- Perform log transformation
- 2. Exon-Level DABG Summary File: D:\work\example\test\data\colon\core\rma-exon-core.dabg.summary.txt (text) | Select... (button)
- Filter on DABG, retain summarized values if at least 10 out of 20 arrays have p-value <= 0.05
- 3. Log stabilization factor = 0 (text input, circled in red)

Buttons at the bottom right: Reset All (red), Submit -> Array Grouping (grey).

Submit to [Step 2!!!](#)

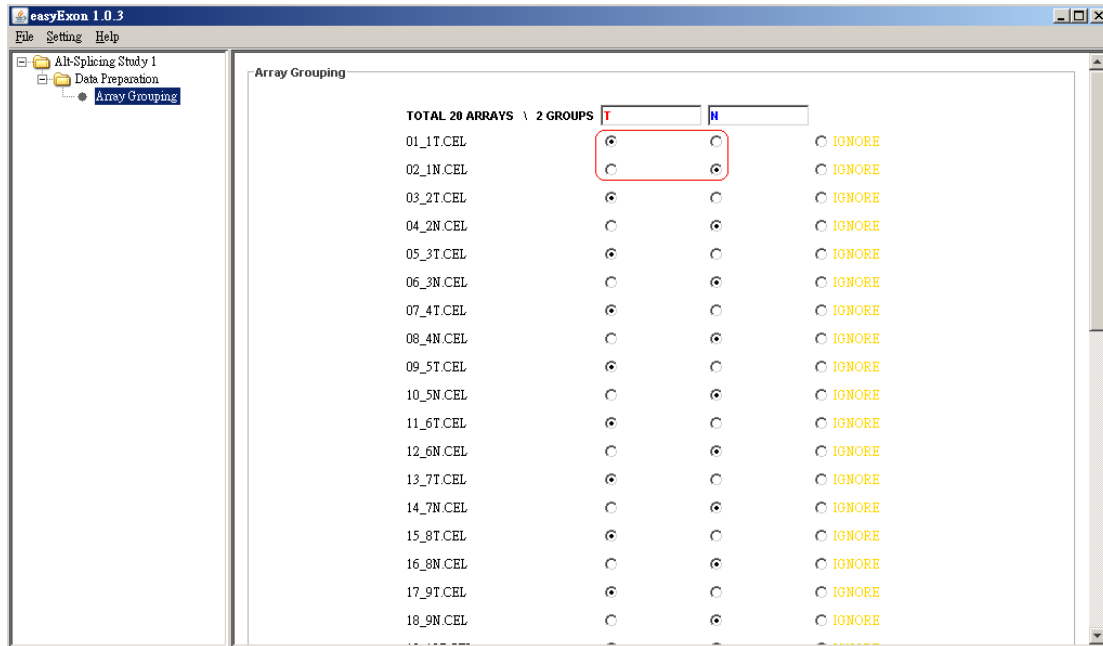


This screenshot is identical to the one above, but the 'Submit -> Array Grouping' button is circled in red.

Step 2: Array Grouping and Data Processing

Step 2-1: Array Grouping

Array samples are assigned into two groups based on experimental design. The group name can be arbitrarily modified for easy recognition. For samples that don't want to be included in the latter analyses, they can be assigned into the "IGNORE" group.

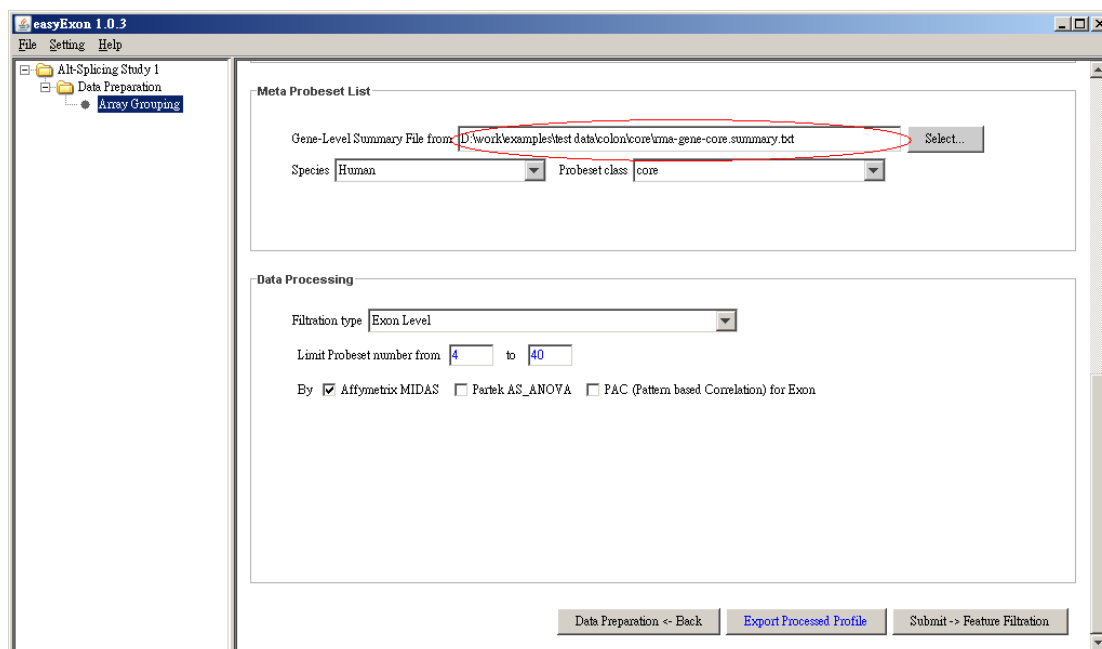


Step 2-2: Meta Probeset List Information

The information needed in this section is used for calculation of splicing index.

Step 2-2-1: Select Gene-level summary file

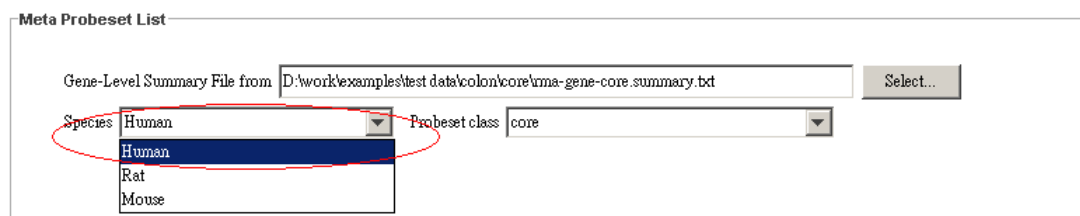
User who selects Expression Console at the [Step 1.b](#) needs to set the path of gene-level summary file. Otherwise, skip this step.



Step 2-2-2: Select meta probeset definitions

User who chooses NO at the [Step 0](#) needs to select the Species and Probeset class for meta probeset definitions. Otherwise, skip this step.

Species selection (Human, Rat, or Mouse)



Probeset class selection (core, full, extended, or defined by user)

Meta Probeset List

Gene-Level Summary File from

Species Probeset class

- core
- extended
- full
- defined by user

When user selects “defined by user,” user needs to select the meta-probesets file location.

Meta-probesets file format:

The First row of the data contains “probeset_id”, “transcript_cluster_id”, and “probeset_list”. The first column contains probeset_ids; the second column is transcript_cluster_id. The third column is probeset_list with corresponding transcript_cluster_id.

E.g.:

```
probeset_id<Tab>transcript_cluster_id<Tab>probeset_list<Tab>...  
3096575<Tab>3096575<Tab>3096630 3096631 3096632<Tab>...  
2949118<Tab>2949118<Tab>2949119 2949120 2949121 2949130<Tab>...
```

Meta Probeset List

Gene-Level Summary File from

Species Probeset class

Group probesets by

Step 2-3: Data Processing

There are two types of differential expression filtrations in easyExon: Exon level and Gene level.

(i) Exon Level

Choose Exon Level and follow the [Step 2-3.a](#).

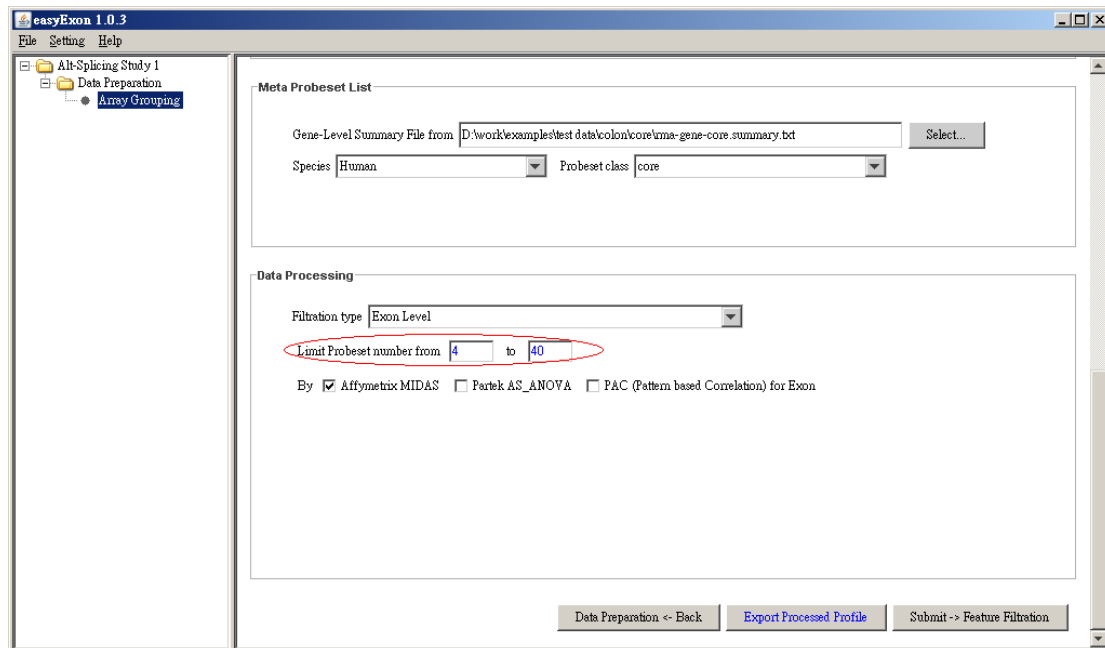
(ii) Gene Level

Choose Gene Level and follow the [Step 2-3.b](#).

The screenshot shows the 'easyExon 1.0.3' application window. On the left is a tree view with 'Alt-Splicing Study 1' expanded to show 'Data Preparation' and 'Array Grouping'. The main area is titled 'Data Processing'. At the top, 'Meta Probeset List' includes a text field for 'Gene-Level Summary File from' (D:\work\examples\test data\colon\core\lma-gene-core.summary.txt) and a 'Select...' button. Below are dropdowns for 'Species' (Human) and 'Probeset class' (core). The 'Filtration type' dropdown is highlighted with a red oval and shows 'Exon Level' selected. Below it, 'Limit Probes' shows 'Gene Level' selected. At the bottom, there are checkboxes for 'By' with options: Affymetrix MIDAS, Partek AS_ANOVA, and PAC (Pattern based Correlation) for Exon. Navigation buttons at the bottom are 'Data Preparation <- Back', 'Export Processed Profile', and 'Submit -> Feature Filtration'.

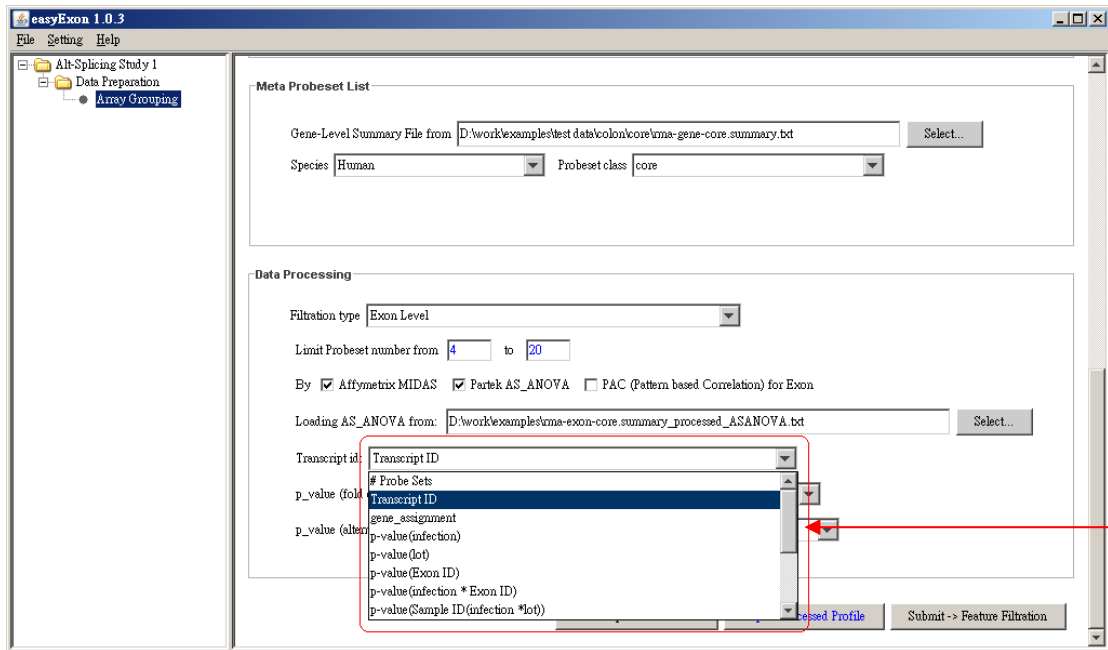
Step 2-3.a: Exon Level Statistical Filtration

Before performing the statistical filtration, user sets the limitation number of probesets in a transcript cluster. For the transcript cluster with probesets greater than the setting number, only the setting number of probesets will be included in the latter analysis.



There are three statistical methods for Exon level filtration in easyExon: Affymetrix MIDAS (Microarray Detection of Alternative Splicing), Partek AS ANOVA, and PAC (Pattern-Based Correlation) for Exon. User can use multiple selections to choose the more than one method at the same time.

- (i) Affymetrix MIDAS
Default selection for statistical filtration. No other information is needed.
- (ii) Partek AS ANOVA
User needs to set the path of AS ANOVA file, and sets the attributes of Transcript id and p -values from attribute list, which is the first row of the data, and each column is an attribute. Attribute list differs from AS ANOVA files.



- (iii) PAC (Pattern based correlation) for Exon
No other information is needed.

Last, go to the [Step 2-4](#).

Step 2-3.b: Gene Level Statistical Filtration

easyExon will perform t-test filtration on gene level in this step. No other information is needed.

The screenshot shows the 'easyExon 1.0.3' application window. The left sidebar shows a project tree with 'Alt-Splicing Study 1' expanded to 'Data Preparation' and 'Array Grouping' selected. The main area contains a table of CEL files:

File Name	Radio 1	Radio 2	Radio 3	Action
15_8T.CEL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	IGNORE
16_8N.CEL	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	IGNORE
17_9T.CEL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	IGNORE
18_9N.CEL	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	IGNORE
19_10T.CEL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	IGNORE
20_10N.CEL	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	IGNORE

Below the table is the 'Meta Probeset List' section with the following fields:

- Gene-Level Summary File from:
- Species:
- Probeset class:

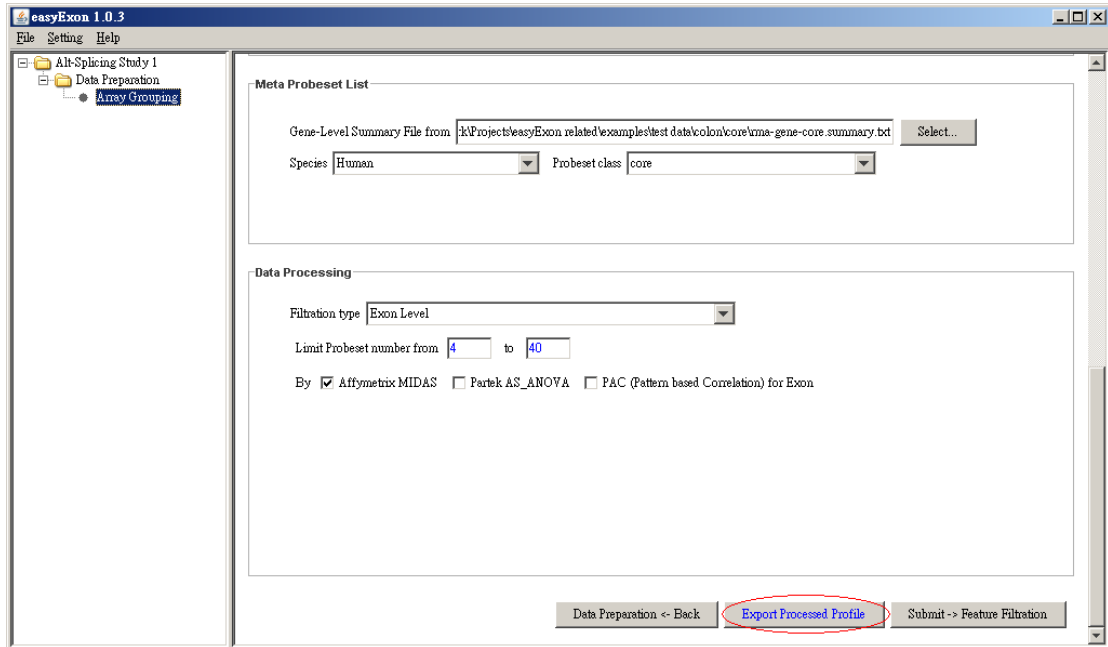
The 'Data Processing' section has:

- Filtration type:
- By t-test

At the bottom, there are three buttons: 'Data Preparation <- Back', 'Export Processed Profile', and 'Submit -> Feature Filtration'.

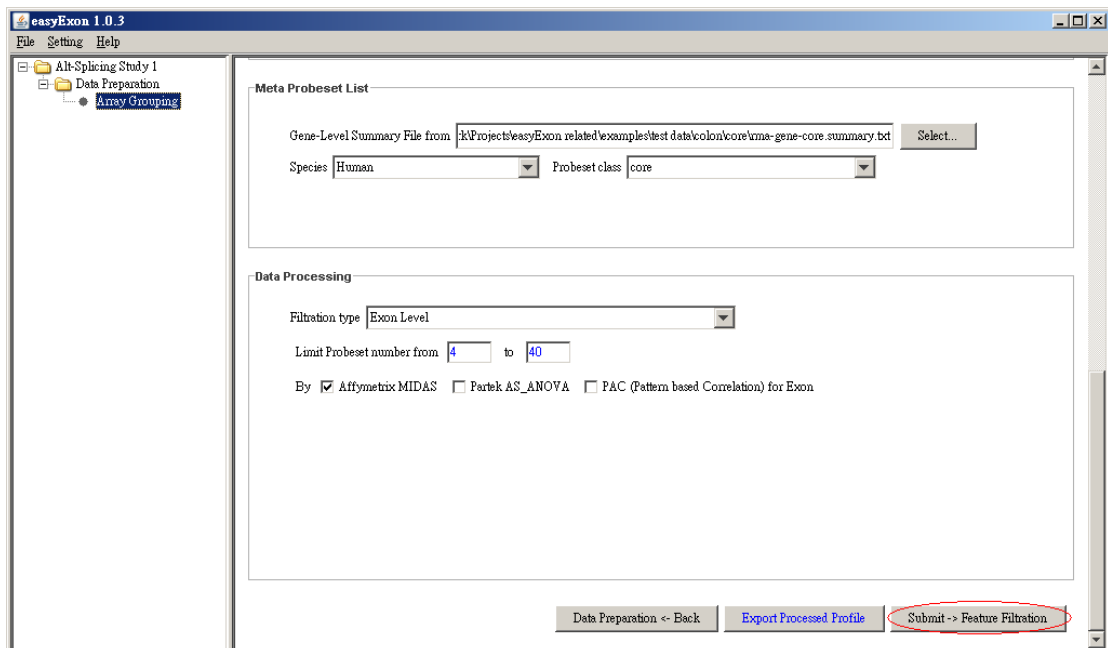
Step 2-4: Export Processed Profile (Optional)

Export the processed file with log stabilization.



The screenshot shows the 'easyExon 1.0.3' application window. The left sidebar contains a tree view with 'Alt-Splicing Study 1' expanded, showing 'Data Preparation' and 'Array Grouping'. The main panel is divided into two sections: 'Meta Probeset List' and 'Data Processing'. In the 'Meta Probeset List' section, the 'Gene-Level Summary File from' field contains the path 'k:\Projects\easyExon-related\example\test data\colon\core\lma-gene-core.summary.bt', and the 'Species' is set to 'Human' and 'Probeset class' to 'core'. The 'Data Processing' section has 'Filtration type' set to 'Exon Level', 'Limit Probeset number from' set to '4' to '40', and the 'By' options include 'Affymetrix MIDAS' (checked), 'Partek AS_ANOVA', and 'PAC (Pattern based Correlation) for Exon'. At the bottom, three buttons are visible: 'Data Preparation <- Back', 'Export Processed Profile' (highlighted with a red oval), and 'Submit -> Feature Filtration'.

Submit to [Step 3!!!](#)



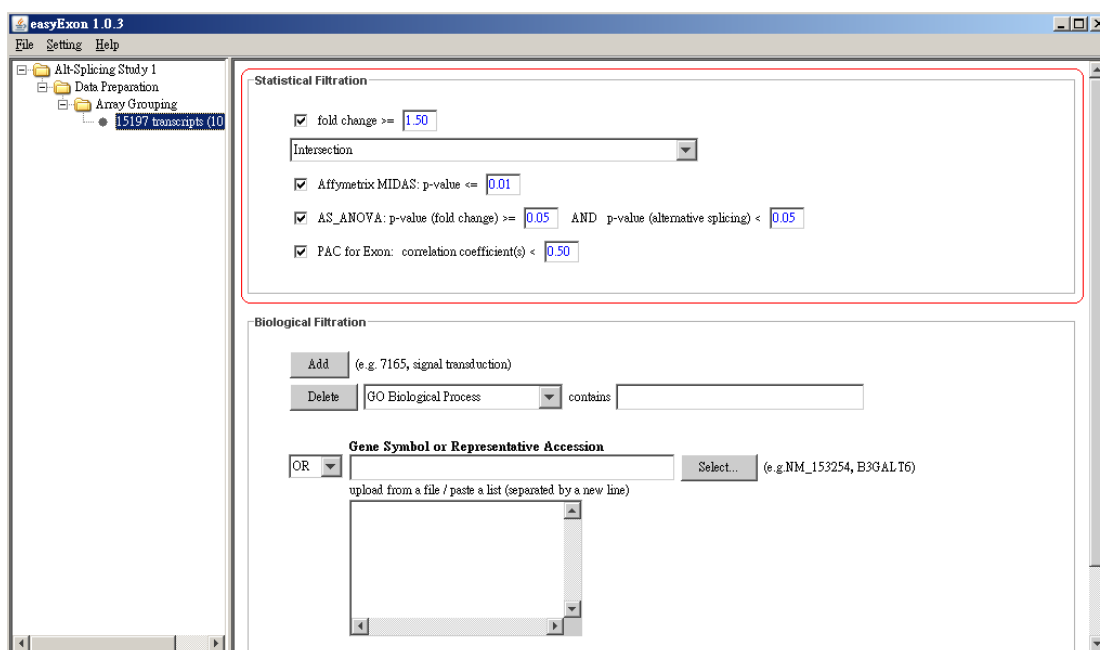
This screenshot is identical to the one above, showing the 'easyExon 1.0.3' interface with the same settings. In this view, the 'Submit -> Feature Filtration' button at the bottom right is highlighted with a red oval, indicating the next step in the process.

Step 3: Feature Filtration

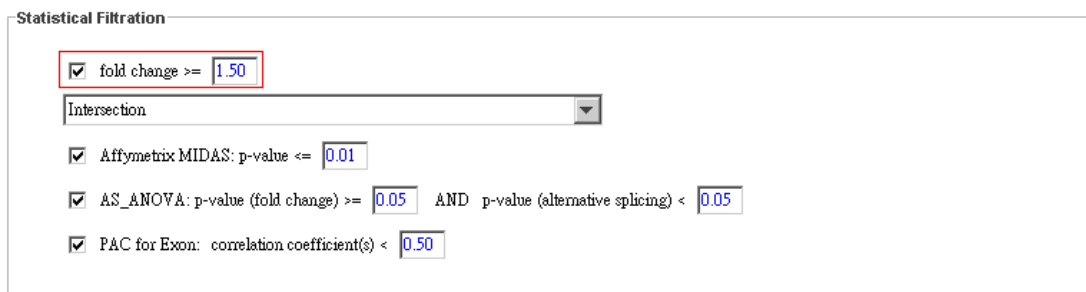
When user chooses Exon Level at the [Step 2-3.a](#), please follow the [Step 3-1.a](#). Otherwise, please follow the Step3-1.b.

Step 3-1.a: Statistical Filtration (Exon Level)

When user chooses method(s) for statistical filtration at the [Step 2-3.a](#), user needs to set the criteria for probeset filtration.



First, user needs to set the fold change.



When user chooses Affymetrix MIDAS at the [Step 2-3.a](#), user needs to set the p -value (calculated by MIDAS).

Statistical Filtration

fold change \geq

Intersection

Affymetrix MIDAS: p -value \leq

AS_ANOVA: p -value (fold change) \geq AND p -value (alternative splicing) $<$

PAC for Exon: correlation coefficient(s) $<$

When user chooses AS_ANOVA at the [Step 2-3.a](#), user needs to set the p -value for fold change and p -value for alternative splicing. Otherwise, skip this step.

Statistical Filtration

fold change \geq

Intersection

Affymetrix MIDAS: p -value \leq

AS_ANOVA: p -value (fold change) \geq AND p -value (alternative splicing) $<$

PAC for Exon: correlation coefficient(s) $<$

When user chooses PAC for Exon at the [Step 2-3.a](#), user needs to set the dissimilarity value of correlation coefficients (the default value is less than 0.5).

Statistical Filtration

fold change \geq

Intersection

Affymetrix MIDAS: p -value \leq

AS_ANOVA: p -value (fold change) \geq AND p -value (alternative splicing) $<$

PAC for Exon: correlation coefficient(s) $<$

When user chooses more than one probeset filtration methods, user can select different ways to combine the filtered results.

Statistical Filtration

fold change \geq

AS_ANOVA: p-value (fold change) \geq AND p-value (alternative splicing) $<$

PAC for Exon: correlation coefficient(s) $<$

User can jump to the [Step 3-2](#).

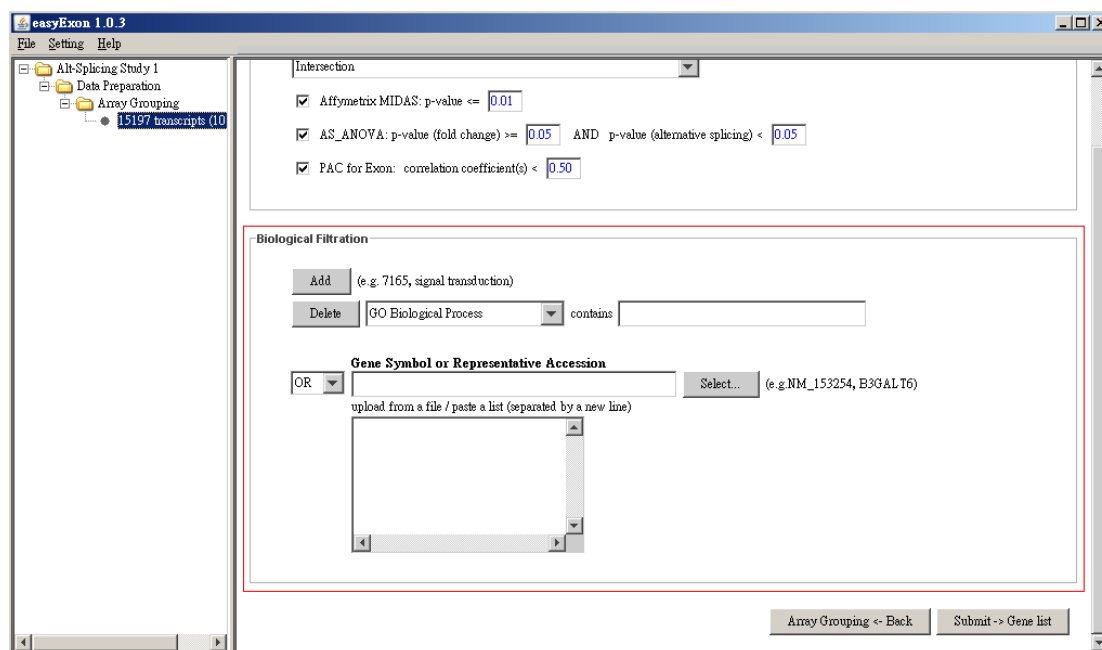
Step 3-1.b: Statistical Filtration (Gene Level)

User sets the significant level of p -value (calculated by t -test) and the minimum value of fold change.

The screenshot shows the 'easyExon 1.0.3' software interface. On the left is a tree view with folders for 'Alt-Splicing Study 1', 'Data Preparation', and 'Array Grouping', and a file named '15197 transcripts (10)'. The main window is titled 'Statistical Filtration' and contains a checkbox labeled 'p-value <= 0.01 AND fold change >= 1.50'. Below this is the 'Biological Filtration' section, which includes an 'Add' button with the text '(e.g. 7165, signal transduction)', a 'Delete' button, a dropdown menu set to 'GO Biological Process', and a 'contains' field. There is also a section for 'Gene Symbol or Representative Accession' with an 'OR' dropdown, a 'Select...' button, and a text area for uploading a list. At the bottom right, there are two buttons: 'Array Grouping <- Back' and 'Submit -> Gene list'.

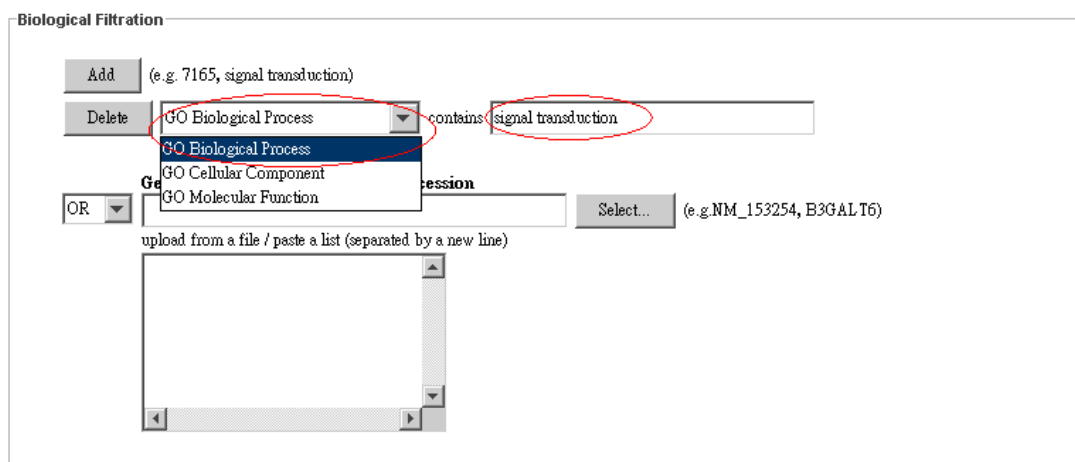
Step 3-2: Biological Filtration

Transcript cluster filtration tool aims to help user narrow down the number of transcripts based on (i) GO term and (ii) gene name or accession number. User can add queries for transcript cluster filtration.



(i) Query by GO term

User can choose the GO term (e.g. Biological Process), and type the key word in the text field (e.g. signal transduction).



When there are more than one queries, user may click the “Add” button to add a new GO query and choose “AND” or “OR” for the query. Ten queries are allowed.

The screenshot shows the 'Biological Filtration' interface. At the top, there is an 'Add' button (circled in red) with the text '(e.g. 7165, signal transduction)'. Below it are two rows of query fields. The first row has a 'Delete' button, a dropdown menu set to 'GO Biological Process', the text 'contains', a text input field with 'signal transduction', and a dropdown menu set to 'OR' (circled in red). The second row has a 'Delete' button, a dropdown menu set to 'GO Biological Process', and the text 'contains'. Below these is a section titled 'Gene Symbol or Representative Accession' with a dropdown menu set to 'OR', a text input field, a 'Select...' button, and the text '(e.g. NM_153254, B3GALT6)'. Below this is a text input field with the placeholder 'upload from a file / paste a list (separated by a new line)' and a large text area with scrollbars.

When user wants to delete one query, user may click the “Delete” button in front of the query.

The screenshot shows the 'Biological Filtration' interface. At the top, there is an 'Add' button with the text '(e.g. 7165, signal transduction)'. Below it are three rows of query fields. The first row has a 'Delete' button, a dropdown menu set to 'GO Biological Process', the text 'contains', a text input field with 'signal transduction', and a dropdown menu set to 'OR'. The second row has a 'Delete' button (circled in red), a dropdown menu set to 'GO Biological Process', and the text 'contains'. The third row has a 'Delete' button, a dropdown menu set to 'GO Biological Process', and the text 'contains'. Below these is a section titled 'Gene Symbol or Representative Accession' with a dropdown menu set to 'OR', a text input field, a 'Select...' button, and the text '(e.g. NM_153254, B3GALT6)'. Below this is a text input field with the placeholder 'upload from a file / paste a list (separated by a new line)' and a large text area with scrollbars.

- (ii) User can query Gene Symbols or Accession numbers of interest by uploading a file or using text area.

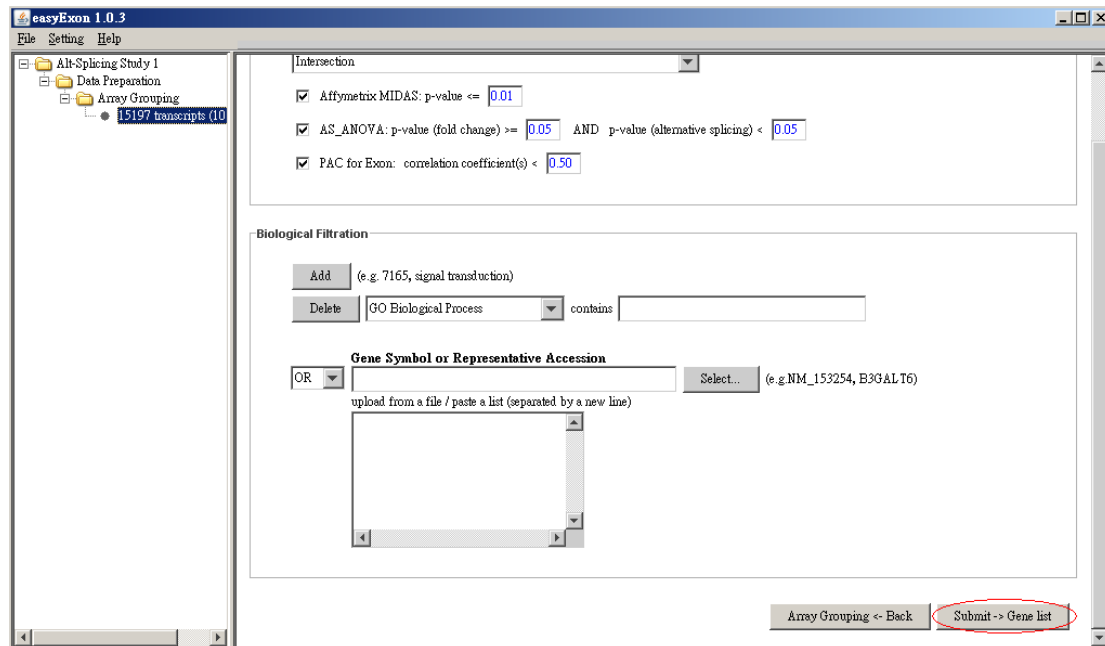
When user sets the GO query above, user may choose “AND” or “OR” to combine with Gene Symbol’s queries.

The screenshot shows the 'Biological Filtration' interface. At the top, there are 'Add' and 'Delete' buttons. The 'Delete' button is next to a dropdown menu set to 'GO Biological Process' and a text input field containing 'signal transduction'. Below this, the 'Gene Symbol or Representative Accession' section is highlighted with a red circle. It features a dropdown menu set to 'OR', a 'Select...' button, and a text input field containing '(e.g. NM_153254, B3GALT6)'. Below the text input field is a text area with the instruction 'upload from a file / paste a list (separated by a new line)'. The text area is currently empty.

Type the key words in the text file or the text area below. Each key word is separated by a new line.

The screenshot shows the 'Biological Filtration' interface. At the top, there are 'Add' and 'Delete' buttons. The 'Delete' button is next to a dropdown menu set to 'GO Biological Process' and a text input field containing 'signal transduction'. Below this, the 'Gene Symbol or Representative Accession' section is highlighted with a red box. It features a dropdown menu set to 'OR', a 'Select...' button, and a text input field containing '(e.g. NM_153254, B3GALT6)'. Below the text input field is a text area with the instruction 'upload from a file / paste a list (separated by a new line)'. The text area contains the text 'NM_153254' and 'B3GALT6'. Red arrows point to the 'Select...' button and the text area, with labels 'upload file here' and 'text area' respectively.

Submit to [Step 4!!!](#)

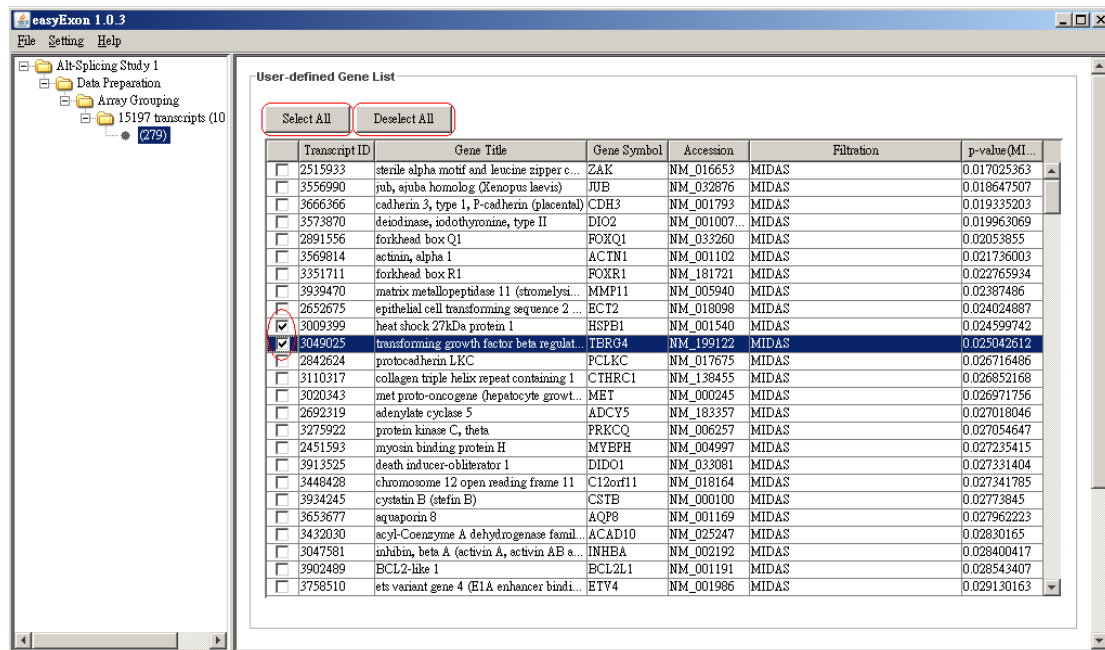


Step 4: Gene List

Step 4-1: User-defined Gene List

Further select genes of interest. Genes in the narrowed down list will be displayed with log scale intensity in the [Step 5](#). The corresponding number of selected genes will be shown in highlighted in the directory tree of the left panel.

User can use the “Select All” or “Deselect All” button to further selection.



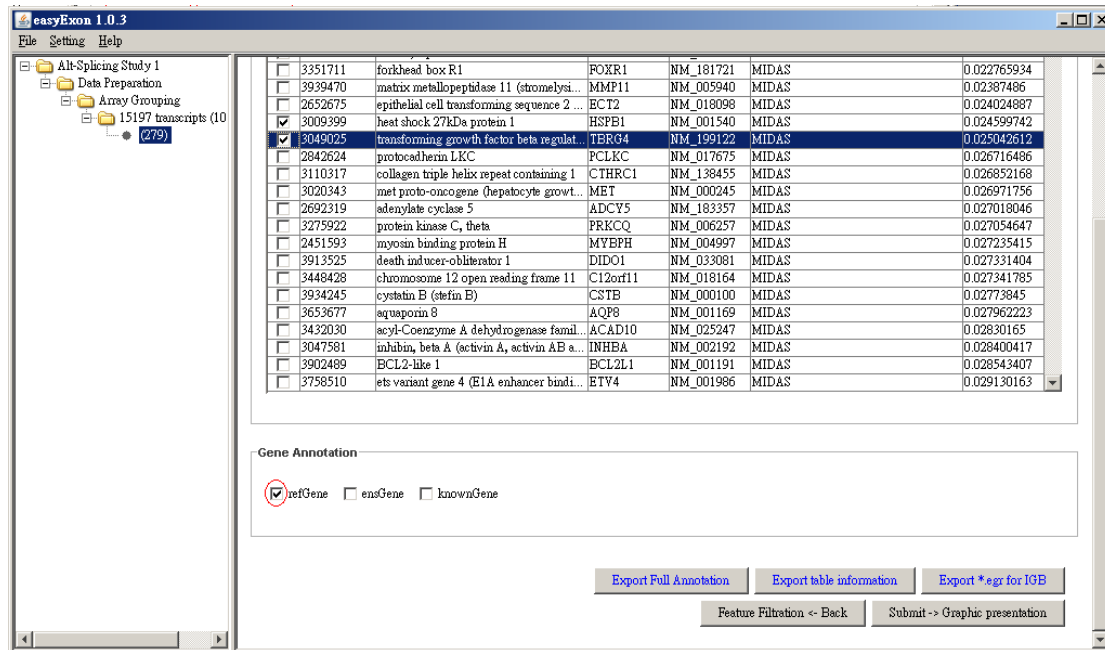
The screenshot shows the 'easyExon 1.0.3' application window. On the left, a directory tree shows 'All-Splicing Study 1' > 'Data Preparation' > 'Array Grouping' > '15197 transcripts (10)' > '(279)'. The main window is titled 'User-defined Gene List' and contains a table with the following data:

Transcript ID	Gene Title	Gene Symbol	Accession	Filtration	p-value(MI...)
<input type="checkbox"/>	sterile alpha motif and leucine zipper c...	ZAK	NM_016653	MIDAS	0.017025363
<input type="checkbox"/>	jub, aruba homolog (Xenopus laevis)	JUB	NM_032876	MIDAS	0.018647507
<input type="checkbox"/>	cadherin 3, type 1, P-cadherin (placental)	CDH3	NM_001793	MIDAS	0.019335203
<input type="checkbox"/>	deiodinase, iodothyronine, type II	DIO2	NM_001007...	MIDAS	0.019963069
<input type="checkbox"/>	forkhead box Q1	FOXQ1	NM_033260	MIDAS	0.02053855
<input type="checkbox"/>	actinin, alpha 1	ACTN1	NM_001102	MIDAS	0.021736003
<input type="checkbox"/>	forkhead box R1	FOXK1	NM_181721	MIDAS	0.022765934
<input type="checkbox"/>	matrix metalloproteinase 11 (stromelysi...	MMP11	NM_005940	MIDAS	0.02387486
<input type="checkbox"/>	epithelial cell transforming sequence 2 ...	ECT2	NM_018098	MIDAS	0.024024887
<input type="checkbox"/>	heat shock 27kDa protein 1	HSPB1	NM_001540	MIDAS	0.024599742
<input checked="" type="checkbox"/>	transforming growth factor beta regulat...	TERG4	NM_199122	MIDAS	0.025042612
<input type="checkbox"/>	protocadherin LKC	PCLKC	NM_017675	MIDAS	0.026716486
<input type="checkbox"/>	collagen triple helix repeat containing 1	CTHRC1	NM_138455	MIDAS	0.026852168
<input type="checkbox"/>	met proto-oncogene (hepatocyte growt...	MET	NM_000245	MIDAS	0.026971756
<input type="checkbox"/>	adenylate cyclase 5	ADCY5	NM_183357	MIDAS	0.027018046
<input type="checkbox"/>	protein kinase C, theta	PRKCQ	NM_006257	MIDAS	0.027054647
<input type="checkbox"/>	myosin binding protein H	MYBPH	NM_004997	MIDAS	0.027235415
<input type="checkbox"/>	death inducer-oblierator 1	DIDO1	NM_033081	MIDAS	0.027331404
<input type="checkbox"/>	chromosome 12 open reading frame 11	CL2orf11	NM_018164	MIDAS	0.027341785
<input type="checkbox"/>	cystatin B (stefin B)	CSTB	NM_000100	MIDAS	0.02773845
<input type="checkbox"/>	aquaporin 8	AQP8	NM_001169	MIDAS	0.027962223
<input type="checkbox"/>	acyl-Coenzyme A dehydrogenase famil...	ACAD10	NM_025247	MIDAS	0.02830165
<input type="checkbox"/>	inhibin, beta A (activin A, activin AB a...	INHBA	NM_002192	MIDAS	0.028400417
<input type="checkbox"/>	BCL2-like 1	BCL2L1	NM_001191	MIDAS	0.028543407
<input type="checkbox"/>	ets variant gene 4 (E1A enhancer bindi...	ETV4	NM_001986	MIDAS	0.029130163

Step 4-2: Gene Annotation

Note: User who chooses “defined by user” for probeset class at the [Step 1.a](#) or the [Step 2-2-2](#) needs to skip this step. In addition, user who chooses “Gene Level” at the [Step 2-3](#) skips this step, too.

User selects the gene annotation definition(s) for graphic presentation.

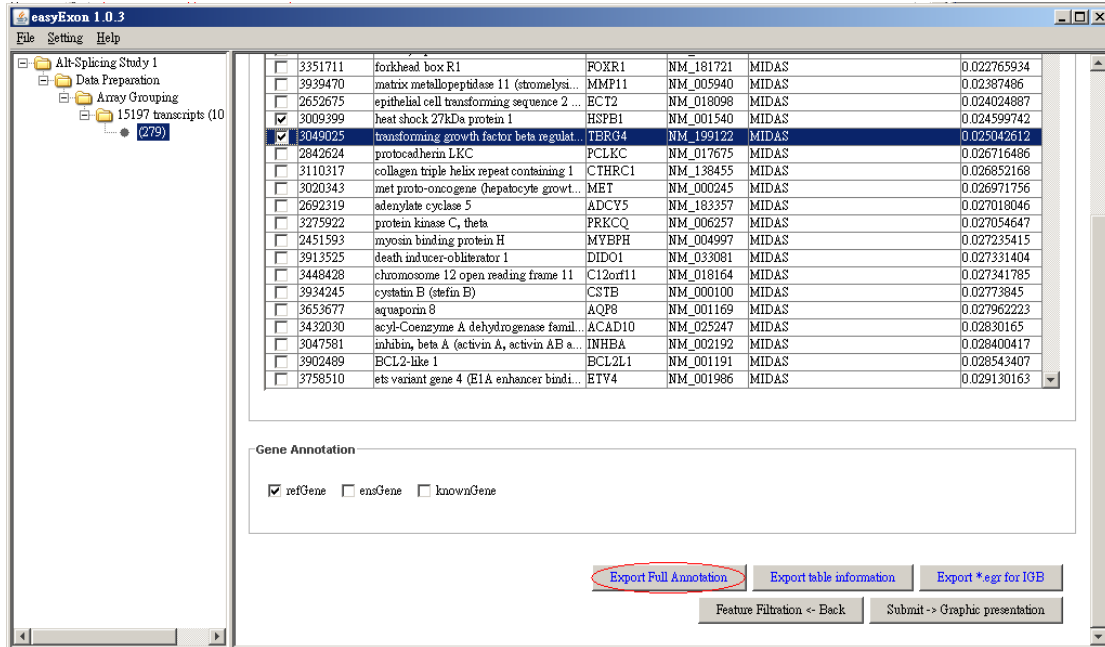


The screenshot shows the 'easyExon 1.0.3' application window. On the left is a tree view with folders for 'Alt-Splicing Study 1', 'Data Preparation', 'Array Grouping', and '15197 transcripts (10)'. The main area contains a table of genes with columns for checkboxes, gene IDs, gene names, symbols, accession numbers, MIDAS values, and scores. The row for 'transforming growth factor beta regulat...' (TERG4) is selected. Below the table is a 'Gene Annotation' section with three radio buttons: 'refGene' (selected), 'ensGene', and 'knownGene'. At the bottom are buttons for 'Export Full Annotation', 'Export table information', 'Export *.egr for IGB', 'Feature Filtration <- Back', and 'Submit -> Graphic presentation'.

Checkbox	Gene ID	Gene Name	Symbol	Accession	MIDAS	Score
<input type="checkbox"/>	3351711	forkhead box R1	FOXK1	NM_181721	MIDAS	0.022765934
<input type="checkbox"/>	3939470	matrix metalloproteinase 11 (stromelysi...	MMP11	NM_005940	MIDAS	0.02387486
<input type="checkbox"/>	2652675	epithelial cell transforming sequence 2 ...	ECT2	NM_018098	MIDAS	0.024024887
<input type="checkbox"/>	3009399	heat shock 27kDa protein 1	HSPB1	NM_001540	MIDAS	0.024599742
<input checked="" type="checkbox"/>	3049025	transforming growth factor beta regulat...	TERG4	NM_199122	MIDAS	0.025042612
<input type="checkbox"/>	2842624	protocadherin LKC	PCLKC	NM_017675	MIDAS	0.026716486
<input type="checkbox"/>	3110317	collagen triple helix repeat containing 1	CTHRC1	NM_138455	MIDAS	0.026852168
<input type="checkbox"/>	3020343	met proto-oncogene (hepatocyte growt...	MET	NM_000245	MIDAS	0.026971756
<input type="checkbox"/>	2692319	adenylate cyclase 5	ADCY5	NM_183357	MIDAS	0.027018046
<input type="checkbox"/>	3275922	protein kinase C, theta	PRKCQ	NM_006257	MIDAS	0.027054647
<input type="checkbox"/>	2451593	myosin binding protein H	MYBPH	NM_004997	MIDAS	0.027235415
<input type="checkbox"/>	3913525	death inducer-obliterator 1	DIDO1	NM_033081	MIDAS	0.027331404
<input type="checkbox"/>	3448428	chromosome 12 open reading frame 11	C12orf11	NM_018164	MIDAS	0.027341785
<input type="checkbox"/>	3934245	cystatin B (stefin B)	CSTB	NM_001000	MIDAS	0.02773845
<input type="checkbox"/>	3653677	aquaporin 8	AQP8	NM_001169	MIDAS	0.027962223
<input type="checkbox"/>	3432030	acyl-Coenzyme A dehydrogenase fami...	ACAD10	NM_025247	MIDAS	0.02830165
<input type="checkbox"/>	3047581	inhibin, beta A (activin A, activin AB s...	INHBA	NM_002192	MIDAS	0.028400417
<input type="checkbox"/>	3902489	BCL2-like 1	BCL2L1	NM_001191	MIDAS	0.028543407
<input type="checkbox"/>	3758510	ets variant gene 4 (E1A enhancer bindi...	ETV4	NM_001986	MIDAS	0.029130163

Step 4-3: Export Full Annotation (Optional)

User may export full annotation by click the button “Export Full Annotation.”



Step 4-4: Export Table Information (Optional)

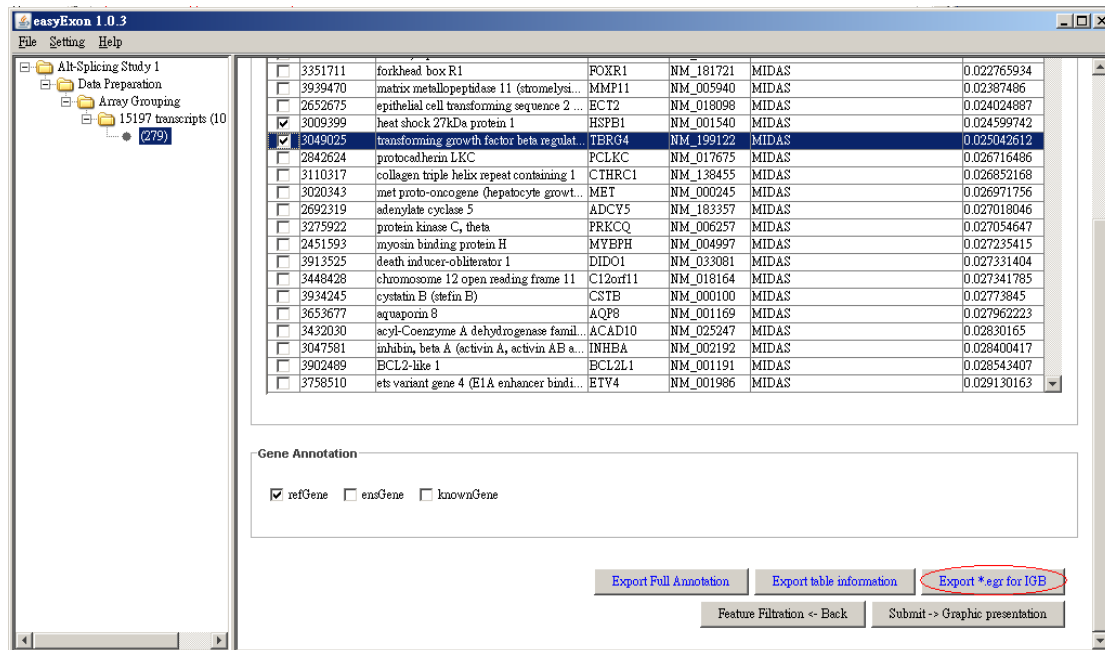
User may export table information by click the button “Export table information.”

The screenshot shows the easyExon 1.0.3 software interface. On the left, there is a tree view showing the project structure: All-Splicing Study 1, Data Preparation, Array Grouping, and 15197 transcripts (10). The main window displays a table of gene annotations. The table has columns for gene ID, gene name, gene symbol, gene accession, gene type, and a numerical value. The row for 'transforming growth factor beta regulat...' is selected. Below the table, there is a 'Gene Annotation' section with three checkboxes: 'refGene' (checked), 'ensGene', and 'knownGene'. At the bottom, there are three buttons: 'Export Full Annotation', 'Export table information' (highlighted with a red circle), and 'Export *.egr for IGB'. Below these buttons are two more buttons: 'Feature Filtration <- Back' and 'Submit -> Graphic presentation'.

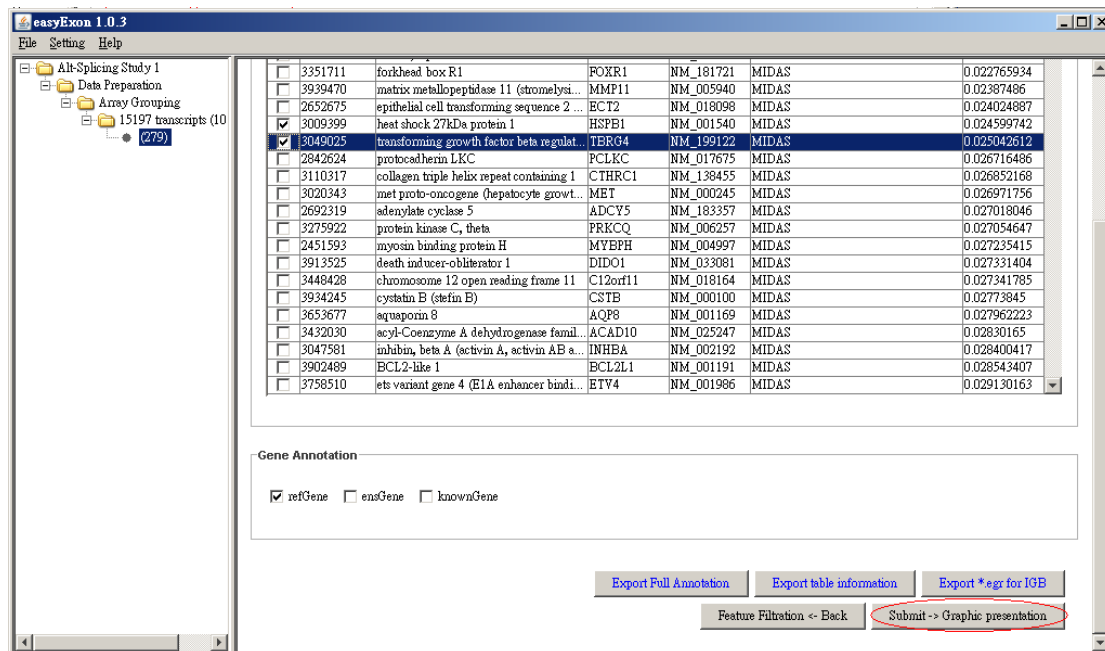
Gene ID	Gene Name	Gene Symbol	Gene Accession	Gene Type	Value
3351711	forkhead box R1	FOXR1	NM_181721	MIDAS	0.022765934
3939470	matrix metalloproteinase 11 (stromelysi...	MMP11	NM_005940	MIDAS	0.02387486
2652675	epithelial cell transforming sequence 2 ...	ECT2	NM_018098	MIDAS	0.024024887
3009399	heat shock 27kDa protein 1	HSPB1	NM_001540	MIDAS	0.024599742
3049025	transforming growth factor beta regulat...	TERG4	NM_199122	MIDAS	0.025042612
2842624	protocadherin LKC	PCLKC	NM_017675	MIDAS	0.026716486
3110317	collagen triple helix repeat containing 1	CTHRC1	NM_138455	MIDAS	0.026852168
3020343	met proto-oncogene (hepatocyte growt...	MET	NM_000245	MIDAS	0.026971756
2692319	adenylate cyclase 5	ADCY5	NM_183357	MIDAS	0.027018046
3275922	protein kinase C, theta	PRKCQ	NM_006257	MIDAS	0.027054647
2451593	myosin binding protein H	MYBPH	NM_004997	MIDAS	0.027235415
3913525	death inducer-obliterator 1	DIDO1	NM_033081	MIDAS	0.027331404
3448428	chromosome 12 open reading frame 11	C12orf11	NM_018164	MIDAS	0.027341785
3934245	cystatin B (stefin B)	CSTB	NM_000100	MIDAS	0.02773845
3653677	aquaporin 8	AQP8	NM_001169	MIDAS	0.027962223
3432030	acyl-Coenzyme A dehydrogenase fami...	ACAD10	NM_025247	MIDAS	0.02830165
3047581	inhibin, beta A (activin A, activin AB s...	INHBA	NM_002192	MIDAS	0.028400417
3902489	BCL2-like 1	BCL2L1	NM_001191	MIDAS	0.028543407
3758510	ets variant gene 4 (E1A enhancer bindi...	ETV4	NM_001986	MIDAS	0.029130163

Step 4-5: Export *.egr for IGB (Optional)

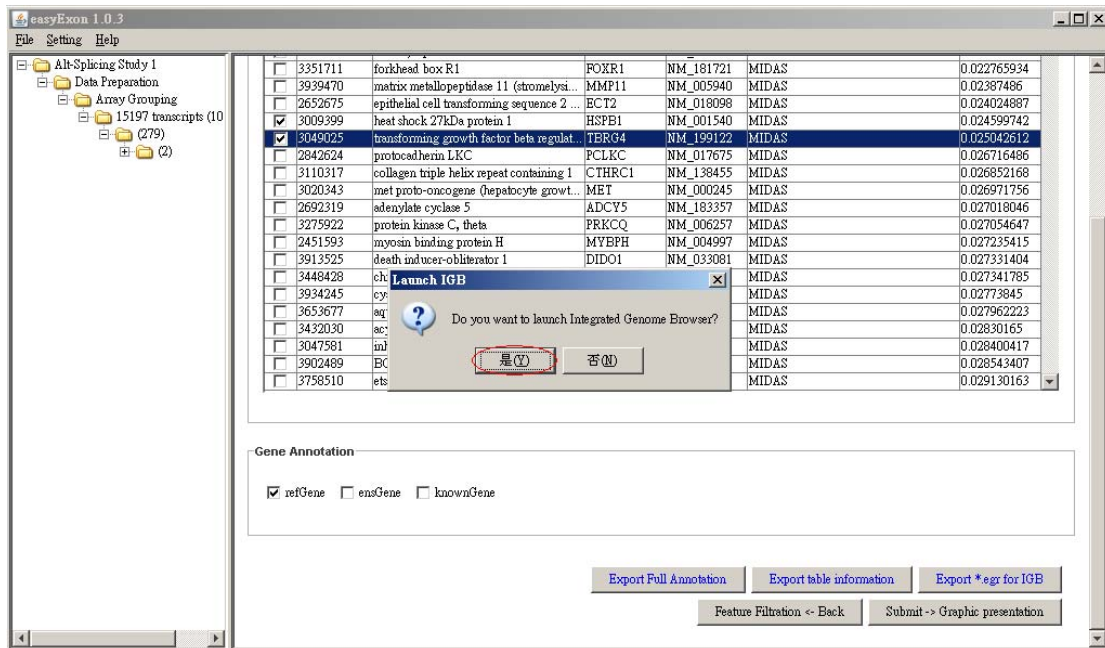
User may export a *.egr file for Integrated Genome Browser (IGB) by clicking the button “Export *.egr for IGB.”



Submit to [Step 5!!!](#)



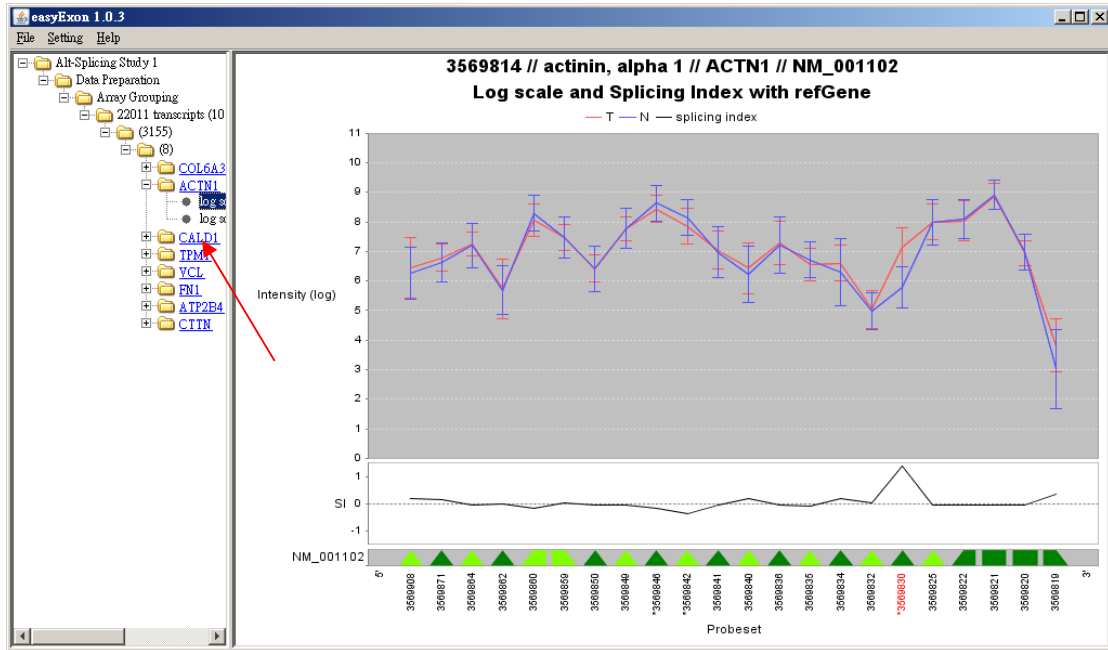
When user needs to launch IGB for browsing the transcript information, user needs to choose “YES” to launch the IGB.



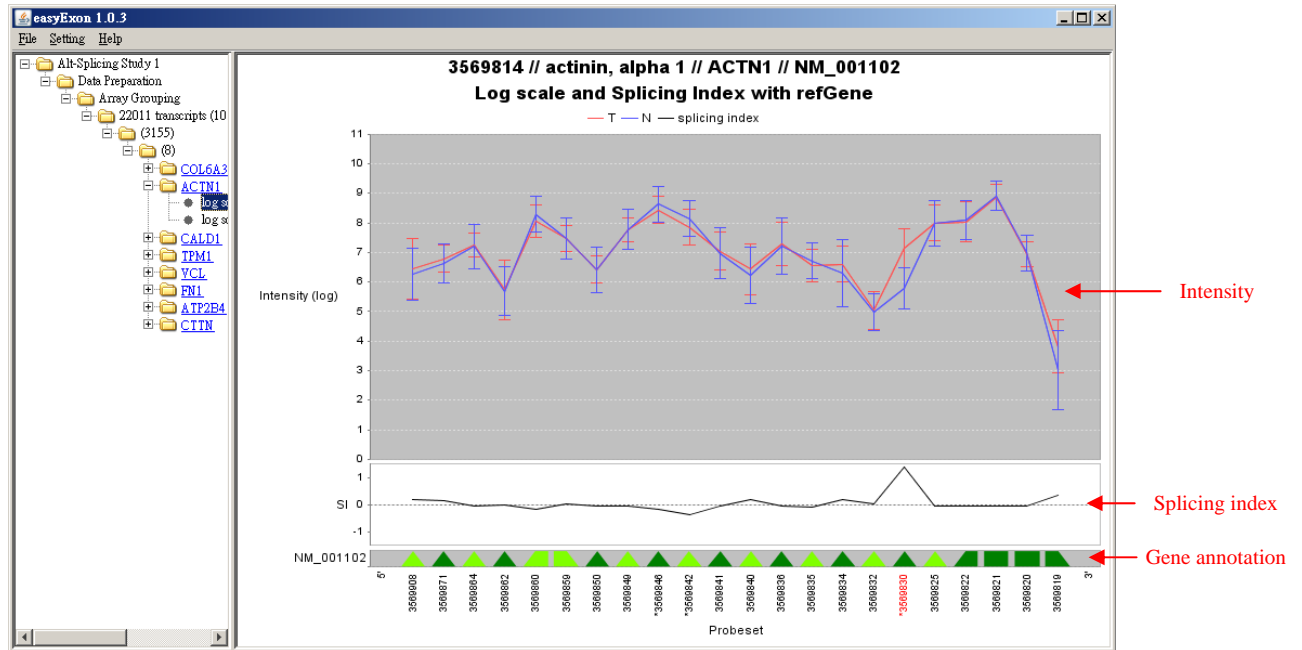
Step 5: Graphic Presentation

Step 5.a: Graphic Presentation for Exon Level

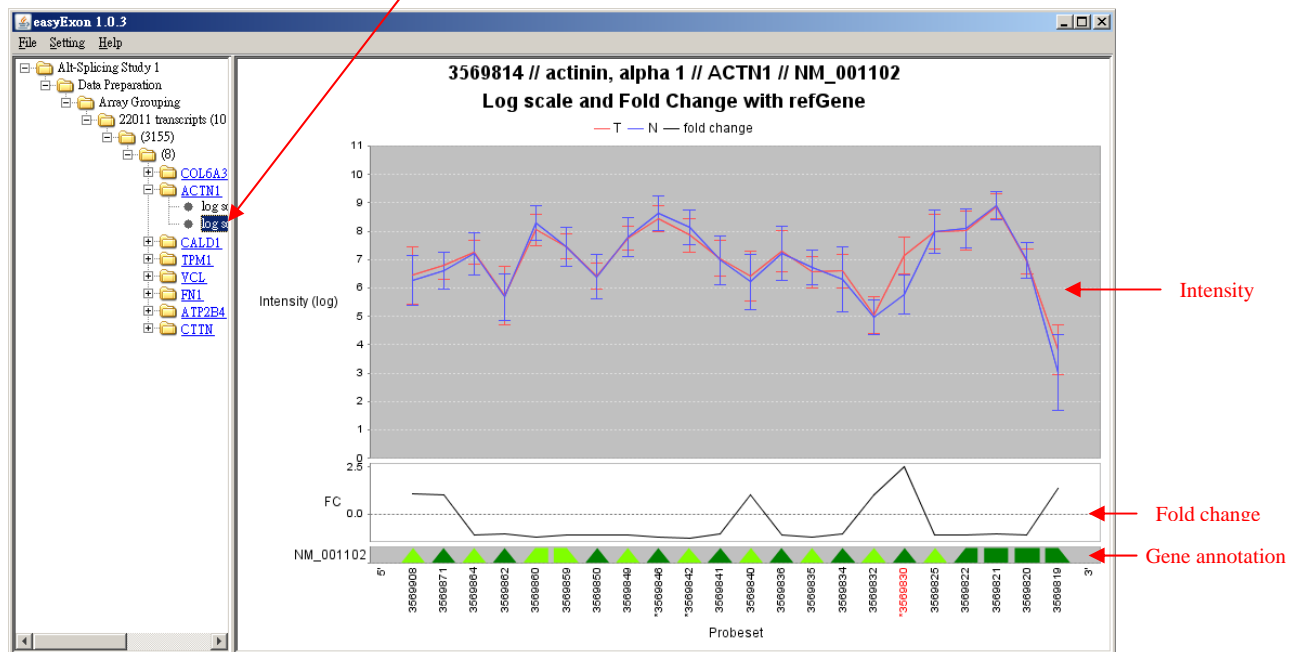
When user launches IGB, user may click the hyper link in the left panel to browse the transcript in IGB.



When user clicks the gene of interest in the left panel, a graphic representation for exons in the corresponding transcript cluster will be displayed in the right panel. The x-axis is the probeset id for each exon, and y-axis represents log scale intensity and splicing index. Neighboring probesets with the same color in gene annotation means that those probesets are in the same exon.



User may select log scale and fold change on the tree of left panel, intensity of log scale and fold change will be displayed on the right panel.



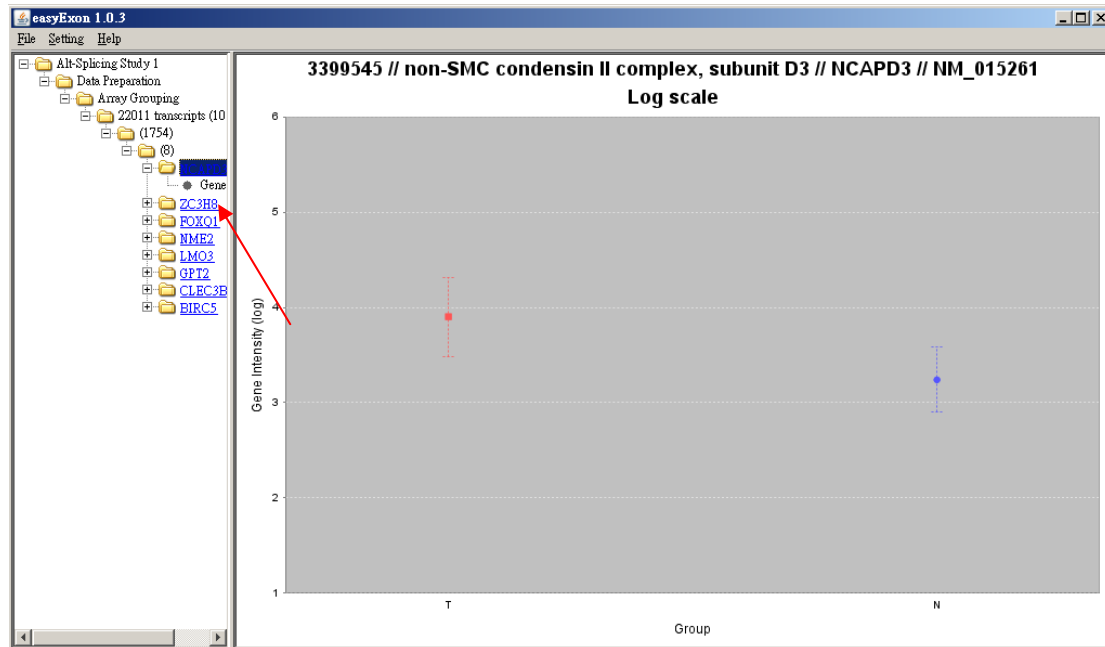
User can modify the Filtration ([Step 3](#)), and gene signature and gene annotation ([Step 4](#)) repeatedly.

Note: Labels for probeset id of x-axial.

- (i) Probeset id labeled in red: fold change is greater than the threshold set in [Step 3-1.a](#).
- (ii) Probeset id labeled in light gray: p -value from DABG file is greater than the threshold set in [Step 1.b](#).
- (iii) Probeset id labeled with “*”: p -value computed by MIDAS is smaller than the threshold set in [Step 3-1.a](#).
- (iv) Probeset id labeled with “PAC”: Patter based correlation is smaller than the threshold set in [Step 3-1.a](#).

Step 5.b: Graphic Presentation for Gene Level

When user launches the IGB, user may click the hyper link to browse the transcript in the IGB.



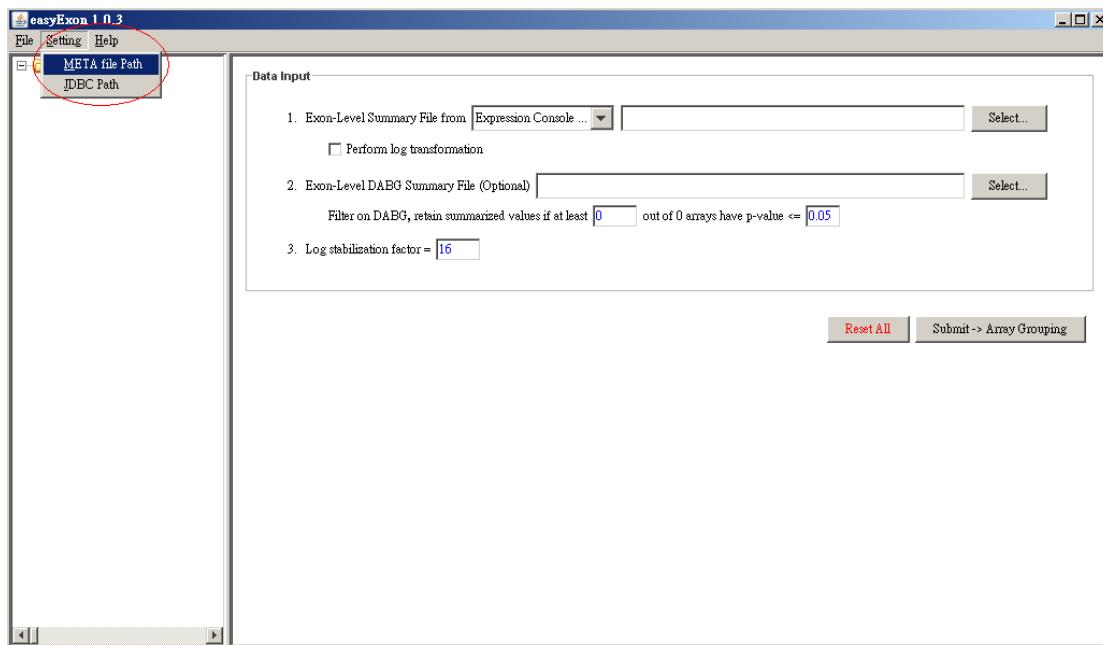
User can modify the Filtration ([Step 3](#)), and gene signature and gene annotation ([Step 4](#)) repeatedly.

2. Environment Setting

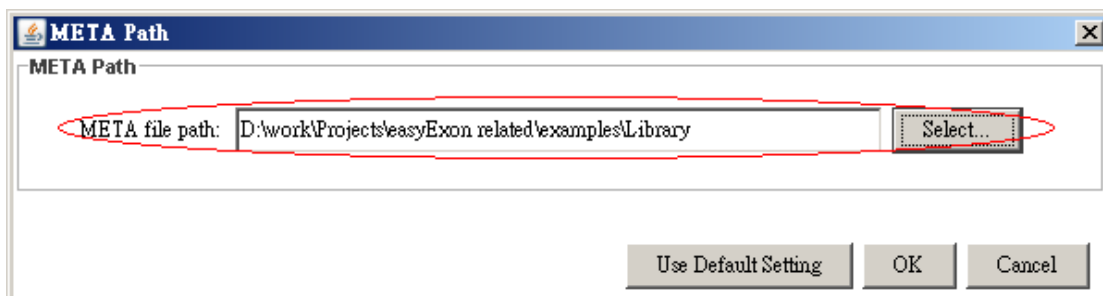
User may download our database and meta probeset files from our homepage (<http://microarray.ym.edu.tw:8080/easyexon/index.jsp?mode=support>) and set the path of those downloading files and database.

2-1: Setting the Path of Meta Probeset Files

User may set the path of meta probeset files by clicking “Setting” on Menu Bar and select “META file Path.”

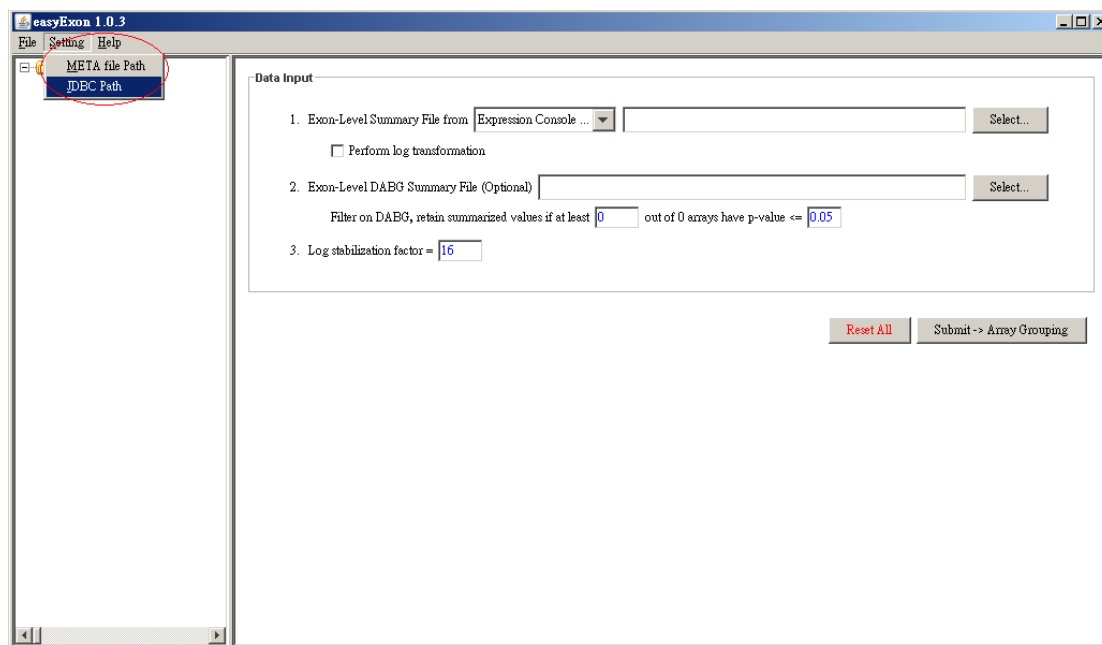


Select the path of meta probeset files directory which includes *.mps files.

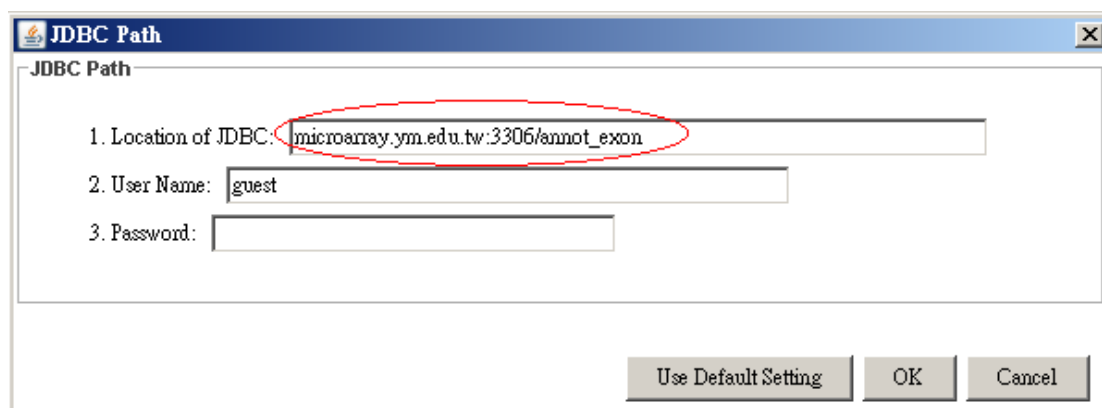


2-2: Setting the Path of Database

User may set the path of database by clicking “Setting” on Menu Bar and select “JDBC Path.”



Type the location of database which includes IP address (or domain name), port number, and the name of database. (E.g.: microarray.ym.edu.tw:3306/annot_exon)



Type the user name and password.

JDBC Path

JDBC Path

1. Location of JDBC: microarray.ym.edu.tw:3306/annot_exon

2. User Name: guest

3. Password: *****

Use Default Setting OK Cancel