

Importing Affymetrix CEL Files with TAC for BRB-ArrayTools

Version 1.0

By

Dr. Qian Xie

BRB-ArrayTools Development Team

arraytools@emmes.com

Jan. 2021

Contents

1. Introduction	1
2. Download and install TAC Software.....	2
3. Convert CEL files to TXT file with TAC	4
4. Import the TXT file in BRB-ArrayTools	7
5. Examples of other array types	11
5.1. Affymetrix Mouse Gene 1.0 ST Array.....	12
5.2. Affymetrix HT HG-U133+ PM Array	13

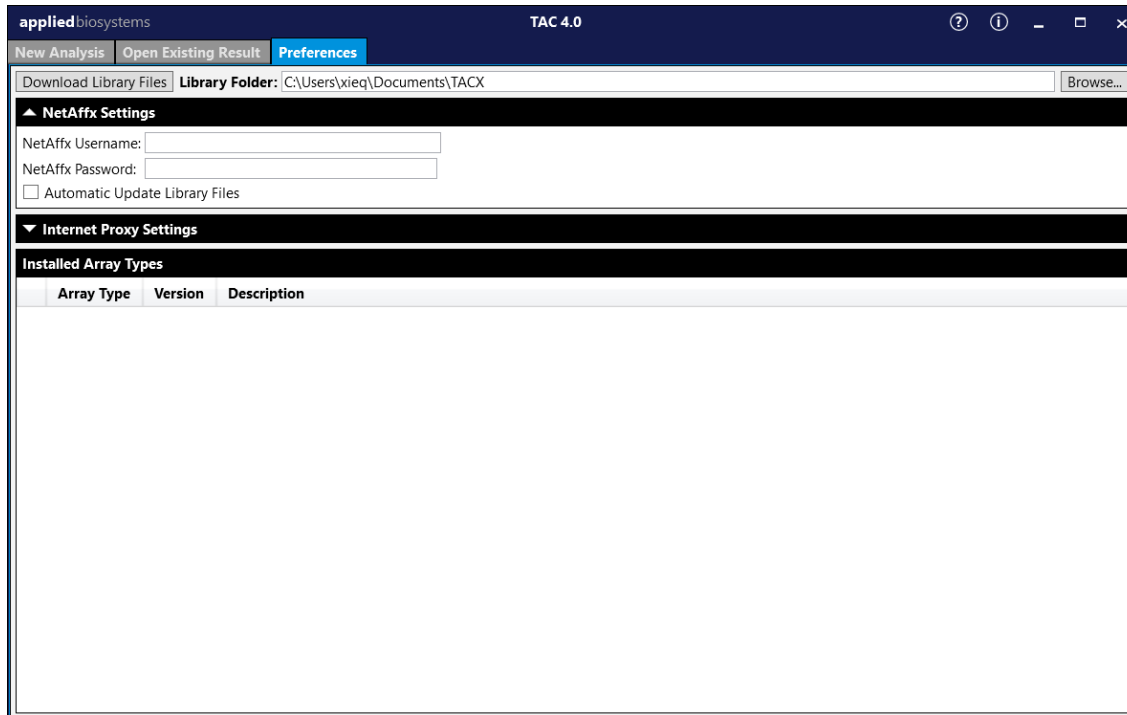
1. Introduction

[Affymetrix Expression Console Software is now part of the Transcriptome Analysis Console \(TAC\) Software](#). In this instruction, you use the TAC software to convert Affymetrix CEL files to a probe summarization TXT file and then import the TXT file in BRB-ArrayTools to collate a BRB-ArrayTools project.

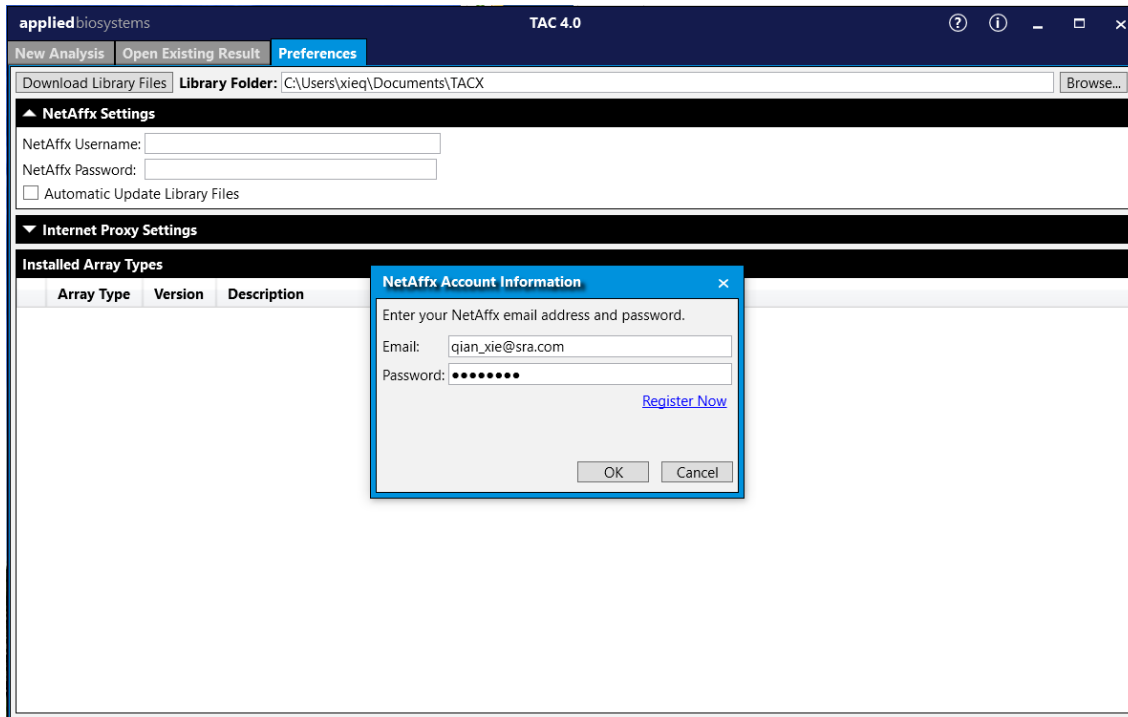
2. Download and install TAC Software

You can download the TAC software from the website <https://www.thermofisher.com/us/en/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/affymetrix-transcriptome-analysis-console-software.html>. Current version is TAC 4.0.2.15 (Nov. 2020). After installing the TAC software, you need to set up the software by clicking the tab “Preferences”:

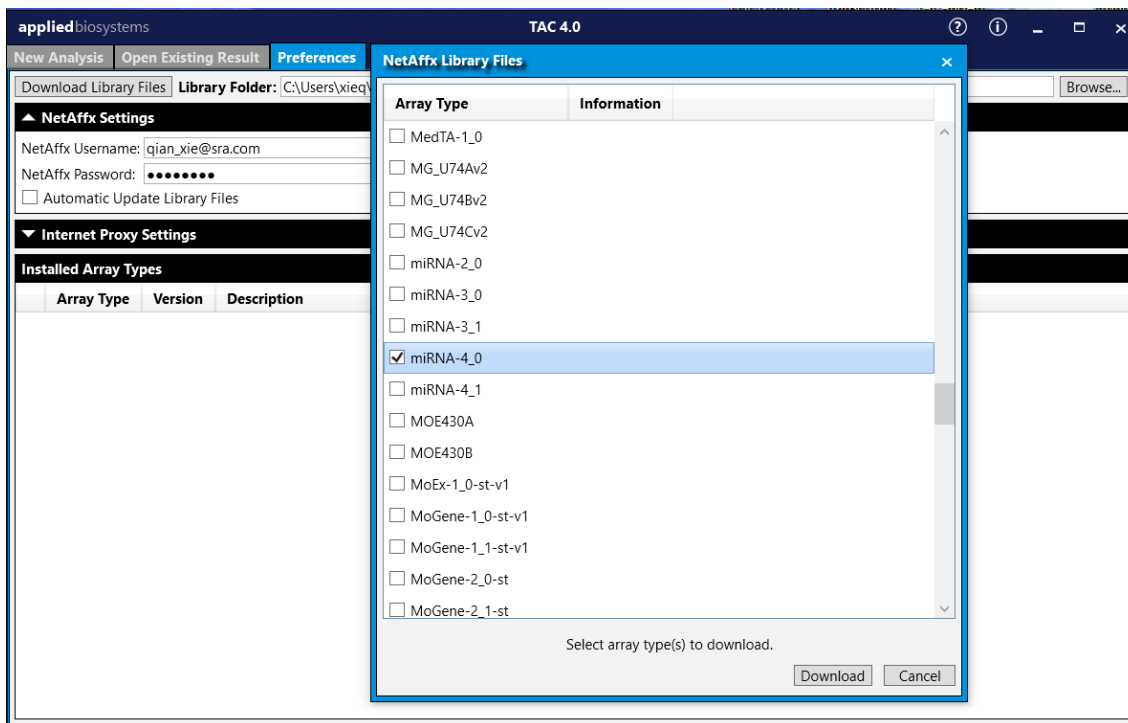
- Set Library Folders: C:/Users/%username%/Documents/TACX



- Download the Library files for your array type:
 - Click the “Download Library Files” button, a NetAffx Account Information window pops up. If you have an account, you may enter your account information here and click “OK”. If you do not have an account, you need to click “Register Now” to create an account in your default browser and then enter here.



- The NetAffx Account Information will automatically fill in the NetAffx Settings field and a window pops up for selecting array types. Select miRNA-4_0 and click the “Download” button. The library files for the array type will be downloaded and the relevant information will keep under the “Installed Array Types” field.

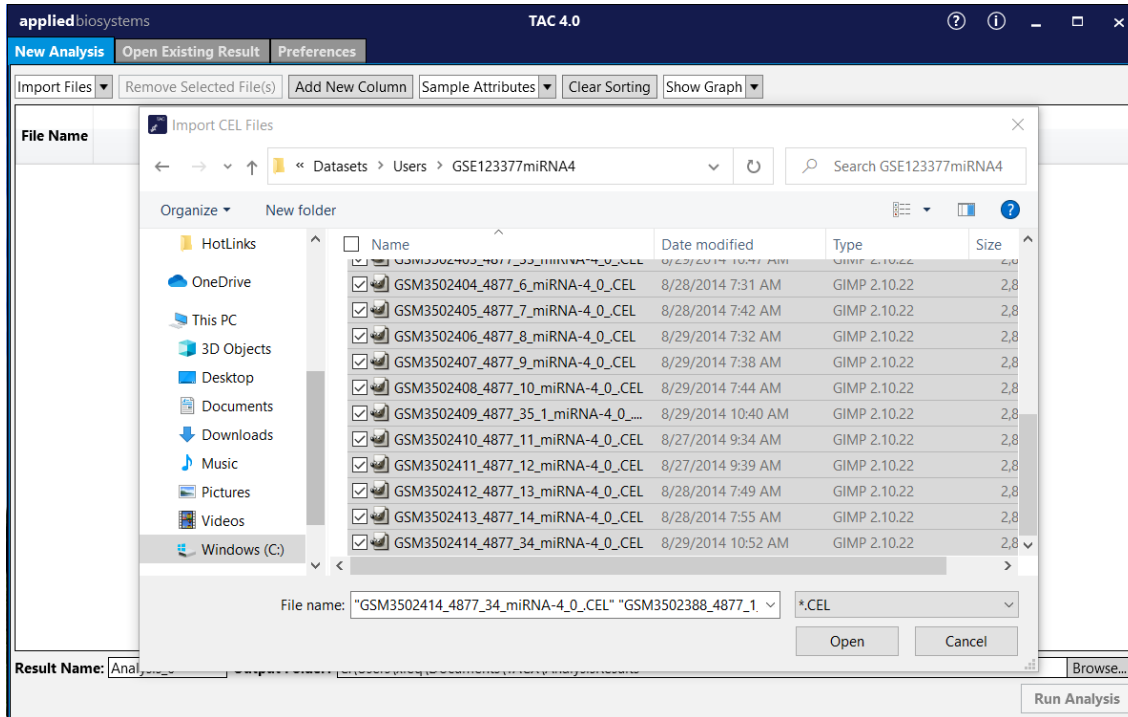


3. Convert CEL files to TXT file with TAC

You keep all CEL files in a folder. Here uses the NCBI GEO GSE123377 dataset as an example. You may download the dataset <https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE123377&format=file>.

You click the tab “New Analysis” to start a new study.

- Import Files > Import CEL Files: a file browser window pops up > Select CEL files (NCBI GEO GSE123377) > Open.



- In the column “Condition”, change Type: Comparison and assign the column values. Half of the files are “1” and another half of the files are “2”. These assignments are random. It is for the program to process. Click the lower-left button “Run Analysis”.

applied biosystems TAC 4.0

New Analysis | Open Existing Result | Preferences

Array Type: miRNA-4_0 Analysis Type: Expression (Gene) Summarization: RMA+DABG (All Organisms) Version: version 1

Import Files | Remove Selected File(s) | Add New Column | Sample Attributes | Clear Sorting | Show Graph

File Name (27)	Condition	
	Type:	Comparison
<input type="checkbox"/> GSM3502388_4877_1_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502389_4877_2_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502390_4877_3_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502391_4877_4_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502392_4877_5_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502393_4877_16_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502394_4877_17_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502395_4877_18_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502396_4877_19_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502397_4877_20_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502398_4877_21_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502399_4877_22_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502400_4877_23_miRNA-4_0_CEL	1	
<input type="checkbox"/> GSM3502401_4877_24_miRNA-4_0_CEL	2	
<input type="checkbox"/> GSM3502402_4877_32_miRNA-4_0_CEL	2	

Result Name: Analysis_1 Output Folder: C:\Users\xieq\Documents\TAC\AnalysisResults

Algorithm Settings | Comparison Setup Wizard | Run Analysis

- When the analysis finishes, a new window pops up with the results.

applied biosystems Analysis_1_tac

Sample QC View | Summary View | Gene View

Sample Table

Apply View | Filters | Show/Hide Columns | Export | Add Column | Reanalyze Samples | Create Line Graph

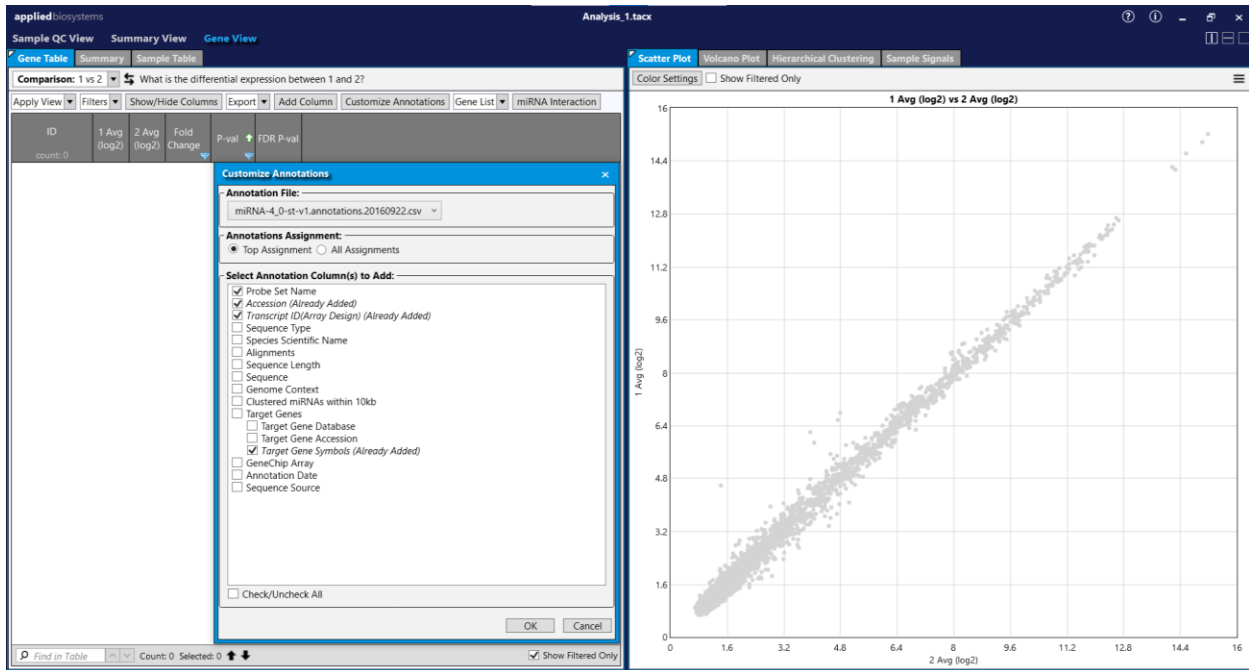
File Name	Hybridization Controls (P) Threshold	Hybridization Controls (S) Threshold	Spike-In Controls Threshold	Condition
GSM3502388_4877_1_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502389_4877_2_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502390_4877_3_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502391_4877_4_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502392_4877_5_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502393_4877_16_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502394_4877_17_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502395_4877_18_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502396_4877_19_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502397_4877_20_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502398_4877_21_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502399_4877_22_miRNA-4_0_...	Pass	Pass	Pass	1
GSM3502400_4877_23_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502401_4877_24_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502402_4877_32_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502403_4877_33_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502404_4877_6_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502405_4877_7_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502406_4877_9_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502407_4877_9_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502408_4877_10_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502409_4877_35_1_miRNA-4_...	Pass	Pass	Pass	2
GSM3502410_4877_11_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502411_4877_12_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502412_4877_13_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502413_4877_14_miRNA-4_0_...	Pass	Pass	Pass	2
GSM3502414_4877_34_miRNA-4_0_...	Pass	Pass	Pass	2

PCA Mapping 41.1% (CHP)

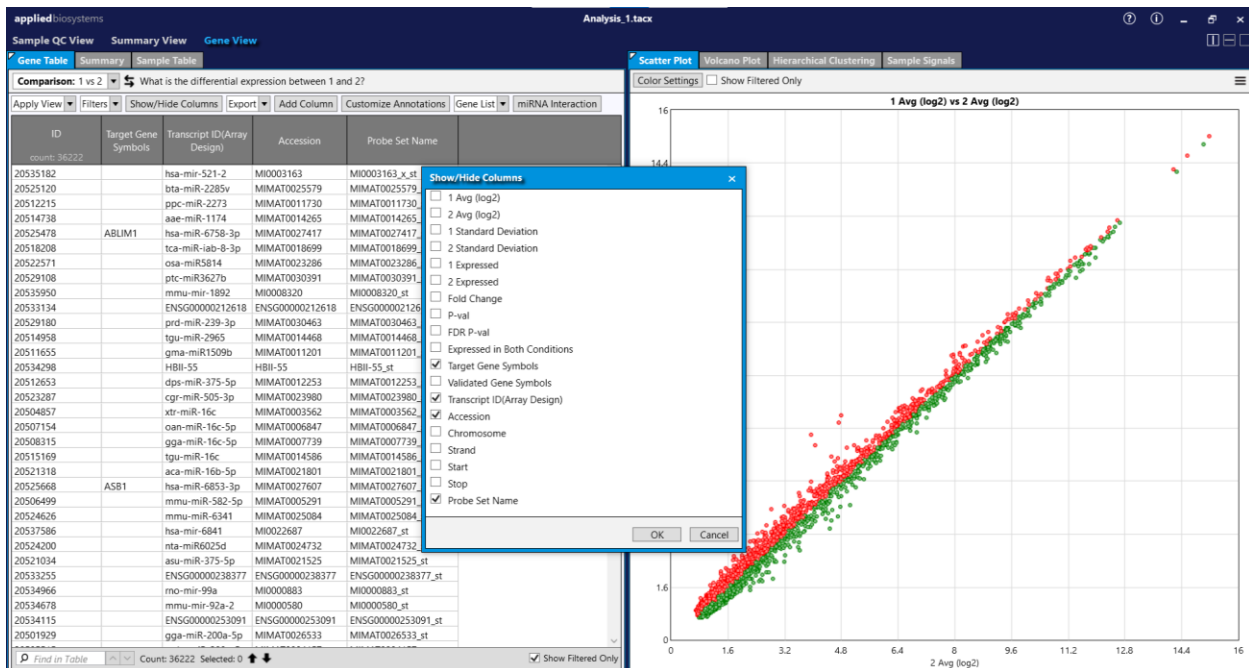
Condition: 1 (blue), 2 (red)

Left Mouse Action: Rotate | Right Mouse Action: Lasso | Zoom: | Size: |

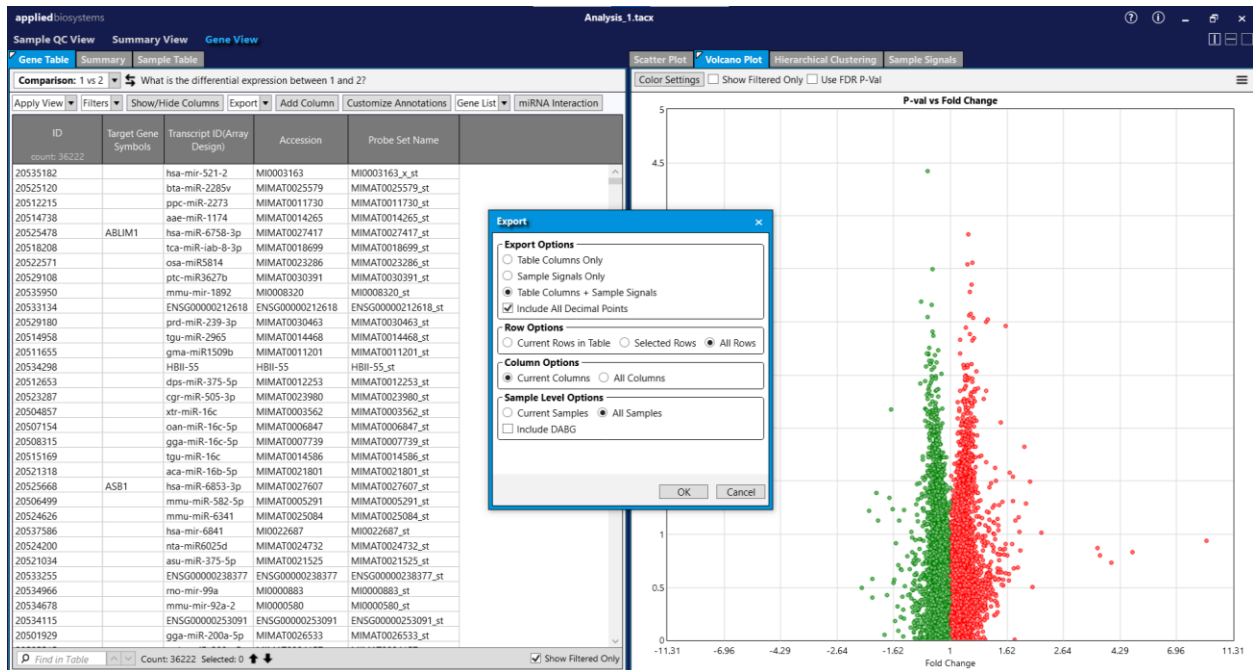
- Click the “Gene View” in the new window. And click “Customize Annotations” to select information columns for output. In the popped up window, check “Top Assignment” to pick the first subfield in each annotation column and select gene information columns you need. Click “OK”.



- Click “Show/Hide Columns”, a window will pop up. You can select columns for output. Click “OK”.



- Click “Export” > Export..., an Export window pops up. You can select the output contents. Click “OK”.

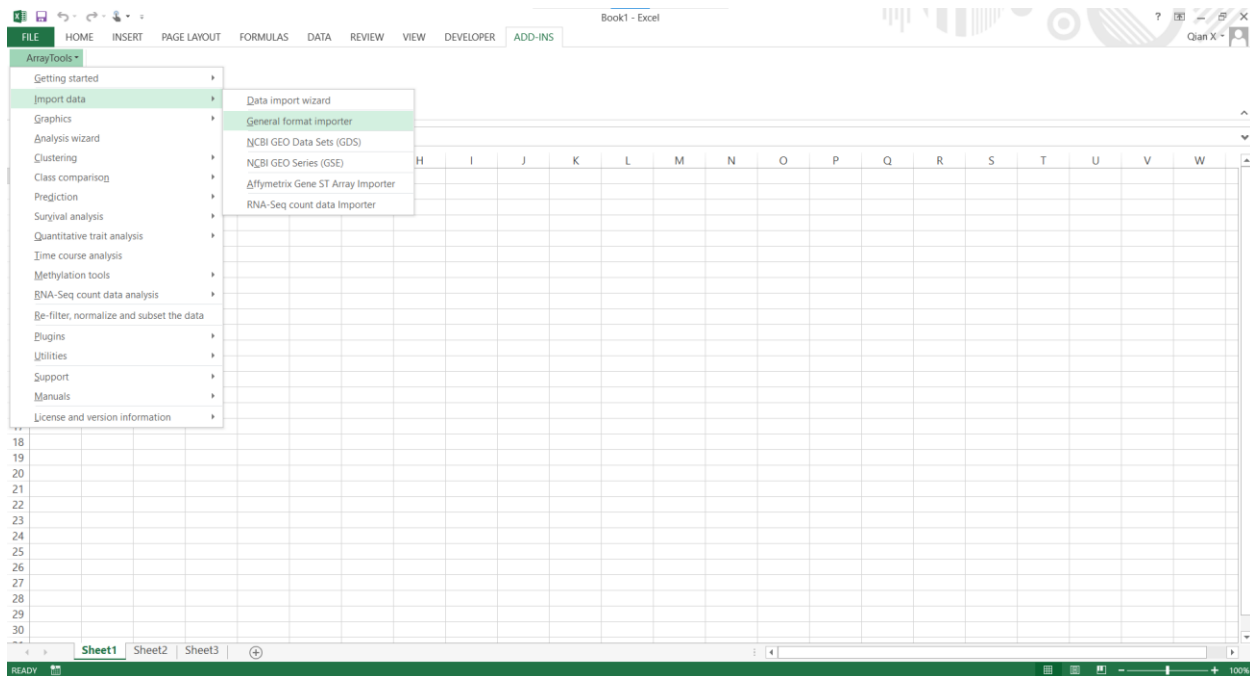


- A file browser window pops up. Give an Export File name: GSE123377.txt. Click “Save”. The probe summarization TXT file will be saved onto your local drive.

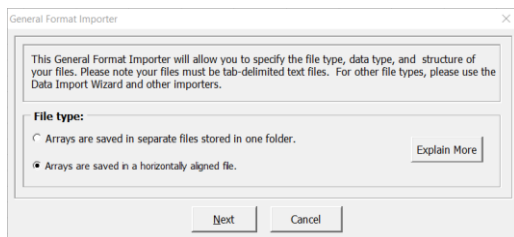
4. Import the TXT file in BRB-ArrayTools

You may modify the title names of the array signals in the TXT file with a text editor before importing it in BRB-ArrayTools. For example, you may shorten the title name “GSM3502388_4877_1_miRNA-4_0_.rma-dabg.chp Signal” to be “GSM3502388”.

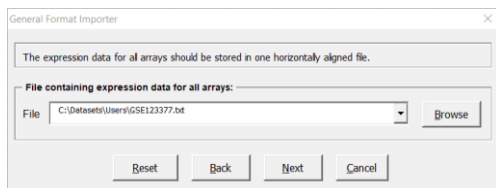
Start Excel and import the TXT file with the BRB-ArrayTools “General format importer”.



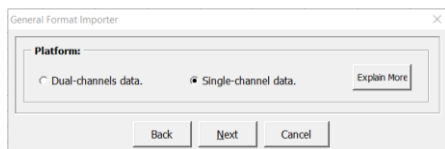
Select "Arrays are saved in a horizontally aligned file." And click "Next".



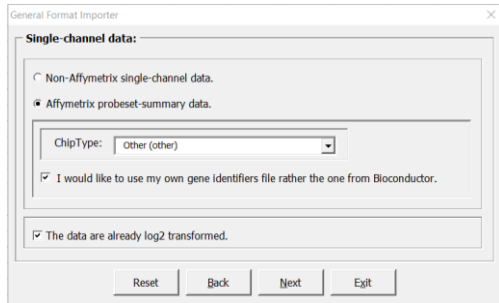
Browse for the TXT file and click "Next".



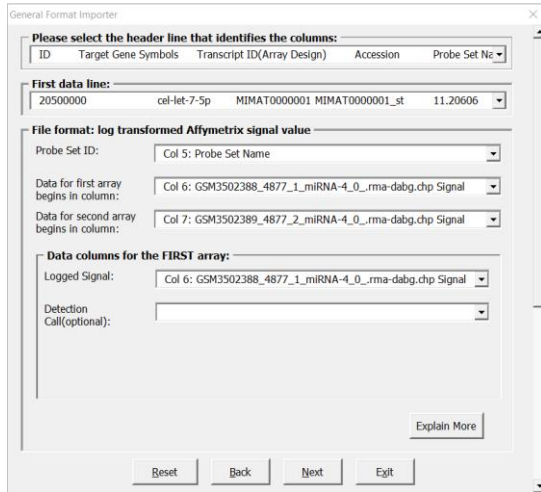
Select "Single-channel data." And click "Next".



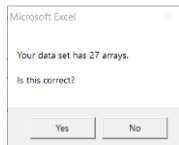
Select "Affymetrix probeset-summary data." And click "Next".



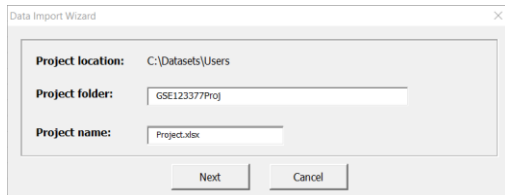
Select correct lines and columns for the fields. And click “Next”.



Confirm by clicking “Yes”.



Give a BRB-ArrayTools project folder name or use the default one and click “Next”.



Select correct columns for different gene identifiers and click “Next”.

Data Import Wizard

Please specify the location of your gene identifiers:

The identifiers are stored alongside the expression data. The identifiers are stored in a separate file.

Please select your Gene identifiers file:

File: Header line #:

Please select the available gene identifiers:

Unique ID (Well or Spot ID, etc.): Col 5: Probe Set Name EntrezId:

Clone ID (IMAGE or ATCC ID, etc.): Gene Name, Title or Description: Col 1: ID

UniGene Cluster ID: GenBank Accession: Col 4: Accession

Gene Symbol: Col 2: Target Gene Symbols Map Location:

Ensembl ID: microRNA ID: Col 3: Transcript ID(Array Desi)

Annotate the project with these gene ids, instead of using the data from SOURCE database. Organisms: Human

Check “I do not have an experiment descriptor file. Please create a template with just array Ids.” And click “Next”.

Data Import Wizard

I do not have an experiment descriptor file. Please create a template with just array ids.

Experiment descriptor file:

File: C:\Datasets\Users\ExpDescFile.xls

Confirm by clicking “OK”.

Refilter, normalize and subset the data

1. Spot filters 2. Normalization 3. Gene filters

Background adjustment is performed before the intensity filtering and the averaging of replicate spots is done on filtered data.

Apply background adjustment.

Intensity Filter:

EXCLUDE the spot if the intensity is below the minimum.

THRESHOLD the intensity at the minimum value if the intensity is below the minimum.

Intensity minimum: 10

Detection Call

EXCLUDE the probe set if the Detection Call contains:

Any one of the following values (comma-separated): A,M,P,No Ca

A

Average the replicate spots within an array.

Use a common reference design.

Confirm by clicking “OK”.

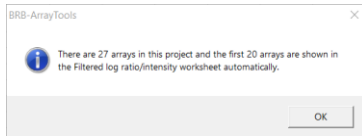
Microsoft Excel

The 'Percent Absent' gene filter will be turned off because your project does not contain the Detection Call data.

Confirm by clicking “OK”.

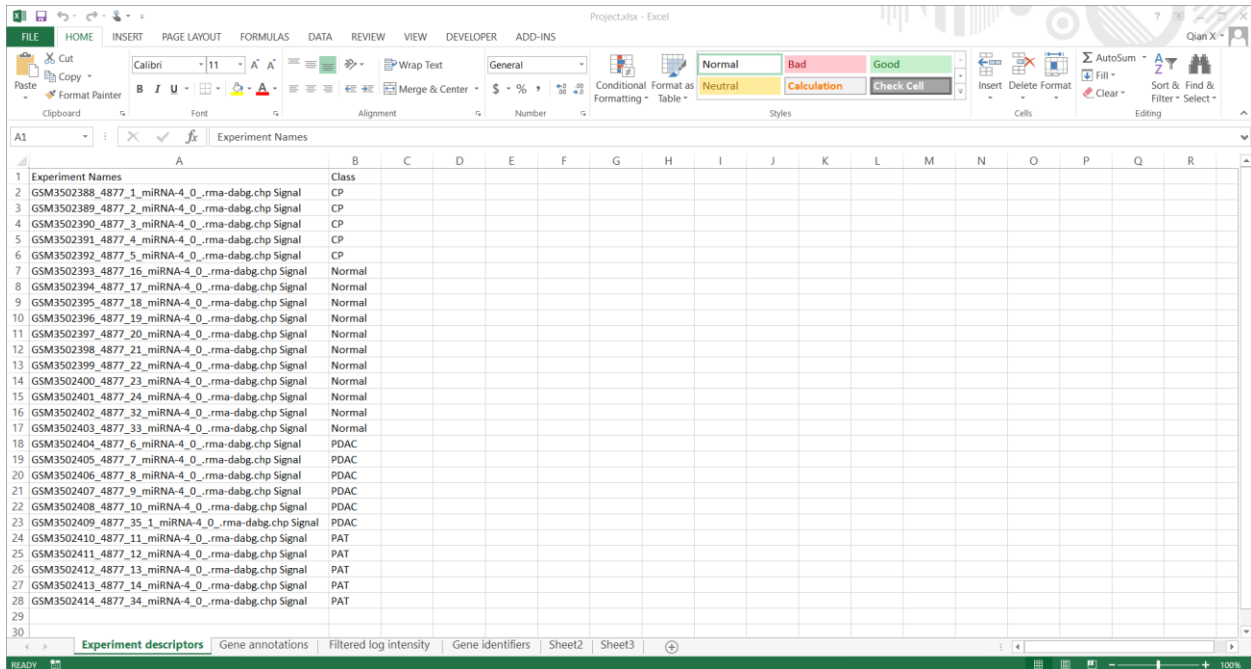


Confirm by clicking “OK”.



Now the BRB-ArrayTools project is collated successfully.

You may edit the worksheet “Experiment descriptors” to add sample information. You can run all the analysis tools in BRB-ArrayTools.



5. Examples of other array types

Other types of arrays data can be imported into BRB-ArrayTools by following the same steps given above. Different array types may have different annotation items for output. In the following examples, popular gene information columns are selected. You may select different columns for your needs.

5.1. Affymetrix Mouse Gene 1.0 ST Array

NCBI GEO GSE54773: Affymetrix Mouse Gene 1.0 ST Array. You can follow the above steps to convert CEL files to a probe summarization TXT file and import it in BRB-ArrayTools. You may pay attention to the choices in the following steps:

Click “Customize Annotations” to select information columns for output in TAC.

Click “Show/Hide Columns” to select output columns for the TXT file in TAC.

Assign different gene identifiers in BRB-ArrayTools.

Data Import Wizard

Please specify the location of your gene identifiers:

The identifiers are stored alongside the expression data. The identifiers are stored in a separate file.

Please select your Gene identifiers file:

File: browse Header line #:

Please select the available gene identifiers:

Unique ID (Well or Spot ID, etc.): Col 1: ID EntrezID: Col 6: Entrez ID

Clone ID (IMAGE or ATCC ID, etc.): Gene Name, Title or Description: Col 4: Description

UniGene Cluster ID: Col 7: UniGene ID GenBank Accession: Col 2: Public Gene IDs

Gene Symbol: Col 3: Gene Symbol Map Location: Col 5: Chromosome

Ensembl ID: microRNA ID:

Annotate the project with these gene ids, instead of using the data from SOURCE database. Organisms: Mouse

Reset Back Next Exit Explain More

5.2. Affymetrix HT HG-U133+ PM Array

NCBI GEO GSE100833: Affymetrix HT HG-U133+ PM Array Plate. You can follow the above steps to convert CEL files to a probe summarization TXT file and import it in BRB-ArrayTools. You may pay attention to the choices in the following steps:

Click “Customize Annotations” to select information columns for output in TAC.

appliedbiosystems Analysis_4.tacx

Sample QC View Summary View Gene View

Gene Table Summary Sample Table

Comparison: 1 vs 2 What is the differential expression between 1 and 2?

Apply View Filters Show/Hide Columns Export Add Column Customize Annotations Gene List miRNA Interaction

ID	1 Avg (log2)	2 Avg (log2)	Fold Change	P-val	FDR P-val	Gene Symbol	Description
230180_PM_at	5.6	6.76	-2.23	6.16E-15	2.61E-13	DDX17	DEAD (Asp-Glu)
232737_PM_s_at	4.34	5.39	-2.06	3.26E-05	0.0002	ENPP3	ectonucleotide
204080_PM_at	2.98	4.24	-2.4	0.0003	0.0012	TINAG	tubulininterstit
241547_PM_at	3.22	4.91	-3.23	0.0008	0.0027	A1CF	APOBEC1 com
207178_PM_s_at	3.94	5.03	-2.13	0.0012	0.0038	FRK	fyn-related Src
236118_PM_at	5.57	6.61	-2.06	0.0012	0.0038	GATA6-A...	GATA6 antisen
239065_PM_at	3.15	4.88	-3.31	0.0030	0.0083	DNAJC22	Dnal (Hsp40) f
223426_PM_s_at	5.43	6.63	-2.29	0.0031	0.0086	EPB41L4B	erythrocyte me
219850_PM_s_at	4.52	5.65	-2.18	0.0032	0.0088	EHF	ets homologou
229337_PM_at	5.24	6.45	-2.31	0.0038	0.0101	USP2	ubiquitin spec
237530_PM_at	2.44	3.77	-2.51	0.0040	0.0108		
204007_PM_at	11.92	9.89	4.09	0.0065	0.0163	FCGR3B	Fc fragment of
210146_PM_x_at	9.96	8.91	2.07	0.0067	0.0167	IL1RL2	leukocyte imm
236279_PM_at	9.4	10.56	-2.23	0.0074	0.0182		
208596_PM_s_at	5.37	7.01	-3.11	0.0079	0.0193	UGT1A1...	UDP glucuron
221305_PM_s_at	3.69	5.51	-3.53	0.0079	0.0194	UGT1A1...	UDP glucuron
206755_PM_at	3.81	5.31	-2.83	0.0099	0.0235	CYP2B6	cytochrome P4
230573_PM_at	5.04	6.2	-2.22	0.0101	0.0239	SGK2	serum/glucoc
205863_PM_at	9.34	7.65	3.22	0.0105	0.0248	S100A12	S100 calcium b
1552367_PM_a...	3.53	5.26	-3.32	0.0122	0.0280	SCIN	scinderin
204379_PM_s_at	7.05	8.09	-2.07	0.0131	0.0299	FGFR3	fibroblast gro
230914_PM_at	6.95	8	-2.08	0.0147	0.0328	HNF4A	hepatocyte nu
215125_PM_s_at	6.01	7.24	-2.35	0.0152	0.0339	UGT1A1...	UDP glucuron
207781_PM_s_at	4.06	5.4	-2.53	0.0165	0.0364	ZNF711	zinc finger pro
209949_PM_at	10.45	9.28	2.26	0.0191	0.0413	NCF2	neutrophil cys
207126_PM_x_at	5.66	6.96	-2.47	0.0212	0.0452	UGT1A1...	UDP glucuron
205568_PM_at	11.23	8.7	5.76	0.0243	0.0507	AQP9	aquaporin 9
204532_PM_x_at	5.25	7.01	-3.4	0.0251	0.0520	UGT1A1...	UDP glucuron
230435_PM_at	2.92	5.13	-4.61	0.0267	0.0548	SLC30A10	solute carrier f
204006_PM_s_at	9.23	8.01	2.33	0.0267	0.0549	FCGR3A...	Fc fragment of
213558_PM_at	4.94	6.51	-2.97	0.0287	0.0583	PCLO	piccolo presyn
230238_PM_at	5.18	6.37	-2.28	0.0293	0.0594	SCWAHA	sonosdowah a

Customize Annotations

Annotation File: HT_HG-U133_Plus_PM.na36.annot.csv

Annotations Assignment: Top Assignment All Assignments

Select Annotation Column(s) to Add:

- GeneChip Array
- Species Scientific Name
- Annotation Date
- Sequence Type
- Sequence Source
- Transcript ID/Array Design (Already Added)
- Target Description
- Representative Public ID
- Archival UniGene Cluster
- UniGene ID
- Genome Version
- Alignments
- Gene Title
- Gene Symbol (Already Added)
- Chromosomal Location
- UniGene Cluster Type
- Ensembl
- Entrez Gene
- SwissProt
- EC
- OMIM
- RefSeq Protein ID
- RefSeq Transcript ID
- RGD Name
- FlyBase
- AgI
- WormBase
- MGI Name
- RGD Name
- Check/Uncheck All

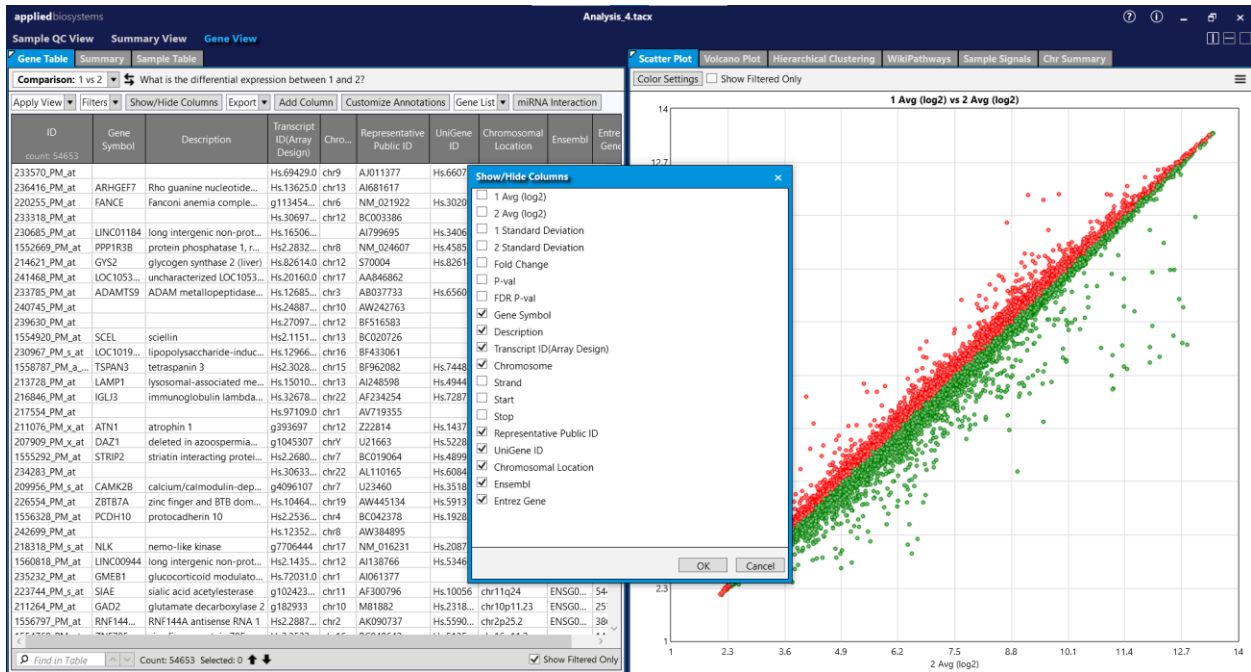
OK Cancel

1 Avg (log2) vs 2 Avg (log2)

1 2.3 3.6 4.9 6.2 7.5 8.8 10.1 11.4 12.7 14

2 Avg (log2)

Click “Show/Hide Columns” to select output columns for the TXT file in TAC.



Assign different gene identifiers in BRB-ArrayTools.

Data Import Wizard

Please specify the location of your gene identifiers:

The identifiers are stored alongside the expression data. The identifiers are stored in a separate file.

Please select your Gene identifiers file:

File: browse Header line #:

Please select the available gene identifiers:

Unique ID (Well or Spot ID, etc.): Col 1: ID EntrezID: Col 10: Entrez Gene

Clone ID (IMAGE or ATCC ID, etc.): Gene Name, Title or Description: Col 3: Description

UniGene Cluster ID: Col 7: UniGene ID GenBank Accession: Col 6: Representative Public ID

Gene Symbol: Col 2: Gene Symbol Map Location: Col 8: Chromosomal Location

Ensembl ID: Col 9: Ensembl microRNA ID: Col 4: Transcript ID(Array Desi

Annotate the project with these gene ids, instead of using the data from SOURCE database. Organisms: Human

Reset Back Next Exit Explain More