

Using Cvt2Mae to Convert User-Defined Array Data for MAExplorer

<http://www.lecb.ncifcrf.gov/Cvt2Mae>

Peter F. Lemkin(1), Greg Thornwall (2), Bob Stephens(3)

(1) LECB/NCI/FCRDC, (2) SAIC/FCRDC, (3) ABCC/FCRDC

DRAFT - Revised: 01-28-2002

Cvt2Mae version 0.60

Accessing Arrays with MAExplorer

- MAExplorer works with any arrays using the schema (see Appendix C of MAExplorer Reference Manual for details)
- All data files are tab-delimited text files
- Databases could be constructed with tools like Excel for editing user data into the schema format
- The Cvt2Mae array data converter “Wizard” tool converts non-standard **<User-defined>** academic or commercial data to MAExplorer format
- Affymetrix, Incyte, GenePix, Scanalyze, and other array data formats may be converted using predefined “Array Layouts”

S.1 MAExplorer Data Schema

- MAExplorer works with any array data using our data schema
- The schema is described in detail in MAExplorer Reference Manual Appendix C.
- Data Schema: tab-delimited experiment data files:
 1. GIPO (Gene In Plate Orders or “array print” file)
 2. List of hybridized samples in database
 3. Configuration data describing the array and conventions
 4. Separate spot quantification data files
- The Cvt2Mae “wizard” tool converts user array data to this schema

S.1.1 MAExplorer GIPO or Print File

- GIPO file maps a spot on the array to a particular gene
- Contains:
 1. location or grid-geometry
 2. one or more genomic identifiers (e.g., Clone ID, GenBank ID, LocusID, or simply Location etc.)
 3. gene description as Gene Name (or other description)
 4. Optional: global spot quality (QualCheck)
 5. optional: plate coordinates for clones

S.1.2 MAExplorer Samples Database File

- List of hybridized samples file SamplesDB.txt file contains:
 1. full sample description
 2. base file name of quantification file (without .quant file extension)
 3. optional sample ID number
 4. other data you wish to carry with the samples (used in array reports)

S.1.3 MAExplorer Configuration Database File

- Configuration data file MaeConfig.txt describes particular type of array and hybridization labeling you are using. This includes:
 - grid-geometry - # of replicate fields, grids, rows/grid, columns/grid
 - spot hybridization labeling - intensity or ratio data, dye names
 - various presentation options - use pseudo-array or actual (x,y) coordinates, etc.

S.1.4 MAExplorer Spot Quantification Files

- Separate spot quantification data files (with .quant file extension) are used for each hybridized sample
- ^{33}P or biotin labeled samples are specified as one hybridization intensity information per file
- Fluorescent Cy3/Cy5-dye labeled samples are specified as two channels of hybridization intensity information per file
- Intensity background data is optional
- Spot quality (QualCheck) data is optional
- Grid-coordinates are specified the same as for GIPO file

S.2 Assumptions About User Data - Array Layout

- User data is tab-delimited ASCII text files (could generate with Excel)
- If the array geometry (#fields, grids, rows/grid, columns/grid) is known, that geometry may be used in MAExplorer
- Otherwise, a pseudo-array geometry is generated for visual use in MAExplorer from the total # of spots in the user data
- An Array Layout describes the user data. It may be edited and saved for subsequent use in converting other array data files of the same type
- The <User-defined> array layout gives users complete flexibility in describing the array

S.3 Example of tab-delimited GIPO Data

The screenshot shows a Microsoft Excel spreadsheet titled "genericGIPOforScanalyze.gal". The data is organized into a table with columns labeled A through I. The first 37 rows contain block information, and the last 15 rows (38-53) provide a detailed list of rows for each block.

Block	Row	Column	ID	Name	
1	ATF	1			
2	35	5			
3	Type=GenePix ArrayList V1.0				
4	BlockCount=32				
5	BlockType=0				
6	Block1= 3350, 12320, 130, 21, 210, 20, 210				
7	Block2= 7840, 12320, 130, 21, 210, 20, 210				
8	Block3= 12330, 12320, 130, 21, 210, 20, 210				
9	Block4=16820, 12320, 130, 21, 210, 20, 210				
10	Block5= 3350, 16810, 130, 21, 210, 20, 210				
11	Block6= 7840, 16810, 130, 21, 210, 20, 210				
12	Block7= 12330, 16810, 130, 21, 210, 20, 210				
13	Block8=16820, 16810, 130, 21, 210, 20, 210				
14	Block9= 3350, 21300, 130, 21, 210, 20, 210				
15	Block10= 7840, 21300, 130, 21, 210, 20, 210				
16	Block11= 12330, 21300, 130, 21, 210, 20, 210				
17	Block12=16820, 21300, 130, 21, 210, 20, 210				
18	Block13= 3350, 25790, 130, 21, 210, 20, 210				
19	Block14= 7840, 25790, 130, 21, 210, 20, 210				
20	Block15= 12330, 25790, 130, 21, 210, 20, 210				
21	Block16=16820, 25790, 130, 21, 210, 20, 210				
22	Block17= 3350, 30280, 130, 21, 210, 20, 210				
23	Block18= 7840, 30280, 130, 21, 210, 20, 210				
24	Block19= 12330, 30280, 130, 21, 210, 20, 210				
25	Block20=16820, 30280, 130, 21, 210, 20, 210				
26	Block21= 3350, 34770, 130, 21, 210, 20, 210				
27	Block22= 7840, 34770, 130, 21, 210, 20, 210				
28	Block23= 12330, 34770, 130, 21, 210, 20, 210				
29	Block24=16820, 34770, 130, 21, 210, 20, 210				
30	Block25= 3350, 39260, 130, 21, 210, 20, 210				
31	Block26= 7840, 39260, 130, 21, 210, 20, 210				
32	Block27= 12330, 39260, 130, 21, 210, 20, 210				
33	Block28=16820, 39260, 130, 21, 210, 20, 210				
34	Block29= 3350, 43750, 130, 21, 210, 20, 210				
35	Block30= 7840, 43750, 130, 21, 210, 20, 210				
36	Block31= 12330, 43750, 130, 21, 210, 20, 210				
37	Block32=16820, 43750, 130, 21, 210, 20, 210				
38	Block	Row	Column	ID	Name
39	1	1	1	IMAGE:820126	ESTs, Moderately similar to AF151830 1
40	1	1	2	IMAGE:820126	ESTs, Moderately similar to AF151830 1
41	1	1	3	IMAGE:820132	Lmo2--LIM only 2
42	1	1	4	IMAGE:820132	Lmo2--LIM only 2
43	1	1	5	IMAGE:820144	RIKEN cDNA 0610009M14 gene
44	1	1	6	IMAGE:820144	RIKEN cDNA 0610009M14 gene
45	1	1	7	IMAGE:820139	ESTs, Weakly similar to UBP4 MOUSE UBI
46	1	1	8	IMAGE:820139	ESTs, Weakly similar to UBP4 MOUSE UBI
47	1	1	9	IMAGE:820161	Itpk6-pendin--inositol hexakisphosphat
48	1	1	10	IMAGE:820161	Itpk6-pendin--inositol hexakisphosphat
49	1	1	11	IMAGE:820188	ESTs, Highly similar to KIAA1423 prote
50	1	1	12	IMAGE:820188	ESTs, Highly similar to KIAA1423 prote
51	1	1	13	IMAGE:820398	RIKEN cDNA 5430405N12 gene
52	1	1	14	IMAGE:820398	RIKEN cDNA 5430405N12 gene
53	1	1	15	IMAGE:820402	ESTs

S.3.1 Example of tab-delimited User-Defined Data

Microsoft Excel - genericData1.DAT

File Edit View Insert Format Tools Data Window Help Acrobat

W2 =

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	HEADER	SPOT	GRID	TOP	LEFT	BOT	RIGHT	ROW	COL	CH1	CH1B	CH1AB	CH2I	CH2B	CH2AB	SPIX	BGPIX	EDGE	RAT2	MRAT
2	REMARK	SOFTWARE	ScanAlyze																	
3	REMARK	SOFTVERS	2.44																	
4	REMARK	CH1 IMAGE	4_Cy3m																	
5	REMARK	CH2 IMAGE	4_Cy5m																	
6	REMARK	GRID FILE	C:generic.SAG																	
7	REMARK	DATE	10/1/2002																	
8	REMARK	TIME	11:45:03																	
9	SPOT	1	1	90	116	103	129	1	1	2690	2196	3150	2122	806	1706	137	1392	0	2.664	1.819
10	SPOT	2	1	90	137	103	150	1	2	2763	2174	3051	995	670	1598	137	1302	0	0.552	0.555
11	SPOT	3	1	90	158	103	171	1	3	6861	2162	2957	1454	548	1480	137	1302	0	0.193	0.266
12	SPOT	4	1	90	179	103	192	1	4	8181	2046	2843	1221	512	1461	137	1302	0	0.116	0.288
13	SPOT	5	1	90	200	103	213	1	5	3139	1912	2772	1060	446	1346	137	1302	0	0.5	0.58
14	SPOT	6	1	90	221	103	234	1	6	3105	1864	2685	1185	406	1259	137	1302	0	0.628	0.474
15	SPOT	7	1	90	242	103	255	1	7	2286	2032	2966	1273	380	1253	137	1302	0	3.516	0.827
16	SPOT	8	1	90	263	103	276	1	8	3374	2234	3197	1165	392	1234	137	1302	0	0.678	0.62

genericData1 / Ready

Microsoft Excel - genericData1.DAT

File Edit View Insert Format Tools Data Window Help Acrobat

AJ16 =

	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	REGR	CORR	LFRAT	CH1GTB1	CH2GTB1	CH1GTB2	CH2GTB2	CH1EDGEA	CH2EDGEA	FLAG	CH1KSD	CH1KSP	CH2KSD	CH2KSP	
2															
3															
4															
5															
6															
7															
8															
9	0.021	0.0227	0.1211	0.438	0.4818	0.2774	0.4161	0.3673	0.4854	0	0.1081	1.01E-01	0.2106	2.42E-05	
10	0.031	0.0447	0.0589	0.4672	0.3796	0.3212	0.3504	0.3587	0.4844	0	0.05403	8.53E-01	0.2482	3.26E-07	
11	0.022	0.0471	0.029	0.781	0.5036	0.6423	0.4234	0.3311	0.51	0	0.3562	2.10E-14	0.2351	1.62E-06	
12	-0.044	-0.101	18.64	0.854	0.4526	0.7591	0.4088	0.3139	0.5587	0	0.4484	1.39E-22	0.2808	4.07E-09	
13	-0.007	-0.008	50.75	0.5474	0.4453	0.4672	0.3796	0.329	0.5362	0	0.1123	8.15E-02	0.3135	2.96E-11	
14	-0.021	-0.032	25.62	0.562	0.4818	0.4453	0.4234	0.3673	0.5413	0	0.1034	1.33E-01	0.2351	1.62E-06	
15	-0.005	-0.007	132.2	0.3796	0.4526	0.2701	0.4161	0.3813	0.5888	0	0.1274	3.27E-02	0.2874	1.59E-09	
16	0.032	0.053	0.0511	0.562	0.5182	0.4526	0.4672	0.3439	0.5401	0	0.1013	1.48E-01	0.2678	2.51E-08	

genericData1 / Ready

I. Procedure: Convert Data for Array Layouts

1. Select the Chip Set array layout if in list, otherwise pick <User-defined>)
2. Select separate GIPO file if needed using the “Browse GIPO file” .
 - 2.1 Repeatedly select 1 or more input files using the “Browse input files”
3. You may edit or change various array layout parameters at this time
 - 3.1 you may edit the array layout with “Edit Layout”
 - 3.2 you may “Assign GIPO fields” in user data file
 - 3.3 you may “Assign Quantification fields” in user data file
 - 3.4 if you changed any array layout parameters, you may save it with “Save Layout”

I. Procedure: continued...

4. Select the project output directory (i.e., folder) to save generated files
5. Press “Run” to convert the data
6. Press “Done” when it is finished.
7. Go to the project directory and then to the MAE sub-directory, click on the Start.mae file to start MAExplorer on the new data

1. Initial State of Cvt2Mae Program

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files:

2.1 Edit array layout and map fields:

2.2 Samples to use '<<file>> sample name':

3. Select Project Output Folder:

Vendor

Layout name

Spots/microarray

Project output folder:

MAExplorer startup File:

4. Edit and Run

Status:

2. Selecting <User-defined> Chipset Array-Layout

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Mult. samples/file

GenePixPro3 - generic
Scanalyze - generic
Affymetrix - generic
Incyte - generic
Affymetrix - Mouse
Affymetrix - Human
Affymetrix - Mouse, use Genomic Descriptions

2.1 Edit array layout and map fields:
 Expert assign-mode

2.2 Samples to use '<<file>> sample name':

Vendor	<input type="text" value="?"/>
Layout name	<input type="text" value="<User-defined>"/>
Spots/microarray	<input type="text" value="0"/>

3. Select Project Output Folder:

Project output folder:

MAExplorer startup File:

4. Edit and Run

Status:

2.1 For this example, select “separate GIPO” and deselect “Mult. Samples/file”

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Separate GIPO Mult. samples/file

2.1 Edit array layout and map fields:
 Expert assign-mode

2.2 Samples to use '<<file>> sample name':

Vendor	<input type="text" value="?"/>
Layout name	<input type="text" value="<User-defined>"/>
Spots/microarray	<input type="text" value="0"/>

3. Select Project Output Folder:

Project output folder:

MAExplorer startup File:

4. Edit and Run

Status:

3.1 Select GIPO Input File with “Browse GIPO file”

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4, press 'Run' to convert your data to MAExplorer format.

1. Select Chipset: <User-defined> **Remove Layout**

2. Select Input Data Files: **Browse input file name** **Separate GIPO** **Browse GIPO file** **Mult. samples/file**

2.1 Edit array layout and map fields: **Edit Layout** **Assign GIPO fields** **Assign Quant fields**
Save Layout **Expert assign-mode**

2.2 Samples to use '<<file>> sample name': **Analyze input files** **Remove sample** **Rename sample**

3. Select Project

4. Edit and Run **Run - do conversion** **Abort** **Reset**

Status:

Enter sepearte input GIPO file to convert

Look in: ScanalyzeData

- genericData1.DAT
- genericData2.DAT
- genericGIPOforScanalyze.gal

File name: genericGIPOforScanalyze.gal **Open**

Files of type: All Files (*.*) **Cancel**

?
<User-defined>
0

3.2 Specify GIPO Field Names for Grid, Row & Column

Specify separate GIPO file fields [X]

Specify GIPO file field names for (Grid, Row, Columns). Case is ignored.

Enter GIPO 'Grid' field name from pull-down list or type it

Block [Block] [v]

Block

Grid

Array Block

Array Grid

Enter GIPO 'Row' field name from pull-down list or type it

Row [Row] [v]

Row

Engter GIPO 'Column' field name from pull-down list or type it

Col [Col] [v]

Column

Ok **Cancel**

3.3 Select Files with “Browse input file” Name

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset: <User-defined> **Remove Layout**

2. Select Input Data Files: **Browse input file name** **Separate GIPO** **Browse GIPO file** **Mult. samples/file**

genericData1.DAT

2.1 Edit array layout and map fields: **Edit Layout** **Assign GIPO fields** **Assign Quant fields**
Save Layout **Expert assign-mode**

2.2 Samples to use '<<file>> sample name': **Analyze input files** **Remove sample** **Rename sample**

<<C:\Temp\ScanalyzeData\genericData1.DAT>> [genericData1.DAT]

3. Select Project

4. Edit and Run

Select next input file to convert (you may use 'ALL' or 'ALL.<ext>')

Look in: ScanalyzeData

- genericData1.DAT
- genericData2.DAT
- genericGIPOforScanalyze.gal

File name: genericData2.DAT **Open**

Files of type: All Files (*.*) **Cancel**

Abort **Reset**

Status: *Either continue adding input files (step 2), or define Output Folder (step 3) when done adding files.*

4. Continue Adding Input Files If Needed

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Separate GIPO Mult. samples/file

genericData1.DAT
genericData2.DAT

2.1 Edit array layout and map fields:
 Expert assign-mode

2.2 Samples to use '<<file>> sample name':

<<C:\Temp\ScanalyzeData\genericData1.DAT>> [genericData1.DAT]
<<C:\Temp\ScanalyzeData\genericData2.DAT>> [genericData2.DAT]

Vendor	<input type="text" value="?"/>
Layout name	<input type="text" value="<User-defined>"/>
Spots/microarray	<input type="text" value="13448"/>

3. Select Project Output Folder:

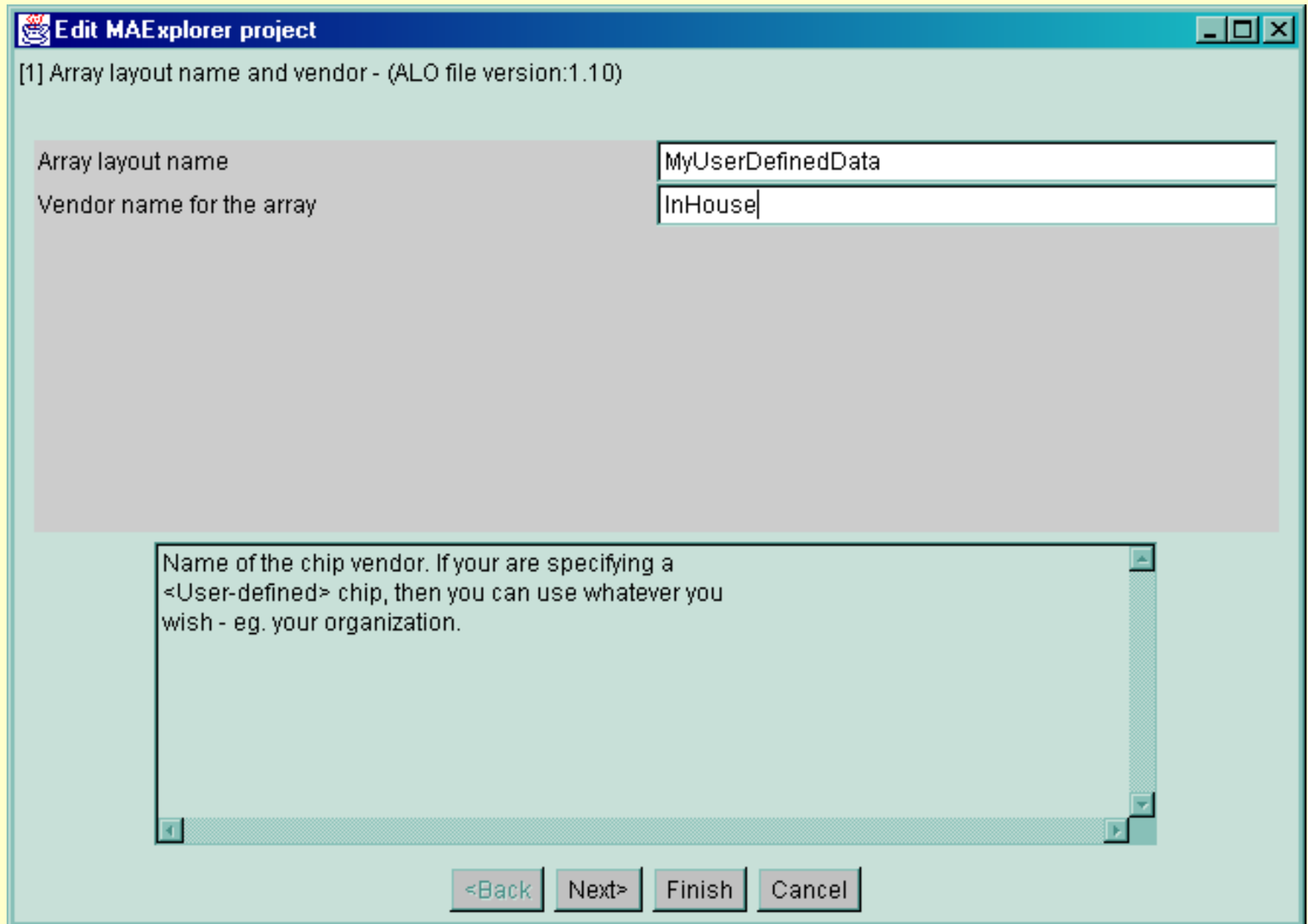
Project output folder:

MAExplorer startup File:

4. Edit and Run

Status:

5.1 Edit Layout 'Wizard' Values for This Array



The image shows a screenshot of a software dialog box titled "Edit MAExplorer project". The dialog box has a title bar with standard window controls (minimize, maximize, close) and a subtitle "[1] Array layout name and vendor - (ALO file version:1.10)".

Inside the dialog, there are two input fields:

- The first field is labeled "Array layout name" and contains the text "MyUserDefinedData".
- The second field is labeled "Vendor name for the array" and contains the text "InHouse".

Below the input fields, there is a text box containing the following text:

Name of the chip vendor. If your are specifying a <User-defined> chip, then you can use whatever you wish - eg. your organization.

At the bottom of the dialog, there are four buttons: "<Back", "Next>", "Finish", and "Cancel".

5.2 Edit Layout 'Wizard' - Grid Geometry.

Deselect "Specify layout by #"

The screenshot shows a dialog box titled "Edit MAExplorer project" with a subtitle "[2] Grid geometry data". It contains several input fields and checkboxes. The "Specify array layout by Grid-geometry OR by # spots/array" checkbox is checked, and the text "Use # spots (BELOW), else grid-geometry (ABOVE)" is highlighted in red. Below the input fields is a text box with instructions and a set of navigation buttons at the bottom.

Number of duplicated spot Fields in array	1
Number of Grids per Field	12
Number of spots per Grid Row	28
Number of spots per Grid Column	41
Use Mol.Dynamics 'NAME-GRC' else (Grid,Row,Col)	<input type="checkbox"/> Use Mol.Dyn. 'NAME-GRC' else above explicit (Grid,Row
Specify array layout by Grid-geometry OR by # spots/array	<input checked="" type="checkbox"/> Use # spots (BELOW), else grid-geometry (ABOVE)
Maximum number of spots in array	13448

If you specify the array layout by Grid-geometry (ABOVE), then enter (#Fields, #Grids, #Grid-rows,#Grid-cols).
If you specify the layout by the maximum number of spots in the array (BELOW), it will estimate a pseudo-layout that the spots will fit on the this array for visualization purposes. It does not correspond to the actual array layout which you do not have to enter.

<Back Next> Finish Cancel

5.2.1 Edit Layout 'Wizard' - Grid Geometry. Enter (Grid, Rows/Grid, Columns/Grid) Values

Edit MAExplorer project

[2] Grid geometry data

Number of duplicated spot Fields in array	1
Number of Grids per Field	32
Number of spots per Grid Row	21
Number of spots per Grid Column	20
Use Mol.Dynamics 'NAME-GRC' else (Grid,Row,Col)	<input type="checkbox"/> Use Mol.Dyn. 'NAME-GRC' else above explicit (Grid,Row)
Specify array layout by Grid-geometry OR by # spots/array	<input type="checkbox"/> Use # spots (BELOW), else grid-geometry (ABOVE)
Maximum number of spots in array	13448

If you specify the array layout by Grid-geometry (ABOVE), then enter (#Fields, #Grids, #Grid-rows,#Grid-cols).
If you specify the layout by the maximum number of spots in the array (BELOW), it will estimate a pseudo-layout that the spots will fit on the this array for visualization purposes. It does not correspond to the actual array layout which you do not have to enter.

<Back Next> Finish Cancel

5.3 Edit Layout 'Wizard' Input Data File Row Values. Verify Row Where Field Names Defined

The screenshot shows a dialog box titled "Edit MAExplorer project" with a subtitle "[3] Input file starting rows data". It contains a list of options on the left and corresponding input fields on the right. Below this is a text area with a scroll bar and a set of navigation buttons at the bottom.

(Optional) Row containing a list sample names	0
Row containing a list of quantitative file Field names	1
First row containing quantitative file Data	9
Row containing opt. separate GIPO file Field names	38
First row containing opt. separate GIPO file Data	39
(Optional) Comment token	
(Optional) Initial keyword for each data row	

Number of row that contains the names of the data file Field names.
Eg. grid, row, column, GeneBank ID, GeneName, Clone ID, etc.
[Row #s start at row 1.]
Data from row #1 in file[C:\Temp\ScanalyzeData\genericData1.DAT]
Current Field name column[1] = 'HEADER'
Current Field name column[2] = 'SPOT'
Current Field name column[3] = 'GRID'
Current Field name column[4] = 'TOP'
Current Field name column[5] = 'LEFT'

<Back Next> Finish Cancel

5.3.1 Edit Layout 'Wizard' Input GIPO File Row Values. Verify Row Where Field Names Defined

Edit MAExplorer project

[3] Input file starting rows data

(Optional) Row containing a list sample names	0
Row containing a list of quantitative file Field names	1
First row containing quantitative file Data	9
Row containing opt. separate GIPO file Field names	38
First row containing opt. separate GIPO file Data	39
(Optional) Comment token	
(Optional) Initial keyword for each data row	

Number of row that contains the names of optional GIPO file Field names in the file. Eg. grid, row, column, GeneBank ID, GeneName, Clone ID, etc.
[Row #'s start at row 1.]
Data from row #38 in file[C:\Temp\ScanalyzeData\genericGIPOforScanalyze.gal]
Current GIPO Field name column[1] = 'Block'
Current GIPO Field name column[2] = 'Row'
Current GIPO Field name column[3] = 'Column'
Current GIPO Field name column[4] = 'ID'
Current GIPO Field name column[5] = 'Name'

<Back Next> Finish Cancel

5.4 Edit Layout 'Wizard' Ratio or Intensity Values. Select "background" and "Use Ratio"

The screenshot shows a dialog box titled "Edit MAExplorer project" with a sub-header "[4] Ratio fluorescence data". The dialog is divided into two main sections. The top section contains configuration options for data type and background. The bottom section contains a text box with explanatory text and navigation buttons.

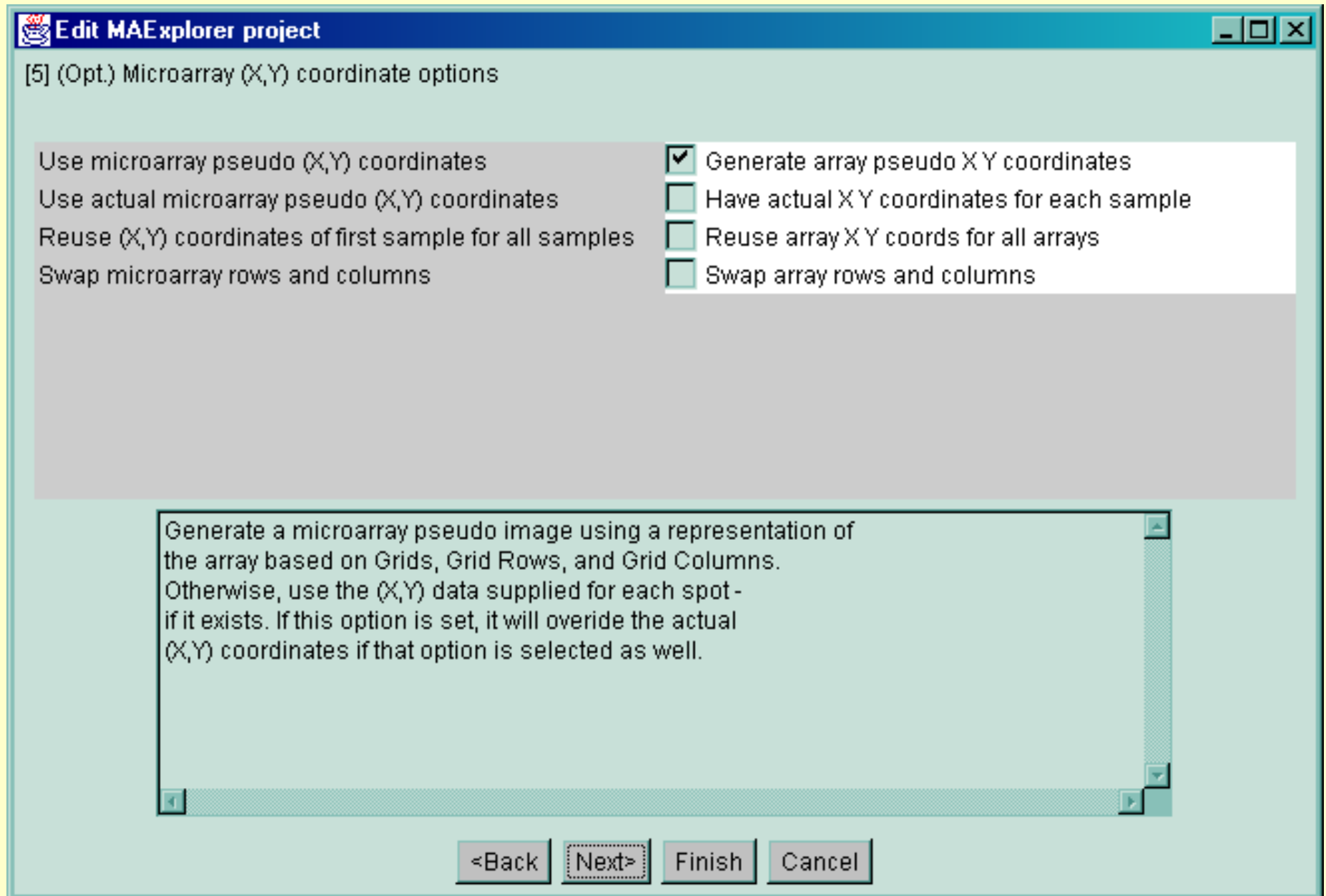
Ratio (i.e. Cy3,Cy5) or Intensity Data
If Ratio data, use (Cy5/Cy3) else (Cy3/Cy5)
Fluorescent dye for intensity 1 (if ratio data)
Fluorescent dye for intensity 2 (if ratio data)
Have background intensity data

Use Ratio else Intensity data
 Use (Cy5/Cy3) else (Cy3/Cy5)
Cy3
Cy5
 Has background data

Data for MAExplorer is either ratio data such as Cy3/Cy5, or intensity data such as P33, etc.

<Back Next> Finish Cancel

5.5 Edit Layout 'Wizard' optional (X,Y) Coordinate Values for your data



5.6 Edit Layout 'Wizard' Genomic ID Values.

Select "Clone ID"

Edit MAExplorer project

[6] (Opt.) Genomic Identifier options

Has Location data	<input type="checkbox"/>	User data file has Location data
Has Clone ID data	<input checked="" type="checkbox"/>	User data file has Clone ID data
Has GenBank data	<input type="checkbox"/>	User data file has GeneBank data
Has UniGene ID data	<input type="checkbox"/>	User data file has UniGene data
Has dbEST data	<input type="checkbox"/>	User data file has dbEST data
Has LocusLink data	<input type="checkbox"/>	User data file has LocusLink data
Has SwissProt data	<input type="checkbox"/>	User data file has SwissProt data
Has Plate data	<input type="checkbox"/>	User data file has Plate data
Get Genomic IDs from 'Description'	<input type="checkbox"/>	Get Genomic IDs from 'Description'

The user data file has I.M.A.G.E 'Clone ID' data.

<Back Next> Finish Cancel

5.7 Edit Layout 'Wizard' Gene Names Description. Selecting "QualCheck"

Edit MAExplorer project

[7] (Opt.) Gene names (or description) options

Has Gene Class user data	<input type="checkbox"/>	User data file has Gene Class data
Has UniGene Name user data	<input type="checkbox"/>	User data file has UniGene Name data
Has separate per-spot QualCheck user data per-sample	<input checked="" type="checkbox"/>	User data has separate per-spot QualCheck data
Has 'GIPO' QualCheck user data for entire DB	<input type="checkbox"/>	User data file has 'GIPO' QualCheck data

The user data file has 'Quant' QualCheck data. This data is on a per-spot basis for each array hybridization. The code (see MAExplorer Reference Manual Appendix C Table C.4.2) may be used to flag bad spots or missing spot data.

<Back Next> Finish Cancel

5.8 Edit Layout 'Wizard' Calibration Values. Define UniGene Species prefix

Edit MAExplorer project

[8] (Opt.) DNA Calibration and user plate names, UniGene species name

Name of calibration DNA (if in database)

Name of researcher's special clones (if in database)

Name of empty wells

Name species (opt)

Name UniGene Species prefix (opt) or select from

- Hs
- Mm**
- At
- Bt
- Dr
- Hv
- Os
- Rn

UniGene species prefix (Mouse Mm, Human Hs, etc.). This is used in querying Genomic Web databases. If you do not see the prefix you want in the choice menu, type it in.

<Back Next> Finish Cancel

5.9 Edit Layout 'Wizard' Database Name Values. Define Optional Names for Database

The screenshot shows a dialog box titled "Edit MAExplorer project" with a standard Windows window border. The main area is light blue and contains a list of labels on the left and corresponding text input fields on the right. The labels are: "Your name of the created database (opt)", "Your name of the database subset (opt)", "Generic project name for all samples (opt)", and "Name of spot quantification program (opt)". The input fields contain the following text: "My user-defined database", "My user-define database subset", "My Data", and "Scanalyze". Below this list is a larger text area with a scroll bar, containing the text: "Name of the program used to quantitate the spot data from the sample images." At the bottom of the dialog are four buttons: "<Back", "Next>", "Finish", and "Cancel".

Edit MAExplorer project

[9] (Opt.) Database and data quantification program

Your name of the created database (opt) My user-defined database

Your name of the database subset (opt) My user-define database subset


Generic project name for all samples (opt) My Data

Name of spot quantification program (opt) Scanalyze

Name of the program used to quantitate the spot data from the sample images.

<Back Next> Finish Cancel

5.10 Edit Layout 'Wizard' HP-X,-Y Class Names

 Edit MA Explorer project

[10] (Opt.) Hybridized sample (X,Y) 'set' class names

Default name of X samples 'set'	<input type="text" value="HP-X 'set'"/>
Default name of Y samples 'set'	<input type="text" value="HP-Y 'set'"/>

This is the name for the samples assign to the 'X set'.

5.11 Edit Layout 'Wizard' Default Thresholds

Edit MAExplorer project

[11] (Opt.) Default data Filtering thresholds

Default cluster similarity threshold [0 : 1000]	10.0
Default # genes in highest/lowest	100
Default # clusters for K-means clustering [1 : 1000]	6
Default p-value threshold (for t-tests) [0.0 : 1.0]	0.05
Default Coeff. Of Variation threshold [0.0 : 1.0]	0.2
Default absolute difference threshold [0.0 : 4.0]	0.2

Default p-Value used in the t-Test data Filter.
This is the initial value shown in popup sliders.

<Back **Next>** Finish Cancel

6. Other Options - Assigning User Data Fields to MAExplorer Fields

- GIPO (Gene In Plate Order or “array print” table) - assigns genes to positions on the array as well as GeneBank ID, Clone ID, LocusID (if available), Gene Name, etc.
- Quant data - assigns names of quantified data in the user file to MAExplorer data (e.g. Cy3 intensity to RawIntensity1, Cy5 to RawIntensity2, etc).

6.1 “Assign user fields to GIPO fields”

Assign user fields to GIPO fields

grid	GRID
grid row	ROW
grid col	COL
Clone ID	ID
GeneName	Name

Done Cancel Help Allow duplicates

Assign user fields to GIPO fields

grid	GRID
grid row	ROW
grid col	COL
Clone ID	ID
GeneName	Name

Done Cancel Help

- CH2EDGEA
- FLAG
- CH1KSD
- CH1KSP
- CH2KSD
- CH2KSP
- ID
- Name

6.2 “Assign user fields to Quant fields”

The image shows a software dialog box with a title bar containing a logo and standard window controls. The main title is "Assign user fields to Quantitation fields". The dialog contains a list of user fields on the left and corresponding quantitation fields in dropdown menus on the right. At the bottom, there are buttons for "Done", "Cancel", and "Help", along with a checked checkbox labeled "Allow duplicates".

User Field	Quantitation Field
grid	GRID
grid row	ROW
grid col	COL
Cy3	CH1I
Cy5	CH2I
QualCheck	FLAG
Cy3Bkg	CH1B
Cy5Bkg	CH2B

Buttons: Done, Cancel, Help, Allow duplicates

7. Optional “Save Layout” to Array Layout Database After Edit Layout and Assign fields

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Separate GIPO Mult. samples/file

genericData1.DAT
genericData2.DAT

2.1 Edit array layout and map fields:
 Expert assign-mode

2.2 Samples to use '<<file>> sample name':

<<C:\Temp\ScanalyzeData\genericData1.DAT>> [genericData1.DAT]
<<C:\Temp\ScanalyzeData\genericData2.DAT>> [genericData2.DAT]

Vendor	InHouse
Layout name	MyUserDefinedData
Spots/microarray	13448

3. Select Project Output Folder:

Project output folder:

MAExplorer startup File:

4. Edit and Run

Status:

8. Specifying “Create new project folder” Option Where Generated Database Will Be Saved

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Separate GIPO Mult. samples/file

genericData1.DAT
genericData2.DAT

2.1 Edit array layout and map fields:
 Expert assign-mode

2.2 Samples to use '<<file>> sample name':

<<C:\Temp\ScanalyzeData\genericData1.DAT>> [genericData1.DAT]
<<C:\Temp\ScanalyzeData\genericData2.DAT>> [genericData2.DAT]

Vendor	InHouse
Layout name	MyUserDefinedData
Spots/microarray	13448

3. Select Project Output Folder:

Project output folder:
MAExplorer startup File:

4. Edit and Run

Status:

8.1 Specifying New “Project Output Folder”

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Separate GIPO Mult. samples/file

genericScanalyzeData2.DAT
genericScanalyzeData1.DAT

2.1 Edit array files:

2.2 Samples to convert:

<<C:\Temp\gen
<<C:\Temp\gen

3. Select Project Output Folder:

Project output folder:

MAExplorer startup File:

4. Edit and Run:

Status:

Select the Project Folder to save converted data

Save in:

Name	Size	Type	Modified
------	------	------	----------

File name:

Save as type:

in-house
MyUserDefinedData
13448

8.2 “Project Output Folder” & MAE startup file

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Separate GIPO Mult. samples/file

genericScanalyzeData2.DAT
genericScanalyzeData1.DAT

2.1 Edit array layout and map fields:
 Expert assign-mode

2.2 Samples to use '<<file>> sample name':

<<C:\Temp\genericScanalyzeData2.DAT>> [genericScanalyzeData2.DAT]
<<C:\Temp\genericScanalyzeData1.DAT>> [genericScanalyzeData1.DAT]

Vendor	InHouse
Layout name	MyUserDefinedData
Spots/microarray	13448

3. Select Project Output Folder:

Project output folder:
MAExplorer startup File:

4. Edit and Run

Status:

9. Conversion in Process After Pressing “RUN”

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Separate GIPO Mult. samples/file

genericScanalyzeData2.DAT
genericScanalyzeData1.DAT

2.1 Edit array layout and map fields:
 Expert assign-mode

2.2 Samples to use '<<file>> sample name':

<<C:\Temp\genericScanalyzeData2.DAT>> [genericScanalyzeData2.DAT]
<<C:\Temp\genericScanalyzeData1.DAT>> [genericScanalyzeData1.DAT]

Vendor	InHouse
Layout name	MyUserDefinedData
Spots/microarray	13448

3. Select Project Output Folder:

Project output folder: C:\Temp\junk\
MAExplorer startup File: C:\Temp\junk\MAE\Start.mae

4. Edit and Run

Status: ==> re-reading row #1800 [C:\Temp\genericScanalyzeData2.DAT]
For sample #1 [genericScanalyzeData2.DAT]

10. Notification that Conversion is Finished

Cvt2Mae: convert array data to MAExplorer files - Version: 01-23-2002 V.0.60 (Beta)

Enter data for steps 1, 2, and 3. Then 4. press 'Run' to convert your data to MAExplorer format.

1. Select Chipset:

2. Select Input Data Files: Separate GIPO Mult. samples/file

genericScanalyzeData2.DAT
genericScanalyzeData1.DAT

2.1 Edit array layout and map fields:
 Expert assign-mode

2.2 Samples to use '<<file>> sample name':

<<C:\Temp\genericScanalyzeData2.DAT>> [genericScanalyzeData2.DAT]
<<C:\Temp\genericScanalyzeData1.DAT>> [genericScanalyzeData1.DAT]

Vendor	InHouse
Layout name	MyUserDefinedData
Spots/microarray	13448

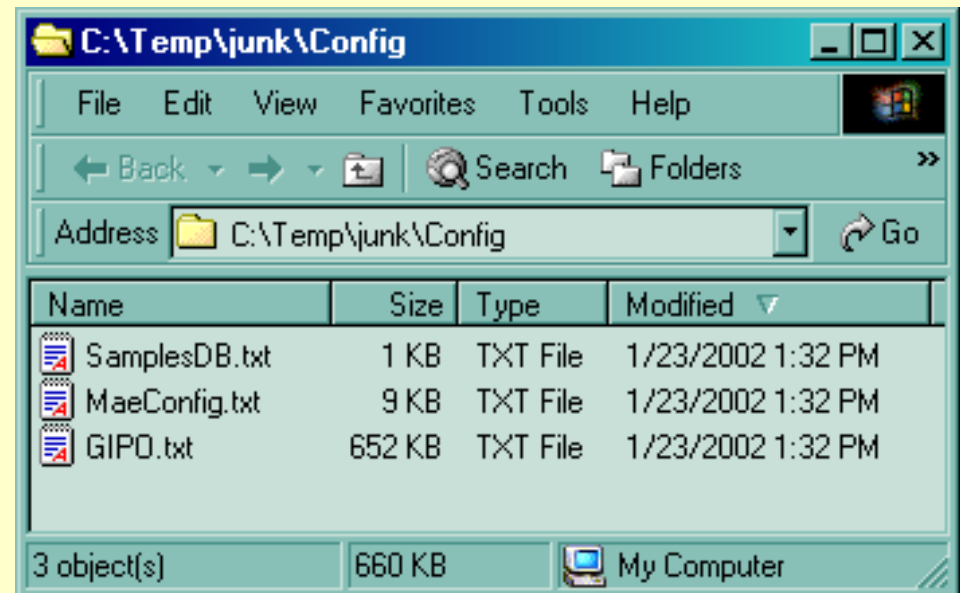
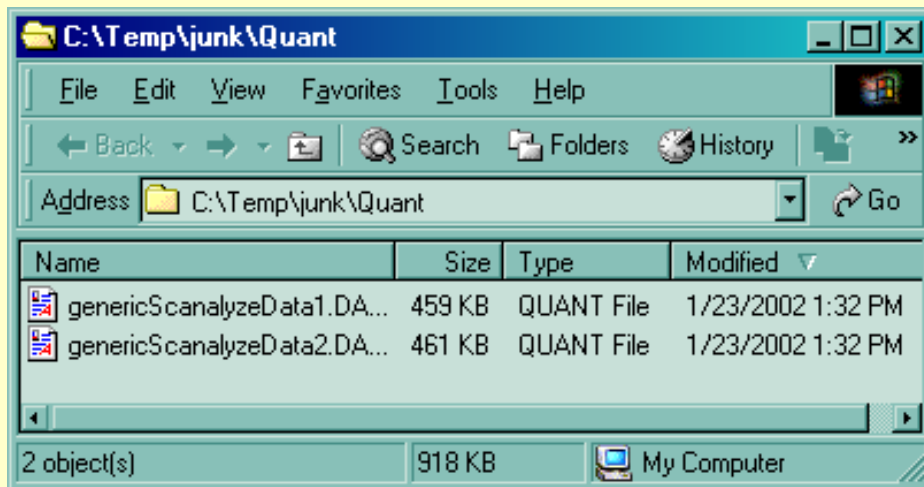
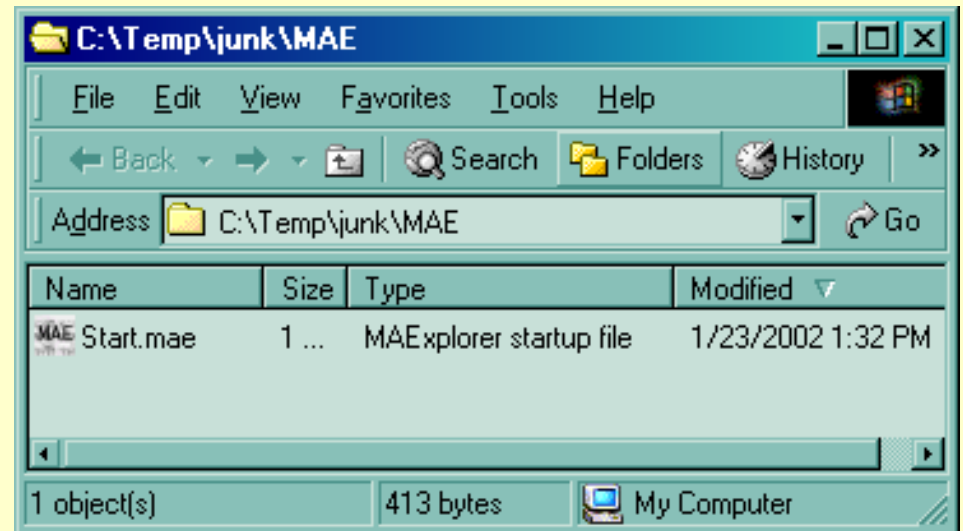
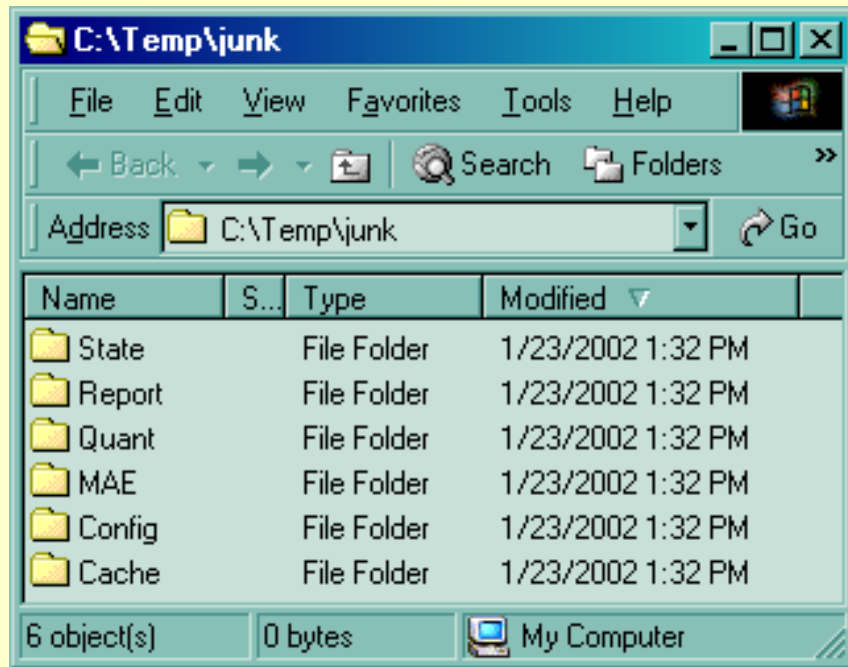
3. Select Project Output Folder:

Project output folder:
MAExplorer startup File:

4. Edit and Run

Status: **==> Finished writing out data files. Press 'Done' to exit**
To start MAExplorer, go to project folder & click on Start.mae.

11. MAExplorer Data Created By Cvt2Mae



12. Running MAExplorer on the Converted Data

