

Agilent Feature Extractor Plug-In for GeneSpring GX

FE Plug-In Quick Start Guide

This document describes the new Feature Extractor Plug-In for GeneSpring GX.

What is the Feature Extractor Plug-In?

The plug-in enhances the importing of data files produced by Agilent Feature Extraction software. The plug-in allows you to import a different set of columns from a FE data file and a number of FEPARAM fields as Sample Attributes, and to convert Agilent flag information into the GeneSpring GX flag information.

The source code for the plug-in is available and is located in the source directory of the Programs folder in the GeneSpring data folder. It is released under the LGPL license.

What is in this Quick Start guide?

- "Installation of the Plug-In"
- "Appendix A Flags"
- "Appendix B Agilent File Format"
- "Appendix C Agilent Normalization Scenario for 1-Color Data"
- "Appendix D Agilent Normalization Scenario for 2-Color Data"
- "Appendix E Extracted Sample Attributes"



Installation of the Plug-In

The FE plug-in for GeneSpring GX can be installed easily by dragging and dropping the Installer ZIP file onto an open main window of GeneSpring GX. The file is a special GeneSpring ZIP file that allows the easy installation of the new functionality into GeneSpring GX.

1 Download the installer ZIP file from the Agilent Technologies Web site at

http://www.chem.agilent.com/scripts/generic.asp?|page-34734&indcol=Y&prodcol=Y

- **2** Save the ZIP file to your local hard drive.
- **3 3**Start GeneSpring GX.
- **4 4**Drag the ZIP file onto the main GeneSpring GX window.
- **5** A dialog box displays.

Install Enhanced Agilent FE Import	×		
This will install the Enhanced Agilent FE Import			
Plug-in. GeneSpring will quit when this operation is			
finished. Do you wish to proceed?			
Yes No			

6 Press **OK** to continue or **Cancel** to cancel the installation process.

7 A **Caution** dialog box may display indicating that the installation will overwrite certain files. This may happen if previously the plug-in had been installed or a normalization scenario with the name **Agilent FE** was created. You can abort the installation by pressing **No** and determine if the files can be safely overwritten.

🕵 Caution	
This upgrade will overwrite some existing files, which	
were most recently installed on Apr 14, 2005 11:05:01	
AM. Do you wish to proceed?	
Yes No	

8 After the installation is complete, a new dialog box displays indicating that the installation was successful.

Upgrade Complete
Upgrade complete. Click Ok to
quit GeneSpring.

- **9** Press **OK** to continue.
- **10** GeneSpring GX closes automatically.

The plug-in is now installed correctly, and Agilent FE data can be loaded.

Loading data produced by Agilent FE

- **1** Start GeneSpring GX.
- 2 If you do not yet have the Agilent genome in GeneSpring GX, download the genome from the Agilent Web site by selecting **File**

> Import Genome. See the User Manual for more information on importing or creating genomes in GeneSpring GX.



- 3 Select File > Import data.
- **4** In the **Select One Data File** window, select an FE output file to import from the file-chooser dialog box provided.

5 The **Import Data: Define File Format and Genome** window displays.

File Format
Choose File Format: Agilent
Genome
Select the genome (set of genes on the array) for this data. If your genome does not appear on the list, you can create a new one by selecting Create a New Genome.
Select Genome
🗆 🔄 Genomes or Arrays
⊖-⊡ Agilent LS_ Human 1A
🗄 🔄 Demo Chips
- 🔁 Demo Human
-B Demo Rat
-B reast
C Create a New Genome
Choose a Name:
Next Cancel Help

6 GeneSpring GX tries to determine the data file's format is by examining the contents. If it is successful, it reports the name in the "Choose File Format:" drop-down box.

The drop-down menu may contain more than one data format, and it will always contain the "Custom" data format. GeneSpring GX can recognize many different data file formats, but it will only report the names of data files that are consistent with the current data files. If the Agilent FE output file contains 2-color data, the format is called *Agilent*. If the file contains 1-color data, the format is called *Agilent FE (1 Color)*.

More than the two data formats, *Agilent* and *Custom*, may be available in the drop-down menu if you or the administrator have created a Recognized Custom format before. This Recognized Custom Format will most likely be chosen as the default.

- 7 Verify that the **Select Genome** field indicates the correct genome for loading the data files, and press **Next** to continue.
- 8 The Import Data: Selected Files dialog box displays to allow you to add more files to be imported.

You can now choose multiple files to import by navigating to the correct folder on the left. Select the files in the middle window and press **Add** to add the files to the list on the right. Press **Next** to continue.

NM-C	0	ko.	3	Colorito d Files:	
C1	- 1	iama A		Name	•
		814702164_10011521014590_502_40		U614702384_16011521019517_802_4	ā
rectories.	1	014702004_10011521014592_002_A0		US14702364_16011521019516_802_4	a
1 P	Diulio	1014702004_10011521014500_002_A0	and the second second	U014702364_16011521019515_502_4	٥
P-	Vlark+	SELECTION_16011621014594_002_A0	Add ==	UE14702364_16011621019514_802_f	٥
	Savet	IS14107384_16011521016504_S07_A0	A	LS14702384_16011521019513_802_4	a
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	a Vener	814702364_10011521010510_902_A0	Harrison	U614702384_16011521019298_502_4	0
6		014702004_10011521016511_002_A0	-+ Remove	UE14702364_16011521019297_502_4	0
		IS14702164_16011521016512_902_AD		UE14702064_16011621019295_802_F	a
6	N3 🛄 -	STATE2164_16010421010413_802_A0	*** Reneve Al	U614702384_16011621019294_802_4	a
	Xanop	1614762364_16011521016514_502_A0		US14702364_16011521016517_502_4	al
2-0	B #COH	014702004_10011521010515_002_A0			n'
0 🗆 A	mersher.	014702304 10011621016518 002 NO.*			1
	•	•		1 Samples have multiple lites (subchips	9

NOTE: If you want to attach image files to Samples, such as JPG images of the arrays, you must place the image files and data files in the same folder.

9 After you select all of the files that you want to load, the Import Data: Preprocess Data Files dialog box displays.

NOTE: To ensure that the FE plug-in will be used, be sure that the **Choose File Format** drop-down box is set to "Agilent" or "Agilent FE (1-Color)" as appropriate to the data type.

🏽 Import Data: Preprocess Data Files 🛛 🗐 🖾
Do you want to preprocess your data files before importing your samples?
Previous Next Cancel Help

10 Select **Enhanced Agilent FE Import** from the drop-down menu, and click **Next**.

🌊 Import Data: Preprocess Data Files	
Do you want to preprocess your data files before importing your sar	mples?
None	-
None Enhanced Agilent FE Import	
Frevious Ivext Cancer Help	

11 The plug-in imports the correct columns—for example, for 2-color data, rProcessedSignal is imported as SIGNAL and gProcessedSignal as CONTROL, and it also allows you to convert Agilent flags to GeneSpring flags.

If you choose to use the Enhanced Agilent FE Import plug-in as directed in step 10, the Agilent FE flag information is converted to GeneSpring flag information.

The **Flag Import** dialog box displays with two questions. How you answer the questions determines *how* GeneSpring GX will interpret the FE flags:

- No No: Ignore all of the FE flags.
- Yes Yes: Import all of the FE flags.
- **Yes No**: Import only FE flags that deal with spot information. This is the default.

• **No – Yes**: Import only FE flags that deal with background information.

Click **Advanced...** if you want to import only specific FE flags. The **Advanced Flag Import** dialog box displays with a more granular set of choices. See step 12.

Flag Import	
Do you want GeneSpring to use the spot information in FE to flag the data?	Yes 💌
Do you want to include the background readings in the flag settings?	No 💌
OK Advanced >> Help	

The three buttons at the bottom of the dialog box offer the options

- **OK** continues the import process by displaying the **Suggested Normalization** window and starting to read and convert the data files.
- **Advanced** displays the **Advanced Flag Import** dialog box where you can change the default conversion settings.
- Help displays the appropriate Help page (URL).
- **12** Click **Advanced...** to access the **Advanced Flag Import** dialog box where you can change the conversion settings.

to you want GeneSpring to use the spot infor	mation in	FE to flag th	e data? Yes _
to you want to include the background readi	ngs in the	flag settings	? No
Spot Problems			
How do you want GeneSpring to flag feature	es with one	e of the follo	wing problems:
	Present	Marginal	Absent
Feature is saturated	0	0	۲
Feature is not uniform	С	0	۲
Feature is not positive and significant	С	0	۲
Feature is a population outlier	С	(•	0
Feature is manually marked	С	(0
Background is not uniform	0	(0
Background reading is a population outlier	0	æ	С
Mark all control probes as Absent (this v	vill exclude	e them from	the normalization)

You can respond **Yes** or **No** to the prompt, *Do you want GeneSpring to use the spot information in FE to flag the data?*

- If you respond Yes, the first five choices in the **Spot Problems** box are enabled along with the **Mark all control probes as Absent** check box.
- If you respond **No**, the entire **Spot Problems** box is disabled along with the *Do you want to include the background readings in the flag setting*? prompt. The new data file that the plug-in creates will contain a **GeneSpring Flags** column where everything is marked **P** for **Present**.

You can also respond **Yes** or **No** to the prompt *Do you want to include the background readings in the flag settings?*

- If you respond **No**, the last two choices in the **Spot Problems** box are disabled (**Background is not uniform** and **Background reading is a population outlier**) and both options are imported as **Present**.
- If you respond **Yes**, the last two choices are enabled and set to **Marginal**.

Alternatively, you can change the conversion setting for each FE flag value by selecting the appropriate radio buttons and pressing **OK** to begin importing.

See appendix A for a detailed explanation of the flag logic and how the settings in the **Advanced Flag Import** dialog box influence the outcome.

13 A window displaying a progress bar displays indicating that the preprocessing is active.



14 The **Suggested Normalizations for Agilent FE** dialog box displays automatically after you have completed the Flag Import settings. The version of this dialog box that displays depends on whether the plug-in recognized the file as being in 1-color or 2-color file format.

For 1-color data



For 2-color data



15 Here the import procedure is the same as the normal import procedure. Continue to import the samples.

K Import Dat	a: Sample Attributes				668
	Please selec	t when for sample a	bilicites.		
2	Sample Name	1			New Atribute
Allolitudo Narso		Arrayticsage	Astha	Experiment Ty	Colorado de Maleria
Attribute Units					EDUSTION PORTS
Nameric		80	ne .	89	Celeter Atticute
1	Processed US14702394_16011521020744_502_A01.M				
2	Processed-U014702304_16011521020745_002_A0164				Resista Terf
3	Processes-US18702384_16811621020844_582_A01.bd				Preparate range
4	Processe# U014702304, 16011521020065, 002, A01.M				
5	Processed-US14702384_16011521021071_502_401.01				Fill Desserve Davn
					Cort
	PTRABUS.	Net. Carcel	нер		

Agilent FE normalization scenarios

When all samples are loaded and used to build an experiment, we recommend that you apply the **Agilent FE** or **Agilent FE (1-Color)** scenario.

🕵 New Exper	iment Checklist		
You are al should se choose yo the button	most finished creating your experime t up its normalizations, experimental j ur default experiment interpretation. Y s below. Alternatively, you may find the	nt. Before you begin analysis,) parameters, and error model, a /ou may reach these windows em in the Experiments menu.	you and using
New Experimen	t Checklist		
	Define Normalizations	Normalizations	
0	Define Parameters	Parameters	
	Define the Default Interpretation	Experiment Interpretation	
	Define the Error Model	Error Model	
	Close		

To apply the normalization scenario, click the **Normalizations button** in the **New Experiment Checklist** dialog box or click **Experiments > Experiment Normalization**.

Contraction of the second seco		Ondor of Normalizations to Perform		
And wear period on weat of	1	Per Spot and Per Chip. Intensity dependent (Lowess) inormalization		Travers
Choose a Nermalization Step				
start with pre-ner wisible diverses				Mave Up
Octa Transformation: GAGE transform				Nove Drawn
Data Transformation. Real Time PCR transform				
Data Transformation: Subfract background based c				Use Defaults
Deta Transformation: Det me as uraments less than				
Data Transformation: Transform from log to linear v				
Dota Transformation: Dye pwap				
Per Spot. Divide by control channel				
Data Transformation: Resarve control channel				
Per Soct and Per Chip: Intensity dependent Cupyres				
Per Chip: Normalge to a median or percentile				
Por Chip: Normalize to positive control genes	- 1h	Recommended Ander And Text Press stations	10.7	
Per Chip: Normalize to a constantivalue		energenergenergenergenergenergenergener		
Por Oane: Normalize to specific camples	Ųs	e a Daved Buenaria Dave As Doemaria		
Per Owne: Normalize to median	-946	ringi		
Per Chip and Per Genix Median polititing		No warsings.		
· · · · · · · · · · · · · · · · · · ·				

The default two-color data normalization scenario is the "Per Spot and Per Chip: Intensity dependent (Lowess) normalization". Since Agilent Feature Extraction already applies Lowess normalization to the data, there is no need for a second Lowess normalization (technically, however, it would not be detrimental).

For Agilent FE 2-color data, a more appropriate normalization scenario is to simply apply a ratio calculation of the two channels, followed by a per-gene normalization to ensure all ratio's for each gene are normalized to 1.

During installation of the plug-in, a **Saved Scenario** was installed to perform this normalization. To apply the normalization scenario:

1 Select Use a Saved Scenario...

🕵 Select a Norma	lization Scenario)	
	Agilent FE No Normaliza Standard On Standard Two	ations e Color o Color	
Load Scenario	Delete Scenario	Rename Scenario	Close

2 Select the Agilent FE or Agilent FE (1-color) scenario from the list and press Load Scenario.

See appendix B for details on the Agilent FE Normalization scenario.

3 The new normalization scenario is loaded and displayed in the Normalization window.

And Alexandrative Property 1		Order of Normalization	as to Perform	-	inspect
AND NOT TRACED IN SIZE **		1 Per Spot Dwide by con	éról channéi		Culture
Choose a Normalization Step	4	2 Per Chip: Normalize to	503 percentle		Centa
Etad with pre-normalized values		3 Por Gone: Nervicibe h	medion		MONP UP
Data Transformation SAGE transform					Mirve Dones
Data Transformation: Real Time FCR transform					
Data Transformation: Dubtract background backed on negative controls					Use Default
Data Transformation: Set measurements less than 8.01 to 0.01					
Data Transformation: Transform from log to linear values					
Data Transformation Dys ewap					
Per Spot. Divide by control channel					
Data Transformation: Reserve central channel					
Per Spot and Per Chip Intensity dependent (Lowesto normalization					
Per Chip: Normalize to a median or percentile					
Fer Chip: Normalize to positive control games		Like Recommended Order	Get Text Description	-	
Per Chip: Normalize to a constant/value		Contraction in the other	Germanzeningene		
Per Genal Normalize to operific samples		Use a Sared Scenario	Save As Scenario		
Per Cene: Normalize to median		Warnings			
For Chip and Por Gone: Median polishing			No warnings.		

4 Press **OK** to apply the new normalization scenario to the experiment.



After the Agilent FE normalization scenario is applied, the data is ready to be analyzed in GeneSpring GX in the usual manner.

Appendix A – Flags

Agilent FE output data contain a large number of flags that can be used to indicate various problems with a feature on an array. Usually, this flag information can be easily used in GeneSpring GX after the data files are imported (using the "Filter on Data file" filter). However, there a number of features in GeneSpring GX that cannot use this flag information unless the information is first converted into special GeneSpring flags. The Agilent FE plug-in converts the complex set of FE flags into three-levels of GeneSpring flags: Absent (A), Marginal (M). or Present (P). The flag columns from Agilent FE output and their meaning are provided in table 1.

Flag description	Column headers in Agilent FE output		
Feature is saturated	glsSaturated, rlsSaturated		
Feature is not uniform	glsFeatNonUnifOL, rlsFeatNonUnifOL		
Feature is not positive and significant	glsPosAndSignif, rlsPosAdSignif		
Feature is a population outlier	glsFeatPopnOL, rlsFeatPopnOL		
Feature is manually marked	IsManualFlag		
Background is not uniform	glsBGNonUnifOL, rlsBGNonUnifOL		
Background reading is population outlier	glsBGPopnOL, rlsBGPopnOL		
Is Control Probe	Control Type		

Table 1Flag columns in Agilent FE output. The g and r used as the first letter of the column header nameindicate the Green (Cy3) and Red (Cy5) channels, respectively

Flag logic

To map the complex 14 binary flag columns (there are 4096 flag combinations) to the simple three-value GeneSpring flags (A, M, and P), a mapping that depends on the seriousness of the flags was applied. Certain flag values carry more weight that other flag values. The most serious flags are mapped as "Absent," medium serious flags are mapped as "Marginal," and all other features are mapped as "Present." Flag values in the FE output are *independent* flags, that is, a feature can be saturated, but uniform and a background population outlier. GeneSpring GX will keep the Lowest GeneSpring flag value (low is A, next is M then followed by P).

For example, if a feature is not flagged as either saturated or non-uniform, but flagged as a population outlier and not manually marked or a control probe (and not making use of the Background mapping), it will have a "score" of P,P,M,P,P,P,P and will therefore receive the GeneSpring flag "M" or "Marginal" (table 2).

Flag description	Value in FE output	GeneSpring flag
Feature is saturated	0	Pass
Feature is not uniform	0	Pass
Feature is not positive and significant	1	Pass
Feature is a population outlier	1	Marginal
Feature is manually marked	0	Pass
Background is not uniform	0	Pass
Background reading is population outlier	0	Pass
Is Control Probe	0	Pass
Result		Marginal

 Table 2
 Feature is a population outlier and gets marked Marginal since this is the lowest of flag scores

A second example concerns a feature that is flagged as a saturated feature and as a population outlier (and all other flags are not set in FE output). It gets marked as "A" because the feature is saturated. Since this is the lowest possible flag value in GeneSpring GX, the entire feature is flagged as "A" or "Absent" (table 3).

Flag description	Value in FE output	GeneSpring flag
Feature is saturated	1	Absent
Feature is not uniform	0	Pass
Feature is not positive and significant	0	Absent
Feature is a population outlier	1	Marginal
Feature is manually marked	0	Pass
Background is not uniform	0	Pass
Background reading is population outlier	0	Pass
Is Control Probe	0	Pass
		Absent

Table 3Feature is both saturated and a population outlier and gets flagged Absent since that is the lowest of
the flag scores

If the background information is used to determine the flag values, those features that are marked as either having a non-uniform background or are background-population outliers, get marked as "M" or "Marginal" probes by default, or whatever is set in the dialog box (table 4).

Flag description	Value in FE output	GeneSpring flag
Feature is saturated	0	Pass
Feature is not uniform	0	Pass
Feature is not positive and significant	1	Pass
Feature is a population outlier	0	Pass
Feature is manually marked	0	Pass
Background is not uniform	0	Pass
Background reading is population outlier	1	Marginal
Is Control Probe	0	Pass
		Marginal

 Table 4
 Feature is a background population outlier and gets marked Marginal since this is the lowest of flag scores

Appendix B – Agilent File Format

Agilent FE transfers these fields to GeneSpring GX.

	Agilent FE fields					
GeneSpring information	For 2-color data	For 1-color data				
Gene Name	ProbeName (If column is empty, use SystematicName. If that column is empty, use FeatureNum.)	Same				
Signal	rProcessedSignal	gProcessedSignal				
Reference (a.k.a. "Control" within GeneSpring)	gProcessedSignal	N/A				
User Provided Signal Precision	LogRatioError is converted using the following calculation. For a given gene in the FE output, let: RPS = rProcessedSignal (imported as Signal) GPS = gProcessedSignal (imported as Control) LRE = logRatioError Then import the following: $signalprecision = GPS \sqrt{(e^{LRE^2} - 1)}$	gProcessedSigError				
Description	Description	Same				
GenBank ID	GenBank	Same				
Flags	The preprocessor needs to translate the eight different flag settings.	Same				

All of the FEPARAMS and STATS fields are loaded as Sample

Attributes. The names of the Sample Attributes are the names provided on the FEPARAMS (or STATS) row. The values are the values displayed below the field name in the DATA row.

The type in the TYPE row can be used to determine if a Sample

FE type: integer or float → GeneSpring: numeric
 FE type: text → GeneSpring: non-numeric

Attribute is numeric or not.

Table 5 Interpreting the Agilent file format	
---	--

Sample Attributes

Same

Appendix C – Agilent Normalization Scenario for 1-Color Data

For optimal use of the expertise that is built into Agilent Feature Extraction software, Agilent recommends that you use a normalization scenario that is different than the default one-color scenario. This recommended normalization scenario makes use of the converted flag information that allows you to exclude the control and otherwise flagged probes.

The recommended Agilent FE(1-Color) normalization involves three steps:

The Agilent FE (1 Color) normalization scenario has three steps:

- **1** Data transformation: Set measurements from < 0.0 to 0.0
- **2** Per chip: Normalize to the median or percentile
- **3** Per gene: Normalize to median

Each step is discussed in detail.

Add Normalization Clan >>			Order of Normalizations to Perform	۸	Inspect
Add Normalization Step 22		1	Data Transformation: Set measurements less than 0.01 to 0.01		Delete
Choose a Normalization Step	^	2	Per Chip: Normalize to 50th percentile		
Start with pre-normalized values		3	Per Gene: Normalize to median		Move Up
Data Transformation: SAGE transform					Move Down
Data Transformation: Real Time PCR transform					
Data Transformation: Subtract background based on negative controls					Use Defaults
Data Transformation: Set measurements less than 0.01 to 0.01					
Data Transformation: Transform from log to linear values					
Data Transformation: Dye swap					
Per Spot: Divide by control channel					
Data Transformation: Reserve control channel					
Per Spot and Per Chip: Intensity dependent (Lowess) normalization					
Per Chip: Normalize to a median or percentile				-	
Per Chip: Normalize to positive control genes			Use Recommended Order Get Text Description		
Per Chip: Normalize to a constant value			Lies a Payod Peoparia		
Per Gene: Normalize to specific samples			Viewinge		
Per Gene: Normalize to median			wannigs		
Per Chip and Per Gene: Median polishing					
	-				

Data transformation: Set measurements from < 0.01 to 0.01

The data transforming normalization involves applying a mathematical modification to a variables values. During normalization, GeneSpring GX recalculates the data values and uses them in subsequent analyses.

The Set Measurements option sets any measurements less than a specified cutoff value to the cutoff value. By default, this value is 0.01.

• For Agilent FE (1 Color) data, set the cutoff value to the default by typing **Cutoff: 0.01** in **Cutoff** text box and click **OK**.

🌊 Data Transformation: Set measurements less than 0.01 to 🔳 🗖 🔀
This sets values below the cutoff to the cutoff. Cutoff: 0.01
Apply only to Specific Samples
OK Cancel Help

By default, this cutoff value is set to 10, but for this example scenario we have lowered this value to 0.01 to allow lower values.

Per chip: Normalize to the median or a percentile

Per-chip normalizations control the intensity of chip-wide variations. These variations may result from inconsistent washing, inconsistent sample preparation, or other microarray production or microfluidics imperfections.

The Normalize to a Median or Percentile option lets you divide all of the measurements on each chip by a specified percentile value. By default, this value is 50.0%.

• For Agilent FE (1 Color) data, type 50.0% in the percentile text box if the default value is not already displayed.

- Check the **Use only measurements flagged** box and set the drop-down menu to **Present Only**
- Click the Never apply extra background correction button

Click **OK** when done.

Normalize each	chip to the 50 th Percentile of the measurements taken from that chip
Restrict Measureme	ents used in the Calculation of the Percentile
Vse only meas	urements flagged Present Only
Use only meas	urements with Raw Signal 💌 values of at least 10
Background Correct	tion
	Never apply extra background correction
	Always apply extra background correction
	C If needed apply extra background correction

Identical chips from different samples may show different intensities due to external factors. You can offset this by asuming that all genes were functioning normally when sampled, and that their intensities are normal for what they were doing. By calling that normal state 1 for every chip in the experiment, you can normalize these variations in gene intesities.

These normalizations are only made on genes considered normal and flagged Present. Genes flagged Marginal or Absent are ignored to avoid skewing the data.

Per gene: Normalize to the median

The last step in normalizing Agilent FE (1 Color) data involves normalizing the expression value for each gene across all arrays.

• For Agilent FE (1 Color) data, set the cutoff value to 0.01 in **Raw** Signal data measurement values, and click **OK**.

Normaliz	e each gene to th	e median of the	measurement	s for that gene.	
Cutoff If the median is be Cutoff Value:	Now the cutoff, the	e cutoff value wi	I be used inste	ad. (See Help for de	tails.)
Cutoff Value:	1.0E-6 in	Raw Signal	m	asurement values	
Apply only to Specif	c Samples				
added and a abream	e eanipree				

There are two options that can be set in this normalization step: (1) a cutoff value can be set to ensure that values lower than the cutoff will be replaced with the cutoff value and (2) the cutoff value can be applied to either the normalized value or the Raw signal value.

The recommended normalization step has a cutoff set for the Raw signal value of $1.0E^{-6}$. This, together with the cutoff value for the Control value of " $1.0E^{-6}$," ensures that all normalized values are sensible.

Normalization completed

When you have completed the three-steps for normalizing Agilent FE (1 Color) data, the following will have occurred:

- Values below 0.01 were set to 0.01.
- Each measurement was divided by the 50.0th percentile of all measurements in that sample.
- Each gene was divided by the median of its measurements in all samples. If the median of the raw values was below 10 then each measurement for that gene was divided by 10 if the numerator was above 10, otherwise the measurement was thrown out.

Appendix D – Agilent Normalization Scenario for 2-Color Data

For optimal use of the expertise that is built into Agilent Feature Extraction software, Agilent recommends that you use a normalization scenario that is different than the default two-color scenario usually used. This recommended normalization scenario also makes use of the converted flag information that allows you to exclude the control and otherwise flagged probes.

The recommended normalization involves three steps:

- **1** Per spot: Divide by the control channel
- **2** Per chip: Normalize to the 50th percentile
- **3** Per gene: Normalize to the median

Each step will be discussed in detail.

K Experiment Normalizations: Demonstration E	xpe	riment		🛛
Add Normalization Oten sa		Order of Normalizations to Perform	-	inspect
Mod Normalization step >>	1	Per Spot. Divide by control channel		Delete
Choose a Normalization Step	2	Per Chip: Normalize to 50th percentile		
Start with pre-normalized values	3	Per Gene: Normalize to median		Move Up
Data Transformation: SAGE transform				Move Down
Data Transformation: Real Time PCR transform				
Data Transformation: Subtract background based c				Use Defaults
Data Transformation: Set measurements less than				
Data Transformation: Transform from log to linear v				
Data Transformation: Dye swap				
Per Spot: Divide by control channel				
Data Transformation: Reserve control channel			-	
Per Spot and Per Chip: Intensity dependent (Lowes	F	Use Recommended Order Get Text Description		
Per Chip: Normalize to a median or percentile				
Per Chip: Normalize to positive control genes	-	Use a Saved Scenano Save As Scenano		
Per Chip: Normalize to a constant value		warnings		
Per Gene: Normalize to specific samples				
Per Gene: Normalize to median				
Per Chip and Per Gene: Median polishing				
×				
		OK Cancel Help		

Per spot: Divide by control channel

GeneSpring GX extracts two columns from the Agilent FE output file that represent the red and green channel signal values, after processing in Agilent Feature Extraction. These two channels, called *rProcessedSignal* (for the red or Cy5 channel data) and *gProcessedSignal* (for the green or Cy3 channel data) are loaded respectively as the *Signal* and *Control* channels in GeneSpring GX.

Any normalization scenario in GeneSpring GX involves the ratio between the Signal and Control channels. The first step in the normalization scenario is to create this ratio with the normalization step "Per Spot: Divide by control channel."

This divides two color ex	the Signal Chann periments if you do	el by the Co not use th	ntrol Channel. e intensity dep	This is recor andent norm	nmended for alization.
Cutoff					
If the Control Ch	iannel is very low a	cutoff value	will be used i	nstead. (See	Help for details
		Cuton: II	oeto		

The Control Channel cutoff value is the only option that you can set for this normalization step. If the Control Channel value is below the cutoff value, the cutoff value is used for the control channel to ensure that sensible ratios are calculated.

By default, this cutoff value is set to 10, but for this example scenario we have lowered this value to $1.0E^{-6}$ to allow lower values for the control channel.

Per chip: Normalize to 50th percentile

The second normalization step normalizes the expression value to the median of the expression values on the array. This ensures that the expression values for each array can be compared across chips by centering each of the distributions for each array around 1, making comparison across arrays possible.

When you determine the median value for the distribution of signal values, you can exclude from the calculation each probe that has been flagged as being unreliable. This ensures that the median calculation and normalization is not skewed by outliers or control probes. It also ensures that normalized values are sensible.

Normalize each	a chip to the 50 th Percentile of the measurements taken from that chip
Restrict Measurem	ents used in the Calculation of the Percentile
Vse only meas	urements flagged Present Only
Use only meas	urements with Raw Signal values of at least 1.0e-6
Background Correc	tion
	Never apply extra background correction
	C Always apply extra background correction
	C If needed apply extra background correction

The recommended scenario only considers probes that are marked with a "Present" or P GeneSpring flag. This is indicated by the Normalization-step option "Use only measurements flagged Present Only."

Preprocessing the data using the Agilent FE plug-in's default settings results in the *exclusion* of probes that are control probes, saturated, feature population outliers, non-uniform features, or manually flagged. To include the features that are only flagged to be marginal, the option can be changed by selecting **Present or Marginal** from the drop-down list.

NOTE: The features are only excluded in calculating the median value. The features are not excluded from the experiment and continue to be available for review or analysis. The excluded features will also have the same normalization applied to when compared to the non-excluded probes.

Per gene: Normalize to median

The last normalization step normalizes the expression value for each gene across all the arrays around 1. This allows you to compare the genes across all the conditions or arrays, regardless of the actual normalized expression value (ratio).

K Per Gene: Normalize to median	
Normalize each gene to the median of the measurements for that gene.	
Cutoff	
If the median is below the cutoff, the cutoff value will be used instead. (See Help for Cutoff Value: 1.0E-6 in Raw Signal reasurement value)	details.) es
Apply only to Specific Samples	
OK Cancel Help	

There are two options that can be set in this normalization step: (1) a cutoff value can be set to ensure that values lower than the cutoff will be replaced with the cutoff value and (2) the cutoff value can be applied to either the normalized value or the Raw signal value.

The recommended normalization step has a cutoff set for the Raw signal value of $1.0E^{-6}$. This, together with the cutoff value for the Control value of $1.0E^{-6}$, ensures that all normalized values are sensible.

Appendix E – Extracted Sample Attributes

The Agilent FE plug-in extracts a number of fields from the Feature Extraction output file and stores them as Sample Attributes in GeneSpring GX. The fields that are extracted from the FE output file are located in line 2 of the FE output file marked "FEPARAMS" and in line 6 marked "STATS."

=	Aicrosoft Ex	cel - US14702384_16011	521021078_502_	AD1.txt				1	
18	Ele Edit	View Insert Format To	ols Data Window	v Help	T	pe a que	stion for h	elp 🖌 💌	. 8 ×
Lo	11.4	- Local			101 6 10		al de de	1 m . A.]
2	M A	a sha	- 10 - B X		21 9 74		1 14 14	· · · ·	· 🗕 ·
1	230	S 20 5 5 2 8 61 7	Reply with Change	s End Revier	New				
-	R2C1 .	A FEPARAMS							
	1	2	3	4	5	6	1	8	9
1	TYPE	text	text	text	text	text	integer	integor	feat
2	FEPARAMS	FeatureExtractor UserName	FeaturaExtractor Co	FeatureExtractor	SpotFinde	SpetFinde	SporFinde	SpotFinder	SpotF
3	DATA	deng	CORBAMITE	Friday, October	1482760-	Version A	105	215	
-4									
6	TYPE	feat	float	fost	integer	float	fost	float	intoge
6	STATS	gDarkOffsetAverage	gDarkOffsetMedian	gDarkOffsetStdD	gDark Offs	rDarkOffse	DarkOffse	rDarkOffse	rDarki
7	DATA	16 503	17	4.45244	1000	16.67	17	4.16128	
8	-								
9	TYPE	irtoger	intagor	integer	text	text	text.	integor	intege
10	FEATURES	FeatureNum	Row	Col	SwissProt	GenBank	GenPept	ProbeUID	Contri
11	DATA	1	1	1				0	
12	DATA	2	1	2				1	
13	DATA	3	1	3			AAFG2490	2	
14	DATA	4	1	4			AAP62941	3	
10	DATA	0	1	0			CAC36455	4	
10	DATA	/			010513			0	
1/	DATA	8		0	Q10513		ANDERES	0	
10	DATA						AAC 220022		
-20	DATA	13		13			CAD10178		
21	DATA	14	1	14			CAPITO INC	5	
22	DATA	15	1	15			AAFOTTN	1 10	
255				1.10	mainti		rearing		-
14 .	· · · /USIA	1/02394_16011521021078_9	02_A/	J¢					N 1
Rea	ty .								

All items from the FEPARAMS are extracted as Sample Attributes. The names of the Sample Attributes are the same as the name of the FEPARAMS or STATS fields..

www.agilent.com

In this book

This book contains brief instructions to help you get started with your Agilent Feature Extractor Plug-In for GeneSpring GX.

The information includes instruction on

- Installing the Feature Extraction Plug-In
- Loading and normalizing data produced by Feature Extraction Software
- Identifying flags related to specific types of problems

© Agilent Technologies, Inc. 2005

Printed in USA First Edition, October 2005

