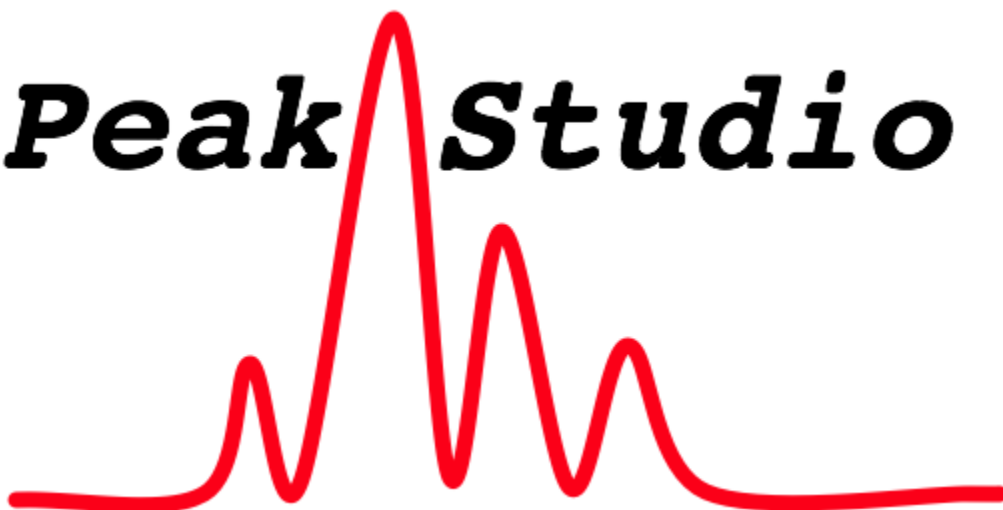


# *Peak Studio*



©2012 Fodor Lab UNCC

## **Contents**

<b>Chapter 1</b>	<b>Using PeakStudio</b>
	<b>Introduction</b>
	<b>Installation</b>
	<b>System Requirements</b>
	<b>Running from Command Line</b>
	<b>Terms used with PeakStudio</b>
	<b>An overview of the workflow</b>
<b>Chapter 2</b>	<b>Dropdown menus in PeakStudio</b>
	<b>File</b>
	<b>Edit</b>
	<b>View</b>
	<b>Axis</b>
	<b>Analysis</b>
	<b>Help</b>
<b>Chapter 3</b>	<b>PCA in PeakStudio</b>
<b>Appendix A</b>	<b>Peak Diagram and Peak Calling</b>
<b>Appendix B</b>	<b>QC Number</b>
<b>Appendix C</b>	<b>Automated Peak Adjusting</b>
<b>Acknowledgements</b>	

# Chapter 1

## Introduction

Welcome to Peak Studio, a program for viewing and analyzing fragment analysis files generated by ABI capillary electrophoresis instruments. PeakStudio is an open source program developed at UNCC in the department of Bioinformatics and Genomics by Jon McCafferty.

PeakStudio was developed in order to help make more objective decisions about fragment analysis files. It allows users to view any file with a .fsa file extension, assign sizing to peaks, manually edit sizing calls and generate PCA plots to analyze the grouping of data.



(PeakStudio screen shot)

## Installation

To install Peak Studio, consult the requirements listed below, then download the .jar file. To use the program, you can double click on the icon or run from a command line (details below).



PeakStudio.jar  
(Windows icon)



PeakStudio.jar  
(Mac icon)

## System Requirements

### Mac OS X

- Mac OS X 10.5 (Leopard) or higher
- Intel processor
- 1 GB Ram
- 4 MB available disk space
- Java version 6

Peak Studio was built and tested on Mac OS X 10.6.4 (Snow Leopard) with 2.8GHz Intel Core 2 Duo processor and 2 GB of RAM.

### Windows

- Windows XP
- Intel processor
- 1 GB RAM
- 4 MB available disk space
- Java version 6

Peak Studio was tested on Windows XP with a 3.0GHz Intel Pentium processor and 2GB of RAM.

## Running from command Line

### Mac OS X

Open a terminal window, found in the Applications/Utilities folder. From here, change to the directory where PeakStudio.jar is located using "cd yourpath/" (without the quotation marks) and run using this command: `user$ "java -Xms128m -Xmx1024m -jar PeakStudio.jar"` (without the quotation marks).

### Windows

The Command Prompt application can be found in All Programs \ Accessories, or accessed by clicking Start then Run and typing "cmd" and clicking OK. From here, change to the directory where PeakStudio.jar is located using "cd yourpath\" (without the quotation marks) and run using this command: `C:\path> java -Xms128m -Xmx1024m -jar PeakStudio.Jar` (without the quotation marks).

Before using PeakStudio, we recommend reading through the user manual and/or watching our tutorial videos at (<http://www.fodorlab.uncc.edu/PeakStudioPage.html>) . In the next section, we will discuss some terms used with PeakStudio that may be unfamiliar.

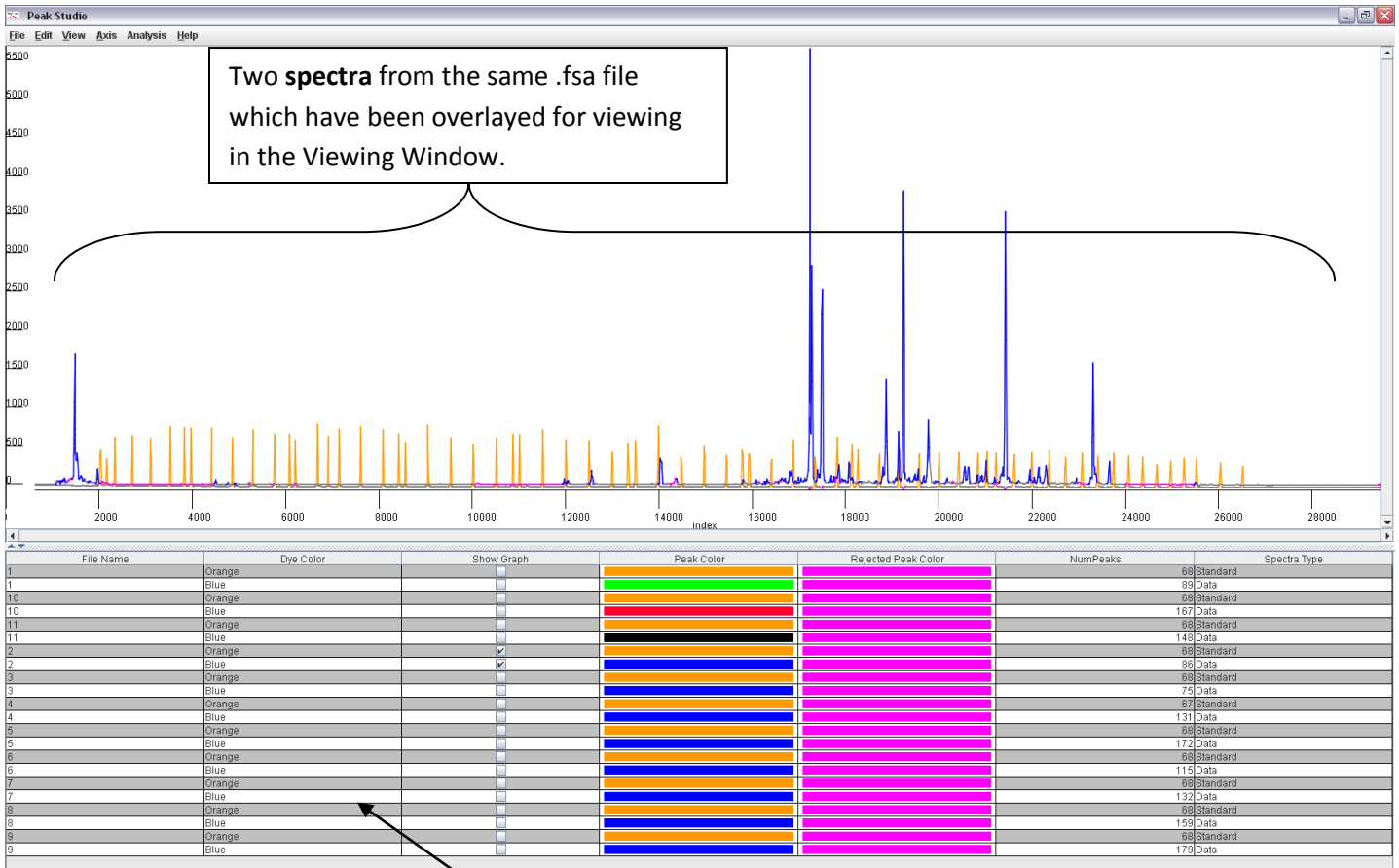
## Terms used with PeakStudio

This manual contains some terms that are specific to PeakStudio and some which have a particular meaning for our purposes in regard to PeakStudio. A list follows with several of these terms, which will help as you read through the rest of the manual:

**Bin** – Bins are used to group peaks for analysis. Smaller bin sizes are more exclusive while larger bin sizes cause peaks of similar size to be grouped with each other.

**Spectra** – In the context of PeakStudio, the word “spectra” is used to refer to one of the 5 possible colors captured by the CCD camera in the genetic analyzer. Every .fsa file contains the data from the different colors (channels), but a spectra is one color separated from the others.

**Table View** – This is the bottom portion of the PeakStudio window. Table View is essentially a customizable table that lists details regarding samples that have been imported for analysis.



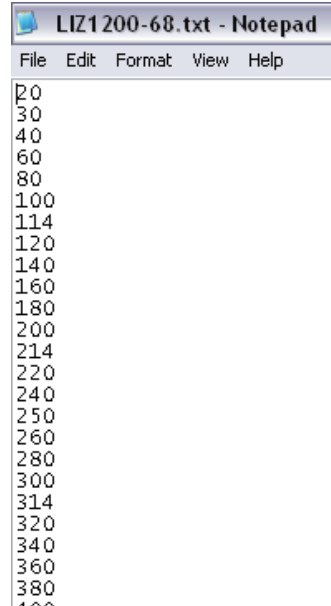
TableView

## An overview of the workflow

PeakStudio has been designed to provide a straightforward method of data viewing and analysis. Below is a walkthrough of a typical project.

### To begin

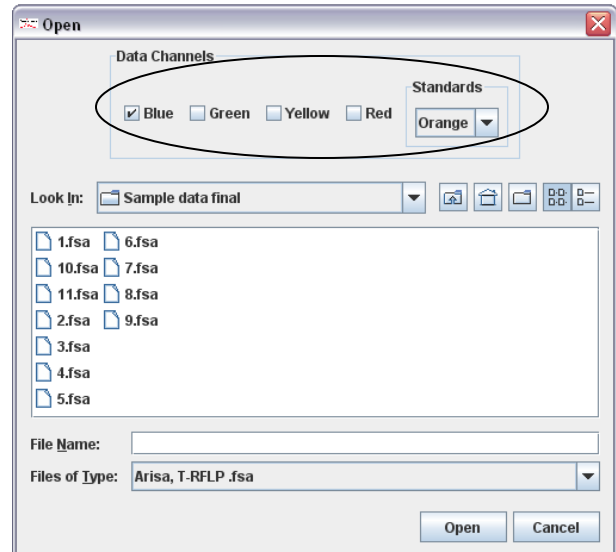
You should have the PeakStudio.jar file downloaded on your computer (or the source code). You will need .fsa files to analyze and a size standard file. The size standard file indicates which size standard fragments were used and should be in a .txt file format that contains a list of each size separated by a return (each size on its own line).









```
LIZ1200-68.txt - Notepad
File Edit Format View Help
20
30
40
60
80
100
114
120
140
160
180
200
214
220
240
250
260
280
300
314
320
340
360
380
400
```

### Import samples

At the upper left of the window go to **File, Open Spectra** and select the file(s) to be analyzed. You must select the color of the dye labeled products (in ARISA, this is typically Blue for FAM) and the color of the size standard dye (by default this is Orange for LIZ). If you are using other dyes for your products or size standards, you can specify that now. Click the **Open** button to open them into the Peak Studio window.

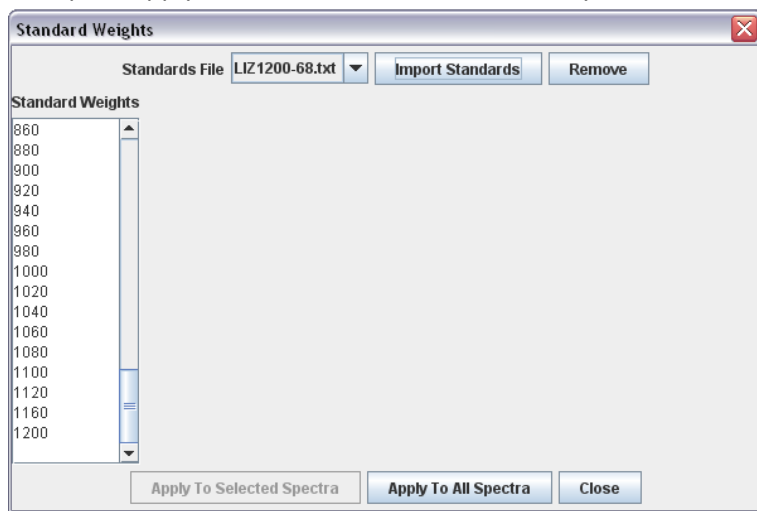


Files are displayed in the **Table View** (bottom portion of the PeakStudio window). Each .fsa file that you import will generate a separate spectra for each dye previously specified. With ARISA for example, there should be two spectra; one for the blue, FAM labeled products and one for the orange, LIZ labeled size standard fragments.












File Name	Dye Color	Show Graph	Peak Color
1	Orange	<input type="checkbox"/>	
1	Blue	<input type="checkbox"/>	
10	Orange	<input type="checkbox"/>	
10	Blue	<input type="checkbox"/>	
11	Orange	<input type="checkbox"/>	
11	Blue	<input type="checkbox"/>	

*Add a standards file*

Go to **Edit, Standards**, click the **Import Standards** button, then find your standards file and **apply**. You can opt to apply the standards to all the files, or you can select certain files in TableView.



Once you have applied a size standard file to the spectra, you will want to determine whether the peak calling algorithm has identified the size standard peaks correctly. Begin by checking the NumPeaks column. If you are using LIZ-1200 from ABI, you should have 68 size standard peaks. Notice that sample 4 only contains 67 identified size standard peaks. We will correct this shortly.

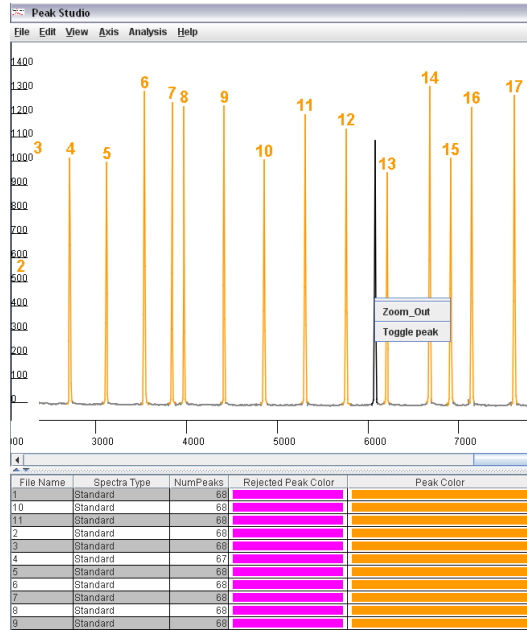
File Name	Spectra Type	NumPeaks	Peak Color
1	Standard	68	
10	Standard	68	
11	Standard	68	
2	Standard	68	
3	Standard	68	
4	Standard	67	
5	Standard	68	
6	Standard	68	
7	Standard	68	
8	Standard	68	
9	Standard	68	



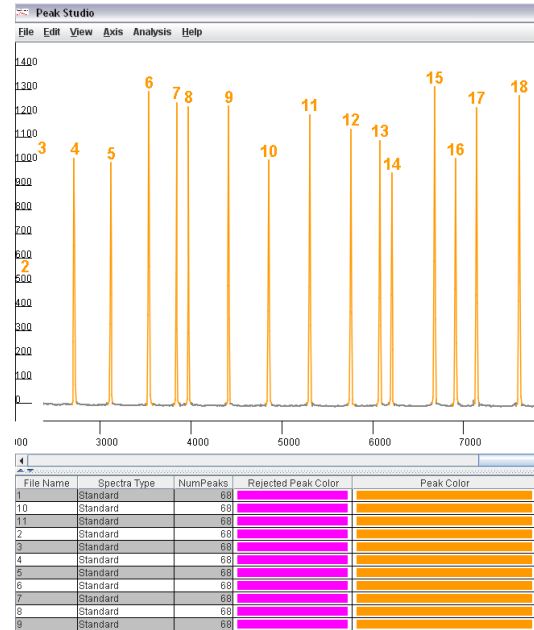
Right click anywhere in the Table View and select **Show columns**, then click on **QC number**. Generally, the lower the number, the more confidence that peaks have been called correctly. As a general guideline, a QC between 0.18 and 0.30 is good (see Appendix B). Any column in Table View can be used to sort the data by clicking on the header.

Click the check box for **Show Graph** to display the spectra. If you prefer, you can go to **Axis**, and check **Basepair** to display the spectra in basepair, rather than raw data format. Also in the **Axis** dropdown, you can choose to **Display** the X and Y axes. To see which peaks have been assigned with each size, go to **View, Show Peak Numbers**.

If you find that you are missing peaks, or that they have been incorrectly identified, you can manually correct this problem. Double click on the misidentified peak and then right click; this will bring up the option to **Toggle Peak**, which allows a peak to be reassigned from accepted to rejected or vice versa.

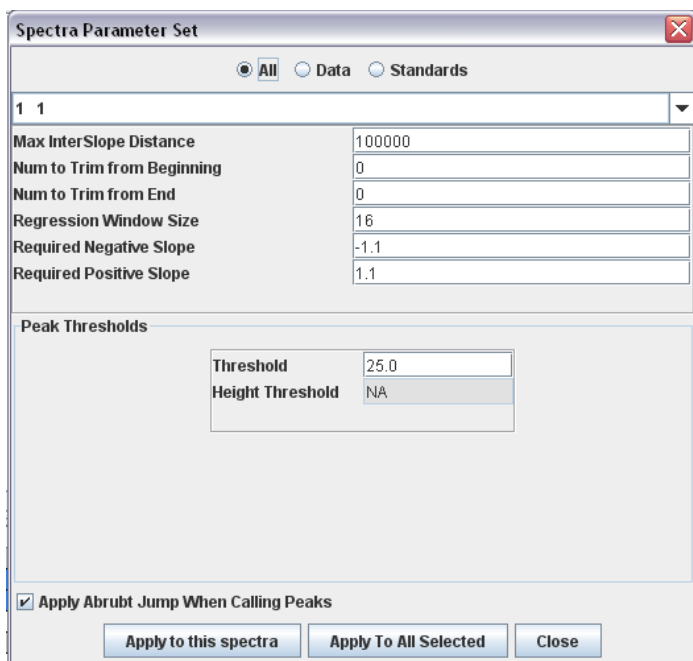


Before toggling the algorithm has missed this peak



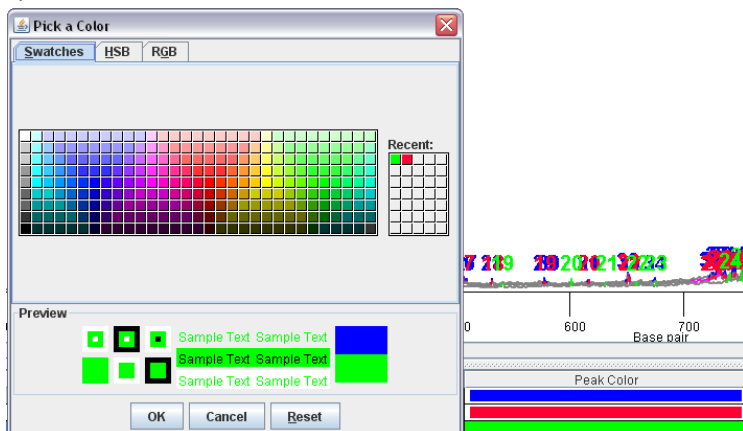
After toggling the peak is now correctly assigned

You can also choose to edit peak calling parameters by going to **Edit** then **Parameter Set** and then allow the software to automatically call peaks again. The modified peak calling parameters can be applied to a particular spectra, or applied to all selected spectra.

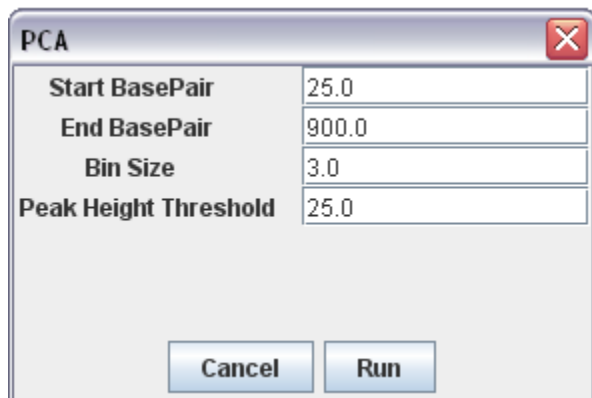


### Analyze your data

Once the standards spectra are acceptable you can choose to remove them from Table View by right clicking in Table View and selecting **Spectra**, then uncheck the box for your standards color. The sizing will remain in effect, but you will only see sample data now. As was the case for standards spectra, the **Show Graph** check box brings up each data spectra in the window. You can show one or many spectra at a time. Sometimes when viewing multiple spectra it can be helpful to change the display color. This is easily achieved by clicking on the **Peak Color** box for the spectra, then selecting a new color in the pop up window.



Now you can determine how closely related your samples are to each other using PCA. To do this, go to **Analysis** and click on **PCA**. This will bring up a window where you can specify which range of bases you would like to compare, how large your bin size should be, and what threshold to use for cutting off peaks. In capillary electrophoresis, very small and very large fragments are often not as reliably sized as those in the middle, so we often limit our analysis to what we consider to be the best fragments for those samples.



Start BasePair	25.0
End BasePair	900.0
Bin Size	3.0
Peak Height Threshold	25.0

Cancel Run

Note the Start BasePair of 25 and End BasePair of 900. Depending on the range of fragments of interest, you may want to make that range narrower or wider. Using the current setting, bins would be 25 – 28, 28 – 31, etc. Peak Height Threshold indicates the minimum y-axis value (Relative Fluorescence Unit) required for a peak to be considered.

Once you click **Run** a new window will pop up with a graph of your PCA. This allows you to visually examine the grouping of samples. You can mouse over data points to reveal which sample is represented or go to **View**, then click the **Name** radio button and all the names will be displayed on the graph. (Hint: you may find it useful to color code your samples to make them easier to visually inspect on the PCA graph. To do this, go back to **Table View** and click on the color of the sample you want to change. You can assign any color to any sample.) It is possible to export the data used to generate the PCA for further analysis by going to **Export**, then **Matrix**, then select **Data**.

## Chapter 2

### Dropdown menu options

#### File Dropdown Menu

**Open Spectra (Ctrl + O)** – Use this function to open one or many .fsa files

**Open Project (Ctrl + P)** – Use this function to open a .svaz project generated by PeakStudio

**Save (Ctrl + S)** – Saves the current project as a .svaz file

**Save As** – Save the current project to another location or with another name

**Import** – Allows the user to import metadata to associate with samples.

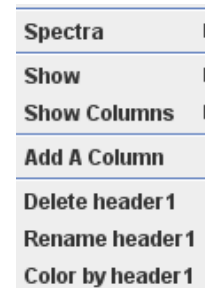
When importing metadata into PeakStudio, the file must be formatted as a tab delimited text file and it must be formatted using the following template:



The first column must be titled “sample name” and the remaining metadata columns can contain any title you choose. The sample names included in the metadata file must match the sample names loaded in PeakStudio.

sample name	anyheader1	anyheader2	anyheader3
1	2	A	U
2	4	S	J
3	4	X	K
4	8	A	K
5	6	D	K
6	12	D	J
7	12	A	U
8	2	S	U
9	2	A	U
10	4	S	U
11	7	X	J

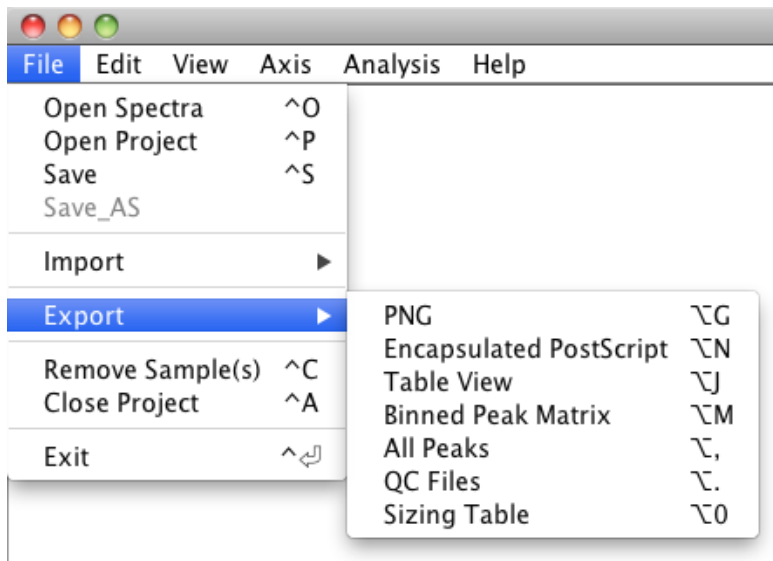
Once metadata have been assigned to the spectra, right clicking in a header column in Table View allows the spectra to be colored based on the data in that column.



File Name	header1	Peak Color	header2	header3
4	8	██████████	A	K
1	2	██████████	A	U
9	2	██████████	A	U
7	12	██████████	A	U
4	8	██████████	A	K
1	2	██████████	A	U
9	2	██████████	A	U
7	12	██████████	A	U
5	6	██████████	D	K
6	12	██████████	D	J
5	6	██████████	D	K
6	12	██████████	D	J
10	4	██████████	S	U
2	4	██████████	S	J
8	2	██████████	S	U
10	4	██████████	S	U
2	4	██████████	S	J
8	2	██████████	S	U
11	7	██████████	X	J
3	4	██████████	X	K
11	7	██████████	X	J
3	4	██████████	X	K

Spectra which have been colored by header2, then sorted by header2

## Export



**PNG** – Export a .png screenshot picture of the spectra currently displayed

**Encapsulated PostScript** – Export a .eps screenshot picture of the spectra currently displayed

**Table View** – Export a spreadsheet of the samples’ details in Table View

	A	B	C	D	E	F	G
1	File Name	Dye Color	Show Gra	Peak Colo	Rejected f	NumPeak	Spectra Type
2	1	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
3	1	Blue	FALSE	(0,0,255)	(255,0,255	89	Data
4	10	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
5	10	Blue	FALSE	(0,0,255)	(255,0,255	167	Data
6	11	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
7	11	Blue	FALSE	(0,0,255)	(255,0,255	148	Data
8	2	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
9	2	Blue	FALSE	(0,0,255)	(255,0,255	86	Data
10	3	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
11	3	Blue	FALSE	(0,0,255)	(255,0,255	75	Data
12	4	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
13	4	Blue	FALSE	(0,0,255)	(255,0,255	131	Data
14	5	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
15	5	Blue	FALSE	(0,0,255)	(255,0,255	172	Data
16	6	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
17	6	Blue	FALSE	(0,0,255)	(255,0,255	115	Data
18	7	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
19	7	Blue	FALSE	(0,0,255)	(255,0,255	132	Data
20	8	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
21	8	Blue	FALSE	(0,0,255)	(255,0,255	159	Data
22	9	Orange	FALSE	(255,153,0	(255,0,255	68	Standard
23	9	Blue	FALSE	(0,0,255)	(255,0,255	179	Data

**Binned Peak Matrix** – Export the matrix of peaks and bins for any stats

	A	B	C	D	E	F
1		BP_25-28	BP_28-31	BP_31-34	BP_34-37	BP_58-61
2	1	0	0	33	0	0
3	10	0	54	0	0	0
4	11	0	44	0	0	0
5	2	0	0	0	0	0
6	3	0	0	0	0	0
7	4	0	32	0	0	0
8	5	51	45	62	0	35
9	6	35	50	0	0	0
10	7	0	44	0	32	0
11	8	35	47	0	0	0
12	9	42	58	0	0	29

**All Peaks** – Export peak information for a selected file

	A	B	C	D	E	F	G
1	Sample	Channel Number	Type	Peak Number	Location(Index)	Location(Base Pair)	Height
2	1	105	Standard	1	2083	20	667
3	1	105	Standard	2	2216	30	571
4	1	105	Standard	3	2402	40	1061
5	1	105	Standard	4	2764	60	1097
6	1	105	Standard	5	3165	80	1004
7	1	105	Standard	6	3589	100	1378
8	1	105	Standard	7	3893	114	1330
9	1	105	Standard	8	4025	120	1334
10	1	105	Standard	9	4470	140	1329
11	1	105	Standard	10	4913	160	1074

**QC Files** – Export a list of scan numbers and associated peak algorithm assignments

	A	B	C	D	E
1	index	basepair	data	feature	
2	0	-80.0806		7	-
3	1	-80.0319		-5	NON_PEAK
4	2	-79.9833		-1	NON_PEAK
5	3	-79.9347		0	NON_PEAK
6	4	-79.886		-2	NON_PEAK
7	5	-79.8374		-2	NON_PEAK
8	6	-79.7888		2	NON_PEAK
9	7	-79.7401		-4	NON_PEAK
10	8	-79.6915		4	NON_PEAK
11	9	-79.6429		-1	NON_PEAK
12	10	-79.5943		-6	NON_PEAK
13	11	-79.5456		-2	NON PEAK

**Sizing Table** – Export a table that conforms to the format output of GeneMapper software

Sizing Table: 6 columns							
Dye Color,Peak# ; File Name; BP location; Peak Height; Peak Area; Data Point							
	A	B	C	D	E	F	
1	O,1	Ecoli_A08_2012-01-17.fsa	20	455	13131.83	2556	
2	O,2	Ecoli_A08_2012-01-17.fsa	30	364	7322.92	2691	
3	O,3	Ecoli_A08_2012-01-17.fsa	40	604	8715.2	2877	
4	O,4	Ecoli_A08_2012-01-17.fsa	60	664	9987.64	3271	
5	O,5	Ecoli_A08_2012-01-17.fsa	80	636	9194.59	3699	
6	O,6	Ecoli_A08_2012-01-17.fsa	100	820	11409.38	4152	
7	O,7	Ecoli_A08_2012-01-17.fsa	114	814	10406.45	4477	
8	O,8	Ecoli_A08_2012-01-17.fsa	120	776	10430.55	4619	
9	O,9	Ecoli_A08_2012-01-17.fsa	140	805	10530.54	5096	
10	O,10	Ecoli_A08_2012-01-17.fsa	160	639	8571.19	5570	
11	O,11	Ecoli_A08_2012-01-17.fsa	180	765	10403.78	6053	
12	O,12	Ecoli_A08_2012-01-17.fsa	200	726	9826.93	6542	
13	O,13	Ecoli_A08_2012-01-17.fsa	214	692	9465.72	6886	
14	O,14	Ecoli_A08_2012-01-17.fsa	220	624	8417.16	7027	
15	O,15	Ecoli_A08_2012-01-17.fsa	240	622	11025.2	7526	

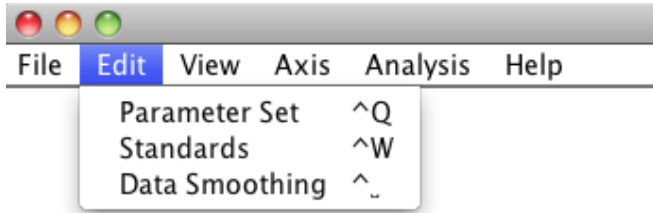
**Remove Sample(s) (Ctrl + C)** – Highlight one or multiple files in the Table view and remove

**Close Project (Ctrl + A)** – Closes the current project, but keeps PeakStudio open

**Exit (Ctrl + Enter)** – Closes PeakStudio



## Edit Dropdown Menu



**Parameter Set (Ctrl + Q)** – Allows the user to adjust peak calling parameters

A screenshot of the 'Spectra Parameter Set' dialog box. The dialog has a title bar with a close button. It contains three radio buttons: 'All' (selected), 'Data', and 'Standards'. Below this is a dropdown menu showing '1 1'. The main area contains several input fields for parameters:

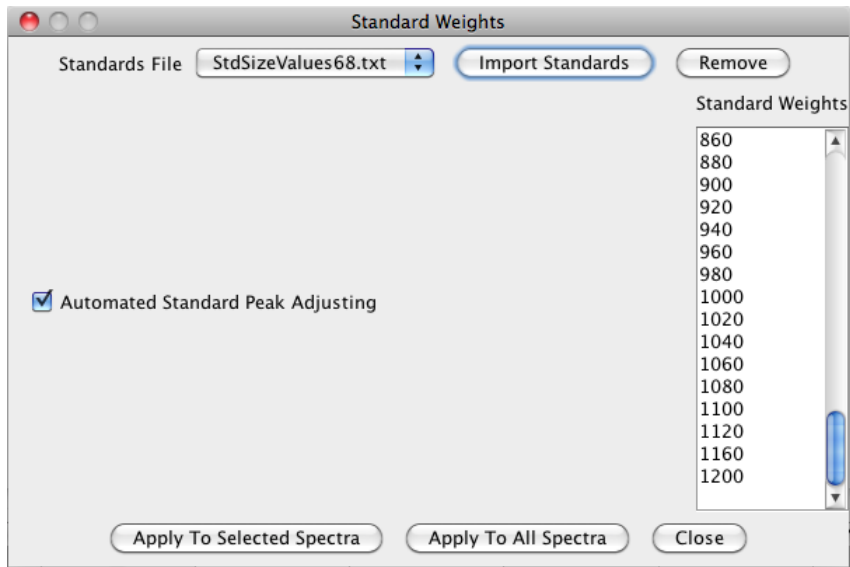
Max Inter Slope Distance	100000
Num to Trim from Beginning	0
Num to Trim from End	0
Regression Window Size	16
Required Negative Slope	-1.1
Required Positive Slope	1.1

Below these fields is a section titled 'Peak Thresholds' with two input fields:

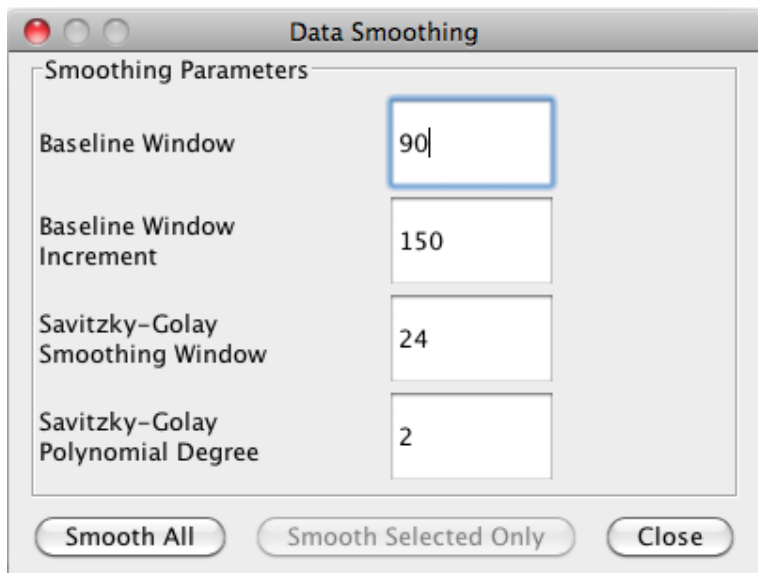
Threshold	25.0
Height Threshold	NA

At the bottom, there is a checked checkbox labeled 'Apply Abrupt Jump When Calling Peaks' and three buttons: 'Apply to this spectra', 'Apply To All Selected', and 'Close'.

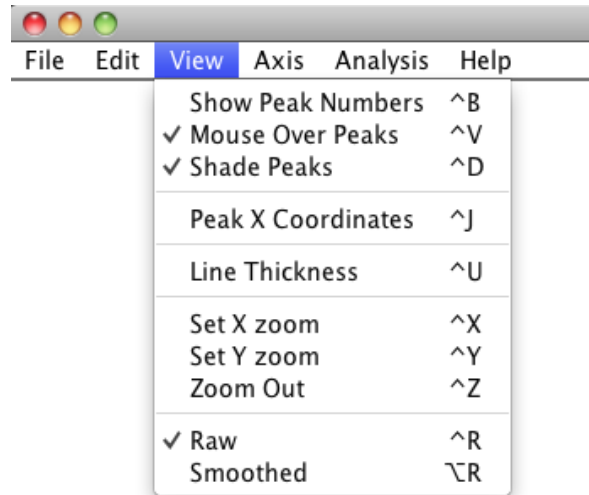
**Standards (Ctrl + W)** – Use this to import a standards file



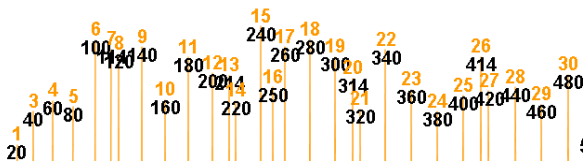
**Data Smoothing (Ctrl + -)** – Use this to smooth data



### View Dropdown Menu



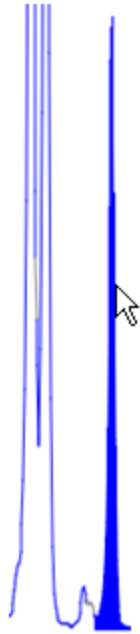
**Show Peak Numbers (Ctrl + B)** – Inserts numbers above peaks which have been called by the peak calling algorithm



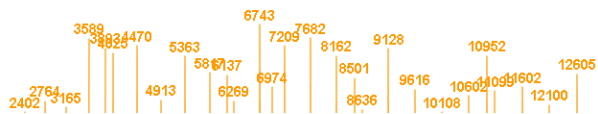
**Mouse Over Peaks (Ctrl + V)** – Displays information box when the mouse cursor is placed on a peak

1  
PEAK  
(14139.59, 294.47)

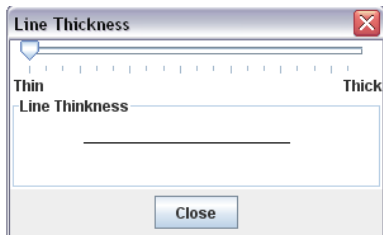
**Shade Peaks (Ctrl + D)** – Shades the area under a peak when the mouse cursor is placed on a peak



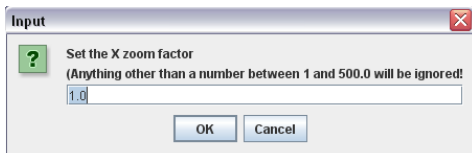
**Peak X Coordinates (Ctrl + J)** – Inserts peak numbers based on base pair or scan number, depending on the current view



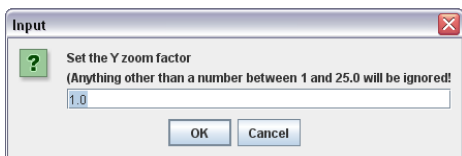
**Line Thickness (Ctrl + U)** – Adjusts the thickness of the displayed spectra



**Set X zoom (Ctrl + X)** – Sets the x axis zoom level



**Set Y zoom (Ctrl + Y)** – Sets the y axis zoom level



**Zoom Out (Ctrl + Z)** – Zooms spectra out to the original image size

**Raw (Ctrl + R)** – View data in raw form

**Smooth (Alt + R)** – View data in smoothed baseline corrected form

### Axis Dropdown Menu

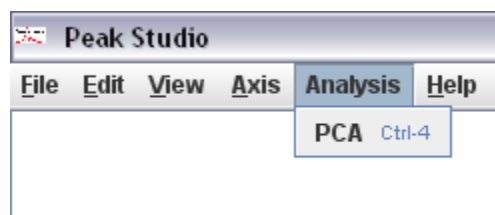
**Display** – Displays the y axis with relative fluorescence units (RFUs) and the x axis with either scan numbers (raw data) or base pair numbers (based on peak calling algorithm)

**BasePair (Ctrl + K)** – Converts the spectral display from scan numbers (raw data) to base pair numbers (based on peak calling algorithm)

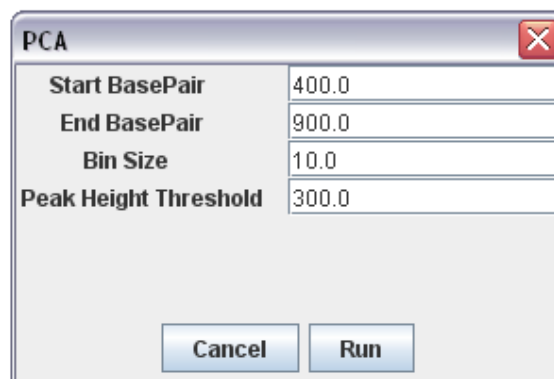


### Analysis Dropdown Menu

**PCA (Ctrl + 4)** – Opens a new window displaying principle component analysis (PCA) on the spectra open in the Table View.



Select the range of basepair values, the size of the bin to use and a value for the peak height threshold, then select Run.

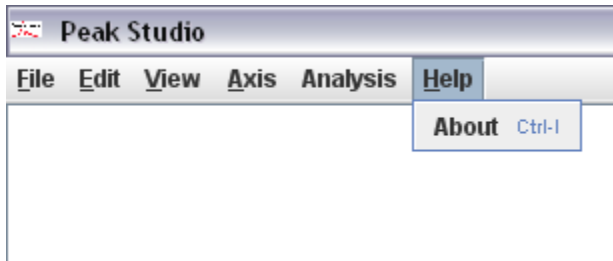


The PCA output opens in a new window. Hovering over data points reveals the identity of each point. Colors are associated with the Peak Color from the PeakStudio viewer window.

See Chapter 3 for description of PCA in PeakStudio.

## *Help Dropdown Menu*

**About (Ctrl + I)** – Contains version information



## Table View

Table view is the bottom panel of the PeakStudio window. Columns can be sorted by clicking on the column headers.

File Name	Dye Color	Show Graph	Peak Color	Rejected Peak Color	NumPeaks	Spectra Type
1	Blue	<input type="checkbox"/>			89	Data
1	Orange	<input type="checkbox"/>			68	Standard
10	Blue	<input type="checkbox"/>			167	Data
10	Orange	<input type="checkbox"/>			68	Standard
11	Blue	<input type="checkbox"/>			148	Data
11	Orange	<input type="checkbox"/>			68	Standard
2	Blue	<input type="checkbox"/>			88	Data
2	Orange	<input type="checkbox"/>			68	Standard
3	Blue	<input type="checkbox"/>			75	Data
3	Orange	<input type="checkbox"/>			68	Standard
4	Blue	<input type="checkbox"/>			131	Data
4	Orange	<input type="checkbox"/>			68	Standard
6	Blue	<input type="checkbox"/>			172	Data
6	Orange	<input type="checkbox"/>			68	Standard

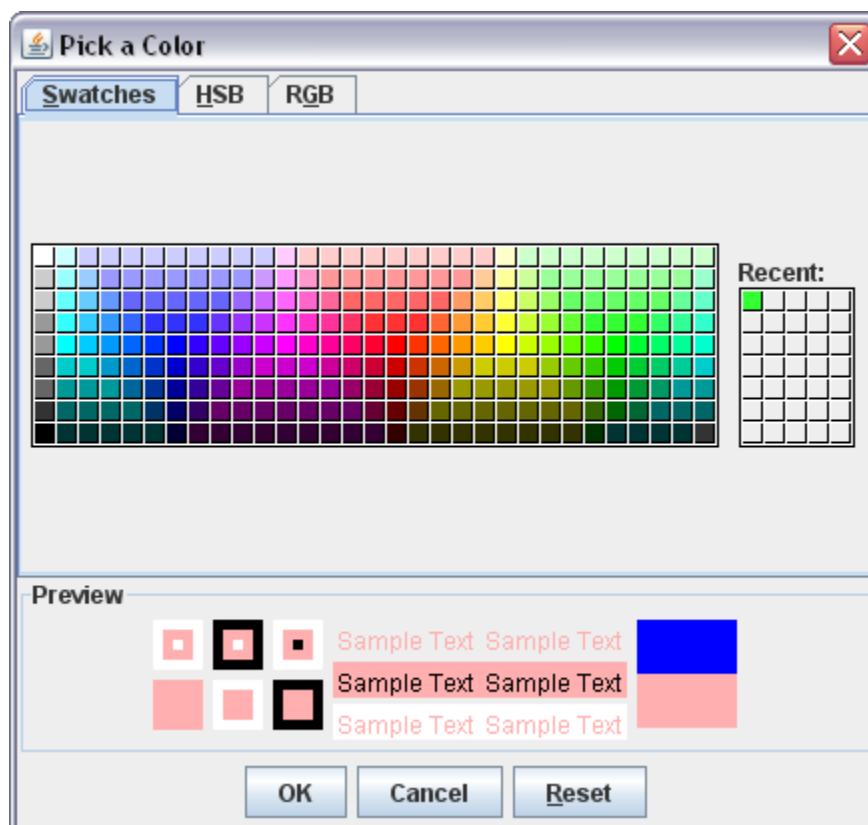
Default columns in Table view are:

**File Name** – Displays the file name associated with each sample

**Data channel** – Displays the data channel associated with the dye used

**Show Graph** – Check the box to display the electropherogram in PeakStudio. One or many graphs can be displayed at a time and peak colors can be adjusted to tell them apart.

**Peak Color** – Peak colors are associated with dye colors, but can be modified by clicking on the color switch in Table view.



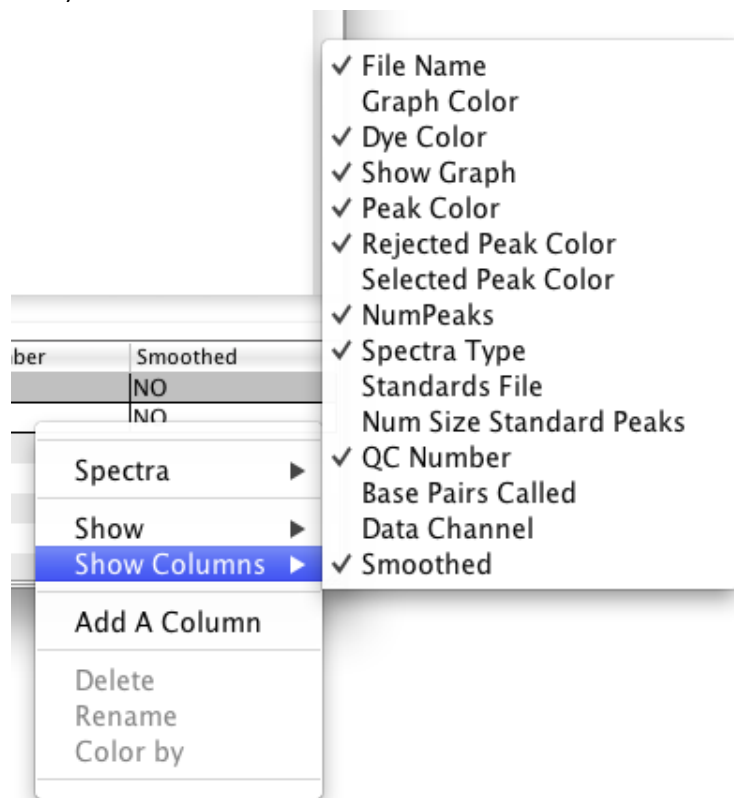
**Rejected Peak Color** – Peaks which are identified by the peak calling algorithm, but are rejected as non-real display as blue on the spectra.



**Spectra Type** – This indicates whether the file represents data from a sample, or the size standard associated with that sample.

**NumPeaks** – The number of real peaks identified by the peak calling algorithm

Additional columns available under Show Columns (right click the mouse when cursor is within Table View) include:



**Graph Color** – The color of the spectra which is not identified by the peak calling algorithm as a peak or rejected peak

**Selected Peak Color** – The color a peak displays when it has been selected by double clicking on it

**QC Status** – Displays “Good Data” or “Bad Data” depending on whether the peak identification algorithm was able to correlate the size standard peaks to their correct sizes

**Standards File** – The file name which was used to assign sizes to spectra is displayed

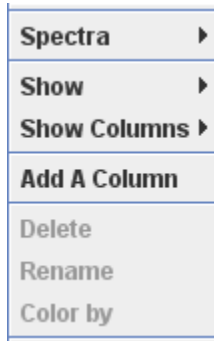
**Num Size Standard Peaks** – Displays the number of peaks in the assigned standards file

**QC Number** – A numerical representation of the quality of peaks based on the interpretation between expected peaks and observed peaks (See appendix B for further detail on QC number)

**Base Pairs Called** – Displays whether basepairs have been called for each sample

**Smoothed** – Lets user know if smoothing algorithm has been applied to the spectra

Additional right click options:



Spectra:

- 1 – blue dye (FAM, HEX)
- 2 – green dye (VIC)
- 3 – yellow dye (NED)
- 4 – red dye (PET, ROX)
- 105 – orange dye (LIZ)

Show:

- Select all – marks all the checkboxes to show all spectra on the screen at one time
- Unselect all – unmarks all the checkboxes to clear the screen of all spectra

Add A Column – Adds a blank column in which you can add your own text

Delete – Deletes a column that was added using the Add A Column function

Rename – Allows you to rename the column that was added

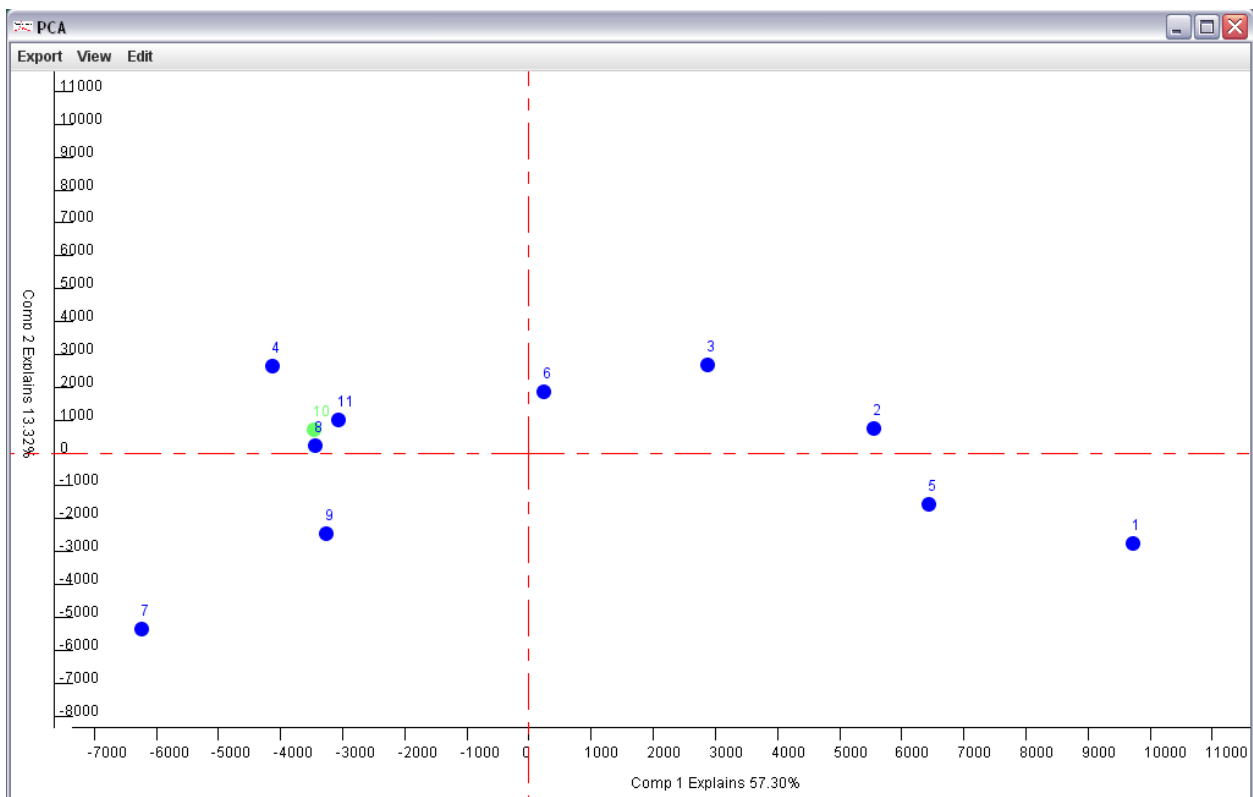
Color by – Changes the color of the spectra associated with the label in the column created.

## Chapter 3

### PCA in PeakStudio

PCA, or Principal Component Analysis, is an algorithm that identifies patterns by revealing similarities in a dataset. This is accomplished by transforming a high dimensional multivariate dataset into a set of principal components, allowing you to project the data onto a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first principal component, the second greatest variance on the second component, and so on. Data can now be visualized in a lower dimensional space creating a more informative view of the dataset.

The Data Matrix used as input to the PCA algorithm is generated with user-defined parameters. Using a start and stop location, bin width, and peak height threshold; the spectra is divided up into bins and the peaks which meet the height threshold are grouped into the bins. A binned matrix with columns representing bins and rows representing spectra is created with the cells containing the sum of peak heights, for each peak over the defined threshold. The PCA algorithm uses this binned data matrix as input.



## Menus in PCA Window

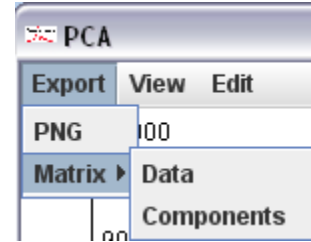
### Export

**PNG** – Export a .png screenshot picture of the PCA currently displayed

### Matrix

**Data** – Export the matrix of peaks and bins for any stats

**Components** – Export the matrix of the transformed data

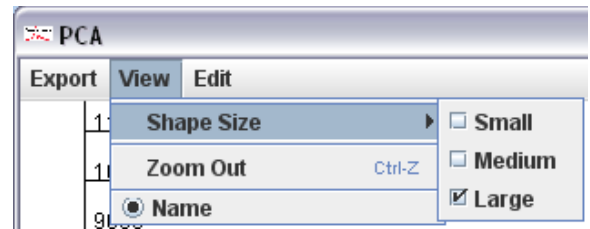


### View

**Shape Size** – This enables the size of data points to be adjusted between 3 levels

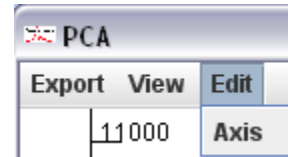
**Zoom Out (Ctrl + Z)** - Zooms PCA window out to the original image size

**Name** – Displays the name of the spectra associated with each data point



### Edit

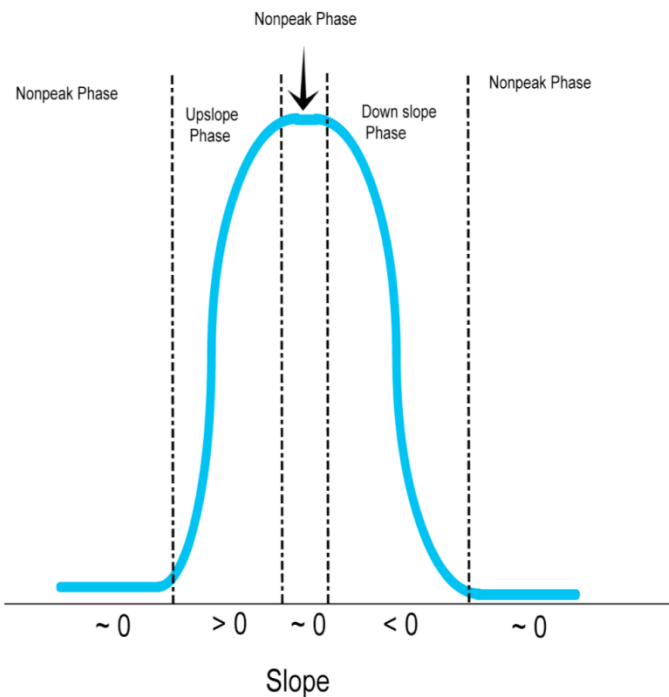
**Axis** – Displays dialog that allows components to be changed along the x and y axes



## Appendix A

### Peak Calling Heuristics

Below is a diagram of a peak, as it is determined by the peak calling algorithm.



### Initial Peak Calling

Accurate identification of peaks is a critical step in ensuring that data is prepared for further analysis. Our peak-calling algorithm applies linear interpolation to separate signals of peaks from that of baseline in raw data from fragment analysis files. The algorithm works by using a configurable parameter set that contains thresholds for values such as slope, and peak heights assigning each data point to one of five possible phases non-peak, peak, up-slope, down-slope or inter-peak. After assigning all the data points to one of the four phases, the peaks can be identified. A peak is recognized as a collection of points that meet the requirement of beginning at an up slope phase and ending at a down slope phase. Taking the difference between the highest and lowest data points in the region containing the peak determines peak heights. If the peak height does not meet the threshold from the parameter set, the region is relabeled as a nonpeak region. Adjusting the parameters allows the user to redefine what constitutes a peak with the resulting peak calls seen in real time. Since any peak-calling algorithm has the potential of missing peaks or miscalling peaks Peak Studio combines automated peak detection with the ability for the user to visually inspect and manually select peaks that need to be adjusted. Through the use of the program, samples that have misidentified peaks can be salvaged by manual user selection of peaks.

## Appendix B

### ***QC Number:***

A quality control method was developed to allow the user to rapidly identify any spectra where the peak-calling algorithm mislabeled peaks. Quality control scores are calculated through a linear interpolation process. We start with a standard spectra and walk through all of the peaks that have been called. Taking a set of 3 peaks at time we use the location of the left and right peak to predict the location of the middle peak. The QC Score is then the sum of the absolute value of the difference between the predicted location and the observed location of the middle peak divided by the number of total peaks called (Equation 1). When the user toggles a peak on or off the QC Score is updated in real time, with lower scores being better. The score represents the overall accuracy of the peak-calling algorithm therefore a smaller score, especially less than 0.5, indicates that the predicted peak was very close to the actual peak that was called. Higher scores are a signal that something is wrong, and the user may have to manually adjust peaks.

$$\text{QC Number} = \frac{\sum | \text{predicted} - \text{observed} |}{\text{number of peaks}} \quad (\text{Equation 1})$$

## Appendix C

### ***Automated Peak Adjusting:***

To increase the user friendliness and the speed at which data can be processed, we implemented a feature that allows the user to automatically adjust peaks in the standard spectra correcting any potentially misclassified data regions. We try to separate background noise from what should be actual peaks by applying filters to the spectral data. Our first filter assumes that because peaks will have larger areas and higher heights than non-peaks they will contribute more to the variation in the distribution of areas under the curve for the spectra. Setting a default threshold of 3 standard deviations (user preference in the parameter set) we can filter out larger peaks from background noise. The second filter is applied if the number of called peaks differs from the number of actual peaks according to the user provided weights. We walk through the current set of peaks and make adjustments by turning peaks on or off trying to minimize the QC score. Our third and final filter is applied if the previous two filters were not successful in correctly adjusting the peaks. This filter uses the current set of features (peaks & non-peaks) and gathers potential peaks by incorporating features that are a default 2 standard deviations away and also requiring that a peak meets a default height threshold of 33% (user preference in the parameter set) the height of the nearest called peak. Post automated adjustment of the standard spectra; the weights should be applied correctly to the called peaks. Understanding that this is a heuristic and will provide the optimal solution but not always the correct solution, the user has the ability to manually adjust any standard spectra that passed the filter steps yet has miscalled peaks. This adjustment procedure is intended to reduce the amount of time a user needs to process data from raw .fsa files to usable data for downstream analysis.

1.	Identify features with SD greater than 3, label them as peaks.
2.	Toggle features on and off, trying to minimize the QC score.
3.	Gather features with SD greater than 2 and whose height meets a threshold of 33% the height of the nearest called peak.
Table 1. Automated Peak Adjusting Filtering Algorithm. SD = standard deviation	

## Acknowledgements

Dr. Michael Thomas Flanagan for making his code publicly available ([www.ee.ucl.ac.uk/~mflanaga](http://www.ee.ucl.ac.uk/~mflanaga)), which we used for smoothing and area calculations in Peak Studio.