Using the Statistical Compare and Fisher Ratio ChromaTOF® Features to Define Variance Prior to Multivariate Analysis in the Small Metabolite Profile of Diabetic Versus Non-Diabetic Urine by GCxGC-TOFMS

John Heim • LECO Corporation; Saint Joseph, Michigan USA

Key Words: Fisher Ratio, Statistical Compare, Multivariate Analysis, Metabolite Profile

1. Introduction

The "Statistical Compare" option available in ChromaTOF® software allows the user to view statistical comparisons as a data processing step for groups of samples. The groups of samples are divided into different subsets or classes. ChromaTOF software aligns the data for the specified group of samples from the data processed peak tables. Upon completion of peak alignment, statistical information (such as minimum, maximum, average, relative standard deviation, etc.) on various peak properties (such as peak height, peak area, retention time, etc.) can be viewed in the *Compound Table* generated by *Statistical Compare*. The software will also compare statistical information from each class and between classes. Additionally, "Fisher Ratios" can be calculated from the Compound Table for each analyte. The Fisher ratio is a statistical calculation that can be used to discover the unknown chemical differences among known classes of complex samples. Statistical Compare results can also be exported as a .csv file and applied to third party software programs for supplemental data reduction such as multivariate analysis.

In this paper, Statistical Compare is utilized to define the small metabolite profile with potentially significant class differences between trimethylsilyl (TMS) derivatized urine from diabetic and non-diabetic subjects analyzed by GCxGC-TOFMS. This research was designed to analyze TMS-derivatized urine samples for the small molecule metabolite profile with the intent to detect possible chemical variations between diabetic diseased state and normal control non-diabetic subjects. It is important to note that this research was conducted solely as a proof of concept study to test the validity of the *Statistical Compare*, *Fisher Ratios*, and .csv file export functions that are available in the ChromaTOF software. The data mining strategy used to distinguish analyte differences between diabetic and non-diabetic sample classes begins by aligning the analytes from the processed data files using the *Statistical Compare* feature. Following the Statistical Compare operation, *Fisher Ratios* are calculated to identify the compounds showing the highest variance. The resulting *Compound Table* is then exported as a .csv file to a third party multivariate analysis software package where PCA and clustering analysis was executed.

2. Experimental Conditions

Morning-fast urine samples were collected from four subjects, two non-diabetic normal controls, one type I diabetic, and one type II diabetic. Samples were stored under refrigeration prior to liquid/liquid extraction with methylene chloride and subsequent derivatization with N,O-bis-(Trimethylsilyl)-trifluoroacetamide (BSTFA). Six 10 mL aliquots from each subject were prepared by acidification with concentrated sulfuric acid to pH 2. The 10 mL aliquots were extracted with 2 mL of methlyene chloride into a 20 mL scintillation vial containing approximately 5 mg sodium sulfate. Derivatization was then carried out with BSTFA by placing 200 µL of extract into a sealed 2 mL auto sampler vial containing approximately 0.5 mg sodium sulfate followed by the addition of 30 µL of dry pyridine and 200 µL BSTFA to each vial. The samples were heated to 60°C for 1 hour and then analyzed by GCxGC-TOFMS.

GCxGC-TOFMS results were generated with a LECO Pegasus® 4D Time-of-Flight Mass Spectrometer (TOFMS). The Pegasus 4D GC-TOFMS instrument was equipped with an Agilent 7890 gas chromatograph featuring a LECO two stage cryogenic modulator and independently temperature controlled secondary oven. LECO ChromaTOF software was used for all acquisition control, data processing, Statistical Compare, and Fisher Ratio calculations. A 30 m x 0.25 mm x 0.25 µm film thickness, Rtx-5ms, (Restek Corp.) GC capillary column was used as the primary column for the GCxGC-TOFMS analysis. In the GCxGC configuration a 1.5 m x 0.18 mm id. x 0.18 µm film thickness, Rtx-200, (Restek Corp.) was placed inside the LECO secondary GC oven which follows the thermal modulator. The Helium carrier gas flow rate was set to 1.5 mL/min at a corrected constant flow via pressure ramps. The primary column was programmed with an initial temperature of 40°C for 1.00 minute and ramped at 6°C/minute to 290°C for 10 minutes. The secondary column temperature program was set to an initial temperature of 50°C for 1.00 minute and then ramped at 6°C/minute to 300°C with a 10 minute hold time. The thermal modulator was set to +25°C relative to the primary oven and a modulation time of 5 seconds was used. The MS mass range was 45-800 m/z with an acquisition rate of 200 spectra per second. The ion source chamber was set to 230°C and the detector voltage was 1750V with an electron energy of -70eV. All of the data was processed with an identical data processing method and signal-to-noise ratio ≥100 prior to conducting Statistical Compare for the diabetic versus non-diabetic sample groups.

Life Science and Chemical Analysis Solutions

3. Results and Discussion

Results of the diabetic versus non-diabetic small molecule metabolite profile study between diabetic and non-diabetic subjects are shown in Figures 1 and 2 below. The total ion chromatograms are depicted as contour plots. These chromatographic examples show visual peak differences between diabetic and non-diabetic sample types. Figures 1 and 2 illustrate the increased peak capacity, improved analyte detectability, and enhanced resolution gained by GCxGC-TOFMS. On average over 1000 peaks were found per sample with a signal-to-noise ratio ≥100 for this study. The red cross hatched area in each contour plot is an unprocessed region developed in the Classifications feature of ChromaTOF software which eliminates unwanted background peaks.

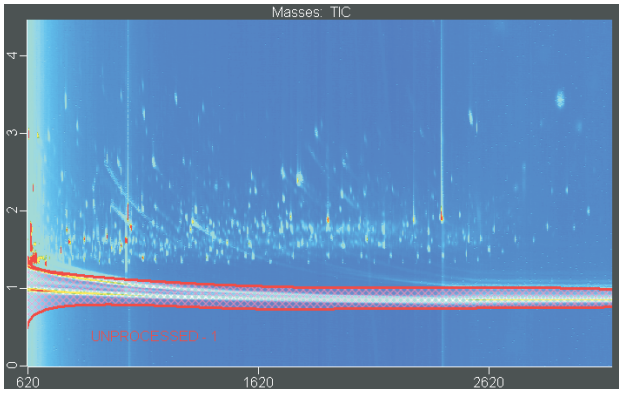


Figure 1. Normal control subject non-diabetic: Contour Plot Total Ion Chromatogram of TMS-derivatized urine sample showing the small molecule metabolite profile for a non-diabetic control sample.

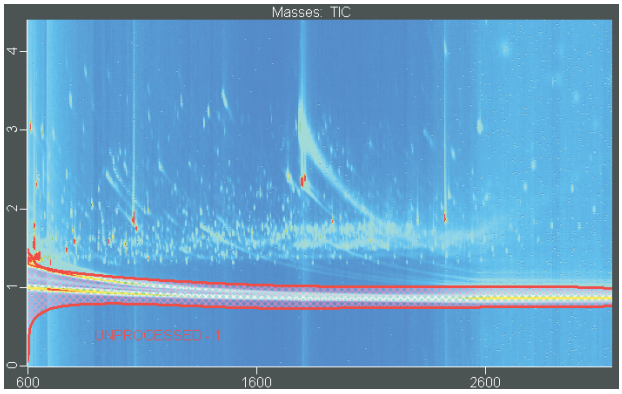


Figure 2. Diseased state subject diabetic: Contour Plot Total Ion chromatogram of TMS-derivatized urine sample showing the small molecule metabolite profile. Notice that peak differences between the non-diabetic control sample and the diabetic sample can be seen by visual observation.

4. Using Statistical Compare in ChromaTOF Software

The following example demonstrates the steps used to generate Statistical Compare data in ChromaTOF software. By following these steps in the Statistical Compare feature of ChromaTOF, the user is able to define the analyte variance between different classes of samples.

STEP 1. CREATE CLASSES

Highlight the Statistical Compare function in the ChromaTOF software database tree. In the database files area, right click and select *New*. Right click on the filename and assign a new name for the Statistical Compare. Next open the *Class Table* window from the Windows menu on the menu bar, click *Statistical Compare* and then open the *Class Table* window.

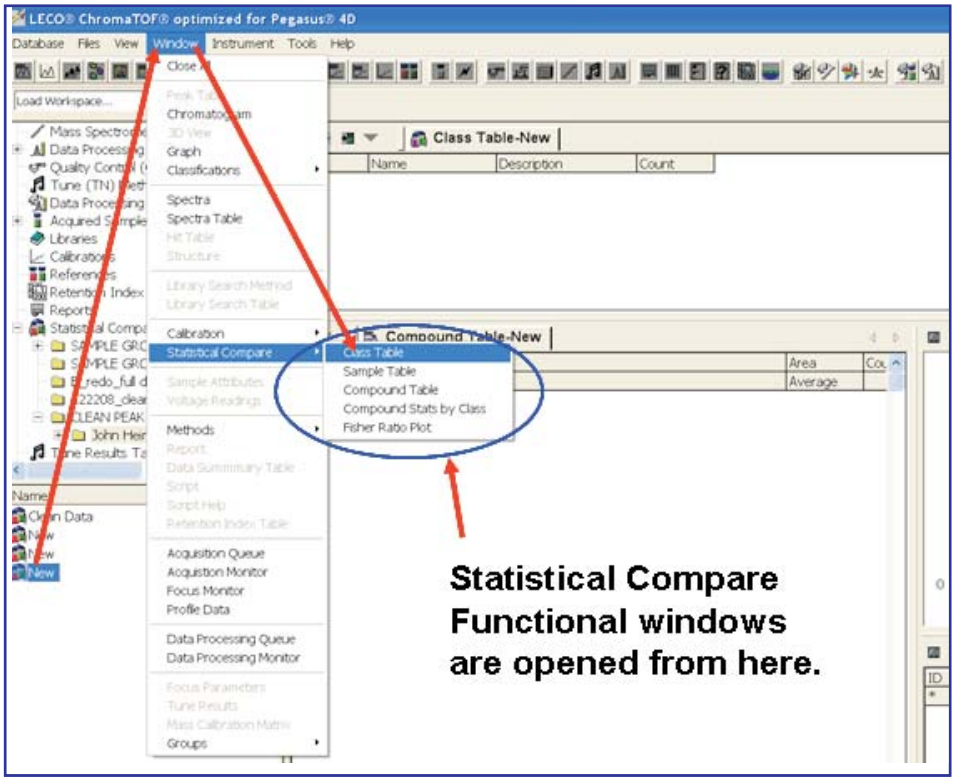


Figure 3. The figure shows how to begin a "new" Statistical Compare operation and open the Class Table from the "Window" menu bar in ChromaTOF.

STEP 2. ADD CLASSES AND SAMPLES TO EACH CLASS

Next Add a *Class* from the *Class Table* by using the (+) button in the *Class Table* toolbar. Highlight each class and add samples by clicking the *Add Samples* button in the *Class Table* toolbar. A database tree of acquired samples will open; place a check mark in the box in front of the sample name which selects the samples to place in a class.

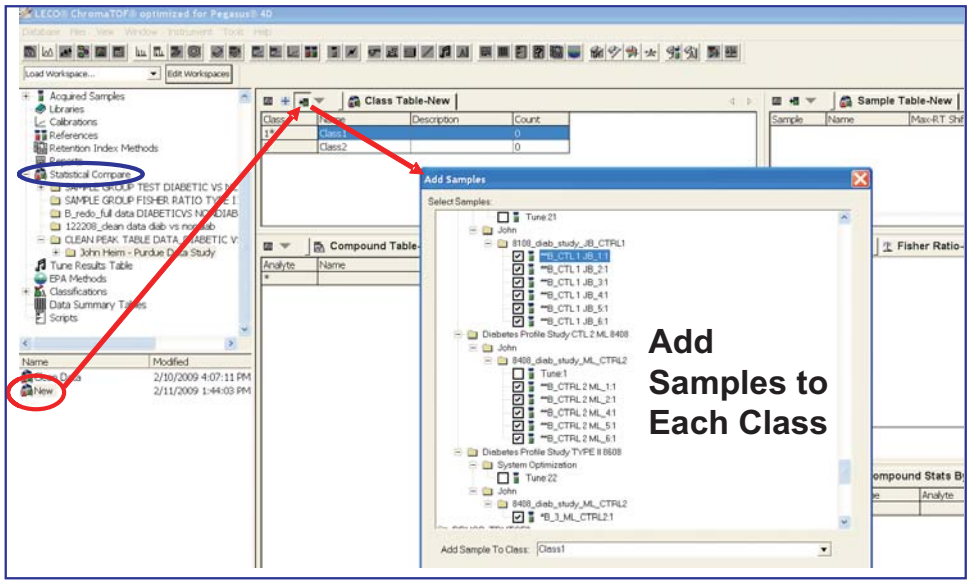


Figure 4. The figure shows the Class table window, the Add samples window, and the Tool bar buttons to use when adding a class and samples to each class.

The completed *Class Table* and *Sample Table* are shown below. The *Sample Table* and *Class Table* are reviewed to make certain that they are correct before continuing the statistical compare process.

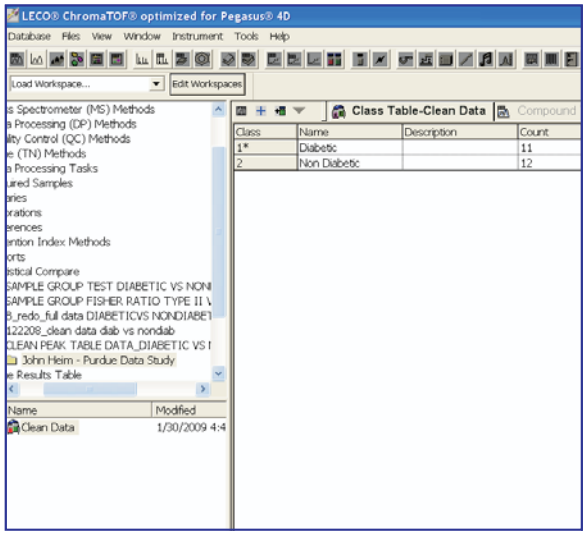
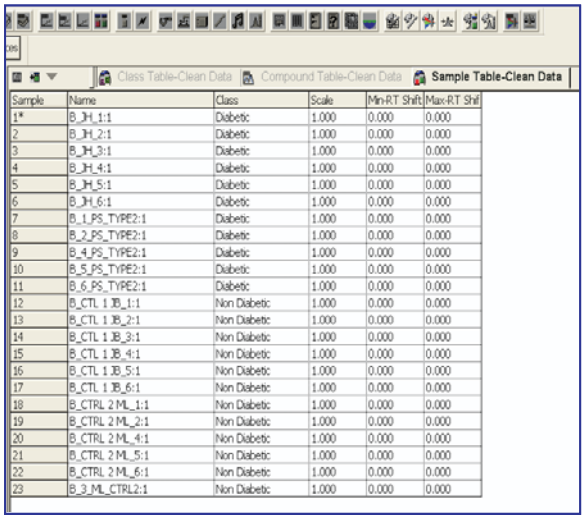


Figure 5. Shown is a completed Class table.



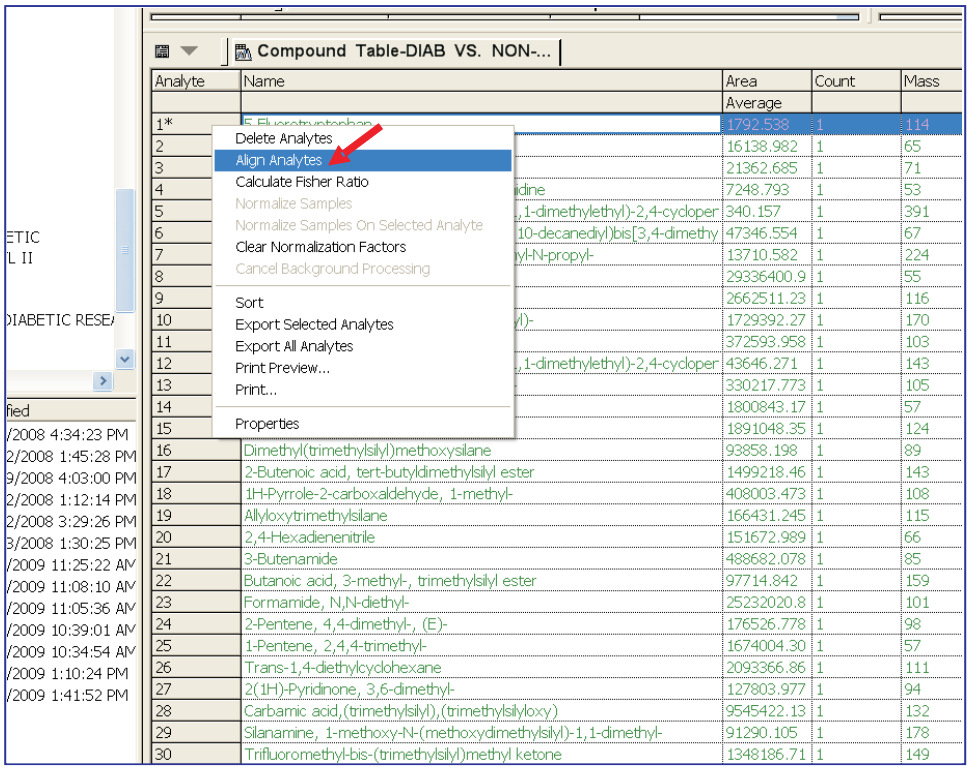
Sample	Name	Class	Scale	Min-RT Shift	Max-RT Shif
1*	B_JH_1:1	Diabetic	1.000	0.000	0.000
2	B_JH_2:1	Diabetic	1.000	0.000	0.000
3	B_JH_3:1	Diabetic	1.000	0.000	0.000
4	B_JH_4:1	Diabetic	1.000	0.000	0.000
5	B_JH_5:1	Diabetic	1.000	0.000	0.000
6	B_JH_6:1	Diabetic	1.000	0.000	0.000
7	B_1_PS_TYPE2:1	Diabetic	1.000	0.000	0.000
8	B_2_PS_TYPE2:1	Diabetic	1.000	0.000	0.000
9	B_4_PS_TYPE2:1	Diabetic	1.000	0.000	0.000
10	B_5_PS_TYPE2:1	Diabetic	1.000	0.000	0.000
11	B_6_PS_TYPE2:1	Diabetic	1.000	0.000	0.000
12	B_CTL 1 JB_1:1	Non Diabetic	1.000	0.000	0.000
13	B_CTL 1 JB_2:1	Non Diabetic	1.000	0.000	0.000
14	B_CTL 1 JB_3:1	Non Diabetic	1.000	0.000	0.000
15	B_CTL 1 JB_4:1	Non Diabetic	1.000	0.000	0.000
16	B_CTL 1 JB_5:1	Non Diabetic	1.000	0.000	0.000
17	B_CTL 1 JB_6:1	Non Diabetic	1.000	0.000	0.000
18	B_CTRL 2 ML_1:1	Non Diabetic	1.000	0.000	0.000
19	B_CTRL 2 ML_2:1	Non Diabetic	1.000	0.000	0.000
20	B_CTRL 2 ML_4:1	Non Diabetic	1.000	0.000	0.000
21	B_CTRL 2 ML_5:1	Non Diabetic	1.000	0.000	0.000
22	B_CTRL 2 ML_6:1	Non Diabetic	1.000	0.000	0.000
23	B_3_ML_CTRL2:1	Non Diabetic	1.000	0.000	0.000

Figure 6. Shown is a completed Sample Table.

Life Science and Chemical Analysis Solutions

STEP 3. ALIGN ANALYTES FROM THE COMPOUND TABLE

From the Windows menu bar select Statistical Compare and open the *Compound Table*. A compound table was generated from all of the sample peak tables. The compound table should appear green. Next right click in the *Compound Table* and select *Align Analytes*. When the align analytes function has finished, the compound table will appear normal (Black). If the compound table appears red, either the compound is undergoing background processing or the Statistical Compare processing has failed. When the *Compound Table* has completed the analyte alignment, the statistics for each compound can be viewed from the *Compound Statistics by Class* table. This table is opened from the Windows menu bar and the drop down menu found under the Statistical Compare option.

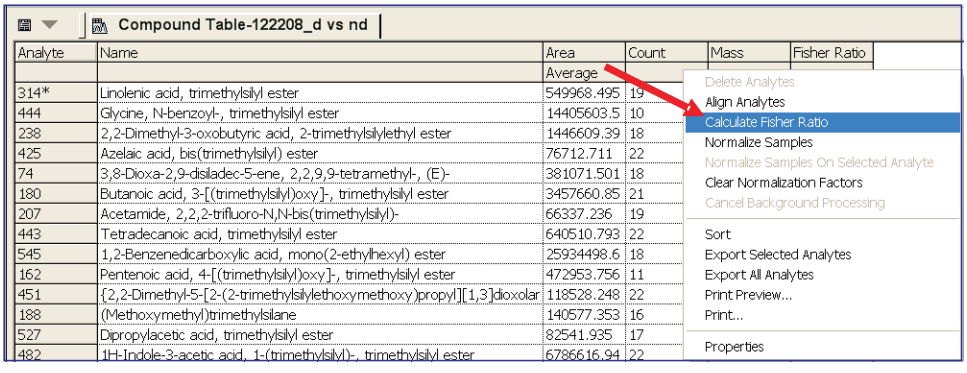


Analyte	Name	Area Average	Count	Mass
1*	5-Fluorotryptophan	1792.538	1	114
2		16138.982	1	65
3		21362.685	1	71
4	idine	7248.793	1	53
5	,1-dimethylethyl)-2,4-cycloper	340.157	1	391
6	10-decanediyl)bis[3,4-dimethy	47346.554	1	67
7	yl-N-propyl-	13710.582	1	224
8		29336400.9	1	55
9		2662511.23	1	116
10	l)-	1729392.27	1	170
11		372593.958	1	103
12	,1-dimethylethyl)-2,4-cycloper	43646.271	1	143
13		330217.773	1	105
14		1800843.17	1	57
15		1891048.35	1	124
16	Dimethyl(trimethylsilyl)methoxysilane	93858.198	1	89
17	2-Butenoic acid, tert-butyldimethylsilyl ester	1499218.46	1	143
18	1H-Pyrrole-2-carboxaldehyde, 1-methyl-	408003.473	1	108
19	Allyloxytrimethylsilane	166431.245	1	115
20	2,4-Hexadienenitrile	151672.989	1	66
21	3-Butenamide	488682.078	1	85
22	Butanoic acid, 3-methyl-, trimethylsilyl ester	97714.842	1	159
23	Formamide, N,N-diethyl-	25232020.8	1	101
24	2-Pentene, 4,4-dimethyl-, (E)-	176526.778	1	98
25	1-Pentene, 2,4,4-trimethyl-	1674004.30	1	57
26	Trans-1,4-diethylcyclohexane	2093366.86	1	111
27	2(1H)-Pyridinone, 3,6-dimethyl-	127803.977	1	94
28	Carbamic acid,(trimethylsilyl),(trimethylsilyloxy)	9545422.13	1	132
29	Silanamine, 1-methoxy-N-(methoxydimethylsilyl)-1,1-dimethyl-	91290.105	1	178
30	Trifluoromethyl-bis-(trimethylsilyl)methyl ketone	1348186.71	1	149

Figure 7. A partial Compound Table is illustrated above for the Statistical Compare of the diabetic versus non-diabetic urine GCxGC-TOFMS study.

STEP 4. CALCULATE FISHER RATIO

A Fisher Ratio calculation can be calculated as an additional feature to Statistical Compare after the analyte alignment is completed. The Fisher Ratio method can be used to derive the unknown chemical differences among known classes of complex samples. The Fisher Ratio is calculated by the difference of the analyte means from different classes divided by the difference of the analyte variance between different classes. The numerical value of the Fisher Ratio is related to the degree of variance by the size of the number. The higher the Fisher Ratio numerical value, the greater the class variance is for a particular compound. To calculate the Fisher Ratio from the Compound Table, right click in the *Compound Table* and select *Calculate Fisher Ratio* from the drop down menu. For further explanation of Fisher ratios, refer to Pierce, K. M., Hoggard, J. C., Hope, J. L., Rainey, P. M., Hoofnagle, A. N., Jack, R. M., Wright, B. W., and Synovec, R. E.; "Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts." *Anal. Chem.*, **2006**, 78 (14), 5068 - 5075.



Analyte	Name	Area Average	Count	Mass	Fisher Ratio
314*	Linolenic acid, trimethylsilyl ester	549968.495	19		
444	Glycine, N-benzoyl-, trimethylsilyl ester	14405603.5	10		
238	2,2-Dimethyl-3-oxobutyric acid, 2-trimethylsilylethyl ester	1446609.39	18		
425	Azelaic acid, bis(trimethylsilyl) ester	76712.711	22		
74	3,8-Dioxa-2,9-disiladec-5-ene, 2,2,9,9-tetramethyl-, (E)-	381071.501	18		
180	Butanoic acid, 3-[(trimethylsilyl)oxy]-, trimethylsilyl ester	3457660.85	21		
207	Acetamide, 2,2,2-trifluoro-N,N-bis(trimethylsilyl)-	66337.236	19		
443	Tetradecanoic acid, trimethylsilyl ester	640510.793	22		
545	1,2-Benzenedicarboxylic acid, mono(2-ethylhexyl) ester	25934498.6	18		
162	Pentenoic acid, 4-[(trimethylsilyl)oxy]-, trimethylsilyl ester	472953.756	11		
451	{2,2-Dimethyl-5-[2-(2-trimethylsilylethoxymethoxy)propyl][1,3]dioxolar	118528.248	22		
188	(Methoxymethyl)trimethylsilane	140577.353	16		
527	Dipropylacetic acid, trimethylsilyl ester	82541.935	17		
482	1H-Indole-3-acetic acid, 1-(trimethylsilyl)-, trimethylsilyl ester	6786616.94	22		

Figure 8. Shown above is a completed Compound Table illustrating the drop down menu. The red arrow indicates the Calculate Fisher Ratio option.

Completed Compound Table with Fisher Ratios

The figure below shows the completed Compound Table with Fisher Ratios calculated and sorted from highest to lowest Fisher Ratio. The table reflects the columns selected in the table properties for Area average, Count of the samples in which the analyte was found, the unique Mass for each analyte, and the Fisher Ratio for each analyte.

Compound Table-122208_d vs nd

Analyte	Name	Area	Count	Mass	Fisher Ratio
		Average			
314	Linolenic acid, trimethylsilyl ester	549968.495	19	122	22176
444	Glycine, N-benzoyl-, trimethylsilyl ester	14405603.5	10	206	5449.9
238	2,2-Dimethyl-3-oxobutyric acid, 2-trimethylsilylethyl ester	1446609.39	18	160	4992.8
425	Azelaic acid, bis(trimethylsilyl) ester	76712.711	22	317	4706.2
74	3,8-Dioxa-2,9-disiladec-5-ene, 2,2,9,9-tetramethyl-, (E)-	381071.501	18	129	4508.3
180	Butanoic acid, 3-[(trimethylsilyl)oxy]-, trimethylsilyl ester	3457660.85	21	147	4207.0
207	Acetamide, 2,2,2-trifluoro-N,N-bis(trimethylsilyl)-	66337.236	19	198	4051.1
443	Tetradecanoic acid, trimethylsilyl ester	640510.793	22	117	3903.6
545	1,2-Benzenedicarboxylic acid, mono(2-ethylhexyl) ester	25934498.6	18	149	3844.9
162	Pentenoic acid, 4-[(trimethylsilyl)oxy]-, trimethylsilyl ester	472953.756	11	143	3679.7
451	{2,2-Dimethyl-5-[2-(2-trimethylsilylethoxymethoxy)propyl][1,3]dioxolar	118528.248	22	131	3613.6

Figure 9. Shown above is a partial completed Compound Table showing the calculated Fisher Ratio for each analyte calculated as numerical value. A large Fisher Ratio value is related as having greater variance than a smaller Fisher Ratio value.

View The Fisher Ratio Plot

The *Fisher Ratio Plot* shows a graphical representation of the Fisher Ratios for the analytes selected from the *Compound Table*. The *Fisher Ratio plot* for each compound is displayed as an intensity line. Compounds with the highest Fisher Ratios and variation are shown graphically as the largest intensity values in the plot. The vertical y-axis shows the intensity scale of the values calculated for Fisher Ratios. The horizontal x-axis is the retention time scale in seconds. Each analyte is displayed at the retention time of elution. The number placed at the top of the Fisher Ratio line is the analyte identification number from the *Compound Table*. The red vertical line indicates the highest Fisher Ratio calculated for Analyte 314 from the *Compound Table* for the diabetic versus non-diabetic urine GCxGC-TOFMS study. Analyte 314 showing the highest variance between the diabetic and non-diabetic classes was identified by NIST library search as Linolenic acid, trimethylsilyl ester.

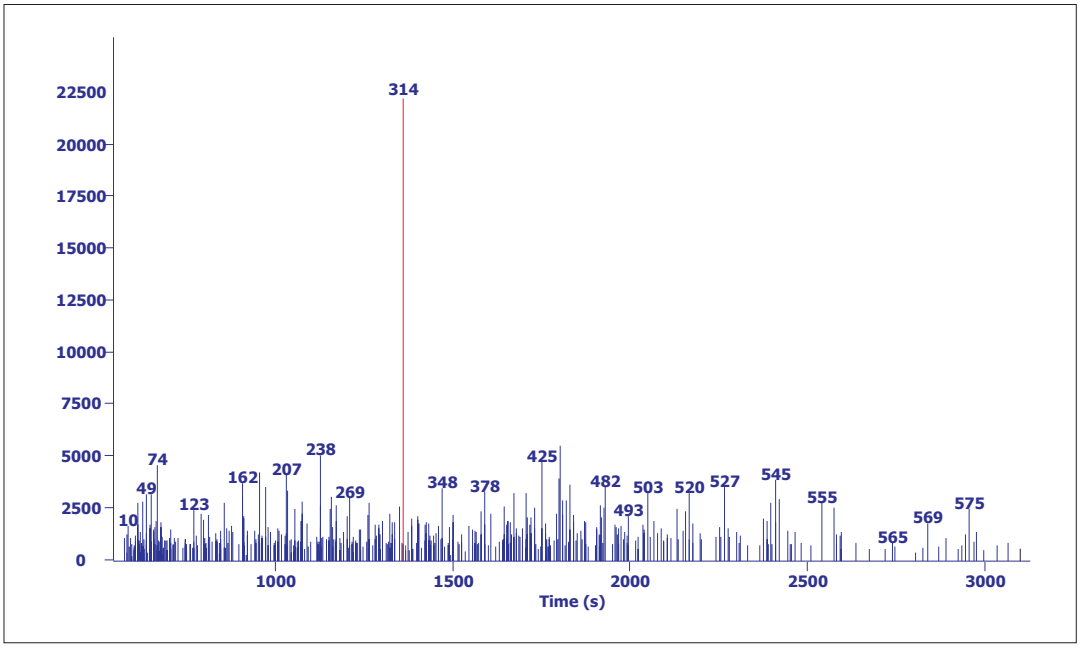


Figure 10. The Fisher Ratio plot shown above graphically represents unknown chemical differences between the normal control non-diabetic sample group and the diabetic diseased state sample group.

LECO®
Life Science and Chemical Analysis Solutions

Display Compound Statistics By Class

Additional statistical information can be viewed by opening the *Compound Stats by Class* option from the drop down menu of the *Statistical Compare* tab in the Windows menu bar. The example below shows the Compound statistics for Analyte 314, Linolenic acid, trimethylsilyl ester which also had the highest calculated variance and Fisher Ratio. Different statistical calculations can be set in the properties as columns and consequently viewed in the compound statistics table consequently providing additional information and insight into the unknown chemical differences between different sample classes. The table below shows the signal-to-noise ratios for Analyte 314 indicating that Linolenic acid trimethylsilyl ester has approximately 20 times higher signal-to-noise response in Class 1 (diabetic) than in Class 2 (non-diabetic). This example illustrates that potentially significant data results can be detected by utilization of the *Compound Stats by Class* feature in Statistical Compare.

Compound Stats By Class-1222...

ID	Name	S/N	Count
		Average	
314-1	Class1	4195.407	10
314-1-1	B_JH_2:1	5417.957	
314-1-2	B_JH_3:1	4750.749	
314-1-3	B_JH_4:1	5115.776	
314-1-4	B_JH_5:1	3751.567	
314-1-5	B_JH_6:1	4795.219	
314-1-6	B_1_PS_TYPE2:1	3933.214	
314-1-7	B_2_PS_TYPE2:1	3203.246	
314-1-8	B_4_PS_TYPE2:1	3104.403	
314-1-9	B_5_PS_TYPE2:1	3791.019	
314-1-10	B_6_PS_TYPE2:1	4090.920	
314-2	Class2	211.541	9
314-2-11	B_CTL 1 JB_1:1	149.858	
314-2-12	B_CTL 1 JB_2:1	122.867	
314-2-13	B_CTL 1 JB_3:1		
314-2-14	B_CTL 1 JB_4:1	809.461	
314-2-15	B_CTL 1 JB_5:1		
314-2-16	B_CTL 1 JB_6:1		
314-2-17	B_CTRL 2 ML_1:1	105.032	
314-2-18	B_CTRL 2 ML_2:1	129.824	
314-2-19	B_CTRL 2 ML_4:1	120.896	
314-2-20	B_CTRL 2 ML_5:1	222.299	
314-2-21	B_CTRL 2 ML_6:1	127.434	
314-2-22	B_3_ML_CTRL2:1	116.197	
314	Total	2308.313	19

Figure 11. The Compound Statistics table by class can be generated and viewed in a separate window of ChromaTOF. Specific parameters can be selected from the Properties tab, such as statistics for retention time, peak height, concentration, etc. The example above shows the diabetic (Class 1) and the non-diabetic (Class 2) results showing the average signal-to-noise ratio for each sample. The results indicate which samples that Analyte 314 was found in and the numerical signal-to-noise value. The Name/Class column shows the sample name for each analysis and the Count column lists in how many samples the analyte was found for a particular class.

STEP 5. EXPORTING RAW DATA FROM THE COMPOUND TABLE AS A (.csv) FILE

The raw data for the analytes of interest can be exported as a .csv file. All of the wanted fields of interest need to be displayed in the *Compound Table*. The software will only include the visible fields in the export. It is possible to export all or selected analytes from the compound table. To export a .csv file, right click anywhere inside the *Compound Table* grid and select *Export Selected Analytes* from the drop down menu. Only the highlighted analytes selected will be exported in this case. Alternatively, right click inside the grid and select *Export All Analytes* if that is the desired option.

Compound Table-122208_d vs nd

Analyte	Name	Area	Count	Mass	Fisher Ratio
		Average			
314*	Linolenic acid, trimethylsilyl ester	549968.495	19	122	22176
444	Glycine, N-benzoyl-, trimethylsilyl ester				5449.9
238	2,2-Dimethyl-3-oxobutyric acid, 2-trimethylsilylethyl ester				4992.8
425	Azelaic acid, bis(trimethylsilyl) ester				4706.2
74	3,8-Dioxa-2,9-disiladec-5-ene, 2,2,9,9-tetramethyl-, (E)-				4508.3
180	Butanoic acid, 3-[(trimethylsilyl)oxy]-, trimethylsilyl ester				4207.0
207	Acetamide, 2,2,2-trifluoro-N,N-bis(trimethylsilyl)-				4051.1
443	Tetradecanoic acid, trimethylsilyl ester				3903.6
545	1,2-Benzenedicarboxylic acid, mono(2-ethylhexyl) ester				3844.9
162	Pentenoic acid, 4-[(trimethylsilyl)oxy]-, trimethylsilyl ester				3679.7
451	[2,2-Dimethyl-5-[2-(2-trimethylsilylethoxymethoxy)propyl][1,3]dio				3613.6
188	(Methoxymethyl)trimethylsilane				3500.2
527	Dipropylacetic acid, trimethylsilyl ester				3469.0
482	1H-Indole-3-acetic acid, 1-(trimethylsilyl)-, trimethylsilyl ester				3462.3
348	Benzoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, ethyl ester				3432.8

Delete Analytes
Align Analytes
Calculate Fisher Ratio
Normalize Samples
Normalize Samples On Selected Analyte
Clear Normalization Factors
Cancel Background Processing
Sort
Export Selected Analytes
Export All Analytes
Print Preview...
Print...
Properties

Figure 12. The Compound Table above illustrates how to export the Statistical Compare results from the compound Table to a .csv file of the raw data.

Example of the Exported .csv File Displayed in Excel

The partial .csv file of the diabetic versus non-diabetic small molecule metabolite profile is shown below. The Excel spreadsheet displays the analyte Peak Name, Sample Name, Class, and Peak Area as columns for every sample and class from the Statistical Compare analysis. The columns shown are a result of what statistical information was set to be displayed in the *Compound Table*. This particular .csv file was designed to be used in the commercially available peripheral multivariate analysis program, Miner3D.

Microsoft Excel - TEST 2 diab vs non13009.csv

File Edit View Insert Format Tools Data Window Help Adobe PDF

A42 13,13-Dimethyl-3,6,9-trioxa-13-silatetradecan-1-ol

	A	B	C	D	E	F	G	H	I
1		B_JH_1:1		B_JH_2:1		B_JH_3:1		B_JH_4:1	
2	Peak	Class	Area	Class	Area	Class	Area	Class	Area
3	Linolenic acid, trimethylsilyl ester	Diabetic	406541.1	Diabetic	1020652	Diabetic	1081791	Diabetic	952796.2
4	3,8-Dioxa-2,9-disiladec-5-ene, 2,2,9,9-tetramethyl-, (E)-	Diabetic	990014.8	Diabetic	1142043			Diabetic	764.895
5	Propanoic acid, 2-[(trimethylsilyl)oxy]-, trimethylsilyl ester					Diabetic	51874.35		
6	2,2-Dimethyl-3-oxobutyric acid, 2-trimethylsilylethyl ester							Diabetic	174642.9
7	Glycine, N-benzoyl-, trimethylsilyl ester			Diabetic	24819.5				
8	Azelaic acid, bis(trimethylsilyl) ester	Diabetic	57557.59	Diabetic	55082.65	Diabetic	53988.82	Diabetic	52485.76
9	Acetamide, 2,2,2-trifluoro-N,N-bis(trimethylsilyl)-	Diabetic	11498.9	Diabetic	13772.69	Diabetic	27717.77	Diabetic	20245.64
10	1,2-Benzenedicarboxylic acid, mono(2-ethylhexyl) ester	Diabetic	19126549	Diabetic	18519912	Diabetic	17554930	Diabetic	24612411
11	Pentenoic acid, 4-[(trimethylsilyl)oxy]-, trimethylsilyl ester	Diabetic	74588.91	Diabetic	72996.18	Diabetic	83847.46		
12	{2,2-Dimethyl-5-[2-(2-trimethylsilylethoxymethoxy)propyl][1,3	Diabetic	14279.52	Diabetic	24021	Diabetic	29858.31	Diabetic	31505.82
13	1H-Indole-3-acetic acid, 1-(trimethylsilyl)-, trimethylsilyl este	Diabetic	1379351	Diabetic	1455209	Diabetic	1275428	Diabetic	1278597
14	(Methoxymethyl)trimethylsilane								
15	Dipropylacetic acid, trimethylsilyl ester								
16	Benzoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, ethyl es	Diabetic	256269.8	Diabetic	300549.3	Diabetic	253671.6	Diabetic	298356.2
17	Arachidonic acid, trimethylsilyl ester					Diabetic	3431.35		
18	Butanoic acid, 3-[(trimethylsilyl)oxy]-, trimethylsilyl ester	Diabetic	230726.9			Diabetic	235220.1	Diabetic	436996.7
19	1,2-Benzenedicarboxylic acid, ethyl trimethylsilyl ester	Diabetic	26892.63	Diabetic	27822.48	Diabetic	25672.19	Diabetic	25118.88
20	7,10-Epoxytricyclo[4.2.1.1(2,5)]decane, 1-trimethylsilyl-			Diabetic	3670.191	Diabetic	10697.73	Diabetic	4531.661
21	Benzenepropanoic acid, 3-[(trimethylsilyl)oxy]-, trimethylsilyl ester								

Figure 13. Illustrated above is an exported .csv file partial Excel spreadsheet of the Statistical Compare results for the diabetic and non-diabetic urine GCxGC-TOFMS study. Subsequently, the compounds of highest variance by their Fisher Ratios were then loaded into multivariate analysis programs. The columns displayed in the figure above show the peak compound name, the sample name with the sample Class designation, and the peak area for each analyte.

STEP 6. MULTIVARIATE ANALYSIS DEVELOPED FROM STATISTICAL COMPARE RESULTS EXPORTED AS A .CSV FILE

Multivariate analysis is based on multivariate statistics which involves observation and analysis of more than one statistical variable at a time. The technique is used to perform studies across multiple dimensions while taking into account the effects of all variables on the responses of interest. This study applied the ChromaTOF features *Statistical Compare* and *Fisher Ratios* to a data set of twenty-three samples from diseased and non-diseased state subjects that determined the analytes with the highest variation across the sample population from two distinct classes. The Statistical Compare results generated a *Compound Table* sorted by variance using the *Fisher Ratio* method which was then exported as a .csv in Excel format and applied to several multivariate analysis platforms.

The exported .csv file was subsequently loaded as a spreadsheet containing the top 430 analytes according to their Fisher Ratios into the third party multivariate analysis software package Miner3D. The data was submitted to an Eigenvector analysis before an optimized PCA Plot was calculated. PCA analysis was conducted on the variables of analyte identification, class, (diseased or non-diseased), and analyte peak area. Following PCA analysis, K-means clustering was applied using the Miner3D software. The figure below shows the Miner3D workspace and the loadings for the PCA vectors.

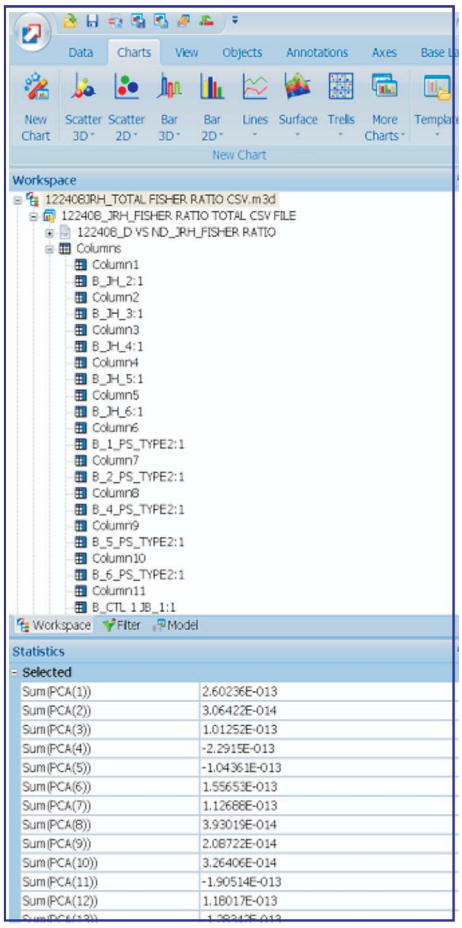


Figure 14. Shows the Miner3D workspace with the .csv file data loaded and the multivariate analysis statistics from the results of the ChromaTOF Statistical Compare analysis.

LECO®
Life Science and Chemical Analysis Solutions

The three-dimensional graph shown below in Figure 15 shows the Statistical Compare data results from the small molecule metabolite profile study of the diabetic versus non-diabetic GCxGC-TOFMS analysis. The results were developed in the commercially available Miner3D software using PCA and K-means clustering analysis. The graph shows clear differences as well as similarities in the small molecule metabolites found in both diabetic and non-diabetic TMS derivatized urine analyzed by GCxGC-TOFMS.

Miner3D PCA Plot With K-means Clustering Showing Analyte and Class Variation Between the Diabetic Versus Non-Diabetic Small Metabolite Profile

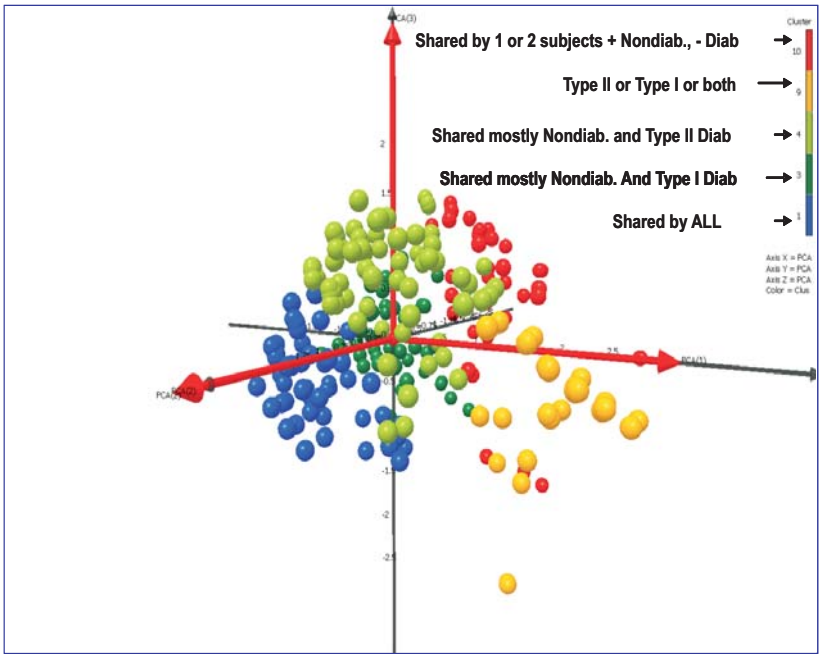


Figure 15. The results from Statistical Compare exported as a .csv file with PCA and Clustering analysis are shown in the figure above. Each colored sphere in the graph represents a different metabolite compound. The color chart on the top right indicates that specific groups of analytes can be attributed to either diabetic, non-diabetic, or both classes. This graphical representation illustrates the value of the Statistical Compare feature of ChromaTOF software to facilitate the data mining process in finding potentially significant biomarkers.

5. Conclusions

This paper presents a step-by-step approach using diabetic versus non-diabetic GCxGC-TOFMS metabolomic data to instruct the user on how to perform *Statistical Compare* as well as export useful result tables to peripheral multivariate software that will provide additional meaningful data reduction capabilities. A comprehensive GCxGC-TOFMS analysis accompanied by statistical comparison targeting high variance data through *Fisher Ratios* along with multivariate PCA and Clustering analysis was demonstrated. This exploratory research presents an optimized GCxGC-TOFMS analysis followed by a data mining strategy using preliminary statistical methods prior to multivariate analysis that establishes a viable strategy which can identify significant metabolite variation in complex biological samples from distinct classes.

The results presented from this study demonstrate that significantly increased analytical performance is achieved by utilizing comprehensive multidimensional gas chromatography coupled with time-of-flight mass spectrometry (GCxGC-TOFMS) for the characterization of small molecule metabolite profiles. TOFMS provides the non-skewed mass spectra and fast acquisition needed to deconvolute complex overlapping peaks as well as the data density required to characterize the narrow peaks (<100 ms) GCxGC. Several new ChromaTOF software features such as Statistical Compare and Fisher Ratio calculations were applied in this metabolomic study. These features allow the user to find significant unknown chemical differences among known classes of complex samples. The new features available in LECO's ChromaTOF software were used to align a large set of data and define the highest variance for analytes between diseased and non-diseased state subjects. Furthermore, it was demonstrated that the results from Statistical Compare and Fisher Ratio calculations can be exported quite simply into multivariate analysis programs whereby PCA and Clustering analysis can be applied.

LECO Corporation • 3000 Lakeview Avenue • St. Joseph, MI 49085 • Phone: 800-292-6141 • Fax: 269-982-8977
info@leco.com • www.leco.com • ISO-9001:2000 • No. FM 24045 • *LECO is a registered trademark of LECO Corporation.*

Form No. 203-821-372 6/09-REV0 © 2009 LECO Corporation