Pattern Recognition for Discrimination of Dyslipidemic States

Gerard G. Dumancas, Mary Muriuki, A. David Marais, Neil Purdie, and Lisa Reilly

Abstract—**The Centers for Disease Control and Prevention reported that the diseases of the heart, cancer, stroke, Alzheimer's disease, and diabetes are among the top leading causes of deaths in the US for 2007. These diseases are known to be caused by a variety of factors including cholesterol and polyunsaturated fatty acids (PUFAs). This paper shows how pattern recognition using an unsupervised clustering algorithm (principal component analysis (PCA)) can provide a direct method of discriminating dyslipidemic patients according to Fredrickson's Classification of Dyslipidemias obtained from the spectral data. The spectral data were obtained by utilizing a mature, patented reagent system based on its selectivity to the -CH=CH-CH_2- group in a wide variety of lipids in human serum samples. The simple colorimetric assay used is rapid, rugged, and inexpensive that produces a characteristic molar absorbance spectra for cholesterol, ω-3 (methyl esters of linolenic, eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA) fatty acids), and ω-6 (methyl esters of linoleic, conjugated linoleic acid (CLA), and arachidonic fatty acids). The assay is reacted with serum samples of known dyslipidemias. An independent full factorial design simulation of synthetic serum mixtures containing different levels of the lipids in chloroform solutions was also done. Pattern recognition using an unsupervised clustering algorithm (PCA) was applied to both the spectral data of synthetic and actual serum samples. The full factorial design of synthetic mixtures of human serum in chloroform solutions accomplished the discrimination of eight clusters. Each cluster corresponded to specific levels of lipids prepared. Application of the assay to real serum samples revealed ten clusters with each corresponding to a disease state according to Fredrickson's Classification of Dyslipidemias. The results demonstrated that the novel chemical approach could provide a good agreement between clinical chemistry and pattern recognition results. This study shows how pattern recognition coupled with the assay can provide discrimination of dyslipidemic states in dyslipidemic individuals in faster and cheaper ways.**

Index Terms- **dyslipidemia, Fredrickson's, hierarchical clustering, pattern recognition, principal component analysis**

Manuscript received January 25, 2011; revised January 29, 2011. This work was supported in part by the Oklahoma State University (OSU) Technology Business Assessment Group (TBAG).

Gerard G. Dumancas is with the OSU Chemistry Department, 002 Physical Science Building, Stillwater, OK 74078, USA (Phone: 405-744-5948; fax: 405-744-6007; email: gerard.dumancas@okstate.edu).

Mary Muriuki is with the OSU Chemistry Department, 002 Physical Science Building, Stillwater, OK 74078, USA (email: marywm@okstate.edu).

A. David Marais is with the Groote Schuur Heart Hospital and Clinic, University of Cape Town, 7925, South Africa (email: david.marais@uct.ac.za).

Neil Purdie is with the OSU Chemistry Department, 018 Physical Science Building, Stillwater, OK 74078, USA (email: neil.purdie@okstate.edu)

Lisa Reilly is with the Department of Physical Sciences, Bethany College, Bethany, WV 26032, USA (email: LReilly@bethanywv.edu)

I. INTRODUCTION

Dyslipidemias are serious and costly health problems worldwide known to be major risk factors for cardiovascular diseases, diabetes mellitus, and atherosclerosis [1], [2]. Prevention in the form of early detection of dyslipidemia and risk modification via drug and non-drug procedures, particularly among the high-risk group, is, therefore imperative.

Hyperlipidemia is the most common form of dyslipidemia, which is characterized by abnormally elevated levels of any or all lipids (cholesterol or triglyceride levels or both) and/or lipoproteins in the blood. It also includes any decreased lipid levels. Hyperlipoproteinemia is a specific form of hyperlipidemia characterized by abnormally elevated concentrations of specific lipoproteins in the plasma.

Cholesterol and triglyceride analyses are the simplest means for detecting hyperlipoproteinemia. They also provide some information about the type of hyperlipoproteinemia because the proportion of these lipids varies from one lipoprotein family to another. Knowledge of the concentrations of cholesterol and triglycerides permits the distinction of three general types of hyperlipidemia that roughly corresponds to certain types of hyperlipoproteinemias. Frederickson, et al developed a classification scheme to correlate hyperlipidemias with the abnormal lipoprotein serum patterns (Table I) [3].

The methods of diagnosis for the abovementioned disease states usually involve electrophoretic methods, ultracentrifugation, and enzymatic tests [3]-[6]. For quantitative determination of total cholesterol in blood serum using the enzymatic test, several factors affect the analysis. These include the specificity of the enzymes used (cholesterol esterase and cholesterol oxidase), conditions of analysis and composition of the reaction mixture, and interferences from components of blood serum [5]. Other methods of determining cholesterol and triglyceride concentrations have included capillary gas chromatography [6]. Chromatographic methods, however, require extended amounts of time to carry out the analyses and are not ideal in typical clinical settings.

The original objective of this study was to introduce a mature, patented reagent system selective to the -CH=CH-CH_2- group that might provide a rapid diagnostic test for screening dyslipidemic individuals.

The assay was determined to simultaneously quantitate cholesterol and six polyunsaturated fatty acids (PUFAs) in synthetic mixtures and in human serum without any need for analytical separation [7]. The assay reagent produced a characteristic molar absorbance spectrum specific for each lipid analyte according to their degree of unsaturation. The method was shown to simultaneously quantitate cholesterol and PUFAs in human serum [8], [9].

ISBN: 978-988-19251-4-5 (revised on 20 September 2011)
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

In this study, pattern recognition diagrams are included to discriminate among different dyslipidemic conditions of patients from selected samples with predetermined historical backgrounds. With additional specimens, correlations between assays and patients can be confirmed with each cluster represented as a biomarker for that disease.

The goal of pattern recognition is to recognize an obscure property in a collection of objects (i.e. samples) from indirect measurements (i.e. spectra) made on the objects [10].

An unsupervised learning algorithm by principal component analysis (PCA) is employed in the pattern recognition used in this study. In PCA, complicated relationships are reduced to simple ones by projecting the data (i.e. spectra) from multidimensional space to two or three dimensions [11]. The first two principal components are then plotted against each other and hierarchical clustering analysis was performed to group together data points.

II. MATERIALS AND METHODS

Human Serum Samples

Serum samples were collected from patients who were admitted to Cape Town Clinic and Lipidology Laboratory at the University of Cape Town (UCT), South Africa. These patients were identified and treated for their respective lipid disorders. The anonymous samples from UCT were from volunteers who had already requested a lipid profile and had given consent. No attempt was made to solicit samples nor was any extensive information derived from the medical records. Subjects fasted for at least 12 hours prior to the collection of the sample. A venous blood sample was collected into a Vacutainer™ red and grey capped separation tube. After inversion of the tube five times to mix the blood and the components of the collection tube, the sample was centrifuged at 3400 revolutions per minute (RPM) for 15 minutes. The collection tube contained a clotting activator which takes approximately 30 minutes to activate and a floating gel that separates the red blood cells from the serum during the centrifugation step. The serum, which was the top layer in the tube, was then transferred to a 10 mL glass vial with a screw cap. The experimental assay was completed within three days of receiving the sample. Samples were stored in a refrigerator at 2-4°C and were allowed to return to room temperature prior to analyses. UCT samples were also drawn from patients with normal to elevated cholesterol levels. For serum sample analysis, a 10 µL sample of serum was added to a 13 x 100 mm borosilicate disposable test tube to which 1.0 mL of 98 % acetyl chloride (AC) (Acros) was added, and then shaken. A 40 µL aliquot of perchloric acid (PA) (70% American Chemical Society (ACS) reagent grade, GFS) was then carefully added down the inside of the test tube and slowly introduced to the acetyl chloride, sample mixture.

The reaction starts on the first contact with PA. The solution was shaken by hand for twenty seconds to allow for the release of the small amount of hydrochloric acid (HCl) gas from the reaction test tube. The test tube was covered with a Teflon cap and placed into a centrifuge and spun for 3 minutes at 3400 RPM. After centrifugation, precipitated proteins were separated as a small plug not removed, and the reagent solution was transferred to a 10.0 mm pathlength optical glass cuvette that was fitted with a Teflon stopper for the remaining time. Absorbance spectra were measured after 15 minutes on an HP8452A Spectrophotometer. A 5.0 second integration time and 2 nm spectral resolution were used to measure the absorbance data over the range of 350-550 nm. The blank for each reaction was pure AC. The reagent mixture of AC with PA did produce a slight color at 15 minutes. Due to the possibility of variability and small absorbance value, AC was substituted as the blank.

Synthetic Mixtures

Methyl esters of ω-6 fatty acids (linoleic, conjugated linoleic acid (CLA), and arachidonic), ω-3 fatty acids (α-linolenic, eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA)) and free cholesterol in chloroform solutions were all used to prepare synthetic mixtures of serum. The mixtures were prepared by full factorial design (n=128) using the *SAS-JMP* Software Package [12]. All of the standards were 90 to 99 % pure based on gas chromatographic analysis and were all purchased from Sigma-Aldrich. Stock solutions for each of the analytes with maximum total concentrations of 0.02 M and 0.04 M were prepared. The stock solutions were used to prepare mixtures to limit the maximum spectral response to ranges between 0.2 and 1.2 absorbance units. The inclusion of water was taken into account in this study. Serum normally consists of 97 % water [13]. With the sample size of serum being 10 µL, approximately 9.7 µL of water was added to the reagents in cases where synthetic mixtures are analyzed. The final experimental assay involved the addition of 10 µL of distilled water as the first step, followed by 1.0 mL AC, 10 µL chloroform mixture sample, and finally 40 µL PA. The final steps of the assay remained the same as serum in order to maintain constancy during the 15-minute reaction period.

Pattern Recognition

The pattern recognition technique used in this study, PCA, is an unsupervised multivariate statistical method useful for reducing multidimensional data down to 2 or 3 dimensions that can readily be comprehended. The graphical representations presented utilize the first 2 or 3 principal components as the axes. Using PCA, the resulting principal components were plotted versus each other to produce 2- and 3- dimensional representations of the data to determine if any clustering patterns were separable. PCA was used for both the prepared mixtures and serum samples. If a pattern was seen, then hierarchical cluster analysis was used to group together the data points using the *JMP* software package [12]. In cluster analysis, the process will start with one piece of data and combines groups based on distances from one another in the principal component space [11]. The cluster analysis in this study was agglomerative hierarchical with Ward's method being used for the distances.

ISBN: 978-988-19251-4-5 (revised on 20 September 2011) WCE 2011
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

III. RESULTS AND DISCUSSIONS

PCA was first performed in spectral data (128 samples x 101 variables matrix) obtained by full factorial design of synthetic serum mixtures in chloroform solutions consisting of cholesterol, and six PUFAs. Full factorial design was used in this study because they are extremely easy to set up and easy to analyze [14]. Fig. 1 shows the full factorial design spectra. The spectral data matrix is decomposed in PCA by singular value decomposition (SVD) algorithms according to the equation below [15]:

$$\underset{nxp}{X} = \underset{nxnc}{C}\ \underset{ncxp}{S^T} \qquad (1)$$

where X is an n (=128) spectra at p (=101) wavelengths; C, 128x7 concentration matrix; S^T, 7x101 matrix of the pure spectra (n=128 is the number of mixture spectra, nc(=7) is the number of components, and p(=101) is the number of wavelengths). Eq. (1) shows the decomposition of the spectral matrix in real factors, a product of S^T of the spectra with a matrix C of concentration profiles.

By decomposing matrix X with a PCA as many significant principal components should be found as there are chemical species in the mixtures [15].

The decomposition in the wavelength space, for a system with seven components is given by:

$$\underset{nxp}{X} = \underset{nx7}{T^*}\ \underset{7xp}{V^{*T}} + \underset{nxp}{E} \qquad (2)$$

Eq. (2) shows the decomposition of the spectral matrix in abstract factors T^* and V^{*T} (E is the error). The score matrix T^* gives the location of the spectra defined by the seven principal components.

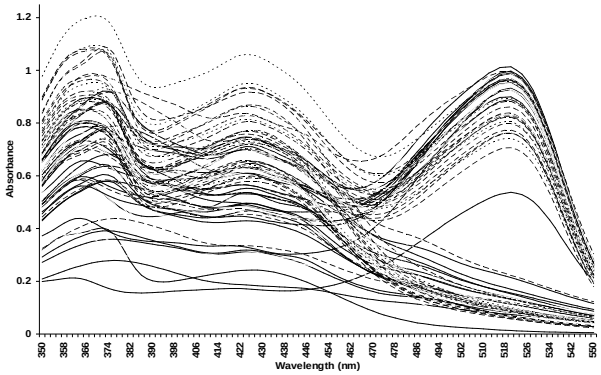

Fig. 1. Full factorial design (n=128) of synthetic serum samples in chloroform solutions.

In this study, all spectral wavelengths (350-550 nm) were used and consequent variable reduction was performed as discussed earlier. For 101-dimensional space, feature reduction was performed to a 2-dimensional data and the resulting principal components, PC1 and PC2 were plotted against each other. Generally, it is possible to perform 101 PCs in this study because the number of variables (m = 101) is less than the number of samples (n = 128). However, as one continues extracting PCs until m PCs are obtained, these will contain less and less variation and less information as well as significant loss of information. As such, the first two PCs were used in this study for they retain the largest information and the most variations [15].

After plotting the first two PCs, agglomerative hierarchical clustering using Ward's algorithm was performed. The method starts with each point as its own cluster. At each step the clustering process calculates the Ward's distance between each cluster and combines the two clusters that are closest together. This combining continues until all the points are in one final cluster [16]. Specifically, for clusters S_{w1} and S_{w2} whose cardinalities are N_{w1} and N_{w2} and centroids c_{w1} and $c_{w1,}$ respectively, Ward's distance is defined as:

$$dw(S_{w1}, S_{w2}) = \frac{N_{w1} N_{w2}}{N_{w1} + N_{w2}} d(c_{w1}, c_{w2}) \qquad (3)$$

where $d(c_{w1}, c_{w2})$ is the squared Euclidean distance between c_{w1} and c_{w2} [17]. Fig. 2 shows the dendogram resulting from hierarchical clustering. Each number corresponds to the clusters containing the samples of similar observations. Ward's method was used in this study for it leads to well-structured dendograms [11]. There are no completely satisfactory methods for determining the number of population clusters for any type of cluster analysis [18], [19]. In this study, the number of clusters was determined by using a Scree Plot found below the dendogram in Fig. 2. The place where the Scree Plot changes from a sharp downward slope to a more level slope is an indication of the number of clusters [20]. Eight clusters were identified corresponding to different levels of the prepared lipids according to the full factorial design of synthetic sets. The resulting plot of PC1 versus PC2 and the resulting clusters with 0.90 confidence density ellipses are shown in Fig. 3.

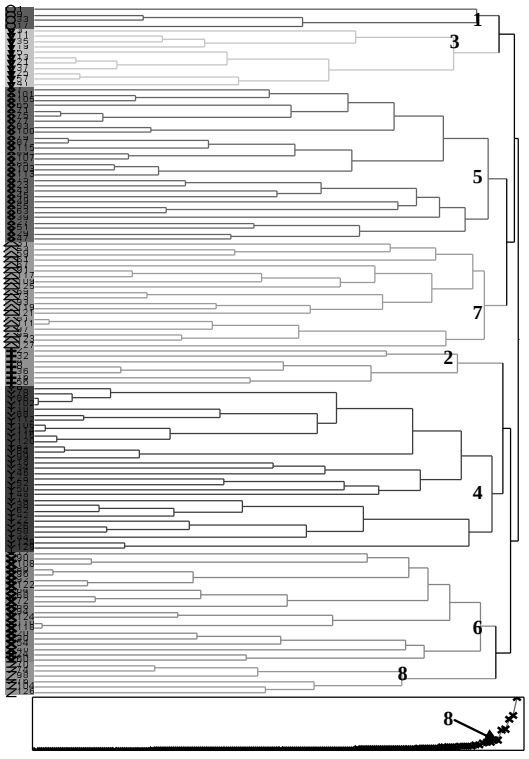

Fig. 2. Dendogram of full factorial design synthetic sets. Diagram below it is the Scree Plot with eight clusters.

ISBN: 978-988-19251-4-5 (revised on 20 September 2011) WCE 2011
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

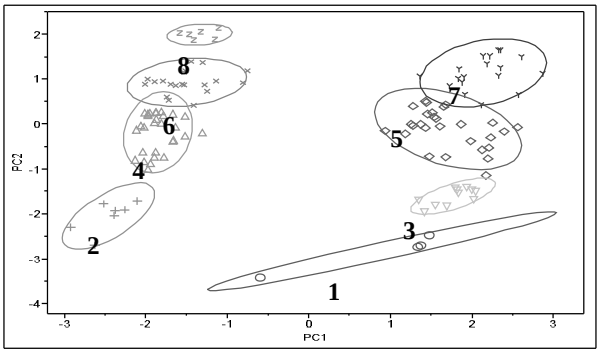

Fig. 3. Prepared synthetic mixtures of lipids in chloroform solutions. Each cluster corresponds to different levels of lipids prepared.

Cluster 1 has high cholesterol; average arachidonic, EPA, and DHA; and an absence of linoleic, linolenic, and CLA. Cluster 2 has high linoleic and EPA; average to high linoleic and arachidonic; average DHA; and an absence of cholesterol and CLA. Cluster 3 has high cholesterol; average to high linoleic, linolenic; average arachidonic, EPA, and DHA; and an absence of CLA. Cluster 4 has the absence to low cholesterol; low, average, and high linoleic, linolenic, arachidonic, EPA, DHA; and low to average CLA. Cluster 5, on the other hand has high cholesterol; low, average, and high linoleic and linolenic; low to average arachidonic, EPA, and DHA; and average to high CLA. Cluster 6 an absence of cholesterol; low to average linoleic; low, average, and high linolenic, arachidonic, EPA, and DHA; and average to high CLA. Cluster 7 has high cholesterol; low to average linoleic, arachidonic, EPA; and low, average, and high linolenic, DHA and CLA. Cluster 8 has the absences of cholesterol and EPA; low to average linoleic, linolenic, arachidonic, and DHA; and average to high CLA.

The separation suggests that the pattern can be used to examine the patterns generated from different mixtures.

A set of data collected by UCT personnel on samples from patients was analyzed using similar methods. After hierarchical clustering, the Scree Plot was able to identify ten clusters as shown in Fig. 4. The medical staff of the hospital in Cape Town determined that the clusters correspond with the dyslipidemic states of the patients. Table I shows the history of patients corresponding to clusters in Fig. 5.

The general trend that is seen in Fig. 5, to this point, is that from top to bottom one goes from the Type V (top) through to the Type IIs in clusters 3, 4, and 5 and end with control patients being distributed in 8 and 9. The clusters were associated according to Fredrickson's Classification of Lipid Disorders [6]. Cluster 1 contains the Type V patient pattern. Cluster 2 is Type III and Type IV (2 of each). Separation will improve with an increase in the sample size of these types and with the inclusion of the fatty acid profiles. Cluster 3 is a Type IIa pattern with a majority considered to be familial hypercholesterolemia (FH) (8/14) and familial combined hyperlipidemia (FCH) (4/14). Cluster 4 is a combination of Type IIa, IIb with controls with cholesterol values over 220 mg/dL. Cluster 5 is a combination of IIa, IIb, and some patients on various stages of treatments.

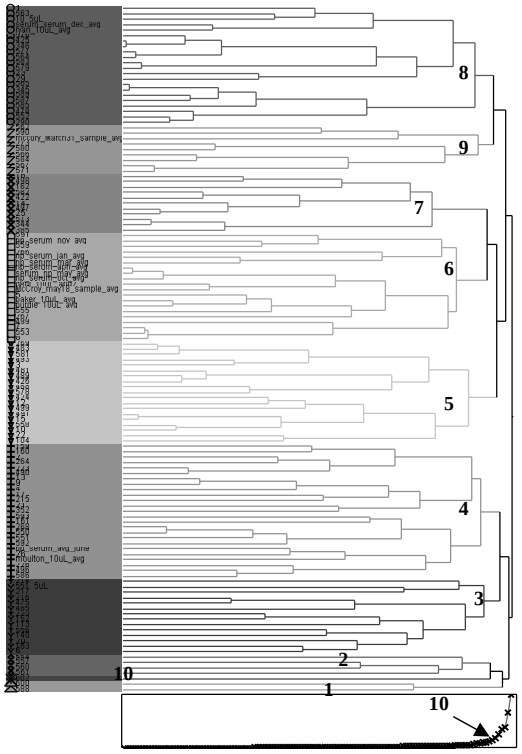

Fig. 4. Dendogram of UCT samples. Diagram below is the Scree Plot with ten clusters.

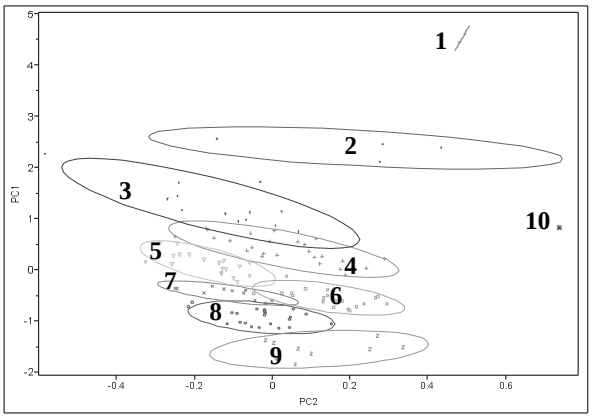

Fig. 5. PC1 vs PC2 of serum samples.

Cluster 6 is a combination of controls with cholesterol over 200 mg/dL. Cluster 7 is a cluster of samples with cholesterol between 193 and 240 mg/dL, statin treated patients, controls, diet induce d hypercholesterolemia, and people under altered diet treatment for hypercholesterolemia. Cluster 8 and 9 are clusters of controls and some statin treatment patients. The one sample in cluster 10 is a sample of a type V patient in cluster 1 that was run with half the required volume of serum due to the extremely high absorbance of the entire spectra. Although the collaborators were able to complete an initial analysis of the clusters, all of the following groups are subject to change as the fatty acid profile is increased and as more information about the medical history is gathered. Further collection of data and patients' histories including diet, treatments, and other health conditions will have to be collected to determine if this pattern recognition can be

ISBN: 978-988-19251-4-5 (revised on 20 September 2011)
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

used as a screening tool for dyslipidemias in a clinical setting.

Fig. 6 shows the average spectra for the clusters in Fig. 5. It is apparent that the corresponding spectra shows recognizable clusters that are different from one patient type to the other. One good example is the average spectrum for cluster 1, which is enormously high corresponding to Type V patients.

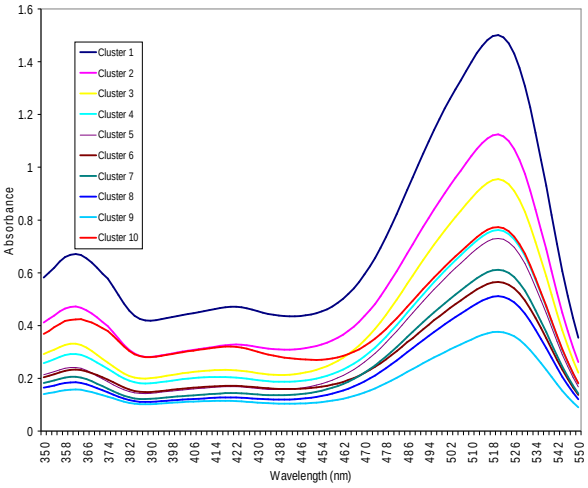

Fig. 6. Average spectra for the clusters in Fig. 5.

TABLE I. HISTORY OF PATIENTS CORRESPONDING TO CLUSTERS IN FIG. 5.

Cluster	Assay Code	Treatment	Dysproteinemia	Diagnosis	Cluster	Assay Code	Treatment	Dysproteinemia	Diagnosis
1	6	New	type V / FCH	V	6	3	New	mixed hyperlipidemia	IV
1	551	New	severe hypertriglyceridemia secondary to diabetes milletus, diet, obesity	V	6	4	New	consider FCH	II a
2	10	New	Dys-Beta	III	6	7	New	moderate dyslipidemia	II a
2	13	New	Dys-Beta	III	6	8	New	moderate hypercholesterolemia	II a
2	21	New	IV , FCH	IV	6	9	New	considered for FH	II- FH
2	163	New	IV , FCH	IV	6	10.5	New	Dys-Beta, confirmed	III
3	5	New	FH	II a	6	344	New	mild mixed hyperlipidemia with hypo-alpha	II b
3	12	New	Dys-Beta	III	6	475	New	diabetes, possibly underlying lipolytic deffect	V
3	160	New	possible FCH	II a	6	496	New	FH	II a
3	264	New	FCH is possible, high levels on excellent diet	II a	6	497	New	FH	II a
3	266	New	likely FH	II a	6	791	Normal (N)	N	N
3	267	New	could be FH but does not have Xanthomata	II a	6	793	N	N	N-treated
3	385	New	serve primary low density lipoprotein (LDL), could be FCH but no family history to support	II a	6	781	N	N	N
3	422	New	homo FH	II a	6	782	N	N	N
3	425	New	possible FH but also consider FCH	II a	6	783	N	N	N
3	426	New	serve primary LDL, could be FCH but no family history to support	II a	6	784	N	N	N
3	485	New	achilles tendon suggestive of FH	II a	6	785	N	N	N
3	499	New	dysbeta or FH	II b	6	786	N	N	N
3	564	trail, treated	FH	II a - tt	6	787	N	N	N
3	571	new	FH Afrikaner	II a	6	788	N	N	N
4	2	first visit	likely low density lipoprotein receptor mutation/ B-100 genes	II a	7	25	new, diet control only	hypercholesterolemia, primary LDL	II a
4	14	New	consider FH/FCH	II a	7	104	new, diet control only	hypercholesterolemia, primary LDL	II a
4	20	New	consider FH	II a	7	113	New	very, very much the diet	II b
4	140	New	moderate mixed hyperlipidemia, could be polygenic, FCH (needs family history)	II b	7	162	New	diet, Increased triglycerides	II b
4	152	New	FH	II a	7	217	New, father of assay # 215 and 216	FH	II a
4	161	New	consider FCH or possibly polygenic	II a	7	352	New	Polygenic	II a
4	223	New	Polygenic and diet aggravation	II a	7	555	trail, treated	FH	II a - tt
4	268	New	FH most likely though total cholestetol (TC) low for FH	II a	7	559	trail, treated	FH	II a - tt
4	269	New	FH Afrikans I and apo E2/E2	II a	7	560	trail, treated	FH	II a - tt
4	273	New	Moderate hypercholesterolemia	II a	7	581	N	N	N
4	424	New	IIa could be FCH	II a	7	588	N	N	N
4	478	New	diabetes control, possibly lipolytic	V	8	1	New, but on treatment	likely moderate mixed hyperlipidaemia	II
4	481	New	could be secondary dyslipidemia	II a	8	16	New	FH	II a
4	483	New	moderately severe, Hyper-LDL-cholesterolemia	II b	8	159	New	Normolipidemic	N
4	488	New	FH phenotype	II b	8	214	New	severe hypertriglyceridaemia	V
4	489	New	IIa likely primary, could be FCH	II a	8	216	New	FH	II a

ISBN: 978-988-19251-4-5 (revised on 20 September 2011)
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

4	490	New	IIa, could be FCH considering family history	II a	8	225	New	no major gene defect, dietary	II a
4	491	New	severe hypercholesterolemia, particularly related to renal disease	II	8	554	trail, treated	FH	II a – tt
4	493	New	severe hypertriglyceridemia	V	8	556	trail, treated	FH	II a – tt
4	494	New	primary LDL hypercholesterolemia, could be FCH but no family history	II a	8	557	trail, treated	FH	II a – tt
4	498	New	FCH or polygenic	II a	8	562	trail, treated	FH	II a – tt
4	558	trail, treated	FH	II a – tt	8	565	trail, treated	FH	II a – tt
4	592	N	N	N	8	573	New	mixed hyperlipidemia, possible dys-Beta	II b
4	728	New	FCH or polygenic	IIa	8	576	N	N	N
4	733	New	FCH or polygenic	IIa	8	577	N	N	N
5	15	New	FCH/ FH	II a	8	582	N	N	N
5	17	New	FH / familial fefective apolipoprotein B-100	II a	8	583	N	N	N
5	22	New	diet and increased triglycerides	II b	8	585	N	N	N
5	23	New	moderate mixed hyperlipidemia, could be polygenic, FCH (needs family history)	II b	8	589	N	N	N
5	26	New	FH	II a	8	591	N	N	N
5	29	New	dysbetalipoproteineamia, likely 2 or Apolipoprotein E2/E2	III	8	593	N	N	N
5	215	New	FH	II a	8	771	N	N	N
5	226	New	FH	II a	8	773	N	N	N
5	345	New	could be FCH, but no family history known	II b	9	290	New	severe hypertriglyceridemia	V
5	550	New	IIa pattern, no clinical signs of FH, FCH possible	II	9	346	New	isolated hypertriglyceridemia	IV
5	552	trail, treated	FH	II a – tt	9	553	trail, treated	FH	II a – tt
5	561	trail, treated	FH	II a – tt	9	578	N	N	N
5	563	trail, treated	FH	II a – tt	9	579	N	N	N
5	566	trail, treated	FH	II a – tt	9	580	N	N	N
5	567	trail, treated	FH	II a – tt	9	586	N	N	N
5	584	N	N	N	9	590	N	N	N
5	587	N	N	N	9	775	N	N	N
5	600	New	FH	II a	10	551	New	severe hypertriglyceridemia secondary to diabetes milletus, diet, obesity	V
5	602	New	FH	II a					

IV. CONCLUSIONS

A cluster diagram created from PCA data derived from serum spectra had ten clusters which suggest separation based on dyslipidemias. Prepared mixtures also yielded separations based on the components present and the concentration ranges of each of the components. Both cluster analyses suggest that the spectrum itself can offer a benefit as a possible screening tool for dyslipidemias. Further studies with an in depth analysis of medical histories must be completed before any final conclusion about the clustering can be established. Once completed, the assay will be a useful clinical assay for a wide variety of research projects including further studies on the clustering and possible pattern recognition methods for the screening of dyslipidemias. The main advantages of the assay are reduction in time and costs to carry out the analyses. This would be most appropriate in a typical clinical setting.

ACKNOWLEDGMENTS

We sincerely acknowledge the receipt of serum samples and assistance of the medical staff at the University of Cape Town's Lipid Clinic.

REFERENCES

[1]. S. Fiorucci, S. Cipriani, F. Baldelli, A. Mencarelli, "Bile acid-activated receptors in the treatment of dyslipidemia and related disorders," *Progress in Lipid Research*, vol. 49, pp. 171–185, 2010.

[2]. D.G. Smith, "Epidemiology of dyslipidemia and economic burden on the healthcare system," *Am J Manag Care*, vol. 13, no. (3 suppl), pp. s69-s71, 2007.

[3]. D.S. Fredrickson, "An International classification of hyperlipidemias and hyperlipoproteinemias," *Annals of Internal Medicine*, vol. 75, no. 3, pp. 471-472, 1971.

[4]. C.G. Magnussen, O.T. Raitakari, R. Thomson, M. Juonala, D.A.Patel, J.S.A. Viikari, "Marniemi, J.; Srinivasan, S. R.; Berenson, G. S.; Dwyer, T.; Venn, A. Utility of currently recommended pediatric dyslipidemia classifications in predicting dyslipidemia in adulthood: Evidence from the childhood determinants of adult health (CDAH) study, cardiovascular risk in Young Finns Study, and Bogalusa Heart Study," *Journal of the American Heart Association*, vol. 117, pp.32-42, 2007.

[5]. V.N. Malakhov, V.V. Pedchenko, "Enzymatic methods for quantitative determination of total cholesterol in blood serum," *Vopr Med Khim*, vol. 37, no. 4, pp. 85-91, 1991.

[6]. **A.** Lohninger, P. Preis, L. Linhart, S.V. Sommoggy, M. Landau, E. Kaiser, "Determination of plasma free fatty acids, free cholesterol, cholesteryl esters, and triacylglycerols directly from total lipid extract by capillary gas chromatography," *Analytical Biochemistry*, vol. 2, pp. 243-250, 1990.

[7]. G.G. Dumancas, M. Muriuki, N. Purdie, L. Reilly, "Simultaneous spectrophotometric and chemometric determination of lipids in synthetic mixtures and human serum," *Lipid Technology*, vol. 21, no. (5/6), pp. 127-130, 2009.

[8]. G.G. Dumancas, M. Muriuki, N. Purdie, L. Reilly, "Chemometric algorithms for the direct determination of lipids in synthetic mixtures and human serum," *Journal of Biotech Research*, vol. 2, pp. 34-43, 2010.

[9]. G.G. Dumancas, M. Kimani, N. Purdie, L. Reilly, "Partial least squares (PLS1) algorithm for quantitating cholesterol and polyunsaturated fatty acids in human serum," *Journal of Biotech Research*, vol. 2, pp. 121-130, 2010.
[10]. M.A. Sharaf, D.L. Illman, B.R. Kowalski, *Chemometrics*. Wiley-Interscience, 1986, pp.183-188.
[11]. M. Otto, *Chemometrics: Statistics and computer application in analytical chemistry*. Wiley-VCH, 1999, pp. 155, 201-203.
[12]. SAS. *JMP 8 Statistical Discovery Software, 4.0*, SAS Institute, Inc.: Cary, NC, USA, 2008.
[13]. B. Zak, J.D. Artiss, "Decreased water: increased solids: distorted serum concentrations," *Microchemical Journal*, vol. 72, no. 3, pp. 235-240, 2002.
[14]. A.C. Atkinson, A.N. Donev, *Optimum experimental designs*. Oxford, 1992, p. 62.
[15]. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Data Handling in Science and Technology Volume 20 Handbook of Chemometrics and Qualimetrics: Part A*. Elsevier Science, 1997, pp. 519-527, 541.
[16] SAS JMP. *JMP 8 Statistics and graphics guide volume 1*, SAS Institute, Inc.: Cary, NC, USA, 2008.
[17]. B. Mirkin, B., *Clustering for data mining: a data recovery approach* Chapman & Hall/CRC, 2005, p. 114.
[18]. B.S. Everitt, "Unresolved problems in cluster analysis," *Biometrics*, vol. 35, pp. 169–181, 1979.
[19]. J.A. Hartigan, "Statistical theory in clustering," *Journal of Classification*, vol. 2, pp. 63–76, 1985.
[20]. J. Sall, L. Creighton, A. Lehman, *JMP Start statistics 3rd Ed.: A guide to statistics and data analysis using JMP and JMP IN software*. SAS Institute, Inc.: Cary, NC, USA, 2005.

Date of modification: September 6, 2011
Brief description of changes: Address and email of Dr. Neil Purdie changed to OSU Chemistry Department, 018 Physical Science Building, Stillwater, OK 74078, USA (email: neil.purdie@okstate.edu)