

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
15 December 2011 (15.12.2011)

PCT

(10) International Publication Number
WO 2011/154259 A1

(51) International Patent Classification:
G01N 33/68 (2006.01) *G06K 9/00* (2006.01)

(21) International Application Number:
PCT/EP2011/058596

(22) International Filing Date:
25 May 2011 (25.05.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/344,198 9 June 2010 (09.06.2010) US

(71) Applicant (for all designated States except US): **GALDERMA RESEARCH & DEVELOPMENT** [FR/FR]; 2400 Route des Colles, Les Templiers, F-06410 Biot (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ANDRES, Philippe** [FR/FR]; 126, avenue des Termes, F-06530 Peymeinade (FR). **ZUGAJ, Didier** [FR/FR]; Cap Soleil, bâtiment 10, 10, rue de la Soulane, F-06250 Mougins le Haut (FR). **LAFFET, Gilbert** [FR/FR]; 23 Les bastidons, 254, chemin des cabot, F-06410 Biot (FR). **SCHOOT, Bernardus Maria** [FR/FR]; 212 Avenue Georges Pompidou, F06220 Golfe Juan-Vallauris (FR). **MARTEL, Philippe** [FR/FR]; Villa de Green Side n° 9, 210, Avenue Saint Philippe, F-06410 Biot (FR).

(74) Agent: **ALLAB, Myriam**; L'OREAL, D.I.P.I., 25-29 Quai Aulagnier, F-92665 ASNIERES-SUR-SEINE Cedex (FR).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: Personal Identification Based on Sebum Composition

(57) Abstract: Methods and systems for personal identification or authentication based on sebum compositions are described. The methods and systems are based on the discovery that an individual has a relatively reproducible and distinct global profile of sebum compositions from a statistical clustering analysis, such as a principal component analysis (PCA), of chromatograms obtained from chromatographic analyses of sebum samples from the individual.



WO 2011/154259 A1

Personal Identification Based on Sebum Composition

BACKGROUND OF THE INVENTION

[0001] Sebum is the oily and waxy substance secreted by sebaceous glands in the skin. It lubricates the skin and hair of animals, has the waterproof function, and can also inhibit the growth of some microorganisms, such as some bacteria, on the skin. In adults with normal active sebaceous glands, sebum is the predominant lipid on most areas of the skin surface. The amount of sebum on a particular area of the skin reflects the size and density of distribution of the sebaceous glands in that area. In human, the amount of surface lipid, as well as the density of sebaceous glands, is found in decreasing order on the face, back, chest, abdomen, arms and legs. The palms and soles, though devoid of sebaceous glands, normally carry some sebum transferred from other areas.

[0002] Sebum is a complex mixture of several lipid classes, such as glycerides, wax esters, squalene, free fatty acids, sterol esters, sterols, saturated hydrocarbons, etc. The major constituents of sebum from sebum-rich areas varies little in normal subjects of the same species. However, the rate of sebum synthesis varies widely in normal adults. Also, the amount of each sebum constituent varies. For example, a sebum sample may include about 5% to 20% of squalene, about 50% to 70% of a mixture of triglycerides of linear fatty acids having a chain of 12 to 22 carbon atoms and of linear fatty acids having a chain of 12 to 22 carbon atoms, about 15% to 25% of esters of linear fatty acids and of linear fatty alcohols having chains of 12 to 22 carbon atoms, about 0.5% to 3% of esters of cholesterol whose acid fraction comprises a chain of 12 to 22 carbon atoms, and about 0% to 5% of cholesterol having the ratio of unsaturated fatty chains to saturated fatty chains lies between 10 and 0.1.

[0003] Effects of various factors on sebum have been studied. For example, it was reported that changing the rate of sebaceous lipogenesis, e.g., by oral administration of estrogens or androgens, or inhibitors of prostaglandin biosynthesis, did not change the composition of the sebum synthesized. See Downing and Strauss, *The J. Invest. Dermatology*, 62:228-244, 1974. The percentage amount of certain constituents in sebum, such as wax ester, squalene, and cholesterol, have changed in patients with different degrees of pellagra as compared with that in controls in the same geographical area, Transvaal, South Africa. Dogliotti, et al., *British Journal of Dermatology*, Volume 97 Issue 1, Pages 25 – 28. Published Online: 29 Jul 2006. The changes, which have been observed in starved but otherwise normal subjects, were reversed with adequate refeeding and treatment with nicotinamide. Whether the changes are due to the lack of energy or the absence of specific components in the diet is unknown. The amount of sebum on the surface of the skin has been used for studying a person's skin type, e.g., dry, normal or oily. Squalene in

hair has been used for the improved interpretation of fatty acid ethyl ester concentrations with respect to alcohol misuse. Auwärter et al., Forensic Science International, Volume 145, Issue 2, Pages 149-159.

[0004] Chromatographic analysis can be used for separating and identifying constituents contained within a sebum sample, e.g., by the absorbance, transmission, reflectance or emission of each separated constituent as a function of time of the illuminating or scattered light transmitted through the mixture. Such measurements of constituents within the sebum sample as "peaks" in a plot, e.g., a function of time of the illuminating or scattered light transmitted through in the plot, is generally referred to as a chromatogram of the chromatographic analysis. The chromatogram contains a number of peaks of various constituents within the sebum sample. Thus, the chromatogram corresponds to a multi-dimensional data space, with each of the peaks defining a dimension. In order to compare the compositions of different sebum samples, different multi-dimensional data spaces must be compared, which can be mathematically challenging.

[0005] Biometrics, such as fingerprint, voice, face, palm, etc., have been used for personal identification and/or verification. However, technologies have also been developed to circumvent the existing personal identification methods and devices.

[0006] There is a need of a novel method and system for personal identification and/or verification that are difficult to circumvent. Such method and system based on sebum composition are described in the present application.

BRIEF SUMMARY OF THE INVENTION

[0007] It is now surprisingly discovered that an individual has a relatively reproducible and distinct global profile of sebum compositions from a statistical clustering analysis, such as a principal component analysis (PCA), of chromatograms obtained from chromatographic analyses of sebum samples from the individual.

[0008] Accordingly, in one general aspect, embodiments of the present invention relate to a method of identifying or authenticating a subject. The method comprises:

(a) obtaining a test chromatogram from a chromatographic analysis of a test sebum sample taken from the subject;

(b) analyzing the test chromatogram using a statistical clustering analysis algorithm to obtain a test vector of p variables, wherein p is less than the number of peaks in the test chromatogram; and

(c) comparing the test vector with a reference vector of p variables, wherein the reference vector is specific to a reference subject and is obtained from analyzing a reference chromatogram using the statistical clustering analysis algorithm, and the reference chromatogram

is obtained from the chromatographic analysis of a reference sebum sample taken from the reference subject,

wherein a close proximity between the test vector and the reference vector identifies or authenticates the subject as the reference subject.

5 **[0009]** In one embodiment of the present invention, the analyzing step comprises:

- (a) saving into a computer the test chromatogram;
- (b) selecting significant peaks from the test chromatogram; and
- (c) compressing the significant peaks selected in (b) into the test vector of p variables

using the statistical clustering analysis algorithm, $1 \leq p \leq$ the number of significant peaks; and

10 wherein the sub-steps of saving, selecting, compressing and the step of comparing are accomplished with the computer.

[0010] Preferably, the significant peaks are identified as relevant to the specificity of the reference vector to the reference subject.

15 **[0011]** In an embodiment of the present invention, the statistical clustering analysis algorithm is a principal component analysis (PCA) algorithm performing a computer assisted process comprising:

- (a) providing a covariance matrix of the significant peaks;
- (b) calculating the eigenvectors and eigenvalues of the covariance matrix;
- (c) ranking the eigenvalues and corresponding eigenvectors in a decreasing order; and

20 (d) compressing the significant peaks into a vector of p eigenvectors corresponding to the top p eigenvalues.

[0012] In a preferred embodiment of the present invention, the cumulative value of the top p eigenvalues is at least 70% of the sum of total eigenvalues.

[0013] In another preferred embodiment of the present invention, $p = 3$.

25 **[0014]** Other aspects of the present invention include systems and program products relating to methods according to embodiments of the present invention.

[0015] Other aspects, features and advantages of the invention will be apparent from the following disclosure, including the detailed description of the invention and its preferred embodiments and the appended claims.

30 BRIEF DESCRIPTION OF THE DRAWINGS

[0016] The foregoing summary, as well as the following detailed description of the invention, will be better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, there are shown in the drawings embodiments of the invention. It should be understood, however, that the invention is not limited to the precise arrangements and
35 instrumentalities shown.

[0017] In the drawings:

[0018] Figure 1A shows multiple chromatograms resulting from GC/MS analysis of sebum samples obtained from 5 human subjects;

[0019] Figure 1B shows multiple chromatograms resulting from GC/MS analysis of sebum samples obtained from a single human subject at 8 visits;

[0020] Figure 2 shows a chromatogram before (upper) and after (lower) baseline correction;

[0021] Figure 3A is the average chromatogram of 47 normalized chromatograms;

[0022] Figure 3B is a bar graph illustrating the intensity of each selected peak: the peaks identified with a circle are the average peaks that are significant in terms of intensity (higher than a predetermined threshold value, e.g., 2%, with a retention time > 2min) and also are validated in term of relevance concerning classes of lipid they belong to; the PCA is next conducted only on this set of peaks;

[0023] Figure 4 shows peak localization of the peak with a retention time of about 12.067 minutes from multiple chromatograms of an individual: the maximum intensity of each chromatogram was computed in a window of retention time, e.g., 12 to 12.162 minutes, the window was automatically determined with the extraction of minimal values on the average chromatogram; and

[0024] Figure 5 shows a 3 dimensional graph of the 3 most significant components from a PCA analysis of sebum samples obtained from 5 different subjects, collected at multiple visits, with 1 week intervals between two samplings.

DETAILED DESCRIPTION OF THE INVENTION

[0025] Various publications, articles and patents are cited or described in the background and throughout the specification; each of these references is herein incorporated by reference in its entirety. Discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is for the purpose of providing context for the present invention. Such discussion is not an admission that any or all of these matters form part of the prior art with respect to any inventions disclosed or claimed.

[0026] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention pertains. Otherwise, certain terms used herein have the meanings as set in the specification. All patents, published patent applications and publications cited herein are incorporated by reference as if set forth fully herein. It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise.

[0027] In one general aspect, the invention relates to a method of characterizing a sebum sample using a statistical clustering analysis, such as a principal component analysis (PCA). The PCA (or any other statistical clustering analysis) is preformed on peaks, preferably selected

significant peaks, from chromatograms obtained from a chromatographic analysis of the sebum sample.

[0028] It has been discovered that a statistical clustering analysis, such as a principal component analysis (PCA), can be used to reduce the dimensionality of a chromatogram obtained from a chromatographic analysis of a sebum sample. Instead of a large number of peaks, the chromatogram can be characterized by a much reduced number of variables, such as a vector of three or more variables, that reveals the internal structure of the multi-dimensional data space of the chromatogram in a way that best explains the variance in the data space.

[0029] Accordingly, a sebum sample can be characterized by a method comprising:

- (a) obtaining a chromatogram from a chromatographic analysis of the sebum sample;
- (b) analyzing the chromatogram using a statistical clustering analysis algorithm to obtain a vector of p variables, wherein p is less than the number of peaks in the chromatogram; and
- (c) characterizing the sebum sample as the vector.

[0030] In one embodiment of the present invention, the analyzing step comprises:

- (a) saving into a computer the chromatogram;
- (b) selecting significant peaks from the chromatogram;
- (c) compressing the significant peaks into the vector of p variables using the statistical clustering analysis algorithm, $1 \leq p \leq$ the number of significant peaks; and

wherein the sub-steps of saving, selecting, compressing and the step of characterizing are accomplished with the computer.

[0031] In one embodiment of the present invention, a plurality of sebum samples can be compared by a method comprising:

- (a) obtaining a plurality of chromatograms, each chromatogram obtained from a chromatographic analysis of each of the plurality of sebum samples, respectively;
- (b) analyzing each of the plurality of chromatograms using a statistical clustering analysis algorithm to obtain a plurality of vectors, each vector having p variables, wherein p is less than the number of peaks in each of the plurality of chromatogram;
- (c) comparing the plurality of sebum samples by comparing the plurality of vectors.

[0032] In an embodiment of the present invention, the analyzing step comprises:

- (a) saving into a computer the plurality of chromatograms;
- (b) selecting significant peaks from each of the plurality of chromatograms; and
- (c) compressing the significant peaks for each of the plurality of chromatograms into the vector of p variables using the statistical clustering analysis algorithm, $1 \leq p \leq$ the number of significant peaks;

wherein the sub-steps of saving, selecting and compressing, and the step of comparing are accomplished with the computer.

[0033] As used herein, the term "subject" means any animal, preferably a mammal, most preferably a human, from whom a sebum sample will be or has been obtained for analysis according to embodiments of the invention. The term "mammal" as used herein, encompasses any mammal. Examples of mammals include, but are not limited to, cows, horses, sheep, pigs, cats, dogs, mice, rats, rabbits, guinea pigs, monkeys, humans etc., more preferably, a human.

[0034] The sebum sample used in embodiments of the present invention can be collected from a subject using any methods known in the art in view of the present disclosure. For example, the sebum sample can be collected by a non-invasive method, such as by an absorption tape, from the skin surface of a subject. In an embodiment of the present invention, the sebum sample is collected from the forehead of a human subject using any known method for sebum collection in view of the present disclosure.

[0035] The sebum sample can be collected from many body locations on the subject. For example, in a human, the sebum sample can be collected from one or more locations of the forehead, nose, cheek, shoulder, neck, or top of the back of the human.

[0036] The collected sebum sample can be analyzed by various chromatographic analysis, includes, but is not limited to, gas chromatography (GC), gas chromatography-mass spectrometry (GC-MS), high-performance liquid chromatography (HPLC), high-performance liquid chromatography-mass spectrometry (HPLC-MS), liquid chromatography (LC), liquid chromatography-mass spectrometry (LC-MS), thin layer chromatography (TLC), thin layer chromatography-mass spectrometry (TLC-MS), etc.

[0037] The chromatogram obtained from the chromatographic analysis contains various peaks, each of which represents one or more constituents of the sebum sample, selected from the group consisting of free fatty acids, squalene, cholesterol, waxes, cholesterol esters, diglycerides and triglycerides and various minor unknown constituents. Depending on the method of the chromatographic analysis and the constituents in the sebum sample, the peaks can be completely or partially separated from one another. Each of the peaks defines a dimension of the data space for a chromatogram. Accordingly, a chromatogram containing n peaks generally defines an n-dimensional data space.

[0038] The chromatogram can be corrected to remove signals and information that are not due to the composition of the analyzed sebum sample, using any of a variety of known correction methods in view of the present disclosure. Examples of such undesired signals and information include various instrumental effects, e.g., that relating to the transmission of optical elements, the sensitivity of the detector, and any other non-desired sample effects due to the instrument utilized to collect the chromatograms, such as fluorescence in the case of Raman chromatograms. However, one skilled in the art will appreciate that uncorrected chromatograms may also be utilized

to practice the inventive method, without departing from the spirit and scope of the present invention.

[0039] After the chromatogram has been corrected to remove instrumental artifacts, such as by baseline subtraction with a control sample, eventually, the significant peaks within the chromatogram are selected to further increase the signal-to-noise ratio and reduce the variability among different chromatograms.

[0040] As used herein, "significant peaks" refers to the peaks of significant intensity that can be reproducibly detected from one sebum sample or peaks of significant intensity that can be consistently detected from a plurality of sebum samples by the chromatographic analysis. To select significant peaks from a plurality of sebum samples, the chromatograms from the plurality of sebum samples are aligned using methods known in the art in view of the present disclosure. The chromatograms from the plurality of sebum samples can also be processed to generate an average chromatogram, which can then be used for the selection of significant peaks shared by the plurality of sebum samples.

[0041] Different sebum samples can share a number of significant peaks, indicating that these sebum samples can share a number of common constituents. However, the sebum samples can be different in intensity of one or more of the shared significant peaks, thus different in amount of one or more common constituents.

[0042] In a preferred embodiment of the present invention, the identity of the constituent or constituents within each of the significant peaks is not required to be known.

[0043] It is readily understood that different numbers of significant peaks can be selected from a chromatogram, depending on the criteria used for significant peak selection, such as the threshold signal-to-noise ratio, the range of retention time of interest, the relevance concerning classes of lipid the significant peaks belong to, etc. In one embodiment, the number of significant peaks is equal to the number of total peaks in a chromatogram. In another embodiment, the number of significant peaks is less than the number of total peaks in a chromatogram.

[0044] In the cases where a peak selection is applied, and m significant peaks are selected from the n peaks of a chromatogram, the dimensionality of the data space for the chromatogram is reduced from n to m . Although m is smaller than n , it may still be a relatively large number. A statistical clustering analysis, such as a principal component analysis (PCA), reduces the m -dimensional data space to a data space with much reduced dimensionality.

[0045] PCA (or any other statistical clustering analysis) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables, such as principal components in PCA. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. PCA involves the calculation of the eigenvalue

decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA supplies the user with a lower-dimensional picture, a "shadow" of this object when viewed from its most informative viewpoint.

[0046] PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA is theoretically the optimum transform for given data in least square terms.

[0047] Given a set of points in Euclidean space, the first principal component (the eigenvector with the largest eigenvalue) corresponds to a line that passes through the mean and minimizes sum squared error with those points. The second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted out from the points. Each eigenvalue indicates the portion of the variance that is correlated with each eigenvector. Thus, the sum of all the eigenvalues, or the sum of total eigenvalues, is equal to the sum squared distance of the points with their mean divided by the number of dimensions. PCA essentially rotates the set of points around their mean in order to align with the first few principal components. This moves as much of the variance as possible (using a linear transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be highly correlated and may be dropped with minimal loss of information. PCA is often used in this manner for dimensionality reduction. PCA has the distinction of being the optimal linear transformation for keeping the subspace that has largest variance.

[0048] In an embodiment of the present invention, a statistical clustering analysis, such as a PCA, compresses the m significant peaks into a vector of p variables, such as p principal components using a PCA algorithm, wherein $p \leq m$.

[0049] The PCA algorithm provides an $m \times m$ covariance matrix of the significant peaks; calculates the eigenvectors and eigenvalues of the covariance matrix; ranks the eigenvalues and corresponding eigenvectors in a decreasing order; and compresses the significant peaks into a vector of the p eigenvectors corresponding to the top p eigenvalues.

[0050] Depending on the structure of the data of the chromatogram and the selected significant peaks subject to PCA, the eigenvalues vary. If the data do not share any correlation (e.g. when data are randomly distributed), it is not possible to exhibit some relevant projection, thus the explained variance, or the cumulative eigenvalue for the top p eigenvalues relative to the sum of total eigenvalues, would be low.

[0051] In embodiments of the present invention, the cumulative eigenvalue for the top p eigenvalues is about 40% to 60%, 60% to 80%, 80% to 90%, or 90% to 95% of the sum of total eigenvalues.

[0052] In preferred embodiments of the present invention, the cumulative eigenvalue for the top p eigenvalues is at least about 70%, 75%, 80%, 85%, 90% or 95% of the sum of total eigenvalues.

[0053] The choice of p for a threshold explained variance can be determined by computing:

$$\frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^N \lambda_i} > Threshold$$

wherein λ_i is the i^{th} eigenvalue (reordered in a decreasing order); N is the total number of peaks, e.g., the m significant peaks, analyzed by PCA.

[0054] In a preferred embodiment of the present invention, $p = 3$ for a threshold of 0.9 (90% of the explained variance), i.e., the cumulative eigenvalue for the top 3 eigenvalues is at least 90% of the sum of total eigenvalues.

[0055] The high signal-to-noise ratio of the significant peaks helps to ensure that the principal components identified by PCA correspond to interesting dynamics of the data set viewed from its most informative viewpoint.

[0056] After the m significant peaks are compressed into a vector of p variables, the sebum sample is then characterized by a p dimensional data space, instead of the original n dimensional data space, $p \ll n$.

[0057] In a preferred embodiment of the present invention, the sebum sample is characterized by a vector of three variables, e.g., three principal components, resulting from PCA (or any other statistical clustering analysis) of all peaks or at least the m significant peaks of the chromatograms obtained from chromatographic analysis of the sebum sample.

[0058] Because PCA involves only rotation and scaling, it has simplified the algebraic computation on the data set for the chromatogram, thus has provided a simplified global or overall profile analysis of any sebum sample.

[0059] Another general aspect of the invention relates to a method of comparing a plurality of sebum samples using a statistical clustering analysis, such as a principal component analysis.

[0060] In one embodiment, chromatograms obtained from chromatographic analyses of the plurality of sebum samples are stored into a computer, if necessary, significant peaks from each of the chromatograms are selected, PCA (or any other statistical clustering analysis) is performed on all the selected significant peaks to reduce each of the chromatograms into a vector of p principal

components, and the plurality of sebum samples are compared by comparing the vector for each of the chromatograms.

[0061] In an embodiment of the present invention, the significant peaks are shared by the plurality of sebum samples. However, shared significant peaks are not required so long as the same criteria are used in selecting the significant peaks from each of the plurality of sebum samples.

[0062] After the statistical clustering analysis, such as a PCA, each sebum sample is represented by a vector of p variables. In one embodiment of the present invention, $p = 3$. Thus, PCA (or any other statistical clustering analysis) has also simplified the algebraic computation for the comparison of different sebum samples.

[0063] Embodiments of the present invention also relate to systems and computer products for characterizing a sebum sample or comparing different sebum samples using a PCA (or any other statistical clustering analysis). Any known computer systems or computer products can be adopted for the present use in view of the present disclosure.

[0064] In one embodiment, the present invention relates to a system for characterizing a sebum sample, the system comprising a computer system capable of:

- (a) storing a chromatogram from a chromatographic analysis of the sebum sample;
- (b) analyzing the chromatogram using a statistical clustering analysis algorithm to obtain a vector of p variables, wherein p is less than the number of peaks in the chromatogram; and
- (c) characterizing the sebum sample as the vector.

[0065] The system can further comprise an apparatus for chromatographic analysis of the sebum sample to obtain the chromatogram.

[0066] In another embodiment, the present invention relates to a system for comparing a plurality of sebum samples, the system comprising a computer system capable of:

- (a) storing a plurality of chromatograms, each chromatogram obtained from a chromatographic analysis of each of the plurality of sebum samples, respectively;
- (b) analyzing each of the plurality of chromatograms using a statistical clustering analysis algorithm to obtain a plurality of vectors, each vector having p variables, wherein p is less than the number of peaks in each of the plurality of chromatograms;
- (c) characterizing the plurality of sebum samples with the plurality of vectors, respectively; and
- (d) comparing the plurality of sebum samples by comparing the plurality of vectors.

[0067] The system can further comprise an apparatus for chromatographic analysis of each of the plurality of sebum samples to obtain the plurality of chromatograms.

[0068] In another embodiment, the present invention relates to a program product for characterizing a sebum sample using a statistical clustering analysis algorithm. The program

product comprises one or more computer programs stored on a recordable medium. The one or more computer programs are capable of:

(a) saving into a computer a chromatogram obtained from a chromatographic analysis of the sebum sample;

5 (b) selecting significant peaks from the chromatogram;

(c) compressing the significant peaks into a vector of p variables using the statistical clustering analysis algorithm, $1 \leq p \leq$ the number of significant peaks; and

(d) characterizing the sebum sample as the vector.

[0069] In another embodiment, the present invention relates to a program product for comparing a plurality of sebum samples using a statistical clustering analysis algorithm. The program product comprises one or more computer programs stored on a recordable medium. The one or more computer programs are capable of:

15 (a) saving into a computer a plurality of chromatograms, each of the plurality of chromatograms being obtained from a chromatographic analysis of each of the plurality of sebum samples;

(b) selecting significant peaks from each of the plurality of chromatograms;

(c) compressing the significant peaks for each of the plurality of chromatograms into a vector of p variables using the statistical clustering analysis algorithm, $1 \leq p \leq$ the number of significant peaks; and

20 (d) characterizing the plurality of sebum samples with a plurality of vectors obtained in (c), whereby the plurality of sebum samples are compared by comparing the plurality of vectors.

[0070] It is readily appreciated by those skilled in the art that embodiments of the present invention may also be used for the analysis of analytical results obtained from other types of sebum analysis. For example, a statistical clustering analysis algorithm, such as a PCA, can also be utilized for the analysis and comparison of spectra obtained from mass spectrometry analyses of sebum samples.

[0071] It is now surprisingly discovered that an individual has a relatively reproducible and distinct global profile of sebum composition from a statistical clustering analysis, such as a principal component analysis (PCA), of chromatograms obtained from a chromatographic analysis of sebum samples from the individual. This provides a method of personal identification based on the unique global profile of sebum composition for each individual.

[0072] Because of the complexity of sebum compositions, such personal identification system would be difficult to temper with, and more reliable.

[0073] Accordingly, one general aspect of the present invention relates to a method of identifying or authenticating a subject. The method comprises:

(a) obtaining a test chromatogram from a chromatographic analysis of a test sebum sample taken from the subject;

(b) analyzing the test chromatogram using a statistical clustering analysis algorithm to obtain a test vector of p variables, wherein p is less than the number of peaks in the test chromatogram; and

(c) comparing the test vector with a reference vector of p variables, wherein the reference vector is specific to a reference subject and is obtained from analyzing a reference chromatogram using the statistical clustering analysis algorithm, and the reference chromatogram is obtained from the chromatographic analysis of a reference sebum sample taken from the reference subject,

wherein a close proximity between the test vector and the reference vector identifies or authenticates the subject as the reference subject.

[0074] As used herein, the test vector and the reference vector are in "a close proximity" when the two vectors are in a close proximity in one or more of the p variables for each of the two vectors. Methods known in the art can be used to calculate or measure the proximity between the test vector and the reference vector in view of the present disclosure. For example, the close proximity between the test vector and the reference vector can comprise a close proximity between one or more of the p variable of the test vector and the reference vector. In one embodiment of the present invention, the test vector and the reference vector are in "a close proximity" when the two vectors are in a close proximity in at least the top two or more of the p variables.

[0075] The required close proximity between the test vector and the reference vector are determined based on parameters obtained from statistical analysis, reproducibility analysis, verification analysis, etc. The parameters can vary depending on factors such as the classes of subjects, e.g., classified by various factors such as animal species, races, genders, ages, etc., types of exposed drug, types of diseases, etc. The parameters can also vary depending on the different purposes. For example, different parameters may be required for personal identification, drug exposure detection, detection or monitoring of a disease, drug screening, etc.

[0076] In one embodiment, the test vector and the reference vector are "in a close proximity" when each of the p variables of the test vector is in a close proximity with each of the corresponding p variables of the reference vector. However, the test vector and the reference vector can be "in a close proximity" when less than all of each of the p variables of the test vector is in a close proximity with each of the corresponding p variables of the reference vector.

[0077] In one embodiment of the present invention, the analyzing step comprises:

(a) saving into a computer the test chromatogram;

(b) selecting significant peaks from the test chromatogram; and

(c) compressing the significant peaks selected in (b) into the test vector of p variables using the statistical clustering analysis algorithm, $1 \leq p \leq$ the number of significant peaks;

wherein the sub-steps of saving, selecting, compressing and the step of characterizing are accomplished with the computer.

5 **[0078]** In an embodiment of the present invention, the method utilizes a plurality of reference vectors obtained from statistical clustering analysis of a plurality of reference chromatograms, which are obtained from the chromatographic analysis of a plurality of reference sebum samples taken from a plurality of reference subjects.

10 **[0079]** In an embodiment of the present invention, the test vector and the plurality of vectors are obtained from the statistical clustering analysis, preferably PCA analysis, of selected significant peaks from each of the test chromatogram and the plurality of reference chromatograms, respectively. The significant peaks can be, but not required to be, shared by the test sebum sample and the one or more reference sebum samples, so long as the same criteria are used in selecting the significant peaks from each of the test chromatogram and reference chromatogram.
15 Information about the shared significant peaks is pre-stored into the computer, which is then used to assist the statistical clustering analysis of the test and/or reference chromatograms.

[0080] Preferably, the significant peaks are identified as relevant to the specificity of a vector to a subject, such as the specificity of a reference vector to a reference subject. Such information about the significant peaks can be based on analyses of multiple sebum samples from one subject,
20 analyses of multiple sebum samples from multiple subjects, and/or the combination of both. The information can be stored in the computer to assist the analysis according to embodiments of the present invention.

[0081] It is readily appreciated by those skilled in the art that there can be more than one vectors specific to a sebum sample, for example, depending on the methods of chromatographic analysis, and the selected significant peaks. Thus, it is necessary that the same analytical method
25 is used to obtain the test and reference chromatograms, and the same criteria are used in selecting the significant peaks from each of the test chromatogram and reference chromatogram.

[0082] In an embodiment of the present invention, to improve the reliability of the result, different test vectors obtained from the same test sebum sample using different analytical methods
30 and criteria can be compared with different reference vectors from the same reference serum using the corresponding analytical methods and criteria.

[0083] In an embodiment of the present invention, the subject and the reference subject can be different animal species.

35 **[0084]** In another embodiment of the present invention, the subject and the reference are the same animal species, preferably, human beings, more preferably, the same human subject.

[0085] In an embodiment of the present invention, the method further comprises the use of information input and data entry apparatus to minimize or assist searching of stored data, such as by use of a personal identification number or code, a photo, or other means.

[0086] In another embodiment of the present invention, the method further comprises the use of one or more other methods of personal identification or authentication, such as by fingerprint, written signature, retinal configuration, voice recognition, physical dimensions of the individual, features of the individual, etc.

[0087] Embodiments of the present invention also relate to systems and program products for personal identification.

[0088] In one embodiment, the present invention relates to a system for identifying or authenticating a subject, the system comprising a computer system capable of:

(a) storing a test chromatogram from a chromatographic analysis of a test sebum sample taken from the subject;

(b) analyzing the test chromatogram using a statistical clustering analysis algorithm to obtain a test vector of p variables, wherein p is less than the number of peaks in the test chromatogram; and

(c) identifying or authenticating the subject by comparing the test vector with a reference vector of p variables, wherein the reference vector is specific to a reference subject and is obtained from analyzing a reference chromatogram using the statistical clustering analysis algorithm, and the reference chromatogram is obtained from the chromatographic analysis of a reference sebum sample taken from the reference subject.

[0089] The system can further comprise an apparatus for chromatographic analysis of the test sebum sample to obtain the test chromatogram.

[0090] In another embodiment, the present invention relates to a program product for identifying or authenticating a subject. The program product comprises one or more computer programs stored on a recordable medium, and the one or more computer programs are capable of:

(a) saving into a computer a test chromatogram and a reference chromatogram each obtained from a chromatographic analysis of a sebum sample taken from the subject and a reference subject, respectively;

(b) selecting significant peaks from each of the test chromatogram and the reference chromatogram;

(c) compressing the significant peaks from each of the test chromatogram and the reference chromatogram into a test vector and a reference vector, respectively, using a statistical clustering analysis algorithm, wherein the reference vector is specific to the reference subject, and each of the vectors comprises p variables, $1 \leq p \leq$ the number of significant peaks; and

(e) identifying or authenticating the test subject by comparing the test vector with the reference vector.

[0091] Another embodiment of a program product for identifying or authenticating a subject comprises one or more computer programs stored on a recordable medium, and the one or more computer programs are capable of:

(a) saving into a computer a reference vector of p variables, wherein the reference vector is specific to a reference subject, and is obtained from a statistical clustering analysis of significant peaks from a reference chromatogram obtained from a chromatographic analysis of a sebum sample taken from the reference subject, $1 \leq p \leq$ the number of significant peaks;

(b) saving into the computer a test chromatogram obtained from the chromatographic analysis of a sebum sample taken from the subject;

(c) selecting the significant peaks from the test chromatogram;

(d) compressing the significant peaks from the test chromatogram into a test vector of p variables using the statistical clustering analysis algorithm; and

(e) identifying or authenticating the subject by comparing the test vector with the reference vector.

[0092] In a preferred embodiment of the present invention, the statistical clustering analysis algorithm is a PCA algorithm as that described herein.

[0093] Preferably, the reference sebum sample and the test sebum sample are taken from the same body location on the reference subject and the test subject, which can be the same or different subjects.

[0094] The reference vector can be obtained from a stored reference chromatogram. The test chromatogram and the reference chromatogram can be stored and analyzed at the same or different time points to obtain the reference and test vectors, respectively. For example, one or more reference chromatograms can be pre-stored in the computer. Preferably, the test chromatogram and the reference chromatogram are aligned during the analysis.

[0095] It is readily appreciated by those skilled in the art that, instead of saving and analyzing a reference chromatogram, a reference vector of p variables can be directly stored and used for comparison with the test vector.

[0096] This invention will be better understood by reference to the non-limiting example that follows, but those skilled in the art will readily appreciate that the example is only illustrative of the invention as described more fully in the claims which follow thereafter.

EXAMPLE

Characterization of Sebum Composition by PCA

[0097] An absorbent tape was used for the *in vivo* sampling of sebum samples from the skin of human subjects. Sebum samples were collected from 5 human subjects, male, 18-60 year old, at 11 visits. At each visit, the middle portion of the forehead of the subject was first treated with ethyl alcohol. Three sebum samples were collected using an absorbent tape from the forehead 30 minutes after treatment. The samples were stored at below -60°C until sample extraction and analysis. The total amount of sebum was more than 140µg/cm² from each collection.

[0098] Sebum enriched sebutapes were extracted in 3 mL HPLC grade heptane by vortexing for 1 minute in 5 mL flask. The organic solvent was transferred to a glass vial, evaporated under nitrogen (about 30 min) at room temperature. The samples were reconstituted with 70µL of methanol (50 volumes) / dichloromethane (50 volumes) mix and then transferred to an injection vial.

[0099] The sebum samples were subject to chromatographic analysis using a gas chromatography/mass spectrometry (GC/MS) system.

[00100] In this Example, 47 chromatograms resulting from the GC/MS analysis of sebum samples from 5 human subjects at 11 visits were obtained. Each of the chromatograms contains more than 70 peaks. The relative intensity or strength of the peaks, which corresponds to the relative amount of respective constituents in the analyzed sebum samples, often varied from one chromatogram to another. The variations were observed among chromatograms obtained from different subjects (Fig. 1A) or from the same subject at different visits (Fig. 1B). In particular, the measurements of fatty acids posed a problem, because of their instability and the difficulties to analyzed them. Thus, it was difficult to characterize and compare different sebum samples based on direct comparison of all peaks in chromatograms.

[00101] Baseline regression and subtraction were performed to correct each of the chromatograms (Fig. 2). The corresponding peaks in the chromatograms were aligned. Significant peaks were selected from the baseline corrected and aligned chromatograms (Figs 3A and 3B).

[00102] In particular, functions in the Bioinformatic Toolbox of MATLAB were used for data processing.

[00103] 1. Resampling the input signal (function msresample)

[00104] This step was used for resampling the raw mass chromatogram. The output chromatogram can have $N=10000$ samples with a spacing that increases linearly within the range of the raw signal.

[00105] 2. Peaks alignment method (function msalign)

[00106] The alignment was performed by scaling and shifting the domain mass/charge (MZ) such that the cross-correlation between the input chromatogram and a synthetic target signal was maximum. The synthetic target signal was built with Gaussian pulses centered at the locations specified by the user defined vector P. After obtaining the new MZ mass-charge domain,

MSALIGN calculated the realigned signal by shape-preserving piecewise cubic interpolation of the shifted input signal to the original MZ vector.

[00107] 3. Baseline correction (function msbackadj)

[00108] The variable background (baseline) of a mass spectrometry signal was adjusted by following steps: estimating the background within multiple shifted windows of width 200 (MZ scale); regressing the varying baseline to the window points using a spline approximation; and removing the background of the input signal.

[00109] 4. Peak extraction (function mspeak)

[00110] Peaks in each chromatogram were selected by following steps: smoothing the chromatogram signals using the undecimated wavelet transform with the Daubechies filter banks; assigning peak locations; and eliminating peaks that do not satisfy specific criteria.

[00111] Peaks in the chromatograms were localized or identified by their corresponding retention times (Fig. 4). PCA was performed to reduce the dimensionality of the data space for each chromatogram.

[00112] In particular, PCA was performed with a chromatogram having m dimensions, m = number of selected significant peaks. An $m \times m$ covariance matrix of the selected significant peaks was provided. The eigenvectors and eigenvalues of the covariance matrix were calculated. The eigenvalues and corresponding eigenvectors were ranked in a decreasing order. The 3 eigenvectors corresponding to the top 3 eigenvalues were chosen as the three principal components C1, C2 and C3. The dimensionality of the chromatogram was thus further reduced from m to three, which best explain the variance in the multi-dimensional data space of the chromatogram.

[00113] In this Example, the over 70 peaks in each of the 47 chromatograms were first reduced to 16 significant peaks, which were further reduced to 3 variables, i.e., three principal components, by PCA. The cumulative value of the top three eigenvalues is 90.9% of the sum of total eigenvalues. Thus, the three variables, C1, C2 and C3 well accounted for the variability in the multi-dimensional data space of the chromatogram and best explain the variance in the over 70 dimension data space. However, it is readily apparent to those skilled in the art that numbers of eigenvectors different from 3 can also be used in the PCA.

[00114] As shown in Fig. 5, PCA reduced each chromatogram into a single dot in a three-dimensional space of the three variables, C1, C2 and C3. The PCA profiles of multiple chromatograms measured from an individual, such as 1-1, 1-2 ...and 1-10 from subject 1, clustered in a region. This region is distinct or separate from that of another individual, such as the region containing 3-1, 3-2... and 3-8 from subject 3. This indicates that an individual has a relatively reproducible and distinct global profile of sebum compositions, as analyzed by PCA on chromatograms of chromatographic analyses of sebum samples from the individual.

[00115] It will be appreciated by those skilled in the art that changes could be made to the embodiments described above without departing from the broad inventive concept thereof. It is understood, therefore, that this invention is not limited to the particular embodiments disclosed, but it is intended to cover modifications within the spirit and scope of the present invention as defined

5 by the appended claims.

CLAIMS

1. A method of identifying or authenticating a subject, the method comprising:

(a) obtaining a test chromatogram from a chromatographic analysis of a test sebum
5 sample taken from the subject;

(b) analyzing the test chromatogram using a statistical clustering analysis algorithm to
obtain a test vector of p variables, wherein p is less than the number of peaks in the test
chromatogram; and

(c) comparing the test vector with a reference vector of p variables, wherein the
10 reference vector is specific to a reference subject and is obtained from analyzing a reference
chromatogram using the statistical clustering analysis algorithm, and the reference chromatogram
is obtained from the chromatographic analysis of a reference sebum sample taken from the
reference subject,

wherein a close proximity between the test vector and the reference vector identifies or
15 authenticates the subject as the reference subject.

2. The method of claim 1, wherein the analyzing step comprising:

(a) saving into a computer the test chromatogram;

(b) selecting significant peaks from the test chromatogram; and

(c) compressing the significant peaks selected in (b) into the test vector of p variables
20 using the statistical clustering analysis algorithm, $1 \leq p \leq$ the number of significant peaks;

wherein the sub-steps of saving, selecting, compressing and the step of comparing
are accomplished with the computer.

3. The method of claim 2, wherein the significant peaks are identified as relevant to the
specificity of the reference vector to the reference subject.

4. The method of claim 2, wherein the statistical clustering analysis algorithm is a
principal component analysis that performs a computer assisted process comprising:

(a) providing a covariance matrix of the significant peaks;

(b) calculating the eigenvectors and eigenvalues of the covariance matrix;

(c) ranking the eigenvalues and corresponding eigenvectors in a decreasing order; and

(d) compressing the significant peaks into a vector of p eigenvectors corresponding to
30 the top p eigenvalues.

5. The method of claim 4, wherein the cumulative value of the top p eigenvalues is at
least 70% of the sum of total eigenvalues.

6. The method of claim 4, wherein $p = 3$.

7. The method of claim 2, further comprising:

- (a) saving into the computer the reference chromatogram;
- (b) selecting significant peaks from the reference chromatogram, and
- (c) compressing the significant peaks selected in (b) into the reference vector of p variables using the statistical clustering analysis algorithm, $1 \leq p \leq$ the number of significant peaks;

5 and

wherein the sub-steps of saving, selecting, compressing and the step of comparing are accomplished with the computer.

8. The method of claim 7, wherein the significant peaks are identified as relevant to the specificity of the reference vector to the reference subject.

10 9. The method of claim 2, further comprising storing the reference vector of p variables into the computer, wherein the identifying or authenticating step comprises comparing the test vector with the stored reference vector.

10. The method of claim 1, further comprising correcting the test chromatogram by baseline subtraction prior to the step of analyzing.

15 11. The method of claim 1, wherein the subject is a human.

12. The method of claim 1, wherein the test sebum sample and the reference sebum sample are taken from the same body location from the subject and the reference subject, respectively.

20 13. The method of claim 12, wherein the body location is selected from the group consisting of the forehead, nose, cheek, shoulder, neck and top of the back of each of the subject and the reference subject, respectively.

25 14. The method of claim 1, wherein the chromatographic analysis is selected from the group consisting of gas chromatography (GC), gas chromatography-mass spectrometry (GC-MS), high-performance liquid chromatography (HPLC), high-performance liquid chromatography-mass spectrometry (HPLC-MS), liquid chromatography (LC), liquid chromatography-mass spectrometry (LC-MS), thin layer chromatography (TLC), and thin layer chromatography-mass spectrometry (TLC-MS).

30 15. The method of claim 1, wherein the step of comparing comprises comparing the test vector with a plurality of reference vectors of p variables, wherein each of the plurality of reference vectors is specific to each of a plurality of reference subjects.

16. A system for identifying or authenticating a subject, the system comprising a computer system capable of:

- (a) storing a test chromatogram from a chromatographic analysis of a test sebum sample taken from the subject;

(b) analyzing the test chromatogram using a statistical clustering analysis algorithm to obtain a test vector of p variables, wherein p is less than the number of peaks in the test chromatogram; and

(c) identifying or authenticating the subject by comparing the test vector with a reference vector of p variables, wherein the reference vector is specific to a reference subject and is obtained from analyzing a reference chromatogram using the statistical clustering analysis algorithm, and the reference chromatogram is obtained from the chromatographic analysis of a reference sebum sample taken from the reference subject.

17. The system of claim 16 further comprising an apparatus for chromatographic analysis of the test sebum sample to obtain the test chromatogram.

18. A program product for identifying or authenticating a subject, the program product comprising one or more computer programs stored on a recordable medium, the one or more computer programs being capable of:

(a) saving into a computer a test chromatogram and a reference chromatogram each obtained from a chromatographic analysis of a sebum sample taken from the subject and a reference subject, respectively;

(b) selecting significant peaks from each of the test chromatogram and the reference chromatogram;

(c) compressing the significant peaks from each of the test chromatogram and the reference chromatogram into a test vector and a reference vector, respectively, using a statistical clustering analysis algorithm, wherein the reference vector is specific to the reference subject, and each of the vectors comprises p variables, $1 \leq p \leq$ the number of significant peaks; and

(e) identifying or authenticating the test subject by comparing the test vector with the reference vector.

19. A program product for identifying or authenticating a subject, the program product comprising one or more computer programs stored on a recordable medium, the one or more computer programs being capable of:

(a) saving into a computer a reference vector of p variables, wherein the reference vector is specific to a reference subject, and is obtained from a statistical clustering analysis of significant peaks from a reference chromatogram obtained from a chromatographic analysis of a sebum sample taken from the reference subject, $1 \leq p \leq$ the number of significant peaks;

(b) saving into the computer a test chromatogram obtained from the chromatographic analysis of a sebum sample taken from the subject;

(c) selecting the significant peaks from the test chromatogram;

(d) compressing the significant peaks from the test chromatogram into a test vector of p variables using the statistical clustering analysis algorithm; and

(e) identifying or authenticating the subject by comparing the test vector with the reference vector.

Figure 1A

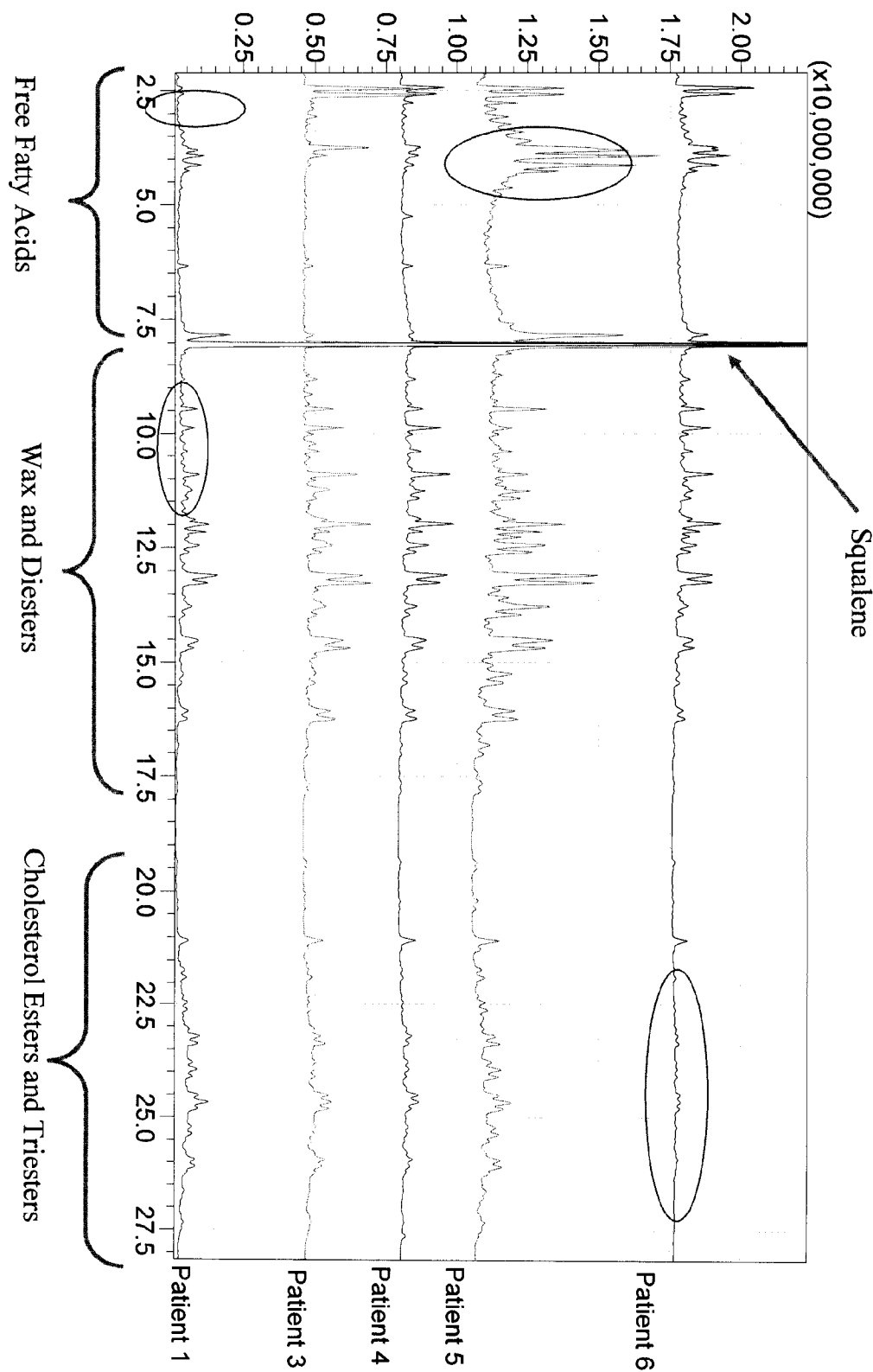


Figure 1B

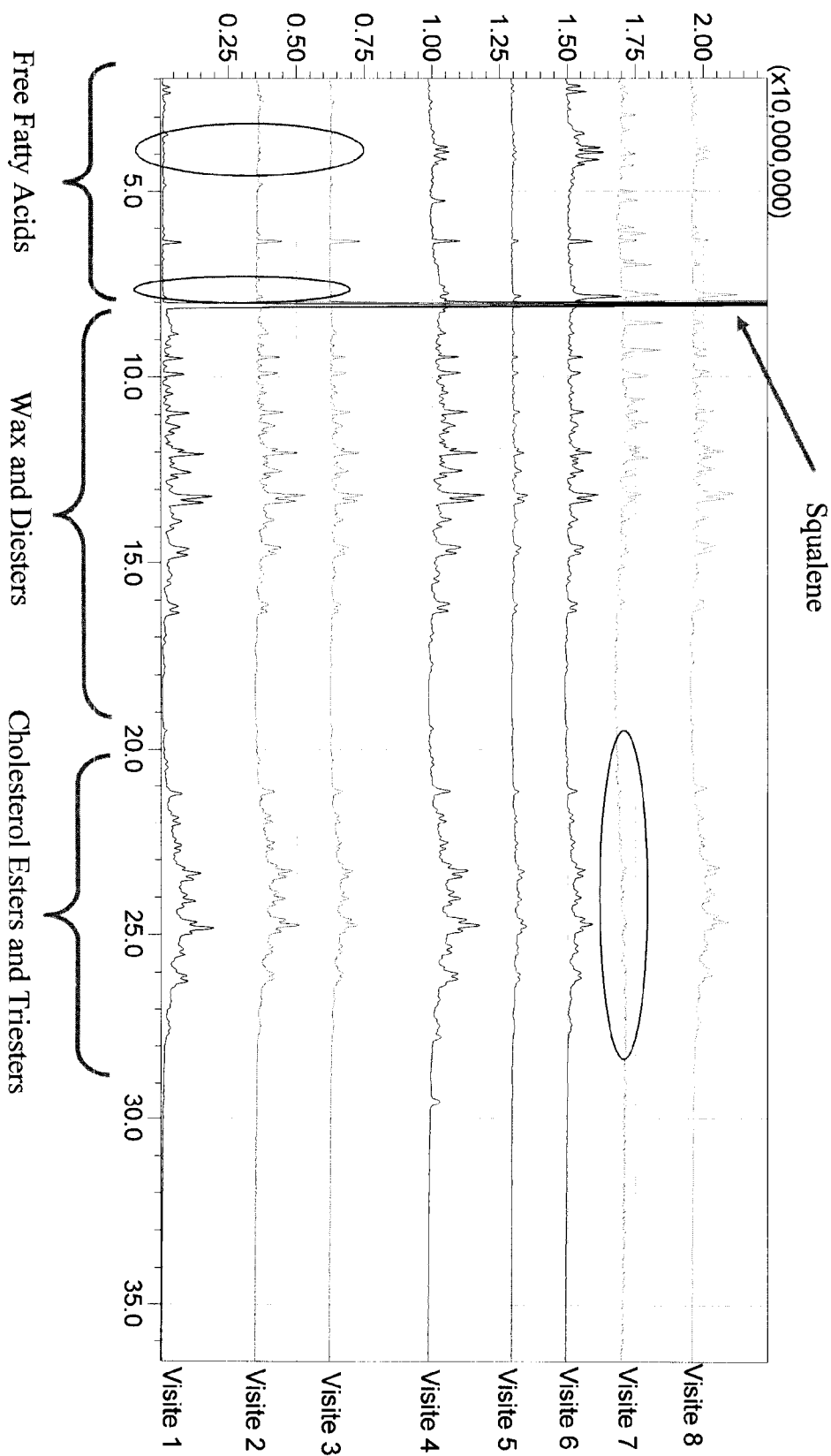


Figure 2

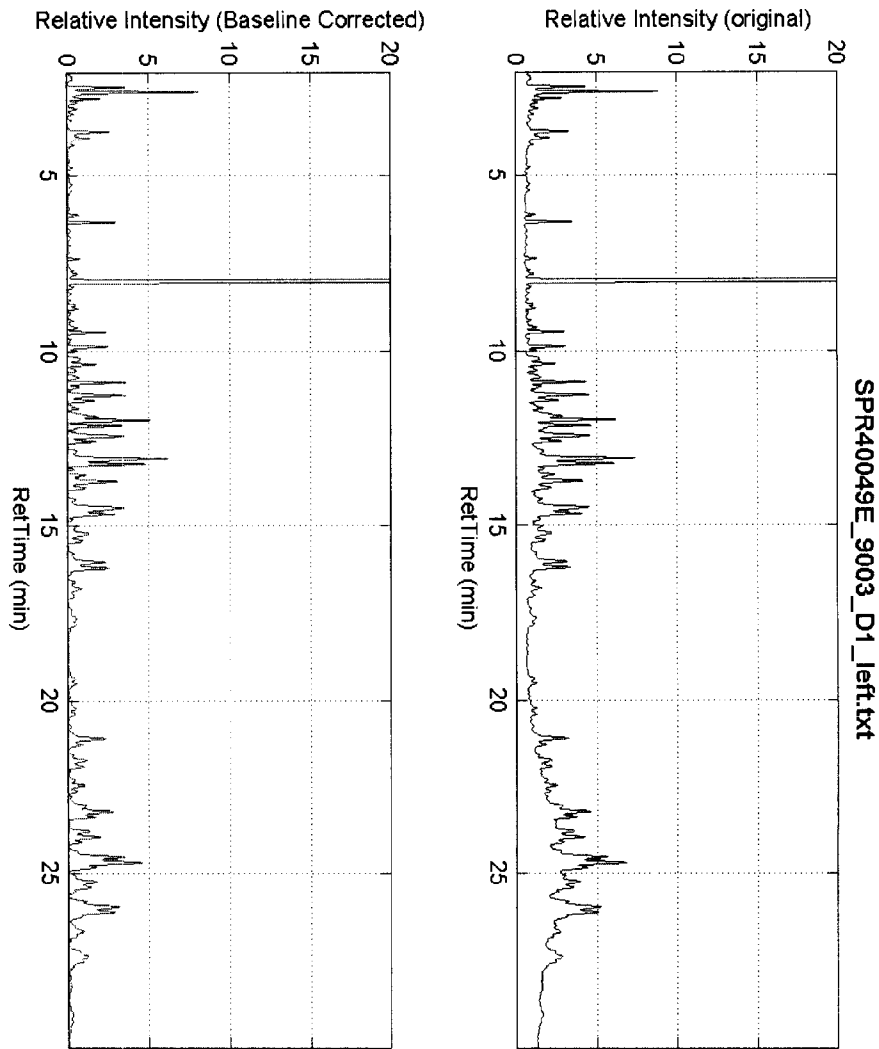


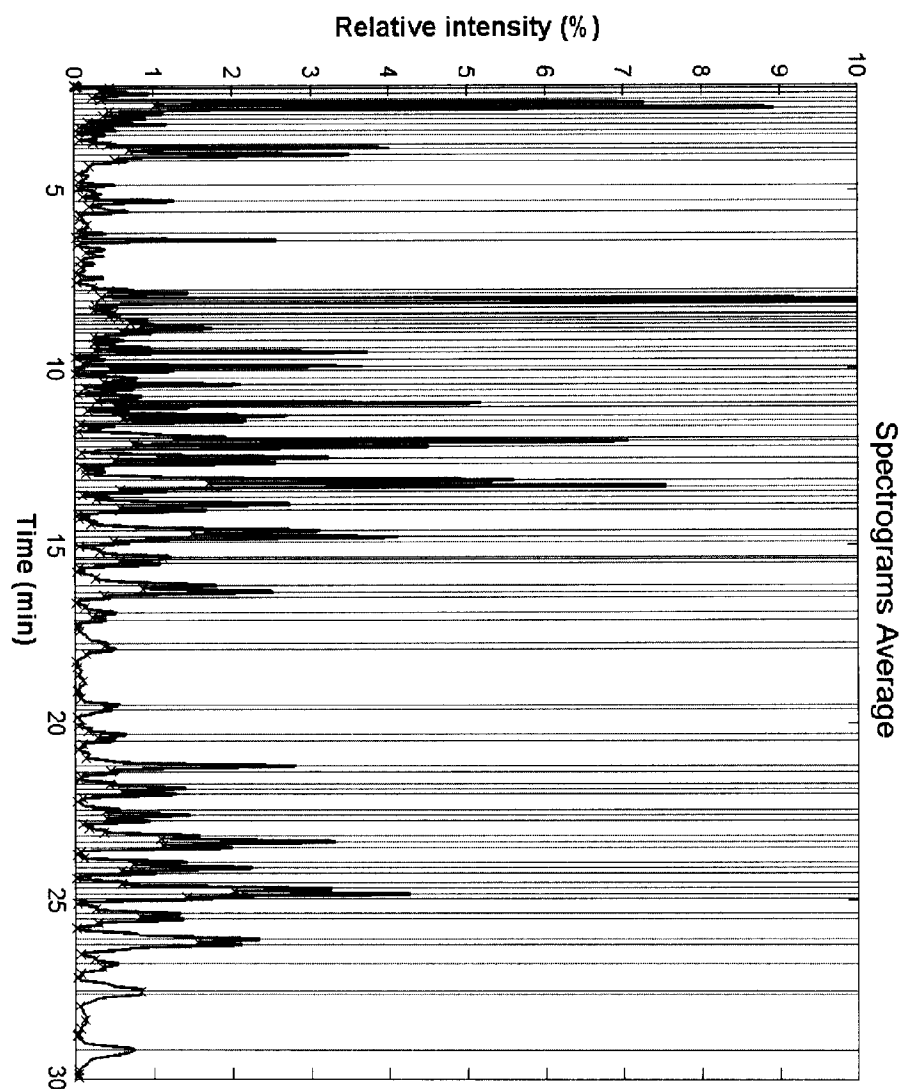
Figure 3A

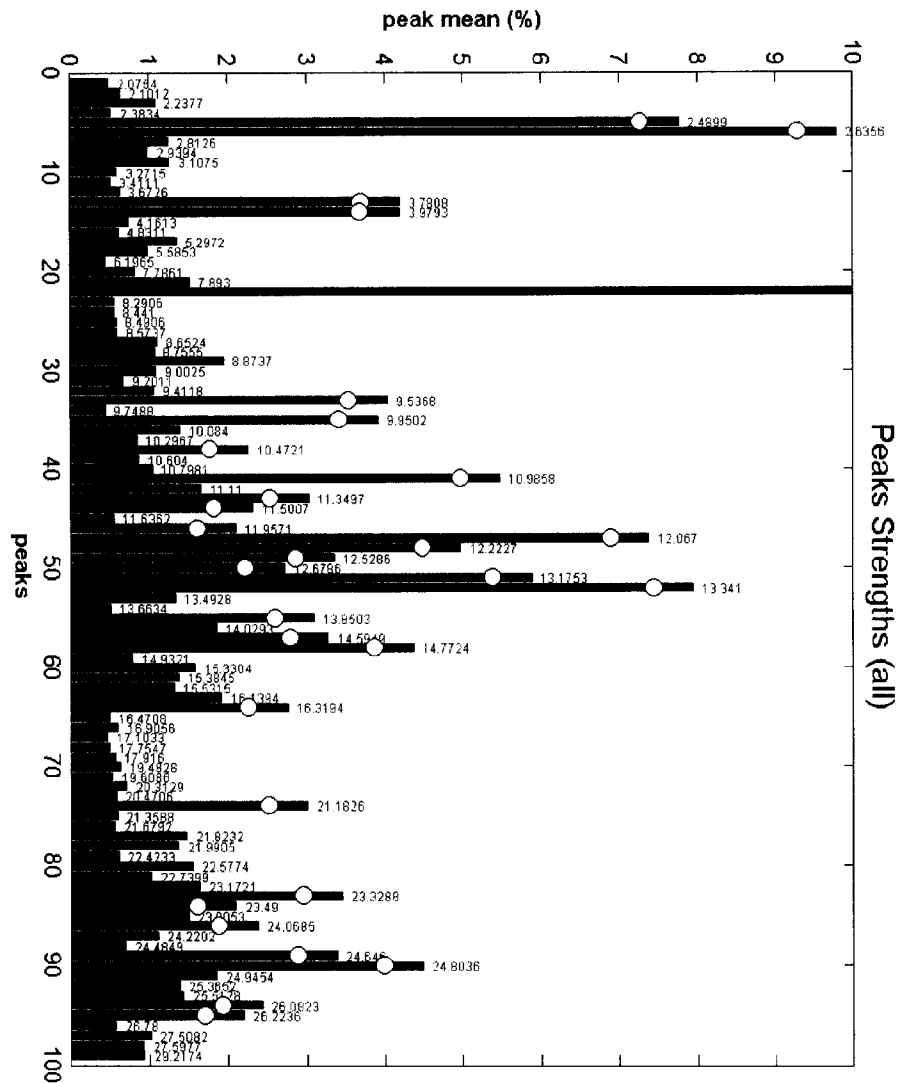
Figure 3B

Figure 4

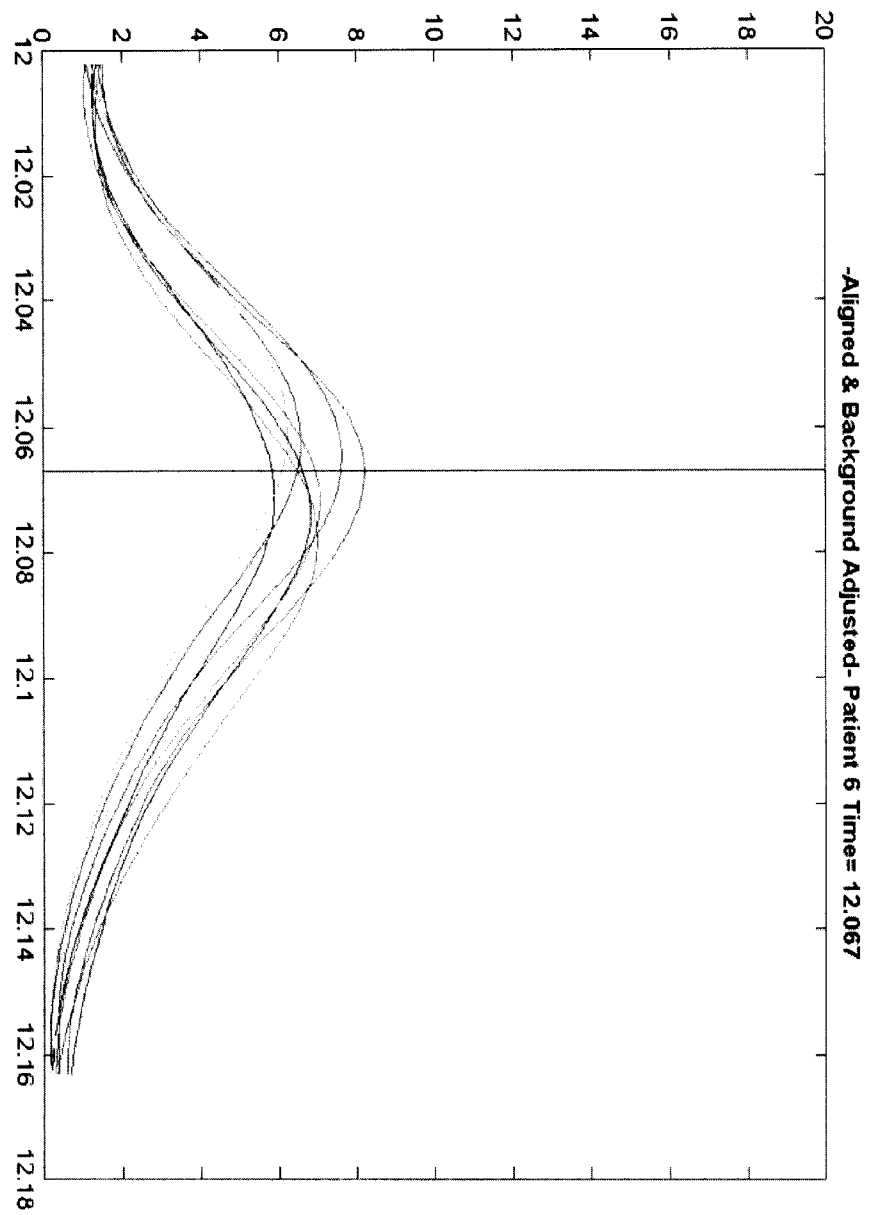
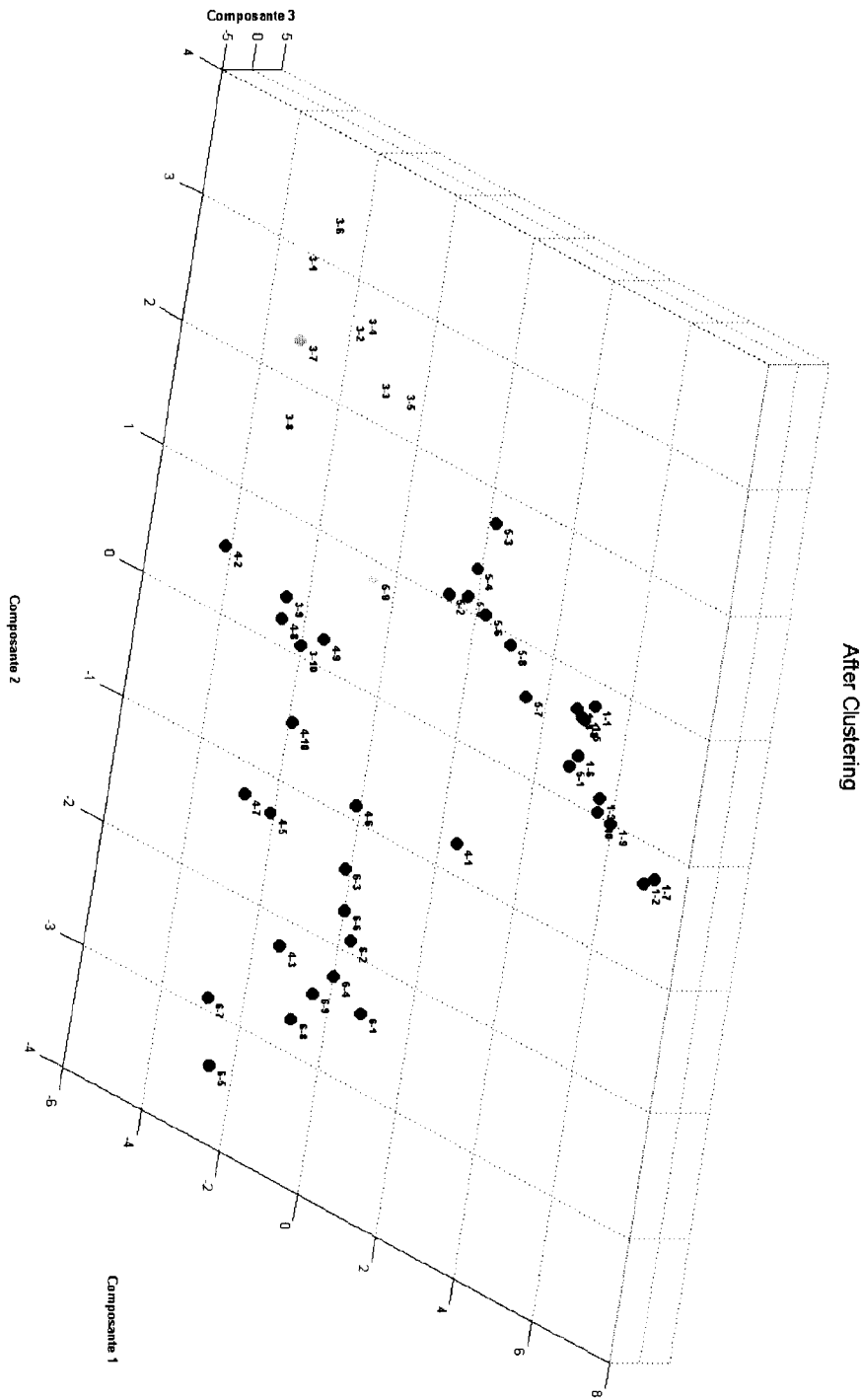


Figure S



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/058596

A. CLASSIFICATION OF SUBJECT MATTER
INV. G01N33/68 G06K9/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G01N G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, COMPENDEX, EMBASE, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	APRIL HEMMILA ET AL: "Fourier Transform Infrared Reflectance Spectra of Latent Fingerprints: A Biometric Gauge for the Age of an Individual", JOURNAL OF FORENSIC SCIENCES, vol. 53, no. 2, 1 March 2008 (2008-03-01), pages 369-376, XP55003944, ISSN: 0022-1198, DOI: 10.1111/j.1556-4029.2007.00649.x the whole document in particular page 369, column 1, paragraph 2 page 370, column 1, paragraph 2 first and second full paragraph; page 376, column 1 table 1 ----- -/--	1-15

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

1 August 2011

Date of mailing of the international search report

16/08/2011

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Thumb, Werner

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/058596

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PAUL J TRIM ET AL: "Matrix-assisted laser desorption/ionisation mass spectrometry imaging of lipids in rat brain tissue with integrated unsupervised and supervised multivariant statistical analysis", RAPID COMMUNICATIONS IN MASS SPECTROMETRY, HEYDEN, LONDON, GB, vol. 22, no. 10, 30 May 2008 (2008-05-30), pages 1503-1509, XP002587368, ISSN: 0951-4198, DOI: DOI:10.1002/RCM.3498 [retrieved on 2008-04-17]	16-19
Y	abstract page 1505, column 2, paragraph 2 - page 1506, column 1, paragraph 3	1-15
X	YI ET AL: "Evaluation of the antimicrobial mode of berberine by LC/ESI-MS combined with principal component analysis", JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS, NEW YORK, NY, US, vol. 44, no. 1, 13 April 2007 (2007-04-13), pages 301-304, XP022026246, ISSN: 0731-7085, DOI: DOI:10.1016/J.JPBA.2007.02.018	16-19
Y	abstract section 2.3	1-15
X	TIM P. SANGSTER ET AL: "Investigation of analytical variation in metabonomic analysis using liquid chromatography/mass spectrometry", RAPID COMMUNICATIONS IN MASS SPECTROMETRY, vol. 21, no. 18, 30 September 2007 (2007-09-30), pages 2965-2970, XP55003846, ISSN: 0951-4198, DOI: 10.1002/rcm.3164	16-19
Y	the whole document page 2966, column 1, paragraph 3	1-15
Y	WO 2006/092689 A1 (WARNER LAMBERT CO [US]; KINDT KRISTIAN ERICH [US]; SZEKELY-KLEPSE GAB) 8 September 2006 (2006-09-08) abstract page 5, line 1 - line 25 page 8, line 7 - line 20 page 9, line 3 - line 16 page 14, line 11 - line 29 examples 1,2	1-15

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/058596

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SCOTT C. GREEN ET AL: "Variation in Sebum Fatty Acid Composition Among Adult Humans.", JOURNAL OF INVESTIGATIVE DERMATOLOGY, vol. 83, no. 2, 1 August 1984 (1984-08-01) , pages 114-117, XP55003870, ISSN: 0022-202X, DOI: 10.1111/1523-1747.ep12263287 the whole document	1-15
Y	----- DE 101 64 553 A1 (UNIV SCHILLER JENA [DE]) 3 July 2003 (2003-07-03) the whole document	1-15
A	----- Lindsay Smith: "A tutorial on principal component analysis", ' 26 February 2002 (2002-02-26), XP002654664, Retrieved from the Internet: URL:http://www.cs.otago.ac.nz/cosc453/stud ent_tutorials/principal_components.pdf [retrieved on 2001-07-29] the whole document -----	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2011/058596

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2006092689	A1	08-09-2006	NONE
DE 10164553	A1	03-07-2003	NONE