



(51) International Patent Classification:
A61B 5/103 (2006.01)

(21) International Application Number:
PCT/CN2014/071853

(22) International Filing Date:
30 January 2014 (30.01.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/759,808 1 February 2013 (01.02.2013) US
61/793,532 15 March 2013 (15.03.2013) US

(71) Applicant: THE CHINESE UNIVERSITY OF HONG KONG [CN/CN]; Shatin, N.T., Hong Kong (CN).

(72) Inventor: ZEE, Benny Chung-Ying; Room 301 Pi Ch'iu Building, Shatin, N.T., Hong Kong (CN).

(74) Agent: INSIGHT INTELLECTUAL PROPERTY LIMITED; Suite 19A, Tower A, InDo Building, No. 48A Zhichun Road, Haidian, Beijing 100098 (CN).

(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, BN, 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, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG).

Published:

— with declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority

(54) Title: SYSTEMS AND METHODS USING EXHALED BREATH FOR MEDICAL DIAGNOSTICS AND TREATMENT

(57) Abstract:



WO 2014/117747 A2

SYSTEMS AND METHODS USING EXHALED BREATH FOR MEDICAL DIAGNOSTICS AND TREATMENT

FIELD

[0001] The present application relates to a system and method using exhaled breath for diagnosis and prognosis of various medical conditions and diseases. Exhaled breath samples are analyzed by selected ion flow tube mass spectrometry (SIFT-MS) and the full mass spectrum is analyzed using a penalized regression model to select masses highly correlated with a disease state. The selected masses are then analyzed by regression methods for diagnosis and treatment.

BACKGROUND

[0002] Exhaled breath has been long envisioned as an ideal translational biomarker for disease association primarily because breath samples are easy to obtain, safe, and non-invasive. Compounds in exhaled breath reflect the status of metabolic process of the human body and can be regarded as a fingerprint of the health status of an individual. In the past, metabolites of interest such as acetone, isoprene, and others represent only a small portion of the metabolites in exhaled gas (trace gases) and were difficult to quantify. Recent advances in technology have enabled quick and accurate identification and quantification of breath components through gas mass spectrometry. As many as 300 compounds in exhaled breath can be quantified in a minute or so using selected ion flow tube mass spectrometry (SIFT-MS). To date, only a limited number of clinical studies have been conducted using mass spectrometry, although previous studies have established a correlation of certain single exhaled breath species to diabetes, kidney function, and lung cancer. Thus, previous exhaled breath studies have focused on only one or two gases, and very few studies have used the

full scan mass data in the analysis. However, it is clear that using single species for clinical diagnostics has several limitations.

[0003] Selected ion flow tube mass spectrometry (SIFT-MS) technology has made possible fast, accurate, and stable mass spectrometry analysis of exhaled breath [1]. Previous breath research has indicated strong correlation between exhaled breath compounds and a wide range of diseases, such as diabetes [2,3,4], cancer [5,6], hepatic disease [7], renal failure [8,9], etc. [1,10]. The rationale behind the association is that excreted gas from metabolic processes is eventually expelled in breath. For example, breath acetone is mostly derived from fatty acid degradation and is an important indicator of diabetes; and the thiol compounds such as methanethiol and ethanethiol that increase in cirrhosis are mostly due to emission of intestinal bacteria [9]. Evidence suggests that breath compounds can be used as fingerprints of human health status, and therefore breath analysis can be a useful part of routine medical diagnosis because the method is convenient and non-invasive. Meanwhile on technology realization side, SIFT-MS provides real-time profiling of gases including in exhaled breath with precision down to parts by billion (ppb) [1]. The apparatus provides two modes: the multiple-ion-monitoring (MIM) mode, which has been used for most current studies depended [2,11]; and the full-scan (FS) mode, which enables quantification of nearly 200 identified metabolites in a single analysis, and provides information about additional as yet unidentified metabolites for possible association analysis.

[0004] Although the full-scan mode clearly provides much richer information, few studies have been done on the data it generates, at least partly due to the early stage of the medical application of SIFT-MS and partly because of the difficulty in the problem, i.e., the number of full-scanned masses is much larger than the size of study cohort.

SUMMARY

[0005] To fully utilize the complex information provided by SIFT-MS, we provide a two-layer statistical analysis method incorporating a penalized regression (lasso) method [12] to identify and select masses that are highly correlated with a disease state, and subsequently analyze the selected masses by regression methods for clinical diagnosis and treatment. The results demonstrate, for example, that key indicators of dialysis efficacy, such as blood level creatinine and urea, can be accurately predicted based the statistical analysis of multiple compounds in exhaled breath as measured using SIFT-MS.

[0006] In a first aspect, methods of diagnosing a disease in a subject are disclosed, the methods comprising providing a sample obtained from the exhaled breath of a subject; obtaining at least one mass spectrum of the sample using selected ion flow tube mass spectrometry (SIFT-MS); analyzing the at least one mass spectrum using a multivariate statistical method; and diagnosing the presence or absence of the disease in the subject based on results of the analysis.

[0007] In a second aspect, methods of diagnosing diabetes in a subject, the method comprising providing a sample obtained from the exhaled breath of a subject; obtaining at least one mass spectrum of the sample using selected ion flow tube mass spectrometry (SIFT-MS), wherein obtaining at least one mass spectrum comprises obtaining at least one mass spectrum comprises obtaining a first mass spectrum using NO^+ as the precursor ion, and obtaining a second mass spectrum using H_3O^+ as the precursor ion; analyzing the at least one mass spectrum using a multivariate statistical method, wherein using the multivariate statistical method comprises analyzing the selected masses using a linear regression method, wherein, the selected masses for the first mass spectrum comprise m/z 150, m/z 152, m/z 173, m/z 29, m/z 52, m/z 14=26; and the selected masses for the second mass spectrum comprise m/z 156, m/z 39, and m/z 108; and diagnosing the presence or absence of diabetes in the subject based on results of the analysis.

[0008] In a third aspect, methods of treating a disease in a patient are provided,

comprising determining that the patient has the disease using the method of claim 1; and administering a treatment to the patient for treating the disease.

[0009] In a fourth aspect, methods of treating a disease in a patient are provided, comprising requesting a test according to the method of claim 1 to determine whether a patient has the disease; and administering a treatment to the patient to treat the disease if the patient is diagnosed as having the disease.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The drawings described herein are for illustration purposes only. The drawings are not intended to limit the scope of the present disclosure.

[0011] **FIG. 1** shows a schematic diagram of a SIFT-MS apparatus.

[0012] **FIG. 2** shows a schematic of a cross-validation process used to assess the efficacy of the statistical model.

[0013] **FIGS. 3A-3D** show the correlation between ammonia, acetone, TMA, and creatinine and the predicted values of creatinine based on multivariate analysis of SIFT-MS breath samples using H_3O^+ as the precursor ion.

[0014] **FIG. 4** shows SIFT mass spectra of a breath sample using independently, H_3O^+ , NO^+ , or O_2^+ as the ion precursor.

[001] **FIGS. 5A-5C** are scatter plots showing the relationship between breath acetone and other clinical parameters body mass index (BMI), plasma glucose levels, and HbA1c.

[0015] **FIGS. 6A-6D** are scatter plots showing the correlation between breath acetone and the amplitude at various amu in the SIFT-MS spectrum using H_3O^+ as the precursor ion. Correlations for 57 amu, 77 amu, 142 amu, and 165 amu, are provided.

[0016] **FIG. 7** show correlations between a predicted creatinine value based on the statistical analysis of a full SIFT-MS mass spectrum using H_3O^+ as the precursor ion, and the measured blood creatinine value.

[0017] FIG. 8 show comparisons of the plasma urea value and predicted urea value based on multivariate analysis of SIFT-MS breath using a H_3O^+ precursor ion.

[0018] Reference is now made to certain embodiments of systems and methods using exhaled breath for medical diagnostics. The disclosed embodiments are not intended to be limiting of the claims. To the contrary, the claims are intended to cover all alternatives, modifications, and equivalents.

DETAILED DESCRIPTION

[0019] Many studies give support for the century old notion that exhaled breath metabolites serve as fingerprints of health status of people. The rationale behind the association of breath compounds and health status is that normal and abnormal gastrointestinal processes excrete gases with distinct profiles [4, 27]. For example, Rooth and Ostenson (1966) [28] indicated that breath acetone reflects the concentration of arterial acetone, which is a direct indication of deficiency of carbohydrates in muscle and adipose tissue and thus diabetes status. Breath ammonia is mostly derived from deamination of amino acid reflects problems in kidney function [4, 29]. Isoprene produced from body cholesterol synthesis and expelled in breath is found to be significantly elevated in patients with end-stage renal failure [30].

[0020] Diagnostic tests based on breath samples are non-invasive and potentially inexpensive methods as compared to blood or urine tests and thereby provide an opportunity for rapid diagnosis of diseases such as lung cancer [1, 20], breast cancer [2], pulmonary tuberculosis [3], diabetes [4], cirrhosis [4]. In a human breath, approximately 200 volatile organic compounds can be detected and only a few compounds and their concentration variations were found to be correlated to those illnesses. For example, breath concentration of acetone was shown to be significantly correlated to diabetes [4, 5]. The test also assists in recognition of early-stage diabetes or pre-diabetes. A study showed that a marked differences in glucose-derived breath CO_2 kinetics between individuals with normal

and impaired glucose tolerance within 60 minutes [6]. The findings support that breath test screening is a potential tool to assess the risk of diabetes and the associated complications at an early stage but needs further development to increase the practicality of the approach.

[0021] To accurately measure compounds in exhaled breath, the traces gases need to first be identified and then quantified. Quantitative measurement of breath began in the 1960's, and technology development in recent years has made measurements more and more accurate and reliable. Current technologies may be divided into three main approaches:

- (1) Gas chromatography (GC);
- (2) Proton transfer reaction mass spectrometry (PTR-MS); and
- (3) Selected ion flow tube mass spectrometry (SIFT-MS).

[0022] GC is a very common method and includes many techniques, however all GC approaches need to perform a complicated pre-concentration step such as chemical, cryogenic or adsorptive treatment, and many organic compounds can be lost during the process. The other two categories, PTR-MS and SIFT-MS, are more recent and share similar principles for detecting volatile organic compounds (VOCs) using chemical ionization. PTR-MS [35] is developed by Hansel, Jordan and Lindinger et al (1995) [31] and uses a carrier gas H_3O^+ to perform proton transfer with the reactant compounds.

[0023] SIFT-MS [32] uses three types of carrier gases that perform proton transfer and charge transfer to count the product ions. Both methods possess the advantages of being free from a pre-concentration step, are free from interference by the nitrogen, oxygen, carbon-dioxide and water vapor in exhaled breath, and exhibit high sensitivity down to part per billion (ppb). Measurements can also be carried out frequently and rapidly. A major difference between the two methods is that PTR uses single carrier ions to capture the VOCs, thereby forming isomer product ions, that can be indistinguished by mass-to-charge ratio alone and require additional measurement steps to resolve. SIFT-MS employs three types of precursor ions and is able to recover the reactant ions by crosschecking because a

particular compound will form different products using different precursor ions. Thus, SIFT-MS offers a more complete mass profile and is more convenient for clinical application. Incorporating multiple volatile organic compounds (VOCs) in the analysis can give much better prediction of disease status and therefore SIFT-MS analysis of exhaled breath represents a very promising method for clinical application.

[0024] A working scheme of a SIFT-MS system is shown in **FIG. 1**. An ion source creates positive precursor ions and injects the ions into a helium carrier gas, which convects along a flow-tube as the cold ions reach the temperature of the carrier gas. The gas to be analyzed is introduced into the carrier gas via an entry port into the flow-tube in a controlled way. At the downstream end of the flow tube, gases are sampled into a quadrupole mass spectrometer, where mass analysis is performed and ions are counted. The count rate is directly proportional to the concentration of the trace gases, and the counting precision ranges from 10 parts per billion (ppb) to 10 parts per million (ppm). [1] Breath samples from a subject can be collected, for example, through a mouth piece connected directly to the flow-tube, or collected in a Tedlar bag for subsequent analysis. One of the key advantages of SIFT-MS is that sample collection is non-invasive and convenient such that even patients in very weak condition can easily provide breath for analysis.

[0025] The precursor ions are chosen such that the ions only react with trace gases and not with major components in the air such as oxygen, water, and nitrogen gas. Suitable precursor ions for SIFT-MS include H_3O^+ , NO^+ , and O_2^+ . H_3O^+ reacts with most organic compounds through proton transfer: a trace gas (M) reacts with H_3O^+ to produce H_2O and a positive trace gas (MH^+) ion with the mass-to-charge ratio increased by one. NO^+ reacts with most organic compounds through several mechanisms, including charge transfer (producing M^+), hydride ion transfer (MH^+), and hydroxide ion transfer (MOH^+). Some of these reactions may occur in parallel. When isotopic compounds occur in one of the precursor reactions, the other precursor ions provide an important way to cross check and

identify the target compounds. For example, O_2^+ reacts with trace gases via a charge transfer producing M^+ . In addition to reacting with organic compounds, O_2^+ also reacts with some small molecules that do not react with H_3O^+ and/or NO^+ , thereby providing valuable information to construct a more complete mass spectrum. [1]

[0026] Because most trace gases can react with two or more precursor ions, shifting between precursor ions can both the identity and quantity of the reactant gas can be determined. The multiple-ion-monitoring (MIM) mode is used in this way to simultaneously monitor several selected trace gas product ions. If target gases are known and clearly identified, this mode is most suitable for their detection and quantification. If, however, the gas of interest is not known, and one would like to identify potential gases that might be associated with disease status, then the full-scan (FS) mode is appropriate. For each precursor ion, the FS mode produces a full range of mass profiles from m/z 10 to m/z 300 and their counts. Although the FS mode provides a rich amount of information and offers exciting opportunities to study disease associations, few existing studies have been done on such data, at least in part due to the available sample size being less than the number of variables. [1]

[0027] The SIFT-MS instrument calibration is done by automated routine validation procedures and requires little operational adjustment. The accuracy of a SIFT-MS instrument is tested to be better than 10% for compounds quantity from 10 ppb to 20 ppm for a current model Voice200TM of SIFT-MS instrument (Syft Technologies Ltd, New Zealand). The Voice200TM weighs 212 kg, has dimensions less than 1 m³ [13], and being portable can be conveniently used in hospitals and clinics.

[0028] Medical application of SIFT-MS for breath analysis is in the early stages of development. Most studies use pre-selected gases to establish correlation with disease. Davies, Spanel and Smith (1997) and Endre et al. (2011) [29, 32] used SIFT-MS to measure selected breath compounds in end-stage renal disease patients who have undergone dialysis,

and reported that breath ammonia had a significant correlation to blood creatinine and urea level; however, not all changes before and after a dialysis session could be tracked. Among the few studies that use multiple compounds, Wehinger et al (2007) [34] used a PTR-MS method to analyze the breath samples of 17 lung cancer patients and 170 controls and found that mass 31 and mass 43 were significantly elevated in lung cancer subjects. Phillips et al (1999) [20] conducted a study of 108 patients using GC-MS, and identified 22 volatile organic compounds by forward stepwise discriminant analysis to classify lung cancer cases from controls. The sensitivity reached 71.7% with specificity of 66.7%. The study of Phillips et al. indicates great potential of breath mass spectrometry in disease classification. As pointed out by Risby and Solga (2006) [27] in a review, the clinical application of breath analysis using mass spectrometry will only be widely accepted when technology development reaches a certain stage. Currently, SIFT-MS instruments have become portable, easy to use, fast and accurate, thereby facilitating its use in clinical applications. In-depth cross-sectional or longitudinal studies are needed to develop reliable diagnostic methods of breath mass spectrometry. Contrary to the studies mentioned above, many existing studies report weak or no association of breath compounds with disease. The reason is likely because only single gases such as acetone and ammonia were analyzed. In fact, proper selection of multiple breath VOCs from hundreds of predictors are an important and essential step before an accurate and reliable diagnostic model can be established, and it is in this key aspect that the statistical methods disclosed herein address.

[0029] However, even with instrument calibration, systematic variation due to subject's condition and differences among measurements may be present. Normalization methods can be a part of the analysis. After removing systematic variation, more information from the exhaled breath compounds can be revealed. Methods for data normalization include, for example, centering the compounds by mean value and adjusting the mass value by

benchmark precursor compounds in the mass spectrum. Both methods retain the variance of individual masses, and remove systematic bias due to external influences.

[0030] After SIFT-MS of breath samples are obtained the mass spectra are analyzed using multivariate statistical methods.

[0031] Pearson's correlation test is performed on a SIFT-MS spectrum, and a p-value < 0.05 is considered significant. Statistical software R [14] can be used to perform the analysis disclosed herein.

[0032] The number of compounds detected in the SIFT-MS full-scan mode far exceeds the number of measurements, and renders traditional tools such as linear regression ineffective, which is a common case in medical studies in which the number of patients is limited. To analyze the SIFT-MS data a multivariate statistical method can be employed such as, for example, the lasso by Tibshirani (1996) [12], which is a penalized regression method that is widely used in bioinformatics to identify genetic markers associated to a trait and with perform predictions.

[0033] The basis of the penalized regression (lasso) model is briefly summarized as follows. Given (x_i^T, y_i) , $i=1, \dots, N$, where $x_i^T = (x_{i1}, \dots, x_{ip})^T$ are the predictor variables, and y_i are the responses, lasso minimizes least square estimates $(\hat{\alpha}, \hat{\beta})$ subject to a linear inequality constraint:

$$(\hat{\alpha}, \hat{\beta}) = \arg \min \left\{ \sum_{i=1}^N (y_i - \sum_{j=1}^p \beta_j x_{ij} - \alpha)^2 \right\} \quad \text{s.t.} \quad \sum_j |\beta_j| \leq t$$

Equation 1

where t is a tuning parameter to be determined by cross-validation. The lasso is essentially a linear regression model with a first order penalty on the coefficients. Because of the absolute first order constraint, the model can effectively put some coefficients to exact zero, achieving a variable selection purpose. For the same reason, the lasso can address the situation that the number of parameters is greater than the number of subjects.

[0034] In Equation 1, when $t = 0$, the model is equivalent to multiple linear regression; and when t is small, a larger penalty is imposed and more coefficients will be put to zero. Thus, the lasso is suitable for analyzing the full-mass scan data, for it selects a small set of variables out of the large total number of possible variables that are more than the number of samples. After variable selection, multiple linear regression can be used to make predictions and to obtain a significance level on the estimated coefficients. The lasso can be performed in the R environment using a statistical package such as *cv.glmnet* and *glmnet* [40]. Cross-validation selection of t (in the R package, it is a λ which inversely related to t) avoids an over-fitting problem in which a model perfectly performs on the dataset it is estimated, but performs poorly on an independent dataset. With over-fitting ruled out, no variable would be selected if predictor variables have no association with their response. Using the R environment [12] automated cross-validation package *cv.glmnet* [15] to estimate the tuning parameter, the selected tuning parameter can be entered into a lasso package such as *glmnet* [15] to calculate the lasso model.

[0035] **FIG. 2** shows how patient data was partitioned to first develop a statistical model, and then to analyze a sub-set of data to verify the accuracy of the model. To objectively and accurately assess the prediction model estimated by lasso, 5 breath samples were randomly set apart as independent test cases (**FIG. 2**), which were not involved in model building. The remaining 35 breath samples are used as a training set. The samples are further randomly divided to form 5 cross-validation training and cross-validation test sets (**FIG. 2**). The tuning parameter selection is done within a cross-validation training set. Variables having non-zero coefficients more than twice in the cross-validation groups are selected, and are also referred to as lasso selected masses or selected masses. Prediction can be performed using the lasso selected masses. In practice, one can directly perform lasso on all samples to fully utilize the available information.

[0036] Usually when analyzing full-scan masses, the precursor ions themselves and their products with water vapor can also be measured [16]. The lasso naturally sets most of these coefficients to zero along with the coefficients for the masses of other non-influential breath compounds. The only exceptions are two precursor ions (mass 73 and mass 30), which appear in the lasso selected masses for precursor H_3O^+ and O_2^+ , though neither one is significant.

[0037] We have compared the results obtained using the precursor measurements separately and jointly in our working paper (Wang et al. 2012a), and found that the results give similar test performance but that a model based on the combined data selects fewer predictor variables. Thus, the combined data set provides a greater opportunity for increased prediction accuracy. Several alternative analysis methods have been evaluated, including the lasso, t-test, multiple linear regression, and the sliced inverse regression (SIR). The methods are selected because they can (1) handle the situation that the number of variables is larger than the number of samples; (2) deal with continuous response variables; and (3) are simple and robust. These are among the key issues to address in practical SIFT-MS breath compound analysis.

[0038] Alternative statistical analysis methods may also be used to analyze SIFT-MS breath samples including, for example, T-test and multiple linear regression, sliced inverse regression (SIR), and texture and fractal analysis.

[0039] Student's t-test first can be performed on each gas compound with the phenotype. It is expected that some underlying gases like acetone, ammonia and isoprene will have strong correlation to biomarkers of renal function such as blood creatinine and urea. The T-test is an easy and robust way to select gases. The number of markers can be determined by cross-validation and based on the p-value. Then a multiple linear regression can be applied on the selected biomarkers to build a model and obtain predictions.

[0040] Another method that can be used to analyze SIFT-MS data is the sliced inverse regression (SIR) [41], [42]. SIR is a nonparametric regression method that uses local smoothing of the response variable. After standardizing X , an inverse regression of X on sliced Y is performed. The inverse regression, $E(X|Y)$, converts a high-dimensional regression problem of Y on X to many simple regressions of X on Y . In estimating $E(X|Y)$, the range of Y is divided into small intervals (sliced) to increase computational efficiency. Next, a principal components analysis will be performed on $E(X|Y)$, and the principal components (PCs) are returned. The coefficient of the first SIR effective direction to select variables. The SIR is similar to Fisher's discriminant analysis, but has the advantage of making use of continuous phenotype information.

[0041] Useful texture and fractal analysis models include, for example, statistical texture analysis, high order spectral (HOS) analysis, and fractal analysis. Using these models, useful interactions among the important features can be evaluated to determine if there are complex relationships that can provide information in breath samples useful for medical diagnostics.

[0042] Multivariate statistical methods such as the lasso can be used to analyze SIFT-MS data from exhaled breath for medical diagnosis and treatment. In certain embodiments, the methods can be used for the diagnosis of ST-elevation cardiovascular diseases including myocardial infarction, non-ST-elevation myocardial infarction, and angina, diabetes, pre-diabetes, renal function, kidney diseases, cancers, stroke, infections including influenza, common cold, bacterial infections, tuberculosis, and human papillomavirus, asthma, drug addiction, Chinese medicine diagnostic system and outcome measures, gastroenterological diseases, neuro-degenerative disorders including dementia and Alzheimer's disease, mental illness and depression. In certain embodiments, the disease to be diagnosed is selected from kidney disease such as renal failure, heart attack, and diabetes. Application of the disclosed statistical methods to the diagnosis and treatment of these diseases is provided as follows.

[0043] Kidney disease is characterized by five stages from the first stage, with few symptoms, to the worst case of end stage kidney disease (ESKD) in which patients need to rely on haemodialysis or renal replacement to sustain life. Because the symptoms of early impaired renal function are not obvious and not easy to notice by self-examination, kidney disease is usually discovered at later stages with relatively few intervention options. The current testing method relies solely on the blood creatinine level or the urine protein concentration, which can be considered a gold standard for testing renal function. However, the test is invasive and it can take hours to several days to receive results. As a consequence, patients do not have frequent voluntary examinations. Because early detection of deteriorating renal function would have a major impact on slowing or preventing kidney disease and enable effective intervention, a fast, non-invasive testing method for assessing renal function is needed. This is especially important for diabetes patients. In our pilot study, we have found that the blood level creatinine and urea can be accurately predicted using multiple breath compounds. This development demonstrates that SIFT-MS breath analysis provides a convenient and fast method for assessing kidney disease, and suggests that similar methods can be applied to the diagnosis of diabetes and for other diseases.

[0044] Plasma creatinine and urea level are the two basic indicators of renal function and dialysis efficacy. For the first time, the two indicators can be accurately estimated non-invasively by full-scan mode SIFT-MS. Previous studies have shown that there is a significant correlation between breath chemicals and dialysis efficacy using SIFT-MS multiple-ion-monitoring (MIM) mode [11]. For example, ammonia, and TMA were found to be correlated with plasma creatinine and urea levels. Utilizing full-scan (FS) provides additional information and improves the predictions, but the analysis is complicated by the difficulty of extracting information from the large amount of collected data. Using multivariate statistical methods, the data from SIFT-MS can be efficiently and accurately analyzed. Using multivariate statistical methods such as the lasso method described herein,

a small number of masses can be selected by finding robust estimators and then analyzing the selected masses using a linear regression model. The predicted values, both on the training set and the test set, track the dialysis effect closely (**FIG. 3D**). For precursor H_3O^+ , the correlation coefficient is 0.96 between the true creatinine values and the predicted values by lasso selected masses.

[0045] Breath ammonia and acetone levels are significantly elevated in patients with end-stage renal disease (ESRD) compared to patients without renal disease [8,9]. Consistent with previous studies, ammonia product ions (m/z 18 for H_3O^+ and m/z 17 for O_2^+) and acetone product ions (m/z 58) are found among the lasso selected masses by both precursor H_3O^+ and precursor O_2^+ when predicting creatinine and urea levels. Ammonia is derived mostly from deamination of amino acid and can be directly caused from renal failure; while acetone is produced by increasing catabolism of fatty acid, induced from relative deficiency of carbohydrate absorption. The increased amounts of both metabolites, which are eventually expired from exhaled breath, reflects problems associated with the kidney function. The correlations of plasma creatinine levels with exhaled ammonia, acetone, TMA, and the predicted value are shown in **FIGS. 3A-3D**.

[0046] Isoprene ions (m/z 68) are also found to be significantly associated with blood creatinine and urea using SIFT-MS full-mass mode. Isoprene is not only selected, but it is also found to be very significant in the regression model, the p-values are 0.0213 and 0.0018, in respective regression analysis of creatinine and urea by precursor NO^+ (Table7). Breath isoprene results from body cholesterol synthesis [17]. It was observed that dialyzed patients exhibit significantly higher isoprene level than pre-dialysis, however, the isoprene levels of renal failure patients (before dialysis) is higher than subjects without renal failure. It is possible that isoprene may be an indicator of psychological stress resulting from tissue injury [16].

[0047] Many metabolites that are found to be significantly associated with creatinine and urea are of special interest for further research into their metabolic pathway related to renal function (Table 6 and Table 7). These include, for example, acetylene (m/z 26), methional (m/z 104), trichlorobenzene (m/z 180), ethyl nonanoate (m/z 185) by precursor NO^+ , and m/z 148 (nicotine) by precursor H_3O^+ . Also, acetonitrile (m/z 41), which is an indicator of smoking, is identified by precursor O_2^+ in creatinine analysis. Moreover, nicotine is found to be significant in both urea and creatinine associated markers. Thus, smoking may have a correlation with kidney disorders.

[0048] To assess the methods provided by the present disclosure for the diagnosis of renal failure, a pilot dataset was used. The pilot dataset was a subset of the phenotype data used in Endre et al. (2011) [36] and included patients with end-stage renal disease (ESRD), attending dialysis sessions in a home-dialysis training center in Christ Church, New Zealand. The blood urea and creatinine values, breath VOC full-scan masses, breath acetone and trimethylamine (TMA) were measured before and after each dialysis session. There were in total 40 measurements from 5 patients. The full-scan masses data ranged from mass-to-charge ratio (m/z) 10 to m/z 200, and included 191 substances in total. A penalized linear regression method such as the lasso [37] was used to build prediction models for blood urea and creatinine levels from the full-scan mass data of breath. A training set was used for model building, and an independent test set with 5 measurements was separated out for evaluation.

[0049] The predicted values were not only very accurate on training set, but also quite good on the independent test set. The changes in blood level measurements were be accurately tracked, even for extremely high and low levels of creatinine and urea. The predicted values and the original values for creatinine are compared in Table 1.

Table 1. Prediction of creatinine value on an independent test set (H_3O^+) by

breath compounds MAPE* = 10.38%.

Sample ID	Predicted Creatinine	Measured creatinine	Deviation	Dev%
DS20104 1	257.85	235.0	22.85	9.7%
DS20301 0	918.08	893.0	25.08	2.8%
DS20301 1	515.50	420.0	95.50	22.7%
DS20303 1	325.75	284.0	41.75	14.7%
DS20305 0	617.66	630.0	-12.34	2%

* MAPE: mean absolute percentage error

[0050] For the precursor ion H_3O^+ , 14 compounds were selected from 191 VOCs, and the R-square on the training set is 85.4%. The mean absolute percentage error (MAPE) was used to evaluate prediction accuracy of the continuous response variable (Equation 1), where n is the number of training or testing set subjects, and $\hat{y}$ is the predicted value.

$$MAPE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| / |y_i| \quad \text{Equation 1}$$

[0051] Cross-validation (CV) was performed 5 times, and the averaged MAPE on test cases is 10.38%, with standard deviation 8.63%. The same procedure was done on precursor NO^+ and O_2^+ and the averaged MAPE of test sets for three precursors was 17.1% with a standard deviation 3.1%. A plot of the predicted values of training set and test set is shown in **FIG. 7**. As shown in **FIG. 7**, the predicted values closely follow the trend before and after each dialysis session. The correlation between predicted and true creatinine is 0.96 (p-value $< 10^{-21}$), implying that SIFT-MS breath analysis is a promising non-invasive method for tracking dialysis efficacy. For the prediction of urea, on average 17 VOCs were

selected, and the average MAPE from 5 group-CV was 17.57% with a standard deviation of 22.5%.

[0052] The masses identified for final prediction not only include the common markers that are known to be associated with renal disease such as acetone (m/z 58), ammonia (m/z 18), and isoprene (m/z 68), but also include other significant biomarkers (p-value <0.05) including nicotine (m/z 148), m/z 111, m/z 144, propyl butanoate (m/z 130) and isoflurane (m/z 185). From the nicotine identified, it is possible that smoking has a certain effect on renal disease. Other compounds may be related to unknown metabolism pathways and are candidates for further study of their relation to renal disease physiology.

[0053] The change in creatinine levels before and after each dialysis session is shown in **FIGS. 3A-3D**. The dashed red line is the measured blood creatinine value, and the solid black line is the predicted creatinine value based on SIFT-MS breath analysis. As shown in **FIGS. 3A-3D** that the predicted values closely follow the trend of change after each dialysis session. The prediction is made using a model based on 14 compounds, and the correlation between predicted and blood creatinine is 0.96, implying that breath analysis as a useful non-invasive method for tracking dialysis efficacy.

[0054] Classification of heart attack patients compared to controls was also determined. Patients who presented to an emergency department with chest pains were recruited. Among the 38 patients recruited, 6 are diagnosed with Non ST Segment Myocardial Infarction (Non-STEMI) – an acute coronary syndrome without ST segment elevation; 8 patients are diagnosed with unstable angina (UA) – an acute coronary syndrome which may progress to heart attack; and 24 were diagnosed as not having heart attacks. Non-STEMI is the most severe type of heart disease in the study cohort. The 14 patients with non-STEMI and UA were combined as heart attack case group, and the other 24 patients were included in the control group. Breath samples were collected and full-scanned masses of breath compounds were measured by SIFT-MS. Similar statistical methods as used in the dialysis

study to analyze the data and make classifications. The average error rate of three precursor ions for the training set was 21.2%, and for the test set was 25%. Interestingly, the associated biomarkers mostly have large mass-to-charge ratios and do not exist naturally, but rather are manmade chemicals used in household cleaning, pesticide (2-chloroethyl ether, m/z 143), explosives (2,3-dimethyl-2,3-dinitrobutane, m/z 176), etc. For example, 1,1,1-trichloroethane, a toxic chemical and have an effect on cardiac arrhythmias, was found to be significantly correlated with heart attack (p -value = 0.00162) (Toxicological Profile for 1,1,1-Trichloroethane). Bromobenzene (m/z 157, p -value=0.0018) was also identified, which can cause nerve and liver damage [38], and was previously reported to be associated with heart disease [39]. The heavy chemical compounds identified in human breath reflect that environmental exposure to toxic substances may have an effect in inducing heart diseases. This study confirms that breath samples contain useful information for disease sub-typing. The studies also show that it is much more efficient to use multiple breath compounds rather than a single metabolite in building prediction models than based on a single metabolite; and that advanced statistical methods can be used to effectively derive information from the SIFT-MS full-scan mass data of exhaled breath.

[0055] The use of the methods provided by the present disclosure to diagnose diabetes was also evaluated. Diabetes can lead to cardiovascular diseases, kidney failure, leg amputations, blindness, and stroke [7]. Type 2 diabetes is the most common form of diabetes among Hong Kong adults. The annual per-capita health care expenditure is about four-fold for people with diabetes compared with the general population and thus diabetes pose a substantial burden to Hong Kong healthcare system [8]. Type 2 diabetes is currently affecting around one in ten people and more than half of type 2 diabetic patients remain undiagnosed in Hong Kong [9]. Hence, there is a need to develop a reliable and inexpensive method for screening the diabetic or pre-diabetic patients in order to reduce the proportion

of undiagnosed patients and initiate life style or behavioral intervention as early as possible to reduce both the medical and financial burdens on society.

[0056] The main purpose of screening is to distinguish an asymptomatic individual at high risk from an individual at low risk for diabetes. According to previous findings [10], diabetes subjects have a recognizable asymptomatic stage and would benefit from early diagnosis. Screening and early intervention are shown to be cost-effective in the United States health care system [18]. Typical screening methods include questionnaires and biochemical tests. Questionnaires are easy to use but may perform poorly as a standalone test. Herman developed a questionnaire that is capable to detect undiagnosed subjects with sensitivity and specificity of 79% and 65% respectively, and a positive predicted value (PPV) of only 10% [11]. The instrument has been adopted in a community program by the American Diabetes Association (ADA) [12]. On the other hand, biochemical tests have higher accuracy but can be inconvenient for individuals as they required blood drawing or finger prick. Glucose [13], HbA1C [14, 15], and fructosamine levels [14] are common test indicators for diabetes screening. Although oral glucose tolerance test (OGTT) (2-hour plasma glucose concentration post consuming glucose load as the measurement) is the gold standard, the ADA recommends fast plasma glucose (FPG) test because of ease of administration, convenience, acceptability to patients, and lower cost. Generally, the sensitivity and the specificity of the screening tests ranged from 40-65% and >90% respectively [10].

[0057] SIFT mass spectrum from 50 diabetes patients were collected by Syft Technologies Ltd (New Zealand) (SYFT). Each sample contained signal intensity in counts per second (cps) for three kinds of precursor ions (H_3O^+ , NO^+ and O_2^+) over a range of atomic mass unit (amu) of m/z 10 to m/z 180. Each subject repeated breath sessions for 20 times to 50 times to obtain an average value for a subject and be able to control the intra-subject variations. The intensity data was analyzed using SAS 9.2 [25] and R [26]. An

example of SIFT mass spectra obtained for the three precursor ions from a breath sample is shown in **FIG. 4**.

[0058] Among the 50 subjects, acetone levels were collected from 38. The relationship between breath acetone and other clinical parameters body mass index (BMI), plasma glucose levels, and HbA1c were analyzed by scatter plots and shown in **FIG. 5A**, **FIG. 5B**, and **FIG. 5C**, respectively.

[0059] According to the results, solely analyzing the breath acetone may not show a clear trend for the association of diabetic parameters. Additional SIFT-MS intensity data was further evaluated to determine if a relationship to acetone level (ppb) could be established. Stepwise linear regression was used to carry out variable selections. As shown in Table 2, the following reagents and products were highly significant (p-value<0.001) for acetone levels.

Table 2.

Reagents	Product Mass	Regression R-square
H ₃ O ⁺	57, 77, 142, 165	0.95
NO ⁺	30	
O ₂ ⁺	-	

[0060] Scatter plots of these relationships are shown in **FIGS. 6A-6D**. As demonstrated by the results presented in **FIGS. 6A-6D**, acetone exhibits a positive correlation to products of reagent H₃O⁺ especially to atomic mass unit (amu) 77 and a negative correlation to amu 30 of reagent NO⁺. The R-square which measures the goodness of fit was larger than 90%. It is highly likely that acetone level can be predicted by SIFT breath compounds using all spectrometry data.

[0061] The correlation of SIFT-MS products on other clinical parameters of diabetes was also assessed. The results from stepwise linear regression analysis of SIFT-MS data with the clinical parameters, body mass index (BMI), the glucose level, and HbA1c, typically associated with diabetes are presented in Table 3.

Table 3.

Dependents	Reagents	Product Mass	Regression R-square
BMI	H ₃ O ⁺	156	0.75
	NO ⁺	150, 152, 173	
	O ₂ ⁺	-	
Glucose level	H ₃ O ⁺	30, 140	0.35
	NO ⁺	-	
	O ₂ ⁺	-	
HbA1c	H ₃ O ⁺	39, 108	0.73
	NO ⁺	29, 52, 126	
	O ₂ ⁺	-	

With the exception of the glucose level, the regression model of SIFT-MS exhibits good correlation with the BMI and HbA1c showing more than 70% goodness of fit.

[0062] The results support the conclusion that the analysis of SIFT-MS data of breath samples can be used to predict acetone, BMI, and HbA1c on a real time basis. Because these are known factors (biomarkers) associated with diabetes, SIFT-MS can be used as a tool for screening purposes.

EXAMPLES

[0063] Embodiments provided by the present disclosure are further illustrated by reference to the following examples. It will be apparent to those skilled in the art that many

modifications, both to materials, and methods, may be practiced without departing from the scope of the disclosure.

Example 1

[0064] Five adult hemodialysis patients were recruited from a home-dialysis training center. The patients were dialyzed 4 days per week (Monday, Tuesday, Thursday, and Friday) utilizing Hemoflow F8HPS dialyzer (Fresenius Medical Care AG, Bad Homburg, Germany). Each dialysis session started between 8 AM and 9 AM with 5 hours duration. The blood creatinine and urea level, weight, the use of medical inhalers, oral hygiene measures, and change of health status of the patients were recorded before and after each dialysis session. Plasma creatinine and urea were measured by Architect c8000 (Abbott Laboratories, IL, USA) using heparinized blood samples. The breath mass concentration was recorded by a Voice200TM SIFT-MS (Syft Technologies Ltd. Christchurch, New Zealand) in both selected ion monitoring mode and whole mass scan mode. The participants directly breathed into a heated inlet of the SIFT-MS, and also filled Tedlar bags (1 litre) before and 30 minutes after dialysis.

[0065] The Friday session of patient DS202 and Tuesday session of DS203 were incomplete, and additional Monday and Tuesday sessions for DS203 were performed one month later. In all the data for 20 sessions (40 measurements of pre- and post-dialysis) were available for analysis.

[0066] The blood level of creatinine and urea for each patient was measured. Breath ammonia, acetone and trimethylamine (TMA) were obtained by SIFT-MS MIM mode. For each precursor ion H_3O^+ , NO^+ and O_2^+ , the full-scan mode of SIFT detected the counts over a spectrum of 186 compounds (mass-to-charge ratio ranges from m/z 15 to m/z 200). The measurement for each compound was repeated several times and resulted in 6 to 50 recordings for each sample. The repeated recordings were averaged to obtain a single value for each compound per measurement.

[0067] When blood level creatinine is used as a response variable (Y), there are 14 lasso selected masses identified from full-scan masses by precursor H_3O^+ . Linear regression on the 14 masses resulted in a very significant model with adjusted $R^2 = 85.4\%$ and the model p-value is 10^{-7} . The prediction based on test set is shown in Table 4.

Table 4. Prediction of creatinine value on test set (H_3O^+) MAPE* = 10.38%.

Sample ID	Predicted Y	True Y	Deviation	Dev%
DS201041	257.849	235	22.849	9.7%
DS203010	918.075	893	25.075	2.8%
DS203011	515.496	420	95.496	22.7%
DS203031	325.747	284	41.747	14.7%
DS203050	617.658	630	-12.342	2%

* MAPE: mean absolute percentage error

[0068] The prediction accuracy was evaluated using the mean absolute percentage error (MAPE), that is, the percentage of absolute difference of predicted Y from true Y, which is 10.38% in this case. The predicted values of training set are shown in **FIG. 7** and demonstrated that the test set aligned with true Y. Independent test sets were randomly drawn an additional four (4) times, and the average MAPE was 17.1% with a standard deviation 3.1%.

[0069] The complete regression output is presented in Table 5.

Table 5: Regression of creatinine on lasso selected masses (precursor H_3O^+)

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-571.100	148.400	-3.849	0.0010	**
Mass 22	25.760	9.733	2.647	0.0155	*
Mass 23	17.390	10.760	1.616	0.1218	
Mass 30	0.313	0.307	1.019	0.3202	
Mass 58	0.260	0.317	0.822	0.4207	

Mass 73	0.001	0.003	0.203	0.8410	
Mass 75	0.047	0.159	0.295	0.7713	
Mass 77	0.002	0.002	1.003	0.3279	
Mass 108	-0.337	0.208	-1.618	0.1214	
Mass 130	1.641	0.568	2.890	0.0091	**
Mass 149	0.111	0.092	1.214	0.2387	
Mass 151	0.550	0.296	1.855	0.0784	.
Mass 154	0.464	0.499	0.931	0.3629	
Mass 160	0.033	0.353	0.094	0.9263	
Mass 198	0.283	0.544	0.520	0.6085	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'

Residual standard error: 87.95 on 20 degrees of freedom

Multiple R-squared: 0.9144, Adjusted R-squared: 0.8544

F-statistic: 15.25 on 14 and 20 DF, p-value: 1.045e-07

[0070] Three compounds exhibited a regression coefficient significance < 0.10, corresponding to mass 30 (NO), mass 130 (C₇H₁₄O₂, C₈H₁₈O), and mass 151, for which the associated compounds are listed in Table 6.

Table 6. Significant product ions in predicting creatinine

Precursor	Significant product ions	Mass	Regression P-value
H ₃ O ⁺	deuterium oxide (20)	Mass 22	0.023
	acetylene (26)		
	tertiary butyl propanoate,	Mass 130	0.044
	butyl propanoate, propyl		
	butanoate, isopropyl butanoate,		
	pentyl acetate, ethyl		
	2-methylbutanoate,		
	2-methylbutyl acetate, isopentyl		
	acetate		
	ethyl pentanoate (130.21)		
	2-octanol, 1-octanol (130.27)		
	bromodifluoromethane		

	(130.93)		
	DEN (150.3)		
	ethylene glycol dinitrate	Mass 151	0.074
	(152.04)		
	acetylene(26.04)	Mass 25	0.0139
	methyl isocyanate (57.051)		
	2-methylaziridine (57.09)	Mass 57	0.0080
	furan (68.08)		
	isoprene (68.14)	Mass 68	0.0213
	1,5-pentanediol (104.17)		
	styrene (104.17)	Mass 104	0.0022
	methional (104.18)		
	ethyl hexanoate, isobutyl		
NO ⁺	butanoate, butyl butanoate, hexyl	Mass 144	0.0017
	acetate (144.24)		
	DEN (150.3)		
	ethylene glycol dinitrate	Mass 151	0.0077
	(152.04)		
	EDME (181.26)		
	1,2,4-trichlorobenzene	Mass 180	0.0011
	(181.46)		
	isoflurane (184.5)		
	perfluorobenzene (186.07)	Mass 185	0.0001
	ethyl nonanoate (186.33)		
	nitrogen molecule (28.1)		
	ethane (28.06)	Mass 28	0.027975
	1-butanamine, diethyl amine		
	(73.16)	Mass 73	0.000753
	dimethyl methyl phosphonate		
	(124.08)		
O ₂ ⁺	p-hydroxybenzyl alcohol,		
	guaiacol (124.15)	Mass 124	0.091926
	2-chloroethyl ethyl sulfide		
	(124.65)		
	2-isobutylthiazole (141.26)		
	methyl iodide (141.94)	Mass 141	0.001562
	1-phenyl-1,2-propanedione	Mass 148	0.003737

(148.17)		
methyl chavicol , cuminal		
(148.23)		
nicotine (148.26)		
1-iodopropane,		
2-iodopropane (170)		
gamma-decalactone,		
delta-decalactone, 1-octen-3-yl	Mass 170	0.056358
acetate (170.28)		
2-undecanone (170.34)		
dodecane (170.4)		

Predictions on creatinine - precursor NO^+ and O_2^+

[0071] Precursor NO^+ has 18 lasso selected masses associated with creatinine, the linear regression adjusted $R^2 = 96.2\%$ and MAPE on test set is 15.43%. Precursor O_2^+ has 14 lasso selected masses and adjusted $R^2 = 89.6\%$, test set MAPE = 18.3%. The results indicate that full-scanned masses can be used to build accurate prediction models for renal failure.

Furthermore, the decrease of blood creatinine after each dialysis can be tracked (**FIG. 7**), which is not the case using single breath compounds in the MIM mode.

[0072] For precursor H_3O^+ , using a 15-mass linear regression model, the fitted adjusted R^2 is 97.7%, and MAPE is 13.2% on an independent test set. The predictions closely follow the change of urea level before and after dialysis (**FIG. 8**). Significant product ions associated with urea prediction and the corresponding compounds are summarized in Table 7.

Table 7. Significant product ions in predicting urea

Precurs	Significant product ions	Mass	Regression
---------	--------------------------	------	------------

or		p-value
H ₃ O ⁺	ethanedial (58.03), propylene oxide (58.08), acetone(58.09), propanal (58.09), butane (58.15), isobutane (58.15)	Mass 58 0.000133
	Catechol (110.12), 5-methylfurfural (110.12), 2, 4-fluorotoluene (110.15), dichloropropene (110.97)	Mass 110 0.00141
	tertiary butyl propanoate, butyl propanoate, propyl butanoate, isopropyl butanoate, pentyl acetate, ethyl 2-methylbutanoate, 2-methylbutyl acetate, isopentyl acetate, ethyl pentanoate (130.21), 1,2-octanol (130.27), Bromodifluoromethane (130.93)	Mass 130 0.000108
	p-isopropenyl toluene (132.23)	Mass 132 0.026792
	isoflurane (184.5), perfluorobenzene (186.07), ethyl nonanoate (186.33)	Mass 185 0.032358
	menthyl acetate (198.35), tetradecane (198.46)	Mass 198 0.017251
	nitrous acid (47)	Mass 47 0.003542
NO ⁺	furan (68.08)	Mass 68 0.001848
	isoprene (68.14)	
	1,5-pentanediol (104.17)	
	styrene (104.17)	Mass 104 0.019284
	methional (104.18)	
	EDME (181.26)	
	1,2,4-trichlorobenzene (181.46)	Mass 180 0.020484
O ₂ ⁺	isoflurane (184.5)	
	perfluorobenzene (186.07)	Mass 185 0.000341
	ethyl nonanoate (186.33)	
	ethylene glycol (62.07)	
	dimethyl sulfide, ethyl mercaptan, chloroethene (62.15)	Mass 62 0.02199

1-phenyl-1,2-propanedione (148.17)		
methyl chavicol , cuminal (148.23)	Mass 148	0.00941
nicotine (148.26)		
2,3-dimethyl-2,3-dinitrobutane (176.17)	Mass 176	0.03739

[0073] The analysis of SIFT-MS spectra using the precursors NO^+ and O_2^+ give similar results: the linear regression R^2 are 99.12% and 94.80%, and test set MAPE are 15.49% and 24.03%, respectively (Table 8).

Table 8. Summary of prediction on urea by 3 precursors using full-scan masses.

Precursor	Number of lasso selected masses	Linear regression adjusted R^2 using lasso selected masses	Test set MAPE*
H_3O^+	14	97.67%	13.20%
NO^+	25	99.12%	15.49%
O_2^+	12	94.80%	24.03%

* MAPE: mean absolute percentage error from true Y

[0074] The lasso analysis of combined masses from the three precursor ions was also evaluated. Performing lasso on combined masses from all 3 precursor ions provided similar result (Table 9): the MAPE for creatinine test set is 15.69%, while the average MAPE from 3 separate models is 14.69%. Although a little less than the averaged MAPE, this model used only 8 variables, which is even more efficient than running 3 separate predictions. Also, the 8 variables are all subsets of the separate selections. The result for urea is similar with a MAPE of 16.6%, which is even better than the averaged MAPE 17.75% from

separate models. The results based on an analysis of the combined masses and using the separate models is compared in Table 9.

Table 9. Comparison of combined masses from 3 precursors and separate models.

	Combined masses		Separate models	
	MAPE	# of Variables	Average MAPE	Average # of variables
Creatinine	15.69%	8	14.69%	15
Urea	16.6%	16	17.57%	17

[0075] These results demonstrate that facilitated by appropriate statistical methods using SIFT-MS full-scan mode, breath compounds can be used as an accurate predictor of blood level creatinine and urea. Predictors are obtained by applying lasso in cross-validation groups, some 200 compounds in breath are reduced to about 20 lasso selected masses, which are then passed to linear regression for prediction. The significant compounds, i.e., the selected masses, obtained in full-scan masses reproduced the metabolites reported in previous studies using a few candidate gases. In addition, the study suggested a number of new metabolites in exhaled breath that may relate to renal disease, such as acetylene, methional, trichlorobenzene and ethyl nonanoate. These compounds can be evaluated for their association with metabolic pathways in kidney function. The analytical methods disclosed herein applied to SIFT-MS full-scan mode of exhaled breath can be used to accurately track the dialysis efficacy, conveniently and non-invasively.

[0076] Methods provided by the present disclosure further include methods of treating a disease in a patient. Using the methods for analyzing SIFT-MS data obtained from exhaled breath as disclosed herein, a patient can be diagnosed as having a disease. A physician or

health care provider can, for example, request that the diagnosis be performed.

Subsequently, if a patient is diagnosed as having a disease, the patient may be treated for the disease. The treatment may include, for example, administering a medication, such as a therapeutically effective amount of a medication, and/or undertaking a procedure to treat the disease. Treating or treatment refers to reducing, minimizing, and/or preventing the disease in a patient.

[0077] Finally, it should be noted that there are alternative ways of implementing the embodiments disclosed herein. Accordingly, the present embodiments are to be considered as illustrative and not restrictive. Furthermore, the claims are not to be limited to the details given herein, and are entitled their full scope and equivalents thereof.

References

1. Phillips et al., Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomark.* 3(2):95-109 (2007).
2. Phillips et al., Prediction of breast cancer using volatile biomarkers in the breath. *Breast Cancer Res Treat.* Sep;99(1):19-21 (2006).
3. Phillips et al., Breath biomarkers of active pulmonary tuberculosis. *Tuberculosis (Edinb).* Mar;90(2):145-51 (2010).
4. Manolls. The diagnostic potential of breath analysis. *Clin chem.* 29(1):5-15 (1983).
5. Storer et al., Measurement of breath acetone concentrations by selected ion flow tube mass spectrometry in type 2 diabetes. *J Breath Res.* Dec;5(4):046011 (2011).
6. Dillon et al., Novel noninvasive breath test method for screening individuals at risk for diabetes. *Diabetes Care.* Mar;32(3):430-5 (2009).

7. Roglic et al., The burden of mortality attributable to diabetes: realistic estimates for the year 2000. *Diabetes Care*. 28:2130-5 (2005).
8. Leung and Lam, Diabetic complications and their implications on health care in Asia. *Hong Kong Med J*. 6: 61-8 (2000).
9. Wong and Wang, Prevalence of type 2 diabetes mellitus of Chinese populations in Mainland China, Hong Kong and Taiwan. *Diabetes Res Clin Pract*. 73:126-34 (2006).
10. Engelgau et al., Screening for Type 2 Diabetes. *Diabetes Care*. Oct;23(10):1653-80 (2000).
11. Herman et al., A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes. *Diabetes Care*. Mar;18(3):382-7 (1995).
12. American Diabetes Association: Screening for type 2 diabetes (Position Statement). *Diabetes Care*. 23(Suppl. 1):S20–S23 (2000).
13. Andersson et al., A model for early diagnosis of type 2 diabetes mellitus in primary health care. *Diabetes Care*. Mar;10(2):167-73 (1993).
14. Ko et al., Combined use of a fasting plasma glucose concentration and HbA1c or fructosamine predicts the likelihood of having diabetes in high-risk subjects. *Diabetes Care*. Aug;21(8):1221-5 (1998).
15. Ko et al., Supplement to the use of a paired value of fasting plasma glucose and glycated hemoglobin in predicting the likelihood of having diabetes. *Diabetes Care*. 1998;21:2032–2033.
16. Davies and Day., Screening for non-insulin-dependent diabetes mellitus (NIDDM): how often should it be performed? *J Med Screen*. Apr;1(2):78-81 (1994).
17. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report

- of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*. 20:1183–1197 (1997).
18. CDC Diabetes Cost-Effectiveness Study Group: The cost-effectiveness of screening for type 2 diabetes. *JAMA* 280:1757–1763 (1998).
 19. Smith, . Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. *Mass Spectrom Rev* 24:661-700 (2005).
 20. Phillips et al., Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet*. Jun 5;353(9168):1930-3 (1999).
 21. Turner et al., A longitudinal study of breath isoprene in healthy volunteers using selected ion flow tube mass spectrometry (SIFT-MS). *Physiol Meas*. Jan;27(1):13-22 (2006).
 22. Endre et al., Breath ammonia and trimethylamine allow real-time monitoring of haemodialysis efficacy. *Physiol Meas*. 2011 Jan;32(1):115-30.
 23. Obuchowski, Sample size calculations in studies of test accuracy. *Stat Methods Med Res*. Dec;7(4):371-92 (1998).
 24. American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*. 2012. 35:S64-S71.
 25. SAS Institute Inc. Cary, NC, USA.
 26. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
 27. Risby and Solga, Current status of clinical breath analysis. *Applied Physics*. B 85, 421-426 (2006).
 28. Rooth and Ostenson, Acetone in alveolar air and the control of diabetes. *Lancet ii*.

- 1102-1105 (1966).
29. Davies et al., Quantitative analysis of ammonia on the breath of patients in end-stage renal failure. *Kidney International* 52: 223-228 (1997).
 30. Hyspler et al., Determination of isoprene in human expired breath using solid-phase microextraction and gas chromatography-mass spectrometry. *Journal of Chromatography B* 739: 183-190 (2000).
 31. Hansel et al., Proton-transfer reaction mass spectrometry – Online trace gas analysis at the ppb level. *International Journal of Mass Spectrometry and Ion Processes* 150, 609-619 (1995).
 32. Smith and Spanel, Selected ion flow tube mass spectrometry (SIFT-MS) for online trace gas analysis. *Mass Spectrometry Reviews* 24(5) 661-700 (2005).
 33. SIFT-MS Voice 200 data sheet. Syft Technologies Ltd 2012.
 34. Wehinger et al., Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas. *International Journal of Mass Spectrometry* 265 49-59 (2007).
 35. Proton-transfer reaction mass spectrometry – Online trace gas analysis at the ppb level. *International Journal of Mass Spectrometry and Ion Processes* 150, 609-619 (1995).
 36. Endre et al., Breath ammonia and trimethylamine allow real-time monitoring of haemodialysis efficacy. *Physiology Measure.* 32 115-130 (2011).
 37. Tibshirani, Regression shrinkage and selection via the Lasso. *Journal of the royal statistical Society Series B – Methodological* 58: 267-288 (1996).
 38. Casini et al., Lipid-Peroxidation and Cellular-Damage in Extrahepatic Tissues of Bromobenzene-Intoxicated Mice. *American Journal of Pathology* 123: 520-531

(1986).

39. Dunnick et al., Critical pathways in heart function: Bis(2-chloroethoxy)methane-induced heart gene transcript change in F344 rats. *Toxicologic Pathology* 34: 348-356 (2006).
40. Friedman et al., Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33: 1-22 (2010).
41. Li, Sliced inverse regression for dimension reduction (with discussion). *J Am Stat Assoc* 1991, 86:316-342 (1991).
42. Wang et al., New insights into old methods for identifying causal rare variants, *BMC Proceedings* 5(Suppl 9):S50 (2011).
43. Wang et al., Bioinformatics facilities breath analysis by Selected Ion Flow Tube Mass Spectrometry (SIFT-MS): a study of dialysis efficacy. Working paper (2012a).
44. Wang et al., Classification of disease types using full-scanned masses by selected ion flow tube - mass spectrometry (SIFT-MS). Working paper (2012b).

CLAIMS

What is claimed is:

1. A method of diagnosing a disease in a subject, the method comprising:
providing a sample obtained from the exhaled breath of a subject;
obtaining at least one mass spectrum of the sample using selected ion flow tube mass spectrometry (SIFT-MS);
analyzing the at least one mass spectrum using a multivariate statistical method;
and
diagnosing the presence or absence of the disease in the subject based on the analysis.
2. The method of claim 1, wherein obtaining the at least one mass spectrum comprises obtaining more than one mass spectra, each of the more than one mass spectra obtained using a different precursor ion.
3. The method of claim 2, wherein the precursor ion is selected from H_3O^+ , NO^+ , and O_2^+
4. The method of claim 2, wherein the obtaining more than one mass spectrum comprises obtaining a mass spectrum using H_3O^+ as the precursor ion, obtaining a mass spectrum using NO^+ as the precursor ion, obtaining a mass spectrum using O_2^+ as the precursor ion; and a combination of any of the foregoing.
5. The method of claim 1, wherein the at least one mass spectrum is obtained over a range of m/z 10 to m/z 300.

6. The method of claim 1, wherein analyzing the at least one mass spectrum comprises analyzing selected masses using a linear regression method.
7. The method of claim 6, wherein the selected masses are correlated with the disease.
8. The method of claim 6, wherein the selected masses number at least 3.
9. The method of claim 6, wherein the selected masses are identified using a multivariate statistical analysis to correlate a mass spectrum with the disease.
10. The method of claim 9, wherein the multivariate statistical analysis is selected from penalized regression analysis (lasso), texture analysis, spectral analysis, fractal analysis, and a combination of any of the foregoing.
11. The method of claim 9, wherein correlating a mass spectrum with a disease comprises correlating a level of a biomarker of the disease.
12. The method of claim 1, wherein using the multivariate statistical method comprises:
using a regression method to identify selected masses for subsequent analysis; and
analyzing the selected masses using the linear regression method.
13. The method of claim 12, wherein the regression method comprises using a penalized regression method (lasso).

14. The method of claim 1, wherein using the multivariate statistical method to correlate comprises analyzing the at least one mass spectrum of the exhaled breath from a population of subjects correlated to the presence or absence of the disease.

15. The method of claim 1, wherein using the multivariate statistical method to correlate comprises analyzing the at least one mass spectrum of the exhaled breath from a population of subjects correlated to a biomarker of the disease.

16. The method of claim 1, wherein the disease is renal failure and the biomarker is selected from creatinine, urea, and a combination thereof.

17. The method of claim 1, wherein the disease is diabetes and the biomarker is selected from acetone, body mass index, glucose levels, HbA1c, and a combination of any of the foregoing.

18. The method of claim 1, wherein diagnosing comprises comparing the analyzed at least one mass spectrum from the subject with a corresponding analysis of exhaled breath samples from a population of healthy patients.

19. The method of claim 1, wherein the disease is selected from a cardiovascular disease, diabetes, pre-diabetes, a renal disease, a kidney disease, a cancer, stroke, an infectious disease, asthma, drug addiction, a gastroenterological disease, a neuro-degenerative disease, Alzheimer's disease, a cognitive disease, and depression.

20. The method of claim 1, wherein the disease is selected from renal failure and

diabetes.

21. The method of claim 1, wherein:

the disease is renal failure;

obtaining at least one mass spectrum comprises obtaining a first mass spectrum using NO^+ as the precursor ion, obtaining a second mass spectrum using H_3O^+ as the precursor ion, and obtaining a third mass spectrum using O_2^+ as the precursor ion;

the selected masses for the first mass spectrum comprise m/z 26, m/z 104, m/z 180, and m/z 185;

the selected masses for the second mass spectrum comprise m/z 148; and

the selected masses for the third mass spectrum comprise m/z 41.

22. The method of claim 1, wherein:

the disease is diabetes;

obtaining at least one mass spectrum comprises obtaining a first mass spectrum using NO^+ as the precursor ion, and obtaining a second mass spectrum using H_3O^+ as the precursor ion;

the selected masses for the first mass spectrum comprise m/z 150, m/z 152, m/z 173, m/z 29, m/z 52, m/z 14=26; and

the selected masses for the second mass spectrum comprise m/z 156, m/z 39, and m/z 108.

23. A method of diagnosing renal failure in a subject, the method comprising:

providing a sample obtained from the exhaled breath of a subject;

obtaining at least one mass spectrum of the sample using selected ion flow tube mass spectrometry (SIFT-MS), wherein obtaining at least one mass spectrum comprises

obtaining a first mass spectrum using NO^+ as the precursor ion, obtaining a second mass spectrum using H_3O^+ as the precursor ion, and obtaining a third mass spectrum using O_2^+ as the precursor ion;

analyzing the at least one mass spectrum using a multivariate statistical method, wherein using the multivariate statistical method comprises analyzing the selected masses using a linear regression method, wherein,

the selected masses for the first mass spectrum comprise m/z 26, m/z 104, m/z 180, and m/z 185;

the selected masses for the second mass spectrum comprise m/z 148; and

the selected masses for the third mass spectrum comprise m/z 41; and

diagnosing the presence or absence of renal failure in the subject based on the analysis.

24. A method of diagnosing diabetes in a subject, the method comprising:

providing a sample obtained from the exhaled breath of a subject;

obtaining at least one mass spectrum of the sample using selected ion flow tube mass spectrometry (SIFT-MS), wherein obtaining at least one mass spectrum comprises obtaining at least one mass spectrum comprises obtaining a first mass spectrum using NO^+ as the precursor ion, and obtaining a second mass spectrum using H_3O^+ as the precursor ion;

analyzing the at least one mass spectrum using a multivariate statistical method, wherein using the multivariate statistical method comprises analyzing the selected masses using a linear regression method, wherein,

the selected masses for the first mass spectrum comprise m/z 150, m/z 152, m/z 173, m/z 29, m/z 52, m/z 14=26; and

the selected masses for the second mass spectrum comprise m/z 156, m/z 39, and m/z 108; and

diagnosing the presence or absence of diabetes in the subject based on the analysis.

25. A method of treating a disease in a patient, comprising:
determining that the patient has the disease using the method of claim 1; and
administering a treatment to the patient for treating the disease.

26. A method of treating a disease in a patient, comprising:
requesting a test according to the method of claim 1 to determine whether a patient
has the disease; and
administering a treatment to the patient to treat the disease if the patient is diagnosed
as having the disease.

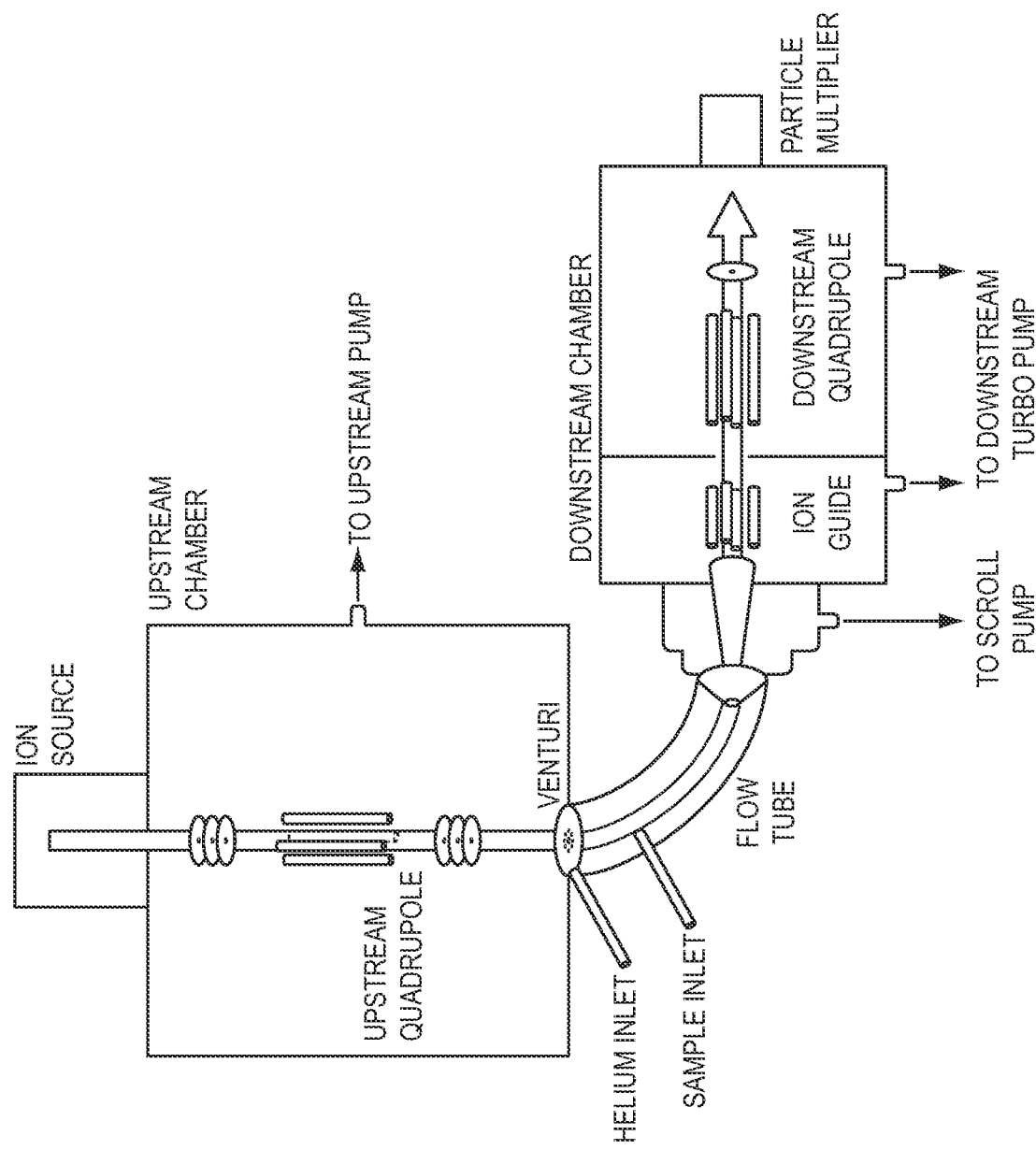


FIG. 1

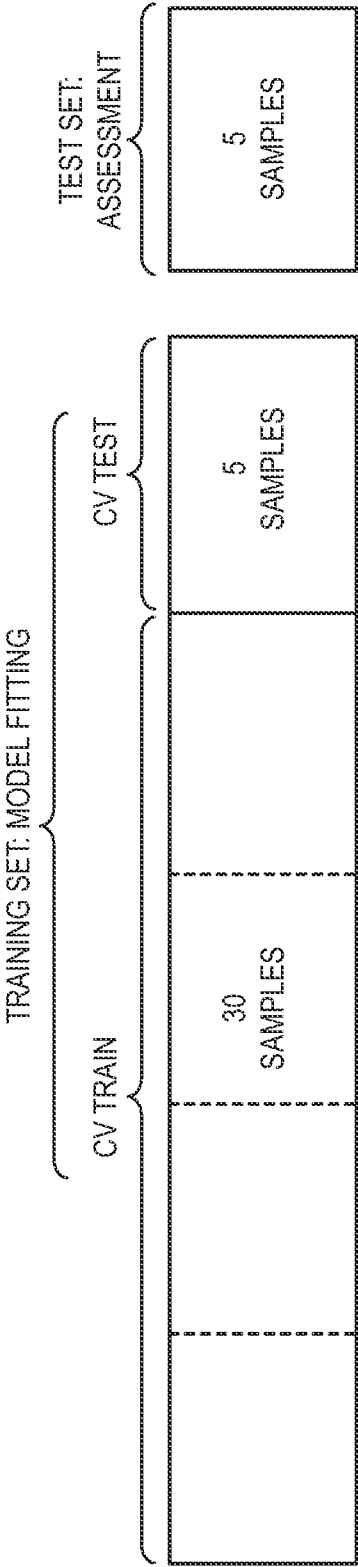


FIG. 2

3/10

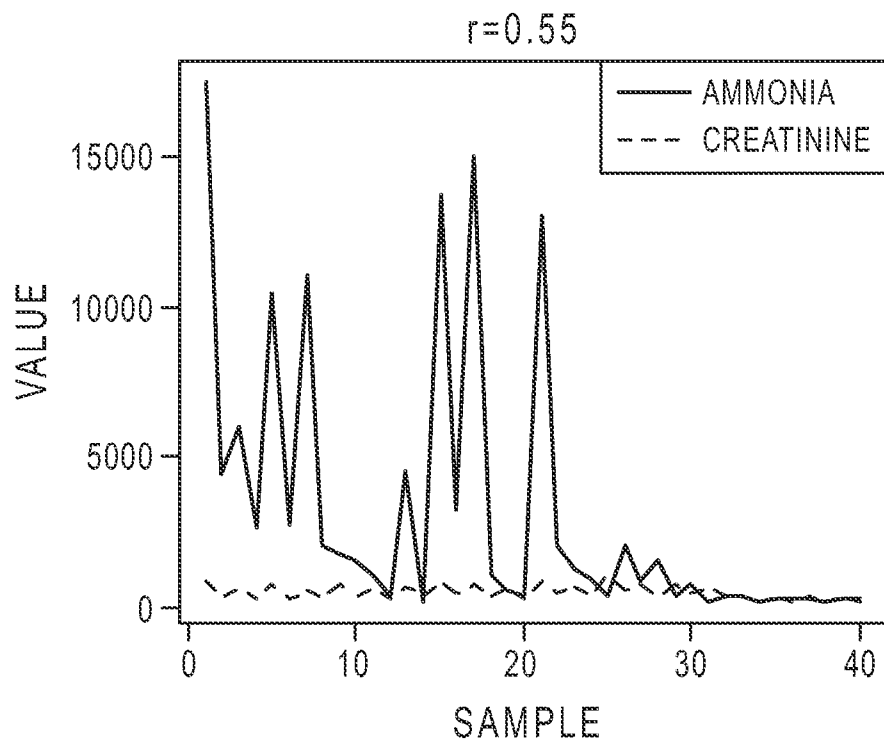


FIG. 3A

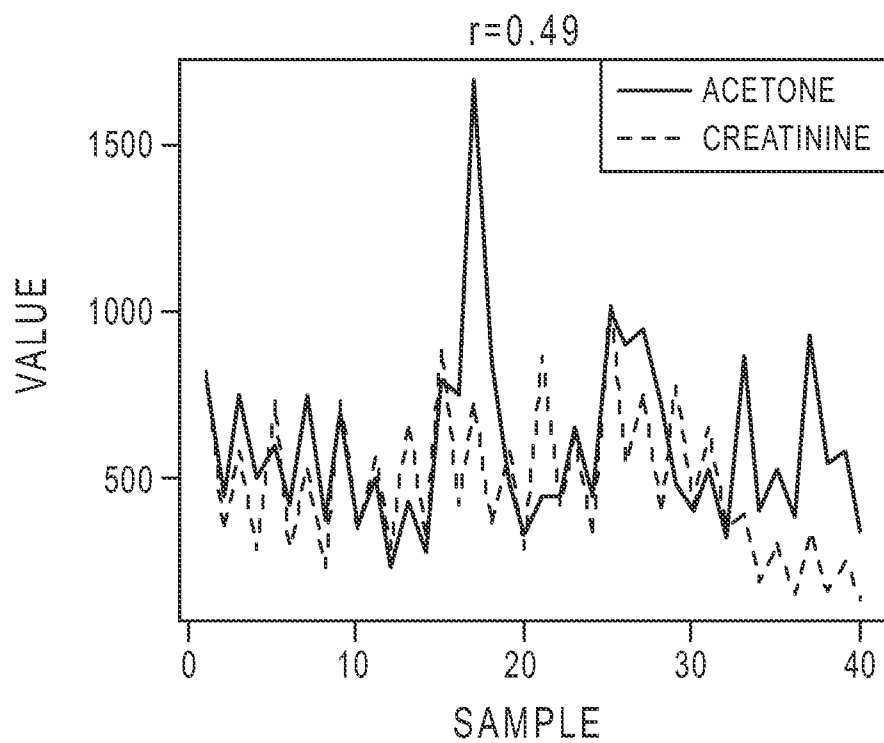


FIG. 3B

4/10

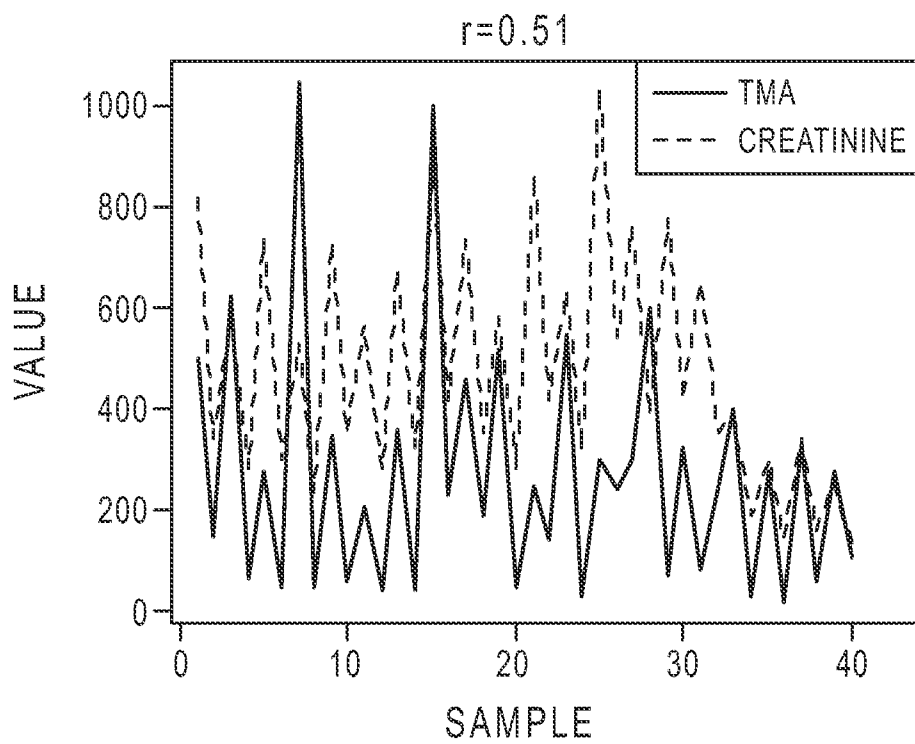


FIG. 3C

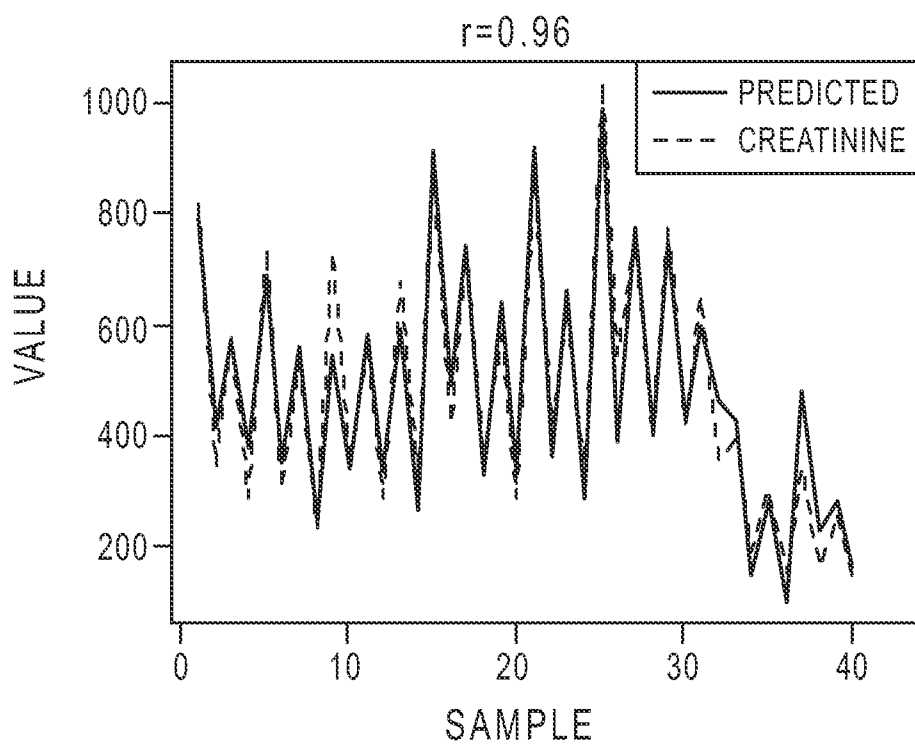


FIG. 3D

5/10

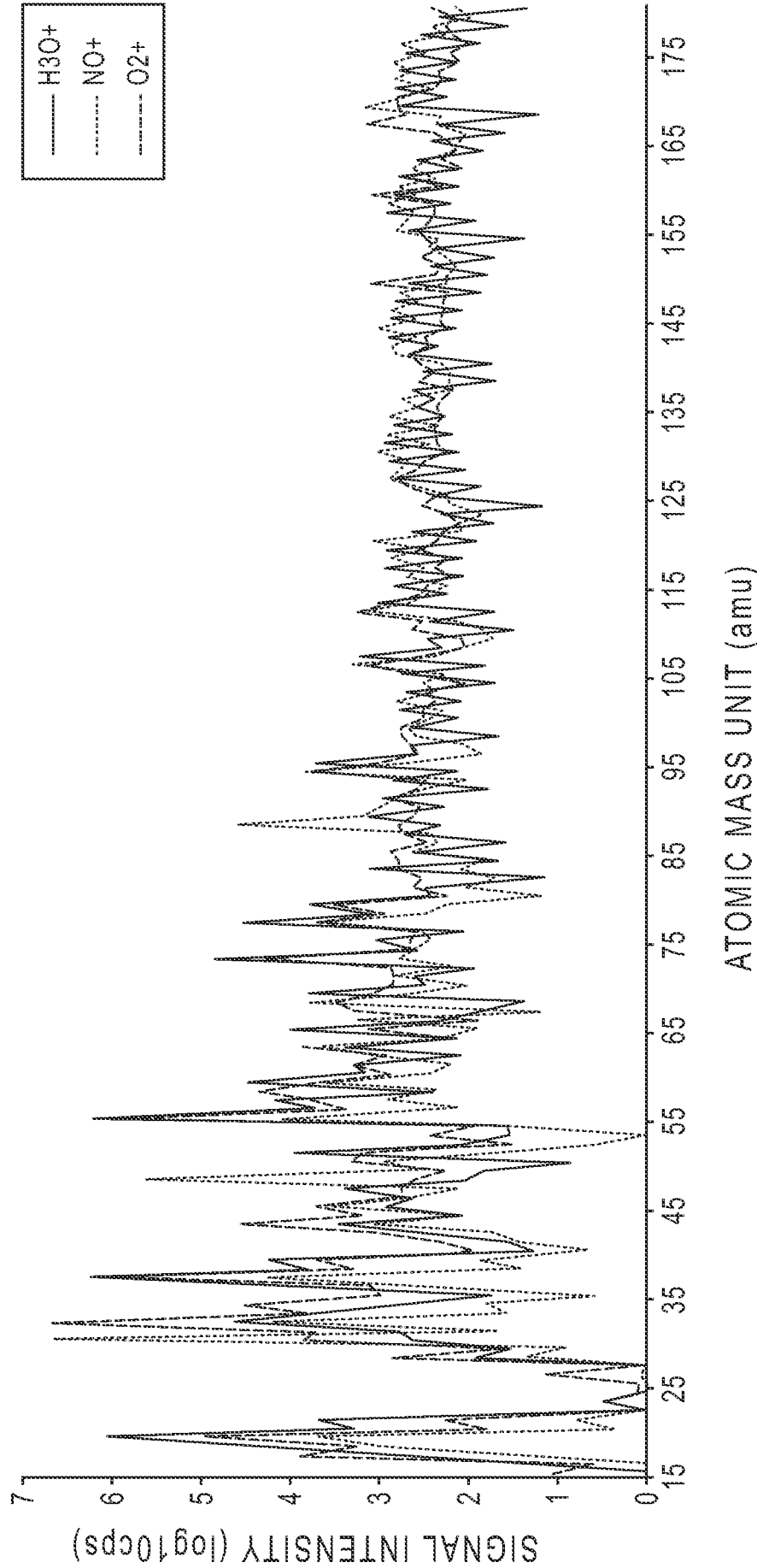


FIG. 4

FIG. 5A

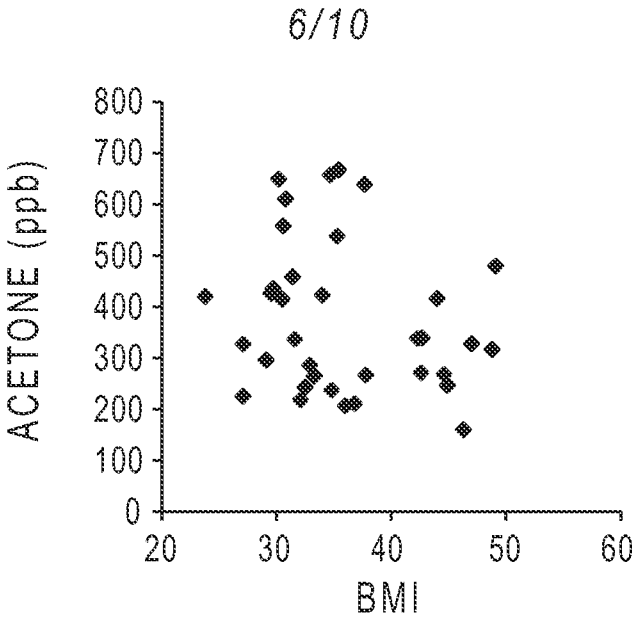


FIG. 5B

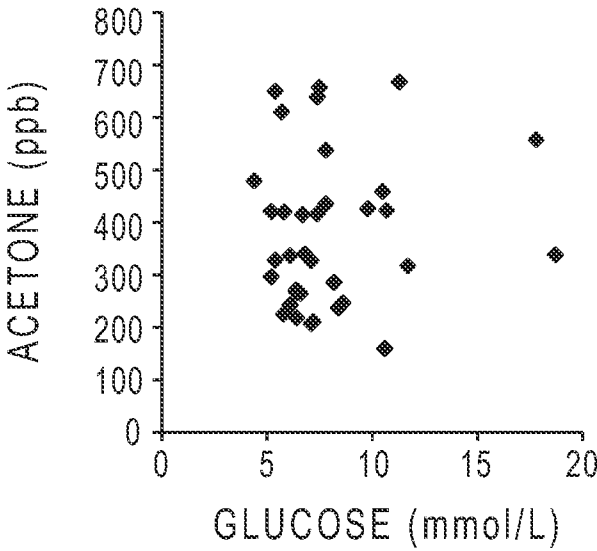
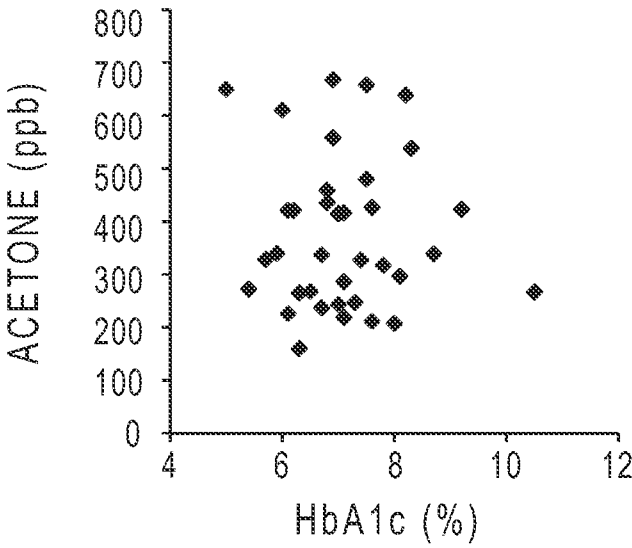


FIG. 5C



7/10

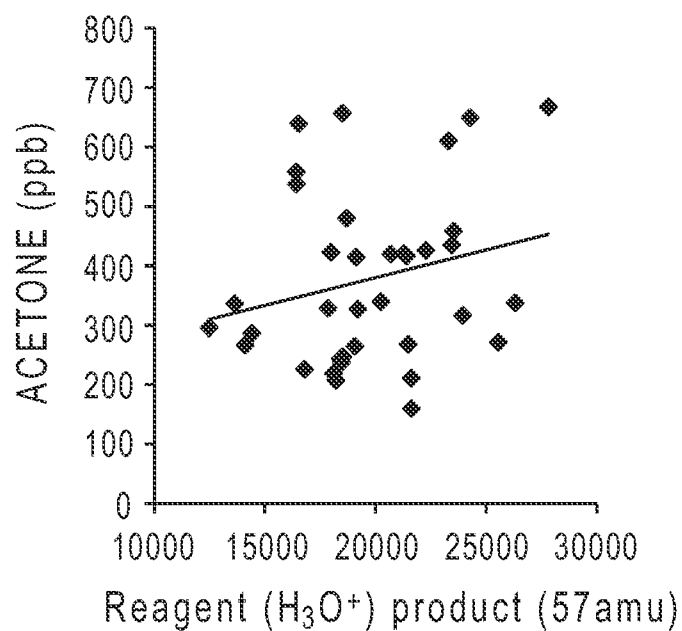


FIG. 6A

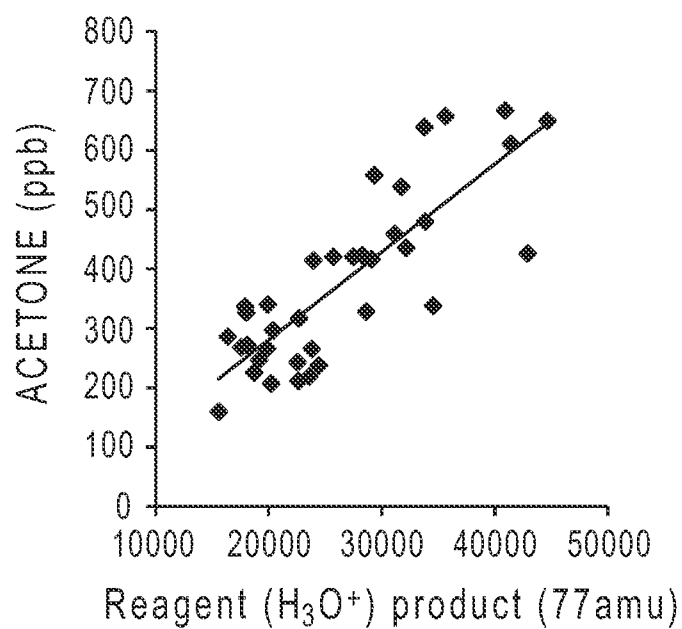


FIG. 6B

8/10

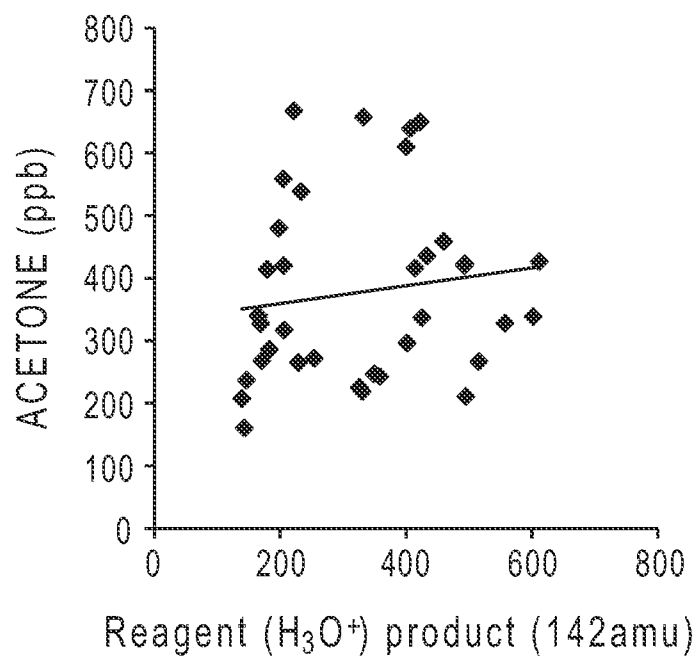


FIG. 6C

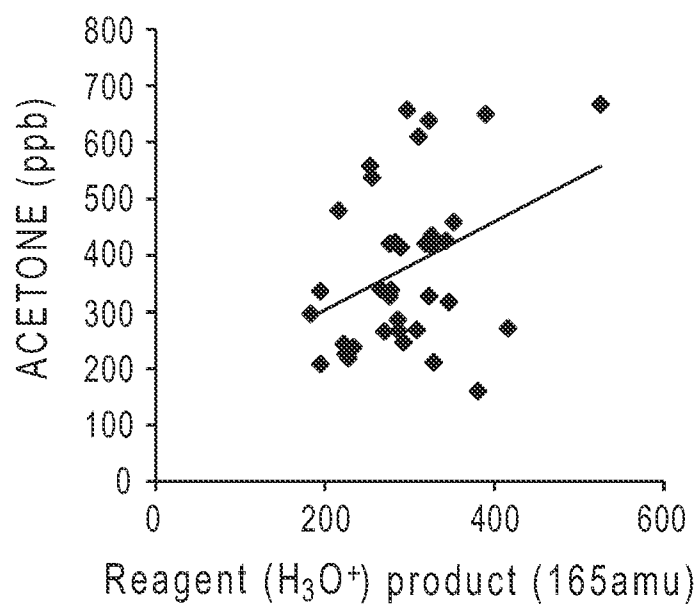


FIG. 6D

9/10

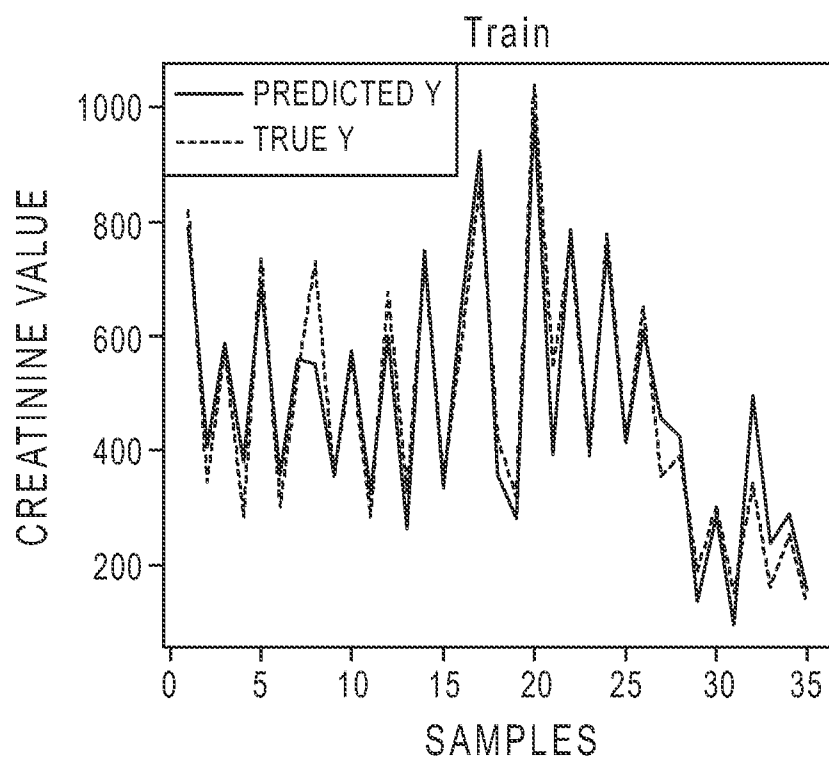


FIG. 7A

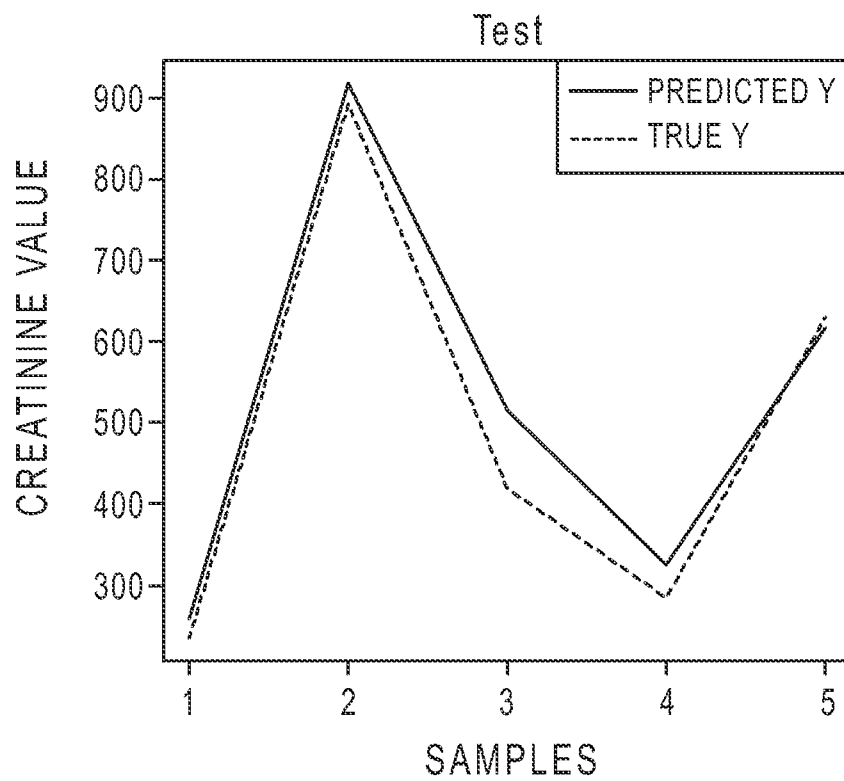


FIG. 7B

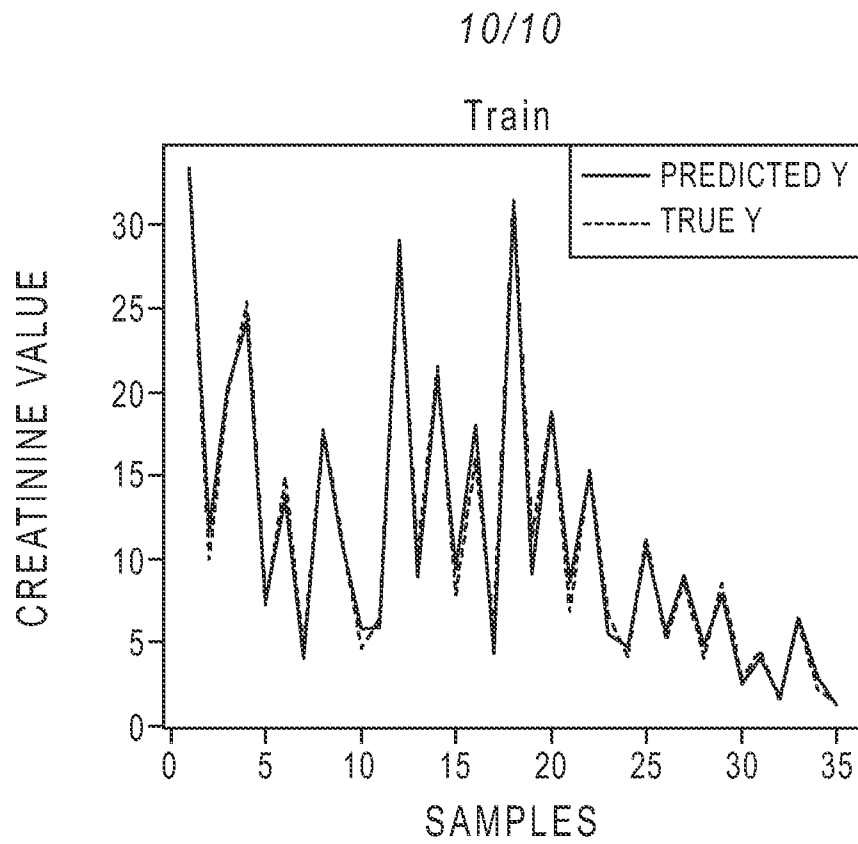


FIG. 8A

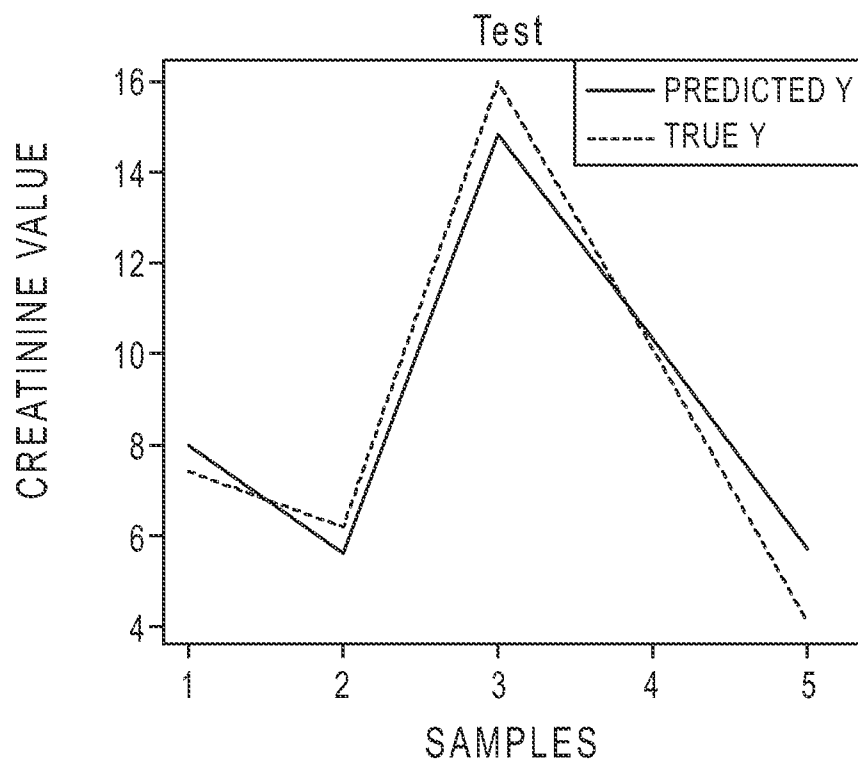


FIG. 8B

PATENT COOPERATION TREATY

PCT

DECLARATION OF NON – ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

(PCT Article 17(2)(a), Rules 13 *ter*.1(c) and 39)

Applicant's agent's reference 14C50076	IMPORTANT DECLARATION	Date of mailing (<i>day/month/year</i>) 03 Apr. 2014 (03.04.2014)
International application No. PCT/CN2014/071853	International filing date (<i>day/month/year</i>) 30 Jan. 2014(30.01.2014)	(Earliest)Priority date(<i>day/month/year</i>) 01 Feb. 2013(01.02.2013)
International Patent Classification (IPC) or both national classification and IPC A61B 5/103 (2006.01) i		
Applicant THE CHINESE UNIVERSITY OF HONG KONG		

This International Searching Authority hereby declares, according to Article 17(2)(a), that **no international search report will be established** on the international application for the reasons indicated below.

1. ☒ The subject matter of the international application relates to:
 - a. ☐ scientific theories
 - b. ☐ mathematical theories
 - c. ☐ plant varieties
 - d. ☐ animal varieties
 - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes
 - f. ☐ schemes, rules or methods of doing business
 - g. ☐ schemes, rules or methods of performing purely mental acts
 - h. ☐ schemes, rules or methods of playing games
 - i. ☒ methods for treatment of the human body by surgery or therapy
 - j. ☐ methods for treatment of the animal body by surgery or therapy
 - k. ☒ diagnostic methods practised on the human or animal body
 - l. ☐ mere presentations of information
 - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:

☐ the description
 ☐ the claims
 ☐ the drawings
3. ☐ A meaningful search could not be carried out without the sequence listing; the applicant did not, within the prescribed time limit:

☐ furnish a sequence listing on paper complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in a form and manner acceptable to it.
☐ furnish a sequence listing in electronic form complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in a form and manner acceptable to it.
☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13ter.1(a) or (b).

4. Further comments:

The subject matter of claims 1-22 relates to a method of diagnosing a disease in a subject, the subject matter of claim 23 relates to a method of diagnosing renal failure in a subject, the subject matter of claim 24 relates to a method of diagnosing diabetes in a subject, and the subject matter of claims 25-26 relates to a method of treating a disease in a patient. Therefore, according to the criteria set out in Rule 39.1(iv), the subject matter of claims 1-26 relates to subject matter for which an international search is not required.

Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer GAO, Hong Telephone No. (86-10) 62413522
---	--