Spectroscopic approach for dynamic bioanalyte tracking with minimal concentration information

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
COMPUTATIONAL MODELS
RAMAN SPECTROSCOPY

Received
9 July 2014

Accepted
14 October 2014

Published
12 November 2014

Correspondence and requests for materials should be addressed to R.R.D. (rrdasari@mit.edu)

* These authors contributed equally to this work.

† Current address: EMC Corporation, Hopkinton, MA, 01748.

Spectroscopic approach for dynamic bioanalyte tracking with minimal concentration information

Nicolas Spegazzini[1,2]*, Ishan Barman[3,4]*, Narahara Chari Dingari[1]*†, Rishikesh Pandey[1], Jaqueline S. Soares[5], Yukihiro Ozaki[2] & Ramachandra Rao Dasari[1]

[1]Laser Biomedical Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, [2]Department of Chemistry, School of Science and Technology, Kwansei Gakuin University, Sanda, Hyogo 669-1337, Japan, [3]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA, [4]Department of Oncology, Johns Hopkins University, Baltimore, Maryland 21287, USA, [5]Departamento de Física, Universidade Federal de Ouro Preto, Ouro Preto, MG 35400-000, Brazil.

Vibrational spectroscopy has emerged as a promising tool for non-invasive, multiplexed measurement of blood constituents - an outstanding problem in biophotonics. Here, we propose a novel analytical framework that enables spectroscopy-based longitudinal tracking of chemical concentration without necessitating extensive *a priori* concentration information. The principal idea is to employ a concentration space transformation acquired from the spectral information, where these estimates are used together with the concentration profiles generated from the system kinetic model. Using blood glucose monitoring by Raman spectroscopy as an illustrative example, we demonstrate the efficacy of the proposed approach as compared to conventional calibration methods. Specifically, our approach exhibits a 35% reduction in error over partial least squares regression when applied to a dataset acquired from human subjects undergoing glucose tolerance tests. This method offers a new route at screening gestational diabetes and opens doors for continuous process monitoring without sample perturbation at intermediate time points.

Blood constituent (analyte) monitoring forms a substantial component of medical diagnostics, ranging from critical-care to point-of-care testing. The concentration levels of these analytes are tightly controlled under normal circumstances and thus any deviation from the well-established ranges can be immediately correlated with an abnormality in body function. Formulation and advance of non-invasive, continuous measurement strategies for such analytes - particularly glucose in diabetic patients[1,2] - is highly desirable, given the significant challenges and inconvenience associated with multiple blood withdrawals per day. Furthermore, such a measurement technology would significantly aid neonatal and ICU patient monitoring as well as the screening for pre-diabetes and gestational diabetes. Currently, the latter pathological conditions are diagnosed via functional loading tests (*e.g.* the oral glucose tolerance test (OGTT)[3]), where the insulin action is monitored by discrete finger-prick measurements over the duration of a few hours following an initial glucose stimulus.

To address this unmet clinical need for non-invasive, continuous measurement of blood analytes, vibrational spectroscopy, especially infrared (IR) absorption and Raman[4–6], has been proposed by researchers due to its ability to quantify biochemical composition of the blood-tissue matrix without necessitating addition of exogenous labels. Raman spectroscopy, in particular, has been exploited due to its exquisite chemical specificity emanating from the characteristic frequency shifts of the photons following its interaction with the matrix molecule(s). This provides an inherent advantage in targeted analysis of a specific bioanalyte as the congestion among the broad overlapping features in IR absorption spectra often washes out the information of interest. To gainfully employ spectroscopic techniques in bioanalyte concentration prediction, chemometric methods, such as partial least squares (PLS) regression[7] and support vector regression (SVR)[8], are employed to develop calibration models from representative samples. The multivariate calibration models are then used in combination with the spectrum acquired from a prospective sample to compute the bioanalyte concentration in that sample.

Despite promising measurements of clinically relevant analytes (*e.g.* glucose, urea and cholesterol) in aqueous solutions[9] and whole blood samples[10], the translation of spectroscopic techniques to *in vivo* measurements in humans has proven to be challenging. The primary impediments for clinical translation has been attributed to sample-to-sample variability in optical properties, such as those due to variations in skin-layer thickness and

hydration state[11], and in physiological characteristics[12]. In view of the substantial inter-person variance, an alternate route in establishing the potential of vibrational spectroscopy would be to perform time-lapse measurements (in a continuous or semi-continuous manner) on a single individual. Specifically, it would be beneficial if temporal evolution of the concentration profile could be obtained solely from spectral acquisitions without resorting to (intermediate) concentration measurements. This would allow for minimum sample perturbation be it in a biomedical setting or in chemical reaction monitoring. Although the utility of such a protocol, which can function with little or no concentration information, is indisputable, there is currently a lack of analytic frameworks that can operate solely based on the acquired spectroscopic and sample-specific kinetic information.

In this article, we propose a novel analytical formulation that enables spectroscopy-based prediction of analyte information, without necessitating reference concentration information for the development of the calibration model. The proposed framework is hereafter referred to as the improved concentration independent calibration (iCONIC) approach. We seek to solve this inverse concentration estimation problem by incorporating the kinetic model of the system to guide the spectroscopy-based concentration estimates. In other words, the kinetic model of the process provides a guide to the "missing" concentration piece of the inverse problem of concentration estimation. While the fundamental principles of the iCONIC approach are generalizable to any spectroscopy-based quantification study, this work focuses on the development and application of the iCONIC framework using non-invasive glucose monitoring as the paradigm. Here, we characterize the physiological lag between the blood and interstitial fluid (ISF) glucose concentrations using a two-compartment mass transfer framework, which has been employed to model the analyte transport by us and others[13–15]. Inspired by indirect implicit calibration ideas[16], minimization of the spectral information and the output of the kinetic model is then pursued in the concentration domain. The spectroscopic calibration step is executed inside the kinetic parameter estimation loop in an iterative fashion. This considerably alleviates the rigidity associated with prior methods that sought to determine a simultaneous solution to the kinetic modeling and the spectroscopic calibration components[15].

Using concentration datasets obtained from a series of OGTTs in human subjects, we demonstrate the potential of the iCONIC approach in estimating blood glucose concentrations. We show that the iCONIC estimates conform more closely to the measured values in relation to the predictions computed from conventional PLS calibration that shows larger deviations. Additionally, this study also provides quantitative insights into the subject's physiological lag characteristics potentially offering a new tool for the personalized assessment of diabetes onset and progression. Collectively, these findings open the door for a diverse range of spectroscopic monitoring applications - especially in clinical practice where obtaining intermediate concentration information is always challenging and often impossible.

Results

Dynamic bioanalyte tracking by calibration-free approach. Motivated by the need for a spectroscopy-based monitoring algorithm that can work with limited or no reference concentration inputs, we explore the powerful, yet relatively underutilized, idea of indirect implicit calibration[16,17] and report its first application to quantitative biological spectroscopy. Spectroscopy-based inference of concentration of system constituents belongs to the class of inverse, ill-posed problems, in the sense that there can be multiple solutions that are consistent with the experimental data[18]. Additionally, tracking the temporal evolution of a constituent necessitates analysis of the spectral time series often incorporating conservation equations of differential nature and constitutive equations of algebraic nature into the spectroscopic calibration framework. Continuous spectroscopy-based non-invasive glucose monitoring offers a representative case study, due to the physiological dynamics of glucose transport between the blood and ISF compartments. Specifically, the time lag between the two glucose levels gives rise to an inconsistency in classical spectroscopic calibration models, as the spectroscopic measurements primarily probe ISF glucose while blood glucose values are used as reference inputs[19]. This problem is particularly exacerbated when measurements are performed during rapid changes in glucose levels such as immediately after a meal ingestion (as is the case for OGTT) or insulin administration. Fig. 1 schematically illustrates the spectroscopic measurement process and shows how the photons interact with the glucose molecules in the two distinct compartments. We introduce here a novel route to address this important problem – and related class of monitoring applications – by minimizing the residual between two concentration profiles, namely the profile computed from the kinetic model and that obtained from transformation of the spectral information to the concentration domain.

In the proposed iCONIC approach, calibration of the acquired spectra is done using the concentrations calculated with iteratively improved kinetic parameter(s) – and thus does not require the actual measurement of the reference concentration values as detailed below. Here, the calculated concentrations, $\hat{\mathbf{C}}$, are considered to be "measured" variables and the residual-minimization is performed in concentration units. This represents a multivariate calibration framework with "floating data"[16]. Using concentration values $\hat{\mathbf{C}}$ (which is a function of the kinetic parameters, $\mathbf{k}$) and the recorded calibration spectral matrix $\mathbf{Y}$, one can compute the corresponding regression matrix $\mathbf{B}$ using the least-squares solution to equation (1):

$$\hat{\mathbf{C}} = \mathbf{YB} + \mathbf{E} \tag{1}$$

where $\mathbf{E}$ denotes the noise (error) in measurements.

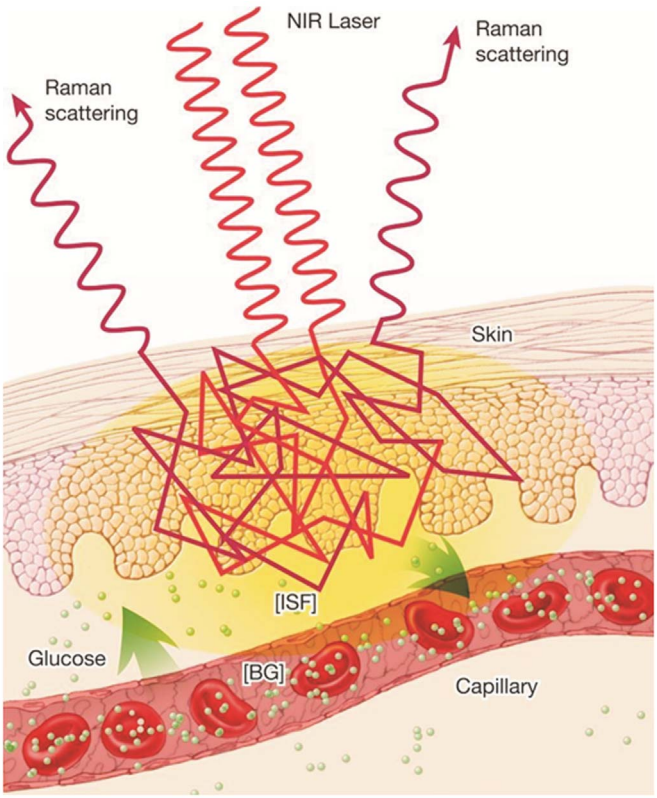

Figure 1 | A schematic illustration of the Raman spectroscopic measurement process for *in vivo* continuous glucose monitoring.

Given the underdetermined nature of the system (as it has more variables, i.e. wavelengths, than equations, i.e. number of calibration data points), solution of the above equation implies calculation of a suitable pseudo-inverse of $\mathbf{Y}$, such that $\hat{\mathbf{B}} = \mathbf{Y}^{+}\hat{\mathbf{C}}$ where $\hat{\mathbf{B}}$ represents the regression matrix estimate. The regression matrix estimate is obtained using singular value decomposition (SVD), PLS or principal component regression (PCR). The calibrated concentration profiles $\hat{\mathbf{C}}_{cal}$ is then determined by substitution in equation (1):

$$\hat{\mathbf{C}}_{cal} = \mathbf{Y}\mathbf{Y}^{+}\hat{\mathbf{C}} \tag{2}$$

This formulation is employed to iteratively obtain the estimates of the kinetic parameters, $\mathbf{k}$, by minimizing the following residual:

$$\mathbf{Q} = \left\|\hat{\mathbf{C}} - \hat{\mathbf{C}}_{cal}\right\| = \left\|\hat{\mathbf{C}} - \mathbf{Y}\mathbf{Y}^{+}\hat{\mathbf{C}}\right\| = \left\|\left(\mathbf{I} - \mathbf{Y}\mathbf{Y}^{+}\right)\hat{\mathbf{C}}\right\| \tag{3}$$

Equation (3), notably, employs two altogether different concentration profiles: $\hat{\mathbf{C}}$, the concentrations computed based on the conservation equations governing the dynamic process; and $\hat{\mathbf{C}}_{cal}$, the spectroscopy-based concentration estimate obtained from the calibration step. Evidently, both concentrations are dependent on the current value of the kinetic parameters.

In order to reduce the impact of spectral baseline fluctuations and improve the contribution of each component of the spectral data during fitting, SVD of the spectral dataset $\mathbf{Y}$ is used to isolate the important time-trace information. Detailed explanation of the specific SVD procedure and its ability to alleviate the pernicious effect of baseline shifts is provided in Supplementary Notes. Reducing $\mathbf{Y}$ to $\mathbf{Y}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{*}$ (where $\mathbf{U}$ is the abstract time-trace matrix of concentration information, $\boldsymbol{\Sigma}$ is the diagonal matrix containing the singular values and $\mathbf{V}^{*}$ is the abstract matrix of basis spectra), and replacing it in equation 3, we obtain:

$$\mathbf{Q} = \left\|\left(\mathbf{I} - \mathbf{U}\mathbf{U}^{T}\right)\hat{\mathbf{C}}\right\| \tag{4}$$

While equation (4) represents the general framework for analysis of any time-resolved spectral data recorded from a dynamic system, the ensuing solution formalism is specialized for spectroscopy-based non-invasive monitoring of blood glucose. Here, the modeled concentration ($\hat{\mathbf{C}}$) and regression ($\hat{\mathbf{B}}$) matrices are replaced by the corresponding ISF glucose-specific vectors ($\hat{\mathbf{c}}_{ISF}$,$\hat{\mathbf{b}}_{ISF}$). Since the acquired spectral data are representative of the ISF glucose concentrations, this ensures consistency in the developed calibration models. Additionally, to remove any remaining ambiguity in the inversion problem, as well as to rule out unphysical and implausible solutions, a secondary convex goal is added by means of a regularization parameter λ[20]. This ensures that the minimization procedure converges on a robust solution in the sense that small variations in the spectral dataset do not cause large variations in the computed kinetic parameters, $\mathbf{k}$, and the resultant regression matrix.

$$\mathbf{Q}_{reg} = \left\|\left[\left(\mathbf{I} - \mathbf{U}\mathbf{U}^{T}\right)\hat{\mathbf{c}}_{ISF}\right]\right\|^{2} + \lambda\|\mathbf{k}\|^{2} \rightarrow \min \tag{5}$$

Where $\hat{\mathbf{c}}_{ISF}$ is assessed from the reverse form of the mass conservation-based model that governs the blood and ISF glucose relationship, as detailed in the Methods section. The residual of equation (5), $\mathbf{Q}_{reg}$, is minimized using the Newton-Gauss-Levenberg/Marquardt (NGLM) algorithm[21] for identification of the optimal kinetic parameters, $\mathbf{k}_{opt}$ (see Supplementary Note 1).

Solution of equation (5) yields the set of optimal ISF glucose concentrations (via $\mathbf{k}_{opt}$), which in turn is used to calculate the ISF glucose-specific regression vector $\hat{\mathbf{b}}_{ISF}$. Using this regression vector in conjunction with the spectrum measured at the prediction time point ($\mathbf{s}_{pred}$), one can predict the ISF glucose concentration:

$$c_{ISF,pred} = (\mathbf{s}_{pred})^{T}\hat{\mathbf{b}}_{ISF} \tag{6}$$

The set of predicted ISF glucose concentrations can be transformed using the forward form of the physiological glucose dynamics model and knowledge of the kinetic parameters $\mathbf{k}_{opt}$ to construct the corresponding blood glucose estimates.

Calculation of Blood Glucose concentration. Fig. 2 shows the mean and ±1 standard deviation (SD) of representative Raman spectra acquired from a human volunteer undergoing an OGTT. The SD to mean ratio of the intensity values over the fingerprint region of the spectrum, 300–1700 cm^{-1}, ranges from 0.03 to 0.1. The tissue spectral signatures can be attributed to the presence of Raman-active components (such as from blood analytes, collagen I and III, structural proteins in the epidermis and dermis, and sub-cutaneous lipids) and endogenous fluorophores. While the near-infrared (NIR) excitation considerably reduces the autofluorescence levels, the presence of a broad background can still be observed in the acquired spectra. The strongest Raman peak is observed at *ca.* 1445 cm^{-1} and other prominent features are located at approximately 859, 938, 1004, 1273, 1302 and 1655 cm^{-1}, which is consistent with prior *in vivo* tissue observations[22]. Expectedly, the Raman bands of glucose are masked in the myriad signals of other constituents and cannot be uniquely assigned by visual inspection alone. This necessitates the use of multivariate algorithms to identify the subtle changes and to link such changes to the glucose concentrations at different time points.

Here, we have used the iCONIC approach to predict the glucose concentrations with only the first reference concentration from each subject being used to develop the model. To understand the efficacy of the proposed method in comparison with more established approaches, PLS calibration was also used to estimate the glucose concentrations based on the acquired Raman spectra. Since PLS calibration (or any other analogous implicit calibration technique such as PCR and SVR[8]) method requires significantly more reference concentrations to build a model, a cross-validation procedure is implemented to test the predictive power of the model. While the leave-one-out cross-validation routine (LOOCV), explained in Methods section, avoids some of the pitfalls encountered in autoprediction, it may yet result in an apparently functional model (due to "overtraining") that cannot be used for prospective prediction[23]. Nevertheless, given the problem constraints, the PLS LOOCV pro-

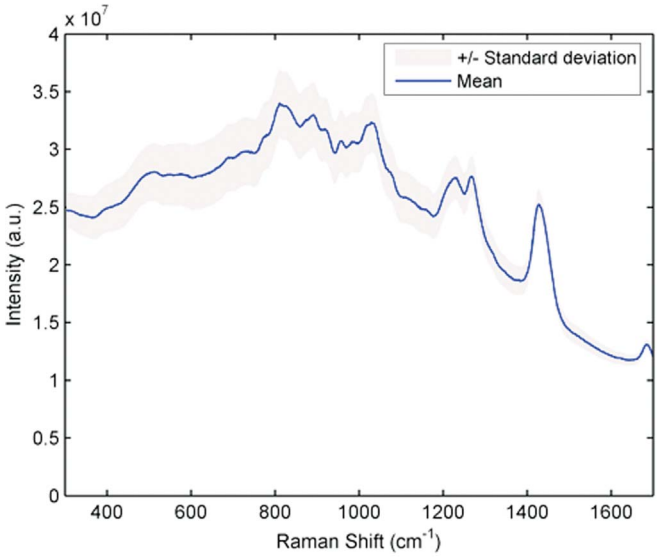

Figure 2 | Representative Raman spectra acquired from a human subject undergoing OGTT. The thick line shows the mean value and the shaded area represent ±1 standard deviation.

cedure provides the best yardstick for comparison while also highlighting the need of an essentially "calibration-free" approach.

Fig. 3 displays the results of iCONIC prediction (red diamond) and PLS LOOCV (black circle) in a representative subject, where the blue squares depict the measured blood glucose values. The measured concentrations show the expected rise in glucose levels due to ingestion of the sugar-rich drink followed by the subsequent recovery to (nearly) euglycemic levels owing to the normal insulin response. The recovery would be delayed or absent if a diabetic subject were tested. We observe that the iCONIC model (root mean squared error of prediction, RMSEP = 5.14 mg/dL) exhibits significantly better prediction accuracy in comparison with the PLS LOOCV estimation (root mean squared error of cross validation, RMSECV = 13.64 mg/dL). The better estimation using the iCONIC approach can be attributed to two factors, namely suitable correction for the physiological lag between blood and ISF glucose, and the isolation of the baseline shifts and system drifts (Supplementary Note 2). The effect of the former can be viewed in the initial 60 minutes when the glucose levels of the subject rise sharply. During this time frame, the iCONIC predictions match considerably better with the reference concentrations, in relation to the PLS estimates, by appropriately modeling the transient discrepancies. As noted in previous studies[15,24], conventional calibration methods exhibit systematic errors during rapid excursions, even in the presence of a positive correlation between blood and ISF glucose.

To better illustrate the predictive power when multiple human subject data sets are included in the analysis, the results of the iCONIC model are plotted on the Clarke error grid (Fig. 4)[25], a widely used method for quantifying the clinical usefulness of glucose predictions. Predictions in zones A and B are regarded as acceptable, and predictions in zones C, D, and E are considered to be potentially dangerous if used for clinical management. The RMSEP and the R^2 value (coefficient of determination) are computed to be 0.54 mM (1 mM of glucose = 18 mg/dL) and 0.97, respectively. Critically, all the glucose predictions over the entire human subject dataset reside in the clinically acceptable regions – even when the glucose levels are relatively low (4–6 mM). This result is of great value as a key benefit of a continuous glucose monitoring system is the real-time detection of hypoglycemic states. A common motif in diabetes care is the lack of immediate knowledge regarding low blood glucose excursions in over-medicated patients resulting in serious consequences including diabetic coma. The ability to non-invasively and continuously estimate blood glucose trends that is predictive of both hypoglycemic and hyperglycemic blood glucose excursions would address this pressing need.

Table 1 summarizes the results of the iCONIC model predictions (viewed in Fig. 4) as well as the corresponding PLS LOOCV estimates. We observe that the reduction in error on application of the iCONIC model, when compared to the PLS model estimates, ranges from nearly 18% to 59% with an average value of 35.5% computed over the 8 subjects. Even when compared with our previous dynamic concentration correction (DCC) model, which provided on average a 16% reduction in prediction error with respect to the corresponding PLS models for the same dataset[21], the iCONIC model demonstrates much better predictive power.

Discussion

Our findings suggest that vibrational spectroscopy in combination with the proposed iCONIC approach can provide continuous glucose tracking information without necessitating substantial invasive blood glucose measurements. In the following, we discuss the validity and efficacy of this information content including the characterization of the glucose diffusion process.

The glucose diffusion process, which was previously modeled using a single lumped parameter[13–15], is now more correctly characterized using a two-parameter model (k_1, k_2). This ensures that the rate of glucose uptake by the subcutaneous tissue is also addressed in the mass diffusion process. In each of the volunteers, k_1 had a larger numerical value in relation to k_2 signifying that the blood glucose rise was faster than the return to euglycemic levels. This is consistent with typical observations in glucose tolerance studies where the increase in blood glucose levels following ingestion of glucose solution is rapid in relation to the subsequent insulin-mediated glucose clearance from the blood (by the cells) and, thus, the corresponding return to normal blood glucose levels. One would anticipate that subjects with impaired glucose tolerance would exhibit significant changes in the determined rate constants, especially k_2.

Critically, this allows us to model situations where, during the time of decreasing glucose levels, ISF glucose may fall in advance of blood glucose and reach nadir values that are lower than the corresponding blood glucose levels[26,27]. Some studies have indicated that ISF glucose levels can remain below blood glucose concentrations for fairly long period of time following correction of insulin-induced hypoglyce-

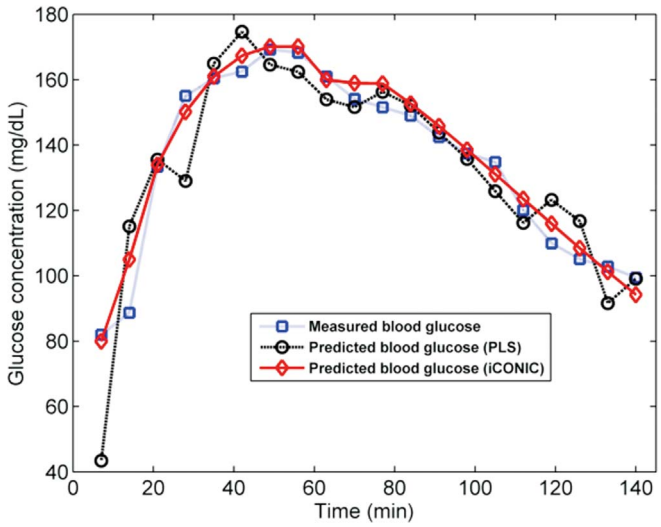

Figure 3 | Plot of prospective prediction (iCONIC, red diamond), LOOCV (PLS, black square) and reference glucose concentrations (blue squares) for a representative human subject.

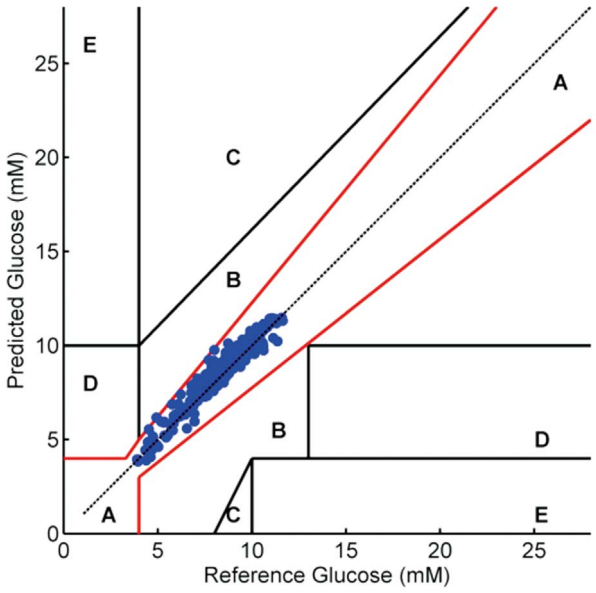

Figure 4 | Blood glucose predictions of the iCONIC model for the complete human subject dataset shown on the Clarke Error Grid.

Table 1 | Summary of PLS LOOCV and iCONIC prediction results for the human subject dataset

		Conventional PLS Model		ICONIC Model			
Subject	No. of Data Points	RMSECV (mM)	RMSEP (mM)	k1	k2	λ	Change in error (%)
1	25	1.17	**0.68**	0.0200	0.0192	0.89	42.15
2	26	0.72	**0.57**	0.0185	0.0178	0.77	20.39
3	26	1.28	**0.53**	0.0157	0.0151	0.67	58.87
4	20	0.76	**0.37**	0.0245	0.0240	0.36	51.31
5	32	0.56	**0.46**	0.0125	0.0120	0.55	17.62
6	25	0.80	**0.64**	0.0230	0.0217	0.38	19.47
7	26	0.83	**0.44**	0.0220	0.0214	0.40	46.55
8	28	0.79	**0.57**	0.0155	0.0148	0.25	27.84

mia[28]. While our limited observations do not appear to support such reports, these findings could be explicated by the so-called push-pull phenomenon, according to which the glucose is pushed from the blood to the ISF compartment during the rising phases and the glucose is recruited from the ISF to the surrounding cells during the falling phases. If this were true for a given process, our model would re-calibrate itself by adjusting the corresponding k_2 value.

Another pertinent question relates to the demonstration of causality of glucose concentration to the acquired spectral information, especially as the intrinsic glucose signal is significantly smaller than that of several other blood-tissue matrix constituents. Moreover, time-dependent physiological processes or variations specific to an instrument that happen to be correlated with the glucose levels have often been found to dominate classical implicit calibration models, especially for non-specific measurement modalities[29]. To investigate the robustness of our predictions to chance correlations (spurious factors), an *F*-test was used to compare the squared error of prediction (SEP) to the standard deviation of the glucose concentrations within the prediction data set (SDP) and, therefore, to assess if the variability of the predicted concentrations is greater than would be expected by chance. Here, from the values listed in Table 2, the *F*-value for the PLS LOOCV estimates and the iCONIC predictions was calculated to be 4.99 and 13.28, respectively. For the PLS computation, the SEP was replaced with SECV. Clearly, both sets of F-values are statistically significant. Thus, the null hypothesis that the variance of errors of glucose predictions is same as the variance of the reference glucose concentrations can be rejected. Table 2 also lists the results for linear regression analysis of the prediction points for the PLS and iCONIC cases. The y-intercept for the iCONIC predictions is lower, the slope is closer to unity, and R^2 value is higher. These results strongly indicate that the current glucose predictions are based on the spectroscopic properties of glucose rather than on chance variations or correlative response between glucose and other matrix constituents. Notably, since the iCONIC models do not use the standard input of an array of reference concentrations, the possibility of building an apparently functional model based on incidental correlations is largely eschewed.

Difference plot analysis was also performed (Supplementary Fig. S1) for further comparison of the methods. The Bland-Altman plots of Supplementary Fig. S1 enable the investigation of the presence of any systematic difference between the reference and Raman measurements and to identify possible outliers. Here, the mean difference is the estimated bias and is found to be 0.5 mg/dL and −4.73 mg/dL for the PLS and iCONIC estimates, respectively. The corresponding 2SD limits are determined to be 31.9 mg/dL (PLS) and 17.1 mg/dL (iCONIC). As per ISO 15197 guidelines, these 2SD limits should be less than 15 mg/dL for glucose concentration below 75 mg/dL and should be lower than 20% for any value higher than 75 mg/dL. Our findings therefore suggest that the combination of the iCONIC approach and Raman spectroscopy provide clinically viable predictions, especially in terms of single-individual prediction.

The results presented in this manuscript provide a proof-of-concept validation of the untapped potential of such a broad and widely generalizable approach. Specific to the problem of glucose monitoring, we envision that the iCONIC model can predict impending hypo- and hyperglycemic excursions potentially allowing the diabetic patient to take necessary corrective action. It can also be gainfully employed in studying physiological changes (for example, in micro- and macro-vasculature) due to the onset of diabetes via its ability to characterize the glucose transport process in the circulation system and in the ISF. A large cohort of normal human volunteers and diabetic patients is currently being studied to test the feasibility of this method across different ages, ethnicities and, critically, in subjects with impaired glucose tolerance characteristics. While the present day standard of care primarily involves interpretation of changes in blood glucose, we believe that in specific cases measurement of ISF glucose levels may be more important clinically such as the persistence of impaired cognition for prolonged periods of time after correction of hypoglycemia.

In the current study, we have proposed the potential of a spectroscopic method for tracking bioanalytes in a dynamic system with minimal *a priori* concentration information. The ability of the iCONIC approach to make accurate predictions in clinical datasets acquired from human subjects is demonstrated in the presence of myriad non-analyte specific variations. The performance metrics of the iCONIC algorithm exceed that of the conventional PLS calibration method, which we attribute to its twin advantages of accounting for the physiological lag between blood and ISF glucose, and avoiding the baseline shifts and system drifts. While the initial pilot studies performed here provide the foundation, further clinical investigations - in single-center and subsequently in multi-center settings - will be pursued to validate the approach. Furthermore, the iCONIC formulation can be readily extended to quantify analytes using other spectroscopic signatures such as infrared absorption and thermal emission, which offer higher sensitivity in comparison to Raman acquisitions[30–32].

Table 2 | Performance characteristics for the PLS and iCONIC models (* indicates SECV is the correct error metric for PLS and is used here)

				Regression analysis of the prediction data		
Model	SDP (mg/dL)	SEP (mg/dL)	Model size	β_0 (mg/dL)	β_1	R^2
PLS	35.57	15.92*	208	13.57	0.9057	**0.9001**
iCONIC	35.57	9.76	208	6.33	0.9893	**0.9718**

Given our findings and the inherent non-invasive nature of vibrational spectroscopy, the combined method would be appropriate as a real-time clinical adjunct for continuous monitoring of glucose and other blood analytes, *e.g.* creatinine, urea and bilirubin, in critical care patients and in neonates, where frequent blood withdrawal is particularly problematic. Application of this minimally perturbative approach would also lay the foundation for a novel blood withdrawal-free spectroscopic assay for glucose tolerance testing in the near future. Moreover, the scope of application of this method extends beyond *in vivo* diagnostics to microfluidics investigations as well as recalcitrant industrial process monitoring, where intermediate sampling of the specimen would compromise its identity.

Methods

Clinical Studies on Human Subjects. To test the capability of the iCONIC approach in predicting concentrations from time-resolved spectra, clinical datasets comprised of blood glucose concentrations and Raman spectra are used. These datasets, which were detailed in one of prior reports[33], were collected from healthy human volunteers undergoing OGTT. Raman spectra were recorded at regular 5 min intervals from the forearms of these volunteers. For Raman spectral acquisition, an 830 nm diode laser (Process Instruments) was used as an excitation source with an average power of *ca.* 300 mW in a $\sim$ 1 mm^2 spot. On the detection end, an f/1.8 spectrograph (Kaiser Optical Systems) was coupled to a liquid nitrogen-cooled CCD (1340 $\times$ 1300 pixels, Roper Scientific). Blood was drawn every 10 min and analyzed using a clinical glucose system (HemoCue, Inc.) to evaluate the subject's response. This study protocol was approved by the MIT Committee on the Use of Humans as Experimental Subjects and written informed consent was obtained from each of the volunteers in the study. All the experiments were carried out in accordance with the approved guidelines by the Committee. Data sets from volunteers exhibiting motional artifacts, inadequate SNR in the acquired spectra, and impaired glucose tolerance characteristics are excluded from our analysis. Additionally, two subjects who underwent double OGTT were also not considered in this study.

Data Analysis. To address clinical concerns arising from glucose monitoring in the subcutaneous interstitial fluid, we have previously developed a DCC method[21]. This method, which incorporates the mass transfer equations governing the diffusion of glucose between the blood and ISF compartments into the spectroscopic framework, provides a greater degree of consistency with the acquired spectra in the calibration model. Here, we re-formulate the solution method to allow for subcutaneous uptake of glucose in the ISF compartment by the cells.

Briefly, the transport of glucose from the blood to the ISF compartment occurs by a diffusion process across an established concentration gradient[19,20]. As detailed in the literature, this process can be mathematically written in the form of the following equation for the glucose component in the ISF space:

$$\frac{d}{dt}(V_{ISF}\, c_{ISF}) = k_{21}\, V_{BG}\, c_{BG} - (k_{12} + k_{02})\, V_{ISF}\, c_{ISF} \qquad (7)$$

where c_{BG}, c_{ISF} are the concentrations of glucose in the blood and ISF compartments, respectively; V_{BG}, V_{ISF} are the volume of blood and ISF in the probed region; k_{21} and k_{12} are the forward and reverse flux rates for glucose transport across the capillaries; and k_{02} is the rate of glucose uptake into the surrounding tissue. This equation can be re-written to the following form by reducing the additional parameters into a two-parameter (k_1, k_2) system:

$$\frac{d}{dt}(c_{ISF}) = k_1\, c_{BG} - k_2\, c_{ISF} \qquad (8)$$

Re-arranging equation (8) forms the forward iCONIC model that is used to compute the blood glucose concentrations based on the ISF glucose values and knowledge of the system parameters. Additionally, integrating equation (8) provides the reverse form of the iCONIC model, which is plugged into equation (5) that in turn performs minimization of the objective function.

For the conventional approach, PLS models were created based on the number of loading vectors that provide the least error in cross-validation[10]. The PLS models did not explicitly address the physiological dynamics issue. Here, LOOCV approach was used to provide concentration estimates, because of the limited number of data points available per individual. In LOOCV, the data from a particular time point is eliminated, and the PLS model developed on all the other points is used to predict the concentration at that time point optimizing agreement with the reference measurement.

1. Brownlee, M. Biochemistry and molecular cell biology of diabetic complications. *Nature* **414**, 813–820 (2001).
2. Zimmet, P., Alberti, K. G. M. M. & Jonathan, S. Global and societal implications of the diabetes epidemic. *Nature* **414**, 782–788 (2001).
3. American Diabetes Association: Standards of Medical Care in Diabetes – 2009. *Diabetes Care* **32**, S13–S61 (2009).
4. Khalil, O. S. Spectroscopic and Clinical Aspects of Noninvasive Glucose Measurements. *Clin. Chem.* **45**, 165–177 (1999).
5. Heise, H. M., Bittner, A. & Marbach, R. Near infrared reflectance spectroscopy for noninvasive monitoring of metabolites. *Clin. Chem. Lab. Med.* **38**, 13–145 (2000).
6. Chaiken, J. *et al.* Effect of hemoglobin concentration variation on the accuracy and precision of glucose analysis using tissue modulated, noninvasive, in vivo Raman spectroscopy of human blood: a small clinical study. *J. Biomed. Opt.* **10**, 031111 (2005).
7. Haaland, D. M. & Thomas, E. V. Partial least squares methods for spectral analysis 1. Relation to other quantitative calibration methods and the extraction of quantitative information. *Anal. Chem.* **60**, 1193–1202 (1988).
8. Barman, I. *et al.* Rapid and accurate determination of tissue optical properties using least-squares support vector machines. *Biomed. Opt. Exp.* **2**, 592–599 (2011).
9. Berger, A. J., Wang, Y. & Feld, M. S. Rapid, noninvasive concentration measurements of aqueous biological analytes by near-infrared Raman spectroscopy. *App. Opt.* **35**, 209–212 (1996).
10. Enejder, A. M. K. *et al.* Blood analysis by Raman spectroscopy. *Opt. Lett.* **27**, 2004–2006 (2002).
11. Barman, I., Singh, G. P., Dasari, R. R. & Feld, M. S. Turbidity-corrected Raman spectroscopy for blood analyte detection. *Anal. Chem.* **81**, 4233–4240 (2009).
12. Boyne, M. S., Silver, D. M., Kaplan, J. & Saudek, C. D. Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor. *Diabetes* **52**, 2790–2794 (2003).
13. Schmidtke, D. W., Freeland, A. C., Heller, A. & Bonnecaze, R. T. Measurement and modeling of the transient difference between blood and subcutaneous glucose concentrations in the rat after injection of insulin. *Proc. Natl. Acad. Sci. USA.* **95**, 294–299 (1998).
14. Freeland, A. C. & Bonnecaze, R. T. Inference of blood glucose concentrations from subcutaneous glucose concentrations: applications to glucose biosensors. *Ann. Biomed. Eng.* **27**, 525–537 (1999).
15. Barman, I., Kong, C. R., Singh, G. P., Dasari, R. R. & Feld, M. S. Accurate spectroscopic calibration for noninvasive glucose monitoring by modeling the physiological glucose dynamics. *Anal. Chem.* **82**, 6104–6114 (2010).
16. Taavitsainen, V. M. & Haario, H. Rapid estimation of chemical kinetics by implicit calibration. *I. J. Chemometrics* **15**, 215–239 (2001).
17. Spegazzini, N., Siesler, H. W. & Ozaki, Y. Sequential identification of model parameters by derivative double two-dimensional correlation spectroscopy and calibration-free approach for chemical reaction systems. *Anal. Chem.* **84**, 8330–8339 (2012).
18. Hansen, P. C. *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion.* (SIAM: Philadelphia, 1998).
19. Thennadil, S. N. *et al.* Comparison of glucose concentration in interstitial fluid, and capillary and venous blood during rapid changes in blood glucose levels. *Diabetes Technol. Ther.* **3**, 357–365 (2001).
20. Tikhonov, A. N. & Arsenin, V. Y. *Solutions of Ill-Posed Problems* (Wiley: New York, 1977).
21. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **2**, 164–168 (1944).
22. Lui, H., Zhao, J., McLean, D. & Zeng, H. Real-time Raman spectroscopy for in vivo skin cancer diagnosis. *Cancer Res.* **72**, 2491–500 (2012).
23. Brereton, R. G. *Applied Chemometrics for Scientists,* 145–220 (Wiley: Chichester West Sussex, 2007).
24. Cengiz, E. & Tamborlane, W. V. A tale of two compartments: interstitial versus blood glucose monitoring. *Diabetes Technol. Ther.* **11**, S11–S16 (2009).
25. Clarke, W. L., Cox, D., Gonder-Frederick, L. A., Carter, W. & Pohl, S. L. Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose. *Diabetes Care* **10**, 622–628 (1987).
26. Sternberg, F. *et al.* Does fall in tissue glucose precede fall in blood glucose? *Diabetologia* **39**, 609–612 (1996).
27. Caplin, N. J., O'Leary, P., Bulsara, M., Davis, E. A. & Jones, T. W. Subcutaneous glucose sensor values closely parallel blood glucose during insulin-induced hypoglycaemia. *Diabet. Med.* **20**, 238–241 (2003).
28. Aussedat, B. *et al.* A user-friendly method for calibrating a subcutaneous glucose sensor-based hypoglycaemic alarm. *Biosens. Bioelectron.* **12**, 1061–1071 (1997).
29. Arnold, M. A., Burmeister, J. J. & Small, G. W. Phantom Glucose Calibration Models from Simulated Noninvasive Human Near-Infrared Spectra. *Anal. Chem.* **70**, 1773–1781 (1998).
30. Sämann, A. *et al.* Non-invasive blood glucose monitoring by means of near infrared spectroscopy: investigation of long-term accuracy and stability. *Exp. Clin. Endocrinol. Diabetes* **108**, 406–413 (2000).
31. Olesberg, J. T., Liu, L., Zee, V. V. & Arnold, M. A. In Vivo Near-Infrared Spectroscopy of Rat Skin Tissue with Varying Blood Glucose Levels. *Anal. Chem.* **78**, 215–223 (2006).
32. Malchoff, C. D., Shoukri, K., Landau, J. I. & Buchert, J. M. A novel non-invasive blood glucose monitor. *Diabetes Care* **25**, 2268–2275 (2002).
33. Enejder, A. M. K. *et al.* Raman spectroscopy for noninvasive glucose measurements. *J. Biomed. Opt.* **10**, 031114 (2005).

Acknowledgments

This research was supported by Grant No. 126004 from Kwansei Gakuin University, National Institute of Biomedical Imaging and Bioengineering (9P41EB015871-27), JHU

Whiting School of Engineering and ME Department startup fund (I.B.) and CNPq (J.S.S.). We acknowledge Ms. I. Kayama's contributions for the schematic illustration (Fig. 1).

Author contributions

N.S., I.B., N.C.D., Y.O. and R.R.D. designed research; N.S., I.B., N.C.D. and J.S. performed research; N.S., I.B., N.C.D. and R.P. analyzed data; and all authors co-wrote the paper.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Spegazzini, N. *et al.* Spectroscopic approach for dynamic bioanalyte tracking with minimal concentration information. *Sci. Rep.* **4**, 7013; DOI:10.1038/srep07013 (2014).