

RESEARCH

Open Access



Glycated lysine-141 in haptoglobin improves the diagnostic accuracy for type 2 diabetes mellitus in combination with glycated hemoglobin HbA_{1c} and fasting plasma glucose

Sandro Spiller^{1,2}, Yichao Li³, Matthias Blüher⁴, Lonnie Welch³ and Ralf Hoffmann^{1,2*}

Abstract

Background: Recent epidemiological studies indicate that only 30–50% of undiagnosed type 2 diabetes mellitus (T2DM) patients are identified using glycated hemoglobin (HbA_{1c}) and elevated fasting plasma glucose (FPG) levels. Thus, novel biomarkers for early diagnosis and prognosis are urgently needed for providing early and personalized treatment.

Methods: Here, we studied the glycation degrees of 27 glycation sites representing nine plasma proteins in 48 newly diagnosed male T2DM patients and 48 non-diabetic men matched for age (range 35–65 years). Samples were digested with trypsin and enriched for glycated peptides using boronic acid affinity chromatography. Quantification relied on mass spectrometry (multiple reaction monitoring) using isotope-labelled peptides as internal standard.

Results: The combination of glycated lysine-141 of haptoglobin (HP K141) and HbA_{1c} provided a sensitivity of 94%, a specificity of 98%, and an accuracy of 96% to identify T2DM. A set of 15 features considering three glycation sites in human serum albumin, HP K141, and 11 routine laboratory measures of T2DM, metabolic syndrome, obesity, inflammation, and insulin resistance provided a sensitivity of 98%, a specificity of 100%, and an accuracy of 99% for newly diagnosed T2DM patients.

Conclusions: Our studies demonstrated the great potential of glycation sites in plasma proteins providing an additional diagnostic tool for T2DM and elucidating that the combination of these sites with HbA_{1c} and FPG could improve the diagnosis of T2DM.

Keywords: Biomarker, Fasting plasma glucose (FPG), Glycation sites, HbA_{1c}, Multiple reaction monitoring (MRM), Plasma proteins, Protein glycation, Type 2 diabetes mellitus

Background

Diabetes mellitus (DM) refers to a group of metabolic disorders characterized by elevated blood glucose concentrations. The majority (90–95%) of DM patients have type 2 DM (T2DM), which is characterized by peripheral insulin resistance and an inability of pancreatic beta cells to compensate for that by increasing insulin secretion [1].

Current diagnostic criteria of the World Health Organization (WHO) for the diagnosis of T2DM include glycated hemoglobin (HbA_{1c}) levels ($\geq 6.5\%$ (48 mmol/mol)), elevated fasting plasma glucose (FPG) concentrations (FPG ≥ 7.0 mmol/L) and/or plasma glucose concentrations 2 h after a 75 g oral glucose load (≥ 11.1 mmol/L) in the context of a standardized oral glucose tolerance test (OGTT) [2, 3]. The American Diabetes Association (ADA) defines a HbA_{1c} level of $\geq 6.5\%$ or a fasting plasma glucose (FPG) level of ≥ 126 mg/dL (≥ 7 mmol/L) or 2-h plasma glucose level of ≥ 200 mg/dL (≥ 11.1 mmol/L) during a 75-g oral glucose tolerance test (OGTT), or a

*Correspondence: bioanaly@rz.uni-leipzig.de

² Center for Biotechnology and Biomedicine (BBZ), Universität Leipzig, Deutscher Platz 5, 04103 Leipzig, Germany

Full list of author information is available at the end of the article

random plasma glucose of ≥ 200 mg/dL (≥ 11.1 mmol/L) as equally valid diagnostic criteria for type 2 diabetes [1]. However, epidemiological studies revealed that compared to OGTT glucose values the use of either HbA_{1c} or FPG only identifies ~30–50% of previously undiagnosed T2DM patients [4–8]. Moreover, categories of increased risk for diabetes (prediabetes) include a HbA_{1c} level of 5.7–6.4% or a FPG level between 100 and 125 mg/dL (5.6–6.9 mmol/L) or 2-h plasma glucose level between 140 and 199 mg/dL (7.8–11.0 mmol/L) in the OGTT [1]. For the distinction of patients with prediabetes, HbA_{1c} values have also been shown to be less sensitive compared to FPG and 2 h OGTT glucose [7]. HbA_{1c} also plays a critical role in the management of diabetic patients, as it reflects the average plasma glucose level for the preceding 3-month period and correlates well with micro- and macrovascular complications [4–6]. FPG is the preferred diagnostic parameter, because it is simple, inexpensive, relatively risk-free, and correlates with diabetic complications like retinopathy [7]. However, even using longitudinally measured HbA_{1c} levels, the prediction or retrospective attribution of individual T2DM associated complications or mortality are not possible with this parameter of chronic hyperglycemia [8–12]. Therefore, there is an unmet need to better predict the individual chronic hyperglycemia-related diabetes outcomes.

Glycated proteins, such as hemoglobin in erythrocytes and serum albumin as major plasma protein, are recognized as markers of hyperglycemia due to the high sensitivity towards even slightly elevated plasma glucose levels [13, 14]. However, only methods determining global protein glycation degrees have been established, whereas recent evidence suggests specific glycation sites in plasma proteins as potential biomarkers for early diagnoses of DM [14–17]. In contrast to the long half-life time of hemoglobin in erythrocytes, plasma proteins vary in half-life times from hours to several weeks, which might allow selecting a small set of protein glycation sites resembling more closely short to medium term fluctuations of plasma glucose levels.

Here we quantified 27 glycation sites in nine plasma proteins after tryptic digestion by mass spectrometry and evaluated their diagnostic value alone and in combination with current WHO criteria of HbA_{1c} and FPG levels. Selection of glycation sites relied on a list of 52 candidates we previously identified and studied [17], by judging each for the biomarker potential and ease of peptide synthesis. A single-blind study using plasma samples from 48 newly diagnosed T2DM patients and 48 non-diabetic individuals clearly indicated different glycation degrees characteristic of each glycation site with glycated Lys141 of haptoglobin providing the best sensitivities (~94 and ~78%) and specificity

(~98%) when combined with HbA_{1c} and FPG levels, respectively.

Methods

Blood samples

Blood samples were obtained from 48 newly diagnosed male T2DM patients and 48 non-diabetic men matched for age (range 35–65 years) in the context of a study of parameters of insulin resistance (Additional file 1: Table S1). Anthropometric and laboratory chemistry parameters were measured or calculated as previously described [18, 19]. The study was approved by the Ethics Committee of Universität Leipzig (approval no: 159-12-21052012), and performed in accordance to the declaration of Helsinki. All subjects gave written informed consent before taking part in this study. In our study, we defined individuals with T2DM according to the criteria of the American Diabetes Association [1]:

- A HbA_{1c} level of 6.5% or higher, or
- A fasting plasma glucose (FPG) level of 126 mg/dL (7 mmol/L) or higher (fasting was defined as no caloric intake for at least 8 h), or
- A 2-h plasma glucose level of 200 mg/dL (11.1 mmol/L) or higher during a 75-g oral glucose tolerance test (OGTT), or
- A random plasma glucose of 200 mg/dL (11.1 mmol/L) or higher in a patient with classic symptoms of hyperglycemia (i.e., polyuria, polydipsia, polyphagia, weight loss) or hyperglycemic crisis

The group of T2DM patients has been further categorized into those with HbA_{1c} levels patients below ($n = 23$) and equal or higher than 6.5% (48 mmol/mol) ($n = 25$). For those patients with HbA_{1c} < 6.5% (48 mmol/mol), the diagnosis of T2DM has been established on the basis of repeated measurements of fasting plasma glucose (>7.0 mmol/L) or 2 h oral glucose tolerance test glucose concentrations >11.1 mmol/L [1]. Since patients were newly diagnosed, none of the individuals in the T2DM receive anti-hyperglycemic medication. All men had a body mass index (BMI) larger than 25.0 kg/m² and were grouped in overweight (BMI 25–30 kg/m²; 15 without and 10 with T2DM) and obese individuals (BMI ≥ 30 kg/m²). EDTA blood samples were collected between 8 a.m. and 9 a.m. after a 12-h fasting period, centrifuged (500×g, 5 min) and analyzed within 1 h after blood drawing for routine laboratory parameters or stored at –80 °C after the cell debris was removed by filtration (Rotilabo® syringe filter) for the analyses of the glycation sites. Plasma insulin and proinsulin were measured with an enzyme immunometric assay for the IMMULITE® automated analyzer (Diagnostic Products Corporation, Los

Angeles, CA, USA). Serum high-sensitive CRP (C-reactive protein) was measured by immunonephelometry (Dade-Behring, Milan, Italy). HbA_{1c}, plasma glucose, serum total- high-density lipoprotein (HDL)-, low-density lipoprotein (LDL)-cholesterol, triglycerides, and free fatty acids were measured as previously described [18].

Peptide quantification

Glycation sites previously identified in plasma samples of patients with diabetes [17] were quantified by electrospray ionization mass spectrometry (ESI-MS) on a QTRAP 4000 (AB Sciex, Darmstadt, Germany) coupled on-line to reversed-phase high-performance liquid chromatography (RP-HPLC) using timed multiple reaction monitoring (MRM) (Table 1 and Additional file 1: Table S2). Briefly, small molecules and peptides were removed

from plasma by ultrafiltration (5 kDa cut-off). The concentrated sample was digested with trypsin (37 °C, 18 h, 5% w/w), spiked with a concentration-balanced mixture of ¹³C,¹⁵N-labelled glycation peptides as internal standard, enriched for glycation peptides by boronic acid affinity chromatography (BAC), and desalted by solid phase extraction (SPE) using optimized protocols [15, 20–22]. The internal standard was added after tryptic digestion to provide the highest accuracy by compensating variations, such as nonspecific degradation, during sample preparation [23], especially as trypsin could cleave some standard peptides at Lys/Arg-Pro-motifs or glycation residues.

Peptides were loaded on a C₁₈-column (Advance-Bio Peptide Mapping column, pore size 120 Å, length 150, 2.1 mm internal diameter, 2.7 μm particles, Agilent Technologies, Böblingen, Germany) coupled on-line to

Table 1 Glycated peptides of different plasma proteins quantified in tryptic digests of type 2 diabetes and control plasma samples

#	Sequence	Protein symbol (Accession number, glycation site)	t _R (min)	Q1 m/z (±0.2)	Q3 m/z (±0.2)
1	TCVADESAENCD K SLHTLFGDK	HSA (P02768; K64)	13.8	887.1	869.1
2	SLHTLFGD K LCTVATLR	HSA (P02768, K73)	17.9	698.7	680.7
3	ETYGEMADCCA K QEPER	HSA (P02768, K93)	8.8	746.0	136.1
4	ETYGEMADCCA K QEPER	HSA (P02768, K93)	7.9	751.3	136.1
5	AAFTECCQAAD K AACLLPK	HSA (P02768, K174)	15.9	763.0	120.0
6	AACLLP K LDELREDEGK	HSA (P02768, K181)	16.2	664.0	646.0
7	AEFAEVS K LVTDLTK	HSA (P02768, K233)	18.6	907.0	880.0
8	ADLA K YICENQDSISSK	HSA (P02768, K262)	13.6	1052.5	1025.5
9	TYETTLE K CCAAADPHECYAK	HSA (P02768, K359)	9.9	670.8	237.1
10	VFDEF K PLVEEPQNLIK	HSA (P02768, K378)	18.6	1104.1	1077.1
11	K VPQVSTPTLVEVSR	HSA (P02768, K414)	14.9	601.3	900.5
12	K QTALVELVK	HSA (P02768, K525)	13.1	645.9	603.9
13	EQL K AVMDDFAAFVEK	HSA (P02768, K545)	18.7	668.3	120.1
14	K LVAASQAALGL	HSA (P02768, K574)	16.3	652.4	501.8
15	VQW K VDNALQSGNSQESVTEQDSK	IGKC (P01834, K41)	12.7	947.1	941.1
16	DSTYLSSTLTL S KADYEK	IGKC (P01834, K75)	16.4	757.7	751.7
17	VYACEVTHQGLSSPV T KSFNR	IGKC (P01834, K99)	12.6	848.1	830.1
18	QV K DNENVVNEYSSELEK	FGB (P02675, K163)	12.8	762.7	72.1
19	S KAIGYLNTGYQR	A2M (P01023, K1003)	10.5	816.9	1255.6
20	ALLAYAFALAGNQD K R	A2M (P01023, K1162)	18.5	628.7	86.1
21	K CSTSSLLEACTFR	TF (P02787, K683)	13.8	911.4	884.4
22	ADSSPV K AGVETTTTPSK	IGLC (P01842, K50)	8.7	613.0	70.1
23	A KVQPYLDDFQK	APOA1 (P02647, K120)	12.5	807.4	765.4
24	K WQEE M ELYR	APOA1 (P02647, K131)	10.1	530.6	524.6
25	AVGD K LPECEAVCGKPK	HP (P00738, K141)	9.4	674.0	656.0
26	M KGLIDEVNQDFTNR	FGA (P02671, K71)	14.9	653.3	894.4
27	SSSYS K QFTSSTSYNR	FGA (P02671, K581)	8.7	664.6	658.6

The timed multiple reaction monitoring relied on the retention time of each peptide in RP-HPLC and a specific precursor/fragment ion pair (Q1/Q3 mass range)

t_R retention time, C, M, and K carbamidomethylated cysteine, methionine sulfoxide, and fructosamine lysine, HSA human serum albumin, IGKC Ig kappa chain c region, FGB fibrinogen beta chain, A2M alpha-2-macroglobulin, TF serotransferrin, IGLC Ig lambda chain C region, APOA1 apolipoprotein A-I precursor, HP haptoglobin, FGA fibrinogen alpha chain

ESI–MS. Eluents A and B were water and acetonitrile, respectively, containing both formic acid (0.1%, v/v). Elution was achieved by a linear gradient starting 3 min after sample injection from 3 to 10% eluent B within 1 min, to 20% eluent B within 10 min and to 95% eluent B in 7 min. The flow rate was 0.3 mL/min and the column temperature was set to 60 °C. Quantification relied on timed MRM using specific transitions of each targeted peptide and isotope-labelled internal peptide standards synthesized on solid phase in-house. Quantification was performed by integrating individual peaks in extracted ion chromatograms (XICs) using Analyst 1.6 software (AB Sciex) relative to the coeluting isotope-labelled peptides.

Statistics and bioinformatics

The samples from individuals with or without T2DM were evaluated by different statistical tests (Kolmogorow-Smirnow, Mann–Whitney, and *t* test) and calculation of Spearman rank correlation coefficients using Prism 6 (GraphPad software; La Jolla, USA). Receiver operating characteristic (ROC) analysis and screening for variable combination relied on the Excel-add-in Multi-base 2015 (Numerical Dynamics) and Prism 6 software, respectively.

The 48 diabetic and 48 samples were classified by a decision tree algorithm using HbA_{1c} in combination with each glycosylated peptide (Additional file 1: Table S6). The same technique was also applied to FPG (Additional file 1: Table S7). The decision tree algorithm was implemented using Scikit-Learn [24]. Accuracies were evaluated using nested tenfold cross validation [25]. To find the best feature set for classification, a support vector machine-recursive feature elimination (SVM-RFE) method [26] was applied on all glycosylated peptides and clinical parameters, including HbA_{1c}, FPG, BMI, etc. Feature normalization and missing value imputation were performed using WEKA toolkit [27]. The support vector machine was implemented using Scikit-Learn [24]. Accuracies and area under the curve (AUC) values were evaluated using nested tenfold cross validation [25]. Hierarchical clustering was performed on 48 diabetic samples using the “hclust” function in R software version 3.2.1 [28]. To find the subclasses in diabetic samples, the expectation–maximization algorithm in Scikit-Learn [24] was applied. The clustering stability score [29] and elbow criterion [30] were used to find the optimal number of subclasses.

Results

The timed MRM method optimized for quantification of the targeted 27 glycosylated peptides (Table 1) obtained by tryptic digestion from plasma provided limits of

detection (LODs) and quantification (LOQs) in the low to high nanomolar range (Additional file 1: Table S3). Intra- and interday precisions showed typically coefficients of variation (CVs) below 20% (Additional file 1: Table S3). The quantities of all 27 glycosylated peptides normalized to the total protein content of each plasma sample were significantly higher in T2DM than in the control group ($P < 0.05$, Additional file 1: Figure S1). However, the normalized quantities of all glycosylated peptides showed a notable overlap. Interestingly, diabetic groups subdivided by an HbA_{1c} threshold of 6.5% (48 mmol/mol) showed very similar average glycation degrees for all peptides, indicating that the glycation degree of hemoglobin in erythrocytes showed only a weak correlation with the glycation levels of serum proteins providing a rationale that they may have different diagnostic and prognostic values.

Spearman's correlation coefficients (r_s) calculated for each glycosylated peptide amount and the established diagnostic parameters indicated moderate correlations between a few glycation sites and BMI ($-0.54 < r_s < -0.38$, $P < 0.001$), body fat ($-0.49 < r_s < -0.36$, $P < 0.01$), and C-peptide levels ($-0.45 < r_s < -0.37$, $P < 0.001$) (Additional file 1: Table S4). Correlations between peptide glycation levels and HbA_{1c} were weak ($0.11 < r_s < 0.34$, $0.001 < P < 0.3$) to moderate ($0.37 < r_s < 0.46$, $P < 0.001$) and also weak ($-0.08 < r_s < 0.30$, $0.001 < P < 0.65$) for FPG (Additional file 1: Table S4). Besides FPG ($r_s = 0.40$, $P < 0.001$) and free fatty acids (FFA, ($r_s = 0.57$, $P < 0.001$)) HbA_{1c} showed for further diagnostic parameters only weak correlations ($-0.35 < r_s < 0.35$, $0.001 < P < 0.95$). However, FPG showed moderate correlations with fasting plasma insulin (FPI, $r_s = 0.40$, $P < 0.001$), proinsulin ($r_s = 0.38$, $P < 0.01$), homeostasis model assessment as an index of insulin resistance (HOMA-IR, $r_s = 0.56$, $P < 0.001$), and fasting plasma free fatty acids (FFA) ($r_s = 0.56$, $P < 0.001$).

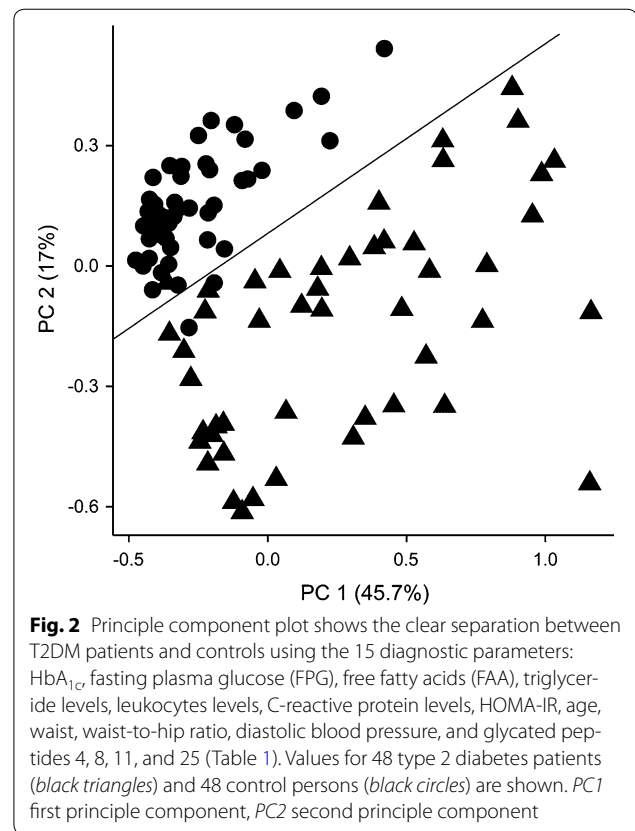
A ROC curve analysis for samples from T2DM patients relative to HbA_{1c} (cut-off 6.5% (48 mmol/mol)) and FPG (cut-off 7.0 mmol/L) provided for all glycosylated peptides maximal sensitivities and specificities of 79 and 88%, respectively, and AUCs larger than 79% for certain cut-off concentrations (Additional file 1: Table S5). In comparison, sensitivities, specificities, and AUCs of HbA_{1c} were 52, 100, and 86%, respectively, as well as 40, 100, and 84% for FPG using the aforementioned cut-offs. A ROC curve analysis determined the best cut-off values maximizing sensitivity and selectivity as 6.0% (42 mmol/mol) for HbA_{1c} (77 and 94%) and 5.69 mmol/L for FPG (75 and 81%).

As the diagnostic accuracies of HbA_{1c} (76% for cut-off 6.5% (48 mmol/mol), 86% for cut-off 6.0%), FPG (70% for cut-off 7.00 mmol/L, 78% for cut-off 5.69 mmol/L), and the evaluated glycosylated peptides (60 to 73%) were

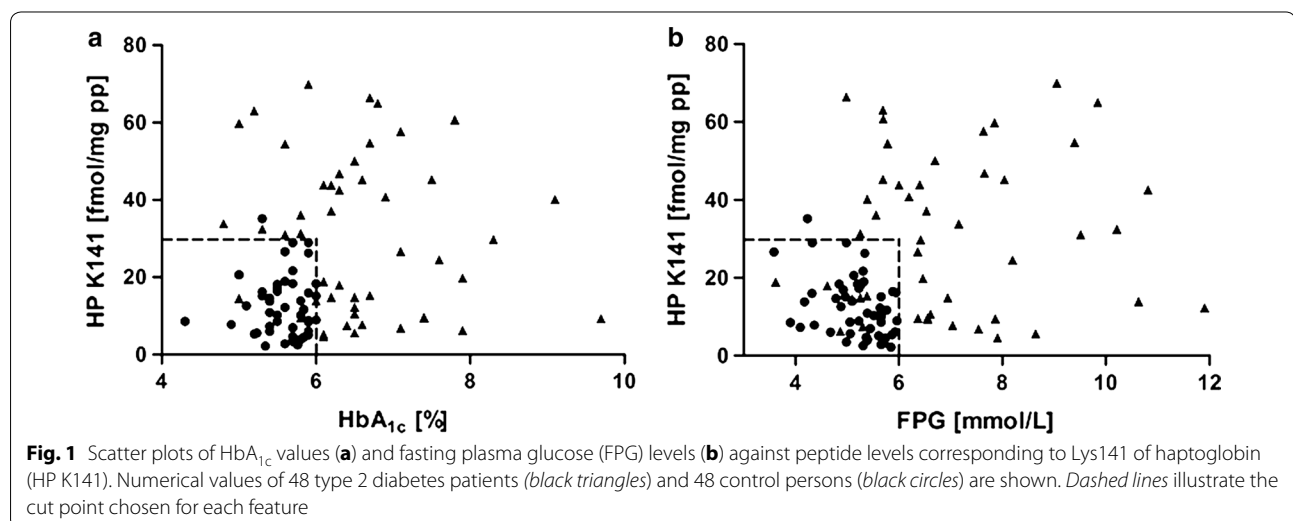
insufficient, the data sets were screened by decision tree algorithm for variable combinations of HbA_{1c} or FPG with any glycosylated peptide for optimal sensitivity, specificity, and accuracy using different cut-off points. HbA_{1c} in combination with glycosylated peptides yielded mostly sensitivities from 75 to 81%, specificities from 96 to 100%, and accuracies from 87 to 90% (Additional file 1: Table S6). Most interestingly, analyses revealed a haptoglobin peptide glycosylated at Lys141 (HP K141; sequence No. 25) which provided in combination with HbA_{1c} a sensitivity of 94%, a specificity of 98%, and an accuracy of 96% for cut-off points 6.0% (42 mmol/mol, HbA_{1c}) and 30 fmol/mg (HP K141) (Fig. 1a). When HP K141 was combined with FPG (6.0 mmol/L), sensitivity to detect T2DM was 78%, specificity 98%, and accuracy 88% (Fig. 1b), whereas the other glycosylated peptides typically provided slightly lower accuracies (Additional file 1: Table S7). Since the decision tree algorithm did not deliver all cut-off values for all combinations, manual confirmation was carried out. Combining manually HbA_{1c} or FPG with glycosylated peptides provided slightly enhanced diagnostic accuracies for all combinations and could confirm all cut points provided by the algorithm with only small variations (Additional file 1: Tables S8 and S9). Noteworthy, the combination of HP K141 and FPG (6.0 mmol/L) was improved to a sensitivity of 83%, a specificity of 98%, and an accuracy of 91%.

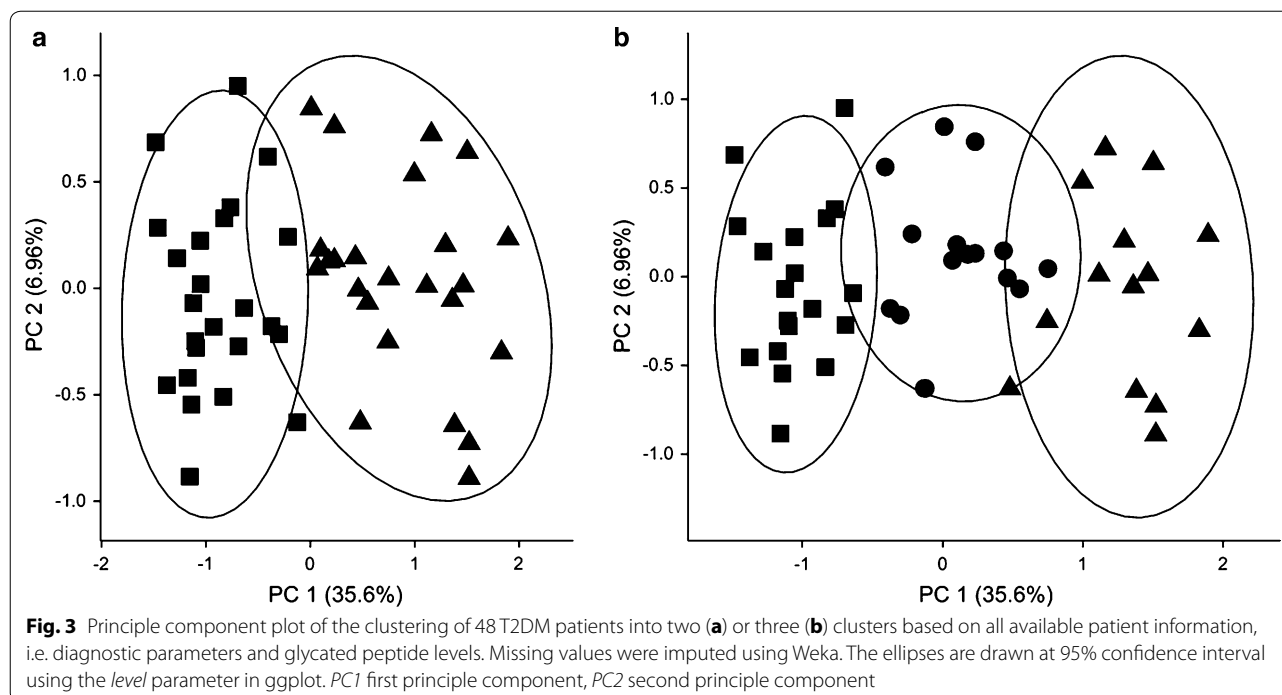
The SVM-RFE method was applied to find a set of diagnostic parameters and peptide amounts for maximizing the classification of T2DM patients and controls. It revealed a set of 15 features providing a sensitivity of 98%, a specificity of 100%, and an accuracy of 99% (Fig. 2).

The cluster analysis of the 48 T2DM plasma samples was applied to an expectation-maximization (EM) algorithm considering all glycosylated peptides and clinical



parameters available (27 glycosylated peptides + 38 clinical parameters). The result showed two or maximal three clusters (Fig. 3). Glycosylated peptide levels were distinct among different groups, however, clinical parameters, including HbA_{1c} and FPG, were similar. This suggests that glycosylated peptides can be more useful to detect subtle differences among diabetic patients. A principle





component plot is used to visualize the distribution of clusters (Fig. 3). A cluster stability test considering the elbow criterion identified three clusters as optimal number. This was further confirmed by hierarchical clustering that also suggested three groups of T2DM patients (Additional file 1: Figure S2).

Discussion

Protein glycation as diagnostic criterion has been addressed in many studies for decades [13–17]. Surprisingly, only glycation of intracellular hemoglobin is currently considered as biomarker to monitor patients assuming that HbA_{1c} levels closely reflect average glucose blood levels over the previous 3 months. Plasma proteins have been mostly analyzed for their global glycation degrees [especially human serum albumin (HSA)], which can be attributed to the lack of specific immunoassays, i.e. antibodies recognizing individual glycation sites. As an example, measurement of fructosamine determines the fraction of total serum proteins (mainly serum albumin) that have undergone glycation. Because albumin has a half-life of approximately 3 weeks, the plasma fructosamine concentration reflects relatively recent changes in blood glucose and is therefore not commonly used to monitor diabetes treatment [31].

Additionally, mass spectrometry has been typically applied for mapping glycation sites in serum proteins, whereas quantitative studies on distinct glycation sites in larger patient cohorts are missing. Besides the analytical

challenges, this is probably attributed to the general assumption that glycation as a non-enzymatic reaction depends only on the glucose concentration and thus the glycation levels of individual sites will change at similar degrees and thus can be judged from global glycation degrees. Here, we could show that the glycation levels of 27 sites in nine plasma proteins provide significantly different sensitivities, specificities and accuracies for classifying T2DM patients and controls. Most importantly, these glycation degrees correlated only slightly to HbA_{1c} and other clinical parameters like FPG. Additionally, sensitivities of the newly identified glycation sites were mostly better than for HbA_{1c}, while specificities were lower demonstrating that T2DM diagnosis might benefit from combining currently applied clinical parameters with plasma protein glycation sites analyzed in the context of this study. Despite studying 27 glycation sites, it was our aim to identify the diagnostically most relevant modification to keep the number of biomarkers and thus the costs of the envisaged diagnostic tools low. Different statistical analyses identified glycation of Lys141 in haptoglobin as the best parameter in combination with HbA_{1c} and FPG. The obtained sensitivity (94%), specificity (98%), and accuracy (96%) of HP K141 in combination with HbA_{1c} exceeded the corresponding values of HbA_{1c} significantly. This suggests that in addition to HbA_{1c} measurements, glycation of Lys141 in haptoglobin could be used as a diagnostic tool in patients with T2DM and those with a high risk to develop T2DM (Fig. 1a). Further

studies are necessary to test whether combination of these two parameters of chronic hyperglycemia may better predict individual risk for the development of diabetes complications. However, an earlier diagnosis of T2DM may be important for individual patients' outcomes, as it has been demonstrated that a "bad glycemetic memory" may contribute to a higher risk of diabetes complications even after periods of well-controlled hyperglycemia [32]. The negative correlation between HP K141 and both BMI and C-peptide appears interesting, as it might indicate higher glycation levels in T2DM patients are caused by both diabetes-specific, but also obesity-associated metabolic alterations, which subsequently may cause chronic hyperglycemia. Negative correlations between glycation levels and BMI suggest that adipose tissue in patients with obesity may have a higher capacity to take up excess glucose before it contributes to glycation of circulating proteins. However, this hypothesis has to be formally proven by glucose distribution studies in lean versus obese animal models under hyperglycemic conditions.

Although our initial intention was to identify a single biomarker, statistical evaluation of the data revealed a promising features selection matrix using three glycation sites in HSA (K93, K262, and K414) besides HP K141 representing short-to-medium term glucose fluctuations in combination with twelve routine parameters typically used to characterize T2DM (FPG, HbA_{1c}, fasting insulin), metabolic syndrome (triglycerides and blood pressure), obesity (BMI, waist circumference, waist-to-hip ratio), inflammation (leukocytes, C-reactive protein), and insulin resistance (HOMA-IR), and age [33]. This feature set provided an extremely high accuracy of 99% for the sample cohort providing the best diagnostic value that has been reported to the best of our knowledge. For example, a feature set reported in 2009 relied on 160 individuals (Danish Inter99 prospective study), who progressed from initially non-diabetic to diabetes during the following five years. The predictive values of 58 biomarkers (selected from presumed diabetes-associated pathways) together with six routine clinical parameters were evaluated by statistical learning methods [34]. The best features, i.e. six biomarkers (adiponectin, C-reactive protein, ferritin, interleukin-2 receptor A, glucose, and insulin), were used in a PreDx diabetes risk score (DRS) model providing an AUC of 0.76 that increased to 0.78 when family history, age, BMI, and waist circumference were added. A later report indicated an AUC of 0.838 [35]. The PreDx DRS model outperformed single variables like FPG (~0.73) and HbA_{1c} (~0.66) significantly. In this respect the 15 feature set selected here considering only four new biomarkers besides routine laboratory measures, provides a better diagnostic accuracy with on the AUC of 0.99375 (accuracy 99%).

Another intriguing result of statistical evaluation and hierarchical clustering was the separation of diabetes patients in three subgroups based on all patient information including all 27 glycation sites (Fig. 3; Additional file 1: Figure S2). At this stage, the clinical relevance remains open, but it is compelling to speculate that the subgroups might represent distinct subgroups of T2DM. This hypothesis is supported by a recent report that three T2DM subgroups have been dissected through topological analysis of patient similarity [36]. In this analysis, one T2DM subtype was characterized by diabetic nephropathy and retinopathy, another showed enriched risk for malignancies and cardiovascular diseases, whereas the third subtype correlated most strongly with cardiovascular diseases, neurological diseases, allergies, and specific infections. Based on this data, future longitudinal studies, which assess the risk of these disease entities in addition to the glycation sites described in our study should be performed to test the hypothesis that these three subtypes could be distinguished by protein glycation markers.

These subgroups may also respond differently to multifactorial T2DM treatment strategies or may predict the progression of the disease itself and the individual risk for developing diabetes complications. Noteworthy, our study has some limitations with regard to the small size of our discovery cohort and the limited transferability into clinical practice. Although we only included newly diagnosed T2DM patients at the earliest possible time point and without antidiabetic treatment, we cannot relate the time point of diagnosis to a defined metabolic state. In addition, we could not exclude individual differences in the duration and severity of the prediabetic phase including chronic effects of intermittent glucose and lipid toxicity. Moreover, the suggestion that glycated Lys141 in haptoglobin may improve diagnostic accuracy for T2DM needs to be tested in prospective studies and cohorts representing the common population.

In conclusion, the 27 glycation sites quantified here provided sensitivities up to 79% and specificities of up to 88% to distinguish T2DM samples relative to age- and BMI-matched control samples using specific cut-offs for each glycation site. The cut-off values of HbA_{1c} (6.5% (48 mmol/mol)) and FPG (7.0 mmol/L) recommended by the WHO showed better specificities of 100%, but lower sensitivities of only 52 and 40%, respectively. Interestingly, lowering the cut points to 6.0% (42 mmol/mol) and 5.69 mmol/L improved the sensitivity significantly to 77 and 75%, respectively, while the selectivity was reduced to 94 and 81%, respectively. This is in good agreement with recent reports suggesting that lower HbA_{1c} and FPG cut points will dramatically increase diagnostic sensitivity [37–41]. Remarkably, plasma proteins appeared to

follow other glycation kinetics than hemoglobin, which might be related to their different environment (i.e. blood versus erythrocytes). Thus, glycation sites in plasma proteins may provide an additional diagnostic tool, as confirmed here by combining the glycation levels HbA_{1c} and HP K141 providing high sensitivity (94%) and specificity (98%). The advantage of combining these two markers are the different half-life times of the corresponding proteins, i.e. 3–4 months for intracellular hemoglobin and two to four days for haptoglobin ($t_{1/2} = 2-4$) [42] being sensitive to long- and short-term fluctuations of blood glucose concentrations. Furthermore, the combination of 15 features consisting of established clinical parameters and several glycosylated plasma proteins allowed dividing the newly diagnosed patients with T2DM into three different subgroups supporting the hypothesis that heterogeneity of T2DM phenotypes maybe due differentially affected pathways including significant differences in protein glycation.

Additional file

Additional file 1. Table S1. Characterization of type 2 diabetes patients and matched non-diabetic persons enrolled in this study. **Table S2.** Parameters and settings used for the RP-HPLC-ESI-QqLIT-MS operating in scheduled multiple reaction monitoring (MRM) mode. **Table S3.** Precision, sensitivity and linearity parameters for glycosylated peptides. **Table S4.** Spearman rank correlation coefficients (rS) and corresponding P values (P) of the statistical relation between glycosylated peptides and several diagnostic parameters. **Table S5.** Receiver operating characteristic (ROC) parameters for peptide levels of all 27 glycosylated peptides quantified in tryptic digests of plasma samples obtained from 48 type 2 diabetes patients and 48 controls. For comparison, ROC parameters of HbA_{1c} and fasting plasma glucose (FPG) are listed. **Table S6.** Evaluation metrics for classification of type 2 diabetes patients and controls by combining the levels of 27 glycosylated peptides in tryptic plasma digests and corresponding HbA_{1c} levels calculated by Decision Tree classifier from Scikit-learn package. **Table S7.** Evaluation metrics for classification of type 2 diabetes patients and controls by combining the levels of 27 glycosylated peptides in tryptic plasma digests and corresponding FPG levels. **Table S8.** Evaluation metrics for classification of type 2 diabetes patients and controls by combining the levels of 27 glycosylated peptides in tryptic plasma digests and corresponding HbA_{1c} levels. Variable cut points were optimized manually for best classification. **Table S9.** Evaluation metrics for classification of type 2 diabetes patients and controls by combining the levels of 27 glycosylated peptides in tryptic plasma digests and corresponding fasting plasma glucose (FPG) levels. Variable cut points were optimized manually for best classification. **Figure S1.** Quantification of glycosylated peptides in tryptic plasma digests obtained from type 2 diabetes patients and non-diabetic controls using internal calibration. **Figure S2.** Dendrogram resulting from a hierarchical clustering of 48 type 2 diabetes patient samples considering all available patient information, i.e. diagnostic parameters and glycosylated peptide levels.

Abbreviations

ADA: American Diabetes Association; AUC: area under the curve; BAC: boronic acid affinity chromatography; BMI: body mass index; CRP: C-reactive protein; CV: coefficient of variation; DM: diabetes mellitus; EM: expectation-maximization; ESI-MS: electrospray ionization mass spectrometry; FFA: free fatty acids; FPG: fasting plasma glucose; HbA_{1c}: glycosylated hemoglobin; HDL: high-density lipoprotein; HOMA-IR: homeostasis model assessment as an index of insulin resistance; HP K141: glycosylated lysine-141 of haptoglobin; HSA: human serum

albumin; LDL: low-density lipoprotein; LOD: limit of detection; LOQ: limit of quantification; MRM: multiple reaction monitoring; OGTT: oral glucose tolerance test; ROC: receiver operating characteristic; RP-HPLC: reversed-phase high-performance liquid chromatography; SPE: solid phase extraction; SVM-RFE: support vector machine-recursive feature elimination; T2DM: type 2 diabetes mellitus; WHO: World Health Organization; XIC: extracted ion chromatogram.

Authors' contributions

SS researched data, contributed to discussion, wrote the manuscript, and reviewed/edited the manuscript. YL and LW researched data, contributed to discussion and reviewed/edited the manuscript. MB and RH contributed to discussion, wrote the manuscript, and reviewed/edited the manuscript. All authors read and approved the final manuscript.

Author details

¹ Institute of Bioanalytical Chemistry, Faculty of Chemistry and Mineralogy, Universität Leipzig, Leipzig, Germany. ² Center for Biotechnology and Biomedicine (BBZ), Universität Leipzig, Deutscher Platz 5, 04103 Leipzig, Germany. ³ School of Electrical Engineering and Computer Science, Ohio University, Athens, OH, USA. ⁴ Department for Internal Medicine, Clinic for Endocrinology and Nephrology, University Hospital Leipzig, Universität Leipzig, Leipzig, Germany.

Acknowledgements

We thank Uta Greifenhagen, Dr. Andrej Frolov, and Dr. Ravi Chand Bollineni for the help with sample preparation and helpful discussions and Dr. Daniel Knappe, Dr. David Singer, and Tina Goldbach for their support in synthesizing the peptides.

Competing interests

The authors declare that they have no competing interests.

Availability of supporting data

The datasets during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Universität Leipzig (approval no: 159-12-21052012), and performed in accordance to the declaration of Helsinki. Written informed consent was obtained from all participants in this study.

Funding

Financial support from Deutsche Forschungsgemeinschaft (HO-2222/7-1 to RH and SFB 1052: Obesity Mechanisms, B01 to MB) is gratefully acknowledged.

Received: 5 November 2016 Accepted: 18 March 2017

Published online: 28 March 2017

References

1. American Diabetes Association (ADA). Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2013;36:67–74.
2. World Health Organization (WHO) and International Diabetes Federation (IDF). Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. Geneva: World Health Organization; 2006.
3. World Health Organization (WHO). Abbreviated report of a WHO consultation. Use of glycosylated hemoglobin (HbA_{1c}) in the diagnosis of diabetes mellitus. Geneva: World Health Organization; 2011.
4. Little RR, Rohlfing CL, Sacks DB. Status of hemoglobin A1c measurement and goals for improvement: from chaos to order for improving diabetes care. *Clin Chem*. 2011;57:205–14.
5. Tanaka S, Tanaka S, Iimuro S, Yamashita H, Katayama S, Akanuma Y, Yamada N, Araki A, Ito H, Sone H, Ohashi Y. Predicting macro- and microvascular complications in type 2 diabetes: the Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine. *Diabetes Care*. 2013;36:1193–9.

6. Leal J, Hayes AJ, Gray AM, Holman RR, Clarke PM. Temporal validation of the UKPDS outcomes model using 10-year posttrial monitoring data. *Diabetes Care*. 2013;36:1541–6.
7. Cox ME, Edelman D. Tests for screening and diagnosis of type 2 diabetes. *Clin Diabetes*. 2009;27:132–8.
8. Qiao Q, Dekker JM, de Vegt F, Nijpels G, Nissinen A, Stehouwer CD, Bouter LM, Heine RJ, Tuomilehto J. Two prospective studies found that elevated 2-hr glucose predicted male mortality independent of fasting glucose and HbA1c. *J Clin Epidemiol*. 2004;57:590–6.
9. Ray KK, Seshasai SR, Wijesuriya S, Sivakumaran R, Netherrcott S, Preiss D, Erqou S, Sattar N. Effect of intensive control of glucose on cardiovascular outcomes and death in patients with diabetes mellitus: a meta-analysis of randomised controlled trials. *Lancet*. 2009;373:1765–72.
10. Currie CJ, Peters JR, Tynan A, Evans M, Heine RJ, Bracco OL, Zagar T, Poole CD. Survival as a function of HbA(1c) in people with type 2 diabetes: a retrospective cohort study. *Lancet*. 2010;375:481–9.
11. McEwen LN, Karter AJ, Waitzfelder BE, Crosson JC, Marrero DG, Mangione CM, Herman WH. Predictors of mortality over 8 years in type 2 diabetic patients: translating research into action for diabetes (TRIAD). *Diabetes Care*. 2012;35:1301–9.
12. Faerch K, Vistisen D, Johansen NB, Jorgensen ME. Cardiovascular risk stratification and management in pre-diabetes. *Curr Diabetes Rep*. 2014;14:493.
13. Lyons TJ, Basu A. Biomarkers in diabetes: hemoglobin A1c, vascular and tissue markers. *Transl Res*. 2012;159:303–12.
14. Rondeau P, Bourdon E. The glycation of albumin: structural and functional impacts. *Biochimie*. 2011;93:645–58.
15. Frolov A, Hoffmann R. Identification and relative quantification of specific glycation sites in human serum albumin. *Anal Bioanal Chem*. 2010;397:2349–56.
16. Zhang Q, Monroe ME, Schepmoes AA, Clauss TRW, Gritsenko MA, Meng D, Petyuk VA, Smith RD, Metz TO. Comprehensive identification of glycosylated peptides and their glycation motifs in plasma and erythrocytes of control and diabetic subjects. *J Proteome Res*. 2011;10:3076–88.
17. Frolov A, Blüher M, Hoffmann R. Glycation sites of human plasma proteins are affected to different extents by hyperglycemic conditions in type 2 diabetes mellitus. *Anal Bioanal Chem*. 2014;406:5755–63.
18. Kloting N, Fasshauer M, Dietrich A, Kovacs P, Schon MR, Kern M, Stumvoll M, Bluher M. Insulin-sensitive obesity. *Am J Physiol Endocrinol Metab*. 2010;299:E506–15.
19. Kannt A, Pfenninger A, Teichert L, Tonjes A, Dietrich A, Schon MR, Kloting N, Bluher M. Association of nicotinamide-N-methyltransferase mRNA expression in human adipose tissue and the plasma concentration of its product, 1-methylnicotinamide, with insulin resistance. *Diabetologia*. 2015;58:799–808.
20. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics*. 2006;5:573–88.
21. Frolov A, Hoffmann R. Analysis of amadori peptides enriched by boronic acid affinity chromatography. *Ann N Y Acad Sci*. 2008;1126:253–6.
22. Spiller S, Frolov A, Hoffmann R. Quantification of specific glycation sites in human serum albumin as prospective type 2 diabetes mellitus biomarkers. *Protein Pept Lett*. 2017;24. doi:10.2174/0929866524666170202124120.
23. Kuzyk MA, Smith D, Yang J, Cross TJ, Jackson AM, Hardie DB, Anderson NL, Borchers CH. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol Cell Proteomics*. 2009;8:1860–77.
24. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
25. Chen L, Xuan J, Wang C, Shih Ie M, Wang Y, Zhang Z, Hoffman E, Clarke R. Knowledge-guided multi-scale independent component analysis for biomarker identification. *BMC Bioinform*. 2008;9:416.
26. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn*. 2002;46:389–422.
27. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *ACM SIGKDD Explor NewsL*. 2009;11:10–8.
28. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna: The R Foundation for Statistical Computing; 2011. ISBN: 3-900051-07-0. <http://www.R-project.org/>.
29. von Luxburg U. Clustering stability: an overview. *Found Trends Mach Learn*. 2009;2:235–74.
30. Ketchen DJ, Shook CL. The application of cluster analysis in strategic management research: an analysis and critique. *Strateg Manag J*. 1996;17:441–58.
31. Danese E, Montagnana M, Nouvenne A, Lippi G. Advantages and pitfalls of fructosamine and glycated albumin in the diagnosis and treatment of diabetes. *J Diabetes Sci Technol*. 2015;9:169–76.
32. Bianchi C, Del Prato S. Metabolic memory and individual treatment aims in type 2 diabetes—outcome-lessons learned from large clinical trials. *Rev Diabetes Stud*. 2011;8:432–40.
33. Meigs JB. Multiple biomarker prediction of type 2 diabetes. *Diabetes Care*. 2009;32:1346–8.
34. Kolberg JA, Jørgensen T, Gerwien RW, Hamren S, McKenna MP, Moler E, Rowe MW, Urdea MS, Xu XM, Hansen T, et al. Development of a type 2 diabetes risk model from a panel of serum biomarkers from the Inter99 cohort. *Diabetes Care*. 2009;32:1207–12.
35. Urdea M, Kolberg J, Wilber J, Gerwien R, Moler E, Rowe M, Jorgensen P, Hansen T, Pedersen O, Jorgensen T, Borch-Johnsen K. Validation of a multimarker model for assessing risk of type 2 diabetes from a five-year prospective study of 6784 Danish people (Inter99). *J Diabetes Sci Technol*. 2009;3:748–55.
36. Li L, Cheng WY, Glicksberg BS, Gottesman O, Tamler R, Chen R, Bottinger EP, Dudley JT. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med*. 2015;7:311–25.
37. Buell C, Kermah D, Davidson MB. Utility of A1C for diabetes screening in the 1999–2004 NHANES population. *Diabetes Care*. 2007;30:2233–5.
38. Kramer CK, Araneta MRG, Barrett-Connor E. A1C and diabetes diagnosis: the Rancho Bernardo Study. *Diabetes Care*. 2010;33:101–3.
39. Carson AP, Reynolds K, Fonseca VA, Muntner P. Comparison of A1C and fasting glucose criteria to diagnose diabetes among U.S. adults. *Diabetes Care*. 2010;33:95–7.
40. Lorenzo C, Wagenknecht LE, Hanley AJ, Rewers MJ, Karter AJ, Haffner SM. A1C between 5.7 and 6.4% as a marker for identifying pre-diabetes, insulin sensitivity and secretion, and cardiovascular risk factors: the Insulin Resistance Atherosclerosis Study (IRAS). *Diabetes Care*. 2010;33:2104–9.
41. Zhou X, Pang Z, Gao W, Wang S, Zhang L, Ning F, Qiao Q. Performance of an A1C and fasting capillary blood glucose test for screening newly diagnosed diabetes and pre-diabetes defined by an oral glucose tolerance test in Qingdao, China. *Diabetes Care*. 2010;33:545–50.
42. Carter K, Worwood M. Haptoglobin: a review of the major allele frequencies worldwide and their association with diseases. *Int J Lab Hematol*. 2007;29:92–110.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

