

RESEARCH

Open Access



Aptamer-based search for correlates of plasma and serum water T₂: implications for early metabolic dysregulation and metabolic syndrome

Vipulkumar Patel^{1,2}, Alok K. Dwivedi³, Sneha Deodhar¹, Ina Mishra^{1,2} and David P. Cistola^{1,2*} 

Abstract

Background: Metabolic syndrome is a cluster of abnormalities that increases the risk for type 2 diabetes and atherosclerosis. Plasma and serum water T₂ from benchtop nuclear magnetic resonance relaxometry are early, global and practical biomarkers for metabolic syndrome and its underlying abnormalities. In a prior study, water T₂ was analyzed against ~ 130 strategically selected proteins and metabolites to identify associations with insulin resistance, inflammation and dyslipidemia. In the current study, the analysis was broadened ten-fold using a modified aptamer (SOMAmer) library, enabling an unbiased search for new proteins correlated with water T₂ and thus, metabolic health.

Methods: Water T₂ measurements were recorded using fasting plasma and serum from non-diabetic human subjects. In parallel, plasma samples were analyzed using a SOMAscan assay that employed modified DNA aptamers to determine the relative concentrations of 1310 proteins. A multi-step statistical analysis was performed to identify the biomarkers most predictive of water T₂. The steps included Spearman rank correlation, followed by principal components analysis with variable clustering, random forests for biomarker selection, and regression trees for biomarker ranking.

Results: The multi-step analysis unveiled five new proteins most predictive of water T₂: hepatocyte growth factor, receptor tyrosine kinase FLT3, bone sialoprotein 2, glucokinase regulatory protein and endothelial cell-specific molecule 1. Three of the five strongest predictors of water T₂ have been previously implicated in cardiometabolic diseases. Hepatocyte growth factor has been associated with incident type 2 diabetes, and endothelial cell specific molecule 1, with atherosclerosis in subjects with diabetes. Glucokinase regulatory protein plays a critical role in hepatic glucose uptake and metabolism and is a drug target for type 2 diabetes. By contrast, receptor tyrosine kinase FLT3 and bone sialoprotein 2 have not been previously associated with metabolic conditions. In addition to the five most predictive biomarkers, the analysis unveiled other strong correlates of water T₂ that would not have been identified in a hypothesis-driven biomarker search.

Conclusions: The identification of new proteins associated with water T₂ demonstrates the value of this approach to biomarker discovery. It provides new insights into the metabolic significance of water T₂ and the pathophysiology of metabolic syndrome.

(Continued on next page)

* Correspondence: david.cistola@ttuhsc.edu

¹Nanoparticle Diagnostics Laboratory, Institute for Cardiovascular & Metabolic Diseases, University of North Texas Health Science Center, Fort Worth, TX 76107, USA

²Center of Emphasis in Diabetes & Metabolism, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, El Paso, TX 79905, USA

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



(Continued from previous page)

Keywords: Metabolic syndrome, Insulin resistance, Nuclear magnetic resonance relaxometry, Transverse relaxation time constant, Hepatocyte growth factor, Glucokinase regulatory protein, Receptor tyrosine kinase FLT3, SOMAscan aptamer assay, Random forests, Classification and regression tree analysis

Background

Metabolic syndrome (MetS) is a cluster of clinical findings that includes increased waist circumference, high blood pressure, high blood glucose, high triglycerides and/or low HDL-cholesterol [1, 2]. The criteria for MetS differ depending on the guideline used, but a widely used consensus requires that at least three of the five criteria are met [3]. Metabolic syndrome is associated with a two-fold increased risk for atherosclerosis, a five-fold increased risk for type 2 diabetes [2] and an increased risk for some forms of cancer [4]. The prevalence of MetS is high among the U.S. population: one third of adults and half of those ≥ 60 years of age [5, 6]. Previously called insulin resistance syndrome [7], the pathophysiological factors that drive MetS include insulin resistance, inflammation and ectopic lipid deposition [1, 2].

In an observational cross-sectional study of 72 non-diabetic human subjects, we discovered that plasma and serum water T_2 detect MetS-associated abnormalities with high sensitivity and specificity [8]. Measured using benchtop nuclear magnetic resonance relaxometry [9], T_2 refers to the time constant for the decay or “relaxation” of the transverse component of the NMR signal. Water T_2 is sensitive to the rotational diffusion of protein-bound and unbound water molecules and serves as a surveillance system for shifts in blood proteins and lipoproteins. One example is the shifts that occur with an acute phase response, which increase the levels of some globulins, while decreasing albumin. As globulins are higher molecular weight than albumin, the net effect is to slow the rotational mobility of bound water and decrease water T_2 [8, 9].

Fasting hyperinsulinemia (insulin resistance), dyslipidemia and inflammation each have independent and additive contributions to the lowering of water T_2 [8]. Hence, water T_2 captures a global view of an individual's metabolic health status with just one measurement. It shows promise as a screening test for the early detection of poor metabolic health to prevent diabetes and cardiovascular disease [8]. However, the role of water T_2 in probing metabolic health and elucidating the pathophysiology of MetS has not been fully explored.

The initial search for metabolic correlates of water T_2 was conducted using 130 strategically-selected blood biomarkers that measure different aspects of metabolic health status [8]. Biomarker selection was based on investigator-driven hypotheses and priorities. While the prior search yielded a wealth of information, it could

have been limited by selection bias. Therefore, the search for new correlates of water T_2 was broadened 10-fold to probe a random library of 1310 plasma proteins using a DNA-based modified aptamer assay developed by SomaLogic, Inc. In this manuscript, we report the results of the SOMAscan analysis of plasma samples from non-diabetic subjects who participated in Phase 2 of the prior study.

Target-specific single-stranded DNA aptamers can be generated in a relatively short time and with substantially less cost than antibodies. Therefore, this technology is gaining recognition as a tool for biomarker discovery [10–18]. A major advantage is that aptamer-based assays are highly multiplexed and can measure hundreds-to-thousands of proteins from biofluids without the need for isolation or pre-treatment [10].

While a broad evaluation of biomarkers is important, it creates challenges related to high-dimension data analysis on a comparatively small number of subjects. Given the large number of proteins measured, the use of statistical correlation alone to identify associations with water T_2 would increase the probability of false positives. To circumvent this problem, we applied a systematic multi-step method for dimension reduction starting with bivariate correlations, followed by principal components analysis with variable clustering, random forest variable selection, and classification and regression tree analysis or CART. The results identified new predictors of plasma and serum water T_2 and provided new insights into biomarkers for metabolic health.

Methods

Human subject recruitment, blood collection and processing

Human subject research was performed under a protocol approved by the Institutional Review Board of the University of North Texas Health Sciences Center, Fort Worth. A screening interview was completed by each subject prior to obtaining informed consent, and a full medical history was obtained after enrollment. The inclusion criteria were adults ages 18 and up, weighing at least 110 pounds. The exclusion criteria were active acute or chronic illness (history/diagnosis or CRP ≥ 10), diabetes (history/diagnosis or fasting glucose ≥ 125 mg/dl or HbA1c $\geq 6.5\%$), confirmed or suspected pregnancy, history of bleeding disorders or difficulty giving blood, or not fasting for at least 12 h.

The fasting blood draw was scheduled for 7:00 AM. During the visit, the nurse-phlebotomist recorded routine physical measurements such as height, weight, abdominal waist circumference and blood pressure. In addition, a urine sample was analyzed for microalbuminuria using Chemstrip Micral (Roche Diagnostics, Inc.). The blood samples were centrifuged right after venipuncture using a two-step procedure [8]. For NMR analysis, the freshly drawn and centrifuged samples were analyzed immediately. For SOMAscan assays, the plasma obtained from Phase 2 subjects was biobanked at -80°C for several months prior to analysis.

Benchtop NMR relaxometry measurements

The ^1H NMR data for plasma and serum samples were recorded using a Bruker mq20 Minispec benchtop relaxometer operating at 0.47 T, corresponding to 20 MHz for ^1H . Samples were pipetted into a 3 mm coaxial insert inside of a 10 mm NMR tube (Norell NI10CCI-B, Norell, Inc., Morganton, North Carolina, USA). The sample height was 1 cm, corresponding to a total volume of $\sim 50\ \mu\text{L}$. A modified Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence was used for T_2 measurement, as detailed elsewhere [8, 9]. The recycle delay was set to $5 \times T_1$ to achieve essentially complete spin relaxation prior to the next round of the pulse sequence. Sixteen scans were signal averaged in each experiment, for a data collection time of 3 min. The data were collected in triplicate. To extract and resolve T_2 values, the raw CPMG decay curves were analyzed using a discrete inverse Laplace transform algorithm as implemented in XpFit [9, 19]. The number of exponential terms was fixed to three for all samples. Water T_2 was the dominant term, accounting for $> 90\%$ of the total CPMG signal intensity [9].

SOMAscan proteomics assay

Frozen biobanked plasma samples were shipped overnight on dry ice to SomaLogic, Inc. (Boulder, Colorado, USA) for SOMAscan analysis. The relative concentrations of 1310 plasma proteins were quantified using a proprietary SOMAscan proteomics assay [10]. This assay is based on the selective binding of single-stranded nucleic acid aptamers called SOMAmers (Slow Off-rate Modified Aptamers) to target proteins. The SOMAmer library for target selection was developed using the SELEX method [20, 21].

Statistical analysis strategy

The search for SOMAscan-detected proteins most predictive of water T_2 was carried out in four steps: (1) screening the variables using bivariate correlations between protein biomarkers and plasma or serum water T_2 , (2) grouping the correlated proteins into statistically-related clusters and identifying the most representative

variable in each cluster, (3) selecting the most predictive variables using an iterative multi-variable random forests analysis, and (4) defining the interactions of the most predictive biomarkers and their final associations with plasma or serum water T_2 levels. The general scheme is illustrated in Fig. 1, and each of the steps is explained further below.

Correlation and variable cluster analysis

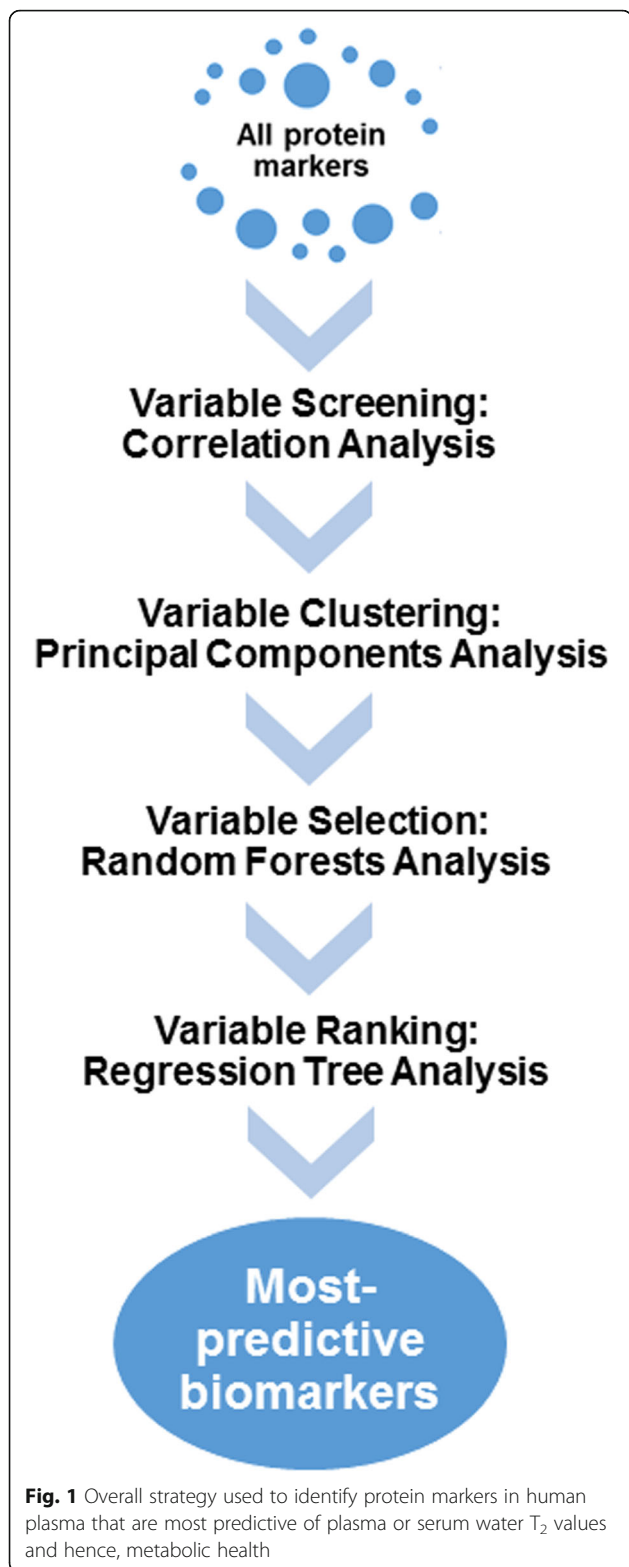
First, the 1310 SOMAscan-derived biomarkers were analyzed using the Shapiro-Wilk normality test in R 3.1.4 statistical software [22]. Based on this analysis, only 408 ($\sim 31\%$) of the variables followed a normal distribution. Therefore, the correlations of plasma or serum T_2 with SOMAscan-derived biomarkers were screened using non-parametric Spearman rank correlation coefficients (ρ values). The screening criterion was $|\rho| \geq 0.3$. In the pre-selection of variables associated with water T_2 , we focused on the effect size (correlation coefficient), not statistical significance (p -value), in order to limit false negatives.

To reduce the dimensionality of the search at the screening stage, we applied principal components analysis with variable clustering as implemented in JMP Pro 12.1.0 (SAS, Inc., Cary, North Carolina, USA). The algorithm identified variable clusters, as well as the most representative variable in each cluster [23]. Variable clustering is not to be confused with conventional cluster analysis, which identifies clustering across subjects, as opposed to clustering across measured variables. It has advantages over factor analysis for dimension reduction and has been recently used in clinical research [24]. In addition, variable clustering reduces the difficulty in interpreting the output of conventional principal component analysis [23]. For each cluster, the variable corresponding to the largest squared correlation with its cluster component was identified as the most representative variable and used for the next step of statistical analysis.

Random forests and CART analysis

The most representative variables from all clusters were used as independent variables, and water T_2 as a dependent variable, to construct two random forest models: one for plasma water T_2 and another for serum water T_2 . Random forests, developed by Leo Breiman and colleagues, is a powerful non-parametric machine learning algorithm to make predictions from the data [25]. In addition, it can be used as a tool to select variables, in this case SOMAscan-derived protein markers, based on their importance in predicting plasma or serum water T_2 . This analysis was performed using the package randomForest in R 3.1.4 [22, 26].

The randomForest algorithm generated regression trees based on a statistical resampling or bootstrap method. It started with a randomly-selected subset of the original data, i.e., a learning set containing approximately one third



of the protein variables. Each learning set was used to create a regression tree, where the first branch contained the protein variable that showed the maximum difference in T_2 between the two branches, with an approximately

equal number of subjects in each branch. Similarly, additional branches were created until each variable in the learning set was incorporated into the tree. Through bootstrapping, a total of 1000 regression trees were created from 1000 randomly-selected learning sets. This method ensured the stability of the results by repeating the association analysis a large number of times. For each subject, the T_2 value predicted from all 1000 trees was averaged and compared to the experimentally determined T_2 for that subject. Finally, the mean squared error was calculated by comparing the predicted vs. observed T_2 values across all subjects.

To select the most predictive variables, the random forests analysis was repeated after removing all trees containing a given variable. Then the remaining trees were used to predict the T_2 value for a given subject, and the mean squared error was calculated to quantify predicted vs. observed T_2 across all subjects. This process was performed recursively by leaving out trees containing one variable at a time and calculating a new mean squared error. The *percent change* in mean squared error before and after leaving out each variable was computed, and the variables were ranked by the percent change. By convention, protein variables with $\geq 5\%$ change in mean squared error after being removed from the random forests model were selected as the top predictors of water T_2 . Note that the use of the 5% threshold was somewhat arbitrary, and proteins falling just below this threshold also are predictive of water T_2 .

Using the most predictive variables, two final regression trees were constructed using classification and regression tree analysis or CART: one for plasma and one for serum water T_2 . The CART analysis explores the possible interactions across all the selected variables by determining the most appropriate binary classification of each variable. The regression trees were constructed by identifying variables that maximized the T_2 difference while keeping the number of subjects in each branch approximately equal. The branching was stopped when the number of subjects in each branch was $< 25\%$ of the total number of subjects in the study.

Multiple regression analysis

As a cross check on the most predictive variables identified by random forest, the variables were used to generate multiple linear regression models, with plasma or serum water T_2 as the outcome variable. The models were constructed using the stepwise tools in JMP Pro v14.0, and acceptable models met the following criteria [8]: (1) all predictor variables were statistically significant at $\alpha = 0.05$, (2) the models were not overfit, as assessed by k-fold cross validation, and (3) the adjusted R^2 was maximized.

Results

Characteristics of the human subject cohort

The study population consisted of asymptomatic individuals without active acute or chronic disease (Table 1). There were approximately equal numbers of males and females. The mean values for clinical lab tests fell within their reference ranges, although some individuals had values outside the normal range. By American Diabetes Association criteria, 15 of the 41 Phase 2 subjects had prediabetes based on HbA_{1c} and/or fasting glucose levels; none had overt diabetes. Using the harmonized criteria [3], 9 of 41 subjects met the definition of MetS. By water T₂ criteria, 19 of the 41 subjects had hyperinsulinemia/insulin resistance using the cut points established by Robinson et al. [8]. Five of the 19 (26%) had compensatory hyperinsulinemia (early metabolic dysregulation) and did not meet the criteria for either prediabetes or MetS.

Bivariate correlations and variable clustering analysis

Figure 2 provides a schematic overview of the results from each stage of statistical analysis for plasma (left side) and serum water T₂ (right side). The correlation analysis revealed 311 and 269 protein markers for plasma and serum T₂, respectively, using a Spearman ρ absolute-value threshold of 0.3. The full lists of 311 and 269 protein markers with correlation coefficients are provided in Additional file 1: Tables S1 and S2, respectively.

The correlated variables were further subjected to dimension reduction using variable clustering. The clustering algorithm revealed 55 and 47 clusters for plasma and serum water T₂, respectively. Additional file 1: Tables S3 and S4 list all of the clusters, as defined by their most representative variables, for plasma and serum T₂--correlated biomarkers, respectively.

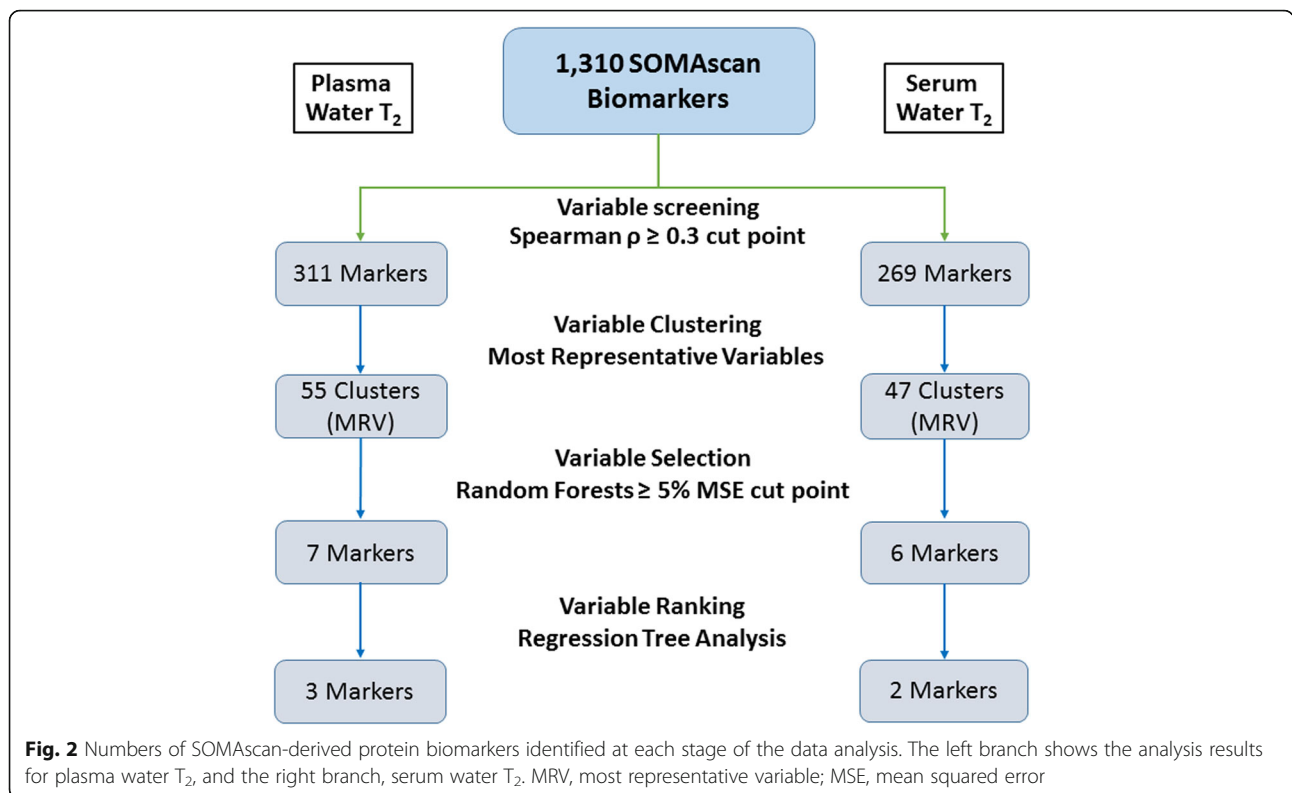
Table 1 Characteristics of the human study population ($n = 41$)

Parameter	Mean \pm S.D.	Range	Reference Values ^a
Age	36.5 \pm 12.5	23–61	n/a
Gender	n/a	19 female, 22 male	n/a
Body-Mass Index (kg/m ²)	26.5 \pm 5.5	19.1–45.1	< 25 normal weight, 25–30 overweight, > 30 obese
Plasma T ₂ (ms)	771.5 \pm 58.3	631–887	\geq 745.0 ^b
Serum T ₂ (ms)	817.4 \pm 5.2	706–908	\geq 811.8 ^b
Glucose (mg/dL)	90.6 \pm 7.7	71–109	< 100 non-diabetic 100–125 (pre-diabetic)
HbA _{1c} (%)	5.5 \pm 0.3	4.7–6.1	< 5.7 (non-diabetic) 5.7–6.4 (pre-diabetic)
Insulin C-peptide (ng/mL)	1.8 \pm 0.8	0.7–5.1	0.8–3.9 (> 2.85, IR ^c)
Insulin (μ U/mL)	8.7 \pm 6.6	2.2–40.1	2.0–19.6 (> 12.2, IR ^c)
Total serum protein (g/dL)	7.2 \pm 0.4	6.3–8.0	6.1–8.1
Serum albumin (g/dL)	4.5 \pm 0.3	3.6–5.1	3.6–5.1
Serum globulins (g/dL)	2.7 \pm 0.3	1.9–3.3	1.9–3.7
Triglycerides (mg/dL)	123 \pm 63.1	50–321	< 150
Total cholesterol (mg/dL)	185.0 \pm 45.0	97–291	< 200
HDL-C (mg/dL)	51.7 \pm 12.7	31–78	\geq 40 (male); \geq 50 (female)
LDL-C (mg/dL)	110.1 \pm 35.9	50–191	< 130
WBC count ($\times 10^3/\mu$ L)	6.6 \pm 1.7	3.9–11.2	3.8–10.8
Neutrophil count ($\times 10^3/\mu$ L)	3.6 \pm 1.2	1.8–7.2	1.5–7.8
hs-CRP (mg/L)	2.6 \pm 2.7	0.1–9.6	< 3.0 (low/average CV risk) 3.0–10.0 (high CV risk) > 10.0 (infection/illness)
Sodium (mmoles/L)	138.2 \pm 2.6	131–143	135–146
Potassium (mmoles/L)	4.1 \pm 0.3	3.7–4.8	3.5–5.3
Total CO ₂ , (mmoles/L)	24.2 \pm 2.3	18–28	19–30

^aReference values from Quest Diagnostics and Atherotech, except where noted

^bCutoff for normoglycemic population established in previous study [8]

^cInsulin cutoff from McAuley et al. [67]; insulin C-peptide cutoff established by linear regression with insulin



Random forests and CART analysis

The most representative variable from each cluster was selected for random forests analysis. This analysis yielded 7 proteins most predictive for plasma water T₂ (Table 2) and 6 for serum water T₂ (Table 3). Each protein displayed a percent increase in mean squared error $\geq 5\%$ after trees containing this protein were removed from the random forests model. As shown in Tables 2 and 3, glucokinase regulatory protein and

receptor-type tyrosine protein kinase FLT3 were top predictors of both plasma and serum water T₂.

As revealed by CART analysis, the final regression tree for plasma water T₂ included three biomarkers: hepatocyte growth factor receptor, receptor tyrosine kinase FLT3 (fms-like tyrosine kinase 3), and bone sialoprotein 2 (Fig. 3). The final regression tree for serum water T₂ included two protein markers: endothelial cell-specific molecule 1 and glucokinase regulatory protein (Fig. 4).

Table 2 Most predictive biomarkers and cluster members for plasma water T₂

Protein Name (Uniprot ID) ^a	% Inc. MSE ^a	Cluster Members
Hepatocyte growth factor (P14210)	9.51	R-spondin-2, Galectin-7
Glucokinase regulatory protein (Q14397)	9.44	Low-density lipoprotein receptor-related protein 1 soluble, T-lymphocyte activation antigen CD86,
Receptor-type tyrosine-protein kinase FLT3 (P36888)	8.34	Complement C4b ^b , Discoidin domain-containing receptor 2, Serine/threonine-protein kinase PAK 6, Heterogeneous nuclear ribonucleoprotein A/B
Ephrin-B2 (P52799)	7.52	Leucine-rich repeat transmembrane protein FLRT3, Amphoterin-induced protein 2, Ephrin-A5, NT-3 growth factor receptor, Kallikrein-8, Interleukin-1 receptor type 1, Iduronate 2-sulfatase, CD109 antigen, Cell adhesion molecule 1, SLIT and NTRK-like protein 5, Ephrin type-A receptor 2, Endoglin, Interleukin-22 receptor subunit alpha-2, OX-2 membrane glycoprotein, Semaphorin-6B, Semaphorin-6A, Interferon alpha/beta receptor 1
Bone sialoprotein 2 (P21815)	5.22	Fibrinogen ^b , Alpha-1-antichymotrypsin, Antithrombin-III, Endothelial cell-specific molecule 1, Serotransferrin
Histone-lysine N-methyl-transferase EHMT2 (Q96KQ7)	5.13	Metalloproteinase inhibitor 1, Metalloproteinase inhibitor 2, Delta-like protein 4
Fibroblast growth factor 2 (P09038)	5.09	Fibroblast growth factor 4

^aThe most predictive biomarkers are defined as those with $\geq 5\%$ increase in mean squared error (MSE)

^bIn a prior study, complement C4 (C4c) and fibrinogen were strongly associated with plasma water T₂ [8]

Table 3 Most predictive biomarkers and cluster members for serum water T₂

Protein Name (Uniprot ID) ^a	% Inc. MSE ^a	Cluster Members
Endothelial cell-specific molecule 1 (Q9NQ30)	11.0	Bone sialoprotein 2, Bone morphogenetic protein 10
Glucokinase regulatory protein (Q14397)	8.0	2'-5'-oligoadenylate synthase 1, T-lymphocyte activation antigen CD86
Lactadherin (Q08431)	7.4	Hepatocyte growth factor receptor, Alpha-2-macroglobulin, Adrenomedullin, N terminal pro BNP
Vascular cell adhesion protein 1 (P19320)	6.1	Secreted frizzled-related protein 3, L-selectin
Receptor-type tyrosine-protein kinase FLT3 (P36888)	5.6	Serum amyloid P-component, Discoidin domain-containing receptor 2, Serine/threonine-protein kinase PAK 6, Heterogeneous nuclear ribonucleoprotein A/B
Semaphorin-6A (Q9H2E6)	5.2	Osteopontin, Leucine-rich repeat transmembrane protein FLRT3, Ephrin-B2, Ephrin-A5, NT-3 growth factor receptor, Kallikrein-8, Interleukin-1 receptor type 1, Iduronate 2-sulfatase, CD109 antigen, Cell adhesion molecule 1, Brother of CDO, SLIT and NTRK-like protein 5, Endoglin, Neuropilin-1

^aThe most predictive biomarkers are defined as those with ≥5% increase in mean squared error (MSE)

Multiple regression analysis

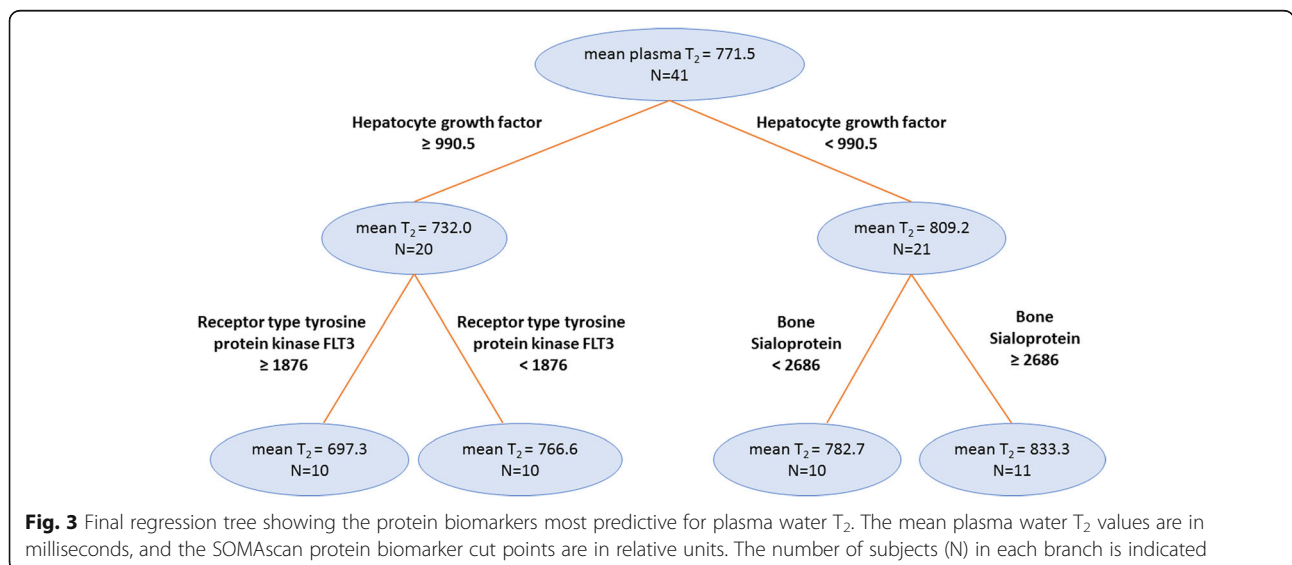
As a validation check for the random forest results, we tested the variables listed in Tables 2 and 3 as predictor variables in multiple linear regression models, with plasma or serum water T₂ as the outcome variable. The best model for plasma water T₂ incorporated hepatic growth factor, receptor-type tyrosine protein kinase FLT3 and bone sialoprotein, yielding an adjusted R² of 0.52. These three predictor variables accounted for over half of the variation in plasma water T₂. For serum water T₂, the best model incorporated endothelial cell specific molecule 1, receptor-type tyrosine protein kinase FLT3 and semaphorin 6A, yielding an adjusted R² of 0.47. Thus, the results from random forests are consistent with those obtained from a different method.

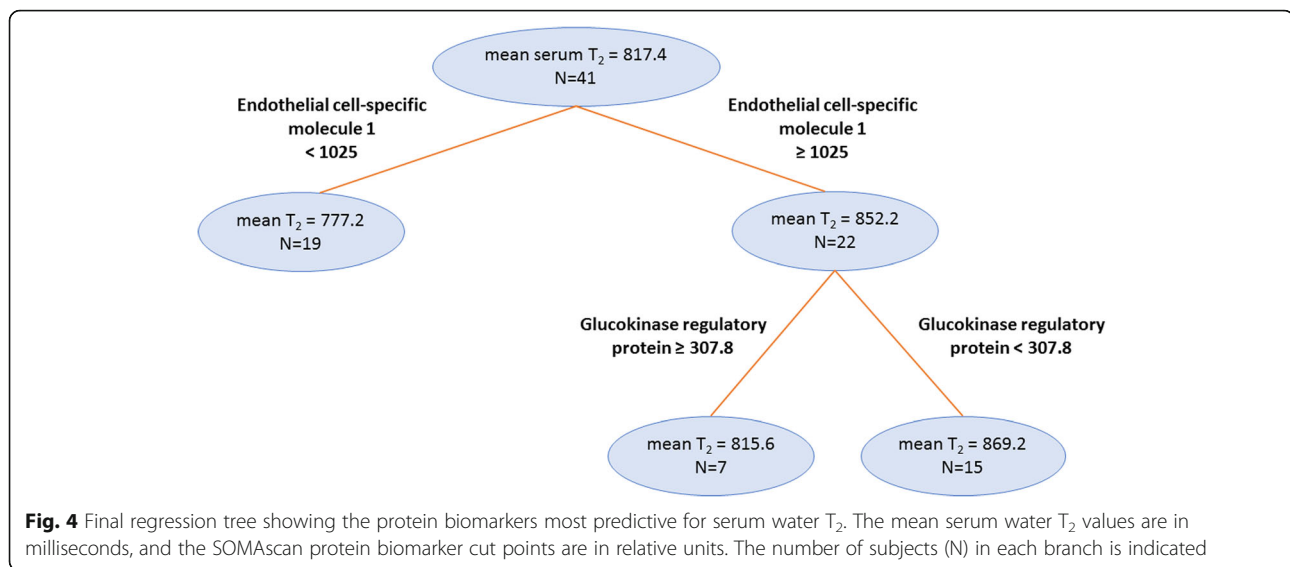
Discussion

For the first time, a highly multiplexed SOMAscan assay was used in an unbiased search for new correlates of

plasma and serum water T₂. Using this discovery strategy, we identified proteins in the SOMAmer library that were most predictive of water T₂ and hence, metabolic health [8]. The dimensionality was reduced using a systematic multi-step procedure that incorporated principal components analysis with variable clustering, random forests, and classification and regression trees. The analysis unveiled five proteins most predictive of plasma and serum water T₂, as well as six other proteins that emerged from the random forests analysis as strong predictors. All are new hits, as none of these proteins were included or considered in the prior hypothesis-driven biomarker search for correlates of water T₂.

Three proteins were most predictive of plasma water T₂: hepatocyte growth factor, receptor tyrosine kinase FLT3 (fms-like tyrosine kinase 3) and bone sialoprotein 2. The latter two proteins have not been previously associated with metabolic conditions or diabetes. However, FLT3 is implicated in inflammation, immunity and





autoimmune diseases and is overexpressed in leukemia [27, 28]. Also known as CD135, FLT3 is involved in development of immune cells in bone marrow and peripheral lymphoid tissue [29, 30]. In particular, FLT3 regulates the growth of hematopoietic stem cells and the development/homeostasis of dendritic cells in lymphoid tissue [29, 30]. Activation of the receptor by mutation leads to proliferation, resistance to apoptosis and prevention of differentiation, leading to myeloid leukemia.

Hepatic growth factor has been implicated in diabetes-related conditions [31–35]. It is elevated in overt type 2 diabetes [35] as well as diabetes-associated coronary artery disease and cerebral infarction [31, 33]. Most relevant to the current study are recent results from the multi-ethnic study of atherosclerosis (MESA), a longitudinal human cohort study. The MESA results revealed that *elevated levels of HGF predict incident type 2 diabetes* [36]. The current observation of a strong inverse association between plasma water T_2 and HGF is consistent with this finding, as low water T_2 detects early metabolic conditions thought to lead to type 2 diabetes, namely insulin resistance, subclinical inflammation and dyslipidemia [8]. In addition, water T_2 is strongly and inversely correlated with complement C3, C4, fibrinogen, and haptoglobin, markers predictive of incident type 2 diabetes [8, 37].

The hepatic growth factor *receptor*, also known as MET, is part of a tyrosine kinase signaling complex that functions in cell growth and survival, angiogenesis and tissue regeneration [38–40]. It is expressed in cells of mesenchymal origin, including epithelial and endothelial cells, neurons, hepatocytes, adipocytes, myocytes and pancreatic cells. The receptor is cell-membrane associated (c-MET), but a soluble ectodomain (s-MET) is shed and circulates in plasma [41]. The receptor is upregulated in cancer, and

both c-MET and s-MET have been investigated as biomarkers of malignancy, metastasis and tumor progression [40, 42–44].

In this study, s-MET (soluble HGF receptor) displayed positive Spearman correlations with plasma and serum water T_2 (+0.45 and +0.44, respectively; $p < 0.01$; Additional file 1: Tables S1 and S2). Those correlations were opposite in sign to those for the receptor ligand HGF. Like HGF, MET was among the variables predictive of water T_2 , but at 4.2%, was just below the 5% mean-squared error threshold employed in Tables 2 and 3. Thus, high HGF and low soluble HGF receptor are associated with low water T_2 and poor metabolic health.

In the pancreas, HGF/MET signaling is necessary for beta-cell regeneration [45]. A pancreas-specific knockout of the MET gene in mice accelerates the onset of diabetes [46]. Also, hepatocyte growth factor signaling is thought to be a mediator of beta cell proliferation in obesity [47]. Moreover, hypoxia-inducible factor (HIF1), which is associated with obesity and sleep apnea, is a transcriptional regulator of MET [48–50]. Thus, the expression of MET appears to be increased under conditions of metabolic dysregulation that place high secretory demand on beta cells, such as obesity, insulin resistance and tissue hypoxia. A decreased ability to upregulate MET under these conditions may hasten the demise of beta cells and accelerate the onset of type 2 diabetes.

The current observation of an association between plasma water T_2 and HGF/MET reinforces the notion that low plasma water T_2 is a biomarker of metabolic dysregulation and poor metabolic health, even in individuals without prediabetes or metabolic syndrome [8]. Given the association of plasma water T_2 with other proteins that predict future type 2 diabetes and atherosclerosis, namely fibrinogen, complement C3 and C4,

haptoglobin, α 1-acid glycoprotein (orosomuroid) and apolipoprotein B, the association of water T_2 with HGF provides further evidence that plasma water T_2 is a biomarker of the metabolic dysregulation that precedes type 2 diabetes and cardiovascular disease [8].

Bone sialoprotein 2, named according to its high sialic acid content, is expressed during the development of bone and cementum [51]. The function of this protein is unknown but believed to serve as a nucleation site for hydroxyapatite crystals [52]. Expression of this protein is regulated by hormones, growth factors and cytokines [53]. As shown in Table 2, fibrinogen is a member of this protein cluster (Table 2) and likely mediates the statistical association between plasma water T_2 and bone sialoprotein 2. Fibrinogen is the fourth most abundant protein in plasma. Changes in its level directly affects plasma water T_2 [8]. Endothelial cell specific molecule 1 is in that cluster as well.

The CART regression tree analysis for *serum* water T_2 yielded two biomarkers: endothelial cell specific molecule 1 and glucokinase regulatory protein (GKRP). Endothelial cell specific molecule 1 (ESM-1 or endocan) is involved in angiogenesis and plays a role in lung-endothelial cell-leukocyte interactions [54, 55]. It has recently been implicated in subclinical atherosclerosis in type II diabetes patients [56]. In addition, ESM-1 is involved or implicated in prostate cancer [57], endothelial injury in respiratory distress syndrome [58], oral cancer [59], erectile dysfunction [60], and pulmonary infection [61]. Note that the ESM-1 cluster for serum water T_2 includes bone sialoprotein 2, but not fibrinogen (Table 3). Serum water T_2 is unaffected by fibrinogen levels, as this protein is absent in serum.

Glucokinase regulatory protein is a well-known inhibitor of glucokinase and a key regulator of liver glucose uptake and metabolism [62–65]. Normally, GKRP is an intracellular protein localized within hepatocytes. As shown here, increased GKRP levels in plasma and serum were associated with a lowering of T_2 values and a worsening of metabolic health, specifically insulin resistance and glucose intolerance. This observation implies that GKRP is leaking from hepatocytes into the circulation, perhaps reflective of early liver damage. None of the subjects in this study have a history of liver disease, but that does not rule out the possibility of subclinical hepatic steatosis or steatohepatitis. This interpretation is supported by the positive correlation between GKRP and ALT observed in these subjects (Spearman $\rho = 0.47$, $p = 0.0024$). Alanine aminotransferase (ALT) is an established marker of liver damage. Plasma and serum T_2 are correlated with both GKRP (this study) and ALT [8].

Study limitations

For two reasons, this study utilized a relatively small number of subjects. First, biobanked samples were available only from Phase 2 of the initial biomarker discovery study

for plasma and serum water T_2 [8]. Second, the SomaScan analysis was expensive, placing practical constraints on its application. At first glance, a small sample size might lead to concerns about statistical power. However, power depends not only on sample size, but also effect size. In this study, the effect size was remarkably large for the most predictive variables identified by random forest. A power calculation for $N = 41$ revealed that a power of 0.8 would be achieved for absolute values of correlation coefficients $>|0.425|$ at $\alpha = 0.05$. By comparison, the Spearman coefficient for hepatocyte growth factor and plasma water T_2 was -0.52 , and the Huber M-value correlation was -0.67 . Likewise, the Spearman and Huber correlation coefficients for endothelial cell specific protein 1 and serum water T_2 were 0.58 and 0.70, respectively. Thus, the current analysis was sufficiently powered because of the large effect sizes, which more than compensated for the relatively small N .

The small sample size placed a practical lower limit on the initial biomarker screening step, possibly generating false negatives by failing to detect some biomarkers that are more weakly, but significantly associated with water T_2 . Therefore, future studies with larger N may unveil additional biomarkers that are less predictive, but still significant contributors to water T_2 . Also, a future study with a different group of subjects will be important for validating the most predictive variables discovered in the current study.

The NMR analysis was performed using freshly-drawn plasma and serum. However, the SOMAscan analysis was performed, by necessity, using one-time frozen-thawed biobanked plasma. Changes in some plasma proteins may have occurred during the freeze-thaw process and could have impacted the analysis. Such changes, if occurred, were likely to be minor, as biobanked plasma and serum are generally stable after one freeze-thaw cycle [66].

Conclusion

The SOMAscan results and multi-stage regression analyses yielded new correlates and predictors of plasma and serum water T_2 that were not previously identified in a hypothesis-driven biomarker search. These new predictors broadened our understanding of the biomarker network and the information content of plasma and serum water T_2 . In addition, the discovery of biomarkers correlated with water T_2 provided new insights into the pathophysiology of metabolic syndrome and the early metabolic dysregulation that precedes type 2 diabetes and cardiovascular disease.

Additional file

Additional file 1: Table S1. Spearman correlation of plasma water T_2 with SOMAscan biomarkers. **Table S2.** Spearman correlation of serum water T_2 with SOMAscan biomarkers. **Table S3.** Clusters of plasma T_2 -correlated SOMAscan biomarkers. **Table S4.** Clusters of serum T_2 -correlated SOMAscan biomarkers. (DOCX 86 kb)

Abbreviations

ADA: American Diabetes Association; AHA: American Heart Association; CART: Classification and regression tree; CPMG: Carr-Purcell-Meiboom-Gill pulse sequence used to measure T_2 ; ESM-1: Endothelial cell-specific molecule 1; FLT3: fms-like tyrosine kinase 3; GKRP: Glucokinase regulatory protein; HGF: Hepatocyte growth factor; MESA: Multi-ethnic Study of Atherosclerosis; MetS: Metabolic syndrome; MSE: Mean squared error; NMR: Nuclear magnetic resonance; SELEX: Systematic evolution of ligands by exponential enrichment; ss-DNA: Single stranded DNA; T_2 : Transverse relaxation time

Acknowledgements

We thank Molly McNamara and colleagues at SomaLogic, Inc. for facilitating the contractual collaboration and performing the SOMAScan assay of plasma samples.

Funding

This work was supported by institutional start-up funds (to D.P.C.) from the University of North Texas Health Science Center, Fort Worth and the Texas Tech University Health Sciences Center El Paso, as well as a grant from the Garvey Texas Foundation.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

VP was a part of the experimental team (also including SD, IM and DPC) that recruited subjects, collected and analyzed the blood samples. He carried out the statistical analyses and drafted the manuscript. The multi-step analysis strategy was devised by AKD, who provided biostatistics expertise throughout the project. IM collected and analyzed the NMR T_2 data. As Principal Investigator, DPC initiated and oversaw all aspects of the project, established the contractual collaboration with SomaLogic, Inc., contributed to the data analysis and edited the manuscript. All authors viewed and edited a penultimate version of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This study was approved and reviewed annually by the Institutional Review Board of the University of North Texas Health Science Center, Fort Worth (Protocol 2013–205). All enrolled subjects gave prior consent to participate by signing an IRB-approved consent form, after being made aware of benefits and risks of participation.

Consent for publication

Not applicable.

Competing interests

The University of North Texas Health Science Center, Fort Worth has pending patent applications for the use of plasma and serum water T_2 for metabolic health screening, with D.P.C. as an inventor.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Nanoparticle Diagnostics Laboratory, Institute for Cardiovascular & Metabolic Diseases, University of North Texas Health Science Center, Fort Worth, TX 76107, USA. ²Center of Emphasis in Diabetes & Metabolism, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, El Paso, TX 79905, USA. ³Division of Biostatistics & Epidemiology, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, El Paso, TX 79905, USA.

Received: 12 July 2018 Accepted: 29 August 2018

Published online: 17 September 2018

References

- Sperling LS, Mechanick JL, Neeland JJ, Herrick CJ, Després J, Ndumele CE, Vijayaraghavan K, Handelsman Y, Puckrein GA, Araneta MRG, Blum KQ, Collins KK, Cook S, Dhurandhar NV, Dixon DL, Egan BM, Ferdinand DP, Herman LM, Hessen SE, Jacobson TA, Pate RR, Ratner RE, Brinton EA, Forker AD, Ritzenthaler LL, Grundy SM. The CardioMetabolic health alliance: working toward a new care model for the metabolic syndrome. *J Am Coll Cardiol*. 2015;66:1050–67.
- Grundy SM. Metabolic syndrome update. *Trends Cardiovasc Med*. 2016;26:364–73.
- Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, Fruchart JC, James WP, Loria CM, Smith SC Jr, International Diabetes Federation Task Force on Epidemiology and Prevention, National Heart, Lung, and Blood Institute, American Heart Association, World Heart Federation, International Atherosclerosis Society, International Association for the Study of Obesity. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation*. 2009;120:1640–5.
- Esposito K, Chiodini P, Colao A, Lenzi A, Giugliano D. Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. *Diabetes Care*. 2012;35:2402–11.
- Aguilar M, Bhuket T, Torres S, Liu B, Wong RJ. Prevalence of the metabolic syndrome in the United States, 2003–2012. *JAMA*. 2015;313:1973–4.
- Moore JX, Chaudhary N, Akinyemiju T. Metabolic syndrome prevalence by race/ethnicity and sex in the United States, National Health and nutrition examination survey, 1988–2012. *Prev Chronic Dis*. 2017;14:E24.
- Reaven GM. The insulin resistance syndrome: definition and dietary approaches to treatment. *Annu Rev Nutr*. 2005;25:391–406.
- Robinson MD, Mishra I, Deodhar S, Patel V, Gordon KV, Vintimilla R, Brown K, Johnson L, O'Bryant S, Cistola DP. Water T_2 as an early, global and practical biomarker for metabolic syndrome: an observational cross-sectional study. *J Transl Med*. 2017;15:258–017. 1359–5.
- Cistola DP, Robinson MD. Compact NMR relaxometry of human blood and blood components. *Trends Anal Chem*. 2016;83:53–64.
- Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, Carter J, Dalby AB, Eaton BE, Fitzwater T, Flather D, Forbes A, Foreman T, Fowler C, Gawande B, Goss M, Gunn M, Gupta S, Halladay D, Heil J, Heilig J, Hicke B, Husar G, Janjic N, Jarvis T, Jennings S, Katilius E, Keeney TR, Kim N, Koch TH, Kraemer S, Kroiss L, Le N, Levine D, Lindsey W, Lollo B, Mayfield W, Mehan M, Mehler R, Nelson SK, Nelson M, Nieuwlandt D, Nikrad M, Ochsner U, Ostroff RM, Otis M, Parker T, Pietrasiewicz S, Resnicow DJ, Rohloff J, Sanders G, Sattin S, Schneider D, Singer B, Stanton M, Sterkel A, Stewart A, Stratford S, Vaught JD, Vrkljan M, Walker JJ, Watrobka M, Waugh S, Weiss A, Wilcox SK, Wolfson A, Wolk SK, Zhang C, Zichi D. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One*. 2010;5:e15004.
- Belongie KJ, Ferrannini E, Johnson K, Andrade-Gordon P, Hansen MK, Petrie JR. Identification of novel biomarkers to monitor beta-cell function and enable early detection of type 2 diabetes risk. *PLoS One*. 2017;12:e0182932.
- De Groote MA, Nahid P, Jarlsberg L, Johnson JL, Weiner M, Muzanyi G, Janjic N, Sterling DG, Ochsner UA. Elucidating novel serum biomarkers associated with pulmonary tuberculosis treatment. *PLoS One*. 2013;8:e61002.
- De Groote MA, Higgins M, Hraha T, Wall K, Wilson ML, Sterling DG, Janjic N, Reeves R, Ochsner UA, Belknap R. Highly multiplexed proteomic analysis of Quantiferon supernatants to identify biomarkers of latent tuberculosis infection. *J Clin Microbiol*. 2017;55:391–402.
- Hathout Y, Brody E, Clemens PR, Cripe L, DeLisle RK, Furlong P, Gordish-Dressman H, Hache L, Henricson E, Hoffman EP, Kobayashi YM, Lorts A, Mah JK, McDonald C, Mehler B, Nelson S, Nikrad M, Singer B, Steele F, Sterling D, Sweeney HL, Williams S, Gold L. Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. *Proc Natl Acad Sci U S A*. 2015;112:7153–8.
- Kiddle SJ, Steves CJ, Mehta M, Simmons A, Xu X, Newhouse S, Sattler M, Ashton NJ, Bazenet C, Killick R, Adnan J, Westman E, Nelson S, Soininen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Curtis C, Breen G, Williams SC, Lovestone S, Spector TD, Dobson RJ. Plasma protein biomarkers of Alzheimer's disease endophenotypes in asymptomatic older twins: early cognitive decline and regional brain volumes. *Transl Psychiatry*. 2015;5:e584.
- Mehan MR, Ayers D, Thirstrup D, Xiong W, Ostroff RM, Brody EN, Walker JJ, Gold L, Jarvis TC, Janjic N, Baird GS, Wilcox SK. Protein signature of lung cancer tissues. *PLoS One*. 2012;7:e35157.
- Mehan MR, Williams SA, Siegfried JM, Bigbee WL, Weissfeld JL, Wilson DO, Pass HI, Rom WN, Muley T, Meister M, Franklin W, Miller YE, Brody EN,

- Ostroff RM. Validation of a blood protein signature for non-small cell lung cancer. *Clin Proteomics*. 2014;11:32. 0275-11-32. eCollection 2014.
18. Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, Miller YE, Pass HI, Rom WN, Siegfried JM, Stewart A, Walker JJ, Weisfeld JL, Williams S, Zichi D, Brody EN. Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. *PLoS One*. 2010;5:e15003.
 19. Goldin A: XPFit - eXponential Fitting Software [www.softscientific.com/science/xpfit.html]. Accessed 5 Dec 2015.
 20. Brody EN, Gold L. Aptamers as therapeutic and diagnostic agents. *J Biotechnol*. 2000;74:5–13.
 21. Gold L. Oligonucleotides as research, diagnostic, and therapeutic agents. *J Biol Chem*. 1995;270:13581–4.
 22. Team RC. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2015.
 23. Parker R: Variable clustering in JMP [<https://community.jmp.com/t5/JMP-Blog/Variable-clustering-in-JMP/ba-p/30261>]. Accessed 31 May 2018.
 24. Gonzalez R, Suppes T, Zeitzer J, McClung C, Tamminga C, Tohen M, Forero A, Dwivedi A, Alvarado A. The association between mood state and chronobiological characteristics in bipolar I disorder: a naturalistic, variable cluster analysis-based study. *Int J Bipolar Disord*. 2018;6:5. 017-0113-5.
 25. Breiman L. Random Forests. *Mach Learning*. 2001;45:5–32.
 26. Liaw A, Wiener M: Classification and Regression by randomForest [<https://CRAN.R-project.org/doc/Rnews/>]. Accessed 13 May 2018.
 27. Whartenby KA, Small D, Calabresi PA. FLT3 inhibitors for the treatment of autoimmune disease. *Expert Opin Investig Drugs*. 2008;17:1685–92.
 28. Ley K. *Physiology of inflammation*. New York: Springer; 2013.
 29. Waskow C, Liu K, Darasse-Jèze G, Guernonprez P, Ginhoux F, Merad M, Shengelia T, Yao K, Nussenzweig M. The receptor tyrosine kinase Flt3 is required for dendritic cell development in peripheral lymphoid tissues. *Nat Immunol*. 2008;9:676.
 30. Hannum C, Culpepper J, Campbell D, McClanahan T, Zurawski S, Bazan JF, Kastelein R, Hudak S, Wagner J, Mattson J. Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of haematopoietic stem cells and is encoded by variant RNAs. *Nature*. 1994;368:643–8.
 31. Konya H, Miuchi M, Satani K, Matsutani S, Tsunoda T, Yano Y, Katsuno T, Hamaguchi T, Miyagawa J, Namba M. Hepatocyte growth factor, a biomarker of macroangiopathy in diabetes mellitus. *World J Diabetes*. 2014; 5:678–88.
 32. Shinoda K, Ishida S, Kawashima S, Wakabayashi T, Matsuzaki T, Takayama M, Shinmura K, Yamada M. Comparison of the levels of hepatocyte growth factor and vascular endothelial growth factor in aqueous fluid and serum with grades of retinopathy in patients with diabetes mellitus. *Br J Ophthalmol*. 1999;83:834–7.
 33. Satani K, Konya H, Hamaguchi T, Umehara A, Katsuno T, Ishikawa T, Kohri K, Hasegawa Y, Suehiro A, Kakishita E, Namba M. Clinical significance of circulating hepatocyte growth factor, a new risk marker of carotid atherosclerosis in patients with type 2 diabetes. *Diabet Med*. 2006;23:617–22.
 34. Morishita R, Moriguchi A, Higaki J, Ogihara T. Hepatocyte growth factor (HGF) as a potential index of severity of hypertension. *Hypertens Res*. 1999; 22:161–7.
 35. Kulseng B, Borset M, Espevik T, Sundan A. Elevated hepatocyte growth factor in sera from patients with insulin-dependent diabetes mellitus. *Acta Diabetol*. 1998;35:77–80.
 36. Bancks MP, Bielinski SJ, Decker PA, Hanson NQ, Larson NB, Sicotte H, Wassell CL, Pankow JS. Circulating level of hepatocyte growth factor predicts incidence of type 2 diabetes mellitus: the multi-ethnic study of atherosclerosis (MESA). *Metabolism*. 2016;65:64–72.
 37. Engstrom G, Hedblad B, Eriksson KF, Janzon L, Lindgarde F. Complement C3 is a risk factor for the development of diabetes: a population-based cohort study. *Diabetes*. 2005;54:570–5.
 38. Bottaro DP, Rubin JS, Faletto DL, Chan AM, Kmieciak TE, Vande Woude G, Aaronson SA. Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product. *Science*. 1991;251:802.
 39. You WK, McDonald DM. The hepatocyte growth factor/c-met signaling pathway as a therapeutic target to inhibit angiogenesis. *BMB Rep*. 2008; 41:833–9.
 40. Matsumoto K, Umitsu M, De Silva DM, Roy A, Bottaro DP. Hepatocyte growth factor/MET in cancer progression and biomarker discovery. *Cancer Sci*. 2017;108:296–307.
 41. Athauda G, Giubellino A, Coleman JA, Horak C, Steeg PS, Lee MJ, Trepel J, Wimberly J, Sun J, Coxon A, Burgess TL, Bottaro DP. C-met ectodomain shedding rate correlates with malignant potential. *Clin Cancer Res*. 2006;12: 4154–62.
 42. Jiang W. Hepatocyte growth factor and the hepatocyte growth factor receptor signalling complex as targets in cancer therapies. *Curr Oncol*. 2007; 14:66–9.
 43. Lv H, Shan B, Tian Z, Li Y, Zhang Y, Wen S. Soluble c-met is a reliable and sensitive marker to detect c-met expression level in lung cancer. *Biomed Res Int*. 2015;2015:626578.
 44. Gao HF, Yang JJ, Chen ZH, Zhang XC, Yan HH, Guo WB, Zhou Q, Gou LY, Dong ZY, Wu YL. Plasma dynamic monitoring of soluble c-met level for EGFR-TKI treatment in advanced non-small cell lung cancer. *Oncotarget*. 2016;7:39535–43.
 45. Alvarez-Perez J, Ernst S, Demirci C, Casinelli GP, Mellado-Gil J, Rausell-Palamos F, Vasavada RC, Garcia-Ocaña A. Hepatocyte growth factor/c-met signaling is required for Beta-cell regeneration. *Diabetes*. 2014;63:216.
 46. Mellado-Gil J, Rosa TC, Demirci C, Gonzalez-Pertusa JA, Velazquez-Garcia S, Ernst S, Valle S, Vasavada RC, Stewart AF, Alonso LC, Garcia-Ocana A. Disruption of hepatocyte growth factor/c-met signaling enhances pancreatic beta-cell death and accelerates the onset of diabetes. *Diabetes*. 2011;60:525–36.
 47. Linnemann AK, Baan M, Davis DB. Pancreatic beta-cell proliferation in obesity. *Adv Nutr*. 2014;5:278–88.
 48. Pennacchietti S, Michieli P, Galluzzo M, Mazzone M, Giordano S, Comoglio PM. Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene. *Cancer Cell*. 2003;3:347–61.
 49. Boccaccio C, Comoglio PM. Invasive growth: a MET-driven genetic programme for cancer and stem cells. *Nat Rev Cancer*. 2006;6:637–45.
 50. Sun K, Tordjman J, Clement K, Scherer PE. Fibrosis and adipose tissue dysfunction. *Cell Metab*. 2013;18:470–7.
 51. Bianco P, Fisher LW, Young MF, Termine JD, Robey PG. Expression of bone sialoprotein (BSP) in developing human tissues. *Calcif Tissue Int*. 1991;49: 421–6.
 52. Hunter GK, Goldberg HA. Nucleation of hydroxyapatite by bone sialoprotein. *Proc Natl Acad Sci U S A*. 1993;90:8562–5.
 53. Ogata Y. Bone sialoprotein and its transcriptional regulatory mechanism. *J Periodont Res*. 2008;43:127–35.
 54. Bechard D, Gentina T, Delehedde M, Scherpereel A, Lyon M, Aumercier M, Vazeux R, Richet C, Degand P, Jude B, Janin A, Fernig DG, Tonnel AB, Lassalle P. Endocan is a novel chondroitin sulfate/dermatan sulfate proteoglycan that promotes hepatocyte growth factor/scatter factor mitogenic activity. *J Biol Chem*. 2001;276:48341–9.
 55. Lassalle P, Molet S, Janin A, Heyden JV, Tavernier J, Fiers W, Devos R, Tonnel AB. ESM-1 is a novel human endothelial cell-specific molecule expressed in lung and regulated by cytokines. *J Biol Chem*. 1996;271:20458–64.
 56. Lv Y, Zhang Y, Shi W, Liu J, Li Y, Zhou Z, He Q, Wei S, Liu J, Quan J. The association between Endocan levels and subclinical atherosclerosis in patients with type 2 diabetes mellitus. *Am J Med Sci*. 2017;353:433–8.
 57. Lai CY, Chen CM, Hsu WH, Hsieh YH, Liu CJ. Overexpression of endothelial cell-specific molecule 1 correlates with Gleason score and expression of androgen receptor in prostate carcinoma. *Int J Med Sci*. 2017;14:1263–7.
 58. Yu YY, Huang LL, Xu XP. Endothelial cell specific molecule-1: a novel biomarker for endothelial injury in acute respiratory distress syndrome. *Zhonghua Nei Ke Za Zhi*. 2017;56:863–5.
 59. Yang WE, Hsieh MJ, Lin CW, Kuo CY, Yang SF, Chuang CY, Chen MK. Plasma levels of endothelial cell-specific Molecule-1 as a potential biomarker of oral Cancer progression. *Int J Med Sci*. 2017;14:1094–100.
 60. Karabakan M, Bozkurt A, Akdemir S, Gunay M, Keskin E. Significance of serum endothelial cell specific molecule-1 (Endocan) level in patients with erectile dysfunction: a pilot study. *Int J Impot Res*. 2017;29:175–8.
 61. Perrotti A, Chenevier-Gobeaux C, Ecarot F, Barrucand B, Lassalle P, Dorigo E, Chocron S. Relevance of endothelial cell-specific molecule 1 (Endocan) plasma levels for predicting pulmonary infection after cardiac surgery in chronic kidney disease patients: the Endolung pilot study. *Cardiorenal Med*. 2017;8:1–8.
 62. Matschinsky FM. Glucokinase as glucose sensor and metabolic signal generator in pancreatic beta-cells and hepatocytes. *Diabetes*. 1990;39:647–52.
 63. Raimondo A, Rees MG, Gloyn AL. Glucokinase regulatory protein: complexity at the crossroads of triglyceride and glucose metabolism. *Curr Opin Lipidol*. 2015;26:88–95.

64. Grimsby J, Coffey JW, Dvorozniak MT, Magram J, Li G, Matschinsky FM, Shiota C, Kaur S, Magnuson MA, Grippo JF. Characterization of Glucokinase regulatory protein-deficient mice. *J Biol Chem*. 2000;275:7826–31.
65. Lloyd DJ, St Jean DJ, Jr KRJ, Wahl RC, Michelsen K, Cupples R, Chen M, Wu J, Sivits G, Helmering J, Komorowski R, Ashton KS, Pennington LD, Fotsch C, Vazir M, Chen K, Chmait S, Zhang J, Liu L, Norman MH, Andrews KL, Bartberger MD, Van G, Galbreath EJ, Vonderfecht SL, Wang M, Jordan SR, Veniant MM, Hale C. Antidiabetic effects of glucokinase regulatory protein small-molecule disruptors. *Nature*. 2013;504:437–40.
66. Mitchell BL, Yasui Y, Li CI, Fitzpatrick AL, Lampe PD. Impact of freeze-thaw cycles and storage time on plasma samples used in mass spectrometry based biomarker discovery projects. *Cancer Inform*. 2005;1:98–104.
67. McAuley KA, Williams SM, Mann JI, Walker RJ, Lewis-Barned NJ, Temple LA, Duncan AW. Diagnosing insulin resistance in the general population. *Diabetes Care*. 2001;24:460–4.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

