Research Article
Volume 14:4, 2020
DOI: 10.37421/jmgm.2020.14.461

Journal of Molecular and Genetic Medicine

ISSN: 1747-0862

Open Access

Biomarkers for Development of Glucocorticoid-Induced Diabetes Mellitus – A Metabolomics-Based Prediction Model

Klarskov CK[1]*, Havelund JF[2], Zegers FD[3], Færgeman NJ[2], Schultz HH[4], Persson F[5], Debrabant B[3], Bjergaard UP[1,6] and Kristensen PL[1]

[1]Nordsjællands Hospital, Dyrehavevej 29, 3400 Hillerød, Capital Region of Denmark
[2]University of Southern Denmark, Center for Bioanalytical Sciences, Campusvej 55, 5230 Odense M, Denmark
[3]University of Southern Denmark, IST - EBB/Epidemiologi, Biostatistik og Biodemografi, J.B. Winsløws Vej 9, 5000 Odense C, Denmark
[4]Herlev-Gentofte University Hospital, Herlev Ringvej 75, 2730 Herlev, Capital Region of Denmark
[5]Steno Diabetes Center Copenhagen, Niels Steensens vej 2, 2820 Gentofte, Capital Region of Denmark
[6]University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Capital Region of Denmark

Abstract

Background: Glucocorticoid-induced diabetes mellitus (GIDM) is a serious side effect of glucocorticoid (GC) treatment that is associated with both increased mortality and morbidity, but not all patients develop GIDM when treated with GC. The reason is not known, and clinical risk factors predictive of type 2 diabetes do not predict GIDM. Previous metabolomics studies have found specific metabolic disturbances prior to clinical type 2 diabetes. This could also be true for GIDM. The primary aim of this study was to investigate whether distinct metabolic patterns in patients treated with high dose GC can predict development of GIDM.

Material and methods: Serum from 116 patients about to be treated with or in the first days of treatment with high-dose GC (>100 mg prednisolone equivalent) was analyzed with liquid chromatography-mass spectrometry (LC-MS) based nontargeted metabolomics. Clinical data were collected at baseline and through a 3-week follow-up period. 52 patients developed GIDM and 64 did not (control group). A logistic regression model and a predictive model was build and differences in the metabolome due to treatment with GC was tested in serum from patients without GC treatment (n=6) and patients with GC treatment (n=107).

Results and discussion: At univariate analysis three metabolites were associated with the development of GIDM. These metabolites could not be annotated to specific metabolites. A multi-metabolite approach could not predict GIDM, and this is different from previous findings in T2DM. This supports the hypothesis that the etiology of T2DM and GIDM is different. The biological significance of our finding remains unknown, but with the rapid development in the field of metabolomics and databases with increasing numbers of characterized metabolites, these metabolites may be identified.

Conclusion: Our data indicate that the typical metabolic shifts in T2DM are not the same in GIDM. This supports the hypothesis that GIDM may have a pathophysiology different from T2DM. Furthermore, our data suggest that there is potential for identifying patients at risk of GIDM before clinical manifestation.

Keywords: Glucocorticoids • Diabetes • Metabolomics • Biomarkers • Metabolites

Abbreviations: ALAT: Alanine Aminotransaminase • ASAT: Aspartate Aminotransferase • BMI: Body mass index • CNS: Central Nervous System • CRP: C-reactive Protein • ECOG: Eastern Cooperative Oncology Group • e-GFR: estimated Glomerular Filtration Rate • FA: Formic Acid • GC: Glucocorticoid • GC-MS: Gas Chromatography-mass Spectrometry • GIDM: Glucocorticoid-Induced Diabetes Mellitus • HbA1c: Hemoglobin A1c • HDL: High-density Lipoprotein • INR: International Normalized Ratio • LCFA: Long Chain Fatty Acid • LC-MS: Liquid Chromatography-Mass Spectrometry • LDH: Lactate Dehydrogenase • LDL: Low-density Lipoprotein • MD: Medical Doctor • NVEK: National Committee on Health Research Ethics • QC: Quality Control • T2DM: Type 2 Diabetes Mellitus • TSH: Thyrotropin • UPLC: Ultra Performance Liquid Chromatography • WHO: World Health Organization • VLDL: Very-low-density Lipoprotein

Introduction

Glucocorticoid-induced diabetes mellitus (GIDM) is a well-known and potentially fatal side effect of glucocorticoid (GC) treatment that is associated with both increased mortality and morbidity [1]. In the Danish population, three per cent receive oral steroids at any given time [2]. GCs affect glucose metabolism in several ways including induction of insulin resistance. Numerous mechanisms contribute (to an unknown extent) to this, e.g. down-regulation of the glucose transporting proteins, increased gluconeogenesis in the liver, and inhibition of beta cell insulin secretion [3]. Early identification of people at risk of GIDM is important to prevent hyperglycemic symptoms and reduce morbidity and mortality, especially because efficient insulin-based treatment of GIDM exists [4,5]. The risk of developing GIDM is affected by the dose, duration, and the accumulated dose of GC treatment [6]. However, recent studies have been conflicting in relation to identification of clinical risk factors for developing GIDM (4). Patients with GIDM weigh less and have less family

***Address for Correspondence:** Dr. Carina Kirstine Klarskov, Department of Endocrinology and Nephrology, Nordsjællands Hospital, Dyrehavevej 29, 3400 Hillerød, Capital Region of Denmark, E-mail: Carina.kirstine.klarskov@regionh.dk*

***Copyright:** © 2020 Klarskov CK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

Received 21 August 2020; **Accepted** 24 August 2020; **Published** 31 August 2020

history of diabetes than patients with type 2 diabetes mellitus (T2DM) and does not develop retinopathy in the same speed as T2DM albeit the same diabetes duration and HbA1c - this could indicate that the etiology for GIDM is different from that of T2DM [7].

Metabolomics is an advancing field in the characterization (identification and quantification) of metabolites in biological systems and represents a real-time functional "fingerprint" of the biological system that is investigated [8]. Metabolomics has emerged as a powerful tool to detect early disease states before clinical manifestation and one study has shown significant changes in the metabolome during GC treatment within a very short time period [9]. The use of metabolomics has already led to many findings including metabolome shifts several years before the diagnosis of diabetes [10,11]. Differences in metabolites such as branched-chain and aromatic amino acids, fatty acids and lipoprotein lipids have been associated with increased risk of developing diabetes [12]. In a cohort of 5,271 Finnish 31-year-old men and women, a multi-metabolite score was associated with a future ten-fold increased diabetes risk [13]. It is unknown whether this association between specific metabolites and incident type 2 diabetes can be transferred to GIDM.

The aim of the project was to identify novel biomarkers for developing GIDM during GC therapy by a *post hoc* metabolomics analysis of a prospective observational study of cancer patients commencing high-dose GC therapy (equivalent of ≥100 mg prednisolone pr. day) [4].

Materials and Methods

Objectives

The primary objective was to investigate the combined potential of all metabolic biomarkers to predict development of GIDM as well as to identify individual metabolites associated with a higher risk for GIDM. Furthermore, we examined if there is a significant difference in metabolomic markers between GC-exposed and non-exposed individuals.

Population

In our prospective observational study, the most important inclusion criteria were a diagnosis of cancer (solid tumors, myeloma or malignant lymphoma) verified histologically and high dose GC treatment with oral prednisolone ≥ 100 mg/day or injection of methylprednisolone ≥ 80 mg/day. The most important exclusion criteria was known type 1 or 2 diabetes (for full description, see [4]). Data were collected on demographics, family history of diabetes, medical history, vital signs, laboratory tests, and information about GC treatment. Biobank specimen collection was optional [4].

For inclusion in the present *post hoc* metabolomics analysis, serum in the biobank and data on GIDM status were needed. One patient was excluded due to a prior diagnosis of GIDM (based on baseline HbA1c measurement). Of the 140 patients, 116 met all inclusion criteria and none of the exclusion criteria and were included in the final analysis. Of the 116 patients, 52 developed GIDM (45%).

Ethical considerations and informed consent

The project was approved by The Danish National Committee on Health Research Ethics with Journal-nr.:H-19002730 and by The Danish Data Monitoring Board with Journal No. VD- 2018-250, with I-Suite no. 6491 and transfer of data from the prospective observational study and biobank with journal no. NOH-2015-022, with I-Suite no. 03811. The Biofluids were collected according to current rules (The Danish Health Act Chapter 5 and 7: consent and the right to self-determination over biological material).

Metabolite extraction from serum for LC-MS

Serum was analyzed with nontargeted metabolomics by liquid chromatography-mass spectrometry (LC-MS) profiling.

Sample preparation: 200 µl serum was thawed on ice before addition of 800 µl extraction solvent (100% methanol). Four empty samples (blanks) were prepared the same way. The samples were vortexed for 30 sec and left at -20°C for 30 min before centrifugation at 16,000 g for 20 min at 4°C. Supernatants (750 µl) were transferred to new tubes and lyophilized. The lyophilized samples were resuspended in 20 µl 1% formic acid (FA).

Metabolomics analysis and data processing: 5 µl were injected using a Vanquish Horizon UPLC (Thermo Fisher Scientific, Germering, Germany) and compounds separated on a Zorbax Eclipse Plus C18 guard (2.1 × 50 mm and 1.8 µm particle size, Agilent Technologies, Santa Clara, CA, USA) and an analytical column (2.1 × 150 mm and 1.8 µm particle size, Agilent Technologies, Santa Clara, CA, USA)) kept at 40°C. The analytes were eluted using a flow rate of 400 µL/min and the following composition of eluent A (0.1% formic acid) and eluent B (0.1% formic acid, acetonitrile) solvents: 3% B from 0 to 1.5 min, 3-40% B from 1.5 to 4.5 min, 40-95% B from 4.5 to 7.5 min, 95% B from 7.5 to 10.1 min and 95 to 3% B from 10.1 to 10.5 min before equilibration for 3.5 min with the initial conditions. Using a 6-port valve and a secondary pump, the guard wash backflushed from 9-10 min with a flow of 1 ml/min with 95% B. The flow from the UPLC was coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) for mass spectrometric analysis in both positive and negative ion mode using the following general settings for MS1 mode: resolution: 120.000, AGC target: 3e6, maximum injection time: 200 ms, scan range 65-975 m/z and lock mass: 391.28429/112.98563 (pos/neg mode). For compound fragmentation MS1/ddMS2 mode was used with the following general settings: resolution: 60.000/15.000, AGC target: 1e6/1e5, maximum injection time: 50/100 ms, scan range 65-975 m/z, loop count: 5, isolation with: 2 m/z and normalized collision energy: 35/38 (pos/neg). A pooled sample was generated and used for quality control (QC) and compound fragmentation. The sequence was constructed with analysis of the blanks first followed by system equilibrations runs using QC sample, MS1/ddMS2 runs for compound fragmentation on QC sample, another QC run and, finally, analysis of the samples in randomized order in MS1 mode with a QC sample injected every third run. A MS2 inclusion list was generated from the blanks and the last equilibration runs using Compound Discover v. 3.0 (Thermo Fisher Scientific) to ensure fragmentation of the later extracted relevant features. The features found in the blank were removed from the inclusion list if they were not > 5 x more abundant in the QC samples.

Raw data was processed with MzMine (v 2.42) [14]. In brief, the following modules were used: Mass detection, ADAP chromatogram builder, ADAP deconvolution, Join aligner, Isotopic peak grouper, Gap filling (same RT and m/z range) and Identification in local spectra database search; all with 5 ppm mass tolerance and 0.25 RT tolerance when possible. Final peak list included features found in at least 75% of the samples, which had at least 2 peaks in an isotope pattern. Compounds were annotated at Metabolomics Standards Initiative (MSI) level 3 using local MS/MS spectra databases of National Institute of Standards and Technology 17 (NIST17) and Mass Bank of North America (MoNA) [15]. MSI level 4 annotations where achieved by searching the molecular formula in Human Metabolite Database [16]. After compound annotation, the datasets were corrected for signal drift using statTarget [17]. Finally, the signal was normalized using the QC sample, auto-scaled and log2 transformed in Metaboanalyst [18].

Statistical methods

For the primary objective our analyses rely on logistic regression modelling with development of GIDM as outcome. A patient was classified as having GIDM if diabetes (defined in our prospective observational study in cancer patients as two random plasma glucose measurements ≥11.1 mmol/L) developed during GC therapy [4]. We used age, gender, body mass index (BMI) and baseline HbA1c as covariates in all analyses. The logistic model including only these demographic and clinical covariates (no metabolites) is referred to the basic logistic model. Before including metabolites into the analysis, we tested the fit of this basic logistic model using Stukel's goodness of fit test [19]. For the analysis of differences in metabolomic markers between GC-exposed and non-exposed individuals linear modelling was applied.

There were 5 missing values for BMI and 4 missing values for baseline HbA1c, whereas the rest of the clinical covariates were complete. The missing values were imputed using *k*-nearest neighbor with 5 nearest neighbors. In total, there were 2,214 different metabolites requiring no further preparation as there were no missing values for all metabolites for all patients.

Univariate association analysis

For each metabolite, the basic logistic model was extended by the metabolite and the p-value for each metabolite was stored. These p-values were corrected using Benjamini-Hochberg's correction for multiple testing, and a metabolite was considered associated with GIDM if it had a significance level of the false discovery rate (FDR) ≤ 5% [20]. Pearson correlations between the 10 metabolites with the lowest FDR's were obtained and plotted (Figure 1). The FDR is a method of conceptualizing the rate of type I errors in null hypothesis testing when conduction multiple comparisons.

Predictive modeling

To elucidate which metabolites in combination can be used to predict the development of GIDM, we extended the basic logistic model by simultaneously including all metabolites. We used LASSO logistic regression together with stability selection as described in Meinshausen et al. and the improved sampling approach as described in Shah et al. [21,22].

The procedure applies LASSO logistic regression to n random subsamples of the whole data, such that on average q predictors are selected. Overall selection probabilities are estimated for each predictor and the final predictors were chosen as those with a selection probability larger than 0.6. This approach allows to control the expected number of predictors with low selection probabilities (ENSP), i.e., the number of false discoveries. Into this stability selection we have forced inclusion of the clinical variables described above. The parameters q=60 and θ=0.6 were used on n=100 subsamples, implying that the final model consisted of the metabolites that were present in at least 60% of the subsamples. To assess the prediction performance of the final model we calculated the area under the ROC curve (AUC) using the whole dataset [23].

Association between GC exposure and metabolites

To find metabolomic markers that systematically differ between GC-exposed and non-exposed individuals, the cohort was divided into two groups based on whether they had their blood sample taken before (n=6) or after (n=107) their first GC treatment. Patients that had received GC the same day as blood sampling (n=3) were left out due to the unknown relationship in time between GC intake and blood sampling, which could have been just minutes apart and thus in reality still unexposed to GCs. Before this subdivision, exclusion of individuals with missing data and no information on when GC treatment was commenced in relation to the time blood was drawn, was also excluded. For each metabolite separately, we regressed metabolite levels against sex, age, BMI, baseline HbA1c, and whether or not the blood sample was taken before the first GC treatment. Prior to this analysis metabolite levels were log-transformed to improve normality assumptions underlying the regression-based testing. The p-value and standard error for the before/after-effect were extracted and stored together with the coefficient. Benjamini-Hochberg's correction for multiple testing was applied, and effects considered significant if FDR ≤ 5% [20].

All computations were made in R version 3.6.1. Imputation of missing values was done with the function kNN from the package VIM. Stukel's goodness of fit test was made using gof from the package LogisticDx. The correlation plot was made with the function chart. Correlation from the package Performance Analytics. The predictive modelling used the function stabsel from the package stabs, K-fold cross validation was made using the caret-package and AUC for the model on the test data was made with the package pROC [24-30].

Results

Univariate association analysis

The basic logistic model with the covariates as described above and no interactions or quadratic effects was tested using Stukel's goodness of

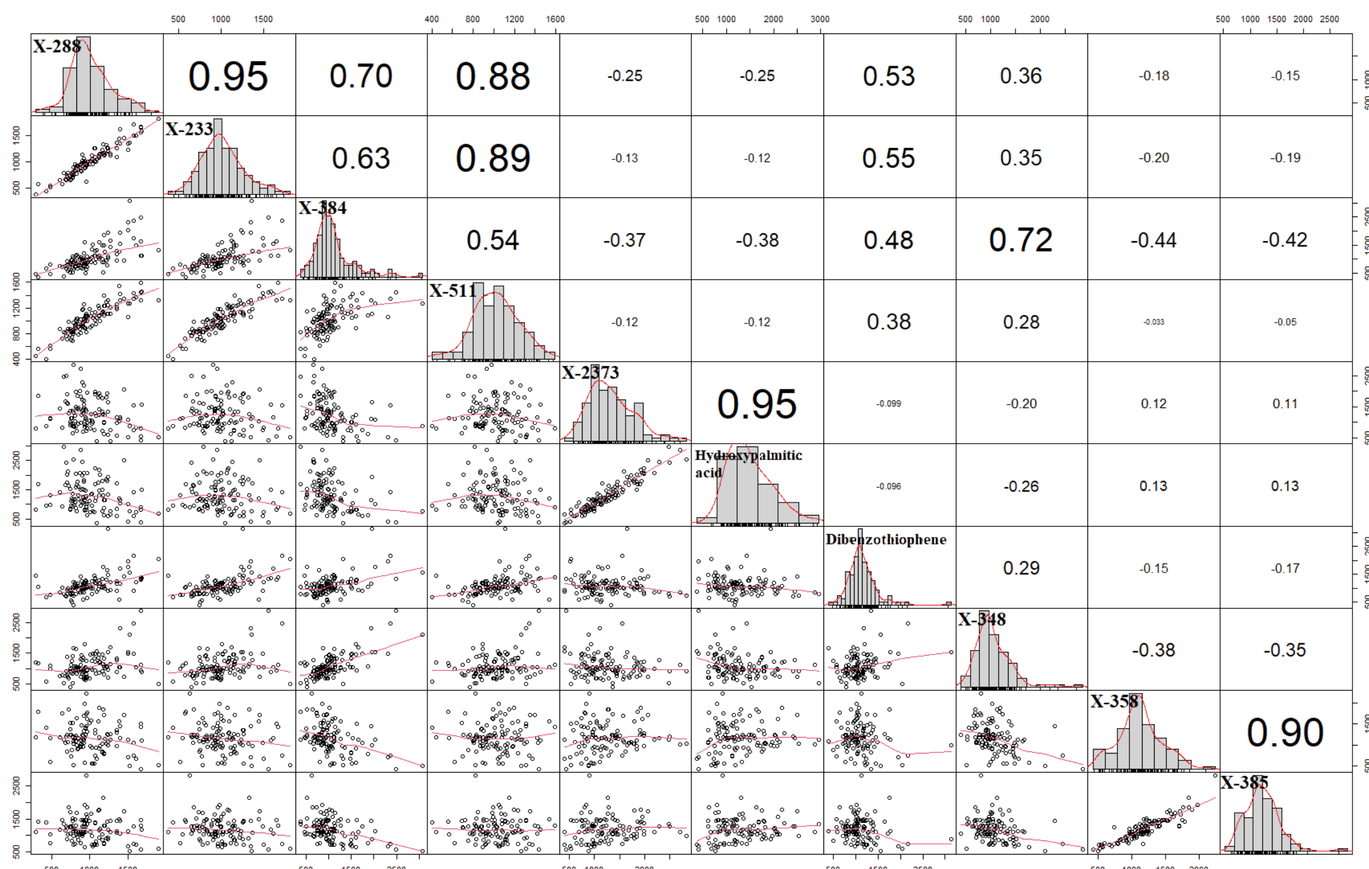

Figure 1. Pairwise scatterplots of the 10 topmost significant associated metabolites as well as Pearson correlation coefficients to illustrate the degree of linear dependencies between the metabolites.

fit test, which showed no significant *p*-value, indicating a reasonable fit. The main results (top 10 metabolite candidates) from the univariate association analysis are presented in Tables 1 and 2. Two of the top ten metabolites were identified from metabolomic databases (β-hydroxy-palmitic acid and dibenzothiophene), but had a FDR >5% and were therefore not significantly associated with the development of GIDM. Three metabolites (X-288, X-233 and X-384) were associated with development of GIDM with the odds ratios 1.004, 1.004, and 1.004 respectively and an FDR of 2.9%. That is, an increase in one unit of just one of these metabolites was associated with 0.4% increased odds of developing GIDM. Figure 1 contains pairwise scatterplots of the 10 topmost metabolites as well as Pearson correlation coefficients to illustrate the degree of linear dependencies between the metabolites. Especially, the three associated metabolites (top left corner) show high correlations ($P \geq 0.63$), with X-288 and X-233 being almost perfectly correlated (P=0.95).

Predictive modeling

In Figure 2, the metabolites showing the highest selection frequencies are displayed. The frequencies indicate how often a metabolite was selected into the prediction model during the stability selection procedure. No metabolites exceeded the desired selection probability of 0.6, hence no metabolites contributed consistently to predicting the development of GIDM.

Table 3 shows the final prediction model including the four clinical covariates (age, gender, BMI and baseline HbA1c). The model yielded an AUC of 0.713 indicating moderate prediction performance. The model showed that for every mmol/mol increase in HbA1c, the odds of developing GIDM increased by 16% (95% CI: 6.9 – 128), *P*=0.001.

Changes in the metabolome due to GCs

We found a high number of metabolites associated with GC exposure. Totally, 22 metabolites had FDR ≤ 5% (Table 4). Ten of them were identifiable and 12 were not. Among the identifiable metabolites were for example drug metabolites such as beta-hydroxy prednisolone.

Discussion

Univariate association analysis

Two metabolites from the top-ten candidates were identified in the metabolomic databases (Table 2). These metabolites (β-hydroxy-palmitic acid, a long chain fatty acid and dibenzothiophene, which can derive from several sources, including being a keratolytic drug derivative) were however not significantly associated with the risk of developing GIDM. 3 metabolites, which were significantly associated with the development of GIDM (X-288, X-233 and X-384) could not be identified in metabolomics databases and were therefore not components of well-described metabolic pathways. The lack of associations between well-known metabolites and GIDM was surprising since some metabolites have been demonstrated to be associated with the risk of developing T2DM: Guasch-Ferré et al. conducted a systematic review of metabolite studies in prediabetes and T2DM with metabolomics data from more than 8,000 individuals and concluded that several blood amino acids are consistently associated with the risk of developing T2DM [8]. Since all amino acids are identifiable in the metabolome databases as well as metabolites from well-known metabolic pathways, our 3 unidentified metabolites significantly

Table 1. Baseline characteristics of participants.

Variables	No diabetes (n=64)	Diabetes per definition (n=39)	Treatment-requiring diabetes (n=13)	Total (n=116)
Males (%)	36 (56.2)	22 (56.4)	10 (76.9)	68 (58.6)
Age (Mean (SD))	69.45 (10.81)	66.87 (8.32)	68.62 (11.13)	68.49 (10.06)
BMI (Mean (SD))	25.39 (5.54)	24.08 (4.68)	23.82 (3.86)	24.79 (5.12)
Performance status (%)				
0	4 (6.2)	1 (2.6)	0 (0.0)	5 (4.3)
1	31 (48.4)	13 (33.3)	2 (15.4)	46 (39.7)
2	16 (25.0)	17 (43.6)	5 (38.5)	38 (32.8)
3	13 (20.3)	8 (20.5)	6 (46.2)	27 (23.3)
Primary tumor (%)				
Lung cancer	18 (28.1)	11 (28.2)	3 (23.1)	32 (27.6)
Breast cancer	15 (23.4)	7 (17.9)	0 (0.0)	22 (19.0)
Prostate cancer	13 (20.3)	11 (28.2)	5 (38.5)	29 (25.0)
Other cancers	18 (28.1)	10 (25.6)	5 (38.5)	33 (28.4)
Lung metastases (%)	33 (51.6)	27 (69.2)	11 (84.6)	71 (61.2)
Brain metastases (%)	62 (96.9)	39 (100.0)	12 (92.3)	113 (97.4)
Liver metastases (%)	52 (81.2)	31 (79.5)	10 (76.9)	93 (80.2)
Bone metastases (%)	1 (1.6)	1 (2.6)	0 (0.0)	2 (1.7)
Pancreatic metastases (%)	64 (100.0)	39 (100.0)	13 (100.0)	116 (100.0)
Bone marrow metastases (%)	62 (96.9)	38 (97.4)	13 (100.0)	113 (97.4)
Lymph node metastases (%)	45 (70.3)	28 (71.8)	7 (53.8)	80 (69.0)
Alcohol abuse now or earlier (%)	11 (17.7)	9 (23.1)	2 (15.4)	22 (19.3)
Smoking now or earlier (%)	44 (68.8)	27 (69.2)	8 (61.5)	79 (68.1)
Atherosclerosis (%)	6 (9.4)	4 (10.3)	2 (15.4)	12 (10.3)
Hypertension (%)	25 (39.1)	16 (41.0)	4 (30.8)	45 (38.8)
Liver disease (%)	2 (3.1)	1 (2.6)	0 (0.0)	3 (2.6)
Pancreatic disease (%)	1 (1.6)	1 (2.6)	0 (0.0)	2 (1.7)
Hospitalization (%)				
Outpatient	32 (50.0)	15 (38.5)	0 (0.0)	47 (40.5)
Inpatient	20 (31.2)	18 (46.2)	10 (76.9)	48 (41.4)
Alternating	12 (18.8)	6 (15.4)	3 (23.1)	21 (18.1)
Prednisolone start dose (mg) (mean (SD))	238 (84)	302 (369)	269 (60)	263 (224)
Cumulated dose of prednisolone (mg) (mean (SD))	3832 (1032)	3714 (1344)	3975 (811)	3806 (1122)
Previous diabetes	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)

Table 2. Top 10 metabolites in the univariate association analysis ordered after lowest FDR and *p*-value. 3 metabolites had significant FDR's and thus increase the odds of developing GIDM.

S. No	Metabolite	Coefficient	Standard error	OR	p value	FDR	ID
1	X-288	0.004	0.001	1.004	3.30E-05	2.90E-02	
2	X-233	0.004	0.001	1.004	3.70E-05	2.90E-02	
3	X-384	0.004	0.001	1.004	3.90E-05	2.90E-02	
4	X-551	0.005	0.001	1.005	1.80E-04	1.00E-01	
5	X-2373	-0.002	0.001	0.998	1.60E-03	3.70E-01	
6	β-hydroxy-palmitic acid[a]	-0.002	0	0.998	2.10E-03	3.70E-01	LCFA
7	Dibenzothiophene[b]	0.003	0.001	1.003	2.10E-03	3.70E-01	Ketolytic drug
8	X-348	0.002	0.001	1.002	2.50E-03	3.70E-01	
9	X-358	-0.002	0.001	0.998	2.70E-03	3.70E-01	
10	X-385	-0.002	0.001	0.998	2.80E-03	3.70E-01	

Note: [a]Annotated at metabolomics standard initative (MSI) level 3
[b]Annotated at metabolomics standard initative (MSI) level 4
Metabolites with X-(numbers) are unidentified. LCFA=long chain fatty acid.

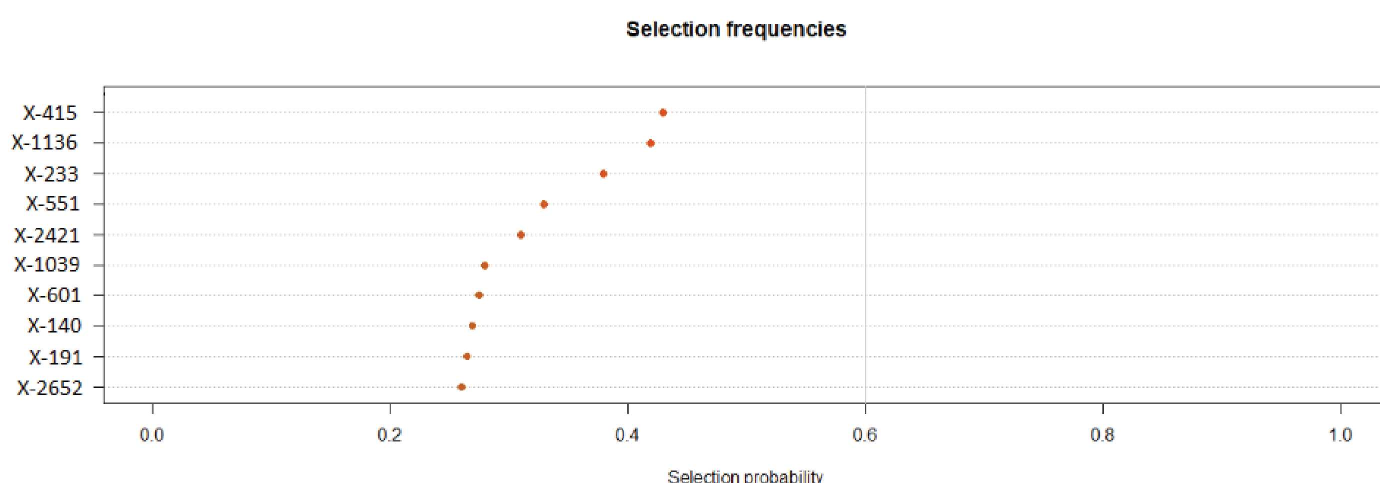

Figure 2. The 10 metabolites with the highest selection frequencies for the multivariable prediction. The frequencies indicate how often a metabolite was selected into the prediction model during the stability selection procedure. As visualized, the cut-off at 0.6 was not achieved.

Table 3. Prediction model for the development of GIDM.

Predictors	OR	Standard error	95%-CI	P value
Intercept	0.088	2.413	0.001 – 9.125	0.314
Age	0.968	0.021	0.928 – 1.007	0.111
Gender	1.076	0.420	0.471 – 2.463	0.862
BMI	0.938	0.042	0.861 – 1.016	0.128
HbA1c	1.163	0.046	1.069 – 1.282	0.001

Note: Observations 116. AUC for whole dataset: 71.3 % (61.9 % - 80.6 %)

Table 4. 22 metabolites significantly associated with treatment with high dose glucocorticoid. 107 patients had bloods samples taken after the initiation of glucocorticoids (exposed) and 6 patients had blood samples taken before the initiation of glucocortoids (non-exposed).

S. No	Metabolite	Coefficient	exp(coefficient)	Standard error	*P* value	FDR	ID
1	Cascarillin 1[b]	1.853	0.3016.38	6.3820.3	1.30E-081.30E-08	2.60E-052.60E-05	Phytochemical
2	X-776	-3.025	0.5020.0486	0.0490.5	2.40E-082.40E-08	2.60E-052.60E-05	
3	Cascarillin 2[b]	-2.633	0.4820.0719	0.0720.48	3.00E-073.00E-07	2.20E-040.00022	Phytochemical
4	β-Hydroxyprednisolone[b]	-2.409	0.4770.0899	0.0900.48	1.80E-061.80E-06	1.00E-030.001	Drug-derivative
5	X-167	-2.177	0.4550.113	0.1130.46	5.60E-065.60E-06	2.50E-030.0025	
6	Glu-Phe (Glutamic acid-phenylalanine dipeptide)[a].	-0.615	0.1310.541	0.5410.13	7.40E-067.40E-06	2.70E-030.0027	Dipeptide
7	X-1133	-2.024	0.4390.132	0.1320.44	1.10E-051.10E-05	3.50E-030.0035	
8	Isomontanolide[b]	-2.568	0.5840.0767	0.0770.58	2.60E-052.60E-05	7.30E-030.0073	Phytochemical
9	Lysylglycine (Lycine-glycine dipeptide)[a].	-1.459	0.3470.232	0.2320.35	5.30E-055.30E-05	1.30E-020.013	Dipeptide
10	X-2964	-2.135	0.5180.118	0.1180.52	7.40E-057.40E-05	1.60E-020.016	
11	X-4005	1.957	0.4777.08	7.0760.48	8.00E-058.00E-05	1.60E-020.016	
12	X-287	-2.437	0.5970.0874	0.0870.6	8.70E-058.70E-05	1.60E-020.016	

13	X-792	-2.108	0.5190.121	0.1210.52	9.40E-059.40E-05	1.60E-020.016	
14	Nigakilactone[b]	-2.141	0.5410.118	0.1180.54	1.40E-040.00014	2.20E-020.022	Phytochemical
15	X-851	1.472	0.3824.36	4.3590.38	2.00E-040.0002	2.90E-020.029	
16	X-4641	-1.962	0.5220.141	0.1410.52	2.80E-040.00028	3.70E-020.037	
17	Cortisolsulfate[b]	-2.324	0.6190.0979	0.0980.62	2.90E-040.00029	3.70E-020.037	Organic compound – sulfated steroid
18	X-1199	-0.885	0.2430.413	0.4130.24	4.30E-040.00043	4.70E-020.047	
19	Dehydrocorticosterone[b]	-1.406	0.3870.245	0.2450.39	4.40E-040.00044	4.70E-020.047	Inactive form of corticosterone
20	Dimethylprostaglandin[a]	0.358	0.0991.43	1.4300.099	4.60E-040.00046	4.70E-020.047	Synthetic prostaglandin analogue
21	X-3388	-0.788	0.2180.455	0.4550.22	4.60E-040.00046	4.70E-020.047	
22	X-1857	1.493	0.4144.45	4.4530.41	4.70E-040.00047	4.70E-020.047	

Note: [a]Annotated at metabolomics standard initative (MSI) level 3
[b]Annotated at metabolomics standard initative (MSI) level 4
Metabolites with X-(number) are unidentified

associated with development of GIDM cannot be amino acids or other well-known metabolites, which may suggest that GIDM and T2DM/prediabetes have important pathophysiological differences. New studies comparing the metabolome of patients with GIDM and patients with T2DM are needed to explore this. Metabolomics databases are in rapid development and new information are being added daily, however, it is only a minor percentage of metabolites from untargeted metabolomics that are identifiable - as in our study [31].

As visualized by the scatter plots in Figure 1, there are very strong correlations between the 3 metabolites X-288, X-233 and X-384 that were shown to be associated with GIDM in the univariate analysis. An explanation could be that the metabolites are very close in the underlying metabolic network. However, it has been shown that metabolite correlations do not necessarily correspond to proximity in the biochemical network as non-neighboring metabolites can also be highly correlated. Camacho et al. found by simulations that strong correlations are likely due to chemical equilibrium and may be due to stronger mutual control of a single enzyme or a variation of a single enzyme level much above others [32].

Predictive modeling

Multi-metabolite profiles have been shown to predict development of various diseases including diabetes [10,13,33]. During our model selection procedure, we allowed up to 60 factors (metabolites and 4 clinical covariates) to be simultaneously selected into the prediction model for developing GIDM. No metabolites were found to consistently contribute to predicting GIDM. This is different from T2DM where multi-metabolite risk scores have been developed. Comprehensive metabolic profiling of patients who develop GIDM may help to better understand the pathophysiology behind GIDM and subsequently possibly target preventive interventions for individuals at risk of developing GIDM and target treatment for patients who have already developed GIDM.

Changes in the metabolome due to GCs

At least 22 metabolites were associated with GC exposure. Drug metabolites such as beta-hydroxy prednisolone and metabolites in the steroid hormone pathways were expected to be associated with GC exposure, which indicates that the trustworthiness of the metabolomics method in our study is good. The significance of the association between the di peptides and the phytochemicals and GIDM is unclear. Dipeptides are often short-lived compounds from breakdown of proteins to amino acids. Plasma levels of amino acids are correlated to plasma glucagon levels, which are an important driver for hyperglycemia, and it has been demonstrated in mechanistic studies that prednisolone can stimulate glucagon secretion. Hence, our findings could indicate that GC exposure elicit hyperglycemia via amino acid driven hyperglucagonemia.

In a targeted metabolomics study Bordag et al. found, that treatment with GCs lead to immediate changes (within 6 hours) in the metabolome in healthy males without evident clinical manifestations [9]. A total of 150 out of 214 metabolites had changed at some time point after GC (4 collection points in 24 hours) and 9 was significantly changed at all 4 time points. The inter-person variability was high and remained uninfluenced by the intervention. Thus, the 22 metabolites associated with GC exposure in our study are reflections of both the difference in time from GC exposure and interindividual differences. Since the study by Bordag et al. used a targeted metabolomics method, it was not possible to directly compare the unidentified metabolites in our study with the metabolites in the Bordag study. In the Bordag study however, 18-Hydroxycorticosterone, Corticosterone and Cortisol were downregulated after Dexamethasone intake, which would be expected since Dexamethasone normally inhibits the pathway of steroid hormone syntheses, and in our study other metabolites in the same steroid pathway (Cortisolsulfate and Dehydrocorticosterone) were also downregulated. This indicates consistency between the two studies.

Strengths and limitations

The strength of the current study is that it is the largest metabolomics study in patients with GIDM and that the majority of patients had blood drawn at the same time of day (10.30-12.30), which is important since the metabolome is known to exhibit circadian rhythmicity [34,35].

Since the metabolome consists of very different compounds like lipids and amino acids that differ both in the presence in tissues and in concentrations, no single metabolomic method can currently measure the whole metabolome and there are interlaboratory variations. Furthermore, the metabolome can be altered by exogenous substances like foods or endogenous by metabolism, in a subject specific manner. In the GC exposure analysis, we had a small sample of n=6. 22 metabolites were associated with GC exposure, but it's possible, that we had found more metabolites in a larger sample. As previously discussed, no data on concurrent medications and nutrition were recorded and this could bias our results.

Conclusion

This metabolomics study in a cohort of cancer patients treated with GC shows significant differences in the metabolome between patients developing GIDM and those who do not. Furthermore, changes in the metabolome due to GC treatment were observed in 22 metabolites which support previous findings. The results of our study suggest that there could be a potential for using metabolomics to better understand the development of GIDM (in contrast to development of T2DM) and perhaps in future to be able to identify patients at risk of GIDM before clinical manifestation. This could lead to preventative measures with the potential of lowering the increased mortality and morbidity related to GIDM. It could also form the basis for mechanistic and clinical studies to prevent GIDM during GC therapy by either personal risk stratification to GC-

sparing treatments or by the development of new drugs to treat the metabolic disturbances before clinical manifestation of GIDM.

Acknowledgments

This study was conducted with support from a Research Grant from Nordsjællands Hospital and The Foundation of Harald Jensen and Wife.

References

1. http://www.ncbi.nlm.nih.gov/pubmed/28555464%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/28555464
2. http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2016-015237
3. Oyer, DS, A Shah and S Bettenhausen. "How to Manage Steroid Diabetes in the Patient with Cancer". *J Support Oncol* 9(2006): 479-483.
4. Schultz, Helga, Svend Aage Engelholm, Eva Harder and Ulrik Pedersen-Bjergaard, et al. "Glucocorticoid-Induced Diabetes in Patients with Metastatic Spinal Cord Compression". *Endocrine Connections* 2(2018): 719-726.
5. Schultz, Helga, Ulrik Pedersen-Bjergaard, Andreas Kryger Jensen and Svend Aage Engelholm, et al. "The Influence on Survival of Glucocorticoid Induced Diabetes in Cancer Patients with Metastatic Spinal Cord Compression". *Clin Transl Radiat Oncol* 11(2018): 19-25.
6. van Raalte, DH, Diamant M. "Steroid diabetes: From Mechanism to Treatment?" *Neth J Med* 72(2014): 62-72.
7. Simmons, Lisa R, Lynda Molyneaux, Dennis K Yue and Elizabeth L Chua. "Steroid-Induced Diabetes: Is It Just Unmasking of Type 2 Diabetes?" *ISRN Endocrinol* 2(2012): 1-5.
8. Guasch-Ferré, Marta, Adela Hruby, Estefanía Toledo and Clary B Clish, et al. "Metabolomics in Prediabetes and Diabetes : A Systematic Review and Meta-analysis". *Diabetes Care* 39(2016): 833-846.
9. Bordag, Natalie, Sebastian Klie, Kathrin Jürchott and Janine Vierheller, et al. "Glucocorticoid (dexamethasone)-Induced Metabolome Changes in Healthy Males Suggest Prediction of Response and Side Effects". *Sci Rep* 5(2015): 1-12.
10. Floegel, Anna, Norbert Stefan, Zhonghao Yu and Kristin Mühlenbruch, et al. "Identification of Serum Metabolites Associated with Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. *Diabetes* 62(2013): 2-4.
11. Walford, Geoffrey A, Bianca C Porneala, Marco Dauriz and Jason L Vassy, et al. "Metabolite Traits and Genetic Risk Provide Complementary Information for the Prediction of Future Type 2 Diabetes". *Diabetes Care* 37(2014): 2508-2514.
12. Gonzalez-Franquesa, Alba, Alison M. Burkart, Elvira Isganaitis, and Mary-Elizabeth Patti. "What Have Metabolomics Approaches Taught Us about Type 2 Diabetes?" *Curr Diab Rep* 16(2016):2-8.
13. Ahola-Olli, Ari V, Linda Mustelin, Maria Kalimeri and Johannes Kettunen, et al. "Circulating Metabolites and the Risk of Type 2 Diabetes: A Prospective Study of 11,896 Young Adults from four Finnish Cohorts". *Diabetologia* 62(2019): 2298-2309.
14. Pluskal, Tomáš, Sandra Castillo, Alejandro Villar-Briones and Matej Orešič. "MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data". *BMC Bioinformatics* 11(2010): 395.
15. Salek, Reza M, Christoph Steinbeck, Mark R Viant and Royston Goodacre, et al. The Role of Reporting Standards for Metabolite Annotation and Identification in Metabolomic Studies. *Gigascience* 2(2013): 2047-217X.
16. Wishart, David S, Yannick Djoumbou Feunang, Ana Marcu and An Chi Guo, et al. "HMDB 4.0: The Human Metabolome Database for 2018". *Nucleic Acids Res* 46(2018): D608-617.
17. Luan, Hemi, Fenfen Ji, Yu Chen and Zongwei Cai. "statTarget: A Streamlined Tool for Signal Drift Correction and Interpretations of Quantitative Mass Spectrometry-Based Omics Data". *Anal Chim Acta* 1036(2018): 66-72.
18. Chong, Jasmine, David S Wishart and Jianguo Xia. "Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis". *Curr Protoc Bioinforma* 68(2019): e86.
19. Stukel TA. Generalized Logistic Models. *J Am Stat Assoc* 83(1988): 426-431.
20. Benjamini, Yoav and Yosef Hochberg. Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser* B 57(1995): 289-300.
21. Meinshausen, Nicolai and Peter Bühlmann. Stability Selection. *J R Stat Soc Ser B* 72(2010): 417-473.
22. Shah, Rajen D and Richard J Samworth. Variable Selection with Error Control: Another Look at Stability Selection. *J R Stat Soc Ser B Stat Methodol* 75(2013): 55-80.
23. Swets, John A. Measuring the Accuracy of Diagnostic Systems. *Science* 240(1988): 1285-1293.
24. R Core, Team R. A Language and Environment for Statistical. R Foundation for Statistical Computing, Vienna, Austria.
25. Kowarik, Alexander and Matthias Templ. Imputation with the R package VIM. *J Stat Softw* 74(2016): 1-16.
26. Dardis, C. LogisticDx: Diagnostic Tests for Models with a Binomial Response. *R Packag Ver* 2(2015): 2-6.
27. Peterson, Brian G, Peter Carl, Kris Boudt and Ross Bennett, et al. "Performance Analytics: Econometric Tools for Performance and Risk Analysis." *R Packag Ver* 1(2014): 107.
28. https//CRANR-project.org/package=stabs.
29. https://CRAN.R-project.org/package=caret.
30. Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, et al. pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformat* 12(2011): 77.
31. Chaleckis, Romanas, Isabel Meister, Pei Zhang and Craig E. Challenges, Progress and Promises of Metabolite Annotation for LC–MS-Based Metabolomics. *Curr Opin Biotechnol* 55(2019): 44-50.
32. Camacho, Diogo, Alberto De La Fuente and Pedro Mendes. The Origin of Correlations in Metabolomics Data. *Metabolomics* 1(2005): 53-63.
33. Geidenstam, Nina, Yu-Han H Hsu, Christina M Astley, Josep M Mercader, et al. Using Metabolite Profiling to Construct and Validate a Metabolite Risk Score for Predicting Future Weight Gain. *PLoS One* 14(2019): e0222445.
34. Wewer, Albrechtsen Nicolai J, Anders E Junker, Mette Christensen and Sofie Hædersdal, et al. Hyperglucagonemia Correlates with Plasma Levels of Non-Branched-Chain Amino Acids in Patients with Liver Disease Independent of Type 2 Diabetes. *Am J Physiol Gastrointest Liver Physiol* 314(2018): G91-96.
35. Dyar, Kenneth A and Kristin L Eckel-Mahan. Circadian Metabolomics in Time and Space. *Front Neurosci* 11: 1-10.

How to cite this article: Klarskov, Carina Kirstine, Jesper F Havelund, Floor Dijkstra Zeger and Nils J Færgeman, et al. "Biomarkers for Development of Glucocorticoid-Induced Diabetes Mellitus – A Metabolomics-Based Prediction Model." *J Mol Genet Med* 14 (2020): 461 doi: 10.37421/jmgm.2020.14.461