Abstract. Type 2 diabetes mellitus (DM) is a progressive disease and the rate of progression from non-diabetes to DM varies considerably between individuals, ranging from a few months to many years. It is important to understand the mechanisms underlying the progression of diabetes. In the present study, a high-resolution metabolomics (HRM) analysis was performed to detect potential biomarkers and pathways regulating the mode of onset by comparing subjects who developed and did not develop type 2 DM at the second year in a 3-year prospective cohort study. Metabolic profiles correlated with progression to DM were examined. The subjects (n=98) were classified into four groups: Control (did not develop DM for 3 years), DM (diagnosed with DM at the start of the study), DM onset at the third year and DM onset at the second year. The focus was on the comparison of serum samples of the DM groups with onset at the second and third year from the first year, where these two groups had not developed DM, yet. Analyses involved sample examination using liquid chromatography-mass spectrometry-based HRM and multivariate statistical analysis of the data. Metabolic differences were identified across all analyses with the affected pathways involved in metabolism associated with steroid biosynthesis and bile acid biosynthesis. In the first year, higher levels of cholesterol (mass-to-charge ratio (m/z) 369.35, (M+H-H2O)+), 25-hydroxycholesterol [m/z 403.36, (M+H)+], 3α,7α-dihydroxy-5β-cholestane [m/z 443.33, (M+K)+], 4α-methylzymosterol-4-carboxylate [m/z 425.34, (M+H-H2O)+], and lower levels of 24,25-dihydrolanosterol [m/z 429.40, (M+H)+] were evident in the group with DM onset at the second year compared with those in the group with DM onset at the third year. These results, with a focus on the cholesterol biosynthesis pathway, point to important aspects in the development of DM and may aid in the development of more effective means of treatment and prevention.

Introduction

Type 2 diabetes mellitus (DM) is characterized by an increase in glucose levels or hyperglycemia due to pancreatic cell dysfunction and insulin resistance (1,2). In 2011, the prevalence of DM in Koreans aged 20-79 years was 7.5% (adjusted to world population). The prevalence is estimated to increase to 8.7% by 2030 (3). Since DM is a multifactorial disease influenced by genetic as well as environmental risk factors (4), the exact etiology of DM remains elusive (5,6). Furthermore, the rate of progression varies markedly between individuals (7) and little is known about the mode of onset. It is important to understand the mechanisms or biomarkers that are associated with the rapid progression of diabetes. Delaying the onset of DM may specifically improve therapies and understanding the characteristics of those who progress slowly may aid in the management of patients with DM (8).

Recent technologies, including high-resolution metabolomics (HRM), have enabled the discovery of potential biomarkers that may be beneficial for the diagnosis, management and treatment of various diseases (9-11). HRM generates comprehensive metabolic profiles by simultaneously measuring thousands of low-molecular weight metabolites in biological fluids, cells and tissues associated with certain diseases (12-14). The sensitivity and selectivity of HRM are best suited to the analysis of highly complex metabolite mixtures, such as biological extracts. HRM may also be used for the identification of potentially affected metabolites and pathways with the aid of the metabolite databases and in human metabolic pathway analyses, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (15).

Metabolic profiling allows for the exploration of different types of metabolites that may be affected by the progression or development of diabetes (16). Various metabolites have been correlated with insulin resistance and diabetes prediction in studies that mainly focused on a general comparison of control and case groups of DM during follow-up (17-21). A recent
study from India reported that 45.1% of subjects with normal glucose tolerance developed dysglycemia. Various predictors of progression included advancing age, family history of diabetes and cholesterol levels (22). Thus, when performing metabolomics analyses, differences in the onset of DM among subjects must be taken into account, and an inter-patient variability in the affected metabolites is likely. The present study used HRM to detect low-molecular weight metabolites by comparing samples collected from subjects who subsequently developed DM. The aim was to identify factors associated with the mode of onset of DM even prior to the development of symptoms.

Materials and methods

Materials. Liquid chromatography-mass spectrometry (LC-MS) grade water (Tedia, Fairfield, OH, USA), acetonitrile (Burdick & Jackson, Muskegon, MI, USA) and formic acid (Fuka, St. Louis, MO, USA) were used as the mobile phase. For quality control, three isotopes were used: Caffeine (3-methyl-13C), L-methionine (13C5, 15N) and N,N-diethyl-M-toluamide (Cambridge Isotope Laboratories, Inc., Tewksbury, MA, USA).

Human participants. This pilot study was reviewed and approved by the Korea University Institutional Review Board (Sejong City, Korea) and conformed to the board’s ethical guidelines (no. KU-IRB-15-19-A-1). Written informed consent was obtained from all participants. The subjects were from the Korean Cancer Prevention Study (KCPS-II Biobank), a pool of 159,844 participants who voluntarily underwent private health examinations and had provided informed consent in one of 11 centers located in Seoul and Gyeonggi provinces in South Korea from 2004 to 2013. Among all participants (age, 30-60 years) were selected. Those with missing data on essential or metabolic syndrome-associated variables were excluded.

The included participants with or without type 2 DM were classified into four groups based on fasting blood sugar levels, with serum samples collected in a span of 3 years. G1NN was the negative control group comprising non-diabetic (NDM) patients for the 3-year cohort study. G2DDD represented the positive control group comprising individuals who already had DM at the start of the study. G3NND and G4NDD represented patients who developed DM in the first year (baseline) and G4NDD represented patients who developed DM in the second year (Fig. 1A). Type 2 DM was defined as fasting blood sugar levels of >125 mg/dl (Fig. 1B). The subjects’ full demographics in the first year (baseline) are provided in Table I. Measurements of the body mass index (BMI) and fasting blood sugar were performed by utilizing COBAS INTEGRA 800 and 7600 Analyzers (Hitachi, Tokyo, Japan) (23). Data were analyzed for significance using SPSS® 24 statistical software (IBM Corp., Armonk, NY, USA).

LC-MS. The samples (50-µl aliquots) were diluted with 200 µl acetonitrile and centrifuged at 16,000 g for 5 min at 4°C to remove any protein (24). The samples were then randomized and analyzed using ultra performance LC (C18 Synchronis aQ, 1.9 µm, 100x2.1 mm; Thermo Fisher Scientific, Inc., Waltham, MA, USA) coupled with a quantitative time-of-flight MS using a model 6550 apparatus (Agilent Technologies, Santa Clara, CA, USA). The mobile phases were water and acetonitrile, and contained 0.1% formic acid. LC was run using the following gradient program: 95% water for 1 min, a linear decrease to 55% water over 8 min, a descending gradient to 10% water over 3 min, a 1.5-min hold and return to 95% water over 0.1 min. Detection of the mass/charge ratio (m/z) of ions set from 50 to 1,000 with a resolution of 20,000 over 15 min, LC runs with data extraction using the apLCMS algorithm provided a minimum of 6,000 reproducible features, with the mass accuracy being sufficient to allow for the prediction of the elemental composition in numerous instances.

Statistical analyses for metabolic profiling. After processing the data using apLCMS (25), all of the features of the samples were retrieved. The features from the LC-MS analyses were log2 transformed and quantile normalized prior to applying bioinformatics. Statistical analysis was performed on the extracted data using MetaboAnalyst 3.0 (http://www.metaboanalyst.ca/) for additional statistical tests, including principal component analysis (PCA), orthogonal signal correction/partial least squares-discriminant analysis (OPLS-DA), hierarchical cluster analysis (HCA) and Manhattan plot, to specifically differentiate between the profiles of all four categories in one analysis. Feature intensities were first log-transformed and auto-scaled prior to the performance of PCA and OPLS-DA with a confidence level of 95%. Statistical analyses, which included univariate analysis, Manhattan plot and false discovery rate adjusted P-value (FDR) (26), were performed to determine the metabolites that were significantly different between G1NND and G4NDD in the first year. The metabolic profiles were differentiated using Limma 2-HCA to distinguish between the two groups based on their metabolites (27).

Scheme of analysis. Utilizing samples from all 3 years, all groups were analyzed by multivariate analysis to discern metabolic differences. The present study focused on exploring the difference in metabolic profiles in the first year between G1NND and G4NDD, where the two groups were initially NDM but subsequently developed DM in the prospective cohort. Thus, a Manhattan plot and HCA were used to examine G3NND and G4NDD in the first year, looking for potential discriminating metabolites concerning diabetes progression. In addition, it was examined whether the metabolic changes were similar in the G1NND and G4NDD groups in the second and third year compared with those in the first year.

Annotation using METLIN and KEGG metabolic pathway analysis. An m/z feature is defined by m/z, ion intensity and retention time. The m/z values were annotated using the METLIN Mass Spectrometry Database (https://metlin.scripps.edu) to identify the metabolites (28). Annotated features from the METLIN database are mapped on human metabolic pathways using the KEGG database. The human metabolic pathways of the KEGG database (http://www.genome.jp/kegg/tool/map_pathway2.html) were used for mapping the significant features from the Manhattan plot with an FDR of q=0.05. The matched features were displayed as black dots in the pathway maps to determine which pathways were affected by each case condition.
**Results**

**Statistical and group comparison.** The four group categories were first analyzed using the unsupervised, multivariate statistical PCA procedure to determine whether any metabolic differences were present. PCA did not provide a good separation of the four groups (Fig. 2A). HCA exhibited a tendency to separate subjects into four groups based on auto-scaled and log-transformed features (Fig. 2B). In addition, supervised OPLS-DA was then used to analyze the data. As displayed in Fig. 2C, a score plot was generated to illustrate the metabolic differences throughout the four stages of DM acquisition. This OPLS-DA model demonstrated a clear separation of groups with an R² value of goodness of fit of 0.956 and a Q² value of predictive ability of the model of 0.835. A clear separation was expected to be evident between G1NNN, G2DDD, G3NND and G4NDD, indicating that groups could be distinguished based on certain serum metabolites in their metabolic profiles. Further analyses focused on the comparison between G3NND and G4NDD in order to observe the effect of a different onset of DM on the metabolic profile.

A good separation between G1NND and G4NDD was clearly observed for each year on the OPLS-DA score plot (Fig. 3A), indicating that different onset times of DM caused metabolic alterations. A subsequent examination focused only on the first year, where the two groups had not developed DM, yet. A Manhattan plot and HCA were used to detect significant metabolites based on which the two groups could be distinguished. The comparison was significantly different with an FDR of q=0.05. The FDR is a multiple testing correction applied in a statistical comparison to decrease the occurrences of false positives. The plot provides a visual presentation of the statistical test result with the application of multiple testing corrections. The x-axis displays the m/z values with a range of 50-1,000, while the y-axis displays the -logP values. The FDR criteria were indicated by the dashed lines. Features above the line (green dots) are considered to be significant between each of the two-way comparisons (samples and metabolites in HCA). Among a total of 3,462 features, the two-way comparison between the first year of G1NND and G4NDD revealed 183 to be statistically significant (Fig. 3B).
A two-way HCA of the comparisons' significant features indicated a clear separation of one group from the other, as displayed by the upper bar (Fig. 3C). The top label indicates the separation distance among samples and the left-hand labels indicate the clustering among metabolites, which statistically differentiated between the two groups.

The significant features were annotated using the METLIN database and mapped on the KEGG human metabolic pathways. Upon mapping the 183 significant features, 34 were matched to human metabolic pathways. Among the possible affected pathways, the two of greatest interest were the primary bile acid biosynthesis and steroid biosynthesis metabolism pathways (data not shown). Additional analyses for the second and third year of G3NND and G4NDD were performed to observe the profiles of the significant metabolites. Among the 183 significant features in the first year, 80 features remained significant in the second year and 90 features were significant in the third year.

Identification of potential biomarkers to distinguish between G3NND and G4NDD. Endogenous compounds were identified and annotated from the affected pathways (Table II). From the primary bile acid biosynthesis pathway, cholesterol, 25-hydroxycholesterol and 3α,7α-dihydroxy-5β-cholestane in the G3NND-first year group were significantly different (P<0.05) from those of the G4NND-first year group, even though patients in the two groups were NDM at the time. Similarly, in the steroid biosynthesis pathway, cholesterol, 24,25-dihydrolanosterol and 4α-methylzymosterol-4-carboxylate were significantly altered in the G3NND-first year group compared with those in the G4NND-first year group, even though patients in the two groups were NDM at the time. Similarly, in the steroid biosynthesis pathway, cholesterol, 24,25-dihydrolanosterol and 4α-methylzymosterol-4-carboxylate were significantly altered in the G3NND-first year group compared with those in the G4NND-first year group prior to development of DM (P<0.05) (Figs. 4 and 5). These results indicated that only cholesterol [m/z 369.35, (M+H-H2O)+] affected the two pathways. However, when NDM progressed to DM, cholesterol and 4α-methylzymosterol-4-carboxylate [m/z 425.34, (M+H-H2O)+] were not significantly different in the second and third year (P<0.05). In the second year, 25-hydroxycholesterol [m/z 403.36, (M+H)+] and 3α,7α-dihydroxy-5β-cholestane [m/z 443.33, (M+K)+] were still significantly altered (P<0.05). 24,25-Dihydrolanosterol [m/z 429.40, (M+H)+] was significantly decreased (P<0.05) throughout the 3-year time course (first, second and third year) (Figs. 6 and 7).
Table II. Significant compounds associated with the two affected pathways.

<table>
<thead>
<tr>
<th>Pathway/compound</th>
<th>m/z</th>
<th>Adduct</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary bile acid biosynthesis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholesterol</td>
<td>369.3507</td>
<td>(M+H-H2O)+</td>
<td>2.62x10^-4</td>
</tr>
<tr>
<td>25-Hydroxycholesterol</td>
<td>403.3593</td>
<td>(M+H)+</td>
<td>4.40x10^-5</td>
</tr>
<tr>
<td>3α,7α-Dihydroxy-5β-cholestane</td>
<td>443.3325</td>
<td>(M+K)+</td>
<td>1.77x10^-3</td>
</tr>
<tr>
<td><strong>Steroid biosynthesis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholesterol</td>
<td>369.3507</td>
<td>(M+H-H2O)+</td>
<td>2.62x10^-4</td>
</tr>
<tr>
<td>24,25-Dihydrolanosterol</td>
<td>429.4042</td>
<td>(M+H)+</td>
<td>2.57x10^-3</td>
</tr>
<tr>
<td>4α-Methylzymosterol-4-carboxylate</td>
<td>425.3428</td>
<td>(M+H-H2O)+</td>
<td>6.08x10^-4</td>
</tr>
</tbody>
</table>

m/z, mass-to-charge ratio.

Figure 4. Steroid biosynthesis and primary bile acid biosynthesis pathway identified as key pathways affecting the progression to DM from a non-diabetic status. Green boxes indicate the 24-dehydrocholesterol reductase enzyme, regulated by 24-dehydrocholesterol reductase. Yellow boxes indicate the increase in concentration. The blue box indicates a decrease in concentration. G4NDD, patients that progressed to DM at the second year. DM, diabetes mellitus.

Figure 5. Significant metabolites altered in the first year between G3NND and G4NDD. *P<0.05. G3NND, patients that progressed to DM at the third year; G4NND, patients that progressed to DM at the second year. DM, diabetes mellitus; m/z, mass-to-charge ratio.
Validation of cholesterol was performed by selected reaction monitoring by detecting specific precursor-product ion transition using tandem MS mode and comparing the fragmentation pattern to its chemical standard, when available. Cholesterol ion fragmentation was detected: m/z 369.35 → m/z 147.11, m/z 161.13 and m/z 95.08 (Fig. 8).

Discussion

The present study explored metabolic profiling by comparing the mode of onset of type 2 DM in a 3-year prospective cohort study. The results indicated the dysregulation of certain pathways and metabolites prior to the development of DM. The baseline value of fasting blood sugar in the G2NND-first year and G4NDD-first year groups indicated that the values in subjects who had not yet developed DM were significantly different from those who had diabetes in the G2DDD group. In addition, five m/z values, which were mainly assigned to sterol compounds, exhibited statistical significance regarding the mode of onset (P<0.05).

The results indicate that the primary bile acid biosynthesis pathway was affected by the time of onset of DM (1 or 2 years prior to DM). Certain metabolites of the primary bile acid pathway tended to be higher in the G4NDD group-first year compared with those in the G2NND group-first year, such that 25-hydroxycholesterol and 3α,7α-dihydroxy-5β-cholestanate...
were elevated. This was in agreement with other studies reporting that bile acid synthesis was stimulated in association with DM, obesity and insulin signaling (29-31). The intensity of cholesterol, which is the precursor of bile acid biosynthesis, was elevated.

Cholesterol is a fundamental raw material for the cell. It acts as the building block for the cell membrane and other organelles (32). It serves as a precursor of numerous biosynthetic processes in human metabolic pathways, including steroid hormone and bile acid biosynthesis (33). Fig. 4 (adapted from the KEGG database) depicts the central function of cholesterol as a bridge connecting the two pathways.

In the present study, cholesterol was significantly higher in the G4NDD group-first year, in comparison with that in the G3NND group-first year (P<0.05). This result indicated that cholesterol biosynthesis is modulated by DM progression. Fig. 4 presents part of the steroid biosynthesis pathway involving the generation of cholesterol, which stimulates primary bile acid synthesis in the G4NDD group-first year. In the present study, cholesterol production in the G4NDD group tended to be higher than that in the G3NND group, and the metabolite 4α-methylzymosterol-4-carboxylate was also elevated in the first year. In addition, 24,25-dihydrolanosterol was lower, explaining for the boost in cholesterol production. Of note, differences in this metabolite were significant in the first, second and third years (P<0.05).

This upregulated downstream effect of cholesterol production may be due to the activity of one enzyme, δ24-sterol reductase (enzyme ID 1.3.1.72, also known as 24-dehydrocholesterol reductase). This enzyme participates in numerous processes and is regulated by the DHCR24 gene (34,35). This gene may be linked to diabetes progression. Berisha et al (36) indicated that DHCR24 is one of the transcripts affected by bariatric surgery in obese DM patients, with enzyme activity being correlated with changes of body weight, fasting plasma glucose and glycosylated hemoglobin content. Although focusing on endometrial carcinoma, another study revealed that the enzyme encoded by DHCR24 was induced by insulin stimulation via signal transducer and activator of transcription 3, which sensitizes to insulin signaling (37,38). This mechanism is possibly associated with the pathology of diabetes due to its link with hyperinsulinemia (39,40). Thus, these results indicated that the dysregulation of the metabolism, at the genomics to metabolomics level, had started to develop one year prior to the onset of DM. Cholesterol detected in the present LC-MS analysis was different from that determined

Figure 8. Identification and validation of cholesterol. (A) EIC of standard, serum and spike sample. (B) The spike peak of cholesterol was increased at 13 min. The cholesterol fragmentation was observed from standard and sample with collision energy of 20 and 10 V. The red arrows represent the fragmented cholesterol from the standard and sample. ESI, electron spray ionization; CID, collision-induced dissociation; EIC, extracted ion chromatogram; rt, retention time; frag, fragmentor voltage.
by routine biochemical analysis, since the latter was the total cholesterol value comprising triglyceride, low-density lipoprotein (LDL) and high-density lipoprotein (HDL) values. In addition, the adduct of cholesterol, (M+H-H2O)+, had an m/z of 369.35 in the present study, while lipoproteins of LDL and HDL containing cholesterol have different m/z values (41).

The present study aimed to utilize HRM to further assess the mode of onset of DM. Primary bile acid biosynthesis and steroid biosynthesis metabolism, which focus on cholesterol biosynthesis, have important roles in different modes prior to the onset of DM. A future study on a larger population should be performed to validate the clinical value of the present results.

Acknowledgements

The authors would like to thank Dr Karan Uppal from Emory University School of Medicine (Atlanta, GA, USA) for providing the R-package to run the Metabolomics Wide Association Study. ADP gratefully acknowledges the Indonesia Endowment Fund for Education (grant no. A-3468/LPDP.3/2015) for the financial support of his master degree program. This study was supported by the Korea Health Industry Development Institute (grant no. HI14C2686) and the National Foundation of Korea (grant nos. NRF-2017R1A2B4003890 and NRF-2017M3A9F1031229).

References