**Study population**

The cross-sectional case-control study included participants with screen-detected T2D with or without DPN and healthy controls. The cohort is a sub-population of participants that attended a clinical follow-up examination between October 2015 and June 2016 at the Aarhus study-site (n=97) as part of the Danish arm of the Anglo-Danish-Dutch study of Intensive Treatment in People with Screen-Detected Diabetes in Primary Care (ADDITION). Briefly, ADDITION-Denmark enrolled participants (aged 40-69 years) with screen-detected, previously undiagnosed T2D by stepwise screening in primary care between 2001 and 2006. Following screening, general practices were randomized to deliver either routine care for T2D patients or intensive multifactorial target-driven care for T2D patients until the trial concluded in 2009 (*1, 2*).

The cohort T2D participants (n=97) in the current study underwent follow-up examination after a mean of 13 years after enrolling in ADDITION-Denmark (*3*). Participants were assessed for DPN using nerve conduction studies (NCS), which identified 48 with DPN and 49 without DPN. Prior to examination, participants fasted from midnight the previous day until their examination. Anthropometric measurements, DPN assessment, and blood sample collection were taken for each participant. This study additionally includes data from a healthy control group of gender- and age-matched study-subjects without T2D, from the same geographical area and derived from the original screening cohort of the ADDITION study (n=9). Healthy control participants were also evaluated at the time of examination to confirm that they still had no diabetes and no other neurological diseases.

**DPN definition**

DPN was defined using NCS criteria outlined by Dyck et al. (*i.e*., criteria number 8 of sum scores of Z-scores of six nerve parameters, as described below) (*4, 5*). NCS were performed using Keypoint.Net EMG equipment (Dantec, Skovlunde, Denmark). The Z-scores of in-house age- and height-matched normative data were used to calculate sum scores from the average of the following six parameters: the conduction velocity of the peroneal, tibial, and median (or ulnar) nerves, minimum F-wave latencies of the tibial and median (or ulnar) nerves, and sensory nerve action potential amplitude of the sural nerve. Ulnar nerves were assessed instead of the median nerve when electrophysiological signs of carpal tunnel syndrome were present. NCS sum Z-scores >2.0 were considered abnormal and indicating DPN, whereas sum Z-scores ≤2.0 indicated no signs of DPN.

**Plasma processing**

At the mean 13 year follow up after T2D diagnosis, blood samples were collected from patients the same day as anthropometrics and DPN assessment. Plasma was collected in purple EDTA tubes with 10 mL of buffy coat, inverted 10 times, incubated for 30-90 minutes at room temperature, and centrifuged at 3000 rpm for 10 minutes. The plasma supernatant was collected by aspirating plasma to approximately 5 mm above the buffy coat and plasma samples were stored in 0.5 mL aliquots at -80°C prior to metabolomics analysis.

**Co-variate assessment**

For each participant, clinical examination and biochemical measures were performed on the same day, including assessments of anthropometrics, blood pressure, and metabolic measures from blood (total cholesterol, HDL cholesterol, triglycerides, HbA1c) and urine samples (creatine). The general practitioners provided prescribed medication records (statins, glucose-lowering drugs, antihypertensives, beta-blockers, and aspirin).

## Metabolon Platform

**Sample Accessioning:** Following receipt, samples were inventoried and immediately stored at -80oC. Each sample received was accessioned into the Metabolon LIMS system and was assigned by the LIMS a unique identifier that was associated with the original source identifier only. This identifier was used to track all sample handling, tasks, results, etc. The samples (and all derived aliquots) were tracked by the LIMS system. All portions of any sample were automatically assigned their own unique identifiers by the LIMS when a new task was created; the relationship of these samples was also tracked. All samples were maintained at -80oC until processed.

**Sample Preparation:**Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

**QA/QC:** Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Tables 1 and 2 describe these QC samples and standards. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections, as outlined in Figure 1.

**Table 1:** **Description of Metabolon QC Samples**

|  |  |  |
| --- | --- | --- |
| **Type** | **Description** | **Purpose** |
| MTRX | Large pool of human plasma maintained by Metabolon that has been characterized extensively. | Assure that all aspects of the Metabolon process are operating within specifications. |
| CMTRX | Pool created by taking a small aliquot from every customer sample. | Assess the effect of a non-plasma matrix on the Metabolon process and distinguish biological variability from process variability. |
| PRCS | Aliquot of ultra-pure water | Process Blank used to assess the contribution to compound signals from the process. |
| SOLV | Aliquot of solvents used in extraction. | Solvent Blank used to segregate contamination sources in the extraction. |

**Table 2:** **Metabolon QC Standards**

|  |  |  |
| --- | --- | --- |
| **Type** | **Description** | **Purpose** |
| RS | Recovery Standard | Assess variability and verify performance of extraction and instrumentation. |
| IS | Internal Standard | Assess variability and performance of instrument. |

****

**Figure 1.**  **Preparation of client-specific technical replicates.** A small aliquot of each client sample (colored cylinders) is pooled to create a CMTRX technical replicate sample (multi-colored cylinder), which is then injected periodically throughout the platform run. Variability among consistently detected biochemicals can be used to calculate an estimate of overall process and platform variability.

**Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS):** All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same afore mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MSn scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

**Bioinformatics:** The informatics system consisted of four major components, the Laboratory Information Management System (LIMS), the data extraction and peak-identification software, data processing tools for QC and compound identification, and a collection of information interpretation and visualization tools for use by data analysts. The hardware and software foundations for these informatics components were the LAN backbone, and a database server running Oracle 10.2.0.1 Enterprise Edition.

**LIMS:**  The purpose of the Metabolon LIMS system was to enable fully auditable laboratory automation through a secure, easy to use, and highly specialized system. The scope of the Metabolon LIMS system encompasses sample accessioning, sample preparation and instrumental analysis and reporting and advanced data analysis. All of the subsequent software systems are grounded in the LIMS data structures. It has been modified to leverage and interface with the in-house information extraction and data visualization systems, as well as third party instrumentation and data analysis software.

**Data Extraction and Compound Identification:** Raw data was extracted, peak-identified and QC processed using Metabolon’s hardware and software. These systems are built on a web-service platform utilizing Microsoft’s .NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (*m/z)*, and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

**Curation:** A variety of curation procedures were carried out to ensure that a high quality data set was made available for statistical analysis and data interpretation. The QC and curation processes were designed to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artifacts, mis-assignments, and background noise. Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

1. S. J. Griffin *et al.*, Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): a cluster-randomised trial. *Lancet (London, England)* **378**, 156-167 (2011).

2. A. Sandbaek *et al.*, Effect of early multifactorial therapy compared with routine care on microvascular outcomes at 5 years in people with screen-detected diabetes: a randomized controlled trial: the ADDITION-Europe Study. *Diabetes care* **37**, 2015-2023 (2014).

3. T. Lauritzen *et al.*, The ADDITION study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with Type 2 diabetes detected by screening. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* **24 Suppl 3**, S6-11 (2000).

4. P. J. Dyck, R. E. Carter, W. J. Litchy, Modeling nerve conduction criteria for diagnosis of diabetic polyneuropathy. *Muscle & nerve* **44**, 340-345 (2011).

5. A. M. Kural *et al.*, The Utility of a Point-of-Care Sural Nerve Conduction Device for Detection of Diabetic Polyneuropathy– A cross-sectional study. (2018).