nature portfolio

Corresponding author(s): Prof. Karsten Suhre & Dr. Anna Halama

Last updated by author(s): 2024/06/5

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

This study consists of data collected across multiple platforms and multiple softwares were used (see method section). Briefly:
1) The obtained blood samples were analyzed within four hours of blood collection at the Department of Laboratory Medicine and Pathology of HMC with the Cobas® 6000 (Roche Diagnostics, Basel, Switzerland)
2) The integrity and quantity of the isolated RNA was measured using Qubit RNA HS Assay Kit (high sensitivity, 5 to 100 ng quantification range) Assay Kit and Qubit 3.0 fluorometer (Life Technologies) according to the manufacturer's protocol. The total RNA was depleted of rRNA and Globin using the NEBNext rRNA & Globin Depletion Kit for Human/Mouse/Rat (New England BioLabs, Ipswich, MA). The depleted RNA was used to generate strand-specific libraries with BIOO NEXTflex Rapid Directional RNA-Seq Kit (Bioo-Scientific, Austin, TX). Library quality and quantity were analyzed with the Bioanalyzer 2100 (Agilent, Santa Clara, CA) on a High Sensitivity DNA chip. 10 libraries were then pooled in equimolar ratios and paired-end sequenced at 75bp on one lane of an Illumina HiSeq 4000 (Illumina, San Diego, CA).
3) The PCR data were processed using Exiqon GenEx qPCR analysis software (version 6). The inter-plate calibration was performed using the mean value of UniSp3 interplate calibrator. The samples with a high degree of hemolysis were identified after monitoring of calculated ΔCt between hsa-miR-23a-3p and hsa-miR-451a. The samples with ΔCt >7 were removed from the analysis. Only microRNA assays with Ct ≤ 35, expressed in at least 60% of the samples were counted and the remaining samples were removed from the analysis.
4) The method employed protein-capture by Slow Offrate Modified Aptamers (SOMAmer). The primary data were submitted to Somalogic for normalization of raw intensities, across-batch calibration and steps of quality control.
5) Proteomics measurements based on the Olink® technology (Olink Proteomics AB, Uppsala, Sweden) at the WCM-Q Proteomics Core. We used two different Olink® panels, namely Cardiometabolic and Metabolism, for measurements of 184 unique proteins. Supplier-provided optimized antibody pairs labeled individually with oligonucletotides (PEA probes) were used.
6) Samples were analyzed by Genos Ltd. (Zagreb, Croatia) using ultra-performance liquid chromatography (UPLC) glycoprofiling. The fluorescence was measured at 420 nm with excitation at 330 nm using Waters Acquity UPLC H-class system consisting of a fluorescence (FLR) detector set with 250 nm excitation and 428 nm emission wavelengths.

The data processing was performed using an automatic processing method enabling to obtain chromatograms separated into 39 peaks. The fluorescence was measured at 420 nm with excitation at 330 nm using Waters Acquity UPLC H-class system consisting of a fluorescence (FLR) detector set with 250 nm excitation and 428 nm emission wavelengths. The data processing was performed using an automatic processing method enabling to obtain chromatograms separated into 39 peaks.
7) The IgG isolation and measurements were conducted by Genos. For the separation and measurements nanoACQUITY UPLC system (Waters, Milford Massachusetts, USA), consisting of binary pump, auxillary pump, autosampler maintained at 10 °C and column oven compartment set at 30 °C coupled to and the Bruker Compact Q-TOF-MS were used. Spectra were recorded from m/z 600 to 1900 with 2 averaged scans at a frequency of 0.5 Hz. Per sample the total analysis time was 15 min. The nanoACQUITY UPLC system and the Bruker Compact Q-TOF-MS were operated under HyStar software version 3.2.
8) The purification, separation and measurements of IgA and IgG was conducted at Leiden University Medical Center. The sample separation and measurements were conducted on Ultimate 3000 RSLCnano system (Dionex/Thermo Scientific) equipped with an Acclaim PepMap 100 trap column (particle size 5 µm, pore size 100 Å, 100 µm × 20 mm) and an Acclaim PepMap C18 nano analytical column (particle size 2 µm, pore size 100 Å, 75 µm × 150 mm) coupled to a quadrupole-TOF-MS (Impact HD; Bruker Daltonics).
9) The measurements were conducted at Metabolon Inc, deploying HD2 platform based on ultra-high-performance liquid chromatography-mass spectrometry (UPLC-MS) and gas chromatography-mass spectrometry (GC-MS) technology. The separation and measurements of the sample aliquots designated for LC-MS were performed on Waters ACQUITY UPLC in-line to Thermo Scientific Q-Exactive high resolution/ accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and an Orbitrap mass analyzer. In the MS analysis, the scan range varied between methods but fell within the range of 70-1000 m/z. The GC- obtained samples were performed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer using electron impact ionization (EI), and the MS scan range was from 50-750 m/z.
10) The Esamples were used to conduct metabolic profiling at Metabolon Inc on technologically advanced, in comparison with HD2, HD4 platform enabling for increased sensitivity and accurate detection of more metabolites. The main technical difference between HD2 and HD4 platforms was replacement of GC-MS with hydrophilic interaction chromatography (HILIC) method.
11) The samples were measured at the Metabolomics Platform of the Helmholtz Center Munich using AbsoluteIDQTM kit p150 (Biocrates Life Science AG, Innsbruck, Austria) Metabolites were measured in positive and negative multiple reaction monitoring (MRM) scan mode by direct infusion to an API 4000 triple quadrupole system (SCIEX Deutschland GmbH, Darmstadt, Germany) equipped with a 1200 Series HPLC (Agilent Technologies Deutschland GmbH, Böblingen, Germany) and a HTC PAL auto sampler (CTC Analytics, Zwingen, Switzerland) controlled by the software Analyst 1.6.2. The metabolite concentrations were calculated using internal standards and the MetIDQ software provided with AbsoluteIDQTM kit, and are reported in µmol/L. For the lipid molecules including PC, lysoPC, SM, and AC, measured with AbsoluteIDQTM kit the information on the sum of the carbons of the fatty acid chains is provided but not the fatty acid chain actual composition.
12) The samples were used for in depth profiling of lipids, which was conducted at Metabolon Inc. deploying LipidyzerTM platform of AB Sciex Pte technology. The samples were analyzed in both positive and negative mode electrospray using Sciex SelexIon-5500 QTRAP. . Lipid class concentrations were calculated from the sum of all molecular species within a class, and fatty acid compositions were determined by calculating the proportion of each class comprised by individual fatty acids.
13) 1H-NMR spectra analysis of urine samples was conducted at Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Germany. Bruker DRX-400 NMR spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) operating at 400.13 MHz 1H frequency equipped with 4 mm selective inverse flow probe (FISEI, 120 µL active volume) was used to record the spectra. The obtained spectra were processed within TOPSPIN 1.3 (Bruker BioSpin GmbH) and the metabolites annotation and quantification was conducted in semi-automated manner by spectral pattern matching using Chnomx NMR suit 7.0 (Chenomx Inc.).
14) The metabolite quantification was conducted by a high-throughput NMR metabolomics platform (Nightingale Ltd, Helsinki, Finland). The measurements were conducted on Bruker AVANCE III 500 MHz and Bruker AVANCE III HD 600 MHz spectrometers.The extracted lipid (LIPID) data was evaluated in full automation with the 600 MHz instrument.
15) DNA methylation profiling was performed using the Illumina Infinium HumanMethylation450 (450K) BeadChip array. Genome Studio software integrated controls dashboard was used to conduct quality control. The arrays were scanned with the Illumina iScan system. Genome Studio (version 2011.1) with methylation module (version 1.9.0) was used to process the raw image data.
16) Illumina Omni 2.5 array (version 8) was used

Data analysis

Multiple statistical analysis were conducted for this study. Exact description is provided in the method section. The statistical strategies used in this study are stored along with the source code in github (https://github.com/karstensuhre/comics). All statistical analyses were conducted using R (version 4.1.0 and above) and Rstudio (version 1.4.1717 and above). If not otherwise stated, the omics data was converted "as-received" into R Summarized Experiment format, representing processed final data. The saliva metabolomics data has been further normalized by saliva osmolality, and the urine metabolomics data has been normalized by urine creatinine.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The RNA, miRNA, proteomics (measured by OLINK and SOMA), glycomics (measuring total N-glycans, IgA and IgG), lipoproteomics, lipidomics (depicted by Lipidyzer™at Metabolon platform and Biocrates), and metabolomics (conducted in plasma, saliva and urine) data generated in this study have been deposited in the Figshare database under link [https://doi.org/10.6084/m9.figshare.25975627.v2]. The DNA and methylation data are available under restricted access because the informed consent given by the study participants does not cover posting of participant genotype and methylation data in public databases, access can be obtained by researcher affiliated with a research institution on an individual basis from the corresponding author (Karsten Suhre or Anna Halama Weill Cornell Medicine – Qatar, Doha, Qatar). Access is subject to approval by the institutional research board of Weill Cornell Medicine – Qata. The raw spectra data from metabolomics and glycomics are protected and are not available due to data privacy laws. The processed glycomics (measuring total N-glycans, IgA and IgG), lipoproteomics, lipidomics (depicted by Lipidyzer™at Metabolon platform and Biocrates), and metabolomics (conducted in plasma, saliva and urine) data are available at Figshare

database under link: https://doi.org/10.6084/m9.figshare.25975627.v2. The RNA, miRNA, proteomics (measured by OLINK and SOMA), glycomics (measuring total N-glycans, IgA and IgG), lipoproteomics, lipidomics (depicted by Lipidyzer™at Metabolon platform and Biocrates), and metabolomics (conducted in plasma, saliva and urine) data generated in this study are provided in the Supplementary Information/Source Data file.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Information regarding sex was obtained through questionnaires. We refer to sex as currently understood biological differences between females and males.
Reporting on race, ethnicity, or other socially relevant groupings	Information regarding ethnicity was obtained through questionnaires. Based on the country of birth of participants parents and grandparents, subjects were divided into three major ethnicity groups: Arabs, South Asians, and Filipinos.
Population characteristics	The cohort consists of 193 females and 198 males. The average participants age was 46.5 years (s.d. = 12.9) and the average BMI was 29.7 kg/m2 (s.d. = 6.0). This cohort includes 195 participants with T2D and 196 without T2D.
Recruitment	The subjects were enrolled in the framework of the Qatar Metabolomics Study on Diabetes (QMDiab), a cross-sectional diabetes case-control study at the Dermatology Department of HMC in Doha, Qatar.
Ethics oversight	The study was approved by the Institutional Review Boards of HMC and Weill Cornell Medicine, Qatar (WCM-Q) (research protocol #11131/11).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This cohort consists of 391 participants and is the first and the largest cohort characterized by 18 technologically different omics platforms enabling deep multiomics phenotyping. We initially defined this cohort size to conduct case-control metabolomics study on type 2 diabetes, based on which we recruited participants. With the technological development we were able to conduct the multiomics measurements using 18 different platforms. Here we report on 34,000 statistically significant (after Bonferroni correction) trait-trait links in biofluids in this cohort.
Data exclusions	Inclusion criteria were a primary form of T2D (cases) and absence of major systemic disorders (controls). Data from five participants were excluded from the analysis due to incomplete records.
Replication	Our study was conducted across 18 technologically different platforms which resulted in determination of 6,304 quantitative molecular traits with 1,221,345 genetic variants, methylation at 470,837 DNA CpG sites, and gene expression of 57,000 transcripts in 391 subjects, which makes it not feasible to replicate. Nevertheless, our cohort was used for replication of multiple associations conducted on the single omics layers.
Randomization	Due to the study nature randomization by participant enrollment was not required. However, it is important to mention that we randomized sample distribution for each measurement.
Blinding	This is cross-sectional case-control study for which blinding to group allocation is not a relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging