Systematic Review Protocol

PROSPERO ID: CRD42025648346

Version 2 – 28/05/2025

Index

38
39

Title

Exploring Biomarkers in Type 2 Diabetes Mellitus versus Normoglycemia Identified through High-throughput Proteomics: A Systematic Review and Meta-Analysis Protocol

Contact information

- **Julia García Currás**
  - Email: julia.gcurras@udc.es; julia.garcia@biostatech.com
  - Affiliation: (1) Biostatech, Advice, Training & Innovation in Biostatistics; (2) CITIC, Computer Architecture Group (Universidade da Coruña).
  - Address: (1) Rúa das Hedras, 6, 2ºH 15895 Ames, A Coruña, Spain; (2) Rúa Maestranza 9, 15001 A Coruña, A Coruña, Spain.
  - Phone Number: (+34) 672 771 319
- **Raquel Pérez Lois**
  - Email: loisperezraquel@gmail.com
  - Affiliation: Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Universidade de Santiago de Compostela (USC); CIBER Fisiopatología de la Obesidad y Nutrición (CIBERobn).
  - Address: Complexo Hospitalario Universitario de Santiago de Compostela. Travesía da Choupana s/n, Edificio Consultas Externas. Planta -2. Santiago de Compostela 15706; Av Monforte de Lemos 3-5, 28029 Madrid, Spain
  - Phone Number: (+34) 981955708
- **María P. Pata**
  - Email: mariapata6@biostatech.com
  - Affiliation: Biostatech, Advice, Training & Innovation in Biostatistics.
  - Address: Rúa das Hedras, 6, 2ºH 15895 Ames, A Coruña, Spain.
  - Phone Number: (+34) 672 771 319
- **Guillermo López Taboada**:
  - Email: guillermo.lopez.taboada@udc.es
  - Affiliation: Computer Architecture Group, CITIC - Universidade da Coruña (UDC)
  - Address: Campus de Elviña, 15071 A Coruña, Spain

Amendment of protocol, version 1

The following document represents an update of the original protocol published in PROSPERO on February 4th, 2025. All changes from protocol version 1 are listed below:

1. Addition of an author who contributed to the discusion of the review (Guillermo López Taboada).
2. Reduction in the sources of information.
3. Refinement of exclusion criteria: additional proteomics techniques and book chapters were added to the exclusions.
4. Refinement of diagnostic criteria for defining the T2D case group: American Diabetes Association (ADA) of Worl Health Organization (WHO) was followed, based on fasting blood glucose (%), HbA1c (%) or 2 hours oral glucose tolerance test (OGTT).
5. Changes to the final search strategy: only one author conducted the main search, and a subset of potentially relevant literature was peer reviewed.
6. Update of the risk of bias (quality) assessment section: the Newcastle–Ottawa Scale was deemed inappropriate for the included studies. A modified version of the QUADOMICS tool was used instead. The assessment was carried out by a single reviewer.
7. Inclusion of an omics-based meta-analytic approach in addition to the random-effects meta-analysis.
8. Exclusive inclusion of the final R packages used during the review, which were modified throughout the protocol.

Funding sources/sponsors

This work has been funded by a predoctoral grant to Julia García Currás (Ref. 23_IN606D_2022_2707220, GAIN, Xunta de Galicia, 2022-2026).

Conflicts of Interest

Julia García Currás is pursuing an industrial PhD at the company Biostatech, Advice, Training & Innovation in Biostatistics. María P. Pata works in Biostatech, Advice, Training & Innovation in Biostatistics.

Anticipated start date

Provisional: 10/02/2024.

Anticipated completion date

4 months after starting the systematic review (10/06/2024).

Review question

Which proteins exhibit differential expression between individuals with type 2 diabetes and normoglycemic controls, as identified through high-throughput proteomic analysis?

In these questions we can see the PICO(S) elements:

- Patients: Type 2 diabetic individuals.
- Intervention: No intervention.
- Comparison: Normoglycemic individuals (control group).
- Outcome: Differential proteins or biomarkers and the magnitude of the difference.
- Studies: Observational studies, including cohort studies and case-control studies.

Type of study: diagnostic accuracy studies.

We will conduct this systematic review in accordance with PRISMA guidelines.

Searches

Sources of information: PubMed central (PMC), Scopus, WOS (Web of Science).

Search dates: 10/02/2024 – 14/02/2024

Restrictions on the search:

- Language: English
- Type of article: excluding reviews and book chapters.
- Period of publication: from 2000 to 2025. DIA and DDA strategies in mass spectrometry were implemented and optimally used at the beginning of the current century.
- Searches will be rerun prior to the final analysis.

- Unpublished studies will not be considered in this systematic review.

Inclusion/exclusion criteria for search results review:

Exclusion criteria *a priori*:

1. Languages other than English.
2. Interventional studies.
3. Reviews without original data and book chapters
4. Non-human studies.
5. Peer-reviewed scientific papers without a list of differential proteins.
6. Targeted proteomics (SRM, MRM); techniques for quantifying individual protein samples (ELISA, western blot); targeted techniques with quantification based on DNA (SOMAscan) or qPCR (Olink); quantification using gels and staining/fluorescent tags (2D-DiGE).
7. Diabetes types other than type 2: gestational, type 1.

Inclusion criteria *a priori*:

1. High-throughput proteomic analysis of proteome data: DDA (shotgun) and DIA techniques.
2. Case control, cohort, and observational studies with original data.
3. Studies with a control group of normoglycemic individuals (adults aged ≥ 18 years).
4. Human studies.

Condition or domain being studied

Type 2 diabetes.

Participants/population

Case group:

- Inclusion: Adults (≥ 18 years old) of any gender diagnosed with type 2 diabetes based on ADA or WHO criteria, which include the assessment of FPG percentage, HbA1c percentage and a 2h OGTT test.
- Exclusion: Gestational or type 1 diabetes, presence of known comorbidities (mainly cancer).

Intervention(s), exposure(s)

(Mandatory field): For reviews of qualitative studies give details of the focus of the review.

This systematic review will investigate proteins associated with type 2 diabetes identified through high-throughput proteomic technologies. The exposure of interest is type 2 diabetes, with differential protein expression being assessed in individuals with type 2 diabetes compared to normoglycemic controls. The review will focus on studies that perform quantitative proteomic analysis, including DDA and DIA techniques, followed by downstream analysis of protein quantification data.

Comparator(s)/control

Control group:

- Inclusion: Adults (≥18 years old) of any gender with a normal status of glycated haemoglobin, determined following the ADA or WHO criteria.
- Exclusion: Presence of any known disease and individuals under 18 years old.

Types of study to be included

As we are not assessing any type of intervention, clinical trials will not be included in the final documents of our search. Therefore, we will focus primarily on observational studies, including cohort and case-control studies, which will be included in this systematic review.

Context

The focus of this review will be on proteins associated with type 2 diabetes that can be identified through proteomic technologies after the downstream analysis of quantification data.

On one hand, proteomic techniques, which have been in use since the early 2000s, allow for the global assessment and comparison of proteomes from different conditions. On the other hand, the prevalence of type 2 diabetes has been steadily increasing in developed countries over the last 40 years, mainly due to sedentary lifestyles and increased sugar intake, making it a key target for biomedical research.

Thus, we consider it necessary to gather the results from various studies focusing on the proteomic status of type 2 diabetic patients in order to construct a comprehensive understanding of the proteomic alterations associated with this condition. Comprehensive proteomic profiling may offer new insights into the dysregulated metabolic environment of type 2 diabetes, and in the future, could serve as a valuable tool for personalized medicine. This underscores the need for a better understanding of circulating protein patterns at the early stages of type 2 diabetes, as well as the dynamics of protein patterns during changes in metabolic status.

Finally, we also plan to compare these results with those obtained from our own data, in which different diabetic conditions are analyzed.

Main outcome(s)

The main outcome of this systematic review will be the differentially expressed proteins, that is, the proteomic biomarkers that define a type 2 diabetic condition compared to a control group. These proteomic biomarkers must be identified through differential quantification analysis between groups, and the quantification data must have been obtained using high-throughput proteomics techniques. All biomarkers will be presented with both their protein symbol and their UniProt identifier.

Measures of effect

The principal measure of effect will be the log fold change. Other available measures related to differential protein expression, like logFC, p-values, t-values or group means (+- standard deviations) will be standardized to log fold change. Description of outcomes:

1) **Log_2 Fold Change (log_2FC):** A metric that measures the ratio of protein levels between type 2 diabetic patients and normoglycemic controls, expressed in logarithmic terms. It is the logarithm (base 2) of the ratio of mean protein quantities**.**
2) **P-values / adjusted p-values:** Derived from the statistical test, p-values or their adjusted versions are necessary to estimate the z-score. The z-score (the number of standard deviations a data point is from the mean) is used to evaluate the consistency and significance of protein quantification changes across studies**.**

Additional outcome(s)

Not applicable.

Data extraction (selection and coding)

Study selection

Julia García Currás will prepare the research equation and will perform the search in those databases listed in the search strategy section. Afterward, all results will be combined and after a first filter perform by Julia García Currás, a final subset of article will be independently screen for inclusion by the previous researcher and Raquel Pérez Lois. Researchers will be blinded to each other's decisions before screening is completed. Disagreements will be resolved by a third part, María P. Pata. This entire process will be conducted using some R packages for systematic reviews (*litsearch*) in R software (v>4.4.0). Other tools for managing spreadsheets (Microsoft Excel) may be also used in this step, as well as reference management software (Mendeley).

Data extraction

The main outcome will be protein biomarkers with differential quantities between groups identified in each study. Protein names for these biomarkers will be converted to UniProt identifiers and presented alongside with the protein symbol for better interpretation. For each of them, the following data will be extracted:

- Log_2FC, $log_{10}FC$
- Ratio of intensities between groups of study
- Adjusted p-value
- P-value
- Mean and standard deviation for groups

In case of missing data, raw proteomic data will be downloaded and reanalyzed.

In addition, we will retrieve the following information from each study:

- Publication:
    - Publication date
    - Journal

- Study design, proteomic methods and analysis:
    - Type of design (cohorts, case-control, observational
    - Sample size
    - Type of sample (blood plasma, blood serum, urine, saliva, tissue biopsy)
    - Proteomic technique (DDA or DIA).
    - Software of downstream data analysis (R, Perseus, Python…).
    - Normalization method for intensities.
    - Imputation method for missing data
    - Statistical test in differential expression analysis (t-test, welch test, eBayes approximation).
- Demographic information:
    - Age (mean, SD)
    - Gender (percentage of females and males)
    - Ethnic group (Caucasian, Sub-Saharan African, East Asian, Native American, Polynesian, Dravidian)
    - Weight (mean, SD)
    - Height (mean, SD)
    - Body Mass Index (mean, SD)
- Clinical data from type 2 diabetic group (summary statistics for each study):
    - Duration of diabetes (mean, SD)
    - Treatment (percentage of patients with each treatment)
    - Fasting blood glucose levels, % (mean, SD)
    - HbA1c levels, % (mean, SD)
- Lifestyle factors (when available):
    - Smoking
    - Alcohol consumption
    - Diet

Julia García Currás will extract the data from the selected studies.

Regarding missing data, study investigators will be contacted for unreported data or additional details. The absence of information on logFC, biomarker ratios or any measure that can be standardized to logFC will result in exclusion from the study. For the rest of the variables, missing data will directly impact on their quality in the risk of bias assessment. The final data will be recorded in an Excel spreadsheet, and R software will be used to reanalyse raw quantification matrix to generate study outcomes. Additionally, some R packages, such as *tabulaPDF, juicr*, may also be used to retrieve information from figures or tables.

Risk of bias (quality) assessment

At the **study level**, the risk of bias will be assessed using the **QUADOMICS** tool, an adaptation of the QUADs tool, which focuses on four key aspects: data quality, information quality, analysis quality, and interpretation quality. For each study, we will evaluate the integrity and accuracy of the experimental data, ensuring that proper experimental conditions were followed and that the data acquisition methods were robust. Additionally, we will assess the completeness and consistency of the associated metadata, such as experimental protocols and sample handling. The quality of the bioinformatic analysis will be examined by reviewing the algorithms and statistical methods used, ensuring that appropriate techniques were applied. Finally, we will evaluate the interpretation of the results, considering the biological and clinical context, and ensuring that the conclusions drawn are based on sound data analysis. This

comprehensive approach will enable us to critically assess the reliability and potential biases in the proteomic data of the studies included in the review.

Quality assessment will be performed by Julia García Currás.

Strategy for data synthesis

A meta-analysis will be carried out using R (v > 4.4.0) to assess the average fold change and adjusted p-value of each selected protein biomarker across all included studies. Only those proteins found in at least two different studies will be included in this initial analysis. The previous metrics for each protein across studies will be aggregated considering the group size using a weighted p-value combination as a variant of the Fisher's method and a weighted average of the logFC, respectively. Both approaches were implemented in the *Amanida* R package. The contribution of each study will be determined by calculating the percentage of proteins from that study relative to the total number of proteins involved in the aggregation. To assess the influence of individual studies on the results of this omics-based meta-analysis, an influence analysis will be performed: data will be reanalysed repeatedly, each time excluding one study, followed by calculation of the Mean Absolute Percentage Error (MAPE) between the new FC and the original FC (previously estimated using all studies). This calculation will be performed only for the subset of proteins included in each reanalysis. Higher MAPE values indicate greater differences between the aggregated logFC from the full meta-analysis and those from the leave-one-out study subsets.

In addition, proteins previously mentioned as relevant biomarkers for T2D in the scientific literature and share by a substantial number of articles will be analysed using a hierarchical random-effects model with restricted maximum likelihood estimation, to explore potential differences between the diabetic and control groups. The logFC and corresponding 95% confidence interval will be computed for each individual study. Additionally, an overall summary of the effect size will be reported, reflecting the pooled data from the studies in each group. Heterogeneity test and measures (eta^2 and Higgin´s I^2) will be obtained.

The goodness of fit will be evaluated by sensitivity analyses based on outlier and influential case diagnostics plots (standardized residuals, hat values and Cook distance), and by Likelihood profile plots to ensure identifiability of variance components. The assessment of publication bias will be evaluated in detail through contour enhanced funnel plots, trim and fill analysis, the three parameter model selection method, and the Rosenberg method.

The *Amanida*, *Meta*, and *metafor* R packages will be used for this part of the analysis.

In addition, we will use the STRING tool through its implementation in R, via the *rbioapi* package, to build the interaction network of all the biomarkers retrieved in this study. Using this information, we will also perform an overrepresentation analysis with a hypergeometric test, a universal background, and both Gene Ontology and KEGG databases to determine the main functions and pathways related to the final pool of biomarkers, using the *clusterProfiler* R package.

Statistical significance will be determined at an alpha level of 0.05 or lower.

Analysis of subgroups or subsets

The proteomic acquisition mode and type of sample will be considered for moderator analysis.

Dissemination plans

This systematic review is part of a doctoral project and will be included in the final report. Additionally, a paper will be submitted to a leading journal in biomedicine as soon as we complete the analysis.

Type and method of review

Diagnostic, Meta-analytic, Systematic review

Language

English.

Country

Spain

Keywords

Type 2 diabetes; proteomics; biomarker; data independent acquisition; shotgun proteomics; logFC; meta-analysis; systematic review.

Current review status

Reporting phase: all analyses were conducted, and a scientific article will be written to spread the results of the current work.