nature research

Corresponding author(s): ANDRZEJ KROLEWSKI
MONIKA NIEWCZAS NMED-A91304B

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	SAS v. 9.4, Cary NC; proteomic measurements outsourced to the vendor: Somalogic Inc, Boulder CO; Affymetrix U133A array; O-link
Data analysis	SAS. v.9.4, Cary, NC; Affymetrix U133A array (ArrayQualityMetrics R package, RMA16 algorithm by just RMA() function), GeneSpring GX software, version 12.6 (Agilent Technologies).

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All global proteomic profiling data coming from the prospective study of 3 cohorts followed for ESRD risk supporting our findings are provided in the Supplementary

information of this article. Raw datasets may be available to investigators upon reasonable requests. A web-based database of all the proteomic and metabolomic profiling data of the Joslin Kidney Study will be launched soon to allow single query-based searches of the compounds and their associations with renal outcomes.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	There are no prior studies using this untargeted proteomic technology to study determinants of diabetic kidney disease, so we had no prior knowledge regarding the potential effect size. Nevertheless, our previously published targeted study of a population comparable in size revealed the association between the biomarker and progression to ESRD in the strength of a p value: p<10-12 (PMID: 22266663). Our untargeted proteomic study imposes a Bonferroni correction for a number of biomarkers measured (n=194) and that results in alpha: α = 2.5 x 10-4. Our current study is sufficiently powered to detect associations of the biomarker with the renal outcome of comparable strength to the previously reported and even those biomarkers with smaller strength of associations. Please note that our study population comprises 3 independent prospective cohorts.
Data exclusions	There were no data exclusions.
Replication	Our prospective study design comprised 3 independent cohort studies of subjects with both diabetes types and varying ethnic ancestries. We rigorously adjusted our results for multiple testing: Bonferroni correction in the discovery panel, nominal alpha = 0.01 in the validation panel and nominal alpha = 0.05 in the confirmation panel. Our signature comprising 17 proteins was significantly robust in the discovery and validation panel (by design) and 15 out of 17 proteins were confirmed further in the confirmation panel.
Randomization	Randomization was not applicable to our study design. Confounding was controlled in the study design phase (restriction to certain stages of chronic kidney disease or albuminuria) together with evaluation of potential confounders or mediators in the multivariable models.
Blinding	The Somalogic research team performing outsourced proteomic measurements were blinded to the caseness status of the samples. Samples were balanced by caseness in-house prior to sending the specimens. Otherwise blinding was not applicable.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	Discovery panel (n=219): Joslin Kidney Study, Type 1 Diabetes, mean age: 45yrs, 48% male, overt diabetic kidney disease; Validation panel (n=144): Joslin Kidney Study, Type 2 Diabetes, mean age 60yrs, 65% male. overt diabetic kidney disease; Confirmation panel (162): Pima Indian Study, Type 2 Diabetes, mean age 44yrs, 29% male, early diabetic kidney disease. 1KGP: glomeruli (n=23), Type 2 Diabetes, mean age 63 yrs, male 44%, overt diabetic kidney disease; 1KGP: tubules (n=37), Type 2 Diabetes, mean age 66 yrs, male 54%, overt diabetic kidney disease; Baricitinib trial (placebo group: n=25, male 75%; 4 mg baricitinib: n=17, male 60%), overt diabetic kidney disease.
Recruitment	The Joslin Kidney Study includes subjects with diabetes who were in CKD stages 3 when enrolled between 1991 and 2006 and were followed for 8 years (median) to ascertain onset of ESRD and rates of renal function decline. Pima Indians from the Gila River Indian Community in Arizona participated in a longitudinal study of the natural history of diabetes and its complications. We selected 310 subjects from this detailed study who had type 2 diabetes, measured GFR and were in CKD stage 1-2 and were

enrolled into the study between 1994 and 2008. They were followed for 11 years (median) to ascertain onset of ESRD and to determine rates of renal function decline. Our study design relied on prospective observation of the three cohorts, therefore the concern of the self-selection bias with regard to the risk of outcomes has been eliminated. Our three-cohort had certain characteristics attributed to their ethnicity/race or geographic location (North America), thus there is a possibility that our findings may not be generalizable to patients in different continents/regions or in patients with non-diabetic kidney diseases. Study participants of the baricitinib, phase II trial (NCT01683409) included subjects with type 2 diabetes, eGFR of 25-70 ml/min/1.73m2 and severely increased albuminuria recruited at 40 sites in Japan, Mexico and the USA. Therefore, the results that we obtained in patients participating in this trial can be generalized to all patients with T2D.