




# Comparison of serum and urinary biomarker panels with albumin/creatinine ratio in the prediction of renal function decline in type 1 diabetes

Marco Colombo<sup>1</sup> · Stuart J. McGurnaghan<sup>2</sup> · Luke A. K. Blackburn<sup>2</sup> · R. Neil Dalton<sup>3</sup> · David Dunger<sup>4,5</sup> · Samira Bell<sup>6</sup> · John R. Petrie<sup>7</sup> · Fiona Green<sup>8</sup> · Sandra MacRury<sup>9</sup> · John A. McKnight<sup>10</sup> · John Chalmers<sup>11</sup> · Andrew Collier<sup>12</sup> · Paul M. McKeigue<sup>1</sup> · Helen M. Colhoun<sup>2,13</sup>  · on behalf of the Scottish Diabetes Research Network (SDRN) Type 1 Bioresource Investigators

Received: 2 September 2019 / Accepted: 2 December 2019 / Published online: 8 January 2020  
© The Author(s) 2020

## Abstract

**Aims/hypothesis** We examined whether candidate biomarkers in serum or urine can improve the prediction of renal disease progression in type 1 diabetes beyond prior eGFR, comparing their performance with urinary albumin/creatinine ratio (ACR).

**Methods** From the population-representative Scottish Diabetes Research Network Type 1 Bioresource (SDRNT1BIO) we sampled 50% and 25% of those with starting eGFR below and above  $75 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$ , respectively ( $N = 1629$ ), and with median 5.1 years of follow-up. Multiplexed ELISAs and single molecule array technology were used to measure nine serum biomarkers and 13 urine biomarkers based on our and others' prior work using large discovery and candidate studies. Associations with final eGFR and with progression to  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$ , both adjusted for baseline eGFR, were tested using linear and logistic regression models. Parsimonious biomarker panels were identified using a penalised Bayesian approach, and their performance was evaluated through tenfold cross-validation and compared with using urinary ACR and other clinical record data.

**Results** Seven serum and seven urine biomarkers were strongly associated with either final eGFR or progression to  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$ , adjusting for baseline eGFR and other covariates (all at  $p < 2.3 \times 10^{-3}$ ). Of these, associations of four serum biomarkers were independent of ACR for both outcomes. The strongest associations with both final eGFR and progression to  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  were for serum TNF receptor 1, kidney injury molecule 1, CD27 antigen,  $\alpha$ -1-microglobulin and syndecan-1. These serum associations were also significant in normoalbuminuric participants for both outcomes. On top of

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00125-019-05081-8>) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

✉ Helen M. Colhoun  
helen.colhoun@igmm.ed.ac.uk

<sup>1</sup> Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

<sup>2</sup> MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK

<sup>3</sup> Evelina London Children's Hospital, Guy's and St Thomas' National Health Service Foundation Trust, London, UK

<sup>4</sup> Department of Paediatrics, University of Cambridge, Cambridge, UK

<sup>5</sup> Wellcome Trust–MRC Institute of Metabolic Science, University of Cambridge, Cambridge, UK

<sup>6</sup> Renal Unit, Ninewells Hospital, Dundee, UK

<sup>7</sup> Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

<sup>8</sup> Research & Development Support Unit, Dumfries & Galloway Royal Infirmary, Dumfries, UK

<sup>9</sup> Department of Diabetes and Cardiovascular Science, University of Highlands and Islands, Inverness, UK

<sup>10</sup> Western General Hospital, NHS Lothian, Edinburgh, UK

<sup>11</sup> Diabetes Centre, Victoria Hospital, NHS Fife, Kirkcaldy, UK

<sup>12</sup> School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, UK

<sup>13</sup> Public Health, NHS Fife, Kirkcaldy, UK

## Research in context

### What is already known about this subject?

- Measuring current eGFR using serum creatinine is important for predicting renal function decline
- Urinary albumin/creatinine ratio (ACR) measurement is in widespread use clinically alongside serum creatinine for predicting future renal status, but requires urine samples, which can be inconvenient for people with diabetes, and in practice many people end up without the recommended annual test
- Many individual biomarkers have been analysed and reported as associated with renal outcomes

### What is the key question?

- How do novel candidate serum and urine biomarkers compare with urinary ACR in predicting progression of renal disease in type 1 diabetes?

### What are the new findings?

- Of 22 biomarkers tested, 14 were strongly associated with a measure of eGFR progression, and for four this was independent of ACR
- A parsimonious panel of serum biomarkers outperformed ACR for predicting future eGFR and progression to eGFR  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$
- Measuring ACR together with the parsimonious serum biomarker panel did not increase prediction beyond that achieved with the serum panel alone

### How might this impact on clinical practice in the foreseeable future?

- If our findings are validated, serum biomarkers could provide an alternative to urinary ACR in predicting progression of renal disease that is more convenient and acceptable for people with diabetes, potentially resulting in more complete annual assessment

baseline covariates, the  $r^2$  for prediction of final eGFR increased from 0.702 to 0.743 for serum biomarkers, and from 0.702 to 0.721 for ACR alone. The area under the receiver operating characteristic curve for progression to  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  increased from 0.876 to 0.953 for serum biomarkers, and to 0.911 for ACR alone. Other urinary biomarkers did not outperform ACR.

**Conclusions/interpretation** A parsimonious panel of serum biomarkers easily measurable along with serum creatinine may outperform ACR for predicting renal disease progression in type 1 diabetes, potentially obviating the need for urine testing.

**Keywords** Clinical science · Epidemiology · Metabolomics · Nephropathy · Proteomics

### Abbreviations

ACR	Albumin/creatinine ratio
AUC	Area under the receiver operating characteristic curve
CD27	CD27 antigen
CKD	Chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
DKD	Diabetic kidney disease
ESRD	End-stage renal disease
IQR	Interquartile range
KIM-1	Kidney injury molecule 1
MCP-1	Monocyte chemoattractant protein 1
SIMOA	Single molecule array
TNFR1	TNF receptor 1

### Introduction

To facilitate the early identification of those most at risk of decline in future renal function, most clinical guidelines recommend the use of regular serum creatinine measurement and equations such as the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) or Modification of Diet in Renal Disease (MDRD) equations for estimation of GFR [1]. The best predictor of future end-stage renal disease (ESRD) is the current eGFR and past eGFR trajectory [2]. As we recently reviewed [3], while serum cystatin C-based eGFR equations have also been proposed, at present creatinine-based equations remain the most widely used for estimating GFR [4–8]. Guidelines also advocate regular urinary testing for albumin/creatinine ratio (ACR) since there is extensive

evidence that albuminuria is a strong risk factor for progression of diabetic kidney disease (DKD). However, it has been widely discussed that ACR lacks specificity and sensitivity for progressive decline in eGFR. For example, a poor positive predictive value was reported, with only about a third of those with microalbuminuria having progressive renal function decline [9]. Albumin excretion also had low sensitivity in that only about half of those with progressive renal function decline were albuminuric [9]. A further practical limitation of ACR is that obtaining urine samples for ACR adds to clinic workload and can be inconvenient and difficult for people with diabetes, resulting in suboptimal uptake. In Scotland, for example, while 86% of those with type 1 diabetes had eGFR in the past year, only 62% had ACR measured [10].

An important question is therefore whether any other prognostic biomarkers can improve on ACR or even replace it for predicting renal function decline. In a recent paper [11] we reported a large-scale discovery experiment where we assessed 297 biomarkers in two type 1 diabetes cohorts with starting eGFRs of  $75 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  or less. We found many highly statistically significant biomarker associations with future eGFR, and that prediction of future eGFR using ACR and baseline eGFR could be improved with biomarkers. However, most of the improvement could be achieved by measuring just serum kidney injury molecule 1 (KIM-1) and CD27 antigen (CD27), a member of the TNF superfamily.

Following on from that study, we extended the evaluation to (1) include study participants with higher starting eGFRs; (2) measure the best biomarkers from our first study i.e. KIM-1 and CD27, but also measure new candidate biomarkers chosen from recent reports from large well-conducted studies; and (3) evaluate a more sensitive serum KIM-1 assay using single molecule array (SIMOA) technology, hypothesising that it would be detectable before microalbuminuria in those progressing to renal function decline. The focus of our analysis was to assess how well these biomarkers predict future eGFR and whether they could outperform or add to the prediction gained with ACR across a wide range of initial eGFRs.

## Methods

**Participants** The Scottish Diabetes Research Network Type 1 Bioresource (SDRNT1BIO) [12] is a prospective cohort study comprising 6127 people with a clinical diagnosis of type 1 diabetes mellitus, representing 25% of all adults with type 1 diabetes in Scotland, recruited between December 2010 and November 2013. On the day of recruitment (which we refer to as the study day or biosample date), clinical measurements and blood and urine samples were obtained in which serum creatinine and urinary ACR were measured. From electronic healthcare records we extracted routine health-related data retrospective and prospective to study day, as described [12].

For this study we selected 1629 individuals with eGFR  $30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  or above at study biosample date and with at least three prospective eGFR determinations over a period of at least 2 years or incident ESRD. These were a random sample of 50% and 25% of those with starting eGFR below and above  $75 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$ , respectively.

The study was performed in accordance with the Declaration of Helsinki; all participants gave their written consent and the study protocol was approved by the local ethics and data governance committee.

**Renal outcomes and covariate data** eGFR was calculated with the CKD-EPI equation [13] using serum creatinine values directly measured and retrieved retrospectively and prospectively from medical records. These excluded readings concurrent with hospital admissions. A summary measure of the historical eGFR was obtained by computing a weighted average of all retrospective eGFR records for each person, with weights inversely related to the amount of time leading to the biosample date. Participants with no retrospective eGFR data had their historical eGFR imputed to study day eGFR. Final eGFR was defined as the median eGFR reading of the last 6 months of follow-up. Initiation of renal replacement therapy was considered to indicate an achieved eGFR of  $10 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  and all subsequent readings were censored. The decline of renal function was estimated by fitting a simple linear regression model to the serial prospective eGFR determinations of each person. We also defined binary clinically significant progression categories of progression to  $<30$  and  $<45 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$ .

ACR was measured in paired urine samples with the first taken at study day and the second several days later using the ADVIA 2400 immunoturbidimetric method for albumin and the ADVIA 2400 enzymatic method for creatinine (Siemens, Munich, Germany). These data were used for adjusting for ACR in the analyses here. In addition, longitudinal urinary ACR data were captured from the routine clinical laboratory data. Clinical record data close to study day were highly correlated ( $r = 0.73$ ) with the direct measurement. At any time point, albuminuric status was defined based on the most recent available albuminuria measurement provided there was no contradictory record of that stage in the preceding or subsequent 90 days; i.e. someone who transitioned from normo- to microalbuminuria but then had another normoalbuminuria measurement within 90 days was assigned as having been normoalbuminuric across that period, such that transient changes in albuminuria readings were ignored. Participants were classified as normo-, micro- or macroalbuminuric according to their ACR falling in the intervals 0–3.39, 3.39–33.9 or above 33.9 mg/mmol, based on two out of three consecutive measurements before baseline.

Retrospective and prospective covariate data including drug exposures were obtained from the electronic healthcare record SCI-Diabetes as described previously [12].

**Biomarkers measured and analysed** Serum and urine biomarkers were measured on samples stored at  $-80^{\circ}\text{C}$  with no prior freeze/thaw cycles. The assay methods used and the quality control performance are summarised in electronic supplementary material (ESM) Table 1. The candidate biomarkers were chosen either because we had already shown these proteins to be predictive of eGFR decline at chronic kidney disease (CKD) stage 3 or worse (serum KIM-1, CD27,  $\alpha$ -1-microglobulin, thrombomodulin) or because they were reported from other well-conducted studies as predictive of renal disease progression (EGF and its ratio to monocyte chemotactic protein 1 [MCP-1], EGF receptor, cystatin C, macrophage inhibitory protein, matrix metalloproteinase 8, TNF receptor 1 [TNFR1]) or as strong candidates from known biology of DKD (syndecan-1), or because they are routinely measured on the same multiplex panel as a candidate (the remainder).

We used a combination of assays using (1) the Luminex platform (Austin, TX, USA) at the Clinical Laboratory Improvement Amendments (CLIA)-certified Myriad RBM laboratory (Austin, TX, USA); (2) a high-sensitivity SIMOA assay for KIM-1 at Myriad RBM that we had found detected KIM-1 in samples KIM-1 negative on their standard Luminex assay; (3) R&D Systems (Minneapolis, MN, USA) Luminex assay at the Immunoassay Biomarker Core Laboratory, University of Dundee (Dundee, UK).

Intraclass correlations were computed over 48 blinded duplicate aliquots to evaluate the reproducibility of the measurements obtained. Biomarkers were excluded from the analysis if their intraclass correlation was less than 0.4, or if over 99% of the readings in the dataset were identical due to falling below the detection threshold. Accordingly, nine serum and 13 urinary biomarkers were included in the final analyses.

Values below the detection limit were imputed to half the detection threshold. All urine biomarkers were normalised to urinary creatinine.

**Univariate analysis** To test for associations with renal outcomes, biomarkers were evaluated independently in linear models for final eGFR and logistic regression models for progression adjusted for age, sex, diabetes duration, study day eGFR and length of follow-up (basic covariates). To allow comparison with ACR, we reran the basic and biomarker models including ACR at biosample date. We further adjusted models for BMI, systolic BP, diastolic BP,  $\text{HbA}_{1c}$ , HDL-cholesterol, total cholesterol, smoking status and a weighted summary of prior eGFR from retrospective records (full covariates). Prior to fitting models, all continuous covariates and biomarkers were Gaussianised and standardised to zero mean and unit standard deviation. Associations were declared significant at Bonferroni-corrected  $p = 0.05/22 = 2.3 \times 10^{-3}$ .

**Construction of parsimonious panels of biomarkers** Urine and serum biomarker sets were modelled independently from each other. As previously described [11], we adopted a Bayesian modelling approach based on hierarchical shrinkage priors, in which the clinical covariates used to control for confounding in the models were assigned a weakly informative Gaussian prior (which induces some regularisation as in ridge regression), while biomarkers were penalised through the regularised horseshoe prior (which heavily shrinks regression coefficients toward zero unless they are informative) to promote sparsity [14]. A similar approach was also adopted elsewhere for biomarker selection in the context of type 2 diabetes mellitus [15]. The hierarchical shrinkage approach was implemented using the Stan Bayesian inference framework [16], which uses Markov chain Monte Carlo (MCMC) to sample the posterior distribution of the parameters given the data and the model. The specific models implemented are available in the R package *hsstan* (version 0.6: <https://CRAN.R-project.org/package=hsstan>).

We evaluated the predictive performance of models on withdrawn data using tenfold cross-validation. For each set of baseline covariates used, we reported the difference in log-likelihood (computed on the observations withdrawn for testing from tenfold cross-validation) between the model with baseline covariates and biomarkers and the model including only the clinical covariates. For models of achieved eGFR we computed the  $r^2$  as the squared Pearson correlation coefficient between observed and predicted outcome on the test folds. For models of progression, we reported the area under the receiver operating characteristic curve (AUC) and the expected information for discrimination,  $\Lambda$ , expressed in bits [17]. This measure quantifies the gain in information that a set of biomarkers provides over and beyond the baseline clinical covariates. The expected information for discrimination was computed with the R package *wevid* (version 0.6.2: <https://CRAN.R-project.org/package=wevid>).

To recover a parsimonious model, we then applied projection predictive variable selection [18]. This approach is based on projecting the high-dimensional draws from the posterior of the model containing all biomarkers (full model) onto lower-dimensional subspaces corresponding to sparse candidate submodels [19, 20]. Predictions made by each candidate submodel are compared with those obtained by the full model: their discrepancy is evaluated using the Kullback–Leibler divergence, which measures the information lost when a smaller submodel is used to approximate a more complex one. By recursively choosing in a forward-selection fashion the submodel that minimises the Kullback–Leibler divergence from the full model, one can construct a series of candidate models of growing complexity. We evaluated each candidate model in terms of its relative contribution towards the performance of the full model, and plotted the relative explanatory power obtained by biomarker panels of different sizes.



## Results

**Participant characteristics** Table 1 reports the summary characteristics of those included stratified by CKD stage. The median (interquartile range [IQR]) of follow-up was 5.1 (4.4, 5.7) years. For the evaluation of progression status there were a median (IQR) of 9 (6, 14) measurements of eGFR prospectively. The median change in eGFR was a fall of  $0.8 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  per year. Overall, 41 and 74 participants, respectively, progressed to  $<30$  and  $<45 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  (corresponding to 2.5% and 4.7% of participants not already below these thresholds). Among the 1404 normoalbuminuric individuals at baseline, 23.1% developed albuminuria during follow-up.

**Biomarkers explored** ESM Table 1 shows the median, IQR and range of the biomarkers analysed in this study. All candidate biomarkers showed good intraclass correlation coefficients ( $\geq 0.75$ ) when measured on blinded duplicate samples. ESM Table 2 lists the biomarkers that were excluded from the analysis.

The correlation matrix for the biomarkers ordered by hierarchical clustering is shown in Fig. 1. Four distinct clusters are seen; for example, we see that cystatin C is moderately correlated with  $\alpha$ -1-microglobulin and clusterin, but only weakly correlated with other biomarkers. TNFR1 and CD27 were very strongly correlated with each other and moderately correlated with KIM-1.

**Table 1** Participant characteristics at study day stratified by CKD stage

Covariate	Missing	G1 ( $n = 835$ )	G2 ( $n = 601$ )	G3 ( $n = 191$ )	All ( $N = 1629$ )
<b>Main characteristics</b>					
Age (years)	–	41.4 (30.4, 49.8)	54.8 (46.0, 63.5)	63.6 (55.3, 71.1)	48.3 (37.9, 59.2)
Sex (female), %	–	41.9	55.2	57.6	48.6
Diabetes duration (years)	–	19.0 (10.2, 28.3)	25.4 (16.5, 36.1)	32.1 (21.1, 42.4)	22.4 (12.7, 33.1)
<b>Observability</b>					
Retrospective study length (years)	–	5.4 (4.5, 6.0)	5.5 (5.0, 6.2)	5.5 (5.1, 6.0)	5.4 (4.8, 6.1)
Length of follow-up (years)	–	5.0 (4.3, 5.6)	5.2 (4.6, 5.7)	5.3 (4.5, 5.9)	5.1 (4.4, 5.7)
Retrospective creatinine readings (n)	–	9 (6, 14)	12 (9, 18)	16 (10, 24)	11 (7, 17)
Prospective creatinine readings (n)	–	8 (5, 11)	11 (7, 15)	15 (9, 24)	9 (6, 14)
<b>Kidney function</b>					
ACR (mg/mmol)	21	0.4 (0.2, 0.8)	0.4 (0.3, 1.4)	0.9 (0.4, 4.7)	0.4 (0.2, 1.1)
ACR category (normo/micro/macro),	21	91.9/7.1/1.1	85.9/10.2/3.9	71.7/15.8/12.5	87.3/9.2/3.5
Prevalent micro- or macroalbuminuria, %	–	8.1	14.1	28.8	12.8
Incident micro- or macroalbuminuria, %	–	16.9	21.1	28.8	19.9
eGFR ( $\text{ml min}^{-1} [1.73 \text{ m}]^{-2}$ )	–	104.4 (96.5, 114.6)	74.1 (68.0, 81.8)	51.1 (43.1, 56.0)	90.7 (70.1, 104.9)
Weighted historical eGFR ( $\text{ml min}^{-1} [1.73 \text{ m}]^{-2}$ )	115	107.6 (99.1, 116.6)	83.2 (76.6, 91.2)	61.3 (54.5, 70.6)	93.1 (79.2, 107.9)
Prospective eGFR slope ( $\text{ml min}^{-1} [1.73 \text{ m}]^{-2} \text{ year}^{-1}$ )	–	−0.7 (−2.1, 0.7)	−0.7 (−2.7, 0.7)	−1.3 (−3.5, 0.7)	−0.8 (−2.5, 0.7)
<b>Other covariates</b>					
HbA <sub>1c</sub> (mmol/mol)	3	70 (62, 81)	68 (60, 78)	68 (60, 80)	69 (61, 80)
HbA <sub>1c</sub> (%)	3	8.6 (7.8, 9.6)	8.4 (7.6, 9.3)	8.4 (7.6, 9.5)	8.5 (7.7, 9.5)
HDL-cholesterol (mmol/l)	54	1.5 (1.2, 1.8)	1.6 (1.3, 1.9)	1.5 (1.2, 1.9)	1.5 (1.2, 1.8)
LDL-cholesterol (mmol/l)	863	2.5 (2.0, 3.0)	2.3 (1.9, 2.9)	2.0 (1.6, 2.4)	2.4 (1.9, 2.9)
Total cholesterol (mmol/l)	21	4.6 (4.0, 5.2)	4.6 (4.0, 5.1)	4.2 (3.7, 4.9)	4.5 (4.0, 5.2)
BMI ( $\text{kg/m}^2$ )	9	26.0 (23.3, 29.5)	26.9 (24.5, 30.1)	27.1 (24.0, 31.0)	26.6 (23.8, 29.9)
Diastolic BP (mmHg)	6	75 (68, 81)	76 (68, 81)	70 (63, 80)	75 (68, 81)
Systolic BP (mmHg)	6	128 (119, 137)	133 (123, 146)	137 (124, 148)	130 (120, 141)
Ever smoker, %	–	61.9	65.2	72.3	64.3
On any anti-hypertensive treatment, %	–	28.1	58.7	85.9	46.3
On ACEi or ARB, %	–	26.1	52.9	77.5	42.1

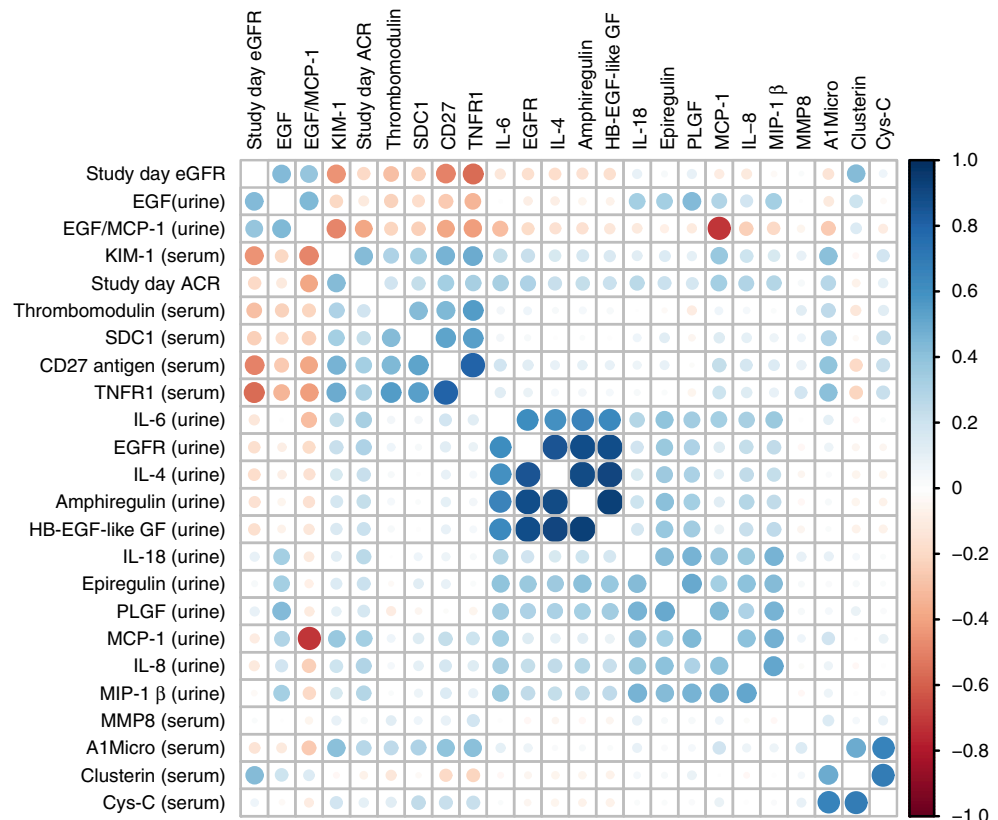
We report frequency (as %) for categorical variables and median (IQR) for continuous variables

CKD stages are defined according to ranges of eGFR in  $\text{ml min}^{-1} [1.73 \text{ m}]^{-2}$ : G1:  $>90$ ; G2: 60–90; G3: 30–60

Participants at stage G4 (eGFR = 15–30;  $n = 2$ ) are not reported as a separate column

ACEi, ACE inhibitor; ARB, angiotensin receptor blocker

**Fig. 1** Correlation matrix ordered by hierarchical clustering. A1Micro,  $\alpha$ -1-microglobulin; Cys-C, cystatin C; GF, growth factor; HB, heparin-binding; MMP8, matrix metalloproteinase-8; PLGF, placenta growth factor; SDC1, syndecan 1



**Univariate associations with final eGFR** Seven serum and six urine biomarkers, including the ratio of urinary EGF to MCP-1, were strongly associated with final eGFR in models adjusted for basic covariates including baseline eGFR (all at  $p < 2.3 \times 10^{-3}$ ; Table 2). Of these, nine remained significantly associated with final eGFR after further adjustment for ACR at biosample date. The most strongly associated biomarkers, TNFR1, KIM-1, CD27, syndecan-1,  $\alpha$ -1-microglobulin and urinary EGF/MCP-1 ratio, were also significantly associated with final achieved eGFR when restricted to those normoalbuminuric at biosample date and even when adjusted for ACR in this normoalbuminuric range (data not shown). Furthermore, other than  $\alpha$ -1-microglobulin, these biomarkers were also highly significantly associated with final eGFR at follow-up adjusted for ACR at baseline when restricted to those with eGFR  $\geq 90 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  at baseline (data not shown). We confirmed that the associations were similar when adjusted for use of ACE inhibitors or angiotensin receptor blockers.

Using progression status to  $<30$  or  $<45 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  as the outcome, a very similar pattern was seen for serum biomarkers (ESM Tables 3 and 4). For example, adjusted for basic covariates and ACR, the odds of progression to  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  were 5.80-fold for every 1 SD in Gaussianised TNFR1, and 2.05-fold per SD of KIM-1 (ESM Table 3). Among the urinary biomarkers, EGF/MCP-1 ratio was associated with progression, but when adjusted for ACR the association remained significant for progression to

$<45$  but not  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$ . When restricted to those with normo- or microalbuminuria at baseline, a very similar pattern of association with progression was seen. Among those with macroalbuminuria at baseline, significant associations with outcomes were not found, but the sample size was very small for testing this.

**Panels of biomarkers for predicting eGFR** Table 3 summarises the cross-validated performance of the linear regression models for prediction of final eGFR using all of the serum or urine biomarkers (in models with hierarchical shrinkage priors) on top of clinical covariates, with and without further inclusion of ACR at biosample date. Starting from basic covariates, the  $r^2$  for prediction of final eGFR increased from 0.702 to 0.743 for serum and to 0.721 for urine biomarkers, compared with an increase to 0.722 for ACR alone. Thus, the serum biomarkers alone outperform ACR alone. As shown in Table 3, the model including serum biomarkers with ACR is better than ACR alone.

Similarly, from Table 4 the AUC for progression to  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  was 0.911 using ACR with basic covariates, but was 0.953 with serum biomarkers and basic covariates, and did not increase further after adding ACR to serum biomarkers. Using the expected information for discrimination,  $\Lambda$ , Table 4 shows that serum biomarkers contain almost one extra bit of information for the prediction of progression to  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  than does ACR (4.06 vs 3.23 bits).

**Table 2** Associations of each biomarker (considered separately) with final eGFR

Biomarker	Basic covariates		Basic covariates + ACR		Full covariates	
	$\beta$ (95% CI)	<i>p</i> value	$\beta$ (95% CI)	<i>p</i> value	$\beta$ (95% CI)	<i>p</i> value
<b>Serum biomarkers</b>						
TNFR1	−0.20 (−0.24, −0.17)	$1.4 \times 10^{-35}$	−0.17 (−0.20, −0.14)	$3.0 \times 10^{-24}$	−0.12 (−0.16, −0.09)	$3.2 \times 10^{-14}$
KIM-1	−0.18 (−0.21, −0.16)	$4.8 \times 10^{-35}$	−0.14 (−0.17, −0.11)	$1.7 \times 10^{-19}$	−0.14 (−0.17, −0.11)	$7.0 \times 10^{-20}$
CD27	−0.17 (−0.20, −0.14)	$7.1 \times 10^{-29}$	−0.14 (−0.17, −0.11)	$1.8 \times 10^{-18}$	−0.11 (−0.13, −0.08)	$1.4 \times 10^{-12}$
$\alpha$ -1-microglobulin	−0.12 (−0.14, −0.09)	$1.1 \times 10^{-17}$	−0.08 (−0.11, −0.06)	$7.4 \times 10^{-10}$	−0.07 (−0.10, −0.05)	$2.9 \times 10^{-08}$
Syndecan 1	−0.12 (−0.15, −0.09)	$5.1 \times 10^{-17}$	−0.09 (−0.12, −0.06)	$1.0 \times 10^{-10}$	−0.07 (−0.10, −0.04)	$1.0 \times 10^{-07}$
Thrombomodulin	−0.10 (−0.13, −0.07)	$2.0 \times 10^{-11}$	−0.07 (−0.10, −0.05)	$2.6 \times 10^{-07}$	−0.05 (−0.07, −0.02)	$3.1 \times 10^{-04}$
Cystatin C	−0.05 (−0.07, −0.02)	$5.9 \times 10^{-04}$	−0.03 (−0.06, 0.00)	$2.8 \times 10^{-02}$	−0.03 (−0.06, −0.01)	$1.3 \times 10^{-02}$
Matrix metalloproteinase-8	−0.04 (−0.07, −0.01)	$2.6 \times 10^{-03}$	−0.03 (−0.06, −0.01)	$1.0 \times 10^{-02}$	−0.02 (−0.04, 0.00)	$1.1 \times 10^{-01}$
Clusterin	0.01 (−0.02, 0.04)	$4.5 \times 10^{-01}$	0.02 (−0.01, 0.04)	$2.9 \times 10^{-01}$	0.00 (−0.02, 0.03)	$7.9 \times 10^{-01}$
<b>Urine biomarkers</b>						
EGF/MCP-1 ratio	0.16 (0.13, 0.19)	$1.7 \times 10^{-27}$	0.12 (0.09, 0.15)	$6.1 \times 10^{-15}$	0.10 (0.07, 0.13)	$1.1 \times 10^{-12}$
MCP-1	−0.10 (−0.13, −0.08)	$2.2 \times 10^{-14}$	−0.06 (−0.09, −0.04)	$4.6 \times 10^{-06}$	−0.07 (−0.09, −0.04)	$4.8 \times 10^{-07}$
IL-8	−0.07 (−0.10, −0.05)	$6.5 \times 10^{-07}$	−0.03 (−0.06, 0.00)	$5.3 \times 10^{-02}$	−0.02 (−0.05, 0.01)	$1.5 \times 10^{-01}$
EGF	0.07 (0.04, 0.10)	$1.0 \times 10^{-06}$	0.07 (0.04, 0.10)	$6.3 \times 10^{-06}$	0.05 (0.02, 0.07)	$1.1 \times 10^{-03}$
EGF receptor	−0.05 (−0.08, −0.03)	$1.0 \times 10^{-04}$	−0.01 (−0.04, 0.01)	$3.6 \times 10^{-01}$	−0.01 (−0.03, 0.02)	$6.1 \times 10^{-01}$
IL-18	−0.05 (−0.08, −0.02)	$5.0 \times 10^{-04}$	−0.01 (−0.03, 0.02)	$7.0 \times 10^{-01}$	0.00 (−0.03, 0.02)	$7.1 \times 10^{-01}$
IL-6	−0.04 (−0.07, −0.01)	$2.6 \times 10^{-03}$	0.00 (−0.02, 0.03)	$7.8 \times 10^{-01}$	0.00 (−0.02, 0.03)	$7.2 \times 10^{-01}$
Macrophage inflammatory protein-1 $\beta$	−0.04 (−0.07, −0.01)	$5.2 \times 10^{-03}$	0.00 (−0.02, 0.03)	$8.4 \times 10^{-01}$	0.00 (−0.02, 0.03)	$8.4 \times 10^{-01}$
Amphiregulin	−0.04 (−0.06, −0.01)	$1.0 \times 10^{-02}$	0.00 (−0.03, 0.02)	$7.5 \times 10^{-01}$	0.01 (−0.02, 0.03)	$6.6 \times 10^{-01}$
Placenta growth factor	−0.03 (−0.05, 0.00)	$5.3 \times 10^{-02}$	0.00 (−0.02, 0.03)	$8.2 \times 10^{-01}$	0.00 (−0.02, 0.03)	$9.5 \times 10^{-01}$
IL-4	−0.02 (−0.05, 0.00)	$8.2 \times 10^{-02}$	0.01 (−0.02, 0.03)	$6.9 \times 10^{-01}$	0.02 (−0.01, 0.04)	$2.3 \times 10^{-01}$
Epiregulin	−0.02 (−0.05, 0.00)	$8.6 \times 10^{-02}$	0.01 (−0.02, 0.04)	$4.7 \times 10^{-01}$	0.01 (−0.02, 0.03)	$4.8 \times 10^{-01}$
Heparin-binding EGF-like growth factor	−0.02 (−0.05, 0.01)	$1.2 \times 10^{-01}$	0.01 (−0.02, 0.03)	$6.5 \times 10^{-01}$	0.02 (−0.01, 0.04)	$2.2 \times 10^{-01}$

Regression coefficients are per unit SD of Gaussianised biomarker

Basic clinical covariates: age, sex, diabetes duration, study day eGFR, length of follow-up

Full clinical covariates: age, sex, diabetes duration, study day eGFR, length of follow-up, ACR, BMI, diastolic BP, systolic BP, HbA<sub>1c</sub>, HDL-cholesterol, total cholesterol, smoking status, weighted average of historical eGFR

The same conclusion is reached using progression to  $<45 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$ , and also when restricted to those with normo- or microalbuminuria at baseline. Among those with macroalbuminuria, however, there was no evidence of serum biomarkers outperforming or improving prediction

beyond ACR. There was no evidence of urinary biomarkers outperforming ACR for progression in any stratum.

By applying the projection predictive variable selection approach, we determined a ranking of the biomarkers within each platform. Figure 2 displays the ranking of the first few

**Table 3** Cross-validated performance of models for prediction of final eGFR

Model	Basic covariates		Basic covariates + ACR		Full covariates	
	$\Delta$ Loglik	$r^2$ (95% PI)	$\Delta$ Loglik	$r^2$ (95% PI)	$\Delta$ Loglik	$r^2$ (95% PI)
Clinical covariates only	–	0.702 (0.700, 0.704)	–	0.722 (0.720, 0.724)	–	0.758 (0.756, 0.761)
Serum biomarkers	120.9	0.743 (0.740, 0.746)	73.3	0.746 (0.743, 0.749)	56.3	0.775 (0.772, 0.777)
Urine biomarkers	54.2	0.721 (0.718, 0.724)	21.6	0.729 (0.726, 0.732)	19.2	0.764 (0.761, 0.767)

Basic and full clinical covariates are listed in the footnotes to Table 2

$\Delta$ Loglik, difference in test log-likelihood (natural logarithm) with respect to the model containing only clinical covariates; PI, posterior uncertainty interval

**Table 4** Cross-validated performance of models for prediction of final eGFR being <30 or <45 ml min<sup>-1</sup> [1.73 m]<sup>-2</sup>, overall and stratified by albuminuric status at study day

Model	Basic covariates			Basic covariates + ACR			Full covariates		
	ΔLoglik	AUC (95% PI)	Λ	ΔLoglik	AUC (95% PI)	Λ	ΔLoglik	AUC (95% PI)	Λ
Final eGFR <30 ( <i>n</i> = 1627, 41 events)									
Clinical covariates only	–	0.876 (0.858, 0.890)	2.18	–	0.911 (0.895, 0.924)	3.23	–	0.929 (0.912, 0.943)	3.93
Serum biomarkers	35.3	0.953 (0.940, 0.965)	4.06	11.2	0.952 (0.939, 0.965)	4.08	0.8	0.940 (0.920, 0.956)	4.28
Urine biomarkers	3.2	0.879 (0.852, 0.901)	2.52	–13.8	0.892 (0.866, 0.913)	2.84	–0.2	0.929 (0.912, 0.943)	3.92
Final eGFR <30 in normo-/microalbuminuric ( <i>n</i> = 1571, 18 events)									
Clinical covariates only	–	0.788 (0.737, 0.836)	1.32	–	0.793 (0.740, 0.845)	1.43	–	0.818 (0.757, 0.873)	2.10
Serum biomarkers	9.3	0.861 (0.807, 0.909)	2.20	7.3	0.856 (0.799, 0.908)	2.14	–2.6	0.815 (0.760, 0.871)	2.20
Urine biomarkers	–0.3	0.786 (0.732, 0.840)	1.31	–1.2	0.787 (0.732, 0.840)	1.32	0.3	0.819 (0.761, 0.873)	2.10
Final eGFR <30 in macroalbuminuric ( <i>n</i> = 56, 23 events)									
Clinical covariates only	–	0.750 (0.692, 0.809)	1.33	–	0.801 (0.743, 0.852)	1.95	–	0.771 (0.697, 0.843)	2.56
Serum biomarkers	4.3	0.830 (0.755, 0.896)	2.31	1.6	0.835 (0.760, 0.895)	2.33	–8.6	0.764 (0.685, 0.839)	2.46
Urine biomarkers	–1.0	0.770 (0.697, 0.835)	1.45	–4.7	0.765 (0.686, 0.833)	1.41	–7.6	0.758 (0.677, 0.834)	2.24
Final eGFR <45 ( <i>n</i> = 1573, 74 events)									
Clinical covariates only	–	0.840 (0.830, 0.849)	1.54	–	0.889 (0.881, 0.897)	2.31	–	0.901 (0.891, 0.911)	2.68
Serum biomarkers	33.6	0.899 (0.887, 0.911)	2.49	8.3	0.907 (0.894, 0.918)	2.59	5.4	0.914 (0.901, 0.926)	2.96
Urine biomarkers	15.3	0.879 (0.866, 0.890)	1.98	–5.5	0.890 (0.877, 0.902)	2.22	–0.3	0.901 (0.891, 0.912)	2.68
Final eGFR <45 in normo-/microalbuminuric ( <i>n</i> = 1529, 53 events)									
Clinical covariates only	–	0.857 (0.843, 0.870)	1.70	–	0.866 (0.854, 0.878)	1.93	–	0.876 (0.858, 0.891)	2.33
Serum biomarkers	16.3	0.891 (0.877, 0.904)	2.31	9.9	0.890 (0.875, 0.904)	2.30	5.0	0.891 (0.872, 0.908)	2.63
Urine biomarkers	–2.4	0.852 (0.834, 0.869)	1.71	–9.7	0.849 (0.831, 0.867)	1.75	–1.6	0.872 (0.851, 0.890)	2.34
Final eGFR <45 in macroalbuminuric ( <i>n</i> = 44, 21 events)									
Clinical covariates only	–	0.538 (0.429, 0.640)	0.17	–	0.529 (0.416, 0.634)	0.19	–	0.561 (0.472, 0.644)	0.25
Serum biomarkers	–10.1	0.510 (0.400, 0.627)	0.07	–9.3	0.512 (0.395, 0.634)	0.07	–30.1	0.536 (0.437, 0.634)	0.22
Urine biomarkers	–2.7	0.541 (0.422, 0.646)	0.18	–0.1	0.540 (0.428, 0.658)	0.17	–94.1	0.479 (0.379, 0.578)	–0.35

Basic and full clinical covariates are listed in the footnotes to Table 2

ΔLoglik, difference in test log-likelihood (natural logarithm) with respect to the model containing only clinical covariates; PI, posterior uncertainty interval; Λ, expected information for discrimination in bits

serum biomarkers when predicting final eGFR or progression to <30 ml min<sup>-1</sup> [1.73 m]<sup>-2</sup>, starting from the basic set of covariates. We also treated baseline ACR as another biomarker (Fig. 2b,d). For final eGFR, TNFR1 and KIM-1 together account for more than 75% of the information contained in the whole set of serum biomarkers independently of ACR, with clusterin and α-1-microglobulin making a small contribution beyond this (Fig. 2a,b). For progression to <30 ml min<sup>-1</sup> [1.73 m]<sup>-2</sup>, KIM-1 and CD27 (which is very strongly correlated with TNFR1) contain most of the predictive information (Fig. 2c). KIM-1, CD27 and syndecan 1 also contribute to the total prediction on top of ACR (Fig. 2d).

## Discussion

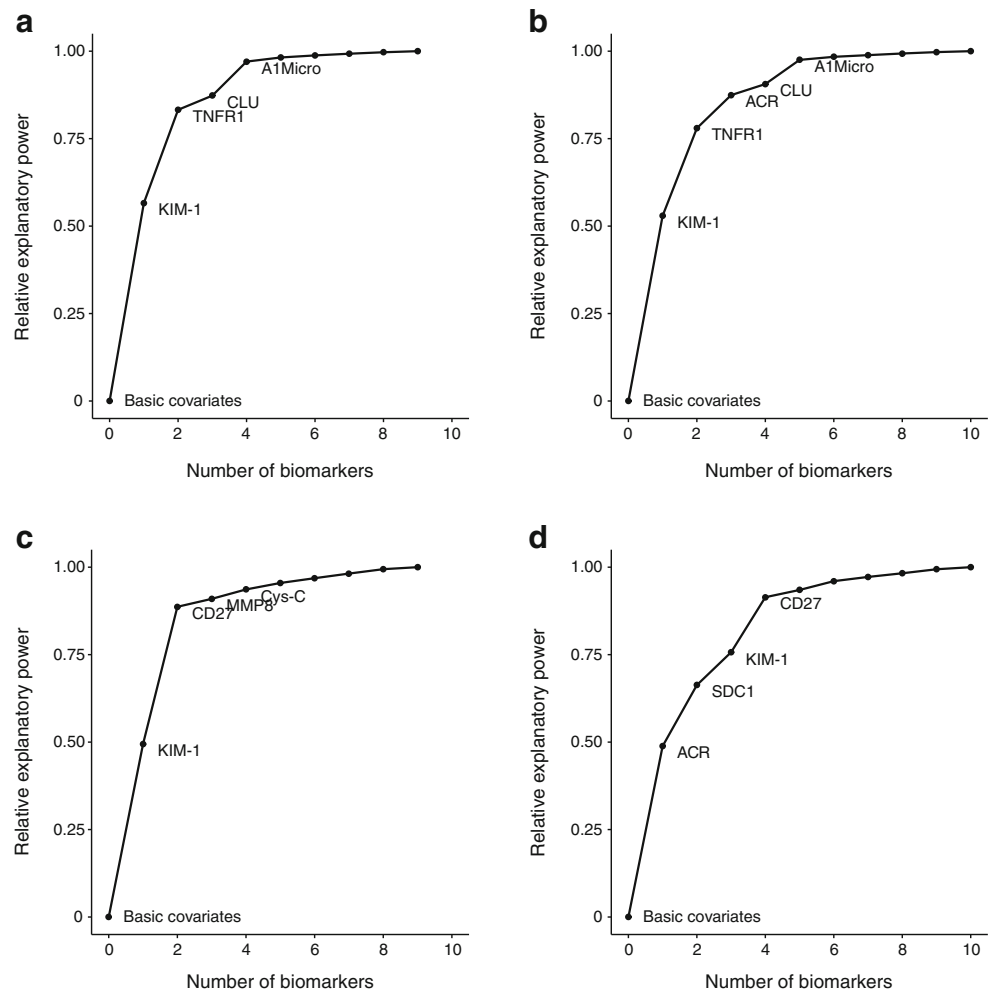
The key findings from this study across the range of CKD stages were that serum biomarkers improve the prediction of

future eGFR and progression to <30 ml min<sup>-1</sup> [1.73 m]<sup>-2</sup> beyond baseline eGFR. As in our past studies [11, 21, 22], a large number of the biomarkers evaluated showed highly significant associations with eGFR and its decline. However, almost all of the predictive information could be obtained using just a few of these intercorrelated biomarkers. Here, we found KIM-1 measured by SIMOA assay and either CD27 or TNFR1 to contain most of the predictive information. CD27 and TNFR1 are very highly correlated ( $r = 0.80$ ) and are both members of the TNF superfamily and therefore likely interchangeable as predictive biomarkers. Beyond these, there is some predictive information in syndecan-1, clusterin or α-1-microglobulin.

An important emphasis of the study reported here was a comparison of biomarkers with the performance of the widely accepted measurement ACR. In this regard, we found that a parsimonious panel of serum biomarkers performed better than ACR for predicting eGFR and eGFR progression to



**Fig. 2** Contribution of biomarker sets to prediction of final eGFR (**a, b**) or final eGFR <30 (**c, d**) when starting from a model containing basic covariates (age, sex, diabetes duration, study day eGFR and length of follow-up). Variable selection was based on: serum biomarkers (**a, c**); and serum biomarkers and ACR (**b, d**). A1Micro,  $\alpha$ -1-microglobulin; CLU, clusterin; Cys-C, cystatin C; MMP8, matrix metalloproteinase-8; SDC1, syndecan 1



<30 ml min<sup>-1</sup> [1.73 m]<sup>-2</sup>. Indeed, serum biomarkers contained almost double the prediction information (measured in bits) than ACR for progression to <30 ml min<sup>-1</sup> [1.73 m]<sup>-2</sup>. The serum biomarkers were predictive of final eGFR in those with starting eGFR above 90 ml min<sup>-1</sup> [1.73 m]<sup>-2</sup>, suggesting that they rise before ACR does and may be more sensitive to early renal damage. Of note, the urinary biomarkers did not consistently outperform ACR.

These data challenge the place of ACR in our clinical management of people with diabetes. Most clinical guidelines now recommend annual measurement of ACR in people with diabetes. Our data suggest that measuring just KIM-1 and TNFR1 or CD27 in the same sample might obviate the need for doing a urine sample collection and then measuring ACR. Collecting urine samples is time intensive in outpatient clinics and many people miss out on annual screening because they do not bring or cannot produce a urine sample. As we note above, while in Scotland we achieve high rates of annual eGFR measurement, ACR capture rates are much lower [10]. Additional differences in reporting rates for ACR than eGFR exist elsewhere, e.g. in the National Diabetes Audit in

England and Wales and in reports from Denmark [23, 24]. Clearly, further validation is required before advocating replacing ACR. Longitudinal analyses of repeat measurements of these biomarkers, for example, are needed to establish their variability and thresholds for specific risk levels warranting clinical action (akin to the microalbuminuria thresholds).

Replication in different cohorts across the range of eGFR, using a variety of assay methods, and testing for prediction of ESRD are also needed. Moreover, at the present time these biomarkers are much more expensive to assay than ACR and are not routinely available on many commonly used clinically certified assay platforms. Some may argue that ACR is also measured because it is a biomarker of more widespread vascular disease and mortality independently of renal function: this should also be shown for candidate replacement serum biomarkers. Nonetheless, at the very least our data suggest that further studies to evaluate replacing urinary ACR with serum biomarkers are warranted, given the logistic and predictive advantages they may offer.

Serum biomarkers might be useful on top of ACR rather than instead of it. In regard to the latter, it is important to note that these biomarkers also added to the prediction of final eGFR or progression even when ACR was included in the model. Furthermore, they also predicted progression to  $<30 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  in those normoalbuminuric or microalbuminuric at baseline, and did so independently of ACR. This suggests they may be useful even when ACR has already been measured and found not to be in the macroalbuminuric range. Once macroalbuminuria has been detected, however, they do not improve prediction further. Such people with macroalbuminuria would in any case be managed uniformly as a high-risk group, and it is questionable whether any further risk discrimination among them would alter clinical practice.

Our study has used eGFR and its progression to  $<30$  and  $<45 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  as endpoints. Of course, as we recently reviewed, creatinine-based eGFR is itself an imperfect measure of underlying GFR, especially at early stages of renal function decline [3]. Ideally, biomarker studies would also evaluate prediction of harder endpoints such as ESRD. However, to evaluate prediction from early stages of renal function decline to ESRD requires large unselected cohorts of people with type 1 diabetes to be followed for many years for sufficient ESRD endpoints to accrue. Such studies are logistically challenging and usually have very few ESRD endpoints in those with  $\text{eGFR} > 60 \text{ ml min}^{-1} [1.73 \text{ m}]^{-2}$  at baseline [25, 26]. Furthermore, it requires that we generalise from samples taken several decades ago to the contemporary state of diabetes. Therefore, studies such as ours with intermediate endpoints are also needed. It should be noted that the imprecision with which eGFR quantifies underlying GFR means that power to detect biomarker associations is reduced rather than false positive associations being detected. Our study will accrue ESRD endpoints as we continue follow-up, eventually allowing associations with this endpoint to be confirmed. Further analyses will also attempt to disentangle the contribution of prediction of intervening acute renal failure events to overall prediction of decline as events accrue. Another limitation of our study is that we were not able to assess incremental prediction on top of the validated kidney failure risk equation [27], as calcium and phosphate data were not available in most participants.

A key strength of our study was the deliberate use of samples from a wide range of starting eGFR. Another key strength of our study is the use of advanced statistical methods that avoid over-optimistic assessments of prediction. These include cross-validation and use of penalised models that account for the high number of analytes being evaluated and can better handle correlations between predictors.

**Conclusions** In summary, a parsimonious panel of serum biomarkers measurable along with creatinine may outperform

ACR for predicting renal disease progression from early CKD stages in type 1 diabetes, and with further development might obviate the need for urine testing.

**Acknowledgements** We thank the SDRN Type 1 Bioresource investigators: J. Chalmers (Diabetes Centre, Victoria Hospital, Kirkcaldy, UK), A. Collier (School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, UK), C. Fischbacher (Information Services Division, NHS National Services Scotland, Edinburgh, UK), F. Green (Research & Development Support Unit, Dumfries & Galloway Royal Infirmary, Dumfries, UK), R. Lindsay (British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, UK), J. McKnight (Western General Hospital, NHS Lothian, Edinburgh, UK), S. MacRury (Highland Diabetes Institute, Raigmore Hospital, NHS Highland, Inverness, UK), C. Palmer (Cardiovascular and Diabetes Medicine, University of Dundee, Dundee, UK), A. Patrick (Royal Infirmary of Edinburgh, NHS Lothian, Edinburgh, UK), D. Pearson (JJR Macleod Centre for Diabetes, Endocrinology and Metabolism, Aberdeen Royal Infirmary, Aberdeen, UK), J. Petrie (Institute of Cardiovascular & Medical Sciences, University of Glasgow, Glasgow, UK) and S. Thekkepat (David Matthews Diabetes Centre, Monklands Hospital, Airdrie, UK). We thank the nurses across the participating sites in SDRN Type 1 Bioresource, and those with diabetes who took part. The SDRN Type 1 Bioresource was originally set up under Ethics Ref. 10/S1402/43, PAC 15/13, and is now running under Public Benefit and Privacy Panel for Health and Social Care (PBPP) Ref. 1819-0315. Some of the data were presented as an abstract at the 55th Annual Meeting of the EASD in Barcelona, September 2019.

**Data availability** We do not have governance permissions to share individual level data on which these analyses were conducted since they derive from clinical record data. However, for any bona fide requests to audit the validity of the analyses, the verifiable research pipeline that we operate means that one can request to view the analyses being run and the same resulting tabulations by contacting the corresponding author. We are also happy to share summary statistics for those wishing to conduct meta-analyses with other studies.

**Funding** This study was supported by funding from the JDRF (Ref. 3-SRA-2016-332-M-R), the Chief Scientist Office (Ref. ETM/47) and Diabetes UK (Ref. 10/0004010), and by in-kind contribution from the Scottish Diabetes Research Network.

**Duality of interest** HMC has received grants, personal fees and non-financial support from Eli Lilly & Company, grants from Pfizer Inc., grants from Boehringer Ingelheim, grants from AstraZeneca LP, grants, personal fees and non-financial support from Sanofi, non-financial support from Novartis Pharmaceuticals, personal fees and non-financial support from Regeneron, has received grants and holds shares at Roche Pharmaceuticals, is a shareholder in Bayer, has received non-financial support from Sanofi Aventis, outside the submitted work. RND is a shareholder of SpOtOn Clinical Diagnostics Ltd. MC, SJM, LAKB, DD, SB, JRP, FG, SMacR, JAMcK, JC, AC and PMM declare that there is no duality of interest associated with their contribution to this manuscript. The study sponsors had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

**Contribution statement** MC carried out the primary statistical analyses and drafted and reviewed/edited the manuscript. SJM contributed to data generation, undertook quality control of the clinical diabetes datasets and reviewed the manuscript. LAKB contributed to harmonisation and databasing of data and contributed to manuscript revision. RND contributed to data generation and commented on data and manuscript drafts. DD contributed to the interpretation of data, critical discussion and

revision of the manuscript. SB, JRP, FG, SMacR, JAMcK, JC and AC contributed to data generation and reviewed and edited the manuscript. PMM contributed to study design, analysis and drafting of the manuscript. HMC conceived and designed the study, and contributed to analysis and data interpretation and drafting of the manuscript. All authors approved the manuscript for publication. HMC is the guarantor of this work.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Michels WM, Grootendorst DC, Verduijn M, Elliott EG, Dekker FW, Krediet RT (2010) Performance of the Cockcroft-Gault, MDRD, and new CKD-EPI formulas in relation to GFR, age, and body size. *Clin J Am Soc Nephrol* 5(6):1003–1009. <https://doi.org/10.2215/CJN.06870909>
2. Jones RH, Hayakawa H, Mackay JD, Parsons V, Watkins PJ (1979) Progression of diabetic nephropathy. *Lancet* 1(8126):1105–1106
3. Colhoun HM, Marcovecchio ML (2018) Biomarkers of diabetic kidney disease. *Diabetologia* 61(5):996–1011. <https://doi.org/10.1007/s00125-018-4567-5>
4. Stevens LA, Coresh J, Schmid CH et al (2008) Estimating GFR using serum cystatin C alone and in combination with serum creatinine: a pooled analysis of 3,418 individuals with CKD. *Am J Kidney Dis* 51(3):395–406. <https://doi.org/10.1053/j.ajkd.2007.11.018>
5. Barr EL, Maple-Brown LJ, Barzi F et al (2017) Comparison of creatinine and cystatin C based eGFR in the estimation of glomerular filtration rate in indigenous Australians: the eGFR study. *Clin Biochem* 50(6):301–308. <https://doi.org/10.1016/j.clinbiochem.2016.11.024>
6. Menon V, Shlipak MG, Wang X et al (2007) Cystatin C as a risk factor for outcomes in chronic kidney disease. *Ann Intern Med* 147(1):19–27. <https://doi.org/10.7326/0003-4819-147-1-200707030-00004>
7. Krolewski AS, Warram JH, Forsblom C et al (2012) Serum concentration of cystatin C and risk of end-stage renal disease in diabetes. *Diabetes Care* 35(11):2311–2316. <https://doi.org/10.2337/dc11-2220>
8. Stevens LA, Schmid CH, Greene T et al (2009) Factors other than glomerular filtration rate affect serum cystatin C levels. *Kidney Int* 75(6):652–660. <https://doi.org/10.1038/ki.2008.638>
9. Krolewski AS (2015) Progressive renal decline: the new paradigm of diabetic nephropathy in type 1 diabetes. *Diabetes Care* 38(6):954–962. <https://doi.org/10.2337/dc15-0184>
10. Scottish Diabetes Data Group (2017) Scottish Diabetes Survey 2017. Available from <http://www.diabetesinscotland.org.uk/Publications/SDS%202017.pdf>. Accessed 22 Aug 2019
11. Colombo M, Valo E, McGurnaghan SJ et al (2019) Biomarker panels associated with progression of renal disease in type 1 diabetes. *Diabetologia* 62(9):1616–1627. <https://doi.org/10.1007/s00125-019-4915-0>
12. Akbar T, McGurnaghan S, Palmer CN et al (2017) Cohort profile: Scottish Diabetes Research Network Type 1 Bioresource study (SDRNT1BIO). *Int J Epidemiol* 46(3):796–796i. <https://doi.org/10.1093/ije/dyw152>
13. Levey AS, Stevens LA, Schmid CH et al (2009) A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150(9):604–612. <https://doi.org/10.7326/0003-4819-150-9-200905050-00006>
14. Piironen J, Vehtari A (2017) Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron J Stat* 11(2):5018–5051. <https://doi.org/10.1214/17-EJS1337SI>
15. Kammer M, Heinzel A, Willency JA et al (2019) Integrative analysis of prognostic biomarkers derived from multiomics panels helps discrimination of chronic kidney disease trajectories in people with type 2 diabetes. *Kidney Int* 96(6):1381–1388. <https://doi.org/10.1016/j.kint.2019.07.025>
16. Carpenter B, Gelman A, Hoffman MD et al (2017) Stan: a probabilistic programming language. *J Stat Softw* 76(1):1–32. <https://doi.org/10.18637/jss.v076.i01>
17. McKeigue P (2018) Quantifying performance of a diagnostic test as the expected information for discrimination: relation to the C-statistic. *Stat Methods Med Res* 28(6):1841–1851. <https://doi.org/10.1177/0962280218776989>
18. Piironen J, Vehtari A (2017) Comparison of Bayesian predictive methods for model selection. *Stat Comput* 27(3):711–735. <https://doi.org/10.1007/s11222-016-9649-y>
19. Goutis C, Robert CP (1998) Model choice in generalised linear models: a Bayesian approach via Kullback-Leibler projections. *Biometrika* 85(1):29–37
20. Dupuis JA, Robert CP (2003) Variable selection in qualitative models via an entropic explanatory power. *J Stat Plan Inference* 111(1):77–94. [https://doi.org/10.1016/S0378-3758\(02\)00286-0](https://doi.org/10.1016/S0378-3758(02)00286-0)
21. Looker HC, Colombo M, Hess S et al (2015) Biomarkers of rapid chronic kidney disease progression in type 2 diabetes. *Kidney Int* 88(4):888–896. <https://doi.org/10.1038/ki.2015.199>
22. Colombo M, Looker HC, Farran B et al (2019) Serum kidney injury molecule 1 and  $\beta$ 2-microglobulin perform as well as larger biomarker panels for prediction of rapid decline in renal function in type 2 diabetes. *Diabetologia* 62(1):156–168. <https://doi.org/10.1007/s00125-018-4741-9>
23. Knudsen ST, Mosbech TH, Hansen B, König E, Johnsen PC, Kamper A-L (2012) Screening for microalbuminuria in patients with type 2 diabetes is incomplete in general practice. *Danish Medical Journal* 59:A4502
24. Healthcare Quality Improvement PHE (2019) National Diabetes Audit, 2017–18 Report 1: Care processes and treatment targets. Available at <https://files.digital.nhs.uk/88/F1E544/National%20Diabetes%20Audit%202017-18%20Full%20Report%201%2C%20Care%20Processes%20and%20Treatment%20Targets.pdf>. Accessed 21 Nov 2019
25. Nowak N, Skupien J, Niewczasz MA et al (2016) Increased plasma kidney injury molecule-1 suggests early progressive renal decline in non-proteinuric patients with type 1 diabetes. *Kidney Int* 89(2):459–467. <https://doi.org/10.1038/ki.2015.314>
26. Pilemann-Lyberg S, Rasmussen DGK, Hansen TW et al (2019) Markers of collagen formation and degradation reflect renal function and predict adverse outcomes in patients with type 1 diabetes. *Diabetes Care* 42(9):1760–1768. <https://doi.org/10.2337/dc18-2599>
27. Tangri N, Stevens LA, Griffith J et al (2011) A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 305(15):1553–1559. <https://doi.org/10.1001/jama.2011.451>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.