An Early Prediction Model for Chronic Kidney Disease

Jing Zhao

Tianjin Medical University

Yuan Zhang

Tianjin Medical University

Jiali Qiu

Tianjin Medical University

Xiaodan Zhang

Tianjin Medical University

Fengjiang Wei

Tianjin Medical University

Jiayi Feng

Tianjin Medical University

Chen Chen

Tianjin Medical University

Kai Zhang

Tianjin Medical University

Shuzhi Feng

Tianjin Medical University

Yaogang Wang

Tianjin Medical University

Wei-Dong Li (✉ liweidong98@tmu.edu.cn)

Tianjin Medical University <https://orcid.org/0000-0003-0521-7412>

Research Article

Keywords: prediction model, chronic kidney disease, nested case-control study, biomarkers, risk factors

Posted Date: June 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-513054/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Identifying individuals predisposed to chronic kidney disease (CKD) is crucial for intervention and treatment, however, risk equations were relatively rare among the middle-aged and elderly in China.

Methods: Eventually, 116 CKD patients and 232 their healthy counterparts from “Tianjin Medical University Chronic Disease Cohort (N=21750)” were included in this study. One nested case-control set (N = 348; followed up for at least five years) was used to develop and internally validate the predictive model. The clinical, demographic and laboratory features were collected and subjected to logistic regression analyses. Using natural logarithm odds ratio value weighting methods, several risk factors (genetic and non-genetic) were selected to construct predictive models. The final comprehensive prediction model is the arithmetic sum of the two optimal models.

Results: We found that transforming growth factor- β (TGF- β), and asymmetric dimethylarginine (ADMA) were effective biomarkers for CKD, the area under the curve (AUC), specificity and sensitivity for the non-genetic equation were 0.889, 0.851 and 0.770, respectively, in the genetic equation, 0.643, 0.794 and 0.838, respectively, and 0.894, 0.827, 0.801, respectively, in the comprehensive prediction model. After internal verification of Bootstrap, its AUC value was 0.820, indicating it showed a favorable predictive performance.

Conclusions: A comprehensive prediction model was established, which may help early identify individuals who are most likely to develop CKD. Its feasibility in the clinic warrants further investigation.

1. Introduction

Chronic kidney disease (CKD), especially its complications, has posed a serious threat on public health worldwide. According to statistics from the Global Burden of Disease Study[1], in 2017, nearly 697.5 million cases of all-stage CKD were recorded, for a global prevalence of 9.1%, posted an increase of 29.3% from 1990. Meanwhile, the global all-age mortality rate from CKD increased by 41.5% between 1990 and 2017. A cross-sectional study showed that the prevalence of chronic kidney disease in China was about 10.8%[2], which means that there were approximately 119.5 million CKD patients in China.

So far, certain risk factors are highly associated with chronic kidney disease, including age [3], being female[4], obesity[5], and Diabetes Mellitus[6]. Recently, several biomarkers were found that associated with CKD. A few previous researches have shown that the elevated ADMA (asymmetric dimethylarginine) level could cause renal damage[7] [8]. Several studies have pointed that ADMA is a powerful biomarker for predicting CKD mortality[9, 10]. It has also shown that NFAL (neutrophil gelatinase-associated lipocalin) expression levels appear to correlate with the degree of renal dysfunctions, which may help to identify patients at high risk for a more rapid decline in renal function[11]. Furthermore, the decrease of serum CysC (Cystatin C) is correlated with the decrease of eGFR (estimated glomerular filtration rate) concentration[12]. It has been speculated that CysC could be used together with serum creatinine as a new biomarker, or as a substitute for serum creatinine to better identify the occurrence of kidney disease in general population[13, 14]. TGF- β (Transforming growth factor- β) is the main regulator of tubular interstitial fibrosis[15], meanwhile, TGF- β signaling can influence a few important renal injury responses in other growth factor signaling pathways[16, 17], ultimately affecting the onset of CKD[18]. Previous studies have reported that more than 50 single nucleotide polymorphisms (SNPs) were associated with renal function indexes or CKD worldwide[19].

The treatment of chronic kidney disease and renal failure is costly and rarely effective. Whereas, less than 5% of patients with early CKD report awareness of their disease[20]. Once CKD can be diagnosed, glomerular damage has reached over 50% and usually irreversible. Effective prediction of chronic kidney disease can be immensely useful in this aspect. Therefore, several CKD prediction models for different populations were [21–24] introduced. Lately, a research developed equations for predicting CKD based on 34 multinational cohorts[25]. Nevertheless, few models considered both genetic and non-genetic risk factors. Although many prediction models reached high prediction power in a relatively large population, an early prediction (at least when eGFR > 60ml/ (min·1.73m²) is essential for CKD treatment and prevention. In this study, we developed genetic, non-genetic (including biomarkers), and comprehensive risk score prediction models for CKD in a nested case-control study.

2. Methods

2.1. Study design and population

This research was designed as a nested case-control study involving 348 participants from the “Tianjin Medical University Chronic Diseases Cohort”. The cohort was established in 2006, with an initial number of 2,068 people for an annual physical examination. By the end of 2018, a total of 21,750 people had been recruited to the cohort, with the longest follow-up period of 13 years. We had collected demographic markers, laboratory markers, and genotyping results for 110 loci (including 380 cases with genome-wide genotyping data). We screened

patients met the following criteria: (i) with a follow-up period of at least 5 years; (ii) no CKD at the first physical examination; (iii) having blood samples and other important information among whom 1804 were eligible, 116 were selected as the case group and 232 were selected as the control group with gender and age $\pm$ 3 years matching, therefore a total of 348 subjects were included.

This study has been reviewed and approved by the Ethics Committee of Tianjin Medical University, and all participants have signed informed consent.

2.2. Diagnostic criteria

The diagnostic criteria for CKD were eGFR < 60ml/ (min $\cdot$ 1.73m²) or positive proteinuria ($\geq$ 1+). Glomerular filtration rate is estimated using the simplified Chinese MDRD equation[26]. The determination of diabetes mellitus (DM) is based on the diagnostic criteria of diabetes published by the World Health Organization (WHO) in 1999 that fasting plasma glucose $\geq$ 7.0 mmol /L and/or 2 h postprandial glucose $\geq$ 11mmol/L. Obesity is defined as body mass index (BMI) $\geq$ 28kg/m² according to the recommendation of "Guidelines for the Prevention and Control of Overweight and Obesity among Chinese Adults" [27] by the Ministry of Health. Hypertension were defined as systolic blood pressure (SBP) $\geq$ 140mmHg and/or diastolic blood pressure (DBP) $\geq$ 90mmHg or a self-reported history of physician-diagnosed hypertension. The diagnostic criteria for hyperuricemia (HUA) [28] were blood uric acid level $\geq$ 420 μ mol/L in men and $\geq$ 360 μ mol/L in women.

2.3. Measurements of biomarkers

After twelve hours' fasting, participants' venous blood samples were collected into non-anticoagulant blood collection tubes at 7:30-9:00 am, standing at room temperature for half an hour and then centrifuged at 3000 rpm at 4 °C for 10 min to separate serum. The serum was stored at -80 °C before analysis. Levels of fasting plasma glucose, serum creatinine, urea nitrogen, serum uric acid, total cholesterol, triglyceride, alanine aminotransferase, total protein, albumin, globulin, total bilirubin, direct bilirubin were determined using Hitachi automatic biochemical analyzer. Cystatin C (CysC), transforming growth factor beta (TGF- β), asymmetric dimethylarginine (ADMA) and neutrophil gelatinase-associated lipocalin (NGAL) were measured by ELISA (enzyme linked immunosorbent assay) kits (Shanghai Huyu Biotechnology Co., LTD).

2.4. Selection of CKD-related non-genetic/genetic risk factors

We incorporated 21 potential risk factors including several biomarkers into the univariate Cox proportional hazard mode (**Supplementary, Table S3**), and then significant factors were taken as explanatory variables and incorporated into multivariate Cox proportional hazard regression model, finally we got five non-genetic risk factors (**Table 2; Figure 1**)

After obtaining part of the data access of the UK-Biobank database, we used PLINK to perform genome-wide association analysis (GWAS) for renal function related indicators, including eGFR, SCr, and CysC. The results of the GWAS were shown in the Manhattan plot (**Supplementary, Figure S1**). Combined with the results of previous studies, a total of 10 SNP loci on 10 genes were screened (**Supplementary, Table S1**). Meanwhile, after integrating information from GWAS databases, the UCSC Genomic bioinformatics Database, and GWAS results for kidney function related phenotypes in Asia or China [29-31], SNP loci with both high Genotype Relative Risk (GRR) and genome wide polygenetic score (GPS) for CKD were selected. Finally, we selected a total of 27 SNP loci from 24 genes to construct of the genetic risk model for CKD (**Supplementary, Table S2**). The 27 SNPs that selected in this study were genotyped in 348 nested case-control subjects using a matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) platform. Hardy-Weinberg equilibrium (HWE) were checked for all 27 SNPs, we deleted 2 SNPs which failed HWE, therefore, genotyping data for 25 SNPs were documented.

2.5. Developing prediction models

In this study, genetic risk score (GRS) models and non-genetic risk score (NGRS) models were built from the weights of natural logarithms (β) of different risk factors' OR values. The combined effects of each non-genetic or genetic factors were calculated in a weighted way, and the optimal combination method was selected to develop the prediction model of CKD. GRS equation was established based on the different contributions of each candidate SNP site to the pathogenesis of CKD. Each SNP site was considered as a potential risk factor for CKD. Different weights for the contribution to the onset of CKD were determined by different OR (or β) values from logistic regression analysis to establish several combinations and screen for the optimal combination. Using a weighted-genetic risk score(wGRS),

$wGRS = \sum_1^i \beta_i G_i$ (β_i is the weight of the i th SNPs, and G_i is the number of alleles at the i th SNPs, assign a value of 0,1,2), the weight is the natural logarithm of the odds ratio (OR) of SNPs, it could be an estimated effect (β coefficient). For each individual, $wGRS$ is the sum of the number of risk alleles weighted by the OR (β) value of each SNP site in Logistic regression. See Formula (1) for details.

$$GRS = \sum_1^i \beta_i G_i$$

$$\text{Logit}P = \alpha + \beta(GRS)$$

$$= \alpha + \beta \sum_1^i \beta_i G_i \quad (1)$$

In the above formula, to fix the weight in advance, we used the value of log-converted single-risk alleles in studies with large sample sizes and high reliability (e.g., meta-analysis) as the weight in the actual model construction.

The building principle of the non-genetic risk score model is the same as that of the GRS. That is, according to the different contributions of the identified CKD-related non-genetic risk factors (e.g. normal high value of TGF- β , the elderly) to the incidence of CKD, different OR (or β) values of logistic regression analysis are used to determine different weights for the onset of CKD, establish different combinations and select the optimal combination. The weighted non-genetic risk score ($wNGRS$) was used, $wNGRS = \sum_1^i \beta_i S_i$ (β_i is the weight of the i th corresponding non-genetic risk factor in the risk of developing CKD, and S_i is the i th corresponding non-genetic risk factor), the weight β takes the natural logarithm of the OR value obtained by logistic regression analysis of different risk factors. For every individual, the $wNGRS$ is the sum of risk factors weighted by the OR (β) value of different non-genetic risk factors in logistic regression, See Formula (2) for details.

$$NGRS = \sum_1^i \beta_i S_i$$

$$\text{Logit}P = \alpha + \beta(NGRS)$$

$$= \alpha + \beta \sum_1^i \beta_i S_i \quad (2)$$

In the above formula, S represents the set vector of a group of non-genetic risk factors (S_i represents the state of the i th non-genetic risk factors, if the individual has the risk factor, the value is 1; if not, the value is 0). The β value used in this study was the β value of each non-genetic risk factor in Logistic regression analysis.

The construction of the comprehensive risk scoring model integrates the optimal GRS model and the NGRS model, which is the sum of the two models. See formula (3) for details.

$$\text{Logit}P = \alpha + \beta(GRS + NGRS)$$

$$= \alpha + \beta(\sum_1^i \beta_i G_i + \sum_1^i \beta_i S_i) \quad (3)$$

2.6. Prediction model evaluation

The evaluation of the constructed GRS model, NGRS model and comprehensive predictive model adopts the receiver operating characteristic curve (ROC) area under curve (AUC) method. The MedCalc software was used to determine the optimal cut point of the ROC curve and the sensitivity and specificity at the optimal cut point. Finally, the evaluation of the prediction effectiveness of the constructed CKD prediction model is realized. The constructed GRS model, NGRS model and comprehensive prediction model were internally validated in a nesting case-control study using Bootstrap five-fold cross-validation. All data analyses were performed using SPSS21.0 software. Statistical significance was determined with a threshold P value of <0.05 .

3. Results

In this nested case-control study, 348 participants (all had $\text{eGFR} \geq 60 \text{ ml/ (min} \cdot 1.73 \text{ m}^2)$ at baseline) were included (116 cases, 232 controls, subjects who reached $\text{eGFR} < 60 \text{ ml/ (min} \cdot 1.73 \text{ m}^2)$ during the 5-year follow-up were considered as “cases”) to build a 5-year risk prediction model for the onset of CKD. Baseline characteristics of the included participants in the nested case-control study were described in Table 1. The levels of fasting plasma glucose (FPG), total cholesterol (TC), urea nitrogen (BUN), serum creatinine (SCr), total protein (TP), globulin (GLB), systolic blood pressure (SBP), Cystatin C (CysC), transforming growth factor- β (TGF- β), and asymmetric dimethylarginine (ADMA) in the CKD group were significantly higher than those in controls. The age of CKD group was significantly higher than that of the non-CKD group, incidences of type 2 diabetes and hyperuricemia were higher than that of on the non-CKD group (Table 1). In addition, triglyceride (TG), serum uric acid (SUA) and body mass index (BMI) levels in CKD group were higher than those in non-CKD group, but there were not statistically significant.

3.1. Non-genetic risk factors for CKD

Cox proportional risk regression model showed that age, diabetes mellitus, normal high value of urea, normal high value of TGF- β , and ADMA were independent risk factors for CKD (Table 2; **Supplementary, Table S3**). Kaplan-Meier survival analyses showed that the elderly, normal high value of urea nitrogen, normal high value of TGF- β , normal high value of ADMA, and diabetes (we defined age ≥ 60 years as the elderly, took the higher quartile of other measurement data as their normal high values) were significantly associated with chronic kidney disease onsets in our cohort (Fig. 1).

3.2. Non-genetic risk score (NGRS) prediction model for CKD

A total of 5 predictors, including age, diabetes mellitus, normal high value of BUN, normal high value of TGF- β , and ADMA were included in the non-genetic prediction model for CKD. Among four models (**Supplementary Material S1; Table S4; Table S5**) that we constructed based on those 5 risk factors, the model CKDNGRS4 yielded the highest C statistic (0.889; 95% *CI*, 0.851–0.925) and the highest OR value (4.113; 95% *CI*, 3.039–5.566) (Table 3; Fig. 2). The prediction equation was $\text{logitP} = 1.84 \times S1 + 1.137 \times S2 + 0.84 \times S3 + 0.497 \times S4 + 0.603 \times S5$, while $S1 = \text{TGF-}\beta \text{ normal high value (0: } < 1.011 \text{ pg/mL; 1: } \geq 1.011 \text{ pg/mL)}$, $S2 = \text{ADMA normal high value (0: } < 0.019 \mu\text{mol/L; 1: } \geq 0.019 \mu\text{mol/L)}$, $S3 = \text{diabetes (0: unaffected; 1: affected)}$, $S4 = \text{BUN normal high value (0: } < 5.9 \text{ mmol/L; 1: } \geq 5.9 \text{ mmol/L)}$, $S5 = \text{elderly (0: } < 60 \text{ yrs; 1: } \geq 60 \text{ yrs)}$. The sensitivity of the model was 0.851, while the specificity was 0.770.

3.3. Genetic risk score (GRS) prediction model for CKD

By integrating the results of CKD-related genetic loci in the UK-Biobank subjects (**Supplementary, Table S1**) and previous studies, 25 SNPs were analyzed for their correlation with CKD by Logistic regression analysis (**Supplementary Table S7**). Seventeen SNPs were selected in CKD genetic prediction model, including 7 SNPs derived from UK-Biobank. A total of 14 models were constructed (**Supplementary Table S8; Material S2**). Combining the results of regression analysis and survival analysis (**Supplementary Table S9; Table S10**), we found that the CKDGRS14 is the best prediction model. The area under the ROC curve (AUC) of the model was 0.643 (95% *CI*, 0.578–0.709), the sensitivity of the model was 0.794, while the specificity was 0.838. The OR value was 2.363 (95% *CI*, 1.518–3.679) (Table 3; Fig. 2). The genetic risk prediction equation for CKD was

$$\text{logitP} = 0.577 \times \text{rs17319721Gi} + (-0.183) \times \text{rs700233} + (-0.362) \times \text{rs671Gi} + (0.286) \times$$

$$\text{rs11864909Gi} + 1.099 \times \text{rs653178Gi} + 0.255 \times \text{rs3752462Gi} + 0.228 \times \text{rs13146355Gi} + 0.253$$

$$\times \text{rs881858Gi} + (-0.24) \times \text{rs1153849Gi} + (-0.234) \times \text{rs3770636Gi} + (0.178) \times \text{rs504915Gi} +$$

$$0.149 \times \text{rs16853722Gi} + 0.683 \times \text{rs12917707Gi} + (-0.133) \times \text{rs1731274Gi}.$$

3.4. Comprehensive prediction model for CKD

Through analyzing and screening, CKDNGRS4 and CKDGRS14 are optimal non-genetic predictive model and genetic predictive model, respectively. The final comprehensive predictive model is the arithmetic sum of the two model. It was $\text{logitP} = 0.577 \times \text{rs17319721Gi} + (-0.183) \times \text{rs700233} + (-0.362) \times \text{rs671Gi} + (-0.286) \times$
 $\text{rs11864909Gi} + 1.099 \times \text{rs653179Gi} + 0.255 \times \text{rs3752462Gi} + 0.228 \times \text{rs13146355Gi} + 0.253 \times \text{rs881858Gi} + (-0.24) \times \text{rs1153849Gi} +$
 $(-0.234) \times \text{rs3770636Gi} + (-0.178) \times \text{rs504915Gi} +$

$$0.149 \times \text{rs16853722Gi} + 0.683 \times \text{rs12917707Gi} + (-0.133) \times \text{rs1731274Gi} + 1.84 \times S1 + 1.137 \times$$

$S2+0.84\times S3+0.497\times S4+0.603\times S5$. The predictive power of CKD comprehensive prediction model was higher than either of the nongenetic and genetic prediction models: the AUC was 0.894 (95% *CI*, 0.857-0.931), OR was 3.758 (95% *CI*, 2.827-4.997), while the sensitivity was 0.827 and the specificity was 0.801 (Table 3, Figure 2).

3.5. Internal validation

In the nested case-control study, Bootstrap five-fold cross validation was carried out for different prediction models of CKD onset. After verification results were averaged, the AUC values of non-genetic, genetic, and comprehensive prediction models of CKD were 0.786, 0.692, and 0.820, respectively.

4. Discussion

Early prediction of CKD is challenging. Decades of researches showed that diabetic nephropathy, primary glomerulonephritis, hypertension, interstitial nephritis, and polycystic kidney could all induce CKD. The awareness of CKD is notoriously low, once CKD had developed, usually the treatment was limited until last remedies of dialysis and renal transplantations are needed for ESRD. The eGFR is a sensitive indicator of renal function, however, it is not an early predictor of CKD. Although many biomarkers were tested for CKD, reappraisal in large sample sized prospective cohort studies are needed. Seeking an early, sensitive, easy to perform and cost-effective prediction model

We carried out a nested case-control study for CKD prediction out of the "Tianjin Medical University Chronic Disease Cohort"[32, 33], with strong pertinence, facilitated prediction of the 5-year probability of chronic kidney disease onset in this area. The average age of subjects was 63 years, thus, those individuals are more likely to develop CKD than younger subjects.

We combined traditional laboratory indicators, multiple biomarkers that related to renal function, and SNP loci to develop CKD prediction models. In the NGRS model, we not only included some indicators which were used in other studies such as diabetes and age[25, 34, 35], several biomarkers, especially TGF- β and ADMA, were also employed as early CKD predictors in the model.

Although hundreds of associations were found among CKD and susceptibility genes, large sample-sized GWAS also yielded very significant results, seems genetic factors only provided a little improvement of the prediction model. Given a certain SNP, the genetic relative risk (GRR) of could be high, however, its contribution to CKD risks in general population was limited. All 17 SNPs employed in our study were from GWASs out of UK Biobank and other large cohorts, however, the AUC of the genetic risk model (GRS) was only 0.643, had only given a marginal improvement of AUC in the comprehensive model (from 0.889 to 0.894). A study in Japan showed that genetic predictors do not contribute significantly to the improvement of the prediction efficiency of the comprehensive prediction model (29). Although certain SNPs had very significant associations with CKD in large sample sized GWASs (i.e., high genetic relative risk, GRR), their contribution to the phenotype variance might be limited.

Several biomarkers were tested and included in our prediction model. The plasma TGF- β level, alone with ADMA, provided better prediction value than the more direct glomerular filtration indicator cystatin C. In our previous study, we found that TGF- β pathway genes were highly expressed in kidney of very early stage of diabetic nephropathy renal biopsies, long before renal fibrosis and decreased filtration occurred. Indeed, screening early biomarkers before eGFR decreasing may give CKD predictions several years earlier, although early treatment could be another obstacle to overcome.

Recently, numerous predictive models have been established and came into use in the clinic for decision-making. Among them, there exist several models estimating the risk of prevalent and incident CKD[22, 34–37]. However, due to differences in race, lifestyle, and geographic environment, it's still necessary to develop an effective predictive model for chronic kidney disease in different ethnic groups, which can help to identify people with higher CKD risks earlier, thus to improve health care by allocating resources to those individuals who benefit most from it, while preventing the potential abusing of health care resources by individuals who at low risks.

This study has a few limitations. First, the research on CKD-related biomarkers was carried out in a nested case-control study that selected from a cohort of chronic diseases, the sample size was relatively small, therefore results from the study may had certain deviations; Second, our risk prediction model only focused on the onset of chronic kidney disease, but did not assess the progression of chronic kidney disease to renal failure or other complications; Third, participants who make up the "Tianjin Medical University Chronic Disease Cohort" were mostly teachers and government employees that worked in urban area. This group of people were more self-disciplined and paid more attention to health. Whether our prediction model could be applied to other groups of people needs more external validations. Our future studies will detect more renal function related biomarkers in larger cohorts in order to validating and improving the prediction model for CKD.

5. Conclusions

Age, diabetes, normal high value of creatinine, TGF- β and ADMA are independent indicators for CKD. Besides, an integrated predictive model was also established to evaluate individuals' risk of CKD at an early stage, therefore, early and appropriate intervention can be exerted to avoid getting worse and even irreversible.

Abbreviations

CKD: chronic kidney disease;

TGF- β :transforming growth factor- β

ADMA: asymmetric dimethylarginine;

AUC: the area under the curve;

NFAL: neutrophil gelatinase-associated lipocalin;

CysC: Cystatin C;

eGFR: estimated glomerular filtration rate;

SNP: single nucleotide polymorphisms;

DM: diabetes mellitus;

WHO: World Health Organization;

BMI: body mass index;

SBP: systolic blood pressure;

HUA: hyperuricemia;

NGAL: neutrophil gelatinase-associated lipocalin;

ELISA: enzyme linked immunosorbent assay;

GWAS: genome-wide association analysis;

GRR: genotype relative risk;

GPS: genome wide polygenetic score;

Declarations

Ethics approval and consent to participate

This study has been reviewed and approved by the Ethics Committee of Tianjin Medical University, and all participants have signed informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by National Natural Science Foundation of China (92046014, 91746205), and Beijing-Tianjin-Hebei Jointed Research Program (19JCZDJC64700).

Authors' contributions

Y.Z., J.Z., J.Q., X.Z. researched data and wrote the manuscript. W-D.L. designed the study, researched data and edited the manuscript. Y.W. designed the study and contributed to discussion. F.W. and J.F. researched data. S.F., C.C., K.Z. collected clinical data and samples.

Acknowledgements

We thank all subjects who took part in this study. This work was supported by National Natural Science Foundation of China (92046014, 91746205), and Beijing-Tianjin-Hebei Jointed Research Program (19JCZDJC64700).

All authors have reviewed the manuscript.

Availability of data and material

All data from the “Tianjin Medical University Chronic Diseases Cohort” are available upon request. Part of genotyping data used in this study is available from UK Biobank.

References

1. Bikbov B, Purcell CA, Levey AS, Smith M, Abdoli A, Abebe M, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*. 2020;395(10225):709–33. doi:10.1016/S0140-6736(20)30045-3.
2. Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, et al. Prevalence of chronic kidney disease in china: a cross-sectional survey. *The Lancet*. 2012;379(9818):815–22. doi:https://doi.org/10.1016/S0140-6736(12)60033-6.
3. O Sullivan ED, Hughes J, Ferenbach DA. Renal aging: causes and consequences. *J Am Soc Nephrol*. 2017;28(2):407–20. doi:10.1681/asn.2015121308.
4. Carrero JJ, Hecking M, Chesnaye NC, Jager KJ. Sex and gender disparities in the epidemiology and outcomes of chronic kidney disease. *Nat Rev Nephrol*. 2018;14(3):151–64. doi:10.1038/nrneph.2017.181.
5. Yun HR, Kim H, Park JT, Chang TI, Yoo TH, Kang SW, et al. Obesity, metabolic abnormality, and progression of ckd. *Am J Kidney Dis*. 2018;72(3):400–10. doi:10.1053/j.ajkd.2018.02.362.
6. Tsai W, Wu H, Peng Y, Ko M, Wu M, Hung K, et al. Risk factors for development and progression of chronic kidney disease. *Medicine*. 2016;95(11):e3013. doi:10.1097/MD.0000000000003013.
7. Fliser D, Kronenberg F, Kielstein JT, Morath C, Bode-Boger SM, Haller H, et al. Asymmetric dimethylarginine and progression of chronic kidney disease: the mild to moderate kidney disease study. *J Am Soc Nephrol*. 2005;16(8):2456–61. doi:10.1681/ASN.2005020179.
8. Ravani P, Tripepi G, Malberti F, Testa S, Mallamaci F, Zoccali C. Asymmetrical dimethylarginine predicts progression to dialysis and death in patients with chronic kidney disease: a competing risks modeling approach. *J Am Soc Nephrol*. 2005;16(8):2449–55. doi:10.1681/ASN.2005010076.
9. Endogenous nitric oxide synthase inhibitor. a novel marker of atherosclerosis.
10. Tripepi G, Mattace RF, Sijbrands E, Seck MS, Maas R, Boger R, et al. Inflammation and asymmetric dimethylarginine for predicting death and cardiovascular events in esrd patients. *Clin J Am Soc Nephrol*. 2011;6(7):1714–21. doi:10.2215/CJN.11291210.
11. Viau A, El KK, Laouari D, Burtin M, Nguyen C, Mori K, et al. Lipocalin 2 is essential for chronic kidney disease progression in mice and humans. *J Clin Invest*. 2010;120(11):4065–76. doi:10.1172/JCI42004.
12. Ferguson MA, Waikar SS. Established and emerging markers of kidney function. *Clin Chem*. 2012;58(4):680–9. doi:10.1373/clinchem.2011.167494.
13. Filler G, Bökenkamp A, Hofmann W, Le Bricon T, Martínez-Brú C, Grubb A. Cystatin c as a marker of gfr—history, indications, and future research. *Clin Biochem*. 2005;38(1):1–8. doi:10.1016/j.clinbiochem.2004.09.025.
14. Ferguson TW, Komenda P, Tangri N. Cystatin c as a biomarker for estimating glomerular filtration rate. *Curr Opin Nephrol Hy*. 2015;24(3):295–300. doi:10.1097/MNH.0000000000000115.
15. Meng X, Nikolic-Paterson DJ, Lan HY. Tgf- β : the master regulator of fibrosis. *Nat Rev Nephrol*. 2016;12(6):325–38. doi:10.1038/nrneph.2016.48.
16. Tazat K, Hector-Greene M, Globe GC, Henis YI. Tbetariii independently binds type i and type ii tgf-beta receptors to inhibit tgf-beta signaling. *Mol Biol Cell*. 2015;26(19):3535–45. doi:10.1091/mbc.E15-04-0203.

17. Wang S, Wilkes MC, Leof EB, Hirschberg R. Noncanonical tgf- β pathways, mtorc1 and abl, in renal interstitial fibrogenesis. *Am J Physiol-Renal*. 2010;298(1):F142–9. doi:10.1152/ajprenal.00320.2009.
18. Nlandu-Khodo S, Neelisetty S, Phillips M, Manolopoulou M, Bhavé G, May L, et al. Blocking tgf-beta and beta-catenin epithelial crosstalk exacerbates ckd. *J Am Soc Nephrol*. 2017;28(12):3490–503. doi:10.1681/ASN.2016121351.
19. Genome-wide association studies in kidney transplantation advantages and constraints.
20. Chen TK, Knicely DH, Grams ME. Chronic kidney disease diagnosis and management. *JAMA*. 2019;322(13):1294. doi:10.1001/jama.2019.14745.
21. Carrillo-Larco RM, Miranda JJ, Gilman RH, Medina-Lezama J, Chirinos-Pacheco JA, Muñoz-Retamozo PV, et al. Risk score for first-screening of prevalent undiagnosed chronic kidney disease in peru: the cronicas-ckd risk score. *Bmc Nephrol* 2017;18(1). doi:10.1186/s12882-017-0758-4.
22. Tangri N, Inker LA, Hiebert B, Wong J, Naimark D, Kent D, et al. A dynamic predictive model for progression of ckd. *Am J Kidney Dis*. 2017;69(4):514–20. doi:10.1053/j.ajkd.2016.07.030.
23. Ma J, Yang Q, Hwang S, Fox CS, Chu AY. Genetic risk score and risk of stage 3 chronic kidney disease. *Bmc Nephrol* 2017;18(1). doi:10.1186/s12882-017-0439-3.
24. Imai E, Horio M, Watanabe T, Iseki K, Yamagata K, Hara S, et al. Prevalence of chronic kidney disease in the Japanese general population. *Clin Exp Nephrol*. 2009;13(6):621–30. doi:10.1007/s10157-009-0199-x.
25. Nelson RG, Grams ME, Ballew SH, Sang Y, Azizi F, Chadban SJ, et al. Development of risk prediction equations for incident chronic kidney disease. *JAMA*. 2019;322(21):2104. doi:10.1001/jama.2019.17379.
26. Ma YC, Zuo L, Chen JH, Luo Q, Yu XQ, Li Y, et al. Improved gfr estimation by combined creatinine and cystatin c measurements. *Kidney Int*. 2007;72(12):1535–42. doi:https://doi.org/10.1038/sj.ki.5002566.
27. Qasim A, Turcotte M, de Souza RJ, Samaan MC, Champredon D, Dushoff J, et al. On the origin of obesity: identifying the biological, environmental and cultural drivers of genetic risk among human populations. *Obes Rev*. 2018;19(2):121–49. doi:10.1111/obr.12625.
28. Bardin T, Richette P. Definition of hyperuricemia and gouty conditions. *Curr Opin Rheumatol*. 2014;26(2):186–91. doi:10.1097/BOR.000000000000028.
29. Okada Y, Sim X, Go MJ, Wu J, Gu D, Takeuchi F, et al. Meta-analysis identifies multiple loci associated with kidney function–related traits in east Asian populations. *Nat Genet*. 2012;44(8):904–9. doi:10.1038/ng.2352.
30. O'Seaghdha CM, Fox CS. Genome-wide association studies of chronic kidney disease: what have we learned? *Nat Rev Nephrol*. 2011;8(2):89–99. doi:10.1038/nrneph.2011.189.
31. Ried JS, Jeff MJ, Chu AY, Bragg-Gresham JL, van Dongen J, Huffman JE, et al. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. *Nat Commun* 2016;7(1). doi:10.1038/ncomms13357.
32. Zhao J, Zhang Y, Wei F, Song J, Cao Z, Chen C, et al. Triglyceride is an independent predictor of type 2 diabetes among middle-aged and older adults: a prospective study with 8-year follow-ups in two cohorts. *J Transl Med* 2019;17(1). doi:10.1186/s12967-019-02156-3.
33. Wei F, Sun N, Cai C, Feng S, Tian J, Shi W, et al. Associations between serum uric acid and the incidence of hypertension: a Chinese senior dynamic cohort study. *J Transl Med* 2016;14(1). doi:10.1186/s12967-016-0866-0.
34. Chien K, Lin H, Lee B, Hsu H, Lee Y, Chen M. A prediction model for the risk of incident chronic kidney disease. *The American Journal of Medicine*. 2010;123(9):836–46.e2. doi:10.1016/j.amjmed.2010.05.010.
35. O'Seaghdha CM, Lyass A, Massaro JM, Meigs JB, Coresh J, D'Agostino RB, et al. A risk score for chronic kidney disease in the general population. *The American Journal of Medicine*. 2012;125(3):270–7. doi:10.1016/j.amjmed.2011.09.009.
36. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. *Plos Med*. 2012;9(11):e1001344. doi:10.1371/journal.pmed.1001344.
37. Fujii R, Hishida A, Nakatochi M, Furusyo N, Murata M, Tanaka K, et al. Association of genetic risk score and chronic kidney disease in a Japanese population. *Nephrology*. 2019;24(6):670–3. doi:10.1111/nep.13479.

Tables

Table 1. Baseline characteristics of subjects in the nested case-control study

	Total (n=348)	CKD group (n=116)	Non-CKD group (n=132)	P value
Men(%)	260(74.7%)	84(70%)	176(77.2%)	<0.001
Age(year)	63.27±10.09	63.96±7.74	63.33±7.14	0.947
FPG(mmol/L)	5.11±1.04	5.28±1.13	5.02±0.98	<0.001
TC(mmol/L)	5.04±0.88	5.22±0.85	4.95±0.89	0.006
TG(mmol/L)	1.75±1.34	1.81±1.65	1.72±1.15	0.508
BUN(mmol/L)	5.45±1.11	5.94±1.11	5.20±1.03	<0.001
SCr(μmmol/L)	83.11±14.32	89.82±14.39	79.58±12.99	0.009
SUA(μmmol/L)	335.6±76.33	344.9±81.27	330.8±73.40	0.100
TP(g/L)	75.41±4.30	75.94±5.17	75.13±3.74	0.005
ALB(g/L)	45.74±2.62	45.57±2.69	45.84±2.59	0.938
GLB(g/L)	29.68±3.41	30.38±3.97	29.32±3.01	0.003
ALT(IU/L)	25.93±3.30	24.44±10.39	26.71±8.60	0.030
TBIL(μmol/L)	14.38±4.63	13.52±4.12	14.83±4.83	0.013
DBIL(μmol/L)	2.30±1.25	2.50±1.35	2.19±1.19	0.037
BMI(kg/m ²)	24.63±3.16	25.07±3.54	24.39±2.91	0.057
SBP(mmHg)	138.8±19.72	147.7±19.65	134.2±18.14	<0.001
DBP(mmHg)	77.42±12.45	77.10±12.76	77.59±12.31	0.729
Hypertension(%)	64(18.4%)	21(17.5%)	43(18.9%)	0.669
Type II diabetes(%)	19(5.5%)	10(8.3%)	9(3.9%)	<0.001
Hyperuricemia(%)	52(14.9%)	23(19.2%)	29(12.7%)	0.038
CysC(mg/L)	1.079±0.64	1.32±0.99	0.95±0.76	<0.001
TGF-β(pg/mL)	13.23±5.18	17.70±3.22	10.88±4.41	<0.001
ADMA(μmol/L)	101.1±64.80	118.6±46.47	91.89±70.99	0.004
NGAL(μg/L)	16.55±7.31	14.88±7.72	17.43±6.95	0.087

Abbreviations: FPG, fasting plasma glucose; TC, total cholesterol; TG, triglyceride; BUN, urea nitrogen; SCr, serum Creatinine; SUA, serum uric acid; TP, total protein; ALB, albumin; GLB, globulin; ALT, alanine aminotransferase; TBIL, total bilirubin; DBIL, direct bilirubin; BMI, body Mass Index; SBP, systolic blood pressure; DBP, diastolic blood pressure; CysC, Cystatin C; TGF-β, transforming growth factor-β; ADMA, asymmetric dimethylarginine; NGAL, neutrophil gelatinase-associated lipocalin.

Data are expressed as mean SD, percentage (number), or median (interquartile range); t-test or Mann–Whitney rank sum test was used for the continuous variables.

Table 2. Non-genetic multivariate Cox regression analyses and non-genetic risk models (NGRS)

Variables	β	SE	χ^2	HR	95%CI	P-value	NGRS model	OR (model)	95%CI (model)	P-value (model)
Women	0.216	0.207	1.094	1.241	0.828-1.861	0.296				
Normal high value of TGF- β^a	0.945	0.195	3.553	2.572	1.756-3.766	<0.001				
							1. Normal high value of TGF- β , Normal high value of ADMA	3.634	2.723-4.850	<0.001
Normal high value of ADMA ^b	1.222	0.244	4.999	3.394	2.102-5.479	<0.001				
							2. Normal high value of TGF- β , Normal high value of ADMA, Diabetes	3.703	2.775-4.942	<0.001
Diabetes	0.742	0.272	7.450	2.100	1.233-3.578	0.006				
							3. Normal high value of TGF- β , Normal high value of ADMA, Diabetes, Normal high value of BUN	3.917	2.910-5.273	<0.001
Normal high value of BUN ^c	0.693	0.197	12.335	2.00	1.359-2.946	<0.001				
							4. Normal high value of TGF- β , Normal high value of ADMA, Diabetes, Normal high value of BUN, The elderly	4.113	3.039-5.566	<0.001
The elderly ^d	1.055	0.256	16.940	2.872	1.738-4.746	<0.001				

TGF-β, Transforming growth factor-β; ADMA, Asymmetric dimethylarginine; BUN, Urea nitrogen; NGRS, Non-genetic risk score; HR, hazard ratio; *CI*, confidence interval.

^aDefined as the serum concentration of TGF-β≥1.011pg/ml

^bDefined as the serum concentration of ADMA≥0.019μmol/L

^cDefined as the serum concentration of BUN≥5.9mmol/L

^dDefined as the age of the participants≥60years

Table 3. Logistic regression analysis and prediction power comparison of non-genetic (NGRS), genetic (GRS), and comprehensive models for CKD

Models	Logistic regression analysis			ROC curve		
	OR	95% <i>CI</i>	<i>P</i> -value	AUC	95% <i>CI</i>	<i>P</i> -value
NGRS4 ^a	4.113	3.039-5.566	<0.001	0.889	0.851-0.925	<0.001
GRS14 ^b	2.363	1.518-3.679	<0.001	0.643	0.578-0.709	<0.001
Comprehensive model ^c	3.758	2.827-4.997	<0.001	0.894	0.857-0.931	<0.001

ROC, receiver operating characteristic; OR, odds ratio; *CI*, confidence interval; AUC, area under curve; NGRS4, non-genetic risk score model 4; GRS14, genetic risk score model 14.

^aNGRS4=1.84×S1+1.137×S2+0.84×S3+0.497×S4+0.603×S5 (S_i represents the state of the *i*th non-genetic risk factors, if the individual has the risk factor, the value is 1; if not, the value is 0. S1=TGF-β normal high value(0:<1.011pg/mL; 1:1.011pg/mL), S2=ADMA normal high value(0:<0.019μmol/L; 1:≥0.019μmol/L), S3=diabetes(0:unaffected; 1:affected), S4=BUN normal high value(0:<5.9mmol/L; 1:≥5.9mmol/L), S5=elderly(0:<60 yrs; 1:≥60 yrs).

^bGRS14=0.577×rs17319721Gi+(-0.183)×rs700233Gi+(-0.362)×rs671Gi+(-0.286)×rs11864909Gi+1.099×rs653178Gi+0.255×rs3752462Gi+0.228×rs13146355Gi+0.253×rs881858Gi+(-0.24)×rs1153849Gi+(-0.234)×rs3770636Gi+(-0.178)×rs504915Gi+0.149×rs16853722Gi+0.683×rs12917707Gi+(-0.133)×rs1731274Gi (Gi is the number of alleles at the *i*th SNPs, assign a value of 0,1,2).

^cComprehensive model=NGRS4+GRS14

Figures

Figure 1

Kaplan-Meier survival curve of CKD cumulative incidence in 348 subjects of the nested case-control study. (a) The elderly; (b) Normal high value of urea nitrogen (BUN); (c) Normal high value of transforming growth factor- β (TGF- β); (d) Normal high value of asymmetric dimethylarginine (ADMA); (e) Diabetes.

Figure 2

ROC curves of the non-genetic (NGRS4), genetic (GRS14), and comprehensive model for CKD prediction. NGRS4: The No.4 non-genetic risk score model; GRS14: The No.14 non-genetic risk score model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplement05.10.2021.docx](#)