JASN Journal of the American Society of Nephrology

Authors: Han, Miyeun; Moon, Sungji; Lee, Sangjun; Kim, Kyungsik; An, Woo Ju; Ryu, Hyunjin; Kang, Eunjeong; Ahn, Jung-Hyuck; Sung, Hye Youn; Park, Yong Seek; Lee, Seung Eun; Lee, Sang-Ho; Jeong, Kyung Hwan; Ahn, Curie; Kelly, Tanika N.; Hsu, Jesse Y.; Feldman, Harold I.; Park, Sue K.; Oh, Kook-Hwan

Title: Novel Genetic Variants Associated with Chronic Kidney Disease Progression

Running title: GWAS in eGFR decline

Manuscript Type: Original Article -Clinical Research

Manuscript Category:

Funders: This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (2017M3A9E4044649) and a grant of the MD-PhD/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea.

Financial Disclosure: The authors have nothing to disclose

Study Group/Organization Name:

Study Group Members' Names:

Total number of words: 3414

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Abstract:

Background: The incidence of chronic kidney disease (CKD) is associated with genetic factors. However, genetic markers associated with the progression of CKD have not been fully elucidated.

Methods: We conducted a genome-wide association study among 1,738 CKD patients, mainly from the KoreaN Cohort Study for Outcomes in Patients With Chronic Kidney Disease (KNOW-CKD). The outcome was estimated glomerular filtration rate (eGFR) slope. We performed a replication study for discovered SNPs with $P < 10\text{-}6$ in 2,498 CKD patients from the Chronic Renal Insufficiency Cohort (CRIC) study. Several expression quantitative trait loci (eQTL) studies, pathway enrichment analyses, exploration of epigenetic architecture, and predicting disruption of transcription factor binding sites explored potential biological implications of the loci. We developed and evaluated the effect of polygenic risk scores (PRS) on incident CKD outcomes.

Results: SNPs in two novel loci, *TPPP* and *FAT1-LINC02374,* were replicated (rs59402340 in *TPPP, P*discovery = 7.11×10-7, *P*CRIC = 8.13×10-4, *P*meta = 7.23×10-8; rs28629773 in *FAT1-LINC02374, P*discovery = 6.08×10-7, *P*CRIC =4.33×10-2, *P*meta = 1.87×10-7). The eQTL studies revealed that the replicated SNPs regulated the expression level of nearby genes associated with kidney function. Furthermore, these SNPs were near gene enhancer regions and predicted to disrupt the binding of transcription factors. PRS based on the independently significant top 22 SNPs were significantly associated with CKD outcomes.

Conclusion: This study demonstrates that SNP markers in the *TPPP* and *FAT1-LINC02374* loci could be predictive markers for the decline of eGFR in CKD patients. Copyright © 2023 by the American Society of Nephrology

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Significance Statement

Estimated glomerular filtration rate (eGFR) slope has been used as a surrogate outcome for progression of CKD. However, genetic markers associated with eGFR slope among CKD patients were unknown. We aimed to identify genetic susceptibility loci associated with eGFR slope. A two-phase genome wide association study identified SNPs in *TPPP* and *FAT1-LINC02374, and 22 of them were used to derive polygenic risk scores that mark the* decline of eGF by disrupting binding of nearby transcription factors. This work is the first to identify the impact of *TPPP and FAT1-LINC02374* on CKD progression, providing predictive markers for the decline of eGFR in CKD patients.

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Journal of the American Society of Nephrology Publish Ahead of Print
DOI: 1681/ASN.0000000000000066

Novel Genetic Variants Associated with Chronic Kidney Disease Progression

Miyeun Han[1†], Sungji Moon[2,3,4,5†], Sangjun Lee[2,3,6], Kyungsik Kim[2,3,6], Woo Ju An[2,3,7], Hyunjin Ryu[8], Eunjeong Kang[9], Jung-Hyuck Ahn[10], Hye Youn Sung[10], Yong Seek Park[11], Seung Eun Lee[11], Sang-Ho Lee[12], Kyung Hwan Jeong[12], Curie Ahn[1,8], Tanika N. Kelly[13], Jesse Y. Hsu[14], Harold I. Feldman[14], Sue K. Park[2,3,7*], Kook-Hwan Oh[8*]

[1]Department of Internal Medicine, National Medical Center, Seoul, Korea

[2]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

[3]Cancer Research Institute, Seoul National University, Seoul, Korea

[4]Interdisciplinary Program in Cancer Biology, Seoul National University College of Medicine, Seoul, Korea

[5]Genomic Medicine Institute, Seoul National University College of Medicine, Seoul, Korea

[6]Department of Biomedical Science, Seoul National University Graduate School, Seoul, Korea

[7]Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, Seoul, Korea

[8]Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea

[9]Department of Internal Medicine, Ewha Womans University Seoul Hospital, Ewha Womans University College of Medicine, Seoul, Korea

[10]Department of Biochemistry, Ewha Womans University College of Medicine, Seoul, Korea

[11]Department of Microbiology, School of Medicine, Kyung Hee University, Seoul, Korea

Copyright © 2023 by the American Society of Nephrology

[12]Department of Internal Medicine, Kyung Hee University School of Medicine, Seoul, Korea

[13]Department of Epidemiology, Tulane University, New Orleans, Louisiana, USA

[14]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA

[†]These authors contributed equally to this work

Short running title: GWAS in eGFR decline

Word count for abstract: 241

Word count for main text: 3414

Correspondence to: Kook-Hwan Oh, M.D., Ph.D. and Sue K. Park, M.D., MPH., Ph.D.

Kook-Hwan Oh, M.D., Ph.D.

Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea, 103, Daehak-ro, Jongro-gu, Seoul, 03080, Korea

Telephone: 82-2-2072-0776; Fax: 82-2-741-4876, E-mail: khoh@snu.ac.kr

Sue K. Park, M.D., MPH., Ph.D. Department of Preventive Medicine, Seoul National University College of Medicine, 103, Daehak-ro, Jongro-gu, Seoul, 03080, Korea

Telephone: 82-2-740-8338; Fax: 82-2-747-4830, E-mail: suepark@snu.ac.kr

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Abstract

Background: The incidence of chronic kidney disease (CKD) is associated with genetic factors. However, genetic markers associated with the progression of CKD have not been fully elucidated.

Methods: We conducted a genome-wide association study among 1,738 CKD patients, mainly from the KoreaN Cohort Study for Outcomes in Patients With Chronic Kidney Disease (KNOW-CKD). The outcome was estimated glomerular filtration rate (eGFR) slope. We performed a replication study for discovered SNPs with $P < 10^{-6}$ in 2,498 CKD patients from the Chronic Renal Insufficiency Cohort (CRIC) study. Several expression quantitative trait loci (eQTL) studies, pathway enrichment analyses, exploration of epigenetic architecture, and predicting disruption of transcription factor binding sites explored potential biological implications of the loci. We developed and evaluated the effect of polygenic risk scores (PRS) on incident CKD outcomes.

Results: SNPs in two novel loci, *TPPP* and *FAT1-LINC02374,* were replicated (rs59402340 in *TPPP,* $P_{discovery} = 7.11\times10^{-7}$, $P_{CRIC} = 8.13\times10^{-4}$, $P_{meta} = 7.23\times10^{-8}$; rs28629773 in *FAT1-LINC02374*, $P_{discovery} = 6.08\times10^{-7}$, $P_{CRIC} = 4.33\times10^{-2}$, $P_{meta} = 1.87\times10^{-7}$). The eQTL studies revealed that the replicated SNPs regulated the expression level of nearby genes associated with kidney function. Furthermore, these SNPs were near gene enhancer regions and predicted to disrupt the binding of transcription factors. PRS based on the independently significant top 22 SNPs were significantly associated with CKD outcomes.

Conclusion: This study demonstrates that SNP markers in the *TPPP* and *FAT1-LINC02374* loci could be predictive markers for the decline of eGFR in CKD patients.

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Introduction

Chronic kidney disease (CKD) is a public health problem that is growing rapidly, affecting more than 10% of adults worldwide[1]. CKD can progress to end stage kidney disease (ESKD), necessitating dialysis or transplantation, and increases the risk of cardiovascular morbidity and mortality. The clinical course of CKD progression is variable. Though CKD usually gets worse over time, some patients show stable kidney function, or even improvement in kidney function over time[2, 3]. The slope of estimated glomerular filtration rate (eGFR) represents the annual change in eGFR and is associated with future risk of ESKD disease[4]. The eGFR slope is a validated surrogate endpoint of renal outcome in randomized controlled trials of CKD[5], and the current treatment strategy is alleviating eGFR slope[6]. CKD progression is influenced by several factors, such as old age, ethnicity, baseline eGFR, proteinuria, blood pressure, and use of angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker medication affect the progression of CKD[7, 8]. Genetic factors could also affect renal disease progression. For instance, genetic variants in the *APOL1* increase the risk of many different types of kidney disease in black individuals, such as hypertension-attributed ESKD or focal segmental glomerulosclerosis[9].

Genome-wide association studies (GWASs) are important for mapping the risk loci for complex diseases by investigating the associations of genetic markers across the genome with the disease or trait of interest[10]. Using GWASs, several genetic variants related to baseline eGFR were identified recently[11, 12] and were confirmed through multiple experimental follow-up projects[13, 14]. However, few studies have evaluated the genetic causes of progression to ESKD. Although the Chronic Renal Insufficiency Cohort (CRIC) Study discovered that a single nucleotide polymorphism (SNP) in *LINC00923* is associated with CKD progression in individuals with nondiabetic CKD, it was limited by lack of replication in equivalent cohorts[15].

Copyright © 2023 by the American Society of Nephrology

In South Korea, there are high incidence and prevalence of ESKD, although the prevalence of CKD is relatively low[1]. Indeed, Asian patients show a higher tendency to progress to ESKD compared with other racial groups[16, 17]. Therefore, we need genetic studies on progression or exacerbation of CKD, especially in Asian population.

Here, we present a GWAS on the change in eGFR among participants with established CKD in Korea and trans-ethnic replication analysis from another independent cohort, CRIC, to improve our understanding of the genetic contribution to CKD progression. We identified two novel loci, and subsequently validated them from CKD Genetics (CKDGen) Consortium studies regarding kidney associated traits. We explored potential biological implications of the loci through prioritizing genes from expression quantitative trait loci (eQTL) studies and conducted functional enrichment analysis for the candidate genes. Furthermore, we explored epigenetic architectures around the loci and predicted their disruptiveness of transcription factor binding site to understand their regulatory function. Additionally we conducted several polygenic risk score analyses regarding CKD progression with this CKD cohort (Fig. 1).**Results**

Characteristics of study participants

The clinical and demographic characteristics of the 1,738 patients in the discovery cohort are shown in Table 1. The mean age was 54.9 years, and 63.3% of patients were male. 44.3% had diabetes, and 96.3% had hypertension. The mean eGFR was 48.9 ± 28.5 mL/min/1.73 m^2. We divided the patients into diabetes mellitus (DM) (N = 771) and non-diabetes mellitus groups (non-DM) (N = 967). The patients with DM were older and had higher systolic blood pressure, higher body mass index, lower hemoglobin, lower albumin, and higher proteinuria. Patients with DM showed lower baseline eGFR and steeper eGFR decline compared with patients in the non-DM group during a median follow-up (interquartile range) of 3.3 (2.0–5.1) years (Supplementary Fig. 1).

Copyright © 2023 by the American Society of Nephrology

Discovery GWAS and replication study

We conducted three GWASs for 7,734,192 SNPs on eGFR slope (Total, DM, and non-DM group), which showed no genomic inflation (lambda values of 0.98 for total group, 1.00 for DM group, and 0.99 for non-DM group, Supplementary Fig. 2–3). Fifteen SNPs from five loci in the total group, 39 SNPs from 10 loci in the DM group, and 21 SNPs from eight loci in the non-DM group reached $P < 10^{-6}$ (Fig. 2, Supplementary Figure 4, Supplementary Table 1–3). One SNP (rs201948830) within *ADGRV1* showed the top significant signal with a *P*-value of 5.24×10^{-9} in all the GWAS results.

We performed replication analyses of the selected SNPs with $P < 10^{-6}$ associated with eGFR slope in CRIC study composed of CKD patients (N=2,498) of 1,339 European (511 patients with DM and 828 patients with non-DM) and 1,159 African descent (576 patients with DM and 583 patients with non-DM)[15]. Six loci with 18 SNPs fulfilled the criteria for candidate loci from the replication analyses ($P <$ 0.05, consistent direction of effect, and combined $P <$ discovery P) (Supplementary Table 1–3, Supplementary Fig. 5). Of these, two loci at an intronic region of *TPPP* (5p15.33) and an intergenic region between *FAT1* and *LINC02374* (4q35.2) were eventually replicated with minor allele frequency (MAF) ≥ 1% in CRIC study. One SNP marker, rs59402340, within *TPPP* from non-DM GWAS had the strongest GWAS signal in CRIC study (European American, non-DM group) among replicated SNPs for eGFR slope ($\beta_{discovery}$ (Standard error, SE) = -0.92 (0.18), $P_{discovery} = 7.11\times10^{-7}$; β_{CRIC} (SE) = -0.35 (0.10), $P_{CRIC} = 8.13\times10^{-4}$; $\beta_{combined}$ (SE) = -0.49 (0.09), $P_{combined} = 7.23\times10^{-8}$) (Table 2, Supplementary Table 2). One SNP in *FAT1-LINC02374* loci from GWAS conducted in total group was also replicated from African American with non-DM of CRIC study (rs28629773; $\beta_{discovery}$ (SE) = 0.2 (0.10), $P_{discovery} = 6.08\times10^{-7}$; β_{CRIC} (SE) = 0.37 (0.07), $P_{CRIC} = 4.33\times10^{-2}$; $\beta_{combined}$ (SE) = 0.31 (0.06), $P_{combined} = 1.87\times10^{-7}$) (Table 2 and Supplementary Table 4).

The subsequent validation studies for *TPPP* and *FAT1-LINC02374* loci were based on CKDGen studies (longitudinal GWAS for kidney function decline, N = 63,558; cross-sectional GWAS for eGFR and blood urea nitrogen (BUN), N=765,348)[18, 19]. Of the significant SNPs within *TPPP* from discovery

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

GWAS, rs35914637 was validated (eGFR trait, $P_{discovery}=4.59\times10^{-7}$, $P_{CKDGen}=1.17\times10^{-3}$) (Table 3, Supplementary Table 5). Six SNPs near *FAT1* locus were validated by cross-sectional GWAS and longitudinal GWAS in CKDGen (rs4861741 for BUN, $P_{discovery} = 4.91\times10^{-7}$, $P_{CKDGen} = 1.76\times10^{-2}$; rs6553037 for eGFR change, $P_{discovery} = 6.03\times10^{-7}$, $P_{CKDGen} = 2.16\times10^{-2}$; Table 3, Supplementary Table 1). Of note, rs28629773 was significant SNP in both CRIC and CKDGen ($P_{discovery} = 6.08\times10^{-7}$, $P_{CRIC} = 4.33\times10^{-2}$, $P_{CKDGen} = 2.63\times10^{-2}$; Supplementary Table 4–5).

The linkage disequilibrium (LD) plots for the SNPs on each locus across the multiple ethnicities are shown in Supplementary Fig. 6–7. We examined the SNPs on *TPPP* and previously reported SNPs nearby *TPPP* associated with baseline eGFR[20, 21], which showed low r^2 with each other (Supplementary Fig. 4). We additionally performed statistical fine-mapping using susieR[22] for two loci with each discovery GWAS summary statistics. Five SNPs in credible sets showed highest posterior inclusion probability (PIP) of 1.00 in *FAT1-LINC02374*, however, no SNPs were captured in a credible set in *TPPP* (Supplementary Table 6, Supplementary Fig. 8).

Association with secondary renal outcomes

We secondarily looked at the association between replicated loci and secondary renal outcomes such as 30% reduction of eGFR in 3 years, renal event, and renal replacement therapy (RRT) (Fig. 3. and Supplementary Table 5). Regional plots for each outcome showed significant signals on each locus, especially around *TPPP* (Fig. 3A). Rs59402340 in *TPPP* showed association with 30% reduction of eGFR in 3 years (OR=2.00, $P=7.12\times10^{-4}$), renal event (OR=2.28, $P=9.31\times10^{-5}$), and RRT (OR=2.41, $P=5.38\times10^{-5}$). Rs6553037 in *FAT1-LINC02374* showed association with 30% reduction of eGFR in 3 years (OR=1.22, $P =6.79\times10^{-3}$), renal event (OR=1.30, $P=6.31\times10^{-4}$), and renal replacement therapy (OR=1.23, $P=0.01$).

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Prioritizing genes from eQTL studies and functional enrichment analysis

We conducted a series of post-GWAS analyses for finding functional relevance of loci we found (Fig. 4). First, we derived eQTL association *P*-values for SNPs in *TPPP* and *FAT1-LINC02374* from NEPTUNE cohort (tubulointerstitium and glomerulus)[23] and GTEx (cortex of kidney, whole blood, and cultured fibroblast)[24] eQTL summary statistics. With the nominal *P*-value cut-off of 0.05, twenty-seven genes had significant eQTL association with SNPs in *TPPP* locus, and three genes did in *FAT1-LINC02374* locus (Supplementary Table 7, and Supplementary Fig. 9). Notably, 11 SNPs in *TPPP* were significantly associated with *NKD2* in tubulointerstitium of the nephrotic kidney (rs59402340, $P = 1.64\times10^{-2}$). *PDCD6* (rs35914637, $P = 1.92\times10^{-3}$, tubulointerstitium) and *SLC12A7* (rs57422616, $P = 1.43\times10^{-2}$, glomerulus) were top significant eQTL associated genes with the SNPs from NEPTUNE cohort. In addition, *SLC9A3-AS1* ($P = 8.83\times10^{-11}$, cultured fibroblast; $P = 5.72\times10^{-9}$, blood) and *RP11-325I22.3* of RNA gene ($P = 3.72\times10^{-4}$, cortex of kidney) were significantly associated with *TPPP* SNPs from GTEx (Supplementary Table 8). The expression of *FAT1* gene was significantly regulated by rs4861741 in tubulointerstitium from nephrotic kidney ($P = 1.10\times10^{-2}$). Reported biological functions of genes nearby candidate loci and eQTL associated genes were described in Supplementary Table 9.

We could not find any significant eQTL SNPs from a large-scale meta-analysis eQTL and a cohort of the tubule and glomerulus-specific eQTL database at genome-wide multiple comparisons corrected significance level [25, 26]. We also referred to single-cell RNA sequencing data for nearby genes of top significant loci and eQTL-associated genes in kidney tissue. *ADGRV1* seems expressed in glial cell type and *FAT1* in podocyte according to single cell RNA-Sequencing data of iPSC-derived human kidney Organoids[27] (Supplementary Fig. 10).

Secondly, we performed functional enrichment analysis using these eQTL-associated genes from Gene Ontology (GO)[28, 29], Reactome[30], and Kyoto Encyclopedia of Genes and Genomes (KEGG) database[31], which showed that some of these eQTL genes were enriched in transmembrane transporter activity (Supplementary Table 10). Notably, *SLC12A7, SLC9A3,* and *SLC12A7* genes were significantly

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

enriched in secondary active transmembrane transporter activity of molecular function in GO (q-value=2.83×10^{-2}) and in transport of inorganic cations/anions and amino acids/oligopeptides in Reactome (q-value=6.51×10^{-4}). In the KEGG pathway, *SLC6A19* and *SLC9A3* were involved in mineral absorption (q-value=1.32×10^{-2}).

Exploration for epigenetic architectures and disruption of transcription factor binding

To understand the regulatory function of the non-coding SNPs we found, we further explored the epigenetic architectures near the risk SNPs in *TPPP* and *FAT1-LINC02374* loci using UCSC Genome Browser[32]. Histone activity of H3K27ac and H3K4me1, which indicate enhancer activity, showed a peak near both sets of risk SNPs in *TPPP* and *FAT1-LINC02374* loci and located near gene enhancer regulatory elements (Fig. 4B-C). We also referred open chromatin accessibility by cell types in kidney from snATAC-seq (single nucleus assay for transposase-accessible chromatin using sequencing) near two loci. The open chromatin accessibility around *TPPP* showed a similar pattern with the histone activity, and *FAT1-LINC02374* loci showed a weak peak in the podocyte (Supplementary Fig. 11)[25].

We finally conducted the transcription factor binding site disruption analysis implemented by motifbreakR[33] using HOmo sapiens COmprehensive MOdel Collection (HOCOMOCO) database[34] and Encode-motif database[35]. Twelve SNPs on *TPPP* and 6 SNPs near *FAT1* showed strong effect on the affinity of transcription factor (Supplementary Table 11), which could explain their regulatory function on the eQTL associated genes. For instance, rs59402340 of *TPPP* and rs4861741 of *FAT1-LINC02374* were predicted to be associated with binding of MZF1 and SOX17/ POU1F1, respectively (Fig. 4D).

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Development and evaluation eGFR slope based polygenic risk score from this CKD cohort

We developed polygenic risk scores (PRS) from the GWAS for eGFR slope and evaluated the effect of the PRS on incident CKD outcomes with two methods: (i) selection of the top 22 independent significant SNPs or 6 SNPs replicated in CRIC study from the total, DM, and non-DM GFR slope GWAS results and conducting validation analysis for CKD outcomes in the same cohort (1,738 patients), and (ii) five-fold leave-one-group-out (LOGO) meta-analysis among total patients with clumping and thresholding (C+T) method[36] and performing validation study in remaining subgroups (Supplementary Figure 12). First, PRS based on the independently significant top 22 SNPs were significantly associated with CKD outcomes (Figure 5, Supplemental Figure 13–14). Of note, top 5 percentile of PRS showed 1.91 (95% CI, 1.29–2.83) times increased odds ratio of receiving renal replacement therapy compared with others and 51% increased odds ratio of decline of eGFR by 30% in 3 years (Figure 5, Supplemental Figure 14). The results from 6 SNPs were shown in the Supplementary Table 11. Last, by C+T method the results showed 1.65 (95% CI, 1.42–1.92) times increased odds of 30% reduction of eGFR in 3 years per 1-SD increment of PRS, 1.87 (95% CI, 1.54–2.26) for renal event and 1.68 (95% CI, 1.38–2.11) for renal replacement therapy, respectively (Table 4, Supplemental Figure. 14).

Discussion

In this study, we performed GWAS to identify genetic variation associated with eGFR decline, as a measure of CKD progression. We identified the risk SNPs in *TPPP* and *FAT1-LINC02374* loci associated with eGFR decline among Korean CKD patients and replicated them in an independent CKD cohort with significant association with CKD binary outcomes. We further validated two loci in CKDGen studies. The eQTL *P*-values of the SNPs in the kidney, fibroblast, and blood suggested that they could regulate the expression level of several genes associated with the kidney function. Furthermore, the risk SNP variants were located near gene enhancer regions, disrupting the binding of transcription factors. This series of results suggest the variants on *TPPP* and *FAT1-LINC02374* loci are causally related with the eGFR slope among CKD patients.

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Although cross-sectional GWASs on eGFR have been performed in more than one million subjects[37], studies on CKD progression are rare, possibly because of difficulties in collecting study samples and clinical information. To the best of our knowledge, this is the first study performed in Asian CKD cohorts with replication in the independent CKD cohort with different ethnicities to find genetic variants for eGFR slope. In our study, GWAS was conducted by dividing the study population into total, DM, and non-DM, because our objective was to identify genetic markers in a common pathway related to CKD progression regardless of DM[38].

SNPs around *TPPP* reported to be associated with the kidney function in previous studies, have not received much attention. For instance, rs13159523 near downstream of *TPPP* was reported as a variant to be associated with eGFR in trans-ethnic meta-analysis[18, 20]. Also, another trans-ethnic GWAS of eGFR in 280,722 participants showed rs413988 and rs62330329 near upstream of *TPPP* were risk variants with eGFR in DM and non-DM participants, respectively[21]. Interestingly, the locus identified in the present study with CKD population is distinct from previously reported ones based on relatively low r^2 (Supplementary Fig 4).

The TPPP protein is a microtubule-associated protein that binds to the tubulin and microtubules, promoting tubulin polymerization and modulating their dynamics. Oligodendrocytes from *TPPP* knockout mice have aberrant microtubule branching and polarity. *TPPP* knockout mice exhibit hypomyelination with shorter, thinner myelin sheaths and motor coordination deficits[39]. In a rodent model of CKD, high salt intake induced phosphorylation changes of TPPP in the hypothalamic paraventricular nucleus, suggesting a role in CKD-induced paraventricular nucleus damage[40]. A specific mechanism of *TPPP* in the kidney disease has not been confirmed. One experimental study reported that ischemia/reperfusion injury induced deacetylation of the microtubules, and that inhibition of microtubule dynamics hindered repair of injured tubular epithelial cells, leading to an acceleration of fibrosis[39]. Based on those findings, we postulate that alteration in *TPPP* can lead to vulnerability toward ischemic injury, resulting in the kidney damage.

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Our identified GWAS signals of the *TPPP* gene lie in non-coding regions, located in gene enhancers, regulating several genes associated with kidney function in the kidney tissues, fibroblasts, and blood cells. Recent studies have suggested that genetic variants in gene enhancers or promoters might have regulatory functions affecting regulation of nearby genes[41]. Of the genes thought to be regulated by *TPPP* in the study, the *NKD2* expressed in myofibroblasts, has been reported to have a functional role in kidney fibrosis[42]. In addition, the *SLC6A19, SLC9A3,* and *SLC12A7* genes were reported to have a critical role in transport of ions, which is important in the kidney function[43]. The specific role of the *TPPP* gene in CKD progression, including whether the actual *TPPP* gene functionally regulates nearby or distant genes, is under exploration. Other genes identified from the eQTL analysis need further investigation in relation to the kidney functions.

Another replicated SNPs were on the intergenic region near *FAT1*, which encodes FAT atypical cadherin 1 and has been described as a tumor suppressor or oncogenic gene[44]. The *FAT1* was reported to be associated with glomerulotubular nephropathy[45]. Recessive mutation in *FAT1* causes nephrotic syndrome, hematuria, and tubular ectasia; loss of *FAT1* causes defects in migration and cell-to-cell adhesion in fibroblasts and cultured podocytes. Notably, the loci identified in the present study were intergenic, between *FAT1* and *LINC02374*.

Other SNPs on the candidate loci of *ADGRV1*, *HS3ST4-C16orf82*4, *SLC2A1*, and *FAM124B-SERPINE2* were rare variants (MAF < 0.1 %) in CRIC study, which necessitates larger study population to confirm their replicability. Of note, the most significant SNP in our discovery cohort was located in *ADGRV1* (*GPR98*), which encodes a member of the G-protein coupled receptor superfamily. Adhesion G protein-coupled receptors play important roles in kidney development. A recent study revealed *ADGRV1* was enriched in the medulla of the adult human kidney, and was associated with kidney disease in the Nephroseq database [46]. Further research is needed to validate the role of *ADGRV1* in the kidney disease.

We also constructed several PRS models with eGFR-associated SNPs, which showed significant

Copyright © 2023 by the American Society of Nephrology

associations with binary CKD outcomes. The performances of the PRS models from top 22 significant SNPs either replicated 6 SNPs were estimated in the same cohort that were used for the discovery of PRS because there are not available large-scale East Asian CKD cohort with both the genotypes and longitudinal CKD outcomes, which could induce inflation of performance of PRS [47-49]. Therefore, we also used the LOGO cross-validation method to avoid sample overlap in the PRS development and validation[36, 48]. Nevertheless, validation studies for evaluating the performance of PRS from other independent CKD patients cohort in East Asia are warranted. Since the KNOW-CKD Cohort, on which the present study is based, has almost finished recruitment of another 1,500 CKD subjects for Phase II and started longitudinal follow-up, we could replicate the PRS with this population afterwards. We additionally conducted if there is collider bias regarding adjustment of baseline eGFR, which was recently explored in large-scale meta-analysis (Supplemental Fig. 16.)[50, 51] . We could not find the substantial existence of collider bias of baseline eGFR among our CKD cohort, probably due to the difference in the main composition of study population: mainly for established CKD patients in the present study and the general population in the previous study [50].

Our study had limitations. The statistical power was limited owing to the relatively small population. The statistical power from the results of rs61464149 in *TPPP* (Non-DM) for eGFR slope was 62% and 81% at the significance level of 5×10^{-8} and 10^{-6}, respectively (Supplementary Table 1–3). To minimize the probability of false positive results, we tried to replicate the selected SNPs in another independent cohort and focused on the restricted loci with subsequent validation analyses. We could not undertake replication studies in other Asian CKD cohorts but, instead, did in the European American and African American CKD cohorts, because there were few Asian CKD cohorts with available genomic information. We conducted replication studies on the same SNPs (r^2 =1) to address this ethnicity mismatch between discovery and replication cohorts[52]. Accounting for importance of replicability in ancestrally matched cohort for GWAS results[53], further replication studies with large-scale Asian CKD populations are needed. The follow-up duration, median 3.3 years, was relatively short to confirm the

Copyright © 2023 by the American Society of Nephrology

renal outcome. In addition, our validation analysis for replicated loci were based on *in-silico* approach, rather than an experimental validation. Therefore, further studies including *in-vivo* or *in-vitro* functional investigation for candidate loci with their related genes and analysis of physiological mechanisms in the kidney, are warranted. eQTL datasets and epigenetic structure used in this study were obtained from the normal kidney (GTEx) or kidney from nephrotic syndrome (NEPTUNE). Gene expression or epigenomic structure in CKD might be different from these in GTEx or NEPTUNE, which requires the further investigation such as CKD specific eQTL studies or exploration of epigenetic structures of the kidneys from CKD patients[54].

This study suggests that genetic variations in *TPPP* and *FAT1-LINC02374* loci are associated with CKD progression. These variations appear to be involved in CKD progression by affecting the activity of transcription factor binding and the regulation of adjacent or distant genes. Our GWAS for CKD progression in established CKD patients can provide imperative insights into the genomic contribution of SNPs to eGFR decline in CKD patients and a credible basis for future genomic studies based on various clinical outcomes of CKD.

Methods

Study subjects

A total of 2,426 subjects with chronic kidney disease was screened; 2,238 from the KoreaN Cohort Study for Outcomes in Patients With Chronic Kidney Disease (KNOW-CKD)[55] and 188 patients with biopsy-proven diabetic nephropathy from Seoul National University Hospital Human Biobank or Kyung Hee University Medical Center. After excluding subjects with polycystic kidney disease (N=364), those without genomic DNA available (N=104), and those whose eGFR slope could not be calculated owing to loss of follow up (N=63), genomic DNA for 1,895 subjects underwent quality control (QC) testing. Of these, 157 samples did not pass QC, resulting in inclusion of 1,738 patients in the discovery GWAS. We

Copyright © 2023 by the American Society of Nephrology

acquired written informed consent and blood samples from all patients, and the study was approved by the Institutional Review Board of Seoul National University Hospital (C-1704-025-842).

Genotyping and quality control

Genomic DNA was extracted from peripheral blood leukocytes isolated from whole blood collected in ethylenediaminetetraacetic acid-coated tubes. Genotyping data were produced using the KoreanChip (K-CHIP, Affymetrix® Axiom™ KORV1.1, Santa Clara, CA, USA) obtained from the K-CHIP Consortium. K-CHIP was designed by Center for Genome Science, Korea National Institute of Health, Korea. Detailed information for K-CHIP has been published previously[56]. The K-CHIP is produced by the Korea Biobank Array Project and is designed to overcome pre-existing multi-ethnic genome chips, which are not suitable for genomic research of Asian races[56]. Although Korean Association REsources data, based on Affymetrix 5.0, have a limitation of low genomic coverage[56], the K-CHIP platform is optimized for the Korean genomic structure, with the highest number of nonsynonymous SNPs in East Asians. Many GWASs have been conducted using K-CHIP[57].

QC was conducted using the K-CHIP QC protocol[56]. Samples matching any of the following criteria were excluded: (i) Dish QC (DQC) < 0.82, (ii) call rate < 97%, (iii) related or cryptically related individuals (identical by descent > 0.8), and (iv) sex inconsistences. SNPs were excluded if they had (i) a missing call rate > 5%, (ii) a minor allele frequency (MAF) < 1%, and (iii) a *P*-value of Hardy-Weinberg equilibrium test < 10^{-6}. Consequently, 745,176 autosomal SNPs remained for the association analysis, imputed by minimac3[58] with a reference panel of all population in phase 3 from the 1000 Genomes Project[59], and phasing with shapeit_v2.12[60]. Finally, 7,734,192 SNPs remained after imputation QC, excluding SNPs with r^2 values < 0.7 or MAF < 1%.

Copyright © 2023 by the American Society of Nephrology

Outcome variables

We used eGFR slope as the primary outcome in this study, rather than development of ESKD or mortality. eGFR slope, which represents annual change in eGFR, is associated with future risk of ESKD disease, cardiovascular disease, and death in general or in a CKD cohort[4]. The eGFR slope has been used as a surrogate outcome for renal progression and is also used as a valid renal endpoint in randomized controlled trials[5]. Secondary outcomes for CKD progression as binary trait included 1) 30% reduction of eGFR in 3 years, 2) renal event defined as the composite of doubling of creatinine, eGFR halving or initiation of renal replacement therapy and 3) renal replacement therapy defined as initiation of kidney transplantation or dialysis.

The eGFR was calculated using the four-variable Chronic Kidney Disease Epidemiology Collaboration equation[61]. The eGFR slope was defined as the eGFR change over time. We assumed that the fixed effect according to the time of repeated eGFR measurement ('Time' variable) can be different for each patient and thus both the fixed effect and the random effect of the time of repeated eGFR affected to random eGFR slope. Under the assumption, we used a mixed model to estimate eGFR slope and the 'Time' variable was set for both fixed and random effects [62, 63] . Hypertension was defined as systolic BP ≥140 mm Hg, diastolic BP ≥90 mm Hg, or a previous diagnosis. Diabetic mellitus was defined as serum hemoglobin A1c ≥6.5%, fasting glucose ≥126 mg/dl, or a previous diagnosis of diabetes.

Statistical analysis

Descriptive statistics on the baseline characteristics of study population were calculated based on DM and non-DM. Results were expressed as mean (standard deviation) or as numbers and percentages. The comparisons were made using the independent t-test for continuous variables and the chi-squared test for categorical variables. A *P*-value < 0.05 was considered statistically significant. The primary outcome was eGFR slope, calculated as mentioned above. We conducted principal component (PC) analysis for

Copyright © 2023 by the American Society of Nephrology

population stratification (Supplementary Fig. 3). To explore genetic loci associated with eGFR slope, we conducted multivariable linear regression analysis adjusting for age, sex, and first 5 PC using an additive model using PLINK tools (version 2.0)[64] . Manhattan plots and quantile-quantile (QQ) plots were drawn by using qqman R package[65, 66]. Annotations for each SNP were conducted via ANNOVAR[67], and descriptions for each gene summary were referred from GeneCards[68]. To identify independent genomic loci, we clumped GWAS results with ±150 Kbp, *P*-value threshold of 10^{-6}, and r^2 threshold of 0.1 calculated from 1000 Genome phase 3 population. Linkage disequilibrium (LD) plots for *TPPP* and *FAT1-LINC02374* loci across multiple ethnicities were displayed using gaston R package[69] calculating LD association by LDlinkR[70]. For the added clinical perspective, for selected loci, we used a logistic regression model for binary traits with same covariates as initial model. For the sensitivity analysis, we additionally added prognostic variables estimated at baseline including urine albumin-to-creatinine ratio either baseline eGFR on the initial regression model. Furthermore, for the results from adjustment of baseline eGFR, we additionally assessed the collider bias, which was recently explored in similar study design [50, 51]. We calculated post-hoc statistical power with genpwr [71] (Supplementary Table 1).

Trans-ethnic replication analyses from Chronic Renal Insufficiency Cohort (CRIC) study

We conducted trans-ethnic replication studies with CRIC[72] study to find causal variants for eGFR slope among CKD patients[73] for the SNPs with discovery GWAS $P < 10^{-6}$ of genome-wide suggestive significance. The eGFR slopes were calculated as the same method in the discovery study and linear regression on eGFR slope was conducted for the candidate SNPs among 2,498 subjects in CRIC study (by ethnicity and by diabetes mellitus) as previously reported[15]. Criteria for the candidate loci were based on (i) $P < 0.05$, (ii) consistent direction of effect, and (iii) combined *P*-values < discovery *P*-values. Consequently, we conducted inverse variance-weighted meta-analyses to estimate combined β (beta coefficient values), standard errors, and their *P*-values of the exact same SNPs[74] from candidate loci assuming a fixed effect model using a meta R package[75]. We estimated the heterogeneity between the

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

results derived from two cohorts in meta-analysis using Higgins I^2 and Cochran's Q test[76] (Supplementary Table 2). We produced the regional plots using LocusZoom around the top significant SNPs for each independent locus[77]. For the definite replication, we additionally restricted candidate loci to those with minor allele frequency (MAF) ≥ 1% in CRIC study considering the relatively small number of the study population.

Validation studies with CKDGen

We conducted additional validation studies for *TPPP* and *FAT1-LINC02374* loci from the two publicly available CKDGen Consortium GWAS summary statistics, one of which reported GWAS results on cross-sectional eGFR or BUN in general population[18], and the other study was regarding eGFR change in a general population with or without CKD at baseline. Criteria of validation from CKDGen studies were based on (i) $P < 0.05$ and (ii) consistent direction of effect on deteriorating kidney function.

Fine-mapping

Fine-mapping is the process by which a trait-associated region from a GWAS is analyzed to identify the particular genetic variants that are likely to causally influence the examined trait[78]. We conducted fine-mapping analysis using the Sum of Single Effects (SuSiE) using summary statistics method in each *TPPP* and *FAT1-LINC02374* region from the discovery GWAS summary statistics. Selected regions were spanning 100 Kbp around the top significant SNPs. LD matrices for the SNPs in these regions were produced by LDlinkR[70]. Based on iterative Bayesian stepwise selection (IBSS), a SuSiE regression model was estimated to calculate posterior inclusion probability (PIP) for a SNP, which is the probability of inclusion of the SNP in a causal association[78]. Credible sets were created through iterative model fitting,

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

which is performed by ranking the SNPs from the largest to the smallest PIP. Credible sets allow probabilistic selection of potential causal variants to prioritize to identify multiple causal variants[78]. We implemented fine-mapping analysis using susieR (v0.11.92)[79].

Prioritizing genes from expression quantitative trait loci (eQTL) studies

To investigate the functional mechanism of SNP markers in *TPPP* and *FAT1-LINC02374* loci, we prioritized genes from publicly available eQTL studies of Nephrotic Syndrome Study Network (NEPTUNE)[23] and Genotype-Tissue Expression (GTEx) Project[24]. We acquired eQTL results of tubulointerstitium and glomerulus from nephrotic syndrome patients from NEPTUNE. In addition, we accessed to the eQTL dataset of the cortex of normal kidney and the blood from the GTEx. For an evaluation of potential mechanisms of kidney disease progression, we also referred to the eQTL datasets of cultured fibroblasts from the GTEx database considering a crucial role of fibroblast in kidney disease progression[42]. The significance of eQTL association of the SNPs set as a nominal $P < 0.05$. We also referred to the association the SNPs from other human kidney eQTL analyses for tubule and glomeruli and a meta-analysis of kidney eQTL studies from Susztaklab Kidney Biobank[25, 26]. For prioritized genes, we referred each gene of expression pattern in single cell levels from single-cell RNA-sequencing data of iPSC-derived human kidney organoids (RID: 16-KDDC)[27].

ACCEPTED

Functional enrichment analysis

For the interpretation of biological functions of genes associated with eQTL of validated SNPs, we performed enrichment analyses using the Gene Ontology (GO)[28, 29], Reactome[30], and Kyoto Encyclopedia of Genes and Genomes (KEGG)[31] database using clusterProfiler[80]. Significant enrichment cut-off for *q*-value of false discovery rate was set as 0.05.

Copyright © 2023 by the American Society of Nephrology

Identification of epigenetic architecture

To investigate the functional relevance of the non-coding SNPs from *TPPP* and *FAT1-LINC02374* we found, we explored the epigenetic architecture around these SNPs using UCSC Genome Browser[32, 41, 81]. We acquired the histone marks from histone activity of monomethylation of lysine 4 on histone H3 protein subunit (H3K4Me1), trimethylation of lysine 4 on histone H3 protein subunit (H3K4Me3), and acetylation of lysine 27 on histone H3 protein subunit (H3K27AC) using chromatin immunoprecipitation-sequencing (ChIP-seq) data from the seven cell lines from Encyclopedia of DNA Elements (ENCODE)[82]. In addition, we identified regulatory elements (gene enhancers and promotors) from the GeneHancer[83]. We also explored open chromatin accessibility by cell types in human adult kidney organoids from snATAC-seq (Single nucleus assay for transposase-accessible chromatin using sequencing) data around the two loci[27].

Transcription factor binding site (TFBS) disruption analysis

To determine whether the risk variants were disrupting motif-binding sites, we conducted transcription factor (TF)-binding disruption analysis using motifbreakR[33] with default parameters and a threshold *P*-value of 10^{-4}. This tool predicts the effects of variants on TF-binding motifs using probability matrices to determine the likelihood of observing a particular nucleotide at a specific position within a TF-binding site based on HOCOMOCO and Encode-motif database[33-35]. We selected the variants with strong effect on TFBS disruption.

Polygenic risk score (PRS) analysis

We constructed several PRS analyses[47]. First, top significant 22 SNPs from total, DM, non-DM GWAS were derived to estimate PRS as a commonly used approach[84] using following equation; $-\sum(\beta \times$

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

dosage), which were standardized into z-scores. We evaluated if this eGFR slope-associated PRS were also associated with binary CKD outcomes in the same total cohort of 1,738 CKD patients. We calculated hazard ratios with Cox proportional regression model to calculate risk of eGFR outcome (decline of eGFR by 30% to baseline in a 3-year, renal event, and renal replacement therapy) of upper 30% of PRS compared with other 70%, 20% vs 80%, 10% vs 90%, 5% vs 95% with cumulative event curves, and of 1-SD increment of PRS. Both crude (PRS only) and multivariable model including PRS, age, sex, 5 PC, diabetes mellitus, systolic blood pressure, body mass index, and urine albumin creatine ratio were used. We also constructed PRS with 6 replicated independent SNPs from CRIC study and evaluated this PRS on the risk of CKD outcomes.

The validation of PRS is preferred in the independent cohort. However, there is no other relevant and independent CKD cohorts in East Asia[49]. Therefore, we additionally used a 5-fold leave-one-group-out (LOGO) cross-validation method with C+T (clumping and thresholding) to avoid overlapping of the subjects for PRS development and PRS validation[48,36]. We split the 1,738 patients to 5-group by random, and conducted GWAS for eGFR slope and subsequently meta-analyzed the 4 results with inverse-variance weighting and fixed effect model (5 times meta-analysis). The remaining subgroups were subsequently divided into test and validation datasets. Clumping was conducted with r^2=0.1 and maximum distance of 150Kbp among all of SNPs and thresholds of P-values included 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 5×10^{-3}, 1×10^{-2}, 5×10^{-2}, 1×10^{-1}, 5×10^{-1}, and 1. We calculated mean R^2 from 5 results in each set of P-value cutoff. We chose a reasonable R^2 (0.09) with $P=5\times10^{-3}$, which included mean 6,094 variants. In each C+T method, after excluding the C+T criterion in which the R^2 value was calculated as near zero, the C+T criterion with the highest increment % (89%) between two C+T criteria from $P=10^{-3}$ to P=1 was $P=5\times10^{-3}$, which had mean 6,094 SNPs. In addition, this criterion showed reasonable mean R^2 of 0.09 based on low expected explanation of PRS on CKD[4]. For the reason, the optimal PRS model was constructed with SNPs selected by C+T method with $P=5\times10^{-3}$ in this study. Finally, we calculated hazard ratios for each binary outcomes with same method above and meta-analyzed the 5 results.

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Data Sharing Statement:

Genome-wide association summary statistics generated during this study are available from the corresponding author upon reasonable request. Summary statistics generated during this study are available from Supplemental Table 1–3.

URLs

Plink2: https://www.cog-genomics.org/plink/2.0/

ANNOVAR: https://annovar.openbioinformatics.org/en/latest/

R: https://www.r-project.org/

LocusZoom: http://locuszoom.sph.umich.edu/

LDlink: https://analysistools.cancer.gov/LDlink/?tab=home

Genecards: https://www.genecards.org/

CKDGen: https://ckdgen.imbi.uni-freiburg.de/

UCSC genome browser: https://genome.ucsc.edu/

NephQTL: http://nephqtl.org/

GTEx: https://www.gtexportal.org/home/

ENCODE: https://www.encodeproject.org/

Human kidney Open chromatic atlas: https://susztaklab.com/Human_snATAC/index.php

Human Kidney eQTL Atlas: https://susztaklab.com/Kidney_eQTL/index.php

Human single cell expression from iPSC derived kidney organoid:

https://www.rebuildingakidney.org/chaise/record/#2/RNASeq:Study/RID=16-1YZM

Copyright © 2023 by the American Society of Nephrology

Funding:

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (2017M3A9E4044649) and a grant of the MD-PhD/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea.

Acknowledgments:

We appreciate KNOW-CKD researchers for their providing biospecimens and clinical information of the patients. The Biospecimens for 188 patients with biopsy proven diabetes nephropathy in Seoul National University Hospital were provided by the Biobank of Seoul National University Hospital, a member of the Korea Biobank Network. We also appreciate CRIC investigators for their kind collaborative research. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 08/23/21. We appreciate the permission of use of data from Susztaklab Kidney Biobank from Dr. Katalin Susztak.

Author Contributions:

Kook-Hwan Oh: Conceptualization, Writing – review & editing

Miyeun Han: Conceptualization, Writing – original draft

Sangjun Lee: Formal analysis, Validation

Kyungsik Kim: Formal analysis, Validation

Woo Ju An: Formal analysis, Validation

Hyunjin Ryu: Data curation, Resources

Eunjeong Kang: Data curation, Resources

Copyright © 2023 by the American Society of Nephrology

Jung-Hyuck Ahn: Investigation, Supervision

Hye Youn Sung: Investigation, Supervision

Yong Seek Park: Methodology, Supervision

Seung Eun Lee: Methodology, Supervision

Sang-Ho Lee: Data curation, Resources

Kyung Hwan Jeong: Data curation, Resources

Curie Ahn: Conceptualization, Writing – review & editing

Tanika Kelly: Investigation, Validation

Jesse Hsu: Investigation, Validation

Harold Feldman: Investigation, Writing – review & editing

Sue Park: Conceptualization, Writing – review & editing

Sungji Moon: Formal analysis, Writing – original draft

Disclosures:

K. Oh reports Research Funding: Fresenius Medical Care, Korea. J. Hsu reports Advisory or Leadership Role: AJKD Statistics/Epidemiology Editor, and PLOS ONE Statistical Advisory Board. H. Feldman reports the following:

Employer: Perleman School of Medicine, University of Pennsylvania, PCORI; Consultancy: Kyowa Hakko Kirin Co,

Ltd., National Kidney Foundation (Ongoing); Honoraria: InMed, Inc.; Advisory or Leadership Role: Steering

Committee Chair Chronic Renal Insufficiency Cohort Study, Member of Advisory Board of the National Kidney

Foundation, American Journal of Kidney Disease (AJKD) Editor in Chief; and Speakers Bureau: InMed, Inc.

Copyright © 2023 by the American Society of Nephrology

Supplemental Table of Contents:

Supplementary Fig. 1. eGFR slope of total, diabetes, and non-diabetes group.

Supplementary Fig. 2. Plots for principal components (PCs) (N=1,738)

Supplementary Fig. 3. Quantile-Quantile plots for eGFR slope of (A) total, (B) diabetes, and (C) non-diabetes group.

Supplementary Fig. 4. Manhattan plots for DM (A) and non-DM (B) patients GWAS for eGFR slope.

Supplementary Fig. 5. Regional association plots for significant GWAS loci with P < 10-6 confirmed by CRIC study.

Supplementary Fig. 6. Linkage disequilibrium plot for selected SNP on TPPP locus with extra 3 SNPs of rs13159523, rs62330329, and rs413988.

Supplementary Fig. 7. Linkage disequilibrium plot for selected SNPs on FAT1-LINC02374 locus.

Supplementary Fig. 8. Fine-mapping results for (A) TPPP and (B) FAT1-LINC02374.

Supplementary Fig. 9. Prioritizing genes from eQTL datasets.

Supplementary Fig. 10. Expression profile from single cell RNA-Seq data of iPSC-derived Human Kidney Organoids (https://www.rebuildingakidney.org/about/usage.html) (A) tSNE graph for cell type specificity in kidney organoid (B) ADGRV1, FAT1, and, NKD2 single gene expression by single cell level (Abbreviation: tSNE, t-Distributed Stochastic Neighbor Embedding).

Supplementary Fig. 11. Open chromatin accessibility by the cell types in human adult kidneys around TPPP and FAT1-LINC02374 (http://www.susztaklab.com/Human_snATAC/index.php).

Supplementary Fig. 12. Evaluation of PRS from leave-one-group-out (LOGO) with C+T (Clumping + Thresholding) method (A) Brief method for leave-one-group-out (LOGO) meta-analysis and cross validation (B) Mean R2 for five results of LOGO cross-validation at different threshold P-value, such as 1, 0.5, 0.1, 0.05, 0.01, 5x10-3, 1x10-3, 1x10-4, 1x10-5, and 1x10-6. (C)

Supplementary Fig. 13. Distributions of polygenic risk score (PRS) constructed by 22 single nucleotide polymorphisms (Abbreviation: PRS, polygenic risk score; σ, standard deviation).

Supplementary Fig. 14. Evaluation of PRS from top significant 22 SNPs on renal event and renal

Copyright © 2023 by the American Society of Nephrology

replacement therapy by tail based cutoff value (Abbreviation: PRS, polygenic risk score).

Supplementary Fig. 15. Evaluation for collider bias of baseline eGFR.

Supplementary Table 1. GWAS summary statistics for eGFR slope in the KNOW-CKD: Results in total CKD patients

Supplementary Table 2. GWAS summary statistics for eGFR slope in the KNOW-CKD: Results in diabetic CKD patients

Supplementary Table 3. GWAS summary statistics for eGFR slope in the KNOW-CKD: Results in non-diabetic CKD patients

Supplementary Table 4. Results of genetic variants related to eGFR slope in GWAS data from the Chronic Renal Insufficiency Cohort study (CRIC) study

Supplementary Table 5. Association with genetic variants on TPPP and FAT1-LINC02374 on the risk of CKD progression (3-year 30% decline of eGFR, 6-year renal events and renal replacement therapy)

Supplementary Table 6. Fine-mapping results for identifying causal variants on TPPP and FAT1-LINC02374 for eGFR slope in the KNOW-CKD

Supplementary Table 7. GWAS-eQTL integration results for identifying risk genetic variants on TPPP and FAT1-LINC02374 genes for eGFR slope

Supplementary Table 8. Summary of genes nearby candidate loci and eQTL linked genes with TPPP and FAT1-LINC02374

Supplementary Table 9. Results of functional enhancement analysis for eQTL in TPPP and FAT1-LINC02374 associated genes to confirm biological function

Supplementary Table 10. Prediction for disruptiveness of transcription factor (TF) binding sites in TPPP and FAT1-LINC02374 gene using motifbreakR

Supplementary Table 11. Comparison of methods for deriving the PRS according the p-value cutoff and R2

Supplementary Table 12. CKD progression associated with the PRS composed of 6 replicated SNPs in the KNOW-CKD original dataset

Copyright © 2023 by the American Society of Nephrology

SDC-http://links.lww.com/JSN/D632

References:

1. Xie Y, Bowe B, Mokdad AH, Xian H, Yan Y, Li T, et al.: Analysis of the Global Burden of Disease study highlights the global, regional, and national trends of chronic kidney disease epidemiology from 1990 to 2016. *Kidney Int,* 94**:** 567-581, 2018 10.1016/j.kint.2018.04.011

2. Xie Y, Bowe B, Xian H, Balasubramanian S, Al-Aly Z: Renal Function Trajectories in Patients with Prior Improved eGFR Slopes and Risk of Death. *PLoS One,* 11**:** e0149283, 2016 10.1371/journal.pone.0149283

3. Orlandi PF, Xie D, Yang W, Cohen JB, Deo R, Ricardo AC, et al.: Slope of Kidney Function and Its Association with Longitudinal Mortality and Cardiovascular Disease among Individuals with CKD. *J Am Soc Nephrol,* 31**:** 2912-2923, 2020 10.1681/ASN.2020040476

4. Oshima M, Jun M, Ohkuma T, Toyama T, Wada T, Cooper ME, et al.: The relationship between eGFR slope and subsequent risk of vascular outcomes and all-cause mortality in type 2 diabetes: the ADVANCE-ON study. *Diabetologia,* 62**:** 1988-1997, 2019 10.1007/s00125-019-4948-4

5. Greene T, Ying J, Vonesh EF, Tighiouart H, Levey AS, Coresh J, et al.: Performance of GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Statistical Simulation. *J Am Soc Nephrol,* 30**:** 1756-1769, 2019 10.1681/ASN.2019010009

Copyright © 2023 by the American Society of Nephrology

6. Inker LA, Heerspink HJL, Tighiouart H, Levey AS, Coresh J, Gansevoort RT, et al.: GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Meta-Analysis of Treatment Effects of Randomized Controlled Trials. *J Am Soc Nephrol,* 30**:** 1735-1745, 2019 10.1681/ASN.2019010007

7. Melsom T, Nair V, Schei J, Mariani L, Stefansson VTN, Harder JL, et al.: Correlation Between Baseline GFR and Subsequent Change in GFR in Norwegian Adults Without Diabetes and in Pima Indians. *Am J Kidney Dis,* 73**:** 777-785, 2019 10.1053/j.ajkd.2018.11.011

8. Anderson AH, Xie D, Wang X, Baudier RL, Orlandi P, Appel LJ, et al.: Novel Risk Factors for Progression of Diabetic and Nondiabetic CKD: Findings From the Chronic Renal Insufficiency Cohort (CRIC) Study. *Am J Kidney Dis,* 77**:** 56-73 e51, 2021 10.1053/j.ajkd.2020.07.011

9. Friedman DJ, Pollak MR: APOL1 Nephropathy: From Genetics to Clinical Applications. *Clin J Am Soc Nephrol,* 16**:** 294-303, 2021 10.2215/CJN.15161219

10. Hildebrandt F: Genetic kidney diseases. *Lancet,* 375**:** 1287-1295, 2010 10.1016/S0140-6736(10)60236-X

11. Kottgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, et al.: Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet,* 41**:** 712-717, 2009 10.1038/ng.377

Copyright © 2023 by the American Society of Nephrology

12. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al.: Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun,* 7**:** 10023, 2016 10.1038/ncomms10023

13. Trudu M, Janas S, Lanzani C, Debaix H, Schaeffer C, Ikehata M, et al.: Common noncoding UMOD gene variants induce salt-sensitive hypertension and kidney damage by increasing uromodulin expression. *Nat Med,* 19**:** 1655-1660, 2013 10.1038/nm.3384

14. Khalili H, Sull A, Sarin S, Boivin FJ, Halabi R, Svajger B, et al.: Developmental Origins for Kidney Disease Due to Shroom3 Deficiency. *J Am Soc Nephrol,* 27**:** 2965-2973, 2016 10.1681/ASN.2015060621

15. Parsa A, Kanetsky PA, Xiao R, Gupta J, Mitra N, Limou S, et al.: Genome-Wide Association of CKD Progression: The Chronic Renal Insufficiency Cohort Study. *J Am Soc Nephrol,* 28**:** 923-934, 2017 10.1681/ASN.2015101152

16. Wen CP, Matsushita K, Coresh J, Iseki K, Islam M, Katz R, et al.: Relative risks of chronic kidney disease for mortality and end-stage renal disease across races are similar. *Kidney Int,* 86**:** 819-827, 2014 10.1038/ki.2013.553

17. Barbour SJ, Er L, Djurdjev O, Karim M, Levin A: Differences in progression of CKD and mortality amongst Caucasian, Oriental Asian and South Asian CKD patients. *Nephrol Dial Transplant,* 25**:** 3663-3672, 2010 10.1093/ndt/gfq189

Copyright © 2023 by the American Society of Nephrology

18. Stanzick KJ, Li Y, Schlosser P, Gorski M, Wuttke M, Thomas LF, et al.: Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat Commun,* 12**:** 4350, 2021 10.1038/s41467-021-24491-0

19. Gorski M, Tin A, Garnaas M, McMahon GM, Chu AY, Tayo BO, et al.: Genome-wide association study of kidney function decline in individuals of European descent. *Kidney Int,* 87**:** 1017-1029, 2015 10.1038/ki.2014.361

20. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al.: A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet,* 51**:** 957-972, 2019 10.1038/s41588-019-0407-x

21. Hellwege JN, Velez Edwards DR, Giri A, Qiu C, Park J, Torstenson ES, et al.: Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. *Nat Commun,* 10**:** 3842, 2019 10.1038/s41467-019-11704-w

22. Gao Wang AS, Peter Carbonetto,Matthew Stephens: A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 82**:** 1273-1300, 2020

23. Gillies CE, Putler R, Menon R, Otto E, Yasutake K, Nair V, et al.: An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *The American Journal of Human Genetics,* 103**:** 232-244, 2018 10.1016/j.ajhg.2018.07.004

Copyright © 2023 by the American Society of Nephrology

24. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al.: The Genotype-Tissue Expression (GTEx) project. *Nature Genetics,* 45**:** 580-585, 2013 10.1038/ng.2653

25. Liu H, Doke T, Guo D, Sheng X, Ma Z, Park J, et al.: Epigenomic and transcriptomic analyses define core cell types, genes and targetable mechanisms for kidney disease. *Nat Genet,* 54**:** 950-962, 2022 10.1038/s41588-022-01097-w

26. Sheng X, Guan Y, Ma Z, Wu J, Liu H, Qiu C, et al.: Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. *Nat Genet,* 53**:** 1322-1333, 2021 10.1038/s41588-021-00909-9

27. Oxburgh L, Carroll TJ, Cleaver O, Gossett DR, Hoshizaki DK, Hubbell JA, et al.: (Re)Building a Kidney. *J Am Soc Nephrol,* 28**:** 1370-1378, 2017 10.1681/ASN.2016101077

28. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res,* 49**:** D325-d334, 2021 10.1093/nar/gkaa1113

29. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics,* 25**:** 25-29, 2000 10.1038/75556

30. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al.: The reactome pathway knowledgebase. *Nucleic Acids Res,* 48**:** D498-d503, 2020 10.1093/nar/gkz1031

31. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res,* 28**:** 27-30, 2000 10.1093/nar/28.1.27

Copyright © 2023 by the American Society of Nephrology

32. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al.: The human genome browser at UCSC. *Genome Res,* 12**:** 996-1006, 2002 10.1101/gr.229102

33. Coetzee SG, Coetzee GA, Hazelett DJ: motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics,* 31**:** 3847-3849, 2015 10.1093/bioinformatics/btv470

34. Kulakovskiy IV, Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, et al.: HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Research,* 41**:** D195-D202, 2012 10.1093/nar/gks1089

35. Kheradpour P, Kellis M: Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Research,* 42**:** 2976-2987, 2013 10.1093/nar/gkt1249

36. Sakaue S, Kanai M, Karjalainen J, Akiyama M, Kurki M, Matoba N, et al.: Trans-biobank analysis with 676,000 individuals elucidates the association of polygenic risk scores of complex traits with human lifespan. *Nat Med,* 26**:** 542-548, 2020 10.1038/s41591-020-0785-8

37. Tin A, Kottgen A: Genome-Wide Association Studies of CKD and Related Traits. *Clin J Am Soc Nephrol*, 2020 10.2215/CJN.00020120

38. Diabetic kidney disease. *Nat Rev Dis Primers,* 1**:** 15038, 2015 10.1038/nrdp.2015.38

39. Fu MM, McAlear TS, Nguyen H, Oses-Prieto JA, Valenzuela A, Shi RD, et al.: The Golgi Outpost Protein TPPP Nucleates Microtubules and Is Critical for Myelination. *Cell,* 179**:** 132-146 e114, 2019 10.1016/j.cell.2019.08.025

Copyright © 2023 by the American Society of Nephrology

40. Li L, Li J, Tan L, Qiu M, Zhang M, Li A: Salt-induced phosphoproteomic changes in the hypothalamic paraventricular nucleus in rats with chronic renal failure. *Brain Res,* 1669**:** 1-10, 2017 10.1016/j.brainres.2017.05.023

41. Tak YG, Farnham PJ: Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin,* 8**:** 57, 2015 10.1186/s13072-015-0050-4

42. Kuppe C, Ibrahim MM, Kranz J, Zhang X, Ziegler S, Perales-Paton J, et al.: Decoding myofibroblast origins in human kidney fibrosis. *Nature,* 589**:** 281-286, 2021 10.1038/s41586-020-2941-1

43. Takemon Y, Chick JM, Gerdes Gyuricza I, Skelly DA, Devuyst O, Gygi SP, et al.: Proteomic and transcriptomic profiling reveal different aspects of aging in the kidney. *Elife,* 10, 2021 10.7554/eLife.62585

44. Katoh M: Function and cancer genomics of FAT family genes (review). *Int J Oncol,* 41**:** 1913-1918, 2012 10.3892/ijo.2012.1669

45. Gee HY, Sadowski CE, Aggarwal PK, Porath JD, Yakulov TA, Schueler M, et al.: FAT1 mutations cause a glomerulotubular nephropathy. *Nat Commun,* 7**:** 10822, 2016 10.1038/ncomms10822

46. Cazorla-Vazquez S, Engel FB: Adhesion GPCRs in Kidney Development and Disease. *Front Cell Dev Biol,* 6**:** 9, 2018 10.3389/fcell.2018.00009

Copyright © 2023 by the American Society of Nephrology

47. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al.: Improving reporting standards for polygenic scores in risk prediction studies. *Nature,* 591**:** 211-219, 2021 10.1038/s41586-021-03243-6

48. Choi SW, Mak TS, O'Reilly PF: Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc,* 15**:** 2759-2772, 2020 10.1038/s41596-020-0353-1

49. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM: Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet,* 14**:** 507-515, 2013 10.1038/nrg3457

50. Gorski M, Rasheed H, Teumer A, Thomas LF, Graham SE, Sveinbjornsson G, et al.: Genetic loci and prioritization of genes for kidney function decline derived from a meta-analysis of 62 longitudinal genome-wide association studies. *Kidney Int,* 102**:** 624-639, 2022 10.1016/j.kint.2022.05.021

51. Khan A, Kiryluk K: Kidney disease progression and collider bias in GWAS. *Kidney Int,* 102**:** 476-478, 2022 10.1016/j.kint.2022.06.018

52. Studies N-NWGoRiA, Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, et al.: Replicating genotype-phenotype associations. *Nature,* 447**:** 655-660, 2007 10.1038/447655a

53. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al.: Genome-wide association studies. *Nature Reviews Methods Primers,* 1**:** 59, 2021 10.1038/s43586-021-00056-9

Copyright © 2023 by the American Society of Nephrology

54. Umans BD, Battle A, Gilad Y: Where Are the Disease-Associated eQTLs? *Trends in Genetics,* 37**:** 109-124, 2021 https://doi.org/10.1016/j.tig.2020.08.009

55. Oh KH, Park SK, Park HC, Chin HJ, Chae DW, Choi KH, et al.: KNOW-CKD (KoreaN cohort study for Outcome in patients With Chronic Kidney Disease): design and methods. *BMC Nephrol,* 15**:** 80, 2014 10.1186/1471-2369-15-80

56. Moon S, Kim YJ, Han S, Hwang MY, Shin DM, Park MY, et al.: The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. *Sci Rep,* 9**:** 1382, 2019 10.1038/s41598-018-37832-9

57. Wang J, Leung KS, Chow SK, Cheung WH: Inflammation and age-associated skeletal muscle deterioration (sarcopaenia). *J Orthop Translat,* 10**:** 94-101, 2017 10.1016/j.jot.2017.05.006

58. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al.: Next-generation genotype imputation service and methods. *Nature Genetics,* 48**:** 1284-1287, 2016 10.1038/ng.3656

59. Fairley S, Lowy-Gallego E, Perry E, Flicek P: The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Research,* 48**:** D941-D947, 2019 10.1093/nar/gkz836

60. Delaneau O, Marchini J, Zagury J-F: A linear complexity phasing method for thousands of genomes. *Nature Methods,* 9**:** 179-181, 2012 10.1038/nmeth.1785

Copyright © 2023 by the American Society of Nephrology

61. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al.: A new equation to estimate glomerular filtration rate. *Ann Intern Med,* 150**:** 604-612, 2009 10.7326/0003-4819-150-9-200905050-00006

62. Shou H, Hsu JY, Xie D, Yang W, Roy J, Anderson AH, et al.: Analytic Considerations for Repeated Measures of eGFR in Cohort Studies of CKD. *Clin J Am Soc Nephrol,* 12**:** 1357-1365, 2017 10.2215/CJN.11311116

63. Oh KH, Kang M, Kang E, Ryu H, Han SH, Yoo TH, et al.: The KNOW-CKD Study: What we have learned about chronic kidney diseases. *Kidney Res Clin Pract,* 39**:** 121-135, 2020 10.23876/j.krcp.20.042

64. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ: Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience,* 4**:** 7, 2015 10.1186/s13742-015-0047-8

65. Turner SD: qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv***:** 005165, 2014 10.1101/005165

66. Team RC: R: A language and environment for statistical computing. 2013

67. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res,* 38**:** e164, 2010 10.1093/nar/gkq603

68. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al.: GeneCards Version 3: the human gene integrator. *Database,* 2010, 2010 10.1093/database/baq020

Copyright © 2023 by the American Society of Nephrology

69. 46th European Mathematical Genetics Meeting (EMGM) 2018, Cagliari, Italy, April 18-20, 2018: Abstracts. *Human Heredity,* 83**:** 1-29, 2018 10.1159/000488519

70. Myers TA, Chanock SJ, Machiela MJ: LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Front Genet,* 11**:** 157, 2020 10.3389/fgene.2020.00157

71. Moore CM, Jacobson SA, Fingerlin TE: Power and Sample Size Calculations for Genetic Association Studies in the Presence of Genetic Model Misspecification. *Hum Hered,* 84**:** 256-271, 2019 10.1159/000508558

72. Feldman HI, Appel LJ, Chertow GM, Cifelli D, Cizman B, Daugirdas J, et al.: The Chronic Renal Insufficiency Cohort (CRIC) Study: Design and Methods. *J Am Soc Nephrol,* 14**:** S148-153, 2003 10.1097/01.asn.0000070149.78399.ce

73. Marigorta UM, Navarro A: High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet,* 9**:** e1003566, 2013 10.1371/journal.pgen.1003566

74. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al.: Replicating genotype–phenotype associations. *Nature,* 447**:** 655-660, 2007 10.1038/447655a

75. Balduzzi S, Rücker G, Schwarzer G: How to perform a meta-analysis with R: a practical tutorial. *Evidence Based Mental Health,* 22**:** 153-160, 2019 10.1136/ebmental-2019-300117

76. Higgins JP, Thompson SG: Quantifying heterogeneity in a meta-analysis. *Stat Med,* 21**:** 1539-1558, 2002 10.1002/sim.1186

Copyright © 2023 by the American Society of Nephrology

77. Boughton AP, Welch RP, Flickinger M, VandeHaar P, Taliun D, Abecasis GR, et al.: LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics,* 37**:** 3017-3018, 2021 10.1093/bioinformatics/btab186

78. Schaid DJ, Chen W, Larson NB: From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet,* 19**:** 491-504, 2018 10.1038/s41576-018-0016-z

79. Wang G, Sarkar A, Carbonetto P, Stephens M: A simple new approach to variable selection in regression, with application to genetic fine-mapping. *bioRxiv***:** 501114, 2020 10.1101/501114

80. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al.: clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation,* 2**:** 100141, 2021 https://doi.org/10.1016/j.xinn.2021.100141

81. Qiu C, Huang S, Park J, Park Y, Ko Y-A, Seasock MJ, et al.: Renal compartment–specific genetic variation analyses identify new pathways in chronic kidney disease. *Nature Medicine,* 24**:** 1721-1731, 2018 10.1038/s41591-018-0194-4

82. Consortium EP: An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**:** 57-74, 2012 10.1038/nature11247

83. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al.: GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford),* 2017**:** bax028, 2017 10.1093/database/bax028

Copyright © 2023 by the American Society of Nephrology

84. Yu Z, Jin J, Tin A, Kottgen A, Yu B, Chen J, et al.: Polygenic Risk Scores for Kidney Function and Their Associations with Circulating Proteome, and Incident Kidney Diseases. *J Am Soc Nephrol*, 2021 10.1681/ASN.2020111599

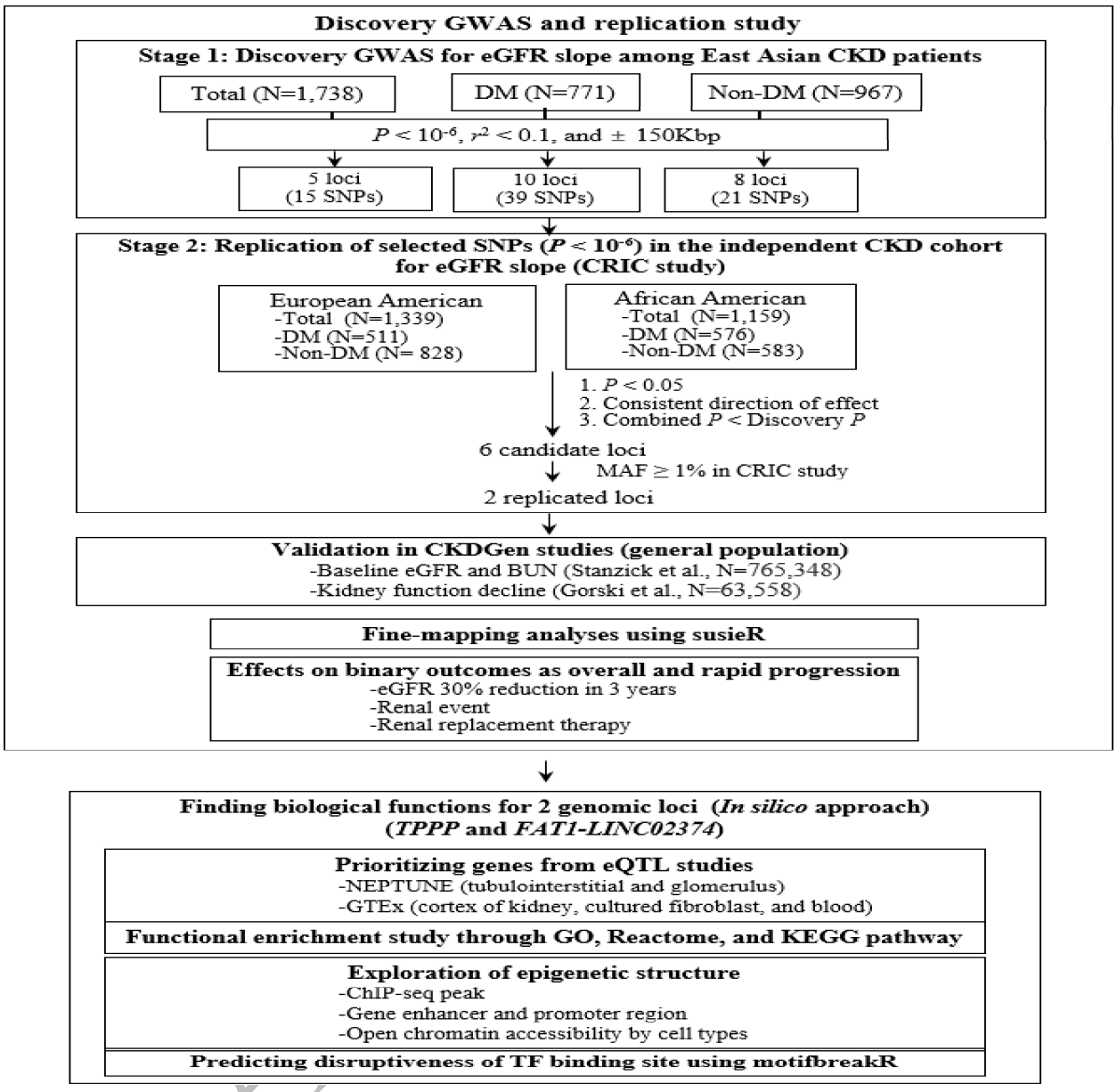
ACCEPTED

Figure legends

Figure 1. Overall study flow.

GWAS, Genome-wide association study; eGFR, estimated glomerular filtration rate; CKD, Chronic kidney disease; KNOW-CKD, Korean cohort study for outcome in patients with chronic kidney disease; DM, diabetes mellitus; CRIC, Chronic renal insufficiency cohort; MAF, minor allele frequency; CKDGen, Chronic Kidney Disease Genetics; BUN, blood urea nitrogen; eQTL, expression quantitative trait loci; NEPTUNE, Nephrotic Syndrome Study Network; GTEx, Genotype-Tissue Expression; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ChIP-seq, chromatin immunoprecipitation-sequencing

Copyright © 2023 by the American Society of Nephrology

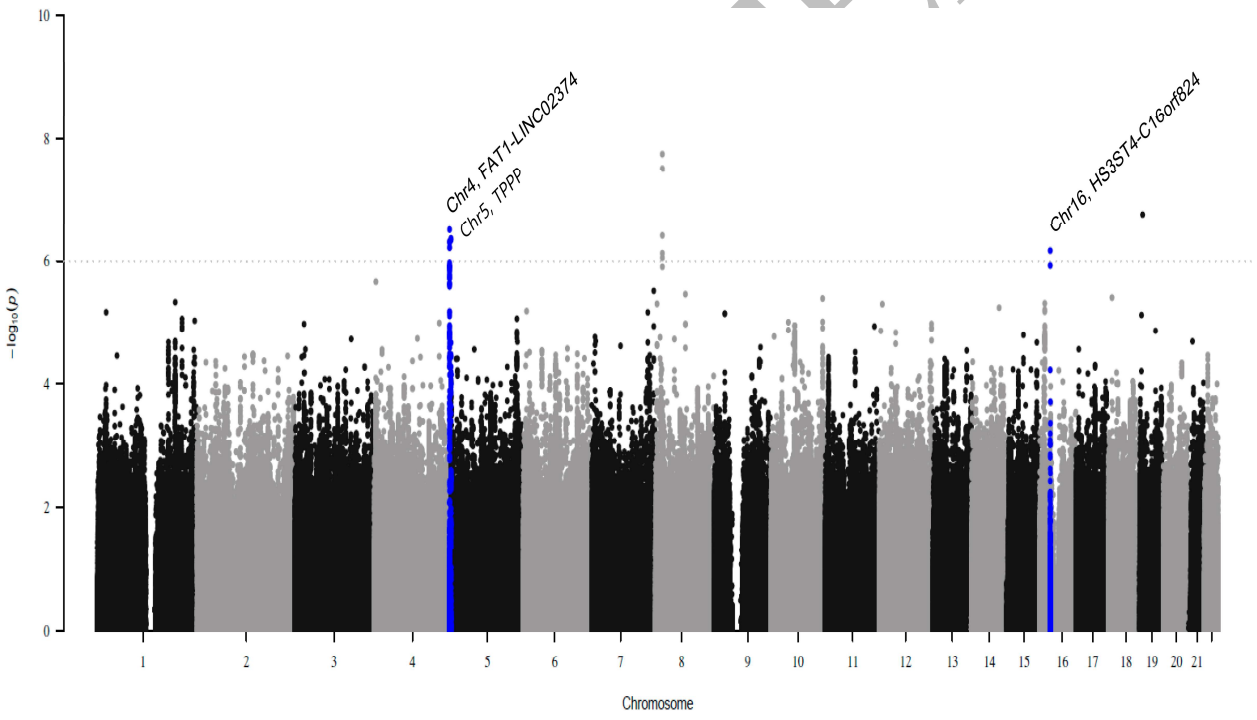


Copyright © 2023 by the American Society of Nephrology

Figure 2. Manhattan plots for discovery genome wide association studies for eGFR slope among CKD patients in Total (N=1,738)

The x axis shows genomic position (chromosomes 1–22), and the y axis shows statistical significance as $-\log_{10}(P\text{-value})$. The horizontal dotted grey lines indicate the thresholds of *P*-value of 10^{-6}. SNPs in genomic loci with *P*-value < 10^{-6} and replicated or candidate loci from CRIC study were colored red.

eGFR, estimated glomerular filtration rate; CKD, chronic kidney disease; CRIC, Chronic Renal Insufficiency Cohort study



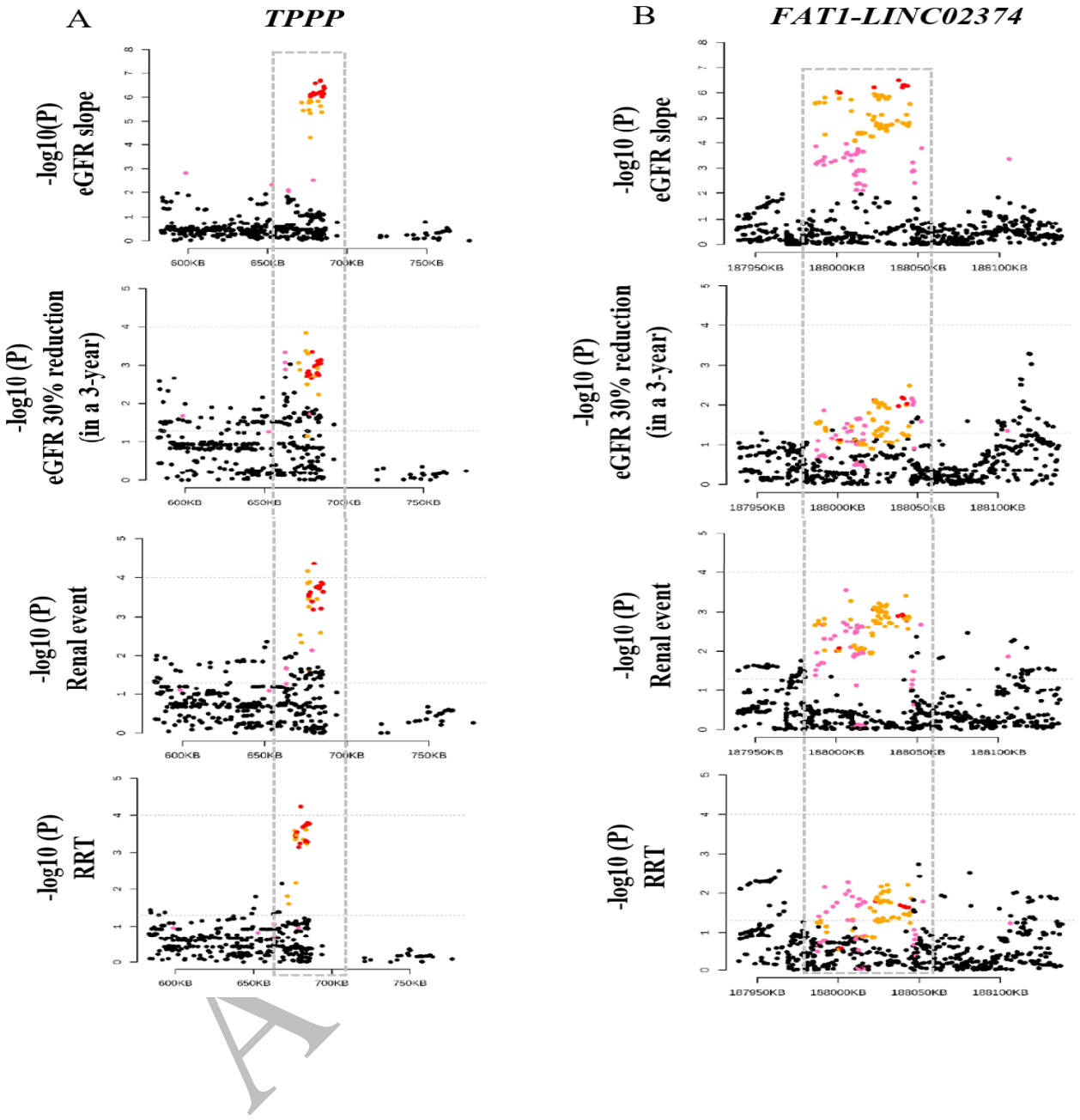
Copyright © 2023 by the American Society of Nephrology

Figure 3. Associations between binary outcomes as chronic kidney disease progression and the *TPPP* and *FAT1-LINC02374* loci. (A) Regional associations of SNPs on *TPPP* on eGFR slope, 30% reduction of eGFR in a 3 year, renal event, and renal replacement therapy (red dots, $P < 10^{-6}$; orange dots, $P < 10^{-4}$; pink dots, $P < 0.05$ in eGFR slope GWAS among non-DM group). The x axis shows genomic position around *TPPP* loci, and the y axis shows statistical significance as $-\log_{10}$(*P*-value) for each trait. (B) Regional associations of SNPs on *FAT1-LINC02374* in eGFR slope, 30% reduction of eGFR in a 3-year, renal event, and renal replacement therapy in overall period (red dots, $P < 10^{-6}$; orange dots, $P < 10^{-4}$; pink dots, $P < 0.05$ in eGFR slope GWAS among total group). The x axis shows genomic position around *FAT1-LINC02374* loci, and the y axis shows statistical significance as $-\log_{10}$(*P*-value) for each trait.

SNP, single nucleotide polymorphism, eGFR, estimated glomerular filtration rate; RRT, renal replacement therapy, DM, diabetes mellitus; GWAS, genome-wide association study

ACCEPTED

Copyright © 2023 by the American Society of Nephrology



Copyright © 2023 by the American Society of Nephrology

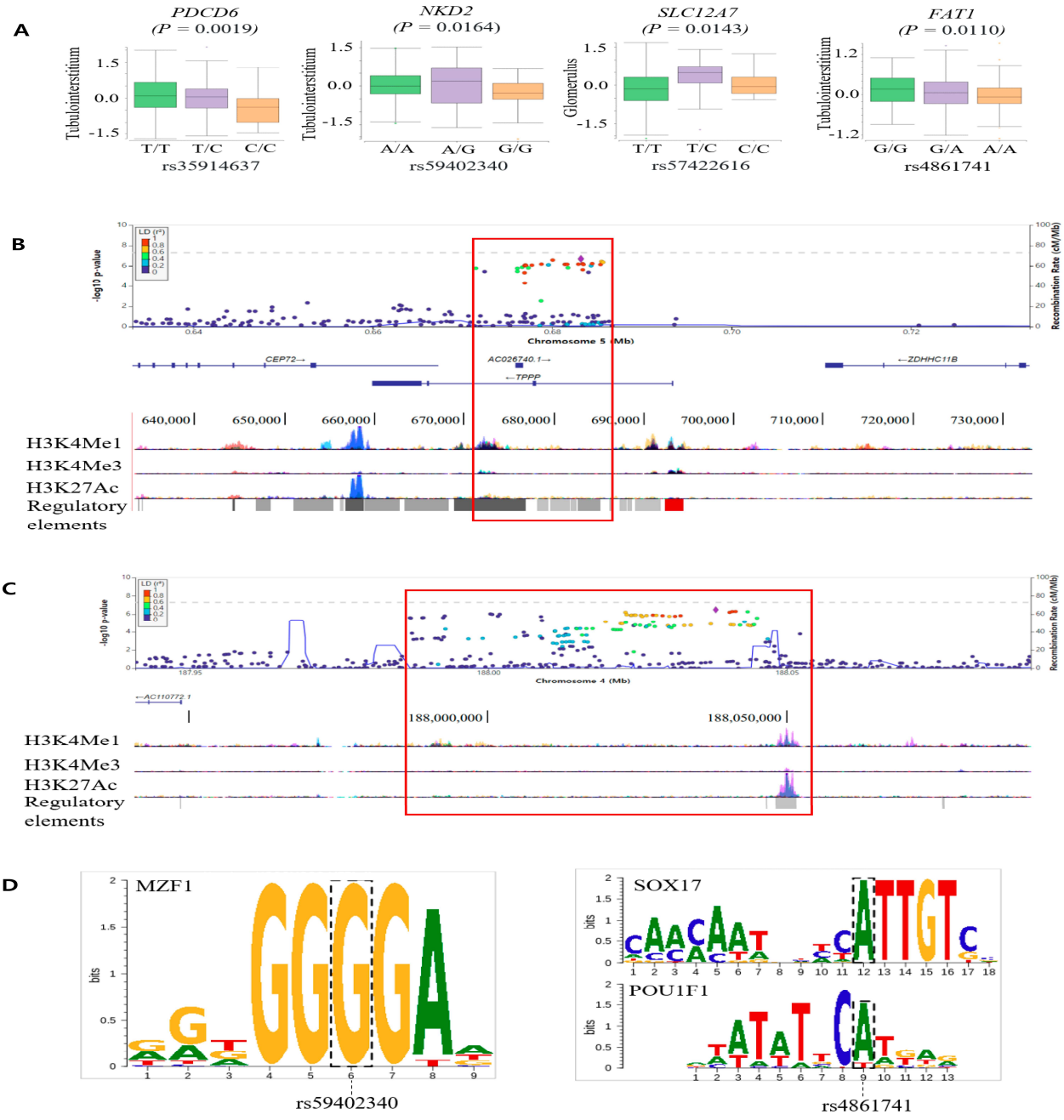
Figure 4. Finding for functional consequences for *TPPP* and *FAT1-LINC02374* loci.

(A) eQTL association plots for selected pairs of SNPs and genes from NEPTUNE cohort. The y-axis show rank normalized, adjusted expression for each gene by each pair of alleles on the x-axis. Histone activities and regulatory elements around (B) *TPPP* and (C) *FAT1-LINC02374* locus. Top, regional association plots for each locus. The plots show the $-\log_{10}$(*P*-value) of each SNP on the y-axis from GWAS on eGFR slopes in non-DM group for *TPPP* locus and total group for *FAT1-LINC02374* locus with genomic positions on the X-axis. Genes overlapping the locus are displayed below the plot. SNPs are colored by their LD value r^2 estimated from 1000 Genome phase 3 Population with the top significant SNP marked as a purple diamond. Middle, histone activities for H3K4Me1, H3K4Me3, and H3K27AC of 7 cell lines from ENCODE. Bottom, gene regulatory elements from GeneHancer. Darkness indicates confidence score for enhancers, red ones indicate promotors (Deeper ones show higher confidence scores) (D) Left, position weight matrix for MZF1 motif. rs59402340 (*TPPP*) are predicted to disrupt binding affinity of MZF1 from HOCOMOCO database. Right, position weight matrix for SOX17 and POU1F1 motifs. rs4861741 (*FAT1-LINC02374*) are predicted to disrupt their binding from HOCOMOCO database.

eQTL, expression quantitative trait loci; SNP, single nucleotide polymorphism; GWAS, genome-wide association study; NEPTUNE, Nephrotic Syndrome Study Network; DM, diabetes mellitus; LD, linkage disequilibrium; H3K4Me1, monomethylation of lysine 4 on histone H3 protein subunit; H3K4Me3, trimethylation of lysine 4 on histone H3 protein subunit; H3K27AC, acetylation of lysine 27 on histone H3 protein subunit; ENCODE, Encyclopedia of DNA Elements; HOCOMOCO, HOmo sapiens COmprehensive MOdel COllection

ACCEPTED

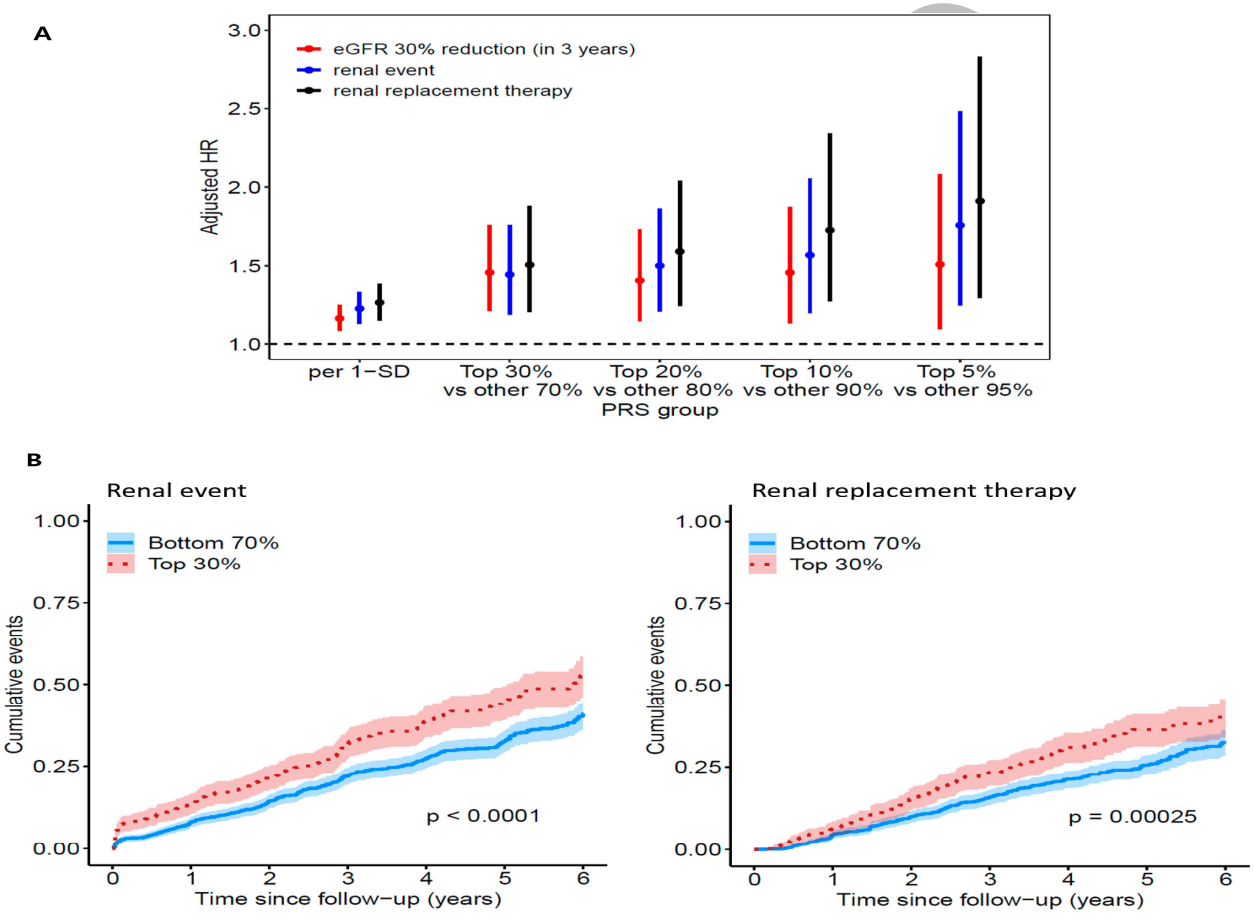
Copyright © 2023 by the American Society of Nephrology



Copyright © 2023 by the American Society of Nephrology

Figure 5. Effect of polygenic risk score derived from top 22 significant SNPs from eGFR slope on CKD outcomes. (A) Multivariable adjusted hazard ratio of binary CKD outcomes by 1-SD increment of PRS and tail-based cut-offs (B) Cumulative incidence curve for top 30% versus others with renal events and receiving renal replacement therapy

eGFR, estimated glomerular filtration rate; CKD, chronic renal disease; SD, standard deviation; RRT, renal replacement therapy, GWAS, genome-wide association study



Copyright © 2023 by the American Society of Nephrology

Table 1. Selected characteristics of total CKD, diabetic CKD, and non-diabetic CKD patients in the discovery cohort[1]

	CKD, total (N=1,738)	Diabetes (N=771)	Non-diabetes (N=967)	
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	***P***
Age at baseline	54.9 (12.1)	58.1 (10.6)	52.4 (12.7)	< 0.001
Systolic BP (mmHg)	127.7 (17.7)	131.9 (20.3)	124.8 (14.9)	< 0.001
Diastolic BP (mmHg)	76.3 (11.7)	76.1 (13.1)	76.6 (10.7)	0.416
Body mass index (kg/m^2)	24.7 (6.0)	25.4 (7.8)	24.1 (4.0)	< 0.001
White blood cells (1,000/mm^3)	6.53 (2.32)	6.55 (2.65)	6.51 (2.02)	0.737
Hemoglobin (g/dL)	12.5 (3.5)	12.0 (4.3)	13.0 (2.5)	< 0.001
Albumin (mg/dL)	4.0 (0.6)	3.9 (0.7)	4.2 (0.5)	< 0.001
24-hour urine protein (g)	1.8 (2.5)	2.6 (3.2)	1.1 (1.5)	< 0.001
eGFR (ml/min/1.73m^2)	48.9 (28.5)	42.0 (25.3)	54.4 (29.8)	< 0.001
eGFR slope	-2.1 (2.1)	-2.5 (2.1)	-1.7 (1.9)	< 0.001
	Median (IQR)	**Median (IQR)**	**Median (IQR)**	***P***
Follow up (years)	3.3 (2.0–5.1)	2.6 (1.6–4.3)	4.1 (2.6–5.4)	< 0.001
	N (%)	**N (%)**	**N (%)**	***P***
Sex (male)	1,101 (63.3)	517 (67.1)	584 (60.4)	0.004
Hypertension	1,674 (96.3)	730 (94.7)	944 (97.6)	0.001

Abbreviations: CKD, chronic kidney disease; SD, standard deviation; BP, blood pressure; eGFR, estimated glomerular filtration rate; IQR, interquartile range

1. CKD patients derived from the Korean cohort study for outcomes in patients with chronic kidney disease (KNOW-CKD) and diabetic CKD patients proven by biopsy in Seoul National University Hospital Human Biobank and Kyung Hee University Medical Center

Copyright © 2023 by the American Society of Nephrology

Table 2. Replication results [1] from the CRIC study for the top significant GWAS signals in 6 independent loci associated with eGFR slope in the discovery cohort

SNP	Alleles	Nearby gene(s)	Cytoband	Position	Study	Group	EAF (%)	Beta	SE	P [2]	Meta-analysis [3]		
											Beta	SE	P
Replicated loci													
rs61464149	A/G	*TPPP*	5p15.33	683210	Discovery	N-DM	5.4	-1.09	0.21	2.56×10^{-7}			
					CRIC (EA)	N-DM	2.3	-0.53	0.23	1.87×10^{-2}	-0.83	0.15	6.49×10^{-8}
rs56697078	C/T	*TPPP*	5p15.33	683207	Discovery	N-DM	5.4	-1.09	0.21	2.56×10^{-7}			
					CRIC (EA)	N-DM	2.5	-0.52	0.22	1.88×10^{-2}	-0.82	0.15	7.82×10^{-8}
rs57422616	T/C	*TPPP*	5p15.33	683182	Discovery	N-DM	5.3	-1.10	0.21	2.74×10^{-7}			
					CRIC (EA)	N-DM	3.7	-0.45	0.18	1.17×10^{-2}	-0.72	0.14	1.36×10^{-7}
rs112671796	G/A	*TPPP*	5p15.33	678808	Discovery	N-DM	6.9	-0.93	0.19	7.90×10^{-7}			
					CRIC (EA)	N-DM	6.7	-0.35	0.12	4.22×10^{-3}	-0.53	0.10	2.93×10^{-7}
rs6886251	A/G	*TPPP*	5p15.33	685395	Discovery	N-DM	6.2	-0.90	0.18	8.79×10^{-7}			
					CRIC (EA)	N-DM	8.5	-0.35	0.1	9.51×10^{-4}	-0.48	0.09	9.61×10^{-8}
rs59402340	A/G	*TPPP*	5p15.33	684463	Discovery	N-DM	6.2	-0.92	0.18	7.11×10^{-7}			
					CRIC (EA)	N-DM	8.6	-0.35	0.10	8.13×10^{-4}	-0.49	0.09	7.23×10^{-8}
rs55957764	G/A	*TPPP*	5p15.33	682926	Discovery	N-DM	6.2	-0.92	0.18	8.09×10^{-7}			
					CRIC (EA)	N-DM	8.2	-0.34	0.11	1.57×10^{-3}	-0.48	0.09	1.73×10^{-7}
rs56352513	C/T	*TPPP*	5p15.33	681456	Discovery	N-DM	6.2	-0.92	0.18	8.44×10^{-7}			
					CRIC (EA)	N-DM	8.2	-0.33	0.11	1.72×10^{-3}	-0.48	0.09	2.01×10^{-7}
rs59183188	C/T	*TPPP*	5p15.33	681425	Discovery	N-DM	6.2	-0.92	0.18	8.44×10^{-7}			
					CRIC (EA)	N-DM	8.2	-0.33	0.11	1.72×10^{-3}	-0.48	0.09	2.02×10^{-7}
rs73030873	C/T	*TPPP*	5p15.33	681675	Discovery	N-DM	6.2	-0.92	0.18	8.27×10^{-7}			
					CRIC (EA)	N-DM	8.2	-0.33	0.11	1.71×10^{-3}	-0.48	0.09	1.96×10^{-7}
rs6863036	G/T	*TPPP*	5p15.33	685666	Discovery	N-DM	6.1	-0.93	0.18	5.38×10^{-7}			
					CRIC (EA)	N-DM	8.1	-0.34	0.11	1.64×10^{-3}	-0.48	0.09	1.45×10^{-7}
rs6862648	G/A	*TPPP*	5p15.33	685504	Discovery	N-DM	5.9	-0.93	0.18	4.56×10^{-7}			
					CRIC (EA)	N-DM	8.2	-0.34	0.11	1.49×10^{-3}	-0.48	0.09	1.17×10^{-7}
rs57966672	G/A	*TPPP*	5p15.33	683483	Discovery	N-DM	7.4	-0.90	0.18	8.60×10^{-7}			
					CRIC (EA)	N-DM	11.0	-0.31	0.1	2.23×10^{-3}	-0.45	0.09	3.37×10^{-7}
rs28629773	G/C	*FAT1-LINC02374*	4q35.2	188040427	Discovery	Total	46.4	0.37	0.07	6.08×10^{-7}			
					CRIC (AA)	N-DM	37.5	0.20	0.10	4.33×10^{-2}	0.31	0.06	1.87×10^{-7}
Candidate loci													
rs74798667	T/C	*ADGRV1*	5q14.3	90034503	Discovery	DM	3.0	-1.90	0.32	7.01×10^{-9}			
					CRIC (AA)	Total	< 0.1	-13.16	5.28	1.29×10^{-2}	-1.94	0.33	2.26×10^{-9}
rs79983000	C/T	*HS3ST4-C16orf8*24	16p12.1	26320278	Discovery	Total	7.9	0.71	0.14	7.14×10^{-7}			
					CRIC (AA)	DM	0.1	8.42	4.11	4.10×10^{-2}	0.72	0.14	4.34×10^{-7}
rs138320114	G/A	*SLC2A1*	1p34.2	43399030	Discovery	DM	1.3	-2.42	0.48	5.40×10^{-7}			
					CRIC (AA)	DM	< 0.1	-70.64	34.37	4.03×10^{-2}	-2.45	0.48	4.03×10^{-7}
rs150696522	C/T	*FAM124B-SERPINE2*	2q36.1	225168505	Discovery	DM	1.3	-2.97	0.55	7.26×10^{-8}			
					CRIC (AA)	N-DM	< 0.1	-92.97	39.51	1.90×10^{-2}	-3.05	0.55	3.00×10^{-8}
					CRIC (AA)	Total	0.1	-54.66	27.73	4.90×10^{-2}	-3.05	0.55	2.91×10^{-8}

Abbreviations: CRIC, Chronic Renal Insufficiency Cohort study; GWAS, genome-wide association study; eGFR, estimated glomerular filtration rate; SNP, single nucleotide polymorphism; CKD, chronic kidney disease; CRIC, Chronic renal insufficiency cohort; DM, diabetes mellitus; EAF, effect allele frequency; SE, standard error; EA, European American; AA, African American; DM, diabetes mellitus group; N-DM, non-diabetes mellitus group

1. Criteria for candidate loci were defined as 'P-value in the CRIC < 0.05', 'a consistent direction of effect between discovery cohort and CRIC', and 'meta-analysis P-value < Discovery P-value'. The criteria for the replication loci were defined as 'the above criteria + EAF ≥ 1%'.

2. Adjusted for age, sex and 5 principal components.

3. Fixed effect model

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Table 3. Replication results [1] from the CKDGen [2] consortium for the top significant GWAS signals in *TPPP* and *FAT1-LINC02374* loci associated with eGFR slope in the discovery cohort

SNP	Alleles	Nearby gene(s)	Cytoband	Position [3]	Study/Ethnicity	Phenotype	EAF (%)	Beta	SE	*P* [4]
rs35914637	T/C	*TPPP*	5p15.33	677348	Discovery-total	eGFR slope	7.6	-0.70	0.14	4.59×10^{-7}
					CKDGen/Trans-ethnic	Baseline eGFR	24.9	-0.0012	0.0003	1.15×10^{-3}
					CKDGen/EA	Baseline eGFR	25.4	-0.0012	0.0003	1.17×10^{-3}
rs13103811	A/G	*FAT1-LINC02374*	4q35.2	188038110	Discovery-total	eGFR slope	54.6	-0.381	0.074	3.07×10^{-7}
					CKDGen/EA	Baseline BUN	76.6	0.0022	0.0010	1.97×10^{-2}
rs11132426	G/A	*FAT1-LINC02374*	4q35.2	188041427	Discovery-total	eGFR slope	53.4	-0.381	0.075	4.79×10^{-7}
					CKDGen/EA	Baseline BUN	77.6	0.0022	0.0010	1.80×10^{-2}
rs4861741	G/A	*FAT1-LINC02374*	4q35.2	188040879	Discovery-total	eGFR slope	53.4	-0.38	0.08	4.91×10^{-7}
					CKDGen/EA	Baseline BUN	76.6	0.0022	0.0010	1.76×10^{-2}
rs6553037	G/C	*FAT1-LINC02374*	4q35.2	188022986	Discovery-total	eGFR slope	45.7	0.353	0.070	6.03×10^{-7}
					CKDGen/EA	eGFR change	9.2	-0.0617	0.0268	2.16×10^{-2}
					CKDGen/EA	Baseline BUN	22.0	-0.0017	0.0008	3.20×10^{-2}
rs28629773	G/C	*FAT1-LINC02374*	4q35.2	188040427	Discovery-total	eGFR slope	46.4	0.375	0.075	6.08×10^{-7}
					CKDGen/EA	Baseline BUN	22.8	-0.0021	0.0010	2.63×10^{-2}

Abbreviations: CKDGen, Chronic Kidney Disease Genetics; CKDGen, Chronic Kidney Disease Genetics; GWAS, genome-wide association study; eGFR, estimated glomerular filtration rate; SNP, single nucleotide polymorphism; EAF, effect allele frequency; SE, standard error; BUN, blood urea nitrogen

1. We selected SNPs related to baseline eGFR and BUN levels from GWAS statistics profiles in the CKDGen consortium and the selection criteria were as 'P-value the CKDGen < 0.05' and 'EAF ≥ 1%'.

2. We used the summary statistics from published CKDGen Consortium using general population: for baseline eGFR and BUN (Stanzick et al., 2021 (N=765,348) [#Ref. 18]); for eGFR change from baseline to the last follow-up (Gorski et al., 2015 (N=63,558) [#Ref. 19])

3. In discovery cohort, GWAS P-values were adjusted for age, sex and 5 principal components.

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

Table 4. CKD progression associated with the PRS [1] composed of 22 top significant SNPs and mean 6,094 SNPs selected by C+T method in the KNOW-CKD

	Validation of PRS from 22 top significant SNPs in the total cohort								PRS with mean 6,094 SNPs by C+T method (LOGO Cross-validation) [2]							
	Crude				Multivariable [3]				Crude				Multivariable [2]			
	Beta	**SE**	***P***	**R^2**	**Beta**	**SE**	***P***	**R**	**Beta**	**SE**	***P***	**R^2**	**Beta**	**SE**	***P***	**R^2**
eGFR slope: Per 1-SD	-0.63	0.05	9.76×10^{-37}	0.098	-0.54	0.04	1.06×10^{-32}	0.249	-1.04	0.07	6.34×10^{-56}	0.241	-0.83	0.07	1.70×10^{-34}	0.341
	HR	**95% CI**	***P***	**C**	**HR**	**95% CI**	***P***	**C**	**HR**	**95% CI**	***P***	**C**	**HR**	**95% CI**	***P***	**C**
Rapid deterioration [4]																
Per 1-SD	1.28	1.19–1.38	1.07×10^{-10}	0.568	1.16	1.08–1.25	3.35×10^{-5}	0.784	2.01	1.76–2.30	5.12×10^{-25}	0.685	1.65	1.42–1.92	1.35×10^{-10}	0.829
Top 30% vs. other 70%	1.59	1.32–1.90	6.09×10^{-7}		1.46	1.21–1.76	7.28×10^{-5}		3.39	2.61–4.40	4.38×10^{-20}		2.48	1.86–3.31	6.25×10^{-10}	
Top 20% vs. other 80%	1.65	1.35–2.01	9.23×10^{-7}		1.41	1.14–1.73	1.20×10^{-3}		3.74	2.82–4.96	7.00×10^{-20}		2.65	1.90–3.68	6.86×10^{-9}	
Top 10% vs. other 90%	1.80	1.41–2.31	2.85×10^{-6}		1.46	1.13–1.88	3.52×10^{-3}		5.20	3.67–7.36	1.40×10^{-20}		2.81	1.83–4.32	2.13×10^{-6}	
Top 5% vs. other 95%	2.20	1.62–3.00	5.92×10^{-7}		1.51	1.09–2.08	1.21×10^{-2}		4.97	3.14–7.85	6.70×10^{-12}		2.89	1.64–5.09	2.39×10^{-4}	
Long-term progression [5]																
Per 1-SD	1.28	1.18–1.39	7.32×10^{-9}	0.565	1.22	1.13–1.33	1.08×10^{-6}	0.862	2.03	1.74–2.36	5.63×10^{-20}	0.684	1.87	1.54–2.26	1.41×10^{-10}	0.892
Top 30% vs. other 70%	1.46	1.21–1.77	9.82×10^{-5}		1.44	1.19–1.76	2.45×10^{-4}		3.08	2.31–4.12	2.50×10^{-14}		2.75	1.94–3.91	1.55×10^{-8}	
Top 20% vs. other 80%	1.59	1.29–1.96	1.64×10^{-5}		1.50	1.21–1.87	2.67×10^{-4}		3.16	2.28–4.38	4.18×10^{-12}		2.76	1.80–4.23	3.06×10^{-6}	
Top 10% vs. other 90%	1.77	1.36–2.31	1.95×10^{-5}		1.57	1.20–2.05	1.10×10^{-3}		6.12	4.15–9.01	5.03×10^{-20}		4.52	2.59–7.88	1.01×10^{-7}	
Top 5% vs. other 95%	2.31	1.66–3.21	5.94×10^{-7}		1.76	1.24–2.48	1.44×10^{-3}		6.95	4.09–11.8	6.94×10^{-13}		7.50	3.50–16.06	2.12×10^{-7}	
ESRD [6]																
Per 1-SD	1.24	1.12–1.36	1.31×10^{-5}	0.556	1.26	1.15–1.39	1.20×10^{-6}	0.909	1.81	1.54–2.13	1.48×10^{-12}	0.653	1.68	1.34–2.11	7.85×10^{-6}	0.923
Top 30% vs. other 70%	1.39	1.12–1.73	3.26×10^{-3}		1.51	1.20–1.88	3.35×10^{-4}		2.61	1.87–3.64	1.80×10^{-8}		2.27	1.48–3.49	1.83×10^{-4}	
Top 20% vs. other 80%	1.49	1.17–1.90	1.30×10^{-3}		1.59	1.24–2.04	2.69×10^{-4}		2.72	1.87–3.95	1.41×10^{-7}		2.61	1.57–4.33	2.06×10^{-4}	
Top 10% vs. other 90%	1.74	1.29–2.35	2.97×10^{-4}		1.72	1.27–2.34	4.68×10^{-4}		4.80	3.09–7.45	2.88×10^{-12}		3.37	1.76–6.44	2.33×10^{-4}	
Top 5% vs. other 95%	2.07	1.41–3.03	1.79×10^{-4}		1.91	1.29–2.83	1.18×10^{-3}		5.26	2.94–9.38	2.00×10^{-8}		8.01	3.34–19.2	3.07×10^{-6}	

Abbreviations: CKD, chronic kidney disease; PRS, Polygenic risk score; SNP, Single nucleotide polymorphism; C+T method, LD-clumping and P-value thresholding method; KNOW-CKD, KoreaN Cohort Study for Outcomes in Patients With Chronic Kidney Disease; eGFR, estimated glomerular filtration rate;
SE, Standard error; HR, hazard ratio; CI, confidence interval; C, C-statistic; ESRD, End-stage renal disease
1. The PRS was standardized to a normal distribution with mean=0 and SD=1. The analyses were performed in 1,550 CKD patients without no missing all covariates.
2. C+T method with LOGO (leave-one-group-out) cross-validation was that to select SNPs in PRS model was used with the method for the calculation of PRS using informed LD clumping and P-value thresholding in 4 data of 5-split datasets and then performed with GWAS discovery analysis in the half of leave-one-group-out (LOGO) and cross-validation in the other half of LOGO. This shows the result in C+T with p-value cutoff=0.005 and R^2<0.1.

ACCEPTED

Copyright © 2023 by the American Society of Nephrology

3. Adjusted for age, sex, 5 component principals, baseline eGFR, urine albumin creation ratio, status of diabetes, systolic blood pressure, and body mass index, and subjects with missing covariates were excluded. The follow-up times until the 6-year for renal event and renal replacement therapy, and 3-year for 30% reduction of eGFR trait.
4. Rapid deterioration was defined as 30% reduction of eGFR in 3-year from the baseline eGFR
5. Long-term progression of CKD was defined as new renal events across total follow-up period, a composite outcome of the first occurrence of a 50% decline in eGFR from the baseline eGFR, 2-fold increase in serum creatinine, or the start of ESRD (initiation of renal replacement therapy, such as dialysis or kidney transplantation) across the follow-up period.
6. ESRD was defined as the onset of renal replacement therapy including dialysis or kidney transplantation

ACCEPTED

Copyright © 2023 by the American Society of Nephrology