



OPEN

## Epigenome-wide association study of diabetic chronic kidney disease progression in the Korean population: the KNOW-CKD study

Hye Youn Sung<sup>1,10</sup>, Sangjun Lee<sup>2,3,4,10</sup>, Miyeun Han<sup>5,10</sup>, Woo Ju An<sup>2,3,6</sup>, Hyunjin Ryu<sup>7</sup>, Eunjeong Kang<sup>8</sup>, Yong Seek Park<sup>9</sup>, Seung Eun Lee<sup>9</sup>, Curie Ahn<sup>5,7</sup>, Kook-Hwan Oh<sup>7</sup>✉, Sue K. Park<sup>2,3,6</sup>✉ & Jung-Hyuck Ahn<sup>1</sup>✉

Since the etiology of diabetic chronic kidney disease (CKD) is multifactorial, studies on DNA methylation for kidney function deterioration have rarely been performed despite the need for an epigenetic approach. Therefore, this study aimed to identify epigenetic markers associated with CKD progression based on the decline in the estimated glomerular filtration rate in diabetic CKD in Korea. An epigenome-wide association study was performed using whole blood samples from 180 CKD recruited from the KNOW-CKD cohort. Pyrosequencing was also performed on 133 CKD participants as an external replication analysis. Functional analyses, including the analysis of disease-gene networks, reactome pathways, and protein–protein interaction networks, were conducted to identify the biological mechanisms of CpG sites. A phenome-wide association study was performed to determine the associations between CpG sites and other phenotypes. Two epigenetic markers, cg10297223 on *AGTR1* and cg02990553 on *KRT28* indicated a potential association with diabetic CKD progression. Based on the functional analyses, other phenotypes (blood pressure and cardiac arrhythmia for *AGTR1*) and biological pathways (keratinization and cornified envelope for *KRT28*) related to CKD were also identified. This study suggests a potential association between the cg10297223 and cg02990553 and the progression of diabetic CKD in Koreans. Nevertheless, further validation is needed through additional studies.

Diabetic chronic kidney disease (CKD) is a common cause of end-stage kidney disease. In Korea, almost 50% of end-stage kidney disease (ESKD) cases are caused by diabetic CKD<sup>1</sup>. Diabetic CKD has a poor prognosis, showing increased mortality and rapid progression to ESKD compared to non-diabetic CKD<sup>2,3</sup>.

The etiology of diabetic CKD is multifactorial, including both genetic and environmental factors. High blood glucose levels and blood pressure, prolonged activation of the renin–angiotensin–aldosterone system, and obesity are risk factors associated with the progression of diabetic CKD. However, most of the variability remains unaccounted for by conventional risk factors<sup>4</sup>. For instance, many patients with diabetic CKD with poor glycemic control do not develop renal complications. This discrepancy can be attributed to genetic or epigenetic factors. Genetic codes explain only a fraction of diabetic CKD development, and epigenetic programming, remodeling, and post-translational modifications, such as advanced glycation end products, have been regarded recently as possible physiological mechanisms<sup>5</sup>.

<sup>1</sup>Department of Biochemistry, Ewha Womans University College of Medicine, 25 Magokdong-ro 2-gil, Gangseo-gu, Seoul 07804, South Korea. <sup>2</sup>Department of Preventive Medicine, Seoul National University College of Medicine, 103, Daehak-ro, Jongro-gu, Seoul 03080, South Korea. <sup>3</sup>Cancer Research Institute, Seoul National University, Seoul, South Korea. <sup>4</sup>Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, South Korea. <sup>5</sup>Department of Internal Medicine, National Medical Center, Seoul, South Korea. <sup>6</sup>Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, Seoul, South Korea. <sup>7</sup>Department of Internal Medicine, Seoul National University Hospital, 103, Daehak-ro, Jongro-gu, Seoul 03080, Republic of Korea. <sup>8</sup>Department of Internal Medicine, Ewha Womans University Seoul Hospital, Ewha Womans University College of Medicine, Seoul, South Korea. <sup>9</sup>Department of Microbiology, School of Medicine, Kyung Hee University, Seoul, South Korea. <sup>10</sup>These authors contributed equally: Hye Youn Sung, Sangjun Lee and Miyeun Han. ✉email: khoh@snu.ac.kr; suepark@snu.ac.kr; ahnj@ewha.ac.kr

Epigenetic changes, including DNA methylation, histone modification, and miRNA regulation, play major roles in the pathogenesis of diabetic CKD. DNA methylation directly impacts human genome function and serves as a critical regulatory function<sup>6</sup>. Epigenetic changes are thought to be important in determining a predisposition to diabetic CKD. In addition to elevated glucose levels, reactive oxygen species, hypoxia, inflammation, cytokines, drugs, nutrition, and physical activity can also modify epigenetic profiles<sup>5</sup>. Several studies have investigated the DNA methylation profile in diabetic CKD<sup>7–9</sup>. However, most studies were case–control studies, and DNA methylation profiles of renal function decline in diabetic CKD are scarce.

The Korean Cohort Study for Outcomes in Patients with Chronic Kidney Disease (KNOW-CKD) is the largest CKD cohort in Korea to establish the clinical course, risk factors, and adverse outcomes of CKD. More than 50 articles have been published on clinical markers, such as anemia, mineral bone disease, quality of life, and serum biomarkers, such as FGF23, adiponectin, and hepcidin, of CKD in the Korean population using the KNOW-CKD cohort<sup>10</sup>. However, epigenetic biomarkers for the decline in the estimated glomerular filtration rate (eGFR) have not been assessed in the KNOW-CKD cohort. In this study, we aimed to identify epigenetic biomarkers associated with the rapid decline in the eGFR observed in KNOW-CKD subjects using blood samples of diabetic CKD.

## Material and methods

**Ethics and inclusion statement.** We acquired written informed consent and blood samples from all participants, and the study was approved by the Institutional Review Boards of Seoul National University Hospital (H-1805-168-948). We also confirmed that all methods were carried out in accordance with relevant guidelines and regulations.

**Data source and study population.** The KNOW-CKD cohort was used to perform an EWAS for diabetic CKD progression. The KNOW-CKD study is a prospective multicenter cohort involving 2238 participants with specific causes of chronic kidney disease (CKD) grouped into glomerulonephritis (GN) (n = 810), diabetic nephropathy (DN) (n = 519), hypertensive nephropathy (HTN) (n = 409), polycystic kidney disease (PKD) (n = 364), and unclassified (n = 136) (Supplementary Fig. 1)<sup>10</sup>. Subgroups were defined by pathologic diagnosis, where available, otherwise by clinical diagnosis<sup>11</sup>. GN was identified by the presence of glomerular hematuria or albuminuria, with or without an underlying systemic disease. DN was diagnosed based on albuminuria in individuals with type 2 diabetes mellitus and diabetic retinopathy. HTN was determined by hypertension history and absence of systemic illness associated with renal damage. PKD was diagnosed using unified ultrasound criteria. Other causative diseases were classified as ‘unclassified.’ The KNOW-CKD cohort has been described in detail elsewhere<sup>11</sup>.

The population was divided into non-progression and progression, based on the eGFR slope of  $-2.6$  ml/min/1.73 m<sup>2</sup>/year, which was calculated as the median the eGFR slope in the DN from the KNOW-CKD cohort. In addition, the threshold of eGFR slope for HTN was  $-2.1$  ml/min/1.73 m<sup>2</sup>/year, while it was  $-2.6$  ml/min/1.73 m<sup>2</sup>/year for DN in a previous study<sup>12</sup>.

This study was a follow-up to the genome-wide association study (GWAS) of the KNOW-CKD group<sup>13</sup>. Out of 519 participants with DN from KNOW-CKD, 434 individuals passed the quality control (QC) for GWAS (Supplementary Fig. 1).

We estimated the appropriate number of participants to be included in the EWAS based on power calculation<sup>14</sup>. Statistical power was estimated from a minimum of 20 to a maximum of 200 participants, assuming a 1:1 ratio of progression to non-progression, 780,000 total CpG sites, 800 targeted CpG sites, minimum detection of  $|\Delta M\text{-value}| = 0.01$ , limma method, FDR threshold of 0.05, and 100 simulations (Supplementary Fig. 2).

Based on power calculation, we attempted to perform the EWAS on 200 participants matched by sex, age, and baseline eGFR between progression and non-progression groups. However, 20 out of 200 participants failed the final epigenomic sample QC process. Therefore, 180 participants (progression: 93, non-progression: 87) were included in the EWAS.

We also performed the pyrosequencing analysis as validation. A total of 78 individuals from DN in KNOW-CKD, excluding those in EWAS, matched by age, sex, and baseline eGFR who passed the QC of pyrosequencing (Supplementary Fig. 1; Supplementary Table 1) were selected. In addition, 55 individuals (41 progression, 14 non-progression) diagnosed with DN from biopsy in the Seoul National University Hospital (SNUH) Human Biobank were included for pyrosequencing analysis (Supplementary Fig. 1). Finally, pyrosequencing analysis was performed based on a total of 133 participants (Supplementary Fig. 1).

**Outcome measurement.** The eGFR was calculated using the four-variable Chronic Kidney Disease Epidemiology Collaboration equation<sup>15</sup>. The eGFR slope in the KNOW-CKD cohort was calculated based on linear mixed models with random intercepts using MIXED procedures in SAS software (SAS Institute, Inc., Cary, North Carolina)<sup>16</sup>. The eGFR slope was estimated using creatinine values, measured at every time point; the initiation of the cohort, 6 months after the initiation of the cohort, and at least one follow-up since 2011, every 1–7 years. Only participants that had had eGFR measured at least three times were included. The LMM was fitted, where follow-up time was the dependent variable, and eGFR was the independent variable.

The fixed effect was the effect of “time” on eGFR. The fixed effect represents the average change in eGFR over time in all participants. The random effects are the participant-specific intercepts and slopes of the association between “time” and eGFR. The random intercept represents the variation among the participants in their baseline level of eGFR. The random slope term for “time” captures the variation among the participants in their rate of change of eGFR over time. Therefore, LMM was fitted, allowing each participant to have their own baseline level of eGFR and rate of change in eGFR over time<sup>13,17</sup>.

$$eGFR_{ij} = \beta_0 + \beta_1 time_{ij} + u0_i + u1_i time_{ij} + \varepsilon_{ij}$$

$eGFR_{ij}$ : eGFR slope of the  $i$ -th subject at the  $j$ -th observation time point,  $time_{ij}$ : follow-up time of the  $i$ -th subject at the  $j$ -th observation time point,  $\beta_0$ : fixed effect intercept,  $\beta_1$ : fixed effect intercept (eGFR slope),  $u0_i$ : random effect intercept of the  $i$ -th subject,  $u1_i$ : random effect slope of the  $i$ -th subject,  $\varepsilon_{ij}$ : error term (residual).

Hypertension was defined as a systolic blood pressure of  $\geq 140$  mmHg, diastolic blood pressure of  $\geq 90$  mmHg, or past medical history. Diabetes mellitus was defined as serum hemoglobin A1C  $\geq 6.5\%$ , fasting blood glucose  $\geq 126$  mg/dl, or a past medical history. CKD progression was defined as an eGFR slope  $< -2.6$  mL/min/1.73 m<sup>2</sup>/year.

**Epigenome-wide DNA methylation profiling.** Genomic DNA was extracted from leukocytes in the peripheral blood of all samples. Comparison of methylation profiles among the primary outcomes (CKD progression vs. non-progression) was performed using the Illumina Infinium MethylationEPIC platform. The microarray-based DNA methylation levels for individuals were profiled using the Illumina Infinium MethylationEPIC BeadChip kits, which features  $> 850,000$  cytosine-phosphate-guanine (CpG) sites in enhancer regions, gene bodies, promoters, and CpG islands. The DNA methylation array was imaged using a standard Illumina procedure with an Illumina iScan scanner (Illumina, Inc., San Diego, CA, USA).

**Quality control and EWAS.** We performed quality control of DNA methylation data extracted from raw intensity data (IDAT), including the signal intensities for each of the probes on the chip with over 1 million probes. To minimize the unintended variation within and between samples, we implemented quantile normalization, which considers the methylated and unmethylated signal intensities separately. We excluded probes with a detection P-value of  $> 0.05$ , which can be considered a low-quality signal from all samples. The detection P-value was calculated using the “m + u” method, which compares the total DNA signal (methylated + unmethylated) at each site to the background signal level which is estimated using negative control sites, assuming a normal distribution<sup>18</sup>. CpG sites that failed in one or more samples are filtered based on the detection P-value. We removed the probes on the X or Y chromosome, in addition to the probes affected by single nucleotide polymorphisms (SNPs) without the specification of a certain minor allele frequency<sup>19</sup>. In addition, non-specific binding probes that mapped to multiple locations on the genome were filtered<sup>20</sup>. The annotation was performed by an Illumina Infinium MethylationEPIC BeadChip (EPIC chip), which is a microarray platform designed to DNA methylation across over 860,000 CpG sites in human genome. Finally, 784,864 out of 1,051,815 probes passed the quality control and were included in the EWAS (Supplementary Fig. 3).

CpG sites associated with CKD progression were identified using linear regression models implemented in the limma package in R with an empirical Bayesian framework<sup>21</sup>. The methylation levels at each CpG probe are represented as M-values. The beta-value is a commonly used measure of DNA methylation that ranges from 0 to 1, representing the proportion of methylation at a given CpG site<sup>22</sup>. Conversely, the M-value, or logit-transformed beta-value, is the log<sub>2</sub> ratio of the intensities of methylated versus unmethylated probes, calculated as<sup>22</sup>:

$$M_i = \log_2 \left( \frac{\max(y_{i,methyl}, 0) + \alpha}{\max(y_{i,methyl}, 0) + \alpha} \right)$$

The M-value has the advantage of being symmetrical around zero, and it is often used in statistical analyses, as it allows for more accurate measurement of differential methylation between groups. The M-value ranges from  $-\infty$  to  $\infty$ , with values close to zero indicating low methylation and increasingly negative or positive values indicating higher levels of hypomethylation or hypermethylation, respectively<sup>22</sup>.

Since the differences in the various cell types of the whole blood between progression and non-progression can lead to false differentiated methylation regions, the effects of cell proportion on the results of EWAS should be considered<sup>23</sup>. Therefore, in addition to the original model without adjustment for the blood cell proportions (referred to as Model 1), we have also adjusted for blood cell proportions, including T lymphocytes, B cells, monocytes, NK cells, and neutrophils, to remove the false CpG sites from the differences of cell proportions based on the Houseman method (referred to as Model 2). Furthermore, we adjusted for body mass index (BMI) and smoking status (yes/no) as covariates (referred to as Model 3). All summary statistics are provided in Supplementary Table 2.

All results in this study are methylation differences in the primary outcome, diabetic CKD progression versus non-progression. Following the implementation of an epigenome-wide significance threshold of  $< 0.05$  using the false discovery rate (FDR)<sup>24</sup>, the number of CpG sites were reduced from a total of 784,864 to 9,809, 8,900, and 8,690 in Model 1, Model 2, and Model 3, respectively (Supplementary Fig. 3; Supplementary Fig. 4).

Subsequently, CpG sites without an annotation for gene symbols were removed (7252, 6537, and 6328 CpG sites in the Model 1, Model 2, and Model 3, respectively). We only selected CpG sites located in the promoter regions (promoters were defined as regions located between 0 and 1500 bp upstream of transcriptional start sites (TSS), 5'UTR, and the 1st exon). There were 3837, 3462, and 3261 CpG sites in the Model 1, Model 2, and Model 3, respectively. In addition, CpG sites located in the shelf and shore regions of the CpG island (CGI) were selected (843, 774, and 742 CpG sites in the Model 1, Model 2, and Model 3, respectively). Furthermore, CpG sites with a more restricted FDR threshold (FDR  $< 0.005$ ) were selected to perform pyrosequencing and in-silico functional analysis (197, 157, and 157 CpG sites in the Model 1, Model 2, and Model 3, respectively).

We used the top five percentile  $|\Delta M\text{-value}|$  as the threshold in the distribution of  $|\Delta M\text{-value}|$  to exclude false positive CpG sites. Since  $|\Delta M\text{-value}|$  has a left skewed distribution, we determined the top five percentile values based on a non-parametric bootstrapping resampling method (alpha = 0.05, the number of resampling = 1000)

(Supplementary Fig. 5). The top five percentile values of the quantile were estimated (quantile [95% CI] = 0.2954 [0.2915, 0.3002]) based on the resampling distribution. In addition, the top five percentile values from the EWAS with the adjustment for blood cell proportions were estimated as quantile [95% CI] = 0.3007 [0.2958, 0.3045] (Supplementary Fig. 5). Furthermore, the top five percentile values from the EWAS with the adjustment for not only blood cell proportions but also BMI and smoking status were estimated as quantile [95% CI] = 0.0382 [0.0371, 0.0393] (Supplementary Fig. 5). Therefore, we selected highly significant CpG sites with an  $|\Delta M\text{-value}| \geq 0.30$  in Model 1 and 2 (17 and 15 CpG sites in Model 1 and Model 2), with an  $|\Delta M\text{-value}| \geq 0.038$  in Model 3 (12 CpG sites in Model 3), respectively.

Quality control and genome-wide analyses were conducted using the Minfi package from the Bioconductor platform in R<sup>25,26</sup>.

**Pyrosequencing: replication analysis.** The 17 candidate CpG sites used in pyrosequencing analysis were selected from the results of EWAS with no adjustment for blood cell proportions (Supplementary Fig. 3; Supplementary Fig. 6). The pyrosequencing primer was designed using the PyroMark Assay Design SW 2.0 software (QIAGEN) under the following three conditions: (1) maximum amplicon length < 200 bp, (2) primer set score  $\geq 75$ , and (3) primers attached to CpG sites were excluded (Supplementary Fig. 6). Ultimately, only 11 out of 17 CpG sites (cg11513352, cg10297223, cg22773662, cg03503634, cg04089320, cg20746451, cg14279121, cg15280188, cg11508872, cg21285133, and cg02990553) within genes *DOC2A*, *AGTR1* (Angiotensin II receptor type 1), *MIEF1*, *TRAF6*, *EMB*, *SMARCD1*, *OSBPL9*, *ASPSCR1*, *RAB14*, *ANP32E*, and *KRT28* (Keratin 28), respectively) were available to undergo pyrosequencing analysis under these three conditions. The primer sequences used in this study are listed in Supplementary Table 3.

For each assay, bisulfite-converted DNA was amplified using PCR, using the instructions provided by the manufacture of by the PyroMark PCR kit (QIAGEN). The PCR product was bound to magnetic streptavidin beads. Quality control of the pyrosequencing data was performed using the PyroMark Q48 software. All samples passed the quality control process. Sequencing was performed on a PyroMark Q48 Autoprep system using the PyroMark Q48 Advanced CpG Reagents (QIAGEN) according to the manufacturer's instructions.

The percentage of DNA methylation at specific CpG sites was estimated using the PyroMark Q48 Autoprep 2.4.2 software (QIAGEN) and exported to the R statistical environment. Subsequently, linear regressions were performed for each CpG site covered by the assay, as well as for the average methylation value across the region.

We performed a linear regression analysis with the average methylation level of methylated cytosines as the dependent variable and progression/non-progression as the independent variable to select CpG sites that show differential methylation levels between these two groups (progression/non-progression)<sup>27</sup>. The beta estimation in the regression was used to calculate the difference in methylation levels between the two groups (Supplementary Table 3).

**Phenome-wide association study.** We performed the PheWAS for cg10297223 and cg02990553 CpG sites based on the variables (phenotypes) in KNOW-CKD cohort. Based on a total of 1,028 variables in KNOW-CKD cohort, we excluded 719 variables with a missing rate over 10%. Of the remaining 309 variables, including in the PheWAS, 144 were continuous and 165 were categorical variables. The association between each  $|M\text{-value}|$  of the CpG sites and the phenotype was estimated using linear or logistic regression models according to the continuous or categorical phenotypes, respectively. The statistical significance threshold for PheWAS was also set at  $FDR < 0.05$  using the Benjamini–Hochberg method<sup>24</sup>.

**In silico functional analysis.** We further performed functional annotation analysis, such as the analysis of disease-gene network (DGN), reactome (RA) pathways, and protein–protein interaction (PPI) network, to identify the biological mechanisms of CpG sites. DNG has been used to identify cross-phenotypes associated with selected genes from CpG sites using DisGeNET<sup>28</sup>. We also used the RA database to annotate gene sets for biological pathways<sup>29</sup>. The PPI network was constructed using the Search Tool for the Retrieval of Interacting Genes (STRING; <http://string.embl.de/>) with a confidence score  $\geq 0.99$  to identify the functional interactions between proteins<sup>30</sup>. Statistical significance was determined by a false discovery rate (FDR)-corrected P-value of  $< 0.05$ . Furthermore, we identified the Expression Quantitative Trait Methylation (eQTM) based on a human whole-blood epigenome-wide association study from the Human Kidney eQTM by Susztak Lab (available on [https://susztaklab.com/Kidney\\_meQTL/index.php](https://susztaklab.com/Kidney_meQTL/index.php))<sup>31</sup>. Network illustrations from the functional analyses were constructed using the Cytoscape software (version 3.9.1) via Rcy3<sup>32,33</sup>.

**DNA methyltransferase (DNMT) inhibitor treatment.** HEK 293 cells were cultured in Dulbecco's modified Eagle's medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher Scientific), 100 U/ml penicillin (Thermo Fisher Scientific), and 100  $\mu\text{g}/\text{ml}$  streptomycin (Thermo Fisher Scientific) in an atmosphere containing 95% humidified air and 5%  $\text{CO}_2$  at 37 °C. To demethylate methylated CpG sites, HEK 293 cells were treated with increasing concentrations (0, 5, 10, and 20  $\mu\text{M}$ ) of 5-aza-2'-deoxycytidine (Sigma-Aldrich, St. Louis, MO, USA) for 72 h, which was replaced daily. Inhibition of methylation was examined by pyrosequencing analysis, and changes in *AGTR1* expression were measured by reverse-transcription quantitative polymerase chain reaction (RT-qPCR).

**RNA preparation and reverse-transcription quantitative polymerase chain reaction (RT-qPCR).** Total RNA was extracted from HEK 293 cells using the RNeasy Mini Kit (Qiagen, Valencia, CA, USA), according to the manufacturer's protocol. One microgram of total RNA was converted to cDNA using Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and oligo-(dT)12–18 primers (both from

Thermo Fisher Scientific) according to the manufacturer's instructions. qRT-PCR was performed in a 20  $\mu$ l reaction mixture containing 1  $\mu$ l cDNA, 10  $\mu$ l SYBR Premix EX Taq (Takara Bio, Otsu, Japan), 0.4  $\mu$ l Rox reference dye (50 $\times$ , Takara Bio), and 200 nM primers for each gene. The following primer sequences were used in this study:

*AGTR1* (forward), 5'-GCCCTTTGGCAATTACCTATGT-3';  
*AGTR1* (reverse), 5'-CGTGAGTAGAAACACACTAGCGT-3';  
*GAPDH* (forward), 5'-AATCCCATCACCATCTTCCA-3';  
*GAPDH* (reverse), 5'-TGGACTCCACGACGT ACTCA-3'.

The reactions were run on a 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) at 95  $^{\circ}$ C for 30 s, followed by 40 cycles at 95  $^{\circ}$ C for 3 s and 60  $^{\circ}$ C for 30 s, and a single cycle at 95  $^{\circ}$ C for 15 s, 60  $^{\circ}$ C for 60 s, and 95  $^{\circ}$ C for 15 s to generate dissociation curves. All PCR reactions were performed in triplicate, and the specificity of the reaction was determined by melting curve analysis. Comparative quantification of each target gene was performed based on the cycle threshold (Ct) normalized to *GAPDH*, using the  $\Delta\Delta$ Ct method<sup>34</sup>.

## Results

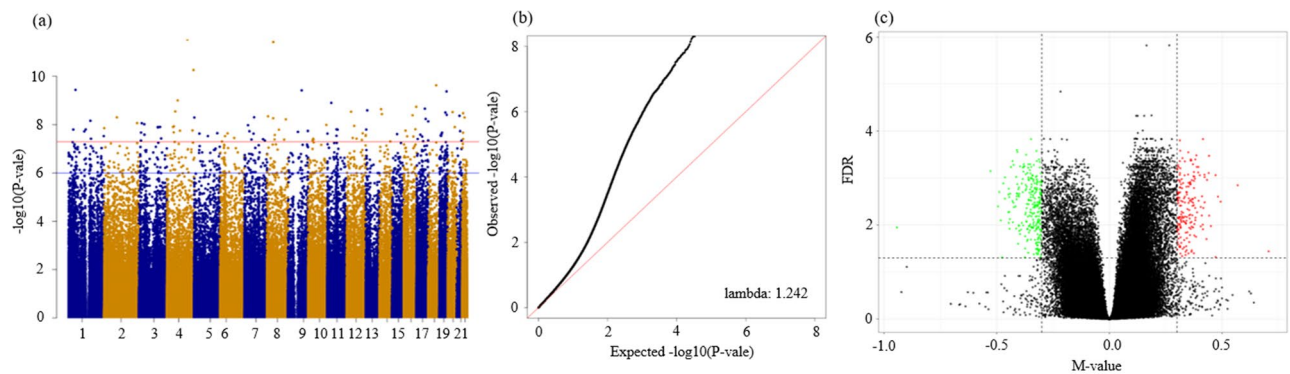
**General characteristics of study population.** Table 1 shows the clinical and demographic characteristics of 180 diabetic participants with CKD (93 progression versus 87 non-progression) based on the KNOW-CKD cohort study. The mean age was 59.1 years, men accounted for 65% of the participants, and most participants had hypertension (98.3%). Urine albumin, urine protein, UACR, UPCR, 24-h urine protein, and 24-h urine phosphorus levels were higher in the progression group than in the non-progression group (Table 1).

Furthermore, Supplementary Table 1 shows the general characteristics of 78 DN from the KNOW-CKD cohort and 55 from the biopsy in SNUH Human Biobank for pyrosequencing analysis. Diastolic blood pressure (BP), Urine albumin, urine protein, UACR, UPCR, and 24-h urine protein levels were higher in the progression group than in the non-progression group among the 78 participants. However, only a limited number of variables were investigated among the 55 participants.

**Differentially methylated CpG sites.** In the results with no adjustment for blood cell proportion, 17 CpG sites remained based on the FDR < 0.005 and  $|\Delta$  M-value  $\geq$  0.3 threshold (Fig. 1; Supplementary Fig. 3). According to the results of the adjustment for blood cell proportion with the same threshold, 15 CpG sites were identified (Supplementary Fig. 3). Based on the results of the adjustment for blood cell proportion, BMI, and smoking status, 12 CpG sites remaining based on the FDR < 0.005 and  $|\Delta$  M-value  $\geq$  0.038 threshold (Supplementary Fig. 3). Of the 15 CpG sites, only 14 CpG sites (cg20746451, cg01490296, cg10297223, cg02990553, cg06205244, cg21285133, cg04089320, cg22773662, cg03503634, cg15280188, cg21285782, cg14279121,

	Total (N = 180)	Progression (N = 93)	Non-progression (N = 87)	P-value
	Mean (SD)	Mean (SD)	Mean (SD)	
Age at baseline	59.1 (8.0)	58.6 (8.4)	59.6 (7.5)	0.43
Systolic BP (mmHg)	131.4 (15.0)	131.9 (15.3)	130.9 (14.7)	0.68
Diastolic BP (mmHg)	74.5 (9.7)	74.1 (9.5)	74.9 (9.9)	0.58
Body mass index (kg/m <sup>2</sup> )	25.0 (3.0)	25.1 (2.7)	25.0 (3.4)	0.73
White blood cells (/mm <sup>3</sup> )	7124.0 (2067.6)	7036.3 (2121.8)	7217.7 (2016.0)	0.56
Hemoglobin (g/dL)	11.8 (1.7)	11.7 (1.7)	11.9 (1.6)	0.37
Urine albumin (mg/dL)	1169.3 (1444.0)	1803.0 (1688.4)	491.9 (626.1)	< 0.01
Urine protein (mg/dL)	167.7 (207.4)	257.0 (245.1)	72.3 (86.1)	< 0.01
UACR	1.3 (1.6)	2.0 (1.8)	0.5 (0.8)	< 0.01
UPCR	1.9 (2.4)	2.9 (2.7)	0.8 (1.1)	< 0.01
24-h urine protein (g)	1.9 (2.2)	3.0 (2.6)	0.8 (0.8)	< 0.01
24-h urine phosphorus (g)	0.9 (1.5)	1.2 (2.0)	0.5 (0.2)	< 0.01
eGFR (ml/min/1.73 m <sup>2</sup> )	37.4 (13.2)	37.4 (13.0)	37.4 (13.5)	0.98
eGFR slope (ml/min/1.73 m <sup>2</sup> /year)	- 3.0 (2.2)	- 4.7 (1.5)	- 1.2 (1.1)	< 0.01
	<b>Median (IQR)</b>	<b>Median (IQR)</b>	<b>Median (IQR)</b>	<b>P-value</b>
Follow up (years)	3.9 (2.3)	3.3 (1.9)	4.3 (3.0)	< 0.001
	<b>N (%)</b>	<b>N (%)</b>	<b>N (%)</b>	<b>P-value</b>
Sex (male)	117 (65.0)	62 (66.7)	55 (63.2)	0.62
Hypertension	177 (98.3)	93 (100.0)	84 (96.6)	0.11

**Table 1.** General characteristics of diabetic chronic kidney disease based on the KoreaN cohort study for Outcome in patients With Chronic Kidney Disease (KNOW-CKD) cohort study. BP, blood pressure; eGFR, estimated glomerular filtration rate; IQR, interquartile Rang.



**Figure 1.** Visualization of methylated probes. (a) Manhattan and (b) quantile–quantile plots of the epigenome-wide association study for CKD progression. (c) Volcano plot showing differentially methylated CpG sites. The x-axis presents the M-value of the difference in signal intensity between the primary outcome for each probe. The y-axis represents the  $-\log_{10}$  (P-value). Significant CpG sites ( $FDR < 0.05$  and  $|\Delta M\text{-value}| > 0.30$ ) are highlighted in red and green. CpG sites highlighted as red and blue are those that were hypermethylated and hypomethylated compared to the non-progression, respectively. FDR, false discovery rate.

cg11508872, cg26296769) were identical and of the 12 CpG sites, only seven CpG sites (cg20746451, cg01490296, cg10297223, cg02990553, cg06205244, cg21285133, cg04089320) were identical to the original 17 CpG sites, respectively (Table 2; Supplementary Fig. 3).

The results of pyrosequencing provided information about the proportion of methylated cytosines in each DNA sample, as well as information about the average level of methylation at individual CpG sites<sup>35</sup>. Of the 11 candidate CpG sites available for pyrosequencing primer design, six genes (*DOC2A*, *MIEF1*, *EMB*, *SMARCA1*, *ASPC1*, and *ANP32E*) could not be considered validated due to the inconsistent direction of effect size with the discovery results (Supplementary Table 3). Out of the remaining five genes, three genes, *TRAF6*, *OSBPL9*, and *RAB14*, were difficult to validate due to the small effect sizes despite the consistent direction of effect size with the discovery results of EWAS. cg10297223 on *AGTR1* (EWAS:  $\Delta M\text{-value} = 0.365$ ,  $FDR = 3.18E-03$ , pyrosequencing: Beta (SE) = 0.788 (0.397), P-value =  $4.90E-02$ ) was considered potentially validated, whereas cg02990553 on *KRT28* (EWAS:  $\Delta M\text{-value} = 0.350$ ,  $FDR = 2.84E-04$ , pyrosequencing: Beta (SE) = 0.459 (0.912), P-value =  $6.10E-01$ ) only demonstrated a consistent direction of effect size without sufficient evidence for statistical validation by pyrosequencing analysis, respectively (Table 2; Fig. 2). In addition, both cg10297223 and cg02990553 were associated with seven phenotypes (24-h urine protein, 24-h urine phosphorus, urine albumin, urine protein, UPCR, UACR, and eGFR slope) based on the PheWAS (Table 3).

**In silico functional analysis.** Based on the functional analysis of the DGN, *AGTR1* and *STOX1* were associated with elevated systolic ( $FDR = 3.22E-02$ ) and diastolic BP ( $FDR = 3.22E-02$ ) (Fig. 3). In addition, *AGTR1* and *NT5C2* were associated with pre-hypertension ( $FDR = 3.53E-02$ ), and *AGTR1* and *KCNC4* were associated with adverse events associated with cardiac arrhythmia ( $FDR = 3.59E-02$ ). *AGTR1* was also associated with *TRAF6* based on the PPI network. *KRT28* was involved in the biological pathways of developmental biology, keratinization, and the formation of the cornified envelope ( $FDR = 3.39E-39$ ) based on the RA pathways (Fig. 3; Supplementary Table 4). In addition, four CpG sites (g15280188, cg14279121, cg04089320, and cg11513352) among a total of 17 top CpG sites were identified as eQTM (Supplementary Table 5).

**AGTR1 expression was regulated by epigenetic DNA methylation.** To determine whether the expression of *AGTR1* mRNA was epigenetically modulated, we treated HEK 293 cells with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine. The expression of *AGTR1* mRNA was quantified by RT-qPCR and the methylation status of the CpG site (cg10297223) within *AGTR1* was determined by pyrosequencing analysis. After treatment with 5-aza-2'-deoxycytidine, the expression of *AGTR1* mRNA was significantly restored ( $\sim 1.67$ -fold) in a dose-dependent manner, which occurred concurrently with the decreased methylation status of the *AGTR1* promoter CpG site (Fig. 4). These results indicate that *AGTR1* expression is regulated by a DNA methylation-dependent mechanism.

## Discussion

In our study, EWAS was performed to select CpG sites that were differentially methylated during the progression of diabetic CKD in the Korean population. External replication analysis was performed using pyrosequencing, focusing on the top-ranked candidate CpG sites. Consequently, cg10297223 on *AGTR1* and cg02990553 on *KRT28* were found to be significant CpG markers, and gene-level functional analysis was performed to confirm that the two CpG sites share biological mechanisms with diabetic CKD progression based on existing knowledge or hypotheses.

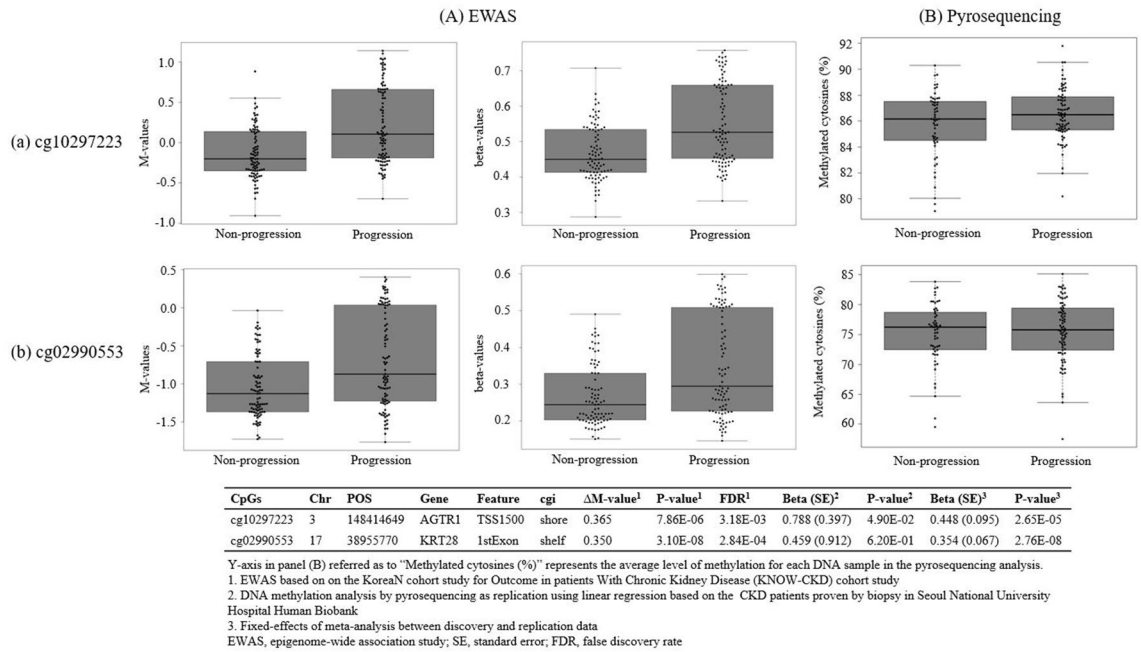
The *AGTR1* is a G-protein-coupled transmembrane receptor located at the end of the renin–angiotensin–aldosterone system (RAAS) cascade<sup>36</sup>. The RAAS cascade is a major regulator of systemic arterial blood

CpGs	Chr	Position	Gene	Feature	cgi	$\Delta M$ -value	P-value	FDR
No adjustment for blood cell proportions (17 CpG sites)								
cg02990553 <sup>a</sup>	17	38955770	<i>KRT28</i>	1stExon	Shelf	0.350	3.10E-08	2.84E-04
cg21285133 <sup>a</sup>	1	150209464	<i>ANP32E</i>	TSS1500	Shore	0.305	1.68E-07	5.30E-04
cg11508872 <sup>a</sup>	9	123964527	<i>RAB14</i>	TSS200	Shore	0.313	4.93E-07	7.81E-04
cg01490296	10	13205619	<i>MCM10</i>	5'UTR	Shore	0.366	1.04E-06	1.10E-03
cg15280188 <sup>a</sup>	17	79934854	<i>ASPSCR1</i>	TSS1500	Shore	-0.343	3.03E-06	1.91E-03
cg14279121 <sup>a</sup>	1	52194898	<i>OSBPL9</i>	TSS1500	Shore	-0.325	3.22E-06	1.96E-03
cg04089320 <sup>a</sup>	5	49737641	<i>EMB</i>	TSS1500	Shore	-0.384	3.37E-06	2.00E-03
cg03503634 <sup>a</sup>	11	36531973	<i>TRAF6</i>	TSS200	Shore	-0.463	6.03E-06	2.74E-03
cg26296769	6	43025862	<i>KLC4</i>	TSS1500	Shore	0.304	7.43E-06	3.08E-03
cg10297223 <sup>a</sup>	3	148414649	<i>AGTR1</i>	TSS1500	Shore	0.365	7.86E-06	3.18E-03
cg20746451 <sup>a</sup>	4	95127703	<i>SMARCAD1</i>	TSS1500	Shore	-0.426	8.61E-06	3.35E-03
cg22773662 <sup>a</sup>	22	39899355	<i>MIEF1</i>	5'UTR	Shore	0.352	8.75E-06	3.37E-03
cg06205244	6	86299452	<i>SNX14</i>	5'UTR	Shelf	0.306	8.84E-06	3.40E-03
cg23715505	10	115613921	<i>NHLRC2</i>	TSS1500	Shore	-0.324	1.07E-05	3.80E-03
cg24577191	19	18344621	<i>PDE4C</i>	5'UTR	Shore	0.318	1.10E-05	3.86E-03
cg11513352 <sup>a</sup>	16	30022537	<i>DOC2A</i>	5'UTR	Shore	0.406	1.23E-05	4.12E-03
cg21285782	1	154530958	<i>UBE2Q1</i>	1stExon	Shore	-0.330	1.28E-05	4.21E-03
Adjustment for blood cell proportions (15 CpG sites)								
cg14279121	1	52194898	<i>OSBPL9</i>	TSS1500	Shore	-0.324	4.79E-06	2.82E-03
cg21285133	1	150209464	<i>ANP32E</i>	TSS1500	Shore	0.309	1.97E-07	7.21E-04
cg21285782	1	154530958	<i>UBE2Q1</i>	1stExon	Shore	-0.338	1.16E-05	4.52E-03
cg10297223	3	148414649	<i>AGTR1</i>	TSS1500	Shore	0.368	9.04E-06	3.90E-03
cg20746451	4	95127703	<i>SMARCAD1</i>	TSS1500	Shore	-0.444	5.27E-06	2.94E-03
cg04089320	5	49737641	<i>EMB</i>	TSS1500	Shore	-0.386	3.72E-06	2.53E-03
cg26296769	6	43025862	<i>KLC4</i>	TSS1500	Shore	0.302	1.18E-05	4.58E-03
cg06205244	6	86299452	<i>SNX14</i>	5'UTR	Shelf	0.309	9.81E-06	4.08E-03
cg11508872	9	123964527	<i>RAB14</i>	TSS200	Shore	0.313	6.49E-07	1.04E-03
cg01490296	10	13205619	<i>MCM10</i>	5'UTR	Shore	0.371	1.15E-06	1.37E-03
cg03503634	11	36531973	<i>TRAF6</i>	TSS200	Shore	-0.475	4.69E-06	2.80E-03
cg08054244	14	88789549	<i>KCNK10</i>	1stExon	Shore	0.444	5.19E-06	2.92E-03
cg02990553	17	38955770	<i>KRT28</i>	1stExon	Shelf	0.354	3.49E-08	3.91E-04
cg15280188	17	79934854	<i>ASPSCR1</i>	TSS1500	Shore	-0.345	3.48E-06	2.46E-03
cg22773662	22	39899355	<i>MIEF1</i>	5'UTR	Shore	0.361	7.03E-06	3.41E-03
Adjustment for blood cell proportions, BMI, and smoking status (12 CpG sites)								
cg21285133	1	150209464	<i>ANP32E</i>	TSS1500	Shore	0.045	1.87E-07	6.42E-04
cg24962873	3	11312541	<i>ATG7</i>	TSS1500	Shore	0.039	5.99E-07	9.81E-04
cg10297223	3	148414649	<i>AGTR1</i>	TSS1500	Shore	0.060	7.28E-06	3.56E-03
cg07321536	4	39459897	<i>LIAS</i>	TSS1500	Shore	-0.040	3.30E-06	2.38E-03
cg20746451	4	95127703	<i>SMARCAD1</i>	TSS1500	Shore	-0.065	4.10E-06	2.66E-03
cg06205244	6	86299452	<i>SNX14</i>	5'UTR	Shelf	0.052	1.06E-05	4.40E-03
cg19933320	7	64125401	<i>ZNF107</i>	TSS1500	Shore	0.039	2.69E-06	2.12E-03
cg10551778	8	359909	<i>FBXO25</i>	5'UTR	Shelf	0.039	1.78E-07	6.32E-04
cg01490296	10	13205619	<i>MCM10</i>	5'UTR	Shore	0.062	1.60E-06	1.61E-03
cg24299813	11	60685836	<i>TMEM109</i>	5'UTR	Shelf	0.045	6.00E-07	9.81E-04
cg02990553	17	38955770	<i>KRT28</i>	1stExon	Shelf	0.060	4.26E-08	3.97E-04
cg22921692	19	9930722	<i>FBXL12</i>	TSS1500	Shore	0.041	2.23E-08	2.92E-04

**Table 2.** Candidate CpG sites for primer design for the external validation based on pyrosequencing analysis. Chr, chromosome; cgi, CpG island; CpG, Cytosine-phosphate-Guanine; BMI, body mass index; FDR, false positive rate. <sup>a</sup>Eleven out of 17 CpG sites were available for primer design.

pressure, fluid, and electrolyte balance<sup>37</sup>, primarily functions in the second stage of the embryo, and plays essential roles in neonates, maintenance of peripheral vascular resistance, and renal blood flow<sup>38</sup>.

In addition, the interaction between *AGTR1* and angiotensin II, which is released from mesangial cells, has been demonstrated to activate the inflammatory cascade by regulating protein kinase C and the mitogen-activated



**Figure 2.** DNA methylation analysis by epigenome-wide association study as discovery and pyrosequencing as validation for the progression of chronic kidney disease (CKD) in diabetic CKD patients. Beeswarm and box plots shows the DNA methylation values of two CpG sites. (A) The M-values and beta-values of epigenome-wide association study based on EPIC BeadChip for (a), cg02990553 on *AGTR1* (b), and cg10297223 on *KRT28*. (B) The percentage of differentially methylated CpG sites using pyrosequencing were generated for (a), cg02990553 on *AGTR1* (b), and cg10297223 on *KRT28*.

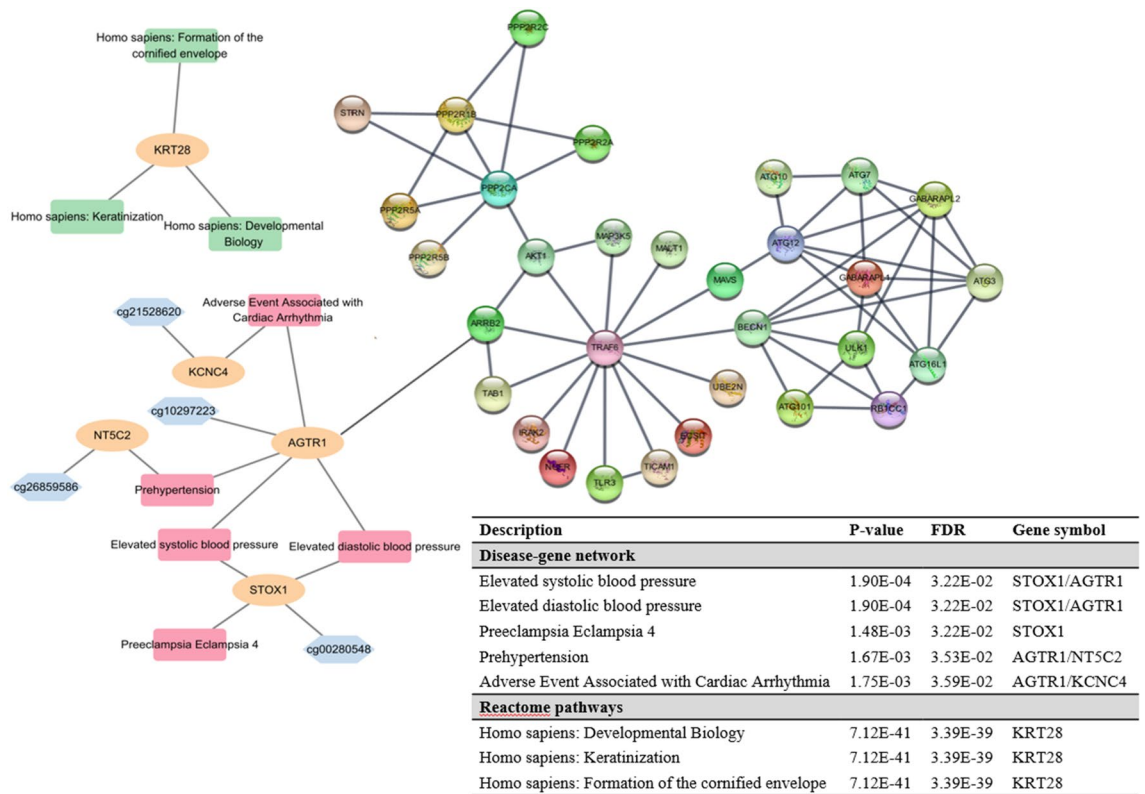
Phenotypes	Effect size	P-value	FDR
cg10297223 ( <i>AGTR1</i> )			
24-h urine protein (mg/day)	1566.788	3.726E-07	2.787E-05
24-h urine phosphorus (mg/day)	1070.956	2.299E-07	2.787E-05
Urine albumin	824.337	9.602E-06	2.394E-04
Urine protein	119.347	8.045E-06	2.149E-04
UPCR	1.408	3.809E-06	1.476E-04
UACR	951.300	2.330E-06	1.089E-04
eGFR slope	- 1.015	3.106E-04	6.489E-03
cg02990553 ( <i>KRT28</i> )			
24-h urine protein (mg/day)	2058.406	2.936E-07	2.787E-05
24-h urine phosphorus (mg/day)	1231.686	3.947E-06	1.476E-04
Urine albumin	1161.956	8.292E-07	5.169E-05
Urine protein	162.965	1.548E-06	8.268E-05
UPCR	1.968	3.355E-07	2.787E-05
UACR	1362.769	8.483E-08	2.787E-05
eGFR slope	- 1.511	2.223E-05	5.224E-04

**Table 3.** Phenome-wide association study based on M-values of cg10297223 (*AGTR1*) and cg02990553 (*KRT28*). FDR, false discovery rate; UPCR, urinary protein-to-creatinine ratio; UACR, urinary albumin-to-protein ratio; eGFR, estimated glomerular filtration rate.

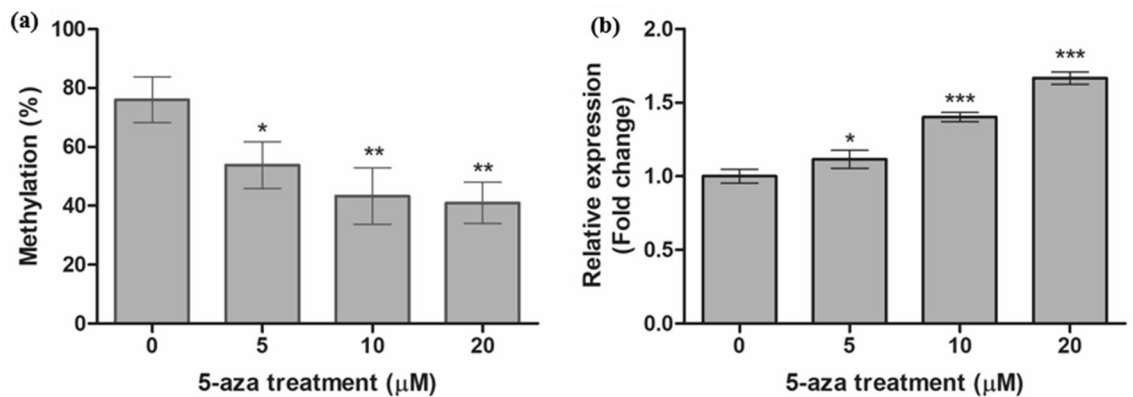
protein kinase (MAPK) pathway<sup>39</sup> and induce the expression of growth factors and proliferative cytokines to sustain the generation of nephrotoxic reactive oxygen, resulting in inflammation, fibroblast formation, and collagen deposition<sup>40</sup>. Angiotensin II also activates signaling of the NF-κB pathways, which are activated by TNF-receptor-associated factor (TRAF), leading to inflammation<sup>41-43</sup>.

Therefore, it has been suggested that pathogenic mutations leading to the absence or defects in *AGTR1* can induce fatal phenotypes<sup>44</sup>. Moreover, chronic activation of the RAAS is recognized as a critical factor in CKD





**Figure 3.** Functional analysis for cg10297223 on *AGTR1* and cg02990553 on *KRT28* which are associated with diabetic CKD. Blue, orange, red, and green nodes indicate CpG sites, gene symbols, disease-gene networks, and Reactome pathways, respectively. Nodes with molecular structure indicate PPI networks. CKD, chronic kidney disease; PPI, protein-protein interaction.



**Figure 4.** Modulation of *AGTR1* mRNA expression following demethylation in HEK 293 cells. HEK 293 cells were treated for 72 h with various concentration of 5-aza-2'-deoxycytidine. After treatment, demethylation of *AGTR1* promoter CpG site (cg10297223) was confirmed by pyrosequencing analysis (a) and the expression of *AGTR1* mRNA was measured by RT-qPCR (b). Data are presented as the mean ± SD from three independent experiments. Statistical analyses were performed using one-way ANOVA with Dunnett's multiple comparison post-test for comparing significance with untreated control (\*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001). 5-aza, 5-aza-2'-deoxycytidine.

progression<sup>45,46</sup>. In addition, a non-functioning RAAS cascade results in kidney damage under neonatal or hypoxic conditions<sup>47</sup>.

A study on *AGTR1*-related CKD in RAAS at the GWAS level was also reported in a systematic review and meta-analysis<sup>48</sup>. In the Chronic Renal Insufficiency Cohort study based on Caucasian and African American populations in 2015, the association between RAAS-related genes with CKD was reported, but *AGTR1* was not found to be significantly associated with CKD<sup>49</sup>. However, another GWA study in an African American population reported an association between *AGTR1* and diabetic ESKD<sup>50</sup>. Nevertheless, previous studies have only

reported an association between the alleles of SNPs in *AGTR1* and CKD, and studies on gene activation, such as gene expression or methylation of *AGTR1*, especially in the Korean population, have not yet been described.

*KRT28* encodes a member of the type I (acidic) keratin family, which belongs to the superfamily of intermediate filament (IF) proteins<sup>51</sup>. Previous studies have reported that a cornified envelope or keratinization, which is associated with *KRT28*, is also associated with CKD.

The components of the cornified envelope were considerably reduced in participants with CKD compared to that of the control group in a previous study<sup>52</sup>. It has been reported that treatment with emollients can reduce the thickness and density of scales and noticeably improve the quality of life of CKD<sup>53</sup>.

Moreover, acquired perforating dermatosis (APD), which is caused by chronic friction leading to epithelial proliferation, abnormal keratinization, and decreased blood supply due to microangiopathy, is often associated with underlying systemic diseases, such as diabetes mellitus and CKD<sup>54</sup>. APD most often occurs after starting dialysis in CKD<sup>55</sup>. Kidney damage is known to affect wound healing<sup>56</sup>. Research data on rats also showed an exacerbating effect of CKD on wound healing, which is mediated by the disruption of keratinization and delayed granulation<sup>56</sup>. In addition, veiled chronic inflammatory conditions, low rates of angiogenesis, and cell proliferation also contribute to poor wound healing<sup>57</sup>. Although several *KRT* series genes related to keratinization have been reported in previous studies, *KRT28* (particularly as an epigenetic marker in Korean populations) was implicated for the first time in our study<sup>58</sup>.

Previous studies based on similar hypotheses have been reported for populations other than Koreans. The previous study has reported the enhancement of renal regulatory regions and their correlation with gene expression changes, including epidermal growth factor, related to kidney damage and impaired function using methylation probes<sup>59</sup>. Another study reported similar correlation results for individuals receiving kidney transplants or dialysis, demonstrating the ability to analyze transplant recipients alongside individuals receiving dialysis to improve the performance of future EWAS for ESKD<sup>60</sup>.

Our study had several limitations that need to be acknowledged. First, although we selected candidate CpG sites for performing external replication analysis using pyrosequencing, epigenome-wide replication analysis could not be performed owing to the lack of Korean or Asian-based CKD cohorts with epigenomic databases. Nevertheless, since the KNOW-CKD cohort, which forms the basis of the current study, has almost completed the recruitment of an additional 1500 CKD participants for phase II and has started follow-up (<https://clinicaltrials.gov/ct2/show/NCT03929900>), we will be able to conduct epigenome-wide validation analysis in the future<sup>13</sup>.

Second, because the DNA samples used in our study were derived from peripheral blood samples, there is limited information on the association between whole blood DNA methylation profiles and kidney tissue-specific DNA methylation differentiation, in part due to the heterogeneity of cell types within the kidney. However, a previous study suggested that blood DNA methylation analysis is valuable because it can reflect changes in DNA methylation in the tissues associated with the phenotypes<sup>61</sup>. Nevertheless, the establishment of a biobank of kidney biopsies is needed to improve tissue-specific DNA methylation analysis for kidney disease in the future<sup>61</sup>.

Although we used the threshold of  $|\Delta M\text{-value}|$  as 0.3 and 0.0038 for EWAS with/without adjustment of blood cell proportions and with adjustment of blood cell proportions, BMI, and smoking status, respectively, in order to exclude false positive CpG probes, there is a possibility that CpG probes with a small difference were not considered as candidate CpG sites for the validation due to the high threshold of the  $|\Delta M\text{-value}|$ .

There was a possibility that cg10297223 and cg02990553 could be validated in the pyrosequencing analysis in our study. However, although cg10297223 had a significant raw P-value (P-value < 0.05), both CpG sites had no statistical significance in FDR or Bonferroni correction (Supplementary Table 3). In future, we hope to validate this finding, utilizing the KNOW-CKD phase II cohort which, as already mentioned, will include an additional 1500 CKD participants—the recruitment of these participants is almost completed<sup>13</sup>.

Since the Human Kidney eQTM results gathered by Susztak Lab were sampled from a different ethnic group than the ethnicity of the cohort used in our study<sup>31</sup>, there may be an association between Korean-specific epigenetics markers and gene expression that has not yet been identified.

Furthermore, the use of 5-aza-2'-deoxycytidine results in the demethylation of CpGs throughout the genome of cells, making it challenging to apply this treatment for the causal analysis of effects. Therefore, caution should be exercised when interpreting the results, as the observed changes may not be directly attributable to the targeted CpG sites<sup>62</sup>.

Despite these limitations, our study had several strengths. First, although epigenome-wide replication analysis could not be performed, functional annotation analysis was conducted in silico to elucidate the biological mechanisms of the CpG sites identified in our study. Moreover, based on PheWAS, CpG sites associated with diabetic CKD in our study were confirmed to be appreciably associated with different phenotypes related to CKD progression.

Second, although the DNA samples used in our study were whole blood DNA samples, a gene or that of the same family reported as GWAS-level in previous studies was also identified in our findings. Moreover, our study demonstrated that epigenetic markers affect gene expression and proteomic production more in the central dogma than at the GWAS-level. Furthermore, since our CpG markers were extracted from a clinically accessible peripheral blood sample, they can be used as diagnostic markers in the future.

We have identified two epigenetic markers (cg10297223 on *AGTR1* and cg02990553 on *KRT28*) that show a potential association with diabetic CKD progression in the Korean population. Based on functional annotations and PheWAS, both genes with CpG sites may offer insights into the activation of genetic markers in diabetic CKD, suggesting that cg10297223 and cg02990553 could be considered as potential clinical biomarkers. Nevertheless, further studies are necessary to validate the association between whole blood and kidney tissue-specific DNA methylation.

## Data availability

All the methylation array data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE230652. The datasets of KNOW-CKD cohort are available from the corresponding author upon reasonable request. Summary statistics estimated in this study are available from Supplementary Table 2.

Received: 21 November 2022; Accepted: 18 May 2023

Published online: 20 May 2023

## References

- Hong, Y. A. *et al.* Trends in epidemiologic characteristics of end-stage renal disease from 2019 Korean Renal Data System (KORDS). *Kidney Res. Clin. Pract.* **40**, 52–61. <https://doi.org/10.23876/j.krcp.20.202> (2021).
- Anderson, A. H. *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. <https://doi.org/10.1053/j.ajkd.2020.07.011> (2021).
- Lee, M. J., Ha, K. H., Kim, D. J. & Park, I. Trends in the incidence, prevalence, and mortality of end-stage kidney disease in South Korea. *Diabetes Metab. J.* **44**, 933–937. <https://doi.org/10.4093/dmj.2020.0156> (2020).
- Thomas, M. C. *et al.* Diabetic kidney disease. *Nat. Rev. Dis. Primers* **1**, 15018. <https://doi.org/10.1038/nrdp.2015.18> (2015).
- Thomas, M. C. Epigenetic mechanisms in diabetic kidney disease. *Curr. Diab. Rep.* **16**, 31. <https://doi.org/10.1007/s11892-016-0723-9> (2016).
- Smyth, L. J., McKay, G. J., Maxwell, A. P. & McKnight, A. J. DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigenetics* **9**, 366–376. <https://doi.org/10.4161/epi.27161> (2014).
- Bell, C. G. *et al.* Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Med. Genomics* **3**, 33. <https://doi.org/10.1186/1755-8794-3-33> (2010).
- Lecamwasam, A. *et al.* DNA methylation profiling identifies epigenetic differences between early versus late stages of diabetic chronic kidney disease. *Nephrol. Dial. Transplant.* **36**, 2027–2038. <https://doi.org/10.1093/ndt/gfaa226> (2021).
- Sapienza, C. *et al.* DNA methylation profiling identifies epigenetic differences between diabetic patients with ESRD and diabetes patients without nephropathy. *Epigenetics* **6**, 20–28. <https://doi.org/10.4161/epi.6.1.13362> (2011).
- Oh, K. H. *et al.* The KNOW-CKD study: What we have learned about chronic kidney diseases. *Kidney Res. Clin. Pract.* **39**, 121–135. <https://doi.org/10.23876/j.krcp.20.042> (2020).
- Oh, K.-H. *et al.* KNOW-CKD (KoreaN cohort study for outcome in patients with chronic kidney disease): Design and methods. *BMC Nephrol.* **15**, 1–9 (2014).
- Heaf, J. G. & Mortensen, L. S. Uraemia progression in chronic kidney disease stages 3–5 is not constant. *Nephron Clin. Pract.* **118**, c367–374. <https://doi.org/10.1159/000323391> (2011).
- Han, M. *et al.* Novel genetic variants associated with chronic kidney disease progression. *J. Am. Soc. Nephrol.* <https://doi.org/10.1681/asn.000000000000066> (2023).
- Graw, S., Henn, R., Thompson, J. A. & Koestler, D. C. pwrEWAS: A user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC Bioinform.* **20**, 218. <https://doi.org/10.1186/s12859-019-2804-7> (2019).
- Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612. <https://doi.org/10.7326/0003-4819-150-9-200905050-00006> (2009).
- Bing, D. & He, X. Linear mixed models in clinical trials using PROC MIXED. *PharmaSUG2010—Paper SP07* (2010).
- Grams, M. E. *et al.* Evaluating glomerular filtration rate slope as a surrogate end point for ESKD in clinical trials: An individual participant meta-analysis of observational data. *J. Am. Soc. Nephrol.* **30**, 1746–1755. <https://doi.org/10.1681/asn.2019010008> (2019).
- Heiss, J. A. & Just, A. C. Improved filtering of DNA methylation microarray data by detection p values and its impact on downstream analyses. *Clin. Epigenet.* **11**, 15. <https://doi.org/10.1186/s13148-019-0615-3> (2019).
- Chen, Y. A. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–209. <https://doi.org/10.4161/epi.23470> (2013).
- Paul, D. S. *et al.* Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. *Nat. Commun.* **7**, 13555. <https://doi.org/10.1038/ncomms13555> (2016).
- Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* <https://doi.org/10.2202/1544-6115.1027> (2004).
- Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinform.* **11**, 587. <https://doi.org/10.1186/1471-2105-11-587> (2010).
- Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **13**, 86. <https://doi.org/10.1186/1471-2105-13-86> (2012).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).
- Dirks, R. A., Stunnenberg, H. G. & Marks, H. Genome-wide epigenomic profiling for biomarker discovery. *Clin. Epigenet.* **8**, 1–17 (2016).
- Lai, C. Q. *et al.* Epigenome-wide association study of triglyceride postprandial responses to a high-fat dietary challenge. *J. Lipid Res.* **57**, 2200–2207. <https://doi.org/10.1194/jlr.M069948> (2016).
- Xie, C. *et al.* Differential methylation values in differential methylation analysis. *Bioinformatics* **35**, 1094–1097. <https://doi.org/10.1093/bioinformatics/bty778> (2019).
- Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845–d855. <https://doi.org/10.1093/nar/gkz1021> (2020).
- Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–622. <https://doi.org/10.1093/nar/gkn863> (2009).
- Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–d612. <https://doi.org/10.1093/nar/gkaa1074> (2021).
- Sheng, X. *et al.* Systematic integrated analysis of genetic and epigenetic variation in diabetic kidney disease. *Proc. Natl. Acad. Sci. USA* **117**, 29013–29024. <https://doi.org/10.1073/pnas.2005905117> (2020).
- Gustavsen, J. A., Pai, S., Isserlin, R., Demchak, B. & Pico, A. R. RCy3: Network biology using Cytoscape from within R. *F1000Res* **8**, 1774. <https://doi.org/10.12688/f1000research.20887.2> (2019).
- Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) method. *Methods* **25**, 402–408. <https://doi.org/10.1006/meth.2001.1262> (2001).
- Delaney, C., Garg, S. K. & Yung, R. Analysis of DNA methylation by pyrosequencing. *Immunosenescence Methods Protocols*, 249–264 (2015).

36. Campbell, C. Y. *et al.* Associations between genetic variants in the ACE, AGT, AGTR1 and AGTR2 genes and renal function in the Multi-ethnic Study of Atherosclerosis. *Am. J. Nephrol.* **32**, 156–162. <https://doi.org/10.1159/000315866> (2010).
37. Peach, M. J. Renin-angiotensin system: Biochemistry and mechanisms of action. *Physiol. Rev.* **57**, 313–370. <https://doi.org/10.1152/physrev.1977.57.2.313> (1977).
38. Demirgan, E. B. *et al.* AGTR1-related renal tubular dysgeneses may not be fatal. *Kidney Int Rep* **6**, 846–852. <https://doi.org/10.1016/j.ekir.2020.11.033> (2021).
39. Lai, K. N. *et al.* Mesangial expression of angiotensin II receptor in IgA nephropathy and its regulation by polymeric IgA1. *Kidney Int.* **66**, 1403–1416. <https://doi.org/10.1111/j.1523-1755.2004.00874.x> (2004).
40. Forrester, S. J. *et al.* Angiotensin II signal transduction: An update on mechanisms of physiology and pathophysiology. *Physiol. Rev.* **98**, 1627–1738. <https://doi.org/10.1152/physrev.00038.2017> (2018).
41. Chen, X. Q. *et al.* Pioglitazone inhibits angiotensin II-induced atrial fibroblasts proliferation via NF- $\kappa$ B/TGF- $\beta$ 1/TRIF/TRAF6 pathway. *Exp. Cell Res.* **330**, 43–55. <https://doi.org/10.1016/j.yexcr.2014.08.021> (2015).
42. Qiu, W. *et al.* Sublytic C5b-9 complexes induce proliferative changes of glomerular mesangial cells in rat Thy-1 nephritis through TRAF6-mediated PI3K-dependent Akt1 activation. *J. Pathol.* **226**, 619–632. <https://doi.org/10.1002/path.3011> (2012).
43. Zhu, Y., Xue, Z. & Di, L. Regulation of MiR-146a and TRAF6 in the diagnose of lupus nephritis. *Med. Sci. Monit.* **23**, 2550–2557. <https://doi.org/10.12659/msm.900667> (2017).
44. Parsa, A. *et al.* Common variants in Mendelian kidney disease genes and their association with renal function. *J. Am. Soc. Nephrol.* **24**, 2105–2117. <https://doi.org/10.1681/asn.2012100983> (2013).
45. Ruster, C. & Wolf, G. Renin-angiotensin-aldosterone system and progression of renal disease. *J. Am. Soc. Nephrol.* **17**, 2985–2991 (2006).
46. Wolf, G., Butzmann, U. & Wenzel, U. O. The renin-angiotensin system and progression of renal disease: From hemodynamics to cell biology. *Nephron Physiol.* **93**, P3–13. <https://doi.org/10.1159/000066656> (2003).
47. Uematsu, M. *et al.* A further case of renal tubular dysgenesis surviving the neonatal period. *Eur. J. Pediatr.* **168**, 207–209. <https://doi.org/10.1007/s00431-008-0743-9> (2009).
48. Smyth, L. J., Cañadas-Garre, M., Cappa, R. C., Maxwell, A. P. & McKnight, A. J. Genetic associations between genes in the renin-angiotensin-aldosterone system and renal disease: A systematic review and meta-analysis. *BMJ Open* **9**, e026777. <https://doi.org/10.1136/bmjopen-2018-026777> (2019).
49. Kelly, T. N. *et al.* The role of renin-angiotensin-aldosterone system genes in the progression of chronic kidney disease: Findings from the Chronic Renal Insufficiency Cohort (CRIC) study. *Nephrol. Dial Transplant.* **30**, 1711–1718. <https://doi.org/10.1093/ndt/gfv125> (2015).
50. Palmer, N. D. *et al.* Evaluation of candidate nephropathy susceptibility genes in a genome-wide association study of African American diabetic kidney disease. *PLoS ONE* **9**, e88273. <https://doi.org/10.1371/journal.pone.0088273> (2014).
51. Ho, M. *et al.* Update of the keratin gene family: Evolution, tissue-specific expression patterns, and relevance to clinical disorders. *Hum. Genomics* **16**, 1. <https://doi.org/10.1186/s40246-021-00374-9> (2022).
52. Agarwal, P. *et al.* Chronic kidney disease-associated pruritus. *Toxins* <https://doi.org/10.3390/toxins13080527> (2021).
53. Balaskas, E. *et al.* Randomized, double-blind study with glycerol and paraffin in uremic xerosis. *Clin. J. Am. Soc. Nephrol.* **6**, 748–752. <https://doi.org/10.2215/cjn.05490610> (2011).
54. Morton, C. A., Henderson, I. S., Jones, M. C. & Lowe, J. G. Acquired perforating dermatosis in a British dialysis population. *Br. J. Dermatol.* **135**, 671–677 (1996).
55. Goel, V., Sil, A. & Das, A. Cutaneous manifestations of chronic kidney disease, dialysis and post-renal transplant: A review. *Indian J. Dermatol.* **66**, 3–11. [https://doi.org/10.4103/ijd.IJD\\_502\\_20](https://doi.org/10.4103/ijd.IJD_502_20) (2021).
56. Seth, A. K., De la Garza, M., Fang, R. C., Hong, S. J. & Galiano, R. D. Excisional wound healing is delayed in a murine model of chronic kidney disease. *PLoS ONE* **8**, e59979. <https://doi.org/10.1371/journal.pone.0059979> (2013).
57. Kursh, E. D., Klein, L., Schmitt, J., Kayal, S. & Persky, L. The effect of uremia on wound tensile strength and collagen formation. *J. Surg. Res.* **23**, 37–42. [https://doi.org/10.1016/0022-4804\(77\)90188-3](https://doi.org/10.1016/0022-4804(77)90188-3) (1977).
58. Snider, N. T. Kidney keratins: Cytoskeletal stress responders with biomarker potential. *Kidney Int.* **89**, 738–740. <https://doi.org/10.1016/j.kint.2015.12.040> (2016).
59. Gluck, C. *et al.* Kidney cytosine methylation changes improve renal function decline estimation in patients with diabetic kidney disease. *Nat. Commun.* **10**, 2461. <https://doi.org/10.1038/s41467-019-10378-8> (2019).
60. Smyth, L. J. *et al.* Assessment of differentially methylated loci in individuals with end-stage kidney disease attributed to diabetic kidney disease: An exploratory study. *Clin. Epigenet.* **13**, 99. <https://doi.org/10.1186/s13148-021-01081-x> (2021).
61. Gu, H. F. Genetic and epigenetic studies in diabetic kidney disease. *Front. Genet.* **10**, 507. <https://doi.org/10.3389/fgene.2019.00507> (2019).
62. Neja, S. A. Site-specific DNA demethylation as a potential target for cancer epigenetic therapy. *Epigenet. Insights* **13**, 2516865720964808. <https://doi.org/10.1177/2516865720964808> (2020).

## Acknowledgements

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (2017M3A9E4044649).

## Author contributions

H.Y.S., S.L. and M.H. conceptualized, performed the research, and wrote the original manuscript. H.Y.S., M.H. and H.R. curated the database. W.J.A., W.J.A., H.R., E.K., Y.S.P., S.E.L. and C.A. reviewed and edited the manuscript. K-H.O., S.K.P. and J-H.A. acquired fundings, reviewed and edited the manuscript, and supervised the research. All authors finally approved the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions relating to the accuracy or integrity of any part of the work are accurately resolved.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-35485-x>.

**Correspondence** and requests for materials should be addressed to K.-H.O., S.K.P. or J.-H.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2023