Published in final edited form as:

Am J Kidney Dis. 2015 February ; 65(2): 217–222. doi:10.1053/j.ajkd.2014.09.019.

Genome-Wide Association Studies in Nephrology: Using Known Associations for Data Checks

Matthias Wuttke, MD¹, Franz Schaefer, MD², Craig S. Wong, MD, MPH³, and Anna Köttgen, MD, MPH¹

¹Division of Nephrology, University Medical Center Freiburg, Freiburg, Germany

²Division of Pediatric Nephrology, University Medical Center Heidelberg, Heidelberg, Germany

³Department of Pediatrics, University of New Mexico Children's Hospital, Albuquerque, New Mexico, US

Abstract

Prior to conducting genome-wide association studies (GWAS) of renal traits and diseases, systematic checks to ensure data integrity and analytical workflow should be conducted. Using positive controls (ie, known associations between a single-nucleotide polymorphism [SNP] and a corresponding trait) allows for identifying errors that are not apparent solely from global evaluation of summary statistics. Strong genetic control associations of chronic kidney disease (CKD), as derived from GWAS, are lacking in the non-African CKD population; thus, in this perspective we provide examples of and considerations for using positive controls among patients with CKD. Using data from individuals with CKD who participated in the CRIC (Chronic Renal Insufficiency Cohort) Study or PediGFR (Pediatric Investigation for Genetic Factors Linked to Renal Progression) Consortium, we evaluated 2 kinds of positive control traits: traits unrelated to kidney function (bilirubin, body height) and those related to kidney function (cystatin C, urate). For the former, the proportion of variance in the control trait that is explained by the control SNP is the main determinant of the strength of the observable association, irrespective of adjustment for kidney function. For the latter, adjustment for kidney function can be effective in uncovering known associations among patients with CKD. For instance, in 1,092 participants of the PediGFR Consortium, the p-value for association of cystatin C concentrations and rs911119 in the *CST3* gene decreased from 2.7×10^{-3} to 2.4×10^{-8} upon adjustment for serum creatinine-based estimated glomerular filtration rate. In this perspective, we give recommendations for the appropriate selection of control traits and SNPs that can be used for data checks prior to conducting GWAS among patients with CKD.

© 2014 by the National Kidney Foundation, Inc. All rights reserved.

Address for correspondence: Anna Köttgen, MD MPH, Dept. of Internal Medicine IV, University Medical Center Freiburg, Berliner Allee 29, 79110 Freiburg, Germany, anna.koettgen@uniklinik-freiburg.de.

Financial Disclosure: The authors declare that they have no other relevant financial interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Index words

genome-wide association study (GWAS); single-nucleotide polymorphism (SNP); genetic marker; positive control; data checking; systematic error; chronic kidney disease (CKD); renal trait

Genome-wide association studies (GWAS) investigate millions of genetic markers per person to identify genomic regions in which genetic variation associates with a trait or disease. For each single-nucleotide polymorphism (SNP), summary statistics (eg, a p-value) are calculated for its association with the phenotype. The resulting GWAS file contains millions of lines, which makes visual data and plausibility checks challenging. While there are excellent tools for systematically checking genome-wide summary statistics, including allele frequencies and computed association statistics for their global distribution^{1,2}, other systematic errors can go unnoticed. For example, incorrect association results may arise from a mismatch between the genotypes and phenotypes of the individuals (ie, inadvertent scrambling of the data). Because most SNPs are not expected to show an association with the phenotype of interest, such errors would escape global checks of summary statistics. Hence, additional checks to reliably assess the integrity of data and analytical workflow are required.

These considerations highlight the importance of using a positive control, one or several genomic markers that are known to reproducibly associate with an available trait or phenotype. A SNP that is a suitable positive control should have an effect strong enough to be detected in as little as a few hundred samples. As outlined in Box 1, a practical approach is to survey the GWAS Catalog of the National Human Genome Research Institute (NHGRI) for phenotypes and diseases available in a given study and to identify SNPs that previously have shown genome-wide significant associations ($\sim p < 1 \times 10^{-7}$ or $< 1 \times 10^{-8}$, depending on the study). The SNPs should have been replicated successfully and should have shown associations in samples of the same ancestry as the data to be analyzed. We suggest reviewing the cited publications from the GWAS Catalog to select the SNP(s) that explain the largest amount of the trait variance. If this is not reported, summary statistics usually can help guide the selection of SNPs with the largest effect estimate and lowest p-values in a given study. Lastly, it is advantageous to have control SNPs represented on a given genotyping chip so that imputation is not required.

Finding a good positive control is challenging for GWAS in the field of kidney disease. In individuals of African descent, variants in the *APOL1* gene have been shown to associate strongly with focal segmental glomerulosclerosis (FSGS), hypertension-attributed end-stage renal disease, and CKD from a variety of causes^{3,4}. Therefore, these markers might serve as positive controls. Since these variants are ancestry specific, data checks in samples that are not of African ancestry require the use of other positive controls.

Using quantitative control phenotypes such as biomarker concentrations generally is recommended due to the superior statistical power to detect associations with continuous phenotypes as compared to binary phenotypes (Box 1). Further, many of them are widely available. However, reduced kidney function influences blood concentrations of many

biomarkers by altering their production, metabolism, and/or elimination. As a result, the genetic influence on marker concentrations can become less apparent.

A feasible and straightforward solution would be to use a biomarker with extrarenal production and, at least partially, extrarenal elimination, or a phenotypic trait unaffected by decreased kidney function. As an alternative, when evaluating the positive control association, it may be feasible to adjust for glomerular filtration rate (GFR) to reduce the effect of reduced kidney function on the biomarker blood concentration. In the following paragraphs, and summarized in Table 1, we present several examples and considerations.

One suitable marker with extrarenal production and largely extrarenal elimination is bilirubin. Serum concentrations of bilirubin reflect the balance between its production and elimination. Although liver and kidney disease can coexist, CKD has not been described as having a major effect on hepatic function, and serum bilirubin values of patients with CKD are usually in the reference range. Polymorphisms in *UGT1A*, which encodes the bilirubin UDP (uridine diphosphate)-glucuronosyltransferase, were identified initially as associated with bilirubin concentrations in some cohorts forming the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) Consortium⁵. In this study, the SNP rs6742078 explained 18% of the variance in bilirubin concentrations, with a p-value of $<5 \times 10^{-324}$ and a 0.23-unit higher log(total serum bilirubin concentration in $\mu\text{mol/l}$) per T allele among 9,464 individuals studied. Effect sizes of this magnitude should be easily detectable at genome-wide significance even in samples of smaller size.

Row A of Table 1 shows the association result for rs6742078 with log(serum bilirubin concentration) in 1,527 participants of European American ancestry of the CRIC (Chronic Renal Insufficiency Cohort) Study⁶. Allele frequency, effect direction, and effect size were consistent with previously published results, with an association p-value of 6×10^{-32} . Additional adjustment for serum creatinine-based estimated GFR (eGFR_{cr}) did not change the effect size of the SNP on bilirubin concentrations and left the p-value essentially unchanged (4×10^{-32} ; Table 1, row B). Panels A and B of Figure 1 show a graphical representation of this example, highlighting the feasibility of using biomarkers with extrarenal metabolism as positive controls in association studies of renal phenotypes. This strategy also increases confidence in genotyping quality, correct data handling, and the approach to generate association statistics.

Body height should not be affected by CKD of adult onset. Because this anthropometric measure is available in many studies, it represents another phenotype that may serve as a positive control. Ideally, body height should be obtained by standardized measurement rather than self-report. We evaluated the association between height and rs6440003 (the SNP, in the gene *ZBTB38*, which showed the strongest association with body height in one of the first GWAS of the trait)⁷ among 1,549 participants of European American ancestry in the CRIC Study (Table 1, row C). Although allele frequency, as well as effect direction and magnitude, were consistent with previous findings, the association was only of borderline significance ($P = 0.08$). As expected, adjustment for eGFR_{cr} did not change the evaluated association. The SNP only explained 0.3% of the variance in body height in the original publication⁷; therefore, it is likely that this association would not be detected in studies of

small size and thus is not an ideal positive control. This example highlights the importance of finding a variant where the proportion of explained phenotypic variance is large. Furthermore, study-specific characteristics need to be taken into consideration. For example, height is not suitable as a positive control in children with CKD because CKD influences growth, and height in children varies as a function of age.

Cystatin C is a protein encoded by *CST3* and ubiquitously found in nuclei. Because of its clearance by the kidney, it is commonly used as a biomarker of kidney function and data on its concentration may be available in studies of patients with kidney disease. Studies in the general population found a sizeable effect of variants in *CST3* (eg, rs911119) on serum concentrations of cystatin C⁸. Since a marker that is cleared by the kidney like cystatin C is affected substantially by reduced kidney function, associations between the marker and a corresponding SNP are expected to strengthen when adjusting for eGFR_{cr}. In the PediGFR Consortium, formed by 3 studies of CKD in children (CKiD [Chronic Kidney Disease in Children] Study⁹, ESCAPE [Effect of Strict Blood Pressure Control and ACE Inhibition on the Progression of CRF in Pediatric Patients] Trial¹⁰ and 4C [Cardiovascular Comorbidity in Children with Chronic Kidney Disease] Study¹¹), the association between serum cystatin C concentrations and rs911119 was nominally significant ($P = 0.003$; Table 1, row E). As shown in Figure 1 (panels C and D), adjusting for eGFR_{cr} substantially decreases the p-value to 2×10^{-8} (Table 1, row F), accompanied by a reduction in the standard error arising from the variation in cystatin C concentrations caused by reduced clearance from the kidneys. This example illustrates that even a phenotype affected by kidney function can be used as a positive control, as long as data are adjusted for GFR.

Uric acid is another biomarker with extrarenal production that may be available as part of a biomarker panel in studies of kidney function and disease. We evaluated rs7442295 in the urate transporter gene *SLC2A9*, a SNP that showed the strongest association with serum urate concentrations in a previous population-based study among individuals of central European ancestry¹², explaining 3.5% of the urate variance. The association between the SNP and serum urate in the 4C Study was of the same direction as the published association but not statistically significant. Adjustment for eGFR_{cr} did not lead to appreciable changes in the association p-value (Table 1, rows G and H).

This observation might be due to the more complex homeostasis of urate compared to cystatin C. As a housekeeping gene, *CST3* produces cystatin C at a constant rate. However, urate concentrations are influenced by several factors; for example, urate concentrations are affected by purine metabolism (of which it is the end product), renal transport (which is complex and strongly influenced by genetic variation), and environmental influences like dietary intake. All of these factors may be affected in the setting of CKD. Therefore, statistically adjusting for eGFR_{cr} alone may not be adequate to account for all determinants of urate variability in a way that allows for the detection of a genetic effect on urate concentrations in subjects with CKD. Conversely, cystatin C is freely filtered at the glomerulus and largely undergoes catabolism after tubular reabsorption; the remainder of cystatin C is eliminated in the urine. Because blood concentrations of this biomarker are mainly affected by its excretion, the association of cystatin C level and *CST3* SNP become more apparent upon adjusting for eGFR_{cr}.

If sufficient biomaterial and financial resources are available, phenotyping a trait that can be used to assess the presence of known associations can add to the value of a study. Cystatin C is an obvious choice because many CKD studies already have measured (or are interested in measuring) this marker, as it is required for using the latest CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) GFR estimating equation¹³. Purely from a quality control perspective, measuring bilirubin may be preferable because of the existence of a polymorphism that explains an exceptionally large amount of the variation in serum bilirubin concentrations. An added advantage bilirubin is that it is an economical biomarker to measure.

In conclusion, the conduct of systematic data checks prior to conducting GWAS in CKD populations remains challenging because of the lack of previously validated positive controls in populations of non-African ancestry. This brief perspective highlights alternative approaches for assuring data integrity by using extrarenal and renal biomarkers and traits as positive controls.

Acknowledgments

Support: The CKiD Study was conducted by the CKiD Consortium Investigators and is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), and the National, Heart, Lung and Blood Institute (NHLBI) (grants U01-DK-66143, U01-DK-66174, U01DK-082194, U01-DK-66116).

The ESCAPE trial and the 4C Study were conducted by the ESCAPE Clinical Research Network. Support for the 4C Study was received from the KfH Foundation for Preventive Medicine, the European Renal Association–European Dialysis and Transplant Association Research Programme, and the German Federal Ministry of Education and Research (reference number: 01EO0802).

The genotyping data used for the PediGFR consortium studies (CKiD, 4C, and ESCAPE) were supported by NIDDK grant RO1-DK-082394. Additional support was provided by the Seventh Framework Programme of the European Union (EURENomics, grant 2012-305608).

The CRIC Study was conducted by the CRIC Study Investigators and supported by the NIDDK. The data from the CRIC study reported in this article were supplied by the NIDDK Central Repositories. This article was not prepared in collaboration with Investigators of the CRIC or CKiD studies and does not necessarily reflect the opinions or views of these studies, the NIDDK Central Repositories, the NIDDK, the NICHD, or the NHLBI.

Drs Wuttke and Köttgen were supported by the Emmy Noether Programme of the German Research Foundation (KO 3598/2-1); Dr Wuttke was supported additionally by the Else Kröner-Fresenius-Stiftung (2013_Kolleg.03) of Bad Homburg, Germany.

References

1. van der Most PJ, Vaez A, Prins BP, et al. QCGWAS: A flexible R package for automated quality control of genome-wide association results. *Bioinformatics*. Jan 13.2014
2. Fuchsberger C, Taliun D, Pramstaller PP, Pattaro C. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics*. Feb 1; 2012 28(3):444–445. [PubMed: 22155946]
3. Parsa A, Kao WH, Xie D, et al. APOL1 risk variants, race, and progression of chronic kidney disease. *The New England journal of medicine*. Dec 5; 2013 369(23):2183–2196. [PubMed: 24206458]
4. Genovese G, Friedman DJ, Ross MD, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*. Aug 13; 2010 329(5993):841–845. [PubMed: 20647424]
5. Johnson AD, Kavousi M, Smith AV, et al. Genome-wide association meta-analysis for total serum bilirubin levels. *Human molecular genetics*. Jul 15; 2009 18(14):2700–2710. [PubMed: 19414484]

6. Feldman HI, Appel LJ, Chertow GM, et al. The Chronic Renal Insufficiency Cohort (CRIC) Study: Design and Methods. *Journal of the American Society of Nephrology : JASN*. Jul; 2003 14(7 Suppl 2):S148–153. [PubMed: 12819321]
7. Weedon MN, Lango H, Lindgren CM, et al. Genome-wide association analysis identifies 20 loci that influence adult height. *Nature genetics*. May. 2008; 40(5):575–583.
8. Kottgen A, Glazer NL, Dehghan A, et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nature genetics*. Jun; 2009 41(6):712–717. [PubMed: 19430482]
9. Furth SL, Cole SR, Moxey-Mims M, et al. Design and methods of the Chronic Kidney Disease in Children (CKiD) prospective cohort study. *Clinical journal of the American Society of Nephrology : CJASN*. Sep; 2006 1(5):1006–1015. [PubMed: 17699320]
10. Wuhl E, Trivelli A, Picca S, et al. Strict blood-pressure control and progression of renal failure in children. *The New England journal of medicine*. Oct 22; 2009 361(17):1639–1650. [PubMed: 19846849]
11. Querfeld U, Anarat A, Bayazit AK, et al. The Cardiovascular Comorbidity in Children with Chronic Kidney Disease (4C) study: objectives, design, and methodology. *Clinical journal of the American Society of Nephrology : CJASN*. Sep; 2010 5(9):1642–1648. [PubMed: 20576824]
12. Doring A, Gieger C, Mehta D, et al. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nature genetics*. Apr; 2008 40(4):430–436. [PubMed: 18327256]
13. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *The New England journal of medicine*. Jul 5; 2012 367(1):20–29. [PubMed: 22762315]
14. Levey AS, Coresh J, Greene T, et al. Expressing the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate with standardized serum creatinine values. *Clinical chemistry*. Apr; 2007 53(4):766–772. [PubMed: 17332152]
15. Schwartz GJ, Schneider MF, Maier PS, et al. Improved equations estimating GFR in children with chronic kidney disease using an immunonephelometric determination of cystatin C. *Kidney international*. Aug; 2012 82(4):445–453. [PubMed: 22622496]
16. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature protocols*. Sep; 2010 5(9):1564–1573.
17. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. Nov 1; 2012 491(7422):56–65. [PubMed: 23128226]
18. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nature methods*. Feb; 2012 9(2):179–181. [PubMed: 22138821]
19. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics*. Jun.2009 5(6):e1000529. [PubMed: 19543373]
20. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. Jun 7; 2007 447(7145):661–678. [PubMed: 17554300]
21. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC bioinformatics*. 2010; 11:288. [PubMed: 20509871]
22. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. Sep 15; 2010 26(18):2336–2337. [PubMed: 20634204]
23. Kottgen A, Pattaro C, Boger CA, et al. New loci associated with kidney function and chronic kidney disease. *Nature genetics*. May; 2010 42(5):376–384. [PubMed: 20383146]
24. Wallace C, Newhouse SJ, Braund P, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *American journal of human genetics*. Jan; 2008 82(1):139–149. [PubMed: 18179892]

Box 1**Workflow to Select Phenotypes and Genetic Markers to Assess Known Associations for Quality Control Purposes**

Step 1: Select one, or preferably several, available control phenotypes. Preference should be given to phenotypes/traits that are continuous and measured (eg, biomarker concentrations) rather than to those that are self-reported. Ideally, the chosen biomarker is not generated in the kidney and does not exhibit net renal secretion or reabsorption.

Step 2: Look for previous GWAS of the corresponding phenotype/trait in the GWAS Catalog (www.genome.gov/gwastudies) and in PubMed. Ensure that the published association was found among individuals of the same ancestry as your study population, that the GWAS was sufficiently powered (large sample size) and the findings were replicated.

Step 3: Among significantly associated markers (typically $P < 5 \times 10^{-8}$), select the one that explains the largest amount of phenotype/trait variance. If this is not reported in the original publication, select for large effect size estimates and low p-values instead. If several markers can be considered, prefer those with high minor allele frequencies and ones that have been genotyped (rather than imputed) in your own study.

Step 4: In your study, to the extent possible model the association between control trait and marker in the same way as was done in the original report, including trait transformation and units. Ensure the modeled allele and strand match those in the published report of the association.

Step 5: Compare direction and effect size of your association to the published result. Also assess whether the p-value meets statistical significance in your study, but (especially in smaller studies) do not expect the p-value to be as low as the ones initially published, which often originate from very large meta-analyses.

Step 6: If the blood concentrations of the chosen biomarker might be influenced by kidney function, rerun the association analyses adjusting for eGFR.

Step 7: If the positive control does not show the expected direction of association or the magnitude differs substantially, attempt to evaluate at least a second control trait. A typical mistake that can cause the repeated absence of known associations (and is not identified in any other data checks such as quality control, exploratory data analysis, data cleaning of phenotype and genotype information, and repetition of association analyses using a different statistical program) is a mismatch of the order of individuals in the phenotype and in the genotype file. This mistake results in the random shuffling of genotypes and phenotypes, giving rise to null associations.

Abbreviations: GWAS, genome-wide association study; eGFR, estimated glomerular filtration rate.

Box 2**Methods**

Genotype and phenotype data sets were acquired from the ESCAPE, 4C, CKiD and CRIC studies^{6,9-11}. Serum creatinine-based eGFR was calculated using the Modification of Diet in Renal Disease [MDRD] Study equation¹⁴ for adults and the enzymatic serum creatinine equation by Schwartz et al¹⁵ for children. After stringent data cleaning according to standard protocols¹⁶, imputation was performed using the public 1000 Genomes data as a reference panel (phase 1, release v3, release date 2010-11-23, ALL subset)¹⁷⁻¹⁹. Association analyses were conducted with the software SNPtest v2.5²⁰. Age, sex, significantly ($p < 0.05$) associated principal components, and optionally eGFR_{cr} were used as covariates. Metaanalysis of results across the PediGFR studies (ESCAPE, 4C, CKiD) was performed using the software GWAMA²¹. Regional association plots were created using the software LocusZoom²².

Figure 1.

Effects of eGFRcr adjustment on SNP associations with blood bilirubin and cystatin C concentrations. See Box 2 for detailed methods.

X-axis: genomic position; y-axis: logarithmic association p-value and recombination rate calculated from the HapMap rel 22 data. The control SNP selected from previous publications is displayed in purple; the color-coding indicates the pair-wise correlation (r^2 from 1000 Genomes EUR panel, March 2012 release) of each SNP with the control SNP.

Panel A and B: Associations at the UGT1A locus with bilirubin, without (A) and with (B) conditioning on GFRcr

Panel C and D: Associations at the CST3 locus with cystatin C, without (C) and with (D) conditioning on GFRcr

Table 1
Association of known SNPs for different phenotypes with and without adjustments for eGFR_{cr}.

Row	SNP	Gene	Phenotype, Unit	Adj eGFR _{cr}	Effect /other allele	Effect Size	SE	p	Study	N	Published Effect Size	Reference
A	rs6742078	UGT1A	Bilirubin (total), log units	N	T/G	0.29	0.02	6×10^{-32}	CRIC ^a	1527		
B	rs6742078	UGT1A	Bilirubin (total), log units	Y	T/G	0.29	0.02	4×10^{-32}	CRIC ^a	1527	0.23	5
C	rs6440003	ZBTB38	Body Height, cm	N	A/G	0.42	0.24	0.08	CRIC ^a	1549		
D	rs6440003	ZBTB38	Body Height, cm	Y	A/G	0.41	0.24	0.09	CRIC ^a	1549	0.48*	7
E	rs911119	CST3	Cystatin C, mg/dl	N	T/C	0.13	0.04	3×10^{-03}	PediGFR	1092		
F	rs911119	CST3	Cystatin C, mg/dl	Y	T/C	0.14	0.03	2×10^{-08}	PediGFR	1092	positive**	23
G	rs7442295	SLC2A9	Uric acid, mg/dl	N	G/A	-0.11	0.07	0.11	4C	535		
H	rs7442295	SLC2A9	Uric acid, mg/dl	Y	G/A	-0.12	0.07	0.08	4C	535	-0.34***	24

See Box 2 for detailed methods. All models were adjusted for age, sex, and principal components; additional adjustment for eGFR_{cr} as indicated. SNP effect allele frequencies: 0.33 (rs6742078), 0.43 (rs6440003), 0.81 (rs911119), 0.24 (rs7442295).

^a comprises patients of European American ancestry

* 1 SD in these studies is equal to 6.82 cm; the published effect size of 0.07 SD is equal to 0.48 cm

** rs911119 T allele decreases cystatin C-based eGFR and thus increases cystatin C

*** Urate unit conversion: the published effect size of 0.02 mmol/l is equal to 0.34 mg/dl

Abbreviations: adj, adjusted; SNP, single-nucleotide polymorphism; eGFR_{cr}, serum creatinine-based estimated glomerular filtration rate; CRIC, Chronic Renal Insufficiency Cohort; PediGFR, Pediatric Investigation for Genetic Factors Linked to Renal Progression; 4C, Cardiovascular Comorbidity in Children with Chronic Kidney Disease; SE, standard error; rs, reference SNP identifier; SD, standard deviation.