

# Epigenome-wide association study of serum urate reveals insights into urate co-regulation and the *SLC2A9* locus

Elevated serum urate levels, a complex trait and major risk factor for incident gout, are correlated with cardiometabolic traits via incompletely understood mechanisms. DNA methylation in whole blood captures genetic and environmental influences and is assessed in transethnic meta-analysis of epigenome-wide association studies (EWAS) of serum urate (discovery,  $n = 12,474$ , replication,  $n = 5522$ ). The 100 replicated, epigenome-wide significant ( $p < 1.1E-7$ ) CpGs explain 11.6% of the serum urate variance. At *SLC2A9*, the serum urate locus with the largest effect in genome-wide association studies (GWAS), five CpGs are associated with *SLC2A9* gene expression. Four CpGs at *SLC2A9* have significant causal effects on serum urate levels and/or gout, and two of these partly mediate the effects of urate-associated GWAS variants. In other genes, including *SLC7A11* and *PHGDH*, 17 urate-associated CpGs are associated with conditions defining metabolic syndrome, suggesting that these CpGs may represent a blood DNA methylation signature of cardiometabolic risk factors. This study demonstrates that EWAS can provide new insights into GWAS loci and the correlation of serum urate with other complex traits.

Serum urate concentrations represent a complex genetic trait, influenced by genetic variation in many genomic regions<sup>1–3</sup>, and by environmental influences such as intake of purine-rich food, alcohol, fructose, and certain diuretic medications<sup>4,5</sup>. Elevated serum urate levels, hyperuricemia, can lead to monosodium urate crystal deposition and thereby cause gout<sup>6</sup>, the most common type of inflammatory arthritis among adults. Moreover, observational studies show that serum urate levels are associated with various cardiometabolic risk factors including hypertension and obesity<sup>7–9</sup>. Gaining insights into the regulation of serum urate levels is important not only to uncover targets to develop urate-lowering and anti-inflammatory therapies to treat gout, but may also provide insights into shared pathways with cardiometabolic risk factors.

Previous studies reported the estimated heritability of serum urate levels in the general population as 30–70%<sup>3,10,11</sup>. Large-scale genome-wide association studies (GWAS) of serum urate levels have identified urate-associated genetic variants at >200 loci<sup>1–3,12</sup>. Genes in the identified loci suggest two major themes: the first is related to urate transport in the kidney and intestine, which determines net urate excretion, and the second is related to genetic co-regulation. The genes in loci most strongly associated with serum urate levels encode for urate transporters (e.g., *SLC2A9*, *SLC22A12*, *ABCG2*, *SLC22A11*, *SLC17A1*) or of substrates that may be exchanged for urate (e.g., *SLC16A9*), as well as for regulatory proteins of urate transporters (e.g., *PDZK1*)<sup>13,14</sup>. While the underlying causal variant rs2231142 at *ABCG2*, a missense variant causing reduced function, has been identified<sup>15</sup>, the causal mechanisms underlying the strongest genetic association signal with serum urate, which maps to the urate transporter gene *SLC2A9*, remain largely unknown<sup>16</sup>.

Regarding the potential genetic co-regulation between serum urate levels and multiple cardiometabolic risk factors such as triglyceride levels, blood pressure, obesity, and insulin resistance, we observed high genetic correlation between these risk factors and serum urate based on genome-wide genetic association statistics<sup>3</sup>. Such genetic co-regulation could explain the observed association of serum urate and cardiometabolic risk factors from epidemiological studies. It may, at least in part, be mediated by transcription factors with major regulatory roles in both liver and kidney such as *HNF1A* and *HNF4A*<sup>3</sup>. Genetic variants in these transcription factors are associated not only with serum urate levels, but also with impaired glucose handling and type 2 diabetes mellitus as well as serum cholesterol and triglyceride levels<sup>17–19</sup>, suggesting that coordinated gene regulation may have joint effects on serum urate and hepatic metabolism.

DNA methylation is an important mechanism of gene regulation and may reveal new insights into the biological processes that influence serum urate levels. We thus performed epigenome-wide association studies (EWAS) of serum urate levels with two scientific aims: first, to detect CpGs associated with serum urate levels and to investigate whether differential methylation may connect genetic risk variants of unknown molecular mechanism with serum urate levels. Second, to evaluate whether urate-associated differentially methylated CpGs are associated with cardiometabolic risk factors, pointing toward shared regulation of the implicated genes.

Here we perform meta-analyses of EWAS of serum urate levels followed by replication, using data from 17,996 individuals from four ancestry groups. Downstream characterization of urate-associated CpGs includes association with differential gene expression, Mendelian randomization (MR), and mediation as well as enrichment analyses. We show that significantly associated CpGs explain 11.6% of the serum urate variance in a dataset not included in the EWAS. Differential methylation of CpGs in *SLC2A9* has causal effects on serum urate levels and mediates the effect of known urate-associated genetic variants. Urate-associated

CpGs show associations with several cardiometabolic traits, consistent with their observational relationships with serum urate. Our study generates both independent and complementary insights to those obtained from GWAS of serum urate levels.

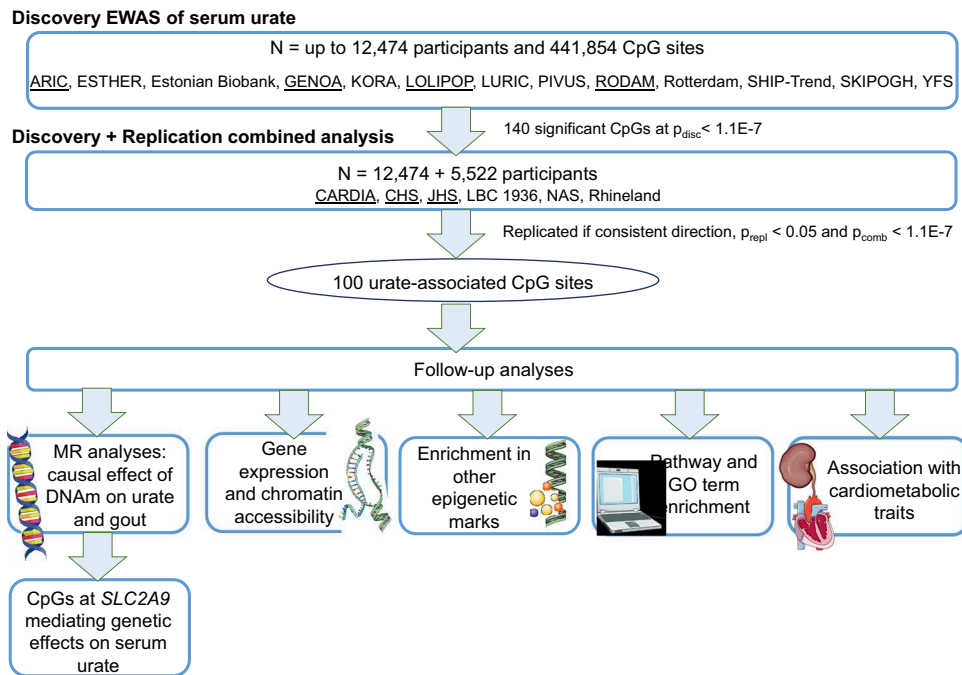
## Results

**Characteristics of the study participants.** The discovery analysis included up to 12,474 participants from 16 cohorts (European ancestry [EA]: 6968, African Americans [AA]: 2101, South Asian ancestry [SA]: 2720, and Sub-Saharan Africans [SSA]: 685). The replication analysis included 5522 participants from 8 cohorts with one study (the Normative Aging Study) including only men (EA: 3338, AA: 2184). Across these cohorts, the median of the average age within each cohort was 56 (25th, 75th percentile: 51, 63), the median of the proportion of men was 47% (25th, 75th percentile: 39%, 49%), and the median of mean serum urate levels was 5.3 mg/dL (25th, 75th percentile: 5.1, 5.6; Supplementary Data 1 and Supplementary Note 1). Study-specific methods and information on blood cell type proportions used to estimate associations between urate and DNA methylation independent of cell type composition are reported in the Methods section and Supplementary Data 2 and 3. A flowchart of the meta-analyses and follow-up characterization is presented in Fig. 1.

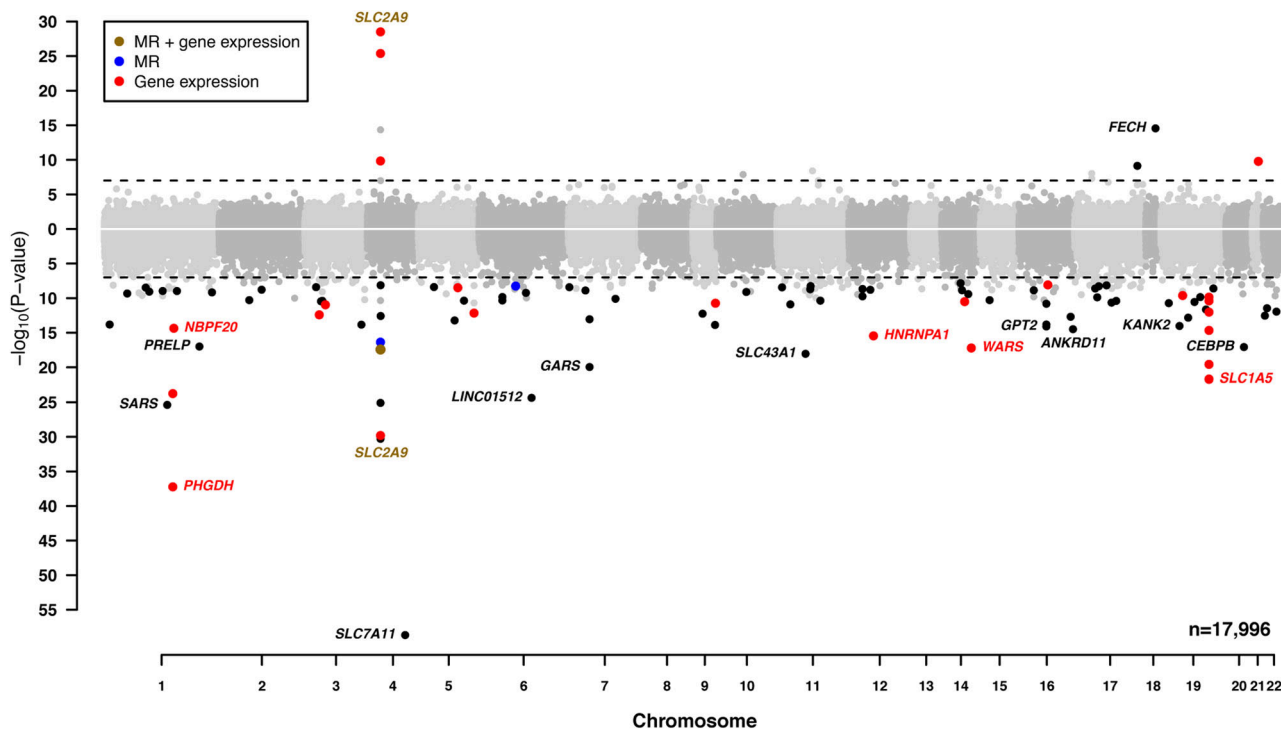
**Discovery and replication of urate-associated CpGs.** The discovery analysis identified 140 significant CpGs ( $p < 1.1E-7 = 0.05/441,854$  CpGs analyzed), of which 100 replicated (consistent effect direction,  $p < 0.05$  in the replication analysis, and overall meta-analysis  $p < 1.1E-7$ ; Supplementary Data 4 and Fig. 2). The estimated inflation factor of the meta-analysis combining discovery and replication was 1.06 (Supplementary Fig. 1). The study-specific results had a median inflation factor of 0.99 (min: 0.86, max: 1.06, Supplementary Data 2), and were corrected prior to any meta-analysis when  $> 1$ . The heterogeneity of most replicated CpGs was low to moderate: the median  $I^2$  from the meta-analysis of the 17 EA studies was 11% (25th, 75th percentile: 0.0, 31.3), from the five AA studies was 0% (25th, 75th percentile: 0.0, 35.1), and from the overall meta-analysis of the four ancestry groups was 30.2% (25th, 75th percentile: 0.0, 60.0, Supplementary Data 4). Of the 32 CpGs that showed heterogeneity  $> 50%$  in the meta-analysis of the four ancestry groups, 90% (29 CpGs) had effect estimates in the same direction across the ancestry groups (Supplementary Fig. 2A–AF). Ancestry-specific results of CpGs with  $p < 1.1E-7$  are reported in Supplementary Data 5 [EA], 6 [AA], and 7 [SA] if present, and with  $p < 1E-5$  otherwise (Supplementary Data 8 [SSA]).

The follow-up analyses focused on the replicated CpGs from the overall meta-analysis, in which the CpG with the lowest  $p$  value was at *SLC7A11* (cg06690548,  $p = 2.18E-59$ ). Among the ten CpGs with the lowest  $p$  values, two mapped to *PHGDH* (cg14476101,  $p = 5.87E-38$ ; cg16246545,  $p = 1.77E-24$ ), which encodes for 3-phosphoglycerate dehydrogenase, and five to *SLC2A9* (cg21795255, cg20479063, cg11266682, cg00071950, cg13841979;  $p < 1E-25$ ), which encodes the urate transporter GLUT9 and is the locus with the largest effect size in GWAS of serum urate<sup>3</sup>. The probe cg21795255 at *SLC2A9* had a common SNP at the extension base of the DNA methylation probe (rs3796839, minor allele frequency of 46% in GnomAD)<sup>20</sup> and was excluded from all follow-up analyses, which therefore included 99 replicated CpGs annotated to 81 genes. Of these, 24 genes contained replicated CpGs with significant findings from our follow-up analyses and are featured in Table 1.

**Heritability of urate-associated CpGs and explained serum urate variance.** DNA methylation can be influenced by genetics and environmental factors<sup>21</sup>. Heritability of DNA methylation levels at a



**Fig. 1 Flowchart of analyses.** MR Mendelian randomization, GO Gene Ontology. Icons were downloaded from smart.servier.com under the Creative Commons Attribution 3.0 Unported License.



**Fig. 2 CpGs in the EWAS of serum urate from the combined meta-analysis of discovery and replication results ( $n = 17,996$ ).** The CpGs are ordered by their chromosomal position on the x-axis with their  $-\log_{10}(p$  value) of the association on the y-axis. CpGs with positive and negative effect estimates are plotted in the upper and lower panels, respectively. The dotted horizontal lines represent the level of significance corrected for multiple testing (two-sided  $p < 1.1E-7$ ). Black: replicated CpGs; brown: gene with replicated CpGs associated with gene expression in monocytes and significant causal effects for serum urate; blue: replicated CpGs with significant causal effects on serum urate or gout; red: replicated CpGs associated with gene expression in monocytes or whole blood. Gene names are displayed if the gene had at least one CpG with  $p$  value  $< 1E-14$ . MR Mendelian randomization.

**Table 1 Genes having replicated CpG(s) associated with gene expression in whole blood or monocytes, with significant causal effects for urate or gout, associated with cardiometabolic traits, or annotated as a GWAS locus for urate or gout.**

Chr	Gene	Function of encoded protein	No. of replicated CpG(s)	Direction of effect of the replicated CpGs				Causal effect	GWAS locus (urate, gout)
				Serum urate	Gene expression in whole blood	Gene expression in monocyte	Cardiometabolic traits		
1	<i>LRRC8D</i>	Anion channel component	1	↓		↓			
1	<i>NBPF20/ PDZK1</i>	Neuroblastoma breakpoint family member/ scaffolding protein incl. urate transporters	1	↓			↓	Urate, gout	
1	<i>PHGDH</i>	Enzyme in L-serine biosynthesis	2	↓	↓	↓	↓		
1	<i>SARS</i>	Enzyme catalyzing L-serine transfer to tRNA	1	↓			↓		
3	<i>CLEC3B</i>	Encodes tetranectin	1	↓		↓			
3	<i>PARP3</i>	Encodes poly(ADP-ribose)transferase 3	1	↓	↓	↓			
4	<i>SLC2A9<sup>a</sup></i>	Urate transporter	7	↑↓		↑↓		Urate, gout	
4	<i>SLC7A11</i>	Component of cysteine/glutamate exchanger	1	↓			↓		
5	<i>PURA</i>	DNA-binding protein	1	↓		↑			
6	<i>LINC01512</i>	RNA gene		↓			↓		
7	<i>GARS</i>	Enzyme catalyzing glycine transfer to tRNA	2	↓			↓		
7	<i>SH2B2</i>	Adapter protein for tyrosine kinase receptor family members	1	↓			↓		
9	<i>UAP1L1</i>	Encodes UDP-N-Acetylglucosamine Pyrophosphorylase 1 Like 1	1	↓	↑	↑			
11	<i>CPT1A</i>	Enzyme in mitochondrial uptake of long-chain fatty acids	1	↓			↓		
11	<i>HRASLS2</i>	(PLAAT2) enzyme with phospholipase A1/2 and acyltransferase activities	1	↓				Urate	
11	<i>SLC43A1</i>	Large neutral amino acid transporter	1	↓			↓		
11	<i>WEE1</i>	Ser/Thr tyrosine kinase	1	↓			↓		
14	<i>GALC</i>	Enzyme in hydrolyzation of glycolipids	1	↓		↓			
14	<i>WARS</i>	Enzyme catalyzing tryptophan transfer to tRNA	1	↓	↓				
16	<i>NOD2</i>	Protein involved in immune response	1	↓		↑	↓		
19	<i>BCL3</i>	Transcriptional activator	1	↓			↓		
19	<i>NFIX</i>	Transcription factor	1	↓		↑			
19	<i>SLC1A5</i>	Neutral amino acid transporter	6	↓	↓	↓	↓		
21	<i>ABCG1</i>	Member of the ABC transporter family, phospholipid efflux	1	↑	↓	↓	↓		

The arrows indicate the effect direction of DNA methylation at each replicated CpG.

Gene function was obtained from Uniprot, EntrezGene, and/or PubMed.

Chr chromosome.

<sup>a</sup>At *SLC2A9*, the up and down arrows indicate that some CpGs in this gene were associated with serum urate and gene expression in the positive direction and some in the negative direction.

CpG estimates the proportion of variance in DNA methylation levels at the CpG that can be attributed to additive genetics; it does not refer to germline inheritance of DNA methylation. Across three datasets from populations of EA, the heritability estimates of the replicated CpGs varied (Supplementary Data 9). The mean heritability estimates for the 99 replicated, urate-associated CpGs was higher than the mean heritability across all CpGs assessed in each of the three studies: 0.36 (urate-associated CpGs) vs. 0.16 (HM450K array) based on Hannon et al.<sup>22</sup>, 0.54 vs. 0.19 for van Dongen et al.<sup>23</sup>, and 0.52 vs. 0.19 for McRae et al.<sup>24</sup>. Among the seven replicated sites at *SLC2A9*, the mean heritability combining the three datasets ranged from 0.39 at cg03725404 to 0.93 at cg11266682. These observations suggest that the DNA methylation at some of the urate-associated CpGs may reflect the effect of common genetic variants on serum urate levels.

The proportion of serum urate variance explained by the replicated, urate-associated CpGs was estimated in a separate study sample using the coefficient of determination ( $R^2$ ) obtained from linear regression (Methods). Compared to an age- and sex-adjusted model, the replicated CpGs explained an additional 11.6% of the serum urate variance. In a recent GWAS of serum urate based on data from 457,690 individuals, replicated genetic index variants explained 7.7% of the age- and sex-adjusted serum urate variance<sup>3</sup>. The higher proportion of urate variance explained by the replicated CpGs as compared to common genetic variants may partly reflect environmental influences on serum urate levels.

**Causal effects of CpGs on serum urate and gout.** To evaluate whether the urate-associated CpGs might causally affect serum urate levels, we used two-sample MR analysis, where genetic variants associated with DNA methylation of a CpG (methylation quantitative trait locus, meQTL) in *cis* (<500 kb) were used as proxies or genetic instruments of the CpGs. *Cis* meQTLs were available for 27 of the 99 replicated CpGs from a meta-analysis of cohorts of EA in the Genetics of DNA Methylation Consortium (GoDMC, Methods). Combined with summary statistics of a GWAS of serum urate among EA individuals, we found evidence for significant causal effects of DNA methylation on serum urate levels at four CpGs at *SLC2A9* ( $p < 1.9E-3 = 0.05/27$ ) based on the multiplicative random effect inverse-variance weighted method, our primary method (Methods, Table 2). In sensitivity analysis, all significant causal effects were supported by three or more other methods that are robust to pleiotropy (Supplementary Data 10). After additionally removing the meQTLs that were correlated with any of the five GWAS index SNPs at *SLC2A9* among EA persons (Methods), cg13841979 was the only CpG with  $\geq 4$  meQTLs for MR analysis and showed nominally significant causal effects on serum urate levels ( $p = 2.25E-2$ , Supplementary Data 10). This attenuation of causal effect is consistent with the observation that DNA methylation at cg13841979 mediated the effects of two urate GWAS index SNPs on serum urate, as reported below.

As expected, the causal effects of the four CpGs on serum urate were consistent with their effect direction from EWAS. At cg11266682, a promoter-associated CpGs, higher DNA methylation levels were associated with higher serum urate levels (effect size: 0.21 mg/dL per standard deviation [SD] of DNA methylation beta value,  $p = 8.8E-04$ ). In contrast, for three CpGs further downstream of the transcription, higher DNA methylation levels were associated with lower serum urate levels (effect size:  $-0.65$  to  $-0.46$  mg/dL per SD of DNA methylation beta value,  $p < 2.1E-4$ ). At all four CpGs, little evidence of pleiotropy, a threat to MR assumptions, was detected based on the Egger intercept test (all  $p > 0.14$ ). The observed heterogeneity of the meQTL effects on serum urate levels ( $p$  heterogeneity  $< 7.54E-239$ ; Table 2 and Supplementary Figs. 3A to 3D) suggests complex genetic

**Table 2 CpGs with significant causal effects on serum urate or gout from Mendelian randomization analysis.**

Probe ID	Chr	Position (b37)	Nearest gene	No. of meQTLs used for MR analysis	Odds ratio	Effect size <sup>c</sup>	SE	p value	p value for pleiotropy	Q statistic for heterogeneity	p value for heterogeneity	
<i>Causal effects of CpGs on serum urate</i>												
cg02387843	4	9892887	<i>SLC2A9</i>	6	-	-0.52	0.14	2.14E-04	2.40E-01	1109.2	7.54E-239	
cg13841979 <sup>a</sup>	4	9990048	<i>SLC2A9</i>	10	-	-0.46	0.12	1.38E-04	1.49E-01	1836.5	<1E-300	
cg03725404 <sup>b</sup>	4	9998017	<i>SLC2A9</i>	8	-	-0.65	0.14	4.36E-06	1.95E-01	1461.0	1.53E-312	
cg11266682 <sup>a,b</sup>	4	10021025	<i>SLC2A9</i>	11	-	0.21	0.06	8.80E-04	3.33E-01	3173.1	<1E-300	
<i>Causal effects of CpG on gout</i>												
cg03725404 <sup>b</sup>	4	9998017	<i>SLC2A9</i>	8	0.43	-0.85	0.14	2.98E-09	1.01E-01	105.8	1.52E-20	

Odds ratio and log(odds ratio) of CpGs on gout were estimated per SD in rank-based transformed DNA methylation beta levels. P values from MR analysis were two-sided and obtained from inverse-variance multiplicative random effect method. P value for heterogeneity was based on Cochran's Q test.

The *cis* methylation QTLs (meQTLs) used as instruments for the urate-associated CpGs were provided by GoDMC ( $N \leq 27,750$ ). The significance level was set at  $1.9E-3 (=0.05/27)$ .

<sup>a</sup>CpGs with significant mediating effect for two urate GWAS index SNPs (Supplementary Data 12 and 13).

<sup>b</sup>CpGs associated with gene expression of *SLC2A9* in monocytes (Supplementary Data 14).

<sup>c</sup>Effect size of CpGs on serum urate estimated in mg/dL per SD in rank-based transformed DNA methylation beta levels.

influences on DNA methylation in this region. Together, these observations at *SLC2A9* are consistent with a causal effect of differential DNA methylation on serum urate levels.

The same 27 CpGs that were tested for causal effects on serum urate were also tested for their causal effects on gout. The CpG cg03725404 at *SLC2A9* showed a significant causal effect on gout ( $p < 1.9E-3 = 0.05/27$ , Table 2). Consistent with its causal effect on lower serum urate, cg03725404 conferred lower odds for gout, with an odds ratio of 0.43 per SD of DNA methylation beta value ( $p = 2.98E-9$ ). In sensitivity analysis, the causal effect of cg03725404 for gout was supported by three MR methods robust to pleiotropy (Supplementary Data 11). The scatter plots of the effects of the meQTLs on DNA methylation and gout along with the regression slope from the MR methods showed consistency among MR methods (Supplementary Fig. 4). The meQTLs included in the MR analysis for serum urate and gout were independent of each other (Supplementary Figs. 5A–D and 6, respectively). Leave-one-out analysis showed that the significant causal effects of DNA methylation on serum urate and gout were not driven by any single meQTL at any of the CpGs (Supplementary Fig. 7A–D for serum urate and Supplementary Fig. 8 for gout). Forest plots of the effects of the meQTLs showed that the majority of the meQTLs supported the significant causal effects, despite the presence of heterogeneity (Supplementary Fig. 9A–D for serum urate and Supplementary Fig. 10 for gout). Together, these findings show robust results for the performed MR analyses and support a causal effect of differential DNA methylation at *SLC2A9* on serum urate levels and gout risk. The reverse MR analysis that assessed the potential causal effects of serum urate on DNA methylation levels did not detect any significant findings among the 99 tested CpGs (Supplementary Note 2).

**CpGs at *SLC2A9* mediate genetic effects.** Given that four CpGs at *SLC2A9* had significant causal effects on serum urate, we hypothesized that these causal effects may mediate genetic effects at this locus. We previously identified four independent intronic or intergenic independent index SNPs in two neighboring 1 Mb regions at *SLC2A9* in a large-scale GWAS meta-analysis of serum urate of EA populations<sup>3</sup>. Therefore, we first tested whether the four CpGs mediated the effects of any of four independent SNPs on serum urate among EA participants of the ARIC study ( $n = 637$ ). Two CpGs (cg11266682 and cg13841979) had significant mediating effects ( $p < 0.0125$ ) for two of the independent index SNPs (rs10017305 and rs6825187; Supplementary Data 12). We next assessed these findings in two additional studies, KORA (Cooperative Health Research in the Region of Augsburg,  $n = 1636$ ) and SHIP (Study of Health in Pomerania,  $n = 223$ ), and observed significant evidence for DNA methylation as a mediator of the effects of these SNPs on serum urate levels. From the meta-analysis of all three studies, the mediating effects ranged from 21% by cg13841979 for rs10017305 (mediating effect: 0.05 mg/dL,  $p = 4.7E-07$ ) to 43% by cg11266682 for rs6825187 (mediating effect: 0.10 mg/dL,  $p = 3.0E-5$ , Supplementary Data 13).

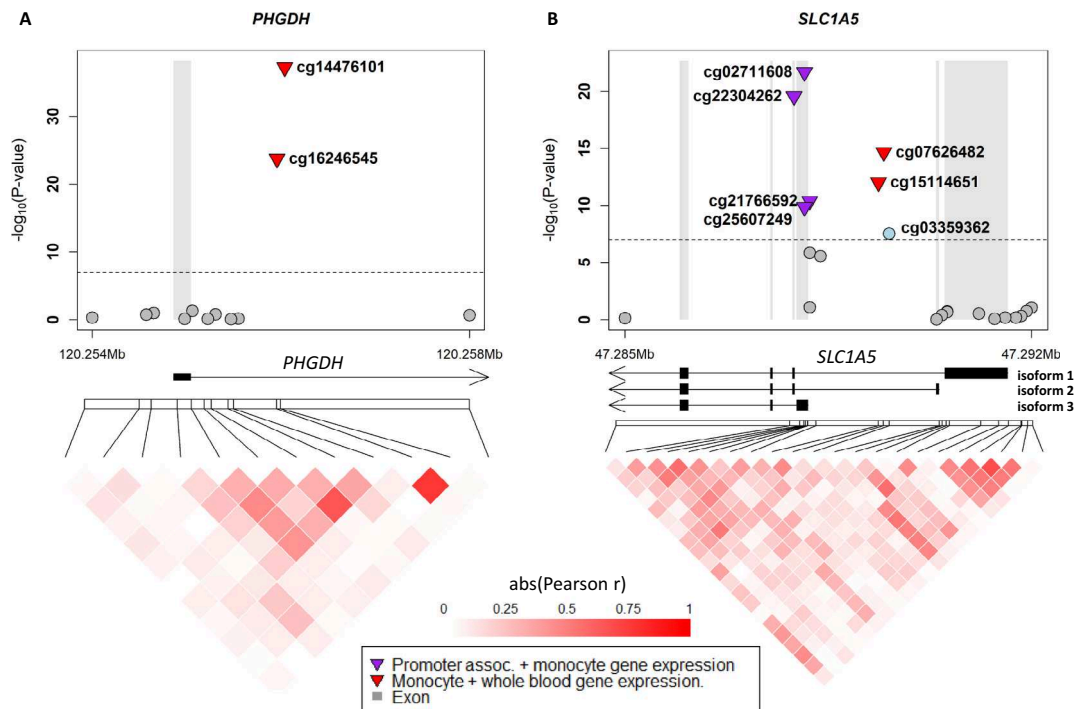
**Urate-associated CpGs, gene expression, and chromatin accessibility.** An important function of DNA methylation is the regulation of gene expression<sup>25</sup>. To determine whether the urate-associated CpGs were associated with gene expression in blood cells, we used summary statistics of DNA methylation and gene expression from monocytes and whole blood. Gene expression in monocytes is particularly relevant for *SLC2A9*, which is mainly expressed in monocytes and myeloid dendritic cells among the blood cell types (Supplementary Fig. 11)<sup>26</sup>. Of the 99 replicated CpGs, 26 CpGs were significantly associated with gene expression in *cis* in monocytes ( $p < 5E-4$ ; Supplementary Data 14)<sup>27</sup>. Some of the genes harbored multiple urate-associated CpGs (two at *PHGDH*, five at *SLC2A9*, six at *SLC1A5*). At *PHGDH* and

*SLC1A5*, higher DNA methylation levels were associated with lower gene expression, with CpGs mapping to intron 1 or intron 2 of the gene, depending on the reference isoform (Fig. 3). The inverse association between DNA methylation levels in intron 1 and gene expression has been reported to be conserved across vertebrates and is found in multiple human tissues<sup>28</sup>. Interestingly, at *SLC2A9*, the effect direction on gene expression for the five CpGs differed by genomic location (Fig. 4A, B). The three CpGs annotated as promoter-associated and located at introns 1 or 2 of isoform 1 (cg02734326, cg00071950, cg11266682) showed inverse associations with gene expression (e.g., cg11266682: effect  $-0.037$ ,  $p = 4.16E-15$ ; Fig. 4A and Supplementary Data 14). These three promoter-associated CpGs were associated with higher serum urate levels in EWAS (Fig. 4C). In contrast, cg03725404, located further downstream in the direction of transcription, also showed inverse associations with gene expression (effect  $-0.024$ ,  $p = 4.75E-6$ , Fig. 4B) but was associated with lower serum urate levels in EWAS (Fig. 4C). Both cg11266682 and cg03725404 were also shown to have significant causal effects on serum urate, as reported above. The different directions of association across CpGs in *SLC2A9* indicate complex regulation. Assuming that DNA methylation affects gene expression levels, which in turn influence serum urate levels, the observation that higher DNA methylation levels at the three promoter-associated CpGs in *SLC2A9* were associated with lower gene expression in monocytes and higher serum urate levels (Fig. 4B, C) allows for the inference that *SLC2A9* expression in monocytes due to differential methylation at these promoter-associated CpGs has an inverse relationship with serum urate levels. As depicted in Fig. 5A, cg11266682 was associated with lower gene expression levels but showed a positive, causal effect on serum urate levels. In contrast, cg03725404, a CpG further downstream in the direction of transcription, was also associated with lower gene expression but had an inverse, causal effect on serum urate levels (Fig. 5B).

For gene expression in whole blood, eight CpGs showed significant associations (Supplementary Data 15). For all seven CpGs with significant associations with gene expression in both monocytes and whole blood (two in *PHGDH*, one in *PARP3*, two in *UAP1L1*, and two in *SLC1A5*), the directions of effect in the two gene expression datasets were consistent. CpGs at *SLC2A9* were not associated with gene expression in whole blood.

There are two main transcripts of *SLC2A9*. Isoform 1 encodes the long version of the GLUT9 protein, which has been reported to localize to the basolateral membrane of renal proximal tubule cells, where it mediates urate reabsorption<sup>29</sup>. We investigated whether the DNA methylation pattern of the urate-associated CpGs in the kidney may be similar to those observed in whole blood. Whole genome bisulfite sequencing (WGBS) of kidney samples ( $n = 5$ ) showed lower DNA methylation levels at the urate-associated CpGs in the promoter-associated region of isoform 1, and the pattern of DNA methylation level aligned with gene expression levels from RNA-sequencing (Fig. 6A). We further studied whether the urate-associated CpGs mapped into open chromatin regions, where transcription factor binding may occur. ATAC-seq data from primary human kidney tissue ( $n = 3$  each from micro-dissected cortex and medulla) showed that the three promoter-associated CpGs, and—to a lesser extent cg20479063—mapped to DNA sequences with high chromatin accessibility in the kidney (Fig. 6B). Taken together, the data from the kidney samples suggest that the urate-associated CpGs at *SLC2A9* are in a regulatory region with potential effects on the expression of *SLC2A9*.

Lastly, we investigated differential gene expression in kidney tissue using two resources. Of the six urate-associated CpGs in *SLC2A9* that were associated with differential *SLC2A9* expression



**Fig. 3 Associations of CpGs at *PHGDH* and *SLC1A5* with serum urate levels.** For both *PHGDH* (A) and *SLC1A5* (B), the upper part shows chromosomal positions on the x-axis and the  $-\log_{10}(p\text{-value})$  on the y-axis, and the lower part shows the correlations between DNA methylation levels of the CpGs. The replicated CpGs are labeled. Promoter-associated annotation was based on the HM450K annotation file. The gene models were based on RefSeq curated genes. Associations of DNA methylation with monocyte gene expression were from Kennedy et al. BMC Genomics 2018 and with whole blood from the KORA study (Methods). The correlations between the CpGs were generated using DNA methylation data from 804 European American participants of the ARIC study. Gene models were based on Genbank RefSeq (Accession numbers. *PHGDH*, NM\_006623; *SLC1A5* isoform 1, NM\_005628; isoform 2, NM\_001145144; isoform 3 NM\_001145145). assoc. associated.

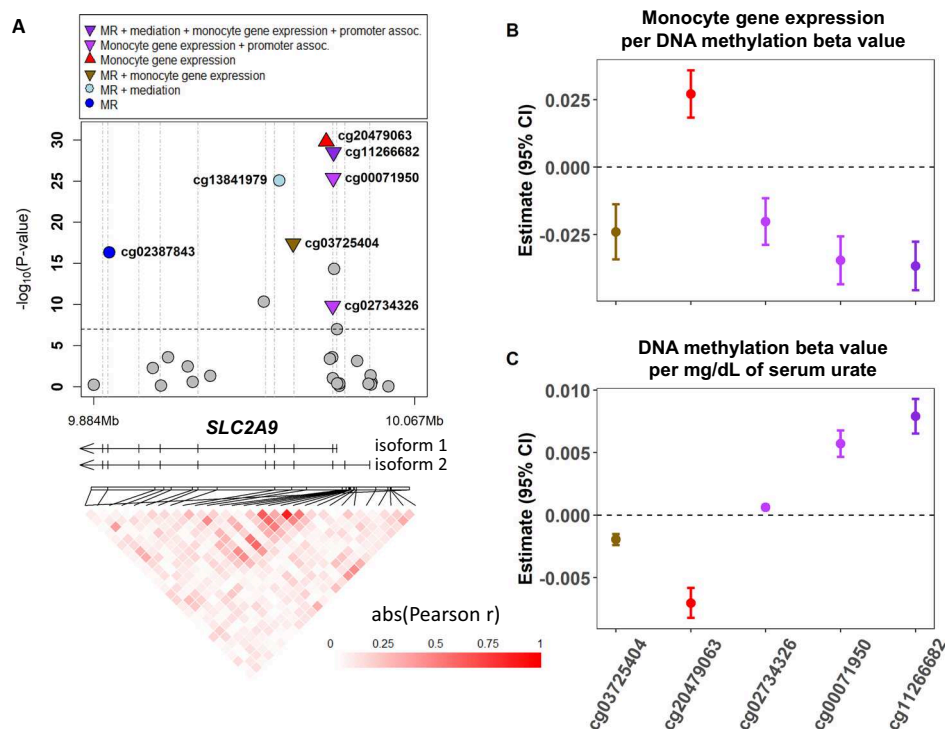
in monocytes, four did not show significant associations based on DNA methylation and gene expression levels quantified from micro-dissected human tubule tissue, and two did not meet quality control criteria (Methods, Supplementary Data 16). In the future, investigation of specific *SLC2A9* isoforms and/or of specific kidney cell types may shed light on the potential different effects of DNA methylation on gene expression levels in the kidney and monocyte. Second, we evaluated whether the genes implicated by the urate-associated CpGs from our study showed evidence for differential expression in kidney or intestine of a humanized mouse model of hyperuricemia (Methods). Of the genes nearest to the urate-associated CpGs, four genes were differentially expressed at false discovery rate (FDR) < 0.05 (*ANKRD11*, *CEBPB*, *SLC7A11*, and *UAP1L1*). *SLC7A11* had lower expression in the intestine of mice with hyperuricemia ( $\log_2$  fold change:  $-1.34$ , FDR =  $2.9E-2$ ). *ANKRD11* had lower expression in the kidney ( $\log_2$  fold change:  $-0.24$ , FDR =  $5.6E-3$ ; Supplementary Data 17). These findings are consistent with the urate-associated CpGs being downstream of factors influencing serum urate levels in mice.

**Enrichment of urate-associated CpGs in DNase I hypersensitive sites and transporters.** The urate-associated CpGs were enriched in DNase I hypersensitive sites of all blood cells tested (FDR  $q < 0.01$ ), with the strongest enrichment observed in CD3 cells ( $q < 1e-9$ ). Among histone modifications, H3K4me1, often found at active and primed enhancers, and H3K4me3, often found at active promoters, were enriched in all blood cells tested. While the former were most enriched in primary monocytes and primary hematopoietic stem cells, H3K4me3 showed the strongest enrichment in primary hematopoietic stem cells (Supplementary Figs. 12 and

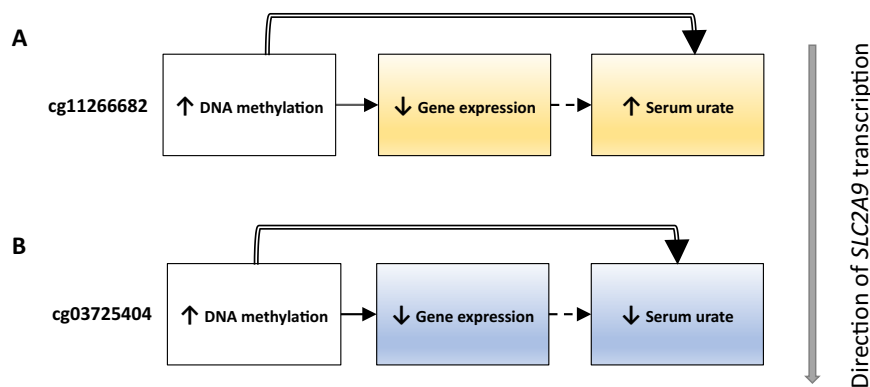
13)<sup>30,31</sup>. In contrast, H3K9me3 and H3K27me3, generally associated with heterochromatin<sup>30–33</sup>, were not significantly enriched in any blood cells or tissues tested (Supplementary Figs. 14 and 15). In addition, among transcription factor binding sites, the urate-associated CpGs were most enriched in POLR2A, the largest sub-unit of RNA polymerase II, the major enzyme synthesizing mRNA in eukaryotes ( $p < 1E-11$ , Supplementary Fig. 16). Together, the observed enrichments suggest a role of the urate-associated CpGs in transcriptional regulation.

We also assessed the enrichment of genes implicated by urate-associated CpGs in Gene Ontology (GO) terms, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathways (Methods). After correction for multiple testing, significant enrichment was observed for 55 terms (Supplementary Data 18). The terms or pathways with the strongest enrichment were related to transmembrane transport of organic acids, carboxylic acids, amino acids, organic and inorganic anions, as well as processes in leukocytes and myeloid cells (Fig. 7). These terms suggest that differentially methylated CpGs capture information about small molecule transport in urate homeostasis.

**Urate-associated CpG sites and other traits.** Our prior work on the genetics of serum urate revealed significant genetic correlations between serum urate and many cardiometabolic traits<sup>3</sup>. Using published EWAS summary statistics, we investigated whether these relationships were also observed when studying epigenetic variation by examining the association between replicated urate-associated CpGs and other cardiometabolic traits (Methods, Supplementary Data 19). Liver and kidney traits were also included due to the important role of these two organs in urate production and excretion<sup>34</sup>. Most of these traits were



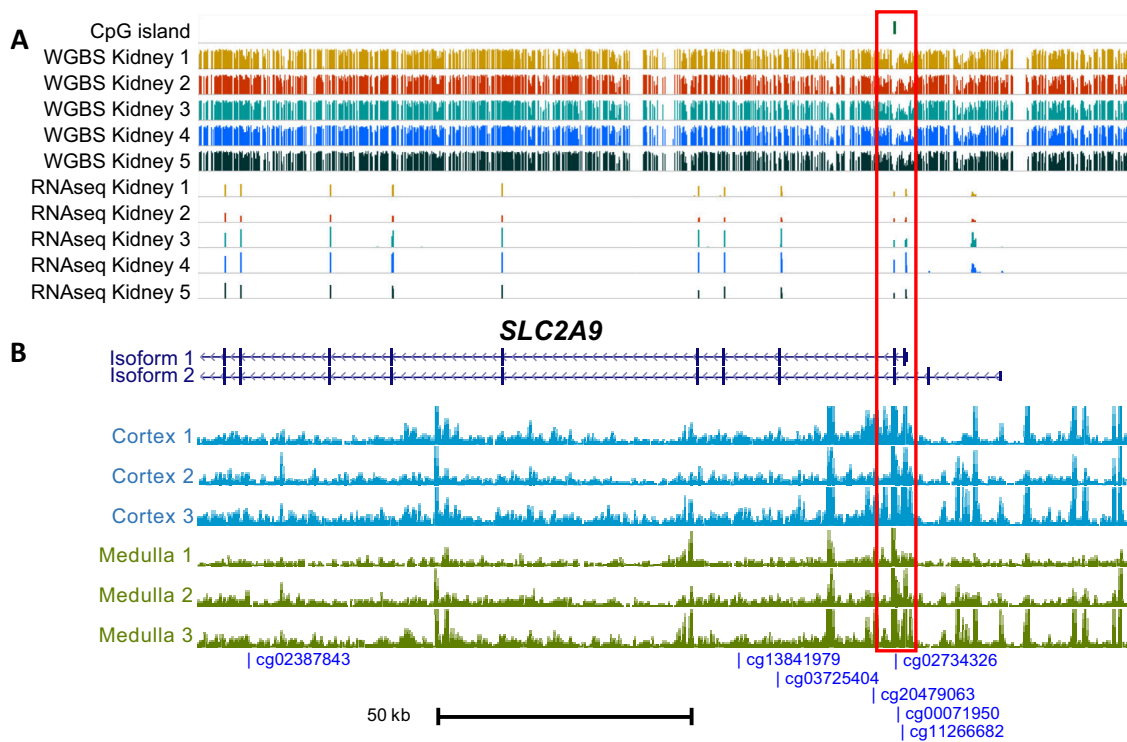
**Fig. 4 Associations of CpGs at *SLC2A9* with serum urate levels and gene expression in monocytes.** CpGs in the *SLC2A9* region (A) with the effect size of DNA methylation on gene expression in monocytes at five replicated CpGs (B) and the effect size of serum urate on DNA methylation levels at these same five CpGs (C). All labeled CpGs were replicated. In the legend, MR indicates CpGs with a significant causal effect on serum urate levels based on Mendelian randomization analysis. Colors of the estimates in panels B and C match the color legend in panel A. Promoter-associated annotation was based on the Illumina HM450K annotation file. The gene models were based on RefSeq curated genes. The correlations between the CpGs were generated using DNA methylation data from 804 European Americans participants of the ARIC study. The association between DNA methylation and monocyte gene expression in panel B is based on Kennedy et al. BMC Genomics 2018 ( $n = 1202$ ). The DNA methylation estimates in panel C are based on the meta-analysis combining discovery and replication cohorts in the present study ( $n = 17,996$ ). Genbank RefSeq accession numbers: isoform 1 (NM\_020041), isoform 2 (NM\_001001290). MR Mendelian randomization, assoc. associated.



**Fig. 5 Conceptual figure summarizing patterns of relationships between DNA methylation at *SLC2A9*, its expression in monocytes, and serum urate levels.** The figure focuses on the two CpGs with significant causal effects on serum urate and association with gene expression in monocytes. The inferred relationship between gene expression and serum urate levels is inverse for the promoter-associated CpG cg11266682 (A, orange shading), and concordant for CpG cg03725404 (B, blue shading). Solid compound arrows indicate both observed cross-sectional association and causal effect of DNA methylation on serum urate based on Mendelian randomization analysis. Solid arrows indicate observed cross-sectional associations. Dashed arrows indicate inferred relationships.

included as covariates in the EWAS of serum urate with the goal of detecting urate-specific DNA methylation signals (Methods). Of the replicated urate-associated CpGs, 17 had significant associations with at least one cardiometabolic trait (Fig. 8 and Supplementary Data 20). Interestingly, the directions of association were always consistent with the correlation between serum urate and these traits in epidemiological studies (Fig. 8)<sup>7,35–37</sup>. A

few urate-associated CpGs showed associations with a particularly high number of cardiometabolic traits: cg14476101 annotated to *PHGDH* (six traits), cg06690548 annotated to *SLC7A11* (seven traits), and cg02711608 annotated to *SLC1A5* (six traits). Lower DNA methylation level were associated with higher serum urate levels, as were as higher levels of body mass index (BMI), blood pressure, triglycerides, liver enzymes, C-reactive protein



**Fig. 6 Whole genome bisulfite sequencing and ATAC-seq data of kidney tissue at *SLC2A9*.** Whole genome bisulfite sequencing of kidney samples ( $n = 5$ ) showed that lower DNA methylation levels at CpGs localized to the promoter-associated region in isoform 1 on *SLC2A9* and aligned with gene expression levels from RNA-sequencing (**A**). ATAC-sequencing data from primary human kidney tissue ( $n = 3$  from micro-dissected cortex and medulla each) showed that the three promoter-associated CpGs, and—to a lesser extent *cg20479063*—mapped to DNA sequences with high chromatin accessibility in the kidney (**B**). Gene models were based on Genbank RefSeq (accession numbers: isoform 1, NM\_020041; isoform 2, NM\_001001290).

(CRP), and incident diabetes in their respective EWAS (Fig. 8). These traits represent components of the metabolic syndrome or are strongly correlated with the components<sup>38</sup>, as is serum urate<sup>37</sup>. Given that the EWAS of the cardiometabolic traits adjusted for white blood cell composition in their analysis, the associations between DNA methylation and traits at the CpGs are independent of cell type composition. Together, these associations at these CpGs may reflect DNA methylation signatures of metabolic syndrome in blood. In contrast, none of the 99 replicated urate-associated CpGs were associated with fasting glucose and insulin among individuals without diabetes, who are less likely to be affected by the metabolic syndrome. The urate-associated CpGs were also not found among those associated with high-density and low-density cholesterol (HDL-C, LDL-C) and estimated glomerular filtration rate (eGFR). In addition, 4 of the 17 CpGs that were associated with serum urate and cardiometabolic traits (*cg16246545* at *PHGDH*, *cg19693031* at *NBPF20/TXNIP*, *cg06690548* at *SLC7A11*, and *cg06690548* at *CPT1A*) were associated with mostly lipid-related metabolites, quantified using nuclear magnetic resonance (Supplementary Data 21)<sup>39</sup>. The associations of these four CpGs were largely observed with triglycerides in various lipid subfractions or lipid metabolites that are part of very low-density lipoprotein and had concordant effect directions as their associations with serum urate.

## Discussion

This large-scale EWAS, based upon data from up to 17,996 participants largely drawn from population-based studies, identified and replicated 99 CpGs, at which differential DNA methylation was significantly associated with serum urate levels. The genes implicated by these CpGs were strongly enriched in terms and/or pathways

related to small molecule transport, including organic anions and amino acids. MR analyses supported significant causal effects of some CpGs on serum urate levels and gout at *SLC2A9*, the strongest GWAS locus for serum urate that encodes a major urate transport protein in the kidney. Mediation analyses supported that genetic variants at *SLC2A9* affect serum urate levels partly via epigenetic mechanisms. Moreover, 17 urate-associated CpGs showed significant associations with numerous cardiometabolic traits, with the direction of association always consistent with the ones between serum urate and the respective cardiometabolic trait reported from observational studies. Differential DNA methylation at these sites may therefore reflect an epigenetic signature of cardiometabolic traits in whole blood.

Results from dedicated EWAS of serum urate levels have not been reported previously. One previous EWAS of blood metabolite levels in one of our contributing studies, KORA F4, examined urate as one of 649 metabolites, and reported significant associations between DNA methylation at *cg00071950* in *SLC2A9* and serum urate levels<sup>40</sup>. This EWAS meta-analysis of measured serum urate levels lends further support to the reproducibility and validity of this prior finding through successful replication in different study populations, association with gene expression, and biological plausibility.

In addition to replicated associations between DNA methylation and serum urate, our study reveals significant causal effects of DNA methylation at some CpGs of *SLC2A9* on serum urate and gout. Prior exposure to serum urate in its soluble or crystal form has been shown to heighten the proinflammatory response of myeloid cells in vitro and in animal models potentially through epigenetic mechanisms<sup>41</sup>. This is also known as urate-induced training immunity. In our study, MR analysis did not identify significant causal effects of serum urate on DNA methylation, but



**Fig. 7** GO terms and KEGG and Reactome pathways that were enriched for urate-associated CpGs. Significance level was set at false discovery rate <math><0.05</math>. GO Gene Ontology, KEGG Kyoto Encyclopedia of Genes and Genomes.

it is conceivable that serum urate might act on other forms of epigenetic mechanisms, such as histone modification<sup>42</sup>.

Serum urate levels show high genetic and observational correlations with kidney function measures and with cardiometabolic traits. Previous large EWAS of kidney function measures focused on eGFR in population-based studies<sup>43</sup> or on changes in kidney function among patients with kidney disease<sup>44</sup>. Despite the strong, inverse genetic and observational correlations between serum urate and eGFR and the well-established role of the kidney in serum urate homeostasis, the replicated urate-associated CpGs in our study did not show significant associations when assessed in EWAS of eGFR. In contrast, many urate-associated CpGs showed significant associations in results from EWAS of cardiometabolic traits, including triglycerides<sup>45</sup>, blood pressure<sup>46</sup>, BMI<sup>47</sup>, liver enzymes<sup>48</sup>, CRP<sup>49</sup>, as well as diabetes<sup>50</sup>. Many of these traits are either conditions that define the metabolic syndrome, or directly related to such conditions (high triglyceride levels, high blood pressure, high BMI, high blood glucose levels)<sup>38</sup>. There has been no robust evidence supporting causal effects of serum urate on cardiometabolic traits<sup>51,52</sup>. Instead, our

observations are consistent with shared gene regulatory programs resulting in a common DNA methylation signature of serum urate and metabolic syndrome in whole blood.

Our study also provides additional insights into a major transporter of serum urate. *SLC2A9* has consistently been detected as the strongest locus in GWAS of serum urate with multiple independent signals<sup>1,3,16,53–57</sup>, but the mechanism underlying the GWAS signal is largely unknown. The association between genetic variants at *SLC2A9* and serum urate levels may be mediated by transcript- or tissue-specific regulatory mechanisms. *SLC2A9* has two described isoforms whose gene products have equivalent urate transport activity but differ in the N-terminal cytosolic portion of the protein<sup>58</sup>. *SLC2A9* shows high complexity in human kidney tissue; the two isoforms of *SLC2A9* have been reported to differ in cellular localization possibly affecting their function in urate homeostasis. Isoform 1 was found to be expressed at the basolateral side of the proximal tubule, while isoform 2 was expressed at the apical side of the collecting duct<sup>29</sup>. The functional consequences of this complexity are not yet clear. In our study, we observed complex relations of DNA

	Urate	BMI	SBP	DBP	TG	HDL-C & LDL-C	Inc. T2D	Insulin & glucose	GGT	ALT	CRP	eGFR
<b>Genetic correlation with urate</b>	1.00	0.37	0.23*	0.25*	0.36	0.37, 0.09	0.36 to 0.44	0.45	N/A	N/A	N/A	-0.26
<b>N for discovery (replication if reported)</b>	12,474 (5,522)	5,387 (4,874)	9,828 (7,182)		3,296		1,264	4,808 (11,750)	731 (719)		8,863 (4,111)	2,264 (2,595)
<b>CpG</b>	<b>Gene</b>											
cg01881899	<i>ABCG1</i>											
cg26470501	<i>BCL3</i>											
cg18120259	<i>C6orf223</i>											
cg00574958	<i>CPT1A</i>											
cg21429551	<i>GARS</i>											
cg19693031	<i>NBPF20</i>											
cg01243823	<i>NOD2</i>											
cg16246545	<i>PHGDH</i>											
cg14476101	<i>PHGDH</i>											
cg03725309	<i>SARS</i>											
cg22103219	<i>SH2B2</i>											
cg22304262	<i>SLC1A5</i>											
cg02711608	<i>SLC1A5</i>											
cg21766592	<i>SLC1A5</i>											
cg11376147	<i>SLC43A1</i>											
cg06690548	<i>SLC7A11</i>											
cg17061862	<i>WEE1</i>											

**Fig. 8 Urate-associated CpGs that were associated with cardiometabolic and kidney traits.** All traits that were included in the lookup are shown, including those without reported associations with urate-associated CpGs. Blue: positive association between DNA methylation and trait levels; Red: inverse association between DNA methylation and trait levels. BMI body mass index, SBP systolic blood pressure, DBP diastolic blood pressure, TG triglycerides, HDL-C high-density lipoprotein cholesterol, LDL-C low-density lipoprotein cholesterol, inc. incident, GGT Gamma-glutamyl transferase, ALT alanine aminotransferase, CRP C-reactive protein, eGFR estimate glomerular filtration rate. Genetic correlations were reported in Tin et al. Nat Genet 2019. \*Genetic correlation between serum urate and hypertension was 0.39. No EWAS of hypertension was available; SBP and DBP were used instead.

methylation levels at *SLC2A9* with gene expression in monocytes and serum urate. Our findings support an inferred inverse relationship between DNA methylation-related *SLC2A9* expression in monocytes and serum urate levels at some CpGs. An inverse relationship between *GLUT9* expression in the kidney and serum urate would be difficult to explain. Loss-of-function mutations at *SLC2A9* in humans result in type 2 familial renal hypouricemia (OMIM# 612076) supporting a role for *GLUT9* as a high-capacity urate transporter in tubular urate reuptake from urine into blood. Therefore, a positive association between *SLC2A9* gene expression in the kidney and serum urate levels would be expected<sup>58</sup>.

However, *SLC2A9* is expressed in a range of extra-renal tissues, with high abundance in the liver and lower abundance in the heart, lung, and intestine<sup>59–61</sup>. In mouse hepatocytes, *GLUT9* is expressed in the basolateral membrane<sup>60</sup>, across which urate uptake from blood into the cell can occur<sup>62</sup>. Therefore, the role of *GLUT9*-mediated transport of urate may potentially affect serum urate levels in different directions, depending on transcript isoform and tissue localization. For example, loss of function mutations in dogs and liver-specific inactivation of *SLC2A9* in a mouse model result in hyperuricemia<sup>62,63</sup>. In addition, basolateral localization of *GLUT9* has also been reported in mouse enterocytes of the small intestine<sup>64</sup>, and enterocyte-specific *Slc2a9* knock-out resulted in reduced intestinal urate excretion and hyperuricemia<sup>65</sup>. A role for *SLC2A9* in intestinal excretion would also be consistent with an inverse relationship between DNA methylation-related *SLC2A9* expression and serum urate levels. Therefore, it is conceivable that the DNA methylation signals we detected at *SLC2A9* in blood represent a signature for urate transport into enterocytes, hepatocytes, and/or other cell types, as a constituent of a urate secretion pathway, a mechanism that would lower serum urate levels. Roles for *SLC2A9* in urate uptake into cells in both secretory and reabsorption pathways may also help explain some of the heterogeneity of associations of SNPs in the *SLC2A9* locus<sup>3</sup>. Regardless of the mechanism, the finding that urate-associated genetic variants at *SLC2A9* could affect serum urate levels via epigenetic mechanisms may be of potential clinical relevance, as each SD higher DNA methylation at CpGs at *SLC2A9* conferred up to 57% lower odds of gout, the most common form of inflammatory arthritis in adults.

Terms or pathways enriched for urate-associated CpGs were mainly related to transmembrane transport of organic acids, carboxylic acids, amino acids, and organic and inorganic anions. This does not only align with uric acid as an organic acid, but also with several genes encoding enzymes and transporters related to amino acid generation and movement across membranes. For example, *PHGDH* encodes 3-phosphoglycerate dehydrogenase, which catalyzes the oxidation of 3-phosphoglycerate to 3-phosphohydroxypyruvate, the committed step in the biosynthesis of L-serine. Inhibition of *PHGDH* alters nucleotide metabolism via disruption of mass balance within central carbon metabolism, leading to changes in the pentose phosphate pathway and the tri-carboxylic acid cycle<sup>66</sup>. In addition, *PHGDH* also has a role in lipid metabolism. Knockdown of *PHGDH* in liver cell lines resulted in lower expression of *LPL*, the gene encoding lipoprotein lipase, and higher expression of *LDLR* and *ABCA1*, both important in lipid transport<sup>67</sup>. The genes *SLC1A5*, *SLC7A11*, and *SLC43A1* encode transport proteins for neutral amino acids<sup>68</sup>, cysteine as well as glutamate<sup>69</sup>, and for branched-chain amino acids<sup>70</sup>, respectively. The CpGs in genes involved in the transport of small molecules such as amino acids, represented the CpGs with the strongest associations in this EWAS of serum urate. These CpGs were also associated with other cardiometabolic traits. Taken together, these results suggest that shared epigenetic regulatory mechanisms may have an important role in urate homeostasis<sup>15,71–73</sup>.

Strengths of our study include its large sample size, the replication of findings in separate study samples, and the use of rigorous statistical methods. The agreement of findings based on multiple layers of complementary evidence and the biological plausibility of many of the results support the validity of our findings. These findings provide new insights into the epigenetic regulation of serum urate levels that are complementary to genetic association studies of serum urate. Potential limitations of our study relate to the quantification of DNA methylation from whole blood or from specific blood cell populations, which may not be representative of other important organs of urate homeostasis such as the kidney or liver. However, serum urate was also measured in blood, where its levels may be influenced by metabolic functions beyond simply reflecting purine metabolism. In addition, serum urate concentrations play an important role in gout,

making blood an attractive target tissue for studies of urate homeostasis and gout.

In summary, meta-analyses of EWAS of serum urate levels and downstream characterization identify urate-associated CpGs that explain a substantial proportion of serum urate variance, mediate some of the genetic effects on serum urate levels, and show associations with several cardiometabolic traits consistent with the observed relationships between serum urate and these traits. These findings constitute complementary evidence to insights from GWAS of serum urate levels.

## Methods

**Study samples and workflow.** The discovery and replication analyses were conducted based on a pre-specified analysis plan. A script (<https://github.com/genepi-freiburg/ckdgen-pheno/tree/ckdgen-ewas-pheno>) was circulated to all participating cohorts (Supplementary Data 1) to generate summary files that were inspected by a central analysis team. All studies were community-based or non-clinical populations. Each cohort conducted study-specific EWAS and uploaded epigenome-wide summary statistics. Before the meta-analysis, we conducted quality control of study-specific results, which included the assessment of the distributions of effect sizes, standard errors, and  $p$  values within each study and across studies. Supplementary Data 2 reports study-specific methods, including the DNA methylation detection  $p$  value and processing pipeline. Supplementary Data 3 reports the white blood cell proportions in each cohort, which were used as covariates to estimate associations between urate and DNA methylation independent of cell type composition. Study research protocols were approved by the respective ethics committees. All participants in all studies provided written informed consent.

**DNA methylation quantification and quality control.** Genomic DNA was extracted from peripheral blood. Levels of DNA methylation were quantified using the Infinium MethylationEPIC BeadChip array (EPIC, six studies), the Illumina Infinium HumanMethylation450K BeadChip array (HM450K, 17 studies), or the Illumina Infinium HumanMethylation27 BeadChip array (HM27K, one study). DNA methylation data preprocessing was performed according to individual study protocols, which included background correction, quantile normalization, probe filtering, sample filtering, SNP matching to the SNP control probe locations, outlier filtering, and assay type correction (Supplementary Data 2).

**Study-specific EWAS.** Ancestry-specific association analyses were conducted within each study. Among the 24 contributing cohorts, 14 cohorts excluded participants who were on urate-lowering medications, whereas the other cohorts did not have this information available (Supplementary Data 1). The methylation levels at each CpG probe were represented as beta values, which can be interpreted as the proportion of CpG sites that were methylated. The beta values were analyzed as the dependent variable with serum urate as the independent variable in a linear regression model adjusting for age, sex, BMI, current smoking, eGFR, HDL-C, systolic blood pressure (SBP), log-transformed levels of CRP and triglycerides, genetic principal components (PC), and estimated or measured blood cell type proportions in order to allow for detecting associations between serum urate and DNA methylation levels independent of differences in cell type composition<sup>74</sup>. Additional study-specific technical covariates included control probe PCs<sup>5</sup>, study center, processing batch, sentrix ID, and sentrix position, as applicable.

**Discovery and replication meta-analysis.** Studies were separated into discovery and replication cohorts by chronological order of contribution to this project (Supplementary Data 1). Only CpG sites common to both the EPIC and HM450K were included in the meta-analysis ( $n = 441,854$ ). CpG probes overlapping with SNPs were annotated using information from Illumina.

Prior to meta-analysis, each set of study-specific results was adjusted for test statistic inflation using the BACON method, which was developed to control for bias and inflation in EWAS using an empirical null distribution approach and assumes that the test statistics are a mixture of three distributions: negative, positive and null associations<sup>75</sup>. The inflation factor is estimated from the distribution of null association. We conducted inverse-variance weighted fixed effects meta-analysis using the R package metafor (version 2.1-0) for discovery, replication, and for combining the results of discovery and replication<sup>76</sup>. CpG sites were excluded if their available sample size was less than half of the sample size in the respective meta-analyses, or if their  $I^2$  heterogeneity estimate was >95%.

The statistical significance threshold for the discovery step was  $p < 1.1E-7$  ( $=0.05/441,854$ ). The replication criteria were:  $P_{\text{discovery}} < 1.1E-7$ ,  $P_{\text{replication}} < 0.05$  with consistent effect direction, and  $P_{\text{combined}} < 1.1E-7$ . For the replicated CpGs, we quantified study heterogeneity within an ancestry using the  $I^2$  from a meta-analysis including all studies within an ancestry, namely EA (17 studies) and AA (5 studies). We further quantified heterogeneity among the four ancestry groups using the  $I^2$  from a transethnic meta-analysis that combined the summary statistics of the meta-analyses of EA, and AA with those from the LOLIPOP study (SA), and the RODAM study (SSA). Given that there was only a single study each including

participants with SA and SSA ancestry, the  $I^2$  from this transethnic meta-analysis might represent both ancestry and study heterogeneities.

**Heritability of replicated CpG sites and explained urate variance.** We used three sources of DNA methylation heritability estimates from whole blood to characterize the replicated urate-associated CpGs: (a) a twin study of 1464 individuals from Great Britain [Hannon et al.], (b) a twin study of 2386 individuals from the Netherlands [van Dongen et al.], and (c) a family-based study of 614 individuals from Australia [McRae et al.]<sup>22–24</sup>. Hannon et al. included 426 monozygotic (MZ) and 306 same-sex dizygotic (DZ) twin pairs to estimate the proportion of the variance of DNA methylation explained by additive genetic (A), shared environment (C), and unshared environmental (E) factors<sup>22</sup>. Van Dongen et al. included 769 MZ and 424 DZ twin pairs for the ACE model analysis<sup>23</sup>. McRae et al. studied 614 individuals from 117 families and used intraclass correlation analysis to partition the variance of DNA methylation into additive genetics and environmental components<sup>24</sup>. DNA methylation levels of these three studies were quantified using the HM450K array.

The proportion of serum urate variance explained was calculated based on data from 1832 participants of the KORA-FF4 study with DNA methylation levels quantified using blood samples collected in 2014. Among the participants of the KORA-FF4 study, 988 were also participants of the KORA F4 cohort, one of the discovery cohorts, with DNA methylation levels quantified using blood samples collected in 2007, 7 years earlier than those in the KORA-FF4 study. Data were available for 97 of the 99 replicated, urate-associated CpGs (missing observations for cg08257009 and cg05201185). To enable comparison of the proportion of explained urate variance to the one obtained from age- and sex-adjusted GWAS summary statistics<sup>3</sup>, two models were computed: a base model of the proportion of urate variance explained by age and sex, and an extended model that additionally included residuals from a regression of the CpGs on sex, age, and cell type composition. The proportion of serum urate variance explained by the replicated CpGs was then calculated as the difference in coefficient of determination ( $R^2$ ) between the two models.

**Forward Mendelian randomization (MR) analysis: causal effects of DNA methylation on serum urate and gout.** To evaluate whether the urate-associated CpGs might have a causal role in influencing serum urate levels or gout, we conducted two-sample MR analyses using meQTLs as genetic instruments of the replicated CpGs. The random assignment of alleles during meiosis mimics treatment allocation in randomized controlled trials. Using meQTLs as proxies or genetic instruments of DNA methylation enables inferences on the causal effects of DNA methylation on serum urate or gout, provided that the meQTLs meet the instrumental variable assumptions: (1) the meQTLs are associated with DNA methylation, (2) the meQTLs are not associated with potential confounders of the association between DNA methylation and serum urate or gout, and (3) the meQTLs are only associated with serum urate or gout through DNA methylation, i.e., without pleiotropic effects<sup>77</sup>. While assumptions 2 and 3 cannot be fully verified, we have addressed these assumptions in our selection criteria of meQTLs and analysis methods as detailed below.

The meQTLs of the CpGs were selected two steps. In the first step, we addressed the three instrumental variable assumptions. To address assumption 1, we selected independent *cis* meQTLs as genetic instruments with strong association with DNA methylation ( $p < 5E-8$ ). To address assumptions 2 and 3, we removed those with genome-wide significant association ( $p < 5E-8$ ) with potential confounders, and with stronger associations with the outcome than with DNA methylation, which suggests pleiotropic effects of a meQTL or may lead to reverse causation. In this first step of instrument selection, we selected *cis* meQTLs for each CpG (within 500 kb of the CpG) from the meta-analysis of the GoDMC with  $MAF > 1\%$  and  $p < 5e-8$  to avoid weak instrument bias<sup>78</sup>. The meQTLs were further pruned based on the following criteria: (a) a clumping threshold of  $r^2 < 0.05$ , (b) associated with potential confounders (BMI, current smoking, HDL-C, SBP, CRP, and triglycerides) at  $p < 5E-8$ , (c) failed Steiger filtering, i.e., displayed stronger association with the outcome than the exposure, and (d) failed harmonization checks of the `harmonise_data` function in the `TwoSampleMR` package, which matches the effect allele of the SNPs associated with the exposure and the outcome and flags palindromic SNPs with allele frequency close to 0.5 as ambiguous<sup>79,80</sup>. The sources of the GWAS summary statistics of the potential confounders are reported in Supplementary Data 22. In the second step, to ensure that the meQTLs for each CpG were indeed conditionally independent, we conducted conditional analysis using Genome-wide Complex Trait Analysis (GCTA) to obtain the conditional  $p$  value for each meQTL controlling for other meQTLs of the same CpG (command: `cojo-cond`)<sup>81</sup>. As the reference panel, we used 13,558 randomly selected individuals of British descent, and 16,969,363 SNPs with  $MAF > 0.01\%$  after quality control as reported previously<sup>3</sup>. We retained only meQTLs with conditional  $p < 5E-8$  as genetic instruments of the CpGs. The median numbers of meQTLs among the 92 replicated CpGs outside of *SLC2A9* was 2 (25th, 75th percentile: 0, 5), whereas it was 20 (25th, 75th percentile: 8, 21) among the 7 CpGs at *SLC2A9*.

The genetic association summary statistics for meQTLs, serum urate, and gout were generated among EA individuals. The meQTL summary statistics were obtained from a meta-analysis of the GoDMC Consortium (max.  $n = 27,750$  from 36 EA studies), which included some of the participating studies of the EWAS of serum urate<sup>82</sup>. DNA methylation in GoDMC was measured in whole blood or cord

blood using the HM450K or EPIC BeadChips. The 1000 Genomes reference panel was used for genotype imputation. The analysis in each cohort had two steps. First, residuals of DNA methylation beta values were generated by adjusting out age, sex, predicted white blood cell type proportions, predicted smoking status based on published results of smoking-associated DNA methylation, and genetic PCs. Next, these residuals were rank transformed and standardized and used for genetic association analysis with adjustment for relatedness in family-based studies. The SNP effect sizes can be interpreted as SD higher DNA methylation per allele. The meta-analysis was conducted in two phases. First, each GoDMC study analyzed the association between all SNPs and all CpGs, returning only associations with  $p < 1E-5$ . Next, these associations with  $p < 1E-5$  were combined to create a candidate list of meQTL associations, which included *cis* associations found in at least one dataset and *trans* associations in at least two datasets. The association statistics of these candidate meQTLs ( $n = 120,212,413$ ) were obtained from all cohorts, and then combined using fixed-effect meta-analyses<sup>82</sup>.

The GWAS summary statistic for serum urate ( $n = 288,649$ ) and gout ( $n = 692,537$ ) were obtained from recent GWAS meta-analyses of EA, samples from the CKDGen Consortium<sup>3</sup>. We required each CpG to have at least four *cis* (<500 kb) meQTLs as genetic instruments. As the primary MR analysis method, we used the multiplicative random effect inverse-variance weighted method as recommended by the guidelines for MR investigation<sup>77</sup>. We used the Egger intercept to evaluate potential pleiotropy, and Cochran's Q test for assessing heterogeneity<sup>78,83,84</sup>. Sensitivity analysis using other MR methods that are robust to pleiotropy were conducted: MR Egger, simple mode, weighted median, and weighted mode<sup>85–87</sup>. To assess whether a causal estimate may be largely driven by a single genetic instrument, we conducted leave-one-out analysis. To assess whether the causal effects of the CpGs at *SLC2A9* on serum urate were partly due to GWAS signals at this locus, we conducted an additional sensitivity analysis excluding meQTLs with  $r^2 > 0.05$  with any of five GWAS index SNPs at *SLC2A9* (rs4447862, rs10017305, rs6825187, rs62286563, and rs73224492) among individuals of EA, reported in Tin et al.<sup>3</sup>. The SNP rs4447862 had the lowest  $p$  value in the *SLC2A9* region, and the other four SNPs were detected as genome-wide significant independent SNPs by the GCTA stepwise model selection command (cojo-slc)<sup>88</sup>. All MR analyses were conducted using the TwoSampleMR package<sup>79</sup>. Calculations of the minimum detectable effect size in the forward MR analysis of DNA methylation on serum urate or gout supported that the MR analysis was well powered (Supplementary Note 3).

**Reverse MR analysis: causal effect of serum urate on DNA methylation.** To evaluate the potential causal effects of serum urate on DNA methylation levels, we conducted two-sample reverse MR analysis. The genetic instruments for serum urate were 123 index SNPs from a recent GWAS meta-analysis of 288,649 EA participants<sup>3</sup>. The associations between these SNPs and DNA methylation levels were generated among 3866 EA participants of the Framingham Heart Study, a sample independent of the EWAS of serum urate<sup>89</sup>. The ARIC study was not included in the reverse MR because most meQTL data were from AA participants, whereas the urate summary statistics were based on data from EA individuals. The DNA methylation levels were quantified as beta values from the HM450K array. The analysis of the association between SNPs and DNA methylation levels adjusted for age, sex, the top 50 methylation PCs, predicted blood cell fractions, and used linear mixed model to account for family structure. The reverse MR did not use the GoDMC meQTL results employed in the forward MR analysis, because the GoDMC meQTL results only contained SNPs associated with DNA methylation levels below a certain significance level, as reported above. Given that the urate index SNPs were selected in separate 1-Mb regions across the genome, we did not conduct GCTA conditional analysis to reassess the independence among these index SNPs. Otherwise, the selection of the SNPs for serum urate, their harmonization, and MR methods for primary and sensitivity analysis were the same as those in the forward MR analysis.

**Mediation analysis.** Given that four CpGs at *SLC2A9* showed significant causal effects for serum urate, we evaluated whether these CpGs may mediate the genetic effects of the EA index SNPs of serum urate at the *SLC2A9* locus<sup>3</sup>. As reported previously, four independent SNPs in two neighboring 1 Mb intervals in the *SLC2A9* region were identified based on GCTA stepwise model selection<sup>3,81,88</sup>. Mediation analyses of the four CpGs for the four index SNPs were first conducted among 637 EA ARIC participants, controlling for age, sex, study center, and ten genetic PCs using the mediation package version 4.5 in R<sup>90</sup>. This mediation method uses simulation to estimate the average causal mediation effect in a potential outcome framework. The significance threshold was set at 0.0125 (0.05/4 independent index SNPs). For the significant index SNP-CpG pairs, mediation analyses were performed in the SHIP-Trend ( $n = 223$ ) and KORA ( $n = 1636$ ) cohorts. The mediation effects of each CpG were then combined using fixed-effect inverse-variance meta-analysis as implemented in metafor<sup>76</sup>.

**Association with gene expression and chromatin accessibility.** We investigated the association of the urate-associated CpGs with gene expression in *cis* in monocytes and whole blood. Monocyte data were obtained from Kennedy et al., who assessed these associations among 1202 individuals from the Multi-Ethnic Study of Atherosclerosis<sup>27</sup>.

DNA methylation levels were quantified using the HM450K array, and gene expression levels were quantified using Illumina HumanHT-12 version 3.0 and 4.0 Expression BeadChips. The regression analysis used log-transformed expression levels as the dependent variable and DNA methylation beta values as the independent variable controlling for age, race, sex, and study site.

The association of DNA methylation levels with gene expression in whole blood was obtained from 713 participants of the KORA study. The DNA methylation levels were quantified using the HM450K array, and gene expression levels were quantified using Illumina HumanHT-12 v3 Expression BeadChips. The log<sub>2</sub>-transformed gene expression values were regressed on the DNA methylation beta values adjusted for sex and age. Prior to the analysis the technical factors as well as the blood cell type proportions were regressed out of the mRNA and DNA methylation levels, and their residuals were included in the final association model. Statistical significance for the association between urate-associated CpGs and gene expression was defined as  $p < 5E-4$  ( $=0.05/99$ ).

To evaluate whether urate-associated CpGs at *SLC2A9* might be associated with gene expression in the kidney, we selected five CpGs at the *SLC2A9* locus that were significantly associated with *SLC2A9* gene expression in monocytes (cg03725404, cg20479063, cg02734326, cg00071950, cg11266682). The DNA methylation and gene expression levels were quantified from tubulo-interstitial kidney tissue of 314 persons. The analysis used gene expression (transcripts per million) as the dependent variable and DNA methylation beta values as the independent variable controlling for age, sex, five genetic PCs, prevalent diabetes and hypertension, BMI, batch factors, bisulfite conversion rate, RNA-sequencing batch, and RNA integrity number.

To explore the epigenetic and gene expression landscape in the *SLC2A9* locus in kidney tissues, we looked up the DNA methylation profiles detected by WGBS and gene expression profiles from RNA-sequencing in five human normal kidneys (GEO accession GSE115098)<sup>91</sup>. The methylation beta values from WGBS and read counts from RNA-sequencing at the *SLC2A9* locus (GRCh37/hg19, chr4:9884000-10067000) were converted to bigwig format. The bigwig files were imported into IGV genome browser for visualization. In addition, we investigated whether the urate-associated CpGs at *SLC2A9* might be mapped into open chromatin regions in the kidney using data generated from ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing). Kidney cortex and medulla tissues were macro-dissected from uninvolved portions of tumor nephrectomy specimens and snap-frozen. ATAC-seq was performed by ActiveMotif (Carlsbad, California) and the subsequent data were mapped to GRCh38 and normalized by read depth. Peak calling was performed using MACS 2.1.0 and filtered against the ENCODE blacklist. Master list generation and set operations with the resulting beta files were performed using bedops. GRCh38 positions were mapped back to GRCh37/hg19 using the online liftOver tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>)

**Differential gene expression in humanized mouse model of hyperuricemia.** To investigate whether the closest genes of the urate-associated CpGs might be differentially expressed in hyperuricemic condition, we obtained gene expression data from a humanized mouse model of hyperuricemia. The orthologous mouse variant to the human ABCG2 Q141K (ABCG2 Q140K) was knocked in to the endogenous *Abcg2* locus using CRISPR/Cas9 on a C57BL/6J background, resulting in hyperuricemia in adult male Q140K+/+ mice as described previously<sup>64</sup>. Animal studies were performed in adherence to the NIH Guide for the Care and Use of Laboratory Animals and approved by the University of Maryland School of Medicine Institutional Animal Care and Use Committee. Mice were housed in groups of two to five per cage on a 12:12 h light/dark cycle; with lights on at 6 a.m., only male mice were used for the analysis. Animal organs were harvested and preserved in RNALater (Sigma R0901) and gradually cooled from 4 °C to -80 °C. RNA was isolated using the RNeasy Plus Mini Kit (Qiagen 74134) per the manufacturer's protocol. RNA was suspended in molecular biology grade water and then quantified with a CLARIOstar plate reader and stored at -80 °C. Illumina Sequencing Libraries were prepared using manufacturer's protocol for NEB Ultra II Directional RNA Library Prep kit with poly-A enrichment (NEB E7760). Samples were sequenced on four flowcell lanes of an Illumina HiSeq4000 75 bp paired end run. Five samples were sequenced in each flowcell lane. Samples were grouped per lane by tissue type. Sample quality assessment, RNA-Seq library preparation, and sequencing were performed by the Genomics Resource Center at the University of Maryland School of Medicine. RNA-Seq data was stored and analyzed on BasePair (<https://app.basepairtech.com/>) for expression count (STAR) and differential expression (DESeq2) analyses. Mouse gene names were mapped to human gene names in HUGO Gene Nomenclature using the Mouse Genome Informatics database. FDR for differential expression was calculated using the Benjamini-Hochberg method. We consider FDR < 0.05 as being significant.

**Enrichment analyses.** To inform the potential functional effects of the urate-associated CpGs, we assessed the enrichment of these CpGs in sites of DNase I or histone modification (H3K4me1, H3K4me3, H3K9me3, H3K27me3), gene sets based on GO terms and pathways in the KEGG and Reactome databases<sup>92–95</sup>. The enrichment analyses of DNase I or histone modification were performed using a local version of eForge version 2.0 (experimentally derived Functional element Overlap analysis of ReGions from EWAS)<sup>96</sup>. Briefly, eForge used data sources from either the ENCODE (125 samples) or Roadmap Epigenomics (299 samples) projects generated by the Hotspot method<sup>97–99</sup>. The overlaps between the input CpGs

and the data sources were compared with those from 1000 random sets of CpGs with matching gene relationship and CpG island annotation. FDR was obtained based on the binomial distribution and the Benjamini–Yekutieli method for multiple testing correction. Our input included 1018 urate-associated CpGs with  $p < 1E-05$  in the combined meta-analysis of the discovery and replication cohorts. We performed 10,000 resampling runs with an active proximity filter and considered FDR  $< 0.01$  as significant, i.e., enrichment  $> 99$  percentile.

Enrichment in gene sets or pathways was performed using the methylGSA package and R version 3.6.1<sup>100</sup>. The enrichment test method (methylglm) was a functional class scoring method implemented using logistic regression accounting for the number of CpG sites per gene and the autosomal background that overlaps the HM450K and EPIC arrays. Gene sets or pathways with 100–500 genes were tested (default setting). We considered a gene set or pathway to be significantly enriched at FDR  $< 0.05$  correcting for multiple testing within each database using the Benjamini and Hochberg method<sup>101</sup>.

**Relationship of urate-associated CpGs with cardiometabolic traits.** Our prior work on the genetics of serum urate revealed significant genetic correlations between serum urate and many cardiometabolic traits and showed that genes mapping into urate-associated loci were highly enriched for expression in kidney and liver<sup>3</sup>. Other than gout, cardiometabolic traits with high genetic correlations ( $r^2 > 0.35$ ) included triglycerides, HDL cholesterol, diabetes, fasting insulin, hypertension, and BMI. To gain insights on whether these relationships were also observed when studying DNA methylation in blood, we investigated whether the urate-associated CpGs were also identified in EWAS of the respective cardiometabolic traits as well as in EWAS of eGFR and liver enzymes. We also included CRP, given that CRP is a marker of inflammation, a component of metabolic syndrome<sup>38</sup>. When more than one EWAS were available for a trait, EWAS were prioritized for the larger sample size. The primary focus was placed on the interpretation of the direction of association of DNA methylation with serum urate and the other traits. The CpGs included in the lookup were those replicated in the EWAS of urate and also considered significant in the EWAS of the other traits based on criteria of the respective study (Supplementary Data 19). All studies included in this lookup controlled for age, sex, and white blood cell type proportions for detecting independent association between DNA methylation and the trait. More details are provided in Supplementary Data 19. When a study included replication and provided meta-analysis results of discovery and replication, the effects from the meta-analysis combining discovery and replication were used.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Disclaimer.** The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the US Department of Health and Human Services.

## Data availability

The summary statistics from the meta-analysis are available from the CKDGen Consortium website (<https://ckdgen.imbi.uni-freiburg.de>). Additional data and programming code that support the findings of this study such as code used to call on software are available from the authors upon request.

## Code availability

The script for generating the phenotypes used in the EWAS is available via GitHub (<https://github.com/genepi-freiburg/ckdgen-pheno-ewas>)<sup>102</sup>. EWAS QC, meta-analysis, and postprocessing were implemented in R v4.0.1 using metafor v2.4.0, qqman v0.1.4, limma v3.42.2, openxlsx v4.1.5, car v3.0.8, bacon v1.16.0, mutoss v0.1.12, methylGSA v1.6.1, ggplot2 v3.3.3, SeSAMe v1.10.5, rmeta v3.0, mediation v4.5.0, TwoSampleMR R v0.5.6, and Genome-wide Complex Trait Analysis (GCTA) v1.93.2beta.

Received: 17 March 2021; Accepted: 8 November 2021;

Published online: 09 December 2021

## References

- Kottgen, A. et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat. Genet.* **45**, 145–154 (2013).
- Kanai, M. et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
- Tin, A. et al. Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat. Genet.* **51**, 1459–1474 (2019).
- Roddy, E. & Choi, H. K. Epidemiology of gout. *Rheum. Dis. Clin. North Am.* **40**, 155–175 (2014).
- Choi, H. K., Atkinson, K., Karlson, E. W., Willett, W. & Curhan, G. Purine-rich foods, dairy and protein intake, and the risk of gout in men. *N. Engl. J. Med.* **350**, 1093–1103 (2004).
- Dalbeth, N. et al. Gout. *Nat. Rev. Dis. Prim.* **5**, 69 (2019).
- Feig, D. I., Kang, D. H. & Johnson, R. J. Uric acid and cardiovascular risk. *N. Engl. J. Med.* **359**, 1811–1821 (2008).
- Sundström, J. et al. Relations of serum uric acid to longitudinal blood pressure tracking and hypertension incidence. *Hypertension* **45**, 28–33 (2005).
- Juraschek, S. P., Miller, E. R. III & Gelber, A. C. Body mass index, obesity, and prevalent gout in the United States in 1988–1994 and 2007–2010. *Arthritis Care Res. (Hoboken)* **65**, 127–132 (2013).
- Emmerson, B. T., Nagel, S. L., Duffy, D. L. & Martin, N. G. Genetic control of the renal clearance of urate: a study of twins. *Ann. Rheum. Dis.* **51**, 375–377 (1992).
- Rice, T. et al. Heterogeneity in the familial aggregation of fasting serum uric acid level in five North American populations: the Lipid Research Clinics Family Study. *Am. J. Med. Genet.* **36**, 219–225 (1990).
- Major, T. J., Dalbeth, N., Stahl, E. A. & Merriman, T. R. An update on the genetics of hyperuricaemia and gout. *Nat. Rev. Rheumatol.* **14**, 341–353 (2018).
- Mandal, A. K. & Mount, D. B. The molecular physiology of uric acid homeostasis. *Annu. Rev. Physiol.* **77**, 323–345 (2015).
- Futagi, Y., Narumi, K., Furugen, A., Kobayashi, M. & Iseki, K. Molecular characterization of the orphan transporter SLC16A9, an extracellular pH- and Na<sup>+</sup>-sensitive creatine transporter. *Biochem. Biophys. Res. Commun.* **522**, 539–544 (2020).
- Woodward, O. M. et al. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc. Natl Acad. Sci. USA.* **106**, 10338–10342 (2009).
- Vitart, V. et al. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat. Genet.* **40**, 437–442 (2008).
- Kooner, J. S. et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.* **43**, 984–989 (2011).
- Voight, B. F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
- Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
- Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
- Kilpinen, H. & Dermitzakis, E. T. Genetic and epigenetic contribution to complex traits. *Hum. Mol. Genet.* **21**, R24–R28 (2012).
- Hannon, E. et al. Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. *PLoS Genet.* **14**, e1007544–e1007544 (2018).
- van Dongen, J. et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat. Commun.* **7**, 11115 (2016).
- McRae, A. F. et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol.* **15**, R73 (2014).
- Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465–476 (2008).
- Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
- Kennedy, E. M. et al. An integrated -omics analysis of the epigenetic landscape of gene expression in human blood cells. *BMC Genomics* **19**, 476 (2018).
- Anastasiadi, D., Esteve-Codina, A. & Piferrer, F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* **11**, 37 (2018).
- Kimura, T., Takahashi, M., Yan, K. & Sakurai, H. Expression of SLC2A9 isoforms in the kidney and their localization in polarized epithelial cells. *PLoS One* **9**, e84996 (2014).
- Local, A. et al. Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat. Genet.* **50**, 73–82 (2018).
- Sharifi-Zarchi, A. et al. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics* **18**, 964 (2017).
- Young, M. D. et al. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res.* **39**, 7415–7427 (2011).
- Nicetto, D. et al. H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science* **363**, 294–297 (2019).
- Maiuolo, J., Oppedisano, F., Gratteri, S., Muscoli, C. & Mollace, V. Regulation of uric acid metabolism and excretion. *Int J. Cardiol.* **213**, 8–14 (2016).
- Kodama, S. et al. Association between serum uric acid and development of type 2 diabetes. *Diabetes Care* **32**, 1737–1742 (2009).
- Krishnan, E., Kwoh, C. K., Schumacher, H. R. & Kuller, L. Hyperuricemia and incidence of hypertension among men without metabolic syndrome. *Hypertension* **49**, 298–303 (2007).

37. Choi, H. K. & Ford, E. S. Prevalence of the metabolic syndrome in individuals with hyperuricemia. *Am. J. Med.* **120**, 442–447 (2007).
38. Grundy Scott, M., Brewer, H. B., Cleeman James, I., Smith Sidney, C. & Lenfant, C. Definition of metabolic syndrome. *Circulation* **109**, 433–438 (2004).
39. Gomez-Alonso, M. D. C. et al. DNA methylation and lipid metabolism: an EWAS of 226 metabolic measures. *Clin. Epigenetics* **13**, 7 (2021).
40. Petersen, A.-K. et al. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum. Mol. Genet.* **23**, 534–545 (2014).
41. Cabău, G., Crișan, T. O., Klück, V., Popp, R. A. & Joosten, L. A. B. Urate-induced immune programming: consequences for gouty arthritis and hyperuricemia. *Immunol. Rev.* **294**, 92–105 (2020).
42. Crișan, T. O. et al. Soluble uric acid primes TLR-induced proinflammatory cytokine production by human primary cells via inhibition of IL-1Ra. *Ann. Rheum. Dis.* **75**, 755–762 (2016).
43. Chu, A. Y. et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat. Commun.* **8**, 1286 (2017).
44. Gluck, C. et al. Kidney cytosine methylation changes improve renal function decline estimation in patients with diabetic kidney disease. *Nat. Commun.* **10**, 2461 (2019).
45. Dekkers, K. F. et al. Blood lipids influence DNA methylation in circulating cells. *Genome Biol.* **17**, 138 (2016).
46. Richard, M. A. et al. DNA methylation analysis identifies loci for blood pressure regulation. *Am. J. Hum. Genet.* **101**, 888–902 (2017).
47. Wahl, S. et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81–86 (2017).
48. Nano, J. et al. Epigenome-wide association study identifies methylation sites associated with liver enzymes and hepatic steatosis. *Gastroenterology* **153**, 1096–1106.e2 (2017).
49. Lighthart, S. et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* **17**, 255 (2016).
50. Cardona, A. et al. Epigenome-wide association study of incident type 2 diabetes in a british population: EPIC-Norfolk Study. *Diabetes* **68**, 2315–2326 (2019).
51. Li, X. et al. MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. *Ann. Rheum. Dis.* **77**, 1039–1047 (2018).
52. Li, X. et al. Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. *BMJ* **357**, j2376 (2017).
53. Doring, A. et al. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat. Genet.* **40**, 430–436 (2008).
54. Yang, Q. et al. Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ. Cardiovasc. Genet.* **3**, 523–530 (2010).
55. Matsuo, H. et al. Genome-wide association study of clinically defined gout identifies multiple risk loci and its association with clinical subtypes. *Ann. Rheum. Dis.* **75**, 652–659 (2016).
56. Li, S. et al. The GLUT9 gene is associated with serum uric acid levels in Sardinia and Chianti cohorts. *PLoS Genet.* **3**, e194 (2007).
57. Tin, A. et al. Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum. Mol. Genet.* **20**, 4056–4068 (2011).
58. Anzai, N. et al. Plasma urate level is directly regulated by a voltage-driven urate efflux transporter URATv1 (SLC2A9) in humans. *J. Biol. Chem.* **283**, 26834–26838 (2008).
59. Phay, J. E., Hussain, H. B. & Moley, J. F. Cloning and expression analysis of a novel member of the facilitative glucose transporter family, SLC2A9 (GLUT9). *Genomics* **66**, 217–220 (2000).
60. Keembiyehetty, C. et al. Mouse glucose transporter 9 splice variants are expressed in adult liver and kidney and are up-regulated in diabetes. *Mol. Endocrinol.* **20**, 686–697 (2006).
61. Witkowska, K. et al. hGLUT9 as a novel urate transporter: its role in liver urate handling and functional study of SLC2A9 SNPs. *FASEB J.* **23**, 797.4–797.4 (2009).
62. Preitner, F. et al. Glut9 is a major regulator of urate homeostasis and its genetic inactivation induces hyperuricosuria and urate nephropathy. *Proc. Natl Acad. Sci. USA.* **106**, 15501–15506 (2009).
63. Bannasch, D. et al. Mutations in the SLC2A9 gene cause hyperuricosuria and hyperuricemia in the dog. *PLoS Genet.* **4**, e1000246 (2008).
64. Hoque, K. M. et al. The ABCG2 Q141K hyperuricemia and gout associated variant illuminates the physiology of human urate excretion. *Nat. Commun.* **11**, 2767 (2020).
65. DeBosch, B. J., Kluth, O., Fujiwara, H., Schürmann, A. & Moley, K. Early-onset metabolic syndrome in mice lacking the intestinal uric acid transporter SLC2A9. *Nat. Commun.* **5**, 4642 (2014).
66. Reid, M. A. et al. Serine synthesis through PHGDH coordinates nucleotide levels by maintaining central carbon metabolism. *Nat. Commun.* **9**, 5442 (2018).
67. Karabegović, I. et al. Epigenome-wide association meta-analysis of DNA methylation with coffee and tea consumption. *Nat. Commun.* **12**, 2830 (2021).
68. Kanai, Y. et al. The SLC1 high-affinity glutamate and neutral amino acid transporter family. *Mol. Asp. Med.* **34**, 108–120 (2013).
69. Fotiadis, D., Kanai, Y. & Palacín, M. The SLC3 and SLC7 families of amino acid transporters. *Mol. Asp. Med.* **34**, 139–158 (2013).
70. Boday, S., Fotiadis, D., Stoeger, C., Kanai, Y. & Palacín, M. The small SLC43 family: facilitator system I amino acid transporters and the orphan EEG1. *Mol. Asp. Med.* **34**, 638–645 (2013).
71. Nigam, S. K. & Bhatnagar, V. The systems biology of uric acid transporters: the role of remote sensing and signaling. *Curr. Opin. Nephrol. Hypertens.* **27**, 305–313 (2018).
72. Takada, T. et al. ABCG2 dysfunction increases serum uric acid by decreased intestinal urate excretion. *Nucleosides Nucleotides Nucleic Acids* **33**, 275–281 (2014).
73. Yano, H., Tamura, Y., Kobayashi, K., Tanemoto, M. & Uchida, S. Uric acid transporter ABCG2 is increased in the intestine of the 5/6 nephrectomy rat model of chronic kidney disease. *Clin. Exp. Nephrol.* **18**, 50–55 (2014).
74. Houseman, E. A. et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **13**, 86 (2012).
75. van Iterson, M., van Zwet, E. W., Heijmans, B. T. & the, B. C. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* **18**, 19 (2017).
76. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* **36**, 1–48 (2010).
77. Burgess, S. et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res.* **4**, 186 (2020).
78. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
79. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
80. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
81. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol. Biol.* **1019**, 215–236 (2013).
82. Min, J. L. et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.* **53**, 1311–1321 (2021).
83. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
84. Cochran, W. G. The comparison of percentages in matched samples. *Biometrika* **37**, 256–266 (1950).
85. Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat. Med.* **36**, 4705–4718 (2017).
86. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
87. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
88. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
89. Huan, T. et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat. Commun.* **10**, 4267 (2019).
90. Imai, K., Keele, L. & Yamamoto, T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat. Sci.* **25**, 51–71 (2010).
91. Park, J. et al. Functional methylome analysis of human diabetic kidney disease. *JCI Insight* **4**, e128886 (2019).
92. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
93. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
94. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
95. Fabregat, A. et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
96. Breeze, C. E. et al. eFORGE: a tool for identifying cell type-specific signal in epigenomic data. *Cell Rep.* **17**, 2137–2150 (2016).
97. Sabo, P. J. et al. Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc. Natl Acad. Sci. USA.* **101**, 16837 (2004).
98. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).

99. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
100. Ren, X. & Kuan, P. F. methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinformatics* **35**, 1958–1959 (2019).
101. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Methodol.* **57**, 289–300. <http://www.jstor.org/stable/2346101> (1995).
102. Schlosser, P., Teumer, A., Tin, A. & Köttgen, A. Phenotype generation code: epigenome-wide association studies identify DNA methylation associated with kidney function and damage. *genepi-freiburg/ckdgen-pheno-ewas*. *GitHub*. <https://doi.org/10.5281/zenodo.5578240> (2021).

## Acknowledgements

Study-specific acknowledgements and funding sources are listed as Supplementary Note 1.

## Author contributions

Drafting of manuscript: Anna Köttgen, Pascal Schlosser, Alexander Teumer, Chris H. L. Thio, Owen M. Woodward, Adrienne Tin. Bioinformatics: Shreeram Akilesh, Nasir A. Aziz, Xu Gao, Peter Henneman, Anselm Hoppmann, Roby Joehanes, Christine Ladd-Acosta, Hongbo Liu, Ake T. Lu, Pamela R. Matias-Garcia, Daniel L. McCartney, Karlijn A.C. Meeks, Christoph Nowak, Pascal Schlosser, Andrea Venema, Antoine Weihs. Critical review of manuscript: Adebowale A. Adeyemo, Charles Agyemang, Nasir A. Aziz, Andrea Baccarelli, Hermann Brenner, Jan Bressler, Monique M. B. Breteler, Cristian Carmeli, Loyal Chaker, Josef Coresh, Adolfo Correa, Simon R. Cox, Graciela E. Delgado, Kai-Uwe Eckardt, Arif B. Ekici, Karlhans Endlich, James S. Floyd, Xu Gao, Allan C. Gelber, Mohsen Ghanbari, Christian Gieger, Philip Greenland, Megan L. Grove, Sarah E. Harris, Peter Henneman, Christian Herder, Lifang Hou, Silva Kasela, Marcus E. Kleber, Wolfgang Koenig, Jaspal S. Kooner, Florian Kronenberg, Anna Köttgen, Christine Ladd-Acosta, Daniel Levy, Lars Lind, Marie Loh, Stefan Lorkowski, Riccardo E. Marioni, Pamela R. Matias-Garcia, Daniel L. McCartney, Karlijn A. C. Meeks, Lili Milani, Winfried März, Matthias Nauck, Christoph Nowak, Holger Prokisch, Bruce M. Psaty, Alex P. Reiner, Pascal Schlosser, Joel Schwartz, Sanaz Sedaghat, Jennifer A. Smith, Nona Sotoodehnia, Hannah R. Stocker, Johan Sundström, Katalin Susztak, Alexander Teumer, Chris H. L. Thio, Adrienne Tin, Uwe Völker, Melanie Waldenberger, Juliane Winkelmann, Yanan Zheng, Johan Ärnlöv. Statistical methods and analysis: Nasir A. Aziz, Cristian Carmeli, Graciela E. Delgado, Eliza Fraszczyk, Xu Gao, Sahar Ghasemi, Franziska Grundner-Culemann, Gibran Hemani, Peter Henneman, Anselm Hoppmann, Steve Horvath, Mikko A. Hurme, Shih-Jen Hwang, Roby Joehanes, Silva Kasela, Marcus E. Kleber, Anna Köttgen, Brigitte Kühnel, Christine Ladd-Acosta, Terho Lehtimäki, Dan Liu, Hongbo Liu, Marie Loh, Ake T. Lu, Pamela R. Matias-Garcia, Daniel L. McCartney, Karlijn A. C. Meeks, Josine L. Min, Pashupati P. Mishra, Christoph Nowak, Scott M. Ratliff, Pascal Schlosser, Sanaz Sedaghat, Brenton R. Swenson, Alexander Teumer, Chris H. L. Thio, Adrienne Tin, Andrea Venema, Antoine Weihs, Zhi Yu, Wei Zhao, Yanan Zheng. Functional studies: Victoria L. Halperin Kuhns, Hongbo Liu, Katalin Susztak, Owen M. Woodward. Phenotyping: Charles Agyemang, Murielle Bochud, Hermann Brenner, Monique M. B. Breteler, Simon R. Cox, James S. Floyd, Sarah E. Harris, Christian Herder, Sharon L. R. Kardia, Silva Kasela, Wolfgang Koenig, Anna Köttgen, Joyce B. J. van Meurs, Annette Peters, Bruce M. Psaty, Scott M. Ratliff, Johan Ärnlöv. Interpretation of results: Shreeram Akilesh, Nasir A. Aziz, Andrea Baccarelli, Hermann Brenner, Jan Bressler, Loyal Chaker, James S. Floyd, Xu Gao, Allan C. Gelber, Wolfgang Koenig, Anna Köttgen, Christine Ladd-Acosta, Daniel Levy, Pamela R. Matias-Garcia, Christoph Nowak, Bruce M. Psaty, Pascal Schlosser, Joel Schwartz, Ben Schöttker, Nona Sotoodehnia, Katalin Susztak, Alexander Teumer, Chris H. L. Thio, Adrienne Tin, Owen M. Woodward. Study design of an individual contributing study: Adebowale A. Adeyemo, Charles Agyemang, Andrea Baccarelli, Murielle Bochud, Hermann Brenner, Monique M. B. Breteler, Loyal Chaker, Adolfo Correa, Simon R. Cox, Mohsen Ghanbari, Christian Gieger, Gibran Hemani, Peter Henneman, Lifang Hou, Sharon L. R. Kardia, Jaspal S. Kooner, Daniel Levy, Ake T. Lu, Riccardo E. Marioni, Karlijn A.C. Meeks, Josine L. Min, Winfried März, Annette Peters, Holger Prokisch, Bruce M. Psaty, Alex P. Reiner, Joel Schwartz, Jennifer A. Smith, Harold Snieder, Jana V. van Vliet-Ostapchouk, Juliane Winkelmann, Bruce H.R. Wolfenbützel, Johan Ärnlöv. Subject recruitment: Charles Agyemang, Nasir A. Aziz, Andrea Baccarelli, Murielle Bochud, Hermann Brenner, Simon R. Cox, Jaspal S. Kooner, Lars Lind, Karlijn A. C. Meeks, Lili Milani, Matthias Nauck, Annette Peters, Joel Schwartz. Methylation assessment: Charles Agyemang, Andrea Baccarelli, Hermann Brenner, James S. Floyd, Xu Gao, Megan L. Grove, Xin Gao, Sarah E. Harris, Peter Henneman, Steve Horvath, Lifang Hou, Mikko A. Hurme, Silva Kasela, Marcus E. Kleber, Stefan Lorkowski, Daniel L. McCartney, Joyce B. J. van Meurs, Lili Milani, Jennifer A. Smith, Nona Sotoodehnia, Silvia Stringhini, Alexander Teumer, Adrienne Tin, Jana V. van Vliet-Ostapchouk, Melanie Waldenberger, Juliane Winkelmann, Bruce H. R. Wolfenbützel, Wei Zhao, Yanan Zheng. Management of an individual contributing study: Adebowale A. Adeyemo, Nasir A. Aziz, Andrea Baccarelli, Murielle Bochud, Hermann Brenner, Monique M. B. Breteler, Loyal Chaker, Josef Coresh, Tanguy Corre, Adolfo Correa, Simon R. Cox, Christian Gieger, Megan L. Grove, Xin Gao, Sarah E. Harris, Gibran Hemani, Peter Henneman, Lifang Hou, Mikko A. Hurme, Sharon L. R. Kardia, Marcus E. Kleber, Wolfgang Koenig, Jaspal S. Kooner, Anna Köttgen, Terho Lehtimäki, Daniel Levy, Lars Lind, Donald M. Lloyd-Jones,

Marie Loh, Riccardo E. Marioni, Karlijn A. C. Meeks, Joyce B. J. van Meurs, Lili Milani, Josine L. Min, Pashupati P. Mishra, Winfried März, Matthias Nauck, Annette Peters, Bruce M. Psaty, Olli T. Raitakari, Alex P. Reiner, Joel Schwartz, Ben Schöttker, Jennifer A. Smith, Harold Snieder, Hannah R. Stocker, Johan Sundström, Jana V. van Vliet-Ostapchouk, Uwe Völker, Melanie Waldenberger, Juliane Winkelmann, Bruce H. R. Wolfenbützel.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

J. Ärnlöv has served on advisory boards for AstraZeneca and Boehringer Ingelheim, and has received lecturing fees from AstraZeneca and Novartis, all unrelated to the present project. J. Coresh has grants from NIH and is a consultant to healthy.io. J. S. Floyd has consulted for Shionogi Inc. C. Herder reports personal fees from Sanofi and Lilly and grant support from Sanofi outside the submitted work. M. E. Kleber is employed with SYNLAB Holding Deutschland GmbH. O. Woodward has grants from AstraZeneca outside the submitted work. W. Koenig reports personal fees from AstraZeneca, Novartis, Pfizer, The Medicines Company, DalCor, Kowa, Amgen, Corvidia, Daichii-Sankyo, Genentech, Novo Nordisk, Omeicos, Esperion, Berlin-Chemie, Sanofi, and Bristol-Myers Squibb and grants and non-financial support from Abbott, Roche Diagnostics, Beckmann, and Singulex outside the submitted work. S. Lorkowski reports grants and personal fees from Akcea Therapeutics Germany, and personal fees from amedes, AMGEN, Berlin-Chemie, Boehringer Ingelheim Pharma, Daichi Sankyo, Lilly Deutschland, MSD Sharp & Dohme, Novo Nordisk Pharma, Roche Pharma, Sanofi-Aventis, Synlab Holding Deutschland, Unilever, and Upfield, all outside the submitted work. R. E. Marioni has received payment from Illumina for presentations. W. März reports grants from Siemens Healthineers, grants and personal fees from Aegerion Pharmaceuticals, grants and personal fees from AMGEN, grants from Astrazeneca, grants and personal fees from Sanofi, grants and personal fees from Alexion Pharmaceuticals, grants and personal fees from BASF, grants and personal fees from Abbott Diagnostics, grants and personal fees from Numares AG, grants and personal fees from Berlin-Chemie, grants and personal fees from Akcea Therapeutics, grants from Bayer Vital GmbH, grants from bestbion dx GmbH, grants from Boehringer Ingelheim Pharma GmbH Co KG, grants from Immundiagnostik GmbH, grants from Merck Chemicals GmbH, grants from MSD Sharp and Dohme GmbH, grants from Novartis Pharma GmbH, grants from Olink Proteomics, other from Synlab Holding Deutschland GmbH, all outside the submitted work. B. Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. J. Sundström reports ownership in companies providing services to Itrium, Amgen, Janssen, Novo Nordisk, Eli Lilly, Boehringer, Bayer, Pfizer, and AstraZeneca, outside the submitted work. B. Kühnel is currently an employee of Regeneron Genetics Center. All other authors have nothing to declare.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27198-4>.

**Correspondence** and requests for materials should be addressed to Adrienne Tin or Anna Köttgen.

**Peer review information** *Nature Communications* thanks William Cookson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Adrienne Tin<sup>1,2,98</sup>, Pascal Schlosser<sup>2,3,98</sup>, Pamela R. Matias-Garcia<sup>4,5,6</sup>, Chris H. L. Thio<sup>7</sup>, Roby Joehanes<sup>8,9</sup>, Hongbo Liu<sup>10</sup>, Zhi Yu<sup>11,12,13</sup>, Antoine Weihs<sup>14</sup>, Anselm Hoppmann<sup>3</sup>, Franziska Grundner-Culemann<sup>3</sup>, Josine L. Min<sup>15,16</sup>, Victoria L. Halperin Kuhns<sup>17</sup>, Adebowale A. Adeyemo<sup>18</sup>, Charles Agyemang<sup>19</sup>, Johan Ärnlöv<sup>20,21</sup>, Nasir A. Aziz<sup>22,23</sup>, Andrea Baccarelli<sup>24</sup>, Murielle Bochud<sup>25</sup>, Hermann Brenner<sup>26,27,28,29</sup>, Jan Bressler<sup>30</sup>, Monique M. B. Breteler<sup>22,31</sup>, Cristian Carmeli<sup>25,32</sup>, Loyal Chaker<sup>33,34</sup>, Josef Coresh<sup>2</sup>, Tanguy Corre<sup>25</sup>, Adolfo Correa<sup>1</sup>, Simon R. Cox<sup>35</sup>, Graciela E. Delgado<sup>36</sup>, Kai-Uwe Eckardt<sup>37,38</sup>, Arif B. Ekici<sup>39</sup>, Karlhans Endlich<sup>40,41</sup>, James S. Floyd<sup>42,43,44</sup>, Eliza Fraszczyk<sup>7</sup>, Xu Gao<sup>24,45</sup>, Xin Gào<sup>26</sup>, Allan C. Gelber<sup>2,46</sup>, Mohsen Ghanbari<sup>33</sup>, Sahar Ghasemi<sup>14,41,47</sup>, Christian Gieger<sup>4,5</sup>, Philip Greenland<sup>48</sup>, Megan L. Grove<sup>30</sup>, Sarah E. Harris<sup>35</sup>, Gibran Hemani<sup>15,16</sup>, Peter Henneman<sup>49</sup>, Christian Herder<sup>50,51,52</sup>, Steve Horvath<sup>53,54</sup>, Lifang Hou<sup>48</sup>, Mikko A. Hurme<sup>55</sup>, Shih-Jen Hwang<sup>8,56</sup>, Sharon L. R. Kardia<sup>57</sup>, Silva Kasela<sup>58</sup>, Marcus E. Kleber<sup>36,59</sup>, Wolfgang Koenig<sup>60,61,62</sup>, Jaspal S. Kooner<sup>63,64,65</sup>, Florian Kronenberg<sup>66</sup>, Brigitte Kühnel<sup>4,5</sup>, Christine Ladd-Acosta<sup>2</sup>, Terho Lehtimäki<sup>67,68,69</sup>, Lars Lind<sup>70</sup>, Dan Liu<sup>22</sup>, Donald M. Lloyd-Jones<sup>48</sup>, Stefan Lorkowski<sup>71,72</sup>, Ake T. Lu<sup>53</sup>, Riccardo E. Marioni<sup>73</sup>, Winfried März<sup>36,72,74,75</sup>, Daniel L. McCartney<sup>73</sup>, Karlijn A. C. Meeks<sup>18,19</sup>, Lili Milani<sup>58</sup>, Pashupati P. Mishra<sup>67,68,69</sup>, Matthias Nauck<sup>41,76</sup>, Christoph Nowak<sup>20</sup>, Annette Peters<sup>5,77</sup>, Holger Prokisch<sup>78,79</sup>, Bruce M. Psaty<sup>42,43,44,80</sup>, Olli T. Raitakari<sup>81,82,83</sup>, Scott M. Ratliff<sup>57</sup>, Alex P. Reiner<sup>43</sup>, Ben Schöttker<sup>26,27</sup>, Joel Schwartz<sup>84</sup>, Sanaz Sedaghat<sup>85</sup>, Jennifer A. Smith<sup>57,86</sup>, Nona Sotoodehnia<sup>44</sup>, Hannah R. Stocker<sup>26,27</sup>, Silvia Stringhini<sup>25</sup>, Johan Sundström<sup>70,87</sup>, Brenton R. Swenson<sup>44,88</sup>, Joyce B. J. van Meurs<sup>34</sup>, Jana V. van Vliet-Ostaptchouk<sup>89</sup>, Andrea Venema<sup>49</sup>, Uwe Völker<sup>41,90</sup>, Juliane Winkelmann<sup>78,91,92,93</sup>, Bruce H. R. Wolffenbuttel<sup>89</sup>, Wei Zhao<sup>57</sup>, Yinan Zheng<sup>48</sup>, The Estonian Biobank Research Team\*, The Genetics of DNA Methylation Consortium\*, Marie Loh<sup>94,95</sup>, Harold Snieder<sup>7</sup>, Melanie Waldenberger<sup>4,5,61</sup>, Daniel Levy<sup>8,9</sup>, Shreeram Akilesh<sup>96</sup>, Owen M. Woodward<sup>17</sup>, Katalin Susztak<sup>10</sup>, Alexander Teumer<sup>41,47,97,99</sup> & Anna Köttgen<sup>2,3,99</sup>

<sup>1</sup>Department of Medicine, University of Mississippi Medical Center, Jackson 39216 MS, USA. <sup>2</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. <sup>3</sup>Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany. <sup>4</sup>Research Unit Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg D-85764 Bavaria, Germany. <sup>5</sup>Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg D-85764 Bavaria, Germany. <sup>6</sup>TUM School of Medicine, Technical University of Munich, Munich, Germany. <sup>7</sup>Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. <sup>8</sup>Framingham Heart Study, Framingham, MA, USA. <sup>9</sup>Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. <sup>10</sup>Department of Medicine and Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia 19104 PA, USA. <sup>11</sup>Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. <sup>12</sup>Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. <sup>13</sup>Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. <sup>14</sup>Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. <sup>15</sup>MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. <sup>16</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. <sup>17</sup>Department of Physiology, University of Maryland School of Medicine, Baltimore, MD, USA. <sup>18</sup>Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. <sup>19</sup>Department of Public and Occupational Health, Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands. <sup>20</sup>Department of Neurobiology, Care Sciences and Society (NVS), Family Medicine and Primary Care Unit, Karolinska Institutet, Huddinge, Sweden. <sup>21</sup>School of Health and Social Studies, Dalarna University, Falun, Sweden. <sup>22</sup>Population Health Sciences, German Centre for Neurodegenerative Diseases (DZNE), Bonn, Germany. <sup>23</sup>Department of Neurology, Faculty of Medicine, University of Bonn, Bonn, Germany. <sup>24</sup>Laboratory of Environmental Precision Health, Mailman School of Public Health, Columbia University, New York, NY, USA. <sup>25</sup>Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland. <sup>26</sup>German Cancer Research Center (DKFZ), Division of Clinical Epidemiology and Aging Research, Heidelberg, Germany. <sup>27</sup>Network Aging Research, Heidelberg University, Heidelberg, Germany. <sup>28</sup>Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. <sup>29</sup>German Cancer Consortium, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>30</sup>Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston 77030 TX, USA. <sup>31</sup>Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Bonn, Germany. <sup>32</sup>Population Health Laboratory, University of Fribourg, Fribourg, Switzerland. <sup>33</sup>Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. <sup>34</sup>Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. <sup>35</sup>Lothian Birth Cohorts Group, Department of Psychology, The University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. <sup>36</sup>Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. <sup>37</sup>Department of Nephrology and Hypertension, University of Erlangen-Nürnberg, Erlangen, Germany. <sup>38</sup>Department of Nephrology and Medical Intensive Care, Charité – Universitätsmedizin Berlin, Berlin, Germany. <sup>39</sup>Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany. <sup>40</sup>Department of Anatomy and Cell Biology, University Medicine Greifswald,



Greifswald, Germany. <sup>41</sup>DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany. <sup>42</sup>Department of Medicine, University of Washington, Seattle 98101 WA, USA. <sup>43</sup>Department of Epidemiology, University of Washington, Seattle 98101 WA, USA. <sup>44</sup>Cardiovascular Health Research Unit, University of Washington, Seattle 98101 WA, USA. <sup>45</sup>Department of Occupational and Environmental Health Sciences, School of Public Health, Peking University, Beijing, China. <sup>46</sup>Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA. <sup>47</sup>Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. <sup>48</sup>Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>49</sup>Department of Clinical Genetics, Amsterdam Reproduction & Development Research Institute, Amsterdam University Medical Centres, University of Amsterdam, Amsterdam, the Netherlands. <sup>50</sup>Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>51</sup>German Center for Diabetes Research, Munich-Neuherberg, Germany. <sup>52</sup>Division of Endocrinology and Diabetology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>53</sup>Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles 90095 CA, USA. <sup>54</sup>Biostatistics, Fielding School of Public Health, UCLA, Los Angeles, CA, USA. <sup>55</sup>Department of Microbiology and Immunology, Faculty of Medicine and Health Technology, Tampere University, Tampere 33014, Finland. <sup>56</sup>Division of Intramural Research, Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. <sup>57</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor 48109 MI, USA. <sup>58</sup>Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. <sup>59</sup>SYNLAB MVZ Humangenetik Mannheim, Mannheim, Germany. <sup>60</sup>Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. <sup>61</sup>DZHK (German Centre for Cardiovascular Research), Partner site Munich Heart Alliance, Munich, Germany. <sup>62</sup>Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm, Germany. <sup>63</sup>National Heart and Lung Institute, Imperial College London, London, UK. <sup>64</sup>Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Southall, UK. <sup>65</sup>Imperial College Healthcare NHS Trust, London, UK. <sup>66</sup>Institute of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria. <sup>67</sup>Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. <sup>68</sup>Finnish Cardiovascular Research Centre, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. <sup>69</sup>Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. <sup>70</sup>Department of Medical Sciences, Uppsala University, Uppsala, Sweden. <sup>71</sup>Institute of Nutritional Sciences, Friedrich Schiller University Jena, Jena, Germany. <sup>72</sup>Competence Cluster for Nutrition and Cardiovascular Health (nutriCARD) Halle-Jena-Leipzig, Jena, Germany. <sup>73</sup>Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK. <sup>74</sup>Synlab Academy, SYNLAB Holding Deutschland GmbH, Mannheim and Augsburg, Germany. <sup>75</sup>Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. <sup>76</sup>Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. <sup>77</sup>Ludwig-Maximilians Universität München, Munich, Germany. <sup>78</sup>Institute of Human Genetics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany. <sup>79</sup>Department of Computational Health, Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany. <sup>80</sup>Department of Health Services, University of Washington, Seattle 98101 WA, USA. <sup>81</sup>Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. <sup>82</sup>Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. <sup>83</sup>Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland. <sup>84</sup>Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>85</sup>Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. <sup>86</sup>Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. <sup>87</sup>The George Institute for Global Health, University of New South Wales, Sydney, NSW, Australia. <sup>88</sup>Institute for Public Health Genetics, University of Washington, Seattle, WA, USA. <sup>89</sup>Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. <sup>90</sup>Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. <sup>91</sup>Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany. <sup>92</sup>Chair Neurogenetics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany. <sup>93</sup>Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. <sup>94</sup>Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. <sup>95</sup>Department of Epidemiology and Biostatistics, Imperial College London, London, UK. <sup>96</sup>Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. <sup>97</sup>Department of Population Medicine and Lifestyle Diseases Prevention, Medical University of Białystok, Białystok, Poland. <sup>98</sup>These authors contributed equally: Adrienne Tin, Pascal Schlosser. <sup>99</sup>These authors jointly supervised this work: Alexander Teumer, Anna Köttgen. \*Lists of authors and their affiliations appear at the end of the paper. ✉email: [atin@umc.edu](mailto:atin@umc.edu); [anna.koettgen@uniklinik-freiburg.de](mailto:anna.koettgen@uniklinik-freiburg.de)

## The Estonian Biobank Research Team

Lili Milani <sup>58</sup>

A full list of members and their affiliations appears in the Supplementary Information.

## The Genetics of DNA Methylation Consortium

Josine L. Min <sup>15,16</sup> & Gibran Hemani <sup>15,16</sup>