Environmental Epigenetics, 2025, **11(1)**, 1–10
DOI: https://doi.org/10.1093/eep/dvaf001
Advance Access Publication 17 January 2025
Research Article

Longitudinal leucocyte DNA methylation changes in Mesoamerican nephropathy

Amin Oomatia[1,*], Olga Chervova[2], Ali M. Al-Rashed[1], Evangelia-Theano Smpokou[1], Simone Ecker[2], Neil Pearce[3], Brianna Heggeseth[4], Dorothea Nitsch[3], Andres Cardenas[5], Stephan Beck[2], Marvin Gonzalez-Quiroz[1,6], Ben Caplin[1]

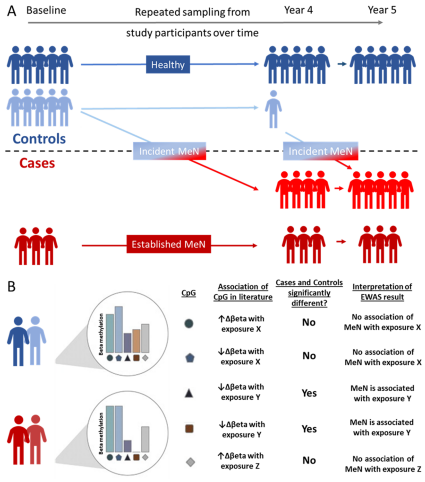
[1]Centre for Kidney and Bladder Health, University College London, London NW3 2PF, United Kingdom
[2]UCL Cancer Institute, University College London, London WC1E 6DD, United Kingdom
[3]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London WC1E 7HT, United Kingdom
[4]Department of Data Sciences, Macalester College, St. Paul, MN 55105-1899, United States
[5]Department of Epidemiology and Population Health, Stanford School of Medicine, Stanford, CA 94305-5405, United States
[6]Department of Environmental and Occupational Health, UT School of Public Health San Antonio, The University of Texas Health Science Centre at San Antonio, San Antonio, TX 78249, United States

*Corresponding author. Centre for Kidney and Bladder Health, University College London Medical School, Rowland Hill Street, London NW3 2PF, United Kingdom. E-mail: amin.oomatia@ucl.ac.uk

Abstract

Mesoamerican nephropathy (MeN) is a leading cause of morbidity and mortality in Central America, yet its aetiology remains unclear. Environmental exposures including heat stress, pesticides, and heavy metals have all been suggested as possible causes or exacerbating factors of the disease, but intermittent and cumulative exposures are difficult to capture using conventional biomonitoring. Locus-specific differential DNA-methylation (DNAm) which is known to occur in association with these environmental exposures can be readily measured in peripheral blood leucocytes, and therefore have the potential to be used as biomarkers of these exposures. In this study, we aimed first to perform a hypothesis-free epigenome-wide association study of MeN to identify disease-specific methylation signatures, and second to explore the association of DNAm changes associated with potentially relevant environmental exposures and MeN onset. Whole-blood epigenome-wide DNAm was analysed from a total of 312 blood samples: 53 incident cases (pre- and post-evidence of disease onset), 61 matched controls and 16 established cases, collected over a 5-year period. Mixed-effect models identified three unique differentially methylated regions that associated with incident kidney injury, two of which lie within the intron of genes (*Amphiphysin* on chromosome 7, and *SLC29A3* chromosome 10), none of which have been previously reported with any other kidney disease. Next, we conducted a hypothesis-driven analysis examining the coefficients of CpG sites reported to be associated with ambient temperature, pesticides, arsenic, cadmium, and chromium. However, none showed an association with MeN disease onset. Therefore, we did not observe previously reported patterns of DNA methylation that might support a role of pesticides, temperature, or the examined metals in causing MeN.

Graphical abstract



Received 4 August 2024; Revised 9 December 2024; Accepted 10 January 2025

Background

Mesoamerican nephropathy (MeN), first reported in 2002, is highly prevalent and a leading cause of death among working-age men in several Central American countries [1, 2]. Affected individuals, who often present with end-stage kidney disease, are usually young male agricultural workers who have no traditional risk factors for kidney disease, such as hypertension or diabetes [3–6]. Given that MeN is usually seen in rural communities and is associated with agricultural work, environmental exposures such as heat stress, heavy metals, and pesticides have all been hypothesized as potentially relevant aetiological agents [5–9]. However, there remains substantial debate surrounding the likely primary cause [10, 11].

Previous work by our group suggests that levels of urinary metals/metalloids and pesticides do not differ at baseline between those who go on to experience renal decline and those who maintain normal renal function over follow-up [12]. However, biosample measurements in our study, as in other studies, have been cross-sectional and therefore may not accurately capture cumulative exposures. Relying on self-reporting is an alternative strategy for certain exposures such as heat or agrochemicals, but this is prone to bias and misclassification [13, 14].

Changes in DNA methylation (DNAm) in whole blood can be cumulative and persistent [15]. Patterns of DNAm can give important insights into the mechanistic pathways underlying disease [16–18]. Furthermore, DNAm can occur at specific loci in response to environmental exposures with existing data demonstrating differential DNAm associated with many of the exposures potentially relevant in MeN, including pesticides, metals, and temperature [19–28]. It may therefore be possible to explore associations between potentially causative environmental factors such as pesticides, ambient temperature, and heavy metals, and MeN using locus-specific DNAm as a biomarker of exposure, while reducing the misclassification that occurs with self-report or intermittent biosampling.

This study aimed to: (i) determine whether MeN is associated with a distinct DNAm signature in leucocytes; and (ii) explore associations between MeN and the differential DNAm associated with potentially relevant various environmental exposures, including pesticides, metals/metalloids, and heat stress. We therefore performed an epigenome-wide association study (EWAS) using mixed-effect modelling on whole blood DNA collected from individuals at multiple time points from three groups: (i) individuals suspected of having incident MeN; (ii) matched healthy individuals from the same source population; (iii) individuals with established kidney disease presumed to be due to MeN. Using the data from these models, we then examined whether individual CpGs known to differentially methylate in association with pesticides, ambient temperature, and several heavy metals were also differentially methylated between cases and controls, which would suggest these exposures were associated with MeN. A visual representation of the study design can be found in Fig. 1.

Results

Results from pipeline pre-processing, quality control and blood cell type fractions

After all preprocessing and clean-up steps, our dataset consisted of 312 samples and 832 813 CpGs per sample. A detailed breakdown of samples and CpGs failing quality control can be found in Supplementary Fig. S1. There was no significant difference in calculated cell type fractions between cases and controls (see Supplementary Fig. S2).

Final study population

After removal of samples failing quality control, the final study population consisted of 127 samples from 53 individuals with incident MeN, matched for age, sex, and smoking status to 149 samples from 61 healthy controls. Additionally, 36 samples from 16 participants with established MeN, which were unmatched, were also included. In total, across the three groups, there were 113 samples at baseline, 104 samples from year four and 93 samples from year five. The distributions of samples by participant case status, year of collection, and classification in EWAS models are shown in Table 1. Key demographics of the participants are shown in Table 2.

Results from mixed-effect models identify no significant DMPs but demonstrate four differentially methylated regions associated with MeN

Model A included 203 healthy samples (149 healthy controls and 54 incident cases pretransition to disease) vs 73 diseased samples (all incident cases post-transition to disease). Model B included the same 203 healthy samples vs 109 diseased samples (73 samples from incident cases post-transition to disease and 36 samples from established cases). QQ plots suggested a good fit for both models ($\lambda = 0.96$, $\lambda = 0.98$) with no inflation. Neither model identified any statistically significant DMPs [false discovery rate (FDR) <0.05, see Fig. 2 for a visual representation]. A list of the top 20 CpGs from each model is given in Supplementary Tables S3 and S4. The *P*-values from the 832 813 DMPs from the mixed-effect models were then analysed using *ipDMR* to identify any genomic regions associated with MeN. Between models A and B, four differentially methylated regions (DMRs) were identified, with the DMRs on chromosomes 5 and 10 being identified in both models. Details of all four DMRs can be found in Table 3.

Hypothesis-driven analysis of exposure-associated CpGs returned no significant association with either incident or established MeN

Our literature review identified 10 relevant and suitable studies from which a list of exposure-associated CpGs could be extracted. These were: heat (one), arsenic (four), cadmium (one), chromium (one), and pesticides (three), from which a list of exposure-associated DMPs were generated [19–28]. Details of the exposure metrics, number of samples, and numbers of reported DMPs from each study can be found in Supplementary Table S5. *ΔBeta* and *P*-values for these CpGs were then extracted from the methylation profiles from both MeN EWAS models and collated for each exposure. For all the examined exposures, the number of the considered DMPs associated with MeN EWAS case status in models was less than would be expected by chance alone (Table 4). Complete lists of both the reported and study *ΔBeta* and *P*-values for each of the exposure-associated DMPs can be found in Supplementary Data Files 1 and 2.

Power calculations suggested adequate power in the MeN EWAS methylation profiles to detect the reported exposure-associated methylation signatures after restricting the data to the CpGs reported in the searched literature. The majority of CpGs associated with arsenic, chromium, and pesticides in both EWAS models had 80% or greater power to detect the reported *ΔBeta*. The majority of CpGs associated with increased ambient temperature had 80% or more power to detect a *ΔBeta* associated with a 2°C

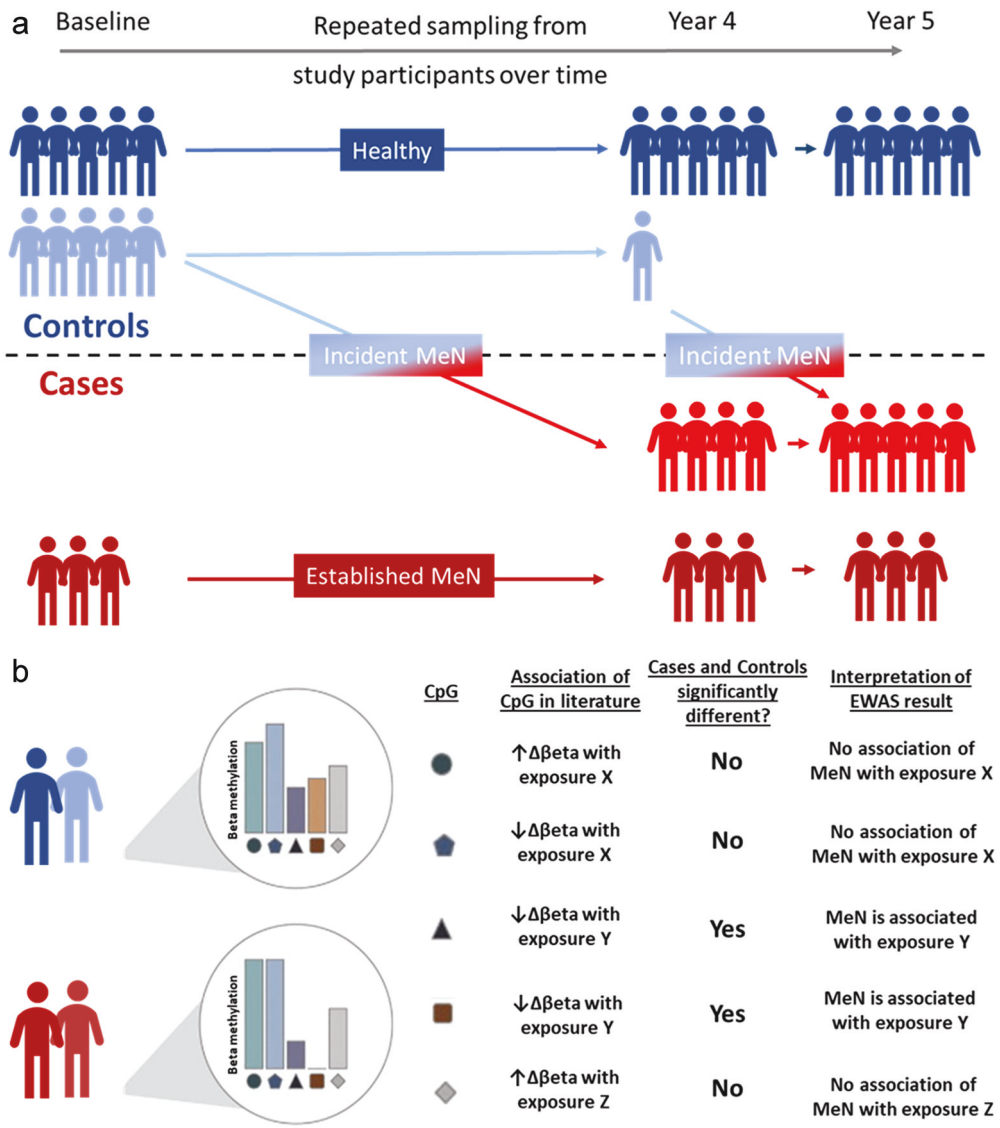


Figure 1. Visual representation of the study design to identify the longitudinal DNAm changes associated with MeN (a), and how this data was used to look for evidence of past environmental exposures that may associate with the disease. (a) Schematic of the MeN EWAS study population. By using repeat samples of whole blood collected at baseline, year four and year five from the same individuals, we performed mixed-effect modelling to compare the DNAm changes that occurred between cases and controls, as well as those that occurred within participants as they transitioned from a healthy to diseased state over time. Samples were derived from three groups: (i) healthy individuals (dark blue)—those who remained healthy throughout study follow-up, (ii) Incident MeN cases—those who were healthy at baseline (light blue) but subsequently experienced a sustained loss of kidney function (light red), and (iii) established cases (dark red)—those who had evidence of kidney disease at baseline and throughout follow-up. Samples were classified as 'cases' or 'control' depending on the kidney health of the individual at the time of sampling. We created two study models. Model A consisted of 203 control samples (149 healthy controls and 54 pre-transition incident cases) vs 73 case samples (all from post-transition incident cases). Model B consisted of the same 203 healthy samples vs 109 case samples (73 post-transition cases and 36 established cases). (b)—Visual representation of how CpGs known to differentially methylate with various environmental exposures were examined in the methylation profiles generated from the EWAS. A list of candidate CpGs for each exposure was collated from a comprehensive literature review. Figure B modified from Paul and Beck (2014) [66] under the Creative Commons BY 3.0 license.

Table 1. Breakdown of samples by participant case status, year of collection, and classification in EWAS model Incident MeN cases were either pre or post-transition from a healthy state to a diseased state

	No. samples at each timepoint				EWAS classification	
Individual's case status	**Baseline**	**Year 4**	**Year 5**	**Total**	**Model A**	**Model B**
Healthy individual	54	51	44	149		
Incident MeN (pretransition)	46	8	0	54	Control ($n = 203$)	Control ($n = 203$)
Incident MeN (post-transition)	0	35	38	73	Case ($n = 73$)	
Established MeN	13	12	11	36	-	Case ($n = 109$)

Table 2. Demographics of study participants Healthy participants were selected to match the age, sex and smoking status of participants with incident MeN. There were no females with established kidney disease in the Colt cohort

	Incident case ($n = 53$)	Healthy ($n = 61$)	Established case ($n = 16$)
Age at recruitment Mean ± SD (years)	24.5 ± 3.4	24.9 ± 4.2	24.7 ± 3
Males	48 (91%)	50 (82%)	16 (100%)
Smokers	25 (47%)	27 (44%)	7 (43%)
eGFR at recruitment, mean ± SD (ml/min/1.73 m^2)	115.8 ± 9.9	122.7 ± 9.2	67.5 ± 15.8
No. samples	127	149	36
Baseline (year 0)	45	53	13
Baseline (year 0.5)	1[a]	1	0
Year 4	43[Δ]	51	12
Year 5	38	44	11

[a]One of the participants who transitioned was recruited at year 0.5, rather than year 0. Therefore, their matched control was also an individual recruited at year 0.5 [34]. [Δ]Of the 43 incident case samples at year 4, 35 were from participants who had transitioned to a diseased state. All samples at year 5 from incident cases reflected a diseased state. NB: The number of both incident cases and healthy individuals was initially 57. However, after sample processing, the HMM was revised resulting in four participants initially classified as incident cases being reclassified as healthy individuals. Therefore, the final number of incident cases included in the EWAS study population was 53, and healthy individuals was 61.

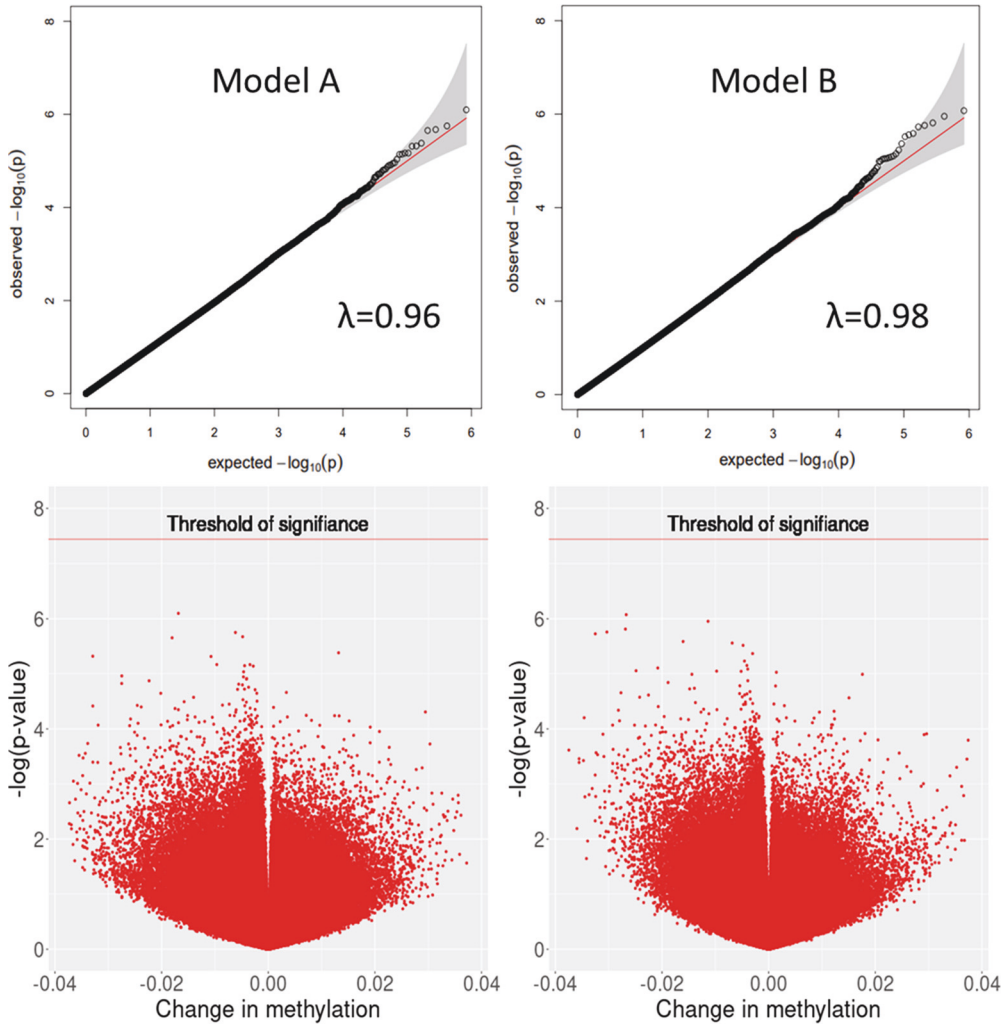


Figure 2. QQ plot and Volcano plots of both mixed-effect models. QQ plots demonstrate inflation λ values close to 1, which suggests appropriate fit and minimal inflation. No CpGs reached a threshold of statistical significance adjusted for multiple testing ($q < 0.05$, $P < 3.6 \times 10^{-7}$). Model A consisted of 73 cases vs 203 controls. Model B consisted of 109 cases vs 203 controls.

temperature difference, but not a 1°C difference. No power calculation could be performed for organophosphates or cadmium because reference *ΔBeta* values were not reported. A detailed breakdown of the number of DMPs with >80% power for detection for each study can be found in Supplementary Table S6. A complete list of the power to detect each reported exposure-associated DMP in each EWAS model can be found in Supplementary Data File 3.

Table 3. ipDMR analysis of models A and B

Chromosome	DMR start	DMR end	DMR FDR Model A	Mean % Δβ Model A	DMR FDR Model B	Mean % Δβ Model B	No. probes	Locus	Nearest genes
5[a]	2350- 7134	2350- 7350	3.21×10^{-7}	1.63	1.51×10^{-7}	1.61%	2	Promotor: EH38E2361519	*PRDM9, CDH12, CDH10*
7	3846- 4820	3846- 5502	2.22×10^{-7}	−0.39	-	-	2	Intron	*Amphiphysin*
10	7132- 3366	7132- 3467	4.72×10^{-9}	−1.14	1.51×10^{-7}	−1.26%	3/2[b]	Intron	*SLC29A3*
14	1012- 25 832	1012- 26 243	-	-	1.51×10^{-7}	−2.04%	2	Enhancer: EH38E1743083	*DIO3, RTL1, DLK1*

[a]One of the two CpGs in the DMR on chromosome 5: cg 25 336 267, was present in the list of probes with potential for multimapping on the EPIC BeadChip as identified by McCartney *et al.* (2016) [65].
[b]The DMR on chromosome 10 in model A was significant at three probes, but only two probes in Model B. DMR start and end positions pertain to genome build from GRCh38/hg38.

Table 4. List of exposure-associated DMPs and their association with CpG from both models of the MeN EWAS

Exposure	Study	Exposure-associated CpGs that also associated with MeN	
		Model A	Model B
Temperature (over 365 days)[a]	Xu *et al* (2020) [26]	0/18	0/18
Metals: Arsenic	Argos *et al.*(2015) [19]	1/3	1/3
Metals: Arsenic	Ameer *et al.*(2017) [28]	0/21	0/21
Metals: Arsenic	Demanelis *et al.*(2019) [21]	2/33	0/33
Metals: Arsenic	Bozack *et al.* (2020) [20]	0/20	2/20
Metals: Cadmium	Domingo-Relloso *et al.* (2020) [22]	1/3	0/3
Metals: Chromium	Feng *et al.*(2020) [23]	1/19	1/19
Pesticides: Organophosphate (OP)	Paul *et al.* (2018) [25]	0/4	0/4
Pesticides: Pyrethroid	Furlong *et al.* (2020) [24]	0/46	0/46
Pesticides:	Hoang *et al.* (2021) [27]	5/154	7/154
Acetochlor		0/20	1/20
Atrazine		0/1	0/1
Dicamba		0/6	1/6
Glyphosate		0/1	0/1
Heptachlor		0/6	0/6
Malathion		0/1	0/1
Mesotrione		5/67	4/67
Metolachlor		0/6	0/6
Picloram		0/46	1/46

For a CpG to be deemed as associated it had to be both congruent in the direction of methylation change and have a raw *P*-value of <.05.
[a]Screening of DMPs associated with shorter temperature windows of preceding 28, 90, and 180 days also demonstrated no association with CpGs from either MeN EWAS model.

Proof of concept: smoking-associated CpGs

To validate the approach of screening for exposure association in the EWAS cohort, a further mixed-effect model examining DNA methylation changes associated with self-reported smoking was performed using the array data from the same 312 whole blood samples. This data were then screened for the smoking-associated CpGs extracted from a Central American cohort, Cardenas *et al.* (2022) [29]. In total, there were 143 samples from smokers vs 169 samples from nonsmokers in the smoking EWAS. Key demographic information, and sample distribution by year of smokers vs nonsmokers, can be found in Supplementary Table S7. The calculated cell-type fractions between smokers and nonsmokers differed for CD4+ T-cells, neutrophils and eosinophils, but were adjusted for using the top three principal components of cell-type fraction (see Supplementary Fig. S8). In total, 23 out of 45 smoking-associated DMPs by Cardenas *et al.* (2022) were associated with smoking status in our cohort with the same direction of effect and $P < .05$. A further 22 DMPs had the same direction of effect but did not reach statistical significance. A list of the smoking-associated DMPs, with the corresponding *ΔBeta and P*-values from both the reference study and the smoking EWAS, is listed in Supplementary Table S9.

Discussion

To our knowledge, this is the first EWAS study of Mesoamerican nephropathy to be conducted. We identified three regions that differentially methylate early in the development of MeN, and a fourth region that associates when established cases are included. Additionally, we found that DMPs reported to be associated with ambient temperature, several pesticides, cadmium, arsenic, and chromium were not associated with either incident or established MeN cases in our study population. These findings therefore do not provide evidence supporting a role of increased ambient temperature, or the metals and pesticides examined, in the initiation of MeN.

The biological significance of the DMRs we have identified remains unclear and has not been reported in other EWAS of chronic kidney disease (CKD), suggesting they are unique to MeN [30–32]. The caveat here is that participants in our study are earlier in the progression of their kidney disease than in most CKD studies. The DMR identified on chromosome 14, detected only in model B, is noteworthy as it may regulate the nearby *RTL1*, *DLK1*, and *DIO3* genes. The locus encompassing these genes has previously been identified in a GWAS of estimated glomerular filtration rate (eGFR) decline among Hispanic participants [33]. The fact that this DMR was only identified in model B, which included individuals with established kidney disease—and thus worse renal function, further suggests that it may represent a genuine difference in DNA methylation. It is not, however, possible to infer functional consequences of DNA methylation at this DMR from our study, nor is it possible to determine if the methylation changes in leucocytes are correlated in renal tissue.

Some studies investigating the cause of Mesoamerican nephropathy have implicated arsenic, chromium, cadmium, pesticides, and temperature, from indirect measurements, such as self-reported heat and pesticide exposure, or measurements of metals from water sources [5–9, 34]. However, we did not find evidence of association of loci with differential DNA methylation previously associated with these exposures with the development of MeN in either of our mixed-effect models. DNA methylation

changes represent an additional means of measuring an individual's past exposures [35], and our estimates suggest our analyses had at least 80% power to detect the majority of the reported effect sizes of exposure-associated DMPs, should they have been present.

Specifically, the MeN EWAS cohort was powered to detect *ΔBetas* associated with a 2°C temperature difference between cases and controls, which is comparable to the ambient temperature difference experienced by agricultural workers at the highest risk of MeN [36, 37]. For arsenic, the EWAS cohort was powered to detect a *ΔBeta* associated with a change of one μg/L of serum, or one μg/g of urine (corrected for creatinine), which is quantitatively smaller than the levels associated with renal dysfunction reported in numerous studies [38–40]. For cadmium, the EWAS was powered to detect a *ΔBeta* associated with one μg/g of urine (corrected for creatinine), which is the level at which the odds ratio of developing tubular proteinuria doubled in one study looking at early kidney disease and low-dose cadmium exposure [41]. Pesticide exposure was often determined through either self-reported exposure or geospatial cumulative exposure scoring in the reference studies we used, which means it is not possible to quantify the concentration of each pesticide that the reported *ΔBetas* may be expected to associate with. However, given the association of MeN with only five (model A) and seven (model B) out of a total of 205 pesticide-associated DMPs examined (Table 4), numbers that would be expected to occur by chance, this does not provide evidence supporting a role of pesticides in the initiation of the disease. Our findings are consistent with previous results published by our group exploring associations between kidney decline and measured metals and pesticides in the baseline bio-samples of study participants [12].

A major strength of our study is that we have serial longitudinal whole-blood DNA methylation measures from study participants. This has meant that we have been able to take into account person-to-person variation in CpG methylation, and focus on the changes associated with the earliest signs of disease onset within individuals. Furthermore, the case definition of incident and established cases are based on multiple repeat estimates of kidney function, and therefore are likely to be less prone to misclassification than if derived from single measures.

Our work also has some limitations which are important to consider. Despite the use of repeat eGFR measures and Hidden-Markov model (HMM) to empirically derive case definitions of MeN, there is a possibility of residual phenotype misclassification. A key assumption that underpins the HMM in determining the case status of an individual is that all instances of early sustained eGFR loss in the study population represent early MeN. However, this assumption is not unreasonable given that the study participants are otherwise young and healthy, have no risk factors for kidney disease such as hypertension or diabetes, and are not found to have new medical diagnoses during follow-up, but are known to be at high risk of MeN. Furthermore, as yet, it has not been conclusively shown that individuals with 'incident MeN' necessarily progress to get established chronic kidney disease of unknown cause (CKDu) (an eGFR <60 ml/min/1.73 m^2) [42]. However, given that there are many 'exacerbating' factors which can drive the progression of kidney disease regardless of the cause of the initial injury, such as high blood pressure, smoking, or high salt intake [43, 44], studying the DNAm changes that occur with early disease is more informative in trying to understand the epigenetic changes associated with MeN itself.

Another limitation of this study, given that only 312 samples from 130 individuals were included, is that we may have had limited power to detect epigenome-wide signals beyond statistical thresholds which account for multiple testing. Furthermore, we acknowledge that the studies of methylation changes associated with ambient temperature, metals, and pesticides are likely to be underpowered to detect more modest differential methylation (see Supplementary Table S5 for number of participants in each study) but also conducted in populations genetically and geographically distinct from our cohort. It has been well described that underlying genetic variation can modify differential methylation at individual loci [45, 46]. Therefore, the absence of the exposure-associated methylation signatures seen in the MeN EWAS data may be due, at least in part, to differences between the reference and study populations.

While we validated our approach of screening for known exposure-associated DMPs using data from mixed-effect models by performing a smoking EWAS in our study cohort, several factors likely contributed to our ability to detect smoking-associated methylation changes in our study data. First, the study by Cardenas *et al.* (2022) was conducted in a Costa Rican population, which is likely genetically similar to our study cohort based in Nicaragua, therefore minimizing the effect of genetic variation in diminishing certain individual loci signal strength [29, 46]. Second, the effect of smoking on DNA methylation is quite pronounced in our EWAS, the mean effect of smoking was a difference of ∼3.5% at each of the smoking associated CpG (see supplementary table S9) making these changes easier to detect. Lastly, smoking-related cell type proportion changes are known to accentuate some of the specific smoking-associated CpG changes [47]. Therefore, despite adjusting for calculated white-cell fractions in our analysis, residual smoking-related differences in proportions of cell types may account for some of the observed smoking-associated differential CpG methylation.

It is also important to consider the study designs of the reference studies themselves. The studies that examined methylation changes associated with pesticide exposure inferred exposure based on geospatial proximity, or participant questioning of exposure [24, 25, 27]. The studies that examined arsenic, cadmium, and chromium used biosample exposure measurements, but at a single time point [19–22, 28]. It is therefore possible that by using these exposure metrics, methylation changes that would be associated with these exposures have not been accurately identified. Similarly, the mean ambient temperature range examined by Xu *et al.*, 11.3–25.1°C, may not reflect the intermittent high environmental temperatures experienced by our study population which may be associated with different patterns of DNA methylation change [26, 48]. Furthermore, ambient temperature does not necessarily correlate with an individual's experience of heat stress, an exposure metric that remains extremely difficult to objectively quantify.

Lastly, although estimates from our calculations suggest that we were 80% powered to detect a difference had it been there, we did not have the power to 'exclude' a difference, as this would require a substantially larger sample size. Therefore, our findings do not support any of the exposures examined but cannot altogether exclude them either.

In conclusion, our study has provided the first available data on DNA methylation changes in whole blood associated with Mesoamerican nephropathy. The four DMRs we have identified warrant replication and further investigation in future studies. We have also utilized the DNA methylation profiles of a population at risk of MeN to look for evidence of past exposures to pesticides, arsenic, cadmium, chromium, and heat. This analysis demonstrated that DNA methylation changes associated with these

exposures did not associate with either incident or established cases, which therefore does not support a role for the involvement of these factors as the first trigger of onset of MeN. Further analytical studies using other methodologies and exploring a full range of possible aetiological factors are therefore needed to identify the cause of the disease. Currently, there are no suitable reference studies for some of the other environmental agents potentially relevant to MeN, such as personal heat stress, silicon nanoparticles, lead, or nickel [49–51]. If and when such studies are published in the future, the methylation profiles we have generated in this study can be re-examined.

Methods

Study population

The Colt cohort is a community-based longitudinal study that began in 2014 and is based in nine rural communities in Nicaragua, funded by the Colt Foundation. The rationale and description of the study design, along with follow-up data, have been previously published [34, 52, 53]. Briefly, the individuals described in this report were selected from a cohort of 350 population-representative participants ages 18–30 years old using a nested case-control approach. All potential participants with a self-reported diagnosis of kidney disease, diabetes, or hypertension at baseline were excluded. Participants were followed up every 6 months for the first 2 years, and annually thereafter for a total of 7 years with clinical measurements, blood and urine sampling, and questionnaire data collected at each visit. Whole blood DNA was available at baseline, and years four and five of follow-up.

Outcome assessment of kidney health

For each individual in the cohort, the eGFR, a measure of kidney function, was calculated for each study visit using serum creatinine [52, 54]. MeN has no diagnostic tests, and remains a disease of exclusion, but is one of the diseases classified as CKDu. Current international guidelines classify individuals with established renal dysfunction (an eGFR< 60 ml/min/1.73 m^2), who reside in areas at high risk of the disease, and in whom other known causes of kidney disease have been excluded as cases of CKDu [42]. However, these criteria exclude individuals with early disease (abnormal kidney function but an eGFR > 60 ml/min/1.73 m^2). Given the range of potential kidney disease exacerbating factors that individuals are exposed to in the at-risk areas (e.g. nonsteroidal anti-inflammatory drug use or episodes of dehydration), uncovering the aetiology of MeN likely requires observation of early signs of kidney damage. We have therefore developed a case definition for incident disease using a HMM approach described elsewhere [53]. This model generates an estimated probability of each individual having a healthy or diseased kidney state at each study visit conditional on the eGFR measurements. We define transition to a disease state at the first study visit when the estimated probability of a healthy state is less than 0.5. A visual representation of example eGFR trajectories of participants in each of these classifications can be found in Supplementary Fig. S10.

The outcome definitions for the purposes of this study are: (i) Incident MeN case—individuals who were in a healthy kidney state at recruitment but transitioned to a diseased state by year five, (ii) Healthy participants—those who remained in a healthy kidney state at all study visits, and (iii) Established MeN case—those who started and remained in a diseased state from baseline through to year five. Incident cases transitioned from healthy to diseased at different study visits. A breakdown of the number of participants transitioning at each study visit can be found in Supplementary Table S11.

EWAS study design, and participant and sample selection

Using whole blood collected from study participants at baseline, years four and five, we sought to determine 'longitudinal' changes in DNA methylation that occur both between cases and controls, as well as those that occur within participants as they develop MeN (Fig. 1a). We therefore performed an EWAS on serial samples using a nested case-control design. Specifically, *all* individuals with incident MeN were matched to healthy participants by age, sex, and smoking status. Where available, at least two samples were taken from each incident case and their respect matched healthy control: one at baseline and one at the end of follow-up, either year four or year five. A selection of individuals, randomly chosen, had samples at all three time points. Additionally, samples from 16 individuals classified as having established MeN were also selected for analysis. The criteria for selecting these individuals are discussed further in Supplementary Material S12.

We created two mixed-effect models to study DNA methylation changes associated with MeN. Model A compared samples taken from incident cases and controls, with the sample being classified as a case or control based on the state of the individual at the time the blood sample was drawn. Model B included samples from established cases alongside incident cases, as well as healthy individuals, with the sample again being classified as healthy or diseased based on the state of the individual 'at the time the blood sample was drawn'.

DNA extraction and quantification of DNA methylation using the Illumina MethylationEPIC array

Genomic DNA was extracted from a total of 320 whole blood EDTA samples from 130 participants (mean 2.6 samples per individual) using a QIAmp DNA Blood Midi kit (Qiagen, Cat. #51185). Genomic DNA from each sample (500ng) first underwent bisulfite conversion and were then profiled using the Infinium MethylationEPIC BeadChip™ at the Zayed Centre for Research into Rare Diseases, University College London [55]. To avoid potential batch effects on downstream analysis, samples were systematically rotated to ensure that cases and controls, as well as those from different years, were spread randomly but evenly across all plates (see Supplementary Fig. S13 for a visual representation and more detailed description).

Preprocessing and clean-up of methylation data

Array data files were preprocessed and underwent stringent quality control measures as per established pipelines by our group, using the *ENmix* R package (V1.30.3) [56–59]. A description and flow diagram of the specific steps performed can be found in Supplementary Fig. S1.

Mixed-effect modelling

We fitted mixed-effect models using the R *lme4* library (https://github.com/lme4) [59]. To adjust for varying cell composition in test samples, cell type fractions were empirically derived using the R package *Epidish* [60]. Principal components were then calculated and the top three, which accounted for nearly 80% of the variation in cell-type fraction, were used as fixed-effects to adjust for cell-type variation.

In both mixed-effect models, for each CpG we used methylation proportion (*Beta* values) as the dependent variable. Age of individuals at sampling, sex, smoking status, and the top three cell-type principal components were included as fixed effects. Sample slide and array position (to adjust for residual technical variation: 'batch effect'), participant ID and year of sample were included as random effects. Finally, *P*-values were adjusted for multiple testing using a false discovery rate of 0.05 with the Benjamini & Hochberg method [61]. The regression equations describing the mixed-effect models are provided in Supplementary Fig. S14. Differentially methylated regions (DMRs) associated with MeN were identified from the mixed-effect models using *ipDMR*, a tool incorporated into the *ENmix* R package, that autocorrelates *P*-values of individual CpGs from EWAS data [62]. Each DMR had a maximum window size of 1000 base pairs, contained at least two probes and was below a FDR threshold of 0.05. DMR position was annotated to the nearest genomic element from ENCODE database [63, 64].

Hypothesis-driven analysis of exposure-associated DMPs

The difference in DNAm *Beta* values between cases and controls (*ΔBeta*) and *P*-values for CpGs from the mixed-effect models were then examined at loci previously reported to be associated with environmental exposures taking a hypothesis-driven approach. To compile a comprehensive list of DMPs associated with heat stress, pesticide, or heavy metal exposure, three separate NCBI PubMed literature searches for human whole blood EWAS data, from either 450 K or EPIC Illumina arrays were performed. The search criteria and results for the literature reviews are described in more detail in Supplementary Table S15. For an exposure to be deemed as associated with MeN, we stipulated that the majority of exposure-associated DMPs should display congruency in the direction of methylation change (e.g. increased or decreased) and also reach statistical significance (unadjusted *P*-value $\leq .05$) in cases vs controls.

To explore if a negative association was true, and not due to the insufficient study power, we performed a *post-hoc* power calculation. In particular, we first estimated the effect size detectable by the two-sample Welch's *t*-test with 80% power with a significance threshold $\alpha = 0.05$. This was done by using the R pwr library (https://cran.r-project.org/web/packages/pwr/). We then calculated the actual effect sizes for the subset of the CpGs from the list of previously associated DMPs and our dataset metrics (case/control group sizes, groups' means and standard deviations) to confirm that we are adequately powered to detect a positive association (at least 80% power).

Methodological validation: smoking-associated CpGs

We additionally validated the approach of screening for exposure association in our cohort by fitting another mixed-effect model to examine DNA methylation changes associated with self-reported smoking in the cohort using the same array data as the MeN EWAS. This 'smoking EWAS' was adjusted identically for the fixed and random effects as the MeN EWAS, except smoking was used as the independent variable of interest. We then examined the *ΔBeta* and *P*-values for DMPs associated with smoking in the study reported by Cardenas *et al.* (2022) [29].

Supplementary data

Supplementary data is available at *EnvEpig* online.

Conflict of interest: None declared.

Funding

This study was funded by grants from the Colt Foundation and the UK Medical Research Council (MRC/UKRI).

Data availability

The original data underlying this article cannot be shared publicly due to the privacy of individuals who participated in the study. Nonidentifiable or summary data will be shared on reasonable request to the study steering group through the corresponding author.

Ethics approval and consent to participate

Ethical approval for the Colt cohort was received from the institutional review boards at Universidad Nacional Autonoma de Nicaragua, Leon (Acta No.116, Ano 2014; Acta No.71, Ano 2018), the London School of Hygiene and Tropical Medicine (Ref: 8643) and University College London (Ref: 14 175).

References

1. Ordunez P, Saenz C, Martinez R *et al*. The epidemic of chronic kidney disease in Central America. *Lancet Glob Health* 2014;**2**:e440–1.
2. Trabanino RG, Aguilar R, Silva CR *et al*. End-stage renal disease among patients in a referral hospital in El Salvador. *Rev Panam Salud Publica* 2002;**12**:202–06.
3. Ramirez-Rubio O, McClean MD, Amador JJ *et al*. An epidemic of chronic kidney disease in Central America: an overview. *J Epidemiol Community Health* 2013;**67**:1–3.
4. Cerdas M. Chronic kidney disease in Costa Rica. *Kidney Int Suppl* 2005;**68**:S31–S33.
5. Peraza S, Wesseling C, Aragon A *et al*. Decreased kidney function among agricultural workers in El Salvador. *Am J Kidney Dis* 2012;**59**:531–40.
6. Brooks DR, Ramirez-Rubio O, Amador JJ. CKD in Central America: a hot issue. *Am J Kidney Dis* 2012;**59**:481–84.
7. Wesseling C, Van Wendel De Joode B, Crowe J *et al*. Mesoamerican nephropathy: geographical distribution and time trends of chronic kidney disease mortality between 1970 and 2012 in Costa Rica. *Occup Environ Med* 2015;**72**:714–21.
8. Pan American Health Organization and World Health Organization. *Final Report from the 152nd Session of the Executive Committee*. 2013 https://iris.paho.org/handle/10665.2/4577 (03 June 2024, date last accessed).
9. García-Trabanino R, Jarquín E, Wesseling C *et al*. Heat stress, dehydration, and kidney function in sugarcane cutters in El Salvador—a cross-shift study of workers at risk of Mesoamerican nephropathy. *Environ Res* 2015;**142**:746–55.
10. González-Quiroz M, Pearce N, Caplin B *et al*. What do epidemiological studies tell us about chronic kidney disease of undetermined cause in Meso-America? A systematic review and meta-analysis. *Clin Kidney J* 2018;**11**:496–506.
11. Pearce N, and Caplin B. Let's take the heat out of the CKDu debate: more evidence is needed. *Occup Environ Med* 2019;**76**:357–359.
12. Smpokou ET, González-Quiroz M, Martins C *et al*. Environmental exposures in young adults with declining kidney function in a population at risk of Mesoamerican nephropathy. *Occup Environ Med* 2019;**76**:920–26.
13. Taylor SC, Nadeau K, Abbasi M *et al*. The ultimate qPCR experiment: producing publication quality, reproducible data the first time. *Trends Biotechnol* 2019;**37**:761–74.

14. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. *Int J Behav Healthc Res* 2011;**2**:320–32.
15. Ladd-Acosta C, Fallin MD. DNA methylation signatures as biomarkers of prior environmental exposures. *Curr Epidemiol Rep* 2019;**6**:1–13.
16. Johnson AA, Akman K, Calimport SRG *et al*. The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Res* 2012;**15**:483–494.
17. Salameh Y, Bejaoui Y, and El Hajj N. DNA methylation biomarkers in aging and age-related diseases. *Front Genetics* 2020;**11**:171.
18. Jin Z, Liu Y. DNA methylation in human diseases. *Genes Dis* 2018;**5**:1–8.
19. Argos M, Chen L, Jasmine F *et al*. Gene-specific differential DNA methylation and chronic arsenic exposure in an epigenome-wide association study of adults in Bangladesh. *Environ Health Perspect* 2015;**123**:64–71.
20. Bozack AK, Domingo-Relloso A, Haack K *et al*. Locus-specific differential dna methylation and urinary arsenic: An epigenome-wide association study in blood among adults with low-to-moderate arsenic exposure. *Environ Health Perspect* 2020;**128**:1–12.
21. Demanelis K, Argos M, Tong L *et al*. Association of arsenic exposure with whole blood DNA methylation: An epigenome-wide study of Bangladeshi adults. *Environ Health Perspect* 2019;**127**:057011.
22. Domingo-Relloso A, Riffo-Campos AL, Haack K *et al*. Cadmium, smoking, and human blood DNA methylation profiles in adults from the strong heart study. *Environ Health Perspect* 2020;**128**:06700.
23. Feng L, Guo X, Li T *et al*. Novel DNA methylation biomarkers for hexavalent chromium exposure: an epigenome-wide analysis. *Epigenomics* 2020;**12**:221–33.
24. Furlong MA, Paul KC, Yan Q *et al*. An epigenome-wide association study of ambient pyrethroid pesticide exposures in California's central valley. *Int J Hygiene Environ Health* 2020;**229**:113569–.
25. Paul KC, Chuang YH, Cockburn M *et al*. Organophosphate pesticide exposure and differential genome-wide DNA methylation. *Sci Total Environ* 2018;**645**:1135–43.
26. Xu R, Li S, Li S *et al*. Ambient temperature and genome-wide DNA methylation: a twin and family study in Australia. *Environ Pollut* 2021;**285**:117700.
27. Hoang TT, Qi C, Paul KC *et al*. Epigenome-wide DNA methylation and pesticide use in the agricultural lung health study. *Environ Health Perspect* 2021;**129**:1–11.
28. Ameer SS, Engström K, Hossain MB *et al*. Arsenic exposure from drinking water is associated with decreased gene expression and increased DNA methylation in peripheral blood. *Toxicol Appl Pharmacol* 2017;**321**:57–66.
29. Cardenas A, Ecker S, Fadadu RP *et al*. Epigenome-wide association study and epigenetic age acceleration associated with cigarette smoking among Costa Rican adults. *Sci Rep* 2022;**12**:4277.
30. Breeze CE, Batorsky A, Lee MK *et al*. Epigenome-wide association study of kidney function identifies trans-ethnic and ethnic-specific loci. *Genome Med* 2021;**13**:1–16.
31. Schlosser P, Tin A, Matias-Garcia PR *et al*. Meta-analyses identify DNA methylation associated with kidney function and damage. *Nat Commun* 2021;**12**:7174.
32. Chu AY, Tin A, Schlosser P *et al*. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun* 2017;**8**:1–12.
33. Horimoto ARVR, Xue D, Cai J *et al*. Genome-wide admixture mapping of estimated glomerular filtration rate and chronic kidney disease identifies European and African ancestry-of-origin loci in Hispanic and Latino individuals in the United States. *J Am Soc Nephrol* 2022;**33**:77–87.
34. Gonzalez-Quiroz M, Smpokou E-T, Silverwood RJ *et al*. Decline in kidney function among apparently healthy young adults at risk of Mesoamerican nephropathy. *J Am Soc Nephrol* 2018;**29**:2200–12.
35. Martin EM, Fry RC. Environmental influences on the epigenome: exposure- associated DNA methylation in human populations. Annual review of public health. *Ann Rev Public Health* 2018;**39**:309–33.
36. Petropoulos ZE, Keogh SA, Jarquín E *et al*. Heat stress and heat strain among outdoor workers in El Salvador and Nicaragua. *J Expos Sci Envir Epidemiol* 2023;**33**:622–30.
37. Crowe J, Wesseling C, Solano BR *et al*. Heat exposure in sugarcane harvesters in Costa Rica. *Am J Ind Med* 2013;**56**:1157–64.
38. Chen Y, Parvez F, Liu M *et al*. Association between arsenic exposure from drinking water and proteinuria: results from the health effects of Arsenic Longitudinal Study. *Int J Epidemiol* 2011;**40**:828–35.
39. Zheng LY, Umans JG, Tellez-Plaza M *et al*. Urine arsenic and prevalent albuminuria: evidence from a population-based study. *Am J Kidney Dis* 2013;**61**:385–94.
40. Zheng L, Kuo CC, Fadrowski J *et al*. Arsenic and chronic kidney disease: a systematic review. *Curr Environ Health Rep* 2014;**1**:192–207.
41. Järup L, Hellström L, Alfvén T *et al*. Low level exposure to cadmium and early kidney damage: the OSCAR study. *Occup Environ Med* 2000;**57**:668–72.
42. Mendley SR, Levin A, Correa-Rotter R *et al*. Chronic kidney diseases in agricultural communities report from a workshop. *Kidney Int* 2019;**96**:1071–9.
43. Yu HT. Progression of chronic renal failure. *Arch Intern Med* 2003;**163**:1417–29.
44. Kazancioğlu R. Risk factors for chronic kidney disease: an update. *Kidney Int Suppl (2011)* 2013;**3**:368–71.
45. Fraser HB, Lam LL, Neumann SM *et al*. Population-specificity of human DNA methylation. *Genome Biol* 2012;**13**:R8.
46. Breeze CE, Beck S, Berndt SI *et al*. The missing diversity in human epigenomic studies. *Nat Genet* 2022;**54**:737–39.
47. Su D, Wang X, Campbell MR *et al*. Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. *PLoS One* 2016;**11**:e0166486.
48. Climate Change Knowledge Portal. *Nicaragua – Climatology*. https://climateknowledgeportal.worldbank.org/country/nicaragua/climate-data-historical (03 June 2024, date last accessed).
49. Fischer RSB, Unrine JM, Vangala C *et al*. Evidence of nickel and other trace elements and their relationship to clinical findings in acute Mesoamerican nephropathy: a case-control analysis. *PLoS ONE* 2020;**15**:e0240988.
50. Mascarenhas S, Mutnuri S, Ganguly A. Deleterious role of trace elements - Silica and lead in the development of chronic kidney disease. *Chemosphere* 2017;**177**:239–49.
51. Ekong EB, Jaar BG, Weaver VM. Lead-related nephrotoxicity: a review of the epidemiologic evidence. *Kidney Int* 2006;**70**:2074–84.
52. González-Quiroz M, Camacho A, Faber D *et al*. Rationale, description and baseline findings of a community-based prospective cohort study of kidney function amongst the young rural population of Northwest Nicaragua. *BMC Nephrol* 2017;**18**:1–8.
53. Gonzalez-Quiroz M, Heggeseth B, Camacho A *et al*. Population-level detection of early loss of kidney function: 7-year follow-up of a young adult cohort at risk of Mesoamerican nephropathy. *Int J Epidemiol* 2023.
54. Levey AS, and Stevens LA. Estimating GFR using the ckd Epidemiology Collaboration (CKD-EPI) creatinine equation: more accurate

GFR estimates, lower CKD prevalence estimates, and better risk predictions. *Am J Kidney Dis* 2010;**55**:622–7.
55. Illumina. Infinium HD Methylation Assay Protocol Guide (15019519). 2015.
56. Ecker S. *DNA Methylation Data Preprocessing Summary [Internet]*. GitHub. 2020 (27October 2023, date last accessed). https://github.com/ucl-medical-genomics/DNA_methylation_data_processing/blob/master/DNA_methylation_preprocessing_summary.md
57. Xu Z, Niu L, Taylor JA. *The ENmix User's Guide* https://www.bioconductor.org/packages/release/bioc/vignettes/ENmix/inst/doc/ENmix.html#15_Differentially_methylated_regions_(DMRs) (27 October 2023, date last accessed).
58. Xu Z, Niu L, Taylor JA. The ENmix DNA methylation analysis pipeline for Illumina BeadChip and comparisons with seven other preprocessing pipelines. *Clin Epigenetics* 2021;**13**:216.
59. RCT. R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing. 2021. https://www.R-project.org/ (27 October 2023, date last accessed).
60. Zheng SC, Breeze CE, Beck S *et al*. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* 2018;**15**:1059–66.
61. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B* 1995;**57**:289–300.
62. Xu Z, Xie C, Taylor JA *et al*. ipDMR: identification of differentially methylated regions with interval P-values. *Bioinformatics* 2021;**37**:711–13.
63. Luo Y, Hitz BC, Gabdank I *et al*. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* 2020;**48**:D882–d9.
64. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74.
65. McCartney DL, Walker RM, Morris SW *et al*. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data* 2016;**9**:22–24.
66. Paul DS, Beck S. Advances in epigenome-wide association studies for common diseases. *Trends Mol Med* 2014;**20**:541–43.

Environmental Epigenetics, 2025, 11(1), dvaf001 , DOI: https://doi.org/10.1093/eep/dvaf001, Advance Access Publication 17 January 2025, Research Article
Received 4 August 2024; Revised 9 December 2024; Accepted 10 January 2025