Human Molecular Genetics, 2025, Vol. 34, 7, 611–616
https://doi.org/10.1093/hmg/ddae202
Advance access publication date 17 January 2025
Original Article

Genome-wide association study of urinary cadmium levels in current smokers from the multiethnic cohort study

Shannon M. Sullivan [1,2,*], Sharon E. Murphy[3,4], Daniel O. Stram[5], Lynne R. Wilkens[6], Christopher A. Haiman[5], Loïc Le Marchand[6], Irina Stepanov[2,4], S. Lani Park[6]

[1]Department of Laboratory Medicine and Pathology, University of Minnesota, 1200 Washington Ave S Suite 175, Minneapolis, MN 55415, United States
[2]Division of Environmental Health Sciences, School of Public Health, University of Minnesota, 1260 Mayo Building, MMC 807420 Delaware St. SE Minneapolis, MN 55455, United States
[3]Department of Biochemistry Molecular Biology and Physics, University of Minnesota, 123 Snyder Hall1475 Gortner Ave St. Paul, MN 55108, United States
[4]Masonic Cancer Center, University of Minnesota, 2231 6th St SE Minneapolis, MN 55455, United States
[5]Keck School of Medicine, Department of Preventative Medicine, University of Southern California, 1975 Zonal Avenue, Los Angeles, CA 90033, United States
[6]Epidemiology Program, University of Hawaii Cancer Center, 701 Ilalo Street, Honolulu, HI 96813, United States

*Corresponding author. Department of Laboratory Medicine and Pathology, University of Minnesota, 1200 Washington Avenue South Suite 175, Minneapolis, MN 55455, United States. Fax: 612-626-2696. Email: sull0401@umn.edu

Abstract

Background: Cadmium (Cd), classified as an International Agency for Research on Cancer (IARC) Group 1 human carcinogen, is present in cigarette smoke. Recent studies have illustrated the potential role of genetics in influencing Cd biomarker levels.
Methods: We conducted a genome-wide association study (GWAS) of urinary Cd levels in 1977 current smokers from the Multiethnic Cohort Study, comprising participants from five different racial and ethnic groups. Linear regression models were adjusted for age at urine collection, sex, self-reported race/ethnicity, and the top ten leading principal components.
Results: Among the 11 710 497 single nucleotide polymorphisms (SNP) analyzed, no associations with urinary Cd reached genome-wide significance ($P < 5.0 \times 10^{-8}$). Notably, five variants demonstrated suggestive associations with urinary Cd levels ($P < 1.0 \times 10^{-6}$). Lead variants included: rs10097646 in the *SCARA* gene at 8q13.2 ($P = 2.62 \times 10^{-7}$); rs7444817 in the *NIPBL* gene at 5p13.2 ($P = 3.10 \times 10^{-7}$), rs830422 in the *SPINK4* gene at 9q13.2 ($P = 4.89 \times 10^{-7}$); chrX:145489901 in the *SLC9A7* gene at Xq121.1 ($P = 5.38 \times 10^{-7}$); and rs73074456 at 5p13.3 ($P = 5.86 \times 10^{-7}$).
Conclusions: Our GWAS of urinary Cd levels in a diverse population of people who smoke, revealed suggestive associations with variants in *SCARA5*, *NIPBL*, *SPINK4*, *SLC9A7*, and 5p13.3. These findings underscore the potential role of genetic factors in understanding and mitigating the health risks associated with internal dose of carcinogens, particularly in the context of tobacco-related carcinogens.

Keywords: urinary cadmium; cigarette smoking; cadmium exposure; genetic variants; biomarkers

Introduction

Cadmium (Cd) is a non-essential, toxic metal found in the environment from both natural (e.g. geological deposits, forest fires) and anthropogenic sources (e.g. mining, smelting, and industrial processes) [1]. Exposure to Cd is linked to several tobacco-related cancers, including lung and pancreatic cancers [2–5]. Exposure to Cd occurs through ingestion and inhalation, with smoking being the major contributor [6]. Following exposure, Cd is transported in the blood and accumulates mainly in the kidney where it is slowly excreted in urine [7]. The inter-individual variability in urinary Cd levels among people who smoke is well-documented, however, the contribution of genetic factors to variability remains unclear [8–15].

Genome-wide association studies (GWAS) of blood Cd, a short-term exposure biomarker, have identified several genetic variants linked to Cd absorption, internal dose, and excretion [16, 17]. However, these studies carry limitations, as blood Cd levels may be significantly influenced by recent exposures and may not capture a broader representation of genetic diversity due to the exclusion of various racial and ethnic groups. Long-term exposure to Cd can be measured using urinary Cd, which has been shown to accurately reflect total body burden due to it's long biological half-life in the body [5, 18, 19]. Notably, there are no published genome-wide associating findings for urinary Cd [20], highlighting the need for further research on the genetic determinants influencing urinary Cd levels.

To address these gaps, we used a GWAS approach to investigate genetic variants associated with urinary Cd levels in 1977 current smokers representing five racial and ethnic populations in the Multiethnic Cohort (MEC) Study.

Materials and Methods

Study population

The MEC is an ongoing prospective study of > 215 000 men and women, who were between the ages of 45 to 75 at recruitment, belonging to the following five racial and ethnic groups: African

Received: February 24, 2024. **Revised:** August 12, 2024. **Accepted:** December 23, 2024

American, Japanese American, Latino, Native Hawaiian, and White. Study recruitment has been described in further detail [21]. Briefly, potential participants were identified between 1993 and 1996 in Hawaii and California (mainly Los Angeles County) through voter registration lists, drivers' license files, and Health Care Financing Administration data. At enrollment, participants completed a 26-page self-administered questionnaire, including demographics, diet, smoking, medical and reproductive histories, and other lifestyle factors. Demographic questions included detailed information on participants' and their parents' places of birth, as well as their ethnic or racial backgrounds [22]. Population groups were categorized based on The Surveillance, Epidemiology and End Results (SEER) hierarchy [23]. Baseline questionnaire indicated a wide range of smoking prevalence for the entire MEC; the prevalence of current smokers was highest among African Americans (19%), followed by Native Hawaiians (17%), Whites (13%), Latinos (11%), and Japanese Americans (9%) [24].

Approximately ten years after study enrollment, a randomly selected subset of MEC participants provided a blood and overnight (participants recruited in Hawaii) or first-morning urine sample (participants recruited California) for genetic and biomarker analyses. In addition, participants completed questionnaires that included average daily cigarette smoking during the past two weeks, smoking duration, and medication records.

The present study was conducted within a subgroup of participants who were current cigarette smokers and lung cancer-free at the time of blood and urine collection (which occurred between January 1997 and December 2006) and had complete data on genetics (n = 2239) [25] and urinary Cd (n = 1977; Supplementary Fig. S1) [15]. Approval for the MEC study protocol was approved by the Institutional Review Boards (IRBs) at the University of Southern California (IRB Study #HS-16-00719), University of Hawaii (IRB Study # CHS17649) and University of Minnesota (IRB Study #00003366). Prior to participation, written consent was obtained from all study participants.

Urinary cadmium and total nicotine equivalents

Details of the analytic methods used to determine the levels of urinary biomarkers of Cd and total nicotine equivalents (TNE; a biomarker of nicotine uptake) in overnight or first-morning urine have been previously published [15, 26]. Briefly, urinary Cd was measured by inductively-coupled plasma mass spectrometry (ICP-MS) [15]. TNE (the molar sum of nicotine *N*-oxide, total nicotine, total cotinine, and total 3-hydroxycotinine [HCOT]) was measured by liquid chromatography–tandem mass spectrometry (LC-MS/MS) [26]. The detection limits were 13 ng/ml for nicotine, 20 ng/ml for cotinine, and 0.12 ng/ml Cd. The coefficient of variation (CV) for the assays across runs were 16.7% for nicotine, 10.1% for cotinine, and 3.1% for Cd. Cd Biomarker levels below the limit of detection (LOD; n = 15, 0.76%) were replaced with LOD/2.

Genotyping, quality control, and imputation

Details of the genotyping, quality control, imputation, and principal components analysis (PCA) of these samples have been previously published [25]. Briefly, blood lymphocyte DNA was extracted using a Q1Aamp DNA blood extraction kit (Qiagen Inc., Valencia, CA) and extracted samples were genotyped using the Illumina Human1M-Duo BeadChip (Illumina, San Diego). Genotyping quality control consisted of removing individual samples where ≥ 2% of genotypes were not called/missing (n = 8), removing single nucleotide polymorphisms (SNPs) ≤ 98% call rate (n = 67 761), known duplicate samples (n = 25), excluding samples with close relatives based on estimated identical by descent (IBD) status (n = 59), and samples with mismatched/conflicting sex (n = 7). Resulting in a total of 1 131 426 SNPs with typed genotypes for 2239 samples [25].

To extend the genotype analysis, variants were imputed using SHAPEIT [27] and IMPUTE2 [28] to a reference panel from the 1000 Genomes Project [29] Phase I integrated variant set (March 2012; version 3). Post-imputation, SNPs were filtered with an IMPUTE2 info quality score < 0.30 or minor allele frequency (MAF) < 1% by racial and ethnic group resulting in a total of 11 710 497 SNPs available for final analysis.

Statistical analysis

Principal components (PC) were previously estimated on the combined multiethnic sample using a random sample of 19 059 autosomal SNPs with frequency ≥ 2% over the five racial/ethnic groups to capture population substructure [25].

SNP associations with urinary Cd levels were estimated using linear regression models of log-transformed urinary Cd, adjusting for age at urine collection, sex (male/female), self-reported race/ethnicity (African American, Native Hawaiians, Whites, Latinos, and Japanese Americans), and the top 10 leading PCs (for population stratification adjustment). Therefore, the estimated beta coefficient and 95% confidence interval (CI) can be interpreted as a one-unit change (ng/mL) in log-transformed Cd for one-unit increase in the SNP dosage value. Quantile-quantile (QQ) plots and genomic inflation factor (lambda, λ) were used to assess the adequacy of the model. Genome-wide statistical significance was based on a threshold of $P < 5.0 \times 10^{-8}$ and was considered suggestive at $P < 1.0 \times 10^{-6}$. Sensitivity analysis with additional adjustment for (1) TNE and (2) log-transformed creatinine was conducted to assess whether the associations were independent of this biomarker of nicotine uptake and urine dilution, respectively. While adjustment for creatinine may help normalize variation in urinary biomarkers, several factors affect urinary creatinine levels, including sex, muscle mass, age, and racial and ethnic group. We have demonstrated that in this multiethnic population of smokers, expressing urinary biomarker concentrations per mg creatinine can result in an inflated denominator leading to underestimates concentrations of the urinary biomarker [30]. Therefore, our primary model is unadjusted for creatinine.

To evaluate independent SNP effects, a conditional analysis including the lead SNP as an additional covariate in the linear regression model was conducted that included less correlated lead SNPs ($P < 1.0 \times 10^{-6}$). We further tested deviations from Hardy–Weinberg Equilibrium (HWE) for all top loci by self-reported racial and ethnic group. Stratified analyses were conducted for (1) self-reported racial and ethnic data and (2) sex for chromosome X only. PLINK (https://www.cog-genomics.org/plink2) was used for the statistical analysis. R (version 3.1.2; https://www.r-project.org) was used to create the Manhattan and QQ plots. LocusZoom was used to generate regional association plots [31].

Results

A total of 1977 people who smoke cigarettes with available genotyping data and urinary Cd measurements were included in this study. Significant variation in smoking history (pack-years), urinary total nicotine equivalents (TNE), and Cd levels were observed across racial and ethnic groups (Table 1) [15]. No genome-wide significant associations ($P < 5.0 \times 10^{-8}$) with urinary Cd were identified. Notably, five loci (*SCARA5*, *NIPBL*,

Table 1. Main characteristics of multiethnic cohort study current smokers at time of urine collection stratified by race/ethnicity (n = 1977).

	All n = 1977	African Americans n = 285	Native Hawaiians n = 296	Whites n = 390	Latinos n = 418	Japanese Americans n = 588
Sex						
Males, n (%)	9235 (47.3)	88 (30.9)	109 (36.8)	173 (44.4)	221 (52.9)	344 (58.5)
Females, n (%)	1042 (52.7)	197 (69.2)	187 (63.2)	217 (55.6)	197 (47.1)	244 (41.5)
	Median (Interquartile range—25th and 75th percentile)					
Age at urine collection, years	63.7 (59.3, 69.5)	64.5 (59.8, 69.1)	61.0 (56.9, 65.9)	62.5 (59.2, 69.3)	65.7 (61.7, 70.8)	63.4 (59.1, 69.8)
Age at smoking initiation, years	22.5 (18.5, 28.0)	24.3 (19.5, 30.5)	21.5 (17.5, 27.5)	21.5 (18.0, 26.5)	23.5 (19.5, 30.5)	21.5 (17.8, 26.0)
Pack-Years of smoking	23.5 (11.6, 38.3)	18.3 (10.2, 32.7)	26.1 (14.7, 44.5)	34.5 (18.3, 47.5)	14.2 (7.0, 26.0)	25.1 (15.4, 38.3)
Urinary TNE, nmol/mL	32.3 (19.6, 52.9)	44.5 (27.8, 71.1)	30.1 (19.3, 46.3)	36.0 (22.0, 58.1)	32.7 (20.9, 53.7)	27.3 (15.7, 42.9)
Urinary Cd, ng/mL	0.60 (0.36, 1.02)	0.84 (0.48, 1.26)	0.57 (0.36, 0.96)	0.48 (0.30, 0.78)	0.72 (0.42, 1.08)	0.54 (0.30, 0.84)

Abbreviations: Pack-years of smoking – calculated based on self-report number of cigarette packs per day x number of years smoked, TNE – total nicotine equivalents, Cd – cadmium

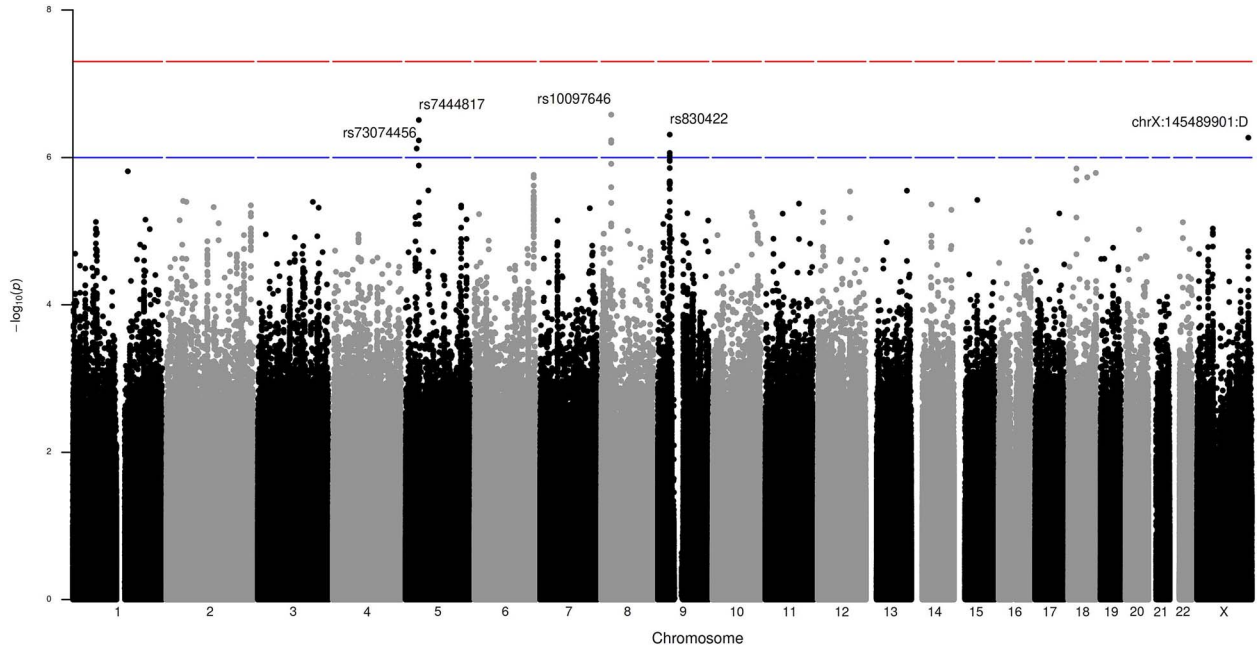

Figure 1. Genome-wide association study (GWAS) of urinary cadmium (Cd; natural log transformed) in current smokers from the multiethnic cohort study. Manhattan plot of the −log10 (*P*-values) from the test of association between genetic variants and continuous urinary Cd (natural log transformed), adjusted for self-reported race/ethnicity, sex, age at urine collection, and principal components (PC) 1–10 plotted as a function of the chromosomal position. The horizontal red line is defined as a $P < 5.0 \times 10^{-8}$ (genome-wide significance) and the blue line is defined as a $P < 1.0 \times 10^{-6}$ (genome-wide suggestive significance).

SPINK4, *SLC9A7*, and a variant on 5p13.3) demonstrated suggestive significance ($P < 1.0 \times 10^{-6}$; Fig. 1; Table 2). The Manhattan and quantile-quantile (Q-Q) plots are provided (Fig. 1; Supplementary Fig. S2), with the latter showing negligible evidence of test statistic inflation (genomic inflation factor [λ] = 0.996). Regional association plots for the top autosomal loci that reached the suggestive significance threshold are provide in Supplementary Fig S3. Lead SNPs were confirmed using conditional analyses (Supplementary Table S1). We further evaluated Hardy Weinberg Equilibrium (HWE) within each self-reported racial and ethnic group for the four autosomal lead variants and in chromosome X for females only. We found no evidence for violation of HWE ($P > 0.030$ for all 22 tests).

In the sensitivity analyses adjusting for TNE, the effect estimates for the five lead SNPs were in the same direction, however, slightly attenuated in magnitude compared to the original model (largest $P = 4.79 \times 10^{-4}$; Supplementary Table S2). Similarly, the log-creatinine-adjusted sensitivity analysis also indicated a reduction in effect sizes and significance levels (largest $P = 1.62 \times 10^{-3}$; Supplementary Table S2).

For chromosome X, given X chromosome copy number varies by sex, we conducted a sex-stratified analysis of the chrX:145489901-Cd association to determine if the larger proportion of females (n = 1042) compared to males (n = 935) might be influencing this chromosome X finding. The results demonstrated that the effect size remained consistent across both the overall and sex-stratified analysis. However, the p-value in the sex-stratified analysis was less significant, with females showing a larger reduction compared to males ($P = 2.50 \times 10^{-3}$ versus $P = 1.07 \times 10^{-4}$, respectively; Supplementary Table S3). No SNP reached the suggestive threshold of $P < 1.0 \times 10^{-6}$ in the sex-stratified analysis of chromosome X ($P > 1.34 \times 10^{-5}$).

We further examined whether the genetic effects observed for the top variants in the overall GWAS varied by racial and ethnic group. However, no evidence of heterogeneity was found across racial and ethnic groups ($P > 0.143$; Supplementary Table S4). In

Table 2. Genome-wide association study results for the association of genetic variants with urinary cadmium (natural log). The table includes the top five single nucleotide polymorphisms that reached suggestive significance ($P < 1.0 \times 10^{-06}$).

Cytoband	SNP	CHR	Position (GRCh37)	A1	A2	Located in/near	INFO	FRQ A1	BETA	SE	P
8q13.2	rs10097646	8	27 833 007	T	C	*SCARA5*	0.840	0.203	−0.17	0.03	2.62E-07
5p13.2	rs7444817	5	36 901 614	T	C	*NIPBL*	0.901	0.073	0.26	0.05	3.10E-07
9q13.2	rs830422	9	33 219 724	A	C	*SPINK4*	0.990	0.215	−0.16	0.03	4.89E-07
Xq121.1	chrX:145489901:D	X	145 489 901	TATATATATATAA	T	*SLC9A7*	0.867	0.114	−0.17	0.03	5.38E-07
5p13.3	rs73074456	5	30 941 598	G	A		0.912	0.022	0.24	0.05	5.86E-07

Model adjusted for age at urine collection, sex (male/female), self-reported race/ethnicity (where appropriate; African American, Native Hawaiian, White, Latino, and Japanese American), and the top 10 leading principal components (PC1-10). **Abbreviations:** SNP - Single nucleotide polymorphism; CHR - Chromosome; A1 - Allele 1, Reference allele; A2 - Allele 2, Alternate allele; INFO - genotype imputation quality score. The info score is an estimated quality of the ratio between the observed and expected statistical information [28]; FRQ - frequency of A1 allele in the population; BETA - beta for the A1-allele effect on urinary cadmium (natural log); SE - standard error; P-value for the association between the SNP and urinary cadmium (natural log)

the ethnic-specific GWAS analyses, one variant was significantly associated with urinary Cd levels in the Japanese population (rs146245782 located *in EXD3*, $P = 6.60 \times 10^{-9}$; Supplementary Fig. S4, Supplementary Table S4). However, there are no SNPs in LD with this variant, and it could not be evaluated in the other racial/ethnic groups due to its rarity. Additionally, a corresponding signal for *SCARA5* (rs1148291) reached suggestive significance in the African American-stratified analysis ($P = 3.38 \times 10^{-6}$; Supplementary Fig. S4A). Three lead SNPs from the overall analysis (rs4744237, rs696750, and chrX:12178184) reached suggestive significance in the Native Hawaiian-stratified analysis ($P < 6.85 \times 10^{-7}$; Supplementary Fig. S4C). Three loci (rs117059503, 113 056 833:D, and rs57692239) reached suggestive significance in the Latino-stratified analysis ($P < 5.25 \times 10^{-7}$; Supplementary Fig. S4G) however none of these reached suggestive significance in any other racial and ethnic group, Supplementary Table S4). Finally, no associations were observed in this study for loci previously identified in GWAS studies of blood Cd or smoking (Supplementary Table S5; smallest $P = 0.0002$).

Discussion

In this multiethnic population of people who smoke, we identified suggestive associations between variants in *SCARA5*, *NIPBL*, *SPINK4*, *SLC9A7* and the 5p13.3 region with urinary Cd levels. Notably, two genomic regions (*SCARA5* and *SLC9A7*) have previously been linked to metal uptake and delivery pathways, reinforcing the potential genetic influence on urinary Cd levels in individuals who smoked at time of biospecimen collection. These findings highlight the importance of considering genetic factors in understanding and mitigating the health risks associated with carcinogen exposure, particularly in the context of tobacco use.

The most robust signal for urinary Cd was observed in the *SCARA5* (scavenger receptor class A member 5) gene, predominately expressed in epithelial cells related to mucosal surfaces. Notably, *SCARA5*-associated pathways include the mediation of non-transferrin-dependent delivery of iron, a process crucial for specific kidney cell type development [32]. This is particularly relevant given the suggested metal–metal interaction between iron and Cd, where iron-deficiency increases absorption of Cd, due to similar mechanisms for divalent metal absorption [33, 34]. However, we could not evaluate this relationship in individuals with iron deficiencies in our study population. Additionally, reduced *SCARA5* expression has been observed in various cancer cells, including lung, liver, and renal cancer, with upregulation hindering tumor growth and metastasis [35–38].

The second strongest signal was identified in the *NIPBL* (NIPBL cohesion loading factor) gene, a protein coding gene involved in chromosome cohesion and regulation of gene expression. Diseases associated with this gene include a rare congenital disorder called Cornelia De Lange Syndrome [39]. Additional evidence suggests that *NIPBL* mutations might play a role in tumorigenesis of cancers with microsatellite instability, such as colorectal and gastric cancers [40]. Further investigation is required to understand its relationship with Cd or related pathways. A second independent signal was also identified on chromosome 5, in the 5p13.3 region, but this location does not map to any known protein-coding genes. However, we speculate this signal might be an artifact, given the lack of linkage disequilibrium between the lead and neighboring SNPs. Future studies may provide insight into the relevance of this locus.

The third signal was located in the *SPINK4* (serine peptidase inhibitor, Kazal type 4) gene, which encodes an enzyme inhibitor for serine proteases, essential for various physiological processes, including digestion, blood clotting, and immune response [41]. Recent evidence suggests a potential association between lower *SPINK4* expression and adverse outcomes in colorectal cancer [42]. Interestingly, some studies have suggested a potential association between Cd exposure and increased risk of colorectal cancer although evidence is not yet conclusive [43,44]. This suggests a potential pathway through which Cd exposure could influence carcinogenesis.

The fourth strongest signal was located near the *SLC9A7* (Solute Carrier Family 9, Sodium/Hydrogen Exchanger, Isoform 7), a gene expressed in the brain, skeletal muscle, and secretory tissues that regulates sodium (Na+) and hydrogen (H+) exchange across cellular membranes [45]. However, we speculate this signal might be an artifact, given the lack of linkage disequilibrium between the lead and neighboring SNPs. Furthermore, when we performed a sex-stratified analysis of chromosome X, we did not see a signal in either sex group.

Two previous GWAS studies of blood Cd have yielded diverse outcomes [16, 17]. One study in a Swedish cohort identified an association between serum Cd and the 6q14.1 locus in the *CD109* gene, suggesting either heightened clearance or reduced Cd absorption [16]. In contrast, the second GWAS in Swedish women did not identify a genome-wide significant association with erythrocyte Cd [17]. Yet, when the analysis was stratified to never smokers only (n = 1728), two independent regions were identified: a cluster of 13 variants on locus 18.q13.3 located in *XKR9* gene and one on locus 18.p11.31 located in the *DLGAP1* gene. We integrated these specific SNPs and smoking-related SNPs from prior literature into a candidate SNP analysis, however, we could not replicate previous findings with urinary Cd. Likely, these discrepancies stem from differences in biomarkers—blood Cd reflects short-term exposure, whereas urinary Cd reflects longer-term body burden. Additionally, our sample included

smokers, and the substantial Cd input from smoking could potentially attenuate and dilute the signal even after adjusting for smoking measures, which we observed in our overall analysis when we further adjusted for a measure of internal smoking dose (TNE).

To the best of our knowledge, this is the first GWAS to study urinary Cd levels. This study strengths include its use of an established, racially and ethnically diverse population, accounting for a range of confounders, including age, sex, smoking, and population stratification, while providing valuable insights into a diverse population of people who smoke [46]. However, our study includes a few limitations. First, the sample size for our race and ethnic-specific groups were limited (n = 285–588); thus, population specific analyses were only explored. Nonetheless, existing research underscores that many loci that narrowly miss the conventional threshold of genome-wide significance ($P < 5 \times 10^{-8}$) truly correlate with the trait of interest when sample size is increased [47]. Second, while this GWAS was performed in a diverse population, findings may not be generalizable to all multiethnic populations across the United States. Replication in an independent diverse cohort of people who smoke is warranted. Third, while the suggestive association on *SCARA5* is interesting given its known role in iron delivery, we were unable to further assess this relationship in individuals with iron deficiencies within our study population. Lastly, understanding the potential biological mechanisms underlying identified associations is needed.

In summary, our GWAS identified *SCARA5, NIPBL, SPINK4,* and *SLC9A7* as genes that may contain variants associated with urinary Cd biomarker levels and internal dose in people who smoke. Future studies are warranted to further investigate the role of these genes and genetics more generally on urinary Cd levels.

Acknowledgements

The authors gratefully acknowledge the time and effort of all the Multiethnic Cohort (MEC) study participants.

Supplementary data

Supplementary data is available at *HMG Journal* online.

Conflict of interest statement: The authors declare no conflicts of interest.

Funding

This work was supported by the National Institute of Health/-National Cancer Institute [grant numbers: P01 CA138338, University of Minnesota; U01 CA164973, University of Hawaii; and P30 CA071789, University of Hawaii]. S.M. Sullivan was also supported in part by grants from the National Institute for Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention [5T42 OH008434, University of Minnesota]. The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Data availability

Investigators interested in accessing the raw de-identified data may apply through the data sharing process of the Multiethnic Cohort (MEC) study (see "Data Sharing" on the MEC website: https://www.uhcancercenter.org/for-researchers/mec-data-sharing). Documentation of IRB approval is required for all projects that request the use of MEC data.

The data generated as part of this study was funded by the NIH and will be shared in accordance with the NIH Genomic Data Sharing policies via the database of Genotypes and Phenotypes (dbGaP). The summary statistics from the overall GWAS of urinary Cd are available at: 10.6084/m9.figshare.27666297.

References

1. International Agency for Research on Cancer (IARC). *Cadmium and cadmium compounds*, Vol. **100C**. Lyon, France: IARC Monographs, 2012, 121–145.
2. Nordberg GF, Bernard A, Diamond GL. *et al.* Risk assessment of effects of cadmium on human health (IUPAC technical report). *Pure Appl Chem* 2018;**90**:755–808.
3. Richter P, Faroon O, Pappas RS. Cadmium and cadmium/zinc ratios and tobacco-related morbidities. *Int J Environ Res Public Health* 2017;**14**:1154.
4. Satarug S, Garrett SH, Sens MA. *et al.* Cadmium, environmental exposure, and health outcomes. *Environ Health Perspect* 2010;**118**:182–190.
5. Järup L, Åkesson A. Current status of cadmium as an environmental health problem. *Toxicol Appl Pharmacol* 2009;**238**:201–208.
6. Pappas RS, Polzin GM, Zhang L. *et al.* Cadmium, lead, and thallium in mainstream tobacco smoke particulate. *Food Chem Toxicol* 2006;**44**:714–723.
7. Nordberg M, Nordberg GF. Metallothionein and cadmium toxicology—historical review and commentary. *Biomol Ther* 2022;**12**:360.
8. Paschal DC, Burt V, Caudill SP. *et al.* Exposure of the U.S. population aged 6 years and older to cadmium: 1988-1994. *Arch Environ Contam Toxicol* 2000;**38**:377–383.
9. Menke A, Muntner P, Silbergeld EK. *et al.* Cadmium levels in urine and mortality among U.S. adults. *Environ Health Perspect* 2009;**117**:190–196.
10. Riederer AM, Belova A, George BJ. *et al.* Urinary cadmium in the 1999–2008 U.S. National Health and nutrition examination survey (NHANES). *Environ Sci Technol* 2013;**47**:1137–1147.
11. Yassin AS, Martonik JF. Urinary cadmium levels in the U S working population, 1988-1994. *J Occup Environ Hyg* 2004;**1**:324–333.
12. Adams SV, Newcomb PA. Cadmium blood and urine concentrations as measures of exposure: NHANES 1999-2010. *J Expo Sci Environ Epidemiol* 2014;**24**:163–170.
13. Tellez-Plaza M, Navas-Acien A, Caldwell KL. *et al.* Reduction in cadmium exposure in the United States population, 1988-2008: the contribution of declining smoking rates. *Environ Health Perspect* 2012;**120**:204–209.
14. Adams SV, Passarelli MN, Newcomb PA. Cadmium exposure and cancer mortality in the third National Health and nutrition examination survey cohort. *Occup Environ Med* 2011;**69**:153–156.
15. Cigan SS, Murphy SE, Alexander BH. *et al.* Ethnic differences of urinary cadmium in cigarette smokers from the multiethnic cohort study. *Int J Environ Res Public Health* 2021;**18**:2669.
16. Ng E, Lind PM, Lindgren C. *et al.* Genome-wide association study of toxic metals and trace elements reveals novel associations. *Hum Mol Genet* 2015;**24**:4739–4745.
17. Borné Y, Söderholm M, Barregard L. *et al.* Genome wide association study identifies two loci associated with cadmium in erythrocytes among never-smokers. *Hum Mol Genet* 2016;**25**:2342–2348.

18. Nordberg GF, Nogawa K, Nordberg M. Cadmium. In: Nordberg GF, Fowler BA, Nordberg M. (eds). *Handbook on the Toxicology of Metals*. (Fourth Edition) Cambridge, MA, USA: Academic Press, 2015, 667–716.
19. Vacchi-Suzzi C, Kruse D, Harrington J. *et al.* Is urinary cadmium a biomarker of long-term exposure in humans? A review. *Curr Environ Health Rep* 2016;**3**:450–458.
20. Buniello A, MacArthur JAL, Cerezo M. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;**47**:D1005–D1012.
21. Kolonel LN, Henderson BE, Hankin JH. *et al.* A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;**151**:346–357.
22. Multiethnic Cohort Study | Cohort Composition and Questionnaires. https://www.uhcancercenter.org/for-researchers/mec-questionnaires (accessed Jul 24, 2024).
23. Race and Hispanic Ethnicity Changes - SEER Documentation. https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity/index.html (accessed Aug 7, 2024).
24. Stram DO, Park SL, Haiman CA. *et al.* Racial/ethnic differences in lung cancer incidence in the multiethnic cohort study: An update. *J Natl Cancer Inst* 2019;**111**:811–819.
25. Patel YM, Stram DO, Wilkens LR. *et al.* The contribution of common genetic variation to nicotine and cotinine Glucuronidation in multiple ethnic/racial populations. *Cancer Epidemiol Biomarkers Prev* 2015;**24**:119–127.
26. Murphy SE, Park S-SL, Thompson EF. *et al.* Nicotine N-glucuronidation relative to N-oxidation and C-oxidation and UGT2B10 genotype in five ethnic/racial groups. *Carcinogenesis* 2014;**35**:2526–2533.
27. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2012;**9**:179–181.
28. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;**5**:e1000529.
29. Altshuler DL, Durbin RM, Abecasis GR. *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010;**467**:1061–1073.
30. Park SL, Carmella SG, Ming X. *et al.* Variation in levels of the lung carcinogen NNAL and its glucuronides in the urine of cigarette smokers from five ethnic groups with differing risks for lung cancer. *Cancer Epidemiol Biomarkers Prev* 2015;**24**:561–569.
31. Boughton AP, Welch RP, Flickinger M. *et al.* LocusZoom.Js: interactive and embeddable visualization of genetic association study results. *Bioinformatics* 2021;**37**:3017–3018.
32. Li JY, Paragas N, Ned RM. *et al.* Scara5 is a ferritin receptor mediating non-transferrin iron delivery. *Dev Cell* 2009;**16**:35–46.
33. Goyer RA. Toxic and essentail metal interactions. *Annu Rev Nutr* 1997;**17**:37–50.
34. Gallagher CM, Chen JJ, Kovach JS. The relationship between body iron stores and blood and urine cadmium concentrations in US never-smoking, non-pregnant women aged 20-49 years. *Environ Res* 2011;**111**:702–707.
35. Liu J, Hu G, Chen D. *et al.* Suppression of SCARA5 by Snail1 is essential for EMT-associated cell migration of A549 cells. *Oncogenesis* 2013;**2**:e73.
36. Huang J, Zheng D-L, Qin F-S. *et al.* Genetic and epigenetic silencing of SCARA5 may contribute to human hepatocellular carcinoma by activating FAK signaling. *J Clin Invest* 2010;**120**:223–241.
37. Xu Z, Hong Z, Ma M. *et al.* Rock2 promotes RCC proliferation by decreasing SCARA5 expression through β-catenin/TCF4 signaling. *Biochem Biophys Res Commun* 2016;**480**:586–593.
38. Yan N, Zhang S, Yang Y. *et al.* Therapeutic upregulation of class a scavenger receptor member 5 inhibits tumor growth and metastasis. *Cancer Sci* 2012;**103**:1631–1639.
39. Borck G, Redon R, Sanlaville D. *et al.* NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. *J Med Genet* 2004;**41**:e128.
40. Kim MS, An CH, Chung YJ. *et al.* NIPBL, a cohesion loading factor, is somatically mutated in gastric and colorectal cancers with high microsatellite instability. *Dig Dis Sci* 2013;**58**:3376–3378.
41. Metsis M, Cintra A, Solfrini V. *et al.* Molecular cloning of PEC-60 and expression of its mRNA and peptide in the gastrointestinal tract and immune system. *J Biol Chem* 1992;**267**:19829–19832.
42. Wang X, Yu Q, Ghareeb WM. *et al.* Downregulated SPINK4 is associated with poor survival in colorectal cancer. *BMC Cancer* 2019;**19**:1–15.
43. Naji S, Issa K, Eid A. *et al.* Cadmium induces migration of colon cancer cells: roles of reactive oxygen species, P38 and Cyclooxygenase-2 | cell Physiol Biochem. *Cell Physiol Biochem* 2019;**52**:1517–1534.
44. Sun M, Jiang Z, Gu P. *et al.* Cadmium promotes colorectal cancer metastasis through EGFR/Akt/mTOR signaling cascade and dynamics. *Sci Total Environ* 2023;**899**:165699.
45. Xu H, Ghishan FK, Kiela PR. SLC9 gene family: function, expression, and regulation. *Compr Physiol* 2018;**8**:555–583.
46. Wojcik GL, Graff M, Nishimura KK. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 2019;**570**:514–518.
47. Chen Z, Boehnke M, Wen X. *et al.* Revisiting the genome-wide significance threshold for common variant GWAS. *G3 (Bethesda)* 2021;**11**:jkaa056.

Human Molecular Genetics, 2025, Vol. 34, 7, 611–616
https://doi.org/10.1093/hmg/ddae202
Original Article