Urinary Metabolite Risk Biomarkers of Lung Cancer: A Prospective Cohort Study

Majda Haznadar[1], Qiuyin Cai[2], Kristopher W. Krausz[3], Elise D. Bowman[1], Ezra Margono[1], Rintaro Noro[4], Matthew D. Thompson[3], Ewy A. Mathé[5], Heather M. Munro[6], Mark D. Steinwandel[6], Frank J. Gonzalez[3], William J. Blot[2,6], and Curtis C. Harris[1]

Abstract

Background: Lung cancer is a major health burden causing 160,000 and 1.6 million deaths annually in the United States and worldwide, respectively.

Methods: While seeking to identify stable and reproducible biomarkers in noninvasively collected biofluids, we assessed whether previously identified metabolite urinary lung cancer biomarkers, creatine riboside (CR), N-acetylneuraminic acid (NANA), cortisol sulfate, and indeterminate metabolite 561+, were elevated in the urines of subjects prior to lung cancer diagnosis in a well-characterized prospective Southern Community Cohort Study (SCCS). Urine was examined from 178 patients and 351 nondiseased controls, confirming that one of four metabolites was associated with lung cancer risk in the overall case–control set, whereas two metabolites were associated with lung cancer risk in European-Americans.

Results: OR of lung cancer associated with elevated CR levels, and adjusted for smoking and other potential confounders, was 2.0 [95% confidence interval (CI), 1.2–3.4; $P = 0.01$]. In European-Americans, both CR and NANA were significantly associated with lung cancer risk (OR = 5.3; 95% CI, 1.6–17.6; $P = 0.006$ and OR=3.5; 95% CI, 1.5–8.4; $P = 0.004$, respectively). However, race itself did not significantly modify the associations. ROC analysis showed that adding CR and NANA to a model containing previously established lung cancer risk factors led to a significantly improved classifier ($P = 0.01$). Increasing urinary levels of CR and NANA displayed a positive association with increasing tumor size, strengthening a previously established link to altered tumor metabolism.

Conclusion and Impact: These replicated results provide evidence that identified urinary metabolite biomarkers have a potential utility as noninvasive, clinical screening tools for early diagnosis of lung cancer. *Cancer Epidemiol Biomarkers Prev; 25(6); 978–86. ©2016 AACR.*

Introduction

Lung cancer continues to be a significant burden in the United States and worldwide (1), with still over 40 million smokers in the United States, approximately 23% of men and 19% of women ages 35 to 64, and even higher percentages among lower socio-economic groups (2). A substantial challenge with this disease is that at present, more than 61% of patients are diagnosed in later stages III and IV, when therapeutic options are limited, and the 5-year survival is only 4%. In contrast, the 5-year survival of patients with early stage and localized lung cancer is about 50%, and is thus considerably improved compared with later stages (1). These exceptionally bleak statistics provide strong motivation to search for biomarkers that could aid currently established screening methods for early detection of lung cancer. To date, no such biomarkers are clinically utilized.

A seminal report by the National Lung Screening Trial (NLST) in 2011 showed for the first time a significant 20% reduction in lung cancer mortality from low-dose CT (LDCT) screening when compared with the chest X-ray (3). Since then, there have been a number of recommendations for screening high-risk populations of ages 55–79 years, with a 30-pack-year history of smoking (4–7), triggering coverage by Medicare and private health insurers under provisions of the Affordable Care Act (8). However, the major problem with LDCT is a high false-positive rate of 96%, which can lead to unnecessary downstream diagnostic procedures that may result in patient anxiety, higher health care spending, and increased mortality (3). Despite encouraging results of the NLST, there are concerns related to the cost benefit of LDCT for lung cancer screening, risks of radiation exposure, overdiagnosis (9), and expected epidemic rise in indeterminate pulmonary nodules (IPN) diagnoses, only some of which present a higher risk for developing lung cancer (10). The use of complementary and accurate biomarkers would be a useful strategy to ameliorate most of these issues. However, most of the available molecular

[1]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland. [2]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee. [3]Laboratory of Metabolism, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland. [4]Nippon Medical School, Tokyo, Japan. [5]Biomedical Informatics, The Ohio State University College of Medicine, Columbus, Ohio. [6]International Epidemiology Institute, Rockville, Maryland.

Corresponding Authors: Curtis C. Harris, Laboratory of Human Carcinogenesis, National Cancer Institute, Building 37, Room 3068, 37 Convent Drive, MSC 4258, Bethesda, MD 20892-4258. Phone: 301-496-2048; Fax: 301-496-0497; E-mail: Curtis_Harris@nih.gov; and William J. Blot, International Epidemiology Institute, 1455 Research Blvd, Suite 550, Rockville, MD 20850. Phone: 301-424-1054; E-mail: blotw@iei.us

AACR

factors incorporated in diagnostic algorithms that guide therapy decisions are established by using invasively collected tissue specimens (11, 12). Noninvasively collected biospecimens, such as urine and blood, are gaining a significant interest as methods for measuring robust biomarkers of risk and prognosis (13–15).

While several circulating biomarkers have been evaluated in association with lung cancer risk (16–23), a large majority are tobacco-related carcinogens (24–27), which are markers of exposure and do not present targets for potential therapeutic interventions. Other studies have evaluated demographic and smoking-related behavioral information to aid in the selection criteria for lung cancer screening, much of which is self-reported (28). Nevertheless, to increase the specificity of LDCT, the ultimate goal is to identify stable, reproducible, and noninvasively measured biomarkers. If said biomarkers are also products of deregulated tumor metabolism, they may have an ability to distinguish benign from malignant nodules. Despite intensive research and major advances in the field, there are still a number of challenges before biomarker panels can be used routinely in clinical practice, the most important of which is a necessity to validate them. The National Cancer Institute (NCI) previously created the Early Detection Research Network to promote biomarker discovery, validation, and translation into clinical practice (29). To fully assess and develop reliable biomarkers for monitoring asymptomatic patients for lung cancer, there is an urgent need to evaluate their utility in prediagnostic samples.

To that extent, we have evaluated a previously identified panel of urinary metabolite lung cancer biomarkers (30) in a well-characterized prospective Southern Community Cohort Study (SCCS): creatine riboside (CR), N-acetylneuraminic acid (NANA), cortisol sulfate, and a glucuronidated, indeterminate metabolite referred to as 561+. Previously, novel CR and NANA were found to be overexpressed in tumor when compared with the adjacent nontumor tissue, with levels positively correlated between the tissue and urine (30). Therefore, if their association with lung cancer is validated, these metabolites present promising molecular markers for clinical evaluation as a method to aid LDCT, especially considering that metabolites are more proximal to disease phenotype than other "omics" markers.

Materials and Methods

Study subjects

With recruitment completed in 2009, the Southern Community Cohort Study (SCCS) comprises a population of adults ages 40–79 years who reside in the southeastern United States and belong to historically underrepresented groups in major cancer research studies (e.g., African-Americans; ref. 31), with a high prevalence of behaviors associated with increased disease risk (e.g., smoking). Key strengths of this cohort are that subjects come from similar socioeconomic backgrounds (lower compared with other established cohorts), thereby diminishing this factor as a potential confounder, and an established Biospecimen Repository, containing frozen stored blood, buccal cells, and urine.

Detailed procedures for data linkage, processing and quality control have been established with the 12 state cancer registries covering the SCCS catchment area, providing the primary means of identifying incident cancer diagnoses. Although reporting lags are common, the registries provide nearly complete and unbiased ascertainment of cancers diagnosed among the participants after their entry into the SCCS. Three of the registries are SEER registries,

and all 12 registries have North American Association of Central Cancer Registries gold- ($N= 11$) or silver-level ($N= 1$) certification. Cohort member deaths are identified through annual linkages with both the National Death Index (NDI) and Social Security Administration (SSA), well-established and reliable means of identifying deaths in the United States (32, 33). In addition, social security numbers (an ideal linking criterion) were collected from 97% of SCCS participants. Informed consent was obtained from all study participants. This study was approved by the Institutional Review Boards of the involved institutions.

At the commencement of the study, there were 854 incident lung cancer cases and 178 of them had available urine samples for the metabolomics analysis, thereby forming the case group for the nested case–control analysis. Corresponding individually matched controls with stored urine specimens were randomly selected in a 2:1 ratio to cases using incidence density sampling among cohort members matched by age (± 2 years), sex, race, CHC recruitment site, menopausal status (women), and date of sample collection (± 6 months), the latter to ensure similar specimen storage durations (Table 1). Of note, we conducted an analysis to compare the main variables used for matching or adjustment of analyses between the cases who had and those who did not have urine samples. We found no significant differences

Table 1. Demographic and clinical characteristics of the nested SCCS case–control sample set

Characteristics	Cases	Controls	Total	P^a
N	178	351	529	0.83
Age; mean $\pm$ SD	57.7 $\pm$ 8.6	57.3 $\pm$ 8.5	57.4 $\pm$ 8.6	
Race; n(%)				
African-American	96 (54)	186 (53)	282 (53)	0.92
European-American	74 (42)	146 (42)	220 (42)	
Other	8 (4)	12 (3)	20 (4)	
Refused		1 (0)	1 (0)	
Missing		6 (2)	6 (1)	
Gender; n(%)				
Male	101 (57)	194 (56)	295 (56)	0.88
Female	77 (43)	152 (44)	229 (44)	
Smoking status; n(%)				
Current	127 (73)	140 (42)	267 (52)	<0.0001
Former	39 (23)	99 (29)	138 (27)	
Never	7 (4)	97 (29)	104 (20)	
Histology; n(%)				
Adenocarcinoma	59 (33)			
Squamous cell carcinoma	36 (20)			
Non-small cell carcinoma	19 (11)			
Small cell carcinoma	29 (16)			
Large cell carcinoma	9 (5)			
Other[b]	13 (7)			
Unknown	13 (7)			
Stage[c]; n(%)				
Stage occult	2 (1)			
I	16 (9)			
II	13 (7)			
III	41 (23)			
IV	77 (43)			
Unknown	29 (16)			

[a]Two-sided χ^2 test.
[b]Other includes: adenosquamous carcinoma, carcinoid tumor malignant, carcinoma, neoplasm, and neuroendocrine carcinoma.
[c]Only cases identified through cancer registry contain staging information. The original staging was based on the 6th ($N = 115$) and 7th ($N = 50$) edition of the AJCC staging manual. Staging presented in the table is based on the 7th edition, wherein all cases with sufficient tumor size, metastases status, and lymph node involvement data previously staged based on the 6th edition were restaged based on the 7th edition for consistency (N total 7th edition staging = 149).

except in the race distribution ($P = 0.03$), with a higher number of African-American subjects in those without urine samples when compared with those with (65%, vs. 54%), and self-reported prior history of chronic obstructive pulmonary disease (COPD; $P = 0.05$). Therefore, we believe that the results presented herein are generalizable to the entire SCCS cohort.

Sample processing

SCCS participants were asked to provide urine samples starting in November of 2004. Participants were not required to fast before giving the samples. One-time urine specimens were collected from participants at Community Health Centers, refrigerated, and shipped overnight to Vanderbilt University (Nashville, TN) for processing and frozen storage at $-80°C$. Participants donated approximately 60 mL of urine which was mixed with a small amount of ascorbic acid as a preservative. Samples were deproteinated using 50% acetonitrile containing 10 μmol/L chloropropamide as internal standard. Supernatants were transferred into 96-well sample plates. All pipetting and dilutions were performed using a MICROLAB STARLET automated liquid handler (Hamilton Robotics), with microcentrifugation steps performed manually ($14,000 \times g$ for 15 minutes at $4°C$).

Ultra performance liquid chromatography-mass spectrometry

Mass spectrometry (MS) was performed on a XEVO G2 ESI QTOF mass spectrometer operating in electrospray ionization (ESI) positive [monitoring CR, mass to charge ratio (m/z) = 264.1196 and retention time (RT) = 0.4 minutes, and 561+, m/z = 561.3435 and RT = 6.3] and negative modes (monitoring NANA, m/z = 308.0982, RT = 0.4 minutes, and cortisol sulfate, m/z = 441.162, RT = 5.5). Detailed methods have previously been reported (30), and are given in some detail in the Supplementary Methods.

Statistical analyses

A χ^2 test was performed on the dichotomized levels of CR and NANA across smoking, gender, and race characteristics, to test differences in levels across these strata. Metabolite abundances were dichotomized into high (>75th percentile) and low (≤75th percentile) groups based on the distribution of abundances in the controls from the current study population. Test for trend was performed to assess trends in metabolite levels across the tumor size categories (34).

Conditional logistic regression analyses were performed to calculate ORs and accompanying 95% confidence intervals (CIs) for lung cancer associated with high metabolite levels (normalized area under the peak of corresponding chromatographic peaks). In this hypothesis-testing analysis, type I error rates of 0.05 were applied in assessing the significance of each relationship. Covariate adjustments were made by adding terms for age at enrollment, body mass index (BMI), cigarette smoking status (current/former/never), amount smoked among smokers (pack-years), education and income level, prior history of COPD, and family history of lung cancer (yes/no/don't know). Interaction analysis to test whether race modifies observed associations was conducted using likelihood ratio test.

SCCS mortality follow-up data was used to conduct survival analyses for death due to lung cancer among cases in relation to categorical variables of dichotomized metabolite abundances using Cox regression modeling. Multivariate models were computed by adjusting for cigarette smoking status (current/former/never), amount smoked among smokers (pack-years), age at

diagnosis, sex, race, stage (7th edition of the AJCC staging manual), and histology.

Subgroup analyses were performed by stratifying on race (European- vs. African-American subjects), and on reported preference for smoking menthol cigarettes (yes vs. no). Sensitivity analyses were performed by removing cases diagnosed within two years of cohort enrollment ($n = 55$), by removing cases and controls who had reported a previous cancer diagnosis (any tumor type; $n = 60$), and by excluding those individuals who did not answer the question regarding the family history of lung cancer ($n = 91$).

In exploratory analyses, we used ROC to assess the predictive value of identified metabolites in lung cancer diagnosis using the roccomp function. Models were built using conditional logistic regression on the continuous abundances of CR and NANA, as well as on the demographic, smoking-related and behavioral risk factors, and the combinations thereof. Difference between the predictive abilities of the model containing CR and NANA, and model without the metabolites was tested using rocreg function.

Mann–Whitney (two-sample Wilcoxon rank-sum) test was used to test differences in levels of urinary menthol glucuronide between those who reported and those who did not report a preference for smoking menthol cigarettes.

All reported P values are two-sided, and all P values less than or equal to 0.05 were considered statistically significant. All analyses were conducted in STATA (Stata Statistical Software Release 13.1).

Results

Risk

CR, NANA, cortisol sulfate, and 561+, previously identified as diagnostic and prognostic biomarkers in a case–control study (30) were assessed for their relationship with lung cancer risk in prospectively collected urine specimens from the SCCS cohort. Of note, these specimens were collected before clinically detectable disease, and relevant demographic and clinical characteristics are given in Table 1. Out of four metabolites, only CR was associated with lung cancer in the overall case–control set, while NANA showed borderline significance; cortisol sulfate and 561+ were not associated with either risk or survival (Supplementary Tables S1 and S2). Therefore, focus of the current study is on metabolites CR and NANA.

Initially, potential correlates of CR and NANA were assessed in the control subjects alone, as the presence of the disease may introduce confounding. As illustrated in Table 2, high levels of CR and NANA (dichotomized on the basis of the 75th percentile of control abundances) are significantly higher in current- when compared with former- and never-smokers. Furthermore, high levels of NANA but not CR are more frequent in former- compared with never-smokers. However, correlation of CR and NANA levels with reported pack-years (measure of amount smoked) in former- and current smokers is not observed (Supplementary Fig. S1), indicating that perhaps the amount smoked does not influence the metabolite levels, consistent with the previous report (30). In addition, high levels of CR are significantly more elevated in males when compared with females, and in African-American when compared with European-American controls. These associations were not seen for NANA. Taking into account possible confounders, all subsequent analyses are adjusted for smoking status, race, and other potential confounding factors.

Table 2. Associations of smoking, gender, and race with high and low levels of (A) creatine riboside and (B) NANA in controls only

A				
Creatine riboside	**High *n* (%)**	**Low *n* (%)**	**χ^2**	**P**
Smoking				
Current	53 (37.9)	87 (62.1)	21.2	**<0.0001**
Former	15 (15.2)	84 (84.9)		
Never	16 (16.5)	81 (83.5)		
Gender				
Male	60 (30.9)	134 (69.1)	8.7	**0.003**
Female	26 (17.1)	126 (82.9)		
Race				
European Americans	21 (14.4)	125 (85.6)	16.4	**<0.0001**
African Americans	63 (33.9)	123 (66.1)		

B				
NANA	**High *n* (%)**	**Low *n* (%)**	**χ^2**	**P**
Smoking				
Current	43 (30.7)	97 (69.3)	6.2	**0.05**
Former	26 (26.3)	73 (73.7)		
Never	16 (16.5)	81 (83.5)		
Gender				
Male	51 (26.3)	143 (73.7)	0.71	0.40
Female	34 (22.4)	118 (77.6)		
Race				
European Americans	35 (24.0)	111 (76.0)	0.15	0.70
African Americans	48 (25.8)	138 (74.2)		

NOTE: Levels were dichotomized into high and low based on the 75th percentile of the control abundances. Bold text designates a significant P value.

Next, conditional logistic regression analysis was conducted. High levels of CR are associated with increased lung cancer risk after adjusting for age at enrollment, BMI, smoking status, pack-years, income, education level, prior history of COPD, family history of lung cancer ($OR_{adjusted} = 2.0$; 95% CI = 1.2–3.4; $P = 0.01$), while NANA displays borderline significance ($OR_{adjusted} = 1.6$; 95% CI = 1.0–2.6; $P = 0.08$; Table 3). Of note, we conducted a sensitivity analysis wherein we removed those

subjects who did not answer the question regarding family history of lung cancer. This analysis indicates that both, CR and NANA are significantly associated with lung cancer risk in the overall cohort (overall: $OR_{adjusted} = 1.9$; 95% CI, 1.0–3.6; $P = 0.4$ and $OR_{adjusted} = 2.0$; 95% CI, 1.1–3.6; $P = 0.02$, respectively; Supplementary Table S3). Stratification by race suggests that these associations are stronger in European-Americans ($OR_{adjusted} = 5.3$; 95% CI = 1.6–17.6 and 3.5; 95% CI = 1.5–8.4, respectively; $P \leq 0.004$) when compared with African-Americans ($OR_{adjusted} = 1.1$; 95% CI, 0.6–2.3 and 0.9; 95% CI = 0.4–1.8, respectively; $P = $ ns; Table 3), although the likelihood ratio tests of significance for differing ORs by race are not significant ($P_{interaction} = 0.25$ for CR and 0.11 for NANA). Furthermore, quartile analysis indicates that CR levels are significantly associated with lung cancer risk in the highest when compared with the lowest quartile ($OR_{adjusted} = 2.0$; 95% CI, 1.1–3.8; $P = 0.03$; Supplementary Table S4).

CR and NANA were previously shown as elevated in the stage I tumor when compared with the adjacent nontumor tissue, linking them directly to altered tumor metabolism (30). Hence it is plausible that these endogenous metabolites are involved with early processes contributing to neoplastic transformation, such as inflammation. To test this hypothesis, we evaluated whether the levels of the markers are predictive of lung cancer risk two or more years before diagnosis by excluding the cases diagnosed within two years of the cohort enrollment (Table 4). While the ORs remain in the same direction for both CR and NANA, the associations are attenuated and no longer significant in the overall cohort after the adjustment for a number of putative confounders, possibly due to insufficiently powered analysis. However, the associations with high CR and NANA remain significant in European-Americans when individuals diagnosed within two years of cohort enrollment are removed ($OR_{adjusted} = 6.7$; 95% CI, 1.6–27.6 and 3.8; 95% CI, 1.3–11.5, respectively; $P \leq 0.02$; Table 4). Next, aware of a possibility that CR and NANA

Table 3. Conditional logistic regression analysis of creatine riboside and NANA in all, European-, and African-American subjects

	Univariate				Multivariate			
Metabolite[a]	**N (%) Cases**	**N (%) Controls**	**OR (95% CI)**	**P^b**	**N (%) Cases**	**N (%) Controls**	**OR (95% CI)[c]**	**P^b**
All								
Low								
Referent_cr	109 (61)	262 (75)	1.00		101 (60)	237 (25)	1.00	
Referent_nana	109 (61)	262 (75)	1.00		101 (60)	231 (74)	1.00	
High								
Creatine riboside	69 (39)	87 (25)	**2.0 (1.3–3.0)**	**0.001**	66 (40)	77 (25)	**2.0 (1.2–3.4)**	**0.01**
NANA	69 (39)	87 (25)	**2.0 (1.3–3.0)**	**0.001**	66 (40)	83 (26)	1.6 (1.0–2.6)	0.08
European-Americans								
Low								
Referent_cr	51 (69)	123 (85)	1.00		49 (68)	116 (87)	1.00	
Referent_nana	40 (54)	109 (76)	1.00		38 (53)	99 (74)	1.00	
High								
Creatine riboside	23 (31)	21 (15)	**2.7 (1.3–5.4)**	**0.005**	23 (32)	18 (13)	**5.3 (1.6–17.6)**	**0.006**
NANA	34 (46)	35 (24)	**2.8 (1.5–5.2)**	**0.001**	34 (47)	35 (26)	**3.5 (1.5–8.4)**	**0.004**
African-Americans								
Low								
Referent_cr	52 (54)	122 (66)	1.00		48 (53)	113 (66)	1.00	
Referent_nana	64 (67)	137 (74)	1.00		59 (66)	124 (73)	1.00	
High								
Creatine riboside	44 (46)	63 (34)	1.6 (1.0–2.7)	0.06	42 (47)	58 (34)	1.1 (0.6–2.3)	0.73
NANA	32 (33)	48 (26)	1.4 (0.8–2.4)	0.20	31 (34)	47 (27)	0.9 (0.4–1.8)	0.69

NOTE: Bold text designates significant results.
[a]Levels dichotomized into high and low based on the 75th percentile of population control abundances (low = referent).
[b]Statistically significant; P-value < 0.05.
[c]Multivariate conditional logistic regression adjusted for age, BMI, income, education level, prior history of COPD, family history of lung cancer, smoking status, and pack years. Individually matched controls selected in 2:1 ratio to cases using incidence density sampling matched by age (±2 years), sex, race, CHC recruitment site, menopausal status (women), and date of sample collection (±6 months).

Haznadar et al.

Table 4. Conditional logistic regression analysis of creatine riboside and NANA in all, European-, and African-American subjects after removal of cases diagnosed within two years ($N = 55$) of cohort enrollment

Metabolite[a]	Univariate				Multivariate			
	N (%) Cases	N (%) Controls	OR (95% CI)	P[b]	N (%) Cases	N (%) Controls	OR (95% CI)[c]	P[b]
All								
Low								
Referent_cr	79 (64)	178 (74)	1.00		73 (64)	163 (74)	1.00	
Referent_nana	76 (62)	183 (76)	1.00		70 (61)	163 (74)	1.00	
High								
Creatine riboside	44 (36)	62 (26)	**1.7 (1.0–2.7)**	**0.04**	42 (36)	56 (26)	1.6 (0.9–3.1)	0.14
NANA	22 (45)	23 (24)	**2.2 (1.3–3.6)**	**0.003**	45 (39)	56 (26)	1.7 (0.9–3.1)	0.10
European-Americans								
Low								
Referent_cr	33 (67)	82 (85)	1.00		31 (66)	77 (87)	1.00	
Referent_nana	27 (55)	73 (76)	1.00		25 (53)	66 (74)	1.00	
High								
Creatine riboside	16 (33)	14 (15)	**3.0 (1.3–7.3)**	**0.01**	16 (34)	12 (13)	**6.7 (1.6–27.6)**	**0.008**
NANA	22 (45)	23 (24)	**3.1 (1.4–7.0)**	**0.007**	22 (47)	23 (26)	**3.8 (1.3–11.5)**	**0.02**
African-Americans								
Low								
Referent_cr	44 (62)	90 (65)	1.00		40 (61)	83 (65)	1.00	
Referent_nana	47 (66)	104 (75)	1.00		43 (65)	94 (74)	1.00	
High								
Creatine riboside	27 (38)	48 (35)	1.2 (0.6–2.1)	0.64	26 (39)	44 (35)	0.6 (0.3–1.5)	0.31
NANA	24 (34)	34 (25)	1.6 (0.8–3.1)	0.15	23 (35)	33 (26)	1.0 (0.4–2.5)	0.98

NOTE: Bold text designates significant results.
[a]Levels dichotomized into high and low based on the 75th percentile of population control abundances (low = referent).
[b]Statistically significant; P-value < 0.05.
[c]Multivariate conditional logistic regression adjusted for age, BMI, income, education level, prior history of COPD, family history of lung cancer, smoking status, and pack years. Individually matched controls selected in 2:1 ratio to cases using incidence density sampling matched by age ($\pm$ 2 years), sex, race, CHC recruitment site, menopausal status (women), and date of sample collection ($\pm$6 months).

may be associated with other cancer types, sensitivity analysis was performed wherein the cases ($N = 28$) and controls ($N = 32$) who had reported a previous cancer diagnosis at the time of the enrollment were removed. The ORs associated with high CR remain significantly elevated (OR_adjusted = 2.0; 95% CI, 1.1–3.6; $P = 0.03$), while NANA is only significant in the unadjusted model, but the effect size remains similar after the adjustment (OR_adjusted = 1.4; 95% CI, 0.8–2.5; $P = 0.28$). Both metabolites are significantly associated with lung cancer risk in European-Americans (OR_adjusted = 19.3; 95% CI, 2.6–142.7) and 5.4 (95% CI, 1.5–20.1), respectively; $P \leq 0.01$; Supplementary Table S5).

Finally, to evaluate the ability of CR and NANA alone and in combination to classify lung cancer in this nested case–control set, conditional logistic regression models and ROC analysis were performed. A model containing variables previously found to be most robust in selecting high-risk individuals for downstream screening comprising smoking status, pack years, age, BMI, income and education levels, previous history of COPD, and family history of lung cancer (28), was compared with a model also comprising CR and NANA, resulting in a significant improvement indicated by the increased AUC (0.78 compared with 0.80 after the addition of the metabolites, $P = 0.01$; Fig. 1A). Of note, a selected cut-off point with the most robust ability to correctly classify subjects, at 50% sensitivity and 86% specificity, leads to a positive predictive value (PPV) of 63%, a negative predictive value (NPV) of 79%, and correct classification of 75% subjects. We assessed the ability of the models to classify lung cancer cases in European-Americans also, due to observed stronger associations in this racial group. We detected a significant improvement in the AUC after the addition of CR and NANA to the model containing risk factors alone (from 0.84 to 0.90, respectively; $P = 0.002$), with a selected cut-off point leading to a correct classification of 84% of subjects (PPV = 75%; NPV = 89%; Fig. 1B).

We have previously shown that a preference for smoking menthol cigarettes was no more harmful than smoking non-menthol cigarettes, as assessed in a larger nested case–control set in the SCCS (35). Considering that urinary menthol glucuronide may be a biomarker for smoking menthol cigarettes, the association between the levels of menthol glucuronide and lung cancer risk was investigated. Menthol glucuronide was dichotomized into high and low categories based on the 75th percentile of the control abundances. No significant association between urinary menthol glucuronide and lung cancer risk were observed in either all subjects, or across the racial strata, after adjustment for amount smoked (Supplementary Table S6). In concordance with our previous report (35), we also did not observe a significant association between a preference for smoking menthol cigarettes and lung cancer risk in this smaller nested case–control set (Supplementary Table S7). Of note, a strong association between the reported preference for smoking menthol cigarettes and urinary menthol glucuronide levels was observed [fold change (FC) = 1.9; $P < 0.0001$ in current- and former-smokers (Supplementary Fig. S2A), and FC = 2.5; $P < 0.0001$ in current-smokers only (Supplementary Fig. S2B)], with similar associations observed in European- and African-Americans.

Survival

On the basis of the previous data from the NCI-MD case–control study that showed high levels of CR and NANA to be associated with worse survival (30), the same analysis was performed in this study. Cox regression analysis adjusted for age at diagnosis, smoking status, pack-years, stage (the 7th edition of the AJCC staging manual) and histology showed no significant associations (Supplementary Table S8; Supplementary Fig. S3). This finding suggests that CR and NANA levels prior to clinically detectable disease are not associated with survival.
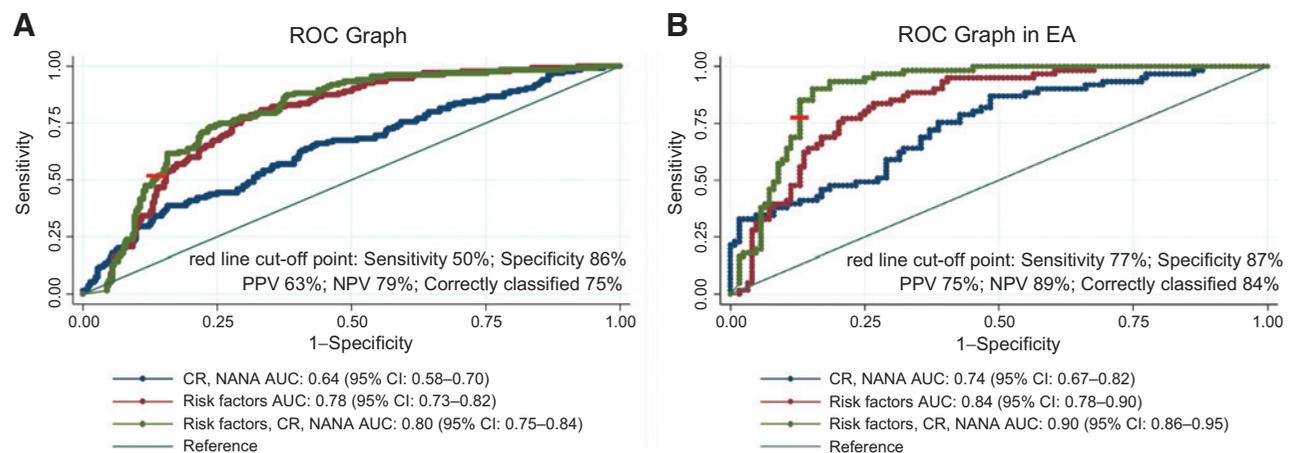
Figure 1.
ROC analysis of selected demographic, smoking, and behavioral characteristics associated with lung cancer risk, CR, NANA, and combinations thereof in all subjects (A), and European-Americans (EA) (B). CR, creatine riboside; NANA, N-acetylneuraminic acid; PPV, positive predictive value; NPV, negative predictive value. Risk factors model comprises: age (age at enrollment), education (achieved level of education), income (household income), smoking [smoking status (current-, former-, never-smoker)], pack-years (amount smoked indicated by pack-years), BMI (body mass index), COPD (history of chronic obstructive pulmonary disease), and family history of lung cancer.

No significant associations between either urinary menthol glucuronide (Supplementary Table S9), nor reported preference for smoking menthol cigarettes (Supplementary Table S10) and lung cancer survival were found in this study.

Levels across tumor sizes

Considering previous evidence that CR and NANA are tumor-specific metabolites linked to altered tumor metabolism (30), it is plausible that their levels may increase with increasing tumor size. For a subset of cases with exact tumor size available ($N = 108$), the prevalence of high CR rose with increasing tumor size ($P_{trend} = 0.03$, Fig. 2A). While the results are not significant for NANA, a similar trend was observed across the tumor size strata (Fig. 2B).

Discussion

Currently, there are no biomarkers that can successfully complement available screening modalities for detecting early-stage lung cancer. LDCT has been proven to reduce lung cancer mortality by 20%, as compared with chest X-ray (3). While no lesion goes undetected by LDCT, most are nonmalignant, leading to an astonishingly high false discovery rate of 96%. A panel comprising four metabolites detected in the urine was found to be most significantly associated with both lung cancer status and survival, and reproducible and stable in stored urine over a long period of time (30). The goal of the current study was to investigate whether the four metabolites, creatine riboside (CR), N-acetylneuraminic acid (NANA), cortisol sulfate, and indeterminate compound designated as 561+, were predictive of lung cancer risk in prediagnostic urine specimens, one of the crucial steps for validating promising biomarkers that may have clinical utility (29). Of these four, only CR was predictive of lung cancer risk before clinically detectable disease, with NANA displaying borderline significance. In our earlier work, we did not detect elevated levels of cortisol sulfate and 561+ in tumor tissue, which also tended to lower the importance of the associations in urine. CR and NANA, however,

were significantly elevated in tumor when compared with matched, adjacent nontumor lung tissue, strengthening their importance as markers of lung cancer risk. NANA has previously been characterized and is the most common form of sialic acid that plays several roles in biology, and is thought to protect cancer cells from immunosurveilance (36). Creatine riboside is a novel metabolite previously reported for the first time by our group (30). While the characterization of this metabolite is of special importance for future studies, we speculate that CR may be a product of both, high creatine within the tumor, and high phosphate flux, as a result of a pronounced energy requirement of fast dividing tumor when compared with normal cells (37). Future studies should elucidate what role, if any, CR may play in neoplastic transformation.

While we were unable to conduct an adjusted conditional logistic analysis between CR and NANA and early stage I-II lung cancer due to a small number of patients in this group, univariate analysis suggest that these metabolites may be robust in predicting early stage lung cancer (data not shown), consistent with our previous report. The utility of these metabolites in aiding early lung cancer diagnosis remains to be established by investigating their usefulness in distinguishing benign from malignant nodules detected by LDCT.

Although the levels of the two metabolites were higher among current than nonsmokers (and levels of NANA but not CR were higher among former than never smokers), the associations between CR and lung cancer risk held after adjusting for cigarette smoking status and amount smoked. In both this and in the previous report, we did not observe any correlations between the metabolites and the amount smoked (cigarettes per day and pack years, respectively). Furthermore, a previous study showed that amino acid and lipid pathways are most significantly affected by the exposure to tobacco-smoke and reversible after smoking cessation; CR and NANA, however, are not members of these pathways (38). Thus, the levels of these metabolites may not be affected by immediate exposure to tobacco smoke, and they are not likely markers of exposure. While NANA displayed borderline
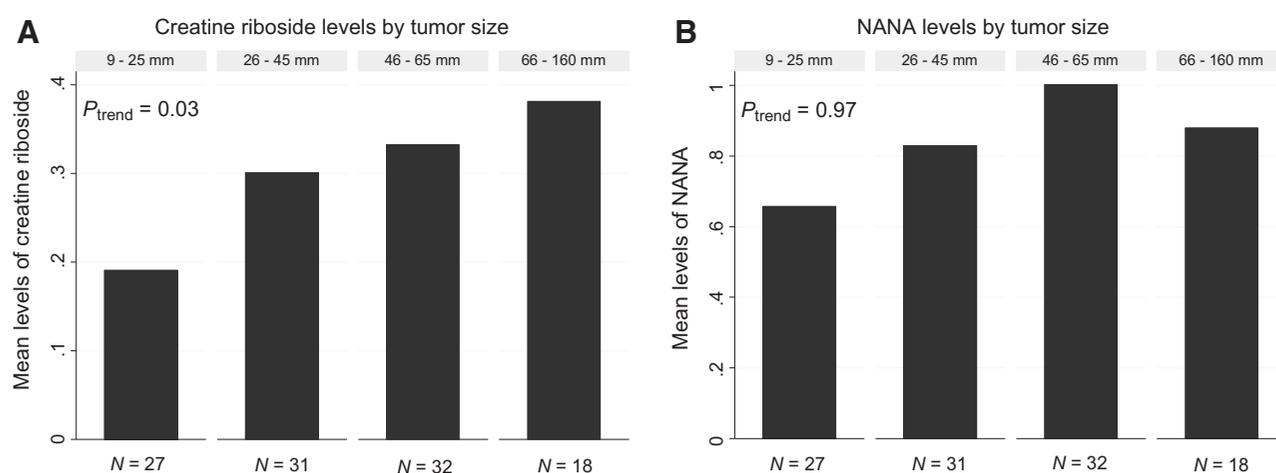
Haznadar et al.

Figure 2.
Mean levels of creatine riboside (A) and NANA (B) in SCCS lung cancer cases across tumor size strata.

significance after adjustment in the overall cohort, the levels of both metabolites were significantly associated with lung cancer risk in European-Americans. Although race did not seem to significantly modify the effect of the associations between metabolites and lung cancer risk, the associations were stronger in European- than African-Americans. The lack of observed formal interaction may be a result of decreased power, or factors associated with race (such as lifestyle), which could not be accounted for in this study. In our previous case–control study, racial differences did not attenuate associations with lung cancer status, as disease presence may override any other contributing factors, such as genetic backgrounds. Future investigation may elucidate whether or to what extent racial differences in the association with lung cancer risk may exist. It is possible that genetic variants may contribute to racial differences, as they play a role in the metabolism of both exogenous (25) and endogenous metabolites (39, 40). In addition, lifestyle and external exposures may exacerbate the observed differences, factors that could not be controlled in this study.

To address other causes that may contribute to health disparities in lung cancer, a preference for smoking menthol cigarettes, which is much higher among African- than European-Americans (41–43) was investigated. We have previously reported that smoking menthol cigarettes, preferred by African-Americans, does not lead to an increased risk for developing lung cancer in the SCCS (35). In the current study, we took an additional step and measured urinary menthol glucuronide as a biomarker of exposure to menthol from tobacco smoke. Consistent with the previous report, no significant associations were observed in relationship with lung cancer risk. We are the first group to our knowledge to report this observation.

Metabolic markers have gained increased interest recently, as they are recognized as proximal markers to the disease phenotype in comparison to other "omics" markers (44–46). The footprint that products of altered tumor metabolism may leave in noninvasively collected biospecimens serves as a great opportunity for biomarker discovery. This "liquid biopsy" practice would be a high-throughput, noninvasive and inexpensive approach to allow not only cancer detection, but also patient monitoring during treatment, and would provide an

ideal therapeutic strategy for precision medicine (47). There is a great necessity to identify and validate a complementary tool to the solid biopsy, which may also allow for delivery of targeted therapy. Using very precise mass spectrometry for measuring lung cancer risk markers, such as CR and NANA, presents a feasible opportunity to develop reproducible, inexpensive and high-throughput diagnostic tools to refine algorithms for early detection of lung cancer.

Caveats of the study include a relatively limited sample size, as well as the inability to control for the exogenous effects on metabolism. We adjusted for the effects of smoking, although residual confounding cannot be ruled out, but other unmeasured factors may have contributed to the differences in CR and NANA levels between those with and without lung cancer. A major strength of the study is a well-matched nested case–control sample set with its close matching on demographic variables and in methods in and timing of collection of urine samples, as well as similar socioeconomic backgrounds of the subjects enrolled in the cohort.

Overall, the results of this study suggest that CR and NANA are associated with lung cancer risk in prospective samples, more strongly in European-Americans, and therefore may have clinical utility for disease screening or diagnosis. Considering an expected forthcoming rise in the detection of indeterminate pulmonary nodules as a result of LDCT screening, it is crucial that noninvasive biomarkers be developed, with ability to distinguish malignant from indolent nodules. Future biomarker studies in LDCT screening trials are needed to determine the clinical utility of these and other biomarkers of NSCLC.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin 2015;65:5–29.
2. Fedewa SA, Sauer AG, Siegel RL, Jemal A. Prevalence of major risk factors and use of screening tests for cancer in the United States. Cancer Epidemiol Biomarkers Prev 2015;24:637–52.
3. National Lung Screening Trial Research, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395–409.
4. Jaklitsch MT, Jacobson FL, Austin JH, Field JK, Jett JR, Keshavjee S, et al. The American Association for Thoracic Surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups. J Thorac Cardiovasc Surg 2012;144: 33–8.
5. Boiselle PM, Chiles C, Patz E, Tammemagi M, Wood DE. Expert opinion: United States Preventive Services Task Force recommendation on screening for lung cancer. J Thorac Imaging 2014;29:197.
6. Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. Ann Intern Med 2013;159:411–20.
7. de Koning HJ, Meza R, Plevritis SK, ten Haaf K, Munshi VN, Jeon J, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. Ann Intern Med 2014;160:311–20.
8. Wood DE. The importance of lung cancer screening with low-dose computed tomography for Medicare beneficiaries. JAMA Int Med 2014;174: 2016–8.
9. Gulati S, Mulshine JL. Lung cancer screening guidelines: common ground and differences. Translat lung cancer res 2014;3:131–8.
10. Pinsky PF, Nath PH, Gierada DS, Sonavane S, Szabo E. Short- and long-term lung cancer risk associated with noncalcified nodules observed on low-dose CT. Cancer Prev Res 2014;7:1179–85.
11. Devesa SS, Bray F, Vizcaino AP, Parkin DM. International lung cancer trends by histologic type: male:female differences diminishing and adenocarcinoma rates rising. Int J Cancer 2005;117:294–9.
12. Reck M, Heigener DF, Mok T, Soria JC, Rabe KF. Management of non-small-cell lung cancer: recent developments. Lancet 2013;382:709–19.
13. Hassanein M, Callison JC, Callaway-Lane C, Aldrich MC, Grogan EL, Massion PP. The state of molecular biomarkers for the early detection of lung cancer. Cancer Prev Res 2012;5:992–1006.
14. Gogalic S, Sauer U, Doppler S, Preininger C. Bladder cancer biomarker array to detect aberrant levels of proteins in urine. Analyst 2015;140:724–35.
15. Debmalya Barh AC, Verma Mukesh, Gunduz Mehmet. Cancer biomarkers: non-invasive early diagnosis and prognosis. 1 ed: Boca Raton, FL: Taylor & Francis LLC; 2014.
16. Esmaeilsabzali H, Beischlag TV, Cox ME, Parameswaran AM, Park EJ. Detection and isolation of circulating tumor cells: principles and methods. Biotechnol Adv 2013;31:1063–84.
17. Friedlander TW, Premasekharan G, Paris PL. Looking back, to the future of circulating tumor cells. Pharmacol Therapeut 2014;142:271–80.
18. Han Y, Su C, Liu Z. Methods for detection of circulating cells in non-small cell lung cancer. Front Biosci 2014;19:896–903.
19. Pine SR, Mechanic LE, Ambs S, Bowman ED, Chanock SJ, Loffredo C, et al. Lung cancer survival and functional polymorphisms in MBL2, an innate-immunity gene. J Natl Cancer Inst 2007;99:1401–9.
20. Pine SR, Mechanic LE, Enewold L, Chaturvedi AK, Katki HA, Zheng YL, et al. Increased levels of circulating interleukin 6, interleukin 8, C-reactive protein, and risk of lung cancer. J Natl Cancer Inst 2011;103:1112–22.
21. Ronald JA, Chuang HY, Dragulescu-Andrasi A, Hori SS, Gambhir SS. Detecting cancers through tumor-activatable minicircles that lead to a detectable blood biomarker. Proc Natl Acad Sci U S A 2015;112:3068–73.
22. Wen CP, Zhang F, Liang D, Wen C, Gu J, Skinner H, et al. The ability of bilirubin in identifying smokers with higher risk of lung cancer: a large cohort study in conjunction with global metabolomic profiling. Clin Cancer Res 2015;21:193–200.
23. Nolen BM, Lomakin A, Marrangoni A, Velikokhatnaya L, Prosser D, Lokshin AE. Urinary protein biomarkers in the early detection of lung cancer. Cancer Prev Res 2015;8:111–9.
24. Boffetta P, Clark S, Shen M, Gislefoss R, Peto R, Andersen A. Serum cotinine level as predictor of lung cancer risk. Cancer Epidemiol Biomarkers Prev 2006;15:1184–8.
25. Park SL, Carmella SG, Ming X, Vielguth E, Stram DO, Le Marchand L, 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;24:561–9.
26. Yuan JM, Butler LM, Gao YT, Murphy SE, Carmella SG, Wang R, et al. Urinary metabolites of a polycyclic aromatic hydrocarbon and volatile organic compounds in relation to lung cancer development in lifelong never smokers in the Shanghai Cohort Study. Carcinogenesis 2014;35: 339–45.
27. Yuan JM, Koh WP, Murphy SE, Fan Y, Wang R, Carmella SG, et al. Urinary levels of tobacco-specific nitrosamine metabolites in relation to lung cancer development in two prospective cohorts of cigarette smokers. Cancer Res 2009;69:2990–5.

28. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013; 368:728–36.

29. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–61.

30. Mathe EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, Bowman ED, et al. Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. Cancer Res 2014;74:3259–70.

31. Hargreaves MK AC, Blot WJ. Community health centers: their role in the treatment of minorities and in health disparities research. New York, NY: McGraw-Hill; 2006.

32. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major US mortality databases. Ann Epidemiol 2002;12:462–8.

33. Wojcik NC, Huebner WW, Jorgensen G. Strategies for using the national death index and the social security administration for death ascertainment in large occupational cohort mortality studies. Am J Epidemiol 2010;172:469–77.

34. Cuzick J. A Wilcoxon-type test for trend. Stat Med 1985;4:87–90.

35. Blot WJ, Cohen SS, Aldrich M, McLaughlin JK, Hargreaves MK, Signorello LB. Lung cancer risk among smokers of menthol cigarettes. J Natl Cancer Inst 2011;103:810–6.

36. Schauer R. Sialic acids: fascinating sugars in higher animals and man. Zoology 2004;107:49–64.

37. Warburg O. On the origin of cancer cells. Science 1956;123:309–14.

38. Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, et al. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. BMC Med 2013;11:60.

39. Setiawan VW, Haiman CA, Stanczyk FZ, Le Marchand L, Henderson BE. Racial/ethnic differences in postmenopausal endogenous hormones: the multiethnic cohort study. Cancer Epidemiol Biomarkers Prev 2006;15: 1849–55.

40. Patel MJ, Batch BC, Svetkey LP, Bain JR, Turer CB, Haynes C, et al. Race and sex differences in small-molecule metabolites and metabolic hormones in overweight and obese adults. OMICS 2013;17:627–35.

41. Benowitz NL, Dains KM, Dempsey D, Wilson M, Jacob P. Racial differences in the relationship between number of cigarettes smoked and nicotine and carcinogen exposure. Nicotine Tob Res 2011;13:772–83.

42. DeSantis C, Naishadham D, Jemal A. Cancer statistics for African Americans, 2013. CA Cancer J Clin 2013;63:151–66.

43. Jones MR, Apelberg BJ, Tellez-Plaza M, Samet JM, Navas-Acien A. Menthol cigarettes, race/ethnicity, and biomarkers of tobacco use in U.S. adults: the 1999–2010 National Health and Nutrition Examination Survey (NHANES). Cancer Epidemiol Biomarkers Prev 2013;22:224–32.

44. Wikoff W, Grapov D, Fahrmann J, DeFelice B, Rom W, Pass H, et al. Metabolomic markers of altered nucleotide metabolism in early stage adenocarcinoma. Cancer Prev Res 2015;8:410–8.

45. Lemjabbar-Alaoui H, McKinney A, Yang YW, Tran VM, Phillips JJ. Glycosylation alterations in lung and brain cancer. Adv Cancer Res 2015;126: 305–44.

46. Marien E, Meister M, Muley T, Fieuws S, Bordel S, Derua R, et al. Non-small cell lung cancer is characterized by dramatic changes in phospholipid profiles. Int J Cancer 2015;137:1539–48.

47. Alix-Panabieres C, Pantel K. Circulating tumor cells: liquid biopsy of cancer. Clin Chem 2013;59:110–8.

Cancer Epidemiology, Biomarkers & Prevention

Urinary Metabolite Risk Biomarkers of Lung Cancer: A Prospective Cohort Study

Majda Haznadar, Qiuyin Cai, Kristopher W. Krausz, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/1055-9965.EPI-15-1191
Supplementary Material	Access the most recent supplemental material at: http://cebp.aacrjournals.org/content/suppl/2016/03/24/1055-9965.EPI-15-1191.DC1

Cited articles	This article cites 45 articles, 15 of which you can access for free at: http://cebp.aacrjournals.org/content/25/6/978.full#ref-list-1
Citing articles	This article has been cited by 2 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/25/6/978.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/25/6/978. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.