Pre-Diagnostic Genome-Wide DNA Methylation in Blood and Risk of Bladder Cancer

Kristina Jordahl

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Parveen Bhatti, Chair

Amanda I Phipps

J Emily White

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

Pre-Diagnostic Genome-Wide DNA Methylation in Blood and Risk of Bladder Cancer

Kristina Jordahl

Chair of the Supervisory Committee:

Parveen Bhatti

Epidemiology

Background: Differential DNA methylation as measured in blood is a promising marker of bladder cancer susceptibility. In fact, established bladder cancer risk factors such as smoking and various germline genetic variants may promote carcinogenesis in bladder tissue through processes that are detectable as differential DNA methylation in blood.

Methods: Genome-wide methylation was measured in pre-diagnostic blood samples, using the Illumina Infinium HumanMethylation450 Bead Array, among 440 bladder cancer cases with the transitional cell carcinoma (TCC) subtype and 440 matched controls from the Women's Health Initiative (WHI) cohort. After normalization and probe filtering, methylation measurements at 361,184 CpG sites remained for each study participant. We used conditional logistic regression models adjusted for potential confounders and for cell type composition to test for associations between the methylation level at each remaining CpG site and bladder cancer risk (Chapter 1). For each of three smoking-associated CpG sites, we conducted a regression-based mediation analysis to assess whether current smoking affects bladder cancer risk through differential methylation at the CpG site, while accounting for the smoking-CpG interaction using the four-way decomposition approach (Chapter 2). Since there are four single nucleotide polymorphisms (SNPs) that are known to be associated with bladder cancer and are also associated with

proximal DNA methylation changes, we performed a regression-based, multiple-mediator mediation analysis for each SNP, which allowed us to determine whether the associated methylation changes at CpG sites are a path through which the SNP causes bladder cancer (Chapter 3).

Results: Increased methylation at cg22748573, located in a CpG island within the 5'-UTR/first exon of the *CITED4* gene, was associated with an 82% decreased risk of bladder cancer (*OR* = 0.18, q-value = 0.05) (Chapter 1). The result was robust to sensitivity analyses accounting for time between enrollment and diagnosis, race, tumor subtype, and secondhand smoke exposure. Most of the excess relative risk (*ERR*) associated with current smoking for a 30 pack-year smoking history as compared to never smoking was mediated through cg05575921 in an enhancer-like regulatory element within *AHRR* and cg19859270 in the first exon of *GPR15* (Chapter 2). The largest components were the mediated interactions for both cg05575921 (*ERR* component = 2.29, p = 0.05; percent of *ERR* = 72%, p = 0.02) and cg19859270 (*ERR* component = 1.89, p-value = 0.05; percent of *ERR* = 72%, p-value = 0.04), where the mediated interaction capture the indirect effect of smoking through differential methylation as well as the smoking-CpG interaction. There was little evidence that the effect of smoking on bladder cancer risk is mediated through cg03636183. Though not statistically significant, our results suggest that large proportions of the modest effects of rs401681 (NIE = 1.05; NIE percent = 98.5%) and of rs2294008 (NIE = 1.10; NIE percent = 77.6%) on bladder cancer risk occur through their associated CpG sites (Chapter 3). Based on exploratory analyses, cg27028750, which is located in a long terminal repeat element, may drive the indirect effect for rs401681. The effect of rs2294008 may be primarily mediated by cg24023258 near *LY6K* and by cg17252645 in *LY6D* among non-smokers and primarily mediated by cg03405983 and cg17888033 in *LYNX1* and by cg06565975 near *SLURP1* among smokers. There was little evidence supporting mediation through changes in DNA methylation for the associations of rs8102137 and rs798766 with bladder cancer risk.

Conclusions: While results need to be validated in additional prospective studies, differential methylation in the promoter region of *CITED4*, as measured in blood, is a promising marker of bladder cancer susceptibility. If confirmed, smoking may have effects on bladder cancer related to changes in *AHRR* and *GPR15* expression. Our results also suggest that for some SNPs associated with bladder cancer, nearby methylation changes may be part of the underlying mechanisms of effect.

Supplemental File: As supplementary information, “KM Jordahl Supplementary Table 1-1.xlsx” contains Supplementary Table 1.1 and provides the full results for all 361,184 CpG sites from the genome-wide study of DNA methylation and bladder cancer risk described in Chapter 1.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful for supervision from many incredible mentors. As my dissertation committee members, Parveen Bhatti, Amanda Phipps, Emily White, Timothy Randolph, and Michael Wu provided invaluable guidance and insight into the study design, approach to data analysis, and interpretation of results for my dissertation chapters. I owe special thanks to my dissertation chair, Parveen Bhatti, who helped me expand my skills beyond statistical analysis and who provided an exceptional amount of support for all of my endeavors. Other members of our research group also contributed to these dissertation chapters, where Xiaoling Song performed the bisulfite conversion and plating of DNA samples, Cassandra Sather ran the DNA methylation assays, Sushma Thomas genotyped genetic variants, Lesley Tinker was our WHI sponsor, and Karl Kelsey assisted with the interpretation of our results based on his expertise in the field of epigenetics.

I am also indebted to my network of family members, including my parents, step-parents, aunts, uncles, siblings, cousins, and grandparents, to my close friends in Seattle, to my dear friend Christyn, and to my life partner for encouraging me throughout of the process of completing my PhD.

This research would not have been possible without the participation of the women in the Women's Health Initiative and the measurement of genome-wide DNA methylation levels in their blood samples through an American Cancer Society Grant (125299-RSG-13-100-01-CCE). I am thankful for financial support for the first three years of my PhD from a pre-doctoral fellowship on the Cancer Prevention Training Grant (R25 CA094880), which also provided research funds for the SNP genotyping project that fulfilled my data collection requirement, and for the last year of my PhD from a position as a project manager on an American Cancer Society Grant (125299-RSG-13-100-01-CCE).

Chapter 1. GENOME-WIDE DNA METHYLATION IN PRE-DIAGNOSTIC BLOOD AND BLADDER CANCER RISK IN THE WOMEN'S HEALTH INITIATIVE

1.1 Prior Publication

Chapter 1 has been published in Cancer Epidemiology, Biomarkers & Prevention (Jordahl et al. 2018) and involves collaboration with the following co-authors: Timothy W. Randolph, Xiaoling Song, Cassandra L. Sather, Lesley F. Tinker, Amanda I. Phipps, Karl Kelsey, Emily White, and Parveen Bhatti. For the purposes of this dissertation, Kristina Jordahl was solely responsible for the analysis of data and for the writing of all sections of the manuscript except for parts of section 1.3.4. Contributions were also made to the conception and design, development of methodology, and administrative support for this study. The published journal article can be found online at http://cebp.aacrjournals.org/content/cebp/27/6/689.full.pdf.

1.2 Introduction

According to 2016 estimates, urinary bladder cancer is the 12th most frequently diagnosed cancer among women in the US and is responsible for approximately 4,570 deaths among women each year (American Cancer Society 2016a). The incidence of bladder cancer increases with age ("Cancer of the Urinary Bladder - SEER Stat Fact Sheets" 2016), and it is most frequently diagnosed among women between the ages of 55 and 84 (SEER 18 data; 2009-2013). The most common histology of bladder cancer is transitional cell carcinoma (TCC), which originates in the urothelial cells of the inner lining of the bladder (American Cancer Society 2016b) and accounts for 94% of bladder cancer cases (Zhang Y et al. 2012). Based on prognostic characteristics, TCC is further divided into non-muscle invasive (NMIBC) and muscle-invasive (MIBC) subtypes (American Cancer Society 2016a), which represent 75% and 25% of all newly diagnosed disease, respectively (Sanli et al. 2017). Perhaps due to the similarity between urinary tract infections and early bladder cancer symptoms that cause delays in seeking clinical evaluation, women are 21% more likely to be diagnosed with advanced bladder cancer than men (Cárdenas-Turanzas M et al. 2006; Dobruch et al. 2016). Though men are more likely to develop bladder cancer,

once diagnosed, the risk of death from bladder cancer is approximately 1.5 times greater for women, even after accounting for race and other prognostic factors (Scosyrev et al. 2009). As such, there is a need for screening indicators that would aid in the identification of high-risk subgroups, particularly among women.

Differential DNA methylation is a particularly promising biomarker of cancer susceptibility because of its well-established links to human cancer (Robertson 2005; Kulis and Esteller 2010). Altered methylation is known to both silence tumor suppressor genes and to activate oncogenes (Kulis and Esteller 2010), and changes in DNA methylation are common and early events in bladder carcinogenesis (Besaratinia and Tommasi 2013). For example, Wolff et al. found that 526 loci were differentially hypermethylated in invasive urothelial tumors as compared to urothelial tissue from cancer-free participants and that non-invasive and invasive urothelial tumors shared 117 differentially hypermethylated loci, which indicates that these changes are associated with and present during early bladder cancer (Wolff et al. 2010).

Only one previous study has evaluated the association of genome-wide methylation, as measured in blood, with bladder cancer. While it identified 9 strongly associated loci, the study considered only a limited number of total loci (~27k), included only 112 cases and 118 controls and used post-diagnostic blood samples collected an average of 1 year after diagnosis, where DNA methylation levels may have been influenced by the presence of cancer or by cancer treatment (Zhang Y et al. 2012).

To our knowledge, ours is the first study of genome-wide DNA methylation, as measured in pre-diagnostic blood, and bladder cancer risk. In addition to a much larger sample size, we utilized the Illumina Infinium HumanMethylation450 Bead Array to capture methylation at a greater number of CpG sites across the genome. Our study focuses entirely on women, who are an understudied population more likely to experience adverse bladder cancer outcomes that would particularly benefit from identification of biomarkers of bladder cancer susceptibility.

1.3 MATERIALS AND METHODS

1.3.1 WHI Study Setting

The present study was nested in the WHI, a prospective study including 161,808 postmenopausal women across the US who were between the ages of 50 and 79 at enrollment (Hays et al. 2003). Women were recruited from 1993 to 1998 through 40 clinical centers in 24 states and the District of Columbia. There are two arms of the WHI, the clinical trials (CT) study and the observational study (OS), with 68,132 and 93,676 participants, respectively. The CT involved concurrent randomized controlled trials of hormone therapy, dietary modification, and calcium/vitamin D, which ended in 2005. Those not able or willing to participate in the clinical trials were asked to participate in the OS. After 2005, WHI participants were invited to enroll in the WHI Extension Studies, which tracked health outcomes for another 10 years.

1.3.2 Study Subjects

Bladder cancer cases were identified through annual medical update questionnaires, and confirmed by blinded and trained physician adjudicators using standard criteria based on pathology, cytology, and operative reports and on hospital discharge information. As of September 2012, 618 WHI participants had been diagnosed with bladder cancer. We restricted our study to the 584 participants diagnosed with the most common TCC subtype of bladder cancer. Cancer-free controls were matched to cases on year of enrollment, age at enrollment (+/- 2 years), follow-up time greater than or equal to their matched case, trial component, and DNA extraction method (5-prime, phenol, or PureGene). After excluding cases lacking an eligible control, we also removed cases who reported diagnosis of any cancer before baseline and those who did not have a sufficient amount of baseline DNA. Selected cases had a median of 7.22 years between enrollment and bladder cancer diagnosis. We also removed the only American Indian/Alaskan Native case and her matched control to allow for accurate estimation of the race/ethnicity covariate and a genetically identical case-control pair. As a result, our final sample size included 440 case-control pairs, where 228 cases were OS participants and 212 were CT participants. Of the 440 cases, 105 were classified as having MIBC.

1.3.3 Data and Biospecimen Collection

During initial screening for the WHI, basic demographic information was collected, and eligible women were invited for a clinic visit. Staff collected physical measurements and blood specimens at baseline. Blood samples were taken after at least 12 hours of fasting, divided into aliquots, centrifuged, and stored at -70°C as buffy coats. Self-administered questionnaires were also collected at baseline to gather information related to risk factors for various health outcomes of interest, including detailed data on smoking. Participants were asked to report whether they had ever smoked at least 100 cigarettes in their lifetime, whether they currently smoked cigarettes, whether they had ever smoked to lose weight, their average number of cigarettes currently or previously smoked per day, and their number of years as a regular smoker, from which smoking status (current, former, never) and pack-years were computed. For OS participants, information on second-hand smoke exposure was captured at baseline by asking the participant if she lived with an inside-smoker as a child, lived with an inside-smoker as an adult, and if she worked in a place where people smoked.

1.3.4 DNA Methylation Array

We used the Illumina 450K Infinium HumanMethylation Bead Array to interrogate methylation status at approximately 485,577 CpG sites. The array covers RefSeq genes (99%), CpG islands (96%), CpG island shores (92%), and CpG island shelves (86%) with an average of 17.2 CpG sites per gene region and an average of 5.63 sites per CpG island (Bibikova et al. 2011).

DNA was previously extracted from buffy coats using 5-prime, phenol, or PureGene methods and stored at -70°C. DNA was quantified by Picogreen (Invitrogen, Carlsbad, CA). A bisulfite conversion kit (Zymo Research) was used to treat 500ng of DNA with sodium bisulfate, which converts unmethylated cytosine to uracil. Based on standard Illumina protocols, a 4µl aliquot of bisulfite converted, denatured and neutralized DNA was enzymatically fragmented for 60 minutes at 37°C, precipitated with isopropanol, air dried, and re-suspended in hybridization buffer. Samples were stored at -80°C for no more than one month before being assayed with the Illumina array. DNA samples were incubated on the beadchips in a hybridization oven at 48°C for 16-24 hours. DNA samples were then run on the Illumina (San Diego, CA) Infinium HumanMethylation450 Bead Array. Washed beadchips were extended and stained in capillary

flow-through chambers and then scanned using Illumina iScan. Twenty-two blind duplicate sample pairs were randomly placed among study samples for quality control assessment.

We used the M-value to measure methylation at each CpG site to reduce the heteroscedasticity as compared to the β-value (Du et al. 2010). The M-value is the base-2 logit of the β-value, where the β-value is the ratio of the methylated signal over the total signal and can be interpreted as the percent of methylation at a specific site (Dedeurwaerder et al. 2014).

1.3.5 Methylation Data Processing

After reading in the raw image files and checking for failed samples, we performed normalization in two steps. First, we corrected the probe intensity levels for background fluorescence, which avoids potential bias resulting from a difference in red/green dye intensity, using the normal-exponential convolution using out-of-band probes (noob) method (Wright et al. 2016). Next, we performed functional normalization to adjust the marginal distribution of methylation levels using summarized information from the control and out of band probes (Fortin et al. 2014). This method effectively isolates unwanted variation, including batch effects, and performs well in studies of cancer (Fortin et al. 2014). After normalization, we sequentially excluded probes that had detection p-values greater than or equal to 0.01 in at least 10% of samples (707 probes), probes with a beadcount less than 3 in at least 10% of samples (215 probes), single nucleotide polymorphism (SNP)-related probes (87,706 probes), cross-reactive probes (24,939 probes), probes located on the sex chromosomes (9,578 probes), and non-CpG sites (1,183 probes). SNP-related probes were excluded because they may impact probe binding (Y. Chen et al. 2013). We removed probes with any SNP present at the CpG interrogation or single nucleotide extension sites as identified by the Illumina annotation, and probes within 10 base pairs of a common SNP (minor allele frequency > 1%) based on 1000 Genomes data from 'Illumina450ProbeVariants.db' (Butcher 2013). Previously identified cross-reactive probes were excluded based on the Chen et al. and Nordlund et al. manuscripts (Y. Chen et al. 2013; Nordlund et al. 2013). We performed these steps using R (version 3.4.0) and the 'minfi' and 'wateRmelon' R packages (Aryee et al. 2014; Pidsley et al. 2013).

1.3.6 Imputation of Missing Data

We used multivariate imputation by chained equations (MICE) to impute values for missing covariate data (Buuren and Groothuis-Oudshoorn 2011). This approach specifies conditional densities for each variable with missing information, draws imputed values by iterating over these conditional densities, and accounts for the uncertainty in predicted values by imputing multiple datasets with complete covariate information.

We generated 5 datasets with complete covariate information using this approach. We used default imputation model specifications for the imputation process and used 10 iterations per dataset to improve convergence. As the first step, we multiply imputed missing values for the pack-years and race/ethnicity covariates. Our predictor set for the initial phase of imputation included these variables as well as: age at baseline, WHI arm (CT, OS), ever smoking status, cell type composition, case-control status, and methylation level at 123 CpG sites. To identify these 123 CpG sites, we started by including the top 100 sites associated with current smoking and the top 100 sites correlated with pack-years in current smokers based on a large study by Ambatipudi et al. (Ambatipudi et al. 2016). There were 155 unique sites from these lists, but only 123 were available in our data after applying our processing pipeline. As the second step, within each partially imputed dataset, we added smoking status at baseline to the predictor set and then singly imputed missing values for current or former smoking status among smokers (pack-years > 0) and assumed never smoking status among non-smokers (pack-years = 0).

1.3.7 Statistical Analyses

We used conditional logistic regression models to assess the association between bladder cancer case-control status and each continuous M-value after adjusting for potential confounders, including race/ethnicity (Asian/Pacific Islander, black/African American, Hispanic/Latino, non-Hispanic white, other), smoking status (never, former, current), pack-years of smoking (continuous), as well as the estimated proportion of CD4+ T-cells (continuous), CD8+ T-cells (continuous), natural killer cells (continuous), granulocytes (continuous), and B-cells (continuous). The relative contribution of the six major white blood cell types were reconstructed using the Houseman et al. method (Houseman et al. 2012). We included race/ethnicity as a covariate because a multidimensional scaling plot (MDS) plot, based on the β-values

from 1,000 CpG sites with the most variation, clearly showed that race/ethnicity is a substantial source of variability in our data (Figure 1.1). To account for multiple testing, we calculated q-values by adjusting p-values using the false-discovery rate (FDR). Associations with bladder cancer at a q-value less than or equal to 0.05 were deemed statistically significant.

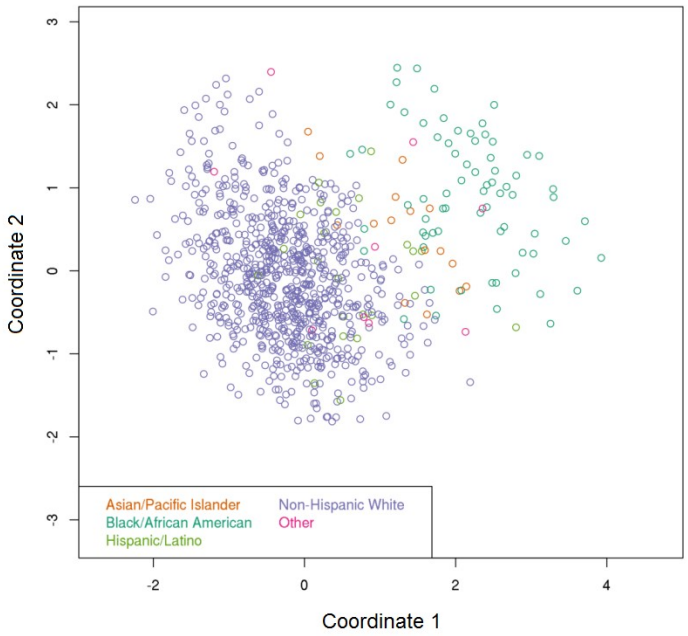

Figure 1.1 MDS plot based on the 1,000 most variable positions, showing race/ethnicity is the most substantial source of variability in our data

Since we performed multiple imputation, the conditional logistic regression analysis for each site was repeated in all of the imputed datasets. To determine the combined parameter estimates and their variance, the results from each of the five complete data analyses were pooled using the approach of Rubin (Rubin 1987) as implemented in the 'mice' R package (Buuren and Groothuis-Oudshoorn 2011).

1.3.8 Sensitivity Analyses

To address the potential influence of early, undiagnosed cancer at the time of blood draw on DNA methylation, we performed a sensitivity analysis restricted to cases diagnosed at least 3 years after enrollment and to their matched controls. To address any possible residual confounding by race/ethnicity,

we also completed an analysis restricting to case-control pairs with matching race/ethnicity. We performed an analysis restricted to cases with the MIBC subtype and to their matched controls to explore the impact of clinically heterogeneous bladder cancer subtypes on our primary results. To address the potential impact of secondhand smoke exposure, we restricted to OS participants and additionally adjusted for secondhand smoke exposure at home as a child (yes, no), secondhand smoke exposure at home as an adult (yes, no), and secondhand smoke exposure at work (yes, no). Given the importance of smoking as a potential confounder, we completed analyses restricted to never smokers. In addition, we completed an unmatched, genome-wide analyses with the M-value of each CpG site as the outcome to check for associations with smoking status (current, never), pack-years of smoking, and second-hand smoke exposure. The absence of associations among our top hits with these smoking variables would help address concerns that our primary findings were driven by residual confounding from incomplete adjustment for tobacco smoke exposure.

1.4 RESULTS

1.4.1 Quality Control for Methylation Data

Based on the 65 SNP probes included on the Illumina array, the genetic distance between the 21 blind duplicate pairs in our study ranged from 0.005 to 0.026, verifying that these samples were genetically identical. The Pearson correlation coefficients between the methylation measurements (β-values) across all 361,184 loci were at least 0.995 for each duplicate pair, indicating suitable precision of the assay.

There was marginally lower concordance in duplicates located on different plates as compared to duplicates located on the same plate, which were likely due to residual batch effects that were not fully addressed through function normalization. Also, cg22748573 was a low outlier in four of the five 100%-methylated samples and may have slightly deflated methylation levels as measured by our assays. However, all case-control pairs were measured on the same plate, and these observed differences are not expected to materially affect study results.

When assessing genetic distance among study samples, we discovered a case-control pair that appeared to be genetically identical. These samples had the same methylation levels across the 65 array probes that measure SNP genotype. Since these samples had the same genotype for a relatively large

number of SNPs, this suggested that these samples came from the same person. We evaluated the possibility that a plating error occurred and affected the measurement of subsequent samples, particularly those occurring on the same plate. However, we found no issue with four QC samples on the same plate that had duplicates on other plates. The genetic distance for these QC pairs, based on the 65 SNP probes, ranged from 0.01 to 0.03. In addition, there were 0%- and 100%-methylated samples included on the plate and DNA methylation array measurements of these samples were consistent with measurements of 0%- and 100%-methylated samples on other plates. As such, we are confident that the issue was confined to the single case-control pair and, as mentioned previously, we dropped those samples from the study.

1.4.2 Genome-wide Analysis of DNA Methylation

The distribution of selected variables is presented by case-control status in Table 1.1. A greater proportion of cases were white, and cases were more likely to be past and current smokers and had greater pack years of smoking than controls. While OS cases were equally likely to be exposed to secondhand smoke at home as children and at work, they were slightly more likely to be exposed to secondhand smoke at home as adults. Inferred cell type composition was similar across cases and controls.

Table 1.1 Distribution of relevant clinical characteristics among bladder cancer cases and controls nested within the Women's Health Initiative

	Cases	Controls
Age (mean, SD)	65.12 (7.04)	65.12 (7.03)
Race (N, %)		
Asian/Pacific Islander	4 (1%)	12 (3%)
Black/African American	25 (6%)	45 (10%)
Hispanic/Latino	8 (2%)	16 (4%)
Non-Hispanic White	400 (91%)	360 (82%)
Other	2 (<1%)	7 (1%)
Missing	1 (<1%)	0 (0%)
Smoking (N, %)		
Never Smoked	156 (35%)	236 (54%)
Past Smoker	218 (50%)	178 (40%)
Current Smoker	56 (13%)	19 (4%)
Missing	10 (2%)	7 (2%)
Pack-Years (N, %)		
Never Smoker	156 (36%)	236 (54%)
< 5	36 (8%)	68 (15%)
5 - < 20	63 (14%)	46 (10%)
>= 20	162 (37%)	78 (18%)
Missing	23 (5%)	12 (3%)
Child Secondhand Smoke Exposure at Home (N, %)[a]		
No	80 (18%)	78 (18%)
Yes	141 (32%)	145 (33%)
Don't Know	4 (1%)	4 (1%)
Missing	215 (49%)	213 (48%)
Adult Secondhand Smoke Exposure at Home (N, %)[a]		
No	48 (11%)	70 (16%)
Yes	178 (40%)	155 (35%)
Missing	214 (49%)	215 (49%)
Secondhand Smoke Exposure at Work (N, %)[a]		
No	46 (10%)	54 (12%)
Yes	180 (41%)	173 (39%)
Missing	214 (49%)	213 (49%)

Continued on page 11

Table 1.1 Distribution of relevant clinical characteristics among bladder cancer cases and controls nested within the Women's Health Initiative (continued)

	Cases	Controls
Cell Type Composition (mean, SD)		
B-Cells	0.06 (0.03)	0.06 (0.03)
CD8 T-Cells	0.09 (0.05)	0.09 (0.04)
CD4 T-Cells	0.19 (0.08)	0.17 (0.07)
Natural Killer Cells	0.09 (0.05)	0.09 (0.04)
Granulocytes	0.50 (0.13)	0.51 (0.13)
Monocytes	0.11 (0.03)	0.11 (0.03)

[a]Information about exposure to secondhand smoke was only available for participants in the Observational Study arm, so the 212 case-control pairs from the Clinical Trials arm were missing this data
[b]Cell type composition was calculated using the Houseman method (Houseman et al. 2012)

As shown in Table 1.1, 35 participants were missing pack-years of smoking data and 17 of those missing pack-years information were also missing smoking status data. A single participant was missing race/ethnicity data.

The imputed datasets are summarized in Table 1.2. Results for the 10 CpG sites with the lowest q-values in association with bladder cancer are summarized in Table 1.3, while results for all 361,184 CpG sites are provided in Supplementary Table 1.1. These results are also summarized as a volcano plot in Figure 1.2, with the top association highlighted.

After adjusting for multiple comparisons, increased methylation at cg22748573 was statistically significantly associated with a reduced risk of bladder cancer (*OR* = 0.18 per unit increase in M-value, q-value = 0.05). cg22748573 is located within the chr1:41326949-41328285 CpG island in the 5-'UTR of the CBP/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 4 (*CITED4*) gene, and it had an average methylation level of 6% in controls. Although not statistically significant, increased methylation at cg20010635 was strongly associated with increased risk of bladder cancer (*OR* = 8.42 per unit increase in M-value, q-value = 0.40). The cg20010635 CpG site is located in the body of the calmodulin binding transcription activator 1 (*CAMTA1*) gene and had a high average methylation of 95% among controls.

Table 1.2 For each imputed dataset, distribution of relevant clinical characteristics among bladder cancer cases and controls nested within the Women's Health Initiative

	Imputed Dataset 1		Imputed Dataset 2		Imputed Dataset 3		Imputed Dataset 4		Imputed Dataset 5	
	Cases	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**
Race (N, %)										
Asian/Pacific Islander	4 (1%)	12 (3%)	4 (1%)	12 (3%)	4 (1%)	12 (3%)	4 (1%)	12 (3%)	4 (1%)	12 (3%)
Black/African American	25 (6%)	45 (10%)	25 (6%)	45 (10%)	25 (6%)	45 (10%)	25 (6%)	45 (10%)	25 (6%)	45 (10%)
Hispanic/Latino	9 (2%)	16 (4%)	8 (2%)	16 (4%)	8 (2%)	16 (4%)	8 (2%)	16 (4%)	8 (2%)	16 (4%)
Non-Hispanic White	400 (91%)	360 (82%)	401 (91%)	360 (82%)	401 (91%)	360 (82%)	401 (91%)	360 (82%)	401 (91%)	360 (82%)
Other	2 (<1%)	7 (1%)	2 (<1%)	7 (1%)	2 (<1%)	7 (1%)	2 (<1%)	7 (1%)	2 (<1%)	7 (1%)
Smoking (N, %)										
Never Smoked	159 (36%)	237 (54%)	160 (36%)	239 (54%)	162 (37%)	239 (54%)	159 (36%)	238 (54%)	160 (36%)	238 (54%)
Past Smoker	222 (51%)	180 (41%)	221 (50%)	178 (41%)	220 (50%)	180 (41%)	221 (50%)	178 (41%)	220 (50%)	178 (41%)
Current Smoker	59 (13%)	23 (5%)	59 (14%)	23 (5%)	58 (13%)	21 (5%)	60 (14%)	24 (5%)	60 (14%)	24 (5%)
Pack-Years (N, %)										
Never Smoker	159 (36%)	237 (54%)	160 (36%)	239 (54%)	162 (37%)	239 (54%)	159 (36%)	238 (54%)	160 (36%)	238 (54%)
< 5	37 (9%)	70 (16%)	37 (8%)	69 (16%)	38 (9%)	68 (15%)	38 (9%)	69 (16%)	40 (9%)	68 (15%)
5 - < 20	67 (15%)	51 (11%)	65 (15%)	50 (11%)	66 (15%)	51 (12%)	67 (15%)	49 (11%)	66 (15%)	51 (12%)
>= 20	177 (40%)	82 (19%)	178 (41%)	82 (19%)	174 (39%)	82 (19%)	176 (40%)	84 (19%)	174 (40%)	83 (19%)

Table 1.3 Top 10 CpG sites in tests of association between methylation levels from pre-diagnostic blood and bladder cancer risk based on conditional logistic regression models adjusted for race/ethnicity, smoking status, pack-years of smoking, and cell type composition

CpG Site	Location[a]	Gene	Gene Group[b]	Relation to CpG Island[c]	*OR*	95% CI[d]	P-value	Q-value[e]
cg22748573	chr1:41327924	*CITED4*	5'UTR;1stExon	Island	0.18	(0.09, 0.34)	1.5e-07	0.05
cg20010635	chr1:7811409	*CAMTA1*	Body	OpenSea	8.42	(3.48, 20.33)	2.2e-06	0.40
cg25955565	chr14:71863756	*SNORD56B*	TSS1500	OpenSea	9.29	(3.59, 24.08)	4.5e-06	0.41
cg26492847	chr11:96063104	*MAML2*	Body	OpenSea	7.24	(3.01, 17.38)	9.5e-06	0.41
cg06414161	chr16:53534040	*AKTIP*	Body	N_Shelf	8.93	(3.36, 23.72)	1.1e-05	0.41
cg27627854	chr11:108158382	*ATM*	Body;1stExon	OpenSea	4.19	(2.21, 7.98)	1.2e-05	0.41
cg02393449	chr10:70743010	*DDX21*	3'UTR	OpenSea	6.91	(2.89, 16.53)	1.4e-05	0.41
cg26695157	chr12:109985283			OpenSea	3.71	(2.05, 6.72)	1.5e-05	0.41
cg17803089	chr6:35419793	*FANCE*	TSS1500	N_Shore	0.18	(0.08, 0.39)	1.8e-05	0.41
cg26729242	chr10:96980135	*C10orf129*	Body	OpenSea	3.00	(1.81, 4.96)	1.9e-05	0.41

[a]Chromosome and chromosomal coordinates of CpG (Build 37)
[b]Functional region of gene as indicated in v1.2 Illumina annotation: TSS1500 = 200-1500 bases upstream of the transcription start site; 5'UTR = Within the 5 prime untranslated region; 1stExon = First segment of gene coding for peptide sequence; Body = Between the ATG and stop codon; 3'UTR = Between the stop codon and poly A signal; Multiple listings indicate a locus in a region with multiple splice variants
[c]Position relative to CpG island as indicated in v1.2 Illumina annotation: Island = Within CpG island (CG content > 50%, Obs/Exp CpG ratio > 0.60, and length > 200 bps); OpeanSea = Non-island region; Shore = 0-2 kb flanking CpG Island; Shelf = 2-4 kb flanking CpG Island
[d]95% confidence interval for odds ratio estimate
[e]Q-values represent p-values adjusted for multiple testing using the FDR method

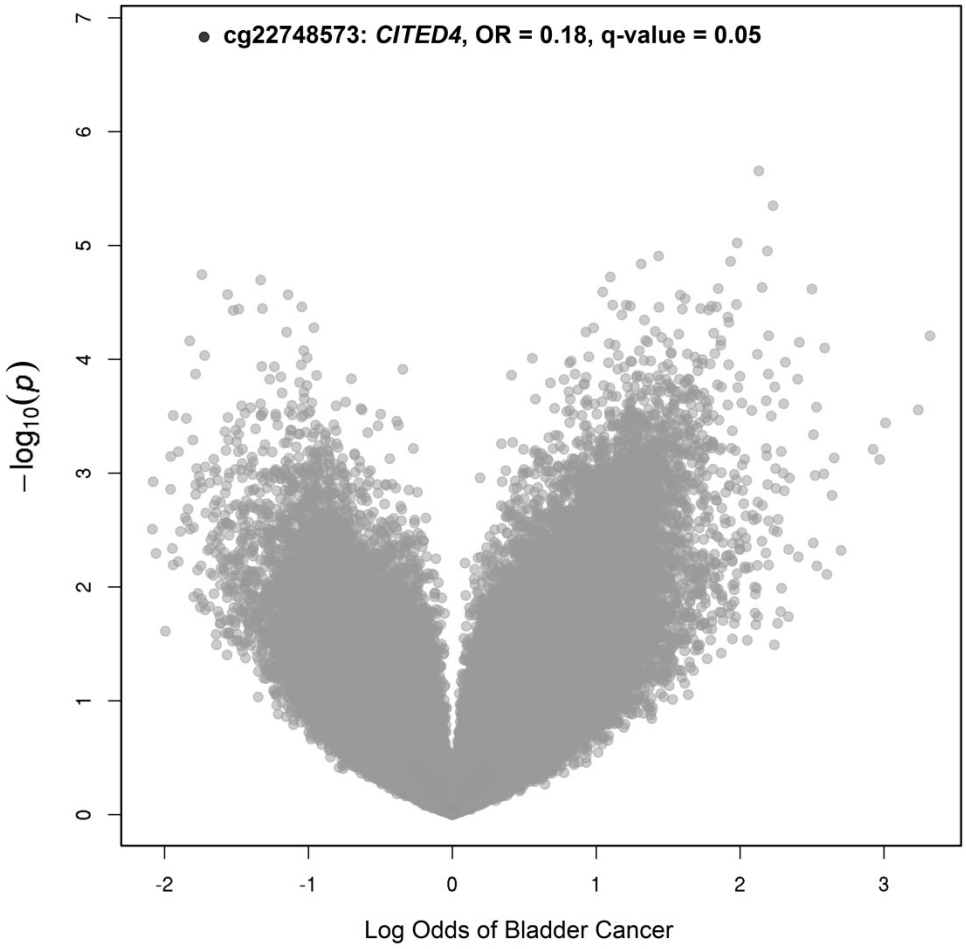

Figure 1.2 Volcano plot of results from the analysis of 361,184 CpG sites in association with risk of bladder cancer

1.4.3 Sensitivity Analyses

Most of the sensitivity analyses yielded odds ratios for cg22748573 that were similar to or slightly stronger than the results from the main analysis (*OR* = 0.18), suggesting little bias in the odds ratio estimate. However, as expected, statistical significance was reduced due to the decreased sample size in each of these analyses. When examining the potential impact of undiagnosed disease, we excluded 74 case-

control pairs where the case was diagnosed with bladder cancer within 3 years of study entry, and cg22748573 remained the most significant finding (*OR* = 0.15, q-value = 0.06). In the race/ethnicity sensitivity analysis, we included only the 337 case-control pairs with matching race/ethnicity, and cg22748573 was no longer significantly associated with bladder cancer; however, despite the more than 20% reduction in sample size, cg22748573 still had the 3rd lowest p-value (*OR* = 0.19, q-value = 0.47). With very small numbers in various non-white race/ethnicity groups, the MIBC analysis had to be restricted to 96 case-control pairs with white or black race/ethnicity for our statistical models to converge. While the analysis produced null results, the cg22748573 locus in *CITED4* was the 2nd most significant locus and the magnitude of the association was much stronger than in the overall TCC analysis (*OR* = 0.03, q-value = 0.71). After restricting to OS participants and additionally imputing the secondhand smoke exposure variables for 8 subjects, the sensitivity analysis for secondhand smoke exposure was also null, but had cg22748573 as the 11th most significant locus (*OR* = 0.13, q-value = 0.87). With only 85 case-control pairs in the analysis restricted to never smokers, cg22748573 was the 530th most significant locus (*OR* = 0.08, q-value = 0.93), but the magnitude of the association with bladder cancer was stronger than the result from our primary analysis. In our analyses of associations between genome-wide DNA methylation and the various tobacco smoke exposure variables, no statistically significant associations were observed with any of the differentially methylated CpG sites highlighted in Table 1.3.

1.5 DISCUSSION

We identified a differentially methylated locus in pre-diagnostic blood that was statistically significantly associated with an 82% reduction in bladder cancer risk. The cg22748573 locus resides in a CpG island within the putative promoter region of *CITED4*. Though methylation is often tissue-specific, methylation levels are more likely to be conserved across various tissues in CpG islands. Price et al. observed that probes differentially methylated between tissues were depleted in Illumina-annotated islands (Price et al. 2013), and Eckhardt et al. reported that only 13% of tissue-specific differentially methylated regions in the 5'-UTR were located within CpG islands (Eckhardt et al. 2006). Though speculative, this suggests that the methylation status of cg22748573 in blood may reflect that of normal bladder tissue.

There is evidence that *CITED4* promoter hypermethylation is associated with better cancer outcomes (Tews et al. 2007), which is unexpected, since *CITED* genes have been identified as tumor suppressors (Fox et al. 2004; Huang et al. 2011; Rogers et al. 2016), and methylation of CpG loci in promoter regions is typically associated with transcriptional silencing (Peter A. Jones 2012; Rakyan et al. 2011; Laird 2010). However, based on data from subjects in The Cancer Genome Atlas (TCGA) with both Illumina array methylation and Illumina HiSeq RNA-Seq gene expression data (Díez-Villanueva, Mallona, and Peinado 2015), we found that methylation at cg22748573 was not significantly associated with *CITED4* expression in normal bladder tissue (Spearman correlation = -0.22, p-value = 0.31). Therefore, the importance of the methylation change at cg22748573 may not be its impact on *CITED4* expression. Instead, this methylation change may be a marker of decreased susceptibility to the development of bladder cancer through other mechanisms. As an example, hypermethylation of cg22748573 has been closely associated with deletion of 1p and 19q in oligodendrogliomas (Tews et al. 2007), where tumors with 1p/19q tumor deletions often lack *TP53* mutations (Reifenberger and Louis 2003) and have better patient prognoses (Boots-Sprenger et al. 2013; Tews et al. 2007).

Tews et al. observed *CITED4* hypermethylation in glioma patients to be associated with longer recurrence-free and overall survival, but hypermethylation was not observed in matched leukocyte DNA (Tews et al. 2007). However, no comparisons between normal brain tissue and blood were made. Given the genomic instability of tumors, one would expect the correlation with DNA methylation levels to be lower for tumor than for normal tissue. Even if blood does not reflect the methylation state of *CITED4* in bladder tissue, differential methylation of *CITED4*, a gene which has been shown to play a central role in blood cell differentiation in mice (Yahata et al. 2002), may reflect systemic changes due to factors like inflammation or environmental exposures that are associated with bladder cancer. In this context, differential methylation of *CITED4* could still be a valuable marker of susceptibility to bladder cancer.

While not statistically significant at the genome-wide level, the second-most significant association was between increased methylation at the cg20010635 locus in the gene body of *CAMTA1* and increased bladder cancer risk. *CAMTA1* was initially identified as a putative tumor suppressor gene because it is located in a region of the genome that is frequently deleted in cancers (Katoh and Katoh 2003; Barbashina et al. 2005). However, recent research suggests that *CHD5* may be the relevant tumor

suppressor gene in the region and that the C-terminus of *CAMTA1* acts as an oncogene when fused with *WWTR1* in epithelioid hemangioendothelioma (Tanas et al. 2011). Since *CAMTA1* is usually only expressed in the brain (Tanas et al. 2011), its role in bladder carcinogenesis is uncertain.

Marsit et al. conducted the first study of loci-specific genome-wide DNA methylation of bladder cancer in blood. This study identified the top 9 loci associated with bladder cancer among 112 cases and 118 controls, of which the most notable were in *BRD7*, *TBCA*, and *COX7C* (Marsit et al. 2011). While *CITED4* was not among these top sites, this study was comparatively underpowered and utilized post-diagnostic blood samples.

Ours is the first study we are aware of that investigates the promising potential link between site-specific genome-wide DNA methylation in blood and future risk of bladder cancer. Being nested in the WHI, we are also the only study of this kind to use pre-diagnostic samples, and we benefited from a relatively large sample of cases and controls with detailed baseline information on potential confounders. Even with pre-diagnostic blood samples, it is possible that methylation measurements in blood reflected early tumor development (Xu et al. 2013). However, our sensitivity analysis restricted to cases diagnosed at least 3 years after enrollment suggests that the association observed with cg22748573 is not driven by the presence of early cancer. Sensitivity analyses also suggested that residual confounding due to inadequate control of exposure to tobacco smoke did not drive our primary result.

Our results suggest that subtypes of TCC may share some methylation susceptibility markers; however, we were underpowered to specifically examine genome-wide associations between DNA methylation and MIBC. We will explore efforts to pool data with future studies of DNA methylation and bladder cancer in an effort to increase power to evaluate subtype-specific effects.

Ideally, pre-diagnostic measurements of DNA methylation would be measured in the tissue of interest. Accessing bladder tissue, however, involves invasive procedures. While urine samples can provide epithelial cells from the bladder, they also include epithelial cells from the rest of the urinary tract. The contribution of cells from the various tissues in the urinary tract can be highly variable (Cheng et al. 2017), and urine samples were only collected from three of the clinical sites in the WHI.

Overall, we discovered differential methylation of cg22748573 in pre-diagnostic blood to be strongly and significantly associated with bladder cancer risk among post-menopausal women. If

confirmed in additional prospective studies, molecular studies should be conducted to determine the exact role of this locus in bladder carcinogenesis. Future prospective studies should involve both men and women and should evaluate the contribution of methylation measurements at this locus to risk prediction models that account for smoking history and germline genetic polymorphisms previously associated with bladder cancer risk.

Chapter 2. DIFFERENTIAL DNA METHYLATION AS A POTENTIAL MEDIATOR OF THE ASSOCIATION BETWEEN CIGARETTE SMOKING AND BLADDER CANCER RISK IN THE WOMEN'S HEALTH INITIATIVE

2.1 INTRODUCTION

Cigarette smoking has a well-established causal association with bladder cancer (Centers for Disease Control and Prevention (US) 2010). Based on results from the NIH-AARP cohort of 186,134 women, smoking accounts for an estimated 52% of bladder cancer incidence in US women between the ages of 50 and 71 (Freedman et al. 2011a).

Known pathways linking smoking to cancer involve exposure to more than 60 carcinogens, formation of DNA adducts, and accumulation of somatic mutations (Centers for Disease Control and Prevention (US) 2010). For bladder cancer, smoking primarily affects disease risk through exposure to aromatic amines (Besaratinia A, Cockburn M, and Tommasi S 2013), which are genotoxic chemical carcinogens that induce DNA adducts and increase the frequency of mutations (Besaratinia and Tommasi 2013). This exposure likely occurs when smoking-related compounds are excreted in urine (Florescu et al. 2009), but this established mechanism does not explain why bladder tissue is particularly susceptible to the tumorigenic effects of aromatic amines (Besaratinia A, Cockburn M, and Tommasi S 2013). This unexplained aspect of tumorigenesis in bladder tissue suggests that as yet unknown mechanisms may influence bladder cancer risk; elucidation of these additional mechanisms would improve our understanding of bladder carcinogenesis and may reveal new avenues for bladder cancer prevention, screening, and treatment.

DNA methylation at specific loci have been reliably associated with past and current smoking (Shenker, Ueland, et al. 2013; Gao et al. 2015; Philibert et al. 2013). Thus, a promising hypothesis is that smoking has indirect effects on bladder cancer through changes in DNA methylation at smoking-associated CpG sites, as was observed in a previous study of lung cancer. Fasanelli et al. observed that hypomethylation of cg05575921 and of cg03636183 in blood explained approximately 37% of the increased risk of lung cancer induced by smoking (Fasanelli et al. 2015). Changes in DNA methylation can alter or reflect alteration of the expression of nearby genes. Shifts in expression may directly affect

oncogenes or tumor suppressor genes or, as observed in adenocarcinomas, may promote smoking-specific transversion mutations (Stueve et al. 2017). Our study investigates the hypothesis that methylation changes at smoking-associated CpG sites, as measured in DNA from pre-diagnostic blood samples, mediate the association between smoking and bladder cancer risk.

2.2 Materials and Methods

2.2.1 Study Participants

Data for the current analyses were drawn from our case-control study of pre-diagnostic DNA methylation and bladder cancer nested in the Women's Health Initiative (WHI) (Jordahl et al. 2018). Briefly, the WHI includes 161,808 postmenopausal women recruited from 1993 to 1998 across the US (Hays et al. 2003). There are two arms of the WHI, the clinical trials (CT) and the observational study (OS). The CT involved concurrent randomized controlled trials of hormone therapy (HRT) and dietary modification (DM), and subsequently calcium/vitamin D (CaD). Those not eligible or willing to participate in the CT were asked to participate in the OS. After 2005, WHI participants were invited to enroll in the WHI Extension Studies, which tracked health outcomes for another 10 years. The time from enrollment to end of follow-up was based on the number of days between enrollment and death or last contact. A total of 440 cases with transitional cell carcinoma of the bladder and 440 cancer-free controls matched on year of enrollment, age at enrollment (+/- 2 years), follow-up time greater than or equal to their matched case, trial component, and DNA extraction method were included in our study (Jordahl et al. 2018).

Since smoking-associated methylation changes are reversible (Ambatipudi et al. 2016; Fasanelli et al. 2015), we expect the associations under investigation to be attenuated in former smokers. Therefore, in order to most effectively examine the relevant causal relationships, we further restricted our analyses to the 210 cases and 256 controls that were current or never smokers with pack-years data. To allow for the covariate adjustment of our models, we additionally dropped the few participants whose DNA was extracted using the BioServe or salt methods (N = 4 subjects) and who were missing data on race/ethnicity or education level (N = 5 subjects), which left 206 cases (53 current smokers, 153 never smokers) and 251 controls (18 current smokers, 233 never smokers) for the mediation analyses.

2.2.2 Data and Biospecimen Collection

Basic demographic information, including age and race/ethnicity, was requested during the screening process (Hays et al. 2003). Through baseline questionnaires, participants reported if they had smoked at least 100 cigarettes ever (yes, no), currently smoked cigarettes (yes, no), or had ever smoked to lose weight (yes, no), and this information was used to determine smoking status. Participants also disclosed the average number of cigarettes currently or previously smoked per day (<1, 1-4, 5-14, 15-24, 25-34, 35-44, 45+) and their number of years as a regular smoker (<5, 5-9, 10-19, 20-29, 30-39, 40-49, 50+). Pack-years were calculated from these questions by dividing the approximate midpoint of the cigarettes per day category (0.5, 2.5, 10, 20, 30, 40, 50) by 20 and then multiplying by the approximate midpoint of the years of regular smoking category (2.5, 7, 15, 25, 35, 45). For current smokers with years as a regular smoker that was over 50 or missing, years of regular smoking was estimated based on the categorical age at which they started smoking regularly, if available. Also at baseline, blood samples were collected after at least 12 hours of fasting and stored at -70°C as buffy coats.

2.2.3 DNA Methylation Array

As described previously, we used the Illumina 450K Infinium HumanMethylation Bead Array to interrogate methylation status at approximately 485,577 CpG sites among bladder cancer cases and controls (Jordahl et al. 2018). We used the M-value to measure methylation at each CpG site to improve the heteroscedasticity of methylation levels (Du et al. 2010). The M-value is the base-2 logit of the β-value, where the β-value is the ratio of the methylated signal over the total signal and can be interpreted as the percent of methylation at a specific site (Dedeurwaerder et al. 2014). After reading in the raw image files and checking for failed samples, we performed background correction and functional normalization of the methylation data (Jordahl et al. 2018). The CpG sites included in the current analysis passed our quality control steps, which excluded any CpG sites that were undetected in at least 10% of samples, had a beadcount less than 3 in at least 10% of samples, were in or near common SNPs (minor allele frequency greater than 1%), were classified as cross-reactive probes, or were located on the sex chromosomes (Jordahl et al. 2018).

2.2.4 Identifying Smoking-Associated Loci

A systematic review, published in 2015, summarized information from the substantial body of research that has evaluated associations between smoking and methylation in blood (Gao et al. 2015). The review included 14 methylome-wide studies and reported a total of 1,460 CpG sites associated with smoking. Three CpG loci stood out as the most robustly associated loci, since they were reported in at least 10 of the 14 included methylome-wide studies. These sites were cg05575921 (located in the aryl-hydrocarbon receptor repressor gene; *AHRR*), cg03636183 (located in the F2R like thrombin receptor 3 gene; *F2RL3*), and cg19859270 (located in the G-protein receptor 15 gene; *GPR15*). The associations of these loci with smoking have been verified in two subsequent large-scale studies (Ambatipudi et al. 2016; Joehanes et al. 2016).

2.2.5 Validation and Exploratory Analyses

To validate the previously reported associations between the selected CpG sites and current smoking in our study population, we tested the association between current smoking and genome-wide DNA methylation using adjusted linear regression models based on methylation β-values and then examined the results for the three CpG sites of interest. The β-value, rather than the M-value, was used to allow direct comparisons with results from previous studies. We used the empirical Bayes approach to test the significance of these associations and adjusted for multiple testing using the FDR method (Li et al. 2015).

To evaluate the potential impact of differences in blood cell type composition on our results, we used the reference-free method to adjust the genome-wide associations with current smoking for latent variables related to cell mixture (Houseman, Molitor, and Marsit 2014), which also required adjusted linear regression models with β-values as the outcomes. The approach uses methods developed for surrogate variable analysis to estimate and adjust for latent variables by creating a methylation profile for each cell type or subtype based on the common variation between the coefficients and residuals from the linear model (Houseman, Molitor, and Marsit 2014). As described by Houseman et al. (Houseman, Molitor, and Marsit 2014), we used the random matrix theory method to estimate the latent variable dimension. Standard errors were estimated based on 100 bootstrap samples and p-values were additionally adjusted for multiple testing using the FDR method.

2.2.6 Mediation Analysis Overview

Separate mediation analyses were conducted for each of the three CpG sites (cg05575921, cg03636183, cg19859270) as a mediator, with baseline current smoking as the exposure, and incident bladder cancer as the outcome. Smoking was included as a continuous pack-years variable, and involved only current smokers (pack-years > 0) and never smokers (pack-years = 0).

We estimated causal effects using a regression-based approach for dichotomous outcomes (T. J. VanderWeele and Vansteelandt 2010; T. VanderWeele 2015), which is based on a counterfactual framework for causal inference. This method uses a logistic regression model for the outcome (2.1) and a linear regression model for the mediator (2.2), where *A* is the exposure, *M* is the mediator, and *Y* is the outcome of interest. Since this is a case-control study, the linear regression model was fit only among controls.

$$\text{logit}\left(P\left(Y=1 \mid a,m,c\right)\right)=\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\theta_4' c \tag{2.1}$$

$$E[M \mid a,c]=\beta_0+\beta_1 a+\beta_2' c \tag{2.2}$$

The models were adjusted for a set of covariates (*c*) that included our matching covariates and potential confounders. The matching variables were WHI arm (HRT/DM, HRT, DM, OS), age at baseline (continuous), year of enrollment (continuous: 1994-1995 = 1, 1996 = 2, 1997 = 3, 1998 = 4), follow-up time (continuous), and DNA extraction method (5-prime, phenol). We then considered the web of edges between possible confounders in the directed acyclic graph (DAG) shown in Figure 2.1. The DAG illustrates that, in addition to the matching variables, it is necessary to adjust for race/ethnicity (Asian/Pacific Islander, black/African American, Hispanic/Latino, non-Hispanic white, other) and education (<high school, high school, post-high school training, college degree, post-college training).

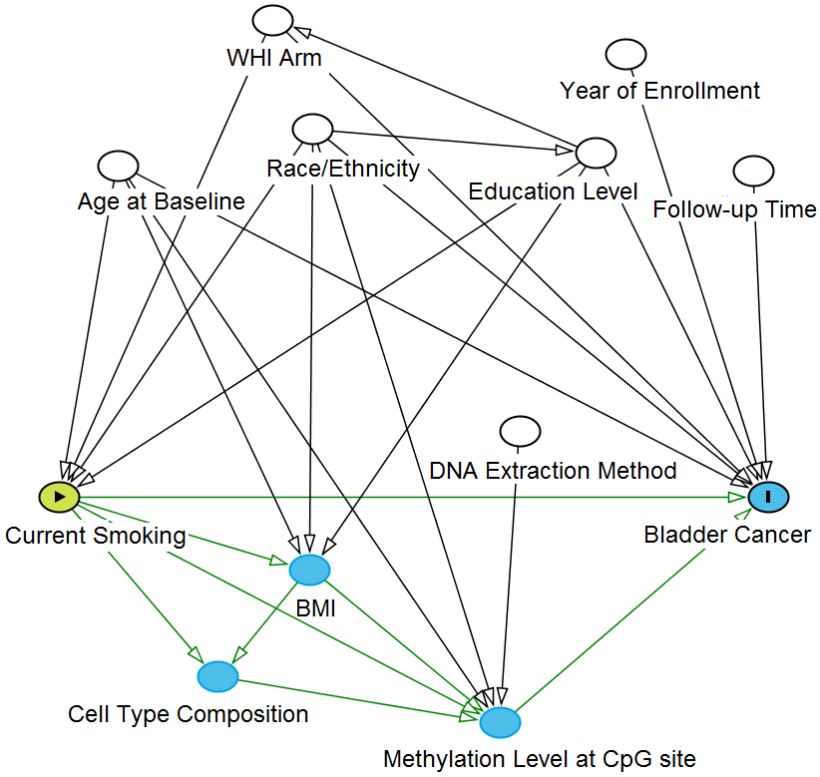

Figure 2.1 Directed acyclic graph (DAG) illustrating potential sources of confounding for our mediation analyses, where variables shown as white circles are included in the set of adjustment covariates

2.2.7 Statistical Analyses

To account for an interaction between the exposure and the mediator, we used the four-way decomposition method, which addresses interaction and mediation simultaneously and provides the most easily interpretable mediation results in the presence of an important additive exposure-mediator interaction (T. VanderWeele 2015). In this study, an interaction occurs when the effect of the methylation change at the CpG-site mediator depends the participant's smoking status. This regression-based method deconstructs the total effect (TE) of smoking into four components, which are the relative risks for the controlled direct effect (CDE), the reference interaction (INT_{ref}), the mediated interaction (INT_{med}), and the pure indirect effect (PIE) (T. VanderWeele 2015). As shown in Table 2.1, these causal effects of smoking on bladder cancer risk capture all possible combinations of mediation and interaction, where

mediation occurs when the exposure causes the mediator and interaction occurs when the change in the mediator is necessary for the exposure to have an effect (T. VanderWeele 2015).

Table 2.1 Description of causal effects for the four-way decomposition approach to mediation analysis

Effect	Mediation	Interaction
CDE	No	No
INT_{ref}	No	Yes
INT_{med}	Yes	Yes
PIE	Yes	No

Using the coefficients from the regression models in equations (2.1) and (2.2), VanderWeele derived equations (2.3), (2.4), (2.5), and (2.6) for estimating the relevant causal effects as components of the excess relative risk (*ERR*) of bladder cancer associated with smoking, where the ERR^{TE} is defined on the ratio scale as the relative risk for the total effect of smoking on bladder cancer minus one. In these equations, we set smoking to an exposed level of 30 pack-years of smoking as compared to an unexposed level of zero pack-years of smoking, and m^* represents setting the mediator to the level observed in non-smoking controls. We chose the exposure level based on evidence from clinical trials related to lung cancer screening, which indicate that 30 pack-years of smoking is clinically relevant for cancer outcomes (Moyer 2014). To calculate the proportion of *ERR*, each component of the *ERR* is divided by the total *ERR* (ERR^{TE}). Confidence intervals and p-values were calculated for the component and proportion estimates based on the delta method (T. VanderWeele 2015).

$$\begin{aligned} ERR^{CDE} \approx & \exp\left\{30\theta_1 + \theta_2 m^* + 30\theta_3 m^* - \theta_2\left(\beta_0 + \beta_2' c\right) + \tfrac{1}{2}\theta_2^2\sigma^2\right\} \\ & - \exp\left\{\theta_2 m^* - \theta_2\left(\beta_0 + \beta_2' c\right) - \tfrac{1}{2}\theta_2^2\sigma^2\right\} \end{aligned} \tag{2.3}$$

$$\begin{aligned} ERR^{INT_{ref}} \approx & \exp\left\{30\left(\theta_1 + \theta_3\left(\beta_0 + \beta_2' c + \theta_2\sigma^2\right)\right) + 450\theta_3^2\sigma^2\right\} - 1 \\ & - \exp\left\{30\theta_1 + \theta_2 m^* + 30\theta_3 m^* - \theta_2\left(\beta_0 + \beta_2' c\right) - \tfrac{1}{2}\theta_2^2\sigma^2\right\} \\ & + \exp\left\{\theta_2 m^* - \theta_2\left(\beta_0 + \beta_2' c\right) - \tfrac{1}{2}\theta_2^2\sigma^2\right\} \end{aligned} \tag{2.4}$$

$$\begin{aligned} ERR^{INT_{med}} \approx & \exp\left\{30\left(\theta_1 + \theta_2\beta_1 + \theta_3\left(\beta_0 + 30\beta_1 + \beta_2' c + \theta_2\sigma^2\right)\right) + 450\theta_3^2\sigma^2\right\} \\ & - \exp\left\{30\theta_2\beta_1\right\} - \exp\left\{30\left(\theta_1 + \theta_3\left(\beta_0 + \beta_2' c + \theta_2\sigma^2\right)\right) + 450\theta_3^2\sigma^2\right\} + 1 \end{aligned} \tag{2.5}$$

$$ERR^{PIE} \approx \exp\left\{30\theta_2\beta_1\right\} - 1 \tag{2.6}$$

The four causal effects are calculated based on specific levels of the mediator and covariates. We set the indicator variables for each category of race/ethnicity, education level, WHI study arm, and DNA extraction method equal to the proportions reported among controls in Table 2.2. Similarly, we set age (65.46), follow-up time in days (5131.51), and year of enrollment (2.38) to their averages in controls. We restricted to non-smoking controls to calculate average M-values in the absence of smoking-related hypomethylation, and used an m^* of 3.98 for cg05575921, 1.55 for cg03636183, and 3.84 for cg19859270.

Within this framework, the ERR^{CDE} captures the excess risk of bladder cancer attributable to the direct effect of smoking if there was no smoking-CpG interaction and the mediator was fixed to m^*, while controlling for covariates. The ERR^{INTref} captures the excess risk of bladder cancer attributable to the change in the effect of smoking that would be observed only if smoking and mediator hypomethylation occurred together when the mediator is fixed to m^*, while controlling for covariates. The ERR^{INTmed} captures the excess risk of bladder cancer attributable both to the change in the effect of smoking that would be observed only if smoking and mediator hypomethylation occurred together and to the changes in mediator methylation caused by smoking, while controlling for covariates. The ERR^{PIE} captures the excess risk of bladder cancer that would be observed if the changes in mediator methylation caused by smoking occurred in never smokers, while controlling for covariates.

These analyses were conducted using SAS® software (version 9.3, SAS Institute Inc., Cary, NC, USA) using four-way decomposition code published by VanderWeele (T. VanderWeele 2015).

2.3 RESULTS

Table 2.2 shows the distribution of relevant variables by case-control status. A much higher proportion of cases were current smokers and the methylation levels of the mediators were lower among cases. In addition, compared to controls, a greater proportion of cases were white.

Table 2.2 Distribution of relevant demographic and clinical characteristics among bladder cancer cases and controls nested within the Women's Health Initiative

	Cases	Controls
Smoking Status (N, %)		
Never Smoker	153 (74%)	233 (93%)
Current Smoker	53 (26%)	18 (7%)
Mediators, M-values (mean, SD)		
cg03636183	1.38 (0.54)	1.49 (0.36)
cg05575921	3.20 (1.36)	3.83 (0.84)
cg19859270	3.62 (0.49)	3.79 (0.28)
WHI Arm (N, %)		
OS	108 (52%)	131 (52%)
CT: HRT and DM	14 (7%)	16 (6%)
CT: HRT only	26 (13%)	29 (12%)
CT: DM only	58 (28%)	75 (30%)
Age, years (N, %)		
< 50-59	58 (28%)	57 (23%)
60 - 69	85 (41%)	116 (46%)
70 - 79+	63 (31%)	78 (31%)
Year of Enrollment (N, %)		
1994 - 1995	50 (24%)	61 (24%)
1996	74 (36%)	79 (32%)
1997	50 (24%)	65 (26%)
1998	32 (16%)	46 (18%)
Follow-up Time, years (mean, SD)	13.29 (3.89)	14.05 (3.29)
DNA Extraction Method (N, %)		
5-prime	197 (96%)	240 (96%)
Phenol	9 (4%)	11 (4%)
Race (N, %)		
Asian/Pacific Islander	3 (2%)	10 (4%)
Black/African American	12 (6%)	25 (10%)
Hispanic/Latino	4 (2%)	13 (5%)
Non-Hispanic White	186 (90%)	196 (78%)
Other	1 (<1%)	7 (3%)
Education (N, %)		
< High School	9 (4%)	12 (5%)
High School	39 (19%)	60 (24%)
Post-High School Training	76 (37%)	89 (35%)
College Degree	23 (11%)	28 (11%)
Post-College Training	59 (29%)	62 (25%)

Abbreviations: OS = observational study; CT = clinical trials; HRT = hormone therapy clinical trial; DM = dietary modification clinical trial

We also verified previously reported associations between smoking and methylation β-values at cg05575921 (regression coefficient = -0.23; q-value = 1.3e-114), cg03636183 (regression coefficient = -0.13; q-value = 9.2e-66), and cg19859270 (regression coefficient = -0.05; q-value = 1.1e-64). After reference-free adjustment for cell mixture, we observed similar associations between smoking and β-values at cg05575921, cg03636183, and cg19859270. The estimate for the association between current smoking and bladder cancer for 30 pack-years of smoking in current smokers as compared to never smoking was a total relative risk of 4.65.

The results from the mediation analysis for each of cg05575921, cg03636183, and cg19859270 are presented in Table 2.3.

Table 2.3 Estimated mediating effects of selected CpG sites for the association between 30 pack-years of smoking in current smokers as compared to never smoking and bladder cancer risk, using four-way decomposition to account for the interaction between smoking and each CpG site

		Component of *ERR*[1]		Percent of *ERR*[1]	
CpG Site	**Causal Effect of Smoking**	**Estimate**	**95% CI**	**Estimate**	**P-value**
cg05575921	CDE	0.18	(-1.94, 2.29)	5%	0.86
	INT_{ref}	0.09	(0, 0.17)	3%	0.05
	INT_{med}	2.29	(0.02, 4.55)	72%	0.02
	PIE	0.64	(-0.38, 1.66)	20%	0.21
	TE	3.19	(0.43, 5.95)	100%	-
cg03636183	CDE	2.06	(-1.47, 5.58)	77%	0.09
	INT_{ref}	0.03	(-0.15, 0.21)	1%	0.74
	INT_{med}	0.90	(-1.16, 2.95)	34%	0.45
	PIE	-0.32	(-0.60, -0.03)	-12%	0.15
	TE	2.67	(0.21, 5.13)	100%	-
cg19859270	CDE	0.46	(-1.66, 2.58)	17%	0.64
	INT_{ref}	0.10	(-0.03, 0.23)	4%	0.13
	INT_{med}	1.89	(0.04, 3.75)	72%	0.04
	PIE	0.18	(-0.38, 0.74)	7%	0.52
	TE	2.63	(0.24, 5.03)	100%	-

[1]Mediation models were adjusted for race/ethnicity, education, WHI arm, age at baseline, year of enrollment, follow-up time, and DNA extraction method
Abbreviations: *ERR* = excess relative risk; CDE = controlled direct effect; INT_{ref} = reference interaction; INT_{med} = mediated interaction; PIE = pure indirect effect; TE = total effect

As derived from Table 2.3, the relative risk estimates for the total effect on bladder cancer for 30 pack-years of smoking in current smokers as compared to never smoking varied somewhat between the models from 3.63 to 4.19 (all $p \leq 0.004$). Most of the *ERR* associated with 30 pack-years of smoking in current smokers as compared to never smoking was mediated through cg05575921 [92% (INT_{med} + PIE)] and cg19859270 [79% (INT_{med} + PIE)]. There was little evidence that smoking is mediated through cg03636183.

For cg03636183, the largest component of the *ERR* associated with current smoking was the controlled direct effect (CDE = 2.06, p = 0.25; 77%, p = 0.09), which captures the effect of smoking on bladder cancer risk that does not interact with cg03636183 and is not mediated by cg03636183. Interestingly, the pure indirect effect of cg03636183 appeared to be protective in the absence of smoking (PIE component = -0.32, p = 0.03; -12%, p = 0.15).

The largest component of the *ERR* of bladder cancer due to current smoking was the mediated interaction for both cg05575921 (INT_{med} = 2.29, p = 0.05; 72%, p = 0.02) and cg19859270 (INT_{med} = 1.89, p-value = 0.05; 72%, p-value = 0.04). This captures the effect of smoking through differential methylation of the mediator that is both caused by smoking and has effects on bladder cancer that depend on the presence of smoking.

2.4 DISCUSSION

The estimate for the total relative risk of bladder cancer associated with 30 pack-years of smoking in current smokers as compared to never smoking was exactly the same as the adjusted hazard ratio for current smoking and bladder cancer reported for women in the demographically similar NIH-AARP Diet and Health study (Freedman et al. 2011b). These estimates are comparable because our calculations are based on 30 pack-years of smoking, which is close to our population average of 38 pack-years among current smokers. The total relative risk estimates in the mediation analyses were different from 4.65 and from each other because they are based on specific levels of the covariates and of m^*, which is the M-value where the mediator is considered absent. The effect sizes for the association between smoking and methylation levels at our selected loci were consistent with those reported by previous methylome-wide

studies (Joehanes et al. 2016; Zeilinger et al. 2013; Tsaprouni et al. 2014; Shenker, Polidoro, et al. 2013; Sun et al. 2013; Wan et al. 2012; Harlid et al. 2014).

Overall, our study indicates that the effects of smoking are substantially mediated through cg05575921 and cg19859270 and that mediation through cg03636183 is negligible. To our knowledge, there are no prior studies of smoking and bladder cancer that quantify mediation through CpG sites. Based on the current literature, there is particularly strong support for a smoking-related biological mechanism involving mediation through cg19859270. Compared to non-smokers, smokers had decreased methylation at cg19859270, as observed in our study and reported by previous methylome-wide studies of smoking (Joehanes et al. 2016; Zeilinger et al. 2013; Tsaprouni et al. 2014; Sun et al. 2013; Wan et al. 2012; Harlid et al. 2014). cg19859270 is located in the first exon of *GPR15,* a gene coding for a chemoattractant receptor that regulates migration of T-cells and that may promote chronic inflammation (Kõks et al. 2015). According to recent research, methylation of the first exon is closely associated with transcriptional silencing and may be even more strongly associated with low gene expression than promoter methylation (Brenet et al. 2011). As a result, hypomethylation of *GPR15* may increase *GPR15* expression in blood from smokers.

Rather than an active causal change in methylation at cg19859270, there is evidence suggesting that the methylation change passively reflects a shift in the immune system. Specifically, smoking may trigger an inflammation response that substantially increases the proportion of T-cells expressing GPR15 (GPR15+ T-cells). In a study by Bauer et al., these GPR15+ T-cells were detected as an overall decrease in *GPR15* methylation and increase in *GPR15* expression in the blood of smokers (Bauer et al. 2015). In fact, the Bauer et al. study demonstrated that smoking was no longer significantly associated with decreased methylation of cg19859270 after adjustment for GPR15+ T-cell subtype (Bauer et al. 2015). This highly significant differential expression of *GPR15* in the blood of current as compared to never smokers was later replicated (fold-change = 5.8, q-value = 0.004) (Kõks et al. 2015). This inflammation response may contribute to tumor initiation and progression through a variety of mechanisms, including the promotion of mutations and neoplastic growth (Hanahan and Weinberg 2011; Qian and Pollard 2010). Further, since the causal effect of smoking through cg19859270 was observed as a mediated interaction,

our results suggest that a pro-inflammatory change involving GPR15+ T-cells is particularly carcinogenic in the context of other effects related to current smoking.

The cg05575921 locus occurs in the *AHRR* gene and was found to be hypomethylated among smokers as compared to nonsmokers in our study and in prior methylome-wide studies of smoking (Joehanes et al. 2016; Zeilinger et al. 2013; Tsaprouni et al. 2014; Shenker, Polidoro, et al. 2013). Since *AHRR* is expressed across all tissues (Fasanelli et al. 2015) and smoking appears to have a similar effect on *AHRR* methylation across a variety of tissue types (Shenker, Polidoro, et al. 2013), the hypomethylation of cg05575921 observed in the blood of smokers may also occur in bladder tissue. The cg05575921 locus is located in an enhancer-like regulatory element within *AHRR*, where we would expect hypomethylation to be associated with enhancer activation and increased *AHRR* expression (Qu et al. 2017). This relationship is consistent with The Cancer Genome Atlas (TCGA) data, which shows a positive correlation between cg05575921 hypomethylation and *AHRR* expression in normal bladder tissue (Spearman correlation = 0.36) (Díez-Villanueva, Mallona, and Peinado 2015).

Though *AHRR* is a putative tumor suppressor gene (Zudaire et al. 2008), its expression down-regulates the aryl hydrocarbon receptor (AHR) and may have harmful effects on bladder tissue in the presence of persistent AHR activation caused by smoking. Through AHR suppression, AHRR may increase inflammation by reducing the expression of anti-inflammatory proteins like COX2 (Awji et al. 2015) or by enhancing T-cell production of IFN-γ (Brandstätter et al. 2016) in chronically inflamed tissue. This proposed mechanism is consistent with the strong mediated interaction of approximately 72% that we observed in our analyses, since the effects of the methylation change at cg05575921 would depend on the continued presence of smoking.

Even though smoking was highly associated with hypomethylation of cg03636183, we found no evidence that the association between smoking and bladder cancer risk is mediated through cg03636183. We do note, however, that this locus is in the north shore of an intrageneic CpG island in *F2RL3*, which encodes the thrombin protease-activated receptor-4 protein (PAR-4) (Gao et al. 2015). PAR-4 is a G-protein coupled receptor that, after activation by thrombin, increases leukocyte recruitment, extravasion, and migration to inflammatory sites (Vergnolle et al. 2002).

A wide variety of well-powered studies have described the association between smoking and genome-wide DNA methylation, but none have evaluated whether changes in DNA methylation underlie the increased risk of bladder cancer associated with smoking. Our study is the first to use mediation analyses to investigate smoking-associated CpG sites as potential mediators of the association between smoking and risk of bladder cancer. We are uniquely positioned to conduct mediation analyses, since our study is nested in a prospective cohort that includes pre-diagnostic blood samples and comprehensive baseline information on potential confounders. However, the sample size for these mediation analyses was relatively limited because there were not many current smokers in the WHI and former smokers were not able to be included.

Conditional on baseline covariates, mediation analysis requires four key assumptions of no-confounding of the exposure-outcome, mediator-outcome, and exposure-mediator relationships, and also requires that there are no mediator-outcome confounders caused by the exposure. To identify possible confounders of the exposure-outcome, mediator-outcome, and exposure-mediator relationships, we carefully reviewed the literature, constructed a detailed DAG for these analyses (Figure 2.1), and adjusted our models for all relevant covariates. Though BMI and cell type composition had the potential to violate the last assumption, they were not considered mediator-outcome confounders because neither has been consistently associated with bladder cancer risk (Larsson et al. 2008; Bhaskaran et al. 2014; Reeves et al. 2007; Holick et al. 2007; Song et al. 2014; Koestler et al. 2016).

We acknowledge error in our measurement of smoking and of methylation as another possible source of bias. This error was introduced since pack-years were calculated based on categories of average cigarettes per day and of years as a regular smoker. Also, we only have a single measure of DNA methylation at baseline that may imprecisely capture the methylation levels for the period that is etiologically relevant for our cases. The errors in our measures of smoking and methylation are likely to be non-differential and are expected to attenuate the reported direct and indirect effect estimates. However, it is possible that methylation changes more accurately capture the smoking exposure than self-report information from category-based questions, which would mean that our indirect effects may include some of the direct effects of smoking and may be somewhat overstated.

We also note that the methylation changes at each of cg05575921, cg03636183, and cg19859270 are closely associated with smoking and, as a result, are highly correlated with each other. As a result, the mediated causal effects estimated from the individual models for these CpG sites likely capture many of the same causal paths. Since we modelled the CpG sites individually, we have a limited ability to combine these markers. While a mediation analysis including all three CpG sites in the same model would be useful, methodology for extending the four-way decomposition approach to allow for multiple mediators has not yet been described. When interpreting our results, it is also important to keep in mind that m^* is fixed to a specific level that is not the maximum observed M-value, so the CDE picks up some of the effect of the INT_{ref}.

Despite the likely importance of the GPR15+ T-cell subtype for the smoking-associated differential methylation of cg19859270, the association between current smoking and cg19859270 methylation was only slightly attenuated after adjustment for cell mixture. However, this result should be interpreted with caution, since available cell type adjustment methods are limited and currently lack reference data for white blood cell subtypes.

We observed that a large proportion of the effect of current smoking on bladder cancer is mediated through methylation changes at cg05575921 and cg19859270, which are particularly harmful in the context of continued smoking. These results may indicate the promotion of chronic inflammation through a higher proportion of GPR15+ T-cells in blood or through increased expression of *AHRR* in bladder tissue. Further investigation of these possible mechanisms has the potential to expand our understanding of the relationship between bladder cancer and smoking, which is its strongest known risk factor.

Chapter 3. DNA METHYLATION CHANGES AS MEDIATORS FOR KNOWN ASSOCIATIONS OF GERMLINE GENETIC VARIANTS AND BLADDER CANCER RISK IN THE WOMEN'S HEALTH INITIATIVE

3.1 INTRODUCTION

Bladder cancer has been the subject of multiple well-powered genome-wide association studies (GWAS), which have identified a robust set of single nucleotide polymorphisms (SNPs) associated with risk of bladder cancer (Figueroa et al. 2014). As is the case for most complex diseases, the mechanisms underlying the associations between these SNPs and bladder cancer risk remain largely unknown. We propose that mediation through proximal changes in DNA methylation may provide information about the downstream effects of these SNPs.

Our study focuses on rs798766, rs401681, rs2294008, and rs8102137. These SNPs have been associated with bladder cancer in GWAS and are also methylation quantitative loci (mQTL), since they affect patterns of DNA methylation at nearby CpG sites (Smith et al. 2014). Based on queries of the Accessible Resource for Integrated Epigenomic Studies (ARIES) mQTL database (Gaunt et al. 2016), 50% of the eight SNPs associated with bladder cancer were mQTL in blood, as compared to only 34% of total SNPs across the entire genome. This is consistent with GWAS results for other complex traits, which are often enriched for *cis* mQTL (Gaunt et al. 2016), suggesting that mQTL SNPs associated with complex diseases like bladder cancer may have effects caused or reflected by differential DNA methylation. Our study is the first to formally investigate the potential mediation of the association between mQTL SNPs and bladder cancer risk through methylation changes in blood at mQTL-associated CpG sites. This study represents an important step toward integrating genetic and epigenetic data in the context of bladder cancer and toward improving our understanding of the mechanisms related to genetic variants involved in susceptibility to bladder cancer.

3.2 Materials and Methods

3.2.1 Study Participants

Data for the current analyses were drawn from our case-control study of pre-diagnostic DNA methylation and bladder cancer that was nested in the Women's Health Initiative (WHI) (Jordahl et al. 2018). Briefly, the WHI includes 161,808 postmenopausal women recruited from 1993 to 1998 across the US (Hays et al. 2003). There are two arms of the WHI, which include the clinical trials (CT) arm and the observational study (OS) arm. The CT arm involved concurrent randomized controlled trials of hormone therapy (HRT), dietary modification (DM), and, subsequently, calcium/vitamin D supplementation (CaD). Those not eligible or willing to participate in these clinical trials were asked to participate in the OS. After 2005, WHI participants were invited to enroll in the WHI Extension Studies, which tracked health outcomes for another 10 years. Follow-up time was calculated as the time between enrollment and death or last contact.

Within the WHI, we identified 440 women who were diagnosed with transitional cell carcinoma of the bladder during study follow-up and selected 440 cancer-free controls matched on year of enrollment, age at enrollment (+/- 2 years), follow-up time greater than or equal to their matched case, trial component, and DNA extraction method (Jordahl et al. 2018).

3.2.2 Data and Biospecimen Collection

Basic demographic information, including age and race/ethnicity, was requested during the screening process (Hays et al. 2003). Through baseline questionnaires, participants reported if they had smoked at least 100 cigarettes ever (yes, no), currently smoked cigarettes (yes, no), or had ever smoked to lose weight (yes, no); this information was used to determine smoking status. Questionnaires also included items about the average number of cigarettes currently or previously smoked per day and the number of years as a regular smoker, which were used to calculate pack-years of smoking. Blood samples were collected at baseline after at least 12 hours of fasting and stored at -70°C as buffy coats.

3.2.3 DNA Methylation Array

As previously described, we used the Illumina 450K Infinium HumanMethylation Bead Array to interrogate methylation status at approximately 485,577 CpG sites among bladder cancer cases and controls (Jordahl et al. 2018). We used the M-value to measure methylation at each CpG site to improve the heteroscedasticity of methylation levels (Du et al. 2010). The M-value is the base-2 logit of the β-value, where the β-value is the ratio of the methylated signal over the total signal and can be interpreted as the percent of methylation at a specific site (Dedeurwaerder et al. 2014). After reading in the raw image files and checking for failed samples, we performed background correction and functional normalization of the methylation data (Jordahl et al. 2018). The CpG sites included in the current analyses passed our quality control (QC) steps, which excluded any CpG sites that were undetected in at least 10% of samples, had a beadcount less than 3 in at least 10% of samples, were in or near common SNPs (minor allele frequency greater than 1%), were classified as cross-reactive probes, or were located on the sex chromosomes (Jordahl et al. 2018).

3.2.4 Identifying mQTL SNPs Associated with Bladder Cancer

A meta-analysis of bladder cancer GWAS conducted in 2014 combined all previous studies to verify known loci and reported genome-wide significant associations ($p < 5e-8$) for eight SNPs: rs710521, rs798766, rs401681, rs1495741, rs2294008, rs9642880, rs8102137, and rs1014971 (Figueroa et al. 2014). The meta-analysis included an average of 22 studies involving an average of 11,131 cases and 50,634 controls for each reported SNP association. We used the ARIES mQTL database to assess whether each of these eight SNPs was also an mQTL (Gaunt et al. 2016). This publicly available dataset includes approximately 1,000 mother-offspring pairs from the Avon Longitudinal Study of Parents and Children (ALSPAC). Associations between each of 8,282,911 directly genotyped or imputed common SNPs and each of 395,625 CpG sites on the Illumina Infinium HumanMethylation450 Bead Array that passed their QC are provided in the mQTL database. To most closely match the population of older women included in the WHI, we only used the methylation data measured in peripheral blood samples collected from mothers at the middle age time point (mean age = 47.45; n = 742). For each SNP, we searched for associated CpG sites in the MatrixEQTL database and then restricted to SNP-CpG results

with a p-value below 1e-14 to control the false positive rate at 0.2% (Gaunt et al. 2016). We also restricted to CpG sites that passed our QC steps, leaving four mQTL that were associated with methylation changes at a total of ten CpG sites. Notably, the risk allele is the minor allele for all selected SNPs except for rs401681.

3.2.5 SNP Genotyping

Taqman® SNP Genotyping Assays were used to determine the genotypes for each mQTL SNP. For the assay, 5 ng of genomic DNA was aliquoted into 384-well plates and dried down. Each assay was combined with Taqman® Genotyper master mix, run under universal PCR conditions, and analyzed with the ABI 7900 HT Taqman® Real-Time instrument.

Prior to running the study samples, 90 samples representing 30 parent-parent-child trios from a population of Utah residents with European ancestry (CEPH) were genotyped to assess performance of the genotyping assays. Assay accuracy was verified by comparing genotypes to publicly available genotype data for these samples from SNP500 Cancer (Packer et al. 2004) and dbSNP (Sherry et al. 2001) and by assessing inheritance errors. As another quality control step, two external control samples from the HapMap project were included on each plate to confirm reliability and reproducibility of the genotyping across the study plates. Inter-plate duplicates were also included. Samples with weak signals and outliers were repeated once. There was 100% concordance with all duplicates. All SNPs had a greater than 99% call rate and were in Hardy-Weinberg equilibrium.

3.2.6 Statistical Analyses

Mediation analyses were conducted with each of the four mQTL (rs798766, rs401681, rs2294008, and rs8102137) as the exposure, all corresponding CpG sites as the mediators, and incident bladder cancer as the outcome. We estimated causal effects using a regression-based approach for dichotomous outcomes (T. VanderWeele and Vansteelandt 2014; T. J. VanderWeele and Vansteelandt 2010), which is based on a counterfactual framework for causal inference. This method uses a logistic regression model for the outcome (3.1) and a linear regression model for the mediator
(3.2), where *A* represents the exposure, *M* represents the mediator(s), and *Y* represents the outcome of

interest. Based on our study design, the logistic regression model was fit among cases and controls (n = 836) and the linear regression model was fit only among controls (n = 424).

$$\text{logit}\left(P\left(Y=1|\, a,\boldsymbol{m},\boldsymbol{c}\right)\right)=\theta_0+\theta_1 a+\sum_{i=1}^{K}\theta_2^{(i)}m^{(i)}+\theta_4'\boldsymbol{c} \tag{3.1}$$

$$E[M^{(i)}\mid a,\boldsymbol{c}]=\beta_0^{(i)}+\beta_1^{(i)}a+\beta_2^{(i)\prime}\boldsymbol{c}\ \text{ for } i=1,\ldots,K \tag{3.2}$$

The models were adjusted for a set of covariates (***c***) that included our matching variables and potential confounders. Specifically, we adjusted our analyses for WHI arm (HRT/DM, HRT, DM, OS), age at baseline (continuous), year of enrollment (continuous: 1994-1995 = 1, 1996 = 2, 1997 = 3, 1998 = 4), follow-up time (continuous), DNA extraction method (5-prime, phenol), race/ethnicity (Asian/Pacific Islander, black/African American, Hispanic/Latino, non-Hispanic white, other), smoking status (never, former, current), and pack-years of smoking (continuous).

With the coefficients from regression models (3.1) and (3.2), equations (3.3) and (3.4) were used to estimate the odds ratio (*OR*) per risk allele for the natural direct effect (NDE) and the natural indirect effect (NIE) (T. VanderWeele and Vansteelandt 2014).

$$OR^{NDE}=\exp\left(\theta_1\right) \tag{3.3}$$

$$OR^{NIE}=\exp\left(\sum_{i=1}^{K}\beta_1^{(i)}\theta_2^{(i)}\right) \tag{3.4}$$

Effect estimates and confidence intervals were estimated using the bootstrapping approach based on 1000 bootstrap samples. The NDE captures the effect of the exposure on the outcome that does not act through the CpG site mediators, while the NIE captures the effect of the exposure on the outcome that acts through the CpG site mediators. The total effect (TE) is the sum of the NDE and NIE, although this relationship is not exact when using the bootstrapping approach. The method allows the effects of individual mediators to cancel one another, since each SNP affects its associated CpG sites simultaneously that may mediate an increase in bladder cancer risk if the direction of effect is consistent.

To improve understanding of the mechanisms underlying any observed NIE, we also conducted exploratory mediation analyses for individual SNP-CpG pairs, stratified by smoking status (never, ever). The models are based on the same equations described above (T. J. VanderWeele and Vansteelandt

2010). For these models, however, effect estimates and confidence intervals were calculated using the delta method (T. J. VanderWeele and Vansteelandt 2010).

The mediation analyses were conducted based on a macro provided by Valeri and VanderWeele (Valeri and VanderWeele 2013) using SAS® software (versions 9.3 and 9.4, SAS® Institute Inc., Cary, NC, USA). For multiple mediators, the macro was modified to produce multiple-mediator effect estimates based on existing equations (T. VanderWeele and Vansteelandt 2014) and to provide bootstrap-based confidence intervals.

3.3 RESULTS

Table 3.1 provides the distribution of demographic and clinical characteristics by case-control status for our study population. Compared to controls, cases had a greater proportion of women with white race/ethnicity, were more likely to be past and current smokers, and were more likely to have a heavier smoking history of at least 30 pack-years. The average follow-up time was slightly shorter for cases than for controls.

Table 3.1 Distribution of relevant demographic and clinical characteristics among bladder cancer cases and controls nested within the Women's Health Initiative

	Cases	Controls
WHI Arm (N, %)		
OS	213 (52%)	218 (52%)
CT: HRT and DM	24 (6%)	26 (6%)
CT: HRT only	55 (13%)	56 (13%)
CT: DM only	120 (29%)	124 (29%)
Age (N, %)		
< 50 - 59	97 (24%)	100 (24%)
60-69	194 (47%)	203 (48%)
70 - 79+	121 (29%)	121 (28%)
Year of Enrollment (N, %)		
1994 - 1995	105 (25%)	109 (26%)
1996	125 (30%)	127 (30%)
1997	114 (28%)	116 (27%)
1998	68 (17%)	72 (17%)

Continued on page 40

Table 3.1 Distribution of relevant demographic and clinical characteristics among bladder cancer cases and controls nested within the Women's Health Initiative (continued)

	Cases	Controls
Follow-up Time (mean, SD)	13.37 (3.77)	13.95 (3.36)
DNA Extraction Method (N, %)		
5-prime	397 (96%)	408 (96%)
Phenol	15 (4%)	16 (4%)
Race (N, %)		
Asian/Pacific Islander	4 (1%)	12 (3%)
Black/African American	24 (6%)	40 (9%)
Hispanic/Latino	7 (2%)	16 (4%)
Non-Hispanic White	375 (91%)	349 (82%)
Other	2 (<1%)	7 (2%)
Smoking (N, %)		
Never Smoked	155 (38%)	234 (55%)
Past Smoker	204 (49%)	172 (41%)
Current Smoker	53 (13%)	18 (4%)
Pack-Years (N, %)		
Never Smoker	155 (38%)	234 (55%)
> 0 - < 30	145 (35%)	137 (32%)
≥ 30	112 (27%)	53 (13%)

Abbreviations: OS = observational study; CT = clinical trials; HRT = hormone therapy clinical trial; DM = dietary modification clinical trial

For each mQTL, there was a range of one to five mQTL-associated CpG sites (Table 3.2). All of our selected mQTL were *cis*-acting, since they were within 1 Mb of their associated CpG sites. According to ARIES, these mQTL are common SNPs with risk allele frequencies ranging from 0.20 to 0.56 and are mostly located in or near CpG islands. These SNPs have effect sizes ranging from a 0.4% to 3% difference in median proportion methylated between homozygote groups. These SNPs are associated with an increase in methylation for three CpG sites and a decrease in methylation for seven CpG sites.

Table 3.2 Annotation information for selected mQTL and their associated CpG sites

SNP	SNP Location	RA (RAF)	CpG	CpG Location	CpG Gene	CpG Gene Group[a]	Relation to CpG Island[b]	ARIES Direction[c]	ARIES Effect Size[d]
rs798766	chr4:1734239	T (0.20)	cg00006948	chr4:1768889			Island	+	0.004
rs401681	chr5:1322087	C (0.56)	cg27028750	chr5:1349422			S_Shelf	+	0.059
			cg26209169	chr5:1316264				-	0.038
rs2294008	chr8:143761931	T (0.45)	cg06565975	chr8:143823917	*SLURP1*	TSS200	S_Shelf	-	0.041
			cg03405983	chr8:143858548	*LYNX1*	5'UTR TSS200 1stExon	Island	-	0.009
			cg24023258	chr8:143781297	*LY6K*	TSS1500	N_Shore	+	0.004
			cg17888033	chr8:143858414	*LYNX1*	5'UTR 1stExon	Island	-	0.020
			cg17252645	chr8:143867129	*LY6D*	Body		-	0.042
rs8102137	chr19:30296853	C (0.33)	cg16836589	chr19:30303674	*CCNE1*	1stExon Body	Island	-	0.011
			cg27475126	chr19:30303651	*CCNE1*	1stExon Body	Island	-	0.049

[a]Functional region of gene as indicated in v1.2 Illumina annotation: TSS1500 = 200-1500 bases upstream of the transcription start site; TSS200 = 0-200 bases upstream of the transcription start site; 5'UTR = Within the 5 prime untranslated region; 1stExon = First segment of gene coding for peptide sequence; Body = Between the ATG and stop codon; 3'UTR = Between the stop codon and poly A signal; Multiple listings indicate a locus in a region with multiple splice variants

[b]Position relative to CpG island as indicated in v1.2 Illumina annotation: Island = Within CpG island (CG content > 50%, Obs/Exp CpG ratio > 0.60, and length > 200 bps); OpeanSea = Non-island region; Shore = 0-2 kb flanking CpG Island; Shelf = 2-4 kb flanking CpG Island

[c]Direction of association between mQTL and CpG site reported in the ARIES mQTL database

[d]Effect size between mQTL and CpG site reported in the ARIES mQTL database, where effect size is defined as the difference in median proportion methylated between homozygote groups

Abbreviations: RA = risk allele; RAF = risk allele frequency

The results from the mediation analyses for rs798766, rs401681, rs2294008, and rs8102137 are presented in Table 3.3. Only rs8102137 had a statistically significant total effect, and none of the indirect effects were statistically significant. However, our results suggest that a large proportions of the modest effects of rs401681 (NIE = 1.05; NIE percent = 98.5%) and of rs2294008 (NIE = 1.10; NIE percent = 77.6%) on bladder cancer risk occur through mQTL-associated CpG sites. For the rs401681 NIE, women with C risk allele at rs401681 and C-allele-associated methylation levels at cg27028750 and cg26209169 have a 5% increased risk of bladder cancer as compared to women with the C allele and T-allele-associated methylation levels at cg27028750 and cg26209169, while adjusting for covariates. The NIE interpretation is similar for rs2294008 based on a T risk allele, 10% increased risk of bladder cancer, and cg06565975, cg03405983, cg24023258, cg17888033, and cg17252645 as mediators. There was little evidence supporting mediation through changes in DNA methylation for the associations of rs8102137 and rs798766 with bladder cancer risk.

Table 3.3 For each mQTL SNP identified by genome-wide association studies of bladder cancer, estimated mediating effects of mQTL-associated CpG site(s) for the association between the SNP and bladder cancer risk using mediation models that allow for multiple mediators

SNP	Region	Gene	RA	Mediator(s)	Effect of SNP[1]	*OR*	95% CI
rs798766	4p16.3	*TMEM129-TACC3-FGFR3*	T	cg00006948	NDE	1.16	(0.86, 1.50)
					NIE	0.95	(0.81, 1.08)
					TE	1.10	(0.85, 1.38)
rs401681	5p15.33	*TERT-CLPTML*	C	cg27028750 cg26209169	NDE	1.01	(0.76, 1.33)
					NIE	1.05	(0.89, 1.25)
					TE	1.05	(0.85, 1.30)
rs2294008	8q24.3	*PSCA*	T	cg06565975 cg03405983 cg24023258 cg17888033 cg17252645	NDE	1.04	(0.78, 1.35)
					NIE	1.10	(0.90, 1.33)
					TE	1.13	(0.91, 1.38)
rs8102137	19q12	*CCNE1*	C	cg16836589 cg27475126	NDE	1.36	(1.05, 1.74)
					NIE	1.02	(0.93, 1.11)
					TE	1.38	(1.09, 1.74)

[1]Mediation models were adjusted for race/ethnicity, smoking status, pack-years of smoking, WHI arm, age at baseline, year of enrollment, follow-up time, and DNA extraction method
Abbreviations: RA = risk allele; CI = confidence interval; NDE = natural direct effect; NIE = natural indirect effect; TE = total effect

Our exploratory analyses in Table 3.4 suggest that the total effects of rs401681 and rs2294008 on bladder cancer risk were stronger among smokers. Based on these analyses, mediation through cg27028750 may contribute more than cg26209169 to the possible indirect effect of rs401681 on bladder cancer risk. For rs2294008, the NIE may be responsible for a larger proportion of the total effect in non-smokers than in smokers. The results also suggest that the effect of rs2294008 is primarily mediated by cg24023258 and cg17252645 among non-smokers and by cg06565975, cg03405983, and cg17888033 among smokers.

Table 3.4 Within the non-smoker and smoker subgroups, estimated mediating effects of each mQTL-associated CpG site in an individual mediation model for the association between the SNP and bladder cancer risk

				Non-Smokers		**Smokers**	
SNP	**Mediator**	**CpG Gene**	**Effect of SNP**[1]	***OR***	**95% CI**	***OR***	**95% CI**
rs401681	cg27028750		NDE	0.91	(0.65, 1.28)	1.10	(0.78, 1.55)
			NIE	1.05	(0.88, 1.26)	1.03	(0.85, 1.23)
			TE	0.96	(0.72, 1.29)	1.13	(0.84, 1.51)
	cg26209169		NDE	0.94	(0.69, 1.30)	1.14	(0.83, 1.56)
			NIE	1.02	(0.89, 1.17)	0.99	(0.87, 1.12)
			TE	0.96	(0.72, 1.29)	1.13	(0.84, 1.51)
rs2294008	cg06565975	*SLURP1*	NDE	1.07	(0.76, 1.50)	1.15	(0.84, 1.59)
			NIE	0.98	(0.82, 1.18)	1.04	(0.88, 1.23)
			TE	1.05	(0.79, 1.40)	1.20	(0.91, 1.58)
	cg03405983	*LYNX1*	NDE	1.12	(0.81, 1.56)	1.15	(0.86, 1.54)
			NIE	0.94	(0.81, 1.08)	1.04	(0.95, 1.14)
			TE	1.05	(0.79, 1.41)	1.19	(0.90, 1.57)
	cg24023258	*LY6K*	NDE	0.91	(0.66, 1.26)	1.20	(0.88, 1.63)
			NIE	1.14	(0.98, 1.32)	1.00	(0.87, 1.14)
			TE	1.04	(0.78, 1.39)	1.19	(0.90, 1.58)
	cg17888033	*LYNX1*	NDE	1.05	(0.76, 1.46)	1.14	(0.84, 1.54)
			NIE	1.00	(0.85, 1.17)	1.05	(0.92, 1.19)
			TE	1.05	(0.79, 1.40)	1.19	(0.90, 1.58)
	cg17252645	*LY6D*	NDE	0.97	(0.69, 1.35)	1.15	(0.83, 1.59)
			NIE	1.08	(0.91, 1.29)	1.04	(0.86, 1.26)
			TE	1.05	(0.79, 1.40)	1.20	(0.91, 1.59)

[1]Mediation models were adjusted for race/ethnicity, smoking status, pack-years of smoking, WHI arm, age at baseline, year of enrollment, follow-up time, and DNA extraction method
Abbreviations: CI = confidence interval; NDE = natural direct effect; NIE = natural indirect effect; TE = total effect

3.4 DISCUSSION

Our results suggest that the effects of germline genetic variation in rs798766 and rs8102137 on bladder cancer risk are unlikely to be mediated by changes in methylation, which is consistent with the putative mechanisms underlying the carcinogenicity of these SNPs. rs798766 is in an intron of encoding transforming, acidic coiled-coil containing protein 3 (*TACC3*) and has been previously linked to activating somatic mutations in nearby fibroblast growth factor receptor 3 (*FGGR3*), which occur in 74% of Ta-stage bladder tumors (Kiemeney et al. 2010). rs8102137 is located 6kb upstream of cyclin E1 gene (*CCNE1*) (Kohaar et al. 2011), which controls cell cycle progression to the S-phase (Siu, Rosner, and Minella 2012) and is elevated in bladder tumors as compared to adjacent normal bladder tissue (Kohaar et al. 2011). Kohaar et al. previously reported a lack of association between the rs8102137 genotype and overall *CCNE1* mRNA levels (Kohaar et al. 2011). Instead of mediation through changes in overall *CCNE1* expression, there is convincing evidence that rs8102137 influences *CCNE1* splicing, since it is associated with increased expression of a *CCNE1* transcript lacking exon 7 in bladder tissue (Y. P. Fu et al. 2013).

Our main analysis suggests that most of the effect of rs401681 on bladder cancer risk occurs indirectly through methylation changes at mQTL-associated CpG sites. The rs401681 SNP is in an intron of the cisplatin resistance related protein CRR9p (*CLPTM1L*) gene. Based on our exploratory analyses, the effect of rs401681 may primarily occur through an increase in methylation at cg27028750, which is not located within a known gene or CpG island. However, cg27028750 is located in a long terminal repeat element (MER-50), and genomic repeats are often targets for Polycomb repression (Leeb et al. 2010). Although data for normal bladder tissue is unavailable, the rs401681 C allele has also been associated with an increased risk of lung cancer (Pintarelli et al. 2017), and publically-available information from the Roadmap Epigenomics project (Zhou and Wang 2012; Zhou et al. 2011) indicates that the region surrounding cg27028750 is weakly repressed by Polycomb-group proteins in lung tissue.

Previous research supports a lack of association between rs401681 and differential gene expression. A study by Rafnar et al. found no association between rs401681 and RNA expression of *TERT* or *CLPTM1L* in blood or adipose tissue (Rafnar et al. 2009). In addition, an expression quantitative trait loci (eQTL) analysis in tumor-adjacent lung tissue failed to identify any associations between rs401681 and expression levels of 10,821 genes (Pintarelli et al. 2017).

Based on these observations, it is possible that rs401681 promotes or reflects cancer-related epigenetic switching, where the repression of genes related to early development and cellular determination shifts from reversible silencing by Polycomb repressive complexes to more permanent silencing by DNA methylation (Gal-Yam et al. 2008; P A Jones et al. 1990; Schlesinger et al. 2007; Widschwendter et al. 2007). This switch is not expected to cause changes in gene expression, but instead affects epigenetic plasticity (Gal-Yam et al. 2008). Recent research suggests these changes predispose cells to cancer, possibly by maintaining a stem-cell-like phenotype that initiates abnormal growth and malignant transformation (Widschwendter et al. 2007).

We also observed that a substantial portion of the effect of the rs2294008 T allele in *PSCA* may act on bladder cancer risk through methylation changes at mQTL-associated CpG sites located in or near four of the other nine genes in the Ly6 cluster on chromosome 8 (*LY6K*, *LY6D*, *LYNX1*, *SLURP1*) (X. W. Fu, Song, and Spindel 2015). *PSCA* is a GPI-anchored cell surface antigen in the Ly6 family (Reiter et al. 1998). Interestingly, the rs2294008 variant has been shown to create an alternative translation start site for *PSCA* that extends the signal peptide from 11 to 20 amino acids (Kohaar et al. 2013).

Like *PSCA*, *LY6K* and *LY6D* are Ly-6 family members that are GPI-anchored proteins expressed on the cell surface (Kong and Park 2012) and show increased expression in bladder cancer (Luo et al. 2016). However, these genes are more clearly oncogenic, since they are consistently upregulated across many cancers (Luo et al. 2016). *LY6K* has been implicated in cell growth, migration, and invasion in bladder cancer cell lines (de Nooij-van Dalen et al. 2003). Its increased expression has been identified as a promising diagnostic biomarker in blood for lung and esophageal carcinomas (Ishikawa et al. 2007; Zhang et al. 2012) and has been associated with reduced five-year overall survival in bladder cancer patients (Luo et al. 2016; Lee et al. 2010). In contrast, *LY6D* is expressed almost exclusively in human squamous and transitional epithelium normal tissues and cancers (Quak et al. 1990), which means it may be important for the transitional cells primarily involved in bladder cancer. Although methylation changes at CpG sites outside of promoter-related CpG islands are extremely difficult to interpret without expression data, we expect that the differential methylation at cg24023258 and cg17252645 is related to increased expression of *LY6K* in blood and bladder tissue and *LY6D* in bladder tissue.

According to our exploratory analyses, rs2294008 also appears to have an effect on bladder cancer through a separate pathway in smokers that involves GPI-anchored Ly6/Neurotoxin 1 (*LYNX1*) and secreted Ly6/uPAR related protein 1 (*SLURP1*). Recent research suggests that *PSCA*, *LYNX1*, and *SLURP1* bind to (Ibañez-Tallon et al. 2002; Moriwaki et al. 2007) and modulate (X. W. Fu, Song, and Spindel 2015) the α7 subunit of nicotinic acetylcholine receptors (α7-nAChRs), which are expressed in normal bladder tissue from rats (Kim et al. 2015; Beckel et al. 2006). There is a well-established connection between nicotine and tumorigenesis through cascades triggered by nAChRs that promote cell proliferation and survival (X. W. Fu, Song, and Spindel 2015; Singh, Pillai, and Chellappan 2011), where *in vivo* experiments primarily implicate α7-nAChR in nicotine-related cancer development and progression (Singh, Pillai, and Chellappan 2011). Specifically, research by Chen at al. suggests that stimulation of nAChRs promotes bladder tumorigenesis through a mechanism involving activation of ERK1/2 and Stat3 signaling, increased *Cyclin D1* expression, and enhanced cell proliferation (R.-J. Chen et al. 2008).

Since *LYNX1* negatively regulates α7-nAChRs (X. W. Fu, Song, and Spindel 2015), methylation changes associated with increased risk of bladder cancer are likely associated with downregulation of *LYNX1*. In fact, *LYNX1* has already been established as a tumor suppressor gene in the context of lung cancer, where it reduces the downstream effects of nAChR signaling by constraining the ability of chronic nicotine exposure to increase nAChR levels in normal lung tissue (X. W. Fu, Song, and Spindel 2015). The methylation changes at cg03405983 and cg17888033 also have a particularly convincing connection to transcriptional silencing (Peter A. Jones 2012; Rakyan et al. 2011; Laird 2010), since hypermethylation occurs at these sites in the chr8:143858279-143859411 CpG island within the putative promoter region of *LYNX1*. The direction of change for *SLURP1* expression is more difficult to predict because *SLURP1* has pro-apoptotic properties, but also positively regulates and is regulated by α7-nAChRs (X. W. Fu, Song, and Spindel 2015; Pettersson et al. 2009; Arredondo, Chernyavsky, and Grando 2007). However, decreased methylation at cg06565975 may cause or reflect changes in *SLURP1* expression.

Our study is the first to examine possible mediation through methylation for top bladder cancer GWAS hits that are also mQTL. We leveraged information from previous GWAS and an existing mQTL database and used a prospective study design to explore the mechanisms underlying known associations between genetic variants and bladder cancer risk. Our sample size was not ideal for examining genetic

effects, since SNPs have relatively small effect sizes. As a result, only rs8102137 had a statistically significant direct and total effect, and we lacked power for many of the other associations in our main and exploratory analyses. We were also limited by the ARIES sample size, which may not have been large enough to detect all relevant mQTL-associated mediators. Despite these limitations, the associations between our selected genetic variants and bladder cancer (OR^{TE} estimates) were similar to those reported previously, and our results provide suggestive evidence of indirect effects through changes in DNA methylation that warrant further exploration in larger-scale studies of bladder cancer.

Conditional on the baseline covariates, the assumptions of mediation analyses require no-confounding of the exposure-outcome, mediator-outcome, and exposure-mediator relationships, and also require that there are no mediator-outcome confounders caused by the exposure. Analyses involving genetic variants are slightly less prone to confounding because, aside from race/ethnicity, genotype is likely unaffected by most factors relevant for methylation or bladder cancer. We also adjusted for a comprehensive set of covariates to address mediator-outcome confounding. We intentionally excluded cell type composition from the adjustment covariates, since it is another possible mediator that contributes to the indirect effects in our analyses.

We also assumed that there was no interaction between the mQTL SNP and its mQTL-associated mediators. We note that there are potential, though non-significant, interactions between rs798766 and cg00006948 and between rs2294008 and cg03405983 that might improve these models in more flexible approaches. However, the inclusion of an interaction term in single-mediator mediation models produced the same conclusions for rs798766 and rs2294008 in the main and relevant exploratory analyses.

As another possible source of bias, we acknowledge error in our measurement of methylation because we only have a single measure of DNA methylation at baseline, which may imprecisely capture the methylation levels for the period that is etiologically relevant for our cases. However, the error in our measures of methylation is likely to be non-differential and is expected to attenuate the reported indirect effect estimates.

Since DNA methylation can be tissue specific, ideally, we would have conducted these analyses based on methylation data from normal bladder tissue rather than from blood. However, in a previous

study of blood and brain tissue, substantial overlap was observed for mQTL from the two sites (Smith et al. 2014), indicating that at least some SNP-CpG relationships may be consistent across tissue type.

Our study suggests that a substantial proportion of the effects of rs401681 and of rs2294008 on bladder cancer risk may be mediated through methylation changes at nearby CpG sites. If confirmed by larger-scale studies, our results point to a connection between rs401681 in *CLPTM1L* and a possible nearby repression shift marked by cg27028750. A robust link between rs2294008 and risk of bladder cancer through cg24023258 may identify *LY6K* expression as a serum biomarker of bladder cancer susceptibility in non-smokers. Among smokers, further investigation of the role of rs2294008 in α7-nAChR signaling is warranted.

REFERENCES

Ambatipudi, Srikant, Cyrille Cuenin, Hector Hernandez-Vargas, Akram Ghantous, Florence Le Calvez-Kelm, Rudolf Kaaks, Myrto Barrdahl, et al. 2016. “Tobacco Smoking-Associated Genome-Wide DNA Methylation Changes in the EPIC Study.” *Epigenomics* 8 (5): 599–618. https://doi.org/10.2217/epi-2016-0001.

American Cancer Society. 2016a. “Cancer Facts and Figures 2016.” Atlanta: American Cancer Society.

———. 2016b. “What Is Bladder Cancer?” 2016. http://www.cancer.org/cancer/bladdercancer/detailedguide/bladder-cancer-what-is-bladder-cancer.

Arredondo, Juan, Alexander I. Chernyavsky, and Sergei A. Grando. 2007. “SLURP-1 and -2 in Normal, Immortalized and Malignant Oral Keratinocytes.” *Life Sciences* 80 (24–25): 2243–47. https://doi.org/10.1016/j.lfs.2007.01.003.

Aryee, Martin J., Andrew E. Jaffe, Hector Corrada-Bravo, Christine Ladd-Acosta, Andrew P. Feinberg, Kasper D. Hansen, and Rafael A. Irizarry. 2014. “Minfi: A Flexible and Comprehensive Bioconductor Package for the Analysis of Infinium DNA Methylation Microarrays.” *Bioinformatics* 30 (10): 1363–69. https://doi.org/10.1093/bioinformatics/btu049.

Awji, Elias G., Hitendra Chand, Shannon Bruse, Kevin R. Smith, Jennifer K. Colby, Yohannes Mebratu, Bruce D. Levy, and Yohannes Tesfaigzi. 2015. “Wood Smoke Enhances Cigarette Smoke–Induced Inflammation by Inducing the Aryl Hydrocarbon Receptor Repressor in Airway Epithelial Cells.” *American Journal of Respiratory Cell and Molecular Biology* 52 (3): 377–86. https://doi.org/10.1165/rcmb.2014-0142OC.

Barbashina, Violetta, Paulo Salazar, Eric C. Holland, Marc K. Rosenblum, and Marc Ladanyi. 2005. “Allelic Losses at 1p36 and 19q13 in Gliomas: Correlation with Histologic Classification, Definition of a 150-Kb Minimal Deleted Region on 1p36, and Evaluation of *CAMTA1* as a Candidate Tumor Suppressor Gene.” *Clinical Cancer Research* 11 (3): 1119–28.

Bauer, Mario, Gunter Linsel, Beate Fink, Kirsten Offenberg, Anne Maria Hahn, Ulrich Sack, Heike Knaack, Markus Eszlinger, and Gunda Herberth. 2015. “A Varying T Cell Subtype Explains Apparent Tobacco Smoking Induced Single CpG Hypomethylation in Whole Blood.” *Clinical Epigenetics* 7 (August): 81. https://doi.org/10.1186/s13148-015-0113-1.

Beckel, Jonathan M., Anthony Kanai, Sun-Ju Lee, William C. de Groat, and Lori A. Birder. 2006. “Expression of Functional Nicotinic Acetylcholine Receptors in Rat Urinary Bladder Epithelial Cells.” *American Journal of Physiology. Renal Physiology* 290 (1): F103–10. https://doi.org/10.1152/ajprenal.00098.2005.

Besaratinia A, Cockburn M, and Tommasi S. 2013. “Alterations of DNA Methylome in Human Bladder Cancer.” *Epigenetics : Official Journal of the DNA Methylation Society* 8 (10): 1013–22.

Besaratinia, Ahmad, and Stella Tommasi. 2013. “Genotoxicity of Tobacco Smoke-Derived Aromatic Amines and Bladder Cancer: Current State of Knowledge and Future Research Directions.” *The FASEB Journal* 27 (6): 2090–2100. https://doi.org/10.1096/fj.12-227074.

Bhaskaran, Krishnan, Ian Douglas, Harriet Forbes, Isabel dos-Santos-Silva, David A Leon, and Liam Smeeth. 2014. “Body-Mass Index and Risk of 22 Specific Cancers: A Population-Based Cohort Study of 5·24 Million UK Adults.” *The Lancet* 384 (9945): 755–65. https://doi.org/10.1016/S0140-6736(14)60892-8.

Bibikova, Marina, Bret Barnes, Chan Tsan, Vincent Ho, Brandy Klotzle, Jennie M. Le, David Delano, et al. 2011. “High Density DNA Methylation Array with Single CpG Site Resolution.” *Genomics*, New Genomic Technologies and Applications, 98 (4): 288–95. https://doi.org/10.1016/j.ygeno.2011.07.007.

Boots-Sprenger, Sandra H. E., Angelique Sijben, Jos Rijntjes, Bastiaan B. J. Tops, Albert J. Idema, Andreana L. Rivera, Fonnet E. Bleeker, et al. 2013. “Significance of Complete 1p/19q Co-Deletion, IDH1 Mutation and MGMT Promoter Methylation in Gliomas: Use with Caution.” *Modern Pathology* 26 (7): 922–29. https://doi.org/10.1038/modpathol.2012.166.

Brandstätter, Olga, Oliver Schanz, Julia Vorac, Jessica König, Tetsushi Mori, Toru Maruyama, Markus Korkowski, et al. 2016. “Balancing Intestinal and Systemic Inflammation through Cell Type-

Specific Expression of the Aryl Hydrocarbon Receptor Repressor." *Scientific Reports* 6 (May): 26091. https://doi.org/10.1038/srep26091.
Brenet, Fabienne, Michelle Moh, Patricia Funk, Erika Feierstein, Agnes J. Viale, Nicholas D. Socci, and Joseph M. Scandura. 2011. "DNA Methylation of the First Exon Is Tightly Linked to Transcriptional Silencing." *PLOS ONE* 6 (1): e14524. https://doi.org/10.1371/journal.pone.0014524.
Butcher, L. 2013. "Illumina450ProbeVariants.Db: Annotation Package Combining Variant Data from 1000 Genomes Project for Illumina HumanMethylation450 Bead Chip Probes." *R Package*.
Buuren, Stef van, and Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45 (3). http://www.jstatsoft.org/v45/i03.
"Cancer of the Urinary Bladder - SEER Stat Fact Sheets." 2016. 2016. http://seer.cancer.gov/statfacts/html/urinb.html.
Cárdenas-Turanzas M, Cooksley C, Pettaway CA, Sabichi A, Grossman HB, and Elting L. 2006. "Comparative Outcomes of Bladder Cancer." *Obstetrics and Gynecology* 108 (1): 169–75.
Centers for Disease Control and Prevention (US). 2010. *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General*. Publications and Reports of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US). http://www.ncbi.nlm.nih.gov/books/NBK53017/.
Chen, Rong-Jane, Yuan-Soon Ho, How-Ran Guo, and Ying-Jan Wang. 2008. "Rapid Activation of Stat3 and ERK1/2 by Nicotine Modulates Cell Proliferation in Human Bladder Cancer Cells." *Toxicological Sciences: An Official Journal of the Society of Toxicology* 104 (2): 283–93. https://doi.org/10.1093/toxsci/kfn086.
Chen, Yi-an, Mathieu Lemire, Sanaa Choufani, Darci T. Butcher, Daria Grafodatskaya, Brent W. Zanke, Steven Gallinger, Thomas J. Hudson, and Rosanna Weksberg. 2013. "Discovery of Cross-Reactive Probes and Polymorphic CpGs in the Illumina Infinium HumanMethylation450 Microarray." *Epigenetics* 8 (2): 203–9. https://doi.org/10.4161/epi.23470.
Cheng, Timothy H. T., Peiyong Jiang, Jacqueline C. W. Tam, Xiao Sun, Wing-Shan Lee, Stephanie C. Y. Yu, Jeremy Y. C. Teoh, et al. 2017. "Genomewide Bisulfite Sequencing Reveals the Origin and Time-Dependent Fragmentation of Urinary CfDNA." *Clinical Biochemistry* 50 (9): 496–501. https://doi.org/10.1016/j.clinbiochem.2017.02.017.
Dedeurwaerder, Sarah, Matthieu Defrance, Martin Bizet, Emilie Calonne, Gianluca Bontempi, and François Fuks. 2014. "A Comprehensive Overview of Infinium HumanMethylation450 Data Processing." *Briefings in Bioinformatics* 15 (6): 929–41. https://doi.org/10.1093/bib/bbt054.
Díez-Villanueva, Anna, Izaskun Mallona, and Miguel A. Peinado. 2015. "Wanderer, an Interactive Viewer to Explore DNA Methylation and Gene Expression Data in Human Cancer." *Epigenetics & Chromatin* 8: 22. https://doi.org/10.1186/s13072-015-0014-8.
Dobruch, Jakub, Siamak Daneshmand, Margit Fisch, Yair Lotan, Aidan P. Noon, Matthew J. Resnick, Shahrokh F. Shariat, Alexandre R. Zlotta, and Stephen A. Boorjian. 2016. "Gender and Bladder Cancer: A Collaborative Review of Etiology, Biology, and Outcomes." *European Urology* 69 (2): 300–310. https://doi.org/10.1016/j.eururo.2015.08.037.
Du, Pan, Xiao Zhang, Chiang-Ching Huang, Nadereh Jafari, Warren A. Kibbe, Lifang Hou, and Simon M. Lin. 2010. "Comparison of Beta-Value and M-Value Methods for Quantifying Methylation Levels by Microarray Analysis." *BMC Bioinformatics* 11 (1): 587. https://doi.org/10.1186/1471-2105-11-587.
Eckhardt, Florian, Joern Lewin, Rene Cortese, Vardhman K. Rakyan, John Attwood, Matthias Burger, John Burton, et al. 2006. "DNA Methylation Profiling of Human Chromosomes 6, 20 and 22." *Nature Genetics* 38 (12): 1378–85. https://doi.org/10.1038/ng1909.
Fasanelli, Francesca, Laura Baglietto, Erica Ponzi, Florence Guida, Gianluca Campanella, Mattias Johansson, Kjell Grankvist, et al. 2015. "Hypomethylation of Smoking-Related Genes Is Associated with Future Lung Cancer in Four Prospective Cohorts." *Nature Communications* 6 (December). https://doi.org/10.1038/ncomms10192.
Figueroa, Jonine D., Yuanqing Ye, Afshan Siddiq, Montserrat Garcia-Closas, Nilanjan Chatterjee, Ludmila Prokunina-Olsson, Victoria K. Cortessis, et al. 2014. "Genome-Wide Association Study Identifies Multiple Loci Associated with Bladder Cancer Risk." *Human Molecular Genetics* 23 (5): 1387–98. https://doi.org/10.1093/hmg/ddt519.

Florescu, Ana, Roberta Ferrence, Tom Einarson, Peter Selby, Offie Soldin, and Gideon Koren. 2009. “Methods for Quantification of Exposure to Cigarette Smoking and Environmental Tobacco Smoke: Focus on Developmental Toxicology.” *Therapeutic Drug Monitoring* 31 (1): 14–30. https://doi.org/10.1097/FTD.0b013e3181957a3b.

Fortin, Jean-Philippe, Aurélie Labbe, Mathieu Lemire, Brent W. Zanke, Thomas J. Hudson, Elana J. Fertig, Celia MT Greenwood, and Kasper D. Hansen. 2014. “Functional Normalization of 450k Methylation Array Data Improves Replication in Large Cancer Studies.” *Genome Biology* 15 (12): 503. https://doi.org/10.1186/s13059-014-0503-2.

Fox, Stephen B., José Bragança, Helen Turley, Leticia Campo, Cheng Han, Kevin C. Gatter, Shoumo Bhattacharya, and Adrian L. Harris. 2004. “CITED4 Inhibits Hypoxia-Activated Transcription in Cancer Cells, and Its Cytoplasmic Location in Breast Cancer Is Associated with Elevated Expression of Tumor Cell Hypoxia-Inducible Factor 1alpha.” *Cancer Research* 64 (17): 6075–81. https://doi.org/10.1158/0008-5472.CAN-04-0708.

Freedman, Neal D., Debra T. Silverman, Albert R. Hollenbeck, Arthur Schatzkin, and Christian C. Abnet. 2011a. “Association between Smoking and Risk of Bladder Cancer among Men and Women.” *JAMA* 306 (7): 737–45. https://doi.org/10.1001/jama.2011.1142.

Freedman, Neal D, Debra T Silverman, Albert R Hollenbeck, Arthur Schatzkin, and Christian C Abnet. 2011b. “Association between Smoking and Risk of Bladder Cancer among Men and Women.” *JAMA : The Journal of the American Medical Association* 306 (7): 737–45. https://doi.org/10.1001/jama.2011.1142.

Fu, Xiao Wen, Ping Fang Song, and Eliot R. Spindel. 2015. “Role of Lynx1 and Related Ly6 Proteins as Modulators of Cholinergic Signaling in Normal and Neoplastic Bronchial Epithelium.” *International Immunopharmacology* 29 (1): 93–98. https://doi.org/10.1016/j.intimp.2015.05.022.

Fu, Y. P., I. Kohaar, W. Tang, P. Porter-Gill, J. D. Figueroa, M. Garcia-Colsas, N. Chatterjee, et al. 2013. “Bladder Cancer Susceptibility Variants within CCNE1 Are Associated with MRNA Expression of an Alternative Splicing Form.” In *ASHG 2013 Annual Meeting*.

Gal-Yam, Einav Nili, Gerda Egger, Leo Iniguez, Heather Holster, Steingrímur Einarsson, Xinmin Zhang, Joy C. Lin, Gangning Liang, Peter A. Jones, and Amos Tanay. 2008. “Frequent Switching of Polycomb Repressive Marks and DNA Hypermethylation in the PC3 Prostate Cancer Cell Line.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (35): 12979–84. https://doi.org/10.1073/pnas.0806437105.

Gao, Xu, Min Jia, Yan Zhang, Lutz Philipp Breitling, and Hermann Brenner. 2015. “DNA Methylation Changes of Whole Blood Cells in Response to Active Smoking Exposure in Adults: A Systematic Review of DNA Methylation Studies.” *Clinical Epigenetics* 7 (October): 113. https://doi.org/10.1186/s13148-015-0148-3.

Gaunt, Tom R., Hashem A. Shihab, Gibran Hemani, Josine L. Min, Geoff Woodward, Oliver Lyttleton, Jie Zheng, et al. 2016. “Systematic Identification of Genetic Influences on Methylation across the Human Life Course.” *Genome Biology* 17 (March): 61. https://doi.org/10.1186/s13059-016-0926-z.

Hanahan, Douglas, and Robert A. Weinberg. 2011. “Hallmarks of Cancer: The Next Generation.” *Cell* 144 (5): 646–74. https://doi.org/10.1016/j.cell.2011.02.013.

Harlid, Sophia, Zongli Xu, Vijayalakshmi Panduri, Dale P. Sandler, and Jack A. Taylor. 2014. “CpG Sites Associated with Cigarette Smoking: Analysis of Epigenome-Wide Data from the Sister Study.” *Environmental Health Perspectives* 122 (7): 673–78. https://doi.org/10.1289/ehp.1307480.

Hays, Jennifer, Julie R. Hunt, F. Allan Hubbell, Garnet L. Anderson, Marian Limacher, Catherine Allen, and Jacques E. Rossouw. 2003. “The Women’s Health Initiative Recruitment Methods and Results.” *Annals of Epidemiology* 13 (9 Suppl): S18-77.

Holick, Crystal N., Edward L. Giovannucci, Meir J. Stampfer, and Dominique S. Michaud. 2007. “Prospective Study of Body Mass Index, Height, Physical Activity and Incidence of Bladder Cancer in US Men and Women.” *International Journal of Cancer* 120 (1): 140–46. https://doi.org/10.1002/ijc.22142.

Houseman, Eugene Andres, William P. Accomando, Devin C. Koestler, Brock C. Christensen, Carmen J. Marsit, Heather H. Nelson, John K. Wiencke, and Karl T. Kelsey. 2012. “DNA Methylation Arrays as Surrogate Measures of Cell Mixture Distribution.” *BMC Bioinformatics* 13: 86. https://doi.org/10.1186/1471-2105-13-86.

Houseman, Eugene Andres, John Molitor, and Carmen J. Marsit. 2014. “Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data.” *Bioinformatics* 30 (10): 1431–39. https://doi.org/10.1093/bioinformatics/btu029.

Huang, Katie T., Elena A. Takano, Thomas Mikeska, David J. Byrne, Alexander Dobrovic, and Stephen B. Fox. 2011. “Aberrant DNA Methylation but Not Mutation of CITED4 Is Associated with Alteration of HIF-Regulated Genes in Breast Cancer.” *Breast Cancer Research and Treatment* 130 (1): 319. https://doi.org/10.1007/s10549-011-1657-1.

Ibañez-Tallon, Inés, Julie M Miwa, Hai-Long Wang, Niels C Adams, Gregg W Crabtree, Steven M Sine, and Nathaniel Heintz. 2002. “Novel Modulation of Neuronal Nicotinic Acetylcholine Receptors by Association with the Endogenous Prototoxin Lynx1.” *Neuron* 33 (6): 893–903. https://doi.org/10.1016/S0896-6273(02)00632-3.

Ishikawa, Nobuhisa, Atsushi Takano, Wataru Yasui, Kouki Inai, Hitoshi Nishimura, Hiroyuki Ito, Yohei Miyagi, et al. 2007. “Cancer-Testis Antigen Lymphocyte Antigen 6 Complex Locus K Is a Serologic Biomarker and a Therapeutic Target for Lung and Esophageal Carcinomas.” *Cancer Research* 67 (24): 11601–11. https://doi.org/10.1158/0008-5472.CAN-07-3243.

Joehanes, Roby, Allan C. Just, Riccardo E. Marioni, Luke C. Pilling, Lindsay M. Reynolds, Pooja R. Mandaviya, Weihua Guan, et al. 2016. “Epigenetic Signatures of Cigarette Smoking.” *Circulation: Cardiovascular Genetics* 9 (5): 436–47. https://doi.org/10.1161/CIRCGENETICS.116.001506.

Jones, P A, M J Wolkowicz, W M Rideout, F A Gonzales, C M Marziasz, G A Coetzee, and S J Tapscott. 1990. “De Novo Methylation of the MyoD1 CpG Island during the Establishment of Immortal Cell Lines.” *Proceedings of the National Academy of Sciences of the United States of America* 87 (16): 6117–21.

Jones, Peter A. 2012. “Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond.” *Nature Reviews. Genetics* 13 (7): 484–92. https://doi.org/10.1038/nrg3230.

Jordahl, Kristina M., Timothy W. Randolph, Xiaoling Song, Cassandra L. Sather, Lesley F. Tinker, Amanda I. Phipps, Karl T. Kelsey, Emily White, and Parveen Bhatti. 2018. “Genome-Wide DNA Methylation in Pre-Diagnostic Blood and Bladder Cancer Risk in the Women’s Health Initiative.” *Cancer Epidemiology and Prevention Biomarkers*, January, cebp.0951.2017. https://doi.org/10.1158/1055-9965.EPI-17-0951.

Katoh, Masuko, and Masaru Katoh. 2003. “Identification and Characterization of FLJ10737 and CAMTA1 Genes on the Commonly Deleted Region of Neuroblastoma at Human Chromosome 1p36.31-P36.23.” *International Journal of Oncology* 23 (4): 1219–24.

Kiemeney, Lambertus A., Patrick Sulem, Soren Besenbacher, Sita H. Vermeulen, Asgeir Sigurdsson, Gudmar Thorleifsson, Daniel F. Gudbjartsson, et al. 2010. “A Sequence Variant at 4p16.3 Confers Susceptibility to Urinary Bladder Cancer.” *Nature Genetics* 42 (5): 415–19. https://doi.org/10.1038/ng.558.

Kim, Hyo Sin, Wang Jin Park, Eun Young Park, Jun Sung Koh, Tae-Kon Hwang, and Joon Chul Kim. 2015. “Role of Nicotinic Acetylcholine Receptor A3 and A7 Subunits in Detrusor Overactivity Induced by Partial Bladder Outlet Obstruction in Rats.” *International Neurourology Journal* 19 (1): 12–18. https://doi.org/10.5213/inj.2015.19.1.12.

Koestler, Devin C., Joseph Usset, Brock C. Christensen, Carmen J. Marsit, Margaret R. Karagas, Karl T. Kelsey, and John K. Wiencke. 2016. “DNA Methylation-Derived Neutrophil-to-Lymphocyte Ratio: An Epigenetic Tool to Explore Cancer Inflammation and Outcomes.” *Cancer Epidemiology and Prevention Biomarkers*, December. https://doi.org/10.1158/1055-9965.EPI-16-0461.

Kohaar, Indu, Patricia Porter-Gill, Petra Lenz, Yi-Ping Fu, Adam Mumy, Wei Tang, Andrea B. Apolo, et al. 2013. “Genetic Variant as a Selection Marker for Anti–Prostate Stem Cell Antigen Immunotherapy of Bladder Cancer.” *JNCI Journal of the National Cancer Institute* 105 (1): 69–73. https://doi.org/10.1093/jnci/djs458.

Kohaar, Indu, Alexandra Scott-Johnson, Yi-Ping Fu, Patricia Porter-Gill, and Ludmila Prokunina-Olsson. 2011. “Functional Exploration of CCNE1 Splicing Forms as a Possible Link to Bladder Cancer Susceptibility.” *Genome Biology* 12 (Suppl 1): P7. https://doi.org/10.1186/gb-2011-12-s1-p7.

Kõks, Gea, Mari-Liis Uudelepp, Maia Limbach, Pärt Peterson, Ene Reimann, and Sulev Kõks. 2015. “Smoking-Induced Expression of the GPR15 Gene Indicates Its Potential Role in Chronic Inflammatory Pathologies.” *The American Journal of Pathology* 185 (11): 2898–2906. https://doi.org/10.1016/j.ajpath.2015.07.006.

Kong, Hyun Kyung, and Jong Hoon Park. 2012. “Characterization and Function of Human Ly-6/UPAR Molecules.” *BMB Reports* 45 (11): 595–603. https://doi.org/10.5483/BMBRep.2012.45.11.210.

Kulis, Marta, and Manel Esteller. 2010. “2 - DNA Methylation and Cancer.” In *Advances in Genetics*, edited by Zdenko Herceg and Toshikazu Ushijima, 70:27–56. Epigenetics and Cancer, Part A. Academic Press. http://www.sciencedirect.com/science/article/pii/B9780123808660600022.

Laird, Peter W. 2010. “Principles and Challenges of Genome-Wide DNA Methylation Analysis.” *Nature Reviews Genetics* 11 (3): 191–203. https://doi.org/10.1038/nrg2732.

Larsson, Susanna C., Swen-Olof Andersson, Jan-Erik Johansson, and Alicja Wolk. 2008. “Diabetes Mellitus, Body Size and Bladder Cancer Risk in a Prospective Study of Swedish Men.” *European Journal of Cancer (Oxford, England: 1990)* 44 (17): 2655–60. https://doi.org/10.1016/j.ejca.2008.07.012.

Lee, Ju-Seog, Sun-Hee Leem, Sang-Yeop Lee, Sang-Cheol Kim, Eun-Sung Park, Sang-Bae Kim, Seon-Kyu Kim, Yong-June Kim, Wun-Jae Kim, and In-Sun Chu. 2010. “Expression Signature of E2F1 and Its Associated Genes Predict Superficial to Invasive Progression of Bladder Tumors.” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 28 (16): 2660–67. https://doi.org/10.1200/JCO.2009.25.0977.

Leeb, Martin, Diego Pasini, Maria Novatchkova, Markus Jaritz, Kristian Helin, and Anton Wutz. 2010. “Polycomb Complexes Act Redundantly to Repress Genomic Repeats and Genes.” *Genes & Development* 24 (3): 265–76. https://doi.org/10.1101/gad.544410.

Li, Dongmei, Zidian Xie, Marc Le Pape, and Timothy Dye. 2015. “An Evaluation of Statistical Methods for DNA Methylation Microarray Data Analysis.” *BMC Bioinformatics* 16 (July): 217. https://doi.org/10.1186/s12859-015-0641-x.

Luo, Linlin, Peter McGarvey, Subha Madhavan, Rakesh Kumar, Yuriy Gusev, and Geeta Upadhyay. 2016. “Distinct Lymphocyte Antigens 6 (Ly6) Family Members Ly6D, Ly6E, Ly6K and Ly6H Drive Tumorigenesis and Clinical Outcome.” *Oncotarget* 7 (10): 11165–93. https://doi.org/10.18632/oncotarget.7163.

Marsit, Carmen J., Devin C. Koestler, Brock C. Christensen, Margaret R. Karagas, E. Andres Houseman, and Karl T. Kelsey. 2011. “DNA Methylation Array Analysis Identifies Profiles of Blood-Derived DNA Methylation Associated With Bladder Cancer.” *Journal of Clinical Oncology* 29 (9): 1133–39. https://doi.org/10.1200/JCO.2010.31.3577.

Moriwaki, Yasuhiro, Ken Yoshikawa, Hiromi Fukuda, Yoshihito X. Fujii, Hidemi Misawa, and Koichiro Kawashima. 2007. “Immune System Expression of SLURP-1 and SLURP-2, Two Endogenous Nicotinic Acetylcholine Receptor Ligands.” *Life Sciences*, Proceedings of the 2nd International Symposium on Non-neuronal Acetylcholine, 80 (24): 2365–68. https://doi.org/10.1016/j.lfs.2006.12.028.

Moyer, Virginia A. 2014. “Screening for Lung Cancer: U.S. Preventive Services Task Force Recommendation Statement.” *Annals of Internal Medicine* 160 (5): 330–38. https://doi.org/10.7326/M13-2771.

Nooij-van Dalen, Arnolda G. de, Guus A. M. S. van Dongen, Serge J. Smeets, Eline J. C. Nieuwenhuis, Marijke Stigter-van Walsum, Gordon B. Snow, and Ruud H. Brakenhoff. 2003. “Characterization of the Human Ly-6 Antigens, the Newly Annotated Member Ly-6K Included, as Molecular Markers for Head-and-Neck Squamous Cell Carcinoma.” *International Journal of Cancer* 103 (6): 768–74. https://doi.org/10.1002/ijc.10903.

Nordlund, Jessica, Christofer L. Bäcklin, Per Wahlberg, Stephan Busche, Eva C. Berglund, Maija-Leena Eloranta, Trond Flaegstad, et al. 2013. “Genome-Wide Signatures of Differential DNA Methylation in Pediatric Acute Lymphoblastic Leukemia.” *Genome Biology* 14 (9): r105. https://doi.org/10.1186/gb-2013-14-9-r105.

Packer, Bernice R., Meredith Yeager, Brian Staats, Robert Welch, Andrew Crenshaw, Maureen Kiley, Andrew Eckert, et al. 2004. “SNP500Cancer: A Public Resource for Sequence Validation and Assay Development for Genetic Variation in Candidate Genes.” *Nucleic Acids Research* 32 (Database issue): D528–32. https://doi.org/10.1093/nar/gkh005.

Pettersson, Ann, Gunnar Nylund, Amir Khorram-Manesh, Svante Nordgren, and Dick S. Delbro. 2009. “Nicotine Induced Modulation of SLURP-1 Expression in Human Colon Cancer Cells.” *Autonomic Neuroscience* 148 (1): 97–100. https://doi.org/10.1016/j.autneu.2009.03.002.

Philibert, Robert A., Steven R. H. Beach, Man-Kit Lei, and Gene H. Brody. 2013. “Changes in DNA Methylation at the Aryl Hydrocarbon Receptor Repressor May Be a New Biomarker for Smoking.” *Clinical Epigenetics* 5 (1): 19. https://doi.org/10.1186/1868-7083-5-19.

Pidsley, Ruth, Chloe C. Y Wong, Manuela Volta, Katie Lunnon, Jonathan Mill, and Leonard C. Schalkwyk. 2013. “A Data-Driven Approach to Preprocessing Illumina 450K Methylation Array Data.” *BMC Genomics* 14 (May): 293. https://doi.org/10.1186/1471-2164-14-293.

Pintarelli, Giulia, Chiara Elisabetta Cotroneo, Sara Noci, Matteo Dugo, Antonella Galvan, Simona Delli Carpini, Lorena Citterio, et al. 2017. “Genetic Susceptibility Variants for Lung Cancer: Replication Study and Assessment as Expression Quantitative Trait Loci.” *Scientific Reports* 7 (February): 42185. https://doi.org/10.1038/srep42185.

Price, E. Magda, Allison M. Cotton, Lucia L. Lam, Pau Farré, Eldon Emberly, Carolyn J. Brown, Wendy P. Robinson, and Michael S. Kobor. 2013. “Additional Annotation Enhances Potential for Biologically-Relevant Analysis of the Illumina Infinium HumanMethylation450 BeadChip Array.” *Epigenetics & Chromatin* 6: 4. https://doi.org/10.1186/1756-8935-6-4.

Qian, Bin-Zhi, and Jeffrey W. Pollard. 2010. “Macrophage Diversity Enhances Tumor Progression and Metastasis.” *Cell* 141 (1): 39–51. https://doi.org/10.1016/j.cell.2010.03.014.

Qu, Ying, Lee Siggens, Lina Cordeddu, Verena I. Gaidzik, Kasper Karlsson, Lars Bullinger, Konstanze Döhner, Karl Ekwall, Sören Lehmann, and Andreas Lennartsson. 2017. “Cancer-Specific Changes in DNA Methylation Reveal Aberrant Silencing and Activation of Enhancers in Leukemia.” *Blood* 129 (7): e13–25. https://doi.org/10.1182/blood-2016-07-726877.

Quak, J. J., A. J. Balm, G. A. van Dongen, J. G. Brakkee, R. J. Scheper, G. B. Snow, and C. J. Meijer. 1990. “A 22-Kd Surface Antigen Detected by Monoclonal Antibody E 48 Is Exclusively Expressed in Stratified Squamous and Transitional Epithelia.” *The American Journal of Pathology* 136 (1): 191–97.

Rafnar, Thorunn, Patrick Sulem, Simon N. Stacey, Frank Geller, Julius Gudmundsson, Asgeir Sigurdsson, Margret Jakobsdottir, et al. 2009. “Sequence Variants at the *TERT-CLPTM1L* Locus Associate with Many Cancer Types.” *Nature Genetics* 41 (2): 221–27. https://doi.org/10.1038/ng.296.

Rakyan, Vardhman K., Thomas A. Down, David J. Balding, and Stephan Beck. 2011. “Epigenome-Wide Association Studies for Common Human Diseases.” *Nature Reviews Genetics* 12 (8): 529–41. https://doi.org/10.1038/nrg3000.

Reeves, Gillian K, Kirstin Pirie, Valerie Beral, Jane Green, Elizabeth Spencer, and Diana Bull. 2007. “Cancer Incidence and Mortality in Relation to Body Mass Index in the Million Women Study: Cohort Study.” *BMJ : British Medical Journal* 335 (7630): 1134. https://doi.org/10.1136/bmj.39367.495995.AE.

Reifenberger, Guido, and David N. Louis. 2003. “Oligodendroglioma: Toward Molecular Definitions in Diagnostic Neuro-Oncology.” *Journal of Neuropathology and Experimental Neurology* 62 (2): 111–26.

Reiter, Robert E., Zhennen Gu, Tetsuro Watabe, George Thomas, Kinga Szigeti, Elizabeth Davis, Matthew Wahl, et al. 1998. “Prostate Stem Cell Antigen: A Cell Surface Marker Overexpressed in Prostate Cancer.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (4): 1735–40.

Robertson, Keith D. 2005. “DNA Methylation and Human Disease.” *Nature Reviews Genetics* 6 (8): 597–610. https://doi.org/10.1038/nrg1655.

Rogers, Michael A., Verena Kalter, Gemma Marcias, Marc Zapatka, Sebastian Barbus, and Peter Lichter. 2016. “CITED4 Gene Silencing in Colorectal Cancer Cells Modulates Adherens/Tight Junction Gene Expression and Reduces Cell Proliferation.” *Journal of Cancer Research and Clinical Oncology* 142 (1): 225–37. https://doi.org/10.1007/s00432-015-2011-5.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316696.

Sanli, Oner, Jakub Dobruch, Margaret A. Knowles, Maximilian Burger, Mehrdad Alemozaffar, Matthew E. Nielsen, and Yair Lotan. 2017. “Bladder Cancer.” *Nature Reviews Disease Primers* 3 (April): 17022. https://doi.org/10.1038/nrdp.2017.22.

Schlesinger, Yeshayahu, Ravid Straussman, Ilana Keshet, Shlomit Farkash, Merav Hecht, Joseph Zimmerman, Eran Eden, et al. 2007. “Polycomb-Mediated Methylation on Lys27 of Histone H3

Pre-Marks Genes for de Novo Methylation in Cancer." *Nature Genetics* 39 (2): 232–36. https://doi.org/10.1038/ng1950.
Scosyrev, Emil, Katia Noyes, Changyong Feng, and Edward Messing. 2009. "Sex and Racial Differences in Bladder Cancer Presentation and Mortality in the US." *Cancer* 115 (1): 68–74. https://doi.org/10.1002/cncr.23986.
Shenker, Natalie S., Silvia Polidoro, Karin van Veldhoven, Carlotta Sacerdote, Fulvio Ricceri, Mark A. Birrell, Maria G. Belvisi, Robert Brown, Paolo Vineis, and James M. Flanagan. 2013. "Epigenome-Wide Association Study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) Identifies Novel Genetic Loci Associated with Smoking." *Human Molecular Genetics* 22 (5): 843–51. https://doi.org/10.1093/hmg/dds488.
Shenker, Natalie S., Per Magne Ueland, Silvia Polidoro, Karin van Veldhoven, Fulvio Ricceri, Robert Brown, James M. Flanagan, and Paolo Vineis. 2013. "DNA Methylation as a Long-Term Biomarker of Exposure to Tobacco Smoke." *Epidemiology (Cambridge, Mass.)* 24 (5): 712–16. https://doi.org/10.1097/EDE.0b013e31829d5cb3.
Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin. 2001. "DbSNP: The NCBI Database of Genetic Variation." *Nucleic Acids Research* 29 (1): 308–11.
Singh, Sandeep, Smitha Pillai, and Srikumar Chellappan. 2011. "Nicotinic Acetylcholine Receptor Signaling in Tumor Growth and Metastasis." *Journal of Oncology* 2011. https://doi.org/10.1155/2011/456743.
Siu, Ka Tat, Marsha Rich Rosner, and Alex C Minella. 2012. "An Integrated View of Cyclin E Function and Regulation." *Cell Cycle* 11 (1): 57–64. https://doi.org/10.4161/cc.11.1.18775.
Smith, Alicia K., Varun Kilaru, Mehmet Kocak, Lynn M. Almli, Kristina B. Mercer, Kerry J. Ressler, Frances A. Tylavsky, and Karen N. Conneely. 2014. "Methylation Quantitative Trait Loci (MeQTLs) Are Consistently Detected across Ancestry, Developmental Stage, and Tissue Type." *BMC Genomics* 15: 145. https://doi.org/10.1186/1471-2164-15-145.
Song, Xin, Eero Pukkala, Tadeusz Dyba, Jaakko Tuomilehto, Vladislav Moltchanov, Satu Männistö, Pekka Jousilahti, and Qing Qiao. 2014. "Body Mass Index and Cancer Incidence: The FINRISK Study." *European Journal of Epidemiology* 29 (7): 477–87. https://doi.org/10.1007/s10654-014-9934-z.
Stueve, Theresa Ryan, Wen-Qing Li, Jianxin Shi, Crystal N. Marconett, Tongwu Zhang, Chenchen Yang, Daniel Mullen, et al. 2017. "Epigenome-Wide Analysis of DNA Methylation in Lung Tissue Shows Concordance with Blood Studies and Identifies Tobacco Smoke-Inducible Enhancers." *Human Molecular Genetics* 26 (15): 3014–27. https://doi.org/10.1093/hmg/ddx188.
Sun, Yan V., Alicia K. Smith, Karen N. Conneely, Qiuzhi Chang, Weiyan Li, Alicia Lazarus, Jennifer A. Smith, et al. 2013. "Epigenomic Association Analysis Identifies Smoking-Related DNA Methylation Sites in African Americans." *Human Genetics* 132 (9): 1027–37. https://doi.org/10.1007/s00439-013-1311-6.
Tanas, Munir R., Andrea Sboner, Andre M. Oliveira, Michele R. Erickson-Johnson, Jessica Hespelt, Philip J. Hanwright, John Flanagan, et al. 2011. "Identification of a Disease-Defining Gene Fusion in Epithelioid Hemangioendothelioma." *Science Translational Medicine* 3 (98): 98ra82-98ra82. https://doi.org/10.1126/scitranslmed.3002409.
Tews, B., P. Roerig, C. Hartmann, M. Hahn, J. Felsberg, B. Blaschke, M. Sabel, et al. 2007. "Hypermethylation and Transcriptional Downregulation of the CITED4 Gene at 1p34.2 in Oligodendroglial Tumours with Allelic Losses on 1p and 19q." *Oncogene* 26 (34): 5010–16. https://doi.org/10.1038/sj.onc.1210297.
Tsaprouni, Loukia G., Tsun-Po Yang, Jordana Bell, Katherine J. Dick, Stavroula Kanoni, James Nisbet, Ana Viñuela, et al. 2014. "Cigarette Smoking Reduces DNA Methylation Levels at Multiple Genomic Loci but the Effect Is Partially Reversible upon Cessation." *Epigenetics* 9 (10): 1382–96. https://doi.org/10.4161/15592294.2014.969637.
Valeri, Linda, and Tyler J. VanderWeele. 2013. "Mediation Analysis Allowing for Exposure–Mediator Interactions and Causal Interpretation: Theoretical Assumptions and Implementation with SAS and SPSS Macros." *Psychological Methods* 18 (2): 137–50. https://doi.org/10.1037/a0031034.
VanderWeele, Tyler. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford, New York: Oxford University Press.

VanderWeele, Tyler J., and Stijn Vansteelandt. 2010. “Odds Ratios for Mediation Analysis for a Dichotomous Outcome.” *American Journal of Epidemiology* 172 (12): 1339–48. https://doi.org/10.1093/aje/kwq332.

VanderWeele, Tyler, and Stijn Vansteelandt. 2014. “Mediation Analysis with Multiple Mediators.” *Epidemiologic Methods* 2 (1): 95–115. https://doi.org/10.1515/em-2012-0010.

Vergnolle, Nathalie, Claudia K. Derian, Michael R. D’Andrea, Martin Steinhoff, and Patricia Andrade-Gordon. 2002. “Characterization of Thrombin-Induced Leukocyte Rolling and Adherence: A Potential Proinflammatory Role for Proteinase-Activated Receptor-4.” *The Journal of Immunology* 169 (3): 1467–73. https://doi.org/10.4049/jimmunol.169.3.1467.

Wan, Emily S., Weiliang Qiu, Andrea Baccarelli, Vincent J. Carey, Helene Bacherman, Stephen I. Rennard, Alvar Agusti, Wayne Anderson, David A. Lomas, and Dawn L. Demeo. 2012. “Cigarette Smoking Behaviors and Time since Quitting Are Associated with Differential DNA Methylation across the Human Genome.” *Human Molecular Genetics* 21 (13): 3073–82. https://doi.org/10.1093/hmg/dds135.

Widschwendter, Martin, Heidi Fiegl, Daniel Egle, Elisabeth Mueller-Holzner, Gilbert Spizzo, Christian Marth, Daniel J. Weisenberger, et al. 2007. “Epigenetic Stem Cell Signature in Cancer.” *Nature Genetics* 39 (2): 157–58. https://doi.org/10.1038/ng1941.

Wolff, Erika M., Yoshitomo Chihara, Fei Pan, Daniel J. Weisenberger, Kimberly D. Siegmund, Kokichi Sugano, Kiyotaka Kawashima, Peter W. Laird, Peter A. Jones, and Gangning Liang. 2010. “Unique DNA Methylation Patterns Distinguish Noninvasive and Invasive Urothelial Cancers and Establish an Epigenetic Field Defect in Premalignant Tissue.” *Cancer Research* 70 (20): 8169–78. https://doi.org/10.1158/0008-5472.CAN-10-1335.

Wright, Michelle L., Mikhail G. Dozmorov, Aaron R. Wolen, Colleen Jackson-Cook, Angela R. Starkweather, Debra E. Lyon, and Timothy P. York. 2016. “Establishing an Analytic Pipeline for Genome-Wide DNA Methylation.” *Clinical Epigenetics* 8: 45. https://doi.org/10.1186/s13148-016-0212-7.

Xu, Zongli, Sophia C. E. Bolick, Lisa A. DeRoo, Clarice R. Weinberg, Dale P. Sandler, and Jack A. Taylor. 2013. “Epigenome-Wide Association Study of Breast Cancer Using Prospectively Collected Sister Study Samples.” *Journal of the National Cancer Institute* 105 (10): 694–700. https://doi.org/10.1093/jnci/djt045.

Yahata, Tetsuro, Hiroko Takedatsu, Sally L. Dunwoodie, José Bragança, Tracey Swingler, Sarah L. Withington, Jingyung Hur, et al. 2002. “Cloning of Mouse Cited4, a Member of the CITED Family P300/CBP-Binding Transcriptional Coactivators: Induced Expression in Mammary Epithelial Cells.” *Genomics* 80 (6): 601–13. https://doi.org/10.1006/geno.2002.7005.

Zeilinger, Sonja, Brigitte Kühnel, Norman Klopp, Hansjörg Baurecht, Anja Kleinschmidt, Christian Gieger, Stephan Weidinger, et al. 2013. “Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation.” *PloS One* 8 (5): e63812. https://doi.org/10.1371/journal.pone.0063812.

Zhang, Bin, Zhengliang Zhang, Xufei Zhang, Xu Gao, Kemp H. Kernstine, and Li Zhong. 2012. “Serological Antibodies against LY6K as a Diagnostic Biomarker in Esophageal Squamous Cell Carcinoma.” *Biomarkers: Biochemical Indicators of Exposure, Response, and Susceptibility to Chemicals* 17 (4): 372–78. https://doi.org/10.3109/1354750X.2012.680609.

Zhang Y, Zhu C, Curado MP, Zheng T, and Boyle P. 2012. “Changing Patterns of Bladder Cancer in the USA: Evidence of Heterogeneous Disease.” *BJU International* 109 (1): 52–56.

Zhou, Xin, Brett Maricque, Mingchao Xie, Daofeng Li, Vasavi Sundaram, Eric A. Martin, Brian C. Koebbe, et al. 2011. “The Human Epigenome Browser at Washington University.” *Nature Methods* 8 (12): 989–90. https://doi.org/10.1038/nmeth.1772.

Zhou, Xin, and Ting Wang. 2012. “Using the Wash U Epigenome Browser to Examine Genome-Wide Sequencing Data.” *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* 0 10 (December): Unit10.10. https://doi.org/10.1002/0471250953.bi1010s40.

Zudaire, Enrique, Natalia Cuesta, Vundavalli Murty, Karen Woodson, Lisa Adams, Nieves Gonzalez, Alfredo Martínez, et al. 2008. “The Aryl Hydrocarbon Receptor Repressor Is a Putative Tumor Suppressor Gene in Multiple Human Cancers.” *The Journal of Clinical Investigation* 118 (2): 640–50. https://doi.org/10.1172/JCI30024.

VITA

Kristina Jordahl was born and has close family ties to cities near Minneapolis, MN and grew up in cities near Boston, MA. She has a Bachelor of Arts in Biology from Brown University and a Master of Science in Biostatistics from the University of Michigan. Before starting the PhD program in Epidemiology at the University of Washington, she worked for two years as a statistician on a variety of epidemiology-focused projects at the Harvard School of Public Health. After completion of her PhD, she plans to continue exploring her interest in multi-dimensional data by working as a post-doctoral research fellow in the Department of Epidemiology at the University of Washington and at the Fred Hutchinson Cancer Research Center.