Cancer Causes & Control (2019) 30:1087–1100
https://doi.org/10.1007/s10552-019-01221-y

ORIGINAL PAPER

Inherited alterations of TGF beta signaling components in Appalachian cervical cancers

Thomas J. Knobloch[1,11] · Juan Peng[2] · Erinn M. Hade[2] · David E. Cohn[3,11] · Mack T. Ruffin IV[4] · Michael A. Schiano[5,6] · Byron C. Calhoun[5,6] · William C. McBee Jr.[7] · Jamie L. Lesnock[7] · Holly H. Gallion[8] · Jondavid Pollock[9] · Bo Lu[1] · Steve Oghumu[1,11] · Zhaoxia Zhang[1] · Marta T. Sears[1] · Blessing E. Ogbemudia[1] · Joseph T. Perrault[10] · Logan C. Weghorst[1] · Erin Strawser[1] · Cecilia R. DeGraffinreid[10] · Electra D. Paskett[1,10,11] · Christopher M. Weghorst[1,11]

Received: 7 November 2018 / Accepted: 15 August 2019 / Published online: 21 August 2019

Abstract

Purpose This study examined targeted genomic variants of transforming growth factor beta (TGFB) signaling in Appalachian women. Appalachian women with cervical cancer were compared to healthy Appalachian counterparts to determine whether these polymorphic alleles were over-represented within this high-risk cancer population, and whether lifestyle or environmental factors modified the aggregate genetic risk in these Appalachian women.

Methods Appalachian women's survey data and blood samples from the Community Awareness, Resources, and Education (CARE) CARE I and CARE II studies ($n = 163$ invasive cervical cancer cases, 842 controls) were used to assess gene–environment interactions and cancer risk. Polymorphic allele frequencies and socio-behavioral demographic measurements were compared using t tests and χ^2 tests. Multivariable logistic regression was used to evaluate interaction effects between genomic variance and demographic, behavioral, and environmental characteristics.

Results Several alleles demonstrated significant interaction with smoking (*TP53 rs1042522, TGFB1 rs1800469*), alcohol consumption (*NQO1 rs1800566*), and sexual intercourse before the age of 18 (*TGFBR1 rs11466445, TGFBR1 rs7034462, TGFBR1 rs11568785*). Interestingly, we noted a significant interaction between "Appalachian self-identity" variables and *NQO1 rs1800566*. Multivariable logistic regression of cancer status in an over-dominant *TGFB1 rs1800469/TGFBR1 rs11568785* model demonstrated a 3.03-fold reduction in cervical cancer odds. Similar decreased odds (2.78-fold) were observed in an over-dominant *TGFB1 rs1800469/TGFBR1 rs7034462* model in subjects who had no sexual intercourse before age 18.

Conclusions This study reports novel associations between common low-penetrance alleles in the TGFB signaling cascade and modified risk of cervical cancer in Appalachian women. Furthermore, our unexpected findings associating Appalachian identity and *NQO1 rs1800566* suggests that the complex environmental exposures that contribute to Appalachian self-identity in Appalachian cervical cancer patients represent an emerging avenue of scientific exploration.

Keywords Cervical cancer · Gene–environment interaction · Genetic association · Polymorphic allele · Appalachia

Introduction

Cancer of the uterine cervix remains a significant cause of cancer death in women worldwide [1–3]. In the U.S. an estimated 13,170 women will be diagnosed with invasive cervical cancer in 2019 and 4,250 will die from the disease. While the incidence of cervical cancer has decreased markedly over the last three decades, these advances have stabilized over the last several years. Globally, nearly 570,000 cases and more than 311,000 deaths are expected with the

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10552-019-01221-y) contains supplementary material, which is available to authorized users.

✉ Thomas J. Knobloch
knobloch.1@osu.edu

Extended author information available on the last page of the article

vast majority (84%) expected to come from less developed regions [1–3].

Despite some of the best healthcare available in the world, health disparities in underserved populations remain an ongoing battle in the U.S. Specifically, Appalachia is a 205,000 square-mile region in the U.S. that follows the contours of the Appalachian Mountain range from New York to Mississippi, touching regions in 13 states along the way [4–6]. Historically characterized as a mostly white, non-Hispanic population living in economically depressed rural settings known for mining, forestry, and chemical manufacturing, the Appalachian region has demonstrated a remarkable recovery over the last 50 years. While the poverty rate in this largely rural region has decreased from 31 to 17.2% during this period, 91 counties are still listed as high poverty with rates greater than 1.5 times the US average [5, 6]. In the U.S., there are increased incidence and death rates of rural Appalachian women from cervical cancer compared to non-Appalachian women, demonstrating a relative risk of 1.29 (95% CI 1.21–1.38). Appalachian women are recognized to participate in lifestyle behaviors such as increased smoking, energy imbalance, and risky sexual behaviors which place them at greater risk for developing cervical cancer [6–9]. Furthermore, fewer Pap tests, elevated rates of Human Papillomavirus (HPV) infection, and an overall diminished access to health care contribute to the increased incidence of cervical cancer in these rural regions [8, 10]. However, it is possible that these factors alone do not fully explain the increased risk of cervical cancer in rural Appalachia. For example, while HPV infections have been causally linked to cervical cancer development, the majority of HPV-infected women do not progress and present malignant lesions. Consequently, other factors must alter biological mechanisms that ultimately tilt the homeostatic balance of the cervical microenvironment toward cancer.

Emerging evidence suggests that the missing risk liability is associated with gene–environment interactions. It is well established that high-penetrance somatic gene mutations are associated with the development of cancers, and this damage is a hallmark feature of multistep epithelial carcinogenesis. But what about inherited genetic variants? Polymorphisms are genetic variants that are detected in > 1% of a population and often of unknown biological significance. Millions of single-nucleotide polymorphisms (SNPs) have been identified by massively parallel sequencing projects, but far fewer have been strongly associated with specific disease states such as cancer, and even less are established as part of a functional mechanism. Consequently, because there is little known about the role of genetic susceptibility and gene–environment interactions on the increased rate of cervical cancer development in Appalachia, we implemented a targeted profiling of genetic variation in Appalachian women in an effort to address this unknown genetic risk component.

The transforming growth factor beta (TGFB) signaling cascade has been well established as a hallmark regulatory mechanism for controlling cell growth, proliferation, and migration [11]. Consequently, when components of this central pathway are impaired, cells become insensitive to TGF-ß-mediated growth controls, and carcinogenic progression occurs [12]. Key members in this signaling series include the secreted cytokines and TGFB ligands (TGFB1, TGFB2, TGFB3) as well as their cognate receptors (TGFBR1, TGFBR2, TGFBR3) at the point of engagement at the cell surface. The events that follow ligand binding, receptor activation, and downstream signaling place TGFB-mediated events at the forefront for both tumor suppression and tumor promotion by altering the cell microenvironment to impact cell growth, differentiation, and migration. Importantly, numerous studies have demonstrated that polymorphic alleles of TGFB pathway components significantly increase the risk of cancer development, including HPV-associated cancers [13, 14]. For example, SNPs in the *TGFB1*, *TGFB2*, *TGFB3* cytokine genes and the *TGFBR1* and *TGFBR2* receptor genes have been associated with altered risk in several cancers, including those of breast, colon, lung, bladder, pancreas, liver, head and neck, ovary, endometrium, and cervix [15–24].

Furthermore, the polymorphic *TGFBR1* gene presents in a number of cancers as a high-frequency, low-penetrance allele (*TGFBR1*6A*) that confers increased cancer susceptibility. The most common *TGFBR1* allele contains an amino acid repeat containing nine alanines (*TGFBR1*9A*) within the coding sequence at the 3′-end of exon 1. The *TGFBR1*6A* hypomorphic allele contains a deletion of three alanines within this repeat region which has been proposed to encode the signal sequence [25]. Studies and meta-analyses have subsequently shown that *TGFBR1*6A* (*i*) is a candidate tumor susceptibility allele, (*ii*) is present in a large proportion of the general population, (*iii*) is significantly elevated (20%) in cases compared to controls, and (*iv*) significantly increases cancer risk by 20–22% [26, 27]. Importantly, at least 14% of the general population carries at least one copy of the *TGFBR1*6A* allele. Recently, Levovitz et al. [14] reported on the fundamental importance of *TGFBR1* as an immune susceptibility gene in HPV-associated cancers, as well as the consistent presence of other TGFB-associated signaling components in HPV-mediated cancers.

Therefore, while the polymorphic alleles within the TGFB signaling cascade have been implicated in modifying cancer susceptibility, the contributions of these factors within a gene–environment model have not been characterized for cervical cancer within the elevated risk Appalachian population. Consequently, a series of pathway-based candidate gene variants previously associated with altered cancer risk were examined within the framework of Appalachian cervical cancer patients versus regionally matched women

presenting with normal Pap tests. In addition, polymorphisms demonstrated to be associated with cervical cancer risk in the general population were characterized for their contribution in this specific at-risk population. Furthermore, key clinicopathologic and socio-demographic features were interrogated in relation to these genetic cofactors in order to explore novel gene–environment interactions within this unique underserved population.

Materials and methods

Study design

As part of the National Institutes of Health, Centers for Population Health and Health Disparities (CPHHD) Initiative, The Ohio State University CPHHD Center was developed to focus on the health of underserved areas in Appalachia. Through this initiative the Community Awareness, Resources, and Education (CARE I [28] and CARE II [29]) studies were established to address high cervical cancer incidence and mortality in the southern North Central Appalachia region (Ohio, West Virginia) and Central Appalachian region (West Virginia, Kentucky) [6].

One source of data and samples was a case–control study in CARE II, which included 163 women diagnosed with invasive cervical cancer (ICC) and 79 control women who had normal Pap tests from Appalachian Ohio, West Virginia, and Kentucky (CARE II 2011–2015) [29]. A second source of data and samples from women with normal Pap tests (control group) were included from southern Appalachian Ohio, CARE I (2005–2009, $n = 763$) [28]. Study methods of CARE I have been described in great detail previously [30].

Women with prevalent or newly diagnosed ICC were recruited during the CARE II study period. Inclusion criteria for participants were (*i*) women residing in Appalachian counties who were ≥ 18 years, (*ii*) spoke English, (*iii*) not cognitively impaired, and (*iv*) able to provide informed consent. Blood samples were collected at the time of recruitment into each study and used for genomic DNA isolation. A self-administered baseline questionnaire was collected from all participants at the time of recruitment. In all studies, blood samples were collected at the time of recruitment into each study and used for genomic DNA isolation. A self-administered baseline questionnaire was collected from all participants at the time of recruitment. Questions included those related to demographics, social environment, general health history, behavioral, psychosocial, and sexual health.

Statistical analysis

Demographic, behavioral, and environmental characteristics were compared across arms using *t* tests and χ^2 tests as appropriate. Several different genetic models of inheritance to estimate cancer risks were defined. For example, *rs1234567890* is a SNP (T/C) with "T" representing the most frequent allele and "C" representing the variant allele which modifies the risk of developing cervical cancer. The risk of developing cancer is dependent on the number of C alleles present in the genotype of interest. Genetic models were defined as follows: (*i*) In a Dominant Model, the presence of a single-risk allele C is sufficient to modify risk. Consequently, both the heterozygous T/C genotype and the homozygous C/C genotype have the same risk, and in summation are compared to the homozygous most frequent allele T/T genotype. (*ii*) In a Co-dominant Model each genotype presents as distinct and non-additive. The heterozygous T/C genotype is compared to the homozygous most frequent allele T/T genotype, and the homozygous risk allele C/C genotype is compared to the homozygous most frequent allele T/T genotype. Co-dominant models allow for a hypo/hypermorphic gradient such that a functional continuum can exist [CC > T/C > T/T or CC < T/C < TT]. (*iii*) In an Over-dominant Model, the heterozygous T/C genotype is compared to the summation of the homozygous most frequent allele genotype T/T and the homozygous risk allele genotype C/C. In this model, the heterozygous T/C allele is dominant to either homozygous allele. (*iv*) In a Recessive Model, the presence of two risk alleles C/C is necessary to modify the risk for cervical cancer. Consequently, both the heterozygous T/C genotype and the homozygous most common allele T/T genotype have the same risk, and the summation of these genotypes is compared to the homozygous variant allele C/C genotype. Of particular interest was the potential effect modification of polymorphism effects by smoking status, alcohol consumption, self-reported Appalachian identity [31, 32], and risky sexual behavior adjusted for age. These were assessed via multivariable logistic regression models. Secondary analyses used multivariable logistic regression models to examine the effects of multiple polymorphisms, significant confounders identified a priori, as well as their interactions. Analyses were performed in SAS version 9.4 (SAS institute, Cary, NC). All *p* values and confidence intervals are not adjusted for multiple comparisons.

Genomic DNA isolation

Venous blood was collected directly into a PreAnalytiX PAXgene Blood DNA Tube for the stabilization and storage of samples prior genomic DNA isolation. All samples were stored at − 70 °C until processed with the PreAnalytiX PAXgene Blood DNA Kit. Purified genomic DNA samples were assessed for quality and quantity using an Agilent NanoDrop spectrophotometer.

Gene variant profiling

A pathway-based targeted genomic variance analysis of 9 SNPs and one polymorphic repeat variant was conducted on blood DNA (Table 1). 100 ng of genomic DNA was used for real-time PCR amplification using predesigned validated TaqMan SNP Genotyping Assays (Life Technologies). Each genotyping assay contained allele-specific VIC- or FAM-dye-labeled probes containing minor groove binder (MGB) probe for improved hybridization stability and non-fluorescent quencher (NFQ) for decreased fluorescence background. Following amplification as suggested by the manufacturer, genotyping data were analyzed using TaqMan Genotyper Software (Life Technologies). The polymorphic repeat in *TGFBR1* rs11466445 was genotyped and assessed for loss of heterozygosity (LOH) as previously reported by us [27].

Results

Characterization of the study participants

Survey assessments and biologicals were obtained from 163 women with cervical cancer and from 832 women without cervical cancer. Table 2 describes all participant demographics and disease status (ICC, normal cervical cytology). Women with ICC tended to be older and married, less often reported consuming alcohol in the last month, and more often reported previous abnormal Pap test than control participants. Frequencies of polymorphic variants are presented in Table 3 and Online Resource 1.

Examination of allele distribution for Hardy–Weinberg equilibrium

No statistically significant deviations from the expected Hardy–Weinberg (H–W) predictions were noted for any of the polymorphic alleles examined (Table 4), with the exception of *NQ01* (p value = 0.02).

Associations between genotypes, behavioral/environmental factors, and cancer status

Several polymorphic alleles demonstrated an interaction after adjusting for age with the recognized behavioral hazards of smoking (*TP53* rs1042522, *TGFB1* rs1800469), alcohol consumption (*NQO1* rs1800566), and sex before the age of 18 (*TGFBR1* rs11466445, *TGFBR1* rs7034462, *TGFBR1* rs11568785). In an over-dominant allele model adjusting on age for *TGFB1* rs1800469, never-smokers with homozygous genotypes A/A + G/G were 2.5-fold (OR 0.4, 95% CI 0.22–0.73, p value = 0.003) less likely to have cervical cancer compared to never-smokers with the heterozygous A/G genotype. This effect was not observed in ever-smokers (interaction p value = 0.02). A significant 3.1-fold increase

Table 1 Candidate polymorphic alleles associated with cervical cancer risk in Appalachian women

Gene ID[a]	Gene name	Variant ID[b]	Type of polymorphism	Cancer associations[c]
TGFB1	Transforming growth factor beta 1	*rs1800469*	Single-nucleotide variant	Cervical, Endometrial, Oropharyngeal, Breast, Lung, Pancreatic, Colorectal, Hepatocellular
TGFB1	Transforming growth factor beta 1	*rs1800470*	Single-nucleotide variant	Cervical, Nasopharyngeal, Breast, Lung, Head and Neck, Prostate
TGFB3	Transforming growth factor beta 3	*rs3917200*	Single-nucleotide variant	Cervical, Vulvar
TGFBR1	Transforming growth factor beta receptor 1	*rs868*	Single-nucleotide variant	Colorectal, Bladder, Hepatocellular
TGFBR1	Transforming growth factor beta receptor 1	*rs7034462*	Single-nucleotide variant	Breast, Colorectal
TGFBR1	Transforming growth factor beta receptor 1	*rs11568785*	Single-nucleotide variant	Colorectal
TGFBR1	Transforming growth factor beta receptor 1	*rs11466445*	Trinucleotide microsatellite	Cervical, Ovarian, Breast, Colorectal, Head and Neck, Bladder
CD83	CD83 molecule	*rs750749*	Single-nucleotide variant	Cervical, Vulvar
NQO1	NAD(P)H quinone dehydrogenase 1	*rs1800566*	Single-nucleotide variant	Cervical, Breast, Colorectal, Prostate, Lung, Bladder, Hepatocellular, Lymphoma
TP53	Tumor protein p53	*rs1042522*	Single-nucleotide variant	Cervical, Endometrial, Ovarian, Breast, Lung, Colorectal, Gastric, Head and Neck, Esophageal, Nasopharyngeal, Hepatocellular, Neuroblastoma, Mesothelioma, Leukemia

[a]Gene ID from HUGO Gene Nomenclature Committee (HGNC)

[b]Variant ID from NCBI dbSNP (https://www.ncbi.nlm.nih.gov/snp)

[c]Citations available via PubMed link within dbSNP

Table 2 Demographics and clinical characteristics of Appalachian study population

Variables	CARE II[a] ICC ($n=163$)	CARE II[b,c] normal cervical cytology controls ($n=79$)	CARE I normal cervical cytology controls ($n=763$)
Age, *mean* (SD)*	52.8 (12.4)	36.9 (11.7)	35.3 (13.2)
White	148 (97%)	51 (94%)	716 (95%)
Perceived stress, *mean* (SD)*	18.2 (7.9)	22.5 (7.8)	17.3 (7.5)
CES-D, *mean* (SD)[d]*	17.7 (13.1)	23.9 (13.9)	15.3 (11.7)
Smoking status			
Current	52 (34%)	23 (45%)	294 (39%)
Former	33 (22%)	6 (12%)	105 (14%)
Ever (current + former)	85 (56%)	29 (57%)	399 (53%)
Never	66 (44%)	22 (43%)	356 (47%)
Consumed alcohol in last month*	46 (32%)	22 (42%)	413 (55%)
Had abnormal pap test*	125 (88%)	27 (54%)	298 (41%)
Have current insurance coverage*	140 (93%)	43 (80%)	170 (80%)
Marital status			
Married	74 (49%)	21 (40%)	337 (45%)
Divorced	36 (24%)	11 (21%)	94 (13%)
Widowed	23 (15%)	2 (4%)	15 (2%)
Separated	2 (1%)	3 (6%)	22 (3%)
Single	14 (9%)	12 (23%)	217 (29%)
Living together	2 (1%)	4 (8%)	57 (8%)
Residence state ($n=987$)*			
OH	32 (22.1%)	0 (0%)	763 (100%)
WV	113 (77.9%)	79 (100%)	0 (0%)
Born state ($n=412$)*			
OH	32 (21.5%)	2 (3.8%)	186 (88.2%)
WV	90 (60.4%)	41 (78.8%)	5 (2.4%)
KY	12 (8.1%)	3 (5.8%)	1 (0.5%)
PA	5 (3.4%)	2 (3.8%)	8 (3.8%)
Other	10 (6.7%)	4 (7.7%)	11 (5.2%)
Parents from Appalachian county ($n=319$)*			
Both parents	95 (84.1%)	27 (73.0%)	118 (69.8%)
One parent	12 (10.6%)	8 (21.6%)	19 (11.2%)
None	6 (5.3%)	2 (5.4%)	32 (18.9%)
Appalachian identity ($n=953$)[e]*			
Yes	99 (76.2%)	41 (93.2%)	219 (29.3%)
No	31 (23.8%)	3 (6.8%)	528 (70.7%)
Sexual behavior (intercourse)			
Age at first sex, *mean* (SD)*	17.4 (3.2)	16.0 (3.6)	16.7 (2.8)
Total partners, *median* (IQR)	4 (5)	6 (6)	5 (6)
Had sex before age 18*	85 (59.9%)	38 (77.6%)	501 (67.9%)
Had $\geq$ 4 partners	87 (66.9%)	29 (74.4%)	425 (62.3%)
Partners ever treated for STD	11 (7.4%)	9 (17.0%)	69 (9.2%)
Self ever treated for STD	21 (14.1%)	7 (13.0%)	75 (10%)

[a]clinicaltrials.gov ID NCT02113514; Community Awareness, Resources and Education (CARE I) [28]

[b]clinicaltrials.gov ID NCT02797600; Community Awareness, Resources and Education (CARE II) [29]

[c]Negative cytology based upon Pap smear at time of blood collection

[d]Center for Epidemiologic Studies Depression Scale (CES-D) [80]

[e]Appalachian Identity, self-identified inclusion based on geographic, cultural, and SES perceptions

*p value <0.05 for comparison among three groups

Table 3 Genotype frequencies of polymorphic variants in the Appalachian study population

Variables	CARE II[a] cases ($n=163$)	CARE II controls ($n=79$)	CARE I[b] normal pap ($n=763$)
TGFB1 rs1800469			
A/A	15 (9%)	3 (4%)	72 (9%)
A/G	75 (46%)	32 (41%)	310 (41%)
G/G	73 (45%)	44 (56%)	381 (50%)
TGFB1 rs1800470			
A/A	52 (32%)	28 (35%)	294 (39%)
A/G	84 (52%)	43 (54%)	348 (46%)
G/G	27 (17%)	8 (10%)	121 (16%)
TGFB3 rs3917200			
A/A	135 (83%)	64 (81%)	644 (84%)
A/G	26 (16%)	14 (18%)	111 (15%)
G/G	2 (1%)	1 (1%)	8 (1%)
TGFBR1 rs868			
A/A	107 (66%)	56 (71%)	495 (65%)
A/G	52 (32%)	23 (29%)	235 (31%)
G/G	4 (2%)	(0%)	33 (4%)
TGFBR1 rs7034462			
C/C	136 (83%)	67 (85%)	635 (83%)
C/T	25 (15%)	12 (15%)	122 (16%)
T/T	2 (1%)	0 (0%)	6 (1%)
TGFBR1 rs11568785			
A/A	135 (83%)	70 (89%)	645 (85%)
A/G	26 (16%)	9 (11%)	113 (15%)
G/G	2 (1%)	0 (0%)	5 (1%)
TGFBR1 rs11466445			
Non 9A	34 (21%)	10 (13%)	141 (19%)
9A	129 (79%)	69 (87%)	618 (81%)
9A/9A	129 (79%)	69 (87%)	618 (81%)
9A/6A	32 (20%)	10 (13%)	136 (18%)
6A/6A	2 (1%)	0 (0%)	5 (0.7%)
CD83 rs750749			
C/C	5 (3%)	3 (4%)	26 (3%)
C/T	49 (30%)	25 (32%)	269 (35%)
T/T	109 (67%)	51 (65%)	468 (61%)
NQO1 rs1800566			
A/A	4 (3%)	1 (1%)	17 (2%)
A/G	56 (34%)	18 (23%)	256 (34%)
G/G	103 (63%)	60 (76%)	490 (64%)
TP53 rs1042522			
C/C	84 (52%)	43 (54%)	398 (52%)
C/G	63 (39%)	32 (41%)	317 (42%)
G/G	16 (10%)	4 (5%)	48 (6%)

[a]clinicaltrials.gov ID NCT02797600; Community Awareness, Resources and Education (CARE II) [29]

[b]clinicaltrials.gov ID NCT02113514; Community Awareness, Resources and Education (CARE I) [28]

Table 4 Allele frequencies of polymorphic variants in the Appalachian study population and association between genotypes and cancer status, adjusted for age

	Case	Control	OR^{a}_{age}	HWE p value[b]
TGFB1 rs1800469				
A/G	75 (46%)	342 (41%)	1.0	0.60
A/A	15 (9%)	75 (9%)	0.73 (0.36–1.47)	
G/G	73 (45%)	425 (50%)	0.75 (0.5–1.12)	
TGFB1 rs1800470				
A/G	84 (52%)	391 (46%)	1.0	0.57
A/A	52 (32%)	322 (38%)	0.79 (0.52–1.2)	
G/G	27 (17%)	129 (15%)	0.85 (0.49–1.48)	
TGFB3 rs3917200				
A/G	26 (16%)	125 (15%)	•	0.19
G/G	2 (1%)	9 (1%)		
A/A	135 (83%)	708 (84%)		
TGFBR1 rs868				
A/G	52 (32%)	258 (31%)	•	
G/G	4 (2%)	33 (4%)		
A/A	107 (66%)	551 (65%)		
TGFBR1 rs7034462				
C/T	25 (15%)	134 (16%)	•	0.89
C/C	136 (83%)	702 (83%)		
T/T	2 (1%)	6 (1%)		
TGFBR1 rs11568785				
A/G	26 (16%)	122 (14%)	•	0.93
G/G	2 (1%)	5 (1%)		
A/A	135 (83%)	715 (85%)		
TGFBR1 rs11466445				
9A/9A	129 (79%)	687 (82%)	1.14 (0.71–1.84)	0.33
9A/6A	32 (20%)	146 (18%)	1.0	
6A/6A	2 (1%)	5 (1%)		
CD83 rs750749				
C/T	49 (30%)	294 (35%)	1.0	0.10
C/C	5 (3%)	29 (3%)	0.98 (0.32–3.07)	
T/T	109 (67%)	519 (62%)	1.29 (0.85–1.96)	
NQO1 rs1800566				
A/G	56 (34%)	274 (33%)	•	0.02
G/G	103 (63%)	550 (65%)		
A/A	4 (2%)	18 (2%)		
TP53 rs1042522				
C/G	63 (39%)	349 (41%)	1.0	0.12
C/C	84 (52%)	441 (52%)	1.20 (0.8–1.79)	
G/G	16 (10%)	52 (6%)	1.27 (0.6–2.67)	

[a]Odds ratios are reported as OR (95% CI)

[b]HWE p values were obtained from Hardy–Weinberg equilibrium of the control group

•Odds ratio not provided due to small cell size

in the odds of cervical cancer was estimated with *TP53* rs1042522 G/G dominant model compared to C/C + C/G genotypes in never-smokers (aOR 3.1, 95% CI 1.1–8.5, p

value = 0.030), but not ever-smokers, with a marked interaction of smoking by genotype (p value = 0.02).

Risky sexual behavior also had an interaction effect on the association between several SNPs and cancer risk. Subjects who had no sex before 18 and with heterozygous non-9A genotype were 2.3-fold (OR 2.3, 95% CI 1.07–4.89, p value = 0.033) more likely to have cervical cancer compared to subjects who had no sex before 18 and with 9A genotype. In a recessive model for *TGFBR1* rs7034462, subjects who had no sex before 18 with C/T + T/T genotype were 2.3-fold (OR 2.31, 95% CI 1.03–5.22, p value = 0.044) more likely to have cervical cancer compared to subjects who had no sex before 18 with C/C genotype. A significant 2.8-fold decrease in the odds of cervical cancer was observed in the over-dominant model in subjects who had no sex before 18 and with C/C + T/T genotype (OR 0.36, 95% CI 0.16–0.84, p value = 0.017). Similar effects were also observed with *TGFBR1* rs11568785. In dominant model, subjects who had no sex before 18 and with A/A genotype were 2.7-fold (OR 0.37, 95% CI 0.16–0.83, p = 0.016) less likely to have cervical cancer compared to subjects who had no sex before 18 and with G/G + A/G genotype. The odds for subjects with A/A + G/G genotype is 3.1-fold (OR 0.32, 95% CI 0.14–0.74, p = 0.0075) less likely compared to subjects with A/G genotype in over-dominant model. All the effects were not observed in subjects who had sex before 18 (all interactions p value < 0.05). Unexpectedly, we noted a marginal interaction with the variables designed to assess self-identified "Appalachian Identity" and the *NQO1* rs1800566 polymorphic allele. Adjusted genetic models for each cancer risk factor are presented in Table 5, Online Resource 1.

Multivariable genetic models for cancer risk estimates

Results from multivariable genetic model with *TGFB1* rs1800469 and *TGFBR1* rs11568785 are presented in Table 6 (Online Resource 2, Online Resource 3). After adjusting for factors identified from previous genetic models, including age, smoking status, and risky sexual behavior, subjects who had no sexual intercourse before age 18 and with homozygous A/A + G/G genotype for *TGFBR1* rs11568785 were 3.03-fold (OR 0.33, 95% CI 0.14–0.78) reduced in odds of having cervical cancer compared to A/G genotype. This effect was not observed in subjects who had sex before age 18 (interaction effect p value = 0.0078). The significant main effect of *TGFB1* rs1800469 indicated that subjects with A/A + G/G genotype had 1.56-fold (OR 0.64, 95% CI 0.42–0.98) reduction in their odds of having cervical cancer compared to A/G genotype (p value = 0.037). Similar effects were observed in multivariable genetic model with *TGFB1* rs1800469 and *TGFBR1* rs7034462 (Table 6, Online Resource 2). After adjusting for age, smoking status, and risky sexual behavior, a significant 2.78-fold (OR 0.36, 95% CI 0.15–0.84) decrease in the odds of cervical cancer was observed in subjects who had no sex before age 18 and with homozygous C/C + T/T genotype compared to heterozygous C/T genotype. Such effect was not observed in subjects with risky sexual behavior (interaction effect p value = 0.0099).

Discussion

People living in the Appalachian Region continue to be at higher risk for the development of cancers, including cervical cancer, compared to the rest of the U.S. [33]. Some risk factors, such as high incidence of tobacco smoking, recovering but still significant levels of poverty, and improving but still poor access to healthcare certainly contribute to these higher incidences of cancer (Table 2). While targeted efforts have narrowed the gap in health disparities between Appalachian and non-Appalachian regions, overall cancer incidence remains elevated in the Appalachian counties. In fact, as suggested by Wilson et al. [33], poorer rates of cervical cancer screening may further underestimate the cancer burden in Appalachian counties. However, another factor that can modulate cancer risk is susceptibility mediated by genomic variation in the form of polymorphic alleles [34] (Table 3). Genetic variants can alter biological responses depending upon the gene–environment interactions present [35]. These studies report for the first time the role of genetic variation as a risk cofactor for the development of invasive cervical cancer in Appalachian women.

The *TGFB1 rs1800469* polymorphism is associated with a decreased risk of developing cervical cancer in Appalachian women

We targeted genomic variants within the TGFB signaling cascade in response to this pathway's established history with EMT and epithelial carcinogenesis. Polymorphism in TGFB cytokines as well as their cognate receptors would allow mechanisms to potentially mediate the susceptibility of cervical epithelia initiation, promotion, and transition into malignant invasive disease (Table 4). The *TGFB1* rs1800469 polymorphism, commonly reported as C-509T, is a single-nucleotide variation in the *TGFB1* gene that functionally alters promoter region activity. The variant T allele is common with a MAF = 0.3680, and *TGFB1* rs1800469 demonstrates an average heterozygosity of 0.465 ± 0.127 (dbSNP) [36]. The C-509T variation has been reported to be associated with disease states ranging from asthma [37] to aplastic anemia [38]. Several studies support the finding that the homozygous C/C genotype is associated with low expression of TGFB1 and the homozygous variant allele T/T genotype is marked by TGFB1 cytokine over-expression,

Table 5 Association between genotypes and cancer status, adjusted for participant age

Interaction effect tested: smoking status							
	Never-smokers ($n=443$)			Ever-smokers ($n=511$)			p value
	Control	Case	OR (95% CI)	Control	Case	OR (95% CI)	
TGFB1 rs1800469 Over-dominant model							
A/G	156 (41%)	38 (58%)	1.0 (Referent)	169 (40%)	32 (38%)	1.0 (Referent)	
A/A-G/G	222 (59%)	27 (42%)	0.4 (0.22–0.73)	258 (60%)	52 (62%)	1.02 (0.59–1.75)	0.02
TP53 rs1042522 Dominant model							
C/C-C/G	361 (95%)	57 (88%)	1.0 (Referent)	395 (93%)	78 (93%)	1.0 (Referent)	
G/G	17 (5%)	8 (12%)	3.08 (1.11–8.50)	32 (7%)	6 (7%)	0.55 (0.19–1.56)	0.02
Interaction effect tested: alcohol consumption							
	No alcohol consumption last month ($n=468$)			Consumed alcohol last month ($n=480$)			p value
	Control	Case	OR (95% CI)	Control	Case	OR (95% CI)	
NQO1 rs1800566 Recessive model							
G/G	246 (66%)	59 (60%)	1.0 (Referent)	276 (64%)	33 (72%)	1.0 (Referent)	
A/G-A/A	124 (34%)	39 (40%)	1.4 (0.83–2.36)	158 (36%)	13 (28%)	0.58 (0.28–1.19)	0.05
NQO1 rs1800566 Over-dominant model							
A/G	116 (31%)	37 (38%)	1.0 (Referent)	148 (34%)	12 (26%)	1.0 (Referent)	
A/A-G/G	254 (69%)	61 (62%)	0.70 (0.41–1.18)	286 (66%)	34 (74%)	1.72 (0.82–3.60)	0.05
Interaction effect tested: sex before 18 years							
	No sex before 18 years ($n=305$)			Sex before 18 years ($n=621$)			p value
	Control	Case	OR (95% CI)	Control	Case	OR (95% CI)	
TGFBR1 rs11466445 Polymorphism							
Non 9A	48 (20%)	17 (30%)	2.29 (1.07–4.89)	93 (17%)	15 (18%)	0.74 (0.37–1.49)	0.034
9A	198 (80%)	40 (70%)	1.0 (Referent)	443 (83%)	68 (82%)	1.0 (Referent)	
TGFBR1 rs7034462 Recessive model							
C/C	207 (83%)	44 (77%)	1.0 (Referent)	449 (83%)	70 (84%)	1.0 (Referent)	
C/T-T/T	41 (17%)	13 (23%)	2.31 (1.03–5.22)	89 (17%)	13 (16%)	0.66 (0.32–1.36)	0.025
TGFBR1 rs7034462 Over-dominant model							
C/T	36 (15%)	12 (21%)	1.0 (Referent)	88 (16%)	13 (16%)	1.0 (Referent)	
C/C-T/T	212 (85%)	45 (79%)	0.36 (0.16–0.84)	450 (84%)	70 (84%)	1.49 (0.72–3.09)	0.013
TGFBR1 rs11568785 Dominant model							
G/G-A/G	37 (15%)	14 (25%)	1.0 (Referent)	81 (15%)	13 (16%)	1.0 (Referent)	
A/A	211 (85%)	43 (75%)	0.37 (0.16–0.83)	457 (85%)	70 (84%)	1.52 (0.73–3.2)	0.012
TGFBR1 rs11568785 Over-dominant model							
A/G	34 (14%)	13 (23%)	1.0 (Referent)	79 (15%)	13 (16%)	1.0 (Referent)	
A/A-G/G	214 (86%)	44 (77%)	0.32 (0.14–0.74)	459 (85%)	70 (84%)	1.46 (0.69–3.09)	0.0083
Interaction effect tested: self-reported Appalachian identity							
	Non-Appalachian identity ($n=561$)			Appalachian identity ($n=357$)			p value
	Control	Case	OR (95% CI)	Control	Case	OR (95% CI)	
NQO1 rs1800566 Recessive model							
G/G	330 (62%)	23 (74%)	1.0 (Referent)	178 (68%)	57 (59%)	1.0 (Referent)	
A/G-A/A	200 (38%)	8 (26%)	0.52 (0.22–1.24)	82 (32%)	40 (41%)	1.55 (0.90–2.68)	0.04

Interaction effects with ≧ 4 partners and parent county were also tested. Only significant interaction effects were reported here. Interaction tests were not performed on specific models if any cell size < 5

Table 6 Multivariable logistic regression results examining the association between *TGFB1–TGFBR1* genotypes and cancer status

	Levels	OR (95% CI)	*p* value
Age[a]	10 years increase	2.75 (2.3–3.28)	< 0.0001
Smoking status	Never-smoker	0.79 (0.51–1.23)	0.3
	Ever-smoker	1.0	
TGFB1 rs1800469	A/A + G/G	0.64 (0.42–0.98)	0.037
	A/G	1.0	
TGFBR1 rs11568785 by risky sexual behavior interaction			0.0078
TGFBR1 rs11568785 when subject had no sex before 18 years	A/A + G/G	0.33 (0.14–0.78)	
	A/G	1	
TGFBR1 rs11568785 when subject had sex before 18 years	A/A + G/G	1.58 (0.74–3.35)	
	A/G	1.0	
Age[b]	10 years increase	2.76 (2.31–3.30)	< 0.0001
Smoking status	Never-smoker	0.77 (0.50–1.20)	0.25
	Ever-smoker	1.0	
TGFB1 rs1800469	A/A + G/G	0.63 (0.41–0.95)	0.029
	A/G	1.0	
TGFBR1 rs7034462 by risky sexual behavior interaction			0.0099
TGFBR1 rs7034462 when subject had no sex before 18 years	C/C + T/T	0.36 (0.15–0.84)	
	C/T	1.0	
TGFBR1 rs7034462 when subject had sex before 18 years	C/C + T/T	1.60 (0.76–3.36)	
	C/T	1.0	

[a]Model with patient's age, smoking status, risky sexual behavior before 18 years of age, *TGFB1* rs1800469 genotype, *TGFBR1* rs11568785 genotype and its interaction with risky sexual behavior before 18 years of age. ($n = 918$)

[b]Model with patient's age, smoking status, risky sexual behavior before age 18, *TGFB1* genotype rs1800469, *TGFBR1* genotype rs7034462 and its interaction with risky sexual behavior before age 18. ($n = 918$)

possibly by enhanced transcription factor binding [37, 39] in the promoter. The heterozygote C/T genotype is often of an intermediate phenotype for TGFB1 expression levels. Interestingly, Jin et al. [40] found that the C/T heterozygote was in fact independently associated with a significantly decreased gastric cancer risk, and in a co-dominant genetic model the T/T + C/T genotypes were significantly associated with a decreased risk of gastric cancer compared to C/C homozygotes.

Within the cervical cancer microenvironment, mesenchymal stromal cells (MSCs) can differentiate and drive tumorigenesis by promoting proliferation, angiogenesis, and migration as part of EMT. The increased expression of TGFB1 and other immune suppressors initiates a microenvironment that is resistant to the cell-mediated tumor response of cytotoxic T cells (CTLs) and natural killer cells. These high levels of TGFB1 coupled with elevated IL-10 lead to suppression of HLA-1 and the necessary trajectory for a cell to escape from immune surveillance and destruction [41]. We report that in an over-dominant model, the C/C + T/T genotypes demonstrate a 2.5-fold decreased risk for developing cervical cancer compared to the heterozygous C/T genotype (Online Resource 3), and that this effect is specific for never-smokers. How genotypes that have previously been associated with either basal levels (C/C) or over-expression (T/T) of TGFB1 confer a decreased risk of cervical cancer in this Appalachian population remains unanswered and may reflect the unique gene–environment interactions of this population.

The TP53 rs1042522 polymorphism is associated with an increased risk of developing cervical cancer in Appalachian women

The p53 protein encoded by *TP53* has been characterized as the "guardian of the genome" due to its intrinsic role in regulating the cell cycle and maintaining genomic integrity [42]. Due to this essential role in suppressing cell growth when genetic damage is detected, the discovery that p53 was mutated in about 50% of all cancers, including cervical cancer, was striking. Common activators of p53 include DNA-damaging chemical exposures, low oxygen conditions, and other metabolic stressors [43]. The amino acid change from proline to arginine that results from the single-nucleotide change in codon 72 of *TP53* rs1042522 is one of the most studied functional polymorphic events in the human genome. The ancestral C allele (encoding proline) demonstrates a diminished proapoptotic activity level compared

to the variant G allele (MAF = 0.4571, dbSNP) [36]. The high-frequency variant G allele (encoding arginine) has been associated with advanced cancer staging and metastatic potential in addition to its enhanced proapoptotic activities [44].

While the impact of *TP53* rs1042522 on cervical cancer promotion remains an active area of debate, consistent with our results, a meta-analysis of 70 case–control studies (14999 cases, 8195 controls) by Jee et al. [45] reported that G/G homozygotes demonstrated an increased risk relative to C/G heterozygotes (OR 1.2, 95% CI 1.1–1.3, p value = 0.001). However, a subsequent meta-analysis with pooled data from 49 case–control studies (7,946 cases, 7,888 controls) proposed that following subgroup analyses, the impact of the G/G genotype was more likely due to small case numbers and deviations from HWE rather than biology [46]. We report that in a dominant model, G/G homozygotes have a 3.1-fold increase in the odds of having cervical cancer compared to either C/C or C/T genotypes. Similarly, in a co-dominant model, individuals with the G/G genotype are 3.65-fold more likely to have cervical cancer than those that are C allele carriers (Online Resource 3). Consistent with non-Appalachian general population studies, the genetic liability affiliated with the *TP53* rs1042522 G risk allele is strongly maintained in Appalachian populations.

HPV and smoking

The human papillomavirus (HPV) is a dsDNA virus whose high-risk variants are causally associated with the vast majority of cervical cancers. While most HPV infections are transient and self-resolve, some become chronic and the presence of HPV in ICC is > 99% [47]. Importantly, the E6 and E7 proteins of HPV are able to functionally interact with the p53 and pRB tumor suppressors, respectively, and alter the regulation of cell cycle progression checkpoints. Specifically, E6 can bind to p53 and promote ubiquitin-mediated proteasomal degradation [48]. Consequently, in the absence of p53, genetically compromised cells are allowed to progress through the cell cycle. The unrepaired DNA damage can become fixed, and subsequent mutations contribute to the initiation of cervical carcinogenesis [48].

Tobacco smoking has been causally associated with the etiology of several cancers, including those of the lung, head and neck, esophagus, stomach, liver, pancreas, bladder, and cervix [49–55]. Exposure to tobacco smoke introduces carbon monoxide, hydrogen cyanide, nicotine, polyaromatic hydrocarbons such as benzo[a]pyrene (B[a]P), tobacco-specific carcinogenic nitrosamines like 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), and N'-nitrosonornicotine (NNN), as well as other toxic or carcinogenic compounds into the body. Some of these compounds (NNN, NNK, B[a]P) form DNA adducts that ultimately damage DNA repair genes, inhibit tumor suppressor genes and activate oncogenes, leading to loss of homeostatic growth control [56, 57]. Other tobacco compounds (nicotine) facilitate deregulated growth by directly binding to cell receptors and promoting inappropriate cell proliferation and migration. Directly or indirectly, often via the generation of ROS and RNS, many of these compounds induce proinflammatory states that provide a favorable tumor microenvironment to facilitate cervical cancer development.

However, debate has continued as to the possible confounding role of HPV-positivity in studies where the virus is considered causally associated. Importantly, a meta-analysis of case–control studies by Plummer et al. [58] supported an independent role for smoking by demonstrating a relative risk of 2.17 (95% CI 1.46–3.22) for HPV-positive cases relative to HPV-positive controls. We report a striking difference in cervical cancer risk and genetic variants in Appalachian women depending upon the smoking status of the cases and controls. For both *TGFB1* rs1800469 and *TP53* rs1042522, the contribution of the genomic variant was seen only in never-smokers. According to the CDC, despite a decrease over the last 10 years, 16.8% of the adult U.S. population currently smokes cigarettes. Importantly, 14.8% of adult women and 26.3% of adults below the poverty level in the general population smoke. In our study, 34.4% of women with cervical cancer were current smokers, while 45% of non-cancer controls reported themselves as current smokers (Table 1). The powerful health-modifying events that occur with tobacco use may explain the inability to detect a cervical cancer risk-modifying component for the *TGFB1* and *TP53* genomic variants in ever-smokers (Online Resource 2).

Interactions between genomic variants, social-behavioral measurements, and clinicopathologic characteristics

We further investigated what additional social-behavioral characteristics may contribute to overall cervical cancer risk as part of Appalachian gene–environment interactions. Using a model including age, smoking status, risky sexual behavior before 18 years of age, *TGFB1* rs1800469 genotype, and *TGFBR1* rs11568785 genotype, we show that there is an association between the *TGFB1* ligand promoter variant rs1800469 and the *TGFBR1* intronic variant rs11568785 only when there was no risky sexual behavior (Table 5, Online Resource 3). Additionally, using a similar model interrogating *TGFB1 rs1800469* and the intronic *TGFBR1 rs7034462*, we demonstrate that there is an association only in the absence of risky sexual behavior (Table 6, Online Resource 3).

The combination of common low-penetrance genetic variants, multifactorial traits, and population heterogeneity

makes definitive genome association studies challenging in defining an accurate cervical cancer genetic risk [59, 60]. In many parts of Appalachian, these challenges are potentially confounded by unique socioeconomic factors, risky social behaviors, healthcare insufficiency, and incompletely surveyed environmental exposome diversity [61–64]. Prior studies have already suggested a strong relationship between socioeconomic status (SES) and depression in Appalachian female smokers [65], and provide a foundational list for potential modifiers and mediators within such an Appalachian exposome [62, 63, 65]. Emerging transdisciplinary geospatial approaches will further define the complex interactions between genomics, environmental exposures, and social behavior risks while engaging community-based participatory research partnerships [61, 64, 66–68]. The complex representation of "Appalachian identity" is as elegantly described by BE Smith [69] as "the impossible necessity of Appalachian studies." Our findings that an interaction effect is present between the NAD(P)H quinone oxidoreductase 1 (*NQO1)* variant rs1800566, self-reported Appalachian identity, and cervical cancer risk (Table 5) provide intriguing insight into the complex dynamics of defining cancer risk in this population.

Abnormal NQO1 activity has been associated with cancer risk for a number of malignancies (bladder, breast, colon, cervical, lung, and pancreas) [70–74] and is likely due to the enzyme's ability to mediate reactive oxygen species availability and stabilize tumor suppressor proteins, including TP53 [75, 76], by inhibiting proteasomal degradation. This ubiquitous but readily inducible enzyme may represent a useful genetic prognostic biomarker of cervical cancer risk in Appalachian women as the meaningful environmental exposures [77, 78] that couple its association with indices of Appalachian self-identity are defined with increased resolution. Again, it is intriguing, but with caution that a possible association between a social identifier such as Appalachian identity and a functional SNP involved with detoxification must be discussed. Previously, the A/A *NQO1* rs1800566 susceptibility allele has been associated with increased cervical cancer risk [79], consistent with an increased OR 1.55 (09.0–2.68) presented here (Table 5) for participant identifying as Appalachian. It would require higher-resolution mapping of geographic localization, workplace risks, access to healthcare services, environmental exposome, and family health history, to reasonably speculate on the mild interaction effect seen (p value = 0.04), while noting the comparatively small numbers available for analysis. One hypothesis would be that self-identified Appalachians have an increased likelihood of exposure to the well-documented cancer risk factors endemic to rural Appalachian regions. The association of Appalachian identity and TP53 genomic variance as a social determinate of health is likely to be as an aggregate biomarker of cancer susceptibility rather than as genetic driver of disease risk per se. It is possible that as a functional SNP the cumulative environmental triggers that permit the TP53 phenotype to be potentially impactful during cervical carcinogenesis may be unique within the Appalachian population.

It should be pointed out that there are notable differences in Appalachian self-identification between CARE I controls and CARE II controls (29.3% vs. 93.2%) (Table 2), and that the pooled data demonstrating an interaction between self-reported Appalachian identity and *NQO1* rs1800566 make use of a comparatively small number of cases for the genetic models (Table 5). These findings further the need to interrogate the behavioral and genetic risk factors potentially contributing to the enhanced cervical cancer risk presented across the Appalachian region, notably in the Central and Southern Appalachian Regional Commission (ARC) subregions addressed in this study. It is important to note that CARE I and CARE II were complementary but non-overlapping studies with distinct primary outcome measures. Consequently, there are different catchment clinics used for participant enrollment for CARE I and CARE II, although all clinics were from the same Appalachian geographic regions as previously noted. Additionally, there are clear differences in age between CARE II controls (36.9 ± 11.7) and ICC cases (52.8 ± 12.4), with the controls between CARE I and CARE II showing only nominal differences. The risk for cervical cancer increases with age with concomitant environmental exposures such as HPV infection and cancer risk-enhancing behaviors (smoking, early sexual activity, and early full-term pregnancy) [47]. Chronic persistent HPV infection increases with age, possibly confounding the associations suggested for the interactions between common low-penetrance susceptibility alleles and cervical cancer: the exposure and opportunity for non-cleared persistent HPV infection is greater in older women. The ability to better match participants would further strengthen the confidence in our conclusions.

Persistent HPV infection, exposure to tobacco smoke and alcohol, and risk-enhancing sexual behaviors clearly drive the cumulative risk for developing cervical cancer. The genetic susceptibility role that common, low-penetrance alleles play in overall risk may ultimately be minor compared to these other dominant risk factors for cervical cancer. However, that role is not trivial, especially in the case of known functional SNP phenotypes that provide both a basis for prognostic genetic biomarker screening, as well as offer the potential for predictive biomarker assessment and possible intervention measures.

Acknowledgments The study was supported by grants from the National Institutes of Health's National Cancer Institute (P50 CA105632, P30 CA016058) and National Center for Advancing Translational Science (UL1TR001070).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Ervik M, Lam F, Ferlay J, Mery L, Soerjomataram I, Bray F (2018) Cancer today. International Agency for Research on Cancer. http://gco.iarc.fr/today/home. Accessed Feb 2019
2. Ferlay J, Soerjomataram I, Dikshit R et al (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136(5):E359–E386. https://doi.org/10.1002/ijc.29210
3. Bray F, Ren J-S, Masuyer E, Ferlay J (2013) Global estimates of cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer 132(5):1133–1145. https://doi.org/10.1002/ijc.27711
4. Appalachian Community Cancer Network (2010) Addressing the cancer burden in Appalachian communities
5. Appalachian Regional Commission (2015) Appalachia then and now: examining changes to the Appalachian region since 1965
6. Appalachian Regional Commission (2019) Appalachian Regional Commission. https://www.arc.gov/index.asp. Accessed Feb 2019
7. Behringer B, Friedell GH (2006) Appalachia: where place matters in health. Prev Chronic Dis. 3(4):A113
8. Behringer B, Friedell GH, Dorgan KA et al (2007) Understanding the challenges of reducing cancer in Appalachia: addressing a place-based health disparity population. Californian J Heal Promot Disparities Soc Justice. 5:40–49
9. Wewers ME, Katz M, Fickle D, Paskett ED (2006) Risky behaviors among Ohio Appalachian adults. Prev Chronic Dis. 3(4):A127
10. Lengerich EJ, Tucker TC, Powell RK et al (2005) Cancer incidence in Kentucky, Pennsylvania, and West Virginia: disparities in Appalachia. J Rural Health. 21(1):39–47
11. Pickup M, Novitskiy S, Moses HL (2013) The roles of TGFβ in the tumour microenvironment. Nat Rev Cancer 13(11):788–799. https://doi.org/10.1038/nrc3603
12. Pasche B, Pennison MJ, Jimenez H, Wang M (2014) TGFBR1 and cancer susceptibility. Trans Am Clin Climatol Assoc 125:300–312
13. Martelossi Cebinelli GC, Paiva Trugilo K, Badaró Garcia S, Brajão de Oliveira K (2016) TGF-β1 functional polymorphisms: a review. Eur Cytokine Netw 27(4):81–89. https://doi.org/10.1684/ecn.2016.0382
14. Levovitz C, Chen D, Ivansson E, Gyllensten U, Finnigan JP, Alshawish S, Zhang W, Schadt EE, Posner MR, Genden EM, Boffetta P, Sikora AG (2014) TGFβ receptor 1: an immune susceptibility gene in HPV-associated cancer. Cancer Res 74(23):6833–6844. https://doi.org/10.1158/0008-5472.CAN-14-0602-T **Epub 2014 Oct 1**
15. Boone SD, Baumgartner KB, Baumgartner RN, Connor AE, Pinkston CM, John EM, Hines LM, Stern MC, Giuliano AR, Torres-Mejia G, Brock GN, Groves FD, Kerber RA, Wolff RK, Slattery ML (2013) Associations between genetic variants in the TGF-β signaling pathway and breast cancer risk among Hispanic and non-Hispanic white women. Breast Cancer Res Treat 141(2):287–297. https://doi.org/10.1007/s10549-013-2690-z (**Epub 2013 Sep 14**)
16. Bellam N, Pasche B (2010) Tgf-beta signaling alterations and colon cancer. Cancer Treat Res 155:85–103. https://doi.org/10.1007/978-1-4419-6033-7_5
17. Di QG, Sun BH, Jiang MM, Du JF, Mai ZT, Zhang X, Zhou LR, Chi YM, Lv J (2017) Polymorphisms of -800G/A and +915G/C in TGF-β1 gene and lung cancer susceptibility. Oncol Lett. 14(1):733–736. https://doi.org/10.3892/ol.2017.6173 (**Epub 2017 May 16**)
18. Wan PQ, Wu JZ, Huang LY, Wu JL, Wei YH, Ning QY (2015) TGF-β1 polymorphisms and familial aggregation of liver cancer in Guangxi, China. Genet Mol Res 14(3):8147–8160. https://doi.org/10.4238/2015.July.27.3
19. Shi Q, Wang X, Cai C, Yang S, Huo N, Liu H (2017) Association between TGF-β1 polymorphisms and head and neck cancer risk: a meta-analysis. Front Genet. 8:169. https://doi.org/10.3389/fgene.2017.00169
20. Hampras SS, Sucheston-Campbell LE, Cannioto R, Chang-Claude J, Modugno F, Dörk T, Hillemanns P, Preus L, Knutson KL, Wallace PK et al (2016) Assessment of variation in immunosuppressive pathway genes reveals TGFBR2 to be associated with risk of clear cell ovarian cancer. Oncotarget. 7(43):69097–69110. https://doi.org/10.18632/oncotarget.10215
21. Yang L, Wang YJ, Zheng LY, Jia YM, Chen YL, Chen L, Liu DG, Li XH, Guo HY, Sun YL, Tian XX, Fang WG (2016) Genetic polymorphisms of TGFB1, TGFBR1, SNAI1 and TWIST1 are associated with endometrial cancer susceptibility in Chinese han women. PLoS ONE 11(5):e0155270. https://doi.org/10.1371/journal.pone.0155270
22. Singh H, Jain M, Mittal B (2009) Role of TGF-beta1 (-509C > T) promoter polymorphism in susceptibility to cervical cancer. Oncol Res 18(1):41–45
23. Al-Harbi NM, Bin Judia SS, Mishra KN, Shoukri MM, Alsbeih GA (2017) Genetic predisposition to cervical cancer and the association with XRCC1 and TGFB1 polymorphisms. Int J Gynecol Cancer. 27(9):1949–1956. https://doi.org/10.1097/IGC.0000000000001103
24. Gautam KA, Pooja S, Sankhwar SN, Sankhwar PL, Goel A, Rajender S (2015) c.29C > T polymorphism in the transforming growth factor-β1 (TGFB1) gene correlates with increased risk of urinary bladder cancer. Cytokine 75(2):344–348. https://doi.org/10.1016/j.cyto.2015.05.017
25. Pasche B, Knobloch TJ, Bian Y et al (2005) Somatic acquisition and signaling of TGFBR1*6A in cancer. J Am Med Assoc 294(13):1634–1646. https://doi.org/10.1001/jama.294.13.1634
26. Zhang H-T, Zhao J, Zheng S-Y, Chen X-F (2005) Is *TGFBR1* *6A really associated with increased risk of cancer? J Clin Oncol 23(30):7743–7744. https://doi.org/10.1200/JCO.2005.02.9108
27. Bian Y, Knobloch TJ, Sadim M et al (2007) Somatic acquisition of TGFBR1*6A by epithelial and stromal cells during head and neck and colon cancer development. Hum Mol Genet 16(24):3128–3135. https://doi.org/10.1093/hmg/ddm274
28. ClinicalTrials.gov. community awareness, resources and education (CARE I): NCT02113514 (2018). https://www.clinicaltrials.gov/ct2/show/NCT02113514.
29. ClinicalTrials.gov. Community awareness, resources and education (CARE II): NCT01299714 (2018). https://www.clinicaltrials.gov/ct2/show/NCT01299714.
30. Reiter PL, Katz ML, Ruffin MT et al (2013) HPV prevalence among women from Appalachia: results from the CARE project. PLoS ONE 8(8):e74276. https://doi.org/10.1371/journal.pone.0074276
31. Weaver R (2016) Appalachia, USA: an empirical note and agenda for future research. J Rural Soc Sci. 31(1):23–52

32. Reiter PL, Katz ML, Ferketich AK, Ruffin MT, Paskett ED (2009) Appalachian self-identity among women in Ohio appalachia. J Rural Commun Psychol E12(1)
33. Wilson RJ, Ryerson AB, Singh SD, King JB (2016) Cancer incidence in Appalachia, 2004–2011. Cancer Epidemiol Biomarkers Prev 25(2):250–258. https://doi.org/10.1158/1055-9965.EPI-15-0946
34. Erichsen HC, Chanock SJ (2004) SNPs in cancer research and treatment. Br J Cancer 90(4):747–751. https://doi.org/10.1038/sj.bjc.6601574
35. Guengerich FP (1998) The environmental genome project: functional analysis of polymorphisms. Environ Health Perspect 106(7):365–368
36. dbSNP: database for short genetic variations scope and access searching for and displaying SNP records
37. Silverman ES, Palmer LJ, Subramaniam V et al (2004) Transforming growth factor-β_1 promoter polymorphism C–509T Is associated with asthma. Am J Respir Crit Care Med 169(2):214–219. https://doi.org/10.1164/rccm.200307-973OC
38. Afify RAA, Salama N (2013) Correlation of transforming growth factor beta-1 gene polymorphisms C-509T and aplastic anemia. Comp Clin Path. 22(4):755–760. https://doi.org/10.1007/s00580-012-1478-6
39. Grainger DJ, Heathcote K, Chiano M et al (1999) Genetic control of the circulating concentration of transforming growth factor type beta1. Hum Mol Genet 8(1):93–97
40. Jin G, Wang L, Chen W et al (2007) Variant alleles of TGFB1 and TGFBR2 are associated with a decreased risk of gastric cancer in a Chinese population. Int J Cancer 120(6):1330–1335. https://doi.org/10.1002/ijc.22443
41. García-Rocha R, Moreno-Lafont M, Mora-García ML et al (2015) Mesenchymal stromal cells derived from cervical cancer tumors induce TGF-β1 expression and IL-10 expression and secretion in the cervical cancer cells, resulting in protection from cytotoxic T cell activity. Cytokine 76(2):382–390. https://doi.org/10.1016/j.cyto.2015.09.001
42. Lane DP (1992) p53, guardian of the genome. Nature 358(6381):15–16. https://doi.org/10.1038/358015a0
43. Gracy G, Sadhna K, Jacqueline J, Deepika K (2014) Highlights of p53 mutation and it's role in cervical cancer metastasis. Int J Biol Med Res 5(1):3772–3779
44. Chen R, Liu S, Ye H et al (2015) Association of p53 rs1042522, MDM2 rs2279744 and p21 rs1801270 polymorphisms with retinoblastoma risk and invasion in a Chinese population. Sci Rep. 5(1):13300. https://doi.org/10.1038/srep13300
45. Jee SH, Won SY, Yun JE, Lee JE, Park JS, Ji SS (2004) Polymorphism p53 codon-72 and invasive cervical cancer: a meta-analysis. Int J Gynecol Obstet. 85(3):301–308. https://doi.org/10.1016/j.ijgo.2003.08.017
46. Klug SJ, Ressing M, Koenig J et al (2009) TP53 codon 72 polymorphism and cervical cancer: a pooled analysis of individual data from 49 studies. Lancet Oncol. 10(8):772–784. https://doi.org/10.1016/S1470-2045(09)70187-1
47. Stumbar SE, Stevens M, Feld Z (2019) Cervical cancer and its precursors: a preventative approach to screening, diagnosis, and management. Prim Care 46(1):117–134. https://doi.org/10.1016/j.pop.2018.10.011 (**Epub 2018 Dec 22**)
48. Wang X, Huang X, Zhang Y (2018) Involvement of human papillomaviruses in cervical cancer. Front Microbiol. 9:2896. https://doi.org/10.3389/fmicb.2018.02896
49. Khani Y, Pourgholam-Amiji N, Afshar M, Otroshi O, Sharifi-Esfahani M, Sadeghi-Gandomani H, Vejdani M, Salehiniya H (2018) Tobacco smoking and cancer types: a review. Biomed Res Ther 5(4):2142–2159. https://doi.org/10.15419/bmrat.v5i4.428
50. Johnson CA, James D, Marzan A, Armaos M (2019) Cervical cancer: an overview of pathophysiology and management. Semin Oncol Nurs. https://doi.org/10.1016/j.soncn.2019.02.003
51. Su B, Qin W, Xue F, Wei X, Guan Q, Jiang W, Wang S, Xu M, Yu S (2018) The relation of passive smoking with cervical cancer: a systematic review and meta-analysis. Medicine 97(46):e13061. https://doi.org/10.1097/MD.0000000000013061
52. Collins S, Rollason TP, Young LS, Woodman CB (2010) Cigarette smoking is an independent risk factor for cervical intraepithelial neoplasia in young women: a longitudinal study. Eur J Cancer 46(2):405–411. https://doi.org/10.1016/j.ejca.2009.09.015 (**Epub 2009 Oct 12**)
53. International Agency for Research on Cancer (2012) IARC monographs on the evaluation of carcinogenic risks to humans. Personal habits and indoor combustions: tobacco smoking, vol 100E. IARC, Lyon
54. Fonseca-Moutinho JA (2011) Smoking and cervical cancer. ISRN Obstet Gynecol. 2011:1–6. https://doi.org/10.5402/2011/847684
55. Hecht SS (2014) It is time to regulate carcinogenic tobacco-specific nitrosamines in cigarette tobacco. Cancer Prev Res. 7(7):639–647. https://doi.org/10.1158/1940-6207.CAPR-14-0095
56. Xue J, Yang S, Seng S (2014) Mechanisms of cancer induction by tobacco-specific NNK and NNN. Cancer 6(2):1138–1156
57. Wieringa HW, van der Zee AG, de Vries EG, van Vugt MA (2016) Breaking the DNA damage response to improve cervical cancer treatment. Cancer Treat Rev 42:30–40. https://doi.org/10.1016/j.ctrv.2015.11.008 (**Epub 2015 Nov 24**)
58. Plummer M, Herrero R, Franceschi S et al (2003) Smoking and cervical cancer: pooled analysis of the IARC multi-centric case–control study. Cancer Causes Control 14(9):805–814
59. Gibson G (2012) Rare and common variants: twenty arguments. Nat Rev Genet 13(2):135–145. https://doi.org/10.1038/nrg3118
60. Saint Pierre A, Génin E (2014) How important are rare variants in common disease? Brief Funct Genomics. 13(5):353–361. https://doi.org/10.1093/bfgp/elu025
61. Woychik R (2014) Where exposure science and citizen science meet. Research Triangle Environmental Health Collaborative. Environ Heal Summit @BULLET Recomm from Res Triangle Environ Heal Collab Environ Heal Summit
62. Langston MA, Levine RS, Kilbourne BJ et al (2014) Scalable combinatorial tools for health disparities research. Int J Environ Res Public Health. 11(10):10419–10443. https://doi.org/10.3390/ijerph111010419
63. Juarez PD, Matthews-Juarez P, Hood DB et al (2014) The public health exposome: a population-based, exposure science approach to health disparities research. Int J Environ Res Public Health. 11(12):12866–12895. https://doi.org/10.3390/ijerph111212866
64. U.S. Department of Health and Human Services (USDHHS), National Institute of Health (NIH) NI of EHS (NIEHS) (2017) Advancing science, improving health: a plan for environmental health research. NIEHS 2012–2017. https://www.niehs.nih.gov/about/strategicplan/strategicplan2012/index.cfm. Accessed 25 Jun 2018
65. Post DM, Gehlert S, Hade EM, Reiter PL, Ruffin M, Paskett ED (2013) Depression and SES in women from Appalachia. J Rural Ment Heal. 37(1):2–15. https://doi.org/10.1037/rmh0000001
66. Oyana TJ, Matthews-Juarez P, Cormier SA, Xu X, Juarez PD (2015) Using an external exposome framework to examine pregnancy-related morbidities and mortalities: implications for health disparities research. Int J Environ Res Public Health. 13(1):ijerph13010013. https://doi.org/10.3390/ijerph13010013
67. Schootman M, Gomez SL, Henry KA et al (2017) Geospatial approaches to cancer control and population sciences. Cancer Epidemiol Biomark Prev 26(4):472–475. https://doi.org/10.1158/1055-9965.EPI-17-0104

68. Korycinski RW, Tennant BL, Cawley MA, Bloodgood B, Oh AY, Berrigan D (2018) Geospatial approaches to cancer control and population sciences at the United States cancer centers. Cancer Causes Control 29(3):371–377. https://doi.org/10.1007/s10552-018-1009-0
69. Smith BE (2015) Representing Appalachia: the impossible necessity of Appalachian studies. In: Berry C, Obermiller PJ, Scott SL (eds) Studying Appalachian studies: making the path by walking. University of Illinois Press, Urbana, IL
70. Yang S, Jin T, Su H-X et al (2015) The association between NQO1 Pro187Ser polymorphism and bladder cancer susceptibility: a meta-analysis of 15 studies. PLoS ONE 10(1):e0116500. https://doi.org/10.1371/journal.pone.0116500
71. Yu H, Liu H, Wang L-E, Wei Q (2012) A functional NQO1 609C > T polymorphism and risk of gastrointestinal cancers: a meta-analysis. PLoS ONE 7(1):e30566. https://doi.org/10.1371/journal.pone.0030566
72. Oh E-T, Park HJ (2015) Implications of NQO1 in cancer therapy. BMB Rep. 48(11):609–617
73. Atia A, Alrawaiq N, Abdullah A (2014) A review of NAD(P)H: quinone oxidoreductase 1 (NQO1); a multifunctional antioxidant enzyme. J Appl Pharm Sci. 4(12):118–122. https://doi.org/10.7324/JAPS.2014.41220
74. Peng Q, Lu Y, Lao X et al (2014) The NQO1 Pro187Ser polymorphism and breast cancer susceptibility: evidence from an updated meta-analysis. Diagn Pathol. 9(1):100. https://doi.org/10.1186/1746-1596-9-100
75. Asher G, Lotem J, Sachs L, Kahana C, Shaul Y (2002) Mdm-2 and ubiquitin-independent p53 proteasomal degradation regulated by NQO1. Proc Natl Acad Sci USA 99(20):13125–13130. https://doi.org/10.1073/pnas.202480499
76. Asher G, Lotem J, Kama R, Sachs L, Shaul Y (2002) NQO1 stabilizes p53 through a distinct pathway. Proc Natl Acad Sci USA 99(5):3099–3104. https://doi.org/10.1073/pnas.052706799
77. Vrijheid M (2014) The exposome: a new paradigm to study the impact of environment on health. Thorax 69(9):876–878. https://doi.org/10.1136/thoraxjnl-2013-204949
78. Wild CP, Scalbert A, Herceg Z (2013) Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. Environ Mol Mutagen 54(7):480–499. https://doi.org/10.1002/em.21777
79. Hu X, Zhang Z, Ma D, Huettner PC, Massad LS, Nguyen L, Borecki I, Rader JS (2010) TP53, MDM2, NQO1, and susceptibility to cervical cancer. Cancer Epidemiol Biomark Prev 19(3):755–761. https://doi.org/10.1158/1055-9965.EPI-09-0886 (**Epub 2010 Mar 3**)
80. Radloff LS (1977) The CES-D scale. Appl Psychol Meas 1(3):385–401. https://doi.org/10.1177/014662167700100306

Affiliations

Thomas J. Knobloch[1,11] · Juan Peng[2] · Erinn M. Hade[2] · David E. Cohn[3,11] · Mack T. Ruffin IV[4] · Michael A. Schiano[5,6] · Byron C. Calhoun[5,6] · William C. McBee Jr.[7] · Jamie L. Lesnock[7] · Holly H. Gallion[8] · Jondavid Pollock[9] · Bo Lu[1] · Steve Oghumu[1,11] · Zhaoxia Zhang[1] · Marta T. Sears[1] · Blessing E. Ogbemudia[1] · Joseph T. Perrault[10] · Logan C. Weghorst[1] · Erin Strawser[1] · Cecilia R. DeGraffinreid[10] · Electra D. Paskett[1,10,11] · Christopher M. Weghorst[1,11]

1 College of Public Health, The Ohio State University, Columbus, OH 43210, USA

2 Department of Biomedical Informatics, Center for Biostatistics, College of Medicine, The Ohio State University, Columbus, OH 43210, USA

3 Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Wexner Medical Center, College of Medicine, The Ohio State University Columbus, Columbus, OH 43210, USA

4 Department of Family and Community Medicine, Milton S. Hershey Medical Center, Penn State University, Hersey, PA 17033, USA

5 Department of Obstetrics & Gynecology, West Virginia University, Charleston, WV 26505, USA

6 Charleston Area Medical Center Health System, Charleston, WV 25302, USA

7 Mon General Health System, Morgantown, WV 26505, USA

8 Pikeville Medical Center, Pikeville, KY 41501, USA

9 Wheeling Hospital, Schiffler Cancer Center, Wheeling, WV 26003, USA

10 Division of Cancer Prevention and Control, Wexner Medical Center, College of Medicine, The Ohio State University Columbus, Columbus, OH 43210, USA

11 The Ohio State University Comprehensive Cancer Center, Columbus, OH 43210, USA