January 2016

Smoking, Genetic Polymorphisms In Dna Repair Genes And Risk Of Thyroid Cancer

Qi Guo
Yale University, qi.guo@yale.edu

Smoking, Genetic Polymorphisms in DNA Repair Genes and Risk of Thyroid Cancer

Qi Guo

YSPH Graduation Thesis for M.P.H. degree

Thesis Adviser: Yawei Zhang, M.D., PhD., MPH.

Second Reader: Ying Chen, M.D., PhD.

Abstract

While the number of current smokers has declined due to legal and regulatory measures and public health awareness, cigarette smoking remains the top risk factor for many diseases, including cancers originated from various tissues. Meanwhile, thyroid cancer incidence has increased significantly in the USA. Former epidemiology studies have investigated the association between smoking and risk of thyroid cancer, and interestingly results from these studies have shown that smoking has protective effect on risk of thyroid cancer. Studies have tapped into genetic polymorphisms and their association with thyroid cancer (and its subtypes) to further examine the susceptibility of certain gene variants carriers, and these studies have covered multiple SNPs of genes, which belong to relevant pathways including xenobiotic metabolism, DNA damage response and repair. Neither did the results from these studies form a consistent opinion about which SNPs/genes/pathways are associated with thyroid cancer among specific population, nor did they give confirmed explanation about molecular mechanisms supporting their results. In the current study, we investigated the association between smoking and thyroid cancer with stratification for 299 SNPs of 340 DNA repair genes. Study subjects included 440 cases diagnosed between 2010 and 2011 from Yale Cancer Center's Rapid Case Ascertainment Shared Resource, and 465 pairing controls in Connecticut. The study was conducted to test the hypothesis that genetic variations in DNA repair genes modify the association between smoking and risk of thyroid cancer. *RPA1* (replication protein A1 gene) rs11867830 have shown statistically significant effect modification for smoking-papillary thyroid cancer association. *ALKBH3* (alkB homolog 3, alpha-ketoglutarate-dependent dioxygenase gene) rs10768995 and rs197371 significantly modified such association for follicular thyroid cancer subtypes. *MAD2L2* (MAD2 mitotic arrest deficient-like 2 gene) rs747863, *RAD54L* (RAD54-like (S. cerevisiae) gene) rs10789488, *TDG* (thymine DNA glycosylase gene) rs2700505, rs322106, and rs322107 are demonstrated to have significant effect modification for overall cases of thyroid cancer.

Key Words

Genetics; Smoking; Thyroid cancer; DNA repair pathway

Introduction

Epidemiology of Thyroid Cancer

Thyroid cancer is the development of malignant tumor of the gland in front of the neck that normally produces thyroid hormones which are important to the normal regulation of the metabolism of the body[1]. The cancer often presents as a nodule, and while many of thyroid gland nodules are benign, the overall incidence of cancer in a nodule is 12%-15%. Main types of thyroid cancer include papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), anaplastic thyroid cancer (ATC), and para-follicular C-cell-derived medullary thyroid carcinoma (MTC). The prevalence of these subtypes among thyroid cancer are approximately 80%, 10%, 3%, and 4%[2]. According to the Surveillance, Epidemiology and End Results Program of National Cancer Institute[3], estimated number of new cases of thyroid cancer in 2015 is 62,450 (47,230 in women, and 15,220 in men), which accounts for 3.8% of all new cancer cases, and the number of new deaths associated with thyroid cancer in 2015 is estimated to be 1,950 (1,080 women and 870 men), accounting for 0.3% of all cancer deaths. Thyroid cancer is ranked as the 8th most common type of cancer in the general population following breast cancer, lung and bronchus cancer, prostate cancer, colon and rectum cancer, bladder cancer, melanoma, and non-Hodgkin lymphoma. It is most common in women than men, and is frequently diagnosed among people aged from 45 to 54.

Established risk factors for thyroid cancer includes radiation exposure[4-6], family history of thyroid disease, female gender and Asian race[7]. Many epidemiology studies also investigated the association of thyroid cancer with cigarette. While it is generally believed that thyroid cancer is not a smoking related neoplasm, several studies results indicated a negative association[8]. Mack et al.[9] conducted a pooled analysis of 14 case control studies including 2725 thyroid cancer cases and 4776 controls in USA, Europe and Asia. Conditional logistic

regression with stratification shows that thyroid cancer risk was reduced in current smokers (OR = 0.6, 95% CI = 0.6–0.7) and former smokers (OR = 0.9, 95% CI = 0.8–1.1). A population-based case control study , Rossing et al.[10] ,also showed that a history of ever having smoked more than 100 cigarettes was associated with a reduced risk of PTC (OR = 0.7, 95% CI = 0.5–0.9); such reduction was most evident in current smokers (OR = 0.5, 95% CI = 0.4–0.7). Galanti et al.[11] showed that women who started smoking before the age of 15 experienced a markedly reduced in risk (OR, 0.38%; 95% CI, 0.18-0.80). Researchers have given explanations for the negative association. Smoking might decrease the thyroid-stimulating hormone (TSH) level which lead to a decreased growth rate of thyroid epithelial cell[12]. A second possible explanation is that smoking decreases the risk of thyroid cancer by decreasing smokers' body weight compared to non-smokers. Some researchers also hypothesized that smoking is related to the decreased level of estrogen, since the thyroid cancer has a higher prevalence among female than male. However, these researches have not yet given clear sufficient illustrations in terms of the microscopic and genetic mechanism of the negative association.

Genetics of Thyroid Cancer and Mechanism of DNA Damage in Thyroid Cells

Different gene mutations are found and investigated for different subtypes of thyroid cancer. Multiple endocrine neoplasia type 2 (MEN2), as one kind of hereditary syndromes, are caused by a mutation in the *RET* proto-oncogene, and could be categorized into three subtypes (MEN2A, familial medullary thyroid carcinoma (FMTC), and MEN2B) which all impart high risk of developing medullary thyroid cancer. The BRAF gene mutation has been much investigated for its association with papillary thyroid carcinoma. Yoram et al.[14] applied polymerase chain reaction (PCR)-restriction enzyme analysis of BRAF exon 15 and among the 35 papillary thyroid carcinomas examined, 24 (69%) were found to carry a thymine-adenine transversion at nucleotide 1796 in the BRAF gene (T1796A). Nikiforova et al.[15] reported that BRAF mutation is identified not only in papillary carcinomas (38%) but also in poorly differentiated carcinomas (13%) and anaplastic carcinomas (10%), but none cases were identified in follicular, Hurthele cell and medullary carcinomas. Other classical oncogenic genetic alterations commonly seen in thyroid cancer include RAS and RET/PTC rearrangement PPARγ fusion oncogene.

Similar to other sites, DNAs in thyroid gland also face both endogenous and exogenous attacks. Endogenous damages mainly are damages from Reactive Oxygen Species (ROS), including peroxides, superoxides, hydroxyl radicals and singlet oxygens. According to K Krohn et al.[16], tobacco smoking would possibly lead to exacerbate oxidative stress in thyrocytes and increase mutagenesis by thiocyanate-ion-induced blockage of the iodine transport. Bio-synthesis of TSH requires high concentrations of hydro peroxides and oxidized iodine. While hydro peroxides give rise to highly reactive hydroxyl radicals that would cause damages including oxidized bases and abasic sites, decreased concentration of iodine actually stimulated generation of excess amount of H_2O_2. Low iodine levels and markers of increased thyroid functionality suggest activation of H_2O_2 generation, which could result in DNA damage and somatic mutation.

Exogenous damages include damages from aromatic substances attacks, which would possibly lead to DNA intercalation, the insertion of specific molecules (ligands) between the planar bases of DNA. Typical ligands include polycyclic aromatics, which are contained in smoke and tar of cigarettes. However, the process of DNA intercalation with toxic substances are mostly likely to happen at the sites of lung cells, and metabolism process of tobacco smoking chemicals shows little evidence that those DNA adducts could happen at thyrocytes.

Smoking related DNA Repair in Thyroid Cancer

Many studies have investigated the association between a biomarker of oxidative stress to DNA repair, urinary 8-hydroxydeoxyguanosine (8-OHdG), and smoking [17-20]. However, although in pathological changed thyrocytes there possibly exists excess oxidative stress that would lead to DNA impairment and production of 8-OHdG, these studies do not dig into relationship between these interaction and thyroid cancer. The reason is that smoking is not yet a considered and clearly understood as a thyroid cancer related risk factor, so these studies do not look into genotyped thyroid cancer patients and the specifically smoking induced repairing mechanisms. An important perspective worth looking into is the biological metabolism of carcinogen contained in smoking and the sites that these chemicals would actually have an effect on DNAs.

Genetic Polymorphism & GXE interaction

While genetic susceptibility researches focus on the association between gene variations and risk of thyroid cancer, gene-environmental researches instead investigate the effect modification on the association between smoking and risk of thyroid cancer, which refers to the testing of whether carrying wild-type homozygous or variant genotypes would result in significantly different associations of such. Despite this difference, genetic susceptibility researches might offers evidence to support interpretations of effect modification. When genetically stratified associations are compared with the raw association between the exposure and outcomes, such comparison would give light to investigation on the genetic susceptibility among specific subgroup of population, for example in this case, smokers.

There have been studies which investigated genetic polymorphisms and risks of thyroid cancer subtypes. In some studies, ionize radiation is acknowledged as a recognized risk factor for thyroid cancer and researchers have tested the hypothesis that polymorphisms in genes, which are potentially associated with ionize radiative attacks from psychological perspectives, are associated with risks of thyroid cancer. Genes of such kind could belong to different pathways (including DNA damage response[21], homologous recombination[22], xenobiotic metabolism[23], and DNA base excision repair[24] etc.), and they were separately studied. As for smoking, Bufalo et al.[25] put xenobiotic metabolism related genes (*CYP* and *GST*) and smoking equally into one multiple logistic regression model as risk factors to differentiated thyroid cancer subtypes and found inverse correlation between cigarette smoking (P=0.0385) and *CYP1A1m1* germline inheritance (P=0.0237) with risks of papillary carcinomas. Similarly, for xenobiotic metabolism related genes, Lemos[26] et al. simply apply the Pearson's chi-square test to examine the differences of *CYP2D6* poor metabolizer genotype and allele between papillary thyroid cancer patients and controls and found a lower frequency of homozygous poor metabolizer genotype carriers (1.6% vs. 5.5%, P=0.037, OR=0.28 (95% CI: (0.09, 0.93))) and *CYP2D6*4* allele carriers (13.4% vs. 21.7%, P=0.002, OR=0.56

(95% CI: (0.39, 0.80))) in the patient group, which implicates the protective effect of the *CYP2D6* poor metabolizer genotype possibly because of the decrease amount of metabolized chemical intermediates.

Different from pre-assuming and selecting specific genes, Neta et al. [27] extended the number and scope of the candidate genes belonging to multiple pathways related to genomic integrity. The study progressively examined the association of SNPs, gene regions, and genomic integrity pathways with risk of papillary thyroid cancer, and found that while after statistical multiple comparisons adjustment, SNPs, gene regions and overall DNA repair pathways do not significantly associate with risks of thyroid cancer, direct reversal of DNA damage and other conserved damage response genes sub-pathways demonstrate significant association. As a gene that belongs to direct reversal of DNA damage pathway, *MGMT* SNPs were reported to be significantly associated with papillary thyroid cancer with a p-value<0.0005 in Neta et al[27] . While no study has reported associations between *MGMT* polymorphisms and thyroid cancer, some studies have been conducted on its polymorphisms and other carcinomas. Zhang et al.[28] reported that although there were no significant associations between candidate *MGMT* SNPs (rs1711646, rs1625649, rs1803965, rs12917) and susceptibility to squamous cell carcinoma of the head and neck (SCCHN) when studied individually, statistically significant risks were found among carriers of 3 or 4 combined genotypes16195 CC, 16286CC, 45996 GT+TT, and 46346 CA+AA genotypes compared to those carrying less than 2 (OR=1.27, 95% CI: (1.11, 2.96)). Ma et al.[29] reported protective effects of several SNPs variants on *MGMT* on esophageal squamous cell carcinoma (ESCC) carcinogenesis and progression. Akbari et al.[30] performed a similar study in a SCCHN high risk region in Iran and found the G>A substitution in *MGMT* rs7087131 variant was associated with decreased risk of ESCC in Turkmen (OR=0.26, *P*=0.01). The function of *MGMT* as a gene encoding a protein which catalyzes transfer of methyl groups from O(6)-alkylguanine and other methylated moieties of the DNA to its own molecule for gene protection[31], and the fact that hypermethylation of multiple genes and genes promoter has been reported to be significantly associated with papillary thyroid cancer[32,33], as well as the methylation effect of PAHs (Polycyclic aromatic hydrocarbons) from smoking[34,35,36,37], might help explain *MGMT*'s effect modification in our study. Moreover, the methylation of *MGMT* itself might

inhibit its direct reversal repairing function[32,33]. However, the SNPs on *MGMT* in our studies which significantly modify the smoking-thyroid cancer association have not yet been reported as risk factors in genetic polymorphism researches, and no former studies have investigated the gene environment interaction for thyroid cancer.

Materials and methods

Population selection

Cases were histologically confirmed, incident thyroid cancer patients [papillary (ICD-O-3: 8050, 8052, 8130, 8260, 8340–8344, 8450, and 8452), follicular (ICD-O-3: 8290, 8330–8332, and 8335), medullary (ICD-O-3: 8345, 8346, and 8510), or anaplastic (ICD-O-3: 8021)] from Connecticut diagnosed between 2010 and 2011. Eligible patients were aged between 21 and 84 years at diagnosis, had no previous diagnosis of cancer, with the exception of nonmelanoma skin cancer, and were alive at the time of interview. Cases were identified through the Yale Cancer Center's Rapid Case Ascertainment Shared Resource (RCA). The RCA acts as an agent of the Connecticut Tumor Registry. The Connecticut Public Health Code requires reporting of all cancer cases in licensed hospitals and clinical laboratories to the Connecticut Tumor Registry. RCA field staff are assigned geographically to survey all of the state's nonpediatric hospitals to identify newly diagnosed cases. Information on the cases identified in the field is sent regularly to the RCA data-entry staff, who enter, verify, and screen these data against the Connecticut Tumor Registry database. The Connecticut Tumor Registry has reciprocal reporting agreements with cancer registries in all adjacent states (and Florida) to identify residents from Connecticut with cancer diagnosed and/or treated in these states. Cases and controls were frequency-matched by age (± 5 years). Distributions of age, sex, and race were similar between the participants and nonparticipants for both cases and controls.

SNPs Selection and Genotyping

After interview 440 thyroid cancer cases and 465 controls donated samples of venipuncture whole blood. Peripheral blood leukocyte DNA was extracted using the Qiagen phenol-chloroform extraction kit according to standard manufacturer protocol. DNA was then genotyped using a custom-made Golden Gate Illumina assay. Genotyping data were successfully obtained for all 440 thyroid carcinoma cases (including 373 papillary thyroid carcinoma cases and 54 follicular subtype) and 465 controls using this protocol. The Golden Gate assay included analysis of 299 single-nucleotide polymorphisms (SNPs) in 340 gene regions involved in DNA repair, based on statistical significance previously demonstrated in the 2011 analysis of Neta et al. (Included in our analysis were all individual SNPs from Neta et al. that had demonstrated $p_{snp}<0.1$, as well as all additional SNPs associated with gene regions with $p_{gene}<0.1$.) Quality-control duplicate samples were also included in the genotyping platform. All duplicate samples yielded a concordance rate of $\geqslant$99%, with no statistically-significant difference observed between replicated samples. Hardy-Weinberg equilibrium (HWE) was assessed in controls for each SNP using a chi-square test. SNPs with a p-value >0.00001 from the chi-square test were considered to be in HWE.

All procedures were performed in accordance with a protocol approved by the Human Investigations Committee at Yale and the Connecticut Department of Public Health. After approval by the hospitals and by each participant's physician (cancer cases), or after selection through random sampling (control population), potential participants were approached by letter and then over the phone. Those who agreed were interviewed by trained study interviewers, either at their homes or at a convenient location.

With regard to smoking exposure, the participants were asked about the following smoking status and history: (a) current smoking status, (b) pack-years of smoking, (c) numbers of cigarettes smoked, (d) smoking duration and (e) age started smoking. The participants were asked whether they had ever smoked. If a participant answered 'yes', he/she was asked how old he/she started smoking, as well as the total number cigarette he/she smoke per year. 'Exposure' to smoking was defined as ever smoking behavior of the respondent, where 'non-

exposure' was defined as non-smoking behavior. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using unconditional logistic regression models to estimate the associations between smoking and the risk of thyroid cancer by histologic subtype and tumor size, and to control for potential confounders. Other confounding factors included would be: age, sex, and education level, family history of thyroid cancer, previous benign thyroid disease, BMI, alcohol consumption, and radiation treatment. Decisions on the covariates to be included in the final model were based on a greater than 10% change in the risk estimates. All analyses were carried out using SAS (version 9.4; SAS Institute Inc., Cary, North Carolina, USA).

Statistical Analysis

Unconditional logistic regression models were used to estimate the associations of smoking and genetic variation of DNA repair pathway genes with risk of thyroid cancer and its subtypes (papillary and follicular). With the assumption that the SNPs we studied were dominant mutation, we combined heterozygous and homozygous variant genotypes of all 299 genes extracted from the studied population to increase statistical power. Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated adjusting for age, gender, race, education, and BMI index.

We created and included the interaction terms of single nucleotide polymorphisms (SNPs) into the logistic regression model to see whether the relationship between smoking and thyroid cancer risk would be modified by genetic variation. We put the interaction term, along with allele genotype variable, smoking and other potential confounding factors into the regression model, and examine the p-value of the coefficient of the interaction term. We used an a priori significance level of 0.01 for each test rather than a Bonferroni correction because the Bonferroni correction is overly conservative when hypothesis tests are correlated. A significance level of 0.01 will increase the type I error rate to the point of certainly identifying some false positive findings. However, it will also reduce the number of false negative findings, and thus the actual alpha is likely to be substantially less than the nominal alpha is this case. If the allele genotype variable is found to have a significant p-value of less than 0.01, we then recognize the SNPs as having significant effect modification on the association

between smoking and thyroid cancer. For these genes we stratified the association between smoking and thyroid cancer by genotype (wild-type homozygous vs. heterozygous + homozygous variant), and the smoking would be identified as having significant association with thyroid cancer if the p-value of its coefficient in the unconditional logistic regression model would be less than 0.05.

Results

Of the 440 patients with thyroid cancer, the majority were diagnosed with papillary thyroid cancer (373, 84.77%), followed by follicular (54, 12.27%), medullary (11, 2.50%), anaplastic (1, 0.23%) and others (1, 0.23%). The average age at diagnosis of thyroid cancer is 51 years old. The patient group has a much higher proportion of female (80.91%) compared to the control group (68.82%). Cases are more likely coming from lower income families, more obese than controls, and have family history of thyroid cancer among first-degree relatives. Moreover, we observed that the proportion of prior benign thyroid disease is lower among cases, and they are less likely to smoke. The characteristics of the studied population are shown in Table 1.

Similar to former epidemiological studies, exposure to smoking has the trend to negatively affect the risk, but did not reach significance (OR=0.80, 95% CI: 0.59, 1.09; p-value: 0.15). When we analyzed the data by histological subtypes, similar association were observed for papillary and follicular thyroid cancer, while the protective association for papillary subtype (OR =0.87, 95% CI: 0.64-1.19; p-value=0.37) is weaker compared to follicular (OR=0.69, 95% CI: 0.36-1.31; p-value=0.25), however, these associations are not statistically significant.

Table 1. Characteristics of thyroid cancer cases and controls

	Cases(440) [N(%)]	**Controls(465)** [N(%)]	**p-value**
Age(years)			
Mean(SD)	50.94(12.14)	54.07(13.08)	0.0006
0-40	89(20.23%)	71(15.27%)	
41-50	121(27.50%)	126(27.10%)	
51-60	133(30.23%)	119(25.59%)	

61-70	75(17.05%)	92(19.98%)	
>70	22(5.00%)	57(12.26%)	
Sex			<0.0001
Male	84(19.09%)	145(31.18%)	
Female	356(80.91%)	320(68.82%)	
Race			0.3706
White	396(90.00%)	427(91.83%)	
Black	16(3.64%)	20(4.30%)	
Asians	7(1.59%)	5(1.08%)	
Others	21(4.77%)	13(2.80%)	
Family income			0.3016
<$25,000	27 (6.19%)	43 (9.35%)	
$25,000~$89,999	121 (27.75%)	113 (24.57%)	
>$90,000	155 (35.55%)	174 (37.83%)	
Confidential	114 (26.15%)	115 (25.00%)	
Not sure	19 (4.36%)	15 (3.26%)	
Missing=9			
Years of education			0.0021
Grades School	4(0.91%)	3(0.65%)	
High School	118(26.94%)	75(16.30%)	
Tech/Trade School	30(6.85%)	23(5.00%)	
College/Univeristy	176(40.18%)	226(49.13%)	
Grad & Prof School	97(22.15%)	120(26.09%)	
Others	13(2.97%)	13(2.83%)	
Missing=7			
Family history of thyroid cancer among first-degree relatives			0.0309
Yes	54(12.27%)	37(7.96%)	
No	386(87.73%)	428(92.04%)	
Prior benign thyroid disease			<0.001
Yes	56(12.73%)	453(97.42%)	
No	384(87.27%)	12(2.58%)	
BMI(kg/m2)			0.0003
0-25	143(32.72%)	192(42.01%)	
26-30	136(31.12%)	155(33.92%)	
>30	158(36.16%)	110(24.07%)	
Missing=11			
Smoking			0.2412
Never	303(69.18%)	302(65.51%)	
Ever	135(30.82%)	159(34.49%)	
Missing=6			
Subtypes			
Papillary	373(84.77%)		
Follicular	54(12.27%)		
Medullary	11(2.50%)		
Anaplastic	1(0.23%)		
Others	1(0.23%)		

[a] Benign thyroid disease included hyperthyroidism, hypothyroidism, goiter, thyroid nodules, and thyroid adenoma.
[b] Ever smoking was defined as ever smoked a total of 100 cigarettes or more.

Table 2. Association between smoking and thyroid cancer

	Overall			
Smoking Exposure[a]	Controls	Cases	OR (95% CI)	P-Value
Never	302	303		
Ever	159	135	0.80 (0.59-1.09)	0.15

[a] Ever smoking was defined as ever smoked a total of 100 cigarettes or more.

Table 3. Association between smoking and thyroid cancer subtypes

	Papillary				Follicular			
Smoking Exposure[a]	Controls	Cases	OR(95%CI)	p-value	Controls	Cases	OR(95%CI)	p-value
Never	346	256			556	39		
Ever	179	115	0.87 (0.64-1.19)	0.37	279	15	0.69 (0.36-1.31)	0.25

[a] Ever smoking was defined as ever smoked a total of 100 cigarettes or more.

Among all 299 DNA repair pathway SNPs studied, statistically significant association for genetic polymorphisms were observed in 5 for thyroid cancer overall, while 1 for papillary and 2 for follicular thyroid cancer. Results are presented in Table 4. These genes were identified first through effect modification analysis, and then through genotype stratification. Only genes with significant p-values in both effect modification analysis and at least one stratum of genotypes were considered as having gene-environment interaction.

5 SNPs in 3 gene regions showed significant gene-environment interaction for overall cases of thyroid cancer. Significant decreased risks of thyroid cancer overall were observed among smokers with wild-type homozygous genotype in 1 gene *MAD2L2* rs747863 CC (OR=0.62, 95% CI (0.43, 0.89), p=0.0095) compared with its heterozygous/homozygous variant TC/TT (OR=1.61, 95% CI (0.84, 3.08), p=0.1482). Significantly decreased risks of thyroid cancer overall were observed among smokers with heterozygous/homozygous variant genotypes (CC/TC) in *RAD54L* rs10789488 (OR=.., 95% CI (), p=), while its wild-homozygous smoking carriers have an

insignificant increased risk of having thyroid cancer (OR=.., 95% CI (), p=). 3 SNPs on the gene TDG (rs2700505, rs322107, rs322106) also demonstrated modifying effect.

Only 1 SNP in 1 gene region significantly modifies the association between smoking and the risk of papillary thyroid cancer. Significantly decreased risks of papillary thyroid cancer were observed among smokers carrying heterozygous/homozygous variant genotypes in *RPA1* rs11867830 TC/CC (OR=0.45, 95% CI: (0.25, 0.81), p-value=0.08), while the carriers of the wild-type homozygous genotype *RPA1* rs11867830 TT (OR=1.15, 95% CI: (0.78, 1.68), p-value=0.478) who smoke have an insignificantly increased risk, which differ from the overall pooled population of study.

2 SNPs in 2 gene regions have demonstrated effect modification for the follicular thyroid cancer subtype. Significant decreased risks of this subtype associated with smoking were observed among carriers of heterozygous/homozygous variant genotype of *RECQL* rs4762 GC/CC (OR=0.33, 95% CI: (0.13, 0.83), p-value=0.019) and *ALKBH3* rs1973717 TC/TT (OR=0.32, 95% CI: (0.13, 0.82), p-value=0.018), while the homozygous genotype carriers of these two genes have an insignificantly increased risk (*RECQL* rs4762 (OR=2.48, 95% CI: (0.80, 7.71), p-value=0.117; *ALKBH3* rs1973717 (OR=2.45, 95% CI (0.78, 7.65), p-value=0.124). Results are shown in Table 5 and 6.

Table 4. Stratified association between smoking and overall cases of thyroid cancer by SNPs

Single Nucleotide Polymorphism	Never Smoke			Ever Smoke		OR (95% CI) [a]	p-value	Interaction Term p-value [b]
	Case	Control	OR=1.00	Case	Control			
MAD2L2 C>T rs747863								
CC	222	212		89	121	0.62 (0.43, 0.89)	0.0095	0.01
TC/TT	80	90		45	38	1.61 (0.84, 3.08)	0.1482	
RAD54L T>C rs10789488								
TT	174	193		84	79	1.17 (0.78, 1.75)	0.4525	0.0021
CC/TC	128	109		50	80	0.45 (0.27, 0.75)	0.0022	
TDG A>G rs2700505								
AA	114	138		68	59	1.44 (0.89, 2.34)	0.1381	0.0076
AG/GG	186	164		67	100	0.54 (0.35, 0.81)	0.0032	
TDG G>A rs322107								
GG	200	212		107	106	1.06 (0.74, 1.52)	0.736	0.0072
AG/GG	102	90		28	53	0.31 (0.16, 0.62)	0.0009	

TDG A>G rs322106							
AA	201	212		107	106	1.06 (0.74, 1.51)	0.7593	0.0057
AG/GG	101	90		27	53	0.29 (0.14, 0.59)	0.0005	

Abbreviations: CI: confidence interval; OR: odds ratio.
[a] *P* value for the interaction between smoking and genotype variations in a logistic regression model, with adjustment for age, sex, race, education, family income, and BMI index
[b] Logistic regression model with adjustment for age, sex, race, education, family income, and BMI index

Table 5. Stratified association between smoking and papillary thyroid cancer by SNPs

Single Nucleotide Polymorphism	Never Smoke			Ever Smoke		OR (95% CI) [a]	p-value	Interaction Term p-Value [b]
	Case	Control	OR=1.00	Case	Control			
RPA1 T>C rs11867830								
TT	84	119		149	242	1.15 (0.78, 1.68)	0.4778	0.007
TC/CC	30	60		107	107	0.45 (0.25, 0.81)	0.0078	

Abbreviations: CI: confidence interval; OR: odds ratio.
[a] *P* value for the interaction between smoking and genotype variations in a logistic regression model, with adjustment for age, sex, race, education, family income, and BMI index
[b] Logistic regression model with adjustment for age, sex, race, education, family income, and BMI index

Table 6. Stratified association between smoking and follicular thyroid cancer by SNPs

Single Nucleotide Polymorphism	Never Smoke			Ever Smoke		OR (95% CI) [a]	p-value	Interaction Term p-value [b]
	Case	Control	OR=1.00	Case	Control			
ALKBH3 G>C rs10768995								
GG	107	9		218	8	2.48 (0.80, 7.71)	0.1168	0.0076
GC/CC	171	6		346	31	0.33 (0.13, 0.83)	0.0187	
ALKBH3 C>T rs1973717								
CC	112	9		229	8	2.45 (0.78, 7.65)	0.1243	0.0082
TC/TT	167	6		331	31	0.32 (0.13, 0.82)	0.0175	

Abbreviations: CI: confidence interval; OR: odds ratio.
[a] *P* value for the interaction between smoking and genotype variations in a logistic regression model, with adjustment for age, sex, race, education, family income, and BMI index
[b] Logistic regression model with adjustment for age, sex, race, education, family income, and BMI index

Interestingly, the significant SNPs identified for overall thyroid cancer cases did not overlap with any of those for papillary or follicular subtypes, and the identified SNPs for these 2 subtypes also are totally different.

Discussion

This study addressed potential modifying effect of SNPs in DNA repair genes on the association between smoking and the risk of differential and overall thyroid cancer. The results of our population-based case-control study are consistent with former epidemiological researches which report reduced risk of thyroid cancer among tobacco consumers, while the association is not statistically significant. By further looking into the genetic level,

we found that genetic variations of specific DNA repair genes could have modified the association between smoking and thyroid cancer, and it helped explain the individual susceptibility to thyroid cancer among smoking population. These effect modifications were observed among genes for multiple DNA repair pathways (shown in Table 7) instead of a single pathway (such as base excision repair, strand breaking repair etc.), and the discovered genes and their mutations are not shown to be associated with thyroid cancer in former studies. The underlying repairing mechanisms of these genes variants against tobacco exposure are not yet supported by sufficient scientific research and evidence and the outcomes in our study might only be significant from a statistical perspective rather than a genetic and psychological one.

Table 7. DNA repair genes with variant genotypes and affected susceptibility to thyroid cancer types

Gene symbol	Gene Full name	Repair pathway and function of encoded protein	Significant thyroid cancer types
RAD54L	RAD54-like (S.cerevisiae)	*Homologous recombination*; Helps homologous recombination related repair of DNA double-strand breaks.	Overall
TDG	thymine DNA glycosylase	*Base excision repair*; Helps remove thymine moieties from G/T mismatches on DNA single strand	Overall
MAD2L2	MAD2 mitotic arrest deficient-like 2 (yeast)	*DNA polymerase*; Makes up the mitotic spindle assembly checkpoint that prevents the onset of anaphase	Overall
RPA1	replication protein A1	*Nucleotide excision repair*; Binds and stabilizes single-stranded DNA, intermediates that from during DNA replication or upon DNA stress	Papillary only
ALKBH3	alkB homolog 3, alpha-ketoglutarate-dependent dioxygenase	*Direct reversal of damage*; Helps defend DNA against alkylating agents, similar to MGMT	Follicular only

Despite genetic susceptibility studies, there has neither been investigation into the DNA repair gene-smoking interaction and risk of thyroid cancer, nor any stratified association between smoking and thyroid cancer by SNPs. Former studies only identified radiation as a risk factor, but smoking as an unclear protective effect. The reason behind these counterintuitive outcomes may be that radiation exposes direct effects on thyroid and it generates carcinogenicity by causing DNA damages which the repairing mechanism might not be sufficient to eliminate. In contrast, while carcinogenic chemicals from smoking would possibly not circulate to thyroid cells, the carcinogens stimulate the systematic DNA repairing mechanisms of the human body, and indirectly decrease

the risk of getting thyroid cancer. Since there is no reference literature on this kind of gene-environment interaction, we could only refer to those genetic susceptibility studies for thyroid cancer to look for genes that we have identified in our research, and if there is not we tried to seek evidence in other types of cancer.

RAD54L

RAD54L (RAD54-like (S. cerevisiae)), encodes a protein which is involved in the homologous recombination and repair of DNA double-strand breaks by inducing a DNA topological changes[37, 38], which shares similarity with Saccharomyces cerevisiae Rad54. As pointed out by Sturgis et al.[39], RAD54 polymorphisms that result in amino acid change have been identified, but reported allele frequencies are less than 2%[40]. Therefore, there are not many studies investigating *RAD54L/RAD54* polymorphism and its relation with thyroid cancer, and the SNP *RAD54L* rs10789488 in our study has never been identified as risk factor of any carcinogenesis before. Carling et al.[41] examined the a well characterized set of parathyroid adenomas for inactivating mutations in the *RAD54* gene and fully sequenced all 18 exons, including the intron–exon junctions of the *RAD54* gene in 12 parathyroid adenomas with LOH at chromosome 1p, and were unable to detect any mutational aberrations or homozygous deletions in the gene. Li et al.[42] examined the effect of *RAD54L* C157T SNP on the survival time of pancreatic cancer, and reported that pancreatic cancer carriers of variant genotypes of *RAD54L* rs1048771 (CT/TT) have significantly shorter survival time, and the SNP in this study is an independent risk factor of survival. The authors pointed out that *RAD54L* rs1048771 is a functionally silent polymorphism so its effect on pancreatic cancer might due to its linkage to other SNPs on *RAD54L* or other genes. Similarly, after identifying a silent C/T transition at nucleotide 2290 in exon 18 on RAD54L, Leone et al.[43] reported that the frequency of the rare allele - T and heterozygotes for the 2290C/T polymorphism in the blood of Spanish meningioma patients and in the Ecuadorian meningioma tumors was significantly higher, but there is lack of evidence that the SNP has any plausible effect on mRNA transcription. Debora et al[44] also suggested RAD54 as a candidate suppressor of breast tumor.

ALKBH3

ALKBH3, as another gene related to direct reversal of damage which also demonstrate its modifying effect for several of its SNPs on smoking and thyroid cancer association, little is known about the relationship between its polymorphism and cancer. In the genetic susceptibility study of Neta et al.[27], SNP *ALKBH3* rs10838192 T>C is significantly associated with papillary thyroid cancer. Carriers of the heterozygous TC and homozygous variant CC genotypes have both respectfully higher risks of having papillary thyroid cancer (TC: OR=1.74(1.25-2.41); CC: OR=2.33(1.05-5.17). However, in this study, neither *ALKBH3* rs10768995 nor rs1973717, which have both shown significant effect modification in our study for follicular cancer, is not significantly associated with development with papillary thyroid cancer (*p*-values > 0.01). Despite the different cancer subtypes, this situation could be explained by that even these SNPs are not directly related to thyroid cancer, they can still modify the association between a risk factor and the carcinoma.

RPA1

RPA1 is a gene region which helps maintain integrity in DNA replication process by binding and stabilization of single-stranded DNA. *Cipollini* et al.[45] reported that RPA1 rs1131636 T>C polymorphism is very unlikely associated with the risk of differentiated thyroid cancer. In Neta et al.[27] it shows a similar result that the examined RPA1 SNPs (including the rs11867830 T>C which has shown significant effect modification for papillary cancer) are not significantly associated with papillary cancer subtype. Michiels et al.[46] found an increasing risk for Head and Neck cancer development for RPA1 haplotype 2 carriers (OR=1.60, 95% CI: 1.02-2.47). Han et al.[47] reported the SNP-SNP interaction and its association with breast cancer and found that the presence of *RPA1* rs11078676 (AG) together with *ERCC* rs50872 was strongly associated with increased risk of breast cancer (OR=2.70, 95%CI: 1.61-4.52).

TDG

TDG, encoded a protein that belongs to the TDG/mug DNA glycosylase family. Thymine-DNA glycosylase (TDG) removes thymine moieties from G/T mismatches by hydrolyzing the carbon-nitrogen bond between the sugar-phosphate backbone of DNA and the mispaired thymine, thymine from C/T and T/T mispairings when at lower activity, and also uracil and 5-bromouracil from mispairings with guanine. This enzyme plays a central role in cellular defense against genetic mutation caused by the spontaneous deamination of 5-methylcytosine and cytosine. Krzesniak ´ et al.[48] reported that the non-silent polymorphism of TDG (199: Ser. 367: Met) did not show any significant association with lung cancer risk.

MAD2L2

MAD2L2 (MAD2 mitotic arrest deficient-like 2) encodes a DNA polymerase which makes up the mitotic spindle assembly checkpoint that prevents the onset of anaphase. Suga et al.[49] screened 999 SNPs in 137 candidate genes on 399 subjects to have found that carriers of haplotypes CG in MAD2L2 would have decreased risk of early skin reactions for irradiated breast. However, none of former studies have reported MAD2L2 rs747863 as significant disease risk indicators.

In fact, while the gene regions might explain such effect modification to some extent, it could not explain the changes brought by multiple SNPs on a gene region. Interestingly, among the significant SNPs identified in our study, *MAD2L2* rs747863, *RAD54L* 10789488, *RPA1* rs11867830, *ALKBH3* rs10768995 and rs1973717 are located in intro regions of the genes, which means they would not be used for mRNA translation or furtherly result in functional amino acid changes. *TDG* rs2700505 is located within an exon but belong to the 5' UTR (5 prime untranslated region) and would not be translated. TDG rs322106 and rs322107 both are categorized as upstream variant 2KB, and it is very likely that they would not induce any functional changes neither since they refer to changes within 1000 bp of the transcript start site on the 5' side, and they are not included into the

mRNA transcription. However, it has now been acknowledged that many important mutations are located in the intron regions, and they might have indirect impacts on splicing and mRNA transcription. While we could not exclude the possibility that the interaction between these SNPs and other functional gene changes would explain such gene-environment interaction, genetic susceptibility studies have already shown lack of evidence that these SNPs would bring out substantial changes. Therefore, the results shown in our research remain to be examined and further researches are still needed.

Acknowledgement

My biggest thanks go to Dr. Yawei Zhang, my thesis advisor, who has given me her sincere guidance, constructive feedbacks and maximum tolerance upon this project. I also earnestly express my thankfulness to her research assistant, Huang Huang, MPH, who has patiently answered my endless questions and has given me hands-on tutorship in data analysis and interpretation. I also thank the second reader of my thesis, Dr. Ying Chen, who has given me so much support in helping me understand these amazing epidemiological concepts and have kindly shared with me some fantastic former researches for reference.

References

[1] "Definition of Thyroid Cancer." *MedicineNet.com.* September 20, 2012. November 10, 2015.

[2] "What is Thyroid Cancer?" *American Cancer Society*. March 17, 2015. November 10, 2015.

[3] Howlader, N., et al. "SEER Cancer Statistics Review (CSR) 1975–2012."Bethesda, MD: National Cancer Institute (2015).

[4] Pacini F, Vorontsova T, Molinaro E, et al.: Prevalence of thyroid autoantibodies in children and adolescents from Belarus exposed to the Chernobyl radioactive fallout. Lancet 352 (9130): 763-6, 1998.

[5] Cardis E, Kesminiene A, Ivanov V, et al.: Risk of thyroid cancer after exposure to ^{131}I in childhood. J Natl Cancer Inst 97 (10): 724-32, 2005.

[6] Tronko MD, Howe GR, Bogdanova TI, et al.: A cohort study of thyroid cancer and other thyroid diseases after the chornobyl accident: thyroid cancer in Ukraine detected during first screening. J Natl Cancer Inst 98 (13): 897-903, 2006.

[7] Iribarren C, Haselkorn T, Tekawa IS, et al.: Cohort study of thyroid cancer in a San Francisco Bay area population. Int J Cancer 93 (5): 745-50, 2001.

[8] Mack WJ, Preston-Martin S (1998) Epidemiology of thyroid cancer. In: Fagin JA, ed. Thyroid Cancer. Boston: Kluwer Academic Publishers, pp. 1–25.

[9] Mack, Wendy J., et al. "A pooled analysis of case–control studies of thyroid cancer: cigarette smoking and consumption of alcohol, coffee, and tea." *Cancer Causes & Control* 14.8 (2003): 773-785.

[10] Rossing, Mary Anne, et al. "Risk of papillary thyroid cancer in women in relation to smoking and alcohol consumption." *Epidemiology* 11.1 (2000): 49-54.

[11] Galanti, Maria Rosaria, et al. "Reproductive history and cigarette smoking as risk factors for thyroid cancer in women: a population-based case-control study." *Cancer Epidemiology Biomarkers & Prevention* 5.6 (1996): 425-431.

[12] Bertelsen, Jette B., and LASZLO HEGEDÜS. "Cigarette smoking and the thyroid." *Thyroid* 4.3 (1994): 327-331.

[13] Ron E (1996) Thyroid Cancer. In: Schottenfeld D, Fraumeni JF, eds. Cancer Epidemiology and Prevention. New York: Oxford University Press, pp. 1000–1021

[14] Cohen, Yoram, et al. "BRAF mutation in papillary thyroid carcinoma." *Journal of the National Cancer Institute* 95.8 (2003): 625-627.

[15] Nikiforova, Marina N., et al. "BRAF mutations in thyroid tumors are restricted to papillary carcinomas and anaplastic or poorly differentiated carcinomas arising from papillary carcinomas." *The Journal of Clinical Endocrinology & Metabolism*88.11 (2003): 5399-5404.

[16] Krohn, Knut, Jacqueline Maier, and Ralf Paschke. "Mechanisms of disease: hydrogen peroxide, DNA damage and mutagenesis in the development of thyroid tumors." *Nature Clinical Practice Endocrinology & Metabolism* 3.10 (2007): 713-720.

[17] Wu, Lily L., et al. "Urinary 8-OHdG: a marker of oxidative stress to DNA and a risk factor for cancer, atherosclerosis and diabetics." *Clinica Chimica Acta* 339.1 (2004): 1-9.

[18] Shen, Han-Ming, et al. "Detection of oxidative DNA damage in human sperm and the association with cigarette smoking." *Reproductive Toxicology* 11.5 (1997): 675-680.

[19] Leanderson, Per, and Christer Tagesson. "Cigarette smoke-induced DNA-damage: role of hydroquinone and catechol in the formation of the oxidative DNA-adduct, 8-hydroxydeoxyguanosine." *Chemico-biological interactions* 75.1 (1990): 71-81.

[20] Erhola, Marina, et al. "Biomarker evidence of DNA oxidation in lung cancer patients: association of urinary 8-hydroxy-2'-deoxyguanosine excretion with radiotherapy, chemotherapy, and response to treatment." *FEBS letters* 409.2 (1997): 287-291.

[21] Akulevich, Natallia M., et al. "Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma."*Endocrine-related cancer* 16.2 (2009): 491-503.

[22] Bastos, Helder Novais, et al. "Association of polymorphisms in genes of the homologous recombination DNA repair pathway and thyroid cancer risk."*Thyroid* 19.10 (2009): 1067-1075.

[23] Gaspar, Jorge, et al. "Combined effects of glutathione S-transferase polymorphisms and thyroid cancer risk." *Cancer genetics and cytogenetics*151.1 (2004): 60-67.

[24] Chiang, Feng-Yu, et al. "Association between polymorphisms in DNA base excision repair genes XRCC1, APE1, and ADPRT and differentiated thyroid carcinoma." *Clinical Cancer Research* 14.18 (2008): 5919-5924.

[25] Bufalo, Natassia E., et al. "Smoking and susceptibility to thyroid cancer: an inverse association with CYP1A1 allelic variants." *Endocrine-Related Cancer*13.4 (2006): 1185-1193.

[26] Lemos, Manuel C., et al. "Genetic polymorphism of CYP2D6 influences susceptibility to papillary thyroid cancer." *Clinical endocrinology* 67.2 (2007): 180-183.

[27] Neta, Gila, et al. "Common genetic variants related to genomic integrity and risk of papillary thyroid cancer." *Carcinogenesis* 32.8 (2011): 1231-1237.

[28] Zhang, Zhengdong, et al. "Polymorphisms of the DNA repair gene MGMT and risk and progression of head and neck cancer." *DNA repair* 9.5 (2010): 558-566.

[29] Ma, Wen-Jing, et al. "DNA polymorphism and risk of esophageal squamous cell carcinoma in a population of North Xinjiang, China." *World J Gastroenterol* 16.5 (2010): 641-647.

[30] Akbari, Mohammad R., et al. "Candidate gene association study of esophageal squamous cell carcinoma in a high-risk region in Iran." *Cancer research* 69.20 (2009): 7994-8000.

[31] "MGMT O-6-methylguanine-DNA Methyltransferase [Homo Sapiens (human)]." *National Center for Biotechnology Information*. N.p., 29 Mar. 2016. Web. <http://www.ncbi.nlm.nih.gov/gene/4255>.

[32] Schagdarsurengin, Undraga, et al. "CpG island methylation of tumor-related promoters occurs preferentially in undifferentiated carcinoma." *Thyroid* 16.7 (2006): 633-642.

[33] Ishida, Eiwa, et al. "DNA hypermethylation status of multiple genes in papillary thyroid carcinomas." *Pathobiology* 74.6 (2007): 344-352.

[34] Pfeifer, Gerd P., et al. "Tobacco smoke carcinogens, DNA damage and p 53 mutations in smoking-associated cancers." *Oncogene* 21.48 (2002): 7435-7451.

[35] Nock, Nora L., et al. "Associations between smoking, polymorphisms in polycyclic aromatic hydrocarbon (PAH) metabolism and conjugation genes and PAH-DNA adducts in prostate tumors differ by race." *Cancer Epidemiology Biomarkers & Prevention* 16.6 (2007): 1236-1245.

[36] Pfeifer, Gerd P., et al. "Tobacco smoke carcinogens, DNA damage and p 53 mutations in smoking-associated cancers." *Oncogene* 21.48 (2002): 7435-7451.

[37] "Genes and Mapped Phenotypes." *National Center for Biotechnology Information*. U.S. National Library of Medicine. Web. 13 Apr. 2016.

[38] Karran, Peter. "DNA double strand break repair in mammalian cells." *Current opinion in genetics & development* 10.2 (2000): 144-150.

[39] Sturgis, Erich M., et al. "Radiation Response Genotype and Risk of Differentiated Thyroid Cancer: A Case-Control Analysis." *The Laryngoscope*115.6 (2005): 938-945.

[40] Mohrenweiser, Harvey W., et al. "Identification of 127amino acid substitution variants in screening 37 DNA repair genes in humans." *Cancer Epidemiology Biomarkers & Prevention* 11.10 (2002): 1054-1064.

[41] Carling, Tobias, et al. "Analysis of the RAD54 gene on chromosome 1p as a potential tumor-suppressor gene in parathyroid adenomas." *International journal of cancer* 83.1 (1999): 80-82.

[42] Li, Donghui, et al. "Single nucleotide polymorphisms of RecQ1, RAD54L, and ATM genes are associated with reduced survival of pancreatic cancer."*Journal of clinical oncology* 24.11 (2006): 1720-1728.

[43] Leone, Paola E., et al. "Implications of a RAD54L polymorphism (2290C/T) in human meningiomas as a risk factor and/or a genetic marker." *BMC cancer*3.1 (2003): 1.

[44] Rasio, Debora, et al. "Characterization of the human homologue of RAD54: a gene located on chromosome 1p32 at a region of high loss of heterozygosity in breast tumors." *Cancer research* 57.12 (1997): 2378-2383.

[45] Cipollini, Monica, et al. "Polymorphisms within base and nucleotide excision repair pathways and risk of differentiated thyroid carcinoma." *DNA Repair* 41 (2016): 27-31.

[46] Michiels, Stefan, et al. "Polymorphism discovery in 62 DNA repair genes and haplotype associations with risks for lung and head and neck cancers."*Carcinogenesis* 28.8 (2007): 1731-1739.

[47]Han, Wonshik, et al. "SNP–SNP interactions between DNA repair genes were associated with breast cancer risk in a Korean population." *Cancer*118.3 (2012): 594-602.

[48] Krześniak, M., et al. "Polymorphisms in TDG and MGMT genes–epidemiological and functional study in lung cancer patients from Poland." Annals of human genetics 68.4 (2004): 300-312.

[49]Suga, Tomo, et al. "Haplotype-based analysis of genes associated with risk of adverse skin reactions after radiotherapy in breast cancer patients."*International Journal of Radiation Oncology* Biology* Physics* 69.3 (2007): 685-693.