JOINT EFFECTS OF ALCOHOL CONSUMPTION AND POLYMORPHISMS IN ALCOHOL AND OXIDATIVE STRESS METABOLISM GENES ON RISK AND SURVIVAL FROM HEAD AND NECK CANCER

Anne M. Hakenewerth

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Public Health (Epidemiology)

Chapel Hill
2011

Approved by:

Andrew Olshan, PhD (Chair)

Amy Herring, ScD

Robert Millikan, DVM, PhD

Kari North, PhD

Ivan Rusyn, MD, PhD

ABSTRACT

ANNE M. HAKENEWERTH: Joint effects of alcohol consumption and polymorphisms in alcohol and oxidative stress metabolism genes on risk and survival from head and neck cancer
(Under the direction of Dr. Andrew Olshan)

Heavy drinking of alcoholic beverages increases risk of developing squamous cell carcinoma of the head and neck (SCCHN; oral, pharyngeal and laryngeal cancer). This study hypothesized that genetic variation in the ethanol metabolism and oxidative stress pathways may influence the occurrence of and survival from SCCHN.

Interview and genotyping data were obtained from 1227 SCCHN cases and 1325 controls from the Carolina Head and Neck Cancer Epidemiology study, a population-based case-control study of head and neck cancer conducted in North Carolina from 2002-2006. Vital status through 2008 was obtained from the National Death Index. A panel of 45 single nucleotide polymorphisms (SNPs) in the ethanol and 19 SNPs in the oxidative stress metabolism pathways were evaluated in relation to the risk of SCCHN using logistic regression and in relation to all-cause and cancer-specific survival using Cox regression. Bonferroni- corrected p-values were also estimated.

Two SNPs were associated with SCCHN risk: *ADH1B* rs1229984 A allele [odds ratio (OR)=0.7, 95% confidence interval (CI)=0.6-0.9)] and *ALDH2* rs2238151 C allele (OR=1.2, 95% CI=1.1-1.4). Three SNPs were associated with cancer risk in anatomic sub-sites: *ADH1B* rs17028834 C allele (larynx, OR=1.5, 95% CI=1.1-2.0), *SOD2* rs4342445 A allele (oral cavity, OR=1.3, 95% CI=1.1-1.6), and *SOD2* rs5746134 T allele (hypopharynx,

OR=2.1, 95% CI=1.2-3.7). Four SNPs in alcohol metabolism genes interacted with level of alcohol consumption: *ALDH2* rs2238151, *ADH1B* rs1159918, *ADH7* rs1154460, and *CYP2E1* rs2249695. No interactions with alcohol were found for oxidative stress pathway SNPs.

Minor alleles of two SNPs in *CYP2E1* – the 'C' allele of both rs3813865 and rs8192772 – were associated with increased hazard of cancer-specific death [hazard ratio (HR)=2.09, 95%CI=1.38-3.18; HR=1.71, 95% CI=1.23-2.37, respectively]. No associations with survival were found for SNPs in *ADH1B*, *ADH1C*, *ADH4*, *ADH7*, *ALDH2*, *GPx2*, *GPx4*, and *CAT*.

Previously unreported associations of SNPs in *ALDH2*, *CYP2E1*, *GPX2*, *SOD1*, and *SOD2* with tumor incidence, and in *CYP2E1* with cancer-specific survival, warrant further investigation. Associations with cancer incidence provide evidence that genetic variation in alcohol and oxidative stress pathways influence SCCHN carcinogenesis.

DEDICATION

To the CHANCE study participants, without whom none of this would have been possible.

I hope this project is worthy of their trust.

and

To the Spirit that moves us.

God whispers to us in our pleasures,
speaks to us in our conscience,
but shouts in our pains:
It is His megaphone to rouse a deaf world.

C.S. Lewis

Feel the fear, and do it anyway.

Susan Jeffers

ACKNOWLEDGEMENTS

I would first like to thank everyone on my dissertation committee – Drs. Andrew Olshan, Robert Millikan, Amy Herring, Kari North, and Ivan Rusyn – for their intellectual guidance, critical reviews, and encouragement. These were invaluable to me throughout the process of preparing the research plan, analyzing and writing about the data, and improving this manuscript through many drafts.

I especially thank my dissertation advisor, Andy Olshan, for his leadership in guiding me through the dissertation process. Due to his heavy administrative duties and my long commute, Andy and I did not often have the opportunity to meet in person, so communicating successfully by e-mail and telephone was crucial. I appreciate Andy for his thoughtful, focused and swift replies to my e-mail questions and emergencies. I sent a lot of those messages late at night, and Andy typically responded very early the next morning, in spite of a more than full load of administrative duties.

I also thank Bob Millikan for his support and perspective as my academic advisor during a challenging first three years in the program, and for inspiring me to focus on cancer epidemiology for my dissertation research. Go raibh mile maith agat!

Many other individuals helped bring this project to fruition. Dr. Jill Barnholtz-Sloan contributed insight and technical expertise on estimating ancestry, and Drs. Mark Weissler and William Funkhouser shared their expert clinical judgments and advice. Joanna Smith, Patricia Basta, and Kimberly Ludwig gave critical assistance in obtaining, cleaning, documenting, and formatting CHANCE data for this project, and student members of the CHANCE research team (Anne Butler, Jessica Jensen, Annah Layman, and Jeanette Stingone) provided additional variables, documentation, analyses, and discussion.

The members of my dissertation support group have been a steadfast source of technical and emotional support. Weekly discussions with such intelligent, kind, generous, and witty colleagues were a regular source of inspiration.

I am thankful for financial support received from the National Institutes of Health Cancer training grant, the National Institute for Occupational Safety and Health training grant, and several years of rewarding employment as a graduate research assistant at NC DETECT.

Finally, I am deeply grateful for my husband Paul's unfailing love, enthusiasm, patience, encouragement, and support through this five-year process. And thanks as well to my daughters Katherine and Mary, and friends and family who rejoiced at this long-anticipated venture and kept me focused on what is truly important. At this juncture, I know they are all looking forward to a conversation with me that does not include the "D" word! Having taken the scenic route to reach this milestone, I am gratified to have arrived and look forward to focusing on a fulfilling career in epidemiology and public health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

ADH	Alcohol dehydrogenase (generic term for family of enzymes)
ADH1b	Alcohol dehydrogenase 1b enzyme (class I), beta polypeptide (italicized version refers to gene)
ADH1c	Alcohol dehydrogenase 1c enzyme (class I), gamma polypeptide (italicized version refers to gene)
ADH4	Alcohol dehydrogenase 4 enzyme (class II), pi polypeptide (italicized version refers to gene)
ADH7	Alcohol dehydrogenase 7 enzyme (class IV), mu or sigma polypeptide (italicized version refers to gene)
ALDH	Aldehyde dehydrogenase (generic term for family of enzymes) (italicized version refers to gene)
ALDH2	Aldehyde dehydrogenase 2 family (mitochondrial) (italicized version refers to gene)
CAT	Catalase enzyme (italicized version refers to gene)
CDK4	Cyclin-dependent kinase 4, an enzyme required for for cell cycle G1 phase progression
CDK6	Cell division protein kinase 6, an enzyme required for cell cycle G1 phase progression and G1/S transition
CHANCE	Carolina Head and Neck Cancer Epidemiology study
Cyclin D1	Enzyme that is a regulatory subunit of CDK4 or CDK6
CYP2E1	Cytochrome P-450, family 2, subfamily E, polypeptide 1 enzyme (italicized version refers to gene)
GPx1	Glutathione peroxidase 1 enzyme (italicized version refers to gene)
GPx2	Glutathione peroxidase 2 enzyme (gastrointestinal) (italicized version refers to gene)
GPx4	Glutathione peroxidase 4 enzyme (phospholipid hydroperoxidase) (italicized version refers to gene)
HPV	Human papillomavirus
IARC	International Agency for Research on Cancer
INHANCE	International Head and Neck Cancer Epidemiology Consortium
LOH	Loss of heterozygosity

p16	Cyclin-dependent kinase inhibitor 2A, (CDKN2A), a tumor suppressor
p53/TP53	Protein 53 or tumor protein 53, a tumor suppressor (italicized version refers to gene)
PTEN	Phosphatase and tensin homolog enzyme (italicized version refers to gene)
Rb	Retinoblastoma
ROS	reactive oxygen species
SCCHN	Squamous cell carcinoma of the head and neck (mouth, pharynx, larynx)
SCC	Squamous cell carcinoma
SOD1	Superoxide dismutase 1 enzyme (copper-zinc SOD) (italicized version refers to gene)
SOD2	Superoxide dismutase 2 enzyme (manganese SOD) (italicized version refers to gene)
UADT	Upper aerodigestive tractp53/TP53 Protein 53 or tumor protein 53, a tumor suppressor (italicized version refers to gene)
PTEN	Phosphatase and tensin homolog enzyme (italicized version refers to gene)
Rb	Retinoblastoma
ROS	reactive oxygen species
SCCHN	Squamous cell carcinoma of the head and neck (mouth, pharynx, larynx)
SCC	Squamous cell carcinoma
SOD1	Superoxide dismutase 1 enzyme (italicized version refers to gene)
SOD2	Superoxide dismutase 2 enzyme (italicized version refers to gene)
TP53	Gene that codes for p53 protein
UADT	Upper aerodigestive tract

CHAPTER 1

BACKGROUND AND REVIEW OF LITERATURE

1.1 INTRODUCTION

Head and neck cancers and their treatment differ from other human cancers in important ways. Not only do organs in the head and neck provide speaking ability, they also provide four of the five senses -- sight, hearing, taste, and smell -- as well as enable the essential life functions of breathing and eating. We interact with our environment primarily through the head and neck. Since cancers occurring in this region and their treatment may therefore result in deficits in any of these organs and their corresponding function, quality of life for people with head and neck cancer is more negatively impacted than for people with most other cancers. Pain, noticeable disfigurement and disability, and depression are common sequelae. Consequently, people with head and neck cancer require more rehabilitation and emotional support compared to those with most other cancers.

1.2 TUMOR CHARACTERISTICS AND DESCRIPTIVE EPIDEMIOLOGY

1.2.1 Tumor characteristics

Cancers of the head and neck are defined by location, histology, and the fact that most are associated with use of tobacco and alcoholic beverages. Cancers of the head and neck (or upper aerodigestive tract/UADT) usually include cancers of the oral cavity (lip, tongue, gum, floor of mouth, other mouth, salivary glands); pharyngeal cancer (nasopharynx, hypopharynx, oropharynx, pharynx NOS); cancers of the nose, nasal cavity

and nasal sinuses; and laryngeal cancer. More than 90% of these cancers are squamous cell carcinoma (SCC). Excluded from the group are cancers of the skin, bone, pituitary gland, pineal gland, thyroid, eye, ear, brain, and meninges.

Nasopharyngeal cancer and cancer of the nose, nasal cavity and nasal sinuses are rare in comparison to cancers of the other upper aerodigestive sites, and are inconsistently associated with tobacco and alcohol use. In addition, non-squamous cell cancers may well have a different etiology and risk factors than SCCs. Therefore these cancers are not included in this study.

This dissertation analyzed case-control data from CHANCE, the Carolina Head and Neck Cancer Epidemiologic study. CHANCE enrolled cases of invasive squamous cell carcinomas of the head and neck (SCCHN), encompassing cancers of the oral cavity including lip, tongue, gum, floor of mouth, and other mouth; cancers of the pharynx including oropharynx, hypopharynx, and pharynx NOS; and laryngeal cancers. It excluded patients with nasopharyngeal cancer and cancers of the nose, nasal cavity, and nasal sinuses. CHANCE enrolled a small number of cases with lip cancer (n=21) and included them in its definition of SCCHN; however, this dissertation project excluded lip cancers from final analyses because the etiology of lip cancer (sunlight exposure in addition to alcohol and tobacco) differs from other SCCHNs.

1.2.1.1 Anatomic distribution

In the United States, the age-adjusted incidence rate of cancer of the oral cavity, larynx, and pharynx (excluding salivary gland and nasopharynx), averaged from 2002-2006, was 12.0 cases per 100,000. Slightly more than half of these occurred in the larynx and tongue, and over 80% occurred in the larynx, tongue, oropharynx, tonsil, gum and other mouth (Table 1).

Table 1. Anatomic distribution of head and neck cancers included in this study, United States 2002-2006

Anatomic Location	Annualized age-adjusted incidence rate per 100,000 (%)*	Cumulative %
Larynx	3.5 (29.2)	29.2
Tongue	2.8 (23.3)	52.5
Oropharynx and tonsil	1.9 (15.8)	68.3
Gum and other mouth	1.5 (12.5)	80.8
Lip	0.8 (6.7)	87.5
Hypopharynx	0.7 (5.8)	93.3
Floor of mouth	0.6 (5.0)	98.3
Other oral cavity and pharynx	0.2 (1.7)	100.0
ALL SITES COMBINED	12.0 (100)	

*SEER incidence data used in calculating the rates are from the 17 SEER areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta, San Jose-Monterey, Los Angeles, Alaska Native Registry, Rural Georgia, California). Rates are age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130).

1.2.1.2 Pathogenesis and premalignant lesions

Oral and many pharyngeal cancers are often clinically preceded by precursor lesions and conditions. The best-characterized such lesions include oral leukoplakia, erythroplakia, and oral submucous fibrosis. Oral leukoplakia and erythroplakia are diagnoses of exclusion; they are defined as adherent white (leukoplakia) or red (erythroplakia) patches on the mucous membranes of the oral cavity that cannot be defined as another type of lesion. Oral submucous fibrosis is a chronic debilitating disease characterized by inflammation and progressive fibrosis of submucosal tissues, resulting in stiffness of the mucosa.

Erythroplakia is much more likely than leukoplakia to contain epithelial dysplasia, thus resulting in higher risk of eventual progression to cancer. Most oral lesions are found to be leukoplakia, in which 4 to 18% have been found to progress to cancer (1). Similarly, an Indian study reported that about 8% of oral submucous fibrosis cases eventually progress to cancer (2). In contrast, although much less common than leukoplakia, almost all erythroplakia lesions are found to be dysplastic or already-cancerous at the time of

diagnosis (1). Oropharyngeal tumors associated with human papillomavirus (HPV) infection almost exclusively occur in tonsillar crypts, where even small tumors often first present as regionally metastasized, with little evidence of precursor lesions.

Laryngeal cancers are associated with a sequence of changes in the laryngeal epithelial mucosa that often includes laryngeal keratosis. These usually white, sometimes red, plaques appear similar to oral leukoplakia, and may present with or without atypia. They are often found near tumors and have a high transformation rate of from 1 to 40% (3) that is correlated with the amount of atypia. However, a significant percentage of laryngeal cancers present at an advanced stage without evidence of precursor lesions.

1.2.1.3 Molecular characteristics

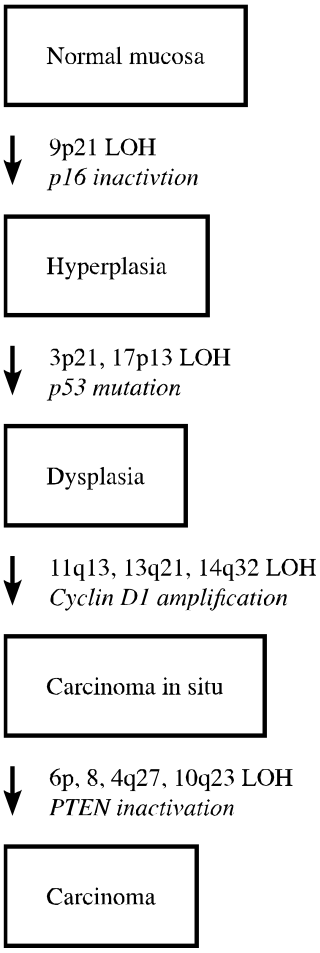
SCCHN has traditionally been regarded as a homogeneous disease requiring a relatively common therapeutic regimen. However, differences in epidemiological risk factors and incidence and mortality trends among SCCHNs arising from different anatomic sites have supported the view that head and neck cancers form a heterogeneous group with distinct etiologies and prognoses that may require distinct treatments.

Tobacco smoking is the most well characterized risk factor for SCCHN. Although heavy alcohol consumption is a documented risk factor in its own right, its effects are usually manifested in its ability to synergistically enhance the effects of tobacco smoking. The risk of cancer development is much higher for those who heavily drink and smoke than would be expected by examining the risks for smoking and drinking alone (4). Although tobacco and alcohol exposure are responsible for the large majority of SCCHNs in the oral cavity, larynx, and hypopharynx, their role in carcinogenesis in the oropharynx is much reduced. Oncogenic HPV infection, especially type 16, has been identified as a causal agent in about 70% of oropharyngeal SCCHN, and in HPV-associated cases, alcohol and tobacco are not associated as risk factors (5).

Much research has investigated the relative order of genetic and epigenetic changes that occur during carcinogenesis. An understanding of the type and order of such changes may provide insight into the carcinogenetic process ending in SCCHN. Early changes may be most common and therefore might be suitable targets for early therapeutic intervention, while later alterations may be less common but might serve as prognostic factors and therapeutic targets for certain subsets of tumors. Clinical and histologic progression from squamous hyperplasia to dysplasia, carcinoma *in situ*, and eventually invasive carcinoma are driven by progressive accretion of chromosomal genetic alterations (Figure 1). The transition from normal mucosa to hyperplasia is driven by loss of heterozygosity (LOH) at chromosome 9p21, which inactivates p16 (CDKN2A), a cell cycle negative regulator. The transition from hyperplasia to dysplasia is accompanied by LOH at 3p21 and 17p13, causing p53 mutation and/or up-regulation. Transition from dysplasia to carcinoma *in situ* is facilitated by LOH at 11q13, 13q21, and 14q32, resulting in Cyclin D1 amplification. Cyclin D1 functions as a regulatory subunit of CDK4 and CDK6, whose activity is required for cell cycle G1/S transition; it has been shown to be positively regulated by interaction with retinoblastoma (Rb) protein. The final transformation to frank carcinoma is mediated by additional LOH at 6p, 8, 4q27, and 10q23, which inactivates the PTEN tumor suppressor gene (5).

In almost all SCCHNs, the tumor-protein p53 (TP53) and Rb gene pathways are disrupted. TP53 regulates the cell cycle and is a tumor suppressor; over 50% of SCCHN tumors contain inactivating *TP53* mutations or chromosomal loss at 17p where the *TP53* gene resides (5). The most common mutation in the Rb pathway is the $p16^{INK4A}$ tumor suppressor gene. Lippman et al. demonstrated a wide range of p53 protein expression in 90% of SCCHN tumors, compared with no expression in any normal oral cavity epithelium (6). Boyle et al. demonstrated that mutation in the *TP53* gene occurred in 19% of

Figure 1. Genetic progression model of SCCHN tumorigenesis*



Clinical and histologic progression from squamous hyperplasia through progressively more advanced stages of dysplasia, resulting eventually in invasive SCCHN, are driven by progressive accumulation of genetic and epigenetic transformations.
* adapted from Pai and Westra (5).

non-invasive lesions (severe dysplasias and carcinoma *in situ*), increasing to 43% of invasive carcinomas (7). Such studies indicate abnormal increases in p53 *expression* may occur early in carcinogenesis, but that p53 *mutations* may occur later in the process. *TP53* mutations have been found to occur more frequently in smokers than in non-smokers (8), though not uniformly so. The pattern of base pair changes in *TP53* is different between smokers and non-smokers, with significantly more mutations occurring in CpG islands in the nonsmokers than in smokers (8).

Although HPV-related SCCHNs also display abnormal amounts of p53 and Rb, they most often do not contain *TP53* or *Rb* mutations, thus implying the presence of alternative (to mutation) oncogenic mechanisms. In most of these tumors, p53 and Rb proteins are bound and degraded by the viral oncoproteins E6 and E7, respectively, rendering upstream mutation of the gene unnecessary to the oncogenic process. (5)

Subsequent tumors, in people with primary SCCHN tumors that were completely surgically removed, are often fatal and develop not only in the head and neck but also in the lungs or esophagus. Ten to forty percent of patients experience such a secondary tumor (9-11). This observation led to the 1953 field cancerization theory of Slaughter (12) which postulates that large portions of the epithelial

surface of the respiratory tract are damaged simultaneously (although not identically) by carcinogen exposure. According to this theory, multiple lesions are thought to develop independently of each other, and therefore would most likely have non-identical mutations. Recent research (13) supports the field cancerization theory: nearby secondary SCCHN tumors often have genetically identical mutations, while more distant secondary tumors almost always have different mutations than the original lesion. The realization that the pathologist or surgeon cannot accurately identify all damaged sections of epithelium that should be removed is fueling a search for biomarkers that could be used in diagnosis and treatment planning.(5)

1.2.2 Descriptive epidemiology

1.2.2.1 Incidence and mortality in the United States

SCCHN is one of the ten most frequently diagnosed cancers in men in the United States; three-quarters of incident cases and deaths occur in men. There were estimated to be 52,140 new cases of oropharyngeal and laryngeal cancer and 11,460 deaths in the U.S. in 2011 (14). SCCHN patients have a relatively poor prognosis compared to those diagnosed with other cancers and experience significant morbidity, including disfigurement and psychosocial sequelae, related to the cancer and its treatment (15, 16). SCCHN incidence increases with age and is more common in all races for men than women by a factor of 2 to 3 for oropharyngeal cancers, and by a factor of 5 to 9 for laryngeal cancers (Table 2). Incidence rates of oropharyngeal cancer are about the same in African-Americans and in whites, while rates of laryngeal cancer are much higher in African-Americans than whites. Compared to both African-Americans and whites, the lowest incidence and mortality rates of SCCHN in both men and women occur in Hispanics, Asian/Pacific Islanders, and Native Americans. Five-year relative survival rate and median survival time for both

oropharyngeal and laryngeal cancers are both lower among African-Americans compared to whites (5-year survival, African-Americans versus whites: 41% vs. 61% for oropharyngeal and 51% vs. 64% for laryngeal cancer) (15).

Table 2. SCCHN incidence and mortality rates by anatomic location, sex and race, United States 2002-2006 (15)

Sex and race	Incidence rate*			Mortality rate**		
	Oral cavity & pharynx	Laryngeal	SCCHN (combined)#	Oral cavity & pharynx	Laryngeal	SCCHN (combined)#
Men – All races	15.4	6.2	21.6	3.9	2.3	6.2
White	15.6	6.2	21.8	3.7	2.1	5.8
Black	16.7	10.5	27.2	6.5	4.7	11.2
Asian/Pacific Islander	10.8	2.7	13.5	3.2	0.7	3.9
Native Americans	9.2	2.4	11.6	3.6	1.9	5.5
Hispanic	9.0	4.9	13.9	2.5	1.9	4.4
Women – All races	6.1	1.3	7.4	1.5	0.5	2.0
White	6.1	1.3	7.4	1.4	0.5	1.9
Black	5.8	2.0	7.8	1.6	0.7	2.3
Asian/Pacific Islander	5.4	0.3	5.7	1.3	0.1	1.4
Native Americans	5.1	---	5.1	1.5	---	1.5
Hispanic	3.5	0.6	4.1	0.8	0.2	1.0

*SEER incidence data used in calculating the rates are from the 17 SEER areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta, San Jose-Monterey, Los Angeles, Alaska Native Registry, Rural Georgia, California). Rates are annualized per 100,000 population and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130).
**US Mortality Files, National Center for Health Statistics, Centers for Disease Control and Prevention. Rates are per 100,000 population and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130).
#SCCHN rates are computed by adding separate rates for oral cavity and pharynx cancers, and laryngeal cancers.

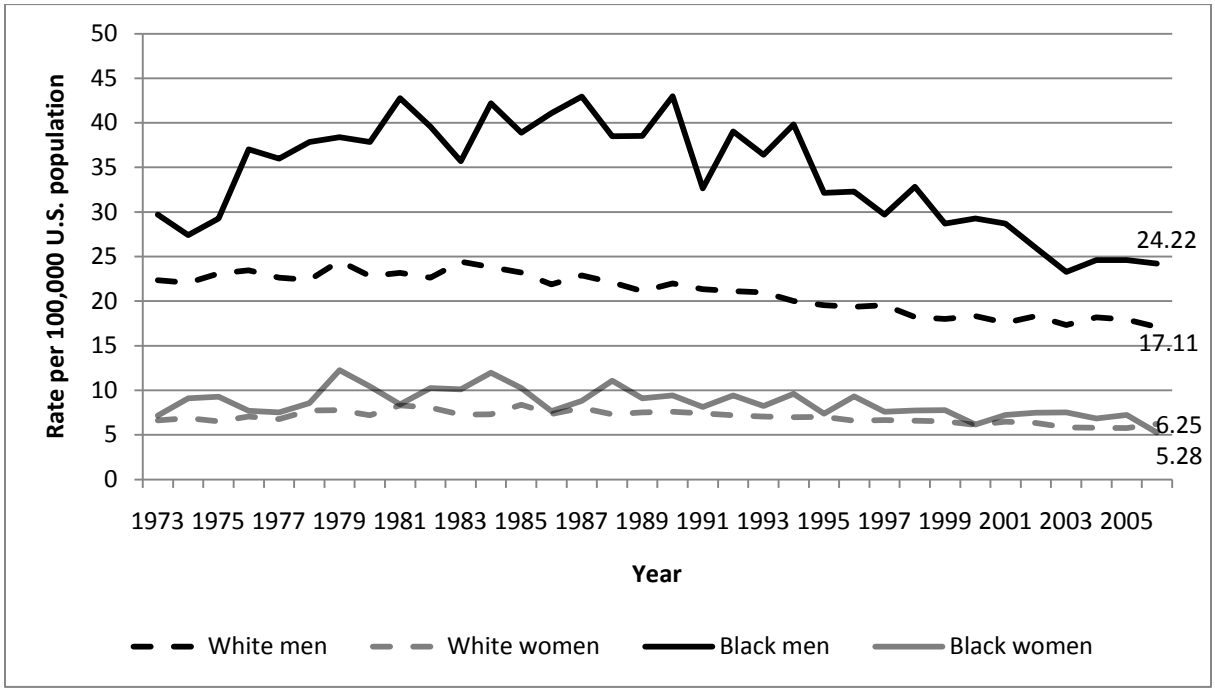
1.2.2.2 Time trends in the United States

In the U.S., between 1984-86 and 1999-01, age-adjusted incidence rates of all first primary tumors of the upper aerodigestive tract (UADT) declined from 13.0 (12.7-13.3) to 10.2 (10.0-10.4). However, there was no significant change in incidence of cancer of the oropharynx (including the tonsils, base of the tongue, and surrounding epithelium); in fact there was an increased incidence in white men younger than 60. During the same time, 5-year relative survival rates for UADT increased from 48.1% to 52.4% in the U.S., driven largely by a dramatic increase in relative survival for cancers of the oropharynx, from 35.3%

to 51.0%.(17) Small increases in relative survival were also observed for cancers of the oral cavity and hypopharynx, but not for laryngeal cancer. These trends are consistent with the hypothesis that that there has been a change in the etiology of oropharyngeal cancer in the U.S. especially in younger men, related to a rise in pharyngeal HPV infection and that accompanies a change in its response to therapy (17).

Ninety-six percent of the population in North Carolina self-declares their race/ethnicity as Black (African-American) or White (European-American), with other race-ethnicities (4% total) each comprising less than 2% of the population. The proportion of the population in North Carolina that is African-American (21.6%) is much larger than in the U.S. as a whole (12.8%), so health disparities that negatively affect African-Americans have a proportionately larger effect in North Carolina. For cancers of the tongue, mouth floor, gum/other, tonsils, oropharynx, hypopharynx, other oral cavity and pharynx, and larynx, incidence rates among African-American men in the U.S. have, since 1973, been much higher than rates in white men and all women (see Figure 2), while rates in white women are lowest. Rates in African-American men climbed steeply from 1973 to the 1980s, then began a sharp decline continuing through the 1990s to the present. This is thought to be related mostly to changing rates of smoking. In contrast, white men experienced a small increase from 1973 to a peak in 1981, after which rates declined consistently until 2000, and remained stable thereafter. Rates in both African-American and white women were much lower than in men from 1973-2005; African-American women experienced higher rates than white women during the 1980s but since then have experienced a slow decline, and currently experience about the same rate as white women.

Figure 2. Time trends of incidence of oropharynx and larynx cancers, U.S., 1973-2005*



* Data from: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 9 Regs Limited-Use, Nov 2008 Sub (1973-2006) , National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2009, based on the November 2008 submission. Oropharyngeal cancer is defined as cancer of the tongue, mouth floor, gum/other, tonsils, oropharynx, hypopharynx, other oral cavity and pharynx, and larynx. Rates are age-adjusted and standardized to the 2000 U.S. population.

1.2.2.3 International incidence rates and trends

Internationally, in both men and women, incidence rates are higher in more developed than in less developed regions. For example, the age standardized rates (world) in men living in more developed regions are 7.9 per 100,000 for oral cavity cancer and 6.9 for laryngeal cancer, compared to 5.7 and 4.3, respectively, in less developed regions. In women, rates are much lower than in men (e.g. 7.9 for men in more developed countries versus 2.4 for women in the same countries) and there are no consistent differences between women living in more and less developed countries. In men, mortality rates are also higher in more developed regions, but in women the reverse is true (18).

Between 1975 and 1995, laryngeal cancer incidence rates declined in both men and women in almost all countries with the exception of Japan. Although incidence rates for pharyngeal, oral cavity, and lip cancers have generally declined in both men and women over the same time period, there have been plateaus or increases in oropharyngeal cancer in the region of the base of the tongue and the tonsils in younger men (age 20-44 years) in the U.S. (19) and many parts of Europe (20). The tongue and tonsils have been shown to be the predominant region of oropharyngeal HPV infection that progresses to cancer (21).

1.3 RISK FACTORS

1.3.1 Heredity

Epidemiologic studies of familial clustering of SCCHN cases provide suggestive evidence for genetic predisposition. Such evidence implies shared genes, shared exposure, or both. In a Brazilian case-control study of 754 cases of SCCHN and 1,507 age- and gender-matched hospital-based controls with non-malignant diseases (22), investigators found that the relative risk (RR) for developing SCCHN was doubled in cases who had a first-degree relative with any cancer, 3.65 (95%CI=1.97-6.76) if the first-degree relative had head and neck cancer, and 8.57 (95% CI=2.72-27.04) if the first-degree relative with head and neck cancer was a sibling. The same authors reported a similar increased risk for developing SCCHN (OR=3.79; 95% CI=1.11-13.0) in a Canadian population if first-degree relatives also had SCCHN (23). A comparable study in Puerto Rico on 342 cases with oral cavity or pharyngeal cancer and 521 controls found an increased OR of 2.6 (95% CI=1.4-4.8) in cases with first-degree relatives with any upper aerodigestive tract (UADT) tumor (24). Ankathil et al. (25) created a detailed pedigree analysis for each of their Indian oral cancer patients, based on a questionnaire eliciting information about family history of cancer and especially SCCHN. They observed, as with other familial cancers, that a family history

of SCCHN was associated with an early age of onset. In addition, family members who did not use tobacco products or smoke were also affected. Such familial aggregation was observed in 1% of their oral cancer cases. The authors therefore postulate the existence of an oral cancer susceptibility gene. Morita et al. (26) compared family cancer history in 167 patients with hypopharyngeal or cervical esophageal cancers versus 167 controls with benign diseases. The odds ratio for cases with family history of UADT tumors in relation to controls was 2.6 (95% CI=1.1-6.3), and these cases were significantly younger than cases whose relatives had another type of cancer or no cancer. In contrast, a study by Goldstein et al. (27) found only non-significant excess risk among those with a family history of oropharyngeal tumors (OR=1.2, 95% CI=0.7-2.3) or esophageal/laryngeal cancers (OR=1.6, 95% CI=0.7-3.8). They found that excess familial risk of oropharyngeal cancers was associated with smoking-related cancers among male but not female relatives, and conclude that it is likely that environmental factors such as smoking and drinking contribute to the familial tendencies they observed; they were unable to discriminate genetic from environmental factors in their study. A recent pooled analysis of 8,967 head and neck cancer (HNC) cases and 13,627 controls from 12 case-control studies found that a family history of HNC in first degree relatives increased the risk of HNC (OR=1.7; 95% CI: 1.2-2.3) (28). The risk was higher when the affected relative was a sibling (OR=2.2, 95% CI: 1.6-3.1) rather than a parent (OR=1.5, 95% CI=1.1-1.8). The OR was also higher in subjects exposed to tobacco, and rose to 7.2 (95% CI=5.5-9.5) in subjects with family history who also used alcohol and tobacco. Interestingly, the study found a small but significant association (OR=1.1, 95% CI=1.0-1.2) with family history of other tobacco-related cancers, particularly with laryngeal cancer (OR=1.3; 95% CI=1.1-1.5). No association was found between family history of non-tobacco-related tumors and risk of HNC.

The Swedish Family-Cancer Database was used to investigate familial clustering of cancers at HPV-associated sites, including cervical, anogenital, UADT and skin (29).

Investigators calculated standard incidence ratios (SIR) for offspring site-specific cancer risks according to site-specific cancers in sibling and parental probands. Risk of UADT cancers in offspring were increased by having a parent or sibling with certain cancers; for women, risk was increased by having a sister with cervical SCC (SIR=1.37; 95% CI=1.03-1.79), a mother with skin SCC (SIR=1.50; 95% CI=1.00-2.18), or a father with UADT SCC (SIR=1.92; 95% CI=1.09-3.12); for men, risk was increased by having a sibling or father with UADT SCC (SIRs=2.36, 1.66) or a mother with vulvar SCC (SIR=2.47; 95% CI=1.41-4.02).

1.3.2 Primary environmental exposures

1.3.2.1 Tobacco

Two International Agency for Research on Cancer (IARC) Working Groups stated that there was sufficient evidence to conclude that tobacco is carcinogenic to humans and that the occurrence of malignant tumors of the respiratory and upper digestive tract is causally related to the smoking of different forms of tobacco (cigarettes, cigars, pipes, bidis, etc.) (30, 31). It is estimated that tobacco smoking confers an increased risk of 4 to 5 times for oral cavity, oropharyngeal, and hypopharyngeal cancers, and 10 times for laryngeal cancers (32).

Tobacco smoke. An increased risk of head and neck cancer in tobacco smokers was reported in the first IARC tobacco monograph in 1986 (30), and reiterated in the updated report issued in 2004 (31). The 2004 monograph reported that from 1987 to 2002 numerous studies measured the association between tobacco smoking and head and neck cancers in various regions of the world including North and South America, Europe, and Asia: three cohort and 16 case-control studies on oral cavity cancer; three cohort and 12 case-control studies on oropharyngeal, hypopharyngeal, or pharyngeal cancers in general; and five cohort and 26 case-control studies on laryngeal cancer. Most studies reported dose-

response relationships between cancer risk and frequency (e.g. cigarettes per day), duration (years), and lifetime consumption (e.g. pack-years).

In an attempt to distinguish the often-correlated effects of lifetime consumption and exposure rate, Lubin et al. (33) used pooled data from 15 case-control studies to compare excess odds ratios (EOR) for lifetime tobacco (smoking) consumption accumulated at high frequency but short duration, with the same lifetime consumption accumulated at lower frequency and longer duration. Above 15 cigarettes per day, the EOR per pack-year decreased, suggesting that more cigarettes per day for shorter duration was less hazardous than fewer cigarettes per day for longer duration. Estimates of EOR per pack-year were similar for oral cavity, pharyngeal and laryngeal cancers, but the effects of cigarettes per day differed and were higher for laryngeal cancers, indicating that increased risk of laryngeal cancer is more affected by cigarettes per day effects rather than pack-years.

While all types of tobacco smoking increase SCCHN risk, smokers of black tobacco experience a higher cancer risk than smokers of blond tobacco, and hand-rolled cigarettes confer a higher risk than manufactured cigarettes (31). Other types of smoked tobacco, including cigars and pipes, also confer increased risk of head and neck cancer, but it hasn't been possible to precisely quantify differences in risk between types of smoked tobacco because of the small number of people who smoke only cigars or pipes. The few studies to examine this question among exclusive users of cigars or pipes did not observe differences in their risk compared to the risk among cigarette smokers (31). Another type of commonly smoked tobacco product in Southeast Asia is bidi. A meta-analysis of 12 case-control studies reported approximately triple the risk of head and neck cancer among bidi users compared to never smokers (34). A multi-center case-control study from India reported that bidi smoking was a stronger risk factor than cigarette smoking for cancers of the hypopharynx (OR 6.80 versus OR 3.82) and supraglottis (OR 7.53 versus OR 2.14) but about the same for cancer of the glottis (35).

The primary hypothesis explaining the risk difference between smokers of black and blond tobacco is that smokers of black tobacco are exposed to more *N*-nitrosamines and aromatic amines. These compounds require metabolic activation by *N*-acetylation and *O*-acetylation to form acetoxy intermediates which can form DNA adducts. Genes coding for *N*-acetyltransferase enzymes (NAT1 and NAT2), which are active in both *N*- and *O*-acetylation, are highly polymorphic, with different variants having strikingly different rates of enzyme activity. Fast acetylators are hypothesized to incur higher SCCHN risk. Although NAT1 and NAT2 polymorphisms have been definitely implicated as increasing risk for bladder cancers, studies examining their role in head and neck cancer have not consistently shown increased risk for SCCHN (36).

Age at which smoking began appears to be associated with risk of head and neck cancer, if frequency and duration aren't taken into account (37-40). This association may appear because smoking of longer duration is probably associated with age at start of smoking. However, studies that investigated age at start and cancer risk, and adjusted estimates for tobacco smoking habits, have shown inconsistent results. For example, a Cuban study that adjusted for tobacco smoking habits reported higher risk among those who began smoking when <17 years of age compared to those who began smoking when older (41), while studies in France (42) and India (43) reported no difference in cancer risk based upon age at start.

Cessation of tobacco smoking has been consistently shown to reduce the risk of many types of head and neck cancer, such that former smokers have consistently lower relative risks compared to current smokers. The 2004 IARC monograph summarized one cohort study and eight case-control studies of oral cancer that reported lower risk for former compared to current smokers (31). Seven of the case-control studies reported a negative trend for risk of oral cancer with time since quitting, and a rapid decline in relative risks compared to nonsmokers after 10 or more years. A similar reduction in risk of pharyngeal

cancer was reported by one cohort study and two case-control studies. A pooled analysis by the International Head and Neck Cancer Epidemiology (INHANCE) Consortium showed that cancer risk 1-4 years after quitting was reduced to 70% of that experienced by current smokers (OR=0.70; 95% CI: 0.61-0.81) with complete risk reduction down to that of nonsmokers after 20 or more years of not smoking (OR=0.23; 95% CI: 0.18-0.31) (44).

Environmental tobacco smoke. In a 2008 pooled analysis of 542 cases and 2,197 controls who reported never using tobacco, no effect of environmental tobacco smoke (ETS) was found for those ever-exposed (45). However, risk increased significantly for those who reported more than 15 years of exposure at home (OR=1.60; 95% CI=1.12-2.28) or at work (OR=1.55; 95% CI=1.04-2.30). The effect of duration of ETS exposure at home was stronger for pharyngeal and laryngeal cancers compared to other sites. Increased risk of similar magnitude also occurred in those who reported using neither tobacco nor alcohol but were exposed to ETS for more than 15 years. A larger 2009 study of 2103 UADT cases and 2221 controls (46) reported increased risk for ever-exposure only at home (OR=1.49; 95% CI=0.90-2.48), only at work (OR=1.79; 95% CI=1.03-3.13, and both home and work (OR=1.68; 95% CI=1.00-2.83).

Smokeless tobacco. Smokeless tobacco products are ingested without burning. Oral products are placed in the mouth, cheek or lip and are sucked (“dipped”) or chewed. In the United States, most oral tobacco products can be classified as loose leaf (air cured and sweetened), plug (heavier tobacco leaves, sweetened and compressed), twist (air- or fire-cured leaves, flavored and twisted), and snuff (finely cut or powdered tobacco; moist snuff contains water, while dry does not). In 2000, 4.4% of US men and 0.3% of women used snuff or chewing tobacco (47). Worldwide a variety of products are used; for example tobacco pastes or powders which are applied to gums or teeth, and products which contain tobacco as one ingredient (e.g. snus; some betel quid products). Use varies widely between countries and within countries based on age, sex, ethnicity, and socioeconomic status.

Epidemiologic data from the United States and Asia report an increased risk of oral cancer (RR=2.6; 95% CI=1.3-5.2), while northern European studies reported no increased risk (RR=1.0; 95% CI=0.7-1.3) (48). In India, more than half of oral cancers are estimated to be attributable to smokeless tobacco products, while in the United States 6.6% of oral cancers in men are estimated to be attributable to such products (48). SCCHN risk of smokeless tobacco users lies between the rates of non-tobacco users and smokers.

Population attributable risk. The INHANCE Consortium analyzed individual-level pooled data from 17 European and American case-control studies (11,221 cases and 16,168 controls). They estimated that the population attributable risk for tobacco alone was 33% (49).

1.3.2.2 Alcohol

The International Agency for Research on Cancer (IARC) concluded in 1988 that there was sufficient evidence of the carcinogenicity of alcohol for cancers of the oral cavity, pharynx, and larynx, as well as esophagus and liver (50). Evidence from many studies of various design and in different international populations has consistently demonstrated that daily consumption of about 50 grams of ethanol increases the risk for oral cavity and pharyngeal cancer three-fold, and the risk for laryngeal cancer two-fold. For these cancers, the effects of drinking and smoking are multiplicative (51).

Independent effect. Due to the small number of alcohol-drinking non-smokers in most populations, estimates from smaller studies of the independent effect of alcohol on risk of head and neck cancer have been difficult to measure precisely. Therefore, in order to estimate the effect of alcohol drinking apart from smoking, the risk of developing head and neck cancer in alcohol drinkers who had never smoked was investigated by the INHANCE Consortium (52). This study pooled individual-level data from 15 case-control studies that included 10,244 head and neck cancer cases and 15,227 controls; of these, 1,072 cases

and 5,775 controls were never-users of tobacco. Their results (Table 3) suggest that the association between alcohol consumption and the risk of head and neck cancer is weak for low frequency drinking (<3 alcoholic drinks per day) and is apparent only at higher frequency and only for pharyngeal and laryngeal cancers. The dose-response trend for frequency of alcohol drinking was strongest for cancers of the pharynx and larynx and weakest for oral cancers. The results of two smaller studies (that were not included in their analysis) also suggested an increased head and neck cancer risk at high frequency of alcohol intake in never smokers (53, 54).

Dose-response. The 2008 IARC monograph on alcohol (55) reported that between 1988 (the publication date of the previous IARC monograph on alcohol) and 2007, 58 epidemiologic studies of head and neck cancer were published in the literature: 5 cohort studies on oral cavity and pharyngeal cancers, and 53 case-control studies (8 on oral cavity cancer, 9 on pharyngeal cancer, 18 on oral cavity and pharyngeal cancers combined, and 18 on laryngeal cancer). The study populations included Europeans, North and Latin Americans, and Asians. Most adjusted for tobacco smoking and consistently reported a strong dose-response relationship between frequency of drinking (i.e. drinks per day or per week) and risk of head and neck cancer. Hashibe et al. (52), using data from the INHANCE Consortium on never-smokers, found no dose-response relationship between duration of alcohol drinking and risk of oral cavity, pharyngeal, and laryngeal cancers (52). Lubin et al. (33) examined total alcohol exposure (drink-years) and exposure rate (drinks per day) and the risk of head and neck cancer in the INHANCE pooled data. Their analyses suggest that for participants who drank fewer than 10 drinks per day, at fixed total alcohol exposures, exposure to higher frequency of drinking (drinks per day) brought higher risk of head and neck cancer than exposure to lower frequency of drinking over longer duration. There were few participants who reported drinking more than 10 drinks per day, so no conclusions could be drawn about such exposure frequencies. The association of alcohol drinking with greater

risk of oral cavity and pharyngeal cancers than of laryngeal cancer was confirmed in their study, and results suggested that these risk differences by site were more attributable to cumulative alcohol consumption (drink-years) rather than frequency (drinks per day).

Age at start of alcohol drinking. Only a few studies have reported on the association between age at start of drinking alcoholic drinks and risk of head and neck cancers. Two early studies which didn't control for smoking or drinking habits reported that starting age wasn't associated with head and neck cancer risk (40, 56). One study which adjusted for tobacco smoking also showed no effect of age when drinking began on head and neck cancer risk (57). Two studies that adjusted for both tobacco and alcohol habits reported the same lack of effect (41, 43).

Cessation of alcohol drinking. There are few published studies on the risk of head and neck cancer associated with discontinuance of alcohol drinking. Such studies may be confounded by indication because pre-cancerous and early lesions of the oral cavity and pharynx could themselves cause alcohol cessation. To summarize four studies from 1988 to 2004, for recent quitters, risk for oral and pharyngeal cancers rises above current drinkers; as the number of years increases, the risk slowly drops below the risk of current drinkers and approaches the levels of non-drinkers after 10-20 years (55). In the INHANCE pooled analysis, a beneficial effect on the risk of head and neck cancer compared with current drinkers was only observed after 20 or more years of quitting alcohol (OR=0.60, CI=0.40–0.89), when the risk was reduced to the level of never drinkers (44). This is in contrast to the benefit for quitting smoking, which is observed as early as 1 to 4 years after quitting.

Table 3. Alcohol drinking frequency and risk of head and neck cancer, overall and by anatomic sub-site, among never-users of tobacco[*]

Drinking variable	All sites[‡]		Oral cavity[€]		Oropharynx and hypopharynx[‡]		Oral cavity and pharynx[€]		Larynx[€]	
	# cases / # controls	OR (95% CI)	# cases / # controls	OR (95% CI)	# cases / # controls	OR (95% CI)	# cases / # controls	OR (95% CI)	# cases / # controls	OR (95% CI)
Frequency, drinks/day										
Never	541/2482	1.00 (ref)	243/2482	1.00 (ref)	153/2482	1.00 (ref)	80/1982	1.00 (ref)	40/1788	1.00 (ref)
<1	164/899	1.04 (0.79-1.38)	44/899	1.14 (0.8-1.63)	73/899	1.39 (0.99-1.96)	25/834	1.08 (0.67-1.75)	15/724	0.92 (0.5-1.69)
1-2	202/1324	1.30 (0.94-1.80)	60/1324	1.64 (1.19-2.25)	83/1324	1.66 (1.18-2.34)	26/1171	1.24 (0.77-1.99)	28/1155	1.26 (0.77-2.07)
3-4	59/536	1.82 (1.10-2.99)	10.536	1.11 (0.57-2.15)	24/536	2.33 (1.37-3.98)	13/495	2.32 (1.24-4.34)	11/480	1.24 (0.62-2.45)
≥5	65/389	2.81 (1.49-5.27)	8/389	1.23 (0.59-2.57)	29/389	5.50 (2.26-13.36)	4/382	0.77 (0.27-2.18)	22/349	2.98 (1.72-5.17)
P_{trend}		0.001		0.032		<0.001		0.891		<0.001
P_{het}†		<0.001		0.913		<0.001		0.202		0.006

*reported by Hashibe et al. (2007) (52)
†Two-sided test for heterogeneity between studies
‡Random-effects model. Study-specific ORs adjusted for age, sex, race/ethnicity, and education level.
€ Fixed-effects model, adjusted for age, sex, race/ethnicity, education level, and study center.

Types of alcohol beverages. Nine studies conducted between 1988 and 2007 examined differences in risk by type of alcoholic beverage, including wine, beer, and liquor (55). Highest risks were observed for beer in North America, wine in Europe, and hard liquors in Latin America. The consensus is that the most common alcoholic beverage in the region studied produced the greatest risk. Because most drinkers consume more than one type of beverage, it is difficult to separate their effects; to overcome this, a recent INHANCE Consortium analysis studied risk in each beverage type among individuals who drank only one type (58). Risk of head and neck cancer was doubled over that of non-drinkers for those who drank 16-30 beer or liquor beverages per week, while risk for drinkers of 16-30 glasses of wine per week was not significantly increased. However, above 30 drinks per week, all alcoholic beverages were associated with increased risk: odds ratios were about 4 for liquor, 5 for beer, and 6 for wine. Confidence intervals overlapped, suggesting no significant differences. Risk estimates by geographic region (North America, Latin America, Europe) for those drinking more than 15 drinks per week suggested somewhat higher risk for beer in North America and Europe, higher risk for liquor in North and Latin America, and higher risk for wine in Europe. Odds ratios were stronger for oral and pharyngeal cancer than for

laryngeal cancer and among smokers compared with never smokers. The pattern of associations across beverage types did not change with anatomic site or smoking status (58).

Alcohol and tobacco interaction. Between 1988 and 2007, two prospective cohort and 15 case-control studies in the United States, Europe, Asia, and Australia described and estimated the joint effects of alcohol drinking and tobacco smoking on risk of oropharyngeal cancer. The studies differed in their method of assessing effect measure modification, ranging from descriptive to estimation of interaction coefficients in multivariate regression models (55). All of the studies found that alcohol ingestion and tobacco smoking have a synergistic effect on risk of oropharyngeal cancer. The large majority of studies (n=14) demonstrated greater than multiplicative interaction, while a few (n=3) found greater than additive but less than multiplicative interaction (55). The largest of the case control studies (1560 cases and 2948 hospital controls) had the most precise estimates; the OR for the highest frequency of smoking in non-drinking or occasional-drinking men was 2.0 (95% CI=1.1-3.6), OR for highest frequency of drinking in men who never smoked was 2.9 (95% CI: 1.1-8.1), and OR for the highest frequencies of both smoking and drinking in men was 20.1 (95% CI: 12.9-31.5), much higher than multiplicative (59). Results in women were similar. The second-largest study (1,114 cases and 1,268 population controls) categorized alcohol intake by frequency and cigarette smoking by frequency only in those who had smoked for 20 or more years; the OR for the highest alcohol intake in non-smokers was 5.8, OR for the highest smoking frequency was 7.4, and OR for the highest frequencies of both smoking and drinking in men was 37.7, or approximately multiplicative (56). One study in India of 591 cases of oral cancer and 582 controls reported no interaction between tobacco (in paan) chewing (as opposed to smoking) in never/current users of alcohol: OR for oral cavity cancer for current drinkers who didn't chew paan was 2.83 (95% CI: 1.58-5.09), OR

for current paan chewers who never drank was 7.31 (95% CI: 3.79-14.10); and OR for current drinkers and paan users was 8.62 (95% CI: 4.12-18.06) (43).

For laryngeal cancer, 16 studies published from 1982 through 2002 computed risk estimates for the highest level of consumption of both tobacco smoking and alcoholic beverage drinking, and concluded that a multiplicative risk model best fit the data (55). For example, Talamini et al. reported in 2002 that the OR for non-smokers in the highest drinking category was 8.5 (95% CI: 2.4-30.2), the OR for non-drinkers in the highest smoking category was 18.9 (95% CI: 5.7-62.7), and the OR for the highest frequencies of both smoking and drinking was 177.2 (95% CI: 65.0-483.3), or approximately multiplicative (4).

In order to estimate the extent of interaction between alcohol and tobacco exposures and the risk of head and neck cancer, as well as population attributable risk (PAR), the INHANCE Consortium analyzed individual-level pooled data from 17 European and American case-control studies (11,221 cases and 16,168 controls). They reported a greater than multiplicative joint effect between ever-use of both tobacco and alcohol: for oral, pharyngeal, and laryngeal cancers combined, the OR for the highest rate of tobacco use (>20 cigarettes/day) in alcohol abstainers was 4.2 (95% CI: 2.4-7.1), the OR for the highest rate of alcohol use (≥3 drinks/day) in never-smokers was 1.9 (95% CI: 1.3-2.9), and the OR for the highest category of use for both alcohol and tobacco was 14.2 (95% CI: 8.3-24.4) (49).

Population attributable risk. The INHANCE Consortium study estimated that the population attributable risk for tobacco or alcohol at 72% (95% CI: 61-79%), of which 4% was due to alcohol alone, 33% to tobacco alone, and 35% to use of both alcohol and tobacco. Interestingly, the total population attributable risk differed by anatomic subsite of the cancer (64% for oral cavity, 72% for pharynx, 89% for larynx), by sex (74% for men, 57% for women), by age (33% for cases <45 years, 73% for cases >60 years), and by

geographic region (84% in Europe, 51% in North America, 83% in Latin America) (49). The researchers noted that a substantial fraction of head and neck cancers in certain sub-groups cannot be attributed to tobacco and alcohol use; in particular, their use does not explain one-third of oral cavity cancers, more than 40% of head and neck cancers in women, and two-thirds of young-onset cases.

Mechanisms of carcinogenesis.

Long-term alcohol intake is positively associated with cancer of the aerodigestive tract as well as cancers of the liver and breast. Numerous epidemiologic studies show that ingestion of all types of alcoholic beverages increases cancer risk, suggesting that ethanol is the common factor that causes this effect. The association with cancer strongly suggests that prolonged alcohol consumption somehow results in DNA mutations. Because there is no evidence that alcohol in and of itself is carcinogenic (55), mutagenic, or clastogenic (50), the mechanism must be indirect. Much persuasive research indicates that the primary means by which alcohol increases risk of head and neck cancer, as well as esophageal cancer, is related to the amount of acetaldehyde produced in the metabolism of alcohol (60-62). Other byproducts of alcohol metabolism, including reactive oxygen species (ROS), are believed to act as additional components. In other organs where alcohol causes cancer, such as liver, colon, rectum, and breast, other mechanisms may become more important, such as changes in methionine and folate metabolism, repression of DNA methylation and DNA repair, changes in retinoid metabolism, suppressed immune function and induction of angiogenesis (61, 63-66).

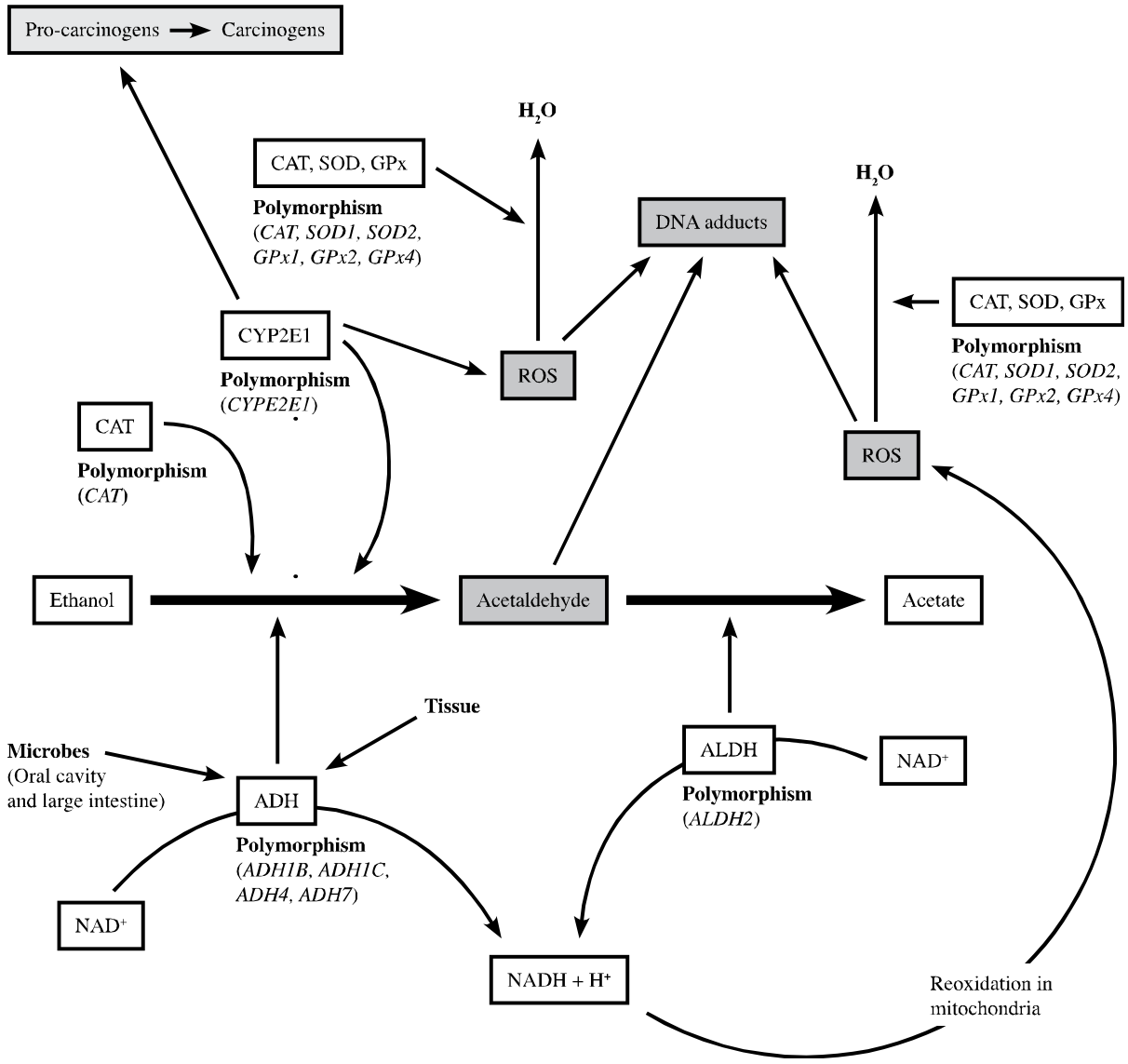
There are three paths by which ethanol is metabolized to acetaldehyde (Figure 3): (1) alcohol dehydrogenase (ADH), (2) cytochrome P450 2E1 (CYP2E1), and (3) catalase (CAT); afterward, acetaldehyde is converted to acetate by aldehyde dehydrogenase (ALDH). In addition to producing acetaldehyde, ADH-mediated ethanol catabolism generates reduced nicotinamide adenine dinucleotide (NADH), which is re-oxidized to NAD^+

in the mitochondria and in the process produces reactive oxygen species (ROS). ADHs are ubiquitous in bacteria and yeasts, so that these UADT organisms are also a significant source of acetaldehyde (50, 61) in the saliva in addition to that produced endogenously in humans. In fact, such organisms can be significant sources of local oxidation of ethanol and production of acetaldehyde in the lower gastrointestinal tract or wherever microbial overgrowth occurs (55). Several human isoforms of ADH and ALDH are genetically polymorphic, which can cause the rate of production and/or oxidation of acetaldehyde to differ between individuals. Likewise, CYP2E1-mediated ethanol catabolism generates a large volume of reactive oxygen species (ROS). *CYP2E1* is also polymorphic with SNPs that affect enzyme activity, and most importantly, chronic ethanol exposure can induce it to produce functional enzyme product even in tissues where it is not constitutively expressed, such as UADT mucosal cells (67). Increased CYP2E1 activity leads not only to increased generation of damaging ROS, which can cause DNA breaks and oxidative base damage, but also to activation of other environmental pro-carcinogens present in tobacco smoke and certain foods that require CYP2E1 to be converted into carcinogens. Ethanol oxidation by CAT appears to be of quantitatively lesser importance, except in the fasted state, in diabetics, and persons eating a high-fat diet (68), although it has recently been shown to be inducible in the presence of alcohol when methyl donor supplements are added to the diet (69).

Acetaldehyde. The first metabolite of ethanol is acetaldehyde, a highly reactive compound that can react with amino groups of proteins resulting in acetaldehyde-protein adducts. Acetaldehyde from alcohol beverages, along with ethanol in such beverages, has been classified by IARC as a Group 1 carcinogen (70). Acetaldehyde can be present in the alcoholic beverage itself (especially in dark liquors) besides being formed *in vivo* from ethanol metabolism (71). Acetaldehyde has been shown to cause laryngeal cancer in inhalation experiments and other UADT tumors in ingestion experiments with animals.

Based on epidemiological findings as well as on research with cell lines *in vitro*, experimental animals (55, 72), and human subjects, acetaldehyde is considered a mutagen (73) that can directly react with DNA to form adducts.

Figure 3. Schematic of ethanol metabolism and its role in carcinogenesis



Ethanol is converted to acetaldehyde by alcohol dehydrogenases (ADH),cytochrome P450 2E1 (CYP2E1), and catalase (CAT). Acetaldehyde is converted to acetate by aldehyde dehydrogenases (ALDH). Reactive oxygen species (ROS) produced during these reactions can genetically damage cells by forming DNA adducts. See text for detailed discussion.
(modified from Seitz & Stickel, 2007 (60))

One highly mutagenic adduct of acetaldehyde that has been studied is 1,N^2-propano-2'-deoxyguanosine (PdG). PdG adducts can exist in either a closed-ring or an open chain form, which may continuously interconvert during replication and transcription. The closed-ring form is more likely to exist in conjunction with single-stranded DNA where it is likely to be mutagenic because it interferes with replication and transcription. The open chain form is more likely to occur in a double-stranded DNA lesion, where the free aldehyde group permits formation of DNA-protein or DNA-DNA cross-links. Because acetaldehyde has been shown to increase chromosomal abnormalities and sister chromatid exchanges in human cells *in vitro*, and is also capable of forming DNA interstrand cross-links (ICLs) that are found in white blood cells of alcoholics but not controls, it has been suggested that ICLs may be the mechanism by which acetaldehyde causes chromosomal aberrations (73).

Much data on acetaldehyde-DNA reactions comes from *in vitro* studies that have provided support for the hypothesis that alcohol-related UADT cancer risk is mediated by acetaldehyde. How this is related to what actually happens *in vivo* has been investigated in a number of studies. Acetaldehyde-fed rats developed severe hyper-regeneration of the upper gastrointestinal mucosa (62), very similar to structural changes that occur after chronic alcohol consumption in humans (74). However, these changes only occurred if the animals had intact salivary glands; after removal of the salivary glands the proliferation disappeared (74), thus supporting the theory that salivary acetaldehyde supports carcinogenesis. Besides being produced by ADH in epithelial cells, and by ADH in the liver where it is released into the bloodstream, acetaldehyde is also produced by oral bacteria. After ingestion of a moderate dose of alcohol, acetaldehyde can be detected in the saliva of healthy volunteers, where the levels are ten to twenty times higher than those in systemic blood, even at higher alcohol intake (61). Interestingly, salivary levels of acetaldehyde can be greatly reduced by using an antiseptic mouthwash before alcohol intake, thus

emphasizing the important role of oral bacteria in acetaldehyde production (61). It has been demonstrated that alcoholics with oropharyngeal cancer who also smoke have very high salivary acetaldehyde concentrations, which may be caused by a smoking-associated transition in bacterial flora from primarily gram-negative to gram-positive species (67). As a general rule, gram positive bacteria are capable of much higher acetaldehyde production than gram negative bacteria. In addition, *Candida albicans*, also frequently present in smokers, converts alcohol into acetaldehyde. Poor oral hygiene, which is common in heavy drinkers, is associated with bacterial overgrowth, periodontal disease, caries, and increased salivary acetaldehyde concentrations (75). Acetaldehyde in saliva would have to enter cells in the basal (regenerative) layer of the epithelium of the aerodigestive tract in order to initiate carcinogenesis. Research in rats that were fed alcohol over 6 months found enlarged nuclei in basal cells in the oral mucosa along with an increased proportion of cells in the S-phase and a reduction of the epithelial thickness indicating mucosal atrophy and hyperproliferation, as compared to controls (76). Finally, acetaldehyde would have to enter cell nuclei in order to damage nuclear DNA; a 1995 study (77) used a very sensitive ^{32}P-postlabeling method to show that N^2-ethyl-deoxyguanosine (a DNA adduct formed from acetaldehyde) becomes detectable in nuclear DNA of liver cells in mice treated for 6 weeks with 10% alcohol in their drinking water. The adduct was not detected in control mice that were not given alcohol. This demonstrates that acetaldehyde can enter cell nuclei *in vivo* to cause DNA lesions.

Reactive oxygen species (ROS). Free radicals are atoms, molecules, or compounds that are chemically unstable due to their distribution of electrons; such entities react easily with other chemicals to produce stable compounds. A ROS is any free radical with at least one oxygen atom in its structure, mainly including superoxide anion radical ($O_2^{\cdot -}$), hydroxyl radical ($^{\cdot}OH$), and peroxide ($O_2^{=}$). Peroxide normally exists in the cell as hydrogen peroxide (H_2O_2) and is considered a precursor of $^{\cdot}OH$. Superoxide, peroxide, and

the hydroxyl radical are considered the primary ROS. Reactions of molecular oxygen that produce ROS along the pathway to water can be summarized as follows (78):

$$O_2 \xrightarrow{e^-} O_2^{-\cdot} \xrightarrow{e^-} O_2^{=} \xrightarrow{2H^+} H_2O_2 \xrightarrow{e^-,\ H^+} H_2O + {}^{\cdot}OH \xrightarrow{e^-,\ H^+} H_2O$$

molecular oxygen → superoxide → peroxide → hydrogen peroxide → water + hydroxyl radical → water

(with enzymes, cofactors, and electrons and protons required along the pathway)

ROS are produced from endogenous sources as by-products of usual and essential metabolic reactions, primarily mitochondrial energy production and detoxification reactions of the cytochrome P-450 enzyme system (which includes CYP2E1 and others). It has been estimated that only about 2-3% of the molecular oxygen consumed by the mitochondrial respiratory process is converted to ROS (78). At low to moderate concentrations ROS play an essential role, along with reactive nitrogen species, in cellular defense against infectious agents, functioning of many cell signaling pathways, and induction of cell division (79). In the amounts normally produced endogenously, several classes of cellular enzymes (described below) are able to chemically reduce most of the ROS to a stable, non-reactive form before they cause oxidative damage (80).

In addition to endogenous sources, ROS can also come from exogenous sources such as cigarette smoke, excess alcohol consumption, and ionizing radiation. When ROS levels increase dramatically during occasions of environmental stress, the resulting oxidative stress can overwhelm the cellular defenses that would normally remove the ROS or repair ROS damage, resulting in significant damage to DNA and other cell structures. ROS can (1) cause permanent structural changes to DNA such as base-pair mutations, deletions, insertions, rearrangements, and amplification; (2) initiate lipid peroxidation; (3) activate cytoplasmic and nuclear signal transduction pathways; and (4) modulate activity of stress proteins and genes that regulate genes related to growth, differentiation, and cell death (80).

When human tissues are subjected to high concentrations of alcohol for a week or more, CYP2E1 is induced (via protein stabilization rather than transcription or translation (81)) and becomes active in processing ethanol to acetaldehyde. These CYP2E1-catalyzed reactions produce large amounts of ROS. Many studies of the effects of administering ethanol to CYP2E1-knockout mice (i.e., mice without CYP2E1 protein) have shown that ethanol-induced oxidative stress and subsequent liver damage is completely prevented in these mice. CYP2E1 also activates many other toxicologically important chemicals besides ethanol, including carbon tetrachloride, acetaminophen, benzene and many halogenated substrates. Toxicity of all these compounds is enhanced after induction of CYP2E1 and reduced by inhibitors of CYP2E1. CYP2E1 is normally a minor pathway of ethanol oxidation, responsible for less than 10% of ethanol conversion to acetaldehyde in the liver. However, this percentage is increased after chronic ethanol exposure induces CYP2E1 protein stabilization (81).

Cells normally protect themselves from ROS damage by three systems: those that suppress the generation of ROS in the first place; those that scavenge and remove ROS after they have been created but before they have damaged the cell; and those that repair damage caused by ROS. The systems that remove ROS after they have been created include the enzymes superoxide dismutase (SOD), glutathione peroxidase (GPx), and catalase (CAT) (78, 82, 83), which remove superoxide and hydrogen peroxide. (See below for detailed descriptions of these enzymes.) Levels of ROS in blood lymphocytes of oral and laryngeal cancer patients, as well as SOD and catalase activities, have been shown in small studies to be altered compared to healthy controls. In addition, SOD and catalase levels in cancerous oral and laryngeal tissues were shown to be altered compared to nearby non-cancerous tissues as well as tissues from control subjects. Some effects appeared to increase along with stage. Direction of effect was inconsistent between studies, which may have been partly due to the low number of subjects as well as differences in measurement

protocols. (84-90) However, these studies do provide tantalizing evidence that anti-oxidant activities are altered during carcinogenesis in the aerodigestive tract.

Summary. Epithelial tissues of the upper aerodigestive tract appear to express ADHs but have relatively low ALDH2 activity, which may make them more susceptible to toxicity mediated by acetaldehyde (either produced locally or from other sources such as saliva or microbes). Chronic drinking of alcoholic beverages induces CYP2E1 metabolism, most prominently in the liver but also at lower levels in the oropharynx and nasal mucosa, although the latter is not well characterized in humans. The relationship between oxidative-stress-induced DNA damage and drinking of alcoholic beverages has not been well established. (51)

Although the exact mechanisms that explain the carcinogenic effects of alcohol are not fully characterized, the direct and well-documented DNA damaging effects of acetaldehyde suggest that the enzymes controlling this reaction, primarily ADH, ALDH and CYP2E1, are excellent candidates for investigation as susceptibility factors for SCCHN. Because DNA-damaging ROS are also produced by these enzymes as by-products of alcohol metabolism, polymorphisms in the genes coding for SOD, GPx, and CAT are also of great interest, because these enzymes are the primary de-activators of ROS, and because there is evidence that their activity is altered in red blood cells and tumor tissue from people with oral and laryngeal cancers. Polymorphisms in any of these genes may affect levels of enzyme activity in UADT epithelial tissues, especially in conjunction with exposures such as alcohol and tobacco, which in turn may affect carcinogenesis and subsequent survival.

1.3.2.3 Human papillomavirus

HPV-associated SCCHNs occur more frequently in younger, male patients, and most often in the oropharynx. In fact, up to 70% of SCCHNs occurring in the oropharynx are associated with the presence of HPV DNA. The most common type is HPV 16, the same

type associated with cervical cancers. These tumors are generally more poorly differentiated, but, perhaps surprisingly, the patients have a better prognosis than those patients whose tumors are not associated with HPV (5, 21). HPV 16-related tumors are not associated with chronic alcohol or tobacco exposure (91).

1.3.3 Other potential factors – nutrition and dental

The association between diet and risk of SCCHN is among the strongest for any cancer. Many studies have found strong, consistent inverse associations between risk of SCCHN and fruit and vegetable consumption that are independent of the effects of alcohol and tobacco (92). In the IARC multinational case-control study of 1670 oral cancer cases, ever-drinkers with the highest quartile of fruit and vegetable intake experienced reduced cancer risk (OR=0.4, 95% CI=0.3-0.6) compared to those in the lowest quartile (93) . This protective effect was not observed among never-drinkers (OR=1.0; 95%CI=0.6-1.6). The authors conclude that vegetables and fruits may modify the carcinogenic effects of alcohol (as well as tobacco). The benefits of a diet high in fruits and vegetables are apparent regardless of the location of tumors in the oral cavity and oropharynx, age (stratified at 60 years), and sex (92). Most studies suggest that diet is associated with risk of oral cancer only in the presence of at least some level of alcohol and tobacco consumption. In a related vein, studies examining the influence of BMI on risk of oral and oropharyngeal tumors suggest that low BMI is not an independent risk factor for SCCHN, but is probably a biomarker of chronic nutritional deficiencies, whether they be secondary to alcohol and tobacco use as they are in developed countries, or whether they be due to dietary insufficiency irrespective of alcohol and tobacco use as they are in less developed countries (92).

Poor oral health is associated with multiple measures of oral hygiene, including frequency of toothbrushing and visits to a dental care provider. Clinical findings such as

mucosal irritation, cavities, and tartar are also associated with a two- to four-fold increased risk, but the strongest, most consistent risk factor is presence/absence of tooth loss as well as number of teeth lost. It appears most likely that poor oral health is an independent risk factor for SCCHN, probably due to type and number of oral bacteria, for which tooth loss serves as a surrogate measure. (92)

1.4 EPIDEMIOLOGIC STUDIES OF GENETIC POLYMORPHISMS AND CANCER INCIDENCE

1.4.1 Polymorphisms of alcohol metabolism genes

1.4.1.1 Alcohol dehydrogenases *(ADH)*

Human ADH is a dimeric protein consisting of two subunits with a molecular weight of 40kD each. It is actually a complex enzyme family and, in humans, five classes are categorized based on structural and kinetic properties. At least seven different genetic loci code for human ADH subunits that combine with other members in the same class as dimers, resulting in more than 20 known ADH isozymes. Such variation is unique to humans. The *ADH1A*, *ADH1B*, and *ADH1C* genes have several variants with different enzymatic activity that are distributed differently across racial groups (Table 4). Such variation was predicted to increase ethanol metabolic rates if the more active forms of the enzyme were present, but this has been hard to demonstrate because any one isozyme constitutes only a fraction of the total alcohol oxidizing capacity in any given tissue and because alcohol elimination rates are variable even among people with the same *ADH* genotype, or even twins. Increased blood levels of acetaldehyde in people with higher activity ADHs has only been demonstrated in those individuals who also have inactive ALDH2 (see below) (55). The ADH isozymes with high K_m for ethanol, for example β3, π, and σ, are predicted to be more active when blood ethanol concentrations are high or in tissues of the UADT that are directly exposed to ethanol in alcoholic drinks, though this

hasn't been tested directly in humans due to ethical concerns (55). The ADH family of enzymes metabolizes a wide variety of substrates besides ethanol, including retinol, other aliphatic alcohols, hydroxysteroids, and lipid peroxidation products. The liver expresses the highest levels of ADH classes I, II, and III, but many other tissues also express ADH, where it may play a role in the toxicity or carcinogenicity of alcohol and its catabolites. Little detail is known about expression of ADH outside the liver, except for gastric ADH, which is reduced with age, in women, and with heavy drinking according to some investigators (55). These variants have been shown to influence drinking levels and therefore the risk of developing alcohol abuse or addiction, and in fact these associations are the strongest and most widely replicated of any gene with risk of alcoholism (94). *ADH* expression can be induced at the mRNA level.

The variant allele *ADH1B*2* encodes a superactive subunit of ADH1B, whose homodimers have 40 times higher maximum velocity than heterodimers ADH1B*1/*2 (95). The enzyme encoded by *ADH1B*1/*1* has only 1% or less of the oxidation capability of enzymes encoded by *ADH1B*1/*2* and *ADH1B*2/*2* (96). *ADH1B*1* is the predominant allele in Europeans and Africans, with *ADH1B*2* frequency not exceeding 15% (63, 97), but *ADH1B*2* is common in Asian populations where its frequency varies from 10% to 90%.

Most studies of *ADH1B* polymorphisms and alcohol exposure have examined the risks of esophageal cancer in Asians. Surprisingly, those studies found that significantly increased risk was associated with the *ADH1B*1* (slow) allele in moderate to heavy drinkers, compared to moderate to heavy drinkers with the *ADH1B*2* (fast) allele (98-102). Three of these studies reported 45% to 65% increased risk in those with the **1/*1* genotype compared to other genotypes (OR=1.56, 95% CI=1.01-2.39 (98); OR=1.65, 95% CI=1.02-2.68 (101); OR=1.45, 95% CI=0.97-2.16 (100)); one reported that the **1/*1* genotype quadrupled cancer risk compared to other genotypes (OR=4.11; 95% CI=2.08-8.12 (102));

and one reported that the **1/*1* genotype quadrupled the cancer risk of **2/*2* (OR=3.99; 95% CI=2.13-7.48 (99)).

Two recent studies in European countries of UADT cancer, and one recent study in Brazil of SCCHN, found the reverse to be true: a protective effect (50% reduced risk) of the slow allele in drinkers compared with drinkers homozygous for the fast allele (103, 104) (Table 5). Two recent Japanese studies reported conflicting results: One reported a doubling in risk due to the fast allele (105), and the other reported a similar effect for the slow allele (106) (Table 5). Most studies of ADH1B *1 and *2 alleles report both a main effect of the gene on head and neck cancer, along with significant interaction with alcohol exposure (Tables 5 and 6); magnitude of interaction varied along the complete spectrum from greater than additive but less than multiplicative, to multiplicative, to greater than two times the expected multiplicative odds ratio. However, a 2003 study in Caucasian Germans found no significant association between *ADH1B* genotype and risk of laryngeal cancer (107).

The results for *ADH1B* polymorphisms don't consistently support the "acetaldehyde hypothesis," which speculates that the much-less-active *ADH1B*1* allele should decrease the risk of cancer in drinkers, because they would have less exposure to the carcinogenic and mutagenic acetaldehyde product. Some researchers have hypothesized that the higher risk for *ADH1B*1/*1* homozygotes may be due to an absence of alcohol flushing (which is caused by acetaldehyde & is an uncomfortable sensation) so that the person is more likely to become a heavy drinker.

The *ADH1C* gene has two alleles in most populations, *ADH1C*1* and *ADH1C*2*. *ADH1C*1* isoenzymes metabolize ethanol into acetaldehyde 2.5 times faster than the *ADH1C*2* allele. In European Caucasians, the frequency of the ADH1C*1 (fast) allele is about 50%, but in Africans the frequency is 75-90% and in Asians even higher (63, 97). Individuals homozygous for the *ADH1C*1* fast allele were found to have elevated salivary acetaldehyde levels, which is hypothesized to cause increased UADT cancer risk (108).

Eight small studies (each n_{cases}<500), most conducted prior to 2006, mostly reported no main effect of *ADH1C* on head and neck cancer, although a few found an effect in heavy drinkers. (Table 5) Since 2006, three of four large studies in European Caucasians and Latin Americans, Brazilian Caucasians and Blacks and Mullatos, and Japanese populations (103-105, 109) reported that genotypes containing the *2 slow allele increase the risk of cancer by 20 to 50%, and two reported greater risks in heavy drinkers (105, 109). Interaction with alcohol exposure was reported as greater than additive in the Japanese population and more than two times the expected multiplicative value in whites and African-Americans (Table 6). Overall, recent large studies on *ADH1C* polymorphisms and head and neck cancer risk have produced consistent evidence of increased risk associated with the slow allele, but a smaller effect than that of the *ADH1B* polymorphisms previously described.

The largest and most comprehensive epidemiologic study to date on the relationships between multiple *ADH* genes and UADT cancers combined participant-level data from five smaller studies conducted in Europe and Latin America (104) (Tables 5 & 6: Hashibe et al. 2008). It found that UADT cancers are associated with rs1229984 *ADH1B*2* (fast) allele (OR=0.56; 95% CI=0.47-0.66) and rs1573496 variant G allele in *ADH7* (OR=0.68; 95% CI=0.60-0.78), but not with rs1984362 in *ADH4* (104). It estimated greater than additive but less than multiplicative interaction. A recent case-control study in Japan (Tables 5 & 6: Oze et al. 2009) found significant independent associations of rs3737482 in *ADH7* and rs4148887 in *ADH4* on risk of UADT cancers, and possibly a small effect of rs1229984 in *ADH1C* that was not completely explained by linkage with *ADH1B*. It reported a suggestive increase in risk with increasing alcohol use related to the *ADH4* SNP (105), and estimated greater than additive interaction.

1.4.1.2 Aldehyde dehydrogenase *(ALDH)*

Two primary ALDH enzymes metabolize the acetaldehyde produced during ethanol oxidation. ALDH1A1 (old name ALDH1) is found in the cytosol, while ALDH2 is found in the mitochondria (Table 4). It is estimated that about 30% of total liver ALDH activity is from ALDH2 with the rest contributed by ALDH1A1, ALDH9A1 and possibly ALDH1B1 (55). Both *ALDH1A1* and *ALDH2* have similar structure with 13 exons, and the proteins they encode are 70% identical in sequence and have very similar structure. *ALDH* transcripts of all four types (Table 4) are found in almost every tissue of the body (55). Low levels of *ALDH* mRNAs have been found in placenta, brain, and pancreas; because these are target organs for alcoholic pathology, these data are consistent with the hypothesis that the presence of ALDHs is protective against acetaldehyde toxicity (55). A polymorphism of *ALDH2* exists that has no enzymatic activity. *ALDH2*1* is the active allele, and *ALDH2*2* is the inactive allele. *ALDH2*2/*2* homozygotes have no enzyme activity, while *ALDH2*1/*2* heterozygotes have 6% residual activity. The *ALDH2*2* mutant allele is prevalent in Asians up to a frequency as high as 40%, but its frequency in European and African populations is less than 5%. In an animal model mimicking the human *ALDH2*2/*2* genotype, knock-out mice lacking ALDH2 activity have reduced alcohol preference, and when forcibly exposed to higher doses of ethanol, they exhibit elevated acetaldehyde in blood, liver, and brain (110). These mice have been used for toxicological studies of alcohol and acetaldehyde (55).

For cancer of the UADT, many studies have reported increased risk of esophageal cancer in Asians with the *ALDH2*2* (inactive) allele who were moderate or heavy drinkers (98-102, 111, 112). Two small studies of cancer (each n_{cases}<250) of the UADT, oral cavity, or oropharynx in Asian populations found significant interaction between *ALDH2**1/*2 genotype and heavy or moderate/heavy alcohol intake (106, 113); in contrast, two other studies in Asians reported no interaction (111, 114). Only the largest of these four studies

(n_{cases}=239) reported a significant main effect of the gene – i.e., a 66% increase in UADT cancer risk associated with the null phenotype. These results are consistent with the hypothesis that reduced ALDH2 activity exposes the body to more mutagenic acetaldehyde. In addition, people with ALDH2 deficiency have higher acetaldehyde levels in their serum and saliva than occur in people with the wild type genotype, and higher frequency of acetaldehyde adducts, sister chromatid exchanges, and micronuclei have been reported in Asian drinkers with the heterozygous genotype compared with the wild type (60).

In central and eastern Europeans, in whom the *ALDH2**2 inactive mutant allele is almost completely absent, one study of three other, previously unstudied, variants in *ALDH2* (103) reported 30% increased risk of UADT cancers as a main effect of the variant allele, and even higher risk in medium to heavy drinkers (defined as more than 2 drinks per week) (Tables 5&6: Hashibe et al. 2006 (103)). The GG genotype of rs886205 in the 5' untranslated region was associated with increased UADT risk compared to the AA genotype (OR=1.29; CI=1.03-1.60), as was the CC genotype of rs440 in intron 6 compared to the TT genotype (OR=1.32; 95% CI=1.06-1.65), and the CC genotype of rs441 in intron 6 compared to the TT genotype (OR=1.33; 95% CI=1.07-1.66). In medium/heavy drinkers, these risks were increased to ORs of 4.38 (95% CI=1.32-14.53), 5.79 (95% CI=1.49-22.52), and 5.79 (95% CI=1.49-22.49), respectively. Effect measure modification was estimated to be greater than multiplicative. Risk was higher for esophageal cancer than for oral cavity, pharynx, and larynx cancers. A second large 2009 study of European Caucasians (115) found no main effect of two additional SNPs in *ALDH2*, rs886205 in the 5'UTR and rs4646777 in intron 8, and no interactive effect with alcohol (Tables 5&6: Canova et al. 2009 (115)), although a 2006 study of 436 cases with alcoholism and 365 controls in the U.S. (116) reported that these two SNPs are associated with alcoholism in both whites and African-Americans.

Table 4. Human ADH and ALDH isozymes (55, 68, 94, 117)

Class	Official gene name (old name)	Allele (old name)	SNP (nucleotide)	Chromo-somal location	Protein subunit	Amino acid differences between alleles	K_m sub-strate (mM)	V_{max} (min^{-1})	Ethnic/ national distribution
Human alcohol dehydrogenase (ADH) isozymes									
I	*ADH1A (ADH1)*	*ADH1A*		4q22	α		4	54	Europe, Africa
	ADH1B (ADH2)	*ADH1B*1 (ADH2*1)*	rs1229984 (A) rs2066702 (C)	4q22	β1	Arg48, Arg370	0.05	9	Europe, Africa
		*ADH1B*2 (ADH2*2)*	rs1229984 (G) rs2066702 (C)		β2	His48, Arg370	0.9	400	Asia
		*ADH1B*3 (ADH2*3)*	rs1229984 (A) rs2066702 (T)		β3	Arg48, Cys370	34	300	Africa, Native American
	ADH1C (ADH3)	*ADH1C*1 (ADH3*1)*	rs1693482 (G) rs698 (A)	4q22	γ1	Arg272, Ile350	1.0	87	All
		*ADH1C*2 (ADH3*2)*	rs1693482 (A) rs698 (G)		γ2	Gln272, Val350	0.63	35	Europe
		*ADH1C*3*	rs35719513 (A)			Thr352*			Native
II	*ADH4*	*ADH4*1*	rs1126671 (A)	4q21-25	π	Ile309	34	40	All
		*ADH4*2*	rs1126671 (G)			Val309			Sweden
III	*ADH5*			4q21-25	X		1000		All
V	*ADH6*			4q21-25	ADH6		30	?	All
IV	*ADH7*			4q23-24	σ,μ		20	1510	All
Human aldehyde dehydrogenase (ALDH) isozymes									
	ALDH1A1 (ALDH1)			9q21.13			30		All
	ALDH2	*ALDH2*1*	rs671 (G)	12q24.2		Glu504	1		All
		*ALDH2*2*	rs671 (A)			Lys504			Asia
			rs886205 (G)						All
			rs886205 (A)						All
			rs440 (C)						All
			rs440 (T)						All
			rs441 (C)						All
			rs441 (T)						All
	ALDH1B1 (ALDH5)	*ALDH1B1*		9p11.1			?		
	ALDH9A1	*ALDH9A1*		9q21.13			30		All

*ADH1C*352Thr has been found in Native Americans as an additional variation on chromosomes that have the Val350 characteristic of ADH1C*2. The protein has not been isolated for study.
NOTE: K_m indicates the concentration of substrate at which the enzyme works at 50% capacity. V_{max} (turnover) indicates how many molecules of product are formed by the enzyme in 1 minute at saturating substrate concentrations.

1.4.1.3 Cytochrome P450-2E1 *(CYP2E1)*

Alcohol can be metabolized by cytochrome P450s, with CYP2E1 being the most active in this regard. CYP2E1 is associated with microsomes in the endoplasmic reticulum, where it reduces molecular oxygen to water as ethanol is oxidized to acetaldehyde. Its K_m for ethanol is about 10mM (fairly high), and CYP2E1 probably assumes a larger role in metabolizing ethanol when blood alcohol levels are high. When metabolizing alcohol,

CYP2E1 generates a large number of ROS including hydroxyl radicals, superoxide anion, hydrogen peroxide, and hydroxyethyl radicals; therefore it is a major source of oxidative stress. (55) CYP2E1 enzyme production is induced by chronic drinking and also by fasting, diabetes, and a high-fat diet. It is expressed at highest concentrations in the liver but is also expressed in numerous other tissues, including esophageal and nasal mucosa among many others. One animal study found that after chronic alcohol feeding, CYP2E1 was found in squamous epithelial cells of the cheek mucosa, tongue, esophagus, and fore-stomach (118). Some individuals experience a very low extent of induction of CYP2E1 after alcohol ingestion, whereas others show a high extent (119); thus, variation in the extent of induction may modulate alcohol-associated carcinogenesis (67), although it is unknown whether these differences are affected by *CYP2E1* polymorphisms. A recent study (69) found that, in mice, dietary supplementation with methyl donor supplements (which are associated with reduction in alcohol-induced liver injury) led to a 35% increase in blood alcohol elimination rate. While the supplements caused no difference in alcohol metabolism in the stomach, there was a tremendous effect in the liver: the catalase-dependent pathway of alcohol metabolism was induced, at the same time that CYP2E1 induction was blunted. This research is consistent with the hypothesis that CYP2E1 alcohol metabolism is responsible for much of the liver damage associated with alcoholism. Four major polymorphisms of *CYP2E1* are the best characterized, but there exist numerous un- or little-studied polymorphisms:

1. the wild-type $RsaI^+$ (c1) and the variant $RsaI^-$ (c2) located in the 5'-flanking region of the *CYP2E1* gene in a region that interacts with HNF1 (also known as *CYP2E1**5A and *5B, RsaI/PstI RFLP, rs2031920). The c2 (T) variant is associated with a 10 times higher transcriptional activity, elevated protein levels, and increased enzyme activity compared to the c1 (C) allele (120). The c2 variant is present in 2-8% of Caucasians and 25-36% of East Asians. A dozen small studies (Table 5), with number of cases

ranging from 75 to 408, in Brazil, China, India, Europe, the U.S., and Japan, have studied the effects of c2 on risk of head and neck cancer (111, 121-131). Eight of these studies reported no main effect or interaction with alcohol. Three studies, conducted in France, India, and Japan (125, 128, 131) reported significantly increased risk of head and neck cancer (ORs=2.6, 2.4, and 3.4 respectively), and the French and Indian studies also reported interactive effects with alcohol. (See Table 6 re: interactions.) The Indian study (125) reported greater than additive but less than multiplicative interaction, and the French study (128) reported more than 2x multiplicative interaction. A study conducted in Brazil (124) reported no main effect of the c2 polymorphism but did report greater than 2x multiplicative interaction with alcohol exposure. Yang et al. reported no main effect of this polymorphism on risk of esophageal cancer in an Asian population, and no interactive effect with alcohol (100).

Several studies reported increased risk for gastric and rectal cancers. A meta-analysis published in 2007 reported a possible increased risk of gastric cancer in Asians homozygous for the c2 allele (132). A Chinese study (133) found that the *CYP2E1* c2/c2 genotype increases susceptibility to rectal (but not colon) cancer (OR=1.64; 95% CI=1.12-2.41) compared to c1 allele carriers, and gene-environment interactions between the *CYP2E1* polymorphism and smoking or alcohol drinking were found to exist for colorectal neoplasia in general.

2. the DraI polymorphism located in intron 6 (also known as *CYP2E1**6, DraI RFLP, T7632A SNP, and rs6413432), for which the variant is found in 10% of Caucasians and about half of East Asians. A study in the Caucasian French population (128) reported a two-fold increased risk of oral cavity/ pharyngeal cancer (OR=2.0; 95% CI=1.0-3.9) and laryngeal cancer (OR=1.8; 95% CI=1.0-3.5) in carriers of the T variant allele compared to AA individuals. (See Tables 5&6.) The highest risk of oral cavity/pharyngeal cancer was observed among the heaviest drinkers (oral

cavity/pharyngeal cancer: OR=5.8, 95% CI=1.9-18.2; laryngeal cancer: OR=4.6, 95% CI=1.5-14.1), with the interactive effect being greater than 2x multiplicative. More precise estimates were not possible due to the rarity of the variant allele in the Caucasian French population. A 2009 Indian study (125) and a 2006 Japanese study (131) reported increased risk of head and neck cancer with the T risk allele (OR=1.55, 2.28 respectively), but no interactive effects with alcohol. Two studies reported no main or interactive effects of DraI on risk of head and neck cancer (121, 126).

3. the -71G>T polymorphism located 5' of the gene, which occurs in about 10% of Caucasians (also known as *CYP2E1**7B, DdeI RFLP, G-71T SNP, and rs6413420). There are no studies of this polymorphism and risk of head and neck cancer.
4. a 96-bp insertion polymorphism located in the regulatory region, which is found in 15% of Asians and 2% of Caucasians. There are no studies of this polymorphism and risk of head and neck cancer.

In summary, studies investigating the effects of these *CYP2E1* polymorphisms on risk of alcoholic complications or cancer risk have small study populations and have produced inconsistent results (55).

1.4.2 Polymorphisms of oxidative stress metabolism genes

1.4.2.1 Catalase *(CAT)*

CAT is expressed in almost all tissues, where it is primarily found in peroxisomes, and also by colonic microorganisms in the lower gastrointestinal tract, where it contributes to formation of acetaldehyde from ethanol. In peroxisomes it primarily catalyzes a reaction between hydrogen peroxide molecules, resulting in the formation of water and O_2, but it can also facilitate the reaction of hydrogen peroxide with hydrogen donors so that the hydrogen peroxide is converted to one molecule of water and the reduced donor (e.g., ethanol)

becomes oxidized. It can be upregulated by oxidative stress (134). In a study using ADH-deficient deermice, ethanol oxidation was highly sensitive to inhibition by a known catalase inhibitor (135). Some Asians do not have active CAT, and a number of SNP polymorphisms in the 5' untranslated region and introns have been reported, but these have not been studied in terms of their influence on enzyme activity or on responses to ethanol (55). A common C/T polymorphism has been identified in the promoter region, and the variant affects transcriptional activity and CAT enzyme levels in red blood cells (136). Associations of *CAT* polymorphisms with risk of SCCHN have not been studied. Polymorphisms of *CAT* were not associated with lung cancer risk in a Chinese population (137), non-Hodgkin's lymphoma in the United Kingdom (138), or prostate cancer in the United States (139).

1.4.2.2 Superoxide dismutase *(SOD)*

SOD enzymes serve an important antioxidant role by catalyzing the conversion of superoxide free radicals into oxygen and hydrogen peroxide. Humans have three forms of SOD, whose genes are located on different chromosomes, and which are active in different locations. The gene for *SOD1*, also known as copper-zinc superoxide dismutase, is located on chromosome 21; its isozyme is a soluble cytoplastmic and intra-membrane mitochondrial protein. The gene for *SOD2*, also known as manganese superoxide dismutase, is located on chromosome 6; its product is active in the mitochondrial matrix. The gene for *SOD3* is located on chromosome 4; its product is a major extracellular antioxidant enzyme expressed in many tissues, especially blood vessels (140, 141). Most studies report no association between *SOD1* polymorphisms and cancer. The only study of *SOD* polymorphisms and risk of head and neck cancer – a 2009 study in European Caucasians (115) – found no association between rs2758346 in *SOD2* and cancer risk (Table 5, last entry). A variant non-synonymous SNP polymorphism in exon 2 of *SOD2* (rs1799725) has been independently associated with 20-30% elevated risk of prostate cancer in Caucasians (Val/Ala versus

Val/Val: OR=1.17; 95% CI=0.97-1.42; Ala/Ala versus Val/Val: OR=1.28; 95% CI=1.03-1.60 (142)). Another study (143) found that white women with the Ala allele were at increased risk of ovarian cancer (OR=2.1; 95% CI=1.1-4.0). And Wang et al. (144) reported that, in a U.S. study population, Ala homozygotes were at increased risk for B-cell lymphomas compared to other genotypes (OR=1.3; 95% CI=1.0-1.6). A 2007 study reported that SNPs in *SOD3* were not associated with prostate cancer in a U.S. population (142).

1.4.2.3 Glutathione peroxidases *(GPx/GPX)*

GPX's are a family of selenium-dependent enzymes that remove hydrogen peroxide. There are four isoenzymes characterized to date. GPX1 was the first identified and is widely expressed in the cytoplasm of human cells, protecting cells against oxidative damage by reducing hydrogen peroxide and other peroxides. The glutathione peroxidase system consists of several components, including the GPx enzyme, glutathione reductase and cofactors glutathione (GSH) and reduced nicotinamide adenine dinucleotide phosphate (NADPH). Associations of *GPx* polymorphisms with risk of SCCHN have not been studied. However, a SNP resulting in a non-synonymous amino acid substitution at codon 198 (Pro198Leu) has been studied for other cancer associations by several different groups. Arsova-Sarafinovska et al. (145) reported an overall protective effect of the variant Leu allele on prostate cancer risk; heterozygous carriers of the variant Leu (T) allele had reduced risk of prostate cancer compared with homozygous wild-type (OR=0.38; 95% CI=0.20-0.75). In a Danish study (146), carriers of the Leu variant were at 1.43-fold higher risk of breast cancer compared with non-carriers (95% CI=1.07-1.92). In a lung cancer study also conducted in Denmark (147), the odds of developing lung cancer were reduced by 40% (OR=0.60; 95% CI=0.35-1.05) when homozygous carriers of the variant allele were compared with the homozygous carriers of the wild type. A Japanese study (148) found that the risk of bladder cancer more than doubled with the Pro/Leu genotype compared with the wild-type Pro/Pro

(OR=2.63; 95% CI=1.45-4.75). However, these results are not consistent in all populations (141). GPX2 (extracellular intestinal enzyme), GPX3 (extracellular and common in plasma), and GPX4 (which has high affinity for lipid hydroperoxides) have not been studied for their association with cancer.

Table 5. Case control studies examining main effects of alcohol and oxidative metabolism genes on risk of SCCHN
(studies examining the same gene polymorphism and effect direction are grouped together and shaded alike)

Authors	Year	Subjects' race # cases/# controls	Gene tested**	SNP, variant & referent groups	OR (95% CI)	Significant at α=0.05 for: Main effect of gene?	Significant at α=0.05 for: Interaction of gene with alcohol exposure?
Garcia et al.(123)	2010	Brazilian Caucasians, Blacks, Mulattos 207/244	*ADH1B*	rs2066702 *3 allele TT+TC (also fast) versus CC	0.60 (0.27-1.32)		
Canova et al.(115)	2009	European Caucasians 1511/1457	*ADH1B*	rs17033 A>G GG versus AA	1.35 (0.46-3.94)		
Hashibe et al.(104)	2008	European Caucasians & Latin Americans 3449/5278	*ADH1B*	rs6413413 TT+TA versus AA	1.07 (0.70-1.63)		
Garcia et al.(123)	2010	Brazilian Caucasians, Blacks, Mulattos 207/244	*ADH1B*	rs1229984 GG+GA (fast) versus AA	0.42 (0.21-0.85)	✓	✓
Oze et al. (105)	2009	Japanese 585/1170	*ADH1B*	rs1229984 GG+GA (fast) versus AA	2.20 (1.46-3.32)	✓	✓
Hashibe et al.(103)	2006	European Caucasians 811/1083	*ADH1B*	rs1229984 AA+GA (fast) versus GG	0.47 (0.32-0.70)	✓	✓
Risch et al.(107)	2003	German Caucasians 245/251	*ADH1B*	rs1229984 GA (fast) versus AA	0.86 (0.41-1.82)		
McKay et al.(149)	2011	Europeans, Latin & North Americans of European descent, African-Americans 8527/11653	*ADH1B*	rs1229984 log-additive model, additional risk for each A allele	0.62 (0.56-0.68)	✓	✓
Hashibe et al.(104)	2008	European Caucasians & Latin Americans 3449/5278	*ADH1B*	rs1229984 AA+GA (slow) versus GG (fast)	0.56 (0.47-0.66)	✓	✓
Asakage et al.(113)	2007	Japanese 96/642	*ADH1B*	rs1229984 AA+AG (slow) versus GG	Not stated; NS		✓
Hiraki et al.(106)	2007	Japanese 239/716	*ADH1B*	rs1229984 AA (slow) versus GG	2.67 (1.51-4.75)	✓	✓
McKay et al.(149)	2011	Europeans, Latin & North Americans of European descent, African-Americans 7890/10938	*ADH1C*	rs698 log-additive model, additional risk from each G allele	1.10 (1.05-1.15)	✓	
Garcia et al.(123)	2010	Brazilian Caucasians, Blacks, Mulattos 207/244	*ADH1C*	rs698 GG+GA (slow) versus AA	1.37 (0.95-1.99)		

Table 5. Case control studies examining main effects of alcohol and oxidative metabolism genes on risk of SCCHN

(studies examining the same gene polymorphism and effect direction are grouped together and shaded alike)

Authors	Year	Subjects' race # cases/# controls	Gene tested**	SNP, variant & referent groups	OR (95% CI)	Significant at α=0.05 for: Main effect of gene?	Significant at α=0.05 for: Interaction of gene with alcohol exposure?
Oze et al. (105)	2009	Japanese 585/1170	*ADH1C*	rs698 each G (slow) allele	1.38 (1.02-1.87)	✓	
Hashibe et al.(104)	2008	European Caucasians & Latin Americans 3449/5278	*ADH1C*	rs698 GG+GA (slow) versus AA	1.16 (1.05-1.29)	✓	
Hashibe et al.(103)	2006	European Caucasians 811/1083	*ADH1C*	rs698 GG (slow) versus AA	1.38 (1.01-1.88)	✓	
Oze et al. (105)	2009	Japanese 585/1170	*ADH1C*	rs1693482 Allelic model (A vs G)	1.37 (1.01-1.85)	✓	
Hashibe et al.(104)	2008	European Caucasians & Latin Americans 3449/5278	*ADH1C*	rs1693482 AA+AG (slow) versus GG	1.20 (1.08-1.32)	✓	
Asakage et al.(113)	2007	Japanese 96/642	*ADH1C*	rs1693482 AA+AG (slow) versus GG	Not stated; NS		
Hashibe et al.(103)	2006	European Caucasians 811/1083	*ADH1C*	rs1693482 AA (slow) versus GG	1.49 (1.08-2.05)	✓	
Peters et al.(109)	2005	EA*, AA* (4%) 521/599	*ADH1C*	*2/*2 (slow) versus *1/*2 & *1/*1	Not stated; NS		✓
Wang et al.(150)	2005	EA*, Other (4%) 348/330	*ADH1C*	*1/*1 (fast) and *1/*2 versus *2/*2	Not stated; NS		
Risch et al.(107)	2003	German Caucasians 245/251	*ADH1C*	*1/*1 (fast) versus *1/*2 and *2/*2	1.06 (0.70-1.62)		
Zavras et al.(151)	2002	Greeks 93/99	*ADH1C*	*1/*1 (fast) versus *2/*2	0.9 (0.3-2.5)		✓
Olshan et al.(152)	2001	EA*, AA* (24%) 182/202	*ADH1C*	*1/*1 (fast) versus *2/*2	0.9 (0.4-1.9)		
Schwartz et al.(153)	2001	EA*, AA* (6%) 333/541	*ADH1C*	(1) *1/*1 (fast) vs *2/*2 (2) *1/*2 vs *2/*2	(1) 1.0 (0.7-1.5) (2) 1.3 (1.0-1.8)	✓	✓
Bouchardy et al.(128)	2000	French Caucasians 250/172	*ADH1C*	*1/*1 (fast) versus *1/*2 and *2/*2	OC/pharynx: 1.4 (0.8-2.3) larynx: 0.7 (0.4-1.3)		
Canova et al.(115)	2009	European Caucasians 1511/1457	*ADH1C*	rs1662058 A>G GG versus AA	1.12 (0.88-1.42)		
Canova et al.(115)	2009	European Caucasians 1511/1457	*ADH1C*	rs2241894 A>G GG versus AA	1.01 (0.72-1.42)		
Oze et al. (105)	2009	Japanese 585/1170	*ADH4*	rs4148887 TT+TC versus CC	1.96 (1.34-2.87)	✓	✓
Hashibe et al.(104)	2008	Europeans & Latin Americans 3449/5278	*ADH4*	rs1984362 TT+TC versus CC	1.03 (0.94-1.14)		
Oze et al. (105)	2009	Japanese 585/1170	*ADH7*	rs3737482 CC+CT versus TT	1.62 (1.18-2.24)	✓	✓
McKay et al.(149)	2011	Europeans, Latin & North Americans of European descent, African-Americans 8545/11657	*ADH7*	rs1573496 log-additive model, additional risk for each C allele	0.74 (0.69-0.80)	✓	
Hashibe et al.(104)	2008	Europeans & Latin Americans 3449/5278	*ADH7*	rs1573496 CC+CG versus GG	0.68 (0.60-0.78)	✓	✓

Table 5. Case control studies examining main effects of alcohol and oxidative metabolism genes on risk of SCCHN

(studies examining the same gene polymorphism and effect direction are grouped together and shaded alike)

Authors	Year	Subjects' race # cases/# controls	Gene tested**	SNP, variant & referent groups	OR (95% CI)	Significant at α=0.05 for: Main effect of gene?	Significant at α=0.05 for: Interaction of gene with alcohol exposure?
Canova et al.(115)	2009	European Caucasians 1511/1457	*ALDH2*	rs886205 (+82A>G) GG versus AA	1.09 (0.72-1.67)		
Hashibe et al.(103)	2006	European Caucasians 811/1083	*ALDH2*	rs886205 (+82A>G) GG+GA versus AA	1.29 (1.03-1.60)	✓	✓
Hiraki et al.(106)	2007	Japanese 239/716	*ALDH2*	rs671 AG (null) versus GG	1.66 (1.20-2.31)	✓	✓
Asakage et al.(113)	2007	Japanese 96/642	*ALDH2*	rs671 AA (null) versus GG	Not stated; NS		✓
Hashimoto et al.(114)	2006	Japanese 192/192	*ALDH2*	rs671 AA+AG (null) versus GG	Not stated; NS		
Katoh et al.(111)	1999	Japanese 92/147	*ALDH2*	rs671 AA+AG (null) versus AA	1.10 (0.61-1.99)		
McKay et al.(149)	2011	Europeans, Latin & North Americans of European descent, African-Americans 8232/11064	*ALDH2*	rs4767364 log-additive model, additional risk for each A allele	1.12 (1.07-1.17)	✓	✓
Canova et al.(115)	2009	European Caucasians 1511/1457	*ALDH2*	rs4646777 G>A AA versus GG	1.08 (0.71-1.63)		
Hashibe et al.(103)	2006	European Caucasians 811/1083	*ALDH2*	rs441 CC+CT versus TT	1.33 (1.07-1.66)	✓	✓
Hashibe et al.(103)	2006	European Caucasians 811/1083	*ALDH2*	rs440 CC+CT versus TT	1.32 (1.06-1.65)	✓	✓
Ruwali et al.(125)	2009	Indians 350/350	*CYP2E1*	rs6413432 (DraI) TT+TA versus AA	1.55 (1.08-2.22)	✓	
Boccia et al.(121)	2008	European Caucasians 210/245	*CYP2E1*	rs6413432 (DraI) TT+TA versus AA	0.87 (0.43-1.76)		
Soya et al.(126)	2008	Indians 408/220	*CYP2E1*	rs6413432 (DraI) TT versus AA	0.94 (0.36-2.46)		
Sugimura et al.(131)	2006	Japanese 122/241	*CYP2E1*	rs6413432 (DraI) TT versus AA	2.28 (1.06-4.91)	✓	
Bouchardy et al.(128)	2000	French Caucasians 250/172	*CYP2E1*	rs6413432 (DraI) TT+TA versus AA	OC/pharynx: 2.0 (1.0-3.9) larynx: 1.8 (1.0-3.5)	✓ ✓	✓ ✓
Garcia et al.(123)	2010	Brazilian Caucasians, Blacks, Mulattos 207/244	*CYP2E1*	rs2031920 (RsaI) TC (fast) versus CC	1.53 (0.76-3.09)		
Tai et al.(127)	2010	Chinese 278/278	*CYP2E1*	rs2031920 (RsaI) TC (fast) versus CC	0.97 (0.66-1.43)		
Olivieri et al.(124)	2009	Brazilian Caucasians, Blacks, Mulattos 153/145	*CYP2E1*	rs2031920 (RsaI) TT (fast) +TC versus CC	Not given		✓
Ruwali et al.(125)	2009	Indian 350/350	*CYP2E1*	rs2031920 (RsaI) TT (fast) +TC versus CC	2.40 (0.98-5.85)	✓	✓
Boccia et al.(121)	2008	European Caucasians 210/245	*CYP2E1*	rs2031920 (RsaI) TT (fast) +TC versus CC	0.72 (0.33-1.63)		
Buch et al.(122)	2008	EA* 197/416	*CYP2E1*	rs2031920 (RsaI) TT (fast) +TC versus CC	Not given		

Table 5. Case control studies examining main effects of alcohol and oxidative metabolism genes on risk of SCCHN

(studies examining the same gene polymorphism and effect direction are grouped together and shaded alike)

Authors	Year	Subjects' race # cases/# controls	Gene tested**	SNP, variant & referent groups	OR (95% CI)	Significant at α=0.05 for: Main effect of gene?	Significant at α=0.05 for: Interaction of gene with alcohol exposure?
Soya et al.(126)	2008	Indians 408/220	*CYP2E1*	rs2031920 (RsaI) TT (fast) +TC versus CC	1.03 (0.36-2.94)		
Sugimura et al.(131)	2006	Japanese 122/241	*CYP2E1*	rs2031920 (RsaI) TT (fast) versus CC	3.38 (1.22-9.36)	✓	
Bouchardy et al.(128)	2000	French Caucasians 250/172	*CYP2E1*	rs2031920 (RsaI) TT (fast) + TC versus CC	OC/pharynx: 2.6 (1.0-6.6) larynx: 1.4 (0.5-4.0)	✓	✓
Katoh et al.(111)	1999	Japanese 92/147	*CYP2E1*	rs2031920 (RsaI) TT (fast) + TC versus CC	1.45 (0.80-2.62)		
Morita et al.(130)	1999	Japanese 145/164	*CYP2E1*	rs2031920 (RsaI) TT (fast) + TC versus CC	Not stated; NS		
Gonzalez et al.(129)	1998	Spanish Caucasians 75/200	*CYP2E1*	rs2031920 (RsaI) TT (fast) + TC versus CC	Not stated; NS		
Soya et al.(126)	2008	Indians 408/220	*CYP2E1*	*1B A1A1 versus A2A2	0.52 (0.16-1.66)		✓
Canova et al.(115)	2009	European Caucasians 1511/1457	*SOD2*	rs2758346 G>A AA versus GG	0.98 (0.78-1.22)		

* AA=African American, EA=European American/white

** This column lists only those genes also being tested in this study (*ADH, ALDH, CYP2E1, SOD, GPX, CAT*)

Table 6. Case control studies examining interaction between alcohol and oxidative metabolism genes and alcohol in production of SCCHN

Authors	Year	Subjects' race	Genes tested**	Evidence of interaction found	Case/control source	# cases/ # controls
McKay et al.(149)	2011	Europeans, Latin & North Americans of European descent, African-Americans	*ADH1B, ADH1C, ADH7, ALDH2*	rs1229984 in *ADH1B* protective in drinkers	15 UADT studies: 11 hospital-based, 4 population-based	8744/11982
Garcia et al.(123)	2010	Brazilian Caucasians, Blacks, Mulattos	*ADH*	rs1229984 in *ADH1B* ↑ HNC risk in heavy alcohol drinkers (>additive, 2x multiplicative)	hospital/clinic	207/244
Tai et al.(127)	2010	Chinese	*CYP2E1*	none	hospital/clinic	278/278
Oze et al. (105)	2009	Japanese	*ADH*	1 in *ADH1B*, 2 SNPs in *ADH4*, 1 in *ADH7* ↑ risk of UADT; greater magnitude in heavy drinkers (>multiplicative)	hospital/clinic	585/1170
Olivieri et al.(124)	2009	Brazilian Caucasians, Blacks, Mulattos	*CYP2E1*	rs2031920/*5B/RsaI ↑risk in alcoholics versus non-alcoholics (2x multiplicative)	hospital/hospital	153/145
Ruwali et al.(125)	2009	Indian	*CYP2E1*	* rs2031920/*5B/RsaI /rs6413432*6/DraI ↑ risk in alcoholics versus non-alcoholics (>additive, < multiplicative)	hospital/hospital	350/350

Table 6. Case control studies examining interaction between alcohol and oxidative metabolism genes and alcohol in production of SCCHN

Authors	Year	Subjects' race	Genes tested**	Evidence of interaction found	Case/control source	# cases/ # controls
Boccia et al.(121)	2008	Caucasian	*CYP2E1*	none	hospital/hospital	210/245
Buch et al.(122)	2008	Caucasian	*CYP2E1*	none	hospital/population	197/416
Hashibe et al.(104)	2008	European & Latin American	*ADH*	rs1229984 in *ADH1B* >*additive,* <*multiplicative* rs1573496 in *ADH7* heterogeneous effect but can't evaluate additive interaction	hospital/hospital	3449/5278
Soya et al.(126)	2008	Indian	*CYP2E1*	**1B* variant homozygote ↑ risk only in regular alcohol users (>additive, <multiplicative)	hospital/hospital	408/220
Asakage et al.(113)	2007	Japanese	*ADH, ALDH*	*ADH1B* rs1229984 > additive	hospital/clinic	96/642
Hiraki et al.(106)	2007	Japanese	*ADH, ALDH*	*ADH1B* Arg/Arg, *ALDH2* Glu/Lys ↑ risk in heavier drinkers (>additive, <multiplicative for both)	hospital/clinic	239/716
Hashibe et al. (103)	2006	European Caucasian	*ADH, ALDH*	*ADH1B**2 ↓ risk in heavier drinkers (>additive, <multiplicative), 3 *ALDH* polymorphisms ↑ risk in heavier drinkers (>2x multiplicative)	hospital & clinic/hospital & clinic	811/1083
Hashimoto et al.(114)	2006	Japanese	*ALDH*	none	hospital/mixed	192/192
Sugimura et al.(131)	2006	Japanese	*CYP2E1*	none	hospital/hospital	122/241
Peters et al.(109)	2005	EA*; AA* (4%)	*ADH*	*ADH1C*2-2* homozygote ↑ risk in heavy drinkers (>additive, >2x multiplicative)	hospital/population	521/599
Wang et al.(150)	2005	EA*; Other (4%)	*ADH1C*	none	hospital/clinic	348/330
Risch et al.(107)	2003	German Caucasian	*ADH1B, ADH1C*	none	population recruits/population	245/251
Zavras et al.(151)	2002	Greek	*ADH*	*ADH1C*2-2* homozygote ↑ risk as alcohol consumption ↑ (>additive, >2x multiplicative)	hospital/hospital	93/99
Olshan et al.(152)	2001	EA*; AA* (24%)	*ADH1C*	none	hospital/hospital	182/202
Schwartz et al.(153)	2001	EA*; AA* (6%)	*ADH*	*ADH1C*2-2* homozygote ↑ risk as alcohol consumption ↑ (>additive, >multiplicative)	population/ population	333/541
Bouchardy et al.(128)	2000	French Caucasian	*ADH, CYP2E1*	*ADH1C* - none *CYP2E1* DraI C allele, RsaI c2 allele both ↑ risk in heaviest drinkers (>2x multiplicative)	hospital/hospital	250/172
Katoh et al.(111)	1999	Japanese	*ALDH CYP2E1*	none	hospital/hospital	92/147
Morita et al.(130)	1999	Japanese	*CYP2E1*	none	unknown/clinic	145/164

Table 6. Case control studies examining interaction between alcohol and oxidative metabolism genes and alcohol in production of SCCHN

Authors	Year	Subjects' race	Genes tested**	Evidence of interaction found	Case/control source	# cases/ # controls
Gonzalez et al.(129)	1998	Caucasian	*CYP2E1*	none	hospital/blood donors	75/200

* AA=African American/black; EA=European American/white
** This column lists only those genes also being tested in this study (*ADH, ALDH, CYP2E1, SOD, GPX, CAT*)

1.5 EPIDEMIOLOGIC STUDIES OF GENETIC POLYMORPHISMS AND SURVIVAL AFTER SCCHN DIAGNOSIS

Of all the *ADH* genes, only *ADH1C* polymorphisms have been studied for effect on survival. Wang et al.(150) found higher recurrence and poorer overall and disease-specific survival among homozygotes for both the fast (*1) and slow (*2) alleles of *ADH1C*, as compared to *1/*2 heterozygotes. Overall and disease-specific hazard of survival in *1 homozygotes was 30% of that observed in heterozygotes ($HR_{overall}$=0.3, 95% CI=0.2-0.6; $HR_{disease\text{-}specific}$=0.3, 95% CI=0.1-0.8), and overall and disease-specific hazard of survival in *2 homozygotes was 40% of that of heterozygotes ($HR_{overall}$=0.4, 95% CI=0.2-0.9; $HR_{disease\text{-}specific}$=0.4, 95% CI=0.2-1.1).

Olivieri et al.(124) reported that, in addition to being present more frequently in alcohol users in cases than alcohol users in controls, the *CYP2E1**5B c2 (highly active) variant was significantly associated with advanced clinical stage (T3 and T4) at diagnosis.

Although many studies (154-165) have analyzed and discussed the significance of somatic mutations and variant protein expression in tumor cells, no studies have tested the germline *ALDH*, *CYP2E1, SOD*, *GPx*, and *CAT* genes for possible survival differences.

CHAPTER 2

AIMS AND METHODS

2.1 SPECIFIC AIMS

The overall goal of this dissertation project was to evaluate the relationship between polymorphisms of genes that, in combination with exposure to alcohol, may modify the risk of developing SCCHN or survival after diagnosis. The genes were selected on the basis of their potential functional importance related to alcohol metabolism, alcohol-induced carcinogenesis and previous epidemiologic literature on these genes and the risk of SCCHN.

The Carolina Head and Neck Cancer Epidemiologic study (CHANCE), a large population-based case-control study (n=1389 cases and 1396 controls), is the parent study upon which this research is based. It was designed to assess modifications of cancer risk due to genes in specific metabolic pathways that have been shown to affect cancer risk in prior studies of head and neck cancer and/or other types of cancer. Such candidate gene studies are designed to estimate the effect of genetic variants on population risk, in contrast to genome-wide association studies that attempt to identify chromosomal regions that are linked to the disease by ranking p-values, and do not estimate risk alterations. This dissertation project evaluated variants of the major genes that are active in metabolizing alcohol in the mouth, pharynx, and larynx in terms of their effect on SCCHN risk and survival: (1) alcohol dehydrogenase (*ADH*) genes; (2) acetaldehyde dehydrogenase (*ALDH*) genes; and (3) the cytochrome P450 gene *CYP2E1*. In addition, I examined variants of the primary oxidative stress genes and their effect on SCCHN risk and survival, including

glutathione peroxidase (*GPx*), superoxide dismutase (*SOD*), and catalase (*CAT*). These genes encode enzyme products that process carcinogenic oxygen free radicals produced primarily in the course of reactions catalyzed by CYP2E1 (166, 167). The prevalence of variants in these genes was compared between case and control groups to determine the risk of SCCHN. In addition, I obtained vital status of all CHANCE case participants from the National Death Index (NDI) in order to compare all-cause and cancer-specific death rates among patients by different polymorphisms of the aforementioned genes.

The primary specific aims of this study were as follows:

1. Estimate the main effects of each genetic polymorphism (SNPs and haplotypes) on the risk of developing SCCHN. *My hypothesis*: The polymorphisms are not associated with SCCHN risk in non-drinkers and light drinkers. If, however, some polymorphisms are found to be independently associated with risk in the absence of alcohol exposure, one may conclude either that study subjects misreported their drinking to be lower than it really is, and/or that a mechanism other than mutagenic alcohol metabolites is affecting cancer risk.
2. Estimate the interactive effect of alcohol exposure with each polymorphism on the risk of developing SCCHN. *My hypothesis*: Some polymorphisms will be associated with SCCHN risk in the presence of alcohol exposure. If, however, no polymorphism*alcohol interaction is found, one may conclude that either my study was under-powered to detect interactions, my study did not identify the important polymorphisms using the SNPs we genotyped, and/or that alcohol does not interact with the major gene polymorphisms that are present in whites and African-Americans.
3. Compare death rates of cases with different gene polymorphisms in order to identify survival differences associated with particular polymorphisms. *My hypothesis*: Polymorphisms associated with SCCHN risk in heavy drinkers at baseline will also be associated with their survival. If, however, this is found not to be the case, one may

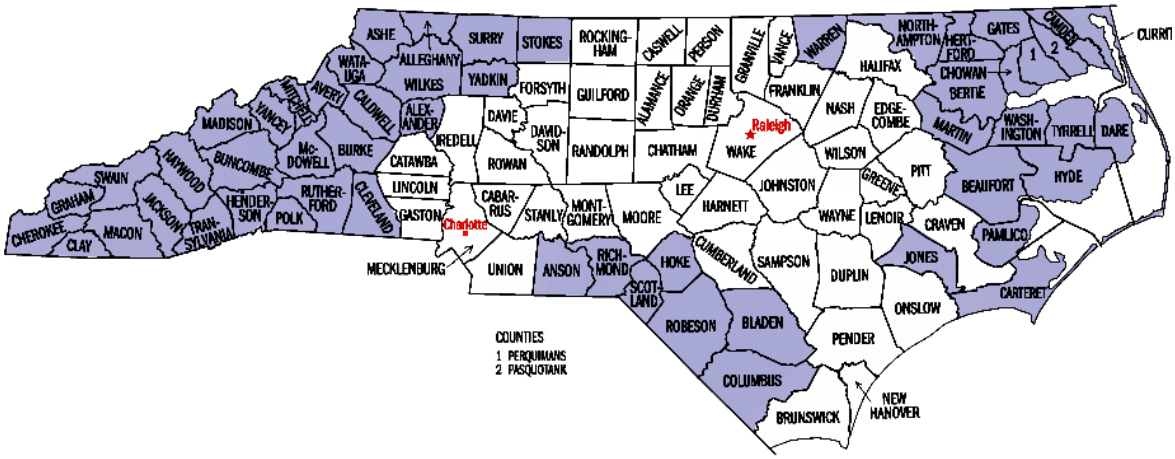
conclude either that the lack of complete follow-up measures of alcohol intake in participants who were heavy drinkers before their cancer diagnosis prevents us from discovering whether behaviors after prognosis are associated with survival, or that the mechanisms that affect cancer incidence are not the same as those that affect mortality once cancer has occurred.

Knowledge of higher-risk genotypes provides new insights about disease mechanisms, which can suggest new strategies for prevention, screening, diagnosis, and treatment. This may be especially useful for African-American men, who develop and die from SCCHN at much higher rates than white men, and women of both races.

2.2 STUDY POPULATION

Study area. The CHANCE geographic area (Figure 4) includes a population that is about one-third African American and two-thirds white, and a mix of rural and urban areas. Since head and neck cancers disproportionately afflict African Americans, and existing studies (Table 6) don't include enough African Americans to enable precise measurement of odds ratios, the CHANCE study population provides valuable information on an under-studied population (168).

Figure 4. Forty-six North Carolina counties from which CHANCE participants were enrolled. Participating counties are highlighted in white.



Examination of data on county of residence compared to county of diagnosis or treatment showed a total of only 5 cases (0.3%) resident in the entire 46-county study area that were diagnosed outside the study area in a single year; therefore almost all cases occurring in the study area during the study enrollment period were identified as eligible for enrollment, rendering the cases population-based. North Carolina incidence and mortality for oral cavity and pharyngeal cancer are higher than the national average (15, 169, 170), such that a sufficient number of cases (>1500) were expected to occur in the 46 counties over the four years of the study.

Cases. Cases were adults aged 20-80 with a newly diagnosed first primary invasive squamous cell carcinoma of the head and neck (pharynx, larynx, oral cavity) between January 1, 2002 and February 28, 2006. To be eligible, cases had to be residents of a 46-county region in North Carolina. People with carcinomas of other histologies or carcinomas at other head and neck sites, or a history of recurrent or second primary tumors, were not eligible.

Specifically included were oral cavity, including lip (mucosa) and pharynx (ICD-O-3 topography codes CO.OO to C14.8), and larynx (C32.0 to C32.9). Excluded were tumors of the salivary glands (C07.9, C08.0 to C08.9), nasopharynx (C11 .0 to C11.9), nasal cavity (C30.0), and nasal sinuses (C31.0 to C31 .9). Tumors with the following histologies were included (ICD-O-3 morphology codes): 8010/3 (epithelial neoplasm, carcinoma, NOS);

8051/3 (verrucous carcinoma, NOS); 8070/3 (SCC, NOS); 8083/3 (basaloid SCC); 8071/3 (SCC, keratinizing, NOS); 8072/3 (SCC, large cell, nonkeratinizing, NOS); 8073/3 (SCC, small cell, nonkeratinizing); 8074/3 (SCC, spindle cell); and 8076/3 (SCC, microinvasive). Excluded histologies included: 8010/0 (epithelial tumor, benign); 8010/2 (carcinoma, in situ, NOS); 8050/0-3 (papillary carcinoma); 8052/0-3 (papillary carcinoma); and 8075/3 (SCC, adenoid).

To minimize recall bias and to include later stage and terminally ill subjects in the study, rapid identification of newly-diagnosed cancer cases was conducted by the North Carolina Central Cancer Registry (CCR) to ensure that pathology reports were sent from hospitals to the CCR more rapidly than usual. The CCR team contacted the cancer registrars of 54 hospitals in the study area on a monthly basis to identify potentially eligible head and neck cancer cases. Pathology reports were usually sent to the CCR within four weeks of diagnosis. Staff entered basic data on age, pathology, tumor site, and physician name from each report into a database, and the information was then downloaded and delivered to the study office at the University of North Carolina at Chapel Hill.

Patient physicians were then notified by mail of their patient's eligibility for the study, and physicians were given the opportunity to refuse to give permission to the investigators to contact their patient. (To promote physician cooperation, all physicians practicing in the study area who were listed as head and neck or ENT surgeons in the North Carolina Medical Society and Medical Board lists were mailed a letter and brochure at the start of the study describing the study and soliciting their support.) After physician notification, potentially eligible cases were mailed a brochure with answers to commonly asked questions about the study, and a letter from the CCR describing the purpose of the Cancer Registry and the CCR's mission to work with researchers seeking to improve the health of North Carolina residents. The introduction letter explained the study purpose, the questionnaire, and biologic specimen collection (blood and/or mouth rinse). The letter

indicated that the study was completely voluntary, that participants would be paid $50 for completing the questionnaire, and that a study nurse would contact them by telephone to answer questions and schedule an appointment. Nurse-interviewers subsequently contacted the subjects to verify eligibility and schedule an interview in the subject's home or other convenient location. Written informed consent for the interview, collection of biologic specimens, and collection of cases' medical records were obtained prior to conducting the interview. Medical records were reviewed to verify that potential cases did indeed have one of the specified head and neck cancer histological types.

Of 1827 eligible SCCHN cases, 1337 (73%) personally completed the interview, and of these, 1313 (96%) also provided a biological sample with sufficient DNA for analysis (Figure 5).

Controls. Potentially eligible controls from the same counties as cases were identified through North Carolina Department of Motor Vehicles (DMV) records. Controls were frequency-matched to cases using random sampling with stratification on age, race, and sex to obtain comparability to the expected case distribution on these factors. Of 3035 potential controls with whom contact was attempted, 1379 (45%) personally completed the interview (Table 7), and of these, 1368 (99%) also provided a biologic sample with sufficient DNA for analysis (Figure 5).

The CHANCE study was very successful in locating eligible cases (97% located) and obtaining their cooperation (81%), yielding an excellent 79% case response rate of self-completion of an interview. As with most recent population-based case-control studies, locating and recruiting controls was more difficult. Among potentially eligible controls, 22% could not be located despite intensive tracing. About 37% of eligible and located controls refused to participate, compared to 17% of cases, which resulted in an adequate contact rate of 78% and cooperation rate of 61% but an overall response rate of 47%. The overall

response rate for African-American controls (36% versus 51% for whites) was lower largely due to the greater number of unlocatable subjects. (Table 7)

The few living participants who only completed the questionnaire without providing DNA (less than 1% of both cases and controls) did not differ significantly in terms of demographics or major risk factors (alcohol, smoking) from those who provided both questionnaire and DNA data. The 4% of cases (compared to 1% of controls) who died before they could complete the questionnaire and provide a DNA sample are also a small portion of the study participants; however, if survival is associated with gene variants as hypothesized, the associations measured in this study may be attenuated due to missing genotyping data for the mostly rapidly fatal cases. Therefore the source population for this proposed study is very similar, though not identical, to the CHANCE study population, and differs mainly because (1) 21 cases of lip cancer were excluded; (2) 46 cases of other race-ethnicity were excluded; and (3) genotype data on a small percentage of rapidly fatal cases are not available. (Reasons for excluding lip cancer and those of “other” race-ethnicity are explained later in this chapter.)

Institutional Review Board. Written informed consent was obtained from all subjects. The CHANCE parent study and this study were approved by the Biomedical Institutional Review Board at the University of North Carolina at Chapel Hill.

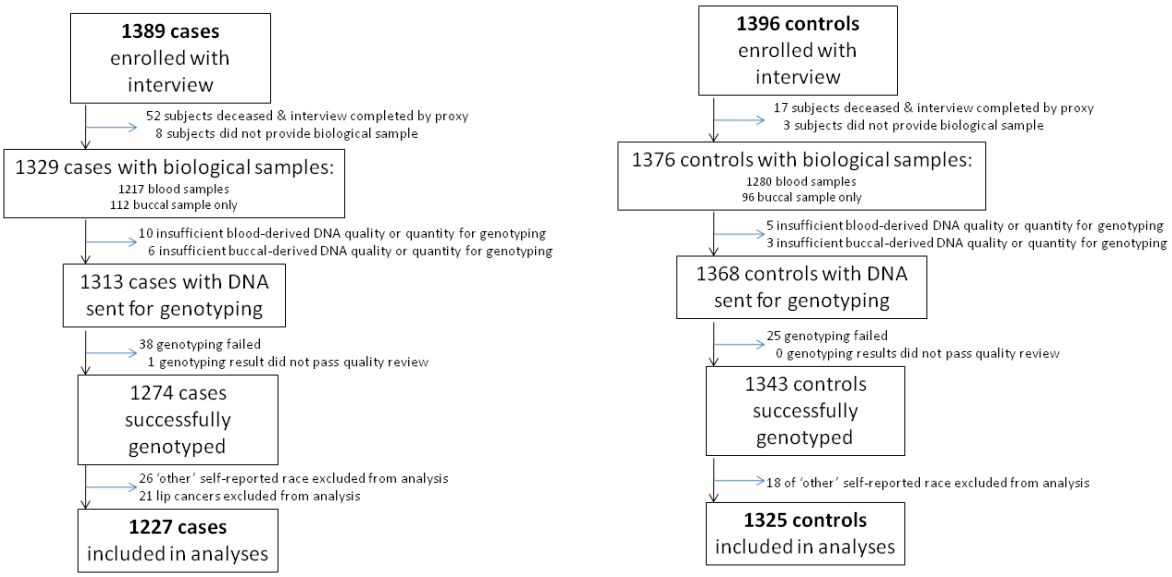
Table 7. CHANCE response rates by race

Study subject status	Total		Whites		African-Americans		Other[1]		Unknown	
	N	%	N	%	N	%	N	%	N	%
CASES										
Sampled	2135		1524		546		39		26	
Ineligible	268	(12.6)	184	(12.0)	69	(12.8)	7	(17.9)	9	(34.6)
Physician refusal	39	(1.8)	29	(1.9)	9	(1.6)	1	(2.6)	0	-
Eligible (A)	1827		1311		468		31		17	
Unlocatable (B)	50	(2.7)	27	(2.1)	19	(4.1)	1	(3.2)	3	(17.6)
Subject refusal	311	(17.0)	224	(17.1)	75	(16.0)	2	(6.5)	10	(58.8)
Deceased, no proxy	77	(4.2)	50	(3.8)	23	(5.0)	0	-	4	(23.5)
Eligible (A')[2]	1698		1231		424		30		13	
Completed interview – self (C')	1337		980		330		27		0	
Completed interview – proxy	52		30		21		1		0	
Total interviews (C)	1389		1011		351		28		0	
Contact rate = (A'-B)/A' (self)	0.971		0.978		0.955		0.967		0.769	
Cooperation rate = C'/(A'-B) (self)	0.811		0.814		0.815		0.931		-	
Response rate = C'/A' (self)	0.788		0.796		0.778		0.900		-	
Response rate = C/A (self+proxy)	0.760		0.770		0.750		0.903		-	
CONTROLS										
Sampled	4049		2803		1170		76		0	
Ineligible	234	(5.8)	165	(5.9)	57	(4.9)	12	(15.8)	0	-
Not initiated	780	(19.3)	406	(14.5)	349	(29.8)	25	(32.9)	0	-
Eligible (A)	3035		2232		764		39		0	
Unlocatable (B)	655	(21.6)	369	(16.5)	272	(35.6)	14	(35.9)	0	-
Subject refusal	875	(28.8)	672	(30.1)	196	(25.7)	7	(17.9)	0	-
Deceased, no proxy	109	(3.6)	77	(3.4)	32	(4.2)	0	-	0	-
Eligible (A')	2909		2141		729		39		0	
Completed interview – self (C')	1379		1100		261		18		0	
Completed interview – proxy	17		14		3		0		0	
Total interviews (C)	1396		1114		264		18		0	
Contact rate = (A'-B)/A' (self)	0.775		0.828		0.627		0.641		-	
Cooperation rate = C'/(A'-B) (self)	0.612		0.621		0.571		0.720		-	
Response rate = C'/A' (self)	0.474		0.514		0.358		0.462		-	
Response rate = C/A (self+proxy)	0.460		0.499		0.346		0.462		-	

[1] American Indian, Asian/Pacific Islander, and Other

[2] A' is the sum of self-completed interview (C'), unlocatable (B), and number of subjects who refused to enroll.

Figure 5. Flowchart of biological sample processing in CHANCE cases and controls



2.3 MEASUREMENTS

2.3.1 Outcomes

2.3.1.1 Incidence of head and neck cancer

See the "Cases" portion of Section 2.2 (Study population), above.

2.3.1.2 Mortality

In May 2011, determination of whether death had occurred in study participants by December 31, 2008, and, if so, the time and cause, was obtained from the National Death Index (NDI), a national file of identified death record information compiled from computer files submitted by State vital statistics offices. The NDI is considered the gold standard for identifying U.S. deaths; it demonstrated the highest sensitivity ratings, as compared to other databases of death information, in a 2002 study (171). Currently, it is also the only source at the national level with a cause of death field useful for research purposes; the discrepancy

rate between coded cause of death in the NDI and specialist-verified coded cause of death is estimated to be about 4%.(171) Death records are added to the NDI file annually, approximately 12 months after the end of a particular calendar year. The CHANCE study collected excellent matching data in multiple fields: social security number (SSN), date of birth (DOB), sex, race, state of residence, and name. Therefore there was a high proportion (75%) of perfect/very close to perfect matches on SSN, DOB, and sex. A small number of partial matches (e.g. a few SSN digits, parts of DOB) were examined case-by-case and a determination was made whether to accept or not. Using NDI vital status as of December 31, 2008, the follow-up time on CHANCE participants ranged from 2.8 to 7 years.

2.3.2 Main exposures

2.3.2.1 Alcohol drinking

Alcohol drinking exposure was measured in CHANCE by self-report answers to questions asked on the in-person nurse-administered questionnaire (Appendix B). Questions about alcohol use were based on previous questionnaires used in head and neck cancer studies and other cancer epidemiology and nutrition studies conducted at UNC. The questions were designed to estimate lifetime history of alcohol consumption prior to the year before diagnosis, and also asked about usual consumption prior to diagnosis of SCCHN. The questionnaire asked about beer, wine, and hard liquor separately. For each, it asked:

(1) Did you drink [beer/wine/hard liquor]?

(2) At what age did you start?

(3) At what age did you stop?

(4) For how many years did you drink [beer/wine/hard liquor] during this period?

(5) How much [beer/wine/hard liquor] did you usually drink?

Per day/week/month/year?

(6) What size did you usually drink?

CHANCE participants include a high number and percentage of never-drinkers (416 or 16%), which enabled precise risk estimates because never-drinkers were used as the referent category. The number and percentage of drinkers of only one type of alcoholic beverage were small (Table 8), and the resulting imprecision of estimates rendered it difficult to compare risks between beverage types. Drinking rate categories have large numbers in each class (Table 8), with the largest numbers in the light category (<3 drinks per day). When comparing cases to controls (Table 8), cases include a higher proportion of past drinkers and a lower proportion of never-drinkers than do controls (though, interestingly, the proportion of current drinkers among cases and controls is about the same). Cases also include a much smaller percentage of wine-only drinkers (1.2% of cases versus 5.3% of controls), and a much higher percentage of moderate and heavy drinkers (53% of cases versus 18% of controls).

In the United States, the typical alcoholic drink (12 ounces of beer, 4 ounces of wine, 1-1.5 ounces of liquor) contains about 12-14 grams of ethanol. By this measure, 5 drinks per day contain about 60-70 grams of ethanol, and 6-7 drinks per day contain about 80 grams of ethanol. Other countries consume different types of alcoholic drinks with different amounts of ethanol; for example, one “go” of sake (180 ml) contains about 27 grams of ethanol, while one beer (half-pint) in the United Kingdom contains about 8 grams (172). For each CHANCE subject, I used U.S. average estimates of ethanol volume in each type of beverage (beer, wine, liquor) to convert number of drinks per week of each type to milliliters of ethanol, and combined that information with duration of drinking to estimate the lifetime total of

Table 8. Distribution of alcohol-drinking behaviors in cases and controls

Category	All study participants (n=2657) # (%)*	Cases (n=1290) # (%)*	Controls (n=1367) # (%)*
Timing (# of participants with complete data on timing: n_{total}=2641, n_{cases}=1281, $n_{controls}$=1360)			
Never-drinkers	416 (15.8%)	124 (9.7%)	292 (21.5%)
Past drinkers	749 (28.4%)	434 (33.9%)	315 (23.2%)
Current drinkers	1476 (55.9%)	723 (56.4%)	753 (55.4%)
Type of alcoholic drink (# of participants with complete data on type of drink: n_{total}=2641, n_{cases}=1281, $n_{controls}$=1360)			
N/A (never drinker)	416 (15.8%)	124 (9.7%)	292 (21.5%)
Beer only	169 (6.4%)	100 (7.8%)	69 (5.1%)
Liquor only	93 (3.5%)	48 (3.7%)	45 (3.3%)
Wine only	88 (3.3%)	16 (1.2%)	72 (5.3%)
Multiple types	1875 (71.0%)	993 (77.5%)	882 (64.9%)
Drinking rate (past or current) (# of participants with complete data on maximum drinking rate: n_{total}=2648, n_{cases}=1283, $n_{controls}$=1365)			
N/A (never drinker)	416 (15.7%)	124 (9.7%)	292 (21.4%)
Light (≤ 2/day)	1313 (49.6%)	483 (37.6%)	830 (60.8%)
Moderate (3-5/day)	396 (15.0%)	248 (19.3%)	148 (10.8%)
Heavy (≥ 6/day)	523 (19.8%)	428 (33.4%)	95 (7.0%)

*Numbers may not add to total n in column header due to missing data for some participants. Percentages are calculated with a denominator equal to the number of participants with data (excluding those with missing data). Percentages in each category may not total exactly 100 due to rounding.

ethanol consumed for each beverage. I then summed the lifetime total for each beverage type to construct an estimate of lifetime ethanol consumption from all alcoholic beverage types. For purposes of comparing CHANCE risk-by-drinking-frequency estimates to those of INHANCE, I standardized CHANCE drink frequencies to the INHANCE definition of 15.6 ml ethanol per drink (of beer, wine, or liquor).

Alcohol variable selection and construction. Based on literature review as summarized by IARC (51) that concluded that frequency of drinking alcoholic beverages is more consistently associated with risk of SCCHN than duration of smoking, it would have been most desirable to construct a single drinking frequency variable for CHANCE subjects that integrates frequency of drinking beer, wine, and liquor. Although this was not possible due to the way the questions were worded, it was possible to estimate lifetime ethanol

consumption in milliliters. Because this variable is not standard in the literature on gene-ethanol interaction, I (a) examined smoothed splines of SCCHN odds versus duration, frequency, and lifetime intake of each alcoholic beverage, to determine general patterns; (b) calculated the correlation between drink frequencies for individual beverages and the lifetime ethanol intake for all beverages combined; and (c) examined smoothed splines of SCCHN odds versus lifetime intake of all beverages combined for purposes of determining a classification scheme that would best capture the odds pattern in CHANCE data.

Splines examined in (a) (Figure 6) indicated that the odds of getting cancer increased monotonically as *frequency* increased from 0 to 100 drinks per week, while the pattern for increasing *duration* was not consistently monotonic for each beverage type. When excluding subjects within the top 5% of lifetime alcohol consumption, cancer odds increased monotonically as lifetime consumption (of all beverage types combined) increased (Figure 7). I concluded that lifetime intake in milliliters of all beverage types combined is a reasonable substitute measure for the frequency of drinking, in terms of its effect on cancer odds.

Figure 6. Splines of log-odds of SCCHN cancer in CHANCE, versus frequency and duration, separately for each beverage (excluding never-drinkers)

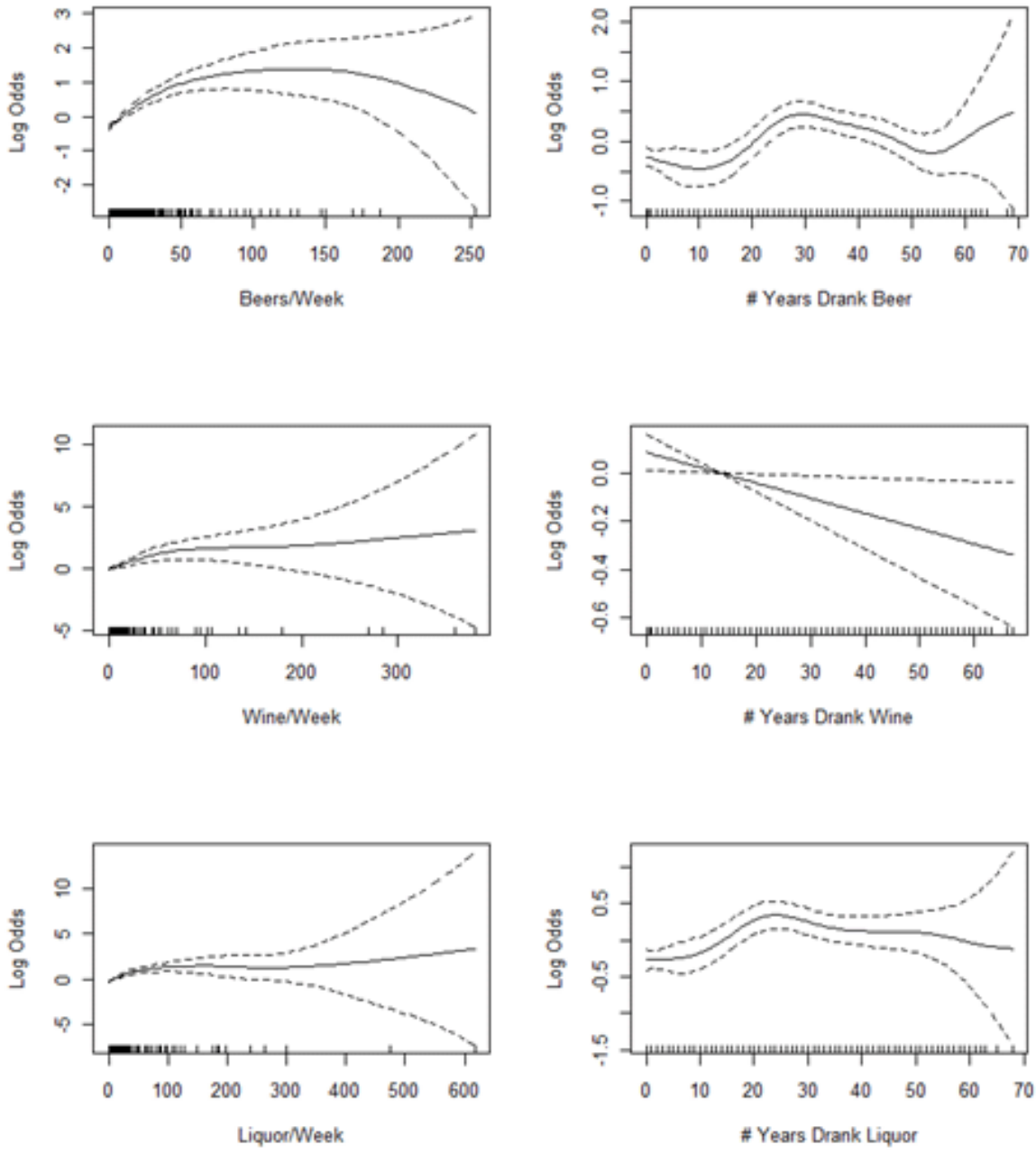
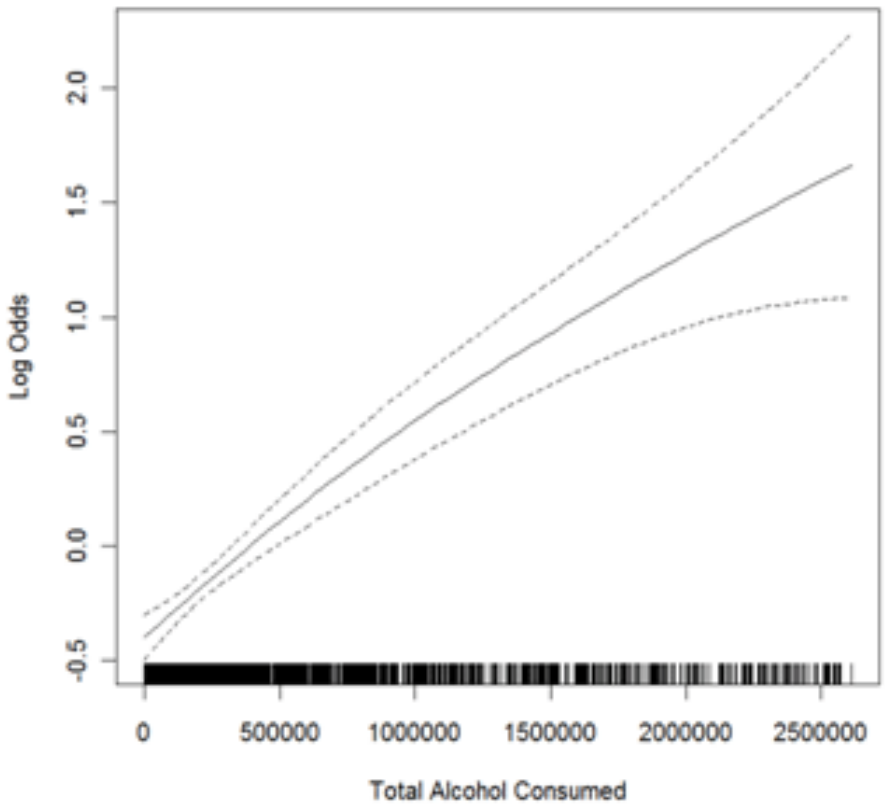


Figure 7. Splines of log-odds of SCCHN cancer in CHANCE, for lifetime intake (excluding never-drinkers and 5% of subjects whose lifetime intake was > 400 drink-years)



Correlations (b) indicated that the frequency of drinking each type of beverage is most highly correlated with the lifetime total consumed of that beverage (0.87 for liquor, 0.91 for beer, 0.93 for wine), while duration variables have very low correlation (0.28 for liquor, 0.36 for beer, 0.23 for wine), indicating that frequency contributes more significantly to the lifetime total variables than duration. In addition, the correlation between each frequency variable and lifetime intake in milliliters (summed over all types) is also relatively high (0.81 for liquor, 0.64 for beer, 0.53 for wine), and is much higher than for duration variables. Thus

the metric available in CHANCE -- overall lifetime ethanol intake -- is highly correlated with the preferred metric of drink frequency.

Finally, in an effort (c) to capture the risk profile for lifetime ethanol intake yet maintain a practical number of categories for estimating interaction with genetic polymorphisms, I explored various categorization schemes, finally settling on non-drinkers plus tertiles of lifetime drinking. It describes the risk profile observed in the splines reasonably well, and the average drink frequency of beer and liquor differs widely between categories (about 2 drinks/week, 7 drinks/week, and 40 drinks/week), thus mimicking frequency categories for which the literature has described very different risks.

Follow-up questions about alcohol use. CHANCE nurse-interviewers attempted to contact all cases about six months after their initial questionnaire, to ask questions about alcohol use that occurred after diagnosis. For each type of alcoholic beverage (beer, wine, liquor), participants were asked if they drank it after being diagnosed with SCCHN, and if so, the duration (days, weeks, months), frequency (number of drinks per day/week/month), and portion size they usually drank. The follow-up survey was administered to 711 of our 1227 cases (58%), but only 392 answered the alcohol questions; i.e. only 26% of cases answered questions about their post-diagnosis alcohol use. Because of the high rate of missingness and lack of understanding about who was interviewed post-diagnosis, when they were interviewed, and why so many did not answer the alcohol questions, I did not use these data in survival analyses.

2.3.2.2 Genetic polymorphisms

Biological sample collection. During the in-person visit by the research nurse-interviewers, who were trained in phlebotomy, biologic samples were obtained with the consent of the subjects. Two 10 ml yellow-topped vacutainer tubes (containing acid citrate dextrose) and one red-topped tube of whole blood (no additives) were obtained by

venipuncture. One yellow-topped tube was used for plasma and collection of mononuclear cells for subsequent DNA extraction. The other yellow-topped tube was used for plasma, buffy coat and packed red blood cell separation. The buffy coat was stored frozen for subsequent DNA extraction. The red-topped tube was used to collect serum that was stored for potential use in future assays. The interviewers transported the blood samples back to the laboratory in coolers, usually within 12 hours. In most cases, specimens were processed within 12-72 hours of collection. If the subject was not willing or able to consent to the blood draw, they were asked to contribute a non-invasive sample of buccal cells by swishing vigorously with 1.5 ounces of mouthwash or saline for 30 seconds, spitting the solution into a container, and repeating. The sealed container was then put in a plastic bag and stored in a cooler.

DNA extraction. DNA was extracted from either the fresh blood specimen from one yellow-topped tube or from buccal cells pelleted and frozen at -80°C using a modified salt procedure with Puregene chemistries. DNA samples were quantified in multi-spectral optical density spectrophotometers. The 260/280 ratio was used to assess sample quality. Ratios of >1.7 for DNA extracted from blood or >1.6 for buccal rinse samples were considered to be quality samples. In addition, each DNA sample was subjected to 0.4% agarose gel electrophoresis to assess DNA size. Greater than 96% of the blood samples' genomic DNA was of high quality as demonstrated by a single large band of DNA with a size greater than 25kb. DNA from some buccal samples was of insufficient quantity for genotyping. DNA was aliquotted into multiple vials which were stored at -80°C for long-term storage.

SNP selection strategies. SNPs tested in CHANCE were from candidate genes identified in the literature as being part of alcohol and oxidative stress metabolic pathways. First, tag SNPs in these candidate genes were chosen using the Genome Variation Server (GVS) (accessed between 2/22/08 and 4/25/08), a local database hosted by the SeattleSNPs Program for Genomic Applications (PGA), with the following parameters: allele

frequency cutoff was set to 10% and monomorphic sites were excluded; the R^2 threshold was set to 0.8; data coverage percent for tag SNPs was set to the default 85%; and data coverage percent for clustering was set to the default 70%. The data merge option selected was "combined samples with common variations." The populations used were primarily the HapMap-Ceu (unrelated only) and HapMap-YRI (unrelated only), Release 2.

For this dissertation project, 75 SNPs were selected in 12 genes: *ADH1B, ADH1C, ADH4, ADH7, ALDH2, CYP2E1, SOD1, SOD2, GPX1, GPX2, GPX4, and CAT*. Sixty-nine (92%) were selected as tag SNPs for candidate genes from the GVS, and the remaining six were not tag SNPs but were forced in because prior research found high association between the SNP and either the development of aerodigestive cancer, survival from breast cancer, alcohol dependence, or interaction with genes in the proposed carcinogenic pathway for vinyl chloride. In addition, 157 ancestry informative markers selected by J. Barnholtz-Sloan of Case Comprehensive Cancer Center at Case Western Reserve University were genotyped to permit adjustment for percent African ancestry; see details below in the "Genetic ancestry" sub-section under section "2.3.3 Covariates." Genotyping of gene-related SNPs and tag SNPs was done on the same plates in the same laboratory, thus eliminating errors due to differences in laboratories and techniques.

Genotyping and quality control. Exposure to selected gene polymorphisms was determined by genotyping of 75 SNPs located in exon, intron, untranslated (UTR) and nearby regions of the *ADH1B*, *ADH1C*, *ADH4*, *ADH7*, *ALDH2*, *CYP2E1*, *SOD1*, *SOD2*, *GPX1*, *GPX2*, *GPX4*, and *CAT* genes (see Table 9). Genotyping was done by the University of North Carolina at Chapel Hill, Mammalian Genotyping Core Facility, using the Illumina GoldenGate genotyping assay with Sentrix Array Matrix and 96-well standard microtiter plates (173), a multiplexing technique that tests for 1,536 SNPs simultaneously. The SNPs in Table 9 that were analyzed in this dissertation project were genotyped as part of a larger panel of 1,536 SNPs.

The Illumina GoldenGate technology has been widely used, for example in phase I of the International HapMap Project (174). Locus specificity is attained by a two-step recognition that includes annealing of both upstream and downstream oligonucleotides to the SNP site, followed by enzymatic mismatch discrimination to attain additional genomic specificity and selectivity for a particular allele. This is followed by allele-specific primer extension and ligation, assay amplification, and hybridization of multiplex PCR amplification products to a universal array of address sequences. Strict washing protocols remove excess and incorrectly hybridized probes. When CHANCE samples were genotyped, technicians were blinded to case/control status of the samples, and samples from both cases and controls were tested on each plate, along with DNA controls and random sample duplicates.

Assay intensity data and genotype cluster images for all SNPs were individually reviewed; as a result, 14 SNPs (including 12 AIMs) from this study were excluded from the dataset (6%) due to inadequate signal intensity or inability to distinguish genotype clusters. All specimens had less than 4% missing SNP results. Blind duplicates of 109 samples were genotyped to verify reliability of genotype calls. Among the 109 samples' paired results, 145,568 non-missing pairs of SNPs were examined for agreement; of these, five pairs (0.003%) were discrepant, but not for any of the SNPs and AIMs to be used in my analyses. Tests of Hardy-Weinberg equilibrium (HWE) were conducted in controls stratified by race; 35 SNPs were judged to be out of HWE due to p-value less than 0.001, including one tag SNP each in *ADH1C* and *ADH4* that were to be analyzed in this dissertation project (Table 9). These two SNPs were excluded from all analyses.

Table 9. List of SNPs genotyped in or near metabolic genes, by gene, in ascending position order

Gene	SNP ID	Alleles (ancestral/ variant)	Average heterozygosity (140)	Functional location within gene	SNP selection method
ADH1B (ADH1B alcohol dehydrogenase 1B (class I), beta polypeptide [Homo sapiens]) Location: 4q21-q23					
	rs12507573	A/C	0.420	vicinity of gene	GVS-tag
	rs1042026	A/G	0.493	exon 9 (3' UTR)	GVS-tag
	rs17033[3]	A/G	0.190	exon 9 (3' UTR)	GVS-tag
	rs7673353	C/T	0.104	intron 8	GVS-tag
	rs17028834	C/T	0.082	intron 7	GVS-tag
	rs1693457	C/T	0.237	intron 5	GVS-tag
	rs4147536[3]	G/T	0.344	intron 3	GVS-tag
	rs1229984	A/G	0.397	missense	(104)
	rs1353621[3]	A/G	0.302	intron 1	GVS-tag
	rs1159918	G/T	0.484	3' near gene	GVS-tag
	rs1229982	G/T	0.346	3' near gene	GVS-tag
ADH1C (alcohol dehydrogenase 1C (class I), gamma polypeptide [Homo sapiens]) Location: 4q21-q23					
	rs2298753	C/T	0.102	exon 9 (3' UTR)	GVS-tag
	rs1614972	C/T	0.495	intron 8	GVS-tag
	rs1391088	A/C	0.073	intron 8	GVS-tag
	rs698[3]	A/(G/T)	0.389	exon 8 (both missense)	(116)
	rs1693482	G/A	0.366	exon 6 (missense)	GVS-tag (116, 175-177)
	rs1631460	C/G	0.239	intron 5	GVS-tag
	rs2241894[3]	A/G	0.487	exon 5 (synonymous)	GVS-tag
	rs3762896[1]	C/T	0.496	intron 3	GVS-tag
	rs11936869	C/G	0.478	intron 1	GVS-tag
ADH4 (alcohol dehydrogenase 4 (class II), pi polypeptide [Homo sapiens]) Location: 4q21-q24; 4q22					
	rs29001227	A/T	0.083	intron 8	GVS-tag
	rs1126672	C/T,A,G	0.214	exon 8 (T synonymous; A,G missense)	GVS-tag
	rs4699710	C/T	0.325	intron 6	GVS-tag
	rs10017466	C/T	0.309	intron 5	GVS-tag
	rs1800759	A/C	0.499	5' near gene	GVS-tag
	rs1800761	A/G/T	0.435	5' near gene	GVS-tag
	rs3762894	C/T	0.471	5' near gene	GVS-tag
	rs4148884[3]	A/G	0.347	5' near gene	GVS-tag
ADH7 (alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide [Homo sapiens]) Location: 4q23-q24					
	rs284787	C/T	0.435	exon 9 (3' UTR)	GVS-tag
	rs894369	C/G	0.325	exon 9 (3' UTR)	GVS-tag
	rs17588403	A/T	0.178	intron 8	GVS-tag

Table 9. List of SNPs genotyped in or near metabolic genes, by gene, in ascending position order

Gene	SNP ID	Alleles (ancestral/ variant)	Average heterozygosity (140)	Functional location within gene	SNP selection method
	rs1154454	C/T	0.315	intron 7	GVS-tag
	rs1154456	C/T	0.347	intron 7	GVS-tag
	rs1154460	A/G	0.485	intron 6	GVS-tag
	rs971074	G/A	0.221	exon 6 (synonymous)	GVS-tag
	rs1573496	G/C	0.172	exon 3 (missense)	(104)
ALDH2 (aldehyde dehydrogenase 2 family (mitochondrial) [Homo sapiens]) Location: 12q24.2					
	rs4767939	A/G	0.487	intron 1	GVS-tag
	rs2238151	C/T	0.377	intron 1	GVS-tag
	rs7312055	A/G	0.164	intron 2	GVS-tag
	rs2158029	A/G	0.411	intron 12	GVS-tag
	rs16941667	C/T	0.134	intron 12	GVS-tag
	rs16941669	G/T	0.165	intron 12	GVS-tag
CYP2E1 (cytochrome P450, family 2, subfamily E, polypeptide 1 [Homo sapiens]) Location: 10q24.3-qter					
	rs3813865	C/G	0.256	5' near gene	GVS-tag
	rs3813867	C/G	0.153	5' near gene	(123, 178, 179)
	rs915906[3]	C/T	0.469	intron 2	GVS-tag
	rs8192772	C/T	0.269	intron 2	GVS-tag
	rs6413419[1]	G/A	0.101	exon 4 (missense)	GVS-tag
	rs915908	A/G	0.235	intron 5	GVS-tag
	rs915909	C/T	0.064	exon 6 (synonymous)	GVS-tag
	rs7092584	C/T	0.416	intron 6	GVS-tag
	rs743535	C/T	0.353	intron 6	GVS-tag
	rs2515641[3]	C/T	0.404	exon 8 (synonymous)	GVS-tag
	rs2249695	C/T	0.494	intron 8	GVS-tag
	rs28969387	A/T	0.031	exon 9 (missense)	GVS-tag
	rs11101812	C/T	0.041	exon 9 (3' UTR)	GVS-tag
SOD1 (superoxide dismutase 1, soluble [Homo sapiens]) Location: 21q22.11					
	rs11910115	A/C	0.070	5' near gene	GVS-tag
	rs4998557	A/G	0.456	intron 1	GVS-tag
	rs10432782	G/T	0.409	intron 2	GVS-tag
	rs2070424	A/G	0.377	intron 3	GVS-tag
	rs1041740	C/T	0.370	intron 4	GVS-tag
SOD2 (superoxide dismutase 2, mitochondrial [Homo sapiens]) Location: 6q25.3					
	rs4342445	A/G	0.390	vicinity of *SOD2*	GVS-tag
	rs2842980	A/T	0.471	3' near gene	GVS-tag
	rs8031	A/T	0.380	intron 5	GVS-tag

Table 9. List of SNPs genotyped in or near metabolic genes, by gene, in ascending position order

Gene	SNP ID	Alleles (ancestral/ variant)	Average heterozygosity (140)	Functional location within gene	SNP selection method
	rs5746134	C/T	0.104	intron 5	GVS-tag
	rs2758331	A/C	0.393	intron 4	GVS-tag
	rs4880[3]	C/T	0.460	vicinity of *SOD2*	GVS-tag
GPX1 (glutathione peroxidase 1 [Homo sapiens]) Location: 3p21.3					
	rs8179172	A/T	0.070	vicinity of *GPX1*	GVS-tag
	rs1800668	C/T	0.274	exon 1 (5' UTR)	GVS-tag
	rs3811699	A/G	0.302	5' near gene	GVS-tag
	rs3448	C/T	0.316	exon 5 (3' UTR)	GVS-tag
GPX2 (glutathione peroxidase 2 (gastrointestinal) [Homo sapiens]) Location: 14q24.1					
	rs11623705	G/T	0.136	3' near gene	GVS-tag
	rs2412065	C/G	0.367	intron 1	GVS-tag
	rs2737844	C/T	0.474	intron 1	GVS-tag
GPX4 (glutathione peroxidase 4 (phospholipid hydroperoxidase) [Homo sapiens]) Location: 19p13.3					
	rs757229	C/G	0.493	5' near gene	(180)
CAT (catalase [Homo sapiens]) Location: 11p13					
	rs1049982	C/T	0.357	exon 1 (5' UTR)	(181) [(2)]

[1]Genotype results questionable due to being out of Hardy-Weinberg equilibrium at the exact p<0.001 level.
[2]Also selected due to unpublished finding from the Carolina Breast Cancer Study.
[3]Genotyping failed.

Cross-checking of genetic data for inconsistencies was accomplished using computerized algorithms implemented in SAS. The algorithms checked that genetic sex determination matched the study sex variable; one sample was discrepant on sex so its genotyping results were discarded. In addition, genotyping of controls that was performed on each plate was determined to be accurate, and there were no identical genotypes among unique samples. Analysts also looked for patterns among samples that were unsuccessful, such as plate location, DNA concentration, volume, or total DNA amount. SNP

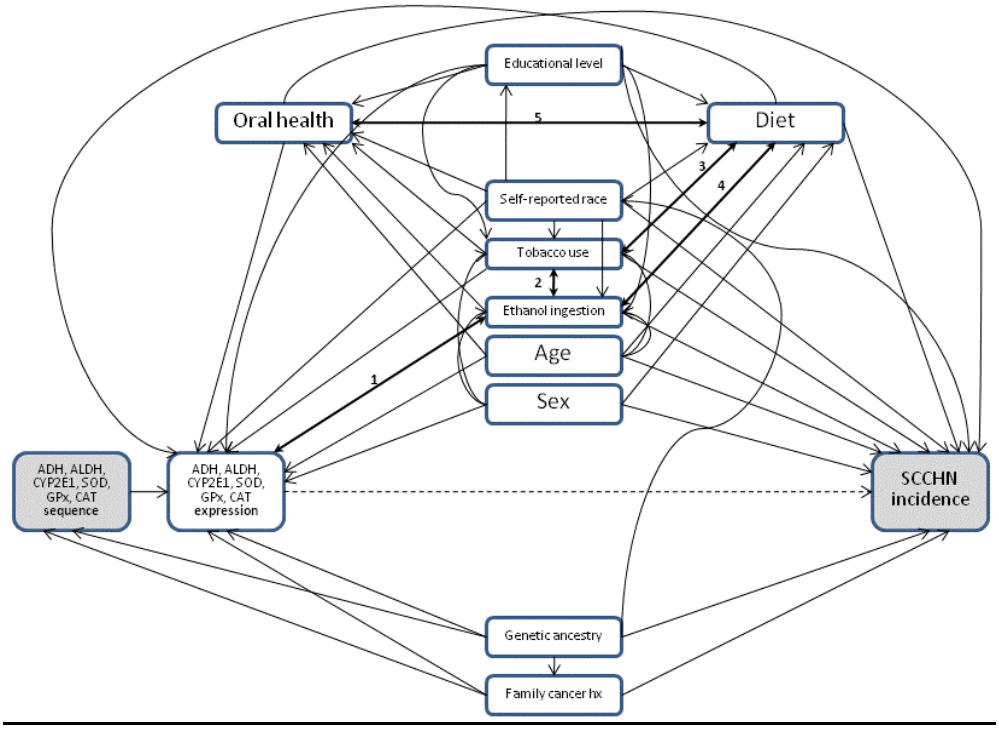
inconsistencies and plate problems were resolved, either by re-testing and calling, else being set to missing.

As part of this dissertation project, statistical tests for Hardy-Weinberg equilibrium (HWE) of each SNP's genotype distribution were conducted separately for each race in the control population, to verify previous HWE tests. DNA samples were not available from 52 cases and 17 controls because they had died before they could be interviewed (their questionnaires were completed by a proxy). This represents 4% of all cases interviewed and 1% of all controls interviewed, and therefore would have only a small effect on effect estimates.

2.3.3 Covariates: potential confounders and effect measure modifiers

A causal diagram for development of SCCHN is shown in Figure 8. There are known bi-directional associations between ethanol/tobacco use and gene expression, ethanol/tobacco use and diet, and oral health and diet. Because all of the factors listed can conceivably affect both gene expression and cancer outcome, I included all these variables in the initial models, using a 10% (0.1) change in log-odds estimate from the full model as the criterion for final variable selection, and used the reduced model in regression analyses (182).

Figure 8. Causal diagram for development of SCCHN



Dotted line indicates hypothesized causal connection between main exposure, gene sequence/expression, and SCCHN incidence. Note bi-directional, mutual effects of (1) ethanol ingestion and gene expression, (2) tobacco use and ethanol ingestion, (3) tobacco use and diet, (4) ethanol ingestion and diet, and (5) oral health and diet.

Genetic ancestry. Case-control studies assume that, barring confounding effects, differences in allele frequency between cases and controls are related directly to the trait of interest. However, differences in allele frequency between cases and controls may be due to systematic differences in ancestry (i.e. population stratification) rather than the association of genes with disease, making ancestry a potential confounder of any association between gene and disease. Population stratification is present in recently admixed populations such as African Americans and Latinos, but also in European Americans and even in historically isolated populations such as Tibeto-Burmese and Icelanders (183). Most residents of North Carolina are of European and African ancestry, with only a small percentage reporting American Indian and Asian ancestry. Because this study excluded the small number of

those reporting “other” race-ethnicity (n=46), all remaining study participants are self-identified as either African American or white. Therefore the appropriate ancestral populations for CHANCE participants are European and West African, whose genetic variations have been measured in the CEU and YRI HapMap populations, respectively.

For population stratification to exist, both (1) the frequency of the SNP allele and (2) the background disease prevalence must vary significantly by race-ethnicity. First, SCCHN disproportionately affects a higher percentage of African-Americans than whites. Second, of the six non-synonymous candidate gene SNPs genotyped in CHANCE for this study, three have very different allele frequencies in HapMap CEU and YRI populations (HapMap Data Rel 27 Phase II+III, Feb09, on NCBI B36 assembly, dbSNP b126):

1. rs1693482 (*ADH1C*): 48% of the CEU population carries the variant T allele, compared to only 5% of the YRI population;
2. rs698 (*ADH1C*): 47% of the CEU population carries the variant C allele for compared to 6% of YRI;
3. rs6413419 (*CYP2E1*): 2% of the CEU population carries the variant A allele compared to 30% of YRI.
4. rs1573496 (*ADH7*): 9% of the CEU population carries the variant G allele compared to 0% of YRI
5. rs28969387 (*CYP2E1*): 0% of the CEU population carries the variant T allele compared to 4% of YRI

Therefore it is possible that genetic ancestry is a confounder in the CHANCE study population.

Sometimes stratifying effect estimates by self-reported race-ethnicity may be sufficient to control for ancestry, but populations in the United States are recently admixed, causing increased inter-individual variation in ancestry. In the United States and many other countries, in addition to genetic ancestry, cultural and behavioral factors influence

individuals' self-identified race-ethnicity, such that race-ethnicity should be recognized as a complex composite construct of genetic and environmental/social factors rather than a simple marker of genetic ancestry. One case-control study conducted in Detroit, Michigan, a highly admixed population with German, Polish, and African origins, reported that adjusting for ancestry provided better control of confounding by ancestry than adjusting for self-reported race (184). Another study (not case-control) analyzed frequency of risk genotypes in six drug metabolizing genes in individuals from eight international populations (including Africans, Europeans, Ashkenazi Jews, Chinese, Pacific Islanders, and Afro-Caribbeans), and found not only that they varied by ancestral group, but that self-reported race was an inaccurate and insufficient marker of these ancestral clusters (185).

Correcting for inter-individual differences in ancestry in the CHANCE study required genotyping additional SNPs. Genomic control (GC) was the first method described in the literature to adjust for potential population stratification. GC uses random genomic SNPs across the genome and calculates a single numeric adjustment to be applied to all association tests. A separate adjustment can be calculated for each race-ethnicity. It considers group-level stratification only (as defined by self-reported race-ethnicity), and controls for Type I error but not Type II. Although ancestry informative markers (AIMs), markers that have large allele frequency differences between ancestral populations, can also be used in genomic control methods, it is not recommended since additional Type I errors will result because the method will over-correct for the large differences in AIM frequencies between different populations.

Because estimating individual ancestry has been shown to provide better control of confounding by ancestry than genomic control, CHANCE genotyped 157 SNP AIMs (Appendix A) on each participant. In order to obtain robust individual ancestry estimates, ideal AIMs would have one fixed, unique allele in each ancestral European and African population. It has been estimated that at least 72 ideal markers would be required to obtain

ancestry estimates with acceptable confidence limits (186). Since ideal markers are relatively rare in the genome, researchers must select the most informative markers from the remaining suboptimal markers (187). Both Bayesian and maximum likelihood estimation (MLE) methods can be used to calculate ancestry using the AIMs, and both methods produce similar results in information-rich scenarios (i.e. markers are informative and ancestral groups are large and accurate), but MLE is computationally much less intensive and does not require inclusion of genotype information from ancestral groups as the Bayesian methods do for estimation stability (188). In addition, interpretation of the results from the Bayesian methods can sometimes be confusing as it is not necessarily readily apparent which of the estimated clusters is related to which ancestral group. Therefore MLE was chosen to estimate ancestry in CHANCE participants.

One disadvantage of MLE methods is that they cannot easily estimate percent ancestry of more than two ancestral groups, though this should not be a major problem in CHANCE. And our AIMs in particular were not selected to identify Asian or Native American ancestry, so that study subjects self-reporting their race as “Other” could not be included in analyses of all participants even if that analysis should be desired.

Jill Barnholtz-Sloan, a CHANCE collaborator at Case-Western Reserve University, selected the CHANCE SNP AIMs to maximize (1) the difference in allele frequencies (delta) between European and African populations in the HapMap data (CEU versus YRI), and (2) the Fisher’s information criterion (FIC) which is the inverse of the maximum likelihood estimation (MLE) of the ancestral proportion and therefore is directly related to the precision of the ancestry estimate (187). FIC is calculated using the allele frequencies and the estimated proportions of mixture of the two populations. AIMs were prioritized based on having the highest delta values and the highest FIC values in the following order: 90% European/10% African, 10% European/90% African, and 50% European/50% African. This

prioritization scheme allowed AIMs to be chosen to represent the whole expected ancestral distribution of this population. Individual estimates of African ancestry were calculated from 141 successfully genotyped AIMs using maximum likelihood estimation methods previously described by Barnholtz-Sloan (184, 189). MLEs were calculated using customized software written by Barnholtz-Sloan and her team; other software that also estimates individual ancestry using MLE, such as FRAPPE (188), could also have been used. AIMs were chosen to differentiate between African and European ancestry only, so individual ancestry proportions for the two groups sum to 1.0. Percent African ancestry of each participant was included as a continuous covariate in full/initial logistic and proportional hazard regression models using the entire study population.

Demographics. Controls were matched to cases on age group, self-reported race, and sex, so were included as stratification variables in conditional logistic regression analyses. As another measure of the sociodemographic factors for which self-reported race-ethnicity is a marker, CHANCE collected information on education level (less than high school, high school or above) and income (either above or below the federal poverty level, based on income and the number of people supported). I included both in full/initial regression models.

Tobacco Use. Because tobacco use is known to be an important risk factor for SCCHN, CHANCE collected detailed self-reported information about each participant's history of cigarette, cigar, pipe, snuff and chewing tobacco use. Subjects were asked the age they started using the specific form of tobacco, current status, if and when they stopped, years of use, and the number used. In addition, information about use of filter brand cigarettes, size of chewing tobacco and snuff packages, and time to completely use a container were obtained. Subjects were asked about smoking among other family members and individuals living in the same household to quantify potential environmental tobacco

smoke (ETS) exposures. With this information, indices of tobacco use were constructed, including average number of cigarettes smoked per week, years of smoking, and pack-years. Based on current consensus in the literature, I adjusted for cigarette smoking by including duration of smoking in regression models. Exposure to environmental tobacco smoke (ETS) at home and at work was included in initial models as two separate dichotomous covariates. Smoking variables that changed the main effect estimate by 10% or more were retained in all subsequent models.

Diet. Because fruit and vegetable intake is the major dietary influence on SCCHN, and intake patterns are altered among heavy drinkers (and *ADH*, *ALDH*, and possibly *CYP2E1* are associated with alcoholism), the CHANCE diet questions focused primarily on fruit and vegetable consumption. However, in order to explore associations between other dietary components that may be related to SCCHN, CHANCE used an assessment that captures total diet. Dietary intake was measured using the 60-item version of a validated food frequency instrument (FFQ), the NCI Health Habits and History Questionnaire (190). This instrument assesses respondents' frequency of consumption and usual serving size for 60 foods. It has been validated with various populations including low income and African-American populations (191). A software package available from the National Cancer Institute, the Dietary Analysis Personal Computer System (DIETSYS.V30), uses this information and an updated database to compute nutrient intake based on nutrient content of each food and serving. Initially I intended to adjust separately for dietary intake of fruits and vegetables using a quintile-based categorization of number of fruits and vegetables eaten daily. However a large percentage of subjects were missing information on fruit and vegetable intake, and those missing the information were often at the high extreme of total caloric intake. I concluded that including these variables would inordinately bias effect estimates, so these variables were excluded from regression analyses.

Anatomic location and clinical factors. I included tumor location in five categories (oral cavity; oropharynx; hypopharynx; larynx; oral cavity, oropharynx, hypopharynx cancer NOS) as recommended by INHANCE (52) in logistic and Cox regression analyses having any SCCHN as the outcome. In secondary analyses, to explore risk factor differences by anatomic location of tumor, I stratified analyses by these five anatomic sub-sites.

Survival analyses took into account TNM stage that was abstracted from patients' medical records, which I divided into three ordinal variables (tumor, node, and metastasis stages). Cancer summary stage was abstracted from TNM staging that was obtained from patients' medical records and was entered into models as an ordinal variable. Recent survival studies on head and neck cancer have usually controlled Cox regression analyses for stage using separate tumor and node variables, only including metastasis if any of their patients were diagnosed with metastatic disease. I included separate T, N, and M variables in my Cox proportional hazards models, but also ran analyses using the single ordinal summary stage variable, and results were similar.

Treatment types were recorded as surgery, radiation, and chemotherapy, which I included in models as three dichotomous variables. Additional clinical information on presence of angiolymphatic invasion was included in Cox regression models as a dichotomous variable. Margin status, although available, was eliminated by backwards elimination and therefore excluded from final models. Information on extracapsular extension was available on less than 15% of subjects, so was excluded from all models.

Four dichotomous variables representing selected co-morbid diseases were included in full/initial Cox models: liver disease and/or hepatitis, cardiovascular and/or lung disease, renal disease and/or diabetes, and other cancer. A numeric variable counting the number of co-morbid diseases was also included in initial logistic and Cox regression models, but did not change effect estimates and was excluded from final models.

Other Factors. Other factors for which CHANCE collected data are oral health and mouthwash use; medical history; and family history of cancer in first degree relatives.

A recent study analyzing CHANCE data to investigate the association between oral health factors and incidence of SCCHN provided support for a modest association of SCCHN with periodontal disease, as measured by self-reported tooth loss indicators (such as tooth mobility, dichotomous, yes or no; having had one or more routine dental visits during the 10 years prior to SCCHN diagnosis, dichotomous, yes or no). However, there was no association between SCCHN and number of natural teeth lost excluding third molars and teeth extracted for orthodontic reasons (coded as a 3-level indicator variable for number of lost teeth: 0-5 (referent), 6-15, and 16-28). Consistent use of mouthwash (dichotomous, yes or no) showed evidence of a protective effect only for pharyngeal cancer (192). In addition, oral health factors are probably altered among heavy drinkers (and heavy drinking is associated with *ADH* and *ALDH* polymorphisms, and possibly *CYP2E1* polymorphisms as well). Therefore I included, as dichotomous variables (yes/no), tooth mobility, routine dental visits, and mouthwash use in initial models, retaining them in all subsequent models if they change effect estimates by >10%. None of these variables were important in SCCHN incidence models, but routine dental visits and tooth mobility variables were retained in some survival models.

Family history of cancer in first degree relatives, especially head and neck cancer, has been shown to increase risk of head and neck cancer in probands (see above, section 1.3.1 “Heredity”) and may well be associated with the genetic polymorphisms under study. Therefore I included in initial models one dichotomous (yes/no) variable for having had any first degree relatives with head and neck cancer. However, this variable did not change effect estimates and was excluded in all final models.

2.4 STATISTICAL POWER

Specific Aim 1: *Main effects of genotype exposure on power of logistic regression analyses:*

(Refer to Table 10 for power estimates based on genotype prevalence and size of study.) At an α level of 0.05, for calculating the main effect of a binary genotype exposure (any minor allele versus none) occurring at a frequency of 0.10 to 0.30 in the entire study population, this study had insufficient power (defined as less than 80%) to detect an OR of 1.2 but sufficient power to detect higher ORs. As it turned out, the minor allele of 80% of our SNPs had population prevalence >10%, so we had power to detect ORs of about 1.4 or higher for most SNPs. Power to detect main effects of genotype in the African-American participants was insufficient for genotype prevalence ≤20% when the OR is less than 2.0, but the study should have been able to detect main effects in the African-American population if the genotype prevalence is >20% (which occurred in 78% of SNPs in African-American subjects) and the OR >=1.8.

(Refer to Table 5 for estimates of main gene effects.) For genotype prevalence of 10% in all study subjects combined, there was sufficient power to detect effects of 5 of 8 SNPs in *ADH1B*, for which previously estimated protective effects ranged from 0.42 to 0.56 and risk effects ranged from 2.20 to 2.67. There was marginal (70-80%) power to detect effects of *ADH1C*, for which previously measured risk effects ranged from 1.20 to 1.49 with one imprecise outlier of 7.39. The study had sufficient power to detect effects of SNPs in *ADH4* and *ADH7* with previously measured protective estimates of about 0.68 and risk estimates ranging from 1.62 to 1.96. The study had sufficient power to detect ORs of 1.30 for *ALDH2* polymorphisms common in European Caucasians, and ORs of 1.8 to 2.6 for *CYP2E1* SNPs common in European Caucasians. Power estimates in whites should be similar to those of the entire group, though power was not sufficient to detect main effects of

ADH1C if OR<1.4. It should be noted that <10% minor allele prevalence occurred in one-third of our *ADH1B* SNPs and half our *CYP2E1* SNPs, so we had insufficient power to detect any effects of those SNPs.

Because essentially no previous studies have measured gene effects in African and African-American populations, effect estimates were unknown in this population prior to this study. Assuming that effect estimates in European Caucasians and European Americans also apply generally to African-Americans, this study would have had insufficient power to detect effect estimates less than 2.0 in African-Americans (e.g. *ADH1C, ADH4, ADH7, and ALDH2*), but sufficient power to detect previously reported effects of some *ADH1B* and *CYP2E1* polymorphisms. Power was insufficient to detect effects of *ADH4* and *ADH7* polymorphisms except those that were present in 30% or more of the African-American population, which was the case for about half of *ADH4* and *ADH7* SNPs.

For oxidative stress genes (*CAT, SOD1, SOD2, GPX1, GPX2,* and *GPX4*) this study had insufficient power to detect main effect ORs of 1.2 to 1.3 in the entire study population and in African-Americans and whites separately. ORs would have to be higher than that reported for most polymorphisms (at least 1.4) to detect a main effect in the entire study population, and at least 1.8 in the African-American participants.

Table 10. Study power for main gene effects

Genotype prevalence	Odds ratio	Study power, all participants (n_{cases}=1274, $n_{controls}$=1343)	Study Power, white participants (n_{cases}=952, $n_{controls}$=1089)	Study Power, African-American participants (n_{cases}=322, $n_{controls}$=254)
10%	1.2	30%	25%	8%
	1.5	92%	84%	27%
	1.8	>99%	>99%	53%
	2.0	>99%	>99%	70%
20%	1.2	48%	39%	12%
	1.5	99%	97%	43%
	1.8	>99%	>99%	76%
	2.0	>99%	>99%	90%
30%	1.2	58%	48%	14%
	1.5	>99%	>99%	52%
	1.8	>99%	>99%	85%
	2.0	>99%	>99%	95%

Calculations were done in Episheet.xls as described in Rothman & Boice (193).
Assumptions: α=0.05 and case/control ratio=1.054

Specific Aim 2: *Interactive effects of genotype and alcohol exposure on power of logistic regression analyses:*

Table 11 contains estimates of study power for detecting alcohol-SCCHN multiplicative interaction. At an α level of 0.05 the study would have >90% power to detect 2X gene-alcohol multiplicative interactions, assuming:

1. The prevalence of heavy drinking at 1 year prior to diagnosis for cases, and reference date for controls, defined as ≥6 drinks per day, is 19.8% (as shown in Table 8).
2. The prevalence of SCCHN in low drinkers (<6 drinks per day) is 0.1%.
3. An OR of 1.5 to 3.0 was assumed for the independent effect of heavy drinking on SCCHN risk, based on the literature review of risk in non-smokers (Table 3).
4. The at-risk genotype had a prevalence of 20% or greater. There would be inadequate power (<80%) when prevalence is 10% or less.

When designing this study I had planned to test for departures from multiplicative interaction, but during the analysis phase, decided to test for departures from additive interaction instead, as it is thought to be more biologically relevant. At that time I re-calculated our power for additive interaction, using the characteristics of my CHANCE-derived study population to set parameters.

Table 12 contains estimates of study power for detecting alcohol-SCCHN additive interaction. At an α level of 0.05 the study would have >80% power to detect 2.5X gene-alcohol additive interactions in 1227 cases and 1325 controls, assuming:

1. The prevalence of non-drinkers was 16%, and the proportion of drinkers in each category of lifetime drinking was about 28%.
2. The minimum excess OR to be detected was set to 2.5. For additive interaction, this is equal to: $(OR_{11}-1) / (OR_{10}-1 + OR_{01}-1)$.

We had sufficient power to detect 2.5x additive interactions only when the genotype distribution was either 50/50 or 40/60. This was true for 39 of our 64 SNPs (60%), including all but one of our *ALDH2* SNPs and all our *CYP2E1* SNPs. As a result we were not able to detect additive interaction, even if present, for most of our SNPs, unless the excess OR was larger than 2.5.

(Refer to Table 6 for a list of studies that investigated gene-alcohol interaction.) Seven studies reported *ADH1B* interaction with alcohol exposure; two reported greater than multiplicative interaction, four reported interactions greater than additive though less than multiplicative, and one reported interaction with no information on type. Only one study reported no interaction of *ADH1B* with alcohol. Three studies reported *ADH1C* interaction with alcohol exposure; all indicated greater than both additive and multiplicative interaction. However four studies reported no interaction of *ADH1C* with alcohol. Four studies reporting *CYP2E1* interactions with alcohol exposure reported greater than 2x multiplicative (1 study),

2x multiplicative (1 study), and greater than additive, less than multiplicative (one study), but seven studies reported no interaction with alcohol. It was often difficult to determine the type of gene-alcohol interaction (greater than additive, greater than multiplicative, or somewhere in between) because often the main effect OR for alcohol exposure was not provided. Our study had sufficient power to detect 2x multiplicative interaction when the variant genotype is present in at least 15% of participants, and sufficient power to detect 2.5x additive interaction when the variant genotype is present in at least 40% of participants.

Table 11. Study power for detecting gene-alcohol multiplicative interactions

Genotype prevalence	Exposure prevalence	Exposure OR	Genotype OR	2x Interaction OR (in the doubly exposed)	Study power
30%	19.8%	3.0	1.5	9.0	96%
		2.5		7.5	96%
		2.0		6.0	96%
		1.5		4.5	96%
20%	19.8%	3.0	1.5	9.0	92%
		2.5		7.5	92%
		2.0		6.0	92%
		1.5		4.5	91%
10%	19.8%	3.0	1.5	9.0	74%
		2.5		7.5	74%
		2.0		6.0	74%
		1.5		4.5	73%

Calculations were done in POWER version 3.0 from the National Cancer Institute (194).
Assumptions: α=0.05, case/control ratio=1.054

Table 12. Study power for detecting gene-alcohol additive interactions

Genotype prevalence	OR_{01} (only exposed to alcohol, not minor allele) = 2.0 OR_{10} (only exposed to minor allele, non-drinker) = 1.0	OR_{01} (only exposed to alcohol, not minor allele) = 2.0 OR_{10} (only exposed to minor allele, non-drinker) = 0.6
50%	0.94	0.81
40%	0.92	0.78
30%	0.87	0.73
20%	0.76	0.61
10%	0.49	0.39

Calculations were done in POWER version 3.0 from the National Cancer Institute (194).
Assumptions: α=0.05, case/control ratio=1.08

Specific Aim 3: *Main effects of genotype exposure on power of survival analysis using cases only (195, 196)*

At an α level of 0.05, assuming frequency of all-cause deaths at 0.394 (based on the actual number of deaths in my study through 2008 from any cause), the study had sufficient power to detect a hazard ratio (HR) of 1.3 or greater for death from any cause (Table 13). Because the frequency of deaths from head and neck cancer is lower – 17.5% – we would have sufficient power to detect a hazard ratio of 1.5 or greater for cancer-specific death.

The only study that has investigated overall and disease-specific hazard of survival in SCCHN patients (150) reported "hazard of survival" ratios of 0.3 to 0.4 which would correspond to a greater than 3.0 "hazard of death" ratio. This study would have good power to detect HRs of 2.0 to 3.0.

Table 13. Study power for main gene effects on survival

Hazard ratio	Death rate = 0.394 (death from any cause)	Death rate = .175 (death from head and neck cancer)
1.2	48%	24%
1.3	78%	45%
1.4	94%	65%
1.5	99%	81%
2.0	100%	100%

Calculations were done in PASS 11 (195)
Assumptions: α=0.05, number of cases=1227

2.5 ANALYSIS PLAN

2.5.1 Issues for all analyses

For use in all regression models, individual estimates of African ancestry were calculated from 157 AIMs using maximum likelihood methods (as described in (184, 189)). AIMs were selected to differentiate only between African and European ancestry, so participants with "Other" race-ethnicity were excluded from the primary analysis. Allele and genotype frequencies, and tests of HWE, were calculated using SAS Genetics (197). To identify ancestry outliers in the populations, I used a box test on the ancestry estimates (198). For application of this method the upper and lower fourth of the ordered observations are determined. Outliers are those values that exceed the upper or lower fourth by more than $1.5f_s$, where f_s is the upper fourth minus the lower fourth (75% #–25% #). When testing the ancestry variable during backwards elimination from a full model, ancestry was not important to any of the SNP models, regardless of whether I used all calculated ancestry values or discarded the outliers identified by this method.

Haplotypes have some advantages over single SNP tests of association, in that they can capture information about multiple causal/interacting *cis*-mutations in the same gene that are relevant to disease risk or survival. Although some authors have suggested that haplotype analysis is much less powerful than single SNP analysis when looking for gene-disease association (199, 200), others (e.g. (201, 202)) suggest that haplotype analysis can be more powerful when the true causal SNPs are not typed (as in this study, since most are tag SNPs) and when there are multiple disease-causing alleles or even interactions. Because haplotyping can only be done on SNPs that are in high linkage disequilibrium (LD), I verified that the SNPs for each gene are indeed highly correlated using HaploView 4.2 (203) to make LD plots in African-American and white controls.

There are two main classes of statistical methods currently used for estimating haplotypes: expectation maximization (EM) based on likelihood theory, and Bayesian estimators (204). For this research I employed Haplo.stats (205), a widely used EM method implemented in S-Plus/R, to estimate haplotypes and test their association with SCCHN risk and survival. Haplo.stats assumes that all subjects are unrelated and that ambiguous haplotypes exist due to unknown linkage phase of genetic markers. It permits inclusion of environmental covariates and specification of haplotype effects as codominant, additive, dominant, or recessive, and compares each haplotype to a common referent. Haplo.stats also provides general linear models functionality for regression of the outcome on haplotypes, and can easily include covariates and interactions in the models.

Correction for multiple statistical tests with highly correlated exposures. Both Bonferroni correction and controlling for false discovery rate (FDR) (206) assume independent test statistics, and although this assumption is false when testing multiple tag SNPs located in the same gene, many researchers use these techniques in an attempt to control Type I error. It has been noted that in settings involving a few tests (e.g. 10-20) up to a few hundred tests, Bonferroni correction is often robust in spite of linkage disequilibrium between some of the markers (207). But in settings with more tests, a Bonferroni correction can be too conservative, thereby causing the false negative rate to be too high. In order to better balance risk of Type I and Type II errors and avoid the extreme conservatism of the Bonferroni and related Family-wise error (FWE) correction methods, some researchers prefer controlling the False Discovery Rate (FDR) (206). The interpretation of p=0.05 in FDR correction means that 5% of the genes considered statistically significant after correction are false positives. FDR can be estimated by PROC MULTTEST in SAS®. In this method (208), the p-values of each SNP are ranked from smallest to largest, and all except the largest p-value are corrected by multiplying by the total number of polymorphisms being tested divided by the p-value's rank; thus the FDR correction becomes more stringent as the p-

value decreases. In contrast to FWE and FDR error correction methods that assume independence of the tests, simulation-based resampling techniques attempt to control for multiple statistical tests by comparing observed p-values with p-values calculated from simple repeated perturbations of the data; of these methods, the permutation test is the most widely used (207). However, permutation methods can't easily incorporate covariates. One recent (2007) study reported a method of computing p-values adjusted for correlated tests that gives similar results to permutation methods with much less computation time and more ease in including covariates (209). Interestingly, haplotyping is also sometimes used to reduce the number of tests performed, although it was not used primarily for that purpose in this project. Recently, several Bayesian methods have been proposed for appropriately controlling Type I error when handling highly correlated exposure data: the false positive report probability (FPRP) by Wacholder et al. in 2004 (210), and an extension of this method referred to by the author as Bayesian false-discovery probability (BFDP) proposed by Wakefield in 2007 (211). FPRP allows the prior probability of the hypothesis, study power, tolerance for a false-positive decision, as well as the p-value, to affect decisions about which findings are noteworthy, and unlike most Bayesian methods, dichotomizes findings into noteworthy and non-noteworthy categories. The BFDP extension of FPRP is designed to reduce Type I errors even further than FPRP and was designed especially for genome-wide association studies with multiple stages, substituting its Bayes factor for p-values in determining which SNPs to select for replication analysis.

For this dissertation project, because the number of statistical tests is in the middle ground where FEW and FDR error corrections might be robust, I compared results of Bonferroni, FDR, and the Bayesian FPRP method, and planned to discuss any differences. Because the results and conclusions were essentially the same for all three methods, I have reported only Bonferroni-corrected p-values for this research.

Bias in measuring alcohol consumption. Reviews on the topic of validity of self-reported alcohol consumption report a confusing set of results (212, 213). This results largely from the fact that there is no “gold standard” to which to compare self-reported alcohol consumption. Researchers investigating self-report of addictive behaviors have concluded that validity is increased if sensitivity of the information desired is minimal, the questionnaire is methodologically sophisticated, the interviewer is skilled at minimizing negative task characteristics (214, 215), and the recall period is sufficiently long (216). Personal characteristics of the respondent are felt to be less important than these factors (216).

Because CHANCE used trained nurse interviewers to ask a standardized script of questions about legal behaviors (smoking and drinking) over a long period of time (lifetime), and because our classifications were quite broad (tertiles of lifetime alcoholic beverage consumption; number of years the subject smoked cigarettes) I believe our measures of smoking and drinking did not significantly bias our effect estimates.

2.5.2 Analyses for specific aims

Specific Aim 1: “Estimate the main effects of each genetic polymorphism (SNPs and haplotypes) on the risk of developing SCCHN.”

Single SNP associations with SCCHN risk, and risk of each of the five sub-sites, were modeled in SAS® using conditional logistic regression. Unconditional logistic regression models, containing the frequency matched variables of age category, sex, race, and their two- and three-way interactions, often did not converge; therefore I used conditional logistic regression models instead.

Almost nothing is known about the genetic models that govern the effects of these SNPs, so I began analysis using a co-dominant genetic model since it provides the most information about inheritance patterns. However, because so many of the SNPs had

insufficient numbers of subjects who were homozygous for the minor allele, which precluded calculation of precise effect estimates, I used instead a dominant genetic model in all regression analyses of SNP effects. These models compared cancer odds in carriers of any minor allele (either 1 or 2 copies) versus those homozygous for the major allele.

A Bonferroni correction for 64 statistical tests (one for each SNP) was applied to the single-test alpha level of 0.05 (see more details, below). Initial/full models included tobacco exposure, proportion of African ancestry, and other potential socioeconomic, health, and exposures as described above. However main effects models for SNPs did not include alcohol consumption, because it may be in the causal pathway between SNP and cancer outcome.

Haplotypes were constructed separately for whites and African-Americans using default D' blocks created in Haploview 4.2 from the study genotyping data. The algorithm (217) constructs 95% confidence limits on D' and each comparison is defined as either "strong LD", "inconclusive" or "strong recombination". A block is created if 95% of informative comparisons are in "strong LD". The method ignores markers with minor allele frequency less than 5%. Assignment of most likely haplotype for individuals with ambiguous haplotype was done using an EM algorithm in haplo.stats (205), with minimum counts set to 10. Estimation of odds ratios and confidence intervals was accomplished using general linear methods implemented in haplo.stats 1.4.4. Bonferroni correction for multiple statistical tests was applied to beta coefficient p-values (13 tests – for 13 haplotypes -- for whites, 12 tests for African-Americans).

Specific Aim 2: "Estimate the interactive effect of alcohol exposure with each polymorphism on the risk of developing SCCHN."

I had originally planned to test for both additive and multiplicative interaction between SNPs and alcohol consumption, but because additive interaction is believed to be more

relevant to biological mechanisms, I decided to calculate and report only measures of additive interaction.

Departures from additive interaction were evaluated by computing interaction contrast ratios (ICRs), confidence intervals, and Bonferroni-corrected p-values. ICRs were calculated using cancer odds ratios of subjects in three categories: (1) the highest drinking category and no minor allele (i.e., those singly exposed to drinking only – OR_{01}); (2) never-drinkers with at least one minor allele (i.e., those singly exposed to only the variant allele – OR_{10}); and (3) subjects in the highest drinking category and at least one minor allele (i.e., those doubly exposed to both alcohol and the variant allele – OR_{11}), compared to never-drinkers homozygous for the major allele (i.e., the referent group that was not exposed to either the variant allele or to drinking – $OR_{00} = 1.0$). ICR is calculated as follows: $ICR=OR_{11} - OR_{01} - OR_{10} + 1$. ICRs significantly different from zero indicate departure from additive interaction.

This study's power to detect interactions was marginal even when using the entire study population. Because haplotypes were constructed and analyzed separately for each race, there was insufficient statistical power to detect within-race interactions, so I did not test haplotypes for interaction with lifetime alcohol consumption.

Specific Aim 3. "Compare death rates of cases with different gene polymorphisms in order to identify survival differences associated with particular polymorphisms."

All deaths that occurred in CHANCE subjects on or before December 31, 2008 were classified as all-cause events. If the initial three digits of the first-listed NDI cause of death code were head and neck cancers (C01-C06, C09, C10, C12-C14, or C32), the death was further classified as a cancer-specific death. For all-cause death, months of survival were calculated by subtracting the date of diagnosis from the date of death, and subjects who were still alive on 12/31/2008 were censored on that date. For cancer-specific death,

months of survival were calculated by subtracting the date of diagnosis from the date of cancer death, while subjects who died of another cause were censored as of the date of death; subjects who were still alive on 12/31/2008 were censored on that date.

To ascertain SNP effects on all-cause and cancer-specific survival after SCCHN diagnosis, I used a dominant genetic model in SAS® 9.2 to calculate Kaplan-Meier survival curves with log-rank tests, and hazard ratios from Cox proportional hazards models (218). Tumor, node, and metastasis stage at diagnosis, presence/absence of angiolymphatic invasion and selected co-morbid diseases, dichotomous treatment variables (surgery, chemotherapy, and radiation), age, sex, self-reported race, percent African ancestry, tobacco exposure, and socioeconomic and health variables were entered into initial models, and covariates were tested and backwards-eliminated from full models by the 10% change rule as described above. Including TNM summary stage as an ordinal variable to control for stage, instead of separate T, N, and M stages, did not substantially change effect estimates. Alcohol consumption was not included in these main effects models because it may be in the causal pathway between SNP and cancer. I also did the same tests separately for whites and African-Americans to determine whether effect estimates were similar for the two races.

2.6 STRENGTHS AND LIMITATIONS

A major advantage of CHANCE is that both cases and controls were drawn from a large population at risk for SCCHN. This improves the generalizability of risk estimates associated with polymorphisms in alcohol-processing and oxidative metabolism genes. The large sample size permitted examination of gene-environment interactions in the overall study with sufficient statistical power for common genotypes. The study gathered extensive questionnaire and clinical data on all major risk factors for SCCHN, and genotyped more than 95% of study participants for polymorphisms in genes that may be related to disease

incidence and survival. Sufficient tag SNPs were genotyped in most of the genes such that haplotype analysis to comprehensively assess common variation over the entirety of each gene can be performed, with the goal of identifying risk haplotypes. Outcome, stage at diagnosis, and treatments were verified with medical records to reduce misclassification of outcome and clinical covariates. African Americans, an under-studied population, make up 22% (615/2785 questionnaires) of participants.

Limitations included low response rate of potential controls, and different response rates between white and African-American controls (50% versus 35%, respectively). The result was fewer African-American controls than there were African-American cases. Because response rate of cases was much higher than that of controls (76% versus 46%, respectively), participation bias may be present, and it is difficult to estimate how that might affect frequency of genotypes. Also, this study excluded enrolled cases who died without personally completing the questionnaire and providing a biological sample, thus eliminating the most rapidly fatal cases. If incidence and survival effects due to SNP polymorphisms are more pronounced in subjects with rapidly-fatal disease, this selection bias would tend to skew our effect estimates toward the null.

A second limitation is that most covariates were measured by self-report at baseline, which can introduce misclassification of risk factors. However, nurse interviewers received extensive training to ask questions in a consistent manner of each subject, in order to reduce errors introduced by the way in which questions were phrased. Also, CHANCE did not measure alcohol and tobacco use after baseline, so could not address whether changing these behaviors after diagnosis improves survival.

Finally, genotyping only tag SNPs, although less expensive than typing other variants, is likely to miss rare variants.

CHAPTER 3

JOINT EFFECTS OF ALCOHOL CONSUMPTION AND POLYMORPHISMS IN ALCOHOL AND OXIDATIVE STRESS METABOLISM GENES ON RISK OF HEAD AND NECK CANCER

3.1 ABSTRACT

Single nucleotide polymorphisms (SNPs) in alcohol metabolism genes are associated with squamous cell carcinoma of the head and neck (SCCHN), and may influence cancer risk in conjunction with alcohol. Genetic variation in the oxidative stress pathway may impact the carcinogenic effect of reactive oxygen species produced by ethanol metabolism. We hypothesized that alcohol interacts with these pathways to affect SCCHN incidence.

Interview and SNP genotyping data were obtained from 2552 white and African-American subjects (1227 cases, 1325 controls) from the Carolina Head and Neck Cancer Epidemiology study, a population-based case-control study of SCCHN conducted in North Carolina from 2002-2006. We estimated odds ratios and 95% confidence intervals, adjusting for matching variables (age, sex, race) and duration of cigarette smoking.

Two SNPs were associated with altered SCCHN risk: *ADH1B* rs1229984 A allele (OR=0.7, 95% CI=0.6-0.9) and *ALDH2* rs2238151 C allele (OR=1.2, 95% CI=1.1-1.4). Three were associated with sub-site tumors: *ADH1B* rs17028834 C allele (larynx, OR=1.5, 95% CI=1.1-2.0), *SOD2* rs4342445 A allele (oral cavity, OR=1.3, 95% CI=1.1-1.6), and *SOD2* rs5746134 T allele (hypopharynx, OR=2.1, 95% CI=1.2-3.7). Four SNPs in alcohol metabolism genes showed evidence of additive interaction with level of alcohol

consumption: *ALDH2* rs2238151, *ADH1B* rs1159918, *ADH7* rs1154460, and *CYP2E1* rs2249695. No alcohol interactions were found for oxidative stress pathway SNPs.

Previously unreported associations of SNPs in *ALDH2*, *CYP2E1*, *GPX2*, *SOD1*, and *SOD2* with SCCHN and sub-site tumors provide evidence that alterations in alcohol and oxidative stress pathways influence SCCHN carcinogenesis, and warrant further investigation.

3.2 INTRODUCTION

Head and neck cancers typically include tumors of the oral cavity, pharynx, larynx, nose, nasal cavity and sinuses, and esophagus. This study focuses specifically on squamous cell cancers of the oral cavity, pharynx, and larynx (SCCHN).

There were an estimated 49,260 new cases and 11,480 deaths from oropharyngeal and laryngeal cancer in the U.S. in 2010 (219). Globally in 2008, oral cavity tumors were among the top 10 incident cancers in men world-wide, and among the top 10 fatal cancers in men in developing countries (220).

SCCHN incidence is higher in men than women, and, in the U.S., in African-Americans and those of low socioeconomic status. A large proportion of this disparity is due to the higher incidence of laryngeal tumors among African-American men (221).

SCCHN is strongly associated with smoking tobacco products and drinking alcoholic beverages, and recently with human papillomavirus infection. It is estimated that 75% of oropharyngeal cancer in the US is due to cigarette smoking and alcohol consumption (56). The effect of these exposures varies by anatomic sub-site, with smoking more associated with laryngeal tumors, and drinking with oral cavity tumors. However only a small fraction of people exposed to these carcinogens will develop SCCHN, suggesting that other factors, including genetic, must be considered. Inherited genetic variation in the metabolism of alcohol has been suggested as a potentially important contributor to SCCHN risk.

Investigation of the association between single nucleotide polymorphisms (SNPs) and SCCHN may help to identify high-risk groups and clarify carcinogenesis pathways.

Many previous studies of the major genes in the alcohol metabolism pathway (*ADH* family, *ALDH2*, *CYP2E1*) have been limited by sample size, and, with one exception (152), none have included a significant percentage of African-Americans. Further, few studies have examined the influence of genetic variation in oxidative stress pathways (e.g., *SOD*, *GPx*, *CAT*). We examined the association between SNPs and haplotypes of genes in the alcohol metabolism and oxidative stress pathways and SNP-alcohol interactions using data from a large North Carolina (N.C.) population-based case-control study of SCCHN, including 22% African-Americans.

3.3 METHODS

3.3.1 Subject enrollment

The Carolina Head and Neck Cancer Epidemiology study (CHANCE) is a population-based case-control study upon which these analyses are based (222).

All cases of squamous cell carcinoma of the oral cavity, pharynx, and larynx diagnosed in 46 N.C. counties between 1/1/2002 through 2/28/2006 were eligible for enrollment. Rapid case identification was conducted by the N.C. Central Cancer Registry. CHANCE cases included ICD-O-3 topography codes C0.00-C14.8, and C32.0-C32.9, excluding salivary gland (C07.9, C08.0-C08.9), nasopharynx (C11.0-C11.9), nasal cavity (C30.0), and nasal sinuses (C31.0-C31.9). ICD-O-3 morphology codes included were 8010/3, 8051/3, 8083/3, 8071/3, 8072/3, 8073/3, 8074/3, and 8076/3. Benign tumors, carcinomas *in situ*, papillary carcinomas, and adenoid carcinomas were excluded. This analysis further excluded 21 lip cancers (C00.3-C00.9, C14.2), 46 of “other” race, and 96 without genotyping data, producing a study composition of 1227 cases and 1325 controls.

Potentially eligible controls from the same counties as cases were identified through N.C. Department of Motor Vehicles records. Controls were frequency-matched to cases using random sampling with stratification on age, race, and sex.

Trained nurse-interviewers conducted an in-person interview with each subject. For this analysis only self-reported, non-proxy data were included. Questions were asked about demographics (age, sex, race-ethnicity, education, income, health insurance), tobacco use, drinking of alcoholic beverages, diet, oral health, medical history, and family history of cancer.

Blood samples were obtained by nurse-interviewers trained in phlebotomy. If the subject was not willing or able to consent to the blood draw, they were asked to contribute a buccal cell sample via mouthrinse.

Written informed consent was obtained from all subjects. The study was approved by the Biomedical Institutional Review Board at the University of North Carolina at Chapel Hill.

3.3.2 Outcome, exposure, and covariate measurement

Outcome. Case tumors were classified into anatomic sub-sites according to the following 5 ICD-O categories used by the International Head and Neck Cancer Epidemiology Consortium (52): (1) oral cavity: C02.0-C02.3, C03.0, C03.1, C03.9, C04.0, C04.1, C04.8, C04.9, C05.0, C06.0-C06.2, C06.8, and C06.9; (2) oropharynx: C01.9, C02.4, C05.1, C05.2, C09.0, C09.1, C09.8, C09.9, C10.0-C10.4, C10.8, and C10.9; (3) oral cavity-oropharynx-hypopharynx NOS: C02.8, C02.9, C05.8, C05.9, C14.0, C14.2, and C14.8; (4) hypopharynx: C12.9, C13.0-C13.2, C13.8, and C13.9; and (5) larynx: C32.0-C32.3, and C32.8-C32.9.

Alcohol and tobacco use. Questions about alcohol use were designed to estimate lifetime history of alcohol consumption, and usual consumption of each beverage type, prior to the year before diagnosis. Questions were asked about beer, wine, and hard liquor separately as follows: (1) Did you drink [beer/wine/hard liquor]? (2) At what age did you

start? (3) At what age did you stop? (4) For how many years did you drink [beer/wine/hard liquor] during this period? (5) How much [beer/wine/hard liquor] did you usually drink? Per day/week/month/year? (6) What size did you usually drink?

To efficiently estimate odds ratios for the interactive effect of genotype and alcohol consumption, a single variable combining all three types of alcohol consumption was derived. As frequency of drinking has demonstrated stronger associations with SCCHN than duration (51), a single frequency measure that included all types of alcoholic beverages would have been optimal, although unavailable in CHANCE. We instead derived a lifetime measure of alcohol intake, in milliliters, for beer, wine, and liquor combined. Using splines, we confirmed that tertiles best represented the risk associated with alcohol intake.

The primary tobacco exposure covariate selected was continuous duration of cigarette smoking rounded to whole years. Dichotomous variables representing additional potential tobacco confounders were: ever use of non-cigarette tobacco products, ever exposed to environmental tobacco smoke (ETS) at work, and ever exposed to ETS at home.

SNPs. For this project, 75 SNPs (69 tag SNPs and 6 candidate SNPs found in prior studies to be associated with cancer incidence or survival, or alcohol dependence) were selected in 12 genes that are part of two metabolic pathways: *ADH1B*, *ADH1C*, *ADH4*, *ADH7*, *ALDH2*, and *CYP2E1* in the alcohol metabolism pathway in the upper aerodigestive tract; and *CAT*, *SOD1*, *SOD2*, *GPX1*, *GPX2*, and *GPX4* in the oxidative stress metabolic pathway. Tag SNPs, chosen to represent the genetic variation within each of the 12 candidate genes (gene and 2000 bp upstream and downstream) were selected using the Genome Variation Server (223), using SNPs that were polymorphic in either CEU or YRI HapMap Release 2 (unrelated only), and with the following parameters: allele frequency cutoff 10%, 0.8 R^2 threshold minimum for variations to belong to the same cluster, 85% minimum data coverage for tag SNPs, 70% minimal data coverage for a variation to be potentially clustered with others.

To control for potential population stratification, we selected 157 ancestry informative markers (AIMs) to maximize (1) the difference in allele frequencies (delta) between European and African populations in the HapMap data (CEU versus YRI), and (2) the Fisher's information criterion (FIC). AIMs were prioritized based on having the highest delta and FIC values in the following order: 90% European/10% African, 10% European/90% African, and 50% European/50% African. This allowed AIMs to represent the entire expected ancestral distribution of the study population. Individual estimates of percentage African ancestry were calculated from 145 successfully genotyped AIMs using maximum likelihood estimation (MLE) methods previously described (184, 189, 224). AIMs were chosen to differentiate only between African and European ancestry, so we derived individual ancestry estimates for the two groups that summed to 1.0.

DNA was extracted from blood or buccal samples collected at time of interview. Genotyping was done by the University of North Carolina at Chapel Hill, Mammalian Genotyping Core Facility, using the Illumina GoldenGate genotyping assay with Sentrix Array matrix and 96-well standard microtiter plates.

Haplotypes using the study data were constructed separately for African-Americans and whites using default D' blocks in Haploview 4.2. The algorithm (217) constructs 95% confidence limits on D' and each comparison is defined as either "strong LD", "inconclusive" or "strong recombination". A block is created if 95% of informative comparisons are in "strong LD". The method ignores markers with minor allele frequency less than 5%. Assignment of most likely haplotype for individuals with ambiguous haplotype was done using an EM algorithm in haplo.stats (205), with minimum counts set to 10. Estimation of odds ratios and confidence intervals was accomplished using unconditional logistic regression implemented in haplo.stats 1.4.4.

SES, oral health: Dichotomous variables representing additional potential confounders were: covered by health insurance on reference date, ever had a routine dental

visit, ever had a loose permanent tooth due to disease, ever used mouthwash, family history of SCCHN, household poverty as defined by federal guidelines for both income and number of persons supported, and highest attained education level.

3.3.3 Statistical analysis

Odds ratios for the independent effects of SNPs and alcohol, and their interactive effects, were computed using conditional logistic regression methods implemented in SAS® 9.2. Odds ratios for the main effects of haplotypes were computed using unconditional logistic regression methods implemented in haplo.stats 1.4.4.

A dominant genetic model (at least one minor allele versus referent of no minor alleles) was used for SNPs instead of the general model because for many of the SNPs, the number of subjects homozygous for the minor allele was too small to permit estimation of precise effect estimates.

Potential covariates were eliminated using step-wise backwards elimination, comparing each reduced model to the full model that included all covariates listed in Table 14. Note that no collinearity was noted between any variables in the full model, with one exception as described below. If a covariate did not change the ln(OR) for any SNP by a difference of at least 0.10, it was eliminated from all models. The final models contained the genetic variant (single SNP or haplotype), coded for a dominant genetic model for SNPs, additive genetic model for haplotypes; categorized lifetime drinking variable; continuous smoking duration variable; and the SNP*drinking interaction term (for SNP genetic variants, but not for haplotypes). The conditional logistic regression used for SNPs by definition takes into account the matching variables of age category, sex, and race. The unconditional logistic regression models used for haplotypes (for each race separately) included, as covariates, sex, age, and their 2-way interaction. Ancestry was not important for the genetic variants studied, probably because self-reported race was already included (as a matching

variable). The ancestry variable also showed evidence of collinearity with race, so for these reasons and for parsimony's sake, ancestry was not included in final models.

A Bonferroni correction was used to adjust p-values to control for Type 1 error introduced by multiple statistical testing, for either 64 tests (for 64 SNPs) or for 12 or 13 tests (for haplotypes).

Departures from additive interaction were evaluated by computing interaction contrast ratios (ICRs) and Bonferroni-corrected confidence intervals for 64 statistical tests. ICRs were calculated using cancer odds ratios of subjects in three categories: (1) the highest drinking category and no minor allele (i.e., those singly exposed to drinking only – OR_{01}); (2) never-drinkers with at least one minor allele (i.e., those singly exposed to only the variant allele – OR_{10}); and (3) subjects in the highest drinking category and at least one minor allele (i.e., those doubly exposed to both alcohol and the variant allele – OR_{11}), compared to never-drinkers homozygous for the major allele (i.e., the referent group that was not exposed to either the variant allele or to drinking – $OR_{00} = 1.0$). ICR is calculated as follows: $ICR=OR_{11} - OR_{01} - OR_{10} + 1$. ICRs significantly different from zero indicate departure from additive interaction.

3.4 RESULTS

3.4.1 Description of study population

Although controls were somewhat older and more likely to be female and white than cases (Table 14), the percentages of cases versus controls in each of the 28 age-sex-race cross-categories, as a proportion of the entire study population, differed by less than 2%. Compared to controls, cases smoked and drank more, and were poorer, less likely to have completed high school or have health insurance, less likely to have ever had a routine dental visit, and more likely to have lost a permanent tooth to disease. Cases were also more likely to have been exposed to ETS at home and work. Mean proportion of African ancestry was slightly higher in cases than controls.

The SNPs in Table 16 that were analyzed for this study were genotyped as part of a larger panel of 1,536 SNPs. Assay intensity data and genotype cluster images for all SNPs were individually reviewed; as a result, 9 tag/candidate SNPs and 12 AIMs (9% of SNPs) were excluded due to inadequate signal or inability to distinguish genotype clusters. Blind duplicates of 109 samples were genotyped to verify call reliability; none of the calls for SNPs analyzed in this study were discrepant. Two tag SNPs, judged to be out of HWE (SAS® PROC ALLELE) in both black and white controls due to an exact p-value <0.001, were eliminated from analysis.

There were no large differences in allele frequencies between cases and controls, when stratified by race (Appendix C). However there are large allele frequency differences between African-Americans and whites.

3.4.2 Cancer risk from alcohol consumption

The odds of developing SCCHN increase monotonically as lifetime alcohol consumption increases (Table 15). Subjects with the lowest level of consumption (up to 133,294 ml) experienced reduced SCCHN odds compared to non-drinkers (OR=0.8, 95% CI=0.6-1.0), a reduction which was driven by laryngeal and oral cavity tumors (OR=0.7, 95% CI=0.4-1.1 and OR=0.4, 95% CI=0.2-0.9, respectively).

Successively higher levels of alcohol consumption were associated with increasing odds. The middle tertile of lifetime consumption (133,294 to 757,550 ml) was associated with 30% higher odds than never-drinkers (OR=1.3, 95% CI=1.0-1.8). The highest lifetime consumption (>=757,550 ml) was associated with tripled odds of SCCHN (OR=3.2, 95% CI=2.3-4.5). In the highest drinking category, all sub-sites were associated with significantly increased odds: doubled odds of laryngeal cancer, and tripled or greater odds for oropharyngeal, oral-cavity-oropharyngeal-hypopharyngeal-NOS cancers, and oral cavity tumors.

3.4.3 Cancer risk from genetic variants

None of the SNP associations with SCCHN or any of the sub-site cancers had a significant Bonferroni-corrected p-value, although five SNPs in *ADH1B*, *ALDH2*, and *SOD2* showed evidence of reduced or increased cancer odds ratios overall and in oral cavity, laryngeal, and hypopharyngeal sub-sites (Table 16; remaining sub-site effects in Appendix D). In *ADH1B*, the rs1229984 A allele was associated with decreased SCCHN odds (OR=0.7, 95% CI=0.6-0.9), and the rs17028834 C allele with increased odds of laryngeal tumors (OR=1.5, 95% CI=1.1-2.0). In *ALDH2*, the rs2238151 C allele showed evidence of association with increased odds of SCCHN (OR=1.1, 95% CI=1.0-1.2), driven largely by an association with laryngeal tumors (OR=1.2, 95% CI=1.1-1.4). In *SOD2*, the rs4342445 A allele was associated with increased odds for oral cavity tumors (OR=1.3, 95% CI=1.1-1.6), and the rs5746134 T allele with odds for hypopharyngeal cancer (OR=2.1, 95% CI=1.2-3.7).

Four haplotypes in *ALDH2*, *CYP2E1*, *GPX2*, and *SOD1* were associated with SCCHN, either in whites or African-Americans, or both (Table 17). One *GPX2* haplotype was significantly associated with decreased odds of SCCHN in whites (OR=0.7, 95% CI=0.5-0.9). Reduced SCCHN odds were found for an *ALDH2* haplotype (OR=0.5, 95% CI=0.3-0.8) in African-Americans and a *CYP2E1* haplotype in whites (OR=0.7, 95% CI=0.6-0.9). The *SOD1* AGGC haplotype was associated with increased odds in whites and reduced odds in African-Americans.

3.4.4 Cancer risk from alcohol interaction with SNPs

Four SNPs showed evidence of possible additive interaction with alcohol consumption (Table 18). All met the following two characteristics: (1) statistically significant or near-significant Bonferroni-corrected p-value (for 64 tests; $p<0.0008$) for departure from additive interaction, and (2) at least 10 cases and 10 controls in each of the three comparison groups OR_{01}, OR_{10}, OR_{11}. For example, rs2238151 in *ALDH2* showed statistically significant evidence of synergistic additive interaction (increased odds for

drinkers with the C allele). Also the T allele at rs1159918 in *ADH1B*, the A allele at rs1154460 in *ADH7*, and the T allele at rs2249695 in *CYP2E1* showed some evidence for greater than additive interaction between alcohol consumption and SNP. (Evaluations of additive interaction with alcohol for remaining SNPs can be found in Appendix E.)

No interactions with alcohol were detected for anatomic sub-sites.

3.4.5 SNP effects by race

SNP effect estimates were similar in whites and African-Americans, with the exception of two SNPs in *SOD1* (rs10432782, rs2070424) that were associated with decreased odds of getting SCCHN in African-Americans and increased odds in whites. These SNPs were part of the *SOD1* haplotype that we found to be associated with different direction of effect in African-Americans and whites. Direction of effect for carrying the minor allele of individual SNPs was consistent with the haplotype: OR (95% CI) for rs10432782 was 0.65 (0.42-1.00) in African-Americans, 1.35 (1.07-1.71) in whites; OR (95% CI) for rs2070424 was 0.52 (0.33-0.83) in African-Americans, 1.47 (1.10-1.97) in whites. The magnitude of the joint effect for the four SNPs found to interact additively with alcohol exposure did not differ between races (data not shown).

3.5 DISCUSSION

Alcohol consumption. Most studies report a strong dose-response relationship between higher levels of drinking, both in lifetime frequency of drinking (e.g. drinks per day) and lifetime alcohol intake (e.g. milliliters of ethanol), and increased cancer risk. However, the type of alcohol beverage associated with cancer risk varies substantially by study. For example, some studies suggest that the most common alcoholic beverage in the study's geographic region studied produces the highest cancer risk (58). There is some evidence that moderate levels of wine consumption produce lower risk than beer and liquor (comparing 16-30 ethanol-standardized drinks per week of each type), but above 30 drinks per week, all types are associated with increased risk (58).

We found a general pattern of association with alcohol intake that is consistent with previous studies (52), with monotonically increasing cancer risk as lifetime consumption increases. Beer and liquor accounted for about 90% of lifetime alcohol consumption in our study population, and those beverages were associated with higher cancer risk than wine consumption (data not shown). This is consistent with the hypothesis that the most commonly drunk alcoholic beverages are associated with the highest risk. We found suggestive evidence of reduced odds of SCCHN for drinkers in the lowest tertile compared to non-drinkers.

Alcohol metabolism genes. Variant *ADH* and *ALDH* alleles coding for either superactive or inactive subunits of ADH and ALDH isozymes are common. Numerous studies in Asian populations have reported an association between several presumably functional variants in *ADH1B*, *ADH1C*, *ADH4*, *ADH7*, and *ALDH2* and SCCHN incidence (105, 106, 111, 113, 114, 127, 130, 131). However, these studies lacked sufficient power to consistently detect interaction between gene and alcohol drinking. In recent years, these variants and others were investigated in larger studies of Europeans, Latin-Americans, and Indians with similar findings (103, 104, 107, 115, 121, 123-126, 128). However, only a few smaller studies examined risk in whites and African-Americans (109, 122, 150, 152, 153), and those included very small numbers of African-Americans.

We discovered an association between rs1229984 in *ADH1B* and SCCHN odds ($OR_{AA+AGvsGG}$=0.7, 95% CI=0.6-0.9). It is the same direction of effect for the A allele as reported in a Japanese study (105)($OR_{GG+GAvsAA}$=2.20, 95% CI=1.46-3.32) and in European Caucasians and Latin-Americans (104) ($OR_{AA+GAvsGG}$=0.56, 95% CI=0.47-0.66), but is the reverse of the effect reported in a few other studies (103, 106, 123) (e.g. (103): $OR_{GG+GAvsAA}$=0.36, 95% CI=0.17-0.77). A recent INHANCE GWAS (149) reported a replicated association of 5 SNPs with upper aerodigestive tract (UADT) cancer (i.e. SCCHN and esophageal cancer), including rs1229984, for which the A allele under a log-additive

genetic model was associated with reduced odds in both the discovery (OR=0.52, 95% CI=0.43-0.64) and the replication phases (OR=0.68, 95% CI=0.60-0.78). The GWAS replication sample included 2,027 CHANCE subjects as 10% of the replication sample.

We found no effect on SCCHN risk of the rs1693482 "slow" allele in *ADH1C* ($OR_{TT+TCvsCC}$=1.05, 95% CI=0.95-1.15). The two largest studies to date of this SNP and SCCHN in European-Caucasians (103) and European-Caucasians and Latin-Americans (104) found increased odds associated with this allele (OR=1.49, 95% CI=1.08-2.05; OR=1.20, 95% CI=1.08-1.32); respectively). Also, all four studies of rs698 "slow" or G allele in Brazilian, Japanese, European-American and Latin-American populations (103-105, 123) reported evidence of 16-38% increased odds. In CHANCE, rs1631460 is in high LD (r^2=0.95) with rs698 in both CEU and YRI HapMap populations, but we found no association between it and SCCHN odds (OR=1.04, 95% CI=0.95-1.14).

No *ADH4* and *ADH7* SNPs were associated with SCCHN, including the rs1573496 C allele in *ADH7* (OR=1.0, 95% CI=0.9-1.1). This is in contrast to the one study that investigated this allele and found it to be associated with reduced odds in Europeans and Latin-Americans (OR=0.7, 95% CI=0.6-0.8) (104).

No *ALDH2* SNPs were associated with SCCHN, and a possible haplotype association was present only in African-Americans (OR=0.5, 95% CI=0.3-0.8). Previous studies of rs886205, a SNP that is polymorphic in Europeans, found conflicting results of no association and increased association for the G allele (103, 115).

Gene interaction with alcohol. We discovered evidence of additive interaction with alcohol of several SNPs in alcohol metabolism pathway genes, although the SNPs we identified were different from those previously reported in the literature. Specifically, we found two SNPs in *ADH1B* and *ADH7* – rs1159918 and rs1154460, respectively – that appear to interact with alcohol. We also found one previously unstudied *ALDH2* SNP, rs2238151, that showed evidence of additive interaction ($OR_{11actual}$=3.3 versus

$OR_{11expected}$=1.4). Whereas previous studies reported that rs1229984 in *ADH1B*, rs4148887 in *ADH4*, rs1573496 in *ADH7*, and rs886205, rs441 (both in high LD with our SNP rs4767939), and rs440 in *ALDH2* interacted with alcohol drinking (103, 104, 106, 113, 123), we did not find evidence for an interaction with these SNPs, probably because we measured alcohol consumption using lifetime alcohol intake instead of drinking frequency.

We also found evidence for greater than additive interaction for *CYP2E1* rs2249695 with alcohol. A recent linkage and association study (225) identified that SNP, among others, to be associated with "tipsiness," or quick response to alcohol challenge. In CHANCE, the T allele (Table 18, last row) appeared to be protective in never-drinkers (OR_{10}=0.6, 95% CI=0.4-0.9) but in the heaviest drinkers the odds of cancer for those with the T allele are 70% higher than expected.

Oxidative stress genes. We found two previously unstudied SNPs in *SOD2* to be associated with specific tumor sites: rs4342445 with oral cavity tumors, and rs5746134 with hypopharynx tumors. One *SOD1* haplotype was associated with SCCHN risk in both races, albeit in different directions, likely related to the effect of three individual SNP effects that differed by race in that gene. Finally, we found a *GPX2* haplotype to be associated with reduced SCCHN risk in whites only. This may indicate that the haplotype is in high LD with an unmeasured causal polymorphism in whites but not in African-Americans.

Only one previous study examined effects on SCCHN incidence of any SNPs in oxidative stress pathways (115); it reported that rs2758346 in *SOD2* (which we did not study) was not associated with SCCHN (OR_{AAvsGG}=0.98, 95% CI=0.78-1.22).

We found no evidence of interaction with alcohol consumption for any oxidative stress SNP.

3.6 CONCLUSIONS

CHANCE is one of the largest studies of head and neck cancer conducted in both African-Americans and whites. This study examined genetic polymorphisms in genes in the alcohol metabolism and oxidative stress biological pathways, and estimated main effects of these polymorphisms along with their interaction with alcohol. We selected tag SNPs to capture most of the variation in the 12 genes studied, rather than studying only missense SNPs within the coding regions.

An inherent limitation of genotyping common tag SNPs, as we did, is that the method is likely to miss rare variants. A second limitation is that, due to small numbers, we could not precisely estimate the interaction between SNPs and alcohol intake for allele frequencies <30% or in relation to anatomic site of tumor.

Our study confirms findings of previous studies that the effects of many polymorphisms in alcohol metabolism pathways are modified by alcohol intake. However, most genetic variants in *ALDH2* and *CYP2E1* have been understudied and warrant additional investigation in light of the new associations that we report.

Our analysis of tag SNPs in *GPX2*, *SOD1*, and *SOD2* has identified several that are associated with SCCHN, hypopharyngeal, and oral cavity tumors. Confirmation of these findings in a variety of populations is warranted.

Table 14. Distribution of non-genetic variables in cases and controls

Variable		Cases (n=1227) n[a]	Cases (n=1227) (col %)	Controls (n=1325) n[a]	Controls (n=1325) (col %)	Chi-square or t-test, unadjusted p-value
Age (years)						
	20-49	239	19.5%	151	11.4%	<0.0001
	50-54	189	15.4%	156	11.8%	
	55-59	207	16.9%	199	15.0%	
	60-64	205	16.7%	202	15.2%	
	65-69	168	13.7%	237	17.9%	
	70-74	135	11.0%	216	16.3%	
	75-80	84	6.8%	164	12.4%	
Sex						
	Male	938	76.4%	924	69.7%	0.0001
	Female	289	23.6%	401	30.3%	
Race						
	White	922	75.1%	1074	81.1%	0.0003
	African-American	305	24.9%	251	18.9%	
Drinking (lifetime ethanol intake in ml)						<0.0001
	never drinkers	117	9.5%	280	21.1%	
	>0 to <134,699	210	17.1%	467	35.2%	
	134,699 to 757,550	318	25.9%	360	27.2%	
	757,550+	505	41.2%	173	13.1%	
	missing	77	6.3%	45	3.4%	
Smoking (duration in years)						<0.0001
	0	160	13.0%	497	37.5%	
	1-19	104	8.5%	266	20.1%	
	20-39	435	35.5%	314	23.7%	
	40-49	295	24.0%	131	9.9%	
	50+	155	12.6%	71	5.4%	
	Missing	78	6.4%	46	3.5%	
Poverty group (at or above, or below, federal poverty guideline)						<0.0001
	>= poverty guideline	816	66.5%	1088	82.1%	
	<poverty guideline	356	29.0%	187	14.1%	
Had a routine dental visit in past 10 years?						<0.0001
	Yes	781	63.7%	1115	84.2%	
	No	438	35.7%	210	15.8%	
Drank alcoholic beverages in prior 20 years?						0.8851
	No	24	2.0%	27	2.0%	
	Yes	1202	98.0%	1298	98.0%	
Ever exposed to environmental tobacco smoke at work						0.0024
	No	316	25.8%	414	31.2%	
	Yes	909	74.1%	911	68.8%	
Ever exposed to environmental tobacco smoke at home						<0.0001
	No	399	32.5%	592	44.7%	
	Yes	827	67.4%	732	55.2%	
Ever used non-cigarette tobacco products						0.1165
	No	754	61.5%	854	64.5%	
	Yes	473	38.5%	471	35.5%	
Had health insurance at reference date						<0.0001
	Yes	1068	87.0%	1250	94.3%	
	No	154	12.6%	74	5.6%	
Highest education level attained						<0.0001
	>=high school	828	67.5%	1123	84.8%	
	<high school	399	32.5%	202	15.2%	
Ever had loose permanent tooth due to disease						<0.0001
	No	765	62.3%	1018	76.8%	
	Yes	455	37.1%	305	23.0%	
Ever regularly used mouthwash						0.8572
	No	502	40.9%	549	41.4%	
	Yes	719	58.6%	775	58.5%	
Family history of SCCHN among 1st degree relatives						0.3848
	No	1206	98.3%	1296	97.8%	
	Yes	21	1.7%	29	2.2%	
Mean % African ancestry		23.8%		19.7%		0.0008

[a] Frequencies for all variables may not sum to the total number of cases and controls, due to missing values

Table 15. Effect of lifetime alcohol consumption on odds of developing cancer

	SCCHN (all 5 sub sites combined)		Oral cavity cancer		oropharyngeal cancer		oral cavity, oropharyngeal, hypopharyngeal cancer NOS		hypopharyngeal cancer		laryngeal cancer	
lifetime alcohol consumption (ml)	# cases/ controls[a]	Adjusted OR[b] (95% CI)	# cases/ controls	Adjusted OR[b] (95% CI)	# cases/ controls	Adjusted OR[b] (95% CI)	# cases/ controls	Adjusted OR[b] (95% CI)	# cases/ controls	Adjusted OR[b] (95% CI)	# cases/ controls	Adjusted OR[b] (95% CI)
missing	74/43		6/43		22/43		15/43		3/43		28/43	
0	117/280	1.00 (ref)	22/280	1.00 (ref)	27/280	1.00 (ref)	23/280	1.00 (ref)	1/280	1.00 (ref)	44/280	1.00 (ref)
>0 to 133,294	209/466	0.75 (0.56-1.02)	19/466	0.45 (0.23-0.89)	69/466	0.87 (0.53-1.44)	48/466	0.93 (0.54-1.62)	5/466	2.25 (0.26-19.84)	68/466	0.67 (0.42-1.08)
133,294 to 757,550	318/360	1.29 (0.95-1.76)	41/360	1.28 (0.68-2.41)	94/360	1.47 (0.89-2.45)	51/360	1.48 (0.83-2.64)	9/360	5.13 (0.61-43.04)	123/360	1.25 (0.78-2.00)
757,550+	505/173	3.22 (2.29-4.52)	84/173	5.34 (2.67-10.67)	120/173	3.47 (2.00-6.04)	86/173	4.49 (2.40-8.39)	36/173	28.74 (3.42-241.40)	179/173	2.26 (1.38-3.70)

[a] Cases and controls do not sum to 1227 and 1325, respectively, because 4 cases and 3 controls are missing information on duration of cigarette smoking.

[b] Conditional logistic regression models conditioned on sex, race, and age category, and adjusted for continuous smoking duration rounded to whole years

Table 16. SNP effects on odds of developing cancer (dominant genetic model)

			SCCHN (all 5 anatomic subsites, combined)				oral cavity cancer				hypopharyngeal cancer				laryngeal cancer			
			# of cases/controls[a]				# of cases/controls				# of cases/controls				# of cases/controls			
Gene	SNP	major/ minor alleles	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]
ALCOHOL METABOLISM GENES																		
ADH1B	rs12507573	C/A	371/386	852/935	0.96 (0.87-1.06)	1.00	57/386	115/935	0.94 (0.78-1.14)	1.00	15/386	39/935	1.03 (0.74-1.42)	1.00	137/386	305/935	0.94 (0.82-1.08)	1.00
	rs1042026	A/G	712/744	511/578	1.00 (0.91-1.10)	1.00	94/744	78/578	1.06 (0.88-1.26)	1.00	36/744	18/578	0.89 (0.65-1.22)	1.00	264/744	178/578	0.96 (0.84-1.10)	1.00
	rs7673353	C/T	1124/1239	99/83	1.01 (0.83-1.23)	1.00	161/1239	11/83	0.88 (0.60-1.30)	1.00	48/1239	6/83	0.84 (0.51-1.40)	1.00	402/1239	40/83	1.04 (0.79-1.36)	1.00
	rs17028834	T/C	1138/1265	85/57	1.28 (1.03-1.59)	1.00	159/1265	13/57	1.50 (1.01-2.25)	1.00	48/1265	6/57	1.71 (1.00-2.95)	1.00	408/1265	34/57	1.49 (1.11-2.01)	**0.54**
	rs1693457	T/C	803/860	419/461	1.01 (0.92-1.11)	1.00	114/860	58/461	0.99 (0.82-1.19)	1.00	36/860	18/461	0.94 (0.69-1.28)	1.00	294/860	148/461	1.02 (0.89-1.17)	1.00
	rs1229984	G/A	1192/1243	31/79	0.72 (0.57-0.91)	**0.35**	169/1243	3/79	0.63 (0.35-1.15)	1.00	53/1243	1/79	0.62 (0.22-1.70)	1.00	429/1243	13/79	0.79 (0.57-1.12)	1.00
	rs1159918	G/T	402/468	821/854	1.07 (0.97-1.18)	1.00	60/468	112/854	1.03 (0.86-1.24)	1.00	14/468	40/854	1.17 (0.83-1.64)	1.00	154/468	288/854	1.05 (0.91-1.21)	1.00
	rs1229982	G/T	699/775	524/547	1.00 (0.91-1.10)	1.00	97/775	75/547	0.95 (0.80-1.14)	1.00	35/775	19/547	0.80 (0.59-1.09)	1.00	241/775	201/547	1.05 (0.92-1.19)	1.00
ADH1C	rs2298753	T/C	1002/1117	221/205	1.12 (1.00-1.27)	1.00	140/1117	32/205	1.12 (0.89-1.41)	1.00	46/1117	8/205	1.16 (0.78-1.75)	1.00	366/1117	76/205	1.10 (0.93-1.31)	1.00
	rs1614972	C/T	539/585	684/737	1.00 (0.92-1.09)	1.00	85/585	87/737	0.93 (0.78-1.10)	1.00	26/585	28/737	0.83 (0.61-1.11)	1.00	192/585	250/737	0.96 (0.85-1.10)	1.00
	rs1391088	C/A	1045/1119	178/203	1.01 (0.89-1.14)	1.00	142/1119	30/203	1.07 (0.85-1.34)	1.00	50/1119	4/203	0.71 (0.41-1.20)	1.00	381/1119	61/203	1.05 (0.88-1.26)	1.00
	rs1693482	C/T	527/585	694/735	1.05 (0.95-1.15)	1.00	75/585	97/735	1.00 (0.83-1.20)	1.00	24/585	30/735	1.18 (0.87-1.60)	1.00	186/585	256/735	1.09 (0.95-1.25)	1.00
	rs1631460	C/G	519/574	703/748	1.04 (0.95-1.14)	1.00	74/574	97/748	0.99 (0.83-1.19)	1.00	24/574	30/748	1.17 (0.86-1.58)	1.00	184/574	258/748	1.08 (0.94-1.23)	1.00
	rs11936869	C/G	611/645	612/676	0.98 (0.90-1.07)	1.00	90/645	82/676	0.96 (0.81-1.14)	1.00	28/645	26/676	0.86 (0.64-1.16)	1.00	221/645	221/676	0.93 (0.82-1.06)	1.00
ADH4	rs29001227	A/T	1119/1253	104/69	1.25 (1.02-1.53)	1.00	157/1253	15/69	1.31 (0.90-1.91)	1.00	48/1253	6/69	1.38 (0.81-2.34)	1.00	401/1253	41/69	1.39 (1.05-1.84)	1.00
	rs1126672	C/T	720/758	503/564	1.01 (0.92-1.10)	1.00	93/758	79/564	1.08 (0.90-1.28)	1.00	38/758	16/564	0.82 (0.60-1.13)	1.00	257/758	185/564	1.06 (0.93-1.21)	1.00
	rs4699710	T/C	640/684	583/638	1.01 (0.93-1.11)	1.00	77/684	95/638	1.12 (0.94-1.33)	1.00	34/684	20/638	0.86 (0.64-1.16)	1.00	230/684	212/638	1.05 (0.93-1.20)	1.00
	rs10017466	T/C	556/632	666/690	1.05 (0.97-1.15)	1.00	67/632	105/690	1.16 (0.97-1.38)	1.00	30/632	24/690	0.90 (0.68-1.20)	1.00	196/632	246/690	1.12 (0.99-1.27)	1.00
	rs1800759	C/A	357/412	866/910	1.02 (0.92-1.12)	1.00	46/412	126/910	1.09 (0.90-1.33)	1.00	13/412	41/910	1.05 (0.75-1.48)	1.00	122/412	320/910	1.09 (0.94-1.25)	1.00
	rs1800761	G/A	750/830	473/492	1.02 (0.93-1.12)	1.00	112/830	60/492	0.92 (0.77-1.10)	1.00	35/830	19/492	0.90 (0.66-1.22)	1.00	276/830	166/492	0.99 (0.87-1.13)	1.00
	rs3762894	T/C	839/906	384/416	0.98 (0.89-1.08)	1.00	126/906	46/416	0.89 (0.74-1.08)	1.00	38/906	16/416	0.89 (0.65-1.23)	1.00	307/906	135/416	0.94 (0.82-1.08)	1.00
ADH7	rs284787	C/T	739/817	484/505	1.06 (0.96-1.16)	1.00	112/817	60/505	0.94 (0.79-1.13)	1.00	29/817	25/505	1.18 (0.88-1.59)	1.00	264/817	178/505	1.09 (0.96-1.25)	1.00
	rs894369	C/G	832/849	391/472	0.92 (0.84-1.01)	1.00	116/849	56/472	0.96 (0.80-1.15)	1.00	42/849	12/472	0.75 (0.54-1.05)	1.00	304/849	138/472	0.93 (0.81-1.06)	1.00
	rs17588403	T/A	837/887	386/435	0.96 (0.87-1.05)	1.00	116/887	56/435	0.95 (0.79-1.14)	1.00	37/887	17/435	0.94 (0.69-1.28)	1.00	297/887	145/435	0.96 (0.84-1.10)	1.00
	rs1154454	T/C	756/808	467/514	0.94 (0.86-1.04)	1.00	117/808	55/514	0.80 (0.66-0.98)	1.00	37/808	17/514	0.69 (0.49-0.97)	1.00	260/808	182/514	1.00 (0.87-1.15)	1.00
	rs1154456	T/C	577/629	646/693	1.03 (0.94-1.13)	1.00	86/629	86/693	0.97 (0.81-1.15)	1.00	26/629	28/693	1.04 (0.78-1.40)	1.00	207/629	235/693	1.07 (0.94-1.21)	1.00
	rs1154460	G/A	329/377	891/944	1.04 (0.95-1.15)	1.00	49/377	123/944	1.03 (0.85-1.25)	1.00	13/377	41/944	1.11 (0.80-1.55)	1.00	118/377	322/944	1.12 (0.97-1.29)	1.00
	rs971074	G/A	943/1027	280/295	1.01 (0.91-1.12)	1.00	131/1027	41/295	1.09 (0.89-1.34)	1.00	39/1027	15/295	1.14 (0.83-1.58)	1.00	348/1027	94/295	1.00 (0.85-1.16)	1.00
	rs1573496	C/G	1056/1117	166/205	0.99 (0.87-1.12)	1.00	147/1117	25/205	1.09 (0.86-1.39)	1.00	48/1117	6/205	0.91 (0.57-1.44)	1.00	383/1117	59/205	1.02 (0.85-1.23)	1.00
ALDH2	rs4767939	A/G	694/806	528/516	1.08 (0.98-1.18)	1.00	101/806	71/516	1.04 (0.87-1.25)	1.00	29/806	25/516	1.10 (0.81-1.49)	1.00	238/806	204/516	1.12 (0.98-1.28)	1.00
	rs2238151	T/C	364/482	855/837	1.13 (1.03-1.25)	**0.84**	51/482	121/837	1.18 (0.97-1.44)	1.00	16/482	38/837	1.00 (0.70-1.42)	1.00	117/482	323/837	1.24 (1.07-1.43)	**0.32**
	rs7312055	G/A	1081/1201	142/121	0.93 (0.78-1.12)	1.00	150/1201	22/121	0.95 (0.67-1.34)	1.00	44/1201	10/121	0.92 (0.57-1.50)	1.00	390/1201	52/121	0.87 (0.67-1.12)	1.00
	rs2158029	G/A	1166/1269	57/53	0.95 (0.75-1.19)	1.00	165/1269	7/53	0.76 (0.48-1.21)	1.00	50/1269	4/53	0.98 (0.52-1.82)	1.00	419/1269	23/53	0.99 (0.72-1.36)	1.00
	rs16941667	C/T	985/1097	238/224	1.10 (0.98-1.23)	1.00	142/1097	30/224	1.11 (0.88-1.40)	1.00	41/1097	13/224	1.24 (0.88-1.75)	1.00	347/1097	95/224	1.17 (0.99-1.37)	1.00
	rs16941669	T/G	984/1084	239/236	1.06 (0.95-1.19)	1.00	134/1084	38/236	1.20 (0.97-1.48)	1.00	45/1084	9/236	1.00 (0.68-1.47)	1.00	358/1084	84/236	1.10 (0.93-1.29)	1.00
CYP2E1	rs3813865	G/C	1098/1215	125/106	1.08 (0.92-1.27)	1.00	155/1215	17/106	1.16 (0.85-1.58)	1.00	48/1215	6/106	1.09 (0.67-1.77)	1.00	393/1215	49/106	1.11 (0.89-1.38)	1.00
	rs3813867	G/C	1139/1237	83/84	0.97 (0.82-1.16)	1.00	160/1237	12/84	0.98 (0.70-1.38)	1.00	51/1237	3/84	0.87 (0.47-1.62)	1.00	405/1237	36/84	1.09 (0.86-1.39)	1.00
	rs8192772	T/C	1023/1114	199/208	1.07 (0.95-1.21)	1.00	142/1114	30/208	1.09 (0.87-1.38)	1.00	44/1114	9/208	1.11 (0.75-1.64)	1.00	363/1114	79/208	1.14 (0.96-1.35)	1.00
	rs915908	G/A	974/1006	249/316	0.93 (0.84-1.04)	1.00	139/1006	33/316	0.92 (0.74-1.14)	1.00	46/1006	8/316	0.77 (0.51-1.15)	1.00	362/1006	80/316	0.87 (0.74-1.02)	1.00
	rs915909	C/T	1164/1271	58/46	1.00 (0.79-1.27)	1.00	164/1271	7/46	0.88 (0.54-1.44)	1.00	50/1271	4/46	1.09 (0.60-1.99)	1.00	421/1271	21/46	0.92 (0.66-1.28)	1.00
	rs7092584	C/T	959/1052	262/269	1.05 (0.95-1.17)	1.00	131/1052	41/269	1.12 (0.91-1.38)	1.00	45/1052	8/269	0.86 (0.58-1.28)	1.00	340/1052	101/269	1.11 (0.96-1.30)	1.00
	rs743535	C/T	966/1071	248/243	1.06 (0.95-1.19)	1.00	134/1071	38/243	1.09 (0.88-1.35)	1.00	40/1071	13/243	1.18 (0.84-1.66)	1.00	342/1071	96/243	1.10 (0.94-1.29)	1.00
	rs2249695	C/T	622/682	600/639	0.93 (0.85-1.03)	1.00	87/682	84/639	0.93 (0.77-1.12)	1.00	24/682	30/639	0.95 (0.68-1.33)	1.00	223/682	219/639	0.93 (0.81-1.08)	1.00
	rs28969387	A/T	1198/1302	25/20	0.98 (0.70-1.37)	1.00	170/1302	2/20	0.69 (0.31-1.56)	1.00	54/1302	0/20	----	1.00	429/1302	13/20	1.08 (0.70-1.68)	1.00
	rs11101812	T/C	1194/1300	26/20	1.11 (0.79-1.56)	1.00	170/1300	2/20	0.79 (0.35-1.77)	1.00	51/1300	3/20	1.79 (0.86-3.71)	1.00	427/1300	14/20	1.38 (0.89-2.13)	1.00
OXIDATIVE STRESS METABOLISM GENES																		
CAT	rs1049982	C/T	521/589	700/726	1.02 (0.93-1.11)	1.00	78/589	94/726	0.95 (0.80-1.14)	1.00	27/589	26/726	0.82 (0.61-1.11)	1.00	182/589	260/726	1.05 (0.92-1.19)	1.00
GPX1	rs8179172	T/A	1172/1271	51/51	0.92 (0.73-1.17)	1.00	166/1271	6/51	0.86 (0.52-1.41)	1.00	51/1271	3/51	0.67 (0.32-1.39)	1.00	423/1271	19/51	0.86 (0.62-1.19)	1.00
	rs1800668	C/T	635/672	588/646	1.04 (0.95-1.14)	1.00	84/672	88/646	1.14 (0.96-1.35)	1.00	29/672	25/646	1.01 (0.76-1.35)	1.00	237/672	205/646	1.03 (0.91-1.17)	1.00
	rs3811699	A/G	609/649	614/673	1.04 (0.95-1.13)	1.00	77/649	95/673	1.16 (0.98-1.38)	1.00	26/649	28/673	1.07 (0.81-1.43)	1.00	230/649	212/673	1.02 (0.90-1.16)	1.00
	rs3448	C/T	648/729	575/593	1.03 (0.95-1.13)	1.00	96/729	76/593	1.02 (0.86-1.21)	1.00	27/729	27/593	1.18 (0.88-1.58)	1.00	222/729	220/593	1.12 (0.99-1.28)	1.00
GPX2	rs11623705	G/T	996/1069	227/253	1.00 (0.89-1.12)	1.00	139/1069	33/253	0.99 (0.79-1.23)	1.00	44/1069	10/253	1.01 (0.70-1.47)	1.00	360/1069	82/253	1.01 (0.86-1.19)	1.00
	rs2412065	G/C	694/733	529/589	0.94 (0.86-1.03)	1.00	96/733	76/589	0.94 (0.79-1.13)	1.00	32/733	22/589	0.81 (0.60-1.10)	1.00	251/733	191/589	0.97 (0.85-1.11)	1.00
	rs2737844	C/T	493/511	725/806	0.89 (0.81-0.99)	1.00	72/511	98/806	0.86 (0.71-1.04)	1.00	20/511	34/806	0.87 (0.63-1.21)	1.00	177/511	265/806	0.91 (0.80-1.05)	1.00
GPX4	rs757229	G/C	300/358	922/963	1.03 (0.93-1.14)	1.00	39/358	133/963	1.08 (0.88-1.32)	1.00	13/358	41/963	0.93 (0.66-1.31)	1.00	103/358	339/963	1.04 (0.90-1.21)	1.00

Table 16. SNP effects on odds of developing cancer (dominant genetic model)

			SCCHN (all 5 anatomic subsites, combined)				oral cavity cancer				hypopharyngeal cancer				laryngeal cancer			
			# of cases/controls[a]				# of cases/controls				# of cases/controls				# of cases/controls			
Gene	SNP	major/ minor alleles	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]
SOD1	rs11910115	A/C	1161/1257	62/65	0.93 (0.74-1.17)	1.00	165/1257	7/65	0.76 (0.48-1.20)	1.00	51/1257	3/65	0.80 (0.41-1.58)	1.00	418/1257	24/65	0.94 (0.69-1.30)	1.00
	rs4998557	G/A	829/959	393/360	1.09 (0.99-1.21)	1.00	123/959	49/360	0.98 (0.80-1.21)	1.00	31/959	23/360	1.28 (0.93-1.78)	1.00	288/959	154/360	1.17 (1.01-1.36)	1.00
	rs10432782	T/G	874/992	348/329	1.07 (0.97-1.18)	1.00	131/992	41/329	0.93 (0.76-1.15)	1.00	34/992	20/329	1.24 (0.90-1.71)	1.00	306/992	136/329	1.13 (0.97-1.31)	1.00
	rs2070424	A/G	1003/1120	220/202	1.06 (0.94-1.20)	1.00	142/1120	30/202	1.02 (0.80-1.30)	1.00	46/1120	8/202	0.85 (0.55-1.31)	1.00	354/1120	88/202	1.09 (0.92-1.29)	1.00
	rs1041740	C/T	682/705	541/617	0.98 (0.90-1.07)	1.00	95/705	77/617	1.00 (0.84-1.20)	1.00	36/705	18/617	0.82 (0.60-1.12)	1.00	250/705	192/617	0.97 (0.85-1.11)	1.00
SOD2	rs4342445	G/A	752/850	471/472	1.11 (1.02-1.22)	1.00	95/850	77/472	1.33 (1.12-1.59)	**0.09**	42/850	12/472	0.84 (0.60-1.18)	1.00	268/850	174/472	1.16 (1.01-1.32)	1.00
	rs2842980	A/T	715/807	508/515	1.02 (0.94-1.12)	1.00	97/807	75/515	1.05 (0.88-1.25)	1.00	31/807	23/515	0.98 (0.73-1.31)	1.00	269/807	173/515	0.97 (0.85-1.10)	1.00
	rs8031	T/A	387/382	836/940	0.98 (0.89-1.08)	1.00	52/382	120/940	1.00 (0.83-1.21)	1.00	16/382	38/940	1.12 (0.80-1.55)	1.00	140/382	302/940	1.02 (0.89-1.17)	1.00
	rs5746134	C/T	1125/1257	98/65	1.20 (0.98-1.48)	1.00	162/1257	10/65	0.98 (0.64-1.50)	1.00	43/1257	11/65	2.14 (1.25-3.67)	**0.34**	409/1257	33/65	1.13 (0.84-1.50)	1.00
	rs2758331	C/A	422/419	801/903	0.99 (0.90-1.09)	1.00	61/419	111/903	0.99 (0.82-1.20)	1.00	17/419	37/903	1.22 (0.88-1.70)	1.00	150/419	292/903	1.04 (0.91-1.20)	1.00

[a] Cases and controls do not sum to 1227 and 1325, respectively, because 4 cases and 3 controls are missing information on duration of cigarette smoking, and because a few subjects lack genotype information for some SNPs

[b] Conditional logistic regression models conditioned on sex, race, and age category, and adjusted for continuous smoking duration rounded to whole years. Odds ratios are for those with one or more copies of the minor allele versus the referent group of those homozygous for the major allele (dominant genetic model).

[c] Bonferroni-corrected for 64 statistical tests; corrected p-values less than 1.00 are highlighted in **bold**.

Table 17. Selected haplotype main effects on SCCHN risk, additive genetic model

Gene (haplotype definition)	haplotype	Race	prevalence (%)	OR (95% CI)[b]
ALDH2	ACAGCT	African-	26%	1.0 (ref)
(rs4767939, rs2238151, rs7312055, rs2158029, rs16941667, rs16941669)	ATGGCT	American	11%	**0.5 (0.3-0.8)**
CYP2E1	GCCC	White	65%	1.0 (ref)
(rs915908, rs7092584, rs743535, rs2249695)	GCCT		10%	0.7 (0.6-0.9)
GPX2	GGC	White	70%	1.0 (ref)
(rs11623705, rs2412065, rs2737844)	GCT		9%	0.7 (0.5-0.9)
	GTAC	White	58%	1.0 (ref)
SOD1	AGGC		6%	1.4 (1.1-1.9)
(rs4998557, rs10432782, rs2070424, rs1041740)	GTAC	African-	52%	1.0 (ref)
	AGGC	American	6%	0.6 (0.4-0.9)

a Criterion for selecting haplotypes for this table: ORs were statistically significant, or nearly so, after Bonferroni correction for multiple testing (13 for White, 12 for African-American). Statistically significant ORs are highlighted in **bold**.

b Unconditional logistic regression models adjusted for matching variables sex, race, age category, and their interactions; and for continuous smoking duration rounded to whole years. The referent group for each OR was the most common haplotype.

Table 18. Additive interactive effects of alcohol with selected SNPs[a]

Gene, SNP, major/minor alleles lifetime ethanol (ml)	homozygous for major allele # cases/ controls	 Adjusted OR (95% CI)[b]	one or two copies of minor allele # cases/ controls	 Adjusted OR (95% CI)[b]	ICR[c] (Bonferroni-corrected CI)
ALDH2, rs2238151, T/C					
never-drinkers	50/95	1.0 (ref)	67/184	**0.7 (0.4-1.1)**	**1.9 (0.1-3.80)**
>0 to <134,699	75/173	0.5 (0.3-0.9)	133/293	0.6 (0.4-1.0)	
134,699 to <757,550	97/138	0.8 (0.5-1.3)	220/222	1.2 (0.8-1.9)	
757,550+	122/65	**1.7 (1.0-2.8)**	381/106	**3.3 (2.0-5.3)**	
ADH1B, rs1159918, G/T					
never-drinkers	49/101	1.0 (ref)	68/179	**0.9 (0.6-1.5)**	1.0 (-0.9, 3.0)
>0 to <134,699	80/168	0.7 (0.4-1.1)	129/298	0.7 (0.5-1.1)	
134,699 to <757,550	107/129	1.1 (0.7-1.8)	211/231	1.3 (0.8-2.0)	
757,550+	140/58	**2.4 (1.4-4.1)**	365/115	**3.3 (2.1-5.4)**	
ADH7, rs1154460, G/A					
never-drinkers	35/65	1.0 (ref)	81/215	**0.6 (0.4-1.1)**	0.9 (-0.6, 2.4)
>0 to <134,699	57/131	0.5 (0.3-0.9)	152/334	0.6 (0.3-0.9)	
134,699 to <757,550	88/111	0.8 (0.5-1.4)	229/249	1.0 (0.6-1.6)	
757,550+	133/54	**1.9 (1.1-3.5)**	371/119	**2.5 (1.5-4.2)**	
CYP2E1, rs2249695, C/T					
never-drinkers	73/136	1.0 (ref)	44/144	**0.6 (0.4-0.9)**	1.2 (-0.6, 3.0)
>0 to <134,699	127/253	0.6 (0.4-0.9)	82/213	0.5 (0.3-0.8)	
134,699 to <757,550	160/178	1.1 (0.7-1.6)	158/181	1.0 (0.6-1.4)	
757,550+	218/92	**2.1 (1.4-3.3)**	286/81	**2.9 (1.8-4.6)**	

[a] Selected SNPs have (1) ICR confidence intervals that either don't include 0 or nearly so, after Bonferroni correction, and (2) genotype information on sufficient numbers of cases and controls (at least 10 each) for calculating each of the three ORs highlighted in **bold** for that SNP**.** If ICR confidence interval appeared significant but numbers of cases and controls were too sparse, SNP was judged to have insufficient evidence of interaction with alcohol, and was not included in this table.

[b] Conditional logistic regression models conditioned on sex, race, and age category, and adjusted for continuous smoking duration rounded to whole years

[c] ICRs were calculated using odds ratios that are highlighted in **bold**. Bonferroni correction of ICR confidence interval for 64 statistical tests.

CHAPTER 4

EFFECTS OF POLYMORPHISMS IN ALCOHOL METABOLISM AND OXIDATIVE STRESS GENES ON SURVIVAL FROM HEAD AND NECK CANCER

4.1 ABSTRACT

Heavy drinking of alcoholic beverages increases risk of developing squamous cell carcinoma of the head and neck (SCCHN). While the mechanisms of alcohol-induced carcinogenesis are complex, alcohol metabolism leads to generation of cytotoxic and mutagenic intermediates including acetaldehyde and reactive oxygen species (ROS). Single nucleotide polymorphisms (SNPs) in alcohol metabolism genes have been associated with SCCHN incidence, and genetic variation in the antioxidant pathways may impact ROS metabolism. We hypothesized that the genes involved in these pathways may influence survival after SCCHN diagnosis.

Interview and SNP genotyping data were obtained from 1227 white and African-American cases from the Carolina Head and Neck Cancer Epidemiology study, a population-based case-control study of head and neck cancer conducted in North Carolina from 2002-2006. Vital status and date and cause of death through 2008 were obtained from the National Death Index. To identify survival differences between alleles of each SNP, we performed Kaplan-Meier log-rank tests and estimated both unadjusted and adjusted hazard ratios and 95% confidence intervals.

Minor alleles of two SNPs in *CYP2E1* – the 'C' allele of both rs3813865 and rs8192772 – were associated with increased hazard of cancer-specific death in both adjusted and unadjusted analyses ($HR_{adjusted}$, 95% CI = 2.09, 1.38-3.18; 1.71, 1.23-2.37),

respectively). Four additional SNPs in *CYP2E1* and four in *GPx1*, *SOD1*, and *SOD2* displayed suggestive differences in allele hazards for all-cause and/or cancer death. No consistent associations with survival were found for SNPs in *ADH1B*, *ADH1C*, *ADH4*, *ADH7*, *ALDH2*, *GPx2*, *GPx4*, and *CAT*.

Polymorphisms in alcohol metabolism and ROS-detoxification genes influence survival in subjects with SCCHN. Previously unreported associations of SNPs in *CYP2E1* warrant further investigation, especially with regard to interaction with type of chemotherapeutic agent used for treatment.

4.2 INTRODUCTION

Head and neck cancers typically include cancers of the oral cavity, pharynx, larynx, nose, nasal cavity and sinuses, and esophagus. There were 49,260 new cases and 11,480 deaths from oropharyngeal and laryngeal cancer in the U.S. in 2010 (219). Globally in 2008, oral cavity tumors were among the top 10 incident cancers in men world-wide, and among the top 10 fatal cancers in men in developing countries (220). Five-year relative survival for laryngeal and oropharyngeal cancer patients averages about 80% for localized cases, 50% for regional cases, and 33% for metastatic cases, with somewhat lower survival for laryngeal compared to oropharyngeal cancers (226).

It is estimated that 75% of new cases of SCCHN in the United States are caused by tobacco use, especially cigarette smoking, and/or drinking of alcoholic beverages (56). In the past few decades the proportional incidence of oropharyngeal tumors associated with carcinogenic human papillomavirus sub-types has risen; such tumors appear to have better prognosis than non-HPV oropharyngeal tumors (227). Multiple studies have reported associations between SCCHN incidence and polymorphisms in alcohol metabolism genes, especially *ADH1B*, *ADH1C*, *ADH4*, *ADH7*, *ALDH2*, and *CYP2E1* ((104, 105, 115, 121-127)). The primary biological mechanism responsible for this effect is hypothesized to be high levels of toxic and mutagenic acetaldehyde, the metabolic intermediate between ethanol

and acetate. Acetaldehyde, when associated with consumption of alcoholic beverages, is classified as a known human carcinogen by the International Agency for Research on Cancer (70). Also, alcohol metabolism through CYP2E1 is known to result in production of excess levels of reactive oxygen species (ROS) (228). The resulting DNA damage from these metabolic pathways may also influence cancer progression and, moreover, variation in pathway genes may modify progression and survival. Therefore, it is also of interest to know whether polymorphisms in genes encoding enzymes protective against oxidative stress (*SOD*, *GPx*, *CAT*) are associated with altered survival in SCCHN patients.

In contrast to the multiple studies on genetic associations with SCCHN incidence, there have been few investigations of the effect of gene polymorphisms in alcohol metabolism and oxidative stress genes on survival of subjects with SCCHN. For example, only two studies investigated whether selected *ADH1C* and *CYP2E1* polymorphisms influence prognosis, and they reported evidence of an association with advanced clinical stages or higher recurrence (124, 150). Further, no studies have examined the influence on post-diagnosis survival of genetic variation in the oxidative stress pathways.

We examined the effect on survival of SNPs in genes in the alcohol metabolism and oxidative stress pathways, using exposure, genetic, clinical, and outcome data from cases included in a large North Carolina (NC) population-based case-control study of SCCHN.

4.3 METHODS

4.3.1 Study population

Cases for this analysis were obtained from the Carolina Head and Neck Cancer Epidemiology study (CHANCE), a population-based case-control study (222).

All cases of squamous cell carcinoma of the oral cavity, pharynx, and larynx diagnosed in 46 N.C. counties between January 1, 2002 and February 28, 2006 were eligible for enrollment. Rapid identification of cases was conducted by the N.C. Central Cancer Registry. CHANCE cases included ICD-O-3 topography codes C0.00 to C14.8, and

C32.0 to C32.9, excluding salivary gland tumors (C07.9, C08.0 to C08.9), nasopharynx (C11.0 to C11.9), nasal cavity (C30.0), and nasal sinuses (C31.0 to C31.9). ICD-O-3 morphology codes included were 8010/3, 8051/3, 8083/3, 8071/3, 8072/3, 8073/3, 8074/3, and 8076/3. Benign tumors, carcinomas *in situ*, papillary carcinomas, and adenoid carcinomas were excluded. This analysis further excluded 21 lip cancers (C00.3-C00.9, C14.2), 26 cases of "other" race, and 115 without genotyping data, producing a study composition of 1227 cases, of which 922 were white/European-American and 305 were black/African-American.

Case tumors were classified into anatomic sub-site according to the following 5 ICD-O categories that are also used by the International Head and Neck Cancer Epidemiology (INHANCE) Consortium (52): (1) oral cavity: C02.0-C02.3, C03.0, C03.1, C03.9, C04.0, C04.1, C04.8, C04.9, C05.0, C06.0-C06.2, C06.8, and C06.9; (2) oropharynx: C01.9, C02.4, C05.1, C05.2, C09.0, C09.1, C09.8, C09.9, C10.0-C10.4, C10.8, and C10.9; (3) oral cavity-oropharynx-hypopharynx NOS: C02.8, C02.9, C05.8, C05.9, C14.0, C14.2, and C14.8; (4) hypopharynx: C12.9, C13.0-C13.2, C13.8, and C13.9; and (5) larynx: C32.0-C32.3, and C32.8-C32.9.

Written informed consent was obtained from all subjects. The study was approved by the Biomedical Institutional Review Board at the University of North Carolina at Chapel Hill.

4.3.2 Outcome assessment

We determined whether death had occurred in study participants by December 31, 2008, and, if so, the date and cause, through linkage with the National Death Index (NDI). The NDI is a national file of identified death record information compiled from computer files submitted by State vital statistics offices. CHANCE collected multiple matching data: social security number (SSN), date of birth (DOB), sex, race, state of residence, and name. Therefore there was a high proportion (76% for 2008 deaths) of perfect/very close to perfect matches on SSN, DOB, and sex. A small number of partial matches (e.g. a few SSN digits,

parts of DOB) were examined and a determination was made whether to accept or not. If the initial three digits of the first-listed cause of death code were C01-C06, C09, C10, C12-C14, or C32, the cause of death was classified as head and neck cancer.

4.3.3 SNP Selection and Genotyping

Blood samples were obtained at the time of questionnaire administration by nurse-interviewers trained in phlebotomy. If the subject was not willing or able to consent to the blood draw, they were asked to contribute a buccal cell sample via mouthrinse.

For this project, 75 SNPs (69 tag SNPs and 6 candidate SNPs found in prior studies to be associated with cancer incidence or survival, or alcohol dependence) were selected in 12 genes that are part of two metabolic pathways: *ADH1B*, *ADH1C*, *ADH4*, *ADH7*, *ALDH2*, and *CYP2E1* in the alcohol metabolism pathway in the upper aerodigestive tract; and *CAT*, *SOD1*, *SOD2*, *GPx1*, *GPx2*, and *GPx4* in the oxidative stress metabolic pathway. Tag SNPs, chosen to represent the genetic variation within each of the 12 candidate genes (gene and 2000 bp upstream and downstream) were selected using the Genome Variation Server (223), using SNPs that were polymorphic in either CEU or YRI HapMap Release 2 (unrelated only), and with the following parameters: allele frequency cutoff 10%, 0.8 R^2 threshold minimum for variations to belong to the same cluster, 85% minimum data coverage for tag SNPs, 70% minimal data coverage for a variation to be potentially clustered with others.

To control for potential population stratification, we selected 157 Ancestry Informative Markers (AIMs) to maximize (1) the difference in allele frequencies (delta) between European and African populations in the HapMap data (CEU versus YRI), and (2) the Fisher's information criterion (FIC). AIMs were prioritized based on having the highest delta and FIC values in the following order: 90% European/10% African, 10% European/90% African, and 50% European/50% African. This allowed AIMs to represent the entire expected ancestral distribution of the study population. Individual estimates of percentage

African ancestry were calculated from 145 successfully genotyped AIMs using maximum likelihood estimation (MLE) methods previously described (184, 189, 224). AIMs were chosen to differentiate only between African and European ancestry, so we derived individual ancestry estimates for the two groups that summed to 1.0.

Genotyping was done by the Mammalian Genotyping Core Facility, University of North Carolina at Chapel Hill, using the Illumina GoldenGate genotyping assay with Sentrix Array matrix and 96-well standard microtiter plates. Technicians were blinded to case/control status of samples, and samples from both cases and controls were tested on each plate, along with DNA controls and random sample duplicates. Assay intensity data and genotype cluster images for all SNPs were individually reviewed to identify those with inadequate signal intensity or indistinguishable genotype clusters. Blind duplicates of 109 samples were genotyped to verify reliability of genotype calls. Tests of Hardy-Weinberg equilibrium (HWE) were conducted in controls stratified by race, and those SNPs for which the p-value was <0.001 were judged to be out of HWE.

4.3.4 Covariate measurement

Trained nurse-interviewers conducted an in-person interview with each subject. For this analysis only self-reported, non-proxy data were included. Questions were asked about demographics (age, sex, race-ethnicity, education, income, health insurance), cigarette smoking, drinking of alcoholic beverages, diet, oral health, medical history, family history of cancer, and textual information on co-morbid conditions.

We made an initial selection of covariates based on *a priori* knowledge regarding their potential association with survival. Covariates obtained from the interview included age, sex, race; duration of cigarette smoking; ever-use of non-cigarette tobacco products, ever exposure to environmental tobacco smoke (ETS) at work, ever exposure to ETS at home; whether the subject was covered by health insurance on diagnosis date, ever had a routine dental visit, ever had a loose permanent tooth due to disease, ever used mouthwash; family

history of SCCHN; household poverty as defined by federal guidelines for both income and number of persons supported, highest attained education level; and presence of selected co-morbid diseases: liver disease including hepatitis, cardiovascular and/or lung disease, renal disease and/or diabetes, and other cancers. Alcohol consumption, although measured, was not included in analyses because it may be an intermediate in the causal pathway between the SNPs and survival.

Clinical information about the tumor and its treatment were abstracted from the subjects' medical record. Pathology reports were reviewed by the study pathologist (WF) and a head and neck cancer surgeon (MW). Tumor and treatment characteristics obtained included TNM and summary stages, margin status, angiolymphatic invasion status, and cancer treatment (surgery, radiation, and/or chemotherapy).

4.3.5 Statistical analysis

For all-cause death, months of survival were calculated by subtracting the date of diagnosis from the date of death, and subjects who were still alive on 12/31/2008 were censored on that date. For head-and-neck-cancer-specific death (hereafter referred to as cancer death or cancer-specific death), months of survival were calculated by subtracting the date of diagnosis from the date of cancer death, while subjects who died of another cause were censored as of the date of death; subjects who were still alive on 12/31/2008 were censored on that date. Kaplan-Meier survival plots, and hazard ratios for the independent effects of SNPs from Cox proportional hazards regression models, were implemented in SAS® 9.2.

Kaplan-Meier plots and log-rank tests for all-cause and cancer-specific survival were constructed for the two alleles of each SNP. Kaplan-Meier plots and log-rank tests were also constructed for all-cause and cancer specific survival for the two races, for all head-and-neck cancers combined as well as separately for each anatomic sub-site.

A dominant genetic model (at least one minor allele versus referent of no minor alleles) was implemented because, for many of the SNPs, the number of subjects homozygous for the minor allele was too small to permit estimation of precise effect estimates.

Potential covariates were selected using a step-wise backwards elimination from a full model (which contained demographic, socioeconomic status (SES), tumor, clinical, co-morbid diseases, oral health, and tobacco exposure variables, as listed in Table 20) and comparison of each reduced model to the full model. There was no collinearity between any of the explanatory variables except possibly between race and ancestry. If eliminating a set of covariates did not change the ln(HR) for any SNP by a difference of at least 0.10, those covariates were eliminated from subsequent models. The final Cox models for all-cause and cancer-specific death included the demographic, SES, tumor, clinical, co-morbid diseases, oral health, and tobacco exposure variables listed in Table 21 footnotes. With self-reported race included in the model the ancestry variable was not important to all-cause or cancer-specific death for the genetic variants studied, so for parsimony, and because ancestry and race showed evidence of collinearity, ancestry was not included in final models.

For head and neck cancer, clinicians consider local-regional-distant stage classification to be inadequate for predicting survival; for example, advanced nodal (N) stage is a worse prognosticator than advanced loci disease (229, 230). Therefore, recent studies of survival in head and neck cancer patients most often control for cancer stage in Cox regression models by including separate variables representing tumor (T), node (N), and metastasis (M) stages. In CHANCE, comparison of two sets of hazard ratios and confidence intervals from Cox regression models that included summary stage, versus models that included separate variables for T, N, and M stages, revealed similar estimates from both methods. Therefore I decided to control for stage using three separate variables.

The angiolymphatic invasion variable was missing in 13.5% of subjects and that, along with a small percent missing in other covariates, resulted in 18% of records being excluded from complete-case analyses. In order to include available information from all cases, and because the data were not missing in a monotone pattern, we used the Markov Chain Monte Carlo method implemented in SAS® 9.2 PROC MI and information about non-missing values from all variables in full models to impute missing values for variables included in final models. This method imputes missing values for binary variables (such as angiolymphatic invasion) as continuous values from a normal distribution, which we did not round to 0 or 1 because the continuous values are less biased. Hazard ratios reported from those analyses are summaries of HRs from Cox regressions using 25 imputed datasets, calculated using SAS® 9.2 PROC MIANALYZE.

Bonferroni correction for 64 statistical tests was used to adjust p-values to control for Type 1 error introduced by multiple statistical testing.

4.4 RESULTS

Mean survival in subjects with SCCHN was 45.5 months (Table 19). Mean/median survival for those with hypopharynx tumors was lower (35/30 months) than for those with other tumor sites (43-47/45-48 months). For subjects who died during the follow-up period, mean/median survival for all-cause death was 26.2/22.4 months, and for cancer-specific death, 22.3/18.7 months.

Genotyping. DNA samples from 1266 CHANCE SCCHN cases of white or African-American race were genotyped. Of these, 38 failed to genotype, and 1 was gender discrepant, leaving 1227 cases with genotyping data. Of the 75 tag and candidate SNPs for which genotyping was attempted, 8 tag SNPs and 1 candidate SNP were excluded from analyses due to inadequate signal intensity or inability to distinguish genotype clusters, and 2 tag SNPs were excluded due to genotypes being out of HWE at the $p<0.001$ level, leaving

64 tag and candidate SNPs available for analyses. Genotype calls for blind duplicates were concordant for all SNPs. All 1227 cases had less than 4% missing SNP results.

Descriptive analyses. Table 20 presents the frequency distributions and hazard ratios for the initial set of covariates within the study population. Eighty percent of the subjects were 50 years of age or older upon diagnosis, three-quarters were male, and 25% were African-American.

For both all-cause and cancer-specific survival, more subjects with oropharynx cancer survived compared to those with oral cavity and hypopharynx tumors (Table 20). Increased survival was also associated with: lower Tumor and Node stages, absence of angiolymphatic invasion, having had surgical cancer treatment, absence of cardiovascular and lung disease, and shorter duration of smoking. In addition, having survived was also associated with being younger, richer, and more educated; absence of cardiovascular or lung disease; having had at least one routine dental visit in the last 10 years, and not having been exposed to environmental tobacco smoke (ETS) at home. Surprisingly, race was associated with all-cause death but not cancer-specific death in univariate analysis.

Examination of univariate hazard ratios for all-cause and cancer-specific death calculated for each potential covariate in separate Cox regression models (without SNPs) (Table 20), indicates that, with only a few exceptions, the same variables that are associated with vital status (yes/no) are also associated with length of all-cause and cancer-specific survival.

SNP hazard ratios and survival plots. Backwards elimination of potential covariates, from 64 full models each containing one SNP (of the 64 studied) and all covariates listed in Table 20, produced a Cox regression model for all-cause death containing the following covariates: SNP being analyzed (one per model), anatomic subsite, T stage, N stage, M stage, angiolymphatic invasion status, treatment (surgery, chemotherapy), co-morbid cardiovascular or lung disease, duration of cigarette smoking, and routine dental care. The

final Cox regression model for cancer death contained all those variables, plus radiation treatment, co-morbid liver disease, sex, race, age category, federal poverty group, exposure to ETS at work, and tooth mobility due to disease.

In unadjusted analyses, most SNPs were not associated with significantly increased hazard of all-cause or cancer death (Table 21). However, two SNPs in *CYP2E1* – rs3813865 and rs8192772 – were statistically associated with increased hazard of cancer death after adjustment for confounders in complete-case analysis and Bonferroni correction of p-values, with a similar pattern noted after adjustment for confounders in multiple imputation analysis (Table 21). Carrying 'C' minor allele of rs3813865 was associated with more than double the hazard of cancer death compared to the 'G' allele. Similarly, carrying 'C' allele of rs8192772 was associated with about 80% increased cancer death hazard compared to the 'T' allele. A non-significant increase in hazard of all-cause death was also noted for these SNPs. Separate analysis of deaths due to head and neck cancer versus deaths due to other causes indicates that there was no association of these SNPs with non-cancer deaths, and the non-significant increase in hazard of all-cause death was caused by cancer deaths (data not shown). Figure 9 presents the Kaplan-Meier survival curves for these SNPs.

There were no statistically significant differences in hazard ratios by race, because imprecise effect estimates precluded detailed study of this question. For example, carrying the *CYP2E1* rs3813865 minor 'C' allele, present in 26% of African-Americans and 5% of whites, conferred an increased HR (95% CI) in all subjects of 2.1 (1.4-3.2), which was 2.8 (1.5-5.2) in African-Americans, and 1.4 (0.7-2.9) in whites. Carrying the minor 'C' allele of rs8192772, present in 22% of African-Americans and 14% of whites, conferred similar increased risks in blacks and whites, with a HR (95% CI) of 1.7 (1.2-2.4) in all subjects, 2.1 (1.1-4.3) in African-Americans, and 1.5 (1.0-2.3) in whites. Although no other SNP HRs achieved statistical significance in the adjusted Cox models, four additional SNPs in

CYP2E1 (rs7092584, rs743535, rs2249695, rs28969387), and four SNPs in the oxidative stress pathway (rs11623705 and rs2412065 in *GPx2*, rs2070424 in *SOD1*, and rs2842980 in *SOD2*) displayed suggestive differences in allele hazards for either or both all-cause and cancer death from Kaplan-Meier and log-rank analysis (Table 22).

To investigate the effect of including alcohol consumption and percent African ancestry variables to the final reduced models, hazard ratios were re-calculated with those variables included. Effect estimates changed very little, if at all (data not shown).

4.5 DISCUSSION

In this study we evaluated the association of SNPs in genes in alcohol metabolism and oxidative stress pathways with survival after SCCHN diagnosis. We found that the minor alleles of several SNPs in *CYP2E1* were significantly associated with increased hazard for death from head and neck cancer.

CYP2E1 is a member of the cytochrome P450 oxidative system that is involved in metabolism of xenobiotics, ethanol, and drugs (231). Although CYP2E1 enzymatically inactivates some substrates, it has also been shown to bioactivate many compounds that are possibly carcinogenic; for example, the tobacco carcinogen N-nitrosamines, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (232). Indeed, the induction of alcohol metabolism, especially the CYP2E1 pathway, has been suggested as one of the key modes of alcohol-induced cancer (233). Laboratory studies have demonstrated that the most common chemotherapy drugs used to treat SCCHN – i.e., platinum-containing chemotherapeutic agents – increase *in vivo* CYP2E1 activity in mice, and *in vitro* CYP2E1 activity in renal and liver cells and liver microsomes (234-237). Moreover, numerous studies report that increased CYP2E1 activity is cytotoxic to liver cells *in vitro* and in mouse models, and decreased activity is protective against liver damage (238). Thus it is plausible that

CYP2E1 polymorphisms may differentially affect CYP2E1 activity when exposed to platinum-containing chemotherapy drugs.

The genes for Class I CYPs are highly conserved. Some polymorphisms have been found to be associated with cancer risk and survival. For example RsaI/rs2031920 and a 96-bp insertion polymorphism have been associated with colorectal and esophageal cancer (232). It is thought that these effects are produced by CYP2E1 inducibility rather than enzyme function. One study suggested that *CYP2E1* polymorphisms play a role in the metabolism of drugs used to treat non-Hodgkin lymphoma, with one allele of rs2070673 associated with longer survival (239), and another reported a significant association between *CYP2E1* rs2031920 wild-type C allele and improved survival from non-small cell lung cancer (240).

Our findings provide evidence that *CYP2E1* gene polymorphisms are associated with survival from head and neck cancer. Similar to other studies reporting that apparently non-functional *CYP2E1* SNPs influence cancer risk, the two SNPs we identified are also in typically non-coding regions of the genome – rs3813865 in the 5' region near the gene, and rs8192772 in intron 2. Interestingly, some studies have suggested that upstream 5' mutations in *CYP2E1* may affect gene expression and inducibility by ethanol (241). If these SNPs are not causal, they may instead be in linkage disequilibrium with a causal polymorphism.

Stratified analysis identified little difference in HR for the minor 'C' allele of rs3813865 in those who received chemotherapy (HR=2.0, 95% CI=1.0-3.8) and those who didn't (HR=2.2, 95% CI=1.3-3.9), but there was a large difference in HR for rs8192772 for those who received chemotherapy (HR=1.2, 95% CI=0.7-2.0) versus those who didn't (HR=2.1, 95% CI=1.4-3.2), indicating that some of the survival effects of *CYP2E1* polymorphisms may be mediated through either the differential effect of SNP alleles on the

metabolism of the chemotherapeutic agent and/or the differential effect of the chemotherapeutic agent on CYP2E1 activity.

Although one previous study reported an association of *CYP2E1* rs2031920 with advanced T stage SCCHN (124), ours is the first study to report the association of SNPs in *CYP2E1* with survival after SCCHN diagnosis. Our results are compatible with the hypothesis that genetic predisposition affects cancer-specific survival after diagnosis, and that at least some of this effect may be mediated by interaction of chemotherapeutic agents with *CYP2E1* polymorphisms. We did not, however, find any other SNPs in the alcohol metabolism pathway, or in the oxidative stress pathway, to be associated with survival, either cancer-specific or overall. We also did not have detailed information on the specific chemotherapeutic agents administered to our subjects; some drugs may interact with CYP2E1 to affect survival and others may not, which would confound our results.

In conclusion, our study found multiple SNPs in *CYP2E1* to be associated with cancer-specific survival. Some had different effects in those who received chemotherapy versus those who didn't, while others had the same effect in both groups. Future genetic survival analyses of head and neck cancer should confirm the association with these SNPs and investigate potentially causal loci. Such research would improve understanding of differential treatment effects on head and neck tumor progression and metastasis.

Table 19. Survival[1] within study population

Variable	Categories	Cases (n=1227) n	(col %)
Vital status			
	Alive	744	60.6%
	Dead	483	39.4%
Cause of death head and neck cancer?			
	No (other cause)	268	21.8%
	Yes	215	17.5%
	N/A (alive)	744	60.6%

	Survival (months)			
	n	Mean	Median	Std dev
All subjects	1227	45.5	46.0	21.8
Oral cavity	172	43.0	44.7	22.6
Oropharynx	333	47.0	46.7	21.5
Hypopharynx	55	35.0	30.0	22.3
Oral cavity - oropharynx - hypopharynx NOS	224	44.0	45.8	22.5
Larynx	443	47.3	47.7	20.7
Subjects who died of any cause	483	26.2	22.4	17.0
Subjects who died of head and neck cancer	215	22.3	18.7	14.1

[1] Vital status as of 12/31/2008, as determined from NDI

Table 20. Descriptives and univariate hazard ratios for non-SNP variables -- one explanatory variable per model (complete case analysis)

			DEATH FROM ALL CAUSES (n=483)			DEATH FROM HEAD & NECK CANCER (n=215)		
Potential covariate (type)	Covariate values	# (%) of total cases	# died/ # survived	HR[1] (95% CI)	unadjusted p-value (BOLD if <0.05)	# died/ # survived	HR[1] (95% CI)	unadjusted p-value (BOLD if <0.05)
DEMOGRAPHICS								
Age category (ordinal)	20-49 years	239 (19.5%)	80/159	1.08 (1.03-1.13)	**0.002**	40/199	1.06 (0.99-1.14)	0.097
	50-54 years	189 (15.4%)	75/114			30/159		
	55-59 years	207 (16.9%)	70/137			33/174		
	60-64 years	205 (16.7%)	79/126			36/169		
	65-69 years	168 (13.7%)	78/90			29/139		
	70-74 years	135 (11.0%)	58/77			26/109		
	75-80 years	84 (6.8%)	43/41			21/63		
Sex (categorical)	Female	289 (23.6%)	111/178	1.00 (ref)		57/232	1.00 (ref)	
	Male	938 (76.4%)	372/566	1.05 (0.85-1.30)	0.644	158/780	0.87 (0.64-1.18)	0.363
Race (categorical)	White/European-American	922 (75.1%)	339/583	1.00 (ref)		160/762	1.00 (ref)	
	Black/African-American	305 (24.9%)	144/161	1.41 (1.16-1.72)	**0.000**	55/250	1.13 (0.83-1.54)	0.427
Proportion African ancestry (continuous) (mean % in those who died)		1227		1.63 (1.25-2.12)	**0.000**		1.18 (0.78-1.79)	0.435
TUMOR/TREATMENT CHARACTERISTICS								
Anatomic subsite (categorical)	oral cavity (referent)	172 (14.0%)	83/89	1.00 (ref)		38/134	1.00 (ref)	
	hypopharynx	55 (4.5%)	35/20	1.59 (1.07-2.36)	**0.021**	15/40	1.44 (0.79-2.62)	0.231
	larynx	443 (36.1%)	169/274	0.71 (0.55-0.93)	**0.012**	70/373	0.65 (0.44-0.97)	**0.033**
	oral cavity-oropharynx-hypopharynx NOS	224 (18.3%)	83/141	0.75 (0.55-1.02)	0.065	41/183	0.81 (0.52-1.26)	0.359
	oropharynx	333 (27.1%)	113/220	0.64 (0.48-0.85)	**0.002**	51/282	0.64 (0.42-0.97)	**0.037**
T stage (ordinal)	T1, T1a, T1b	392 (31.9%)	105/287	1.43 (1.32-1.55)	**0.000**	31/361	1.64 (1.45-1.85)	**0.000**
	T2	409 (33.3%)	151/258			69/340		
	T3	218 (17.8%)	108/110			53/165		
	T4, T4a, T4b	208 (17.0%)	119/89			62/146		
N stage (ordinal)	N0	657 (53.5%)	223/434	1.31 (1.20-1.43)	**0.000**	84/573	1.50 (1.31-1.71)	**0.000**
	N1	156 (12.7%)	59/97			31/125		
	N2, N2a, N2b, N2c	366 (29.8%)	170/196			79/287		
	N3	48 (3.9%)	31/17			21/27		
M stage (categorical)	M0	1219 (99.3%)	476/743	1.00 (ref)		210/1009	1.00 (ref)	
	M1	8 (0.7%)	7/1	7.63 (3.60-16.17)	**0.000**	5/3	10.79 (4.41-26.40	**0.000**
Margin status (categorical)	Negative	396 (32.3%)	151/245	1.00 (ref)		64/332	1.00 (ref)	
	Positive	691 (56.3%)	284/407	1.13 (0.93-1.38)	0.225	128/563	1.19 (0.88-1.60)	0.263
	Missing	140 (11.4%)	48/92			23/117		
Angiolymphatic invasion (categorical)	Absent	954 (77.8%)	371/583	1.00 (ref)		162/792	1.00 (ref)	
	Present	107 (8.7%)	52/55	1.47 (1.10-1.96)	**0.010**	26/81	1.64 (1.08-2.48)	**0.020**
	Missing	166 (13.5%)	60/106			27/139		
Surgical treatment (categorical)	No	537 (43.8%)	243/294	1.00 (ref)		114/423	1.00 (ref)	
	Yes	690 (56.2%)	240/450	0.69 (0.58-0.83)	**0.000**	101/589	0.63 (0.48-0.83)	**0.001**
Radiation treatment (categorical)	No	282 (23.0%)	92/190	1.00 (ref)		36/246	1.00 (ref)	
	Yes	945 (77.0%)	391/554	1.36 (1.08-1.71)	**0.008**	179/766	1.57 (1.10-2.24)	**0.014**
Chemotherapy treatment (categorical)	No	754 (61.5%)	287/467	1.00 (ref)		125/629	1.00 (ref)	
	Yes	473 (38.5%)	196/277	1.17 (0.98-1.41)	0.088	90/383	1.22 (0.93-1.60)	0.157
COMORBID DISEASES								
Comorbid liver disease and/or hepatitis (categorical)	No	1113 (90.7%)	420/693	1.00 (ref)		187/926	1.00 (ref)	
	Yes	105 (8.6%)	57/48	1.63 (1.23-2.15)	**0.001**	24/81	1.49 (0.97-2.27)	0.068
	Missing	9 (0.7%)	6/3			4/5		
Comorbid cardiovascular and/or lung disease (categorical)	No	887 (72.3%)	302/585	1.00 (ref)		145/742	1.00 (ref)	
	Yes	336 (27.4%)	158/178	1.45 (1.20-1.74)	**0.000**	68/268	1.29 (0.97-1.72)	0.085
	Missing	4 (0.3%)	2/2			2/2		
Comorbid renal disease and/or diabetes (categorical)	No	1053 (85.8%)	403/650	1.00 (ref)		186/867	1.00 (ref)	
	Yes	167 (13.6%)	74/93	1.23 (0.96-1.58)	0.098	26/141	0.94 (0.62-1.42)	0.772
	Missing	7 (0.6%)	6/1			3/4		
Comorbid disease: other cancer (categorical)	No	1019 (83.0%)	395/624	1.00 (ref)		177/842	1.00 (ref)	
	Yes	204 (16.6%)	85/119	1.08 (0.85-1.37)	0.521	35/169	0.98 (0.68-1.41)	0.918
	Missing	4 (0.3%)	3/1			3/1		
# comorbid conditions (continuous) (mean # in those who died)		1227		1.21 (1.11-1.32)	**0.000**		1.13 (0.99-1.28)	0.074
OTHER POTENTIAL CONFOUNDERS								
Smoking duration, rounded to nearest whole year (continuous) (mean # years in those who died)		1227		1.02 (1.02-1.03)	**0.000**		1.02 (1.01-1.03)	**0.000**
Alcoholic beverages, lifetime consumption of ethanol in ml, never-users and tertiles of drinking (ordinal)				1.36 (1.23-1.52)	**0.000**		1.24 (1.07-1.44)	**0.005**
Federal poverty guideline (categorical)	At or above	816 (66.5%)	271/545	1.00 (ref)		126/690	1.00 (ref)	
	Below	356 (29.0%)	188/168	1.87 (1.55-2.25)	**0.000**	76/280	1.59 (1.19-2.11)	**0.002**
	Missing	55 (4.5%)	24/31			13/42		
Routine dental visit in last 10 years (categorical)	Yes	781 (63.7%)	240/541	1.00 (ref)		118/663	1.00 (ref)	
	No	438 (35.7%)	238/200	2.06 (1.72-2.46)	**0.000**	94/344	1.63 (1.24-2.14)	**0.000**
	Missing	8 (0.7%)	5/3			3/5		
Environmental tobacco smoke exposure at home (categorical)	No	399 (32.5%)	128/271	1.00 (ref)		59/340	1.00 (ref)	
	Yes	827 (67.4%)	354/473	1.38 (1.13-1.69)	**0.002**	155/672	1.31 (0.97-1.76)	0.081
	Missing	1 (0.1%)	1/0			1/0		
Environmental tobacco smoke exposure at work (categorical)	No	316 (25.8%)	134/182	1.00 (ref)		58/258	1.00 (ref)	
	Yes	909 (74.1%)	348/561	0.87 (0.71-1.06)	0.154	155/754	0.90 (0.66-1.21)	0.478
	Missing	2 (0.2%)	1/1			1/1		

Table 20. Descriptives and univariate hazard ratios for non-SNP variables -- one explanatory variable per model (complete case analysis)

			DEATH FROM ALL CAUSES (n=483)			DEATH FROM HEAD & NECK CANCER (n=215)		
Potential covariate (type)	Covariate values	# (%) of total cases	# died/ # survived	HR[1] (95% CI)	unadjusted p-value (BOLD if <0.05)	# died/ # survived	HR[1] (95% CI)	unadjusted p-value (BOLD if <0.05)
Ever used non-cigarette tobacco product (categorical)	No	754 (61.5%)	292/462	1.00 (ref)		129/625	1.00 (ref)	
	Yes	473 (38.5%)	191/282	1.06 (0.88-1.27)	0.552	86/387	1.08 (0.82-1.42)	0.590
Had health insurance on reference date (categorical)	Yes	1068 (87.0%)	403/665	1.00 (ref)		177/891	1.00 (ref)	
	No	154 (12.6%)	76/78	1.50 (1.17-1.92)	**0.001**	36/118	1.58 (1.11-2.27)	**0.012**
	Missing	5 (0.4%)	1/1			2/3		
Highest education level attained (categorical)	High school or more	828 (67.5%)	286/542	1.00 (ref)		133/695	1.00 (ref)	
	Less than high school	399 (32.5%)	197/202	1.57 (1.31-1.89)	**0.000**	82/317	1.39 (1.06-1.83)	**0.019**
Ever had loose permanent tooth due to disease (categorical)	No	765 (62.3%)	276/489	1.00 (ref)		138/627	1.00 (ref)	
	Yes	455 (37.1%)	202/253	1.29 (1.08-1.55)	**0.005**	74/381	0.93 (0.70-1.24)	0.638
	Missing	7 (0.6%)	5/2			3/4		
Ever regularly used mouthwash (categorical)	No	502 (40.9%)	191/311	1.00 (ref)		82/420	1.00 (ref)	
	Yes	719 (58.6%)	287/432	1.06 (0.89-1.28)	0.501	131/588	1.14 (0.86-1.50)	0.369
	Missing	6 (0.5%)	5/1			2/4		
Family history of H&N cancer in 1st degree relatives (categorical)	No	1206 (98.3%)	474/732	1.00 (ref)		210/996	1.00 (ref)	
	Yes	21 (1.7%)	9/12	1.13 (0.58-2.18)	0.722	5/16	1.39 (0.57-3.37)	0.466

[1] For categorical variables, HRs are in comparison to referent category. For continuous and ordinal variables, HRs are per unit of measurement and are cumulative (e.g. HR per year of age, HR per each additional comorbid disease, HR per each higher stage category)

Table 21. SNP hazard ratios, dominant genetic model

(adjusted p-values < 0.05 are highlighted)

				DEATH FROM ALL CAUSES						DEATH FROM HEAD AND NECK CANCER					
				Model: SNP unadjusted Complete case analysis n=1227 for most[3] SNPs		Model: SNP & covariates[1] Complete case analysis n=1050 for most[3] SNPs		Model: SNP & covariates[1] summary from 25 imputations n=1227 for all SNPs		Model: SNP unadjusted Complete case analysis n=1227 for most[3] SNPs		Model: SNP & covariates[2] Complete case analysis n=1050 for most[3] SNPs		Model: SNP & covariates[2] summary from 25 imputations n=1227 for all SNPs	
Gene	SNP name	major/ minor allele	allele frequency (any minor/only major/missing)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)
ALCOHOL METABOLISM GENES															
ADH1B	rs1250757[3]	C/A	855/372/0	0.99 (0.81-1.20)	1.000	1.18 (0.95-1.46)	1.000	1.10 (0.90-1.35)	1.000	1.03 (0.77-1.39)	1.000	1.05 (0.75-1.47)	1.000	1.11 (0.82-1.50)	1.000
	rs1042026	A/G	512/715/0	0.93 (0.78-1.12)	1.000	1.02 (0.84-1.25)	1.000	1.02 (0.85-1.22)	1.000	1.04 (0.79-1.36)	1.000	1.06 (0.77-1.46)	1.000	1.02 (0.76-1.35)	1.000
	rs7673353	C/T	99/1128/0	1.34 (0.99-1.81)	1.000	1.00 (0.72-1.38)	1.000	1.04 (0.76-1.43)	1.000	1.14 (0.70-1.84)	1.000	1.18 (0.64-2.16)	1.000	1.04 (0.58-1.83)	1.000
	rs17028834	T/C	86/1141/0	1.29 (0.93-1.79)	1.000	0.97 (0.68-1.38)	1.000	1.02 (0.73-1.42)	1.000	1.33 (0.82-2.15)	1.000	1.33 (0.72-2.45)	1.000	1.42 (0.81-2.50)	1.000
	rs1693457	T/C	421/805/1	0.86 (0.71-1.04)	1.000	0.80 (0.65-0.99)	1.000	0.85 (0.70-1.03)	1.000	0.87 (0.65-1.16)	1.000	0.94 (0.68-1.31)	1.000	0.89 (0.66-1.19)	1.000
	rs1229984	G/A	31/1196/0	0.67 (0.35-1.29)	1.000	0.74 (0.30-1.79)	1.000	0.89 (0.46-1.72)	1.000	0.84 (0.34-2.03)	1.000	0.74 (0.18-3.05)	1.000	1.01 (0.41-2.48)	1.000
	rs1159918	G/T	825/402/0	1.03 (0.85-1.25)	1.000	0.86 (0.70-1.06)	1.000	0.94 (0.77-1.14)	1.000	0.95 (0.72-1.26)	1.000	0.91 (0.65-1.27)	1.000	0.95 (0.70-1.28)	1.000
	rs1229982	G/T	527/700/0	1.11 (0.93-1.33)	1.000	0.94 (0.77-1.15)	1.000	1.04 (0.87-1.25)	1.000	0.94 (0.72-1.23)	1.000	0.91 (0.65-1.26)	1.000	0.96 (0.72-1.29)	1.000
ADH1C	rs2298753	T/C	221/1006/0	0.82 (0.64-1.04)	1.000	0.84 (0.64-1.09)	1.000	0.85 (0.67-1.08)	1.000	0.90 (0.63-1.28)	1.000	0.98 (0.65-1.47)	1.000	0.91 (0.63-1.31)	1.000
	rs1614972	C/T	687/540/0	0.93 (0.78-1.11)	1.000	0.85 (0.70-1.03)	1.000	0.88 (0.73-1.05)	1.000	0.87 (0.66-1.13)	1.000	0.85 (0.62-1.16)	1.000	0.87 (0.66-1.14)	1.000
	rs1391088	C/A	178/1049/0	0.79 (0.60-1.04)	1.000	0.77 (0.57-1.05)	1.000	0.82 (0.62-1.08)	1.000	0.80 (0.53-1.20)	1.000	0.69 (0.42-1.14)	1.000	0.85 (0.56-1.31)	1.000
	rs1693482	C/T	695/530/2	0.93 (0.77-1.11)	1.000	1.08 (0.88-1.31)	1.000	1.06 (0.88-1.27)	1.000	0.96 (0.73-1.26)	1.000	1.08 (0.79-1.50)	1.000	1.07 (0.81-1.43)	1.000
	rs1631460	C/G	704/522/1	0.91 (0.76-1.09)	1.000	1.07 (0.88-1.30)	1.000	1.05 (0.88-1.27)	1.000	0.96 (0.73-1.26)	1.000	1.08 (0.79-1.50)	1.000	1.08 (0.81-1.43)	1.000
	rs1193686[9]	C/G	614/613/0	0.92 (0.77-1.10)	1.000	0.89 (0.73-1.08)	1.000	0.89 (0.75-1.07)	1.000	0.78 (0.60-1.03)	1.000	0.83 (0.61-1.14)	1.000	0.84 (0.63-1.11)	1.000
ADH4	rs2900122[7]	A/T	105/1122/0	1.25 (0.93-1.69)	1.000	1.03 (0.74-1.43)	1.000	1.04 (0.76-1.41)	1.000	1.45 (0.95-2.21)	1.000	1.77 (0.99-3.15)	1.000	1.83 (1.08-3.10)	1.000
	rs1126672	C/T	505/722/0	0.91 (0.76-1.09)	1.000	0.94 (0.77-1.15)	1.000	1.00 (0.83-1.21)	1.000	0.80 (0.61-1.06)	1.000	0.70 (0.50-0.98)	1.000	0.79 (0.59-1.05)	1.000
	rs4699710	T/C	585/642/0	0.98 (0.82-1.18)	1.000	1.03 (0.84-1.25)	1.000	1.08 (0.89-1.29)	1.000	0.86 (0.66-1.13)	1.000	0.76 (0.55-1.04)	1.000	0.87 (0.66-1.15)	1.000
	rs1001746[6]	T/C	669/557/1	1.03 (0.86-1.24)	1.000	1.02 (0.84-1.25)	1.000	1.07 (0.89-1.29)	1.000	0.99 (0.76-1.29)	1.000	0.90 (0.66-1.23)	1.000	1.01 (0.77-1.33)	1.000
	rs1800759	C/A	870/357/0	1.15 (0.94-1.41)	1.000	1.04 (0.83-1.29)	1.000	1.07 (0.87-1.31)	1.000	1.02 (0.76-1.37)	1.000	0.93 (0.66-1.32)	1.000	0.95 (0.70-1.30)	1.000
	rs1800761	G/A	474/753/0	1.04 (0.87-1.25)	1.000	1.04 (0.85-1.27)	1.000	1.03 (0.86-1.24)	1.000	0.99 (0.75-1.30)	1.000	1.00 (0.72-1.39)	1.000	1.05 (0.78-1.40)	1.000
	rs3762894	T/C	385/842/0	0.99 (0.81-1.20)	1.000	0.99 (0.80-1.21)	1.000	1.00 (0.82-1.21)	1.000	0.95 (0.71-1.27)	1.000	0.95 (0.68-1.33)	1.000	0.99 (0.74-1.34)	1.000
ADH7	rs284787	C/T	484/743/0	0.93 (0.78-1.12)	1.000	1.01 (0.83-1.24)	1.000	1.05 (0.87-1.26)	1.000	1.14 (0.87-1.49)	1.000	1.24 (0.91-1.70)	1.000	1.32 (1.00-1.74)	1.000
	rs894369	C/G	391/836/0	1.13 (0.93-1.36)	1.000	1.06 (0.86-1.30)	1.000	1.04 (0.86-1.26)	1.000	1.01 (0.76-1.34)	1.000	0.93 (0.67-1.30)	1.000	0.92 (0.68-1.23)	1.000
	rs1758840[3]	T/A	388/839/0	1.00 (0.82-1.21)	1.000	1.09 (0.88-1.33)	1.000	1.06 (0.87-1.29)	1.000	0.80 (0.59-1.08)	1.000	0.82 (0.58-1.16)	1.000	0.78 (0.57-1.06)	1.000
	rs1154454	T/C	469/758/0	1.14 (0.95-1.37)	1.000	0.98 (0.80-1.20)	1.000	0.99 (0.82-1.19)	1.000	1.03 (0.79-1.36)	1.000	0.93 (0.67-1.30)	1.000	0.94 (0.70-1.26)	1.000
	rs1154456	T/C	648/579/0	0.88 (0.73-1.05)	1.000	1.03 (0.84-1.25)	1.000	0.98 (0.82-1.17)	1.000	1.10 (0.84-1.43)	1.000	1.16 (0.84-1.59)	1.000	1.16 (0.87-1.54)	1.000
	rs1154460	G/A	895/329/3	0.87 (0.72-1.06)	1.000	0.89 (0.72-1.11)	1.000	0.84 (0.69-1.02)	1.000	1.01 (0.75-1.37)	1.000	0.96 (0.68-1.35)	1.000	1.01 (0.74-1.37)	1.000
	rs971074	G/A	281/946/0	1.04 (0.85-1.29)	1.000	0.93 (0.74-1.17)	1.000	0.94 (0.76-1.16)	1.000	0.97 (0.70-1.34)	1.000	0.93 (0.64-1.35)	1.000	0.92 (0.66-1.29)	1.000
	rs1573496	C/G	166/1060/1	0.88 (0.68-1.15)	1.000	0.85 (0.64-1.14)	1.000	0.85 (0.65-1.12)	1.000	0.91 (0.61-1.35)	1.000	0.77 (0.48-1.25)	1.000	0.83 (0.55-1.25)	1.000
ALDH2	rs4767939	A/G	529/697/1	1.10 (0.92-1.31)	1.000	0.92 (0.76-1.12)	1.000	0.98 (0.81-1.17)	1.000	1.08 (0.83-1.42)	1.000	0.96 (0.70-1.31)	1.000	1.02 (0.77-1.35)	1.000
	rs2238151	T/C	858/365/4	1.19 (0.97-1.46)	1.000	0.92 (0.74-1.15)	1.000	0.98 (0.80-1.20)	1.000	1.18 (0.87-1.59)	1.000	0.98 (0.67-1.41)	1.000	1.12 (0.81-1.55)	1.000
	rs7312055	G/A	144/1083/0	1.69 (1.32-2.15)	0.002	1.21 (0.92-1.59)	1.000	1.25 (0.97-1.61)	1.000	1.37 (0.92-2.02)	1.000	1.79 (0.99-3.23)	1.000	1.50 (0.88-2.55)	1.000
	rs2158029	G/A	57/1170/0	1.38 (0.94-2.02)	1.000	0.96 (0.64-1.45)	1.000	0.99 (0.67-1.46)	1.000	0.62 (0.28-1.41)	1.000	0.54 (0.21-1.36)	1.000	0.48 (0.21-1.12)	1.000
	rs1694166[7]	C/T	239/988/0	1.05 (0.84-1.32)	1.000	0.95 (0.74-1.21)	1.000	0.98 (0.78-1.23)	1.000	1.24 (0.90-1.71)	1.000	1.12 (0.78-1.62)	1.000	1.19 (0.86-1.64)	1.000
	rs1694166[9]	T/G	240/987/0	0.91 (0.72-1.15)	1.000	0.96 (0.75-1.22)	1.000	0.88 (0.70-1.11)	1.000	0.95 (0.68-1.33)	1.000	0.89 (0.60-1.32)	1.000	0.91 (0.64-1.29)	1.000
CYP2E1	rs3813865	G/C	125/1102/0	1.53 (1.17-1.98)	0.107	1.39 (1.04-1.85)	1.000	1.47 (1.12-1.92)	0.334	1.67 (1.14-2.44)	0.549	2.28 (1.44-3.61)	0.030	2.09 (1.38-3.18)	0.034
	rs3813867	G/C	83/1143/1	0.79 (0.53-1.16)	1.000	0.70 (0.47-1.05)	1.000	0.67 (0.45-0.99)	1.000	0.94 (0.55-1.62)	1.000	0.69 (0.37-1.30)	1.000	0.83 (0.48-1.45)	1.000
	rs8192772	T/C	200/1026/1	1.41 (1.13-1.77)	0.150	1.46 (1.15-1.85)	0.123	1.40 (1.12-1.76)	0.212	1.59 (1.15-2.19)	0.299	1.88 (1.32-2.69)	0.032	1.71 (1.23-2.37)	0.095
	rs915908	G/A	250/977/0	0.73 (0.57-0.93)	0.650	0.85 (0.65-1.11)	1.000	0.78 (0.61-1.00)	1.000	0.91 (0.65-1.27)	1.000	0.98 (0.66-1.45)	1.000	0.92 (0.65-1.30)	1.000
	rs915909	C/T	59/1167/1	1.13 (0.76-1.69)	1.000	0.81 (0.53-1.25)	1.000	0.76 (0.50-1.15)	1.000	1.21 (0.67-2.16)	1.000	1.18 (0.59-2.37)	1.000	1.03 (0.54-1.98)	1.000
	rs7092584	C/T	263/962/2	1.31 (1.07-1.61)	0.612	1.26 (1.01-1.57)	1.000	1.26 (1.02-1.55)	1.000	1.37 (1.01-1.86)	1.000	1.34 (0.96-1.88)	1.000	1.37 (1.01-1.86)	1.000
	rs743535	C/T	249/969/9	1.34 (1.09-1.65)	0.399	1.24 (0.99-1.56)	1.000	1.23 (1.00-1.52)	1.000	1.54 (1.14-2.09)	0.331	1.41 (1.00-2.00)	1.000	1.55 (1.13-2.11)	0.378
	rs2249695	C/T	603/623/1	1.32 (1.10-1.58)	0.166	1.09 (0.89-1.33)	1.000	1.07 (0.89-1.28)	1.000	1.37 (1.05-1.80)	1.000	1.44 (1.02-2.03)	1.000	1.33 (0.98-1.81)	1.000
	rs28969387	A/T	25/1202/0	1.75 (1.05-2.93)	1.000	0.96 (0.53-1.73)	1.000	1.12 (0.66-1.89)	1.000	2.11 (1.04-4.27)	1.000	0.81 (0.28-2.32)	1.000	1.43 (0.67-3.06)	1.000
	rs11101812	T/C	26/1198/3	0.72 (0.36-1.46)	1.000	0.52 (0.25-1.11)	1.000	0.45 (0.22-0.93)	1.000	0.83 (0.31-2.24)	1.000	0.65 (0.20-2.15)	1.000	0.67 (0.24-1.91)	1.000

Table 21. SNP hazard ratios, dominant genetic model

(adjusted p-values < 0.05 are highlighted)

				DEATH FROM ALL CAUSES						DEATH FROM HEAD AND NECK CANCER					
				Model: SNP unadjusted Complete case analysis n=1227 for most[3] SNPs		Model: SNP & covariates[1] Complete case analysis n=1050 for most[3] SNPs		Model: SNP & covariates[1] summary from 25 imputations n=1227 for all SNPs		Model: SNP unadjusted Complete case analysis n=1227 for most[3] SNPs		Model: SNP & covariates[2] Complete case analysis n=1050 for most[3] SNPs		Model: SNP & covariates[2] summary from 25 imputations n=1227 for all SNPs	
Gene	SNP name	major/ minor allele	allele frequency (any minor/only major/missing)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)	HR[4] (95% CI)	Bonferroni-corrected p-value (64 tests) (highlighted if <0.05)
OXIDATIVE STRESS GENES															
CAT	rs1049982	C/T	703/521/3	1.08 (0.90-1.30)	1.000	1.08 (0.89-1.32)	1.000	1.06 (0.89-1.28)	1.000	0.97 (0.74-1.27)	1.000	1.06 (0.77-1.44)	1.000	1.01 (0.77-1.33)	1.000
GPX1	rs8179172	T/A	51/1176/0	1.64 (1.11-2.42)	0.781	1.16 (0.75-1.79)	1.000	1.20 (0.81-1.79)	1.000	0.91 (0.43-1.93)	1.000	0.52 (0.20-1.35)	1.000	0.71 (0.31-1.58)	1.000
	rs1800668	C/T	588/638/1	0.96 (0.81-1.15)	1.000	1.00 (0.82-1.21)	1.000	0.99 (0.83-1.19)	1.000	0.86 (0.65-1.12)	1.000	0.78 (0.57-1.06)	1.000	0.83 (0.63-1.09)	1.000
	rs3811699	A/G	614/613/0	1.00 (0.84-1.20)	1.000	0.96 (0.79-1.17)	1.000	0.98 (0.81-1.17)	1.000	0.89 (0.68-1.16)	1.000	0.77 (0.56-1.05)	1.000	0.84 (0.63-1.10)	1.000
	rs3448	C/T	578/649/0	0.85 (0.71-1.02)	1.000	0.85 (0.70-1.03)	1.000	0.82 (0.68-0.98)	1.000	0.84 (0.64-1.10)	1.000	0.93 (0.68-1.26)	1.000	0.79 (0.60-1.04)	1.000
GPX2	rs1162370[5]	G/T	227/1000/0	0.95 (0.75-1.20)	1.000	0.98 (0.75-1.27)	1.000	0.98 (0.77-1.24)	1.000	1.41 (1.03-1.93)	1.000	1.47 (1.01-2.14)	1.000	1.41 (1.02-1.97)	1.000
	rs2412065	G/C	529/698/0	1.20 (1.00-1.44)	1.000	1.04 (0.86-1.27)	1.000	1.04 (0.87-1.25)	1.000	1.36 (1.04-1.77)	1.000	1.28 (0.93-1.75)	1.000	1.27 (0.96-1.69)	1.000
	rs2737844	C/T	728/494/5	1.28 (1.06-1.54)	0.646	1.09 (0.89-1.34)	1.000	1.03 (0.85-1.25)	1.000	1.28 (0.97-1.69)	1.000	1.30 (0.92-1.84)	1.000	1.21 (0.89-1.64)	1.000
GPX4	rs757229	G/C	925/301/1	1.05 (0.85-1.29)	1.000	1.04 (0.83-1.30)	1.000	1.02 (0.83-1.27)	1.000	0.97 (0.71-1.31)	1.000	1.12 (0.78-1.61)	1.000	0.96 (0.71-1.32)	1.000
SOD1	rs1191011[5]	A/C	62/1165/0	1.26 (0.86-1.83)	1.000	1.16 (0.79-1.72)	1.000	1.09 (0.74-1.59)	1.000	1.05 (0.57-1.93)	1.000	1.31 (0.65-2.63)	1.000	1.18 (0.60-2.31)	1.000
	rs4998557	G/A	395/831/1	1.20 (0.99-1.45)	1.000	1.06 (0.87-1.30)	1.000	1.05 (0.87-1.27)	1.000	1.17 (0.89-1.56)	1.000	1.11 (0.79-1.55)	1.000	1.09 (0.81-1.47)	1.000
	rs1043278[2]	T/G	350/876/1	1.22 (1.00-1.47)	1.000	1.13 (0.92-1.39)	1.000	1.12 (0.92-1.36)	1.000	1.31 (0.99-1.74)	1.000	1.26 (0.91-1.75)	1.000	1.25 (0.94-1.68)	1.000
	rs2070424	A/G	222/1005/0	1.26 (1.01-1.57)	1.000	1.09 (0.86-1.38)	1.000	1.12 (0.89-1.40)	1.000	1.40 (1.01-1.92)	1.000	1.13 (0.78-1.65)	1.000	1.31 (0.93-1.82)	1.000
	rs1041740	C/T	543/684/0	0.91 (0.76-1.09)	1.000	0.97 (0.79-1.18)	1.000	0.97 (0.81-1.16)	1.000	0.95 (0.72-1.24)	1.000	0.88 (0.64-1.21)	1.000	0.89 (0.67-1.18)	1.000
SOD2	rs4342445	G/A	473/754/0	0.91 (0.76-1.10)	1.000	1.04 (0.85-1.27)	1.000	1.01 (0.84-1.22)	1.000	0.78 (0.59-1.04)	1.000	0.89 (0.64-1.24)	1.000	0.86 (0.65-1.16)	1.000
	rs2842980	A/T	510/717/0	1.21 (1.01-1.44)	1.000	1.07 (0.88-1.30)	1.000	1.16 (0.97-1.39)	1.000	1.46 (1.12-1.91)	0.349	1.37 (1.00-1.88)	1.000	1.47 (1.11-1.94)	0.446
	rs8031	T/A	838/389/0	1.01 (0.83-1.22)	1.000	1.06 (0.86-1.31)	1.000	1.03 (0.85-1.25)	1.000	1.18 (0.88-1.58)	1.000	1.12 (0.80-1.57)	1.000	1.12 (0.83-1.53)	1.000
	rs5746134	C/T	100/1127/0	1.31 (0.96-1.77)	1.000	0.93 (0.66-1.31)	1.000	0.96 (0.70-1.31)	1.000	1.33 (0.85-2.09)	1.000	1.09 (0.58-2.05)	1.000	1.26 (0.73-2.20)	1.000
	rs2758331	C/A	803/424/0	0.98 (0.81-1.18)	1.000	1.12 (0.91-1.37)	1.000	1.07 (0.89-1.30)	1.000	1.06 (0.80-1.41)	1.000	1.10 (0.78-1.54)	1.000	1.07 (0.79-1.46)	1.000

[1] Cox models used to calculate hazard ratios for all-cause death included snp coded for dominant model, anatomical site, T stage, N stage, M stage, angiolymphatic invasion status, treatment (surgery, chemotherapy), presence of comorbid cardiovascular or lung disease, duration of smoking cigarettes (years), reception of at least one routine dental visit in past 10 years

[2] Cox models used to calculate hazard ratios for cancer-specific death included snp coded for dominant model and all covariates listed above for all-cause death models, plus the following additional variables: radiation treatment, presence of comorbid liver disease or hepatitis, sex, race, age, federal poverty level, ever exposure to environmental tobacco smoke at work, ever loss of a tooth due to disease

[3] These analyses included only those cases who had complete information on the SNP and any covariates in the model. Two-thirds of SNPs (43 of 64) were successfully genotyped for all cases, and 20% of SNPs (n=13) were missing in only one subject.

[4] HRs represent hazard of any minor allele, compared to hazard of homozygous major allele)

Table 22. Hazard differences between alleles of selected SNPs

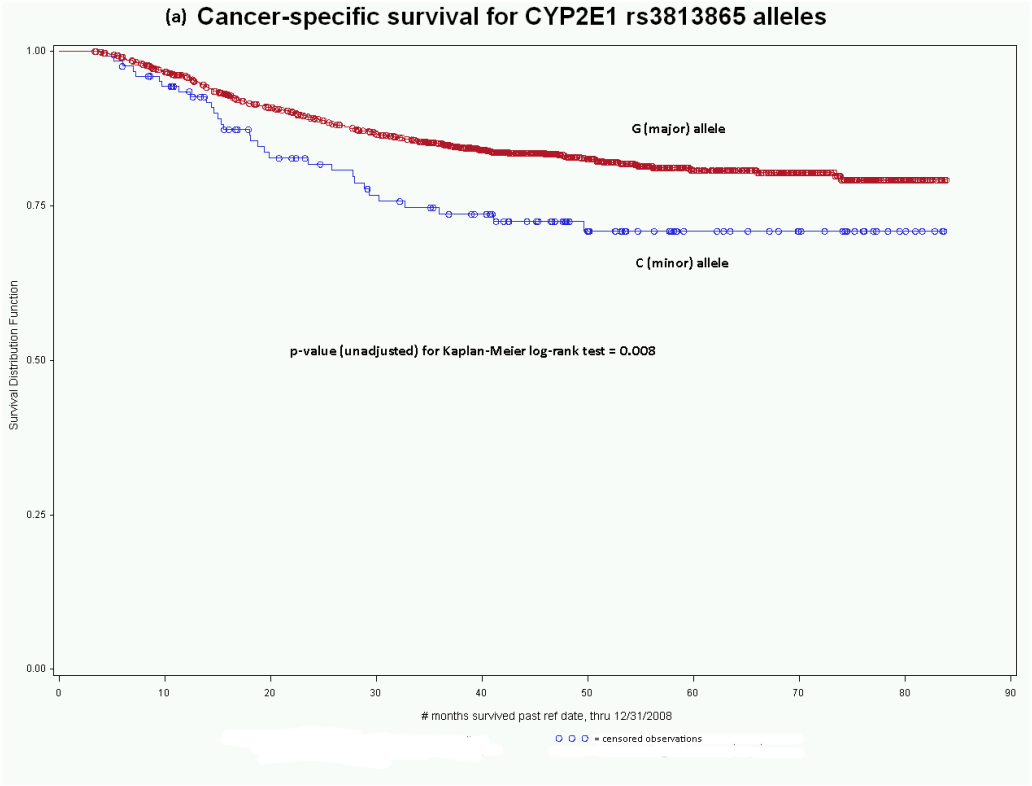
			ALL-CAUSE DEATH			CANCER DEATH		
				p-values from Kaplan-Meier log-rank tests[3]			p-values from Kaplan-Meier log-rank tests[3]	
gene	snp	major/ minor allele	Adjusted HR[1] (95% CI)	Unadjusted	Bonferroni-corrected (64 tests)	Adjusted HR[2] (95% CI)	Unadjusted	Bonferroni-corrected (64 tests)
CYP2E1	rs3813865	G/C	1.47 (1.12-1.92)	0.002	0.099	2.09 (1.38-3.18)	0.008	0.505
	rs8192772	T/C	1.40 (1.12-1.76)	0.002	0.146	1.71 (1.23-2.37)	0.004	0.276
	rs7092584	C/T	1.26 (1.02-1.55)	0.009	0.600	1.37 (1.01-1.86)	0.042	1.000
	rs743535	C/T	1.23 (1.00-1.52)	0.006	0.390	1.55 (1.13-2.11)	0.005	0.309
	rs2249695	C/T	1.07 (0.89-1.28)	0.002	0.159	1.33 (0.98-1.81)	0.021	1.000
	rs28969387	A/T	1.12 (0.66-1.89)	0.031	1.000	1.43 (0.67-3.06)	0.035	1.000
GPX2	rs11623705	G/T	0.98 (0.77-1.24)	0.681	1.000	1.41 (1.02-1.97)	0.030	1.000
	rs2412065	G/C	1.04 (0.87-1.25)	0.044	1.000	1.27 (0.96-1.69)	0.025	1.000
SOD1	rs2070424	A/G	1.12 (0.89-1.40)	0.036	1.000	1.31 (0.93-1.82)	0.040	1.000
SOD2	rs2842980	A/T	1.16 (0.97-1.39)	0.040	1.000	1.47 (1.11-1.94)	0.005	0.331

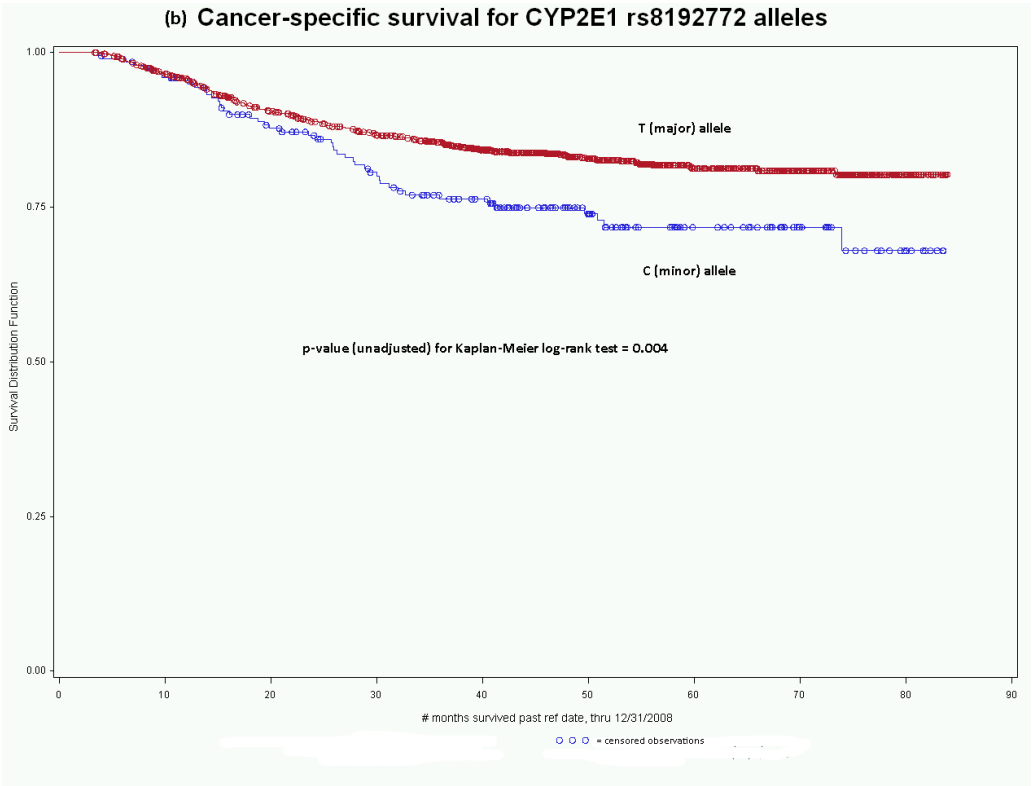
[1] Cox models used to calculate hazard ratios for all-cause death included snp coded for dominant model (any minor allele versus homozygous major allele), anatomical site, T stage, N stage, M stage, angiolymphatic invasion status, treatment (surgery, chemotherapy), presence of comorbid cardiovascular or lung disease, duration of smoking cigarettes (years), reception of at least one routine dental visit in past 10 years. 25 datasets were imputed, and results were summarized as a single HR.

[2] Cox models used to calculate hazard ratios for cancer-specific death included snp coded for dominant model (any minor allele versus homozygous major allele) and all covariates listed above for all-cause death models, plus the following additional variables: radiation treatment, presence of comorbid liver disease or hepatitis, sex, race, age, federal poverty level, ever exposure to environmental tobacco smoke at work, ever loss of a tooth due to disease. 25 datasets were imputed, and results were summarized as a single HR.

[3] Kaplan-Meier log-rank test for differences between the survival curves for the two alleles of each SNP

Figure 9. Kaplan-Meier survival plots for *CYP2E1* rs3813865 and rs8192772 alleles





CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 SPECIFIC AIMS

This dissertation research investigating the association of genetic variation in the alcohol and oxidative stress metabolism pathways with head and neck cancer accomplished the following aims: to measure (1) main effects of genetic variation on risk of developing cancer; (2) interaction of genetic variations with ethanol consumption from alcoholic beverages; and (3) main effects of genetic variation on survival after cancer diagnosis.

5.2 STUDY POPULATION AND DATA COLLECTION

To accomplish these aims, data from a single study population, the Carolina Head and Neck Cancer Epidemiology study (CHANCE), were analyzed. CHANCE was a large population-based case-control study of squamous cell carcinoma of the head and neck conducted in North Carolina (United States) from 2002-2006. It included both white and African-American cases and controls, with controls being frequency matched to cases on sex, race, and age category. Rapid ascertainment techniques were used to identify cases in an effort to capture cases while they were still capable of personally answering the questionnaire and providing a biological sample.

Nurse-administered questionnaires were used to gather self-report information. This study analyzed questionnaire data regarding demographics (sex, age, and race); duration, frequency, size, and type of alcoholic beverages consumed; duration and frequency of cigarette smoking; exposure to environmental tobacco smoke and non-cigarette tobacco products; socioeconomic variables of education, income, and health insurance; oral health; co-morbid disease; and family history of cancer.

Many previous case-control studies measured odds ratios for functional or presumed functional non-synonymous polymorphisms in candidate genes. Because synonymous SNPs and variations in introns and up- and down-stream regions can also affect enzyme function, I wanted to study greater genetic variation across each candidate gene (six genes in each pathway) rather than only non-synonymous SNPs. Therefore, for each gene, tag SNPs were selected that were in high linkage disequilibrium with other SNPs in the same gene, including an additional 2000 bp both up- and down-stream, with the goal of capturing 80% of the genetic variation within and near each gene. This would allow us to also examine the effect of haplotypes on cancer risk. In addition to tag SNPs, six candidate SNPs examined in previous studies of head and neck or other cancers were also genotyped. DNA extracted from blood and buccal samples was genotyped for each SNP; the laboratory successfully genotyped 64 of 75 SNPs (85%), including 59 tag and 5 candidate SNPs.

Because the study population included both whites and African-Americans, it was possible that population stratification could bias effect measure estimates. Self-reported race, which does not take admixture into account, might have been inadequate for purposes of controlling for population stratification in regression analyses. Therefore CHANCE investigators also genotyped ancestry informative markers (AIMs, 157 SNPs) that were selected to be highly divergent between African and European-American HapMap populations (YRI and CEU, respectively) for purposes of constructing an estimate of proportion of African ancestry. The laboratory successfully genotyped 145 of 157 AIMs (92%).

For Aim 3, CHANCE physician investigators abstracted clinical information from subjects' medical records about tumor location, stage, treatment, and other prognostic factors such as presence of angiolymphatic invasion and surgical margin status. Vital status and date and cause of death were obtained from linkage with the National Death Index as of 12/31/2008.

5.3 FINDINGS

5.3.1 Aim 1

The minor allele of two SNPs was associated with altered SCCHN risk: the 'A' allele of *ADH1B* rs1229984 was associated with decreased odds and the 'C' allele of *ALDH2* rs2238151 with increased odds. In addition, minor alleles of three SNPs were associated with increased cancer odds in anatomic sub-sites: the 'C' allele of *ADH1B* rs17028834 with laryngeal cancer, the 'A' allele of *SOD2* rs4342445 with oral cavity tumors, and the 'T' allele of *SOD2* rs5746134 with hypopharyngeal cancer. Previous studies, including a recent genome-wide association study spearheaded by IARC, have also found decreased risk associated with the A allele of *ADH1B* rs1229984, but have not reported the associations I found between *ADH1B* rs17028834 and laryngeal cancer, or between *ALDH2* rs2238151 and SCCHN. No other studies have examined SNPs in *SOD2* in regard to risk of SCCHN or any of its anatomic sub-sites.

Most SNPs studied (59 of 64, or 92%) were not associated with altered cancer risk. No SNPs in *ADH1C*, *ADH4*, *ADH7*, *CYP2E1*, *CAT*, *GPx1*, *GPx2*, *GPx4*, and *SOD1* were associated with risk of developing head and neck cancer. Though other studies have investigated SNPs in *ADH1C*, *ADH4*, *ADH7*, and *CYP2E1*, this is the first study to investigate the association of *CAT*, *GPx1*, *GPx2*, *GPX4*, and *SOD1* with SCCHN.

Because alcohol may be in the causal pathway between SNPs in these genes and SCCHN, I did not adjust for alcohol consumption when estimating odds ratios for SNPs. Sex, race, and age category were included in conditional logistic regression models because controls were frequency matched to cases on these variables. Although I included 13 potential covariates in initial models, including measurements of tobacco exposure, socioeconomic status, health history, and percent African ancestry, the only variable that materially affected odds ratios was years of cigarette smoking. The fact that adjusting for

African ancestry did not change effect estimates provides evidence that self-reported race sufficiently controlled for population stratification in the CHANCE study population.

5.3.2 Aim 2

Many studies have investigated interaction between alcohol metabolism genes and alcoholic beverage consumption. The strongest associations were found in Asian populations, in which the highly prevalent null (non-functional) and 10% functional genotypes of *ALDH2* are markedly associated with increased SCCHN risk in drinkers. Although the prevalence of the null *ALDH2* genotype is very low (<5%) or non-existent in non-Asian populations, including CHANCE, studies in other populations have reported that a common polymorphism of another gene, *ADH1B*, in the alcohol metabolism pathway is associated with risk. The "slow" form of ADH1B, which is about 40 times slower at metabolizing ethanol than the fast form – is associated with reduced risk of SCCHN overall, and synergistically interacts with lifetime alcohol consumption. Conversely, the majority of studies of the fast versus slow form of ADH1C reported no main effect on SCCHN incidence, but some evidence of increased risk in drinkers. It is important to note that the ADH1C fast form is only 2 ½ times faster at processing ethanol than the slow form, which may explain why fewer effects on cancer incidence are noted for *ADH1C* compared to *ADH1B*. Although my findings were consistent with the literature for the slow form of *ADH1B* being associated with reduced SCCHN risk, in contrast to many other studies I found no evidence of its interaction with alcohol consumption. Consistent with most studies of the *ADH1C* fast allele, I found no main effect or interaction with alcohol for that polymorphism.

This study found four tag SNPs in alcohol metabolism genes that showed evidence of additive interaction with the level of alcohol consumption, in the entire population and suggestively among whites and African-Americans separately: *ADH1B* rs1159918, *ADH7* rs1154460, *ALDH2* rs2238151, and *CYP2E1* rs2249695. These SNPs are different from those reported by previous studies of alcohol interaction and SCCHN incidence. The body of

literature supports the hypothesis that polymorphisms of *ADH* and *ALDH* genes produce enzymes that interact differentially with alcohol to affect SCCHN incidence, so it is not surprising that my tag SNPs, due to linkage disequilibrium with causal alleles, would also pick up signals of genetic interaction with alcohol. Evidence of *CYP2E1* interaction with alcohol had been less compelling. However, a 2011 linkage study reported that, consistent with my finding, rs2249695 is associated with "tipsiness" and is a predictor of alcoholism risk.

This is the only study to report having examined interactions of alcohol with SNPs in oxidative stress genes. I found no evidence of interaction for these genes and alcohol. Alcohol metabolism genes interact with alcohol by causing direct DNA damage via acetaldehyde and ROS, as well as discomfort or tipsiness mechanisms that influence the individual's level of drinking. Discomfort mechanisms are unlikely to occur due to the action of oxidative stress genes, which are not known to affect how individuals feel upon drinking, so the effects of these genes on cancer incidence are likely due to direct DNA damage.

5.3.3 Aim 3

With the exception of one study examining a single SNP in *SOD2*, this is the only study to have examined survival associations of polymorphisms in the alcohol metabolism and oxidative stress pathways.

Carrying the C allele for either of two SNPs in *CYP2E1* is associated with increased hazard of head-and-neck-cancer-specific death in both adjusted and unadjusted analyses. Four additional SNPs in *CYP2E1* and four in *GPx1*, *SOD1*, and *SOD2* displayed suggestive differences in allele hazards for all-cause and/or cancer death. No consistent associations with survival were found for SNPs in *ADH1B*, *ADH1C*, *ADH4*, *ADH7*, *ALDH2*, *GPx2*, *GPx4*, and *CAT*.

Adjusting for self-reported race was sufficient to control for confounding by race in Cox proportional hazards models, and including percent African ancestry in models did not change hazard ratio estimates.

Alcohol was excluded from main effects models because it may be in the causal pathway between SNPs and survival. In addition to the potential confounders tested for inclusion in the logistic regression models for odds of developing cancer, initial Cox models for survival included variables representing presence of selected co-morbid diseases (liver, cardiovascular/lung, renal/diabetes, other cancers), TNM stage, margin status, presence of angiolymphatic invasion, and treatment (surgery, radiation, and/or chemotherapy). As expected, both overall and cancer-specific survival were affected by smoking and many other factors, including TNM stage, angiolymphatic invasion, surgical cancer treatment, and co-morbid cardiovascular/lung disease, although margin status, believed to be clinically important, was not associated with survival in either univariate or adjusted analyses. Socioeconomic, health, and other variables that were not important to cancer incidence became important for survival after diagnosis, including income, education level, co-morbid cardiovascular/lung disease, exposure to environmental tobacco smoke at home, and whether or not the subject had at least one routine dental visit in the past 10 years. I find it very interesting that having had a routine dental visit is strongly associated with improved all-cause and cancer-specific survival. Is this a marker for having received preventive health care in general, or for oral health as an independent factor influencing survival? Having had routine dental care is partially correlated (Pearson's correlation coefficients ranging from 0.14 to 0.35) with race, lifetime alcohol drinking, duration of cigarette smoking, poverty level, highest attained education, having had tooth mobility due to disease, and whether the person had health insurance at the time of diagnosis. It is therefore difficult to rule out residual confounding from socioeconomic factors as the source of oral health care's importance in regression models.

Because CYP2E1 metabolizes the platinum-containing chemotherapeutic agents often used to treat SCCHN, and its metabolism of alcohol and xenobiotics is associated with liver damage and cancer in animal models, it is biologically plausible that polymorphisms in this gene might affect enzyme function, treatment efficacy, and thus survival. The two *CYP2E1* SNPs most associated with cancer-specific survival in this study are located in the 5' region near the gene and in intron 2, three suggestively associated SNPs were in introns, and only one suggestively associated SNP was a missense variant. Although it is possible that some of the SNPs I investigated were functional, it is more likely that they are simply markers of other variants that are the true causal polymorphisms.

Suggestive survival effects of SNPs in *GPx1*, *SOD1* and *SOD2* need to be investigated in other and larger studies, as my findings are not statistically significant.

5.3.4 Across specific aims

Polymorphisms in *ADH1B*, *ALDH2*, *CYP2E1*, and *SOD2* were found to have multiple associations with incidence or survival, across more than one specific aim. These associations are summarized below.

Carriers of the minor allele of two SNPs in *ADH1B* were found to have higher odds of developing SCCHN and laryngeal tumors (Aim 1), and a third SNP in this gene interacted with ethanol consumption to increase risk in heavy drinkers (Aim 2). However, no SNPs in *ADH1B* were associated with survival.

Carriers of the minor allele of rs2238151 in *ALDH2* had higher odds of developing SCCHN cancer, primarily due to increased risk of laryngeal tumors (Aim 1), and experienced increased risk if they were heavy drinkers (Aim 2). But this SNP had no effect on survival.

CYP2E1, with its dual roles in both alcohol and drug metabolism, had a haplotype that was associated with reduced odds of developing SCCHN (Aim 1) and that included a SNP which interacted with alcohol to increase risk in heavier drinkers carrying the minor

allele (Aim 2). The minor alleles of several *CYP2E1* SNPs were also associated with worse cancer-specific survival (Aim 3).

SOD2 contains two SNPs for which the minor allele was suggestively associated with higher odds of oral cavity and hypopharyngeal (Aim 1), and one SNP whose minor allele was associated with increased hazard of death (Aim 3).

5.4 STRENGTHS AND LIMITATIONS

5.4.1 For all study aims

CHANCE is one of the largest studies of head and neck cancer conducted in the United States. Its population is drawn from a state with a large African-American population, and as a result it has the largest number of African-Americans ever enrolled in a molecular epidemiologic study of head and neck cancer. Even though population ancestry was determined to be unnecessary in regression models once self-reported race was included, it is an advantage of this study to be able to confirm that residual confounding due to ancestry admixture is probably not confounding the results.

Because I studied mostly tag SNPs, I was able to study polymorphism associations with cancer with greater coverage across the gene. However, because they are selected to capture information about other SNPs with a minor allele frequency of at least 10%, they are likely to miss rare variants. This study was limited by the fact that 11 of the 75 tag SNPs were not successfully genotyped, thereby preventing effect estimation for the gene sections tagged by the missing SNPs. Genotyping failure most affected coverage in *ADH1C* (3 of 9 SNPs failed), *ADH1B* (3 of 11), and *CYP2E1* (3 of 13), and to a lesser extent *ADH4* (1 of 8) and *SOD2* (1 of 6), but did not affect coverage of the *ADH7*, *ALDH2*, *SOD1*, *GPx* family, and *CAT* genes. The laboratory was not able to genotype one of the six candidate (non-tag) SNPs, a missense SNP in *ADH1C* (rs698), because of its low Illumina design score, so I was not able to compare its effects in the CHANCE population to those reported in the literature.

There were a number of cases who died very soon after qualifying for enrollment in CHANCE, and who thereby neither personally responded to the questionnaire nor provided a biological sample. These cases were excluded from this study. If incidence and survival effects due to SNP polymorphisms are more pronounced in subjects with rapidly-fatal disease, this selection bias would tend to skew effect estimates toward the null. However, I found no evidence of SNP associations with T, N, or summary stage in our data. CHANCE also gathered no information on infection with carcinogenic types of human papillomavirus, an important cause of SCCHN especially in the oropharynx. This could alter effect estimates in similar fashion to what would happen if cigarette smoking, a significant cause of SCCHN, were excluded from models.

As with all questionnaire-based studies, CHANCE was dependent on accuracy of subject recall for socially sensitive questionnaire variables such as smoking and drinking habits. Researchers investigating self-report of addictive behaviors have concluded that its validity varies with the sensitivity of the information desired, methodologic sophistication of the questionnaire and of the interviewer (including task characteristics), and personal characteristics of the subject, with the latter believed to be of relatively less importance than other factors (214, 215), The literature suggests that in the U.S., African-Americans under-report all types of substance use, but these findings may not be generalizable (212, 242). It may be related to reduced trust, some of which is ameliorated by in-person interviews versus a more distance telephone interview (243). It is also believed that recall error is minimized when the recall period is sufficiently long (216). Many studies in a variety of populations have compared self-report smoking behavior with serum cotinine levels, with the consensus that self-reported smoking levels are highly correlated with biological assessment. CHANCE used trained nurse interviewers to ask, in person, a standardized script of questions about legal behaviors (smoking and drinking) over a long period of time (lifetime). For this reason and because we detected alcohol and tobacco associations with

SCCHN similar to those observed in other studies, I believe CHANCE measures of smoking and drinking are acceptable for my broad classifications of tertiles of lifetime alcoholic beverage consumption as a primary exposure, and lifetime duration of smoking as a confounder.

Fruit and vegetable intake, in servings per day, was also measured by questionnaire. In addition to subjects who were truly missing this information, many cases reported such high caloric intake that they were considered outliers according to accepted nutrition research practice. Their fruit and vegetable intake variables were also set to missing. As a result, due to truly missing data and suspect outlier data, I felt that the CHANCE fruit and vegetable intake variables were missing in too many subjects, and not-at-random, such that including them in regression models would bias results in unpredictable ways. Therefore I did not include them in my logistic regression models. Because the oxidative stress metabolism genes I studied may well interact with the anti-oxidants in fruits and vegetables consumed, I may have residual confounding and biased effect estimates as a result.

5.4.2 Aim 1

The number of subjects was sufficient to detect SCCHN incidence odds ratios greater than about 1.3. I had sufficient power to detect larger associations with anatomic sub-site tumors, except for hypharyngeal tumors due to the small number of cases. Power to detect main effects in African-American participants was insufficient to detect main effects except when the prevalence of the minor allele was greater than 20% (which was true for 78% of this study's SNPs in African-Americans) and the OR was greater than 1.8; i.e. two-fold effects or greater.

5.4.3 Aim 2

The wording of alcohol consumption questions on the questionnaire precluded construction of a drinking frequency variable that integrated all three beverage types. Instead, I used a lifetime consumption variable that combined information on both frequency

and duration of alcoholic beverage consumption. Based on results from numerous studies, IARC has concluded that drinking frequency is highly associated with SCCHN incidence while duration is only weakly associated; these conclusions were also borne out in CHANCE data. Beer and liquor accounted for 88% of lifetime volume of ethanol consumed by CHANCE subjects, and correlations between each beverage's frequency variable and lifetime ethanol intake in milliliters (summed over all types) are relatively high (0.81 for liquor, 0.64 for beer). Correlation between wine drinking frequency and lifetime ethanol intake is lower (0.53) but wine represented only 12% of average lifetime ethanol intake, making this a proportionally small source of error. Thus the metric available in CHANCE – overall lifetime ethanol intake – is highly correlated with the preferred metric of drink frequency.

This study had low power to detect additive interactions except when the additive effect was quite large. Therefore I did not examine gene-gene interactions and may well have missed gene-alcohol interactions of moderate effect size. Due to even smaller numbers of cases within each of the five sub-sites, I could not assess interaction between SNPs and alcohol intake in relation to anatomic site of tumor.

5.4.4 Aim 3

CHANCE collected excellent information for matching participant records to the National Death Index (NDI), including first and last names, middle initial, social security number, and date of birth. The result was that about three-quarters of identified deaths matched perfectly with a single NDI record. By the end of 2008, 39.4% of cases had died of any cause and 17.5% had died of head and neck cancer. For polymorphisms present in at least 10% of subjects – which includes 36 of 45 SNPs studied in the alcohol metabolism pathway (including 6 of 10 *CYP2E1* SNPs) and 16 of 19 SNPS studied in the oxidative stress pathway – I had at least 80% power to detect hazard ratios of 1.3 or greater for all-cause death and 1.5 or greater for cancer-specific death.

A limitation of my study was that CHANCE collected no data on recurrence (relapse-free survival) or second primary tumors, or on use of alcohol and tobacco after diagnosis. I found that drinking and smoking prior to diagnosis affect all-cause but not head-and-neck-cancer-specific survival (244), and there is extensive literature associating continued smoking post-diagnosis with worse all-cause survival and increased number of second primary head and neck tumors. In contrast, there are very few studies examining whether continuing to drink after diagnosis affects all-cause and cancer-specific survival. Existing studies are small and examined all-cause survival (not head-and-neck-cancer-specific survival) or occurrence of second primary tumors, with conflicting results (245). In any case, lacking data on post-diagnosis drinking and smoking behavior, I was not able to address the question of whether quitting smoking or drinking improves cancer-specific survival.

I had information on whether the subject received chemotherapy as treatment for their cancer, but not on the specific type of agent used. CYP2E1 is known to bio-activate platinum-containing agents, which are commonly used to treat head and neck cancers, but there are differences in CYP2E1 metabolism for different platinum-containing agents, and I could not investigate those.

5.5 IMPLICATIONS AND CONCLUSIONS

This study replicated, in a study population of mixed white and African-American ancestry, previous studies' findings for Asian, European, Indian, and Latin American populations that some polymorphisms in the alcohol metabolism pathway interact with alcohol consumption to affect risk of head and neck cancer. This is the first study to report associations of polymorphisms in oxidative stress genes with SCCHN incidence, and that polymorphisms in *CYP2E1* are associated with cancer-specific survival. Some associations may be causal, but most are probably due to linkage disequilibrium between the measured SNP and the causal polymorphism.

Elucidation of carcinogenesis pathways for head and neck cancers is not immediately useful for public health purposes, but may eventually lead to specific preventive measures and recommendations, and more effective ways of treating cancer patients. In particular, studies designed to explore possible genetic reasons for high incidence rates of laryngeal cancer in African-American men could result in targeted intervention in that population.

Alcohol metabolism genes should continue to be studied in relation to head and neck cancer, in an effort to pinpoint the functional or regulatory polymorphisms that are responsible for observed effects on cancer incidence. Questions asking about alcohol consumption should be worded so that summary measures of drinking frequency (including all beverage types) can be computed. Measuring gene-alcohol interaction and effects of rarer variants will require the statistical power of large consortia and pooling studies such as INHANCE, which combine subjects from multiple smaller studies using similar design and protocols. This will eventually involve gene sequencing as well as enzyme expression and functionality studies, and will facilitate understanding of how these mechanisms influence carcinogenesis.

As part of these investigations, *CYP2E1* variation and its interaction with chemotherapy treatments should be investigated for effects on cancer-specific survival, to replicate my findings and to pinpoint the responsible polymorphisms. This line of investigation may eventually lead to improved treatments. Also, in light of the associations this study found in oxidative stress genes, I suggest that future genetic studies on SCCHN and other cancers include polymorphisms in these genes, to replicate my findings, to explore possible interaction with fruit and vegetable intake, and to further illuminate mechanisms of carcinogenesis.

Finally, future studies of head and neck cancer should measure post-diagnosis drinking behavior, updated at specified intervals, as well as timing of recurrence and second

primary tumors, in order to address the public health question of whether reducing or eliminating drinking after diagnosis will improve all-cause and cancer-specific survival.

Appendix A

List of 157 SNPs genotyped as ancestry informative markers (AIMs)
by the method of C. Tian et al., 2006 (alphabetically ordered)

rs10028057[1]	rs12900262	rs1885167[1]	rs4529792	rs7187359
rs10041728	rs12900552	rs1911999	rs4602918	rs7189172
rs10056388	rs12926237	rs1917028	rs4619931	rs735480
rs1011643	rs12945601	rs1982235	rs4659762	rs7424137
rs10124991	rs12997060	rs1991818	rs4789070	rs7512316
rs10195705	rs1303629	rs2075902	rs4792105	rs7575147[1]
rs10202705	rs13080353[1]	rs2184033	rs4793237	rs7689609
rs10254729	rs13169284	rs222674	rs4811651[1]	rs7788641
rs10255169	rs13173738	rs2246695	rs4823460	rs7810554
rs1043809	rs13178470	rs228768	rs4859147	rs798443
rs10806263	rs13261248	rs2416791	rs4885162	rs8113143
rs10842753	rs13318432	rs2426515	rs4896780[1]	rs833282
rs10908312[1]	rs1335826	rs2451563	rs4923940	rs857440
rs10952147	rs1372115	rs2488465	rs503677	rs870272
rs10962612[1]	rs1372894	rs2593595	rs567357	rs897351
rs11000419	rs1380014	rs2596793	rs6023376	rs9297712
rs11150219	rs1412521	rs2660769	rs6414248	rs9306906
rs1117382	rs1415723	rs2687427	rs645510	rs9416026
rs11223503	rs1426654[1]	rs2777804	rs6491743	rs9416972
rs1125217[1]	rs1462309	rs316598	rs6494466	rs9525462
rs11264110	rs1470608	rs328744	rs6535244	rs9530646
rs11607932	rs1477921	rs33957	rs6556352	rs9543532
rs11652805	rs1490728	rs344454	rs6666101	rs9806307
rs11901793	rs1508061	rs3755446	rs6765491	rs9849733
rs12094678	rs155409[1]	rs3759171	rs6820509	rs9923864
rs12129648	rs16891982	rs3791896	rs6937164	
rs1256197	rs17049450	rs385194	rs7021690	[1]Genotyping
rs1257010	rs17261772	rs3861709	rs7086	failed
rs12594483	rs17269594	rs4143633	rs710052	
rs12612040	rs1733731	rs4149436	rs7107482	
rs12640848	rs17520733	rs4350528	rs7111814	
rs12676654	rs1862819	rs4489979	rs7134682	
rs12692701	rs1870571	rs4506877	rs7161[1]	

Appendix B
Alcohol questions on nurse-administered questionnaire
CHANCE study (p. 1 of 5)

E. Alcohol

Now I have some questions about (your/your _____'s) alcohol consumption. As you answer these questions, keep in mind that we are interested in (your/your ___'s) usual pattern of alcohol use.

. This section was completed by:

① Participant
② Proxy
③ Proxy Assisted

: (Have you/has your ____) ever consumed beer, wine, or hard liquor?

① Yes
② No ⇒ Sec F
⑩ Don't know ⇒ Sec F
⑦ Refused ⇒ Sec F

. (Have you consumed/did your ____ consume) 50 beers or two cases in (your/his/her) lifetime? This includes draft, malt liquor, etc. and home brew.

① Yes
② No ⇒ Q9
⑩ Don't know ⇒ Q9
⑧ Refused ⇒ Q9

4. At what age did (you/your ___) start drinking beer?

Age: ___

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑩ Don't know ⑧ Refused

5. Do you still drink beer?

① Yes ⇒ Q7
② No
⑩ Don't know ⇒ Q7
⑧ Refused ⇒ Q7

6. At what age did (you/did your ____) stop?

Age: __

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑩ Don't know ⑧ Refused

7. For how many years did (you/your ____) not drink beer during this period?

Years: ___

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑩ Don't know ⑧ Refused

Appendix B
Alcohol questions on nurse-administered questionnaire
CHANCE study (p. 2 of 5)

Cases: As I mentioned earlier, sometimes people change their behaviors after they become ill. Therefore, I need to ask about (your/your ____'s) alcohol use before (your/his/her) recent head and neck problem. The next two questions are about beer consumption before the year (ref. date).

8. How much beer did (you/he/she) usually drink? Cases: before the year (ref. date).

a. Beers: __ __

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

b. Per:

① Day ② Week ③ Month ④ Year

c. What size beer did (you/he/she) usually drink?

① 7 oz. ponies
② 12 oz. cans/bottles/glasses
③ Approx. 16 oz. cans
④ Approx. 25 oz. cans
⑤ 40 oz bottles
⑥ Other ⇒ Q8d
⓪ Don't know
⑧ Refused

d. Other (specify): _____

⓪ Don't know ⑧ Refused

Appendix B
Alcohol questions on nurse-administered questionnaire
CHANCE study (p. 3 of 5)

The following questions are about **wine** consumption.

9. (Have you consumed/did your ____ consume) wine 20 times in (your/his/her) lifetime? This includes store bought wine, sherry, port, wine from a bar or restaurant, and homemade.

① Yes
② No ⇒ Q15
⑧ Don't know ⇒ Q15
⑨ Refused ⇒ Q15

10. At what age did (you/your ___) start drinking wine?

Age: ___

⑤ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⑩ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑧ Don't know ⑨ Refused

11. Do you still drink wine?

① Yes ⇒ Q13
② No
⑧ Don't know ⇒ Q13
⑨ Refused ⇒ Q13

12. At what age did (you/did your ___) stop?

Age: ___

⑤ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⑩ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑧ Don't know ⑨ Refused

13. For how many years did (you/your ____) not drink wine during this period?

Years: ___

⑤ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⑩ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑧ Don't know ⑨ Refused

Cases: As I mentioned earlier, sometimes people change their behaviors after they become ill. Therefore, I need to ask about (your/your ____'s) alcohol use before (your/his/her) recent head and neck problem. The next two questions are about wine consumption before the year (ref. date).

14. How much wine did (you/he/she) usually drink?

a. Number of drinks of wine: ___

⑤ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⑩ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

b. Per:

① Day ② Week ③ Month ④ Year

c. What size drink did (you/he/she) usually have?

⑥ Glass (4 oz.) ⑦ Bottle (2 liters)
⑧ Glass (8 oz.) ⑨ Box (3 liters)
⑤ Half carafe (375ml) ⑩ Other⇒ Q14d
⑥ Bottle (750 ml) ⑧ Don't know
⑤ Bottle (1.5 liters) ⑨ Refused

d. Other (specify): _____

⑧ Don't know ⑨ Refused

2758

Appendix B
Alcohol questions on nurse-administered questionnaire
CHANCE study (p. 4 of 5)

The following questions are about **hard liquor** consumption.

15. (Have you consumed/ did your_____ consume) hard liquor 20 times in (your/his/her) lifetime? This includes straight and mixed drinks and moonshine.

① Yes
② No ⇒ Sec F
⑧ Don't know ⇒ Sec F
⑨ Refused ⇒ Sec F

16. At what age did (you/your___) start drinking hard liquor?

Age: __ __

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑧ Don't know ⑨ Refused

17. Do you still drink hard liquor?

① Yes ⇒ Q20
② No
⑧ Don't know ⇒ Q20
⑨ Refused ⇒ Q20

18. At what age did (you/your____) stop?

Age: __ __

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑧ Don't know ⑨ Refused

19. For how many years did (you/your____) not drink hard liquor during this period?

Years: __ __

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⑧ Don't know ⑨ Refused

Cases: As I mentioned earlier, sometimes people change their behaviors after they become ill. Therefore, I need to ask about (you/your ____'s) alcohol use before (your/his/her) recent head and neck problem. The next few questions are about hard liquor consumption before the year (ref. date).

20. How much hard liquor did (you/he/she) usually drink? Cases: before the year (ref. date).

a. Number of drinks:

① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

b. Per

① Day ② Week ③ Month ④ Year

c. What size drink (did you/he/she) usually have?

ⓐ Shot glass ⓗ Half-gallon
ⓒ Half pint ⓛ Gallon
ⓟ Pint ⓞ Other ⇒ Q20d
ⓕ Fifth ⓨ Don't know
ⓠ Quart ⓡ Refused

d. Other (specify): _____

ⓝ Don't know ⓗ Refused

Appendix B
Alcohol questions on nurse-administered questionnaire
CHANCE study (p. 5 of 5)

11. What type of hard liquor did (you/he/she) usually drink? Cases: Remember to think about the time before the year (ref. date).

[SHOW CARD. SELECT ALL THAT APPLY.]

(A)	Gin (Clear)	(J)	Rum (Light)
(B)	Gin (Fruit)	(K)	Rum (Dark)
(C)	Gin (Unsure)	(L)	Rum (Unsure)
(D)	Bourbon	(M)	Tequila (Clear)
(E)	Whiskey	(N)	Tequila (Gold)
(F)	Brandy	(O)	Tequila (unsure)
(G)	Vodka	(P)	Don't know
(H)	Scotch	(Q)	Refused
(I)	Homemade liquor	(R)	Other⇒ Q21a

a. Other (specify): _____

12. Did (you/your [illegible]) usually drink it straight or combine it with a mixer?

(1)	Drank straight	
(2)	Added mixer	
(3)	Drank straight and mixed drinks	
(8)	Don't know	⇒ Sec F
(9)	Refused	⇒ Sec F

We are finished with the questions about alcohol consumption.

APPENDIX C Distribution of SNP genotypes in cases and controls, by race

				Whites				African-Americans			
				Cases (n=922)		Controls (n=1074)		Cases (n=305)		Controls (n=251)	
Gene	SNP	Major/minor allele	Genotype	n	(col %)	n	(col %)	n	(col %)	n	(col %)
ALCOHOL METABOLISM GENES											
ADH1B	rs12507573	C/A	AA	188	20.4%	234	21.8%	56	18.4%	40	15.9%
			AC	462	50.1%	536	49.9%	149	48.9%	127	50.6%
			CC	272	29.5%	304	28.3%	100	32.8%	83	33.1%
			missing	0	0.0%	0	0.0%	0	0.0%	1	0.4%
	rs1042026	A/G	GG	73	7.9%	81	7.5%	3	1.0%	5	2.0%
			GA	382	41.4%	457	42.6%	54	17.7%	36	14.3%
			AA	467	50.7%	536	49.9%	248	81.3%	210	83.7%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs7673353	C/T	TT	0	0.0%	0	0.0%	7	2.3%	12	4.8%
			TC	3	0.3%	6	0.6%	89	29.2%	65	25.9%
			CC	919	99.7%	1068	99.4%	209	68.5%	174	69.3%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs17028834	T/C	CC	1	0.1%	0	0.0%	2	0.7%	5	2.0%
			CT	3	0.3%	2	0.2%	80	26.2%	50	19.9%
			TT	918	99.6%	1072	99.8%	223	73.1%	196	78.1%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1693457	T/C	CC	27	2.9%	27	2.5%	9	3.0%	16	6.4%
			CT	297	32.2%	333	31.0%	88	28.9%	86	34.3%
			TT	597	64.8%	714	66.5%	208	68.2%	148	59.0%
			missing	1	0.1%	0	0.0%	0	0.0%	1	0.4%
	rs1229984	G/A	AA	0	0.0%	3	0.3%	0	0.0%	0	0.0%
			AG	30	3.3%	71	6.6%	1	0.3%	5	2.0%
			GG	892	96.7%	1000	93.1%	304	99.7%	246	98.0%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1159918	G/T	TT	129	14.0%	138	12.8%	161	52.8%	140	55.8%
			TG	419	45.4%	485	45.2%	116	38.0%	93	37.1%
			GG	374	40.6%	451	42.0%	28	9.2%	18	7.2%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1229982	G/T	TT	30	3.3%	45	4.2%	69	22.6%	51	20.3%
			TG	282	30.6%	327	30.4%	146	47.9%	125	49.8%
			GG	610	66.2%	702	65.4%	90	29.5%	75	29.9%
			missing	0.0%	0	0.0%	0	0.0%	0	0.0%	0.0%
ADH1C	rs2298753	T/C	CC	15	1.6%	15	1.4%	0	0.0%	0	0.0%
			CT	186	20.2%	177	16.5%	20	6.6%	14	5.6%
			TT	721	78.2%	882	82.1%	285	93.4%	237	94.4%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1614972	C/T	TT	73	7.9%	96	8.9%	67	22.0%	67	26.7%
			TC	403	43.7%	446	41.5%	144	47.2%	129	51.4%
			CC	446	48.4%	532	49.5%	94	30.8%	55	21.9%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1391088	C/A	AA	3	0.3%	4	0.4%	2	0.7%	2	0.8%
			AC	118	12.8%	161	15.0%	55	18.0%	36	14.3%
			CC	801	86.9%	909	84.6%	248	81.3%	213	84.9%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1693482	C/T	TT	158	17.1%	172	16.0%	6	2.0%	4	1.6%
			TC	437	47.4%	497	46.3%	94	30.8%	64	25.5%
			CC	325	35.2%	403	37.5%	205	67.2%	183	72.9%
			missing	2	0.2%	2	0.2%	0	0.0%	0	0.0%
	rs1631460	C/G	GG	160	17.4%	173	16.1%	6	2.0%	4	1.6%
			GC	443	48.0%	508	47.3%	95	31.1%	65	25.9%
			CC	318	34.5%	393	36.6%	204	66.9%	182	72.5%
			missing	1	0.1%	0	0.0%	0	0.0%	0	0.0%
	rs11936869	C/G	GG	64	6.9%	69	6.4%	53	17.4%	56	22.3%
			GC	354	38.4%	430	40.0%	143	46.9%	122	48.6%
			CC	504	54.7%	575	53.5%	109	35.7%	72	28.7%
			missing	0	0.0%	0	0.0%	0	0.0%	1	0.4%
ADH4	rs29001227	A/T	TT	1	0.1%	1	0.1%	8	2.6%	8	3.2%
			TA	4	0.4%	1	0.1%	92	30.2%	59	23.5%
			AA	917	99.5%	1072	99.8%	205	67.2%	184	73.3%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1126672	C/T	TT	53	5.7%	82	7.6%	6	2.0%	2	0.8%
			TC	386	41.9%	434	40.4%	60	19.7%	46	18.3%
			CC	483	52.4%	558	52.0%	239	78.4%	203	80.9%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs4699710	T/C	CC	70	7.6%	98	9.1%	13	4.3%	5	2.0%
			CT	412	44.7%	458	42.6%	90	29.5%	78	31.1%

APPENDIX C Distribution of SNP genotypes in cases and controls, by race

				Whites				African-Americans			
				Cases (n=922)		Controls (n=1074)		Cases (n=305)		Controls (n=251)	
Gene	SNP	Major/minor allele	Genotype	n	(col %)	n	(col %)	n	(col %)	n	(col %)
			TT	440	47.7%	518	48.2%	202	66.2%	168	66.9%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs10017466	T/C	CC	72	7.8%	99	9.2%	40	13.1%	32	12.7%
			CT	414	44.9%	459	42.7%	143	46.9%	101	40.2%
			TT	436	47.3%	516	48.0%	121	39.7%	118	47.0%
			missing	0	0.0%	0	0.0%	1	0.3%	0	0.0%
	rs1800759	C/A	AA	114	12.4%	150	14.0%	171	56.1%	152	60.6%
			AC	472	51.2%	527	49.1%	113	37.0%	83	33.1%
			CC	336	36.4%	397	37.0%	21	6.9%	16	6.4%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1800761	G/A	AA	31	3.4%	28	2.6%	26	8.5%	22	8.8%
			AG	293	31.8%	343	31.9%	124	40.7%	100	39.8%
			GG	598	64.9%	703	65.5%	155	50.8%	129	51.4%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs3762894	T/C	CC	25	2.7%	20	1.9%	14	4.6%	17	6.8%
			CT	247	26.8%	306	28.5%	99	32.5%	73	29.1%
			TT	650	70.5%	748	69.6%	192	63.0%	161	64.1%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
ADH7	rs284787	C/T	TT	61	6.6%	58	5.4%	6	2.0%	6	2.4%
			TC	339	36.8%	384	35.8%	78	25.6%	57	22.7%
			CC	522	56.6%	632	58.8%	221	72.5%	188	74.9%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs894369	C/G	GG	32	3.5%	43	4.0%	7	2.3%	6	2.4%
			GC	275	29.8%	356	33.1%	77	25.2%	68	27.1%
			CC	615	66.7%	675	62.8%	221	72.5%	176	70.1%
			missing	0	0.0%	0	0.0%	0	0.0%	1	0.4%
	rs17588403	T/A	AA	44	4.8%	31	2.9%	4	1.3%	3	1.2%
			AT	274	29.7%	344	32.0%	66	21.6%	58	23.1%
			TT	604	65.5%	699	65.1%	235	77.0%	190	75.7%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1154454	T/C	CC	24	2.6%	31	2.9%	49	16.1%	48	19.1%
			CT	250	27.1%	309	28.8%	146	47.9%	128	51.0%
			TT	648	70.3%	734	68.3%	110	36.1%	75	29.9%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1154456	T/C	CC	116	12.6%	127	11.8%	7	2.3%	7	2.8%
			CT	438	47.5%	494	46.0%	87	28.5%	68	27.1%
			TT	368	39.9%	453	42.2%	211	69.2%	176	70.1%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1154460	G/A	AA	194	21.0%	218	20.3%	66	21.6%	58	23.1%
			AG	466	50.5%	544	50.7%	169	55.4%	127	50.6%
			GG	259	28.1%	312	29.1%	70	23.0%	65	25.9%
			missing	3	0.3%	0	0.0%	0	0.0%	1	0.4%
	rs971074	G/A	AA	7	0.8%	10	0.9%	8	2.6%	5	2.0%
			AG	165	17.9%	210	19.6%	101	33.1%	70	27.9%
			GG	750	81.3%	854	79.5%	196	64.3%	176	70.1%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1573496	C/G	GG	5	0.5%	9	0.8%	0	0.0%	0	0.0%
			GC	154	16.7%	190	17.7%	7	2.3%	6	2.4%
			CC	762	82.6%	875	81.5%	298	97.7%	245	97.6%
			missing	1	0.1%	0	0.0%	0	0.0%	0	0.0%
ALDH2	rs4767939	A/G	GG	35	3.8%	39	3.6%	52	17.0%	38	15.1%
			GA	297	32.2%	315	29.3%	145	47.5%	125	49.8%
			AA	589	63.9%	720	67.0%	108	35.4%	88	35.1%
			missing	1	0.1%	0	0.0%	0	0.0%	0	0.0%
	rs2238151	T/C	CC	119	12.9%	125	11.6%	257	84.3%	179	71.3%
			CT	436	47.3%	464	43.2%	46	15.1%	71	28.3%
			TT	363	39.4%	482	44.9%	2	0.7%	1	0.4%
			missing	4	0.4%	3	0.3%	0	0.0%	0	0.0%
	rs7312055	G/A	AA	0	0.0%	0	0.0%	28	9.2%	12	4.8%
			AG	5	0.5%	8	0.7%	111	36.4%	101	40.2%
			GG	917	99.5%	1066	99.3%	166	54.4%	138	55.0%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs2158029	G/A	AA	0	0.0%	0	0.0%	6	2.0%	1	0.4%
			AG	6	0.7%	11	1.0%	45	14.8%	41	16.3%
			GG	916	99.3%	1063	99.0%	254	83.3%	209	83.3%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs16941667	C/T	TT	6	0.7%	5	0.5%	6	2.0%	4	1.6%

APPENDIX C Distribution of SNP genotypes in cases and controls, by race

				Whites				African-Americans			
				Cases (n=922)		Controls (n=1074)		Cases (n=305)		Controls (n=251)	
Gene	SNP	Major/minor allele	Genotype	n	(col %)	n	(col %)	n	(col %)	n	(col %)
			TC	153	16.6%	158	14.7%	74	24.3%	58	23.1%
			CC	763	82.8%	911	84.8%	225	73.8%	188	74.9%
			missing	0	0.0%	0	0.0%	0	0.0%	1	0.4%
	rs16941669	T/G	GG	14	1.5%	11	1.0%	2	0.7%	0	0.0%
			GT	174	18.9%	190	17.7%	50	16.4%	36	14.3%
			TT	734	79.6%	872	81.2%	253	83.0%	214	85.3%
			missing	0	0.0%	1	0.1%	0	0.0%	1	0.4%
CYP2E1	rs3813865	G/C	CC	0	0.0%	1	0.1%	4	1.3%	3	1.2%
			CG	45	4.9%	49	4.6%	76	24.9%	53	21.1%
			GG	877	95.1%	1023	95.3%	225	73.8%	195	77.7%
			missing	0	0.0%	1	0.1%	0	0.0%	0	0.0%
	rs3813867	G/C	CC	1	0.1%	0	0.0%	1	0.3%	1	0.4%
			CG	51	5.5%	58	5.4%	30	9.8%	25	10.0%
			GG	870	94.4%	1016	94.6%	273	89.5%	224	89.2%
			missing	0	0.0%	0	0.0%	1	0.3%	1	0.4%
	rs8192772	T/C	CC	6	0.7%	7	0.7%	3	1.0%	2	0.8%
			CT	127	13.8%	146	13.6%	64	21.0%	53	21.1%
			TT	789	85.6%	921	85.8%	237	77.7%	196	78.1%
			missing	0	0.0%	0	0.0%	1	0.3%	0	0.0%
	rs915908	G/A	AA	21	2.3%	27	2.5%	0	0.0%	1	0.4%
			AG	211	22.9%	266	24.8%	18	5.9%	22	8.8%
			GG	690	74.8%	781	72.7%	287	94.1%	228	90.8%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs915909	C/T	TT	0	0.0%	0	0.0%	5	1.6%	2	0.8%
			TC	1	0.1%	4	0.4%	53	17.4%	40	15.9%
			CC	921	99.9%	1070	99.6%	246	80.7%	204	81.3%
			missing	0	0.0%	0	0.0%	1	0.3%	5	2.0%
	rs7092584	C/T	TT	11	1.2%	9	0.8%	4	1.3%	5	2.0%
			TC	168	18.2%	197	18.3%	80	26.2%	58	23.1%
			CC	743	80.6%	868	80.8%	219	71.8%	187	74.5%
			missing	0	0.0%	0	0.0%	2	0.7%	1	0.4%
	rs743535	C/T	TT	11	1.2%	6	0.6%	7	2.3%	4	1.6%
			CT	149	16.2%	176	16.4%	82	26.9%	57	22.7%
			CC	757	82.1%	886	82.5%	212	69.5%	188	74.9%
			missing	5	0.5%	6	0.6%	4	1.3%	2	0.8%
	rs2249695	C/T	TT	34	3.7%	41	3.8%	28	9.2%	23	9.2%
	"tipsy" SNP		TC	292	31.7%	372	34.6%	124	40.7%	103	41.0%
			CC	595	64.5%	661	61.5%	153	50.2%	124	49.4%
			missing	1	0.1%	0	0.0%	0	0.0%	1	0.4%
	rs28969387	A/T	TT	0	0.0%	0	0.0%	0	0.0%	0	0.0%
			TA	2	0.2%	0	0.0%	23	7.5%	20	8.0%
			AA	920	99.8%	1074	100.0%	282	92.5%	231	92.0%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs11101812	T/C	CC	0	0.0%	0	0.0%	2	0.7%	1	0.4%
			CT	2	0.2%	1	0.1%	22	7.2%	18	7.2%
			TT	919	99.7%	1073	99.9%	279	91.5%	230	91.6%
			missing	1	0.1%	0	0.0%	2	0.7%	2	0.8%
OXIDATIVE STRESS METABOLISM GENES											
CAT	rs1049982	C/T	TT	95	10.3%	106	9.9%	92	30.2%	76	30.3%
			TC	396	43.0%	449	41.8%	147	48.2%	123	49.0%
			CC	429	46.5%	514	47.9%	65	21.3%	50	19.9%
			missing	2	0.2%	5	0.5%	1	0.3%	2	0.8%
GPX1	rs8179172	T/A	AA	0	0.0%	0	0.0%	0	0.0%	1	0.4%
			AT	1	0.1%	2	0.2%	50	16.4%	48	19.1%
			TT	921	99.9%	1072	99.8%	255	83.6%	202	80.5%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1800668	C/T	TT	77	8.4%	101	9.4%	15	4.9%	19	7.6%
			TC	381	41.3%	437	40.7%	115	37.7%	91	36.3%
			CC	464	50.3%	533	49.6%	174	57.0%	140	55.8%
			missing	0	0.0%	3	0.3%	1	0.3%	1	0.4%
	rs3811699	A/G	GG	79	8.6%	100	9.3%	23	7.5%	28	11.2%
			GA	379	41.1%	442	41.2%	133	43.6%	105	41.8%
			AA	464	50.3%	532	49.5%	149	48.9%	118	47.0%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs3448	C/T	TT	84	9.1%	83	7.7%	29	9.5%	17	6.8%
			TC	350	38.0%	397	37.0%	115	37.7%	96	38.2%
			CC	488	52.9%	594	55.3%	161	52.8%	138	55.0%

APPENDIX C Distribution of SNP genotypes in cases and controls, by race

				Whites				African-Americans			
				Cases (n=922)		Controls (n=1074)		Cases (n=305)		Controls (n=251)	
Gene	SNP	Major/minor allele	Genotype	n	(col %)	n	(col %)	n	(col %)	n	(col %)
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
GPX2	rs11623705	G/T	TT	13	1.4%	15	1.4%	0	0.0%	0	0.0%
			TG	199	21.6%	220	20.5%	15	4.9%	19	7.6%
			GG	710	77.0%	839	78.1%	290	95.1%	232	92.4%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs2412065	G/C	CC	37	4.0%	55	5.1%	48	15.7%	47	18.7%
			CG	290	31.5%	369	34.4%	154	50.5%	119	47.4%
			GG	595	64.5%	650	60.5%	103	33.8%	85	33.9%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs2737844	C/T	TT	76	8.2%	109	10.1%	209	68.5%	162	64.5%
			TC	361	39.2%	462	43.0%	82	26.9%	75	29.9%
			CC	481	52.2%	498	46.4%	13	4.3%	14	5.6%
			missing	4	0.4%	5	0.5%	1	0.3%	0	0.0%
GPX4	rs757229	G/C	CC	192	20.8%	233	21.7%	104	34.1%	86	34.3%
			CG	484	52.5%	522	48.6%	145	47.5%	123	49.0%
			GG	246	26.7%	318	29.6%	55	18.0%	42	16.7%
			missing	0	0.0%	1	0.1%	1	0.3%	0	0.0%
SOD1	rs11910115	A/C	CC	0	0.0%	0	0.0%	2	0.7%	5	2.0%
			CA	1	0.1%	1	0.1%	59	19.3%	59	23.5%
			AA	921	99.9%	1073	99.9%	244	80.0%	187	74.5%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs4998557	G/A	AA	12	1.3%	14	1.3%	41	13.4%	40	15.9%
			AG	205	22.2%	195	18.2%	137	44.9%	111	44.2%
			GG	704	76.4%	862	80.3%	127	41.6%	100	39.8%
			missing	1	0.1%	3	0.3%	0	0.0%	0	0.0%
	rs10432782	T/G	GG	12	1.3%	14	1.3%	20	6.6%	20	8.0%
			GT	205	22.2%	194	18.1%	113	37.0%	101	40.2%
			TT	704	76.4%	865	80.5%	172	56.4%	130	51.8%
			missing	1	0.1%	1	0.1%	0	0.0%	0	0.0%
	rs2070424	A/G	GG	6	0.7%	5	0.5%	6	2.0%	7	2.8%
			AG	127	13.8%	113	10.5%	83	27.2%	77	30.7%
			AA	789	85.6%	956	89.0%	216	70.8%	167	66.5%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs1041740	C/T	TT	88	9.5%	101	9.4%	6	2.0%	3	1.2%
			TC	387	42.0%	473	44.0%	62	20.3%	43	17.1%
			CC	447	48.5%	500	46.6%	237	77.7%	205	81.7%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
SOD2	rs4342445	G/A	AA	57	6.2%	56	5.2%	5	1.6%	4	1.6%
			AG	340	36.9%	355	33.1%	71	23.3%	59	23.5%
			GG	525	56.9%	663	61.7%	229	75.1%	188	74.9%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs2842980	A/T	TT	42	4.6%	61	5.7%	30	9.8%	24	9.6%
			TA	299	32.4%	338	31.5%	139	45.6%	92	36.7%
			AA	581	63.0%	675	62.8%	136	44.6%	135	53.8%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs8031	T/A	AA	206	22.3%	267	24.9%	31	10.2%	35	13.9%
			AT	467	50.7%	534	49.7%	134	43.9%	107	42.6%
			TT	249	27.0%	273	25.4%	140	45.9%	109	43.4%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs5746134	C/T	TT	0	0.0%	0	0.0%	7	2.3%	3	1.2%
			TC	0	0.0%	4	0.4%	93	30.5%	58	23.1%
			CC	922	100.0%	1070	99.6%	205	67.2%	190	75.7%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	rs2758331	C/A	AA	204	22.1%	264	24.6%	18	5.9%	18	7.2%
			AC	468	50.8%	537	50.0%	113	37.0%	87	34.7%
			CC	250	27.1%	273	25.4%	174	57.0%	146	58.2%
			missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%

APPENDIX D. SNP effects on odds of developing cancer (dominant genetic model)

			Oral cavity, oropharyngeal, hypopharyngeal cancer NOS				Oropharyngeal cancer			
			n cases/controls				n cases/controls			
Gene	SNP	major/ minor alleles	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]	homozygous for major allele	one or two copies of minor allele	Adjusted OR[b] (95% CI)	p-value[c]
ALCOHOL METABOLISM GENES										
ADH1B	rs1250757	C/A	71/386	152/935	0.94 (0.80-1.11)	1.00	91/386	241/935	1.02 (0.88-1.17)	1.00
	rs1042026	A/G	125/744	98/578	1.04 (0.89-1.22)	1.00	193/744	139/578	0.97 (0.84-1.11)	1.00
	rs7673353	C/T	204/1239	19/83	1.09 (0.79-1.50)	1.00	309/1239	23/83	0.97 (0.72-1.29)	1.00
	rs17028834	T/C	211/1265	12/57	1.05 (0.72-1.53)	1.00	312/1265	20/57	1.29 (0.94-1.77)	1.00
	rs1693457	T/C	136/860	86/461	1.13 (0.97-1.32)	1.00	223/860	109/461	1.02 (0.89-1.18)	1.00
	rs1229984	G/A	219/1243	4/79	0.60 (0.36-1.01)	1.00	322/1243	10/79	0.74 (0.52-1.06)	1.00
	rs1159918	G/T	65/468	158/854	1.16 (0.98-1.37)	1.00	109/468	223/854	1.12 (0.97-1.29)	1.00
	rs1229982	G/T	129/775	94/547	0.99 (0.84-1.16)	1.00	197/775	135/547	0.98 (0.85-1.12)	1.00
ADH1C	rs2298753	T/C	180/1117	43/205	1.18 (0.97-1.44)	1.00	270/1117	62/205	1.19 (1.00-1.41)	1.00
	rs1614972	C/T	95/585	128/737	1.02 (0.87-1.19)	1.00	141/585	191/737	1.02 (0.89-1.16)	1.00
	rs1391088	C/A	192/1119	31/203	0.98 (0.79-1.22)	1.00	280/1119	52/203	1.03 (0.86-1.24)	1.00
	rs1693482	C/T	97/585	125/735	1.03 (0.88-1.20)	1.00	145/585	186/735	1.03 (0.90-1.18)	1.00
	rs1631460	C/G	93/574	130/748	1.05 (0.90-1.23)	1.00	144/574	188/748	1.02 (0.89-1.17)	1.00
	rs1193686	C/G	111/645	112/676	0.96 (0.83-1.12)	1.00	161/645	171/676	0.99 (0.87-1.13)	1.00
ADH4	rs29001227	A/T	208/1253	15/69	1.00 (0.70-1.42)	1.00	305/1253	27/69	1.33 (0.99-1.79)	1.00
	rs1126672	C/T	131/758	92/564	1.02 (0.87-1.19)	1.00	201/758	131/564	0.96 (0.83-1.10)	1.00
	rs4699710	T/C	120/684	103/638	0.98 (0.84-1.13)	1.00	179/684	153/638	0.96 (0.84-1.10)	1.00
	rs10017466	T/C	106/632	117/690	1.00 (0.87-1.16)	1.00	157/632	174/690	1.02 (0.89-1.16)	1.00
	rs1800759	C/A	75/412	148/910	0.93 (0.79-1.10)	1.00	101/412	231/910	1.00 (0.87-1.16)	1.00
	rs1800761	G/A	134/830	89/492	1.01 (0.87-1.18)	1.00	193/830	139/492	1.07 (0.94-1.22)	1.00
	rs3762894	T/C	150/906	73/416	1.00 (0.86-1.18)	1.00	218/906	114/416	1.04 (0.91-1.20)	1.00
ADH7	rs284787	C/T	126/817	97/505	1.14 (0.98-1.33)	1.00	208/817	124/505	1.01 (0.88-1.16)	1.00
	rs894369	C/G	153/849	70/472	0.93 (0.80-1.09)	1.00	217/849	115/472	0.96 (0.84-1.10)	1.00
	rs17588403	T/A	157/887	66/435	0.92 (0.78-1.08)	1.00	230/887	102/435	0.94 (0.81-1.08)	1.00
	rs1154454	T/C	134/808	89/514	0.95 (0.81-1.11)	1.00	208/808	124/514	0.94 (0.82-1.09)	1.00
	rs1154456	T/C	108/629	115/693	0.98 (0.84-1.14)	1.00	150/629	182/693	1.08 (0.94-1.24)	1.00
	rs1154460	G/A	61/377	161/944	1.01 (0.86-1.20)	1.00	88/377	244/944	1.06 (0.91-1.23)	1.00
	rs971074	G/A	168/1027	55/295	1.07 (0.90-1.27)	1.00	257/1027	75/295	0.95 (0.81-1.12)	1.00
	rs1573496	C/G	186/1117	36/205	1.08 (0.88-1.33)	1.00	292/1117	40/205	0.84 (0.69-1.02)	1.00
ALDH2	rs4767939	A/G	135/806	88/516	0.99 (0.85-1.16)	1.00	191/806	140/516	1.08 (0.94-1.23)	1.00
	rs2238151	T/C	66/482	156/837	1.13 (0.95-1.34)	1.00	114/482	217/837	1.02 (0.88-1.18)	1.00
	rs7312055	G/A	196/1201	27/121	0.88 (0.65-1.20)	1.00	301/1201	31/121	0.83 (0.64-1.09)	1.00
	rs2158029	G/A	216/1269	7/53	0.74 (0.47-1.14)	1.00	316/1269	16/53	1.06 (0.76-1.47)	1.00
	rs16941667	C/T	189/1097	34/224	0.96 (0.78-1.18)	1.00	266/1097	66/224	1.11 (0.94-1.31)	1.00
	rs16941669	T/G	175/1084	48/236	1.14 (0.95-1.37)	1.00	272/1084	60/236	0.98 (0.83-1.16)	1.00
CYP2E1	rs3813865	G/C	204/1215	19/106	1.02 (0.77-1.36)	1.00	298/1215	34/106	1.11 (0.88-1.40)	1.00
	rs3813867	G/C	207/1237	16/84	1.02 (0.76-1.37)	1.00	316/1237	16/84	0.81 (0.60-1.09)	1.00
	rs8192772	T/C	195/1114	28/208	0.92 (0.73-1.15)	1.00	279/1114	53/208	1.07 (0.90-1.28)	1.00
	rs915908	G/A	164/1006	59/316	1.11 (0.93-1.32)	1.00	263/1006	69/316	0.89 (0.75-1.04)	1.00
	rs915909	C/T	212/1271	11/46	1.11 (0.75-1.62)	1.00	317/1271	15/46	1.02 (0.72-1.45)	1.00
	rs7092584	C/T	181/1052	42/269	0.98 (0.81-1.19)	1.00	262/1052	70/269	1.05 (0.89-1.23)	1.00
	rs743535	C/T	182/1071	41/243	0.99 (0.82-1.20)	1.00	268/1071	60/243	1.00 (0.85-1.19)	1.00
	rs2249695	C/T	108/682	115/639	1.01 (0.86-1.19)	1.00	180/682	152/639	0.90 (0.78-1.04)	1.00
	rs28969387	A/T	219/1302	4/20	1.06 (0.59-1.90)	1.00	326/1302	6/20	0.93 (0.56-1.53)	1.00
	rs11101812	T/C	219/1300	3/20	0.98 (0.52-1.85)	1.00	327/1300	4/20	0.92 (0.52-1.63)	1.00
OXIDATIVE STRESS METABOLISM GENES										
CAT	rs1049982	C/T	100/589	123/726	0.97 (0.831.13)	1.00	134/589	197/726	1.07 (0.94-1.22)	1.00
GPX1	rs8179172	T/A	212/1271	11/51	0.97 (0.661.44)	1.00	320/1271	12/51	0.90 (0.62-1.29)	1.00
	rs1800668	C/T	113/672	110/646	1.06 (0.911.23)	1.00	172/672	160/646	1.02 (0.89-1.16)	1.00
	rs3811699	A/G	109/649	114/673	1.06 (0.911.23)	1.00	167/649	165/673	1.01 (0.89-1.15)	1.00
	rs3448	C/T	113/729	110/593	1.12 (0.971.31)	1.00	190/729	142/593	0.98 (0.86-1.12)	1.00
GPX2	rs11623705	G/T	185/1069	38/253	0.94 (0.771.14)	1.00	268/1069	64/253	1.02 (0.86-1.21)	1.00
	rs2412065	G/C	122/733	101/589	0.98 (0.841.15)	1.00	193/733	139/589	0.94 (0.82-1.08)	1.00
	rs2737844	C/T	88/511	134/806	0.95 (0.811.12)	1.00	136/511	194/806	0.93 (0.81-1.08)	1.00
GPX4	rs757229	G/C	60/358	162/963	0.98 (0.831.16)	1.00	85/358	247/963	1.02 (0.88-1.18)	1.00
SOD1	rs11910115	A/C	208/1257	15/65	1.01 (0.701.45)	1.00	319/1257	13/65	0.86 (0.60-1.22)	1.00
	rs4998557	G/A	155/959	68/360	1.00 (0.841.20)	1.00	232/959	99/360	1.07 (0.92-1.25)	1.00
	rs10432782	T/G	163/992	60/329	0.98 (0.821.17)	1.00	240/992	91/329	1.07 (0.92-1.25)	1.00
	rs2070424	A/G	187/1120	36/202	0.95 (0.761.17)	1.00	274/1120	58/202	1.05 (0.88-1.26)	1.00
	rs1041740	C/T	127/705	96/617	0.94 (0.801.10)	1.00	174/705	158/617	0.99 (0.87-1.14)	1.00
SOD2	rs4342445	G/A	145/850	78/472	1.01 (0.861.18)	1.00	202/850	130/472	1.08 (0.95-1.24)	1.00
	rs2842980	A/T	129/807	94/515	1.06 (0.911.23)	1.00	189/807	143/515	1.09 (0.96-1.25)	1.00
	rs8031	T/A	77/382	146/940	0.91 (0.771.07)	1.00	102/382	230/940	0.99 (0.86-1.14)	1.00
	rs5746134	C/T	207/1257	16/65	1.20 (0.851.68)	1.00	304/1257	28/65	1.41 (1.05-1.89)	1.00
	rs2758331	C/A	86/419	137/903	0.90 (0.771.06)	1.00	108/419	224/903	1.02 (0.88-1.17)	1.00

[a] Cases and controls do not sum to 1227 and 1325, respectively, because 4 cases and 3 controls are missing information on duration of cigarette smoking, and because a few subjects lack genotype information for some SNPs

[b] Conditional logistic regression models conditioned on sex, race, and age category, and adjusted for continuous duration of smoking in years (rounded to whole years). Odds ratios are for those with one or more copies of the minor allele versus the referent group of those homozygous for the major allele (dominant genetic model).

[c] Bonferroni-corrected for 64 statistical tests

APPENDIX E. Odds ratios for developing SCCHN for SNP genotypes, at varying levels of lifetime alcohol consumption

				Odds ratios in comparison to common referent group[a]				Assessment of additive interaction between SNP and lifetime alcohol consumption
				homozygous major allele		at least one minor allele		
Gene	SNP	major/ minor alleles	lifetime ethanol consumption (ml)	n (cases/ controls)[b]	Adjusted odds ratio (95% CI)	n (cases/ controls)[b]	Adjusted odds ratio (95% CI)	ICR[c] (**bold** if statistically significant after Bonferroni correction)
ALCOHOL METABOLISM GENES								
ADH1B	rs12507573	C/A	never drinkers	40 / 72	1.00 (ref)	77 / 208	0.64 (0.39 - 1.06)	-0.19
			>0 to <134,699	68 / 142	0.62 (0.37 - 1.05)	141 / 323	0.52 (0.32 - 0.85)	
			134,699 to <757,550	84 / 118	0.74 (0.44 - 1.27)	234 / 242	1.05 (0.65 - 1.69)	
			757,550+	160 / 47	2.75 (1.56 - 4.86)	345 / 126	2.21 (1.34 - 3.65)	
	rs1042026	A/G	never drinkers	59 / 157	1.00 (ref)	58 / 123	1.20 (0.75 - 1.93)	-0.75
			>0 to <134,699	112 / 253	0.80 (0.53 - 1.20)	97 / 213	0.85 (0.56 - 1.30)	
			134,699 to <757,550	188 / 206	1.48 (0.99 - 2.22)	130 / 154	1.31 (0.86 - 2.01)	
			757,550+	312 / 98	3.75 (2.42 - 5.81)	193 / 75	3.20 (2.02 - 5.08)	
	rs7673353	C/T	never drinkers	115 / 261	1.00 (ref)	2 / 19	0.29 (0.06 - 1.35)	1.06
			>0 to <134,699	199 / 438	0.72 (0.53 - 0.98)	10 / 28	0.71 (0.31 - 1.62)	
			134,699 to <757,550	289 / 346	1.17 (0.85 - 1.60)	29 / 14	3.14 (1.44 - 6.84)	
			757,550+	453 / 159	3.10 (2.19 - 4.39)	52 / 14	3.44 (1.66 - 7.15)	
	rs17028834	T/C	never drinkers	113 / 272	1.00 (ref)	4 / 8	1.13 (0.28 - 4.49)	3.55
			>0 to <134,699	201 / 447	0.75 (0.55 - 1.01)	8 / 19	1.01 (0.40 - 2.54)	
			134,699 to <757,550	303 / 342	1.29 (0.95 - 1.77)	15 / 18	1.39 (0.61 - 3.16)	
			757,550+	453 / 163	3.06 (2.16 - 4.32)	52 / 10	6.73 (3.06 - 14.81)	
	rs1693457	T/C	never drinkers	78 / 177	1.00 (ref)	38 / 103	0.86 (0.52 - 1.41)	1.02
			>0 to <134,699	146 / 295	0.77 (0.54 - 1.11)	63 / 170	0.64 (0.42 - 0.98)	
			134,699 to <757,550	201 / 234	1.21 (0.84 - 1.75)	117 / 126	1.31 (0.87 - 1.98)	
			757,550+	327 / 122	2.85 (1.92 - 4.24)	178 / 51	3.72 (2.34 - 5.93)	
	rs1229984	G/A	never drinkers	109 / 268	1.00 (ref)	8 / 12	1.60 (0.59 - 4.32)	-1.85
	previously studied		>0 to <134,699	206 / 427	0.83 (0.61 - 1.13)	3 / 39	0.15 (0.04 - 0.50)	
			134,699 to <757,550	305 / 339	1.35 (0.98 - 1.85)	13 / 21	1.07 (0.48 - 2.37)	
			757,550+	499 / 169	3.36 (2.38 - 4.75)	6 / 4	2.11 (0.49 - 9.10)	
	rs1159918	G/T	never drinkers	49 / 101	1.00 (ref)	68 / 179	0.90 (0.55 - 1.46)	1.03
			>0 to <134,699	80 / 168	0.67 (0.41 - 1.07)	129 / 298	0.73 (0.47 - 1.14)	
			134,699 to <757,550	107 / 129	1.11 (0.68 - 1.80)	211 / 231	1.26 (0.81 - 1.97)	
			757,550+	140 / 58	2.42 (1.43 - 4.08)	365 / 115	3.34 (2.08 - 5.36)	
	rs1229982	G/T	never drinkers	72 / 168	1.00 (ref)	45 / 112	0.91 (0.56 - 1.47)	0.52
			>0 to <134,699	131 / 270	0.79 (0.54 - 1.16)	78 / 196	0.64 (0.42 - 0.97)	
			134,699 to <757,550	180 / 211	1.21 (0.83 - 1.79)	138 / 149	1.28 (0.85 - 1.93)	
			757,550+	274 / 98	2.91 (1.91 - 4.43)	231 / 75	3.34 (2.14 - 5.21)	
ADH1C	rs2298753	T/C	never drinkers	97 / 238	1.00 (ref)	20 / 42	1.19 (0.63 - 2.27)	-0.56
			>0 to <134,699	169 / 392	0.74 (0.53 - 1.03)	40 / 74	0.96 (0.59 - 1.57)	
			134,699 to <757,550	246 / 303	1.22 (0.87 - 1.71)	72 / 57	1.84 (1.16 - 2.93)	
			757,550+	429 / 144	3.34 (2.32 - 4.81)	76 / 29	2.97 (1.71 - 5.16)	
	rs1614972	C/T	never drinkers	56 / 123	1.00 (ref)	61 / 157	0.84 (0.53 - 1.34)	0.02
			>0 to <134,699	81 / 219	0.52 (0.33 - 0.82)	128 / 247	0.84 (0.55 - 1.28)	
			134,699 to <757,550	155 / 152	1.33 (0.86 - 2.06)	163 / 208	1.05 (0.68 - 1.61)	
			757,550+	216 / 75	2.99 (1.86 - 4.79)	289 / 98	2.85 (1.80 - 4.50)	
	rs1391088	C/A	never drinkers	100 / 235	1.00 (ref)	17 / 45	0.95 (0.50 - 1.82)	0.12
			>0 to <134,699	181 / 392	0.76 (0.55 - 1.06)	28 / 74	0.66 (0.39 - 1.12)	
			134,699 to <757,550	270 / 315	1.24 (0.89 - 1.72)	48 / 45	1.59 (0.95 - 2.67)	
			757,550+	427 / 145	3.18 (2.21 - 4.57)	78 / 28	3.24 (1.87 - 5.62)	
	rs1693482	C/T	never drinkers	44 / 123	1.00 (ref)	73 / 157	1.24 (0.77 - 2.01)	-0.24
	previously studied		>0 to <134,699	84 / 199	0.87 (0.54 - 1.38)	124 / 265	0.85 (0.55 - 1.33)	
			134,699 to <757,550	135 / 164	1.37 (0.86 - 2.18)	182 / 196	1.54 (0.98 - 2.41)	
			757,550+	234 / 77	3.64 (2.22 - 5.96)	271 / 96	3.65 (2.26 - 5.88)	
	rs1631460	C/G	never drinkers	43 / 119	1.00 (ref)	74 / 161	1.23 (0.76 - 1.99)	-0.17
			>0 to <134,699	81 / 195	0.86 (0.53 - 1.37)	128 / 271	0.86 (0.55 - 1.34)	
			134,699 to <757,550	134 / 162	1.38 (0.87 - 2.22)	184 / 198	1.53 (0.97 - 2.41)	
			757,550+	231 / 76	3.61 (2.19 - 5.93)	273 / 97	3.66 (2.26 - 5.93)	
	rs11936869	C/G	never drinkers	61 / 135	1.00 (ref)	56 / 145	0.90 (0.56 - 1.43)	-0.17
			>0 to <134,699	102 / 237	0.64 (0.42 - 0.98)	107 / 228	0.79 (0.52 - 1.20)	
			134,699 to <757,550	170 / 174	1.37 (0.90 - 2.08)	148 / 186	1.08 (0.71 - 1.65)	
			757,550+	242 / 78	3.17 (2.00 - 5.02)	263 / 95	2.90 (1.85 - 4.54)	
ADH4	rs29001227	A/T	never drinkers	113 / 271	1.00 (ref)	4 / 9	1.04 (0.27 - 4.02)	4.55
			>0 to <134,699	199 / 445	0.74 (0.55 - 1.01)	10 / 21	0.98 (0.41 - 2.31)	
			134,699 to <757,550	294 / 334	1.28 (0.94 - 1.76)	24 / 26	1.47 (0.74 - 2.93)	
			757,550+	445 / 163	3.02 (2.13 - 4.27)	60 / 10	7.61 (3.46 - 16.75)	
	rs1126672	C/T	never drinkers	58 / 152	1.00 (ref)	59 / 128	1.19 (0.74 - 1.92)	-0.14
			>0 to <134,699	120 / 263	0.84 (0.56 - 1.27)	89 / 203	0.79 (0.51 - 1.22)	
			134,699 to <757,550	186 / 214	1.37 (0.91 - 2.06)	132 / 146	1.46 (0.95 - 2.24)	
			757,550+	310 / 107	3.47 (2.24 - 5.37)	195 / 66	3.53 (2.21 - 5.63)	
	rs4699710	T/C	never drinkers	49 / 142	1.00 (ref)	68 / 138	1.45 (0.90 - 2.32)	0.09
			>0 to <134,699	108 / 239	0.93 (0.61 - 1.44)	101 / 227	0.91 (0.58 - 1.41)	
			134,699 to <757,550	172 / 183	1.70 (1.10 - 2.63)	146 / 177	1.45 (0.93 - 2.24)	
			757,550+	272 / 100	3.69 (2.33 - 5.85)	233 / 73	4.23 (2.63 - 6.80)	
	rs10017466	T/C	never drinkers	45 / 134	1.00 (ref)	72 / 146	1.46 (0.91 - 2.35)	1.08
			>0 to <134,699	100 / 224	0.93 (0.60 - 1.46)	109 / 242	0.93 (0.60 - 1.45)	
			134,699 to <757,550	154 / 163	1.73 (1.11 - 2.72)	164 / 197	1.49 (0.96 - 2.31)	

APPENDIX E. Odds ratios for developing SCCHN for SNP genotypes, at varying levels of lifetime alcohol consumption

				Odds ratios in comparison to common referent group[a]				Assessment of additive interaction between SNP and lifetime alcohol consumption
				homozygous major allele		at least one minor allele		
Gene	SNP	major/ minor alleles	lifetime ethanol consumption (ml)	n (cases/ controls)[b]	Adjusted odds ratio (95% CI)	n (cases/ controls)[b]	Adjusted odds ratio (95% CI)	ICR[c] (**bold** if statistically significant after Bonferroni correction)
			757,550+	223 / 93	3.26 (2.03 - 5.24)	281 / 80	4.80 (2.98 - 7.72)	
	rs1800759	C/A	never drinkers	35 / 93	1.00 (ref)	82 / 187	1.17 (0.70 - 1.95)	0.66
			>0 to <134,699	70 / 153	0.80 (0.47 - 1.36)	139 / 313	0.86 (0.53 - 1.39)	
			134,699 to <757,550	101 / 104	1.61 (0.95 - 2.73)	217 / 256	1.37 (0.85 - 2.22)	
			757,550+	132 / 55	3.02 (1.72 - 5.31)	373 / 118	3.85 (2.32 - 6.41)	
	rs1800761	G/A	never drinkers	68 / 167	1.00 (ref)	49 / 113	1.13 (0.70 - 1.81)	-0.63
			>0 to <134,699	121 / 300	0.70 (0.48 - 1.04)	88 / 166	0.95 (0.63 - 1.44)	
			134,699 to <757,550	208 / 235	1.39 (0.95 - 2.04)	110 / 125	1.29 (0.84 - 1.98)	
			757,550+	310 / 101	3.57 (2.35 - 5.45)	195 / 72	3.07 (1.95 - 4.83)	
	rs3762894	T/C	never drinkers	82 / 178	1.00 (ref)	35 / 102	0.75 (0.45 - 1.23)	-0.74
			>0 to <134,699	135 / 328	0.62 (0.43 - 0.89)	74 / 138	0.83 (0.55 - 1.27)	
			134,699 to <757,550	226 / 263	1.15 (0.80 - 1.65)	92 / 97	1.22 (0.79 - 1.88)	
			757,550+	348 / 108	3.28 (2.19 - 4.89)	157 / 65	2.29 (1.45 - 3.60)	
ADH7	rs284787	C/T	never drinkers	67 / 166	1.00 (ref)	50 / 114	1.05 (0.65 - 1.68)	0.45
			>0 to <134,699	127 / 299	0.74 (0.50 - 1.08)	82 / 167	0.83 (0.54 - 1.27)	
			134,699 to <757,550	195 / 210	1.36 (0.92 - 2.01)	123 / 150	1.26 (0.83 - 1.92)	
			757,550+	308 / 112	3.11 (2.05 - 4.72)	197 / 61	3.61 (2.26 - 5.74)	
	rs894369	C/G	never drinkers	84 / 181	1.00 (ref)	33 / 99	0.71 (0.42 - 1.17)	0.80
			>0 to <134,699	140 / 303	0.70 (0.48 - 1.00)	69 / 162	0.63 (0.42 - 0.96)	
			134,699 to <757,550	207 / 214	1.26 (0.87 - 1.83)	111 / 146	1.00 (0.67 - 1.50)	
			757,550+	350 / 125	2.75 (1.85 - 4.07)	155 / 48	3.25 (2.02 - 5.22)	
	rs17588403	T/A	never drinkers	71 / 190	1.00 (ref)	46 / 90	1.35 (0.83 - 2.19)	-1.33
			>0 to <134,699	142 / 307	0.87 (0.60 - 1.25)	67 / 159	0.78 (0.51 - 1.21)	
			134,699 to <757,550	226 / 251	1.49 (1.03 - 2.16)	92 / 109	1.32 (0.85 - 2.05)	
			757,550+	347 / 108	3.96 (2.63 - 5.96)	158 / 65	2.97 (1.88 - 4.69)	
	rs1154454	T/C	never drinkers	76 / 175	1.00 (ref)	41 / 105	0.85 (0.52 - 1.39)	0.13
			>0 to <134,699	143 / 276	0.82 (0.57 - 1.19)	66 / 190	0.55 (0.36 - 0.84)	
			134,699 to <757,550	191 / 227	1.16 (0.80 - 1.70)	127 / 133	1.31 (0.86 - 1.98)	
			757,550+	306 / 106	3.04 (2.02 - 4.58)	199 / 67	3.02 (1.91 - 4.77)	
	rs1154456	T/C	never drinkers	54 / 125	1.00 (ref)	63 / 155	0.77 (0.48 - 1.22)	-0.03
			>0 to <134,699	91 / 218	0.62 (0.40 - 0.96)	118 / 248	0.68 (0.44 - 1.04)	
			134,699 to <757,550	149 / 181	1.03 (0.67 - 1.59)	169 / 179	1.20 (0.78 - 1.85)	
			757,550+	256 / 81	2.91 (1.82 - 4.65)	249 / 92	2.65 (1.67 - 4.20)	
	rs1154460	G/A	never drinkers	35 / 65	1.00 (ref)	81 / 215	0.63 (0.37 - 1.07)	0.92
			>0 to <134,699	57 / 131	0.53 (0.30 - 0.92)	152 / 334	0.55 (0.34 - 0.90)	
			134,699 to <757,550	88 / 111	0.83 (0.48 - 1.45)	229 / 249	0.97 (0.59 - 1.60)	
			757,550+	133 / 54	1.93 (1.07 - 3.46)	371 / 119	2.48 (1.47 - 4.20)	
	rs971074	G/A	never drinkers	91 / 219	1.00 (ref)	26 / 61	1.15 (0.65 - 2.03)	0.86
			>0 to <134,699	166 / 357	0.79 (0.56 - 1.11)	43 / 109	0.73 (0.46 - 1.17)	
			134,699 to <757,550	247 / 275	1.34 (0.95 - 1.89)	71 / 85	1.31 (0.84 - 2.05)	
			757,550+	380 / 140	3.13 (2.15 - 4.55)	125 / 33	4.14 (2.47 - 6.94)	
	rs1573496	C/G	never drinkers	98 / 236	1.00 (ref)	19 / 44	1.08 (0.57 - 2.05)	0.46
	previously studied		>0 to <134,699	175 / 386	0.76 (0.55 - 1.06)	34 / 80	0.76 (0.46 - 1.27)	
			134,699 to <757,550	272 / 301	1.31 (0.94 - 1.83)	46 / 59	1.29 (0.78 - 2.13)	
			757,550+	446 / 155	3.20 (2.23 - 4.59)	58 / 18	3.73 (1.98 - 7.05)	
ALDH2	rs4767939	A/G	never drinkers	76 / 174	1.00 (ref)	41 / 106	0.88 (0.54 - 1.44)	1.09
			>0 to <134,699	120 / 289	0.65 (0.44 - 0.94)	89 / 177	0.84 (0.56 - 1.27)	
			134,699 to <757,550	194 / 218	1.21 (0.83 - 1.77)	124 / 142	1.28 (0.85 - 1.93)	
			757,550+	261 / 103	2.69 (1.78 - 4.07)	243 / 70	3.66 (2.35 - 5.71)	
	rs2238151	T/C	never drinkers	50 / 95	1.00 (ref)	67 / 184	0.70 (0.43 - 1.14)	**1.91**
			>0 to <134,699	75 / 173	0.54 (0.34 - 0.88)	133 / 293	0.64 (0.41 - 0.99)	
			134,699 to <757,550	97 / 138	0.81 (0.50 - 1.31)	220 / 222	1.19 (0.76 - 1.86)	
			757,550+	122 / 65	1.66 (0.99 - 2.80)	381 / 106	3.27 (2.02 - 5.28)	
	rs7312055	G/A	never drinkers	111 / 254	1.00 (ref)	6 / 26	0.41 (0.15 - 1.12)	0.97
			>0 to <134,699	201 / 427	0.74 (0.54 - 1.01)	8 / 39	0.32 (0.13 - 0.78)	
			134,699 to <757,550	282 / 328	1.19 (0.87 - 1.64)	36 / 32	1.30 (0.69 - 2.46)	
			757,550+	422 / 152	2.93 (2.06 - 4.17)	83 / 21	3.31 (1.76 - 6.23)	
	rs2158029	G/A	never drinkers	116 / 270	1.00 (ref)	1 / 10	0.28 (0.03 - 2.30)	2.73
			>0 to <134,699	207 / 443	0.77 (0.57 - 1.04)	2 / 23	0.13 (0.03 - 0.60)	
			134,699 to <757,550	302 / 349	1.25 (0.91 - 1.70)	16 / 11	1.79 (0.72 - 4.45)	
			757,550+	470 / 166	3.09 (2.19 - 4.35)	35 / 7	5.10 (2.04 - 12.73)	
	rs16941667	C/T	never drinkers	98 / 233	1.00 (ref)	19 / 47	1.06 (0.57 - 1.99)	0.99
			>0 to <134,699	163 / 390	0.70 (0.50 - 0.98)	46 / 75	1.11 (0.69 - 1.79)	
			134,699 to <757,550	259 / 294	1.29 (0.92 - 1.80)	59 / 66	1.42 (0.89 - 2.29)	
			757,550+	405 / 148	3.12 (2.17 - 4.49)	100 / 25	4.17 (2.39 - 7.28)	
	rs16941669	T/G	never drinkers	97 / 218	1.00 (ref)	20 / 61	0.74 (0.41 - 1.34)	1.45
			>0 to <134,699	162 / 380	0.67 (0.48 - 0.94)	47 / 86	0.88 (0.55 - 1.39)	
			134,699 to <757,550	254 / 302	1.17 (0.84 - 1.64)	64 / 58	1.46 (0.90 - 2.35)	
			757,550+	415 / 149	2.90 (2.01 - 4.18)	90 / 23	4.09 (2.30 - 7.26)	
CYP2E1	rs3813865	G/C	never drinkers	108 / 259	1.00 (ref)	9 / 20	1.24 (0.51 - 3.00)	1.95
			>0 to <134,699	196 / 432	0.76 (0.56 - 1.04)	13 / 34	0.75 (0.37 - 1.54)	
			134,699 to <757,550	289 / 324	1.32 (0.96 - 1.82)	29 / 36	1.27 (0.69 - 2.31)	
			757,550+	437 / 160	3.11 (2.19 - 4.42)	68 / 13	5.31 (2.65 - 10.65)	

APPENDIX E. Odds ratios for developing SCCHN for SNP genotypes, at varying levels of lifetime alcohol consumption

				Odds ratios in comparison to common referent group[a]				Assessment of additive interaction between SNP and lifetime alcohol consumption
				homozygous major allele		at least one minor allele		
Gene	SNP	major/ minor alleles	lifetime ethanol consumption (ml)	n (cases/ controls)[b]	Adjusted odds ratio (95% CI)	n (cases/ controls)[b]	Adjusted odds ratio (95% CI)	ICR[c] (**bold** if statistically significant after Bonferroni correction)
	rs3813867	G/C	never drinkers	110 / 263	1.00 (ref)	7 / 16	0.94 (0.35 - 2.47)	3.88
			>0 to <134,699	197 / 441	0.74 (0.54 - 1.01)	12 / 25	0.80 (0.37 - 1.76)	
			134,699 to <757,550	301 / 329	1.35 (0.98 - 1.85)	17 / 31	0.65 (0.33 - 1.28)	
			757,550+	460 / 165	3.00 (2.11 - 4.25)	44 / 8	6.82 (2.90 - 16.02)	
	rs8192772	T/C	never drinkers	99 / 237	1.00 (ref)	18 / 43	1.07 (0.55 - 2.05)	1.89
			>0 to <134,699	179 / 394	0.75 (0.54 - 1.03)	30 / 72	0.84 (0.50 - 1.41)	
			134,699 to <757,550	264 / 294	1.30 (0.94 - 1.81)	53 / 66	1.29 (0.80 - 2.08)	
			757,550+	413 / 152	3.02 (2.11 - 4.33)	92 / 21	4.97 (2.78 - 8.91)	
	rs915908	G/A	never drinkers	93 / 216	1.00 (ref)	24 / 64	1.04 (0.59 - 1.85)	-1.29
			>0 to <134,699	152 / 352	0.74 (0.53 - 1.04)	57 / 114	0.80 (0.52 - 1.24)	
			134,699 to <757,550	254 / 277	1.35 (0.96 - 1.91)	64 / 83	1.12 (0.71 - 1.77)	
			757,550+	417 / 131	3.56 (2.45 - 5.17)	88 / 42	2.31 (1.39 - 3.82)	
	rs915909	C/T	never drinkers	115 / 270	1.00 (ref)	2 / 9	0.62 (0.12 - 3.24)	0.80
			>0 to <134,699	206 / 449	0.76 (0.56 - 1.03)	3 / 15	0.35 (0.09 - 1.30)	
			134,699 to <757,550	302 / 348	1.26 (0.92 - 1.71)	16 / 12	1.74 (0.71 - 4.24)	
			757,550+	468 / 163	3.15 (2.23 - 4.44)	36 / 9	3.56 (1.56 - 8.17)	
	rs7092584	C/T	never drinkers	92 / 223	1.00 (ref)	25 / 57	1.17 (0.66 - 2.07)	1.06
			>0 to <134,699	167 / 375	0.75 (0.54 - 1.05)	42 / 91	0.89 (0.56 - 1.44)	
			134,699 to <757,550	252 / 279	1.36 (0.97 - 1.92)	65 / 80	1.24 (0.79 - 1.94)	
			757,550+	384 / 142	3.10 (2.14 - 4.49)	120 / 31	4.33 (2.58 - 7.26)	
	rs743535	C/T	never drinkers	95 / 223	1.00 (ref)	22 / 53	0.93 (0.51 - 1.69)	2.26
			>0 to <134,699	169 / 386	0.70 (0.50 - 0.98)	40 / 78	0.91 (0.56 - 1.49)	
			134,699 to <757,550	252 / 279	1.29 (0.92 - 1.81)	60 / 80	1.04 (0.66 - 1.65)	
			757,550+	386 / 147	2.79 (1.93 - 4.04)	116 / 26	4.98 (2.89 - 8.59)	
	rs2249695	C/T	never drinkers	73 / 136	1.00 (ref)	44 / 144	0.56 (0.35 - 0.91)	1.21
	"tipsy" SNP		>0 to <134,699	127 / 253	0.63 (0.43 - 0.94)	82 / 213	0.52 (0.34 - 0.80)	
			134,699 to <757,550	160 / 178	1.06 (0.71 - 1.59)	158 / 181	0.95 (0.63 - 1.43)	
			757,550+	218 / 92	2.12 (1.37 - 3.28)	286 / 81	2.89 (1.83 - 4.55)	
	rs28969387	A/T	never drinkers	115 / 275	1.00 (ref)	2 / 5	1.03 (0.16 - 6.58)	0.05
			>0 to <134,699	206 / 464	0.75 (0.55 - 1.01)	3 / 2	2.25 (0.33 - 15.13)	
			134,699 to <757,550	311 / 353	1.30 (0.95 - 1.77)	7 / 7	1.06 (0.33 - 3.38)	
			757,550+	493 / 170	3.22 (2.29 - 4.54)	12 / 3	3.30 (0.88 - 12.40)	
	rs11101812	T/C	never drinkers	115 / 275	1.00 (ref)	1 / 5	0.41 (0.04 - 3.90)	19.34
			>0 to <134,699	205 / 463	0.74 (0.55 - 1.00)	4 / 3	2.70 (0.52 - 14.05)	
			134,699 to <757,550	312 / 348	1.31 (0.96 - 1.79)	5 / 10	0.67 (0.20 - 2.23)	
			757,550+	489 / 172	3.10 (2.20 - 4.36)	15 / 1	21.85 (2.60 #####)	
OXIDATIVE STRESS METABOLISM GENES								
CAT	rs1049982	C/T	never drinkers	60 / 137	1.00 (ref)	57 / 140	1.10 (0.69 - 1.77)	-0.15
			>0 to <134,699	95 / 216	0.75 (0.49 - 1.15)	114 / 247	0.81 (0.54 - 1.22)	
			134,699 to <757,550	135 / 150	1.41 (0.92 - 2.16)	183 / 209	1.29 (0.85 - 1.94)	
			757,550+	200 / 71	3.33 (2.10 - 5.28)	303 / 102	3.28 (2.12 - 5.07)	
GPX1	rs8179172	T/A	never drinkers	116 / 270	1.00 (ref)	1 / 10	0.14 (0.02 - 1.25)	0.16
			>0 to <134,699	205 / 450	0.73 (0.54 - 0.99)	4 / 16	0.44 (0.13 - 1.46)	
			134,699 to <757,550	305 / 349	1.21 (0.88 - 1.65)	13 / 11	2.15 (0.85 - 5.44)	
			757,550+	475 / 160	3.10 (2.19 - 4.38)	30 / 13	2.40 (1.10 - 5.27)	
	rs1800668	C/T	never drinkers	56 / 146	1.00 (ref)	61 / 132	1.16 (0.73 - 1.85)	0.37
			>0 to <134,699	97 / 222	0.76 (0.50 - 1.16)	112 / 243	0.85 (0.56 - 1.29)	
			134,699 to <757,550	178 / 185	1.43 (0.94 - 2.16)	140 / 175	1.33 (0.87 - 2.04)	
			757,550+	272 / 95	3.23 (2.07 - 5.05)	233 / 77	3.75 (2.37 - 5.95)	
	rs3811699	A/G	never drinkers	55 / 142	1.00 (ref)	62 / 138	1.14 (0.72 - 1.82)	0.37
			>0 to <134,699	95 / 216	0.76 (0.50 - 1.17)	114 / 250	0.85 (0.56 - 1.29)	
			134,699 to <757,550	171 / 176	1.45 (0.95 - 2.21)	147 / 184	1.31 (0.86 - 2.01)	
			757,550+	258 / 91	3.21 (2.04 - 5.04)	247 / 82	3.71 (2.35 - 5.88)	
	rs3448	C/T	never drinkers	60 / 158	1.00 (ref)	57 / 122	1.27 (0.80 - 2.03)	0.32
			>0 to <134,699	116 / 256	0.84 (0.56 - 1.26)	93 / 210	0.84 (0.55 - 1.28)	
			134,699 to <757,550	157 / 192	1.35 (0.89 - 2.03)	161 / 168	1.55 (1.03 - 2.34)	
			757,550+	273 / 105	3.34 (2.17 - 5.14)	232 / 68	3.93 (2.49 - 6.22)	
GPX2	rs11623705	G/T	never drinkers	86 / 232	1.00 (ref)	31 / 48	1.58 (0.90 - 2.77)	-0.76
			>0 to <134,699	166 / 367	0.83 (0.59 - 1.17)	43 / 99	0.83 (0.52 - 1.32)	
			134,699 to <757,550	263 / 292	1.47 (1.05 - 2.08)	55 / 68	1.24 (0.77 - 2.01)	
			757,550+	421 / 142	3.60 (2.47 - 5.24)	84 / 31	3.41 (2.00 - 5.84)	
	rs2412065	G/C	never drinkers	67 / 158	1.00 (ref)	50 / 122	1.00 (0.62 - 1.60)	0.06
			>0 to <134,699	126 / 251	0.85 (0.57 - 1.26)	83 / 215	0.65 (0.43 - 0.99)	
			134,699 to <757,550	186 / 204	1.32 (0.89 - 1.96)	132 / 156	1.25 (0.82 - 1.90)	
			757,550+	278 / 99	3.18 (2.07 - 4.88)	227 / 74	3.24 (2.06 - 5.10)	
	rs2737844	C/T	never drinkers	58 / 117	1.00 (ref)	59 / 161	0.76 (0.48 - 1.22)	0.03
			>0 to <134,699	100 / 173	0.84 (0.54 - 1.30)	108 / 292	0.54 (0.35 - 0.82)	
			134,699 to <757,550	129 / 147	1.09 (0.70 - 1.70)	188 / 211	1.12 (0.73 - 1.72)	
			757,550+	180 / 62	2.84 (1.73 - 4.65)	322 / 111	2.63 (1.66 - 4.16)	
GPX4	rs757229	G/C	never drinkers	35 / 83	1.00 (ref)	82 / 197	0.92 (0.55 - 1.52)	0.09
			>0 to <134,699	58 / 114	0.74 (0.43 - 1.28)	150 / 352	0.69 (0.43 - 1.12)	
			134,699 to <757,550	67 / 109	0.90 (0.52 - 1.56)	251 / 251	1.34 (0.83 - 2.17)	

APPENDIX E. Odds ratios for developing SCCHN for SNP genotypes, at varying levels of lifetime alcohol consumption

				Odds ratios in comparison to common referent group[a]				Assessment of additive interaction between SNP and lifetime alcohol consumption
				homozygous major allele		at least one minor allele		
Gene	SNP	major/ minor alleles	lifetime ethanol consumption (ml)	n (cases/ controls)[b]	Adjusted odds ratio (95% CI)	n (cases/ controls)[b]	Adjusted odds ratio (95% CI)	ICR[c] (**bold** if statistically significant after Bonferroni correction)
			757,550+	116 / 40	3.05 (1.68 - 5.54)	389 / 132	3.05 (1.84 - 5.06)	
SOD1	rs11910115	A/C	never drinkers	116 / 271	1.00 (ref)	1 / 9	0.31 (0.03 - 2.91)	2.22
			>0 to <134,699	205 / 441	0.76 (0.56 - 1.04)	4 / 25	0.27 (0.08 - 0.92)	
			134,699 to <757,550	299 / 340	1.27 (0.93 - 1.73)	19 / 20	1.33 (0.62 - 2.85)	
			757,550+	469 / 165	3.09 (2.19 - 4.35)	36 / 8	4.63 (1.94 - 11.06)	
	rs4998557	G/A	never drinkers	85 / 202	1.00 (ref)	32 / 78	0.91 (0.54 - 1.53)	1.51
			>0 to <134,699	157 / 341	0.75 (0.53 - 1.07)	52 / 122	0.69 (0.44 - 1.09)	
			134,699 to <757,550	208 / 258	1.15 (0.80 - 1.64)	110 / 102	1.53 (1.00 - 2.34)	
			757,550+	328 / 124	2.76 (1.87 - 4.08)	176 / 49	4.18 (2.60 - 6.72)	
	rs10432782	T/G	never drinkers	86 / 208	1.00 (ref)	31 / 72	0.95 (0.56 - 1.61)	1.15
			>0 to <134,699	162 / 351	0.77 (0.55 - 1.09)	47 / 115	0.65 (0.41 - 1.04)	
			134,699 to <757,550	222 / 267	1.20 (0.84 - 1.70)	96 / 93	1.48 (0.96 - 2.27)	
			757,550+	350 / 128	2.91 (1.98 - 4.27)	154 / 44	4.00 (2.48 - 6.46)	
	rs2070424	A/G	never drinkers	95 / 238	1.00 (ref)	22 / 42	1.17 (0.63 - 2.17)	0.01
			>0 to <134,699	181 / 402	0.78 (0.56 - 1.08)	28 / 64	0.76 (0.44 - 1.30)	
			134,699 to <757,550	260 / 302	1.31 (0.94 - 1.83)	58 / 58	1.41 (0.87 - 2.30)	
			757,550+	407 139	3.27 (2.27 - 4.72)	98 / 34	3.45 (2.06 - 5.79)	
	rs1041740	C/T	never drinkers	56 / 156	1.00 (ref)	61 / 124	1.43 (0.89 - 2.28)	-1.09
			>0 to <134,699	104 / 239	0.89 (0.59 - 1.35)	105 / 227	0.90 (0.59 - 1.37)	
			134,699 to <757,550	193 / 191	1.71 (1.13 - 2.59)	125 / 169	1.33 (0.86 - 2.05)	
			757,550+	291 / 94	4.15 (2.65 - 6.49)	214 / 79	3.49 (2.21 - 5.52)	
SOD2	rs4342445	G/A	never drinkers	55 / 177	1.00 (ref)	62 / 103	2.14 (1.33 - 3.44)	0.27
			>0 to <134,699	135 / 296	1.10 (0.74 - 1.63)	74 / 170	0.96 (0.62 - 1.49)	
			134,699 to <757,550	184 / 227	1.61 (1.07 - 2.40)	134 / 133	2.11 (1.38 - 3.24)	
			757,550+	331 / 123	4.04 (2.65 - 6.17)	174 / 50	5.45 (3.33 - 8.90)	
	rs2842980	A/T	never drinkers	74 / 169	1.00 (ref)	43 / 111	0.87 (0.54 - 1.41)	0.89
			>0 to <134,699	121 / 282	0.69 (0.47 - 1.01)	88 / 184	0.76 (0.50 - 1.14)	
			134,699 to <757,550	196 / 222	1.27 (0.87 - 1.87)	122 / 138	1.15 (0.76 - 1.74)	
			757,550+	291 / 108	2.77 (1.84 - 4.19)	214 / 65	3.54 (2.25 - 5.56)	
	rs8031	T/A	never drinkers	38 / 75	1.00 (ref)	79 / 205	0.84 (0.51 - 1.40)	-0.97
			>0 to <134,699	59 / 139	0.58 (0.34 - 0.98)	150 / 327	0.71 (0.44 - 1.15)	
			134,699 to <757,550	101 / 113	1.14 (0.68 - 1.93)	217 / 247	1.15 (0.71 - 1.86)	
			757,550+	155 / 40	3.73 (2.08 - 6.71)	350 / 133	2.61 (1.59 - 4.30)	
	rs5746134	C/T	never drinkers	113 / 264	1.00 (ref)	4 / 16	0.55 (0.16 - 1.87)	4.29
			>0 to <134,699	201 / 447	0.73 (0.54 - 0.99)	8 / 19	0.80 (0.31 - 2.08)	
			134,699 to <757,550	299 / 342	1.25 (0.91 - 1.71)	19 / 18	1.47 (0.68 - 3.18)	
			757,550+	446 / 163	2.96 (2.09 - 4.19)	59 / 10	6.80 (3.03 - 15.26)	
	rs2758331	C/A	never drinkers	38 / 82	1.00 (ref)	79 / 198	0.92 (0.55 - 1.52)	-1.51
			>0 to <134,699	60 / 151	0.58 (0.34 - 0.98)	149 / 315	0.78 (0.48 - 1.24)	
			134,699 to <757,550	108 / 127	1.14 (0.68 - 1.90)	210 / 233	1.26 (0.78 - 2.03)	
			757,550+	179 / 44	4.23 (2.39 - 7.50)	326 / 129	2.64 (1.61 - 4.34)	

[a] Conditional logistic regression models conditioned on sex, race, and age category (age in years: 20-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-80), and adjusted for continous duration of smoking in years (rounded to nearest year)

[b] Cases and controls do not sum to 1227 and 1325, respectively, because (1) 4 cases and 3 controls are missing information on duration of cigarette smoking, (2) 77 cases and 45 controls are missing information on lifetime alcohol consumption, and (3) a few subjects are missing genotype information for some SNPs

[c] ICRs were calculated using cancer odds ratios of subjects in three categories: (1) the highest drinking category and no minor allele (i.e., those singly exposed to drinking only -- OR_{01}); (2) never-drinkers with at least one minor allele (i.e., those singly exposed to only the variant allele -- OR_{10}); and (3) subjects in the highest drinking category and at least one minor allele (i.e., those doubly exposed to both alcohol and the variant allele -- OR_{11}), compared to never-drinkers homozygous for the major allele (i.e., the referent group that was not exposed to either the variant allele or to drinking -- OR_{00} = 1.0). ICRs statistically different from 0 indicate departure from additive interaction. p-values for each ICR (not shown) were Bonferroni-adjusted for 64 statistical tests.

REFERENCES

1. Neville BW, Day TA. Oral Cancer and Precancerous Lesions. CA Cancer J Clin. 2002 July 1, 2002;52(4):195-215.

2. Nair U, Bartsch H, Nair J. Alert for an epidemic of oral cancer due to use of the betel quid substitutes gutkha and pan masala: a review of agents and causative mechanisms. Mutagenesis. 2004 Jul;19(4):251-62.

3. Olshan AF. Cancer of the larynx. In: Schottenfeld D, Fraumeni JF, Jr., editors. Cancer Epidemiology and Prevention. Third ed. New York City: Oxford University Press; 2006. p. 627-37.

4. Talamini R, Bosetti C, La Vecchia C, Dal Maso L, Levi F, Bidoli E, et al. Combined effect of tobacco and alcohol on laryngeal cancer risk: a case-control study. Cancer Causes Control. 2002 Dec;13(10):957-64.

5. Pai SI, Westra WH. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. Annu Rev Pathol. 2009;4:49-70.

6. Lippman SM, Shin DM, Lee JJ, Batsakis JG, Lotan R, Tainsky MA, et al. p53 and retinoid chemoprevention of oral carcinogenesis. Cancer Res. 1995 Jan 1;55(1):16-9.

7. Boyle JO, Hakim J, Koch W, van der Riet P, Hruban RH, Roa RA, et al. The incidence of p53 mutations increases with progression of head and neck cancer. Cancer Res. 1993 Oct 1;53(19):4477-80.

8. Brennan JA, Boyle JO, Koch WM, Goodman SN, Hruban RH, Eby YJ, et al. Association between cigarette smoking and mutation of the p53 gene in squamous-cell carcinoma of the head and neck. N Engl J Med. 1995 Mar 16;332(11):712-7.

9. Jones AS, Morar P, Phillips DE, Field JK, Husband D, Helliwell TR. Second primary tumors in patients with head and neck squamous cell carcinoma. Cancer. 1995 Mar 15;75(6):1343-53.

10. Schwartz LH, Ozsahin M, Zhang GN, Touboul E, De Vataire F, Andolenko P, et al. Synchronous and metachronous head and neck carcinomas. Cancer. 1994 Oct 1;74(7):1933-8.

11. Chuang SC, Scelo G, Tonita JM, Tamaro S, Jonasson JG, Kliewer EV, et al. Risk of second primary cancer among patients with head and neck cancers: A pooled analysis of 13 cancer registries. Int J Cancer. 2008 Nov 15;123(10):2390-6.

12. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer. 1953 Sep;6(5):963-8.

13. van Oijen MG, Slootweg PJ. Oral field cancerization: carcinogen-induced independent events or micrometastatic deposits? Cancer Epidemiol Biomarkers Prev. 2000 Mar;9(3):249-56.

14. American Cancer Society. Cancer Facts & Figures 2011. Atlanta: American Cancer Society, Inc.; 2011.

15. Ries LAG, Melbert D, Krapcho M, Stinchcomb DG, Howlader N, Horner MJ, et al. SEER Cancer Statistics Review, 1975-2005. Bethesda, MD: National Cancer Institute, 2008.

16. Chin D, Boyle GM, Porceddu S, Theile DR, Parsons PG, Coman WB. Head and neck cancer: past, present and future. Expert Rev Anticancer Ther. 2006 Jul;6(7):1111-8.

17. Gupta S, Kong W, Peng Y, Miao Q, Mackillop WJ. Temporal trends in the incidence and survival of cancers of the upper aerodigestive tract in Ontario and the United States. Int J Cancer. 2009 Apr 23;125(9):2159-65.

18. Curado MP, Hashibe M. Recent changes in the epidemiology of head and neck cancer. Curr Opin Oncol. 2009 May;21(3):194-200.

19. Shiboski CH, Schmidt BL, Jordan RC. Tongue and tonsil carcinoma: increasing trends in the U.S. population ages 20-44 years. Cancer. 2005 May 1;103(9):1843-9.

20. Hammarstedt L, Dahlstrand H, Lindquist D, Onelov L, Ryott M, Luo J, et al. The incidence of tonsillar cancer in Sweden is increasing. Acta Otolaryngol. 2007 Sep;127(9):988-92.

21. Adelstein DJ, Ridge JA, Gillison ML, Chaturvedi AK, D'Souza G, Gravitt PE, et al. Head and neck squamous cell cancer and the human papillomavirus: Summary of a National Cancer Institute State of the Science Meeting, November 9-10, 2008, Washington, D.C. Head Neck. 2009 Sep 29;31(11):1393-422.

22. Foulkes WD, Brunet JS, Kowalski LP, Narod SA, Franco EL. Family history of cancer is a risk factor for squamous cell carcinoma of the head and neck in Brazil: a case-control study. Int J Cancer. 1995 Dec 11;63(6):769-73.

23. Foulkes WD, Brunet JS, Sieh W, Black MJ, Shenouda G, Narod SA. Familial risks of squamous cell carcinoma of the head and neck: retrospective case-control study. BMJ. 1996 Sep 21;313(7059):716-21.

24. Brown LM, Gridley G, Diehl SR, Winn DM, Harty LC, Otero EB, et al. Family cancer history and susceptibility to oral carcinoma in Puerto Rico. Cancer. 2001 Oct 15;92(8):2102-8.

25. Ankathil R, Mathew A, Joseph F, Nair MK. Is oral cancer susceptibility inherited? Report of five oral cancer families. Eur J Cancer B Oral Oncol. 1996 Jan;32B(1):63-7.

26. Morita M, Kuwano H, Nakashima T, Taketomi A, Baba H, Saito T, et al. Family aggregation of carcinoma of the hypopharynx and cervical esophagus: special reference to multiplicity of cancer in upper aerodigestive tract. Int J Cancer. 1998 May 18;76(4):468-71.

27. Goldstein AM, Blot WJ, Greenberg RS, Schoenberg JB, Austin DF, Preston-Martin S, et al. Familial risk in oral and pharyngeal cancer. Eur J Cancer B Oral Oncol. 1994 Sep;30B(5):319-22.

28. Negri E, Boffetta P, Berthiller J, Castellsague X, Curado MP, Dal Maso L, et al. Family history of cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. Int J Cancer. 2009 Jan 15;124(2):394-401.

29. Hussain SK, Sundquist J, Hemminki K. Familial clustering of cancer at human papillomavirus-associated sites according to the Swedish Family-Cancer Database. Int J Cancer. 2008 Apr 15;122(8):1873-8.

30. Tobacco smoking. IARC Monogr Eval Carcinog Risks Hum. 1986;38.

31. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Tobacco smoke and involuntary smoking. IARC Monogr Eval Carcinog Risks Hum. 2004;83:1-1438.

32. Vineis P, Alavanja M, Buffler P, Fontham E, Franceschi S, Gao YT, et al. Tobacco and cancer: recent epidemiological evidence. J Natl Cancer Inst. 2004 Jan 21;96(2):99-106.

33. Lubin JH, Purdue M, Kelsey K, Zhang ZF, Winn D, Wei Q, et al. Total exposure and exposure rate effects for alcohol and smoking and risk of head and neck cancer: a pooled analysis of case-control studies. Am J Epidemiol. 2009 Oct 15;170(8):937-47.

34. Rahman M, Sakamoto J, Fukui T. Bidi smoking and oral cancer: a meta-analysis. Int J Cancer. 2003 Sep 10;106(4):600-4.

35. Sapkota A, Gajalakshmi V, Jetly DH, Roychowdhury S, Dikshit RP, Brennan P, et al. Smokeless tobacco and increased risk of hypopharyngeal and laryngeal cancers: a multicentric case-control study from India. Int J Cancer. 2007 Oct 15;121(8):1793-8.

36. Brennan P, Boffetta P. Mechanistic considerations in the molecular epidemiology of head and neck cancer. IARC Sci Publ. 2004(157):393-414.

37. Barra S, Baron AE, Franceschi S, Talamini R, La Vecchia C. Cancer and non-cancer controls in studies on the effect of tobacco and alcohol consumption. Int J Epidemiol. 1991 Dec;20(4):845-51.

38. Franceschi S, Talamini R, Barra S, Baron AE, Negri E, Bidoli E, et al. Smoking and drinking in relation to cancers of the oral cavity, pharynx, larynx, and esophagus in northern Italy. Cancer Res. 1990 Oct 15;50(20):6502-7.

39. Llewellyn CD, Johnson NW, Warnakulasuriya KA. Risk factors for oral cancer in newly diagnosed patients aged 45 years and younger: a case-control study in Southern England. J Oral Pathol Med. 2004 Oct;33(9):525-32.

40. Merletti F, Boffetta P, Ciccone G, Mashberg A, Terracini B. Role of tobacco and alcoholic beverages in the etiology of cancer of the oral cavity/oropharynx in Torino, Italy. Cancer Res. 1989 Sep 1;49(17):4919-24.

41. Garrote LF, Herrero R, Reyes RM, Vaccarella S, Anta JL, Ferbeye L, et al. Risk factors for cancer of the oral cavity and oro-pharynx in Cuba. Br J Cancer. 2001 Jul 6;85(1):46-54.

42. Menvielle G, Luce D, Goldberg P, Bugel I, Leclerc A. Smoking, alcohol drinking and cancer risk for various sites of the larynx and hypopharynx. A case-control study in France. Eur J Cancer Prev. 2004 Jun;13(3):165-72.

43. Balaram P, Sridhar H, Rajkumar T, Vaccarella S, Herrero R, Nandakumar A, et al. Oral cancer in southern India: the influence of smoking, drinking, paan-chewing and oral hygiene. Int J Cancer. 2002 Mar 20;98(3):440-5.

44. Marron M, Boffetta P, Zhang ZF, Zaridze D, Wunsch-Filho V, Winn DM, et al. Cessation of alcohol drinking, tobacco smoking and the reversal of head and neck cancer risk. Int J Epidemiol. 2010 Feb;39(1):182-96.

45. Lee YC, Boffetta P, Sturgis EM, Wei Q, Zhang ZF, Muscat J, et al. Involuntary smoking and head and neck cancer risk: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. Cancer Epidemiol Biomarkers Prev. 2008 Aug;17(8):1974-81.

46. Lee YC, Marron M, Benhamou S, Bouchardy C, Ahrens W, Pohlabeln H, et al. Active and involuntary tobacco smoking and upper aerodigestive tract cancer risks in a multicenter case-control study. Cancer Epidemiol Biomarkers Prev. 2009 Dec;18(12):3353-61.

47. Tomar SL. Trends and patterns of tobacco use in the United States. Am J Med Sci. 2003 Oct;326(4):248-54.

48. Boffetta P, Hecht S, Gray N, Gupta P, Straif K. Smokeless tobacco and cancer. Lancet Oncol. 2008 Jul;9(7):667-75.

49. Hashibe M, Brennan P, Chuang SC, Boccia S, Castellsague X, Chen C, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. Cancer Epidemiol Biomarkers Prev. 2009 Feb;18(2):541-50.

50. Alcohol drinking. IARC Working Group, Lyon, 13-20 October 1987. IARC Monogr Eval Carcinog Risks Hum. 1988;44:1-378.

51. IARC Working Group on the Evaluation of Carcinogenic Risk of Chemicals to Humans. Alcohol Consumption and Ethyl Carbamate. Lyon, France: International Agency for Research on Cancer: Distributed by the Secretariat of the World Health Organization, 2010.

52. Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, Curado MP, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. J Natl Cancer Inst. 2007 May 16;99(10):777-89.

53. Ng SK, Kabat GC, Wynder EL. Oral cavity cancer in non-users of tobacco. J Natl Cancer Inst. 1993 May 5;85(9):743-5.

54. Schlecht NF, Franco EL, Pintos J, Negassa A, Kowalski LP, Oliveira BV, et al. Interaction between tobacco and alcohol consumption and the risk of cancers of the upper aero-digestive tract in Brazil. Am J Epidemiol. 1999 Dec 1;150(11):1129-37.

55. Consumption of alcoholic beverages and ethyl carbamate (urethane). IARC Working Group, Lyon, 6-13 February 2007. IARC Monogr Eval Carcinog Risks Hum. 2008;96 in press.

56. Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, et al. Smoking and drinking in relation to oral and pharyngeal cancer. Cancer Res. 1988 Jun 1;48(11):3282-7.

57. Lissowska J, Pilarska A, Pilarski P, Samolczyk-Wanyura D, Piekarczyk J, Bardin-Mikollajczak A, et al. Smoking, alcohol, diet, dentition and sexual practices in the epidemiology of oral cancer in Poland. Eur J Cancer Prev. 2003 Feb;12(1):25-33.

58. Purdue MP, Hashibe M, Berthiller J, La Vecchia C, Dal Maso L, Herrero R, et al. Type of alcoholic beverage and risk of head and neck cancer--a pooled analysis within the INHANCE Consortium. Am J Epidemiol. 2009 Jan 15;169(2):132-42.

59. Kabat GC, Chang CJ, Wynder EL. The role of tobacco, alcohol use, and body mass index in oral and pharyngeal cancer. Int J Epidemiol. 1994 Dec;23(6):1137-44.

60. Seitz HK, Stickel F. Molecular mechanisms of alcohol-mediated carcinogenesis. Nat Rev Cancer. 2007 Aug;7(8):599-612.

61. Homann N, Jousimies-Somer H, Jokelainen K, Heine R, Salaspuro M. High acetaldehyde levels in saliva after ethanol consumption: methodological aspects and pathogenetic implications. Carcinogenesis. 1997 Sep;18(9):1739-43.

62. Homann N, Karkkainen P, Koivisto T, Nosova T, Jokelainen K, Salaspuro M. Effects of acetaldehyde on cell regeneration and differentiation of the upper gastrointestinal tract mucosa. J Natl Cancer Inst. 1997 Nov 19;89(22):1692-7.

63. Brennan P, Lewis S, Hashibe M, Bell DA, Boffetta P, Bouchardy C, et al. Pooled analysis of alcohol dehydrogenase genotypes and head and neck cancer: a HuGE review. Am J Epidemiol. 2004 Jan 1;159(1):1-16.

64. Poschl G, Seitz HK. Alcohol and cancer. Alcohol Alcohol. 2004 May-Jun;39(3):155-65.

65. Zakhari S. Alcohol metabolism, tobacco and cancer. In: Cho CH, Purohit V, editors. Alcohol, Tobacco and Cancer. Basel, Switzerland: S. Karger AG; 2006. p. 29-47.

66. Brooks PJ, Theruvathu JA. DNA adducts from acetaldehyde: implications for alcohol-related carcinogenesis. Alcohol. 2005 Apr;35(3):187-93.

67. Poschl G, Stickel F, Wang XD, Seitz HK. Alcohol and cancer: genetic and nutritional aspects. Proc Nutr Soc. 2004 Feb;63(1):65-71.

68. Zakhari S. Overview: how is alcohol metabolized by the body? Alcohol Res Health. 2006;29(4):245-54.

69. Powell CL, Bradford BU, Craig CP, Tsuchiya M, Uehara T, O'Connell TM, et al. Mechanism for Prevention of Alcohol-Induced Liver Injury by Dietary Methyl Donors. Toxicol Sci. 2010 (in press) Jan 29.

70. Secretan B, Straif K, Baan R, Grosse Y, El Ghissassi F, Bouvard V, et al. A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol. 2009 Nov;10(11):1033-4.

71. Lachenmeier DW, Monakhova YB. Short-term salivary acetaldehyde increase due to direct exposure to alcoholic beverages as an additional cancer risk factor beyond ethanol metabolism. Journal of experimental & clinical cancer research : CR. [Research Support, Non-U.S. Gov't]. 2011;30:3.

72. Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide. Proceedings of the IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, France, 17-24 February 1998. IARC Monogr Eval Carcinog Risks Hum. 1999;71 Pt 1:1-315.

73. Brooks PJ. DNA damage, DNA repair, and alcohol toxicity--a review. Alcohol Clin Exp Res. 1997 Sep;21(6):1073-82.

74. Simanowski UA, Suter P, Stickel F, Maier H, Waldherr R, Smith D, et al. Esophageal epithelial hyperproliferation following long-term alcohol consumption in rats: effects of age and salivary gland function. J Natl Cancer Inst. 1993 Dec 15;85(24):2030-3.

75. Salaspuro MP. Acetaldehyde, microbes, and cancer of the digestive tract. Crit Rev Clin Lab Sci. 2003 Apr;40(2):183-208.

76. Seitz HK, Stickel F, Homann N. Pathogenetic mechanisms of upper aerodigestive tract cancer in alcoholics. Int J Cancer. 2004 Feb 10;108(4):483-7.

77. Fang JL, Vaca CE. Development of a 32P-postlabelling method for the analysis of adducts arising through the reaction of acetaldehyde with 2'-deoxyguanosine-3'-monophosphate and DNA. Carcinogenesis. 1995 Sep;16(9):2177-85.

78. Wu D, Cederbaum AI. Oxidative stress and alcoholic liver disease. Semin Liver Dis. 2009 May;29(2):141-54.

79. Valko M, Leibfritz D, Moncol J, Cronin MT, Mazur M, Telser J. Free radicals and antioxidants in normal physiological functions and human disease. Int J Biochem Cell Biol. 2007;39(1):44-84.

80. Wu D, Zhai Q, Shi X. Alcohol-induced oxidative stress and cell responses. J Gastroenterol Hepatol. 2006 Oct;21 Suppl 3:S26-9.

81. Jimenez-Lopez JM, Cederbaum AI. CYP2E1-dependent oxidative stress and toxicity: role in ethanol-induced liver injury. Expert Opin Drug Metab Toxicol. 2005 Dec;1(4):671-85.

82. Seitz HK, Stickel F. Risk factors and mechanisms of hepatocarcinogenesis with special emphasis on alcohol and oxidative stress. Biol Chem. 2006 Apr;387(4):349-60.

83. Reddy SP. The antioxidant response element and oxidative stress modifiers in airway diseases. Curr Mol Med. 2008 Aug;8(5):376-83.

84. Akcil E, Caylakli F, Akiner M, Kocak M. Trace element concentrations and superoxide dismutase and catalase activities in benign and malignant larynx tumors. Biol Trace Elem Res. 2004 Dec;101(3):193-201.

85. Durak I, Isik AC, Canbolat O, Akyol O, Kavutcu M. Adenosine deaminase, 5' nucleotidase, xanthine oxidase, superoxide dismutase, and catalase activities in cancerous and noncancerous human laryngeal tissues. Free Radic Biol Med. 1993 Dec;15(6):681-4.

86. Gokul S, Patil V, Jailkhani R, Hallikeri K, Kattappagari K. Oxidant-antioxidant status in blood and tumor tissue of oral squamous cell carcinoma patients. Oral Dis. 2009 Jul 27.

87. Kacakci A, Aslan I, Toplan S, Oysu C, Aslan O, Aydemir B. Significance of the counteracting oxidative and antioxidative systems in the pathogenesis of laryngeal carcinoma. J Otolaryngol Head Neck Surg. 2009 Apr;38(2):172-7.

88. Patel JB, Shah FD, Shukla SN, Shah PM, Patel PS. Role of nitric oxide and antioxidant enzymes in the pathogenesis of oral cancer. J Cancer Res Ther. 2009 Oct-Dec;5(4):247-53.

89. Szuster-Ciesielska A, Hryciuk-Umer E, Stepulak A, Kupisz K, Kandefer-Szerszen M. Reactive oxygen species production by blood neutrophils of patients with laryngeal carcinoma and antioxidative enzyme activity in their blood. Acta Oncol. 2004;43(3):252-8.

90. Yigitbasi OG, Guney E, Haghighi N, Dogan P, Saraymen R, Balkanli S. Oxidant and antioxidant status in larynx squamous cell carcinomas. J Exp Clin Cancer Res. 2000 Dec;19(4):447-51.

91. Applebaum KM, Furniss CS, Zeka A, Posner MR, Smith JF, Bryan J, et al. Lack of association of alcohol and tobacco with HPV16-associated head and neck cancer. J Natl Cancer Inst. 2007 Dec 5;99(23):1801-10.

92. Gillison ML. Current topics in the epidemiology of oral cavity and oropharyngeal cancers. Head Neck. 2007 Aug;29(8):779-92.

93. Kreimer AR, Randi G, Herrero R, Castellsague X, La Vecchia C, Franceschi S. Diet and body mass, and oral and oropharyngeal squamous cell carcinomas: analysis from the IARC multinational case-control study. Int J Cancer. 2006 May 1;118(9):2293-7.

94. Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. Alcohol Res Health. 2007;30(1):5-13.

95. Bosron WF, Li TK. Genetic polymorphism of human liver alcohol and aldehyde dehydrogenases, and their relationship to alcohol metabolism and alcoholism. Hepatology. 1986 May-Jun;6(3):502-10.

96. Bosron WF, Magnes LJ, Li TK. Kinetic and electrophoretic properties of native and recombined isoenzymes of human liver alcohol dehydrogenase. Biochemistry. 1983 Apr 12;22(8):1852-7.

97. Quertemont E. Genetic polymorphism in ethanol metabolism: acetaldehyde contribution to alcohol abuse and alcoholism. Mol Psychiatry. 2004 Jun;9(6):570-81.

98. Boonyaphiphat P, Thongsuksai P, Sriplung H, Puttawibul P. Lifestyle habits and genetic susceptibility and the risk of esophageal cancer in the Thai population. Cancer Lett. 2002 Dec 5;186(2):193-9.

99. Chen YJ, Chen C, Wu DC, Lee CH, Wu CI, Lee JM, et al. Interactive effects of lifetime alcohol consumption and alcohol and aldehyde dehydrogenase polymorphisms on esophageal cancer risks. Int J Cancer. 2006 Dec 15;119(12):2827-31.

100. Yang CX, Matsuo K, Ito H, Hirose K, Wakai K, Saito T, et al. Esophageal cancer risk by ALDH2 and ADH2 polymorphisms and alcohol consumption: exploration of gene-environment and gene-gene interactions. Asian Pac J Cancer Prev. 2005 Jul-Sep;6(3):256-62.

101. Yang SJ, Wang HY, Li XQ, Du HZ, Zheng CJ, Chen HG, et al. Genetic polymorphisms of ADH2 and ALDH2 association with esophageal cancer risk in southwest China. World J Gastroenterol. 2007 Nov 21;13(43):5760-4.

102. Yokoyama A, Kato H, Yokoyama T, Tsujinaka T, Muto M, Omori T, et al. Genetic polymorphisms of alcohol and aldehyde dehydrogenases and glutathione S-transferase M1 and drinking, smoking, and diet in Japanese men with esophageal squamous cell carcinoma. Carcinogenesis. 2002 Nov;23(11):1851-9.

103. Hashibe M, Boffetta P, Zaridze D, Shangina O, Szeszenia-Dabrowska N, Mates D, et al. Evidence for an important role of alcohol- and aldehyde-metabolizing genes in cancers of the upper aerodigestive tract. Cancer Epidemiol Biomarkers Prev. 2006 Apr;15(4):696-703.

104. Hashibe M, McKay JD, Curado MP, Oliveira JC, Koifman S, Koifman R, et al. Multiple ADH genes are associated with upper aerodigestive cancers. Nat Genet. 2008 Jun;40(6):707-9.

105. Oze I, Matsuo K, Suzuki T, Kawase T, Watanabe M, Hiraki A, et al. Impact of Multiple Alcohol Dehydrogenase Gene Polymorphisms on Risk of Upper Aerodigestive Tract Cancers in a Japanese Population. Cancer Epidemiol Biomarkers Prev. 2009 Oct 27.

106. Hiraki A, Matsuo K, Wakai K, Suzuki T, Hasegawa Y, Tajima K. Gene-gene and gene-environment interactions between alcohol drinking habit and polymorphisms in alcohol-metabolizing enzyme genes and the risk of head and neck cancer in Japan. Cancer Sci. 2007 Jul;98(7):1087-91.

107. Risch A, Ramroth H, Raedts V, Rajaee-Behbahani N, Schmezer P, Bartsch H, et al. Laryngeal cancer risk in Caucasians is associated with alcohol and tobacco consumption but not modified by genetic polymorphisms in class I alcohol dehydrogenases ADH1B and ADH1C, and glutathione-S-transferases GSTM1 and GSTT1. Pharmacogenetics. 2003 Apr;13(4):225-30.

108. Li J, Benesova M, Goette K, Fiehn W, Conradt C, Hoermann K, et al. Enhanced risk of upper alimentary tract cancer in patients with the alcohol dehydrogenase 3*1 allele due to

increased salivary acetaldehyde following ethanol ingestion. Gastroenterology; 2001. p. A-160.

109. Peters ES, McClean MD, Liu M, Eisen EA, Mueller N, Kelsey KT. The ADH1C polymorphism modifies the risk of squamous cell carcinoma of the head and neck associated with alcohol and tobacco use. Cancer Epidemiol Biomarkers Prev. 2005 Feb;14(2):476-82.

110. Isse T, Matsuno K, Oyama T, Kitagawa K, Kawamoto T. Aldehyde dehydrogenase 2 gene targeting mouse lacking enzyme activity shows high acetaldehyde level in blood, brain, and liver after ethanol gavages. Alcohol Clin Exp Res. 2005 Nov;29(11):1959-64.

111. Katoh T, Kaneko S, Kohshi K, Munaka M, Kitagawa K, Kunugita N, et al. Genetic polymorphisms of tobacco- and alcohol-related metabolizing enzymes and oral cavity cancer. Int J Cancer. 1999 Nov 26;83(5):606-9.

112. Matsuo K, Hamajima N, Shinoda M, Hatooka S, Inoue M, Takezaki T, et al. Gene-environment interaction between an aldehyde dehydrogenase-2 (ALDH2) polymorphism and alcohol consumption for the risk of esophageal cancer. Carcinogenesis. 2001 Jun;22(6):913-6.

113. Asakage T, Yokoyama A, Haneda T, Yamazaki M, Muto M, Yokoyama T, et al. Genetic polymorphisms of alcohol and aldehyde dehydrogenases, and drinking, smoking and diet in Japanese men with oral and pharyngeal squamous cell carcinoma. Carcinogenesis. 2007 Apr;28(4):865-74.

114. Hashimoto T, Uchida K, Okayama N, Imate Y, Suehiro Y, Ueyama Y, et al. ALDH2 1510 G/A (Glu487Lys) polymorphism interaction with age in head and neck squamous cell carcinoma. Tumour Biol. 2006;27(6):334-8.

115. Canova C, Hashibe M, Simonato L, Nelis M, Metspalu A, Lagiou P, et al. Genetic associations of 115 polymorphisms with cancers of the upper aerodigestive tract across 10 European countries: the ARCAGE project. Cancer Res. 2009 Apr 1;69(7):2956-65.

116. Luo X, Kranzler HR, Zuo L, Wang S, Schork NJ, Gelernter J. Diplotype trend regression analysis of the ADH gene cluster and the ALDH2 gene: multiple significant associations with alcohol dependence. Am J Hum Genet. 2006 Jun;78(6):973-87.

117. Agarwal DP. Genetic polymorphisms of alcohol metabolizing enzymes. Pathol Biol (Paris). 2001 Nov;49(9):703-9.

118. Shimizu M, Lasker JM, Tsutsumi M, Lieber CS. Immunohistochemical localization of ethanol-inducible P450IIE1 in the rat alimentary tract. Gastroenterology. 1990 Oct;99(4):1044-53.

119. Oneta CM, Lieber CS, Li J, Ruttimann S, Schmid B, Lattmann J, et al. Dynamics of cytochrome P4502E1 activity in man: induction by ethanol and disappearance during withdrawal phase. J Hepatol. 2002 Jan;36(1):47-52.

120. Druesne-Pecollo N, Tehard B, Mallet Y, Gerber M, Norat T, Hercberg S, et al. Alcohol and genetic polymorphisms: effect on risk of alcohol-related cancer. Lancet Oncol. 2009 Feb;10(2):173-80.

121. Boccia S, Cadoni G, Sayed-Tabatabaei FA, Volante M, Arzani D, De Lauretis A, et al. CYP1A1, CYP2E1, GSTM1, GSTT1, EPHX1 exons 3 and 4, and NAT2 polymorphisms, smoking, consumption of alcohol and fruit and vegetables and risk of head and neck cancer. J Cancer Res Clin Oncol. 2008 Jan;134(1):93-100.

122. Buch SC, Nazar-Stewart V, Weissfeld JL, Romkes M. Case-control study of oral and oropharyngeal cancer in whites and genetic variation in eight metabolic enzymes. Head and Neck-Journal for the Sciences and Specialties of the Head and Neck. 2008 Sep;30(9):1139-47.

123. Garcia SM, Curioni OA, de Carvalho MB, Gattas GJ. Polymorphisms in alcohol metabolizing genes and the risk of head and neck cancer in a Brazilian population. Alcohol Alcohol. 2010 Jan-Feb;45(1):6-12.

124. Olivieri EH, da Silva SD, Mendonca FF, Urata YN, Vidal DO, Faria Mde A, et al. CYP1A2*1C, CYP2E1*5B, and GSTM1 polymorphisms are predictors of risk and poor outcome in head and neck squamous cell carcinoma patients. Oral Oncol. 2009 Sep;45(9):e73-9.

125. Ruwali M, Khan AJ, Shah PP, Singh AP, Pant MC, Parmar D. Cytochrome P450 2E1 and head and neck cancer: interaction with genetic and environmental risk factors. Environ Mol Mutagen. 2009 Jul;50(6):473-82.

126. Soya SS, Vinod T, Reddy KS, Gopalakrishnan S, Adithan C. CYP2E1 polymorphisms and gene-environment interactions in the risk of upper aerodigestive tract cancers among Indians. Pharmacogenomics. 2008 May;9(5):551-60.

127. Tai J, Yang M, Ni X, Yu D, Fang J, Tan W, et al. Genetic polymorphisms in cytochrome P450 genes are associated with an increased risk of squamous cell carcinoma of the larynx and hypopharynx in a Chinese population. Cancer Genet Cytogenet. 2010 Jan 1;196(1):76-82.

128. Bouchardy C, Hirvonen A, Coutelle C, Ward PJ, Dayer P, Benhamou S. Role of alcohol dehydrogenase 3 and cytochrome P-4502E1 genotypes in susceptibility to cancers of the upper aerodigestive tract. Int J Cancer. 2000 Sep 1;87(5):734-40.

129. Gonzalez MV, Alvarez V, Pello MF, Menendez MJ, Suarez C, Coto E. Genetic polymorphism of N-acetyltransferase-2, glutathione S-transferase-M1, and cytochromes P450IIE1 and P450IID6 in the susceptibility to head and neck cancer. J Clin Pathol. 1998 Apr;51(4):294-8.

130. Morita S, Yano M, Tsujinaka T, Akiyama Y, Taniguchi M, Kaneko K, et al. Genetic polymorphisms of drug-metabolizing enzymes and susceptibility to head-and-neck squamous-cell carcinoma. Int J Cancer. 1999 Mar 1;80(5):685-8.

131. Sugimura T, Kumimoto H, Tohnai I, Fukui T, Matsuo K, Tsurusako S, et al. Gene-environment interaction involved in oral carcinogenesis: molecular epidemiological study for metabolic and DNA repair gene polymorphisms. J Oral Pathol Med. 2006 Jan;35(1):11-8.

132. Boccia S, De Lauretis A, Gianfagna F, van Duijn CM, Ricciardi G. CYP2E1PstI/RsaI polymorphism and interaction with tobacco, alcohol and GSTs in gastric cancer susceptibility: A meta-analysis of the literature. Carcinogenesis. 2007 Jan;28(1):101-6.

133. Gao CM, Takezaki T, Wu JZ, Chen MB, Liu YT, Ding JH, et al. CYP2E1 Rsa I polymorphism impacts on risk of colorectal cancer association with smoking and alcohol drinking. World J Gastroenterol. 2007 Nov 21;13(43):5725-30.

134. Hunt CR, Sim JE, Sullivan SJ, Featherstone T, Golden W, Von Kapp-Herr C, et al. Genomic instability and catalase gene amplification induced by chronic exposure to oxidative stress. Cancer Res. 1998 Sep 1;58(17):3986-92.

135. Bradford BU, Seed CB, Handler JA, Forman DT, Thurman RG. Evidence that catalase is a major pathway of ethanol oxidation in vivo: dose-response studies in deer mice using methanol as a selective substrate. Arch Biochem Biophys. 1993 May 15;303(1):172-6.

136. Forsberg L, Lyrenas L, de Faire U, Morgenstern R. A common functional C-T substitution polymorphism in the promoter region of the human catalase gene influences transcription factor binding, reporter gene transcription and is correlated to blood catalase levels. Free Radic Biol Med. 2001 Mar 1;30(5):500-5.

137. Ho JC, Mak JC, Ho SP, Ip MS, Tsang KW, Lam WK, et al. Manganese superoxide dismutase and catalase genetic polymorphisms, activity levels, and lung cancer risk in Chinese in Hong Kong. J Thorac Oncol. 2006 Sep;1(7):648-53.

138. Lightfoot TJ, Skibola CF, Smith AG, Forrest MS, Adamson PJ, Morgan GJ, et al. Polymorphisms in the oxidative stress genes, superoxide dismutase, glutathione peroxidase and catalase and risk of non-Hodgkin's lymphoma. Haematologica. 2006 Sep;91(9):1222-7.

139. Choi JY, Neuhouser ML, Barnett M, Hudson M, Kristal AR, Thornquist M, et al. Polymorphisms in oxidative stress-related genes are not associated with prostate cancer risk in heavy smokers. Cancer Epidemiol Biomarkers Prev. 2007 Jun;16(6):1115-20.

140. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 2007 Jan;35(Database issue):D26-31.

141. Klaunig JE, Kamendulis LM, Hocevar BA. Oxidative stress and oxidative damage in carcinogenesis. Toxicol Pathol. 2010;38(1):96-109.

142. Kang D, Lee KM, Park SK, Berndt SI, Peters U, Reding D, et al. Functional variant of manganese superoxide dismutase (SOD2 V16A) polymorphism is associated with prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer study. Cancer Epidemiol Biomarkers Prev. 2007 Aug;16(8):1581-6.

143. Olson SH, Carlson MD, Ostrer H, Harlap S, Stone A, Winters M, et al. Genetic variants in SOD2, MPO, and NQO1, and risk of ovarian cancer. Gynecol Oncol. 2004 Jun;93(3):615-20.

144. Wang SS, Davis S, Cerhan JR, Hartge P, Severson RK, Cozen W, et al. Polymorphisms in oxidative stress genes and risk for non-Hodgkin lymphoma. Carcinogenesis. 2006 Sep;27(9):1828-34.

145. Arsova-Sarafinovska Z, Matevska N, Eken A, Petrovski D, Banev S, Dzikova S, et al. Glutathione peroxidase 1 (GPX1) genetic polymorphism, erythrocyte GPX activity, and prostate cancer risk. Int Urol Nephrol. 2009;41(1):63-70.

146. Ravn-Haren G, Olsen A, Tjonneland A, Dragsted LO, Nexo BA, Wallin H, et al. Associations between GPX1 Pro198Leu polymorphism, erythrocyte GPX activity, alcohol consumption and breast cancer risk in a prospective cohort study. Carcinogenesis. 2006 Apr;27(4):820-5.

147. Raaschou-Nielsen O, Sorensen M, Hansen RD, Frederiksen K, Tjonneland A, Overvad K, et al. GPX1 Pro198Leu polymorphism, interactions with smoking and alcohol consumption, and risk for lung cancer. Cancer Lett. 2007 Mar 18;247(2):293-300.

148. Ichimura Y, Habuchi T, Tsuchiya N, Wang L, Oyama C, Sato K, et al. Increased risk of bladder cancer associated with a glutathione peroxidase 1 codon 198 variant. J Urol. 2004 Aug;172(2):728-32.

149. McKay JD, Truong T, Gaborieau V, Chabrier A, Chuang SC, Byrnes G, et al. A Genome-Wide Association Study of Upper Aerodigestive Tract Cancers Conducted within the INHANCE Consortium. PLoS Genet. 2011 Mar;7(3):e1001333.

150. Wang D, Ritchie JM, Smith EM, Zhang Z, Turek LP, Haugen TH. Alcohol dehydrogenase 3 and risk of squamous cell carcinomas of the head and neck. Cancer Epidemiol Biomarkers Prev. 2005 Mar;14(3):626-32.

151. Zavras AI, Wu T, Laskaris G, Wang YF, Cartsos V, Segas J, et al. Interaction between a single nucleotide polymorphism in the alcohol dehydrogenase 3 gene, alcohol consumption and oral cancer risk. Int J Cancer. 2002 Feb 1;97(4):526-30.

152. Olshan AF, Weissler MC, Watson MA, Bell DA. Risk of head and neck cancer and the alcohol dehydrogenase 3 genotype. Carcinogenesis. 2001 Jan;22(1):57-61.

153. Schwartz SM, Doody DR, Fitzgibbons ED, Ricks S, Porter PL, Chen C. Oral squamous cell cancer risk in relation to alcohol consumption and alcohol dehydrogenase-3 genotypes. Cancer Epidemiol Biomarkers Prev. 2001 Nov;10(11):1137-44.

154. Fan CY. Genetic alterations in head and neck cancer: interactions among environmental carcinogens, cell cycle control, and host DNA repair. Curr Oncol Rep. 2001 Jan;3(1):66-71.

155. Geisler SA, Olshan AF, Weissler MC, Cai J, Funkhouser WK, Smith J, et al. p16 and p53 Protein expression as prognostic indicators of survival and disease recurrence from head and neck cancer. Clin Cancer Res. 2002 Nov;8(11):3445-53.

156. Ghosh S, Ghosh A, Maiti GP, Alam N, Roy A, Roy B, et al. Alterations of 3p21.31 tumor suppressor genes in head and neck squamous cell carcinoma: Correlation with progression and prognosis. Int J Cancer. 2008 Dec 1;123(11):2594-604.

157. Almadori G, Bussu F, Paludetti G. Should there be more molecular staging of head and neck cancer to improve the choice of treatments and thereby improve survival? Curr Opin Otolaryngol Head Neck Surg. 2008 Apr;16(2):117-26.

158. Almadori G, Bussu F, Paludettii G. Predictive factors of neck metastases in laryngeal squamous cell carcinoma. Towards an integrated clinico-molecular classification. Acta Otorhinolaryngol Ital. 2006 Dec;26(6):326-34.

159. Chen Y, Chen C. DNA copy number variation and loss of heterozygosity in relation to recurrence of and survival from head and neck squamous cell carcinoma: a review. Head Neck. 2008 Oct;30(10):1361-83.

160. De Herdt MJ, Baatenburg de Jong RJ. HGF and c-MET as potential orchestrators of invasive growth in head and neck squamous cell carcinoma. Front Biosci. 2008;13:2516-26.

161. Georgolios A, Batistatou A, Manolopoulos L, Charalabopoulos K. Role and expression patterns of E-cadherin in head and neck squamous cell carcinoma (HNSCC). J Exp Clin Cancer Res. 2006 Mar;25(1):5-14.

162. Hunter KD, Parkinson EK, Harrison PR. Profiling early head and neck cancer. Nat Rev Cancer. 2005 Feb;5(2):127-35.

163. Lim SC. Expression of c-erbB receptors, MMPs and VEGF in head and neck squamous cell carcinoma. Biomed Pharmacother. 2005 Oct;59 Suppl 2:S366-9.

164. Sebastian S, Grammatica L, Paradiso A. Telomeres, telomerase and oral cancer (Review). Int J Oncol. 2005 Dec;27(6):1583-96.

165. Thomas GR, Nadiminti H, Regalado J. Molecular predictors of clinical outcome in patients with head and neck squamous cell carcinoma. Int J Exp Pathol. 2005 Dec;86(6):347-63.

166. Seitz HK, Becker P. Alcohol metabolism and cancer risk. Alcohol Res Health. 2007;30(1):38-41, 4-7.

167. Seitz HK, Meier P. The role of acetaldehyde in upper digestive tract cancer in alcoholics. Transl Res. 2007 Jun;149(6):293-7.

168. Ragin CC, Langevin S, Rubin S, Taioli E. Review of studies on metabolic genes and cancer in populations of African descent. Genet Med. 2010 Jan;12(1):12-8.

169. North Carolina Central Cancer Registry. 2002-2006 North Carolina cancer incidence rates by race and ethnicity per 100,000 population. Raleigh, NC: North Carolina Center for Health Statistics 2009.

170. North Carolina Central Cancer Registry. Cancer mortality rates by race and sex per 100,000 population: 2002, 2003, 2004, 2005, 2006. Raleigh, NC: North Carolina Center for Health Statistics 2009.

171. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major US mortality databases. Ann Epidemiol. 2002 Oct;12(7):462-8.

172. Morgan TR, Mandayam S, Jamal MM. Alcohol and hepatocellular carcinoma. Gastroenterology. 2004 Nov;127(5 Suppl 1):S87-96.

173. Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, et al. High-throughput SNP genotyping on universal bead arrays. Mutat Res. 2005 Jun 3;573(1-2):70-82.

174. International HapMap C. A haplotype map of the human genome. Nature. 2005 Oct 27;437(7063):1299-320.

175. Kuo PH, Kalsi G, Prescott CA, Hodgkinson CA, Goldman D, van den Oord EJ, et al. Association of ADH and ALDH genes with alcohol dependence in the Irish Affected Sib Pair Study of alcohol dependence (IASPSAD) sample. Alcohol Clin Exp Res. 2008 May;32(5):785-95.

176. Luo X, Kranzler HR, Zuo L, Wang S, Schork NJ, Gelernter J. Multiple ADH genes modulate risk for drug dependence in both African- and European-Americans. Hum Mol Genet. 2007 Feb 15;16(4):380-90.

177. Macgregor S, Lind PA, Bucholz KK, Hansell NK, Madden PA, Richter MM, et al. Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. Hum Mol Genet. 2009 Feb 1;18(3):580-93.

178. Li Y, Marion MJ, Zipprich J, Santella RM, Freyer G, Brandt-Rauf PW. Gene-environment interactions between DNA repair polymorphisms and exposure to the carcinogen vinyl chloride. Biomarkers. 2009 May;14(3):148-55.

179. Cotterchio M, Boucher BA, Manno M, Gallinger S, Okey A, Harper P. Dietary phytoestrogen intake is associated with reduced colorectal cancer risk. J Nutr. 2006 Dec;136(12):3046-53.

180. Udler M, Maia AT, Cebrian A, Brown C, Greenberg D, Shah M, et al. Common germline genetic variation in antioxidant defense genes and survival after diagnosis of breast cancer. J Clin Oncol. 2007 Jul 20;25(21):3015-23.

181. D'Souza A, Kurien BT, Rodgers R, Shenoi J, Kurono S, Matsumoto H, et al. Detection of catalase as a major protein target of the lipid peroxidation product 4-HNE and the lack of its genetic association as a risk factor in SLE. BMC Med Genet. 2008;9:62.

182. Weng HY, Hsueh YH, Messam LL, Hertz-Picciotto I. Methods of covariate selection: directed acyclic graphs and the change-in-estimate procedure. Am J Epidemiol. 2009 May 15;169(10):1182-90.

183. Tiwari HK, Barnholtz-Sloan J, Wineinger N, Padilla MA, Vaughan LK, Allison DB. Review and evaluation of methods correcting for population stratification with a focus on underlying statistical principles. Hum Hered. 2008;66(2):67-86.

184. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, Schwartz AG. Examining population stratification via individual ancestry estimates versus self-reported race. Cancer Epidemiol Biomarkers Prev. 2005 Jun;14(6):1545-51.

185. Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, et al. Population genetic structure of variable drug response. Nat Genet. 2001 Nov;29(3):265-9.

186. Reed TE. Number of gene loci required for accurate estimation of ancestral population proportions in individual human hybrids. Nature. 1973 Aug 31;244(5418):575-6.

187. Pfaff CL, Barnholtz-Sloan J, Wagner JK, Long JC. Information on ancestry from genetic markers. Genet Epidemiol. 2004 May;26(4):305-15.

188. Tang H, Peng J, Wang P, Risch NJ. Estimation of individual admixture: analytical and study design considerations. Genet Epidemiol. 2005 May;28(4):289-301.

189. Barnholtz-Sloan JS, McEvoy B, Shriver MD, Rebbeck TR. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. Cancer Epidemiol Biomarkers Prev. 2008 Mar;17(3):471-7.

190. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol. 1986 Sep;124(3):453-69.

191. Coates RJ, Eley JW, Block G, Gunter EW, Sowell AL, Grossman C, et al. An evaluation of a food frequency questionnaire for assessing dietary intake of specific carotenoids and vitamin E among low-income black women. Am J Epidemiol. 1991 Sep 15;134(6):658-71.

192. Divaris K, Olshan AF, Smith J, Bell ME, Weissler MC, Funkhouser WK, et al. Oral health and risk for head and neck squamous cell carcinoma: the Carolina Head and Neck Cancer Study. Cancer Causes Control. 2010 Jan 5.

193. Rothman KJ, Boice Jr. JD. Epidemiologic analysis with a programmable calculator. U.S. Department of Health Education and Welfare, Public Health Service, National Institutes of Health; 1979.

194. Garcia-Closas M, Lubin JH. Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. Am J Epidemiol. 1999 Apr 15;149(8):689-92.

195. PASS Statistical Software. 11 ed: NCSS; 2011.

196. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. Stat Med. 1998 Jul 30;17(14):1623-34.

197. SAS. Statistical Analysis Software, Version 9.2, Cary, North Carolina. 2007.

198. Devore JL. Probability and statistics for engineering and the sciences. Belmont, CA: Thomson/Brooks/Cole; 2008.

199. Chapman JM, Cooper JD, Todd JA, Clayton DG. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. Hum Hered. 2003;56(1-3):18-31.

200. Roeder K, Bacanu SA, Sonpar V, Zhang X, Devlin B. Analysis of single-locus tests to detect gene/disease associations. Genet Epidemiol. 2005 Apr;28(3):207-19.

201. Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet. 2001 Apr;9(4):291-300.

202. Ziegler A, Konig IR. A statistical approach to genetic epidemiology : concepts and applications. Weinheim: Wiley-VCH; 2006.

203. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263-5.

204. Salem RM, Wessel J, Schork NJ. A comprehensive literature review of haplotyping software and methods for use with unrelated individuals. Hum Genomics. 2005 Mar;2(1):39-66.

205. Sinnwell J, Schaid D. haplo.stats: Statistical analysis of haplotypes with codominant traits and covariates when subjects are unrelated and haplotype is ambiguous 1.4.4 ed2009.

206. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.

207. Rice TK, Schork NJ, Rao DC. Methods for handling multiple testing. Adv Genet. 2008;60:293-308.

208. Silicon Genetics. Multiple testing corrections. http://www.silicongenetics.com/Support/GeneSpring/GSnotes/analysis_guides/mtc.pdf; 2003.

209. Conneely KN, Boehnke M. So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. Am J Hum Genet. 2007 Oct 26;81(6).

210. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst. 2004 Mar 17;96(6):434-42.

211. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet. 2007 Aug;81(2):208-27.

212. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. Addiction. 2003 Dec;98 Suppl 2:1-12.

213. Midanik LT. Validity of Self-reported Alcohol Use: a literature review and assessment. British Journal of Addiction: Wiley-Blackwell; 1988. p. 1019-30.

214. Babor TF, Brown J, Del Boca FK. Validity of self-reports in applied research on addictive behaviors: fact or fiction? Behavioral Assessment. 1990;12:5-31.

215. Babor TF, Stephens RS, Marlatt GA. Verbal report methods in clinical research on alcoholism: response bias and its minimization. J Stud Alcohol. 1987 Sep;48(5):410-24.

216. Gmel G, Rehm J. Measuring alcohol consumption. Contemporary Drug Problems. [Article]. 2004 Fall2004;31(3):467-540.

217. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002 Jun 21;296(5576):2225-9.

218. Cox DR, Oakes D. Analysis of Survival Data. New York: Chapman and Hall; 1984.

219. American Cancer Society. Cancer Facts & Figures 2009. Atlanta: American Cancer Society, Inc.; 2009.

220. American Cancer Society. Global Cancer Facts & Figures. 2nd ed. Atlanta: American Cancer Society; 2011.

221. SEER. Relative survival rates for oral cavity-pharynx and larynx cancers by survival time: all ages, all races, both sexes, 1988-2004. [10/18/2009]; Available from: http://seer.cancer.gov/faststats/selections.php.

222. Divaris K, Olshan AF, Smith J, Bell ME, Weissler MC, Funkhouser WK, et al. Oral health and risk for head and neck squamous cell carcinoma: the Carolina Head and Neck Cancer Study. Cancer Causes Control. 2010 Apr;21(4):567-75.

223. NHLBI Program for Genomic Applications [database on the Internet]. SeattleSNPs, Seattle, WA. [cited 2/22/08 - 4/25/08]. Available from: http://pga.gs.washington.edu.

224. Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, Brennan DJ, et al. FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. Carcinogenesis. [Research Support, N.I.H., Extramural]. 2010 Aug;31(8):1417-23.

225. Webb A, Lind PA, Kalmijn J, Feiler HS, Smith TL, Schuckit MA, et al. The Investigation into CYP2E1 in Relation to the Level of Response to Alcohol Through a Combination of Linkage and Association Analysis. Alcohol Clin Exp Res. 2011 Oct 19.

226. Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, et al. SEER Cancer Statistics Review, 1975-2008. Bethesda, MD: National Cancer Institute, 2008.

227. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tan PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. 2010 Jul 1;363(1):24-35.

228. Lu Y, Cederbaum AI. CYP2E1 and oxidative liver injury by alcohol. Free Radic Biol Med. 2008 Mar 1;44(5):723-38.

229. Brandwein-Gensler M, Smith RV. Prognostic indicators in head and neck oncology including the new 7th edition of the AJCC staging system. Head Neck Pathol. 2010 Mar;4(1):53-61.

230. Takes RP, Rinaldo A, Silver CE, Piccirillo JF, Haigentz M, Jr., Suarez C, et al. Future of the TNM classification and staging system in head and neck cancer. Head & Neck. [Review]. 2010 Dec;32(12):1693-711.

231. Lieber CS. Cytochrome P-4502E1: its physiological and pathological role. Physiol Rev. 1997 Apr;77(2):517-44.

232. Trafalis DT, Panteli ES, Grivas A, Tsigris C, Karamanakos PN. CYP2E1 and risk of chemically mediated cancers. Expert Opin Drug Metab Toxicol. 2010 Mar;6(3):307-19.

233. Koop DR. Alcohol metabolism's damaging effects on the cell: a focus on reactive oxygen generation by the enzyme cytochrome P450 2E1. Alcohol Res Health. 2006;29(4):274-80.

234. Ando Y, Shimizu T, Nakamura K, Mushiroda T, Nakagawa T, Kodama T, et al. Potent and non-specific inhibition of cytochrome P450 by JM216, a new oral platinum agent. Br J Cancer. 1998 Nov;78(9):1170-4.

235. Liu H, Baliga M, Baliga R. Effect of cytochrome P450 2E1 inhibitors on cisplatin-induced cytotoxicity to renal proximal tubular epithelial cells. Anticancer Res. 2002 Mar-Apr;22(2A):863-8.

236. Lu Y, Cederbaum AI. Cisplatin-induced hepatotoxicity is enhanced by elevated expression of cytochrome P450 2E1. Toxicol Sci. 2006 Feb;89(2):515-23.

237. Mašek V, Anzenbacherová E, Machová M, Brabec V, Anzenbacher P. Interaction of antitumor platinum complexes with human liver microsomal cytochromes P450. Anti-Cancer Drugs. 2009;20(5):305-11.

238. Gonzalez FJ. Role of cytochromes P450 in chemical toxicity and oxidative stress: studies with CYP2E1. Mutat Res. 2005 Jan 6;569(1-2):101-10.

239. Han X, Zheng T, Foss FM, Lan Q, Holford TR, Rothman N, et al. Genetic polymorphisms in the metabolic pathway and non-Hodgkin lymphoma survival. Am J Hematol. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. 2010 Jan;85(1):51-6.

240. Haque AK, Au W, Cajas-Salazar N, Khan S, Ginzel AW, Jones DV, et al. CYP2E1 polymorphism, cigarette smoking, p53 expression, and survival in non-small cell lung cancer: a long term follow-up study. Appl Immunohistochem Mol Morphol. 2004 Dec;12(4):315-22.

241. Neafsey P, Ginsberg G, Hattis D, Johns DO, Guyton KZ, Sonawane B. Genetic polymorphism in CYP2E1: Population distribution of CYP2E1 activity. J Toxicol Environ Health B Crit Rev. 2009;12(5-6):362-88.

242. Langenbucher J, Merrill J. The validity of self-reported cost events by substance abusers. Limits, liabilities, and future directions. Eval Rev. 2001 Apr;25(2):184-210.

243. AQUILINO WS. INTERVIEW MODE EFFECTS IN SURVEYS OF DRUG AND ALCOHOL USE: A FIELD EXPERIMENT. Public Opinion Quarterly. 1994 June 20, 1994;58(2):210-40.

244. Farshadpour F, Kranenborg H, Calkoen EV, Hordijk GJ, Koole R, Slootweg PJ, et al. Survival analysis of head and neck squamous cell carcinoma: Influence of smoking and drinking. Head & Neck. 2011 Jun;33(6):817-23.

245. Gritz ER, Demark-Wahnefried W. Health behaviors influence cancer survival. J Clin Oncol. 2009 Apr 20;27(12):1930-2.