ENVIRONMENTAL CARCINOGENESIS: CHARACTERIZATION OF RISING INCIDENCE AND DISCOVERY OF NOVEL BIOMARKERS OF PATHOGENESIS AND PROGNOSIS

by
Shama Virani

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Toxicology)
in The University of Michigan
2014

Doctoral Committee:
Assistant Professor Laura M. Rozek, Chair
Associate Professor Niladri Basu
Associate Professor Michele L. Cote, Wayne State University
Assistant Professor Rafael Meza
Professor Thomas G. Robins

DEDICATION

I would like to dedicate this work to the populations that served these studies. Specifically, I would like to dedicate my work on female breast cancer incidence to the female breast cancer patients of Songkhla, Thailand. I would like to dedicate the findings from the cadmium exposure study to the residents of the Mae Sot District in Tak Province, Thailand. Finally, I would like to dedicate the work done on novel biomarkers of survival to the head and neck cancer patients of the University of Michigan's Head and Neck Specialized Program of Research Excellence. These participants are extraordinarily generous to provide the large amount of data and biological samples that were used for my studies. Without them, this work would not have been possible.

ACKNOWLEDGEMENTS

Over the past five years, my experience as a doctoral student has been shaped by a large number of individuals who have provided mentorship, support, and encouragement throughout my time at the University of Michigan. This work would not have been possible without this support. First, I would like to acknowledge my dissertation chair and mentor, Dr. Laura Rozek, who has been incredibly encouraging and supportive throughout my Ph.D. career. She has not only mentored my research projects, but has also contributed greatly to my professional development. I am grateful for her honesty and support that has helped me recognize career path and future goals.

I would like to thank my committee members, Dr. Niladri Basu, Dr. Michele Cote, Dr. Rafael Meza and Dr. Thomas Robins, have all shown me how to think of my research in different perspectives, allowing me to better understand my research goals and their public health impact. Dr. Basu provided me with valuable guidance on heavy metals exposure and led me to better understand how human exposures can be quantitatively assessed. Dr. Michele Cote's insight into cancer epidemiology was useful in developing study questions and interpreting the results of my studies. Dr. Meza taught me sophisticated methodologies of cancer epidemiology that allowed me to analyze my data with a novel approach. Dr. Robins' expertise in global public health was invaluable to my training and research project development.

I am fortunate enough to have had the opportunity to conduct much of my dissertation work in Thailand. This work would have not been possible without the support and mentorship of my Thai collaborators, Dr. Hutcha Sriplung at Prince of Songkla University, Dr. Werawan

Ruangyuttikarn at Chiang Mai University and my Japanese collaborator, Dr. Muneko Nishijo at Kanazawa Medical University, all of with whom I spent time with in their respective countries. Dr. Sriplung offered knowledge on cancer epidemiology in Thailand and continues to inspire me to think of new ways to approach public health. Dr. Ruangyuttikarn provided guidance and support in the development of my research project. Dr. Nishijo hosted me in Japan and showed me the history behind my work on cadmium exposure. I could not have understood the importance of my work without these experiences.

I am grateful to Dr. Gregory Wolf, P.I. of the HN- SPORE program, to provide me with the opportunity to work with the University of Michigan Head and Neck Specialized Program of Research Excellence (HN-SPORE) and to support and guide me in my research. I also could not have completed this project without the efforts of the HN-SPORE team. This includes Lisa Peterson, the Project 3 manager who provided me with data and assisted with data cleanup; Emily Bellile, who provided me with biostatistics support; and Andrea Haddad and Alisha Virani, who are responsible for the majority of the HN-SPORE subject recruitment.

I would like to recognize the EHS administrative staff for their incredible support, specifically Kelsey Hargesheimer, Sue Crawford, Katherine Wood, and Patrice Sommerville. This amazing administrative team provided me with assistance in booking travel and keeping track of academic requirements, among other activities.

There are few people who understand doctoral student dilemmas and it was a great experience to share these with my labmates, Dr. Anna Arthur Parker, Justin Colacino, and Kathryn Demanelis. I am also thankful for the support and encouragement I have received from numerous other peers and friends within the EHS department, School of Public Health, and in Ann Arbor.

I am incredibly blessed to have the love and support of my family and friends during this challenging and rewarding period of my life. Specifically, I would like to acknowledge my mother and father, Yasmin and Hamid Virani, my sisters, Sheila and Alisha Virani and my childhood friend, Saeeda Merchant. These people have been the best support group throughout this time, offering continuous support whenever I needed it, endless optimism when I thought I couldn't do it, and nonstop encouragement if I was ever doubtful of my path. I am very fortunate and appreciative of their continued happiness and excitement for my success.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

ABSTRACT

This dissertation applied molecular epidemiology to address the relationship between the environment and cancer. The World Health Organization has identified cancer as the leading cause of death worldwide, and according to the International Agency for Research on Cancer, the global burden of cancer lies primarily in low- and middle-income countries (LMICs), and this disparity will become more extreme by 2035. Many LMICs are undergoing an epidemiologic transition involving a shift in health burden from infectious to chronic diseases. Because we know that the majority of cancers are sporadic and caused by a combination of genetic and environmental factors, it is crucial to understand the effects of the surrounding environment on cancer incidence, induction and outcomes.

The first objective of this dissertation was to characterize female breast cancer incidence trends in southern Thailand. Thailand is a LMIC in an epidemiologic transition. Women are adopting lifestyle changes that affect their risk of developing breast cancer, including reduced parity, later age at first birth and fewer years of breastfeeding. Incidence rates of breast cancer increased by almost 300% from 1990 to 2010. Both period and birth cohort effects played a role in shaping the increase in incidence. Three distinct incidence projection methods consistently suggested that incidence rates will continue to increase in the future with incidence for women age 50 and above increasing at a higher rate than for women below 50. This is the first study to utilize Thai cancer registry data in identifying relevant trends in a region-specific manner. The findings from this study identify opportunities for breast cancer prevention and

future research both by identifying priorities for screening and developing hypotheses for population-based studies.

Cadmium is a Class 1 carcinogen (IARC), however its mechanism of toxicity is unknown. My second objective was to determine epigenetic changes associated with this exposure. There are significant areas of environmental cadmium exposure in Thailand. Environmental cadmium exposure is associated with specific DNA methylation, and this relationship depended on the gender of the individual. Identification of these biomarkers of cadmium in a human population is a logical first step to identifying the mechanism of carcinogenesis.

Epigenetic changes are a hallmark of cancers, but the relationship between environmental and epidemiologic factors and these changes is not well understood. Tobacco, alcohol and human papillomavirus are known etiologic environmental factors of head and neck cancer. Novel relationships between methylation markers, survival and recurrence were identified in a large cohort of HNC patients. These associations differed based upon etiologic environmental factors. Identification of these significant epigenetic markers should open new horizons for interventions directed at reversible gene alterations and new therapeutic targets.

Overall, this research provides an understanding of the environmental contribution to cancer incidence, pathogenesis and prognosis. It offers a basis for future studies aimed at developing targeted interventions to address the rise in cancer incidence in Thailand, toxicity of environmental exposures and determining individualized treatment therapies to promote patient survival.

CHAPTER 1
Introduction

Theme

Cancer is a leading cause of death worldwide, with 14.1 million new cancer and 8.2 million deaths in 2012[1]. Approximately 19% of cancer cases worldwide are attributed to environmental factors[2]. The expected increase in global population is projected to shift 60-70% of the global cancer burden on low- and middle-income countries, where life expectancy is increasing with a concurrent decrease in mortality from infectious disease. In high-income countries, cancer mortality has stayed stable[1]. To address these issues, the theme of this dissertation is to characterize epidemiologic and molecular factors in cancer incidence, pathogenesis, and prognosis.

There is a significant environmental contribution to the development of cancer. Cancer rates vary substantially across countries[4], both by and within cancer site. While some variability may be due to genetic differences between populations, the majority of human cancers do not show simple patterns of inheritance[5], indicating the important role of environmental and lifestyle factor. This is illustrated by cancer incidence patterns from geographic variations and of migrant populations, which change to reflect rates of the adoptive country after one or two generations. For example, Japanese immigrants in the US had increased incidences of cancers of the breast, uterine corpus, and ovary for woman and prostate for men[6] similar to those rates in the US. Similarly, Chinese populations who generally had low incidence rates of prostate and breast cancers in China saw increased rates of these upon migration to the US[7]. Shifts in incidence

rates have also been observed in countries that have undergone rapid, substantial, economic development. For example, colorectal cancer rates of younger Japanese generations, born after 1930, mimic those of their US Caucasian counterparts[8]. Therefore, as LMICs undergo epidemiologic transitions that drive lifestyle changes, we can expect to see significant increases in cancer incidence rates that have previously been associated with developed countries, such as breast, colon, and prostate cancers.

Establishment of cancer registries has been a critical first step in public health planning, cancer prevention strategies, health care planning and patient care[10 11]. Registries collect information on age, gender, cancer site, and stage of disease, for all cancer cases in a defined geographical area[10]. These data are used in descriptive studies by providing statistics that can be used to identify patterns and variations across gender and age groups for specific cancer sites. Understanding patterns of cancer trends is important in generating etiological hypotheses that focus future studies and strategies of prevention, mechanism or treatment.. Characterization of these population incidence trends is necessary to evaluate the growing burden of cancer in LMICs.

The basic principles of multistage carcinogenesis and tumor biology are that cancer is a multistage process that involves sequential dysregulation of cellular pathways. These key features of carcinogenesis arise from the interaction of genetic alterations with acquired epigenetic abnormalities. Epigenetics is the study of heritable changes in gene expression that do not involve changes to the DNA sequence[16]. All cells in an individual are genetically identical but become structurally and functionally diverse due to differential expression of genes. Epigenetic mechanisms, such as DNA methylation, histone/chromatin modification, and non-coding RNAs interpret genetic code and regulate gene expression to promote development and

cell differentiation[17-21]. Therefore, the epigenome could be considered to be a link between genotype and phenotype[22].

DNA methylation is the most extensively characterized epigenetic modification. It involves the methylation of the fifth carbon of a cytosine nucleotide to create 5-methylcytosine (5mC). The methyl group of 5mC lies in the major groove of the double helix and can interfere with transcription factor binding to prevent gene expression[23-25]. Cytosine pairs with guanine by means of a phosphate group, and this dinucleotide (CpG) has been a major focus of epigenetic research because of its capacity to directly silence gene expression, particularly with respect to tumor-suppressor genes in carcinogenesis. CpG sites are unevenly distributed throughout the genome, concentrating in repetitive sequences such as tandem and interspersed repeats, distal gene regulatory regions, and at 5' promoter ends of genes, called CpG islands[26, 27]. CpG islands occupy approximately 60% of human gene promoters, most of which are constitutively expressed genes, such as housekeeping and regulatory genes[28]. A CpG island is generally defined as a 1000-kb stretch of DNA with GC content greater than 50%. The normal hypomethylated pattern of CpG islands is found to be consistent across various types of somatic tissues despite tissue-specific differences, illustrating that DNA methylation of these islands is not used as a regulatory mechanism of gene expression[23, 27, 28]. Therefore, when a CpG island becomes aberrantly methylated, it can have detrimental effects by stably silencing the associated gene[29]. The cancer cell genome is characterized by hypermethylation of CpG islands in promoter regions[27, 30, 31]. Hypermethylation in these regions promotes the progression of tumorigenesis by silencing tumor-suppressor genes[32-34]. Suppression of p16, a cell-cycle regulator, occurs in essentially all common human cancers[35]. Inactivating these tumor suppressors directly promotes tumorigenesis due to lack of control over cellular processes. In addition to tumor-suppressor

genes, hypermethylation of other classes of genes such as DNA repair genes and transcription factors can indirectly lead to tumorigenesis through silencing of further downstream targets or accumulation of genetic errors[36, 37]. Therefore, hypermethylation of CpG islands in cancers can affect multiple pathways to promote carcinogenesis. Although it may appear that hypomethylation and hypermethylation in cancer are opposing forces, the patterns usually coexist within the same tumor, though in different genomic regions. Further, the epigenetic abnormalities that occur because of hypo- and hypermethylation can interact in various ways to produce distinct subtypes of cancer, contributing to the complexity of the cancer cell epigenome.

Just as the epigenome is influenced by intrinsic signals, it is also shaped by external environmental stimuli. DNA methylation changes have been associated with a variety of environmental factors, including tobacco smoke, sunlight, air pollution, asbestos, physical activity, diet and alcohol consumption[17, 38-41]. A pivotal study analyzing epigenetic changes in monozygotic twins found that genetically identical twins were almost indistinguishable in terms of their epigenome in early years of life; however older monozygotic twins had remarkably different patterns of DNA methylation and, consequently, gene expression profiles[42], suggesting that epigenetic changes are influenced by extrinsic factors encountered throughout life and can explain phenotypic differences between monozygotic twins. A separate study examined monozygotic twins longitudinally and found that individual differences in methylation were not stable over time, suggesting that environmental influences account for these changes differentially across the genome[43]. DNA methylation profiles occur in response to environmental stimuli and therefore, provide information about biological predisposition for adverse health outcomes due to environmental exposures. DNA methylation tends to precede chromosomal instability and genetic alterations, may be reversible and is involved in early and precancerous

stages[44-47]. Additionally, these patterns are stable but not irreversible and remain flexible as the surrounding environment changes, making them desirable candidates for intermediate and prognostic biomarkers.

Specific Aims

Identification of human carcinogens is difficult when exposures are diffuse and chronic, as they tend to be in environmental exposures. The projected increase in cancer rates attributable to environmental factors highlights the need to determine novel biomarkers that predict risk of developing disease in response to exposures and to monitor disease progression. DNA methylation is a critical epigenetic mechanism incorporating the genome with the environment to regulate phenotype plasticity. It can control gene expression in a way that is stably propagated over multiple cell divisions, but is also flexible enough to respond to environmental influences. This intermediate position between stability and plasticity renders epigenetic information highly useful as biomarkers for monitoring cellular states that offer insight into potential mechanisms between exposure and disease.

Aim 1. *Identify and analyze population incidence trends of pre- and post-menopausal female breast cancer in Southern Thailand and project incidence rates into the future using mathematical modeling.*

As outlined above, LMICs are expected to exhibit an increase in cancer incidence rates as they undergo lifestyle changes to mimic those of developed countries. It is necessary to characterize these incidence trends to reveal changes in population rates. Therefore, in this aim, female breast cancer rates in southern Thailand are analyzed using various vigorous statistical

methods for comparative analyses and projected into the future using mathematical modeling. The hypotheses examined in this aim include:

Hypothesis 1.1: Epidemiologic trends are influenced by a combination of period and cohort effects due to the epidemiologic transition of the country.

Hypothesis 1.2: Breast cancer incidence will continue to increase in the future, particularly in postmenopausal women.

Aim 2. *Quantify epigenetic variation of candidate genes in a population of control and cadmium exposed subjects living in and around the Mae Sot District of Thailand and determine their association with urinary cadmium levels and renal biomarkers.*

Biomarkers of exposure are necessary to understand the biological changes that occur in response to environmental exposure. A sub-district in northern Thailand consisting of thirteen villages is highly exposed to environmental cadmium due to runoff from a long-standing zinc mine. Cadmium is listed as a carcinogen based on studies conducted in occupational settings with acute exposures. However, environmental cadmium exposures are chronically low and there is limited evidence as to the mechanism by which cadmium exerts its toxic effects in non-occupational exposures. This aim focuses on identification of biomarkers of exposure and investigates the following hypotheses:

Hypothesis 2.1: DNA methylation markers are associated with cadmium exposure as measured by urinary cadmium levels.

Hypothesis 2.2: DNA methylation markers are associated with markers of adverse health outcomes.

Aim 3: *Identify whether the association between exposure to viral/chemical carcinogens and tumor recurrence or survival is mediated through epigenetic mechanisms in head and neck cancer patients.*

Biomarkers of cancer are used to monitor disease progression, inform prognosis and guide therapy. In head and neck squamous cell carcinoma, patient characteristics such as, tobacco and alcohol use, human papillomavirus, patient age, comorbidity status and gender are generally used to identify high risk groups. However tumor heterogeneity makes it difficult to speculate disease trajectory, highlighting a need to identify biomarkers that encompass population characteristics, respond to environmental stimuli and provide information about tumor biology. The following hypotheses addressed the identification of these markers in this aim:

Hypothesis 3.1: Tumor DNA methylation profiles are associated with epidemiologic and clinicopathologic characteristics.

Hypothesis 3.2: DNA methylation profiles predict patient survival and recurrence from head and neck squamous cell carcinoma.

Conceptual Framework

Figure 1.2 outlines the conceptual framework for this dissertation research. Population-based epidemiology is needed to characterize cancer rates that are important for specific populations. Breast cancer incidence rates in southern Thailand will be characterized in Specific Aim 1. Once a cancer is found to affect a specific population, it is necessary to understand how

the environment affects induction of cancer using molecular events. The association between cadmium exposure and carcinogenesis will be investigated in Specific Aim 2. The environment can also play a role in prognosis in patients with cancer as well. The relationship between the environment and prognosis of head and neck squamous cell carcinoma will be explored in Specific Aim 3.

References

[1]Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser SM, C , Rebelo M, Parkin D, Forman D, Bray F. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide*. IARC CancerBase No 11 [Internet].International Agency for Research on Cancer. Accessed on March 10. 2013. Available From: http://globocan.iarc.fr

[2]World Health Organization (WHO). *Environmental and occupational cancers*. Accessed on March 19, 2014. 2011. Available From: http://www.who.int/mediacentre/factsheets/fs350/en/#

[3]Global Initiative for Cancer Registry Development. Facts, Figures, A Future. In: International Agency for Research on Cancer, ed., 2012.

[4]National Cancer Institute (NCI), National Institute of Environmental Health Sciences (NIEHS). Cancer and the Environment. In: US Department of Health and Human Services, ed., vol. NIH Publication No. 03–2039 2003.

[5]Perera FP, Weinstein IB. Molecular epidemiology: recent advances and future directions. *Carcinogenesis* 2000;**21**: 517-24.

[6]Dunn JE. Cancer epidemiology in populations of the United States--with emphasis on Hawaii and California--and Japan. *Cancer research* 1975;**35**: 3240-5.

[7]Yu H, Harris RE, Gao YT, Gao R, Wynder EL. Comparative epidemiology of cancers of the colon, rectum, prostate and breast in Shanghai, China versus the United States. *International journal of epidemiology* 1991;**20**: 76-81.

[8]Yiu HY, Whittemore AS, Shibata A. Increasing colorectal cancer incidence rates in Japan. *International journal of cancer Journal international du cancer* 2004;**109**: 777-81.

[9]Das A. Ch 2. Cancer Registry Databases: An Overview of Techniques of Statistical Analysis and Impact on Cancer Epidemiology. In: Verma M. *Cancer Epidemiology*ed. Totowa, NJ: Humana Press, 2009: 31-49.

[10]Jensen M, Storm HH. Purposes and uses of cancer registration. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. *Cancer Registration: Principles and Methods (Scientific Publication No 95)*ed.: International Agency for Research on Cancer, 1991: 7-21.

[11]Muir CS, Demaret E, Boyle P. The cancer registry in cancer control: an overview *The Role of the Registry in Cancer Control (IARC Scientific Publications No 66)*ed. Lyons, France: International Agency for Research on Cancer, 1985: 13-26

[12]Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics* 2004;**1**: 182-8.

[13]Boffetta P. Biomarkers in cancer epidemiology: an integrative approach. *Carcinogenesis* 2010;**31**: 121-6.

[14]Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. *Nature reviews Cancer* 2005;**5**: 845-56.

[15]Lowry LK. Role of biomarkers of exposure in the assessment of health risks. *Toxicology letters* 1995;**77**: 31-8.

[16]Wu C, Morris JR. Genes, genetics, and epigenetics: a correspondence. *Science* 2001;**293**: 1103-5.
[17]Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* 2003;**33 Suppl**: 245-54.

[18]Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell* 2007;**128**: 669-81.

[19]Kouzarides T. Chromatin modifications and their function. *Cell* 2007;**128**: 693-705.

[20]Ruthenburg AJ, Li H, Patel DJ, Allis CD. Multivalent engagement of chromatin modifications by linked binding modules. *Nature reviews Molecular cell biology* 2007;**8**: 983-94.

[21]Ozanne SE, Constancia M. Mechanisms of disease: the developmental origins of disease and the role of the epigenotype. *Nature clinical practice Endocrinology & metabolism* 2007;**3**: 539-46.

[22]Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 2007;**447**: 425-32.

[23]Feinberg AP, Tycko B. The history of cancer epigenetics. *Nature reviews Cancer* 2004;**4**: 143-53.

[24]Esteller M. Epigenetics in cancer. *The New England journal of medicine* 2008;**358**: 1148-59.

[25]Virani S, Colacino JA, Kim JH, Rozek LS. Cancer epigenetics: a brief review. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources* 2012;**53**: 359-69.

[26]Bird A. DNA methylation patterns and epigenetic memory. *Genes & development* 2002;**16**: 6-21.

[27]Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic acids research* 1982;**10**: 2709-21.

[28]Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 2008;**454**: 766-70.

[29]Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007;**128**: 683-92.

[30]Jones PA, Laird PW. Cancer epigenetics comes of age. *Nature genetics* 1999;**21**: 163-7.

[31]Meehan R, Lewis J, Cross S, Nan X, Jeppesen P, Bird A. Transcriptional repression by methylation of CpG. *Journal of cell science Supplement* 1992;**16**: 9-14.

[32]Fan S, Zhang X. CpG island methylation pattern in different human tissues and its correlation with gene expression. *Biochemical and biophysical research communications* 2009;**383**: 421-5.

[33]Hatziapostolou M, Iliopoulos D. Epigenetic aberrations during oncogenesis. *Cellular and molecular life sciences : CMLS* 2011;**68**: 1681-702.

[34]Illingworth RS, Bird AP. CpG islands--'a rough guide'. *FEBS letters* 2009;**583**: 1713-20.

[35]Liggett WH, Jr., Sidransky D. Role of the p16 tumor suppressor gene in cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1998;**16**: 1197-206.

[36]Alvarez-Nunez F, Bussaglia E, Mauricio D, Ybarra J, Vilar M, Lerma E, de Leiva A, Matias-Guiu X. PTEN promoter methylation in sporadic thyroid carcinomas. *Thyroid : official journal of the American Thyroid Association* 2006;**16**: 17-23.

[37]Esteller M, Toyota M, Sanchez-Cespedes M, Capella G, Peinado MA, Watkins DN, Issa JP, Sidransky D, Baylin SB, Herman JG. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is associated with G to A mutations in K-ras in colorectal tumorigenesis. *Cancer research* 2000;**60**: 2368-71.

[38]Bjornsson HT, Fallin MD, Feinberg AP. An integrated epigenetic and genetic approach to common human disease. *Trends in genetics : TIG* 2004;**20**: 350-8.

[39]Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, et al. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS genetics* 2009;**5**: e1000602.

[40]Langevin SM, Houseman EA, Christensen BC, Wiencke JK, Nelson HH, Karagas MR, Marsit CJ, Kelsey KT. The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. *Epigenetics : official journal of the DNA Methylation Society* 2011;**6**: 908-19.

[41]Gronniger E, Weber B, Heil O, Peters N, Stab F, Wenck H, Korn B, Winnefeld M, Lyko F. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS genetics* 2010;**6**: e1000971.

[42]Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, et al. Epigenetic differences arise during the

lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America* 2005;**102**: 10604-9.

[43]Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, Mill J. A longitudinal study of epigenetic variation in twins. *Epigenetics : official journal of the DNA Methylation Society* 2010;**5**: 516-26.

[44]Kanai Y, Hirohashi S. Alterations of DNA methylation associated with abnormalities of DNA methyltransferases in human cancers during transition from a precancerous to a malignant state. *Carcinogenesis* 2007;**28**: 2434-42.

[45]Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends in genetics : TIG* 2000;**16**: 168-74.

[46]Ramchandani S, Bhattacharya SK, Cervoni N, Szyf M. DNA methylation is a reversible biological signal. *Proceedings of the National Academy of Sciences of the United States of America* 1999;**96**: 6107-12.

[47]Gonzalgo ML, Jones PA. Mutagenic and epigenetic effects of DNA methylation. *Mutation research* 1997;**386**: 107-18.

Figure 1.1 Global burden of cancer by year 2035.

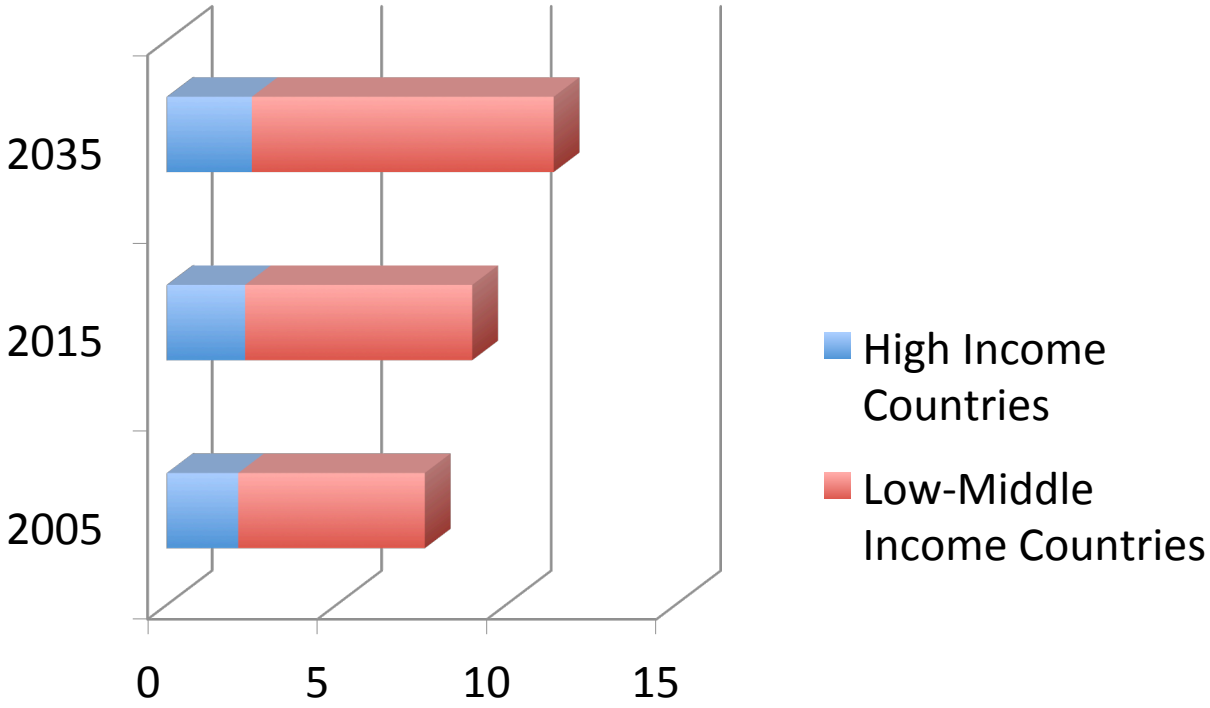
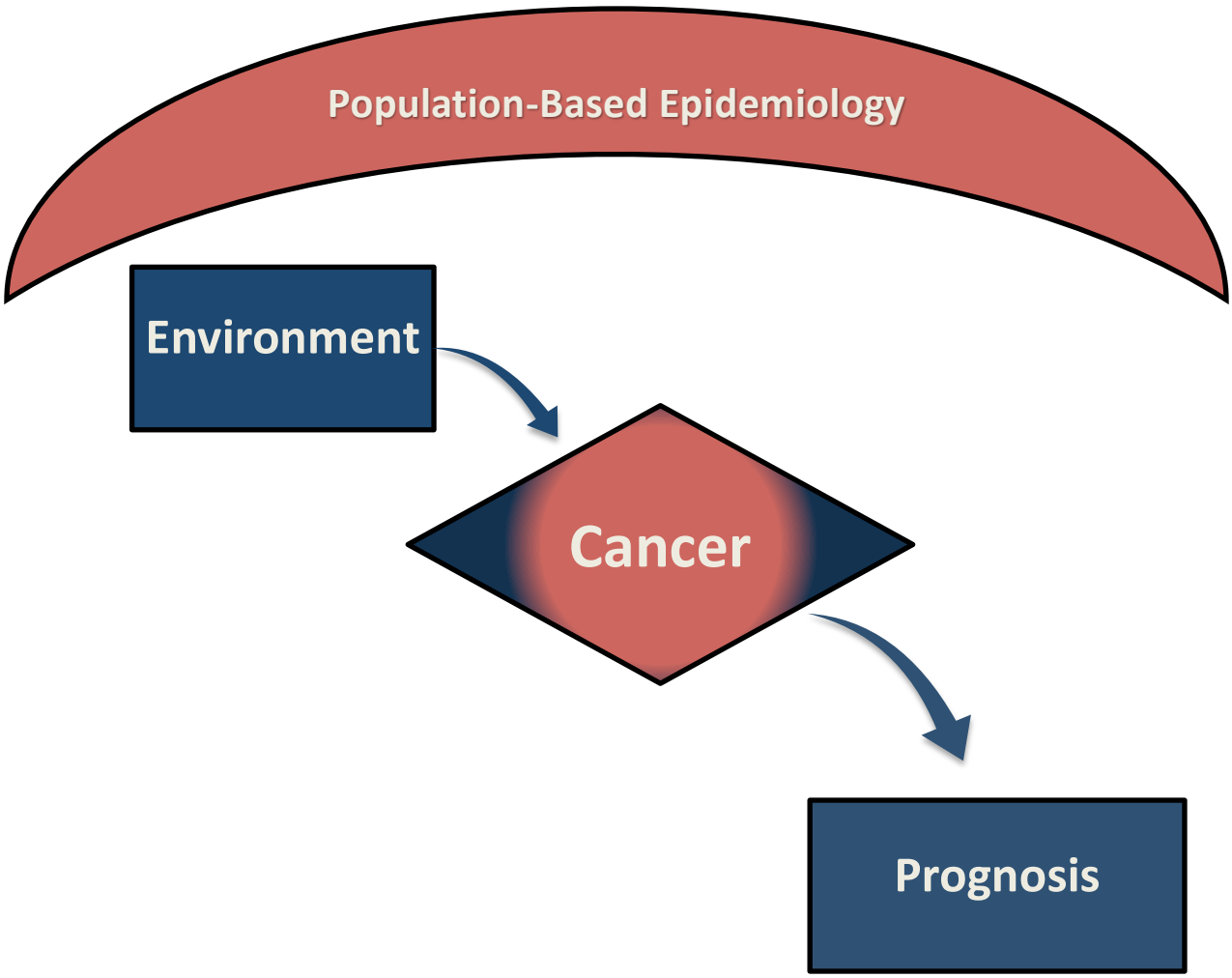


Figure adapted from: Global Initiative for Cancer Registry Development. Facts, Figures, A Future. In: International Agency for Research on Cancer, ed., 2012.

Figure 1.2. Conceptual framework.



CHAPTER 2

Escalating burden of breast cancer in southern Thailand: analysis of 1990-2010 incidence and prediction of future trends

ABSTRACT

Background: Thailand is undergoing an epidemiologic transition, with decreasing incidence of infectious diseases and increasing rates of chronic conditions, including cancer. Breast cancer has the highest incidence rates among females both in the southern region of Thailand and throughout Thailand. However, there is a lack of research on the epidemiology of this and other cancers.

Methods: Here we use cancer incidence data from the Songkhla Cancer Registry to characterize and analyze the incidence of breast cancer in Southern Thailand. We use joinpoint analysis, age-period-cohort models and nordpred analysis to investigate the incidence of breast cancer in Southern Thailand from 1990-2010 and project future trends from 2010-2029.

Results: We found that age-adjusted breast cancer incidence rates in Southern Thailand increased by almost 300% from 1990 to 2010 going from 10.0 to 27.8 cases per 100,000 person-years. Both period and cohort effects played a role in shaping the increase in incidence. Three distinct incidence projection methods consistently suggested that incidence rates will continue to increase in the future with incidence for women age 50 and above increasing at a higher rate than for women below 50.

Conclusions: To date, this is the first study to examine Thai breast cancer incidence from a regional registry. This study provides a basis for future planning strategies in breast cancer prevention and to guide hypotheses for population-based epidemiologic research in Thailand.

INTRODUCTION

Thailand is undergoing an epidemiologic transition, with decreasing rates of mortality due to infectious diseases and increasing rates of chronic conditions, including cancer. Worldwide cancer incidence is projected to rise 70% by the year 2030, with the largest burden on low- and middle – income countries (LMICs)[1]. Breast cancer poses a particular problem over the next decades as LMICs are increasingly adopting characteristics of a Western lifestyle. There is a strong association between Western lifestyle factors, such as diet and parity, and the incidence of breast cancer[2-5]. Since mammographic screening is often not available in LMICs, accurate incidence predictions are crucial to target resources to prevent and control breast cancer.

Breast cancer incidence rates are increasing throughout Thailand[6]. From 1998-2000, the age-standardized incidence rate (ASR) was 20.5 cases and increased to 30.7 cases per 100,000 person-years in 2008 [1, 7]. However, the epidemiologic basis of breast cancer in Thailand is not well characterized. The regions of Thailand vary dramatically in terms of population characteristics, risk factor exposures, and incidence rates[6]. Southern Thailand consists of a population of unique ethnic and cultural make-up. The Thai National Statistics office estimates that Muslims make up approximately 30% of the population of southern Thailand[8]. Muslims in the Songkhla province are nonetheless predominantly of Thai ethnicity. This results in a unique population since religiosity is known to be correlated with distinct lifestyle characteristics and potentially distinct risks for cancer and other diseases.

Registry data have been used in other LMICs to identify cancer trends, inform resource planning and guide hypotheses for population-based epidemiologic research[9-13]. A cancer registry was established in the Songkhla Province in 1989 to characterize the cancer incidence in southern Thailand. We investigated incidence rates of female breast cancer data from the

Songkhla registry from 1990-2010 using joinpoint regression and age-period-cohort models. The goal of these analyses was to characterize, for the first time, the breast cancer incidence trends in the province by calendar year, birth-cohort, age of diagnosis, and to project female breast cancer rates in southern Thailand to 2029 (Thai calendar: 2572), for all women and separated by pre and post-menopausal women. Each of the main regions of Thailand differ in their cancer incidence profiles and therefore, it is essential to analyze incidence rates by region[6, 12]. This is the first study to utilize Thai cancer registry data in identifying relevant trends in a region-specific manner. Our analyses highlight the utility of carefully collected cancer surveillance data in LMICs, and identify opportunities for breast cancer prevention and future research.

METHODS

Region

Songkhla, Thailand is a southern province occupying an area of 7,393 square kilometers on the eastern side of the Malaysian Peninsula (Figure 2.1). Muslims make up approximately 30% of the population of southern Thailand over 15 years of age [8]. The average life expectancy in the Thailand for females is 77.5 years[14].

Cancer Registry

The Songkhla Registry covers sixteen districts in southern Thailand. The population of southern Thailand at the 2010 census was 8.9 million people of which 4.4 million were females [15]. This registry actively compiles cancer cases from 23 sources including community hospitals, private hospitals, and the population registration office. The number of undetected cases is difficult to

estimate due to remote villages with limited access to health facilities and the use of traditional Thai medicine in lieu of health care services.

Female breast cancer cases were extracted from the Songkhla Cancer Registry from 1990 through 2010 using ICD-10 codes C50.X. Information included age at diagnosis, date of diagnosis and stage at diagnosis. In situ cases were excluded. Person-years were calculated from census data[15-17]. Age-specific incidence rates were calculated for eighteen age groups (0-4, 5-9, …, 80-84, and ≥85) and twenty-one calendar periods from 1990 to 2010 (1-year intervals). This stratification resulted in data for 90 single-year birth cohorts from 1902 through 1992.

Trend Analysis

We standardized age-adjusted breast cancer incidence rates in Songkhla, Thailand from 1990-2010 to the World Health Organization (WHO) world population[18, 19]. We then evaluated trends in incidence using the Joinpoint-Regression Program version 4.0.4[20]. Joinpoint regression identifies statistically significant trend change points (joinpoints) and the rate of change (annual percent change) in each trend segment using a Monte Carlo permutation method. Analyses were conducted for all females, and then for females younger than 50, and females 50 years old or older to determine differences in incidence trends above and below the mean age of menopause[21, 22].

Age-Period-Cohort Models

To investigate the effects of age, calendar year, and birth-cohort on the incidence of breast cancer, we fit age-period cohort models to the incidence rates. Age-specific incidence rates were calculated for 5-year age groups. We used the "classical" method of analysis, which fits a log-

linear model with a Poisson distribution to the observed data to estimate age, period, and cohort effects in a multiplicative APC model as follows:

$$log\lambda_{a,p} = f(a) + g(p) + h(c)$$

where the expected log-incidence rates $\lambda_{a,p}$ is assumed to be equal to a linear combination of effects that adjust for age *a,* period *p,* and birth-cohort *c*, with *c=p-a.* To address the well-known non-identifiability problem of APC models, we fit two-effects models (Age-Period and Age-Cohort) and then fit the remaining effect (cohort or period) to the respective model's residuals using natural splines to reduce random variation [23]. These are referred to as the AP-C and AC-P models.

Comparative Modeling for Prediction of Incidence Rates

We used three independent methods to project the incidence rates of breast cancer in the Songkhla province; joinpoint, nordpred and age-period-cohort model projections.

Joinpoint. Each best-fit joinpoint model was separated into its linear and residual components. The residuals described the curvature while the linear component illustrated the secular drift of the trend. Natural spline models were fit to each separate component and extrapolated out to 2029. To reduce the effect of drift in projected incidence rates, the linear component of the trend was attenuated by 5% each year from 2015-2019, 10% each year from 2020-2024, and 15% each year from 2025-2029. Residual and linear components were then added to give total age-

adjusted incidence rates with linear attenuation. Separate projections were made for all females, females below age 50 and female at or above age 50.

Nordpred. We used the nordpred R-package to project the breast cancer incidence in Songkhla. Nordpred fit an APC model to the data and then calculated world-standardized incidence rates for eighteen age groups (from 0-4 to ≥85, 5-year intervals) and 5-year interval periods (1990-1994,…, 2004-2009). Trends based upon all of the observed data were then extrapolated out to four 5-yr periods, ending in 2029. To avoid overestimation of cases from the multiplicative model, a power function in Nordpred was used to attenuate the linear drift by 25%, 50%, and 75%, respectively, for the second (2015-2019), third (2020-2024) and fourth (2025-2029) 5-year projection periods.

AP-C. The third prediction approach used a spline model fit to the AP-C model period effect estimates across all 18 age groups. Similar to the joinpoint projections, the linear and residual components of the period effects were separated out and projected individually to 2029. The linear component of each model was attenuated as done in joinpoint. Residual and attenuated linear components were added to yield period effects to year 2029. Incidence rates from 2010-2029 were calculated using the AP-C model age-effects, and the projected period effects, as well as projected population counts by age[15-17, 24]. Models were run with the same method for all females, females under age 50 and females at or above age 50.

AC-P. The fourth prediction model used a spline model fit to the estimated cohort effect estimates from the AC-P model. Linear and residual components were separated and projected to

2029, and incidence rates were calculated as described above. In all cases, the estimated period effects of the AC-P models were almost identical to 1. We therefore projected the period effects for these models as 1, and applied the corresponding attenuation effects as was done for joinpoint and AP-C model projections.

Joinpoint Regression Program and the R-statistical software were used for trend analysis and prediction (Epi 1.1.44 and NORDPRED, R version 3.0.1)[23, 25, 26].

RESULTS

There were 2,545 cases of total female breast cancer diagnosed in Songkhla Province from 1990-2010. Of these, 1,280 occurred in females under the age of 50 and 1,265 cases in females over the age of 50. Stage distribution by each age group is shown (Figure 2.2). The underlying trends in stage classification for both age groups show an increasing percentage of cases across periods for local and regional cancers and a decrease in unknown cancers. Because this decrease in unknown cases across time would create bias any trend analysis, stage was not used as a parameter in our analyses of female breast cancer rates in Songkhla.

Joinpoint Analysis

Breast cancer incidence rates increased from an age-standardized rate (ASR) of 10.0 in 1990 to an ASR of 27.8 cases per 100,000 person-years in 2010, an increase of 4.1% per year (Figure 2.3, Table 2.1). Overall, the incidence rate of female breast cancer increased at an annual percentage change (APC) of 1.55% from 1990-1997 and then increased by 15.08% per year from 1998-2000, although due to the low number of cases and the short period this trend was not

statistically significant. From 2001-2010, there was a significant 1.90% increase per year (Figure 2.3a.). Women under the age of 50 had a significant increase in incidence from 1990-1995 (APC: 10.73%) (Figure 2.3b), but the overall trend of incidence rates was low compared to women over 50. Breast cancer incidence for this group increased from an ASR of 23.98 cases in 1990 to 74.40 cases per 100,000 in 2010, a 4.63% increase per year (p-value<0.05) (Figure 2.3c). Comparison of both age group show the increase in incidence trend for all females was primarily driven by incidence rates of women over the age of 50 (Figure 2.3d).

Age-Period-Cohort Analyses

Descriptive Analysis. Incidence rates for all periods peak at the 45- 50 age group and again at the 75- 80 year age group (Figure 2.4a). Rates by age vary little by period except for the 25-29 year age group that shows an increase in incidence rates from 2000-2004 and the 85+ age group that increases sharply from 1995 to 2009. As expected, older age groups and birth cohorts have greater incidence rates as compared to younger subjects (Figure 2.4b, 2.4c, 2.4d). For a given age, there seems to be an increasing trend by birth year (Figure 2.4d).

Model fits. AC-P and AP-C models were fit for all females and females grouped by age (<50 and 50+). The Akaike Information Criteria for each model is shown in Table 2.2 [27]. The AC-P model for all females shows an age effect that increases with age (Figure 2.5a, left). The crude rate, which corresponds to the birth-cohort of 1948, increases sharply from age 32 to 52, going from 6.0 to 59.7 cases per 100,000. From age 57, the crude rate climbs from 76.4 to its peak at 140.3 cases per 100,000 person-years at age 82. In terms of cohort effects, the first cohort, 1902, has a 0.12 times lower risk of breast cancer as compared to those born in 1948 (Figure 2.5a, center). Conversely, those born in 1976 have a risk about two times higher

compared to those born in 1948. Cohorts after this year have low observations and large confidence intervals and therefore their higher relative risks must be interpreted with caution. Estimated period effects for this model are close to the relative risk of 1 (Figure 2.5a, right).

The AP-C model for all females shows an age effect that increases sharply until age 57 where it peaks with a crude incidence rate corresponding to the year 1992, of 42.5 cases per 100,000 person-years (Figure 2.5a, left). This trend decreases down to 28.4 cases per 100,000 for age 67, but then peaks again at 30.1 cases per 100,000 for at age 77. The period effect remains close to the reference level until 1998 when the risk for breast cancer increases and continues to increase up to 1.9 times greater than the reference in 2010 (Figure 2.5a, right). The cohort effect for the AP-C model remains close to 1 (Figure 2.5a, center).

The estimated age-period and cohort effects for the models by age (<50 and 50+) are consistent with those from the all female model (Figure 2.5b and 2.5c).

Projections

We used a variety of approaches to assess the validity of the projection of breast cancer rates in Songkhla through variability across model predictions[28]. We first used the joinpoint models to project incidence trends to 2029 for all females and for females grouped by age (Figure 2.6). Observed data from 2000-2010 was used as the basis for all projections to avoid influence from the notable drop in rates from 1995-2000. Projections based on all years (1990-2010) were also made for comparison and are shown in Figure 2.7. We found that incidence rates are expected to continue to increase to 29.2 cases per 100,000 person-years or 241 cases per year, in 2029. Females below the age of 50 are expected to reach 12.0 cases per 100,000 person-

years in 2029 (62 cases) while females at or above age 50 are predicted to reach 103.8 cases per 100,000 person-years, or 323 cases per year, in 2029 (Figure 2.6a, 2.6b).

The second approach used Nordpred to project trends up to the year 2029. Incidence rates for all females are expected to reach 30.7 cases per 100,000 person-years, or 252 cases per year, for period 2025-2029. For females below age 50, rates are expected to rise to 13.7 cases per 100,000 person-years, or 72 cases per year, while females at or above age 50 are expected to have an incidence rate of 91.3 cases per 100,000 person-years, or 273 cases per year, for the period 2025-2029 (Figure 2.6c, 2.6d).

The third approach projected rates using the AP-C and AC-P models. For the AC-P model, data was used for the cohorts from 1902 to 1972. While the estimated incidence rates did not differ greatly in a sensitivity analysis (shown in Figure 2.7), using this approach with censored observed data was deemed more appropriate due to small sample sizes in later cohorts. Incidence rates for all females reach 38.9 cases per 100,000 person-years (321 cases), 17.7 cases per 100,000 person-years (91 cases) for women under age 50, and 107.0 cases per 100,000 person-years (333 cases) for women at or above age 50 in year 2029 (Figure 2.6e, 2.6f).

According to the projected AP-C model, incidence rates for all females are expected to continue to increase to 51.4 cases per 100,000 person-years, or 424 cases per year, in 2029. Females below the age of 50 are expected to reach 28.3 cases per 100,000 person-years, or 146 cases in 2029, while females at or above 50 incidence rates are projected reach 195.6 cases per 100,000 person-years, or 609 cases, in 2029.

For comparison purposes, we have repeated the joinpoint and AC-P projections using all available data as the basis for prediction and repeated the Nordpred analysis using data from only the last ten years to project incidence rates and cases into the future. The joinpoint projections

increase slightly for all females and for the age groups. Our original analysis forecasted rates to be 29.2, 12.0 and 103.8 cases per 100,000 for all females, females below age 50 and females at or above age 50, respectively, in 2029. This translates to 241, 62 and 323 cases per year, respectively. Including all the observed data, these incidence rates decreased slightly to 28.7, 14.0 and 105.6 cases per 100,000 for all females, females below age 50 and females at or above age 50, translating to 238, 72, and 328 cases per year, respectively (Figure 2.7a, 2.7b). Although the results do not change much through inclusion of all data, our limitation of observed data to base projections off of is appropriate due to the uncharacteristic drop in incidence rates from 1995-2000. This is likely due to improper identification of cancer cases.

Using observed data for the last 10 years to project future rates is a common method in Nordpred analysis due to the rational that recent trends in incidence rates are more likely to be an influence in future rates. Using the recent trend as a basis for future predictions incidence rates are projected to decrease to 24.3 cases per 100,000 person-years for all females for the period 2025-2029, or 200 cases per year (Figure 2.7c, 2.7d). Although this was included for comparison purposes, the use of historical trend for projections is a more accurate depiction of future projections since there nothing significant was introduced, such as screening, which would impact future incidence rates of female breast cancer.

Cohorts by which to base projections on for the AC-P model began in 1902 and were limited to 1972 due to the small number of cases (≤10) from cohorts after 1972. To compare results, the projection was repeated using all available cohorts with cases (1902-1988). This approach changed the projected incidence rate for all females from 38.9 to 46.0 cases per 100,000 person-years, or 321 to 380 cases in 2029. The incidence rate for females below age 50

changed from 17.7 to 28.7 cases per 100,000 person-years, or 91 to 148 cases in 2029. Results for the women at or above age 50 did not change (Figure 2.7e, 2.7f).

DISCUSSION

This study showed that breast cancer incidence in southern Thailand has increased significantly since 1990, likely due to a combination of changes in demographics and the risk profile of the population, as well as increases in surveillance and breast cancer awareness. Projections from various models consistently show that the burden of breast cancer in Southern Thailand will continue to rise in the near future, potentially reaching a rate of about 29-51 cases per 100,000 PY for all females in 2029, although the magnitude of the increase could hypothetically be affected by future healthcare planning and other cancer burden control measures.

These data highlight the utility of surveillance data in LMICs with increasing rates of cancer. A strength of this study is the carefully collected data in Songkhla province. The Thai NCI has put considerable resources into training surveillance staff through standardization of ICD-O coding protocols, data reporting, collection and entry. Thus these are high quality data to construct robust and informative models. However, the registry data are limited due to the lack of information on biomarkers, religion, and other lifestyle characteristics. To date, there has not been comprehensive collection of this information at the population-level. Our research group is actively developing protocols to collect population-based data to identify novel risk factors for breast cancer in this population, to augment the utility of the existing registry data.

Generally, incidence rates for all females and for both age groups above and below the mean age at menopause (<50 years and ≥50 years) show increases in breast cancer incidence

from 1990-2010. As expected, the older age group has a higher incidence rate, contributing significantly to the overall trend for all females. The only significant increase in incidence for females under age 50 was during 1990-1995. Younger women in southern Thailand tend to be more aware of early detection measures, such as breast self-examinations, which could explain this significant increase. However, risk profiles for younger women tend to be low which may explain why the overall APC remained lower than that for all females and females at or above age 50. However, a limitation of this study is that we do not have case ascertainment information since the start of the registry. It has likely changed over time and may contribute somewhat to the increase in incidence rates over time. Nonetheless, we expect the impact to be limited since the registry follows very strict protocols for case identification.

It is difficult to ascertain if female breast cancer trends for southern Thai women are influenced predominantly by period or cohort effects. From Figure 5, we can see that both the AC-P model and the AP-C model fit the data well as the third parameter in each model hovers around the relative risk of 1. Age-specific rates in Figure 4 do not show proportionality between periods or cohorts, indicating the incidence trend is not shaped by either effect alone. Rather, the analysis shows that both period and cohort effects are both relevant in shaping the trends and so we chose to present both models and to make projections with each. Similar results have been shown in other populations[29, 30]. However, consistent with reports from other countries [31], there are indications of a stronger cohort effect. The AIC of the AC model is lower than that of the AP model, demonstrating a better fit for this model to the data (Table 2). The age effect estimates of the AC-P model exhibit the well-known 'Clemmensen's hook' where incidence rates increase exponentially until around age 50, and dip slightly before rising again at a lower rate. This phenomenon has been observed for breast cancer incidence and mortality across countries and is

thought to be due to the overlap of rates between pre- and post-menopausal women[32, 33]. The projections with the AC-P model are also more consistent with the two other methods than those from the AP-C model.

The ASR for all women in southern Thailand increased from 10.0 in 1990 to 27.8 cases per 100,000 person-years in 2010. Compared to a high income country, such as the United States where the ASR in 2010 was 126 cases per 100,000 person-years, this incidence is still relatively low, although this is likely a consequence of the lack of population-level breast cancer screening throughout Thailand[34]. In comparison to other LMICs, Thailand does still fall on the lower end of breast cancer rates. The ASRs from 1998-2002 in Korea, Taiwan, Hong Kong and Singapore were 37.2, 59.7, 69.1 and 90.1 cases per 100,000 person-years, respectively[35]. However, Thailand is undergoing an epidemiologic transition in lifespan and lifestyle characteristics combined with a shift from infectious to chronic diseases. For example, the average life expectancy in Thailand for females is expected to increase from 77.5 years in 2010 to 80.1 years in 2020[14]. Concordantly, the percentage of southern Thai women aged 50 and over increased from 15.3% in 1990 to 18.1% and then to 23.0% in 2000 and 2010 [15-17]. Our registry data show that the risk of breast cancer is highest for this age group and our projections suggest that it will continue to increase at a higher rate compared to women below 50. There is also considerable evidence that southern Thai women are adopting a Western lifestyle. The percentage of southern Thai women who are overweight (BMI $\geq$ 25 kg/m^2) increased from 36.3% in 2004 to 40.7% in 2009. The average percentage of southern Thai women with diabetes increased from 5.06% in 2004 to 6.0% in 2009. Hypertension in southern Thai women increased from 20.9% in 2004 to 21.4% in 2009[36]. Considering parity, the total fertility rate in Thai population has been gradually decreasing from 6 in 1970 to 3 in 1985, 2 in 1998 and 1.6 in 2010[37].

This shift in risk profiles of southern Thai women makes it necessary to focus on early detection and awareness of breast cancer. Young women especially are aware of the importance of screening. Within the past year, 66.9% and 12.3% of southern Thai women aged 15-59 years have conducted self breast examinations and clinical examinations, respectively, while the national adherence for these techniques is 60.7% and 17.9%[36]. The rates of breast self-examination in Thailand are higher in comparison with other LMICs such as Malaysia (46.8%) as well as with more developed countries such as Greece (21.4%) and the United Kingdom (24.8%) [38, 39]. In contrast, clinical examination rates are low with respect to other countries[36]. Mammography is not an established early detection measure in southern Thailand due to a general lack of trained radiologists to read mammographies and that this diagnostic tool is not covered by Thai universal health care. Therefore, it is crucial to identify population factors in breast cancer trends in Thailand to create targeted prevention strategies.

To assess the future breast cancer burden in Southern Thailand, we projected out the observed trends to the year 2029 using a variety of methods. All approaches have advantages and limitations; therefore, we decided to include them all with the intention of drawing conclusions based upon commonalities from various methods. There is a clear agreement from all our models that female breast cancer incidence rates will continue to increase. Predictions made with the aggregated 5-year periods using nordpred were similar to joinpoint predictions made with 1 year periods. Projected rates were 29.2 to 30.7 cases per 100,000 person-years, 12.0 to 13.7 cases per 100,000 person-years and 91.3 to 103.8 cases per 100,000 person-years for all females, females less than 50 and females at or above 50, respectively, in 2029. For purposes of understanding future breast cancer incidence trends, the projections these models are thought to be a lower bound of future rates.

The APC models used 1-year periods and projected 38.9 to 51.4 cases per 100,000 person-years for all females, 17.6 to 28.3 cases per 100,000 person-years for females under 50, and 107.0 to 195.6 cases per 100,000 person-years at or above 50. These ranges are wide; however they give an indication of the potential magnitude of breast cancer incidence rates in the future. With the current lack of early detection measures, such as mammography, we can expect that future incidence rates of breast cancer will increase only towards the lower limits of the projections and likely be late-stage and have poor survival rates. However, if early detection measures were introduced into universal health care coverage, future rates may increase to the upper end of these projections, but present at earlier stages and increase the probability of responding to treatment. Specifically, the increase in the proportion of women 50 and above in the population combined with their elevated risk for breast cancer as shown in this paper, identifies this group for targeted intervention.

CONCLUSIONS

This study provides the first in-depth look at the epidemiology of breast cancer in southern Thailand. Because of the changing risk profile of the women in this region, these findings need to be extended to characterize the population in terms of diet, lifestyle, and genetic factors. Future research should be directed towards strategies to control the burden of breast cancer in Thailand.

References

[1]Ferlay J, Shin, HR., Bra,y F., Forman, D., Mathers, C., and Parkin DM. *GLOBOCAN 2008 Cancer Incidence and Mortality Worldwide* IARC CancerBase. 05/14/2013. 2010. Available From: http://globocan.iarc.fr/

[2]Althuis MD, Dozier JM, Anderson WF, Devesa SS, Brinton LA. Global trends in breast cancer incidence and mortality 1973-1997. *International journal of epidemiology* 2005;**34**: 405-12.

[3]Kruk J. Association of lifestyle and other risk factors with breast cancer according to menopausal status: a case-control study in the Region of Western Pomerania (Poland). *Asian Pacific journal of cancer prevention : APJCP* 2007;**8**: 513-24.

[4]Cui X, Dai Q, Tseng M, Shu XO, Gao YT, Zheng W. Dietary patterns and breast cancer risk in the shanghai breast cancer study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2007;**16**: 1443-8.

[5]Layde PM, Webster LA, Baughman AL, Wingo PA, Rubin GL, Ory HW. The independent associations of parity, age at first full term pregnancy, and duration of breastfeeding with the risk of breast cancer. Cancer and Steroid Hormone Study Group. *Journal of clinical epidemiology* 1989;**42**: 963-73.

[6]Khuhaprema T, Attasara, P., Sriplug, H., Wiangnon, S., Sumitsawan, Y., Sangrajrang, S. *Cancer in Thailand Vol. VI, 2004-2006*. Bangkok: Ministry of Public Health, National Cancer Institute, 2012.

[7]Khuhaprema T SP, Sriplung H, Wiangnon S, Sumitsawan Y, Attasara P. *Cancer in Thailand Vol. IV, 1998-2000*. Bangkok: Ministry of Public Health, Ministry of Education, 2007.

[8]National Statistical Office of Thailand. *Population Survey*. Office of the Prime Minister. 6/5/2013. 2005. Available From: http://service.nso.go.th/nso/nsopublish/service/survey/cult48.pdf

[9]Bhurgri Y. Karachi Cancer Registry Data--implications for the National Cancer Control Program of Pakistan. *Asian Pacific journal of cancer prevention : APJCP* 2004;**5**: 77-82.

[10]Jarlbaek L, Christensen L, Bruera E, Gilsa Hansen D. The epidemiology of long- and short-term cancer survivors. A population-based cohort study exploring denominators for rehabilitation and palliative care programs. *Acta Oncol* 2013.

[11]Jedy-Agba E, Curado MP, Ogunbiyi O, Oga E, Fabowale T, Igbinoba F, Osubor G, Otu T, Kumai H, Koechlin A, Osinubi P, Dakum P, et al. Cancer incidence in Nigeria: a report from population-based cancer registries. *Cancer epidemiology* 2012;**36**: e271-8.

[12]Sriplung H, Sontipong S, Martin N, Wiangnon S, Vootiprux V, Cheirsilpa A, Kanchanabat C, Khuhaprema T. Cancer incidence in Thailand, 1995-1997. *Asian Pacific journal of cancer prevention : APJCP* 2005;**6**: 276-81.

[13]Vaktskjold A, Lebedintseva JA, Korotov DS, Tkatsjov AV, Podjakova TS, Lund E. Cancer incidence in Arkhangelskaja Oblast in northwestern Russia. The Arkhangelsk Cancer Registry. *BMC cancer* 2005;**5**: 82.

[14]Office of the National Economic and Social Development Board. *Life Expectancy at Birth years 2508-2573*. 6/5/2013. Available From: http://social.nesdb.go.th/SocialStat/StatReport_Final.aspx?reportid=87&template=1R2C&yeartype=M&subcatid=4

[15]National Statistic Office. *2010 Population and Housing Census*. Office of the Prime Minister: Bangkok. 2013.

[16]National Statistical Office. *1990 Population and Housing Census*. Office of the Prime Minister: Bangkok. 1994.

[17]National Statistical Office. *2000 Population and Housing Census*. Office of the Prime Minister: Bangkok. 2002.

[18]Ahmad OE B-PC, Lopex AD, Murray CJL, Lozano R, Inoue M., Age standardization of rates: a new WHO standard. World Health Organization, 2000.

[19]Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in medicine* 2000;**19**: 335-51.

[20]Joinpoint Regression Program, Version 4.0.4 - May 2013: Statistical Methodology and Applications Branch, Surveillance Research Program, National Cancer Institute.

[21]Thomas F, Renaud F, Benefice E, de Meeus T, Guegan JF. International variability of ages at menarche and menopause: patterns and main determinants. *Human biology* 2001;**73**: 271-90.

[22]Henderson KD, Bernstein L, Henderson B, Kolonel L, Pike MC. Predictors of the timing of natural menopause in the Multiethnic Cohort Study. *American journal of epidemiology* 2008;**167**: 1287-94.

[23]Carstensen B, Plummer, M., Laara, E., Hills, M. *Epi: A Package for Statistical Analysis in Epidemiology*. R package version 1.1.49. 2013. Available From: http://CRAN.R-project.org/package=Epi

[24]Office of the National Economic and Social Development Board of Thailand. *Population Projections for Thailand 2553-2573.* 6/5/2013. 2013. Available From: http://www.nesdb.go.th/

[25]RStudio. RStudio: Integrated development environment for R (Version 0.97.336) Boston, MA, 2012.

[26]Møller B FH, Hakulinen ., Tryggvadottir L, Storm HH, Talback M, Haldorsen T. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer* 2002: 11 Suppl 1:S1–S96.

[27]Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;**19**: 716.

[28]Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2003;**6**: 9-17.

[29]Viel JF, Rymzhanova R, Fournier E, Danzon A. Trends in invasive breast cancer incidence among French women not exposed to organized mammography screening: an age-period-cohort analysis. *Cancer epidemiology* 2011;**35**: 521-5.

[30]Dhillon PK, Yeole BB, Dikshit R, Kurkure AP, Bray F. Trends in breast, ovarian and cervical cancer incidence in Mumbai, India over a 30-year period, 1976-2005: an age-period-cohort analysis. *British journal of cancer* 2011;**105**: 723-30.

[31]Moolgavkar SH, Stevens RG, Lee JA. Effect of age on incidence of breast cancer in females. *Journal of the National Cancer Institute* 1979;**62**: 493-501.

[32]De Waard F, Baanders-Vanhalewijn EA, Huizinga J. The Bimodal Age Distribution of Patients with Mammary Carcinoma; Evidence for the Existence of 2 Types of Human Breast Cancer. *Cancer* 1964;**17**: 141-51.

[33]Anderson DE. Genetic study of breast cancer: identification of a high risk group. *Cancer* 1974;**34**: 1090-7.

[34]Howlader N NA, Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA. SEER Cancer Statistics Review, 1975-2010 Bethesda, MD: National Cancer Institute.

[35]Shin HR, Joubert C, Boniol M, Hery C, Ahn SH, Won YJ, Nishino Y, Sobue T, Chen CJ, You SL, Mirasol-Lumague MR, Law SC, et al. Recent trends and patterns in breast cancer incidence among Eastern and Southeastern Asian women. *Cancer causes & control : CCC* 2010;**21**: 1777-85.

[36]National Health Examination Survey Office. *Report of the National Health Examination Survey in Thai Population, 2008-2009*. Health System Research Institute. 6/4/2013. Available From: http://www.nheso.or.th/content_detail.php?con_id=14

[37]United Nations Population Fund. *Impact of Demographic Change in Thailand*. 6/5/2013. 2011. Available From: http://countryoffice.unfpa.org/thailand/?publications=3297

[38]Chouliara Z, Papadioti-Athanasiou V, Power KG, Swanson V. Practice of and attitudes toward breast self-examination (BSE): a cross-cultural comparison between younger women in Scotland and Greece. *Health care for women international* 2004;**25**: 311-33.

[39]Loh SY, Chew SL. Awareness and practice of breast self examination among malaysian women with breast cancer. *Asian Pacific journal of cancer prevention : APJCP* 2011;**12**: 199-202.

Table 2.1: Joinpoint trends in female breast cancer incidence rates in Songkhla, Thailand

		Trend 1		Trend 2		Trend 3		Trend 4	
Age	N	Years	APC	Years	APC	Years	APC	Years	APC
All	2545	1990-1997	1.55	1998-2000	15.08	2001-2010	1.90*		
<50	1280	1990-1995	10.73*	1996-1998	-7.92	1999-2000	30.83	2001-2010	0.1
>50	1265	1990-2010	4.63*						

Abbreviation: APC Annual Percentage Change

*APC is significantly different from zero $p<0.05$

Table 2.2: Akaike information criteria (AIC) values for the AC, AC and APC models relative (difference) to the Age only model**

Model	AIC*
Age-Period	-161.68
Age-Cohort	-163.63
Age-Period-Cohort	-184.93

*'-2 x log(likelihood) + 2 x number of estimated parameters.

**Relative values that weight the goodness of fit of the model to empirical data. The lower the AIC, the better the model fit.

Figure 2.1: Map of Thailand. Songkhla province is highlighted

Figure 2.2: Stage distribution across five-year periods for women below age 50 (**2a**) and women at or above age 50 (**2b**). Age-adjusted standardized rates by stage for each year for women below age 50 (**2c**) and women at or above age 50 (**2d**). The percentage of local cases showed the largest increase in women below the age of 50 (**2a**) from 2000-2009, while the percentage of regional cases increased steadily over the time period. Both local and regional cases increased steadily for women at or above age 50 (**2b**). Percentage of distant and unknown cases decreased steadily for both age groups. Age-adjusted rates for both age groups show an increase in incidence of local cancers and a notable increase for regional cancers, while incidence rates of unknown cancers decrease (**2c, 2d**). NOTE: Y-axis have different ranges between **2c** and **2d**.

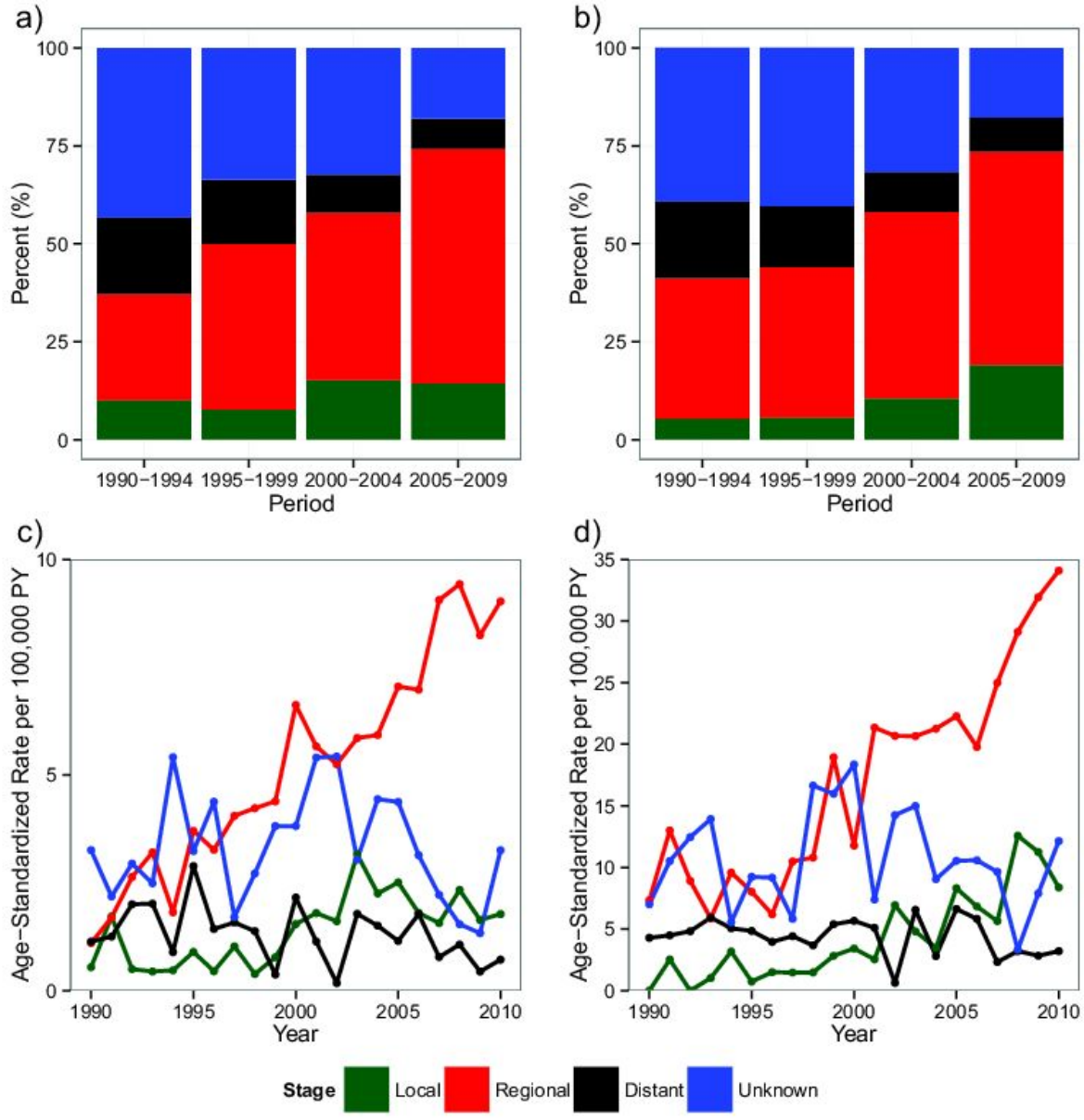


Figure 2.3: Trend analysis of age-adjusted breast cancer incidence trends for females from 1990-2010. Trend analysis of age-adjusted breast cancer incidence trends for females from 1990-2010. (*) denotes an APC significantly different from zero. **a)** Age-adjusted incidence trends for all females **b)** Age-adjusted incidence trends for females under 50 years of age **c)** Age-adjusted incidence rates for females age 50 years and older d) Comparison of age-adjusted incidence trends for females below and at or above age. NOTE: Y-axis have different ranges.

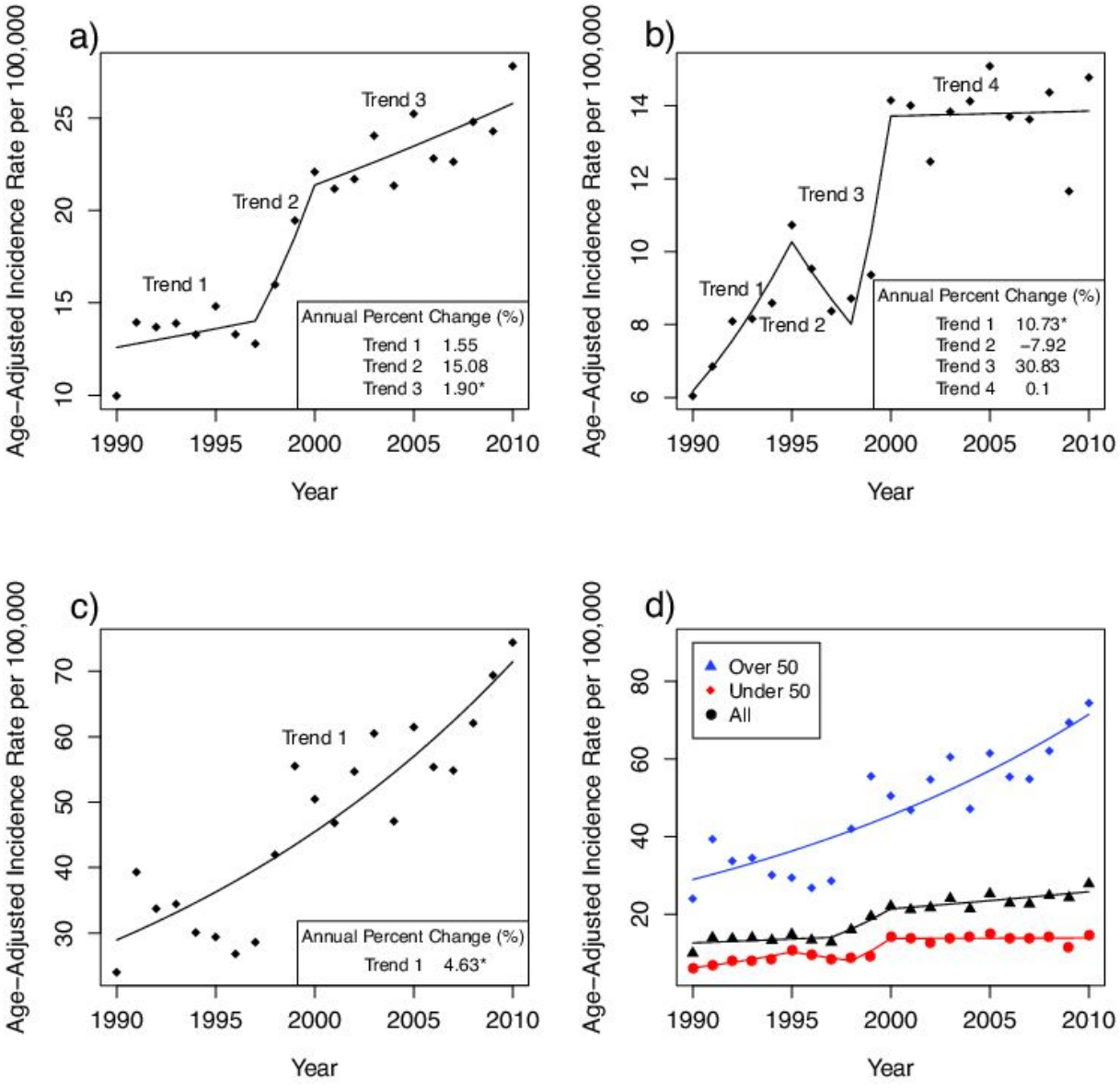


Figure 2.4: Incidence of female breast cancer in women per 100,000 person-years by age, period and birth cohort. Incidence rates by period **(a)**, birth cohort **(b)**, and age **(c,d)**.

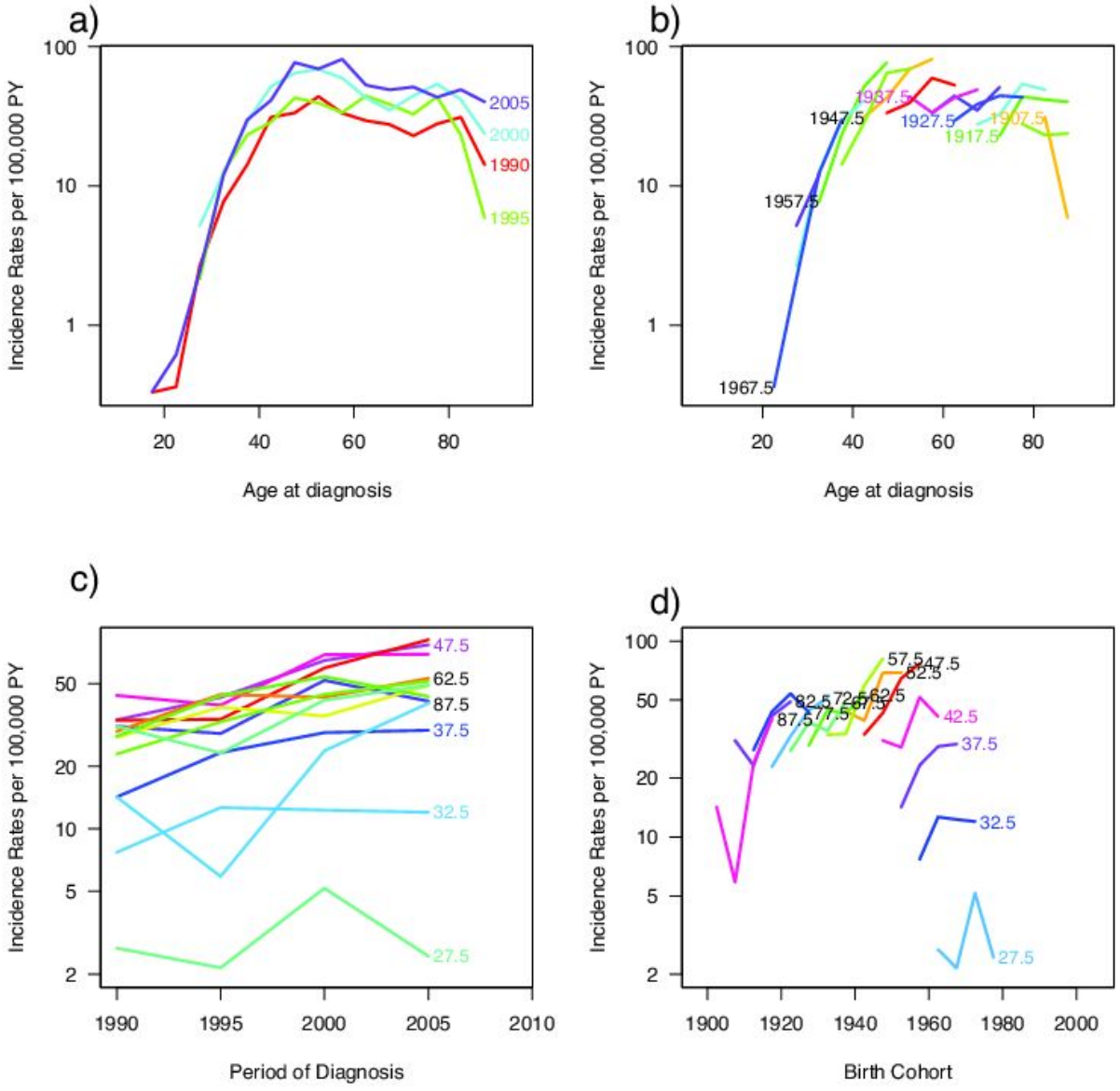


Figure 2.5: APC trend analysis. AC-P (blue) and AP-C (red) models for **a)** all females, **b)** females under the age of 50 and **c)** females at or above the age of 50. Incidence rates are plotted in the log scale (left y-axis).

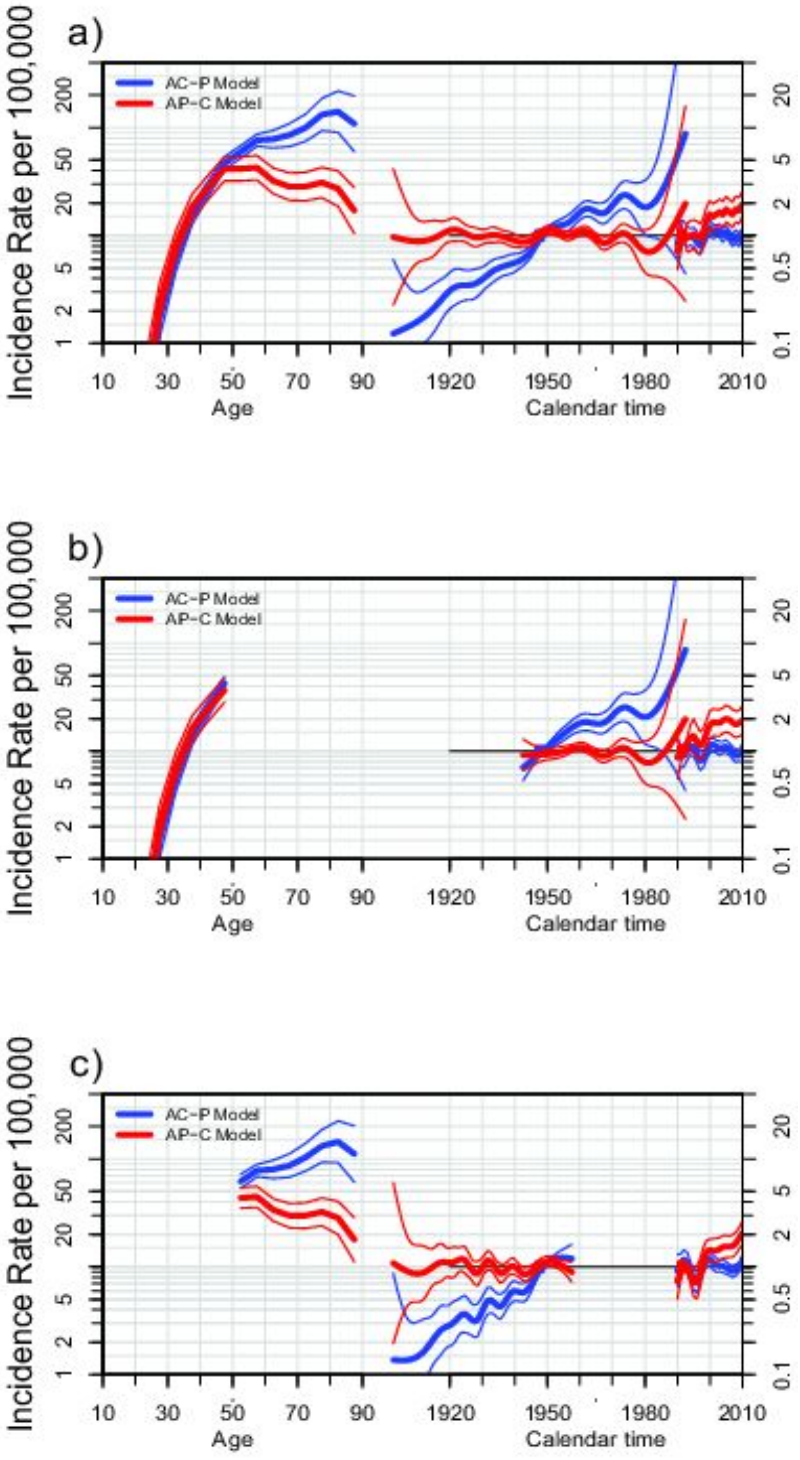


Figure 2.6: Breast Cancer incidence trend projections to 2029. **a)** Rate projections for all females and by age group using the joinpoint model **b)** Case projection for all females from the joinpoint model **c)** Rate projections using nordpred **d)** Case projections for all females using the projection from the nordpred model **e)** Rate projections from the AP models **f)** Case projections for all females from the APC models

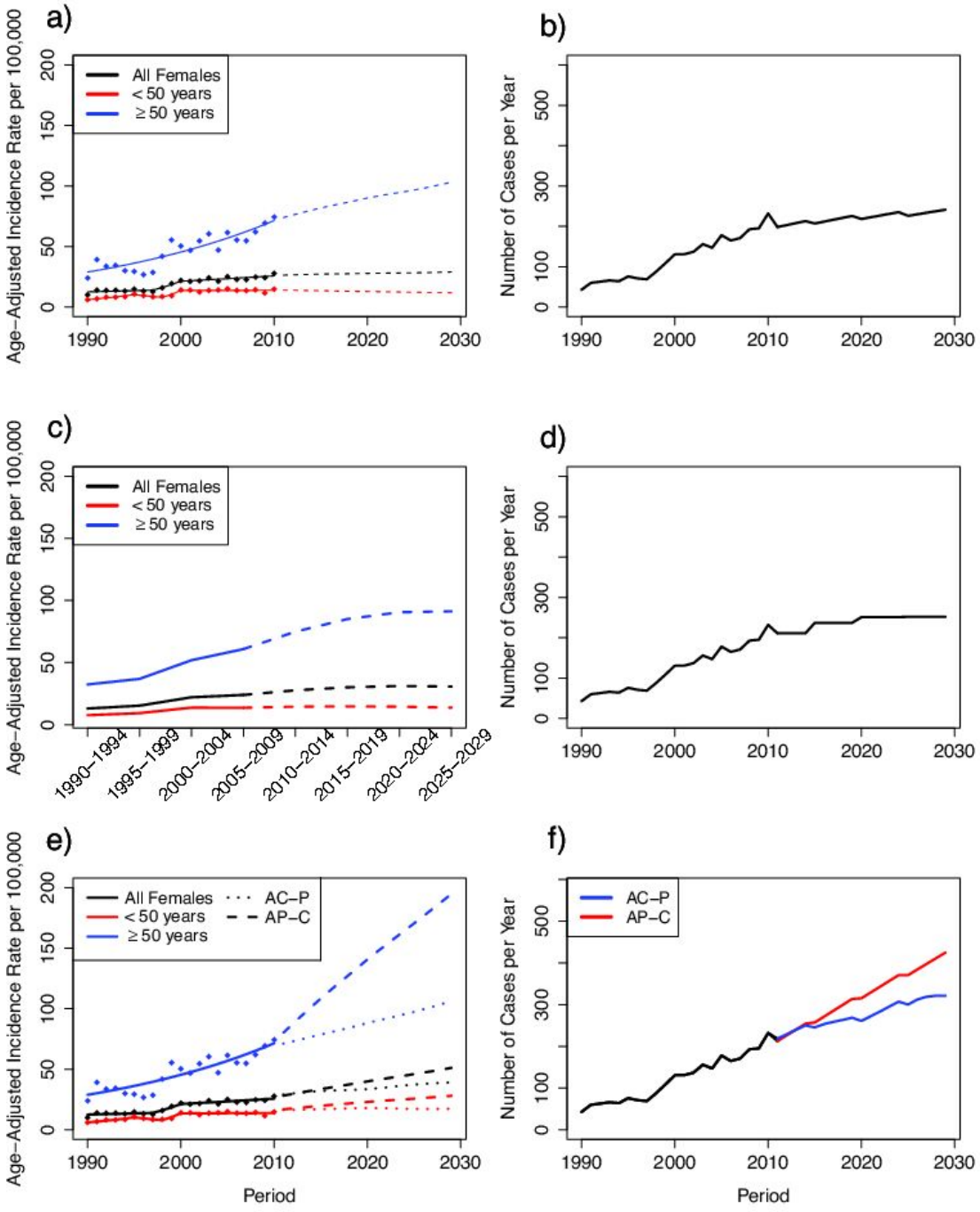
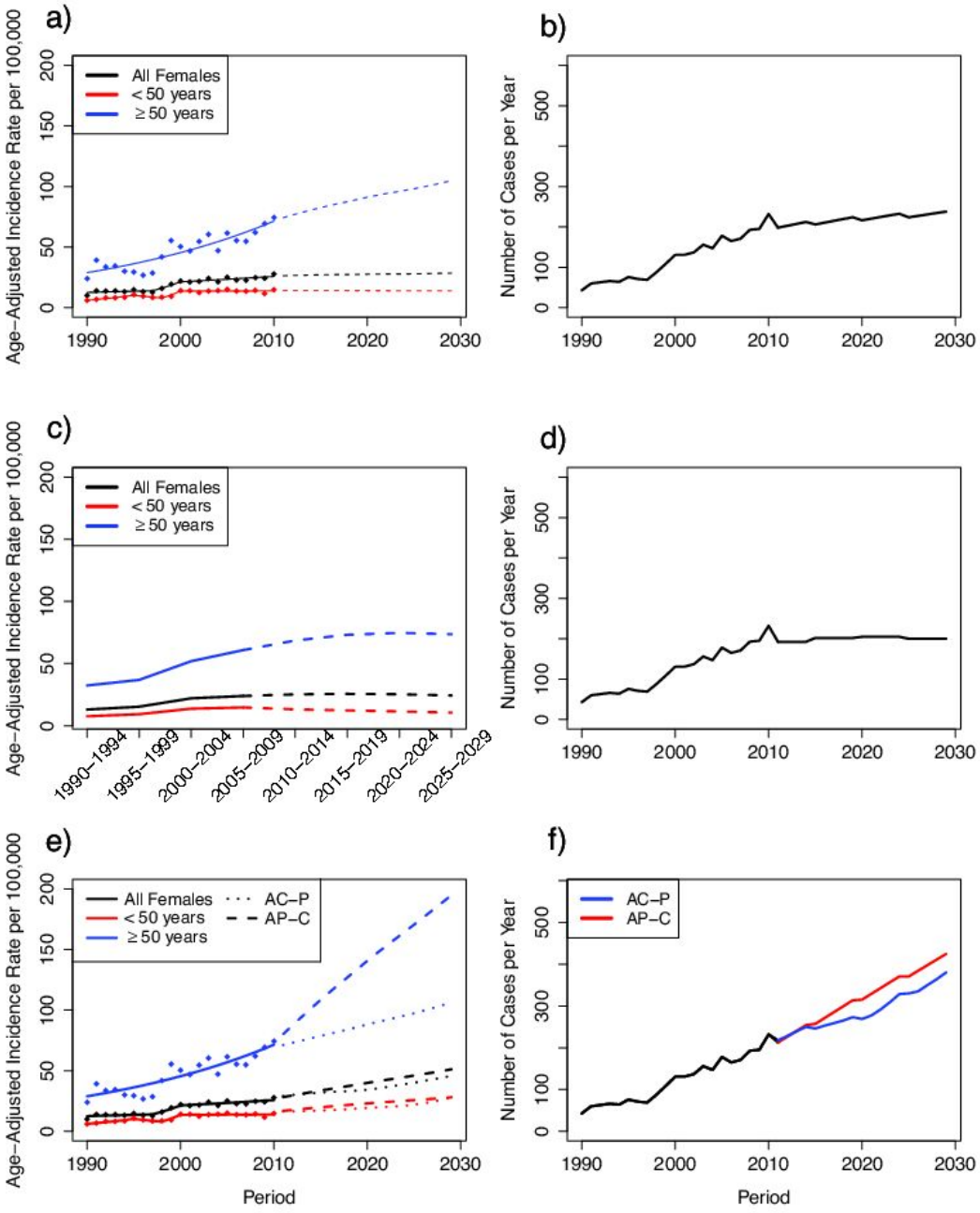


Figure 2.7: Rate and case projections for models included for comparison purposes. **a)** Rate projections for all females and by age group using the joinpoint model with all observed data included **b)** Case projection for all females from the joinpoint model with all observed data included **c)** Rate projections using the Nordpred model, but only including the last 10 years for the observed data **d)** Case projections for all females using the projection from the Nordpred model using only the recent trend from the last 10 years **e)** Rate projections from the APC models (AP-C and AC-P). The AC-P model used all available cohort information to base predictions off of **f)** Case projections for all females from the APC models in e).



CHAPTER 3
Biomarkers of effect from Cd exposure that may indicate early adverse health outcomes including carcinogenesis

ABSTRACT

Background: Cadmium (Cd) is a known human carcinogen, however, much of what is known about the biological effects of cadmium (Cd) comes from in vitro and animal studies. Evidence from these models has linked Cd to epigenetic changes, although there is limited evidence in human populations.

Objectives: The objective of this study is to determine the association between DNA methylation markers and renal biomarkers with Cd exposure, taking into account sex differences, to identify potential biomarkers of effect resulting from Cd exposure that may lead to carcinogenesis.

Methods: One hundred and sixty-nine residents from known exposure areas of Mae Sot, Thailand and one hundred residents from non-exposed areas nearby were surveyed in 2012. Demographic information and urine/blood samples were collected for measurement of urinary Cd (UCd), methylation of Cd-related markers (*DNMT3*, *MGMT*, LINE-1, *MT2A*) and renal marker levels (NAG, serum creatinine).

Results: UCd levels were about 7 times higher in the exposed compared to the non-exposed population (unexposed median: 1.04 μg/L, exposed median: 7.41 μg/L, p <0.001). *MGMT* methylation was significantly lower in the exposed population. UCd was significantly associated with all renal biomarkers. Sex differences were seen in exposure groups for *DNMT*, LINE-1 and *MGMT* methylation. Several markers were associated with both UCd and renal markers by

gender including *MT2A* and *DNMT* in females and LINE-1 and *DNMT* in males. Mediation analysis revealed LINE-1 methylation slightly mediated the effect of urinary Cd on B2MG in males.

Conclusions: Environmental Cd exposure induces DNA methylation changes that are dependent upon sex. DNA methylation changes are associated with altered levels of renal biomarkers, which may precede Cd-induced carcinogenesis

Abbreviations: Cd (cadmium), UCd (urinary cadmium), β2MG (beta 2-microglobulin), NAG (N-acetyl-beta-D-glucosaminidase), UCr (urinary creatinine), SCr (serum creatinine)

INTRODUCTION

Cadmium (Cd) is a biologically toxic transition metal found in low levels in nature and high levels from anthropogenic sources[1, 2]. Non-occupational exposures occur from tobacco smoke and ingestion of contaminated food and water and are associated with kidney damage including chronic renal diseases, osteoporosis, cardiovascular diseases, multiple cancers and osteomalacia[1-4] both in the US[5-7] and globally[8-10]. Cd is known to sequester in the liver and kidney and replace calcium in the bone[11, 12]. Occupational exposures first linked acute Cd exposure to lung, nasal, breast, kidney and pancreatic cancers[13-17] and the International Agency for Research on Cancer (IARC) has classified cadmium as a Group 1 human carcinogen[3, 4]. However, less is known about the carcinogenicity of Cd at low-to-moderate levels of exposure. The US Third National Health and Nutrition Examination Survey (NHANES) revealed an association of urinary Cd with cancer mortality over a 13-year follow-up period[18]. The associations found in men were for urinary Cd and mortality from lung and pancreas and with non-Hodgkin lymphoma cancers whereas for women the association was found for leukemia, lung, ovaries, and uterine cancers[18]. Cd exposure has been associated with breast and endometrial cancers in other populations as well[5, 19-21].

Although the association between UCd and cancer incidence/mortality is well-described, the mechanism by which Cd exerts its biologic effects is unclear. Although there is a lack of mechanistic studies in humans, in vitro and animal studies have provided evidence linking Cd to carcinogenesis (Table 3.1)[5, 22-25]. Cd weakly binds to DNA, is indirectly genotoxic and poorly mutagenic, indicating that it does not initiate disease through direct interaction with DNA[26, 27]. Cd is known to cause aberrant expression of a variety of genes including the oncogenes c-myc and c-jun[28], translation elongation factor-1 and 3[29], and AP-1 and MAP kinases, which are

involved in signaling pathways[30]. Specifically, Cd has been shown to affect DNA methylation, although the direction of effect is unclear. Many studies have implicated these epigenetic changes in the process of tumorigenesis[28-30 23]. Inactivation of tumor suppressor genes *p16* and *RASSF1A* in Cd-transformed malignant human prostate cells was found to be due to silencing of promoter regions from hypermethylation and correlated with DNMT3B overexpression[23]. Similarly, increased DNA methylation and DNMT activity were seen in human embryo fibroblasts after two months of Cd exposure[31] . However, Huang et al. found that DNA was hypomethylated in a leukemia K562 cell line only 24 and 48 hours after Cd exposure[32]. Cd exposure to normal human breast epithelial cell line MCF-10A for 40 weeks resulted in transformation of cells to basal-like breast carcinoma cells, with DNA hypomethylation and overexpression of c-myc and KRAS, as is commonly seen in aggressive breast cancers[22]. Cd was found to induce hypermethylation of caspase-8 in the liver of mice and *RASAL1* and *KLOTHO* in a Chinese population exposed to low levels in the environment[33, 34]. This evidence from previous studies suggests that changes in DNA methylation occur in response to Cd exposure.

It is difficult to assess initial molecular events in response to Cd exposure in cancer patients because cancer takes a long time to develop and there are likely numerous biological pathways that become dysregulated to initiate a cancerous phenotype. However, the low-dose exposure levels that have been associated with various cancers overlap with those shown to induce adverse kidney and bone defects (Table 3.2; Table 3.3). This offers an intermediate outcome with which to associate initial molecular events in response to Cd exposure that may fall along the pathway to cancer.

Renal dysfunction is well characterized in response to Cd exposure and is diagnosed with discrete clinical parameters measuring levels of specific renal biomarkers. Candidate gene

studies have identified DNA methylation markers associated specifically with Cd-induced kidney damage. Increased DNA methylation and DNMT activity were seen in human embryo fibroblasts after two months of Cd exposure[31]. Hypermethylation of tumor suppressor *RASAL1* and renal fibrosis inhibitor *KLOTHO*, was found to be strongly associated with both blood and urinary Cd levels and with renal fibrogenesis[34-36]. Additionally, Cd-induced gene-specific DNA hypermethylation has been reported for genes involved in cell cycle regulation, DNA repair, apoptosis and cell proliferation[33, 37-39]. Hypermethylation can prevent transcription of these important genes and promote initiation of disease. These studies show that Cd induces epigenetic alterations, which may fall along the molecular pathway of Cd exposure and renal dysfunction, and eventually, cancer.

Mae Sot District in the Tak Province of Thailand is located on the country's northwest border and is an area of high cadmium contamination due to a long standing zinc mine (Figure 3.1). Several subdistricts have been determined to have Cd levels in soil and rice exceeding Thai standards of 0.15mg/kg of soil and 0.043mg/kg of rice[40]. These standards reflect levels at which human exposure may result in adverse health effects[41]. A study conducted in Mae Sot in 2004 found a subset of residents that had urinary Cd (UCd) levels ≥5ug/g creatinine to also have irreversible renal dysfunction, indicated by clinical parameters of renal marker levels[12, 42]. UCd levels above this have been shown to induce preclinical renal dysfunction and have been deemed as a human exposure standard by the WHO[43]. However, a separate study found significant changes in renal biomarkers at much lower exposures, below UCd levels of 0.5ug/g creatinine, leading to concern about the effects of chronic, low-level exposures to Cd[44]. Sex differences in body burden and health outcomes in response to Cd exposure have been widely reported[45-47]. However, epigenetic changes in response to Cd exposure have not been considered within this

context, mostly because much of what we know about Cd induced epigenetic changes come from animal model or in vitro studies. Additionally, epigenetic changes associated with biomarkers of Cd-induced renal dysfunction are currently unknown. The objective of this study is to determine the sex-specific associations between UCd and methylation of selected genes and between these genes and renal biomarkers to clarify the role methylation may play initially in the pathway from Cd exposure to adverse renal outcomes and eventually, to cancer. These genes include *MT2A* (metallothionein 2A), *DNMT3* (DNA methyltransferase) and *MGMT* (DNA repair protein), all of which have been found to be altered in the presence of cadmium in animal model and in vitro studies[23, 39, 48, 49] . Additionally, methylation of LINE-1, a retrotransposon, was determined as a measure of genomic stability[22] . Renal biomarkers including NAG (N-acetyl-beta-D-glucosaminidase), and SCr (serum creatinine), were chosen based on their use as clinical indicators of renal dysfunction. Because Cd exerts different biological effects based upon sex, we hypothesize that methylation markers are associated with both UCd and renal biomarkers and differ by sex.

METHODS

Study Population

One hundred and sixty-nine subjects were selected from Mae Ku, Mae Tao and Phra Thad Padang subdistricts, areas in the Mae Sot District known to be contaminated with Cd due to their proximity to the Mae Tao creek[50]. These subjects were selected from a larger sample of 700 subjects who participated in a health impact survey conducted in 2007[51, 52]. The subjects in the Cd polluted area were at least 40 years old and were selected based on levels of β_2-MG and Cd in urine collected in 2007. The exposed population consisted of 47 men and 37 women with marked

renal tubular dysfunction (β_2-MG ≥ 1000 μg/g Cr) and high UCd (UCd ≥ 5 μg/g Cr), and 36 men and 48 women with no clear renal dysfunction (β_2-MG < 300 μg/g Cr) and low UCd exposure (UCd < 5 μg/g Cr). One hundred subjects aged 40 years and above were chosen from the Mae Kasa subdistrict, a non-contaminated area of Mae Sot. Trained health workers interviewed each participant about demographic characteristics, occupation, residency time, smoking status, and alcohol consumption. Medical history of hypertension, diabetes, and urinary stones was obtained from medical records at Mae Sot General Hospital. The research ethical committee of the Faculty of Medicine, Chiang Mai University, approved this study. (Approval No. 004/2012)

Sample Collection and Processing

Twenty-five mL of morning urine were collected from each participant in Cd-free polyethylene containers. Urine was immediately tested qualitatively for pH and chemical specific gravity (SG) using paper indicator strips (Ames test, Siemens, Germany) on site. One drop of 0.5N sodium hydroxide was added to one of the aliquots if it had a pH of ≤5 for prevention of further degradation of β2MG in an acid condition.

Ten milliliters of fasting venous blood was obtained by cubital venipuncture, and then divided into two Cd-free polyethylene tubes with and without EDTA. All samples were transported on ice to Mae Sot General Hospital's Laboratory within two hours of collection. Whole blood with EDTA and urine samples were stored at -20°C until analysis. Creatinine in sera obtained from blood in tubes without EDTA was determined by the Jaffe reaction method using an automated analyzer (Konelab 30, Thermo Electron Corporation, Finland) in Mae Sot hospital. White blood cell counts were measured using an automated hematology analyzer (HmX Hematology Analyzer, Beckman Coulter, USA). Frozen urine samples were transported with dry ice to

Kanazawa Medical University, Japan for urinary biomarker measurements. The concentration of urinary β2MG was measured via enzyme immunoassay using a latex agglutination immunoassay (Eiken Chemical, Japan); urinary NAG was measured via a colorimetric assay using the NAG test kit (Shionogi Pharmaceuticals, Japan); urinary creatinine (Cr) was measured via an enzyme assay using the Cica liquid-Stest kit (Kantokagaku Reagent Division, Ltd., Japan). Frozen whole blood with EDTA and urine samples were transported on dry ice to the University of Michigan for DNA collection and cadmium measurements.

Urinary Cadmium Measurement

Specific gravity of each sample was determined with a refractometer (PAL-10S, Atago Inc., USA). Urinary Cd was measured at the Michigan Department of Community Health. Briefly, urine samples were diluted 1:10 with a diluent composed of 2.0% nitric acid, internal standards and 0.05% Triton X, and Cd concentrations were determined using ICP-MS (DRCII , PerkinElmer, USA). The analytical accuracy using QMEQAS08U urinary standard reference material (Institut national de santé publique du Québec, INSPQ) was 101.1% (n=8) and all samples were above the analytical detection limit of 0.15 ug/L.

DNA Sample Preparation

DNA was extracted from 300uL of whole blood using the QiaAMP DNA Mini Kit (Qiagen, Valencia, CA, USA). DNA concentration and purity was quantified using the Nanodrop (ThermoScientific, Wayne, MI, USA) and DNA stored at -20C until methylation analyses. Sodium bisulfite modification was performed on 500ng of genomic DNA using the Epitect

Bisulfite Kit ((Qiagen, Valencia, CA, USA) according to the manufacturer's recommended protocol.

DNA Methylation Measurements

Methylation assays for 3 genes - *MT2A*, *DNMT3* and *MGMT*, were designed using PyroMark Assay Design 2.0 software. LINE-1 was measured using a previously published assay[53]. Bisulfite singleplex PCR amplification was performed using FastStart Taq Polymerase (Roche Diagnostics, Indiana, USA) with a forward and reverse primer concentration of 0.2 mM and 10ng/uL of bisulfite-converted DNA. Fifteen microliters of each PCR product was combined with the respective sequencing primer and methylation analysis by pyrosequencing was conducted for each assay as previously described[54]. Complete coverage of all samples for every methylation marker selected was not possible due to low quantity of total extracted DNA.

Statistical Analysis

Correlation between Cr-adjusted, SG-adjusted and unadjusted values was calculated using the Spearman method. All renal biomarkers and urinary cadmium levels were adjusted by specific gravity (SG) to adjust for all dilution-related variation of U-Cd[55]. Urinary markers were standardized to the median specific gravity of the control population with the following formula: $B_c = B[(1.017 - 1)/SG - 1)]$, where B_c is the SG-corrected urinary biomarker, B is the observed biomarker level, 1.017 is the median SG of the unexposed population, and SG is the specific gravity of the individual's urine sample [56]. Nonparametric Kruskal-Wallis tests were used to compare total exposed and unexposed populations and exposure populations by sex. Findings were determined to be statistically significant at p-value<0.05.

Linear regression models were performed to determine the associations of urinary cadmium with each methylation marker. Each marker was also tested as a predictor of each renal marker. Methylation changes were standardized to an interquartile range (IQR) increase in UCd levels. Methylation units were in 1% increments. Due to departures from normality, UCd, all methylation markers and all renal marker values were log-transformed for regression analysis and models were interpreted as percent increases in predictor associated with percent changes in outcome. Because age is associated with cadmium and renal marker levels, all models were adjusted by age. Sex is associated with cadmium and affects levels of renal markers; therefore, these models were stratified by sex.

Model Selection for Multivariable Models

To determine covariates to be included in final versions for each renal model, best-fit models were chosen separately for the exposed and unexposed population. A stepwise algorithm was first used to determine a multivariable model for each renal marker in each group, adding in variables in both backward and forward modes based on Akaike information criterion (AIC). Covariates in the full model included: urinary Cd, age, occupation, smoking status, gender, white blood cell count, hgb, BMI, alcohol use, and history of diabetes, hypertension and urinary stones. Variables included as predictors in each model by exposure group were then combined into one final model for each renal marker and run with the entire study population to accurately assess relevant predictors of renal marker levels. A priori predictors such as age and medical history were not significant in all models and were therefore, reintroduced into the final models since they are known to be associated with cadmium exposure and methylation. Final models were stratified by gender.

Causal Mediation Analysis

To test whether methylation markers may mediate the effects of urinary cadmium on renal biomarkers, a mediation analysis was conducted in four steps. Methylation markers were tested individually as each marker represents a separate potential mechanism of Cd toxicity. First, the association between urinary Cd and each renal biomarker was tested for significance (α=0.05) in a regression equation. Second, the association between urinary Cd and each methylation marker was tested for significance in a regression equation. Third, the association between each methylation marker and each renal marker was tested for significance with urinary Cd in the model. Any methylation marker that did not reach significance at the α=0.05 level was not considered as a mediator. Upon determination of potential mediators, methylation markers were tested for mediation effects using the final best-fit multivariable models. Using the mediation package in R, the average causal mediation effect (ACME), average direct effect (ADE) and the total effect was first calculated in the multivariable model without the methylation marker to determine the effect UCd had on the renal biomarker alone. Then, the methylation marker that was found to be a potential mediator was added to the model to determine its mediation effect (Figure 3.2). Significance of the mediation effect was measured by comparison to the sampling distribution created using the bootstrap method, a non-parametric method that uses resampling with replacement (n=2000) to create a sampling distribution of indirect effects.

R-statistical software was used for all analyses (R version 3.0.1)

RESULTS

Correlation among UCd measurements with different correction methods

Correlation of various correction methods was measured over the total population. Correlation coefficients for unadjusted UCd and SG-adjusted UCd and for SG-adjusted and Cr-adjusted UCd were similar (*rho*=0.95, $p<0.0001$). The correlation coefficient for unadjusted UCd and Cr-adjusted UCd was slightly lower (*rho*=0.87, $p<0.0001$).

Comparison of all markers between exposed and non-exposed groups

Median levels of UCd were nearly seven times higher in the exposed population than the unexposed population ($p <0.001$) (Table 3.4). The ranges of exposure between populations slightly overlap, however the exposed population has a much wider range with higher exposure levels than the unexposed population. There were significant differences in urinary markers by exposure group including UCd, specific gravity, β2MG, NAG, and Scr. Subjects in the Cd exposed population had significantly lower levels of SG than those in the exposed population. Levels of β2MG, NAG and SCr were all significantly higher, while *MGMT* methylation was significantly lower in the Cd exposed population as compared to the unexposed population.

Demographically, the proportion of regular smokers was highest in the Cd exposed population, of which 66% were male, whereas the unexposed population had similar proportions of never, former and regular smokers (Table 3.5). Half of the Cd exposed population had hypertension (HT) compared to forty percent of the unexposed population. The majority of the unexposed population was farmers (74%) compared to only fifty percent in the Cd exposed population.

Comparison of all markers by sex

Within the unexposed area, women had significantly higher levels of UCd as compared to men. In these women, levels of β2MG were significantly higher; however, levels of Scr, UCr and *DNMT* methylation were significantly lower than unexposed men (Table 3.6). In contrast, Cd-exposed women had significantly lower levels of UCd compared to Cd-exposed men, as well as lower levels of NAG, SCr, and LINE-1 methylation.

Exposed men had significantly higher levels of UCd, β2MG, NAG, SCr, and *DNMT* methylation, but lower MGMT methylation levels compared to unexposed men. Similar trends were seen in women across exposure group. Exposed women showed significantly higher levels of UCd, β2MG, NAG and SCr, but lower levels of *MGMT* methylation, compared to unexposed women (Table 3.6).

Associations between UCd and methylation markers

Regression analyses adjusted for age showed distinct associations of methylation markers to UCd (Figure 3.3). For the total population, *MGMT* methylation decreased by 16.6% with an interquartile (IQR) increase in UCd levels ($p<0.001$). In men, a 24.7% decrease in *MGMT* methylation was significantly associated with an IQR increase in UCd ($p=0.003$). There was also a significant increase in men of about 1% in LINE-1 methylation with an IQR increase in UCd ($p=0.04$). In women, *DNMT* methylation decreased by 8.5% and *MT2A* methylation decreased by 7.4% with an IQR increase in UCd ($p=0.05$, $p=0.03$, respectively). Methylation markers were not correlated with each other.

Associations between methylation markers and renal biomarkers

Renal biomarkers were significantly associated with UCd and several methylation markers (Table 3.7). All changes in renal biomarkers reflected a 1% increase in methylation. In the total population and populations separated by sex, all renal biomarkers were positively associated with an IQR increase in UCd. In the total population, a 1% increase in LINE-1 methylation was associated with a 15.1% increase in β2MG (p =0.01) while an increase in *MGMT* methylation was associated with a 0.8% decrease in β2MG (p =0.02). In males, LINE-1 methylation was associated with a 32.7% increase in β2MG (p <0.001), *MT2A* methylation was associated with a 1.4% increase in β2MG (p =0.01), while *MGMT* methylation was associated with a 0.9% decrease in β2MG (p =0.04). In females, a 1% increase in *DNMT* methylation was associated with a 1.5% decrease in β2MG (p =0.04).

LINE-1 methylation was associated with a 4% increase in NAG in the total population (p =0.01). In males, NAG increased by 5% and 0.3% with a 1% increase in LINE-1 and *MT2A* methylation, respectively (p =0.04; p =0.04, respectively). In females, *MT2A* methylation was associated with a 0.7% decrease in NAG (p=0.007).

In the total population, LINE-1 and *MGMT* methylation were associated with a 1.2% increase and 0.1% decrease in SCr, respectively (p =0.03; p =0.03). In males, LINE-1 methylation was associated with a 2.5% increase in SCr (p =0.001), while in females, *MGMT* methylation was associated with a 0.1% decrease in SCr (p =0.04).

Predictors of Renal Markers

Out of the final models for the unexposed population, UCd was fit as a predictor only for NAG. In the exposed population, UCd was a significant predictor for all renal markers

(Table 3.8). Predictors from exposed and unexposed models were combined for each renal marker (Table 3.9) These final three models were then stratified by gender (Table 3.10). All models included UCd, which was significantly associated with all renal markers in both males and females. In the B2MG model, white blood cell count (WBC) and presence of DM were significant predictors in females only. Presence of urinary stones is a significant predictor of B2MG in males only. Age was significantly associated with B2MG in both males and females. In the NAG model, WBC and DM were significant predictors in females only. Presence of urinary stones was a significant predictor of NAG in both males and females. In the SCr model, age had similar magnitudes of association in both males and females, but the association was only significant in males. There was a significant association of WBC with SCr in females only and DM was associated with SCr in males only.

Causal Mediation combining methylation markers with best-fit markers

Causal mediation analysis was used to establish if methylation of these markers fall on the causal pathway between Cd exposure and renal dysfunction. UCd was significantly associated with all renal biomarkers in the total population. However, the only methylation marker that was significantly associated with UCd was *MGMT*. Testing this marker for significance in the presence of UCd with each renal biomarker revealed no significant association of MGMT with any renal biomarker. Therefore, no potential mediators were found to be tested in the whole population. This analysis was repeated stratifying by gender. LINE-1 methylation was found to be a potential mediator for the effects of UCd on B2MG and SCr levels in men. *MT2A* was a potential mediator of the effect of UCd on NAG levels in women.

Mediation analyses testing the indirect effect of the potential mediators are shown in Figure 3.4. For mediation of LINE-1 with B2MG, the estimated ACME is statistically different from zero as are the estimated average direct (ADE) and total effects (Fig 3.4a). Therefore, we find that LINE-1 methylation slightly mediated the effect of urinary Cd on B2MG in males. This effect, however, was small with a point estimate of 0.07 and *p*-value of 0.04, while the effect of urinary Cd on B2MG is comparatively larger (β=0.93; *p*-value<0.001). For mediation of LINE-1 with SCr, the estimated ACME is not statistically significant, but the estimated ADE and total effects are, although the effect estimates are small (β=0.056, *p*-value<0.001; β=0.063, p-value<0.001, respectively). Therefore, we find that LINE-1 methylation is not a mediator of the effect of urinary Cd on SCr in males. Mediation testing for MT2A with NAG revealed that only the ADE and total effects are significant (Fig 3.4b). Therefore, MT2A is not a mediator of the effect of urinary Cd on NAG in females.

DISCUSSION

The toxic effects of environmental Cd exposure have been studied as early as the 1960s with the outbreak of itai-itai disease in Japan. Since then, Cd has been shown to have a variety of adverse effects on human health, the most notable being proximal tubular renal dysfunction and bone disorders, such as osteomalacia and osteoporosis[9, 12, 44, 50, 57, 58]. However, the pathway by which Cd induces these health outcomes is unknown.

Our study shows the expected differences in UCd and renal biomarkers between groups with high and little to no environmental Cd exposure. The exposed group has significantly higher levels of UCd and renal markers compared to the unexposed group. Although our unexposed population had higher UCd levels (mean: 1.04 μg/L) , compared to the US population (mean:

0.25 μg/L)[59-61], this group provides a good control to measure Cd exposure in demographically similar Thai populations.

We report sex-specific differences in UCd and renal biomarkers based on exposure group. Women living in the unexposed area had higher levels of UCd, but lower levels of all renal biomarkers, compared to men in the same area. Although the area is defined as unexposed, there is likely some source of Cd exposure at very low concentrations. Women, especially those with low iron stores, tend have higher body burdens of Cd as compared to males, regardless of exposure levels[45, 46]. Within those living in high Cd-exposed areas, women had lower levels of UCd and renal biomarkers as compared to men living in the same area. This is potentially due to the large proportion of male smokers living in Cd-exposed areas, since cadmium is found in tobacco.

We also report differences in methylation markers by exposure group and sex and identified sex-specific associations between an IQR increase in UCd and DNA methylation. LINE-1 methylation was higher in exposed males compared to exposed females and we found that LINE-1 methylation increased with higher levels of UCd in men. These results corroborate literature that shows that, in general, men tend to have higher methylation of LINE-1 compared to women, even in the presence of cadmium exposure[62]. In the total population we found significantly lower levels of *MGMT* methylation in the exposed group, indicating Cd exposure may induce hypomethylation of this DNA repair gene. By sex, we observed *MGMT* methylation to be significantly hypomethylated in exposed males and females compared to unexposed males and females, respectively. We found an inverse relationship between Cd exposure and *MGMT* methylation in the total population and men specifically, which supports the hypothesis that Cd-induced oxidative stress can lead to DNA strand breaks, leading to activation of *MGMT* as it is a

DNA repair gene that targets damage induced by oxidative stress[63, 64]. As Cd exposure increases, methylation of this gene decreases; potentially making it available for transcription. *DNMT* was significantly hypermethylated in women compared to men within the unexposed group and we found that *DNMT* methylation decreased with an IQR increase in UCd in women. Our findings support in vitro studies that found hypomethylation of this gene and consequent overexpression[23, 65] as well as a previous human study conducted on environmentally-exposed Argentinian women[49]. In women, an increase in Cd exposure is associated with a decrease in *MT2A* methylation. *MT2A* encodes metallothionein, a protein that binds free Cd and sequesters it in the liver and kidney. Women tend to have higher body burden of Cd compared to men and this points to a potential mechanism of increased MT2A expression by which women may sequester larger amounts for longer periods of time.

Several methylation markers associated with UCd were also found to be associated with renal biomarkers. These renal biomarkers are all significantly associated with UCd and are used as clinical parameters to diagnose renal tubular dysfunction, a well-known health outcome of Cd exposure. β2MG is the most common marker used to detect renal dysfunction. It is found on the surface of white blood cells and is readily filtered through the glomerulus and reabsorbed through the proximal tubules. In renal tubular dysfunction, β2MG is not reabsorbed and is excreted in the urine in high amounts. NAG is a lysosomal enzyme found in high concentrations in renal proximal tubular cells. Upon renal tubular damage, this enzyme is released and excreted in high amounts in the urine. High levels of creatinine in the serum are indicative of defects in glomerular filtration rate, a symptom of renal dysfunction.

In men, LINE-1 methylation was associated with UCd and with all three of these renal markers, β2MG, NAG, and SCr, indicating this retroposon may play a role in the mechanism

between Cd exposure and renal dysfunction. *DNMT* methylation was significantly higher in exposed males compared to unexposed males, but the lack of association with UCd and any renal markers among males suggests that *DNMT* methylation in men may be related to factors other than Cd exposure. *MT2A* methylation was not associated with UCd in males, but it was associated with β2MG and NAG, indicating it might be useful as a marker of renal dysfunction only. Conversely, *MGMT* methylation was associated with UCd and β2MG in males, suggesting it may fall along the pathway of cadmium toxicity. In women, *MT2A* methylation was inversely associated with NAG, while *DNMT* methylation was inversely associated with both β2MG and SCr. Methylation of both of these genes decreased with increasing levels of UCd, and may be related to renal tubular damage in women. *MGMT* methylation was also associated with SCr in women only, suggesting it may be related to renal glomerular alteration.

The sex differences in methylation markers may be indicative of the sex-specific differences in pathways of Cd toxicity, which is best demonstrated by the itai-itai disease outbreak from severe Cd poisoning in Japan. Out of 195 victims officially diagnosed with itai-itai disease between 1967 and 2008, only three were male[66]. Itai-itai disease is characterized by osteomalacia and renal tubular dysfunction, which may be exacerbated in women through physiological changes such as pregnancy and menopause[67, 68]. Women have a higher probability of developing iron deficiency and, since the method of uptake of Cd is similar to iron, they accumulate higher body burdens[69]. Although specific mechanisms are unknown, female risk factors such as increased gut absorption, skeletal changes during pregnancy and menopause, and variation in estrogen levels may play a role in exacerbating the toxic effects of Cd exposure[46, 68]. Additionally, the differences seen in smoking status by sex may also induce gender differences.

Mediation analysis in this study showed that only LINE-1 was a causal mediator of the effect of Cd on B2MG, although the effect estimate was small. A limitation of causal mediation analysis is that it tests for direct causality. Knowing that biological mechanisms rarely involve only one marker, it is necessary to look at multiple markers involved in the same or similar mechanisms and assess them as multiple mediators. Previous literature has shown epigenetic aberration of genes such as *hMSH2*, *ERCC1*, *XRCC1*, *hOGG1*, in response to Cd and our study adds to this by the significant differences seen in the methylation of markers by exposure group and gender. A more likely scenario to consider when assessing epigenetic changes in response to Cd exposure would be to assess multiple mediators along the pathway of Cd exposure and adverse health outcomes (Figure 3.5).

Specific gravity (SG) was used to adjust urinary biomarker levels for dilution effects since creatinine is a clinical parameter of renal dysfunction and has been shown to be inversely correlated with Cd measurements in the soil[70-72.] To avoid potential bias, individual urinary biomarkers were standardized to the median SG of the unexposed population only.

The association of methylation markers with UCd and renal biomarkers implicates aberrant methylation of these genes in the pathogenesis of Cd exposure to renal dysfunction and potentially to cancer. These relationships are sex-specific, and suggest that the pathway of exposure to disease may differ by sex. A limitation of this cross-sectional study is that it was not able to assess cancer as a health outcome. This limitation was addressed by using renal dysfunction as a proxy since Cd exposure levels associated with kidney effects and various cancers are similar (Table 3.2; Table 3.3). Other limitations include lack of dietary information and measurements of other metals that may confound the results presented here.

CONCLUSIONS

Much of what is known about the mechanism behind Cd toxicity is from animal models or in vitro cellular experiments. Our study utilizes a population with a wide range of relevant Cd levels and extensively characterized methylation biomarkers. To our knowledge, this is the first study to examine such ranges within the context of DNA methylation in a population-based study. As with many epigenetic epidemiology studies, we were limited to measuring methylation in circulating cells from blood. As such, the conclusions we can draw regarding the direct effects of Cd on epigenetic may not be extrapolated to relevant tissues, such as the kidneys and liver. However, studies such as these are a logical first step to identifying the mechanisms of Cd toxicity, and may identify useful biomarkers that indicate underlying health consequences of Cd exposure.

References

[1]ATSDR (Agency for Toxic Substances and Disease Registry), DRAFT TOXICOLOGICAL PROFILE FOR CADMIUM, 2008.

[2]WHO (World Health Organization), EXPOSURE TO CADMIUM: A MAJOR PUBLIC HEALTH CONCERN, 2010.

[3]IARC (International Agency for Research on Cancer), International Agency for Research on Cancer Monographs, Beryllium, Cadmium, Mercury, and Exposures in the Glass Industry, 1993.

[4]IARC (International Agency for Research on Cancer), CADMIUM AND CADMIUM COMPOUNDS, 2012.

[5]Gallagher CM, Chen JJ, Kovach JS. Environmental cadmium and breast cancer risk. *Aging* 2010;**2**: 804-14.

[6]Navas-Acien A, Tellez-Plaza M, Guallar E, Muntner P, Silbergeld E, Jaar B, Weaver V. Blood cadmium and lead and chronic kidney disease in US adults: a joint analysis. *American journal of epidemiology* 2009;**170**: 1156-64.

[7]Tellez-Plaza M, Navas-Acien A, Menke A, Crainiceanu CM, Pastor-Barriuso R, Guallar E. Cadmium Exposure and All Cause and Cardiovascular Mortality in the US General Population. *Environmental health perspectives* 2012.

[8]Haswell-Elkins M, Imray P, Satarug S, Moore MR, O'Dea K. Urinary excretion of cadmium among Torres Strait Islanders (Australia) at risk of elevated dietary exposure through traditional foods. *Journal of exposure science & environmental epidemiology* 2007;**17**: 372-7.

[9]Swaddiwudhipong W, Limpatanachote P, Nishijo M, Honda R, Mahasakpan P, Krintratun S. Cadmium-exposed population in Mae Sot district, Tak province: 3. Associations between urinary cadmium and renal dysfunction, hypertension, diabetes, and urinary stones. *J Med Assoc Thai* 2010;**93**: 231-8.

[10]Weaver VM, Kim NS, Lee BK, Parsons PJ, Spector J, Fadrowski J, Jaar BG, Steuerwald AJ, Todd AC, Simon D, Schwartz BS. Differences in urine cadmium associations with kidney outcomes based on serum creatinine and cystatin C. *Environmental research* 2011;**111**: 1236-42.

[11]Liang Y, Lei L, Nilsson J, Li H, Nordberg M, Bernard A, Nordberg GF, Bergdahl IA, Jin T. Renal function after reduction in cadmium exposure: an 8-year follow-up of residents in cadmium-polluted areas. *Environmental health perspectives* 2012;**120**: 223-8.

[12]Limpatanachote P, Swaddiwudhipong W, Nishijo M, Honda R, Mahasakpan P, Nambunmee K, Ruangyuttikarn W. Cadmium-exposed population in Mae Sot District, Tak Province: 4 bone mineral density in persons with high cadmium exposure. *J Med Assoc Thai* 2010;**93**: 1451-7.

[13]Jarup L, Bellander T, Hogstedt C, Spang G. Mortality and cancer incidence in Swedish battery workers exposed to cadmium and nickel. *Occupational and environmental medicine* 1998;**55**: 755-9.

[14]Park RM, Stayner LT, Petersen MR, Finley-Couch M, Hornung R, Rice C. Cadmium and lung cancer mortality accounting for simultaneous arsenic exposure. *Occupational and environmental medicine* 2012;**69**: 303-9.

[15]Pollan M, Gustavsson P. High-risk occupations for breast cancer in the Swedish female working population. *American journal of public health* 1999;**89**: 875-81.

[16]Il'yasova D, Schwartz GG. Cadmium and renal cancer. *Toxicology and applied pharmacology* 2005;**207**: 179-86.

[17]Schwartz GG, Reis IM. Is cadmium a cause of human pancreatic cancer? *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2000;**9**: 139-45.

[18]Adams SV, Passarelli MN, Newcomb PA. Cadmium exposure and cancer mortality in the Third National Health and Nutrition Examination Survey cohort. *Occupational and environmental medicine* 2012;**69**: 153-6.

[19]McElroy JA, Shafer MM, Trentham-Dietz A, Hampton JM, Newcomb PA. Cadmium exposure and breast cancer risk. *Journal of the National Cancer Institute* 2006;**98**: 869-73.

[20]Julin B, Wolk A, Bergkvist L, Bottai M, Akesson A. Dietary cadmium exposure and risk of postmenopausal breast cancer: a population-based prospective cohort study. *Cancer research* 2012;**72**: 1459-66.

[21]Nagata C, Nagao Y, Nakamura K, Wada K, Tamai Y, Tsuji M, Yamamoto S, Kashiki Y. Cadmium exposure and the risk of breast cancer in Japanese women. *Breast cancer research and treatment* 2013;**138**: 235-9.

[22]Benbrahim-Tallaa L, Tokar EJ, Diwan BA, Dill AL, Coppin JF, Waalkes MP. Cadmium malignantly transforms normal human breast epithelial cells into a basal-like phenotype. *Environmental health perspectives* 2009;**117**: 1847-52.

[23]Benbrahim-Tallaa L, Waterland RA, Dill AL, Webber MM, Waalkes MP. Tumor suppressor gene inactivation during cadmium-induced malignant transformation of human prostate cells correlates with overexpression of de novo DNA methyltransferase. *Environmental health perspectives* 2007;**115**: 1454-9.

[24]Qu W, Diwan BA, Reece JM, Bortner CD, Pi J, Liu J, Waalkes MP. Cadmium-induced malignant transformation in rat liver cells: role of aberrant oncogene expression and minimal role of oxidative stress. *International journal of cancer Journal international du cancer* 2005;**114**: 346-55.

[25]Rossman TG, Roy NK, Lin WC. Is cadmium genotoxic? *IARC scientific publications* 1992: 367-75.

[26]Waalkes MP. Cadmium carcinogenesis in review. *Journal of inorganic biochemistry* 2000;**79**: 241-4.

[27]Hartwig A. Role of DNA repair inhibition in lead- and cadmium-induced genotoxicity: a review. *Environmental health perspectives* 1994;**102 Suppl 3**: 45-50.

[28]Abshire MK, Buzard GS, Shiraishi N, Waalkes MP. Induction of c-myc and c-jun proto-oncogene expression in rat L6 myoblasts by cadmium is inhibited by zinc preinduction of the metallothionein gene. *Journal of toxicology and environmental health* 1996;**48**: 359-77.

[29]Joseph P, Lei YX, Whong WZ, Ong TM. Oncogenic potential of mouse translation elongation factor-1 delta, a novel cadmium-responsive proto-oncogene. *The Journal of biological chemistry* 2002;**277**: 6131-6.

[30]Huang C, Zhang Q, Li J, Shi X, Castranova V, Ju G, Costa M, Dong Z. Involvement of Erks activation in cadmium-induced AP-1 transactivation in vitro and in vivo. *Molecular and cellular biochemistry* 2001;**222**: 141-7.

[31]Jiang G, Xu L, Song S, Zhu C, Wu Q, Zhang L, Wu L. Effects of long-term low-dose cadmium exposure on genomic DNA methylation in human embryo lung fibroblast cells. *Toxicology* 2008;**244**: 49-55.

[32]Huang D, Zhang Y, Qi Y, Chen C, Ji W. Global DNA hypomethylation, rather than reactive oxygen species (ROS), a potential facilitator of cadmium-stimulated K562 cell proliferation. *Toxicology letters* 2008;**179**: 43-7.

[33]Wang B, Li Y, Tan Y, Miao X, Liu XD, Shao C, Yang XH, Turdi S, Ma LJ, Ren J, Cai L. Low-dose Cd induces hepatic gene hypermethylation, along with the persistent reduction of cell death and increase of cell proliferation in rats and mice. *PloS one* 2012;**7**: e33853.

[34]Zhang C, Liang Y, Lei L, Zhu G, Chen X, Jin T, Wu Q. Hypermethylations of RASAL1 and KLOTHO is associated with renal dysfunction in a Chinese population environmentally exposed to cadmium. *Toxicology and applied pharmacology* 2013;**271**: 78-85.

[35]Bechtel W, McGoohan S, Zeisberg EM, Muller GA, Kalbacher H, Salant DJ, Muller CA, Kalluri R, Zeisberg M. Methylation determines fibroblast activation and fibrogenesis in the kidney. *Nature medicine* 2010;**16**: 544-50.

[36]Sun CY, Chang SC, Wu MS. Suppression of Klotho expression by protein-bound uremic toxins is associated with increased DNA methyltransferase expression and DNA hypermethylation. *Kidney international* 2012;**81**: 640-50.

[37]Zhang J, Fu Y, Li J, Wang J, He B, Xu S. Effects of subchronic cadmium poisoning on DNA methylation in hens. *Environmental toxicology and pharmacology* 2009;**27**: 345-9.

[38]Zhu H, Li K, Liang J, Zhang J, Wu Q. Changes in the levels of DNA methylation in testis and liver of SD rats neonatally exposed to 5-aza-2'-deoxycytidine and cadmium. *Journal of applied toxicology : JAT* 2011;**31**: 484-95.

[39]Zhou ZH, Lei YX, Wang CX. Analysis of aberrant methylation in DNA repair genes during malignant transformation of human bronchial epithelial cells induced by cadmium. *Toxicological sciences : an official journal of the Society of Toxicology* 2012;**125**: 412-7.

[40]Zarcinas BA, Pongsakul P, McLaughlin MJ, Cozens G. Heavy metals in soils and crops in Southeast Asia. 2. Thailand. *Environ Geochem Health* 2004;**26**: 359-71.

[41]Pongsakul P, Attajarusit S. Assessment of heavy metal contaminations in soils. . *Thai Journal of Soils and Fertilizers* 1999: 71-82.

[42]Swaddiwudhipong W, Limpatanachote P, Mahasakpan P, Krintratun S, Padungtod C. Cadmium-exposed population in Mae Sot District, Tak Province: 1. Prevalence of high urinary cadmium levels in the adults. *J Med Assoc Thai* 2007;**90**: 143-8.

[43]WHO JEFCA, Evaluation of certain food additives and contaminants. Seventy-third meeting of the Joint FAO/WHO Expert Committee on Food Additives, 2010.

[44]Satarug S, Moore MR. Adverse health effects of chronic exposure to low-level cadmium in foodstuffs and cigarette smoke. *Environmental health perspectives* 2004;**112**: 1099-103.

[45]Satarug S, Ujjin P, Vanavanitkun Y, Baker JR, Moore MR. Influence of body iron store status and cigarette smoking on cadmium body burden of healthy Thai women and men. *Toxicology letters* 2004;**148**: 177-85.

[46]Vahter M, Berglund M, Akesson A, Liden C. Metals and women's health. *Environmental research* 2002;**88**: 145-55.

[47]Nishijo M, Satarug S, Honda R, Tsuritani I, Aoshima K. The gender differences in health effects of environmental cadmium exposure and potential mechanisms. *Molecular and cellular biochemistry* 2004;**255**: 87-92.

[48]Arita A, Costa M. Epigenetics in metal carcinogenesis: nickel, arsenic, chromium and cadmium. *Metallomics : integrated biometal science* 2009;**1**: 222-8.

[49]Hossain MB, Vahter M, Concha G, Broberg K. Low-level environmental cadmium exposure is associated with DNA hypomethylation in Argentinean women. *Environmental health perspectives* 2012;**120**: 879-84.

[50]Simmons RW, Pongsakul P, Saiyasitpanich D, Klinphoklap S. Elevated levels of cadmium and zinc in paddy soils and elevated levels of cadmium in rice grain downstream of a zinc mineralized area in Thailand: implications for public health. *Environ Geochem Health* 2005;**27**: 501-11.

[51]Ruangyuttikarn W, Panyamoon A, Nambunmee K, Honda R, Swaddiwudhipong W, Nishijo M. Use of the kidney injury molecule-1 as a biomarker for early detection of renal tubular dysfunction in a population chronically exposed to cadmium in the environment. *SpringerPlus* 2013;**2**: 533.

[52]Nambunmee K, Honda R, Nishijo M, Swaddiwudhipong W, Nakagawa H, Ruangyuttikarn W. Bone resorption acceleration and calcium reabsorption impairment in a Thai population with high cadmium exposure. *Toxicology mechanisms and methods* 2010;**20**: 7-13.

[53]Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 2004;**32**: e38.

[54]Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nature protocols* 2007;**2**: 2265-75.

[55]Suwazono Y, Akesson A, Alfven T, Jarup L, Vahter M. Creatinine versus specific gravity-adjusted urinary cadmium concentrations. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals* 2005;**10**: 117-26.

[56]Meeker JD, Hu H, Cantonwine DE, Lamadrid-Figueroa H, Calafat AM, Ettinger AS, Hernandez-Avila M, Loch-Caruso R, Tellez-Rojo MM. Urinary phthalate metabolites in relation to preterm birth in Mexico city. *Environmental health perspectives* 2009;**117**: 1587-92.

[57]Noonan CW, Sarasua SM, Campagna D, Kathman SJ, Lybarger JA, Mueller PW. Effects of exposure to low levels of environmental cadmium on renal biomarkers. *Environmental health perspectives* 2002;**110**: 151-5.

[58]Inaba T, Kobayashi E, Suwazono Y, Uetani M, Oishi M, Nakagawa H, Nogawa K. Estimation of cumulative cadmium intake causing Itai-itai disease. *Toxicology letters* 2005;**159**: 192-201.

[59]Centers for Disease Control (CDC), National Health and Nutrition Examination Survey data, protocols, and analytic guidelines. . U.S. Department of Health and Human Services, 2011.

[60]Centers for Disease Control (CDC), Fourth National Report on Human Exposure to Environmental Chemicals: Centers for Disease Control and Prevention. Dept. of Health and Human Services, 2009.

[61]Centers for Disease Control (CDC), Fourth National Report on Human Exposure to Environmental Chemicals: Updated Tables. Centers for Disease Control and Prevention. Department of Health and Human Services, 2013.

[62]Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, Vishwanatha JK, Santella RM, Morabia A. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics : official journal of the DNA Methylation Society* 2011;**6**: 623-9.

[63]Stohs SJ, Bagchi D. Oxidative mechanisms in the toxicity of metal ions. *Free radical biology & medicine* 1995;**18**: 321-36.

[64]Dally H, Hartwig A. Induction and repair inhibition of oxidative DNA damage by nickel(II) and cadmium(II) in mammalian cells. *Carcinogenesis* 1997;**18**: 1021-6.

[65]Takiguchi M, Achanzar WE, Qu W, Li G, Waalkes MP. Effects of cadmium on DNA-(Cytosine-5) methyltransferase activity and DNA methylation status during cadmium-induced cellular transformation. *Experimental cell research* 2003;**286**: 355-65.

[66]Matsunami J. Kadomiumu hiagai hyakunen: Kaiko to tenbo. [in Japanese; A hundred years of cadmium poisoning: recollection and prospects. *Katsura Shobo, Toyama* 2010.

[67]Horiguchi H, Teranishi H, Niiya K, Aoshima K, Katoh T, Sakuragawa N, Kasuya M. Hypoproduction of erythropoietin contributes to anemia in chronic cadmium intoxication: clinical study on Itai-itai disease in Japan. *Archives of toxicology* 1994;**68**: 632-6.

[68]Kakei M, Sakae T, Yoshikawa M. Combined effects of estrogen deficiency and cadmium exposure on calcified hard tissues: animal model relating to itai-itai disease in postmenopausal women. *Proceedings of the Japan Academy Series B, Physical and biological sciences* 2013;**89**: 340-7.

[69]Gunshin H, Mackenzie B, Berger UV, Gunshin Y, Romero MF, Boron WF, Nussberger S, Gollan JL, Hediger MA. Cloning and characterization of a mammalian proton-coupled metal-ion transporter. *Nature* 1997;**388**: 482-8.

[70]Akesson A, Lundh T, Vahter M, Bjellerup P, Lidfeldt J, Nerbrand C, Samsioe G, Stromberg U, Skerfving S. Tubular and glomerular kidney effects in Swedish women with low environmental cadmium exposure. *Environmental health perspectives* 2005;**113**: 1627-31.

[71]Barbier O, Dauby A, Jacquillet G, Tauc M, Poujeol P, Cougnon M. Zinc and cadmium interactions in a renal cell line derived from rabbit proximal tubule. *Nephron Physiology* 2005;**99**: p74-84.

[72]Staessen JA, Lauwerys RR, Ide G, Roels HA, Vyncke G, Amery A. Renal function and historical environmental cadmium pollution from zinc smelters. *Lancet* 1994;**343**: 1523-7.

[73]Nawrot T, Plusquin M, Hogervorst J, Roels HA, Celis H, Thijs L, Vangronsveld J, Van Hecke E, Staessen JA. Environmental exposure to cadmium and risk of cancer: a prospective population-based study. *The lancet oncology* 2006;**7**: 119-26.

[74]Kriegel AM, Soliman AS, Zhang Q, El-Ghawalby N, Ezzat F, Soultan A, Abdel-Wahab M, Fathy O, Ebidi G, Bassiouni N, Hamilton SR, Abbruzzese JL, et al. Serum cadmium levels in pancreatic cancer patients from the East Nile Delta region of Egypt. *Environmental health perspectives* 2006;**114**: 113-9.

[75]Zeng X, Jin T, Jiang X, Kong Q, Ye T, Nordberg GF. Effects on the prostate of environmental cadmium exposure--a cross-sectional population study in China. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* 2004;**17**: 559-65.

[76]Vinceti M, Venturelli M, Sighinolfi C, Trerotoli P, Bonvicini F, Ferrari A, Bianchi G, Serio G, Bergomi M, Vivoli G. Case-control study of toenail cadmium and prostate cancer risk in Italy. *The Science of the total environment* 2007;**373**: 77-81.

[77]van Wijngaarden E, Singer EA, Palapattu GS. Prostate-specific antigen levels in relation to cadmium exposure and zinc intake: results from the 2001-2002 National Health and Nutrition Examination Survey. *The Prostate* 2008;**68**: 122-8.

[78]Kellen E, Zeegers MP, Hond ED, Buntinx F. Blood cadmium may be associated with bladder carcinogenesis: the Belgian case-control study on bladder cancer. *Cancer detection and prevention* 2007;**31**: 77-82.

[79]Satarug S, Garrett SH, Sens MA, Sens DA. Cadmium, environmental exposure, and health outcomes. *Environmental health perspectives* 2010;**118**: 182-90.

[80]Akesson A, Bjellerup P, Lundh T, Lidfeldt J, Nerbrand C, Samsioe G, Skerfving S, Vahter M. Cadmium-induced effects on bone in a population-based study of women. *Environmental health perspectives* 2006;**114**: 830-4.

[81]Satarug S, Nishijo M, Ujjin P, Vanavanitkun Y, Moore MR. Cadmium-induced nephropathy in the development of high blood pressure. *Toxicology letters* 2005;**157**: 57-68.

[82]Teeyakasem W, Nishijo M, Honda R, Satarug S, Swaddiwudhipong W, Ruangyuttikarn W. Monitoring of cadmium toxicity in a Thai population with high-level environmental exposure. *Toxicology letters* 2007;**169**: 185-95.

[83]Gallagher CM, Kovach JS, Meliker JR. Urinary cadmium and osteoporosis in U.S. Women >or= 50 years of age: NHANES 1988-1994 and 1999-2004. *Environmental health perspectives* 2008;**116**: 1338-43.

[84]Schutte R, Nawrot TS, Richart T, Thijs L, Vanderschueren D, Kuznetsova T, Van Hecke E, Roels HA, Staessen JA. Bone resorption and environmental exposure to cadmium in women: a population study. *Environmental health perspectives* 2008;**116**: 777-83.

[85]Wu X, Liang Y, Jin T, Ye T, Kong Q, Wang Z, Lei L, Bergdahl IA, Nordberg GF. Renal effects evolution in a Chinese population after reduction of cadmium exposure in rice. *Environmental research* 2008;**108**: 233-8.

[86]Thomas LD, Hodgson S, Nieuwenhuijsen M, Jarup L. Early kidney damage in a population exposed to cadmium and other heavy metals. *Environmental health perspectives* 2009;**117**: 181-4.

Table 3.1. Key Molecular Studies in Determining Potential Mechanisms of Actions of Cadmium in Carcinogenesis

Model	Dose	Duration	Results	References
TRL 1215 rat liver cells	1.0 and 2.5uM	1 week and 10 weeks	Inhibited *DNMT* activity and global DNA hypomethylation at 1 week, enhanced *DNMT* activity and global DNA hypermethylation at 10 weeks	Takiguchi et al., 2003[65]
Human embryo lung fibroblasts	1.2 and 1.5uM	2 months	Concentration-dependent hypermethylation of DNA. Correlated with increased *DNMT* activity and overexpression of *DNMT1, DNMT3a* and *DNMT3b*	Jiang et al., 2008[31]
Chronic myelogenous leukemia K562 cells	2.0uM	24 & 48 hours	Global hypomethylation of DNA	Huang et al, 2008[32]
Human breast epithelial cell line MCF-10A	2.5 µM	40 weeks	Malignant transformation to basal-like breast carcinoma cells. Global DNA hypomethylation, *c-myc* and *k-ras* overexpression, reduced BRCA1 expression, increased CK5 and p63 expression	Benbrahim-Tallaa et al., 2009[22]
Human prostate epithelial cells	10uM	10 weeks	Malignant transformation of cells, global DNA hypomethylation , overexpression of DNMT3b, promoter hypermethylation and reduced expression of *RASSF1A* and *p16* tumor suppressor genes	Benbrahim-Tallaa et al., 2007[23]

Table 3.2. Exposure levels associated with cancer.	
Cancer/study population, reference	**Exposure/risk estimate**
Lung, Belgium, n = 994, 15-year observation[73]	Hazard ratios of 1.7, 2.6, and 1.6 were attributed to a 2-fold increase in body burden, living in high-exposure area, and a 2-fold increase in soil cadmium, respectively.
Pancreas, Egypt, n = 31 cases, 52 controls[74]	ORs of 1.12 and 3.25 were attributed to elevated serum cadmium and farming occupation, respectively.
Breast, United States, n = 246 cases, 254 controls[19]	OR of 2.3 when comparing urinary cadmium < 0.26 versus ≥ 0.58 μg/g creatinine
Endometrium, Sweden, n = 30,210, 16-year observation[70]	OR of 2.9 was attributed to cadmium intake > 15 μg/day.
Prostate, China; n = 297[75]	Dose response between body burden and abnormal serum PSA levels
Prostate, Italy, n = 45 cases, 58 controls[76]	OR of 4.7 when comparing nail cadmium content in the lowest versus the highest quartile
Prostate, United States, n = 422[77]	An increase of urinary cadmium to 1 μg/g creatinine associated with a 35% increase in serum PSA
Urinary bladder, Belgium, n = 172 cases, 395 controls[78]	OR of 5.7 when comparing blood cadmium in the lowest versus the highest tertile
Abbreviations: OR, odds ratio; PSA, prostate-specific antigen.	

Table adapted from Satarug, et al, Environ Health Perspect. Feb 2010; 118(2): 182–190[79]

Table 3.3. Exposure levels associated with kidney and bone effects.	
Study population, age, reference	**Exposure/outcomes**
Sweden, n = 820, 53–64 years of age[70, 80]	Blood and urinary cadmium at 0.38 μg/L and 0.67 μg/g creatinine were associated with tubular impairment. Urinary cadmium at 0.8 μg/g creatinine was associated with glomerular impairment. Increased body burden of cadmium was associated with lowered bone mineral density, decreased serum parathyroid hormone and bone metabolism.
Thailand, n = 200, 16–60 years of age[81]	A 3-fold increase in body burden associated with 11%, 32%, and 61% increases the probability of having high blood pressure, renal injury, and tubular impairment.
Thailand, n = 224, 30–87 years of age[82]	OR for tubular impairment was 10.6, comparing urinary cadmium 1–5 versus > 5 μg/g creatinine.
United States, n = 4,258, ≥ 50 years of age[83]	A 1.43-fold increase in osteoporosis risk, comparing urinary cadmium 1 versus < 0.5 μg/g creatinine
Belgium, n = 294, mean age 49.2 years of age[84]	A 2-fold increase in body burden associated with increased bone resorption, urinary calcium loss, decreased proximal forearm bone density, and low serum parathyroid hormone.
China, n = 148, 3-year observation[85]	Progressive tubular and glomerular impairment was observed among those with urinary cadmium > 10 μg/g creatinine.
United Kingdom, n = 160, 18–86 years of age[86]	Risk for early renal effects[a] was increased by 2.6-fold and 3.6-fold, comparing urinary cadmium 0.3 versus < 0.5 versus ≥ 0.5 μg/g creatinine.
United States, n = 14,778, > 20 years of age[6]	Risk for albuminuria was 2.34 and risk for lowered glomerular filtration rate was 1.98, comparing those in the highest versus lowest quartiles of blood cadmium and lead.
Abbreviations: OR, odds ratio [a]Early renal injury was defined as urinary NAG > 2 IU/g creatinine.	

Table adapted from Satarug, et al, Environ Health Perspect. Feb 2010; 118(2): 182–190[79]

Table 3.4. Descriptive Statistics of Study Population

	Unexposed (n=100)	Exposed (n=169)	
	Mean (SD)	Mean (SD)	*p*-value
Age (years)	61.0 (11.5)	61.2 (11.8)	0.93
	Median (IQR)	Median (IQR)	*p*-value
UCd (µg/L)	1.0 (0.65, 1.60)	7.4 (4.4, 10.2)	<0.001
Specific Gravity	1.017 (1.013, 1.020)	1.014 (1.011, 1.018)	0.01
β2MG (µg/L)	235.3(98.6, 411.2)	1171.3 (199.0, 6715.0)	<0.001
NAG (U/L)	5.54 (3.7, 9.4)	8.84 (5.2, 12.7)	0.0002
SCr (mg/dL)	0.9 (0.8, 1.1)	1.1 (0.9, 1.3)	<0.001
UCr (g/L)	1.3 (1.0, 1.6)	1.4 (1.1, 1.7)	0.08
LINE-1 (%)	82.5 (79.0, 81.1)	80.2 (79.4, 81.1)	0.87
MT2A (%)	23.2 (20.9, 28.8)	23.1 (20.8, 25.0)	0.1
DNMT (%)	1.4 (1.1, 1.9)	1.4 (1.2, 1.9)	0.78
MGMT (%)	2.3 (1.8, 3.1)	1.7 (1.4, 2.1)	<0.001
WBC (10^3cells/uL)	8.4 (6.9, 9.9)	8.3 (6.9, 9.9)	0.57

Abbreviations: UCd (urinary cadmium); β2MG (β2-microglobulin); NAG (N-acetyl-beta-D-glucosaminidase);SCr (serum creatinine); WBC (white blood cell); n (number of subjects); SD (standard deviation); IQR (Interquartile Range).

Table 3.5. Demographics of Study Population

	Unexposed (n=100)			Exposed (n=169)		
	All	**Male**	**Female**	**All**	**Male**	**Female**
Smoking	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Never	36 (36)	6 (12)	30 (60)	66 (39.1)	11 (13.3)	55 (64)
Former	33 (33)	21 (42)	12 (24)	32 (18.9)	17 (20.5)	15 (17.4)
Regular	31 (31)	23 (46)	8 (16)	71 (42.0)	55 (66.3)	16 (18.6)
Diabetes Mellitus (DM)						
No	88 (88)	45 (90)	43 (86)	158 (93.5)	80 (96.4)	78 (90.7)
Yes	12 (12)	5 (10)	7 (14)	11 (6.5)	3 (3.6)	8 (9.3)
Hypertension (HT)						
No	61 (61)	29 (58)	32 (64)	93 (55)	49 (59)	44 (51.2)
Yes	39 (39)	21 (42)	18 (36)	76 (45)	34 (41)	42 (48.8)
Urinary Stones						
No	95 (95)	47 (94)	48 (96)	140 (82.8)	68 (81.9)	72 (83.7)
Yes	5 (5)	3 (6)	2 (4)	29 (17.2)	15 (18.1)	14 (16.3)
Drinker						
Never	50 (50)	6 (12)	44 (88)	82 (48.5)	19 (22.9)	63 (73.3)
Former	26 (26)	23 (46)	3 (6)	33 (19.5)	22 (26.5)	11 (12.8)
Regular	24 (24)	21 (42)	3 (6)	54 (32.0)	42 (50.6)	12 (14)
Occupation						
Agriculture	74 (74)	36 (72)	38 (76)	85 (50.3)	51 (61.4)	34 (39.5)
Other	26 (26)	14 (28)	12 (24)	84 (49.7)	32 (38.6)	52 (60.5)

Table 3.6. Descriptive Statistics of Study Population by Sex

	Unexposed (n=100)			Exposed (n=169)				
	Male *(n=50)*	**Female** *(n=50)*	***p*-value[a]**	**Male** *(n=83)*	**Female** *(n=86)*	***p*-value[b]**	***p*-value[c] (M)**	***p*-value[d] (F)**
	Mean (SD)	*Mean (SD)*		*Mean (SD)*	*Mean (SD)*			
Age (years)	62.5 (12.1)		0.21	62.7 (11.7)	59.7 (11.8)	0.09	0.91	0.96
	Median (IQR)	*Median (IQR)*		*Median (IQR)*	*Median (IQR)*			
UCd (µg/L)	0.9 (0.5, 1.2)	1.4 (0.9, 2.0)	<0.001	8.0 (5.2, 10.5)	6.4 (3.6, 9.5)	0.04	<0.001	<0.001
Specific Gravity	1.018 (1.013, 1.019)	1.016 (1.012, 1.020)	0.3	1.014 (1.012, 1.019)	1.014 (1.011, 1.017)	0.12	0.08	0.08
β2MG (µg/L)	277.2 (133.7, 678.5)	164.1 (87.1, 315.6)	0.03	1434.4 (269.8, 6323.1)	735.6 (154.8, 7073.8)	0.5	<0.001	<0.001
NAG (U/L)	5.7 (4.5, 10.4)	5.4 (3.4, 7.4)	0.12	9.8 (6.8, 13.5)	7.8 (4.6, 12.5)	0.03	0.003	0.02
SCr (mg/dL)	1.0 (0.9, 1.1)	0.8 (0.8, 1.3)	<0.001	1.2 (1.0, 1.3)	1.0 (0.8, 1.2)	<0.001	0.001	0.0002
UCr (g/L)	1.3 (1.2, 1.6)	1.1 (1.0, 1.5)	0.003	1.4 (1.2, 1.8)	1.3 (1.0, 1.6)	0.008	0.33	0.08
LINE-1 (%)	80.6 (79.0, 81.1)	80.4 (79.0, 81.2)	0.64	80.4 (79.7, 81.4)	79.7 (78.9, 80.7)	0.001	0.18	0.4
MT2A (%)	22.8 (20.7, 24.9)	23.8 (21.5, 26.2)	0.09	22.7 (20.7, 24.4)	23.5 (21.0, 25.5)	0.69	0.65	0.06
DNMT (%)	1.3 (0.9, 1.8)	1.7 (1.3, 2.1)	0.008	1.5 (1.2, 1.9)	1.4 (1.2, 1.9)	0.66	0.03	0.08
MGMT (%)	2.4 (2.0, 3.4)	2.2 (1.7, 2.7)	0.22	1.6 (1.3, 2.2)	1.7 (1.4, 2.0)	0.91	0.0002	0.003
WBC (10^3cells/uL)	8.7 (7.0, 9.9)	8.3 (6.8, 10.4)	0.74	8.5 (7.2, 10.3)	8.1 (6.9, 9.8)	0.13	0.92	0.37

Abbreviations: UCd (urinary cadmium); β2MG (β2-microglobulin); NAG (N-acetyl-beta-D-glucosaminidase);SCr (serum creatinine); WBC (white blood cell); n (number of subjects); SD (standard deviation); IQR (Interquartile Range); M (males); F (females). *p*-values are given for comparisons between males and females within the [a]unexposed group, [b]exposed group, and between exposure groups within [c]males and [d]females.

Table 3.7. Associations between Urinary Cadmium, Methylation and Renal Markers

		Urinary Cadmium (µg/L)[a]		
		All	**Men**	**Women**
	N	IQR Change (95% CI)	IQR Change (95% CI)	IQR Change (95% CI)
LINE-1[c]	259	0.5 (-0.3, 1.2)	0.9 (0, 1.7)	0.03 (-0.9, 0.9)
MT2A[c]	269	-3.2 (-9.2, 3.2)	1.3 (-10.9, 15.2)	-7.4 (-12.9, -1.7)
DNMT[c]	264	-1.8 (-8.6, 5.5)	2.2 (-10.1, 16.2)	-8.5 (-16.1, -0.1)
MGMT[c]	226	-16.6 (-25.1, -7.2)	-24.7 (-36.5, -10.7)	-8.9 (-19.3, 2.9)
		β2MG (µg/L)[b]		
		All	**Men**	**Women**
	N	Change (95% CI)[d]	Change (95% CI)	Change (95% CI)
UCd[c]	269	1044.6 (646.2, 1655.9)	1130.7 (547.8, 2238.1)	1090.3 (548.8, 2083.6)
LINE-1[d]	259	15.1 (3.3, 79.9)	32.7 (13, 158.1)	2.7 (-11.8, 26)
MT2A[d]	269	0.4 (-0.6, 32.3)	1.4 (0.3, 53.4)	-1.4 (-3.1, 5.8)
DNMT[d]	264	-0.9 (-1.8, 1)	-0.3 (-1.5, 60.3)	-1.5 (-2.8, -3.1)
MGMT[d]	226	-0.8 (-1.5, -8.2)	-0.9 (-1.7, -3.4)	-0.7 (-1.9, 23.3)
		NAG (U/L)[b]		
		All	**Men**	**Women**
	N	IQR Change (95% CI)	IQR Change (95% CI)	IQR Change (95% CI)
UCd[c]	269	60.5 (39.1, 85.1)	65.2 (33.4, 104.6)	61.6 (34.7, 93.8)
LINE-1[d]	259	4.0 (0.9, 7.3)	5.1 (0.4, 10)	2.5 (-1.8, 7)
MT2A[d]	269	-0.04 (-0.3, 0.2)	0.3 (0, 0.6)	-0.7 (-1.1, -0.2)
DNMT[d]	264	0.05 (-0.2, 0.3)	0.2 (-0.2, 0.5)	-0.004 (-0.4, 0.4)
MGMT[d]	226	-0.1 (-0.3, 0.1)	-0.1 (-0.3, 0.2)	-0.1 (-0.4, 0.2)
		Serum Creatinine (mg/dL)[b]		
		All	**Men**	**Women**
	N	IQR Change (95% CI)	IQR Change (95% CI)	IQR Change (95% CI)
UCd[c]	269	13.6 (9.6, 17.7)	16.5 (7, 26.9)	11.4 (4.9, 18.4)
LINE-1[d]	259	1.2 (0.1, 2.3)	2.5 (1, 4.1)	0.1 (-1.4, 1.6)
MT2A[d]	269	0.01 (-0.1, 0.1)	0.05 (0, 0.1)	0 (-0.2, 0.1)
DNMT[d]	264	-0.1 (-0.2, 0.02)	-0.01 (-0.1, 0.11)	-0.1 (-0.3, 0)
MGMT[d]	226	-0.1 (-0.1, -0.01)	-0.1 (-0.1, 0.01)	-0.1 (-0.2, -0.01)

[a]predictor ; [b]outcome; [c]Percent change associated with IQR increase in UCd ; [d]Percent change in renal marker level with a 1% increase in methylation; *All variables except age log-transformed; **all models are adjusted by age

Table 3.8. Best-fit Models for Exposed and Unexposed Populations

	UNEXPOSED N=100						EXPOSED N=169					
Predictors	**β2MG(μg/L)**		**NAG(U/L)**		**SCr (mg/dL)**		**β2MG(μg/L)**		**NAG(U/L)**		**SCr (mg/dL)**	
	β (SE)	*p*	β (SE)	*p*	β (SE)	*p*	β (SE)	*p*	β (SE)	*p*	β (SE)	*p*
UCd (μg/L)			0.28 (0.19)	0.14			1.77 (0.22)	<0.001	0.46 (0.06)	<0.001	0.07 (0.03)	0.0
Age (years)	0.05 (0.01)	<0.001	0.02 (0.005)	0.002	0.003 (0.0006)	<0.001	0.07 (0.01)	<0.001	0.01 (0.003)	<0.001	0.004 (0.001)	0.0
Former Smoker	0.82 (0.29)	0.006									0.04 (0.05)	0.3
Regular Smoker	0.46 (0.28)	0.1									-0.07 (0.04)	0.0
Former Drinker												
Regular Drinker												
Male			0.24 (0.12)	0.05	0.10 (0.01)	<0.001					0.12 (0.04)	0.0(
WBC (10^3cells/uL)	1.16 (0.45)	0.01					1.36 (0.49)	0.006	0.50 (0.12)	<0.001	0.14 (0.06)	0.0
Hgb							-0.20 (0.08)	0.01			-0.02 (0.01)	0.1
DM	0.97 (0.36)	0.008	0.53 (0.17)	0.002			1.07 (0.52)	0.04			0.27 (0.06)	<0.0
HT							0.60 (0.27)	0.03				
Urinary Stone							0.93 (0.33)	0.005	0.28 (0.08)	<0.001	0.07 (0.04)	0.0
AIC	316.6		170.7		-254.3		649.2		182		-58.7	
AdjR2	0.3692		0.18		0.4826		0.5325		0.4689		0.272	
p-value	<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	

Abbreviations: UCd (urinary cadmium); β2MG (β2-microglobulin); NAG (N-acetyl-beta-D-glucosaminidase);SCr (serum creatinine); WBC (white blood cell); N (number of subjects); SE (standard error); Hgb (hemoglobin levels); DM (diabetes mellitus); HT (hypertension);

Table 3.9. Combined predictors from both Populations						
	β2MG(µg/L)		NAG(U/L)		SCr(mg/dL)	
Predictors	β (SE)	*p*-value	β (SE)	*p*-value	β (SE)	*p*-value
UCd (µg/L)	1.30 (0.13)	<0.001	0.27 (0.04)	<0.001	0.07 (0.01)	<0.001
Age (years)	0.06 (0.009)	<0.001	0.02 (0.003)	<0.001	0.003 (0.001)	<0.001
Former Smoker	0.38 (0.26)	0.14			0.01 (0.03)	0.64
Regular Smoker	0.19 (0.23)	0.41			-0.06 (0.03)	0.04
Former Drinker						
Regular Drinker						
Male			0.13 (0.06)	0.03	0.12 (0.03)	<0.001
WBC (10^3cells/uL	1.20 (0.37)	0.001	0.37 (0.11)	<0.001	0.10 (0.04)	0.02
Hgb	-0.13 (0.06)	0.05			-0.01 (0.008)	0.05
DM	1.11 (0.34)	0.001	0.34 (0.10)	0.001	0.13 (0.04)	<0.001
HT	0.52 (0.20)	0.01	0.27 (0.09)	0.002		
Urinary Stone	0.88 (0.28)	0.002			0.06 (0.03)	0.08
AdjR2	0.5314		0.35		0.2728	
p-value	<0.001		<0.001		<0.001	

Abbreviations: UCd (urinary cadmium); β2MG (β2-microglobulin); NAG (N-acetyl-beta-D-glucosaminidase);SCr (serum creatinine); WBC (white blood cell); N (number of subjects); SE (standard error); Hgb (hemoglobin levels); DM (diabetes mellitus); HT (hypertension);

Table 10. Combined Predictors from both Populations Stratified by Gender												
	Males N=133						Females N=136					
	β2MG		**NAG**		**SCr**		**β2MG**		**NAG**		**SCr**	
Predictors	β (SE)	*p*	β (SE)	*p*	β (SE)	*p*	β (SE)	*p*	β (SE)	*p*	β (SE)	*p*
UCd	1.05 (0.16)	<0.001	0.22 (0.05)	<0.001	0.07 (0.02)	<0.001	1.59 (0.20)	<0.001	0.33 (0.06)	<0.001	0.07 (0.02)	0.003
Age	0.04 (0.01)	0.001	0.01 (0.004)	<0.001	0.003 (0.001)	0.007	0.09 (0.01)	<0.001	0.01 (0.004)	0.002	0.003 (0.002)	0.14
Former Smoker	0.46 (0.44)	0.29			0.05 (0.04)	0.23	-0.18 (0.37)	0.63			0.002 (0.05)	0.97
Regular Smoker	0.07 (0.40)	0.87			-0.02 (0.04)	0.54	0.14 (0.37)	0.71			-0.05 (0.04)	0.27
Former Drinker												
Regular Drinker												
Male												
WBC	0.97 (0.55)	0.08	0.31 (0.17)	0.07	0.06 (0.05)	0.27	1.57 (0.51)	0.002	0.43 (0.16)	0.007	0.14 (0.06)	0.03
Hgb	-0.17 (0.09)	0.05			-0.009 (0.009)	0.28	-0.17 (0.11)	0.12			-0.02 (0.01)	0.11
DM	0.98 (0.55)	0.08	0.14 (0.17)	0.42	0.24 (0.05)	<0.001	1.27 (0.43)	0.004	0.43 (0.13)	0.002	0.06 (0.05)	0.26
HT	0.45 (0.29)	0.12	0.002 (0.09)	0.98	0.05 (0.03)	0.07	0.50 (0.29)	0.08	0.12 (0.09)	0.2	-0.007 (0.04)	0.83
Urinary Stone	0.98 (0.38)	0.01	0.25 (0.12)	0.04	0.06 (0.04)	0.11	0.88 (0.44)	0.05	0.29 (0.14)	0.03	0.03 (0.05)	0.64
AdjR2	0.4764		0.2729		0.3855		0.5892		0.3849		0.1181	
p-value	<0.001		<0.001		<0.001		<0.001		<0.001		0.002	

Abbreviations: UCd (urinary cadmium); β2MG (β2-microglobulin); NAG (N-acetyl-beta-D-glucosaminidase);SCr (serum creatinine); WBC (white blood cell); N (number of subjects); SE (standard error); Hgb (hemoglobi levels); DM (diabetes mellitus); HT (hypertension);

Figure 3.1 Map of Thailand. Mae Sot District is marked (red star)

Figure 3.2. Mediation Analysis. The total effect of the independent variable (IV) is calculated on the dependent variable (DV) alone. Then the Average Direct Effect (ADE) of the IV on DV is calculated with the potential mediator in the model. The average causal mediation effects (ACME) are calculated to determine the effect the mediator has on the relationship between the IV and DV. Adding the ACME and the ADE gives the total effect.

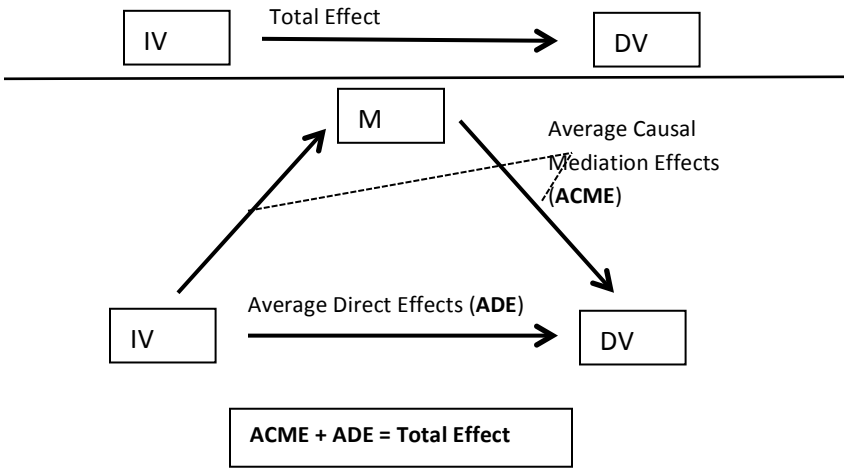


Figure 3.3. Percent change for each methylation marker per IQR increase in urinary Cd for the total population and by sex. In the total population, MGMT methylation decreases with an IQR increase in urinary Cd (purple). In males, an IQR increase in urinary Cd is associated with an increase in LINE-1 and a decrease in MGMT methylation (light blue). In females, an IQR increase in urinary Cd is associated with a decrease in MT2A and DNMT methylation (dark blue).

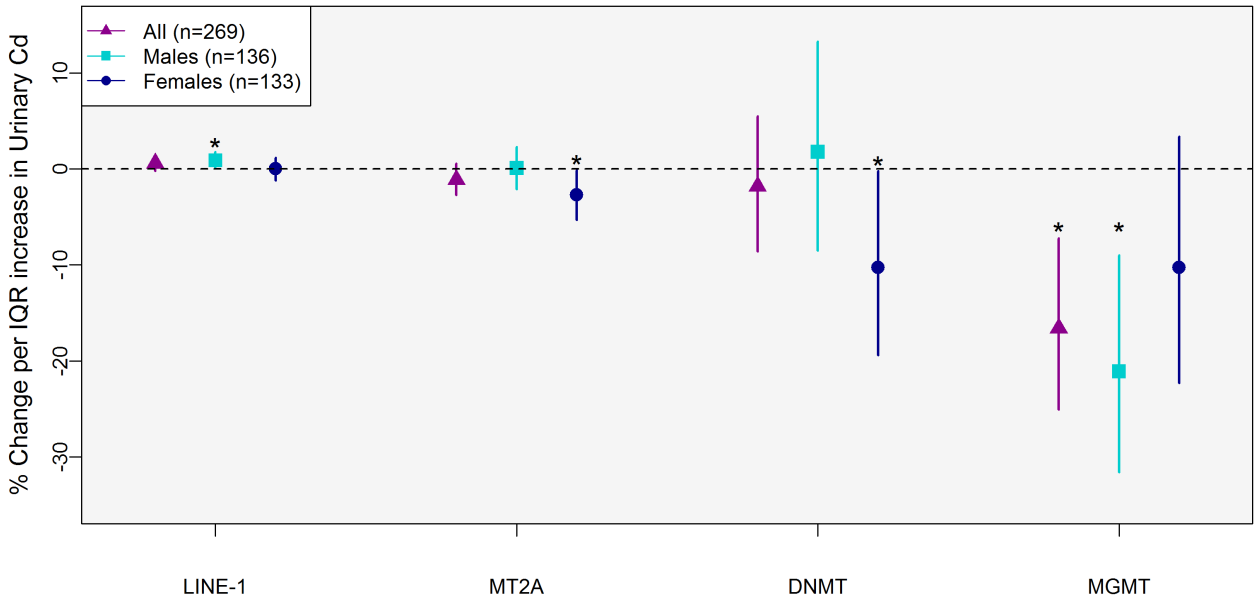


Figure 3.4. Mediation analysis. a) Males. LINE-1 is a mediator of the effects of Cd on B2MG (black), but not of the effects of Cd on blood creatinine (red) b) Females. No mediators were found.

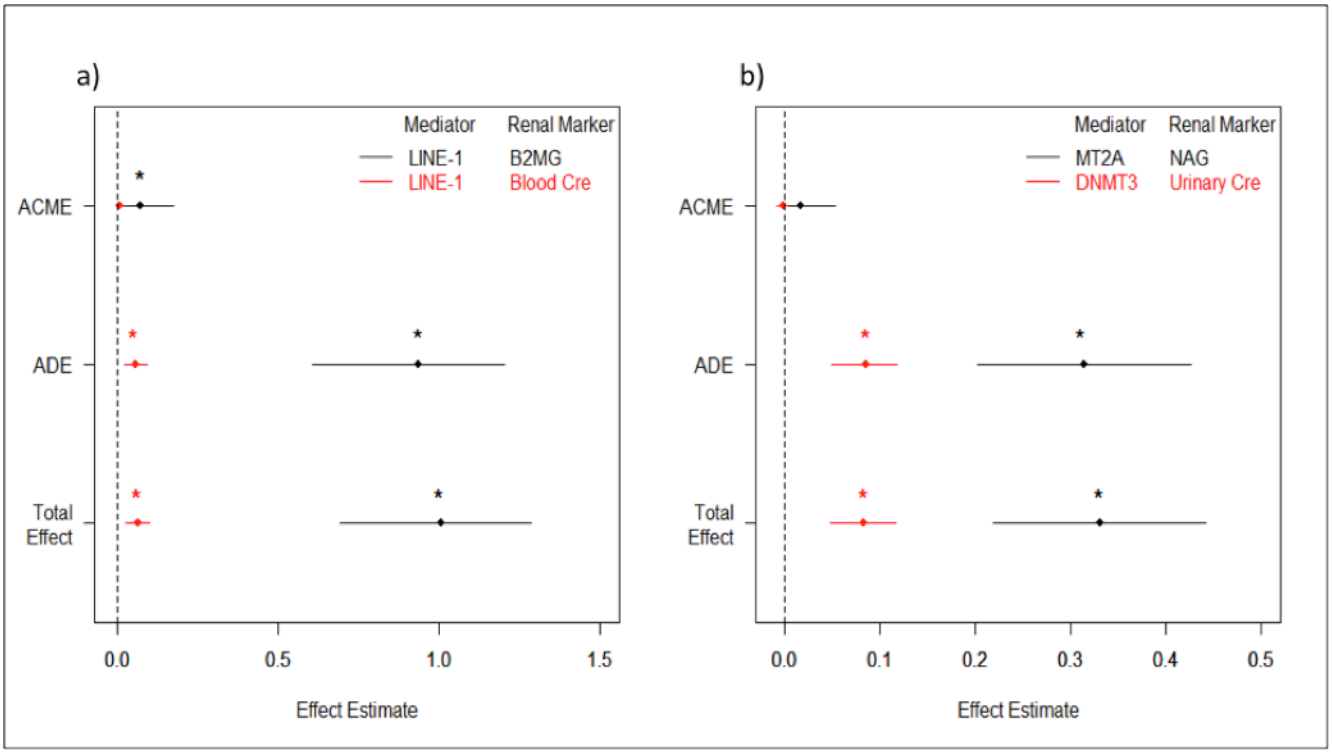
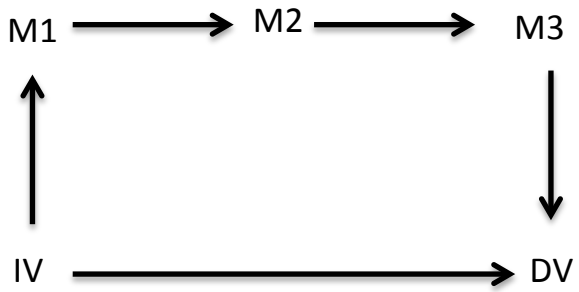


Figure 3.5. A better way to assess mediation along the pathway from Cd exposure to adverse health outcomes would be to consider multiple mediators as biological mechanisms rarely involve only one marker



CHAPTER 4
Gene Methylation markers are associated with recurrence and survival in patients with head and neck squamous cell carcinoma

ABSTRACT

Background: The overall 5-year survival rate of head and neck squamous cell carcinoma (HNSCC) has remained at about 50-60% for the past decade, primarily due to loco-regional or metastatic recurrence. HPV-associated HNSCCs have a unique risk profile, clear etiologic mechanism, and an improved prognosis as compared to non-HPV associated HNSCCs. However, even within the HPV+ group, there is heterogeneity in survival time, with up to 20% progressing with distant metastases. Thus, there is strong interest in identifying prognostic markers in these patients related to tumor biology and epidemiologic characteristics.

Methods: Six gene methylation markers were analyzed using DNA extracted from previously untreated tumor tissue collected from a well-characterized, unselected cohort of HNSCC patients at the University of Michigan. Survival analysis was conducted to determine the association of these markers with overall survival time and recurrence-free survival time using Kaplan-Meier curves and stratified Cox proportional hazards models.

Results: In the multivariable Cox proportional hazards model, those with low methylation of *NDN* and *CD1A* had 1.6 and 1.3 times the odds of a death event, respectively ($p < 0.01$; $p < 0.05$, respectively). Multivariable Cox models stratified by HPV status revealed that HPV+ patients

with low methylation of *CD1A* had 3.3 times higher odds of a death event and 2.1 times higher odds of a recurrence event (p <0.01, p <0.01) . However, the same patients with low *CCNA1* methylation had 0.31 times lower odds of a recurrence event (p <0.05). HPV- patients with low *NDN* methylation had 1.6 times higher odds of a death event (p <0.05).

Conclusions: The novel methylation markers identified in this study offer new, specific, epigenetic molecular differences within the setting of the generalized hypermethylation phenotype associated with HPV status and warrant further investigation. The findings support biological implications of epigenetic markers on patient survival and their potential clinical utility in identifying unique subsets of patients with varied outcomes.

INTRODUCTION

Head and neck cancer is the 6th most common cancer in the world with at least 90% being squamous cell carcinomas and approximately 600,000 new cases each year [1-3]. Sites of the head and neck are shown in Figure 4.1. Heavy tobacco and alcohol use are well established risk factors but more recently human papilloma virus (HPV) infection has been identified as a new etiologic factor for head and neck squamous cell carcinomas (HNSCCs). The overall 5-year survival rate has remained at about 50-60% for the past decade, primarily due to loco-regional or metastatic recurrence, which present in 35% to 55% of patients within two years (Figure 4.2)[1, 4]. This is partially due to the fact that almost 60% of patients are diagnosed after the disease has locally advanced, but also due to pathological, clinical and epidemiological heterogeneity[5, 6] and frequent association of significant co-morbidities.

The incidence of HPV-associated HNSCC has steadily increased, especially in younger patients, while incidence of non-HPV associated HNSCC has declined in recent years[7, 8]. HPV-associated HNSCCs have a unique risk profile, clear etiologic mechanism, and an improved prognosis as compared to non-HPV associated HNSCCs[7, 9-14]. HPV+ patients tend to have cancers almost exclusively located in the oropharynx, to be younger, with a higher socioeconomic status and a less profound use of alcohol and tobacco[12, 15, 16]. Studies show a 60-80% reduction in mortality in HPV+ patients compared to patients with non-HPV associated HNSCC[17-19] regardless of treatment modality or tumor stage (Figure 4.3). Even within the HPV+ group, however, there is heterogeneity in survival time, with up to 20% progressing with distant metastases. Distinct patterns of chromosomal aberrations and copy number variation have been found to be associated with shorter survival times [20, 21]. Changes in gene expression patterns of tumor suppressors and oncogenes have been used to classify subtypes of HNSCC with improved

prognosis[22-24]. Thus, there is strong interest in identifying potential prognostic markers for patients with HNSCC. A useful approach would be use epigenetic factors, particularly for DNA methylation, as prognostic markers. CpG methylation changes precede genetic aberrations and gene expression, providing an underlying mechanism for tumor heterogeneity.

It has been demonstrated that epigenetic alterations are part of the causal pathway in tumorigenesis[25]. Aberrant methylation profiles have implicated the involvement of many genes in the development and progression of HNSCC. Changes in methylation have been found in genes involved in various pathways including cell cycle control[26-28], signal transduction[22, 26, 29-32], apoptosis[26, 33], cell-cell adhesion[33, 34], immune response[35, 36], and epithelial-mesenchymal transition[28, 37] pathways. Some of these findings have been extended to propose prognostic markers using discovery-based[27, 38, 39] or candidate gene based approaches[28, 30]. However, these studies have a limited ability to detect strong survival associations due to small sample sizes. Large sample cohorts with defined clinic-pathological and epidemiological characteristics are needed to provide reliable prognostic biomarkers[40]. Combining methylation information of clinical characteristics known to affect survival with other biomarkers is crucial to understanding the differences in survival rates by these characteristics and how they may be targeted for intervention. Few studies have taken a comprehensive approach in observing methylation of prognostic biomarkers across known clinical and epidemiological indicators of survival.

It is unclear whether the prognostic advantage of HPV-associated tumors is due to tumor biology, epidemiologic characteristics, or a combination of the two. Our group recently completed a discovery-based approach to determine novel prognostic epigenetic biomarkers[27] and here we test identified markers for their association with tumor recurrence and survival. This takes advantage of a well-characterized, unselected cohort of HNSCC patients at the

University of Michigan, with extensive epidemiologic, clinical, and survival information that were treated by a single group of clinicians with a homogenous treatment approach. This allows careful consideration of the epigenetic biomarkers in the context of epidemiologic and clinico-pathologic characteristics that also influence survival. Identification of significant epigenetic markers of biologic tumor behavior and outcome should open new horizons for interventions directed at reversible gene alterations and new therapeutic targets.

METHODS

Recruitment. The University of Michigan's Head and Neck Specialized Program of Research Excellence (SPORE) approaches every incident, previously untreated head and neck squamous cell carcinoma (HNSCC) patient to participate in our longitudinal epidemiology studies. This unselected study population represents 28% of incident HNSCC cases in the state of Michigan. From November 2008 through June 2012, subjects were screened for eligibility and 92% of subjects approached signed a written, informed consent. After consent, 513 subjects were asked to complete a baseline epidemiologic questionnaire of demographics, epidemiologic characteristics, and behavior modules. Research assistants collected paraffin-embedded (FFPE) tissue blocks and detailed pathophysiological and clinical data annually until death or lost to follow-up for patients with tumor of the oral cavity, oropharynx, hypopharynx and larynx. This study was approved by the Institutional Review Board of the University of Michigan Medical School (HUM00042189).

Tissue acquisition. The FFPE tissue blocks were collected from three possible sources: (1) a biopsy done at an outside hospital, (2) a biopsy done at the University of Michigan hospital,

and/or (3) a surgery done at the University of Michigan hospital. All possible tumor specimen blocks were collected for each subject, except if blocks were unable to be obtained due to reasons of outside hospital refusal to provide blocks or a block was unavailable for research purposes. Tissue acquired from the three sources yielded a possible sample for 88% of subjects.

Study Population. A pathologist (JM) confirmed tumor histology and screened representative blocks for areas of >70% cellularity and minimal necrosis. Tumors with insufficient histology, inadequate tissue, or yielded insufficient amounts of DNA for methylation assays were excluded from analysis. There remained 72% of subjects who had sufficient tissue which yielded methylation results. There were 15 subjects whose sites were not analyzed (unknown primary, nasopharynx, salivary gland, skull bones), 7 subjects with equivocal HPV status, and 1 subject lost to follow up. The final count of study subjects is 346 and represents 67% of eligible participants screened for this analysis.

Follow-up. All patients were followed prospectively at designated intervals by our clinicians in clinics at the University of Michigan or through contact with referring physicians. Median follow period for survival was 27.4 months and for 24.9 months for recurrence. Deaths were captured through the Social Security Death Index, yearly survey updates, notification from family, and medical record reviews. Survival time and events were censored as of 4/30/13. Recurrence and persistent disease events were updated annually during a chart review occurring at every subject's yearly anniversary of their date of initial diagnosis.

Microdissection/DNA Extraction/Bisulfite Conversion. Designated areas of FFPE tissue were microdissected from unstained slides and DNA was extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. DNA concentration and purity was measured with a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA)). Sodium bisulfite modification was performed on 250ng of DNA using the Epitect Bisulfite Kit ((Qiagen, Valencia, CA) according to the manufacturer's recommended protocol.

HPV testing. Tumor HPV status was determined by an ultrasensitive method using real-time competitive polymerase chain reaction and matrix-assisted laser desorption/ionization time of flight mass spectroscopy with separation of products on a matrix loaded silicon chip array, as described in Tang et al.[41]. Multiplex PCR amplification of the E6 region of 15 discrete high risk HPV types (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 and 73), and human GAPDH control was run to saturation followed by shrimp alkaline phosphatase quenching. Amplification reactions included a competitor oligo identical to each natural amplicon except for a single nucleotide difference. Probes that identify unique sequences in the oncogenic E6 region of each type were used in multiplex single base extension reactions extending at the single base difference between wild-type and competitor HPV so that each HPV type and its competitor were distinguished by mass when analyzed on the MALDI-TOF mass spectrometer as described previously[42-45].

Methylation Analysis. *CCNA1, NDN, DCC and CD1a* were chosen for further study from a previous analysis that found them to be differentially methylated in head and neck cancer

patients by HPV status [27]. Methylation assays for 3 genes, *DCC, CD1a,* and *NDN*, were designed using PyroMark Assay Design 2.0 software. *CCNA1* was sequenced using the Sequenom EpiTyper, a MALDI-TOF mass spectrometry based platform. Bisulfite singleplex PCR amplification was performed using FastStart Taq Polymerase (Roche Diagnostics, Indiana, US) with a forward and reverse primer concentration of 0.2 mM and 10ng/uL of bisulfite-converted DNA. Fifteen microliters of each PCR product was combined with the respective sequencing primer and methylation analysis by pyrosequencing was conducted for each assay as previously described[46]. Complete coverage of all samples for every methylation marker selected was not possible due to low quantity of total extracted DNA.

Statistical Analysis. Univariate and bivariate analyses were conducted on all methylation markers and pathological and epidemiological characteristics. Kaplan-Meier analysis was employed to graphically visualize time-to-event outcomes for the probability of overall survival (OST) and recurrence-free survival time (RFT). Methylation of each marker was divided into quartiles with quartile 1 containing the lowest values and quartile 4 containing the highest. Statistical differences in curves were tested using the log-rank test. All methylation markers except *NDN* were log-transformed due to departures from normality to satisfy model assumptions.

Cox proportional hazard models were used to adjust for identified prognostic factors including patient age, tumor stage, disease site (and comorbidity score for OST), and HPV status. Stratified Cox proportional hazard models were used to vary the baseline hazard function by HPV status to account for non-constant hazards observed in our data by HPV status. Odds of a death or recurrence were based on methylation values standardized to the interquartile range

(IQR) of each respective marker. The interpretation of the hazard ratio was a comparison of those with methylation in the 25th percentile compared to those in the 75th percentile.

A methylation score was calculated to create a composite measurement of methylation across all candidate genes. Methylation for each gene less than 10% was given a score of zero. The median of the remaining samples for each gene was calculated and methylation values less than the median were given a score of 1 while methylation values above the median were given a score of 2 for each gene. These scores were added up across all genes for each patient to generate a composite score of methylation. The composite scores were divided according to distribution across scores and by quartiles and analyzed by HPV status and for OST and RFT.

RESULTS

Population Characteristics

The mean age of the population was 59.7 years with 75% males (Table 4.1). Cancer sites were mostly oropharyngeal and oral cavity (36% each) while glottic cancers made up about 24% of cases and only 3% presented in the hypopharynx. Sixty-one percent of cases were stage IV and HPV(-). Most patients had mild comorbidity status (46%), while 26% had moderate comorbidity scores and 8% had severe scores. Forty-two percent of patients were classified as current smokers, having quit within the past 12 months, while 36% were former smokers and 22% were never smokers. Most patients were former alcohol users (66%), having quit greater than 12 months ago. (Table 4.1).

Patients with HPV+ tumors were on average younger than patients with HPV(-) tumors (mean age = 61.4 years, SD: 12.3 years for HPV (-) patients and 57 years, SD: 9.6 years, for HPV (+) patients (Table 4.2). The ratio of male to female patients was more extreme in the group

of patients with HPV (+) tumors (89% HPV (+) male vs. 66% HPV (-) male, p-value<0.0001). The majority of HPV (-) patients had cancers of the oral cavity (52%) whereas the majority of HPV (+) patients had cancer of the oropharynx (78%). The highest proportion of patients was stage IV with none or mild comorbidities regardless of HPV status, although more patients with HPV (+) tumors were diagnosed with stage IV tumors than those patients with HPV (-) tumors (77% vs 51%) primarily due to a higher frequency of patients with N2 neck disease that is so common in HPV related cancers. Most HPV (-) patients were current smokers (48%) while HPV (+) patients had lower proportions of current (32%) smokers which were similar to the frequency of former (37%) and never (31%) smokers. Pack-years of cigarettes use only and all types of tobacco use were higher in HPV (-) patients (mean: 38.4 pack-years; 41.8 pack-years, respectively) compared to HPV (+) patients (29.9 pack-years; 29.6 pack-years, respectively; p-value=0.02, p-value=0.005). Alcohol use, post treatment status and two year recurrence and survival rates were also significantly different by HPV status (Table 4.2).

Patient Characteristics and Survival/Recurrence

Kaplan-Meier curves with age categorized into quartiles revealed that the oldest quartile (Q4) had the worst probability of overall survival time (OST) compared to other ages (p<0.001), whereas there were similarly high probabilities of recurrence free survival time (RFT) for the two lowest quartiles of age and similarly low probabilities of RFT for the two highest quartiles of age, although these differences were not significant (p=0.09) (Figure 4.4). Patients who were HPV (-) had significantly poorer probabilities of both OST and RFT (p<0.001; p=0.001, respectively) (Figure 4.5). Smoking status was not significantly associated with either probability of OST or RFT, however trends revealed that never smokers had a better probability of OST than

former and current smokers (p=0.16). There was no difference in trend by smoking status with regard to probability of RFT (p=0.98) (Figure 4.6). Patients with HNSCC in their hypopharynx had the worst probabilities of OST and RFT compared to other disease sites. Following this, cancers of the larynx, oral cavity and oropharynx had increasing probabilities of OST and RFT (p<0.001; p<0.001, respectively) (Figure 4.7). There was no difference on probability of OST by stage of disease, although the trend showed earlier stages with slightly better probabilities (p=0.22). However, there was a clear separation in probability of RFT with the earlier stages clustering at around 0.85 compared the later stages clustered around 0.65 (p=0.05) (Figure 4.8). Patients with the highest ACE score denoting their comorbidity status had the lowest probabilities of OST and RFT. Probability of OST increased with decreasing ACE score (p<0.001) while probability of RFT was not significantly different for ACE scores 0-3 (p=0.10) (Figure 4.9).

Tumor Methylation and Survival/Recurrence

Quartiles of methylation show differential effects for survival and recurrence

We categorized methylation by quartile within markers, with Q1 indicating the lowest quartile of methylation, and compared probabilities of overall and recurrence-free survival between quartiles (Figure 4.10). Both *CD1A* and *NDN* were significantly associated with OST in the total population. Methylation of *CD1A* in the lowest quartile (Q1) was associated with the lowest probability of OST whereas higher methylation in Q2, Q3 and Q4 clustered around a probability of OST of about 0.75 (p =0.007, Figure 4.10). Methylation of *NDN* in Q1 and Q3 were associated with poor probability of OST compared to methylation in Q2 and Q4 (p-value=0.001). *CCNA1*, *DCC* and *NDN* methylation were significantly associated with RFT in the

total population (Figure 4.11). *CCNA1* and *DCC* depicted a U-shaped association where the lowest and highest quartiles, Q1 and Q4, had the lowest probabilities of RFT while the intermediate quartiles, Q2 and Q3, had similarly higher probabilities of RFT (p=0.04; p=0.01, respectively). *NDN* revealed a linear trend of increasing probability of RFT with increasing quartile (p=0.009).

Quartiles of methylation across HPV status

In HPV (+) patients, only *CD1A* was associated with OST. Low methylation of *CD1A* in Q1 had the lowest probability of OST compared to methylation in all other quartiles (p=0.002) (Figures 4.12 through 4.15). *CCNA1* was associated with RFT in these patients. High methylation in Q4 of *CCNA1* had the worst probability of RFT, while methylation in Q2 had the best probability of RFT. Methylation of this gene in Q1 and Q3 had similar probabilities of RFT.

No markers were associated with probability of OST within HPV (-) patients, however, *DCC* methylation was significantly associated with probability of RFT. The trend of the quartiles of *DCC* methylation was nonlinear; intermediate levels of methylation in Q2 and Q3 were associated with higher probabilities of RFT while the highest and lowest quartiles were associated with the lowest probabilities of RFT (p=0.03) (Figures 4.12 through 4.15).

NDN and CD1A methylation are novel markers of survival in HNSCC patients

Univariate cox proportional hazards models for survival show that patients in the 25th percentile of methylation of *CCNA1* had a 1.43 times higher odds of a death event than those in

the 75th percentile (95% CI: 1.04-1.96) (Table 4.3). Those with *NDN* methylation in the 25th percentile had almost twice the odds of a death event compared to those in the 75th percentile of methylation (95% CI: 1.42-2.71). Patients with both *CD1A* and *DCC* methylation in the 25th percentile had about 1.5 times higher odds of a death event than those in the 75th percentile of methylation (95% CI: 1.24-1.95; 95% CI: 1.10-1.98, respectively). Multivariable Cox proportional hazards models adjusting for site, stage, HPV status, age and comorbidity score showed that only significant associations of NDN and CD1A gene methylation with survival persisted in the total population (Table 4.3). Those with *NDN* methylation in the 25th percentile had a 1.6 times higher odds of a death event compared to those with methylation in the 75th percentile (95% CI: 1.13-2.39). Patients with *CD1A* methylation in the 25th percentile had a 1.3 times higher odds of a death event compared to those with methylation in the 75th percentile upon adjustment (95% CI: 1.01-1.71). These results indicate that for the total population, hypermethylation of *NDN* and *CD1A* are associated with better patient survival.

Univariate cox proportional hazards models for recurrence reveal that patients with *NDN* methylation in the 25th percentile had a 1.6 times higher odds of a recurrence event than those in the 75th percentile (95% CI: 1.17-2.17) (Table 4.3). Those that have *CD1A* methylation and *DCC* in the 25th percentile had about 1.4 and 1.46 times higher odds of a recurrence event, respectively, as compared to those in the 75th percentile (95% CI: 1.08-1.73; 95% CI: 1.09-1.94). These associations did not persist upon adjustment for other variables.

Novel methylation markers of survival and recurrence differ by HPV status

To determine the extent to which HPV status may play a role in these findings, stratified Cox models were used to measure associations separately within each groups (Table 4.3).

Divergent associations were found in patients based on their HPV status. In HPV (+) patients, those with *CD1A* methylation in the 25th percentile had 3.34 times higher odds of a death event than those in the 75th percentile (95% CI: 1.88-5.93). The same patients had twice the odds of a recurrent event if they fell into the 25th percentile of *CD1A* methylation compared to the 75th percentile. These results indicate that hypermethylation of *CD1A* is associated with better survival and lower recurrence. Conversely, HPV (+) patients had a 0.31 times lower odds of a recurrent event comparing those in the 25th percentile to those in the 75th percentile of *CCNA1* methylation, indicating that hypomethylation of this gene is associated with lower recurrence.

Within HPV (-) patients, *NDN* was significantly associated with survival. Patients with *NDN* methylation in the 25th percentile had 1.58 times higher odds of a death event compared to those in the 75th percentile of methylation (95% CI: 1.03-2.42), indicating hypermethylation of this gene is associated with better survival in HPV (-) patients only.

Methylation Score indicates Survival and Recurrence

Methylation score was used as a composite measurement of overall methylation status across all genes. A higher methylation score indicated tumors were methylated across all six genes. When patients were separated out either by their composite score or by quartiles of distribution of composite scores, the proportion of HPV (-) patients decreased and the proportion of HPV (+) patients increased with increasing score/quartile (Figure 4.16a and 4.16b). These results indicate that HPV (+) tumors tend to be more methylated than HPV (-) tumors.

Probability of OST was greatest for patients with the highest methylation scores and decreased with each score (Figure 4.16c). The same trend was seen for probability of RFT

(Figure 4.16d). These results indicate that overall methylation plays an important role in survival and recurrence.

Tumor Methylation and Epidemiologic Characteristics

Methylation is associated with epidemiologic characteristics

Each methylation marker was significantly associated with age and disease site (Table 4.4). HPV status was significantly associated with several markers. Methylation of *CCNA1*, *NDN*, *CD1A*, and *DCC* was higher while methylation of p16 was lower in HPV (+) tumors compared to HPV (-) tumors. Tobacco status was significantly associated with methylation of *NDN*, *CD1A*, and *DCC* (p=0.02; p<0.001; p=0.03, respectively). Interestingly former smokers had lower methylation of these genes compared to never or current smokers. Methylation of *NDN* was higher in patients with higher stage tumors (p<0.001), while a trend for *CCNA1* methylation with stage was U-shaped, with higher methylation at stages I and IV (p=0.01). As comorbidity status increased, methylation of *CD1A* increased (p-value=0.006). Finally, methylation of *NDN* and *DCC* was significantly higher in males compared to females (p =0.003; p =0.002, respectively).

Stratification by HPV status across these epidemiologic characteristics revealed HPV (+) tumors were hypermethylated across *CCNA1*, *NDN*, *CD1A*, and *DCC*, but hypomethylated in *p16* as compared to HPV(-) tumors, regardless of epidemiological trait (data not shown).

DISCUSSION

Changes in methylation patterns are one of the most frequent events in human tumors and epigenetic alterations have been increasingly recognized to play a role in the complex

mechanism of head and neck carcinogenesis. This study is the first to describe novel epigenetic alterations associated with survival in an unselected cohort of patients with HNSCC. DNA hypermethylation of *NDN* and *CD1A* was found to be significantly associated with improved overall survival time in the total study population. Stratification revealed major differences in these associations with survival by HPV status and identified novel methylation markers, including *CCNA1*, to be associated with recurrence in HPV (+) patients only. These unique discoveries raise significant new questions about why these specific epigenetic changes differ among biologically distinct subsets of HNSCC patients (HPV + versus HPV -) and whether these differences are in some mechanistic way linked to HPV status or are simply independent gene factors that are hypermethylated due to other unrelated reasons.

NDN is a maternally imprinted gene that has monoallelic expression. It encodes necdin, a protein that interacts with p53 to induce cell cycle arrest[47]. Differences in Kaplan-Meier survival curves were seen across quartiles of *NDN* methylation, however the trend was not linear. Those in the lowest and third quartile of methylation had better survival than those in the second and fourth quartiles. Although necdin is a p53 target gene involved in cell growth arrest and found to be hypermethylated in cancers, it has recently been implicated to act more as a "switch", promoting quiescence in the steady state but suppressing p53-dependent apoptosis in a stressful state[47-50]. The methylation trend by quartiles may reflect this "switch"-like behavior or could reflect some unknown biologic heterogeneity within the clinical characteristics of patients in these quartiles. Our adjusted Cox model analyses revealed hypermethylation of *NDN* was associated with better survival in the total population, and strongly, in HPV (-) patients specifically, where p53 status is a significant prognostic factor. Although precise mechanisms remain unknown, it would be anticipated that methylation would reduce NDN regulation of p53.

Previous literature has implicated necdin as a putative tumor suppressor; however, our results suggest that this multifunctional protein may be involved in other pathways that facilitate tumorigenesis, making hypermethylation of this gene beneficial for patient survival.

CD1A encodes an immune protein responsible for presenting antigens by dendritic Langerhans cells to T lymphocytes such as natural killer cells. Infiltration of tumors by CD1a positive cells has been associated with aggressive behavior of oral cavity cancers[51, 52]. The difference in Kaplan-Meier survival curves across quartiles of *CD1A* methylation showed that as methylation increased, probability of overall survival increased in the total population and in HPV (+) patients, specifically. This was validated in our adjusted Cox model analyses that revealed hypermethylation was associated with better overall survival for the total population and, for better survival and lower recurrence in HPV (+) patients specifically. Regulation of CD1A involves many factors associated with immune homeostasis in the tumor microenvironment, such as cytokine production[53]. Hypermethylation of this gene may be indicative of a failure in activation of immune suppressive myeloid cells in the tumor environment contributing to better patient survival rather than changes in this gene in the tumor cell component. Likewise, altered regulation of antigen presentation may induce chronic inflammatory responses that which may support immune mechanisms by this gene, which would be protective for survival. This is particularly attractive speculation in view of differences with respect to HPV status in our study and recent correlations and important differences in immunologic status in the patient's peripheral blood and tumor microenvironment according to HPV status[54, 55]. Patients with non-HPV related HNSCC tend to be significantly immunosuppressed and this immune suppression associated with worse outcomes.

Hypomethylation of *CCNA1* was associated with lower odds of recurrent events in HPV (+) patients only. *CCNA1* a cell cycle regulator that binds to retinoblastoma, E2F transcription factor and p21 family proteins to promote cell cycle progression. This pathway is particularly important regulator of cell proliferation in HPV infection. Our results suggest that decreased methylation of this cell cycle regulator is protective against recurrence events. Our findings are consistent with previous findings in HPV (+) HNSCC patients[27, 56]. Recently, a study found that HPV induced cyclin A1 overexpression could occur despite promoter hypermethylation, and this overexpression was associated with a lower recurrence rate[56].

Considering *CD1A* is an immune gene and *CCNA1* is involved in response pathways to HPV infection, it is within reason that these genes would play a larger role in cancers with a viral etiology. On the other hand, *NDN*, which is involved in cell cycle regulation via p53 pathways, would more likely be involved in regulation of altered cell cycle pathways due to mutation events from chemical carcinogenesis. Clearly, HPV (-) cancers are much more likely to harbor p53 mutations and inactivations while wild type p53 is a common finding in HPV (+) cancers that are associated with better outcomes. It remains unclear if these significant associations with survival are due to treatment efficacy or reflect epigenetic variations due to co-morbidities or health behaviors. In contrast to other reports from large outcome studies, there were not significant differences in survival outcomes with respect to comorbidities suggesting that the epigenetic changes more likely were reflective of differences in tumor biology rather than co-morbidity. Larger epigenetic studies of smoking, diet and comorbidities in our group are currently underway.

Epidemiologic and clinical characteristics have generally been used to understand cancer phenotype, determine prognosis and inform treatment plans for patients. Epidemiologic factors

such as smoking history, nutrition, and comorbidity are well known significant prognostic factors for overall survival and indicate the importance of including such factors in studies of new molecular markers. In the last decade, the clinical importance of better understanding tumor biology has emerged through validation of HPV status as a significant molecular predictor of patient survival and recurrence. Differences in patient outcomes according to HPV status are so dramatic that many investigators believe they reflect a new and unique phenotype that could justify significant de-intensification of therapy. The novel methylation markers identified in this study offer new, specific, epigenetic molecular differences within the setting of the generalized hypermethylation phenotype associated with HPV status and warrant further investigation. The findings support biological implications of epigenetic markers on patient survival and their potential usefulness in identifying unique subsets of patients with varied outcomes. Several markers show expected associations with patient characteristics.

Hypomethylation of the immune marker, *CD1A*, is found in all HPV (-) tumors, but most hypomethylated in laryngeal tumors. Due to this gene's involvement in immune function, it is not surprising to see that this gene is indicative of comorbidity status. HPV status and disease site are associated with methylation of *p16*, which encodes a cell cycle regulator and known tumor suppressor. Compared to HPV (+) tumors where this gene is expressed and the protein is inactivated, *p16* is generally lost through chromosomal deletion, mutation or promoter hypermethylation in HPV (-) tumors[57-59]. This gene was hypermethylated in our HPV (-) patients and most methylated in oral cavity tumors compared with other sites, potentially as another mechanism by which to inactivate expression of this protein. *DCC* encodes a tumor suppressor and methylation of this gene is associated with HPV status, site, tobacco status and gender. This gene is hypomethylated in HPV (-) patients and concurrently, is the least methylated in oral

cavity tumors. However, this gene, located on 18q is commonly found to deleted due to loss of heterozygosity (LOH) in HNSCC and therefore, may be inactivated through this pathway in HPV (-) patients[60]. It is hypermethylated in HPV (+) patients, which is consistent with previous literature in HNSCC, although many previous studies did not examine *DCC* methylation by HPV status[32, 61-63]. *CCNA1* was hypermethylated in our HPV (+) patients and was the most methylated in oropharyngeal tumors. This is consistent with previous findings in HPV (+) HNSCC patients[27, 56]. Recently, a study found that HPV induced cyclin A1 overexpression despite promoter hypermethylation, and this overexpression was associated with a lower recurrence rate[56]. *CCNA1* was also associated with stage, specifically hypomethylation in stages II and III, perhaps indicating that this gene is important in progression of tumorigenesis. *NDN* was hypomethylated in HPV (-) tumors, but both oral cavity and larynx tumors were the least methylated sites. As stage increased, *NDN* methylation increased, indicating inhibition of this gene with progression of tumorigenesis. Promoter methylation of *NDN* was associated with the most epidemiological characteristics, making it a suitable marker to represent these traits.

Overall, HPV (+) tumors tended to be more methylated as shown by composite methylation scores. Highly methylated tumors had a higher probability of OST compared to lowly methylated tumors, as expected. The association of methylation score with probability of RFT followed this trend as well, indicating that general methylation status of a tumor is indicative of survival and recurrence.

Several of associations of methylation marks with characteristics such as stage, gender and tobacco status were novel. *NDN*, *DCC*, and *CD1A* all showed former smokers had lower methylation of these genes compared to never and current smokers. Because total pack years is inversely associated with these genes, perhaps the reason why former smokers stand out is

duration of exposure or exposure at an early age is integral in the initiation of processes that permits carcinogenesis.

CONCLUSIONS

Our cohort shows the expected associations established in previous literature, such as the relationships between stage, site, and HPV status with overall survival time and the expected population characteristics of a HNSCC cohort established by previous studies. For example, this study population consists of a majority of male patients, a higher number of cases diagnosed in later stages, larger proportion of HPV (-) tumors, and a mean age of about 60 years. Separation by HPV status reveals characteristics that are also expected from etiologic differences. HPV (+) patients are slightly younger, most HPV (-) tumors are in the oral cavity while most HPV (+) tumors are in the oropharynx, pack years are higher in HPV (-) patients and HPV (+) patients have longer recurrence free and overall survival times. These attributes provides assurance that the new associations discovered with this cohort are meaningful and can be extrapolated to the patient population.

References

[1]Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. *Nature reviews Cancer* 2011;**11**: 9-22.

[2]Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA: a cancer journal for clinicians* 2005;**55**: 74-108.

[3]Hoffman HT, Karnell LH, Funk GF, Robinson RA, Menck HR. The National Cancer Data Base report on cancer of the head and neck. *Archives of otolaryngology--head & neck surgery* 1998;**124**: 951-62.

[4]Wise-Draper TM, Draper DJ, Gutkind JS, Molinolo AA, Wikenheiser-Brokamp KA, Wells SI. Future directions and treatment strategies for head and neck squamous cell carcinomas. *Translational research : the journal of laboratory and clinical medicine* 2012;**160**: 167-77.

[5]Forastiere AA, Ang KK, Brizel D, Brockstein BE, Burtness BA, Cmelak AJ, Colevas AD, Dunphy F, Eisele DW, Goepfert H, Hicks WL, Jr., Kies MS, et al. Head and neck cancers. *Journal of the National Comprehensive Cancer Network : JNCCN* 2008;**6**: 646-95.

[6]Jung AC, Job S, Ledrappier S, Macabre C, Abecassis J, de Reynies A, Wasylyk B. A poor prognosis subtype of HNSCC is consistently observed across methylome, transcriptome, and miRNome analysis. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013;**19**: 4174-84.

[7]Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: a virus-related cancer epidemic. *The lancet oncology* 2010;**11**: 781-9.

[8]Gillison ML. Human papillomavirus and prognosis of oropharyngeal squamous cell carcinoma: implications for clinical research in head and neck cancers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2006;**24**: 5623-5.

[9]Fakhry C, Gillison ML. Clinical implications of human papillomavirus in head and neck cancers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2006;**24**: 2606-11.

[10]Joseph AW, D'Souza G. Epidemiology of human papillomavirus-related head and neck cancer. *Otolaryngologic clinics of North America* 2012;**45**: 739-64.

[11]Psyrri A, Seiwert TY, Jimeno A. Molecular pathways in head and neck cancer. *American Society of Clinical Oncology educational book / ASCO American Society of Clinical Oncology Meeting* 2013: 246-55.

[12]Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, Viscidi R. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *Journal of the National Cancer Institute* 2008;**100**: 407-20.

[13]Chaturvedi AK, Engels EA, Anderson WF, Gillison ML. Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008;**26**: 612-9.

[14]Gillison ML, Lowy DR. A causal role for human papillomavirus in head and neck cancer. *Lancet* 2004;**363**: 1488-9.

[15]Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W, Liu L, Lynch CF, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011;**29**: 4294-301.

[16]Ragin CC, Modugno F, Gollin SM. The epidemiology and risk factors of head and neck cancer: a focus on human papillomavirus. *Journal of dental research* 2007;**86**: 104-14.

[17]Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, Zahurak ML, Daniel RW, Viglione M, Symer DE, Shah KV, Sidransky D. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *Journal of the National Cancer Institute* 2000;**92**: 709-20.

[18]Schwartz SR, Yueh B, McDougall JK, Daling JR, Schwartz SM. Human papillomavirus infection and survival in oral squamous cell cancer: a population-based study. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* 2001;**125**: 1-9.

[19]Weinberger PM, Yu Z, Haffty BG, Kowalski D, Harigopal M, Brandsma J, Sasaki C, Joe J, Camp RL, Rimm DL, Psyrri A. Molecular classification identifies a subset of human papillomavirus--associated oropharyngeal cancers with favorable prognosis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2006;**24**: 736-47.

[20]Ashman JN, Patmore HS, Condon LT, Cawkwell L, Stafford ND, Greenman J. Prognostic value of genomic alterations in head and neck squamous cell carcinoma detected by comparative genomic hybridisation. *British journal of cancer* 2003;**89**: 864-9.

[21]Bauer VL, Braselmann H, Henke M, Mattern D, Walch A, Unger K, Baudis M, Lassmann S, Huber R, Wienberg J, Werner M, Zitzelsberger HF. Chromosomal changes characterize head and neck cancer with poor prognosis. *J Mol Med (Berl)* 2008;**86**: 1353-65.

[22]Misawa Y, Misawa K, Kanazawa T, Uehara T, Endo S, Mochizuki D, Yamatodani T, Carey TE, Mineta H. Tumor suppressor activity and inactivation of galanin receptor type 2 by aberrant promoter methylation in head and neck cancer. *Cancer* 2013.

[23]Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X, Shockley WW, Weissler MC, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer cell* 2004;**5**: 489-500.

[24]Walter V, Yin X, Wilkerson MD, Cabanski CR, Zhao N, Du Y, Ang MK, Hayward MC, Salazar AH, Hoadley KA, Fritchie K, Sailey CG, et al. Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. *PloS one* 2013;**8**: e56823.

[25]Baylin SB, Herman JG, Graff JR, Vertino PM, Issa JP. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Advances in cancer research* 1998;**72**: 141-96.

[26]Laytragoon-Lewin N, Chen F, Castro J, Elmberger G, Rutqvist LE, Lewin F, Turesson I, Lundgren J. DNA content and methylation of p16, DAPK and RASSF1A gene in tumour and distant, normal mucosal tissue of head and neck squamous cell carcinoma patients. *Anticancer research* 2010;**30**: 4643-8.

[27]Colacino JA, Dolinoy DC, Duffy SA, Sartor MA, Chepeha DB, Bradford CR, McHugh JB, Patel DA, Virani S, Walline HM, Bellile E, Terrell JE, et al. Comprehensive analysis of DNA methylation in head and neck squamous cell carcinoma indicates differences by survival and clinicopathologic characteristics. *PloS one* 2013;**8**: e54742.

[28]De Schutter H, Geeraerts H, Verbeken E, Nuyts S. Promoter methylation of TIMP3 and CDH1 predicts better outcome in head and neck squamous cell carcinoma treated by radiotherapy only. *Oncology reports* 2009;**21**: 507-13.

[29]Rettori MM, de Carvalho AC, Bomfim Longo AL, de Oliveira CZ, Kowalski LP, Carvalho AL, Vettore AL. Prognostic significance of TIMP3 hypermethylation in post-treatment salivary rinse from head and neck squamous cell carcinoma patients. *Carcinogenesis* 2013;**34**: 20-7.

[30]Misawa K, Kanazawa T, Misawa Y, Uehara T, Imai A, Takahashi G, Takebayashi S, Cole A, Carey TE, Mineta H. Galanin has tumor suppressor activity and is frequently inactivated by aberrant promoter methylation in head and neck cancer. *Translational oncology* 2013;**6**: 338-46.

[31]Worsham MJ, Chen KM, Ghanem T, Stephen JK, Divine G. Epigenetic Modulation of Signal Transduction Pathways in HPV-Associated HNSCC. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* 2013;**149**: 409-16.

[32]Schussel J, Zhou XC, Zhang Z, Pattani K, Bermudez F, Jean-Charles G, McCaffrey T, Padhya T, Phelan J, Spivakovsky S, Brait M, Li R, et al. EDNRB and DCC salivary rinse hypermethylation has a similar performance as expert clinical examination in discrimination of oral cancer/dysplasia versus benign lesions. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013;**19**: 3268-75.

[33]Ha PK, Califano JA. Promoter methylation and inactivation of tumour-suppressor genes in oral squamous-cell carcinoma. *The lancet oncology* 2006;**7**: 77-82.

[34]Hwang S, Mahadevan S, Qadir F, Hutchison IL, Costea DE, Neppelberg E, Liavaag PG, Waseem A, Teh MT. Identification of FOXM1-induced epigenetic markers for head and neck squamous cell carcinomas. *Cancer* 2013.

[35]Langevin SM, Koestler DC, Christensen BC, Butler RA, Wiencke JK, Nelson HH, Houseman EA, Marsit CJ, Kelsey KT. Peripheral blood DNA methylation profiles are indicative of head and neck squamous cell carcinoma: an epigenome-wide association study. *Epigenetics : official journal of the DNA Methylation Society* 2012;**7**: 291-9.

[36]Bennett KL, Lee W, Lamarre E, Zhang X, Seth R, Scharpf J, Hunt J, Eng C. HPV status-independent association of alcohol and tobacco exposure or prior radiation therapy with promoter methylation of FUSSEL18, EBF3, IRX1, and SEPT9, but not SLC5A8, in head and neck squamous cell carcinomas. *Genes, chromosomes & cancer* 2010;**49**: 319-26.

[37]Shaw RJ, Liloglou T, Rogers SN, Brown JS, Vaughan ED, Lowe D, Field JK, Risk JM. Promoter methylation of P16, RARbeta, E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing. *British journal of cancer* 2006;**94**: 561-8.

[38]Poage GM, Butler RA, Houseman EA, McClean MD, Nelson HH, Christensen BC, Marsit CJ, Kelsey KT. Identification of an epigenetic profile classifier that is associated with survival in head and neck cancer. *Cancer research* 2012;**72**: 2728-37.

[39]Marsit CJ, Christensen BC, Houseman EA, Karagas MR, Wrensch MR, Yeh RF, Nelson HH, Wiemels JL, Zheng S, Posner MR, McClean MD, Wiencke JK, et al. Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma. *Carcinogenesis* 2009;**30**: 416-22.

[40]Mikeska T, Bock C, Do H, Dobrovic A. DNA methylation biomarkers in cancer: progress towards clinical implementation. *Expert review of molecular diagnostics* 2012;**12**: 473-87.

[41]Tang AL, Hauff SJ, Owen JH, Graham MP, Czerwinski MJ, Park JJ, Walline H, Papagerakis S, Stoerker J, McHugh JB, Chepeha DB, Bradford CR, et al. UM-SCC-104: a new human papillomavirus-16-positive cancer stem cell-containing head and neck squamous cell carcinoma cell line. *Head & neck* 2012;**34**: 1480-91.

[42]Kumar B, Cordell KG, Lee JS, Worden FP, Prince ME, Tran HH, Wolf GT, Urba SG, Chepeha DB, Teknos TN, Eisbruch A, Tsien CI, et al. EGFR, p16, HPV Titer, Bcl-xL and p53, sex, and smoking as indicators of response to therapy and survival in oropharyngeal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008;**26**: 3128-37.

[43]Worden FP, Kumar B, Lee JS, Wolf GT, Cordell KG, Taylor JM, Urba SG, Eisbruch A, Teknos TN, Chepeha DB, Prince ME, Tsien CI, et al. Chemoselection as a strategy for organ preservation in advanced oropharynx cancer: response and survival positively associated with HPV16 copy number. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008;**26**: 3138-46.

[44]Maxwell JH, Kumar B, Feng FY, McHugh JB, Cordell KG, Eisbruch A, Worden FP, Wolf GT, Prince ME, Moyer JS, Teknos TN, Chepeha DB, et al. HPV-positive/p16-positive/EBV-negative nasopharyngeal carcinoma in white North Americans. *Head & neck* 2010;**32**: 562-7.

[45]Maxwell JH, Kumar B, Feng FY, Worden FP, Lee JS, Eisbruch A, Wolf GT, Prince ME, Moyer JS, Teknos TN, Chepeha DB, McHugh JB, et al. Tobacco use in human papillomavirus-positive advanced oropharynx cancer patients related to increased risk of distant metastases and tumor recurrence. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2010;**16**: 1226-35.

[46]Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nature protocols* 2007;**2**: 2265-75.

[47]De Faveri LE, Hurst CD, Platt FM, Taylor CF, Roulson JA, Sanchez-Carbayo M, Knowles MA, Chapman EJ. Putative tumour suppressor gene necdin is hypermethylated and mutated in human cancer. *British journal of cancer* 2013;**108**: 1368-77.

[48]Asai T, Liu Y, Nimer SD. Necdin, a p53 target gene, in normal and cancer stem cells. *Oncotarget* 2013.

[49]Lafontaine J, Rodier F, Ouellet V, Mes-Masson AM. Necdin, a p53-target gene, is an inhibitor of p53-mediated growth arrest. *PloS one* 2012;**7**: e31916.

[50]Lafontaine J, Tchakarska G, Rodier F, Mes-Masson AM. Necdin modulates proliferative cell survival of human cells in response to radiation-induced genotoxic stress. *BMC cancer* 2012;**12**: 234.

[51]Goldman SA, Baker E, Weyant RJ, Clarke MR, Myers JN, Lotze MT. Peritumoral CD1a-positive dendritic cells are associated with improved survival in patients with tongue carcinoma. *Archives of otolaryngology--head & neck surgery* 1998;**124**: 641-6.

[52]Goncalves AS, Costa NL, Arantes DA, de Cassia Goncalves Alencar R, Silva TA, Batista AC. Immune response in cervical lymph nodes from patients with primary oral squamous cell carcinoma. *Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* 2013;**42**: 535-40.

[53]Coventry B, Heinzel S. CD1a in human cancers: a new role for an old molecule. *Trends in immunology* 2004;**25**: 242-8.

[54]Wansom D, Light E, Worden F, Prince M, Urba S, Chepeha DB, Cordell K, Eisbruch A, Taylor J, D'Silva N, Moyer J, Bradford CR, et al. Correlation of cellular immunity with human papillomavirus 16 status and outcome in patients with advanced oropharyngeal cancer. *Archives of otolaryngology--head & neck surgery* 2010;**136**: 1267-73.

[55]Wansom D, Light E, Thomas D, Worden F, Prince M, Urba S, Chepeha D, Kumar B, Cordell K, Eisbruch A, Taylor J, Moyer J, et al. Infiltrating lymphocytes and human papillomavirus-16--associated oropharyngeal cancer. *The Laryngoscope* 2012;**122**: 121-7.

[56]Weiss D, Koopmann M, Basel T, Rudack C. Cyclin A1 shows age-related expression in benign tonsils, HPV16-dependent overexpression in HNSCC and predicts lower recurrence rate in HNSCC independently of HPV16. *BMC cancer* 2012;**12**: 259.

[57]Hafkamp HC, Speel EJ, Haesevoets A, Bot FJ, Dinjens WN, Ramaekers FC, Hopman AH, Manni JJ. A subset of head and neck squamous cell carcinomas exhibits integration of HPV 16/18 DNA and overexpression of p16INK4A and p53 in the absence of mutations in p53 exons 5-8. *International journal of cancer Journal international du cancer* 2003;**107**: 394-400.

[58]Reed AL, Califano J, Cairns P, Westra WH, Jones RM, Koch W, Ahrendt S, Eby Y, Sewell D, Nawroz H, Bartek J, Sidransky D. High frequency of p16 (CDKN2/MTS-1/INK4A) inactivation in head and neck squamous cell carcinoma. *Cancer research* 1996;**56**: 3630-3.

[59]Sartor MA, Dolinoy DC, Jones TR, Colacino JA, Prince ME, Carey TE, Rozek LS. Genome-wide methylation and expression differences in HPV(+) and HPV(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. *Epigenetics : official journal of the DNA Methylation Society* 2011;**6**: 777-87.

[60]Papadimitrakopoulou VA, Oh Y, El-Naggar A, Izzo J, Clayman G, Mao L. Presence of multiple incontiguous deleted regions at the long arm of chromosome 18 in head and neck cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 1998;**4**: 539-44.

[61]Carvalho AL, Chuang A, Jiang WW, Lee J, Begum S, Poeta L, Zhao M, Jeronimo C, Henrique R, Nayak CS, Park HL, Brait MR, et al. Deleted in colorectal cancer is a putative conditional tumor-suppressor gene inactivated by promoter hypermethylation in head and neck squamous cell carcinoma. *Cancer research* 2006;**66**: 9401-7.

[62]Park HL, Kim MS, Yamashita K, Westra W, Carvalho AL, Lee J, Jiang WW, Baek JH, Liu J, Osada M, Moon CS, Califano JA, et al. DCC promoter hypermethylation in esophageal squamous cell carcinoma. *International journal of cancer Journal international du cancer* 2008;**122**: 2498-502.

[63]Ogi K, Toyota M, Ohe-Toyota M, Tanaka N, Noguchi M, Sonoda T, Kohama G, Tokino T. Aberrant methylation of multiple genes and clinicopathological features in oral squamous cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2002;**8**: 3164-71.

Table 4.1: Patient Clinical and Epidemiological Characteristics, N=346

Characteristic		N	(%)	Mean (std), range
Age at Dx (years)				59.7 (11.5), 25-93
Gender	Male	259	75%	
	Female	87	25%	
Cancer Site	Larynx/Glottic	83	24%	
	Oral Cavity	126	36%	
	Oropharynx	125	36%	
	Hypopharynx	12	3%	
Cancer Stage	I/Cis	47	14%	
	II	35	10%	
	III	52	15%	
	IV	212	61%	
Comorbidities (ACE)	None	91	26%	
	Mild	158	46%	
	Moderate	70	20%	
	Severe	27	8%	
HPV status	Positive	135	39%	
	Negative	211	61%	
Tobacco Use	Current (within past 12 months)	145	42%	
	Former (quit > 12 months)	125	36%	
	Never	76	22%	
Pack Years, n=257	(cigs only)			35.6 (30.0), 0.1-171.0
Pack Years, n=264	(sum of cigs, cigars, pipe)			37.7 (33.3), 0.07-242.9
Alcohol Use, n=244	Never	29	8%	
	Former (quit >12 months)	227	66%	
	Current (within past 12 months)	90	26%	
Treatment, n=335	Surgery alone	67	20%	
	Radiation alone	34	10%	
	Surgery + Radiation	34	10%	
	Radiation + Chemotherapy	138	41%	
	Surgery + Radiation + Chemotherapy	39	12%	
	No Treatment prior to death	23	7%	
Persistent Disease[a]		29	8%	
Median Follow-up for Survival		27.4 months		
Median Follow-up for Recurrence		24.0 months		
Recurrence[b]		81		
Death		78		
KM estimate 2 year RFT[c]			75%	
KM estimate 2 year OST[d]			79%	

a). Disease considered persistent if patient never became disease free.
b). Recurrence of the HNSCC in the primary, regional and/or a distant location. Patients whose disease never cleared after treatment are included and considered recurrent with a recurrence time=1 day.
c). OST=Overall Survival Time. Death from any cause considered an event. Overall survival time defined from date of diagnosis by UM physician.
d). RFT=Recurrence Free Time defined as time from diagnosis to recurrence event or end of follow-up. End of follow-up is the last date where patient was reviewed for recurrence. Patients whose disease never cleared after treatment are considered recurrent with a recurrence time=1 day.

Table 4.2: Patient Clinical and Epidemiological Characteristics by HPV Status, N=346

		HPV (–)		HPV (+)		HPV (-)	HPV (+)	
Characteristic		N	(%) among HPV(-)	N	(%) among HPV(+)	Mean (SD)	Mean (SD)	*p*-value
Age at Dx (years)		211		135		61.4 (12.3)	57.0 (9.6)	0.0002
Gender	Male	139	66%	120	89%			<0.001
	Female	72	34%	15	11%			
Cancer Site	Larynx/Glottic	72	34%	11	8%			<0.001
	Oral Cavity	110	52%	16	12%			
	Oropharynx	20	9%	105	78%			
	Hypopharynx	9	4%	3	2%			
Cancer Stage	I/Cis	36	17%	11	8%			<0.001
	II	28	13%	7	5%			
	III	39	18%	13	10%			
	IV	108	51%	104	77%			
Comorbidities (ACE)	None	41	19%	50	37%			0.001
	Mild	99	47%	59	44%			
	Moderate	50	24%	20	15%			
	Severe	21	10%	6	4%			
Tobacco Use	Current (within past 12 months)	102	48%	43	32%			0.001
	Former (quit > 12 months)	75	36%	50	37%			
	Never	34	16%	42	31%			
Pack Years, n=257	(cigs only)					38.4 (25.6)	29.9 (28.8)	0.02
Pack Years, n=264	(cigs, cigars, pipe)					41.8 (34.5)	29.6 (29.4)	0.005
Alcohol Use, n=244	Current (within past 12 months)	63	30%	27	20%			0.01
	Former (quit > 12 months)	126	60%	101	75%			
	Never	22	10%	7	5%			
Treatment, n=335	Surg alone	53	26%	14	11%			<0.001
	Rad alone	23	11%	11	8%			
	Surg + Rad	30	15%	4	3%			
	Rad + Chemo	45	22%	93	70%			
	Surg + Rad + Chemo	29	14%	10	8%			
	No Treatment prior to death	22	11%	1	1%			
Post Treatment Status	Free of Disease	186	88%	131	97%			0.004
	Persistent Disease	25	12%	4	3%			
KM estimate 2 year RFT			69%		84%			0.001
KM estimate 2 year OST			71%		91%			<0.001

-Percentages will not add to 100% horizontally.
% presented is proportion of clinical or epidemiological subgroup among HPV status category.
-p-values calculated by t-test for continuous variables and chi-square test for categorical variables. Log-rank test used for KM analysis.

Table 4.3. Stratified Cox Proportional Hazards Analysis

			Univariable HR (95% CI)	Multivariable HR (95% CI)		
	Gene Methylation	n[a]	Univariable HR (95% CI)[c]	Multivariable HR (95% CI)[c,d]	Stratified Model Among HPV+[c,e]	Stratified Model Among HPV-[c,e]
OST	*CCNA1*[b]	341	1.43 (1.04, 1.96)*	1.13 (0.81, 1.59)	0.62 (0.25, 1.52)	1.25 (0.86, 1.81)
	NDN	346	1.96 (1.42, 2.71)***	1.64 (1.13, 2.39)**	1.89 (0.87, 4.11)	1.58 (1.03, 2.42)*
	CD1A[b]	342	1.55 (1.24, 1.95)***	1.31 (1.01, 1.71)*	3.34 (1.88, 5.93)**	1.13 (0.84, 1.51)
	DCC[b]	344	1.48 (1.10, 1.98)**	1.30 (0.95, 1.78)	1.45 (0.59, 3.60)	1.28 (0.91, 1.79)
	p16[b]	343	1.005 (0.84, 1.18)	1.11 (0.93, 1.32)	0.76 (0.48, 1.19)	1.17 (0.97, 1.42)
	GADD45[b]	340	1.05 (0.80, 1.37)	0.88 (0.66, 1.17)	0.80 (0.43, 1.50)	0.91 (0.66, 1.25)
RFT	*CCNA1*[b]	334	1.06 (0.77, 1.47)	0.81 (0.57, 1.14)	0.31 (0.13, 0.74)*	0.96 (0.66, 1.42)
	NDN	339	1.59 (1.17, 2.17)**	1.40 (0.98,2.00)	1.42 (0.73, 2.78)	1.38 (0.90, 2.11)
	CD1A[b]	335	1.37 (1.08, 1.73)**	1.12 (0.95, 1.62)	2.06 (1.21, 3.49)**	1.07 (0.79, 1.46)
	DCC[b]	337	1.45 (1.09, 1.94)*	1.32 (0.96, 1.83)	0.97 (0.40, 2.34)	1.37 (0.97, 1.94)
	p16[b]	336	0.92 (0.79, 1.08)	1.00 (0.84, 1.19)	0.91 (0.57, 1.46)	1.00 (0.84, 1.20)
	GADD45[b]	333	1.26 (0.97, 1.62)	1.12 (0.87, 1.46)	0.84 (0.49, 1.42)	1.23 (0.91, 1.66)

* p<0.05, ** p<0.01, ***p<0.001.

a) n=number of subjects in models

b) Methylation values are log-transformed to adhere to normality assumption of Cox proportional hazards models.

c) All methylation measurements standardized to interquartile ranges. HRs are interpreted as comparison between those with methylation in the 25th percentile compared to those with methylation in the 75th percentile

d)) Multivariable HR calculated from a multivariable Cox Proportional Hazards model after controlling for: age, HPV status, disease site, stage, and comorbidity score in the OST model and age, HPV status, disease site and stage in the RFT model.

e) Multivariable HR (95% CI) calculated from a stratified multivariable Cox Proportional Hazards model. Models are stratified by HPV status to allow different baseline hazards for HPV+ and HPV- groups and also control for HPV status with a main effect and interaction with methylation marker. OST models control for: age, disease site, stage, and comorbidity score and RFT models control for: age, disease site and stage.

Table 4.4. Methylation Markers by Epidemiological Characteristics												
	CCNA1		NDN		CD1A		DCC		p16		GADD45	
	median (range)	*p*	*median (range)*	*p*	*median (range)*	*p*	*median (range)*	*p*	*median (range)*	*p*	*median (range)*	*p*
Age	26 (3, 90.5)	<0.001	44.2 (10.8, 84.9)	<0.001	69.8 (21.4, 93.9)	<0.001	33.0 (3.0, 85.8)	<0.001	2.3 (0, 70.1)	<0.001	1.4 (0, 16)	<0.001
HPV												
Positive	33.3 (3.0, 76.0)	<0.001	52.7 (14.0, 84.9)	<0.001	77.6 (25.1, 93.9)	<0.001	44.7 (3.1, 85.8)	<0.001	1.9 (0, 59.8)	<0.001	1.4 (0.4, 3.9)	0.8
Negative	20.9 (5.5, 90.5)		40.0 (10.8, 80.3)		65.4 (21.4 (92.2)		26.0 (3.0, 83.3)		2.5 (0, 70.1)		1.4 (0, 16)	
Tobacco												
Current	25.5 (6.3, 76.0)		45.0 (10.8, 84.9)		69.5 (25.1, 91.7)		39.5 (4.2, 83.3)		2.2 (0, 70.1)		1.4 (0, 16)	
Former	25.6 (6.8, 77.0)	0.8	42.2 (11.2, 80.3)	0.02	64.4 (21.4, 90.0)	<0.001	29.3 (3.0, 85.8)	0.03	2.4 (0, 46.6)	0.7	1.4 (0, 3.9)	0.5
Never	26.6 (3, 90.5)		46.4 (14.0, 84.2)		79.7 (47.2, 93.9)		35.7 (3.1, 79.8)		2.2 (0, 26.9)		1.5 (0.3, 3.7)	
Site												
Larynx	24.5 (8.0, 90.5)		41.2 (11.2, 84.9)		56.2 (21.6, 92.2)		26.3 (3.1, 83.3)		2.4 (0, 51.8)		1.4 (0, 3.9)	
Oral Cavity	22.0 (5.5, 77.0)		40.0 (10.8, 62.8)		70.6 (21.4, 90.2)		28.4 (3.0, 67.2)		2.6 (0, 70.1)		1.4 (0, 16.0)	
Oropharynx	32.3 (3.0, 76.0)	<0.001	52.7 (21.8, 84.2)	<0.001	77.5 (21.6, 93.9)	<0.001	42.5 (6.1, 81.6)	<0.001	2.0 (0, 39.9)	<0.001	1.4 (0.4, 3.9)	0.02
Hypopharynx	15.6 (9.0, 64.0)		44.2 (17.2, 91.0)		60.5 (25.1, 90.6)		54.5 (12.3, 85.8)		2.6 90, 59.8)		1.1 (0, 1.6)	
Stage												
I/Cis	27.5 (11.8, 77.0)		38.6 (11.2, 80.3)		68.1 (30.3, 89.4)		26.7 (3.0, 67.2)		2.5 (0, 39.0)		1.3 (0, 3.5)	
II	22.4 (12, 74.3)	0.01	41.2 (17.2, 84.2)	<0.001	68.8 (21.6, 88.3)	0.8	40.8 (7.9, 83.3)	0.1	2.3 (0, 59.8)	0.1	1.4 (0.6, 4.6)	0.3
III	21.3 (6.8, 61.3)		40.4 (15.3, 69.7)		71.2 (21.4, 90.2)		28.8 (3.9, 75.1)		2.6 (0, 70.1)		1.5 (0, 16.0)	
IV	27.6 (3.0, 90.5)		46.8 (10.8, 84.9)		70.5 (21.6, 93.9)		35.9 (3.1, 85.8)		2.2 (0, 51.8)		1.4 (0, 3.9)	
Comorbidities												
None	29.5 (6.3, 74.3)		46.0 (20.2, 84.2)		74.8 (32.4, 93.9)		33.1 (5.6, 75.2)		2.5 (0, 39.9)		1.5 (0.4, 4.6)	
Mild	25.5 (3.0, 77.0)	0.4	43.3 (10.8, 84.9)	0.3	69.5 (21.6, 92.0)	0.006	35.4 (3.0, 85.8)	0.6	2.2 (0, 70.1)	0.4	1.4 (0, 16.0)	0.1
Moderate	23.0 (6.8, 90.5)		44.0 (15.0, 79.3)		69.2 (21.4, 92.2)		29.8 (3.5, 79.8)		2.2 (0, 59.8)		1.3 (0, 3.9)	
Severe	30.5 (13.5,63.3)		44.6 (11.2, 80.3)		65.4 (21.6, 85.1)		43.0 (6.2, 83.3)		1.8 (0, 51.8)		1.2 (0, 3.9)	
Gender												
Male	26.5 (3.0, 90.5)	0.2	45.6 (13.0, 84.9)	0.003	69.9 (21.4, 93.9)	0.4	36.1 (3.1, 85.8)	0.002	2.3 (0, 70.1)	0.9	1.4 (0, 3.9)	0.5
Female	23.8 (6.3, 77.0)		40.5 (10.8, 68.9)		69.5 (21.6, 91.7)		25.4 (3.0, 75.1)		2.1 (0, 46.6)		1.4 (0, 16.0)	

Figure 4.1. Tumor sites of the head and neck

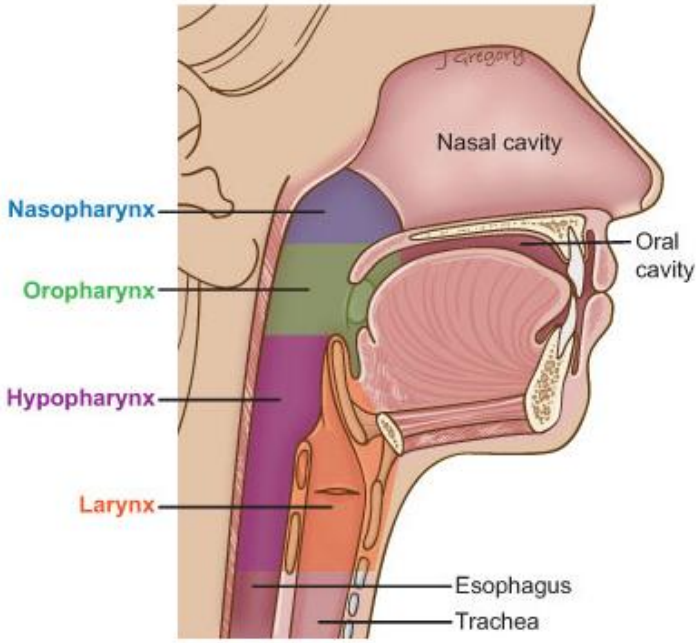


Figure 4.2 Five-year survival by year of diagnosis and tumor site

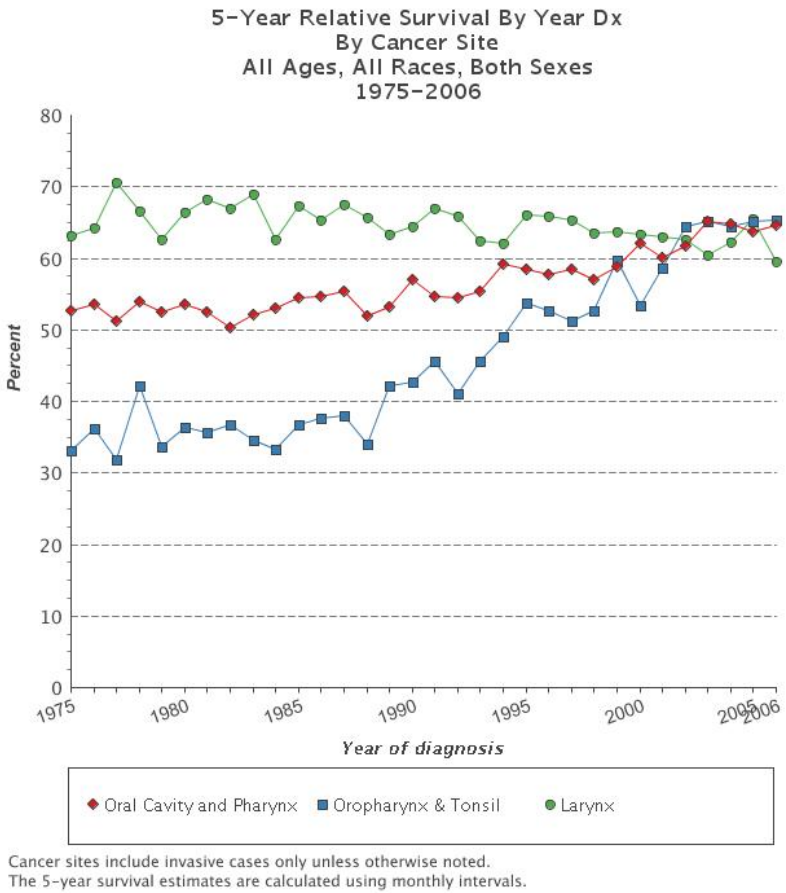


Figure adapted from the National Cancer Institute

Figure 4.3. Distinct clinicopathology of HPV-positive and HPV-negative HNSCCs

Characteristic	HPV-Positive	HPV-Negative
Epidemiology		
Age	Younger	Older
Gender	3:1 Men	3:1 Men
SES	High	Low
Risk Factors	Sexual Behavior	Alcohol and Tobacco
Incidence	Increasing	Decreasing
Clinical		
Anatomic Site	Oropharynx	All sites
Histology	Basaloid	Keratinized
Survival	Improved	Decreasing
Response to Tx	Better	Worse
Recurrence	Lower risk	Higher risk

Figure 4.4. Age is significantly associated with survival. a) The oldest patients have the poorest probability of OST compared to ages in all other quartiles; b) Probability of RFT increases with decreasing quartiles of age

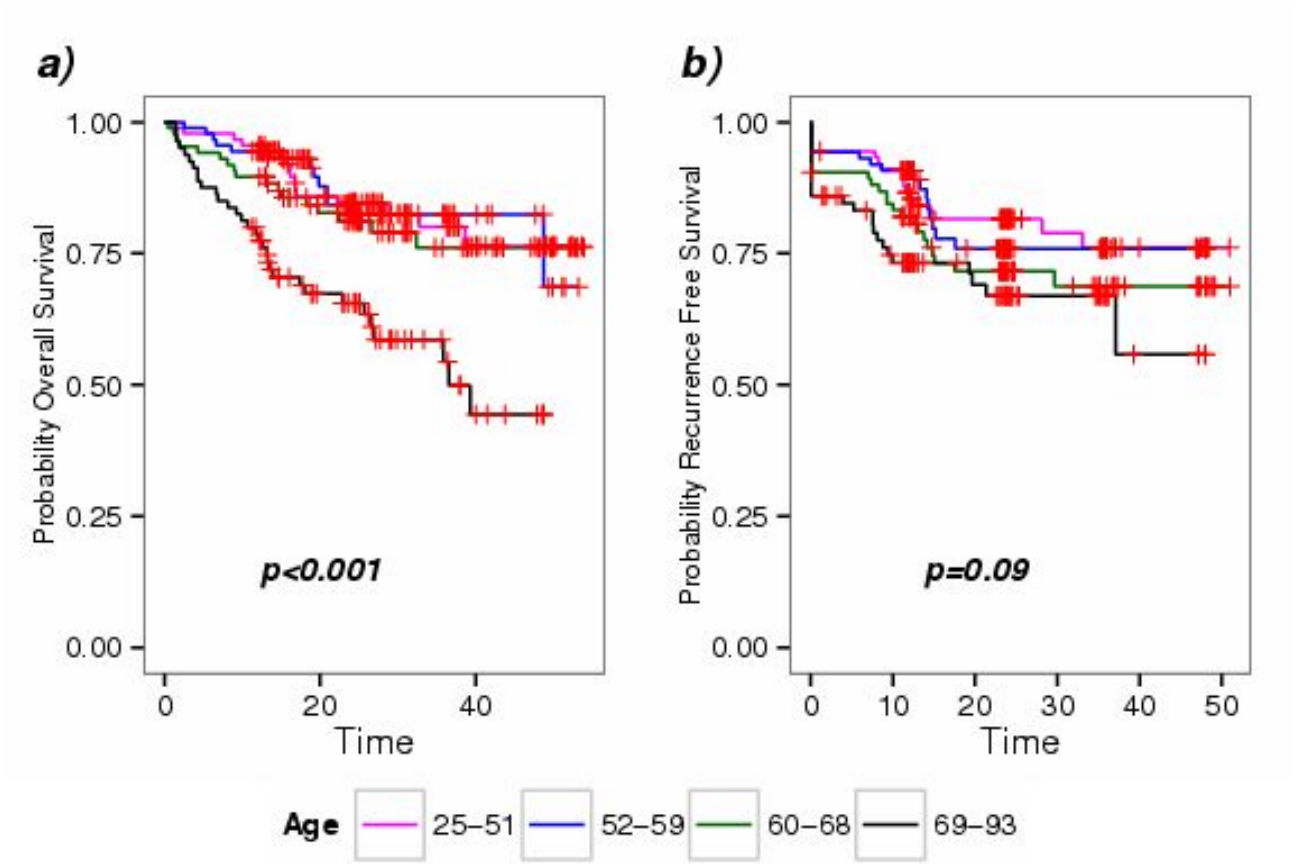


Figure 4.5. HPV(+) patients have significantly better prognosis compared to HPV (-) patients. a) HPV+ patients have a significantly higher probability of OST; b) HPV(+) patients have a significantly higher probability of RFT

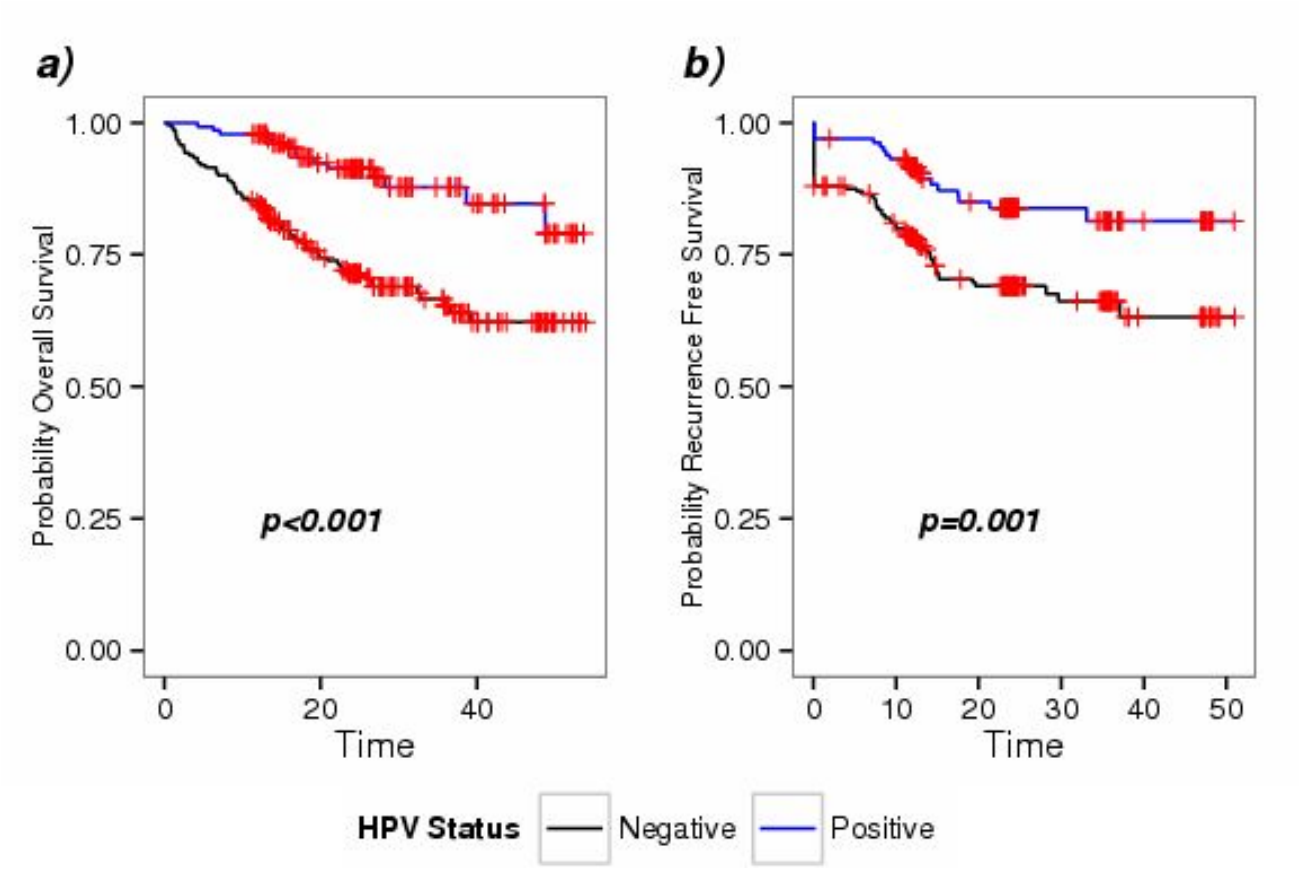


Figure 4.6. Smoking status is not associated with survival or recurrence. a) Never smokers have the best probability of OST, but smoking is not significantly associated with OST; b) Probability of RFT is similar across smoking statuses

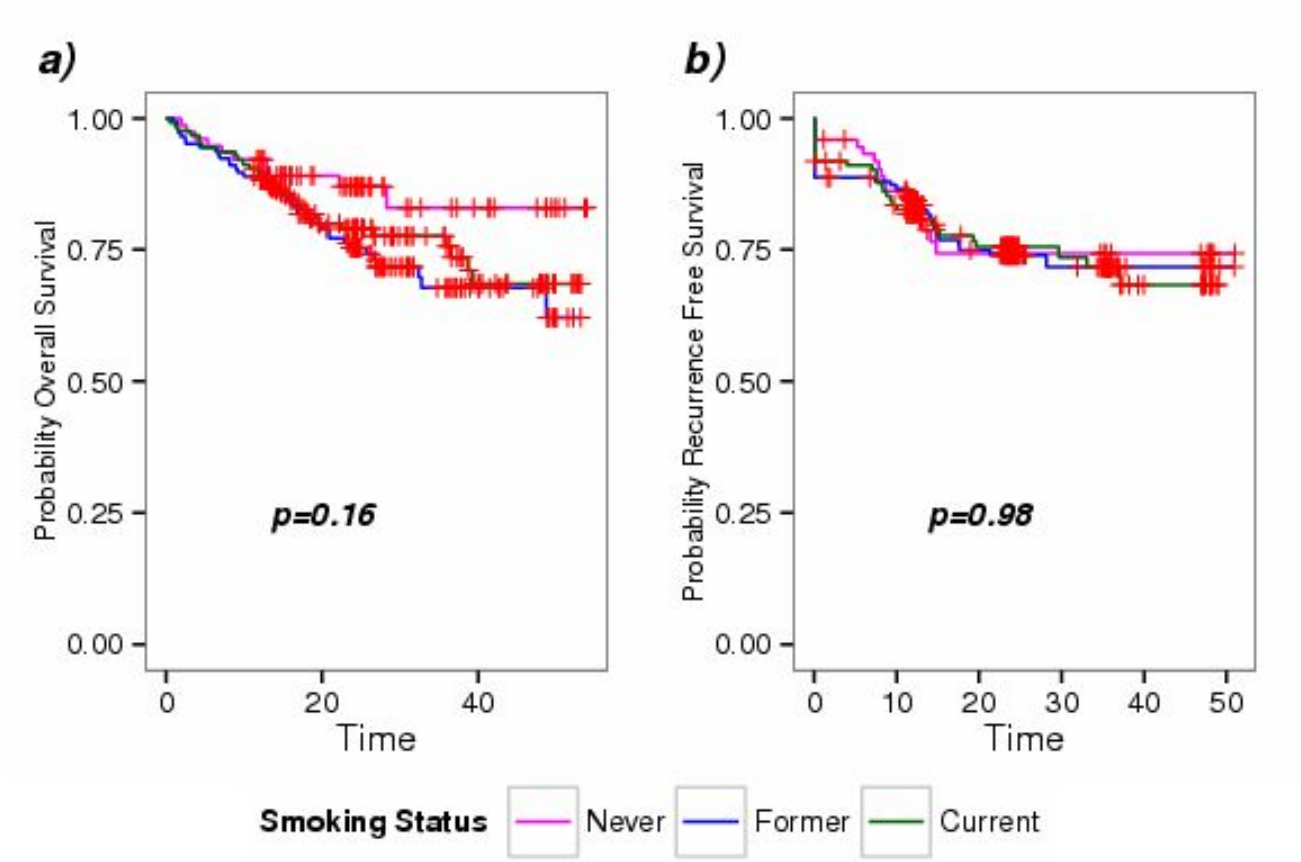


Figure 4.7. Disease site is significantly associated with prognosis. a) Patients with tumors in the hypopharynx have the worst probability of OST while patients with sites of larynx, oral cavity and oropharynx have increasing probabilities of OST; b) Probabilities of RFT follow the same trends as OST

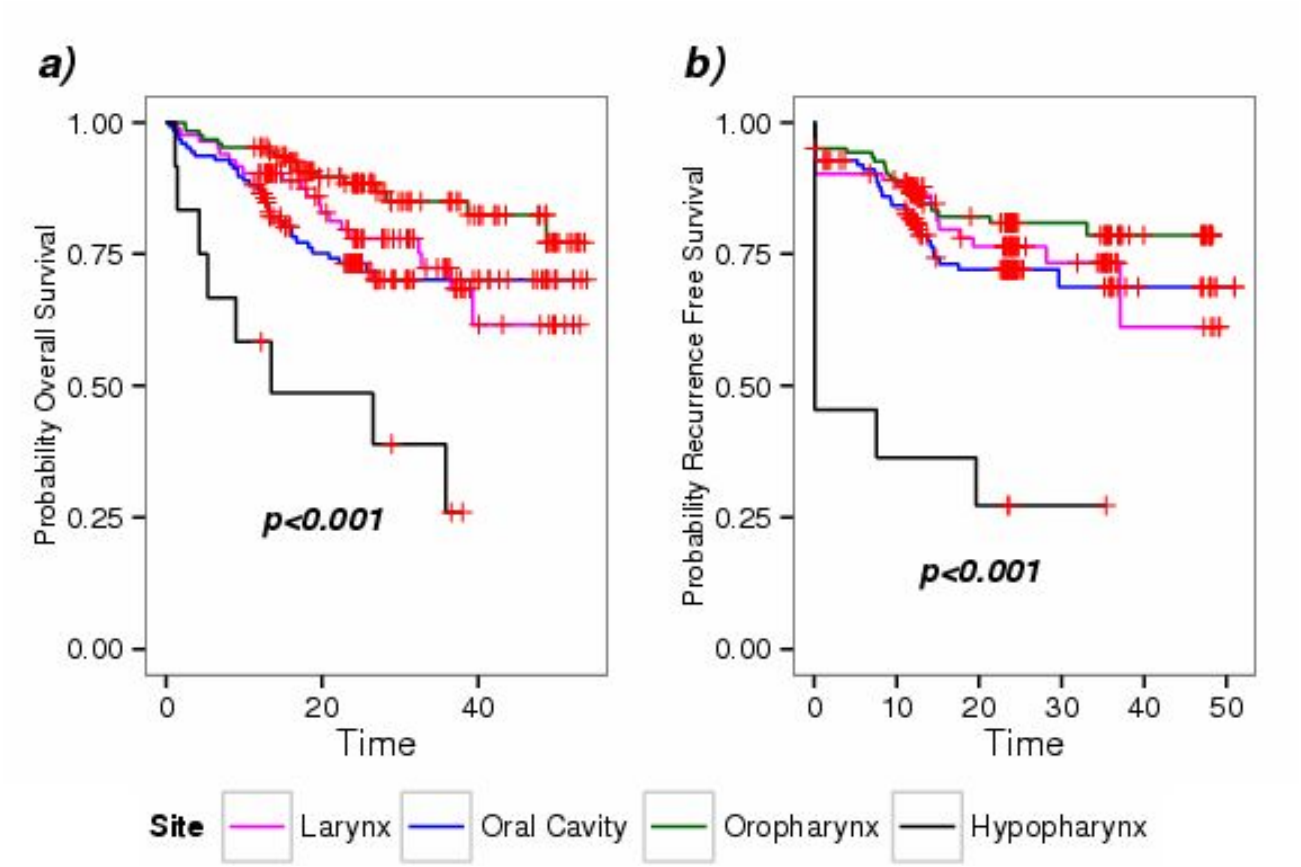


Figure 4.8. Disease stage is not associated with prognosis. a)There is no separation of curves across stage of probability of OST; b) There is a slight separation of curves for probability of RFT, with stages 1 and 2 having slightly better probabilites

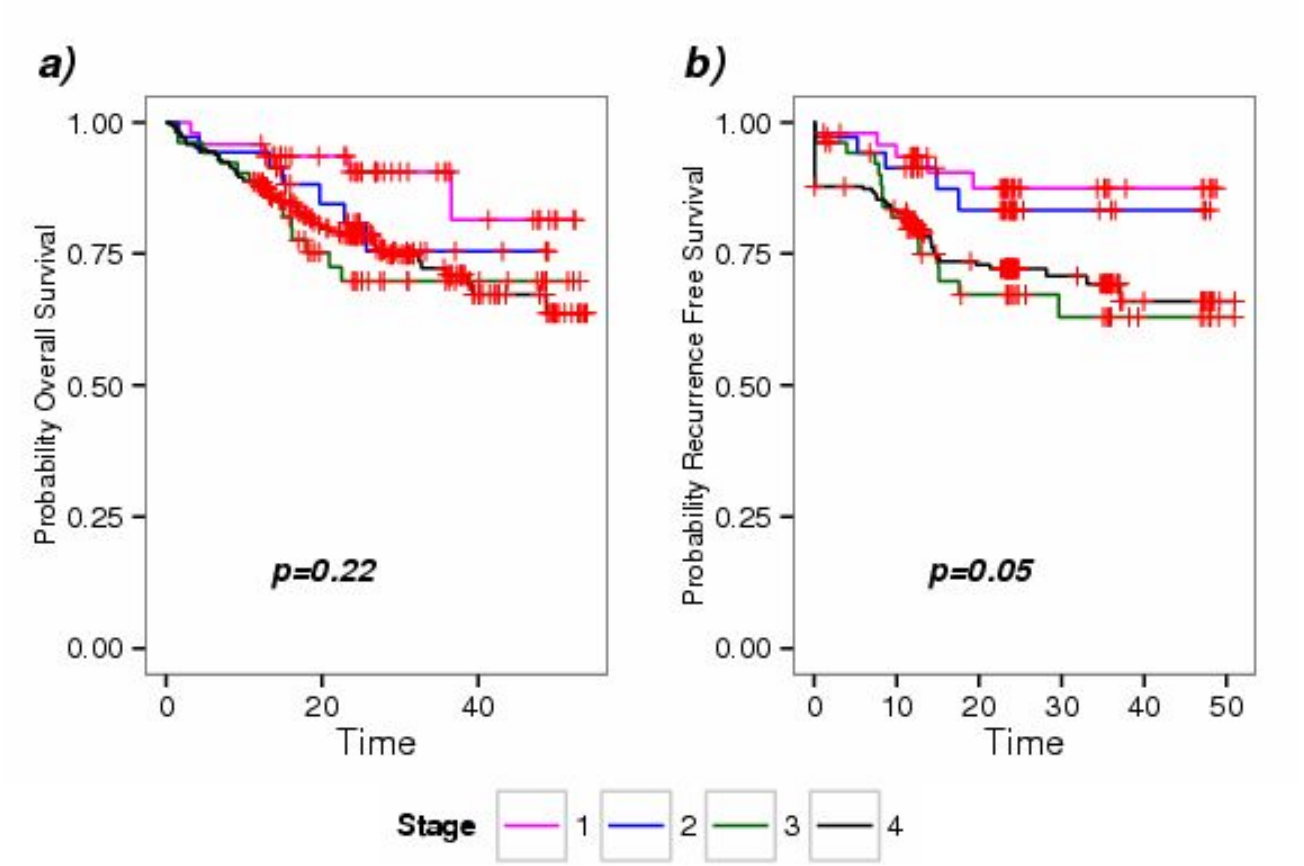


Figure 4.9. Comorbidity score is associated with survival only. a) Probability of OST significantly decreases with increasing score; b) Similar trends are seen for RFT as for OST, but these are not significant

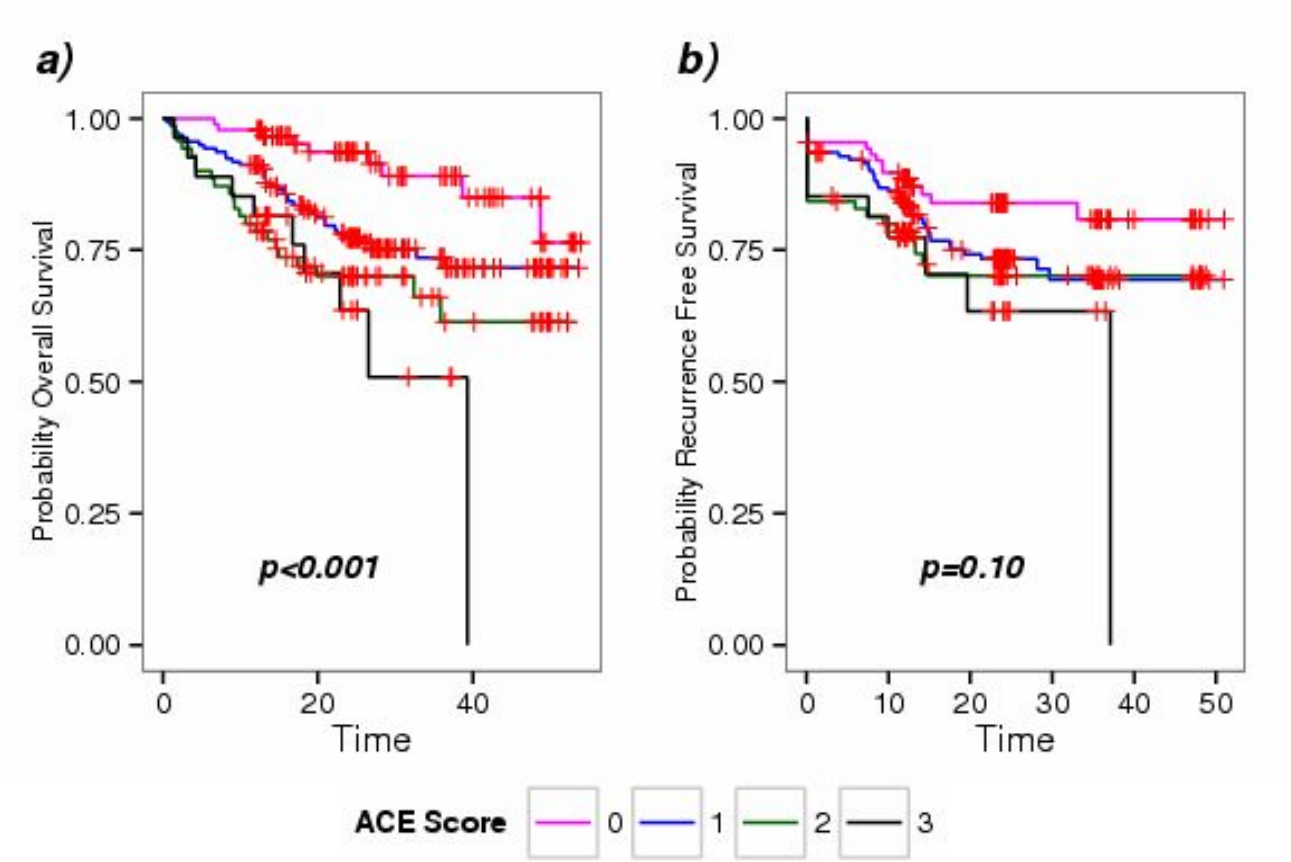


Figure 4.10. Probability of overall survival by quartiles of methylation. *CD1A* and *NDN* are significantly associated with probability of OST. The lowest quartile (Q1) of *CD1A* methylation has the worst probability while the rest of quartiles cluster around a probability of 0.75 (b); Q1 and Q3 of *NDN* have the lowest probabilities of OS, while Q2 and Q4 have the highest

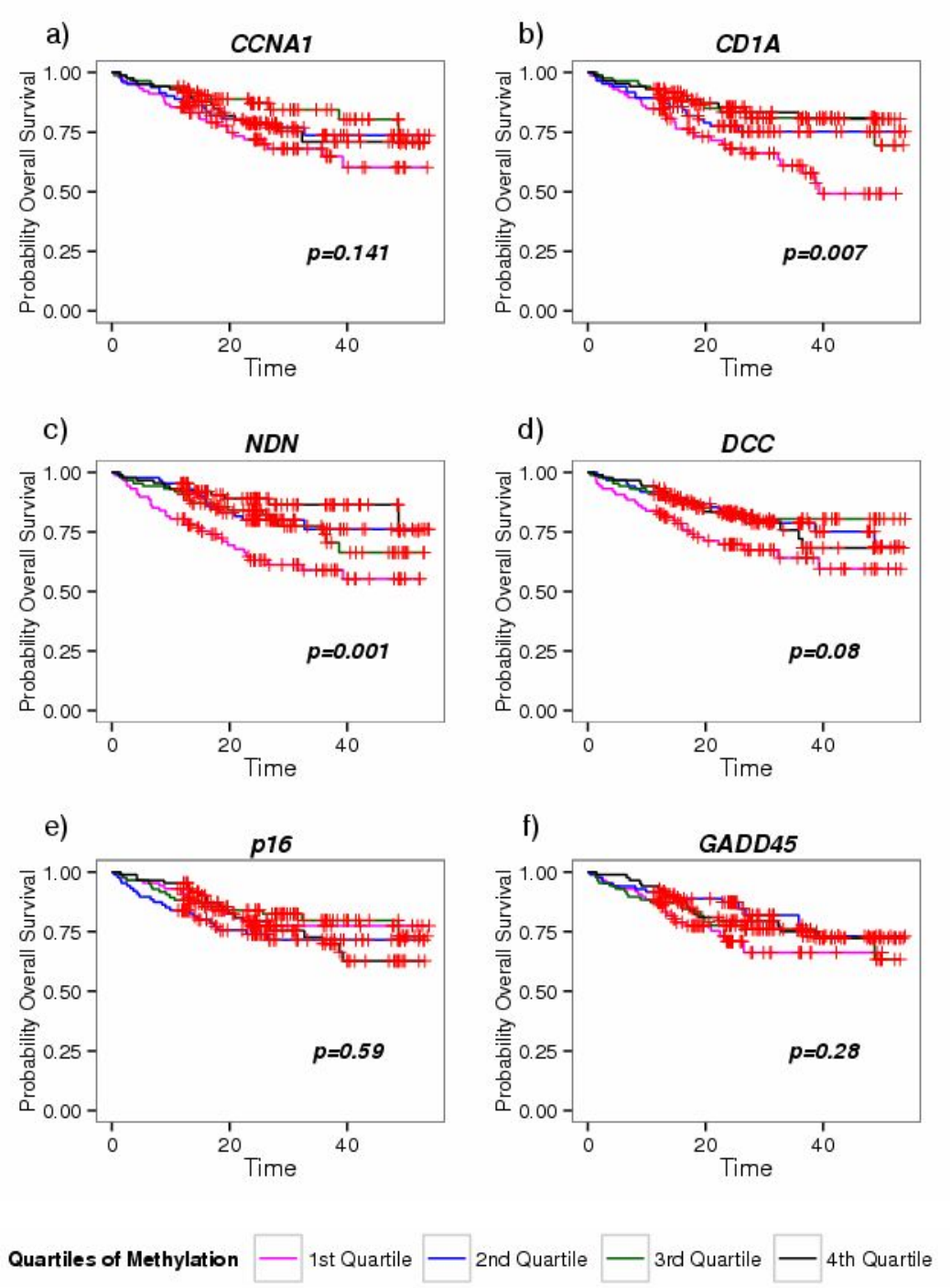


Figure 4.11. Probability of recurrence free survival by quartiles of methylation. *CCNA1*, *NDN* and *DCC* re all significantly associated with probability of RFT; Quartiles of *CCNA1* (a) and *DCC* (d) are associated with probability of RFT in a U-shaped manner where the lowest and highest quartiles have the lowest probabilities and the intermediate quartiles have high probabilities of RFT; Probability of RFT increases with increasing quartile of *NDN* methylation (c)

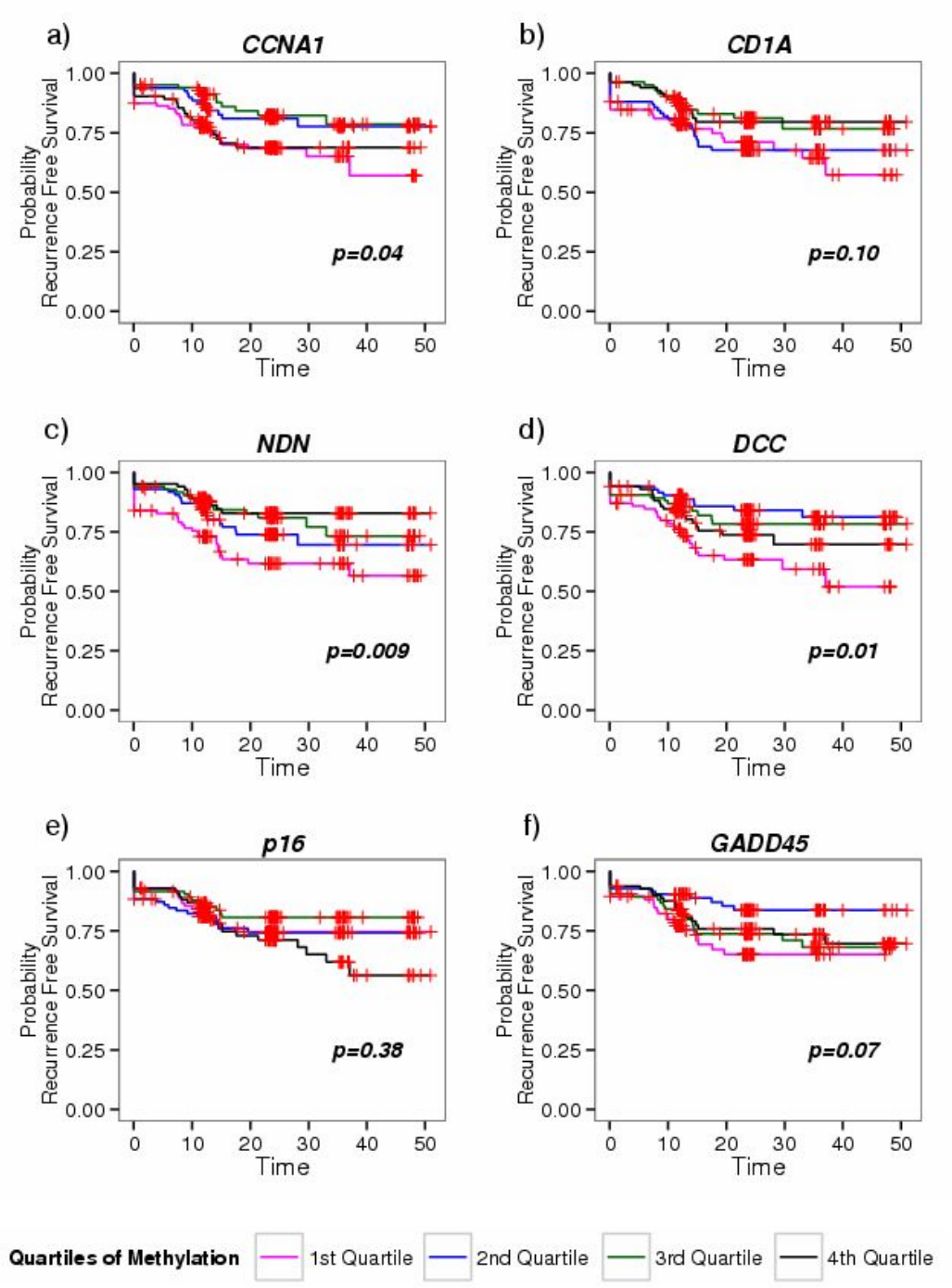


Figure 4.12. Association of *CCNA1*, *CD1A*, and *NDN* with survival, stratified by HPV status. *CD1A* is significantly associated with probability of OST in HPV(+) patients only. Low methylation of *CD1A* in Q1 has the lowest probability of OST compared to all other quartiles (d); No other significant associations are seen for these genes

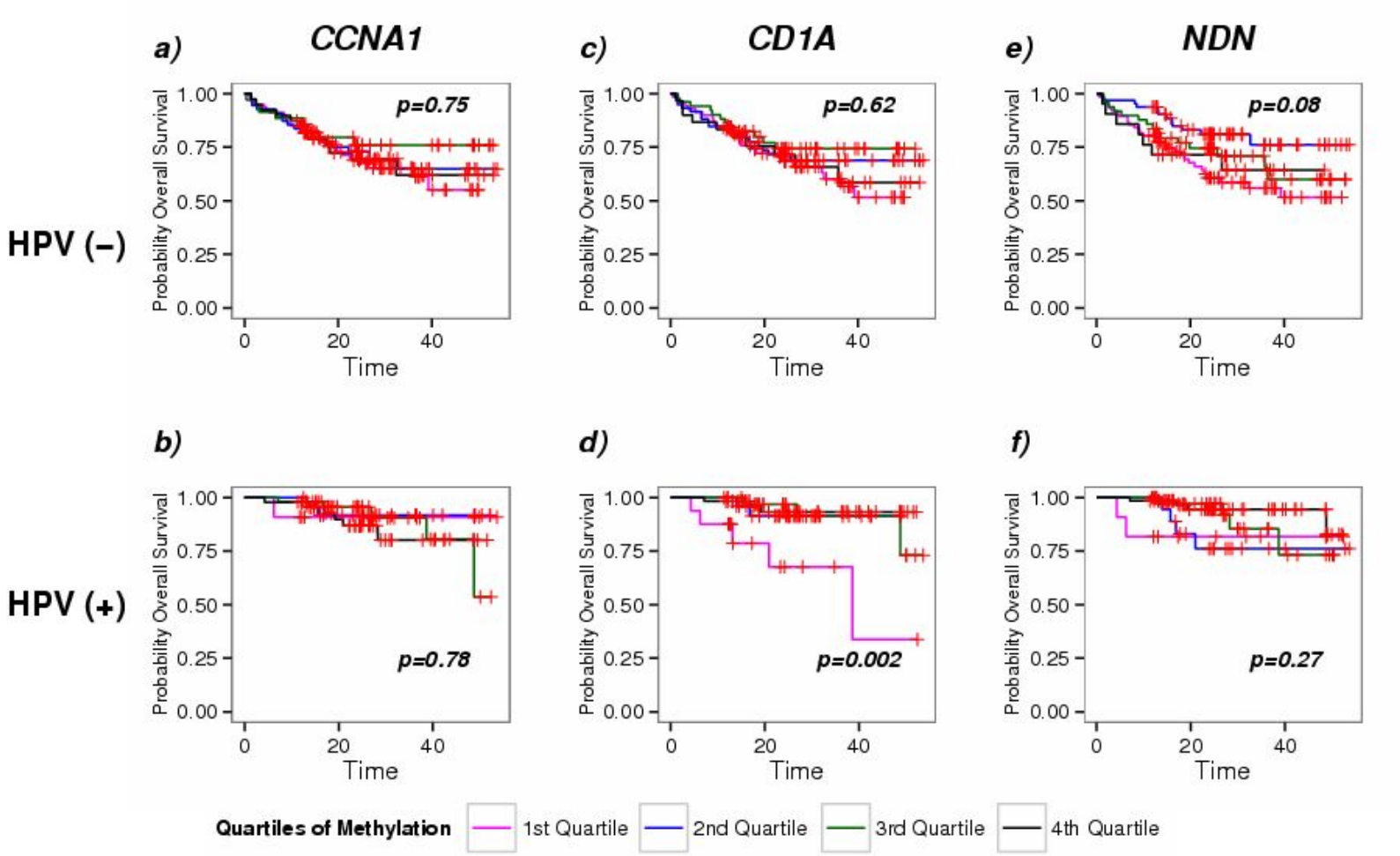


Figure 4.13. Association of *DCC*, *p16*, and *GADD45* with survival, stratified by HPV status. These markers re not significantly associated with probability of OST by HPV status

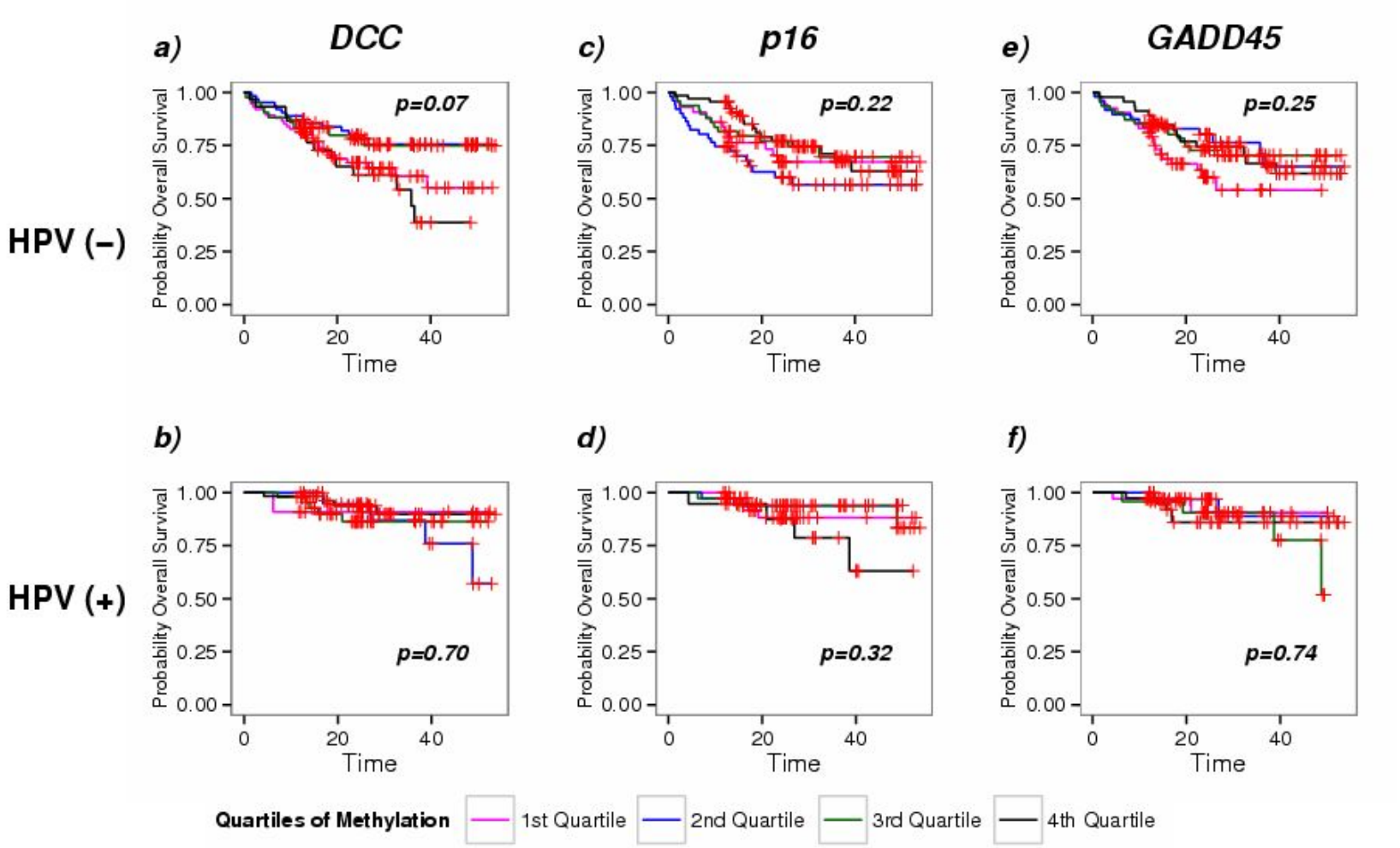


Figure 4.14. Association of *CCNA1*, *CD1A*, and *NDN* with recurrence, stratified by HPV status. *CCNA1* is associated with probability of RFT in HPV (+) patients. *CCNA1* methylation in Q4 has the worst probability, Q2 has the best probability and Q1 and Q3 had similar probabilities of RFT (b)

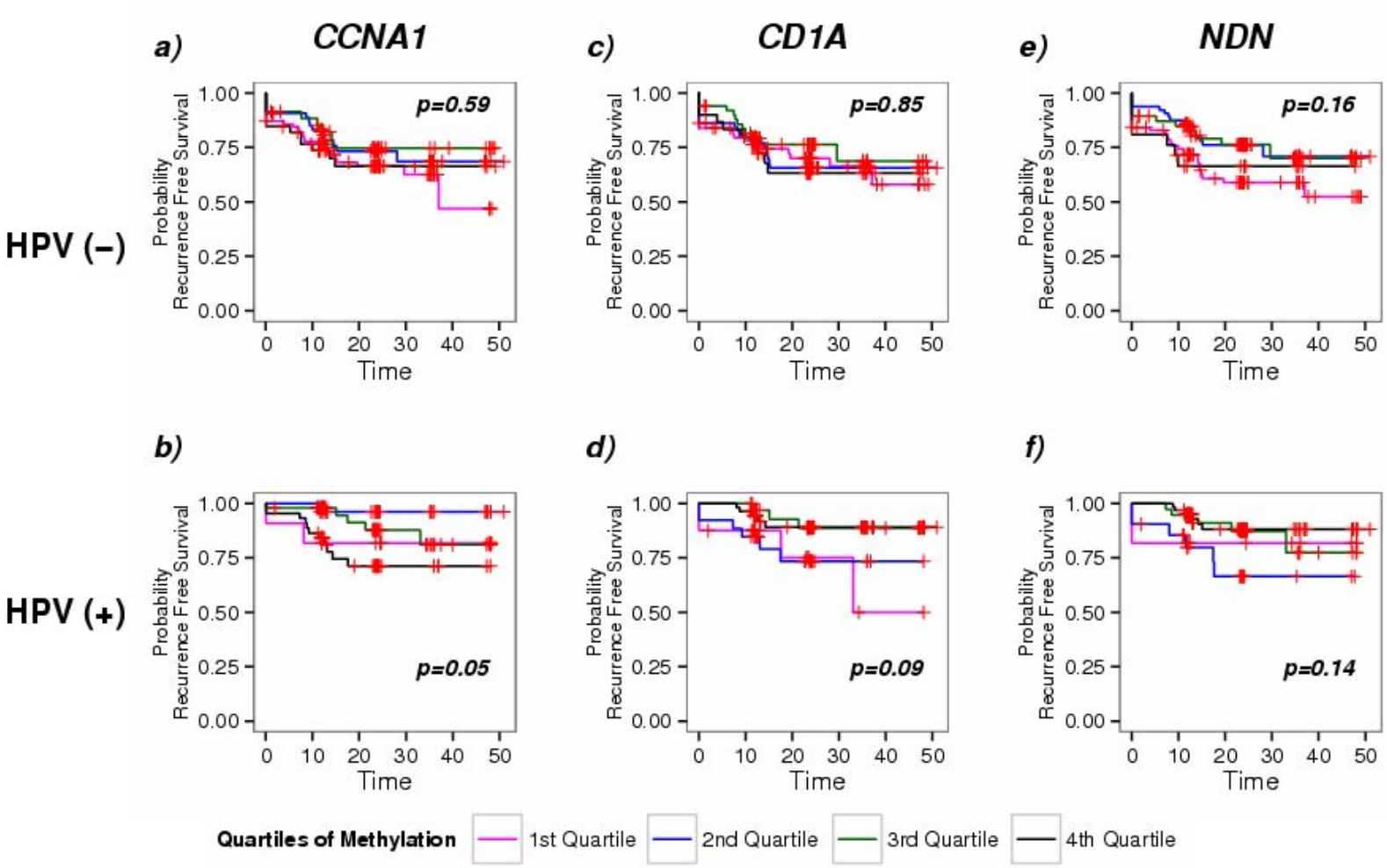


Figure 4.15. Association of *DCC*, *p16*, and *GADD45* with recurrence, stratified by HPV status. *DCC* methylation is significantly associated with probability of RFT; The trend of quartiles of *DCC* methylation are non-linear; intermediate quartiles, Q2 and Q3, have higher probabilities while Q1 and Q4 have the lowest probabilities of RFT

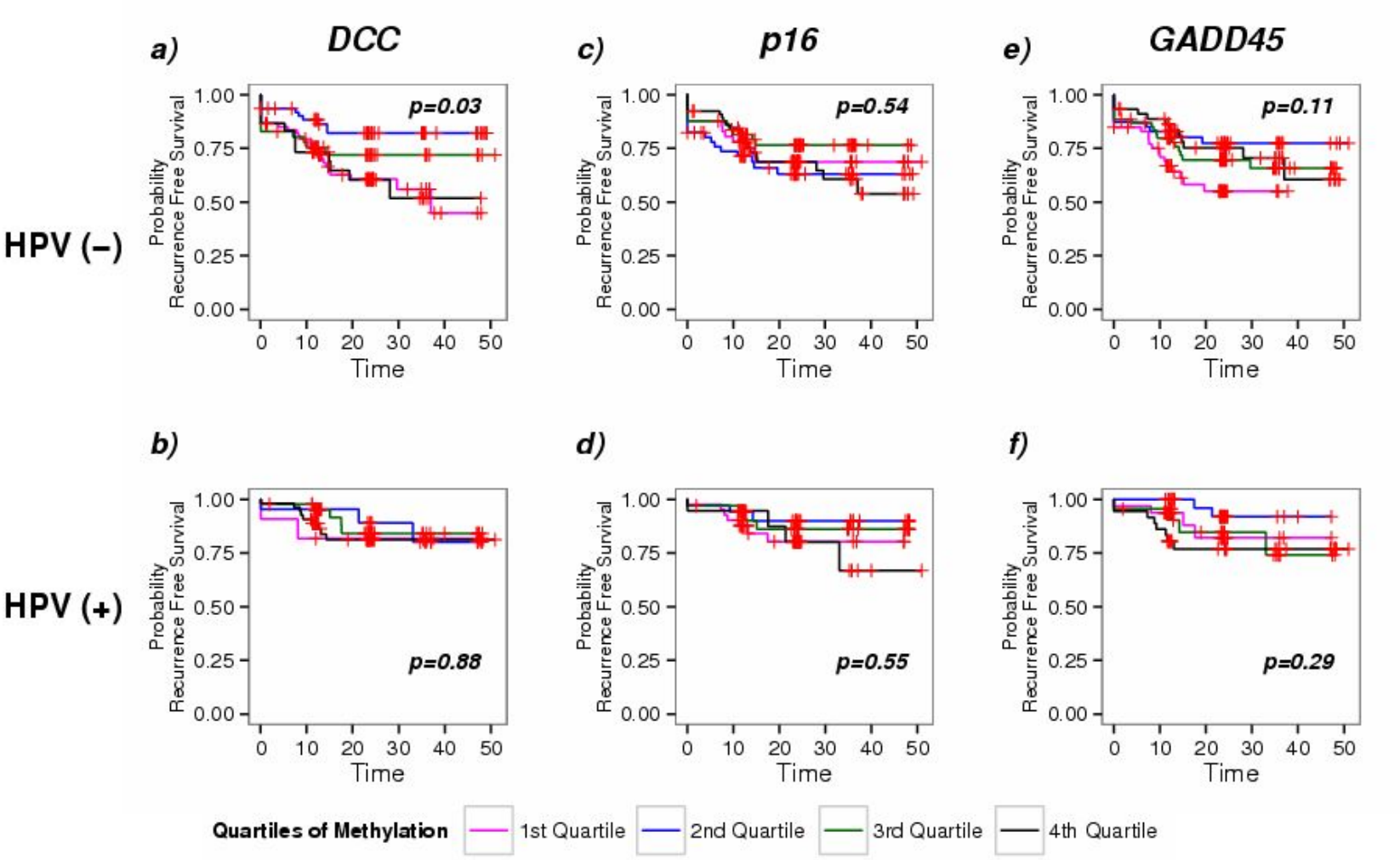
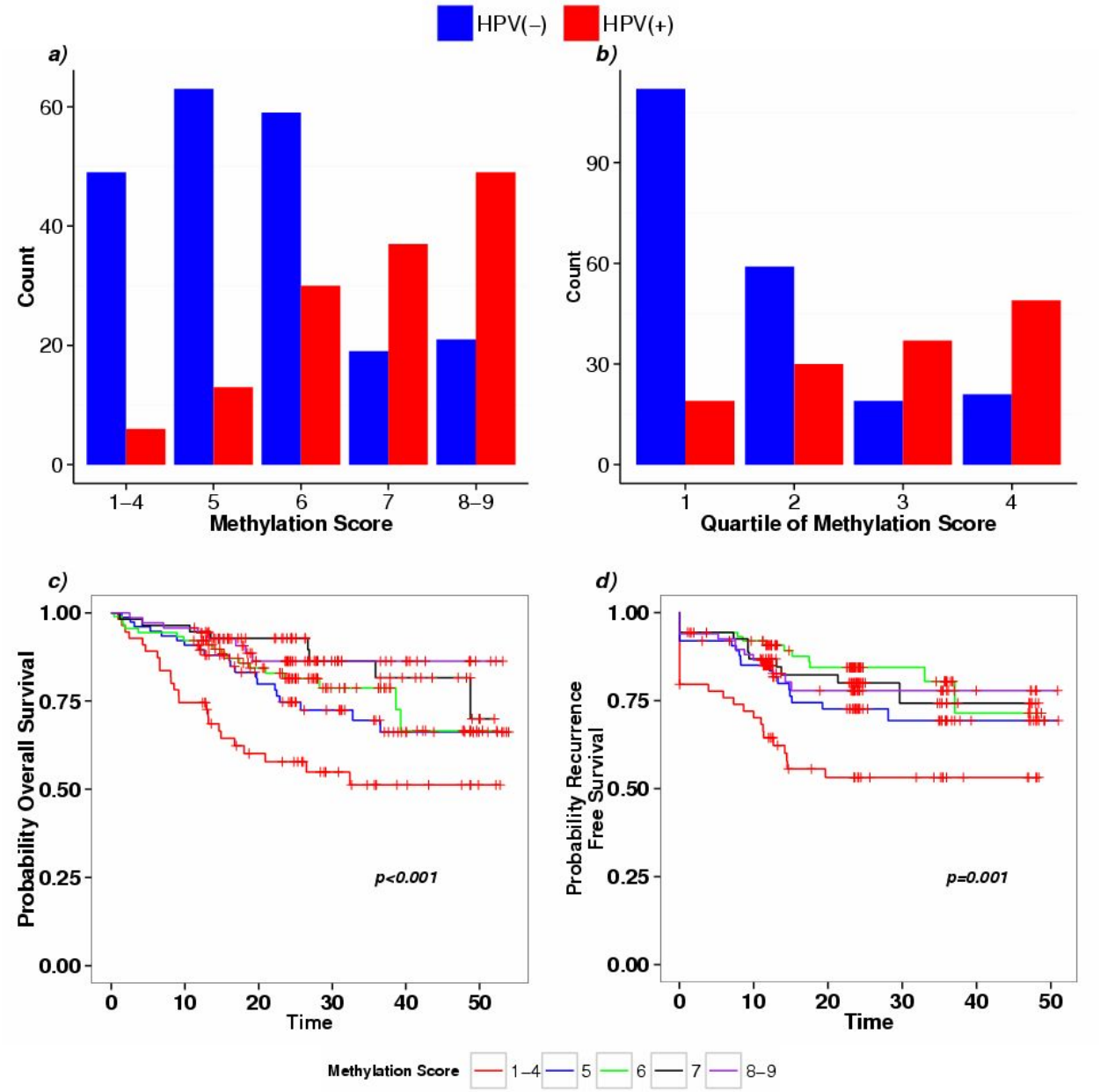


Figure 4.16. Distribution and association of methylation score with survival and recurrence. Number of HPV(-) patients decreases while number of HPV(+) patients increases as methylation score increases (a) and as quartile of methylation score increases (b); Probabilities of OST and RFT significantly decrease as methylation scores decrease (c, d)



CHAPTER 5
Public Health Implications

The work presented in this dissertation provides key findings that have immediate implications in the field of public health. Additionally, it provides a basis for future work for continuous research on environmental influences on cancer through its long-term implications.

BREAST CANCER IN THAILAND

Cancer is a leading cause of death worldwide. Sixty percent of cancer incidence occurs in low and middle income countries (LMICs)[1]. With the burden of cancer shifting to LMICs, it is necessary to provide data on incidence trends to support evidence-based implementation of strategies for cancer prevention, early detection of cancer and management.

This study used cancer registry data from Songkhla, Thailand to identify and forecast incidence trends for female breast cancer by age group.

Immediate Public Health Implications

This work analyzes female breast cancer in southern Thailand providing evidence of late staging of cancers and increased incidence in women at or above age 50. Characterization of incidence trends reveals that both period and cohort effects are significant contributors indicating that there may be separate types of risk factors that are contributing to incidence. Certain risk factors may affect younger generations more, whereas others affect older generations more. Some risk factors that are differential

between younger and older women include parity, adherence to self-breast examinations, changes in diet, age at first menarche, older ages at first birth, post-menopausal obesity, alcohol consumption, exogenous estrogen, and sexual behavior[2]. Some of these may be exacerbated during certain time periods contributing to period effects of the incident trend, while others represent generational lifestyle changes. The level of contribution of these risk factors are shifting as this country undergoes an epidemiological transition from a culturally traditional lifestyle to a more western lifestyle. This study provides characterization of female breast cancer rates in a region-specific manner, which is important as lifestyle factors vary dramatically by region. Additionally, it provides contextual evidence to examine the role of risk factors that change across generations and during specific time periods as lifestyle patterns change due to economic development in this region.

Long-Term Public Health Implications

This analysis consists of female breast cancer in Songkhla province only. However, the Thai NCI has established that breast cancer is the leading cancer in women nationally. Because there are several cancer registries across Thailand, results from this study can be combined with information from other registries to understand the incidence trends of female breast cancer across Thailand. Additionally, registry data from each region can be analyzed separately using this methodology to determine the types of effects that contribute to the region-specific incidence trends, providing a basis to determine differential risk factors by region. Careful characterization of cancer trends will provide the basis for future research to identify important risk factors, design interventions, and create prevention strategies.

CADMIUM IN THAILAND

Heavy metals have long been associated with carcinogenesis. Much of the evidence comes from industrial or occupational settings where exposures tend to be acute. However, environmental exposures to heavy metals also exist, although the exposure tends to be chronically low and may induce distinct pathways of carcinogenesis[3]. This study measured the association between environmental cadmium exposure and methylation markers in the highly exposed population of Mae Sot, Thailand.

Additionally, renal dysfunction was used as a proxy for disease since the process of carcinogenesis takes a long time to develop. Currently, the mechanism by which cadmium exerts its toxic effects is unknown. In vitro and animal studies have implicated the involvement of certain genes, but this has not been evaluated in human population in relevant doses.

Immediate Public Health Implications

This study takes advantage of a unique population with a wide range of environmental exposure to cadmium, to validate methylation markers that have been previously implicated in mechanisms of Cd toxicity. Much of what we know about molecular mechanisms of Cd toxicity comes from in vitro and animal studies. To date, few studies considering molecular events have been conducted in human populations. The results from this study validate specific methylation markers in human populations in relation to both cadmium exposure and renal dysfunction, a known health outcome of cadmium exposure. Importantly, these results confirm sex-specific differences on a molecular level. It has long been known that cadmium affects women differently from men. This is seen primarily from the sex-ratio of itai-itai disease patients and incidence rates of adverse health effects in women compared to men in response to cadmium exposure[4-7] .

The presence of molecular differences implicates distinct events that occur to generate sex-specific health outcomes.

Long-Term Public Health Implications

This work provides a basis for future studies to examine epigenetic changes in particular pathways as potential molecular mechanisms of cadmium toxicity leading to carcinogenesis. The toxicokinetics of cadmium indicate renal dysfunction is the direct health effect linked to cadmium exposure. Additionally, the levels of exposure known to induce renal dysfunction have also been associated with cancer incidence. Several methylation markers were associated only with cadmium exposure, offering potential mechanisms a variety of downstream health effects. Other methylation markers were associated with both cadmium exposure and renal dysfunction, implicating that methylation of these markers fall along the pathway of exposure and disease. It will be necessary to validate these novel biomarkers of exposure in cohorts with cancerous outcomes. The differential methylation markers by sex provide evidence that future interventions should be targeted at men and women separately, focusing on sex-specific behaviors and exposures.

HEAD AND NECK CANCER IN THE U.S.

Cancer is a heterogeneous disease and insufficient understanding has led to little improvement in survival for several types of cancer, including HNSCC. Biomarkers of diseases associated with survival play an important role in addressing high mortality rates of cancer. This study determined the association between methylation markers and patient survival by HPV status in HNSCC.

Immediate Public Health Implications

Currently, subsets of HNSCC are based on HPV status, a known biological factor that contributes to etiology of this cancer. HPV (+) patients have a prognostic advantage in terms of survival, although it is unclear whether this is due to tumor biology, epidemiologic characteristics, or a combination of the two. However, even within the HPV+ group, there is a good deal of heterogeneity in survival time, with up to 20% progressing with distant metastases. This study identifies novel biological factors that are associated with survival and recurrence and exhibit differential methylation by HPV status and epidemiological characteristics. These biomarkers highlight differences in tumor biology that contribute to differential prognosis supporting biological implications of epigenetic markers on patient survival.

Long-Term Public Health Implications

Differences in patient outcomes according to HPV status are so dramatic that many investigators believe they reflect a new and unique phenotype that could justify significant de-intensification of therapy. Methylation markers may incorporate epidemiologic characteristics with tumor biology and therefore are useful in understanding survival differences by HN subtypes. The novel methylation markers identified in this study offer new, specific, epigenetic molecular differences within the setting of the generalized hypermethylation phenotype associated with HPV status and warrant further investigation. These markers offer potential usefulness in identifying unique subsets of patients with varied outcomes, providing an opportunity for targeted, personalized care.

References

[1]Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser SM, C , Rebelo M, Parkin D, Forman D, Bray F. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide*. IARC CancerBase No 11 [Internet].International Agency for Research on Cancer. Accessed on March 10. 2013. Available From: http://globocan.iarc.fr

[2]Hsieh CC, Trichopoulos D, Katsouyanni K, Yuasa S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. *International journal of cancer Journal international du cancer* 1990;**46**: 796-800.

[3]Hayes RB. The carcinogenicity of metals in humans. *Cancer causes & control : CCC* 1997;**8**: 371-85.

[4]Nishijo M, Satarug S, Honda R, Tsuritani I, Aoshima K. The gender differences in health effects of environmental cadmium exposure and potential mechanisms. *Molecular and cellular biochemistry* 2004;**255**: 87-92.

[5]Vahter M, Akesson A, Liden C, Ceccatelli S, Berglund M. Gender differences in the disposition and toxicity of metals. *Environmental research* 2007;**104**: 85-95.

[6]Tsuritani I, Honda R, Ishizaki M, Yamada Y, Kido T, Nogawa K. Impairment of vitamin D metabolism due to environmental cadmium exposure, and possible relevance to sex-related differences in vulnerability to the bone damage. *Journal of toxicology and environmental health* 1992;**37**: 519-33.

[7]Olsson IM, Bensryd I, Lundh T, Ottosson H, Skerfving S, Oskarsson A. Cadmium in blood and urine--impact of sex, age, dietary intake, iron status, and former smoking--association of renal effects. *Environmental health perspectives* 2002;**110**: 1185-90.

CHAPTER 6
Conclusions

This thesis highlights studies that are essential in understanding the trajectory of human environmental exposures inducing and progressing carcinogenic events, both on an epidemiological and a molecular basis.

The study in chapter two focused on the changing environment and induction of cancer, measured epidemiologically through the use of incidence rates from population-based cancer registries. There are many known risk factors for breast cancer, but this health burden is rising the fastest in LMICs. Thailand is one country undergoing an epidemiologic transition in lifespan and lifestyle characteristics combined with a shift from infectious to chronic diseases. This shift in risk profiles of southern Thai women deems it necessary to focus on early detection and awareness of breast cancer. In this study, female breast cancer incidence trends were characterized to be influenced by both period and cohort effects, implicating risk factors associated with birth year and lifestyle/behavioral changes in more recent years. Additionally, keeping the current healthcare infrastructure constant, comparative modeling illustrated that these trends would continue to increase for women of all ages, although trends for women ages 50 and up would accelerate at a much faster rate by 2029. These results provide a basis for future health care planning, focusing on potential integration of mammography or other preventive screening techniques into the universal health care system. Although breast self-examination is practiced, compliance is highest only in younger females, leaving the older female population at risk. Inclusion of mammography into universal healthcare would promote preventative care and

although incidence rates would increase due to capture of more cases, cancers will be staged earlier allowing for more effective treatment and management.

A strength of this study is that it is the first one to characterize breast cancer trends in a specific region of Thailand. Lifestyle factors vary dramatically by region and likely have differential effects on incidence rates, making it necessary to study trends in a region-specific manner. Additionally, cancer trends are characterized separately for low and high risk groups in terms of menopausal status, allowing for evidence to be used in interventions targeted at specific age groups. Although the registry data are limited due to the lack of information on biomarkers, religion and other lifestyle characteristics, this is an inherent limitation as, to date, there has not been comprehensive data collection at the population-level. Regardless, this study provides the first in-depth look at the epidemiology of breast cancer in southern Thailand. Because of the changing risk profile of the women in this region, these findings need to be extended to characterize the population in terms of diet, lifestyle and genetic factors to direct strategies aimed at controlling the burden of breast cancer in Thailand.

Chapter three was centered on environmental exposure to a known carcinogen and the induction of disease, measured molecularly through the use of epigenetic markers as biomarkers of effect. Cadmium was found to be associated with DNA methylation levels of important candidate genes that have been previously shown from in vitro and animal studies to be involved in mechanisms of toxicity. Specifically, cadmium was associated with DNA methylation differentially by sex. These biomarkers of effect were also associated with renal biomarkers indicative of renal dysfunction. The process of carcinogenesis takes years to develop, so renal dysfunction was used as a proxy as this is a known health effect of cadmium and the same levels of exposure that have been shown to induce renal dysfunction have also been associated with

cancer incidence. Urinary cadmium levels, indicative of body burden of this exposure, were associated with biomarkers of effect that fall along the pathways of cadmium toxicity, although these marks differed by sex, indicating that cadmium toxicity may occur through differential mechanisms in males compared to females.

Strengths of this study include the use of a population with a wide range of relevant cadmium levels and extensively characterized methylation biomarkers. This is the first study able to look at such ranges within the context of DNA methylation in a population-based study. However, this cross-sectional study was limited in the amount of information provided as we did not have a part in the creation of the questionnaire and the survey had already been completed by the time these analyses took place. It would have been useful to have measurements on diet, body burden of other heavy metals that coexist with environmental cadmium or longitudinal data that would have allowed for measurement of cancerous outcomes. Dietary information would have been useful to understand the amount of cadmium ingested along with other food items that may affect metabolism or biotransformation of cadmium. Other heavy metals such as lead, nickel, lead and copper, are commonly found with cadmium in the environment and may work synergistically to induce adverse health outcomes. Longitudinal data would have allowed for measurement of cancer incidence, although this limitation was addressed by using renal dysfunction as a proxy since Cd exposure levels associated with kidney effects and various cancers are similar. Finally, methylation was measured in circulating cells of the blood and therefore, the direct effects of Cd on DNA methylation may not be extrapolated to relevant tissues, such as the kidneys and liver. Nevertheless, the results from this study are important in recognizing that differential mechanisms of cadmium toxicity may come into play depending on sex. This has been previously alluded to with the recognition that women tend to sequester larger

amounts of cadmium and suffer more from osteoporosis than men. Additionally, the past outbreak of itai-itai disease from environmental cadmium exposure in Japan had only 3 out of 195 patients that were male, suggesting that cadmium toxicity affects females differently than males. In spite of this, few studies have looked at sex-specific incidence rates of cancer with regards to environmental cadmium exposure, highlighting a need to examine cancer outcomes in this way.

Chapter four discussed prognostic markers, found in diagnosed cancer patients, which indicate overall patient survival and recurrence free survival from head and neck squamous cell carcinoma (HNSCC). The 5-year survival rate has remained at about 50-60% for the past decade. Although HPV-associated cancers are indicative of improved prognosis, there is still heterogeneity in survival time, highlighting a need for prognostic markers of survival patients related to tumor biology and epidemiologic characteristics. Several epigenetic marks were found to be associated with survival and recurrence, both overall and within HPV subsets of HNSCC. These discoveries are novel findings as they have not previously been associated with prognosis and offer new, specific, epigenetic molecular differences within the setting of the generalized hypermethylation phenotype associated with HPV status.

A limitation of this study is that there is no measurement on the persistence of these epigenetic marks. As this is a cross-sectional study using tumor tissue from previously untreated patients, specimens are difficult to come by and radiation/chemotherapy treatments begin after surgery. Additionally, diet, which is known to influence DNA methylation patterns and have been associated with survival in HNSCC patients, is not included in this study. Combining methylation data with diet information would be useful in determining if these DNA methylation marks are truly indicative of prognosis. A strength of this study is the use of a well-characterized,

unselected cohort of HNSCC patients at the University of Michigan, with extensive epidemiologic, clinical and survival information, treated by a single group of clinicians with a homogenous treatment approach. This cohort shows the expected associations established in previous literature, such as the relationships between stage, site and HPV status with overall survival time, and the expected population characteristics of a HNSCC cohort established by previous studies, providing assurance that the new associations discovered with this cohort are meaningful and can be extrapolated to the patient population. The findings from this study support biological implications of epigenetic markers on patient survival and their potential usefulness in identifying unique subsets of patients with varied outcomes.

The changing environment has a profound effect on human health worldwide. Both epidemiological and molecular studies are needed to address the rising global cancer burden that will particularly affect low- and middle-income countries (LMICs). The work presented here addresses this issue using three studies that cover the spectrum of environmentally influenced cancer, from induction to progression to final prognosis. Findings from this work are crucial in understanding the evolution of disease and in providing evidence that can be used in prevention, healthcare planning and clinical treatment.