Application of Genomic and Epigenomic Methods to Understand Environmental and Dietary Factors in Carcinogenesis

by

Justin A. Colacino

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Environmental Health Sciences)
in the University of Michigan
2014

Doctoral Committee:

Associate Professor Laura S. Rozek, Chair
Associate Professor Dana C. Dolinoy
Associate Professor Jun Li
Assistant Professor Maureen Sartor

DEDICATION

I dedicate this work to my parents, Tom and Mari, for all of their love, support, and understanding.

ACKNOWLEDGEMENTS

I would first like to acknowledge my thesis advisor, Laura Rozek. As Laura's first graduate student, I have seen the evolution of her laboratory over the course of her being new faculty to tenured professor. Laura was always a supportive mentor, willing to let me pitch my own ideas but always there to let me know if what I was proposing was too crazy. I learned not only how to be a good scientist from her, but also what's involved in navigating early faculty life. I would also like to acknowledge the other members of my dissertation committee Dana Dolinoy, Jun Li, and Maureen Sartor for their excellent input and support throughout this process. I've learned so much from each of you and I feel lucky to have been mentored by such a bright group of scientists.

I would also like to acknowledge Madhuri Kakarala's mentorship. Madhuri originally inspired me to work with stem cells, and without her support, my research would not have taken the (interesting!) turn that it did. Her mentorship and input was extremely influential in the development of my research program, and for that I am very grateful. Similarly, I would like to acknowledge Max Wicha, who took me under his wing when it looked like my work was going to hit a dead end. Max has been a tremendous mentor as well as a scientific inspiration over the course of the past two and a half years. In particular, Max's lab meetings have been a fantastic source of information and input. I will strive to emulate that setting in my own lab meetings as I transition to my role as an independent researcher. I look forward to continuing to collaborate and work with Max and the rest of his group.

I would also like to thank Dean Brenner, Karen Peterson, Jeff Moyer, Jeremy Taylor, Tom Carey, Martin Philbert, and Greg Wolf. I've received

tremendous input from all of you, either in lab meetings, face-to-face meetings, or in round after round of publication edits. I am a better scientist today because of my interactions with all of you.

Within the Rozek lab, I'd like to particularly acknowledge Anna Arthur and Shama Virani. We went through a lot together and came out (mostly) better on the other side. I'm excited about continuing to work with both of you as we all move forward in our careers (as doctors!) I'd also like to thank our lab managers over the years, Adrienne Van Zomeren-Dohm, Tami Jones, and Katie Rentschler. Obviously nothing would ever get done without you guys, and all of you have been so unbelievably helpful over the years. I'd also like to thank my former and current trainees, Erika Koeppe and Danielle Lee for being so bright, useful, and generally awesome. In the Wicha lab, I'd like to specifically acknowledge Sean McDermott, who mentored me throughout my work presented here in Chapter 4. Sean, I'm pretty sure, is some sort of science wizard, and learning flow cytometry from him made me feel like I developed superpowers. Shawn Clouthier and Tahra Luther in the Wicha lab have also been fantastic, thank you both so much for all of the support over the years.

I would not have been able to accomplish any of this work without the Department of Environmental Health's tremendous staff, especially Patrice Sommerville, Kelsey Hargesheimer, Katherine Wood, Sue Crawford, and Cecilia Young. You all keep the department running, and I'm glad that we get to keep working together in the future.

On a more personal note, I'd like to thank my parents, Tom and Mari Colacino, for the love and support that they have given me over the years. They always encouraged me to follow my dreams, even if those dreams led to me to being a student until I was 32. I know that their input early in my life developed my scientific curiosity and my love of learning, a classic example of the developmental origins hypothesis. I'd also like to thank my wonderful girlfriend Kelly Bakulski, who is an amazing scientist in her own right and one of the best overall humans. I've also been blessed by having a tremendous group of friends

that I made at Michigan. I’m looking forward to years of science, kickball, juice parties, and/or commune planning with Ashley Carpenter, Sean Ferris, Katie Conlon, Lauren Tetz, Jon Bryant-Genevier, Ashley Dehudy, Owen Darr, Muna Nahar, Lisa Marchlewicz, Melissa Smarr, Liv Anderson, Laura Kubik, Zishaan Farooqui, Shayna Liberman, Kevin Boehnke, Liz Pierce, Mike Steinbaugh, Matt Smith, Matt Molusky, Dave Thal, Paul Marinec, Levi Blazer, Paul Moore, Brian Davis, Jen Massie, Bev Piggott, Vani Murthy, Kathryn Demanelis, Luis Rivera, Siying Huang, and Juan Quirindongo. I’d also be remiss to not acknowledge the furry roommates who mostly relieved, but occasionally caused, stress during my time at 806 Pauline: Henrietta, Carl, Hamburger, Fisher, Bugaloo, Zoe, and Malcolm.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Sporadic cancers comprise the vast majority of diagnosed cancer cases, many with a largely environmental etiology. The mechanisms by which specific environmental factors influence cancer risk, however, remain widely uncharacterized. Because sporadic cancers are diagnosed later in life, many incident cancer studies poorly quantify previous exposures or utilize methodologies that may not be appropriate for the study of cancer initiation or prevention. Developing novel methods of studying the role of nutrition and the environment in carcinogenesis will provide essential insight towards the prevention, early identification, and treatment of cancer. Incorporating novel culture methods, including primary tissue culture, will allow for the study of specific and relevant normal cell populations, including stem cells, that may be particularly sensitive to environmental and nutritional factors.

The overarching goal of this dissertation was to develop and apply novel statistical and experimental methods to characterize the roles of nutrition and the environment in carcinogenesis and cancer prevention, with a focus on epigenetic change. In Chapter 2, comprehensive epidemiological and clinical information were paired with DNA methylation profiling of head and neck tumors to identify significant differences in tumor DNA methylation in chemically induced or human papillomavirus induced tumors. In Chapter 3, data on average dietary intake was paired with tumor epigenetic measurements to identify that a head and neck cancer patient’s diet in the year before diagnosis can significantly affect tumor epigenetic profiles, providing a potential mechanism by which diet affects disease prognosis. In Chapter 4, normal human breast stem cells from reduction mammoplasty tissue were treated with the putatively cancer prevention

compounds curcumin and piperine, and a genome-wide screen was conducted to identify the stem cell specific changes induced by these compounds. The results and conclusions presented here reflect the utility of the application of these methods, from cancer molecular epidemiology to normal human in vitro stem cell culture, to understand the role of the environment in cancer.

CHAPTER 1. INTRODUCTION

1.1 Background and Motivation

Sporadic cancers comprise the vast majority of diagnosed breast and head and neck cancer cases, many with a largely environmental etiology (Lichtenstein et al. 2000). Smoking, alcohol consumption, and poor diet are estimated to be the leading risk factors for cancer worldwide (Danaei et al. 2005). The shifting epidemiological profile of these diseases provides additional evidence that the environment plays a significant role in cancer development. In the United States, the epidemiology of head and neck cancer has undergone a radical shift in the past decade. The traditional risk factors for the development of head and neck cancer are tobacco smoking, alcohol consumption, and poor diet (Lewin et al. 1998). Since 1999, however, there has been a significant increase in the incidence of oropharangeal cancer in non-smoking males under the age of 50 (Marur et al. 2010). In developing countries, including those of southeast and central Asia, the probability of developing breast cancer more than doubled between 1980 and 2010 (Forouzanfar et al. 2011). Some of the increased incidence can undoubtedly be attributed to lengthening lifespan and improvements in detection over this time period. However, these dramatic shifts in cancer rates, paired with the stable nature of the human genome over this time period, suggest a strong role of the environment in cancer.

Despite the changing epidemiology of cancer worldwide, the mechanisms by which specific environmental factors influence cancer risk remain widely uncharacterized. Since cancer is diagnosed later in life and it takes decades for most cancers to progress from initiation to neoplasia, many cancer studies poorly quantify previous exposures or utilize methodologies that may not be appropriate for the study of cancer initiation or prevention. For example, cancer epigenetic epidemiology studies often either poorly characterize environmental exposures or

focus on a limited number of epigenetic target sites. *In vitro* studies of cancer have traditionally relied on the use of immortalized cells or cells derived from tumor tissue. These studies often do not account for interindividual variability in response and may be limited in their ability to produce physiologically relevant data about effects of environmental exposures. Developing novel methods for studying the role of nutrition and the environment in carcinogenesis will provide essential insight towards the prevention, early identification, and treatment of cancer. Incorporating novel culture methods, including primary tissue culture, will allow for the study of environmental and nutritional factors in specific normal cell populations that may be at increased risk for cancer development, including stem cells.

1.2 Genetic and Epigenetic Causes of Tumor Heterogeneity

The tumor is a complex microenvironment that forms from the interaction of cancer cells, tumor associated stroma, blood vasculature, and immune cells (Hanahan and Weinberg 2000, 2011). Cellular heterogeneity within a tumor, reflected as phenotypically distinct cellular subpopulations, has long been described in a range of tumor types. In colorectal cancer, for example, tumor heterogeneity has been well-characterized as the gradual sequential accumulation of genetic and epigenetic alterations throughout tumorigenesis. As normal colonic epithelium progresses to a carcinoma, mutations in *APC*, *KRAS, DCC*, and *TP53* are accumulated, leading to a dramatic phenotypic change in cells that accumulate the range of mutations (Fearon and Vogelstein 1990). This serial acquisition of genetic mutations that confers increasing fitness can lead to expansion of clonal populations with phenotypes increasingly divergent from normal cells. This hypothesis has been elegantly illustrated in recent experiments sequencing the genomes of tumors at ultrahigh depths to understand their clonal evolution. These studies have identified that even within a tumor, there can be substantial genetic variation. For example, sequencing analyses of 21 individual breast cancers identified that tumors are comprised of a diverse number of genetic subclones, with individual mutations only found in subsets of cells (Nik-Zainal et al. 2012). Tumors evolve to thrive from the selective pressure from their

surroundings, and the genomic instability in a tumor allows for hyper-accelerated evolution. Thus, within a tumor, individual cells can be more or less likely to survive therapy based on both their mutational profile and their interactions with the microenvironment.

While the genetic contribution to cancer is undoubtedly a necessary part of tumorigenesis, the contribution of epigenetic factors to cancer is also becoming clear (Feinberg and Tycko 2004). Epigenetic modifications cause heritable changes in gene expression without changing the underlying genetic code. Since the hundreds of different cell types in the human body, from a liver cell to a skin cell, all share the same basic genetic code, epigenetic modifications control the gene expression required to establish and maintain each of these cell types (Reik 2007). Thus, epigenetic modifications are essential for establishing cellular identity during and throughout development, as cells progress from pluripotent, to multipotent, to fully differentiated. Epigenetic modifications are environmentally labile, varying over the lifetime and by tissue. They are also potentially reversible, representing a distinct mechanism for organisms to adapt to their environment (Feinberg and Irizarry 2010). The effect of environmental exposures on epigenetic regulation in human populations, however, remains poorly understood.

There are a number of epigenetic mechanisms of gene regulation that have been identified, and all of these are dysregulated in cancer (Virani et al. 2012). DNA methylation is by far the best characterized, and involves the covalent addition of a methyl group to a cytosine in a cytosine-guanine pair (CpG). The methyl group of 5-methylcytosine lies in the major groove of DNA and can directly interfere with binding of transcription factors, or can attract a class of proteins known as “methyl binding proteins” which can also block transcription factors (Fuks et al. 2003). A well-characterized role of DNA methylation takes place in the context of “CpG Islands”, dense stretches of CpGs typically found in promoter regions of genes. Methylation of CpG islands in gene promoters is usually associated with transcriptional silencing (Merlo et al. 1995). Besides the promoter region of genes, CpGs are distributed unevenly throughout

the genome, where they are overrepresented in repetitive regions and distal gene regulatory regions (Ponger et al. 2001). Tumors have distinct patterns of DNA methylation, where CpGs in repetitive regions, which are normally methylated, become hypomethylated, paired with aberrant methylation of promoter regions of genes, some of which are likely important drivers of carcinogenesis (Esteller and Herman 2002).

Experimental evidence accumulating over the past two decades has shown that within many tumors, the cellular hierarchy imitates that found during the development of normal tissue. Early observations by pathologists identified gross similarities between embryonic tissues and tumors, leading to the original hypothesis that cancers arise from embryo-like cells (Récamier 1829; Virchow 1855). Further work, particularly the seminal research of Becker, McCulloch, and Till in the 1960s, provided the first evidence for tissue-specific stem cells and led to the hypothesis that tumors may arise from dysregulation of this cell population (Becker et al. 1963; Till and McCulloch 1961). Advances in fluorescence activated cell sorting (FACS) have confirmed pathological observations of tumor heterogeneity, identifying that cancers are comprised of distinct cellular compartments that can be differentiated based on cell surface markers. These same techniques have allowed for these different tumor cell populations to be purified, assayed, and characterized. Cancer stem cells at the apex of the tumor hierarchy can differentiate to provide the complex cellular hierarchy found in a tumor (Wicha et al. 2006). Cancer stem cells are also defined by their ability to initiate a new tumor, suggesting an essential role in metastasis. These cells have become an intense subject of research (Kreso and Dick 2014). They have natural defense mechanisms, including increased adaptation to oxidative stress and drug resistance, which makes them resistant to traditional chemotherapeutics (Dean et al. 2005; Conley et al. 2012). In solid tumors, cancer stem cells, similar to normal stem cells, are rare cells that are often quiescent. As chemotherapeutics often target rapidly cycling cell populations, the stem cell population of the tumor can survive, despite reductions in tumor bulk. The identification of these rare cells that potentially drive tumorigenesis highlights the

need to better understand the role of epigenetics in cancer and cellular differentiation.

1.3 Epigenetics of HPV(+) vs HPV(-) Head and Neck Cancers

Head and neck squamous cell carcinomas (HNSCCs), the eighth most commonly diagnosed cancer in the U.S. population, have a complex etiology that includes life style behaviors, classical chemical carcinogenesis, and infection with high risk types of human papillomavirus (HPV). Traditionally, head and neck cancer is associated with a profound history of tobacco and alcohol use, and those with HNSCC experience poor survival compared to other cancers (Marur and Forastiere 2008). Over the last decade, high-risk HPV has emerged as a risk factor for head and neck cancer, particularly of the oropharynx (D'Souza et al. 2007; Gillison et al. 2000). Patients with HPV(+) head and neck cancer have a distinct risk profile, associated with a less remarkable history of tobacco and alcohol use (Gillison et al. 2008), a more beneficial micronutrient profile (Arthur et al. 2011), and improved survival compared to those with HPV(-) tumors (Fakhry et al. 2008)**.**

Both tobacco- and alcohol-related, as well as HPV-associated, head and neck cancers have a well-described multistep model of carcinogenesis (Califano et al. 1996). Broadly, mutations or loss of heterozygosity of major cell cycle regulator genes, such as *p53*, are frequently detected in tobacco and alcohol-related head and neck cancers (Lee et al. 2011; Somers et al. 1992), although mutation at these genes has not consistently been associated with patient survival. Likewise, HPV(+) head and neck cancers are associated with functional inactivation of *p53* and *Rb*, which is mediated by E6 and E7 viral oncoproteins, resulting in overexpression of *p16* (Boyer et al. 1996; Hafkamp et al. 2003; Werness et al. 1990). HPV(+) head and neck cancers have a distinct clinical profile when compared to alcohol and tobacco-related HPV(-) tumors, the former of which are typically more responsive to treatment (Kumar et al. 2008).

Epigenetic modifications are an important mechanism in carcinogenic progression (Hansen et al. 2011), but the epigenetic profiles between HPV(+) and HPV(-) tumors remain poorly characterized, with most studies focusing on

specific loci or global levels of DNA methylation (Marsit et al. 2006; Richards et al. 2009). A handful of epigenome-wide studies of head and neck cancer have focused on differences between normal and tumor tissue, associations with alcohol and tobacco exposure, and associations with global marks of DNA methylation (Marsit et al. 2009; Poage et al. 2011). We reported an epigenome-wide analysis of concurrently measured DNA methylation and gene expression in HPV(+) and HPV(**-**) squamous cell carcinoma cell lines, noting that HPV(+) cell lines have higher amounts of genic methylation as well as increased expression of *DNMT3A* (Sartor et al. 2011). Information about the specific epigenomic differences in DNA methylation based on clinical characteristics, including HPV infection, remain unknown, and require a well-characterized cohort of patient samples. Because of the particular morbidity associated with treatment for head and neck cancer, epigenetic biomarkers of survival may allow for de-escalation of treatment for those most likely to respond.

1.4 Epigenetics and Nutrition in Cancer

Epidemiologic evidence also supports the hypothesis that diet modulates the risk, progression and prognosis of head and neck cancer (Duffy et al. 2009; Lucenteforte et al. 2009; Sandoval et al. 2009). However, the molecular mechanisms by which dietary compounds exert their effects are not entirely understood. Dietary intake, particularly of micronutrients involved in the one-carbon metabolism pathway, has been widely hypothesized to influence epigenomic states by modifying the availability of functional groups required for DNA and histone protein modifications (Oommen et al. 2005). Early evidence from a clinical trial of 8 postmenopausal women fed a low folate diet for three months showed lymphocyte DNA hypomethylation, as quantified by the [^{3}H]-methyl acceptance assay, which was found to be reversible following high level folate supplementation for three weeks (Jacob et al. 1998). A similar study in elderly women also reported overall genomic hypomethylation following seven weeks of folate depletion, without the same recovery following dietary folate supplementation (Rampersaud et al. 2000). Restriction of another key factor in the one-carbon metabolism pathway, vitamin B12, was also found to significantly

decrease the DNA methylation of colonic epithelium in rats (Choi et al. 2004). These findings provide evidence that variation of intake of nutrients relevant to one-carbon metabolism can influence DNA methylation and potentially carcinogenesis.

Epidemiology studies of the effects of diet on epigenetics have typically focused on lifestage specific effects in populations exposed to food scarcity. Individuals exposed prenatally to famine in the Dutch Hunger Winter, during a German-imposed food embargo at the end of World War II, have been studied extensively for epigenetic dysregulation that could be linked to nutrient deprivation *in utero*. Individuals who were exposed to famine during early gestation had significantly lower methylation at the *IGF2* differentially methylated region in whole blood, as measured by quantitative mass spectrometry, than their unexposed same-sex siblings (Heijmans et al. 2008). A follow up study in the same population extended these findings to an additional 15 loci associated with cardiovascular and metabolic disease (Tobi et al. 2009). The proximal promoter region of *INSIGF* was found to be hypomethylated in famine exposed individuals compared to their unexposed sex-matched siblings, while *IL10*, *LEP*, *ABCA1*, *MEG3*, and *GNASAS* were all found to be hypermethylated in famine exposed individuals. These findings suggest that nutrient deprivation can lead to global DNA hypomethylation, but either loci-specific hyper- or hypomethylation based on both exposure timing and sex (Tobi et al. 2009). Interestingly, while nutrient deprivation *in utero* has been associated with an increased risk of developing chronic diseases ranging from Type 2 diabetes (Hales and Barker 1992) to renal disease (Hoy et al. 1999), individuals exposed to famine early *in utero* were found to be less likely to develop a CpG Island Methylator Phenotype (CIMP) colorectal tumor (Hughes et al. 2009). These examples suggest that once methylation and disease associations are established, the use of dietary interventions to modify the epigenome could have a large impact on how we influence human health.

1.5 Epigenetics and Cellular Hierarchies

Breast Development

Cancer can be thought of as a “hijacking” of the normal developmental machinery in a cell. Thus, understanding the role that stem cells play in breast cancer is aided by an understanding of normal breast development and tissue homeostasis throughout the life course. However, knowledge of the molecular mechanisms that govern human breast development is limited due to the difficulty in obtaining non-pathologic samples. Many inferences about human breast development are derived from research in model organisms, most commonly mice, although there have been a small number of studies that examine human fetal and infant breast structure (Anbazhagan et al. 1991; Anbazhagan et al. 1998; Jolicoeur et al. 2003; Jolicoeur 2005). Human breast development begins *in utero* typically around the fifth week of gestation with a thickening within the ectoderm of a 2 to 4 cell layer, termed the mammary band. During the sixth to seventh weeks, the thoracic region thickens to form the mammary crest, followed by involution of the mammary band. Soon after, the mammary crest forms into a nodule that sinks into the surrounding developing mesoderm defining the region that will become the developing nipple. Between 18-21 weeks, epithelial outgrowths form from the breast bud and invade the mesenchyme. These outgrowths quickly form a lumen, characterized by two layers of cubiodal epithelial cells, which over the course of the next few weeks form the early side branches, which define the tree-like structure of the mammary ductwork.

Interestingly, at birth, there is considerable, and currently unexplained, variation noted in breast structure between individuals (Anbazhagan et al. 1991). Some individuals are born with a full mammary ductwork, complete with terminal ductal lobular units and lobules, while other individuals are born with a simple series of dead-end tubes. Whether these differences can be attributed to gestational age at birth, *in utero* nutrition, the maternal hormonal environment, or other exogenous factors is currently unknown. Regardless of the developmental status of the mammary gland at birth, during the first few months of infancy, the glands undergo an involution similar to that observed during menopause, leading to mammary glands that are similar between the sexes. Leading up to puberty,

the breast tissue grows in proportion relative to other tissues, with an elongation of the rudimentary ducts proportionate to the growth of other organs.

The temporality of the gross anatomical and molecular changes that occur in the developing human mammary gland during puberty are currently poorly understood. Throughout puberty, the ducts that were established *in utero* undergo side branching, invading through the surrounding mammary stroma. This invasion is driven first by estrogen and epidermal growth factor (EGF), which control ductal elongation and branching. As puberty progresses and menarche begins, the cyclical fluctuations in the levels of estrogen and progesterone begin to drive the development of mammary lobules and alveoli. Throughout menstruation, monthly cycling of progesterone induces the proliferation of the breast epithelium, with the highest rates of proliferation observed in the luteal phase (Söderqvist et al. 1997). Additionally, throughout the menstrual cycle, morphological changes of the breast can be observed, with the formation of double layered acini during menstrual days 6 -15, followed by an increase in terminal ductal lobule units at days 16-24, and increased vacuolation and apoptosis within lobules in days 25-28 (Ramakrishnan et al. 2002).

During pregnancy, the breast undergoes dramatic remodeling in preparation for lactation, and following a successful term pregnancy, undergoes what amounts to the terminal differentiation of the organ. Early in pregnancy, increased secretion of progesterone from the ovaries leads to dramatic increases in the generation of alveoli, the structures which will secrete milk following birth (Macias and Hinck 2012). Prolactin, produced mainly in the pituitary gland but also in the mammary epithelium itself, is required for mammary ductal sidebranching, alveolar budding, and lobuloalveolar growth (Ormandy et al. 1997). After weaning, the mammary architecture undergoes involution that involves epithelial apoptosis, detachment of alveolar cells, and eventually, collapse of alveolar structure over the course of approximately a week. The involution that occurs during menopause involves the replacement of the mammary glands with adipose, ultimately reducing the epithelial content of the breast substantially.

Stem cells are known to play key roles throughout mammary development. The massive structural changes that occur in the breast, particularly during puberty and pregnancy, require the existence of cells with great proliferative potential that can also differentiate into multiple cell types, including the myoepithelial, luminal, and alveolar lineages of the breast. Serial transplantation assays, where a small tissue fragment is serially transplanted into cleared murine mammary fat pads and allowed to form a full mammary tree, originally identified that there is likely a mammary stem cell population that can regenerate the mammary structure (Daniel et al. 1968). Early studies of mammary development in mice identified that the tip of the mammary terminal end bud contained cap cells that can produce new myoepithelial cells required for ductal branching and morphogenesis, suggesting that the tip of the terminal end buds may be a site of mammary stem cells (Williams and Daniel 1983). A major advance in mammary stem cell biology came in 2006, when a bipotent murine mammary stem cell was characterized by isolating discrete populations of mammary cells utilizing cell surface markers and reimplanting single cells into cleared mammary fat pads to reconstitute a mammary gland (Shackleton et al. 2006). More recently, lineage tracing studies have identified that both the luminal and myoepithelial lineages have distinct populations of stem cells that demonstrate significant self-renewal capacity (Van Keymeulen et al. 2011). Further, an *in vivo* labeling study in mice provided evidence for the existence of long-lived bipotent stem cells in the basal portion of the mammary gland that drive proliferation throughout puberty and play important roles in mammary tissue homeostasis (Rios et al. 2014). These combined results provide strong evidence for the existence of a bipotent mammary stem cell that is essential for mammary development. The identification and isolation of these cells in the normal human breast is an area of great interest, and factors that influence their potential for self-renewal may identify not yet appreciated avenues for cancer prevention.

Identification of Normal and Cancer Stem Cells

A key feature of normal stem cells that has proven essential to isolate and characterize these cells in the breast is the expression of specific surface and

enzymatic markers of stemness. Early studies of the cellular hierarchy of the breast identified that the expression of the cell surface markers MUC-1- to ±/CD-10 ± to +/ESA+ isolated cells with the ability to develop colonies with both luminal and myoepithelial features (Stingl et al. 1998). A follow-up study identified that normal bipotent mammary progenitors are enriched in a cellular subfraction expressing both CD49f (α6 integrin) and EpCAM (Stingl et al. 2001). Recent work has identified that normal bipotent mammary stem cells are further enriched in the CD44+/CD24- fraction of CD49f+/EpCAM+ cells (Ghebeh et al. 2013). In addition to stem cell enriching cell surface protein markers, enzymatic markers of breast stemness have also been identified. Normal breast stem and progenitor cells were found to express high levels of aldehyde dehydrogenase 1 (ALDH1) (Ginestier et al. 2007). ALDH1 expressing cells can be identified using the non-immunological Aldefluor assay, where the substrate, Bodipy-aminoacetaldehyde, is converted intercellularly to fluorescent Bodipy-aminoacetate. Stem cells can also be isolated by exploiting their increased expression of ATP-binding cassette drug transporters (Bunting 2002). By staining cells with Hoechst, a DNA-binding dye that is effluxed by ATP-binding cassette transporters, one can discriminate populations of cells that are high and low Hoechst staining, with the low-Hoechst stained, stem cell enriched, fraction termed the “side population”.

A number of functional assays for the identification and classification of both normal and cancer stem cells have been established. The first series of experiments to identify the presence of tumor initiating cells utilized a transplantation assay of human acute myeloid leukemia cells into severe combined immunodeficient (SCID) mice (Lapidot et al. 1994). The SCID mice were examined for the presence of human leukemia cells, with limiting dilution experiments identifying that approximately 1 in 250,000 cells have the ability to engraft, with the cells most likely to engraft possessing the CD34+CD38- hematopoetic progenitor signature. In the human breast, cancer stem cells were first identified by injecting single cell suspensions of dissociated human breast tumor tissue into non-obese diabetic/severe combined immunodeficient (NOD/SCID) mice (Al-Hajj et al. 2003). Tumor cells were sorted based on cell

surface markers (CD44, CD24, EpCAM) and injected into the mammary fat pads of mice at limiting dilutions, revealing that as few as 200 ESA+/CD44+/CD24- cells were consistently able to form tumors, while injection with 20,000 CD44+/CD24+ cells failed to grow tumors (Al-Hajj et al. 2003). Furthermore, the ESA+CD44+CD24- cell fractions were able to recapitulate the original tumor phenotype after serial transplantations, showing that these cells possess the ability to both proliferate and differentiate into the different cell types that comprised the original tumor.

In addition to the transplantation assays described above, a number of other assays have been utilized to enrich and characterize normal and cancer stem cells. Neural stem cells were first discovered to grow in anchorage independent, serum free culture conditions forming free floating spheroids of neural cells termed neurospheres (Reynolds and Weiss 1996). These culture conditions were later adapted for mammary tissue, where both tumor and normal breast stem/progenitor cells were found to propagate under these conditions, termed mammosphere formation conditions (Dontu et al. 2003). The mammosphere formation assay is useful for characterizing three key aspects of stem cell biology: (1) Proliferation potential; (2) The ability to self-renew; and (3) the ability to differentiate into downstream progeny. Since each mammosphere is initiated by a single cell, proliferation capacity can be assessed by mammosphere size. Mammospheres can then be serially passaged to assess stem cell self-renewal capacity over time. Finally, mammospheres can be stained for known markers of luminal vs. myoepithelial differentiation, or plated into differentiating culture conditions to assay the bipotency of the stem cell population.

Identifying the Source of Cancer Stem Cells

While the evidence for a subpopulation of cancer cells with stem cell like properties is growing, the cell of origin for breast cancer stem cells has not been characterized. There are three competing theories for the origin of breast cancer stem cells. First, normal breast stem cells could acquire genetic and epigenetic changes that confer the ability to inappropriately undergo symmetric self-renewal

and initiate tumorigenesis. A second possibility is that cancer stem cells derive from lineage committed rapidly cycling progenitor cells that undergo mutations that reconfer stem-like properties. A final possibility is that a series of mutations in fully differentiated cells can lead to dedifferentiation to a tumorigenic stem state. As the body of research surrounding both normal and cancer stem cells grows, evidence for each of these potential pathways is accumulating.

In vivo labeling of mammary stem cells revealed that bipotent stem cells are detectable in adult animals, and are essential for both tissue remodeling and homeostasis (Rios et al. 2014). These cells are activated during puberty and pregnancy, and a subpopulation of long lived luminal progenitor cells was found to contribute extensively to ductal remodeling. Importantly, these long lived stem cells have the potential to accumulate the genetic and epigenetic effects over a lifetime's worth of environmental exposures, passing these mutations to their progeny. Additionally, recent evidence suggests that normal breast, as well as breast cancer, stem cells exist in two different states, luminal-like (ALDH+) and basal-like (CD44+/CD24-) (Liu et al. 2013; Van Keymeulen et al. 2011). Luminal stem cells have an epithelial phenotype and are proliferative, while basal stem cells have a mesenchymal phenotype and are invasive. As luminal stem cells have a greater proliferative capacity, their phenotype could therefore be more susceptible to the effects of environmental insult. Thus, the increased risk of breast cancer associated with exposures during windows of susceptibility, such as puberty or pregnancy, could reflect an increased vulnerability of cycling luminal stem cells to oxidative stress and DNA damage resulting from environmental stressors. Support for this hypothesis comes from recent studies of the role of breast stem cells in carcinogenesis. Despite a distinct basal-like tumor phenotype, *BRCA1* mutant tumors were experimentally found to arise from a luminal progenitor compartment (Molyneux et al. 2010). Transformation of EpCAM+ luminal breast cells was able to give rise to a range of breast tumors that recapitulate many subtypes observed in humans, while transformation of basal CD10+ cells solely gave rise to rare subset of tumors that resemble the claudin-low subtype (Keller et al. 2012). Luminal breast progenitors are also

more prone to have shortened telomeres, and can induce a telomere specific DNA damage response (Kannan et al. 2013). These data point to the luminal stem/progenitor compartment as a major source of tumor initiating cells, however, there is also evidence to suggest that dedifferentiation of lineage-committed may also play a role in the formation of cancer stem cells. In a landmark 2011 paper, Robert Weinberg's laboratory reported that both normal and cancer nonstem cells can spontaneously dedifferentiate into stem-like cells (Chaffer et al. 2011). A floating population of hTERT-immortalized human mammary epithelial cells was isolated and found to be enriched in CD44+/CD24-/EpCAM- and CD44-/CD24+/EpCAM- cells. When the CD44- fraction of these cells was isolated and plated into 2D culture, the cells were found to increasingly form CD44+ cells. Similar experiments conducted with primary cells found that after 12 days in culture, 6% of the CD44- cells had been stochastically converted to CD44+ cells (Chaffer et al. 2011). These striking results suggest that cell type, at least in culture, is very plastic, and that differentiated cells can convert to stem-like cells with relative ease. These results, however, stand in contrast to the low conversion efficiency of induced pluripotent stem cells (iPSCs). Normal human fibroblasts exposed to the required transcription factors to induce pluripotency (Oct3/4, Sox2, Klf4, c-Myc) have very low dedifferentiation efficiency, less than 1 in 1000 (Takahashi et al. 2007). The difference could potentially lie in the ease of transition from a differentiated breast cell to a breast stem cell, instead of a differentiated skin cell to a pluripotent stem cell. Additionally, there is the possibility that the different stem cell types in the breast (luminal-like vs. basal-like) interconvert, and that the "dedifferentiation" observed is actually due to a state transition between two types of stem cell. Recently, computational modeling of the effect of dedifferentiation on cancer stem cell populations found that estimated rate of dedifferentiation was heavily dependent on stem cell homeostasis (Jilkine and Gutenkunst 2014). In a model where stem cell regulation is tightly conserved, with each stem cell undergoing asymmetric self-renewal, dedifferentiation plays a minor role in the generation of cancer stem cells. If stem cell homeostasis is not tightly conserved, and stem cells are

allowed to symmetrically self-renew, dedifferentiation can lead to dramatic expansion of the stem cell population. These results point to key gaps in our understanding of normal stem cell homeostasis that are reflected in the currently unknown origin of cancer stem cells.

1.6 Targeting Stem Cells with Curcumin for Cancer Prevention

Curcumin is a dietary polyphenol derived from the rhizomes of turmeric (*curcuma longa*) which has been widely used in traditional Indian and Chinese medicine for treatment of a range of diseases, including inflammatory conditions, diabetes, and rheumatoid arthritis (Noorafshan and Ashkani-Esfahani 2013). Preclinical models have implicated curcumin as a potential agent for both the prevention and treatment of cancers. The promise of curcumin as a preventive agent and as a potential adjuvant to traditional cancer chemotherapy has led to considerable interest in translating these preclinical findings to the clinic (Gupta et al. 2013). To this regard, there are a number of completed and ongoing clinical trials examining the safety and efficacy of curcumin for treatment of a number of cancer types, including breast. A dose-escalation study identified that glucoronidated and sulfated metabolites of curcumin were detectable in the plasma of advanced colorectal cancer patients receiving 3.6g curcumin/day, while the parent compound was rarely detected (Sharma et al. 2004). Additionally, 3.6g/day curcumin treatment was found to significantly reduce the concentrations of the inflammatory mediator PGE2 in isolated blood leukocytes treated with liposaccharide (LPS), although basal circulating PGE2 levels were unchanged. These findings suggest that curcumin treatment inhibits PGE2 induction, which may represent a mechanism by which curcumin acts as an anti-inflammatory compound. A trial examining the effects of treatment with a combination of curcumin (480 mg) and the dietary flavonol quercitin (20mg) orally, three times per day, in 5 patients with familial adenomatous polyposis (FAP) found a significant reduction in both the number and size of the polyps (Cruz–Correa et al. 2006). Oral curcumin (4g/day for 30 days) also significantly reduced the number of aberrant crypt foci in 41 smokers who had 8 or more aberrant crypt foci, as diagnosed by colonoscopy, at baseline (Carroll et al.

2011). These results suggest that curcumin is effective at inhibiting the growth of early cancer-related lesions in the colon. Whether these results can be extended to other organs, such as the breast, where orally ingested curcumin may not be as bioavailable is unknown.

The wide range of diseases that can be treated with curcumin (Gupta et al. 2013) reflect the pleiotropic action of the compound. Relevant to tumorigenesis, curcumin has been shown to act as a potent anti-inflammatory and antioxidant compound. Curcumin downregulates the production of the pro-inflammatory cytokines TNF-α and IL-1β at a range of doses (5 – 20 μM) (Chan 1995; Cho et al. 2007), potentially through the inhibition of the binding of the AP-1 and NF-κB transcription factors (Bierhaus et al. 1997; Xu et al. 1996). Curcumin treatment also induces expression of the antioxidant enzyme Heme Oxygenase 1 (HO-1), through the dissociation of NRF2 from the NRF2-KEAP1 complex (Balogun et al. 2003). Curcumin is also a direct antioxidant, with the ability to scavenge hydroxyl radicals (Reddy and Lokesh 1994), nitrogen dioxide radicals (Rao 1994), and ferrous iron chelation (Ak and Gülçin 2008).

Curcumin has also been previously shown to inhibit stem cell self-renewal, a potential mechanism by which curcumin may exert its cancer preventive effects. Curcumin was found to inhibit primary and secondary mammosphere formation of cells isolated from normal human mammoplasty reduction tissue in a dose dependent manner (Kakarala et al. 2010). Increasing concentrations of curcumin were also shown to reduce the expression of the breast stem cell marker ALDH1A1, representing a functional readout of inhibition of stem cell self-renewal (Kakarala et al. 2010). Kakarala et al. also found that curcumin decreased the number of cells with nuclear beta-catenein, suggesting that curcumin also inhibits the activation of the Wnt signaling pathway (Kakarala et al. 2010). A number of studies in other cell types besides breast have also reported inhibition of Wnt signaling by curcumin, suggesting that this mechanism is conserved across tissues (Li and Zhang 2014). Curcumin was also found to reduce the size of the side population (Hoescht dye positive) in rat glioma cells in a dose dependent manner, with significant effects observed at 5uM treatment

(Fong et al. 2010). A study of the effects of curcumin on six esophageal squamous cell carcinoma cell lines found that cells which survived a 60uM curcumin treatment had significantly fewer ALDH1A1 cells, as quantified by immunocytochemistry (Almanaa et al. 2012). Additionally, curcumin treatment surviving cells were able to form tumorspheres at a much lower rate than the parental cell line, suggesting that curcumin treatment induced a change, potentially epigenetic, that limited the number of cancer stem cells in this population. Together, these results point to the efficacy in curcumin to limit stem cell self-renewal in both normal and cancer cells. The mechanism by which curcumin targets normal breast stem cells remains to be fully characterized, but likely involves modulation of Wnt signaling.

A major hurdle in the use of curcumin as a chemopreventive agent in a breast cancer susceptible patient population is the limited bioavailability of orally ingested curcumin. Curcumin is rapidly metabolized and eliminated, leading to low serum levels and poor tissue distribution, which limits its use as a therapeutic agent (reviewed in (Anand et al. 2007)). A number of different strategies to enhance the bioavailability of curcumin have been studied, with one of the most promising approaches being the use of adjuvants in concurrent dosing with curcumin. Piperine, an alkaloid compound derived from black pepper, has been shown to increase the bioavailability of curcumin up to 2000% in a study of human volunteers (Shoba et al. 1998). The mode of action of piperine's influence on the metabolism of curcumin likely involves inhibition of drug and nutrient metabolizing enzymes, including CYP3A4 and P-glycoprotein (Bhardwaj et al. 2002). Besides influencing the bioavailability of nutritive compounds such as curcumin, piperine also has anticarcinogenic activities, reducing cancer incidence in rodent models of lung carcinogenesis (Pradeep and Kuttan 2002; Selvendiran et al. 2004). Kakarala et al previously showed in a preclinical *in vitro* model that both curcumin and piperine, when given individually or in combination, can limit breast stem cell self renewal while remaining non-toxic to normal differentiated cells (Kakarala et al. 2010). The mechanisms by which both piperine and curcumin combine to affect influence breast self-renewal capacity are not known.

1.7 Research Objectives

The overall goal of this dissertation is to apply novel genomic and epigenomic methods in the context of epidemiological and *in vitro* stem cell studies to understand epigenetic mechanisms of carcinogenesis and cancer prevention in two cancer types with strong environmental influences, breast and head and neck. The specific aim of the work presented in Chapter 2 was to quantify the epigenetic differences between HPV(+) and HPV(-) head and neck tumors. The working hypothesis for this aim was that HPV-induced head and neck tumors have a distinct epigenetic profile, which reflects the viral etiology of the disease. This work was published in *PLOS ONE* in 2013 (Colacino et al. 2013). The specific aim of the work presented in Chapter 3 was to identify dietary factors associated with head and neck tumor DNA methylation, with the hypothesis for the aim being that increased intake of dietary nutrients involved in the one carbon metabolism pathway would influence the methylation profile of the tumor. The work in Chapter 3 was published in *Epigenetics* in 2012 (Colacino et al. 2012). Finally, the specific aim of the research summarized in Chapter 4 was to comprehensively characterize the effects of the cancer preventive compounds curcumin and piperine on normal human breast stem cells. The hypothesis for this aim was that curcumin and piperine treatment of normal breast cells will induce molecular changes that are involved in regulating breast stem cell self-renewal. These studies provide insight into the etiology of head and neck and breast cancers and provide novel routes of prevention and treatment of these deadly diseases.

1.8 References

Ak T, Gülçin İ. 2008. Antioxidant and radical scavenging properties of curcumin. Chemico-biological interactions 174:27-37.

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. 2003. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 100:3983-3988.

Almanaa TN, Geusz ME, Jamasbi RJ. 2012. Effects of curcumin on stem-like cells in human esophageal squamous carcinoma cell lines. BMC complementary and alternative medicine 12:195.

Anand P, Kunnumakkara AB, Newman RA, Aggarwal BB. 2007. Bioavailability of curcumin: problems and promises. Mol Pharm 4(6): 807-818.

Anbazhagan R, Bartek J, Monaghan P, Gusterson B. 1991. Growth and development of the human infant breast. American journal of anatomy 192:407-417.

Anbazhagan R, Osin PP, Bartkova J, Nathan B, Lane EB, Gusterson BA. 1998. The development of epithelial phenotypes in the human fetal and infant breast. The Journal of pathology 184:197-206.

Arthur AE, Duffy SA, Sanchez GI, Gruber SB, Terrell JE, Hebert JR, et al. 2011. Higher micronutrient intake is associated with human papillomavirus-positive head and neck cancer: A case-only analysis. Nutrition and cancer 63:734-742.

Balogun E, Hoque M, Gong P, Killeen E, Green C, Foresti R, et al. 2003. Curcumin activates the haem oxygenase-1 gene via regulation of nrf2 and the antioxidant-responsive element. Biochem J 371:887-895.

Becker AJ, McCulloch EA, Till JE. 1963. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells.

Bhardwaj RK, Glaeser H, Becquemont L, Klotz U, Gupta SK, Fromm MF. 2002. Piperine, a major constituent of black pepper, inhibits human P-glycoprotein and CYP3A4. J Pharmacol Exp Ther 302(2): 645-650.

Bierhaus A, Zhang Y, Quehenberger P, Luther T, Haase M, Müller M, et al. 1997. The dietary pigment curcumin reduces endothelial tissue factor gene expression by inhibiting binding of ap-1 to the DNA and activation of nf-kappa b. Thrombosis and haemostasis 77:772-782.

Boyer SN, Wazer DE, Band V. 1996. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. Cancer Research 56:4620-4624.

Bunting KD. 2002. Abc transporters as phenotypic markers and functional regulators of stem cells. STEM CELLS 20:11-20.

Califano J, van der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S, et al. 1996. Genetic progression model for head and neck cancer: Implications for field cancerization. Cancer Res 56:2488-2492.

Carroll RE, Benya RV, Turgeon DK, Vareed S, Neuman M, Rodriguez L, et al. 2011. Phase iia clinical trial of curcumin for the prevention of colorectal neoplasia. Cancer Prevention Research 4:354-364.

Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, et al. 2011. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proceedings of the National Academy of Sciences 108:7950-7955.

Chan MM-Y. 1995. Inhibition of tumor necrosis factor by curcumin, a phytochemical. Biochemical pharmacology 49:1551-1556.

Cho J-W, Lee K-S, Kim C-W. 2007. Curcumin attenuates the expression of il-1ß, il-6, and tnf-α as well as cyclin e in tnf-α-treated hacat cells: Nf-κb and mapks as potential upstream targets. Int J Mol Med 19:469-474.

Choi SW, Friso S, Ghandour H, Bagley PJ, Selhub J, Mason JB. 2004. Vitamin b-12 deficiency induces anomalies of base substitution and methylation in the DNA of rat colonic epithelium. J Nutr 134:750-755.

Colacino JA, Arthur AE, Dolinoy DC, Sartor MA, Duffy SA, Chepeha DB, et al. 2012. Pretreatment dietary intake is associated with tumor suppressor DNA methylation in head and neck squamous cell carcinomas. Epigenetics 7:883-891.

Colacino JA, Dolinoy DC, Duffy SA, Sartor MA, Chepeha DB, Bradford CR, et al. 2013. Comprehensive analysis of DNA methylation in head and neck squamous cell carcinoma indicates differences by survival and clinicopathologic characteristics. PLoS One 8:e54742.

Conley SJ, Gheordunescu E, Kakarala P, Newman B, Korkaya H, Heath AN, et al. 2012. Antiangiogenic agents increase breast cancer stem cells via the generation of tumor hypoxia. Proc Natl Acad Sci U S A 109:2784-2789.

Cruz–Correa M, Shoskes DA, Sanchez P, Zhao R, Hylind LM, Wexner SD, et al. 2006. Combination treatment with curcumin and quercetin of adenomas in familial adenomatous polyposis. Clinical Gastroenterology and Hepatology 4:1035-1038.

D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, et al. 2007. Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med 356:1944-1956.

Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M. 2005. Causes of cancer in the world: Comparative risk assessment of nine behavioural and environmental risk factors. Lancet 366:1784-1793.

Daniel CW, De Ome KB, Young J, Blair PB, Faulkin Jr L. 1968. The in vivo life span of normal and preneoplastic mouse mammary glands: A serial

transplantation study. Proceedings of the National Academy of Sciences of the United States of America 61:53.

Dean M, Fojo T, Bates S. 2005. Tumour stem cells and drug resistance. Nat Rev Cancer 5:275-284.

Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, et al. 2003. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev 17:1253-1270.

Duffy SA, Ronis DL, McLean S, Fowler KE, Gruber SB, Wolf GT, et al. 2009. Pretreatment health behaviors predict survival among patients with head and neck squamous cell carcinoma. J Clin Oncol 27:1969-1975.

Esteller M, Herman JG. 2002. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *The Journal of pathology* 196: 1-7.

Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. 2008. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 100:261-269.

Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. Cell 61:759-767.

Feinberg AP, Tycko B. 2004. The history of cancer epigenetics. Nature Reviews Cancer 4:143-153.

Feinberg AP, Irizarry RA. 2010. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A* 107 Suppl 1: 1757-64

Fong D, Yeh A, Naftalovich R, Choi TH, Chan MM. 2010. Curcumin inhibits the side population (sp) phenotype of the rat c6 glioma cell line: Towards targeting of cancer stem cells with phytochemicals. Cancer letters 293:65-72.

Forouzanfar MH, Foreman KJ, Delossantos AM, Lozano R, Lopez AD, Murray CJ, et al. 2011. Breast and cervical cancer in 187 countries between 1980 and 2010: A systematic analysis. The Lancet 378:1461-1484.

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. 2003. The methyl-cpg-binding protein mecp2 links DNA methylation to histone methylation. Journal of Biological Chemistry 278:4035-4040.

Ghebeh H, Sleiman GM, Manogaran PS, Al-Mazrou A, Barhoush E, Al-Mohanna FH, et al. 2013. Profiling of normal and malignant breast tissue show cd44high/cd24low phenotype as a predominant stem/progenitor marker when used in combination with ep-cam/cd49f markers. BMC cancer 13:289.

Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, et al. 2000. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer I 92:709-720.

Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. 2008. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst 100:407-420.

Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. 2007. Aldh1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell 1:555-567.

Gupta SC, Patchva S, Aggarwal BB. 2013. Therapeutic roles of curcumin: Lessons learned from clinical trials. The AAPS journal 15:195-218.

Hafkamp HC, Speel EJ, Haesevoets A, Bot FJ, Dinjens WN, Ramaekers FC, et al. 2003. 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. Int J Cancer 107:394-400.

Hales CN, Barker DJ. 1992. Type 2 (non-insulin-dependent) diabetes mellitus: The thrifty phenotype hypothesis. Diabetologia 35:595-601.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell 100:57-70.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: The next generation. Cell 144:646-674.

Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. 2011. Increased methylation variation in epigenetic domains across cancer types. Nat Genet 43:768-775.

Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, et al. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci U S A 105:17046-17049.

Hoy WE, Rees M, Kile E, Mathews JD, Wang Z. 1999. A new dimension to the barker hypothesis: Low birthweight and susceptibility to renal disease. Kidney Int 56:1072-1077.

Hughes LA, van den Brandt PA, de Bruine AP, Wouters KA, Hulsmans S, Spiertz A, et al. 2009. Early life exposure to famine and colorectal cancer risk: A role for epigenetic mechanisms. PLoS One 4:e7951.

Jacob RA, Gretz DM, Taylor PC, James SJ, Pogribny IP, Miller BJ, et al. 1998. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. J Nutr 128:1204-1212.

Jilkine A, Gutenkunst RN. 2014. Effect of dedifferentiation on time to mutation acquisition in stem cell-driven cancers. PLoS computational biology 10:e1003481.

Jolicoeur F, Gaboury LA, Oligny LL. 2003. Basal cells of second trimester fetal breasts: Immunohistochemical study of myoepithelial precursors. Pediatric and Developmental Pathology 6:398-413.

Jolicoeur F. 2005. Intrauterine breast development and the mammary myoepithelial lineage. Journal of mammary gland biology and neoplasia 10:199-210.

Kakarala M, Brenner DE, Korkaya H, Cheng C, Tazi K, Ginestier C, et al. 2010. Targeting breast stem cells with the cancer preventive compounds curcumin and piperine. Breast Cancer Res Treat 122:777-785.

Kannan N, Huda N, Tu L, Droumeva R, Aubert G, Chavez E, et al. 2013. The luminal progenitor compartment of the normal human mammary gland constitutes a unique site of telomere dysfunction. Stem Cell Reports 1:28-37.

Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, et al. 2012. Defining the cellular precursors to human breast cancer. Proc Natl Acad Sci U S A 109:2772-2777.

Kreso A, Dick JE. 2014. Evolution of the cancer stem cell model. Cell Stem Cell 14:275-291.

Kumar B, Cordell KG, Lee JS, Worden FP, Prince ME, Tran HH, et al. 2008. 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 26:3128-3137.

Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. 1994. A cell initiating human acute myeloid leukaemia after transplantation into scid mice.

Lee SH, Lee NH, Jin SM, Rho YS, Jo SJ. 2011. Loss of heterozygosity of tumor suppressor genes (p16, rb, e-cadherin, p53) in hypopharynx squamous cell carcinoma. Otolaryngol Head Neck Surg 145:64-70.

Lewin F, Norell SE, Johansson H, Gustavsson P, Wennerberg J, Biörklund A, et al. 1998. Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck. Cancer 82:1367-1375.

Li Y, Zhang T. 2014. Targeting cancer stem cells by curcumin and clinical applications. Cancer Lett 346:197-205.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. 2000. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from sweden, denmark, and finland. N Engl J Med 343:78-85.

Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. 2013. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports 2:78-91.

Lucenteforte E, Garavello W, Bosetti C, La Vecchia C. 2009. Dietary factors and oral and pharyngeal cancer risk. Oral Oncol 45:461-467.

Macias H, Hinck L. 2012. Mammary gland development. Wiley Interdisciplinary Reviews: Developmental Biology 1:533-557.

Marsit CJ, McClean MD, Furniss CS, Kelsey KT. 2006. Epigenetic inactivation of the sfrp genes is associated with drinking, smoking and hpv in head and neck squamous cell carcinoma. International Journal of Cancer 119:1761-1766.

Marsit CJ, Christensen BC, Houseman EA, Karagas MR, Wrensch MR, Yeh RF, et al. 2009. Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma. Carcinogenesis 30:416-422.

Marur S, Forastiere AA. 2008. Head and neck cancer: Changing epidemiology, diagnosis, and treatment. Mayo Clin Proc 83:489-501.

Marur S, D'Souza G, Westra WH, Forastiere AA. 2010. Hpv-associated head and neck cancer: A virus-related cancer epidemic. The lancet oncology 11:781-789.

Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, et al. 1995. 5′ cpg island methylation is associated with transcriptional silencing of the tumour suppressor p16/cdkn2/mts1 in human cancers. Nature medicine 1:686-692.

Molyneux G, Geyer FC, Magnay F-A, McCarthy A, Kendrick H, Natrajan R, et al. 2010. Brca1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell 7:403-417.

Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. 2012. The life history of 21 breast cancers. Cell 149:994-1007.

Noorafshan A, Ashkani-Esfahani S. 2013. A review of therapeutic effects of curcumin. Current pharmaceutical design 19:2032-2046.

Oommen AM, Griffin JB, Sarath G, Zempleni J. 2005. Roles for nutrients in epigenetic events. J Nutr Biochem 16:74-77.

Ormandy CJ, Camus A, Barra J, Damotte D, Lucas B, Buteau H, et al. 1997. Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. Genes & development 11:167-178.

Poage GM, Houseman EA, Christensen BC, Butler RA, Avissar-Whiting M, McClean MD, et al. 2011. Global hypomethylation identifies loci targeted for hypermethylation in head and neck cancer. Clin Cancer Res 17:3579-3589.

Ponger L, Duret L, Mouchiroud D. 2001. Determinants of CpG islands: expression in early embryo and isochore structure. *Genome Res* 11: 1854-60

Pradeep CR, Kuttan G. 2002. Effect of piperine on the inhibition of lung metastasis induced B16F-10 melanoma cells in mice. Clin Exp Metastasis 19(8): 703-708.

Ramakrishnan R, Khan SA, Badve S. 2002. Morphological changes in breast tissue with menstrual cycle. Modern pathology 15:1348-1356.

Rampersaud GC, Kauwell GP, Hutson AD, Cerda JJ, Bailey LB. 2000. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. Am J Clin Nutr 72:998-1003.

Rao M. 1994. Curcuminoids as potent inhibitors of lipid peroxidation. Journal of Pharmacy and Pharmacology 46:1013-1016.

Récamier JCA. 1829. Recherches sur le traitement du cancer: Par la compression méthodique simple ou combinée, et sur l'histoire générale de la même maladie:Gabon.

Reddy ACP, Lokesh BR. 1994. Studies on the inhibitory effects of curcumin and eugenol on the formation of reactive oxygen species and the oxidation of ferrous iron. Molecular and cellular biochemistry 137:1-8.

Reik W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447:425-432.

Reynolds BA, Weiss S. 1996. Clonal and population analyses demonstrate that an egf-responsive mammalian embryonic cns precursor is a stem cell. Developmental biology 175:1-13.

Richards KL, Zhang B, Baggerly KA, Colella S, Lang JC, Schuller DE, et al. 2009. Genome-wide hypomethylation in head and neck cancer is more pronounced in hpv-negative tumors and is associated with genomic instability. PLoS One 4:e4941.

Rios AC, Fu NY, Lindeman GJ, Visvader JE. 2014. In situ identification of bipotent stem cells in the mammary gland. Nature.

Sandoval M, Font R, Manos M, Dicenta M, Quintana MJ, Bosch FX, et al. 2009. The role of vegetable and fruit consumption and other habits on survival following the diagnosis of oral cancer: A prospective study in spain. Int J Oral Maxillofac Surg 38:31-39.

Sartor MA, Dolinoy DC, Jones TR, Colacino JA, Prince ME, Carey TE, et al. 2011. Genome-wide methylation and expression differences in hpv(+) and hpv(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. Epigenetics 6:777-787.

Selvendiran K, Banu SM, Sakthisekaran D. 2004. Protective effect of piperine on benzo(a)pyrene-induced lung carcinogenesis in Swiss albino mice. Clin Chim Acta 350(1-2): 73-78.

Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat M-L, et al. 2006. Generation of a functional mammary gland from a single stem cell. Nature 439:84-88.

Sharma RA, Euden SA, Platton SL, Cooke DN, Shafayat A, Hewitt HR, et al. 2004. Phase i clinical trial of oral curcumin biomarkers of systemic activity and compliance. Clinical Cancer Research 10:6847-6854.

Shoba G, Joy D, Joseph T, Majeed M, Rajendran R, Srinivas PS. 1998. Influence of piperine on the pharmacokinetics of curcumin in animals and human volunteers. Planta Med 64(4): 353-356.

Söderqvist G, Isaksson E, von Schoultz B, Carlström K, Tani E, Skoog L. 1997. Proliferation of breast epithelial cells in healthy women during the menstrual cycle. American journal of obstetrics and gynecology 176:123-128.

Somers KD, Merrick MA, Lopez ME, Incognito LS, Schechter GL, Casey G. 1992. Frequent p53 mutations in head and neck cancer. Cancer Res 52:5997-6000.

Stingl J, Eaves CJ, Kuusk U, Emerman J. 1998. Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. Differentiation 63:201-213.

Stingl J, Eaves CJ, Zandieh I, Emerman JT. 2001. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. Breast cancer research and treatment 67:93-109.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131:861-872.

Till JE, McCulloch EA. 1961. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiation research 14:213-222.

Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, et al. 2009. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet 18:4046-4053.

Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, et al. 2011. Distinct stem cells contribute to mammary gland development and maintenance. Nature 479:189-193.

Virani S, Colacino JA, Kim JH, Rozek LS. 2012. Cancer epigenetics: A brief review. ILAR J 53:359-369.

Virchow R. 1855. Editorial. Arch Pathol Anat Physiol Klin Med 8:23.

Werness BA, Levine AJ, Howley PM. 1990. Association of human papillomavirus types 16 and 18 e6 proteins with p53. Science 248:76-79.

Wicha MS, Liu S, Dontu G. 2006. Cancer stem cells: An old idea—a paradigm shift. Cancer research 66:1883-1890.

Williams JM, Daniel CW. 1983. Mammary ductal elongation: Differentiation of myoepithelium and basal lamina during branching morphogenesis. Developmental biology 97:274-290.

Xu Y, Pindolia K, Janakiraman N, Chapman R, Gautam S. 1996. Curcumin inhibits il1 alpha and tnf-alpha induction of ap-1 and nf-kb DNA-binding activity in bone marrow stromal cells. Hematopathology and molecular hematology 11:49-62.

CHAPTER 2. COMPREHENSIVE ANALYSIS OF DNA METHYLATION IN HEAD AND NECK SQUAMOUS CELL CARCINOMA INDICATES DIFFERENCES BY SURVIVAL AND CLINICOPATHOLOGIC CHARACTERISTICS

2.1 Abstract

Head and neck squamous cell carcinoma (HNSCC) is the eighth most commonly diagnosed cancer in the United States. The risk of developing HNSCC increases with exposure to tobacco, alcohol and infection with human papilloma virus (HPV). HPV-associated HNSCCs have a distinct risk profile and improved prognosis compared to cancers associated with tobacco and alcohol exposure. Epigenetic changes are an important mechanism in carcinogenic progression, but how these changes differ between viral- and chemical-induced cancers remains unknown. CpG methylation at 1505 CpG sites across 807 genes in 68 well-annotated HNSCC tumor samples from the University of Michigan Head and Neck SPORE patient population were quantified using the Illumina Goldengate Methylation Cancer Panel. Unsupervised hierarchical clustering based on methylation identified 6 distinct tumor clusters, which significantly differed by age, HPV status, and three year survival. Weighted linear modeling was used to identify differentially methylated genes based on epidemiological characteristics. Consistent with previous *in vitro* findings by our group, methylation of sites in the *CCNA1* promoter was found to be higher in HPV(+) tumors, which was validated in an additional sample set of 128 tumors. After adjusting for cancer site, stage, age, gender, alcohol consumption, and smoking status, HPV status was found to be a significant predictor for DNA methylation at an additional 11 genes, including *CASP8* and *SYBL1*. These findings provide insight into the epigenetic regulation of viral vs. chemical carcinogenesis and could provide novel targets for

development of individualized therapeutic and prevention regimens based on environmental exposures.

2.2 Introduction

Head and neck squamous cell carcinomas (HNSCCs), the eighth most commonly diagnosed cancer in the U.S. population, have a complex etiology that includes life style behaviors, classical chemical carcinogenesis, and infection with high risk types of human papillomavirus (HPV). Traditionally, head and neck cancer is associated with a profound history of tobacco and alcohol use, and poor survival compared to other cancers (Marur and Forastiere 2008). Over the last decade, high-risk HPV has emerged as a risk factor for head and neck cancer, particularly of the oropharynx (D'Souza et al. 2007; Gillison et al. 2000). Patients with HPV(+) head and neck cancer have a distinct risk profile, associated with a less remarkable history of tobacco and alcohol use (Gillison et al. 2008), a more beneficial micronutrient profile (Arthur et al. 2011), and improved survival compared to those with HPV(-) tumors (Fakhry et al. 2008).

Both tobacco- and alcohol-related, as well as HPV-associated, head and neck cancers have a well-described multistep model of carcinogenesis (Califano et al. 1996). Broadly, mutations or loss of heterozygosity of major cell cycle regulator genes, such as *p53*, are frequently detected in tobacco and alcohol-related head and neck cancers (Lee et al. 2011; Somers et al. 1992), although mutation at these genes has not consistently been associated with patient survival. Likewise, HPV(+) head and neck cancers are associated with functional inactivation of *p53* and *Rb*, which is mediated by E6 and E7 viral oncoproteins, resulting in overexpression of *p16* (Boyer et al. 1996; Hafkamp et al. 2003; Werness et al. 1990). Conversely, HPV(+) head and neck cancers have a distinct clinical profile when compared to alcohol and tobacco-related HPV(-) tumors, the former of which are typically more responsive to treatment (Kumar et al. 2008b).

Epigenetic modifications are an important mechanism in carcinogenic progression (Hansen et al. 2011), but the epigenetic profiles between HPV(+) and HPV(-) tumors remain poorly characterized, with most studies focusing on specific loci or global markers of DNA methylation (Marsit et al. 2006; Richards et

al. 2009). A handful of epigenome-wide studies of head and neck cancer have focused on differences between normal and tumor tissue, associations with alcohol and tobacco exposure, and associations with global marks of DNA methylation (Marsit et al. 2009; Poage et al. 2011).

Recently, we reported an epigenome-wide analysis of concurrently measured DNA methylation and gene expression in HPV(+) and HPV(-) squamous cell carcinoma cell lines, noting that HPV(+) cell lines have higher amounts of genic methylation as well as increased expression of *DNMT3A* (Sartor et al. 2011). Information about the specific epigenome-wide differences in DNA methylation based on clinical characteristics, including HPV infection, remain unknown, and require a well-characterized cohort of patient samples. In this study, a comprehensive methylation bead array was used to measure DNA methylation at 1505 CpG sites across 807 genes in both HPV(-) and HPV(+) head and neck cancer in tumor samples collected from the ongoing patient cohort in the University of Michigan Head and Neck Specialized Program of Research Excellence (SPORE). In addition, important survival differences by epigenetic profiles are identified as described. Findings from this study provide insight into the epigenetic regulation of viral vs. chemical carcinogenesis and provide novel targets for development of individualized therapeutic regimens based on environmental exposures.

2.3 Methods

Design

Subjects for this study were obtained from a prospective, cohort study of patients enrolled in the University of Michigan Head and Neck Cancer SPORE. Newly diagnosed patients were recruited, provided informed consent, and followed quarterly for 2 years and then yearly thereafter. In addition tumor samples were collected. Institutional Review Board approval was approved from all participating sites including the Institutional Review Boards of the University of Michigan Medical School and the Institutional Review Board for Human Subject Research at the Veterans Affairs Ann Arbor Healthcare System.

Study Population

Individuals eligible for participation included patients diagnosed with first primary head and neck cancer between January 1, 2003 and December 31, 2005 completed an epidemiologic questionnaire, and had paraffin-embedded tumors available for analysis with adequate residual tissue for microdissection (N=82). The epidemiologic questionnaire included questions about lifestyle behaviors, including smoking and drinking. Clinical characteristics included tumor site and stage, comorbidities, depression, quality of life, and recurrence status, as well as treatment modalities. Tumor blocks were re-cut for uniform histopathologic review and microdissection, with the first and last slides in a series of 12 reviewed by a qualified pathologist (JM) to confirm the original diagnosis and to circle areas for DNA extraction. Percent cellularity was estimated for each tumor and areas with >70% cellularity of cancer were designated for use in the analyses.

Laboratory Methods

FFPE tissue, DNA isolation, bisulfite conversion

Regions identified for DNA extraction were cored from the formalin fixed paraffin embedded (FFPE) tissue blocks using an 18 gauge needle. Isolation of DNA from cored tissue samples was performed using the QIAamp DNA FFPE Tissue Kit (Qiagen, Valencia, CA) modified to include overnight incubation at 56^{o}C in lysis buffer. DNA concentration and purity were confirmed via NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA). Sodium bisulfite modification was performed on 500ng to 1µg of extracted DNA using the EZ DNA Methylation kit (Zymo Research, Orange, CA) following the manufacturer's recommended protocol.

HPV testing

HPV status was determined by an ultra-sensitive 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. (Tang et al. 2012) . 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 (Kumar et al. 2008a; Maxwell et al. 2010a; Maxwell et al. 2010b; Worden et al. 2008).

Bead Array Methods

The commercially available Illumina Goldengate® Methylation Cancer Panel was used to detect DNA methylation patterns in tumor samples. The Cancer Panel measure DNA methylation at 1505 CpG sites located in known CpG islands across 807 genes related to cancer, including oncogenes, tumor suppressor genes, imprinted genes, and genes involved in cell cycle regulation, DNA repair, apoptosis and metastasis. Bead arrays were processed at the University of Michigan DNA Sequencing Core Facility according to the manufacturer's protocol. Briefly, bisulfite converted tumor DNA was hybridized to the bead array as described previously (Bibikova et al. 2006), and bead arrays were imaged using Illumina BeadArray Reader software. Raw bead array fluorescence data were initially analyzed using Illumina BeadStudio Methylation software, which converts fluorescence values of the methylated (Cy5) and unmethylated (Cy3) alleles into an average methylation value at a specific probe using the formula $\beta = [Max(Cy5,0)]/[Max(Cy5,0) + Max(Cy3,0) + 100]$, ranging from completely unmethylated ($\beta = 0$) to completely methylated ($\beta \approx 1$). For each probe, background fluorescence, as estimated from a set of negative controls, was subtracted. Fourteen of the 82 samples (17.1%) failed on the array were excluded from further analyses, resulting in a final sample size of 68 tumors. Methylation at specific CpG probes on the Goldengate BeadArray has been shown to be biased by probe thermodynamic properties (Kuan et al. 2010). Known biases include probe length and GC content, which can affect the melting

temperature of the probes as well as probe fluorescence intensities. Thus, we used the method proposed by Kuan et al. to normalize our average β values based on probe length and GC content (Kuan et al. 2010).

Detection p-values on the Goldengate BeadArray are calculated based on fluorescence signal at a probe compared to background fluorescence and represent the (1-probability) that a signal is stronger than background fluorescence. The weighted methodology proposed by Kuan et al. was used to develop sample and site weights based on p-values of detection, with sites and samples with lower p-values of detection given higher weights. Both samples and sites with larger detection p-values are generally considered less reliable and were down-weighted in further gene specific analyses.

Site Specific Validation

DNA methylation of four CpG sites in the promoter region of *CCNA1* was quantified in an additional sample of 128 pretreatment head and neck tumors using the Sequenom EpiTyper, a MALDI-TOF mass spectrometry based platform. DNA was extracted from FFPE tumors, HPV status was determined, and the DNA was bisulfite converted as described above. Bisulfite PCR amplification was performed using FastStart *Taq* Polymerase (Roche Diagnostics, Indiana, US) with a forward and reverse primer concentration of 0.2 μM and an annealing temperature of 48C and 45 PCR cycles. The primer sequences, including the forward and T7 promoter tags required for Sequenom analysis were: 5'-AGGAAGAGAGATGTATTTTGGATTTTTTATTGGGG (forward primer) and 5'-CAGTAATACGACTCACTATAGGGAGAAGGCTAAAAAAACATTCTAACAAACCTCCA (reverse primer). Methylation analyses were performed at the University of Michigan Sequencing Core Facility following the manufacturer's recommended protocol.

Statistical Methods

Unsupervised hierarchical clustering was performed using the Euclidean distance metric and the Ward clustering method in the *hclust* package in R version 2.10.1.(Wang and Zhu 2008). All 68 tumor samples were included in the

hierarchical clustering algorithm. To minimize sex-specific effects, we excluded CpG sites on the sex chromosomes. The cluster analysis was performed using three different cutoffs for inclusion of individual CpG sites; the 50%, 25%, and 10% of CpG sites with the highest variance in methylation across samples. Clinical characteristics were evaluated across clusters based on cluster membership using non-parametric rank-based and exact statistics. For survival analyses, death was considered an “event”; survival time was censored at 3 years (1095 days). The Kaplan-Meier method was used to estimate overall survival and the log-rank test was used to test differences in survival distributions amongst the subtypes using the R *survival* package. Differences in age were compared using analysis of variance (ANOVA). Fisher’s Exact test was performed to test differences in cancer site, stage, and HPV status. Cox proportional hazards models were constructed to test the association between methylation at each CpG site on the Goldengate BeadArray and survival, adjusting for HPV status, gender, age, disease stage, cancer site, smoking status, and problem drinking using the *coxph* function in the R *survival* package. Individuals with a tumor testing positive for any strain of HPV were considered HPV positive. Age was treated as a continuous variable, while disease stage and cancer site were treated as categorical variables. Smoking was categorized, as an ordinal variable, into never smoker, past former smoker (quit more than 12 months ago), recent former smoker (quit in previous 12 months), and current smoker. Problem drinking was defined as a score of greater than 8 on the validated Alcohol Use Disorders Identification Test, as previously described (Duffy et al. 2009). Due to the simultaneous testing of multiple proportional hazard models, we controlled the false discovery rate by calculating the false discovery rate q-value (Storey and Tibshirani 2003). Q-values were calculated using the *qvalue* R package.

Overall site specific methylation differences between HPV(+) and HPV(-) tumors were compared by calculating the difference in the mean methylation per CpG site in HPV(+) and HPV(-) tumors. The effects of clinical characteristics on DNA CpG methylation measured on the Goldengate array were examined using

Limma in R 2.10.1 (Smyth 2004). Sample weights generated with *LumiWCluster* based on detection p-values across samples were used in the *lmFit* function from the *Limma* package to downweight samples with higher detection p-values. CpG sites were identified as differentially methylated between HPV(+) and HPV(-)tumors, adjusting for cancer site (oral cavity, nasopharynx, oropharynx, hypopharynx, larynx), cancer stage, sex and age. An empirical Bayes method (using the *eBayes* function in *Limma*) was used to rank CpG sites in order of significance of differential methylation. Additionally, *Limma* was used to examine methylation differences between the case cluster with significantly worse survival compared to the remaining cases by comparing methylation for individuals in the cluster with worst survival to all other individuals. For CCNA1 validation, mean methylation was calculated across the 4 sites measured by the EpiTyper and compared across HPV(+) and HPV(-) tumors using the Wilcoxon rank-sum test. Additionally, a multiple linear regression model was constructed with mean CCNA1 methylation as the independent variable and HPV status as the main predictor, adjusting for age, sex, tumor site, and tumor stage. Gene Set Enrichment Analysis (GSEA) was used to identify common pathways and chromosomal locations for genes identified as significant ($p<0.05$) in the *Limma* analysis (Subramanian et al. 2005). Statistically significant genes were ranked by t-value and input into GSEA as a ranked list. The full list of genes assayed on the Goldengate BeadArray were input into GSEA as a chip platform file, which provided the background for the enrichment analysis. Weighted enrichment statistics were calculated by the GSEA software, using a minimum analyzed gene set size of 5.

2.4 Results

Descriptive statistics: Study samples

The mean age of the 68 subjects was 57 years (range: 28 - 82 years)**;** 75% of the subjects were male. The majority of the HNSCCs were from the oropharynx (47%), oral cavity (25%), and larynx (19%), with a large proportion of cancers diagnosed as late stage (22% stage III and 62% stage IV, **Table 2.1**). Approximately one-third of the tumors tested positive for HPV (20 HPV-16, 2

HPV-18, 1 HPV-35 and 1 HPV-59). The majority of the patients were former (60%) or current (24%) cigarette smokers and 34% screen positive for problem drinking. All patients were treated in a standardized fashion with single modality treatment for patients with early stage tumors (Stage I/II) and combined chemotherapy and radiation and in some cases surgery for patients with advanced (Stage III/IV) cancers. Median follow up for the entire cohort was 60 months (95% CI: 59.9, 60.0).

General clustering: cluster characteristics

Excluding CpG sites located on the sex chromosomes (n=84), and limiting the cluster analysis to the 50% of CpG sites with the most variance (n=711), six distinct clusters were identified (**Figure 2.1)**. Clusters by epidemiological and clinical characteristics were assessed first. Individuals who grouped in Cluster 3 tended to be older (mean age = 61.6 years, **Table 2.2**) and were significantly more likely to be HPV positive (62%, p = 0.02). There was no significant difference in the proportion of individuals who were problem drinkers in each of the clusters. Tumor samples from individuals grouped into Cluster 5 were more likely to have widespread DNA hypomethylation, while tumor tissue from individuals in Clusters 3 and 4 tended to have higher levels of methylation in the most differentially methylated genes. A similar distribution of epidemiological characteristics was observed across clusters when including only the 25% of CpG sites with the greatest variance, which revealed 3 distinct clusters, with HPV (p=0.004) and age (p=0.04) remaining statistically significantly different. These differences were not observed when restricting the analysis to only the top 10% most variable CpG sites, , where 4 distinct tumor clusters were observed, and neither age (p=0.41) nor HPV (p=0.07) were statistically significantly different across clusters. There was no clear clustering of the tumors from the HPV-18, HPV-35, or HPV-59 individuals.

Next, cluster membership was characterized by survival. Three year survival was compared between the six clusters (**Figure 2.2**). Overall, individuals in Cluster 3 had the best three year survival (86%) while individuals in Cluster 5 had the worst overall survival (25%). Cluster membership was found to be a

significant predictor of three year survival (p=0.02). HPV(+) cases were found to have statistically significant better three year survival than HPV(**-**) cases (p=0.03). Interestingly, Cluster 3 had the highest proportion of Stage IV disease, the highest proportion of HPV(+) tumors, and the best three-year survival, while Cluster 5 had the lowest proportion of Stage IV disease and the worst survival. This aligns with previous findings that HPV positive tumors have a better prognosis, leading to the increased survival rates observed for Cluster 3 (Kumar et al. 2008b). This also aligns with the observation that many HPV-positive patients present with advanced nodal disease.

CpG Site-Specific Methylation Differences by HPV Status

Plotting average differences in methylation at each site showed that HPV(+) tumors tended to be hypermethylated at more sites than HPV(-) tumors (**Figure 2.3**). In order to better understand how HPV infection affects the DNA methylation profile in head and neck cancer, associations between methylation at each of the 1505 CpG sites on the Goldengate array and HPV status were calculated. Thirteen individual CpG sites on the array were found to be significantly associated with the HPV status of the tumor with a q-value <0.05 (**Table 2.3**)**.** The top hit, a CpG site located slightly downstream of the transcription start site of *CCNA1* in a CpG island, was found to be significantly more methylated in HPV(+) tumors ($p = 1.8x10^{-6}$). This finding corroborates our recent analysis of epigenome-wide DNA methylation differences in HPV(+) and HPV(-) cell lines where *CCNA1* was found to be a major interaction hub following bioinformatic analyses (Sartor et al. 2011). CpG sites in *GRB7*, *CDH11*, *RUNX1T1, SYBL1,* and *TUSC3* were also found to be significantly more methylated in HPV(+) tumors. CpG sites in *SPDEF, RASSF1, STAT5A, MGMT, ESR2, JAK3,* and *HSD17B12* were found to be significantly hypomethylated in HPV(+) tumors (**Table 2.4)**.

CCNA1 Site Specific Validation

To validate our findings of increased *CCNA1* methylation HPV(+) tumors, we quantified methylation at 4 CpG sites in the promoter region of *CCNA1* in an additional 128 pretreatment head and neck tumors. Mean *CCNA1* methylation

was significantly higher in HPV(+) tumors (p=0.0005). After adjusting for age, sex, tumor site, and tumor stage, HPV(+) tumors were found to be, on average, 9.6% more methylated at the *CCNA1* promoter compared to HPV(-) tumors (p=0.029).

Gene Set Enrichment Analysis (GSEA)

To analyze if specific gene sets or pathways display differential epigenetic regulation in HPV(+) versus HPV(-) tumors, a GSEA of the genes associated with HPV status was conducted. An analysis of Gene Ontology (GO) Biological Processes significantly enriched for differentially methylated genes implies that three gene sets associated with cell cycle regulation were hypomethylated in HPV(+) tumors (**Table 2.5**). Specific genes included in these gene sets that were significantly less methylated in HPV(+) ($p < 0.05$) include *RASSF1*, *CDK10*, *CHFR*, *RUNX3*, *APC*, and *CDKN2A* (*p16*). An analysis of enriched gene sets from the Kyoto Encyclopedia of Genes and Genomes (KEGG) found that genes associated with Neuroactive Ligand Receptor Interactions were hypermethylated in HPV(+) tumors (**Table 2.5**). The specific genes from this KEGG pathway include *GRPR*, *MC2R*, *GABRA5*, *PRSS1*, *NTSR1*, and *F2R.* Additionally, genes from the enriched KEGG set JAK-STAT Signaling Pathway were found to be hypomethylated in HPV(+) tumors, specifically *STAT5A*, *JAK3*, *OSM*, *MPL*, and *EPO*. None of these pathways were significantly enriched ($q<0.05$) after adjusting for multiple comparisons, however.

CpG Sites Associated with Survival

Cox Proportional Hazards Regression was used to determine whether methylation at individual CpG sites is associated with three year survival rates. Significantly associated genes (p-value<0.05) are listed in **Table 2.6**. While no individual CpG site was found to have a false discovery rate less than 0.15, methylation at a number of genes was found to be potentially associated with survival, including *NOTCH1*, *UGT1A1*, and *IL-6.*

After comparing survival by cluster, where we noted that cases in Cluster 5 had significantly worse survival as well as apparent widespread differences in methylation, we conducted a post hoc analysis to identify specific genes

differentially methylated in that cluster. After adjusting for other clinical covariates, including age, sex, cancer site, stage, smoking, problem drinking, and HPV status, a substantial number of genes were found to be differentially methylated in Cluster 5 compared to all other clusters (**Table 2.7**). Gene set enrichment analysis identified pathways, molecular functions, and a chromosomal region significantly differentially methylated in Cluster 5 cases (**Table 2.8**). Genes located in chromosome 7q21 were found to be significantly hypomethylated in Cluster 5 cases. Biological processes associated with negative regulation of cellular metabolism as well as homeostatic processes were found to be enriched with genes hypomethylated in this cluster. An analysis of molecular functions identified dysregulation of nucleotide binding, particularly purine and adenyl nucleotide binding as well as kinase and phosphorus transferase activity.

2.5 Discussion

Using an epidemiologically well characterized sample of head and neck cancer patients with a high proportion of HPV(+) cases, we confirmed a distinct epigenetic profile in HPV(+) head and neck cancers when compared to HPV(-) cancers. This has been previously noted by others for global methylation (Richards et al. 2009), candidate gene methylation (Weiss et al. 2011) and by Marsit et al. using the same platform as this study (Marsit et al. 2009). Other studies have described the association between methylation and traditional risk factors for HPV(-) head and neck cancer such as smoking and alcohol use (Smith et al. 2007).

Our prior work has shown how epigenetic profiles and expression patterns correspond to these divergent mechanisms of carcinogenesis in HPV(+) and HPV(-) cell lines (Sartor et al. 2011). The findings of this study expand upon our prior cell line work, identifying numerous loci in tumor samples that are differentially methylated between HPV(+) and HPV(-) tumors, particularly those involved in cell cycle regulation and JAK-STAT signaling. The top differentially methylated site on the array between HPV(+) and HPV(-) tumors was seven bases downstream from the transcription start site of *CCNA1* with an average

percent methylation level of 10% in HPV(-) tumors and 31% in HPV(+) tumors. This was also one of our top ranked genes in HNSCC cell lines (Sartor et al. 2011), and was noted by other groups as differentially methylated (Weiss et al. 2011) and differentially expressed (Weiss et al. 2012) in HPV (+) HNSCC, indicating that methylation and expression of this gene could likely be important both mechanistically and as a biomarker for HPV-associated HNSCC. *CCNA1* is an important regulator of the cell cycle and is required for S phase and passage through G2 (Girard et al. 1991). Other genes involved in cell cycle regulation tended to be hypomethylated in HPV(+) compared to HPV(-) HNSCC, indicating that regulation of these pathways may be important for HPV(+) head and neck carcinogenesis. This hypomethylated set of genes included many genes that have previously been shown to be methylated in head and neck cancer, including *RASSF1* (Paluszczak et al. 2011), *CHFR* (Toyota et al. 2003), *RUNX3* (Tsunematsu et al. 2009), *APC* (Uesugi et al. 2005), and *CDKN2A* (*p16*) (El-Naggar et al. 1997). These results are of particular importance to studies of biomarkers for head and neck cancer, which frequently do not take HPV status into account (Demokan et al. 2010; Guerrero-Preston et al. 2011; Viet and Schmidt 2008).

These analyses represent essentially a sizeable candidate-gene study, and the large number of loci allowed for initial pathway and positional analyses of the methylated CpGs. This was particularly useful when evaluating the contribution of epigenetic modifications to the prediction of survival, where methylation at single genes or sites did not predict survival time in this cohort. Hierarchical cluster analysis identified one set of patients with particularly worse survival solely based on methylation. Notably, this cluster did not include any HPV(+) cases, and contained the lowest proportion of males of all clusters (63%). This cluster had significant hypomethylation of 7q21, a region amplified in multiple cancers (Holzmann et al. 2004; Takada et al. 2005). This region has been identified as containing a placental-specific imprinted gene region (Monk et al. 2008), which is epigenetically inactivated in prostate carcinoma (Ribarska et

al. 2010). Thus, epigenetic regulation of this region may also play a role in a subset of head and neck cancers.

These and other epigenetic studies have strong implications for head and neck cancer research, particularly in light of recent reports on the complex landscape of head and neck cancer research (Agrawal et al. 2011; Stransky et al. 2011). For example, the mutation rate of HPV-associated tumors was reported to be much lower than HPV(-) tumors by exome sequencing (from 2 to 5 times less likely to harbor mutations). The results of this study indicate that HPV-associated tumors are likely driven to a larger extent by methylation changes than HPV(-) tumors. Additionally, it is intriguing to hypothesize that methylation could serve as a complementary mechanism of inactivation in many known candidate tumor suppressor genes. For example, methylation of *NOTCH1* was the strongest predictor of survival in this study (p=0.0002), and was also identified as frequently mutated in head and neck tumors in Stransky et al. and Agrawal et al. Interestingly, truncating mutations in *NOTCH1* indicate a tumor suppressor function as opposed to activating mutations seen in other cancers, and methylation of this gene also indicates a tumor suppressor function. Loss of heterozygosity (LOH) at the *NOTCH1* locus has also been reported for a small number of tumors (Agrawal et al. 2011). The significance of the mechanism of inactivation remains to be clarified, but given the stable, yet potential reversible nature and variable levels of epigenetic modifications, this may have direct implications for treatment and therapy. Longitudinal epigenetic phenotyping of tumor methylation profiles during treatment could provide insight to the degree to which DNA methylation marks are labile to chemotherapy, radiation, or dietary intervention. These results also emphasize the importance of simultaneous evaluation of molecular mechanisms in tumors in conjunction with epidemiologic characteristics, and future studies will benefit from the careful existing comprehensive studies of molecular alterations in HNSCC.

This study has a number of limitations. The population size was relatively small, however, the technology used was able to detect differences in promoter methylation in a large number of genes associated with cancer. The Goldengate

cancer panel, however, does not provide a measurement of promoter methylation in other genes with less well characterized functions, nor does it measure methylation at other genomic features, such as intergenic regions, which could provide information about genomic structure and stability. While the sample was representative of the patients seem in the institutions from which participants were recruited, women and particularly minorities were under-represented. Future planned studies will include a more diverse patient population and a more comprehensive view of the cancer epigenome, integrating epigenetic and transcriptional measures.

2.6 Conclusions

Clinically and pathologically relevant subsets of tumors defined by methylation status have been identified in many cancer types, most notably the CpG Island Methylator Phenotype (CIMP) in colorectal cancer (Toyota et al. 1999). These CIMP tumors exhibit a distinct somatic profile of microsatellite instability and BRAF mutations, with divergent epidemiologic characteristics compared to non-CIMP tumors including a survival advantage (Samowitz et al. 2005a; Samowitz et al. 2005b). Array-based profiling of acute myeloid leukemias using the GoldenGate panel identified clinically relevant subgroups defined by epigenetic modifications, although there was not a strong association between these clusters and survival (Wilop et al. 2011). In this study we investigated the likelihood of identifying a clinically relevant subset of head and neck tumors defined by CpG methylation, taking advantage of a well-established patient cohort at the University of Michigan with well-annotated survival and epidemiologic data. Our sample was representative of the overall cohort regarding age, gender, smoking history, and alcohol consumption. We examined the epigenetic differences between HPV(+) and HPV(-) tumors, following from our recent work in cell lines showing evidence for divergent pathways of carcinogenesis and the well-described epidemiologic differences between individuals with differential HPV tumor status (Sartor et al. 2011). Further, we were able to evaluate survival in this cohort in light of their epigenetic profile (as defined by cluster status), HPV status and other epidemiologic characteristics.

2.7 References

Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, et al. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in notch1. Science 333:1154-1157.

Arthur AE, Duffy SA, Sanchez GI, Gruber SB, Terrell JE, Hebert JR, et al. 2011. Higher micronutrient intake is associated with human papillomavirus-positive head and neck cancer: A case-only analysis. Nutrition and cancer 63:734-742.

Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. 2006. High-throughput DNA methylation profiling using universal bead arrays. Genome Res 16:383-393.

Boyer SN, Wazer DE, Band V. 1996. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. Cancer Research 56:4620-4624.

Califano J, van der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S, et al. 1996. Genetic progression model for head and neck cancer: Implications for field cancerization. Cancer Res 56:2488-2492.

D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, et al. 2007. Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med 356:1944-1956.

Demokan S, Chang X, Chuang A, Mydlarz WK, Kaur J, Huang P, et al. 2010. Kif1a and ednrb are differentially methylated in primary hnscc and salivary rinses. Int J Cancer 127:2351-2359.

Duffy SA, Ronis DL, McLean S, Fowler KE, Gruber SB, Wolf GT, et al. 2009. Pretreatment health behaviors predict survival among patients with head and neck squamous cell carcinoma. Journal of clinical oncology 27:1969-1975.

El-Naggar AK, Lai S, Clayman G, Lee JK, Luna MA, Goepfert H, et al. 1997. Methylation, a major mechanism of p16/cdkn2 gene inactivation in head and neck squamous carcinoma. Am J Pathol 151:1767-1774.

Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. 2008. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 100:261-269.

Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, et al. 2000. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer I 92:709-720.

Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. 2008. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst 100:407-420.

Girard F, Strausfeld U, Fernandez A, Lamb NJ. 1991. Cyclin a is required for the onset of DNA replication in mammalian fibroblasts. Cell 67:1169-1179.

Guerrero-Preston R, Soudry E, Acero J, Orera M, Moreno-Lopez L, Macia-Colon G, et al. 2011. Nid2 and hoxa9 promoter hypermethylation as biomarkers for prevention and early detection in oral cavity squamous cell carcinoma tissues and saliva. Cancer Prev Res (Phila) 4:1061-1072.

Hafkamp HC, Speel EJ, Haesevoets A, Bot FJ, Dinjens WN, Ramaekers FC, et al. 2003. 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. Int J Cancer 107:394-400.

Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. 2011. Increased methylation variation in epigenetic domains across cancer types. Nat Genet 43:768-775.

Holzmann K, Kohlhammer H, Schwaenen C, Wessendorf S, Kestler HA, Schwoerer A, et al. 2004. Genomic DNA-chip hybridization reveals a higher incidence of genomic amplifications in pancreatic cancer than conventional comparative genomic hybridization and leads to the identification of novel candidate genes. Cancer Res 64:4428-4433.

Kuan PF, Wang S, Zhou X, Chu H. 2010. A statistical framework for illumina DNA methylation arrays. Bioinformatics 26:2849-2855.

Kumar B, Cordell KG, Lee JS, Worden FP, Prince ME, Tran HH, et al. 2008a. Egfr, p16, hpv titer, bcl-xl and p53, sex, and smoking as indicators of response to therapy and survival in oropharyngeal cancer. J Clin Oncol 26:3128-3137.

Kumar B, Cordell KG, Lee JS, Worden FP, Prince ME, Tran HH, et al. 2008b. 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 26:3128-3137.

Lee SH, Lee NH, Jin SM, Rho YS, Jo SJ. 2011. Loss of heterozygosity of tumor suppressor genes (p16, rb, e-cadherin, p53) in hypopharynx squamous cell carcinoma. Otolaryngol Head Neck Surg 145:64-70.

Marsit CJ, McClean MD, Furniss CS, Kelsey KT. 2006. Epigenetic inactivation of the sfrp genes is associated with drinking, smoking and hpv in head and neck squamous cell carcinoma. International Journal of Cancer 119:1761-1766.

Marsit CJ, Christensen BC, Houseman EA, Karagas MR, Wrensch MR, Yeh RF, et al. 2009. Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma. Carcinogenesis 30:416-422.

Marur S, Forastiere AA. 2008. Head and neck cancer: Changing epidemiology, diagnosis, and treatment. Mayo Clin Proc 83:489-501.

Maxwell JH, Kumar B, Feng FY, McHugh JB, Cordell KG, Eisbruch A, et al. 2010a. Hpv-positive/p16-positive/ebv-negative nasopharyngeal carcinoma in white north americans. Head Neck 32:562-567.

Maxwell JH, Kumar B, Feng FY, Worden FP, Lee JS, Eisbruch A, et al. 2010b. Tobacco use in human papillomavirus-positive advanced oropharynx cancer patients related to increased risk of distant metastases and tumor recurrence. Clin Cancer Res 16:1226-1235.

Monk D, Wagschal A, Arnaud P, Muller PS, Parker-Katiraee L, Bourc'his D, et al. 2008. Comparative analysis of human chromosome 7q21 and mouse proximal chromosome 6 reveals a placental-specific imprinted gene, tfpi2/tfpi2, which requires ehmt2 and eed for allelic-silencing. Genome Res 18:1270-1281.

Paluszczak J, Misiak P, Wierzbicka M, Wozniak A, Baer-Dubowska W. 2011. Frequent hypermethylation of dapk, rarbeta, mgmt, rassf1a and fhit in laryngeal squamous cell carcinomas and adjacent normal mucosa. Oral Oncol 47:104-107.

Poage GM, Houseman EA, Christensen BC, Butler RA, Avissar-Whiting M, McClean MD, et al. 2011. Global hypomethylation identifies loci targeted for hypermethylation in head and neck cancer. Clin Cancer Res 17:3579-3589.

Ribarska T, Ingenwerth M, Goering W, Engers R, Schulz WA. 2010. Epigenetic inactivation of the placentally imprinted tumor suppressor gene tfpi2 in prostate carcinoma. Cancer Genomics Proteomics 7:51-60.

Richards KL, Zhang B, Baggerly KA, Colella S, Lang JC, Schuller DE, et al. 2009. Genome-wide hypomethylation in head and neck cancer is more pronounced in hpv-negative tumors and is associated with genomic instability. PLoS One 4:e4941.

Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, et al. 2005a. Evaluation of a large, population-based sample supports a cpg island methylator phenotype in colon cancer. Gastroenterology 129:837-845.

Samowitz WS, Sweeney C, Herrick J, Albertsen H, Levin TR, Murtaugh MA, et al. 2005b. Poor survival associated with the braf v600e mutation in microsatellite-stable colon cancers. Cancer research 65:6063.

Sartor MA, Dolinoy DC, Jones TR, Colacino JA, Prince ME, Carey TE, et al. 2011. Genome-wide methylation and expression differences in hpv(+) and hpv(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. Epigenetics 6:777-787.

Smith IM, Mydlarz WK, Mithani SK, Califano JA. 2007. DNA global hypomethylation in squamous cell head and neck cancer associated with smoking, alcohol consumption and stage. Int J Cancer 121:1724-1728.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3:Article3.

Somers KD, Merrick MA, Lopez ME, Incognito LS, Schechter GL, Casey G. 1992. Frequent p53 mutations in head and neck cancer. Cancer Res 52:5997-6000.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100:9440-9445.

Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. 2011. The mutational landscape of head and neck squamous cell carcinoma. Science 333:1157-1160.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545-15550.

Takada H, Imoto I, Tsuda H, Sonoda I, Ichikura T, Mochizuki H, et al. 2005. Screening of DNA copy-number aberrations in gastric cancer cell lines by array-based comparative genomic hybridization. Cancer Sci 96:100-110.

Tang AL, Hauff SJ, Owen JH, Graham MP, Czerwinski MJ, Park JJ, et al. 2012. Um-scc-104: A new human papillomavirus-16-positive cancer stem cell-containing head and neck squamous cell carcinoma cell line. Head Neck 34:1480-1491.

Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. 1999. Cpg island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A 96:8681-8686.

Toyota M, Sasaki Y, Satoh A, Ogi K, Kikuchi T, Suzuki H, et al. 2003. Epigenetic inactivation of chfr in human tumors. Proc Natl Acad Sci U S A 100:7818-7823.

Tsunematsu T, Kudo Y, Iizuka S, Ogawa I, Fujita T, Kurihara H, et al. 2009. Runx3 has an oncogenic role in head and neck cancer. PLoS One 4:e5892.

Uesugi H, Uzawa K, Kawasaki K, Shimada K, Moriya T, Tada A, et al. 2005. Status of reduced expression and hypermethylation of the apc tumor suppressor gene in human oral squamous cell carcinoma. Int J Mol Med 15:597-602.

Viet CT, Schmidt BL. 2008. Methylation array analysis of preoperative and postoperative saliva DNA in oral cancer patients. Cancer Epidemiol Biomarkers Prev 17:3603-3611.

Wang S, Zhu J. 2008. Variable selection for model-based high-dimensional clustering and its application to microarray data. Biometrics 64:440-448.

Weiss D, Basel T, Sachse F, Braeuninger A, Rudack C. 2011. Promoter methylation of cyclin a1 is associated with human papillomavirus 16 induced head and neck squamous cell carcinoma independently of p53 mutation. Mol Carcinog 50:680-688.

Weiss D, Koopmann M, Basel T, Rudack C. 2012. 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 12:259.

Werness BA, Levine AJ, Howley PM. 1990. Association of human papillomavirus types 16 and 18 e6 proteins with p53. Science 248:76-79.

Wilop S, Fernandez AF, Jost E, Herman JG, Brummendorf TH, Esteller M, et al. 2011. Array-based DNA methylation profiling in acute myeloid leukaemia. Br J Haematol 155:65-72.

Worden FP, Kumar B, Lee JS, Wolf GT, Cordell KG, Taylor JM, et al. 2008. Chemoselection as a strategy for organ preservation in advanced oropharynx cancer: Response and survival positively associated with hpv16 copy number. J Clin Oncol 26:3138-3146.

Figure 2.1. DNA methylation heatmap constructed using unsupervised hierarchical Ward clustering of the 711 CpG sites with the greatest variance across the 68 tumor samples identified six distinct methylation clusters.

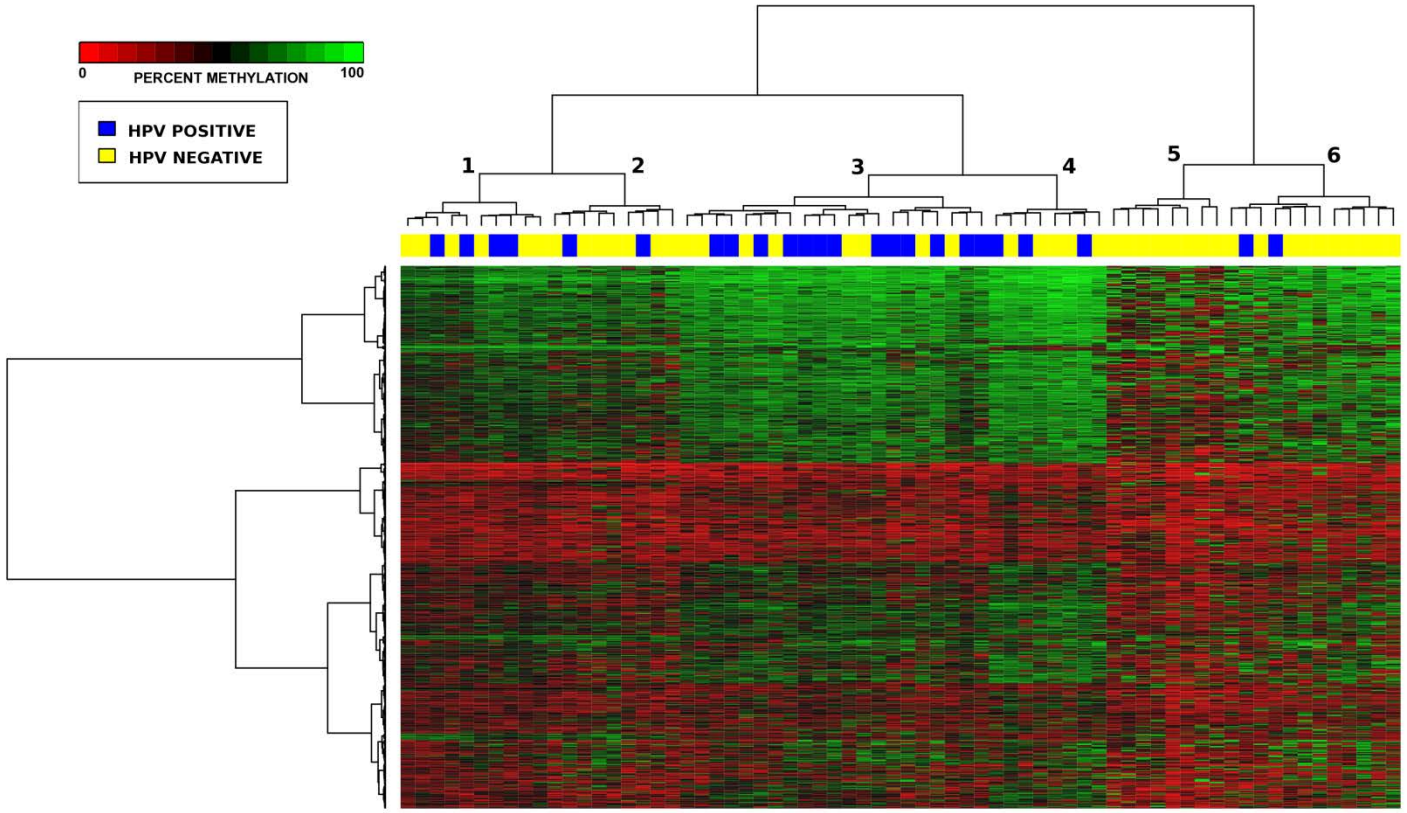

Figure 2.2. Kaplan-Meier survival curves depicting three year survival for each of the six clusters identified via unsupervised hierarchical cluster analysis.

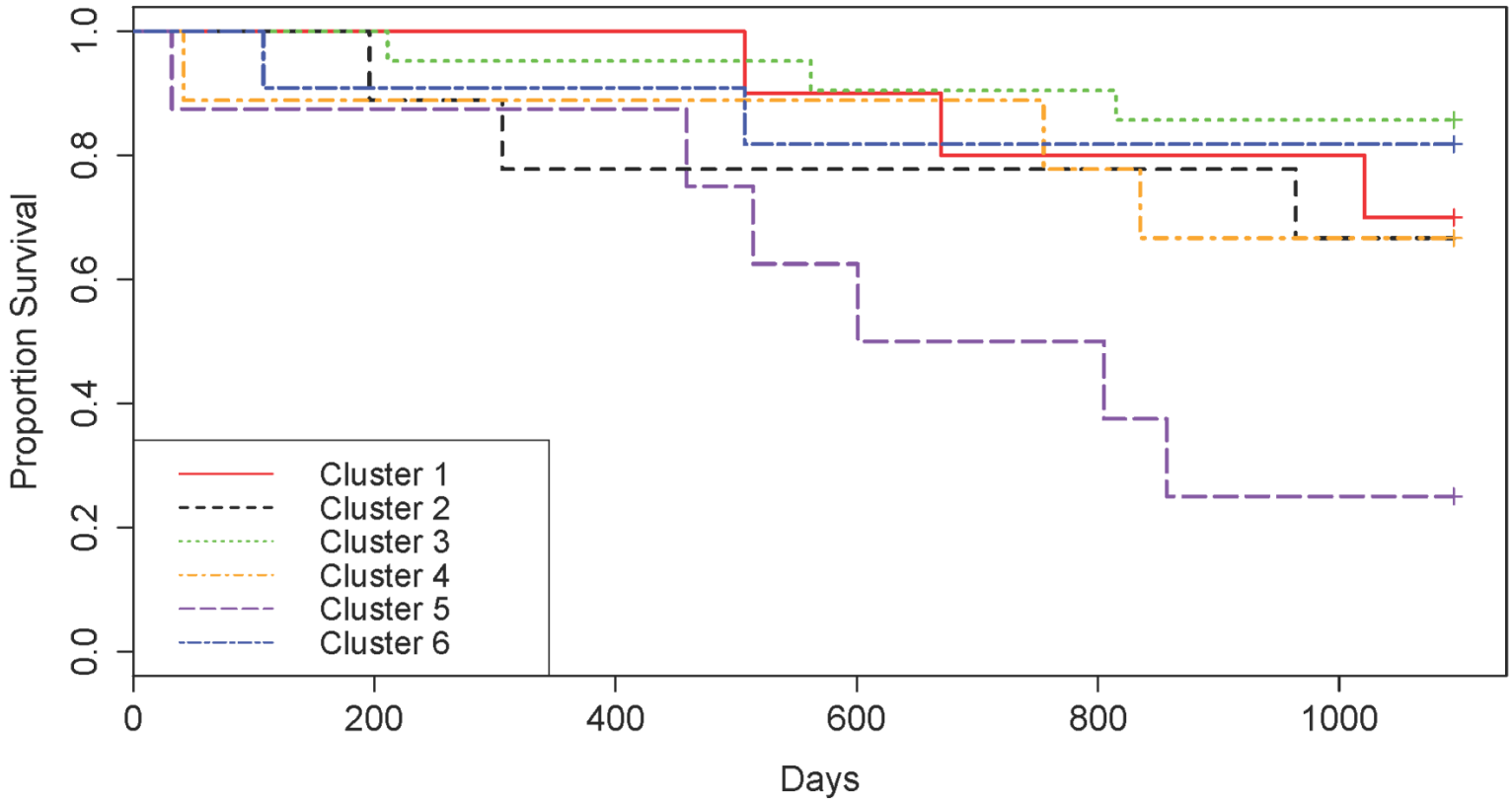

Figure 2.3 Average differences in methylation per CpG site comparing HPV(+) and HPV(-) tumors.

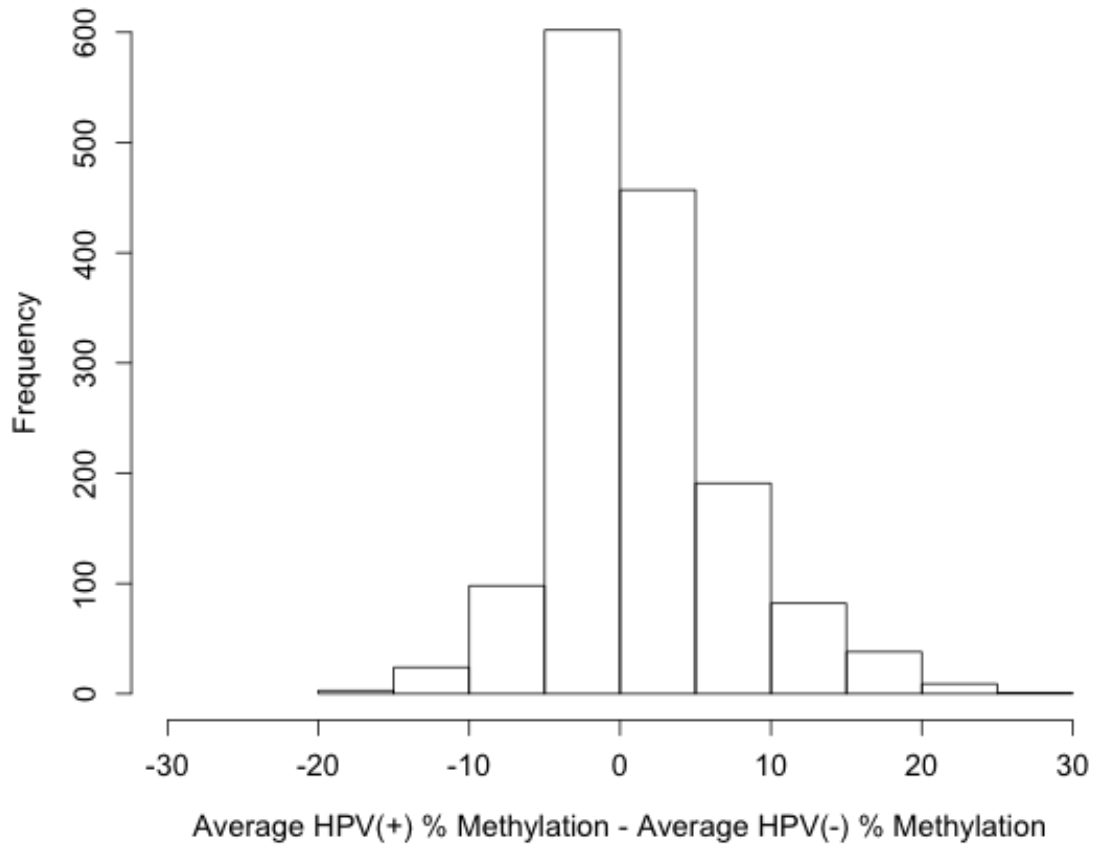

Table 2.1. Clinical characteristics of the study participants (n=68)

Patient Characteristic		N (%)	Mean (SD), Median (range)
Age			57.0(10.0), 55.0 (28-82)
Gender	Male	51 (75%)	
	Female	17 (25%)	
Stage	I and II	11 (16%)	
	III	15 (22%)	
	IV	42 (62%)	
Cancer Site of first Primary	Oral Cavity	17 (25%)	
	Oropharynx	32 (47%)	
	Hypopharynx	4 (6%)	
	Larynx	13 (19%)	
	Other	2 (3%)	
Tumor Tissue HPV (+) Status		24 (35%)	
Smoking Status	Never	11 (16%)	
	Past	41 (60%)	
	Current	16 (24%)	
Pack-years			33.3(37), 25 (0-220)
Non-cigarette Tobacco	(yes/no) ever	12 (18%)	
Alcohol Problem	AUDIT >= 8 and drank within 1 year.	23 (34%)	

Table 2.2. Clinical characteristics of the six clusters identified via unsupervised hierarchical cluster analysis of DNA methylation values.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
N	10	9	21	9	8	11
Age*						
mean(sd)	50.7 (9.1)	55.4 (8.2)	61.6 (9.7)	51.8 (6.4)	57.9 (9.6)	58.9 (11.9)
median, (min-max)	51.5, (28-61)	54, (42-68)	62 (41-82)	53 (43-64)	57.5 (46-72)	64 (41-73)
Male n (%)	7 (70%)	9 (100%)	16 (76%)	6 (67%)	5 (63%)	8 (73%)
Cancer Site n (%)						
OC	2 (20%)	1 (11%)	1 (5%)	3 (33%)	5 (63%)	5 (45%)
OP	6 (60%)	4 (44%)	15 (71%)	4 (44%)	0	3 (27%)
HP	1 (10%)	1 (11%)	1 (5%)	0	0	1 (9%)
LA	1 (10%)	2 (22%)	4 (19%)	2 (22%)	3 (37%)	1 (9%)
OT	0	1 (11%)	0	0	0	1 (9%)
Primary Cancer Stage n (%)						
I and IIIII			3 (14%)			4 (36%)
IV	0	0	2 (10%)	1 (11%)	3 (38%)	2 (18%)
	4 (40%)	3 (33%)	16 (76%)	2 (22%)	2 (25%)	5 (45%)
	6 (60%)	6 (67%)		6 (67%)	3 (38%)	
HPV status n (%)*						
Pos	4 (40%)	1 (11%)	13 (62%)	3 (33%)	0	2 (18%)
Neg	6 (60%)	8 (89%)	8 (38%)	6 (67%)	8 (100%)	9 (82%)
Smoking Status n (%)						
Currently smoke cigarettes	1 (10%)	3 (33%)	5 (24%)	4 (44%)	0	3 (27%)
Past smoker, quit within last year	5 (50%)	4 (44%)	3 (14%)	3 (33%)	5 (63%)	4 (36%)
Past smoker, quit over a year ago	3 (30%)	1 (11%)	7 (33%)	0	3 (38%)	3 (27%)
Never smoked cigarettes	1 (10%)	1 (11%)	6 (29%)	2 (22%)	0	1 (9%)
Problem Drinking n (%)[a]	4 (40%)	5 (56%)	5 (24%)	3 (33%)	3 (38%)	3 (27%)
3 year Overall Survival*	7 (70%)	6 (66%)	18 (86%)	3 (66%)	2 (25%)	9 (82%)
Treatment						

Surgery Only	1 (10%)	0	4 (19%)	1 (11%)	2 (25%)	0
Radiation Only	0	1 (11%)	1 (5%)	0	0	1 (9%)
Surgery and Radiation	0	2 (22%)	3 (14%)	2 (22%)	0	2 (18%)
Radiation and Chemotherapy	4 (40%)	4 (44%)	8 (38%)	4 (44%)	4 (50%)	7 (64%)
Surgery, Radiation and Chemotherapy	5 (50%)	2 (22%)	3 (14%)	2 (22%)	0	2 (18%)

*p<0.05 for difference between clusters

a) Problem drinking defined: AUDIT>8 and drank in past 1 year. Note: n=14 missing AUDIT score

Table 2.3. CpG sites with DNA methylation values significantly associated (Adjusted p<0.05) with HPV status of the tumor. Positive T-Values correspond with higher methylation in HPV(+) while negative T-Values correspond with higher methylation in HPV(-) tumors.

Gene Symbol	Chromosome	CpG Coordinate	Distance to TSS	DNA Strand of Transcription	T-Value	Mean % Difference in Methylation	P-Value	Adjusted P-Value
CCNA1	13	35904640	7	+	5.30	21.3	1.86E-06	0.0028
GRB7	17	35147553	-160	-	4.58	8.0	2.46E-05	0.0161
SPDEF	6	34631953	116	-	-4.51	-3.7	3.20E-05	0.0161
CDH11	16	63713774	-354	-	4.32	18.4	6.08E-05	0.0192
RUNX1T1	8	93176474	145	-	4.31	13.7	6.37E-05	0.0192
RASSF1	3	50353615	-244	+	-4.22	-2.1	8.47E-05	0.0213
STAT5A	17	37693133	42	+	-4.05	-11.5	1.51E-04	0.0318
MGMT	10	131155184	-272	-	-4.01	-3.6	1.73E-04	0.0318
ESR2	14	63830765	66	+	-3.98	-6.4	1.90E-04	0.0318
JAK3	19	17819736	64	+	-3.92	-11.8	2.31E-04	0.0348
SYBL1	X	154763858	-349	+	3.88	12.2	2.71E-04	0.0370
HSD17B12	11	43659026	145	-	-3.83	-0.9	3.14E-04	0.0394
TUSC3	8	15442130	29	-	3.73	6.7	4.28E-04	0.0496

Table 2.4. All CpG sites with DNA methylation values significantly associated ($p<0.05$) with HPV status of the tumor. Positive T-values correspond with sites more highly methylated in HPV(+) while negative T-values correspond with sites more highly methylated in HPV(-) tumors. Adjusted p-values were calculated via the Benjamini-Hochberg Method.

Gene Symbol	Chromosome	CpG Coordinate	Distance to TSS	DNA Strand of Transcription	T-Value	P-Value	Adjusted P-Value
CCNA1	13	35904640	7	+	5.30	1.86E-06	0.0028
GRB7	17	35147553	-160	-	4.58	2.46E-05	0.0161
SPDEF	6	34631953	116	-	-4.51	3.20E-05	0.0161
CDH11	16	63713774	-354	-	4.32	6.08E-05	0.0192
RUNX1T1	8	93176474	145	-	4.31	6.37E-05	0.0192
RASSF1	3	50353615	-244	+	-4.22	8.47E-05	0.0213
STAT5A	17	37693133	42	+	-4.05	1.51E-04	0.0318
MGMT	10	131155184	-272	-	-4.01	1.73E-04	0.0318
ESR2	14	63830765	66	+	-3.98	1.90E-04	0.0318
JAK3	19	17819736	64	+	-3.92	2.31E-04	0.0348
SYBL1	X	154763858	-349	+	3.88	2.71E-04	0.0370
HSD17B12	11	43659026	145	-	-3.83	3.14E-04	0.0394
TUSC3	8	15442130	29	-	3.73	4.28E-04	0.0496
TNFRSF10C	8	23016372	-7	+	-3.66	5.38E-04	0.0578
OSM	22	28993028	-188	+	-3.61	6.39E-04	0.0629
CD1A	1	156490545	-6	+	3.59	6.69E-04	0.0629
PTPRG	3	61522325	40	-	-3.53	8.07E-04	0.0677
CDK10	16	88280653	74	+	-3.52	8.34E-04	0.0677
NEFL	8	24870155	-209	-	3.52	8.55E-04	0.0677
MPL	1	43575405	-657	+	-3.44	0.0011	0.0806
PLAT	8	42184431	-80	+	3.37	0.0013	0.0924
TNFRSF10D	8	23077555	-70	+	-3.37	0.0014	0.0924
SEMA3F	3	50168185	333	-	3.32	0.0016	0.0993
IGF2AS	11	2118120	-203	+	3.31	0.0016	0.0993

SMARCA4	19	10932244	-362	-	-3.28	0.0018	0.1064
CHFR	12	131974892	-635	-	-3.22	0.0021	0.1205
ZNF264	19	62394284	-397	+	3.21	0.0022	0.1206
IRF7	11	605691	236	-	-3.17	0.0024	0.1271
MYLK	3	125085707	132	-	3.17	0.0024	0.1271
MEG3	14	100362306	91	+	3.12	0.0028	0.1381
THBS2	6	169395933	129	+	-3.12	0.0028	0.1381
BCL2L2	14	22845586	-280	+	3.10	0.0030	0.1388
OSM	22	28992874	-34	+	-3.06	0.0034	0.1491
RASSF1	3	50353255	116	+	-3.05	0.0034	0.1491
SEMA3B	3	50280140	96	+	-3.03	0.0036	0.1491
DES	2	219991571	228	-	-3.02	0.0037	0.1491
GRPR	X	16051145	-200	-	3.01	0.0038	0.1491
SHB	9	38059683	-473	-	-3.01	0.0038	0.1491
SPI1	11	47356469	205	+	-3.01	0.0039	0.1491
IRF5	7	128365107	-123	+	-2.97	0.0043	0.1609
CREBBP	16	3871424	-712	-	2.96	0.0044	0.1614
HPSE	4	84475422	-93	+	-2.95	0.0045	0.1624
ICAM1	19	10242393	-386	-	-2.94	0.0046	0.1624
EYA4	6	133603698	-508	+	-2.86	0.0059	0.1984
ADAMTS12	5	33927829	52	-	-2.85	0.0060	0.1984
PTPRO	12	15366383	-371	+	-2.84	0.0061	0.1984
S100A2	1	151804894	36	-	2.84	0.0063	0.1984
CASP8	2	201806900	474	+	-2.82	0.0065	0.1984
CTGF	6	132314848	-693	-	2.82	0.0066	0.1984
E2F5	8	86276358	-516	-	-2.82	0.0066	0.1984
CTSL	9	89530719	-81	+	2.80	0.0069	0.2002
NDN	15	21483412	131	-	2.80	0.0069	0.2002
SHH	7	155297400	328	+	2.79	0.0071	0.2021
RIPK4	21	42060152	166	+	2.78	0.0073	0.2042

DSC2	18	26936285	90	+	2.76	0.0077	0.2073
FRK	6	116488650	-36	+	2.76	0.0077	0.2073
CCNA1	13	35904417	-216	+	2.75	0.0080	0.2107
CSF3R	1	36721104	-8	+	2.72	0.0086	0.2141
CHFR	12	131974758	-501	+	-2.71	0.0087	0.2141
EPHA7	6	94186198	-205	-	-2.71	0.0087	0.2141
RUNX3	1	25164455	-393	-	-2.71	0.0088	0.2141
AXL	19	46416440	-223	-	2.71	0.0089	0.2141
HLA-DOB	6	32893119	-357	-	-2.70	0.0091	0.2141
FGF1	5	142046169	-357	-	2.70	0.0092	0.2141
NOTCH4	6	32299818	4	+	2.69	0.0093	0.2141
RAB32	6	146906835	314	-	2.69	0.0094	0.2141
SMO	7	128616006	57	+	-2.68	0.0096	0.2151
CDKN2A	9	21964709	229	-	-2.67	0.0098	0.2164
MC2R	18	13906560	-1025	+	2.66	0.0101	0.2193
MMP9	20	44070765	-189	+	2.62	0.0111	0.2354
GABRA5	15	24742740	44	-	2.62	0.0112	0.2354
SPI1	11	47357603	-929	+	2.62	0.0113	0.2354
HSPA2	14	64072214	-162	-	-2.60	0.0117	0.2405
SOX2	3	182911870	-546	+	-2.59	0.0121	0.2405
GADD45A	1	67922734	-737	-	2.58	0.0124	0.2405
TNK1	17	7225093	-41	-	2.58	0.0124	0.2405
JAK3	19	17819956	-156	-	-2.58	0.0124	0.2405
CDH11	16	63713623	-203	-	2.58	0.0125	0.2405
FASTK	7	150409141	-257	+	2.57	0.0128	0.2405
MMP14	14	22375425	-208	-	2.57	0.0128	0.2405
EPHB4	7	100263392	-313	-	-2.55	0.0134	0.2487
RHOH	4	39874876	-121	+	-2.54	0.0139	0.2556
COL18A1	21	45649031	-494	-	-2.51	0.0147	0.2619
IGFBP5	2	217268372	144	+	-2.51	0.0149	0.2619

FGF7	15	47502707	-44	+	2.51	0.0150	0.2619
PRSS1	7	142136949	45	-	2.51	0.0150	0.2619
CSF2	5	131437632	248	-	2.50	0.0151	0.2619
SMO	7	128615494	-455	-	2.50	0.0153	0.2619
HDAC6	X	48545533	102	+	2.49	0.0155	0.2629
SERPINE1	7	100556653	-519	+	2.49	0.0158	0.2641
TGFA	2	70634996	-558	+	-2.48	0.0161	0.2661
HRASLS	3	194441259	-353	-	2.47	0.0164	0.2669
TRIM29	11	119514334	-261	+	2.47	0.0165	0.2669
SLIT2	4	19864125	-208	+	2.46	0.0167	0.2669
CD40	20	44179941	-372	-	-2.45	0.0172	0.2700
GNMT	6	43036604	126	+	-2.45	0.0172	0.2700
HOXA11	7	27191320	35	+	-2.44	0.0178	0.2726
PLA2G2A	1	20179228	268	+	-2.43	0.0182	0.2726
PTPNS1	20	1823858	433	-	-2.43	0.0182	0.2726
NOTCH1	9	138559607	452	-	-2.43	0.0182	0.2726
YES1	18	802927	-600	+	-2.43	0.0183	0.2726
JAK3	19	17820875	-1075	-	2.42	0.0185	0.2727
FLT3	13	27573007	-302	+	-2.42	0.0188	0.2745
MET	7	116100028	333	+	2.41	0.0191	0.2757
CD9	12	6179231	-585	-	2.40	0.0194	0.2785
GDF10	10	48059133	39	+	-2.39	0.0200	0.2827
MATK	19	3752874	-64	+	-2.39	0.0201	0.2827
FGFR2	10	123348367	-460	-	2.38	0.0207	0.2886
KCNK4	11	63815280	-171	-	-2.37	0.0209	0.2886
RHOH	4	39874044	-953	-	2.35	0.0222	0.3036
INS	11	2139248	-248	+	2.34	0.0229	0.3089
PLS3	X	114701835	70	+	2.33	0.0230	0.3089
APC	5	112101600	117	-	-2.33	0.0232	0.3095
APC	5	112101203	-280	-	-2.32	0.0237	0.3124

ZNF264	19	62394729	48	-	-2.32	0.0240	0.3144
GUCY2D	17	7846665	-48	-	-2.31	0.0243	0.3151
CD40	20	44180371	58	-	-2.31	0.0247	0.3170
EPHA2	1	16355354	-203	+	2.30	0.0251	0.3170
ETV1	7	13995804	-515	+	-2.30	0.0251	0.3170
MAPK10	4	87593281	26	+	2.30	0.0253	0.3170
ACVR1	2	158404019	-983	+	2.29	0.0259	0.3221
AATK	17	76710603	-709	-	2.27	0.0269	0.3318
IMPACT	18	20260446	-234	-	2.26	0.0273	0.3319
LOX	5	121442166	-313	-	2.26	0.0274	0.3319
MGMT	10	131155175	-281	+	-2.26	0.0276	0.3319
BTK	X	100527943	-105	+	-2.25	0.0281	0.3361
NTSR1	20	60810316	-318	+	2.25	0.0284	0.3361
ASCL2	11	2249367	-609	-	2.24	0.0291	0.3424
MKRN4	X	40578589	249	-	2.22	0.0306	0.3453
ICAM1	19	10243021	242	+	-2.21	0.0309	0.3453
IGSF4C	19	48836364	-533	-	-2.21	0.0311	0.3453
CD9	12	6179312	-504	+	2.21	0.0312	0.3453
TIE1	1	43539317	66	-	2.21	0.0312	0.3453
RUNX3	1	25164035	27	-	-2.20	0.0315	0.3453
SPDEF	6	34632075	-6	-	2.20	0.0315	0.3453
ETV6	12	11694485	430	+	-2.20	0.0319	0.3453
CD86	3	123256908	-3	+	2.20	0.0319	0.3453
PAX6	11	31790576	-1121	+	-2.20	0.0319	0.3453
INHA	2	220144009	1252	+	-2.20	0.0321	0.3453
EPO	7	100156603	244	-	-2.19	0.0323	0.3453
DBC1	9	121171318	204	+	2.19	0.0325	0.3453
FLT3	13	27572379	326	-	-2.19	0.0326	0.3453
KRT5	12	51200818	-308	+	2.19	0.0329	0.3458
IL4	5	132037010	-262	-	2.17	0.0344	0.3592

SLIT2	4	19864444	111	-	-2.16	0.0353	0.3660
MMP10	11	102156418	136	-	2.15	0.0357	0.3680
CDKN1C	11	2864177	-626	+	-2.14	0.0362	0.3706
UBA52	19	18543375	-293	-	-2.14	0.0366	0.3706
MMP14	14	22375620	-13	+	2.14	0.0367	0.3706
EMR3	19	14646849	-39	-	2.13	0.0375	0.3739
GSTM1	1	110031699	-266	+	2.13	0.0375	0.3739
DSC2	18	26936782	-407	-	2.12	0.0378	0.3739
ARHGAP9	12	56169124	-260	+	2.12	0.0383	0.3739
COL1A2	7	93861402	-407	-	2.12	0.0383	0.3739
APOC2	19	50140706	-377	+	2.11	0.0389	0.3746
STK11	19	1156503	-295	-	-2.11	0.0392	0.3746
KIAA1804	1	231529448	-689	-	2.11	0.0392	0.3746
F2R	5	76047454	-88	+	2.11	0.0393	0.3746
IGSF4	11	114880779	-454	+	2.10	0.0398	0.3763
CDH11	16	63713318	102	-	2.10	0.0400	0.3763
RARB	3	25444698	-60	+	2.09	0.0410	0.3790
HIC2	22	20101165	-528	-	2.09	0.0411	0.3790
SEMA3A	7	83662191	-343	+	-2.09	0.0414	0.3790
ITK	5	156540651	166	-	-2.08	0.0416	0.3790
NRAS	1	115061141	-103	-	2.08	0.0421	0.3790
PTGS2	1	184916703	-524	-	2.07	0.0425	0.3790
RIPK2	8	90839296	123	+	-2.07	0.0425	0.3790
RUNX3	1	25164309	-247	+	-2.07	0.0426	0.3790
COL1A1	17	45633997	-5	+	2.07	0.0432	0.3790
TCF4	18	51406615	-175	-	-2.07	0.0432	0.3790
SMAD4	18	46810137	-474	-	-2.07	0.0432	0.3790
THY1	11	118799110	-20	-	-2.06	0.0434	0.3790
COL18A1	21	45649160	-365	-	2.06	0.0436	0.3790
FES	15	89228490	-223	-	-2.06	0.0439	0.3799

GNAS	20	56848248	58	+	2.05	0.0451	0.3873
PTHLH	12	28016940	-757	+	2.05	0.0453	0.3873
PITX2	4	111777933	24	-	2.04	0.0457	0.3887
MCM6	2	136350345	136	+	-2.04	0.0461	0.3896
EPHA8	1	22762135	-456	-	-2.03	0.0470	0.3939
MYCN	2	15998211	77	-	-2.03	0.0471	0.3939
MYOD1	11	17697891	156	+	2.02	0.0476	0.3960
RYK	3	135452769	-493	+	-2.02	0.0481	0.3974
RRAS	19	54835312	-100	-	-2.01	0.0487	0.4004
SEPT9	17	72827370	-374	+	2.01	0.0490	0.4005
MME	3	156280182	29	+	2.00	0.0496	0.4005
GPR116	6	46991675	-850	+	2.00	0.0497	0.4005
DSG1	18	27151891	-159	-	2.00	0.0498	0.4005

Table 2.5. Candidate enriched gene sets for differentially methylated genes associated with HPV status.

Name	Size	Enrichment Score (ES)	Normalized Enrichment Score (NES)	Nominal P-Value	FDR Q-Value
GENE ONTOLOGY - BIOLOGICAL PROCESSES					
Regulation of Cell Cycle	11	-0.54	-1.96	0.007	0.41
Cell Cycle (GO 0007049)	14	-0.43	-1.69	0.036	1.00
Negative Regulation of Cellular Metabolic Process	6	-0.58	-1.63	0.041	1.00
KEGG PATHWAYS					
Neuroactive Ligand Receptor Interaction (HSA04080)	6	0.60	1.72	0.020	0.27
JAK-STAT Signaling Pathway (HSA04630)	9	-0.49	-1.62	0.047	0.40

Table 2.6. CpG sites identified as significantly associated (p<0.05) with three year survival by Cox Proportional Hazards Modeling.

Gene Symbol	Chromosome	CpG Coordinate	Distance to TSS	DNA Strand of Transcription	Hazard Coefficient	Standard Error	P-Value	Q-Value
NOTCH1	9	138559607	452	-	22.29	5.95	0.0002	0.172
TEK	9	27098915	-526	+	-5.66	1.63	0.0005	0.172
UGT1A1	2	234333094	-564	-	-6.63	1.92	0.0005	0.172
IL6	7	22732734	-611	+	4.86	1.45	0.0008	0.172
AHR	7	17304605	-166	-	12.36	3.82	0.0012	0.172
RRAS	19	54835312	-100	-	11.86	3.69	0.0013	0.172
NEO1	15	71130861	-1067	+	22.72	7.17	0.0015	0.172
F2R	5	76046703	-839	+	-4.45	1.41	0.0016	0.172
IGFBP3	7	45928431	-1035	+	11.24	3.58	0.0017	0.172
PLS3	X	114701671	-94	-	8.04	2.58	0.0019	0.172
HS3ST2	16	22732815	-546	+	-6.21	2.00	0.0019	0.172
MST1R	3	49916466	-392	+	13.09	4.22	0.0019	0.172
MYOD1	11	17697891	156	+	-4.41	1.44	0.0021	0.175
FER	5	108110841	-581	+	-4.86	1.64	0.0030	0.215
PLAT	8	42184431	-80	+	12.04	4.06	0.0030	0.215
VAMP8	2	85657987	-241	+	6.68	2.29	0.0035	0.217
ARHGAP9	12	56169124	-260	+	-3.76	1.31	0.0039	0.217
TMEFF1	9	102275718	180	-	7.56	2.63	0.0041	0.217
CYP1B1	2	38157008	-212	+	4.89	1.71	0.0042	0.217
JAG2	14	104706470	-264	+	15.37	5.38	0.0043	0.217
TMEFF2	2	192767395	494	-	-3.70	1.30	0.0043	0.217
ROR1	1	64012296	-6	+	4.60	1.64	0.0049	0.240
RUNX3	1	25164035	27	-	-3.52	1.26	0.0053	0.246
EVI2A	17	26672423	420	+	-3.78	1.37	0.0058	0.246
S100A4	1	151785100	-194	-	-3.35	1.22	0.0059	0.246

BTK	X	100527943	-105	+	-6.03	2.19	0.0060	0.246
CDC25B	20	3724375	-11	-	28.47	10.46	0.0065	0.258
DSP	6	7486833	-36	+	9.05	3.35	0.0069	0.262
CFTR	7	116906881	-372	-	-4.54	1.70	0.0075	0.263
NTSR1	20	60810743	109	+	6.76	2.54	0.0077	0.263
MC2R	18	13906560	-1025	+	3.51	1.32	0.0080	0.263
JAK2	9	4974473	-772	-	8.76	3.30	0.0080	0.263
HTR1B	6	78229607	232	-	-3.58	1.35	0.0081	0.263
RIPK4	21	42060490	-172	+	4.01	1.53	0.0087	0.275
IRAK1	X	152938848	-312	+	5.69	2.18	0.0091	0.278
JAK3	19	17820875	-1075	-	-3.22	1.24	0.0093	0.278
BCL2A1	15	78051825	-1127	-	-3.05	1.18	0.0098	0.279
CPA4	7	129719269	-961	-	-2.81	1.10	0.0105	0.279
FANCA	16	88411572	-1006	-	-2.97	1.16	0.0105	0.279
LAMC1	1	181259642	466	-	35.82	14.07	0.0109	0.279
PTK6	20	61639101	50	+	3.65	1.44	0.0111	0.279
YES1	18	802927	-600	+	22.56	8.92	0.0114	0.279
VAV2	9	135847168	58	+	31.29	12.40	0.0116	0.279
JUNB	19	12762161	-1149	-	7.59	3.01	0.0118	0.279
PCDH1	5	141238392	-264	+	7.53	3.01	0.0124	0.279
SEMA3F	3	50168185	333	-	12.85	5.14	0.0125	0.279
ZIM3	19	62348179	203	+	-4.55	1.82	0.0125	0.279
SNURF	15	22751226	-2	-	-3.91	1.57	0.0127	0.279
FES	15	89228490	-223	-	-5.26	2.11	0.0129	0.279
LYN	8	56955279	353	+	5.58	2.25	0.0131	0.279
NPR2	9	35781788	-618	+	3.91	1.58	0.0133	0.279
MLH1	3	37009602	197	+	20.70	8.43	0.0140	0.284
ICAM1	19	10242393	-386	-	6.95	2.84	0.0142	0.284
ALPL	1	21708178	-278	+	15.42	6.30	0.0143	0.284
ITGA2	5	52321134	120	+	16.56	6.81	0.0151	0.284

RBP1	3	140741330	-150	+	10.78	4.45	0.0153	0.284
EPHA8	1	22762135	-456	-	-4.16	1.71	0.0153	0.284
ATP10A	15	23660487	-524	-	-3.81	1.57	0.0155	0.284
TDGF1	3	46593789	-428	-	-3.13	1.30	0.0161	0.284
MYH11	16	15858391	-22	+	-4.50	1.87	0.0162	0.284
MFAP4	17	19231283	-197	+	-3.45	1.44	0.0163	0.284
MMP19	12	54522728	274	-	-2.74	1.14	0.0167	0.284
MAP3K1	5	56146015	-7	+	12.84	5.37	0.0168	0.284
PRKCDBP	11	6298110	206	+	5.20	2.18	0.0170	0.284
MFAP4	17	19231096	-10	-	-3.96	1.67	0.0177	0.289
SH3BP2	4	2789568	-771	-	10.49	4.44	0.0182	0.289
LRP2	2	169927239	20	+	2.32	0.98	0.0184	0.289
PHLDA2	11	2907848	-622	+	6.22	2.64	0.0184	0.289
SEMA3A	7	83662506	-658	-	-8.23	3.50	0.0187	0.289
PLS3	X	114701835	70	+	2.54	1.08	0.0189	0.289
DDR2	1	160869183	331	+	-2.97	1.27	0.0192	0.289
SMAD2	18	43712069	-848	-	12.81	5.50	0.0199	0.289
PGF	14	74492011	33	+	8.42	3.62	0.0201	0.289
INHA	2	220144009	1252	+	9.52	4.10	0.0203	0.289
BCL3	19	49943942	71	+	4.88	2.11	0.0205	0.289
CTLA4	2	204440932	176	-	-2.54	1.10	0.0205	0.289
IL3	5	131423690	-556	+	-2.18	0.94	0.0208	0.290
GAS7	17	10042445	148	+	-2.54	1.10	0.0213	0.293
MYCN	2	15997670	-464	-	16.94	7.47	0.0233	0.316
EPHX1	1	224079751	152	+	-2.69	1.19	0.0236	0.316
SOD3	4	24404928	-225	+	-2.31	1.03	0.0244	0.321
SLC5A8	12	100128060	60	-	-3.05	1.36	0.0248	0.321
EGR4	2	73374048	70	+	3.77	1.68	0.0251	0.321
LMO1	11	8241717	265	-	-23.80	10.66	0.0255	0.321
BLK	8	11388916	-14	+	-3.08	1.38	0.0258	0.321

MUSK	9	112470652	-308	+	-2.44	1.10	0.0261	0.321
EYA4	6	133603412	-794	+	-2.50	1.13	0.0266	0.321
ZNF264	19	62394729	48	-	7.10	3.20	0.0267	0.321
PPARG	3	12304537	178	-	4.19	1.89	0.0268	0.321
EDN1	6	12398606	-39	-	20.69	9.35	0.0269	0.321
BMPR2	2	202950351	435	+	41.33	18.84	0.0283	0.330
MYLK	3	125086308	-469	-	-5.43	2.48	0.0286	0.330
GLI3	7	42241564	148	-	-3.45	1.58	0.0286	0.330
PWCR1	15	22846906	-811	+	-2.38	1.09	0.0294	0.335
MAF	16	78192938	-826	-	15.89	7.31	0.0297	0.335
MYBL2	20	41728769	-354	+	9.32	4.31	0.0307	0.343
EMR3	19	14646849	-39	-	5.10	2.37	0.0310	0.343
HLA-DPA1	6	33149321	35	-	-2.52	1.18	0.0323	0.353
TNK1	17	7224913	-221	+	-2.98	1.40	0.0327	0.354
IL8	4	74825056	-83	+	-4.95	2.32	0.0331	0.354
XRCC1	19	48772236	-681	-	-2.12	1.00	0.0342	0.362
KIAA1804	1	231529448	-689	-	3.08	1.46	0.0345	0.362
PROK2	3	71916902	0	+	2.16	1.02	0.0349	0.363
GPX1	3	49370989	-194	+	9.52	4.53	0.0356	0.364
EVI1	3	170346817	-30	-	-19.54	9.32	0.0361	0.364
KRAS	12	25295039	82	+	-3.42	1.63	0.0361	0.364
ACVR1	2	158402708	328	-	-2.23	1.07	0.0363	0.364
PLG	6	161043679	406	+	-2.26	1.08	0.0370	0.366
C4B	6	32057622	-191	+	-2.72	1.30	0.0373	0.366
ITGB4	17	71228594	-517	+	-6.47	3.11	0.0375	0.366
LMO1	11	8242151	-169	+	17.98	8.67	0.0380	0.367
S100A4	1	151785793	-887	-	-2.55	1.24	0.0389	0.372
AXL	19	46416440	-223	-	-4.67	2.27	0.0393	0.372
TDGF1	3	46594270	53	-	-1.97	0.96	0.0410	0.376
SEPT9	17	72827686	-58	-	-1.77	0.87	0.0413	0.376

RASA1	5	86600014	107	+	13.88	6.80	0.0413	0.376
SOX1	13	111768896	-1018	-	-6.54	3.21	0.0414	0.376
GSTP1	11	67107788	-74	+	-25.67	12.59	0.0415	0.376
LTA	6	31647858	-214	-	-2.09	1.03	0.0419	0.376
BDNF	11	27700131	-259	-	-2.10	1.03	0.0423	0.376
SMAD2	18	43711929	-708	-	-76.61	37.81	0.0427	0.376
CD34	1	206152086	-780	-	-2.56	1.27	0.0435	0.376
RYK	3	135452769	-493	+	6.38	3.16	0.0435	0.376
FVT1	18	59185663	-225	+	2.17	1.07	0.0436	0.376
CD81	11	2354912	-211	+	-2.66	1.32	0.0444	0.380
RET	10	42892544	11	+	-3.52	1.76	0.0451	0.380
TSC2	16	2038740	140	+	-8.57	4.28	0.0452	0.380
SLC6A8	X	152606393	-193	-	2.76	1.38	0.0460	0.380
MME	3	156279765	-388	+	-3.47	1.74	0.0462	0.380
HHIP	4	145786045	-578	-	-5.15	2.58	0.0463	0.380
CCKAR	4	26101061	79	+	-1.94	0.98	0.0467	0.380
CDK6	7	92300892	256	+	23.63	11.90	0.0471	0.380
HOXA11	7	27191447	-92	-	3.00	1.52	0.0479	0.380
TMEFF1	9	102274912	-626	-	-4.52	2.29	0.0483	0.380
TGFBR3	1	92124672	-429	+	7.68	3.89	0.0486	0.380

Table 2.7. CpG sites identified as significantly differentially methylated between cluster 5 and all other clusters.

Gene Symbol	Chromosome	CpG Coordinate	Distance to TSS	DNA Strand of Transcription	T-Value	P-Value	Adjusted P-Value
FGR	1	27823199	-39	+	-8.71	3.7E-12	5.6E-09
ACTG2	2	73973255	-346	+	-8.20	2.7E-11	2.0E-08
MLF1	3	159771921	243	+	-7.09	2.0E-09	9.8E-07
GFAP	17	40349608	-1214	+	-7.00	2.8E-09	1.1E-06
TNFRSF10C	8	23015767	-612	-	-6.67	1.0E-08	3.0E-06
MMP1	11	102174501	-397	-	-6.54	1.7E-08	4.3E-06
PTHLH	12	28016940	-757	+	-6.38	3.1E-08	6.6E-06
ARHGAP9	12	56169382	-518	-	-6.34	3.6E-08	6.8E-06
MSH3	5	79986037	-13	-	-6.16	7.2E-08	1.2E-05
DNAJC15	13	42495297	-65	+	-5.87	2.2E-07	3.3E-05
NKX3-1	8	23597266	-871	-	-5.78	3.1E-07	4.3E-05
SRC	20	35406205	-297	+	-5.58	6.4E-07	8.0E-05
ZNFN1A1	7	50411546	-179	+	-5.42	1.2E-06	1.4E-04
CSF1R	5	149473102	26	+	-5.33	1.6E-06	1.8E-04
P2RX7	12	120054464	-597	+	-5.19	2.8E-06	2.8E-04
LCN2	9	129951398	-141	-	-5.12	3.6E-06	3.4E-04
CDH17	8	95289955	31	+	-5.01	5.3E-06	4.6E-04
IFNG	12	66840247	-459	-	-5.00	5.5E-06	4.6E-04
NAT2	8	18293024	-11	+	-4.90	8.1E-06	6.4E-04
PLAGL1	6	144371482	-236	-	-4.73	1.5E-05	1.1E-03
THPO	3	185578143	483	+	-4.60	2.3E-05	1.6E-03
FOLR1	11	71578618	368	-	-4.58	2.5E-05	1.7E-03
TNFSF8	9	116732775	-184	+	-4.55	2.7E-05	1.8E-03
CLDN4	7	72882009	-1120	-	-4.49	3.4E-05	2.1E-03
PARP1	1	224663024	-610	-	-4.46	3.8E-05	2.3E-03

CTLA4	2	204439628	-1128	+	-4.35	5.5E-05	3.1E-03
FGF9	13	21143013	-862	-	4.34	5.7E-05	3.1E-03
EPHA1	7	142816226	-119	-	4.34	5.8E-05	3.1E-03
MCF2	X	138552324	195	+	-4.33	5.9E-05	3.1E-03
THPO	3	185579211	-585	-	-4.18	9.8E-05	4.9E-03
SNRPN	15	22619657	-230	-	-4.17	1.0E-04	5.0E-03
EPO	7	100156197	-162	-	-4.16	1.1E-04	5.0E-03
ITGA6	2	172999898	-718	-	-4.12	1.2E-04	5.5E-03
FOSL2	2	28469667	384	-	-4.11	1.3E-04	5.6E-03
SLC5A5	19	17843842	60	+	-4.09	1.3E-04	5.7E-03
SYK	9	92603307	-584	+	-4.07	1.4E-04	5.7E-03
PTHR1	3	46894070	-170	-	-4.07	1.4E-04	5.7E-03
PMP22	17	15110623	-1254	+	-4.07	1.4E-04	5.7E-03
PWCR1	15	22847798	81	-	-4.06	1.5E-04	5.7E-03
GSTM1	1	110031602	-363	+	-4.03	1.6E-04	6.1E-03
LIG3	17	30331029	-622	-	-4.01	1.7E-04	6.3E-03
GFAP	17	40348450	-56	-	-3.99	1.9E-04	6.7E-03
DNMT2	10	17283886	-199	+	-3.98	1.9E-04	6.7E-03
TIMP3	22	31525688	-1114	-	-3.97	2.0E-04	6.7E-03
PCTK1	X	46962653	77	-	3.97	2.0E-04	6.7E-03
LMTK2	7	97573099	-1034	+	-3.95	2.1E-04	6.7E-03
WNT8B	10	102213275	487	+	-3.95	2.1E-04	6.7E-03
NRAS	1	115061050	-12	-	3.95	2.1E-04	6.7E-03
RAD54B	8	95556713	-227	+	-3.94	2.2E-04	6.7E-03
DMP1	4	88790677	194	+	-3.93	2.2E-04	6.7E-03
BRCA1	17	38531829	-835	-	-3.92	2.3E-04	6.8E-03
MMP9	20	44071042	88	-	-3.92	2.4E-04	6.8E-03
LMO2	11	33871206	-794	-	-3.91	2.4E-04	6.8E-03
APOA1	11	116213809	-261	+	-3.89	2.6E-04	7.1E-03
BCR	22	21852206	-346	+	-3.88	2.7E-04	7.2E-03

SPARC	5	151046905	-195	+	-3.87	2.8E-04	7.2E-03
SERPINE1	7	100557361	189	-	-3.87	2.8E-04	7.2E-03
FGF7	15	47502141	-610	+	-3.87	2.8E-04	7.2E-03
WNT1	12	47658660	157	+	3.86	2.8E-04	7.2E-03
SLC22A2	6	160600058	-109	+	-3.86	2.9E-04	7.2E-03
ABCB4	7	86947735	-51	+	-3.85	3.0E-04	7.2E-03
HOXB2	17	43977879	-488	-	-3.85	3.0E-04	7.2E-03
OPCML	11	132907684	-71	+	-3.82	3.3E-04	7.9E-03
DHCR24	1	55126145	-652	-	-3.81	3.4E-04	8.0E-03
WNT1	12	47658424	-79	-	3.80	3.5E-04	8.0E-03
IPF1	13	27391427	-750	+	-3.80	3.5E-04	8.0E-03
ZIM3	19	62349100	-718	-	-3.78	3.7E-04	8.2E-03
SERPINB2	18	59704983	-939	+	-3.78	3.7E-04	8.2E-03
CDH17	8	95290362	-376	+	-3.77	3.9E-04	8.4E-03
BMPR1A	10	88506464	88	+	-3.76	3.9E-04	8.5E-03
DDR2	1	160868109	-743	-	-3.75	4.1E-04	8.7E-03
PTHR1	3	46894276	36	-	-3.74	4.2E-04	8.7E-03
GNG7	19	2604493	-903	+	-3.71	4.7E-04	9.6E-03
ZIM3	19	62348833	-451	-	-3.70	4.8E-04	9.7E-03
ELK1	X	47394808	156	+	-3.69	4.9E-04	9.7E-03
IGF2AS	11	2118327	4	+	3.69	4.9E-04	9.7E-03
EGF	4	111053257	-242	-	-3.69	5.0E-04	9.7E-03
CYP2E1	10	135190441	-416	+	-3.65	5.6E-04	1.1E-02
PDGFA	7	525644	-78	+	-3.65	5.6E-04	1.1E-02
LEFTY2	1	224196262	-719	+	-3.64	5.9E-04	1.1E-02
MMP2	16	54070610	21	-	3.62	6.2E-04	1.2E-02
AGXT	2	241456655	-180	+	-3.60	6.5E-04	1.2E-02
EGFR	7	55053959	-260	-	-3.60	6.5E-04	1.2E-02
AATK	17	76709831	63	-	-3.60	6.6E-04	1.2E-02
TRIM29	11	119514208	-135	+	-3.59	6.8E-04	1.2E-02

PYCARD	16	31121902	-150	+	-3.58	6.9E-04	1.2E-02
VAMP8	2	85658114	-114	+	-3.57	7.2E-04	1.2E-02
CREBBP	16	3871424	-712	-	-3.57	7.2E-04	1.2E-02
NPR2	9	35781313	-1093	+	-3.56	7.6E-04	1.3E-02
IL1A	2	113259329	113	-	-3.55	7.7E-04	1.3E-02
KLK11	19	56224392	-1290	+	-3.53	8.1E-04	1.3E-02
SLC22A18	11	2877055	-472	-	-3.53	8.3E-04	1.3E-02
MPL	1	43576000	-62	+	-3.53	8.3E-04	1.3E-02
TMPRSS4	11	117452424	-552	+	-3.52	8.4E-04	1.3E-02
LMO2	11	33870264	148	+	-3.52	8.5E-04	1.3E-02
TFF2	21	42644733	-557	-	-3.50	8.8E-04	1.4E-02
NDN	15	21483412	131	-	-3.50	9.0E-04	1.4E-02
RARA	17	35717896	-1076	-	-3.50	9.1E-04	1.4E-02
EPHA2	1	16355354	-203	+	-3.49	9.1E-04	1.4E-02
WNT8B	10	102212572	-216	-	-3.49	9.4E-04	1.4E-02
KRAS	12	25295039	82	+	-3.47	9.8E-04	1.5E-02
HIC1	17	1904448	-565	-	3.47	9.9E-04	1.5E-02
TSG101	11	18505322	-257	-	-3.45	1.0E-03	1.5E-02
HIF1A	14	61231504	-488	+	-3.45	1.1E-03	1.5E-02
ABCB4	7	86947255	429	+	-3.44	1.1E-03	1.5E-02
S100A2	1	151806116	-1186	+	-3.44	1.1E-03	1.5E-02
ACTG2	2	73973146	-455	-	-3.42	1.1E-03	1.6E-02
PIK3R1	5	67557911	-307	+	-3.42	1.1E-03	1.6E-02
IRAK3	12	64869271	-13	+	-3.42	1.1E-03	1.6E-02
GDF10	10	48059133	39	+	3.42	1.2E-03	1.6E-02
PTPRH	19	60412481	173	+	-3.41	1.2E-03	1.6E-02
JAK3	19	17820875	-1075	-	-3.40	1.2E-03	1.6E-02
GABRG3	15	25343690	-75	+	-3.40	1.2E-03	1.6E-02
KRT5	12	51200314	196	-	-3.39	1.2E-03	1.6E-02
SRC	20	35406602	100	-	-3.39	1.3E-03	1.6E-02

BCAP31	X	152644153	.	+	-3.38	1.3E-03	1.7E-02
ASB4	7	94953168	-52	-	-3.37	1.3E-03	1.7E-02
ZIM2	19	62043777	110	+	-3.37	1.4E-03	1.7E-02
IL12B	5	158690034	25	+	-3.36	1.4E-03	1.7E-02
RET	10	42892087	-446	+	3.35	1.4E-03	1.8E-02
CPA4	7	129718965	-1265	-	-3.34	1.5E-03	1.8E-02
MCF2	X	138553543	-1024	-	-3.34	1.5E-03	1.8E-02
CDK10	16	88280653	74	+	-3.33	1.5E-03	1.8E-02
VAMP8	2	85657987	-241	+	-3.33	1.5E-03	1.8E-02
ZP3	7	75864584	-220	+	-3.33	1.5E-03	1.8E-02
MLH3	14	74587814	72	+	-3.33	1.5E-03	1.8E-02
DIO3	14	101097351	-90	+	3.33	1.5E-03	1.8E-02
PSCA	8	143759274	359	+	-3.31	1.6E-03	1.8E-02
TES	7	115637635	-182	+	3.31	1.6E-03	1.8E-02
CCR5	3	58097	-630	-	-3.31	1.6E-03	1.8E-02
KLF5	13	72531130	-13	+	3.31	1.6E-03	1.8E-02
WRN	8	31009351	-969	+	-3.31	1.6E-03	1.8E-02
IAPP	12	21417365	280	+	-3.30	1.6E-03	1.9E-02
STK23	X	152699886	182	-	-3.30	1.7E-03	1.9E-02
GUCY2F	X	108612196	-255	+	-3.29	1.7E-03	1.9E-02
THBS1	15	37660072	-500	+	-3.27	1.8E-03	2.0E-02
GLA	X	100549950	-343	-	-3.27	1.8E-03	2.0E-02
EPHX1	1	224078241	-1358	-	-3.27	1.8E-03	2.0E-02
IFNG	12	66839976	-188	+	-3.27	1.8E-03	2.0E-02
LTA	6	31648100	28	-	-3.26	1.9E-03	2.0E-02
UNG	12	108019628	-170	+	-3.24	2.0E-03	2.1E-02
CXCL9	4	77147397	268	-	-3.22	2.1E-03	2.2E-02
TDGF1	3	46594270	53	-	-3.22	2.1E-03	2.2E-02
ETS1	11	127897118	253	-	3.22	2.1E-03	2.2E-02
RUNX3	1	25164455	-393	-	-3.22	2.1E-03	2.2E-02

EPHA2	1	16355491	-340	-	-3.22	2.1E-03	2.2E-02
SIN3B	19	16800611	-607	+	-3.22	2.1E-03	2.2E-02
SFTPB	2	85749512	-689	-	-3.21	2.2E-03	2.2E-02
ZIM2	19	62043909	-22	+	-3.20	2.2E-03	2.2E-02
JAK3	19	17819956	-156	-	-3.20	2.2E-03	2.3E-02
TMPRSS4	11	117453059	83	+	-3.19	2.3E-03	2.3E-02
KRAS	12	25295772	-651	+	3.18	2.4E-03	2.3E-02
TMEFF2	2	192767395	494	-	-3.17	2.4E-03	2.4E-02
CHI3L2	1	111571578	-226	+	-3.16	2.5E-03	2.4E-02
PTK2	8	142081249	-735	-	-3.16	2.5E-03	2.4E-02
ODC1	2	10506328	-424	+	3.14	2.6E-03	2.5E-02
CD34	1	206151645	-339	-	-3.14	2.6E-03	2.5E-02
MMP7	11	101906629	59	+	-3.14	2.7E-03	2.6E-02
KLK11	19	56223205	-103	-	-3.13	2.7E-03	2.6E-02
SFTPC	8	22075126	13	+	-3.12	2.8E-03	2.6E-02
GABRA5	15	24742740	44	-	-3.12	2.8E-03	2.6E-02
ACVR1	2	158404019	-983	+	-3.10	2.9E-03	2.7E-02
HDAC11	3	13496268	-556	+	3.10	3.0E-03	2.7E-02
IL16	15	79262162	-93	-	-3.10	3.0E-03	2.7E-02
PI3	20	43235518	-1394	-	-3.08	3.1E-03	2.9E-02
GPATC3	1	27099954	-410	-	-3.08	3.1E-03	2.9E-02
ELK3	12	95111824	-514	+	-3.07	3.3E-03	2.9E-02
WNT5A	3	55496328	43	+	3.06	3.3E-03	3.0E-02
DDR2	1	160869183	331	+	-3.04	3.5E-03	3.1E-02
TRIP6	7	100301891	-1090	+	-3.03	3.7E-03	3.2E-02
FGF7	15	47502707	-44	+	-3.03	3.7E-03	3.2E-02
TGFB3	14	75517184	58	-	-3.02	3.7E-03	3.3E-02
HBII-52	15	22966310	-659	+	-3.01	3.8E-03	3.3E-02
EPS8	12	15834038	-437	+	-3.01	3.9E-03	3.3E-02
TCF7L2	10	114700008	-193	-	-3.01	3.9E-03	3.3E-02

ITGB4	17	71229255	144	+	-3.00	4.0E-03	3.4E-02
DLL1	6	170441784	-386	+	2.98	4.2E-03	3.6E-02
MST1R	3	49916161	-87	-	-2.98	4.2E-03	3.6E-02
DSG1	18	27152342	292	+	-2.97	4.3E-03	3.6E-02
UGT1A7	2	234254572	-751	-	-2.97	4.3E-03	3.6E-02
B3GALT5	21	39950794	-330	+	-2.97	4.4E-03	3.6E-02
ACTG2	2	73973699	98	-	-2.96	4.4E-03	3.6E-02
CARD15	16	49288249	-302	-	-2.95	4.6E-03	3.7E-02
UGT1A1	2	234333094	-564	-	-2.95	4.6E-03	3.7E-02
ALK	2	29997753	183	-	2.93	4.8E-03	3.9E-02
MMP9	20	44070765	-189	+	-2.93	4.8E-03	3.9E-02
ALOX12	17	6840213	85	-	-2.93	4.8E-03	3.9E-02
GABRG3	15	25343888	123	-	-2.93	4.9E-03	3.9E-02
MAS1	6	160247307	-657	-	-2.92	5.0E-03	3.9E-02
CHD2	15	91243972	-451	+	-2.92	5.0E-03	3.9E-02
TNFRSF10C	8	23016372	-7	+	-2.92	5.0E-03	3.9E-02
GNG7	19	2603280	310	-	-2.91	5.1E-03	4.0E-02
CDH17	8	95290518	-532	+	-2.90	5.2E-03	4.1E-02
CSF3	17	35425456	242	-	-2.90	5.2E-03	4.1E-02
DSG1	18	27151891	-159	-	-2.90	5.3E-03	4.1E-02
FZD9	7	72485870	-175	+	-2.90	5.3E-03	4.1E-02
HLA-DPA1	6	33149321	35	-	-2.89	5.4E-03	4.1E-02
ICAM1	19	10243021	242	+	-2.89	5.4E-03	4.1E-02
PLSCR3	17	7239318	-751	-	-2.88	5.5E-03	4.2E-02
NOS2A	17	23151970	-288	-	-2.87	5.7E-03	4.3E-02
DIRAS3	1	68288993	55	-	-2.86	5.8E-03	4.3E-02
DSC2	18	26936285	90	+	-2.86	5.9E-03	4.4E-02
DIRAS3	1	68289793	-745	+	-2.85	6.1E-03	4.5E-02
SFN	1	27062338	118	+	-2.85	6.1E-03	4.5E-02
OSM	22	28993028	-188	+	-2.84	6.2E-03	4.5E-02

HOXA9	7	27171422	252	-	-2.84	6.2E-03	4.5E-02
STK23	X	152699680	-24	+	-2.84	6.3E-03	4.5E-02
APOC1	19	50109355	-406	-	-2.83	6.3E-03	4.6E-02
TAL1	1	47468847	-817	+	-2.83	6.4E-03	4.6E-02
KRT1	12	51361244	-798	-	-2.82	6.5E-03	4.6E-02
BCAM	19	50004278	100	-	-2.82	6.5E-03	4.6E-02
SYBL1	X	154764230	23	-	-2.82	6.5E-03	4.6E-02
MMP8	11	102100779	89	-	-2.81	6.7E-03	4.7E-02
TESK2	1	45729672	-252	-	-2.81	6.7E-03	4.7E-02
ATP10A	15	23660110	-147	+	-2.81	6.7E-03	4.7E-02
TNFSF8	9	116732333	258	-	-2.81	6.8E-03	4.7E-02
MSH3	5	79986053	3	+	-2.80	6.9E-03	4.8E-02
GNAS	20	56848104	-86	+	-2.80	6.9E-03	4.8E-02
ERBB2	17	35097860	-59	-	-2.80	6.9E-03	4.8E-02
PTGS1	9	124173130	80	+	-2.79	7.0E-03	4.8E-02
TNFSF10	3	173723965	-2	-	-2.79	7.1E-03	4.8E-02
TRIM29	11	119513884	189	+	-2.79	7.2E-03	4.9E-02
ITGA2	5	52321134	120	+	-2.78	7.4E-03	5.0E-02
EDNRB	13	77447813	-148	-	-2.77	7.4E-03	5.0E-02
KRT13	17	36916067	-676	+	-2.77	7.5E-03	5.0E-02

Table 2.8. Significantly enriched gene sets for genes identified as differentially methylated in cases from the cluster identified with worst survival (Cluster 5) compared to all other cases.

Name	Size	Enrichment Score (ES)	Normalized Enrichment Score (NES)	Nominal P-Value	FDR Q-Value
POSITIONAL GENE SETS					
CHR 7q21	5	-0.68	-1.70	0.015	0.148
GENE ONTOLOGY - BIOLOGICAL PROCESSES					
Negative Regulation of Cellular Metabolic Process	12	0.51	2.29	0.000	0.043
Negative Regulation of Metabolic Process	12	0.51	2.21	0.000	0.046
Homeostatic Process	10	0.56	2.20	0.004	0.032
Chemical Homeostasis	7	0.56	1.89	0.018	0.154
Interphase	5	0.67	1.88	0.006	0.129
Interphase of Mitotic Cell Cycle	5	0.67	1.87	0.024	0.121
Regulation of Biological Quality	21	0.32	1.78	0.032	0.166
Cell Fate Commitment	5	0.62	1.73	0.037	0.184
GENE ONTOLOGY - MOLECULAR FUNCTION					
Purine Ribonucleotide Binding	11	0.48	1.98	0.000	0.192
Purine Nucleotide Binding	11	0.48	1.97	0.013	0.098
Nucleotide Binding	11	0.48	1.96	0.012	0.071
ATP Binding	10	0.47	1.83	0.007	0.122
Adenyl Nucleotide Binding	10	0.47	1.83	0.011	0.098
Adenyl Ribonucleotide Binding	10	0.47	1.82	0.012	0.082
Kinase Activity	29	0.27	1.81	0.028	0.073
Transferase Activity - Transferring Phosphorus Containing Groups	29	0.27	1.72	0.030	0.098
Protein Kinase Activity	28	0.26	1.71	0.027	0.095
Magnesium Ion Binding	5	0.62	1.69	0.024	0.090
Phosphotransferase Activity - Alcohol Group as Acceptor	28	0.26	1.62	0.036	0.114
Protein Serine Threonine Kinase Activity	14	0.34	1.59	0.034	0.122

CHAPTER 3. PRETREATMENT DIETARY INTAKE IS ASSOCIATED WITH TUMOR SUPPRESSOR DNA METHYLATION IN HEAD AND NECK SQUAMOUS CELL CARCINOMAS

3.1 Abstract

Background**:** Diet is associated with cancer prognosis, including head and neck cancer (HNC), and has been hypothesized to influence epigenetic state by determining the availability of functional groups involved in the modification of DNA and histone proteins.

Objective**:** To describe the association between pretreatment diet and HNC tumor DNA methylation.

Methods**:** Information on usual pretreatment food and nutrient intake was estimated via food frequency questionnaire (FFQ) on 49 HNC cases. Tumor DNA methylation patterns were assessed using the Illumina Goldengate Methylation Cancer Panel. First, a methylation score, the sum of individual hypermethylated tumor suppressor associated CpG sites, was calculated and associated with dietary intake of micronutrients involved in one-carbon metabolism and antioxidant activity, and food groups abundant in these nutrients. Second, gene specific analyses using linear modeling with empirical Bayesian variance estimation were conducted to identify if methylation at individual CpG sites was associated with diet. All models were controlled for age, sex, smoking, alcohol, and HPV status.

Results**:** Individuals reporting in the highest quartile of folate, vitamin B12, and vitamin A intake, compared to those in the lowest quartile, had significantly less tumor suppressor gene methylation (p=0.04, 0.04, and p=0.01, respectively), as did patients reporting the highest cruciferous vegetable intake (p=0.04). Gene specific analyses, identified differential associations between DNA methylation and vitamin B12 and vitamin A intake when stratifying by HPV status.

Conclusions: These preliminary results suggest that intake of folate, vitamin A, and vitamin B12 may be associated with the tumor DNA methylation profile in HNC, and enhance tumor suppression.

3.2 Introduction

Head and neck cancer (HNC) is the sixth most common type of cancer and accounts for approximately 5% of all new cancer diagnoses worldwide each year (Argiris et al. 2008; Ferlay et al. 2010). These malignancies develop from the epithelial tissue of the oral cavity, nasopharynx, oropharynx or larynx, and more than 90% are of squamous cell histologic type. HNC is most often associated with extensive lifetime exposure to tobacco and alcohol consumption, although high-risk human papillomavirus (HPV) has recently emerged as the primary etiologic factor responsible for an increasing number of oropharyngeal tumors (Chaturvedi et al. 2011; Maxwell et al. 2010).

Epidemiologic evidence also supports the hypothesis that diet modulates risk, progression and prognosis of head and neck cancer (Duffy et al. 2009; Lucenteforte et al. 2009; Sandoval et al. 2009). However, the molecular mechanisms by which dietary compounds exert their effects are not entirely understood. Epigenetic dysregulation is a key mechanism in tumorigenesis that may be influenced by dietary intake. Aberrant DNA methylation in cancer is typically characterized by gene promoter specific hypermethylation, leading to gene silencing, and hypomethylation of intergenic regions, leading to overall genomic instability (Ehrlich 2002; McKay and Mathers 2011).

Dietary intake has been widely hypothesized to influence epigenetic state by determining the availability of functional groups involved in the covalent modification of DNA and histone proteins (Burdge et al. 2007; Oommen et al. 2005). Specifically, nutrients that act as methyl donors, such as methionine and folate, or nutrients that act as cofactors in one carbon metabolism pathway, such as betaine, vitamin B12, and choline, are required for establishment and maintenance of methylation marks both on DNA and histone protein tails (Cooney et al. 2002; Dobosy et al. 2008; Keyes et al. 2007; Waterland and Jirtle 2003; Waterland et al. 2006). A deficiency or excess of any of the nutrients involved in one-carbon metabolism can potentially alter the availability of S-

adenosylmethione (SAM), the universal donor involved in DNA methylation, and ultimately alter epigenetic state (McKay and Mathers 2011). In addition to diet, it has been shown that in smokers, chronic DNA damage coupled with reduced DNA repair capacity may induce gene promoter methylation (Leng et al. 2008). Given these findings we hypothesize that, in addition to micronutrients involved in one-carbon metabolism, micronutrients involved in cell growth and differentiation, such as vitamin A, or with antioxidant activity such as vitamin C, vitamin E and β-carotene may indirectly influence DNA methylation by reducing DNA damage (Stidley et al. 2010).

The objective of this study was to test the hypothesis that pretreatment dietary intake of methyl donors, antioxidants, and foods abundant in these micronutrients is associated with tumor DNA methylation in head and neck cancer. A better understanding of how dietary intake is associated with tumor DNA methylation in head and neck cancer will enhance our knowledge of disease pathogenesis and may allow for the development of individualized medical nutrition therapy regimens.

3.3 Methods

Design

This was a cross-sectional study of patients enrolled in the University of Michigan Head and Neck Cancer Specialized Program of Research Excellence (SPORE). The independent variables were micronutrients involved in one-carbon metabolism, antioxidant nutrients, and food groups abundant in these micronutrients. Confounding variables included age, sex, smoking, alcohol problem and HPV-status. The outcome variable was DNA methylation. Our analyses were limited to foods and nutrients we hypothesized are associated with DNA methylation—namely those involved in one-carbon metabolism or with antioxidant activity. A two-tiered analysis was conducted; first we observed whether dietary intake is associated with methylation at CpG sites in tumor suppressor genes overall and second, we observed whether DNA methylation at specific CpG sites is associated with dietary intake.

Study Population

Sixty-eight newly diagnosed cases of head and neck cancer were identified and recruited at two sites participating in the University of Michigan Head and Neck SPORE: the University of Michigan Health System and the Ann Arbor Veterans Administration (VA) Medical Center. Individuals eligible for participation included patients diagnosed with first primary head and neck cancer between January 1, 2003 and December 31, 2005 who signed written, informed consent, completed an epidemiologic questionnaire that includes questions on demographic and behavioral data, and had a paraffin-embedded tumor available for analysis with adequate residual tissue for microdissection. Exclusion criteria included: 1) < 18 years of age; 2) pregnant; 3) non-English speaking; 4) diagnosed as mentally unstable; or 5) a diagnosis of another non-upper aerodigestive tract cancers (such as thyroid or skin cancer). For the purposes of the current analysis, patients were excluded if they had not completed a baseline FFQ (*n* = 19), yielding a final sample size of 49 participants. Excluded patients did not differ significantly from included patients in respect to age, sex, tumor site, cancer stage or smoking status. This study was approved as being within the ethical standards of the Institutional Review Boards of the University of Michigan and the Ann Arbor VA Medical Center.

Tumor blocks were recut for uniform histopathologic review and microdissection, with the first and last slides of a series of 12 reviewed by a qualified pathologist (JM and ND) to confirm the original diagnosis and to circle areas for DNA extraction. Percent cellularity was estimated for each tumor, and areas with >70% cancer cellularity were designated for use in the analyses.

Measures

FFPE tissue, DNA isolation, and bisulfite conversion

Regions that were identified for DNA extraction were cored from the formalin fixed paraffin embedded (FFPE) tissue blocks using an 18 gauge needle. Isolation of DNA from the cored tissue samples was performed using the QIAamp DNA FFPE Tissue Kit (Qiagen, Valencia, CA) with a modification to the manufacturer's recommended lysis protocol (incubation overnight at 56^{o}C in lysis buffer). DNA concentration and purity was confirmed via NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA). We performed sodium

bisulfite modification on 500ng to 1µg of DNA using the EZ DNA Methylation kit (Zymo Research, Orange, CA) following the manufacturer's recommended protocol.

Bead Array Methods

The commercially available Illumina Goldengate® Methylation Cancer Panel was utilized to detect DNA methylation patterns in tumor samples. The Cancer Panel interrogates 1505 CpG sites located in known CpG islands across 807 genes related to cancer, including oncogenes, tumor suppressor genes, imprinted genes, and genes involved in cell cycle regulation, DNA repair, apoptosis and metastasis. Bead arrays were run at the University of Michigan DNA Sequencing Core Facility according to the manufacturer's protocol. Briefly, bisulfite converted tumor DNA was hybridized to the bead array as described previously(Bibikova et al. 2006), and bead arrays were imaged using Illumina Cancer Panel Reader software. Raw bead array fluorescence data was initially analyzed using Illumina BeadStudio Methylation software, which converts fluorescence values of the methylated (Cy5) and unmethylated (Cy3) alleles into an average methylation value at a specific probe using the formula $\beta = [\mathrm{Max}(\mathrm{Cy5},0)]/[\mathrm{Max}(\mathrm{Cy5},0) + \mathrm{Max}(\mathrm{Cy3},0) + 100]$ with $\beta \epsilon (0,1)$.

Methylation at specific CpG probes on the Goldengate Cancer Panel has been shown to be biased by probe thermodynamic properties (Kuan et al. 2010). Known biases include probe length and GC content, which can affect the melting temperature of the probes as well as probe fluorescence intensities. Thus, we used the method proposed by Kuan et al. to normalize our average β values based on probe length and GC content (Kuan et al. 2010).

Detection p-values on the Goldengate Cancer Panel are calculated based on fluorescence signal at a probe compared to background fluorescence and represent one minus the probability) that a signal is stronger than background fluorescence. The weighted methodology proposed by Kuan et al. was used to develop sample and site weights based on p-values of detection. Both samples and sites with larger detection p-values are generally considered less reliable and were down-weighted in further gene specific analyses.

Dietary Intake Estimation

Dietary components of interest in this study included antioxidant micronutrients, micronutrients involved in one-carbon metabolism, and food groups serving as rich sources of these micronutrients including cruciferous vegetables, green leafy vegetables, dark yellow vegetables, refined grains, whole grains, red meat, and legumes. Pre-treatment dietary intake was collected using a self-administered, semi-quantitative FFQ, designed to assess respondents' usual dietary intake from food and supplements over the year prior to diagnosis. The reproducibility and validity of this 131-item questionnaire has been reported previously (Rimm et al. 1992; Willett 1998; Willett et al. 1985). The FFQ includes given standard portion sizes for each item (i.e. 1 orange or 3-5 oz chicken) which allowed participants to choose their average frequency of consumption over the past year from a list of 1-9 choices ranging from "almost never" to "≥6 times per day". Daily energy and nutrient intake was estimated by summing intakes from each food based on the given standard portion size, reported frequency of consumption, and nutrient content of each food item (Willett 1998). Daily food group servings were estimated by summing the frequency weights of each food item based on reported daily frequencies of consumption.

Covariates

Each participant completed a self-administered health questionnaire at baseline. This questionnaire was designed to collect demographic data as well as data on smoking, alcohol use, physical activity, sleep, comorbidities, depression and quality of life. Age, sex, smoking and alcohol data were collected from the health questionnaire. Smoking data permitted analysis by "current-past-never" smoking. Alcohol consumption was measured using the previously validated Alcohol Use Disorders Identification Test (AUDIT) (Saunders et al. 1993). An AUDIT score ≥8 was considered problem drinking. Tumor site and stage was recorded from operative notes and surgical pathology forms at baseline. As we have previously shown HPV status to be associated with both tumor DNA methylation profiles at specific gene promoter regions (Sartor et al. 2011b) and micronutrient intake (Arthur et al. 2011), we considered HPV status as a potential confounder in statistical analyses.

HPV status was determined by an ultra-sensitive 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. 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.

Statistical Methods

Methylation Score

To create a measurement of stochastic methylation across genes that could be associated with dietary intake, we developed a novel method of calculating a methylation score based on methylation of genes associated with specific cellular functions. Since the Goldengate Cancer Panel quantifies methylation of cancer-related genes with various functions, we first subsetted genes associated with our main cellular function of interest, tumor suppression. To identify genes associated with tumor suppression, we conducted a simply query of the National Center for Biotechnology Information (NCBI) Gene database for “tumor suppressor” and extracted Gene IDs. A total of 444 individual CpG sites across 237 genes on the Goldengate Cancer Panel were linked to tumor suppression. Next, for each study participant, the methylation score was defined as the number of tumor suppressor gene-associated CpG sites with normalized average β values above 0.5. A β value cutoff of 0.5 was established as a methylation level likely to result in decreased expression of tumor suppressor genes. Thus, the methylation score ranged from 0 to 444. We infer that a higher methylation score is associated with higher levels of DNA

methylation of promoter regions of genes associated with tumor suppression and cell cycle regulation.

Micronutrients were adjusted for total energy intake using the residual method (Willett et al. 1997). Descriptive statistics, including means, medians, maxima and minima were calculated for both the energy adjusted nutrients and the methylation score. Associations between energy-adjusted micronutrient intake, food group intake, and methylation score were assessed with Spearman rank correlation due to the typical non-normal distribution of micronutrient and food group intake. Additionally, we used linear regression models with methylation score as the outcome and energy-adjusted micronutrient or food group intake as the main predictor, categorizing dietary variables into quartiles, and adjusting for age, sex, smoking status, problem alcohol drinking, and HPV status, as well as analyzing for trend by treating the quartiles as a continuous variable in regression analysis. To examine the robustness of our findings, we performed a sensitivity analysis in which we categorized micronutrient intake into tertiles as opposed to quartiles and adjusted for the same covariates. Regression diagnostic plots were examined to ensure that associations observed were not being driven by influential points. All analyses were performed in SAS software, Version 9.2 or R Version 2.13.0.

Site Specific Analyses

To determine whether methylation at specific sites of the genome may be labile to effects due to dietary intake, we calculated associations between DNA methylation at specific CpG sites and dietary intake, stratifying by HPV status. To limit the number of multiple comparisons conducted, we restricted our analysis to micronutrients identified as significant in the tumor suppressor gene methylation score analysis following sensitivity analysis, vitamin B12 and vitamin A, and compared site specific DNA methylation between highest and lowest quartile of intake.

The association between micronutrient intakes and individual CpG site DNA methylation at the 1505 CpG sites measured on the Goldengate Cancer Panel were examined using the *Limma* package in R 2.10.1 (Smyth 2004). Sample weights based on detection p-values across samples were used in the

lmFit function from the *Limma* package to downweight samples with higher detection p-values. An empirical Bayes method (using the *eBayes* function in *Limma*) was used to shrink standard errors to a common value and to rank CpG sites in order of differential methylation. Multiple comparisons were accounted for using the q-value method previously described (Storey and Tibshirani 2003). Micronutrient intakes were categorized into quartiles, and differences in individual CpG site methylation between the highest and lowest quartile of intake were calculated, adjusting for age, sex, smoking status, and problem drinking.

Based on the site specific analyses, we used Gene Set Enrichment Analysis (GSEA) to identify common pathways and chromosomal locations for genes identified as differentially methylated by micronutrient intake (Subramanian et al. 2005). All sites were ranked by t-value and input into GSEA as a ranked list. The full list of genes assayed on the Goldengate Cancer Panel was input into GSEA as a chip platform file, which provided the background for the enrichment analysis. Weighted enrichment statistics were calculated by the GSEA software, using a minimum gene set size of 5 to account for the limited number of genes assayed by the Goldengate Cancer Panel. Enrichment at the nominal p-value of 0.05 and an adjusted q-value of 0.25 was considered statistically significant.

3.4 Results

Descriptive Statistics: Study Sample

The mean age of the 49 study participants with both dietary intake and methylation data was 57.3 years (range 41-75 years); 80% of the participants were male. The majority of the cancer sites were oropharynx (53%), evidenced by the relatively high proportion of HPV(+) cases (43%). Approximately 85% of the study participants reported being current or former smokers, with an estimated one-third reporting an alcohol problem. Clinical characteristics are displayed in **Table 3.1**.

Dietary Intake Associations with Tumor Suppressor Methylation

Reported average daily intakes of micronutrients and food group servings reported by study participants are presented in **Table 3.2**. The distribution of the

tumor suppressor methylation score for all subjects is shown in **Figure 3.1**. **Table 3.3** shows associations across quartiles of micronutrient and food group intake and tumor suppressor DNA methylation. Higher antioxidant, methyl donor and food group intake tended to be negatively associated with tumor suppressor methylation score, the sum of individual hypermethylated tumor suppressor associated CpG sites, suggesting an inverse relationship between DNA methylation at promoter regions of these genes and micronutrient intake. The regression coefficients in **Table 3.3** can be interpreted as the number of CpG sites of genes associated with tumor suppression activity that are methylated more than 50% compared to individuals in the lowest quartile of intake. Specifically, individuals in the lowest quartiles of dietary folate, vitamin B12, and vitamin A intake were found to have significantly higher tumor suppressor methylation scores compared to individuals in the highest quartiles of intake. Additionally, tests for trend across quartiles of intake for dietary folate, vitamin B12, and vitamin A intake indicated statistically significantly higher methylation score with each decreasing quartile (**Table 3.3**). Individuals in the lowest quartile of cruciferous vegetable intake had a significantly higher tumor suppressor methylation score than individuals in the highest quartile of intake, and the test for trend across quartiles was also statistically significant (p=0.013). Interestingly, individuals consuming the highest servings per day of refined grains had significantly higher tumor suppressor methylation score than individuals consuming the lowest servings per day of refined grains. However the test for trend across quartiles of intake for refined grains was not statistically significant.

In a sensitivity analysis, in which we categorized micronutrient intake into tertiles as opposed to quartiles and adjusted for the same covariates, our findings were similar to the findings reported above in that individuals in the lowest tertile of vitamin B12, vitamin A, and cruciferous vegetable intake had significantly higher tumor suppressor methylation score than those in the highest tertile of intake ($\beta = 18.4$, $\beta = -18.2$, and $\beta = 18.2$, respectively; data not shown). The tests for trend across tertiles of intake for the relationship between vitamin B12, vitamin A, and cruciferous vegetables and tumor suppressor methylation score were also significant ($P_{trend} = 0.015$, $P_{trend} = 0.034$, and $P_{trend} = 0.022$,

respectively). The inverse association between tumor suppressor methylation score and food folate was no longer statistically significant (β = -14.7 for tertile 3 versus tertile 1, P_{trend} = 0.074).

CpG Site Specific Associations with Vitamin A and Vitamin B12 Intake

With respect to vitamin A intake, there were no significant associations with site-specific DNA methylation for HPV(+) individuals (all adjusted p-values >0.15). In contrast, a number of CpG sites were differentially methylated for HPV(-) individuals with the highest levels of both vitamin A and vitamin B12 intake compared to those with the lowest levels of intake. Interestingly, vitamin B12 intake was also associated with hypomethylation of a CpG site associated with *RARB* (retinoic acid receptor beta) in HPV(+) individuals after adjusting for multiple comparisons (adjusted p-value <0.05). The ten sites identified as the most significantly differentially methylated by both vitamin A and vitamin B12 intake in HPV(-) individuals, as well as the one site differentially methylated by vitamin B12 intake in HPV(+) individuals are shown in **Table 3.4**. Consistent with the findings of the methylation score analysis, the most differentially methylated CpG sites were, by and large, hypomethylated in tumors from individuals in the highest quartile of vitamin A and vitamin B12 intake compared to tumors from individuals in the lowest quartiles of intake. There was one exception, *CHGA* (Chromogranin A), which was hypermethylated in tumors from HPV(-) negative individuals who consumed the largest amounts of vitamin B12. CpG sites associated with *C4B* (Complement Component 4B), *GML* (Glycosylphosphatidylinositol anchored molecule like protein), *MLF1* (Myeloid Leukemia Factor 1), and *THPO* (Thrombopoietin) were significantly associated with methylation in tumors from HPV(-) individuals based on both vitamin A and vitamin B12 intake.

Table 3.5 presents the results of the ranked Gene Set Enrichment Analysis (GSEA). An enrichment analysis, by genomic location, of genes differentially methylated by vitamin B12 intake in HPV(+) cancers revealed a trend towards hypomethylation of chromosome 12p12 with increased vitamin B12 intake. The same chromosomal region was identified as hypomethylated in tumors from HPV(-) individuals with the highest levels of vitamin A intake, as was

chromosome 8p21. While there was no consistent enrichment of Gene Ontology Biological Processes or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways based on dietary intake for HPV(-) tumors, for HPV(+) tumors there was a significant enrichment of processes involved in immune function based on vitamin B12 intake.

3.5 Discussion

To our knowledge, this is the first study to comprehensively examine the association between dietary intake and promoter methylation of genes related to head and neck cancer. These novel findings suggest diet is significantly associated with epigenetic events that occur in head and neck cancer. While previous studies have indicated deficiencies of folate, vitamin B12 and vitamin A are associated with increased risk of developing head and neck cancer (Almadori et al. 2005; Lucenteforte et al. 2009; Nacci et al. 2008; Pelucchi et al. 2003) as well as with poorer prognosis in patients diagnosed with the disease (Kawakita et al. 2011), our study has shown a decrease in tumor suppressor methylation in the highest quartile of vitamin B12 and vitamin A intake. Epigenetic dysregulation, such as the silencing of tumor suppressor genes via promoter hypermethylation, is a major event in malignant disease.

Though the underlying biological mechanism related to how folate and vitamin B12 deficiencies may be associated with DNA hypermethylation has not been elucidated, we have two suggested hypotheses. First, folate and vitamin B12 play important roles in the synthesis of DNA and DNA damage repair (Blount et al. 1997; Choi and Mason 2000; Metz et al. 1968). Increases in DNA methylation have been observed in an experimental model inducing *in vitro* DNA damage, in the form of DNA strand breaks, and homology directed DNA repair (Cuozzo et al. 2007). An increased rate of DNA strand breaks was also observed in both animal models and cell culture models of folate deficiency (Branda and Blickensderfer 1993; Pogribny et al. 1995), potentially providing a mechanism that links folate deficiency, DNA strand breaks, and DNA methylation. Second, while the DNA methylation changes that occur in cancer have been well characterized, there is also interest in describing the DNA methylation changes

that occur with aging, as these changes could reflect what occurs in the early stages of carcinogenesis. A recent study has described that DNA methylation of two genes, *ERα* and *SFRP1*, in colon tissue increases with age and also with folic acid supplementation status (Wallace et al. 2010). In our study, we not only observed a decrease in tumor suppressor methylation in the highest quartile of vitamin A, vitamin B12 and food folate intake, but also a trend towards increased tumor suppressor methylation in the individuals who consumed the highest levels of refined grains. During the refinement process these grains are fortified with folic acid, niacin, riboflavin, thiamin and iron. It is unclear how these nutrients may be associated with increased DNA methylation and more research is necessary. The relative bioavailability of these nutrients could modulate the normal stochastic changes in methylation that occur due to aging and may be reflected in tumor DNA methylation.

Other studies support our finding that increased consumption of folate and folate-rich foods are associated with decreased methylation of genes associated with tumor suppression. Another study conducted in head and neck cancer patients reported increased methylation of $p16^{INK4a}$, a gene involved in cell cycle regulation that is commonly lost or silenced in head and neck cancer via epigenetic mechanisms, in those consuming the lowest levels of folate compared to the highest levels (Kraunz et al. 2006).

Based on our findings that folate intake was inversely associated with tumor suppressor DNA methylation, it is not surprising that high cruciferous vegetable consumption was associated with lower methylation levels, given this food group is a rich source of folate. However, it is interesting to note that cruciferous vegetables are also rich sources of glucosinolates such as sulforphane and indole-3-carbinol, both of which are believed to exhibit anti-cancer activities. Though there is little evidence suggesting involvement of sulforaphane in DNA methylation, this compound has been shown to act as a histone deacetylase (HDAC) inhibitor. HDACs have been shown to be widely overexpressed in a number of cancers and are involved in gene silencing by inducing a more condensed chromatin state, leaving DNA less accessible for transcription (Ropero and Esteller 2007). HDAC inhibitors, such as sulforaphane,

have been shown to be promising cancer therapies because of their ability to induce expression of epigenetically silenced genes (Marks et al. 2004).

The finding that vitamin A intake is significantly associated with decreased DNA methylation in HPV(-) tumors is unexpected, considering vitamin A does not play a direct role in the one-carbon metabolism pathway. However, it is possible vitamin A can alter DNA methylation by influencing the availability and activity of methyltransferases. Two studies have reported rats treated with all-*trans*-retinoic acid (ATRA), a vitamin A metabolite, displayed increased levels and activity of glycine *N*-methyltransferase (GNMT), the enzyme involved in the conversion of S-adenosyl methionine (SAM) to S-adenosylhomocysteine (SAH) (Ozias and Schalinske 2003; Rowling et al. 2002). Increased GNMT activity led to a reduction of SAM, thus potentially resulting in a reduction of methylation. Similarly, Das et al reported demethylation of 402 gene promoters in neuroblastoma cells treated with ATRA. These cells also exhibited down-regulation of the methyltransferases, *DNMT1* and *DNMT3B*, which the authors proposed as the potential mechanism by which ATRA exerts its' effect on DNA methylation(Das et al. 2010).

The GSEA findings of this study confirm the previous work of our group and others describing the differences in tumor biology between HPV(+) and HPV(-) HNCs (Arthur et al. 2011; Sartor et al. 2011a). In this study, we identified that gene sets involved in immune function, including cytokine production and antigen processing and presentation, were significantly enriched based on DNA methylation differences in HPV(+) tumors from individuals with high and low levels of vitamin B12 intake. There was no consistent KEGG or Gene Ontology Biological Processes enriched in HPV(-) tumors, based on either vitamin A or vitamin B12 intake. A positional enrichment analysis identified the same chromosomal region, chromosome 12p12, as differentially methylated in HPV(+) tumors based on vitamin B12 intake as HPV(-) tumors based on vitamin A intake. This region has previously been identified as a candidate tumor suppressor region in head and neck and other cancers (Gunduz et al. 2005), and may represent a region of the genome that is more labile to the epigenetic influences of dietary intake.

Results of this preliminary study should be interpreted in light of several limitations. Our small sample size poses analytic limitations in detecting associations of interest and the complexity of statistical tests used. Although we have adjusted for several covariates, there are no doubt other unmeasured sources of confounding we were not able to consider. The FFQ is susceptible to measurement error and potential unquantified systematic biases that may have led to misclassification of dietary intake. The FFQ was designed to assess usual dietary intake over the course of the year prior to diagnosis with HNC. Thus we cannot make any conclusions about the association between lifetime dietary exposures and tumor DNA methylation. While the Goldengate Cancer Panel provides a comprehensive measurement of methylation across 1505 CpG sites in promoter regions of genes with known roles in cancer, this technology does not provide an unbiased assessment of the epigenome. Finally, due to these limitations, the findings of this study should be validated in another set of tumors from similar individuals.

In conclusion, dietary intake of micronutrients involved in one-carbon metabolism such as folate and vitamin B12, as well as other micronutrients such as vitamin A, may be significant regulators of epigenetic events occurring in head and neck cancer. These findings have potential clinical implications for the treatment of head and neck cancer. The idea that dietary interventions could potentially reprogram the epigenome in such a way as to optimize the likelihood of positive disease outcomes is appealing. Further research is necessary to better understand the complex role these micronutrients play in regulating the epigenetic process, as well as how nutrients may differentially regulate this process according to HPV-status. Such studies could give rise to highly specific randomized controlled trials and ultimately to the development of individualized medical nutrition therapy regimens that may improve prognosis in the head and neck cancer population.

3.6 References

Almadori G, Bussu F, Galli J, Cadoni G, Zappacosta B, Persichilli S, et al. 2005. Serum levels of folate, homocysteine, and vitamin b12 in head and neck squamous cell carcinoma and in laryngeal leukoplakia. Cancer 103:284-292.

Argiris A, Karamouzis MV, Raben D, Ferris RL. 2008. Head and neck cancer. Lancet 371:1695-1709.

Arthur AE, Duffy SA, Sanchez GI, Gruber SB, Terrell JE, Hebert JR, et al. 2011. Higher micronutrient intake is associated with human papillomavirus-positive head and neck cancer: A case-only analysis. Nutrition and cancer:1.

Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. 2006. High-throughput DNA methylation profiling using universal bead arrays. Genome Res 16:383-393.

Blount BC, Mack MM, Wehr CM, MacGregor JT, Hiatt RA, Wang G, et al. 1997. Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: Implications for cancer and neuronal damage. Proc Natl Acad Sci U S A 94:3290-3295.

Branda RF, Blickensderfer DB. 1993. Folate deficiency increases genetic damage caused by alkylating agents and gamma-irradiation in chinese hamster ovary cells. Cancer Res 53:5401-5408.

Burdge GC, Hanson MA, Slater-Jefferies JL, Lillycrop KA. 2007. Epigenetic regulation of transcription: A mechanism for inducing variations in phenotype (fetal programming) by differences in nutrition during early life? Br J Nutr 97:1036-1046.

Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, et al. 2011. Human papillomavirus and rising oropharyngeal cancer incidence in the united states. J Clin Oncol.

Choi SW, Mason JB. 2000. Folate and carcinogenesis: An integrated scheme. J Nutr 130:129-132.

Cooney CA, Dave AA, Wolff GL. 2002. Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. J Nutr 132:2393S-2400S.

Cuozzo C, Porcellini A, Angrisano T, Morano A, Lee B, Di Pardo A, et al. 2007. DNA damage, homology-directed repair, and DNA methylation. PLoS Genet 3:e110.

Das S, Foley N, Bryan K, Watters KM, Bray I, Murphy DM, et al. 2010. Microrna mediates DNA demethylation events triggered by retinoic acid during neuroblastoma cell differentiation. Cancer Res 70:7874-7881.

Dobosy JR, Fu VX, Desotelle JA, Srinivasan R, Kenowski ML, Almassi N, et al. 2008. A methyl-deficient diet modifies histone methylation and alters igf2 and h19 repression in the prostate. Prostate 68:1187-1195.

Duffy SA, Ronis DL, McLean S, Fowler KE, Gruber SB, Wolf GT, et al. 2009. Pretreatment health behaviors predict survival among patients with head and neck squamous cell carcinoma. J Clin Oncol 27:1969-1975.

Ehrlich M. 2002. DNA methylation in cancer: Too much, but also too little. Oncogene 21:5400-5413.

Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008: Globocan 2008. Int J Cancer 127:2893-2917.

Gunduz M, Nagatsuka H, Demircan K, Gunduz E, Cengiz B, Ouchida M, et al. 2005. Frequent deletion and down-regulation of ing4, a candidate tumor suppressor gene at 12p13, in head and neck squamous cell carcinomas. Gene 356:109-117.

Kawakita D, Matsuo K, Sato F, Oze I, Hosono S, Ito H, et al. 2011. Association between dietary folate intake and clinical outcome in head and neck squamous cell carcinoma. Ann Oncol.

Keyes MK, Jang H, Mason JB, Liu Z, Crott JW, Smith DE, et al. 2007. Older age and dietary folate are determinants of genomic and p16-specific DNA methylation in mouse colon. J Nutr 137:1713-1717.

Kraunz KS, Hsiung D, McClean MD, Liu M, Osanyingbemi J, Nelson HH, et al. 2006. Dietary folate is associated with p16(ink4a) methylation in head and neck squamous cell carcinoma. Int J Cancer 119:1553-1557.

Kuan PF, Wang S, Zhou X, Chu H. 2010. A statistical framework for illumina DNA methylation arrays. Bioinformatics 26:2849-2855.

Leng S, Stidley CA, Willink R, Bernauer A, Do K, Picchi MA, et al. 2008. Double-strand break damage and associated DNA repair genes predispose smokers to gene methylation. Cancer Res 68:3049-3056.

Lucenteforte E, Garavello W, Bosetti C, La Vecchia C. 2009. Dietary factors and oral and pharyngeal cancer risk. Oral Oncol 45:461-467.

Marks PA, Richon VM, Miller T, Kelly WK. 2004. Histone deacetylase inhibitors. Adv Cancer Res 91:137-168.

Maxwell JH, Kumar B, Feng FY, Worden FP, Lee JS, Eisbruch A, et al. 2010. Tobacco use in human papillomavirus-positive advanced oropharynx cancer patients related to increased risk of distant metastases and tumor recurrence. Clin Cancer Res 16:1226-1235.

McKay JA, Mathers JC. 2011. Diet induced epigenetic changes and their implications for health. Acta Physiol (Oxf) 202:103-118.

Metz J, Kelly A, Swett VC, Waxman S, Herbert V. 1968. Deranged DNA synthesis by bone marrow from vitamin b-12-deficient humans. Br J Haematol 14:575-592.

Nacci A, Dallan I, Bruschini L, Traino AC, Panicucci E, Bruschini P, et al. 2008. Plasma homocysteine, folate, and vitamin b12 levels in patients with laryngeal cancer. Arch Otolaryngol Head Neck Surg 134:1328-1333.

Oommen AM, Griffin JB, Sarath G, Zempleni J. 2005. Roles for nutrients in epigenetic events. J Nutr Biochem 16:74-77.

Ozias MK, Schalinske KL. 2003. All-trans-retinoic acid rapidly induces glycine n-methyltransferase in a dose-dependent manner and reduces circulating methionine and homocysteine levels in rats. J Nutr 133:4090-4094.

Pelucchi C, Talamini R, Negri E, Levi F, Conti E, Franceschi S, et al. 2003. Folate intake and risk of oral and pharyngeal cancer. Ann Oncol 14:1677-1681.

Pogribny IP, Basnakian AG, Miller BJ, Lopatina NG, Poirier LA, James SJ. 1995. Breaks in genomic DNA and within the p53 gene are associated with hypomethylation in livers of folate/methyl-deficient rats. Cancer Res 55:1894-1901.

Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. 1992. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 135:1114-1126; discussion 1127-1136.

Ropero S, Esteller M. 2007. The role of histone deacetylases (hdacs) in human cancer. Mol Oncol 1:19-25.

Rowling MJ, McMullen MH, Schalinske KL. 2002. Vitamin a and its derivatives induce hepatic glycine n-methyltransferase and hypomethylation of DNA in rats. J Nutr 132:365-369.

Sandoval M, Font R, Manos M, Dicenta M, Quintana MJ, Bosch FX, et al. 2009. The role of vegetable and fruit consumption and other habits on survival following the diagnosis of oral cancer: A prospective study in spain. Int J Oral Maxillofac Surg 38:31-39.

Sartor MA, Dolinoy DC, Jones TR, Colacino JA, Prince ME, Carey TE, et al. 2011a. Genome-wide methylation and expression differences in hpv(+) and hpv(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. Epigenetics 6:777-787.

Sartor MA, Dolinoy DC, Jones TR, Colacino JA, Prince ME, Carey TE, et al. 2011b. Genome-wide methylation and expression differences in hpv(+) and hpv(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. Epigenetics 6.

Saunders JB, Aasland OG, Babor TF, FUENTE JR, Grant M. 1993. Development of the alcohol use disorders identification test (audit): Who collaborative

project on early detection of persons with harmful alcohol consumption-ii. Addiction 88:791-804.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3:Article3.

Stidley CA, Picchi MA, Leng S, Willink R, Crowell RE, Flores KG, et al. 2010. Multivitamins, folate, and green vegetables protect against gene promoter methylation in the aerodigestive tract of smokers. Cancer Res 70:568-574.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100:9440-9445.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545-15550.

Wallace K, Grau MV, Levine AJ, Shen L, Hamdan R, Chen X, et al. 2010. Association between folate levels and cpg island hypermethylation in normal colorectal mucosa. Cancer Prevention Research 3:1552.

Waterland RA, Jirtle RL. 2003. Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol 23:5293-5300.

Waterland RA, Dolinoy DC, Lin JR, Smith CA, Shi X, Tahiliani KG. 2006. Maternal methyl supplements increase offspring DNA methylation at axin fused. Genesis 44:401-406.

Willett W. 1998. Nutritional epidemiology. 2nd ed. New York:Oxford University Press.

Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. 1985. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 122:51-65.

Willett WC, Howe GR, Kushi LH. 1997. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr 65:1220S-1228S; discussion 1229S-1231S.

Figure 3.1. Histogram of the tumor suppressor methylation score, the sum of CpG sites with β values greater than 0.5.

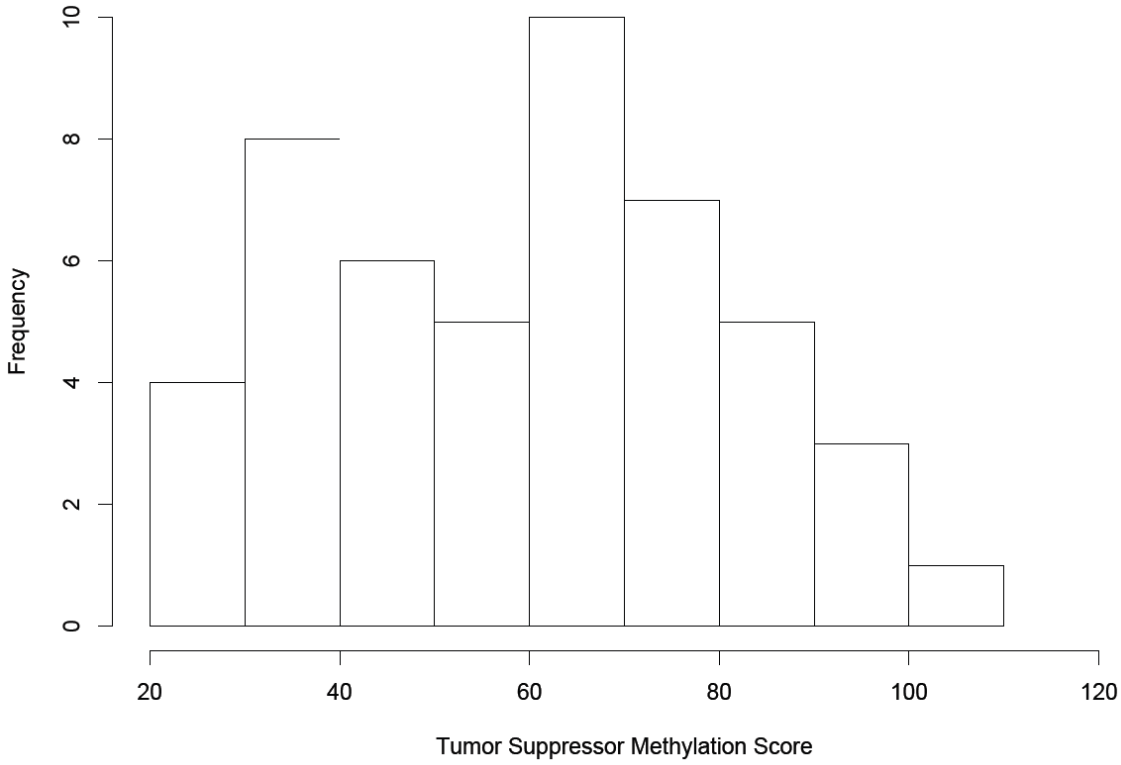

Table 3.1. Clinical characteristics of the study participants (n=49)

Patient Characteristic		N (%)	Mean (SD), Median (range)
Age			57.3(8.6), 55.0 (41-75)
Gender	Male	39 (80%)	
	Female	10 (20%)	
Stage	1	1 (2%)	
	2	6 (12%)	
	3	12 (24%)	
	4	30 (61%)	
Cancer Site of first Primary	Oral Cavity	9 (18%)	
	Oropharynx	26 (53%)	
	Hypopharynx	2 (4%)	
	Larynx	10 (20%)	
	Other	2 (4%)	
Tumor Tissue HPV (+) Status		21 (43%)	
Smoking Status	Never	9 (18%)	
	Past	32 (65%)	
	Current	8 (16%)	
Alcohol Problem	AUDIT >= 8 and drank within 1 year.	16 (32%)	

Table 3.2. Summary of mean daily micronutrient and food group servings by quartile of consumption

Dietary Component	Quartile 1 n = 12	Quartile 2 n = 12	Quartile 3 n = 13	Quartile 4 n=12
Micronutrients (energy-adjusted)				
Total Folate (µg/d)	325	441	706	1294
Food Folate (µg/d)	202	269	330	422
Vitamin A (IU/d)	3869	6705	10594	222111
Vitamin B6 (mg/d)	1.8	2.5	3.8	10.5
Vitamin B12 (µg/d)	4.7	8.5	13.1	32.2
Vitamin C mg/day	64	114	243	1086
Vitamin D (IU/d)	77	183	462	835
Vitamin E (mg/d)	5.4	8.9	24.6	288
β-carotene (µg/d)	1022	1913	3353	9178
Methionine (g/d)	1.4	1.8	2.3	2.8
Food Groups (servings/d)				
Cruciferous Vegetables	0.02	0.1	0.22	0.62
Green leafy Vegetables	0.07	0.15	0.3	0.94
Dark Yellow & Orange Vegetables	0.03	0.09	0.19	0.45
Refined Grains	0.32	0.67	1.17	2.9
Whole Grains	0.07	0.36	0.86	2.11
Red Meat	0.18	0.4	0.66	1.43
Legumes	0.07	0.27	0.27	0.71

Table 3.3. Results of unadjusted and adjusted linear regression showing associations between tumor suppressor methylation score and quartile of energy adjusted micronutrient intake or intake of food group, setting Quartile 1 (lowest intake) as the reference category

Micronutrient	Quartile 2 β (S.E)	Quartile 3 β (S.E)	Quartile 4 β (S.E)	P-Value for Trend
		Unadjusted		
Dietary Folate	-4.3 (8.6)	-7.0 (8.6)	-16.8 (8.4)	**0.047**
Total Folate	1.3 (8.7)	-1.0 (8.7)	-6.2 (8.7)	0.450
Vitamin A	-8.6 (8.6)	-7.9 (8.6)	-13.3 (8.6)	0.146
Vitamin B6	-7.1 (8.6)	-1.0 (8.6)	-11.4 (8.6)	0.301
Vitamin B12	-7.2 (8.0)	-6.6 (8.2)	**-23.9 (8.2)**	**0.008**
Vitamin C	0.8 (8.6)	2.0 (8.6)	-8.4 (8.6)	0.389
Vitamin D	-8.8 (8.6)	-11.2 (8.6)	-10.2 (8.6)	0.218
Vitamin E	-1.8 (8.8)	-3.9 (8.8)	-2.4 (8.8)	0.727
Beta Carotene	-3.3 (8.6)	-7.5 (8.6)	-13.1 (8.6)	0.111
Methionine	1.5 (8.5)	11.7 (8.5)	-4.4 (8.5)	0.932
Cruciferous Vegetables	-4.7 (8.0)	**-16.5 (8.0)**	**-18.1 (8.5)**	**0.014**
Green Leafy Vegetables	0.7 (8.5)	-6.1 (8.9)	-6.6 (8.9)	0.343
Dark Yellow Vegetables	-5.3 (9.3)	-5. 3 (9.0)	-10.1 (8.4)	0.245
Refined Grains	12.3 (8.5)	9.3 (8.8)	14.3 (8.7)	0.150
Whole Grains	9.1 (8.7)	8.0 (9.3)	-1.5 (8.8)	0.808
Red Meat	**-17.8 (8.2)**	-11.3 (8.5)	-12.4 (8.0)	0.185
Legumes	-5.6 (8.3)	-5.4 (9.6)	-9.8 (9.0)	0.307
		Adjusted*		
Dietary Folate	-3.1 (8.9)	-4.0 (9.5)	**-18.7 (8.9)**	**0.043**
Total Folate	4.6 (9.1)	-3.2 (9.6)	-9.1 (9.2)	0.258
Vitamin A	-13.3 (9.8)	-15.1 (9.9)	**-21.1 (9.7)**	**0.040**
Vitamin B6	-5.5 (9.6)	-2.2 (9.3)	-12.9 (9.2)	0.219

Vitamin B12	-7.6 (8.8)	-11.0 (8.7)	**-27.6 (8.8)**	**0.003**
Vitamin C	0.2 (8.9)	-1.9 (9.4)	-16.6 (9.3)	0.089
Vitamin D	-11.4 (9.3)	-15.6 (9.5)	-18.1 (9.8)	0.059
Vitamin E	-4.3 (9.2)	-11.4 (9.8)	-12.2 (9.9)	0.175
Beta Carotene	-3.4 (9.2)	-12.5 (9.0)	-15.6 (9.4)	0.055
Methionine	3.9 (9.1)	9.6 (9.1)	-7.9 (9.7)	0.687
Cruciferous Vegetables	-6.8 (8.5)	-16.2 (8.5)	**-20.4 (9.2)**	**0.013**
Green Leafy Vegetables	4.8 (9.0)	-8.7 (9.4)	-10.9 (9.5)	0.183
Dark Yellow Vegetables	-6.8 (9.6)	-15.4 (10.8)	-15.4 (9.8)	0.098
Refined Grains	15.3 (9.7)	13.5 (10.3)	**22.6 (10.7)**	0.069
Whole Grains	10.2 (9.7)	0.8 (10.7)	-3.8 (10.9)	0.542
Red Meat	-14.8 (10.3)	-12.8 (10.5)	-14.9 (11.8)	0.189
Legumes	-6.2 (8.9)	-4.5 (10.6)	-14.2 (10.6)	0.227

* Adjusted for age, sex, smoking status, HPV status, and problem drinking. Food group analyses also adjusted for total calorie consumption.
Bold text indicates $p<0.05$

Table 3.4. CpG sites with tumor DNA methylation values significantly different between highest (4th quartile) and lowest (1st quartile) of vitamin B12 and vitamin A intake, stratified by HPV status.

Gene Symbol	Chromosome	CpG Coordinate	Distance to TSS	T-Value	P-Value	Adjusted P-Value
HPV Positive Cases - Methylation by Vitamin B12 Intake						
RARB	3	25444872	114	-5.72	3.18E-05	0.0479
HPV Negative Cases - Methylation by Vitamin B12 Intake						
C4B	6	32057984	171	-5.39	1.79E-05	0.0178
GML	8	14391293 8	-281	-5.23	2.64E-05	0.0178
PTHR1	3	46893982	-258	-5.11	3.55E-05	0.0178
MLF1	3	15977192 1	243	-4.78	8.14E-05	0.0306
THPO	3	18557814 3	483	-4.58	0.0001	0.0341
CHGA	14	92459297	52	4.57	0.0001	0.0341
XIST	X	72989344	-31	-4.44	0.0002	0.0350
LCN2	9	12995139 8	-141	-4.41	0.0002	0.0350
SRC	20	35406338	-164	-4.40	0.0002	0.0350
SFN	1	27062338	118	-4.26	0.0003	0.0442
HPV Negative Cases - Methylation by Vitamin A Intake						
PTHLH	12	28016940	-757	-6.55	1.11E-06	0.0017
LMO2	11	33870264	148	-5.80	6.64E-06	0.0050
SIN3B	19	16800704	-514	-5.26	2.48E-05	0.0092
TNFSF8	9	11673277 5	-184	-5.19	2.91E-05	0.0092
C4B	6	32057984	171	-5.17	3.06E-05	0.0092
GML	8	14391293 8	-281	-4.66	0.0001	0.0241
HGF	7	81238681	-1293	-4.65	0.0001	0.0241
MLF1	3	15977192 1	243	-4.59	0.0001	0.0244
THPO	3	18557814 3	483	-4.51	0.0002	0.0244
HOXA11	7	27191320	35	-4.44	0.0002	0.0244

Table 3.5. Significantly enriched gene sets for CpG sites identified as differentially methylated in tumors, stratified by HPV status, from individuals with the highest compared to the lowest quartile of vitamin A and vitamin B12 intake.

Name	Size	Enrichment Score (ES)	Normalized Enrichment Score (NES)	Nominal P-Value	FDR Q-Value
POSITIONAL ENRICHMENT - HPV(+) by Vitamin B12					
Chromosome 12p12	5	-0.87	-1.68	0.003	0.122
KYOTO ENCYCLOPEDIA OF GENES AND GENOMES (KEGG) PATHWAYS - HPV (+) by Vitamin B12					
Type I Diabetes Mellitus	19	-0.63	-1.71	0.001	0.076
Antigen Processing and Presentation	9	-0.71	-1.62	0.016	0.164
Pathogenic Escherichia Coli Infection EHEC	8	0.67	1.87	0.017	0.158
Pathogenic Escherichia Coli Infection EPEC	8	0.67	1.86	0.009	0.081
Adherens Junction	23	0.43	1.71	0.01	0.124
Huntingtons Disease	7	0.61	1.66	0.033	0.121
Epithelial Cell Signaling in Helicobacter Pylori Infection	15	0.43	1.58	0.025	0.139
GENE ONTOLOGY - BIOLOGICAL PROCESSES - HPV(+) by Vitamin B12					
Regulation of Cytokine Production	8	-0.81	-1.82	0.000	0.080
Negative Regulation of Multicellular Organismal Process	6	-0.84	-1.72	0.001	0.233
Regulation of Multicellular Organismal Process	24	-0.60	-1.70	0.002	0.232
Positive Regulation of Cytokine Production	5	-0.87	-1.68	0.000	0.229
Cytokine Production	17	-0.63	-1.67	0.005	0.205
POSITIONAL ENRICHMENT - HPV(-) by Vitamin B12					
Chromosome 17q11	8	-0.68	-1.55	0.024	0.214
Chromosome 7q31	7	0.65	1.85	0.008	0.03
POSITIONAL ENRICHMENT - HPV(-) by Vitamin A					
Chromosome 12p12	5	-0.82	-1.67	0.004	0.192
Chromosome 8p21	8	-0.69	-1.59	0.020	0.218

CHAPTER 4. TRANSCRIPTOMIC EFFECTS OF CURCUMIN AND PIPERINE IN NORMAL HUMAN BREAST STEM CELLS

4.1 Abstract

Curcumin is a dietary polyphenol derived from the rhizomes of turmeric (*curcuma longa*) and a potential agent for both the prevention and treatment of cancers. Recently, curcumin treatment alone, or in combination with piperine, was shown to limit breast stem cell self-renewal while remaining non-toxic to normal differentiated cells. Here, we paired fluorescence activated cell sorting (FACS) with low input high throughput RNA sequencing (RNA-seq) to characterize the genome-wide changes induced specifically in normal breast stem cells following treatment with these compounds. We generated genome-wide maps of the transcriptional changes that occur in luminal (ALDH+) and basal (ALDH+/CD44+/CD24-) normal breast stem cells following treatment with curcumin and piperine. This transcriptome analyses confirm that these compounds target breast stem cell self-renewal in both stem cell populations by down-regulating expression of breast “stemness” genes. We also identified novel genes and pathways targeted by curcumin in these cells, including previously undescribed mechanisms by which curcumin may target Wnt signaling in breast stem cells, including downregulation of *SCD* and upregulation of *CACYBP*. These findings help clarify the mechanisms by which curcumin and piperine target breast stem cell self-renewal, providing insight into pathways involved in stem cell regulation in the normal human breast and novel targets for cancer chemoprevention and treatment efforts.

4.2 Introduction

While there has been a significant reduction in the number of deaths due to breast cancer since 1990, breast cancer remains the second most deadly cancer in U.S. women, with an estimated 39,840 deaths in 2010 (Jemal et al. 2010). The large number of newly diagnosed cases and deaths from breast cancer are indicative of the necessity for the development of novel strategies for the prevention of this deadly disease. The current strategies for prevention of breast cancer in susceptible populations are associated with toxicity or short and long term risks from surgery (Brandberg et al. 2008) or antiestrogen therapy (Fisher et al. 1998; Vogel et al. 2010). Antiestrogen therapy has also been shown to be only effective at preventing estrogen receptor positive (ER+) disease, and ineffective at prevention or treatment of estrogen receptor negative (ER-) disease (Howell 2008). There is a need, therefore, for the identification and development of novel cancer prevention strategies that are non-toxic, safe, and prevent both ER+ and ER- disease.

Curcumin is a dietary polyphenol derived from the rhizomes of turmeric (*curcuma longa*) which has been widely used in traditional Indian and Chinese medicine for treatment of a range of diseases, including inflammatory conditions, diabetes, and rheumatoid arthritis (Noorafshan and Ashkani-Esfahani 2013). Preclinical models have implicated curcumin as a potential agent for both the prevention and treatment of cancers. A major issue with the use of curcumin clinically is the limited bioavailability following ingestion. A number of strategies to increase the bioavailability of curcumin have been attempted, with the use of piperine as an adjuvant treatment showing up to a 20-fold increase in curcumin bioavailability in a clinical trial (Shoba et al. 1998). Recently, the mammosphere model, an *in vitro,* anchorage independent primary breast tissue culture method that enriches for a population of stem and early progenitor cells (Dontu et al. 2003), was used to show that the dietary compounds piperine and curcumin combined limit breast stem cell self-renewal while remaining non-toxic to normal differentiated cells (Kakarala et al. 2010).

Since breast tumors potentially arise from, and are sustained by, a population of progenitor or stem-like cells that harbor dysregulated self-renewal capacity, it is essential to characterize the effects of cancer preventive compounds specifically in

stem and progenitor cells (Al-Hajj et al. 2003; Molyneux et al. 2010). Evidence suggests that normal breast, as well as breast cancer, stem cells exist in two different states, luminal-like and basal-like (Liu et al. 2013; Van Keymeulen et al. 2011). Luminal stem cells have an epithelial phenotype, are proliferative, and express aldehyde dehydrogenase (ALDH+). Basal stem cells have a mesenchymal phenotype, are invasive, and are characterized by CD44+/CD24- surface expression (Liu et al. 2013). How these distinct populations of breast stem/progenitor cells in normal tissue respond to cancer preventive agents has not been comprehensively characterized.

The goal of this study was to extend previous findings that curcumin and piperine targeting normal breast stem cell self-renewal (Kakarala et al. 2010) by comprehensively characterizing the genomic changes that occur specifically in normal breast stem cells following treatment with these compounds. By pairing fluorescence activated cell sorting (FACS) with low input high throughput RNA sequencing (RNA-seq), we generated genome-wide maps of the transcriptional changes that occur in luminal (ALDH+) and basal (ALDH-/CD44+/CD24-) normal breast stem cells following treatment with curcumin and piperine. Our results confirm that these compounds target breast stem cell self-renewal in both stem cell populations by down-regulating expression of breast “stemness” genes. Additionally, we identify novel genes and pathways targeted by curcumin in these cells. These results help clarify the mechanisms by which curcumin and piperine target breast stem cell self-renewal and provide insight into pathways involved in stem cell regulation in the normal human breast.

4.3 Materials and Methods

Materials

Curcumin (98% pure) and piperine (BioPerine; 95% pure piperine) were donated by Sabinsa Corporation (Piscataway, NJ). Curcumin and piperine were diluted in DMSO to form a stock solution for dissolution in cell culture media. MCF7 cells were obtained from American Type Culture Collection (ATCC). SUM149 cells were obtained from Asterland.

Human Normal Breast Tissue Dissociation

Normal (non-pathogenic) breast tissue was isolated from women undergoing voluntary reduction mammoplasty at the University of Michigan hospital. The study protocol was approved by the University of Michigan Institutional Review Board. Breast tissue was mechanically and enzymatically digested as previously described (Dontu et al. 2003; Kakarala et al. 2010).

Mammosphere Formation

Single cells were plated in ultralow attachment plates (Corning) at a density of 100,000 viable cells/mL for primary cells and a density of 20,000 viable cells/mL for SUM149, MCF7, and T47D cells. Primary mammospheres were allowed to form for 7-10 days in serum-free mammary epithelial basal medium (MEBM) supplemented with 1 ug/mL hydrocortisone, 50ug/mL insulin, 20 ng/mL EGF, B27, 20 ug/mL gentamycin, and 1x antibiotic-antimytotic in the presence of either curcumin, piperine, or vehicle (DMSO) control. Each experiment was performed in triplicate and sphere number quantified manually. Previous work had established that 5uM curcumin, individually or in co-treatment with 5uM piperine, was sufficient to significantly inhibit primary and secondary mammosphere formation (Kakarala et al. 2010). To characterize interindividual variation in response to curcumin and piperine, we treated cells isolated from 13 mammoplasty reduction patients and quantified primary sphere formation.

Flow Cytometry and Curcumin Treated Sorted Cell Populations

Primary breast cells from 3 individuals were sorted on a MoFlo Astrios. Cells were first stained for a hematopoetic, fibroblast, and endothelial cell lineage depletion cocktail that consisted of biotinylated antibodies targeted against CD45, HLA-DR, CD14, CD31, CD41, CD19, CD235a, CD56, CD3, CD16, and CD140b (all from eBioscience, except for CD140b (Biolegend) and CD41 (Acris). Next, cells were stained with Alexafluor750-streptavidin, Alexafluor750 LIVE/DEAD Fixable Dead Cell Stain (Invitrogen), Brilliant Violet 421 – CD24 (Biolegend), APC-CD44 (BD), and Aldefluor (Stem Cell Technology). Single color and isotype controls were included for compensation and gating purposes. Aldefluor-positive gating was based on DEAB controls. Viability of cells post sorting was confirmed via trypan blue exclusion. Sorted cell populations were plated in mammosphere formation conditions as described above treated with either 5 µM curcumin, 5 µM piperine, 5 µM curcumin and 5 µM piperine,

or vehicle control. After 24 hours, total RNA was isolated from each treated cell population using the RNEasy Micro Kit (Qiagen) with on column DNase treatment.

High Throughput RNA Sequencing

RNA concentration and quality was determined using a Nanodrop (Thermo) and Bioanalyzer (Agilent). Due to the low level of input RNA due to a small number of input cells following FACS sorting and treatment, we depleted ribosomal RNAs with Ribominus (Life) and prepared sequencing libraries utilizing the SMARTer Stranded RNA-Seq kit (Clontech) following the manufacturer's recommended protocol. Libraries were multiplexed (4 per lane) and sequenced using paired end 50 cycle reads on a HiSeq 2500 (Illumina) at the University of Michigan DNA Sequencing Core Facility. Due to the high redundancy introduced at the beginning of the read by the SMARTer library preparation kit, one lane of PhiX control was included.

RNA-Seq Data Analysis

Raw sequencing read quality was assessed utilizing FastQC. Sequencing reads were concatenated by sample and read in pair using SeqTK. The first three nucleotides of the first read in each read pair were trimmed, as recommended by Clontech, using Prinseq 0.20.3. A splice junction aware build of the human genome (GRCh37) was built using the genomeGenerate function from STAR 2.3.0 (Dobin et al. 2013). Read pairs were aligned to the genome using STAR, using the options "outFilterMultimapNmax 10" and "sjdbScore 2". The aligned reads were assigned to genomic features (GRCh37 genes) using HTSeq-count, with the set mode "union". We conducted differential expression testing on the assigned read counts per gene utilizing edgeR (Robinson et al. 2010). Separate analysis were conducted for each stem/progenitor cell type (luminal and basal), adjusting for study subject as a covariate using the glmLRT. To reduce the dispersion of the dataset due to lowly expressed genes, genes with a mean count less than five across samples were excluded from analysis. Normalized counts per million were estimated utilizing the "cpm" function in edgeR (Robinson et al. 2010). Genes were considered differentially expressed between conditions at a false discovery rate adjusted p-value < 0.05 (Benjamini and Hochberg 1995).

Pathway analyses

Differentially expressed pathways were identified utilizing LRPath (Sartor et al. 2009; Kim et al. 2012). A directional analysis was conducted on all genes by including p-value of the differential expression test as a measure of effect size and log2 fold difference in expression as a measure of effect direction. Input biological pathways included in analysis were Gene Ontology Biological Processes, KEGG pathways, Panther pathways, miRBase, and Transcription Factors. Biological pathways were considered differentially expressed at a false discovery rate adjusted p-value <0.05.

4.4 Results

Curcumin and piperine inhibit mammosphere formation

To confirm and extend previous findings of the inhibitory effect of curcumin treatment on mammosphere formation (Kakarala et al. 2010), we exposed MCF7 cells, SUM149 cells, and primary human breast cells to curcumin and piperine *in vitro*. Curcumin inhibited mammosphere formation in both MCF7 and SUM149 in a dose dependent manner (**Figure 4.1**, panels A and B), as well as decreasing mammosphere size (representative images, **Figure 4.1** panel C). Piperine alone was not found to affect mammosphere formation. Treatment with curcumin and piperine in combination was found to decrease primary mammosphere formation at a greater rate than curcumin treatment alone. In cell lines assayed here, and based on previous results (Kakarala et al. 2010), 5 µM curcumin treatment was found to significantly inhibit mammosphere formation without inducing acute cytotoxicity, as assayed by trypan blue exclusion. We found significantly inhibited primary mammosphere formation in primary breast cells isolated from voluntary mammoplasty patients (n=13) treated with 5 µM curcumin, 5 µM piperine, and both agents simultaneously (**Figure 4.1** panel D).

Transcriptome-wide analysis of curcumin's effects in ALDH+ and ALDH-CD44+CD24- breast cells

Differential expression in ALDH+ and ALDH-/CD44+/CD24- cells

We isolated two distinct populations enriched for breast stem and progenitor cells, ALDH+ (epithelial-like) and ALDH-/CD44+/CD24- (mesenchymal-like), from normal reduction mammoplasty tissue via FACS (**Figure 4.2**). To characterize the baseline transcriptional differences in the two sorted normal breast cell populations, we compared expression between the vehicle control treated cell fractions after 24 hours of

culture. **Figure 4.3** presents a multidimensional scaling plot based on the 500 most differentially expressed genes across the 6 samples (ALDH+ and ALDH-/CD44+/CD24- fractions from 3 subjects), with the two different cell fractions clearly clustering on the first dimension of the leading log fold change. A differential expression analysis identified 1369 genes upregulated in the ALDH-/CD44+/CD24- fraction and 1573 genes upregulated in the ALDH+ fraction (**Figure 4.4**). Amongst the 10 most differentially expressed genes, *TNFRSF11B*, *S100A9*, *PIGR*, *SERPINB7*, *KRT23*, *KRT80*, *TCN1*, and *EHF* were overexpressed in ALDH+ cells, while *CCND2* and *TP63* were overexpressed in ALDH-/CD44+/CD24- cells. Epithelial-mesenchymal transition (EMT) associated genes, such as *BMP4* (FDR = 5.7E-4), *FGF2* (FDR=9.2E-19), *IGFBP4* (FDR=1.6E-3), *SERPINE1* (FDR=1.1E-8), and *SNAI2* (FDR=3.6E-4) were significantly upregulated in ALDH-/CD44+/CD24- cells. Conversely, epithelial-phenotype associated genes, including *CDH1* (FDR=2.7E-17), *EPCAM* (FDR=1.0E-7), *CLDN1* (FDR=3.1E-23), *CLDN3* (FDR=4.4E-20), and *KRT18* (FDR=1.3E-27) were overexpressed in the ALDH+ cells. Pathway analyses identified that previously identified biological processes involved in cell adhesion, mesenchymal cell proliferation, epithelial bud morphogenesis, and Notch signaling were differentially expressed between the two cell types (**Table 4.1**). Other significantly enriched pathways not typically associated with cell stemness or epithelial/mesenchymal phenotype included transition metal ion transport (FDR=8.4E-3), defense response (FDR=1.4E-3), and response to organic cyclic substance (FDR= 3.3E-3).

Transcriptomic changes in ALDH+ and *ALDH-/CD44+/CD24- cells induced by curcumin*

To characterize the effects of curcumin in ALDH+ cells, we compared the transcriptional profiles of the sorted ALDH+ cells cultured for 24 hours with 5 μM curcumin or DMSO. Unlike when comparing the expression profiles of the ALDH+ and ALDH-/CD44+/CD24- cells, there was no clear separation by treatment using MDS for 2 of the 3 samples tested (**Figure 4.5**). One hundred and ninety genes were identified as differentially expressed with curcumin treatment (FDR p-value < 0.05), with 97 genes upregulated and 93 genes downregulated (**Figure 4.6**). The most significantly upregulated genes with curcumin treatment included *HMOX1*, *SRXN1*, *HSPA1A*, *HSPA7*, *HSPA6*, *HSPA1B*, and *UBB,* while *KRT15*, *KRT6A*, and *SCD* were

downregulated. The list of all of the genes identified as differentially expressed by curcumin treatment in the ALDH+ cells, along with the estimated counts per million in each of the curcumin or DMSO treated samples, is shown in **Table 4.2**. LRPath pathway enrichment analysis revealed that curcumin treatment changed expression of genes in pathways involved in response to unfolded protein, protein ligase activity, and development (**Table 4.3**). Additionally, genes involved in sterol metabolism (FDR=5.6E-5), response to calcium ion (FDR=5.6E-5), cell adhesion (FDR=1E-6), cell junction organization (FDR=9.7E-7), and biosynthesis of unsaturated fatty acids (FDR=1.3E-4) were also differentially expressed with curcumin treatment.

In ALDH-/CD44+/CD24- cells, curcumin treatment was found to induce a similar pattern of overall change as in the ALDH+ cells, with 2 of the 3 samples having no clear separation when visualized utilizing MDS **(Figure 4.7**). Interestingly, in contrast to ALDH+ cells, approximately two times more genes were significantly upregulated with curcumin treatment (164) than downregulated (83) in ALDH-/CD44+/CD24- cells (**Figure 4.8; Table 4.4**). The most upregulated genes following curcumin treatment included *HSPA6*, *UBC*, *HSPA7*, *DNAJB1*, *HSPA1B*, *BAG3, HSPA1A*, *HMOX1*, *FBXL14*, and *TAOK3*. Biological pathways involved in response to protein stimulus, protein folding, ribosome biogenesis, non-coding RNA metabolism, and heat response were the most enriched for differentially expressed genes following curcumin treatment (**Table 4.5**). Additional pathways identified as enriched included MAPK signaling (FDR=2.3E-5), bone morphogenesis (FDR=1.2E-4), regulation of cytokine production involved in immune response (FDR=3.2E-4), endocrine system development (FDR=3.0E-4), and regulation of Notch signaling (FDR=1.8E-3).

Between the curcumin and DMSO treated ALDH+ and ALDH-/CD44+/CD24- cells, there were 46 genes differentially expressed in both cell populations. These genes were enriched in biological processes involved in unfolded protein response (FDR=2.1E-16), response to protein stimulus (FDR=1.9E-14), response to organic substance (FDR=5.5E-8), and response to heat (FDR=6.1E-5). This overlap group included a number of genes previously reported as differentially expressed following curcumin treatment, including *HMOX1*, *MIR22HG*, *IGFBP3*, *SOD1*, and *FOSB*. *CACYBP* (calcyclin binding protein), a gene involved in calcium homeostasis was

upregulated in both cell populations, while *S100A6* (calcyclin) was downregulated in the curcumin treated ALDH-CD44+CD24- population, while *S100A8* (S100 calcium binding protein A8) was upregulated in the curcumin treated ALDH+ cells. A number of genes associated with normal or breast cancer "stemness" were also significantly downregulated in the curcumin treated cells. In the ALDH+ cell fraction, *ALDH1A3*, *VIM*, and *PROM1* were downregulated with curcumin treatment, while *TP63*, *ITGA6 (CD49f), NFKB1*, and *JAG1* were downregulated in the ALDH-/CD44+/CD24- cells following curcumin treatment.

Transcriptomic changes in ALDH+ and *ALDH-/CD44+/CD24- cells by piperine treatment*

Despite a significant reduction in primary mammosphere formation observed in the normal breast cells following 5 µM piperine treatment (**Figure 4.1**), no genes were identified as differentially expressed (FDR < 0.05) in the piperine treated ALDH+ or ALDH-CD44+CD24- cells. LRPath analyses identified that in ALDH+ cells, the most enriched pathways in differentially expressed genes with piperine treatment involved DNA repair, transcription, development, and RNA metabolic processes (**Table 4.6**). Additionally, genes targeted by the micro-RNA mir-543 (FDR=0.025), or involved in non-homologous end-joining (FDR=0.008) or MAPK signaling (FDR=0.024) were enriched following piperine treatment. In ALDH-CD44+CD24- cells, pathways involved in translational elongation, non-coding RNA processing, response to amino acid stimulus, and energy generation were enriched following piperine treatment (**Table 4.6**).

Combined effects of curcumin and piperine in breast stem/progenitor cells

Treatment of the ALDH+ cells with curcumin and piperine in tandem revealed many of the same genes to be differentially expressed as with curcumin only treatment. A total of 191 genes were differentially expressed in ALDH+ cells, 82 upregulated and 109 downregulated, with curcumin and piperine co-treatment (**Figure 4.9; Tables 4.7 and 4.8)**. Most of the same pathways that were identified as differentially enriched in curcumin treated vs control cells were also enriched in curcumin and piperine co-treated vs DMSO cells (data not shown). To specifically identify differentially expressed genes and pathways induced by curcumin and piperine co-treatment, we compared the genome-wide expression levels in curcumin and piperine treated ALDH+ or ALDH-CD44+CD24- cells compared to curcumin treatment alone. A number of genes were

identified as differentially expressed in the curcumin and piperine cotreated ALDH+ cells (**Table 4.9**), including the ribosomal protein *RPL26*, ubquitin B (*UBB*), and the calcium homeostasis regulator *CALHM2*. In ALDH-CD44+CD24- cells, all of the significantly differentially expressed genes identified in the curcumin and piperine cotreated cells compared to curcumin only treated cells were genes involved in heat shock response (**Table 4.9**). The addition of piperine to the curcumin treatment affected pathways involved in translation, ribosomal biogenesis, and oxidative phosphorylation in the ALDH+ cell (**Table 4.10**). In ALDH-CD44+CD24- cells, pathways involved in protein folding, calcium ion transport, cytokine signaling, and growth regulation were enriched in the curcumin and piperine co-treated cells compared to the curcumin only treated cells (**Table 4.10**).

4.5 Discussion

Curcumin has shown great promise as a cancer preventive compound in preclinical models, particularly when paired with piperine as an adjuvant to increase bioavailability. Here, we confirm and extend previous findings of curcumin targeting self-renewal of both normal and cancer stem cell populations. Here, we reported, for the first time, genome-wide expression profiles of FACS sorted ALDH+ and ALDH-CD44+CD24- normal breast cells treated with curcumin and piperine. These results highlight the efficacy of pairing flow cytometry sorting of normal stem and progenitor cell populations with low input high throughput RNA-sequencing methods to identify mechanisms of action of cancer preventive compounds in normal cells.

Mounting evidence from the past decade shows that breast tumors likely arise from, and are sustained by, a population of stem-like cells that harbor dysregulated self-renewal capacity (Al-Hajj et al. 2003). Additionally, it was recently shown that cancer stem cells exist in two distinct phenotypic states, an epithelial-like, proliferative state, and a mesenchymal-like state that is quiescent and invasive, with cancer stem cells being able to interconvert between the two states (Liu et al. 2013). These results are consistent with findings from lineage tracing studies in the mouse mammary gland, where distinct stem cell populations were identified that are typically lineage restricted to generating either a luminal or basal progeny (Van Keymeulen et al. 2011), but either can recapitulate a full mammary gland when transplanted into a cleared mammary fat

pad (Keller et al. 2012). Here, we show that the widely used breast stem cell markers ALDH+ and ALDH-/CD44+/CD24- identify distinct epithelial-like and basal-like cell populations, respectively, based on comprehensive gene expression analyses.

The findings that normal breast stem and breast cancer stem cells exist in multiple, interconvertable states has important implications for breast cancer prevention. It has been previously reported that ALDH+ cancer stem cells typically localize in the tumor interior, while CD44+/CD24- stem cells localize at the invasive edge of the tumor (Liu et al. 2013). Thus, ALDH+ stem cells may be responsible for maintaining the growth of the bulk of the tumor, while EMT-like CD44+/CD24- stem cells are responsible for invasion and metastasis. With respect to cancer prevention efforts, recent pathologic evidence suggests that population expansion of ALDH+ cells may be an important early step in carcinogenesis. In histologically normal tissue isolated from *BRCA1/BRCA2* mutation carriers, or women with a family history of breast cancer, increased numbers of ALDH positive cells were observed in the breast ductules compared to control patients (Isfoss et al. 2013). A comparison of benign breast biopsy tissues isolated from women who went on to develop breast cancer, or not, found increased ALDH-1 staining in both epithelial and stromal breast cells isolated from women who later went on to develop breast cancer (Kunju et al. 2011). Additionally, ALDH1A1 tumor staining was strongly associated with early recurrence and metastasis of breast cancer, regardless of ER, PR, or HER2 status (Zhong et al. 2013). CD44+/CD24- breast cancer cells were originally identified as the cells that contained the tumor initiating fraction in breast cancers (Al-Hajj et al. 2003). Further work identified that these cells have a basal phenotype and increased invasive capacity in tumors (Sheridan et al. 2006), and that these cells can be generated through an EMT in immortalized human mammary epithelial cells (Mani et al. 2008). The most efficacious cancer preventive compounds, therefore, should inhibit self-renewal in both of these stem cell phenotypes, particularly if the two phenotypes can interconvert. Here, we show that curcumin downregulated genes associated with breast “stemness”, including *ALDH1A3* and *PROM1* (*CD133*) in ALDH+ cells, and *TP63* and *ITGA6* (*CD49f*) in ALDH-/CD44+/CD24- cells. These results, paired with the functional results from the mammosphere formation assay, show that curcumin likely inhibits stem cell self-renewal of both of these stem cell populations.

In addition to downregulating genes involved in stem cell self-renewal, 24-hour curcumin treatment induced significant expression changes in a large number of genes and pathways in both ALDH+ and ALDH-/CD44+/CD24- cells that have been previously reported in other cell types. Induction of *HMOX1* by curcumin treatment, which we observed in both stem cell fractions following curcumin treatment, has been previously identified in a number of model systems, and is likely mediated through NRF2 induction (McNally et al. 2007; Motterlini et al. 2000; Yang et al. 2009). Pathways involved in unfolded protein response, ligase activity, and response to heat were upregulated following curcumin treatment, corroborating previous findings of curcumin activating a heat shock and proteasome response (Dunsmore et al. 2001; Shen et al. 2007). These results may also point to a novel mechanism by which curcumin could inhibit breast stem cell self-renewal in light of recent findings in hematopoetic stem cells. Specifically, misfolded protein accumulation was shown to strongly induce apoptosis in hematopoetic stem cells, while lineage committed progenitor cells had an adapted response that allowed for their survival (van Galen et al. 2014). These results point to a mechanism by which tissue-wide damage due to the accumulation of unfolded protein in stem cells is limited due to stem cells preferentially undergoing apoptosis in these stress conditions.

Curcumin treatment also modified Notch signaling in ALDH-/CD44+/CD24- cells, potentially through downregulation of *JAG1*, which has previously been reported in curcumin treated esophageal cancer cell lines (Subramaniam et al. 2012). Curcumin also decreased mammosphere formation as well as the CD44+/CD24- fraction of BT-474 cells, in addition to decreasing the number of tubulin microtentacles, which are involved in the reattachment of suspended cells (Charpentier et al. 2014). Microtentacle formation has been shown to be dependent on vimentin and tubulin (Whipple et al. 2008), both of which we identified as downregulated in ALDH+ curcumin treated cells, suggesting these processes may also be important in inhibition of mammosphere formation in normal breast cells.

Due to our cell type specific, genome-wide approach, I was able to identify a number of novel genes and pathways differentially expressed in normal breast stem cells following curcumin treatment. In both ALDH+ and ALDH-/CD44+/CD24- cells,

curcumin treatment significantly downregulated expression of *SCD* (stearoyl-CoA desaturase), an enzyme involved in the synthesis of monounsaturated fatty acids. *SCD* has recently been implicated in carcinogenesis, as fatty acid synthesis is essential for the formation of plasma membranes as well as regulating glycolysis (Igal 2010). Relevant to stem cell function, downregulation of *SCD* has been associated with inhibition of beta-catenin and Wnt signaling in MCF7 and MDA-MB231 cells (Mauvoisin et al. 2013) and has been identified as required for lung cancer initiating cell spheroid production (Noto et al. 2013). We also observed a significant increase in expression of *CACYBP* (calcyclin binding protein) in both stem cell fractions treated with curcumin, with a concurrent decrease in the expression of S100A6 (calcyclin) in ALDH-/CD44+/CD24- cells. *CACYBP* plays a role in the ubquitination and degradation of beta-catenin (Siah-1, SIP, and Ebi collaborate in a novel pathway for beta-catenin degradation linked to p53 responses), and *CACYBP* expression has recently been shown to be regulated by S100A6 (Ning et al. 2012). Through our whole genome transcriptomic approach, we identified novel targets of curcumin that likely play a role in the downregulation of Wnt signaling previously reported by us (Kakarala et al., 2010) and others (reviewed in Li and Zhang 2014). Understanding the roles that *SCD* and *CACYBP* play in the regulation of normal breast stem cell self-renewal and differentiation could provide novel insight into mechanisms of carcinogenesis and provide new avenues for targeting cancer stem cells.

Despite observing a modest, but statistically significant, decrease in primary mammosphere formation following 5 µM piperine treatment, we did not observe any genes as significantly differentially expressed in either normal breast stem cell fraction. Pathway analyses showed, however, that piperine treatment affected pathways associated with non-recombinational DNA repair, RNA translation, and ribosome biogenesis. Co-treatment of curcumin and piperine led to differential expression of many of the same genes identified as differentially expressed with curcumin treatment. Comparison of the curcumin-treated and curcumin and piperine co-treated cells identified differences in expression of pathways associated with translational elongation, protein folding, and calcium transport, among others. Calcium homeostasis has recently been implicated in EMT of breast cancer cells, where intracellular chelation of calcium

was found to inhibit the expression of the EMT markers vimentin, Twist, and N-cadherin (Davis et al. 2014).These findings, paired with the interesting findings about *CACYBP* and *S100A6* above, likely point to a role of calcium homeostasis in the regulation of normal breast stem cells that is modulated by both curcumin and piperine treatment.

A weakness of this study is the lack of comprehensive characterization of molecular markers for normal human breast stem cells, which still remain to be fully characterized. Thus, our sorted cell fractions are enriched for progenitor and stem populations, but are not comprised entirely of stem cells. Here, we utilized two commonly used markers to isolate live normal human breast stem cells, ALDH and CD44+/CD24-, which we have previously shown to isolate distinct populations of breast cancer stem cells. Recent research, however, shows that the ALDH+ fraction of breast cells may represent two distinct populations as well. Two major aldehyde dehydrogenase enzymes have activity detected by Aldefluor, ALDH1A3 and ALDH1A1. A recent study identified that expression of these two enzymes is exclusive in normal human mammary tissue, with ALDH1A3 positive cells localizing in larger ducts, while ALDH1A1 cells typically localizing in branching points (Honeth et al. 2014). In our study, expression of ALDH1A3, but not ALDH1A1, was identified as significantly different between ALDH+ and ALDH-CD44+CD24- cells, suggesting an enrichment of the ALDH1A3+ cellular population in these experiments. Future research will be necessary to better clarify the roles of both the ALDH1A1 and ALDH1A3 populations in breast development and the role of either of these luminal stem cell populations in breast carcinogenesis in order to target these specific breast stem cell populations for cancer prevention efforts.

4.6 Conclusion

The results from this study confirm our previous findings of the effects of curcumin and piperine on normal and cancer stem cell self-renewal. We extended these findings using an experimental paradigm that combines FACS sorting of live, normal human breast cells into stem-enriched fractions, treatment *in vitro*, and whole genome RNA sequencing from these sorted cell populations utilizing low input methods. The use of primary tissues provides novel information about stem cell regulation in the normal human breast that may not be available from studies utilizing cell lines or model

animals. This experimental technique can thus be applied to understand the effects of carcinogens or cancer preventive compounds in specific cell populations. Here, we identify the pathways and individual genes differentially expressed by curcumin treatment, pointing to novel mechanisms of inhibiting stem cell self-renewal and potential biomarkers of curcumin efficacy in clinical trials.

4.7 References

Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. 2003. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 100:3983-3988.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological):289-300.

Brandberg Y, Sandelin K, Erikson S, Jurell G, Liljegren A, Lindblom A, et al. 2008. Psychological reactions, quality of life, and body image after bilateral prophylactic mastectomy in women at high risk for breast cancer: A prospective 1-year follow-up study. J Clin Oncol 26:3943-3949.

Charpentier MS, Whipple RA, Vitolo MI, Boggs AE, Slovic J, Thompson KN, et al. 2014. Curcumin targets breast cancer stem-like cells with microtentacles that persist in mammospheres and promote reattachment. Cancer Res 74:1250-1260.

Davis FM, Azimi I, Faville RA, Peters AA, Jalink K, Putney JW, Jr., et al. 2014. Induction of epithelial-mesenchymal transition (EMT) in breast cancer cells is calcium signal dependent. Oncogene 33:2307-2316.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. 2013. Star: Ultrafast universal rna-seq aligner. Bioinformatics 29:15-21.

Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, et al. 2003. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev 17:1253-1270.

Dunsmore KE, Chen PG, Wong HR. 2001. Curcumin, a medicinal herbal compound capable of inducing the heat shock response. Crit Care Med 29:2199-2204.

Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. 1998. Tamoxifen for prevention of breast cancer: Report of the national surgical adjuvant breast and bowel project p-1 study. J Natl Cancer Inst 90:1371-1388.

Honeth G, Lombardi S, Ginestier C, Hur M, Marlow R, Buchupalli B, et al. 2014. Aldehyde dehydrogenase and estrogen receptor define a hierarchy of cellular differentiation in the normal human mammary epithelium. Breast Cancer Res 16:R52.

Howell A. 2008. The endocrine prevention of breast cancer. Best Pract Res Clin Endocrinol Metab 22:615-623.

Igal RA. 2010. Stearoyl-coa desaturase-1: A novel key player in the mechanisms of cell proliferation, programmed cell death and transformation to cancer. Carcinogenesis 31:1509-1515.

Isfoss BL, Holmqvist B, Jernstrom H, Alm P, Olsson H. 2013. Women with familial risk for breast cancer have an increased frequency of aldehyde dehydrogenase expressing cells in breast ductules. BMC Clin Pathol 13:28.

Jemal A, Siegel R, Xu J, Ward E. 2010. Cancer statistics, 2010. CA Cancer J Clin 60:277-300.

Kakarala M, Brenner DE, Korkaya H, Cheng C, Tazi K, Ginestier C, et al. 2010. Targeting breast stem cells with the cancer preventive compounds curcumin and piperine. Breast Cancer Res Treat 122:777-785.

Keller PJ, Arendt LM, Skibinski A, Logvinenko T, Klebba I, Dong S, et al. 2012. Defining the cellular precursors to human breast cancer. Proc Natl Acad Sci U S A 109:2772-2777.

Kim JH, Karnovsky A, Mahavisno V, Weymouth T, Pande M, Dolinoy DC, et al. 2012. Lrpath analysis reveals common pathways dysregulated via DNA methylation across cancer types. BMC Genomics 13:526.

Kunju LP, Cookingham C, Toy KA, Chen W, Sabel MS, Kleer CG. 2011. Ezh2 and aldh-1 mark breast epithelium at risk for breast cancer development. Mod Pathol 24:786-793.

Li Y, Zhang T. 2014. Targeting cancer stem cells by curcumin and clinical applications. Cancer Lett 346:197-205.

Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. 2013. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports 2:78-91.

Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 133:704-715.

Mauvoisin D, Charfi C, Lounis AM, Rassart E, Mounier C. 2013. Decreasing stearoyl-coa desaturase-1 expression inhibits beta-catenin signaling in breast cancer cells. Cancer Sci 104:36-42.

McNally SJ, Harrison EM, Ross JA, Garden OJ, Wigmore SJ. 2007. Curcumin induces heme oxygenase 1 through generation of reactive oxygen species, p38 activation and phosphatase inhibition. Int J Mol Med 19:165-172.

Molyneux G, Geyer FC, Magnay F-A, McCarthy A, Kendrick H, Natrajan R, et al. 2010. Brca1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell 7:403-417.

Motterlini R, Foresti R, Bassi R, Green CJ. 2000. Curcumin, an antioxidant and anti-inflammatory agent, induces heme oxygenase-1 and protects endothelial cells against oxidative stress. Free Radic Biol Med 28:1303-1312.

Ning X, Sun S, Zhang K, Liang J, Chuai Y, Li Y, et al. 2012. S100a6 protein negatively regulates cacybp/sip-mediated inhibition of gastric cancer cell proliferation and tumorigenesis. PLoS One 7:e30185.

Noorafshan A, Ashkani-Esfahani S. 2013. A review of therapeutic effects of curcumin. Curr Pharm Des 19:2032-2046.

Noto A, Raffa S, De Vitis C, Roscilli G, Malpicci D, Coluccia P, et al. 2013. Stearoyl-coa desaturase-1 is a key factor for lung cancer-initiating cells. Cell Death Dis 4:e947.

Robinson MD, McCarthy DJ, Smyth GK. 2010. Edger: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.

Sartor MA, Leikauf GD, Medvedovic M. 2009. Lrpath: A logistic regression approach for identifying enriched biological groups in gene expression data. Bioinformatics 25:211-217.

Shen SQ, Zhang Y, Xiang JJ, Xiong CL. 2007. Protective effect of curcumin against liver warm ischemia/reperfusion injury in rat model is associated with regulation of heat shock protein and antioxidant enzymes. World J Gastroenterol 13:1953-1961.

Sheridan C, Kishimoto H, Fuchs RK, Mehrotra S, Bhat-Nakshatri P, Turner CH, et al. 2006. Cd44+/cd24- breast cancer cells exhibit enhanced invasive properties: An early step necessary for metastasis. Breast Cancer Res 8:R59.

Shoba G, Joy D, Joseph T, Majeed M, Rajendran R, Srinivas PS. 1998. Influence of piperine on the pharmacokinetics of curcumin in animals and human volunteers. Planta Med 64:353-356.

Subramaniam D, Ponnurangam S, Ramamoorthy P, Standing D, Battafarano RJ, Anant S, et al. 2012. Curcumin induces cell death in esophageal cancer cells through modulating notch signaling. PLoS One 7:e30590.

Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, et al. 2011. Distinct stem cells contribute to mammary gland development and maintenance. Nature 479:189-193.

Vogel VG, Costantino JP, Wickerham DL, Cronin WM, Cecchini RS, Atkins JN, et al. 2010. Update of the national surgical adjuvant breast and bowel project study of tamoxifen and raloxifene (star) p-2 trial: Preventing breast cancer. Cancer Prev Res (Phila) 3:696-706.

Whipple RA, Balzer EM, Cho EH, Matrone MA, Yoon JR, Martin SS. 2008. Vimentin filaments support extension of tubulin-based microtentacles in detached breast tumor cells. Cancer Res 68:5678-5688.

Yang C, Zhang X, Fan H, Liu Y. 2009. Curcumin upregulates transcription factor nrf2, ho-1 expression and protects rat brains against focal ischemia. Brain Res 1282:133-141.

Zhong Y, Lin Y, Shen S, Zhou Y, Mao F, Guan J, et al. 2013. Expression of aldh1 in breast invasive ductal carcinoma: An independent predictor of early tumor relapse. Cancer Cell Int 13:60.

Figure 4.1. The effects of curcumin and piperine treatment on primary mammosphere number and size in cancer cell lines and normal breast cells. Curcumin and piperine were tested at multiple concentrations in cell lines (5C = 5 µM curcumin, 5P = 5 µM piperine, 5C5P = 5 µM curcumin and 5 µM piperine, for example). (A) and (B) The effects of curcumin and piperine on primary mammosphere formation in MCF7 and SUM149 cells, respectively (NT= Not treated). (C) The effects of curcumin on mammosphere size in MCF7 cells. (D) Curcumin and piperine significantly inhibited primary mammosphere formation in normal breast cells (N=13).

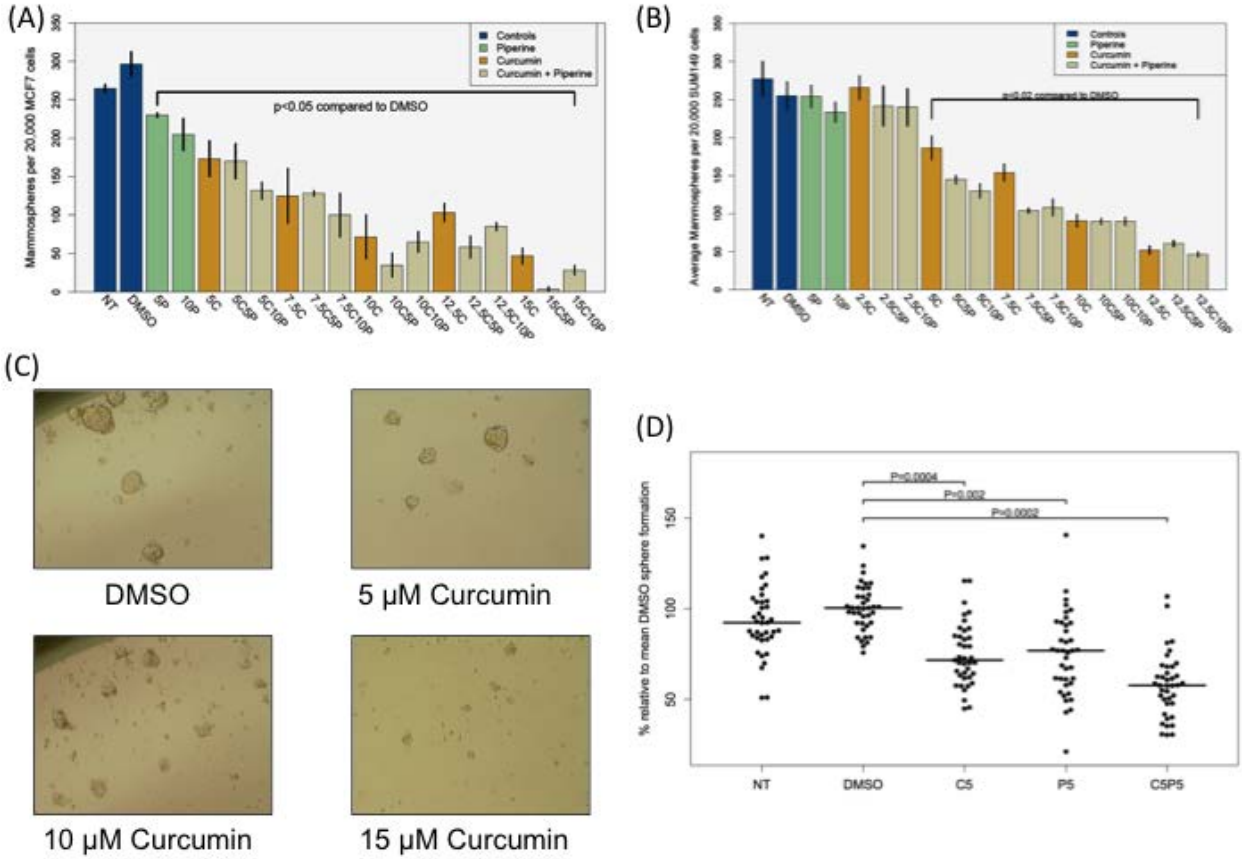

Figure 4.2. FACS gating scheme to isolate live, lineage negative, ALDH(+) and ALDH(-)/CD44(+)/CD24(-) normal breast cells.

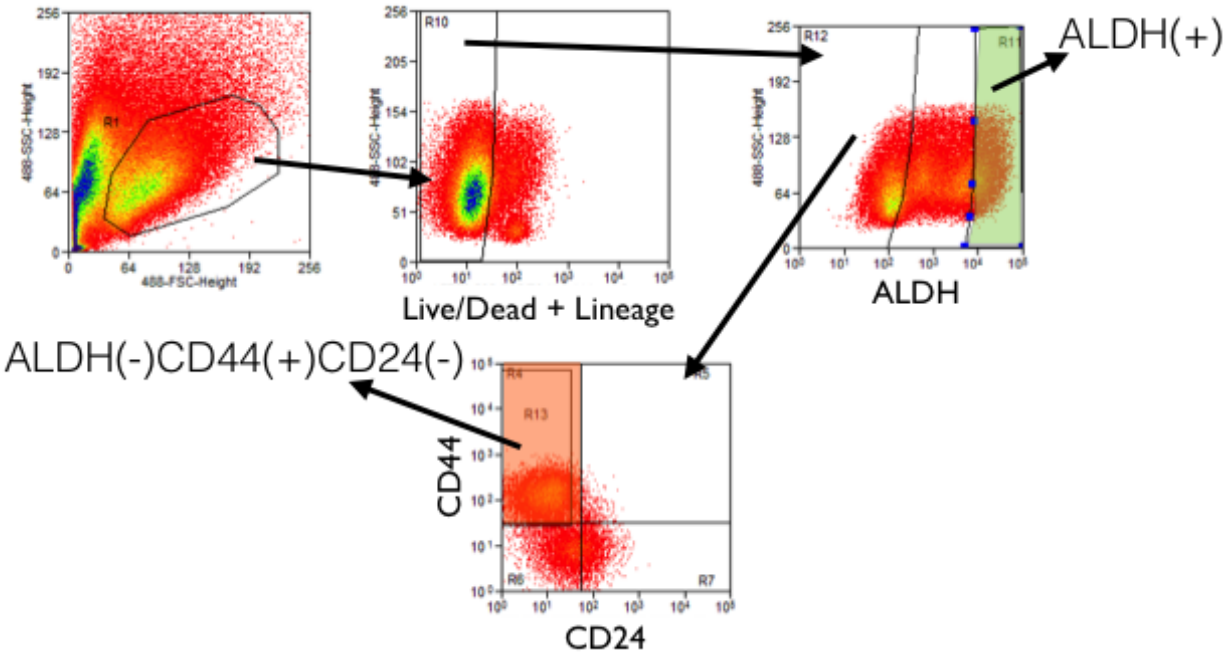

Figure 4.3. Multidimensional scaling plot of the vehicle control treated ALDH+ and ALDH-/CD44+/CD24- cells, based on the top 500 most variable genes.

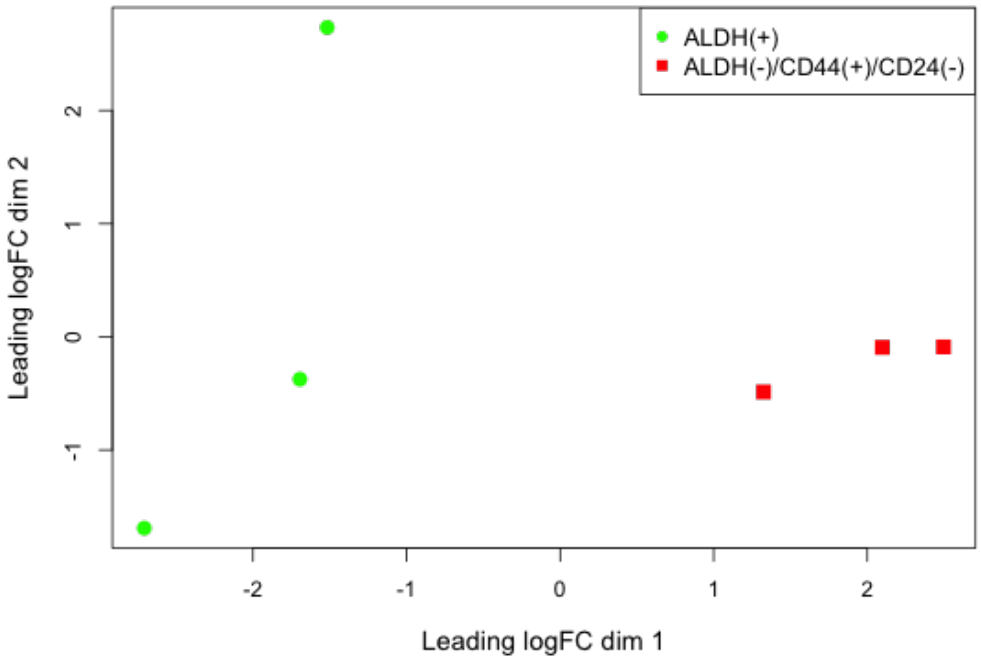

Figure 4.4. False discovery rate (FDR) volcano plot of the log(2) ratio of gene expression between the vehicle control treated ALDH(+) and ALDH(-)CD44(+)CD24(-) cells.

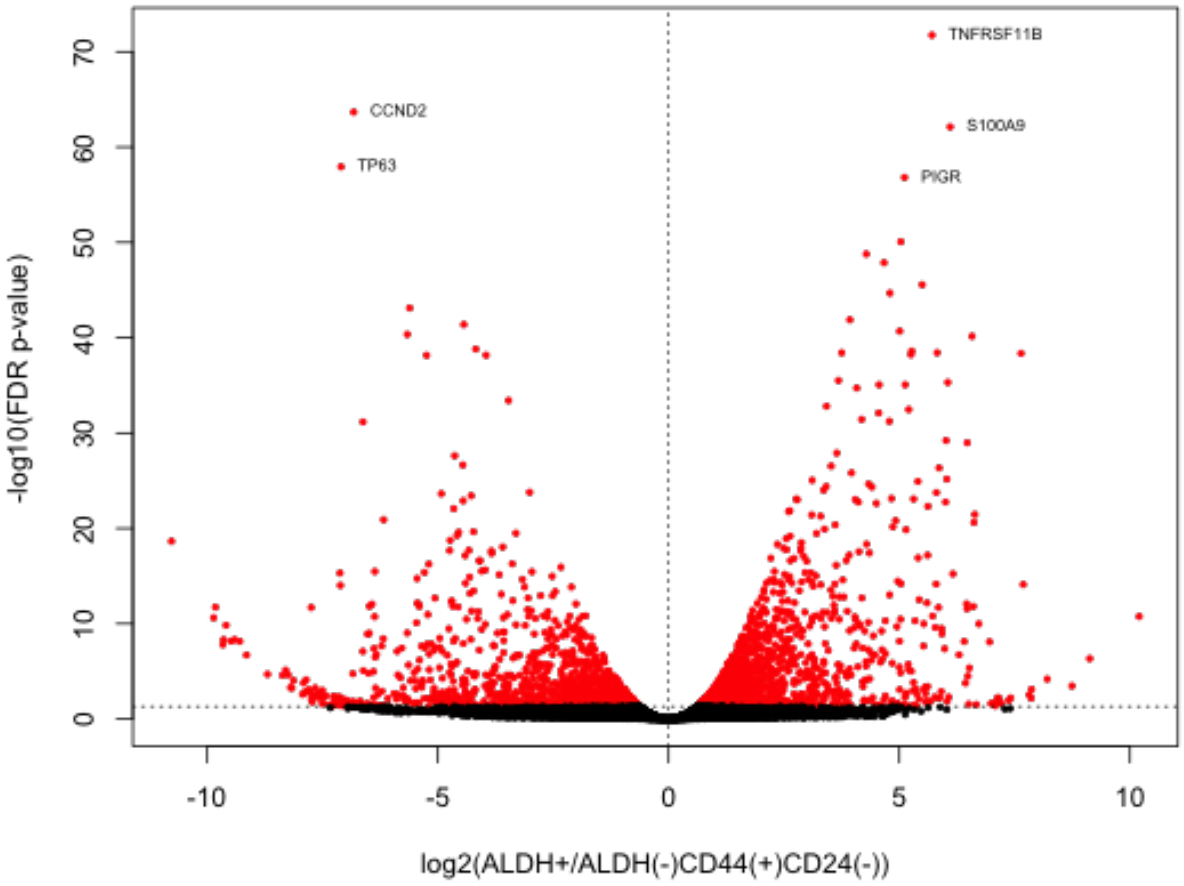

Figure 4.5. Multidimensional scaling plot 5 µM curcumin and DMSO treated ALDH+ cells, by the top 500 most differentially expressed genes.

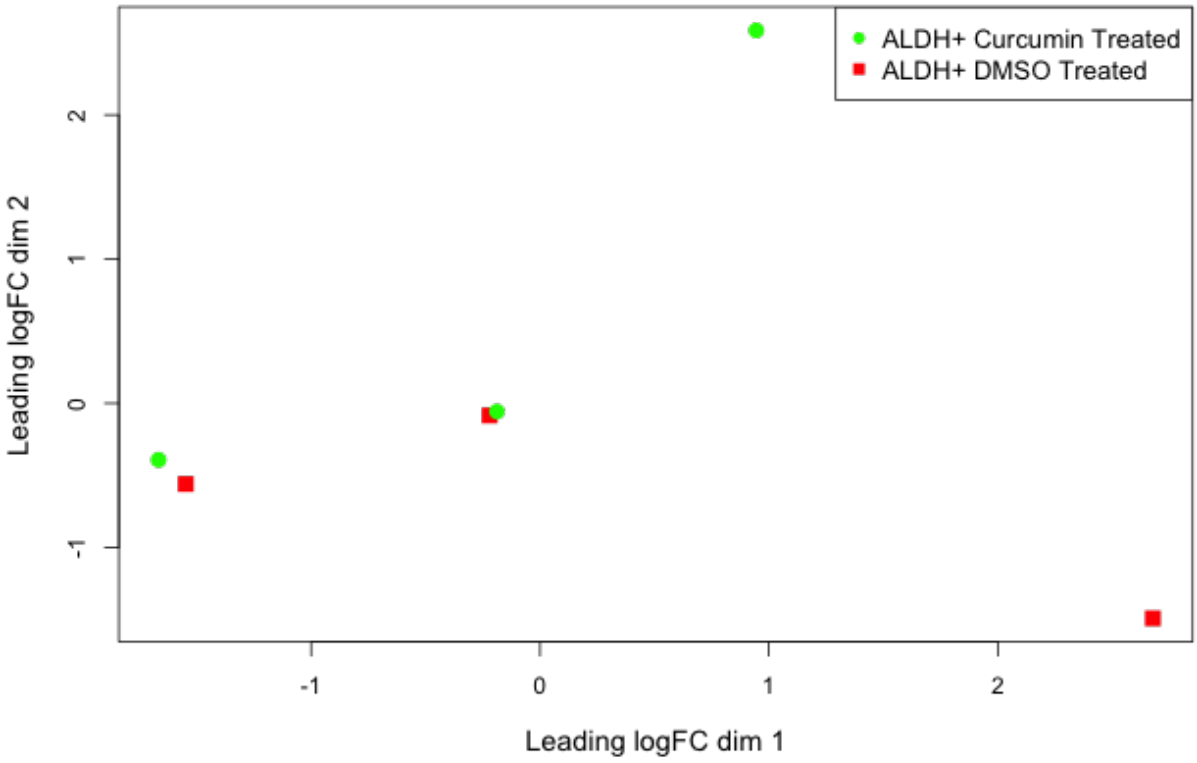

Figure 4.6. FDR volcano plot of the log(2) ratio of gene expression between the 5 µM curcumin and DMSO treated ALDH+ cells.

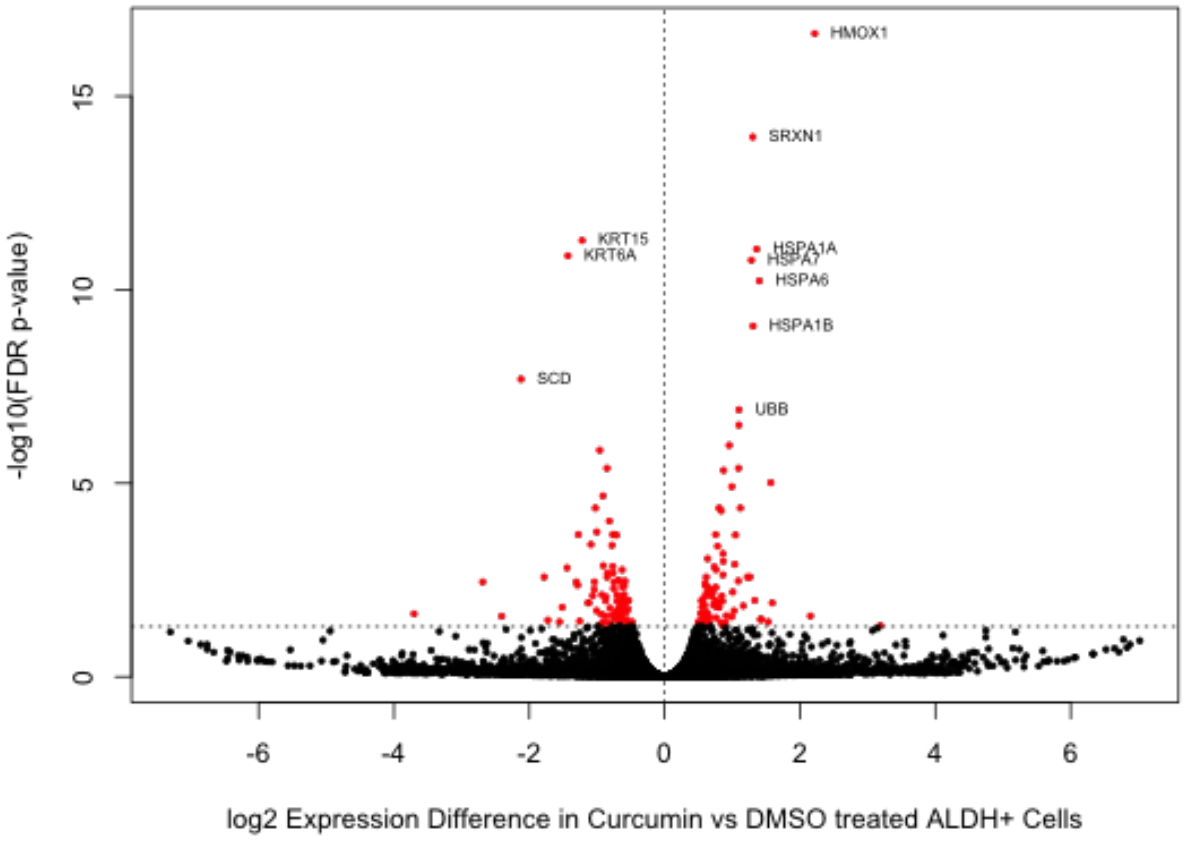

Figure 4.7. Multidimensional scaling plot of the 5 µM curcumin and DMSO treated ALDH-/CD44+/CD24- cells by the top 500 most differentially expressed genes.

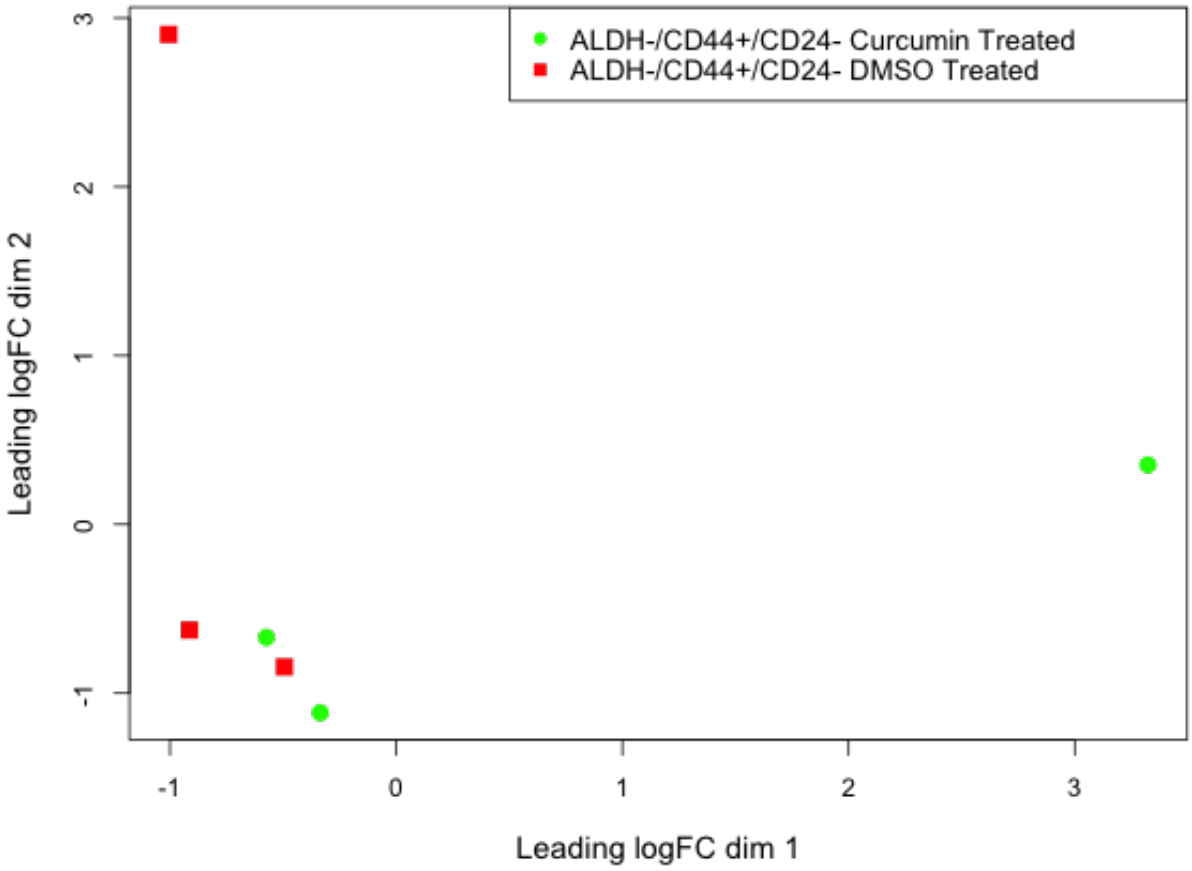

Figure 4.8. FDR volcano plot of the log(2) ratio of gene expression between the 5 µM curcumin and DMSO treated ALDH-/CD44+/CD24- cells.

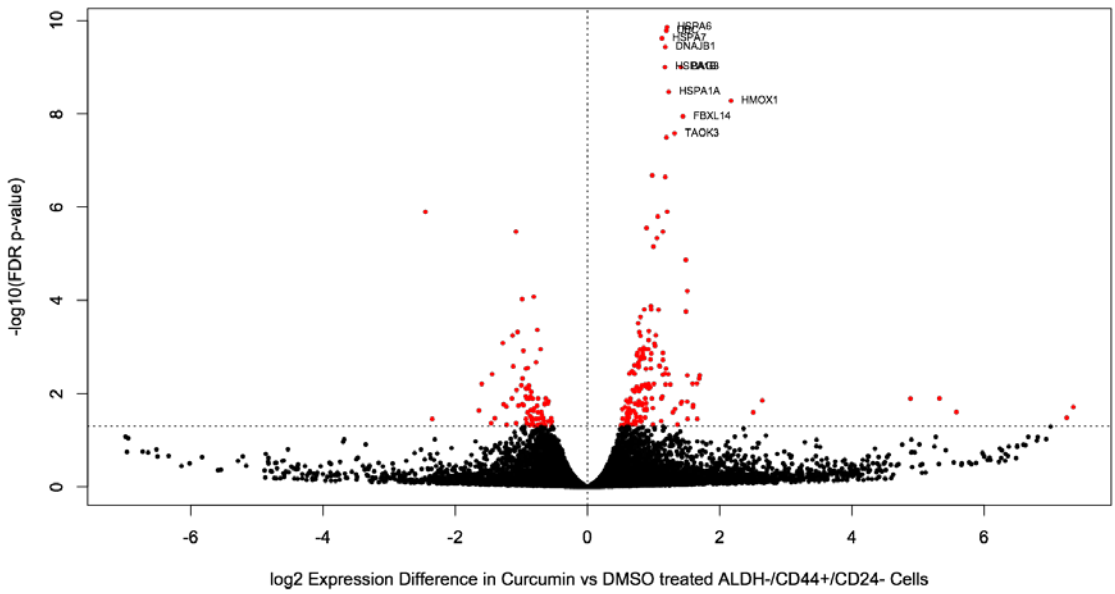

Figure 4.9. FDR volcano plot of the log(2) ratio of gene expression between the 5 µM curcumin and 5 µM piperine co-treated vs. DMSO treated ALDH+ cells.

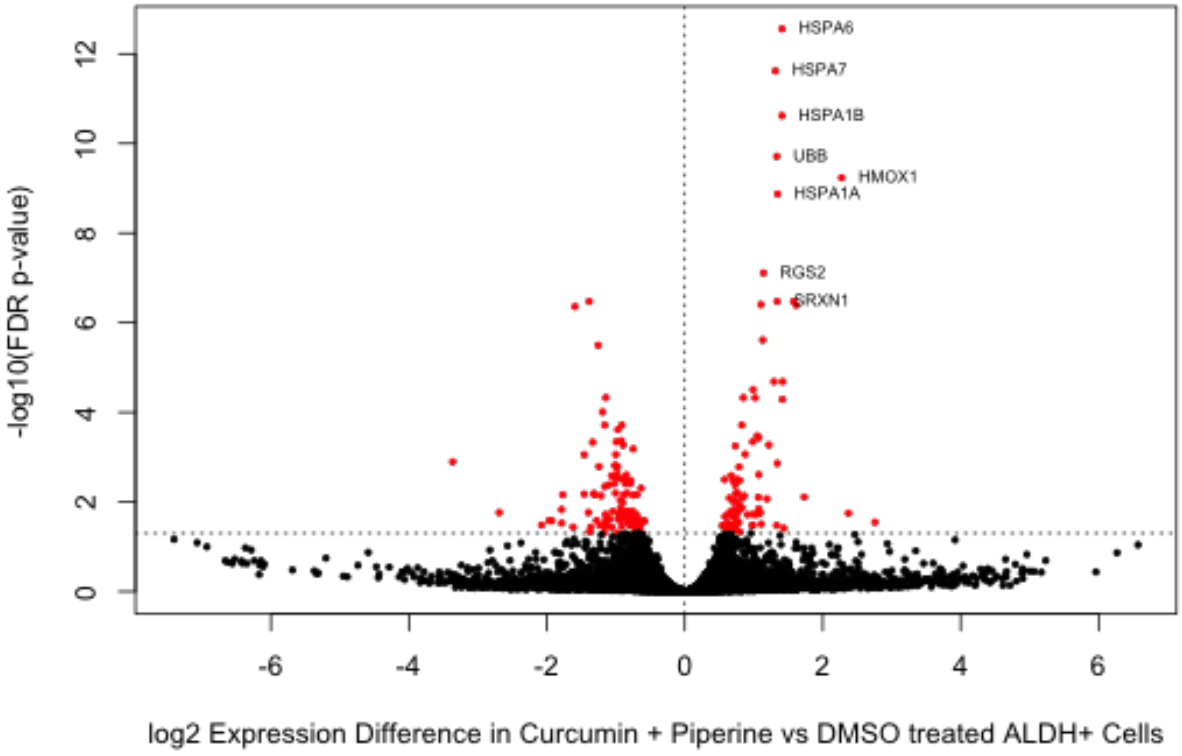

Table 4.1. The 10 most differentially expressed Gene Ontology biological processes identified between the vehicle control treated ALDH+ and ALDH-/CD44+/CD24- cells.

Name	#Genes	OddsRatio	P-Value	FDR
calcium-independent cell-cell adhesion	11	1.33	2.11E-10	5.71E-07
organ formation	21	0.77	1.18E-07	1.60E-04
development of primary female sexual characteristics	61	0.81	6.53E-07	4.73E-04
positive regulation of mesenchymal cell proliferation	20	0.78	8.86E-07	4.73E-04
female sex differentiation	66	0.81	9.43E-07	4.73E-04
morphogenesis of an epithelial bud	11	0.77	1.05E-06	4.73E-04
regulation of mesenchymal cell proliferation	22	0.79	1.59E-06	5.02E-04
regulation of Notch signaling pathway	11	0.77	1.60E-06	5.02E-04
pattern specification process	199	0.85	1.87E-06	5.02E-04
regulation of nervous system development	179	0.85	2.37E-06	5.02E-04

Table 4.2. The 190 genes identified as differentially expressed by curcumin treatment, with the estimated counts per million (CPMs) in each sample tested, in ALDH+ normal breast cells.

Gene	logFC	P Value	Sample 1 CPMs		Sample 2 CPMs		Sample 3 CPMs	
			Curcumin	DMSO	Curcumin	DMSO	Curcumin	DMSO
ABCA3	1.51	1.3E-04	5.31	20.98	10.23	24.05	8.66	22.7
ABHD3	-0.99	2.9E-04	75.91	30.31	41.12	30.42	41.51	17.54
ACAT2	0.91	3.7E-06	33.44	77.24	49.17	72.86	90.69	178.36
ACOT13	-1.26	9.5E-06	29.9	9.91	38.33	21.26	75.15	28.28
ACSL1	0.63	5.1E-05	74.85	141.94	249.07	314.75	315.74	498.39
ACTN1	0.60	1.9E-05	155.89	270.76	437.4	598.69	400.33	587.16
AGRN	1.04	1.4E-05	28.31	49.26	24.9	62.35	36.39	71.24
AHSA1	-0.57	9.8E-05	1651.28	891.56	572.62	440.96	356.66	260.48
AIF1L	0.92	4.4E-04	11.68	22.73	31.71	47.93	35.61	82.52
AKR1C1	-0.86	6.3E-04	126.34	136.11	240.29	113.46	545.7	185.16
ALDH1A3	0.59	1.3E-05	405.39	579.41	869.11	1219.8	2372.65	4008.56
ANKRD11	0.59	1.6E-04	103.16	173.12	183.06	290.15	103.67	133.23
ASNS	0.74	3.2E-04	50.96	91.81	68.29	89.74	40.52	81.57
ATP6V1A	-0.63	4.1E-05	225.08	130.28	150.93	105.1	184.92	123.85
AZI2	-0.65	5.2E-04	129.7	80.73	77.07	49.13	63.74	41.6
BAG3	-0.86	1.6E-06	894.83	390.55	290.6	234.65	203.21	95.71
C20orf194	1.73	3.7E-04	3.54	27.98	11.98	19.83	6.49	20.26
C5orf46	2.41	2.7E-04	0.18	8.74	9.09	17.12	4.13	21.75
CACYBP	-0.51	4.8E-04	1512.91	892.43	565.91	489.61	449.11	301.53
CALD1	0.62	8.6E-05	149.7	285.62	641.12	762.48	334.82	538.35
CBR3	-1.09	1.3E-05	121.92	69.95	34.81	18.71	36.98	11.96
CBS	1.78	9.0E-06	10.97	20.4	9.5	30.97	7.08	47.99
CCDC59	-0.61	2.7E-04	414.06	203.43	193.6	143.8	125.31	95.98
CD24	0.75	4.2E-07	216.94	447.67	827.27	1138.8	1189.97	2020.73
CD9	0.59	9.0E-05	148.64	285.62	385.85	535.87	633.44	822.49
CDH1	0.73	3.2E-04	122.09	344.5	525.41	652.52	651.15	878.09
CDK2AP1	0.70	6.4E-04	28.84	60.62	100.62	113.86	110.16	203.92
CDK6	0.64	2.1E-04	51.85	95.89	162.81	195.24	141.44	246.88
CEP85	-1.22	9.2E-06	94.31	26.23	33.57	19.67	48.59	22.98
CHORDC1	-0.68	2.7E-05	623.04	308.94	386.16	313.96	208.52	125.21
CLDN1	1.30	1.4E-05	13.27	77.82	85.95	150.81	67.08	110.12
CLDN7	0.66	2.3E-04	67.59	148.64	145.04	216.58	206.16	253
CLSTN1	0.76	6.6E-06	57.51	106.09	129.96	178.52	111.15	216.97
CLU	-0.60	1.6E-05	1475.93	842.88	919.73	688.27	1001.51	676.21
CNN2	0.87	6.7E-05	32.2	95.89	81.61	123.58	102.69	146.55
COPS4	-0.80	1.2E-04	96.61	48.38	64.15	41.72	62.16	35.48
CRYAB	-0.81	6.2E-08	1185.38	561.05	480.27	273.67	318.88	222.41
CSRP2	-0.85	7.2E-05	364.87	128.82	67.36	46.1	48.39	34.53
CTSL2	0.52	5.7E-04	100.15	153.01	282.44	349.15	128.66	204.6
CYFIP1	0.67	1.2E-04	44.94	88.89	95.76	145.71	122.16	168.17

CYP51A1	0.71	4.7E-07	92.72	169.92	192.98	281.79	504.59	837.44
DAG1	1.12	8.4E-05	12.39	57.42	50.83	89.34	85.97	121.81
DDAH1	0.94	2.2E-04	15.04	56.25	103.51	139.58	84	129.97
DDX42	0.71	6.9E-05	52.02	100.26	90.5	135.52	75.15	114.88
DHCR7	1.55	4.4E-04	2.65	15.45	19.32	23.25	13.97	58.05
DMD	-0.83	4.6E-04	82.63	33.81	57.75	39.17	58.62	36.98
DNAJA4	-0.80	1.2E-04	265.25	102.01	117.46	93	74.75	46.36
DNAJB1	-0.99	1.5E-08	4078.31	1401.6	1294.73	807.54	893.31	528.29
DNAJB4	-0.95	8.7E-10	1637.83	682.59	468.18	308.14	356.46	178.91
DNAJC10	0.66	2.6E-04	63.88	135.53	146.69	171.51	152.66	250.42
EDN1	0.57	1.9E-04	120.32	208.97	209.3	283.54	128.85	180.13
ELF5	1.13	8.7E-05	10.62	21.86	16.43	41.72	28.92	57.1
EMC4	-0.65	3.6E-04	99.27	75.49	109.71	68.56	123.34	67.57
EPHA4	-1.53	4.4E-04	39.46	8.74	16.12	10.03	14.36	4.08
ERBB3	0.71	1.6E-04	34.15	76.36	112.5	143.64	137.31	217.79
ESRP1	0.62	5.0E-04	40.88	64.41	76.96	122.54	90.89	131.05
FADS2	1.44	4.5E-06	2.65	10.49	16.53	33.92	29.31	84.56
FAM53C	-0.76	3.1E-04	122.27	56.25	64.67	45.86	68.66	43.1
FBXO30	-0.86	7.6E-06	197.47	140.19	180.68	111.55	131.02	49.08
FDFT1	0.54	1.5E-04	84.94	112.21	228.93	334.02	337.77	533.32
FOSB	-0.55	2.2E-04	478.29	337.8	137.91	102.63	120.2	72.46
FTL	-0.84	8.2E-08	678.95	498.09	522.93	268.97	1214.16	567.04
FXR1	-0.50	4.7E-04	461.84	293.49	259.3	202.8	253.57	181.49
GABARAPL1	-0.76	6.7E-05	242.6	214.8	223.97	117.13	247.48	110.8
GABRP	0.93	4.2E-05	17.69	27.98	81.3	118.48	72	209.22
GALNT3	0.59	1.6E-04	94.84	185.36	277.07	357.75	387.74	535.09
GARS	0.69	1.8E-04	86.53	204.6	176.34	253.2	159.93	202.83
GCLM	-0.80	7.8E-05	178.54	153.3	212.19	123.02	300	114.47
GLA	-1.01	3.3E-05	65.65	27.69	40.5	24.44	67.48	32.22
GLIPR1	0.85	6.7E-06	76.44	199.94	297.31	377.81	172.33	308.74
GLRX	-1.05	4.5E-07	38.75	11.37	84.09	55.42	351.74	175.92
GPRC5A	0.69	4.0E-04	25.83	51	45.56	71.1	112.92	158.65
GSR	-1.09	4.3E-09	115.9	67.62	120.25	44.59	182.56	88.37
GUCY1A3	0.78	2.5E-04	29.37	50.42	58.57	79.38	61.38	133.64
HMGA2	1.26	4.1E-04	12.21	24.19	16.12	30.73	8.07	30.04
HMOX1	-2.22	1.7E-21	1658.54	612.05	981.82	203.52	1418.56	184.48
HSP90AA1	-0.76	2.2E-05	10902.15	4549.6	5881.1	4902.4	5422.03	3238.82
HSP90AB1	-0.59	2.3E-04	24201.24	12006.74	8668.8	7316.5	7860.38	5500.05
HSPA1A	-1.36	2.5E-15	3483.76	1018.92	975.21	524.24	593.51	223.77
HSPA1B	-1.31	4.8E-13	5437.45	1475.92	1819.83	946.96	1172.85	546.37
HSPA1L	-1.57	1.1E-08	151.82	35.27	23.04	11.15	30.89	10.6
HSPA2	-0.70	4.9E-05	255.51	138.44	101.14	73.49	67.67	39.83
HSPA4	-0.59	3.9E-04	1646.15	818.99	638.84	563.66	756.98	509.26
HSPA4L	-0.87	2.7E-06	493.86	186.53	233.26	160.44	97.38	61.86
HSPA6	-1.40	2.9E-14	10180.37	2545.27	2828.72	1357.7	1871.41	847.5

HSPA7	-1.28	7.2E-15	3024.76	930.32	248.24	130.74	325.57	140.43
HSPB8	-0.90	4.7E-04	115.72	43.14	39.36	26.67	30.69	18.76
HSPD1	-0.71	3.2E-05	4368.33	1880.17	1378.72	1101.7	874.03	575.2
HSPH1	-0.73	4.1E-06	3030.24	1387.61	1277.79	974.67	703.67	439.25
HTATIP2	-0.76	5.1E-06	129.53	78.11	140.6	91.17	162.29	85.65
HYOU1	0.48	4.4E-04	583.4	939.94	725.62	947.36	547.28	708.02
IER5	-0.78	1.0E-06	2094.54	909.63	890.29	637.47	730.03	461.14
IGFBP3	1.04	2.6E-05	102.81	428.44	395.14	565.65	1678.23	2542.77
ITGB4	0.70	4.1E-05	26.01	34.68	67.98	116.65	175.08	314.72
ITGB8	0.67	1.3E-05	79.27	158.26	285.43	381.72	444.79	681.1
ITPR3	0.57	6.0E-04	62.64	96.18	132.33	195.24	70.62	101.28
KCNN4	0.75	2.5E-05	31.14	63.54	77.89	121.66	113.51	176.19
KDM2A	-0.50	5.0E-04	375.66	232.58	294.01	225.89	234.49	173.61
KIF1B	-0.55	1.3E-04	254.63	169.04	267.25	195.16	241.57	158.11
KLK10	0.65	1.3E-04	16.63	22.15	76.86	122.38	198.69	340.41
KLK7	1.00	3.1E-07	12.21	29.15	83.57	149.14	65.11	130.1
KRT15	1.22	1.1E-15	175.71	463.99	733.06	1337.4	1084.13	2855.18
KRT16	0.82	1.6E-07	139.44	186.82	360.54	704.83	403.08	826.84
KRT17	0.77	3.9E-06	482.36	616.13	1504.96	2528.9	420.98	966.05
KRT5	0.96	1.3E-09	209.15	434.56	503.82	752.84	265.38	631.34
KRT6A	1.43	4.6E-15	12.03	43.43	133.88	304.88	68.85	179.86
KRT7	0.66	7.9E-05	924.2	2008.99	3145.35	3772.7	5244.98	7956.76
LAYN	-3.19	6.1E-04	11.86	2.04	9.19	0.88	0.79	0
LDLR	0.63	5.3E-06	176.59	304.57	313.02	458	357.05	524.22
LOC344887	-1.59	8.7E-05	11.68	4.08	22.93	8.52	29.31	8.29
LPIN1	0.58	4.0E-04	69.89	121.54	166.94	201.61	157.77	255.58
LPP	1.06	4.5E-05	8.49	35.27	114.36	148.1	73.97	147.5
LRP2BP	-2.15	2.6E-04	31.5	4.95	6.3	2.63	5.31	0.82
MACC1	0.60	4.0E-05	149.34	266.1	564.98	873.23	375.74	475.68
MARS	0.53	5.3E-04	141.2	241.32	186.47	254.16	139.87	181.49
MET	0.56	3.5E-04	273.03	528.7	677.58	933.99	841.38	1013.63
MIR22HG	-0.51	6.2E-04	667.63	494.6	316.32	258.94	265.18	150.77
MRPL18	-0.61	9.6E-06	491.03	301.65	323.24	221.75	189.64	125.21
MSMO1	0.87	5.0E-05	25.13	73.45	84.09	113.15	133.57	218.6
MTHFD1L	0.87	2.6E-04	29.37	58.58	32.02	65.85	34.03	50.44
MTHFD2	0.85	4.3E-09	152.71	284.46	170.14	291.02	111.93	207.46
NGDN	-0.51	5.2E-04	475.46	293.2	206.4	157.18	163.08	121.54
NKTR	0.64	2.7E-04	68.48	138.44	126.45	188.23	98.75	126.57
NUDC	-0.66	2.9E-04	458.83	219.76	217.87	197.55	185.51	107.81
OSGIN1	-1.02	1.7E-04	59.1	24.19	44.32	26.04	32.46	15.77
PAIP2	-0.64	2.2E-05	216.76	147.77	180.68	119.99	184.92	107.4
PARP6	0.95	6.2E-04	20.88	59.46	31.61	49.45	20.85	34.8
PASK	3.70	2.1E-04	1.77	8.45	0.1	3.26	0.39	10.06
PDIA4	0.64	4.2E-04	206.32	484.98	529.86	714.46	617.31	745.81
PERP	0.55	7.1E-05	220.48	371.6	449.07	589.14	767.8	1099.68

PFN2	-0.91	2.5E-04	67.06	39.93	42.67	27.31	63.54	24.88
PLAA	-0.72	1.0E-04	93.43	68.49	110.54	65.05	121.97	64.03
POMP	-0.58	5.3E-05	403.97	262.89	300	231.55	366.69	217.52
PPP1R15A	-0.54	6.6E-05	5569.81	3395.73	3351.14	2634.6	2096.46	1410.87
PPP1R3C	-1.16	1.1E-04	58.22	21.57	36.05	18.55	19.08	8.84
PROM1	0.76	4.1E-07	82.28	149.81	206.82	293.89	278.75	527.34
PSAT1	1.28	3.8E-07	19.29	43.43	42.87	85.44	17.9	57.91
PSMA1	-0.60	2.9E-04	120.86	90.93	150.31	99.13	162.49	94.62
PSMA3	-0.63	3.9E-04	222.78	167.59	174.17	135.28	292.72	135.4
PSMA4	-0.60	2.0E-05	290.2	192.65	246.69	155.58	319.67	219.96
PSMC1	-0.59	1.1E-04	194.11	136.4	162.81	99.05	195.54	135.4
PSMC3	-0.54	2.1E-04	231.98	170.79	205.06	143.8	334.23	213.17
PSMD12	-0.69	3.3E-05	208.98	105.21	165.6	107.81	140.07	101.83
PSMD14	-0.63	1.9E-04	378.14	322.64	372.93	248.03	502.62	240.08
PXDN	0.81	5.6E-04	21.41	68.2	112.81	144.04	111.93	160.83
PYGB	0.67	2.6E-04	53.44	96.47	131.82	148.5	214.03	431.36
RGS2	-0.87	5.2E-09	634.71	319.72	168.18	87.98	93.64	58.87
RIT1	-0.60	1.9E-04	155.18	123.58	170.56	98.26	221.9	140.98
RMRP	-0.75	4.3E-07	1431.87	686.96	962.09	690.34	438.88	265.64
RPLP1	-0.62	2.0E-04	241.53	145.14	148.35	120.07	132.59	75.04
RPPH1	-0.62	3.6E-05	2316.43	1493.12	1275.83	1029.1	709.38	377.12
S100A8	-0.74	8.9E-05	272.32	133.19	45.04	31.45	38.75	24.47
SACS	0.57	8.0E-05	282.59	507.13	439.67	597.73	583.28	780.34
SCD	2.12	1.3E-11	9.91	112.79	127.48	289.35	225.84	835.67
SCRIB	0.81	1.3E-04	35.57	73.45	45.66	80.34	45.64	68.11
SDC1	1.09	8.5E-07	17.16	60.04	71.28	119.83	96.79	172.52
SELK	-0.56	4.0E-04	566.41	459.62	353.93	261.32	280.13	146.28
SEMA3E	-1.43	3.5E-04	50.96	12.82	19.94	14.65	20.66	5.44
SERPINH1	-0.64	6.0E-04	427.51	184.49	175.21	134.25	143.8	114.06
SLC1A5	0.69	1.2E-05	91.13	127.66	135.54	248.11	105.44	171.16
SLC38A1	0.75	1.5E-05	202.08	491.39	405.89	556.73	281.9	406.08
SLC43A3	0.61	5.0E-04	97.68	196.44	138.84	210.76	144	168.98
SLC44A2	0.84	4.9E-04	6.72	16.32	41.84	65.45	60.2	101.28
SLC6A14	0.62	2.3E-05	69.19	111.63	170.45	223.58	497.31	863.68
SLC6A9	2.69	1.4E-05	0	5.54	3.31	17.52	6.69	22.84
SLC7A5	0.78	9.2E-07	112.89	173.12	119.83	198.5	83.61	166.54
SLU7	-0.50	4.4E-04	526.07	338.96	252.58	204.79	281.51	189.92
SMARCA1	1.29	1.9E-05	12.03	45.76	16.43	39.33	34.03	56.83
SNRPF	-0.76	1.2E-04	115.72	68.78	60.74	36.07	68.46	40.24
SOD1	-0.64	2.2E-06	1213.34	842.3	764.88	507.84	1023.74	592.46
SORBS2	1.01	1.6E-04	10.97	25.07	24.59	44.11	32.66	66.34
SQLE	0.85	9.0E-06	23.53	44.3	50.21	75.32	102.69	214.66
SQSTM1	-0.67	3.3E-05	2126.92	1706.17	2167.46	1431.2	4024.32	1900.42
SRXN1	-1.30	1.6E-18	339.92	162.34	362.4	141.25	441.25	159.06
STC2	1.02	6.7E-08	57.33	101.43	69.11	128.19	49.57	127.52

TAF7	-0.55	4.2E-04	829.89	442.43	391.63	334.1	364.72	254.09
TGFB2	0.91	2.8E-08	96.79	216.55	134.19	263.32	110.56	167.76
TMEM123	0.62	7.4E-05	81.75	138.15	188.53	244.29	170.56	283.72
TMEM184B	0.97	6.5E-04	14.16	42.84	26.65	43.63	33.05	52.48
TNC	0.75	2.2E-04	233.04	255.9	126.24	221.99	83.02	204.74
TPRKB	-0.84	4.6E-05	91.66	40.22	81.1	46.74	90.1	60.5
TRIB2	-1.04	3.3E-06	78.03	32.64	48.76	25.4	77.9	40.92
TUBA1A	0.63	7.3E-05	111.65	157.68	190.7	253.44	245.11	479.35
TUBB	0.54	6.1E-04	76.62	132.61	128	161.8	175.48	252.05
UBB	-1.10	8.7E-11	10406.87	3492.78	2556.3	1550.9	2549.7	1277.37
UBC	-1.10	2.4E-10	5944.94	1927.68	2413.02	1280.1	1993.77	1181.52
VIM	0.53	6.4E-05	580.39	895.93	486.57	720.59	406.62	539.31
VTRNA1-3	-1.42	3.5E-04	70.6	19.53	16.84	12.58	12.79	2.99
ZFAND2A	-1.12	6.4E-08	380.26	118.04	112.81	76.2	95.8	44.32
ZNF222	-1.33	6.4E-05	92.72	55.96	61.67	34.32	38.16	6.25

Table 4.3. The 10 most enriched Gene Ontology biological processes in curcumin vs DMSO treated ALDH+ cells.

Name	#Genes	OddsRatio	P-Value	FDR
response to unfolded protein	61	0.26	4.97E-14	1.31E-10
response to protein stimulus	96	0.31	1.76E-13	2.33E-10
ectoderm development	143	2.71	2.15E-11	1.89E-08
protein folding	147	0.42	7.20E-11	4.22E-08
positive regulation of ligase activity	73	0.38	1.09E-10	4.22E-08
anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	62	0.37	1.16E-10	4.22E-08
negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	61	0.37	1.42E-10	4.22E-08
negative regulation of ligase activity	65	0.37	1.44E-10	4.22E-08
negative regulation of ubiquitin-protein ligase activity	65	0.37	1.44E-10	4.22E-08
regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	66	0.38	1.94E-10	4.57E-08

Table 4.4. The 247 genes identified as differentially expressed by curcumin treatment, with the estimated counts per million (CPMs) in each sample tested, in ALDH-/CD44+/CD24- normal breast cells.

Gene	logFC	P Value	Sample 1 CPMs		Sample 2 CPMs		Sample 3 CPMs	
			Curcumin	DMSO	Curcumin	DMSO	Curcumin	DMSO
ABCB1	-0.82	6.0E-06	451.14	188.47	260.01	190.3	115.16	68.56
ADSL	0.81	2.2E-04	53.09	152.5	76.77	116.41	189.84	240.75
AFG3L2	0.64	7.5E-04	134.59	305.78	171.64	228.55	209.58	262.07
AGXT	0.93	1.4E-05	43.93	127.08	43.97	69.61	118.48	182.33
AHSA1	-0.77	1.3E-05	1654.44	692.3	1057.16	760.12	763.82	507.33
AHSA2	-0.92	3.7E-05	92.78	52.2	50.9	22.33	50.61	29.77
AKAP1	0.75	7.0E-04	42.87	97.75	48.78	60.82	59.95	99.96
ALAS1	-0.83	3.8E-06	282.97	116.52	231.03	151.33	174.06	113.5
ALDH1A2	-1.51	2.3E-05	98.09	22.29	29.97	24.35	33.36	8.05
ANKRD1	-1.51	9.1E-08	94.5	21.7	37.61	13.78	35.66	17.99
ANO1	0.78	8.8E-06	68.62	130.8	69.99	116.89	88.66	140.92
ANXA1	-0.75	5.7E-06	2676.31	1207.86	3763.32	2571.63	1277.75	864.2
AP4B1	-1.14	2.1E-05	87.07	26.59	32.24	17.7	39.48	21.85
ARID5B	-0.67	1.7E-05	612.53	326.31	560.18	423.83	327.22	199.27
ARL4A	-0.71	5.3E-04	242.36	115.74	147.47	76.86	121.93	112
AXIN2	-0.97	1.0E-04	86.94	31.67	55.99	33.14	25.31	15.37
B4GALNT1	-1.69	2.7E-05	13.54	4.69	18.8	6.89	21.85	5.23
BAG3	-1.41	3.7E-13	983.51	299.52	563.29	321.68	575.82	176.7
BCL10	-0.81	5.5E-05	115.61	59.43	66.88	36.94	75.92	49.39
BEX1	-0.78	6.4E-05	127.95	56.5	182.81	127.34	40.77	25.84
BIRC3	0.65	1.6E-04	210.37	448.69	285.32	391.05	574.21	760.32
BPHL	2.35	4.5E-04	1.46	10.95	0.85	5.94	2.85	8.18
C1orf116	-0.81	5.6E-04	54.42	29.52	70.27	46.09	27.19	14.07
C4orf26	-7.35	2.0E-04	5.57	0	2.69	0	0.46	0
CABLES1	-0.96	4.8E-06	126.09	58.26	133.89	57.14	35.98	24.6

CACYBP	-0.93	9.8E-07	1983.47	699.34	1095.9	736.72	907.64	554.24
CARD10	0.84	4.1E-04	48.31	137.05	57.12	67	68.83	119.78
CASP4	-0.71	6.7E-05	268.24	123.17	193.7	146.7	164.16	106.64
CBR3	-1.09	1.1E-05	64.51	25.42	46.94	21.74	41.64	23.49
CCDC121	-1.22	2.0E-05	50.57	19.94	28.98	11.76	24.66	12.04
CCDC84	-0.89	2.4E-04	71.41	29.52	53.73	38.61	46.38	24.14
CCND2	0.86	8.7E-05	115.61	342.73	185.22	222.61	115.07	195.02
CD109	1.40	4.3E-04	4.51	31.87	15.41	30.41	26.59	40.76
CD83	-0.60	3.1E-04	218.34	145.07	198.51	112.25	135.41	103.69
CDC42EP3	-0.93	4.3E-05	131.13	42.43	121.17	80.66	108.91	71.57
CDK6	0.69	3.4E-04	57.74	119.85	72.39	110.47	82.27	110.56
CDKN2C	-1.51	4.6E-04	20.71	6.84	12.72	3.44	12.56	5.82
CDR2	-0.62	6.3E-04	145.07	84.66	114.24	81.25	75.64	50.57
CHORDC1	-1.14	4.0E-09	998.5	297.17	676.4	393.54	460.65	246.37
CITED2	-0.86	6.5E-06	196.57	82.31	145.21	97.64	86.41	51.22
CLU	-1.19	2.2E-11	1111.72	374.01	859.92	515.66	612.03	253.9
CNN3	-0.59	1.2E-04	620.23	351.72	384.72	278.91	273.3	196
COL2A1	-0.96	2.2E-07	142.81	67.26	115.37	71.51	110.52	51.55
CREM	-0.68	6.7E-04	118.26	55.13	122.44	94.67	88.52	58.94
CRYAB	-0.86	2.7E-07	1850.35	932.19	1431.13	1036.3	1341.48	617.57
CSRP2	-0.84	5.5E-06	595.94	231.09	499.52	355.53	240.22	151.91
CTGF	-0.76	4.5E-06	364.87	185.34	186.63	127.58	73.66	44.16
CYP1B1	2.45	1.2E-09	17.25	44.58	4.52	33.38	2.81	24.86
DAZAP1	0.84	4.8E-04	48.31	157.39	103.21	133.87	116.96	164.53
DEDD2	-0.78	1.1E-06	284.3	148.59	190.31	128.17	200.51	111.74
DKC1	0.72	6.7E-04	75.26	199.62	154.4	212.51	167.16	207.06
DNAJA1	-0.87	1.0E-04	3627.29	1118.71	2465.66	1855.1	1940.9	1344.26
DNAJA4	-0.89	3.0E-09	472.11	240.67	354.74	210.97	223.2	114.94
DNAJB1	-1.18	9.3E-14	5087.95	2006.32	2925.03	1588.54	3464.94	1402.68
DNAJB4	-1.18	1.9E-10	1880.61	582.81	817.93	475.98	876.63	420.78

DOK5	-0.73	5.5E-04	114.28	62.95	66.31	40.63	27.05	17.53
DST	0.60	1.5E-04	1534.06	2855.22	2672.93	3468.83	1952.68	2805.16
DUSP1	-0.71	2.0E-05	356.77	176.94	261	159.77	153.31	114.16
DUSP2	-0.80	7.1E-06	275.54	120.43	156.23	102.87	164.03	108.01
DUSP7	0.99	1.5E-07	32.92	68.62	49.91	103.82	81.62	148.18
EBNA1BP2	0.55	7.7E-04	189	354.07	364.92	475.86	399.92	513.16
EGR1	-0.63	1.9E-05	828.22	521.42	384.86	277.01	594.36	352.68
EGR2	-0.72	5.9E-04	134.98	62.56	71.12	45.02	61.79	46.38
EGR3	-0.83	1.5E-04	101.01	47.31	53.87	33.74	47.44	28.65
EHD3	0.82	4.7E-04	39.42	109.49	35.63	49.65	78.49	113.18
ENDOD1	0.87	3.6E-04	25.35	73.71	52.45	87.43	69.43	90.21
EPHB4	1.00	4.2E-05	28.93	80.16	30.12	61.06	41.92	60.91
FABP4	-1.51	1.3E-04	58.27	16.42	9.33	2.26	7.27	4.38
FAM110C	0.90	1.3E-05	28.54	44.97	27.85	55.95	76.24	155.24
FAM129B	0.60	5.5E-04	60.39	106.55	83.42	121.88	148.2	200.97
FAM210A	-0.83	1.7E-04	144.54	52.79	167.97	100.49	81.44	65.29
FAM210B	0.97	1.7E-04	17.39	32.45	15.98	32.79	30.23	59.01
FAM43A	-1.29	2.9E-04	39.15	13.49	20.36	11.52	10.72	3.73
FBXL14	-1.44	6.4E-12	160.47	43.21	119.19	47.16	85.21	39.19
FBXO30	-0.77	8.9E-06	264.52	178.11	226.93	150.74	169.82	76.41
FEM1C	-0.73	3.1E-04	111.89	59.63	95.15	67.83	45.96	26.56
FLNC	-0.73	6.5E-04	152.9	81.33	51.89	37.18	18.96	10.99
FOS	-0.70	1.9E-05	717.12	341.95	384.43	266.56	221.03	154.78
FOSB	-0.85	3.0E-06	503.17	200.98	232.02	165.59	186.99	111.54
GABARAPL1	-0.75	5.2E-06	326.24	162.66	218.59	153.12	169.46	102.58
GBP3	-1.25	3.9E-05	67.43	24.05	15.27	5.23	24.98	14.52
GFM2	-0.77	1.5E-04	280.19	107.53	219.01	150.03	96.25	73.34
GLA	-0.88	4.5E-05	74.99	31.48	119.61	88.02	81.02	41.54
GNG12	0.59	1.2E-04	138.43	197.86	128.8	202.77	177.37	269.34
GNL3L	1.04	1.8E-04	19.51	76.05	37.18	61.65	61.93	88.19

GPATCH4	0.57	5.5E-04	88.26	145.46	99.82	161.08	188.27	235.91
GPX1	0.88	1.1E-04	23.63	55.52	32.1	60.23	59.68	87.21
GTF3A	0.70	6.7E-04	39.42	72.53	38.03	62.13	60	87.07
HES1	-0.93	5.6E-05	491.89	145.07	309.5	242.09	302.79	187.69
HIST1H4E	1.44	2.0E-05	6.77	30.7	12.02	29.22	17.48	33.82
HLA-F-AS1	-1.18	3.9E-05	47.78	14.27	53.73	22.69	37.77	24.66
HLA-G	-1.13	8.8E-05	159.41	34.8	104.34	79.59	94.55	51.75
HMGA2	1.06	1.0E-06	28.67	69.6	33.51	69.61	46.01	83.35
HMOX1	-2.17	2.7E-12	1853.8	586.14	2398.36	865.01	2371.74	218.96
HSP90AA1	-0.80	4.4E-07	11212.5	5298.9	10838.92	7430.38	8273.98	4820.37
HSP90AB1	-0.78	1.1E-05	26939.85	11425.78	18681.84	13563.72	15235.33	9843.08
HSPA1A	-1.23	1.5E-12	4989.34	1774.64	3052.56	1772.9	2375.19	895.41
HSPA1B	-1.17	3.5E-13	8112.27	3253.48	4774.66	2660.24	4337.29	1700.8
HSPA1L	-1.49	1.9E-08	173.61	34.41	78.33	41.46	70.86	29.96
HSPA2	-0.95	7.4E-06	402.96	127.67	236.54	148.6	133.61	91.46
HSPA4	-0.74	5.0E-05	2102.26	874.91	1407.09	1055.66	1343.36	922.49
HSPA4L	-0.88	1.5E-05	946.74	319.07	685.87	498.79	575.13	369.36
HSPA6	-1.20	8.6E-15	12009.53	4512.17	8307.23	4228	6227.9	2668.56
HSPA7	-1.13	4.5E-14	7223.4	2950.04	1301.61	609.74	1178.97	593.36
HSPA8	-0.79	3.5E-06	10952.89	4701.42	12143.36	8408.95	8163.69	5260.84
HSPB1	-1.05	5.8E-09	310.45	121.41	167.83	107.03	170.7	77.2
HSPB2	-4.89	1.1E-04	5.31	0	9.9	2.61	5.8	0
HSPB8	-1.21	1.1E-09	174.27	52.59	199.07	110.47	171.02	81.64
HSPD1	-0.84	4.7E-06	4937.71	1917.75	3488.6	2335.95	2278.66	1536.85
HSPH1	-1.00	9.3E-09	3715.69	1383.62	2489.98	1585.45	1915.27	1015.98
HTRA1	0.99	1.6E-04	22.3	71.95	62.49	87.43	22.82	40.43
HTRA3	-2.64	1.2E-04	20.84	0.78	8.06	3.8	9.11	1.5
IER5	-0.77	6.3E-07	1694.66	843.23	912.52	548.32	667.93	454.15
IGFBP3	1.28	2.1E-06	149.19	650.85	409.32	500.21	278.45	744.42
IL6	-0.87	4.0E-05	160.47	73.51	56.84	35.04	25.63	14.92

IMPDH2	0.65	1.5E-04	83.62	164.23	148.6	200.99	150.73	218.96
IRAK1	0.91	2.9E-04	27.08	88.17	49.91	78.52	85.07	113.7
ITGA3	0.66	5.1E-04	80.03	179.48	104.06	131.85	153.49	215.76
ITGA6	0.76	1.8E-04	300.49	804.13	398.29	475.51	575.68	884.09
JAG1	0.76	9.0E-07	251.52	370.69	238.52	391.52	339.6	678.08
JMJD1C	-0.60	3.9E-04	1004.08	506.96	746.24	620.66	652.06	447.21
KANK1	0.75	2.9E-04	109.5	296.78	217.6	296.85	166.56	217.46
KBTBD4	-1.59	2.0E-04	49.11	5.08	28.84	16.39	36.12	18.51
KIAA0895	-0.95	4.8E-05	64.37	27.76	61.22	39.32	49.51	24.6
KIF1A	1.27	1.6E-04	76.98	502.46	211.94	363.13	259.08	344.57
KITLG	-0.80	3.8E-04	110.3	45.36	90.35	54.52	53.1	40.1
KLF13	0.78	4.5E-04	44.99	108.7	52.88	74	47.67	71.7
LAD1	1.14	1.0E-04	16.19	75.47	44.25	72.82	71.91	107.94
LIN54	-1.00	3.6E-05	42.61	19.94	55.42	28.15	45.55	23.55
LINC00152	-0.80	1.4E-06	232.67	122.19	146.76	90.16	115.07	67.32
LINC00568	-5.32	9.8E-05	4.91	0.2	3.82	0.12	2.16	0
LMNB2	0.70	2.8E-04	82.03	186.91	88.08	130.07	147.14	188.67
LOC100130899	1.08	6.1E-04	11.15	30.89	14.14	33.14	18.36	27.54
LOC100506305	-0.89	1.5E-04	43	22.48	78.47	53.81	65.89	29.05
LOC400680	-5.58	2.8E-04	5.44	0	3.82	0	1.56	0.13
LRIF1	-0.92	1.8E-06	309.12	113.4	367.61	272.74	262.63	139.87
MAP2	-0.77	1.7E-04	122.24	93.06	125.27	76.74	84.93	37.03
MAP3K8	-0.80	4.9E-04	140.16	59.04	69.28	37.42	57.7	47.69
MBNL2	-0.77	9.7E-06	208.91	103.23	178.71	125.56	117.51	68.63
MCCC2	0.83	6.2E-04	33.71	96.97	49.2	76.86	66.16	85.44
MEX3B	-0.99	6.9E-04	53.62	19.55	33.65	20.31	18.82	10.79
MICB	-1.66	3.5E-05	51.63	6.45	26.16	12	27.38	13.61
MIR22HG	-0.68	6.1E-05	444.37	237.35	336.08	278.91	244.04	132.87
MME	-0.74	3.8E-04	164.98	67.45	160.05	104.3	77.8	61.95
MMP3	-0.51	7.3E-04	617.58	397.47	756.14	504.73	652.7	529.97

MORC4	-0.62	3.2E-04	190.2	100.69	181.82	126.86	126.57	93.81
MRPL18	-1.03	1.3E-06	763.58	223.08	570.5	386.65	393.8	229.17
MRPS26	0.93	5.7E-05	20.44	47.31	34.92	63.91	40.54	66.93
NAP1L2	-1.07	2.9E-07	201.48	73.9	128.66	89.45	88.62	37.16
NEFM	-0.60	2.9E-04	245.28	135.88	228.34	162.98	101.82	73.86
NFKB1	0.55	4.3E-04	142.15	240.09	184.65	239.83	273.62	391.67
NKIRAS2	-0.86	2.4E-04	94.63	42.03	58.25	43.24	45.6	23.29
NOP14	0.63	1.0E-04	91.45	169.12	155.24	223.08	162.74	229.56
NOV	-1.66	4.6E-04	12.61	1.76	22.76	12.59	12.19	4.06
NR4A1	-0.92	3.4E-06	159.27	86.42	86.81	38.37	48.63	29.77
NUAK1	-1.59	3.6E-05	37.56	14.27	12.3	4.75	11.04	2.75
NXT2	-1.14	4.5E-06	137.51	36.36	94.02	56.54	73.52	41.93
PAPD5	-0.78	1.2E-04	167.5	69.41	127.53	95.74	91.51	57.96
PDCD11	0.65	7.0E-04	107.51	251.03	204.73	259.55	184.13	243.1
PDCD2L	1.46	6.1E-04	5.18	35.78	8.91	19.01	17.35	26.89
PDHB	0.82	2.6E-04	29.2	48.1	25.87	43.12	45.92	90.35
PDIA4	0.75	1.0E-04	96.36	241.65	208.55	261.09	325.2	496.28
PDK4	-0.66	2.6E-04	276.6	136.86	135.59	91.23	81.44	61.95
PEG10	-0.88	3.4E-06	211.04	84.66	168.25	110.23	90.27	54.17
PELO	-0.75	4.8E-05	97.55	63.74	113.25	59.75	120.64	73.99
PES1	0.64	7.6E-04	125.16	288.57	196.39	267.03	237.46	290.07
PF4V1	-7.25	4.1E-04	4.91	0	0.99	0	1.2	0
PGF	-0.89	2.2E-04	96.09	34.02	82.85	59.87	53.97	32.51
PIP5K1A	-0.59	5.2E-04	191.52	113.59	172.63	131.14	101.77	66.66
PLEC	0.89	4.3E-05	337.13	1062.79	379.77	516.84	384.14	571.32
PLK2	-0.65	1.1E-04	1022.4	492.1	891.31	664.97	484.99	349.87
PNLDC1	-1.32	2.3E-04	63.05	11.54	33.37	24.23	33.77	15.31
POGK	0.77	4.5E-04	58.53	155.43	71.97	111.07	101.5	125.08
PPP1R14B	0.73	4.4E-04	66.23	168.92	107.03	129.36	143.37	215.23
PPP1R3C	-1.43	1.3E-04	48.31	8.41	42.42	15.32	33.91	25.64

PPP4R4	1.12	1.2E-05	12.87	49.27	25.17	42.53	94.41	160.48
PRKAR1B	1.22	7.1E-04	6.64	29.91	14.28	32.55	24.66	32.97
PRKD2	-1.14	7.6E-06	67.82	28.94	37.89	21.86	42.33	16.16
PRRX1	0.76	7.5E-04	31.59	59.43	28.28	42.41	42.61	72.55
PTGES3	-0.83	6.7E-06	465.47	194.92	290.13	226.88	322.62	177.03
PWP2	1.07	6.0E-05	19.51	75.27	49.49	87.66	66.85	95.84
PYGB	0.67	4.1E-04	36.37	65.89	47.22	68.07	99.8	155.57
RAPGEF2	-0.63	1.3E-04	228.82	160.71	172.78	117.6	126.62	72.16
RASD1	-1.12	1.0E-04	140.16	32.45	82.57	59.04	49.65	28.33
RASSF9	-0.72	1.5E-04	143.48	73.9	97.84	64.98	65.15	42.59
RBPJ	0.67	4.2E-04	38.62	82.51	93.88	134.7	115.3	156.68
RGS2	-1.09	1.2E-05	506.62	122	273.16	166.18	179.3	122.73
RGS4	-1.42	1.5E-04	35.44	13.29	15.13	4.16	13.16	6.48
RRAD	-0.57	2.0E-04	417.56	246.54	299.18	217.62	271.05	194.23
RRP12	0.54	5.6E-04	130.47	216.43	155.67	225.81	260.51	333.84
RRP9	0.93	8.6E-05	34.24	107.73	60.23	83.51	71.78	117.49
S100A2	0.93	5.2E-05	21.63	40.67	38.74	50.84	86.04	235.58
S100A6	0.71	3.4E-06	240.5	358.17	328.3	498.07	779.32	1502.05
SAA1	-0.82	4.3E-05	472.38	172.44	213.07	150.27	136.56	95.51
SCD	0.98	2.7E-05	24.55	68.43	51.18	69.97	57.83	118.8
SCML1	-0.87	4.0E-04	45.92	20.92	57.83	34.09	37.22	22.31
SDCCAG3	0.83	1.1E-04	36.37	95.21	67.44	107.15	93.26	129.07
SDPR	-1.17	1.9E-05	108.44	35.39	44.82	17.46	18.4	12.69
SEC22C	0.81	7.5E-04	30.93	73.9	33.51	57.14	43.2	57.83
SEPP1	-1.11	5.3E-04	44.86	12.12	50.76	38.01	19.42	9.03
SERPINH1	-0.96	2.8E-07	604.17	208.8	564.84	359.09	482.05	293.08
SH2D5	0.89	5.5E-05	66.23	196.29	86.53	129.24	69.15	100.68
SH3BGR	-1.01	2.2E-06	115.21	49.85	69.85	41.81	47.62	22.37
SMO	-1.36	6.9E-04	29.6	6.84	25.03	17.58	12.42	4.32
SNAP23	-0.96	1.4E-05	383.98	119.26	152.27	115.7	238.89	136.6

SNX3	-0.62	2.2E-04	402.43	203.72	282.21	221.18	243.03	168.59
SOD1	-0.53	4.4E-04	1190.03	704.03	801.81	601.18	792.53	592.38
SOGA3	-2.51	2.9E-04	13.67	1.37	11.03	4.51	2.99	0.33
SOX9	-0.99	2.1E-04	165.78	43.21	133.05	108.22	93.54	54.69
SPP1	-0.76	5.3E-05	43.4	24.05	267.22	167.97	20.01	11.64
SPRY2	1.13	1.4E-06	33.05	52.4	22.48	59.39	42.97	107.16
SPTY2D1	-0.61	5.1E-04	285.76	142.92	247.71	204.55	164.26	111.54
SREBF1	0.88	7.2E-04	16.33	41.25	29.13	49.42	33.17	49.2
ST13	-0.62	2.1E-04	847.06	414.87	733.94	583.48	740.12	517.8
STIP1	-0.85	2.9E-06	1761.95	686.63	1264.29	879.62	1189.14	751.61
STK11	0.86	2.4E-04	45.39	131.77	53.3	82.2	67.36	91.2
STK17B	-0.64	3.2E-04	134.05	74.88	139.13	88.02	99.75	73.66
SYNE1	-0.53	2.3E-04	1408.63	967.97	2559.68	1805.21	492.13	339.07
SYNPO2	-0.90	5.0E-05	93.31	58.07	64.9	32.43	32.94	16.22
SYP	1.23	1.9E-04	5.97	29.72	23.75	39.32	34.37	59.99
TAOK3	-1.32	1.6E-11	144.01	46.53	132.2	60.46	85.9	37.22
TCP1	-0.70	2.6E-04	2586.32	1052.23	1907.6	1425.45	1528.51	1171.74
TGFBI	1.64	2.5E-04	1.59	17.79	14.7	22.69	21.26	51.55
THOP1	0.73	7.2E-04	28.14	55.52	45.53	71.87	50.15	74.45
TINAGL1	0.92	1.0E-04	107.38	378.51	413.28	595.12	290.92	395.53
TP63	0.97	3.9E-06	36.63	106.55	70.27	102.75	113.19	203.92
TRAF1	0.92	6.2E-04	61.85	248.88	211.52	296.26	420.99	525.78
TSPYL2	-0.65	1.3E-04	2148.98	1024.27	1721.82	1375.32	1509.87	1021.41
TTC28-AS1	-1.60	1.7E-04	28.14	4.5	17.39	7.01	17.12	8.7
TXNIP	1.08	3.8E-09	44.33	104.01	41.71	89.68	91.79	174.48
TXNL1	-0.55	6.8E-04	341.24	188.28	314.73	260.62	376.18	260.77
UBB	-0.98	1.6E-10	9995.13	4432.59	7484.5	3752.01	5609.25	3297.18
UBC	-1.19	2.0E-14	6809.95	2489.81	5701.03	2840.68	4185.46	1924.34
ULBP1	-1.19	1.4E-05	70.48	27.76	49.34	30.41	19.97	6.87
USPL1	-1.02	2.6E-06	229.62	73.51	127.11	69.85	92.02	61.76

VEGFA	0.81	1.3E-07	315.36	616.83	360.4	530.74	442.16	831.56
VTRNA1-2	-1.49	3.3E-07	63.71	15.05	68.71	34.33	12.33	4.58
VTRNA1-3	-1.70	2.2E-05	141.75	12.51	233.15	112.97	76.7	44.49
WBP5	-0.62	4.0E-04	1385.8	643.81	903.61	690.87	638.76	490.46
WDR3	0.85	6.8E-05	126.36	384.18	267.79	362.3	327.78	474.43
WDR77	0.93	4.6E-04	15	45.94	31.53	55.24	50.06	67.91
ZFAND2A	-1.06	1.6E-09	421.94	150.74	234.14	126.75	196.79	111.08
ZFP106	-0.62	1.4E-04	344.69	182.41	309.5	230.09	191.82	134.9
ZNF131	-0.71	1.1E-05	488.7	240.09	338.48	217.26	319.59	230.67
ZNF184	-0.82	5.3E-04	63.84	26.59	68.43	42.88	54.8	37.49
ZNF701	-0.79	5.3E-04	79.77	33.04	80.59	50.37	59.72	43.5
ZP4	1.60	3.7E-05	3.98	17.01	6.22	14.97	9.8	27.48

Table 4.5. The 10 most enriched Gene Ontology biological processes in curcumin vs DMSO treated ALDH-/CD44+/CD24- cells.

Name	#Genes	Odds Ratio	P-Value	FDR
response to protein stimulus	101	0.22	1.3E-31	3.5E-28
response to unfolded protein	61	0.20	1.3E-29	1.7E-26
response to biotic stimulus	333	0.41	1.7E-20	1.5E-17
ribonucleoprotein complex biogenesis	187	4.12	8.2E-19	5.7E-16
protein refolding	11	0.20	1.5E-18	8.5E-16
protein folding	150	0.38	1.2E-17	5.4E-15
ribosome biogenesis	126	4.54	8.7E-17	3.4E-14
response to organic substance	723	0.53	4.0E-16	1.4E-13
ncRNA metabolic process	221	3.25	3.0E-14	9.1E-12
response to heat	51	0.34	6.5E-14	1.8E-11

Table 4.6. The top 10 most enriched Gene Ontology Biological processes enriched in ALDH+ and ALDH-CD44+CD24- breast cells following 5 µM piperine treatment.

Gene Ontology Biological Processes enriched in piperine treated ALDH+ cells				
Name	#Genes	OddsRatio	P-Value	FDR
non-recombinational repair	14	0.02	1.5E-06	2.4E-03
positive regulation of transcription from RNA polymerase II promoter	295	3.25	1.8E-06	2.4E-03
regulation of transcription from RNA polymerase II promoter	593	2.19	1.7E-05	1.5E-02
mesoderm development	51	8.92	2.2E-05	1.5E-02
V(D)J recombination	10	0.02	3.9E-05	2.1E-02
transcription from RNA polymerase II promoter	733	1.95	6.0E-05	2.6E-02
regulation of tissue remodeling	12	30.46	7.5E-05	2.6E-02
positive regulation of transcription, DNA-dependent	393	2.38	7.9E-05	2.6E-02
positive regulation of gene expression	474	2.20	9.5E-05	2.8E-02
positive regulation of RNA metabolic process	399	2.33	1.1E-04	2.8E-02
Gene Ontology Biological Processes enriched in piperine treated ALDH-/CD44+/CD24- cells				
Name	#Genes	OddsRatio	P-Value	FDR
translational elongation	99	14.51	1.9E-11	5.3E-08
ribonucleoprotein complex biogenesis	186	5.13	4.2E-07	5.7E-04
ribosome biogenesis	126	6.62	7.4E-07	6.8E-04
response to amino acid stimulus	18	42.95	1.2E-06	7.9E-04
ncRNA processing	182	4.84	1.5E-06	8.2E-04
ncRNA metabolic process	222	3.90	6.4E-06	2.9E-03
translation	377	2.84	1.1E-05	4.4E-03
response to amine stimulus	34	16.80	1.4E-05	4.8E-03
generation of precursor metabolites and energy	281	3.15	2.5E-05	7.7E-03
response to organic nitrogen	62	8.88	3.0E-05	8.2E-03

Table 4.7. Genes identified as differentially expressed (FDR<0.05) in curucmin and piperine co-treated ALDH+ cells, compared to DMSO controls.

Gene	logFC	P Value	Sample 1 CPMs		Sample 2 CPMs		Sample 3 CPMs	
			Curcumin +Piperine	DMSO	Curcumin+Piperine	DMSO	Curcumin+Piperine	DMSO
ACAT2	1.00	2.4E-06	29.58	78.26	44.45	72.68	82.83	177.19
ACOT13	-1.07	1.0E-04	19.72	10.04	41	21.21	66.15	28.09
ACSL1	0.70	1.6E-04	74.94	143.82	224.26	314.01	289.89	495.09
ACTN1	0.79	1.8E-05	128.19	274.36	417.59	597.28	326.09	583.28
AGRN	0.93	5.9E-05	29.58	49.91	32.65	62.2	34.83	70.77
AHNAK	0.63	2.5E-05	767.16	1260.44	1914.76	2712.26	1378.88	2223.06
AHSA1	-0.70	1.4E-05	1601.38	903.4	639.11	439.92	430.01	258.76
AKAP12	0.94	1.7E-04	25.64	84.17	45.31	68.79	31.54	57.26
AKR1C1	-0.99	3.8E-04	141.99	137.92	226.99	113.2	636.63	183.94
ALDH1A3	0.75	1.7E-06	370.76	587.1	777.35	1216.88	2106.05	3982.08
ARHGAP29	0.66	4.4E-04	80.86	164.5	243.53	375.33	187.21	257.41
ASNS	0.95	1.3E-04	51.28	93.03	63.44	89.52	29.27	81.03
BAG3	-1.07	7.8E-07	1214.84	395.73	338.9	234.1	200.26	95.08
BTN2A3P	-2.76	3.0E-04	33.53	4.73	6.76	2.62	5.45	0.14
C11orf73	-1.07	4.8E-05	69.03	25.4	45.6	21.37	38.35	22.55
C20orf111	-0.75	1.3E-05	591.64	360.89	327.11	214.08	300.56	157.6
C2CD2	1.12	3.9E-04	5.92	20.97	25.17	38.92	23.83	63.61
C6orf15	1.62	4.6E-04	0	5.91	19.71	37.18	5.22	22.69
CACYBP	-0.79	5.6E-06	1946.51	904.28	697.8	488.45	503.08	299.54
CALD1	0.67	5.2E-04	151.86	289.42	634.66	760.68	291.93	534.8
CBR3	-1.34	4.2E-06	138.05	70.88	35.53	18.67	52.65	11.88
CBS	1.36	4.6E-04	7.89	20.67	19.56	30.9	11.35	47.67
CCL2	-1.07	8.7E-06	347.1	113.4	57.68	38.45	13.62	6.62
CD24	0.69	3.8E-05	268.21	453.62	834.46	1136.09	1088.43	2007.38
CDH1	0.85	1.1E-04	114.38	349.07	478.87	650.98	556.75	872.29
CDH3	0.91	4.3E-04	3.94	34.85	55.24	77.05	97.01	174.62
CDK2AP1	0.83	1.2E-04	21.69	61.43	80.41	113.59	108.81	202.57

CDK6	0.85	1.4E-05	49.3	97.16	130.47	194.78	117.89	245.25
CDKN2AIP	-0.71	2.0E-04	147.91	75.9	101.41	67.6	95.99	61.99
CEP170B	0.79	3.8E-04	35.5	102.77	84.15	125.59	64.9	97.24
CEP85	-1.07	1.4E-04	82.83	26.58	29.63	19.62	48.9	22.82
CHORDC1	-0.62	4.9E-04	522.62	313.04	403.92	313.22	210.92	124.38
CLDN1	0.99	5.4E-06	23.67	78.85	80.99	150.45	64.67	109.39
CLMN	1.02	1.8E-05	9.86	38.39	50.06	82.69	56.05	116.28
CLSTN1	0.85	1.6E-05	47.33	107.5	120.4	178.1	111.08	215.54
CLU	-0.65	5.0E-05	1534.33	854.08	948.24	686.64	1051.66	671.74
CLUH	0.82	1.4E-04	33.53	78.85	59.27	95.4	36.53	61.31
CNN2	0.89	1.3E-05	43.39	97.16	80.55	123.28	71.48	145.58
COL12A1	0.84	3.4E-05	69.03	98.93	121.26	208.04	28.71	61.99
CREM	-1.05	6.8E-07	102.55	52.27	56.96	26.21	66.03	32.01
CRYAB	-0.88	2.3E-06	1396.28	568.5	452.83	273.02	338.68	220.94
CSRP2	-1.29	2.5E-08	512.76	130.53	86.31	45.99	69.66	34.3
CYP1B1	1.77	3.9E-05	23.67	34.26	8.49	27.8	4.31	29.71
CYP51A1	0.90	6.7E-05	51.28	172.17	206.85	281.12	473.13	831.91
DAG1	0.79	3.7E-04	25.64	58.18	64.01	89.13	63.99	121
DEPDC5	-1.33	3.9E-04	33.53	10.93	18.41	8.98	19.63	7.29
DNAJA1	-0.58	1.9E-04	3577.47	2040.69	1664.17	1284.08	1366.97	922.39
DNAJB1	-1.13	2.4E-09	4459.02	1420.21	1414.88	805.64	1000.83	524.8
DNAJB4	-1.02	7.2E-08	1416	691.65	475.42	307.42	468.82	177.73
EIF4EBP1	1.19	1.6E-07	27.61	55.23	30.35	72.6	35.74	82.92
F3	0.65	3.9E-04	67.05	115.77	172.33	275.72	154.31	224.32
FADS2	1.40	1.3E-04	0	10.63	21	33.84	24.51	84
FAM53C	-0.78	7.5E-05	106.5	57	75.23	45.76	72.73	42.81
FBXO30	-0.91	1.7E-04	201.16	142.05	164.99	111.29	151.58	48.75
FLNA	0.76	1.1E-04	495.01	1266.94	1460.2	1931	1202.11	1784.41
FTL	-0.99	4.0E-08	790.83	504.71	534.25	268.33	1396.13	563.29
GAA	1.97	2.5E-04	1.97	17.42	3.45	11.12	2.04	7.56
GABARAPL1	-0.76	2.9E-04	254.41	217.65	206.42	116.85	250.75	110.07
GADD45G	-1.42	2.5E-08	143.97	56.7	39.99	19.46	49.7	13.37
GARS	0.86	1.8E-05	80.86	207.32	173.19	252.61	112.89	201.49
GLA	-1.42	8.1E-08	104.52	28.06	43.01	24.39	98.6	32.01

GLIPR1	0.67	2.2E-04	104.52	202.59	252.6	376.92	207.18	306.7
GLRX	-1.09	1.4E-04	63.11	11.52	77.39	55.29	311.56	174.75
GNB4	1.46	3.6E-05	0	25.99	21	41.23	20.2	50.64
GNE	1.21	4.3E-05	9.86	42.82	17.55	31.77	24.17	56.05
GSR	-0.99	1.0E-06	100.58	68.52	101.12	44.48	187.89	87.78
GUCY1A3	1.00	3.3E-05	15.78	51.09	52.94	79.2	59.45	132.75
HIP1R	0.68	6.8E-04	57.19	71.47	82.71	127.97	93.83	178.94
HMOX1	-2.27	2.1E-13	1416	620.18	1067.78	203.04	1698.16	183.26
HSP90AA1	-0.68	9.7E-06	9056.09	4610.01	6381.51	4890.86	5118.77	3217.43
HSP90AB1	-0.58	1.3E-05	20476.78	12166.18	9580.68	7299.27	8245.28	5463.71
HSPA1A	-1.35	5.8E-13	3346.73	1032.45	978.02	523.01	603.5	222.29
HSPA1B	-1.41	5.2E-15	5251.82	1495.52	1921.52	944.73	1422.34	542.77
HSPA1L	-1.61	3.5E-10	126.22	35.73	32.65	11.12	29.61	10.53
HSPA4L	-0.80	9.6E-05	445.7	189.01	208.58	160.06	107.11	61.45
HSPA6	-1.41	2.0E-17	8683.35	2579.07	2883.57	1354.46	2206.01	841.9
HSPA7	-1.32	3.5E-16	2510.54	942.68	296.04	130.43	355.58	139.51
HSPA8	-0.54	4.0E-04	10939.49	6100.81	6129.2	5252.29	4736.97	3196.22
HSPD1	-0.73	1.5E-06	3358.56	1905.14	1612.96	1099.07	1015.02	571.4
HSPH1	-0.83	3.5E-07	2583.51	1406.04	1442.22	972.37	894.18	436.35
HYOU1	0.72	5.8E-04	368.79	952.42	727.44	945.13	508.3	703.34
ID2	-1.05	1.6E-04	254.41	67.04	46.75	32.57	43.57	25.79
IER5	-0.74	3.3E-05	2001.73	921.71	852.01	635.96	737.61	458.09
IGFBP3	0.98	8.1E-06	136.08	434.13	425.93	564.31	1273.82	2525.97
IGSF3	0.79	2.0E-04	19.72	39.87	53.37	84.04	76.02	136.4
INTS1	1.32	3.5E-05	15.78	52.86	29.92	46.63	14.98	51.72
ITGB4	1.16	3.6E-07	13.81	35.14	67.03	116.37	113.23	312.64
ITGB8	0.70	4.4E-04	65.08	160.36	273.74	380.82	462.92	676.6
JAM3	2.69	1.3E-04	0	16.83	1.01	7.15	1.25	4.59
KCNN4	1.28	2.5E-04	1.97	64.38	85.88	121.38	88.39	175.02
KDR	1.79	1.0E-04	3.94	2.36	7.34	25.9	7.15	32.55
KIAA1644	3.37	3.7E-06	3.94	8.56	0.43	6.59	0.91	14.32
KIFC3	1.08	2.3E-04	11.83	40.75	37.4	57.75	15.66	37.68
KLF6	0.68	2.6E-04	76.91	158.29	140.25	218.45	104.72	147.61
KLK10	1.01	1.1E-05	1.97	22.44	77.1	122.09	161	338.16

KLK5	0.79	2.7E-04	21.69	58.18	120.54	158.47	148.86	281.98
KLK7	1.01	4.7E-06	5.92	29.53	79.26	148.78	70.23	129.24
KRT15	1.38	2.4E-10	136.08	470.16	801.23	1334.28	874.21	2836.32
KRT16	0.91	3.6E-07	112.41	189.3	402.92	703.17	374.42	821.37
KRT17	0.89	1.4E-06	408.23	624.32	1541.9	2522.88	380.21	959.66
KRT5	0.92	9.3E-07	224.82	440.33	504.04	751.07	268.56	627.17
KRT6A	1.14	6.4E-08	21.69	44	162.4	304.16	66.26	178.67
KRT7	0.59	2.8E-04	1173.43	2035.67	3246.93	3763.83	4664.02	7904.2
KRT81	1.23	3.9E-04	3.94	52.27	123.71	149.9	19.97	50.78
LOXL4	1.79	3.2E-04	5.92	5.02	6.04	15.97	8.28	57.26
LPAR6	-2.37	1.4E-04	63.11	9.75	16.97	9.85	6.69	0.27
LPP	0.72	3.6E-04	23.67	35.73	92.64	147.75	83.28	146.53
LRIF1	-0.80	1.4E-05	451.62	212.63	235.48	160.78	124.58	72.66
LYPD3	-0.99	1.6E-04	47.33	21.56	39.41	18.27	46.86	27.15
MACC1	0.85	9.0E-06	106.5	269.63	530.8	871.17	306.57	472.54
ME1	-1.11	3.4E-04	13.81	3.84	40.28	19.86	51.51	26.33
MED6	-0.77	1.3E-04	220.88	108.09	110.19	85.23	108.92	56.18
MIR22HG	-0.69	4.0E-04	684.33	501.17	357.17	258.33	332.44	149.77
MRPL18	-0.73	2.0E-05	534.45	305.66	331.28	221.23	218.3	124.38
MTHFD1L	0.83	3.1E-04	33.53	59.36	35.67	65.69	28.93	50.1
MTHFD2	0.74	4.1E-04	112.41	288.24	221.09	290.34	134.56	206.09
MVD	1.16	2.2E-05	11.83	32.49	19.56	37.57	28.14	67.93
MYH9	0.82	2.0E-04	532.48	1546.02	2198.71	2574.83	1414.28	2364.18
NBEAL2	1.31	3.8E-05	5.92	12.11	24.89	49.09	20.2	65.5
NRIP3	-1.43	5.0E-04	55.22	18.9	21.86	14.54	18.38	3.78
NT5E	0.72	3.3E-04	23.67	56.7	121.26	190.33	116.86	174.62
OSBPL5	1.15	3.0E-04	11.83	31.9	13.09	33.28	17.47	31.47
PDIA4	0.88	3.8E-05	163.69	491.42	476.85	712.78	490.94	740.88
PFDN4	-0.74	1.9E-04	102.55	78.26	95.08	50.6	121.29	68.88
PFN2	-1.22	1.3E-06	65.08	40.46	63.72	27.25	75.34	24.71
PHGDH	1.06	9.5E-06	25.64	47.55	25.6	56.88	30.97	64.28
PIP4K2B	1.17	6.7E-04	7.89	31.9	12.95	19.94	18.49	47.4
PKP3	1.05	1.2E-04	3.94	43.12	47.76	81.74	75.68	124.92
PLEC	0.85	1.2E-04	485.15	1448.86	1486.95	1854.59	939	1519.85

PLXNA1	0.89	1.9E-04	27.61	88.6	80.7	106.6	62.86	123.03
PPL	0.97	2.2E-04	29.58	139.69	311.14	400.36	388.83	684.57
PPP1R3C	-1.19	5.3E-05	61.14	21.85	37.11	18.51	19.63	8.78
PROM1	0.73	2.7E-04	94.66	151.8	221.53	293.2	244.51	523.86
PRSS12	1.15	1.6E-04	7.89	20.67	24.31	43.45	20.2	52.8
PSAT1	1.34	1.1E-06	25.64	44	38.55	85.23	16.11	57.53
PSMA2	-0.70	6.2E-05	234.69	157.41	215.77	136.47	278.77	155.44
PSMA3	-0.74	5.2E-04	191.3	169.81	234.62	134.96	300.67	134.51
PSMA4	-0.76	3.2E-05	280.04	195.21	250.73	155.22	451.57	218.51
PSMA6	-0.57	4.5E-04	709.97	519.77	544.61	370.57	703.68	434.05
PSMB1	-0.65	1.3E-04	508.81	326.33	324.38	225.28	395.64	229.45
PSMC1	-0.71	1.0E-04	183.41	138.21	167.58	98.82	252.34	134.51
PSMC3	-0.73	5.9E-05	230.74	173.06	254.03	143.46	398.02	211.76
PSMD12	-0.62	4.2E-04	159.74	106.61	156.94	107.56	167.24	101.15
PSMD13	-0.62	3.0E-04	301.74	197.28	221.09	151.56	353.66	219.05
PSMD14	-0.78	2.9E-04	374.71	326.92	438.3	247.44	578.31	238.5
RASD1	-0.83	9.0E-05	153.83	83.87	41.43	25.82	39.26	20.53
RGS2	-1.14	3.9E-11	765.19	323.97	176.64	87.78	132.41	58.48
RNASE4	1.93	2.6E-04	0	9.16	1.87	7.23	7.83	22.96
RPL39	-0.83	5.1E-05	126.22	96.28	90.77	44.8	107.79	54.29
RPS6KA2	1.14	2.1E-04	5.92	16.24	36.97	60.77	19.52	54.97
S100A8	-0.86	4.3E-05	315.54	134.96	43.87	31.38	44.93	24.31
SACS	0.71	2.6E-04	216.94	513.86	444.63	596.32	529.52	775.19
SCD	1.59	4.1E-10	21.69	114.29	155.21	288.67	225.11	830.15
SCNN1A	1.46	2.5E-06	3.94	21.56	17.55	34.87	27.23	88.32
SDC1	0.94	1.2E-05	35.5	60.84	73.07	119.55	71.59	171.38
SGPL1	0.82	2.1E-04	21.69	55.52	74.08	106.29	61.84	115.06
SHFM1	-0.65	1.6E-04	297.79	210.27	239.94	142.59	290.8	181.24
SHMT2	0.98	5.3E-06	45.36	134.37	86.16	134.64	63.42	123.44
SLC38A1	0.78	1.5E-05	252.43	497.92	394.57	555.41	215.35	403.39
SLC43A3	0.63	6.5E-04	102.55	199.05	143.42	210.27	120.95	167.87
SLC44A2	1.05	4.7E-04	3.94	16.54	46.75	65.3	37.56	100.61
SLC7A1	0.99	1.0E-06	37.47	83.58	68.62	129.64	33.58	66.85
SLC7A5	0.75	4.1E-05	98.61	175.42	134.93	198.03	89.18	165.44

SNIP1	-0.60	6.1E-04	201.16	134.37	152.62	98.82	92.81	62.12
SOD1	-0.85	7.1E-08	1500.8	853.49	840.21	506.64	1181.46	588.55
SQLE	0.85	1.6E-04	23.67	44.89	52.65	75.15	96.78	213.24
SQSTM1	-0.64	4.4E-04	1993.84	1728.83	2205.62	1427.86	4008.56	1887.86
SREBF1	1.24	5.3E-06	17.75	45.48	43.3	72.37	31.77	103.18
SRXN1	-1.34	2.2E-10	305.68	164.5	335.45	140.92	559.59	158.01
STC2	1.25	3.4E-09	31.55	102.77	65.16	127.89	50.72	126.68
TAGLN	1.16	4.7E-04	7.89	14.47	27.91	46.95	16.79	53.61
TGFB2	0.97	4.7E-07	82.83	219.43	150.75	262.69	92.92	166.65
THAP9-AS1	-0.72	4.0E-04	323.43	137.33	284.96	218.61	121.18	79.41
THSD4	0.93	2.5E-04	13.81	31.01	24.89	45.76	35.63	66.98
TMA7	-0.70	6.0E-04	187.35	160.36	141.69	86.19	181.31	84.95
TMEM132A	0.97	6.8E-04	21.69	54.04	27.91	39	19.74	48.08
TNC	0.95	1.6E-04	199.19	259.29	136.94	221.47	60.81	203.39
TP53INP1	1.38	6.2E-04	7.89	12.11	6.47	24.55	14.41	32.14
TRIM16L	-0.98	3.6E-04	43.39	43.12	86.45	35.67	86.68	33.9
TUBA1A	0.67	3.9E-04	106.5	159.77	183.69	252.84	249.95	476.19
TXN	-0.67	3.9E-04	481.2	371.52	411.98	281.12	576.83	275.23
TXNIP	0.93	9.3E-05	74.94	85.05	76.96	145.37	86.68	240.93
UBB	-1.33	5.6E-14	11966.97	3539.17	2888.17	1547.25	3215.58	1268.93
UBC	-1.10	3.1E-10	5727.11	1953.28	2401.39	1277.09	2120.92	1173.72
VGLL3	1.10	2.0E-05	19.72	46.07	27.62	58.94	18.38	38.08
VIM	0.59	2.6E-04	546.28	907.83	503.18	718.89	369.09	535.74
VTRNA1-3	-1.73	4.6E-05	65.08	19.79	22.44	12.55	20.54	2.97
ZFAND2A	-1.58	2.4E-10	571.92	119.61	141.26	76.02	134.34	44.03
ZNF462	0.72	4.2E-04	86.77	153.86	111.05	145.05	38.12	76.3
ZNF701	-0.79	6.1E-04	86.77	38.69	59.7	35.98	48.9	32.68
ZPLD1	2.08	3.6E-04	0	3.25	7.05	18.83	2.16	13.5

Table 4.8. Genes identified as differentially expressed (FDR<0.05) in curucmin and piperine co-treated ALDH-CD44+CD24- cells, compared to DMSO controls.

Gene	logFC	P Value	Sample 1 CPMs		Sample 2 CPMs		Sample 3 CPMs	
			Curcumin +Piperine	DMSO	Curcumin+Piperine	DMSO	Curcumin+Piperine	DMSO
ABCB1	-0.94	1.2E-06	483.29	188.74	283.11	190.49	127.05	68.59
ABHD3	-0.95	5.1E-04	86.08	28	60.26	44	42.42	24.15
ACTBL2	-1.61	1.9E-04	39.24	8.22	21.86	8.68	4.47	1.9
AFAP1	-0.73	6.4E-04	87.09	43.86	95.81	58.74	66.03	46.47
AFG3L2	0.72	5.2E-04	122.53	306.21	157.65	228.78	207.42	262.19
AGXT	1.01	7.4E-04	28.1	127.26	50.52	69.68	130.37	182.41
AHSA1	-0.85	1.3E-05	1793.67	693.27	996.73	760.88	878.95	507.56
AHSA2	-0.80	5.5E-04	99.49	52.27	35.75	22.35	51.21	29.78
AKAP1	0.95	2.0E-04	31.65	97.89	45.49	60.88	55.88	100.01
AKT1	1.06	1.9E-04	20	80.47	42.64	77.05	73.13	98.37
ALAS1	-0.94	2.2E-06	315.7	116.69	233.87	151.49	194.1	113.56
ALDH1A2	-1.52	9.3E-06	70.38	22.32	37.52	24.38	39.51	8.05
ANKRD1	-1.64	2.8E-08	112.4	21.73	36.24	13.79	41.13	18
ANO1	1.23	3.2E-06	29.37	130.98	72.28	117	73.94	140.98
ANXA1	-0.80	1.6E-05	3025.56	1209.55	3604.26	2574.2	1299.41	864.59
AP4B1	-1.16	8.1E-05	96.2	26.63	26.88	17.72	45.19	21.86
ARFRP1	0.95	6.5E-04	19.75	66.96	34.96	53.51	49.12	72.65
ARID5B	-0.71	1.6E-05	622.28	326.76	619.98	424.26	314.65	199.36
ARL4D	-1.83	2.8E-05	32.4	7.24	18.81	4.04	13.19	5.96
ATP1B1	-1.38	4.6E-04	63.04	9.2	44.9	25.33	28.14	17.28
AXIN2	-0.90	8.7E-04	85.32	31.72	53.47	33.17	23.27	15.38
B4GALNT1	-1.55	4.8E-04	13.16	4.7	17.23	6.9	18.67	5.24
BAG3	-1.40	1.2E-14	994.93	299.94	656.71	322	477.5	176.78
BCL10	-0.75	4.2E-04	109.87	59.52	66.17	36.98	71.31	49.41

BEX1	-0.89	6.8E-05	154.68	56.58	172.03	127.47	45.6	25.85
BIRC3	0.58	5.0E-04	246.33	449.32	288.13	391.44	558.54	760.66
BMP4	-1.09	6.9E-04	36.71	11.75	40.57	21.76	23.41	13.55
BPGM	-0.84	2.3E-04	82.28	35.44	70.51	45.3	65.01	40.19
C2orf44	-0.93	8.1E-04	65.32	22.91	43.82	26.63	41	26.83
C7orf50	1.10	8.0E-04	17.97	85.75	28.26	51.72	65.35	79.72
CABLES1	-1.08	2.4E-07	151.14	58.34	116.2	57.19	44.38	24.61
CACYBP	-1.03	3.6E-07	2186.58	700.32	1119.03	737.46	999.44	554.49
CAMK2N1	0.66	6.8E-04	40.25	77.33	78.09	117.36	104.66	147.39
CASP4	-0.80	3.7E-05	294.43	123.34	206.49	146.85	170.76	106.68
CBR3	-1.07	4.3E-05	66.84	25.45	42.34	21.76	43.3	23.5
CCDC121	-1.54	2.8E-07	77.97	19.97	33.68	11.77	26.93	12.04
CCK	-0.96	2.4E-04	55.7	20.95	68.93	49.46	50.13	24.28
CCND2	0.84	2.2E-04	116.96	343.21	173.8	222.83	126.78	195.11
CD83	-0.68	1.2E-04	256.71	145.27	190.84	112.37	143.29	103.74
CDC42EP3	-0.97	7.7E-04	173.92	42.48	117.38	80.74	93.83	71.6
CDK6	0.68	7.7E-04	60	120.02	86.46	110.58	67.79	110.61
CDR2	-0.68	6.4E-04	166.84	84.77	107.24	81.33	80.03	50.59
CHORDC1	-1.12	1.1E-07	1024.05	297.59	681.03	393.94	427.7	246.48
CITED2	-0.75	1.5E-04	182.02	82.42	149.68	97.74	72.86	51.25
CKAP5	0.80	2.9E-04	54.43	149.77	112.55	162.07	131.45	179.92
CLU	-1.19	2.0E-10	1130.88	374.53	877.09	516.17	583.51	254.01
CNN3	-0.57	4.2E-04	574.43	352.21	400.98	279.19	274.47	196.09
COL2A1	-1.05	1.1E-07	167.09	67.35	120.23	71.58	109.94	51.57
CRYAB	-0.83	4.4E-07	1956.45	933.49	1506.42	1037.34	1131.63	617.85
CSRP2	-0.91	6.4E-06	625.31	231.42	483.59	355.89	274.33	151.97
CTGF	-0.86	2.1E-06	393.42	185.6	192.31	127.71	82.81	44.18
CYB5D2	-1.42	1.9E-05	25.06	9.98	35.15	16.77	38.36	10.6
CYP1B1	2.11	2.2E-05	22.28	44.64	10.83	33.41	1.62	24.87
DEDD2	-0.86	1.2E-06	281.01	148.8	197.73	128.3	227.11	111.79

DKC1	0.84	3.9E-04	64.56	199.89	158.93	212.72	145.93	207.15
DNAJA1	-0.93	4.9E-05	3826.07	1120.27	2639.43	1856.96	1946.38	1344.86
DNAJA4	-0.87	1.3E-07	410.13	241.01	393.3	211.18	223.19	115
DNAJB1	-1.21	3.8E-15	5138.21	2009.13	3130.51	1590.13	3482.24	1403.31
DNAJB4	-1.20	8.2E-10	1893.92	583.63	792.2	476.46	941.8	420.97
DOK5	-0.77	8.6E-04	128.1	63.04	75.63	40.67	22.73	17.54
DST	0.61	3.6E-05	1769.11	2859.22	2376.32	3472.29	1882.11	2806.42
DUSP1	-0.90	2.7E-07	390.89	177.18	271.68	159.93	197.41	114.21
DUSP2	-0.93	4.9E-06	324.05	120.6	147.81	102.97	195.52	108.06
DUSP4	0.57	5.6E-04	267.09	422.5	382.17	600	182.19	237.84
DUSP7	1.03	3.2E-07	32.15	68.72	57.8	103.92	66.98	148.24
EBNA1BP2	0.64	2.8E-04	176.2	354.56	321.21	476.34	397.94	513.39
ECHS1	1.09	1.4E-04	20.76	80.08	31.61	57.43	48.51	70.75
EDN1	-0.73	7.0E-04	201.77	84.77	121.32	83.47	60.82	46.14
EGR1	-0.67	1.7E-05	849.87	522.15	452.97	277.29	539.87	352.84
EGR2	-0.88	3.8E-04	177.21	62.65	84.69	45.07	55.2	46.4
EHD3	1.03	6.5E-04	25.32	109.64	38.01	49.7	71.1	113.23
FADD	1.26	7.7E-05	12.41	59.52	19.01	37.69	44.25	70.1
FAM210A	-0.98	2.6E-04	197.21	52.86	172.52	100.59	80.37	65.32
FAM210B	1.04	2.1E-04	18.48	32.5	15.26	32.82	26.32	59.04
FAM60A	0.64	7.7E-04	126.33	267.44	157.85	215.93	148.84	197.72
FAM83D	-1.31	1.5E-05	70.63	21.54	52.88	32.58	17.66	6.02
FBXL14	-1.47	2.5E-12	153.67	43.27	127.03	47.21	89.1	39.2
FBXL5	0.87	8.5E-04	25.06	78.12	53.17	66.94	69.28	112.64
FBXO30	-0.72	9.4E-05	250.38	178.36	229.83	150.89	160.41	76.45
FEM1C	-0.98	9.4E-07	121.27	59.71	122.4	67.9	56.02	26.57
FLNC	-0.90	3.3E-05	150.89	81.45	61.84	37.22	23.61	11
FOS	-0.80	8.1E-06	779.49	342.42	420.08	266.83	227.86	154.85
FOSB	-0.94	1.8E-07	487.85	201.27	296.3	165.76	182.87	111.59
GABARAPL1	-0.74	7.4E-05	265.32	162.89	206	153.27	217.84	102.63

GADD45G	-0.77	9.3E-05	170.13	102.79	112.45	55.41	58.52	39.86
GATA3	-1.25	7.4E-04	36.96	10.96	25.7	13.08	10.96	5.3
GBP3	-1.19	5.1E-04	58.99	24.08	17.53	5.23	22.12	14.53
GCLM	-0.66	6.1E-04	224.81	184.04	275.62	181.33	250.32	117.48
GEMIN8P4	-1.64	1.8E-04	36.2	5.48	19.4	8.2	16.24	7.72
GFM2	-0.85	6.1E-05	293.67	107.68	207.09	150.18	116.7	73.37
GJA5	-2.95	8.6E-04	11.9	0	11.82	2.62	4.87	1.51
GLA	-1.01	4.2E-06	69.62	31.52	132.44	88.11	102.36	41.56
GLOD4	1.00	2.5E-04	18.73	63.43	42.15	60.64	40.32	71.08
GNG12	0.74	1.2E-05	118.73	198.13	124.76	202.97	156.28	269.46
GPATCH4	0.64	2.6E-04	85.82	145.66	105.46	161.24	160.88	236.01
GPM6B	1.28	1.7E-04	7.59	34.65	34.17	50.18	16.24	39.07
GPRC5C	-1.01	7.4E-04	160.51	39.74	174	169.92	94.51	46.21
H1F0	1.03	1.3E-04	18.99	68.13	62.33	87.4	41.94	76.64
HBEGF	-0.72	3.6E-04	317.72	132.74	247.95	183.12	168.12	120.69
HES1	-1.02	4.1E-04	644.81	145.27	311.96	242.33	282.38	187.78
HIST1H1D	2.34	1.2E-04	0	8.22	6.5	22.83	6.29	17.67
HIST1H1E	0.87	7.6E-04	24.56	56.39	41.56	57.31	24.76	48.96
HIST1H4E	1.17	3.2E-04	9.87	30.74	14.48	29.25	17.73	33.84
HLA-F-AS1	-1.37	7.8E-07	52.15	14.29	46.18	22.71	59.53	24.67
HLA-G	-1.03	1.7E-06	95.19	34.85	125.75	79.67	105.61	51.77
HMGA2	1.01	1.3E-05	29.37	69.7	44.8	69.68	36.87	83.38
HMGCS1	-0.76	7.7E-04	428.1	150.95	333.62	254.22	224.27	169.65
HMOX1	-2.10	9.7E-10	1423.03	586.96	2581.63	865.87	2547	219.06
HSP90AA1	-0.78	9.8E-08	10653.14	5306.31	12275.27	7437.81	7417.89	4822.53
HSP90AB1	-0.83	6.2E-07	27881.19	11441.77	19161.09	13577.29	16278.14	9847.5
HSPA1A	-1.40	1.6E-17	5350.87	1777.12	3694.36	1774.67	2651.32	895.81
HSPA1B	-1.31	3.5E-17	9036.43	3258.03	5445.18	2662.9	4599.87	1701.56

HSPA1L	-1.71	1.5E-09	214.68	34.46	84.59	41.5	84.7	29.98
HSPA2	-0.92	3.3E-06	335.7	127.85	263.02	148.75	134.9	91.5
HSPA4	-0.75	1.1E-04	2181.01	876.13	1343.74	1056.72	1396.36	922.91
HSPA4L	-0.95	7.6E-06	989.62	319.52	708.11	499.29	616.66	369.53
HSPA6	-1.35	2.1E-17	13755.91	4518.48	8603.46	4232.22	7185.97	2669.76
HSPA7	-1.22	2.7E-14	8378.71	2954.17	1208.94	610.34	1340.75	593.63
HSPA8	-0.81	2.0E-06	11210.1	4708	11747.46	8417.36	8476.66	5263.21
HSPB1	-1.11	2.1E-09	331.9	121.58	204.13	107.13	151.14	77.23
HSPB8	-1.22	7.4E-08	194.43	52.67	183.94	110.58	171.7	81.68
HSPD1	-0.92	2.7E-06	5596.95	1920.44	3526.56	2338.29	2379.63	1537.54
HSPH1	-1.02	1.3E-07	4227.58	1385.56	2533.77	1587.04	1756.54	1016.43
HTRA1	1.36	2.0E-05	13.42	72.05	56.52	87.51	18	40.45
HTRA3	-2.77	4.8E-05	22.03	0.78	11.72	3.8	8.19	1.51
IARS2	0.77	4.1E-04	49.37	123.74	57.31	83	102.09	143.79
ICMT	1.30	8.5E-04	6.33	32.3	16.44	47.92	38.16	42.74
IER5	-0.96	8.9E-09	2037.72	844.41	981.37	548.87	773.07	454.35
IGFBP3	1.37	8.0E-06	139.49	651.76	441.74	500.71	233	744.75
IL6	-0.75	5.6E-04	131.9	73.61	60.46	35.08	23.14	14.92
IL8	0.56	7.3E-04	2014.68	3612.59	3816.07	4258.74	6572.77	10532.37
IMP3	0.86	3.2E-04	32.66	61.28	26.98	54.34	38.63	61.2
IMPDH2	0.61	8.7E-04	86.33	164.46	148	201.19	157.36	219.06
IRAK1	1.12	4.1E-05	21.27	88.3	51.4	78.6	64.81	113.75
ITGA3	0.82	1.7E-04	64.56	179.73	94.04	131.99	147.82	215.85
ITGA6	0.77	7.0E-05	345.82	805.26	390.83	475.98	506.32	884.49
ITPA	0.98	4.8E-04	20.51	54.23	17.04	29.96	28.41	47.84
JAG1	0.84	4.6E-07	202.28	371.21	254.16	391.91	335.36	678.39
JUND	-0.58	6.2E-04	509.87	301.31	276.41	212.37	230.16	152.83
KANK1	0.84	1.6E-04	101.52	297.2	190.25	297.15	168.12	217.55
KCNJ2	-1.14	1.1E-04	44.05	20.17	46.48	23.66	19.35	7.66

KCNMA1	0.87	3.2E-04	17.47	45.81	48.05	70.87	57.44	95.43
KIAA1644	2.21	2.0E-04	0	2.35	2.86	9.16	6.56	27.62
KIF1A	1.38	1.3E-04	66.84	503.16	207.18	363.5	238.95	344.72
KITLG	-0.78	3.8E-04	94.43	45.42	96.11	54.58	55.88	40.12
LAD1	1.07	7.1E-05	19.49	75.57	44.02	72.89	68.74	107.99
LDHA	0.72	2.7E-04	151.9	362.59	348.49	432.58	384.61	585.51
LGALS7B	5.86	8.4E-04	0	4.11	0.1	2.73	0	1.44
LIMD1-AS1	-1.85	1.7E-04	24.81	7.05	19.89	10.58	13.4	1.77
LIN54	-1.16	6.0E-05	73.92	19.97	52.58	28.18	39.71	23.56
LINC00152	-0.82	9.0E-06	253.16	122.36	133.23	90.25	120.96	67.35
LINC00472	-0.95	3.3E-04	91.65	32.5	51.11	35.2	39.51	22.19
LINC00568	-5.10	7.9E-04	5.57	0.2	2.46	0.12	1.89	0
LOC100128239	-1.06	4.8E-04	52.4	25.84	25.01	12.13	21.72	10.01
LOC100130899	1.45	3.9E-04	4.56	30.93	14.97	33.17	16.71	27.55
LOC100506305	-1.03	8.7E-05	59.49	22.52	71.79	53.86	71.92	29.06
LRIF1	-0.98	5.8E-05	399.75	113.55	361.39	273.01	234.08	139.93
MAK16	0.85	5.0E-05	80	214.58	148.89	202.73	145.32	237.58
MAP2	-0.72	5.8E-04	115.19	93.19	131.46	76.81	77.94	37.04
MAP3K8	-0.81	7.7E-04	141.01	59.13	72.18	37.46	56.29	47.71
MBNL2	-0.87	2.3E-05	272.4	103.37	184.24	125.68	110.14	68.66
MCCC2	0.80	8.1E-04	35.95	97.11	54.75	76.93	58.93	85.48
MEX3B	-1.12	2.6E-04	57.97	19.58	45.1	20.33	17.05	10.8
MGC16275	-2.67	4.4E-04	22.78	0.2	33.68	24.02	11.64	1.96
MICB	-1.87	3.1E-06	58.48	6.46	30.72	12.01	32.34	13.61
MIR22HG	-0.92	5.4E-06	616.2	237.68	367.2	279.19	265.67	132.93
MME	-0.87	9.2E-05	192.4	67.55	159.03	104.4	89.23	61.98
MORC4	-0.74	2.5E-04	234.68	100.83	168.09	126.99	142.41	93.85
MRPL18	-1.10	1.5E-06	824.81	223.39	606.49	387.04	396.18	229.27
MRPS26	0.96	1.4E-04	18.99	47.38	32.3	63.97	43.7	66.96
MSX1	-1.31	4.6E-04	44.56	13.51	14.38	7.25	14.75	6.35

MSX2	-1.88	2.1E-04	28.1	3.33	25.7	16.53	10.62	2.49
MYBBP1A	0.84	7.8E-04	24.81	68.33	50.12	86.8	61.09	77.17
MYCN	-3.31	6.4E-04	6.08	0.59	7.88	1.66	2.44	0.07
NAP1L2	-1.30	1.5E-09	233.67	74.01	152.04	89.54	105	37.18
NBN	0.95	5.9E-04	19.75	67.35	27.47	39.48	60.21	94.57
NCAPG	-1.55	3.4E-04	36.2	8.03	13.79	7.25	12.38	4.12
NFKB1	0.70	5.3E-05	126.84	240.42	154.3	240.07	268.99	391.85
NKIRAS2	-1.02	4.2E-06	95.95	42.09	80.35	43.28	46	23.3
NOL6	0.98	5.1E-04	22.53	85.75	46.28	65.76	53.65	82.27
NPPC	-1.11	4.1E-04	48.86	17.62	38.4	18.91	14	7.72
NPY1R	-1.64	4.6E-04	43.54	4.89	31.51	21.64	15.02	5.56
NR1D1	1.04	1.0E-04	28.86	55.6	17.82	35.08	18.81	42.94
NR4A1	-0.77	6.9E-04	121.01	86.54	91.78	38.41	44.58	29.78
NUDC	-0.56	6.7E-04	530.13	353.39	610.23	370.15	319.12	247.86
NXT2	-1.33	4.2E-06	189.11	36.42	102.41	56.6	74.42	41.95
OGFOD1	0.69	7.1E-04	59.24	119.62	90.99	111.53	73.07	126.32
PAPD5	-0.71	6.4E-04	158.99	69.5	130.57	95.84	82.88	57.99
PDHB	0.79	8.5E-04	31.65	48.16	20.58	43.16	55.81	90.39
PDIA4	0.68	2.8E-04	112.15	241.99	193.99	261.36	349.02	496.5
PDK4	-0.60	7.4E-04	229.11	137.05	135.6	91.32	87.48	61.98
PEG10	-0.85	1.3E-04	228.35	84.77	141.9	110.34	91.87	54.19
PELO	-0.67	5.2E-04	98.48	63.83	87.84	59.81	130.16	74.02
PLEC	0.72	4.4E-04	418.48	1064.28	395.46	517.36	424.12	571.57
PLK2	-0.69	2.2E-04	1126.33	492.79	863.5	665.64	500.16	350.03
PM20D2	0.98	1.5E-04	40.76	137.64	50.81	71.82	41.4	69.05
PNLDC1	-1.55	5.7E-05	86.58	11.55	40.37	24.26	33.62	15.32
PNMAL1	-1.10	7.8E-04	40.51	18.99	29.64	12.84	14.27	7.07
POGK	0.95	8.6E-04	37.72	155.65	79.47	111.18	93.5	125.14
POLR2F	0.91	3.0E-04	25.57	72.44	46.08	72.65	38.22	59.3
PPP4R4	0.93	1.5E-04	17.47	49.34	31.12	42.57	85.92	160.55

PRKD2	-1.28	9.3E-07	86.84	28.98	43.62	21.88	38.83	16.17
PROM1	0.82	5.7E-04	14.18	24.47	34.47	56.48	53.45	102.3
PRRX1	0.89	2.9E-04	26.08	59.52	26.78	42.45	40.86	72.58
PTGES3	-0.68	4.7E-05	342.28	195.2	331.26	227.11	284.95	177.11
PWP2	0.93	5.6E-05	28.1	75.38	48.74	87.75	64.47	95.88
PYGB	1.03	1.9E-06	24.3	65.98	37.22	68.13	86.6	155.64
RASD1	-1.00	6.6E-04	128.1	32.5	77.1	59.1	45.87	28.34
RASL11A	-1.35	1.7E-04	41.77	10.77	34.37	21.52	17.59	6.35
RASSF9	-0.73	2.4E-04	140.25	74.01	108.42	65.04	61.36	42.61
RBPJ	1.01	5.0E-06	26.58	82.62	82.42	134.84	91.2	156.75
RCAN1	-0.62	2.9E-04	440.51	244.34	198.42	142.93	233.06	159.57
RGS2	-1.22	2.2E-06	548.86	122.17	281.63	166.35	209.52	122.78
RRAD	-0.73	5.7E-05	527.59	246.88	299.94	217.84	300.24	194.32
RRP9	0.96	5.8E-04	28.1	107.88	53.17	83.59	90.66	117.55
RUVBL1	0.85	4.0E-05	63.8	167.39	111.27	157.79	144.3	232.8
S100A16	0.80	4.2E-04	24.3	68.33	55.44	82.05	151.75	207.8
S100A2	1.07	2.4E-06	18.99	40.72	32.59	50.89	86.53	235.68
S100A6	0.69	4.0E-05	276.96	358.67	265.18	498.57	868.53	1502.73
SCD	0.69	6.7E-04	44.3	68.52	44.31	70.04	69.21	118.86
SCML1	-1.06	1.5E-04	66.84	20.95	54.26	34.13	40.93	22.32
SDPR	-1.34	1.3E-06	126.33	35.44	45.79	17.48	22.05	12.7
SEPP1	-1.22	2.7E-04	45.06	12.14	52.49	38.05	23.14	9.03
SEPT11	0.84	7.7E-04	25.06	69.7	51.7	75.98	47.09	69.25
SERPINB4	-2.35	6.7E-04	18.73	2.55	3.94	0.24	13.67	7.46
SERPINH1	-1.09	4.8E-08	671.64	209.1	628.84	359.45	513.35	293.21
SH2D5	0.99	6.4E-04	47.09	196.57	77.99	129.37	85.04	100.73
SH3BGR	-1.25	1.8E-07	152.15	49.92	68.83	41.85	59.94	22.38
SH3BP5L	-0.71	4.5E-04	204.56	94.95	95.12	73.96	92.14	58.05
SLC6A15	1.24	3.0E-04	11.9	38.37	26.78	39.12	7.64	22.32
SMYD2	1.04	1.4E-05	20.76	43.27	25.5	48.87	41.94	90.39

SNAP23	-1.02	5.7E-07	347.34	119.43	181.19	115.81	250.32	136.66
SNX16	-0.94	1.4E-04	82.03	35.44	53.27	34.36	44.04	22.38
SOCS3	-0.64	7.2E-04	265.57	133.33	163.76	114.98	127.8	95.95
SOD1	-0.60	1.5E-04	1165.31	705.02	802.54	601.78	934.22	592.65
SOX9	-1.03	4.8E-05	157.97	43.27	186.7	108.32	76.18	54.72
SPP1	-0.82	5.7E-05	39.49	24.08	265.58	168.13	25.1	11.65
SPRY2	1.08	1.5E-05	34.68	52.47	28.16	59.45	37.48	107.21
ST13	-0.67	5.9E-04	958.48	415.45	702.79	584.07	750.95	518.03
ST18	-3.99	1.0E-04	10.63	0	7.98	1.19	2.3	0.2
STIP1	-0.80	1.2E-05	1657.46	687.59	1190.92	880.5	1230.07	751.95
SYNE1	-0.62	6.9E-05	1646.83	969.32	2444.66	1807.02	531.48	339.23
SYP	1.15	3.2E-04	7.09	29.76	20.48	39.36	38.29	60.02
SYT2	-4.44	4.9E-04	8.1	1.37	2.17	0	3.86	0.07
TAOK3	-1.35	3.3E-08	176.46	46.6	98.27	60.52	100.33	37.24
TCP1	-0.80	3.8E-05	2771.13	1053.71	2032.26	1426.87	1660.07	1172.27
TFRC	0.72	2.7E-04	171.9	425.83	464.29	634.84	559.76	752.8
TGFBI	1.80	1.2E-04	1.01	17.82	10.93	22.71	22.33	51.57
TINAGL1	1.06	9.6E-05	87.59	379.04	364.15	595.72	297.4	395.71
TM4SF1	-0.57	7.2E-04	371.14	211.25	283.8	209.39	342.8	246.61
TNIP2	0.63	5.4E-04	82.78	157.02	87.15	132.22	264.73	345.71
TP63	0.78	1.9E-04	60.25	106.7	81.53	102.85	90.18	204.01
TRAF1	1.12	2.5E-04	49.11	249.23	165.53	296.55	436.02	526.02
TRIM26	-0.82	5.8E-04	103.04	37.98	79.66	57.43	84.84	56.55
TSPYL2	-0.71	2.1E-04	2495.18	1025.71	1759.98	1376.69	1449.4	1021.87
TTLL12	1.10	5.1E-04	19.75	96.33	48.84	84.42	68.4	87.57
TUBB	0.64	7.2E-04	63.29	110.42	58.39	92.27	113.72	156.49
TXNIP	0.96	6.6E-07	61.77	104.16	41.06	89.77	87.68	174.55
TXNL4A	1.04	3.9E-04	15.7	68.13	51.7	76.34	67.18	101.51
TYMP	0.84	3.5E-04	33.67	93.78	34.47	49.94	107.91	158.19
UBB	-1.04	2.8E-13	9580.23	4438.8	7471.73	3755.77	6624.18	3298.67

UBC	-1.23	1.1E-15	6832.89	2493.3	5797.51	2843.52	4436.08	1925.2
USPL1	-1.05	2.6E-06	226.58	73.61	138.65	69.92	90.72	61.78
VEGFA	0.84	2.3E-06	283.29	617.7	386.01	531.27	436.09	831.93
VTRNA1-2	-1.49	2.1E-06	58.48	15.08	63.12	34.36	14.75	4.58
WBP5	-0.69	2.0E-04	1462.78	644.71	936.96	691.56	671.46	490.68
WDR3	0.83	3.6E-05	146.58	384.71	253.56	362.66	315.06	474.64
WDR75	0.59	7.4E-04	94.18	146.45	121.71	190.96	135.44	189.94
XPC	0.75	3.4E-04	56.96	137.05	86.36	134.84	113.79	148.51
ZFAND2A	-1.26	6.4E-10	531.9	150.95	258.29	126.87	216.29	111.13
ZFP106	-0.64	8.4E-04	394.94	182.67	296.6	230.32	185.71	134.96
ZMYND11	0.64	2.0E-04	150.13	278.8	217.43	326.63	193.89	267.62
ZNF131	-0.76	6.7E-05	568.61	240.42	313.53	217.48	331.43	230.78
ZNF184	-0.84	6.7E-04	66.84	26.63	71	42.93	53.78	37.5
ZNF259	0.80	1.5E-04	79.75	209.29	153.91	241.02	189.7	248.12
ZP4	1.34	6.9E-04	7.85	17.03	6.11	14.98	9.2	27.49

Table 4.9. The most differentially expressed genes (q<0.05) between curcumin and piperine treated vs. curcumin only treated ALDH+ or ALDH-CD44+CD24- cells.

Gene	logFC	P Value	Sample 1 CPMs		Sample 2 CPMs		Sample 3 CPMs	
			Curcumin	Curcumin + Piperine	Curcumin	Curcumin + Piperine	Curcumin	Curcumin + Piperine
Genes differentially expressed in ALDH+								
RPL26	0.27	1.6E-07	5662.93	7132.06	3326.19	3839.2	3127.85	3779.93
UBB	0.24	1.8E-06	10310	11863.31	2557.14	2872.6	2526.8	3197.72
CALHM2	7.93	3.4E-05	0	9.78	0	2.43	0	5.19
TPT1	0.27	3.5E-05	3989.86	5071.43	2544.94	2777.75	2599.52	3290.92
RPL32	0.44	8.5E-05	660.89	977.53	420.49	498.46	346.04	497.47
PLEKHG4	-1.85	1.1E-04	28.75	9.78	17.77	6.72	10.92	1.81
DMD	-0.81	1.2E-04	81.87	44.97	57.77	35.05	58.1	31.71
RPL31	0.36	1.4E-04	1232.89	1640.3	638.02	746.84	632.63	854.24
Genes differentially expressed in ALDH-CD44+CD24-								
HSPA6	0.15	9.5E-12	11720.74	13469.66	8051.91	8361.56	6086.89	7010.39
HSPA1B	0.14	1.9E-08	7917.2	8848.39	4627.91	5292.08	4239.09	4487.47
HSP90AB1	0.06	1.6E-07	26292.03	27301.01	18107.64	18622.34	14890.38	15880.39
HSPA1A	0.18	8.4E-06	4869.36	5239.52	2958.74	3590.48	2321.41	2586.54

Table 4.10. The most significantly enriched Gene Ontology Biological Processes identified comparing expression between curcumin and piperine co-treated and curcumin only treated ALDH+ and ALDH-/CD44+/CD24- cells.

Concepts enriched in ALDH+				
Name	#Genes	OddsRatio	P-Value	FDR
translational elongation	98	70.47	4.3E-39	1.1E-35
translation	372	6.69	3.1E-19	4.1E-16
ribosomal small subunit biogenesis	13	50.46	4.2E-09	3.7E-06
hydrogen transport	46	9.73	9.1E-06	4.1E-03
ribosomal large subunit biogenesis	10	33.24	1.2E-05	4.1E-03
oxidative phosphorylation	84	6.18	1.2E-05	4.1E-03
ATP synthesis coupled proton transport	34	11.85	1.4E-05	4.1E-03
energy coupled proton transport, down electrochemical gradient	34	11.85	1.4E-05	4.1E-03
proton transport	44	9.71	1.4E-05	4.1E-03
Concepts enriched in ALDH-CD44+CD24-				
Name	#Genes	OddsRatio	P-Value	FDR
protein refolding	11	18.00	3.3E-08	9.0E-05
negative regulation of proteolysis	21	10.01	8.8E-07	1.2E-03
response to unfolded protein	61	6.52	1.4E-06	1.3E-03
response to protein stimulus	102	4.93	4.1E-06	2.9E-03
calcium ion transport	131	3.69	9.3E-05	4.2E-02
regulation of cytokine-mediated signaling pathway	22	7.10	1.0E-04	4.2E-02
di-, tri-valent inorganic cation transport	168	3.32	1.1E-04	4.2E-02
negative regulation of growth	120	3.75	1.3E-04	4.4E-02
cellular amino acid catabolic process	53	5.08	1.5E-04	4.5E-02
divalent metal ion transport	138	3.37	2.8E-04	7.3E-02

CHAPTER 5. CONCLUSION

5.1 Overview.

The epidemiological distribution of many cancers, including head and neck and breast, is shifting worldwide. In the United States, the epidemiology of head and neck cancer has undergone a radical shift in the past decade, with a rapid increase in the incidence of oropharangeal cancers associated with HPV infection. With respect to breast cancer, incidence rates have more than doubled between 1980 and 2010 in many developing countries. Due to the stable nature of the human genome over this time period, these changing cancer rates likely reflect the strong environmental role in these diseases. How environmental factors influence carcinogenesis for many types of cancer, however, remains poorly understood. Additionally, mechanisms by which dietary factors influence cancer development and pathogenesis remain poorly understood.

Many epidemiological studies of environmental carcinogenesis or characterizing the role of diet in cancer prognosis poorly quantify previous exposures. With respect to environmental epigenetic epidemiology, many studies often either poorly characterize environmental exposures or focus on a limited number of epigenetic target sites. Studies designed to characterize the mechanisms of cancer chemopreventive compounds utilize immortalized or cancer cell lines, which may not recapitulate the effects of these compounds in normal human cells. Developing novel methodology to study the role of nutrition and the environment in carcinogenesis at relevant time stages provides essential insight towards the prevention, early identification, and treatment of cancer. Incorporating novel culture methods, including primary tissue culture, will allow for the study of environmental and nutritional factors in specific normal cell populations that may be at increased risk for cancer development, including stem cells.

Epigenetic changes are now recognized as an important mechanism in carcinogenic progression. These modifications are environmentally labile, vary over the lifetime, and are potentially reversible, making them an attractive target for cancer prevention and treatment. Additionally, epigenetic modifications control the gene expression required to establish and maintain each of the cell types in the body, establishing cellular identity during and throughout development, as cells progress from pluripotent, to multipotent, to fully differentiated.

The overarching goal of this dissertation was to develop and apply novel statistical and experimental methods to characterize the roles of nutrition and the environment in carcinogenesis and cancer exposure, with a focus on epigenetic change. In Chapter 2, comprehensive epidemiological and clinical information was paired with comprehensive DNA methylation profiling of head and neck tumors to identify significant differences in tumor DNA methylation in chemically induced or HPV induced tumors. In Chapter 3, comprehensive data on average dietary intake was paired with tumor epigenetic measurements to identify that a patient's diet in the year before diagnosis can significantly affect tumor epigenetic profiles, providing a potential mechanism by which diet affects disease prognosis. In Chapter 4, we treated normal human breast stem cells from reduction mammoplasty tissue with the cancer prevention compounds curcumin and piperine, and conducted a genome-wide screen to identify the stem cell specific changes induced by these compounds. The results and methods presented here reflect the utility of these methods, from cancer molecular epidemiology to normal human *in vitro* stem cell culture, to understanding the role of the environment in cancer.

5.2. Chapter 2 - Comprehensive analysis of DNA methylation in head and neck squamous cell carcinoma indicates differences by survival and clinicopathologic characteristics.

Head and neck squamous cell carcinoma (HNSCC) is the eighth most commonly diagnosed cancer in the United States. The risk of developing HNSCC increases with exposure to tobacco, alcohol and infection with human papilloma

virus (HPV). HPV-associated HNSCCs have a distinct risk profile and improved prognosis compared to cancers associated with tobacco and alcohol exposure (Fakhry et al. 2008; Gillison et al. 2008). Epigenetic changes are an important mechanism in carcinogenic progression, but how these changes differ between viral- and chemical-induced cancers remains unknown. We previously reported an epigenome-wide analysis of concurrently measured DNA methylation and gene expression in HPV(+) and HPV(-) squamous cell carcinoma cell lines, noting that HPV(+) cell lines have higher amounts of genic methylation (Sartor et al. 2011). In Chapter 2, to translate and extend the findings in cell lines to patients, we conducted an epigenetic epidemiology study in a well-characterized cohort of 68 patient samples.

Unsupervised hierarchical clustering based on methylation identified 6 distinct tumor clusters, which significantly differed by age, HPV status, and three year survival. Weighted linear modeling was used to identify differentially methylated genes based on epidemiological characteristics. Consistent with previous *in vitro* findings by our group, methylation of sites in the *CCNA1* promoter was higher in HPV(+) tumors, which was validated in an additional sample set of 128 tumors. After adjusting for cancer site, stage, age, gender, alcohol consumption, and smoking status, HPV status was a significant predictor for DNA methylation at an additional 11 genes, including *CASP8* and *SYBL1*.

In this study we investigated the likelihood of identifying a clinically relevant subset of head and neck tumors defined by CpG methylation, taking advantage of a well-established patient cohort at the University of Michigan with well-annotated survival and epidemiologic data. Our sample was representative of the overall cohort regarding age, gender, smoking history, and alcohol consumption. We examined the epigenetic differences between HPV(+) and HPV(-) tumors, following from our recent work in cell lines showing evidence for divergent pathways of carcinogenesis and the well-described epidemiologic differences between individuals with differential HPV tumor status (Sartor et al. 2011). Further, we were able to evaluate survival in this cohort in light of their epigenetic profile (as defined by cluster status), HPV status and other

epidemiologic characteristics. These findings provide insight into the epigenetic regulation of viral vs. chemical carcinogenesis and could provide novel targets for development of individualized therapeutic and prevention regimens based on environmental exposures and epigenetic profiles of tumor DNA.

5.3 Chapter 3 - Pretreatment dietary intake is associated with tumor suppressor DNA methylation in head and neck squamous cell carcinomas.

Epidemiologic evidence supports the hypothesis that diet modulates risk, progression and prognosis of head and neck cancer (Duffy et al. 2009; Lucenteforte et al. 2009; Sandoval et al. 2009). The molecular mechanisms by which dietary compounds exert their effects are not entirely understood. Epigenetic dysregulation is a key mechanism in tumorigenesis that may be influenced by dietary intake by determining the availability of functional groups involved in the covalent modification of DNA and histone proteins (Burdge et al. 2007; Oommen et al. 2005). The objective of the study reported in Chapter 4, was to test the hypothesis that pretreatment dietary intake of methyl donors, antioxidants, and foods abundant with these micronutrients is associated with tumor DNA methylation in head and neck tumors isolated from 49 individuals.

We found that individuals reporting in the highest quartile of folate, vitamin B12, and vitamin A intake, compared to those in the lowest quartile, had significantly less tumor suppressor gene methylation, as did patients reporting the highest cruciferous vegetable intake. Gene specific analyses, identified differential associations between DNA methylation and vitamin B12 and vitamin A intake when stratifying by HPV status.

To our knowledge, this was the first study to comprehensively examine the association between dietary intake and promoter methylation of genes related to head and neck cancer. These novel findings suggest diet is significantly associated with epigenetic events that occur in head and neck cancer and may provide a mechanism by which dietary micronutrient intake influences cancer prognosis. These findings have potential clinical implications for the treatment of head and neck cancer. The idea that dietary interventions could potentially reprogram the epigenome in such a way as to optimize the likelihood of positive

disease outcomes is appealing. Such studies could give rise to highly specific randomized controlled trials and ultimately to the development of individualized medical nutrition therapy regimens that may improve prognosis in the head and neck cancer population.

5.4 Chapter 4 - Transcriptomic effects of curcumin and piperine in normal human breast stem cells.

Curcumin is a dietary polyphenol derived from the rhizomes of turmeric (*curcuma longa*) which has been implicated as a potential agent for both the prevention and treatment of cancers. Recently, we showed that curcumin treatment alone, or in combination with piperine, limited breast stem cell self-renewal while remaining non-toxic to normal differentiated cells (Kakarala et al., 2010). In Chapter 4, to extend these findings and identify the stem cell specific mechanisms of action of these compounds, we characterized the genome-wide changes induced specifically in normal breast stem cells following treatment with these compounds by pairing fluorescence activated cell sorting (FACS) with low input high throughput RNA sequencing (RNA-seq).

We generated genome-wide maps of the transcriptional changes that occur in luminal (ALDH+) and basal (ALDH-/CD44+/CD24-) normal breast stem cells following treatment with curcumin and piperine. Our results confirm that these compounds target breast stem cell self-renewal in both stem cell populations by down-regulating expression of breast “stemness” genes including *ITGA6*, *PROM1*, *ALDH1A3*, and *TP63*. Curcumin treatment was also found to significantly increase unfolded protein response, heat shock, and oxidative stress response genes in both stem cell fractions. Additionally, we identified novel genes and pathways targeted by curcumin in these cells, including mechanisms by which curcumin may target Wnt signaling in breast stem cells that have not been previously described, including downregulation of *SCD* and upregulation of *CACYBP*.

These findings help clarify the mechanisms by which curcumin and piperine function as cancer preventive compounds, providing novel targets for cancer chemoprevention and treatment efforts. The use of primary tissues

provides novel information about stem cell regulation in the normal human breast that may not be available from studies utilizing cell lines or model animals. This experimental technique can thus be applied to understand the effects of carcinogens or cancer preventive compounds in specific cell populations. Here, we identify the pathways and individual genes differentially expressed by curcumin treatment, pointing to novel mechanisms of inhibiting stem cell self-renewal and potential biomarkers of curcumin efficacy in clinical trials.

5.5. Public health implications and future research needs.

The public health burden of cancer is estimated to increase worldwide throughout the 21st century, with the majority of the burden in the developing world. As populations age and the rates of mortality from infectious disease decrease, rates of chronic diseases, including cancer, will concurrently increase. Thus, the identification of novel methods of prevention, early detection, and treatment of cancer will become imperative. Of particular importance will be the characterization of environmental and dietary risk factors for cancer and understanding how these interact with genotype to influence disease risk.

With respect to epigenetic differences between HPV(+) and HPV(-) tumors, a growing body of research, including the research presented here, shows dramatic differences between these tumor types. Interestingly, the hypermethylation profile observed in HPV(+) tumors is also generally, but not always, associated with a decreased somatic mutation rate. Future research that combines high throughput epigenomic profiling and genomic mutation analysis across tumors, paired with comprehensive epidemiological data, will identify biomarkers that predict whether tumors are more or less responsive to treatment. The field of nutritional epigenetics, particularly with respect to cancer prognosis, is still in its infancy. Mechanisms by which dietary intake influence epigenetic profiles in people are still poorly understood. For example, in Chapter 3, we identified that increased intake of dietary micronutrients associated with one carbon metabolism is associated with a *decrease* in tumor suppressor DNA methylation. These findings suggest that epigenetic regulation based on micronutrient availability is a complex process involving a number of feedback

loops. How these pathways differ in tumors and normal cells is not understood, but could provide insight into dietary prevention of cancer and modulation of tumor epigenetic profiles by diet.

Finally, characterizing how environmental and dietary factors specifically affect breast stem cells will be essential to understand the molecular origins of breast cancer. We also know, however, that mammary stroma also plays an essential role in the maintenance of the stem cell niche. Future work could investigate the role of toxicant exposed stroma, including adipose and fibroblasts, on stem cell self-renewal in an effort to build an organ level model of disease. Integrating molecular toxicology and epidemiology by culturing tissues isolated from epidemiologically and clinically characterized individuals will allow us to define the molecular basis for known breast cancer risk factors with respect to stem cell regulation. Additionally, isolating and exposing tissues from women with known breast cancer risk factors to carcinogens will allow us to better characterize factors that govern environmental susceptibility. We will also be able to study the interplay of environmental exposures, diet, and stem cell regulation by incorporating nutritional epidemiology and exposure assessment methodologies. These studies will define new markers for the screening of at risk populations and will provide mechanistic insight for the early detection, chemoprevention, and treatment of breast cancers.

5.6 References

Burdge GC, Hanson MA, Slater-Jefferies JL, Lillycrop KA. 2007. Epigenetic regulation of transcription: A mechanism for inducing variations in phenotype (fetal programming) by differences in nutrition during early life? Br J Nutr 97:1036-1046.

Duffy SA, Ronis DL, McLean S, Fowler KE, Gruber SB, Wolf GT, et al. 2009. Pretreatment health behaviors predict survival among patients with head and neck squamous cell carcinoma. J Clin Oncol 27:1969-1975.

Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. 2008. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 100:261-269.

Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. 2008. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst 100:407-420.

Kakarala M, Brenner DE, Korkaya H, Cheng C, Tazi K, Ginestier C, et al. 2010. Targeting breast stem cells with the cancer preventive compounds curcumin and piperine. Breast Cancer Res Treat 122:777-785.

Lucenteforte E, Garavello W, Bosetti C, La Vecchia C. 2009. Dietary factors and oral and pharyngeal cancer risk. Oral Oncol 45:461-467.

Oommen AM, Griffin JB, Sarath G, Zempleni J. 2005. Roles for nutrients in epigenetic events. J Nutr Biochem 16:74-77.

Sandoval M, Font R, Manos M, Dicenta M, Quintana MJ, Bosch FX, et al. 2009. The role of vegetable and fruit consumption and other habits on survival following the diagnosis of oral cancer: A prospective study in spain. Int J Oral Maxillofac Surg 38:31-39.

Sartor MA, Dolinoy DC, Jones TR, Colacino JA, Prince ME, Carey TE, et al. 2011. Genome-wide methylation and expression differences in hpv(+) and hpv(-) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. Epigenetics 6:777-787.