**Molecular Characterization of Esophageal Adenocarcinoma and Factors Influencing Racial Differences in Incidence**

by

Daysha Ferrer-Torres

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cancer Biology)
in the University of Michigan
2017

Doctoral Committee:

Professor David G Beer, Chair
Professor Arul M. Chinnaiyan
Professor Elizabeth Lawlor
Professor Thomas Wang

Daysha Ferrer-Torres

dferrert@umich.edu

ORCID iD:
0000-0002-3576-0347

## DEDICATION

**To my parents**

**Rosario Torres and David Ferrer,**

**for believing in my dreams, my potential, and supporting every step of my career.**

**My success is your success.**

## ACKNOWLEDGEMENTS

First and foremost, to David G. Beer, thank you for your constant support, and for providing the opportunity to be part of an amazing work team. In my five years in the lab, you provided support for my growth as a scientist and provided a foundation that I know will help me be successful in my future endeavors. Thank you for your patience and guidance throughout the years. In addition, thank you to my thesis committee members Arul Chinnaiyan, Elizabeth Lawlor, and Thomas Wang, for being an integral part of my development and education.

To the past members and current members in the Beer lab, especially to Derek Nancarrow and Zhuwen Wang, thank you for providing daily advice and mentorship. Your assistance and guidance in experimental procedures as well as discussions of ideas, results, and presentations helped me achieve the goals I present here today. For that, I am forever thankful.

To Dipankar Ray and Paramita Ray, in the past year, your advice and guidance have been fundamental to my scientific success. Your hard work, dedication, motivation, and attention to experimental detail are an inspiration for me to follow.

The Cancer Biology Program, faculty, and students, who provided much insight and help to guide some of the technical aspects and provided valuable perspective into results of my project. To Ben Chandler and Anjan Saha, thank you for the laughs and the guidance in writing and figure making.

To the amazing support group of friends, that also happen to be all amazing Ph.D. women in science; Shayna Bradford, Jing Ping, Anna Hung-An, and especially Justine Pinskey. Thank you for listening and giving me advice during the good and the bad days in science and life.

Finally, to my family, I am forever grateful for all your endless support in helping me reach my career and personal goals. To my parents, this Ph.D. was possible because of all the sacrifices you made throughout your life for me to be here and finish my dream of becoming a scientist. To my son, Raymar M Santiago-Ferrer, everything I do is for you, thank you for your patience and kindness.

## FINANCIAL ACKNOWLEDGEMENTS

I will like to acknowledge the many sources of funding that made this research possible. First, to the Rackham graduate school for accepting me into the Ph.D. program with a Rackham Merit Fellowship. To the Cancer Biology program-training grant, the Ruth L. Kirschstein National Research Service Award (NRSA) (T32)-Cancer Research Training Grant-National Cancer Institute (T32CA009676), which provided support throughout my first 2 years in the program. To Rackham Graduate School for the Student Research Pre-doctoral grant and Conference Travel Grant. To the National Cancer Institute for their support through a Ruth L. Kirschstein National Research Service Award (NRSA) Individual Predoctoral Fellowship (F31CA200113-02). To Yale Ciencia Academy, for providing me a one-year career development fellowship. Finally, to the support provided by the BETRNet grant awarded to Dr. David G. Beer and others (U54 CA163059 NIH/NCI).

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF APPENDICES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ESCC | Esophageal squamous cell carcinoma |
| GERD | Gastro-esophageal reflux disease |
| BE | Barrett's Esophagus |
| ND-BE | Non-dysplastic Barrett's Esophagus |
| LGD | Low-grade dysplasia |
| HGD | High-grade dysplasia |
| EAC | Esophageal Adenocarcinoma |
| GEJ | Gastro-esophageal junction |
| tEAC | Tubular Esophageal Adenocarcinoma |
| GEJAC | Gastro-esophageal junction adenocarcinoma |
| SD | Standard Deviation |
| FDR | False discovery rate |
| PCA | Principal Component Analysis |
| NE | Normal esophagus (squamous mucosa) |
| NG | Normal gastric |
| C1 | Cluster 1 |
| C2 | Cluster 2 |
| ANOVA | Analysis of variance |
| ROC | Receiver operating characteristics |

| | |
|---|---|
| GO | Gene ontologies |
| GEO | Gene Expression Omnibus |
| qRT-PCR | Quantitative Reverse Transcription Polymerase Chain Reaction |
| PCR | Polymerase Chain Reaction |
| TMA | Tissue microarray |
| Cau | Caucasian (European American) |
| AA | African American |
| AC | Adenocarcinoma |
| GAPDH | Glyceraldehyde-3-Phosphate Dehydrogenase |
| ZNF217 | Zinc Finger Protein 217 |
| ICAM1 | Intercellular Adhesion Molecule 1 |
| CDH11 | Cadherin 11 |
| CLND3 | Claudin-3 |
| RPKM | Reads Per Kilobase of transcript per Million mapped reads |
| AB-PAS | Alcian blue-Periodic acid-Schiff reagent |
| MUC5B | Mucin 5B, Oligomeric Mucus/Gel-Forming |
| MUC3A | Mucin 3A, Cell Surface Associated |
| MUC17 | Mucin 17 |
| DDR | DNA-damage response |
| ATM | ATM Serine/Threonine Kinase |
| MRE11 | MRE11 Homolog, Double Strand Break Repair Nuclease |
| RAD50 | RAD50 Double Strand Break Repair Protein |
| P53BP1 | Tumor Protein P53 Binding Protein 1 |

| | |
|---|---|
| γ-H2AX | Phosphorylated Histone H2AX |
| C3 | Complement C3 |
| HLA-DRB5 | Major Histocompatibility Complex, Class II, DR Beta 5 |
| FC | Fold-Change |
| GSTT2 | Glutathione S-Transferase Theta 2 (Gene/Pseudogene) |
| GSTT2B | Glutathione S-Transferase Theta 2 (Gene/Pseudogene) |
| DDT | D-Dopachrome Tautomerase |
| GSTTP1 | Glutathione S-Transferase Theta Pseudogene 1 |
| GSTT1 | Glutathione S-Transferase Theta 1 |
| PPARG | Peroxisome Proliferator Activated Receptor Gamma |
| Chr | Chromosome |
| kb | Kilo-base |
| bp | Base-pair |
| kd | Kilo-dalton |
| μM | Micro-molar |
| CN | Copy number |
| D and/or DUP | Duplicated |
| P | Promoter |
| DEL | Deletion |
| AFR | Africa |
| EUR | European |
| CDX | Chinese Dai in Xishuangbanna, China |
| CHB | Han Chinese in Bejing, China |

| | |
|---|---|
| JPT | Japanese in Tokyo, Japan |
| KHV | Kinh in Ho Chi Minh City, Vietnam |
| CHS | Southern Han Chinese, China |
| BEB | Bengali in Bangladesh |
| GIH | Gujarati Indian in Houston, TX |
| ITU | Indian Telugu in the UK |
| PJL | Punjabi in Lahore, Pakistan |
| STU | Sri Lankan Tamil in the UK |
| ASW | African Ancestry in Southwest US |
| ACB | African Caribbean in Barbados |
| ESN | Esan in Nigeria |
| GWD | Gambian in Western Division, The Gambia |
| LWK | Luhya in Webuye, Kenya |
| MSL | Mende in Sierra Leone |
| YRI | Yoruba in Ibadan, Nigeria |
| GBR | British in England and Scotland |
| FIN | Finnish in Finland |
| IBS | Iberian populations in Spain |
| TSI | Toscani in Italy |
| CEU | Utah residents with Northern and Western European ancestry |
| CLM | Colombian in Medellin, Colombia |
| MXL | Mexican Ancestry in Los Angeles, California |
| PEL | Peruvian in Lima, Peru |

| | |
|---|---|
| PUR | Puerto Rican in Puerto Rico |
| AMR | Americas |
| SAS | South Asia |
| ESA | East Asia |
| Cum-OOH | Cumene-Hydroperoxide |
| NT | Non-target |
| siRNA | Small (or short) interfering RNA |
| DAPI | 2-(4-amidinophenyl)-1H -indole-6-carboxamidine |
| DMSO | Dimethyl sulfoxide |

**Cell Lines**

| | |
|---|---|
| HET1A | Esophagus, Epithelial; Sv40 Large T Antigen Transfected, Normal |
| HeLa | Cervix, Epithelial, Adenocarcinoma |
| UM-SCC-29 | Head and Neck, squamous |
| UM-SCC-11A | Head and Neck, squamous |
| UM-SCC-11A | Head and Neck, squamous |
| 12-oug4 | Head and Neck, squamous |
| UM-SCC-6 | Head and Neck, squamous |
| SK-MES | Lung; Derived From Metastatic Site: Pleural Effusion, squamous cell cancer |
| FLO | Esophagus, Esophageal Adenocarcinoma |
| OE33 | Esophagus, Esophageal Adenocarcinoma |
| HCT116 | Colon, epithelial, Colorectal Carcinoma |

## ABSTRACT

The incidence of esophageal adenocarcinoma (EAC) has been rising at an alarming rate over the past three decades (>450-fold). Identification of patients with the metaplastic condition termed Barrett's esophagus (BE), is a risk factor for the development of EAC, yet allows patients to undergo endoscopic surveillance biopsy to help detect cancer. Our group and others have identified markers that can potentially identify high-grade dysplasia (HGD) and/or EAC in BE patients. However, we have observed that >30% of patients with cancers at the gastroesophageal junction (GEJ) have no previous history of BE. These cancers are also similarly increasing in incidence thus new approaches are also needed for these patients. Interestingly, the increasing incidence of EAC primarily affects the Caucasian population as compared to the African American (AA) population. **Therefore, in this thesis we aim to understand: 1) whether EAC vs. tumors located at the GEJ are molecularly distinct, 2a) the characterization of molecular events at the transcriptome level associated with the progression from non-dysplastic BE to dysplasia to cancer 2b) as well as biomarkers that identify patients that are at greatest risk for development of EAC and 3) understanding the basis for the difference in the incidence of EAC between CAU and AA.** We were able to show that EAC and GEJAC tumors are molecularly similar, and identified cell surface markers that can be used to detect both non-BE derived GEJs and EACs. Using transcriptional analysis of BE, dysplasia, and EAC we identified splicing has a key dysregulated pathway in BE progression. In addition, we observed an increase in the ATM/DNA-response damage in this progression that, we assume, is associated with loss of protective mucins. Importantly, we have identified a novel genomic

event that reduces the expression of the detoxifying enzyme GSTT2 in the esophagus of Cau populations as compared to AA populations. This body of work adds to our understanding of molecular events in EAC and GEJ and could potentially impact the way we diagnose and treat these cancers in the future.

# CHAPTER 1

## Introduction: Esophageal Cancers; Incidence, Histology, Treatment, and Progression

### Esophageal Cancer

Cancer is the second leading cause of death in the United States (US), with 14.1 million new cases diagnosed globally[1]. In the US, approximately 42% of all cancers cases in men are related to prostate, lung, and colorectal cancers. Similarly, women's most common cancers diagnose include breast, lung, and colorectal[1]. In contrast, esophageal cancers (EC) are not common, with approximately 16,940 new cases diagnosed in 2017[1]. Nevertheless, EC are a significant worldwide health problem because of their poor prognosis[1,2], and they are the seventh leading cause of death in men when compared to other cancers[1]. There are two types of esophageal cancers: esophageal squamous cell cancer (ESCC) and esophageal adenocarcinoma (EAC). It is estimated that in 2017, 16,940 people in the US will be diagnosed with esophageal cancer and greater than 90% of those diagnosed will succumb to the disease[1]. There have been advances in surgical techniques, chemotherapy, and radiotherapy; however, these methods have not significantly modified patient prognosis over the past decades[3–9]. Diagnosis of EAC primarily occurs when it is at an advanced stage, resulting in an overall poor 5-year survival rate of less than 15%[10]. ESCC main risk factors are smoking[11], alcohol[12], and socioeconomic factors[13,14]. Due to large efforts in reducing the smoking incidence, ESCC incidence has significantly decreased over the past decades and continues to decrease[15]. On the contrary, EAC is associated with the risk factors that include; a history of gastro-esophageal reflux disease

(GERD), obesity, Caucasian race, and age over 40[16]. Obesity is becoming a global epidemic, and obesity is associated with the rise in GERD[17], and EAC[18–22]. Although the incidence of EAC is increasing among most developed countries[18,23], there is a significant difference in this incidence by race[1,22,24], with African Americans having a low incidence of this type of cancer[1,22,25,26].

Many patients with EAC have associated with it Barrett's esophagus (BE), a predisposing condition where the normal squamous of the esophagus is replaced by an intestinal-type columnar tissue[27]. Only about 1% of patients with BE will develop EAC[28]. The prevalence of BE is also significantly different between racial populations, with African Americans and Hispanics having a significantly lower frequency of BE when compared to Caucasians[29–32]. Interestingly, when looking at the common risk factors associated with development of EAC, such as obesity and GERD, there are no significant differences between Caucasians and African Americans[29–32]. The reasons for the high incidence of EAC in the Caucasian population vs. African Americans are currently unknown.

Understanding the progression of Barrett's metaplasia to EAC and discovery of biomarkers that could stratify these patients based on the potential risk for developing EAC could potentially have a high impact on patient survival. In this introduction, we focus on the epidemiology, clinical features and molecular mechanisms that have been discovered for new diagnosis and therapeutic approaches for BE and EAC. We also discuss potential new methods to explore and risk stratify patients based on their potential to develop EAC from Barrett's metaplasia.

## Esophageal Adenocarcinomas

### Barrett's history and progression to EAC.

The metaplastic epithelium called Barrett's esophagus (BE) was first described by Norman Barrett where he showed the apparent replacement of the normal squamous epithelium of the esophagus with a columnar mucosa that resembles intestinal epithelium[33,34]. He initially thought this was a congenital condition but it was later shown that it was acquired. GERD was later shown to be associated with Barrett's development[35] and in 1975 a study of 140 cases showed that 8.5% of those patients with BE developed EAC[36]. Only about 10-15% of patients with chronic GERD developed BE, this suggests that further genetic and environmental factors are possibly involved in the development of BE and/or EAC[37].

The process of neoplastic transformation from BE to EAC is thought to be a stepwise process, which involves a transition to low-grade dysplasia (LGD) to high-grade dysplasia (HGD) and finally to EAC (**Figure 1.1**). BE is considered an acquired premalignant lesion of the esophagus, although the exact mechanisms underlying its development and progression are still not understood. Some studies have shown that bile acids can cause injury in the esophageal squamous lining and lead to the development of metaplasia by inducing oxidative stress and DNA damage[38,39]. In addition, patients with central obesity are more predisposed to suffer from increased intra-gastric pressure, which increases gastroesophageal reflux disease (GERD), and subsequent are at higher risk for EAC[40,41]. Nevertheless, up to 40% of patients report no history of chronic GERD and develop EAC[42]. Studies trying to understand the progression from BE to EAC using expression profiles compared to normal tissue, have suggested different pathways activated in BE, but the cell of origin of BE remains known[43]. Patients who suffer from BE are enrolled in surveillance programs where they undergo biopsies of the BE by endoscopy to detect dysplasia by histopathology[44–48]. The process includes identification of either non-dysplastic Barrett's (ND), low-grade dysplasia (LGD), high-grade dysplasia (HGD) and/or EAC[49],

however, some patients with BE under endoscopic surveillance often develop cancer without prior biopsy-based detection of any of these previously mentioned premalignant tissues (**Figure 1.3**).

**GEJAC vs. EAC.**

The appropriate distinction between adenocarcinomas of the esophagus and the stomach has long been debated[50-51] (**Figure 1.2**). The appropriate classification of these tumors that span the gastroesophageal junction (GEJ) remains ambiguous. Due to this uncertainty, most treatment options and therapeutic approaches have relied on the use of anatomy and histology to classify this tumor types[50,52,53] (**Figure 1.2**). One important aspect for elucidating the differences between these tumor types will be to understand their molecular characteristics, which in turn, may help classify each tumor type. Our group and others have focused on the molecular characterization of gastric adenocarcinomas[54], esophageal squamous cell cancer[55], and esophageal adenocarcinomas[55]. The Tumor Cancer Genome Atlas (TCGA) network, performed comprehensive molecular analysis of esophageal cancers subtypes revealing strong molecular signatures that are unique to each type. For example, ESCC is characterized by frequent genomic amplifications of *CCND1*, *SOX2* and *TP63*[55]. On the other hand, EAC is characterized by more *ERBB2, VEGFA, GATA4*, and *GATA6* amplifications. At the molecular level, EAC resembles more of a genomic profile similar to gastric adenocarcinoma, more specifically, those with similar chromosomal instability[55]. The observation that gastric and EAC share characteristics at the molecular level, suggests that they could be considered a single disease entity for future treatment strategies. In addition, others and our laboratory and collaborators have worked on the development of early cancer detection imaging approaches[56–61]. Specifically, the development of

fluorescently-labeled peptide-based imaging agents to enhance endoscopic detection of specific cell surface markers as a means to improve early stage EAC diagnosis[61,62].

EAC arises from the precursor lesion, Barrett's esophagus (BE). Patients diagnosed with BE are enrolled in surveillance programs, where they undergo routine biopsy of the BE tissue to identify dysplastic progression and help to determine appropriate treatment. One of the current problems is the observation that >30% of GEJAC tumors present with no history or evidence of BE[63,64]. Making this group of patients, unlikely to be enrolled in surveillance programs. Therefore, understanding whether BE-derived EAC and non BE-derived GEJAC are molecularly distinct, as well as identification of cell surface markers for GEJAC, will potentially enhance early diagnostic tools and potential treatment for this patient cohort.

**Comprehensive genomic analyses of esophageal cancers.**

At the molecular level, genomic instability has been shown to be a fundamental property of the progression of the BE to higher grades of dysplasia and cancer[65–69]. Whole genome analysis studies have shown that BE samples show low levels of chromosomal instability, which include events such as copy number gains, losses, and loss of heterozygosity (LOH)[70]. The extent of genetic abnormalities correlates with the degree of dysplasia with BE showing less than 2% and increasing to greater than 30% in later dysplastic stages, and tumors having the highest percentage of genomic abnormalities. The most frequent genetic alterations are a loss of chromosome 9p, *TP53* mutations, *APC* and *RB* mutation and loss and overexpression of cyclin D1, Bcl2, and Src[71,69,72]. For many years it was thought that, at the genomic level, esophageal adenocarcinoma progresses through a stepwise progression where loss of *CDKN2A* was first observed followed by inactivation of *TP53* with further increases in genome instability and aneuploidy. A recent study, however, challenges that idea by identifying that genomic instability

is observed later, with *TP53* being an early event, even identified in benign tissue (histologically benign non-dysplastic BE tissue)[73]. The majority of the tumors showed a genome-doubling event, with more amplification events observed, instead of inactivation of tumor suppressors. These key findings suggest EAC does not always progress in the common linear progression once thought, and that careful characterization of the *TP53* gene and genome doubling could function as a potential early detection strategy[73].

Our group and others have extensively characterized EAC at the DNA level[71,74–83,67,66]. More in-depth, we have performed a comprehensive analysis of copy number gain and loss[69], as well as examining the mutational profile[68] of EAC. EAC is characterized as having high copy-number gain and loss, with recurrent focal amplifications centered near known oncogenes such as *ErbB2, EGFR, K-RAS* and focal deletions centered near *CDKN2A* and *FHIT*. In addition, our group demonstrated that EAC presents with more multiple amplification events as compared to either gastric or colon adenocarcinomas[69]. In addition, when characterizing the mutational profile of EAC, our group described a specific mutational signature of EAC, where we observed a high prevalence of A>C transversions at AA dinucleotides[68]. The observation of these specific DNA alterations in EAC and not others cancers, suggests that they may be driven by the unique environment during EAC development, mainly driven by the chronic inflammation associated with GERD. These specific observations and the mutational landscape, suggest that EAC reflects the unique environment of both GERD and chronic inflammation associated with EAC that together increases the frequency of these DNA alterations.

The molecular characterization of EAC using exome sequencing and whole genome sequencing has revealed that most frequent mutations occur in tumor suppressor genes[68], such as *TP53, ARID1A, and SMAD4*[68]. The lack of actionable targets in EAC, such as tyrosine kinases,

and the significant genomic instability and heterogeneity in these tumors, explains the poor advances in treatments in this type of cancer. A recent study, by Secrier *et al.*[67], identified specific subgroups of tumors that were enriched for different targetable potential. For example, they identified a BRCA-signature with prevalent defects in the homologous recombination pathway. In the breast cancer field, there has been a success in targeting tumors containing the BRCA mutated and homologous recombination pathway[84], therefore, this suggests that similar approaches could be used in patients with EAC that have this BRCA signature. Secrier *et al.*[67] also identified a dominant T>G mutational pattern associated with a high mutational load and increased neoantigen burden. Finally, a third group had an increased prevalence of a C>A/T mutational pattern with evidence of an aging imprint. This subgroup was categorized by a dominant T>G mutational pattern, with highest mutational burden, and highest neoantigen load. Therefore, potential therapeutics for this group could include immunotherapies such as CTLA4-targeting agents and/or PD-1\PD-L1 treatments. These sub-clusters provide insight into categorizing EAC patients based on their mutational signature (C/A/T dominant group, the DDR impaired group, and the highly mutagenic group) for specific selection of targeted or other therapies.

**Clinical features of BE and EAC.**

BE patients are most often middle-aged, Caucasian males with a history of chronic GERD, and associated with extensive heartburn and chest pain for a long period of time[85]. Diagnosis of GERD and BE can be difficult, first, because some patients do not complain, or only show some of the common symptoms of GERD and secondly, because when symptoms do appear, their severity may not correlate with severity of the disease. While there is no effective method to predict EAC, the presence of dysplastic BE in the form of LGD or particularly HGD

increases the risk of risk of developing EAC[86]. GERD clinical characteristics are only present in 21% of EAC patients[23]. Most patients present with advanced disease, whose symptoms could be assigned to the direct effects of the regional, local or distant complications of the tumor. Many patients with EAC arrive with complaints of dysphagia (74%) and odynophagia (17%) at the time of diagnosis. There is also often weight loss, and if the patient has lost more than 10% of body mass this is associated with poor prognosis.

**Surveillance and detection of neoplastic progression of BE.**

Predicting which patients with BE will develop EAC is very difficult. The current clinical practice for the detection of the possible presence of EAC is by screening patients with high-risk gastro-esophageal reflux disease through endoscopy screening of the metaplastic BE tissue and to characterize the degree of dysplasia in the biopsy samples[27,87]. Patients are enrolled, based on clinical risk factors, such as long term GERD, male gender, being obese and over the age of 40. The screening program involves undergoing endoscopic-biopsy every 3 months to 2 years depending on the degree of dysplasia, during which 4 quadrant biopsy samples are taken every 1 to 2 cm and then evaluated for histologic changes by expert pathologists[27,87]. Dysplasia grading remains an important basis of surveillance for patients with Barrett's esophagus and many studies have shown that this method is associated with detection of EAC at an earlier stage, and this, in turn, has helped the patient outcome. One of the flaws of this approach is the random sampling of the endoscopic biopsies. Our group has suggested a more targeted approach, and we have developed fluorescently-labeled peptides that can identify the cancer cells. Importantly, a recent study by our group and collaborators shows the potential for detection of EAC *in vivo* through the use of fluorescently-labeled peptides detected by endoscopy after administration into the esophagus[62]. With this new approach fluorescent peptides targeting cell surface proteins

overexpressed in high-risk cells could facilitate the early detection of tumors located in the distal esophagus or gastric cardia.

**Risk factors for progression.**

Multiple studies have identified additional factors at the time of endoscopy that can inform for the potential risk of BE progression. There have been four studies that associated the presence and length of the hiatal hernia to progression[19,85,88,89]. A hiatal hernia is characterized as a portion of the upper stomach inserted through the diaphragmatic hiatus into the chest cavity. Avidan *et al.*[85] and Weston *et al.*[88] reported an association and prediction of progression that was correlated to the size of the hiatal hernia. In contrast, in two larger cohort prospective studies and a case-control study Sikkemma *et al*.[90] and Pohl *et al*.[19] respectively, found no association between hiatal hernia and progression to EAC. It is known that the presence of a hiatal hernia increases the incidence of acid reflux; therefore, it is often associated with BE and progression and this is likely due to increasing chronic reflux as consequence of the protrusion of the stomach into the chest cavity.

The length of BE is an additional clinical characteristic that some studies have associated with an increased risk of progression. Six studies found a range of results in terms of association of the BE length to the likelihood of progression to EAC. Rudolph *et al*.[91] and Bhat *et al*.[92] found in two independent studies that the length of the BE was not significantly correlated with the risk of progression. In contrast, Weston *et al.,* Wong *et al.* and Sikkemma *et al.* found a significant correlation of long segment BE with a higher degree of dysplasia. It is difficult to define a specific cut-off length of BE associated with progression to EAC, however, many of these studies did find and suggest that an increasing length of the BE is associated with malignant progression.

**High-grade dysplasia.**

The risk of progression to EAC from HGD varies across studies, mainly because of the inconsistency in the techniques used to manage the HGD, which include differences in surveillance protocols, endomucosal resection (EMR), and esophagectomy[91,93,94]. Nonetheless, most studies have shown that at the time of HGD removal, there is evidence of EAC arising within the HGD. In a study conducted by Dar *et al*. they found that around 57% of patients undergoing esophagectomy for pathologically-confirmed HGD, also had an invasive foci of EAC[95]. Studies have found that detection of HGD is associated with a high risk of progression yet this poses significant problems since 1) there is inter-observer variability in diagnosis of HGD, particularly in establishing the correct grade as either LGD or HGD, 2) sampling error, where the actual dysplastic tissue may be missed and 3) evidence that EAC can arise without prior detection of the BE-dysplasia sequence.

There are some clues at the time of endoscopy that can identify the potential risk for progression, nevertheless, identification of potential molecular biomarkers that could precisely risk stratify patients with BE and categorize subgroups at risk for development to HGD/EAC could enable more rational tailoring of endoscopic surveillance. Currently, upper endoscopy is the most extensively used tool to visualize and biopsy the esophagus to establish the identification of BE and potential progression. Current surveillance of patient with BE could be potentially enhanced by the use of biomarkers. Understanding and identification of markers that could predict subsequent incomplete response and/or relapse in patients who have undergone endoscopic eradication treatments for BE could potentially also help in the appropriate identification of the right endoscopic surveillance and/or treatment for these patients.

**Treatment.**

When identification of dysplastic BE or EAC occurs, there are a couple of clinical approaches to treating this disease. Patients that are diagnosed with high-grade dysplasia and therefore are at high risk for the development of cancer, require more aggressive approach. Often esophagectomy is the prevailing method to treat cancer and HGD but is associated with a mortality and morbidity frequency of 3-5% and 20-50% respectively[96]. Another alternative approach is endoscopic mucosal resection (EMR), a procedure in which specific regions of the mucosal esophagus lining is resected with a definite histologic diagnosis performed and shown to be curative in some cases[97]. The caveat of this approach is that since it only targets some regions of the lesions, neoplastic recurrence can be observed in up to 20% of cases during follow-up[98]. Because acid reflux is known to play a significant role in the development of dysplasia, most treatment strategies also include acid-blocking drugs or anti-reflux surgical procedures. However, no prospective trials currently exist showing that regulation of reflux symptoms has an impact on the prevention of the development of EAC. There is reported of a 50% reduction in the risk of development of EAC in patients taking non-steroidal anti-inflammatory drugs (NSAID)[99,100]. BE neoplastic progression to EAC is a complex process, and this suggests inflammation is an important factor. Further elucidation of better strategies for reducing cancer risk in patients with BE is currently needed. Additionally, risk stratification is required to maximize surveillance strategies and therapeutic decisions.

## Conclusion

Patients who have gastro-esophageal reflux disease (GERD) often develop Barrett's metaplasia, but only a small subset of patients with BE progress to adenocarcinoma[23,86]. The use of endoscopic surveillance coupled with a histological assessment and other variables can be useful for identifying high-risk individuals. Unfortunately, the random sampling of the biopsies

of the tissue can miss the tissue where the dysplastic and/or cancer cells are arising. The observation that a percentage of tumors in the cardia/GEJ arise without a previous existence of BE, poses another challenge to identify these patients. Therefore, the main goals of this thesis are to elucidate the molecular events that distinguish GEJAC vs EAC and to characterize, at the transcriptome levels, the events that associated with BE dysplastic progression to EAC. In addition, to identifying cell surface overexpressed biomarkers, for the detection of HGD, GEJAC, and EAC. Finally, we aimed at elucidating the genetics events that could be driving the significance difference in the incidence of EAC between Caucasians and African Americans. With the development of new detection techniques for early diagnosis and potential preventive treatments, we expect, in the future, to positively impact and increase the 5-year survival rate for EAC patients.

## Figures

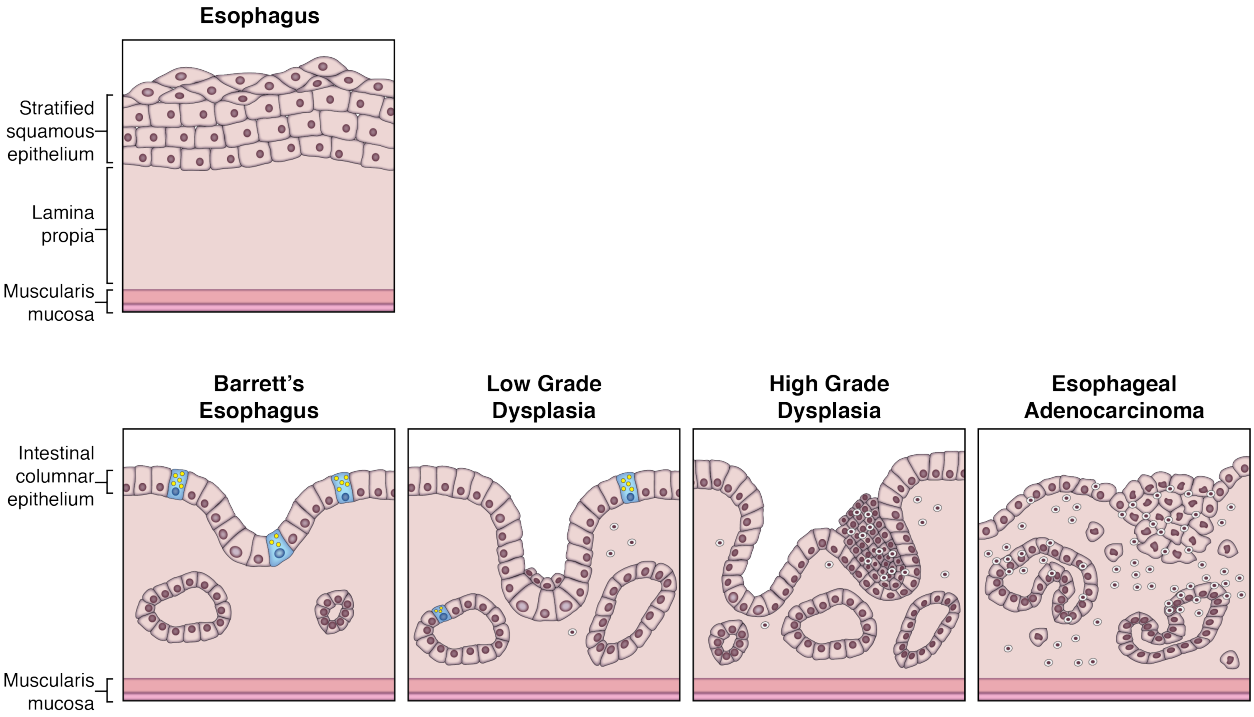



**Figure 1.1 Graphical depiction of histology observed in the replacement of esophageal tissue with Barrett's esophagus.** Esophageal squamous mucosa in the presence of GERD gets replaced with intestinal type tissue (bottom panel), that contains globet cells (blue cells) and produce mucin (yellow dots), which function as a protective mechanism against constant acid reflux. The intestinal-type BE tissue may progress through low-grade dysplasia, and to high-grade dysplasia (HGD) and is associated with loss of glandular BE structure and loss of mucin production. Finally, esophageal adenocarcinoma (EAC) is depicted when abnormal malignant cells have invaded through the basement cell membrane. (modified from Anaparthy and Sharma *et al.* 2014)

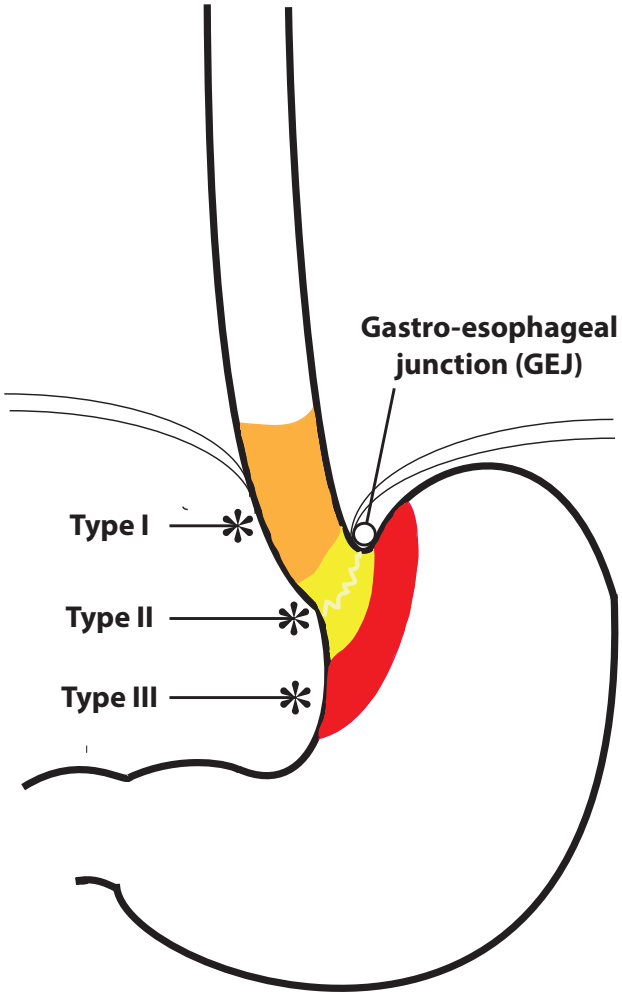



**Figure 1.2 Siewert-Stein classification of the location of esophago-gastric junction (GEJ) tumors. Type I** Adenocarcinoma of distal part of the esophagus (located within between 1-5cm above the anatomic GEJ) **Type II** Adenocarcinoma of the real cardia (within 1cm above and 2cm below the GEJ) **Type III** Adenocarcinoma of the subcardial stomach (2-5cm below GEJ) (Siewert and Stein, 1996).

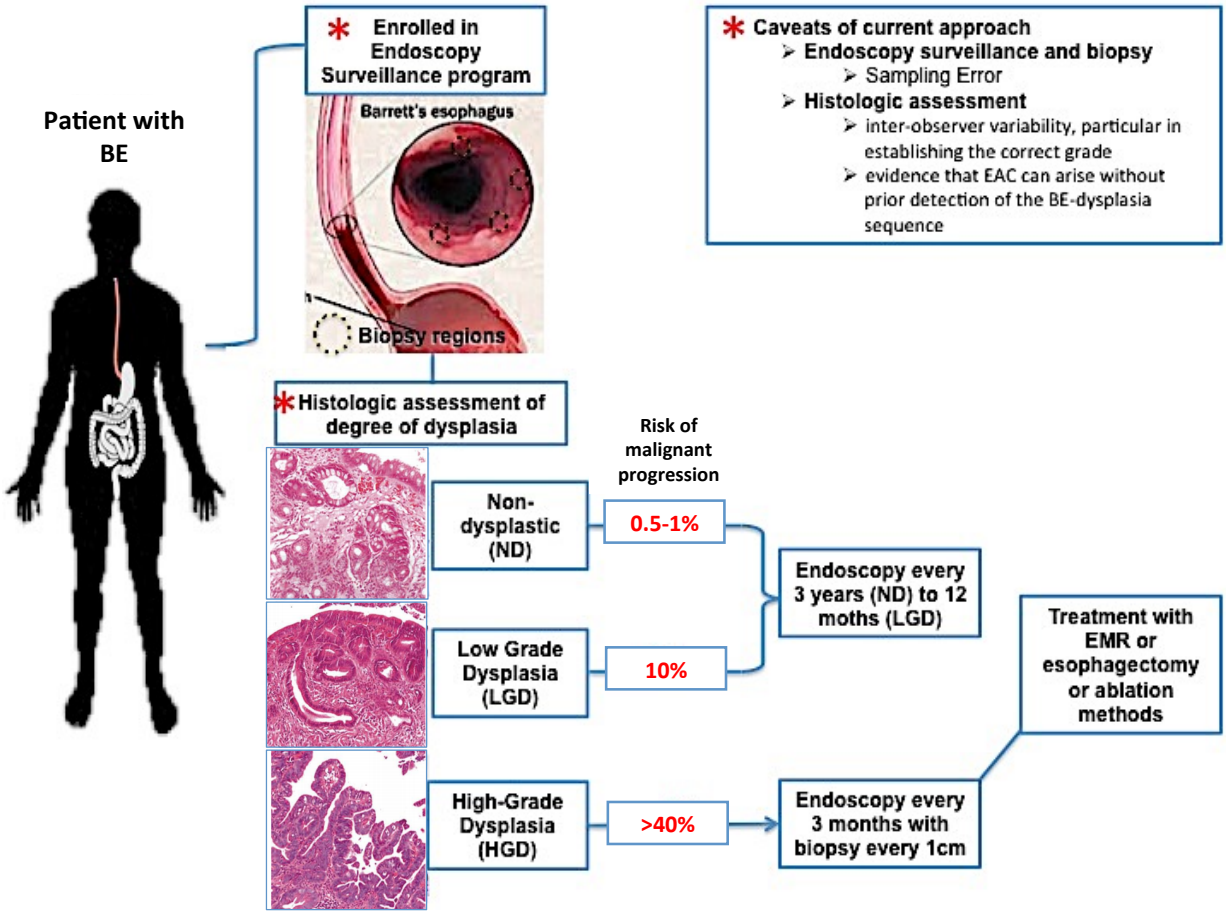



**Figure 1.3. Current methods for the diagnosis of malignant progression to esophageal adenocarcinomas (EAC) in patients with Barrett's esophagus (BE).** Patients that are diagnosed with BE are enrolled in an endoscopic surveillance program where biopsies of the BE tissue are taken and submitted for histopathological assessment of dysplasia grading. ND and LGD have a less risk of progression to EAC and patients are scheduled to receive surveillance every 3 years or 12 months, respectively for subsequent biopsy. Patients with HGD are at much higher risk of malignant progression, therefore, they should undergo endoscopy and biopsy every 3 months, or receive treatment for removal of the HGD regions by surgical or ablation methods. (Some of the pictures and figures presented here were modified from Conteduca V *et al.*[101] and Ong *et al.* 2010[102])

## References


1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. **67,** 7–30 (2017).
2. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136,** E359–E386 (2015).
3. Sampliner, R. E., Garewal, H. S., Fennerty, M. B. & Aickin, M. Lack of impact of therapy on extent of Barrett's esophagus in 67 patients. *Dig. Dis. Sci.* **35,** 93–96 (1990).
4. Williamson, W. A., Ellis, F. H., Gibb, S. P., Shahian, D. M. & Aretz, H. T. Effect of antireflux operation on Barrett's mucosa. *Ann. Thorac. Surg.* **49,** 537–542 (1990).
5. Sagar, P. M. *et al.* Regression and progression of Barrett's esophagus after antireflux therapy. *Br J Surg* **82,** 806–810 (1995).
6. Allum, W. H., Stenning, S. P., Bancewicz, J., Clark, P. I. & Langley, R. E. Long-term results of a randomized trial of surgery with or without preoperative chemotherapy in esophageal cancer. *J. Clin. Oncol.* **27,** 5062–5067 (2009).
7. Cunningham, D. *et al.* Capecitabine and Oxaliplatin for Advanced Esophagogastric Cancer. *N. Engl. J. Med.* **358,** 36–46 (2008).
8. Sudo, K. *et al.* Locoregional failure rate after preoperative chemoradiation of esophageal adenocarcinoma and the outcomes of salvage strategies. *J. Clin. Oncol.* **31,** 4306–4310 (2013).
9. Ku, G. Y. & Ilson, D. H. Chemotherapeutic options for gastroesophageal junction tumors. *Semin. Radiat. Oncol.* **23,** 24–30 (2013).
10. Siewert, J. R., Feith, M. & Stein, H. J. Biologic and clinical variations of adenocarcinoma at the esophago-gastric junction: relevance of a topographic-anatomic subclassification. *J. Surg. Oncol.* **90,** 139–46; discussion 146 (2005).
11. Ca, R. N. De. Independent and Joint Effects of Tobacco Smoking and Alcohol. **664,** 657–664 (1999).
12. Slomko, H., Heo, H. J. & Einstein, F. H. Minireview: Epigenetics of obesity and diabetes in humans. *Endocrinology* **153,** 1025–30 (2012).
13. Cummings, L. C. & Cooper, G. S. Descriptive epidemiology of esophageal carcinoma in the Ohio Cancer Registry. *Cancer Detect. Prev.* **32,** 87–92 (2008).
14. Whiteman, D. C. Esophageal Cancer: Priorities for Prevention. *Curr. Epidemiol. Reports* 138–148 (2014).
15. Quante, M., Abrams, J. A., Lee, Y. & Wang, T. C. Barrett esophagus: What a mouse model can teach us about human disease. *Cell Cycle* **11,** 4328–4338 (2012).
16. Shah, A. K., Saunders, N. a, Barbour, A. P. & Hill, M. M. Early diagnostic biomarkers for esophageal adenocarcinoma--the current state of play. *Cancer Epidemiol. Biomarkers Prev.* **22,** 1185–209 (2013).
17. Team, R. *et al.* World Gastroenterology Organisation Global Guidelines GERD Global Perspective on Gastroesophageal Reflux Disease. **51,** 467–478 (2017).
18. Arnold, M., Laversanne, M., Brown, L. M., Devesa, S. S. & Bray, F. Predicting the Future Burden of Esophageal Cancer by Histological Subtype: International Trends in Incidence up to 2030. *Am J Gastroenterol***,** (2017).
19. Pohl, H. *et al.* Risk factors in the development of esophageal adenocarcinoma. *Am. J. Gastroenterol.* **108,** 200–7 (2013).
20. Rastogi, A. *et al.* Incidence of esophageal adenocarcinoma in patients with Barrett's

esophagus and high-grade dysplasia: a meta-analysis. *Gastrointest. Endosc.* **67,** 394–8 (2008).
21. Kong, C. Y. *et al.* The impact of obesity on the rise in esophageal adenocarcinoma incidence: estimates from a disease simulation model. *Cancer Epidemiol. Biomarkers Prev.* **20,** 2450–6 (2011).
22. Thrift, a P. & Whiteman, D. C. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. *Ann. Oncol.* **23,** 3155–62 (2012).
23. Vaughan, T. L. & Fitzgerald, R. C. Precision prevention of oesophageal adenocarcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **12,** 243–248 (2015).
24. Howlader N, et al. SEER Cancer Statistics Review 1975-2011 National Cancer Institute. *SEER Cancer Stat. Rev.* (2011). at <http://seer.cancer.gov/csr/1975_2011/>
25. Chen, Z., Ren, Y., Du, X. L., Yang, J. & Shen, Y. Incidence and survival differences in esophageal cancer among ethnic groups in the United States. (2017).
26. Pickens, A. & Orringer, M. Geographical Distribution and Racial Disparity in Esophageal Cancer. *Ann. Thorac. Surg.* **4975,** 1367–1369 (2003).
27. Spechler, S. J., Sharma, P., Souza, R. F., Inadomi, J. M. & Shaheen, N. J. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* **140,** 1084–91 (2011).
28. Sabo, E. *et al.* Human Cancer Biology Expression Analysis of Barrett ' s Esophagus ^ Associated High-Grade Dysplasia in Laser Capture Microdissected Archival Tissue. **14,** 6440–6448 (2008).
29. Abrams, J. a, Lee, P. C., Port, J. L., Altorki, N. K. & Neugut, A. I. Cigarette smoking and risk of lung metastasis from esophageal cancer. *Cancer Epidemiol. Biomarkers Prev.* **17,** 2707–13 (2008).
30. El-Serag, HB and Sonnenberg, A. Associations between di V erent forms of gastro-oesophageal reflux disease. 594–599 (1997).
31. Spechler, S. J., Jain, S. K., Tendler, D. A. & Parker, R. A. Racial differences in the frequency of symptoms and complications of gastro-oesophageal reflux disease. *Aliment. Pharmacol. Ther.* **16,** 1795–1800 (2002).
32. El-Serag, H. B. *et al.* Gastroesophageal reflux among different racial groups in the United States. *Gastroenterology* **126,** 1692–1699 (2004).
33. Barrett, N. R. Foreign bodies in the breast. *Br. J. Surg.* **8,** 416–445 (1950).
34. Barrett, N. R. Hiatus hernia. *Br. J. Surg.* 13–16 (1954).
35. Allison PR and Johnstone, A. The oesophagus lined with gastric mucous. (1953).
36. Naef, A. P. Columnar-lined lower esophagus: an acquired lesion with malignant predisposition. *J Thorac Cardiovasc Surg* (1975).
37. Conteduca, V. *et al.* Barrett's esophagus and esophageal cancer: an overview. *Int. J. Oncol.* **41,** 414–24 (2012).
38. Dvorak, K. *et al.* Bile acids in combination with low pH induce oxidative stress and oxidative DNA damage: relevance to the pathogenesis of Barrett's oesophagus. *Gut* **56,** 763–71 (2007).
39. Zhang, H. Y., Hormi-carver, K., Zhang, X., Spechler, S. J. & Souza, R. F. In Benign Barrett ' s Epithelial Cells , Acid Exposure Generates Reactive Oxygen Species That Cause DNA Double-Strand Breaks. 9083–9090 (2009).
40. Whiteman, D. C. *et al.* Combined effects of obesity, acid reflux and smoking on the risk

of adenocarcinomas of the oesophagus. *Gut* **57,** 173–180 (2008).
41. Chow, W. *et al.* Body Mass Index and Risk of Adenocarcinomas of the Esophagus and Gastric Cardia Joseph F . Fraumeni , Jr .* Background : Incidence rates have risen rapidly for esophageal adenocarcinoma and moderately for gastric cardia ad- noma rose with increasing adult. *JNCI J. Natl. Cancer Inst.* **90,** 150–155 (1998).
42. Langergren, J. *et al*. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *NEJM* **339,** 61–68 (1998).
43. Shaheen, N. J. & Richter, J. E. Barrett's oesophagus. *Lancet* **373,** 850–61 (2009).
44. van Sandick, J. W. *et al.* Impact of endoscopic biopsy surveillance of Barrett's oesophagus on pathological stage and clinical outcome of Barrett's carcinoma. *Gut* **43,** 216–222 (1998).
45. Incarbone, R., Bonavina, L., Saino, G., Bona, D. & Peracchia, A. Outcome of esophageal adenocarcinoma detected during endoscopic biopsy surveillance for Barrett's esophagus. *Surg. Endosc. Other Interv. Tech.* **16,** 263–266 (2002).
46. Ferguson, M. K. *et al.* Long-term survival after esophagectomy for Barrett's adenocarcinoma in endoscopically surveyed and nonsurveyed patients. *J. Gastrointest. Surg.* **6,** 29–36 (2002).
47. Fountoulakis, A. *et al.* Effect of surveillance of Barrett's oesophagus on the clinical outcome of oesophageal cancer. *Br. J. Surg.* **91,** 997–1003 (2004).
48. Corley, D. A., Levin, T. R., Habel, L. A., Weiss, N. S. & Buffler, P. A. Surveillance and survival in Barrett's adenocarcinomas: A population-based study. *Gastroenterology* **122,** 633–640 (2002).
49. Booth, C. L. & Thompson, K. S. Barrett's esophagus: A review of diagnostic criteria, clinical surveillance practices and new developments. *J. Gastrointest. Oncol.* **3,** 232–242 (2012).
50. Suh, Y.-S. *et al.* Should Adenocarcinoma of the Esophagogastric Junction Be Classified as Esophageal Cancer? A Comparative Analysis According to the Seventh AJCC TNM Classification. *Ann. Surg.* **255,** 908–915 (2012).
51. Hasegawa, S. *et al.* Esophagus or Stomach? The Seventh TNM Classification for Siewert Type II/III Junctional Adenocarcinoma. *Ann. Surg. Oncol.* 773–779 (2012). doi:10.1245/s10434-012-2780-x
52. Sobin L, Gospodarowicz M, W. C. (eds). *TNM : classification of malignant tumours*. (2009).
53. Rüdiger Siewert, J., Feith, M., Werner, M. & Stein, H. J. Adenocarcinoma of the esophagogastric junction: results of surgical therapy based on anatomical/topographic classification in 1,002 consecutive patients. *Ann. Surg.* **232,** 353–61 (2000).
54. Bass, A. J. *et al.* Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513,** 202–209 (2014).
55. Kim, J. *et al.* Integrated genomic characterization of oesophageal carcinoma. *Nature* (2017). doi:10.1038/nature20805
56. Canto, M., Setrakian, S., Willis, J. & Chak, A. Methylene blue staining of dysplastic and nondysplastic Barrett's esophagus: an in vivo and ex vivo study. (2001).
57. Breyer, H. P., Silva de Barros, S. G., Maguilnik, I. & Edelweiss, M. I. Does methylene blue detect intestinal metaplasia in Barrett's esophagus? *Gastrointest. Endosc.* **57,** 505–509 (2003).
58. Guelrud, M. N m & m. *Gastrointest. Endosc.* **47,** 3–6 (1998).

59. Guelrud, M., Herrera, I., Essenfeld, H. & Castro, J. Enhanced magnification endoscopy: A new technique to identify specialized intestinal metaplasia in Barrett's esophagus. *Gastrointest. Endosc.* **53,** 559–565 (2001).
60. Réaud, S., Croue, A. & Boyer, J. Diagnostic accuracy of magnifying chromoendoscopy with detection of intestinal metaplasia and dysplasia using acetic acid in Barrett's esophagus. *Gastroenterol. Clin. Biol.* **30,** 217–223 (2006).
61. Bird-Lieberman, E. L. *et al.* Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. *Nat. Med.* **18,** 315–321 (2012).
62. Sturm, M., Joshi, B. & Lu, S. Targeted imaging of esophageal neoplasia with a fluorescently labeled peptide: first-in-human results. *Sci. Transl. ...* **5,** (2013).
63. Clark, G. W. B. *et al.* Is Barrett's Metaplasia the source of Adenocarcinomas of the Cardia? *Arch Surg* **129,** 609–614 (1994).
64. Theisen, J. *et al.* Preoperative chemotherapy unmasks underlying Barrett's mucosa in patients with adenocarcinoma of the distal esophagus. *Surg. Endosc. Other Interv. Tech.* **16,** 671–673 (2002).
65. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
66. Weaver, J. M. J. *et al.* Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat. Genet.* **46,** 837–843 (2014).
67. Secrier, M. *et al.* Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat. Genet.* **48,** 1131–1141 (2016).
68. Dulak, A. M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45,** 478–486 (2013).
69. Dulak, A. M. *et al.* Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res.* **72,** 4383–93 (2012).
70. Liu, W., Hahn, H., Odze, R. & Goyal, R. Metaplastic esophageal columnar epithelium with out goblet cells shows DNA content abnormalities similar to goblet cell-containing epithelium. *Am J Gastroenterol.***, 104,** 816–824 (2009).
71. Dulak, A. M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45,** 478–86 (2013).
72. Chaves, P. *et al.* Chromosomal analysis of Barrett's cells: demonstration of instability and detection of the metaplastic lineage involved. *Mod. Pathol.* **20,** 788–96 (2007).
73. Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47,** 1047–1055 (2015).
74. Lin, L. *et al.* A minimal critical region of the 8p22-23 amplicon in esophageal adenocarcinomas defined using sequence tagged site-amplification mapping and quantitative polymerase chain reaction includes the GATA-4 gene. *Cancer Res.* **60,** 1341–1347 (2000).
75. Lin, L. *et al.* Identification and characterization of a 19q12 amplicon in esophageal adenocarcinomas reveals cyclin E as the best candidate gene for this amplicon. *Cancer Res.* **60,** 7021–7027 (2000).
76. Lin, L. *et al.* The hepatocyte nuclear factor 3 alpha gene, HNF3alpha (FOXA1), on

chromosome band 14q13 is amplified and overexpressed in esophageal and lung adenocarcinomas. *Cancer Res.* **62,** 5273–5279 (2002).
77. Miller, C. T. *et al.* Gene Amplification in Esophageal Adenocarcinomas and Barrett ' s with High-Grade Dysplasia Gene Amplification in Esophageal Adenocarcinomas and Barrett ' s with High-Grade Dysplasia. 4819–4825 (2003).
78. Miller, C. T. *et al.* Amplification and Overexpression of the Dual-Specificity Tyrosine- ( Y ) -Phosphorylation Regulated Kinase 2 ( DYRK2 ) Gene in Esophageal and Lung Adenocarcinomas Amplification and Overexpression of the Dual-Specificity Tyrosine- ( Y ) - Phosphorylation R. **2,** 4136–4143 (2003).
79. Miller, C. T. *et al.* Genomic amplification of MET with boundaries within fragile site FRA7G and upregulation of MET pathways in esophageal adenocarcinoma. *Oncogene* 409–418 (2005). doi:10.1038/sj.onc.1209057
80. Lin, J. *et al.* Expression and effect of inhibition of the ubiquitin-conjugating enzyme E2C on esophageal adenocarcinoma. *Neoplasia* **8,** 1062–71 (2006).
81. Lin, J. *et al.* L-Type Amino Acid Transporter-1 Overexpression and Melphalan Sensitivity in Barrett's Adenocarcinoma. *Neoplasia* **6,** 74–84 (2004).
82. Lin, J. *et al.* Melanoma-associated antigens in esophageal adenocarcinoma: Identification of novel MAGE-A10 splice variants. *Clin. Cancer Res.* **10,** 5708–5716 (2004).
83. Dulak, A. M. *et al.* Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res.* **72,** 4383–4393 (2012).
84. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434,** 917–921 (2005).
85. Avidan, B. & Sonnenberg, A. Hiatal hernia size, Barrett's length, and severity of acid reflux are all risk factors for esophageal adenocarcinoma. *Am J Gastroenterol.,* **97,** (2002).
86. Sharma, P. Barrett ' s Esophagus. 1–5 (2009).
87. Playford, R. J. New British Society of Gastroenterology (BSG) guidelines for the diagnosis and management of Barrett's oesophagus. *Gut* **55,** 442 (2006).
88. Weston, A. P. *et al.* Risk stratification of Barrett's esophagus: updated prospective multivariate analysis. *Am. J. Gastroenterol.* **99,** 1657–66 (2004).
89. Sikkema, M. *et al.* Predictors for neoplastic progression in patients with Barrett's Esophagus: a prospective cohort study. *Am. J. Gastroenterol.* **106,** 1231–8 (2011).
90. Sikkema, M., de Jonge, P. J. F., Steyerberg, E. W. & Kuipers, E. J. Risk of esophageal adenocarcinoma and mortality in patients with Barrett's esophagus: a systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* **8,** 235–44; quiz e32 (2010).
91. Rudolph, R. E. *et al.* Effect of segment length on risk for neoplastic progression in patients with Barrett esophagus. *Ann. Intern. Med.* **132,** 612–20 (2000).
92. Bhat, S. *et al.* Risk of malignant progression in Barrett's esophagus patients: results from a large population-based study. *J. Natl. Cancer Inst.* **103,** 1049–57 (2011).
93. Buttar, N., Wang, K. & Sebo, T. Extent of high-grade dysplasia in Barrett's esophagus correlates with risk of adenocarcinoma. *Gastroenterology* 1630–1639 (2001).
94. Schnell, T. G. *et al.* Long-term nonsurgical management of Barrett's esophagus with high-grade dysplasia. *Gastroenterology* **120,** 1607–1619 (2001).
95. Dar, M. S., Goldblum, J. R., Rice, T. W. & Falk, G. W. Can extent of high grade dysplasia in Barrett's oesophagus predict the presence of adenocarcinoma at oesophagectomy ? 486–489 (2003).

96. Yachimski, P., Nishioka, N., Richards, E. & Hur, C. Treatment of Barrett's esophagus with high-grade dysplasia or cancer: predictors of surgical versus endoscopic therapy. *Clin. Gastroenterol.* **6,** 1206–1211 (2008).
97. Fleischer, D. E. *et al.* Endoscopic ablation of Barrett's esophagus: a multicenter study with 2.5-year follow-up. *Gastrointest. Endosc.* **68,** 867–76 (2008).
98. Pech, O. *et al.* Long-term results and risk factor analysis for recurrence after curative endoscopic therapy in 349 patients with high-grade intraepithelial neoplasia and mucosal adenocarcinoma in Barrett's oesophagus. *Gut* **57,** 1200–6 (2008).
99. Liao, L. M. *et al.* Nonsteroidal anti-inflammatory drug use reduces risk of adenocarcinomas of the esophagus and esophagogastric junction in a pooled analysis. *Gastroenterology* **142,** 442–452.e5; quiz e22–3 (2012).
100. Jankowski, J. a & Hooper, P. a. Chemoprevention in Barrett's esophagus: A pill a day? *Gastrointest. Endosc. Clin. N. Am.* **21,** 155–70 (2011).
101. Conteduca, V. *et al.* Barrett's esophagus and esophageal cancer: an overview. *Int. J. Oncol.* **41,** 414–24 (2012).
102. Ong, C.-A. J. Biomarkers in Barrett's esophagus and esophageal adenocarcinoma: Predictors of progression and prognosis. *World J. Gastroenterol.* **16,** 5669 (2010).

# CHAPTER 2

## Genomic Similarities Between Gastroesophageal Junction and Esophageal Barrett's Adenocarcinomas[*1]


### Summary

The current high mortality rate of only a <15% five-year survival for esophageal adenocarcinoma (EAC) reflects frequent presentation at an advanced stage. Recent efforts utilizing fluorescent peptides have identified overexpressed cell surface targets for endoscopic detection of early stage Barrett's-derived EAC. Unfortunately, 30% of distal AC patients present with gastroesophageal junction adenocarcinomas (GEJAC) without premalignant Barrett's metaplasia, thus limiting this early detection strategy. We compared mRNA profiles from 52 EACs (tubular EAC; tEAC) collected above the gastroesophageal junction with 70 GEJACs, 8 normal esophageal and 5 normal gastric mucosa samples. We also analyzed our previously published[22] whole-exome sequencing data in a large cohort of these tumors. Principal component analysis, hierarchical clustering and survival-based analyses demonstrated that GEJAC and tEAC are highly similar, with only modest differences in expression and mutation profiles. The combined expression cohort allowed identification of 49 genes encoding cell surface targets overexpressed in both GEJAC and tEAC. We confirmed that three of these candidates (*CDH11*, *ICAM1,* and *CLDN3*) were overexpressed in tumors when compared to normal esophagus, normal gastric and non-dysplastic Barrett's mucosa, and localized to the surface of tumor cells.


[1] Ferrer-Torres, D. *et al.* Genomic similarity between gastroesophageal junction and esophageal Barrett's adenocarcinomas. *Oncotarget* **5,** (2016)

Molecular profiling of tEAC and GEJAC tumors indicated extensive similarity and related molecular processes. Identified genes that encode cell surface proteins overexpressed in both Barrett's-derived EAC and those that arise in GEJAC without Barrett's metaplasia may allow detection of both types of cancer.

**Introduction**

Over the past three decades, the incidence of GEJAC and EAC in the US has risen at a rate of 7.5% per year[2], with other Western countries reporting similar increases[3-4]. Currently this disease presents within a characteristic demographic, such that approximately 80% of new cases arise within Caucasian males over the age of 40 years[5-6], and while the reasons for the rapid incidence increase are undetermined, it is clear that obesity, smoking and particularly chronic GERD each play an important role. Advances in diagnostic and treatment approaches have improved short-term treatment responses, yet only one in five patients survive 5 years post-diagnosis[7]. The greatest obstacles for improving patient survival include an advanced stage at diagnosis and an incomplete response to chemoradiotherapy[8–10]. Evidence from several small-scale programs suggests an early diagnosis via adequate surveillance can dramatically improve EAC patient survival[11-12], as well as reduce the need for aggressive chemoradiation[8]. Therefore, there is a pressing need to implement efficient, accurate surveillance programs among high-risk populations.

We[13] and others[14-15] are developing fluorescently-labeled peptide-based imaging agents to enhance endoscopic detection of specific cell surface markers as a means to improve early stage EAC diagnosis. EAC arises from the precursor lesion, BE, which becomes dysplastic in a small minority of cases[16]. The presence of BE is currently the key factor for enrollment in existing surveillance programs, with histological evidence of dysplastic progression used as a

trigger for treatment interventions, including surgical or endoscopic mucosal resection (EMR) of HGD. However, 25-30% of EAC cases present with no histological evidence or history of BE[17,18]. Adenocarcinoma of the distal esophagus and junctional cardia do share many characteristics (reviewed by Carr *et al.*[19]) and are currently treated using similar surgical and chemotherapy treatment strategies. At the University of Michigan Hospital, 30% of EACs arise at the GEJ that are not associated with the presence of BE[9]. Others have reported similar findings suggesting that GEJAC tumors make up a high proportion of EAC cases with no history of BE[17]. Here we applied molecular profiling technologies to assess the relationship between GEJAC and EAC, and expression profiling as an initial screen to identify potential cell surface markers for the detection of both GEJAC and EAC regardless of the prior presence of Barrett's mucosa.

## Results

### GEJAC and tEAC Demographics

**Table 2.1** summarizes the key characteristics of the 122 EAC tissues used for expression array analysis. There were no differences associated with gender, BMI, stage, node status, adjuvant treatment or tobacco usage between GEJAC and tEAC (**Table 2.1**), yet GEJAC cases presented at a slightly older mean age. We saw minimal differences in semi-quantitative measures for tumor histological characteristics, including desmoplastic response, differentiation or the degree of lymphocytic invasion within each tumor.

We used detailed pathology records to assess the presence or absence of BE. We found that 77% (54/70) of GEJACs arose in the absence of BE, significantly more frequent (p=3.13e-05) than among tEAC samples (46%; 24/52). Given the strength of this result, the consistency with previous studies[8, 10, 17, 20], and difficulties associated with clear demarcations on the basis of anatomical site of origin in advanced tumors, we chose to also compare the 54 GEJACs without

evidence of underlying BE (perhaps representing GEJAC arising from cardia) to the 28 EACs with histologically confirmed BE. As shown in **Table 2.1**, when comparing these more rigorous subsets there were no differences in clinical characteristics. We also compared GEJAC with and without BE, and tEAC with and without BE across clinical parameters and found that for both tumor subsets the presence of BE was associated with a higher frequency of early stage tumors, as has been published previously[21]. Note that with only 11 cases of GEJAC with evidence of BE, therefore the GEJAC comparison is underpowered (**Table 2.2**).

**Mutation Comparisons of GEJAC and tEAC**

Using whole-exome sequencing data from 149 normal and tumor pairs, those samples with available pathology information assigning tumors as either tEAC (n=53) or GEJAC (n=41) in the original paper were chosen[22]. **Figure 2.1A** shows no significant differences in the total number of non-silent, protein-coding mutations between GEJAC and tEAC tumor groups, while **Figure 2.1B** shows that GEJAC mutations are significantly (p=0.02 Wilcoxon Rank-sum test) less likely to involve the ApA dinucleotide, a signature mutation associated with EAC[22-23]. Although the incidence of all mutations shown in **Figure 2.1** were previously described by Dulak and coworkers[22], specifically comparing the incidence of mutations between GEJAC and tEAC was not presented in that study. Here we show that profiles of certain mutations (identified by Dulak) only slightly differ between GEJAC and tEACs. Among the 26 significantly mutated genes we found no difference in the overall mutation frequency (**Figure 2.1C**; p=0.13 Wilcoxon Rank-sum test) between GEJAC and tEAC, though the mutation profile was significantly different (p=0.047 by paired T-test; **Figure 2.1D**). While the most mutated gene in EAC by WES, *TP53*[22-23], had a similar mutation frequency in both GEJAC and tEAC (75 and 77% respectively), several less frequently mutated genes (<15% of the cohort) showed a noticeably

higher mutation rate in tEAC (*AKAP6*, *TLL1*, *AJAP1*, *ACTL7B*, *F5* and *CNTNAP5*) relative to GEJAC (**Figure 2.1D**), but the per gene mutation counts (ranging from 1 to 9 in either GEJAC or tEAC groups) were too small for individual gene statistical comparisons. Of the top 26 genes, only *MYST3* showed a notably higher mutation rate in GEJAC (9.8%; 4/41) compared to tEAC (<2%; 1/53).

We then considered GEJACs without BE vs tEACs with BE and saw the above results recapitulated, with a significantly lower fraction of ApA mutations in GEJAC without BE (p=0.023 by Wilcoxon Rank-sum test) and a significant difference in the distribution of mutations across the same 26 genes (p=0.04 by paired T-test), as well as similar individual gene profiles to those of the parent dataset listed above (**Figure 2.2**).

**Unsupervised Clustering of 122 Tumors**

We used PCA and unsupervised hierarchical clustering to investigate whether GEJAC represents a distinct, overlapping or indistinguishable subset of EAC, based on whole-genome expression profiling. For PCA we used all 26,613 annotated array elements across 135 mRNA samples (NE=8, NG=5, GEJAC=70, tEAC=52) and found that both types of normal samples were clearly separated from the tumors within the first 3 principal components (PC) (**Figure 2.3**). To improve resolution within the cancer group we repeated PCA using only the 122 tumor samples (**Figure 2.4**). We then overlaid tumor location information, either GEJ or tubular esophagus, (**Figure 2.4A**), and assessed membership across PC1 and PC2, which each accounted for >5% of the total variance (**Figure 2.5**). We performed unsupervised hierarchical clustering by Pearson correlation and complete linkage across all 135 mRNA profiles that resulted in 4 basic clusters; NE and NG groups, as well as two cancer clusters, designated C1 and C2 in **Figure 2.6**. We then used membership in these two cancer clusters as an overlay for PCA and

considered the same two PCs in order to provide a point of comparison (**Figure 2.4B**). We used the Wilcoxon rank sum test to assess whether there was a difference in sample distribution when location (**Figure 2.4A**) or unsupervised hierarchical clustering (**Figure 2.4B**) were used to group tumors. While the GEJAC and tEAC comparison did give a significant different across the first PC (p=0.044) we saw no obvious subgroups or division of samples. By contrast, and as expected, the difference resulting from the unsupervised hierarchical clustering of tumor samples by gene expression was visibly and significantly separated (p=7.1E-16), although still overlapping (**Figure 2.4B**). The results were very similar when only GEJAC without evidence of BE were compared to tEAC with BE using the same procedure outlined above (**Figure 2.7**), demonstrating that the presence or absence of BE was not a key determinant.

**GEJAC and tEAC Expression**

Comparing the expression profiles of GEJAC and tEAC directly resulted in 1,368 differential probesets (ANOVA p-value < 0.01), although only 96 (7%) had a fold-change (FC) difference >1.5. Given the low number of transcripts with meaningful FC shifts in this comparison, gene ontology analysis was conducted on all 1,368 using DAVID (1,183 unique Entrez gene IDs). This identified two over-represented gene categories (hsa05322: Systemic lupus erythematosus and hsa04514: Cell adhesion molecules), however, only one gene, *HLA-DRB3*, (one of four genes common to both ontology categories) had >1.5-fold difference between GEJAC and tEAC.

As a more sensitive comparison, we identified genes that distinguished GEJAC and tEAC from normal (NG plus NE) tissue (ANOVA<0.01 and FC>1.5), then considered either increased (FC>1.5) or decreased (FC<0.67) in the cancer groups relative to normal tissues. As expected, ontology analyses on these lists for GEJAC and tEAC independently, revealed strong differences

compared with normal tissues for both cancer groups, including cell cycle, immune response, extra cellular matrix structural factors, cell adhesion and digestion related categories; all previously reported in association with EAC. To compare the relative strengths of these ontology categories, we plotted the –log base 10 of Benjamini adjusted p values for ontologies over-represented within GEJAC against the corresponding values resulting in tEAC (**Figure 2.4C**). We considered $>10^5$-fold difference between these matched p values (dotted lines marked on **Figure 2.4C**) to indicate a particular ontology category was more strongly represented within one cancer group. We assessed biologically relevant gene categories (relative to normal tissues) that might be more or less represented in GEJAC or tEAC. The majority of biological processes perturbed in EAC were similarly well represented in GEJAC and tEAC (**Figure 2.4C**), however, cell cycle and inflammation-related categories were more strongly represented in tEAC relative to GEJAC.

**Transcripts Associated with Overall Survival**

By applying univariate COX analyses, we identified 1,289 Entrez genes (1,462 transcripts, including unknowns) with log-rank test p values <0.05 to overall survival in our treatment naïve cohort of 116 EACs from patients surviving more than 3 months post-surgery. This was very similar to the 1,331 transcripts (5% of 26,613) expected by chance. Of these just over half, 689 genes (784 transcripts), showed increased expression with increased risk (relative risk >1), which were overrepresented with members of the cadherin gene family residing in chromosomal band 5q31, in addition to a broad group of transcription-related genes. The contrasting set of 601 genes (679 transcripts), where reduced expression was associated with decreased overall survival, were over-represented by structural mitochondrial genes including a subset directly related to cellular respiration.

While no individual genes passed the false discovery adjusted significance threshold of 0.05, one gene, *ZNF217* had an FDR adjusted p=0.054 and a 2.3 hazard ratio (95% confidence interval of 1.6 to 3.3). The next strongest scores were for a cluster of 12 loci with FDR adjusted p values ranging from 0.29-0.3. These genes were associated with modest relative risk contributions of less than 2.5-fold with the majority showing increased expression and increased risk. Among these 13 genes, the highest risk ratio was 5, for the pseudogene *GTF2IP1*, and the lowest was 0.4 to PIGW. Of interest were several zinc finger factors (*ZNF217*, *ZNF117*, *GTF2IP1* and *MEX3D*), though sparse *in silico* evidence links these genes.

As the gene with the strongest correlation to survival in our cohort, we used Kaplan-Meijer plots to compare samples with high and low *ZNF217* expression for all 116 EACs, as well as GEJAC (n=67) and tEAC (n=49) subsets. We also assessed potential dependences on key clinical features using multivariate COX regression analysis (**Figure 2.8A–C**). We used median expression across each tumor cohort to dichotomize high and low mRNA expression and reported log-rank p-value comparisons for each of these groups. These data confirmed a consistent, but modest survival benefit to EAC patients with low *ZNF217* expressing tumors (p=0.0034), with the same trend present in both GEJAC (p=0.0039) and tEAC (p=0.065) subsets.

We identified histological stage, node positivity, smoking history and tumor location as clinical variables with a univariate association with survival (**Figure 2.8D**). In the univariate analyses of tumor location (GEJAC vs tEAC) we saw that GEJAC was associated with a slight but significant improvement in overall survival by log-rank statistic (p=0.0044; **Figure 2.8D**). As can be seen in **Table 2.1** and **Table 2.2**, in our cohort there was a non-significant trend for GEJACs to present at an earlier stage, such that 40% of GEJACs presented with early stage disease (I or II) compared to 27.5% of tEACs. Siewert *et al.*[10, 24] reported similar findings, with

GEJAC (AEG II) showed a shift towards earlier stage at presentation, however, Clark *et al.*[17] and Curtis *et al.*[20] did not. When we restricted our comparison to early stage GEJAC vs early stage tEAC (**Figure 2.9B**), or compare the late stage subsets (**Figure 2.9C**), this relationship to overall survival disappeared (p=0.109 and p=0.169 respectively). We are unsure why, in our cohort, more GEJAC patients have an earlier presentation, however, given the strong correlation of disease stage to overall survival (**Figure 2.8D**) this difference in the distribution of disease stage may explain the improved survival for GEJAC patients. Multivariate analysis demonstrated that *ZNF217* over expression was independent of tumor stage, with other clinical variables having no significant impact on the model, including location (**Figure 2.8D**).

**Cell Surface Markers for GEJAC and EAC**

We used a three-step procedure (**Figure 2.10**) to identify overexpressed cell surface markers potentially useful for endoscope-based[13] detection of both GEJAC and tEAC. Firstly, differential genes distinguishing GEJAC (n=70) from NE and NG expression profiles were identified using both ANOVA ($p<0.01$) and fold-change (>2) thresholds, represented as Venn diagrams in **Figure 2.10**: Step 1. This resulted in 396 transcripts for GEJAC and 534 when the same criteria were applied to tEAC (n=52) of which 359 were common to both lists (91% of the smaller list: **Figure 2.10**: Step 2). We also used 2-fold rather than 1.5-fold to improve the prospect of qRT-PCR validation. Combined, the two lists totaled 571 transcripts, corresponding to 523 Entrez gene IDs. The broad gene ontology category GO:0005887 was used to identify plasma membrane associated factors within our list of genes overexpressed in GEJAC and/or tEAC and found 253 of the 523 encoded cell membrane proteins (**Figure 2.10**: Step 3).

As a final step, we examined our prior BE-EAC progression cohort (GEO series GSE37203[25]) that included Barrett's samples with no dysplasia (BE) (n=9), low-grade dysplasia

(LGD) (n=15), high-grade dysplasia (HGD) (n=7) and EAC (n=15). We identified 684 transcripts overexpressed in EACs compared to BE without dysplasia, which included 151 genes, represented in GO:0005887 (**Figure 2.10**). We then compared this list of genes overexpressed in EAC relative to BE to the list generated above and overexpressed in EAC (GEJAC and tEAC) relative to normal tissues. We found 49 membrane-associated genes that overlapped. Heat maps of these 49 genes for GEJAC vs normal tissues, tEAC vs normal tissues and the GSE37203 progression series (**Figure 2.11A**, **Figure 2.11B,** and **Figure 2.11C** respectively) demonstrated that while expression was collectively higher in each tumor set, relative to non-cancer tissues, each individual gene was high in only a subset of tumors. The 3 genes that passed selection thresholds for GEJAC, but not tEAC, (top genes in **Figure 2.11A**, **2.11B,** and **2.11C**) were also overexpressed in a number of tEACs. Similarly, the last 9 genes listed in each **Figure 2.11** panel passed our expression threshold in tEAC only but were overexpressed in a similar portion of GEJACs. Thus the GEJAC and tEAC group-specific expression trends were very similar across the 49 genes.

While the mean expression for each tumor group represented in each **Figure 2.11** panel was higher than the non-cancer sample groups for each of the 49 genes, each gene had a number of individual cancer samples with expression levels comparable to normal tissues. This lower expressing subset varied for each gene thus to discriminate the majority of EACs from surrounding tissues, multiple genes are required. While some degree of correlated expression was evident among the 49 genes, several had more unique expression profiles, including *CLDN3* and *SLC19A3*, potentially representing valuable additions to a detection panel.

Seventeen of the 49 potential cell surface marker genes identified were previously reported in association with EAC including *PLAU, PTGS2,* and *SPARC* which showed increased

EAC expression, relative to BE, in multiple studies. In addition, we recently showed TGM2 is overexpressed on the surface of EAC cells[26]. In the current study, we chose to validate *CDH11, ICAM1,* and *CLDN3* as examples of potential cell surface markers common to subsets of both GEJAC and tEAC.

Using an expanded cohort including available arrayed samples (7 NE, 5 NG, 58 GEJAC, 46 EAC) together with additional samples (1 NE, 1 NG, 7 BE, 19 LGD and 29 HGD), qRT-PCR was used to confirm overexpression of selected candidate genes in cancer relative to normal and precancerous tissues (**Figure 2.12**). **Figure 2.13** demonstrates that Pearson-correlation analyses *CDH11, ICAM1* and *CLDN3* among Human Gene 2.1 ST arrayed samples indicate consistent correlations between log2 array and relative expression (qRT-PCR) data (rho values of 0.84, 0.81 and 0.89 respectively). While each gene showed a clear difference between NE and either GEJAC or tEAC, differences were less distinct among non-cancer columnar tissues (NG and the BE groups: BE, LGD, HGD) (**Figure 2.12A**) suggesting that these genes would only be useful for distinguishing cancer foci from pre-cancer and normal tissues, rather than markers for identifying high-risk epithelium. Using our TMA with commercially available antibodies as shown in **Figure 2.12B**, cell surface expression of these markers was observed in HGD and EACs, although high-level expression was only observed in a small subset of tumors. For 14 EACs we had matching mRNA and TMA data. Although this overlapping subset was small, there was a trend towards specimens with higher mRNA levels staining strongly for the corresponding protein (data not shown). While protein detection sensitivity may be an issue, there are many biological considerations that influence protein to message ratios, such that mRNA levels can only be considered as a screening tool to identify likely candidates for protein validation.

## Discussion

We found that GEJACs have significantly less histological evidence of BE when compared to tEAC, yet molecular comparisons of these tumor classifications using DNA mutation and mRNA expression profiling suggest only minor differences, even when the presence or absence of BE was taken as a co-discriminator. Minor differences were also observed in both mutation profiles, with less ApA mutations among GEJACs (a recognized characteristic of the EAC mutation profile[22-23]), and over-represented gene ontologies, with less cell cycle and immune response factors overexpressed in GEJAC. Together these observations may suggest that a subset of EACs arise as a result of a more extreme set of conditions requiring more prominent mucosal defense and an increasing the likelihood of initiating the formation of BE (with goblet cells), although resulting tumors arise via the same set of mutagenic triggers.

Although pathology confirmed >70% viable tumor in each cancer specimen is it possible that associated normal tissue present may have masked GEJAC and tEAC cellular differences. In this case, microdissection rather than macrodisection may better discriminate GEJAC and tEAC. However, it should be noted that were true, then our mutation analysis would still have detected the differences between tumor cell types. We believe the subtle differences we observed by both expression and mutation analyses suggest that GEJAC and tEAC cancer cells are similar. As we move into IHC screening of our cell surface markers, we will be able to discern not only whether these proteins localize to the cell surface, but also which cells are staining. Markers that highlight stromal cells, rather than tumor cells will not be prioritized for validation as the relationship between stroma, activated stroma, and tumor cells is still an emerging field of investigation. Expression data revealed 1368 transcripts were significantly different between GEJAC and EAC, and more than expected by chance (266), but only 96 transcripts (7%) demonstrated >1.5-fold

difference between tumor groups. As expected from these minor differences, PCA analysis showed that tumor location (GEJAC or tEAC) was not a strong influence on gene expression profiling.

These results are consistent with epidemiological studies that demonstrate most risk factors for GEJAC and tEAC are shared, with subtle patient differences in obesity-related factors, reflux, and gender (reviewed by 19, 27). A significantly reduced association between BE histology and adenocarcinomas arising at the GEJ has often been observed[8–10,17,20], as we confirm here. The difference in BE rate does not translate into molecular differences suggests that the founding cell type(s), and pathway(s) for GEJAC and tEAC are shared. Molecular investigations of EAC indicate a heterogeneous disorder with different combinations of changes leading to cancer, suggesting the existence of molecular subtypes, as is the case for other common cancers. Unsupervised clustering of expression profiles in **Figure 2.4** demonstrated that the underlying molecular characteristics were much stronger than minor differences attributed to GEJAC vs tEAC. Perhaps the underlying tumor causation spectrum is influenced by tumor location, though the specific investigation of this hypothesis is beyond the scope of the current study. When examining the DNA copy number variations of 27 GEJAC tumors, Isinger-Ekstrand *et al*. (28) also found that junctional AC profiles mirrored those of tEAC and were distinct from changes frequent in non-cardia gastric cancers. We report that both expression profiling and mutation analyses suggest a shared etiology for GEJAC and EAC of the distal esophagus. This holds true whether group distinction was based solely on tumor location, or whether the absence of BE was included as a co-discriminator. It should also be noted that a lack of evidence of BE at the time of surgery does not exclude the possibility that it was either missed (present in esophageal sections other than those reviewed for histology) or that it was overrun by cancer

leaving no evidence at surgery.

The known association of EAC with BE has led to surveillance biopsy protocols with the intent to detect early cancer in these patients. The reduced incidence of BE, however, suggests that a large proportion of individuals at risk for GEJAC are unlikely to be considered for routine screening. The development of novel fluorescently-labeled peptides for endoscopic identification of early cancer in the esophagus[13] has increased the potential for the detection of early Barrett's-associated adenocarcinomas, with the promise of improving patient outcomes. The strong similarity between Barrett's-associated EAC and GEJAC, as shown in the present study, suggests that potentially useful peptides could be developed that would identify cancers of both the lower esophagus and GEJ, regardless of the presence of Barrett's esophagus.

Univariate COX analysis for overall survival against all 26,613 annotated transcripts showed that over expression of *ZNF217* mRNA represented the strongest gene-based risk within our cohort, with both GEJAC and tEAC samples showing support for this association (**Figure 2.8**). While *ZNF217* was the strongest, and just short of FDR-adjusted significance, several other genes show evidence of an association, though as with *ZNF217*, their relative risk contributions were small (<2.5 fold). In other cancer types, both mRNA and protein levels for ZNF217 have been shown to correlate with patient outcomes, including breast, ovarian, colon and prostate cancer types (recently showed by 29, and reviewed by 30, 31). These associations generally correlate with the presence of gain of chromosome 20q, a frequent event in EAC[32–35]. Both ZNF217 protein and mRNA tumor expression have been associated with 20q13 copy number for several tumor types[36–38]. Geppert *et al*.[39] used FISH to demonstrate that the presence of chromosomal gain involving *ZNF217* predicted stage-independent survival in 130 EAC patients. While based on copy number, these data are consistent with our mRNA findings.

Several lines of evidence implicate ZNF217 as a key player in the regulation of the epithelial to mesenchymal transition (EMT), including the discovery of *CDH1* as a direct repression target[40] and that *ZNF217* expression can be directly regulated by several EMT-related miRNAs, including miR-24[41], miR-203[42] and miR200c[43]. In prostate cancer miR-203 exists in a double negative feedback loop with the EMT transcription factor *SNAI2*, along with *ZNF217*[44] and miR-203 was previously shown to differentiate EAC from BE and decreased expression associated with poorer EAC patient outcome[45-46].

Using ChIP-seq analysis Frietze *et al.*[47] showed that ZNF217 associates with the repressive histone mark H3K27ac and H3K4me1. Transgenic models with up regulation of *ZNF217* expression stimulate mesenchymal transition through the activation of Snail1 and Twist[29]. Thus epigenetic remodeling, with ZNF217 as a key component, could be a central feature in explaining the dynamic nature of EMT[48].

Our expression profiling analysis has revealed 49 genes encoding potential cell surface markers that demonstrate transcriptional overexpression in EAC tumors compared to normal and pre-cancerous tissue. We confirmed tumor-specific overexpression for three genes (*CDH11, ICAM1,* and *CLDN3)* using qRT-PCR, and demonstrated protein localization specific to the cell surface of tumor cells by IHC. In addition, we have recently demonstrated that TGM2 was also overexpressed and present on the cell surface of EAC cells[26], while both *PTGS2* (COX2) and *TNFRSF12A* are known to increase during the transition from BE to EAC[49, 50]. The products of several genes from our potential cell surface list are suspected of playing key roles in more general cancer-related activities such as immunosuppression/evasion (*CD14* and *CD86*), cell migration (*ICAM*, *CDH11*) and proliferation (*TGFB1*, *PMEPA1*, *PDGFRL*, *SLC19A3*). Other markers on this list may have confounding issues, for example *OLR1* has shown strong

squamous cell staining at the leading edge of the epithelial surface while *SLC2A3* (GLUT3) expression is known be elevated in the tissue of smokers[51]. These factors, along with protein expression gradient, and overexpression frequency will need to be considered in the construction of a specific panel of markers to aid in the identification of early cancers. Ultimately we aim to apply a multiplexed panel of peptides using multispectral scanning fiber endoscope technology[52] to improve the success of histology-based screening programs for early EAC detection.

## Materials and Methods

### Sample cohort.

All samples were obtained following written, informed patient consent according to the approval and guidelines of the University of Michigan institutional review board. Tissues were obtained from patients undergoing esophagectomy for adenocarcinoma within the University of Michigan Health System between 1991 and 2012, without preoperative radiation or chemotherapy.

A portion of each specimen was immediately frozen in liquid nitrogen and stored at −80°C until use. All resected cancers underwent pathological analysis, and only those indicating adenocarcinomas arising either within 1 cm above and 2 cm below the GEJ (Siewert type II[10]) or within the distal (tubular) esophagus, more than 1 cm above the GEJ (tEAC), were included in this study. A board-certified pathologist (DGT) performed categorical or semi-quantitative histopathological assessment of the sections as follows; tissue type (squamous, BE, cardia, gastric), tumor type (AC), differentiation (well, moderate, poor), desmoplastic response (weak, moderate, high) and inflammatory response (weak, moderate, high). We noted histological evidence of signet ring cells in seven tumors (4 GEJAC and 3 tEAC). Cryostat sectioning was used to select regions containing >70% tumor cellularity prior to DNA or RNA isolation. Height

and weight data, at the time of surgery, were extracted from patient records and used to determine BMI category as follows: 'underweight' (BMI < 18.5), 'normal' (BMI 18.5 to 24.9), 'overweight' (BMI 25 to 29.9) and 'obese' (BMI 30.0 and above). Adjuvant treatment was considered positive when a standard chemo and/or radio treatment commenced within not more than three months after primary resection. Pathology reports were also reviewed regarding the presence of BE. A sample was considered positive for BE when the pathologist noted goblet cells among the columnar tissue at the margin of tumor sections, or when BE was noted in the resected material. Using this information, we have included additional analyses based on comparing the subset of GEJACs with no evidence of BE to the subset of tEACs where the presence of BE was noted as described below.

**Whole exome sequencing comparison of GEJAC and tEAC.**

WES data generated by Dulak *et al.* (22) was used to investigate differential mutation profiling within GEJAC and tEAC subgroups with variant calling, annotation and sample characteristics provided in the original publication. We compared all non-silent mutations observed in GEJAC (n=41) or tEAC (n=52) samples using the Wilcoxon Rank- sum test, as well as a paired Student T-test (two-sided) comparisons of the non-silent mutations within the 26 genes significantly mutated within the entire WES cohort (n=149) as originally identified[22] using the MutSig algorithm[53]. We also conducted analyses in which we compared the subset of GEJAC samples where histology did not note BE (n=35; 85% of the GEJAC mutation cohort) to the subset of tEAC samples where BE was noted (n=42; 81% of the tEAC mutation cohort).

**mRNA profiling.**

Total RNA was purified from normal esophageal squamous (NE; n=8), normal gastric (NG; n=5) epithelium and adenocarcinomas arising at both the gastro-esophageal junction

(GEJAC; n=70), and within the 'tubular' esophagus (tEAC; n=52) using miRNeasy spin columns (Qiagen, Valencia, CA), including on-column DNAse I incubation, according to the manufacturer's instructions. RNA samples with RIN scores greater than 6.0 (Bioanalyzer; Agilent Technologies, Palo Alto, CA), were submitted to the University of Michigan Cancer Center Genomics Core for cDNA synthesis, cRNA amplification (Ambion WT Expression Kit; Life Technologies, Grand Island, NY) and hybridization to Human Gene ST 2.1 arrays (Affymetrix, Santa Clara, CA) according to the manufacturer instructions. Expression values for each gene were estimated using the robust multi-array average (RMA) method[54] in the Bioconductor package[55] and log2-transformed. Analyses were restricted to the 26,613 coding and non-coding genes for which annotation details were available, including HUGO Gene Nomenclature Committee (HGNC) approved gene symbol and Entrez Gene ID.

**Principal component analysis and unsupervised clustering.**

We used Cluster (version 3.0) to perform Principal Component Analysis (PCA) of expression array data to visualize the relationship between sample groups. Mean normalized, batch adjusted log 2 expression data for 26,613 annotated array elements were applied to PCA, either using all 135 samples (8 NE, 5 NG, 70 GEJAC and 52 tEACs) or just the 122 tumor samples. To generate two-dimensional plots, we compared the top principal components (PC), ranked by eigenvalues, which individually explained the highest levels of the total variance, using 5% as a minimum threshold for investigation. Among these components, those that best demonstrated the separation between sample groups were graphed. Typically, this meant the top two PCs were compared.

The software packages Cluster (version 3.0) and Treeview (Java version 3.0 (56)) were used to generate and graph unsupervised hierarchical clustering of the 135 expression profiles

using Pearson correlation with average linkage using 26,613 annotated array elements. Data were normalized to tumor means for each gene to aid in dendrogram visualization. This analysis resulted in normal samples clustering together and tumors separating into two groups, with mixed GEJAC and tEAC membership in each of these clusters (**Figure 2.6**). Given that GEJAC and tEAC groups were not distinct by either PCA or hierarchical clustering, the Pearson correlation cluster membership was overlaid onto the PCA graphs, as a comparison to demonstrate how well the sample cohort could be separated. We used this as a comparison purely to more clearly demonstrate that PCA incompletely discriminated between GEJAC and tEAC.

**Gene ontology analysis of expression array data.**

The arrays were run in two batches, the first batch holding 8 NE, 5 NG, and 35 GEJAC, while the second batch consisted of 52 tEAC and an additional 35 GEJAC. We adjusted for batch effects by adding probe-set specific constants to each value in the second batch such that the probe-set means for GEJAC's in batch 2 agreed with those of batch 1. When fitting a one-way analysis of variance (ANOVA) model with means for each of the four tissue types, we reduce the degrees of freedom in the mean- squared-error and F-tests by 1 to account for this batch adjustment. Mean group expression ratios (typically >1.5 or 2-fold increase/decrease), in combination with an ANOVA test of $p<0.01$, were used to select differentially expressed genes between groups. Enrichment testing for over-represented gene ontology terms was performed using the DAVID website with the appropriate platform- specific background gene list ("HuEx-1_0-st-v2") and default algorithm settings[57-58]. Individual ontology categories with false discovery adjusted (Benjamini) p values <0.05 were reported, though we applied the modular enrichment analysis (MEA) based Functional Annotation Clustering feature built into DAVID to

assess redundant gene categories and to group similar gene sets under appropriate descriptors (59). Both batch- normalized and raw expression data for this experiment were deposited into the Gene Expression Omnibus (GEO series GSE74553).

**Identification of genes associated with overall survival.**

Of the 122 tumors used in this study, there were 2 patients who died from surgical complications within a month of surgery, and a further 4 patients who died within 3 months of surgery (3 GEJAC and 3 tEACs combined). In order to reduce the possibility of surgical complications confounding survival data, we chose to use the identified 116 patients who survived more than 3 months following surgery. For these patients, the average survival time was 38.7 months (range: 3 to 251 months), and an average follow-up time of 94.2 months (range: 18 to 242 months) for surviving participants. Using univariate analyses, we determined that of the available clinical variables stage, node status, tumor location, and smoking status each showed an association to overall survival (**Figure 2.8D**). We applied univariate COX analysis for all 26,613 annotated transcripts and applied FDR adjustment to the resulting log-rank (Mantel-Cox) test p-values. We considered genes with an FDR adjusted p value $<0.05$ to provide a significant association to overall survival. Survival associations were plotted (Kaplan-Meijer) using dichotomized mRNA expression, with cohort median expression as a cutoff. Multivariate analyses were used to assess whether the survival associations for significant genes were independent of stage, node status, tumor location and smoking status.

**Identification of GEJAC/tEAC expressing genes for cell surface proteins.**

To identify cell surface-coding genes selectively overexpressed in both GEJAC and tEAC, we applied a three-step procedure schematically represented in Figure 2.10. In step 1, we asked that GEJAC vs NE and GEJAC vs NG comparisons both gave $p<0.01$ and a fold-change

(FC) >2 and that this also was true for comparisons of tEAC vs NE and tEAC vs NG. In step 2 we selected the subset of genes indicated as being "plasma membrane" by Gene Ontology (GO:0005887), as listed within the COMPARTMENTS subcellular localization database, which resulted in 5 162 potentials among the 26,613 transcripts[60]. For step 3, we then analyzed the resulting list of genes in our previously published, independent data-set of 9 BE, 7 BE+LGD, 8 LGD, 7 HGD, 15 EAC assayed on Affymetrix U133A arrays (GEO Series GSE37203) in order to compare cancer (EAC) and non-dysplastic pre-cancer (BE) expression levels by ANOVA and fold-change[25].

**qRT-PCR validation.**

cDNA synthesis was performed using the High-Performance RT-PCR Kit (Life Technologies, Grand Island, NY) according to the manufacturer's instructions. Total RNA from 58 GEJAC and 46 tumors from the ST 2.0 array, 53 BE (7 without dysplasia, 17 LGD and 29 HGD), 6 NG and 8 NE samples were available for real-time (qRT- PCR) validation of the selected gene transcripts. qRT-PCR reactions primers were designed using Primer-BLAST[61] (*CDH11*: 5'-GCACGAGACCTATCATGCCA-3', 3'-CTGTCTGTGCTTCCACCGAA-5', *ICAM1*: 5'-GTA TGAACTGAGCAATGTGCAAG-3', 3'-GTTCCACCCG TTCTGGAGTC-5', *CLDN3*: 5'- TCGGCCAACACCA TTATCCG-3', 3'-GTACTTCTTCTCGCGTGGGG-5', *ZNF217*: 5'- CTCCGGGCCACTTTACACTT-3', 3'-TCTCT TTTGTGCCATGCTGTT-5') or previously published (GAPDH: 62). Annealing temperatures were determined and optimized using Cepheid SmartCycler (Cepheid, Sunnyvale, CA). Samples were run using the ABI PRISM® 7900HT Sequence Detection System according to the manufacturer's instructions and analyzed using relative quantitation utilizing GAPDH as the reference gene. Technical validation was assessed by correlating (Pearson rho) log2 of relative qRT-PCR expression values with

matched log2 ST 2.1 array data for each validation gene (**Figure 2.13**). GAPDH was chosen because it was highly expressed (mean log2 expression of 7.57 across all samples) with a minimal mean difference between normal and tumor samples (1.03-fold for 13 normal vs 122 tumor samples) within our ST 2.1 array data and is known to be an effective reference for esophageal samples[63].

**Immunohistochemistry and tissue microarray (TMA).**

A TMA was constructed as described by Kononen *et al* (64) containing 122 cores derived from the resected tissue from 73 EAC patients, including 60 tEAC, 3 GEJAC, 22 BE, 9 metastatic lymph nodes, and 14 normal tissues. Five μm sections were used for immunohistochemistry as previously described[25]. CDH11 (Cat# 32-1700, Life Technologies), ICAM1 (Cat# ab53013, Abcam, Cambridge, MA) and CLDN3 (Cat# 18-7340, Thermo Scientific, Pittsburgh, PA) monoclonal antibody were used at dilutions of 1:500, 1:250 and 1:100, respectively, after microwave citric acid epitope retrieval for 20 minutes and lightly counterstaining with hematoxylin. Each sample was then scored 0-3 corresponding to absent, light, moderate, or intense staining.

**Figures**

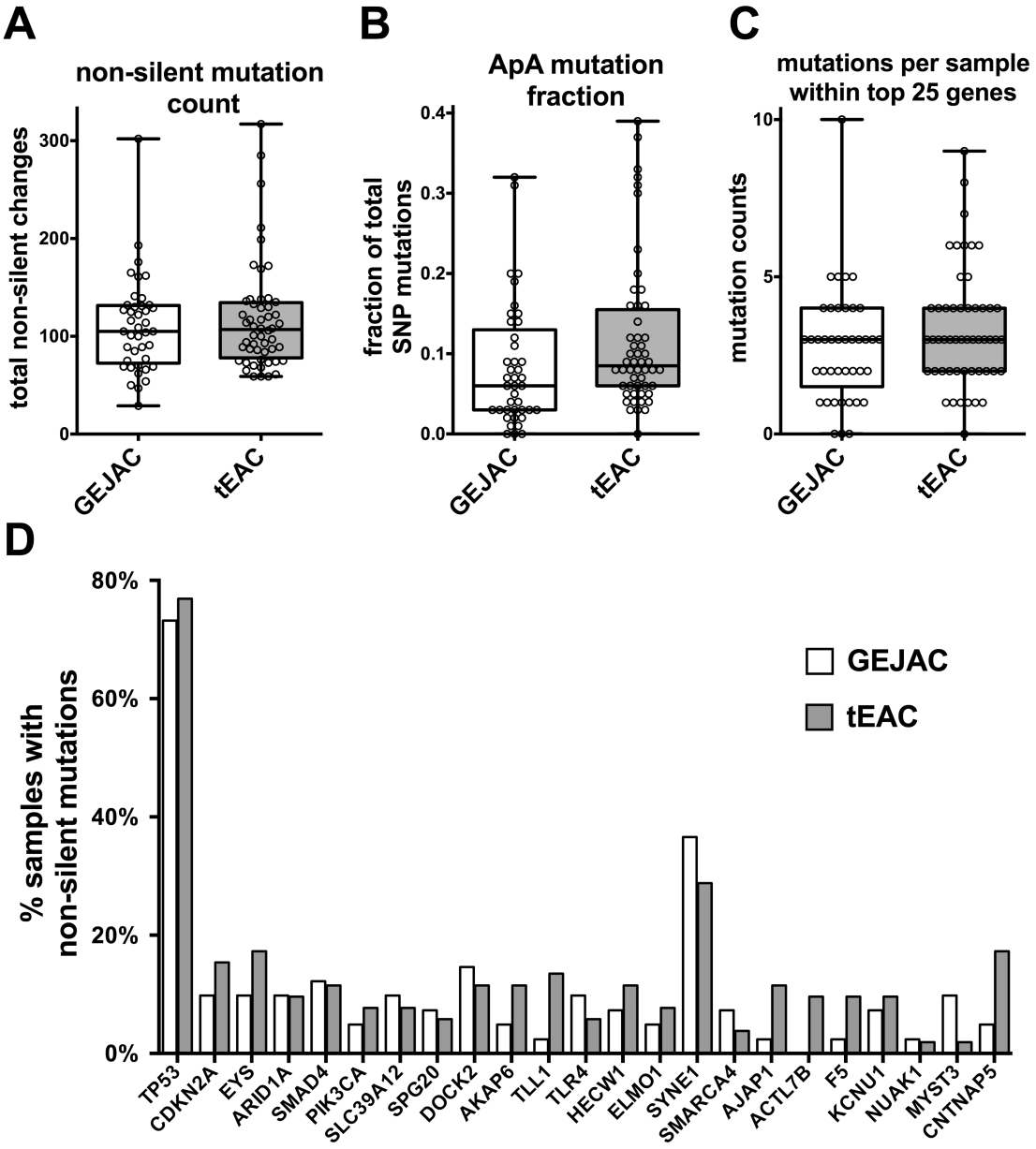



**Figure 2.1: Mutation profiling comparison of GEJAC and tEAC.** Whole exome sequencing data were downloaded for a cohort of 149 normal-tumor pairs, with mutation type and frequency determinations performed as in Dulak *et al.* 2013. When looking at **A)** the total number of non-silent mutations in tEAC vs GEJAC we found no significance difference based on tumor type. There was a modest difference when **B)** only mutations with the ApA dinucleotide profile were considered with the Wilcoxon rank-sum test. When only the originally identified 26 significantly genes were considered there was **C)** no difference in the summated number per sample ($p=0.134$ by Wilcoxon rank-sum test), however **D)** there was significance when the collective mutation profiles for these genes were compared between GEJAC and tEAC by paired T-test.

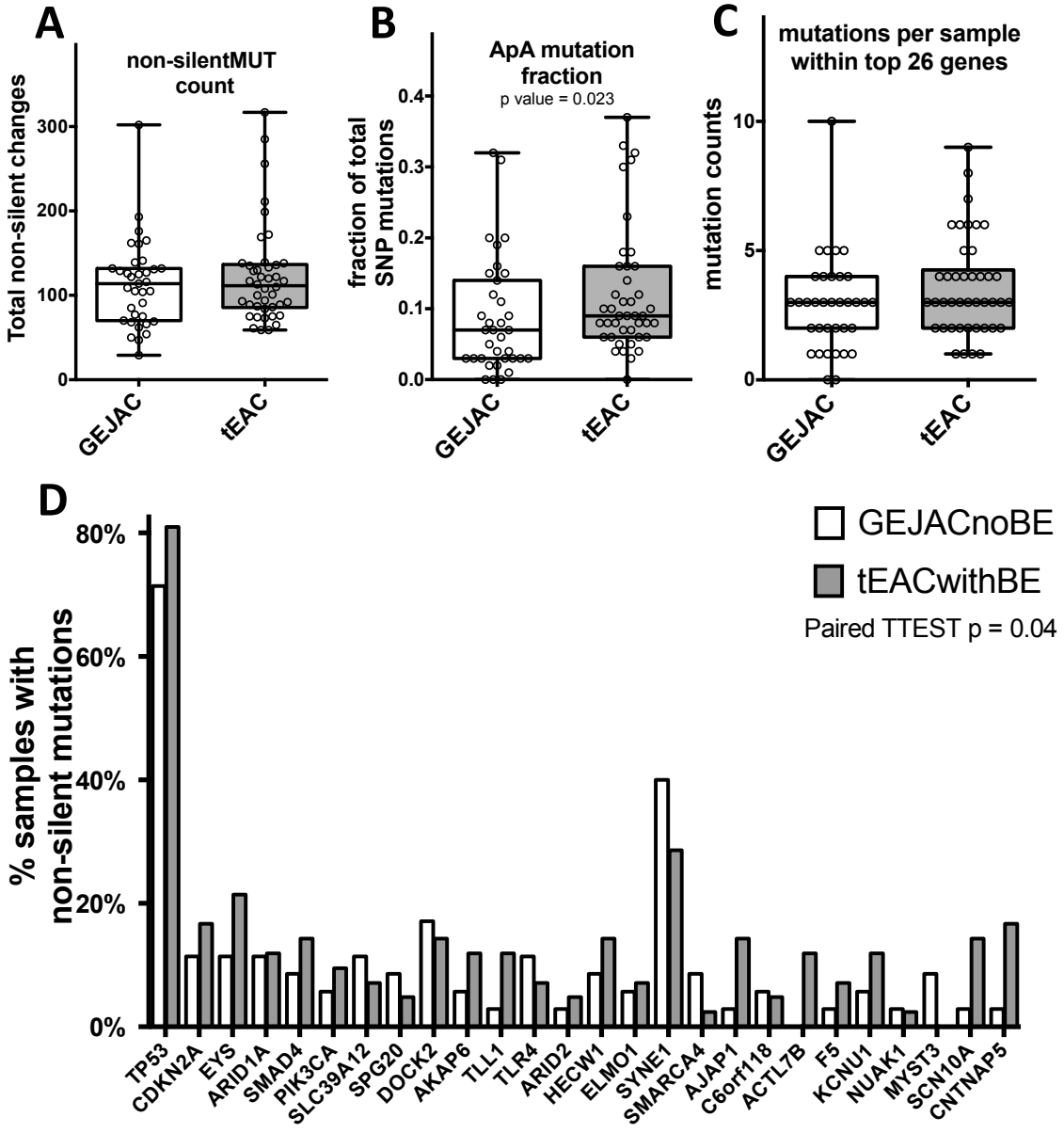



**Figure 2.2: Mutation profiling comparison of GEJAC without BE and tEAC with BE.** Whole exome sequencing data for GEJAC samples without BE histology (GEJACnoBE: n=35) and tEAC samples where BE histology was noted (tEACwithBE: n=42) were extracted from a cohort of 149 normal-tumor pairs, with mutation type and frequency determinations performed as in Dulak *et al.* 2013. When looking at **A)** the total number of non-silent mutations in tEAC with BE vs GEJAC no BE we found no significance differences. There was a modest difference when **B)** only mutations with the ApA dinucleotide profile were considered with the Wilcoxon rank-sum test. When only the originally identified 26 significantly genes were considered there was **C)** no difference in the summated number per sample, however **D)** there was significance when the collective mutation profiles for these genes were compared between GEJAC no BE and tEAC with BE by paired T-test.

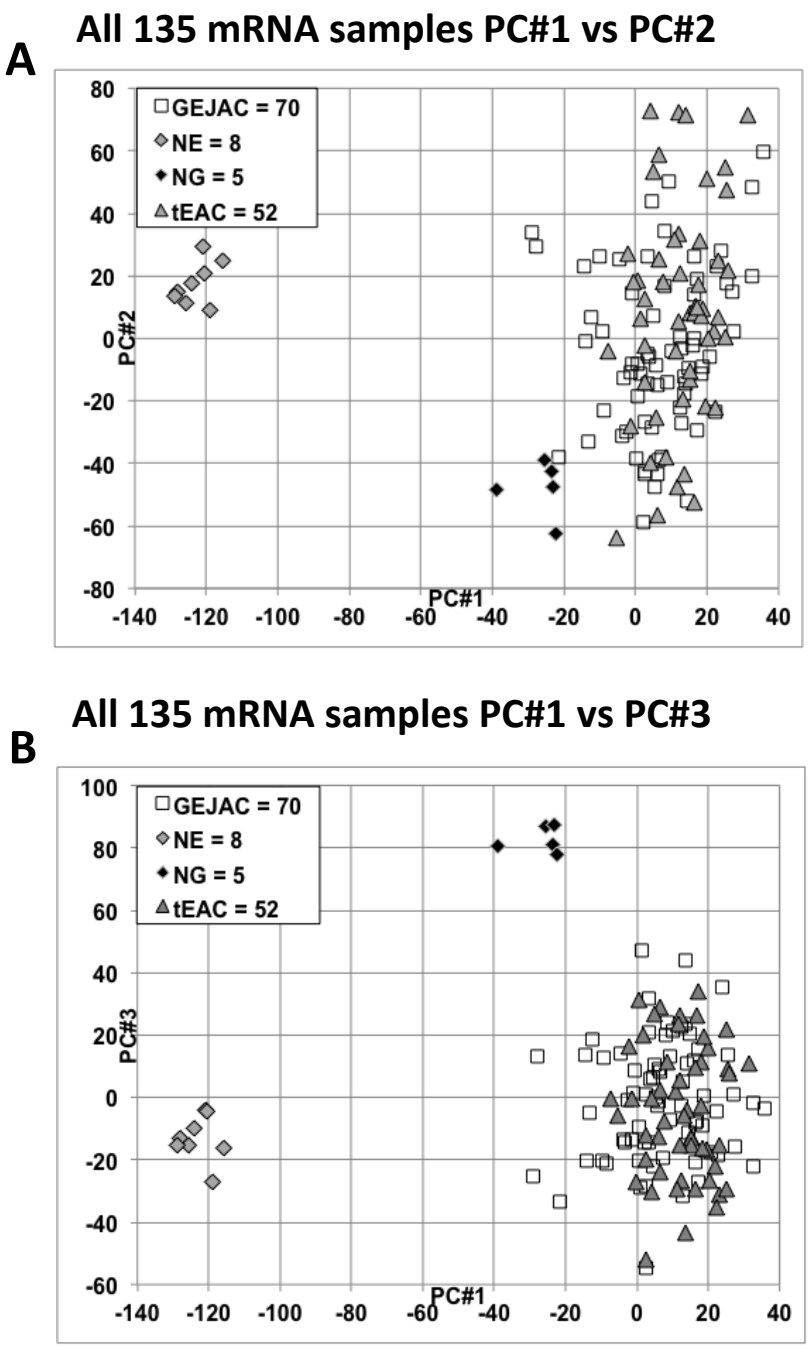



**Figure 2.3: PCA analysis in all 4 mRNA profiling groups.** All annotated probe sets (n=26,613), were standardized by subtracting the all sample cohort mean and dividing by the SD. **A)** The first two and **B)** first and third principal components were plotted and individual samples were assigned to their four histological groupings to demonstrate clear separation of normal tissues (NE and NG) but no separation of tumor groups (GEJAC and tEAC).

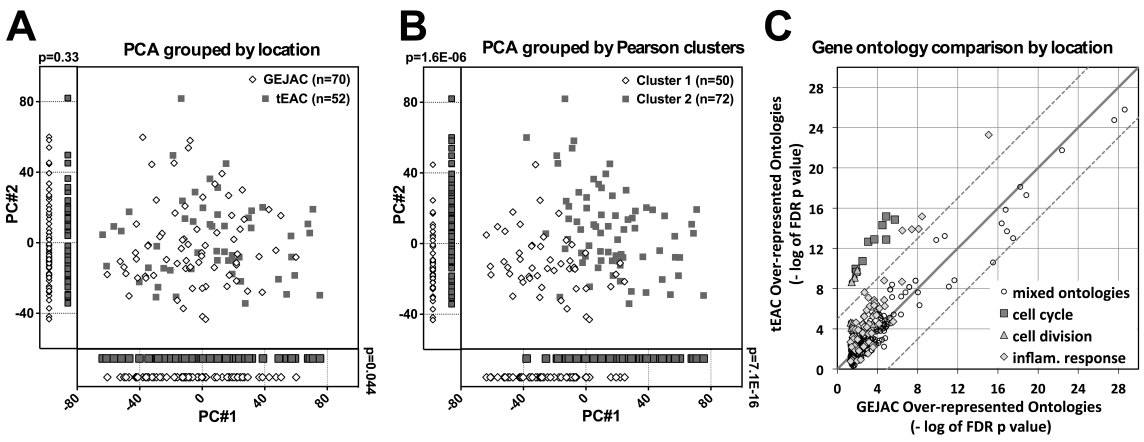



**Figure 2.4: mRNA profiling comparison of GEJAC and tEAC.** All annotated probe sets (n=26 613), were standardized by subtracting the tumor cohort mean and dividing by the SD. The first two principal components (each with variance >5%: **Figure 2.5**) were plotted and individual samples were assigned either **A)** a location (GEJ or tubular esophagus) or **B)** an unsupervised clustering assignment based on Pearson-correlation on the same 26 613 probe sets (**Figure 2.6**) Visual and statistical comparison demonstrates minor expression differences between GEJAC and tEAC compared to class assignment by gene expression. **C)** Gene ontologies significantly over-represented (DAVID generated Benjamini adjusted p values <0.05) in GEJAC comparison to both the normal tissue groups were plotted against their tEAC equivalent using a – log 10 (p value) format. Dotted and continuous lines represent $10^5$ fold and 1:1 ratio markers respectively Results demonstrate that genes related to the cell cycle and broad inflammation ontology categories were more enriched in tEAC relative to normal tissues, compared to GEJAC.

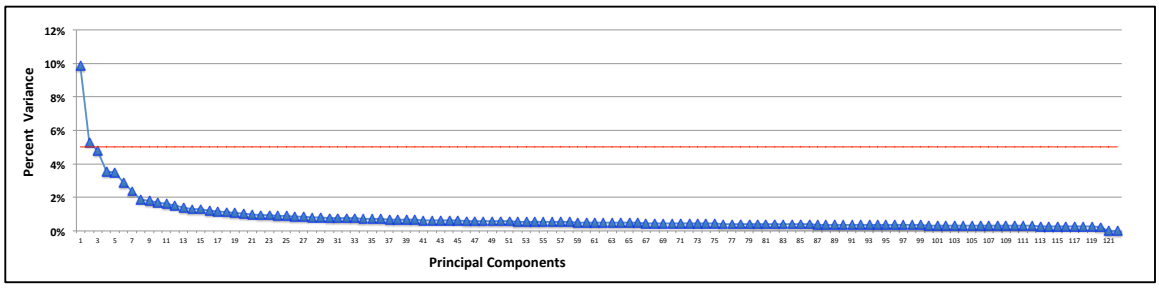



**Figure 2.5: The fraction of variance in each principle component.** All annotated probe sets (n=26,613), were standardized by subtracting the all sample cohort mean and dividing by the SD. The percent of total variance attributed to each principle component were then plotted to show that the top two components (PC#1 and PC#2) each explain more than 5% of the total variance.

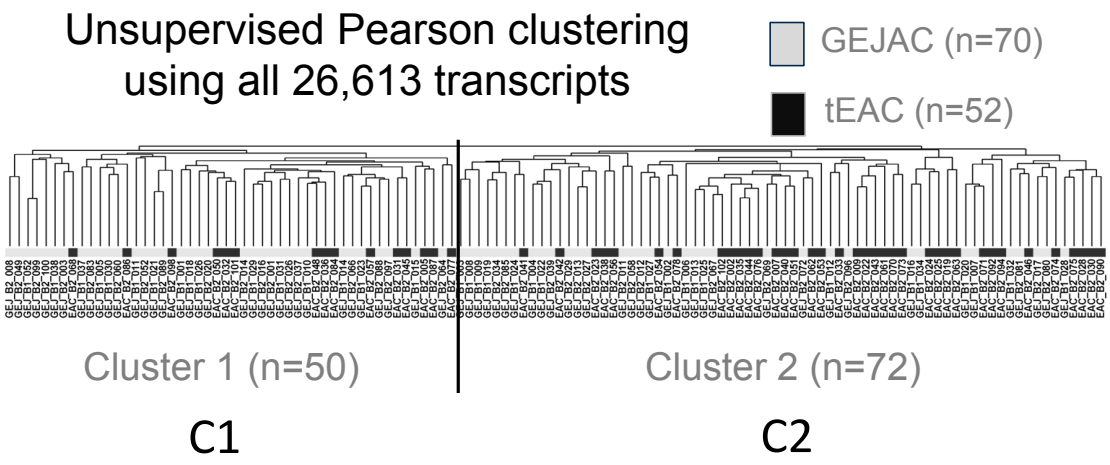



**Figure 2.6: Hierarchical clustering in 122 tumor mRNA samples.** Unsupervised clustering assignment of all 122 tumor samples based on Pearson-correlation and complete linkage on all 26,613 probe sets.

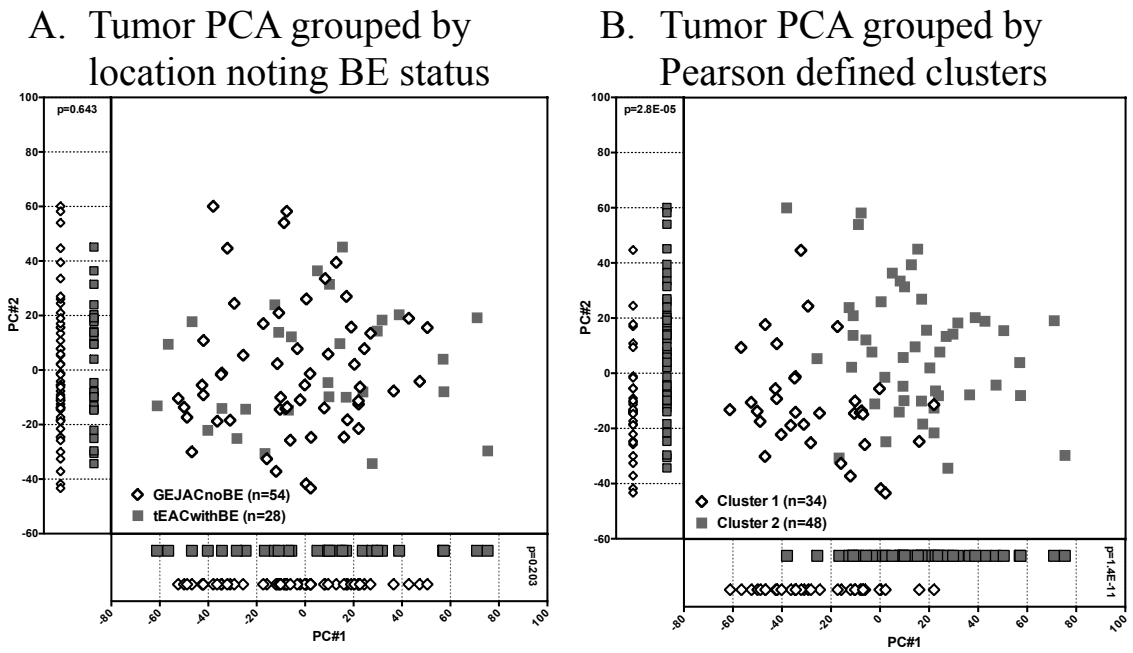
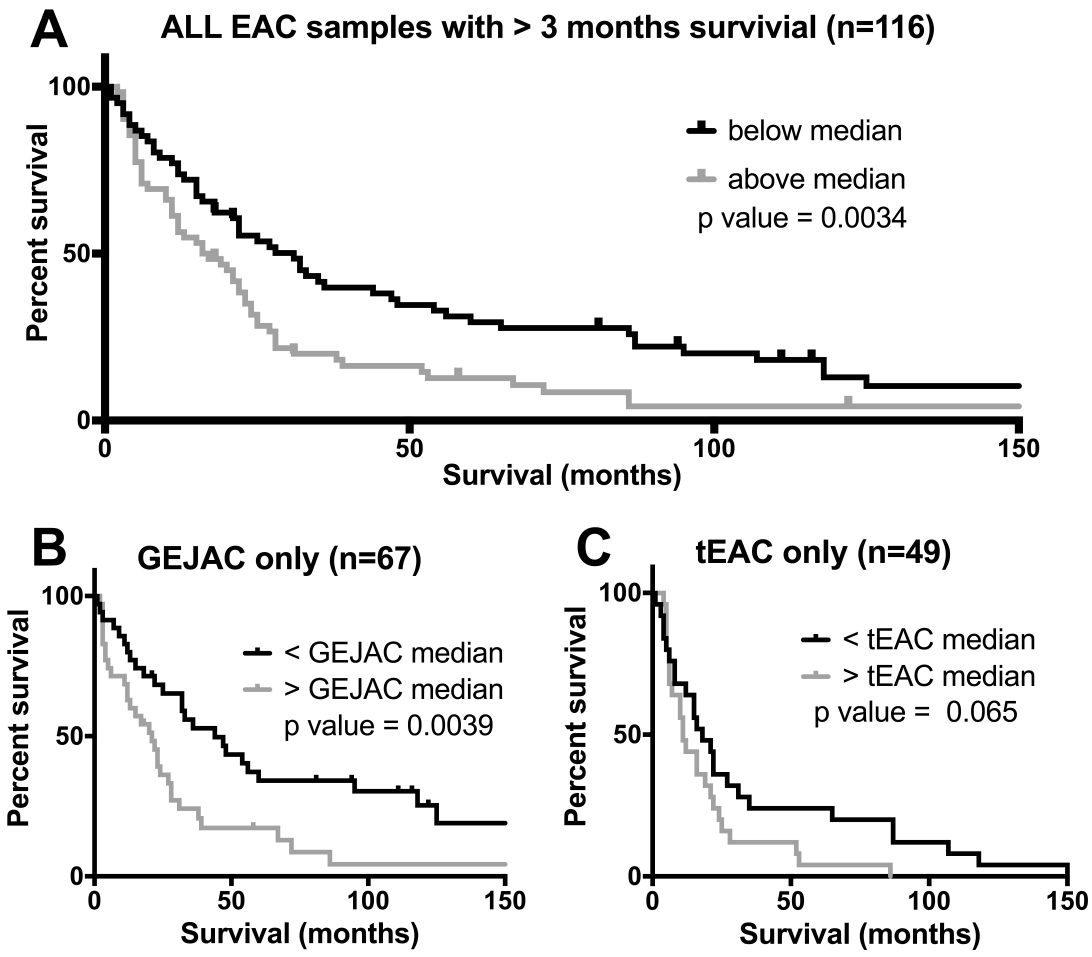


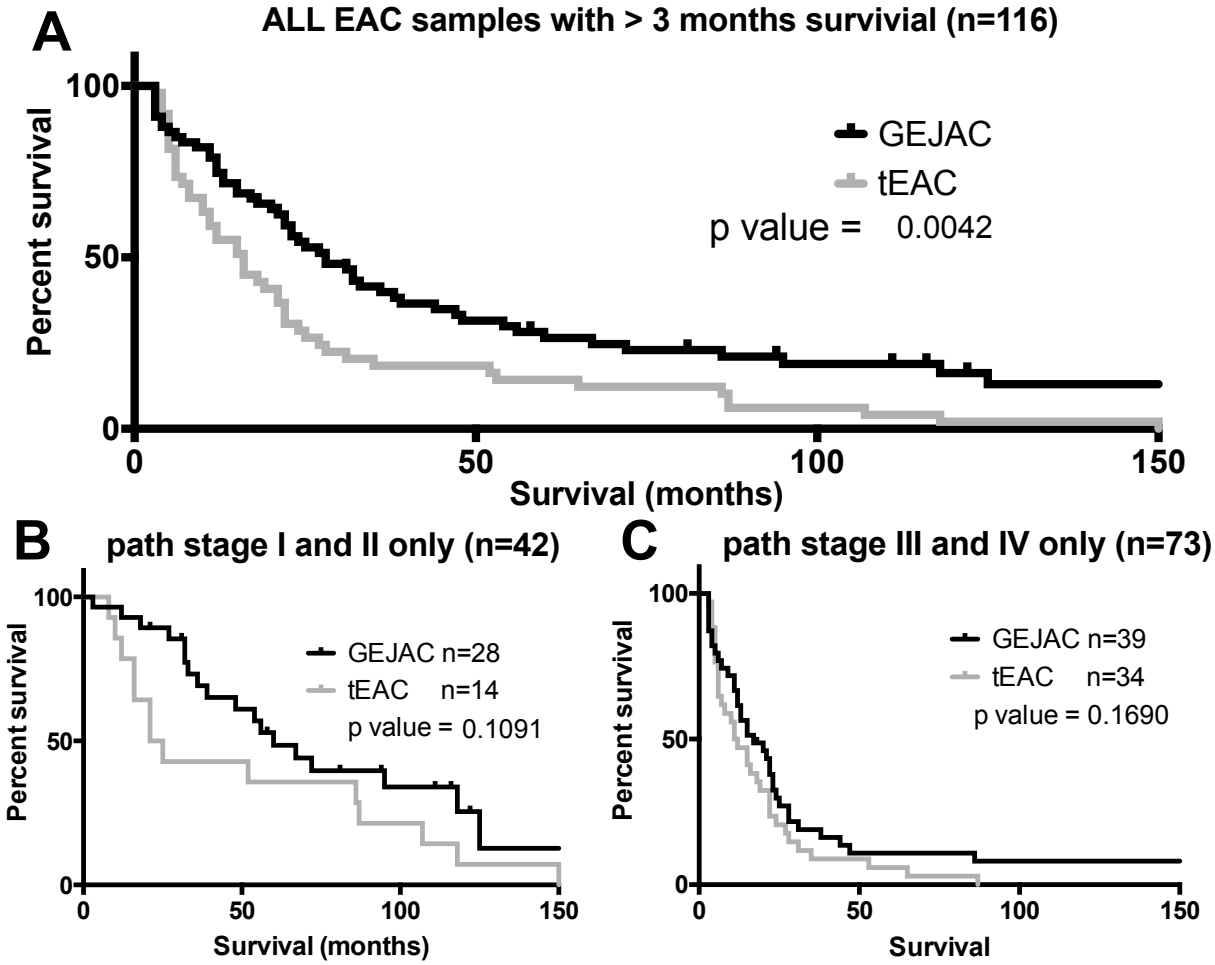

**Figure 2.7: PCA comparison of GEJAC noBE and tEAC with BE.** All annotated probe sets (n=26,613), were standardized by subtracting the tumor cohort mean and dividing by the SD. Only GEJACs with no histological evidence of BE (GEJAC no BE; n=54) and tEACs with BE (tEAC with BE; n=28) were considered. The first two principal components (each with variance >5%: (Figure 2.5) were plotted and individual samples were assigned either **A)** a location (GEJ or tubular esophagus) or **B)** an unsupervised clustering assignment based on Pearson-correlation on the same 26,613 probe sets (Figure 2.6) Visual and statistical comparison demonstrated minor expression differences between GEJAC no BE and tEAC with BE compared to class assignment by gene expression, as was seen when all GEJAC and tEAC samples were compared (Figure 2.4).




**D** **Univariate and multivariate Cox proportional hazard regression analyses for overall survival.**

| | Univariate | | Multivariate | |
|---|---|---|---|---|
| Covariates | Hazard Ratio (95% CI) | P-value | Hazard Ratio (95% CI) | P-value |
| ZNF217 | 2.31 (1.64-3.25) | 1.92E-06 | 1.9 (1.34-2.72) | 0.000374 |
| Stage | 2.5 (1.86-3.41) | 2.70E-09 | 2.4 (1.49-3.87) | 0.000312 |
| Node pos | 2.75 (1.66-4.56) | 8.98E-05 | 0.96 (0.47-1.96) | 0.91 |
| Location | 1.77 (1.2-2.62) | 0.0044 | 1.17 (0.74-1.85) | 0.513 |
| Tobacco | 1.79 (1.13-2.83) | 0.0134 | 1.41 (0.83-2.4) | 0.199 |

**Figure 2.8: Univariate and multivariate analyses of *ZNF217*.** Univariate Kaplan-Meijer estimates with patients survived more than 3 months post-surgery stratified into high or low risk on the basis of median log2 normalized *ZNF217* expression. Plots demonstrate higher *ZNF217* expression as a risk factor whether A) all EACs, B) GEJAC only and C) tEAC only patients are considered. D) Shows tabulated comparisons of univariate and multivariate Cox proportional hazard components for *ZNF217* expression in conjunction with key clinical factors.




**D** **Univariate and multivariate Cox proportional hazard regression analyses for overall survival (Location)**

| | Univariate | | Multivariate | |
|---|---|---|---|---|
| **Covariates** | **Hazard Ratio (95% CI)** | **P-value** | **Hazard Ratio (95% CI)** | **P-value** |
| Location | 1.77 (1.2-2.62) | 0.0044 | 1.36 (0.862-2.14) | 0.187 |
| Stage | 2.5 (1.86-3.41) | 2.70E-09 | 2.2 (1.38-3.48) | 0.000833 |
| Node pos | 2.75 (1.66-4.56) | 8.98E-05 | 1.03 (0.50-2.12) | 0.932 |
| Tobacco | 1.79 (1.13-2.83) | 0.0134 | 1.64 (0.98-2.74) | 0.0589 |

**Figure 2.9: Survival analyses comparing GEJAC and tEAC. A)** Univariate survival analysis shows GEJAC had improved survival over tEAC, however by looking at **B)** early and **C)** late stage tumors separately, as well as **D)** multivariate analysis, indicated this association is dependent upon tumor stage. Kaplan-Meijer plots and log-rank p=values were generated in Prism while tabulate univariate and multivariate analyses were made using the coxph module in R.

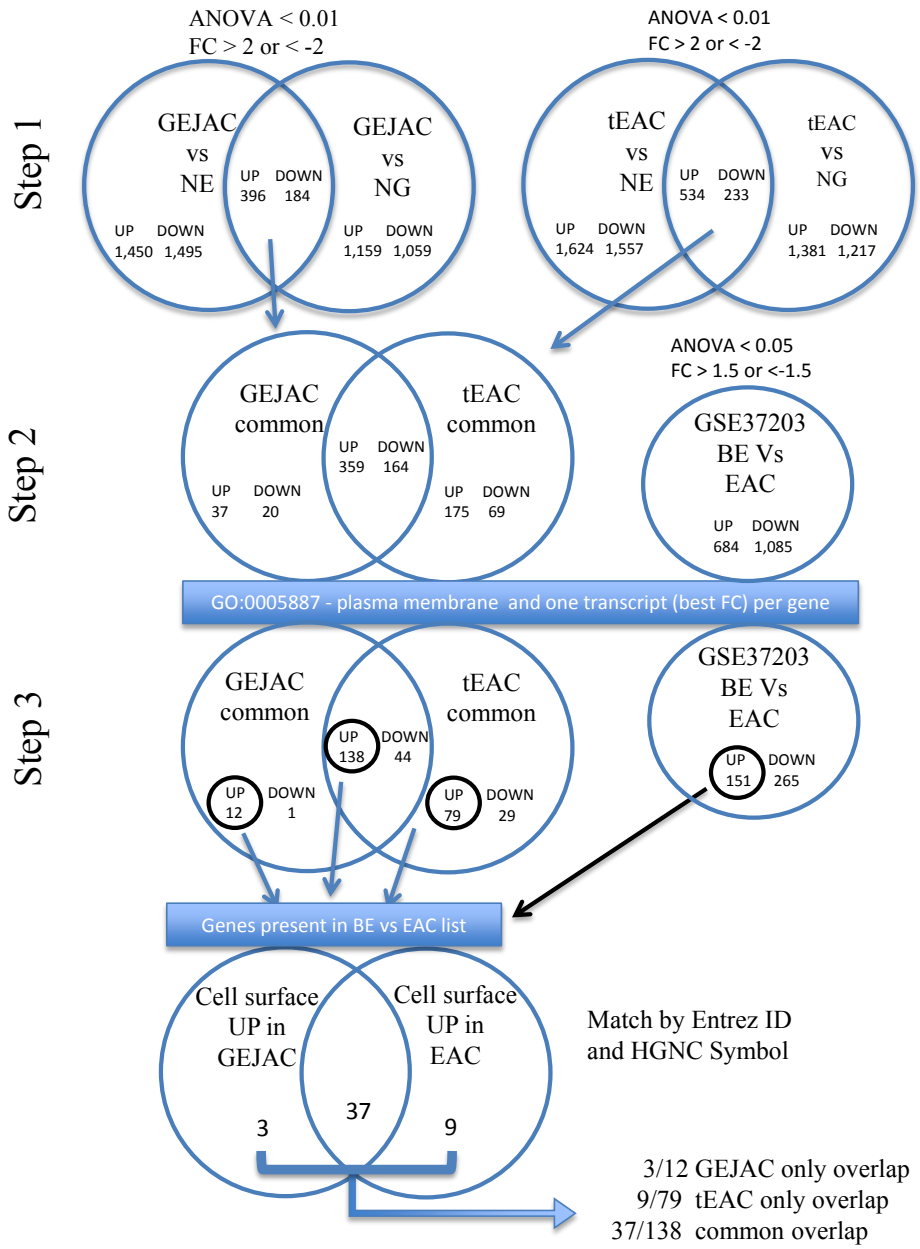



**Figure 2.10: Schematic of the steps used to identify potential cell surface markers for GEJAC and tEAC using expression profiling data.** Beginning with all 26,613 probe sets the schematic demonstrates the bioinformatic steps and associated gene numbers as we combine comparisons to each normal tissue group (NE and NG) to each of the tumor group (GEJAC and tEAC), selected the combined subset with known plasma membrane associations by GO and finally contrasted this list to those similarly overexpressed in EAC relative to BE in our previously published progression cohort (GEO ID: GSE37203). The resulting 49 genes represent potential cell surface factors overexpressed in EAC (both GEJAC and tEAC) relative to surrounding normal and pre-cancerous tissues.

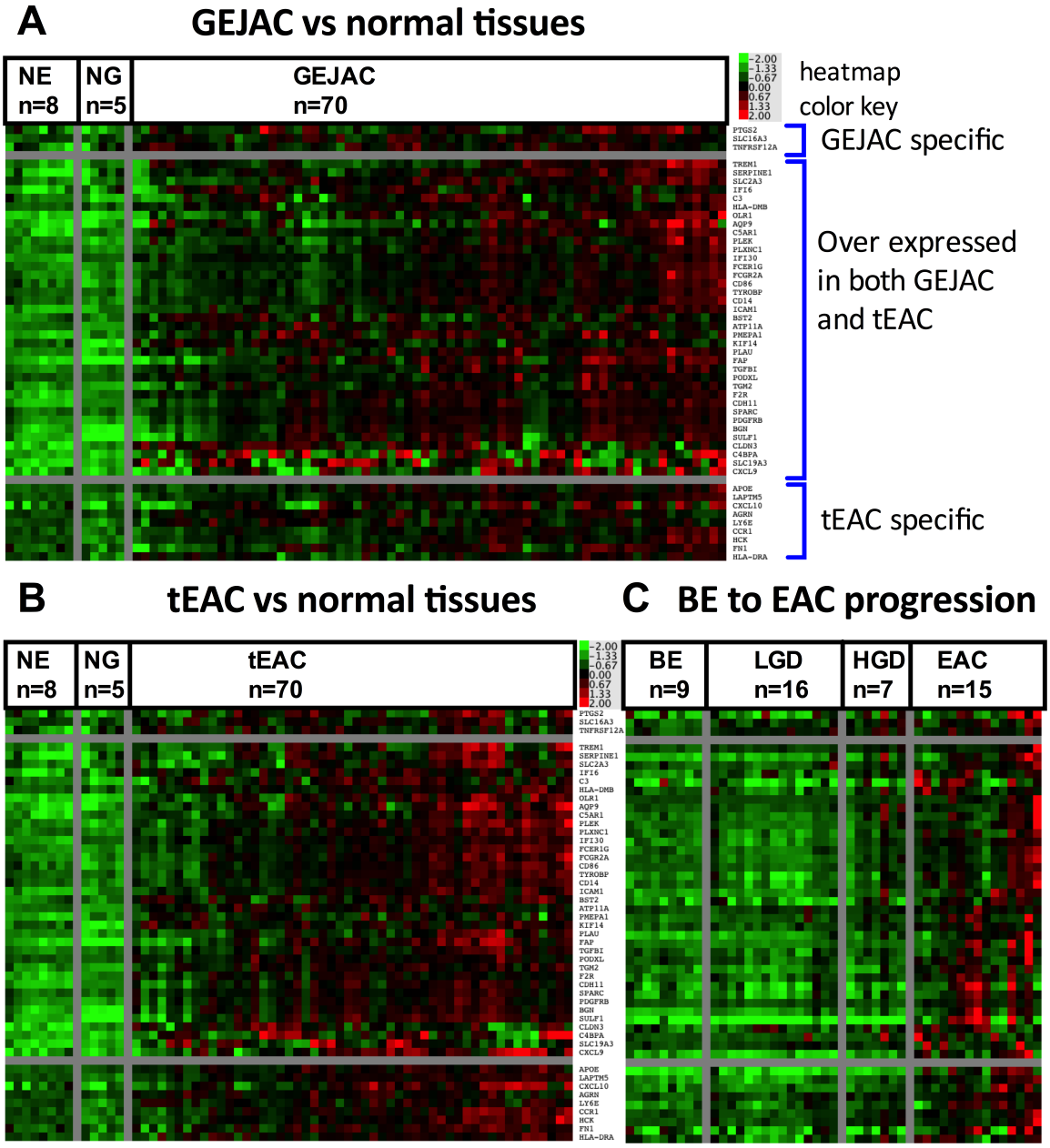



**Figure 2.11: Heatmap of potential cell surface coding genes for GEJAC and tEAC.** ANOVA and fold-change based comparisons, in conjunction with Gene Ontology data, were used to identify 49 genes potentially over-represented in GEJAC and tEAC, as outlined in **Figure 2.10** and Methods. Mean normalized expression was then applied to these data to sort expression patterns across **A)** GEJAC and **B)** tEAC relative to mRNA from normal tissues, as well as **C)** EAC relative to BE samples ordered by histology, taken from Gene Expression Omnibus (GEO) Series ID GSE37203) (Silvers *et al.* 2010). In each figure plate the top three genes only passed the overexpression threshold in GEJAC, the lower nine only passed in tEAC, while the central listed genes were selected in both cancer types. While all genes are generally more highly expressed in tumor groups as compared to the represented non-cancerous tissues, there was considerable variation between tumor samples, with no clear pattern in relation to GEJAC and tEAC.

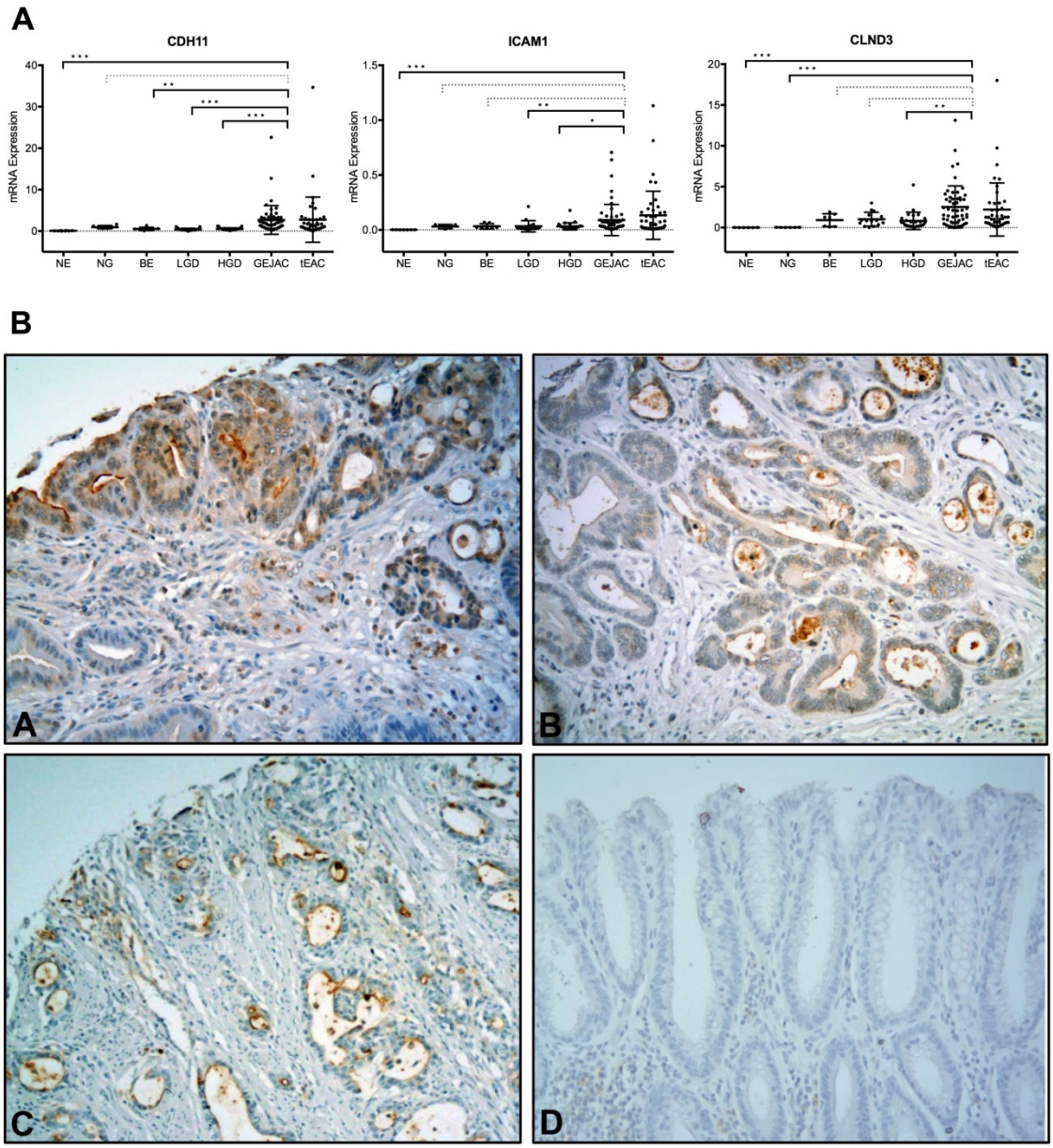



**Figure 2.12: qRT-PCR and protein validation of potential cell surface markers.** Three genes, *CDH11*, *ICAM1* and *CLDN3,* were chosen to confirm mRNA overexpression in EAC tumors, and demonstrate tumor cell surface staining of gene products. A) Relative qRT-PCR expression levels were determined for an esophagus-related panel of cDNA samples using the ABI PRISM® 7900HT technology and *GAPDH* as a reference gene, as described in Methods. GEJAC and tEAC groups were not significantly different for any gene (p=0.12, 0.65 and 0.17 by WMU). The combined cancers were compared to NE, NG, BE and HGD groups (grey no significant comparison, * $p<0.05$, ** $p<0.01$, *** $p<0.001$ by MWU) for each of the 3 genes. B) We then used a TMA containing histologically-confirmed EAC tissues to demonstrate that commercially available antibodies for A. CDH11, B. CLDN3, C. ICAM1 and D. no primary antibody negative control stained cell surface profiles localized to HGD and tumor cells.

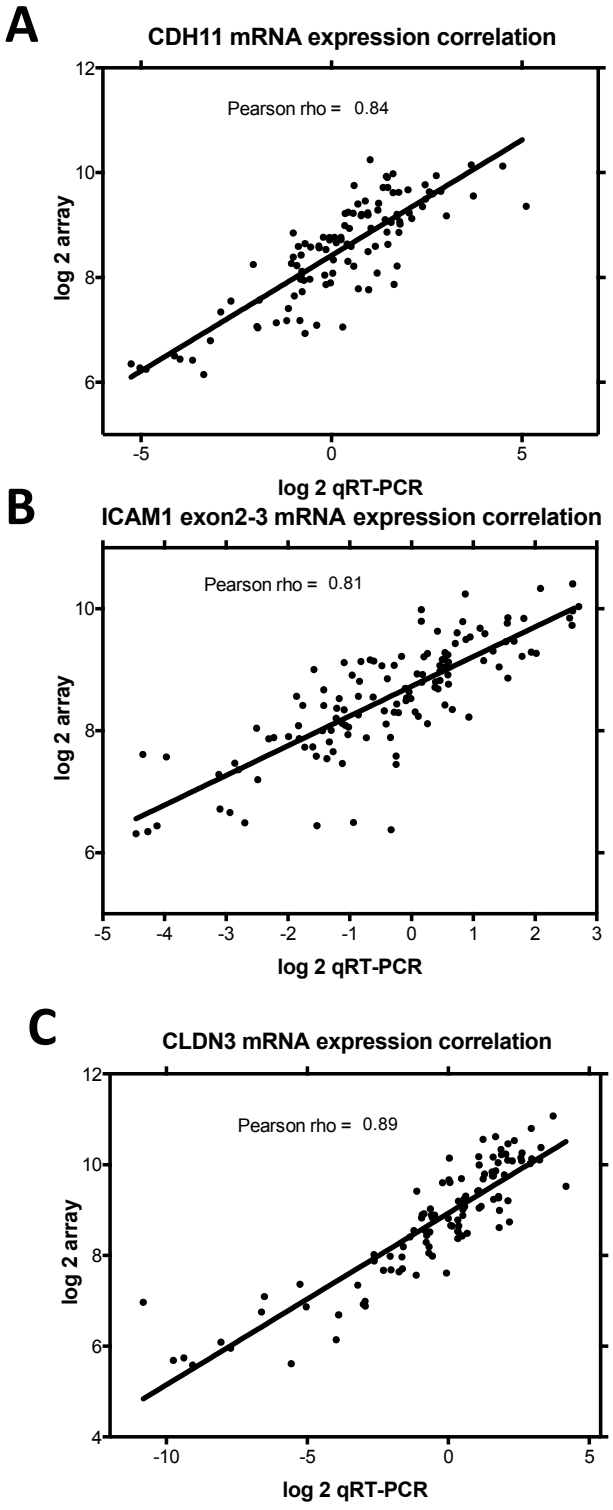



**Figure 2.13: Validation of ST 2.1 array data by qRT-PCR.** Pearson-correlation analyses for **A)** *CDH11*, **B)** *ICAM1* and **C)** *CLDN3* comparing log2 normalized Human Gene 2.1 ST arrayed samples and relative expression (qRT-PCR relative to *GAPDH*) data.

## TABLES

### Table 2.1: Clinical Characteristics

| | | GEJAC and tEAC | | | GEJAC no BE and tEAC with BE | | |
|---|---|---|---|---|---|---|---|
| | | **GEJAC n=70 (100%)** | **tEAC n=52 (100%)** | **p-value** | **GEJAC no BE n=54 (100%)** | **tEAC with BE n=28 (100%)** | **p-value** |
| Age | **median** | 70.3 | 68.1 | 0.266# | 70 | 69.7 | 0.914# |
| | **under 70** | 26 (37.1%) | 29 (55.8%) | | 20 (37.0%) | 14 (50.0%) | |
| | **over 70** | 44 (62.9%) | 23 (44.2%) | 0.0453^ | 34 (63.0%) | 14 (50.0%) | 0.345^ |
| Gender | | | | | | | |
| | **male** | 54 (77.1%) | 47 (90.4%) | | 40 (74.1%) | 25 (89.3%) | |
| | **female** | 16 (22.9%) | 5 (9.6%) | 0.088^ | 14 (25.9%) | 3 (10.7%) | 0.152^ |
| Tumor Stage | | | | | | | |
| | **I** | 9 (12.9%) | 4 (7.8%) | | 6 (11.1%) | 3 (10.7%) | |
| | **II** | 19 (27.1%) | 10 (19.6%) | | 12 (22.2%) | 8 (28.6%) | |
| | **III** | 37 (52.9%) | 30 (58.8%) | | 32 (59.3%) | 14 (50.0%) | |
| | **IV** | 5 (7.1%) | 7 (13.7%) | 0.11@ | 4 (7.4%) | 3 (10.7%) | 1.0@ |
| Node status | | | | | | | |
| | **negative** | 21 (30.0%) | 8 (18.2%) | | 14 (25.9%) | 7 (29.2%) | |
| | **positive** | 49 (70.0%) | 36 (81.8%) | 0.189^ | 40 (74.1%) | 17 (70.8%) | 0.787^ |
| Differentiation | | | | | | | |
| | **well** | 16 (22.9%) | 4 (7.7%) | | 12 (22.2%) | 3 (10.7%) | |
| | **moderate** | 22 (31.4%) | 19 (36.5%) | | 17 (31.5%) | 9 (32.1%) | |
| | **poor** | 32 (45.7%) | 29 (55.8%) | 0.066@ | 25 (46.3%) | 16 (57.1%) | 0.229@ |
| Desmoplasia | | | | | | | |
| | **low** | 25 (35.7%) | 14 (26.9%) | | 19 (35.2%) | 7 (25.0%) | |
| | **moderate** | 21 (30.0%) | 17 (32.7%) | | 15 (27.8%) | 11 (39.3%) | |
| | **high** | 24 (34.3%) | 21 (40.4%) | 0.379@ | 20 (37.0%) | 10 (35.7%) | 0.676@ |
| Lymphocytic infiltration | | | | | | | |
| | **low** | 24 (34.3%) | 12 (23.1%) | | 19 (35.2%) | 7 (25.0%) | |
| | **moderate** | 28 (40.0%) | 16 (30.8%) | | 20 (37.0%) | 9 (32.1%) | |
| | **high** | 18 (25.7%) | 24 (46.2%) | 0.039@ | 15 (27.8%) | 12 (42.9%) | 0.197@ |
| Tobacco usage | | | | | | | |
| | **no** | 19 (28.8%) | 16 (34.0%) | | 15 (29.4%) | 10 (40.0%) | |
| | **yes** | 47 (71.2%) | 31 (66.0%) | 0.68^ | 36 (70.6%) | 15 (60.0%) | 0.438^ |
| BE status | | | | | | | |
| | **no BE** | 54 (77.1%) | 24 (46.2%) | | 100% | | |
| | **+ BE** | 11 (15.7%) | 28 (53.8%) | 3.13E-05^ | | 100% | |
| | **unknown** | 5 (7.1%) | 0 (0%) | | | | |

#, t-test: ^, Fisher Exact Test: @, Mantel-Haenszel Chi-square test of association

**Table 2.2: Clinical characteristics of BE related subtypes for each cancer group**

| | | GEJAC with and without BE | | | tEAC with and without BE | | |
|---|---|---|---|---|---|---|---|
| | | **GEJAC no BE n=54* (100%)** | **GEJAC with BE n=11 (100%)** | **p-value** | **tEAC no BE n=24 (100%)** | **tEAC with BE n=28 (100%)** | **p-value** |
| Age | **median** | 70 | 72.3 | 0.695# | 66.2 | 69.7 | 0.173# |
| | **under 70** | 20 (37.0%) | 4 (36.4%) | | 15 (62.5%) | 14 (50.0%) | |
| | **over 70** | 34 (63.0%) | 7 (63.6%) | 1.00^ | 9 (37.5%) | 14 (50.0%) | 0.412^ |
| Gender | | | | | | | |
| | **male** | 40 (74.1%) | 9 (81.8%) | | 22 (91.7%) | 25 (89.3%) | |
| | **female** | 14 (25.9%) | 2 (18.2%) | 0.718^ | 2 (8.3%) | 3 (10.7%) | 1.00^ |
| Weight category | | | | | | | |
| **under weight** | **BMI < 18.5** | 1 (2.0%) | 0 (0.0%) | | 0 (0.0%) | 1 (4.0%) | |
| **normal weight** | **18.5 – 24.9** | 17 (34.0%) | 4 (36.4%) | | 5 (26.3%) | 6 (24.0%) | |
| **over weight** | **25.0 – 29.9** | 18 (36.0%) | 3 (27.3%) | | 10 (52.6%) | 12 (48.0%) | |
| **obese** | **30.0 and over** | 14 (28.0%) | 4 (36.4%) | 0.722@ | 4 (21.1%) | 6 (24.0%) | 0.906@ |
| Tumor stage | | | | | | | |
| | **I** | 6 (11.1%) | 2 (18.2%) | | 1 (4.3%) | 3 (10.7%) | |
| | **II** | 12 (22.2%) | 6 (54.5%) | | 2 (8.7%) | 8 (28.6%) | |
| | **III** | 32 (59.3%) | 3 (27.3%) | | 16 (69.6%) | 14 (50.0%) | |
| | **IV** | 4 (7.4%) | 0 (0.0%) | 0.0398@ | 4 (17.4%) | 3 (10.7%) | 0.0747@ |
| Node status | | | | | | | |
| | **negative** | 14 (25.9%) | 5 (45.5%) | | 1 (5.0%) | 7 (29.2%) | |
| | **positive** | 40 (74.1%) | 6 (54.4%) | 0.275^ | 19 (95.0%) | 17 (70.8%) | 0.0544^ |
| Differentiation | | | | | | | |
| | **Well** | 12 (22.2%) | 4 (36.4%) | | 1 (4.2%) | 3 (10.7%) | |
| | **moderate** | 17 (31.5%) | 3 (27.3%) | | 10 (41.7%) | 9 (32.1%) | |
| | **Poor** | 26 (46.3%) | 4 (36.4%) | 0.371@ | 13 (54.2%) | 16 (57.1%) | 0.841@ |
| Desmoplasia | | | | | | | |
| | **low** | 19 (35.2%) | 5 (45.4%) | | 7 (29.2%) | 7 (25.0%) | |
| | **moderate** | 15 (27.8%) | 6 (54.6%) | | 6 (25.0%) | 11 (39.3%) | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **High** | 20 (37.0%) | 0 (0.0%) | 0.0837@ | 11 (45.8%) | 10 (35.7%) | 0.793@ |
| Lymphocytic infiltration | | | | | | | |
| | **low** | 19 (35.2%) | 4 (36.4%) | | 5 (20.8%) | 7 (25.0%) | |
| | **moderate** | 20 (37.0%) | 6 (54.6%) | | 7 (29.2%) | 9 (32.1%) | |
| | **High** | 15 (27.8%) | 1 (9.1%) | 0.437@ | 12 (50.0%) | 12 (42.9%) | 0.187@ |
| Adjuvant treatment | | | | | | | |
| | **negative** | 42 (77.8%) | 8 (72.7%) | | 15 (62.5%) | 19 (67.9%) | |
| | **positive** | 12 (22.2%) | 3 (27.3%) | 0.706^ | 8 (33.3%) | 7 (25.0%) | 0.757^ |
| Tobacco usage | | | | | | | |
| | **No** | 15 (29.4%) | 16 (34.0%) | | 15 (29.4%) | 10 (40.0%) | |
| | **Yes** | 30 (70.6%) | 31 (66.0%) | 0.68^ | 36 (70.6%) | 15 (60.0%) | 0.438^ |
| BE status | | | | | | | |
| | **no BE** | 100% | | | 100% | | |
| | **+ BE** | | 100% | | | 100% | |
| | **unknown** | 0 (0.0%) | 0 (0.0%) | | | | |

#, t-test: ^, Fisher's exact test: @, Mantel-Haenszel Chi-square test: *, 5 individuals with unknown BE status were removed

## References


1. Kohler BA, Sherman RL, Howlader N, Jemal A, Ryerson AB, Henry KA, Boscoe FP, Cronin KA, Lake A, Noone AM, Henley SJ, Eheman CR, Anderson RN and Penberthy L. Annual report to the nation on the status of cancer, 1975- 2011, featuring incidence of breast cancer subtypes by race/ ethnicity, poverty, and state. *Journal of the National Cancer Institute*. **107**, (2015).
2. Kroep S, Lansdorp-Vogelaar I, van der Steen A, Inadomi JM and van Ballegooijen M. The Impact of Uncertainty in Barrett's Esophagus Progression Rates on Hypothetical Screening and Treatment Decisions. *Medical decision making*. **35,** 726–733 (2014).
3. Thrift AP and Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: Analysis of period and birth cohort effects on recent trends. *Annals of Oncology*. **23,** 3155–3162 (2012).
4. Wijetunge S and Suraj P. Changing trend of oesophageal adenocarcinoma and cardia carcinoma incidence; review of literature and analysis. *Journal of Diagnostic Pathology*. **6**, 17–27 (2012).
5. Cummings LC and Cooper GS. Descriptive epidemiology of esophageal carcinoma in the Ohio Cancer Registry. *Cancer Detect Prev*. **32**, 87–92 (2008).
6. Sudo K, Taketa T, Correa AM, Campagna M-C, Wadhwa R, Blum Ma, Komaki R, Lee JH, Bhutani MS, Weston B, Skinner HD, Maru DM, Rice DC, Swisher SG, Hofstetter WL and Ajani Ja. Locoregional failure rate after preoperative chemoradiation of esophageal adenocarcinoma and the outcomes of salvage strategies. *Journal of clinical oncology*. **31**, 4306–4310 (2013).
7. Siegel R, Ma J, Zou Z and Jemal A. Cancer statistics, 2014. CA *Cancer J Clin*. 2014; **64**, 9–29 (2014).
8. Davies AR, Gossage Ja, Zylstra J, Mattsson F, Lagergren J, Maisey N, Smyth EC, Cunningham D, Allum WH and Mason RC. Tumor Stage After Neoadjuvant Chemotherapy Determines Survival After Surgery for Adenocarcinoma of the Esophagus and Esophagogastric Junction. *Journal of Clinical Oncology*. **32**, 1–9 (2014).
9. Orringer MB, Marshall B, Chang AC, Lee J, Pickens A and Lau CL. Two thousand transhiatal esophagectomies: changing trends, lessons learned. *Annals of surgery*. **246**, 363–372 (2007).
10. Siewert JR, Feith M, Werner M and Stein HJ. Adenocarcinoma of the esophagogastric junction: results of surgical therapy based on anatomical/topographic classification in 1,002 consecutive patients. *Annals of surgery*. **232**, 353–361 (2000).
11. Verbeek RE, Leenders M, Ten Kate FJW, van Hillegersberg R, Vleggaar FP, van Baal JWPM, van Oijen MGH and Siersema PD. Surveillance of Barrett's esophagus and mortality from esophageal adenocarcinoma: a population-based cohort study. *The American journal of gastroenterology*. **109**, 1215–1222 (2014).
12. Wang KK and Sampliner RE. Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus. *The American journal of gastroenterology*. **103**, 788–797 (2008).
13. Sturm MB, Joshi BP, Lu S, Piraka C, Khondee S, Elmunzer BJ, Kwon RS, Beer DG, Appelman HD, Turgeon DK and Wang TD. Targeted imaging of esophageal neoplasia with a fluorescently labeled peptide: first-in-human results. *Science translational medicine*. **5**,161–184 (2013).

14. Bird-Lieberman EL, Neves Aa, Lao-Sirieix P, O'Donovan M, Novelli M, Lovat LB, Eng WS, Mahal LK, Brindle KM and Fitzgerald RC. Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus. *Nature medicine*. **18**, 315–321 (2012).
15. Thekkek N, Maru DM, Polydorides AD, Bhutani MS, Anandasabapathy S and Richards-Kortum R. Pre-Clinical Evaluation of Fluorescent Deoxyglucose as a Topical Contrast Agent for the Detection of Barrett's-Associated Neoplasia During Confocal Imaging. *Technology in cancer research & treatment.* **10**, 431–441 (2011).
16. Reid BJ, Li X, Galipeau PC and Vaughan TL. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nature reviews Cancer. **10**, 87–101 (2010).
17. Clark GW, Smyrk TC, Burdiles P, Hoeft SF, Peters JH, Kiyabu M, Hinder RA, Bremner CG and DeMeester TR. Is Barrett's metaplasia the source of adenocarcinomas of the cardia? *Archives of surgery*. **129**, 609–614 (1994).
18. Theisen J, Stein HJ, Dittler HJ, Feith M, Moebius C, Kauer WKH, Werner M and Siewert JR. Preoperative chemotherapy unmasks underlying Barrett's mucosa in patients with adenocarcinoma of the distal esophagus. *Surgical endoscopy*. **16**, 671–673
19. Carr JS, Zafar SF, Saba N, Khuri FR and El-Rayes BF. Risk factors for rising incidence of esophageal and gastric cardia adenocarcinoma. *Journal of gastrointestinal cancer*. **44**, 143–151 (2013).
20. Curtis NJ, Noble F, Bailey IS, Kelly JJ, Byrne JP and Underwood TJ. The relevance of the Siewert classification in the era of multimodal therapy for adenocarcinoma of the gastro-oesophageal junction. *Journal of Surgical Oncology*. **109**, 202–207 (2014).
21. Grant KS, DeMeester SR, Kreger V, Oh D, Hagen JA, Chandrasoma P and DeMeester TR. Effect of Barrett's esophagus surveillance on esophageal preservation, tumor stage, and survival with esophageal adenocarcinoma. *J Thorac Cardiovasc Surg*. **146**, 31–37 (2013).
22. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, Bandla S, Imamura Y, Schumacher SE, Shefler E, McKenna A, Carter SL, Cibulskis K, Sivachenko A, Saksena G, Voet D, *et al*. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature genetics*. **45**, 478–486 (2013).
23. Nones K, Waddell N, Wayte N, Patch AM, Bailey P, Newell F, Holmes O, Fink JL, Quinn MC, Tang YH, Lampe G, Quek K, Loffler KA, Manning S, Idrisoglu S, Miller D, *et al*. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nature communications*. **5**, 5224 (2014).
24. Siewert JR, Feith M and Stein HJ. Biologic and clinical variations of adenocarcinoma at the esophago-gastric junction: Relevance of a topographic-anatomic subclassification. *Journal of Surgical Oncology*. **90**,139–146 (2005).
25. Silvers AL, Lin L, Bass AJ, Chen G, Wang Z, Thomas DG, Lin J, Giordano TJ, Orringer MB, Beer DG and Chang AC. Decreased selenium-binding protein 1 in esophageal adenocarcinoma results from posttranscriptional and epigenetic regulation and affects chemosensitivity. *Clinical Cancer Research*. **16**, 2009–2021 (2010).
26. Leicht DT, Kausar T, Wang Z, Ferrer- Torres D, Wang TD, Thomas DG, Lin J, Chang AC, Lin L and Beer DG. TGM2 A Cell Surface Marker in Esophageal Adenocarcinomas. *J Thorac Oncol.* **9**, 872–881 (2014).
27. Buas MF and Vaughan TL. Epidemiology and risk factors for gastroesophageal junction tumors: understanding the rising incidence of this disease. *Seminars in radiation oncology*.

**23**, 3–9 (2013).
28. Isinger-Ekstrand A, Johansson J, Ohlsson M, Francis P, Staaf J, Jönsson M, Borg Å and Nilbert M. Genetic profiles of gastroesophageal cancer: combined analysis using expression array and tiling array–comparative genomic hybridization. *Cancer Genetics and Cytogenetics*. **200**, 120–126 (2010).
29. Littlepage LE, Adler AS, Kouros-Mehr H, Huang G, Chou J, Krig SR, Griffith OL, Korkola JE, Qu K, Lawson DA, Xue Q, Sternlicht MD, Dijkgraaf GJ, Yaswen P, Rugo HS, Sweeney CA, *et al*. The transcription factor ZNF217 is a prognostic biomarker and therapeutic target during breast cancer progression. *Cancer discovery*. **2**, 638–651 (2012).
30. Banck MS, Li S, Nishio H, Wang C, Beutler AS and Walsh MJ. The ZNF217 oncogene is a candidate organizer of repressive histone modifiers. *Epigenetics*. **4**, 100–106 (2009)
31. Quinlan KG, Verger A, Yaswen P and Crossley M. Amplification of zinc finger gene 217 (ZNF217) and cancer: when good fingers go bad. *Biochimica et biophysica acta*. **1775**, 333–340 (2007).
32. Frankel A, Armour N, Nancarrow D, Krause L, Hayward N, Lampe G, Smithers BM and Barbour A. Genome-wide analysis of esophageal adenocarcinoma yields specific copy number aberrations that correlate with prognosis. *Genes, chromosomes & cancer.* **53**, 324–338 (2014).
33. Gu J, Ajani JA, Hawk ET, Ye Y, Lee JH, Bhutani MS, Hofstetter WL, Swisher SG, Wang KK and Wu X. Genome- wide catalogue of chromosomal aberrations in barrett's esophagus and esophageal adenocarcinoma: a high-density single nucleotide polymorphism array analysis. *Cancer prevention research.* **3**, 1176–1186 (2010).
34. Nancarrow DJ, Handoko HY, Smithers BM, Gotley DC, Drew PA, Watson DI, Clouston AD, Hayward NK and Whiteman DC. Genome-wide copy number analysis in esophageal adenocarcinoma using high-density single- nucleotide polymorphism arrays. *Cancer research.* **68**, 4163–4172 (2008).
35. Pasello G, Agata S, Bonaldi L, Corradin A, Montagna M, Zamarchi R, Parenti A, Cagol M, Zaninotto G, Ruol A, Ancona E, Amadori A and Saggioro D. DNA copy number alterations correlate with survival of esophageal adenocarcinoma patients. *Modern pathology.* **22**, 58–65 (2009).
36. Fang Z, Xiong Y, Zhang C, Li J, Liu L, Li M, Zhang W and Wan J. Coexistence of copy number increases of ZNF217 and CYP24A1 in colorectal cancers in a Chinese population. *Oncology letters.* **1**, 925–930 (2010).
37. Ginestier C, Cervera N, Finetti P, Esteyries S, Esterni B, Adelaide J, Xerri L, Viens P, Jacquemier J, Charafe-Jauffret E, Chaffanet M, Birnbaum D and Bertucci F. Prognosis and gene expression profiling of 20q13-amplified breast cancers. *Clinical cancer research.* **12**, 4533–4544 (2006).
38. Vendrell JA, Thollet A, Nguyen NT, Ghayad SE, Vinot S, Bieche I, Grisard E, Josserand V, Coll JL, Roux P, Corbo L, Treilleux I, Rimokh R and Cohen PA. ZNF217 is a marker of poor prognosis in breast cancer that drives epithelial- mesenchymal transition and invasion. *Cancer research.* **72**, 3593–3606 (2012).
39. Geppert CI, Rummele P, Sarbia M, Langer R, Feith M, Morrison L, Pestova E, Schneider-Stock R, Hartmann A and Rau TT. Multi-colour FISH in oesophageal adenocarcinoma-predictors of prognosis independent of stage and grade. *British journal of cancer.* **110**, 2985–2995 (2014).
40. Cowger JJ, Zhao Q, Isovic M and Torchia J. Biochemical characterization of the zinc-finger

protein 217 transcriptional repressor complex: identification of a ZNF217 consensus recognition sequence. *Oncogene.* **26**, 3378–3386 (2007).
41. Szczyrba J, Nolte E, Hart M, Doll C, Wach S, Taubert H, Keck B, Kremmer E, Stohr R, Hartmann A, Wieland W, Wullich B and Grasser FA. Identification of ZNF217, hnRNP-K, VEGF-A and IPO7 as targets for microRNAs that are downregulated in prostate carcinoma. International journal of cancer. **132**, 775–784 (2013).
42. Li Z, Du L, Dong Z, Yang Y, Zhang X, Wang L, Li J, Zheng G, Qu A and Wang C. MiR-203 suppresses ZNF217 upregulation in colorectal cancer and its oncogenicity. *PloS one*. **10**, e0116170 (2015).
43. Bai WD, Ye XM, Zhang MY, Zhu HY, Xi WJ, Huang X, Zhao J, Gu B, Zheng GX, Yang AG and Jia LT. MiR-200c suppresses TGF-beta signaling and counteracts trastuzumab resistance and metastasis by targeting ZNF217 and ZEB1 in breast cancer. *International journal of cancer.* **135**, 1356–1368 (2014).
44. Ding X, Park SI, McCauley LK and Wang CY. Signaling between transforming growth factor beta (TGF-beta) and transcription factor SNAI2 represses expression of microRNA miR-203 to promote epithelial-mesenchymal transition and tumor metastasis. *The Journal of biological chemistry.* **288**, 10241–10253 (2013).
45. Hezova R, Kovarikova A, Srovnal J, Zemanova M, Harustiak T, Ehrmann J, Hajduch M, Svoboda M, Sachlova M and Slaby O. Diagnostic and prognostic potential of miR- 21, miR-29c, miR-148 and miR-203 in adenocarcinoma and squamous cell carcinoma of esophagus. *Diagnostic pathology.* **10**, 42 (2015).
46. Wijnhoven BP, Hussey DJ, Watson DI, Tsykin A, Smith CM, Michael MZ and South Australian Oesophageal Research G. MicroRNA profiling of Barrett's oesophagus and oesophageal adenocarcinoma. *The British journal of surgery*. **97**, 853–861 (2010).
47. Frietze S, O'Geen H, Littlepage LE, Simion C, Sweeney CA, Farnham PJ and Krig SR. Global analysis of ZNF217 chromatin occupancy in the breast cancer cell genome reveals an association with ERalpha. *BMC genomics.* **15**, 520 (2014).
48. Carmona FJ, Davalos V, Vidal E, Gomez A, Heyn H, Hashimoto Y, Vizoso M, Martinez-Cardus A, Sayols S, Ferreira HJ, Sanchez-Mut JV, Moran S, Margeli M, Castella E, Berdasco M, Stefansson OA, *et al*. A comprehensive DNA methylation profile of epithelial-to-mesenchymal transition. *Cancer research*. **74**, 5608–5619 (2014).
49. Shirvani VN, Ouatu-Lascar R, Kaur BS, Omary MB and Triadafilopoulos G. Cyclooxygenase 2 expression in Barrett's esophagus and adenocarcinoma: Ex vivo induction by bile salts and acid exposure. *Gastroenterology*. **118**, 487–496 (2000).
50. Watts GS, Tran NL, Berens ME, Bhattacharyya AK, Nelson MA, Montgomery EA and Sampliner RE. Identification of Fn14/TWEAK receptor as a potential therapeutic target in esophageal adenocarcinoma. *International journal of cancer*. **121**, 2132–2139 (2007).
51. Fonteyne P, Casneuf V, Pauwels P, Van Damme N, Peeters M, Dierckx R and Van de Wiele C. Expression of hexokinases and glucose transporters in treated and untreated oesophageal adenocarcinoma. *Histology and histopathology*. **24**, 971–977 (2009).
52. Miller SJ, Lee CM, Joshi BP, Gaustad A, Seibel EJ and Wang TD. Targeted detection of murine colonic dysplasia in vivo with flexible multispectral scanning fiber endoscopy. *Journal of biomedical optics*. **17**, 021103 (2012).
53. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, *et al*. Mutational heterogeneity in cancer and the search for new cancer-

associated genes. *Nature*. **499**, 214–218 (2013).
54. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B and Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research*. **31**, e15 (2003).
55. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, *et al*. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*. **5**, R80 (2004).
56. Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics*. **20**, 3246—3248 (2004).
57. Huang DW, Sherman BT and Lempicki Ra. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*. **37**, 1–13 (2009).
58. Huang DW, Sherman BT and Lempicki Ra. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. **4**, 44–57 (2009).
59. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC and Lempicki Ra. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*. **8**, R183 (2007).
60. Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R and Jensen LJ. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database *Oxford*. **2014**:bau012 (2014).
61. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S and Madden TL. Primer-BLAST: a tool to design target- specific primers for polymerase chain reaction. *BMC bioinformatics*. **13**, 134 (2012).
62. Lin L, Bass AJ, Lockwood WW, Wang Z, Silvers AL, Thomas DG, Chang AC, Lin J, Orringer MB, Li W, Glover TW, Giordano TJ, Lam WL, Meyerson M and Beer DG. Activation of GATA binding protein 6 (GATA6) sustains oncogenic lineage-survival in esophageal adenocarcinoma. *Proceedings of the National Academy of Sciences of the United States of America*. **109**, 4251–4256 (2012).
63. Rubie C, Kempf K, Hans J, Su T, Tilton B, Georg T, Brittner B, Ludwig B and Schilling M. Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Molecular and cellular probes*. **19**, 101–109 (2005).
64. Kononen J, Bubendorf L, Kallioniemi A, Bärlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G and Kallioniemi OP. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature medicine*. **4**, 844–847 (1998).

# CHAPTER 3

## Transcriptional Characterization of Progression of Barrett's Esophagus to Esophageal Adenocarcinoma


### Summary

Patients with gastroesophageal reflux disease (GERD) often develop a metaplastic condition called Barrett's esophagus (BE), but only a small subset of patients with BE progress to esophageal adenocarcinoma (EAC). The process of neoplastic transformation from Barrett's esophagus (BE) to EAC is a stepwise process involving progression from non-dysplastic BE mucosa to low-grade dysplasia (LGD), to high-grade dysplasia (HGD) and eventually adenocarcinoma. Patients with HGD have a >40% chance of developing EAC, making the HGD lesion the gold standard in the field for identification of EAC risk. Current strategies to diagnose the presence of HGD and EAC are through random biopsy of the esophagus and histologic assessment of the tissue to identify the high-risk tissue. Caveats of this approach include the potential missing of these events due to random sampling. Therefore, there is a current need for the development of new prevention strategies in patients with premalignant Barrett's metaplasia, better methods for early cancer detection, and refining new treatment modalities to improve patient survival. In this work, we sought to characterize the molecular events that are associated with the progression of BE to LGD, HGD, and EAC, and to identify biomarkers for immunohistochemistry as well as a cell surface over-expressed marker in the HGD/EAC tissue that may serve as candidates for peptide-based imaging. Using RNA-sequencing on a progression series of 66 samples with different degrees of dysplasia and cancer (BE → LGD →

HGD → EAC) we have identified deregulation of the splicing pathway, down-regulation of mucin protection, and subsequent increase in the DNA damage response pathway (DDR). In addition, for early diagnosis purposes, we identify two cell surface coding genes, *C3* and *HLA-DRB5,* that are overexpressed in the HGD/EAC and cancer, when compared to non-dysplastic BE or normal gastric and esophagus. ROC curves show that the expression levels of these markers are significantly sensitive and specific to differentiate between non-dysplastic BE vs. HGD/EAC and between normal tissues (esophagus and gastric) vs. tumors (EAC/GEJAC).

**Introduction**

In the United States, it is estimated that in 2013, 17,990 people will be diagnosed with esophageal cancer, and greater than 85% of those diagnosed will succumb to the disease. One of the significant problems EAC patients face is that diagnosis tends to occur at an advanced stage, resulting in an overall poor 5-year survival rate of less than 15% [1]. There are some common risk factors for EAC incidence, such as a history of gastro-esophageal reflux disease (GERD), obesity, male gender, Caucasian and over the age of 40[2]. Many patients with EAC also have associated Barrett's esophagus (BE). BE is a predisposing condition where the normal squamous of the esophagus is replaced by a columnar intestinal-type epithelium, yet, only around 1% of patients with BE are estimated to develop EAC[3]. There is currently no molecular marker(s) that effectively identify patients at highest risk for development of EAC. The current clinical practice for detection of EAC is by screening patients with high-risk GERD using an endoscopic-based sampling of the metaplastic tissue and characterization of the degree of dysplasia in the biopsy samples by pathology[4,5]. Once enrolled in the screening program they undergo endoscopic-biopsy every three months to two years depending on the degree of dysplasia. During the endoscopic procedure, four-quadrant biopsy samples are taken every 1 to 2 cm of the BE mucosa

and evaluated for histologic dysplastic changes by expert pathologists[4,5]. Even though some studies have shown that this method is associated with detection of EAC at an earlier stage, improving patient outcome[7], one of the problems with this approach is the random sampling of the endoscopic biopsies which may miss small regions of HGD or EAC.

Our group has suggested a more targeted approach, where we develop fluorescently-labeled peptides that can target the cancer cells. Importantly, a recent study by our group shows the potential for detection of EAC *in vivo* in human patients, through the use of fluorescently-labeled peptides detected by endoscopy after administration of the peptide into the esophagus[8]. This new method of detection can be used to detect tumors located in the distal esophagus and gastric cardia and we have begun to identify additional cell surface proteins that can be targeted by fluorescent peptides for early cancer identification.

Using transcriptional profiling of BE progression and analysis of cellular processes we show, first, that BE progression is associated with loss of mucin, increased ATM-DNA damage response, and increased splicing. Secondly, we identify two molecular targets (*C3* and *HLA-DRB5*) that significantly distinguish HGD/EAC tissue when compared to non-dysplastic BE. Additionally, we show that these candidates are over-expressed in EAC and GEJAC tumors when compared to normal tissue. In conclusion, we propose these candidates for future peptide target development of the HGD/EAC tissue.

## Results

### Transcriptional Profiling of Barrett's Metaplasia, Dysplasia, and Esophageal Adenocarcinoma Progression

EAC is known to arise from the dysplastic progression of BE (**Figure 3.1a**). Therefore, to identify the processes/pathways deregulated in this dysplastic progression we performed careful

cryostat-sectioning and macro-dissection of the mucosa in BE. We collected samples representing non-dysplastic BE, BE containing different degrees of dysplasia, and EAC samples with >70% tumor cells (BE n=5, LGD n=20, HGD n=29 and EAC n=11). These samples were carefully characterized by a pathologic review preformed by Dr. Scott Owens. The percentage of each histology in 10% increments for each BE sample was defined (**Figure 3.1b**). The macro-dissected sections were used to extract mRNA and we then performed RNA-sequencing (RNA-seq), a comprehensive method for the analysis of the whole transcriptome[9] in this large cohort of pathologically confirmed non-dysplastic, LGD, HGD and EAC tissues (n=66). In addition to the RNA-seq cohort, we examined our previously published progression cohort, where we had BE-LGD-HGD and EAC samples examined by Affymetrix arrays (n=46). Using these complementary methods, we proceeded to identify the pathways that are deregulated in BE-dysplastic progression to cancer and characterize the potential reasons for the deregulation of these pathways. In addition, we identify cell surface coding genes that are over-expressed in HGD/EAC.

**Pathway analysis: Spliceosome.**

After organizing the samples by the degree of dysplasia to EAC (**Figure 3.1b**), we correlated the expression values (RPKM) for the whole gene transcripts (n=57,284) per sample to the degree of dysplasia (**Figure 3.2a**). We identified that 954 genes show positive correlation and over-expressed in HGD and EAC relative to BE and LGD (**Figure 3.2a** right panel). We also observed 1,193 genes that are negatively correlated with BE progression to EAC and are decreased in expression in HGD and EAC as compared to BE and LGD (**Figure 3.2a** left panel). This progression set nominates the genes that are progressively being over-expressed/down-expressed in the progression of BE dysplasia to cancer. To identify the pathways/processes

deregulated in progression, we performed pathway analysis using DAVID-genome browser. We identified spliceosome as the top pathway increased during BE progression to EAC (**Figure 3.2b)**. We confirmed that many of the spliceosome-related genes are also increased using the independent cohort of BE to EAC samples analyzed using Affymetrix arrays (**Figure 3.3a-b**), thus supporting the observation of increased spliceosome pathway during progression of BE to EAC.

The spliceosome machinery is a large complex of proteins responsible for removing introns from transcribed pre-mRNA, and creating isoform diversity. To determine if the splicing machinery is hyperactive in the progression of BE to EAC, we utilized the RNA-seq data, and determined isoform specific expression levels in collaboration with Dr. Hui Jiang and correlated them to the splicing over-expressed genes. We observed a total of 194,872 isoform specific transcripts (**Figure 3.4a**, blue circle). The total cohort of splicing machinery-coding genes is comprised of 35 genes (**Table 3.1**). We calculated a z-score to represent the overall level of expression per sample. We further identified the number of isoforms differentially expressed per sample, and calculated an isoform diversity score. We observed a significant correlation of splicing with increased isoform diversity from BE to EAC (r = 0.04787, $P$ = 4.792e-05)(**Figure 3.3c**). Further, we characterized the isoforms that are positively or negatively correlated with dysplastic progression (**Figure 3.4**). We observed 886 isoform transcripts that significantly change during progression ($P$ <0.05)(**Figure 3.4a**, yellow circle). **Figure 3.3b** depicts an example of a gene transcript, *RNF128*, which has three isoforms. Two isoforms of *RNF128* show unique expression in the progression (**Figure 3.3b**, top and middle panel), and one isoform has non-detectable levels of mRNA (**Figure 3.3b**, bottom panel). We assessed whether isoforms of the same gene show a "switch" in progression, suggesting a deregulation in the ratio of

expression of those isoforms in the dysplastic progression. We identified 241 isoforms depicting a "switch" between their isoforms (**Figure 3.4a**, green circle). Of those, we identified the isoforms that show a significant "switch" and are correlated to progression, with *RNF128* isoforms being one of the most significant (**Figure 3.4a**, orange box)(**Figure 3.4d**). Together, these observations support a role of splicing and increased isoform diversity in the dysplastic progression of BE to EAC.

**Loss of mucin and activation of the DNA damage response.**

We have established that splicing and isoform diversity is increased in HGD and EAC, but the potential mechanism or activator of the splicing machinery leading to more isoform diversity is currently unknown in the context of BE progression to EAC. Recently, the core spliceosome has been found to be a target and effector of non-canonical ATM signaling[10]. Therefore, we further examined the potential events associated with increased spliceosome activity in BE to EAC progression.

A mucus barrier is responsible for the protection of epithelial cells in the context of BE mucosa, that represent the intestinal type cells that are known to be at most risk of developing into dysplasia. We observed that the mucin genes (*MUC5B, MUC3A, MUC17*) are significantly decreased in both RNA-seq and Affymetrix cohorts (**Figure 3.5a-c**) with a corresponding loss of goblet cells in LGD and especially HGD (**Figure 3.5e, arrows)** seen in histopathologic preparations (H/E sections) (**Figure 3.5d-e**). Concurrent with the loss of mucins, which functionally protect cells against acid/bile reflux induced cellular damage, we observe increased mRNA expression of *ATM, MRE11, P53BP1,* and *RAD50* (**Figure 3.6a-d**) in BE progression suggesting activation of the "DNA damage response" (DDR). We further assessed the activation of DDR, by using antibodies to phospho-H2AX (γH2AX) in BE progression samples (**Figure**

**3.7**). We observed γH2AX increased positive staining of dysplastic BE cells but not non-dysplastic glands (**Figure 3.7a**, bottom panel). We further assessed the levels of mucin stains using alcian blue-periodic acid Schiff reagent (**Figure 3.7**). After staining for mucin, in BE tissue, we observe high mucin staining in BE and in mucin producing glands. This staining is reduced or lost in dysplastic areas that also show abundant apoptotic cells (**Figure 3.7**, top panel). Importantly, IHC analysis with γH2AX reveals increased positive staining of dysplastic BE cells but not in non-dysplastic glands. This suggests that loss of protective mucins in dysplastic BE may lead to DNA damage and initiate phosphorylation of H2AX by the ATM kinase.

**Identification of cell surface markers for HGD.**

The presence of BE is often associated with esophageal adenocarcinoma (EAC)[11–15]. Patients with BE have an estimated 30 to 40-fold greater risk of developing EAC than the general population[12,13]. Diagnosis of EAC often occurs at an advanced stage resulting in an overall poor 5-year survival rate of less than 15%. The current clinical practice for detection of early EAC consists of Barrett's patients having multiple biopsies taken by endoscopy and pathological characterization of the degree of dysplasia in each biopsy. This current method has the serious limitation of random sampling which can miss the exact location where dysplastic and/or cancer cells are arising. The malignant potential of this condition is evidenced by the progression of non-dysplastic BE to LGD, HGD and finally to invasive EAC. Most studies have shown that at the time of HGD removal, there is pathological evidence of EAC present within the HGD. Therefore, patients with HGD are the most 'at risk' for developing EAC. Detection and identification of molecular markers that can identify the high-risk patients could potentially have positive effects in the diagnosis of the EAC at the earliest stage of cancer development and

improve patient outcome.

By utilizing cell surface fluorescent peptides targeted at cancer rising cells, we can greatly improve the BE biopsy sampling error[8]. Since HGD represents the tissue with highest cancer risk, we utilized the RNA-seq progression cohort to identify cell surface coding genes that are over-expressed in the HGD/EAC cells when compared to non-dysplastic BE or LGD.

We performed an analysis of variance (ANOVA) across all transcript (n=57,284) in the BE-EAC progression and we identified those that were overexpressed in HGD when compared to non-dysplastic BE (n=954) (**Figure 3.8a-b**). Each transcript/gene was annotated using gene ontologies for 'plasma membrane' and/or "integral to plasma membrane" (n=4,000) coding genes which resulted in n=134 candidates that were both overexpressed in HGD ($p<0.001$, FC>2) and potentially located on the cell surface (**Figure 3.9a-b**). We identified two candidate genes (*C3* and *HLA-DRB5*) that were overexpressed in HGD and EAC when compared to BE and LGD (**Figure 3.9a-b**, top panel, **Figure 3.10a-b**, top panel, respectively). In addition, we looked at the expression levels of *C3* and *HLA-DRB5* in our previously published cohort, of normal (esophagus and gastric) vs. EAC/GEJAC[16]. **Figure 3.11a-b**, top panel, shows that *HLA-DRB5* and *C3* are over-expressed in EAC/GEJAC when compared to normal tissue (FC>1.5, P<0.01). Finally, sensitivity vs. specificity analysis showed that both, *C3* and *HLA-DRB5*, mRNA have a significant capability to distinguish HGD/EAC when compared to BE/LGD (mRNA comparisons of *HLA-DRB5*: ND-BE/LGD vs. HGD/EAC ROC: 0.6629, *P*=0.0240; ND-BE vs. HGD/EAC ROC=0.6863, *P*=0.0166; ND-BE vs. HGD ROC=0.7899, *P*=0.0029) (**Figure 3.10b**, bottom panel). From both candidates, *C3* showed the most significant in sensitivity vs. specificity (mRNA comparisons of *C3*: ND-BE/LGD vs. HGD/EAC ROC: 0.7647, *P*<0.0002; ND-BE vs. HGD/EAC ROC=0.8438, *P*<0.0001; ND-BE vs. HGD ROC=0.8229, *P*=0.0002)

(**Figure 3.9b,** bottom panel), nevertheless, both markers show significance in ROC analysis when we performed a more restricted analysis with just non-dysplastic BE and HGD (*HLA-DRB5* ND-BE vs. HGD ROC=0.7899, *P*=0.0029; *C3* ND-BE vs. HGD ROC=0.8229, *P*=0.0002) (**Figure 3.9b-3.10b**). We further assessed the mRNA levels of *HLA-DRB5* and *C3* in our previously published cohort of the normal esophagus and gastric tissue vs. EAC/GEJAC[16]. We observed that both *HLA-DRB5* and *C3* are significantly over-expressed in EAC/GEJAC when compared to normal tissue (esophagus and gastric) ($P < 0.0001$) (**Figure 3.11 a-b**). Even further, ROC analysis between normal vs. tumor shows that *HLA-DRB5* and *C3* mRNA expression significantly distinguishes between normal tissues vs. EAC/GEJAC. (ROC= 0.7530, 0.8974; $P=0.0028$, $P < 0.0001$, respectively). Finally, we used data from the publicly available domain Human Protein Atlas[17,18], to show that in gastric cancer, both HLA-DRB3 and C3, have high protein expression and this expression localizes mainly to the cytoplasm/membranous (C/M) location (**Figure 3.9-3.10c-d**). Altogether, we show that *C3* and *HLA-DRB5* are over-expressed cell surface coding genes with significant ROC curves to distinguish HGD/EAC from BE/LGD.

## Discussion

### Up-regulation of Splicing Coupled with Decreased Protective Mucins and Activation of the ATM/DNA Damage Response in BE Progression to EAC

It is known that BE progresses to low-grade dysplasia, high-grade dysplasia and then to invasive esophageal adenocarcinoma. Significant genomic instability, high alterations of copy number, and a high rate of *TP53* mutations are all molecular events that characterize EAC from other cancers[19–23]. Here, using two independent patient cohorts and (Affymetrix arrays (n=46) and RNA-seq (n=66) to examine gene expression during the progression of BE to EAC, we have shown that the splicing pathway, as well as increased isoform diversity, is observed in the

dysplastic progression from BE to EAC. It has been observed in other cancers (such as breast, lung, and ovary), that splicing-associated genes were indeed expressed in tumors when compared to normal tissue [24]. These studies did not previously show any association of splicing with more isoform diversity. Therefore, our comprehensive analysis, using RNA-seq, provides new insights into the spectrum of diversity and the level of isoform expression in the context of BE dysplastic progression to EAC. This diversity may allow BE cells to survive under stress for multiple sources such as reflux, inflammation or other causes.

Further, we show the loss of protective mucin and an increase in DDR pathway is associated with the progression of BE to EAC. ATM has been recently implicated in the regulation of splicing [10]. This suggests that loss of protective mucins in dysplastic BE may lead to DNA damage and initiate phosphorylation of H2AX by the ATM kinase. As reported [10], the core spliceosome is a target and effector of non-canonical ATM signaling. Thus we suggest that increased spliceosome pathway during BE progression to EAC reflects both DNA damage and ATM-dependent signaling events in dysplastic BE cells.

In the future, investigation of the function of the differential isoforms, as well as mechanistic characterization of the direct role of ATM in splicing activation, may lead to potential therapeutic approaches, or preventive strategies for reducing BE progression into EAC. Since we observed an increase in DNA damage and splicing, one potential approach will be to treat esophageal Barrett's cell lines, with both DNA damage inducing agents, as well as, bile acid. Using RNA-sequencing we will look at the differential isoform patterns of the treated vs. non-treated. We expect to see an increase in DNA damage response genes, such as *RAD50*, *TP53BP1*, *MRE11*, and *ATM*, as well as, an increase in the isoform diversity pattern of the cells treated vs. non-treated cell lines.

**Identification of cell surface markers for HGD.**

The current clinical practice for detection of early EAC consists of Barrett's patients having multiple biopsies taken by endoscopy and pathological characterization of the degree of dysplasia in each biopsy. This current method has the serious limitation of random sampling which can miss the exact location where dysplastic and/or cancer cells are arising. Dr. Wang in collaboration with our group, has suggested a more targeted imaging-based approach, where we develop fluorescently-labeled peptides that bind to highly overexpressed cell surface proteins found in cancer cells. In a recent first in-human patient study, our group demonstrated the potential for cancer detection of EAC *in vivo* [8]. The approach utilized phage display technology to identify peptides that bind specifically to the plasma membrane of human H460 cancer cells. Confirmation of specific binding was performed using *ex vivo* resected human esophageal specimens containing HGD and EAC. This imaging-based approach greatly reduces the issue of random sampling.

Here, we have proposed two cell surface coding genes, *C3* and *HLA-DRB5*, for the detection of HGD. We show that *C3* and *HLA-DRB5* are over-expressed in the HGD and EAC tissue. Most studies have shown that at the time of HGD removal, there is pathological evidence of EAC rising within the HGD[25]. Therefore, patients with HGD are the most 'at risk' for developing cancer[26]. Although future validation and tissue microarrays are needed, we suggest that *C3* and *HLA-DRB5* coupled with endoscopy could be used as detection markers to identify the high-risk epithelium and potentially help in the diagnosis of the EAC at the earliest stage of cancer development. In addition, it would enhance the current methods of EAC diagnosis and hopefully eradicate the sampling error that current surveillance programs face. This could potentially, not only improve patient outcome but also would be cost effective. The current

protocols enrolling patients in the screening program requires biopsies of the esophagus for assessment of the degree of dysplasia. Our targeted approach would allow stratification of patients at greatest 'risk' of developing cancer from those at lower risk, therefore; only patients that are at greatest risk would require close monitoring.

The identification of molecular markers for endoscopic imaging that identify patients at 'high-risk' for developing cancer is innovative and a new strategy that could improve current medical procedures monitoring patients with BE and improve patient outcome. This method may be able to increase the overall 5-year patient survival of patients with EAC, which is currently very low.

## Materials and Methods

### Patient tissues.

Samples were obtained following informed patient consent according to the approval and guidelines of the University of Michigan institutional review board. Tissues were obtained from patients undergoing esophagectomy for adenocarcinoma within the University of Michigan Health System between 1991 and 2012, without preoperative radiation or chemotherapy. All specimens were collected fresh and frozen in liquid nitrogen and stored at -80 C until use, and with regards to EAC patients, collected prior to any preoperative radiation or chemotherapy.

Cryostat sections of Barrett's mucosa allowed the careful macro-dissection of the epithelium prior to isolation of mRNA and DNA. Tissues from patients undergoing resection for HGD or EAC were histopathologically characterized [27] for the percentage of low-grade (LGD) or high-grade dysplasia (HGD) in 10% increments.

### mRNA extraction and Affymetrix expression analysis.

RNA was isolated from the matching samples used for pathologic review. mRNA extraction was performed using QIAzol Lysis Reagent (Qiagen, Valencia, CA) and purified with RNeasy spin columns (Qiagen), including on-column DNAse I incubation, according to the manufacturer's instructions. RNA samples with RIN scores greater than 7.0 (Agilent Bioanalyzer; Agilent Technologies, Palo Alto, CA) were submitted to the University of Michigan Sequencing Core. Paired-end sequence analysis of 120 million 100bp reads per lane using Illumina sequencers was used per sample. Strand-specific RNAseq libraries were prepared [28].

**RNA sequencing.**

Hui Jiang, Ph.D. developed the pipeline for defining the gene expression levels[28]. Sequence alignment was performed using Bowtie (0.12.8) which was supplied with the set of transcript models annotated in the *Homo sapiens* ensemble database version 72 (GENCODE v.17) with 57,281 annotated genes and 194,871 annotated transcripts. Read alignment was performed for all 66 samples, a total of 3.7 billion, strand-specific read pairs (i.e., 7.4 billion reads) were sequenced, resulting in 2.5 billion (68%) read pairs that aligned uniquely (range: 26-51 million; median: 36 million). Quantification of gene and transcript expression levels was performed using rSeq (0.1.0) and defined as RPKM (reads per kilobase of exon model per million mapped reads) unit. Statistical analysis was performed on log-transformed expression levels [log(RPKM+1)] to avoid taking the logarithm of zero. Data analyses included two sample t-test and Wilcoxon rank-sum test between BE and LGD vs. HGD and EAC as well as between BE and EAC and Spearman correlation test between gene expression levels and disease progression grades. Wilcoxon rank-sum tests were more powerful than t-tests for this data set likely due to non-normality of expression levels. Fold-change (FC) of expression levels between

BE and LGD vs. HGD and EAC was summarized as the ratio of the median expression levels between the two groups. False discovery rates (FDR) were computed using Storey's method.

**Immunohistochemistry.**

Tissue sections were cryostat and mounted. We proceeded to stain using the monoclonal-mouse antibody for γ-H2AX (Cat. No 05-636- Anti-phospho-Histone H2A.X (Ser139), clone JBW301, Millipore). The monoclonal antibody was used at dilutions of 1:500, after microwave citric acid epitope retrieval for 20 minutes and lightly counterstained with hematoxylin.

**Pathway analysis.**

We correlated the level of the transcripts to the progression of dysplasia and identify two groups, one group with positive correlation (n=954) and a group with negative correlation (n=1,193) (**Figure 3.2**). We performed enrichment testing, for the positively or negatively correlated gene transcripts. We performed enrichment testing using the DAVID website with the appropriate platform-specific background gene list ("HuEx-1_0-st-v2") and default algorithm settings. Individual ontology categories with false discovery adjusted (Benjamini) p values $<0.05$ were reported.

**Isoform Diversity**

Dr. Hui Jiang descries that the score measures how evenly distributed the transcript expression levels are, therefore, the more evenly distributed the higher the score. The score takes its maximum value when all the transcripts have the same expression level. In contrast, it takes its minimum value when only one transcript is expressed highly and all other transcripts are not expressed.

**Figures**

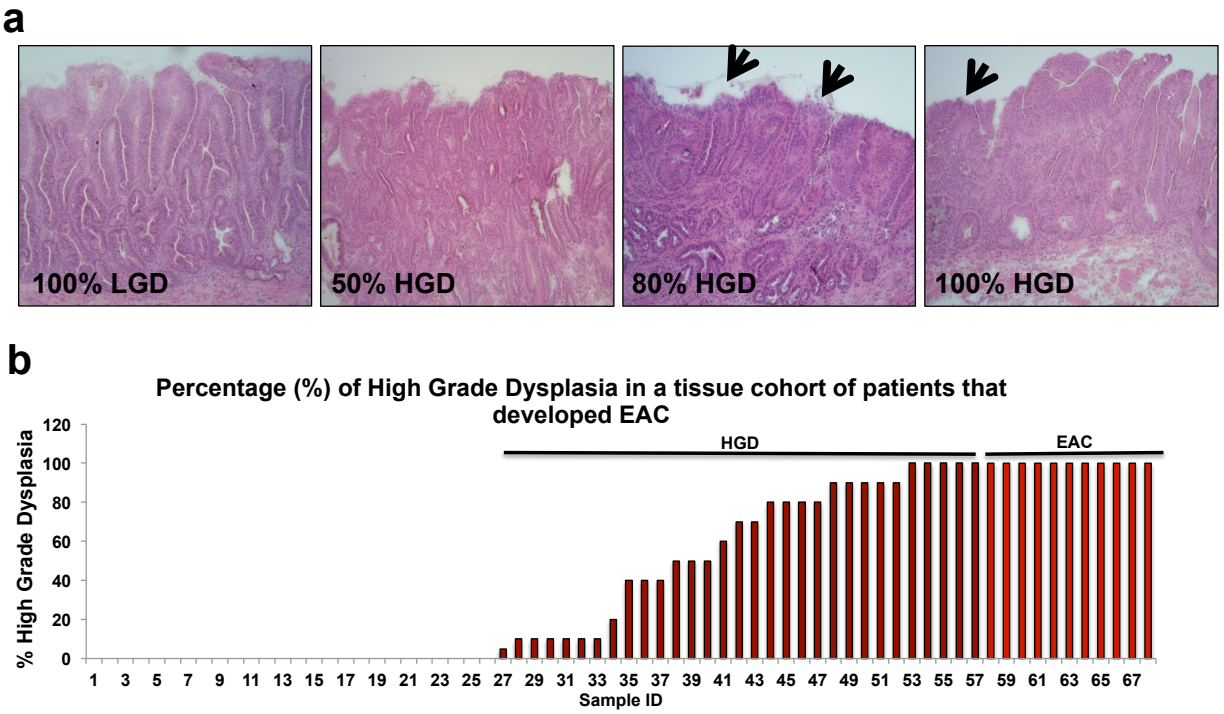



**Figure 3.1 Tissue characterization for dysplastic progression of BE to EAC** (a) Tissues from patients undergoing resection for HGD or EAC were histopathologically characterized for the percentage of low-grade (LGD) or high-grade dysplasia (HGD) in 10% increments. Arrows indicate the HGD tissue/cells are frequently localize in the surface of the epithelium. (b) 68 BE tissues from patients undergoing resection for either HGD (n=29) or invasive EAC (n=11) were utilized for RNA-seq, histologically characterized, and organized based on the percentage (%) of HGD. The LGD samples are also organized by increasing percentage of LGD however only the increasing percentage of HGD is shown, along with 11 associated EACs.

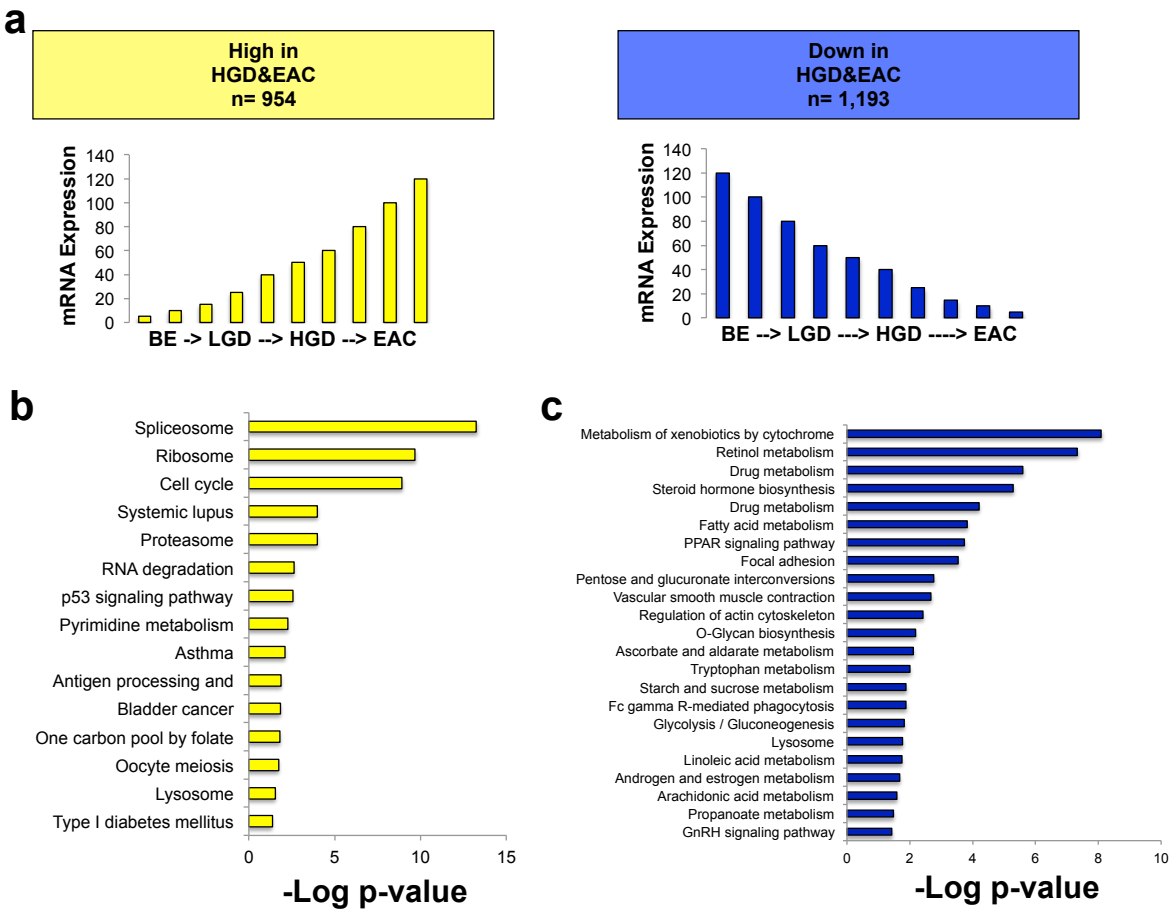



**Figure 3.2 Pathway analysis in BE dysplastic progression to EAC** (a) Identification of transcripts over-expressed (n=954) and down-regulated (n=1193) in the progression from BE to EAC. These transcripts were used for pathway analysis, using DAVID-genome browser (b-c) in BE progression to EAC and reveals spliceosome (b) as the top increased pathway and metabolism of xenobiotics by cytochrome P450 (c) the top reduced pathway.

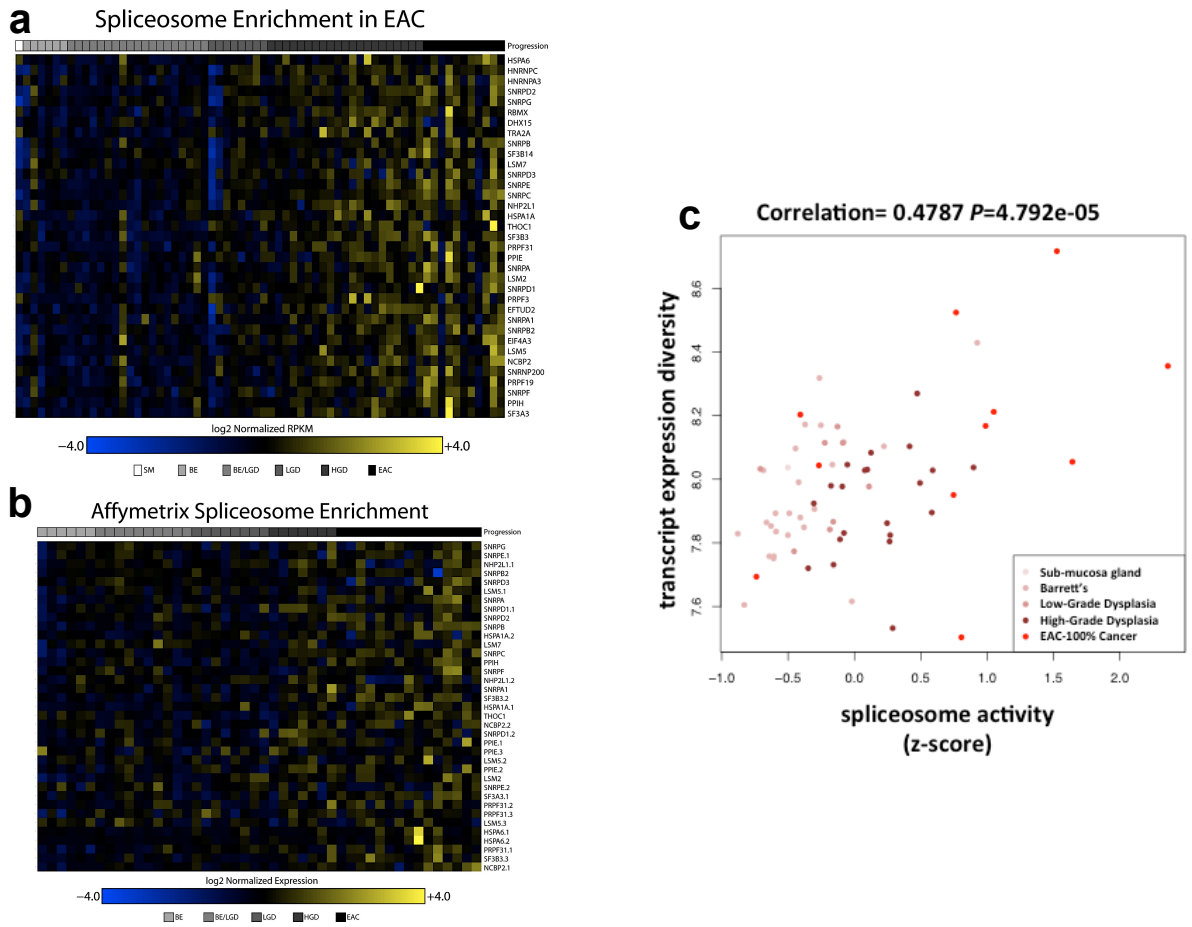



**Figure 3.3 Splicesome and isoform diversity in BE dysplastic progression to EAC** (a) RNAseq and (b) Affymetrix array (46 BE-EAC samples) analysis showing individual increased spliceosome genes during progression of BE to EAC. Affymetrix array of genes with increased expression are shown in yellow. (c) Bioinformatics analysis of RNA-seq data for isoform-variant expression diversity significantly correlated with higher expression of splicing related genes.

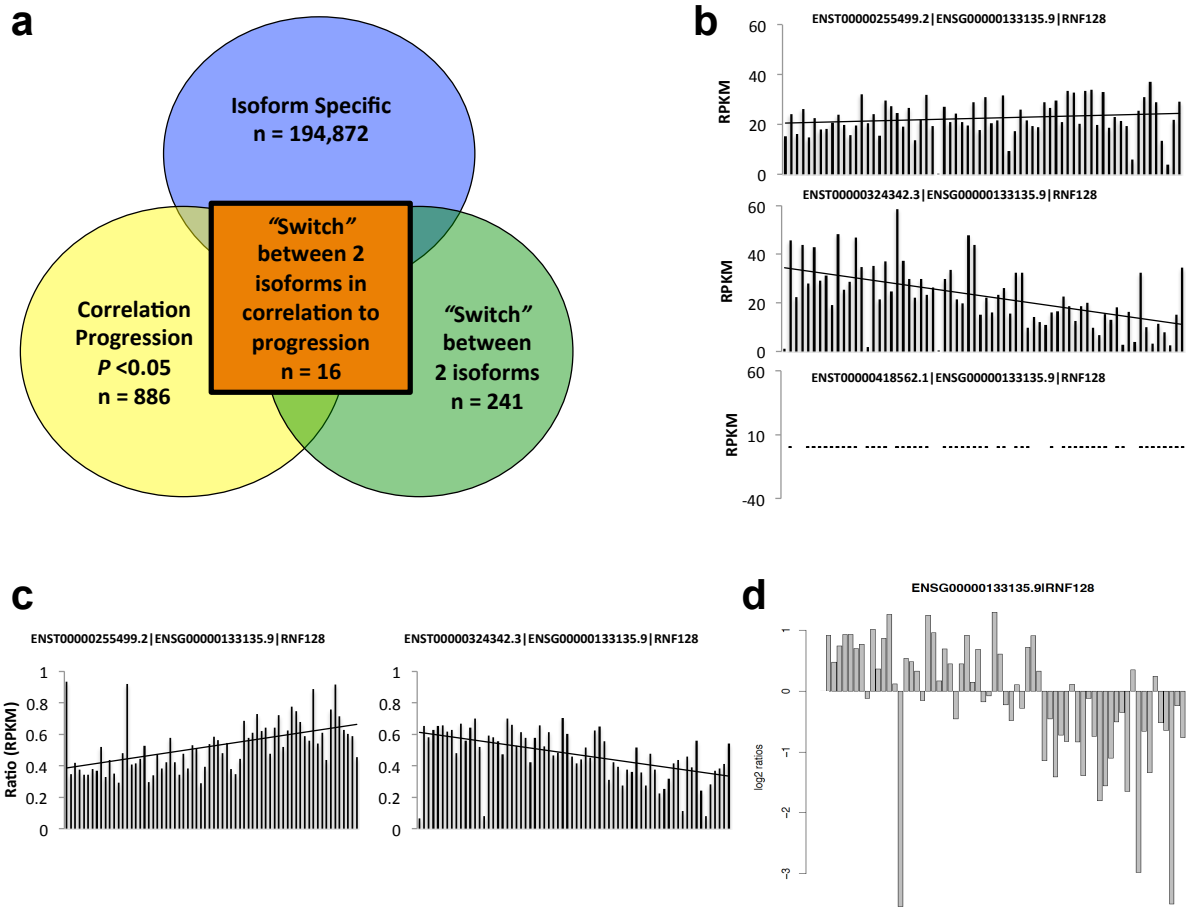



**Figure 3.4 Identification of Isoform specific transcripts that "switch" during BE dysplastic progression to EAC** (a) Using RNA-seq data we calculated RPKMs for isoform specific transcripts (n=194,872), (b) depicts the RPKM of *RNF128* and its three different isoforms. Only two were expressed. We calculated the ratio of each isoform to whole gene transcripts, with (c) depicting *RNF128* ratios and correlation to BE dysplastic EAC progression. We observed (a-yellow) 886 isoform transcripts, which correlate to progression. We further assessed whether any of these isoforms show a "switch" (one isoform going up and/or one going down in BE → EAC progression (c). We identify 16 genes that have different isoforms that show a "switch" in progression (a-orange), with (d) *RNF128* isoform showing a "switch" in isoform expression in BE → EAC progression.

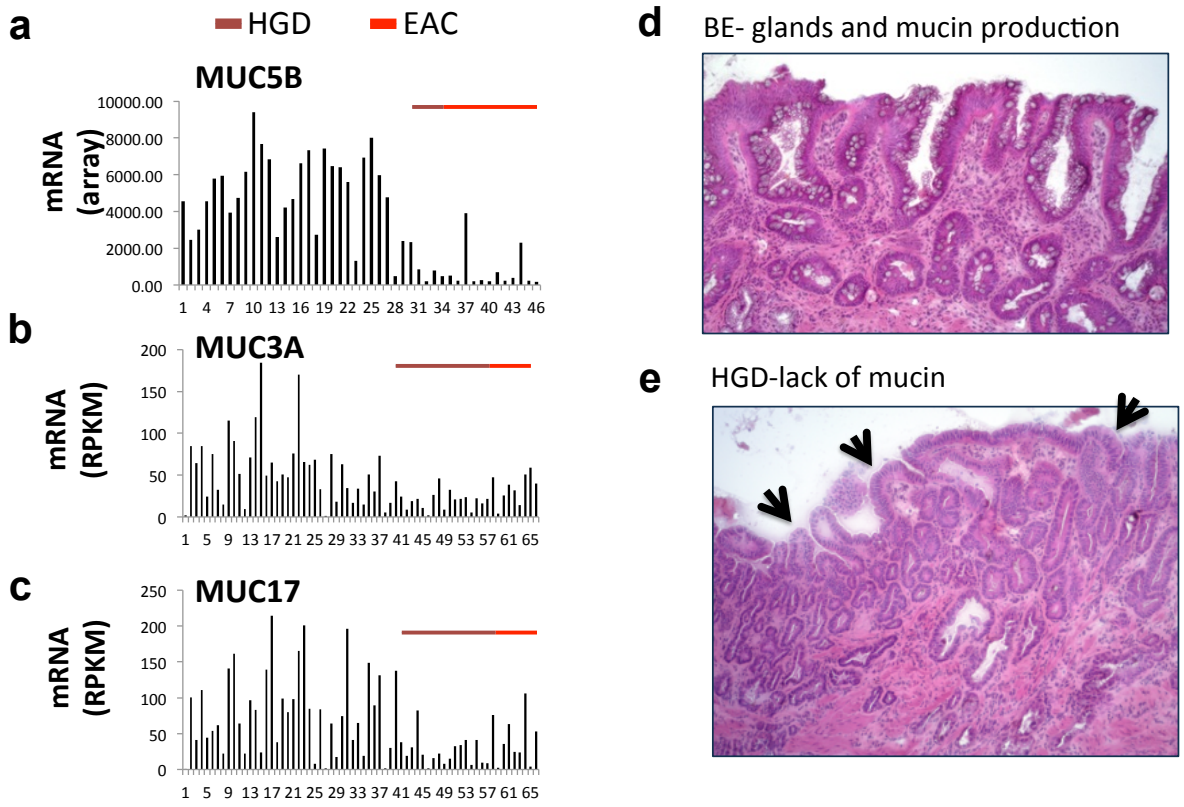



**Figure 3.5 Loss of mucin production in BE progression** (a) Affymetrix and (b-c) RNA-seq analysis showing reduced mucin gene expression (*MUC5AB, 3A, 17*) in progression of BE to EAC and particularly in HGD and EAC. (d-e) Loss of mucin gene expression is consistent with abundant goblet cells in nondysplastic BE that are completely lost in HGD (H/E sections).

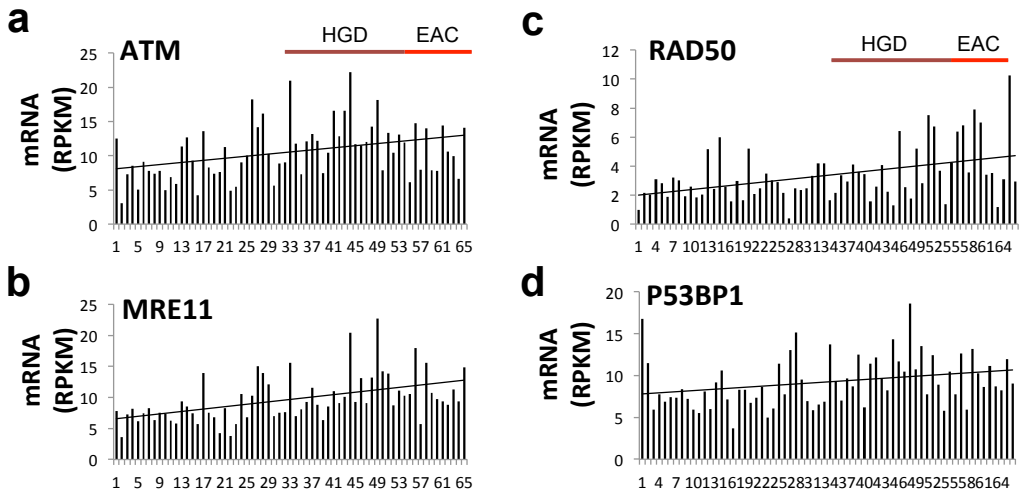



**Figure 3.6 Increased expression of genes that are part of the DNA Damage Response**
(a-d) Over-expression of the ATM/DNA damage response (DDR) pathway genes (*ATM, MRE11, P53BP1,* and *RAD50*) in dysplastic progression to EAC.

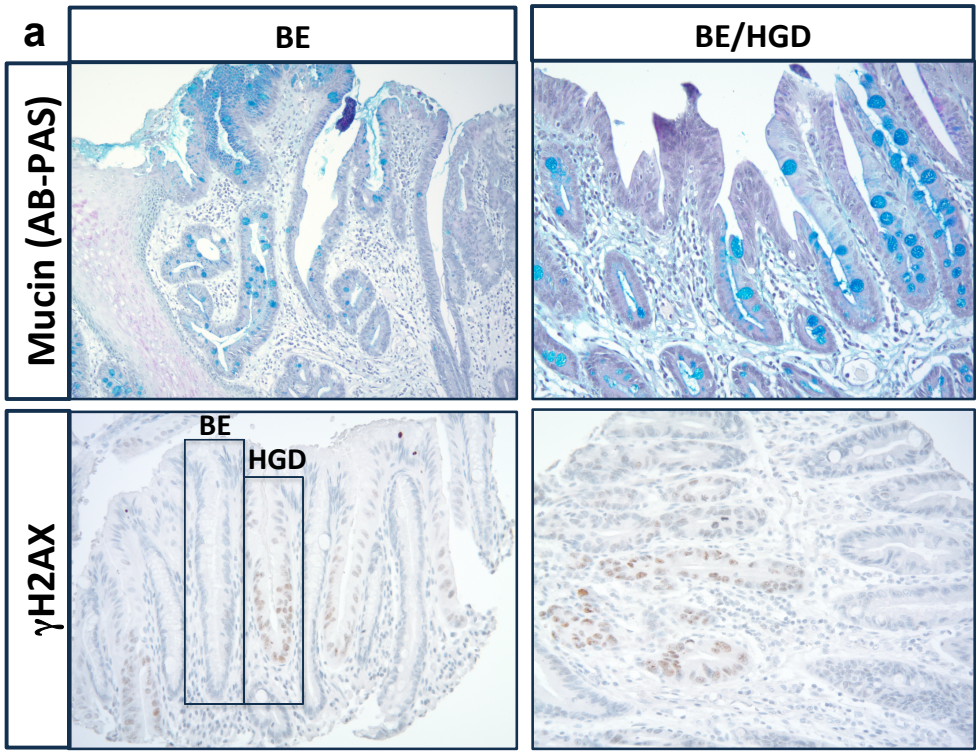



**Figure 3.7 Phosphorylation of H2AX (γ-H2AX) in HGD** (a) Top panel- Mucin staining: abundant blue mucin is present in non-dysplastic BE goblet cells but are lost in dysplastic BE cells. Bottom-panel- γ-H2AX nuclear staining is detected in dysplastic but not non-dysplastic BE cells. Non-dysplastic BE cells clearly show basally located nuclei, mucin-positive goblet cells present, whereas the dysplastic BE cells in neighboring glands have disorganized, stratified nuclei and no mucin.

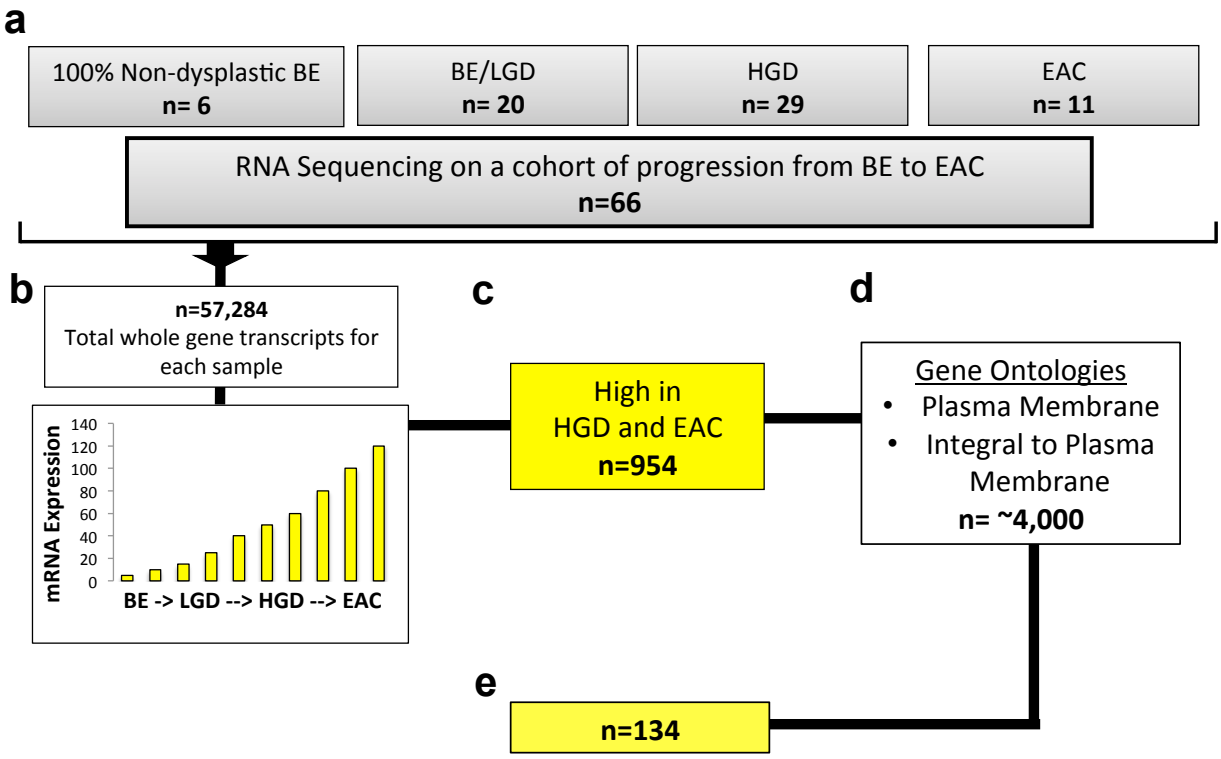



**Figure 3.8 Schematic of methods utilized to identify cell-surface over-expressed genes in HGD and EAC** (a) 66 total samples were used for RNA-sequencing (BE n=6, LGD n=20, HGD n=29, and EAC n=11). (b)(top panel) After RNA sequencing we obtained a total of 57,284 gene transcripts. We manually curated/correlated the transcripts showing an increase in mRNA during progression from BE to EAC (b: bottom panel). (c) We observed a total of 954 transcripts that correlated with progression (BE → EAC). (d) To identify the genes that code for cell surface protein, we use gene ontologies (GO) terms for plasma membrane and integral to plasma membrane. We identified 134 genes that are cell surface and that correlated with progression from BE → EAC.

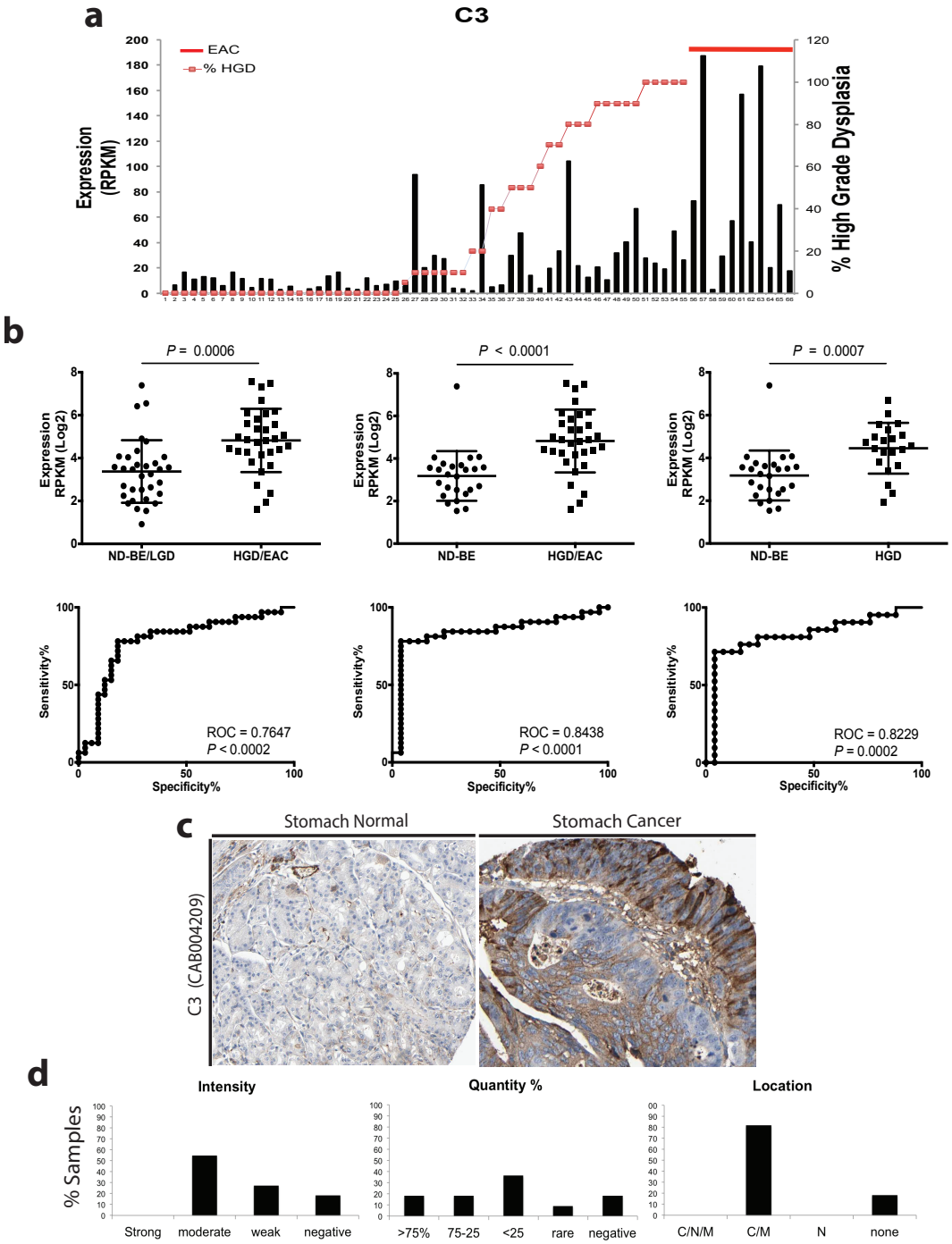



**Figure 3.9 *C3* expression** (a) in BE dysplastic progression to EAC (% of HGD displayed by the red dots, red line reflects 100% EAC samples). (b) Analysis of mRNA expression and sensitivity vs. specificity of *C3* between different stages of progression (ND-BE, LGD, HGD, and EAC). (c) Protein expression of C3 in normal and stomach (d) Quantification of number of samples for intensity, quantification, and location (C=cytoplasm, M=membranous, N=nuclear) of C3 protein in gastric/stomach cancers. (d and c data were obtain from the publicly available data domain The Human Protein Atlas[17,29])

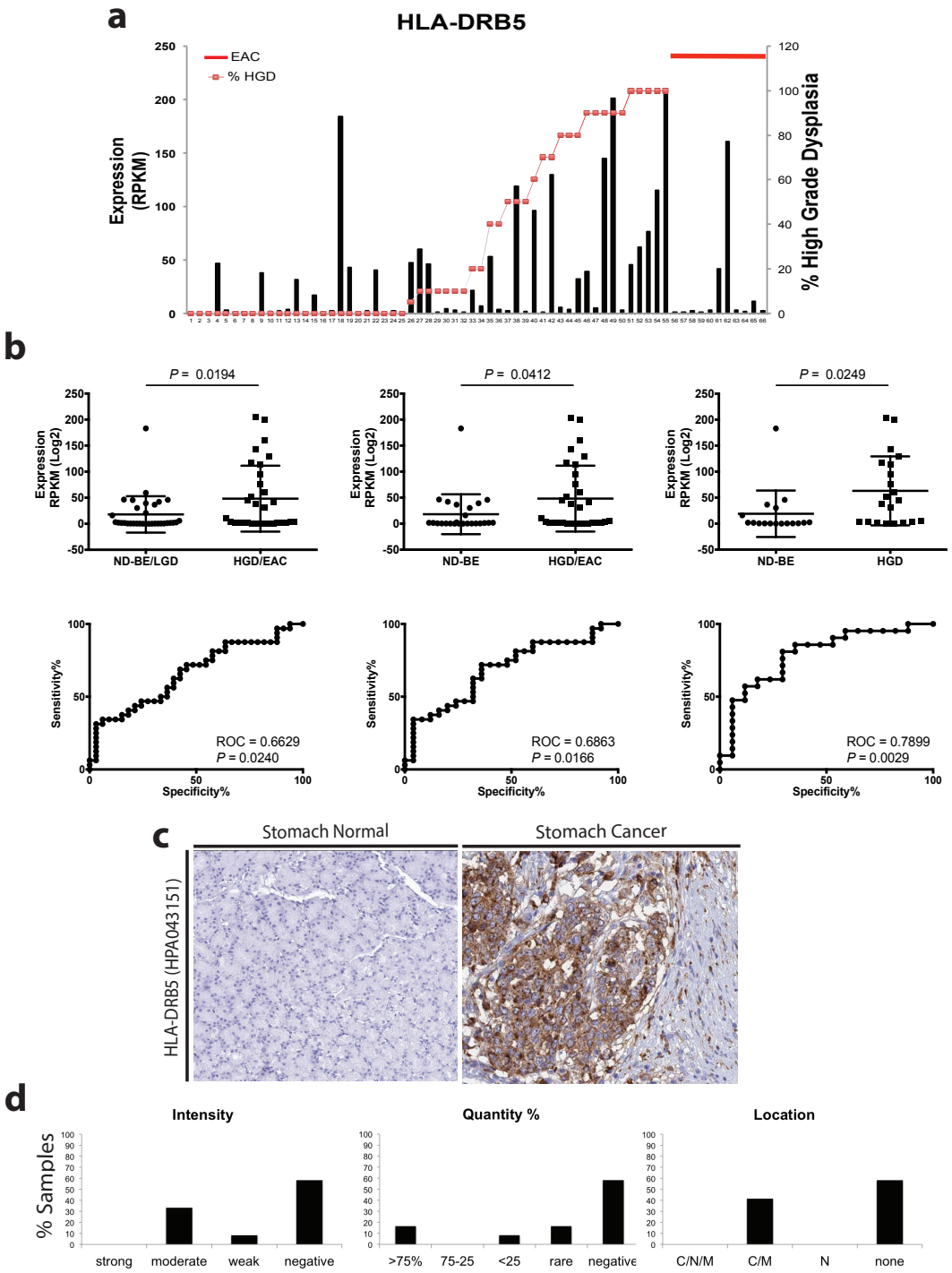



**Figure 3.10 *HLA-DRB5* expression** (a) in BE dysplastic progression to EAC (% of HGD displayed by the red dots, red line reflects 100% EAC samples). (b) Analysis of mRNA expression and sensitivity vs. specificity of *HLA-DRB5* between different stages of progression (ND-BE, LGD, HGD, and EAC). (c) Protein expression of HLA-DRB5 in normal and stomach (d) Quantification of number of samples for intensity, quantification, and location (C=cytoplasm, M=membranous, N=nuclear) of HLA-DRB5 protein in gastric/stomach cancers. (d and c data were obtain from the publicly available data domain The Human Protein Atlas[17,29])

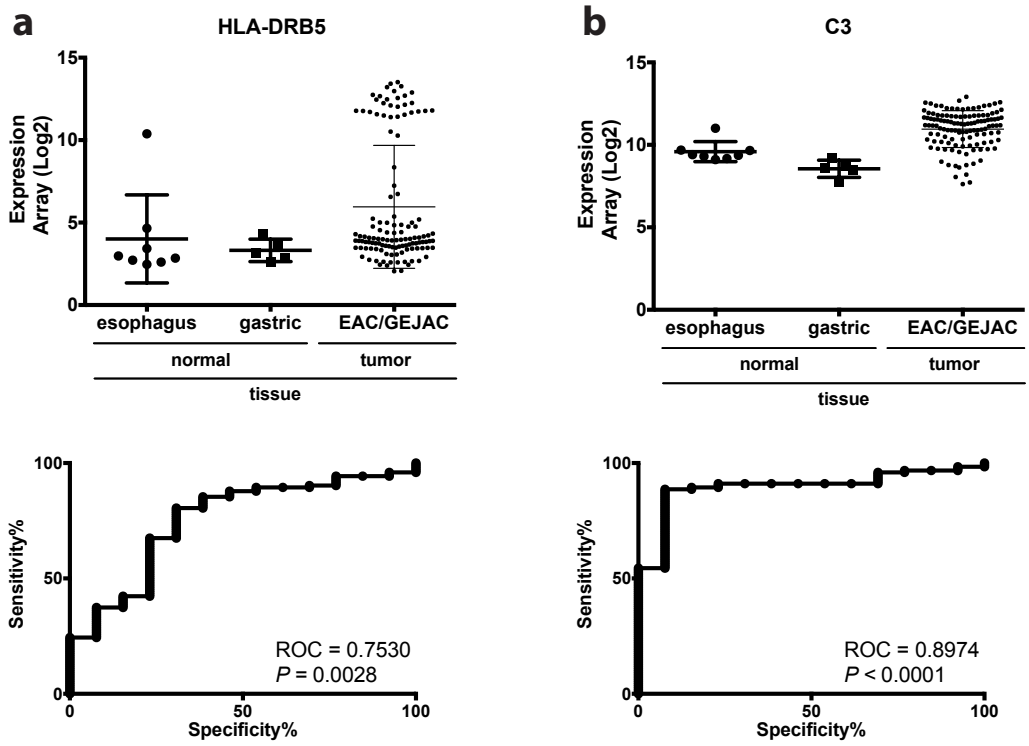



**Figure 3.11 *HLA-DRB5* and *C3* expression in normal (NE/NG) vs. tumor (EAC/GEJAC)** (a-b, top panel) *HLA-DRB5* and *C3* over-expression in EAC/GEJAC when compared to normal tissue (esophagus and gastric) ($P$ <0.0001). (a-b, bottom panel) ROC curves of *HLA-DRB5* and *C3* mRNA expression significantly distinguishes between normal tissues vs. EAC/GEJAC. (ROC= 0.7530, 0.8974; $P$=0.0028, $P$ < 0.0001, respectively).

## Tables

### Table 3.1 Spliceosome-associated gene expression

| Gene ID | Mean Barrett's | LGD | HGD | EAC | Barrett's +LGD | HGD+EAC | Fold Change Barrett's+LGD vs HGD+EAC | p-value |
|---|---|---|---|---|---|---|---|---|
| SNRPD2 | 164.5 | 173.0 | 224.5 | 288.5 | 166.6 | 246.5 | 79.90 | 8.59493E-05 |
| SNRPG | 133.6 | 128.1 | 156.9 | 199.8 | 132.3 | 171.7 | 79.90 | 3.18848E-06 |
| HNRNPC | 285.4 | 310.7 | 346.9 | 355.3 | 291.5 | 349.8 | 58.27 | 0.002201989 |
| SNRPB | 97.0 | 85.6 | 111.3 | 160.0 | 94.2 | 128.0 | 33.81 | 0.00196389 |
| HNRNPA3 | 249.2 | 270.9 | 280.1 | 297.5 | 254.5 | 286.1 | 31.63 | 0.005082472 |
| HSPA1A | 19.1 | 24.5 | 44.4 | 58.3 | 20.4 | 49.2 | 28.82 | 0.032288517 |
| SNRPD1 | 43.8 | 43.7 | 67.6 | 68.3 | 43.8 | 67.8 | 24.10 | 0.080210686 |
| SNRPE | 60.6 | 61.1 | 75.8 | 99.6 | 60.7 | 83.9 | 23.24 | 0.000231132 |
| RBMX | 88.8 | 97.6 | 109.9 | 119.5 | 90.9 | 113.2 | 22.24 | 0.000176419 |
| SNRPC | 60.1 | 61.4 | 73.9 | 96.2 | 60.4 | 81.6 | 21.21 | 2.61935E-05 |
| TRA2A | 80.2 | 94.2 | 105.3 | 102.7 | 83.6 | 104.4 | 20.79 | 0.003912889 |
| DHX15 | 84.8 | 88.5 | 108.2 | 92.4 | 85.7 | 102.8 | 17.04 | 0.004131615 |
| SF3B14 | 94.4 | 89.3 | 104.1 | 121.9 | 93.2 | 110.2 | 17.00 | 0.041963148 |
| LSM5 | 34.7 | 35.6 | 44.6 | 61.4 | 34.9 | 50.3 | 15.45 | 0.002592769 |
| NHP2L1 | 55.7 | 58.3 | 72.4 | 67.7 | 56.3 | 70.8 | 14.47 | 0.000177597 |
| PPIH | 28.7 | 28.6 | 35.9 | 51.0 | 28.7 | 41.1 | 12.38 | 0.007739117 |
| PRPF19 | 33.3 | 32.9 | 41.6 | 50.2 | 33.2 | 44.5 | 11.30 | 0.000831764 |
| LSM7 | 80.7 | 70.5 | 84.5 | 98.1 | 78.2 | 89.2 | 11.01 | 0.215138636 |
| SF3A3 | 25.5 | 30.0 | 35.3 | 41.1 | 26.6 | 37.3 | 10.72 | 0.001274657 |
| EFTUD2 | 30.9 | 32.5 | 42.0 | 42.0 | 31.3 | 42.0 | 10.68 | 5.2821E-06 |
| SNRPA1 | 32.6 | 32.6 | 40.7 | 48.0 | 32.6 | 43.2 | 10.63 | 0.001017551 |
| SF3B3 | 22.6 | 25.1 | 32.1 | 35.8 | 23.2 | 33.4 | 10.16 | 2.20002E-05 |
| SNRNP200 | 36.5 | 38.7 | 43.9 | 51.4 | 37.1 | 46.5 | 9.40 | 0.000299133 |
| SNRPD3 | 71.5 | 74.2 | 77.2 | 87.6 | 72.1 | 80.8 | 8.64 | 0.038522638 |
| NCBP2 | 39.3 | 42.8 | 46.9 | 51.8 | 40.2 | 48.6 | 8.44 | 0.001221452 |
| PPIE | 22.3 | 24.7 | 28.1 | 37.4 | 22.9 | 31.3 | 8.43 | 0.051336455 |
| HSPA6 | 1.6 | 1.8 | 10.4 | 9.2 | 1.7 | 10.0 | 8.33 | 0.168730252 |
| SNRPB2 | 32.0 | 32.9 | 37.9 | 45.1 | 32.2 | 40.4 | 8.19 | 0.000374409 |
| SNRPF | 30.7 | 29.9 | 33.1 | 48.9 | 30.5 | 38.6 | 8.07 | 0.018430127 |
| PRPF3 | 31.2 | 39.8 | 42.5 | 38.3 | 33.3 | 41.1 | 7.78 | 0.001101099 |
| LSM2 | 18.2 | 17.4 | 22.3 | 31.8 | 18.0 | 25.6 | 7.55 | 0.002626582 |
| PRPF31 | 22.3 | 24.5 | 28.4 | 33.2 | 22.8 | 30.0 | 7.19 | 0.000187034 |
| EIF4A3 | 31.7 | 31.4 | 36.8 | 42.6 | 31.6 | 38.8 | 7.19 | 0.012506196 |
| THOC1 | 18.6 | 24.9 | 25.9 | 29.3 | 20.1 | 27.1 | 6.98 | 0.016582256 |
| SNRPA | 18.4 | 19.0 | 22.4 | 29.5 | 18.5 | 24.8 | 6.32 | 0.000972627 |

## References


1. Siewert, J. R., Feith, M. & Stein, H. J. Biologic and clinical variations of adenocarcinoma at the esophago-gastric junction: relevance of a topographic-anatomic subclassification. *J. Surg. Oncol.* **90,** 139–46; discussion 146 (2005).
2. Shah, A. K., Saunders, N. a, Barbour, A. P. & Hill, M. M. Early diagnostic biomarkers for esophageal adenocarcinoma--the current state of play. *Cancer Epidemiol. Biomarkers Prev.* **22,** 1185–209 (2013).
3. Rev, E., Hepatol, G. & Tannapfel, A. Barrett ' s esophagus : can biomarkers predict progression to malignancy? 653–663 (2008).
4. Spechler, S. J., Sharma, P., Souza, R. F., Inadomi, J. M. & Shaheen, N. J. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* **140,** 1084–91 (2011).
5. Playford, R. J. New British Society of Gastroenterology (BSG) guidelines for the diagnosis and management of Barrett's oesophagus. *Gut* **55,** 442 (2006).
7. Kubo, A., Block, G., Quesenberry, C. P., Buffler, P. & Corley, D. a. Effects of dietary fiber, fats, and meat intakes on the risk of Barrett's esophagus. *Nutr. Cancer* **61,** 607–16 (2009).
8. Sturm, M., Joshi, B. & Lu, S. Targeted imaging of esophageal neoplasia with a fluorescently labeled peptide: first-in-human results. *Sci. Transl. Med* **5,** 184ra61 (2013).
9. Levin, J. Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7,** 709–715 (2010).
10. Tresini, M. *et al.* The core spliceosome as target and effector of non-canonical ATM signalling. *Nature* **523,** 53–58 (2015).
11. Spechler, S. J. & Souza, R. F. Barrett's Esophagus. *N. Engl. J. Med.* **371,** 836–845 (2014).
12. Bresalier, R. S. Barrett's Esophagus and Esophageal Adenocarcinoma. *Annu. Rev. Med.* **60,** 221–231 (2009).
13. Sarr, M. G. Barrett ' s Esophagus : Its Prevalence and Association With Adenocarcinoma In Patients With Symptoms of Gastroesophageal Reflux. (1984).
14. Bull, L. M., White, D. L., Bray, M., Nurgalieva, Z. & El-Serag, H. B. Phase I and II enzyme polymorphisms as risk factors for Barrett's esophagus and esophageal adenocarcinoma: A systematic review and meta-analysis. *Dis. Esophagus* **22,** 571–587 (2009).
15. Shaheen, N. J. & Richter, J. E. Barrett's oesophagus. *Lancet* **373,** 850–61 (2009).
16. Ferrer-Torres, D. *et al.* Genomic similarity between gastroesophageal junction and esophageal Barrett's adenocarcinomas. *Oncotarget* **5,** (2016).
17. Uhlen, M. A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Mol. Cell. Proteomics* **4,** 1920–1932 (2005).
18. Cancer, T. & Atlas, G. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499,** 43–9 (2013).

19. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).

20. TCGA. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541** (7636): 139-175 (2017).

21. Weaver, J. M. J. *et al.* Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat. Genet.* **46,** 837–843 (2014).

22. Secrier, M. *et al.* Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat. Genet.* **48,** 1131–1141 (2016).

23. Dulak, A. M. *et al.* Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res.* **72,** 4383–93 (2012).

24. Quidville, V. *et al.* Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. *Cancer Res.* **73,** 2247–2258 (2013).

25. Dar, M. S., Goldblum, J. R., Rice, T. W. & Falk, G. W. Can extent of high grade dysplasia in Barrett's oesophagus predict the presence of adenocarcinoma at oesophagectomy? *Gut,* **52**, 486–489 (2003).

26. Pohl, H. *et al.* Risk factors in the development of esophageal adenocarcinoma. *Am. J. Gastroenterol.* **108,** 200–7 (2013).

27. Varghese, S. *et al.* Analysis of Dysplasia in Patients With Barrett's Esophagus Based on Expression Pattern of 90 Genes. *Gastroenterology* **149,** 1511–1518.e5 (2015).

28. Li, J. & Jiang, H. Robust estimation of isoform expression with RNA-Seq data. *Statistics,* 1–16 (2014).

29. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science (80).* **347,** 1260419–1260419 (2015).

# CHAPTER 4

## Esophageal GSTT2 and Racial Differences in Esophageal Adenocarcinoma


### Summary

Esophageal adenocarcinoma primarily affects Caucasians (Cau) but not African Americans (AA), despite both having a similar incidence of risk factors. Therefore, *we hypothesize that the normal esophageal squamous mucosa in AA has protective mechanisms against GERD-induced damage that differ from Cau*. We used Affymetrix ST 2.1 expression arrays to determine the transcriptional differences between 40 normal AA and CAU esophageal squamous mucosa (NE) tissues (24 with no disease and 16 presenting with a history of GERD, Barrett's esophagus (BE) and/or EAC). Despite minimal differences between the non-diseased NE expression profiles, we identified that a detoxifying enzyme, *GSTT2*, is significantly over-expressed in AA compared to Cau. Two genomic events, a 30kB deletion and a 17bp promoter duplication in the *GSTT2* locus, previously associated with lower expression of *GSTT2*, were also found to correlate with lower *GSTT2* expression in EA compared to AA. These observations were validated using matched genomic sequence and expression data in lymphoblasts from the 1000 Genomes Project. In addition, we identify that the 17 bp promoter non-duplication is highly conserved in African and African decent populations. Furthermore, we demonstrate that GSTT2 is important in the protection of esophagus squamous cells against DNA damage when exposed to genotoxic stress from cumene-hydroperoxide (cum-OOH). Together, these observations

suggest that increased *GSTT2* expression may protect against damage caused by GERD and underlie the low incidence of EAC in AA.

**Introduction**

EAC is known to arise from a metaplasia tissue known as Barrett's esophagus (BE), a pre-disposing condition where the normal squamous of the esophagus (NE) is replaced by columnar intestinal type tissue. It has been extensively observed that EAC primarily affects Caucasians (Cau) but not African Americans (AA), despite having a similar incidence of the key risk factors: obesity, gastroesophageal reflux disease (GERD), and smoking[1,2]. More importantly, it is known that the primary risks factors including GERD, and obesity,[3–6] do not differ between the two ethnic groups[7–9]. Nevertheless, AA show less frequent erosive esophagitis than Cau (24% vs 50%, $P$ =0.03)[9] in addition to a much lower incidence of EAC in this population (AA). There is a lack of understanding as to why the AA population has less erosive esophagitis and therefore a lower incidence of developing EAC.

Transcriptional profiling is the measurement of the entire RNA transcript in a given cell or tissue that is encoded by DNA. We sought to understand the transcriptional (RNA) profile of the normal squamous mucosa of individuals of both populations (Cau vs. AA). Further, we included populations of individuals: disease free (no-history of GERD, BE, and EAC), and diseased individuals (long history of GERD, the presence of BE and development of EAC). Our goal was to utilize transcriptional profiling to identify key differences and pathways that might contribute to the protection of the esophagus in AA that the esophagus of Cau may lack.

Here we report that a detoxifying enzyme responsible for inactivating reactive oxygen species and reducing DNA damage[10], glutathione-s-transferase theta 2 *(GSTT2,* GSTT2*)*, is over expressed in the esophagus of AA relative to Cau. Moreover, we report the strong association of

two genomic events in the *GSTT2* locus (a 30kb deletion[11] and a 17bp promoter duplication[12]), that negatively affect *GSTT2* mRNA expression, are highly frequent in the Cau population. The promoter non-duplication is associated with high levels of *GSTT2* mRNA and is highly conserved in African and African descendant populations. In addition, we show that reduction of GSTT2 in an esophageal squamous cell line makes the cells more susceptible to DNA damage under genotoxic stress. Together, these observations suggest that increased *GSTT2* expression may protect against damage caused by GERD and underlie the low incidence of EAC in AA.

## Results

### Transcriptional Profile of Normal Squamous Mucosa in AA vs. Cau

To determine whether the NE of AA demonstrates a different transcription profile from Cau we used Affymetrix ST 2.1 expression profiling analysis. We obtained two specific cohorts of NE from AA and Cau. One included healthy individuals (no history of BE or EAC) from both ethnic groups (Cau-NE n=12; AA-NE n=12) and the other included the NE of diseased patients with a history of GERD and presence of BE and/or EAC (Cau-NE:B n=8; AA-NE:B n=8)(**Table 4.1**). We performed analysis of variance (ANOVA) across all groups. When comparing NE profiles of AA to Cau only eight genes showed >2-fold change (FC) ($P$<0.01; **Table 4.2**), suggesting minimal differences in the expression profiles between these two cohorts. The *GSTT2* (mRNA) demonstrated the largest FC (AA-NE vs. EA-NE; $P$=0.0004, FC=5.1) between NE samples of the two ethnic groups, while *GSTT2B* transcripts demonstrated the fourth largest FC **(Figure 4.1a; Table 4.2**). Since the Affymetrix probesets do not adequately distinguish between *GSTT2* and *GSTT2B* transcripts, we therefore present mRNA expression as *GSTT2/*2B (detailed in Methods). We observed a similar, though non-significant ($P$=0.07) trend of increase *GSTT2/*2B expression in AA-NE:B compared to Cau-NE:B (**Table 4.2)**. Together these

observations suggest that *GSTT2/2B* expression is lower in the NE of the normal Cau population as well as in individuals who developed BE and/or EAC. We then validated the array results using an extended cohort of non-diseased squamous samples (AA-NE n=21 vs. Cau-NE n=21) with reverse transcription followed by quantitative real-time polymerase chain reaction (qRT-PCR). *GSTT2/2B* was overexpressed in the NE of AA relative to EA (FC=1.4, $P$=0.0014; **Figure 4.2a**) and correlated with the array data (r=0.80, $P$<0.0001; **Figure 4.2b**). Phase II detoxification enzymes, including *GSTT2*, are responsible for conjugating activated endogenous compounds to large anionic groups to facilitate active transport and excretion from the cell[13–18]. *GSTT2* has an active role in protecting cells against the effects of reactive oxygen species (ROS) and thus reducing DNA damage[14–18]. Given its role in detoxification and our observations of increased *GSTT2/2B* expression in the AA population, we hypothesize that enhanced protection against GERD-related DNA damage may underlie racial differences in EAC incidence.

**Genomic events associated with mRNA levels of GSTT2.**

To address further the basis for the differential mRNA expression of *GSTT2/2B* in AA vs CAU we examined the *GSTT2* chromosomal region. Resulting from an inverted chromosomal segmentation duplication, both genes (*GSTT2* and *GSTT2B*) are located in chromosome 22q11.23, with *DDT* and its inverted homolog, DDTL, located between them[24] (**Figure 4.1b)**. *GSTT1*, the only other Theta class GST present in humans, shares 55% protein homology and is located telomeric to both *GSTT2* and *GSTT2B*[24] (**Figure 4.1b**). Strongly differential expression of *GSTT2* and *GSTT2B* ($P$<0.001) (**Figure 4.1a),** but not *DDT* (FC=0.85, $P$=0.013), *GSTTP1* (FC=0.80, $P$=0.16), or *GSTT1* (FC=0.70, $P$=0.22), was observed between AA-NE vs. CAU-NE (**Figure 4.1c-e**). This suggests the increased expression of *GSTT2/2B* in AA is not due to regional DNA amplification and/or chromosomal translocation, but is *GSTT2* and *GSTT2B*

specific. We further assessed other genomic events in the chromosome 22q11.23 locus that may affect regulation of *GSTT2/2B* expression. Zhao et al. characterized a 37kb deletion of *GSTT2B* (**Figure 4.3a; Figure 4.5a**), which is associated with lower levels of *GSTT2* expression[11]. We genotyped both AA and Cau normal squamous cohort samples using the method of Zhao et al.[11] and observed a slightly higher frequency (Fisher Exact *P*=0.03) of the *GSTT2B* deletion in Cau NE (80.6%) relative to AA NE (54.8%; **Figure 4.5b**). Lower *GSTT2* expression was observed among individuals harboring homozygous deletions of this gene, though not significantly different from those homozygously non-deleted (*P*=0.10; **Figure 4.5c**). Indeed using qRT-PCR to examine the expression pattern in combination with GSTT2B deletion genotype using extended cohorts of NE from AA and Cau individuals show that deletion of *GSTT2B* cannot explain the expression differences between the two populations (**Figure 4.5d***)*.

In a study examining tandem duplications across the genome, Marotta *et al*. identified a 17bp tandem duplication within the promoter of *GSTT2* and *GSTT2B*[12] (**Figure 4.3a**). Using a luciferase-linked assay to assess promoter function they showed the presence of the 17bp tandem duplication associates with lower *GSTT2* and *GSTT2B* promoter activity[12]. Utilizing a PCR-based methodology to differentiate the 17bp promoter duplication, we genotyped gDNA from the Cau and AA samples. We observed a significantly higher frequency (Cau NE vs. AA NE: 97.5% to 67.5%, *P*<0.00067) of the promoter duplication in Cau relative to AA (**Figure 4.3b; Figure 4.4a-b**). We then examined matched lymphoblast genotype and mRNA expression data available for a subset of the 1,000 Genomes Project samples[25] and confirmed those individuals of African descent have higher expression of *GSTT2* mRNA than Cau/European populations (*P*=0.030) (**Figure 4.6a**). Furthermore, we confirmed that *GSTT2* promoter duplication levels correlated significantly with lower levels of *GSTT2* expression in both our esophageal cohort

(*P*=0.0058; **Figure 4.3c**) and in the 1000 Genomes cohort (*P*=1.5x10$^{-6}$; **Figure 4.6b**). The *GSTT2B* deletion also strongly correlated with *GSTT2* expression in the 1000 Genomes data (**Figure 4.6c**), however, the most important contribution to low *GSTT2/2B* expression occurs when the 17bp *GSTT2/2B* promoter duplication is homozygous (**Figure 4.6d**) and thus similar to the results seen in the esophagus (**Figure 4.3d**). In both sample sets the 17bp promoter genotype status provided a much better explanation for *GSTT2/2B* expression variance than ethnicity. These results suggest the presence of the promoter duplication is more frequent in Cau than AA and may underlie reduced esophageal *GSTT2/2B* expression in this population. In terms of potential confounding effects, we did not observe any significant differences in *GSTT2/2B* mRNA levels between males and females (**Figure 4.7a, b**) or in relation to the presence of GERD or smoking status (**Figure 4.7c, d**). It is of interest to note that El-Serag *et al*.[9] examined variation in the prevalence of GERD within US populations and reported no difference in the frequency of GERD symptoms (weekly heartburn and/or regurgitation) between different racial groups (AA 29%, Cau 28%, other 25%, *P*=0.80). However, AA participants had a much lower risk of esophagitis (adjusted OR 0.22–0.46, *P*<0.001)[9], which translates into reduced tissue damage for comparable reflux exposure.

**Validation of genomic events associated with *GSTT2* expression using 1000 genomes data.**

We used 1000 Genomes data to examine the distribution of the 17bp *GSTT2/2B* promoter duplication status across world populations. The 17bp *GSTT2/2B* promoter duplication is prevalent in the other super-populations (AMR=94%, EUR=95%, SAS=95% and EAS=97%) when compared to AFR (AFR=73%) (chi$^2$ *P*<0.01; **Figure 4.8a**). In particular West African populations showed the highest frequency, with a >30% compared to other populations (which in

general had average 17bp *GSTT2/2B* promoter duplication frequencies of 1-5%) (**Figure 4.8b**; **Table 4.3**).

**GSTT2 functions to prevent DNA damage in cells undergoing genotoxic stress.**

In the esophageal field, there is a lack of normal esophagus cell lines, with only one immortalized esophagus squamous cell line known as HET1A. Therefore, we utilized additional cell lines of squamous lineage to assess *GSTT2* genotype, mRNA, and protein levels. To identify the effect of the number of *GSTT2* gene copies and the presence of *GSTT2* promoter duplications we genotyped a cohort of cell lines including the normal squamous esophagus (HET1A), and a variety of different squamous cancers (head and neck, and cervical cancer) as well as adenocarcinomas (**Figure 4.9a**).

We observed that the normal esophagus cell line (HET1A) is homozygously deleted for the 37KB fragment, indicating that it only harbors 2 copies of the *GSTT2* gene (**Figure 4.9a,** top panel). On the other hand, HeLa cells are homozygously non-deleted, and therefore harbors 4 copies of *GSTT2* (**Figure 4.9a** top panel). We further assessed whether these gene copies of *GSTT2* harbor the promoter duplication. HET1A is homozygous and HeLa is heterozygous for the promoter duplication, respectively (**Figure 4.9a** bottom panel). Although both cells lines were immortalized from African Americans, based on our current observations, HET1A represents the genotype most frequently observed in the Cau population whereas HeLa cells represent a predominantly AA genotype. We further evaluated both the mRNA and protein levels of *GSTT2*/GSTT2 in HeLa and HET1A. As expected, we observed that HeLa cells have higher levels of endogenous GSTT2 (both at the mRNA and protein levels), with over 2-fold higher levels than HET1A cells (**Figure 4.9b-c**). GSTT2 has been previously shown to protect cells against DNA damage when exposed to the oxidative agent, cumene-hydroperoxide (cum-

OOH)[20]. Based on these observations and the endogenous levels of GSTT2 in the cell lines, we hypothesized that HET1A cells will be more susceptible to DNA damage when exposed to cum-$H_2O_2$ than HeLa cells. DNA damage was measured using immunofluorescence staining of DNA double-strand break marker, gamma-H2AX (γ-H2AX) and quantification of nuclear foci (Methods). First, we determined the efficiency of cum-OOH in producing DNA damage and the cellular toxicity threshold, by performing increasing dose gradient (0, 18, 25, 75, 100, 150μM) and time course (30min, 1hr, 2hr) experiments. We observed >50% DNA damage formation (cells with >10 γH2AX foci) in HET1A cells treated with 100μM cum-OOH for 1hr and without cell death. We then treated HET1A and HeLa cells with 100μM cum-OOH for 1hr and stained for γH2AX and DAPI (nuclei stain). Under this genotoxic stress using cum-OOH treatment, HET1A cells showed >20% more DNA damage than HeLa cells (**Figure 4.9d-c**). Altogether we identified that the *GSTT2* genotype associates with the level of *GSTT2*/GSTT2 mRNA and protein in cell lines, and cells with higher expression of GSTT2 sustain less DNA damage under similar genotoxic stress conditions when compared to low GSTT2 expressing cells (**Figure 4.9d-c**).

We compared two cell lines with different endogenous levels of GSTT2 and found that HET1A is more susceptible to DNA damage from genotoxic stress than HeLa. It is important to highlight that this comparison is made between a normal cell line and a cancer cell line, therefore, the lower DNA damage observed in HeLa may be influenced by cancer-related factors (more resistant to treatments), than GSTT2 specific factors. To elucidate if GSTT2 plays a role in protecting cells against genotoxic stress in HET1A and HeLa, we performed knockdown experiments in both cell lines with and without treatment of cum-OOH. We transfected HET1A and HeLa with a non-targeting (NT) siRNA and four *GSTT2* target siRNAs. We confirmed

successful mRNA and protein knockdown of >50-80% in both cell lines (**Figure 4.10 a-e**). We observed that after knockdown of GSTT2, using four different GSTT2 siRNAs, treated cells have twice the amount of DNA damage when compared to controls (control, NT). Even further, after the knockdown, HeLa cells sustain DNA damage, comparable to HET1A. This data suggests that GSTT2 plays an important role in protecting cells against genotoxic stress and that the relative levels of GSTT2 associates with the amount of DNA damage the cells sustain when exposed to genotoxic stress.

We further assessed whether increasing the levels of GSTT2 could make cells less susceptible to DNA damage. We transfected HET1A cells with a GSTT2-DDK over-expression construct for 24hrs and observed significant over-expression at both the mRNA (**Figure 4.13a**) and protein level (**Figure 4.13b**, top band). After 24hrs transfection we treated the cells with 0 and 100μM cum-OOH. We first observed that cum-OOH increases GSTT2 expression, both endogenously (**Figure 4.13c**) and after transfection (**Figure 4.13d**). We did not observe any significant difference in cellular damage with over-expression of GSTT2 but no treatment with cum-H2O2 (**Figure 4.13e**). This suggests that GSTT2 abundance does not impact or induce DNA damage in cells. Moreover, we observed that the cells with increased GSTT2 expression and receiving cum-OOH treatment contained less DNA damage as compared to negative GSTT2 cells (**Figure 4.13f**). In conclusion, we observe that cells containing increased expression of GSTT2 and exposed to genotoxic stress and cells with high levels of GSTT2 are less susceptible to double strand breaks, than cells negative or low levels of GSTT2.

**GSTT2 protein in NE of AA vs. Cau.**

We obtained NE tissue from individuals that self-identify as either AA or Cau and in addition had GERD and no-GERD. We measured GSTT2 expression levels in the normal

squamous tissue (NE) using Western analysis. We observed that GSTT2 is over-expressed at a significantly higher frequency in the esophagus of AA (7/8) vs Cau (2/5). Quantification of GSTT2 levels normalized to α-tubulin revealed that AA esophagus has greater than 2 FC protein expression as compared to Cau (GERD AA vs Cau $P$=0.0003; GERD-AA vs Non-GERD Cau $P$=0.0002).(**Figure 4.14a-b**). This data supports the observation that GSTT2 is over-expressed, at the mRNA and protein levels, in the esophagus of AA as compared to Cau.

Using matched DNA extracted from the blood of the individuals evaluated by Western, we evaluated the genotype status for the 37kb deletion and 17bp promoter duplication in the *GSTT2/2B* locus (**Figure 4.14c**) and compared to the levels of GSTT2 (**Figure 4.14d-e**). Genotype for *GSTT2/2B* 37kb deletion (-/-)(-/+) and non-deleted (+/+) did not significantly associate with levels of GSTT2 protein (**Figure 4.14d**). In contrast, the *GSTT2/2B* for the promoter duplication (+/+)(-/+) and non-duplicated (-/-) promoter associated with levels of GSTT2 protein (**Figure 4.14e**). In the context of overall genotype for the promoter duplication vs. GSTT2 protein levels, we observed that the homozygous-duplicated status correlates with both, individuals with high and low GSTT2 expression. Nevertheless, when we divide the comparison by the GERD vs. no-GERD, individuals with no-GERD, those with the 17bp promoter duplication have less endogenous expression than individuals with no-GERD no promoter duplication. In the context of GERD, the genotype does not associate distinctly with the levels of GSTT2.

**Esophagus cells treatment with non-steroidal anti-inflammatory (NSAID)**

**Indomethacin.**

NSAIDs have been found to increase the levels of GSTT2 in the esophagus of rats[26]. Therefore, we wanted to test whether using indomethacin, one type of NSAID, had an effect on

the levels of GSTT2. First, we assessed the levels by mRNA. We treated HET1A cells with indomethacin, with cells given media containing either 300μM or 600μM indomethacin as compared to DMSO treated (control). We analyzed the effects of indomethacin in HET1A expression (mRNA) at 6, 24, 48, and 72hrs. We observed at the mRNA level, that indomethacin induced the levels of *GSTT2/2B* by up to 15-fold (600μM; 72hrs) (**Figure 4.15a**). Gene expression can be regulated by different mechanisms including epigenetic regulation. Therefore, we wanted to assess whether epigenetic activation by tri-methylation of histone 3 at lysine 4 (H3K4me3) was observed in indomethacin treated HET1A cells. Acetylated or methylated H3K4 is a marker for transcriptionally active chromatin[27], and tri-methylation of lysine residue 27 on histone 3 (H3K27me3) is a repressive/silent state[28,29]. We performed chromatin immunoprecipitation (ChIP) assays in HET1A cells treated with 600μM of indomethacin (48hrs) or in non-treated cells and measured DNA pull down for three different regions of the *GSTT2/2B* promoter. We observed that after 48hrs of treatment, HET1A shows an enrichment of H3K4me3 when compared to non-treated controls ($P$=0.01) (**Figure 4.15c**). In addition, we observed an enrichment of RNA Pol II at the same sites after indomethacin treatment ($P$=0.0004) (**Figure 4.15c**). In contrast, we do not observe a significant enrichment for the repressive mark H3K27me3 ($P$=0.7339) (**Figure 4.15c**). The enrichment of H3K3me3 together with RNA Pol II correlates with our previous observation of increased expression of *GSTT2/2B* after treatment with indomethacin (**Figure 4.15a**). Together this data suggests that indomethacin is inducing the expression of *GSTT2/2B*, and epigenetic regulation is playing a role in this activation. Nevertheless, we observed that at the protein level, GSTT2 levels did not significantly change following indomethacin treatments (**Figure 4.15b**). This may suggest more potent inducers will be needed to be tested in the future, to increase levels of GSTT2 in HET1A cell lines. Because

GSTT2 is a very stable protein it may be that longer incubation times are needed before increased protein is detected. It is important to note, that significant changes in GSTT2 levels were reported by Van Lieshout *et al*. when rats were treated with different NSAIDs[26], suggesting that the potential environment and model for induction of GSTT2 may be important.

## Discussion

Obesity is a current epidemic in most developed countries. We have observed that the incidence of EAC has skyrocketed in the past three decades, and it is more common among the Cau than AA. *El-Serag et al.* and others have reported no difference in the frequency of GERD symptoms (weekly heartburn and/or regurgitation) as well as other risk factors including obesity, smoking, or age between both ethnic groups[7–9]. Our current findings provide the first evidence for a genetic and a molecular basis to explain this disparity between the two ethnic populations.

We observed a higher frequency of the 17bp *GSTT2/2B* promoter duplication in Cau as compared to AA and that this is associated with lower levels of NE *GSTT2/2B* expression. Interestingly, lower levels of the related family member *GSTT1* has been associated with increased esophageal squamous cancer risk in Asian populations[33]. GSTT2 functions to protect cells against genotoxic stress, and in colon cell studies[10], it has been observed that having high or low levels of GSTT2 contributes to a cell's susceptibility to DNA damage. Here we report that GSTT2 functions in the same manner in the esophagus squamous cell line (HET1A). We observed that after knockdown of GSTT2, cells are more susceptible to DNA damage after genotoxic stress. Further, we observed that the endogenous levels of GSTT2 (high GSTT2 cell line vs. low GSTT2 cell line) influence a cell's susceptibility to DNA damage under genotoxic stress. These observations, suggest that an individual with low levels of GSTT2 and exposed to constant genotoxic stress (for example GERD), may undergo more DNA damage than individuals with high levels of GSTT2.

Our observation, using 1000 genomes and HapMap data, that the non-duplicated promoter is mainly conserved in populations of African descent, suggests a change in selective pressure in favor of the duplication when populations migrated from Africa. Our analysis indicates that the *GSTT2/2B* promoter duplication is observed in the Denisovan genome[32] suggesting an ancient derivation that appeared in populations leaving Africa and is now distributed worldwide. The higher frequency of promoter duplication might be explained by dietary changes necessitated as a consequence of the migration into new ecosystems. This change may not have had significant detrimental health consequences until populations more recently experienced higher obesity levels and the associated rise in GERD incidence, both key risk factors for EAC[31]. The prevalence of GERD in East Asia is estimated to be 2.5-7.8% while populations in the US and Europe, with higher EAC risk, have more frequent GERD (18-28%)[31]. Obesity levels follow a similar pattern, with higher rates among high-risk EAC populations (15-25%) when compared to lower risk populations in Asia (5-15%)[32]. These observations suggest a potential interaction between lower *GSTT2/2B* and obesity that may combine to influence an individual's risk for development of the BE and possibly EAC. Thus among non-African populations (higher at-risk *GSTT2/2B* duplicated promoter), obesity rates may offer at least a partial explanation for the difference in EAC incidences. If true, then the current trend of increasing obesity levels within these populations should be viewed with concern in terms of esophageal adenocarcinoma cancer risk.

Non-steroidal anti-inflammatory (NSAIDs) have been shown to reduce the risk of EAC development by >40-50%[33]. In addition, NSAIDs and apple polyphenols have been observed to induce the levels of GSTT2[10,20,26]. Although it has been suggested that the effect of NSAIDs in reducing the risk of EAC is through inhibition of inflammation, we propose the idea that

NSAIDs also lower an individual's risk to GERD-related damage, through increasing GSTT2 expression as well as inhibiting inflammation. Further studies looking at the effect of NSAIDs in inducing GSTT2 levels in the esophagus are needed to elucidate this hypothesis. Nevertheless, treatment with these or other compounds in high-risk populations with GERD may potentially increase protective *GSTT2/2B* and help reduce the increasing incidence of EAC.

In summary, we report a significant difference in the levels of a key-detoxifying enzyme, GSTT2, which is differentially expressed between Cau and AA populations. In addition, two genetic variants (37kb deletion and 17bp promoter duplication) negatively associate with the levels of *GSTT2* and are highly frequent in the Cau population, but not GSTT2 protein levels in the contact of GERD. Limitations of this study are the low number of non-GERD samples. Although, we have observed that GSTT2 increases when cells are exposed to genotoxic stress (**Figure 4.13c-d**), we cannot differentiate whether the levels of GSTT2 in the AA-GERD group are the endogenous, or induced levels due to the presence of GERD. Comparing the Cau and AA suggests that there is the potential for this to be an induced state of GSTT2, nevertheless, further studies looking at non-GERD individuals from AA and Cau populations will be needed.

In conclusion, esophageal squamous cells with low levels of GSTT2 are more susceptible to DNA damage following treatments that induce genotoxic stress. Moreover, given the potential induction of GSTT2 through treatments with NSAIDs and apple polyphenols, this suggests a potential preventive approach for individuals at greatest risk for the disease. Altogether, these observations suggest that increased *GSTT2/2B* expression may be protective against esophageal damage caused by GERD and might underlie the low incidence of EAC in AA populations.

**Materials and Methods**

**Biopsies of the squamous esophagus.**

Histologically normal biopsies of the esophageal squamous epithelium were collected from consented men of either AA or Cau ethnicity who underwent a research upper endoscopy between 2008 and 2016 at the time of a scheduled colonoscopy performed for colorectal cancer screening at the University of Michigan Health System, University of North Carolina or Ann Arbor Veterans Affairs Medical Center and found to have no Barrett's esophagus or erosive esophagitis (population controls). Samples from the normal squamous epithelium (NE) of AA and Cau individuals who developed BE and/or EAC were obtained from Case Western University and Johns Hopkins University and University of Michigan Health System Samples were collected between 1991 and 2004 using protocols approved by from their respective institutional review boards (IRB). Racial group determination (AA or Cau) was based on the results of questionnaires completed by each recruited participant. For normal (NE) tissue from EAC patients, none have received preoperative radiation or chemotherapy. All specimens were collected fresh and frozen in liquid nitrogen and stored at -80 C until use.

**mRNA extraction and Affymetrix expression analysis.**

RNA was isolated from the normal esophageal squamous mucosa from individuals self-identified as either Cau or AA. mRNA extraction was performed using QIAzol Lysis Reagent (Qiagen, Valencia, CA) and purified with RNeasy spin columns (Qiagen), including on-column DNAse I incubation, according to the manufacturer's instructions. RNA samples with RIN scores greater than 6.0 (Agilent Bioanalyzer; Agilent Technologies, Palo Alto, CA) were submitted to the University of Michigan Sequencing Core for cDNA synthesis and cRNA amplification (Ambion WT Expression Kit; Life Technologies, Grand Island, NY), followed by hybridization to Human Gene ST 2.1 arrays (Affymetrix, Santa Clara, CA) according to the manufacturer's

instructions. Expression values for each gene were determined using the robust multi-array average (RMA) method[34] in the Bioconductor package[35] of the R statistical platform. Analyses were restricted to the 26,613 coding and non-coding genes for which annotation details were available. Analysis of variance was performed across all annotated transcripts**.** 12 AA without Barrett's (AA-NE), 8 AA with Barrett's (AA-NE:B), 12 Cau without Barrett's (Cau-NE), and 8 Cau with Barrett's (Cau-NE:B). *P* values were determined and fold-changes calculated between the four groups and compared between races (**Table 4.2**).

***GSTT2* and *GSTT2B* Probesets.**

Affymetrix probesets do not adequately distinguish between *GSTT2* and *GSTT2B* transcripts, therefore we present mRNA expression as *GSTT2/*2B and only report the more complete 16933088 probeset which covers all 5 exons, rather than 16928115 which only has probes on exons 2 and 4. The regulatory and coding sequences of RefSeq genes *GSTT2* (hg38:chr22:23980123-23983911, RefSeq: NM_000854) and *GSTT2B* (hg38: chr22:23957414-23961186, RefSeq: NM_001080843) are indistinguishable (99.9% reference sequence homology with the difference of a single SNP: rs74487784, which may or may not be unique to the *GSTT2B* form).

**qRT-PCR validation.**

Total RNA from the normal squamous tissue of the 12 AA and 12 Cau examined by ST 2.1 array, as well as additional normal squamous tissue from 8 AA and 8 Cau were then used for real-time (qRT-PCR) validation of selected gene transcripts. Isolated RNA was converted to cDNA using reverse-transcriptase following manufacturer's instructions (Life Technologies, Grand Island, NY). RT-PCR reactions were performed in a final volume of 20 µl containing 2 µl 10x RT buffer, 2 µl 10x RT random primers, 0.8 µl 25x dNTPs, 1 µl reverse transcriptase, 1 µl

RNase inhibitor, and 200 ng to 2 μg RNA, depending on RNA concentration after extraction. Following RT-PCR the reactions were then diluted to 3 ng/μl prior to qRT-PCR quantitation of selected gene transcripts. Primers for qRT-PCR reactions were designed using Primer-BLAST[36] (*GSTT2*: primers for cDNA Exons 1-2 5'-TGGGCCTAGAGCTGTTTCTT-3', 3'-CCAGGCTGTTGATCTGCAAG-5' and *GSTT1* primers for cDNA 5'-CTGGAGTTTGCTGACTCCCTC-3', 3'- GCTCGAAGGGAATGTCGTTCT-5') or previously published (*GAPDH* and/or *B-Actin*)[37]. Optimal annealing temperatures and reaction conditions were confirmed using the Cepheid SmartCycler (Cepheid). Samples were run in duplicate using the ABI PRISM® 7900HT Sequence Detection System according to the manufacturer's instructions. Thermal cycling conditions consisted of 10min of initial denaturation at 95°C, and 40 cycles of 15 sec of denaturation at 95°C, and 1min of annealing/extension at 60°C. Relative expression and fold-change of the genes of interest (*GSTT2* and *GSTT1*) were calculated by the ΔΔ-Ct method (subtracted the average of the Ct's for the reference gene GAPDH from the Ct values of each target gene)[38].

**DNA extraction.**

DNA was extracted from cell lines using Qiagen DNeasy Blood & Tissue Kit following the manufacturer's instructions. Genomic DNA was extracted from normal squamous esophageal biopsies by adding 800 μl of 100% ethanol to the QIAzol and then spun in a table top centrifuge at 4°C for 30min at 1200 rpm. Ten μl of proteinase K (10 mg/μl) and 25 μl SDS (10%) was added to each sample and incubated overnight at 56°C. Tris-extracted phenol (500 μl) was then added to each sample, gently rotated with inversion at room temperature for 3 hrs, spun at room temperature for 20min at 3500 rpm, and the aqueous layer transferred to new tubes. Ice-cold 100% ethanol (800 μl) and 2 μl of sterile 5 M NaCl were added to each tube to precipitate DNA.

Samples were centrifuged at 4°C at 12000 rpm for 15min, air-dried and resuspended in 50 µl TE buffer. Samples were then incubated overnight at 4°C and concentration and DNA purity determined by absorbance using a NanoDrop spectrophotometer.

***GSTT2B* deletion genotype analysis.**

The *GSTT2B* genotype was determined using a three-primer set for PCR as previously described (Zhao *et al*., 2009): *GSTT2B*-6858 5'-CACTCAACACAGTAGCCTCATCGTG-3', *GSTT2B*-6857 5'-TGCCTCCCCTGCCTTATTTC-3', and *GSTT2B-2B* 5'-CCTTCTGAAATGGAGCCTTTG-3'. The PCR reaction contained a total volume of 50 µl and consisted of 25 µl of Promega GoTaq Green Master Mix, 10 pmol of primers *GSTT2B*-6858 and *GSTT2B-2B*, 15 pmol of primer *GSTT2B*-6857, and 25 ng genomic DNA. The thermal cycling conditions for the reaction consisted of initial denaturation for 2min at 95°C, followed by 35 cycles of 30 sec of denaturation at 95°C, 30 sec of annealing at 60°C, and 45 sec of extension at 72°C. PCR products were then separated in 2% agarose gels with ethidium bromide by electrophoresis in 1X TBE buffer. Invitrogen's 100 bp DNA ladder was used as a marker.

***GSTT2/2B* promoter duplication analysis.**

Standard PCR using primers and the protocol from Marotta *et al* (2012) was used. PCR products were loaded on 2% agarose gel (Fisher) and run overnight (>18hrs) at 20 V to completely separate products. Two examples of each promoter genotype were chosen for sequencing based confirmation. PCR products were purified using Qiaquick (Qiagen) columns as per manufacturer's instructions and submitted to the University of Michigan Sequencing Core for Sanger sequencing using the reverse primer. Resulting sequence reads were aligned using Geneious software (version 5.4.6; Biomatters Ltd, Auckland, New Zealand) (**Figure 4.4a**) and manually (**Figure 4.4b**) to match 17bp genotypes to qPCR results.

**1000 Genomes and HapMap data analysis for *GSTT2B* and *GSTT2* promoter genotypes.**

The 17bp promoter duplication is not present in the 1000 Genomes set, primarily because of falling in a segmental duplication region but also due to the duplicative nature of the variant itself. Nevertheless, the duplication itself can be clearly seen from the raw alignment data examined manually. After manual alignment, the duplication appears as a deletion relative to the reference sequence (GRCh37). Subsequently, we could not use the 1000GP variant calls and therefore utilized a counting of reads containing either the duplicated allele or non-duplicated allele:

***Duplication:***

GTGCACGAAGTGGGAGCTCCCGCTGTCTGGCAGCTCCCGCTGTCTGGCAG

***Non-duplicated:***

GTGCACGAAGTGGGAGCTCCCGCTGTCTGGCAGCAGCTGCTCTGCAGGGG

We enumerated sequences that mapped around both paralogs, to take into account alignments that might align to one or another. We extracted RKPM expression data for 110 lymphoblastoid cell lines from the tabulated GEUVADIS RNA-seq data (Express Array ID: E-GEUV-1)[39]. These samples (39 AFR and 77 EUR) match those for whom raw copy number variation data were available. We then used the non-parametric Kruskal–Wallis statistic to investigate the frequency of *GSTT2* copy number variations and expression levels within these matched samples.

**Cell line and treatments cumene-hydroperoxide (cum-OOH).**

HET1A (an immortalized cell line from normal esophageal squamous mucosa) and HeLa (immortalized cell line from cervical cancer) were cultured in Dulbecco's Modified Eagle's

Medium (DMEM)(Cat. No. 11965092) with 10% fetal bovine serum and 1% antibiotic-antiseptic. Cells were plated in 35mm x 10mm polystyrene plates (Cat. No 430165)(Corning Incorporated, NY) with 4 coverslips (6mm) within the plate, at a starting number of 120,000 cells per plate. For knockdown of GSTT2, four siRNAs (siRNA-05 GCUAAGGAUGGUGAUUUC; siRNA-06 GCACCGUGGAUUUGGUCAA; siRNA-07 AGGCUAUGCUGCUUCGAAU, siRNA-08 GACACUGGCUGAUCUCAUG, Darmacon) directed at *GSTT2* were used at a concentration of 10nm. For controls, we use a non-target (NT) siRNA (10nm), as well as treated cells with just the OPTI-MEM plus RNAimax (mock control)(Invitrogen). Cells were plated and transfected using Lipofectamine® RNAiMAX Transfection Reagent (as directed by RNAiMAX protocol, Cat. No. 13778030(Thermo Fisher)) on day 1, with the following groups: mock, NT, si05, si06, si07, and si08. On day 2, we changed the media and proceeded to transfect again. Day 3; we changed the media and incubated cells for 48hrs. After 48hrs, coverslips were transferred to a 24-well plate, cells when then treated with either 0μM or 100μM of cumene-OOH (247502 Sigma-ALDRICH) for 1hr. After treatment, fixed the cells for 20min at -20 °C using 100% cold methanol. After fixing, we washed the cells with PBS three times (5min) and incubated them in PBS at 4 °C until performing the immunofluorescence protocol.

Over-expression of GSTT2 was performed using an OriGene construct for GFP (control) and GSTT2 (RC200040, NM_000854). HET1A cells were plated in 35mm x 10mm polystyrene treates plates at a starting number of 120,000 cells per plate. On day 2, cells were transfected using Fu-Gene transfection reagent at a 1:3 ratio. After 24hr transfection, media was changed, and cells were treated with 0μM or 100μM of cum-OOH for a 1hr period. Cells were fixed and stored using the methods described above.

**Immunofluorescence analysis.**

On day 1, we removed the PBS and cross-linked the protein-DNA with 500µL of 10% phosphate buffered formalin for 20min at room temperature (RT). The cells were washed once with 500µL TBS followed by removal and the cells were permeabilized with 100% cold methanol (-20°C). After permeabilization, the cells were incubated for 1hr in a blocking buffer containing 5% goat serum, 1% BSA, and 0.2% Triton x100 in TBS (500µl of blocking buffer per well). We diluted the primary antibody (γ-H2AX 1:1500, GSTT2-1:500) in TBS with 1% BSA. 150µL for each cover glass was used followed by incubation of the cells overnight at 4 °C, in a humidified chamber. The next day the cells were washed with TBS-T for 5min, 3 times. The cells were then incubated with secondary antibody for 1hr at room temperature (Alexa Fluor 488 goat anti-mouse IgG1 (gamma1)(Cat. No. A21121); Alexa Fluor 594 goat anti-mouse IgG2a (gamma2a)(Cat. No. A21135)(Life Technologies)) at dilutions of 1:50 and 1:200 respectively, in 1% BSA, followed by three 5min washes with TBS-T. Coverslips were mounted onto slides using DAPI mountain solution (Cat. No. P36935, Thermo Fisher) and stored at -20 °C prior to microscopic and photographic imaging.

**Western blotting.**

Cells were incubated and plated as described above (Treatments cum-H2O2 section). After 1 wash with PBS, 40ul of Cell Signaling protein lysis buffer (1X) with proteinase inhibitor cocktail (PIC)(20µl per mL) was added followed by scraping the plates. Lysates were transferred from the plates to a 1.5mL tube and spun for 20min at 4 °C. Supernatant was transferred to a new 1.5mL tube followed by protein quantification using protein assay. All lysates for westerns were prepared at a 20µg concentration, with 20% 2-mercaptoethanol, and 4X sample buffer. Samples were run in SDS Novex gels (4-12% gradient)(Invitrogen), for 2hrs at 125-126V. Nitrocellulose

membranes were used and activated with 100% methanol for 1min. The transfer was performed using Novex Transfer Buffer, and run overnight at 12V. After transfer, membranes were stained with Ponceau red for assessment of transfer efficiency. Blocking was performed by incubating the membranes in 5% milk/TBST or 5% BSA/TBST for 1hr at room temperature. Primary antibodies [(GSTT2 (mouse monoclonal isotype IgG2a κ - Santa Cruz, Cat. No. 514667); γ-H2AX (Anti-phospho-Histone H2A.X (Ser139) isotype IgG1 -Millipore, Cat. No. 05-636)] were diluted in 5% milk/TBST or 5% BSA/TBST and incubated over night at 4 °C. Membranes were then wash 3 times (5min per wash) with TBS-T. Secondary antibodies were diluted in 5% milk/TBST and added to the membranes for 1hr incubations at room temperature. The membranes were washed 3 times with TBST, 5min per wash and then incubated with ECL for 5mins before revealing using X-ray film.

**Indomethacin treatments.**

Indomethacin (Cat No. I7378; Sigma-Aldrich) was diluted to 1M in 100% DMSO. Indomethacin (1M) was added to media use for HET1A (DMEM, 10% FBS) and diluted at concentrations of 300μM and 600μM. Since DMSO was used to dilute indomethacin, we use 0.6% DMSO treated HET1A has control. After we added indomethacin to the media, we followed by an incubation in the water bath (37°C) for 1hr, or until indomethacin went into solution. 300,000 cells per well were plated on day 1 and incubated overnight for attachment. On day 2, media + indomethacin was added to the cells and left for the respective time points (6, 24, 48, and 72hrs). 350μl of QIAzol Lysis Reagent (Qiagen, Valencia, CA) for mRNA, or 40 μl Cell Signaling protein lysis buffer (1X) with proteinase inhibitor cocktail (PIC)(20ul per mL) for protein extractions, were added at each time point and store in -80C° until all time points were

collected. mRNA and protein extractions were performed on all samples following protocols, described above.

**Chromatin immunoprecipitation (ChIP).**

Chromatin immunoprecipitation was performed according to the methods of Gilfillan *et al.* 2012[40] and Krook *et al.* 2016[41]. Summarizing, HET1A cells ($3.0x10^5$ per IP) were digested using Micrococcal nuclease (MNase) (70196Y, Affymetrix, Santa Clara, CA) for 5 minutes at 37°C. Follow by sonication for 20 seconds (Qsonica cup horn sonicator (Qsonica Sonicators, Newtown, CT, USA)), blocked for 1 hour with Dynabeads A+G (10001D and 10003D; Life Technologies, Carlsbad, CA). After blocking, we proceeded to incubate with 1 µg of desired antibody concentrations overnight. The next day they were incubated with Dynabeads A+G for 3 hours, washed (5 minutes wash; 5 x RIPA buffer, 1 x LiCl buffer, 1 x TE buffer), followed by digestion of proteins with Proteinase K for 1 hour at 55°C. Finally, we purified immunoprecipitated DNA according to manufacturer's instructions (Zymo Genomic DNA Clean & Concentrator, D4011). Primer pairs for the *GSTT2/2B* promoter are listed in the appendix. ChIP antibodies were used as per manufacturer's instructions; H3K4me3 Rabbit anti-Human Polyclonal Antibody (49-1005; Life Technologies, Carlsbad, CA), anti-trimethyl-Histone H3 (Lys27) Antibody (07-449; Millipore, Billerica, MA), normal mouse IgG (sc-2025; Santa Cruz Biotechnology, Dallas, TX), Rabbit IgG (ab37415; Abcam, Cambridge, MA).

**Figures**

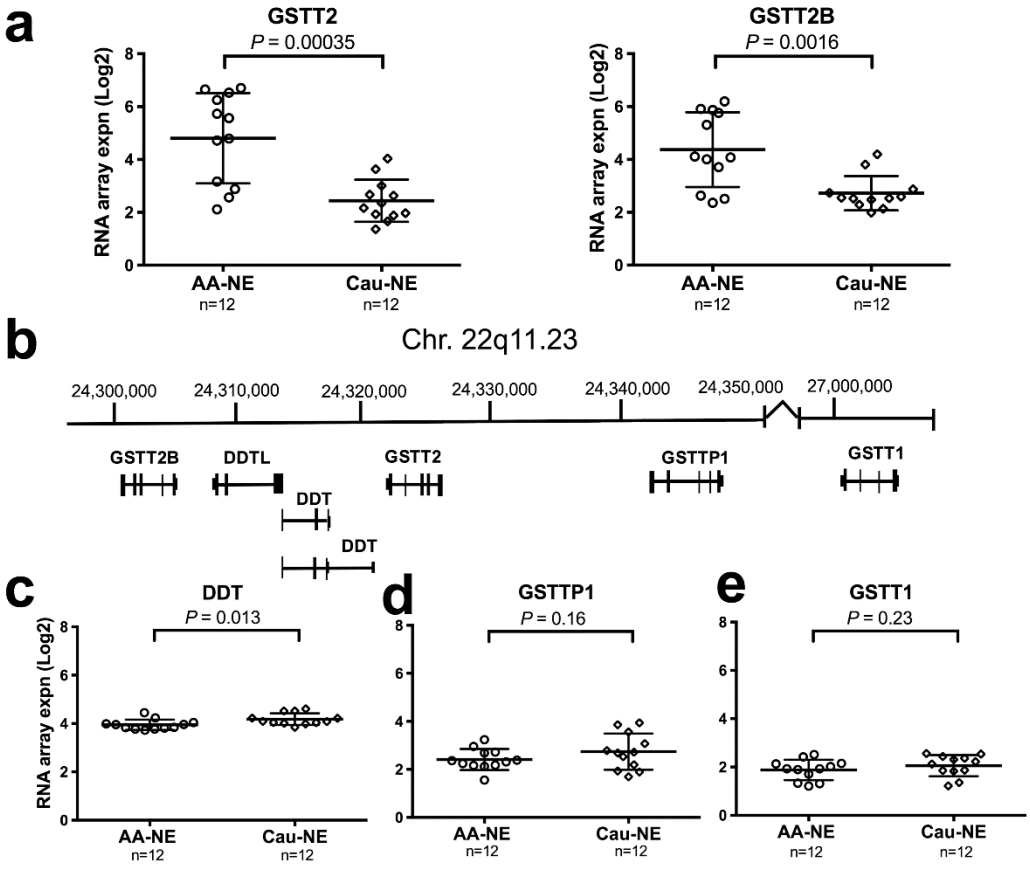



**Figure 4.1. *GSTT2 and GSTT2B* expression in the NE of AA vs Cau** (**a**) Array analysis of normal squamous esophageal tissue (NE) from AA and Cau reveals differential mRNA expression of *GSTT2* (left panel) and *GSTT2B* (right panel), respectively. **(b)** Gene locus of *GSTT2* and *GSTT2B* coding genes on chromosome 22 including GSTT family members *GSTTP1* and *GSTT1*. **(c-e)** Array analysis of normal squamous (NE) tissue from AA and Cau indicates *DDT* or GSTT family members *GSTTP1* and *GSTT1* are not differentially expressed.

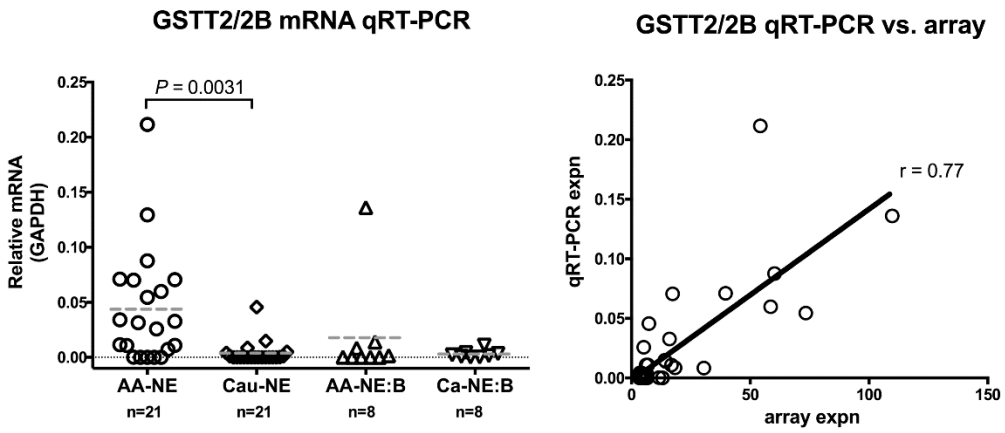



**Figure 4.2 *GSTT2*/2B mRNA array vs. qRT-PCR (a)** Validation of *GSTT2/2B* expression in the NE of AA and Cau with and without Barrett's (B) using qRT-PCR and with a larger cohort confirms differential mRNA expression of *GSTT2/2B* in AA vs. Cau ($P$=0.0031). **(b)** Pearson correlation of mRNA of *GSTT2/2B* between qRT-PCR vs array samples (r=0.77, $P<1.0\text{x}10^{-8}$).

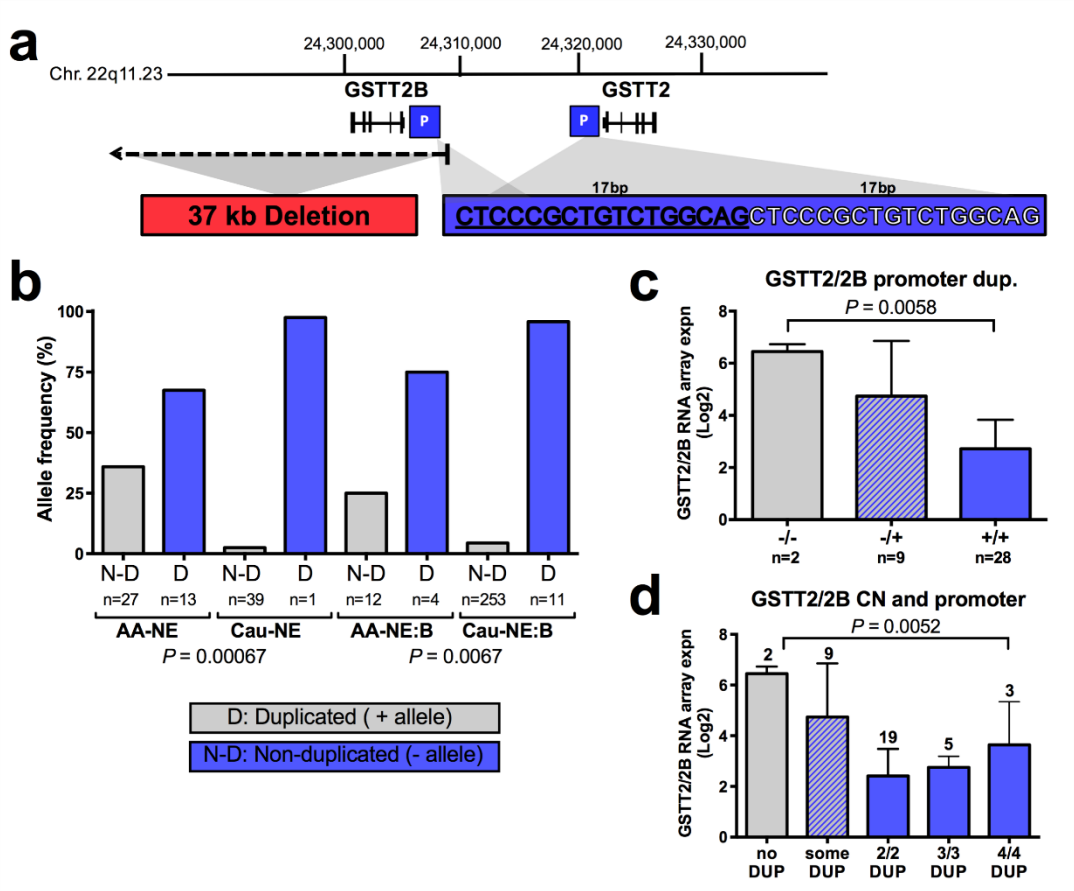



**Figure 4.3** ***GSTT2/2B*** **17bp promoter duplication is associated with reduced mRNA expression (a)** Two genomic events influence *GSTT2/2B* mRNA levels: a 37kb deletion that removes the *GSTT2B* gene, and the 17bp tandem *GSTT2/2B* promoter duplication. **(b)** The 17bp *GSTT2/2B* promoter duplication frequency is significantly lower in AA vs. both Cau and the AA:B and Cau:B disease populations by Fisher Exact test. **(c)** When all squamous samples are combined and analyzed (ANOVA) we observed that the 17bp promoter duplication shows a dose-dependent association with *GSTT2/2B* mRNA expression. **(d)** The combination of the 17bp *GSTT2/2B* promoter duplication plus the deleted *GSTT2B* genotype common to Cau have much lower expression as compared to individuals having at least one copy of the non-duplicated promoter.

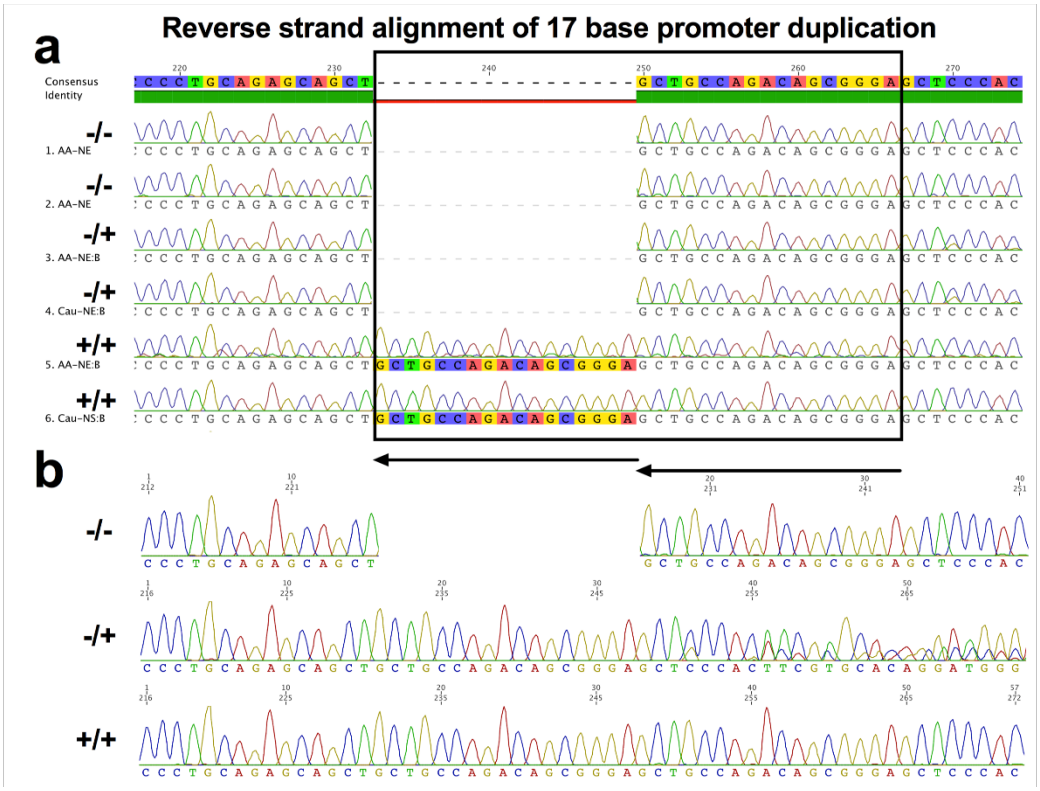



**Figure 4.4 *GSTT2* promoter duplication genotype sequence in AA and Cau** Examples of the *GSTT2/2B* 17bp duplication genotypes using reverse strand sequences showing **(a)** the alignment generated using Geneious and **(b)** manual alignment of raw sequence traces.

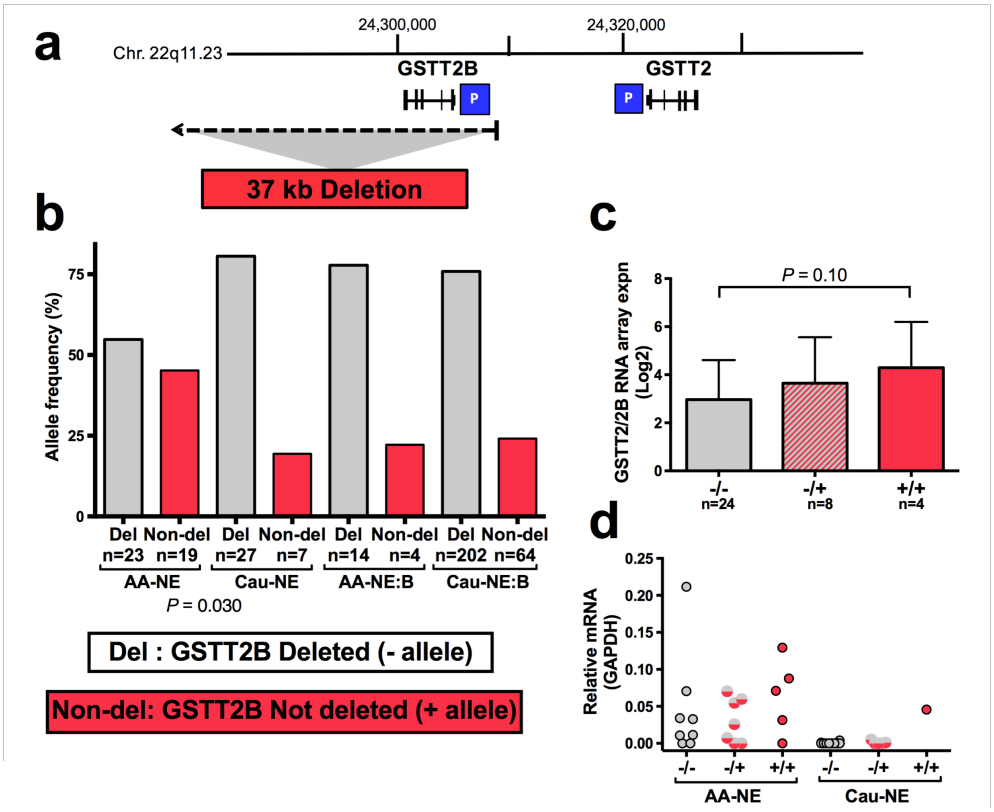



**Figure 4.5 *GSTT2* 37kb deletion in AA vs Cau (a)** *GSTT2/2B* locus structure showing the 37kb deletion that removes the *GSTT2B* locus. **(b)** Allele frequency of the 37kb *GSTT2B* deletion in NE of Cau and AA with disease (B) and without disease (NE). A trend towards an increased incidence of the deletion in Cau was noted (Fisher Exact). Comparisons involving AA-NE:B verse AA:NE or either Cau group were not significant. The comparison between Cau groups was also not significant. **(c)** When all squamous samples were combined a dose-related effect of the *GSTT2B* deletion on *GSTT2/2B* mRNA expression was suggested but did not reach statistical significance (Kruskal–Wallis test) because of low sample numbers. **(d)** The presence or absence of BE did not influence the effect of the *GSTT2B* deletion on *GSTT2/2B* mRNA expression.

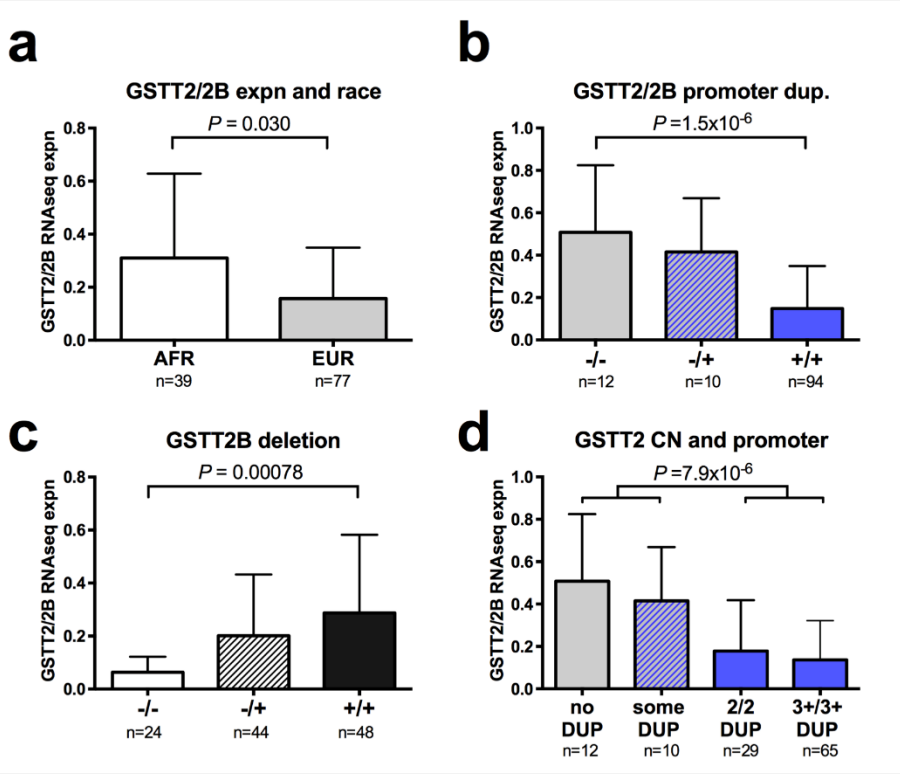



**Figure 4.6 1000 Genomes Data analysis of *GSTT2/2B* genotype and mRNA** We used publicly available, matched RNAseq and DNA copy number data from cultured lymphoblasts from a subset of 1000 Genomes population controls (n=116) from normal African (AFR) and Caucasian (EUR) individuals that show: **(a)** the same trend as our esophageal NE samples of higher average *GSTT2/2B* expression in individuals of African descent. We also confirmed that both the **(b)** *GSTT2/2B* promoter duplication and **(c)** *GSTT2B* deletion have gene dose-related effects upon expression such that **(d)** individuals homozygous for the promoter duplication have much lower *GSTT2/2B* expression than individuals with at least one non-duplicated copy**.**

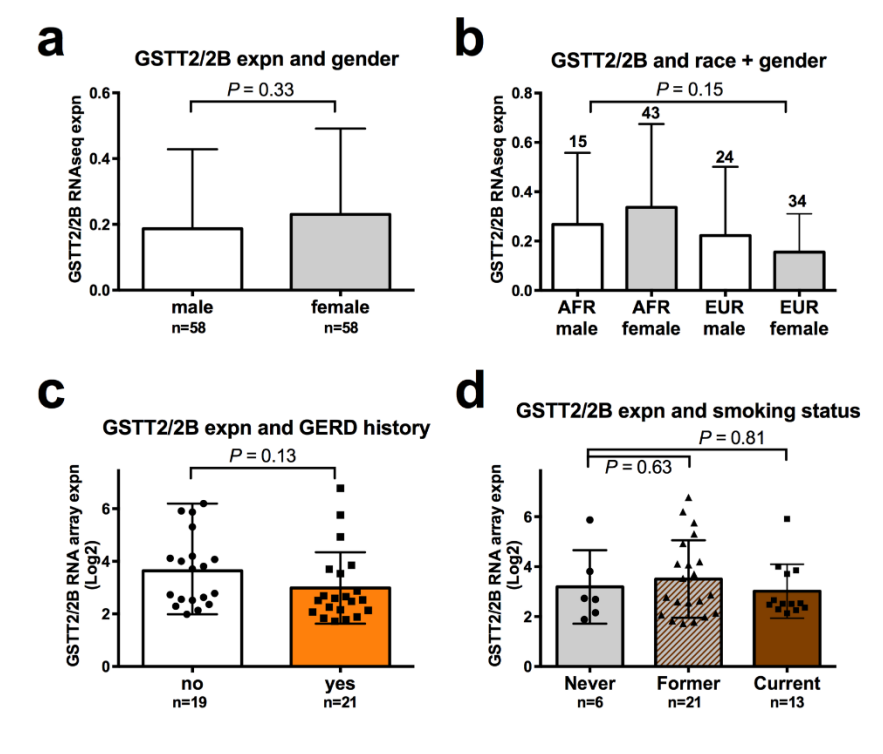



**Figure 4.7 *GSTT2/2B* mRNA levels by gender, GERD history, and smoking status** Using 1000 Genomes population controls (n=116) we observe that differences in *GSTT2/2B* mRNA expression is not significantly different between **(a)** genders but is **(b)** race dependent. Within the 40 arrayed NE samples differences in *GSTT2* expression were not explained in terms of either the presence of GERD nor smoking status (ANOVA).

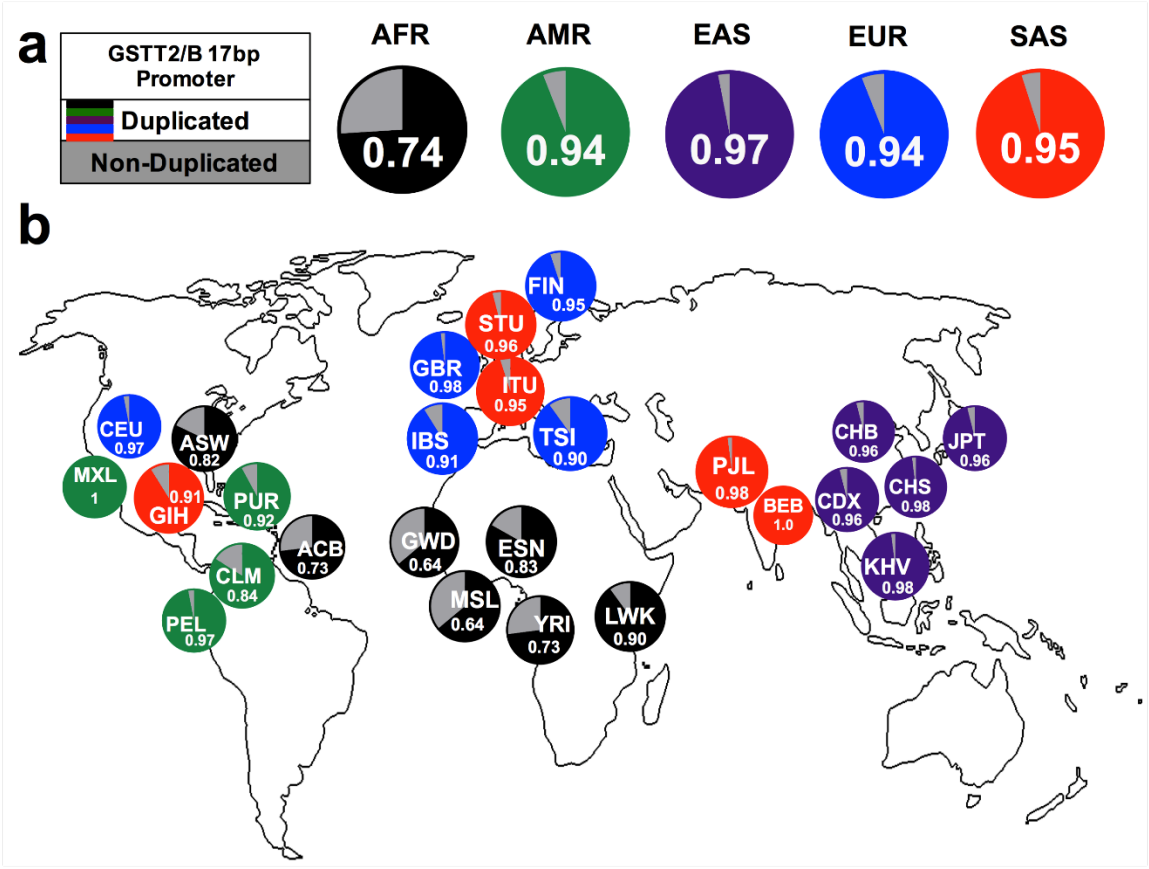



**Figure 4.8 *GSTT2/2B* promoter allele frequency among human populations (a)** Frequency of *GSTT2/2B* promoter duplication and non-duplicated alleles in super populations (1000 Genomes data). **(b)** Frequency of the non-duplicated *GSTT2/2B* promoter is highest among African and African descent populations. The abbreviations for subpopulations are as described (**Table 4.3**).

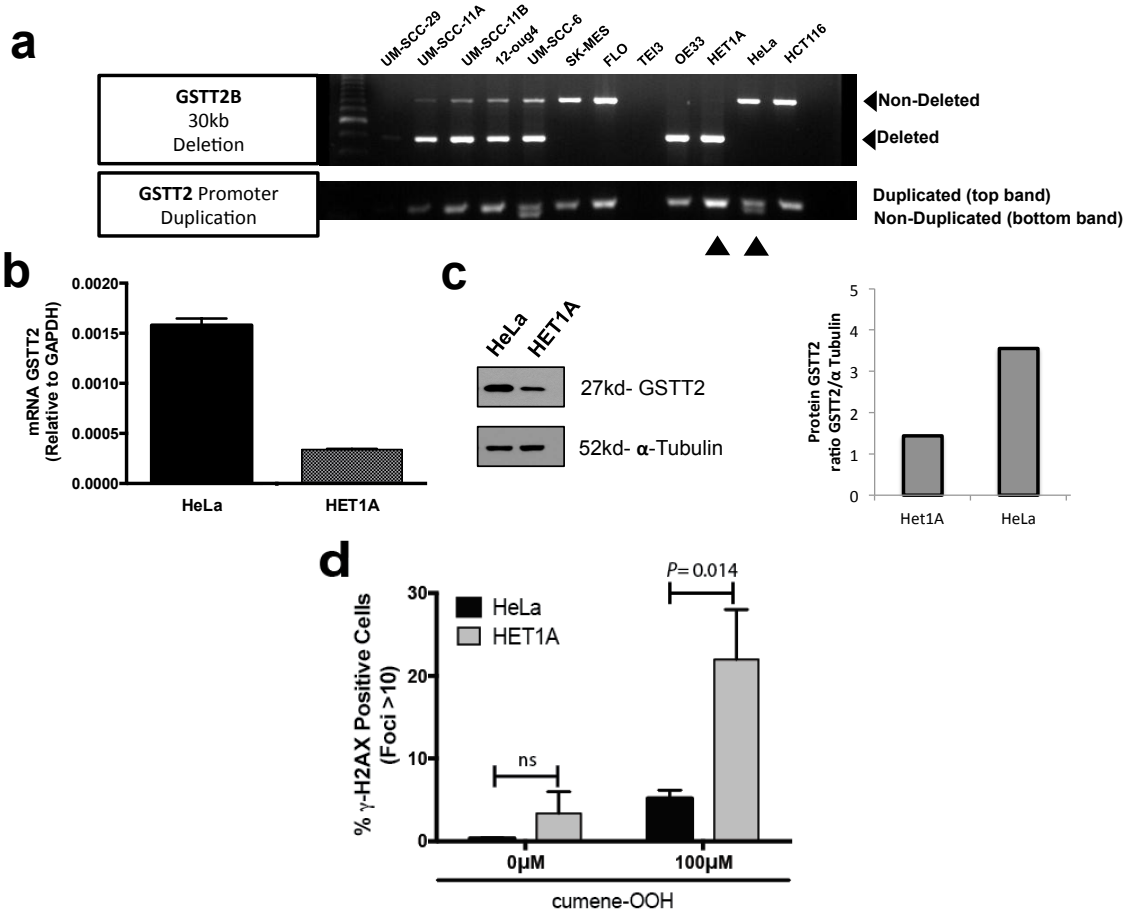



**Figure 4.9 HET1A and HeLa assesment of susceptibility to DNA damage with and without cum-OOH treatments** (a) *GSTT2*/2B 37kb deletion and promoter duplication genotyping in a cohort of 12 cell lines. Bottom triangles highlight the HET1A and HeLa cell lines, respectively. Cells were quantified (400,000) and used to extract both mRNA and protein. (b-c) qRT-PCR and western blots were performed in HeLa and HET1A to measure endogenous *GSTT2*/2B and GSTT2 levels. (d) Quantification of positive foci γ-H2AX in HET1A and HeLa cells. A nuclei with >10 foci was considered a positive cell with DNA damage.

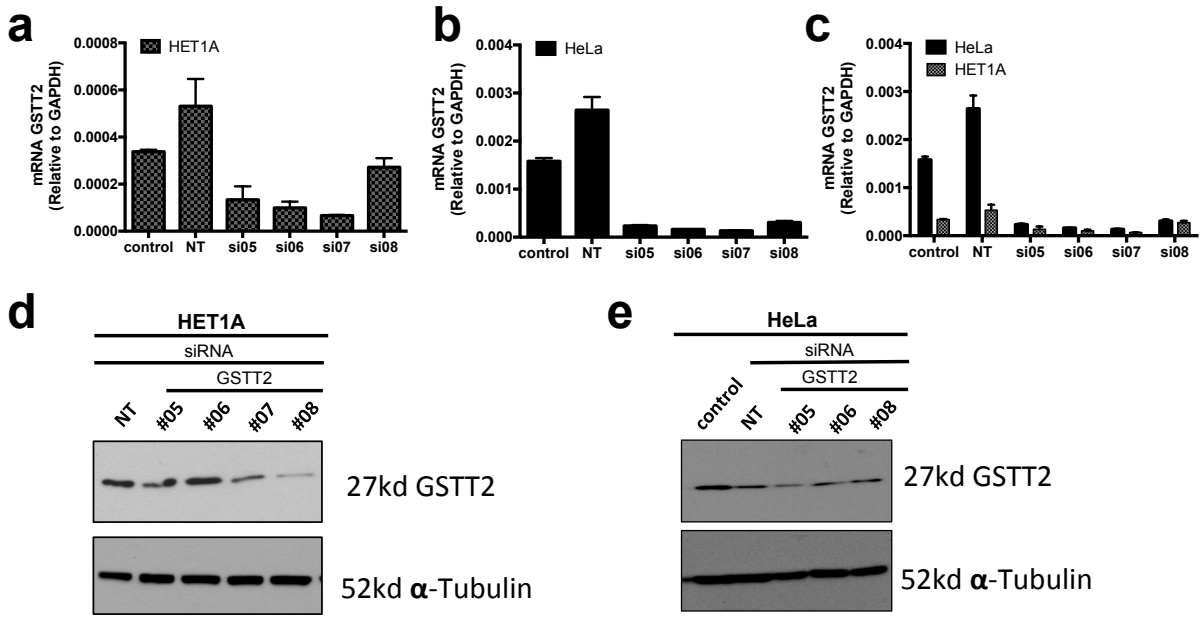



**Figure 4.10 Validation of GSTT2 knockdown in HET1A and HeLa cells** (a-c) Relative *GSTT2/2B* mRNA expression in HET1A and HeLa cells after 48hr transfection with four siRNAs. (d-e) GSTT2 protein levels in HET1A and HeLa cells after 48hr transfection with four siRNAs.

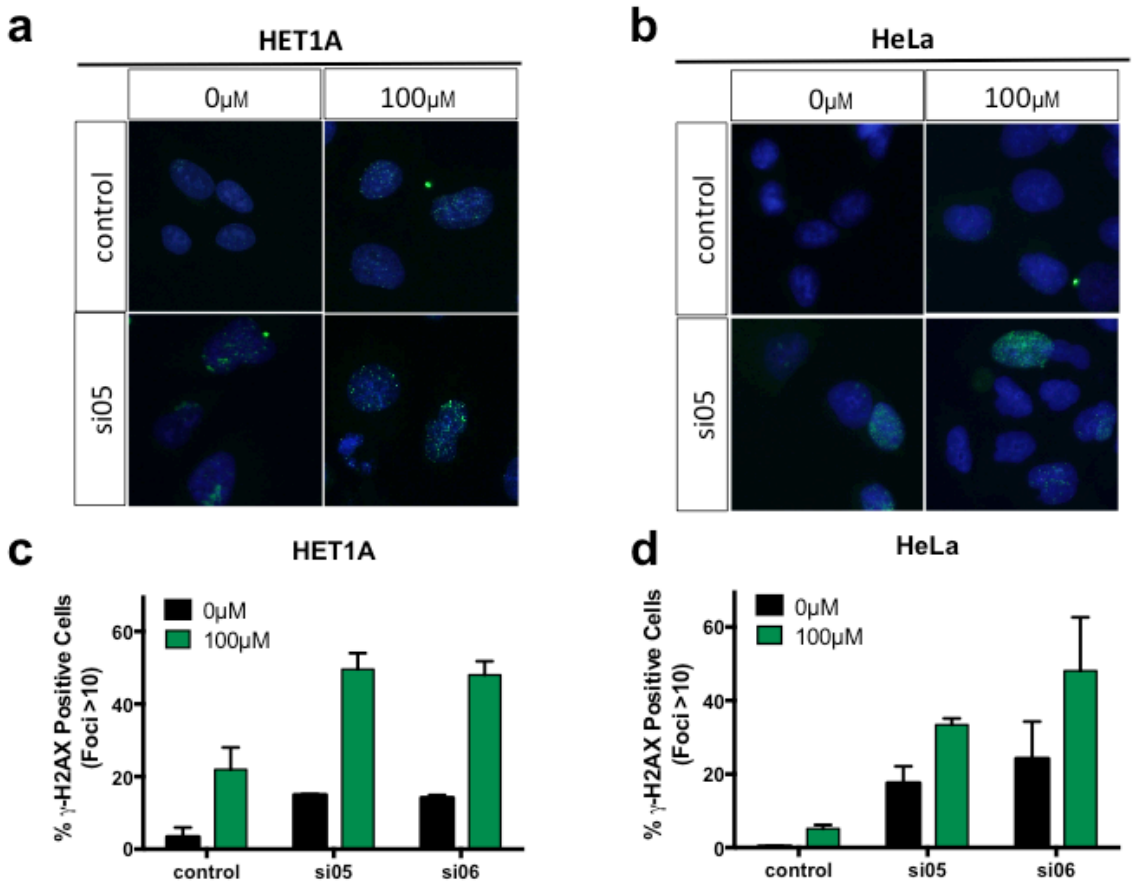



**Figure 4.11 Knockdown of GSTT2 in HET1A and HeLa cells with and without cum-OOH treament** (a-b) HET1A and HeLa cells were treateted with cum-OOH (0μM and 100μM)(1hr), DNA damage was assessed using immunofluorescence staining for γ-H2AX (green) and nuclei (DAPI). (c-d) Quantification of positive foci for γ-H2AX in HET1A and HeLa cells. A nuclei with >10 foci was considered a positive cell with DNA damage.

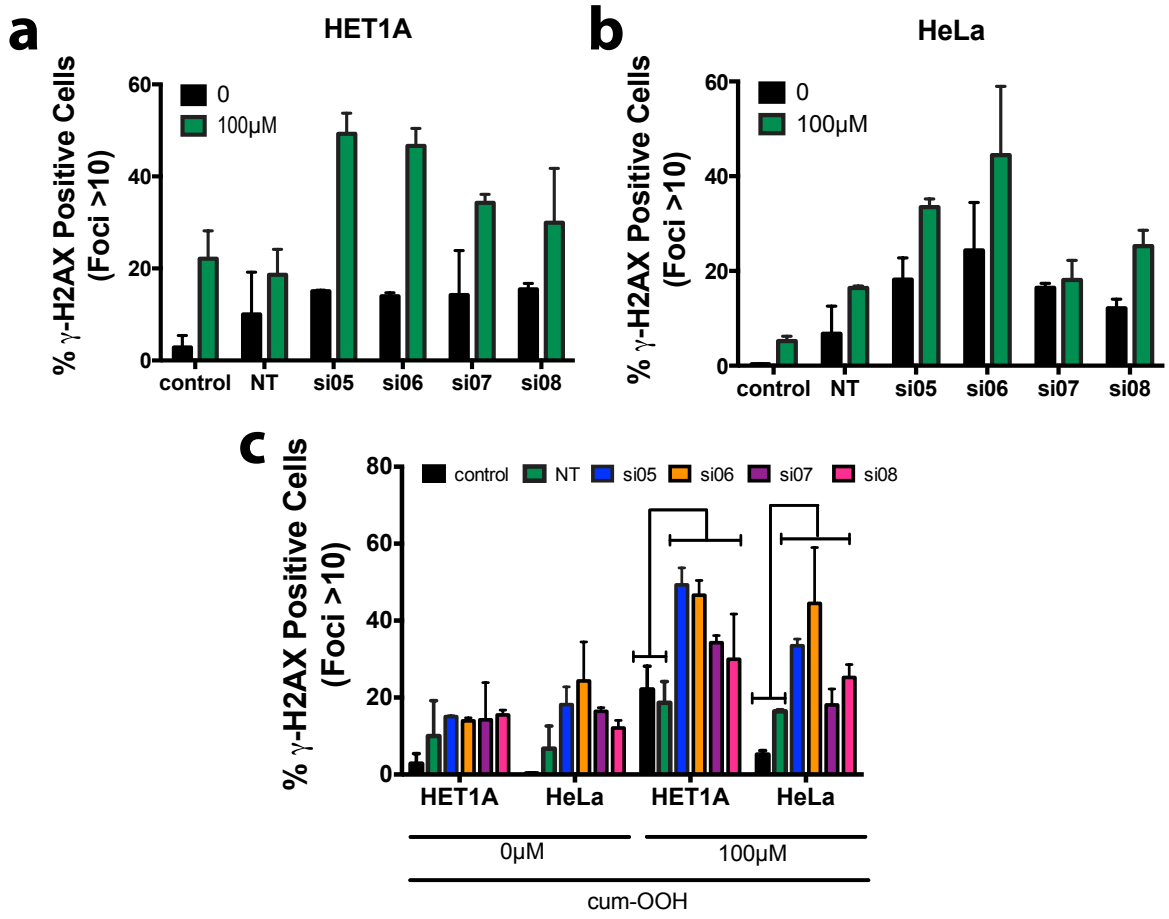



**Figure 4.12 Knockdown of GSTT2 in HET1A and HeLa cells with and without cum-OOH treament** (a-b) HET1A and HeLa cells were treated with cum-OOH (0μM and 100μM)(1hr), DNA damage was assessed using immunofluorescence staining for γ-H2AX (green) and nuclei (DAPI). (c) Quantification of positive foci γ-H2AX in HET1A and HeLa cells. A nuclei with >10 foci was considered a positive cell with DNA damage.

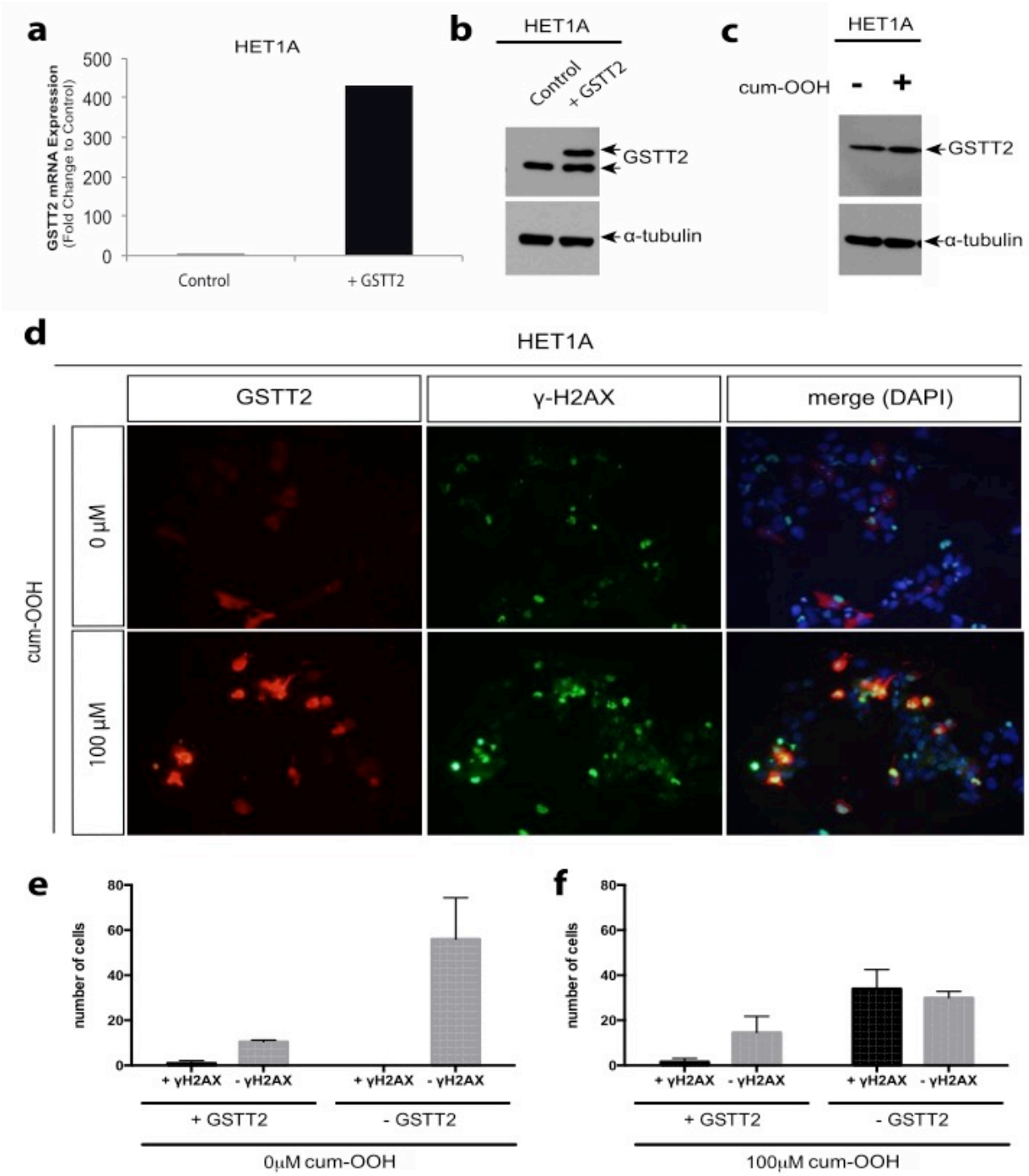



**Figure 4.13 HET1A over-expressing GSTT2 and following genotoxic stress** (a) mRNA levels of *GSTT2* in HET1A (qRT-PCR) cells and *GSTT2* transfected (+ GSTT2) vs. control (GFP-transfected cells). (b) Western blot of HET1A cells transfected with GSTT2 (+GSTT2)(top panel). (c) Western blot showing GSTT2 protein expression in HET1A treated with 0μM or 100μM of cum-OOH for 1hr. (d) Dual immunofluorescence staining of GSTT2 (red) and γ-H2AX in HET1A cells transfected with GSTT2, cells were treated with 100μM of cum-OOH for 1hr. (nuclei staining is depicted in blue using DAPI). (e-f) quantification of foci formation in GSTT2 positve vs negative cells, and with and without cum-OOH treatment.

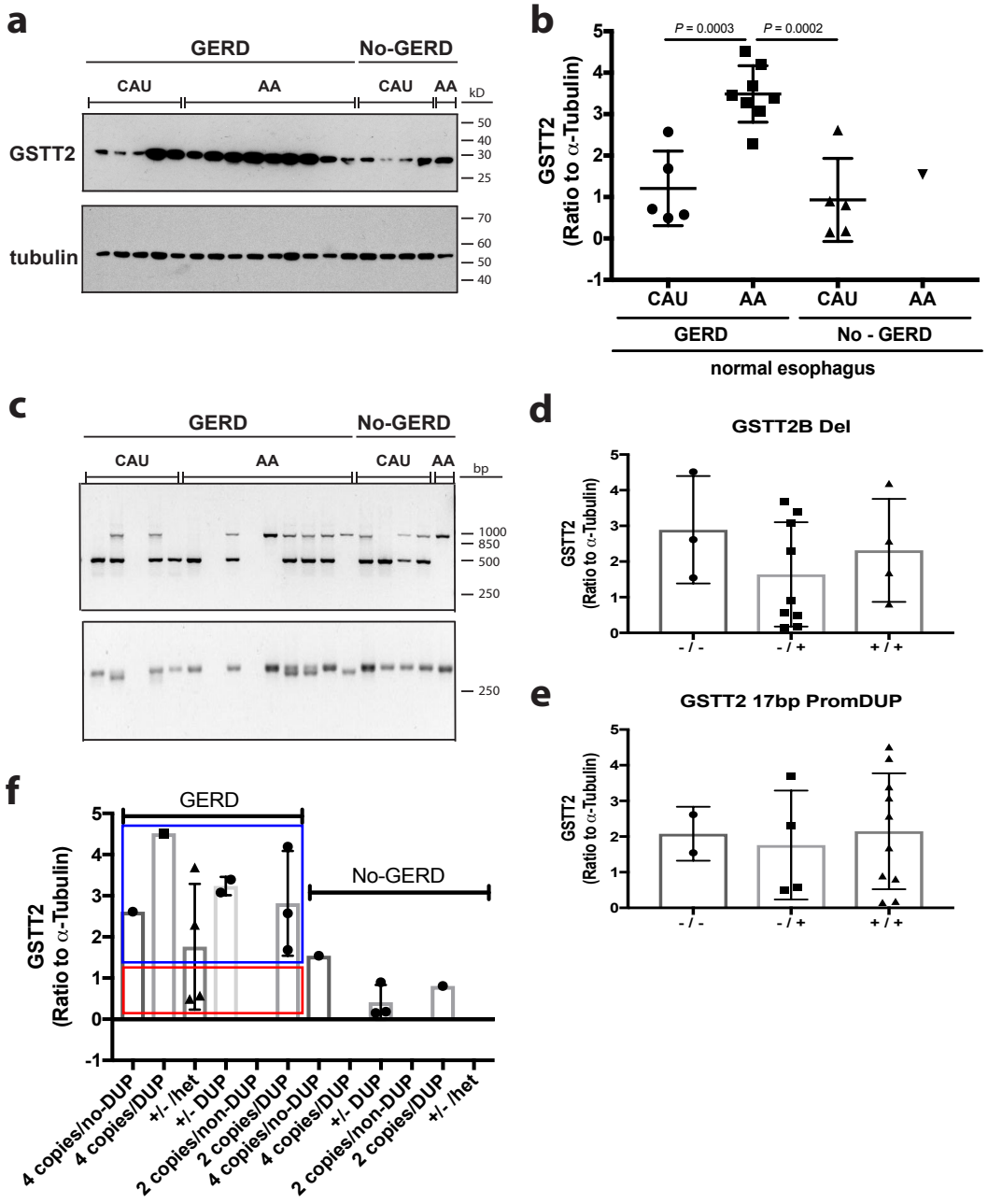



**Figure 4.14 GSTT2 expression in AA vs Cau normal esophagus epithelium in GERD and non-GERD individuals** (a) Western blot depicting GSTT2 expression in a cohort of AA and CAU normal esophagus, alpha-tubulin was used as loading control (b) Quantification of protein expression of GSTT2 in AA and CAU normal esophagus biopsies (GSTT2 expression in each sample relative to alpha-tubulin). (c) Genotype analysis was done by PCR, for the 37kb deletion (top panel: top band (~850bp) Non-Del; bottom band (~510) Deleted), and *GSTT2*/2B promoter duplication (bottom panel: top band (~317bp) DUP; bottom band (~300bp) Non-DUP). (d) Genotype for *GSTT2/2B* 37kb Deletion (-/-)(-/+) and non-deleted (+/+) associated with levels of GSTT2 protein. (e) Genotype for *GSTT2/2B* Promoter Duplication (+/+)(-/+) and non-duplicated

(-/-) associated with levels of GSTT2 protein. (f) Both genotypes matched to protein levels in GERD vs. No-GERD.

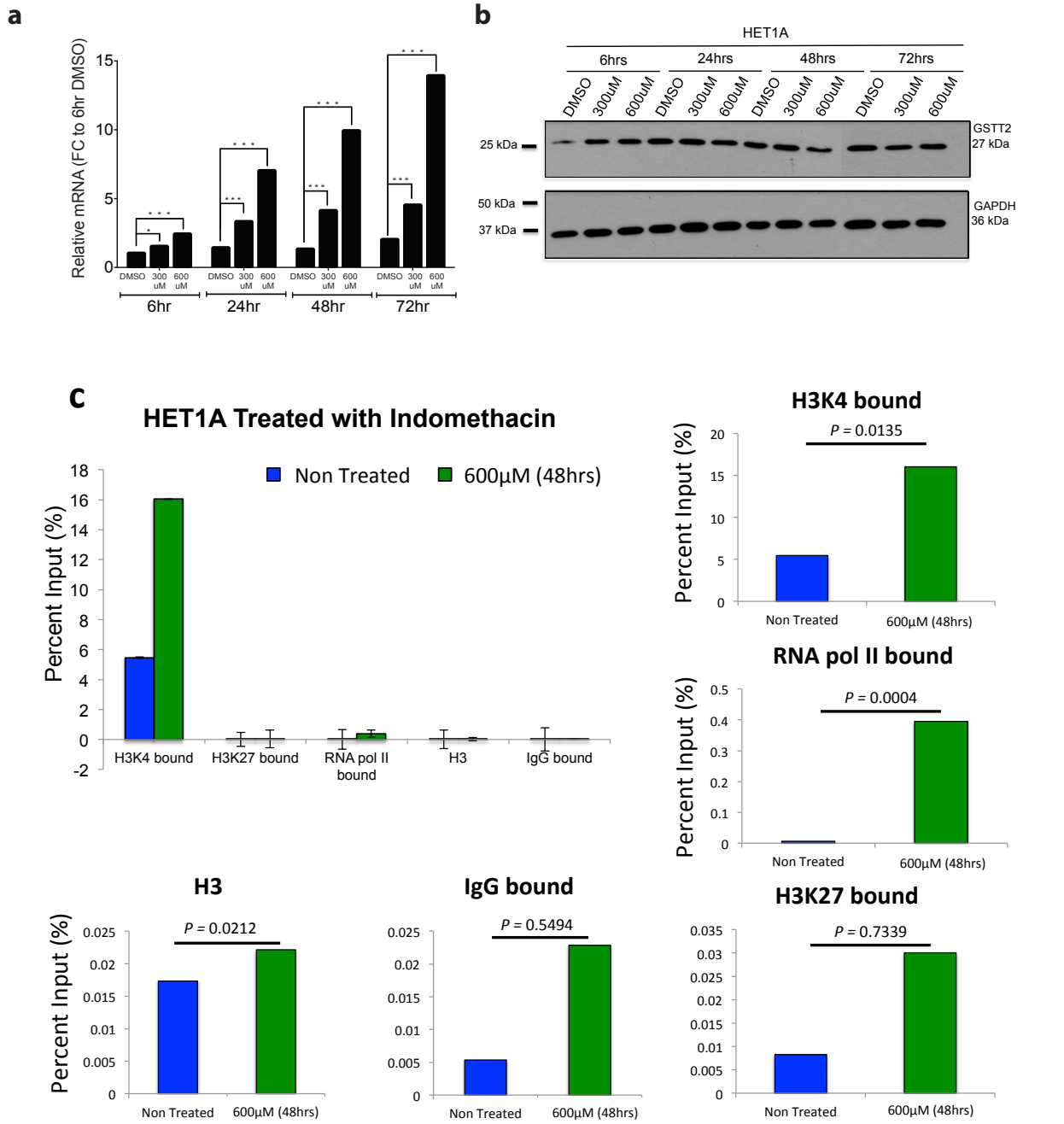



**Figure 4.15 HET1A treatment with Indomethacin (a)** mRNA, (**b**) western, and (**c**) ChIP of HET1A after treatment with Indomethacin (300µM, 600µM), and DMSO (control). HET1A were treated with two different concentrations and samples were collected at different time points (6hr, 24hr, 48hrs, and 72hrs). ChIP experiment was performed using 600µM and harvesting cells after 48hrs incubation with indomethacon (600µM).

# TABLES

## Table 4.1 Summary of Analysis Cohort

| | N | Age >55 years (%) | Gender Male (%) | Tobacco smoked (%) | GERD diagnosed (%) | BMI overweight (%) | BE/EAC path diagnosed (%) |
|---|---|---|---|---|---|---|---|
| AA-NE | 12 | 8 (67%) | 12 (100%) | 11 (92%) | 2 (17%) | 11 (92%) | 0 (0%) |
| Cau-NE | 12 | 5 (42%) | 12 (100%) | 10 (83%) | 4 (33%) | 11 (92%) | 0 (0%) |
| AA-NE:BE | 8 | 7 (88%) | 5 (63%) | 7 (88%) | 8 (100%) | 7 (88%) | 8 (100%) |
| Cau-NE:BE | 8 | 7 (88%) | 6 (75%) | 6 (75%) | 7 (88%) | 6 (75%) | 8 (100%) |

## Table 4.2 All genes >2 FC between AA vs Cau (Affymetrix 2.1 ST Array)

| | | **ANOVA p-values*** | | | | **Fold Change*** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Symbol** | **Name** | **AA-NE vs C-NE** | **AA-NE:B vs C-NE:B** | **AA-NE:B vs AA-NE** | **C-NE:B vs C-NE** | **AA-NE / C-NE** | **AA-NE:B / C-NE:B** | **AA-NE:B / AA-NE** | **C-NE:B / C-NE** |
| **GSTT2** | **glutathione S-transferase theta 2** | **0.0004** | **0.1372** | **0.0479** | **0.8557** | **5.15** | **2.17** | **0.39** | **0.92** |
| IGHD | immunoglobulin heavy constant delta | 0.0076 | 0.0307 | 0.0644 | 0.0039 | 5.07 | 0.21 | 0.29 | 7.24 |
| IGHA1 | immunoglobulin heavy constant alpha 1 | 0.0074 | 5.4E-06 | 0.0169 | 1.0E-06 | 3.77 | 0.05 | 0.27 | 21.53 |
| **GSTT2B** | **glutathione S-transferase theta 2B (gene/pseudogene)** | **0.0016** | **0.07** | **0.1038** | **0.5163** | **3.13** | **2.15** | **0.54** | **0.78** |
| HLA-DPB1 | major histocompatibility complex, class II, DP beta 1 | 0.0072 | 0.9458 | 0.6697 | 0.0486 | 3.13 | 1.03 | 0.82 | 2.5 |
| MS4A1 | membrane-spanning 4-domains, subfamily A, member 1 | 0.0054 | 2.8E-08 | 0.0194 | 2.0E-09 | 2.31 | 0.09 | 0.46 | 12.27 |
| LOC643669 | uncharacterized LOC643669 | 0.0003 | 0.0242 | 0.6023 | 0.655 | 2.13 | 1.73 | 0.89 | 1.1 |
| MIR4518 | microRNA4518 | 0.0071 | 0.8133 | 0.3397 | 0.0731 | 2.11 | 0.93 | 0.75 | 1.71 |

* *P*-values calcualted on Log2 values. Fold changes (FC) were calculated and are presented in non-log values

## Table 4.3 Promoter duplication allele frequency across world populations

| Population Description | Population Code | Super Population | Total samples* | Duplicated | Non-Duplicated | Duplicated frequency | Non- duplicated frequency |
|---|---|---|---|---|---|---|---|
| African Caribbean in Barbados | ACB | AFR | 96 | 135 | 50 | 0.73 | 0.27 |
| African Ancestry in Southwest US | ASW | AFR | 66 | 46 | 10 | 0.82 | 0.18 |
| Esan in Nigeria | ESN | AFR | 99 | 157 | 32 | 0.83 | 0.17 |
| Gambian in Western Division, The Gambia | GWD | AFR | 113 | 131 | 73 | 0.64 | 0.36 |
| Luhya in Webuye, Kenya | LWK | AFR | 116 | 124 | 14 | 0.90 | 0.10 |
| Mende in Sierra Leone | MSL | AFR | 85 | 105 | 59 | 0.64 | 0.36 |
| Yoruba in Ibadan, Nigeria | YRI | AFR | 116 | 110 | 40 | 0.73 | 0.27 |
| Colombian in Medellin, Colombia | CLM | AMR | 95 | 109 | 21 | 0.84 | 0.16 |
| Mexican Ancestry in Los Angeles, California | MXL | AMR | 69 | 43 | 0 | 1.00 | 0.00 |
| Peruvian in Lima, Peru | PEL | AMR | 86 | 212 | 6 | 0.97 | 0.03 |
| Puerto Rican in Puerto Rico | PUR | AMR | 105 | 128 | 11 | 0.92 | 0.08 |
| Chinese Dai in Xishuangbanna, China | CDX | EAS | 99 | 213 | 10 | 0.96 | 0.04 |
| Han Chinese in Bejing, China | CHB | EAS | 106 | 190 | 8 | 0.96 | 0.04 |
| Southern Han Chinese, China | CHS | EAS | 112 | 128 | 3 | 0.98 | 0.02 |
| Japanese in Tokyo, Japan | JPT | EAS | 105 | 156 | 6 | 0.96 | 0.04 |
| Kinh in Ho Chi Minh City, Vietnam | KHV | EAS | 101 | 247 | 6 | 0.98 | 0.02 |
| Utah residents with Northern and Western European ancestry | CEU | EUR | 103 | 162 | 5 | 0.97 | 0.03 |
| Finnish in Finland | FIN | EUR | 100 | 99 | 5 | 0.95 | 0.05 |
| British in England and Scotland | GBR | EUR | 94 | 115 | 2 | 0.98 | 0.02 |
| Iberian populations in Spain | IBS | EUR | 107 | 161 | 15 | 0.91 | 0.09 |
| Toscani in Italy | TSI | EUR | 110 | 191 | 22 | 0.90 | 0.10 |
| Bengali in Bangladesh | BEB | SAS | 86 | 166 | 0 | 1.00 | 0.00 |
| Gujarati Indian in Houston,TX | GIH | SAS | 106 | 152 | 15 | 0.91 | 0.09 |
| Indian Telugu in the UK | ITU | SAS | 103 | 148 | 7 | 0.95 | 0.05 |
| Punjabi in Lahore,Pakistan | PJL | SAS | 96 | 180 | 3 | 0.98 | 0.02 |
| Sri Lankan Tamil in the UK | STU | SAS | 103 | 163 | 6 | 0.96 | 0.04 |

*not all samples yielded data for *GSTT2* promoter genotype status

## References


1. Pickens, A. & Orringer, M. Geographical Distribution and Racial Disparity in Esophageal Cancer. *Ann. Thorac. Surg.* **4975,** 1367–1369 (2003).
2. Thrift, a P. & Whiteman, D. C. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. *Ann. Oncol.* **23,** 3155–62 (2012).
3. Cummings, L. C. & Cooper, G. S. Descriptive epidemiology of esophageal carcinoma in the Ohio Cancer Registry. *Cancer Detect. Prev.* **32,** 87–92 (2008).
4. Lepage, C., Rachet, B., Jooste, V., Faivre, J. & Coleman, M. P. Continuing rapid increase in esophageal adenocarcinoma in England and Wales. *Am. J. Gastroenterol.* **103,** 2694–9 (2008).
5. Pohl, H., Sirovich, B. & Welch, H. G. Esophageal Adenocarcinoma Incidence: Are We Reaching the Peak? *Cancer Epidemiol. Biomarkers Prev.* **19,** 1468–1470 (2010).
6. Sudo, K. *et al.* Locoregional failure rate after preoperative chemoradiation of esophageal adenocarcinoma and the outcomes of salvage strategies. *J. Clin. Oncol.* **31,** 4306–4310 (2013).
7. El-Serag, HB and Sonnenberg, A. Associations between different forms of gastro-oesophageal reflux disease. **41**, 594–599 (1997).
8. Spechler, S. J., Jain, S. K., Tendler, D. A. & Parker, R. A. Racial differences in the frequency of symptoms and complications of gastro-oesophageal reflux disease. *Aliment. Pharmacol. Ther.* **16,** 1795–1800 (2002).
9. El-Serag, H. B. *et al.* Gastroesophageal reflux among different racial groups in the United States. *Gastroenterology* **126,** 1692–1699 (2004).
10. Petermann, A. *et al.* GSTT2, a phase II gene induced by apple polyphenols, protects colon epithelial cells against genotoxic damage. *Mol. Nutr. Food Res.* **53,** 1245–53 (2009).
11. Zhao, Y., Marotta, M., Eichler, E. E., Eng, C. & Tanaka, H. Linkage disequilibrium between two high-frequency deletion polymorphisms: implications for association studies involving the glutathione-S transferase (GST) genes. *PLoS Genet.* **5,** e1000472 (2009).
12. Marotta, M., Piontkivska, H. & Tanaka, H. Molecular trajectories leading to the alternative fates of duplicate genes. *PLoS One* **7,** e38958 (2012).
13. Hayes, J. D. & Pulford, D. J. The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit. Rev. Biochem. Mol. Biol.* **30,** 445–600 (1995).
14. Armstrong, R. N. Structure, Catalytic Mechanism, and Evolution of the Glutathione Transferases. *Chem. Res. Toxicol.* **10,** 2–18 (1997).
15. Dirr, H., Reinemer, P. & Huber, R. X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *Eur. J. Biochem.* **220,** 645–61 (1994).

16. Pickett, C. B. & Lu, A. Y. H. Glutathione s-transferases: gene structure, regulation, and biological function. *Annu. Rev. Biochem.* **58,** 743–764 (1989).

17. Eaton, D. L. & Bammler, T. K. Concise review of the glutathione S-transferases and their significance to toxicology. *Toxicol. Sci.* **49,** 156–164 (1999).

18. Seidegård, J. & Ekström, G. The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics. *Environ. Health Perspect.* **105,** 791–799 (1997).

19. Veeriah, S. *et al.* Intervention with cloudy apple juice results in altered biological activities of ileostomy samples collected from individual volunteers. *Eur. J. Nutr.* **47,** 226–34 (2008).

20. Miene, C., Klenow, S., Veeriah, S., Richling, E. & Glei, M. Impact of apple polyphenols on GSTT2 gene expression, subsequent protection of DNA and modulation of proliferation using LT97 human colon adenoma cells. *Mol. Nutr. Food Res.* **53,** 1254–1262 (2009).

21. Veeriah, S. *et al.* Apple polyphenols modulate expression of selected genes related to toxicological defence and stress response in human colon adenoma cells. *Int. J. Cancer* **122,** 2647–55 (2008).

22. Pool-Zobel, B. L. *et al.* Butyrate may enhance toxicological defence in primary, adenoma and tumor human colon cells by favourably modulating expression of glutathione S-transferases genes, an approach in nutrigenomics. *Carcinogenesis* **26,** 1064–76 (2005).

23. Miene, C., Weise, A. & Glei, M. Impact of polyphenol metabolites produced by colonic microbiota on expression of COX-2 and GSTT2 in human colon cells (LT97). *Nutr. Cancer* **63,** 653–62 (2011).

24. Landi, S. Mammalian class theta GST and differential susceptibility to carcinogens: a review. *Mutat. Res.* **463,** 247–83 (2000).

25. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

26. Van Lieshout, E. M., Tiemessen, D. M., Roelofs, H. M. & Peters, W. H. Nonsteroidal anti-inflammatory drugs enhance glutathione S-transferase theta levels in rat colon. *Biochim. Biophys. Acta* **1381,** 305–11 (1998).

27. Li, B., Carey, M. & Workman, J. L. The Role of Chromatin during Transcription. *Cell* **128,** 707–719 (2007).

28. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).

29. Pietersen, A. M. & van Lohuizen, M. Stem cell regulation by polycomb repressors: postponing commitment. *Curr. Opin. Cell Biol.* **20,** 201–207 (2008).

30. Lawlor, E. R. & Thiele, C. J. Epigenetic changes in pediatric solid tumors: Promising new targets. *Clin. Cancer Res.* **18,** 2768–2779 (2012).

31. El-Serag, H. B., Sweet, S., Winchester, C. C. & Dent, J. Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* **63,** 871–880 (2014).

32. WHO. Global status report on noncommunicable diseases 2014. *World Health* **176** (2014).

33. Liao, L. M. *et al.* Nonsteroidal anti-inflammatory drug use reduces risk of adenocarcinomas of the esophagus and esophagogastric junction in a pooled analysis. *Gastroenterology* **142,** 442-452–3 (2012).

34. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31,** e15 (2003).

35. Gentleman, R. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).

36. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13,** 134 (2012).

37. Lin, L. *et al.* Activation of GATA binding protein 6 (GATA6) sustains oncogenic lineage-survival in esophageal adenocarcinoma. *Proc. Natl. Acad. Sci.* **109,** 4251–4256 (2012).

38. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25,** 402–8 (2001).

39. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501,** 506–11 (2013).

40. Gilfillan, G. D. *et al.* Limitations and possibilities of low cell number ChIP-seq. *BMC Genomics* **13,** 645 (2012).

41. Krook, M. A. *et al.* A bivalent promoter contributes to stress-induced plasticity of CXCR4 in Ewing sarcoma. *Oncotarget* **7,** 61775-788 (2016).

# CHAPTER 5

## Conclusion and Future Directions

The incidence of esophageal adenocarcinoma (EAC) and gastro-esophageal junction adenocarcinomas (GEJAC) has increased at an alarming rate of >600% over the last five decades[1,2]. The increased incidence of EAC is one of the greatest observed for any frequent cancer type, and it is highly associated with the increased epidemic of obesity and gastroesophageal reflux disease (GERD). Furthermore, a recent study predicts that by the year 2030, 1 in 100 Caucasians in the Netherlands and United Kingdom will be diagnosed with EAC in their lifetime[3]. The common risk factor, GERD, is now widely prevalent around the world. GERD is especially high in most developed countries/continents (North America, Australia/Oceania, Northern Europe)[4]. Other continents have started to observe a rise in GERD, mostly in Western and Southern Asia, as well as South America, making it a potential global concern[4].

Patients diagnosed with early stage EAC, have a favorable outcome when surgical resection or mucosal resection is performed[5]. Nevertheless, because the esophagus is highly vascularized, presentation beyond stage 1 is more likely to lead to metastasis, poor response to chemotherapeutics and radiation, as well as poor survival. Although many efforts have been made in implementing multimodal therapy, the 5-year survival rate for esophageal cancer remains low at 5-15%[6,7]. All these observations highlight the importance of developing new,

early diagnostic-detection tools and preventive strategies to improve overall patient survival.

**EAC and GEJAC**

EAC is often associated with histological evidence of BE as compared to GEJAC[8–10]. Nevertheless, we observed that these two tumor types share many molecular characteristics at the mutation, copy number, and transcriptome level[11,12]. Because EAC is associated with the presence of BE, this has led to many efforts in surveillance protocols focused on early cancer detection in these patients[13–17]. GEJAC presents without BE, therefore, these individuals are unlikely to be considered for routine screening.

Efforts have been made in improving the surveillance methods in the screening process of BE patients. One is the development of novel fluorescently-labeled peptides for endoscopic identification of early cancer in the esophagus[18]. In this thesis work, we have shown a strong similarity between Barrett's-associated EAC and GEJAC. These data and identification of cell surface markers that are shared between these cancer types, suggests that peptides could be developed that would identify cancers of both the lower esophagus and GEJ, regardless of the presence of Barrett's esophagus.

We have recently demonstrated that TGM2 is overexpressed and present on the cell surface of EAC cells[19]. In addition, in this thesis work, we confirmed tumor-specific overexpression for three genes (*CDH11, ICAM1,* and *CLDN3)* using qRT-PCR, and demonstrated protein localization specific to the cell surface of tumor cells by IHC. The future goal of these markers will be to apply a multiplexed panel of peptides using multispectral scanning fiber endoscope technology[20] to improve the success of histology-based screening programs for early EAC detection.

**BE dysplastic progression to EAC**

EAC is known to arise from the pre-malignant BE tissue. Moreover, patients with high-grade dysplasia (HGD) have a greater than 40% chance of developing EAC[21]. Therefore, HGD is the current gold standard for identifying "higher risk" of developing EAC in BE patients. In this thesis work, we have presented new insights into the transcriptome-based events that associate with increased risk of BE progression to EAC. We observed an increase in the splicing pathway, as well as an increase in transcript isoform diversity. This is likely to allow cells to increase their diversity of gene products and survive in a hostile environment. Concurrently, we observe a loss of protective mucin, and an increase in the ATM/DNA-damage response pathway (DDR). We further confirmed the activation of γH2AX by immunohistochemistry in HGD but not BE cells where mucin is lost and thus DNA damage is detected. Altogether, we surmised that increased spliceosome pathway during BE progression to EAC reflects both DNA damage and ATM-dependent signaling events in dysplastic BE cells. Further mechanistic studies to understand the role of a diverse isoform variance in HGD/EAC might elucidate potential targets and treatment strategies for reducing the progression of BE to cancer.

Detection of HGD poses significant challenges. Since HGD is associated with a high risk of progression, yet, there are problems in regards to sampling error and inter-observer variability in its diagnosis. We pose the idea of developing targeted fluorescent peptides that are over-expressed in the HGD when compared to non-dysplastic or low-grade dysplastic BE cells. In the current work, we have identified two markers, *C3* and *HLA-DRB5*, as over-expressed in HGD/EAC. Nevertheless, dysplastic BE cells and EACs are heterogeneous[22–24]. Not every person with HGD over-expresses the same cell surface protein-coding genes. Therefore, we created an independent cohort of 160 high-risk patients with pathologically-confirmed HGD from patients progressing to EAC. We have confirmed and pathologically-assessed the

percentage of non-dysplastic BE, LGD or HGD in this independent cohort. In the future, it will be important to validate the expression levels using techniques that are effective and more clinically applicable than quantitative RT-PCR. If successful, candidates could be useful for identifying high-risk patients containing HGD in BE independent of pathologic assessment alone.

Protein localization and expression levels are imperative for understanding the validity of these cell surfaces markers. Although the Human Protein Atlas[25,26] provides evidence that C3 and HLA-DRB5 are abundant in GI cancers such as stomach cancers and are localized abundantly in the plasma membrane. Further, it will be necessary to validate this membrane expression in HGD and EAC tissue. We have made a formalin-fixed, paraffin-embedded tissue microarrays (TMA) containing duplicate cores from 100 patients with HGD that developed EAC. Using antibody's against the candidate proteins of interest it will be important to assess the expression levels, the frequency of overexpression, and patterns of cell surface staining of these specific targets in the HGD/EAC.

Altogether these data highlight new insights into the molecular progression of BE dysplastic progression and identify potential biomarkers for detecting the highest risk epithelium HGD. In the future, we expect to develop better methods for early cancer detection and hopefully find new treatment modalities to improve patient survival with EAC.

**GSTT2 in the esophagus**

We have established that EAC primarily affects Caucasians with African Americans seemingly protected, despite both populations having a similar incidence of risk factors: obesity, and GERD[27,28]. In addition, African Americans have lower rates of esophagitis than Caucasians[29,28,30,31]. In this thesis, using transcriptional profiling, we analyzed a cohort of

African American and Caucasian normal esophageal tissues and identified glutathione S-transferase theta 2 (*GSTT2*) as being the top over-expressed gene in the African American esophagus. In addition, we have shown that GSTT2 protects esophageal cells against DNA damage under genotoxic stress. Acid reflux has been shown to generate reactive oxygen species and can induce DNA double-strand breaks[32]. In the esophageal field, there is a lack of cell lines derived from normal squamous esophagus tissue. Currently, HET1A is the only cell line which was derived from the normal squamous epithelium of an American of African descent and was subsequently immortalized. Not only cell line number poses a problem for reproducibility of experimental approaches, but also the immortalization process produces caveats as to the true nature of a normal cell line vs. a transformed cell line. To better understand the processes that affect the normal squamous tissue and its response to acid reflux leading to DNA damage, it is imperative that we develop better models to study the normal tissue of the human esophagus. Recently, Mau *et al.* developed a system where they use a dual SMAD inhibition protocol for the long term expansion of different epithelial like tissues (such as pseudostratified, stratified and glandular epithelium)[33]. More specifically, they showed that they can expand and culture tissue from mice esophagus and these cells retain their ability to differentiate into stratified esophageal epithelium using an air-liquid interface (ALI) culture[33]. Mau *et al.* approach provides an opportunity to expand and maintain the human-derived normal esophageal tissue in culture, and in turn, examine the functional characterization of GSTT2 role in protecting this tissue against acid-induce DNA damage.

Despite significant racial/population differences in EAC[34], our ability to functionally explore these differences, investigate mechanisms leading to disease, design preventative strategies in the human esophagus is nearly impossible owing to a lack of normal, healthy tissue

derived from squamous esophagus tissue. We have established the importance of GSTT2 in protecting cells against genotoxic stress. Nevertheless, the methods used in this thesis, such as the knock-down using targeted siRNA, may introduce off target effects. In the future, it will be important to perform mechanistic studies involving genetic gain- and loss-of-function approaches to test the hypothesis that GSTT2 provides protection against genotoxic stress in both AA and Cau genetic backgrounds. For example, we could generate stable cell lines from AA and Cau genetic backgrounds that harbor genetically-inducible *GSTT2,* using tet-inducible lentiviral transduction or, use Crispr/Cas9 technology to generate *GSTT2* knockout cell lines. These methods have the potential to elucidate GSST2 function in normal human tissue *in vitro* and *in vivo*.

Finally, we show that *GSTT2*/GSTT2 is inducible using indomethacin (an NSAID), at the mRNA but potentially less so at the protein level. It has been reported that GSTT2 increases in the esophagus of rats when treated with NSAID.[35] Non-steroidal anti-inflammatories (NSAIDs) have been independently shown that they can reduce the risk of EAC by >40% and induce GSTT2[36]. Our results, together with the above rat studies, makes us hypothesize, that the type of model is important for understanding GSTT2 induction ability. Even further, it is unclear, the mechanism by which NSAIDs induce GSTT2. We have identified a binding motif within the promoter of GSTT2, for the transcription factor PPAR-gamma (**Figure 5.1**). Interestingly, it is known that PPAR-gamma is induced by indomethacin[37]. Therefore, future assessment of the effects of treatment of NSAIDs and polyphenols in affecting GSTT2 levels, and investigating whether this effect is mediated through the activation of the transcription factor (TF), PPAR-gamma, will give key insights into the regulatory mechanism of GSTT2 in the setting of NSAID treatments.

Defining the mechanisms by which different human populations respond to injury/insult, will expand our understanding of EAC development and may provide avenues for new preventive treatments that might lower the incidence of EAC in high-risk populations

## Figures

| ID | Name | Motif ID | Gene ID | Family | Sequence | From | To | Dir | Score |
|---|---|---|---|---|---|---|---|---|---|
| T132748_1.02 | PPARD | M1894_1.02 | ENSG00000112033 | Nuclear receptor | GTGGGGCCTGATCCCCTATT | 176 | 195 | R | 12.265 |
| T132760_1.02 | * PPARG | M1894_1.02 | ENSG00000132170 | Nuclear receptor | GTGGGGCCTGATCCCCTATT | 176 | 195 | R | 12.265 |
| T132781_1.02 | PPARA | M1894_1.02 | ENSG00000186951 | Nuclear receptor | GTGGGGCCTGATCCCCTATT | 176 | 195 | R | 12.265 |

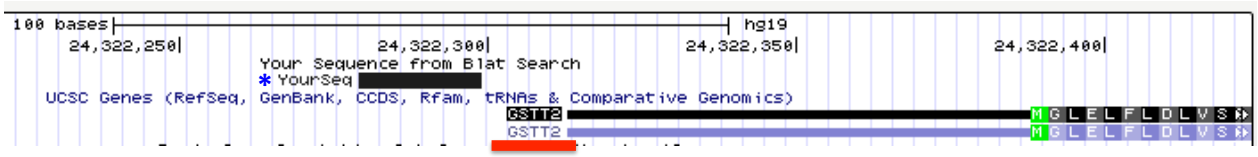



**Figure 5.1** PPAR-gamma DNA sequence binding motif (top panel) and BLAST results demonstrates region within the *GSTT2* promoter where consensus sequence for PPAR-gamma is found (black bar *).

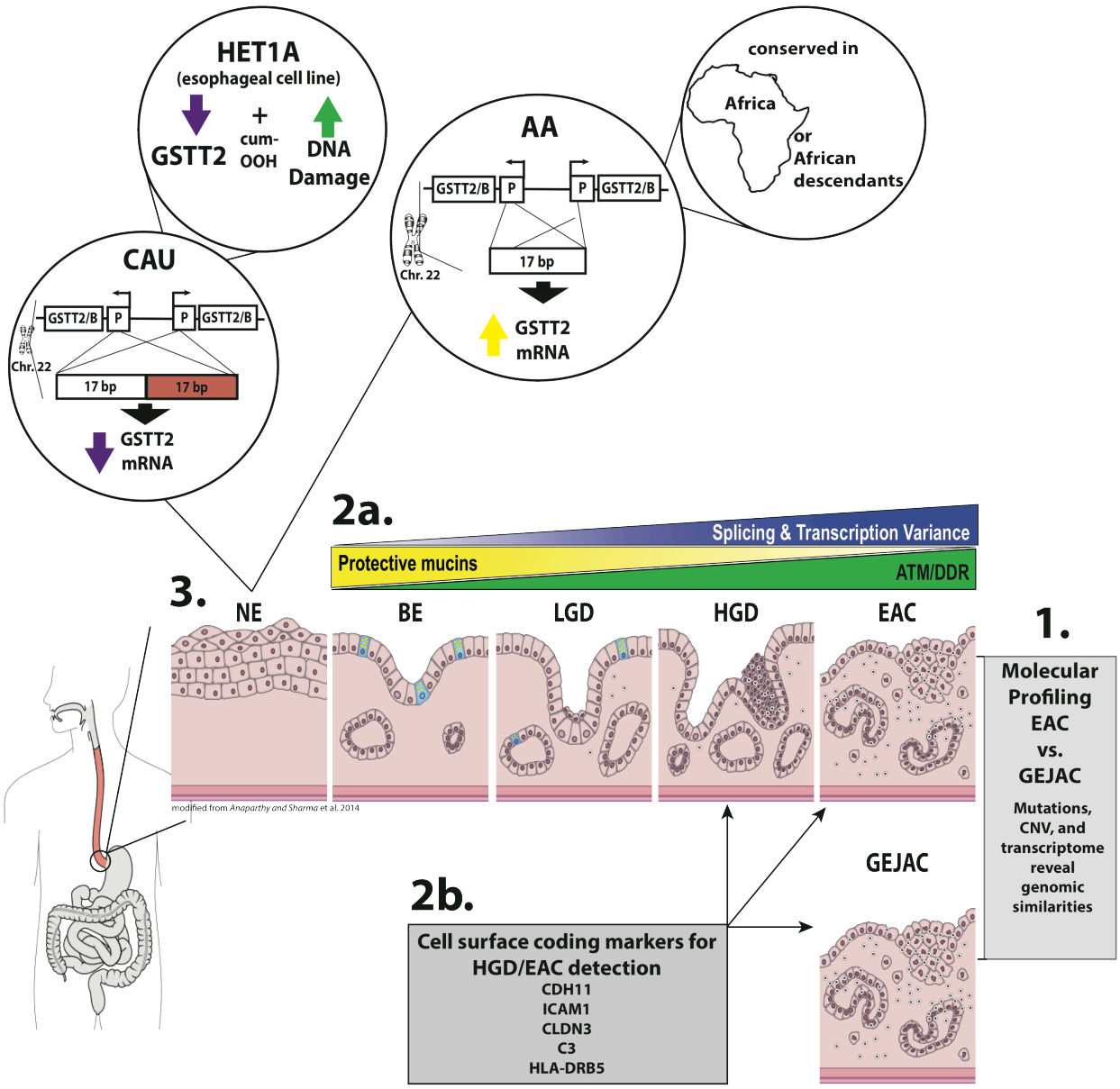



**Figure 5.2 Summary of thesis:** Molecular characterization of esophageal adenocarcinomas and factors influencing racial differences in incidence.

## References


1. Pohl, H. & Welch, H. G. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *J. Natl. Cancer Inst.* **97,** 142–6 (2005).
2. Vaughan, T. L. & Fitzgerald, R. C. Precision prevention of oesophageal adenocarcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **12,** 243–248 (2015).
3. Arnold, M., Laversanne, M., Brown, L. M., Devesa, S. S. & Bray, F. Predicting the Future Burden of Esophageal Cancer by Histological Subtype: International Trends in Incidence up to 2030. *Am J Gastroenterol* (2017).
4. Team, R. *et al.* World Gastroenterology Organisation Global Guidelines GERD Global Perspective on Gastroesophageal Reflux Disease. **51,** 467–478 (2017).
5. Anaparthy, R. & Sharma, P. Progression of Barrett oesophagus: role of endoscopic and histological predictors. *Nat. Rev. Gastroenterol.* **11**, 525–534 (2014).
6. Spechler, S. J. & Souza, R. F. Barrett's Esophagus. *N. Engl. J. Med.* **371,** 836–845 (2014).
7. Bresalier, R. S. Barrett's Esophagus and Esophageal Adenocarcinoma. *Annu. Rev. Med.* **60,** 221–231 (2009).
8. Orringer, M. B. *et al.* Two Thousand Transhiatal Esophagectomies. *Ann. Surg.* **246,** 363–374 (2007).
9. Clark, G. W. B. *et al.* Is Barrett's Metaplasia the source of Adenocarcinomas of the Cardia? *Arch Surg* **129,** 609–614 (1994).
10. Theisen, J. *et al.* Preoperative chemotherapy unmasks underlying Barrett's mucosa in patients with adenocarcinoma of the distal esophagus. *Surg. Endosc. Other Interv. Tech.* **16,** 671–673 (2002).
11. Ferrer-Torres, D. *et al.* Genomic similarity between gastroesophageal junction and esophageal Barrett's adenocarcinomas. *Oncotarget* **5,** (2016).
12. Kim, J. *et al.* Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).
13. van Sandick, J. W. *et al.* Impact of endoscopic biopsy surveillance of Barrett's oesophagus on pathological stage and clinical outcome of Barrett's carcinoma. *Gut* **43,** 216–222 (1998).
14. Incarbone, R., Bonavina, L., Saino, G., Bona, D. & Peracchia, A. Outcome of esophageal adenocarcinoma detected during endoscopic biopsy surveillance for Barrett's esophagus. *Surg. Endosc. Other Interv. Tech.* **16,** 263–266 (2002).
15. Ferguson, M. K. *et al.* Long-term survival after esophagectomy for Barrett's adenocarcinoma in endoscopically surveyed and nonsurveyed patients. *J. Gastrointest. Surg.* **6,** 29–36 (2002).
16. Fountoulakis, A. *et al.* Effect of surveillance of Barrett's oesophagus on the clinical outcome of oesophageal cancer. *Br. J. Surg.* **91,** 997–1003 (2004).
17. Corley, D. A., Levin, T. R., Habel, L. A., Weiss, N. S. & Buffler, P. A. Surveillance and survival in Barrett's adenocarcinomas: A population-based study. *Gastroenterology* **122,** 633–640 (2002).
18. Sturm, M., Joshi, B. & Lu, S. Targeted imaging of esophageal neoplasia with a fluorescently labeled peptide: first-in-human results. *Sci. Transl.* **5,** (2013).
19. Leicht, D. T. *et al.* TGM2: A Cell Surface Marker in Esophageal Adenocarcinomas. *J. Thorac. Oncol.* **9,** 872–881 (2014).

20. Miller, S. J. *et al.* Targeted detection of murine colonic dysplasia in vivo with flexible multispectral scanning fiber endoscopy. *J. Biomed. Opt.* **17,** 21103 (2012).
21. Dar, M. S., Goldblum, J. R., Rice, T. W. & Falk, G. W. Can extent of high grade dysplasia in Barrett's oesophagus predict the presence of adenocarcinoma at oesophagectomy ? 486–489 (2003).
22. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
23. Dulak, A. M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45,** 478–486 (2013).
24. Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47,** 1047–1055 (2015).
25. Uhlen, M. A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Mol. Cell. Proteomics* **4,** 1920–1932 (2005).
26. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science (80- ).* **347,** 1260419–1260419 (2015).
27. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics , 2017. **67,** 7–30 (2017).
28. El-Serag, H. B., Sweet, S., Winchester, C. C. & Dent, J. Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* **63,** 871–880 (2014).
29. El-Serag, H. B. *et al.* Gastroesophageal reflux among different racial groups in the United States. *Gastroenterology* **126,** 1692–1699 (2004).
30. Spechler, S. J., Jain, S. K., Tendler, D. A. & Parker, R. A. Racial differences in the frequency of symptoms and complications of gastro-oesophageal reflux disease. *Aliment. Pharmacol. Ther.* **16,** 1795–1800 (2002).
31. Sharma, P., Wani, S., Romero, Y., Johnson, D. & Hamilton, F. Racial and Geographic Issues in Gastroesophageal Reflux Disease. *Am. J. Gastroenterol.* **103,** 2669–2680 (2008).
32. Zhang, H. Y., Hormi-carver, K., Zhang, X., Spechler, S. J. & Souza, R. F. In Benign Barrett ' s Epithelial Cells , Acid Exposure Generates Reactive Oxygen Species That Cause DNA Double-Strand Breaks. 9083–9090 (2009).
33. Tata, P. R. *et al.* Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells Article Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells. *Stem Cell* **19,** 217–231 (2016).
34. Xie, S.-H., Rabbani, S., Petrick, J., Cook, M. B. & Largergren, J. Racial and Ethnic Disparities in the Incidence of Esophageal Cancer in the Unived States, 1992-2013. *Am. J. Epidemiol.* (2017).
35. Van Lieshout, E. M., Tiemessen, D. M., Roelofs, H. M. & Peters, W. H. Nonsteroidal anti-inflammatory drugs enhance glutathione S-transferase theta levels in rat colon. *Biochim. Biophys. Acta* **1381,** 305–11 (1998).
36. Liao, L. M. *et al.* Nonsteroidal anti-inflammatory drug use reduces risk of adenocarcinomas of the esophagus and esophagogastric junction in a pooled analysis. *Gastroenterology* **142,** 442-452–3 (2012).
37. Lenhard, J. M. Peroxisome Proliferator-activated Receptors alpha and gamma Are Activated by Indomethacin and Other Non-steroidal Anti-inflammatory Drugs. *J. Biol. Chem.* **272,** 3406–3410 (1997).

# APPENDICES

## Appendix 1: Authors Contributions

### CHAPTER 2

**Daysha Ferrer-Torres[1*], Derek J. Nancarrow[2*], Rork Kuick[3], Dafydd G. Thomas[4], Ernest Nadal[5], Jules Lin[2], Andrew C. Chang[2], Rishindra M. Reddy[2], Mark B. Orringer[2], Jeremy M. G. Taylor[3], Thomas D. Wang[6], David G. Beer[2]**

[1]Cancer Biology, Program in Biomedical Science, University of Michigan Medical School, Ann Arbor, Michigan, USA
[2]Section of Thoracic Surgery, Department of Surgery, University of Michigan Medical School, Ann Arbor, Michigan, USA
[3]Center for Cancer Biostatistics, Department of Biostatistics, School of Public Health, Ann Arbor, Michigan, USA
[4]Department of Pathology and Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA
[5]Medical Oncology Department, Catalan Institute of Oncology, Barcelona, Spain
[6]Department of Medicine and Department of Biomedical Engineering, University of Michigan Medical School, Ann Arbor, Michigan, USA
[*]These authors have contributed equally to this work

### CHAPTER 3

**[1]Daysha Ferrer-Torres, [1]Zhuwen Wang, [1]Hee-won Yoon, [1]Derek Nancarrow, [4]Hui Jiang, [3]Scott Owens, Anjan Saha[5], Dafydd Thomas, [2]Ted Lawrence, [1]Andrew Chang, [1, 2]David G. Beer**

[1]Department of Surgery, Thoracic Surgery, University of Michigan, Ann Arbor, MI 48109
[2]Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109
[3]Department of Pathology, University of Michigan, Ann Arbor, MI 48109
[4]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109
[5] Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109

### CHAPTER 4

**Daysha Ferrer-Torres[1], Derek Nancarrow[1], Hannah Steinberg[1], Rork Kuick[5], Melanie A. Krook[3], Jules Lin[1], Andrew C. Chang[1], Rishindra M. Reddy[1], Mark B. Orringer[1], Ryan**

**Mills[4], Marcia I. Canto[6], Joel H. Rubenstein[2], Amitabh Chak[7], Thomas Wang[2], David G. Beer[1]**

[1]Department of Thoracic Surgery, University of Michigan, Ann Arbor, MI 48109
[2]Departments of Internal Medicine, University of Michigan, Ann Arbor 48109
[3]Departments of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor 48109
[4]Departments of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109
[5]Department of Biostatistics, Center for Cancer Biostatistics, School of Public Health, Ann Arbor, MI 48109
[6]Department of Medicine, Gastroenterology and Hepatology, Johns Hopkins University, Baltimore, MD 21287
[7]Department of Medicine, Gastroenterology, Case Western Reserve University, Cleveland, OH 44106

# Appendix 2: List of oligonucleotide primers used

| Chapter | Type | Name | Figure | Primer Name | Sequence |
|---|---|---|---|---|---|
| 2 | cDNA | *CDH11* | Figure 2.12; 2.13 | *CDH11*-cDNA | GCACGAGACCTATCATGCCA |
| 2 | cDNA | *CDH12* | Figure 2.12; 2.14 | *CDH12*-cDNA | CTGTCTGTGCTTCCACCGAA |
| 2 | cDNA | *ICAM1* | Figure 2.12; 2.15 | *ICAM1*-cDNA | GTA TGAACTGAGCAATGTGCAAG |
| 2 | cDNA | *ICAM2* | Figure 2.12; 2.16 | *ICAM2*-cDNA | GTTCCACCCG TTCTGGAGTC |
| 2 | cDNA | *CLDN3* | Figure 2.12; 2.17 | *CLDN3*-cDNA | TCGGCCAACACCA TTATCCG |
| 2 | cDNA | *CLDN4* | Figure 2.12; 2.18 | *CLDN4*-cDNA | GTACTTCTTCTCGCGTGGGG |
| 2 | cDNA | *ZNF217* | Figure 2.8 | *ZNF217*-cDNA | CTCCGGGCCACTTTACACTT |
| 2 | cDNA | *ZNF218* | Figure 2.8 | *ZNF218*-cDNA | TCTCT TTTGTGCCATGCTGTT |
| 4 | cDNA | *GSTT2/2B* cDNA | Figure 4.2; 4.3; 4.5; 4.9; 4.10; 4.13; 4.15 | GSTT2 (E1-E2) | TGTTTCTTGACCTGGTGTCCC |
| 4 | cDNA | *GSTT2/2B* cDNA | Figure 4.2; 4.3; 4.5; 4.9; 4.10; 4.13; 4.15 | GSTT2 (E1-E2) | CCAGGCTGTTGATCTGCAAG |
| 4 | cDNA | *GSTT1* | | *GSTT1-c*DNA | CTGGAGTTTGCTGACTCCCTC |
| 4 | cDNA | *GSTT1* | | *GSTT1*-cDNA | GCTCGAAGGGAATGTCGTTCT |
| 4 | DNA | *GSTT2/2B*_17b p Promoter | Figure 4.3; 4.4; 4.9; 4.14 | GSTT2ProAlleleF | CCCATCCTGTGCACGAAGTG |
| 4 | DNA | *GSTT2/2B*_17b p Promoter | Figure 4.3; 4.4; 4.9; 4.15 | GSTT2ProAlleleR | GCCCTGACCCAGAAACGACTG |
| 4 | DNA | *GSTT2*-37kb Deletion | Figure 4.5; 4.9; 4.14 | GSTT2B-6858 | CACTCAACACAGTAGCCTCATCGTG |
| 4 | DNA | *GSTT2*-37kb Deletion | Figure 4.5; 4.9; 4.14 | GSTT2B-6857 | TGCCTCCCCTGCCTTATTTC |
| 4 | DNA | *GSTT2*-37kb Deletion | Figure 4.5; 4.9; 4.14 | GSTT2B-2B | CCTTCTGAAATGGAGCCTTTG |
| 4 | DNA | *GSTT2/2B*-Promoter (ChIP) | Figure 4.15 | GSTT2_Prom_F71 | AACGAACCCTCAGATGTCCG |
| 4 | DNA | *GSTT2/2B*-Promoter (ChIP) | Figure 4.15 | GSTT2_Prom_R71 | CCCTGACCCAGAAACGACTG |
| 4 | DNA | *GSTT2/2B*-Promoter (ChIP) | Figure 4.15 | GSTT2_PromF_116 | CATCCCACTGGGTGAAACTCT |
| 4 | DNA | *GSTT2/2B*-Promoter (ChIP) | Figure 4.15 | GSTT2_PromR_116 | GCACGGACATCTGAGGGTTC |
| 4 | DNA | *GSTT2/2B*-Promoter (ChIP) | Figure 4.15 | GSTT2_PromF_300 | CCCATCCTGTGCACGAAGTG |
| 4 | DNA | *GSTT2/2B*-Promoter (ChIP) | Figure 4.15 | GSTT2_PromR_300 | GCCCTGACCCAGAAACGACTG |