STUDY NUMBER: CASE 3207

STUDY TITLE: Genetic and Environmental Determinants of Barrett Esophagus and Esophageal Adenocarcinoma

PRINCIPAL INVESTIGATORS:

Amitabh Chak, M.D.
Department of Gastroenterology
University Hospitals Case Medical Center
Case Western Reserve University
11100 Euclid Avenue
Cleveland, OH 44106
(216) 844-6172

CO- INVESTIGATORS: Julian Abrams, M.D., M.S.
Columbia University Medical Center

Marcia Irene Canto, M.D.
Departments of Gastroenterology & Hepatology
The Johns Hopkins Hospital
Blalock 943, 600 North Wolfe Street
Baltimore, Maryland 21287

Dawn Dawson, MD
Department of Pathology
University Hospitals Case Medical Center
Case Western Reserve University

John Dumot, M.D
Department of Gastroenterology
University Hospitals Case Medical Center
Ahuja Medical Center
11100 Euclid Avenue
Cleveland, OH 44106

Gary Falk, M.D. M.S
Hospital of University of Pennsylvania
Division of Gastroenterology
3 Ravdin / 3400 Spruce Street
Philadelphia, PA 19104

William Grady, M.D.
Department of Gastroenterology
Fred Hutchinson Cancer Research Center

University of Washington Medical School
VA Puget Sound Health Care System
Seattle, WA 98109

Katarina B. Greer, MD
Department of Internal Medicine/Division of Gastroenterology
University Hospitals Case Medical Center
Case Western Reserve University

Margaret Kinnard, M.D.
Gastroenterology Section
Lois Stokes VA Medical Center (Wade Park)
Cleveland, Ohio

Rom Leidner, M.D
Case Western Reserve University
10900 Euclid Avenue WRB 3-126
Cleveland, Ohio 44106

Li Li, MD/PhD
Department of Family Medicine
University Hospitals Case Medical Center
Case Western Reserve University

Sanford Markowitz, M.D.
UHCMC
Wolstein Research Bldg.
2103 Cornell Road
Cleveland, OH 44106

Sumeet Mittal, M.D.
Department of Surgery
Creighton University Medical Center
601 North 30th Street, Suite 3700
Omaha, NE 68131-2197

Jean S Wang, MD, PhD
Washington University at St. Louis
St. Louis, MO

Joseph Willis, M.D.
Department of Pathology
University Hospitals Case Medical Center
Case Western Reserve University
Cleveland, OH 44106

G.A. Prasad , M.D.
Department of Gastroenterology
Alfred M430
200 First Street, SW
Rochester, MN 55905

Nicholas Shaheen, M.D.
Division of Gastroenterology and Hepatology
University of North Carolina – Chapel Hill
4102 Bioinformatics Bldg., CB#7080
Chapel Hill, NC

John Vargo, M.D.
Cleveland Clinic Foundation
Taussig Cancer Center, A31
9500 Euclid Avenue
Cleveland, OH 44195

STATISTICIANS: Robert Elston, Ph.D.
Jill Barnholtz-Sloan, PhD
(216) 368-5630

SPONSOR: NIDDK/NCI

CLINICAL FACILITY: University Hospitals Case Medical Center, Hospital University of Pennsylvania ,Cleveland Clinic, Lois Stokes VA Medical Center (Wade Park), Fred Hutchinson Cancer Research Center, Washington University, Johns Hopkins Hospital, Mayo Clinic, University of North Carolina at Chapel Hill, Washington University at St Louis and The Creighton University Hospital

APPROVALS: Protocol Review and Monitoring Committee: Approved
05/20/2010 to 05/12/2011

UHC IRB: Approved through 06/14/2011

Continuing Review CASE 3207 V2 05/06/2010 approved through 05/12/2011

STUDY COORDINATORS:

Beth Bednarchik, R.N.
Department of Gastroenterology
University Hospitals Case Medical Center
Case Western Reserve University
11100 Euclid Avenue
Cleveland, OH 44106-5065

Wendy Brock, R.N.
Department of Gastroenterology
University Hospitals Case Medical Center
Case Western Reserve University
11100 Euclid Avenue
Cleveland, OH 44106-5065

Ellen Neumeister
Department of Gastroenterology
University Hospitals Case Medical Center
Case Western Reserve University
11100 Euclid Avenue
Cleveland, Ohio 44106-5065

Anna Haas, R.N.
Gastroenterology Section
Lois Stokes VA Medical Center (Wade Park)
10701 East Boulevard 111E (W)
Cleveland, OH 44106

Antia Compan, R.N.
Gastroenterology Section
Lois Stokes VA Medical Center (Wade Park)
10701 East Boulevard 111E (W)
Cleveland, OH 44106

Kaysey Orlowski
Clinical Trials Support Office
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue N, LM-200
Seattle, WA 98109

Laura Veri
Department of Gastroenterology
University Hospitals Case Medical Center
Case Western Reserve University
11100 Euclid Avenue
Cleveland, OH 44106-5065

Linda Chessler
Mercy Medical Center Gastroenterology
4360 Fulton Drive NW
Practice GHS, Suite B

Canton, OH 44718

Beth Gifford, RN
Department of Surgery
Creighton University
601 N. 30th Street Suite 3700
Omaha, NE 68131

Eric Murphy
Department of Surgery
Creighton University
601 N 30th Street Suite 3700
Omaha, NE 68131

Verna Scheeler, MLA
Project Manager, Clinical Research
The Johns Hopkins Hospital
Gastroenterology & Hepatology
Blalock 406A
600 North Wolfe Street
Baltimore, MD 21205

Kelly Dunagan
Mayo Clinic
200 First Street SW
Rochester, MN 55905

Amy Pruitt
Div. of Gastroenterology and Hepatology
4102 Bioinformatics Bldg.
CB #7080

Courtney Ziefle
Div. of Gastroenterology and Hepatology
4102 Bioinformatics Bldg.
CB #7080

Julie Starr
Hospital of University of Pennsylvania
3400 Spruce Street
Gound Floor/Centres Bldg
Philadelphia, PA 19104

TABLE OF CONTENTS

1.0 BACKGROUND AND RATIONALE

Whereas the rate of many common cancers has declined, the incidence of esophageal adenocarcinoma (EAC) has increased greater than six-fold over the past 3 decades.[1] Furthermore, the outcome of this cancer remains poor, accounting for over 1 in 50 adult male cancer-related deaths.[2] Although Barrett's esophagus (BE), a precursor of EAC can be easily recognized at endoscopy, current clinical strategies of endoscopic screening and surveillance based on the close association of BE with chronic GERD in adults are woefully inadequate. Nearly 40% of patients who develop EAC have no preceding symptoms of GERD[3, 4] and less than 5% of EAC's are diagnosed in patients whose BE was recognized prior to cancer diagnosis.[5] In recognition of the burden of morbidity and mortality caused by EAC and the clinical need to develop more effective screening, surveillance, and ablative strategies for BE, this translational research proposal aims to develop new molecular approaches for improving the detection, prognostication, and treatment of BE, with the ultimate goal of reducing the mortality associated with this cancer.

There has been a startling rise in the incidence of esophageal adenocarcinoma (EAC) in the United States during the past three decades, especially in white males where the incidence has gone up more than six fold.[1, 6-8] National statistics estimate 16,640 new cases of esophageal cancer, mainly adenocarcinomas, in 2010.[9] Although there have been advances in chemotherapy, radiotherapy, and surgical therapy, the prognosis remains poor with a dismally low five year survival of 16%.[9] Most EACs originate in Barrett's epithelium, a pre-malignant condition in which normal squamous epithelium is replaced by metaplastic specialized intestinal type columnar epithelium.[10-17] Despite endoscopic screening and surveillance efforts,[18] less than 5% of EAC are diagnosed in individuals with previously detected BE.[5] Gastroesophageal reflux disease (GERD) is closely associated with BE and is present in about 10% of patients who have endoscopy for GERD[16, 18-20] but even if esophago-gastro-duodenoscopy (EGD) was recommended in all adults with GERD, our study and others show that this strategy would only be partially effective. Nearly 40% of adenocarcinomas occur in individuals without GERD symptoms.[3, 4] Less than 0.5% of patients with BE progress annually to cancer, yet surveillance consists of frequent endoscopies with random biopsies to try and detect high grade dysplasia and early cancer.[21] Endoscopic surveillance is costly and there are no randomized controlled trials to show that it is efficacious. Our research group found that family history was a risk factor for BE or EAC[22] and demonstrated that familial aggregation of BE and EAC was present in 7% of cases.[23] Furthermore, our segregation analysis discovered that Familial Barrett's Esophagus (FBE) is consistent with autosomal dominant transmission of susceptibility allele(s).[24] The characterization of inheritable risk factors would lead to the development of more effective screening and surveillance programs for BE. Moreover, the identification of germ-line mutations that predispose selected individuals to these diseases would lead to an improved understanding of disease pathogenesis. Our multi-disciplinary researchers now propose genetic and physiological research on BE and EAC that will improve clinical management by identifying susceptibility genes for BE and EAC, leading to biological insights into the esophagitis-metaplasia-dysplasia-cancer progression, providing targets for chemopreventive and therapeutic drug development, and identifying markers for early detection, prognostication, and therapeutic efficacy.

The development of BE in only a subset of those with chronic GERD and its predominance in white males suggests a susceptibility to the development of metaplasia in selected individuals.

Familial aggregation of BE and associated cancers has been reported, which we and others have termed familial Barrett esophagus (FBE)[25-31]. Our preliminary study discovered familial aggregation in over one fifth of patients presenting with these diseases[22]. Further endoscopic screening has enabled the characterization of extended FBE families. These families strongly suggest an autosomal dominant mode of inheritance.

We believe that BE and EAC are complex diseases that develop through various combinations of inherited germ-line and environmentally induced somatic mutations. Prior reports have sometimes included short segment BE (SSBE) as well as esophagogastric junction cancers (EGJAC) as part of a broad familial trait. One of our ***Central Hypotheses*** is that ***FBE represents a genetic susceptibility to the development of BE and EAC, including SSBE and EGJAC***. Furthermore, the term "isolated disease" is used to designate affected patients without a family history, recognizing that their disease cannot be classified as "sporadic" until the mutated gene(s) that cause FBE is identified.

> Compelling evidence also indicates that obesity is a risk factor for many cancers. The dramatic increase in the incidence of EAC has been attributed, at least partially, to a pandemic of obesity.[5,6] Both BE and EAC are strongly associated with gastroesophageal reflux disease (GERD).[15-18] Obesity is also an independent risk factor.[5, 6, 19-26] Obesity is clearly linked with GERD [26-28] and contributes to BE through a GERD dependent mechanism. However, the independent association of visceral adiposity with BE and EAC[19, 20,24-26] strongly indicates a second reflux independent pathway. This association of obesity with BE and EAC could be related either to a dietary factor, genetic susceptibility linked to obesity, or alterations in metabolic mediators. In general, the association between obesity and disease states may be mediated by dietary factors, physical activity, genetic susceptibility, or metabolic alterations related to the obesity state itself. Molecular mechanisms proposed to mediate obesity associated carcinogenesis include oxidative stress/DNA damage, non-inflammatory changes in immune function, inflammatory mediators, hormones and growth factors, and metabolic detoxification factors. The mechanisms by which obesity (or the metabolic syndrome) or specific dietary patterns might lead to the development of BE and EAC need to be explored. The Western diet (high carbohydrate and high animal fat) as well as obesity likely lead to insulin resistance, hyperinsulinemia, and increased levels of insulin like growth factors (IGF).[29-34] Chronically elevated insulin and IGF-1 have been implicated as one mechanism for the tumorigenesis of many obesity associated cancers.[35-44] One ***Hypothesis*** is that ***sustained hyperinsulinemia and high levels of IGF-1 in susceptible individuals contribute to the development of BE and/or its subsequent progression to EAC. Furthermore, genes involved in the regulation of the IGF-1 signaling pathway are also known to be silenced by aberrant DNA methylation and may be selected to be epigenetically silenced in BE and EAC that arise in the setting of high levels of IGF-1 or hyperinsulinemia. Alterations in other hormones such as leptin and adiponectin have also been implicated in obesity associated carcinogenesis.***

This study represents ongoing growing collaborations between investigators at eight institutions who comprise our Familial Barrett Esophagus Consortium (FBEC). Dr. Chak at University Hospitals Case Medical Center will provide overall direction for the familial recruitment and endoscopic screening aspects of this study. Drs. Chak (University Hospitals Case Medical Center), Vargo (Cleveland Clinic), Grady (Fred Hutch Cancer Research Center and University of

Washington), Falk (Hospital of University of Pennsylvania), Canto (Johns Hopkins), Wang (Washington University at St Louis), Prasad (Mayo Clinic), Mittal (Creighton University), Abrams (Columbia Presbyterian), Shaheen (University of North Carolina), and Kinnard (Louis Stokes VA) will direct patient recruitment, biobanking and family ascertainment at each individual site. Dr. Willis (University Hospitals Case Medical Center) will be responsible for supervising review of pathology specimens. Drs. Barnholtz-Sloan and Elston, genetic epidemiologists, will provide guidance in study design and statistical analysis. Drs. Markowitz, Guda, and Leidner at CWRU, Grady and Kaz at FHCRC, and Orlando and Shaheen at UNC will perform the necessary molecular and genetic laboratory studies. In total, this multidisciplinary team represents the range of expertise required for the success of this project.

2.0 OBJECTIVES

Aim 1: To identify susceptibility genes for BE by complementary linkage approaches and whole exome sequencing approaches of genomic DNA from affected distant relative pairs and cancer DNA from FBE pedigrees and subsequently develop models for predicting the presence of BE in family members.

Aim 2: To define epigenetic changes associated with clinical characteristics in BE patients and develop DNA (DNA) testing as a noninvasive method for detecting BE, dysplastic BE, and early EAC.

Aim 3: To identify transcriptional alterations and miR signatures associated with increased risk of BE progressing to EAC by deep sequencing the transcriptome as well as the microRNA fraction isolated from tissue and to develop serological and/or in situ methods for detection and prognostication.

Aim 4: To define physiological and molecular changes in the neosquamous epithelium post-ablation of BE in order to develop more effective therapeutic strategies.

Aim 5: Foster pilot projects at our multiple institutions and develop activities across other centers as well as build on existing interactions with other research consortia such as NCI's Chemoprevention Network (CPN) and the Early Detection and Research Network (EDRN).

Aim 6: Provide Investigators support through core resources with state-of-the-art Bioinformatics, Statistical Methods, and Administrative Services.

3.0 PATIENT SELECTION

3.1 Inclusion Criteria for Probands

3.1.1 Persons diagnosed with BE and/or EAC (or EGJAC)

3.1.2 Male or female over the age of 18 years and able to provide informed consent. Informed consent will be accepted from next of kin if the participant is deceased.

3.2 Inclusion Criteria for Family Members

3.2.1 Relative of a person diagnosed with BE and/or EAC

3.2.2 Male or female over the age of 18 years and able to provide informed consent. Informed consent will be accepted from next of kin if the participant is deceased.

3.2.3 The proband or another relative enrolled in the study gave permission to contact this family member

3.3 Inclusion Criteria for Controls

Male or female, age 18 or older. EGD does not show SSBE, BE, EAC, or EGJAC. Ability to give informed consent. Willing to donate 30 cc of blood and endoscopic mucosal biopsies for research. Colonoscopy performed within three years; pathology negative for adenoma (only for those in whom stool is being collected). No prior surgery for treatment of obesity. Not enrolled in a chemoprevention trial.

Permission will be taken from the endoscopist who performed the endoscopy to contact the patient. Potential study subjects may also be contacted by letter and then phone. Informed consent and a signed HIPAA form will be obtained by a study nurse or study investigator. As much as possible, control subjects will be frequency matched in blocks of 10 for GERD indication, gender, race, and age (+ 5 yrs) with recruited case subjects.

3.4 Exclusion Criterion for Probands

3.4.1 Endoscopic and/or pathologic characteristics that are not consistent with a diagnosis of SSBE, BE, EGJAC, and/or EAC (Table 1).

3.5 Exclusion Criterion for Family Members

3.5.1 The proband's endoscopic and/or pathologic characteristics are not consistent with a diagnosis of SSBE, BE, EGJAC, and/or EAC (Table 1).

Table 1 Criteria for ascertaining diagnosis

Diagnosis	Findings noted on EGD Report	Findings noted at pathology
BE	≥ 3 cm involvement of tubular esophagus on at least one report	Specialized intestinal metaplasia or goblet cells noted
SSBE	< 3 cm of salmon-colored mucosa or length not reported. Biopsies specifically taken from the esophagus.	Same as above
EAC	Mass predominantly involving tubular esophagus	Invasive adenocarcinoma
EGJAC	Mass predominantly involving esophagogastric junction or gastric cardia and the esophagus	Invasive adenocarcinoma

If the endoscopy report does not specify the affected length of suspected Barrett's epithelium or does not clearly state that the biopsies that are subsequently found to have intestinal metaplasia were obtained from the tubular esophagus, the proband will be excluded. Biopsies obtained from the gastroesophageal junction will not be considered.

4.0 REGISTRATION PROCEDURES

To enter eligible patients on study, contact the study coordinator at the Cancer Center Clinical Trials Unit. If applicable, all patients enrolled on study will be entered into the Oncore database. A parallel database will be maintained in Labmatrix, which is a secure protected environment designed for human research. This database is licensed to the PI

5.0 DATA SAFETY MONITORING PLAN

This multicenter investigation aims to identify susceptibility genes associated with FBE, molecular risk factors associated with BE, and biomarkers for diagnosing BE, predicting its progression to EAC and predicting response to therapy through collection of blood, biospecimens, and tissue samples from persons affected with BE and/or EAC, family members and unaffected controls. Detailed questions about personal, diet, and family history will be collected through personal interviews and a questionnaire. Furthermore, the project will perform endoscopic screening of family members of selected patients with BE and EAC to determine the risk of BE; ultrathin unsedated endoscopies (UUE) will be performed. The recruitment, upper endoscopic screening protocols, collection of biopsecimens, and phlebotomy pose minimal risk to participants. Because of this low risk status, the data safety monitoring (DSM) plan for this trial focuses on close monitoring by the principal investigators in conjunction with a safety officer, along with prompt reporting of excessive adverse events and any serious adverse events to the NIH and to the IRB at the respective institution.

Injury/illness survey and adverse events form have been devised to meet the goals of this plan. Safety reports will be sent to Amitabh Chak, M.D. and Jeffry Katz, M.D. (safety officer). Wendy Brock, R.N. will be responsible for assembling the data from all participating centers and producing these reports, as well as assuring that all parties obtain copies of these reports.

The frequency of data review for this study differs according to the type of data and can be summarized in the following table:

Table 2 Data Safety Monitoring Plan

Data type	Frequency of review
Subject accrual (adherence to protocol regarding demographics, inclusion/exclusion)	Quarterly for 2 years
Adverse event rates (injuries) related to endoscopic procedures	Quarterly for 2 years

Qualifications and responsibilities of the Safety Officer

The safety officer for this trial will be Jeffry Katz, M.D. Dr. Katz is board certified in gastroenterology and has experience running epidemiological studies. He understands the types and severity of injuries commonly experienced as a result of upper endoscopy. As Safety Officer, Dr. Katz will review the reports sent by the study coordinator (at the frequency outlined above) and determine whether there is any corrective action, trigger of an ad hoc review, or stopping rule violation that should be communicated to the study investigators, the Case Comprehensive Cancer Center IRB, the Cleveland Clinic IRB, the University Hospitals Case Medical Center IRB, and the NIDDK.

Measurement and reporting of subject accrual, adherence to inclusion/exclusion criteria

Review of the rate of subject accrual, adherence to inclusion/exclusion criteria will occur every 3 months during the study. Because BE and EAC predominantly affect Caucasian males we expect to see a lower recruitment of children, women, and other ethnic groups as indicated in the proposal.

Measurement and reporting of adverse events

We plan to collect injury data from the endoscopic procedures quarterly. We plan to present unblended adverse events data to the principal investigators and the safety officer throughout this trial. Any adverse event will also be reported to the Quality Assurance Committee of the institution where it occurs and the IRB. Adverse event forms that meet the goals of this data safety monitoring plan already exist and will be used by the study staff to report injuries or other adverse events caused by endoscopic procedures. The risks of routine diagnostic upper endoscopy are minimal and include a risk of bleeding, aspiration, bleeding, and adverse reaction to sedative medications. The complication rate for routine sedated upper endoscopic diagnostic procedures is 0.1 to 0.5%[30]. The complication rate for unsedated upper endoscopy (UUE) will be lower (No complications noted in over 200 cases performed by the PI). Any adverse event rate over 0.5 % in 12 months will be reported to all institutional IRBs and the NIH.

Stopping rules

In this minimal risk trial it is unlikely that excess adverse events will occur and require stopping the trial. However, as outlined, we will monitor injury rates in all participants and the safety officer, together with the study investigators, will alert the IRB and the NIH if a larger than reasonably expected injury rate (> 0.5%) should occur in the endoscopy group. It is exceedingly unlikely that any new information will become available during this trial that would necessitate stopping the trial.

6.0 MEASUREMENT OF EFFECT

Not applicable

7.0 STUDY PARAMETERS

Subjects who were previously enrolled in one of the two related protocols at the Cleveland Clinic or University Hospitals Case Medical Center will be notified that this research is being conducted under a new IRB protocol at the newly developed Case Comprehensive Cancer Center IRB. Subjects will be notified of this change by letter (Appendix A). If Dr. Chak or Dr. Vargo sees a subject again in clinic, he or she will be asked to provide informed consent as outlined in the form approved by the Case Comprehensive Cancer Center IRB.

Potential new subjects will be identified in five ways:

1. Patients scheduled for endoscopy or surgery for BE and/or EAC will be approached by an investigator of the study and/or their designee in person or by telephone.
2. Review of the Cleveland Clinic Division of Pathology Tissues/Body Fluids Database/Registry/Repository (CCF IRB#6001), review of the UHCMC pathology records,

or Cancer Database/Registry by an investigator of the study and/or their designee. The patient will be contacted by an investigator of the study and/or their designee by phone.
3. Referrals from collaborating physicians: collaborating physicians will introduce the study to eligible patients and receive verbal consent to be contacted, which they will document and send to an investigator of this study. This patient will then be contacted by an investigator of this study and/or their designee in person or by telephone.
4. Self-referral (through telephone or email contact) upon reading an advertisement of the study (Appendices B, C, D, E, F, G). The participant, once eligibility is determined, will then be provided with a study packet containing the necessary documents for participation and a recruitment letter (Appendix H).
5. Self-referral (through telephone, email, or family member) upon discussions with a participating family member. The participant, once eligibility is determined, will then be provided with a study packet containing the necessary documents for participation (Appendix I).

Upon initial contact, if the subject's endoscopy and/or pathology report is not available, consent (release of information) will be obtained from the subject to obtain this information from medical records. Investigators of the study will then determine if the potential subject meets eligibility criteria. Once eligibility has been established the study will be explained in detail by one of the investigators and/or their designee. Specifically, it will be explained that participation requires the following:

1. Written informed consent to participate in the study. For deceased subjects, the next of kin or other authorized representative will give informed consent.
2. Possible donation of 20-40ml of venous blood from all living subjects and/or donation of biofluids or tissue collected during routine upper endoscopy with biopsy.
3. Signature of "Authorization for Release of Medical Information" form (Appendix G) by affected subjects or their authorized representative to allow researchers to obtain surgical reports, pathology reports, chart summaries, and/or paraffin-embedded tumor blocks and matching H & E slides related to the diagnosis of BE and/or EAC.
4. Detailed interview with an investigator of the study or their designee in which a medical and family history will be obtained in the form of a family pedigree.
5. Completion of the Familial Barrett's Esophagus Questionnaire (FBEQ) that asks questions about reflux symptoms, the Food Frequency Questionnaire, and the Meat Preparation Questionnaire.

This FBEQ obtains details regarding GERD symptoms, environmental exposure history, prior cancer history, family history, height, weight, and prior upper endoscopy. The gastroesophageal reflux questionnaire (GERQ) is a validated measure of reflux symptoms and environmental exposure history developed by the Mayo Clinic[32]. It has served as the foundation for the GERD symptoms section of the study questionnaire. The FBEQ also incorporates detailed questions regarding medication usage, duration of obesity, and family history of BE, EAC, and other cancers. Subjects are asked to speak to their relatives for permission to be contacted. They are then asked to list names, addresses, and phone numbers for all first-degree relatives who gave them permission to be contacted.

A cover letter explaining the study (Appendix I), a consent form, a FBEQ, a release of medical information form, and a self-addressed, post pre-paid return envelope will then be mailed to the

relatives. After one month, those family members who have not responded to the questionnaire will be contacted by phone.

Additionally, family members, for whom the proband has provided us with consent to contact, are eligible for an upper endoscopy. Family members who have had a normal endoscopy within the past 3 years or who have had an endoscopy after the age of 50 years, will not be asked to undergo another endoscopy. Although the procedure is very safe, we feel a repeat invasive test cannot be ethically justified in siblings over age 50 who have had prior endoscopy because BE is unlikely to develop after this age[19]. Family members who meet conventional criteria [i.e. have chronic GERD symptoms greater than 5 years, GERD symptoms that persist despite therapy or alarm symptoms (dysphagia, weight loss, anemia)] will be referred for a standard sedated upper endoscopy, which will be billed to the participant's insurance (Appendix K). Eligible family members of FBE families with 3 or more affected family members may be offered an unsedated ultrathin esophagoscopy (UUE) (Appendix L). This screening esophagoscopy will be provided free of charge at the Department of Gastroenterology at University Hospitals Case Medical Center under the direction of Amitabh Chak, M.D. If mucosal changes consistent with BE are identified on the screening endoscopy then the participant will be referred for a standard sedated upper endoscopy with biopsy, which will be billed to the patient's insurance.

Participants who are eligible for and interested in the UUE will be asked to travel to the Cleveland area for this procedure, the cost of the procedure will be covered by the grant. Participants who do not qualify for the UUE or those who do not want to travel to Cleveland will be sent all materials needed to participate in this research study. These participants will have the option of being in the study without having an endoscopy. Those participants who meet clinical criteria for standard upper endoscopy can elect to have the procedure performed by the physician of their choice (outside of this research study) and have the results released to the study investigators for inclusion in the research. If standard upper endoscopy is indicated, the cost of this clinical procedure will be billed to the patient or the patient's insurance. The cost of travel to Cleveland will not be covered.

At the time of recruitment, obesity measurements such as height, weight, waist and hip circumference will also be recorded. Glucose, insulin, serum levels of IGF-1, IGF binding proteins, and other hormones as needed will be measured from serum collected in the fasting state. Insulin resistance will be estimated by calculating HOMA-IR index which is defined as fasting insulin (μU/ml) times fasting glucose (mmol/l) divided by 22.5. HOMA-IR has been validated in normoglycemic subjects against insulin sensitivity measured directly from the euglycemic-hyperinsulinemic clamp and has been widely used in epidemiological studies[116]. During EGD, four quadrant biopsies will be obtained with the Microvasive (Boston Scientific) "jumbo" spiked-biopsy (Max Capacity) forceps. Biopsy specimens for routine histology will be obtained at 2-cm intervals along the entire length of the Barrett's epithelium. Additional brushings or biopsies may be obtained for research purposes from areas of columnar appearing mucosa and also from normal squamous mucosa at least 3cm proximal to the squamocolumnar junction, normal gastric mucosa, or normal duodenal mucosa for control purposes. A total of up to 12 additional brushings or biopsies will be obtained. The number of total biopsies obtained in patients with BE will depend on the length of abnormal esophageal tissue. In patients with EAC, up to 6 additional biopsies may be obtained for research purposes. In normal control patients, up to 12 brushings and biopsies will be obtained from normal esophagus, stomach and duodenum for control

purposes. Research specimens will be snap frozen in liquid nitrogen or dry ice in the endoscopy suite. These additional brushings and biopsies will continue to be obtained should study subjects return for surveillance endoscopy and will also be obtained from subjects who have undergone ablative therapies for Barrett's esophagus. Subjects will be re-consented only if new information becomes available or a protocol version change has occurred. Specimen will be stained with hematoxylin and eosin (H&E) for routine examination. In select instances, a research biopsy specimen will be placed immediately in cell culture medium in the endoscopy suite, for subsequent establishment of research cell lines, with an emphasis on cases of previously diagnosed dysplasia or frank adenocarcinoma. Activation of insulin/IGF-1R pathway will be confirmed by immunohistostaining for mediators of the insulin resistance pathway such as Tyr-phos-IRS-1, Ki-67, phospho-AKT, and phospho-mTOR. Aberrant methylation will be studied by extracting DNA from snap frozen tissue or paraffin embedded tissue samples and running MSP assays on genes believed to cause progression from GERD to BE to EAC. The transcriptome and miRome will be studied by extracting RNA from snap frozen tissue and performing NextGen sequencing.

Paired stool samples will be used in this study as a preliminary assessment of non-invasive detection of aberrant methylation in fecal DNA as compared to results of biopsy-based tissue assays. Saliva samples will also be collected and the non-invasive detection of aberrant methylation in salivary DNA will be explored. A stool-based assay for vimentin methylation will also be performed by methods which our group has previously optimized for colon cancer detection. Similar assays will be performed on DNA extracted from saliva and brushings from the esophagus. Patients will be asked to provide one stool sample either prior to endoscopy or 6 to 14 days following endoscopy using a home stool collection kit with detailed instructions. Briefly, a collection container is placed on the rim of the toilet bowl and 250 mL of a DNA stabilizing buffer is added before the container is sealed. Time of collection is recorded and the container is shipped overnight at room temperature, using a coded identifier to protect privacy. Upon receipt, samples are processed and kept in frozen storage until time of analysis. A minimum of 50g of stool is necessary for analysis, which has not previously been a limiting factor in large trials of stool testing for colon cancer detection. Subjects will be provided with saliva collection kits. They will be asked to spit into a collection container containing DNA stabilizing buffer either at the time of endoscopy or they will be asked to take the container home and collect the saliva in the morning when they wake up.

Assays for fasting insulin, glucose, IGF-1 and IGFBP1/3 and other adipokinnes are performed at the GCRC lab of the Case Medical Center. Immunohistostaining of tissues will be performed in the lab spaces of Drs. Dawn Dawson and Joe Willis. Dr. Rom Leidner will perform parallel assays in brushings, stool and saliva at CWRU. Dr William Grady will be in charge of studying aberrant gene methylation in tissue at the Fred Hutchinson Cancer Center in Seattle, Washington. Dr. Guda will be responsible for sequencing the genome and the transcriptotme using deidentified samples at the CCCC Genome Sequencing Core or through a commercial sequencer as appropriate. Drs. Orlando and Shaheen will perform molecular and physiological studies of ablated epithelium.

After completion of endoscopy, patients will be asked to complete three self report questionnaires. The GERD questionnaire assesses severity, frequency and duration of

heartburn, dysphagia, and odynophagia.[117] The Arizona Food Frequency Questionnaire and Arizona Cancer Center Meat Preparation Questionnaire are semi-quantitative 153-item food frequency surveys that ask respondents to report how often they usually consume each particular food and how it was prepared over the prior 12-month period and to indicate their usual portion size. Vitamin/mineral supplement use information is collected and modification questions assess dietary behavioral practices related to fat consumption. The questionnaires have been tested for reliability as well as relative validity.[118,119] Questionnaires will take approximately 60 to 90 minutes to complete. All questionnaires need to be returned by mail to study coordinators at respective institutions. Patients may be called by phone to assure completion and return of questionnaire data as well as to assist with any necessary instruction on use and shipping of the home stool collection kit. After completion and return of all three questionnaires, subjects will receive twenty dollars as compensation for their effort and time.

Subjects will be informed that we cannot disclose individual results of genetic testing as per the Clinical Laboratory Improvement Act (CLIA) of 1988. Participants will be informed by letter if this study yields clinically relevant discoveries about the gene or genes underlying BE. At that point, the subject may arrange for genetic counseling at University Hospitals Case Medical Center without cost to learn more about the new discovery. At this time, they can decide whether or not to undergo clinical testing at their own expense. Those individuals who elect not to receive genetic counseling through University Hospitals will be referred to a local medical genetics service. Any costs associated with the services received by the study participant will not be covered by this study.

Participants in the research will be informed that it may be months or years before results are available, and may not receive direct results or benefit from participating in this study. Further, potential subjects will be informed that their sample(s) may not be used for the study if pathology reports indicate they do not meet eligibility criteria. Subjects will be informed that they can withdraw from the study at any time, and if they choose to do so, their samples will be destroyed. Subjects will be responsible for keeping the researchers informed of their phone number and mailing address so that they can be contacted when and if results become available.

DNA and Tissue Bank

Blood, brushings, mucosal biopsies and tumor tissue samples collected on patients will be stored de-identified at the DHI in the BRB or Dr. Markowitz' lab in the Wolstein Building. Blood samples may also be stored as serum, whole blood, lymphocytes, lymphoblastoid lines, DNA, and/or RNA at the DCRU. Per the NIDDK's request, participants' samples will also be stored at the Rutgers University Cell and DNA Repository (RUCDR) in Piscataway, NJ. Samples sent to Rutgers will be anonymized and only brief medical history information will be available to the RUCDR. The samples banked at RUCDR are available to the investigators involved in this study as well as other investigators throughout the nation conducting research on BE and/or EAC.

Genome Scanning

In the absence of any clues as to the nature of the proposed tumor suppressor, the putative gene should be searched for via genetic approaches. Linkage analysis requires the availability of blood samples from a number of families segregating BE and/or EAC. Under the assumption of

dominant inheritance with ~50% penetrance (existent families suggest a penetrance of at least 80%), typically a panel of 10 families each having 4 affected individuals in at least 2 generations would probably be sufficient to map a gene by this approach. There are significant limitations to using fewer families with fewer affected individuals, namely the search would be more difficult and time-consuming and, above all, our ability to interpret the results will be compromised.

Searching for allelic association among unrelated probands with BE/EAC is a powerful alternative method to linkage, but is not likely to succeed other than in well-defined isolated founder populations such as Finns, Jews, Amish, and French-Canadians. As we do not have access to special populations, we conclude that allelic association analysis is not a viable first step. However, with modern advances in gene hunting strategies, we have the option to utilize a combined linkage and association analysis approach, which with modern algorithms may provide a more powerful alternative if sample sizes are inadequate.

Microarray analysis ("Gene Chip") is a molecular research technique that can be used to assess the expression profiles of multiple genes simultaneously. Briefly, this technology involves a platform of unique cDNA elements immobilized on a glass slide. Purified mRNA from tumor, Barrett's epithelium, and normal tissues is reverse transcribed, labeled in some manner (usually with fluorophores), and hybridized to the slide. Data are analyzed by comparing relative fluorescence intensities between normal and tumor cDNA arrays, giving a profile of relative gene expression in these two tissues[33, 34]. These data can be used to identify genomic loci associated with the development of BE and its associated adenocarcinomas.

Massively parallel sequencing (NextGen sequencing) methods offer a new approach for sequencing the entire exome or even the whole genome. These methods are rapidly being adopted for identifying germline mutations in families and somatic mutations in cancers and other diseased tissues. Whole exome sequencing approaches could allow us to identify genes that share mutations in affected relatives in families or genes with two deleterious hits in cancers.

LOH is extensively used in the search for tumor suppressor genes, and on a number of occasions has proven successful. Hundreds, if not thousands of papers have been published on LOH in various tumors. The main problem with LOH as a tool for this purpose is the interpretation of the data. This is because of the remarkable genetic instability displayed by most cancers that can be seen as changes in ploidy, structurally abnormal chromosomes, and loss and gain of DNA in many regions. Thus, in many tumors, LOH is usually seen at so many different sites that it becomes problematic or impossible to determine which one or ones harbor the putative suppressor gene(s). Apart from LOH on 5q, 17p and 18q, no other regions have been fully investigated. Hence, this would be a worthwhile strategy for both BE and EAC.

LOH can be searched for by at least two independent methods, both of which compare tumor tissue and normal tissue for quantitative (CGH method) or qualitative ("test tube" method) differences. In comparative genomic hybridization (CGH), tumor and normal DNA are hybridized to normal human chromosomes on a slide. The amount of DNA from the tumor and normal tissue is compared and assessed along the chromosomes by quantifying the fluorescence. A limitation of this method is that only relatively large regions of chromosomal loss or gain (i.e. as a band or greater) are detectable. In the "test tube" system, microsatellite markers are tested individually for loss of one or both alleles. This is done semi-automatically in the ABI sequencer and the results are analyzed using existing software. With a standard marker panel of 400 fluorescent markers, the study of 5-15 patients can be readily accomplished. A main

problem is that not all markers are informative (heterozygous) in every patient and many markers may fail to amplify if the DNA is from fixed tissue.

Linkage analysis will be performed using a similar panel of SNP markers or the most modern technology available (eg, 100K or 500K SNP chip) scattered throughout the autosomal genome. This is performed semi-automatically on ABI platforms or Illumina Beadstation and results are analyzed using existing software.

This study will be approached via linkage analysis comprising an autosomal genome scan and LOH and microarray analysis, which should complement the genetic analysis. We will also explore selected multiplex pedigrees using whole exome and/or whole genome sequencing. Any areas of LOH would point genetic analysis to those markers in the first instance. However, this may be viewed conservatively as:

If LOH is found, it needs to be confirmed and further defined by studying more samples from multiple families and more markers in the flanking genetic regions. Once confirmed, the hereditary nature of the putative mutation in families can be tentatively tested by linkage analysis even when the families are too small to provide formal proof of linkage. When the existence of a locus is established, the search for the gene will proceed using the well-established genetic techniques of positional cloning and candidate gene approach. If no convincing LOH is found, this may mean that the hypothesis is wrong or that the attempts to detect it failed. However, the Gene Chip microarray technology may provide better resolution and may identify up-regulated as well as down-regulated genes.

Newer platforms are now available that make it feasible to genotype an increased number of SNP's and even sequence the entire human genome. Tissue obtained from selected patients will be grown in culture or xenografts in Dr. Markowitz' laboratory to enable genome wide sequencing as necessary. Attempts will be made to grow these de-identified samples into cell lines. Large segments of the genome from these coded samples will be sequenced either in the Cancer Center genotyping Core using the new Next Generation sequencer or through available commercial sequencing services. The samples will be coded and will contain no identifying information. This advancement in technology make it feasible to identify susceptibility mutations even from a single family.

8.0 DRUG INFORMATION

Not applicable

9.0 STATISTICAL CONSIDERATIONS

Recruitment estimates

The estimated number of eligible patients at institutions that will participate in the present study is shown in Table 1.

Table 1. ***Estimates of newly diagnosed BE, BE undergoing endoscopic surveillance, and new EAC patients encountered annually at participating institutions.***

Hospital Name (PI)	BE (New Dx)	BE (Survey)	EAC
UHC/Wade Park VA (Chak/Kinnard)	**30**	**90**	**28**
Hospital of U Pennsylvania (Falk)	**25**	**75**	**20**
FHCRC/UW Med Ctr (Grady)	**16**	**60**	**26**
Creighton (Mittal)	**15**	**60**	**25**
Mayo Clinic (Prasad)*	**18**	**80**	**61**
Johns Hopkins (Canto)	**50**	**160**	**52**
UNC (Shaheen)	**30**	**100**	**18**
WUSTL (Wang)	**24**	**100**	**24**
TOTAL	**231**	**730**	**256**

**Number of patients reflect cases available to Dr. Prasad at individual Mayo Clinic hospital for accrual.*

Accrual (Isolated Disease and Familial Disease): During the past 4 years our recruitment rate for affected cases has averaged 55% (range, 41% to 63%) at our accrual centers. We will be attempting to recruit from a pool of roughly 1200 BE/EAC cases a year. Therefore, we could annually accrue up to an estimate of 660 (550 BE, 110 EAC) cases. The presence of clinical factors known to be associated with BE and EAC is then determined by means of the FBEQ, a questionnaire that collects details regarding GERD symptoms, environmental exposure history, prior cancer history, family history (specifically of BE, esophageal cancer, and other cancers), height, weight, and prior upper endoscopy, as described in previous studies.[22, 23, 35] Over the past 5 years we have used a detailed family approach, i.e. contacting the proband as well as all family members to assess presence of disease in probands and family members.[23, 35, 36] We will continue to employ a detailed family history approach to enable accurate assessment of familial aggregation if the proband reports a family history of disease and we will concentrate on expanding existing multiplex FBE kindreds to identify affected distant relatives. This modified approach will use resources efficiently and will not greatly affect our accrual of FBE kindreds for subsequent genetic studies. We have ascertained FBE, i.e. confirmed presence of disease in proband and at least one relative, in 7.3% of cases, and suspected FBE, i.e. unable to review medical record of disease in relatives, in another 6.3%.[23] – we estimate that we will accrue up to 599 isolated disease cases, 35 ascertained FBE, and 26 suspected FBE cases annually.

We expect to bank up to 1,500 blood (serum, genomic DNA, plasma) specimens from BE and EAC cases with isolated disease (no reported family history) and fresh tissue from ca. 1,100 BE subjects and 400 EAC subjects at individual institutions. Over the past several years we have also banked ca. 600 cell lines from 211 FBE families in the RUCDR FBEC biobank. To save on banking costs, we are now proposing to bank DNA from all study participants at individual institutions. We will create cell lines at RUCDR for only multiplex families with affected distant relative pairs since these families provide the means for subsequent gene discovery using a whole exome sequencing approach. We will also bank lymphocyte DNA from an additional 400 individuals from ca. 200 new FBE families at our existing genomic DNA FBE biobank at RUCDR during 5 years. This will result in a total of ca. 1000 DNA samples from ca. 500 FBE families. This unique and valuable resource will enable genetic studies for our consortium as well as other investigators.

Objective – Genetic Susceptibility

The general strategy to achieve our objective is to perform a total autosomal linkage genome scan of at least 120 "typical" FBE families. However, because we may actually have more than 2 susceptibility loci, to increase power, we plan to recruit and perform our total genome scan on >250 potentially informative families. Simulations have estimated that 120 is the minimum number of families required to have sufficient power to achieve a mean maximum LOD of 3.0 assuming the following. "Typical" family pedigree structures were operationally defined as follows:

- Affected sib pair
- Affected sib pair and unaffected sib
- Discordant sib pair and affected parent
- Affected avuncular pair
- Affected cousin pair

Where the above individuals have DNA available for genotyping, but the parents or other connecting individuals may not necessarily be available. These assumptions of "typical" pedigree structures are valid because the great majority ($\geq$ 85%) of the families collected by the FBEC thus far have one of these structures. Simulations were performed using SIMLINK[37] for the above pedigree structures and four different autosomal dominant models that are consistent with the observed population prevalence and the roughly calculated sibling risk ratio (λ_s) from our preliminary study[22]. Assuming a recombination fraction of 0.05 (e.g., for a 400-marker genome scan the average inter-marker distance is 9 cM) and genetic homogeneity, our simulations indicate we will need a sample size of 60 "typical" pedigrees to achieve a mean maximum LOD score of 3.0. Under an assumption of genetic heterogeneity, where only 50% of the pedigrees are linked to a particular locus, the required sample size is doubled to 120 "typical" families. This fraction of pedigrees linked to a single locus is a realistic assumption based on our preliminary data noted under section 3.2, where approximately half the families are excluded from linkage to the chromosome 20 critical interval, albeit based on small numbers of families so far genotyped. This estimate of sample size is actually conservative since ~10% of pedigrees have 3-6 affected individuals potentially available for analysis, and will provide more information to detect linkage than the "typical" pedigrees described above. The FBEC has already recruited 93 families and the conduct of studies in Objectives 2 and 3 should result in an additional 210 families, including 10 multiplex multigeneration families with significantly higher LOD scores than used for our conservative simulations. In addition to the families already identified by FBEC, the recruitment of these new families should allow us to easily achieve our projected samples size. Therefore, this study will have sufficient power to detect linkage at the two (or more) putative FBE loci.

Linkage analysis will be performed with the input of FBEC statistical geneticists, Dr. Robert Elston. Dr. Elston and his colleagues are continually developing the program package S.A.G.E. (Statistical Analysis for Genetic Epidemiology), funded by a resource grant from the National Center of Research Resources, which will be used for this analysis. The investigators will perform a total genome linkage analysis paying particular attention to the chromosome 20 critical interval, as well as to uncover other regions of linkage.

Preliminary analysis and data checking will be performed prior to the linkage analysis. First, descriptive statistics about nuclear family data will be produced by the PEDINFO program of the S.A.G.E. package. More than 100 descriptive statistics on pedigree structure can be calculated including means, variances, histograms of family, sibship, and pedigree size, counts of each type of relative pairs, etc. These analyses will enhance our understanding of the dataset, and flag many common errors in pedigree structure. Any errors that are detected will be resolved in collaboration with the Genetic Counselor prior to proceeding with the analysis. Second, genome-wide relationship testing will be performed using the RELTEST program of the S.A.G.E. package. This program currently performs analyses to classify putative sib-pairs, half-sib pairs, parent-offspring pairs, and marital pairs into five different classes: MZ twin pairs, sib pairs, half-sib pairs, parent-offspring pairs, and unrelated pairs. Relationships that have been misclassified will either be reclassified (if possible) or removed from the analysis. Third, we will check for Mendelian inconsistencies using the program MARKERINFO of the S.A.G.E. package. These inconsistencies are sorted by marker and by pedigree. A large number of inconsistencies for a single marker may indicate problems with genotyping for that marker, and the marker may be removed from the analysis. A large number of inconsistencies for a single pedigree may indicate errors in the putative relationships that were not detected with the RELTEST program (which does not test *all* possible relationship types for extended pedigrees).

For our initial linkage analysis, we will use model-based linkage analysis (i.e., the LOD score method) with the same model as in the initial genome scan with pedigree 3090. The assumptions of this model included autosomal dominant inheritance with an age-dependent penetrance function of 90% by age 50 years and a conservative 3% prevalence of BE or EAC in the general population, and a conservative allele frequency for the trait locus of 0.0015. Marker allele frequencies will be estimated from all the family data collected using a maximum likelihood approach implemented in the FREQ program of the S.A.G.E. package. The underlying assumption is that none of the marker alleles are in linkage disequilibrium with a disease locus. While this assumption may not be valid for all marker alleles, it is robust in that linkage disequilibrium will not lead to a false finding of linkage. Parametric, or model-based, linkage analysis is the most powerful linkage analysis method when the model parameters for the *linked* trait locus are correctly specified. The power of this method is most sensitive to misspecification of the mode of inheritance for models with one or more trait loci[38, 39]. Given our strong belief in a dominant mode of inheritance for this disease (based on published reports[22, 25-31]), we believe that model misspecification will not significantly reduce the power of our analysis. Multipoint linkage analysis is more sensitive to model misspecification than two-point linkage analysis[40]. Therefore, we will also compute two-point LOD scores for each marker which will a) provide supporting evidence for a location if more than one marker shows positive evidence of linkage in the region, b) ensure that the multipoint LOD score peaks are not a result of a single marker with substantial genotyping error, and c) ensure that we do not miss any regions because of model misspecification in the multipoint analysis. Our preliminary results suggest there are at least two loci that predispose to FBE. Therefore, the results of the linkage analysis will be used to perform a heterogeneity test, and to estimate the proportion of linked families in candidate regions (α)[41].

We will use the MLOD and LODLINK programs of the S.A.G.E. program package to perform multipoint and two-point linkage analysis, respectively. LODLINK computes model-based two-point LOD scores to obtain a maximum likelihood estimate of the recombination fraction (θ) in the parameter space [0, 0.5], performs tests of homogeneity, and estimates the proportion of families linked to a particular region (α). MLOD performs multipoint model-based linkage analysis. A modified version of the Lander-Green algorithm[42] is implemented in MLOD to

perform the computations[43], which significantly decreases the computation time for multipoint linkage analysis compared to programs using the Elston-Stewart algorithm[44]. The Lander-Green algorithm scales linearly with the number of loci and exponentially with the number of nonfounders in the pedigree, meaning that it is possible to perform multipoint linkage analysis with large numbers of markers on moderately sized pedigrees. We anticipate that the pedigree sizes and structures that are recruited in this study will be well within the limits for this type of analysis. Because of the sensitivity of model-based linkage analysis to misspecification of the model parameters, we will also perform nonparametric linkage analysis. The Genehunter Plus [45] modification of the Genehunter program will be used for the model-free analysis, which corrects the variance of the NPL Z-score test statistic to provide a less conservative test when there is incomplete data. The model-free method implemented in Genehunter Plus is more powerful than several common alternatives [e.g., sib-pair analysis, or the Affected Pedigree Member (APM) method[46, 47]] and is robust to model misspecification.

Analyses of different subgroups will allow us to determine whether: 1) positive findings in the whole sample are found within subgroups, and 2) subgroups lead to positive findings not present in the whole sample. Covariates that will be considered in this analysis include age, gender, reflux symptoms, obesity, and clinical subcategories such as disease status (i.e., BE vs. EAC). These stratified analyses should help control for genetic heterogeneity and increase the probability of identifying disease susceptibility genes. Stratification is one method that is often used to analyze different subgroups. However, it is difficult to incorporate multiple covariates using stratification, because it results in smaller and smaller subgroups, and a loss of power. In addition, it is difficult to incorporate quantitative covariates into the analysis, because arbitrary cutpoints must be used to divide the sample. Therefore we will use an affected relative pair method implemented in the LODPAL program of the S.A.G.E. package that avoids both of these difficulties. This program implements the general conditional logistic model of Olson[48] modified to give the one-parameter model of Goddard et al[49]. In this method, the allele sharing between relatives depends on the value of covariates, and thus we are able to allow for locus heterogeneity due to those covariates. In terms of offspring recurrence risk ratio, conditional on K covariates, x_k, the model is parameterized as $\lambda_1(x) = \exp(\beta + \sum \gamma_k x_k)$; in terms of the recurrence risk ratio for MZ twins, it is parameterized as $\lambda_2(x) = 3.634\lambda_1(x) - 2.634$. Covariate specific sibling recurrence risk ratios can be obtained by the expression $\lambda_s(x) = ¼ + ½ \lambda_1(x) + ¼ \lambda_2(x)$. "LOD" scores are obtained as the likelihood-ratio statistics divided by 4.605 (i.e., $2\log_e 10$). Using this method, one can test both the significance of the contribution of a covariate and the overall evidence of linkage.

Virtual physical mapping and candidate gene analysis

We anticipate that five to six marker regions will "suggest linkage" by chance alone and warrant further investigation in order to maintain our desired level of statistical power. For each positive region, six to eight markers spaced two cM apart on either side of the positive marker will be typed. Further, if putative candidate genes lie within 20 cM of the region with suggestive linkage, then markers close to or within the gene or the candidate gene itself will be typed. Since markers for the second stage cannot be specified *a priori*, exact spacing or marker heterozygosities are unknown.

In the 33-cM chromosome 20 critical interval, haplotypes will be inspected and recombinants noted so as to narrow the critical interval. Similarly, the other novel critical intervals noted from the genome scan will have their respective haplotypes inspected for recombinants to minimize

the critical interval(s). The narrowest critical intervals will be subjected to virtual physical mapping and data mining from the federal and Celera human genome databases (available at the Human Cancer Genetics Program, CCF) for possible candidate genes. There will likely be multiple candidate genes per critical interval, typically 20. Mutation analysis of all candidate genes would prove costly, labor intensive, and be well beyond the resources of this proposal. Therefore, to prioritize the analysis of candidate genes, the putative genes and EST's will be examined for known (or putative) function (e.g. cell cycle, apoptosis) and expression in the esophagus. These genes will be initially targeted for mutation analysis of a selected series of genomic DNA samples from 20 unrelated affected individuals representing the youngest affected from the most classic and largest families and a series of 20 normal controls. Power calculations reveal sufficient power (β=0.88) for a sample size of 20 probands if the putative mutation were present in 10% of such cases. Direct sequence analysis using the ABI3730xl (fee-for-service Genomics Core Facility, CCF's Genomic Medicine Institute) will be used to examine for variation in each candidate gene between affected individuals and controls. Once a variant in a particular candidate gene is observed which appears to be present only in at least one affected individual and not among controls, this gene will be subjected to mutation analysis by direct sequencing of one proband from each of the families and 100 unrelated normal controls matched by region and race.

Analysis For Whole Exome/Whole Genome Approaches

The allelic variants that are not present in the dbSNP and 1000 Genome databases will be compared with respective matched affected relative(s) to identify missense, nonsense, frameshift-causing indels, and splice site alleles that are shared by all affected relatives. The list of shared candidate variants will further be filtered for each pedigree if we have DNA available for other affected informative members whose exomes were not sequenced. We will genotype these other affected members and determine if the variant is also present in those members. The candidate variants for each pedigree will then be explored further by prioritization based on predicted loss of function and prediction of a deleterious change in function using a prediction algorithm such as CHASM.[50] Based on the number of novel loss of function mutations (< 1/400 genes) identified in other studies that have performed whole exome sequencing[51, 52] we expect that any distant relative pair will on the average share no more than two novel loss of function (nonsense or frame shift) mutations. Any such gene with a loss of function variant identified from sequencing a distant relative pair, confirmed in other affected members of the pedigree, will immediately become a candidate for validation as a causative gene. Any variant that causes a deleterious mutation in a gene mutated in other cancers as catalogued in the cancer genome atlas (TCGA) will also become a candidate for validation. Finally any gene that has shared private variants in two or more affected relatives in independent FBE kindreds will also become a candidate for validation.

Genes with two NS/SS/I variant alleles in cancer tissues will be given priority for consideration as candidate causative genes. Any gene sequence from an FBE cancer that has two novel allelic variants will be examined further. Germline DNA will be obtained from the lymphoblastoid cell lines or from the normal squamous epithelium stored on these individuals in the FBEC biorepository at the RUCDR. Using primers designed for each identified variant, the germline DNA will be sequenced using Sanger sequencing. A variant not identified in the germline will be classified as a novel somatic mutation. A gene with one allele that has a loss of function germline mutation and a second allele with a somatic mutation in an FBE cancer will be a strong candidate for the causative gene and will be immediately analyzed for validation. We expect that on the average less than 3 genes per sequenced EAC will meet these criteria. Of course, prior to validating the gene we will determine whether the germ-line mutation is also present in other affected members of the particular pedigree who have DNA specimens banked at RUCDR.

Whole Exome sequencing will be performed from single affected members of selected multiplex pedigrees using NextGen sequencing techniques.
Sample Size: We have calculated our sample size of sequencing 30 exomes based on existing data that an average individual genome contains roughly 650 genes with un-annotated private non-synonymous variants in coding exomes, i.e. that per genome each gene has a 3% likelihood of carrying a private nonsynonymous variant.[51-54] Assuming 20,000 coding genes that are sequenced, and binomial distributions of novel variants in random genes, upon sequencing the whole exome of 30 individuals the expected number of genes for which X or more individuals have rare variants is given in Table 2. Of note the distribution decreased rapidly. Only 37 genes are predicted to be randomly detected as having private non-synonymous variants detected in at least 5 of the 30 sequenced exomes, and only 4.6 genes are predicted to randomly be detected as having private non-synonymous variants in at least 6 of the 30 sequenced exomes.

We also modeled FBE very conservatively as a trait in which 5 major genes with novel variants are each responsible for disease in 15% of multiplex FBE kindreds plus multiple other genes with very rare variants that are responsible for disease in the remaining 25% of multiplex FBE kindreds. Again, assuming that 20,000 coding genes are sequenced in 30 individuals from multiplex FBE kindreds of which 5 are such causal major FBE genes, we expect that 2.3 of these 5 major FBE genes will be detected by each having novel non-synonymous variants present in at least 5 of the 30 sequenced exomes and that 1.4 major FBE genes will be detected by each having novel non-synonymous gene variants present in at least 6 of the 30 sequenced exomes (Table 2).

Table 2. Estimated distributions of random variants and true FBE variants detected in X or more exomes upon sequencing of independent individual exomes from affected members of 30 multiplex FBE kindreds

X	Random Gene Variants	True FBE Variants
1	11979.860	4.962
2	4538.490	4.760
3	1201.382	4.243
4	238.093	3.392
5	36.994	2.378
6	4.652	1.447
7	0.484	0.763
8	0.042	0.349
9	0.003	0.139

Comparing random variants with true FBE variants detected, we see that if we select to carry forward to validation in aim 4 all genes in which novel non-synonymous variants are detected in at least 6 of the 30 kindreds studied, that we will expect to carry forward 1.4 true FBE genes and 5 genes with variants found by chance. If we select to carry forward validation of all genes that are detected in at least 5 of the 30 kindreds studied, then we would carry forward at least 2 true positives, at the cost of 37 false positives.

Because we expect that FBE is genetically heterogeneous, we will search for genes that have novel variants in subsets of all sequenced individuals using a strategy similar to that employed by Ng et al.[52] We will first rank each gene in the genome according to the number of kindreds identified that carry a private coding variant in the gene. Genes will then be prioritized for further study according to:

i) The numbers of kindreds identified that carry clearly inactivating germline variants in the gene (NS/SS/I);
ii) The numbers of kindreds identified that carry either clearly inactivating germline variants or that carry missense variants predicted to inactivate function (using the CHASM prediction algorithm); [50]
iii) The numbers of kindreds identified that carry private germline variants in genes identified as targeted for somatic mutations in human cancers, as recorded in The Cancer Genome Atlas (TCGA) ;[55, 56]
iv) The numbers of kindreds identified that carry private missense variants in a candidate gene.

Priority will be given to genes with identified novel coding mutations that are not in public databases. Unique genes with novel coding variants in the same gene present in the maximum number of independent exomes ($\geq$ 5 shared exomes), especially those that have loss of function mutations or predicted deleterious mutations and are associated with BE or EAC, will become candidates for validation. We plan to test up to 50 candidate genes for validation.
Assuming we sequence 100 cases and 200 controls (all unrelated) and a Bonferroni significance level adjusted for testing 50 genes of 0.001, we provide power calculations for mutation frequency in the controls ranging from 1-3% and in the cases from 10-20% (Table 3).

Table 3. Power for detecting significant differences in mutations between 100 cases and 200 controls at $\alpha = 0.001$.

Control Mutation Freq	Case Mutation Freq	Power
1%	10%	0.64
	15%	0.90
	20%	0.98
2%	10%	0.44
	15%	0.81
	20%	0.96
3%	10%	0.26
	15%	0.68
	20%	0.91

Thus, using our multi-pronged approach we have reasonable power for identifying the rare mutations that are responsible for genetic susceptibility in 15 to 20% of FBE pedigrees.

Validation Strategy One – Case Control Comparison of Germline DNA: We will sequence candidate genes in the FBE cohort and unaffected controls. Depending on costs we will use Sanger sequencing for a small number of candidates, or NextGen sequencing if we need to sequence a larger number. Relevant primers will be designed according to the specific polymorphisms and tested in the affected DNA from which the variant was identified using PCR assays to confirm the robustness of the primers. DNA from affected probands of 100 FBE families and 200 unaffected age, gender, and race matched population controls (DNA is available from over 1000 deidentified annotated population controls seen at the CCCC) will then be selected for sequencing. We will consider a candidate gene as validated first if the specific allelic variants identified initially and found as again present in FBE cases but not in controls, or alternatively if the proportion of deleterious mutations in a candidate gene (as assessed by using the CHASM algorithm[50]) are significantly enriched in FBE cases compared to unaffected controls. A p-value of 0.05 divided by the number of candidate genes tested (Bonferroni correction) will be considered significant. In addition, we will use the cumulative minor-allele test (CAMT) in order to jointly analyze a group of variants within a gene or region for association with FBE [57], which has been shown to be equivalent to other weighted sum statistics for sequencing data such as the weighted sum statistic [58]. DNA from members of FBE pedigrees

who have been phenotyped with endoscopic screening has been banked in the RUCDR biorepository. Deleterious variants that are identifed in an FBE proband will then be tested in other affected and unaffected members of that pedigree.

Validation Strategy Two -- Archived FBE Cancers: Further support for the pathogenic role of candidate genes in which we identify germline alterations will be sought by determining if these genes are also targets for developing somatic mutations in archived familial EAC specimens from the pedigree(s) in which the germline variant(s) of the particular candidate gene were discovered. We are able to extract DNA of up to 600 bp in size from formalin fixed and paraffin embedded material and sequence these DNAs for the purpose of mutation identification in cancers.[59] Furthermore, we have developed the ability to enrich target regions using Agilent's SureSelect capture kit and sequence a large number of genes using NextGen sequencing. Defined PCR primers are already designed for amplification and sequencing of every exon, including splice junctions, in the human genome.[60] Depending on the number of genes that will be candidates for validation and the costs, we will either use Sanger sequencing for a small number of candidate genes or use NextGen sequencing for a larger number of candidate genes. Our current FBE archive contains FFPE blocks from cancers identified in 42 pedigrees. We will extract DNA from these cancers and sequence the candidate genes that had germline mutations in the particular FBE pedigree. The presence of a somatic mutation in the second allele of a gene known to already contain a germline mutation will confirm that gene as a causative susceptibility gene. The germline allelic variant(s) in the gene that were discovered by exome sequencing will be classified as susceptibility allele(s) for FBE.

Validation Strategy Three -- Sporadic Cancers: We will also extract DNA from 100 resected surgical FFPE specimens of EAC maintained in the Case Comprehensive Cancer GI Tissue Core. Over the past 20 years, Co-I Dr. Willis, the Director of this Core, has created an archive of over 200 EAC and EGJAC that were not treated with chemotherapy or radiation prior to surgery. If we use NextGen sequencing to sequence our candidate genes in these cancer specimens, we will individually capture the candidate genes from each specimen. In converting the captured DNAs into Illumina sequencing libraries suitable for analysis on the Illumina Genome Analyzer, we will incorporate identifying tags into the library primers, such that the library from up to six individuals will be uniquely tagged. This will allow all six individuals to be pooled and sequenced on one run of the Illumina sequencer, allowing substantial cost savings. The presence of somatic genetic mutation(s) will be determined by comparing the sequences of the candidate gene with public databases (dbSNP). In cancers where we identify a genetic mutation, we will also sequence germ-line DNA from the normal squamous esophageal tissue contained within the archived resection specimen to confirm that the individual did not contain a germ-line mutation. Since somatic mutations in cancer genomes are rare, candidate genes confirmed to have somatic mutations in more than one sporadic cancer will be considered to be susceptibility genes for BE and EAC. Somatic mutations identified by Next Generation sequencing will be re-confirmed by Sanger sequencing of the DNA extracted from the corresponding sporadic EAC tumor.

Models for Predicting BE in Family Members:

Our published segregation analysis gave evidence for the segregation of dominant alleles.[61] The best model also included a significant polygenic component, which we surmised was simply evidence of heterogeneity across families in the penetrance of the particular causal alleles. Further unpublished analyses of this sample, updated and using imputation to determine the ages of the 40% of family members for whom this was not recorded,[62] indicate that, after making the penetrance age-dependent, the polygenic component is no longer significant. This suggests that once the disease-causing variant in a family is known, prediction of which family members will (without preventive treatment) succumb can be accurately made from knowing the specific genotypes of those individuals, which will have important clinical application. Even if the assumption of penetrance homogeneity is wrong, including knowledge of the variant segregating in the family, together with an appropriate polygenic component, will improve the predictive probability (i.e. make it closer to 0 or 1) that could be calculated for any member of the family. We therefore propose to develop a model, similar the segregation model we have already implemented, that incorporates the genotype at the appropriate segregating locus in the penetrance, together with age and other relevant risk factors. Then, for situations with or without

knowledge of the appropriate genotype for any particular family, we shall build tables for clinical use that predict the recurrence of BE in the relative of a proband with BE.

Assignment of methylation status/data analysis: Tissue DNA samples will be subjected to sodium bisulfite conversion using the EZ96 DNA Methylation Kit (Zymo Research) and eluted in an 18 µl volume with retention of 3 µl for use in post-bisulfite quality control experiments to determine the completeness and recovery of bisulfite conversion in the sample set. The HumanMethylation450 DNA methylation assays (Illumina) studies will be carried out on the converted DNA according to manufacturer's specifications in the FHCRC Genomics Shared Resources Core after DNA amplification using published protocols [63]. Of note, on these arrays there is a pair of probes used for each assay, one designated "M" which binds to a methylated CpG and a second, designated "U," that binds to an unmethylated CpG. The binding score is an average of ~30 background subtracted, replicate measurements. The beta value, the calculated DNA methylation value for each locus is determined as: M / (U + M). U and M intensities will be extracted and beta values and detection p-values (the statistic after comparison of the intensities for each locus versus a panel of negative controls) will be calculated for each locus and sample according to Illumina specifications. In addition, analysis will be done using the M value method given its more statistically robust performance on methylation array data [64]. M value measurements with accompanying detection p-values > 0.05 that are not significantly different from the panel of negative controls will be re-labeled as "N/A." Samples for which <80% of the data points have detection p-value>0.05 will be removed to ensure high quality data.

Statistical analysis of methylation studies: Data will be analyzed by Dr. Chao Jen Wong at FHCRC in collaboration with Dr. Jill Barnholtz-Sloan at CWRU who are members of the TRC-F Bioinformatics Core.

Power: In this aim, we will evaluate the association between the methylation levels of ~450,000 CpG, with the histologic status of the tissue, The comparison groups will be 1) Normal esophagus (SQ) vs. BE, 2) Normal stomach (GA) vs. BE 3) SQ v. EAC, 4) GA vs. EAC, 5) BE vs. BE/LGD, 4) BE vs. BE/HGD, 4) BE+HGD vs. EAC and 5) EAC vs. BE. Power is based on a two-sided two-sample t-test for each site assuming a training dataset of ~50 tumors. In the training set, 100% power will be achieved to detect an effect size (defined as (mean of group 1 –mean of group 2)/SD) at least 1.4 for each test when controlling the false discovery rate at 5% and at least 100 loci are differentially methylated.[65] The effect size for the differentially methylated loci with significance when controlling the false discovery rate at 5% between normal and BE and EAC groups range between 1.4 and 2.2 in our preliminary studies. After the model is developed using the training dataset, the model will be validated using the validation dataset (at least 50 BE and 50 EAC cases, 50 matched normal esophagus tissue samples and 50 normal proximal stomach samples). With a significance level of 0.05, 84.3% power will be achieved to detect an effect size of 0.60, and 99.0% power will be achieved to detect an effect size of 1.0.

Data analysis for methylation studies (Bioinformatics Core)

a. Identification of candidate methylated genes for use as biomarkers: For each CpG site across all samples stratified by BE or EAC we will first assess the mean, minimum and maximum and standard error of the methylation intensity values (i.e. M values), making note of potential outliers and checking for consistency between replicated samples. For each sample we will then average the CpG M values adjacent to each gene, when there are multiple CpG sites assessed per gene, allowing us to obtain a gene-average M value for further analysis. Two-sided t-tests will be utilized to identify genes whose methylation intensity differ between 1) BE and both normal stomach (GA) and normal esophagus (SQ). Those genes that are not methylated in the normal stomach and normal esophagus, as defined as being

<2SD below the mean methylation value of the normal values in both the esophagus and stomach, will then be used in the comparison groups described above in the power analysis. Global tests of genome-wide methylation will be performed; SAS and R will be used for statistical analyses.

The methylation data will be assessed in an unsupervised analysis to detect any large-scale differences in the BE and EAC samples. The data will be filtered to retain only loci with changes >0.5 and having a minimum level of variance within the measurement. These highly variable methylated loci will be clustered using hierarchical clustering, with the Pearson correlation as a distance metric and full linkage. Cluster 3.0 will be used to cluster the data, and Java Treeview will be used to visualize the resulting heatmap. Large-scale patterns, such as large groups of targets similarly methylated or large groups of samples with similar methylation patterns may be observed. [66]. This type of analysis is used to observe broad patterns of methylation in the tissue samples. In addition, we will use Ingenuity Pathway Analysis software and other published approaches to identify functional groups affected [67]. Furthermore, logistic models will be developed to assess whether panels of CpG loci have the potential to accurately distinguish the groups described above.

b. Identification of low and high methylator BE and EAC subgroups: In addition to the analysis described above, we will also assess whether there are high and low methylator molecular subclasses of BE and EAC that may have different clinical behaviors. A threshold for the high and low methylator subgroups will be established using a subset of the BE and EAC cases as a training set. The results of this analysis will then be applied to a test set of samples to validate the scoring system. Individuals with high versus low methylation type lesions will then be compared to determine if these molecular differences reflect differences in histology, presentation, and clinical course (including representation in early versus late lesions as a surrogate indicator of likelihood of disease progression in the two disease subsets).

c. Assessment of methylated genes in BE adjacent to LGD, HGD, and EAC to identify genes that can be used in future longitudinal studies of prognostic methylated genes in BE, LGD, and HGD: Methylated genes identified using the HumanMethylation450 arrays that are present in HGD and EAC but not in BE will be assessed. Those methylated genes present in BE adjacent to HGD and EAC will be candidate biomarkers for longitudinal studies for prognostic markers for BE and LGD.

Methylation analysis comparing FBE and sporadice BE: We will also evaluate the methylation levels of 450,000 CpG's between familial BE cases vs. sporadic BE cases. Power is based on a two-sided two-sample t-test for each site assuming a samples size of 50 familial BE/EAC cases and 50 sporadic BE/EAC cases. With a significance level of 0.05, 84.3% power will be achieved to detect an effect size of 0.60, and 99.0% power will be achieved to detect an effect size of 1.0.

a. Adjustment for clinical factors: The comparison of the sporadic and familial BE cases will be adjusted for known and suspected confounding factors that affect BE incidence and progression risk, including age, sex, ethnicity, tobacco use, obesity, antacid therapy (e.g. proton pump inhibitor use), use of NSAIDs, severity and duration of gastroesophageal reflux disease (GERD) symptoms, and length of BE [68].

b. Analysis of results from methylation arrays: For each CpG site across all samples stratified by familial BE or sporadic BE we will first assess the mean, minimum and maximum and standard error of the methylation intensity values (i.e. M values), making note of potential outliers and checking for consistency between replicated samples. For each sample we will then average the CpG M values adjacent to each gene, since there are multiple CpG sites per gene, allowing us to obtain a gene-average M value for further analysis. Two-sided t-tests will be utilized to identify genes whose methylation M values differ between familial vs. sporadic BE. Global tests of genome-wide methylation will also be performed to determine if the proportion or pattern of methylated genes varies between familial and sporadic BE samples . We will use SAS and R for statistical analyses. The methylation data will be

further explored in an unsupervised analysis to detect any large-scale differences in the BE samples. The data will be filtered to retain only loci with large changes between minimum and maximum differences (e.g. > 0.5) and having a minimum level of variance within the measurement. These highly variable methylated loci will be clustered using hierarchical clustering, with the Pearson correlation as a distance metric and full linkage. Cluster 3.0 will be used to cluster the data, and Java Treeview will be used to visualize the resulting heatmap. Large-scale patterns, such as large groups of targets similarly methylated or large groups of samples with similar methylation patterns may be observed. This type of analysis is used to observe broad patterns of methylation in the tissue samples.

Analysis of coding RNA markers: For RNA profiling studies in cancers, we will characterize the transcriptome of ca.150 EAC cases and their matched uninvolved/healthy nSQ tissues obtained from upper esophagus. For RNA profiling in premalignant lesions, we plan to use an independent study cohort of 90 BE cases with no concurrent dysplasia or cancer, and 75 HGD cases with no concurrent cancer, and their respective matched nSQ tissues. The patients enrolled in this study will be mostly white males (>80%), which is in accordance with the well-recognized white male predominance in patients susceptible to BE [69]. In order to minimize the heterogeneity among our cancer cases, we will exclude cancers of the esophagogastric junction (AEJ), and will primarily focus on intestinal-type EACs. ***Experimental:*** Total RNA, which includes the miRNA fraction, will be extracted from snap frozen EAC biopsies using *mir*Vana kit (Ambion, Austin, TX), and the quality will be assessed using a Bioanalyzer (Agilent technologies, Santa Clara, CA). Only samples with RNA Integrity values (RIN) ranging from 7 - 10, and Abs260/280 ratios >1.9 will be selected for further analysis. Of note, the samples utilized for our pilot studies in this proposal had RIN values ranging from 7.3 to 9.9, while the Abs260/280 ratios from 1.9 to 2.1. Briefly, paired-end poly A RNA libraries will be generated from 1-5 µg aliquots of total RNA using the poly A RNA-sequence library construction kit as per manufacturer's instructions (Illumina, San Diego, CA). As a cost effective measure, we will index the EAC and nSQ RNA libraries using two different indexed adapter tags, respectively. Next, the RNA libraries from each of the matched EAC and nSQ, that have unique index tags, will be pooled together, and deep sequenced using Illumina GAII with a 76bp single-end sequencing protocol at the High-throughput Sequencing Facility. NGS data analysis and quantitation will be performed using RNA-seq module of the Genome Studio package (Illumina). The ratio of RPKM values in matched EAC vs. nSQ will be calculated to obtain RNA fold-change in EACs. Additionally, we will identify a set of house-keeping genes that remain unaltered between nSQ and EAC samples, and will further normalize individual gene expression counts to these house-keeping controls. Gene expression levels in pre-cancerous lesions will be determined using RNA isolated from laser microdissected frozen tissue sections of BE, HGD, and matched uninvolved nSQ biopsies. As a quality control, representative serial sections from each of the biopsies will be reviewed by Dr. Joseph Willis, chief of gastrointestinal pathology, for confirming diagnosis and for demarcating areas of BE and HGD lesions prior to LCM. 24-48 (7µm) frozen sections stained with Cresyl Violet (Ambion) will be laser microdissected to isolate pure cell populations of BE, HGD, and uninvolved nSQ epithelia (Veritas System, Sunnyvale, CA). Next, total RNA which also includes the miRNA fraction will be extracted using RNAaqueous-Micro kit (Ambion), followed by DNAse treatment. RNA concentration will be determined using Qubit (Invitrogen, Carlsbad, CA), and expression analysis of *CANCER-RNAs* in BE and HGD lesions will be performed using 50-100ng of LCM RNA by employing the whole genome DASL assay platform. Results from the DASL assay will be analyzed using the Gene Expression module of the Genome Studio software (Illumina). For quantitating lncRNAs, and examining for alternative splice variants in premalignant lesions, we will employ individual custom designed qPCR assays. ***Statistical:*** The statistical analyses will

be directed by Dr. Jill Barnholtz-Sloan. We will split our total sample size of 150 pairs into a discovery set (n=75) and a validation set (n=75). We will discover all differentially expressed genes between all 75 EAC and nSQ pairs in the discovery and validation set and will rank order all genes according to the following criteria: RNAs altered with ≤ 5% false-discovery rate (FDR) p-value, RNAs altered at least 4-fold within EAC and nSQ pairs, and RNAs deregulated in ≥ 50% of EACs. With 75 matched pairs, assuming an absolute mean difference in log gene expression of 4-fold and a standard deviation of the log ratio of gene expression between groups of 0.66 we have >90% power to detect only 1 false positive out of 10,000 when 50% of the genes are deregulated [70]. Next, we will describe the distribution of the deregulated genes by BE (n=90) or HGD (n=75) status. Looking at the progression model in a pair-wise fashion, i.e. BE vs. HGD or HGD vs. EAC, we will describe the differences in the proportion of deregulated *CAN-RNAs* using a one-sided Fishers exact test. Assuming that 10% of BE (n=90) have a particular gene altered, we have 80% power to detect only 1 false positive out of 10,000 to see 43% or higher of HGDs (n=75) have the same gene deregulated. Now assuming 43% of HGDs (n=75) have a particular *CANCER-RNA* deregulated, we have 80% power to detect only 1 false positive out of 10,000 to see 76% or higher of EACs (n=150) have the same gene deregulated. Recall that the *CANCER-RNAs* were defined as being deregulated in at least 50% of all EACs. We will also be interested in *CANCER-RNAs* that are not detected or are detected in extremely low proportions or are altered inconsistently in one group compared to another. As a proof of principle that we have reasonable power to assess these genes, we calculate that if 1% of BE (n=90) have a particular *CANCER-RNA* deregulated, we have 80% power to detect only 1 false positive out of 10,000 to see 27% or higher of HGDs (n=75) have the same gene deregulated.

Analysis of non-coding RNA Biomarkers -- *Experimental:* miRNA alterations in premalignant and malignant lesions will be determined in the same BE, HGD, and EAC cases used for coding RNA profiling studies. Briefly, 1-5 µg total RNA aliquots from EACs will be used for generating miRNA libraries (Illumina). Next, the miRNA libraries will be deep sequenced using the Illumina GAII. As a cost effective measure, we will incorporate unique indexed adapters in to the miRNA libraries in such a way that up to four samples can be pooled prior to deep sequencing, while still achieving sufficient depth of coverage. Sequence reads that are >15bp long [71, 72], and have a minimum of 6 transcript counts will be subsequently processed for miRNA detection using the miRANALYZER software package [73]. The raw miRNA expression counts obtained from miRANALYZER will be initially normalized to the total number of sequence reads for each sample. In order to minimize technical variability among the samples, we will additionally perform rank-invariant normalization on the miRNA expression data. Furthermore, we believe that our approach of deep sequencing a significant number of normal and cancer samples will likely identify a set of "house-keeping" miRNAs, which can be further incorporated in to our normalization protocol. For quantitating miRNAs in laser microdissected premalignant lesions, we will use 50ng of total RNA aliquots from BE and HGD lesions, and will perform qPCR analysis using miRCURY LNA Universal RT PCR system with pre- or custom-designed primers (Exiqon). Of note, our estimates show that 50ng of total RNA is sufficient to assay up to 130 miRNA species by qPCR. In addition, given the average RNA yield of 1000ng from laser microdissected tissue sections, and the availability of duplicate biopsies from each patient, we expect to have sufficient material for profiling miRNAs and coding RNAs in each of the premalignant lesions. ***Statistical:*** Given that we are performing this analysis on the same paired sample sets, our analytical approach also remains the same as described above, except that miRNAs showing at least 2-fold difference within pairs will be considered significant instead of using a 4-fold cutoff as proposed earlier. We will split our total EAC sample size of 150 pairs into a discovery set (n=75) and a validation set (n=75). We will

discover all differential miRs between all 150 EAC and nSQ pairs where miRNAs are altered with a ≤ 5% false-discovery rate (FDR) p-value, at least 2-fold variation within EAC and nSQ pairs, and in ≥ 50% of EACs in our discovery set (n=75). Assuming 80% power and a significance level of 0.05, we would need a minimum of 24 pairs in order to detect a 2-fold differential miR expression in 50% of the sample pairs [74], therefore we are more than adequately powered for this study. We will use the paired t-test to assess differential miR expression from qPCR analysis in our validation set of 75 pairs. Given that we are performing this analysis on the same paired sample sets, our statistical power for our paired tests to detect 1 in 10,000 as false positives remains the same. We will also look at the progression model in a pair-wise fashion, i.e. BE vs. HGD or HGD vs. EAC, describing the differences in the proportion of altered miRs using a one-sided Fishers exact test with the same statistical power as earlier given that we are using the same BE, HGD and EAC samples.

Analysis of adipokines and environmental factors -- Based on our prior study, [23] and reported BMI from patients in our existing BE database, we expect that the frequency of sustained obesity will be approximately 25 - 40% for cases compared to 10 - 20% for controls. Moreover, for analysis of IGF-1 levels and IGF-1R activation, the frequencies of risk exposures will depend on the quintiles used (i.e., if one compares the highest tertile of IGF-1 to others, the frequency will be 0.33). We plan analysis of 120 subjects in each group but will restrict the primary analysis to a homogeneous population of white males, which should comprise about 75% of the population or 90 subjects in each group. We will divide measurements of growth factors into tertiles, allowing for trend test for disease risk association. Thus, the proportion of cases that will have elevated insulin resistance index, chronic obesity, elevated IGF-1, and other "high risk" exposures we wish to explore is estimated to range from 0.20 to 0.40. Furthermore, we are interested in identifying effects that we anticipate will have odds ratios (ORs) between 2.0 and 3.0. We estimate the magnitude of this effect based on previous data regarding the relationship of obesity and esophageal adenocarcinoma[36-38]. In Table 4 below, sample sizes are estimated across risk exposure frequencies of 0.2 to 0.4, to detect minimum odds ratios of 2.0 to 2.5 with $\alpha = 0.05$, power = 80% ($\beta = 0.2$).

Table 4. Sample size calculations to detect expected Odds Ratios (OR) assuming specific "high risk" exposure frequencies

OR	**Frequency**		
	0.2	**0.3**	**0.4**
2.0	182	141	134
2.5	95	80	77
3.0	64	55	54

We have opted to use a categorical measurement because the relationship between serum measurements of IGF-1 and tissue affects is non-linear. Using a categorical variable allows us to test for trends and determine threshold effects of IGF-1. We propose to collect a sample size of 120 BE cases/120 EAC cases and 120 controls in each group and restrict the primary analysis to a homogeneous group for white males (ca. 90 cases vs. 90 controls in each control group) to answer our primary hypothesis. However, we plan to coninue accrual of cases and controls beyond 120 cases and 120 controls and also collect cases with dysplasia and cancer because we will be able to ask further questions regarding the importance of this pathway in progression to cancer, role of other pathways such as leptin and adiponectin, and also examine questions regarding gene environment interactions. **In this study, we should be able to detect the main effects of the risk exposures using our sample size assuming exposure frequencies are at least 0.2 and OR > 2. This same strategy will be employed for other adipokines such as adiponectin and leptin as appropriate.**

Simple descriptive statistics will be performed on all variables to define the study population and to describe the frequencies of risk factors in the cases (BE/EAC) and controls. All analyses will be performed separately for BE and EAC patients. Patients with EAC will be compared to Barrett's subjects as well as controls. Correlations among all variables will be examined to detect collinearity. Matched variables (age, gender, race, and GERD symptoms) will be compared between case and control groups to detect inadvertent differences. The crude association between insulin resistance, free IGF-1, total IGF-1, total IGF-2, IGFBP-1/IGFBP-3, IGF-1-R/Ki-67 immunohistostaining and case status will be estimated by the odds ratio with ninety-five percent confidence intervals derived from contingency tables. Continuous variables will be explored as such as well as grouped into tertiles. Highest tertiles of HOMA-IR, IGF-1, IGF-BP1/BP3 will be compared to the lowest tertiles. Odds ratios for the crude association of other major variables (BMI, waist-hip ratio,) with case status will be determined. We will control for confounding and explore effect modification. Multivariate logistic regression analysis will be conducted to describe the complex relationships between the covariates, case status, and insulin resistance.

We also propose to explore the relationship between diet and case status (BE/EAC). Controlled basic science experiments show that high carbohydrate intake, lack of caloric restriction, excess alcohol consumption, and deficiency of certain micronutrients hasten progression towards development of cancer[19,120]. Information regarding carbohydrate, fat, alcohol and micronutrient intake will be extracted at the University of Arizona from the Arizona Food Frequency questionnaire using validated computerized algorithms. Association between case status and various macro/micronutrients will be explored by univariate analysis. Factors with biological plausibility and significance by univariate analysis ($p<0.25$) will be included in the multivariate logistic regression equation modeling the relationship between case status, diet, obesity measures and insulin resistance.

10.0 PATIENT CONSENT AND PEER JUDGEMENT

All institutional, NCI, FDA, state and Federal regulations concerning informed consent and peer judgment will be fulfilled.

11.0 RECORDS TO BE KEPT

To ensure confidentiality for patients participating in the study, numbers will be assigned to each subject. The study coordinator assigns a number that is specific to this study and is used on all paperwork and RUCDR assigns its own number. Information that can be used to identify patients will be removed from all blood tubes and tumor samples and replaced with the assigned numbers. Each study coordinator and PI has access to participant identifiers and the assigned numbers at their respective institutions. Questionnaire information will be identified by a code number rather than participant name. All subject records (hard copy) will be stored in locked file cabinets at the institution where the patient was enrolled. All information (FBEQ data, pathology reports, surgical reports, endoscopy reports, pedigrees, and genetic test results) collected about a participant will be entered into Labmatrix, a secure electronic database. The information stored in the database is protected under HIPAA and all individuals with access to the information follow government regulations to protect the privacy of research study subjects. PHI for all study participants will be visible to Dr. Chak.

If participants are patients at the institution where they are recruited, results from genetic research and other information obtained during the study will not be placed in the patient's medical record. If genetic test results are confirmed in a CLIA laboratory it is appropriate to include those results in the patient's medical record.

12.0 REFERENCES

1. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. J Natl Cancer Inst 2005;97:142-6.
2. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin;60:277-300.
3. Chak A, Faulx A, Eng C, Grady W, Kinnard M, Ochs-Balcom H, Falk G. Gastroesophageal reflux symptoms in patients with adenocarcinoma of the esophagus or cardia. Cancer 2006;107:2160-6.
4. Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 1999;340:825-31.
5. Dulai GS, Guha S, Kahn KL, Gornbein J, Weinstein WM. Preoperative prevalence of Barrett's esophagus in esophageal adenocarcinoma: a systematic review. Gastroenterology 2002;122:26-33.
6. Devesa SS, Blot WJ, Fraumeni JF, Jr. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. Cancer 1998;83:2049-53.
7. Blot WJ, Devesa SS, Kneller RW, Fraumeni JF, Jr. Rising incidence of adenocarcinoma of the esophagus and gastric cardia. JAMA 1991;265:1287-9.
8. Pera M, Cameron AJ, Trastek VF, Carpenter HA, Zinsmeister AR. Increasing incidence of adenocarcinoma of the esophagus and esophagogastric junction. Gastroenterology 1993;104:510-3.
9. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ. Cancer statistics, 2008. CA Cancer J Clin 2008;58:71-96.
10. Cameron AJ, Lomboy CT, Pera M, Carpenter HA. Adenocarcinoma of the esophagogastric junction and Barrett's esophagus. Gastroenterology 1995;109:1541-6.
11. Haggitt RC, Tryzelaar J, Ellis FH, Colcher H. Adenocarcinoma complicating columnar epithelium-lined (Barrett's) esophagus. Am J Clin Pathol 1978;70:1-5.
12. Hameeteman W, Tytgat GN, Houthoff HJ, van den Tweel JG. Barrett's esophagus: development of dysplasia and adenocarcinoma. Gastroenterology 1989;96:1249-56.
13. Hirota WK, Loughney TM, Lazas DJ, Maydonovitch CL, Rholl V, Wong RK. Specialized intestinal metaplasia, dysplasia, and cancer of the esophagus and esophagogastric junction: prevalence and clinical data. Gastroenterology 1999;116:277-85.
14. Reid BJ, Blount PL, Rubin CE, Levine DS, Haggitt RC, Rabinovitch PS. Flow-cytometric and histological progression to malignancy in Barrett's esophagus: prospective endoscopic surveillance of a cohort. Gastroenterology 1992;102:1212-9.
15. Ruol A, Parenti A, Zaninotto G, Merigliano S, Costantini M, Cagol M, Alfieri R, Bonavina L, Peracchia A, Ancona E. Intestinal metaplasia is the probable common precursor of adenocarcinoma in barrett esophagus and adenocarcinoma of the gastric cardia. Cancer 2000;88:2520-8.
16. Sharma P, McQuaid K, Dent J, Fennerty MB, Sampliner R, Spechler S, Cameron A, Corley D, Falk G, Goldblum J, Hunter J, Jankowski J, Lundell L, Reid B, Shaheen NJ, Sonnenberg A, Wang K, Weinstein W. A critical review of the diagnosis and management of Barrett's esophagus: the AGA Chicago Workshop. Gastroenterology 2004;127:310-30.
17. Spechler SJ. Clinical practice. Barrett's Esophagus. N Engl J Med 2002;346:836-42.

18. Wang KK, Sampliner RE. Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus. Am J Gastroenterol 2008;103:788-97.
19. Cameron AJ, Lomboy CT. Barrett's esophagus: age, prevalence, and extent of columnar epithelium. Gastroenterology 1992;103:1241-5.
20. Shaheen N, Ransohoff DF. Gastroesophageal reflux, barrett esophagus, and esophageal cancer: scientific review. JAMA 2002;287:1972-81.
21. Wani S, Sharma P. The rationale for screening and surveillance of Barrett's metaplasia. Best Pract Res Clin Gastroenterol 2006;20:829-42.
22. Chak A, Lee T, Kinnard MF, Brock W, Faulx A, Willis J, Cooper GS, Sivak MV, Jr., Goddard KA. Familial aggregation of Barrett's oesophagus, oesophageal adenocarcinoma, and oesophagogastric junctional adenocarcinoma in Caucasian adults. Gut 2002;51:323-8.
23. Chak A, Ochs-Balcom H, Falk G, Grady WM, Kinnard M, Willis JE, Elston R, Eng C. Familiality in Barrett's esophagus, adenocarcinoma of the esophagus, and adenocarcinoma of the gastroesophageal junction. Cancer Epidemiol Biomarkers Prev 2006;15:1668-73.
24. Sun X, Elston R, Barnholtz-Sloan J, Falk G, Grady WM, Kinnard M, Mittal SK, Willis JE, Markowitz S, Brock W, Chak A. A segregation analysis of Barrett's esophagus and associated adenocarcinomas. Cancer Epidemiol Biomarkers Prev 2010;19:666-74.
25. Crabb DW, Berk MA, Hall TR, Conneally PM, Biegel AA, Lehman GA. Familial gastroesophageal reflux and development of Barrett's esophagus. Ann Intern Med 1985;103:52-4.
26. Eng C, Spechler SJ, Ruben R, Li FP. Familial Barrett esophagus and adenocarcinoma of the gastroesophageal junction. Cancer Epidemiol Biomarkers Prev 1993;2:397-9.
27. Fahmy N, King JF. Barrett's esophagus: an acquired condition with genetic predisposition. Am J Gastroenterol 1993;88:1262-5.
28. Gelfand MD. Barrett esophagus in sexagenarian identical twins. J Clin Gastroenterol 1983;5:251-3.
29. Jochem VJ, Fuerst PA, Fromkes JJ. Familial Barrett's esophagus associated with adenocarcinoma. Gastroenterology 1992;102:1400-2.
30. Poynton AR, Walsh TN, O'Sullivan G, Hennessy TP. Carcinoma arising in familial Barrett's esophagus. Am J Gastroenterol 1996;91:1855-6.
31. Prior A, Whorwell PJ. Familial Barrett's oesophagus? Hepatogastroenterology 1986;33:86-7.
32. Locke GR, Talley NJ, Weaver AL, Zinsmeister AR. A new questionnaire for gastroesophageal reflux disease. Mayo Clin Proc 1994;69:539-47.
33. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. Nat Genet 1999;21:15-9.
34. Yue H, Eastman PS, Wang BB, Minor J, Doctolero MH, Nuttall RL, Stack R, Becker JW, Montgomery JR, Vainer M, Johnston R. An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. Nucleic Acids Res 2001;29:E41-1.
35. Chak A, Faulx A, Kinnard M, Brock W, Willis J, Wiesner GL, Parrado AR, Goddard KA. Identification of Barrett's esophagus in relatives by endoscopic screening. Am J Gastroenterol 2004;99:2107-14.
36. Ochs-Balcom HM, Falk G, Grady WM, Kinnard M, Willis J, Elston R, Eng C, Chak A. Consortium approach to identifying genes for Barrett's esophagus and esophageal adenocarcinoma. Transl Res 2007;150:3-17.

37. Ploughman LM, Boehnke M. Estimating the power of a proposed linkage study for a complex genetic trait. Am J Hum Genet 1989;44:543-51.
38. Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J. Effects of misspecifying genetic parameters in lod score analysis. Biometrics 1986;42:393-9.
39. Vieland VJ, Hodge SE, Greenberg DA. Adequacy of single-locus approximations for linkage analyses of oligogenic traits. Genet Epidemiol 1992;9:45-59.
40. Risch N, Giuffra L. Model misspecification and multipoint linkage analysis. Hum Hered 1992;42:77-92.
41. Ott J. Linkage probability and its approximate confidence interval under possible heterogeneity. Genet Epidemiol Suppl 1986;1:251-7.
42. Lander ES, Green P. Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci U S A 1987;84:2363-7.
43. Idury RM, Elston RC. A faster and more general hidden Markov model algorithm for multipoint likelihood calculations. Hum Hered 1997;47:197-202.
44. Elston RC, Stewart J. A general model for the genetic analysis of pedigree data. Hum Hered 1971;21:523-42.
45. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 1997;61:1179-88.
46. Markianos K, Daly M, Kruglyak L. Efficient Multipoint Linkage Analysis through Reduction of Inheritance Space. Am J Hum Genet 2001;68:963-977.
47. Weeks DE, Lange K. The affected-pedigree-member method of linkage analysis. Am J Hum Genet 1988;42:315-26.
48. Olson JM. A general conditional-logistic model for affected-relative-pair linkage studies. Am J Hum Genet 1999;65:1760-9.
49. Goddard KA, Witte JS, Suarez BK, Catalona WJ, Olson JM. Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. Am J Hum Genet 2001;68:1197-206.
50. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, Vogelstein B, Karchin R. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. Cancer Res 2009;69:6660-7.
51. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 2009;106:19096-101.
52. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet;42:790-3.
53. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet;42:30-5.
54. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009;461:272-6.
55. Hao Y, Sood S, Triadafilopoulos G, Kim JH, Wang Z, Sahbaie P, Omary MB, Lowe AW. Gene expression changes associated with Barrett's esophagus and Barrett's-associated adenocarcinoma cell lines after acid or bile salt exposure. BMC Gastroenterol 2007;7:24.

56. Hao Y, Triadafilopoulos G, Sahbaie P, Young HS, Omary MB, Lowe AW. Gene expression profiling reveals stromal genes expressed in common between Barrett's esophagus and adenocarcinoma. Gastroenterology 2006;131:925-33.
57. Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zollner S. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. Am J Hum Genet;87:604-17.
58. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 2009;5:e1000384.
59. Jones S, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, Traulsen A, Nowak MA, Siegel C, Velculescu VE, Kinzler KW, Vogelstein B, Willis J, Markowitz SD. Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci U S A 2008;105:4283-8.
60. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. Science 2006;314:268-74.
61. Sun X, Elston R, Barnholtz-Sloan J, Falk G, Grady WM, Kinnard M, Mittal SK, Willis JE, Markowitz S, Brock W, Chak A. A Segregation Analysis of Barrett's Esophagus and Associated Adenocarcinomas. Cancer Epidemiol Biomarkers Prev.
62. Schnell AH, Elston RC, Hull PR, Lane PR. Major gene segregation of actinic prurigo among North American Indians in Saskatchewan. Am J Med Genet 2000;92:212-9.
63. Thirlwell C, Eymard M, Feber A, Teschendorff A, Pearce K, Lechner M, Widschwendter M, Beck S. Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. Methods;52:248-54.
64. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics;11:587.
65. Jung SH. Sample size for FDR-control in microarray data analysis. Bioinformatics 2005;21:3097-104.
66. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP, Verhaak RG, Hoadley KA, Hayes DN, Perou CM, Schmidt HK, Ding L, Wilson RK, Van Den Berg D, Shen H, Bengtsson H, Neuvial P, Cope LM, Buckley J, Herman JG, Baylin SB, Laird PW, Aldape K. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell;17:510-22.
67. Engreitz JM, Daigle BJ, Jr., Marshall JJ, Altman RB. Independent component analysis: mining microarray data for fundamental human gene expression modules. J Biomed Inform;43:932-44.
68. Prasad GA, Bansal A, Sharma P, Wang KK. Predictors of progression in Barrett's esophagus: current knowledge and future directions. Am J Gastroenterol;105:1490-1502.
69. Prasad GA, Bansal A, Sharma P, Wang KK. Predictors of progression in Barrett's esophagus: current knowledge and future directions. Am J Gastroenterol 2010;105:1490-1502.
70. Wei C, Li J, Bumgarner RE. Sample size for detecting differentially expressed genes in microarray experiments. BMC Genomics 2004;5:87.

71. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 2008;18:610-21.
72. Reid JG, Nagaraja AK, Lynn FC, Drabek RB, Muzny DM, Shaw CA, Weiss MK, Naghavi AO, Khan M, Zhu H, Tennakoon J, Gunaratne GH, Corry DB, Miller J, McManus MT, German MS, Gibbs RA, Matzuk MM, Gunaratne PH. Mouse let-7 miRNA populations exhibit RNA editing that is constrained in the 5'-seed/ cleavage/anchor regions and stabilize predicted mmu-let-7a:mRNA duplexes. Genome Res 2008;18:1571-81.
73. Hackenberg M, Sturm M, Langenberger D, Falcon-Perez JM, Aransay AM. miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 2009;37:W68-76.
74. DNA Microarrays, Part B: Databases And Statistics In: Kimmel A, Oliver B, eds. Methods in Enzymology: Elsevier, 2006.

APPENDIX A

Letter – To be sent to study participants previously enrolled under separate protocols at University Hospitals Case Medical Center and Cleveland Clinic

<Date>

Dear Research Participant,

Thank you for contributing to research investigating the familial factors involved in Barrett esophagus and esophageal cancer. Since you were enrolled in the study <site, IRB# and name> the Case Comprehensive Cancer Center Institutional Review Board (CCCC IRB) was established. This is a new institution shared by Case Western Reserve University, University Hospitals Case Medical Center, and the Cleveland Clinic that facilitates cancer related research at these organizations.

For the ease of administration the current study in which you are enrolled at <site> has been rolled into a new protocol at the CCCC IRB. Please note that this was not due to any research infractions or change in focus of the research, but to simplify the research process for the investigators researching your disease.

If you have any questions or concerns about this change you may contact <study coordinator> at <phone number>. As originally indicated, once we have completed this study we will send you a letter summarizing the overall results. Thank you for your understanding.

Sincerely,

<Investigator>
<Department>
<Site>

APPENDIX C

Study Advertisement – Available on the Clinical Trials website
http://clinicaltrials.gov/ct/show/NCT00288119;jsessionid=DF1EF971B09134778AF05AC055E37AB3?order=26

Information was too large to paste into this document. Please see the website above, or, if preferred, we can mail a hard copy of the website.

APPENDIX D

Study Advertisement - Available on the University Hospitals Case Medical Center website
http://www.uhhospitals.org/case/tabid/1133/ClinicalTrialID/30/Default.aspx

Clinical Trial Listing

Trial Name: **Barrett's Esophagus**

Condition:

Description: This study is designed to determine whether there is an increased incidence of Barrett's esophagus or esophageal cancer in the relatives of persons with Barrett's esophagus or esophageal cancer. The Division of Gastroenterology is contacting siblings, children, and parents of patients with Barrett's esophagus or esophageal cancer, who have had symptoms of gastroesophageal reflux disease (GERD) or heartburn. Using a very thin scope, we can examine the esophagus to check for Barrett's without the need for sedation. This procedure takes no more than 10 minutes.

Criteria: Participants should be parents or siblings of patients with Barrett's esophagus or esophageal cancer.

Contact: For information or to register for screening in this trial please call HealthMatch at (216) 844-5000, or complete and submit the form below.

Participant Confidentiality:
All information regarding participation in Clinical Trials is treated with the utmost confidentiality. University Hospitals Case Medical Center and the Division of Clinical Research recognize and respect the confidentiality of this information. Access to information related to clinical trials and their participants is restricted to authorized clinical trials personnel only.

APPENDIX E

Letter – To be included in packets given to the proband. This information is provided to prospective participants following a complete discussion about the study and obtaining informed consent by phone or in person.

RE: Familial Barrett Esophagus and Esophageal Adenocarcinoma Study

Investigators at the Cleveland Clinic, University Hospitals Case Medical Center, the Veterans Affairs Medical Center at Wade Park, the University of Washington, Fred Hutchinson Cancer Research Center, Creighton University, Mayo Clinic, Johns Hopkins Hospital, University of North Carolina-Chapel Hill, Washington University at St. Louis and the Hospital of University of Pennsylvania are investigating the familial factors involved in Barrett esophagus and esophageal cancer. Barrett esophagus is a pre-cancerous condition that predisposes someone to developing esophageal cancer. Certain families may be more likely to develop these conditions than others.

With the assistance of your physician, we are contacting individuals who have been seen at <site> in the past. You have been identified though our records as having a <personal and/or family> history of Barrett esophagus and/or esophageal cancer. We would like to recruit you to be part of our study. You are under no obligation to participate in this study; this is completely voluntary.

2 copies of the consent form
1 FBE Questionnaire
1 medical release of information form
1 self-address, postage pre-paid envelope

Please read and sign the consent form and complete the questionnaire. If you have been diagnosed with Barrett esophagus and/or esophageal cancer please complete the medical release of information form. The dates of treatment on the medical release form should indicate the time you were initially diagnosed until <now/end of treatment >.

Hopefully with your assistance, we may be able to learn more about Barrett esophagus and be able to diagnose and treat people and families with this disease in order to prevent esophageal cancer. If you have any questions about this research or the enclosed materials please contact <study coordinator> of <site> at <phone number>. We will be happy to answer any questions that you may have. Once the materials have been completed and returned a study investigator will contact you. We look forward to speaking with you in the near future and thank you for your time and consideration.

Sincerely,

<Investigator>
<Department>
<Site>

APPENDIX F

Letter – To be sent to family members whom we have permission from the proband to contact

Investigators at the Cleveland Clinic, University Hospitals Case Medical Center, the Veterans Affairs Medical Center at Wade Park, the University of Washington, Fred Hutchinson Cancer Research Center, Creighton University, Mayo Clinic, Johns Hopkins Hospital, University of North Carolina-Chapel Hill, Washington University at St. Louis and The Hospital of University of Pennsylvania are investigating the familial factors involved in Barrett esophagus and esophageal cancer. Barrett esophagus is a pre-cancerous condition that predisposes someone to developing esophageal cancer. Certain families may be more likely to develop these conditions than others.

You have a family member who has been diagnosed with Barrett esophagus and/or esophageal cancer. He or she told us that you may be interested in learning more about our research and gave us permission to contact you.

With the assistance of your family member, we are contacting you to provide you with information about our research study. You are under no obligation to participate in this study; this is completely voluntary.

We have included the following materials:

An information sheet in the form of Frequently Asked Questions
2 copies of the consent form
1 FBE Questionnaire
1 medical release of information form
1 self-address, postage pre-paid envelope

Please read and sign the consent form and complete the questionnaire. If you have been diagnosed with Barrett esophagus and/or esophageal cancer please complete the medical release of information form. The dates of treatment on the medical release form should indicate the time you were initially diagnosed until <now/end of treatment >.

Hopefully with your assistance, we will learn more about Barrett esophagus. Having a better understanding of this condition will improve our ability to diagnose and treat people and families with this disease, with the ultimate goal of preventing esophageal cancer. If you have any questions about this research or the enclosed materials please contact <study coordinator> of <site> at <phone number>. If we do not hear from you within the next month, we will call you to follow up. Once the materials have been completed and returned a study investigator will contact you to discuss how to proceed. We look forward to speaking with you in the near future and thank you for your time and consideration.

Sincerely,

<Investigator>
<Department>
<Site>

APPENDIX G

AUTHORIZATION FOR RELEASE OF MEDICAL INFORMATION

To request medical records for yourself or your living relatives, this form should be completed by the individual for whom records are requested. For deceased relatives, this form should be completed by the individual's next of kin or the executor of the estate.

Medical records are requested for:

_____born on ____/____/____,
Last Name First Name Middle Initial Mo. Day Year

S.S.#_____/____/_____, who is: (check one) ☐Living ☐Deceased**.**

If deceased, please include date of death**.** ____/____/____
Mo. Day Year

From:

- ☐ The Cleveland Clinic Foundation
- ☐ Case / UH
- ☐ Clinic
- ☐ Other [Provide the name, city, and state of the hospital(s) where treated.]

_____

_____

_____

_____

The following medical information regarding my hospitalization, care and/or treatment on the following dates: _____to _____as an: ☐ Inpatient ☐ Outpatient ☐ Emergency Dept.

Please provide my:

☐ Surgical report ☐ Pathology report ☐ Chart summary ☐ Paraffin-embedded tumor block & matching H & E slide

Purpose of Disclosure: CONFIRMATION/DOCUMENTATION

Please release and furnish to: <Investigator>
<Department>
<Mailing Address>
Cleveland, OH <Zip>

I understand and acknowledge that this authorization extends to all or any part of the information designated above, which may include treatment for physical and mental illness, and/or alcohol/drug abuse, and/or AIDS (Acquired Immunodeficiency Syndrome), and/or may include the results of an HIV test or the fact that an HIV test was performed. Information in the form of

audio, photo or video has been designated above if applicable. I expressly consent to the release of information designated above. This consent is valid for 60 days, unless revoked by my written notice, provided said notice is received prior to release of the above designated information.

_____ _____

Signature of Patient or Next of Kin and relationship to the patient — Date

_____ _____

Witness — Date

APPENDIX H

Letter – To be sent to family members referred for standard upper endoscopy

Thank you for participating in our study of familial factors in the development of Barrett esophagus. As you know, you have been identified as a close relative of a person with Barrett esophagus or esophageal cancer. In an attempt to determine if certain families are more likely to get Barrett esophagus we are evaluating the data from the questionnaires you returned. You have indicated that you have heartburn symptoms or other symptoms of gastroesophageal reflux disease (GERD), which is associated with the development of Barrett esophagus. Thus, we are recommending that you be referred to a gastroenterologist for further evaluation and therapy. We would be happy to see you here in Cleveland, however, if this is not convenient for you or for insurance coverage reasons, you should attempt to obtain a referral to a gastroenterologist in your area.

The gastroenterologist will evaluate your symptoms and inquire about your family history and determine whether further workup is necessary. Medications that effectively treat GERD may be prescribed. If you do undergo an upper endoscopy, we would like to be sent a report of the results for our study. You will need to initiate a request for the endoscopy report with your doctor's office and may need to sign a release form.

Please call <study coordinator> at <phone number> for further information or to schedule an appointment. You may find it helpful to bring this letter to your primary doctor or gastroenterologist to facilitate this process.

Sincerely,

<Investigator>
<Department>
<Site>

APPENDIX I

Letter – To be sent to family members eligible for UUE

Thank you for participating in our study of familial factors in the development of Barrett esophagus. As you know, you have been identified as a relative of a person with Barrett esophagus or esophageal cancer. In an attempt to determine if certain families are more likely to get Barrett esophagus we are evaluating the data from the questionnaire we have sent out. You have indicated that you have no or limited heartburn symptoms. At this time there is no recognized standard indication to examine your esophagus with an endoscope. However, based on your positive family history of Barrett esophagus and/or esophageal cancer there is some evidence that you may be at increased risk for developing Barrett esophagus. As part of this study we are offering an ultrathin endoscopy to evaluate for Barrett esophagus without the use of sedation, thus minimizing the risk of the standard procedure. We would like to offer this option to you.

We realize that travel to the <site> may not be convenient, but if you are interested in receiving a free endoscopy or if you would like any additional information, we would like to hear from you. You should also inform your primary physician about your family history.

Please call <study coordinator> at <phone number> for further information or to schedule an appointment.

Thank you, again.

Sincerely,

<Investigator>
<Department>
<Site>