Germline variation in inflammation-related pathways and risk of Barrett's esophagus and esophageal adenocarcinoma

Matthew F. Buas[1], Qianchuan He[1], Lisa G. Johnson[1], Lynn Onstad[1], David M. Levine[2], Aaron P. Thrift[3], Puya Gharahkhani[4], Claire Palles[5], Jesper Lagergren[6,7], Rebecca C. Fitzgerald[8], Weimin Ye[9], Carlos Caldas[10], Nigel C. Bird[11], Nicholas J. Shaheen[12], Leslie Bernstein[13], Marilie D. Gammon[14], Anna H. Wu[15], Laura J. Hardie[16], Paul D. Pharoah[17], Geoffrey Liu[18], Prassad Iyer[19], Douglas A. Corley[20,21], Harvey A. Risch[22], Wong-Ho Chow[23], Hans Prenen[24], Laura Chegwidden[25], Sharon Love[26], Stephen Attwood[27], Paul Moayyedi[28], David MacDonald[29], Rebecca Harrison[30], Peter Watson[31], Hugh Barr[32], John deCaestecker[33], Ian Tomlinson[5], Janusz Jankowski[34], David C. Whiteman[35], Stuart MacGregor[4], Thomas L. Vaughan[1,36], Margaret M. Madeleine[1,36*]

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

[2]Department of Biostatistics, University of Washington, School of Public Health, Seattle, Washington, USA.

[3]Department of Medicine and Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA

[4]Queensland Institute of Medical Research Berghofer Medical Research Institute Brisbane, Queensland, Australia.

[5]Wellcome Trust Centre for Human Genetics and NIHR Comprehensive Biomedical Research Centre, University of Oxford, Oxford, UK.

[6]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden.

[7]Division of Cancer Studies, King's College London, United Kingdom.

[8]Medical Research Council (MRC) MRC Cancer Unit, Hutchison-MRC Research Centre, University of Cambridge, Cambridge, UK.

[9]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[10]Cancer Research UK, Cambridge Institute, Cambridge, UK.

[11]Department of Oncology, Medical School, University of Sheffield, Sheffield, UK.

[12]Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA.

[13]Department of Population Sciences, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, California, USA.

[14]Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA.

[15]Department of Preventive Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA.

[16]Division of Epidemiology, University of Leeds, Leeds, UK.

[17]Department of Oncology, University of Cambridge, Cambridge, UK ; Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

[18]Pharmacogenomic Epidemiology, Ontario Cancer Institute, Toronto, Ontario, Canada M5G 2M9.

[19]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota, USA.

[20]Division of Research, Kaiser Permanente Northern California, Oakland, California, USA.

[21]San Francisco Medical Center, Kaiser Permanente Northern California, San Francisco, California, USA.

[22]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA

[23]Department of Epidemiology, MD Anderson Cancer Center, Houston, TX, USA

[24]Department of Digestive Oncology, University Hospital Gasthuisberg, Leuven, Belgium

[25]University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

[26]Centre for Statistics in Medicine and Oxford Clinical Trials Research Unit, Oxford, UK

[27]Department of General Surgery, North Tyneside General Hospital, North Shields, UK

[28]Department of Medicine, McMaster University, Hamilton, Ontario, Canada

[29]Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, British Columbia, Canada

[30]Department of Pathology, Leicester Royal Infirmary, Leicester, UK

[31]Department of Medicine, Institute of Clinical Science, Royal Victoria Hospital, Belfast, UK

[32]Department of Upper GI Surgery, Gloucestershire Royal Hospital, Gloucester, UK

[33]Department of Gastroenterology, Leicester General Hospital, Leicester, UK

[34]UClan Medical School, West Lakes Campus, Rm 127 Lindow Centre, Cumbria, England, CA24 3JY

[35]Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane, Australia.

[36]Department of Epidemiology, University of Washington, School of Public Health, Seattle, Washington, USA.

The authors disclose no potential conflicts of interest.

Running title: Inflammation-related germline variation and risk of BE/EA

***Correspondence to** Margaret M. Madeleine, E-mail: mmadelei@fredhutch.org, Phone: (206) 667-4630, Fax: (206) 667-5948

Keywords: Barrett's esophagus, esophageal adenocarcinoma, inflammation, post-GWAS

Word Counts:

Abstract: 245

Article: 3,886

Abbreviations:

EA, esophageal adenocarcinoma

BE, Barrett's esophagus

GERD, gastroesophageal reflux disease

COX, cyclooxygenase

Abstract

Objective:

Esophageal adenocarcinoma (EA) incidence has risen sharply in Western countries over recent decades. Local and systemic inflammation is considered an important contributor to EA pathogenesis. Established risk factors for EA and its precursor, Barrett's esophagus (BE), include symptomatic reflux, obesity, and smoking. The role of inherited genetic susceptibility remains an area of active investigation. Here, we explore whether germline variation related to inflammatory processes influences susceptibility to BE/EA.

Design:

We used data from a genome-wide association study (GWAS) of 2,515 EA cases, 3,295 BE cases, and 3,207 controls. Our analysis included 7,863 single nucleotide polymorphisms (SNPs) in 449 genes assigned to five pathways: cyclooxygenase (COX), cytokine signaling, oxidative stress, human leukocyte antigen, and NFκB. A principal components-based analytic framework was employed to evaluate pathway-level and gene-level associations with disease risk.

Results:

We identified a significant signal for the COX pathway in relation to BE risk (P=0.0059, FDR q=0.03), and in gene-level analyses found an association with *MGST1* (microsomal glutathione-S-transferase 1; P=0.0005, q=0.005). Assessment of 36 *MGST1* SNPs identified 14 variants associated with elevated BE risk (q<0.05). Four of these were subsequently confirmed ($P<5.5 \times 10^{-5}$) in a meta-analysis encompassing an independent set of 1,851 BE cases and 3,496 controls, and are known strong eQTLs for *MGST1*. Three such variants were associated with similar elevations in EA risk.

Conclusion:

This study provides the most comprehensive evaluation of inflammation-related germline variation in relation to risk of BE/EA, and suggests that variants in *MGST1* influence disease susceptibility.

1. What is already known about this subject: 3-4 bullet points

- Local and systemic inflammation are considered important contributors to the pathogenesis of Barrett's esophagus (BE) and esophageal adenocarcinoma (EA)
- Inherited genetic variation in inflammation-related pathways and genes may modulate the functional intensity of inflammatory signaling networks
- Small-scale candidate gene-based studies suggest possible associations between inflammation-related genetic variation and altered risk of BE/EA

2. What are the new findings: 3-4 bullet points

- Germline variation in the cyclooxygenase (COX) pathway is associated with altered risk of BE, based on analysis of the largest-available consortium GWAS dataset for this condition
- The pathway-level association signal for COX and risk of BE appears to be driven in part by variation at the microsomal glutathione S-transferase 1 (*MGST1*) gene locus
- Several intronic SNPs with strong regulatory potential (eQTLs) at the *MGST1* locus exhibit significant associations with risk of BE, with confirmation in a large independent sample set

3. How might it impact on clinical practice in the foreseeable future?

Our data identify the *MGST1* gene as a novel susceptibility locus for Barrett's esophagus. These findings further enhance our understanding of the biological pathways that likely underlie differential susceptibility to this cancer precursor lesion in human populations. Such insights will contribute to the ongoing effort to develop comprehensive population-based risk assessment tools for BE that integrate behavioral, clinical, and genetic data to identify and target individuals at highest risk of disease for preventive interventions.

Introduction

The incidence of esophageal adenocarcinoma (EA) has risen rapidly over recent decades in Western countries [1, 2]. EA typically arises within a metaplastic precursor epithelium known as Barrett's esophagus (BE) [3]. Established risk factors for EA and BE include symptomatic gastroesophageal reflux disease (GERD), abdominal adiposity, tobacco smoking, European ancestry, and male sex [3, 4, 5]. A prevailing conceptual model has linked chronic inflammation and genomic instability to EA pathogenesis [3]. Several exposures associated with elevated disease risk, such as GERD, obesity, and smoking, increase levels of local and systemic inflammation, while use of non-steroidal anti-inflammatory drugs (NSAIDs) and statins has been associated with reduced risk [6, 7, 8]. It remains poorly understood, however, whether and to what extent inherited genetic variation in specific genes and pathways implicated in inflammatory signaling may modulate disease susceptibility and interact with these established risk factors.

A biologic link between chronic inflammation and cancer risk has long been appreciated [9, 10]. Inflammation may act at multiple stages of disease development to disrupt tissue homeostasis, induce aberrant proliferative responses, modulate the tumor microenvironment, and compromise immune surveillance [11, 12]. Inflammatory physiologic changes such as oxidative stress are known to exert downstream genotoxic effects [13], and when sustained over extended periods, can promote the emergence of cancer-initiating mutations. In the esophagus, long-term exposure to gastric acid or bile salts results in the release of pro-inflammatory cytokines (e.g., interleukin-8), activation of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB) and cyclooxygenase-2 (COX2), and direct tissue damage to the squamous epithelium [14, 15, 16]. Cigarette smoking can also expose the esophagus to deleterious toxins while simultaneously inducing systemic inflammatory responses based on activation of cytokine signaling, NFκB activation, and COX pathway stimulation [17, 18, 19]. Abdominal adiposity and obesity have been associated with elevated circulating levels of pro-inflammatory mediators such as tumor necrosis factor-α (TNFα), C-reactive protein (CRP), interleukin-6 (IL-6), and leptin [20].

Recent large-scale GWAS have provided comprehensive assessments of genetic susceptibility to BE and EA [21, 22, 23, 24, 25]. Novel associations have been identified with variants in or near several transcription factors implicated in embryonic esophageal development, a transcriptional co-activator, and

the human leukocyte antigen (HLA) region. It remains likely, however, that additional loci that did not satisfy the commonly used, stringent statistical threshold ($p<5x10^{-8}$) may be involved in modifying disease risk. In this regard, pathway-based analytic methods can offer significant advantages over conventional genome-wide analyses, by simultaneously reducing the number of statistical comparisons and increasing power via aggregating large numbers of low-magnitude signals [26]; such methods allow for the systematic analysis of coherent biological processes most likely implicated in disease etiology.

Given the central role of inflammation in BE and EA pathogenesis, we examined genetic variation in five inflammation-related pathways—COX, cytokine signaling, oxidative stress, HLA, and NFκB—using a principal components analysis (PCA)-based framework. Using genotyping data from the International Barrett's Esophagus and Adenocarcinoma Consortium (BEACON) GWAS of 2,515 EA cases, 3,295 BE cases, and 3,207 controls, we selected 7,863 SNPs in 449 genes and assessed associations with risks of BE and EA in a pre-specified tiered fashion, first at the pathway level, next at the gene level, and ultimately at the SNP level.

Methods

Study population and SNP genotyping

The BEACON GWAS included individuals diagnosed with EA or BE, and control participants pooled from 14 individual studies conducted in Western Europe, Australia, and North America, as described previously [21]. The current analysis employed a pooled dataset [27] that included participants of European ancestry from the BEACON GWAS, additional BE and EA patients from the UK Barrett's Esophagus Gene Study and the UK Stomach and Oesophageal Cancer Study (SOCS), respectively [21], and additional control participants from a hospital-based case-control study of melanoma conducted at the MD Anderson Cancer Center (Houston, TX) [28]. Genotyping of buffy coat or whole blood DNA from all participants was conducted using the Illumina Omni1M Quad platform, in accordance with standard quality control procedures [29]. All participants gave written informed consent, and this project was approved by the ethics review board of the Fred Hutchinson Cancer Research Center. We selected all unrelated participants with <2% missing genotyping calls, as described previously [21]. The final study

sample included 2,515 EA cases, 3,295 BE cases, and 3,207 controls. Three control participants were excluded from analyses involving BE cases, because of familial relation to cases.

Selection of genes in inflammation-related pathways

Five pathways implicated in chronic inflammation were selected for analysis: 1) cyclooxygenase (COX) (n=40 genes), 2) pro- and anti-inflammatory cytokines (n=198 genes), 3) oxidative stress (n=117 genes), 4) HLA (32 genes), and 5) NF-kB (n=125 genes). Selected genes (**Table S1**) were identified based on an extensive survey of the prior literature on inflammation in cancer and EA pathogenesis [11, 30, 31, 32, 33, 34], and as described in public databases (eg. KEGG, Biocarta).

SNP selection

SNPs selected for this study are located in or near (+/- 2.0 kilobases) the genes chosen for analysis. We included only those SNPs that passed Illumina quality measures and standard quality control procedures [29], as previously described [21], and had a minor allele frequency (MAF) of ≥1%. Imputation of missing values for genotyped SNPs was conducted using SHAPEIT [35]. After imposing the above filters, we identified all available Omni1M SNPs located within the selected genes. Segments of 2.0 kb of flanking sequence proximal to the transcriptional start sites and distal to the 3'UTRs were also included, based on gene boundaries defined in hg19/GRChB37. No Omni1M SNPs were available for 16 genes initially selected (cytokines: n=14, oxidative stress: n=2) (**Table S1**). Minor and major alleles were reported throughout using the 'plus' strand designation.

Statistical analysis

We examined each of the five inflammation-related pathways using an application of principal components analysis (PCA) [36] (**Figure S1**). We first constructed a genotype matrix comprising all SNPs assigned to the indicated pathway, inclusive of case patients of the selected type (BE or EA) and all control participants. Individual SNP variables, coded as 0, 1, or 2 minor alleles, were standardized across participants to have a mean of zero and standard deviation (SD) of one. The first N principal components (PCs) that captured ≥50% of the genotypic variance of the pathway were selected. Association between a

given pathway and risk of BE or EA was assessed using the likelihood ratio test (LRT). Two logistic regression models were compared: i) a full model containing N pathway-level PCs ($PC_{1,p}$…$PC_{N,p}$), age, sex, and the first four PCs derived from ancestry-informative markers (AIM) to account for population stratification ($PC_{1,AIM}$-$PC_{4,AIM}$) [27]; and ii) a reduced model containing only age, sex and $PC_{1,AIM}$-$PC_{4,AIM}$. HLA loci were excluded from the set of ancestry-informative markers, as described previously [21]. We selected pathways for which the resulting LRT P value was <0.05, after correction for multiple comparisons (n=5) via the false discovery rate method (FDR) (**Figure S1**, step 1).

To prioritize genes within a selected pathway for further analysis at the gene level, we examined SNP loading factors within the first pathway-level principal component ($PC_{1,p}$). SNPs within $PC_{1,p}$ were rank-ordered by the absolute values of their loading coefficients. The first ten genes represented by these rank-ordered SNPs were advanced to gene-level analysis (**Figure S1**, step 2). PCA was conducted for each of these genes using a genotype matrix comprised of all SNPs assigned to the indicated gene; the first N PCs that captured ≥50% of the genotypic variance were selected. Association between a given gene and risk of BE or EA was assessed as above using the LRT, comparing i) a full model inclusive of the selected gene-level PCs ($PC_{1,g}$-$PC_{N,g}$), age, sex, and $PC_{1,AIM}$-$PC_{4,AIM}$; and ii) a reduced model containing age, sex, and $PC_{1,AIM}$-$PC_{4,AIM}$. Multiple comparisons (n=10) were accounted for via the FDR method (**Figure S1**, step 3).

Genes satisfying FDR q<0.05 were selected for additional analysis at the SNP level. Unconditional logistic regression was used to compute odds ratios (ORs) for risk of BE or EA associated with a given SNP variant, under an additive model (per-allele) with adjustment for age, sex, and $PC_{1,AIM}$-$PC_{4,AIM}$, and correcting for multiple comparisons via the FDR method (**Figure S1**, step 4). Observed associations were visualized graphically using LocusZoom [37].

Gene-environment interactions were investigated using a subset of the overall study population for which exposure data were available; BE/EA case patients from SOCS and the UK BE Gene Study, and control participants from MD Anderson were excluded. Reflux symptoms, BMI, smoking history, and NSAID use were defined as follows: reflux/heartburn: ≥weekly symptoms (yes or no); BMI: <25, 25-29, 30-34, or 35+; smoking: ever or never, or pack-years: 0, 1-14, 15-29, 30-44, or 45+; NSAID use: ever or never. Pack-year smoking history and BMI were modeled as continuous covariates in tests for interaction.

All PCA-based statistical analyses, as well as SNP-based tests and interaction studies, were conducted using STATA/SE version 14 (College Station, TX).

Validation studies

An independent dataset comprised of 1,851 BE patients and 3,496 control participants from the UK, described previously [22], was used for validation studies. Summary statistics for the associations of 13 genotyped SNPs at the *MGST1* locus and risk of BE were extracted and used in a subsequent meta-analysis based on the inverse-variance weighting method [38]. Validation analyses were conducted in R v3.2.1.

Results

Characteristics of study participants

The distributions of demographic and behavioral characteristics among control participants, BE case patients, and EA case patients are shown in **Table 1**. EA cases were somewhat older and more often male compared to controls and BE cases. The percentage reporting ever having smoked cigarettes was higher among BE and EA cases than among controls, and heavy smoking (45+ pack years) was more prevalent among EA cases. Obesity (BMI 30+) and weekly reflux/heartburn were more prevalent among BE and EA cases than among controls. NSAID use appeared similarly common across the three groups.

Pathway-level associations with risk of BE or EA

To obtain a top-level, global assessment of the association between germline variation within five selected inflammation-related pathways (COX, cytokine signaling, oxidative stress, HLA, and NFκB) and risk of BE or EA, we employed a PCA-based approach (**Figure S1**). Based on logistic regression analyses that incorporated a subset of the derived principal components as predictor variables and assessed associations with disease risk, we identified a single significant ($P<0.05$) pathway-level signal for risk of BE: the COX pathway ($P=0.006$) (**Table 2**). This association remained significant after

accounting for multiple comparisons (FDR q=0.03). None of the five pathways examined were found to be associated (P<0.05) with risk of EA.

Gene-level associations with risk of BE

To determine whether the observed pathway-level signal for COX could be further localized to particular individual genes, we undertook gene-level analyses using the same PCA framework (**Figure S1**). Of the 40 genes assigned to the COX pathway, we prioritized 10 for further analysis, as pre-specified in our analysis plan. The top 10 genes were selected based on their contribution to the overall pathway-level genotypic variance, as reflected in rank-ordered SNP loading coefficients in the first principal component. Among these 10 genes assessed for associations with risk of BE (**Table 3**), only a single gene exhibited a significant signal: microsomal glutathione S-transferase 1 (*MGST1*) (P=0.0005, FDR q=0.005). A borderline-significant (P=0.07) association was observed for gene-level variation in *MGST1* and risk of EA (**Table S2**).

SNP-level associations with risk of BE

Individual SNPs located within or in proximity to (± 2.0 kb) the *MGST1* locus were assessed for associations with risk of BE. Among 36 such variants examined, 14 exhibited a significant signal (P<0.05, FDR q<0.05) (**Table 4**). The minor alleles at all 14 SNPs were associated with elevated risk of BE, with ORs ranging in magnitude from 1.10-1.38. The most significant association was for rs4149203 C>T (OR=1.16, P=9.0 x 10^{-5}, q=0.001). A LocusZoom plot of the 36 assessed SNPs revealed a cluster of six associated 3' variants in high linkage disequilibrium (LD, r^2>0.8) with rs4149203 (**Figure 1**). A second cluster of six SNPs satisfying FDR q<0.05 was situated at the 5' end of the *MGST1* locus (**Figure S2**); modest to moderate LD was observed between rs2239676, the top-ranked SNP in this second region, and the other five variants in close proximity. Among the 14 significant SNP-level signals identified for BE, eight were also associated with increased risk of EA (P<0.05, q<0.05), with observed ORs ranging from 1.10 to 1.17 (**Table S3**).

Assessment of top *MGST1* SNPs and risk of BE in an independent study sample

We next evaluated whether any of the 14 *MGST1* variants associated with risk of BE showed similar associations in a large, independent sample set from the UK comprised of 1,851 BE patients and 3,496 control participants. 13 of the 14 SNPs were available for analysis, and four 3' variants exhibited borderline-significant (P<0.10) associations with BE: rs3852575, rs4149204, rs7312090, rs4149203 (**Table 5**). ORs for these SNPs were similar to those obtained in the primary analysis, though slightly reduced in magnitude (1.08 versus 1.16). In a subsequent meta-analysis, the P values for all four of these variants were highly significant ($P<5.5 \times 10^{-5}$), with an additional six SNPs satisfying P<0.05.

Assessment of gene-environment interactions

We next conducted stratified analyses and evaluated interactions between several risk factors for BE/EA (smoking, obesity, reflux, and NSAID use) and the top *MGST1* variants (q<0.01) associated with risk of BE. No statistically significant interactions were observed (data not shown).

Discussion

Chronic inflammation may occur as a result of multiple exposures established as risk factors for BE and EA (gastroesophageal reflux, obesity, smoking) and is thought to represent a common pathway underlying the emergence and progression of these conditions [3, 39]. This study represents the first systematic examination of the relationship between germline genetic variation in inflammation-related pathways—COX, cytokine signaling, oxidative stress, HLA, and NFκB—and risks of BE and EA. Drawing on genetic data from a large consortium-based GWAS [21], we found a significant association between variation in the COX pathway and risk of BE, and identified a gene-level signal for *MGST1*. 14 individual *MGST1* variants were associated with elevated disease risk, including several intronic variants subsequently confirmed ($P<5.5 \times 10^{-5}$) in a meta-analysis encompassing a large independent sample set of additional BE cases and controls.

MGST1 is one of three microsomal glutathione S-transferase (GST) enzymes in humans, and belongs to a larger GST gene family encoding a number of proteins that neutralize oxidative stress through conjugation of endogenous and xenobiotic lipophilic electrophiles with glutathione [40, 41, 42]. MGST1 shares ~40% sequence homology at the amino acid level with prostaglandin E synthase

(PTGES, formerly MGST1L1), a key enzyme that acts downstream of cyclooxygenases to catalyze the production of PGE_2 from PGH_2 [43]. MGST1-3 and PTGES belong to the "MAPEG" super-family of membrane-associated proteins in eicosanoid and glutathione metabolism. Microsomal GST1 is localized to the endoplasmic reticulum and outer mitochondrial membrane, and plays an important role in suppressing lipid peroxidation and protecting mitochondrial integrity [41]. Multiple alternatively spliced transcripts arise from the *MGST1* gene locus, and the *MGST1* promoter region has been shown to be transcriptionally responsive to oxidative stress [40]. Some evidence exists for an association between genetic variation in the *MGST1* gene and altered risk of colorectal cancer in Han Chinese [44].

The 14 *MGST1* SNPs found to be associated with risk of BE in our primary analysis were geographically clustered into two main groups, one at the 3' end of the gene, and the other at the 5' end. The most significant association was for rs4149203 C>T, a 3' intronic variant in strong LD with six other associated 3' SNPs (r^2>0.8). Four of these seven SNPs, including rs4149203, were confirmed in the meta-analysis phase of our validation studies. These variants modify predicted sequence motifs for several transcription factors (eg. POU5F1, SOX, BRCA1, FOXP1) [45] and have been described as strong expression quantitative trait loci (eQTLs) for *MGST1* in whole blood (**Table S4**) [46]. Of interest, *FOXP1* was previously identified as a susceptibility locus for BE/EA [21, 25]. At the 5' end of *MGST1*, rs2239676 C>G was the top signal identified among a cluster of six associated variants. Three of these SNPs lie in close proximity to the *MGST1* transcriptional start site, within a region characterized as active chromatin in esophageal tissue (**Figure S3**). One of these SNPs (rs2975138), as well an upstream variant (rs4149186), represents a strong eQTL for *MGST1* in esophageal mucosa [47]. Given that these 5' variants were not confirmed in the Oxford (UK) dataset, however, their association with BE risk remains questionable.

Our findings suggest that several of the identified variants may play a role in influencing *MGST1* RNA expression levels. Additional studies, however, are warranted to investigate experimentally potential associations between selected variants and altered tissue-specific *MGST1* expression, and to explore a possible causal basis for the observed findings. Data from the Genotype-Tissue Expression Project (GTEx) further indicate that a number of *MGST1* variants not included in our present analysis are also strong cis-eQTLs in esophageal mucosa [47]; imputed genotypes were available for 25 of these 29

eQTLs, but did not show significant (P<0.05) associations with disease risk (data not shown). Whether the other four *MGST1* eQTLs exhibit associations with BE remains to be evaluated. Since BE and EA often arise within an epithelium chronically exposed to refluxate and to cigarette-associated toxins (ie. associated with inflammation), it would be of interest to determine whether *MGST1* plays a protective biological role in counteracting such insults and maintaining tissue homeostasis. In this regard, we note that past studies have provided support for cytosolic glutathione S-transferases (GSTs) and glutathione peroxidases carrying out such functions in the esophagus, and have implicated epigenetic silencing of these genes as a feature of BE/EA pathogenesis [48].

Given that BE is the only known precursor of EA, one expectation is that risk factors linked to altered risk of BE would be associated with similar alterations in risk of EA. In this study, variation in the COX pathway as a whole met the threshold for significance in relation to risk of BE, but not EA. A number of the individual *MGST1* SNPs associated with risk of BE, however, did exhibit similar associations with risk of EA (**Table S3**). With respect to top SNP-level signals, the associated ORs for EA were in the same direction as, and of comparable magnitude to, those observed for BE. This strong level of concordance suggests that the identified variants, if causal, may influence disease risk primarily at the level of BE, rather than progression from BE to EA.

Previous candidate gene-based studies have reported associations between germline variation in *PTGS2* (*COX-2*) and altered risk of EA [49, 50], while independent evidence has supported an inverse association between use of NSAIDs (inhibitors of COX-1 and COX-2 activity) and risk of EA [6, 7]. Our gene-level and SNP-level analyses did not include all genes assigned to the COX pathway (e.g. *PTGS2*), as only a limited subset were advanced for further study based on pre-specified selection criteria (the top 10 genes in PC1, see **Table 3**). It remains possible that associations of disease risk with variation in other COX pathway genes may be evident in our dataset, and contribute in part to the observed pathway-level signal.

One of the strengths of our study was the use of a PCA framework to assess pathway-level and gene-level associations between germline genetic variation and risk of BE or EA. PCA is an effective strategy to reduce data dimensionality [36]. In this report, we adapted PCA to genetic pathway and gene analysis, and implemented a hierarchical strategy to identify genetic variants associated with traits.

Application of PCA to GWAS data offered key advantages over conventional marginal analyses that are based exclusively on evaluation of individual SNPs. First, by aggregating signals across multiple genes (of a given pathway) or across multiple SNPs (of a given gene), the PCA method increased our ability to detect associations characterized by multiple, independent, distributed low-magnitude signals. Second, by reducing the dimensionality of the genotype matrix, PCA appreciably reduced the number of multiple comparisons and effectively increased our statistical power.

Another important strength was the use of pooled data from the BEACON GWAS, which provided the largest sample size to date in the evaluation of inflammation-related germline variation and risks of BE and EA. As a consequence of analyzing both BE and EA, we had the opportunity to compare genetic variation associated with risk of a neoplastic precursor lesion and the cancer that arises from it. While our main findings were limited to BE, individual variants of the *MGST1* gene were also significant for risk of EA. Our pooled assessment of 7,863 SNPs in 449 genes assigned to five pathways significantly expands past candidate gene-based efforts to examine genetic variation in inflammation-related loci in relation to risk of BE and EA. Finally, our use of an external independent dataset from the U.K. helped further strengthen the identified signal at *MGST1*.

This study also had certain limitations. First, while our tiered analysis scheme enabled us to restrict the number of comparisons and boost statistical power, it also narrowed the scope of our analysis and potentially resulted in missed association signals. Variation in four of the five included pathways was not examined at the gene or SNP level, while only 25% of the genes in the COX pathway were advanced beyond pathway-level assessment. Second, given the hierarchical nature of our statistical analysis, whereby we first assessed significance at the pathway level, and then proceeded to the gene level only for 'significant' pathways, the initial P values obtained for individual genes, and subsequently for individual SNPs, should be interpreted as the P values conditional on that pathway (or gene) already being selected, i.e., P(A|B), where B represents the event that a pathway (or gene) is selected, and A represents the event that a gene (or SNP) is significant. This conditional probability framework was well suited to our use of PCA as a discovery-phase approach for identifying novel associations for subsequent confirmation in an independent sample set. Third, missing data for smoking, obesity, reflux, and NSAID use in a sizable fraction of participants reduced our statistical power to evaluate gene-environment

interactions, underscoring the need for further studies using expanded sample sets. Fourth, while our study provided broad coverage of several major biological pathways of probable relevance to BE/EA, it is almost certain that a number of important genes or genomic loci were not included. Cytokine signaling, NFκB activity, and oxidative stress, for example, represent complex processes likely influenced by many hundreds or more gene products and a large number of intergenic loci harboring both enhancer/insulator transcriptional elements and non-coding RNAs. The present analysis was restricted to examining common germline variants located within or in close (2.0-kb) proximity to defined protein-coding genes.

In conclusion, our study represents the most comprehensive evaluation to date of inflammation-related inherited genetic variation in relation to risk of BE and EA. Using a PCA framework for pathway-level and gene-level analyses, we describe evidence for novel associations between variation at the *MGST1* locus and increased risk of BE. It appears plausible that certain associated variants may act to influence expression levels of *MGST1*, a gene with known roles in the cellular response to oxidative stress. Pending further validation in additional study populations, future studies are warranted to fine-map the identified association signals, assess experimentally the functional effects of these variants, and explore the biological role of MGST1 in BE/EA pathogenesis.

Acknowledgments

We thank Terri Watson, Tricia Christopherson, and Paul Hansen for their contributions in project management and organization of biospecimens/data. We thank Dr. Liam Murray (Queen's University Belfast, UK) for access to GWAS data derived from subjects enrolled in his past studies of EA and BE (FINBAR Study). Genotyping data for MD Anderson control participants [42] was obtained from dbGaP through accession number phs000187.v1.p1. Genotyping data generated in the BEACON GWAS has been deposited into dbGaP under accession number phs000869.v1.p1.

Funding

This work was principally supported by the National Institutes of Health [R21DK099804 to M.M.M., and R01CA136725 to T.L.V. and D.C.W.]. Support for studies related to the Oxford dataset was granted by the Esophageal Adenocarcinoma GenE Consortia incorporating the ChOPIN project (grant C548/A5675) and Inherited Predisposition of neoplasia analysis of genomic DNA (IPOD) from AspECT☒and BOSS clinical trials project (grant MGAG1G7R); Cancer Research UK (AspECT, grants C548/A4584 and D9612L00090, and Histological AssessmeNt Determining EpitheliaL Response (HANDEL), grant C548/A9085); AstraZeneca UK educational grant; University Hospitals of Leicester R and D grant; and AspECT (T91 5211 University of Oxford grant HDRMJQ0). Additional funding sources for individual studies included in the BEACON GWAS, and for BEACON investigators, have been acknowledged previously [21].

Competing Interests

J. Jankowski reports being a Takeda Lecturer (Tokyo, 2015) and Chief of Eagle and Aspect Trials. D. McDonald, R. Harrison, and J.Jankowski declare being related. D. Corley, L. Bernstein, and L. Onstad report grants from the National Cancer Institute/National Institutes of Health during the conduct of the study. Dr. MacGregor reports grants from Australian Research Council, grants from Australian National Health and Medical Research Council, during the conduct of the study. RC Fitzgerald is named on patents related to the Cytosponge diagnostic device and associated assays that have been licensed by

the UK Medical Research Council to Coviden (Now Medtronic). Other authors declare no competing interests.

Author contributions

Conception and design: M.M.M., Q.H., M.F.B., L.G.J., T.L.V. Participant recruitment or data acquisition: C.P., J.L., R.C.F., W.Y., C.C., N.C.B., N.J.S., L.B., M.D.G., A.H.W., L.J.H., P.D.P., G.L., P.I., D.A.C., H.A.R., W-H.C., H.P., L.C., S.L., S.A., P.M., D.M., R.H., P.W., H.B., J.C., I.T., J.J., D.C.W., S.M., T.L.V. Analysis and interpretation of data: M.F.B, Q.H., M.M.M., L.G.J., L.O., D.M.L., A.P.T., P.G., T.L.V. Drafting of the manuscript: M.F.B., M.M.M., Q.H., T.L.V. Study supervision: M.M.M. All authors critically revised the manuscript for intellectual content.

References

[1] Thrift AP, Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: Analysis of period and birth cohort effects on recent trends. Ann Oncol. 2012 Dec;23(12):3155–3162.

[2] Vaughan TL, Fitzgerald RC. Precision prevention of oesophageal adenocarcinoma. Nat Rev Gastroenterol Hepatol. 2015 Apr;12(4):243–248. Available from: http://dx.doi.org/10.1038/-nrgastro.2015.24.

[3] Reid BJ, Li X, Galipeau PC, Vaughan TL. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nat Rev Cancer. 2010 Feb;10(2):87–101. Available from: http://www.pubmedcentral.nih.gov/-articlerender.fcgi?artid=2879265&tool=pmcentrez&rendertype=abstract.

[4] Lagergren J, Bergström R, Lindgren A, Nyrén O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med. 1999 Mar;340(11):825–831. Available from: http://-dx.doi.org/10.1056/NEJM199903183401101.

[5] Vaughan TL, Davis S, Kristal A, Thomas DB. Obesity, alcohol, and tobacco as risk factors for cancers of the esophagus and gastric cardia: Adenocarcinoma versus squamous cell carcinoma. Cancer Epidemiology Biomarkers and Prevention. 1995 Mar;4(2):85–92. Available from: http://-www.ncbi.nlm.nih.gov/pubmed/7742727.

[6] Vaughan TL, Dong LM, Blount PL, Ayub K, Odze RD, Sanchez CA, et al. Non-steroidal anti-inflammatory drugs and risk of neoplastic progression in Barrett's oesophagus: A prospective study. Lancet Oncol. 2005 Dec;6(12):945–952. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16321762.

[7] Liao LM, Vaughan TL, Corley DA, Cook MB, Casson AG, Kamangar F, et al. Nonsteroidal anti-inflammatory drug use reduces risk of adenocarcinomas of the esophagus and esophagogastric junction in a pooled analysis. Gastroenterology. 2012 Mar;142(3):442–452. Available from: http://-www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3488768&tool=pmcentrez&rendertype=abstract.

[8] Kantor ED, Onstad L, Blount PL, Reid BJ, Vaughan TL. Use of statin medications and risk of esophageal adenocarcinoma in persons with Barrett's esophagus. Cancer Epidemiology Biomarkers and Prevention. 2012 Mar;21(3):456–461. Available from: http://www.pubmedcentral.nih.gov/-articlerender.fcgi?artid=3297725&tool=pmcentrez&rendertype=abstract.

[9] Rosin MP, Anwar WA, Ward AJ. Inflammation, chromosomal instability, and cancer: the schistosomiasis model. Cancer Res. 1994 Apr;54(7 Suppl):1929s–1933s.

[10] Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002;420(6917):860–867. Available from: http://dx.doi.org/10.1038/nature01322.

[11] Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144(5):646–674. Available from: http://dx.doi.org/10.1016/j.cell.2011.02.013.

[12] Elinav E, Nowarski R, Thaiss CA, Hu B, Jin C, Flavell RA. Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. Nat Rev Cancer. 2013 Nov;13(11):759–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24154716.

[13] Atsumi T, Singh R, Sabharwal L, Bando H, Meng J, Arima Y, et al. Inflammation amplifier, a new paradigm in cancer biology. Cancer Res. 2014 Jan;74(1):8–14. Available from: http://-www.ncbi.nlm.nih.gov/pubmed/24362915.

[14] Zhang F, Altorki NK, Wu YC, Soslow RA, Subbaramaiah K, Dannenberg AJ. Duodenal reflux induces cyclooxygenase-2 in the esophageal mucosa of rats: evidence for involvement of bile acids. Gastroenterology. 2001 Dec;121(6):1391–1399.

[15] McAdam E, Haboubi HN, Griffiths AP, Baxter JN, Spencer-Harty S, Davies C, et al. Reflux composition influences the level of NF-kB activation and upstream kinase preference in oesophageal adenocarcinoma cells. Int J Cancer. 2014;8:527–535.

[16] Matsuzaki J, Suzuki H, Tsugawa H, Watanabe M, Hossain S, Arai E, et al. Bile acids increase levels of microRNAs 221 and 222, leading to degradation of CDX2 during esophageal carcinogenesis. Gastroenterology. 2013 Dec;145(6):1300–1311. Available from: http://www.ncbi.nlm.nih.gov/pubmed/-23933602 http://dx.doi.org/10.1053/j.gastro.2013.08.008.

[17] Gonçalves RB, Coletta RD, Silvério KG, Benevides L, Casati MZ, da Silva JS, et al. Impact of smoking on inflammation: overview of molecular mechanisms. Inflamm Res. 2011 May;60(5):409–424. Available from: http://dx.doi.org/10.1007/s00011-011-0308-7.

[18] Shiels MS, Katki HA, Freedman ND, Purdue MP, Wentzensen N, Trabert B, et al. Cigarette smoking and variations in systemic immune and inflammation markers. J Natl Cancer Inst. 2014 Nov;106(11). Available from: http://dx.doi.org/10.1093/jnci/dju294.

[19] Vlachopoulos C, Aznaouridis K, Bratsas A, Ioakeimidis N, Dima I, Xaplanteris P, et al. Arterial stiffening and systemic endothelial activation induced by smoking: The role of COX-1 and COX-2. Int J Cardiol. 2015 Apr;189:293–298. Available from: http://dx.doi.org/10.1016/j.ijcard.2015.04.029.

[20] Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology. Nat Rev Cancer. 2015 Jul;15(8):484–498. Available from: http://dx.doi.org/10.1038/-nrc3967.

[21] Levine DM, Ek WE, Zhang R, Liu X, Onstad L, Sather C, et al. A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. Nat Genet. 2013 Dec;45(12):1487–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24121790.

[22] Su Z, Gay LJ, Strange A, Palles C, Band G, Whiteman DC, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. Nat Genet. 2012 Oct;44(10):1131–1136. Available from: http://www.pubmedcentral.nih.gov/-articlerender.fcgi?artid=3459818&tool=pmcentrez&rendertype=abstract.

[23] Ek WE, Levine DM, D'Amato M, Pedersen NL, Magnusson PKE, Bresso F, et al. Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's esophagus, and gastroesophageal reflux. J Natl Cancer Inst. 2013 Nov;105(22):1711–1718.

[24] Palles C, Chegwidden L, Li X, Findlay JM, Farnham G, Giner FC, et al. Polymorphisms Near TBX5 and GDF7 Are Associated With Increased Risk for Barrett's Esophagus. Gastroenterology. 2015 Nov;148(2):367–378. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0016508514013341 http://www.ncbi.nlm.nih.gov/pubmed/25447851.

[25] Becker J, May A, Gerges C, Anders M, Veits L, Weise K, et al. Supportive evidence for FOXP1, BARX1, and FOXF1 as genetic risk loci for the development of esophageal adenocarcinoma. Cancer Medicine. 2015;.

[26] Jin L, Zuo XY, Su WY, Zhao XL, Yuan MQ, Han LZ, et al. Pathway-based analysis tools for complex diseases: a review. Genomics Proteomics Bioinformatics. 2014 Oct;12(5):210–220. Available from: http://dx.doi.org/10.1016/j.gpb.2014.10.002.

[27] Buas MF, Levine DM, Makar KW, Utsugi H, Onstad L, Li X, et al. Integrative post-genome-wide association analysis of CDKN2A and TP53 SNPs and risk of esophageal adenocarcinoma.

Carcinogenesis. 2014 Dec;35(12):2740–2747. Available from: http://www.carcin.oxfordjournals.org/cgi/-doi/10.1093/carcin/bgu207.

[28] Amos CI, Wang LEE, Lee JE, Gershenwald JE, Chen WV, Fang S, et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. Hum Mol Genet. 2011 Dec;20(24):5012–5023. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21926416 http://-www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3298855&tool=pmcentrez&rendertype=abstract.

[29] Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet Epidemiol. 2010;34(6):591–602. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20718045.

[30] Schottenfeld D, Beebe-Dimmer J. Chronic inflammation: a common and important factor in the pathogenesis of neoplasia. CA Cancer J Clin. 2006;56(2):69–83.

[31] Konturek PC, Nikiforuk A, Kania J, Raithel M, Hahn EG, Mühldorfer S. Activation of NFkappaB represents the central event in the neoplastic progression associated with Barrett's esophagus: a possible link to the inflammation and overexpression of COX-2, PPARgamma and growth factors. Dig Dis Sci. 2004 Aug;49(7-8):1075–1083.

[32] Poehlmann A, Kuester D, Malfertheiner P, Guenther T, Roessner A. Inflammation and Barrett's carcinogenesis. Pathol Res Pract. 2012 May;208(5):269–280. Available from: http://dx.doi.org/10.1016/-j.prp.2012.03.007.

[33] Abdalla SI, Sanderson IR, Fitzgerald RC. Effect of inflammation on cyclooxygenase (COX)-2 expression in benign and malignant oesophageal cells. Carcinogenesis. 2005 Sep;26(9):1627–1633. Available from: http://dx.doi.org/10.1093/carcin/bgi114.

[34] O'Riordan JM, Abdel-latif MM, Ravi N, McNamara D, Byrne PJ, McDonald GSA, et al. Proinflammatory cytokine and nuclear factor kappa-B expression along the inflammation-metaplasia-dysplasia-adenocarcinoma sequence in the esophagus. Am J Gastroenterol. 2005 Jun;100(6):1257–1264. Available from: http://dx.doi.org/10.1111/j.1572-0241.2005.41338.x.

[35] Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012 Feb;9(2):179–181. Available from: http://dx.doi.org/10.1038/nmeth.1785.

[36] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York: Springer; 2009.

[37] Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010 Sep;26(18):2336–2337. Available from: http://dx.doi.org/10.1093/bioinformatics/btq419.

[38] Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013 Jun;14(6):379–389. Available from: http://dx.doi.org/10.1038/nrg3472.

[39] Tselepis C, Perry I, Dawson C, Hardy R, Darnton SJ, McConkey C, et al. Tumour necrosis factor-alpha in Barrett's oesophagus: a potential novel mechanism of action. Oncogene. 2002 Sep;21(39):6071–6081. Available from: http://dx.doi.org/10.1038/sj.onc.1205731.

[40] Kelner MJ, Bagnell RD, Montoya MA, Estes LA, Forsberg L, Morgenstern R. Structural organization of the microsomal glutathione S-transferase gene (MGST1) on chromosome 12p13.1-13.2. Identification of the correct promoter region and demonstration of transcriptional regulation in response to oxidative stress. J Biol Chem. 2000 Apr;275(17):13000–13006.

[41] Johansson K, Järvliden J, Gogvadze V, Morgenstern R. Multiple roles of microsomal glutathione transferase 1 in cellular protection: a mechanistic study. Free Radic Biol Med. 2010 Dec;49(11):1638–1645. Available from: http://dx.doi.org/10.1016/j.freeradbiomed.2010.08.013.

[42] Nebert DW, Vasiliou V. Analysis of the glutathione S-transferase (GST) gene family. Hum Genomics. 2004 Nov;1(6):460–464.

[43] Jakobsson PJ, Thorén S, Morgenstern R, Samuelsson B. Identification of human prostaglandin E synthase: a microsomal, glutathione-dependent, inducible enzyme, constituting a potential novel drug target. Proc Natl Acad Sci U S A. 1999 Jun;96(13):7220–7225.

[44] Zhang H, Liao LH, Liu SM, Lau KW, Lai AKC, Zhang JH, et al. Microsomal glutathione S-transferase gene polymorphisms and colorectal cancer risk in a Han Chinese population. Int J Colorectal Dis. 2007 Oct;22(10):1185–1194. Available from: http://dx.doi.org/10.1007/s00384-007-0308-9.

[45] Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012 Jan;40(Database issue):D930–D934. Available from: http://dx.doi.org/10.1093/nar/gkr917.

[46] Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013 Oct;45(10):1238–1243. Available from: http://dx.doi.org/10.1038/ng.2756.

[47] "The_GTEx_Consortium". The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015 May;348(6235):648–660. Available from: http://dx.doi.org/-10.1126/science.1262110.

[48] Peng DF, Razvi M, Chen H, Washington K, Roessner A, Schneider-Stock R, et al. DNA hypermethylation regulates the expression of members of the Mu-class glutathione S-transferases and glutathione peroxidases in Barrett's adenocarcinoma. Gut. 2009 Jan;58(1):5–15. Available from: http://-dx.doi.org/10.1136/gut.2007.146290.

[49] Moons LMG, Kuipers EJ, Rygiel AM, Groothuismink AZM, Geldof H, Bode WA, et al. COX-2 CA-haplotype is a risk factor for the development of esophageal adenocarcinoma. Am J Gastroenterol. 2007;102(11):2373–2379. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17581270?dopt=Citation.

[50] Kristinsson JO, van Westerveld P, te Morsche RHM, Roelofs HMJ, Wobbes T, Witteman BJM, et al. Cyclooxygenase-2 polymorphisms and the risk of esophageal adeno- or squamous cell carcinoma. World J Gastroenterol. 2009;15(28):3493–3497. Available from: http://www.ncbi.nlm.nih.gov/pubmed/-19630103.

Table 1. Study participant characteristics[†].

	Controls[#] (n=3207)		BE (n=3295)		EA (n=2515)	
	N	%	N	%	N	%
Age (years)						
<50	726	22.6	449	13.7	189	7.6
50-59	885	27.6	780	23.7	547	21.9
60-69	963	30.0	1011	30.7	884	35.4
70+	633	19.7	1048	31.9	875	35.1
Sex						
Female	880	27.4	806	24.5	320	12.7
Male	2327	72.6	2489	75.5	2195	87.3
BMI						
<25	786	36.3	425	20.7	245	24.6
25-29.99	944	43.6	882	42.9	455	45.7
30-34.99	307	14.2	521	25.3	201	20.2
35+	130	6.0	230	11.2	95	9.5
Smoking status						
No	889	40.9	798	33.7	348	24.7
Yes	1284	59.1	1570	66.3	1062	75.3
Smoking (p-y)[a]						
None	889	41.3	798	44.5	348	32.8
<15	358	16.6	320	17.9	156	14.7
15-29	326	15.1	232	12.9	160	15.1
30-44	273	12.7	198	11.0	173	16.3
45+	309	14.3	244	13.6	225	21.2
NSAID use						
Never	814	44.0	503	42.8	381	46.2
Ever	1038	56.0	672	57.2	444	53.8
Reflux/heartburn[b]						
No	1448	80.6	957	49.0	563	56.2
Yes	349	19.4	996	51.0	438	43.8

[†]Numbers do not add to total subjects due to missing data; [#]3 participants were excluded from the control group for comparison to BE case patients due to relatedness; [a]Pack-years, [b]Weekly symptoms

Table 2. Assessment of pathway-level associations with risk of Barrett's esophagus (BE) or esophageal adenocarcinoma (EA).

			BE			EA		
Pathway	**Genes**	**Variants[a]**	**PCs[b]**	**P[c]**	**q[d]**	**PCs[b]**	**P[c]**	**q[d]**
COX	40	1241	40	0.006	0.03	40	0.20	0.60
Cytokines	184	2622	110	0.10	0.21	109	0.28	0.60
Oxidative stress	115	1958	73	0.13	0.21	73	0.58	0.60
Immune/HLA	32	1036	10	0.59	0.74	10	0.60	0.60
NFκB	125	1681	110	0.84	0.84	109	0.42	0.60

[a]Total number of single nucleotide polymorphisms (SNPs) selected for analysis; [b]Pathway-level principal components (PCs) included in the logistic regression model; [c]Likelihood ratio P value; [d]False discovery rate (FDR) q value.

Table 3. Assessment of first 10 gene-level associations with risk of BE.

	Gene		Variants[a]	PCs[b]	P[c]	q[d]
1	MGST1	Microsomal glutathione S-transferase 1	36	3	0.0005	0.005
2	PTGER3	Prostaglandin E receptor 3 (subtype EP3)	185	4	0.11	0.51
3	PPARG	Peroxisome proliferator-activated receptor gamma	121	3	0.15	0.51
4	TBXAS1	Thromboxane A synthase 1 (platelet)	176	5	0.29	0.58
5	IL12RB2	Interleukin 12 receptor, beta 2	29	3	0.29	0.58
6	CYP19A1	Cytochrome P450, family 19, subfamily A, polypeptide 1	50	3	0.40	0.66
7	MMP2	Matrix metallopeptidase 2	25	3	0.48	0.69
8	PPARA	Peroxisome proliferator-activated receptor alpha	54	3	0.72	0.80
9	MGST2	Microsomal glutathione S-transferase 2	57	4	0.72	0.80
10	PTGES	Prostaglandin E synthase	11	3	1.00	1.00

[a]Total number of SNPs selected for analysis of the indicated gene; [b]Gene-level principal components (PCs) included in the logistic regression model; [c]Likelihood ratio P value; [d]False discovery rate (FDR) q value.

Table 4. Assessment of *MGST1* SNPs (n=36) and risk of BE[#]

					Controls		BE cases					
	SNP	Chr	Position	Alleles[a]	N	MAF[b]	N	MAF[b]	OR[c]	95% CI	P	q[d]
1	rs4149203	12	16514921	T/C	3203	0.308	3288	0.346	1.16	(1.08-1.26)	0.0001	0.001
2	rs3852575	12	16516260	T/C	3203	0.304	3288	0.34	1.16	(1.08-1.25)	0.0001	0.001
3	rs7312090	12	16515945	T/C	3203	0.304	3288	0.34	1.16	(1.07-1.25)	0.0002	0.001
4	rs4149204	12	16515062	C/T	3203	0.307	3288	0.342	1.16	(1.07-1.25)	0.0002	0.001
5	rs4149207	12	16517491	T/C	3203	0.306	3288	0.338	1.14	(1.06-1.23)	0.0008	0.005
6	rs4149208	12	16517581	T/C	3203	0.306	3288	0.338	1.14	(1.06-1.23)	0.0008	0.005
7	rs3759207	12	16516710	C/T	3203	0.31	3288	0.34	1.14	(1.05-1.23)	0.0012	0.006
8	rs4149195	12	16512128	G/A	3203	0.109	3288	0.125	1.20	(1.07-1.35)	0.0013	0.006
9	rs2239676	12	16500448	G/C	3203	0.096	3288	0.113	1.19	(1.06-1.34)	0.0033	0.013
10	rs4149187	12	16500071	G/C	3203	0.098	3288	0.114	1.18	(1.05-1.32)	0.0061	0.022
11	rs2239677	12	16500680	A/G	3203	0.021	3288	0.027	1.38	(1.09-1.75)	0.0077	0.025
12	rs2239675	12	16500265	G/A	3203	0.172	3288	0.187	1.12	(1.02-1.23)	0.0172	0.049
13	rs4149186	12	16498700	C/A	3203	0.215	3288	0.235	1.11	(1.02-1.21)	0.0179	0.049
14	rs2975138	12	16501551	A/G	3203	0.237	3288	0.256	1.10	(1.02-1.20)	0.0192	0.049

[#]Results for n=14 SNPs satisfying FDR q<0.05, [a]Minor/major alleles, [b]Minor allele frequency, [c]Odds ratio, adjusted for age, sex, $PC_{1,AIM}$-$PC_{4,AIM}$ using additive model (per-allele), [d]False discovery rate (FDR)

Table 5. Assessment of *MGST1* SNPs and risk of BE in an independent study sample of 1,851 BE cases and 3,496 control participants (Oxford)[#].

			BEACON		Oxford		Meta-analysis	
	SNP	**Allele[a]**	**OR**	**P**	**OR**	**P**	**OR**	**P**
1	rs4149203	T	1.16	0.0001	1.08	0.0718	1.13	3.46E-05
2	rs3852575	T	1.16	0.0001	1.08	0.0661	1.13	4.04E-05
3	rs7312090	T	1.16	0.0002	1.08	0.0678	1.13	5.12E-05
4	rs4149204	C	1.16	0.0002	1.08	0.0668	1.13	5.25E-05
5	rs4149207	T	1.14	0.0008	1.05	0.2676	1.10	0.0011
6	rs4149208	T	1.14	0.0008	1.05	0.2837	1.10	0.0013
7	rs3759207	C	1.14	0.0012	1.05	0.2649	1.10	0.0015
8	rs4149195	G	1.20	0.0013	1.09	0.2160	1.15	0.0012
9	rs2239676	G	1.19	0.0033	0.99	0.9402	1.10	0.0293
10	rs4149187	G	1.18	0.0061	0.99	0.8894	1.09	0.0461
11	rs2239675	G	1.12	0.0172	1.00	0.9882	1.07	0.0704
12	rs4149186	C	1.11	0.0179	0.99	0.7973	1.05	0.1081
13	rs2975138	A	1.10	0.0192	1.01	0.8523	1.06	0.0605

[#]Results for n=13 SNPs available for analysis among the 14 variants listed in Table 4; [a]Effect allele (all ORs represent per-allele risk estimates under an additive model)

Figure Legends

Figure 1. Regional association plot for n=36 genotyped SNPs at the *MGST1* gene locus. The top-ranked SNP associated with risk of BE is shown in solid purple. SNPs are ordered by genomic location. The color scheme indicates LD between the top-ranked SNP and other SNPs in the region using r^2 values calculated from the 1000 Genomes Project. The y axis shows $-\log_{10}$ (P) values computed from 3295 BE cases and 3204 controls. The recombination rate from CEU (Utah residents of Northern and Western European ancestry) HapMap data (right y axis) is shown in light blue.