



This is a repository copy of *A Newly Identified Susceptibility Locus near FOXP1 Modifies the Association of Gastroesophageal Reflux with Barrett's Esophagus.*

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/110794/>

Version: Accepted Version

---

**Article:**

Dai, J.Y., Tapsoba, J.D.D., Buas, M.F. et al. (15 more authors) (2015) A Newly Identified Susceptibility Locus near FOXP1 Modifies the Association of Gastroesophageal Reflux with Barrett's Esophagus. *Cancer Epidemiology, Biomarkers and Prevention*, 24 (11). pp. 1739-1747. ISSN 1055-9965

<https://doi.org/10.1158/1055-9965.EPI-15-0507>

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



Published in final edited form as:

*Cancer Epidemiol Biomarkers Prev.* 2015 November ; 24(11): 1739–1747. doi:  
10.1158/1055-9965.EPI-15-0507.

## A newly identified susceptibility locus near *FOXP1* modifies the association of gastroesophageal reflux with Barrett' s esophagus

James Y. Dai<sup>1,2</sup>, Jean de Dieu Tapsoba<sup>1</sup>, Matthew F. Buas<sup>1</sup>, Lynn E. Onstad<sup>1</sup>, David M. Levine<sup>2</sup>, Harvey A. Risch<sup>3</sup>, Wong-Ho Chow<sup>4</sup>, Leslie Bernstein<sup>5</sup>, Weimin Ye<sup>6</sup>, Jesper Lagergren<sup>7,8</sup>, Nigel C. Bird<sup>9</sup>, Douglas A. Corley<sup>10,11</sup>, Nicholas J. Shaheen<sup>12</sup>, Anna H. Wu<sup>13</sup>, Brian J. Reid<sup>1,14</sup>, Laura J. Hardie<sup>15</sup>, David C. Whiteman<sup>16</sup>, and Thomas L. Vaughan<sup>1,17</sup>

<sup>1</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

<sup>2</sup>Department of Biostatistics, University of Washington, School of Public Health, Seattle, Washington, 98109 USA

<sup>3</sup>Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut 06510

<sup>4</sup>Department of Epidemiology, MD Anderson Cancer Center, Houston, TX, USA

<sup>5</sup>Department of Population Sciences, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, California, 91010 USA

<sup>6</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden

<sup>7</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

<sup>8</sup>Division of Cancer Studies, King' s College London, London, United Kingdom

<sup>9</sup>Department of Oncology, Medical School, University of Sheffield, Sheffield S10 2RX, UK

<sup>10</sup>Division of Research, Kaiser Permanente Northern California, Oakland, California, 94612 USA

<sup>11</sup>San Francisco Medical Center, Kaiser Permanente Northern California, San Francisco, California, 94115 USA

<sup>12</sup>Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, University of North Carolina, Chapel Hill, North Carolina, 27599 USA

<sup>13</sup>Department of Preventive Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, 90033 USA

<sup>14</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

<sup>15</sup>Division of Epidemiology, University of Leeds, Leeds LS2 9JT, UK

<sup>16</sup>Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane QLD 4006, Australia

---

Corresponding Authors: James Y. Dai, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, P.O. Box 19024, Seattle, WA 98109-1024, phone: 206-667-6364, jdai@fredhutch.org; Thomas L. Vaughan, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, P.O. Box 19024, Seattle, WA 98109-1024, phone: 206-667-5134, tvaughan@u.washington.edu.

The authors disclose no potential conflicts of interest.

<sup>17</sup>Department of Epidemiology, University of Washington, School of Public Health, Seattle, Washington, 98109 USA

## Abstract

**Background**—Important risk factors for esophageal adenocarcinoma (EA) and its precursor, Barrett' s esophagus (BE) include gastroesophageal reflux disease, obesity, and cigarette-smoking. Recently, genome-wide association studies have identified seven germline single nucleotide polymorphisms (SNPs) that are associated with risk of BE and EA. Whether these genetic susceptibility loci modify previously identified exposure-disease associations is unclear.

**Methods**—We analyzed exposure and genotype data from the BEACON Consortium discovery phase GWAS, which included 1516 EA case patients, 2416 BE case patients, and 2187 control participants. We examined the seven newly identified susceptibility SNPs for interactions with body mass index, smoking status, and report of weekly heartburn or reflux. Logistic regression models were used to estimate odds ratios for these risk factors stratified by SNP genotype, separately for BE and EA.

**Results**—The odds ratio for BE associated with at least weekly heartburn or reflux varied significantly with the presence of at least one minor allele of rs2687201 (nominal p-value=0.0005, false discovery rate=0.042). Odds ratios (95% confidence intervals) for weekly heartburn or reflux among participants with 0, 1, or 2 minor alleles of rs2687201 were 6.17 (4.91,7.56), 3.56 (2.85,4.44), and 3.97 (2.47,6.37), respectively. No statistically significant interactions were observed for smoking status and body mass index.

**Conclusion**—Reflux symptoms are more strongly associated with BE risk among persons homozygous for the major allele of rs2687201, which lies ~75 kb downstream of the transcription factor gene *FOXP1*.

**Impact**—The novel gene-exposure interaction discovered in this study provides new insights to the etiology of esophageal adenocarcinoma.

## Keywords

esophageal cancer; gastroesophageal reflux disease; gene-environment interaction; genome-wide association study; smoking

## INTRODUCTION

The incidence of esophageal adenocarcinoma (EA), an infrequent but often lethal disease, has been rising sharply in the past four decades, especially among white populations in developed countries (1–6). The majority of EA cases arise from Barrett' s esophagus (BE), a precursor lesion defined by a specialized columnar metaplasia of the distal esophagus (7). BE and EA share many risk factors, including gastroesophageal reflux disease (GERD), European ancestry, male gender, obesity and tobacco smoking, although the magnitudes of these associations may differ for BE and EA (8–14). Population-based studies in the United States and Australia indicate that GERD symptoms, high body-mass, and tobacco smoking account for approximately 75% of the population risk for EA (10,12).

Less is known about inherited genetic susceptibility to BE and EA. Using the candidate gene approach, a number of genetic association studies have investigated the role of genetic predisposition in several biological pathways (15–21). Recently, large genome-wide association studies (GWAS) have identified several single nucleotide polymorphisms (SNPs) for BE (22–23), including variants in or near *FOXF1*, *TBX5*, *GDF7*, and the major histocompatibility complex (*MHC*) genes. Additional susceptibility loci for BE/EA were also identified in or near *CRTC1*, *BARX1*, and *FOXP1* (24). These loci were close to or passed the stringent genome-wide significance threshold. Notably, the GWAS based on the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON) reported that risk of BE and EA is influenced by many germline variants of small effect and, perhaps not surprisingly, genetic heritability is largely shared between the two diseases (25).

Understanding the interplay between genetic susceptibility and epidemiologic risk factors will improve risk prediction and provide insights into disease pathogenesis, thereby increasing understanding of BE and EA etiology. Some gene-exposure interactions have also been identified using the candidate gene approach, e.g. smoking and variants of *GSTM1*, *GSTT1* and *VEGF* (26–31). Few of the previously reported variants have been replicated and more systematic studies are warranted. The statistical power to detect a genome-wide gene-environment interaction is, however, typically low because of high variability in estimates of interactions and the stringent p-value correction required for genome-wide testing.

An alternative approach is to focus the search for gene-exposure interactions on genetic susceptibility loci with significant genome-wide associations. This approach is mostly driven by statistical considerations to increase the chances of discovery, and has been used successfully in past studies of breast cancer (32). If a genetic variant interacts with an exposure, it is quite plausible that the marginal genetic association, i.e. the genetic association observed across all exposure levels (ignoring gene-exposure interaction) will also be evident. Limiting the search for gene-exposure interactions to SNPs that already demonstrate significant marginal genetic associations decreases the risk of chance associations and reduces the multiple-testing burden (33–35). Furthermore, marginal genetic associations are independent of gene-environment interactions as long as they are assessed in nested models, so that a much reduced multiple-test penalty is required to test for gene-environment interactions among SNPs that rank high in marginal association tests (33). This approach therefore presents a benefit in statistical power when compared to an agnostic genome-wide interaction search.

In this report, using BEACON consortium data, we examine gene-exposure interactions, focusing on the seven SNPs previously shown to be associated with BE or EA at or near genome-wide significance and confirmed in replication studies. The risk factors under investigation include body-mass index, cigarette smoking, and acid reflux or heartburn symptoms.

## Materials and Methods

**Study population and SNP genotyping**—Information from participants with BE and EA, and from associated controls, was collected by BEACON investigators in 14 studies

conducted in Australia, Western Europe and North America. All EA and BE diagnoses were histologically confirmed. Detailed population characteristics and genotyping protocols have been previously described (24). Demographic and exposure data were harmonized into standard variables. Genotyping of DNA from buffy coat or whole blood was performed using the Illumina HumanOmni1-Quad platform. Quality assurance and quality control of genotyping calls followed standard procedures (36). All participants gave written informed consent, and the project was approved by the ethics review boards at each participating center as well as for the study overall at the Fred Hutchinson Cancer Research Center.

All 1516 EA case patients and 2416 BE case patients in the discovery phase of the BEACON GWAS were included in this investigation, together with 2187 control participants. Three control participants were excluded from BE analyses due to familial relation to cases. Although all study sites collected data on age, sex, BMI and smoking status, some did not ascertain history of heartburn or reflux. The missing status of major risk factors under investigation in this report has been tabulated in Supplementary Table 1. The vast majority of missing data occurred because not all variables were included in all study questionnaires.

**Statistical methods**—Seven SNPs confirmed as associated with risk of EA or BE (22–24) were included in our analysis: rs3072 in 2p24.1 (*GDF7*), rs2687201 in 3p14 (*FOXP1*), rs9257809 in 6p21 (*MHC*), rs11789015 in 9q22 (*BARX1*), rs2701108 in 12q24.21 (*TBX5*), rs9936833 in 16q24 (*FOXF1*), and rs10419226 in 19p13 (*CRTC1*). The reported odds ratios for these SNPs can be found in Supplementary Table 2. Three established risk factors were investigated for potential interaction with the SNPs: cigarette-smoking, BMI (weight (kg)/height<sup>2</sup> (m<sup>2</sup>)), and gastroesophageal reflux symptoms. These variables were coded as ever smoking (yes or no), BMI (<25, 25-<30, 30 kg/m<sup>2</sup>) and at least weekly heartburn or weekly reflux (yes or no). The definition of gastroesophageal reflux symptoms has been previously described (50). For each of these variables, logistic regression models were fitted to assess separately the risks of BE and EA. Each model included age, sex, the first four principal components (PC1-PC4) derived using genome-wide SNP data to account for population stratification by ancestry (24), the environmental exposure under investigation, a binary genotype indicator (0/1) for the presence of at least one minor allele of the given SNP, or a continuous genotype variable with discrete values 0, 1, or 2, and the product term of the environment exposure and genotype variable. The statistical significance of the interaction was assessed by a 1 degree-of-freedom (df) statistic for equal odds ratios between participants without the minor allele and participants with one or two copies of the minor allele, or the odds ratio associated with 1 additional minor allele. Although correlated, these two types of interaction tests explore parsimoniously the unknown true functional form of the interactions under investigation. The conservative Bonferroni correction for multiple testing was conducted for the 84 tests of gene-exposure interactions between 7 SNPs and 3 risk variables, two outcomes (BE and EA), as well as interaction by the presence of the minor allele and by the trend interaction test. The false discovery rate (FDR) was computed for each of the nominal p-values for interaction (37). A sensitivity analysis was conducted to assess the impact of missing data in risk factors. Inverse probability weighted logistic

regression models were fitted, in which the weight is the inverse of the probability of being observed given age, gender, site, genotype and the top four principal components.

After examining interactions for each SNP-exposure pair, one at a time, additional analyses were conducted for SNP-exposure pairs that satisfied the criterion of family-wise error rate  $<0.05$  after Bonferroni correction. To control for potential confounding from correlated risk factors, a logistic regression model was fitted to assess the three interaction terms between a SNP and the three risk variables, namely BMI, cigarette-smoking, and at least weekly heartburn or acid reflux, controlling for age, sex, the main effects of the SNP and risk factors, and the top four principal components.

Association analyses were conducted for imputed SNPs based on the 1000 Genomes project around any genotyped SNPs that showed significant interactions with one of the three risk factors. The details of the imputation procedure were presented previously (24).

## RESULTS

Characteristics of participants in this study have been described elsewhere (24). Table 1 lists the associations of the three studied risk factors for EA and BE, adjusted for age and sex. The odds ratios for BE were generally larger than those for EA, except for cigarette smoking.

Figure 1 shows the q-q plot of p-values for each SNP and exposure pair, and for EA and BE separately. The p-values for the two types of interaction tests were split due to potentially high correlation between the two tests. As a comparison, a randomly selected set of seven SNPs from the BEAGESS study was assessed for interactions with the three exposures. The observed p-values were generally smaller than the expected p-values from the random set, possibly because of random noise and correlation of these statistics. Two SNPs had interaction p-values  $<0.05$  (Table 2, equality test and trend test) and deviated significantly from the diagonal line in the q-q plot. The first SNP, rs2687201 (G/T, major and minor plus strand allele), lies near the *FOXP1* gene, and the second, rs10419226 (C/A, major and minor plus strand allele), is located near the *CRTC1* gene. Both SNPs were identified in the BEAGESS study (24).

The upper half of Table 2 shows the odds ratios of the three risk factors for groups with 0, 1, or 2 minor alleles of rs2687201 separately for BE and EA. The interaction between rs2687201 (G/T, major and minor plus strand allele) and GERD in relation to the risk of BE had the smallest p-value (0.0005), which is statistically significant under Bonferroni correction for 84 comparisons ( $P < 5.95 \times 10^{-4}$ ). The presence of at least one minor allele of rs2687201 appeared to decrease the magnitude of disease risk associated with GERD. The interaction of this SNP with ever smoking reached borderline statistical significance, whereas the association of BMI with risk of BE did not appear to vary substantially with rs2687201 genotype. The interactions of this SNP with exposures on the risk of EA were not statistically significant, although the patterns of the odds ratios among the three genotype groups were similar to those for BE; having one or two minor alleles generally decreased the magnitude of risk associated with reflux/heartburn and smoking. The interaction between

BMI and this variant in relation to EA risk was of borderline nominal significance ( $p=0.08$ ), and this association was also in the same direction as that for BE risk. In sensitivity analyses, where the missing data in risk factors were accounted for by the inverse probability weighting method, the interaction between rs2687201 and GERD remained statistically significant (Supplementary Table 3). In a separate analysis, where BE and EA case participants were combined into a single case group, the significance level of the interaction between rs2687201 and GERD decreased (Supplementary Table A4). In results not shown, the interaction odds ratios in the three continents did not differ significantly ( $p=0.54$ ).

The lower half of Table 2 shows the results for rs10419226 (C/A, major and minor plus strand allele). The smallest p-values occurred for the interaction of this SNP with GERD in association with EA risk (nominal  $p=0.01$ ). The FDR for this interaction was 0.28.

In Table 3, the three risk factors were investigated simultaneously in one logistic regression model for their interactions with the rs2687201 variant, the only SNP with an interaction P value satisfying the Bonferroni significance threshold. This analysis helped to differentiate the independent contributions of the risk factors and their interactions since the three risk factors were correlated across the study populations. The p-value for the interaction with GERD remained nominally statistically significant ( $p=0.001$ ), whereas the statistical significance of the interaction between the SNP and ever-smoking decreased when compared to the results in Table 2. These findings suggest that the newly discovered interaction between the SNP and GERD appears unlikely to be significantly confounded by the other two risk factors.

Table 4 shows odds ratios stratified by rs2687201 genotype and GERD status, adjusted for age, sex, and the top four principal components. Participants without minor alleles and without GERD comprised the reference group. The genetic relative risk (GRR) in the absence of GERD was 1.50, a statistically significant elevation of BE risk (nominal p-value = 0.00004). GERD increased the risk of BE substantially, regardless of the genotype group. The odds ratio for GERD-positive participants with at least one copy of the minor allele was less than that observed for GERD-positive participants homozygous for the major allele.

Figure 2 shows the regional interaction associations around rs2687201, including both genotyped and imputed SNPs. Of the 2017 adjacent SNPs assessed, a cluster of adjacent imputed SNPs with correlations ( $r^2$ )  $>0.6$  showed higher levels of significance in their interactions with GERD. The smallest nominal p-value obtained was for rs7638679 ( $P=0.00006$ , Supplementary Table 5), which was much more statistically significant than that observed for rs2687201. The rs7638679 variant is situated ~50 kb away from the *FOXP1* 3'UTR.

## DISCUSSION

We systematically investigated interactions between known risk factors for BE and EA and previously identified genome-wide significant associations in germline susceptibility. We found that rs2687201 (G/T, major and minor plus strand allele), a SNP on chromosome 3 near the *FOXP1* gene, modifies the association between gastroesophageal reflux disease and

risk of BE. None of the SNPs showed evidence of statistically significant interactions with BMI and smoking. As reported in our GWAS analysis of genetic effects alone (24), the rs2687201 polymorphism was associated with increased risk of BE and EA when the two diseases were combined into a single case group (OR=1.14).

The rs2687201 variant (chromosome 3p13) is located 75 kb distal to the *FOXP1* 3'UTR, in a ~1-Mb intergenic region containing several pseudogenes (*RNPC3P1*, *UQCRHP4*, *COX6CP6*, *HMGB1P36*) and a predicted ~130-bp “novel miRNA” locus (*AC096971.1*) (Figure S1). Based on data from the 1000 Genomes Project (38), rs2687201 is in strong linkage disequilibrium (LD) ( $r^2 > 0.8$ ) with ~60 other SNPs, located within 30–60 kb (20 SNPs with  $r^2 > 0.94$ ). rs2687201 is situated within a 1600-bp region characterized as heterochromatin in esophageal tissue, according to chromatin state segmentation data derived from the NIH Roadmap Epigenome Project (39). DNA regulatory motifs for multiple transcriptional regulators (e.g., CPHX, IK-2, ZNF143, HOXA9/10, MEF2, CTCF, RAD21, YY1, and NF- $\kappa$ B) are predicted to be altered by rs2687201 or other nearby variants in high LD ( $r^2 > 0.90$ ) (40). Data derived from cell lines analyzed in the ENCODE project indicate recruitment of several transcription factors (CEBPB, CJUN, P300, RAD21, STAT3) to recognition sequences within 4–8 kb (41).

Further assessment of the 13 imputed SNPs with more significant interaction P values than rs2687201 revealed that several of these variants (eg, rs2597312, rs7611254, rs1522554) are situated in putative enhancer sequences according to data from the Roadmap Epigenome Project (Supplementary Table 6.1); most of the 13 SNPs also appear to alter predicted DNA regulatory motifs. Expression quantitative trait locus (eQTL) analyses using the Genotype-Tissue Expression (GTEx) Project resource (42) provided some evidence that eight of these 13 SNPs may be associated ( $P < 0.05$ ) with altered *FOXP1* expression levels in esophageal mucosa (Supplementary Table 6.2). Cautious interpretation is required, however, given the sizable number of comparisons and relatively weak P values. The precise functional effects of rs2687201 and/or linked variants on expression levels of *FOXP1* or other neighboring loci remain to be determined. While biological characterization will require experimental follow-up studies, a potential regulatory role for several of these SNPs appears plausible.

The *FOXP1* gene encodes a member of the Forkhead box (*FOX*) family of transcription factors, which share an evolutionarily conserved “winged-helix” DNA binding domain and function as versatile regulators of a wide range of biological processes, including development and cancer (43–45). Knockout studies in mice demonstrated that *FOXP1* and *FOXP2* cooperatively regulate lung and esophageal development (46), while human *FOXP1* was first identified as a candidate tumor suppressor gene on chromosome 3p, a region known to exhibit loss of heterozygosity in many tumors, and in pre-malignant epithelial lesions of the oral cavity, breast, and cervix (47). In subsequent studies, expression of *FOXP1* was associated with improved survival in breast cancer (48), but reduced survival in diffuse large B-cell lymphoma (49). *FOXP1* appears to play opposing roles in different tissues and functions as either a tumor suppressor or oncogene depending upon context (50). It is interesting to note that genetic variation (rs9936833) in proximity to another *FOX* family member, *FOXF1*, has also been associated with altered risk of BE (22). While significant GxE interactions were not observed between this variant and reflux, BMI, or

smoking, our present findings for *FOXP1* provide further evidence implicating *FOX* transcription factors in the modulation of disease risk for BE/EA, two conditions arising within an organ and tissue regulated by these same proteins during embryogenesis.

Several previous studies have pursued candidate-based GxE analyses of EA, and described potential interactions between GERD, smoking, or BMI and variants in a number of different genes related to detoxification, angiogenesis, DNA repair, apoptosis, and extracellular matrix degradation (27–31). These studies, however, were all limited by small sample sizes and lack of independent confirmation, and interactions in relation to risk of BE were not examined.

The present study has several strengths. First, our use of genetic and epidemiologic data from a large consortium-based GWAS of EA/BE (24) provided us with greater statistical power to detect GxE interactions than has been available in any previous study. Our decision to focus on the top marginal genetic signals identified in recent genome-wide analyses also eliminated the need to correct for a massive number of comparisons (33). Second, all genotyping from this GWAS was conducted on a single platform, and subjected to stringent quality-control procedures. Third, inclusion of both EA and BE cases in the current analysis enabled parallel assessment of GxE interactions for both an epithelial cancer and its metaplastic precursor lesion.

Our study also has certain limitations. We did not have a replication study for the newly discovered interaction due to lack of relevant exposure data in other studies. A full exploration of variants around the signal SNP and their functional consequences will further elucidate the mechanism of the interaction. The extent of missing data for environmental variables examined (34% missing BMI data among EA cases; 34% and 18% missing reflux data among EA or BE, respectively) considerably reduced statistical power for the indicated analyses, and may have resulted in some falsely negative GxE interactions. While systematic differences cannot be ruled out between included subjects with complete covariate data and those excluded because of missing data, the distributions of age, sex, and race were essentially comparable between these two groups, and most missing values for reflux occurred because certain study sites did not include this variable in their questionnaires. Our ability to detect GxE associations may also have been limited by the manner in which we modeled (or assessed) environmental exposures (e.g., at least weekly heartburn). Inclusion of more precise covariates, such as duration or timing of heartburn relative to diagnosis, might have helped capture additional interactions, but such information was often unavailable. While this analysis focused on seven SNPs previously shown to be associated at the genome-wide level with altered risk of EA or BE, a number of variants not satisfying this stringent threshold in marginal analysis may nonetheless exhibit significant interactions with reflux or BMI. Genome-wide GxE studies would be required to further explore this possibility. Lastly, some potential for measurement error of the exposures examined, especially reflux, should be acknowledged, given our consortium-based pooled study population. These exposures were ascertained across two decades in multiple studies on different continents. Nevertheless, an extensive effort was devoted to ensuring accurate data harmonization, as documented in several recent pooled analyses (51–53).

Our report describes a novel interaction between an intergenic germline polymorphism and weekly reflux symptoms in relation to risk of Barrett's Esophagus, the known precursor of esophageal adenocarcinoma. Further studies will be necessary to validate these findings in external populations and investigate the potential biological basis for differential disease risk associated with reflux in the presence of this variant.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

### Financial support

This work was directly supported by the National Institutes of Health [R01HL114901 to J.Y. Dai and J.D. Tapsoba, P01CA53996 to J.Y. Dai, R01CA136725 to T.L. Vaughan and D.C. Whiteman, T32CA009168 to T.L. Vaughan and M. F. Buas, and K05CA124911 to T.L. Vaughan]. D.C. Whiteman was supported by a Research Fellowship from the National Health and Medical Research Council of Australia.

We thank Stuart MacGregor and Puya Gharakhani for conducting and providing imputed genotypes based on the 1000 genomes project.

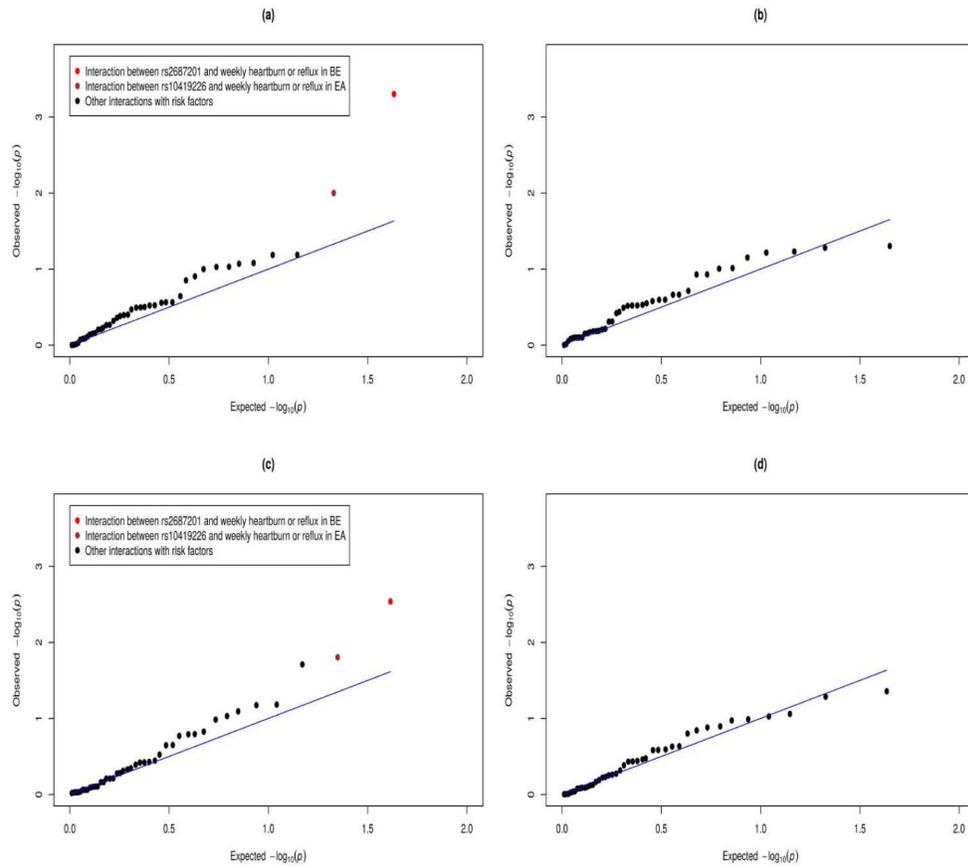
## References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *Int J Cancer*. 2010; 127:2893–917. [PubMed: 21351269]
2. Yousef F, Cardwell C, Cantwell MM, Galway K, Johnston BT, Murray L. The incidence of esophageal cancer and high-grade dysplasia in Barrett's esophagus: a systematic review and meta-analysis. *Am J Epidemiol*. 2008; 168:237–49. [PubMed: 18550563]
3. Blot WJ, Devesa SS, Kneller RW, Fraumeni JF Jr. Rising incidence of adenocarcinoma of the esophagus and gastric cardia. *JAMA*. 1991; 265:1287–9. [PubMed: 1995976]
4. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *J Natl Cancer Inst*. 2005; 97:142–146. [PubMed: 15657344]
5. Brown LM, Devesa SS, Chow WH. Incidence of adenocarcinoma of the esophagus among white americans by sex, stage, and age. *J Natl Cancer Inst*. 2008; 100:1184–7. [PubMed: 18695138]
6. Stavrou EP, McElroy HJ, Baker DF, Smith G, Bishop JF. Adenocarcinoma of the oesophagus: incidence and survival rates in new south wales, 1972–2005. *Med J Aust*. 2009; 191:310–314. [PubMed: 19769552]
7. Reid BJ, Li X, Galiepeau PC, Vaughan TL. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. *Nature Review*. 2010; 10:87–101.
8. Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med*. 1990; 340:825–831. [PubMed: 10080844]
9. Anderson LA, Watson RG, Murphy SJ, Johnston BT, Comber H, Mc Guigan J, et al. Risk factors for Barrett's oesophagus and oesophageal adenocarcinoma: results from the finbar study. *World J Gastroenterol*. 2007; 13:1585–94. [PubMed: 17461453]
10. Engel LS, Chow WH, Vaughan TL, Gammon MD, Risch HA, Stanford JL, et al. Population attributable risks of esophageal and gastric cancers. *Cancer Epidemiol Biomarkers Prev*. 2003; 95:1404–1413.
11. Steevens J, Schouten LJ, Driessen AL, Huysentruyt CJ, Keulemans YC, Goldbohm RA, et al. A prospective cohort study on overweight, smoking, alcohol consumption, and risk of Barrett's esophagus. *Cancer Epidemiol Biomarkers Prev*. 2011; 20:345–358. [PubMed: 21173169]
12. Olsen CM, Pandeya N, Green AC, Webb PM, Whiteman DC. for the Australian Cancer Study. Population attributable fractions of adenocarcinoma of the esophagus and gastrophageal junction. *Am J Epidemiol*. 2011; 174:582–90. [PubMed: 21719746]

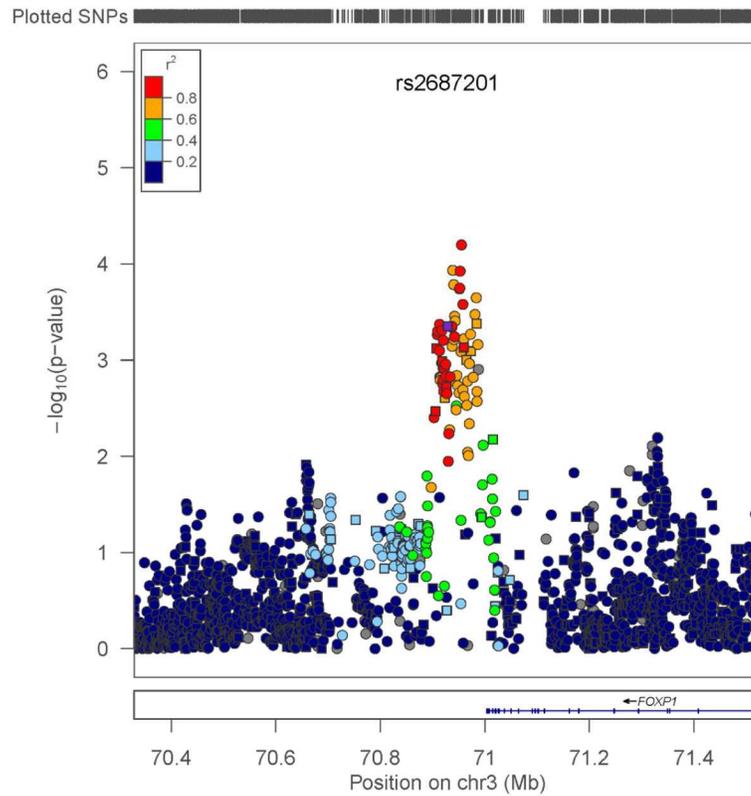
13. Corley DA, Kubo A, Zhao W. Abdominal obesity and the risk of esophageal and gastric cardia carcinomas. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:352–358. [PubMed: 18268119]
14. Johansson J, Hakansson HO, Mellblom L, Kempas A, Johansson KE, Granath F, et al. Risk factors for Barrett' s oesophagus: a population-based approach. *Scand J Gastroenterol.* 2007; 42:148–156. [PubMed: 17327933]
15. Moons LM, Kuipers EJ, Rygiel AM, Groothuismink AZ, Geldof H, Bode WA, et al. COX-2 CA-haplotype is a risk factor for the development of esophageal adenocarcinoma. *Am J Gastroenterol.* 2007; 102:2373–9. [PubMed: 17581270]
16. Kristinsson JO, van Westerveld P, teMorsche RH, Roelofs HM, Wobbes T, Witteman BJ, et al. Cyclooxygenase-2 polymorphisms and the risk of esophageal adeno- or squamous cell carcinoma. *World J Gastroenterol.* 2009; 15:3493–7. [PubMed: 19630103]
17. Babar M, Ryan AW, Anderson LA, Segurado R, Turner G, Murray LJ, et al. Genes of the interleukin-18 pathway are associated with susceptibility to Barrett' s esophagus and esophageal adenocarcinoma. *Am J Gastroenterol.* 2012; 107:1331–41. [PubMed: 22664470]
18. Menke V, van Zoest KP, Moons LM, Hansen B, Pot RG, Siersema PD, et al. NcoI TNF- $\beta$  gene polymorphism and TNF expression are associated with an increased risk of developing Barrett' s esophagus and esophageal adenocarcinoma. *Scand J Gastroenterol.* 2012; 47:378–386. [PubMed: 22243485]
19. Casson AG, Zheng Z, Porter GA, Guernsey DL. Genetic polymorphisms of microsomal epoxide hydroxylase and glutathione S-transferases M1, T1 and P1, interactions with smoking, and risk for esophageal (Barrett) adenocarcinoma. *Cancer Detect Prev.* 2006; 30:423–431. [PubMed: 17064856]
20. Bull LM, White DL, Bray M, Nurgalieva Z, El-Serag HB. Phase I and II enzyme polymorphisms as risk factors for Barrett' s esophagus and esophageal adenocarcinoma: a systematic review and meta-analysis. *Dis Esophagus.* 2009; 22:571–87. [PubMed: 19222528]
21. Doecker J, Zhao ZZ, Pandeya N, Sadeghi S, Stark M, Green AC, et al. Polymorphisms in MGMT and DNA repair genes and the risk of esophageal adenocarcinoma. *Int J Cancer.* 2008; 123:174–180. [PubMed: 18386788]
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; 44:1131–1136. [PubMed: 22961001]
23. Palles C, Chegwidzen L, Li X, Findlay JM, Farnham G, Castro Giner F, et al. Polymorphisms near *TBX5* and *GDF7* are associated with increased risk for Barrett' s esophagus. *Gastroenterology.* 2015; 148:367–78. [PubMed: 25447851]
24. 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; 45:1487–1493. [PubMed: 24121790]
25. Ek WE, Levine DM, D' Amato M, Pedersen NL, Magnusson PK, Bresso F, et al. Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett' s esophagus and gastroesophageal reflux. *JNCI.* 2013; 105:1711–8. [PubMed: 24168968]
26. di Martino E, Hardie LJ, Wild CP, Gong YY, Olliver JR, Gough MS, et al. The NAD(P)H:quinone oxidoreductase I C609T polymorphism modifies the risk of Barrett esophagus and esophageal adenocarcinoma. *Genet Med.* 2007; 9:341–7. [PubMed: 17575500]
27. Zhai R, Liu G, Asomaning K, Su L, Kulke MH, Heist RS, et al. Genetic polymorphisms of VEGF, interactions with cigarette smoking exposure and esophageal adenocarcinoma risk. *Carcinogenesis.* 2008; 29:2330–4. [PubMed: 18780893]
28. Cheung WY, Zhai R, Bradbury P, Hopkins J, Kulke MH, Heist RS, et al. Single nucleotide polymorphisms in the matrix metalloproteinase gene family and the frequency and duration of gastroesophageal reflux disease influence the risk of esophageal adenocarcinoma. *Int J Cancer.* 2012; 131:2478–86. [PubMed: 22422400]
29. Zhai R, Chen F, Liu G, Su L, Kulke MH, Asomaning K, et al. Interactions among genetic variants in apoptosis pathway genes, reflux symptoms, body mass index, and smoking indicate two distinct etiologic patterns of esophageal adenocarcinoma. *J Clin Oncol.* 2010; 28:2445–51. [PubMed: 20385987]

30. Wu IC, Zhao Y, Zhai R, Liu CY, Chen F, Ter-Minassian M, et al. Interactions between genetic polymorphisms in the apoptotic pathway and environmental factors on esophageal adenocarcinoma risk. *Carcinogenesis*. 2011; 32:502–506. [PubMed: 21212151]
31. Zhai R, Zhao Y, Liu G, Ter-Minassian M, Wu IC, Wang Z, et al. Interactions between environmental factors and polymorphisms in angiogenesis pathway genes in esophageal adenocarcinoma risk: a case-only study. *Cancer*. 2012; 118:804–11. [PubMed: 21751195]
32. Prentice RL, Huang Y, Hinds DA, Peters U, Cox DR, Beilharz E, et al. Variation in the FGFR2 gene and the effect of a low-fat dietary pattern on invasive breast cancer. *Cancer Epidemiol Biomarker Prev*. 2010; 19:74–79.
33. Kooperberg C, LeBlanc M. Increasing the power of identifying gene X gene interactions in genome-wide association studies. *Genet Epidemiol*. 2008; 32:255–263. [PubMed: 18200600]
34. Dai JY, Kooperberg C, LeBlanc M, Prentice RL. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika*. 2012; 99:929–44. [PubMed: 23843674]
35. Prentice RL, Huang Y, Hinds DA, Peters U, Cox DR, Beilharz E, et al. Variation in the FGFR2 gene and the effect of a low-fat dietary pattern on invasive breast cancer. *Cancer Epidemiol Biomarker Prev*. 2010; 19:74–79.
36. 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; 34:591–602. [PubMed: 20718045]
37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*. 1995; 57:289–300.
38. 1000 Genomes Project. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 135:0–9.
39. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010; 28:1045–1048. [PubMed: 20944595]
40. 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; 40(Database issue):D930–4. [PubMed: 22064851]
41. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res*. 2013; 41(Database issue):D56–63. [PubMed: 23193274]
42. The GTEx Consortium. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015; 348:648–660. [PubMed: 25954001]
43. Lam EW, Brosens JJ, Gomes AR, Koo CY. Forkhead box proteins: tuning forks for transcriptional harmony. *Nat Rev Cancer*. 2013; 13:482–95. [PubMed: 23792361]
44. Hannenhalli S, Kaestner KH. The evolution of Fox genes and their role in development and disease. *Nat Rev Genet*. 2009; 10:233–40. [PubMed: 19274050]
45. Myatt SS, Lam EW. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer*. 2007; 7:847–59. [PubMed: 17943136]
46. Shu W, Lu MM, Zhang Y, Tucker PW, Zhou D, Morrissey EE. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development*. 2007; 134:1991–2000. [PubMed: 17428829]
47. Banham AH, Beasley N, Campo E, Fernandez PL, Fidler C, Gatter K, et al. The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p. *Cancer Res*. 2001; 61:8820–9. [PubMed: 11751404]
48. Fox SB, Brown P, Han C, Ashe S, Leek RD, Harris AL, et al. Expression of the forkhead transcription factor FOXP1 is associated with estrogen receptor alpha and improved survival in primary human breast carcinomas. *Clin Cancer Res*. 2004; 10:3521–7. [PubMed: 15161711]
49. Banham AH, Connors JM, Brown PJ, Cordell JL, Ott G, Sreenivasan G, et al. Expression of the FOXP1 transcription factor is strongly associated with inferior survival in patients with diffuse large B-cell lymphoma. *Clin Cancer Res*. 2005; 11:1065–72. [PubMed: 15709173]

50. Koon HB, Ippolito GC, Banham AH, Tucker PW. FOXP1: a potential therapeutic target in cancer. *Expert Opin Ther Targets*. 2007; 11:955–65. [PubMed: 17614763]
51. Cook MB, Corley DA, Murray LJ, Liao LM, Kamangar F, Ye W, et al. Gastroesophageal Reflux in Relation to Adenocarcinomas of the Esophagus: A Pooled Analysis from the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON). *PLoS One*. 2014; 9:e103508. [PubMed: 25075959]
52. Hoyo C, Cook NB, Kamangar F, Freedman ND, Whiteman DC, Bernstein L, Brown LM, et al. Body mass index in relation to oesophageal and oesophagogastric junction adenocarcinomas: a pooled analysis from the International BEACON Consortium. *Int J Epidemiol*. 2012; 41:1706–18. [PubMed: 23148106]
53. Cook MB, Kamangar F, Whiteman DC, Freedman ND, Gammon MD, Bernstein L, et al. Cigarette smoking and adenocarcinomas of the esophagus and esophagogastric junction: a pooled analysis from the international BEACON consortium. *J Natl Cancer Inst*. 2010; 102:1344–53. [PubMed: 20716718]



**Figure 1.** Quantile-quantile plots of 42 interaction p-values for seven single nucleotide polymorphisms (SNPs) under investigation. P-values were derived from interaction models in which SNP genotype was modeled as a continuous variable with values of 0/1/2 in (a). A randomly selected set of seven SNPs from the BEAGESS study were also plotted for comparison in (b) for the trend interaction test. P-values from the interaction test with SNP genotypes modeled as 0/1 were displayed in (c), along with the random set of seven SNPs in (d).



**Figure 2.**

Regional association plot of genotyped and imputed single nucleotide polymorphisms (SNPs) in proximity to rs2687201, summarizing evidence of interaction with at least weekly heartburn or reflux in relation to risk of Barrett's esophagus (BE). The top portion of the figure has physical position along the x-axis, and  $-\log_{10}(p\text{-value})$  for the interaction term on the y-axis. Each dot on the plot represents the p-value of the interaction between a SNP (coded as 0/1) and GERD in relation to BE risk. The index SNP is marked as a purple square, other genotyped SNPs are marked with squares, and imputed SNPs are marked by circles. The color scheme represents the pairwise correlation ( $r^2$ ) between a given SNP and the index SNP. Correlation was calculated using the HapMap CEU data. The bottom portion of the figure shows the position of genes across the region.

**Table 1**

Associations of risk factors with Barrett' s esophagus and esophageal adenocarcinoma.

Risk factors	Barrett' s esophagus		Odds ratio (95% confidence interval) <sup>†</sup>	Esophageal adenocarcinoma		Odds ratio (95% confidence interval) <sup>†</sup>
	Case (n)	Control (n)		Case (n)	Control (n)	
Categorized body mass index						
<25	426	785	Referent	246	787	Referent
25–29.9	883	944	1.77 (1.52, 2.06)	457	944	1.44 (1.20, 1.73)
≥30	752	437	3.18 (2.68, 3.76)	296	438	2.20 (1.78, 2.72)
Ever smoked cigarettes						
No	801	888	Referent	348	889	Referent
Yes	1570	1284	1.41 (1.25, 1.60)	1066	1286	1.95 (1.68, 2.27)
Weekly heartburn or reflux						
No	958	1447	Referent	566	1449	Referent
Yes	997	349	4.62 (3.97, 5.37)	438	350	3.46 (2.89, 4.13)

<sup>†</sup> Logistic regression model for association of the risk factor with case-control status adjusted for sex and age.

**Table 2**

Odds ratio (95% confidence interval) estimates for risk factors in the BEACON study according to the number of minor alleles of the indicated single nucleotide polymorphism (SNP) (rs2687201 or rs10419226) \*.

Risk factors	Number of Minor Alleles			p-value <sup>†</sup> (FDR <sup>*</sup> ) for equality test	p-value <sup>§</sup> (FDR) for trend test
	0	1	2		
<b>rs2687201</b>					
<b>Barrett' s esophagus</b>					
Categorized body mass index <sup>§</sup>	1.65 (1.46,1.86)	1.95 (1.72,2.22)	1.74 (1.33,2.26)	0.083 (0.635)	0.224 (0.783)
Ever smoked cigarettes	1.59 (1.33,1.91)	1.35 (1.12,1.62)	0.96 (0.65,1.42)	0.065 (0.909)	0.019 (0.328)
At least weekly heartburn or reflux	6.17 (4.91,7.76)	3.56 (2.85,4.44)	3.97 (2.47,6.37)	0.0005 (0.042)	0.003 (0.122)
<b>Esophageal adenocarcinoma</b>					
Categorized body mass index	1.35 (1.16,1.58)	1.55 (1.32,1.82)	1.79 (1.31,2.46)	0.125 (0.582)	0.081 (0.678)
Ever smoked cigarettes	2.19 (1.75,2.74)	1.85 (1.47,2.33)	1.37 (0.88,2.13)	0.141 (0.627)	0.067 (0.624)
At least weekly heartburn or reflux	4.05 (3.10,5.30)	3.15 (2.42,4.09)	2.48 (1.45,4.27)	0.099 (0.525)	0.066 (0.691)
<b>rs10419226</b>					
<b>Barrett' s esophagus</b>					
Categorized body mass index	1.75 (1.49,2.05)	1.92 (1.70,2.17)	1.55 (1.30,1.85)	0.811 (0.987)	0.373 (0.825)
Ever smoked cigarettes	1.43 (1.14,1.80)	1.45 (1.22,1.74)	1.36 (1.05,1.77)	0.971 (0.995)	0.785 (0.999)
At least weekly heartburn or reflux	4.04 (3.07,5.30)	4.99 (4.03,6.21)	4.74 (3.41,6.48)	0.227 (0.733)	0.407 (0.795)
<b>Esophageal adenocarcinoma</b>					
Categorized body mass index	1.54 (1.253,1.88)	1.54 (1.33,1.79)	1.33 (1.07,1.66)	0.728 (0.956)	0.378 (0.813)
Ever smoked cigarettes	1.76 (1.33,2.34)	1.99 (1.61,2.46)	2.12 (1.54,2.93)	0.415 (0.792)	0.381 (0.799)
At least weekly heartburn or reflux	2.35 (1.67,3.31)	3.81 (2.97,4.89)	4.30 (2.94,6.28)	0.010 (0.280)	0.016 (0.329)

\* FDR: false discovery rate.

<sup>†</sup>One degree of freedom test of equality of the two odds ratios among participants with no minor alleles versus those with at least one minor allele; logistic regression model for case control status including terms for sex, age, PC1-PC4, SNP genotype, the exposure variable together with its interaction with an indicator of at least one minor allele.

<sup>§</sup>P-value associated with the test for no interaction between exposure and SNP; logistic regression model for case control status including terms for sex, age, PC1-PC4, SNP genotype coded as a continuous variable (0,1,2), the exposure variable along with its interaction with the SNP.

<sup>§</sup>Body mass index in three categories: (<25, 25-<30, 30 kg/m<sup>2</sup>).

**Table 3**

Odds ratio (95% confidence interval) estimates for risk factors in the BEACON study in a joint model with all risk factors and their interactions with rs2687201 genotype.

Risk factor	Number of Minor Alleles			p-value <sup>†</sup> for equality test	p-value <sup>§</sup> for trend test
	0	1	2		
<b>Barrett' s esophagus</b>					
Categorized body mass index	1.78 (1.53, 2.08)	2.03 (1.75, 2.37)	1.77 (1.29, 2.43)	0.299	0.554
Ever smoked cigarettes	1.69 (1.34, 2.12)	1.34 (1.07, 1.70)	1.11 (0.68, 1.79)	0.102	0.075
At least weekly heartburn or reflux	7.21 (5.66, 9.19)	4.17 (3.28, 5.31)	4.50 (2.73, 7.43)	0.001	0.004
<b>Esophageal adenocarcinoma</b>					
Categorized body mass index	1.43 (1.17, 1.75)	1.35 (1.11, 1.65)	1.83 (1.26, 2.66)	0.906	0.451
Ever smoked cigarettes	2.05 (1.49, 2.83)	1.58 (1.16, 2.17)	1.10 (0.61, 1.99)	0.127	0.070
At least weekly heartburn or reflux	4.01 (2.95, 5.45)	3.22 (2.38, 4.36)	2.54 (1.39, 4.66)	0.185	0.124

<sup>†</sup>One degree of freedom test of equality of the two odds ratios among participants with no minor alleles versus those with at least one minor allele; logistic regression model for case control status including terms for sex, age, PC1-PC4, SNP genotype, the three exposure variables together with their interactions with an indicator for the presence of at least one minor allele.

<sup>§</sup>P-value associated with the test for no interaction between exposure and SNP; logistic regression model for case control status including terms for sex, age, PC1-PC4, SNP genotype coded as a continuous variable (0,1,2), the three exposure variables along with their interaction with the SNP.

**Table 4**

Odds ratio estimates for four strata defined by rs2687201 genotype and history of gastroesophageal reflux disease (GERD). The stratum with no minor alleles and less than weekly heartburn or reflux is the baseline group.

	Frequency in study population <sup>§</sup>			Odds ratio* (95% confidence interval)
		BE cases (n)	Control (n)	
GERD-negative, no minor alleles	30%	391	733	Referent
GERD-negative, at least one minor allele	34%	567	714	1.50 (1.27, 1.77)
GERD-positive, no minor alleles	16%	460	150	6.17 (4.93, 7.73)
GERD-positive, at least one minor allele	20%	537	199	5.44 (4.42, 6.70)

\* Logistic regression model for association of SNP and GERD with the BE case-control status adjusted for sex, age and four eigenvectors.

<sup>§</sup> BE cases and controls combined.