



Published in final edited form as:

*Nat Genet.* 2013 December ; 45(12): 1487–1493. doi:10.1038/ng.2796.

## A Genome-Wide Association Study Identifies New Susceptibility Loci for Esophageal Adenocarcinoma and Barrett's Esophagus

David M. Levine<sup>1</sup>, Weronica E. Ek<sup>2</sup>, Rui Zhang<sup>1</sup>, Xinxue Liu<sup>3</sup>, Lynn Onstad<sup>4</sup>, Cassandra Sather<sup>5</sup>, Pierre Lao-Sirieix<sup>3</sup>, Marilie D. Gammon<sup>6</sup>, Douglas A. Corley<sup>7</sup>, Nicholas J. Shaheen<sup>8</sup>, Nigel C. Bird<sup>9</sup>, Laura J. Hardie<sup>10</sup>, Liam J. Murray<sup>11</sup>, Brian J. Reid<sup>4,12</sup>, Wong-Ho Chow<sup>13</sup>, Harvey A. Risch<sup>14</sup>, Olof Nyrén<sup>15</sup>, Weimin Ye<sup>15</sup>, Geoffrey Liu<sup>16</sup>, Yvonne Romero<sup>17,18</sup>, Leslie Bernstein<sup>19</sup>, Anna H. Wu<sup>20</sup>, Alan G. Casson<sup>21</sup>, Stephen Chanock<sup>22</sup>, Patricia Harrington<sup>23,24,25</sup>, Isabel Caldas<sup>25</sup>, Irene Debiram-Beecham<sup>3</sup>, Carlos Caldas<sup>25,26</sup>, Nicholas K. Hayward<sup>27</sup>, Paul Pharoah<sup>23,24,25</sup>, Rebecca Fitzgerald<sup>3</sup>, Stuart MacGregor<sup>2</sup>, David C. Whiteman<sup>28</sup>, and Thomas L. Vaughan<sup>4</sup>

<sup>1</sup>Department of Biostatistics, University of Washington, School of Public Health, Seattle, WA, USA

<sup>2</sup>Statistical Genetics, QIMR Berghofer Medical Research Institute, Queensland, Australia

<sup>3</sup>MRC Cancer Cell Unit, Hutchison-MRC Research Centre and University of Cambridge, Cambridge, UK

<sup>4</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

<sup>5</sup>Genomics Shared Resource, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Correspondence should be addressed to: T.L.V. (tvaughan@u.washington.edu).

### URLS

BEACON, <http://beacon.tlvnet.net>;

eQTL resources, <http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl>;

GWASTools, <http://www.bioconductor.org/packages/devel/bioc/html/GWASTools.html>;

HaploReg, <http://www.broadinstitute.org/mammals/haploreg>;

MaCH, <http://www.sph.umich.edu/csg/abecasis/mach/index.html>;

METAL, <http://www.sph.umich.edu/csg/abecasis/metal/index.html>;

R, <http://www.R-project.org>;

RegulomeDB, <http://www.regulomedb.org/index>;

SNPRelate, <http://cran.r-project.org/web/packages/SNPRelate/index.html>;

The Wellcome Trust Case Control Consortium, <http://www.wtccc.org.uk>;

UCSC Genome Resources, <https://genome.cse.ucsc.edu>;

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

### AUTHOR CONTRIBUTIONS

LB, NB, CC, AC, SC, WHC, DC, RF, MG, LH, NH, GL, LM, ON, PP, BR, HR, YR, NS, TV, DW, AW, WY planned the study.

PP, TV, DW obtained funding.

CC, RF, DL, TV, DW oversaw the study.

LB, NB, IC, AC, DC, IRB, RF, MG, LH, NH, GL, LM, ON, BR, YR, NS, TV, DW, AW, WY recruited participants.

PH, PLS, CS processed samples.

PH, PLS, CS undertook genotyping.

WE, DL, XL, SM, LO, PP, RZ analyzed data.

WE, RF, DL, SM, LO, TV, DW drafted manuscript.

All authors critically revised the paper for intellectual content.

TV and DW jointly led the project.

<sup>6</sup>Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, NC, USA

<sup>7</sup>Division of Research and San Francisco Medical Center, Kaiser Permanente Northern California, CA, USA

<sup>8</sup>Division of Gastroenterology and Hepatology, UNC School of Medicine, University of North Carolina, Chapel Hill, NC, USA

<sup>9</sup>Department of Oncology, The Medical School, University of Sheffield, UK

<sup>10</sup>Division of Epidemiology, University of Leeds, Leeds, UK

<sup>11</sup>Centre for Public Health, Queen's University, Belfast, UK

<sup>12</sup>Division of Human Biology; Fred Hutchinson Cancer Research Center; Seattle, WA, USA

<sup>13</sup>Department of Epidemiology, MD Anderson Cancer Center, Houston, TX, USA

<sup>14</sup>Yale University School of Medicine, Department of Epidemiology and Public Health, New Haven, CT, USA

<sup>15</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

<sup>16</sup>Pharmacogenomic Epidemiology, Ontario Cancer Institute, Toronto, ON, Canada

<sup>17</sup>Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA

<sup>18</sup>On behalf of the Romero Registry, Mayo Clinic, Rochester, MD, USA

<sup>19</sup>Department of Populations Sciences, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, CA, USA

<sup>20</sup>Department of Preventive Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA

<sup>21</sup>Department of Surgery, University of Saskatchewan, Saskatoon SK, Canada

<sup>22</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

<sup>23</sup>Department of Public Health, University of Cambridge, Cambridge, UK

<sup>24</sup>Department of Primary Care, University of Cambridge, Cambridge, UK

<sup>25</sup>Department of Oncology, University of Cambridge, Cambridge, UK

<sup>26</sup>Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

<sup>27</sup>Oncogenomics, QIMR Berghofer Medical Research Institute, Queensland, Australia

<sup>28</sup>Cancer Control, QIMR Berghofer Medical Research Institute, Queensland, Australia

---

Esophageal adenocarcinoma is a cancer with rising incidence and poor survival. Most such cancers arise in a specialized intestinal metaplastic epithelium, which is diagnostic of Barrett's esophagus. In a genome-wide association study, we compared esophageal adenocarcinoma cases (n=2,390) and patients with precancerous Barrett's esophagus

(n=3,175) with 10,120 controls in two phases. For the combined case group we identified three new associations. The first is on 19p13 (rs10419226,  $P=3.6\times 10^{-10}$ ) in *CRTCI* (CREB-regulated transcription co-activator), whose aberrant activation has been associated with oncogenic activity. A second is on 9q22 (rs11789015,  $P=1.0\times 10^{-9}$ ) in *BARX1*, which encodes a transcription factor important in esophageal specification. A third is on 3p14 (rs2687201,  $P=5.5\times 10^{-9}$ ) near the transcription factor, *FOXP1*, which regulates esophageal development. We also refine a previously-reported association with Barrett's esophagus near the putative tumor suppressor gene, *FOXF1*, on 16q24, and extend our findings to now include esophageal adenocarcinoma.

A genetic component to the development of Barrett's esophagus and esophageal adenocarcinoma has long been suspected based on prior studies in unrelated individuals and familial disease clusters.<sup>1-9</sup> This study leverages the resources of the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON), and combines high-quality largely population-based epidemiologic studies of esophageal adenocarcinoma and Barrett's esophagus conducted over two decades.

For the discovery phase, we used 1,516 EA cases, 2,416 Barrett's esophagus cases and 3,209 controls, all of European ancestry, after rigorous quality control (QC) procedures (Online Methods) were applied to the genotyping data. The cases and 2,187 of the controls were collected by investigators in BEACON from cohort and case-control studies conducted in Western Europe, Australia, and North America. An additional 1,022 cancer-free controls were obtained from a study of melanoma<sup>10</sup> and included to increase statistical power. All cases were histologically confirmed. The distribution of samples by study is given in Supplementary Table 1 and their demographic characteristics in Supplementary Table 2. All samples were genotyped on the Illumina HumanOmni1-Quad platform.

We performed association analyses on the 922,031 autosomal and X chromosome SNPs that passed QC using an additive logistic regression model implemented in GWASTools<sup>11</sup> including as covariates age, sex and the first four eigenvectors from principal component analysis (PCA). To assess variants not present on the Illumina HumanOmni1-Quad, we performed imputation for each region of interest (Tables 1 and 2). To identify shared genetic susceptibility loci for the two conditions, we treated the esophageal adenocarcinoma and Barrett's esophagus cases as a single phenotype. The results are shown in a Manhattan plot in Figure 1. We also compared Barrett's esophagus and esophageal adenocarcinoma separately against the controls (Supplementary Figure 1). QQ plots are shown in Supplementary Figure 2. The genomic inflation factor  $\lambda$  for analysis of the combined case group was 1.039 (1.084 excluding the first four principal components); no evidence for population stratification in our data set is indicated.

We selected 94 associated ( $p < 10^{-4}$ ) SNPs for replication. Of these, 87 were genotyped in 874 histologically confirmed esophageal adenocarcinoma cases from the Stomach and Oesophageal Cancer Study (SOCS), 759 histologically confirmed Barrett's esophagus cases from the UK Barrett's Oesophagus Gene Study (UK Gene Study) and 6,911 controls, of which 1,711 were from the SEARCH Study and 5,200 from the Wellcome Trust Case Control Consortium 2 (WTCCC2).<sup>12</sup> All SOCS, UK Gene Study and SEARCH samples

self-identified as Caucasians and were genotyped on a Fluidigm 96.96 Dynamic Array IFC. WTCCC2 subjects were of European ancestry as determined by projection onto the first two principal components of a PCA of HapMap individuals and were genotyped on a custom version of the Illumina Human1.2M-Duo array. Replication analysis was done using an additive logistic regression model with sex as a covariate. METAL software<sup>13</sup> was used for meta-analysis of the discovery and replication data sets.

The three loci that reached genome-wide significance ( $P < 5 \times 10^{-8}$ ) in the combined case group meta-analysis are given in Table 1 and results for all replicated SNPs are given in Supplementary Table 3. The most strongly associated SNP for each of the three loci had similar ORs in Barrett's esophagus and esophageal adenocarcinoma. None of the top imputed SNPs showed substantially stronger association than the genotyped SNPs.

The most significant locus was at 19p13 (Figure 2a); rs10419226,  $P_{\text{META}(\text{BE}+\text{EA})} = 3.55 \times 10^{-10}$ , odds ratio (OR) = 1.18, 95% confidence interval (CI) = 1.12 – 1.24. Five imputed SNPs in high LD with rs10419226 ( $r^2 > 0.85$ ) were genome-wide significant in the combined discovery data set and two were also significant in the Barrett's esophagus discovery set (Supplementary Table 4). A second significant locus was at 9q22.32 (Figure 2b) for rs11789015  $P_{\text{META}(\text{BE}+\text{EA})} = 1.02 \times 10^{-9}$ , OR(CI) = 0.83 (0.79–0.88). The third genome-wide significant locus was at 3p13 (rs2687201) near *FOXP1* (Figure 2c) with  $P_{\text{META}(\text{BE}+\text{EA})} = 5.47 \times 10^{-9}$ , OR(CI) = 1.18 (1.12 – 1.25)

A previous study of Barrett's esophagus identified the SNP rs9936833, near the putative tumor suppressor gene *FOXF1*.<sup>14</sup> A subset of the BEACON samples from the present study (2,398 Barrett's esophagus cases, 2,167 controls) were used in the replication analysis of rs9936833 ( $P = 5.13 \times 10^{-4}$ , OR(CI) = 1.16 (1.07 – 1.27)). With the additional samples used here (18 Barrett's esophagus cases and 1,042 controls) the p-value was more significant ( $P_{\text{DISCOVERY}(\text{BE})} = 1.69 \times 10^{-4}$ ), but with no change in the odds ratio (OR(CI) = 1.16 (1.07 – 1.26)). This SNP was associated with esophageal adenocarcinoma ( $P_{\text{DISCOVERY}(\text{EA})} = 2.06 \times 10^{-3}$ , OR(CI) = 1.16 (1.05 – 1.27)) (Table 2).

Examining the regional association plot of chromosome 16 near rs9936833 for the combined data (Figure 2d) we identified four nearby SNPs that had more significant P-values than rs9936833. To test the independence of these associations, this region was fine-mapped using conditional analysis to assess whether the associations were due to one or multiple variants. The discovery and replication results are shown in Table 2, with conditional analysis results in Table 3. Our results indicate that these markers define a set of complex susceptibility alleles; with between two and four independent loci.

The previous study of Barrett's esophagus also identified rs9257809 in the major histocompatibility complex (HLA) as being associated with increased risk of Barrett's esophagus.<sup>14</sup> The significance in our extended data set ( $P_{\text{DISCOVERY}(\text{BE})} = 0.11$ , OR(CI) = 1.11 (0.99 – 1.29)) is not substantially different from the original replication results ( $P = 0.083$ , OR(CI) = 1.13 (0.98 – 1.30)). We found a similar OR for esophageal adenocarcinoma as in Barrett's esophagus, suggesting the previously identified genome-wide significant Barrett's esophagus association also plays a role in esophageal adenocarcinoma risk

( $P_{\text{DISCOVERY(EA)}} = 0.09$ ,  $\text{OR(CI)} = 1.14 (0.97 - 1.32)$ ;  $P_{\text{DISCOVERY(BE+EA)}} = 0.06$ ,  $\text{OR(CI)} = 1.12 (1.01 - 1.27)$ ).

A previous report using BEACON data showed that Barrett's esophagus and esophageal adenocarcinoma risk is influenced by a large number of common genetic variants, and that a large proportion of the genes affecting risk of these two conditions are shared between them.<sup>15</sup> These findings informed our choice of analysis approach with a primary focus on the combined esophageal adenocarcinoma and Barrett's esophagus samples. The utility of this approach was borne out in the results presented here. The ORs after meta-analyses comparing esophageal adenocarcinoma cases vs. controls and Barrett's esophagus cases vs. controls for our top five loci are very similar, whereas direct comparison of the two case types revealed no significant differences (data not shown). Combining the two case types allows these SNPs to clearly achieve genome-wide significance in the combined data sets (Table 1).

The SNPs in 19p13.11, rs10419226 and rs10423674, are intronic *CRTC1* (CREB-regulated transcription co-activator) variants associated with oncogenic activity.<sup>16</sup> Phosphorylation of *CRTC1* is regulated by the tumor suppressor kinase *LKB1*. Down-regulation or loss of *LKB1* expression in human esophageal cancer cell lines and patient samples resulted in activated *CRTC1* signaling and the transcriptional activation of downstream targets including *LYPD3*, which is associated with cancer metastasis leading to the promotion of esophageal cancer cell migration and invasion.<sup>16</sup>

rs10419226 has been shown to be an expression quantitative trait loci (eQTL) for *PIK3R2* in lymphoblastoid cell lines.<sup>17</sup> *PIK3R2* is known to be involved in cancer<sup>18</sup> and is expressed in gastrointestinal tumors.<sup>19</sup> *PIK3R2* is also known to interact with epidermal growth factor (EGF), which plays an important physiological role in the maintenance of esophageal and gastric tissue integrity. The biological effects of salivary EGF includes healing of oral and gastroesophageal ulcers and inhibition of gastric acid secretion.<sup>20</sup> Furthermore the EGF receptor has been found in gastrointestinal tissue and demonstrates increased expression in BE and esophageal adenocarcinoma.<sup>21</sup> The G/G genotype for the SNP EGF A61G is associated with a two- to four-fold increased risk of esophageal adenocarcinoma.<sup>22,23</sup>

There are several SNPs in high LD ( $r^2 > 0.9$ ) with rs10419226. Three of these, rs200331191, rs139340769 and rs8102046, lie in a region of probable promoter and enhancer activities across multiple cell lines.<sup>24</sup> The intronic *CRTC1* SNP rs10423674 influences age at menarche.<sup>25</sup> The basis of this pleiotropic effect is unclear, but may be related to obesity as *CRTC1*-/- mice are hyperphagic and obese.<sup>26</sup> rs10423674 is an eQTL for *PBX4* in lymphoblastoid cell lines.

The nearest gene to the peak SNP on chromosome 3 (rs2687201) is *FOXP1*. The transcription factors FOXP1 and FOXP2 cooperatively regulate lung and esophagus development and FOXP1 is a therapeutic target in cancer.<sup>27,28</sup> The FOX family is overexpressed in esophageal cancer.<sup>14</sup> There are several SNPs in high LD ( $r^2 > 0.8$ ) with rs2687201 which lie within enhancer histone marks. One of them, rs7626449, is at a site

where there is also evidence from DNase-seq for transcription factor binding in esophageal epithelial cells.<sup>29</sup>

The SNP rs11789015 lies in an intron of *BARX1*, a homeobox transcription factor known to be involved in esophageal and trachea differentiation in developing mouse embryos and associated with down-regulation of Wnt pathway activity in stomach morphogenesis and specification.<sup>30</sup> The *BARX1* promoter region is hypermethylated in gastric cancer (GC) cell lines and patient samples, with *BARX1* mRNA expression in GC tissues and cell lines reduced.<sup>31</sup> rs11789015 lies in a region where histone marks denote likely promoter activity. rs11789015 also alters a known regulatory motif for the transcription factor *FOXP1*. A correlated SNP, rs62574346 ( $r^2=0.97$  with rs11789015), resides where there is also evidence from DNase-seq for transcription factor binding in esophageal epithelial cells.<sup>29</sup>

A subset of the BEACON data presented here (Supplementary Figure 3a) formed part of the replication arm of a recent Barrett's esophagus GWAS.<sup>14</sup> A primary finding from that work was a Barrett's esophagus association at 16q24 SNP rs9936833. Here we found clear evidence this locus is also associated in esophageal adenocarcinoma ( $P_{\text{DISCOVERY(EA)}} = 2.06 \times 10^{-3}$ , OR(CI) = 1.16 (1.05 – 1.27), as did a recent small study (316 esophageal adenocarcinoma cases, 602 controls; OR(CI) = 1.21 (0.99 – 1.47)).<sup>32</sup> Two other SNPs near rs9936833, rs2178146 and rs3111601, have stronger and more significant associations in esophageal adenocarcinoma cases (Table 2 and Supplementary Figure 3b).

Since the size and direction of effect of the Barrett's esophagus-associated SNPs at 16q24 were similar in esophageal adenocarcinoma, we used the combined Barrett's esophagus and cancer data to identify other SNPs which are more significantly associated in 16q24 than rs9936833 (Table 2). One of these is rs3111601, which is in high LD ( $r^2 = 0.75$ ) with rs9936833. All of the SNPs in high LD with rs3111601 are intergenic, although rs1979654 ( $r^2=0.64$  with rs3111601) stands out as having excellent regulatory potential across a wide range of cell types and is likely to affect protein binding, chromatin structure and histone modification.<sup>29</sup> There was evidence for additional independent signals in the region at rs3950627 (38kb nearer to *FOXF1*) and rs2178146 (64kb nearer to *FOXF1*), where both had similar P-values to rs3111601 (Table 2 and Figure 2d), but were in only modest LD ( $r^2 < 0.2$ ) with rs3111601. Neither are good regulatory candidates, although rs8045253 (imputation association  $P_{\text{DISCOVERY(BE+EA)}} = 8.04 \times 10^{-5}$ ), in  $r^2=0.63$  with rs3950627, changes a motif for the transcription factor *FOXP1*. Changing the way that *FOXP1* binds to this region is particularly interesting in light of our association findings on chromosome 3.

In summary, we report the first genome-wide association study of esophageal adenocarcinoma, and the first to examine this cancer together with its precancerous lesion, Barrett's esophagus. Consistent with our findings showing extensive polygenic overlap between esophageal adenocarcinoma and Barrett's esophagus,<sup>15</sup> our most significant results were for cancer and pre-cancer combined. Together, these findings suggest that much of the genetic basis for esophageal adenocarcinoma lies in the development of Barrett's esophagus, rather than progression from Barrett's esophagus to esophageal adenocarcinoma. We found three novel genome-wide significant loci for esophageal adenocarcinoma and Barrett's esophagus combined, and extended existing findings at the *FOXF1* and *HLA* loci. One of the

novel regions is chromosome 3p13, near *FOXP1*, a gene encoding a transcription factor which regulates esophageal development. Interestingly, two of the other regions (*BARX1/9q22.32* and *FOXF1/16q24.1*) contain risk associated SNPs which disrupt binding of *FOXP1*. Further dissection of these loci is likely to lead to insights into the etiology of this rapidly fatal cancer.

## Online Methods

### DISCOVERY

**Study subjects**—Cases of Barrett’s esophagus and esophageal adenocarcinoma, together with associated population controls, were collected by investigators in the BEACON consortium. A subset of these individuals with European ancestry from epidemiologic studies conducted in Western Europe, Australia, and North America over the past twenty years were used in the Barrett’s and Esophageal Adenocarcinoma Genetic Susceptibility Study (BEAGESS) study. To increase the statistical power of the study we included additional controls from a hospital based case-control study of melanoma.<sup>10</sup> These controls (“MD Anderson controls”) were cancer-free friends or acquaintances of European ancestry who had accompanied melanoma patients to their clinical visits at the MD Anderson Cancer Center in Houston, Texas. The distribution of samples by study is given in Supplementary Table 1.

Histological confirmation of esophageal adenocarcinoma was carried out for all esophageal adenocarcinoma studies. Similarly, Barrett’s esophagus was histologically confirmed via identification of goblet cells in metaplastic columnar epithelium in a biopsy taken from the esophagus. Age, sex (Supplementary Table 2) and other esophageal adenocarcinoma/ Barrett’s esophagus risk factors were collected by all of the included studies via standardized questionnaires, usually through personal interviews. All recruited participants gave informed consent and this project was approved by the ethics boards of each participating institution.

**Genotyping**—BEAGESS specimens were shipped to the Fred Hutchinson Cancer Research Center (Seattle, WA) where they were processed and genotyped in three batches. In each batch, samples on each genotyping plate were stratified and balanced according to case/control status, study, and gender with samples assigned to plates randomly within those strata. Genotyping of DNA from buffy coat or whole blood was performed using the Illumina HumanOmni1-Quad platform. MD Anderson controls were genotyped using the Illumina HumanOmni1-Quad platform at the Johns Hopkins University Center for Inherited Disease Research (CIDR). SNP annotations were based on version H of the Illumina product files and corresponded to the Genome Reference Consortium GRCh37 release.

**Quality control**—Quality assurance and quality control (QA/QC) of the BEAGESS and MD Anderson data sets were carried out independently by the Genetics Coordinating Center at The University of Washington following standard procedures<sup>33</sup>. QA/QC of the MD Anderson data set was described previously<sup>10</sup>. BEAGESS samples with call rate < 95%, admixture of more than one DNA source, or unexpected relatedness (including unexpected duplicates) or misannotated sex that could not be explained were removed from the data set.

We looked for batch and plate effects using intensity data and allelic frequency and checked for case control associations with different experimental factors. No important batch or plate effects or case control associations with experimental factors were found. We used heterozygosity, sex chromosome intensity data, identity by descent (IBD) analysis and visualization of B allele frequency (BAF) and log R ratio (LRR) plots to identify samples that had one or more of misannotated sex, unexpected relatedness or were sample mixtures.<sup>34</sup> Two additional sample mixtures were removed from the data set. After further sample filtering to keep only unrelated European ancestry samples (see next section), 2,416 Barrett's esophagus cases, 1,516 esophageal adenocarcinoma cases and 2,187 controls remained. These samples were combined with 1,022 European ancestry controls from the MD Anderson data set for discovery analysis.

SNPs were clustered using Illumina's GenomeStudio software and defining SNP clusters using all samples with a call rate > 95%. SNPs that had either a GenTrain score < 0.6 or a cluster separation value < 0.4 had their genotypes set to missing. Additionally, we filtered SNPs that were intensity only, had a missing call rate > 5%, had a Hardy Weinberg equilibrium p-value over controls  $\geq 10^{-4}$ , had a discordance among any of the duplicate pairs, or had a Mendelian error in either BEAGESS families or HapMap trios. These filters were combined with similar filters calculated for the MD Anderson data set.<sup>10</sup> Additionally, we removed a further 344 SNPs that were discordant in the same HapMap control samples (n=3) run in both the BEAGESS and MD Anderson data sets. After QA/QC a total of 926,923 SNPs remained for analysis.

**Principal Components Analysis**—We performed Principal Component Analysis (PCA) as a two-step process using the SNPRelate software. First, we used PCA to define a homogeneous set of European ancestry samples in the BEAGESS data set. We did this by running PCA on a set of 6,248 unrelated (except for six two-person families) subjects each of which was an EA case, a BE case, or a control. A majority of these subjects (~98%) self-identified their race as “White” and a scatterplot of all subjects along the axes of the first two eigenvectors showed the majority of samples formed a tight cluster (Supplementary Figure 4). Therefore, we computed the means and standard deviations (SD) of the first two eigenvectors and defined any sample that fell within a two SD rectangle of both eigenvector means to be of homogeneous European ancestry (n=6,125).

Second, we ran PCA on the combined data set (n=7,147) consisting of the BEAGESS European ancestry samples (n=6,125) and the similarly defined set of MD Anderson controls (n=1,022). The intent here was to identify eigenvectors to include as covariates in our model to adjust for population differences that were present in the remaining European-ancestry-only samples.<sup>35,36</sup> For this analysis we selected 65,774 SNPs that were non-monomorphic, autosomal, passed quality control, had missing call rate < 5%, minor allele frequency > 5%, did not have an LD value > 0.2 between any two SNPs in a sliding window of 500K bases, and were not in the LCT gene (2q21), HLA region or polymorphic regions on chromosomes eight (8p23) and seventeen (17q21.31). We included the first four eigenvectors as covariates in the association test model to account for population stratification by ancestry since they were significantly correlated with case-control status and a scree plot showed that the variance accounted for by each eigenvector flattened out



after these four eigenvectors (data not shown). To check that only genome-wide variation was detected we computed the absolute value of the correlation coefficient of each eigenvector against the genotypes for each SNP. We observed one small region of high correlation ( $\rho = -0.4$ ) between eigenvector one with chromosome two which may indicate long-range LD with the LCT gene.

**Statistical Analysis**—After excluding six related samples and six other samples that had missing call rate  $> 2\%$ , we ran a case-control analysis of the remaining 7,135 samples: 3,928 cases (1,514 EA and 2,414 BE) vs. 3,207 controls. We used an additive logistic regression model with case status regressed on each SNP's genotype score (coded as 0, 1, or 2 for BB, AB and AA) including covariates age, sex and the first four PCA eigenvectors to compute the odds ratios (OR) and 95% confidence intervals (95% CI) relating risk of EA or BE to a given SNP variant. To test SNPs on the X chromosome, male genotypes were coded as 0 and 2 and female genotypes as 0, 1 and 2. After filtering SNPs that did not pass quality control and SNPs with a minor allele frequency  $< 1\%$  the  $\lambda$  value was 1.04. The QQ plot is shown in Supplementary Figure 2. We also compared Barrett's esophagus and esophageal adenocarcinoma cases separately against the controls using the same model. The corresponding Manhattan and QQ plots are shown in Supplementary Figures 1 and 2, respectively. Analysis was carried out in the R statistical programming language<sup>37</sup> using the Bioconductor packages GWASTools<sup>11</sup> and SNPRelate.<sup>38</sup>

Using the combined BE + EA discovery data set we performed a step-wise series of nested logistic regression analyses to test the independence of the associations in 16q24 near rs9936833. We used the same logistic regression model and covariates as in our primary analysis, and also fitted rs3950627 as a covariate since it was the most significant SNP 100kb up or downstream of rs9936833. This conditional analysis identified rs2178146 as the most significant SNP 100kb up or downstream of rs9936833. We repeated this analysis four more times identifying and adding to each successive model rs2178146, rs3111601, rs1490865, and rs13332095, respectively, stopping when the p-value of the most significant remaining SNP had  $P > 0.01$  (Table 3).

**Imputation**—To assess the impact of variants not present on the Illumina HumanOmni1-Quad, we imputed genotypes using the MaCH<sup>39</sup> software and a European reference panel from the 1000 Genomes project. At each region in Tables 1 and 2 imputation was done in two steps. First, haplotypes were estimated in a pre-phasing step. Second, missing alleles for additional SNPs were imputed onto these phased haplotypes using Minimac<sup>39</sup> and a publicly available reference panel of haplotypes from European ancestry populations. SNPs with very different allele frequencies (a Chi Square statistic  $> 40$  in a test for difference in allele frequency) between the BEACON data and the reference panel were removed prior to the second step. SNPs with MACH imputation  $r^2 < 0.3$  (a measure of imputation quality) and SNPs with a minor allele frequency  $< 1\%$  were also removed. Association analysis between imputed SNPs and disease status was performed using the same regression model as for genotyped SNPs, but with dosage probabilities as predictors, rather than the actual genotype calls. All association tests were two-sided. Linkage disequilibrium (LD) calculations ( $r^2$ ) were computed with the discovery data when the two SNPs being compared were both

genotyped, otherwise the European ancestry samples from Phase 1 of the 1000 genomes project were used.

## REPLICATION

**SNP selection**—We selected post-QA/QC SNPs for replication that had a p-value from discovery  $< 10^{-4}$  and a minor allele frequency  $> 1\%$ . This yielded 406 SNPs: 179 from Barrett's esophagus and esophageal adenocarcinoma vs. controls, 105 from esophageal adenocarcinoma vs. controls and 122 from Barrett's esophagus vs. controls, of which 321 were unique. A subset of these SNPs ( $n=111$ ) were selected via LD pruning with PLINK using the command `clump`<sup>40</sup>; if a SNP was in LD  $> 0.5$  with any other SNP(s) in the list, the SNP with lowest p-value was selected for replication. For each of the ten SNPs on this list with the smallest p-value, we selected an extra “proxy” SNP to include in case the top SNP was not successfully genotyped in the replication set. These proxy SNPs were ones in high LD with the top SNP, but with a less significant p-value. We visually examined cluster plots of all SNPs and kept only those that were high quality. SNPs were rank ordered by p-value and replication attempted for the top 94.

**Study Subjects**—The replication cohort consisted of Barrett's esophagus cases, esophageal adenocarcinoma cases and two control sets. Barrett's esophagus cases were identified at endoscopy with a confirmed histopathological diagnosis of intestinal metaplasia from the UK Barrett's Oesophagus Gene Study. esophageal adenocarcinoma cases were selected from the Stomach Oesophageal Cancer Study and had an ICD coding of malignant neoplasm of the esophagus (C15) and a pathological diagnosis of adenocarcinoma. One set of controls came from the SEARCH study which ascertains eligible cases of breast, ovary, prostate, colorectal, melanoma and endometrial cancer from the UK Eastern Cancer Registration and Information Centre. Controls were ascertained by frequency matching on age (five-year age bands) and sex to the esophageal adenocarcinoma and Barrett's esophagus cases excluding individuals with a past history of cancer (excluding non-melanoma skin cancer). All recruited participants gave informed consent and the studies have been approved by the relevant institutional ethics review board. The other set of controls was from the Wellcome Trust Case Control Consortium 2 (WTCCC2).

**Genotyping**—Barrett's esophagus cases, esophageal adenocarcinoma cases and SEARCH controls were genotyped using the Fluidigm<sup>TM</sup> high-throughput platform and Fluidigm 96.96 Dynamic Arrays<sup>TM</sup> according to the manufacturer's instructions and read using the Fluidigm EP1<sup>TM</sup>. Each array is capable of running 96 samples against 96 SNP assays. Cases and controls were plated out in sets of 96 samples and combined into 384-well arrays for genotyping with the cases and controls mixed on each 384-well plate. Genotypes were automatically called using the BioMark Genotyping Analysis<sup>TM</sup> software, but all cluster plots were also checked manually and adjusted as necessary. The WTCCC2 controls were genotyped on a custom version of the Illumina Human1.2M-Duo array.

**Quality Control**—We filtered Barrett's esophagus cases, esophageal adenocarcinoma cases and SEARCH controls with low call rate, inconsistent gender, were duplicates or had self-reported ethnicity of “non-white” or “missing”. This left 759 Barrett's esophagus cases,

874 esophageal adenocarcinoma cases and 1711 SEARCH controls. We excluded SNPs with missing call rate  $\geq 5\%$ , with significant differential missing call rates in cases and controls ( $p < 5 \times 10^{-4}$ ), with low minor allele frequency (defined as  $< 1\%$ ), and with significant departure from Hardy-Weinberg equilibrium ( $p < 0.0005$ ), leaving 87 post-QC SNPs. We applied standard sample and SNP exclusion criteria to the WTCCC2 controls keeping 5,190 post-QC European ancestry samples. There were 67 post-QC SNPs in the WTCCC2 controls in common with the 87-post QC SNPs from the BE cases, EA cases and SEARCH controls.

**Statistical analysis**—Each of the 87 SNPs was run using an additive logistic regression model with case status regressed on the SNPs genotype and including sex as a covariate. The analysis focus was on the comparison of BE and EA cases against controls, but we also ran each case type separately against the controls. The final data set used for replication consisted of 759 BE cases and 874 EA cases. For 67 SNPs the control set consisted of 6,911 samples; 1711 SEARCH controls and 5200 WTCCC2 controls. For 20 SNPs that were not genotyped in the WTCCC2 data only the 1,711 SEARCH controls were used. The R statistical programming language was used for all analyses.

## META-ANALYSIS

We used the inverse variance-based method in the METAL software<sup>13</sup> to perform a meta-analysis of the discovery and replication data sets. This approach weights the effect size estimates ( $\beta$ -coefficients from the discovery and replication regression models) by their standard error estimates and calculates an overall Z-score and p-value. This was done separately for each sample set.

## BIOINFORMATICS/FUNCTIONAL GENOMICS

Each region of interest was interrogated using the tools eQTL browser, HaploReg, RegulomeDB, and the UCSC Genome Browser.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Patricia Christopherson, Paul Hansen, Lorna Nolan, and Terri Watson for their efforts in project management and organization of biospecimens and data; and Cathy Laurie, Cecelia Laurie and Bruce Weir for helpful suggestions regarding analyses.

This work was primarily funded by a US NIH grant (R01CA136725) to TV and DW. TV is also supported by grant K05CA124911. DW is supported by a Future Fellowship (FT0990987) from the Australian Research Council. SM is supported by an Australian NHMRC Career Development Award. NKH is supported by Research Fellowships from the National Health and Medical Research Council of Australia. YR is supported in part by a grant from the National Institutes of Health (NIDDK 02956), the Robert Wood Johnson Foundation Harold Amos Medical Faculty Development Program, and the Fraternal Order of the Eagles.

The Swedish Esophageal Cancer Study was funded by grants (R01 CA57947-03) from the National Cancer Institute, and the Swedish Cancer Society (4559-B01-01XAA, 4758-B02-01XAB). The Kaiser Permanente Study was supported by US NIH grants R01DK63616, R01CA59636) and from the California Tobacco Related Research Program (3RT-0122 and 10RT-0251). The UK Barrett's oesophagus gene study was funded by a Medical Research Council Programme grant, the UK SOCS study was funded by CRUK as well as funding from the Cambridge

NIHR biomedical research centre and the Cambridge Experimental Cancer Medicine Centre. The SEARCH grant was funded by CRUK grants C490/A10119 and C490/A10124.

Genotyping of MD Anderson controls (C. Amos, PI) were performed through the University of Texas MD Anderson Cancer Center (UTMDACC) and the Center for Inherited Disease Research (CIDR), supported in part by NIH grants R01CA100264, P30CA016672, and R01CA133996, the UTMDACC NIH SPORE in Melanoma 2P50CA093459, as well as by the Marit Peterson Fund for Melanoma Research. CIDR is supported by contract HHSN268200782096C.

This study made use of data generated by the Wellcome Trust Case Control Consortium: Funding for the project was provided by the Wellcome Trust under award 076113; a full list of the investigators who contributed to the generation of the data is available from the website (see URLs).

The Romero Registry is supported in part by the American Digestive Health Foundation “Endoscopic Research Award,” the American College of Gastroenterology “Junior Faculty Development Award,” the Glaxo Wellcome Inc. Institute for Digestive Health “Clinical Research Award,” and the Miles and Shirley Fiterman Center for Digestive Diseases at Mayo Clinic, Rochester, Minnesota. The Romero Registry also has charitable gifts from five industry partners (Affymetrix, AstraZeneca, Santarus, Takeda and Wyeth); The US Multicenter Study was funded by U01CA57949, U01CA57983, and U01CA57923.

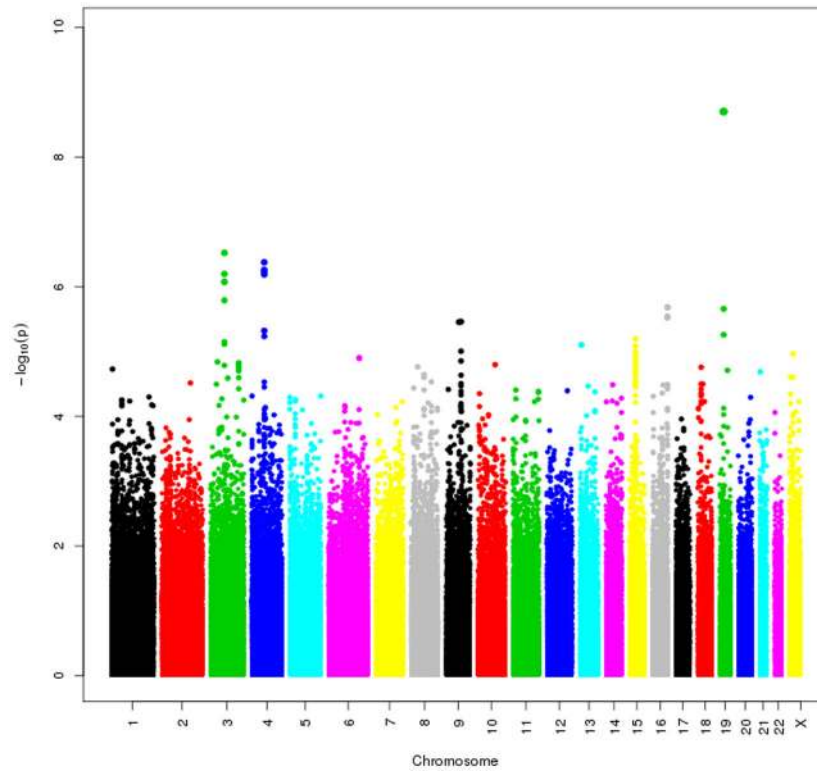
Members of the Romero Registry: (<sup>1</sup>Mayo Clinic Florida, <sup>2</sup>Mayo Clinic Rochester, <sup>3</sup>Mayo Clinic Arizona): Sami R. Achem, M.D.<sup>1</sup>, Yvonne Romero, M.D.<sup>2</sup>, David A. Ahlquist, M.D.<sup>2</sup>, Steven R. Alberts, M.D.<sup>2</sup>, Jeffrey A. Alexander, M.D.<sup>2</sup>, Mark S. Allen, M.D.<sup>2</sup>, Amindra S. Arora, MBBChir<sup>2</sup>, Jonathan B. Ashman, M.D., Ph.D.<sup>3</sup>, Pamela J. Atherton<sup>2</sup>, Lisa A. Boardman, M.D.<sup>2</sup>, Ernest P. Bouras, M.D.<sup>1</sup>, Vicki A. Bryhn<sup>2</sup>, Patrick A. Burch, M.D.<sup>2</sup>, George E. Burdick, M.D.<sup>2</sup>, Navtej S. Buttar, M.D.<sup>2</sup>, John K. Camoriano, M.D.<sup>3</sup>, John R. Cangemi, M.D.<sup>1</sup>, Stephen D. Cassivi, M.D.<sup>2</sup>, Frances K. Cayer<sup>1</sup>, Amy C. Clayton, M.D.<sup>2</sup>, Michael D. Crowell, M.D., Ph.D.<sup>3</sup>, Julie M. Cunningham, Ph.D.<sup>2</sup>, Mariza de Andrade, Ph.D.<sup>2</sup>, Piet de Groen, M.D.<sup>2</sup>, Giovanni De Petris, M.D.<sup>3</sup>, Claude Deschamps, M.D.<sup>2</sup>, Kenneth R. DeVault, M.D.<sup>1</sup>, Robert B. Diasio, M.D.<sup>2</sup>, John K. DiBaise, M.D.<sup>3</sup>, Eric S. Edell, M.D.<sup>2</sup>, Sharon Elcombe<sup>2</sup>, Charles Erlichman, M.D.<sup>2</sup>, Douglas O. Faigel, M.D.<sup>3</sup>, Tom R. Fitch, M.D.<sup>3</sup>, David E. Fleischer, M.D.<sup>3</sup>, Jean C. Fox, M.D.<sup>2</sup>, Amy E., Foxx-Orenstein, D.O.<sup>2</sup>, Dawn Francis, M.D.<sup>2</sup>, Mary B. Fredericksen<sup>2</sup>, Evanthia Galanis, M.D.<sup>2</sup>, Debra M. Geno<sup>2</sup>, Axel Grothey, M.D.<sup>2</sup>, Michael G. Haddock, M.D.<sup>2</sup>, Kevin C. Halling, M.D., Ph.D.<sup>2</sup>, Denise M. Harnois, D.O.<sup>1</sup>, Tracy W. Hilton<sup>1</sup>, Timothy Hobday, M.D.<sup>2</sup>, Lesley A. Houghton, Ph.D.<sup>1</sup>, Prasad A. Iyer, M.D.<sup>2</sup>, Dawn E. Jaroszewski, M.D.<sup>3</sup>, Aminah Jatoi, M.D.<sup>2</sup>, Robert B. Jenkins, M.D., Ph.D.<sup>2</sup>, Elizabeth A. Johnson, M.D.<sup>1</sup>, Rajni Katipamula, M.D.<sup>2</sup>, David A. Katzka, M.D.<sup>2</sup>, Sharon F. Kaufman<sup>2</sup>, Andrew P. Keaveny, M.D.<sup>1</sup>, Daniel A. Keller<sup>2</sup>, George P. Kim, M.D.<sup>1</sup>, Benjamin R. Kipp, Ph.D.<sup>2</sup>, Dora M. Lam-Himlin, M.D.<sup>2</sup>, Stephen M. Lange, M.D.<sup>1</sup>, Louis Lanza, M.D.<sup>3</sup>, Shauna N. Legrand<sup>1</sup>, Paul J. Limburg, M.D.<sup>2</sup>, Wilma L. Lingle, Ph.D.<sup>2</sup>, Wanguo Liu, Ph.D.<sup>2</sup>, G. Richard Locke, III, M.D.<sup>2</sup>, David S. Loeb, M.D.<sup>1</sup>, Lori S. Lutzke<sup>2</sup>, James A. Martenson, Jr., M.D.<sup>2</sup>, Robert R. McWilliams, M.D.<sup>2</sup>, Robert C. Miller, M.D.<sup>2</sup>, Joseph A. Murray, M.D.<sup>2</sup>, Cuong C. Nguyen, M.D.<sup>3</sup>, Francis C. Nichols, III, M.D.<sup>2</sup>, John A. Odell, M.D.<sup>1</sup>, Rahul Pannala, M.D.<sup>3</sup>, Harshita R. Paripati, M.D.<sup>2</sup>, Alexander S. Parker, Ph.D.<sup>1</sup>, Shabana F. Pasha, M.D.<sup>2</sup>, Bret T. Petersen, M.D.<sup>2</sup>, Gloria M. Petersen, Ph.D.<sup>2</sup>, Michael F. Picco, M.D.<sup>1</sup>, Henry C. Pitot, M.D.<sup>2</sup>, J. Fernando Quevedo, M.D.<sup>2</sup>, Massimo Raimondo, M.D.<sup>1</sup>, Elizabeth Rajan, M.D.<sup>2</sup>, Francisco (Pancho) C. Ramirez, M.D.<sup>2</sup>, Renee M. Root<sup>2</sup>, Helen J. Ross, M.D.<sup>3</sup>, Joseph Rubin, M.D.<sup>2</sup>, Schuyler O. Sanderson, M.D.<sup>2</sup>, Daniel J. Sargent, Ph.D.<sup>2</sup>, Daniel J. Schaid, Ph.D.<sup>2</sup>, K. Robert Shen, M.D.<sup>2</sup>, Qian Shi, Ph.D.<sup>2</sup>, Frank A. Sinicrope, M.D.<sup>2</sup>, Jeff A. Sloan, Ph.D.<sup>2</sup>, Dan C. Smith, M.D.<sup>1</sup>, David I. Smith, Ph.D.<sup>2</sup>, Thomas C. Smyrk, M.D.<sup>2</sup>, Mark E. Stark, M.D.<sup>1</sup>, Nicholas Talley, M.D., Ph.D.<sup>2</sup>, Stephen N. Thibodeau, Ph.D.<sup>2</sup>, Michael D. Van Norstrand, M.D., Ph.D.<sup>2</sup>, Michael B. Wallace, M.D.<sup>1</sup>, Kenneth K. Wang, M.D.<sup>2</sup>, Richard M. Weinshilboum, M.D.<sup>2</sup>, Dennis Wigle, M.D., Ph.D.<sup>2</sup>, Herbert C. Wolfsen, M.D.<sup>1</sup>, Louis M. Wong Kee Song, M.D.<sup>2</sup>, Timothy A. Woodward, M.D.<sup>1</sup>, Tsung-The Wu, M.D., Ph.D.<sup>2</sup>, Harry H. Yoon, M.D.<sup>2</sup>, Alan R. Zinsmeister, Ph.D.<sup>2</sup>.

## References

1. Jochem VJ, Fuerst PA, Fromkes JJ. Familial Barrett's esophagus associated with adenocarcinoma. *Gastroenterology*. 1992; 102:1400–1402. [PubMed: 1551547]
2. Crabb DW, et al. Familial gastroesophageal reflux and development of Barrett's esophagus. *Ann Intern Med*. 1985; 103:52–54. [PubMed: 4003988]
3. Poynton AR, Walsh TN, O'Sullivan G, Hennessy TP. Carcinoma arising in familial Barrett's esophagus. *Am J Gastroenterol*. 1996; 91:1855–1856. [PubMed: 8792715]
4. Romero Y, et al. Familial aggregation of gastroesophageal reflux in patients with Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterology*. 1997; 113:1449–1456. [PubMed: 9352846]
5. Melzer E, Shtoyerman R, Appelman Z, Kashtan H. Familial Barrett's adenocarcinoma. *Am J Gastroenterol*. 2006; 101:677. [PubMed: 16542304]

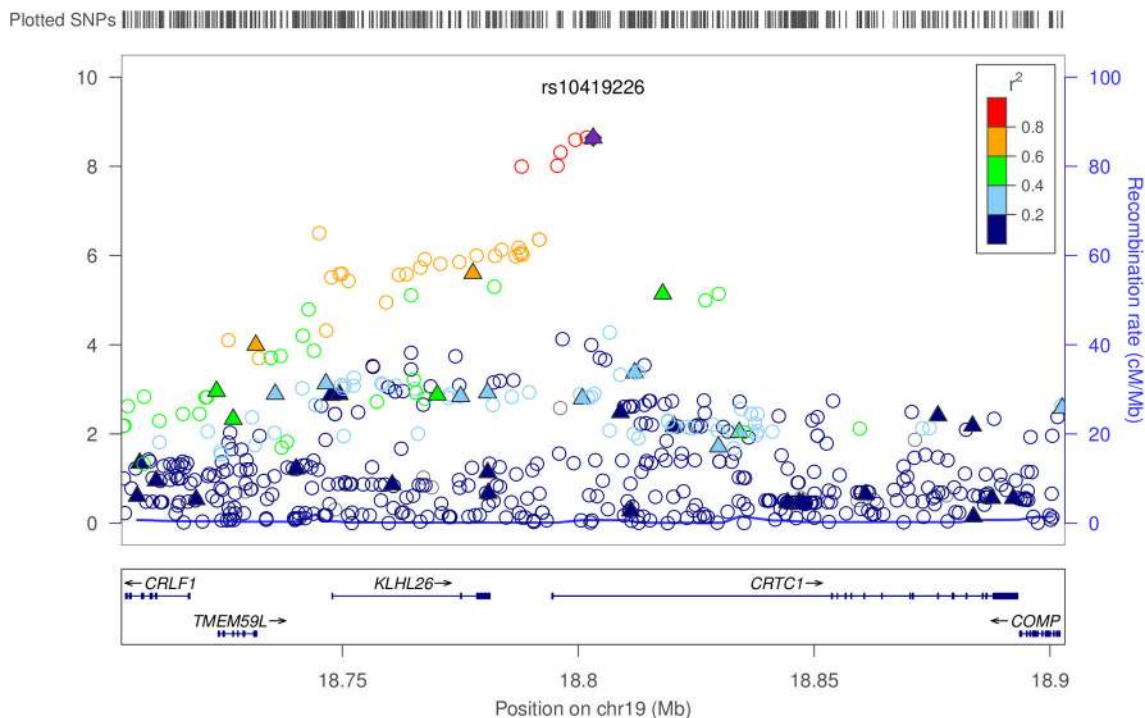
6. Groves C, Jankowski J, Barker F, Holdstock G. A family history of Barrett's oesophagus: another risk factor? *Scand J Gastroenterol.* 2005; 40:1127–1128. [PubMed: 16211720]
7. Trudgill NJ, Kapur KC, Riley SA. Familial clustering of reflux symptoms. *Am J Gastroenterol.* 1999; 94:1172–1178. [PubMed: 10235188]
8. Chak A, et al. Identification of Barrett's esophagus in relatives by endoscopic screening. *Am J Gastroenterol.* 2004; 99:2107–2114. [PubMed: 15554988]
9. Fitzgerald RC. Complex diseases in gastroenterology and hepatology: GERD, Barrett's, and esophageal adenocarcinoma. *Clin Gastroenterol Hepatol Off Clin Pr. J Am Gastroenterol Assoc.* 2005; 3:529–537.
10. Amos CI, et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Hum Mol Genet.* 2011; 20:5012–5023. [PubMed: 21926416]
11. Gogarten SM, et al. GWASTools: an R/Bioconductor package for quality control and analysis of Genome-Wide Association Studies. *Bioinforma Oxf Engl.* 2012;10.1093/bioinformatics/bts610
12. Strange A, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat Genet.* 2010; 42:985–990. [PubMed: 20953190]
13. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma Oxf Engl.* 2010; 26:2190–2191.
14. Su Z, 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]
15. Ek WE, et al. Large genetic component and overlap in Barrett's esophagus and esophageal adenocarcinoma. *JNCI.* Forthcoming.
16. Gu Y, et al. Altered LKB1/CREB-regulated transcription co-activator (CRTC) signaling axis promotes esophageal cancer cell migration and invasion. *Oncogene.* 2012; 31:469–479. [PubMed: 21706049]
17. Montgomery SB, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature.* 2010; 464:773–777. [PubMed: 20220756]
18. Cheung LWT, et al. High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov.* 2011; 1:170–185. [PubMed: 21984976]
19. Cortés I, et al. p85 $\beta$  phosphoinositide 3-kinase subunit regulates tumor progression. *Proc Natl Acad Sci U S A.* 2012; 109:11318–11323. [PubMed: 22733740]
20. Venturi S, Venturi M. Iodine in evolution of salivary glands and in oral health. *Nutr Health.* 2009; 20:119–134. [PubMed: 19835108]
21. Jankowski J, Hopwood D, Wormsley KG. Expression of epidermal growth factor, transforming growth factor alpha and their receptor in gastro-oesophageal diseases. *Dig Dis Basel Switz.* 1993; 11:1–11.
22. Menke V, et al. Functional single-nucleotide polymorphism of epidermal growth factor is associated with the development of Barrett's esophagus and esophageal adenocarcinoma. *J Hum Genet.* 2012; 57:26–32. [PubMed: 22129558]
23. Lanuti M, et al. A functional epidermal growth factor (EGF) polymorphism, EGF serum levels, and esophageal adenocarcinoma risk and outcome. *Clin Cancer Res Off J Am Assoc Cancer Res.* 2008; 14:3216–3222.
24. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489:57–74. [PubMed: 22955616]
25. Elks CE, et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet.* 2010; 42:1077–1085. [PubMed: 21102462]
26. Altarejos JY, et al. The Creb1 coactivator Crtc1 is required for energy balance and fertility. *Nat Med.* 2008; 14:1112–1117. [PubMed: 18758446]
27. Koon HB, Ippolito GC, Banham AH, Tucker PW. FOXP1: a potential therapeutic target in cancer. *Expert Opin Ther Targets.* 2007; 11:955–965. [PubMed: 17614763]
28. Shu W, et al. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Dev Camb Engl.* 2007; 134:1991–2000.

29. Boyle AP, et al. Annotation of functional variation in personal genomes using Regulome DB. *Genome Res.* 2012; 22:1790–1797. [PubMed: 22955989]
30. Woo J, Miletich I, Kim BM, Sharpe PT, Shivdasani RA. Barx1-mediated inhibition of Wnt signaling in the mouse thoracic foregut controls tracheo-esophageal septation and epithelial differentiation. *PLoS One.* 2011; 6:e22493. [PubMed: 21799872]
31. Agarwal R, et al. Abstract LB-384: BARX1, CYGB, and other epigenetic targets of gastric cancer. *Cancer Res.* 2012; 72:LB-384–LB-384.
32. Dura P, et al. Barrett associated MHC and FOXF1 variants also increase esophageal carcinoma risk. *Int J Cancer J Int Cancer.* 201310.1002/ijc.28160
33. Laurie CC, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol.* 2010; 34:591–602. [PubMed: 20718045]
34. Peiffer DA, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* 2006; 16:1136–1148. [PubMed: 16899659]
35. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–909. [PubMed: 16862161]
36. Novembre J, et al. Genes mirror geography within Europe. *Nature.* 2008; 456:98–101. [PubMed: 18758442]
37. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2006. at <<http://www.R-project.org>>
38. Zheng X, et al. A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinforma Oxf Engl.* 201210.1093/bioinformatics/bts606
39. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010; 34:816–834. [PubMed: 21058334]
40. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]



**Figure 1.**

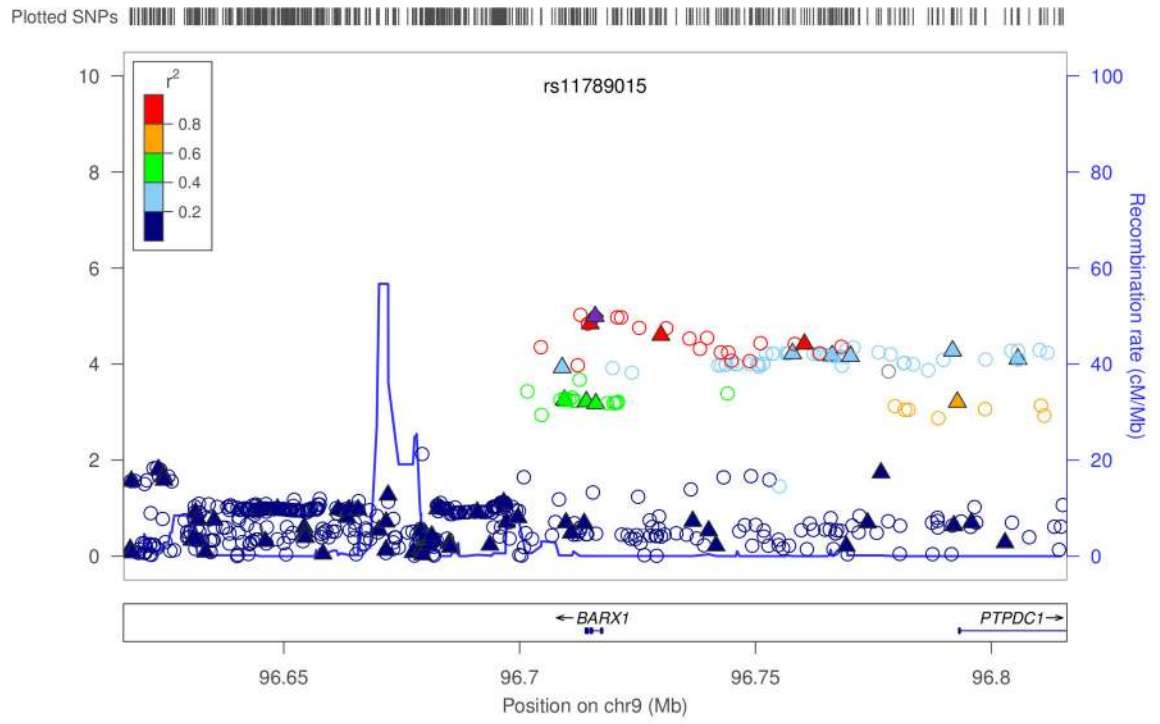
Plot of genome-wide association results from the discovery data for the combined Barrett's esophagus and esophageal adenocarcinoma cases using an additive logistic regression model with age, sex and the first four eigenvectors from principal components analysis as covariates. Results are shown for 3,928 cases (2414 Barrett's esophagus, 1514 esophageal adenocarcinoma) and 3,207 controls for 801,552 autosomal and X chromosome SNPs that passed quality control and have a minor allele frequency > 1%. Chromosomes are delineated by alternating colors, as labeled on the x-axis. The y-axis shows the  $-\log_{10}$  P-values.



**Figure 2.**

Regional association plots showing genotyped and imputed SNPs from the discovery data for the combined Barrett's esophagus + esophageal adenocarcinoma cases for three newly discovered loci (a–c) and one previously identified locus (d). Genotyped SNPs are indicated by solid triangles, and imputed SNPs are indicated by hollow circles. The top-ranked SNP at each locus is shown as a solid purple diamond, except in (d) where it is rs9936833. SNPs are ordered by genomic location. The color scheme indicates linkage disequilibrium between the top ranked SNP and other SNPs in the region using the  $r^2$  value calculated from the 1000 genomes project. The y-axis is the  $-\log_{10}$  p-value computed from 3,928 cases (2414 Barrett's esophagus, 1514 esophageal adenocarcinoma) and 3,207 controls. Imputation P values for all SNPs are plotted. Note that imputed and genotyped P-values for genotyped SNPs differ slightly because for the imputed result, the analysis was based on dosage scores, whereas with genotyped SNPs, the hard genotype calls are used. The recombination rate from CEU HapMap data (right side y axis) is shown in light blue. (a) Chromosome 19p13 region. (b) Chromosome 9q22 region. (c) Chromosome 3p13 region. (d) Chromosome 16q24 region.





**Figure 3.**

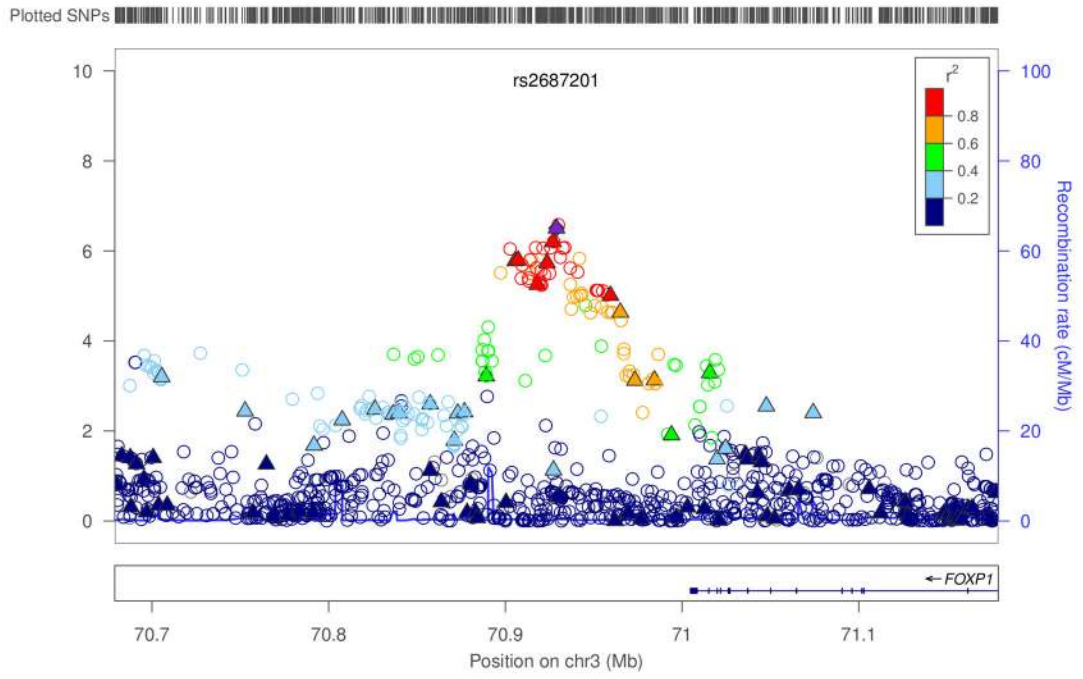


Figure 4.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

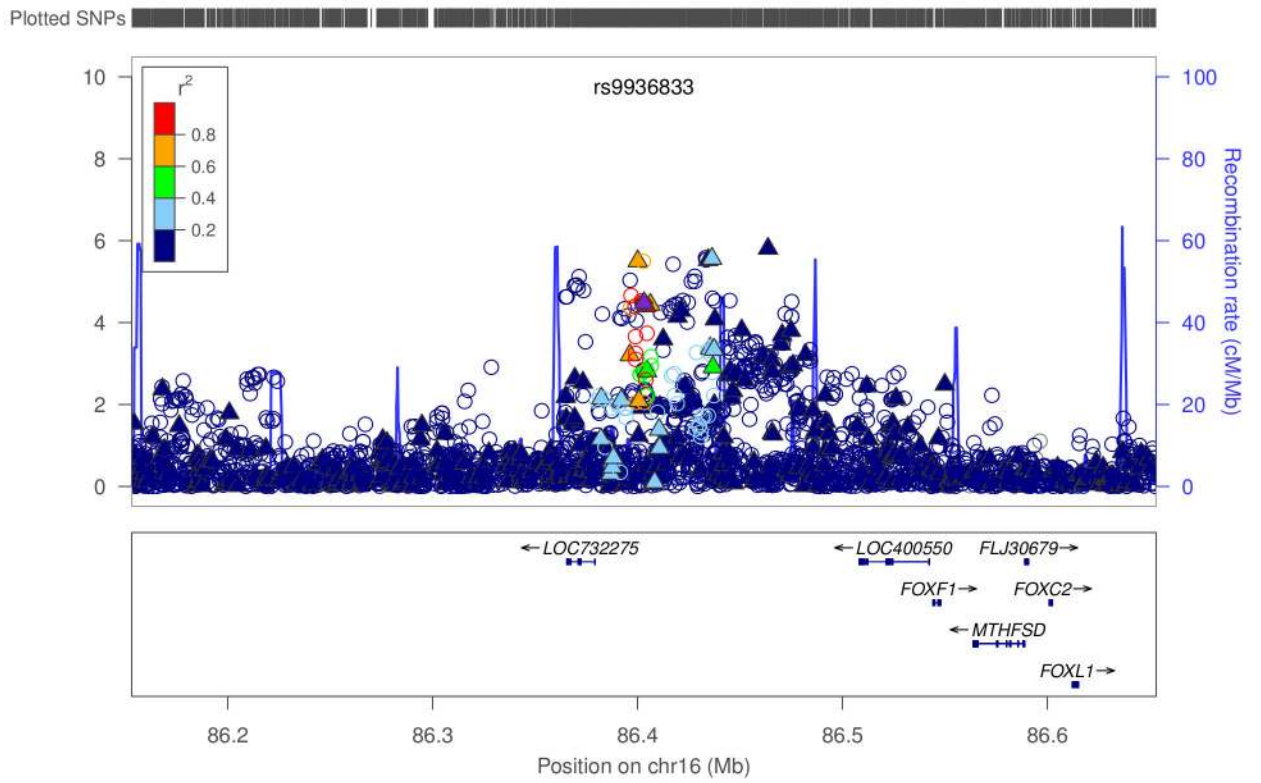


Figure 5.

**Table 1**

Top five newly identified SNPs associated with Barrett’s esophagus (BE) and esophageal adenocarcinoma (EA). Shown are the discovery, replication and meta-analysis results for Barrett’s esophagus v. controls, esophageal adenocarcinoma v. controls and the combined Barrett’s esophagus and esophageal adenocarcinoma (BE+EA) v. controls. Each cell contains the association p-value, the odds ratio (OR) and 95% confidence interval (CI) for the minor allele, and the frequency of the minor allele in cases and controls. The slight variation in the number of discovery controls reflects the use of only unrelated samples for analysis, although six two-person families are present in the data set.

	Cases/Cntrls	SNP	rs2687201	rs11789015	rs6479527	rs10419226	rs10423674
			Chr.	9	9	19	19
		Position	70928930	96716028	96858411	18803172	18817903
		CytoBand	3p13	9q22	9q22	19p13	19p13
		Nearest Gene	FOXPI	BARX1	PTPDC1	CRTC1	CRTC1
		Minor*/Major plus strand allele	T/G	G/A	T/C	A/C	T/G
<b>Barrett’s Esophagus</b>	<b>Discovery</b>	<b>P<sub>DISCOVERY</sub>(BE) OR (95% CI) MAF Case/Cntrl</b>	2.27 × 10 <sup>-5</sup> 1.19 (1.10 – 1.29) 0.335/0.300	2.34 × 10 <sup>-4</sup> 0.85 (0.78 – 0.93) 0.264/0.292	9.16 × 10 <sup>-5</sup> 0.86 (0.79 – 0.93) 0.468/0.505	2.32 × 10 <sup>-8</sup> 1.24 (1.15 – 1.34) 0.492/0.442	6.84 × 10 <sup>-6</sup> 0.83 (0.77 – 0.90) 0.303/0.338
	<b>Replication</b>	<b>P<sub>REPLICATION</sub>(BE) OR (95% CI) MAF Case/Cntrl</b>	2.48 × 10 <sup>-2</sup> 1.14 (1.02 – 1.28) 0.336/0.305	6.60 × 10 <sup>-3</sup> 0.84 (0.74 – 0.95) 0.247/0.279	1.29 × 10 <sup>-1</sup> 0.92 (0.82 – 1.02) 0.461/0.481	1.22 × 10 <sup>-1</sup> 1.09 (0.98 – 1.21) 0.486/0.462	5.67 × 10 <sup>-2</sup> 0.89 (0.80 – 1.00) 0.312/0.338
	<b>Meta-analysis</b>	<b>P<sub>META</sub>(BE) OR (95% CI) MAF Case/Cntrl</b>	2.00 × 10 <sup>-6</sup> 1.18 (1.10 – 1.26) 0.335/0.303	5.08 × 10 <sup>-6</sup> 0.85 (0.79 – 0.91) 0.260/0.283	4.74 × 10 <sup>-5</sup> 0.88 (0.82 – 0.93) 0.467/0.489	5.54 × 10 <sup>-8</sup> 1.19 (1.12 – 1.26) 0.491/0.456	1.92 × 10 <sup>-6</sup> 0.85 (0.80 – 0.91) 0.306/0.338
<b>Esophageal Adeno-carcinoma</b>	<b>Discovery</b>	<b>P<sub>DISCOVERY</sub>(EA) OR (95% CI) MAF Case/Cntrl</b>	3.27 × 10 <sup>-6</sup> 1.26 (1.14 – 1.39) 0.343/0.300	1.96 × 10 <sup>-4</sup> 0.83 (0.75 – 0.91) 0.257/0.292	1.30 × 10 <sup>-4</sup> 0.84 (0.77 – 0.92) 0.463/0.504	1.20 × 10 <sup>-5</sup> 1.23 (1.12 – 1.34) 0.486/0.442	3.15 × 10 <sup>-3</sup> 0.86 (0.78 – 0.95) 0.316/0.338
	<b>Replication</b>	<b>P<sub>REPLICATION</sub>(EA) OR (95% CI) MAF Case/Cntrl</b>	2.31 × 10 <sup>-2</sup> 1.14 (1.02 – 1.27) 0.335/0.305	2.11 × 10 <sup>-4</sup> 0.80 (0.71 – 0.90) 0.238/0.279	4.43 × 10 <sup>-3</sup> 0.86 (0.77 – 0.95) 0.445/0.481	1.20 × 10 <sup>-2</sup> 1.14 (1.03 – 1.26) 0.497/0.462	7.59 × 10 <sup>-5</sup> 0.80 (0.71 – 0.89) 0.289/0.338
	<b>Meta-analysis</b>	<b>P<sub>META</sub>(EA) OR (95% CI) MAF Case/Cntrl</b>	5.76 × 10 <sup>-7</sup> 1.20 (1.12 – 1.29) 0.335/0.303	1.80 × 10 <sup>-7</sup> 0.81 (0.75 – 0.88) 0.254/0.283	2.03 × 10 <sup>-6</sup> 0.85 (0.79 – 0.91) 0.460/0.489	8.35 × 10 <sup>-7</sup> 1.19 (1.11 – 1.27) 0.494/0.456	1.46 × 10 <sup>-6</sup> 0.83 (0.77 – 0.90) 0.298/0.338
<b>Combined</b>	<b>Discovery</b>	<b>P<sub>DISCOVERY</sub>(BE+EA) OR (95% CI)</b>	3.01 × 10 <sup>-7</sup>	9.88 × 10 <sup>-6</sup>	3.47 × 10 <sup>-6</sup>	2.00 × 10 <sup>-9</sup>	5.52 × 10 <sup>-6</sup>

Cases/Cntrls		SNP	rs2687201	rs11789015	rs6479527	rs10419226	rs10423674
		Chr.	3	9	9	19	19
		Position	70928930	96716028	96858411	18803172	18817903
		CytoBand	3p13	9q22	9q22	19p13	19p13
		Nearest Gene	FOXPI	BARX1	PTPDC1	CRTC1	CRTC1
		Minor*/Major plus strand allele	T/G	G/A	T/C	A/C	T/G
		MAF Case/Cntrl	1.21 (1.12 – 1.30) 0.338/0.300	0.84 (0.78 – 0.91) 0.261/0.292	0.85 (0.80 – 0.91) 0.466/0.504	1.23 (1.15 – 1.32) 0.490/0.442	0.84 (0.79 – 0.91) 0.308/0.338
<b>Replication</b>	1,633/6,911	<b>P<sub>REPLICATION(BE+EA)</sub> OR (95% CI) MAF Case/Cntrl</b>	2.64 × 10 <sup>-3</sup> 1.14 (1.05 – 1.24) 0.336/0.305	2.25 × 10 <sup>-5</sup> 0.82 (0.75 – 0.90) 0.242/0.279	3.51 × 10 <sup>-3</sup> 0.89 (0.82 – 0.96) 0.452/0.481	7.18 × 10 <sup>-3</sup> 1.11 (1.03 – 1.20) 0.492/0.462	7.75 × 10 <sup>-5</sup> 0.84 (0.77 – 0.92) 0.300/0.338
<b>Meta-analysis</b>	5,561/10,118	<b>P<sub>META(BE+EA)</sub> OR (95% CI) MAF Case/Cntrl</b>	5.47 × 10 <sup>-9</sup> 1.18 (1.12 – 1.25) 0.335/0.303	1.02 × 10 <sup>-9</sup> 0.83 (0.79 – 0.88) 0.257/0.283	5.84 × 10 <sup>-8</sup> 0.87 (0.82 – 0.91) 0.464/0.489	3.55 × 10 <sup>-10</sup> 1.18 (1.12 – 1.24) 0.492/0.456	1.75 × 10 <sup>-9</sup> 0.84 (0.80 – 0.89) 0.302/0.338

\* Coded allele.

**Table 2**

SNPs in region 16q24 on chromosome 16 near the FOXP1 gene associated with Barrett's esophagus and esophageal adenocarcinoma. Shown are the discovery, replication and meta-analysis results for SNPs 100kb up or downstream of rs9936833, a previously identified susceptibility loci for Barrett's esophagus.<sup>14</sup> This table contains rs9936833, SNPs that had more significant P-values than rs9936833 in the combined Barrett's esophagus and esophageal adenocarcinoma data set ( $P_{DISCOVERY(BE+EA)}$ ), or were the most significant SNP in conditional analysis, in this region after fitting other SNPs (Table 3).

	Cases/Cntrls	SNP	rs1490865		rs3111601		rs9936833		rs1728400		rs3950627		rs2178146		rs13332095	
			Position	G/A	G/A	G/A	G/A	A/C	G/T	G/A	G/A	A/G				
Barrett's Esophagus	2,416/3,206	Discovery	$P_{DISCOVERY(BE)}$	$7.00 \times 10^{-2}$	$1.22 \times 10^{-4}$	$1.69 \times 10^{-4}$	$3.14 \times 10^{-6}$	$2.88 \times 10^{-6}$	$3.65 \times 10^{-4}$	$1.29 \times 10^{-2}$						
			OR (95% CI)	1.09 (0.99 – 1.19)	1.17 (1.08 – 1.27)	1.16 (1.07 – 1.26)	1.20 (1.11 – 1.30)	0.87 (0.80 – 0.94)	1.17 (1.03 – 1.33)							
			MAF Case/Cntrl	0.258/0.241	0.332/0.303	0.397/0.366	0.473/0.434	0.396/0.424	0.1105/0.093							
Barrett's Esophagus	759/6,911	Replication	$P_{REPLICATION(BE)}$	$5.01 \times 10^{-1}$	$1.04 (0.93 – 1.17)$				$3.67 \times 10^{-1}$							
			OR (95% CI)						0.95 (0.85 – 1.06)							
			MAF Case/Cntrl		0.324/0.313			0.396/0.408								
Barrett's Esophagus	3,175/10,117	Meta-analysis	$P_{META(BE)}$		$4.24 \times 10^{-4}$				$6.24 \times 10^{-4}$							
			OR (95% CI)		1.13 (1.05 – 1.20)				0.89 (0.84 – 0.95)							
			MAF Case/Cntrl		0.330/0.310			0.396/0.413								
Esophageal Adeno-carcinoma	1,516/3,209	Discovery	$P_{DISCOVERY(EA)}$	$8.60 \times 10^{-1}$	$5.64 \times 10^{-5}$	$2.06 \times 10^{-3}$	$4.94 \times 10^{-3}$	$4.70 \times 10^{-3}$	$1.31 \times 10^{-5}$	$1.42 \times 10^{-3}$						
			OR (95% CI)	0.99 (0.89 – 1.10)	1.22 (1.11 – 1.35)	1.16 (1.05 – 1.27)	1.14 (1.04 – 1.25)	1.14 (1.04 – 1.25)	0.81 (0.74 – 0.89)	1.28 (1.10 – 1.48)						
			MAF Case/Cntrl	0.243/0.241	0.343/0.303	0.397/0.366	0.460/0.434	0.494/0.465	0.380/0.424	0.112/0.093						
Esophageal Adeno-carcinoma	874/6,911	Replication	$P_{REPLICATION(EA)}$		$1.67 \times 10^{-1}$				$4.22 \times 10^{-2}$							
			OR (95% CI)		1.08 (0.97 – 1.20)				0.90 (0.81 – 1.00)							
			MAF Case/Cntrl		0.330/0.313			0.385/0.408								
Esophageal Adeno-carcinoma	2,390/10,120	Meta-analysis	$P_{META(EA)}$		$8.49 \times 10^{-5}$				$4.37 \times 10^{-6}$							
			OR (95% CI)		1.16 (1.08 – 1.24)				0.85 (0.79 – 0.91)							
			MAF Case/Cntrl		0.331/0.310			0.392/0.413								
Combined	3,928/3,207	Discovery	$P_{DISCOVERY(BE+EA)}$	$1.95 \times 10^{-1}$	$3.02 \times 10^{-6}$	$3.44 \times 10^{-5}$	$2.87 \times 10^{-6}$	$2.07 \times 10^{-6}$	$2.10 \times 10^{-6}$	$7.69 \times 10^{-4}$						
			OR (95% CI)	1.05 (0.97 – 1.14)	1.19 (1.11 – 1.28)	1.16 (1.08 – 1.24)	1.18 (1.10 – 1.26)	1.18 (1.10 – 1.26)	0.84 (0.79 – 0.91)	1.21 (1.08 – 1.36)						
			MAF Case/Cntrl	0.252/0.242	0.336/0.303	0.397/0.366	0.468/0.434	0.500/0.465	0.390/0.425	0.1108/0.093						
Combined	1,633/6,911	Replication	$P_{REPLICATION(BE+EA)}$		$2.24 \times 10^{-1}$				$5.04 \times 10^{-2}$							
			OR (95% CI)		$2.24 \times 10^{-1}$											
			MAF Case/Cntrl													

SNP		Cases/Cntrls	Minor*/Major plus strand allele	rs1490865	rs3111601	rs9936833	rs1728400	rs3950627	rs2178146	rs13332095
Position	86387275			86400081	86403118	86434446	86436343	86463695	86465590	
				G/A	G/A	G/A	A/C	G/T	G/A	A/G
			MAF Case/Cntrl		1.05 (0.97 - 1.15) 0.327/0.313				0.92 (0.85 - 1.00) 0.390/0.408	
			<b>P</b> <sub>META(BE+EA)</sub> <b>OR</b> (95% CI) MAF Case/Cntrl		1.56 × 10 <sup>-5</sup> 1.13 (1.07 - 1.19) 0.331/0.310				1.14 × 10 <sup>-6</sup> 0.88 (0.83 - 0.92) 0.394/0.413	
		Meta-analysis	5,561/10,118							

\* Coded allele.

**Table 3**

Step-wise conditional analysis to test for independent SNP signals in region 16q24. Each SNP is the most significant SNP 100kb up or downstream of rs9936833 after fitting all SNPs in the previous rows as additional covariates in the same logistic regression model used in the primary analysis. Starting with the most significant SNP in this region, rs3950627, this methodology fit four other SNPs (rows 2–5) before stopping when the p-value of the most significant remaining SNP had  $P > 0.01$ . The last row is the P-value of an association test of rs9936833 in each of these models.

rsID	Fitted SNP(s)			rs3950627		rs3950627 + rs2178146		rs3950627 + rs2178146 + rs3111601		rs3950627 + rs2178146 + rs3111601 + rs1490865	
	position	p-value	OR (CI)	p-value	OR (CI)	p-value	OR (CI)	p-value	OR (CI)	p-value	OR (CI)
rs3950627	86436343	0.0000021	1.18 (1.10 – 1.26)								
rs2178146	86463695	0.0000021	0.84 (0.79 – 0.91)	0.00119	0.88 (0.82 – 0.95)						
rs3111601	86400081	0.0000030	1.19 (1.11 – 1.28)	0.00225	1.13 (1.05 – 1.22)	0.00478	1.12 (1.04 – 1.21)				
rs1490865	86387275	0.1949656	1.05 (0.97 – 1.14)	0.02464	1.10 (1.01 – 1.19)	0.02582	1.10 (1.01 – 1.19)	0.00140	1.15 (1.05 – 1.25)		
rs13332095	86465590	0.0007695	1.21 (1.08 – 1.36)	0.00147	1.20 (1.07 – 1.34)	0.01658	1.15 (1.03 – 1.30)	0.00990	1.17 (1.04 – 1.31)	0.00490	1.18 (1.05 – 1.33)
rs9936833	86403118	0.0000344	1.16 (1.08 – 1.24)	0.01529	1.10 (1.02 – 1.19)	0.02265	1.09 (1.01 – 1.18)	0.84929	0.99 (0.86 – 1.14)	0.69573	1.03 (0.89 – 1.19)