Sex-Specific Genetic Associations for Barrett’s Esophagus and Esophageal Adenocarcinoma


1Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA.
2Institute for Genomic Statistics and Bioinformatics, Medical Faculty, University of Bonn, Germany.
3Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.
4Centre for Public Health, Queen’s University Belfast, Belfast, Northern Ireland.
5Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
6Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
7Department of Population Sciences, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, CA, USA.
8Institute of Human Genetics, University of Bonn, Bonn, Germany.
9Department of Epidemiology, MD Anderson Cancer Center, Houston, TX, USA.
10Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA.
11Pharmacogenomic Epidemiology, Ontario Cancer Institute, Toronto, Ontario, Canada M5G 2M9.
12Cancer Research UK, Cambridge Institute, Cambridge, UK.
13Department of Oncology, University of Cambridge, Cambridge, UK.

14Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

15Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA.

16Department of Medicine II, Sana Klinikum, Offenbach, Germany.

17Department of Internal Medicine II, Evangelisches Krankenhaus, Düsseldorf, Germany.

18Department of Gastroenterology and Interdisciplinary Endoscopy, Vivantes Wenckebach-Klinikum, Berlin, Germany.

19Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University Hospital, Magdeburg, Germany.

20Department of General, Visceral and Transplantation Surgery, University of Heidelberg, Heidelberg, Germany.

21Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, University of Hamburg, Hamburg, Germany.

22Department of General, Visceral and Cancer Surgery, University of Cologne, Cologne, Germany.

23Department of Internal Medicine and Gastroenterology, Elisabeth Hospital, Essen, Germany.

24Department of Interdisciplinary Endoscopy, University Hospital Hamburg-Eppendorf, Hamburg, Germany.

25Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany.

26Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA.

27Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

28Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden.

29School of Cancer and Pharmaceutical Sciences, King’s College London, London, UK.

30Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, Chapel Hill, NC, USA.

31Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA.

32San Francisco Medical Center, Kaiser Permanente Northern California, San Francisco, CA, USA.

33Department of Visceral, Transplant, Thoracic and Vascular Surgery, University Hospital of Leipzig, Leipzig, Germany.
Medical Research Council (MRC) Cancer Unit, Hutchison-MRC Research Centre, University of Cambridge, Cambridge, UK.

Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.

Center for Human Genetics, University Hospital of Marburg, Marburg, Germany.

Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA.

Grant support: This work was primarily funded by National Institutes of Health (R01CA136725). The funders of the study had no role in the design, analysis, or interpretation of the data, nor in writing or publication decisions related to this article. JD and QTO were supported by a Research Training Grant from the Cancer Prevention and Research Institute of Texas (CPRIT; RP160097). PG was supported by a grant from National Health and Medical Research Council of Australia (1123248). GL was supported by the Alan B. Brown Chair in Molecular Genomics and by the CCO Chair in Experimental Therapeutics and Population Studies. The University of Cambridge received salary support for PDP from the NHS in the East of England through the Clinical Academic Reserve. BJR was supported by a grant (P01CA91955) from the National Institutes of Health /National Cancer Institute. NJS was supported by a grant (P30 DK034987) from the National Institutes of Health. TLV was supported by National Institutes of Health Established Investigator Award K05CA124911. SM was supported by an Australian Research Council Future Fellowship. DCW was supported by a Research Fellowship from the National Health and Medical Research Council of Australia (APP1058522). MBC was supported by the Intramural Research Program of the United States National Cancer Institute, NIH, DHHS. DAC was supported by the United States National Institutes of Health grants R03 KD 58294, R21DK077742, and RO1 DK63616 and National Cancer Institute grant R01CA136725. CM was supported by the BONFOR-program of the Medical Faculty, University of Bonn (O-147.0002).
**Abbreviations:** BE, Barrett’s esophagus; BEACON, Barrett’s and Esophageal Adenocarcinoma Consortium; CI, confidence interval; EA, esophageal adenocarcinoma; GERD, gastroesophageal reflux disease; GWAS, genome-wide association study; OR, odds ratio.

**Correspondence to:** Aaron P. Thrift, Ph.D., Baylor College of Medicine, One Baylor Plaza, MS: BCM307, Room 621D, Houston, Texas 77030-3498. Tel: 713-798-9107; Fax: 713-798-3658; E-mail: aaron.thrift@bcm.edu

**Conflicts of interest:** The authors disclose no potential conflicts of interest.

**Specific author contributions:** APT had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: JD, CM, DAC, MBC, DCW, TLV, APT. Statistical analysis and interpretation of data: JD, CM, ST, QTO, OB, JS, IG, MK, PK, AB, JS, APT. Acquisition of data and obtained funding: LAA, AHW, WY, NCB, LB, WHC, MDG, GL, CC, PDP, HAR, AM, CG, MA, MV, TS, JI, AH, BS, YV, WP, HN, TR, PGI, BJR, JL, NJS, DAC, RCF, DCW, TLV, JS. Data preparation: JD, CM, ST, PG. Drafting of the manuscript: JD, APT. Critical review of the manuscript for important intellectual content: JD, CM, ST, QTO, PG, LAA, AHW, WY, LB, OB, JS, IG, WHC, MDG, GL, CC, PDP, HAR, AM, CG, MA, MV, TS, JI, AH, BS, YV, WP, HN, TR, MK, PK, AB, PGI, BJR, SM, JL, NJS, DAC, RCF, MBC, DCW, TLV, JS, APT. Study supervision: APT. All authors read and approved the final version for submission.

**Word count:** 5,866 with reference
ABSTRACT

Background and Aims: Esophageal adenocarcinoma (EA) is characterized by a strong and yet unexplained male predominance (with a male-to-female ratio in incidence of up to 8:1). Genome-wide association studies (GWAS) have identified more than 20 susceptibility loci for EA and its premalignant lesion, Barrett’s esophagus (BE). However, the sex differences in genetic associations with BE/EA remain largely unknown.

Methods: Given strong genetic overlap, BE and EA cases were combined into a single case group for analysis and were compared with population-based controls. We performed a sex-specific GWAS of BE/EA in three separate studies and then meta-analyzed using a fixed-effects inverse variance-weighting approach (6,758 male cases, 7,489 male controls, 1,670 female cases and 6,174 female controls). A series of downstream analyses were conducted separately in males and female to identify genes associated with BE/EA and the genetic correlations between BE/EA and other traits.

Results: Meta-analysis of sex-specific GWAS identified three novel independent genome-wide significant loci for BE/EA, including one variant at chromosome 12p12.3 (rs35827298, MGST1-LMO3, \( P = 1.28 \times 10^{-8} \)) detected in males only, and two variants at chromosome 8p (rs13259457, PRSS55-RP1L1, \( P = 6.65 \times 10^{-9} \), and rs17321041, DPYSL2, \( P = 4.98 \times 10^{-8} \)) detected in females only. We also observed strong genetic correlations of BE/EA with reflux disease in males and obesity in females.

Conclusions: The identified novel sex-specific variants associated with BE/EA could improve our understanding of the genetic architecture of the disease and the reasons for the male predominance.

Keywords: Esophageal adenocarcinoma; Barrett’s esophagus; sex difference; genome-wide association study; interaction
INTRODUCTION

Esophageal adenocarcinoma (EA) is a highly fatal cancer with increasing incidence in many Western populations during the last four decades.\textsuperscript{1, 2} EA is characterized by a strong male predominance, with a male-to-female ratio in incidence of up to 8:1 in the United States.\textsuperscript{3-5} While much attention has been given to the striking increase in EA incidence among white males (10-fold increase since 1973 in the U.S.),\textsuperscript{6, 7} EA incidence has also increased among white females such that the sex ratio has remained relatively stable over time.\textsuperscript{2} Males are also more likely than females to develop Barrett’s esophagus (BE), the only known premalignant lesion for EA.\textsuperscript{8}

Reasons for the male predominance in BE and EA remain poorly understood, but likely reflects differing prevalence of environmental risk factors, sex hormonal and reproductive factors and differing effects of underlying susceptibility alleles between males and females.\textsuperscript{9} Gastroesophageal reflux disease (GERD),\textsuperscript{10, 11} obesity,\textsuperscript{12} and smoking\textsuperscript{13, 14} are established risk factors for BE and EA. These factors, however, explain very little of the sex differences in BE and EA incidence, as their prevalence and effect-size of associations with the risk of BE and EA are not dissimilar between sexes.\textsuperscript{13, 15-20} Previous studies have made efforts to evaluate the influence of sex hormonal exposures and reproductive factors on the risk of EA.\textsuperscript{21, 22} However, these studies were limited by small numbers of EA cases in women, inherent limitations of observational studies, and thus results to date have been mixed.\textsuperscript{3}

Findings from large genomic studies suggest that sex-specific genetic architecture could contribute to sex differences in human disease.\textsuperscript{23, 24} Although genome-wide association studies (GWAS) have identified over 20 susceptibility loci for BE and/or EA,\textsuperscript{25-28} little is known about whether genetic effects differ between males and females. We therefore conducted a sex-specific genome-wide meta-analysis to identify sex-specific susceptibility loci for BE and EA.
METHODS

Study population

We used data from EA cases, BE cases and controls from studies in Europe, North America, and Australia, the details of which we have described in full elsewhere.\textsuperscript{25, 29} Namely, genome-wide genotype data were obtained from three GWAS: the international Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON; \url{http://beacon.tlvnet.net/}), as well as studies from Cambridge (United Kingdom) and Bonn (Germany).\textsuperscript{25, 27, 28, 30} The BEACON GWAS included 1,508 EA cases, 2,406 BE cases, and 2,177 controls from 15 epidemiologic studies conducted in North America, Western Europe, and Australia.\textsuperscript{27} An additional 4,541 controls from dbGaP (phs000187.v1.p1, phs000196.v2.p1 and phs000524.v1.p1) were merged with the BEACON GWAS to increase the statistical power.\textsuperscript{25} The Cambridge studies included 873 BE cases from the UK Barrett's Esophagus Gene Study, 995 EA cases from the UK Stomach and Oesophageal Cancer Study, and 3,408 controls from the WTCCC2 including the National Blood Service (UKBS) and 1958 birth (58C) studies.\textsuperscript{25} The Bonn GWAS included 1,609 EA cases, 1,037 BE cases, and 3,537 controls.\textsuperscript{25, 30} All participants were of European ancestry. Patients with BE were identified by histopathological diagnosis of intestinal metaplasia, and patients with EA had a histopathological diagnosis of adenocarcinoma. Each contributing study complied with their institutional review board requirements and all participants gave informed consent.

Genotyping and imputation

Genotyping of buffy coat or whole blood DNA from all participants was performed on high-density SNP arrays (Illumina, San Diego, CA, USA), in accordance with standard quality control (QC) procedures for each participating study.\textsuperscript{25, 31} All genotyped samples and variants met the following inclusion criteria: per variant and per sample missingness ≤ 3%; SNPs with a minor allele frequency (MAF) > 1%; SNPs with $P \geq 0.0001$ in controls and $P \geq 5 \times 10^{-10}$ in BE and EA cases for Hardy-Weinberg equilibrium, no familial relationships, extreme heterozygosity rate, or outliers. Sex was confirmed genetically for all subjects as part of QC. For the imputation of autosomal chromosomes, we used SHAPEIT version 2.12\textsuperscript{32, 33} for
phasing of the genotyped SNPs and IMPUTE2 version 2.3.1\textsuperscript{34} for imputation of missing SNPs, using the 1000 Genomes Phase 3 haplotypes (October, 2014 release) as a reference panel.\textsuperscript{25} An additional flag of -chrX was added when running imputation for the X chromosome. Post-imputation QC was conducted in each study by excluding SNPs with imputation quality score < 0.4 or minor allele count (MAC, 2*MAF*sample size) < 20.\textsuperscript{35} We included in the analysis only those SNPs that passed the post-imputation QC in all three studies (i.e., BEACON, Cambridge and Bonn).

**Genome-wide meta-analysis**

Given the substantial genetic overlap between BE and EA ($r_g$=1.0),\textsuperscript{25,36} we performed genetic association tests comparing a combined BE and EA case group (‘BE/EA’) with controls to maximize statistical power. Sex-stratified logistic regression analyses were conducted separately in the BEACON, Cambridge and Bonn datasets using SNPTTEST version 2.5.4-beta3,\textsuperscript{37,38} adjusted for study-specific top four principal components for autosomal chromosomes. Variants were defined as allele dosages based on an additive model.\textsuperscript{25} The X chromosome was analyzed separately using ‘newml’ in SNPTTEST.\textsuperscript{37,38} We did the sex-specific meta-analysis with the fixed-effects inverse variance–weighting approach in METAL version 2011-03-25, and considered a standard genome-wide significance threshold of $5 \times 10^{-8}$.\textsuperscript{39} To evaluate the independence of associated SNPs with $P_{meta} < 5 \times 10^{-8}$ in sex-specific analysis, we performed a stepwise selection procedure implemented in GCTA-COJO (GCTA version 1.93.0beta).\textsuperscript{40} COJO enables conditional and joint association analysis using the sex-specific meta-analysis summary statistics. We used the BEACON dataset comprising 10,632 participants as the reference data to estimate linkage disequilibrium (LD) patterns. Only SNPs with $P < 5 \times 10^{-8}$ in both the sex-specific meta-analysis and the conditional analysis were deemed statistically significant and reported.

In addition to the analyses stratified by sex, we examined for SNP-by-sex interactions. We first assessed study-specific interactions using SNPTTEST version 2.5.4-beta3 with method ‘newml’, adjusted for study-specific top four principal components,\textsuperscript{37,38} then pooled the study-specific estimates using a joint 2
degree-of-freedom (2df) meta-analysis (JMA) of the main SNP effect and the SNP-by-sex interaction effect. The JMA accounts for the covariance between the SNP regression estimate and the SNP-by-sex regression estimate and has greater statistical power than other methods to detect interaction effects. This approach can identify SNPs either having a significant main effect or an interaction effect, or only showing significant associations when the interaction effect is considered.

Genomic control correction was applied on all meta-analyses to account for population stratification or unaccounted for relatedness. We used quantile-quantile (Q-Q) plots and the genomic inflation factor ($\lambda$) to detect potential population stratification in each study as well as in the meta-analyses. We used R 3.6.1 to generate Q-Q and Manhattan plots and LocusZoom version 1.4 to generate regional plots.

**Functional annotation**

To explore the biological relevance of identified loci, we performed functional annotation using FUMA v1.3.5e. All SNPs in LD (at $r^2 \geq 0.6$) with the significant SNPs, with $P_{\text{meta}} < 0.05$ in the sex-specific meta-analysis, and with MAF > 0.01 were selected for annotation, which were obtained from ANNOVAR categories, combined annotation-dependent depletion (CADD) scores and RegulomeDB. CADD score is the score of deleteriousness of SNP predicted by 63 functional annotations with the score $> 12.37$ as the threshold to be deleterious and the score $> 20$ indicating the variant is ranked in the top 1% of highest scoring variants. RegulomeDB score is a categorical score (ranging from 1a to 7) based on expression quantitative trait loci (eQTLs) and chromatin marks. The lower of the RegulomeDB score, the more likely it is that the SNP has a regulatory function. We also investigated the biological impact of the susceptibility loci using HaploReg v4.1 and Genotype-Tissue Expression (GTEx) v8 for gene-expression and regulatory data derived from ENCODE, Roadmap projects and other resources.

**Gene-based and pathway-enrichment analyses**
Gene-based analysis was conducted in MAGMA (v1.07) using the per-SNP summary statistics obtained from the sex-specific meta-analysis. SNPs were assigned to genes based on their position according to the NCBI 37.3 build with no additional boundary placed around the genes. LD between SNPs was estimated using reference data from the 1000 Genomes Project Phase 3 European ancestry samples. It resulted in 18,856 protein-coding genes (each containing at least one SNP from the meta-analysis) being tested for males, and 18,863 genes for females. We used a Bonferroni-corrected p-value to account for multiple comparisons ($P_{\text{gene}} < 2.65 \times 10^{-6}$ [0.05/18,863]). The derived gene-based $P$ values were further used for gene-set analysis implemented in MAGMA. Competitive $P$ value for a specific gene set was computed for 12,165 predefined gene sets curated from MsigDB (version 7.0), considering known biological and metabolic pathways from Gene Ontology (9,996 gene sets), BIOCARTA (289 gene sets), KEGG (186 gene sets), PID (196 gene sets) and REACTOME (1,499 gene sets). The Bonferroni-corrected significant threshold for gene-set analysis was set to $0.05/12,165 = 4.11 \times 10^{-6}$.

**Transcriptome-wide association analysis**

We applied metaXcan methods to integrate the summary statistics from the sex-specific meta-analysis with eQTLs information to map genes whose predicted expression levels in esophageal tissues were associated with disease risk. SNP weights and their respective covariance for three relevant esophageal tissues (esophagus gastroesophageal junction, esophagus mucosa and esophagus muscularis) were obtained from PredictDB Data Repository (http://predictdb.org/) based on GTEx v8 database. The total number of genes tested for these tissues were 6,241, 8,448 and 8,160, respectively (total across all tissues, 22,849). Genes with $P < 2.19 \times 10^{-6}$ (0.05/22,849) were considered to have gene expression profiles statistically significantly associated with BE/EA.

**Cross-trait genetic correlations**

LD score regression was applied to estimate genetic correlations ($r_g$) between BE/EA and a range of diseases, disorders and human traits (767 total) based on GWAS summary statistics obtained from
publicly available databases in LD-Hub (http://ldsc.broadinstitute.org/). Calculations were conducted separately for males and females. The Bonferroni corrected significant threshold was $P < 6.52 \times 10^{-5} (0.05/767)$. 
RESULTS

Sex-specific susceptibility loci for BE/EA

The sex-specific meta-analysis included 6,758 male cases (i.e., BE and EA cases combined), 7,489 male controls, 1,670 female cases and 6,174 female controls (Supplementary Table 1). In total, 9,353,006 SNPs for males and 9,350,284 SNPs for females were tested. The results of the sex-specific meta-analysis are summarized in Figure 1. As evidenced by the Q-Q plots (Supplementary Figure 1), there was no evidence for hidden substructure or cryptic relatedness in the sex-specific analysis (λ ranged from 1.01 to 1.08).

Table 1 presents the independent loci meeting genome-wide significance in the conditional analysis ($P_{cojo} < 5 \times 10^{-8}$). In males, we identified three loci independently associated with BE/EA, of which rs35827298 on chromosome 12p12.3 within MGST1 and LMO3 ($P_{cojo} = 1.28 \times 10^{-8}$, $P$ for heterogeneity=0.89, Figure 2A) was not previously reported in the sex-combined GWAS.\textsuperscript{25-28} In females, we identified two novel loci independently associated with BE/EA: rs13259457 on chromosome 8p23.1 near PRSS55 and RP11 (P\textsubscript{cojo} = 6.65 \times 10^{-9}, $P$ for heterogeneity=0.43, Figure 2B), and rs17321041 on chromosome 8p21.2 within DPYSL2 ($P_{cojo} = 4.98 \times 10^{-8}$, $P$ for heterogeneity=0.62, Figure 2C). None of these loci were associated with BE/EA risk in the other sex group ($P > 0.05$) (Table 1). In the interaction analysis, both female-specific loci exhibited a strong SNP-by-sex interaction effect ($P_{G \times S} = 4.25 \times 10^{-5}$ and 4.91\times10^{-6} for rs13259457 and rs17321041, respectively) but a non-significant SNP main effect ($P_{SNP} = 0.22$ and 0.71 for rs13259457 and rs17321041, respectively), though only rs13259457 was statistically significant in the 2df JMA of main and interaction effects ($P_{JMA}=1.87 \times 10^{-8}$ and 3.15\times10^{-7} for rs13259457 and rs17321041, respectively) (Table 2). While carrying a rs13259457 G allele was associated with 54% increased risk of BE/EA in females, there was no SNP effect in males ($P$ for heterogeneity = 0.003, Supplementary Table 2). Similar results were observed for rs17321041 ($P$ for heterogeneity < 0.001, Supplementary Table 2). The associations for male-specific loci in the 2df JMA all exhibited significant main effects, which were consistent with previous reports.\textsuperscript{25-28} Besides rs13259457 and rs35827298, we identified six other
statistically significantly associated loci in the 2df JMA of main and interaction effects (Supplementary Table 3), all of which have been previously reported in the sex-combined GWAS.\textsuperscript{25-28}

**Functional annotation**

In FUMA, most of the SNPs in the susceptibility loci were intronic or intergenic (Supplementary Table 4). Twenty-two SNPs in males were predicted to be potential deleterious variants (CADD score \(> 12.37\)), with the highest probability of a deleterious protein effect observed for rs6938505 (CADD score = 21.2), a proxy of the SNP rs12660153 (\(P = 4.42 \times 10^{-8}\), \(r^2 = 0.68\)) in males. Using HaploReg v4.1 database, the three newly identified SNPs all mapped to a regulatory region, showing evidence of transcription factor binding sites and effects on regions marked by histone modifications within promoter and/or enhancer and by DNase hypersensitivity (Supplementary Table 5). From the eQTL analyses, we found that rs35827298, rs13259457 and 17321041 regulated the expression of several genes in various tissues, including the genes previously reported to be associated with the risk of BE and EA, such as \textit{MGST1}\textsuperscript{55} (rs35827298, \(P = 4.5\times10^{-5}\)) (Supplementary Table 5). Specifically, rs17321041 was \textit{cis}-eQTL of \textit{SDAD1P1} and \textit{PNMA2} in the three esophageal relevant tissues (esophagus gastroesophageal junction, esophagus mucosa and esophagus muscularis) with \(P\) ranging from \(6.5\times10^{-5}\) to \(7.9\times10^{-9}\).

**Implicated genes and pathways**

Gene-based analysis in MAGMA identified seven genes in males (\textit{TPPP, MGST1, ALDH1A2, KLHL26, SLC22A3, ISL1 and CRTCl}) and three in females (\textit{TENM4, BLK} and \textit{MSRA}) associated with BE/EA (Supplementary Table 6). Among them, \textit{MGST1} (\(P_{\text{gene}} = 3.79 \times 10^{-7}\)), \textit{KLHL26} (\(P_{\text{gene}} = 1.39 \times 10^{-6}\)), \textit{SLC22A3} (\(P_{\text{gene}} = 1.89 \times 10^{-6}\)), \textit{ISL1} (\(P_{\text{gene}} = 1.84 \times 10^{-6}\)) and \textit{TENM4} (\(P_{\text{gene}} = 5.36 \times 10^{-7}\)) were not previously reported in the sex-combined GWAS.\textsuperscript{25-28} While one gene (\textit{SLC9A3}) that was not identified in MAGMA gene-based analysis was significantly associated with BE/EA in esophageal tissues in males (\(P_{\text{gene}} = 1.19\times10^{-6}\)) in metaXcan, there were no additional genes identified for females.
In the MAGMA gene-set analysis, no pathways were significantly associated with BE/EA risk after Bonferroni correction ($P < 4.11 \times 10^{-6}$); however, the most significant pathways were different between males and females (top 10 listed in Supplementary Table 7). In males, the most significant pathways were observed for homologous recombination ($P = 2.08 \times 10^{-5}$) and response to vitamin A ($P = 5.21 \times 10^{-5}$). In females, regulation of coagulation exhibited the strongest association ($P = 3.27 \times 10^{-5}$).

**Cross-trait associations**

Using LD score regression, we observed significant $r_g$ of BE/EA with GERD ($r_g = 0.701$, $P = 1.43 \times 10^{-7}$), hiatus hernia ($r_g = 0.748$, $P = 1.96 \times 10^{-6}$), and taking medication for heartburn ($r_g = 0.473$, $P = 2.19 \times 10^{-5}$) in males. In females, the most highly correlated trait was fat percentage, though fat percentage and other traits were not significantly associated with BE/EA in females after Bonferroni correction ($P < 6.52 \times 10^{-5}$). Supplementary Table 8 lists those traits that had genetic correlations with BE/EA at a false discovery rate $< 0.05$. 


DISCUSSION

EA is characterized by a striking male predominance. While the reasons for this male predominance are not yet fully understood, it appears unlikely to be explained by differences between males and females in the prevalence of established risk factors. This study was the first attempt to scan sex-specific genetic associations with the risk of EA and its precursor lesion, BE, with the aim of assessing whether sex-specific genetic architecture could contribute to sex differences in BE/EA. We identified two female-specific independent loci located in PRSS55-RP1L1 (rs13259457 at 8p23.1) and DPYSL2 (rs17321041 at 8p21.2), and three male-specific independent loci located in KHDRBS2-MTRNR2L9 (rs112894788 at 6q11.1), MGST1-LMO (rs35827298 at 12p12.3) and CRTC1 (rs2003476 at 19p13.11) associated with risk of BE/EA. Three of these loci (8p23.1, 8p21.2 and 12p12.3) were not identified in previous BE/EA GWASs, which pooled together males and females.25-28

SNP rs13259457 lies ~23kb downstream of PRSS55 (serine protease 55) and ~28kb downstream of RP1L1 (retinitis pigmentosa-1-like 1). PRSS55, also known as TSP1 (thrombospondin 1), encodes a secreted protein that possesses important biological functions in tumorigenesis and metastasis.56-58 Secreted TSP1 from esophageal cancer cells is mediated by Rab37 to inhibit p-FAK/p-paxillin/p-ERK migration signaling and promotes the neovasculature of tumor microenvironment and tumor progression.59 Studies have shown that TSP1 was upregulated in the metaplasia-dysplasia-EA sequence in 150 micro-dissected esophageal stroma.60 Furthermore, an expression analysis in a cohort of BE patients showed that increased expression of TSP1 was associated with almost 4-fold higher risk of progression to EA and poorer survival outcomes in patients with EA.60 The associations between RP1L1 and cancer are largely unknown. Limited studies have reported RP1L1 mutations in gastric cancer,61 and linked RP1L1 mutations with dopamine-agonist resistance in prolactinoma.62 Of interest, rs13259457 exhibited strong interactions with sex, with the effects of risk allele G on BE and EA significantly different by sex. The biological mechanism needs further investigation.
SNP rs17321041 is an intronic variant of DPYSL2 (dihydropyrimidinase like 2), also known as CRMP2 (collapsin response mediator protein-2), a gene that encodes a multifunctional adaptor protein that promotes microtubule assembly. CRMP2 expressed in tumor tissues and its phosphorylation induced by CDK5 has been observed in the nuclei of tumor cells. CRMP2 interacts with neurofibromatosis type 1 (Nf1), a tumor suppressor gene that produces neurofibromin functioning as a negative regulator of Ras. Expression and phosphorylation of CRMP2 and CRMP2-neurofibromin interaction have been further linked to cancer proliferation and progression, including breast, glioma, lung, and lymphoma. In addition, rs17321041 was recently reported to be associated with diastolic blood pressure by GWAS, suggesting potential shared genetic predisposition between BE/EA and hypertension as they have common risk factors (i.e., obesity). eQTL analysis showed that rs17321041 is cis-eQTL of PNMA2 (paraneoplastic antigen Ma2) in esophageal tissues. The functions of PNMA2 are not clear. Limited evidence have indicated its role in tumorigenesis, and reported PNMA2 expression could be a biomarker for gastrointestinal neuroendocrine carcinomas and small intestine neuroendocrine tumors.

SNP rs35827298 is located ~50kb downstream from MGST1 (microsomal glutathione S-transferase 1) and 133kb downstream from LMO3 (LIM domain only 3), and acts as a cis-eQTL of MGST1. MGST1 is a member of the superfamily of membrane-associated proteins in eicosanoid and glutathione metabolism (MAPEG). MGST1 encodes a protein localized to the endoplasmic reticulum and outer mitochondrial membrane that is responsible for detoxifying electrophilic xenobiotics and neutralizing oxidative stress. Previous studies have shown that MGST1 is upregulated in various cancer types, such as lung, prostate, brain and colorectal, and linked its expression to tumorigenesis and apoptosis. Recently, Buas and colleagues investigated genetic variation in inflammation-related pathways and found that MGST1 variants were associated with increased risk of BE and EA in European populations. LMO members are important regulators in cell fate determination and differentiation. The oncogenic role of LMO3 was first described in neuroblastoma, where LMO3 deregulated expression was associated with genesis and progression of neuroblastoma. Since then, increasing evidence suggests that LMO3 could promote...
cancer cell proliferation, invasion and metastasis in hepatocellular carcinoma, gastric cancer, clear cell renal cell carcinoma, and lung cancer. These oncogenic roles may be due to the interactions between LMO3 and p53, HEN2, LATS1 or microRNAs.

Gene-based and enrichment analysis implicated additional genes associated with BE/EA, including SLC22A3, SLC9A3 and ISL1. Both SLC22A3 (solute carrier family 22 member 3) and SLC9A3 (solute carrier family 9 member A3) belong to the solute carrier superfamily and were recently reported to be candidate genes of BE and EA in a study investigating eQTLs in esophageal tissues. In addition, downregulation of SLC22A3 (solute carrier family 22 member 3) could drive early tumor invasion and metastasis in familial esophageal cancer. ISL1 (LIM homeobox 1) encodes LIM-homeodomain transcription factor and promotes tumor progression in multiple cancer types, such as gastric and breast.

A recent GWAS of smoking status also implicated ISL1, suggesting ISL1 may be a common susceptibility gene between BE/EA and their risk factor, smoking.

Strengths of the current study include the advantage of large worldwide consortium of BE and EA with high-quality case-control and cohort parent studies, providing us with a rare opportunity to perform the first-ever genome-wide sex-specific analysis for BE and EA. Our comprehensive downstream analysis revealed the shared genetic predisposition between BE/EA and their established risk factors (i.e., smoking, obesity). Our evaluation of genetic correlations between BE/EA and other phenotypes highlighted overlap with GERD in males and obesity in females. A recent study that investigated shared genetic components of obesity-related traits and BE/EA also revealed strong genetic correlation for BMI only in females. These findings may provide valuable additional clues regarding male predominance of BE and EA, and provide sex-specific early detection and treatment strategy. Our study has several limitations. First, our sample size may not provide sufficient power for genome-wide interaction analysis. Therefore, we applied joint meta-analysis of SNP main and SNP-by-sex interaction. This method could provide optimal statistical power. Second, we do not have data of sex hormone exposure. It is possible that genetic
variants interacted with sex hormonal factors to confer the risk of BE/EA differently between sex groups. Third, although we conducted extensive functional prediction for the identified variants, validation of these biological functions in cell lines and animal models are lacking. Further experimental studies are needed to reveal the underlying biological mechanisms. In addition, we focused on common genetic variation (MAF>1%) represented on genotyping platforms which often exhibit moderate genetic effects. Further large-scale studies based on whole-exome or whole-genome sequencing would be required to identify additional sex-specific associations with rare variants.

In conclusion, our meta-analysis identified three novel loci associated with the risk of BE/EA in a sex-dependent manner. This finding warrants further mechanistic investigations to how these variants influence the sex differences in BE/EA, and their interaction with other factors that differ by sex (e.g., sexual hormone and immune responses). The identified sex-specific genetic determinants of BE and EA may further guide risk prediction for these diseases specifically targeted for males and females.
REFERENCES

38. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.


Acknowledgements

The following UK hospitals participated in sample collection through the Stomach and Oesophageal Cancer Study (SOCS) collaboration network: Addenbrooke's Hospital, University College London, Bedford Hospital, Hinchinbrooke Hospital, Peterborough City Hospital, West Suffolk Hospital, Norfolk and Norwich University Hospital, Churchill Hospital, John Radcliffe Hospital, Velindre Hospital, St Bartholomew's Hospital, Queen's Hospital Burton, Queen Elisabeth Hospital, Diana Princess of Wales, Scunthorpe General Hospital, Royal Devon & Exeter Hospital, New Cross Hospital, Belfast City Hospital, Good Hope Hospital, Heartlands Hospital, South Tyneside District General Hospital, Cumberland Infirmary, West Cumberland Hospital, Withybush General Hospital, Stoke Mandeville Hospital, Wycombe General Hospital, Wexham Park Hospital, Southend Hospital, Guy's Hospital, Southampton General Hospital, Bronglais General Hospital, Aberdeen Royal Infirmary, Manor Hospital, Clatterbridge Centre for Oncology, Lincoln County Hospital, Pilgrim Hospital, Grantham & District Hospital, St Mary's Hospital London, Croydon University Hospital, Whipps Cross University Hospital, Wansbeck General Hospital, Hillingdon Hospital, Milton Keynes General Hospital, Royal Gwent Hospital, Tameside General Hospital, Castle Hill Hospital, St Richard's Hospital, Ipswich Hospital, St Helens Hospital, Whiston Hospital, Countess of Chester Hospital, St Mary's Hospital IOW, Queen Alexandra Hospital, Glan Clwyd Hospital, Wrexham Maelor Hospital, Darent Valley Hospital, Royal Derby Hospital, Derbyshire Royal Infirmary, Scarborough General Hospital, Kettering General Hospital, Kidderminster General Hospital, Royal Lancaster Infirmary, Furness General Hospital, Westmorland General Hospital, James Cook University Hospital, Friarage Hospital, Stepping Hill Hospital, St George's Hospital London, Doncaster Royal Infirmary, Maidstone Hospital, Tunbridge Hospital, Prince Charles Hospital, Hartlepool Hospital, University Hospital of North Tees, Ysbyty Gwynedd, St. Jame's University Hospital, Leeds General Infirmary, North Hampshire Hospital, Royal Preston Hospital, Chorley and District General, Airedale General Hospital, Huddersfield Royal Infirmary, Calderdale Royal Hospital, Torbay District General Hospital, Leighton Hospital, Royal Albert Edward Infirmary, Royal Surrey County Hospital, Bradford Royal Infirmary, Burnley General Hospital, Royal Blackburn Hospital, Royal
Sussex County Hospital, Freeman Hospital, Royal Victoria Infirmary, Victoria Hospital Blackpool, Weston Park Hospital, Royal Hampshire County Hospital, Conquest Hospital, Royal Bournemouth General Hospital, Mount Vernon Hospital, Lister Hospital, William Harvey Hospital, Kent and Canterbury Hospital, Great Western Hospital, Dumfries and Galloway Royal Infirmary, Poole General Hospital, St Hellier Hospital, North Devon District Hospital, Salisbury District Hospital, Weston General Hospital, University Hospital Coventry, Warwick Hospital, George Eliot Hospital, Alexandra Hospital, Nottingham University Hospital, Royal Chesterfield Hospital, Yeovil District Hospital, Darlington Memorial Hospital, University Hospital of North Durham, Bishop Auckland General Hospital, Musgrove Park Hospital, Rochdale Infirmary, North Manchester General, Altnagelvin Area Hospital, Dorset County Hospital, James Paget Hospital, Derriford Hospital, Newham General Hospital, Ealing Hospital, Pinderfields General Hospital, Clayton Hospital, Dewsbury & District Hospital, Pontefract General Infirmary, Worthing Hospital, Macclesfield Hospital, University Hospital of North Staffordshire, Salford Royal Hospital, Royal Shrewsbury Hospital, Manchester Royal Infirmary.