Polymorphisms Near *TBX5* and *GDF7* Are Associated With Increased Risk for Barrett’s Esophagus


1Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2Plymouth University Peninsular School of Medicine and Dentistry, Plymouth, Devon, UK; 3Centre of Biostatistics, Bioinformatics and Biomarkers, Plymouth University Peninsular Schools of Medicine and Dentistry, Plymouth, Devon, UK; 4Department of Gastroenterology and Hepatology, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands; 5Department of Digestive Oncology, University Hospital Gasthuisberg, Leuven, Belgium; 6Department of Pathology, Leicester Royal Infirmary, Leicester, UK; 7Department of Cellular Pathology, Warwick Hospital, Warwick, UK; 8Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, British Columbia, Canada; 9Department of Gastroenterology, Wansbeck General Hospital, Ashington, Northumberland, UK; 10William Harvey Research Institute, The Ernest Cook Vascular & Microvascular Unit, Centre for Clinical Pharmacology, St Bartholomew’s Hospital, London, UK; 11Centre for Statistics in Medicine and Oxford Clinical Trials Research Unit, Oxford, UK; 12Centre for Digestive Diseases, Queen Mary University of London, London, UK; 13Department of Gastroenterology, Leicester General Hospital, Leicester, UK; 14Department of Oncology, New Cross Hospital, Royal Wolverhampton Hospitals NHS Trust, Wolverhampton, UK; 15Department for Gastroenterology, Leicester Royal Infirmary, Leicester, UK; 16Gloucestershire Royal Hospital, Great Western Road, Gloucester, UK; 17Department of Upper GI Surgery, Gloucestershire Royal Hospital, Gloucester, UK; 18Department of Medicine, McMaster HC, Hamilton Ontario, Canada; 19School of Medicine, Dentistry, and Biomedical Sciences, Centre for Public Health, Queens University Belfast, N. Ireland; 20University of Southampton, Southampton General Hospital, Tremona Road, Southampton, UK; 21Lancashire Teaching Hospitals NHS Foundation Trust, Royal Preston Hospital, Lancashire, UK; 22General Surgery, Countess of Chester Hospital, Chester, UK; 23Southampton University Hospitals NHS Trust, Southampton, UK; 24Wolfson Digestive Diseases Centre, Queens Medical Centre, Nottingham, UK; 25Worcestershire Acute Hospitals NHS Trust, Alexandra Hospital, Redditch, UK; 26Department of Gastroenterology, Royal Cornwall Hospital, Truro, Cornwall, UK; 27Macclesfield General Hospital, Macclesfield, Cheshire, UK; 28Department of Gastroenterology, Sandwell General Hospital, Lyndon, West Bromwich, UK; 29Department of Gastroenterology, University Hospital of Coventry, Coventry, UK; 30Department of Gastroenterology, Leeds General Infirmary, Leeds, UK; 31School of Biomedical and Healthcare Sciences, University of Plymouth, Plymouth, UK; 32Gastroenterology, Royal Albert Edward Infirmary NHS Trust, Wigan, UK; 33GI Science Centre, Salford Royal NHS Foundation Trust, University of Manchester, Salford, UK; 34Department of Gastroenterology and Hepatology, Department of Internal Medicine, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands; 35Forth Valley Royal Hospital, Larbert, Scotland, UK; 36Norfolk and Norwich University Hospitals NHS Foundation Trust, Norfolk and Norwich University Hospital, Norwich, UK; 37Burnley General Hospital, Burnley, Lancashire, UK; 38Department of Gastroenterology, University Hospital of Northern BC, Prince George, British Columbia, Canada; 39Bradford Teaching Hospitals NHS Foundation Trust, Bradford Royal Infirmary, Bradford, UK; 40Royal United Hospital Bath NHS
BACKGROUND & AIMS: Barrett’s esophagus (BE) increases the risk of esophageal adenocarcinoma (EAC). We found the risk to be BE has been associated with single nucleotide polymorphisms (SNPs) on chromosome 6p21 (within the HLA region) and on 16q23, where the closest protein-coding gene is FOXP1. Subsequently, the Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON) identified risk loci for BE and esophageal adenocarcinoma near CRTC1 and BARX1, and within 100 kb of FOXP1. We aimed to identify further SNPs that increased BE risk and to validate previously reported associations. METHODS: We performed a genome-wide association study (GWAS) to identify variants associated with BE and further analyzed promising variants identified by BEACON by genotyping 10,158 patients with BE and 21,062 controls. RESULTS: We identified 2 SNPs not previously associated with BE: rs3072 (2p241; odds ratio [OR] = 1.14; 95% CI: 1.09–1.18; P = 1.8 × 10−11) and rs2701108 (12q24.21; OR = 0.90; 95% CI: 0.86–0.93; P = 7.5 × 10−4). The closest protein-coding genes were respectively GDF7 (rs3072), which encodes a ligand in the bone morphogenetic protein family, and TBX5 (rs2701108), which encodes a transcription factor that regulates esophageal development and cardiac valves. Our data also supported in BE cases 3 risk SNPs identified by BEACON (rs2687201, rs11789015, and rs10423674). Meta-analysis of all data identified another SNP associated with BE and esophageal adenocarcinoma: rs3784252, within ALDH1A2 (OR = 0.90; 95% CI: 0.87–0.93; P = 3.72 × 10−8). CONCLUSIONS: We identified 2 loci associated with risk of BE and provided data to support a further locus. The genes we found to be associated with risk for BE encode transcription factors involved in thoracic, diaphragmatic, and esophageal development or proteins involved in the inflammatory response.

Keywords: EAC; Intestinal Metaplasia; Susceptibility; Cancer.
Our GWAS previously identified 2 SNPs, on chromosomes 6p21 (rs9257809; \( P = 4.1 \times 10^{-9}\)) and 16q24 (rs9936833; \( P = 2.7 \times 10^{-10}\)), that are associated with BE.\(^{11}\) One of these loci lies within the HLA region and the other, close to FOXF1, which is involved in esophageal structure and development. Both SNPs have also been shown to be associated with risk of EAC.\(^{14}\) More recently, in a combined analysis of BE and EAC cases, the Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON) identified susceptibility SNPs within CRTC1 and BARX1, and near FOXP1.\(^{12}\) The last 2 of these genes are known to be involved in esophageal development.\(^{15,16}\)

We aimed to identify further BE predisposition SNPs from our GWAS by performing wider and deeper independent replication of SNPs that already had promising disease associations.

**Methods**

**Patient and Sample Collection Criteria**

Figure 1 outlines this study and the numbers of samples that contributed to each phase. As described previously,\(^{11}\) Discovery Phase cases were diagnosed with histologically confirmed BE and ascertained through the UK-based Aspirin Esomeprazole Chemoprevention Trial (AspECT) in Barrett’s Metaplasia, a clinical trial of proton-pump inhibitor (esomeprazole) and aspirin as preventive agents for progression of BE to EAC.\(^{17}\) Replication Phase UK, Irish, Dutch, and Belgian patient samples were obtained from the Chemoprevention Of Premalignant Intestinal Neoplasia (ChOPIN) genetic study and the Esophageal Adenocarcinoma GenEtics (EAGLE) consortium.\(^{1}\) Replication Phase patients were diagnosed with BE with lengths of \( \geq 1 \) cm (C1M1) circumferential disease or \( \geq 2 \) cm tongue patterns (C0M2), according to the Prague criteria.\(^{18}\) Patient collection was in accordance with British
Society of Gastroenterology criteria for BE\textsuperscript{19} and followed verification of endoscopic findings and proven BE on histopathologic records. Presence of EAC at presentation or subsequently was recorded, but was not an inclusion criterion.

**Sample Sets**

**Discovery Phase.** BE patients (n = 1852) were UK participants in the AspECT study (Chief Investigator: Jankowski), HANDEL study (Chief Investigator: Jankowski), ChOPIN study (Chief Investigator: Jankowski), and population controls of white Caucasian origin (n = 5172) were from the common Wellcome Trust Case Control Consortium 2 (WTCCC2) set.\textsuperscript{11}

**Replication Phase 1.** UK Replication 1 totaled 1105 BE patients from ChOPIN and 6819 controls. The controls comprised People of the British Isles (Chief Investigator: W. Bodmer) (n = 2578) and WTCCC2 (Chief Investigator: P. Donnelly) samples (n = 4241) that were not genotyped in the Discovery Phase. The Dutch replication samples consisted of 473 BE patients and 1780 controls from the University Medical Centre, Groningen. An additional 64 Dutch cases and 206 controls, provided since 2012 from Nijmegen and Rotterdam as part of EAGLE, were genotyped for the 7 SNPs taken into Replication Phase 3.

**Replication Phase 2.** UK Replication 2 comprised 1765 BE patients from the ChOPIN study. Controls (n = 1586) were from the Colorectal Tumour Gene Identification (CoRGI) Consortium\textsuperscript{22} (Chief Investigator: Tomlinson), comprising spouses or partners unaffected by cancer and without a family history (to 2\textsuperscript{nd}-degree relative level) of colorectal neoplasia. All were of white UK ethnic origin. The Irish replication samples were 245 BE patients and 473 controls of white Caucasian origin from St James’s Hospital and Mater Misericordiae University Hospital, Dublin. Healthy donor controls were provided by Trinity Biobank.

**Replication Phase 3.** UK Replication 3 comprised 997 BE patients from the ChOPIN study and 974 female controls from the Genetics of Lobular Carcinoma In Situ in Europe (GLACIER) study (Chief Investigators: Sawyer, Roylance) with no personal or family history of breast cancer and of white Caucasian origin.\textsuperscript{24} The Belgian replication samples consisted of 362 cases and 848 controls from Leuven. Finally, 3295 BE patients and 3204 controls predominantly of northern European descent from the BEACON consortium GWAS (Chief Investigators: Vaughan, Whiteman, Levine) were included.\textsuperscript{12} All studies received ethical board approval (details in Supplementary Material). Two SNPs described in Su et al\textsuperscript{11} had been genotyped previously in Replication Phase 2 and BEACON/BEAGESS samples. All other Replication Phase 3 samples were new to this study.

**Genotyping**

For all samples, genomic DNA was extracted from peripheral blood. Various genotyping methods were used, depending on the phase of the study and on pre-existing data from some sample sets. In brief, Discovery Phase genotyping was performed using the Illumina 660W-Quad array for cases and a custom Human 1.2M-Duo array for controls at the Wellcome Trust Sanger Institute.\textsuperscript{21} Replication Phase 1 genotyping was performed using the Illumina Immunochip at the Wellcome Trust Sanger Institute\textsuperscript{11} or as described in Trynka et al.\textsuperscript{22} In Replication Phase 2, samples underwent custom genotyping for SNPs that met one of the following criteria: $P_{\text{association}} < 10^{-4}$ in combined Discovery and Replication Phase 1 analysis (n = 63); $P_{\text{association}} < 10^{-4}$ in Discovery Phase, but not included in Replication Phase 1 (n = 12); and $P_{\text{association}} < 10^{-4}$ in a sex-stratified analysis of the Discovery phase (n = 5); and candidate polymorphisms previously reported as associated with BE and not well tagged by the Discovery Phase or Immunochip arrays, specifically, $MSR1$ p.Arg293Gly,\textsuperscript{23} and variants in $IGF1R$ and $GHR$\textsuperscript{24} (Supplementary Table 1). Sequenom iPLEX assays were successfully designed for 65 of these SNPs and genotyping was performed at the Wellcome Trust Sanger Institute. Genotypes were assigned using MassArray TyperAnalyzer 4.0 (Sequenom). Samples with sex discrepancies between manifests or with overall call rates <95% were excluded, as were SNPs with call rates of <95%. Where a SNP was in the top 40 of the prioritized SNPs (by $P$ value) and had failed at the design stage of the iPLEX (n = 3), genotyping was performed by KASPar in the full Replication Phase 2 sample set. Seventy-seven of two hundred and forty-five Irish cases passing quality control were genotyped using iPLEX assays, and the other cases were genotyped using the Immunochip. Eighteen iPLEX SNPs were not analyzed in the Irish cohort, as the SNPs were not present on the Immunochip. The samples were also genotyped for 4 SNPs after publication of Levine et al.\textsuperscript{12}

Replication Phase 3 samples (Supplementary Table 2) were genotyped using KASPar for 7 SNPs prioritized after analysis of the previous phases. The samples were also genotyped for 4 SNPs after publication of Levine et al.\textsuperscript{12} All SNPs had call rates >95%. Sample exclusions were as for Replication Phase 2.

We had previously demonstrated >99% concordance between genome-wide array, Immunochip, and KASPar assays for other SNPs.\textsuperscript{11} Sequenom call rate was >96%. For samples analyzed only by KASPar, genotyping QC was tested using duplicate DNA samples within studies and SNP assays, together with direct sequencing of subsets of samples to confirm genotyping accuracy. For all SNPs, >98% concordant results were obtained.

**Association Analysis**

Case-control analysis was performed using frequentist tests under a missing data logistic regression model, as implemented in SNPTEST (version 2.4.1). Principal component analysis was performed for all samples typed on GWAS arrays (Discovery Phase) and has been described in Su et al.\textsuperscript{13} As described previously, principal component 1 (PC1) was included as a covariate in all analyses of the Discovery Phase. Each SNP was tested as a quantitative explanatory variable, coded as 0, 1, 2. We used GWAMA (version 2.1) to implement fixed inverse variance-based methods for meta-analysis.\textsuperscript{25} The software tests for heterogeneity of effects between studies\textsuperscript{26} and enables sex-specific meta-analysis.

**Replication and Validation of Single Nucleotide Polymorphisms From BEACON/BEAGESS Meta-analysis**

In order to examine the 4 genome-wide significant BE + EAC SNPs and 83 other SNPs with $P_{\text{assoc}} < 10^{-4}$ in the BEACON/BEAGESS data, we performed association testing using
AspECT Discovery Phase cases. Because our Discovery Phase controls overlapped with those used by Levine et al.,12 we used 1898 white European controls from colorectal cancer GWAS studies CoRGI and Colon Cancer Family Registry.27 Genotypes were imputed where necessary, with strict cut-offs for imputation quality.28,29

Other Analyses
Details of imputation, fine mapping, pathway analyses, estimation of heritability, and URLs are provided in the Supplementary Material.

Results
Identification of Two New Barrett’s Esophagus Predisposition Single Nucleotide Polymorphisms
In order to identify further loci associated with BE, we prioritized 65 SNPs (Supplementary Table 1) with the best evidence of association with BE from our previous GWAS Discovery Phase and Replication Phase 1 (details in Methods).11 These SNPs were genotyped in an additional 1765 cases and 1586 controls from the UK and in an Irish cohort of 245 cases and 473 controls (Replication Phase 2, previously used to genotype rs9257809 and rs9936833 described in Su et al.11). After meta-analysis of these new data together with Discovery Phase and Replication Phase 1, seven SNPs showing evidence of associations with BE risk at \( P_{\text{meta}} < 5 \times 10^{-6} \) were identified and genotyped in Replication Phase 3 samples (Table 1). After Replication Phase 3, two SNPs—rs3072 and rs2701108 on chromosome 2p24 and 12q24, respectively—reached the level of significance conventionally used for GWAS (\( P = 5 \times 10^{-8} \)) (Table 1). Combined \( P_{\text{meta}} \) values were \( 1.8 \times 10^{-11} \) for rs2701108 (OR = 1.14; 95% CI: 1.09–1.18) and \( 7.5 \times 10^{-10} \) for rs2701108 (OR = 0.90; 95% CI: 0.86–0.93), derived from a total sample of 10,158 BE cases and 21,062 controls (Supplementary Table 3). The associations remained at or near genome-wide significance upon restricting the analysis to the 8521 cases with histologically proven intestinal metaplasia (rs3072: \( P = 1.3 \times 10^{-5} \); OR = 1.13; 95% CI: 1.09–1.17; rs2701108: \( P = 6.2 \times 10^{-4} \); OR = 0.90; 95% CI: 0.86–0.94). There was no evidence of sex heterogeneity for either SNP (Supplementary Table 4).

In Silico Fine Mapping and Annotation of the Chromosome 2p24 and 12q24 Loci
rs3072 lies between 2 genes, mapping 7.5 kb downstream of \( GDF7 \) (also known as \( BMP12 \)) and 6.5 kb downstream of \( C2orf43 \) (Figure 2). rs2701108 is 117 kb downstream of \( GDF7 \) and 270 kb upstream of \( RBM19 \). We imputed in our Discovery Phase all SNPs in 1-Mb regions flanking each of the lead SNPs. At chromosome 2p24, rs3072 remained the most strongly associated SNP, but at chromosome 12q24, rs1920562 was more strongly associated with disease risk (\( P_{\text{Discovery}} = 1.4 \times 10^{-5} \); OR = 0.84) than the lead genotyped SNP (\( P_{\text{Discovery}} = 1.4 \times 10^{-3} \); OR = 0.88). rs1920562 (linkage disequilibrium [LD] with rs2701108; \( r^2 = 0.6 \)) lies 131 kb downstream of \( TBX5 \) and 256 kb upstream of \( RBM19 \). Nonsynonymous SNPs in the genes flanking the signals on chromosomes 2 and 12 were not in strong LD (\( r^2 < 0.4 \); \( D' < 0.8 \)) with the lead genotyped or imputed SNPs, suggesting that the functional variants may have effects on gene expression and regulation rather than protein sequence. Haploreg20,21 and Annovar31 were used to annotate SNPs in strong LD (\( r^2 > 0.4 \)) with the 2 lead tagging SNPs.

rs3072, which may alter a GATA binding motif, lies within a region of histone modifications, such as H3K4Me1, which mark enhancers (data from lymphoblastoid cell line (LCL) GM12878). Three other SNPs in LD with rs3072 map to the enhancer region detected in GM12878. One of these, rs7255, maps to a site of high evolutionary conservation/constraint; another SNP, rs9306894, whilst not at a conserved site (Supplementary Table 5), is predicted as “likely to affect protein binding and linked to expression of a gene target” according to RegulomeDB.32 We examined associations between SNPs in this region and gene expression in The Cancer Genome Atlas (TCGA) EAC data.33 Genotypes were only available for rs9306894 in the chromosome 2 locus and gene expression data had been obtained using RNASeq. After correcting for copy number, we determined associations between rs9306894 genotype and total RNA levels for expression quantitative trait locus (eQTL) analysis and bias in allelic expression of coding SNPs (allele-specific expression [ASE] analysis). There was no significant association with expression of the closest genes \( GDF7 \), \( HS1BP3 \) and \( C2orf43 \) (\( P_{\text{eQTL}} > 0.20; P_{\text{ASE}} > 0.38; n = 62 \)) and no genome-wide association with expression of any other gene was present (\( q > 0.05 \), details not shown). In public data sets based on monocytes34 and on lymphoblastoid cell lines and adipose tissue,35 \( C2orf43 \) is the suggested target of rs9306894 following eQTL studies (GENevar; \( P = 7 \times 10^{-4} \)). rs9306894 genotype was not associated with \( GDF7 \) expression in these cell types.

rs2701108 itself is not likely to be a functionally regulatory SNP, but rs1920562, which showed the strongest regional association after imputation, is a more promising candidate (Supplementary Table 6). This SNP maps to a highly conserved base and a region containing enhancer marks in human embryonic stem cells (h1-ESC) and lung fibroblasts (NHFL). rs1920562 and an additional SNP (rs1247938) in moderate LD (\( r^2 = 0.52 \)) with rs2701108, are highlighted by RegulomeDB as being the most likely SNPs in this region to affect protein binding. CTCF and RAD21 binding are predicted to be affected by rs1247938 and the ability of IKZF1 to bind is predicted to be altered by rs1920562. Expression analyses were performed for the rs2701108 region, in the same way as for rs9306894. However, none of the three rs2701108 region SNPs was associated with \( TBX5 \), \( TBX3 \) or \( RBM19 \) expression in the TCGA data (\( P_{\text{eQTL}} > 0.39; P_{\text{ASE}} > 0.43; n = 62 \)), was an eQTL in whole-transcriptome analysis, or was an eQTL in the public databases (details not shown).

Pathway/Geneset Enrichment Analysis
Improved Gene Set Enrichment Analysis for Genome Wide Association Study (iGSEA4GWAS) and SNP ratio test respectively found 26 and 34 pathways significantly
Table 1. Meta-analysis of Discovery and Replication Phase Sample Sets for SNPs Taken Into Replication Phase 3

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr (build 37)</th>
<th>Position</th>
<th>Minor/MAF (cases/controls)</th>
<th>Discovery MAF</th>
<th>Discovery Chr</th>
<th>Replication phase 1 (rep 1)</th>
<th>Replication phase 2 (rep 2)</th>
<th>Meta 1 (discovery + rep 1)</th>
<th>Meta 2 (meta 1 + rep 2)</th>
<th>Replication phase 3 (rep 3)</th>
<th>Final meta (meta 2 + rep 3)</th>
<th>Final meta $I^2$</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3072</td>
<td>2</td>
<td>20878406</td>
<td>G/A</td>
<td>0.41/0.36</td>
<td>1.23</td>
<td>(1.14–1.33)</td>
<td>(1.02–1.19)</td>
<td>$2.64 \times 10^{-7}$</td>
<td>$1.16 \times 10^{-8}$</td>
<td>$8.13 \times 10^{-13}$</td>
<td>$1.23 \times 10^{-9}$</td>
<td>0.42</td>
<td>8</td>
</tr>
<tr>
<td>rs6751791</td>
<td>2</td>
<td>35581997</td>
<td>A/G</td>
<td>0.51/0.48</td>
<td>1.15</td>
<td>(1.06–1.23)</td>
<td>(1.07–1.25)</td>
<td>$5.03 \times 10^{-4}$</td>
<td>$2.68 \times 10^{-11}$</td>
<td>$1.14 \times 10^{-12}$</td>
<td>$1.09 \times 10^{-15}$</td>
<td>1.08</td>
<td>8</td>
</tr>
<tr>
<td>rs2731672</td>
<td>5</td>
<td>176842474</td>
<td>A/G</td>
<td>0.27/0.24</td>
<td>1.18</td>
<td>(1.09–1.28)</td>
<td>(1.04–1.24)</td>
<td>$1.64 \times 10^{-11}$</td>
<td>$1.14 \times 10^{-14}$</td>
<td>$1.14 \times 10^{-16}$</td>
<td>$1.07 \times 10^{-18}$</td>
<td>0.63</td>
<td>8</td>
</tr>
<tr>
<td>rs2701108</td>
<td>12</td>
<td>114674261</td>
<td>A/G</td>
<td>0.38/0.41</td>
<td>0.88</td>
<td>(0.81–0.95)</td>
<td>(0.81–0.94)</td>
<td>$4.40 \times 10^{-13}$</td>
<td>$1.10 \times 10^{-12}$</td>
<td>$5.68 \times 10^{-15}$</td>
<td>$1.07 \times 10^{-16}$</td>
<td>0.14</td>
<td>8</td>
</tr>
<tr>
<td>rs189247</td>
<td>15</td>
<td>97586630</td>
<td>A/G</td>
<td>0.41/0.37</td>
<td>1.18</td>
<td>(1.09–1.27)</td>
<td>(1.05–1.23)</td>
<td>$5.67 \times 10^{-11}$</td>
<td>$1.10 \times 10^{-10}$</td>
<td>$6.12 \times 10^{-13}$</td>
<td>$1.09 \times 10^{-14}$</td>
<td>0.20</td>
<td>8</td>
</tr>
<tr>
<td>rs2043633</td>
<td>16</td>
<td>5819274</td>
<td>C/A</td>
<td>0.37/0.41</td>
<td>0.85</td>
<td>(0.79–0.92)</td>
<td>(0.82–0.95)</td>
<td>$6.04 \times 10^{-11}$</td>
<td>$2.49 \times 10^{-12}$</td>
<td>$4.74 \times 10^{-15}$</td>
<td>$1.08 \times 10^{-13}$</td>
<td>0.58</td>
<td>8</td>
</tr>
<tr>
<td>rs12985909</td>
<td>19</td>
<td>18439383</td>
<td>G/A</td>
<td>0.48/0.45</td>
<td>1.12</td>
<td>(1.04–1.21)</td>
<td>(1.04–1.21)</td>
<td>$2.94 \times 10^{-13}$</td>
<td>$2.73 \times 10^{-13}$</td>
<td>$1.99 \times 10^{-15}$</td>
<td>$1.07 \times 10^{-15}$</td>
<td>0.00</td>
<td>8</td>
</tr>
</tbody>
</table>

NOTE. For each phase, association data show (top to bottom) OR, (95% CI), and $P$ value. Results are presented with respect to the minor allele. rs3072 and rs2701108 reached genome-wide significance and thus are shown in bold. In BEACON, rs7598399 was used as a proxy for rs6751791 ($r^2 = 1$) and rs189247 was imputed from 4 genotyped SNPs (rs991757, rs2670927, rs2670930, and rs234540) with accuracy approximately 98%. The Dutch extension samples were analyzed with the previously described Dutch replication samples as part of Rep 1. The $P$ value threshold for including a SNP in Phase 2 was $1 \times 10^{-4}$ and that for inclusion in Phase 3 was $5 \times 10^{-6}$. Chr, chromosome; $I^2$, heterogeneity index; MAF, minor allele frequency; N, number of studies.
Organ Development (Supplementary Figure 1). The 5 genes implicated by the BE susceptibility SNPs here are all involved in development at a cellular, embryonic, organ, and organism level. Bone morphogenetic protein 4 was the most significant upstream regulator ($P_{\text{overlap}} = 1.99 \times 10^{-6}$).

Barrett’s Esophagus Heritability

Genome-wide haplotype-tagging SNP data on the 1852 cases and 5172 controls in the Discovery Phase of this study were used in Genome-wide Complex Trait Analysis to estimate the proportion of variation in risk of BE that can be explained by common genetic variants. In line with our previous, disease score test analysis, we found a statistically significant component of BE risk to be polygenic (9.99% [SE 1.2%]). This was a lower estimate than that recently derived by Ek et al.13

Replication Testing of Previously Reported Barrett’s Esophagus Susceptibility Single Nucleotide Polymorphisms at Candidate Loci

Using a systematic review, we identified 26 polymorphisms reported in the literature to be associated with BE (Supplementary Table 8). In our Discovery Phase samples, 20 of 26 SNPs were directly genotyped or were in strong LD ($r^2 > 0.7$) with a directly genotyped SNP. Only one of these SNPs showed a nominally significant association ($P < .05$) in our data (rs909253, proxy for rs1041981, $r^2 = 0.93$; OR = 1.12; $P = .005$). This SNP was also present on the Immunochip and showed additional evidence of replication ($P_{\text{meta}} = 3.1 \times 10^{-4}$, OR = 1.07). rs909253 maps to a highly conserved base (based on SiPhy score) in an intron of LTA (tumor necrosis factor-β) where histone marks associated with both promoters and enhancers are present in lymphoblastoid cell lines (LCLs).30 PBX3, PU1, POL2, YY1, and nuclear factor-κB have all been found to bind here in ENCODE ChIP-seq experiments. No data were available for this SNP in RegulomeDB. We were able to genotype 3 other candidate SNPs previously reported for BE susceptibility (rs41341748, rs2715425, rs6898743) on the Sequenom iPLEX panel used to genotype UK Replication 2. None of these SNPs showed associations with BE risk (Supplementary Table 1). Because rs41341748 (MSR1 p.Arg293Gly) has a low minor allele frequency (<5%) and consequently our power to detect an association was relatively low, we additionally genotyped it in UK Replication 3. After meta-analysis of Replication 2 and 3, we remained unable to replicate the previously reported association between this SNP and BE risk (OR = 1.07; 95% CI: 0.70–1.43; $P = .79$).

Assessment of Previously Reported Barrett’s Esophagus + Esophageal Adenocarcinoma Single Nucleotide Polymorphisms and Meta-analysis With BEAGESS Data

Three new genome-wide significant BE + EAC loci (4 SNPs) were recently identified by Levine et al.12 in a combined analysis of EAC and BE: rs10419226 and
rs10423674 in *CRTC1*, rs11789015 in *BARX1* and rs2687201 within 100 kb of *FOXP1*. None of these associations was genome-wide significant at $P < 5 \times 10^{-8}$ when Levine et al restricted their analysis to BE cases alone, although one SNP, rs10419226, within *CRTC1*, reached $P = 5.5 \times 10^{-8}$. In our datasets, only rs10423674 had been directly genotyped, but the remaining SNPs were all reliably imputed in our Discovery Phase samples (Info scores >0.95). However, the controls used in the replication phase of the Levine et al study overlapped entirely with the controls used in our Discovery phase (WTCCC2 controls) and we therefore used alternative UK controls from the GoRGI study (see Methods).

Of the 4 Levine SNPs, rs10423674, 1 of 2 SNPs in *CRTC1*, and rs2687201, near *FOXP1*, were supported in this study ($P = 0.02$; OR = 1.14; 95% CI: 1.03–1.27 and $P = 0.05$; OR = 0.94; 95% CI: 0.88–1.00, respectively). There was also some support for rs11789015, near *BARX1* ($P = 0.07$; OR = 0.90; 95% CI: 0.81–1.01). However, the association at rs10419226, within *CRTC1*, was not replicated in our data ($P = 0.87$; OR = 1.01; 95% CI: 0.91–1.11). All 4 SNPs still reached genome-wide significance ($P < 5 \times 10^{-8}$) upon meta-analysis of our BE and Levine’s BE + EAC datasets. In a BE-only meta-analysis, the associations improved with the inclusion of our data for 3 out of the 4 SNPs, with 1 (rs2687201, *FOXP1*) reaching genome-wide significance for BE (Table 2).

We then addressed the other 87 other SNPs with $P_{assoc} < 10^{-4}$ in the Levine data (Supplementary Table 3 of Levine et al). Of these, 73 were directly genotyped in our samples or could be imputed with an IMPUTE2 info score of >0.95. Of the 10 SNPs that could not be imputed with high quality, only one had $P_{assoc} < 10^{-5}$ in the original Levine data; we therefore genotyped this SNP (rs11771429) using KASP primers in our cases and controls. We did not obtain genotypes for the remaining 9 SNPs. On performing a meta-analysis of the Levine BE + EAC cases with our UK Discovery Phase, 4 SNPs (rs1497205, rs254348, rs3784262, and rs4523255) provided strong evidence that rs3072 or rs2701108 (or other SNPs in strong LD) were eQTLs or influenced ASE. The absence of these associations is typical for GWAS or cancer or precancerous traits. Even for the “prototypic” multi-cancer SNP rs6983267, convincingly demonstrating the effects of SNP alleles on gene expression has required considerable additional work in a variety of systems, and even now, consistent eQTL and ASE associations have not been shown.\(^{41-44}\) The likely major reason for the lack of eQTLs at GWAS SNPs is that the SNPs have their effects in a restricted set of cells or at a particular time. There is, for example, evidence that the forkhead box (FOX) proteins are most strongly expressed during embryogenesis, and that the levels of these transcription factors are critical for proper development.\(^{45-47}\) Given this, our first choice in searching for eQTLs would be cells in the developing human thorax. Unfortunately, such sample collections do not currently exist.

We showed rs2687201 (*FOXP1*) to be associated with disease in a BE-only analysis. Our data generally support the report by Levine et al of associations between BE + EAC and SNPs on chromosome 3 (*FOXP1*), chromosome 9 (*BARX1*), and one of the SNPs on chromosome 19 (*CRTC1*). However, we were not able to replicate the association observed for another SNP (rs10419226) in *CRTC1*. For this last SNP, the meta-analysis showed evidence of significant heterogeneity between the BEACON/BEAGESS data and our data (Table 2), and in the absence of clear reasons for this difference, we caution against drawing firm conclusions here.

We found another SNP, rs3784262 (*ALDH1A2*), to be formally associated with BE + EAC upon meta-analysis of our data with the Levine BE + EAC dataset. *ALDH1A2* encodes retinaldehyde dehydrogenase 2, which catalyzes the synthesis of retinoic acid and may also be involved in signal transduction at the esophageal mucosa in BE. Given this, our second choice in searching for eQTLs would be cells in the developing human thorax.
Table 2. Meta-analysis With Our Data for 4 BE/EAC SNPs and 4 Other Selected SNPs with \( P < 1 \times 10^{-4} \) from Levine et al.\textsuperscript{12}

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Position</th>
<th>Nearby genes</th>
<th>Minor/ major allele</th>
<th>BE+/− EAC</th>
<th>Levine et al\textsuperscript{12} meta OR (95% CI)</th>
<th>This study meta OR (95% CI)</th>
<th>This study + Levine et al\textsuperscript{12} meta OR (95% CI)</th>
<th>( P_{\text{Levine et al}} )</th>
<th>( P_{\text{This Study}} )</th>
<th>( P_{\text{meta}} )</th>
<th>Meta I\textsuperscript{2}</th>
<th>No. of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2687201</td>
<td>3</td>
<td>70928930</td>
<td>FOXP1</td>
<td>T/G</td>
<td>BE</td>
<td>1.18 (1.10–1.26)</td>
<td>1.14 (1.03–1.27)</td>
<td>1.18 ( \times 10^{-2} )</td>
<td>1.16 (1.10–1.23)</td>
<td>4.61 ( \times 10^{-8} )</td>
<td>0.00</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>rs11789015</td>
<td>9</td>
<td>96716028</td>
<td>BARX1</td>
<td>G/A</td>
<td>BE</td>
<td>0.85 (0.79–0.91)</td>
<td>0.90 (0.81–1.01)</td>
<td>6.63 ( \times 10^{-2} )</td>
<td>0.86 (0.81–0.92)</td>
<td>1.38 ( \times 10^{-6} )</td>
<td>0.00</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>rs10419226</td>
<td>19</td>
<td>18803172</td>
<td>CRTC1</td>
<td>A/C</td>
<td>BE</td>
<td>1.19 (1.12–1.26)</td>
<td>1.01 (0.91–1.11)</td>
<td>8.65 ( \times 10^{-1} )</td>
<td>1.13 (1.08–1.20)</td>
<td>2.14 ( \times 10^{-6} )</td>
<td>0.82</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>rs10423674</td>
<td>19</td>
<td>18817903</td>
<td>CRTC1</td>
<td>T/G</td>
<td>BE</td>
<td>0.85 (0.80–0.91)</td>
<td>0.94 (0.88–1.00)</td>
<td>4.88 ( \times 10^{-2} )</td>
<td>0.89 (0.85–0.93)</td>
<td>2.99 ( \times 10^{-7} )</td>
<td>0.40</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>rs1497205</td>
<td>4</td>
<td>76169067</td>
<td>PARM1, RCHY1</td>
<td>C/T</td>
<td>BE</td>
<td>0.86 (0.80–0.92)</td>
<td>0.92 (0.87–0.98)</td>
<td>7.59 ( \times 10^{-1} )</td>
<td>0.90 (0.86–0.94)</td>
<td>2.57 ( \times 10^{-6} )</td>
<td>0.00</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>rs254348</td>
<td>16</td>
<td>65980789</td>
<td>T/C</td>
<td>BE/EAC</td>
<td>BE</td>
<td>0.87 (0.82–0.93)</td>
<td>1.28 ( \times 10^{-5} )</td>
<td></td>
<td>0.90 (0.86–0.94)</td>
<td>3.68 ( \times 10^{-7} )</td>
<td>0.00</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>rs3784262</td>
<td>15</td>
<td>58253106</td>
<td>ALDH1A2</td>
<td>G/A</td>
<td>BE</td>
<td>0.89 (0.84–0.94)</td>
<td>1.40 ( \times 10^{-5} )</td>
<td></td>
<td>0.92 (0.89–0.96)</td>
<td>2.81 ( \times 10^{-5} )</td>
<td>0.53</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>rs4523255</td>
<td>8</td>
<td>8713038</td>
<td>MFHAS1</td>
<td>A/G</td>
<td>BE</td>
<td>1.13 (1.06–1.21)</td>
<td>1.07 (1.01–1.12)</td>
<td>2.11 ( \times 10^{-2} )</td>
<td>1.09 (1.05–1.14)</td>
<td>2.48 ( \times 10^{-5} )</td>
<td>0.36</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

NOTE. The minimum meta-analysis comprised the Levine et al.\textsuperscript{12} Discovery and Replication Phases and our Discovery Phase (with amended controls, as described in Methods). rs10423674 was additionally genotyped in our UK and Dutch Replication Phase 1. rs1497205, rs254348, rs3784262, and rs4523255 were genotyped in our UK and Dutch Replication Phase 1 and UK Replication Phase 2. rs3784262 was also genotyped in Irish Replication Phase 2 samples and UK and Belgium Replication Phase 3 samples. For rs10419226, which shows evidence of inter-study heterogeneity, random effects model \( P \) values for BE and BE/EAC are .10 and .04, respectively. Chr, chromosome.
alcohol metabolism. Of the candidate SNPs we assessed (Supplementary Table 8), we found supporting evidence, albeit short of genome-wide significance, for rs909253 ($P = 3.1 \times 10^{-5}$), mapping to an intronic region of LTA within the HLA region, but not in LD with rs9257809, the other HLA BE SNP.

We previously reported that our original GWAS provided evidence that multiple common variants, each with small effects contribute to BE susceptibility. Ek et al recently estimated that the heritability of BE is 35% (SE 6%). We also found that the heritability of BE is highly significant, but explains only 9.99% of BE risk (SE 1.2%). Our GWAS consisted of UK cases and controls, and the BEACON/BEAGESS samples used by Ek et al originated from 3 continents (Europe, North America, and Australia). Cryptic population stratification could perhaps explain the larger estimate of heritability obtained using the BEACON/BEAGESS GWAS. In addition, we used software to calculate LD-adjusted kinships, such that the SNPs used in the heritability analysis were weighted according to local LD structure. It has been found that heritability estimation from genome-wide SNPs is highly sensitive to uneven LD; causal SNPs in regions of high LD can lead to overestimation of heritability and conversely causal SNPs in regions of low LD can result in an underestimation of heritability.

Although the BE GWAS have not yet identified the functional SNPs in each region or their gene targets, the information generated already permits the generation of hypotheses regarding processes that may be involved in BE. First, transcription factors involved in development and structure of the thorax, diaphragm, and esophagus may be important: the SNPs near FOXF1, FOXP1, BARX1, and TBX5 might act in this way and the genes appear to be functionally related (Supplementary Figure 1). Second, the inflammatory response may be important: the SNPs within the HLA region (rs9257809 and, perhaps, rs909253) might act in this way and pathway analysis provided suggestive evidence of a role for type 1 diabetes genes in BE etiology. A plausible, testable hypothesis is that these 2 groups of SNPs respectively influence the tendency to gastroesophageal reflux disease, perhaps through thoracic and diaphragmatic structure (hiatal hernia defect), and the inflammatory response to the refluxed gastric acid. Given the limited scope for clinical intervention in the former processes, we await with interest the outcome of trials such as AspECT that target the inflammatory response to gastric reflux.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of Gastroenterology at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2014.10.041.

References


Author names in bold designate shared co-first authorship. Received May 6, 2014. Accepted October 21, 2014.

Reprint requests
Address requests for reprints to: Janusz Jankowski, MBChB, MSc, MD, PhD, FRCP, AQAF, FACG, SFHEA, CPgME, University Hospitals Coventry & Warwickshire NHS Trust and Warwick Medical School, University of Warwick, Warwickshire, England, CV2 2DX. e-mail: J.Jankowski@warwick.ac.uk; fax: +44 (0)24 7696 8653; Claire Palles, PhD, Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford,
Acknowledgments
The authors would like to thank the WTCCC2 consortium, the BEACON consortium and the AspECT, BOSS, and ChOPIN trial teams (comprising the EAGLE consortium) and participants in all these studies. The Discovery Phase and Immunochip replication were funded by the Wellcome Trust IPOD grant (084722/Z/08/Z). In addition, sample collection for all phases was core supported by the Cancer Research UK funded AspECT, ChOPIN and Handel trials. CoRGI and GLACIER were also funded by Cancer Research UK. In addition, the authors thank AstraZeneca for an educational grant for the tissue and blood collection. Core funding to the Wellcome Trust Centre for Human Genetics was provided by the Wellcome Trust (090532/Z/09/Z). We also thank Liam J. Murray and Wong-Ho Chow, who provided samples as part of the BEACON consortium. The authors would also like to thank the Experimental Cancer Medicine Centre–supported University of Southampton Faculty of Medicine Tissue Bank. The Cancer Genome Atlas (TCGA) esophageal carcinoma data were also gratefully used for expression analyses. Finally, the authors want to thank Dr Vincent Plagnol of University College London for advice on statistical analyses and presentation.

Conflicts of interest
This author discloses the following: Janusz Jankowski is Chief Investigator of AspECT and ChOPIN trials and was an AstraZeneca consultant from 2002 to 2012. The remaining authors disclose no conflicts.

Funding
This work was supported by Cancer Research UK (Chopin Grant C548, AspECT Grant A4584, AZ Educational Grant J2), Peninsula Schools of Medicine and Dentistry School (Grant code - “PUPSMED -JJ2013-1”), Wellcome Trust (084722/Z/08/Z and 090532/Z/09/Z), and an AstraZeneca UK educational grant.