

# BASIC AND TRANSLATIONAL—ALIMENTARY TRACT

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



Claire Palles,<sup>1,\*</sup> Laura Chegwidden,<sup>2,\*</sup> Xinzhong Li,<sup>3</sup> John M. Findlay,<sup>1</sup> Garry Farnham,<sup>2</sup> Francesc Castro Giner,<sup>1</sup> Maikel P. Peppelenbosch,<sup>4</sup> Michal Kovac,<sup>1</sup> Claire L. Adams,<sup>2</sup> Hans Prenen,<sup>5</sup> Sarah Briggs,<sup>1</sup> Rebecca Harrison,<sup>6</sup> Scott Sanders,<sup>7</sup> David MacDonald,<sup>8</sup> Chris Haigh,<sup>9</sup> Art Tucker,<sup>10</sup> Sharon Love,<sup>11</sup> Manoj Nanji,<sup>12</sup> John deCaestecker,<sup>13</sup> David Ferry,<sup>14</sup> Barrie Rathbone,<sup>15</sup> Julie Hapeshi,<sup>16</sup> Hugh Barr,<sup>17</sup> Paul Moayyedi,<sup>18</sup> Peter Watson,<sup>19</sup> Barbara Zietek,<sup>12</sup> Neera Maroo,<sup>12</sup> Laura Gay,<sup>12</sup> Tim Underwood,<sup>20</sup> Lisa Boulter,<sup>20</sup> Hugh McMurtry,<sup>21</sup> David Monk,<sup>22</sup> Praful Patel,<sup>23</sup> Krish Ragnunath,<sup>24</sup> David Al Dulaimi,<sup>25</sup> Iain Murray,<sup>26</sup> Konrad Koss,<sup>27</sup> Andrew Veitch,<sup>14</sup> Nigel Trudgill,<sup>28</sup> Chuka Nwokolo,<sup>29</sup> Bjorn Rembacken,<sup>30</sup> Paul Atherfold,<sup>31</sup> Elaine Green,<sup>32</sup> Yeng Ang,<sup>33,34</sup> Ernst J. Kuipers,<sup>35</sup> Wu Chow,<sup>36</sup> Stuart Paterson,<sup>37</sup> Sudarshan Kadri,<sup>15</sup> Ian Beales,<sup>38</sup> Charles Grimley,<sup>39</sup> Paul Mullins,<sup>40</sup> Conrad Beckett,<sup>41</sup> Mark Farrant,<sup>42</sup> Andrew Dixon,<sup>43</sup> Sean Kelly,<sup>44</sup> Matthew Johnson,<sup>45</sup> Shahjehan Wajed,<sup>46</sup> Anjan Dhar,<sup>47</sup> Elinor Sawyer,<sup>48</sup> Rebecca Roylance,<sup>49</sup> Lynn Onstad,<sup>50</sup> Marilie D. Gammon,<sup>51</sup> Douglas A. Corley,<sup>52</sup> Nicholas J. Shaheen,<sup>53</sup> Nigel C. Bird,<sup>54</sup> Laura J. Hardie,<sup>55</sup> Brian J. Reid,<sup>50,56</sup> Weimin Ye,<sup>57</sup> Geoffrey Liu,<sup>58</sup> Yvonne Romero,<sup>59,60</sup> Leslie Bernstein,<sup>61</sup> Anna H. Wu,<sup>62</sup> Alan G. Casson,<sup>63</sup> Rebecca Fitzgerald,<sup>64</sup> David C. Whiteman,<sup>65</sup> Harvey A. Risch,<sup>66</sup> David M. Levine,<sup>67</sup> Tom L. Vaughan,<sup>50</sup> Auke P. Verhaar,<sup>4</sup> Jan van den Brande,<sup>68</sup> Eelke L. Toxopeus,<sup>69</sup> Manon C. Spaander,<sup>4</sup> Bas P. L. Wijnhoven,<sup>69</sup> Luc J. W. van der Laan,<sup>69</sup> Kausilia Krishnadath,<sup>70</sup> Cisca Wijmenga,<sup>71</sup> Gosia Trynka,<sup>71</sup> Ross McManus,<sup>72</sup> John V. Reynolds,<sup>73</sup> Jacintha O'Sullivan,<sup>73</sup> Padraic MacMathuna,<sup>74</sup> Sarah A. McGarrigle,<sup>73</sup> Dermot Kelleher,<sup>75</sup> Severine Vermeire,<sup>5</sup> Isabelle Cleynen,<sup>5</sup> Raf Bisschops,<sup>5</sup> Ian Tomlinson,<sup>1</sup> and Janusz Jankowski<sup>76,77</sup>

<sup>1</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; <sup>2</sup>Plymouth University Peninsula School of Medicine and Dentistry, Plymouth, Devon, UK; <sup>3</sup>Centre of Biostatistics, Bioinformatics and Biomarkers, Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, Devon, UK; <sup>4</sup>Department of Gastroenterology and Hepatology, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands; <sup>5</sup>Department of Digestive Oncology, University Hospital Gasthuisberg, Leuven, Belgium; <sup>6</sup>Department of Pathology, Leicester Royal Infirmary, Leicester, UK; <sup>7</sup>Department of Cellular Pathology, Warwick Hospital, Warwick, UK; <sup>8</sup>Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, British Columbia, Canada; <sup>9</sup>Department of Gastroenterology, Wansbeck General Hospital, Ashington, Northumberland, UK; <sup>10</sup>William Harvey Research Institute, The Ernest Cooke Vascular & Microvascular Unit, Centre for Clinical Pharmacology, St Bartholomew's Hospital, London, UK; <sup>11</sup>Centre for Statistics in Medicine and Oxford Clinical Trials Research Unit, Oxford, UK; <sup>12</sup>Centre for Digestive Diseases, Queen Mary University of London, London, UK; <sup>13</sup>Department of Gastroenterology, Leicester General Hospital, Leicester, UK; <sup>14</sup>Department of Oncology, New Cross Hospital, Royal Wolverhampton Hospitals NHS Trust, Wolverhampton, UK; <sup>15</sup>Department for Gastroenterology, Leicester Royal Infirmary, Leicester, UK; <sup>16</sup>Gloucestershire Royal Hospital, Great Western Road, Gloucester, UK; <sup>17</sup>Department of Upper GI Surgery, Gloucestershire, Royal Hospital, Gloucester, UK; <sup>18</sup>Department of Medicine, McMaster HC, Hamilton Ontario, Canada; <sup>19</sup>School of Medicine, Dentistry, and Biomedical Sciences, Centre for Public Health, Queens University Belfast, NI; <sup>20</sup>University of Southampton, Southampton General Hospital, Tremona Road, Southampton, UK; <sup>21</sup>Lancashire Teaching Hospitals NHS Foundation Trust, Royal Preston Hospital, Lancashire, UK; <sup>22</sup>General Surgery, Countess of Chester Hospital, Chester, UK; <sup>23</sup>Southampton University Hospitals NHS Trust, Southampton, UK; <sup>24</sup>Wolfson Digestive Diseases Centre, Queens Medical Centre, Nottingham, UK; <sup>25</sup>Worcestershire Acute Hospitals NHS Trust, Alexandra Hospital, Redditch, UK; <sup>26</sup>Department of Gastroenterology, Royal Cornwall Hospital, Truro, Cornwall, UK; <sup>27</sup>Macclesfield General Hospital, Macclefield, Cheshire, UK; <sup>28</sup>Department of Gastroenterology, Sandwell General Hospital, Lyndon, West Bromwich, UK; <sup>29</sup>Department of Gastroenterology, University Hospital of Coventry, Coventry, UK; <sup>30</sup>Department of Gastroenterology, Leeds General Infirmary, Leeds, UK; <sup>31</sup>Department of Clinical Pharmacology University of Oxford, Oxford, UK; <sup>32</sup>School of Biomedical & Healthcare Sciences, Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK; <sup>33</sup>Gastroenterology, Royal Albert Edward Infirmary NHS Trust, Wigan, UK; <sup>34</sup>GI Science Centre, Salford Royal NHS Foundation Trust, University of Manchester, Salford, UK; <sup>35</sup>Department of Gastroenterology and Hepatology, Department of Internal Medicine, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands; <sup>36</sup>Forth Valley Royal Hospital, Larbert, Scotland, UK; <sup>37</sup>Norfolk and Norwich University Hospitals NHS Foundation Trust, Norfolk and Norwich University Hospital, Norwich, UK; <sup>38</sup>Burnley General Hospital, Burnley, Lancashire, UK; <sup>39</sup>Head of Gastroenterology, University Hospital of Northern BC, Prince George, British Columbia, Canada; <sup>40</sup>Bradford Teaching Hospitals NHS Foundation Trust, Bradford Royal Infirmary, Bradford, UK; <sup>41</sup>Royal United Hospital Bath NHS

Trust, Royal United Hospital, Avon, Bath, Somerset, UK; <sup>42</sup>Kettering General Hospital NHS Foundation Trust, Kettering General Hospital, Rothwell Road, Kettering, Northants, UK; <sup>43</sup>York Teaching Hospital NHS Foundation Trust, York, UK; <sup>44</sup>Luton and Dunstable University Hospital NHS Foundation Trust, Luton, Bedfordshire, UK; <sup>45</sup>Department of Thoracic and Upper Gastrointestinal Surgery, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK; <sup>46</sup>County and Durham and Darlington NHS Foundation Trust, Bishop Auckland, County Durham, UK; <sup>47</sup>Guy's and St Thomas' NHS Foundation Trust, London, UK; <sup>48</sup>Barts Cancer Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London, UK; <sup>49</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Centre, Seattle, Washington; <sup>50</sup>Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, North Carolina; <sup>51</sup>Division of Research and San Francisco Medical Center, Kaiser Permanente Northern California, California; <sup>52</sup>Division of Gastroenterology and Hepatology, UNC School of Medicine, University of North Carolina, Chapel Hill, North Carolina; <sup>53</sup>Department of Oncology, The Medical School, University of Sheffield, UK; <sup>54</sup>Division of Epidemiology, University of Leeds, Leeds, UK; <sup>55</sup>Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington; <sup>56</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; <sup>57</sup>Princess Margaret Cancer Centre, Ontario Cancer Institute, Toronto, ON, Canada; <sup>58</sup>Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota; <sup>59</sup>Registry, On behalf of the Romero; <sup>60</sup>Department of Population Sciences, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, California; <sup>61</sup>Department of Preventive Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California; <sup>62</sup>Department of Surgery, University of Saskatchewan, Saskatoon, Canada; <sup>63</sup>MRC Cancer Cell Unit, Hutchison-MRC Research Centre and University of Cambridge, Cambridge, UK; <sup>64</sup>Cancer Control, QIMR Berghofer Medical Research Institute, Queensland, Australia; <sup>65</sup>Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut; <sup>66</sup>Department of Biostatistics, University of Washington School of Public Health, Seattle, Washington; <sup>67</sup>Department of Gastroenterology, Tergooi Hospital, Hilversum, The Netherlands; <sup>68</sup>Department of Surgery, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, The Netherlands; <sup>69</sup>Department of Gastroenterology and Hepatology, Academic Medical Centre, Amsterdam, The Netherlands; <sup>70</sup>Department of Genetics, University Medical Centre Groningen and University of Groningen, The Netherlands; <sup>71</sup>Department of Clinical Medicine & Institute of Molecular Medicine, Trinity Centre for Health Sciences, Trinity College Dublin, St James's Hospital, Dublin, Ireland; <sup>72</sup>Department of Surgery, Trinity Centre for Health Sciences, Trinity College Dublin, St. James' Hospital, Dublin, Ireland; <sup>73</sup>Gastrointestinal Unit, Mater Misericordiae University Hospital, University College Dublin, Dublin, Ireland; <sup>74</sup>Faculty of Medicine, Imperial College London, South Kensington Campus, London, UK; <sup>75</sup>Faculty of Medicine, Imperial College, South Kensington Campus, London, UK; <sup>76</sup>University Hospitals Coventry & Warwickshire NHS Trust, Warwickshire, England; <sup>77</sup>Warwick Medical School, University of Warwick, Warwickshire, England

**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 (2p24.1; odds ratio [OR] = 1.14; 95% CI: 1.09–1.18;  $P = 1.8 \times 10^{-11}$ ) and rs2701108 (12q24.21; OR = 0.90; 95% CI: 0.86–0.93;  $P = 7.5 \times 10^{-9}$ ). The closest protein-coding genes were respectively *GDF7* (rs3072), which encodes a ligand in the bone morphogenetic protein pathway, and *TBX5* (rs2701108), which encodes a transcription factor that regulates esophageal and cardiac development. 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: rs3784262, within *ALDH1A2* (OR = 0.90; 95% CI: 0.87–0.93;  $P = 3.72 \times 10^{-9}$ ). **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.

Barrett's esophagus (BE) is a common premalignant condition that affects up to 2% of the adult population in the Western world.<sup>1</sup> BE comprises the second stage in the esophagitis–metaplasia–dysplasia–adenocarcinoma sequence. BE confers a 2%–4% lifetime risk of esophageal adenocarcinoma (EAC).<sup>1</sup> Chronic gastric acid reflux is the predominant etiologic factor for BE. In addition, BE co-occurs with conditions such as intestinal metaplasia, hiatal hernia, obesity, and hypercholesterolemia.<sup>2–5</sup> Several factors, including the degree of acid reflux, hiatal hernia size, and the percentage of intestinal metaplasia-positive glands, can affect the progression to cancer. A role for genetics in the pathogenesis of gastroesophageal reflux disease, including BE and EAC, has been implicated on the basis of 3 observations: concordance in monozygous and dizygous twins<sup>6–8</sup>; the increased risk of disease in those with a positive family history<sup>9,10</sup>; and, recently, the identification of single nucleotide polymorphisms (SNPs) associated with BE in Genome-Wide Association Studies (GWAS).<sup>11,12</sup> The proportion of variation in BE risk explained by common variants has been estimated to be 35%.<sup>13</sup>

\*Authors share co-first authorship.

**Abbreviations used in this paper:** ASE, allele-specific expression; BE, Barrett's esophagus; BEACON, Barrett's and Esophageal Adenocarcinoma Consortium; CI, confidence interval; EAC, esophageal adenocarcinoma; eQTL, expression quantitative trait locus; GWAS, genome-wide association study; LD, linkage disequilibrium; OR, odds ratio; PC, principal component; SNP, single nucleotide polymorphism; TCGA, The Cancer Genome Atlas.

**Keywords:** EAC; Intestinal Metaplasia; Susceptibility; Cancer.

© 2015 by the AGA Institute Open access under CC BY license.  
0016-5085

<http://dx.doi.org/10.1053/j.gastro.2014.10.041>

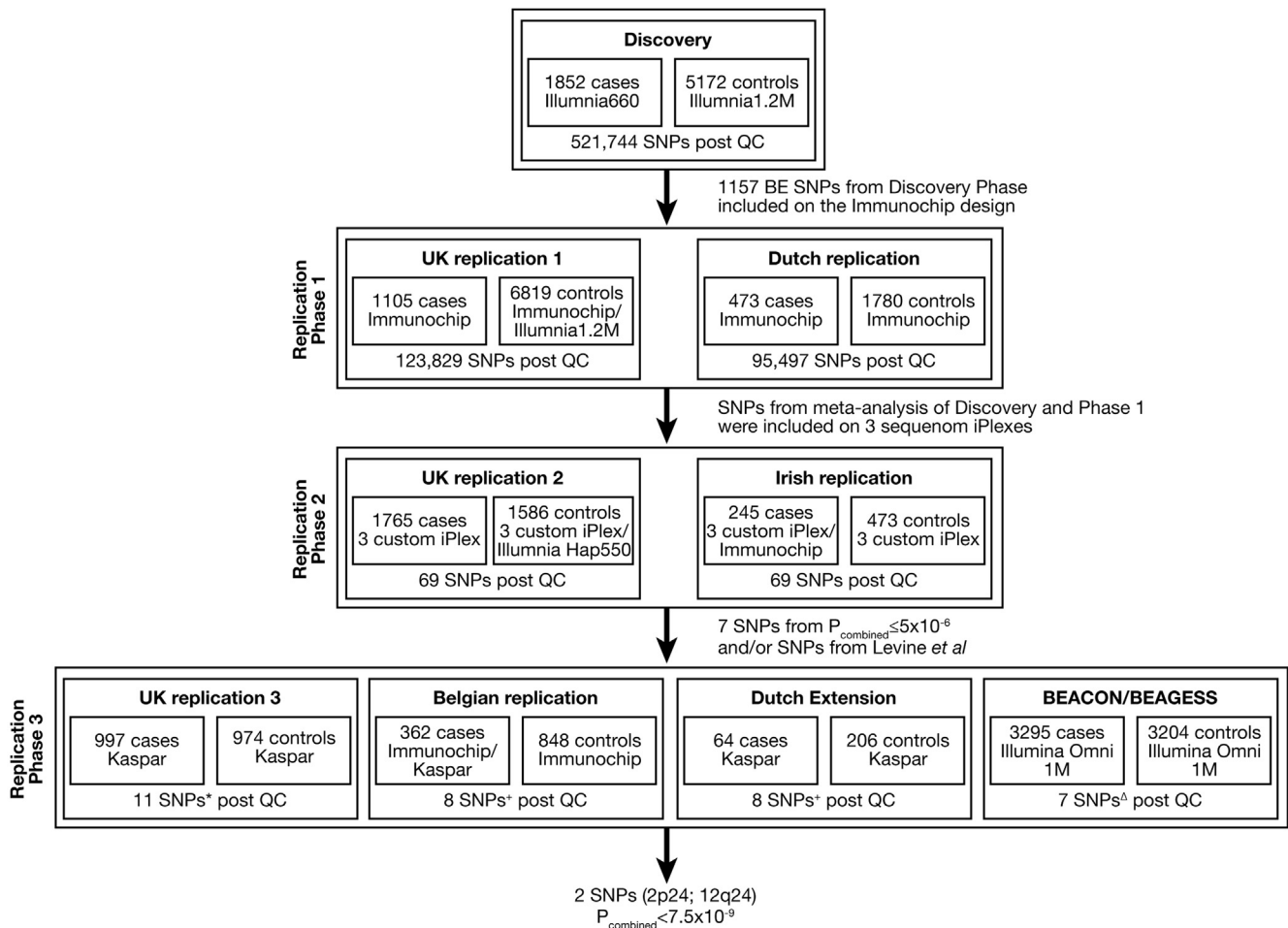
Our GWAS previously identified 2 SNPs, on chromosomes 6p21 (rs9257809;  $P = 4.1 \times 10^{-9}$ ) and 16q24 (rs9936833;  $P = 2.7 \times 10^{-10}$ ), that are associated with BE.<sup>11</sup> One of these loci lies within the HLA region and the other, close to *FOXP1*, which is involved in esophageal structure and development. Both SNPs have also been shown to be associated with risk of EAC.<sup>14</sup> 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*.<sup>12</sup> The last 2 of these genes are known to be involved in esophageal development.<sup>15,16</sup>

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,<sup>11</sup> 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.<sup>17</sup> 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.<sup>1</sup> 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.<sup>18</sup> Patient collection was in accordance with British



**Figure 1.** Outline of the phases of this study and the SNPs analyzed. Two SNPs described in Su *et al*<sup>11</sup> had previously been genotyped in Replication Phase 2 and BEACON/BEAGESS samples. All other replication phase 3 samples are new to this study, as is the genotyping of additional SNPs in phases 2 and 3. Dutch Replication (Phase 1 Replication) and the Dutch extension (Phase 3 Replication) is one cohort in our analyses for the SNPs taken through to Replication Phase 3. \*11 SNPs: Our SNPs: rs3072, rs6751791, rs2731672, rs2701108, rs189247, rs2043633, and rs12985909 and Levine *et al*<sup>12</sup> SNPs: rs1497205, rs254348, rs3784262, and rs4523255. +8 SNPs: Our SNPs: rs3072, rs6751791, rs2731672, rs2701108, rs189247, rs2043633, and Levine *et al* SNP: rs3784262. <sup>Δ</sup>7 SNPs: Our SNPs: rs3072, rs6751791, rs2731672, rs2701108, rs189247, rs2043633.

Society of Gastroenterology criteria for BE<sup>19</sup> 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.<sup>11</sup>

**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<sup>20</sup> (Chief Investigator: Tomlinson), comprising spouses or partners unaffected by cancer and without a family history (to 2<sup>nd</sup>-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, Roynance) with no personal or family history of breast cancer and of white Caucasian origin.<sup>21</sup> 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.<sup>12</sup> All studies received ethical board approval (details in [Supplementary Material](#)). Two SNPs described in Su et al<sup>11</sup> 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.<sup>11</sup> Replication Phase 1 genotyping was performed using the Illumina ImmunoChip at the Wellcome Trust Sanger Institute<sup>11</sup> or as described in Trynka et al.<sup>22</sup>

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,<sup>23</sup> and variants in *IGF1R* and *GHR*<sup>24</sup> ([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.<sup>12</sup>

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.<sup>12</sup> 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.<sup>11</sup> 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.<sup>11</sup> 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.<sup>25</sup> The software tests for heterogeneity of effects between studies<sup>26</sup> 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,<sup>12</sup> we used 1898 white European controls from colorectal cancer GWAS studies CoRGI and Colon Cancer Family Registry.<sup>27</sup> Genotypes were imputed where necessary, with strict cut-offs for imputation quality.<sup>28,29</sup>

### 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).<sup>11</sup> 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<sup>11</sup>). 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 rs3072 (OR = 1.14; 95% CI: 1.09–1.18) and  $7.5 \times 10^{-9}$  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:  $P = 1.3 \times 10^{-9}$ ; OR = 1.13; 95% CI: 1.09–1.17; rs2701108:  $P = 6.2 \times 10^{-8}$ ; 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 *TBX5* 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. Haploregv2<sup>30</sup> and Annovar<sup>31</sup> 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.<sup>32</sup> We examined associations between SNPs in this region and gene expression in The Cancer Genome Atlas (TCGA) EAC data.<sup>33</sup> 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* ( $P_{\text{eQTL}} > .20$ ;  $P_{\text{ASE}} > .38$ ;  $n = 62$ ) and no genome-wide association with expression of any other gene was present ( $q > .05$ , details not shown). In public data sets based on monocytes<sup>34</sup> and on lymphoblastoid cell lines and adipose tissue,<sup>35</sup> *C2orf43* is the suggested target of rs9306894 following eQTL studies (GENovar;  $P = 7 \times 10^{-4}$ ). 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 (NHLF). rs1920562 and an additional SNP (rs1247938) in moderate LD ( $r^2 = 0.52$ ) with rs2701108, are highlighted by Regulome DB 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 ( $P_{\text{eQTL}} > .39$ ;  $P_{\text{ASE}} > .43$ ;  $n = 62$ ), was an eQTL in whole-transcriptome analysis, or was an eQTL in the public databases (details not shown).

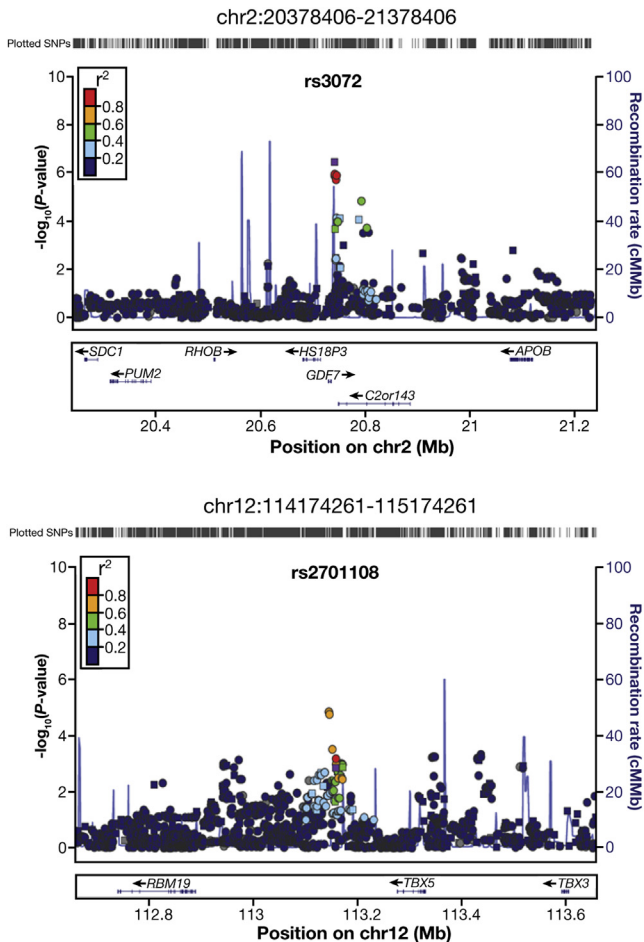
### Pathway/Geneset Enrichment Analysis

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

SNP	Chr	Position (build 37)	Minor/major	Discovery MAF (cases /controls)	Discovery	Replication phase 1 (rep 1)	Meta 1 (discovery + rep 1)	Replication phase 2 (rep 2)	Meta 2 (meta 1+ rep 2)	Replication phase 3 (rep 3)	Final meta (meta 2 + rep 3)	Final meta I <sup>2</sup>	N
<b>rs3072</b>	<b>2</b>	<b>20878406</b>	<b>G/A</b>	<b>0.41/0.36</b>	<b>1.23</b> <b>(1.14–1.33)</b> <b>2.64 × 10<sup>-7</sup></b>	<b>1.10</b> <b>(1.02–1.19)</b> <b>1.33 × 10<sup>-2</sup></b>	<b>1.16</b> <b>(1.10–1.23)</b> <b>8.13 × 10<sup>-8</sup></b>	<b>1.13</b> <b>(1.03–1.24)</b> <b>7.00 × 10<sup>-3</sup></b>	<b>1.16</b> <b>(1.10–1.21)</b> <b>2.27 × 10<sup>-9</sup></b>	<b>1.11</b> <b>(1.04–1.17)</b> <b>1.02 × 10<sup>-3</sup></b>	<b>1.14</b> <b>(1.09–1.18)</b> <b>1.75 × 10<sup>-11</sup></b>	<b>0.42</b>	<b>8</b>
rs6751791	2	35581997	A/G	0.51/0.48	1.15 (1.06–1.23) 5.03 × 10 <sup>-4</sup>	1.16 (1.07–1.25) 1.45 × 10 <sup>-4</sup>	1.15 (1.09–1.21) 2.68 × 10 <sup>-7</sup>	1.07 (0.97–1.16) 1.64 × 10 <sup>-1</sup>	1.13 (1.08–1.18) 2.91 × 10 <sup>-7</sup>	0.99 (0.93–1.05) 7.99 × 10 <sup>-1</sup>	1.08 (1.04–1.12) 7.65 × 10 <sup>-5</sup>	0.60	8
rs2731672	5	176842474	A/G	0.27/0.24	1.18 (1.09–1.28) 1.64 × 10 <sup>-4</sup>	1.14 (1.04–1.24) 3.87 × 10 <sup>-3</sup>	1.16 (1.09–1.23) 2.54 × 10 <sup>-6</sup>	1.09 (0.99–1.21) 8.20 × 10 <sup>-2</sup>	1.14 (1.08–1.20) 8.33 × 10 <sup>-7</sup>	0.95 (0.89–1.02) 1.81 × 10 <sup>-1</sup>	1.07 (1.03–1.12) 1.66 × 10 <sup>-3</sup>	0.63	8
<b>rs2701108</b>	<b>12</b>	<b>114674261</b>	<b>G/A</b>	<b>0.38/0.41</b>	<b>0.88</b> <b>(0.81–0.95)</b> <b>1.00 × 10<sup>-3</sup></b>	<b>0.87</b> <b>(0.81–0.94)</b> <b>4.40 × 10<sup>-4</sup></b>	<b>0.87</b> <b>(0.83–0.92)</b> <b>1.51 × 10<sup>-6</sup></b>	<b>0.89</b> <b>(0.81–0.97)</b> <b>1.10 × 10<sup>-2</sup></b>	<b>0.88</b> <b>(0.84–0.92)</b> <b>5.68 × 10<sup>-8</sup></b>	<b>0.93</b> <b>(0.87–0.99)</b> <b>1.42 × 10<sup>-2</sup></b>	<b>0.90</b> <b>(0.86–0.93)</b> <b>7.48 × 10<sup>-9</sup></b>	<b>0.14</b>	<b>8</b>
rs189247	15	97586630	A/G	0.41/0.37	1.18 (1.09–1.27) 5.67 × 10 <sup>-5</sup>	1.14 (1.05–1.23) 1.25 × 10 <sup>-3</sup>	1.15 (1.09–1.22) 2.91 × 10 <sup>-7</sup>	1.10 (1.00–1.20) 4.90 × 10 <sup>-2</sup>	1.14 (1.09–1.19) 6.12 × 10 <sup>-8</sup>	0.96 (0.90–1.02) 1.73 × 10 <sup>-1</sup>	1.10 (1.06–1.14) 3.55 × 10 <sup>-7</sup>	0.20	8
rs2043633	16	5819274	C/A	0.37/0.41	0.85 (0.79–0.92) 6.04 × 10 <sup>-5</sup>	0.88 (0.82–0.95) 9.80 × 10 <sup>-4</sup>	0.87 (0.82–0.92) 2.49 × 10 <sup>-7</sup>	0.88 (0.80–0.96) 5.00 × 10 <sup>-3</sup>	0.87 (0.83–0.91) 4.74 × 10 <sup>-9</sup>	0.99 (0.94–1.05) 8.39 × 10 <sup>-1</sup>	0.92 (0.88–0.95) 2.25 × 10 <sup>-6</sup>	0.58	8
rs12985909	19	18439383	G/A	0.48/0.45	1.12 (1.04–1.21) 2.94 × 10 <sup>-3</sup>	1.12 (1.04–1.21) 2.73 × 10 <sup>-3</sup>	1.12 (1.06–1.18) 2.45 × 10 <sup>-5</sup>	1.11 (1.01–1.21) 2.70 × 10 <sup>-2</sup>	1.12 (1.07–1.17) 1.99 × 10 <sup>-6</sup>	1.07 (1.01–1.13) 2.63 × 10 <sup>-2</sup>	1.10 (1.06–1.14) 3.28 × 10 <sup>-7</sup>	0.00	8

NOTE. For each phase, association data show (top to bottom) OR, (95% CI), and  $P_{\text{assoc}}$ . 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<sup>2</sup>, I<sup>2</sup> heterogeneity index; MAF, minor allele frequency; N, number of studies.



**Figure 2.** Regional plots of association (left y-axis) and recombination rates (right y-axis) for the chromosomes 2p24 and 12q24 loci after imputation. The lead genotyped SNP is marked with a purple square. Imputed SNPs are plotted as circles and genotyped SNPs as squares.

enriched in cases at False Discovery Rate-corrected  $P < .05$ . Genetic Genomics Analysis of Complex Data (Gengen) did not identify any pathways with corrected  $P < .05$ , but 10 pathways had  $P < .25$  (Supplementary Table 7). Three pathways (type 1 diabetes mellitus, KEGG antigen processing and presentation, and KEGG autoimmune thyroid disease) were identified by all methods and the SNPs mapping to each pathway were subjected to set-based association tests using PLINK, producing empirical  $P$  values of .0021, .025, and .0317, respectively. The SNPs within these pathways that showed replication at  $P < .05$  in Replication Phase 1 all mapped to chromosome 6p21, either close to or within HLA genes. Upon removal of the HLA genes from all the pathways, 20/26 pathways originally with False Discovery Rate  $P < .05$  according to the iGSEA4GWAS approach remained significant, but only 1/34 pathways identified by SNP ratio test remained significant. No HLA-depleted pathways were even suggestive of enrichment (all  $P > .25$ ) by Gengen. The top networks identified by Ingenuity Pathway Analysis were Cardiovascular System Development and Function, Embryonic Development and

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,<sup>11</sup> 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.<sup>13</sup>

### 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- $\beta$ ) where histone marks associated with both promoters and enhancers are present in lymphoblastoid cell lines (LCLs).<sup>30</sup> PBX3, PU1, POL2, YY1, and nuclear factor- $\kappa$ B 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<sup>23</sup> 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<sup>12</sup> 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<sup>12</sup> study overlapped entirely with the controls used in our Discovery phase (WTCCC2 controls) and we therefore used alternative UK controls from the CoRGI 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<sup>12</sup>). 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 KASPar 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) showed  $P_{meta} < 10^{-5}$  and were not strongly correlated with 1 of the 4 BE + EAC SNPs reported previously. We therefore genotyped these 4 SNPs in our Replication Phase samples; rs3784262 (within *ALDH1A2*) was associated with BE + EAC (OR = 0.90; 95% CI: 0.87–0.93;  $P = 3.72 \times 10^{-9}$ ). No SNP was formally associated with BE alone (Table 2). eQTL and ASE analysis (see Results - In Silico Fine Mapping and Annotation of the Chromosome 2p24 and 12q24 Loci) did not show associations for the rs3784262 proxy rs7165247 in TCGA data or other public data sets (details not shown).

## Discussion

We have added 2 new BE predisposition SNPs, rs3072 on chromosome 2p24 and rs2701108 on chromosome 12q24, to the 2 BE SNPs on chromosome 6p21 (HLA region) and chromosome 16q23 (near *FOXF1*) that we reported previously.<sup>11</sup> Both of the new SNPs remained at or very near genome-wide significance when analysis was restricted to cases with intestinal metaplasia. In silico fine mapping

provided evidence that rs3072 and/or 1 of 3 nearby SNPs might be functional because they map to putative enhancer regions. The nearby gene, *GDF7*, is the best functional candidate, because this encodes the BMP12 protein and the BMP pathway has previously been implicated in the development of BE.<sup>36</sup> The importance of this pathway in BE is also suggested in Ingenuity Pathway Analysis, where bone morphogenetic protein 4 acts upstream of proteins encoded by genes close to the BE predisposition SNPs. *GDF7* plays a role in the neural system and tendon/ligament development and repair,<sup>37,38</sup> and also regulates Hedgehog and Wnt signaling pathways that impact on esophageal development through *FOXF1* and *TBX5*. In the chromosome 12q24 region, rs1920562 (the top imputed SNP) provided the strongest association signal and maps to a possible enhancer. Gene expression analysis did not suggest the target of the chromosome 12q24 variation, although *TBX5* is a very strong functional candidate. It is involved in cardiac development and its deficiency causes thoracic malformations and abnormalities of the diaphragmatic musculature,<sup>39,40</sup> which could predispose patients to hiatus hernia and acid reflux, 2 subphenotypes of BE.

Messenger RNA expression analysis using TCGA EAC data and public data from leukocytes and adipocytes provided little 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.<sup>41–44</sup> 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.<sup>45–47</sup> 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<sup>12</sup> 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



**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<sup>12</sup>

SNP	Chr	Position	Nearby genes	Minor/ major allele	BE+/ -EAC	Levine et al meta OR (95% CI)	$P_{Levine\ et\ al}$	This study meta OR (95% CI)	$P_{This\ Study}$	This study + Levine et al <sup>12</sup> meta OR (95% CI)	$P_{meta}$	Meta $I^2$	No. of studies
rs2687201	3	70928930	FOXP1	T/G	BE	1.18 (1.10–1.26)	$2.00 \times 10^{-6}$	1.14 (1.03–1.27)	$1.18 \times 10^{-2}$	1.16 (1.10–1.23)	$4.61 \times 10^{-8}$	0.00	3
					BE/EAC	1.18 (1.12–1.25)	$5.47 \times 10^{-9}$			1.17 (1.11–1.23)	$6.70 \times 10^{-10}$	0.00	3
rs11789015	9	96716028	BARX1	G/A	BE	0.85 (0.79–0.91)	$5.08 \times 10^{-6}$	0.90 (0.81–1.01)	$6.63 \times 10^{-2}$	0.86 (0.81–0.92)	$1.38 \times 10^{-6}$	0.00	3
					BE/EAC	0.83 (0.79–0.88)	$1.02 \times 10^{-9}$			0.85 (0.81–0.89)	$1.14 \times 10^{-10}$	0.00	3
rs10419226	19	18803172	CRTC1	A/C	BE	1.19 (1.12–1.26)	$5.54 \times 10^{-8}$	1.01 (0.91–1.11)	$8.65 \times 10^{-1}$	1.13 (1.08–1.20)	$2.14 \times 10^{-6}$	0.82	3
					BE/EAC	1.18 (1.12–1.24)	$3.55 \times 10^{-10}$			1.14 (1.09–1.19)	$1.17 \times 10^{-8}$	0.82	3
rs10423674	19	18817903	CRTC1	T/G	BE	0.85 (0.80–0.91)	$1.92 \times 10^{-6}$	0.94 (0.88–1.00)	$4.88 \times 10^{-2}$	0.89 (0.85–0.93)	$2.99 \times 10^{-7}$	0.40	5
					BE/EAC	0.84 (0.80–0.89)	$1.75 \times 10^{-9}$			0.88 (0.84–0.91)	$4.87 \times 10^{-11}$	0.49	5
rs1497205	4	76169067	PARM1, RCHY1	C/T	BE	0.86 (0.80–0.92)	$2.86 \times 10^{-5}$	0.92 (0.87–0.98)	$7.59 \times 10^{-1}$	0.90 (0.86–0.94)	$2.57 \times 10^{-6}$	0.00	6
					BE/EAC	0.87 (0.82–0.93)	$1.28 \times 10^{-5}$			0.90 (0.86–0.94)	$3.68 \times 10^{-7}$	0.00	6
rs254348	16	65980789		T/C	BE	0.88 (0.83–0.94)	$1.15 \times 10^{-4}$	0.95 (0.91–1.01)	$8.88 \times 10^{-2}$	0.93 (0.89–0.97)	$5.49 \times 10^{-4}$	0.53	6
					BE/EAC	0.89 (0.84–0.94)	$1.40 \times 10^{-5}$			0.92 (0.89–0.96)	$2.81 \times 10^{-5}$	0.53	6
rs3784262	15	58253106	ALDH1A2	G/A	BE	0.85 (0.80–0.90)	$3.62 \times 10^{-7}$	0.93 (0.89–0.98)	$5.13 \times 10^{-3}$	0.91 (0.87–0.94)	$1.37 \times 10^{-6}$	0.12	9
					BE/EAC	0.88 (0.83–0.92)	$6.72 \times 10^{-7}$			0.90 (0.87–0.93)	$3.72 \times 10^{-9}$	0.16	9
rs4523255	8	8713038	MFHAS1	A/G	BE	1.13 (1.06–1.21)	$2.46 \times 10^{-4}$	1.07 (1.01–1.12)	$2.11 \times 10^{-2}$	1.09 (1.05–1.14)	$2.48 \times 10^{-5}$	0.36	6
					BE/EAC	1.13 (1.07–1.20)	$4.15 \times 10^{-5}$			1.09 (1.05–1.14)	$9.24 \times 10^{-6}$	0.46	6

NOTE. The minimum meta-analysis comprised the Levine et al<sup>12</sup> 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.<sup>48</sup> Of the candidate SNPs we assessed (Supplementary Table 8), we found supporting evidence, albeit short of genome-wide significance, for rs909253 ( $P = 3.1 \times 10^{-4}$ ), 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.<sup>11</sup> Ek et al<sup>13</sup> 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.<sup>49</sup>

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.<sup>1,17</sup>

## Supplementary Material

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

## References

- Jankowski J, Barr H, Wang K, et al. Diagnosis and management of Barrett's oesophagus. *BMJ* 2010; 341:c4551.
- Jankowski J, Odze R. Biomarkers in gastroenterology; between hype and hope comes histopathology. *Am J Gastroenterol* 2009;104:1093–1096.
- Duggan C, Onstad L, Hardikar S, et al. Association between markers of obesity and progression from Barrett's esophagus to esophageal adenocarcinoma. *Clin Gastroenterol Hepatol* 2013;11:934–943.
- Iyer PG, Borah BJ, Heien HC, et al. Association of Barrett's esophagus with type II diabetes mellitus: results from a large population-based case-control study. *Clin Gastroenterol Hepatol* 2013;11:1108–1114.e5.
- DeVault K, McMahon BP, Celebi A, et al. Defining esophageal landmarks, gastroesophageal reflux disease, and Barrett's esophagus. *Ann N Y Acad Sci* 2013; 1300:278–295.
- Cameron AJ, Lagergren J, Henriksson C, et al. Gastroesophageal reflux disease in monozygotic and dizygotic twins. *Gastroenterology* 2002;122:55–59.
- Mohammed I, Cherkas LF, Riley SA, et al. Genetic influences in gastro-oesophageal reflux disease: a twin study. *Gut* 2003;52:1085–1089.
- Lembo A, Zaman M, Jones M, et al. Influence of genetics on irritable bowel syndrome, gastro-oesophageal reflux and dyspepsia: a twin study. *Aliment Pharmacol Ther* 2007;25:1343–1350.
- Groves C, Jankowski J, Barker F, et al. A family history of Barrett's oesophagus: another risk factor? *Scand J Gastroenterol* 2005;40:1127–1128.
- Chak A, Ochs-Balcom H, Falk G, et al. Familiality in Barrett's esophagus, adenocarcinoma of the esophagus, and adenocarcinoma of the gastroesophageal junction. *Cancer Epidemiol Biomarkers Prev* 2006;15:1668–1673.
- Su Z, Gay LJ, Strange A, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. *Nat Genet* 2012;44:1131–1136.
- Levine DM, Ek WE, Zhang R, et al. A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. *Nat Genet* 2013;45:1487–1493.
- Ek WE, Levine DM, D'Amato M, et al. Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's esophagus, and gastroesophageal reflux. *J Natl Cancer Inst* 2013;105:1711–1718.
- Dura P, van Veen EM, Salomon J, et al. Barrett associated MHC and FOXF1 variants also increase esophageal carcinoma risk. *Int J Cancer* 2013;133:1751–1755.
- Woo J, Miletich I, Kim BM, et al. Barx1-mediated inhibition of Wnt signaling in the mouse thoracic foregut controls tracheo-esophageal septation and epithelial differentiation. *PLoS One* 2011;6:e22493.
- Shu W, Lu MM, Zhang Y, et al. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* 2007;134:1991–2000.
- Jankowski J, Barr H. Improving surveillance for Barrett's oesophagus: AspECT and BOSS trials provide an evidence base. *BMJ* 2006;332:1512.
- Sharma P, Dent J, Armstrong D, et al. The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. *Gastroenterology* 2006;131:1392–1399.

19. Bennett C, Vakil N, Bergman J, et al. Consensus statements for management of Barrett's dysplasia and early-stage esophageal adenocarcinoma, based on a Delphi process (BADCAT). *Gastroenterology* 2012;143:336–346.
20. Houlston RS, Cheadle J, Dobbins SE, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet* 2010;42:973–977.
21. Petridis C, Shinomiya I, Kohut K, et al. Germline *CDH1* mutations in bilateral lobular carcinoma in situ. *Br J Cancer* 2013;100:1053–1057.
22. **Trynka G, Hunt KA**, Bockett NA, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011;43:1193–1201.
23. Orloff M, Peterson C, He X, et al. Germline mutations in *MSR1*, *ASCC1*, and *CTHRC1* in patients with Barrett esophagus and esophageal adenocarcinoma. *JAMA* 2011;306:410–419.
24. McElholm AR, McKnight AJ, Patterson CC, et al. A population-based study of IGF axis polymorphisms and the esophageal inflammation, metaplasia, adenocarcinoma sequence. *Gastroenterology* 2010;139:204–212.
25. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinform* 2010;11:288.
26. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–560.
27. Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near *CDKN1A*, *POLD3* and *SHROOM2* influences colorectal cancer risk. *Nat Genet* 2012;44:770–776.
28. Delaneau O, **Zagury JF, Marchini J**. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013;10:5–6.
29. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
30. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–D934.
31. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
32. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–1797.
33. Dulak AM, Stojanov P, Peng S, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* 2013;45:478–486.
34. Zeller T, Wild P, Szymczak S, et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010;5:e10693.
35. Nica AC, Parts L, Glass D, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* 2011;7:e1002003.
36. Castillo D, Puig S, Iglesias M, et al. Activation of the *BMP4* pathway and early expression of *CDX2* characterize non-specialized columnar metaplasia in a human model of Barrett's esophagus. *J Gastrointest Surg* 2012;16:227–237; discussion 237.
37. Berasi SP, Varadarajan U, Archambault J, et al. Divergent activities of osteogenic *BMP2*, and tenogenic *BMP12* and *BMP13* independent of receptor binding affinities. *Growth Factors* 2011;29:128–139.
38. Currie DS, Cheng X, Hsu CM, et al. Direct and indirect roles of CNS dorsal midline cells in choroid plexus epithelia formation. *Development* 2005;132:3549–3559.
39. Arora R, Metzger RJ, Papaioannou VE. Multiple roles and interactions of *Tbx4* and *Tbx5* in development of the respiratory system. *PLoS Genet* 2012;8:e1002866.
40. Mori AD, Bruneau BG. *TBX5* mutations and congenital heart disease: Holt-Oram syndrome revealed. *Curr Opin Cardiol* 2004;19:211–215.
41. Pomerantz MM, Ahmadiyeh N, Jia L, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with *MYC* in colorectal cancer. *Nat Genet* 2009;41:882–884.
42. Tuupainen S, Turunen M, Lehtonen R, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 2009;41:885–890.
43. Ling H, Spizzo R, Atlasi Y, et al. *CCAT2*, a novel non-coding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res* 2013;23:1446–1461.
44. Takatsuno Y, Mimori K, Yamamoto K, et al. The rs6983267 SNP is associated with *MYC* transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. *Ann Surg Oncol* 2013;20:1395–1402.
45. Mahlapuu M, Enerback S, Carlsson P. Haploinsufficiency of the forkhead gene *Foxf1*, a target for sonic hedgehog signaling, causes lung and foregut malformations. *Development* 2001;128:2397–2406.
46. Shaw-Smith C. Genetic factors in esophageal atresia, tracheo-esophageal fistula and the VACTERL association: roles for *FOXF1* and the 16q24.1 *FOX* transcription factor gene cluster, and review of the literature. *Eur J Med Genet* 2010;53:6–13.
47. Wang DH, Tiwari A, Kim ME, et al. Hedgehog signaling regulates *FOXA2* in esophageal embryogenesis and Barrett's metaplasia. *J Clin Invest* 2014;10:1172–1178.
48. Gyamfi MA, Kocsis MG, He L, et al. The role of retinoid X receptor alpha in regulating alcohol metabolism. *J Pharmacol Exp Ther* 2006;319:360–368.
49. Speed D, Hemani G, Johnson MR, et al. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* 2012;91:1011–1021.

---

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, AGAF, FAGG, 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,

Oxford, England, OX3 7BN, United Kingdom. e-mail: [cpalles@well.ox.ac.uk](mailto:cpalles@well.ox.ac.uk); fax: +44(0)1865287501; or Ian Tomlinson, BM, BCh, MA, PhD, Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, England, OX3 7BN, United Kingdom. e-mail: [iant@well.ox.ac.uk](mailto:iant@well.ox.ac.uk); fax: +44(0)1865287501.

#### 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 JJ2), 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.