Integrative post-genome-wide association analysis of CDKN2A and TP53 SNPs and risk of esophageal adenocarcinoma

Matthew E.Buas1,2,6, David M.Levin1, Karen W.Makar2, Heidi Utsugi2, Lynn Onstad3, Xiaohong Li4, Patricia C.Galipeau5, Nicholas J.Shaheen6, Laura J.Hardie6, Yvonne Romero7,8, Leslie Bernstein9, Marilie D.Gammon10, Alan G.Casson11, Nigel C.Bird12, Harvey A.Risch13, Weimian Ye14, Geoffrey Liu15, Douglas A.Corley16,17, Patricia L.Blount1, Rebecca C.Fitzgerald18, David C.Whiteman19, Anna H.Wu20, Brian J.Reid21,4 and Thomas L.Vaughan12

1Department of Epidemiology, University of Washington, School of Public Health, Seattle, WA 98109, USA, 2Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, 3Department of Biostatistics, University of Washington, School of Public Health, Seattle, WA 98109, USA, 4Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, 5Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA, 6Division of Epidemiology, University of Leeds, Leeds LS2 9JT, UK, 7Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN 55905, USA, 8The Romero Registry, Mayo Clinic, Rochester, MN 55905, USA, 9Department of Populations Sciences, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, CA 91010, USA, 10Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, NC 27599, USA, 11Department of Surgery, University of Saskatchewan, Saskatoon S7N 5E5, Canada, 12Department of Oncology, Medical School, University of Sheffield, Sheffield S10 2RX UK, 13Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT 06510, USA 14Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 171 77, Sweden, 15Pharmacogenomic Epidemiology, Ontario Cancer Institute, Toronto, Ontario M5G 2M9, Canada, 16Division of Research, Kaiser Permanente Northern California, Oakland, CA, 94612 USA, 17San Francisco Medical Center, Kaiser Permanente Northern California, San Francisco, CA 94115 USA, 18Medical Research Council (MRC) Cancer Cell Unit, Hutchison-MRC Research Centre and University of Cambridge, Cambridge CB2 0XZ UK, 19Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane Queensland 4006, Australia and 20Department of Preventive Medicine, University of Southern California/ Norris Comprehensive Cancer Center, Los Angeles, CA 90033, USA

*To whom correspondence should be addressed. Tel: +206 667 2927; Fax: +206 667 4787; Email: mbuas@fhcrc.org
Correspondence may also be addressed to Thomas L. Vaughan. Tel: 206 667 5134; Fax: 206 667 4787; Email: tvbaughan@u.washington.edu

Incidence of esophageal adenocarcinoma (EA) in Western countries has increased markedly in recent decades. Although several risk factors have been identified for EA and its precursor, Barrett’s esophagus (BE), including reflux, Caucasian race, male gender, obesity, and smoking, less is known about the role of inherited genetic variation. Frequent somatic mutations in the tumor suppressor genes TP53 and CDKN2A were recently reported in EA tumors, while additional putative risk alleles (19–21) may nonetheless be important factors in modifying disease risk, if, e.g. their main effects were operative only in certain subgroups of the overall population.

To capitalize further on the availability of GWAS data from the International Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON), we have been conducting a series of targeted post-GWAS analyses to assess potential associations between genetic variation within prespecified genes or pathways and risk of EA. While various genome-wide significant findings have been reported previously (16), they remain likely that a substantial number of additional single nucleotide polymorphisms (SNPs) that did not satisfy the highly stringent (Bonferroni) statistical threshold (P < 5 x 10−8) may nonetheless be important factors in modifying disease risk, if, e.g. their main effects were operative only in certain subgroups of the overall population. Targeted reuse of existing GWAS data is one approach for identifying additional putative risk alleles (17,18).

On a genomic level, the emergence of EA is a complex process in which chromosomai instability and specific somatic chromosomal alterations (SCAs) appear to act as key drivers of neoplastic progression (19–21). An extensive body of work has implicated the tumor suppressors CDKN2A and TP53 in EA pathogenesis (22–28). Based on findings from the Seattle Barrett’s Esophagus Project, a longitudinal cohort study of >400 BE patients, Reid and colleagues have demonstrated that loss of heterozygosity (LOH) at 9p (CDKN2A locus) and 17p (TP53 locus) at baseline endoscopy are significantly associated with risk of progression from BE to EA (29,30). When combined with tetraploidy and aneuploidy, these two lesions (9p/17p LOH)

Abbreviations: BE, Barrett’s esophagus; BEACON, Barrett’s and Esophageal Adenocarcinoma Consortium; BMI, body mass index; EA, esophageal adenocarcinoma; GWAS, genome-wide association study; LOH, loss of heterozygosity; PCA, principal component analysis; SCA, somatic chromosomal alteration; SNPs, single nucleotide polymorphism; UTR, untranslated region.

P = 0.0186, q = 0.3059, and rs4074785 C>T (intronic, odds ratio 0.85, P = 0.0248, q = 0.3059). None of the TP53 single nucleotide polymorphisms reached nominal significance. Two of the CDKN2A variants identified were also associated with reduced risk of progression from BE to EA, when assessed in a prospective cohort of 408 BE patients: rs2518720 (hazard ratio 0.57, P = 0.0095, q = 0.0285) and rs3088440 (hazard ratio 0.34, P = 0.0368, q = 0.0552). In vitro functional studies of rs3088440, a single nucleotide polymorphism located in the seed sequence of a predicted miR-663b binding site, suggested a mechanism whereby the G>A substitution may attenuate mir-663b-mediated repression of the CDKN2A transcript. This study provides the first evidence that germline variation at the CDKN2A locus may influence EA susceptibility.

Introduction

Esophageal adenocarcinoma (EA) is a rare but often lethal disease that represents a growing public health problem (1). EA typically arises from a squamous-to-columnar metaplastic precursor lesion known as Barrett’s esophagus (BE (2)). While symptomatic gastroesophageal reflux disease, Caucasian race, male gender, obesity, and smoking are established risk factors for EA and BE (3), the role of inherited genetic variation and its interplay with environmental factors remains an active area of investigation. Initial small-scale individual studies based on candidate-gene approaches linked altered risk of EA or BE to DNA polymorphisms in genes implicated in a wide range of biological pathways: inflammation (PTGS2, IL-18, TNF-β), detoxification (GSTM1, GSTT1, GSTP1, NQO1), DNA repair (MGMT), angiogenesis (VEGF, MMP1, MMP3), and apoptosis (CASP7, CASP9) (4–14). Interactions have also been described between smoking and variants of GSTM1, GSTT1, and VEGF (8,12), and between gastroesophageal reflux disease and variants of MMP1 and MMP3 (13), in relation to risk of EA. More recently, large-scale consortium-based genome-wide association studies (GWAS) of BE and EA identified several polymorphisms significantly associated with disease risk (15,16). These alterations included variants located in three transcription factors (BARX1, FOXP1, FOXP1), a transcriptional co-activator (CRTC1), and the major histocompatibility complex locus.

To capitalize further on the availability of GWAS data from the International Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON), we have been conducting a series of targeted post-GWAS analyses to assess potential associations between genetic variation within prespecified genes or pathways and risk of EA. While various genome-wide significant findings have been reported previously (16), they remain likely that a substantial number of additional single nucleotide polymorphisms (SNPs) that did not satisfy the highly stringent (Bonferroni) statistical threshold (P < 5 x 10−8) may nonetheless be important factors in modifying disease risk, if, e.g. their main effects were operative only in certain subgroups of the overall population. Targeted reuse of existing GWAS data is one approach for identifying additional putative risk alleles (17,18).

On a genomic level, the emergence of EA is a complex process in which chromosomal instability and specific somatic chromosomal alterations (SCAs) appear to act as key drivers of neoplastic progression (19–21). An extensive body of work has implicated the tumor suppressors CDKN2A and TP53 in EA pathogenesis (22–28). Based on findings from the Seattle Barrett’s Esophagus Project, a longitudinal cohort study of >400 BE patients, Reid and colleagues have demonstrated that loss of heterozygosity (LOH) at 9p (CDKN2A locus) and 17p (TP53 locus) at baseline endoscopy are significantly associated with risk of progression from BE to EA (29,30). When combined with tetraploidy and aneuploidy, these two lesions (9p/17p LOH)
were found to strongly predict progression to EA (RR = 38.7, 95% CI 10.8–138.5 (30)). EA is believed to develop within genomically altered fields (or regions) of metaplastic BE epithelial cells (26,31). Early loss of one or both alleles of CDKN2A generally precedes subsequent TP53 mutation and LOH, leading to the evolution of tetraploid and aneuploid cell populations of increasing size and diversity, which ultimately give rise to cancer (32–36). A recent whole-genome sequencing analysis of ~150 EA tumor-normal pairs complemented these foundational studies and reported that CDKN2A and TP53 were the two most significantly mutated genes in EA, among a pool of 26 mutated genes identified [false-discovery rate (FDR) q < 0.1 (37)].

As somatic alterations at the CDKN2A and TP53 genomic loci in the setting of BE appear to be important events along the pathway to cancer, we hypothesized that inherited genetic variants within these two genes may be associated with altered risk of EA. In this study, using data from our consortium-based GWAS conducted on 2515 EA cases and 3207 controls (16), we selected all 37 successfully genotyped SNPs located within or in proximity to (± 2kb) CDKN2A (24 SNPs) or TP53 (13 SNPs) and assessed their association with risk of EA.

Subjects and methods

Study population and SNP genotyping

The Barrett’s and Esophageal Adenocarcinoma Genetic Susceptibility Study included individuals with EA, with BE, and control subjects pooled from 14 individual studies conducted in Western Europe, Australia, and North America over the past 20 years. Detailed study population characteristics and genotyping protocols have been described previously (16). Briefly, all EA and BE cases were confirmed by histologic examination, and a set of population controls was drawn from the included BEACON studies to serve as a comparison group for both EA and BE subjects. Data for EA/BE risk factors such as age, sex, race, body mass index (BMI), and smoking history were collected by all of the included studies via standardized questionnaires, usually through personal interviews. BE and EA cases also were included from the UK Barrett’s Esophagus Gene Study and the UK Stomach and Oesophageal Cancer Study (SOCS), respectively (16). To increase statistical power, additional controls (MD Anderson controls) were selected from a hospital-based case-control study of melanoma (38), and included cancer-free friends or acquaintances, of white European ancestry, who had accompanied patients with melanoma to their clinic visits at the MD Anderson Cancer Center in Houston, Texas. Genotyping of buffy coat or whole blood DNA from all participants was conducted using the Illumina Omni1M Quad platform, in accordance with standard quality control procedures (39). All recruited participants gave written informed consent, and this project was approved by the ethics review board of the Fred Hutchinson Cancer Research Center (FHCRC).

The primary analyses in this study employed a pooled dataset comprised of all EA case and control subjects drawn from these multiple sources, with secondary analyses also encompassing available BE subjects (Supplementary Table S1, available at Carcinogenesis Online). To define a homogeneous set of European ancestry samples, principal component analysis (PCA) was run using SNPRelate (15). Selection of unrelated subjects with a missing genotyping call rate < 2%, the final study population for our primary analyses included 2515 EA cases and 3207 controls. An additional 3295 BE cases were also assessed using the same control group, except for the exclusion of three control subjects due to familial relation to cases. Subsequent analyses employed a prospective cohort of 413 BE patients (primarily of European ancestry) from the Seattle Barrett’s Esophagus Project (40). After exclusion of five subjects with no follow-up visits, 55 of the remaining 408 subjects developed EA over a mean follow-up time of 7.85 person-years (SD = 5.26). None of the EA cases that developed in the cohort members were included in the GWA analyses of EA.

SNP selection

SNPs analyzed in this study are located within the chromosomal regions defined by the HUGO (41) gene boundaries for CDKN2A and TP53, or within additional 2-kb flanking regions located both upstream and downstream. We excluded from consideration SNPs that failed Illumina quality measures or standard quality control procedures (39). Specifically, SNPs were excluded if any of the following criteria were satisfied: (i) Illumina GenTrain score < 0.6 or cluster separation < 0.4; (ii) > 5% missing call rate over samples; (iii) discordant genotype calls in any pair of duplicate study samples; (iv) Mendelian error in either one of the HapMap QC trios or a small number of families identified in the BEACON data; (v) significant departure from Hardy–Weinberg equilibrium (P < 10−6); (vi) minor allele frequency < 1%. After imposing the above filters, 37 Omni1M SNPs were available for analysis (Supplementary Table S2, available at Carcinogenesis Online). Minor and major alleles were reported throughout using the Illumina ‘design’ strand designation.

Statistical analysis

Unconditional multivariate logistic regression was used to compute odds ratios (OR) for risk of EA (or BE) associated with a given SNP variant, using an additive model (per-allele), while adjusting for age, sex, and the first four eigenvectors (ev) derived from PCA to account for population stratification by ancestry. We conducted using the software package EIGENSOFT. To correct for multiple comparisons when assessing statistical significance, we used the Bonferroni and Benjamini–Hochberg FDR methods (42). Stratified analyses by smoking history and BMI were conducted for nominal hits (P < 0.05) identified in the primary analysis, and evidence for interaction was assessed by including a product term in the logistic regression model. Smoking history and BMI were defined categorically in stratified analyses (smoking: ever/never, or pack-years: 0, >0 and <15, 15–29, 30–44, 45+; BMI: <25, 25–29, 30–34, 35+) or continuously (pack-years, BMI) to test for interaction. These selected SNPs were also assessed in the Seattle BE cohort for association with risk of progression from BE to EA. Cox proportional hazards regression was conducted to calculate hazard ratios for cancer progression associated with selected SNP variants, while adjusting for age at baseline and sex. The proportional hazards assumption was verified via goodness-of-fit testing based on Schoenfeld residuals.

The relationship between germline CDKN2A genotype and levels of SCAs detected at baseline esophageal biopsy was also assessed, using a pool of 248 subjects from the Seattle BE Project (21). Levels of total SCA (Mb) in each biopsy sample (including homozygous deletions, copy loss, copy neutral loss of heterozygosity, balanced copy gain of both alleles, and allele-specific copy gain) were quantified as a metric of whole-genome integrity. Welch’s t-tests were conducted to assess whether mean total SCA (or the log of this value) at baseline endoscopy differed significantly by germline genotype. When fewer than five subjects were homozygous for a given minor allele, these subjects were combined with heterozygotes to form a single exposure category. Total SCA at baseline was defined as the maximum level of SCA detected in any of multiple biopsies collected at this initial time point. All statistical analyses were conducted using STATA/SE version 13.1 (College Station, TX).

Functional analysis of the CDKN2A polymorphism rs3088440 G>A

Multiple bioinformatic algorithms (TargetScan (43), miRANDA (44), Patrocles (45)) were used to predict which miRNAs are likely to target the CDKN2A 3’ untranslated region (UTR), and to prioritize these putative miRNA regulators according to the proximity of their predicted 3’UTR binding sites and the rs3088440 polymorphism. Luciferase reporter assays were conducted in 293T cells. The HEK293T cell line was obtained from Dr. Muneech Tewari (Fred Hutchinson CRC) at the outset of this study, expanded, and frozen in multiple aliquots. To ensure low passage, a new aliquot was thawed approximately every 2–4 weeks for the duration of the study (6 months). Cell line authentication was performed on the original cell passage obtained and a frozen stock used at the conclusion of the study, with identical STR profiles obtained. The cell lines were tested using the CODIS technology with 15 loci and Amelogenin (AmpFLSTR Identifier kit
from Applied Biosystems, Inc.). The STR profile was >85% identical to the 293T STR profile reported by the ATCC (ATCC-CRL-3126; www.atcc.org). Reporter plasmids containing either the ‘wildtype’ (rs3088440 G) or polymorphic (rs3088440 A) sequence of the CDKN2A 3’UTR downstream of Renilla luciferase were cotransfected with synthetic miR-663b or scrambled control miRNA, and pGL3-Firefly luciferase (Promega, Madison, WI). A negative control reporter containing no 3’UTR (Empty), and a positive control reporter containing six multimerized consensus miR-663b binding sites 3’ to the luciferase open reading frame (6xMRE), were included in all experiments. Renilla plasmid and miRNA reagents were obtained from SwitchGear Genomics (Carlsbad, CA). Cultures were transfected at ~70% confluence and maintained for 48 h prior to harvesting of lysates for analysis, using the Dual-Luciferase Reporter Assay System (Promega, Madison, WI), according to the manufacturer’s protocols. Three replicate wells were included for each condition. Renilla luciferase activity was normalized to firefly luciferase (pGL3-luc).

Results

Subject characteristics

The distributions of demographic and behavioral characteristics among subjects included in the primary analysis are shown in Table I. EA cases were older (mean 65.0 years) and more likely to be male (87.3%) relative to controls (mean 58.4 years, 72.6% male). The percentage of subjects reporting ever having smoked cigarettes was higher among EA cases (75.3%) compared with controls (59.1%), as was the prevalence of heavy smoking (45+ pack years: 21.2% versus 14.3%) and obesity (BMI ≥30: 29.7% versus 20.2%).

Association of individual SNPs with risk of EA

Of the 37 SNPs selected at the CDKN2A (n = 24) or TP53 (n = 13) gene loci, three were nominally associated (P < 0.05) with reduced risk of EA (Figure 1, Table II; Supplementary Table S3, available at Carcinogenesis Online). All three of these variants are located within the CDKN2A locus—two are intronic, and the third resides in the 3’UTR. The per-allele ORs for these three polymorphisms ranged in magnitude from 0.84 to 0.90: rs2518720 C>T (OR 0.90, P = 0.0121, q = 0.3059), rs3088440 G>A (OR 0.84, P = 0.0186, q = 0.3059), and rs4074785 C>T (OR 0.85, P = 0.0248, q = 0.3059). After correction for multiple comparisons (n = 37), these three variants would be designated as significant when allowing for a FDR of ≥0.31, without reaching the more stringent Bonferroni statistical threshold (α = 0.05/37 = 0.0014). In a separate analysis of BE subjects versus the same controls (demographics described in Supplementary Table S4 is available at Carcinogenesis Online), none of these identified SNPs reached nominal significance (Supplementary Table S5, available at Carcinogenesis Online).

Based on a subset of study subjects (Supplementary Table S1, available at Carcinogenesis Online: BEAGESS) for whom additional behavioral and demographic risk factors had been ascertained, stratified analyses were conducted to determine whether the associations between CDKN2A rs2518720, rs3088440, or rs4074785 and risk of EA were modified by smoking history (Supplementary Table S6A, B, available at Carcinogenesis Online) or BMI (Supplementary Table S6C, available at Carcinogenesis Online). No relationship was found with smoking history. In the four (relatively small) strata defined by BMI, a suggestion of the observed inverse association of rs3088440 G>A was apparent only among overweight or obese subjects. Similar results were observed for rs4074785 C>T. A borderline-significant interaction was observed between rs4074785 and BMI (P = 0.0588).

Association of individual SNPs with risk of progression from BE to EA

Using data obtained from 408 subjects enrolled in the Seattle BE cohort, we next evaluated whether any of the three SNPs found to be associated with risk of EA (P < 0.05) in our post-GWAS analysis were also associated with altered risk of progression from BE to cancer (Table III). Two of these three variants were observed to be associated with reduced risk of disease progression: rs2518720 (HR 0.57, P = 0.0095, q = 0.0285) and rs3088440 (HR 0.34, P = 0.0368, q = 0.0552). The third SNP reached borderline significance (rs4074785: HR 0.42, P = 0.0567, q = 0.0567).

Functional characterization of rs3088440

We elected to pursue functional follow-up studies of rs3088440 G>A, as bioinformatic analysis revealed that this 3’UTR polymorphism resides within, and may disrupt, a predicted binding site for the small noncoding RNA molecule miR-663b (Figure 2, upper panel). SNPs located within miRNA binding sites can alter the efficiency of miRNA:miRNA pairing, leading to changes in miRNA-mediated regulation of target transcripts, and many such variants have been linked to altered risk of cancer (46). Using standard luciferase reporter assays in cultured cells, we sought to ascertain (i) whether miR-663b can target the CDKN2A 3’UTR in vitro, and (ii) whether rs3088440 G>A may attenuate miRNA-mediated repression, if present (Figure 2, lower panel). We first demonstrated that treatment with synthetic miR-663b, relative to scrambled miRNA, resulted in ~35% repression of a positive control template (6xMRE) comprised of the Renilla luciferase open reading frame linked to six multimerized 3’UTR miR-663b binding sites, but only 10% repression of a negative control luciferase template lacking a 3’UTR (P = 0.135). Modest repression (~20%) by synthetic miR-663b (versus scrambled) was also suggested for our ‘wild-type template’, comprised of luciferase linked with the CDKN2A 3’UTR (rs3088440 G) (P = 0.297). By contrast, an analogous reporter containing the rs3088440 G>A variant showed no evidence of repression by miR-663b relative to scrambled control miRNA.

Discussion

In the last two decades, numerous studies have pointed to chromosome instability as a key driver of neoplastic progression (47,48). Studies of BE and EA, in particular, have revealed multiple chromosomal
abnormalities that are strongly linked to risk of cancer development, and have specifically implicated the tumor suppressors CDKN2A and TP53 as important regulators of disease pathogenesis (3,19). These two genes were also found to be the most significantly mutated in recent sequencing-based analyses of EA tumor tissue (37). While this work and related studies have greatly enhanced our understanding of somatic genomic changes that underlie the emergence of EA (49), uncertainty remains as to what extent the initiation and evolution of disease may be influenced or modified by inherited genetic variation. Using data from a recent consortium-based GWAS (16), we assessed the association of 37 SNPs at the CDKN2A and TP53 gene loci and risk of EA. Three CDKN2A variants were found to be nominally associated with reduced risk of EA ($P < 0.05$, $q < 0.31$). Two of the polymorphisms were also associated with decreased risk of progression from BE to EA ($P < 0.05$, $q < 0.06$). These FDR estimates are likely conservative, given high levels of linkage disequilibrium (LD) reported for several of the CDKN2A SNPs analyzed. Functional follow-up studies in vitro were consistent with a model in which one of these variants, rs3088440 G>A, may result in impaired miRNA-mediated repression of CDKN2A transcripts.

Genetic polymorphisms and mutations in the tumor suppressor CDKN2A have been studied previously in the context of multiple case-control studies of several cancers. The rs3088440 G>A polymorphism, in particular, has been examined in studies of melanoma, a cancer that has been linked to lesions at chromosome 9p21 and the CDKN2A locus (50). Findings have been inconsistent—while two studies reported statistically significantly higher frequencies of rs3088440 heterozygotes and the variant (A) allele among melanoma cases (51,52), four other studies were unable to replicate this result (53–56). The reasons for such discrepancies are uncertain, but considerable variation existed with respect to sample sizes, ethnicity of the study populations, and case selection. Adding to the ambiguity, separate reports have described associations between rs3088440 G>A and either improved survival (57) or shorter progression time from primary to metastatic disease (58). In separate work, no evidence of an association was found between rs3088440 and risk of head and neck squamous cell carcinoma (59).

The CDKN2A locus generates several distinct transcripts (60), the most prominent of which are p16INK4a and p14ARF. These transcripts contain unique first exons (located ~20 kb apart at 9p21.3), and shared second and third exons. Our study examined 24 SNPs distributed across the ~30-kb CDKN2A locus that were genotyped on the Omni 1M Illumina BeadChip and passed stringent quality control filters. Of the three variants identified as nominal hits, two are located in intron 1 of the p14ARF transcript (upstream of p16INK4a exon 1), and one resides in the (shared) 3′UTR of p16INK4a/p14ARF. Examination of linkage disequilibrium (LD) in the region based on data from the 1000 Genomes Project revealed strong LD between rs3088440 and risk of head and neck squamous cell carcinoma (59).

A critical observation was that the risk estimate for rs3088440 was much stronger than that of the second variant studied, rs4074785. These two variants are in strong LD with each other ($r^2 = 0.88$), but much weaker LD between rs2518720 and the other two variants (rs2518720/r2 < 0.25). These findings, coupled with the highly similar risk estimates for rs3088440 and rs4074785, are consistent with a single haplotype underlying these two association signals, and a second potential signal represented by rs2518720.

Assuming that the observed associations are not due to chance, uncertainty remains as to whether rs3088440, rs4074785, and rs2518720 represent actual ‘causative’ SNPs, or simply are in LD with other presumably causal polymorphisms. While the location of rs3088440 within a predicted miRNA binding site of the CDKN2A 3′UTR immediately suggested possible functional effects (discussed below), hypothetical consequences of the two intronic SNPs were less apparent. Intronic SNPs can exert diverse functional effects depending upon genomic location and context (61,62). Variants near or within intronic enhancer elements may influence gene transcription levels. Alternatively, SNPs in proximity to splice junctions or associated regulatory sequences can act downstream of transcription and affect alternative splicing. In silico analyses based on chromatin state

**Fig. 1.** Regional association plots for genotyped SNPs at the CDKN2A (left) and TP53 (right) gene loci. The top-ranked SNP at each locus is shown as a solid purple circle. SNPs are ordered by genomic location. The color scheme indicates LD between the top-ranked SNP and other SNPs in the region using $r^2$ values calculated from the 1000 Genomes Project. The y axis shows $-\log_{10}(P)$ values computed from 2515 EA cases and 3207 controls. Dashed red lines indicate the threshold for (nominal) statistical significance ($P < 0.05$). Available SNP annotations are depicted as follows: square (untranslated region), triangle (nonsynonymous), circle (none).

**Table II.** CDKN2A SNPs and risk of EA

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Pos</th>
<th>Alleles</th>
<th>Controls</th>
<th></th>
<th>EA</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>MAF$^b$</td>
<td>N</td>
<td>MAF$^b$</td>
<td>OR$^c$ 95% CI</td>
</tr>
<tr>
<td>rs2518720</td>
<td>9</td>
<td>21978979</td>
<td>T/C</td>
<td>3205</td>
<td></td>
<td>2494</td>
<td>0.332</td>
<td>0.90</td>
<td>(0.83–0.98)</td>
<td>0.0121 0.3059</td>
</tr>
<tr>
<td>rs3088440</td>
<td>9</td>
<td>21968159</td>
<td>A/G</td>
<td>3190</td>
<td></td>
<td>2483</td>
<td>0.083</td>
<td>0.84</td>
<td>(0.73–0.97)</td>
<td>0.0186 0.3059</td>
</tr>
<tr>
<td>rs4074785</td>
<td>9</td>
<td>21981583</td>
<td>T/C</td>
<td>3206</td>
<td></td>
<td>2495</td>
<td>0.083</td>
<td>0.85</td>
<td>(0.74–0.98)</td>
<td>0.0248 0.3059</td>
</tr>
</tbody>
</table>

$^a$Minor/major alleles, $^b$minor allele frequency, $^c$odds ratio, adjusted for age, sex, ev1–ev4 using additive model (per-allele), $^d$false discovery rate (FDR).
Table III. CDKN2A SNPs and risk of progression from BE to EA

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Pos</th>
<th>Alleles</th>
<th>No. minor alleles</th>
<th>No. Minor Alleles</th>
<th>HR (95% CI)</th>
<th>P</th>
<th>q</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2518720</td>
<td>9</td>
<td>21978979</td>
<td>T/C</td>
<td>149</td>
<td>152</td>
<td>22</td>
<td>0.57</td>
<td>0.0095</td>
</tr>
<tr>
<td>rs3088440</td>
<td>9</td>
<td>21968159</td>
<td>A/G</td>
<td>283</td>
<td>64</td>
<td>2</td>
<td>0.34</td>
<td>0.0368</td>
</tr>
<tr>
<td>rs4074785</td>
<td>9</td>
<td>21981583</td>
<td>T/C</td>
<td>286</td>
<td>64</td>
<td>2</td>
<td>0.42</td>
<td>0.0567</td>
</tr>
</tbody>
</table>

Minor/major alleles, *g/*m, adjusted for age and sex using additive model (per-allele); false discovery rate (FDR).

Fig. 2. Functional characterization of CDKN2A rs3088440 G>A. The rs3088440 G>A polymorphism (bold) lies within a predicted miRNA-663b binding site in the CDKN2A 3′ UTR (upper panel). Alteration of G(C) to A(T) at this position disrupts the perfect complementarity indicated between the miR-663b seed sequence (5′ nucleotides 2–8) and the CDKN2A mRNA. Luciferase reporter assays were conducted in 293T cells using synthetic miR-663b or scrambled control miRNA (lower panel). Reporter constructs included the Renilla luciferase open reading frame linked to one of four different 3′UTR sequences: none (Empty), an artificial UTR with six multimerized consensus miR-663b response elements (6xMRE), or the full-length CDKN2A 3′ UTR containing either the [G] or [A] rs3088440 variant. Renilla luciferase activity was normalized to firefly luciferase activity in each well, and all conditions were repeated in triplicate. Percent repression by miR-663b relative to miR-scrambled for each indicated reporter was calculated using normalized luciferase activity values as follows: % repression = (miRsc − miR663-b)/(miRsc) × 100. The values presented are averages from two independent experiments, with error bars depicting standard deviations.

segmentation data from the NIH Roadmap Epigenome Atlas indicated that rs2518720 and rs4074785 are both located within an 11.8-kb chromatin segment characterized as ‘weakly repressed’ in esophageal tissue (Supplementary Figure S1, available at Carcinogenesis Online). Since these SNPs did not appear to alter specific DNA binding sites, their potential to directly influence CDKN2A transcription remains uncertain. The location of these variants in the central region of a ~22-kb intron, far from known splice junctions, would also not appear to support a likely role in splicing. Interestingly, however, rs2518720 is in high LD (r² = 0.93) with a SNP 4 kb downstream (rs3814960), located ~100 bp distal to the p16INK4a transcriptional start site. This variant is situated within a 1200-bp chromatin segment classified as ‘transcriptionally-poised’, likely marking the p16INK4a promoter region. Analyses using the Ensembl Variant Effect Predictor (63) and the JASPAR database (64) indicated that rs3814960 lies within a probable 14-bp binding site for the transcription factor EGR1. The functional potential of this linked variant makes it a plausible biological candidate underlying the rs2518720 association.

Since BE is an established risk factor for and the only known precursor lesion of EA, and recent studies have reported significant shared heritability between these conditions (65), we sought to investigate the extent to which variants associated with altered risk of EA could also be linked to BE. In theory, a SNP causally associated with reduced
risk of EA could act by reducing either the risk of BE, or the risk of progression from BE to EA, or both (though less plausible, such a polymorphism could also be associated with opposing effects on risk at these two phases in disease development, if the magnitude of risk reduction at one phase exceeded the magnitude of risk increase at the other). Of the three SNPs found to be associated (\(p < 0.05\)) with reduced risk of EA, none was associated with risk of BE, while two of the three SNPs (rs2518720 and rs3088440) exhibited inverse associations with risk of progression from BE to EA, and the third was borderline significant (Supplementary Table S7, available at Carcinogenesis Online). These findings suggest that, if causally related to cancer risk, the three SNPs identified appear more likely to modulate disease risk at the level of BE progression versus BE development. If validated in larger external study populations, these polymorphisms warrant further evaluation for possible clinical value in the setting of BE risk stratification. Nonsignificant ORs obtained from the BE analysis for rs3088440 G>A (OR = 0.88) and rs4074785 C>T (OR = 0.91) were similar in magnitude and direction to those in the EA analysis, though somewhat closer to the null, suggesting possible weak associations with reduced risk of BE as well. Given our use of a single control group for comparison with both the EA and BE case groups, however, a certain number of shared associations could also be expected from chance alone.

The positioning of the rs3088440 G>A variant within a predicted miR-663b binding site of the CDKN2A 3′UTR suggested the potential for effects on CDKN2A expression levels, and prompted us to pursue functional follow-up studies of this SNP. While allelic expression quantitative trait locus mapping previously reported an association between the G>A polymorphism and an 8% reduction in CDKN2A expression (66), these studies were limited to assessments of RNA in peripheral blood. Using luciferase reporter assays in cultured cells, we sought to assess experimentally whether this variant could be linked to altered expression levels of CDKN2A, which in turn might underlie the observed inverse association with risk of EA. The results presented suggest a scenario in which the CDKN2A transcript may be susceptible to repression by miR-663b in rs3088440[G]-dependent manner. Under such a scheme, the G>A polymorphism would impair binding of miR-663b to the CDKN2A 3′UTR and potentially result in enhanced production of the p16/p14 proteins. The evidence for regulation of CDKN2A by miR-663b, however, remains speculative. First, while repression of the CDKN2A-3′UTR reporter was ~2-fold stronger than that observed using the empty vector control, this disparity in inhibition levels did not reach statistical significance, raising the possibility of nonspecific regulation. Second, luciferase assays employ artificial reporter plasmids and nonphysiologic conditions, making it uncertain whether endogenous levels of miR-663b, if present in esophageal epithelium, would similarly inhibit expression of native CDKN2A transcripts. Nevertheless, rs3088440-sensitive regulation of CDKN2A by miR-663b remains a plausible explanatory hypothesis for further investigation. Of interest, studies implicating miR-663a, formerly ‘miR-663’, in the regulation of p21WAF1/CIP1 and TGFβ1 have suggested pro-oncogenic functions for this related miRNA family member (67,68) (expression of which appears upregulated in esophageal adenocarcinoma versus normal tissue (69)), although anti-proliferative effects have also been described (70). Given the imperfect nature of miRNA-binding site prediction algorithms, miRNAs other than miR-663b may also be capable of targeting the CDKN2A 3′UTR in a manner dependent upon rs3088440 status.

Recent studies have indicated that somatic alterations spanning the CDKN2A locus appear to be relatively frequent in BE patients—both in those who progress and those who do not (21). A prediction implicit in our current study is that among germline rs3088440 heterozygotes (A/G) with somatic 9p LOH, retention of the potentially protective minor allele (A) would be associated with reduced progression risk, relative to retention of the major allele (G). This hypothesis, while not assessed here due to the small number of heterozygous individuals in the Seattle BE cohort, warrants further evaluation.

The p16 protein has an established function in cell cycle checkpoint control by binding to CDK4/CDK6 and blocking the action of cyclin D, but additional roles for this tumor suppressor have been described in the maintenance of centrosome function and genome integrity (71). Intriguingly, the CDKN2A rs3088440 G>A polymorphism, in particular, has recently been linked to reduced risk of chromosomal damage associated with occupational exposure to an industrial chemical carcinogen (72). The extent of whole-genome integrity can be quantified by measuring the total megabases of SCA—defined as homozygous deletions, copy loss, copy neutral LOH, copy gain—which are present in a genome. Recent studies from the Seattle BE cohort revealed that BE patients who progressed to EA had increased levels of total SCA at baseline biopsy, relative to nonprogressors (21). Given these findings, we sought to assess whether the CDKN2A rs3088440 G>A variant, which exhibited an inverse association with risk of EA, might be linked to reduced baseline levels of SCA in BE subjects. Our results are suggestive of such a relationship (Supplementary Figure S2, Table S8, available at Carcinogenesis Online) but were limited by highly skewed SCA distributions and small sample sizes (particularly for the AG/AA genotypes), underscoring the need for additional validation studies with larger numbers of subjects.

Our analyses did not uncover any significant interactions between two established EA risk factors (smoking and obesity) and the top SNPs identified, but we did observe the suggestion of a trend towards more pronounced inverse associations for rs4074785 and rs3088440 among overweight/obese subjects. The precise mechanisms which may underlie reported associations between obesity (and specifically visceral adiposity) and increased risk of EA have not been firmly established, but speculation has focused on the possible role of increased inflammatory signaling, oxidative stress and associated DNA damage (73). Under this model, it appears intriguing that the strongest apparently protective associations for rs3088440 and rs4074785 were observed in subjects with a presumably elevated ‘inflammatory burden’. At the same time, such a trend was not readily seen among the heaviest smokers. Replication studies using larger samples within the indicated strata of BMI are warranted.

Strengths of our study include the use of high-quality genotyping data from the largest published study to date of inherited genetic variation in relation to risk of EA and BE. All included samples were interrogated using the same array platform and were subjected to identical quality control procedures. Inclusion of both case types allowed us to examine whether genetic variants associated with EA were similarly associated with the precursor lesion that gives rise to this cancer. Longitudinal cohort data enabled us to make inferences about constitutive genetic factors that may contribute to the propensity of selected BE patients to progress or not to EA. Despite extensive characterization of the presence and predictive value of somatic alterations at the CDKN2A and TP53 gene loci during EA development, the role of inherited genetic variation in these regions has not been widely investigated.

A primary limitation of our study is statistical power, a familiar challenge in investigations of infrequent diseases such as EA. While both the EA/BE GWAS and the BE cohort employed in our analyses represent the largest studies of their kind currently available, these sample sizes nevertheless limited our ability to conclusively assess associations of modest magnitude, or of those present exclusively in subgroups of the overall population (particularly in the setting of multiple testing). Our null findings for TP53 SNPs, in particular, must be interpreted in this context. It remains to be determined whether our reported associations will meet more stringent statistical thresholds when evaluated in larger study populations. A second potential limitation relates to the absence of (i) matched EA/BE cases for the MD Anderson controls and (ii) matched controls for the UK EA/BE cases. These subjects were not included either because they do not exist (MD Anderson) or were unavailable for genotyping (UK). While we cannot entirely dismiss the resulting concern regarding comparability between included cases and controls, we believe that by restricting subjects to ‘genetically European’ individuals via PCA, and further controlling for population stratification by ancestry through inclusion of PCA-derived eigenvectors in our logistic regression models, our analysis has taken reasonable precautions to ensure that cases and controls from different populations are as comparable as possible. Apart from these concerns, the scope of our analysis was restricted to 37 polymorphisms included on the
Illumina Omni1M array. While these SNPs were selected to provide a high level of coverage of estimated variation within the regions of interest, additional genotyping, sequencing, or imputation in future studies could help further delineate the apparent associations identified.

Our assessment of 37 SNPs marks the most comprehensive examination of constitutive polymorphisms at these two key tumor suppressor loci in relation to risk of EA. While chance findings cannot be ruled out, the results of this study indicate that inherited variants within the CDKN2A gene may be linked to reduced risk of EA, and raise the intriguing possibility that rs3088440 G>A may lead to compromised miRNA-based regulation of the CDKN2A transcript. Further replication in independent study populations, coupled with additional experimental validation of the effects of the identified polymorphisms on gene expression and/or putative miRNA binding, will be needed to substantiate these variants as true modifiers of disease risk.

Supplementary material
Supplementary Tables 1–8 and Figures 1–2 can be found at http://carcin.oxfordjournals.org/.

Funding
National Institutes of Health (R01CA136725 to T.L.V. and D.C.W, T32CA009168 to T.L.V., and K05CA124911 to T.L.V.).

Acknowledgments
We thank Terri Watson, Tricia Christopherson, and Paul Hansen for their contributions in project management and organization of biospecimens/data, and Jessica Arnaudo, David Cowan, and Carissa Sanchez for assistance with data retrieval. We thank Drs. Wong-Ho Chow (MD Anderson Cancer Center, Houston, TX) and Olof Nyren for their contributions to the BEACON consortium, and Liam Murray (Queen’s University Belfast, UK) for access to GWAS data derived from subjects enrolled in his past studies of EA and BE (FINBAR Study). Genotyping data for MD Anderson controls (38) was obtained from dbGaP through accession number phs000187.v1.p1. Additional funding sources for individual studies included in the BEACON GWAS, and for BEACON investigators, have been acknowledged (16). The authors disclose no potential conflicts of interest.

Conception and design: M.F.B. and T.L.V. Subject recruitment: L.B., N.C.B., A.G.C., D.A.C., R.C.F., M.D.G., L.J.H., G.L., P.L.B., H.A.R., B.J.R., Y.R., D.C.W., and T.L.V. was obtained from dbGaP through accession number phs000187.v1.p1. Additional funding sources for individual studies included in the BEACON GWAS, and for BEACON investigators, have been acknowledged (16). The authors disclose no potential conflicts of interest.

Conflict of Interest Statement: None declared.

References

Received June 12, 2014; revised August 28, 2014; accepted September 19, 2014