Genomic regions associated with susceptibility to Barrett's esophagus and esophageal adenocarcinoma in African Americans: The cross BETRNet admixture study

Jean S. Wang  
*Washington University School of Medicine in St. Louis*  
Ilke Nalbantoglu  
*Washington University School of Medicine in St. Louis*  
et al.

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)  
Please let us know how this document benefits you.

**Recommended Citation**  
[https://digitalcommons.wustl.edu/open_access_pubs/6563](https://digitalcommons.wustl.edu/open_access_pubs/6563)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Genomic regions associated with susceptibility to Barrett’s esophagus and esophageal adenocarcinoma in African Americans: The cross BETRNet admixture study


1 Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, United States of America, 2 Division of Gastroenterology and Hepatology, University Hospitals Cleveland Medical Center, Case Western Reserve University School of Medicine, Cleveland, OH, United States of America, 3 Division of Gastroenterology and Hepatology, Johns Hopkins Medical Institutions, Baltimore, MD, United States of America, 4 Department of Gastroenterology and Hepatology, Cleveland Clinic, Cleveland, OH, United States of America, 5 Department of Cardiology and Thoracic Surgery, Johns Hopkins Medical Institutions, Baltimore, MD, United States of America, 6 Center for Esophageal Diseases & Swallowing, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, United States of America, 7 Thoracic Surgery, Department of Surgery, University of Michigan, Ann Arbor, MI, United States of America, 8 Division of Gastroenterology, Washington University School of Medicine, St Louis, MO, United States of America, 9 University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States of America, 10 Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, United States of America, 11 Department of Medicine, Columbia University Medical Center, New York, NY, United States of America, 12 Division of General Medical Sciences (Oncology), Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, United States of America, 13 Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, United States of America, 14 Department of Pathology, University Hospitals Case Medical Center, Case Western Reserve University School of Medicine, Cleveland, OH, United States of America, 15 Department of Pathology, Cleveland Clinic, Cleveland, OH, United States of America, 16 Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, United States of America, 17 Division of Oncology and Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, United States of America

* axc22@cwru.edu

Abstract

Background
Barrett’s esophagus (BE) and esophageal adenocarcinoma (EAC) are far more prevalent in European Americans than in African Americans. Hypothesizing that this racial disparity in prevalence might represent a genetic susceptibility, we used an admixture mapping approach to interrogate disease association with genomic differences between European and African ancestry.

Methods
Formalin fixed paraffin embedded samples were identified from 54 African Americans with BE or EAC through review of surgical pathology databases at participating Barrett’s...
Esophagus Translational Research Network (BETRNet) institutions. DNA was extracted from normal tissue, and genotyped on the Illumina OmniQuad SNP chip. Case-only admixture mapping analysis was performed on the data from both all 54 cases and also on a subset of 28 cases with high genotyping quality. Haplotype phases were inferred with Beagle 3.3.2, and local African and European ancestries were inferred with SABER plus. Disease association was tested by estimating and testing excess European ancestry and contrasting it to excess African ancestry.

Results

Both datasets, the 54 cases and the 28 cases, identified two admixture regions. An association of excess European ancestry on chromosome 11p reached a 5% genome-wide significance threshold, corresponding to $-\log_{10}(P) = 4.28$. A second peak on chromosome 8q reached $-\log_{10}(P) = 2.73$. The converse analysis examining excess African ancestry found no genetic regions with significant excess African ancestry associated with BE and EAC. On average, the regions on chromosomes 8q and 11p showed excess European ancestry of 15% and 20%, respectively.

Conclusions

Chromosomal regions on 11p15 and 8q22-24 are associated with excess European ancestry in African Americans with BE and EAC. Because GWAS have not reported any variants in these two regions, low frequency and/or rare disease associated variants that confer susceptibility to developing BE and EAC may be driving the observed European ancestry association evidence.

Introduction

Esophageal adenocarcinoma (EAC) is the seventh leading cause of death in U.S. males [1] and one of the deadliest cancers worldwide, with 5-year survival rates lower than 20% [2], and in the United States the incidence rate has increased dramatically up to 7-fold over the past three decades [3,4]. EAC tends to arise from Barrett’s esophagus (BE), which replaces the squamous epithelium with columnar-lined metaplastic epithelium in the lower esophagus during healing reflux esophagitis and may progress to dysplasia [5]. BE and EAC are rare in African Americans (AA), with the majority of cases occurring in European Americans (EA) [6–12]. Although EAC occurs at least five-fold more frequently in EAs than AAs [13], the distribution of known risk factors for BE and EAC (e.g. GERD [14], obesity [15,16], etc) are at least as common in AA as EA, suggesting another basis for the racial differences.

Hypothesizing that this racial/ethnic difference in prevalence might represent a genetic susceptibility, we used an admixture mapping approach to interrogate disease association with genetic differences in European and African ancestry using a multi-institutional sample of AA patients with EAC or BE. Although our sample size is limited because of the rareness of BE in AA patients, we demonstrate here—the first and the only admixture mapping study of BE and EAC we know of—how this approach can efficiently find regions of genetic susceptibility for this disease.
Materials and methods

Samples and genotyping

The study was approved by the Institutional Review Board of the University Hospitals of Cleveland. Formalin fixed paraffin embedded (FFPE) samples were identified from 54 AA patients with EAC or BE through review of surgical pathology databases at eight participating BETRNet institutions. The samples were coded and had no identifying information. The codes and identifying information were kept secure at each individual institution and were only available to the institutional investigator who identified the subject, and were not available to any other research personnel. Germline DNA was extracted from normal tissue of the samples, generally from squamous epithelium; in 9 cases the germline DNA was obtained from other non-neoplastic tissues (5 gastric, 2 colonic, 1 vocal cord, and 1 adipose tissue biopsies). DNA samples were genotyped on the Illumina OmniQuad SNP chip. We considered BE and EAC to be part of the same trait, theorizing that at least a proportion of EAC arose from BE.

Of the 54 patient samples, 30 had a genotyping call rate > 0.95 and, after excluding 2 samples because of very low heterozygosity rates (outside the range of 3 standard deviations from the mean heterozygosity), we formed an enriched sample set of 28 samples.

There were 2.3 million SNPs genotyped, but upon additional processing following instructions from Illumina technologists, and restricting SNPs to call rates greater than 96.4%, 378,711 SNPs were available for data analysis.

For the admixture mapping study, the two ancestry populations were selected from the 1000 genome project [17], which included genotype data on 87 Europeans (CEU) and 88 Africans (YRI). After further filtering SNPs that could not match with SNPs in the 1000 genome data, or had no variation in our sample, we finally had 289,112 SNPs on the 22 autosomal chromosomes for admixture mapping.

Admixture mapping analysis

We performed admixture mapping analysis using all the 54 EAC/BE cases as well as using just the cleaned samples comprising 28 cases.

Phases of the 54 and 28 cases were respectively inferred by Beagle 3.3.2 [18], then local African and European ancestries for each dataset were estimated with SABER [19].

Because the dataset of 54 EAC cases has a lower sample call rate than the dataset of 28 cases, we first compared the allele frequencies in the 54 individuals with those estimated from the 1000 genome data (estimated as 0.8 freqYRI+0.2 freqCEU) for the same allele at each SNP. Among the 289,112 autosomal SNPs, 908 SNPs had an allele frequency difference > 0.3. For the dataset of 54 individuals, we did two admixture mapping analyses: one using all the 289,112 SNPs, and the other excluding these 908 SNPs.

Test of excess European or African ancestry

The incidence of esophageal adenocarcinomas in EAs is five times higher than that in AAs [13,20,21], and so any excess of European ancestry at a SNP could be detecting a region harboring susceptibility variants for EAC and BE. Conversely, such a region would not show excess African ancestry. Excess European and African ancestries were estimated and tested by the following method [22].

Assume we have a total of N individuals and L marker loci.

For the i-th individual at marker locus l, (i = 1, 2, ..., N; l = 1, 2, ..., L), let q_{il} be the estimated European ancestry.
The excess European ancestry is estimated as $\Delta \Pi_i = \frac{1}{N} \sum_{i=1}^{N} (q_{il} - M_i)$, where $M_i = \frac{1}{L} \sum_{l=1}^{L} q_{il}$.

Then the Z score test statistic is defined as $Z_i = \frac{\Delta \Pi_i}{SD(\Delta \Pi_i)}$, where $SD(\Delta \Pi_i)$ is the common standard deviation of $\Delta \Pi_i$ estimated over all the markers.

In view of the large numbers, $Z_i$ follows a standard normal distribution, and the test for excess European ancestry was conducted at each locus, using a right tailed test. The test of excess African ancestry, that is, the association of esophageal adenocarcinomas being due to African ancestry, was conducted by using a left tailed test with the same statistic.

Estimating the number of independent tests

The number of independent tests was estimated by the eigenvalue method using the local ancestry of each chromosome [23] and summed up over the 22 autosomal chromosomes. That is, for each chromosome, we calculated the eigenvalues of the correlation matrix of the local ancestry of SNPs on that chromosome, and the effective number of independent tests for that chromosome was estimated as $M = \frac{\sum_{i=1}^{n} \sqrt{\lambda_i}}{\sum_{i=1}^{n}}$, where $\lambda_i$ is the $i$-th eigenvalue and $n$ is the total number of eigenvalues for that chromosome.

Results

There is no significant difference on demographic statistics between the whole samples of 54 individuals and the subset of 28 individuals (Table 1). In addition, the whole sample of 54 individuals has mean proportion (± standard error) of European ancestry 0.32±0.20, and the subset of 28 individuals has mean proportion of European ancestry 0.27±0.14; both are close to estimates of African ancestry in the literature [24,25].

The number of independent tests

By the eigenvalue method, the number of independent tests for 54 individuals estimated using all SNPs is 288.3; after excluding the 908 SNPs it is 281.9. The number of independent tests for the 28 individuals using all SNPs is 252.2. Thus, the corresponding 5% genome-wide thresholds of association tests are respectively for the three scenarios 0.000173 (corresponding to $-\log_{10}(P) = 3.76$), 0.000177 (corresponding to $-\log_{10}(P) = 3.75$), and 0.000198 (corresponding to $-\log_{10}(P) = 3.70$).

| Table 1. Demographic statistics of the admixed population of EAC/BE patients. |
|---------------------------------|----------------|----------------|
|                                 | All samples   | Subset         | P    |
| **Number of individuals**       | 54            | 28             |      |
| **Diagnosis (%)**               |               |                |      |
| EAC                             | 40 (74.1%)    | 21 (75.0%)     | 1*   |
| BE                              | 14 (25.9%)    | 7 (25.0%)      |      |
| **Gender (%)**                  |               |                | 0.35*|
| Male                            | 35 (64.8%)    | 16 (57.1%)     |      |
| Female                          | 19 (35.2%)    | 12 (42.9%)     |      |

* Chi square test of association between the data of 28 individuals and the remaining data of 26 (= 54–28) individuals.

https://doi.org/10.1371/journal.pone.0184962.t001
Excess of European ancestry

The results of genome-wide excess of European ancestry among the AA EAC/BE patients are shown in S1–S3 Figs. From all the 54 patients, removing the 908 SNPs that have allele frequency difference $> 0.3$ did not decrease the admixture mapping signal (S1–S3 Figs; Figs 1 and 2). The highest excess of European ancestry signals that are in common for the 54- and 28-patient datasets are on chromosomes 11 and 8 (Figs 1 and 2). From the 28 patients, the maximum association signal ($-\log_{10}(P)$) on chromosomes 8 and 11 are respectively 2.73 and 4.28. Because the genome-wide significance threshold for this dataset is 3.70, the signal on chromosome 11 reached genome-wide significance (Fig 2). From the 54 patients without the 908 SNPs, the maximum association signals ($-\log_{10}(P)$) on chromosomes 8 and 11 are respectively 3.32 and 3.51, which do not reach the genome-wide significance threshold of 3.75.

The region of excess European ancestry on chromosome 8 is from 90.4 Mb to 129 Mb, with the peak at 106.3 Mb, and the excess of European ancestry at the peak is 0.15 (S1 File). The region of excess European ancestry on chromosome 11 is from 5.6 Mb to 42.4 Mb, with the two peaks at 24.8 Mb and 32.5 Mb respectively, and the maximum mean excesses of European ancestry are 0.18 and 0.2 (S1 File).
Excess African ancestry

No association due to excess African ancestry reached any genome-wide significance threshold for any of the three datasets; the strongest associations due to African ancestry are on chromosomes 15 and 16, but they are far from reaching genome-wide significance level: the association P-values for the three datasets are all $> 0.001$ (S4–S6 Figs).

It is noteworthy that, compared with the association due to European ancestry, fewer and lower association peaks due to African ancestry are found (compare S1–S3 Figs with S4–S6 Figs).

Discussion

Admixture mapping is an effective tool for discovering genetic regions associated with disease. The method interrogates a recently admixed population such as AAs for genetic associations to diseases when prevalence differs markedly between the ancestral Caucasian and African populations [26–30]. Admixture mapping is particularly more powerful than GWAS in detecting regions that harbor rare or multiple different disease variants [31]. Admixture studies are cost effective, require less dense maps, and compared to association studies, they do not assume a disease model [31]. The identification of ancestry informative markers (AIMS) enables such approaches to be successful in diseases such as BE and EAC, whose prevalence are vastly disparate between AAs and EAs [32–38]. In this study, we performed a genome-wide case only admixture mapping association study of BE and EAC using 54 BE/EAC cases and a subset of 28 cases with high genotyping quality. We identified two chromosome regions with excess European ancestry in both datasets; one on chromosome 11p15, which reached genome-wide significance in the 28 cases, and a second one on chromosome 8q22-24. It is not surprising that the results from the two datasets are consistent, but it does suggest that the signals are not caused by genotype errors. We did not find any significant chromosome region with excess African ancestry. Because no common variants have been reported in these two regions in the genome wide association studies of BE/EAC so far, the current result indicates that these two regions with excess European ancestry likely harbor low frequency and/or rare disease associated variants that confer susceptibility to developing BE and EAC.

The fact that there are more and higher admixture mapping peaks of excess European ancestry than of excess African ancestry is consistent with the higher prevalence of BE and EAC in European Americans than in African Americans. Although not conclusive, this is certainly suggestive.

At the admixture peak on chromosome 8, we have a mean excess European ancestry of 0.15; and at the admixture peak on chromosome 11, we have a higher mean excess European ancestry of 0.2. Assuming an additive model for the genetic variants, the estimated relative risks for European alleles in the chromosome 8q and 11p regions would be 1.93 and 2.81, respectively (S1 File). However, because of the large number of regions examined, these are over-estimates due to the winner’s curse.

The admixture mapping region on 8q22-24 overlaps a linkage region that we previously identified in a linkage study [39]. It also overlaps with the 8q24 region, which is associated with multiple epithelial cancers [40–47]. This region is known to contain tissue specific enhancers that drive c-MYC expression in colon cancers [42,48,49]. Moreover, there are other known cancer susceptibility genes located in the two admixture mapping regions we identified. Metadherin (MTDH), also known as Astrocyte elevated gene-1 (AEG-1), locates 7 Mb from the peak of the chromosome 8 region. It plays an important role in carcinogenesis [50], and overexpression of MTDH/AEG-1 can significantly enhance cell proliferation and anchorage-independent growth ability in a variety of cancers [51]. The Wilms’ tumor gene WT1...
locates at the rightmost (the highest) peak on chromosome 11; it is a tumor-suppressor gene [52] that encodes a zinc finger transcription factor regulating transcription of growth factors such as PDGF-A [53], growth factor receptor (IGF-1R) [54] and other genes (RAR-α, c-myc and bcl-2) [55,56]. Myogenic differentiation 1 MYOD1 is 7 Mb to the left of the peak on chromosome 11, and was reported to have frequent hypermethylation in intestinal metaplasia tissue (Barrett’s esophagus) [57]. The oncogene MYC [58] is close to our chromosome 8 admixture mapping region and, because the location of genetic variants for BE/EAC in our study could be inaccurately estimated due to the small sample size, this is another candidate gene for further fine mapping study. The admixture mapping regions on the two chromosomes are large, with more than 200 genes in the chromosome 8 region and more than 300 genes in the chromosome 11 region, making it is hard to locate the susceptibility genes in this kind of study. Moreover, the susceptibility variants could be miRNAs, other non-coding RNAs, or regulatory elements in the regions. The identification of the susceptibility genes and variants will require a large-scale accrual of cases at multiple centers since AA EAC cases are rare.

The results of this study indicate the power of the admixture approach for racially disparate diseases such as BE and EAC. The case only admixture mapping study is a more efficient design for a rare disease such as EAC than case-control design [26] and, because it is based on comparing local ancestry with average local or global ancestry [27, 59], it is robust against any population stratification difference between cases and controls. The rarity of BE and EAC limited our available sample size. Furthermore, the quality of the fragmented DNA extracted from archived FFPE blocks resulted in poor quality SNP calls for nearly half our available samples. Due to the limitations of the small sample size and the DNA quality, our admixture results could miss other regions and the case only admixture mapping could not accurately locate association variants. Despite these limitations, the admixture analysis was able to identify two regions of excess European ancestry that appear to be associated with BE and EAC. It is especially noteworthy that poor sample quality, which would seriously affect association results, has less of a detrimental effect on admixture mapping (Figs 1 and 2); this can be explained by the fact that in admixture mapping larger regions are studied, and within such regions positive and negative errors tend to cancel out. This, of course, comes at the expense of a less precise location of any causal variant.

In conclusion, our admixture mapping association study of BE and EAC identified chromosome 8q22-24 and chromosome 11p15 with excess European ancestry. This is the first admixture analysis to suggest a genetic basis for the racial disparity in the prevalence of BE and EAC. One possible mechanism to explain racial disparity in BE and EAC is that the normal squamous mucosa of esophagus in AA is less susceptible to refluxate damage, and subsequent development of metaplastic BE, than that of EA. This might be secondary to endogenous protection from more active detoxifying enzymes, more extensive mucin production, or the increased expression of protective gene products in AA. A second possible mechanism is that Barrett’s epithelium is more likely to replace damaged squamous epithelium in EA than AA. Thus, admixed AA cases with BE may carry European alleles that more readily replace damaged squamous epithelium with metaplastic BE. Further sequencing of the two regions with a larger sample size will be conducted to locate genetic variants for esophageal adenocarcinoma and Barrett’s esophagus and identify the mechanisms that explain resistance to the development of BE in individuals of African ancestry.

Supporting information

S1 Fig. The genome-wide admixture mapping result for excess European ancestry signals from 54 individuals using all SNPs.

(TIF)
S2 Fig. The genome-wide admixture mapping result for excess European ancestry signals from 54 individuals using SNPs without the 908 SNPs having allele frequencies different from the 1000 genomes.
(TIF)

S3 Fig. The genome-wide admixture mapping result for excess European ancestry signals from the 28 individuals having high genotyping quality.
(TIF)

S4 Fig. The genome-wide admixture mapping result for excess African ancestry signals from 54 individuals using all SNPs.
(TIF)

S5 Fig. The genome-wide admixture mapping result for excess African ancestry signals from 54 individuals excluding the 908 SNPs having different allele frequency from the 1000 genomes.
(TIF)

S6 Fig. The genome-wide admixture mapping result for excess African ancestry signals from the 28 individuals having high genotyping quality.
(TIF)

S1 File. Estimating the association effect size of variants in the two admixture mapping regions.
(DOCX)

Author Contributions

Conceptualization: Sanford D. Markowitz, Robert Elston, Amitabh Chak.


Formal analysis: Xiangqing Sun, Kishore Guda, Xiaofeng Zhu, Robert Elston, Amitabh Chak.

Funding acquisition: Amitabh Chak.


Project administration: Apoorva K. Chandar, Amitabh Chak.


Software: Robert Elston.

Supervision: Robert Elston.

Writing – original draft: Xiangqing Sun, Robert Elston, Amitabh Chak.

Nalbantoglu, Kishore Guda, Sanford D. Markowitz, Xiaofeng Zhu, Robert Elston, Amitabh Chak.

References


