

## A Segregation Analysis of Barrett's Esophagus and Associated Adenocarcinomas

Xiangqing Sun<sup>1</sup>, Robert Elston<sup>1,2</sup>, Jill Barnholtz-Sloan<sup>1,2</sup>, Gary Falk<sup>2,6</sup>, William M. Grady<sup>8</sup>, Margaret Kinnard<sup>3</sup>, Sumeet K. Mittal<sup>9</sup>, Joseph E. Willis<sup>2,4</sup>, Sanford Markowitz<sup>2,7</sup>, Wendy Brock<sup>5</sup>, and Amitabh Chak<sup>2,5</sup>

### Abstract

Familial aggregation of esophageal adenocarcinomas, esophagogastric junction adenocarcinomas, and their precursor Barrett's esophagus (BE) has been termed familial BE (FBE). Numerous studies documenting increased familial risk for these diseases raise the hypothesis that there may be an inherited susceptibility to the development of BE and its associated cancers. In this study, using segregation analysis for a binary trait as implemented in S.A.G.E. 6.0.1, we analyzed data on 881 singly ascertained pedigrees to determine whether FBE is caused by a common environmental or genetic agent and, if genetic, to identify the mode of inheritance of FBE. The inheritance models were compared by likelihood ratio tests and Akaike's A Information Criterion. Results indicated that random environmental and/or multifactorial components were insufficient to fully explain the familial nature of FBE, but rather, there is segregation of a major type transmitted from one generation to the next ( $P < 10^{-10}$ ). An incompletely dominant inheritance model together with a polygenic component fits the data best. For this dominant model, the estimated penetrance of the dominant allele is 0.1005 [95% confidence interval (95% CI), 0.0587-0.1667] and the sporadic rate is 0.0012 (95% CI, 0.0004-0.0042), corresponding to a relative risk of 82.53 (95% CI, 28.70-237.35) or odds ratio of 91.63 (95% CI, 32.01-262.29). This segregation analysis provides epidemiologic evidence in support of one or more rare autosomally inherited dominant susceptibility allele(s) in FBE families and, hence, motivates linkage analyses. *Cancer Epidemiol Biomarkers Prev*; 19(3): 666-74. ©2010 AACR.

### Introduction

The striking increase in the incidence of Barrett's esophagus (BE), esophageal adenocarcinoma (EAC), and esophagogastric junctional adenocarcinoma (EGJAC) over the last few decades implicates in the pathogenesis of these diseases a potential major environmental factor, such as gastroesophageal reflux, obesity, and/or smoking. Yet, in the past 2 decades, there have been numerous

case reports, case series, and cross-sectional studies documenting increased familial risk for these diseases (1-8), raising the hypothesis that there may be an inherited susceptibility to the development of BE and its associated cancers. Indeed, there is strong evidence of a genetic component in the pathogenesis of gastroesophageal reflux (9-11) and obesity (12-14). Just as a change in environmental factors within the past several decades has interacted with undiscovered genetic factors to contribute to the increasing epidemic of obesity in the United States, it is quite plausible that changes in environmental factors are acting together with genetic susceptibility factors to contribute to the rising incidence of BE, EAC, and EGJAC. Of course, the reported increased familial risk of BE and its associated cancers could be related either to shared environmental factors within families or to transmitted genetic susceptibility factors. The purpose of this study was to do a complex segregation analysis on pedigrees of probands with BE, EAC, and EGJAC to determine whether the familial aggregation of these diseases is consistent with the transmission of autosomal genes. An incompletely dominant inheritance model together with a polygenic component was identified. Such a finding on a large sample would provide strong motivation for a linkage analysis of multiplex families as well as provide parameter estimates that would be expected to maximize the power of a model-based linkage analysis.

**Authors' Affiliations:** <sup>1</sup>Division of Genetic and Molecular Epidemiology, Department of Epidemiology and Biostatistics, <sup>2</sup>Case Comprehensive Cancer Center, <sup>3</sup>Division of Gastroenterology, Wade Park VA Medical Center, <sup>4</sup>Department of Pathology, University Hospitals-Case Medical Center, and <sup>5</sup>Division of Gastroenterology, University Hospitals-Case Medical Center, Case Western Reserve University School of Medicine; <sup>6</sup>Gastroenterology and Hepatology, Cleveland Clinic, Cleveland Clinic Lerner College of Medicine; <sup>7</sup>Howard Hughes Medical Institute, Cleveland, Ohio; <sup>8</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center, Department of Medicine, University of Washington Medical School, R&D Service, Puget Sound VA Healthcare System, Seattle, Washington; and <sup>9</sup>Department of Surgery, Creighton University, Omaha, Nebraska

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

**Corresponding Author:** Xiangqing Sun, Department of Epidemiology and Biostatistics, Case Western Reserve University, 2103 Cornell Road, Cleveland, OH 44106. Phone: 216-368-4658; Fax: 216-368-4880. E-mail: xsun@darwin.cwru.edu

doi: 10.1158/1055-9965.EPI-09-1136

©2010 American Association for Cancer Research.

## Materials and Methods

### Recruitment of Probands

The familial BE (FBE) pedigrees were collected as part of an active multicenter study registered at clinicaltrials.gov, identifier NCT00288119, where the ultimate goal of this accrual is to identify putative susceptibility genes. The methodology used in this study to recruit probands and develop the family structures to identify FBE pedigrees has been reported before (8, 15). Briefly, families were ascertained through probands with established or newly diagnosed BE, EAC, or EGJAC at five tertiary care academic hospitals in the United States as previously described (15). Recruitment periods varied among the hospitals, depending on available personnel and institutional review board approval, and ranged from 6 mo to 7 y in duration. All eligible patients seen in the endoscopy suite during the active recruitment period were approached for study entry.

### Determination of Family History

A FBE questionnaire that elicits information about reflux symptoms (patterned on the validated Mayo gastroesophageal reflux questionnaire; ref. 16), relevant covariates, and a detailed family history of BE, esophageal cancer, and other cancers was used in this ongoing study. Probands were defined as the affected member who brought the family to the study, and each family was assumed to be singly ascertained via the sole proband in each family. A FBE questionnaire was administered to these probands by a study nurse or sent them via mail. Mailed questionnaires were followed up with a phone call by the study nurse. Probands were requested to obtain permission for study personnel to contact their first-degree family members and, similarly, for their affected relatives. Questionnaires were also administered to relatives who contacted the study centers and consented to participate. For deceased relatives, the next of kin was asked to complete relevant portions of the questionnaire and provide consent for review of the medical records. Screening endoscopy was offered to family members who had not had previous upper endoscopy. The endoscopic and histologic diagnosis for family members reported to have BE or esophageal cancer was confirmed by reviewing endoscopy and pathology records. For all reported diagnoses that could not be confirmed by review of medical records, the diagnosis of BE or EAC was classified as possible but not confirmed. Each institutional review board for human investigation at participating hospitals approved the study protocol.

### Definition of Trait and Other Variables

EAC was defined as adenocarcinoma from a mass that predominantly involves the tubular esophagus, and EGJAC was defined as adenocarcinoma on histology from a mass that predominantly involves the gastroesophageal junction but does have some esophageal involvement.

The definition of BE required clear and rigorous documentation of measurements of the affected segment of tubular esophagus in the endoscopy report and documented presence of intestinal metaplasia on biopsy. Biopsies from the gastroesophageal junction or an irregular Z-line containing intestinal metaplasia were classified as intestinal metaplasia of the cardia and not considered to be BE. Institutional biopsies and outside biopsies, when obtained, were reviewed by a gastrointestinal pathologist. BE, EAC, and EGJAC have been considered part of a single complex trait termed FBE because these conditions are epidemiologically similar and there is strong evidence that BE is the precursor of nearly all EACs and a substantial proportion of EGJACs (17-20). The disease trait for the analysis done here was defined as BE, EAC, or EGJAC. For the purpose of this segregation analysis, an individual with a confirmed diagnosis of BE, EAC, or EGJAC was considered affected. Individuals with no reported history, or individuals who did not have BE or adenocarcinoma on review of medical records, were considered unaffected. The trait status for individuals with a diagnosis that could not be confirmed was treated as unknown.

### Statistical Analysis

Complex segregation analysis of FBE was done as implemented in the program SEGREG within the S.A.G.E. 6.0.1 program package (21). The models implemented in SEGREG are briefly described here, with special emphasis on the details that are specific to modeling complex segregation of a binary trait.

The segregation model for a binary trait assumes that susceptibility to disease,  $\gamma$ , defined as the probability that an individual is affected with disease, depends on an unobserved latent factor termed type, designated as  $u$ , which can take on one of the three values AA, AB, or BB. If the segregation is Mendelian (in which case  $\gamma$  is interpreted as a penetrance function), the type  $u$  represents a putative genotype that underlies the distribution of the observed phenotype. In the transmission probability model, types are characterized by two parameter sets: type frequencies and transmission parameters (22). If one assumes Hardy-Weinberg equilibrium, the type frequencies in the population can be defined by a single parameter  $q_A$ , the frequency of A. The set of transmission parameters— $\tau_{AA}$ ,  $\tau_{AB}$ , and  $\tau_{BB}$ —represents the probabilities, respectively, that individuals of types AA, AB, and BB will transmit the component A (allele, if the type is a genotype) to offspring. Assuming that mating is random, the transmissions from each parent are independent. For a Mendelian locus, Hardy-Weinberg equilibrium is assumed, and  $\tau_{AA}$ ,  $\tau_{AB}$ , and  $\tau_{BB}$  are equal to 1, 0.5, and 0, respectively. In the specific case that the disease segregation is caused purely by a random environmental factor, there is no transmission from generation to generation, and the three unobserved types transmit equally ( $\tau_{AA} = \tau_{AB} = \tau_{BB}$ ). In addition, a “general” transmission model allows the transmission probabilities  $\tau_{AA}$ ,  $\tau_{AB}$ , and  $\tau_{BB}$

to take on any arbitrary values between 0 and 1. A more restricted general transmission model assumes homogeneity of the phenotypic distribution across generations and must satisfy two conditions: the type frequencies must follow Hardy-Weinberg equilibrium proportions and  $\tau_{AB}$  must be equal to a specific function of the frequency of A,  $\tau_{AA}$  and  $\tau_{BB}$ ; thus, only two of the transmission probabilities can be freely estimated (23).

For a binary trait, the susceptibility  $\gamma$  is defined as the cumulative logistic function,

$$\gamma = \frac{e^{\theta_u(i)}}{1 + e^{\theta_u(i)}},$$

where, conditional on type  $u$ , the logit of the susceptibility for the  $i$ th individual,  $\theta_u(i)$ , can depend on both major type  $u$  and mean-centered covariate values  $x_{i1}, x_{i2}, \dots, x_{ip}$ :

$$\theta_u(i) = \beta_u + \xi_1 x_{i1} + \xi_2 x_{i2} + \dots + \xi_p x_{ip},$$

where  $\beta_u$  is the intercept corresponding to type  $u$  and  $\xi_1, \xi_2, \dots, \xi_p$  are the covariate regression coefficients. This is analogous to estimating covariate coefficients in a linear model simultaneously with all the other parameters in the segregation analysis of a quantitative trait (24).

Two ways to allow for multifactorial components in the analysis of a binary trait are implemented in SEGREG. The first is a multivariate logistic model (MLM) that allows the incorporation of first-order nuclear family residual association parameters into the logit of susceptibility (25). In this model, the residual association parameters include  $\delta_{FM}$  for father-mother,  $\delta_{FO}$  for father-offspring,  $\delta_{MO}$  for mother-offspring, and  $\delta_{SS}$  for sib-sib residual associations. These association parameters can be converted to residual correlations that depend on the expected value of the logit. In multigenerational pedigrees, the class D assumption (26) is used (i.e., that all sib-sib residual associations are equal but not necessarily due to common parentage alone, nor does it necessarily assume that they are equal to the parent-offspring residual association). The second model is the finite polygenic mixed model (FPMM; refs. 27, 28), which assumes that the logit of susceptibility is influenced by a small number,  $v$ , of additive diallelic loci in addition to possible segregation at a single major locus with large effect. In this model, the effect of the finite number of polygenic loci is represented by additive polygenic variance  $\sigma_v^2$  of the logit; that is, within nuclear families, the sibling correlation is assumed to be the same as the parent-offspring correlation and the correlation between  $k$ th degree relatives is  $(1/2)^k - 1$  times that of first-degree relatives. The incorporation of a polygenic component into the transmission model results in the unified model (29), but in SEGREG it is extended to multigenerational pedigrees by assuming a finite number of polygenic loci.

The segregation models studied can be classified according to the number of underlying susceptibility types

in the model: one major susceptibility type, and therefore no segregation; major type models that assume a mixture of two susceptibility types (dominant or recessive, if there is Mendelian segregation); and major type models that assume a mixture of three susceptibility types. In the analysis presented here, because of difficulties in maximizing a general likelihood in the general codominant case (three unrestricted susceptibilities), we restricted our attention to three susceptibilities that are additive on the logistic scale. In each case, additional familial/multifactorial components were investigated. For the one susceptibility case, both a purely random environmental model (no transmission,  $\tau_{AA} = \tau_{AB} = \tau_{BB} = q_A$ ) without any multifactorial component and the same model but with familial/multifactorial components (sometimes called a "sporadic model") were fitted. For the major type models that have two or three susceptibility types, a homogeneous model with no transmission ( $\tau_{AA} = \tau_{AB} = \tau_{BB} = q_A$ ), a Mendelian transmission model ( $\tau_{AA} = 1, \tau_{AB} = 0.5, \tau_{BB} = 0$ ), and two general transmission models were fitted. These general transmission models subsume both the environmental no transmission and the Mendelian transmission models as special cases, and either assume homogeneity of the phenotype distribution across generations ( $\tau_{AB}$  is a function of  $q_A, \tau_{AA}$ , and  $\tau_{BB}$ ) or do not. In the latter case, all three transmission probabilities are freely estimated in the interval 0,1 without any further restriction.

SEGREG also allows computing the likelihood of a no transmission model ( $\tau_{AA} = \tau_{AB} = \tau_{BB}$ ) that is not homogeneous across generations, but results for this model were not reported here because they did not change any of the conclusions. It should be noted that, in contrast to what occurs for a trait that can take on more than two values, for a binary trait the no transmission model, either with or without the assumption of homogeneity across generations (corresponding to  $\tau_{AA} = \tau_{AB} = \tau_{BB} = q_A$  and  $\tau_{AA} = \tau_{AB} = \tau_{BB}$ , respectively), is the same whatever the number of susceptibility types if there is no multifactorial component incorporated into the model. This is so because the phenotypic distribution is then a mixture of Bernoulli random variables, which is simply a Bernoulli random variable.

All parameters were estimated by numerical maximization of the likelihood, with SEs calculated by numerical double differentiation of the log likelihood evaluated at the maximum likelihood estimates of all the model parameters. The 95% confidence interval (95% CI) of an estimate was calculated as estimate  $\pm 1.96 \times$  SE, and the 95% CIs for susceptibility,  $\gamma$ , relative risk, and odds ratio were calculated by the delta method. Various hypotheses about mode of transmission are incorporated into models that impose different restrictions on the unrestricted general model, and cause the likelihood to be smaller than that for the general model. Likelihood ratio tests were used to test the significance of the departure from specified null hypothesis models using the

**Table 1.** Affection status by gender and founder status for a total of 8,835 individuals from FBE families for segregation analysis

	Affected (%)	Unaffected (%)	Unknown (%)	Total
Males	812 (17.8)	3,653 (80.3)	85 (1.9)	4,550
Females	180 (4.2)	4,061 (94.8)	44 (1.0)	4,285
Founders	97 (3.2)	2,898 (96.0)	25 (0.8)	3,020
Nonfounders	895 (15.4)	4,816 (82.8)	104 (1.8)	5,815

asymptotic properties of the likelihood ratio test: when the null hypothesis is not on a boundary of the unrestricted model, twice the difference in  $\ln(\text{likelihood})$  between the two models is asymptotically distributed as  $\chi^2$ , with the number of degrees of freedom equal to the difference in the number of independent parameters estimated. In some cases, where the null hypothesis is on the boundary of the unrestricted model, the asymptotic distribution is a mixture of  $\chi^2$  distributions (30). In other cases, the asymptotic distribution is unknown, and then Akaike's A Information Criterion (AIC; ref. 31) can be used to select the better model. For a fixed number of susceptibility types, the SEGREG output includes  $P$  values based on the appropriate asymptotic distribution of the likelihood ratio criterion, either a  $\chi^2$  distribution or a mixture of  $\chi^2$  distributions, as well as AIC values, which can be compared across results for different numbers of susceptibility types. All tests of statistical significance except for variances were two-sided.

To allow for ascertainment, the likelihood of each pedigree was conditioned on the affection status of all individuals in the proband sampling frame, which for single ascertainment comprise the probands (32).

## Results

### The Population Sample

A total of 881 pedigrees were included in the study and the average pedigree size was 10.03 (SD, 8.47). There were 81 (9.2%) pedigrees with one generation, 152 (17.2%) pedigrees with two generations, 474 (53.8%) pedigrees with three generations, and 174 (19.8%) pedigrees with four or more generations. Of the total 8,835 individuals in the data, 992 (11.2%) were affected with BE, EAC, or EGJAC. As shown in Table 1, the percentage of males affected is about four times that of females (17.8% versus 4.2%; assuming independent observations,  $\chi^2_1 = 410.9$ ,  $P < 2.2 \times 10^{-16}$ ), and the percentage of affected among nonfounders is about five times that among founders (15.4% versus 3.2%; assuming independent observations,  $\chi^2_1 = 294.6$ ,  $P < 2.2 \times 10^{-16}$ ). These imbalances of affected individuals between males versus females and founders versus nonfounders were accounted for in the segregation analysis using (0, 1) coded covariates.

### One Susceptibility Type Models

First, the effect of covariates on the probability of being affected was evaluated in a one susceptibility type model (which assumes no segregation), denoted "no transmission." In addition to sex, founder status was also investigated as a covariate because of the noted heterogeneity of the proportions affected between founders and nonfounders.

Results in Table 2 indicate that the estimated regression coefficients  $\xi_{\text{sex}}$  and  $\xi_{\text{founder}}$  are highly significant (more than four times their SEs) for any model with sex and/or founder as covariate(s), and also by comparing Akaike's AIC values, the full model, including both sex and founder as covariates, fits the data best. Similar

**Table 2.** Parameter estimate in the one susceptibility type no transmission models, with and without (centered) covariates

Parameter	Covariate(s) in the model			
	No covariate	Sex	Founder	Sex and founder
$\beta$ (SE)	-4.24 (0.10)	-4.36 (0.11)	-4.48 (0.13)	-4.58 (0.14)
$\xi_{\text{sex}}$ (SE)*		-1.19 (0.22)		-1.11 (0.22)
$\xi_{\text{founder}}$ (SE)†			-1.80 (0.33)	-1.72 (0.33)
$-2\ln(L)$	1,435.40	1,401.25	1,388.43	1,359.45
d.f.‡	1	2	2	3
$P^{\S}$	$< 10^{-15}$	$10^{-10}$	$7.31 \times 10^{-8}$	
AIC	1,437.41	1,405.25	1,392.43	1,365.45

\*Sex value for male is 0 and for female is 1.

†Founder value for founder is 1 and for nonfounder is 0.

‡Number of functionally independent parameters estimated.

§Compared with the full model including both sex and founder as covariates.

**Table 3.** Parameter estimate in the one susceptibility type (sporadic) model, incorporating various residual familial associations by MLM ( $\delta_{FM} = 0$ )

Parameters	Familial association components			
	$\delta_{FO} = \delta_{MO} = \delta_{SS} = 0$	$\delta_{FO} = \delta_{MO} = \delta_{SS}$	$\delta_{FO} = \delta_{MO}, \delta_{SS}$	$\delta_{FO}, \delta_{MO}, \delta_{SS}$
$\beta$ (SE)	-4.58 (0.14)	-4.80 (0.17)	-4.72 (0.17)	-4.73 (0.17)
$\xi_{sex}$ (SE)*	-1.11 (0.22)	-1.09 (0.22)	-1.09 (0.22)	-1.09 (0.22)
$\xi_{founder}$ (SE)†	-1.72 (0.33)	-1.65 (0.33)	-1.56 (0.34)	-1.56 (0.34)
$\delta_{FO}$ (SE)	0	0.36 (0.19)	0.07 (0.22)	0.07 (0.23)
$\delta_{MO}$ (SE)	0	0.36 (0.19)	0.07 (0.22)	0.11 (0.44)
$\delta_{SS}$ (SE)	0	0.36 (0.19)	0.51 (0.24)	0.51 (0.24)
$-2\ln(L)$	1,359.45	1,353.00	1,350.38	1,350.37
d.f.‡	3	4	5	6
$P^{\S}$	0.028	0.269	0.920	
$P^{\parallel}$	0.011	0.106		
AIC	1,365.45	1,361.00	1,360.38	1,362.37

\*Sex value for male is 0 and for female is 1.

†Founder value for founder is 1 and for nonfounder is 0.

‡Number of functionally independent parameters estimated.

§Compared with model ( $\delta_{FO}, \delta_{MO}, \delta_{SS}$ ).

∥Compared with model ( $\delta_{FO} = \delta_{MO}, \delta_{SS}$ ).

results were obtained by likelihood ratio tests. So, in the following segregation analyses, sex and founder status were always included as covariates in the logit of FBE susceptibility.

Next, the effect of incorporating different familial/multifactorial components in these nonsegregating models was investigated. The results of familial associations in the MLM model are shown in Table 3. Because there were insufficient data to estimate the father-mother association (it was frequently not possible to maximize the likelihood), we set it equal to 0 for all these models. Although the AIC suggests that including separate parent-offspring and sib-sib associations ( $\delta_{FO} = \delta_{MO}, \delta_{SS}$ ) in the model is best, likelihood ratio tests indicated no significant differences among the three associations: the model with three equal familial associations ( $\delta_{FO} = \delta_{MO} = \delta_{SS}$ ) was not rejected when compared with the model with three different association coefficients ( $\delta_{FO}, \delta_{MO}, \delta_{SS}$ ;  $P = 0.269$ ) or with separate parent-offspring and sib-sib associations ( $\delta_{FO} = \delta_{MO}, \delta_{SS}$ ;  $P = 0.106$ ), which suggests that the model with three equal familial multifactorial components fits the data adequately.

The results of the one susceptibility type models incorporating various numbers of additive polygenic loci in the FPMM are shown in Supplementary Table S1. They indicate that the likelihood increased with the number of polygenic loci, but the increase was small, especially after including more than three polygenic loci. Furthermore, the AIC was always much smaller than seen in the MLM models in Table 3, which indicates that including a multifactorial component via the FPMM model fits the data better.

In summary, results of analyses with one susceptibility models indicate that (a) sex and founder status were significant covariates of susceptibility to FBE, (b) including three equal familial/multifactorial components significantly improved the fitting, (c) the multifactorial familial components can be best summarized by an additive polygenic variance, and (d) three additive loci are sufficient to capture this polygenic variance.

### Two Susceptibility Type Models

Our next step is to investigate whether the residual variability in the logit of susceptibility, after allowing for the covariates, is significantly better explained by the incorporation of a further susceptibility type (with unequal transmission probabilities) to decide if there may be a major latent factor—whether environmentally or genetically caused—segregating. The FPMM model incorporating three polygenic loci was used to allow for a multifactorial/polygenic component. The transmission mode was described by the transmission parameters  $\tau_{AA}, \tau_{AB}, \tau_{BB}$ , and  $q_A$ , the frequency of susceptibility allele A. Four modes of transmissions were compared: homogeneous no transmission ( $\tau_{AA} = \tau_{AB} = \tau_{BB} = q_A$ ), Mendelian transmission ( $\tau_{AA} = 1, \tau_{AB} = 0.5, \tau_{BB} = 0$ ), a homogeneous general transmission model ( $\tau_{AA}, \tau_{BB}$ ), and a similar nonhomogeneous general model in which the three transmission probabilities ( $\tau_{AA}, \tau_{AB}, \tau_{BB}$ ) are freely estimated in the interval 0,1. If there is a polygenic component incorporated into the model, for two and three susceptibility types, we denote the homogeneous no transmission model “environmental plus polygenic transmission” because there is a latent major

environmental factor “segregating” in addition to polygenic transmission.

First, homogeneity across generations was tested by comparing the homogeneous general transmission ( $\tau_{AA}$ ,  $\tau_{BB}$ ) model with the general ( $\tau_{AA}$ ,  $\tau_{AB}$ ,  $\tau_{BB}$ ) model. Then, the homogeneous no transmission model and the Mendelian model were each compared with both the homogeneous general ( $\tau_{AA}$ ,  $\tau_{BB}$ ) model and the general model ( $\tau_{AA}$ ,  $\tau_{AB}$ ,  $\tau_{BB}$ ) to test first if there is a major type segregating, and then, if so, whether this type could be Mendelian.

For two susceptibility types, a purely random environmental model was rejected on comparing it with either the general ( $\tau_{AA}$ ,  $\tau_{AB}$ ,  $\tau_{BB}$ ) model or the homogeneous general ( $\tau_{AA}$ ,  $\tau_{BB}$ ) model (Supplementary Table S2A;  $P = 1.12 \times 10^{-13}$  and  $P = 2.70 \times 10^{-12}$ , respectively). Similarly the environmental model incorporating a polygenic component was also rejected (Supplementary Table S2B;  $P = 8.76 \times 10^{-13}$  and  $P = 3.11 \times 10^{-13}$ , respectively). Moreover, purely polygenic transmission together with random environment, represented by the one susceptibility type no transmission model incorporating three polygenic loci, has a much larger AIC value (1,090.49; Supplementary Table S1) than any of the genetic models incorporating three polygenic loci that also incorporate two susceptibility types (Supplementary Table S2B; AIC values < 1,042). These results indicate that neither random environment nor polygenic transmission nor a major random environmental effect together with polygenic transmission is sufficient to fit the data; rather, there is a major type being transmitted in some manner. Furthermore, at the 0.05 level, without a polygenic component (Supplementary Table S2A), homogeneity across generations was rejected on comparing the homogeneous general ( $\tau_{AA}$ ,  $\tau_{BB}$ ) model with the general ( $\tau_{AA}$ ,  $\tau_{AB}$ ,  $\tau_{BB}$ ) model ( $P = 0.002$ ), and both purely dominant and recessive models were rejected on comparing with the general ( $\tau_{AA}$ ,  $\tau_{AB}$ ,  $\tau_{BB}$ ) model ( $P = 0.003$ ). Only under the erroneous homogeneity assumption could dominant and recessive inheritance not be rejected in a model that did not include a polygenic component ( $P = 0.110$  and  $0.126$ , respectively, for dominant and recessive inheritance). After including three polygenic loci (Supplementary Table S2B), however, homogeneity across generations was not rejected ( $P = 0.206$ ), and Mendelian dominant inheritance was also not rejected—whether compared with the homogeneous general ( $\tau_{AA}$ ,  $\tau_{BB}$ ) model or with the general ( $\tau_{AA}$ ,  $\tau_{AB}$ ,  $\tau_{BB}$ ) model ( $P = 0.750$  and  $0.441$ , respectively)—whereas recessive inheritance was rejected ( $P = 0.001$  and  $0.002$ ).

In summary, these results indicate that two susceptibility types with a polygenic component support homogeneity across generations, and Mendelian dominant inheritance with a polygenic component fits the data best.

### Three Susceptibility Type Models

Because the data were not sufficiently informative to consider a general codominant model, only three sus-

ceptibilities that are additive on the logit scale were studied. Results for three additive susceptibility types in Supplementary Table S3 indicate that there is a major type transmitted ( $P < 10^{-12}$ ), but heterogeneity across generations exists no matter whether a polygenic component is incorporated into the model or not ( $P = 0.035$  and  $0.030$ , respectively). Mendelian additive transmission was not rejected when including a polygenic component in the model ( $P = 0.750$  and  $0.118$ , comparing with the homogeneous general and the general model, respectively). These results indicate that homogeneity across generations was not supported by three additive susceptibility types, although Mendelian additive transmission was supported after including a polygenic component. Moreover, the additive inheritance model is not better than the dominant inheritance model on comparing their AIC values (Table 4; 1,032.10 versus 1,031.98). Table 4, collated from Supplementary Tables S2 and S3, gives detailed results of the major genetic models to be compared. Of particular note is the much better fit of dominant over recessive inheritance once a polygenic component is included in the model.

In summary, these results indicate that three additive susceptibility types do not fit better than two susceptibility types, and dominant transmission is sufficient to fit the data.

### Disease Prevalence

Assuming the estimates for Mendelian dominant inheritance together with a polygenic component, the mode of inheritance found to best fit the data (Table 4), the logit of the susceptibility for types AA and AB,  $-2.19$ , corresponds to a penetrance of  $\gamma_{AA} = \gamma_{AB} = 0.1005$  (95% CI, 0.0587-0.1667), whereas the logit of the susceptibility for BB,  $-6.71$ , corresponds to a sporadic rate of  $\gamma_{BB} = 0.0012$  (95% CI, 0.0004-0.0042). Because all covariates were centered, this penetrance and sporadic rate are those at average values of the covariates (i.e., the proportions of females, 48.5%, and founders, 34.2%, in these pedigrees). These results correspond, again for average covariate values, to a relative risk of 82.53 (95% CI, 28.70-237.35) or odds ratio of 91.63 (95% CI, 32.01-262.29). Incorporating the covariates into the logits, and weighting the penetrance and sporadic rate by the estimated genotype frequencies, the disease prevalence was calculated among males and females, and founders and nonfounders; this was done both with and without a polygenic component in the model (Table 5). With a polygenic component, the estimated prevalence was 0.48% in males (95% CI, 0.41-0.54%) and 0.14% in females (95% CI, 0.11-0.16%), and 0.40% in nonfounders (95% CI, 0.34-0.45%) and 0.12% in founders (95% CI, 0.10-0.14%). Combining the sex and founder status, the estimated prevalence was 0.71% in male nonfounders (95% CI, 0.63-0.80%), 0.21% in male founders (95% CI, 0.18-0.25%), 0.21% in female nonfounders (95% CI, 0.18-0.24%), and 0.06% in female founders (95% CI, 0.05-0.07%). Without a polygenic component, the estimated prevalences were higher,

**Table 4.** Parameter estimates and model AIC values under Mendelian transmission of FBE, with and without a polygenic component in the model

Parameters	Two susceptibility		Three susceptibility	
	Nonpolygenic (recessive)	Polygenic (dominant)*	Nonpolygenic (additive)	Polygenic (additive)
$\beta_{AA}$ (SE)	0.07 (0.78)	-2.19 (0.30)	1.93 (0.57)	2.28 (0.59)
$\beta_{AB}$ (SE)	-5.27 (0.22)	-2.19 (0.30)	-1.96 (0.32)	-2.22 (0.31)
$\beta_{BB}$ (SE)	-5.27 (0.22)	-6.71 (0.63)	-5.87 (0.41)	-6.72 (0.62)
$\xi_{sex}$ (SE) <sup>†</sup>	-1.29 (0.29)	-1.30 (0.28)	-1.19 (0.25)	-1.30 (0.28)
$\xi_{founder}$ (SE) <sup>‡</sup>	-1.43 (0.40)	-1.28 (0.39)	-1.43 (0.36)	-1.29 (0.39)
$\sigma_3^2$ (SE)		1.17 (0.50)		1.16 (0.49)
$q_A$	$4.46 \times 10^{-2}$	$1.76 \times 10^{-3}$	$2.08 \times 10^{-3}$	$1.79 \times 10^{-3}$
$-2\ln(L)$	1,301.83	1,019.98	1,301.92	1,020.10
d.f. <sup>§</sup>	5	6	5	6
AIC	1,311.83	1,031.98	1,311.92	1,032.10

\*The best model.

<sup>†</sup>Sex value for male is 0 and for female is 1.

<sup>‡</sup>Founder value for founder is 1 and for nonfounder is 0.

<sup>§</sup>Number of functionally independent parameters estimated.

0.60% in males (95% CI, 0.54-0.66%) and 0.19% in females (95% CI, 0.17-0.21%), and 0.55% in nonfounders (95% CI, 0.49-0.61%) and 0.14% in founders (95% CI, 0.12-0.15%).

## Discussion

The present segregation study analyzed the factors influencing the phenotype transmission step by step. We first confirmed that sex and founder status influence the susceptibility to FBE significantly when included as covariates in the logit of susceptibility. Familial correlations (representing arbitrary multifactorial components) or a polygenic component was also found to contribute to the susceptibility of FBE, and we found that incorporating the latter was better, it being adequate to allow for three additive polygenic loci in the model. Then, various

transmission models were fitted for two susceptibility types and three susceptibility types. After including the significant polygenic component, dominant inheritance is seen to be the best model.

This segregation analysis of families of patients with BE, EAC, or EGJAC provides strong epidemiologic evidence for the segregation of autosomal dominant inheritance at a Mendelian locus that achieves homogeneity across generations. Including a polygenic component also influences the susceptibility to the development of BE or its associated cancers, but it is not sufficient by itself to explain the familial transmission of the disease. In the dominant inheritance model with a polygenic component, the estimated "genetic" susceptibility to develop FBE is limited, in that the overall penetrance of the dominant genotypes,  $\gamma_{AA}$  and  $\gamma_{AB}$ , is only 0.1005 (95% CI, 0.0587-0.1667). However, this penetrance is 82.53 (95% CI, 28.70-237.35) times the sporadic rate ( $\gamma_{BB} = 0.0012$ ; 95% CI, 0.0004-0.0042), and using the estimated allele frequency (0.0018), it is found that this corresponds to an odds ratio of 91.63 (95% CI, 32.01-262.29). It is this difference in magnitude that is key for a successful linkage study. Although the estimated sporadic rate is very low, it is nevertheless significantly different from 0. By comparing the likelihood of this best model to the dominant model with a sporadic rate  $\gamma_{BB}$  fixed at 0 (effectively accomplished by fixing  $\beta_{BB} = -500$ ), the  $P$  value for departure of  $\gamma_{BB}$  from 0 is  $1.11 \times 10^{-13}$ .

Under our best model, the estimated prevalence of disease is higher in nonfounders than in founders and higher in males than in females (0.71% in male nonfounders, 0.21% in male founders, 0.21% in female nonfounders, and 0.06% in female founders; Table 5). There has been an increase in obesity over time, which probably

**Table 5.** The estimated prevalence (and SE) of FBE in males and females, and founders and nonfounders under Mendelian dominant transmission

	Founder	Nonfounder	Mean
A. No polygenic component			
Male	0.24% (0.01%)	0.96% (0.05%)	0.60% (0.03%)
Female	0.07% (0.005%)	0.30% (0.02%)	0.19% (0.01%)
Mean	0.14% (0.01%)	0.55% (0.03%)	
B. Three polygenic loci			
Male	0.21% (0.02%)	0.71% (0.04%)	0.48% (0.03%)
Female	0.06% (0.01%)	0.21% (0.02%)	0.14% (0.01%)
Mean	0.12% (0.01%)	0.40% (0.03%)	

accounts for the change in susceptibility to FBE. Moreover, the incidence of EAC has increased dramatically within the United States in the past 3 decades (33, 34). This increase is partly attributable to an epidemic of obesity (35, 36) and might reflect other unrecognized environmental factors. It is likely that the incidence of BE is also increasing along with the increased incidence of its related cancers. The nonhomogeneity in disease prevalence between founders and nonfounders shown in this study likely reflects the effect of obesity and other environmental factors that have increased the prevalence of disease each generation. Ideally, we would have regressed on year of birth and age to allow for such nonhomogeneity, but founder status was used as a surrogate because year of birth and/or age were missing on 42% of the pedigree members in the data. Similarly, obesity and smoking status were not considered as covariates because they are only available on 16% of the subjects. The difference in FBE susceptibility between males and females is consistent with the known male predominance of BE and its related cancers (37-39).

Our analysis estimated an overall prevalence of BE and its associated cancers to be 0.48% for men and 0.14% in women. Because the prevalence of EAC or EGJAC is only a small fraction of that of BE, this overall prevalence estimate in our study population is essentially an estimate of the prevalence of BE. The estimate is lower than the 1.6% prevalence of BE reported in a Swedish population study (39) and the 6.8% prevalence of BE reported in an American study of patients undergoing colonoscopy (40). The differences in prevalence estimates are likely related to several factors. The prevalence of BE is known to increase with age (39-41). The study by Ronkainen et al. (39) measured prevalence in an adult population, the study by Rex et al. (40) examined an even more selected older population undergoing colonoscopy, whereas our study included individuals of all ages. Furthermore, the prevalence of BE has likely undergone a rapid increase in the past several decades along with the increasing incidence of EAC (34). Thus, our study population, which included founders from several decades ago and included a sizable proportion of individuals who never had endoscopy, is expected to have a lower prevalence of BE than the more comprehensive and more recent studies by Ronkainen et al. (39) and Rex et al. (40). Among those individuals we classified as affected, 66.7% had BE, and the remainder had EAC or EGJAC. On classifying as

unknown those individuals with confirmed EAC or EGJAC, as well as those with a diagnosis that could not be confirmed, the parameter estimates of the dominant model were similar, with (as expected) lower penetrance of the dominant genotypes. In addition, the estimated polygenic variance was 11% larger.

In summary, this segregation analysis provides the first epidemiologic evidence in support of a genetic etiology for the familial aggregation of BE, EAC, and EGJAC. Furthermore, the pattern of disease in FBE families is consistent with a rare autosomally inherited dominant susceptibility allele. Although the segregation model assumed only one such allele, this kind of analysis of a binary trait cannot detect allelic heterogeneity, nor can it detect locus heterogeneity if the penetrances are similar across loci. Indeed, the significant polygenic component could be nothing more than an indication of locus heterogeneity, rather than of polygenic inheritance, and that its estimate increased on classifying as unknown those with EAC or EGAC suggests that FBE as defined here would be a better phenotype for linkage analysis. Whereas the power of a linkage analysis is not affected by allelic heterogeneity, it can be seriously diminished by locus heterogeneity. However, with an odds ratio of more than 91, it is difficult to believe that no locus exists for which the odds ratio is high enough to make linkage detectable. Studies need to be done to discover these loci with putative susceptibility alleles.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Acknowledgments

We thank Denise Buonocore-Sassano, Anna Haas, and Kasey Orłowski for their accrual efforts and Mike Warfe for securing the data in a web-based data repository.

#### Grant Support

National Institute of Diabetes and Digestive and Kidney Diseases USPHS research grants R01 DK070863 and K24 DK002800, National Institute of General Medical Sciences USPHS grant R37 GM028356, National Cancer Center for Research Resources resource grant P41 RR003655, and National Cancer Institute Cancer Center Support Grant P30 CA043703.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 11/04/2009; revised 12/16/2009; accepted 12/23/2009; published OnlineFirst 03/02/2010.

#### References

- Gelfand MD. Barrett's esophagus in sexagenarian identical twins. *J Clin Gastroenterol* 1983;5:251-3.
- Crabb DW, Berk MA, Hall TR, Conneally PM, Biegel AA, Lehman GA. Familial gastroesophageal reflux and development of Barrett's esophagus. *Ann Intern Med* 1985;103:52-4.
- Prior A, Whorwell PJ. Familial Barrett's esophagus? *Hepatogastroenterology* 1986;33:86-7.
- Jochem VJ, Fuerst PA, Fromkes JJ. Familial Barrett's esophagus associated with adenocarcinoma. *Gastroenterology* 1992;102:1400-2.
- Eng C, Spechler SJ, Ruben R, Li FP. Familial Barrett's esophagus and adenocarcinoma of the gastroesophageal junction. *Cancer Epidemiol Biomarkers Prev* 1993;2:397-9.
- Fahmy N, King JF. Barrett's esophagus: An acquired condition with a genetic predisposition. *Am J Gastroenterol* 1993;88:1262-5.
- Poynton AR, Walsh TN, O'Sullivan G, Hennessy TP. Carcinoma arising in familial Barrett's esophagus. *Am J Gastroenterol* 1996;91:1855-6.

8. Chak A, Lee T, Kinnard MF, et al. Familial aggregation of Barrett's esophagus, esophageal adenocarcinoma, and esophago-gastric junctional adenocarcinoma in Caucasian adults. *Gut* 2002; 51:323–8.
9. Cameron AJ, Lagergren J, Henriksson C, Nyren O, Locke GR III, Pedersen NL. Gastroesophageal reflux disease in monozygotic and dizygotic twins. *Gastroenterology* 2002;122:55–9.
10. Mohammed I, Cherkas LF, Riley SA, Spector TD, Trudgill NJ. Genetic influences in gastro-oesophageal reflux disease: a twin study. *Gut* 2003;52:1085–9.
11. Zheng Z, Nordenstedt H, Pedersen NL, Lagergren J, Ye W. Lifestyle factors and risk for symptomatic gastroesophageal reflux in monozygotic twins. *Gastroenterology* 2007;132:87–95.
12. Farooqi S, O'Rahilly S. Genetics of obesity in humans. *Endocr Rev* 2006;27:710–8.
13. Farooqi IS. Monogenic human obesity. *Front Horm Res* 2008;36:1–11.
14. Ranadive SA, Vaisse C. Lessons from extreme human obesity: monogenic disorders. *Endocrinol Metab Clin North Am* 2008;37: 733–51.
15. 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;15:1668–73.
16. Locke GR, Talley NJ, Weaver AL, Zinsmeister AR. A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc* 1994;69: 539–47.
17. Cameron AJ, Lomboy CT, Pera M, Carpenter HA. Adenocarcinoma of the esophagogastric junction and Barrett's esophagus. *Gastroenterology* 1995;109:1541–6.
18. Haggitt RC, Tryzelaar J, Ellis FH, Colcher H. Adenocarcinoma complicating columnar epithelium-lined (Barrett's) esophagus. *Am J Clin Pathol* 1978;70:1–5.
19. Hirota WK, Loughney TM, Lazas DJ, Maydonovitch CL, Rholi V, Wong RK. Specialized intestinal metaplasia, dysplasia, and cancer of the esophagus and esophagogastric junction: prevalence and clinical data. *Gastroenterology* 1999;116:277–85.
20. Ruol A, Parenti A, Zaninotto G, et al. Intestinal metaplasia is the probable common precursor of adenocarcinoma in Barrett esophagus and adenocarcinoma of the gastric cardia. *Cancer* 2000;88: 2520–8.
21. S.A.G.E. 6.0.1: Statistical Analysis for Genetic Epidemiology; 2009. Available from: <http://darwin.cwru.edu/sage/>.
22. Elston RC, Stewart J. A general model for the genetic analysis of pedigree data. *Hum Hered* 1971;21:523–42.
23. Demenais FM, Elston RC. A general transmission probability model for pedigree data. *Hum Hered* 1981;31:93–9.
24. Elston RC, Nambodiri KK, Glueck CJ, Fallat R, Tsang R, Leuba V. Study of the genetic transmission of hypocholesterolemia and hypertriglyceridemia in a 195 member kindred. *Ann Hum Genet* 1975;39:67–87.
25. Karunaratne PM, Elston RC. A multivariate logistic model (MLM) for analyzing binary family data. *Am J Med Genet* 1998;76:428–37.
26. Bonney GE. On the statistical determination of major gene mechanisms in continuous human traits: regressive models. *Am J Med Genet* 1984;18:731–49.
27. Fernando RL, Stricker C, Elston RC. The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. *Theor Appl Genet* 1994;88:573–80.
28. Lange K. An approximate model of polygenic inheritance. *Genetics* 1997;147:1423–30.
29. Lalouel JM, Rao DC, Morton NE, Elston RC. A unified model for complex segregation analysis. *Am J Hum Genet* 1983;35:816–26.
30. Self S, Liang K. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 1987;82:605–10.
31. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974;AC-19:716–3.
32. Elston RC, Sobel E. Sampling considerations in the gathering and analysis of pedigree data. *Am J Hum Genet* 1979;31:62–9.
33. Pera M, Trastek VF, Pairolero PC, Cardesa A, Allen MS, Deschamps C. Barrett's disease: pathophysiology of metaplasia and adenocarcinoma. *Ann Thorac Surg* 1993;56:1191–7.
34. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *J Natl Cancer Inst* 2005;97:142–6.
35. Chow WH, Blot WJ, Vaughan TL, et al. Body mass index and risk of adenocarcinomas of the esophagus and gastric cardia. *J Natl Cancer Inst* 1998;90:150–5.
36. Engel LS, Chow WH, Vaughan TL, et al. Population attributable risks of esophageal and gastric cancers. *J Natl Cancer Inst* 2003; 95:1404–13.
37. Drewitz DJ, Sampliner RE, Garewal HS. The incidence of adenocarcinoma in Barrett's esophagus: a prospective study of 170 patients followed 4.8 years. *Am J Gastroenterol* 1997;92:212–5.
38. Avidan B, Sonnenberg A, Schnell TG, Chejfec G, Metz A, Sontag SJ. Hiatal hernia size, Barrett's length, and severity of acid reflux are all risk factors for esophageal adenocarcinoma. *Am J Gastroenterol* 2002;97:1930–6.
39. Ronkainen J, Aro P, Storskrubb T, et al. Prevalence of Barrett's esophagus in the general population: an endoscopic study. *Gastroenterology* 2005;129:1825–31.
40. Rex DK, Cummings OW, Shaw M, et al. Screening for Barrett's esophagus in colonoscopy patients with and without heartburn. *Gastroenterology* 2003;125:1670–7.
41. Cameron AJ, Lomboy CT. Barrett's esophagus: age, prevalence, and extent of columnar epithelium. *Gastroenterology* 1992;103:1241–5.