Received: 26 July 2021 | Revised: 1 November 2021 | Accepted: 20 January 2022

DOI: 10.1002/gepi.22451

RESEARCH ARTICLE

Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

WILEY

Gene-level association analysis of ordinal traits with functional ordinal logistic regressions

Chi-Yang Chiu[1,2] | **Shuqi Wang**[3] | **Bingsong Zhang**[3] | **Yutong Luo**[3] | **Claire Simpson**[4] | **Wei Zhang**[5] | **Alexander F. Wilson**[2] | **Joan E. Bailey-Wilson**[2] | **Elvira Agron**[6] | **Emily Y. Chew**[6] | **Jun Zhang**[7] | **Momiao Xiong**[8] | **Ruzong Fan**[2,3]

[1]Division of Biostatistics, Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, USA

[2]Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland, USA

[3]Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University Medical Center, Washington, District of Columbia, USA

[4]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee, USA

[5]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

[6]National Eye Institute, National Institute of Health, Bethesda, Maryland, USA

[7]Department of Computer Science and Engineering Technology, University of Maryland Eastern Shore, Princess Anne, Maryland, USA

[8]Human Genetics Center, University of Texas–Houston, Houston, Texas, USA

Correspondence
Ruzong Fan, Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University Medical Center, Washington, DC, USA.
Email: rf740@georgetown.edu

Funding information
US National Science Foundation, Grant/Award Number: DMS-1915904

Abstract

In this paper, we develop functional ordinal logistic regression (FOLR) models to perform gene-based analysis of ordinal traits. In the proposed FOLR models, genetic variant data are viewed as stochastic functions of physical positions and the genetic effects are treated as a function of physical positions. The FOLR models are built upon functional data analysis which can be revised to analyze the ordinal traits and high dimension genetic data. The proposed methods are capable of dealing with dense genotype data which is usually encountered in analyzing the next-generation sequencing data. The methods are flexible and can analyze three types of genetic data: (1) rare variants only, (2) common variants only, and (3) a combination of rare and common variants. Simulation studies show that the likelihood ratio test statistics of the FOLR models control type I errors well and have good power performance. The proposed methods achieve the goals of analyzing ordinal traits directly, reducing high dimensionality of dense genetic variants, being computationally manageable, facilitating model convergence, properly controlling type I errors, and maintaining high power levels. The FOLR models are applied to analyze Age-Related Eye Disease Study data, in which two genes are found to strongly associate with four ordinal traits.

Chi-Yang Chiu and Shuqi Wang contributed equally to this study.

KEYWORDS
association mapping, complex disease, functional data analysis, ordinal traits, rare variants, sequence data

1 | INTRODUCTION

In genetic study, many phenotypes are naturally ordered and discrete values, that is, ordinal traits. For instance, Age-Related Eye Disease Study (AREDS) collects ordinal traits such as eye drusen size, drusen area, age-related macular degeneration (AMD) categories, and AMD severity scale (Age-Related Eye Disease Study Research Group, 1999, 2001, 2005). In addition, both common and rare variants are available in AREDS using a customized exome chip and next-generation sequencing technologies (Fritsche et al., 2013, 2016). It is interesting and important to develop statistical models and software to perform association analysis for ordinal traits.

Two strategies can be applied to an association analysis: (1) using a common variant or a few common variants in an analysis such as genome-wide association studies (GWAS), and (2) using a large number variants of sequencing data for a gene-based analysis. For common variant analysis, a few methods have been developed for the analysis of ordinal traits. German et al. (2020) proposes score test statistics based on ordinal logistic regression (OLR) for GWAS of biobank data. Morris et al. (2010) uses multinomial regression for GWAS on multi-category traits. By using single-nucleotide polymorphism (SNP) data, proportional odds models and modified proportion odds models are proposed to analyze ordinal traits, which are similar to OLR models (Wang et al., 2019; Xue et al., 2019; Zhang & Li, 2016). For the AMD study, cumulative genetic risk score analysis is applied to drusen progression by unadjusted single variant analysis (Hoffman et al., 2016). OLR is well-studied in ordinal categorical data analysis (Agresti, 2010, 2018; Bilder & Loughin, 2014), which can be readily utilized in the analysis of common variants.

The well-studied OLR in statistics cannot be directly applied to rare variant analysis due to the high dimensionality of rare variant data. First, analysis of a single rare variant or a few rare variants at a time may lead to low power and numerous results which are hard to interpret. Second, an analysis by putting all variants in the OLR model directly can cause convergence problems or can lead to high false positives (e.g., additive OLRs in Section 2.2). To analyze a large number of rare variants or a combination of rare and common variants, it is necessary to build novel gene-based models to overcome convergence problems, to control type I error rates properly, to maintain high power, and to be able to make valid interpretation.

To our knowledge, the gene-based analytic methods are very limited for analysis of ordinal traits. For multilocus association testing, Wang, Ma et al. (2018) proposes a permutation based test that combines multiple p values from single SNP level testings for rare variant analysis which needs intensive computation for a large number of variants. Wang, Philip et al. (2018) develops an approach based on Bayesian generalized linear mixed model with the cumulative logit link function which can be computationally expensive.

In this paper, we propose a gene-level association testing procedure to analyze ordinal traits through functional ordinal logistic regression (FOLR). The FOLR procedure is built upon functional data analysis and the basic idea is to treat dense genetic data as stochastic functions (de Boor, 2001; Ferraty & Romain, 2010; Horváth & Kokoszka, 2012; Ramsay et al., 2009; Ramsay & Silverman, 2005; Ross, 1996). The proposed methods can achieve the goal of reducing high dimensionality of dense genetic variants of the next-generation sequencing data via functional regressions. In addition, the genetic effects are treated as a smooth function of genetic positions since very dense sequencing data lead to strong linkage disequilibrium, high dependence of variants, and continuity of the genetic effects. Moreover, the proposed methods are computationally manageable, facilitate model convergence, properly control type I errors, and maintain power levels. The methods are applied to analyze AREDS ordinal traits to show their practical applications.

The organization of the paper is as follows. In Section 2, we introduce additive OLR and FOLR models to perform gene-based association analysis for ordinal traits. After a presentation of OLR models in Section 2.1, we introduce an additive OLR model and point out its problem in Section 2.2, and then we introduce theoretical FOLR models in Section 2.3. The theoretical FOLR models are revised to analyze variant data in Section 2.4. Based on the revised FOLR models, likelihood ratio test (LRT) statistics are built in Section 2.5. Simulation settings and AREDS ordinal traits and genetic data are presented in Sections 2.6, 2.7, and 2.8. In Section 3, we present the results of simulation studies and data analysis. Section 4 provides a discussion with respect to the FOLR models and their usage in gene-based analysis of sequencing data.

2 | METHODS

2.1 | Ordinal logistic regressions

Consider a study with a population sample consisting of ordinal phenotypic traits, covariates, and genetic data. For an individual in the population, let Y denote her/his ordinal trait that has $L \geq 2$ categories $\{1, 2, ..., L\}$, $Z = (z_1, ..., z_c)'$ a $c \times 1$ vector of covariates, and $X = (x(u_1), ..., x(u_m))'$ a genotype vector of m variants at the physical locations $0 \leq u_1 \leq u_2 \leq \cdots \leq u_m$, where $x(u_j)$ $(= 0, 1, 2)$ is the number of minor alleles at the jth variant. To relate trait variable Y to covariates Z and genotpye X, one may use an OLR model with a logistic link

$$\log \frac{P(Y \leq \ell | Z, X)}{1 - P(Y \leq \ell | Z, X)} = \alpha_\ell + Z'\gamma + h(X), \quad \ell = 1, 2, ..., L-1, \tag{1}$$

where α_ℓ are regression intercepts ($\ell = 1, 2, ..., L-1$), γ is a $c \times 1$ vector of fixed effect coefficients of covariates, and h is a function of genotype. The model (1) assumes that regression coefficient vector γ has the same effects and so the genetic effect $h(X)$ on each of the response categories but each category has its own intercept α_ℓ. Therefore, the intercepts α_ℓ increase in $\ell = 1, 2, ..., L-1$, since $P(Y \leq \ell | X, Z)$ increases in ℓ for each fixed value of Z and X. The function h can be modeled parametrically or nonparametrically. The OLR model (1) is also called the proportional odds model or the cumulative logit model (Agresti, 2010; Bilder & Loughin, 2014; Faraway, 2016; German et al., 2020). With the inverse-logistic transformation of the cumulative probabilities, the cell probabilities are obtained by

$$\begin{aligned} P(Y = 1 | Z, X) &= \frac{\exp\{\alpha_1 + Z'\gamma + h(X)\}}{1 + exp\{\alpha_1 + Z'\gamma + h(X)\}}, \\ P(Y = \ell | Z, X) &= \frac{\exp\{\alpha_\ell + Z'\gamma + h(X)\}}{1 + \exp\{\alpha_\ell + Z'\gamma + h(X)\}} \\ &\quad - \frac{\exp\{\alpha_{\ell-1} + Z'\gamma + h(X)\}}{1 + \exp\{\alpha_{\ell-1} + Z'\gamma + h(X)\}}, \\ &\quad 2 \leq \ell \leq L-1, \\ P(Y = L | Z, X) &= 1 - \textstyle\sum_{\ell=1}^{L-1} P(Y = \ell | Z, X). \end{aligned} \tag{2}$$

Let $\{\alpha_l\} = \alpha_1, ..., \alpha_{L-1}$, and $\theta = (\{\alpha_l\}, \gamma, h)$. Given the cell probabilities, one can obtain the likelihood function for an observation as follows:

$$L(\theta | Y, Z, X) = \Pi_{\ell=1}^{L} [P(Y = \ell | Z, X)]^{I(Y=\ell)}, \tag{3}$$

where $I(Y = \ell)$ is the indicator function of $Y = \ell$. Given a sample of data, one may estimate parameters/coefficients through maximum-likelihood estimate procedure.

2.2 | Additive OLR models

For the simplicity, we use the notations $\pi_\ell = P(Y \leq \ell | Z, X)$ and $logit(\cdot)$ as the logistic function in this section and thereafter. To model the function h in the OLR model (1), one may use an additive effect model: $h(X) = \sum_{j=1}^{m} x(u_j)\beta_j$. This gives us the following additive OLR model

$$\text{logit}(\pi_\ell) = \alpha_\ell + Z'\gamma + \sum_{j=1}^{m} x(u_j)\beta_j, \ell = 1, ..., L-1. \tag{4}$$

The additive OLR model (4) has a straightforward model form and it models effects of covariates and genetic variants directly. If the number m of variants is small and all variants are common, it can be useful as German et al. (2020). However, if the number m of genetic variants is large and some variants are rare such as sequencing data, the additive OLR model (4) may run into convergence problems and the power can be low. To overcome the problem, we incorporate a dimension reduction procedure for genotype data in the model fitting by FOLR models.

2.3 | FOLR models

To consider the dimension reduction on genotype data and to handle rare variants, one may use functional regressions to convert the genotype data of an individual into a genetic variant function (GVF) (Fan et al., 2013, 2014). The idea of functional regressions is to view the observed discrete genotypes $(x(u_1), ..., x(u_m))'$ of an individual as a realization of an unobserved stochastic function $X(u)$ (Ross, 1996). Let the physical locations of the m variants be normalized on the unit region $[0, 1]$. To relate the GVF to the trait status adjusting for covariates, we consider the following FOLR model:

$$\text{logit}(\pi_\ell) = \alpha_\ell + Z'\gamma + \int_0^1 X(u)\beta(u)du, \ell = 1, ..., L-1, \tag{5}$$

where $\beta(u)$ is the genetic effect of GVF $X(u)$ at position u, and the other terms are the same as those in the additive OLR model (4). In the FOLR model (5), the GVF $X(u)$ is assumed to be smooth. This assumption can be relaxed by considering the following β-smooth only FOLR model:

$$\text{logit}(\pi_\ell) = \alpha_\ell + Z'\gamma + \sum_{j=1}^{m} x(u_j)\beta(u_j),\ \ell = 1, ..., L-1, \tag{6}$$

where the genetic effect function $\beta(u)$ is assumed to be continuous/smooth and so it is called β-smooth only FOLR model. In model (6), we use the observed genotype data $X = (x(u_1), ..., x_i(u_m))'$ directly rather than the smooth GVF $X(u)$ as in (5).

2.4 | Revised FOLR models

To estimate the genetic effect function $\beta(u)$ in FOLR model (5) and β-smooth only FOLR model (6), one may expand it using B-spline or Fourier basis functions. Specifically, given a series of K_β basis functions $\psi_1(u), ..., \psi_{K_\beta}(u)$, we have the expansion $\beta(u) = (\psi_1(u), ..., \psi_{K_\beta}(u))(\beta_1, ..., \beta_{K_\beta})' = \psi(u)'\beta$, where $\beta = (\beta_1, ..., \beta_{K_\beta})'$ is a $K_\beta \times 1$ vector of coefficients and $\psi(u) = (\psi_1(u), ..., \psi_{K_\beta}(u))'$. One may choose B-spline basis or Fourier basis (Ramsay et al., 2009; Ramsay & Silverman, 2005).

To estimate the GVF $X(u)$ with observed genotypes X, we use an ordinary linear square smoother (Ramsay et al., 2009; Ramsay & Silverman, 2005). Let $\phi_k(u), k = 1, ..., K$, be a series of K basis functions. Let Φ denote the $m \times K$ matrix containing the values $\phi_k(u_j)$, and we let $\phi(u) = (\phi_1(u), ..., \phi_K(u))'$. Using discrete realization $X = (x(u_1), ..., x(u_m))'$, we estimate the GVF $X(u)$ using an ordinary linear square smoother as follows:

$$\hat{X}(u) = (x(u_1), ..., x(u_m))\Phi[\Phi'\Phi]^{-1}\phi(u). \tag{7}$$

Assume that the genetic effect function $\beta(u)$ is expanded by a series of basis functions $\psi_k(u), k = 1, ..., K_\beta$, as $\beta(u) = \psi(u)'\beta$. Replacing $X(u)$ in the FOLR model (5) with $\hat{X}(u)$ in (7) and $\beta(u)$ with the expansion, we have the following revised FOLR:

$$\begin{aligned}\text{logit}(\pi_\ell) &= \alpha_\ell + Z'\gamma + (x(u_1), ..., x(u_m))\Phi[\Phi'\Phi]^{-1}\\ &\quad \int_0^1 \phi(u)\psi'(u)du\beta,\\ &= \alpha_\ell + Z'\gamma + W'\beta,\ \ell = 1, ..., L-1,\end{aligned} \tag{8}$$

where $W' = (x(u_1), ..., x(u_m))\Phi[\Phi'\Phi]^{-1}\int_0^1 \phi(u)\psi'(u)du$. In the statistical packages R, codes to calculate $\Phi[\Phi'\Phi]^{-1}$ and $\int_0^1 \phi(u)\psi'(u)du$ are readily available (Ramsay et al., 2009).

For the β-smooth only FOLR model (6), $\beta(u_j)$ is introduced as genetic effect at the position u_j and $\beta(u)$ is the genetic effect function. Expanding $\beta(u_j)$ with predetermined basis functions, the β-smooth only FOLR model (6) can be revised as follows:

$$\begin{aligned}\text{logit}(\pi_\ell) &= \alpha_\ell + Z'\gamma + \left[\sum_{j=1}^{m} x(u_j)(\psi_1(u_j), ..., \right.\\ &\quad \left.\psi_{K_\beta}(u_j))\right](\beta_1, ..., \beta_{K_\beta})',\\ &= \alpha_\ell + Z'\gamma + W'\beta,\ \ell = 1, ..., L-1,\end{aligned} \tag{9}$$

where $W' = \sum_{j=1}^{m} x(u_j)(\psi_1(u_j), ..., \psi_{K_\beta}(u_j))$.

2.5 | LRT statistics

Given n independent individuals with observed information of ordinal traits, covariate vectors, and genotype vectors as Y_i, Z_i, and X_i, let $\pi_{i0} = 0$ and $\pi_{i\ell} = P(Y_i \le \ell | Z_i, X_i), 1 \le \ell \le L, i = 1, 2, ..., n$, be the cumulative probability of Y_i conditional on the ith subject's covariates and genotpyes, respectively. Based on relation (3), the likelihood function is

$$\begin{aligned}L(\{\alpha_\ell\}, \beta, \gamma) &= \Pi_{i=1}^{n}\Pi_{\ell=1}^{L}[P(Y_i = \ell | Z_i, X_i)]^{I(Y_i=\ell)}\\ &= \Pi_{i=1}^{n}\Pi_{\ell=1}^{L}[\pi_{i\ell} - \pi_{i,\ell-1}]^{I(Y_i=\ell)}.\end{aligned} \tag{10}$$

In the likelihood (10), $\pi_{i\ell}$'s can be obtained with the choice of the additive OLR model (4), FOLR model (8), and β-smooth only FOLR model (9), respectively. With the likelihood function, one can calculate the maximum-likelihood estimate of the parameters and then obtain the LRT statistics.

To test for association between the ordinal trait and the m genetic variants, the null hypothesis is $H_0 : \beta = (\beta_1, ..., \beta_m)' = 0$ for additive OLR model (4), and $H_0 : \beta = (\beta_1, ..., \beta_{K_\beta})' = 0$ for FOLR model (8) and the β-smooth only FOLR model (9). Under the null of no association between the trait and the genotype data, the models (4), (8), and (9) are simplified as follows:

$$\text{logit}(\pi_\ell) = \alpha_\ell + Z'\gamma,\ \ell = 1, 2, ..., L-1. \tag{11}$$

By fitting the FOLR model (8) or (9) and the null model (11), we may test the null $H_0 : \beta = 0$ by a χ^2-distributed LRT statistic with degrees of freedom K_β. For the additive OLR model (4) and the null model (11), the LRT statistic has degrees of freedom m.

One can use the *MASS* R package to fit the proposed models using *polr* function to calculate the LRT statistics and related p values (Venables & Ripley, 2002).

TABLE 1 Notation used in the tables and figures

Notation	Description and interpretation
LRT_FOLR_BS	LRT of FOLR model (8) with the B-Spline basis versus null model (11)
LRT_FOLR_FR	LRT of FOLR model (8) with the Fourier basis versus null model (11)
LRT_beta_BS	LRT of FOLR model (9) with the B-Spline basis versus null model (11)
LRT_beta_FR	LRT of FOLR model (9) with the Fourier basis versus null model (11)

2.6 | Simulation studies

To assess the performance of the proposed LRT statistics of FOLR models, we perform simulation studies to evaluate their empirical type I error rates and power levels. In total, we consider four test statistics presented in Table 1 based on the FOLR model (8) or (9) and the null model (11). A variant is considered to be rare if its MAF is ≤0.03. We consider two scenarios of genetic variants: (1) some variants are common and the rest are rare, and (2) all variants are rare only. We generate ordinal traits with three, five, six, and eight categories, respectively. In each simulated sample, 2000 or 4000 subjects are generated.

2.6.1 | Genetic variants

We generate sequencing data from 10,000 chromosomes over 1 Mb regions under a coalescent model by using the software package COSI (Schaffner et al., 2005; The International HapMap Consortium, 2007). The sequence data are generated using COSI's calibrated best-fit models, and the generated European haplotypes mimic Centre d'Etude du Polymorphisme Humain (CEPH) Utah individuals with ancestry from northern and western Europe in terms of the site frequency spectrum and linkage disequilibrium (LD) patterns. We randomly select subregions of size 6, 12, 18, 24, and 30 kb from the 1 Mb region to draw genetic variants for the simulation studies. In the simulations, roughly 10% variants are common and the rest are rare.

2.6.2 | Type I error simulations

To estimate type I error rates of the LRT statistics, we generate the ordinal trait for each subject using the following null model:

$$\text{logit}(\pi_\ell) = \alpha_\ell + 0.1z_1 + 0.2z_2, \ell = 1, ..., L-1 \tag{12}$$

where $(\alpha_1, \alpha_2) = (-0.7, 0.7)$ for the case with three categories, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (-1.39, -0.41, 0.41, 1.39)$ for the case with five categories, $(\alpha_1, ..., \alpha_5) = (-0.37, 0.31, 0.88, 1.49, 2.35)$ for the case with six categories, $(\alpha_1, ..., \alpha_7) = (-0.60, 0.00, 0.46, 0.88, 1.33, 1.87, 2.67)$ for the case with eight categories, z_1 is a dichotomous covariate taking values 0 and 1 with a probability of 0.5, and z_2 is a continuous covariate from a standard normal distribution $N(0, 1)$.

For each simulation scenario, 10^6 phenotype–genotype data sets are generated to fit the models and to calculate the test statistics and related p values. The empirical type I error rate is calculated as the proportion of 10^6 p values which are smaller than a given α level.

2.6.3 | Empirical power simulations

To evaluate the power of proposed tests, trait status is determined for each individual based upon the genotypes. The data sets under the alternative hypothesis are simulated by randomly selecting subregions to obtain causal variants. For each sample, a subset of m causal variants located in the selected subregion is then randomly selected from those matching the desired minor allele frequency criteria (e.g., either a combination of common and rare or only rare), yielding genotypes $(x(u_1), ..., x(u_m))$. For each data set, the causal variants are the same for all the individuals in the data set, but we allow the causal variants to be different from data set to data set. Then, we generate the ordinal phenotypic traits by the following cumulative logit model,

$$\text{logit}(\pi_\ell) = \alpha_\ell + 0.1z_1 + 0.2z_2 + \beta_1 x(u_1) + \cdots + \beta_m x(u_m), \ell = 1, 2, ..., L-1, \tag{13}$$

where α_ℓ's, z_1 and z_2' are the same as in the type I error model (12), $(x(u_1), ..., x(u_m))'$ are genotypes of m causal variants, and the βs are additive effects for the causal variants defined as follows. For genetic effect sizes, we use the approach in Wu et al. (2011) by setting $|\beta_j| = c|\log_{10}(MAF_j)|/2$, where MAF_j is the MAF of the

TABLE 2 Empirical type I error rates of the LRT statistics when some variants are common and the rest are rare

Number of ordinal categories	Region size (mean # of variants)	Nominal level α	Type I error rate of LRT statistics				
			FOLR model (8)		β-smooth only FOLR model (9)		Additive OLR (4)
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR	
3 categories	6 kb (117)	0.001	0.000999	0.001112	0.001153	0.001113	0.016532
		0.0001	0.000083	0.000112	0.000097	0.000112	0.002134
		0.00001	0.000008	0.000009	0.000009	0.000009	0.000230
	12 kb (235)	0.001	0.000957	0.001037	0.001065	0.001038	0.076013
		0.0001	0.000078	0.000105	0.000086	0.000105	0.014971
		0.00001	0.000008	0.000013	0.000010	0.000013	0.002397
	18 kb (352)	0.001	0.000975	0.001070	0.001070	0.001072	0.200935
		0.0001	0.000100	0.000123	0.000117	0.000123	0.056926
		0.00001	0.000006	0.000020	0.000007	0.000020	0.013102
	24 kb (469)	0.001	0.000968	0.001052	0.001065	0.001053	0.383890
		0.0001	0.000094	0.000113	0.000109	0.000113	0.149493
		0.00001	0.000012	0.000013	0.000013	0.000013	0.046424
	30 kb (586)	0.001	0.000975	0.001059	0.001068	0.001060	0.584945
		0.0001	0.000107	0.000121	0.000116	0.000121	0.299986
		0.00001	0.000007	0.000005	0.000007	0.000005	0.121670
5 categories	6 kb (117)	0.001	0.000857	0.001122	0.001190	0.001124	0.021473
		0.0001	0.000096	0.000126	0.000141	0.000127	0.003275
		0.00001	0.000010	0.000015	0.000015	0.000015	0.000416
	12 kb (235)	0.001	0.000838	0.001123	0.001125	0.001124	0.091105
		0.0001	0.000099	0.000130	0.000132	0.000130	0.021203
		0.00001	0.000012	0.000020	0.000015	0.000020	0.004149
	18 kb (352)	0.001	0.000796	0.001065	0.001094	0.001066	0.230415
		0.0001	0.000081	0.000104	0.000110	0.000104	0.076388
		0.00001	0.000009	0.000017	0.000015	0.000017	0.021094
	24 kb (469)	0.001	0.000808	0.001065	0.001044	0.001072	0.426695
		0.0001	0.000078	0.000112	0.000107	0.000112	0.190394
		0.00001	0.000008	0.000009	0.000008	0.000009	0.069766
	30 kb (586)	0.001	0.000838	0.001098	0.001126	0.001100	0.626729
		0.0001	0.000086	0.000113	0.000114	0.000114	0.360809
		0.00001	0.000008	0.000010	0.000012	0.000011	0.170301
6 categories	6 kb (117)	0.001	0.000734	0.001136	0.001170	0.001136	0.021527
		0.0001	0.000070	0.000120	0.000116	0.000120	0.003424
		0.00001	0.000005	0.000014	0.000006	0.000014	0.000468
	12 kb (235)	0.001	0.000762	0.001096	0.001150	0.001100	0.089623
		0.0001	0.000074	0.000107	0.000113	0.000107	0.021275
		0.00001	0.000004	0.000012	0.000006	0.000012	0.004288
	18 kb (352)	0.001	0.000737	0.001104	0.001121	0.001107	0.225137

(Continues)

TABLE 2 (Continued)

Number of ordinal categories	Region size (mean # of variants)	Nominal level α	Type I error rate of LRT statistics				
			FOLR model (8)		β-smooth only FOLR model (9)		Additive OLR (4)
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR	
		0.0001	0.000076	0.000125	0.000119	0.000125	0.075485
		0.00001	0.000006	0.000019	0.000011	0.000019	0.020952
	24 kb (469)	0.001	0.000676	0.001069	0.001052	0.001075	0.414842
		0.0001	0.000065	0.000105	0.000107	0.000106	0.185763
		0.00001	0.000003	0.000013	0.000010	0.000013	0.068433
	30 kb (586)	0.001	0.000669	0.001077	0.001053	0.001080	0.611524
		0.0001	0.000071	0.000107	0.000112	0.000107	0.350946
		0.00001	0.000003	0.000006	0.000008	0.000006	0.167192
8 categories	6 kb (117)	0.001	0.000712	0.001046	0.001141	0.001046	0.021077
		0.0001	0.000073	0.000117	0.000124	0.000117	0.003380
		0.00001	0.000005	0.000011	0.000013	0.000011	0.000505
	12 kb (235)	0.001	0.000736	0.001067	0.001054	0.001067	0.084392
		0.0001	0.000068	0.000098	0.000099	0.000098	0.020338
		0.00001	0.000006	0.000008	0.000012	0.000008	0.004283
	18 kb (352)	0.001	0.000735	0.001014	0.001064	0.001014	0.210493
		0.0001	0.000076	0.000099	0.000115	0.000099	0.071192
		0.00001	0.000005	0.000006	0.000011	0.000006	0.020263
	24 kb (469)	0.001	0.000727	0.001086	0.001124	0.001088	0.390165
		0.0001	0.000078	0.000132	0.000130	0.000132	0.173864
		0.00001	0.000006	0.000017	0.000012	0.000017	0.064960
	30 kb (586)	0.001	0.000699	0.001054	0.001032	0.001057	0.571122
		0.0001	0.000076	0.000121	0.000118	0.000121	0.323878
		0.00001	0.000007	0.000010	0.000012	0.000010	0.153897

Note: The order of B-spline basis is 4, the number of basis functions of B-spline is $K = K_\beta = 10$, the number of Fourier basis functions is $K = K_\beta = 11$, and sample size is 2000.

jth variant. Two different settings are considered: 5% and 10% of variants in the subregions are chosen as causal variants, respectively. When 5% and 10% of the variants are causal, $c = \log(30)/k$ and $\log(20)/k$, respectively. When some variants are common and the rest are rare and 5% of variants in the subregions are chosen as causal variants, the constants k increase and genetic effect sizes decrease as region sizes increase

$$k = \begin{cases} 1.25 & \text{if region size} = 6 \text{ kb}, \\ 2.25 & \text{if region size} = 18 \text{ kb}, \\ 3.25 & \text{if region size} = 30 \text{ kb}. \end{cases} \quad (14)$$

and when all variants are rare and 10% of variants in the subregions are chosen as causal variants, the constants k increase and genetic effect sizes decrease as region sizes increase

$$k = \begin{cases} 1.15 & \text{if region size} = 6 \text{ kb}, \\ 1.75 & \text{if region size} = 18 \text{ kb}, \\ 2.35 & \text{if region size} = 30 \text{ kb}. \end{cases} \quad (15)$$

For the direction of genetic effects, we consider three situations: (i) all causal variants have positive effects; (ii) 20%/80% causal variants have negative/positive effects; and (iii) 50%/50% causal variants have negative/positive effects. The empirical power are calculated as the proportion of p values which are smaller than a given α level based on 1000 simulations.

TABLE 3 Empirical type I error rates of the LRT statistics when all variants are rare

Number of ordinal categories	Region size (mean # of variants)	Nominal level α	Type I error rate of LRT statistics				
			FOLR model (8)		β-smooth only FOLR model (9)		Additive OLR (4)
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR	
3 categories	6 kb (117)	0.001	0.001147	0.001320	0.001378	0.001322	0.016619
		0.0001	0.000123	0.000141	0.000151	0.000141	0.002030
		0.00001	0.000014	0.000021	0.000018	0.000021	0.000231
	12 kb (235)	0.001	0.001013	0.001113	0.001139	0.001114	0.074101
		0.0001	0.000114	0.000097	0.000129	0.000098	0.014063
		0.00001	0.000013	0.000011	0.000016	0.000011	0.002208
	18 kb (352)	0.001	0.001027	0.001131	0.001153	0.001133	0.191173
		0.0001	0.000113	0.000116	0.000129	0.000116	0.052133
		0.00001	0.000009	0.000006	0.000010	0.000006	0.011337
	24 kb (469)	0.001	0.000988	0.001078	0.001112	0.001079	0.361997
		0.0001	0.000126	0.000133	0.000136	0.000133	0.134576
		0.00001	0.000013	0.000016	0.000015	0.000016	0.039353
	30 kb (586)	0.001	0.001003	0.001113	0.001102	0.001113	0.553245
		0.0001	0.000102	0.000122	0.000114	0.000122	0.268802
		0.00001	0.000009	0.000011	0.000010	0.000011	0.102334
5 categories	6 kb (117)	0.001	0.000862	0.001241	0.001319	0.001248	0.021662
		0.0001	0.000076	0.000122	0.000138	0.000122	0.003250
		0.00001	0.000003	0.000021	0.000012	0.000021	0.000411
	12 kb (235)	0.001	0.000847	0.001165	0.001206	0.001166	0.088747
		0.0001	0.000079	0.000109	0.000116	0.000110	0.020183
		0.00001	0.000009	0.000010	0.000012	0.000010	0.003848
	18 kb (352)	0.001	0.000836	0.001123	0.001182	0.001128	0.218923
		0.0001	0.000095	0.000110	0.000130	0.000110	0.070008
		0.00001	0.000010	0.000014	0.000013	0.000014	0.018698
	24 kb (469)	0.001	0.000794	0.001085	0.001070	0.001086	0.402191
		0.0001	0.000080	0.000104	0.000123	0.000104	0.171852
		0.00001	0.000008	0.000008	0.000010	0.000008	0.059913
	30 kb (586)	0.001	0.000797	0.001089	0.001089	0.001089	0.597675
		0.0001	0.000095	0.000110	0.000123	0.000110	0.327213
		0.00001	0.000010	0.000011	0.000014	0.000011	0.146294
6 categories	6 kb (106)	0.001	0.000725	0.001270	0.001322	0.001273	0.021853
		0.0001	0.000070	0.000134	0.000129	0.000135	0.003400
		0.00001	0.000007	0.000013	0.000013	0.000013	0.000478
	12 kb (212)	0.001	0.000679	0.001124	0.001111	0.001125	0.087204
		0.0001	0.000074	0.000121	0.000130	0.000123	0.019984
		0.00001	0.000007	0.000017	0.000010	0.000017	0.003940
	18 kb (318)	0.001	0.000696	0.001106	0.001115	0.001108	0.212792
		0.0001	0.000075	0.000127	0.000128	0.000127	0.069040

(Continues)

TABLE 3 (Continued)

Number of ordinal categories	Region size (mean # of variants)	Nominal level α	Type I error rate of LRT statistics				
			FOLR model (8)		β-smooth only FOLR model (9)		Additive OLR (4)
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR	
		0.00001	0.000010	0.000007	0.000015	0.000007	0.018673
	24 kb (424)	0.001	0.000729	0.001052	0.001121	0.001055	0.390141
		0.0001	0.000075	0.000117	0.000132	0.000118	0.167493
		0.00001	0.000004	0.000012	0.000012	0.000012	0.059134
	30 kb (530)	0.001	0.000666	0.001081	0.001057	0.001084	0.586116
		0.0001	0.000055	0.000104	0.000090	0.000104	0.322396
		0.00001	0.000003	0.000010	0.000009	0.000010	0.146330
8 categories	6 kb (106)	0.001	0.000648	0.001170	0.001233	0.001173	0.021567
		0.0001	0.000063	0.000103	0.000123	0.000103	0.003381
		0.00001	0.000006	0.000009	0.000014	0.000009	0.000483
	12 kb (212)	0.001	0.000648	0.001021	0.001040	0.001021	0.082130
		0.0001	0.000060	0.000097	0.000110	0.000098	0.019331
		0.00001	0.000007	0.000009	0.000012	0.000009	0.003958
	18 kb (318)	0.001	0.000682	0.001099	0.001072	0.001103	0.198625
		0.0001	0.000062	0.000105	0.000103	0.000105	0.064946
		0.00001	0.000006	0.000009	0.000012	0.000009	0.017900
	24 kb (424)	0.001	0.000716	0.001053	0.001116	0.001053	0.366685
		0.0001	0.000716	0.001053	0.001116	0.001053	0.156551
		0.00001	0.000008	0.000009	0.000011	0.000009	0.056070
	30 kb (530)	0.001	0.000706	0.001043	0.001094	0.001046	0.557364
		0.0001	0.000071	0.000110	0.000112	0.000110	0.303248
		0.00001	0.000008	0.000016	0.000013	0.000016	0.137448

Note: The order of B-spline basis is 4, the number of basis functions of B-spline is $K = K_\beta = 10$, the number of Fourier basis functions is $K = K_\beta = 11$, and sample size is 2000.

2.7 | Real data analysis: Application to AREDS

The proposed FOLR models are applied to analyze AREDS data (Age-Related Eye Disease Study Research Group, 1999, 2005). AREDS is a clinical trial to learn about the risk factors for macular degeneration and cataract, two leading causes of vision loss in older adults. For right eye, a total of 2911 individuals are included in this analysis with recorded demography, in which 1261 individuals are males and 1650 are females. The race composition is provided in Table S.9 and most people are white. The mean age of the 2911 individuals is 68.65 years with a standard deviation 4.92. For left eye, a total of 2914 individuals are included in the analysis, in which 1263 individuals are males and 1651 are females. The mean age of the 2914 individuals is 68.65 years with a standard deviation 4.92. A covariate analysis shown in Table S.10 indicates that age is very significant and race is significant for all the ordinal traits, and gender is significant for some traits. In the analysis, we adjust for age, race and gender as covariates.

Each individual has ordinal phenotypic traits and is genotyped using a customized exome chip (Fritsche et al., 2013, 2016). Two gene regions, CFH and ARMS2, are of primary interest. In each of the two gene regions, single variant analysis shows that some single nucleotide polymorphisms (SNPs) are associated with the risk of macular degeneration and its progression (Seddon et al., 2007). The ordinal traits include eye drusen size, drusen area, AMD categories, and AMD severity scale (Age-Related Eye Disease Study Research Group, 1999, 2001, 2005). The drusen size has six categories ranging from 0

TABLE 4 Empirical type I error rates of the LRT statistics when some variants are common and the rest are rare

Number of ordinal categories	Region size (mean # of variants)	Nominal level α	Type I error rate of LRT statistics				
			FOLR model (8)		β-smooth only FOLR model (9)		Additive OLR (4)
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR	
3 categories	6 kb (117)	0.001	0.000981	0.001104	0.001091	0.001105	0.017605
		0.0001	0.000095	0.000109	0.000108	0.000109	0.002479
		0.00001	0.000012	0.000008	0.000013	0.000008	0.000338
	12 kb (212)	0.001	0.000977	0.001074	0.001059	0.001075	0.070333
		0.0001	0.000087	0.000111	0.000092	0.000111	0.014638
		0.00001	0.000007	0.000016	0.000007	0.000016	0.002580
	18 kb (352)	0.001	0.000912	0.000982	0.001021	0.000984	0.169688
		0.0001	0.000079	0.000090	0.000089	0.000090	0.048336
		0.00001	0.000006	0.000006	0.000007	0.000006	0.011441
	24 kb (424)	0.001	0.000979	0.001048	0.001059	0.001048	0.312307
		0.0001	0.000096	0.000109	0.000105	0.000109	0.116298
		0.00001	0.000011	0.000008	0.000013	0.000008	0.035418
	30 kb (586)	0.001	0.000967	0.001036	0.001057	0.001037	0.476416
		0.0001	0.000105	0.000123	0.000113	0.000123	0.223332
		0.00001	0.000012	0.000009	0.000012	0.000009	0.084792
5 categories	6 kb (117)	0.001	0.000806	0.001003	0.001081	0.001003	0.018362
		0.0001	0.000066	0.000100	0.000097	0.000100	0.002803
		0.00001	0.000006	0.000005	0.000009	0.000005	0.000399
	12 kb (212)	0.001	0.000860	0.001020	0.001096	0.001020	0.067112
		0.0001	0.000075	0.000090	0.000096	0.000090	0.014791
		0.00001	0.000008	0.000010	0.000011	0.000010	0.002865
	18 kb (352)	0.001	0.000850	0.001022	0.001095	0.001022	0.159437
		0.0001	0.000090	0.000109	0.000111	0.000109	0.047421
		0.00001	0.000006	0.000008	0.000009	0.000008	0.012056
	24 kb (424)	0.001	0.000794	0.001019	0.001040	0.001021	0.288968
		0.0001	0.000076	0.000097	0.000098	0.000097	0.109922
		0.00001	0.000010	0.000008	0.000013	0.000008	0.035552
	30 kb (586)	0.001	0.000808	0.001032	0.001053	0.001034	0.424382
		0.0001	0.000080	0.000113	0.000110	0.000113	0.199814
		0.00001	0.000010	0.000011	0.000013	0.000011	0.078334
6 categories	6 kb (117)	0.001	0.000767	0.001118	0.001115	0.001118	0.017892
		0.0001	0.000063	0.000108	0.000106	0.000108	0.002887
		0.00001	0.000008	0.000015	0.000011	0.000015	0.000417
	12 kb (235)	0.001	0.000754	0.001079	0.001046	0.001079	0.063366
		0.0001	0.000076	0.000119	0.000107	0.000119	0.014350
		0.00001	0.000005	0.000008	0.000013	0.000008	0.002795
	18 kb (352)	0.001	0.000736	0.001027	0.001059	0.001028	0.149371

(Continues)

TABLE 4 (Continued)

Number of ordinal categories	Region size (mean # of variants)	Nominal level α	Type I error rate of LRT statistics				
			FOLR model (8)		β-smooth only FOLR model (9)		Additive OLR (4)
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR	
		0.0001	0.000069	0.000094	0.000097	0.000095	0.044462
		0.00001	0.000006	0.000006	0.000008	0.000006	0.011391
	24 kb (469)	0.001	0.000750	0.001014	0.001088	0.001015	0.266247
		0.0001	0.000074	0.000101	0.000110	0.000101	0.100272
		0.00001	0.000006	0.000008	0.000009	0.000008	0.031960
	30 kb (586)	0.001	0.000728	0.001049	0.001066	0.001051	0.383829
		0.0001	0.000065	0.000095	0.000089	0.000096	0.176130
		0.00001	0.000004	0.000012	0.000007	0.000012	0.068118
8 categories	6 kb (117)	0.001	0.000678	0.001039	0.001069	0.001039	0.016471
		0.0001	0.000065	0.000106	0.000091	0.000106	0.002624
		0.00001	0.000007	0.000009	0.000015	0.000009	0.000385
	12 kb (235)	0.001	0.000714	0.001054	0.001029	0.001055	0.056476
		0.0001	0.000067	0.000097	0.000096	0.000097	0.012477
		0.00001	0.000006	0.000009	0.000008	0.000009	0.002449
	18 kb (352)	0.001	0.000723	0.001021	0.001071	0.001021	0.131669
		0.0001	0.000089	0.000117	0.000120	0.000117	0.038182
		0.00001	0.000005	0.000002	0.000008	0.000002	0.009717
	24 kb (469)	0.001	0.000683	0.001081	0.001029	0.001082	0.230455
		0.0001	0.000057	0.000099	0.000094	0.000099	0.084140
		0.00001	0.000002	0.000003	0.000006	0.000003	0.026262
	30 kb (586)	0.001	0.000735	0.001009	0.001059	0.001011	0.324460
		0.0001	0.000069	0.000111	0.000107	0.000111	0.142084
		0.00001	0.000003	0.000009	0.000005	0.000009	0.053246

Note: The order of B-spline basis is 4, the number of basis functions of B-spline is $K = K_\beta = 10$, the number of Fourier basis functions is $K = K_\beta = 11$, and sample size is 4000.

(none) to 5, the drusen area has eight categories ranging from 0 (none) to 7, AMD categories has six levels, and AMD severity scale has 12 levels. The proposed FOLR models are applied to test association between each ordinal trait and each of the two genes.

2.8 | Functional data analysis parameters

In the data analysis and simulations, we use functions from the fda R package to create the basis functions (Ramsay et al., 2014). In the simulations and data analysis presented in the main text, the order of the B-spline basis is 4, the number of B-spline basis functions is $K = K_\beta = 10$, and the number of Fourier basis functions is $K = K_\beta = 11$. To make sure that the results are stable, we examine a wide range of parameters: $6 \le K = K_\beta \le 17$ for B-spline and Fourier basis functions. In the Supporting Information Materials, we present additional simulation results when the order of the B-spline basis is 4, the numbers of B-spline basis functions are $K = K_\beta = 6$ and 16, and the numbers of Fourier basis functions are $K = K_\beta = 7$ and 17, respectively.

3 | RESULTS

3.1 | Empirical type I error rates

In the main text, the order of the B-spline basis is 4, the number of B-spline basis functions is $K = K_\beta = 10$, the number of Fourier basis functions is $K = K_\beta = 11$.

TABLE 5 Empirical type I error rates of the LRT statistics when all variants are rare

Number of ordinal categories	Region size (mean # of variants)	Nominal level α	Type I error rate of LRT statistics				
			FOLR model (8)		β-smooth only FOLR model (9)		Additive OLR (4)
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR	
3 categories	6 kb (106)	0.001	0.001081	0.001173	0.001236	0.001173	0.017894
		0.0001	0.000110	0.000136	0.000130	0.000136	0.002483
		0.00001	0.000008	0.000009	0.000008	0.000009	0.000301
	12 kb (212)	0.001	0.000919	0.001054	0.001032	0.001054	0.070134
		0.0001	0.000101	0.000098	0.000117	0.000098	0.014343
		0.00001	0.000012	0.000008	0.000015	0.000008	0.002519
	18 kb (318)	0.001	0.000984	0.001076	0.001096	0.001077	0.166531
		0.0001	0.000102	0.000120	0.000115	0.000120	0.046609
		0.00001	0.000015	0.000015	0.000017	0.000015	0.010598
	24 kb (424)	0.001	0.001038	0.001041	0.001149	0.001041	0.302994
		0.0001	0.000096	0.000105	0.000113	0.000105	0.110538
		0.00001	0.000013	0.000008	0.000013	0.000008	0.032623
	30 kb (530)	0.001	0.000936	0.001097	0.001045	0.001099	0.462092
		0.0001	0.000104	0.000128	0.000119	0.000128	0.210000
		0.00001	0.000013	0.000015	0.000013	0.000015	0.077271
5 categories	6 kb (106)	0.001	0.000801	0.001081	0.001134	0.001081	0.018780
		0.0001	0.000078	0.000102	0.000112	0.000103	0.002873
		0.00001	0.000007	0.000014	0.000011	0.000014	0.000411
	12 kb (212)	0.001	0.000857	0.001100	0.001129	0.001101	0.066935
		0.0001	0.000083	0.000111	0.000116	0.000111	0.014761
		0.00001	0.000012	0.000013	0.000014	0.000013	0.002865
	18 kb (318)	0.001	0.000807	0.001036	0.001078	0.001038	0.155790
		0.0001	0.000090	0.000117	0.000126	0.000117	0.045385
		0.00001	0.000009	0.000015	0.000013	0.000015	0.011430
	24 kb (424)	0.001	0.000728	0.001038	0.000958	0.001042	0.282982
		0.0001	0.000071	0.000089	0.000090	0.000089	0.105670
		0.00001	0.000005	0.000008	0.000007	0.000008	0.033433
	30 kb (530)	0.001	0.000809	0.001058	0.001077	0.001059	0.434417
		0.0001	0.000094	0.000101	0.000119	0.000101	0.201703
		0.00001	0.000008	0.000009	0.000013	0.000009	0.077891
6 categories	6 kb (106)	0.001	0.000678	0.001091	0.001081	0.001091	0.018541
		0.0001	0.000061	0.000109	0.000105	0.000109	0.002951
		0.00001	0.000008	0.000013	0.000015	0.000013	0.000422
	12 kb (212)	0.001	0.000731	0.001023	0.001044	0.001023	0.063308
		0.0001	0.000077	0.000124	0.000117	0.000124	0.014275
		0.00001	0.000008	0.000011	0.000009	0.000011	0.002704
	18 kb (318)	0.001	0.000731	0.001024	0.001021	0.001026	0.145568

(Continues)

TABLE 5 (Continued)

Number of ordinal categories	Region size (mean # of variants)	Nominal level α	Type I error rate of LRT statistics				
			FOLR model (8)		β-smooth only FOLR model (9)		Additive OLR (4)
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR	
		0.0001	0.000076	0.000103	0.000104	0.000103	0.042510
		0.00001	0.000009	0.000008	0.000011	0.000008	0.010686
	24 kb (424)	0.001	0.000729	0.001076	0.001057	0.001078	0.264657
		0.0001	0.000064	0.000095	0.000098	0.000095	0.098028
		0.00001	0.000008	0.000006	0.000011	0.000006	0.030742
	30 kb (530)	0.001	0.000680	0.001084	0.001044	0.001088	0.411177
		0.0001	0.000072	0.000117	0.000098	0.000118	0.187902
		0.00001	0.000006	0.000021	0.000011	0.000021	0.071838
8 categories	6 kb (106)	0.001	0.000637	0.001016	0.001089	0.001018	0.017008
		0.0001	0.000064	0.000108	0.000099	0.000108	0.002753
		0.00001	0.000004	0.000004	0.000009	0.000004	0.000391
	12 kb (212)	0.001	0.000720	0.001061	0.001079	0.001063	0.056502
		0.0001	0.000070	0.000121	0.000111	0.000120	0.012493
		0.00001	0.000006	0.000013	0.000011	0.000013	0.002400
	18 kb (318)	0.001	0.000742	0.001041	0.001056	0.001042	0.128207
		0.0001	0.000065	0.000098	0.000112	0.000098	0.036692
		0.00001	0.000007	0.000010	0.000009	0.000010	0.009136
	24 kb (424)	0.001	0.000768	0.001022	0.001059	0.001023	0.234431
		0.0001	0.000079	0.000110	0.000106	0.000110	0.084631
		0.00001	0.000007	0.000013	0.000009	0.000013	0.026070
	30 kb (530)	0.001	0.000699	0.001032	0.000984	0.001033	0.368797
		0.0001	0.000069	0.000124	0.000116	0.000124	0.162141
		0.00001	0.000008	0.000016	0.000010	0.000016	0.060753

Note: The order of B-spline basis is 4, the number of basis functions of B-spline is $K = K_\beta = 10$, the number of Fourier basis functions is $K = K_\beta = 11$, and sample size is 4000.

When the sample size is 2000, the empirical type I error rates of the proposed LRT statistics are reported in Tables 2 and 3 at three significance levels $\alpha = 0.001$, 0.0001, and 0.00001. When the sample size is 4000, the empirical type I error rates of the proposed LRT statistics are reported in Tables 4 and 5. In Tables 2 and 4, all variants (common and rare) are used to generate genotype data under null hypothesis, while in Tables 3 and 5 only rare variants are used. Overall, the LRT statistics of FOLR model (8) and the β-smooth only FOLR model (9) control type I error rates correctly. These two models, being both stable under five region sizes of 6, 12, 18, 24, and 30 kb, show very similar results. These four tables also suggest that B-spline and Fourier basis functions provide similar results. The proposed LRT statistics of FOLR models are stable in terms of region sizes, nominal levels, smoothing methods, and basis functions. However, the LRT of additive OLR model (4) inflates type I error rates severely and getting worse when the region size is getting bigger. In summary, the FOLR models (8) and (9) can be used to analyze high dimension sequencing data; however, the additive OLR model (4) can not be used.

When the order of the B-spline basis is 4, the number of B-spline basis functions is $K = K_\beta = 6$, and the number of Fourier basis functions is $K = K_\beta = 7$, the empirical type I error rates of the proposed LRT statistics are reported in Tables S.1, S.2, S.3, and S.4 in the Supporting Information Materials. Similarly to those in

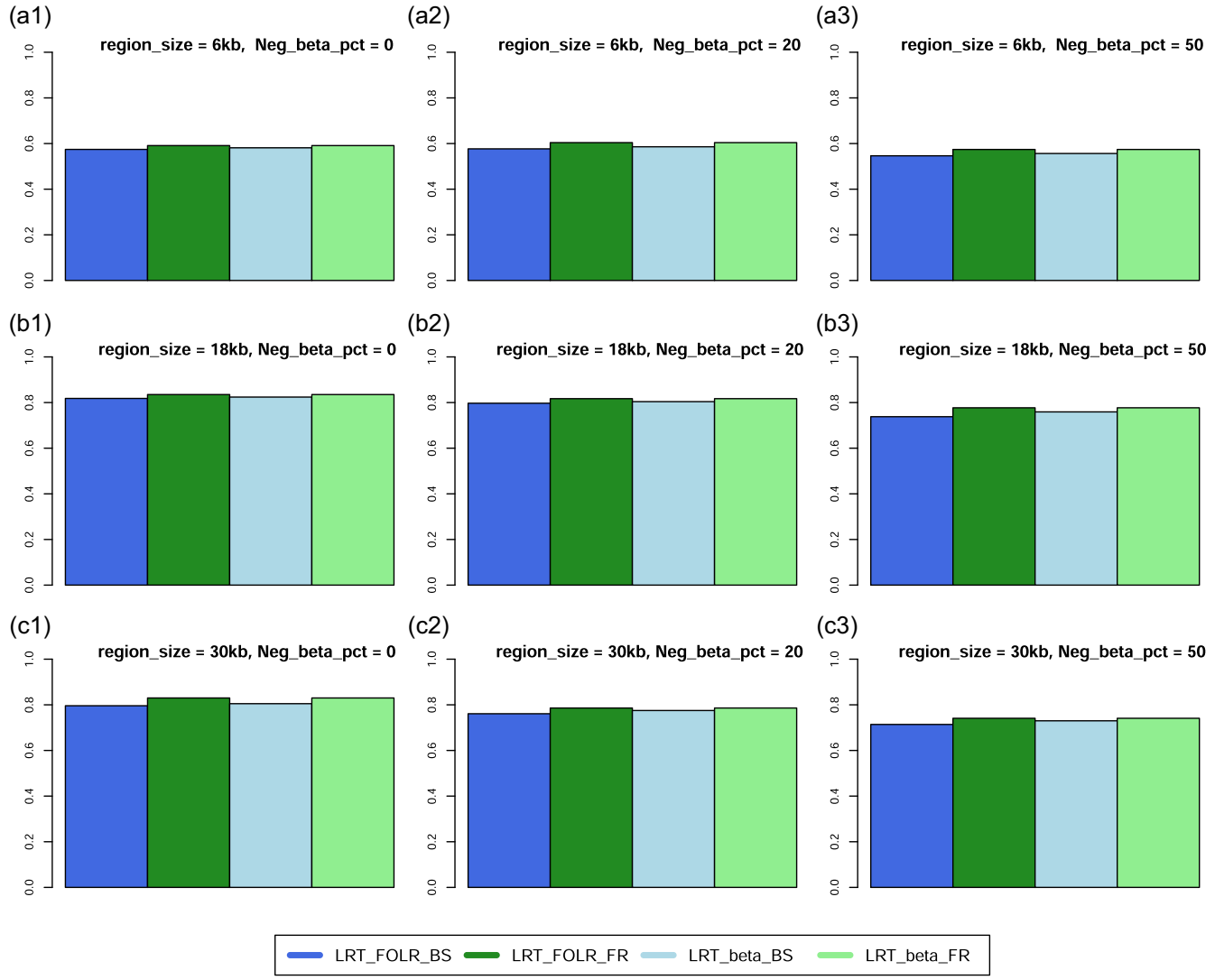

FIGURE 1 The empirical power of the statistics at $\alpha = 0.001$with 2000 subjects and six categories, when some variants are common and the rest are rare, genetic effect sizes are given by (14), and 5% of variants are causal. The order of B-spline basis is 4, the number of basis functions of B-spline is $K = K_\beta = 10$, and the number of Fourier basis functions is $K = K_\beta = 11$. (a1-a3) are power levels when the region size is 6kb and the percentage of negative effect variances is 0%, 20%, and 50%, respectively. (b1-b3) are power levels when the region size is 18kb and the percentage of negative effect variances is 0%, 20%, and 50%, respectively. (c1-c3) are power levels when the region size is 30kb and the percentage of negative effect variances is 0%, 20%, and 50%, respectively. Neg_beta_pct, percentage of causal variants which have negative effects

Tables 2, 3, 4, and 5, the type I error rates of the LRT statistics of FOLR model (8) and the beta-smooth only FOLR model (9) are well controlled.

When the order of the B-spline basis is 4, the number of B-spline basis functions is $K = K_\beta = 16$, and the number of Fourier basis functions is $K = K_\beta = 17$, the empirical type I error rates of the proposed LRT statistics are reported in Tables S.5, S.6, S.7, and S.8 in the Supporting Information Materials. When the sample size is 2000, the type I error rates of the LRT statistics of FOLR model (8) and the beta-smooth only FOLR model (9) can be inflated in Tables S.5 and S.6 when the region size is 6 and 12 kb. When the sample size increases to 4000, the type I error rates are well controlled as shown in Tables S.7 and S.8.

3.2 | Statistical power evaluation

Power performance of the proposed LRT statistics of FOLR is evaluated using data simulated under the alternative hypothesis by relation (13). Since the type I error rates of LRT statistics of FOLR model (8) and the beta-smooth only FOLR model (9) are well-controlled,

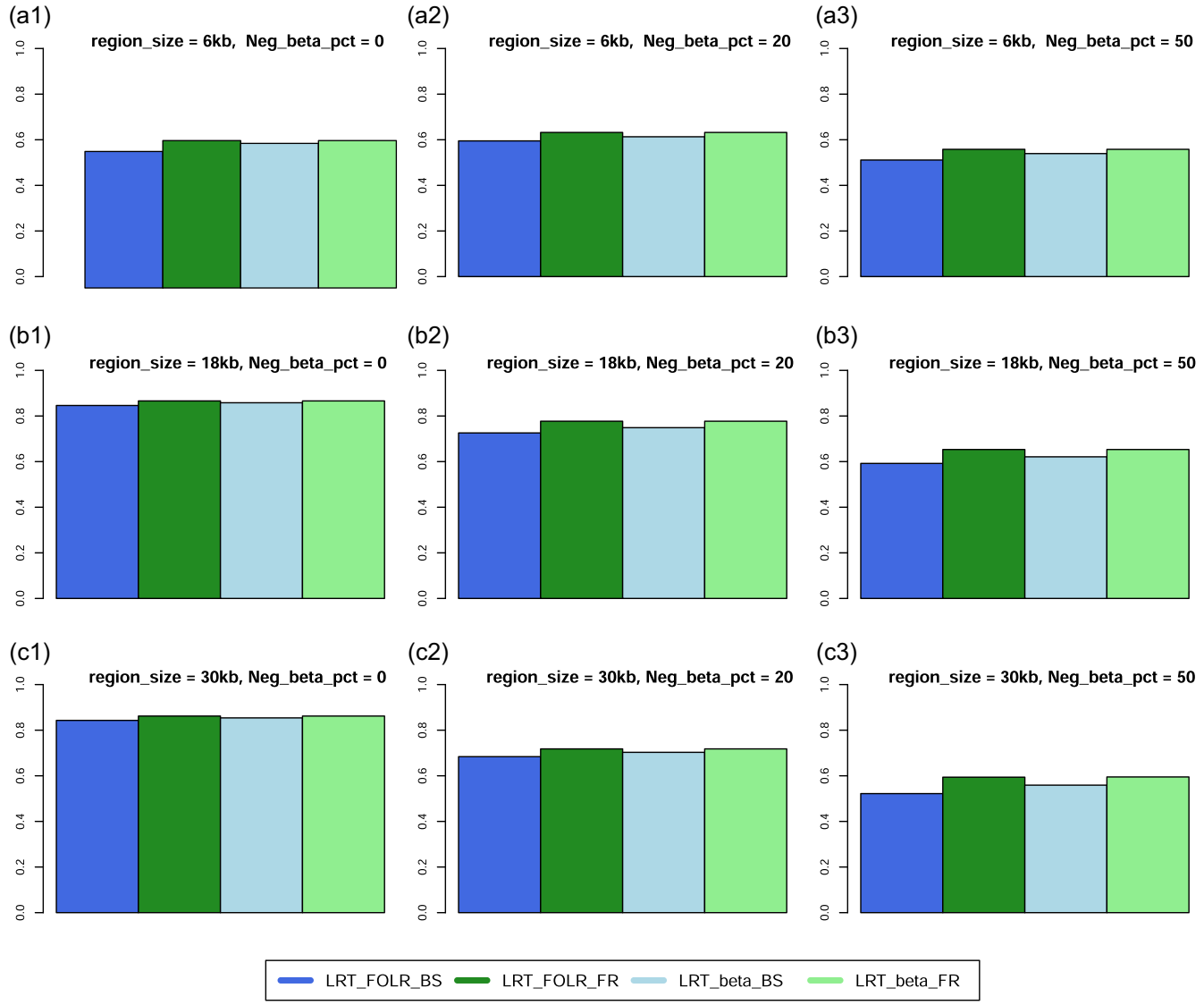

FIGURE 2 The empirical power of the statistics at $\alpha = 0.001$ with 2000 subjects and six categories, when variants are all rare, genetic effect sizes are given by (15), and 10 percent of variants are causal. The order of B-spline basis is 4, the number of basis functions of B-spline is $K = K_\beta = 10$, and the number of Fourier basis functions is $K = K_\beta = 11$. Neg_beta_pct, percentage of causal variants which have negative effects. (a1-a3) are power levels when the region size is 6kb and the percentage of negative effect variances is 0%, 20%, and 50%, respectively. (b1-b3) are power levels when the region size is 18kb and the percentage of negative effect variances is 0%, 20%, and 50%, respectively. (c1-c3) are power levels when the region size is 30kb and the percentage of negative effect variances is 0%, 20%, and 50%, respectively.

the power comparison makes sense. The LRT of additive OLR model (4) inflates type I error rates and so we do not use it for power comparison. When the sample size 2000 and six category traits, the power levels are provided in Figures 1 and 2. In Figure 1, some variants are common and the rest are rare. In Figure 2, all variants are rare. The structure of the two figures is the same, in which there are nine plots representing different combinations of simulation setting. Take Figure 1 as an example to illustrate the structure of each figure. From the left to right, the percentage of negative effect variants ranges from 0%, 20%, to 50%. From the top to bottom, the region sizes are 6, 18, and 30 kb, respectively.

In each plot of the two figures, the power levels of four LRT statistics of FOLR are compared: two are based on B-spline basis functions and two are based on Fourier basis functions by FOLR model (8) and the β-smooth only FOLR model (9), respectively. The four LRT statistics of the FOLR models have similar power. Thus, the proposed LRT statistics are very stable in terms of power performance since they do not strongly depend on whether the genotype data are smoothed or not, or which basis functions are used.

TABLE 6 Association analysis of Age-Related Disease Study (AREDS) left eye data

Gene	Type of variants (number of variants)	Trait	The *p* values of LRT statistics			
			FOLR model (8)		beta-smooth only FOLR model (9)	
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR
ARMS2	Common Only (18)	DrusenSize	3.29E−23	3.29E−23	3.28E−23	1.21E−22
		DrusenSizeBL	1.00E−24	1.00E−24	9.85E−25	3.15E−24
		DrusenArea	2.28E−20	2.28E−20	2.28E−20	8.12E−20
		DrusenAreaBL	1.40E−31	1.40E−31	1.39E−31	8.83E−31
		AMDCAT	1.83E−31	1.83E−31	1.78E−31	1.69E−31
		SevScaleBL	4.30E−40	4.30E−40	4.27E−40	1.84E−39
		SevScaleMax	6.48E−60	6.48E−60	6.43E−60	1.72E−59
		SevScale.1	1.43E−56	1.43E−56	1.39E−56	4.29E−56
		SevScale.2	1.11E−54	1.11E−54	1.09E−54	3.16E−54
	Rare & Common (25)	DrusenSize	5.28E−23	5.28E−23	5.25E−23	3.30E−22
		DrusenSizeBL	6.23E−25	6.23E−25	6.12E−25	3.97E−24
		DrusenArea	7.87E−21	7.87E−21	7.83E−21	2.63E−19
		DrusenAreaBL	1.48E−31	1.48E−31	1.37E−31	1.20E−30
		AMDCAT	2.90E−32	2.90E−32	2.71E−32	3.29E−31
		SevScaleBL	3.06E−40	3.06E−40	3.00E−40	5.91E−40
		SevScaleMax	1.50E−59	1.50E−59	1.50E−59	5.72E−59
		SevScale.1	1.91E−56	1.91E−56	1.90E−56	8.19E−56
		SevScale.2	1.51E−54	1.51E−54	1.51E−54	5.56E−54
	Rare Only (7)	DrusenSize	NA	NA	0.805775	1
		DrusenSizeBL	NA	NA	1	1
		DrusenArea	1	1	0.940895	1
		DrusenAreaBL	NA	NA	0.649658	1
		AMDCAT	NA	NA	0.133268	1
		SevScaleBL	NA	NA	0.127257	1
		SevScaleMax	NA	NA	0.12549	1
		SevScale.1	NA	NA	0.125519	1
		SevScale.2	NA	NA	0.117099	1
CFH	Common Only (59)	DrusenSize	1.06E−36	1.06E−36	1.05E−36	4.59E−33
		DrusenSizeBL	4.06E−30	4.06E−30	3.96E−30	3.50E−28
		DrusenArea	9.18E−25	9.18E−25	9.12E−25	2.04E−22
		DrusenAreaBL	3.66E−35	3.66E−35	3.66E−35	2.77E−33
		AMDCAT	2.01E−31	2.01E−31	2.00E−31	3.79E−30
		SevScaleBL	2.31E−42	2.31E−42	2.24E−42	3.44E−40
		SevScaleMax	1.57E−61	1.57E−61	1.56E−61	3.86E−57
		SevScale.1	1.48E−64	1.48E−64	1.48E−64	1.01E−60
		SevScale.2	8.91E−63	8.91E−63	8.87E−63	5.89E−59

(Continues)

TABLE 6 (Continued)

Gene	Type of variants (number of variants)	Trait	The p values of LRT statistics			
			FOLR model (8)		beta-smooth only FOLR model (9)	
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR
	Rare & Common (162)	DrusenSize	2.03E−37	2.03E−37	2.01E−37	6.27E−36
		DrusenSizeBL	1.06E−30	1.06E−30	8.48E−31	1.16E−29
		DrusenArea	1.23E−24	1.23E−24	8.61E−25	3.83E−23
		rusenAreaBL	6.78E−34	6.78E−34	1.60E−34	8.47E−34
		AMDCAT	3.43E−31	3.43E−31	3.33E−31	2.04E−30
		SevScaleBL	5.17E−42	5.17E−42	4.52E−42	1.42E−40
		SevScaleMax	9.69E−61	9.69E−61	8.03E−61	2.08E−58
		SevScale.1	3.05E−63	3.05E−63	2.39E−63	1.18E−61
		SevScale.2	2.12E−61	2.12E−61	1.48E−61	9.46E−60
	Rare Only (103)	DrusenSize	1.16E−07	1.16E−07	1.08E−07	7.77E−05
		DrusenSizeBL	0.001748	0.001748	0.001542	0.008923
		DrusenArea	0.004748	0.004748	0.00437	0.012091
		DrusenAreaBL	0.014061	0.014061	0.006773	0.015361
		AMDCAT	0.089432	0.089432	0.045615	0.011745
		SevScaleBL	0.001391	0.001391	0.000399	0.001139
		SevScaleMax	9.14E−06	9.14E−06	5.62E−06	0.000188
		SevScale.1	4.19E−05	4.19E−05	2.58E−05	6.11E−05
		SevScale.2	9.94E−05	9.94E−05	6.89E−05	0.000123

Note: The order of B-spline basis is 4, the number of basis functions of B-spline is $K = K_\beta = 10$, and the number of Fourier basis functions is $K = K_\beta = 11$. Notations: DrusenSize is drusen size, DrusenSizeBL is drusen size at baseline, DrusenArea is drusen area, DrusenAreaBL is drusen area at baseline, AMDCAT is AMD categories, SevScaleBL is AMD severity scale at baseline, SevScaleMax is AMD maximal severity scale, SevScale.1 is the 1st measure of AMD severity scale, and SevScale.2 is the 2nd measure of AMD severity scale.

In the Supporting Information Materials, we present two Figures S.1 and S.2 using the number of B-spline basis functions $K = K_\beta = 6$ and the number of Fourier basis functions $K = K_\beta = 7$. The power levels in Figures S.1 and S.2 are lower than those in Figures 1 and 2. Hence, it may draw more information by using the number of B-spline basis functions $K = K_\beta = 10$ and the number of Fourier basis functions $K = K_\beta = 11$.

3.3 | Application to age-related macular degeneration data

We analyze nine ordinal traits of AREDS data by the proposed FOLR model (8) and the β-smooth only FOLR model (9) models. Tables 6, 7, S.11, and S.12 show the results of association analysis of AREDS data for the two genes, CFH and ARMS2. The results of left eye are shown in Tables 6 and S.11, and the results of right eye are shown in Tables 7 and S.12. The data are analyzed three times: (a) all genetic variants; (b) common variants only; and (c) rare variants only. Note that the rare variants here are defined as those with MAF ≤ 0.05, while common variants are referred to those with MAF > 0.05. By considering all genetic variants, two gene regions show significant effects because all the p values of LRT statistics are small. This finding gives support to the argument that the proposed gene-based method can be used in the genome-wide association study of two survival traits.

For the ARMS2 gene, the results of analyzing common variants only exhibits little difference from the results of including all genetic variants by the proposed LRT statistics of FOLR models. This may be due to that there are only seven rare variants which do not provide much results. For the CFH gene, there are 103 rare variants in the gene region and analyzing rare variants only does provide significant results, and there are 59 common variants which provide more significant results than the rare variants. Therefore, both common and rare variants in the CFH gene affect the progression of AMD.

TABLE 7 Association analysis of Age-Related Disease Study (AREDS) right eye data

Gene	Type of variants (number of variants)	Trait	The *p* values of LRT statistics			
			FOLR model (8)		β-smooth only FOLR model (9)	
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR
ARMS2	Common only (18)	DrusenSize	3.31E−23	3.31E−23	3.30E−23	1.24E−22
		DrusenSizeBL	8.86E−25	8.86E−25	8.72E−25	2.79E−24
		DrusenArea	2.04E−20	2.04E−20	2.04E−20	7.23E−20
		DrusenAreaBL	9.47E−32	9.47E−32	9.38E−32	5.78E−31
		AMDCAT	1.69E−31	1.69E−31	1.65E−31	1.64E−31
		SevScaleBL	1.03E−36	1.03E−36	1.02E−36	1.09E−36
		SevScaleMax	4.80E−60	4.80E−60	4.76E−60	1.32E−59
		SevScale.1	1.20E−56	1.20E−56	1.17E−56	3.65E−56
		SevScale.2	9.24E−55	9.24E−55	9.08E−55	2.67E−54
	Rare and common (25)	DrusenSize	5.42E−23	5.42E−23	5.38E−23	3.22E−22
		DrusenSizeBL	5.56E−25	5.56E−25	5.46E−25	3.41E−24
		DrusenArea	7.33E−21	7.33E−21	7.29E−21	2.19E−19
		DrusenAreaBL	9.97E−32	9.97E−32	9.23E−32	7.64E−31
		AMDCAT	2.78E−32	2.78E−32	2.60E−32	3.01E−31
		SevScaleBL	2.55E−37	2.55E−37	2.54E−37	5.59E−37
		SevScaleMax	1.13E−59	1.13E−59	1.13E−59	4.27E−59
		SevScale.1	1.61E−56	1.61E−56	1.61E−56	6.89E−56
		SevScale.2	1.26E−54	1.26E−54	1.26E−54	4.64E−54
	Rare only (7)	DrusenSize	NA	NA	0.805700	1
		DrusenSizeBL	NA	NA	1	1
		DrusenArea	NA	NA	0.942297	1
		DrusenAreaBL	NA	NA	0.645099	1
		AMDCAT	NA	NA	0.133263	1
		SevScaleBL	NA	NA	0.689769	1
		SevScaleMax	NA	NA	0.121721	1
		SevScale.1	NA	NA	0.122950	1
		SevScale.2	NA	NA	0.114734	1
CFH	Common only (59)	DrusenSize	1.10E−36	1.10E−36	1.09E−36	4.68E−33
		DrusenSizeBL	4.45E−30	4.45E−30	4.34E−30	3.97E−28
		DrusenArea	9.86E−25	9.86E−25	9.78E−25	1.95E−22
		DrusenAreaBL	5.90E−35	5.90E−35	5.90E−35	4.85E−33
		AMDCAT	2.32E−31	2.32E−31	2.31E−31	4.91E−30
		SevScaleBL	7.43E−40	7.43E−40	7.39E−40	2.68E−37
		SevScaleMax	2.14E−61	2.14E−61	2.14E−61	4.99E−57
		SevScale.1	1.70E−64	1.70E−64	1.70E−64	1.22E−60
		SevScale.2	1.05E−62	1.05E−62	1.04E−62	7.18E−59

(Continues)

TABLE 7 (Continued)

Gene	Type of variants (number of variants)	Trait	The p values of LRT statistics			
			FOLR model (8)		β-smooth only FOLR model (9)	
			LRT_FOLR_BS	LRT_FOLR_FR	LRT_beta_BS	LRT_beta_FR
	Rare and common (162)	DrusenSize	2.31E−37	2.31E−37	2.29E−37	6.84E−36
		DrusenSizeBL	1.28E−30	1.28E−30	1.01E−30	1.28E−29
		DrusenArea	1.58E−24	1.58E−24	1.10E−24	3.91E−23
		DrusenAreaBL	1.27E−33	1.27E−33	3.04E−34	1.48E−33
		AMDCAT	4.57E−31	4.57E−31	4.43E−31	2.58E−30
		SevScaleBL	4.01E−39	4.01E−39	3.97E−39	6.06E−38
		SevScaleMax	1.49E−60	1.49E−60	1.24E−60	2.71E−58
		SevScale.1	3.99E−63	3.99E−63	3.14E−63	1.38E−61
		SevScale.2	2.79E−61	2.79E−61	1.96E−61	1.11E−59
	Rare only (103)	DrusenSize	1.24E−07	1.24E−07	1.17E−07	8.44E−05
		DrusenSizeBL	0.001753	0.001753	0.001549	0.009216
		DrusenArea	0.004923	0.004923	0.004549	0.012800
		DrusenAreaBL	0.014324	0.014324	0.006894	0.016439
		AMDCAT	0.090875	0.090875	0.046953	0.013201
		SevScaleBL	0.006792	0.006792	0.001884	0.003571
		SevScaleMax	1.00E−05	1.00E−05	6.25E−06	0.000220
		SevScale.1	4.51E−05	4.51E−05	2.81E−05	6.95E−05
		SevScale.2	0.000107	0.000107	7.46E−05	0.000140

Note: The order of B-spline basis is 4, the number of basis functions of B-spline is $K = K_\beta = 10$, and the number of Fourier basis functions is $K = K_\beta = 11$. Notations: DrusenSize is drusen size, DrusenSizeBL is drusen size at baseline, DrusenArea is drusen area, DrusenAreaBL is drusen area at baseline, AMDCAT is AMD categories, SevScaleBL is AMD severity scale at baseline, SevScaleMax is AMD maximal severity scale, SevScale.1 is the first measure of AMD severity scale, and SevScale.2 is the second measure of AMD severity scale.

The results of the LRT statistics of β-smooth only by FOLR model (9) in Tables 6, 7, S.11, and S.12 are similar to the results of the LRT statistics of smoothing both GVFs $X_i(u)$ and genetic effect function $\beta(u)$ by FOLR model (8). This outcome reveals that smoothing GVFs has very limited impact on the data analysis. Similar conclusion is also observed for quantitative, dichotomous, and survival traits in Fan et al. (2013, 2014, 2016).

4 | DISCUSSION

The discrete phenotypic traits, including dichotomous and ordinal traits, are commonly seen in association studies of complex diseases. Ordinal traits are categorical traits with at least three categories where possible values are ordered on a scale. Although a lot of research has been done to analyze dichotomous traits, the ordinal traits have attracted relatively less attention in genetic association studies. One may collapse an ordinal trait to be binary and analyzes the binary trait using logistic regressions. However, ignoring order information may lead to decreased power and loss of information (German et al., 2020). The goal of this article is to develop functional regression based models for gene-based association analyses of ordinal traits without collapsing them.

In this paper, we develop FOLR models to analyze ordinal traits. In the proposed FOLR models, genetic variant data are viewed as stochastic functions of physical positions and the genetic effects are treated as a function of physical positions (Ross, 1996). The FOLR models are built upon functional data analysis which can be revised to analyze the ordinal traits and high dimension genetic data (de Boor, 2001; Ferraty & Romain, 2010; Ramsay et al., 2009; Ramsay & Silverman, 2005). The proposed methods are capable of dealing with dense genotype data which is usually encountered in analyzing the next-generation sequencing data. The methods are flexible and

TABLE 8 Computational efficiency

Region size (Mean # of variants)	6 kb (106)		18 kb (318)		30 kb (530)	
Sample size	2000	4000	2000	4000	2000	4000
Time in second	0.36	0.56	0.57	0.91	0.87	1.66

Note: Running times for one simulated data set calculation which include generating a rare variant data set and analyzing it for each of three region sizes.

can handle three cases: (1) rare variants only, (2) common variants only, and (3) a combination of rare and common variants. Simulation studies show that empirical type I errors of the proposed LRT statistics of FOLR models are well controlled no matter common variants are included or not.

In the data analysis and simulations of this article, we find that almost all models successfully converge (≥99.99%). A wide range of parameters: $6 \le K = K_\beta \le 17$ for B-spline and Fourier basis functions are examined to ensure that the results are valid and stable. We provide results in Tables 2, 3, 4, 5, S.1, S.2, S.3, and S.4 to show that type I error rates are well controlled when $K = K_\beta \le 11$. If the parameter numbers increase to $K = K_\beta = 16$, one needs large sample size to properly control the type I error rates (S.5, S.6, S.7, S.8 in the Supporting Information Materials).

In terms of computation efficiency, we record running times for one simulated data set calculation which include generating a rare variant data set and analyzing it for each of three region sizes, 6, 18, and 30 kb. Table 8 shows the running times of the proposed functional-model-based methods with 10 basis functions by B-spline for a five categorical trait analysis, using a R script running single-threaded on a laptop with a 2.7 GHz Intel Core i7-6820 processor and 16 GB memory. From the results in Table 8, it takes 1.66 s to generate and analyze a data set with 530 rare variants and 4000 subjects, which is the maximum in table. Thus, it is possible to analyze $60 \times 60/1.66 \approx 2169$ such data sets in 1 h. The proposed method is computational feasible.

The proposed methods achieve the goals of analyzing ordinal traits directly, reducing high dimensionality of dense genetic variants, being computationally manageable, facilitating model convergence, properly controlling type I errors, and maintaining high power levels. The proposed models are then applied to analyze AREDS data, in which two genes are found to strongly associate with four ordinal traits. In our previous work, functional regression based models are proved feasible and efficient for association analysis of quantitative, dichotomous, and survival traits. This article fills the gap of no suitable methods to analyze ordinal traits at the gene levels.

Our proposed logistic ordinal regression model requires proportional odds assumption. The reason we use this model is that it is the common used model in analyzing the ordinal data. If the proportional odds assumption is violated, one may need to use other models, such as partial proportional odds models (Peterson & Harrell, 1990). It is interesting in developing methods to investigate the robustness of the methods in our future studies.

ACKNOWLEDGMENTS

Two anonymous reviewers and the editors, Dr. Shete, provided very good and insightful comments for us to improve the manuscript. This study is supported by US National Science Foundation grant DMS-1915904 (Ruzong Fan, Shuqi Wang, and Bingsong Zhang), by the Intramural Research Program of the National Human Genome Research Institute (Ruzong Fan, Chi-Yang Chiu, Alexander F. Wilson and Joan E. Bailey-Wilson), National Institutes of Health (NIH), Baltimore, MD, and by the Intramural Research Program of the National Eye Institute (Elvira Agron and Emily Y. Chew), NIH, Bethesda, MD. This study utilizes the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data are created or analyzed in this study.

COMPUTER PROGRAM

The methods proposed in this paper are implemented using functional data analysis (fda) procedures implemented in the statistical package R (Ramsay et al., 2009). The R codes are available from the web https://sites.google.com/a/georgetown.edu/ruzong-fan/about; for data analysis and simulations.

ORCID

Chi-Yang Chiu http://orcid.org/0000-0002-4837-3194
Momiao Xiong http://orcid.org/0000-0003-0635-5796
Ruzong Fan http://orcid.org/0000-0002-7603-2135

REFERENCES

Age-Related Eye Disease Study Research Group. (1999). The Age-Related Eye Disease Study (AREDS): design implications. AREDS report no. 1. *Control Clinical Trials*, *206*, 573–600.

Age-Related Eye Disease Study Research Group. (2001). The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study Report Number 6. *American Journal of Ophthalmology*, *132*(5), 668–681.

Age-Related Eye Disease Study Research Group. (2005). The Age-Related Eye Disease Study severity scale for age-related macular degeneration: AREDS report no. 17. *Archives of Ophthalmology*, *123*(11), 1484.

Agresti, A. (2010). *Analysis of ordinal categorical data* (Vol. 656). John Wiley & Sons.

Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons.

Bilder, C. R., & Loughin, T. M. (2014). *Analysis of categorical data with R*. CRC Press.

de Boor, C. (2001). *A practical guide to splines, applied mathematical sciences* (Vol. 27, revised version). Springer.

Fan, R. Z., Wang, Y. F., Mills, J. L., Wilson, A. F., Bailey-Wilson, J. E., & Xiong, M. M. (2013). Functional linear models for association analysis of quantitative traits. *Genetic Epidemiology*, *37*, 726–742.

Fan, R. Z., Wang, Y. F., Mills, J. L., Carter, T. C., Lobach, I., Wilson, A. F., Bailey-Wilson, J. E., Weeks, D. E., & Xiong, M. M. (2014). Generalized functional linear models for case-control association studies. *Genetic Epidemiology*, *38*, 622–637.

Fan, R. Z., Wang, Y. F., Qi, Y., Ding, Y., Weeks, D. E., Lu, Z. H., Ren, H., Cook, R. J., Xiong, M. M., Swaroop, A., Chew, E. Y., & Chen W. (2016). Gene-based association analysis for censored traits via functional regressions. *Genetic Epidemiology*, *40*(2), 133–143.

Ferraty, F., & Romain, Y. (2010). *The Oxford handbook of functional data analysis*. Oxford University Press.

Faraway, J. J. (2016). *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. CRC Press.

Fritsche, L. G., Chen, W., Schu, M., Yaspan, B. L., Yu, Y., Thorleifsson, G., Zack, D. J., Arakawa, S., Cipriani, V., Ripke, S., Igo Jr., R. P., Buitendijk, G. H. S., Sim X., Weeks D. E., Guymer R. H., Merriam, J. E., Francis, P. J., Hannum, G., Agarwal, A., ...AMD Gene Consortium. (2013). Seven new loci associated with age-related macular degeneration. *Nature Genetics*, *45*(4), 433–439.

Fritsche, L. G, Igl, W., Bailey, J. N., Grassmann, F., Sengupta, S., Bragg-Gresham, J. L., Burdon, K. P., Hebbring, S. J., Wen, C., Gorski, M., Kim, I. K., Cho, D., Zack, D., Souied, E., Scholl, H. P. N., Bala, E., Lee, K. E., Hunter, D. J., Sardell, R. J., ... Heid, I. M. (2016). A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nature Genetics*, *48*(2), 134–143.

German, C. A., Sinsheimer, J. S., Klimentidis, Y. C., Zhou, H., & Zhou, J. J. (2020). Ordered multinomial regression for genetic association analysis of ordinal phenotypes at Biobank scale. *Genetics Epidemiology*, *44*(3), 248–260.

Hoffman, J. D., Van Grinsven, M. J., Li, C., Brantley, M., McGrath, J., Agarwal, A., Scott, W. K., Schwartz, S. G., Kovach, J., Pericak-Vance, M., Sanchez, C. I., & Haines, J. L. (2016). Genetic association analysis of drusen progression. *Investigative Ophthalmology & Visual Science*, *57*(4), 2225–2231.

Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.

Morris, A. P., Lindgren, C. M., Zeggini, E., Timpson, N. J., Frayling, T. M., Hattersley, A. T., & McCarthy, M. I. (2010). A powerful approach to sub-phenotype analysis in population-based genetic association studies. *Genetics Epidemiology*, *34*(4), 335–343.

Peterson, B., & Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *39*, 205–217.

Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and Matlab*. Springer.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis*, Second Edition. New York: Springer.

Ramsay, J. O., Wickham, H., Graves, S., & Hooker, G. (2014). fda: functional data analysis. *R Package Version*, *2*(4), 142.

Ross, S. M. (1996). *Stochastic processes* 2nd ed. John Wiley & Sons.

Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J., & Altshuler, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. *Genome Research*, *15*, 1576–1583.

Seddon, J. M., Francis, P. J., George, S., Schultz, D. W., Rosner, B., & Klein, M. L. (2007). Association of CFH Y402H and LOC387715 A69S with progression of age-related macular degeneration. *The Journal of American Medical Association*, *297*(16), 1793–1800. https://doi.org/10.1001/jama.297.16.1793

The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*, 851–861.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* 4th ed. Springer.

Verhoeven, VJ, Hysi, PG, Saw, SM, Vitart, V, Mirshahi, A, Guggenheim, JA, Cotch, MF, Yamashiro, K, Baird, PN, Mackey, DA, Wojciechowski, R, Ikram, MK, Hewitt, AW, Duggal, P, Janmahasatian, S, Khor, CC, Fan, Q, Zhou, X, Young, TL, ... Klaver, CC (2012). Large scale international replication and meta-analysis study confirms association of the 15q14 locus with myopia. The CREAM consortium. *Human Genetics*, *131*(9), 1467–1480.

Verhoeven, VJ, Hysi, PG, Wojciechowski, R, Fan, Q, Guggenheim, JA, Höhn, R, MacGregor, S, Hewitt, AW, Nag, A, Cheng, CY, Yonova-Doing, E, Zhou, X, Ikram, MK, Buitendijk, GH, McMahon, G, Kemp, JP, Pourcain, BS, Simpson, CL, Mäkelä, KM, ... Hammond, CJ (2013). Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nature Genetics*, *45*(3), 314–318.

Wang, J., Ding, J., Huang, S., Li, Q., & Pan, D. (2019). A powerful method to test associations between ordinal traits and genotypes. *G3: Genes, Genomes, Genetics*, *9*(8), 2573–2579.

Wang, M., Ma, W., & Zhou, Y. (2018). Association detection between ordinal trait and rare variants based on adaptive combination of *p*-values. *Journal of Human Genetics*, *63*(1), 37–45.

Wang, X. L., Philip, V. M., Ananda, G., White, C. C., Malhotra, A., Michalski, P. J., Karuturi, K. R. M., Acklin, C., Sasner, M., Bennett, D. A., De Jager, P. L., Howell, G. R., & Carter, G. W.

(2018). A Bayesian framework for generalized linear mixed modeling identifies new candidate loci for late-onset alzheimer's disease. *Genetics*, *209*(1), 51–64. https://doi.org/10.1534/genetics.117.300673

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*, *89*(1), 82–93.

Xue, Y., Wang, J., Ding, J., Zhang, S., & Li, Q. (2019). A powerful test for ordinal trait genetic association analysis. *Statistical Applications in Genetics and Molecular Biology*, *18*(2), 20170066.

Zhang, W., & Li, Q. (2016). Incorporating Hardy-Weinberg equilibrium law to enhance the association strength for ordinal trait genetic study. *Annals of Human Genetics*, *80*(2), 102–112.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chiu, C.-Y., Wang, S., Zhang, B., Luo, Y., Simpson, C., Zhang, W., Wilson, A. F., Bailey-Wilson, J. E., Agron, E., Chew, E. Y., Zhang, J., Xiong, M., & Fan, R. (2022). Gene-level association analysis of ordinal traits with functional ordinal logistic regressions. *Genetic Epidemiology*, 46, 234–255. https://doi.org/10.1002/gepi.22451