Hindawi
Computational and Mathematical Methods in Medicine
Volume 2022, Article ID 7843990, 12 pages
https://doi.org/10.1155/2022/7843990

Research Article

Maximal Information Coefficient-Based Testing to Identify Epistasis in Case-Control Association Studies

Yingjie Guo,[1,2] **Zhian Yuan**,[3] **Zhen Liang**,[4] **Yang Wang**,[1] **Yanpeng Wang**,[5] **and Lei Xu**[1]

[1]*School of Electronic and Communication Engineering, Shenzhen Polytechnic, 7098 Liuxian Street, Shenzhen 518000, China*
[2]*Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, No. 4 Block 2 North Jianshe Road, Chengdu 610054, China*
[3]*Research Institute of Big Data Science and Industry, Shanxi University, 92 Wucheng Road, Taiyuan 030006, China*
[4]*School of Life Science, Shanxi University, 92 Wucheng Road, Taiyuan 030006, China*
[5]*Beidahuang Industry Group General Hospital, Harbin, China*

Correspondence should be addressed to Yanpeng Wang; wangyanpeng2013@126.com and Lei Xu; csleixu@szpt.edu.cn

Received 5 December 2021; Revised 12 January 2022; Accepted 27 January 2022; Published 14 February 2022

Academic Editor: Leyi Wei

Interactions between genetic variants (epistasis) are ubiquitous in the model system and can significantly affect evolutionary adaptation, genetic mapping, and precision medical efforts. In this paper, we proposed a method for epistasis detection, called EpiMIC (epistasis detection through a maximal information coefficient (MIC)). MIC is a promising bivariate dependence measure explicitly designed for rapidly exploring various function types equally and for interpreting and comparing them on the same scale. Most epistasis detection approaches make assumptions about the form of the association between genetic variants, resulting in limited statistical performance. Based on the notion that if two SNPs do not interact, their joint distribution in all samples and in only cases should not be substantially different. We developed a statistic that utilizes the difference of MIC as a signal of epistasis and combined it with a permutation resampling strategy to estimate the empirical distribution of our statistic. Results of simulation and real-world data set showed that EpiMIC outperformed previous approaches for identifying epistasis at varying degrees of heredity.

1. Introduction

Genome-wide association studies (GWAS) is an emerging research strategy for discovering associations between genetic variation (e.g., single nucleotide polymorphism (SNP)) and traits like human diseases. More than 71,000 SNPs have been confirmed to be related significantly to diseases [1–3] since the first GWAS study was published in *Science* in 2005 [4]. The majority of these markers, however, are common genetic variants with small effects. Even though the whole genome sequencing enables us to detect several rare variants with large effect, "missing heritability" for the complex disease remains extensive [5–7]. For instance, only 75% of the phenotypic variance of Alzheimer's disease has been explained by known variants [8]. One possible explanation for "missing heritability" is that complex diseases are polygenic, with multiple genes, environmental variables, and interactions involved in their etiology [9, 10]. Genetic interactions are thought to provide a potential answer to the problem of "missing heritability." The solution may be partial, but it may help develop novel gene pathway topologies [11].

Epistasis is generally detected in two ways: biologically and statistically. Bateson and Mendel [12] introduced the concept of biological epistasis, which evaluates the interdependence of genetic variants. It occurs when the effect of one allele on one genetic mutation is dependent on the presence or absence of another genetic mutation and subsequently suppresses or activates the expression of other genes. Several studies reported novel epistasis in diseases. For instance, interactions between SNPs have been associated with pulmonary tuberculosis [13, 14], recurrent miscarriage [15], polycystic ovary syndrome [16], and many more [17–20]. These findings highlight the potential and significance of epistasis research.

Statistical epistasis, coined by Fisher [21], is defined as the deviation from additive effects of genetic mutations at separate loci in terms of their overall contribution to the model. Biological epistasis, on the other hand, refers to the physical interaction of two or more biological components. Studies in model organisms [22–24] have shown that epistasis found using computational and statistical approaches may be physiologically connected in these species. The presence of statistical epistasis, however, does not imply the presence of biological epistasis. Bridging the statistical and biological epistasis gap is a crucial step toward understanding the underlying genetic architecture of complex diseases.

Currently available approaches for detecting statistical epistasis can be divided into three groups depending on their strategy: exhaustive methods, search methods, and machine learning-based methods [25, 26]. Exhaustive methods have evaluated the association of all SNP combinations with phenotypes. Wan et al. [27] proposed BOOST, a multistage exhaustive approach that uses bitwise storage technology to speed up logistic regression test calculation. Zhang et al. [28] developed TEAM that calculates contingency tables by introducing a minimum spanning tree structure to detect pairwise SNP interactions. Ritchie et al. [29] used multifactor dimensionality reduction (MDR) to identify epistasis, which reduced the multiple SNP combinations into one dimension with high risk and low risk. It is one of the widely used methods in this field, and many methods have been developed based on it [30–33].

Exhaustive methods can effectively avoid omitting epistasis detection, but it requires massive computation. Stochastic techniques and heuristic searches are examples of search algorithms. The performance of stochastic methods involves random sampling and probability calculation. BEAM was created by Zhang and Liu [34] to find epistasis by partitioning SNPs into three nonoverlapping groups based on their posterior probability using Markov Chain Monte Carlo sampling. Schork et al. introduced EpiMODE [35], which combined the epistasis module idea with a Gibbs sampling strategy. Heuristic search is an approximation search guided by heuristic information that can reduce the search space and find the optimal solution effectively but may be limited by local optimal solutions. EpiACO [36] and AntEpiSeeker [37], both based on ant colony optimization, are examples of this sort of approach. Epi-GTBN [38] is an epistasis search approach that incorporates genetic algorithms to the Bayesian network heuristic search strategy.

SNP epistasis is also detected using machine learning-based approaches such as the neural network [39], support vector machine [40], random forest, or association rules [41]. SNPrule [42] is an epistasis detection method based on learning predictive rules, and by identifying the predictive rules involved in epistasis, higher-order epistasis may be inferred. EpiForest [43], random Jungle [44], and SNPInterforest [45] are examples of random forests that have been used in GWAS. They treated the random forest output as the most crucial variable set.

This paper introduces EpiMIC (epistasis detection via maximal information coefficient), an epistasis detection method that uses the maximal information coefficient (MIC) to identify marker-level interactions of complex diseases in case-control studies [14, 46]. MIC is a good bivariate dependency measure explicitly designed for rapid exploration of almost all types of data relationships, which means it can detect linear, exponential, and cyclical functions. Specifically, it can detect various function types equally, interpret them, and compare them on the same scale. We establish a statistic that utilizes the difference of MIC as an indicator of the occurrence of epistasis and also use the permutation resampling strategy to learn our statistic's empirical distribution. In simulated data sets with a variety of parameters, our method has demonstrated outstanding performance in finding underlying paired epistasis. Its use of WTCCC (Wellcome Trust Case Control Consortium) rheumatoid arthritis (RA) data shows accurate epistasis detection.

2. Materials and Methods

This section describes the EpiMIC statistical framework. The various parameter choices for simulation studies are presented to assess the power to detect type-I error and pairwise epistasis. Then, using the rheumatoid arthritis data set from the WTCCC database, we evaluated the efficiency of our method in a real-world setting.

2.1. EpiMIC

2.1.1. Preliminaries and Notation. Suppose we have n random samples with a collection of p SNPs, then the observed genotypes $\mathscr{R}^p$ can be represented by a $n \times p$ matrix:

$$G = \left[g_{l,i}\right]_{l\in 1\cdots n, i\in 1\cdots p}, \tag{1}$$

where $g_{l,i}$ is the random variable that models the genotype for SNP i of l^{th} sample. It is a categorical variable with three levels denoted by $g_{l,i} \in \{AA, Aa, aa\} = \{0, 1, 2\}$. The homozygote genotypes are represented by AA and aa, whereas the heterozygote condition is represented by Aa. A and a denote the major and the minor alleles of SNP i, respectively. The value indicates the copy number of SNP i's minor alleles. In case-control studies, $y_i \in \{0, 1\}$ is a categorical label where 0 represents a control subject and 1 represents a case subject.

To determine whether there was a statistical interaction between two SNPs in a case-control study, we developed a statistic through the maximal information coefficient to quantify the strength of pairwise epistasis. We then used the permutation resampling strategy to estimate the statistic's empirical distribution. Intuitively, any pair of SNPs may have an original dependency or not without the phenotypic background. EpiMIC tried to capture the conditional dependency between a pair of SNPs under a disease status, which is the task-related correlation. It is based on the idea that because control samples are frequently picked at random, the epistasis pattern in case samples is more

```
Data: Genotype $G_{(n_1+n_2)\bullet p}$, Phenotype $y$, permutation times $m$
Result: The $p$-value of epistasis for each pair of SNPs
1 for $i = 1$ to $p - 1$ do
2   for $j = i + 1$ to $p$ do
3     Apply MIC to cases and all samples for pair of SNPs $(g_i, g_j)$, to calculate
        $\mathrm{MIC}^{\mathrm{all}}_{n_1+n_2}(g_i, g_j)$ and $\mathrm{MIC}^{D}_{n_2}(g_i, g_j)$;
4     Calculate the difference $\Delta\mathrm{MIC}^0$ between $\mathrm{MIC}^{\mathrm{all}}_{n_1+n_2}(g_i, g_j)$ and $\mathrm{MIC}^{D}_{n_2}(g_i, g_j)$ ;
5     for k= 1 to $m$ do
6       Randomly shuffle label $y$, and generate the new data set;
7       Repeat Steps 3 and 4;
8     end
9       The estimated $p$-value of $\Delta\mathrm{MIC}^0$ is the number of $\Delta\mathrm{MIC}^i$, $i = 1, \cdots, m$, larger than
          $\Delta\mathrm{MIC}^0$, divided by $m$.
10  end
11 end
12 Output all the $p$-value for each pair of SNPs.
```

Algorithm 1: EpiMIC.

representative for understanding the underlying disease etiology. If there is no epistasis between two SNPs for disease, their dependence should show nothing significantly different in both cases and all samples; if they interacted under a disease, their dependency should be significantly different in cases and in all samples. Some methods can be used to calculate the dependency between two variables but may be limited by the functional form; for example, Pearson's correlation only measures linear dependence. Hence, we propose instead quantifying them using the maximal information coefficient.

2.1.2. Maximal Information Coefficient. MIC [46] is an efficient measure of reliance for bivariate associations that encompasses a wide range of functional and not functional associations. It has two heuristic properties: generality and equitability. Generality means that if the sample size is sufficient, MIC covers not only certain types of functions but also various interesting associations or functional relationships that are not well modeled by a function, such as a superposition of functions. Equitability means MIC should assign identical ratings to similarly noisy associations of all kinds.

Let $D \subset \mathscr{R}^2$ be a bivariate finite collection of ordered pairs, with x values and y values of D separated into x bins and y bins, respectively. An (x, y) grid is the term given to such a pair of partitions. The distribution $D|_G$ for a grid G can be determined from the data points in D on the cells of G. It is calculated by taking the probability mass in each cell and dividing it by the proportion of data points in D that fall into that cell. Given a constant D, different grids G produce distinct distributions of $D|_G$. With positive integers (x, y), define $I^*(D, x, y)$ as

$$I^*(D, x, y) = \max I\left(D|_G\right), \tag{2}$$

where the maximum is across all grids (x, y) and $I(D|_G)$ represents the mutual information of $D|_G$.

The characteristic matrix $M(D)$ of a bivariate data collection D is an infinite matrix with elements:

$$M(D)_{(x,y)} = \frac{I^*(D, x, y)}{\log \min \{x, y\}}. \tag{3}$$

With Equation (3), the MIC of a bivariate data set D with n samples and a grid size smaller than $B(n)$ is defined as follows:

$$\mathrm{MIC}(D) = \max_{xy<B(n)} \left\{M(D)_{(x,y)}\right\}, \tag{4}$$

where $\omega(1) < B(n) \le O(n^{1-\varepsilon})$ for some $0 < \varepsilon < 1$. The function $B(n)$ upper binds the sizes of the grids over which MIC searches. Usually, its default setting is $n^{0.6}$ because it works well in practice.

To calculate $M(D)$, it is optimized ideally on all possible grids. But in practice, MIC uses dynamic programming algorithms that optimize only a subset of possible grids for computational efficiency, and it seems to be approaching the true value of MIC.

MIC satisfies the following properties:

(i) Symmetry: $\mathrm{MIC}(X, Y) = \mathrm{MIC}(Y, X)$

(ii) Comparability: $\mathrm{MIC} \in [0, 1]$, $\mathrm{MIC} = 0$ denotes that two variables are independent statistically; $\mathrm{MIC} = 1$ implies a strong association

(iii) Generality: MIC could capture a wide range of relationships

(iv) Equitability: MIC is robust to noisy relationships. It provides the same ratings to similarly noisy associations of various sorts

2.1.3. Illustration of the EpiMIC Framework. Assume there are n samples in a case-control study, with n_2 of them being

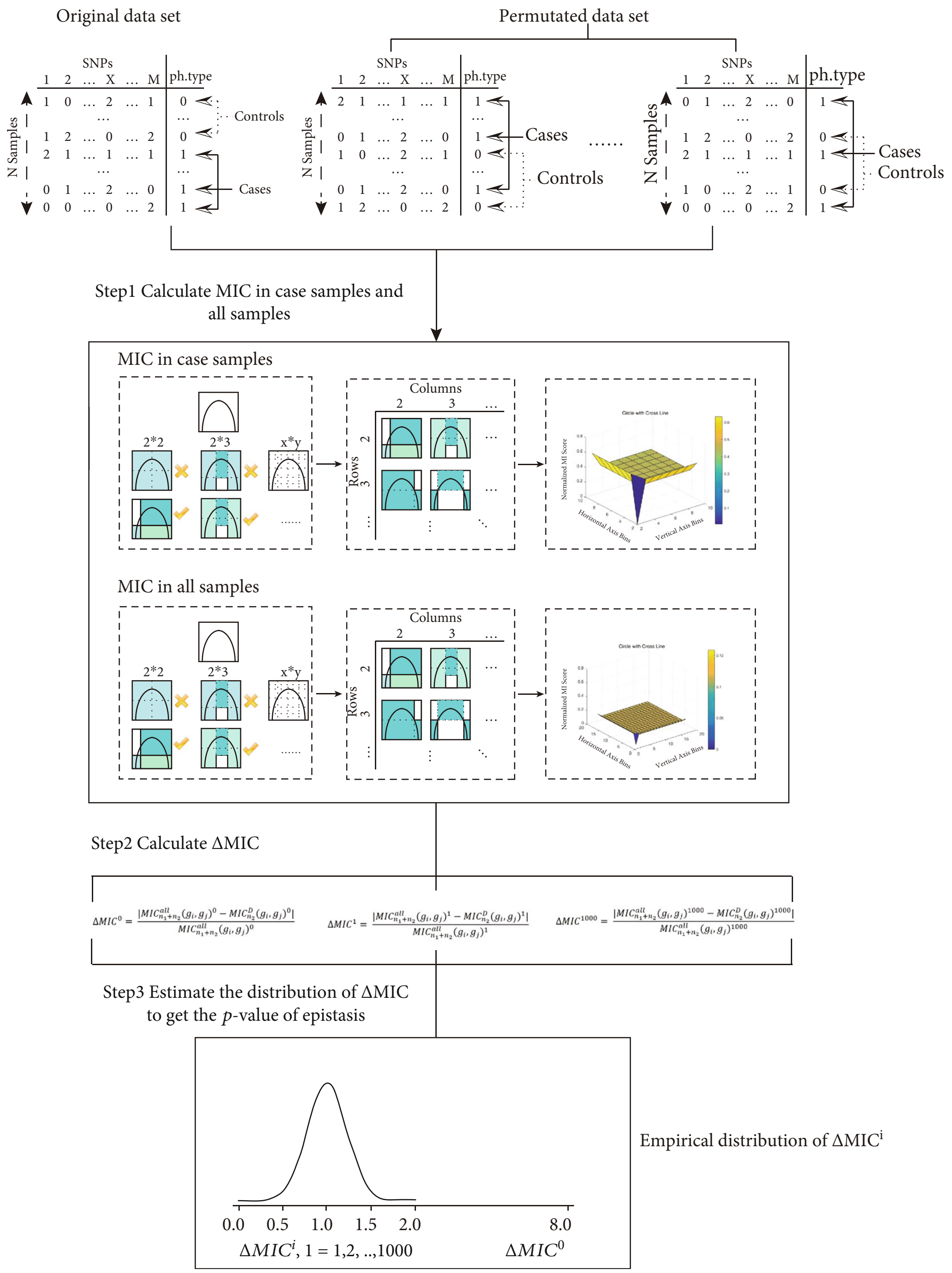
Original data set
Permutated data set
SNPs
1 2 … X … M
ph.type
N Samples
Controls
Cases
Cases
Controls
……
Cases
Controls
Step1 Calculate MIC in case samples and all samples
MIC in case samples
2*2
2*3
x*y
……
Columns
2 3 …
Rows
2 3
Circle with Cross Line
Normalized MI Score
Horizontal Axis Bins
Vertical Axis Bins
MIC in all samples
Step2 Calculate ΔMIC
$\Delta MIC^0 = \frac{|MIC^{all}_{n_1+n_2}(g_i,g_j)^0 - MIC^D_{n_2}(g_i,g_j)^0|}{MIC^{all}_{n_1+n_2}(g_i,g_j)^0}$
$\Delta MIC^1 = \frac{|MIC^{all}_{n_1+n_2}(g_i,g_j)^1 - MIC^D_{n_2}(g_i,g_j)^1|}{MIC^{all}_{n_1+n_2}(g_i,g_j)^1}$
$\Delta MIC^{1000} = \frac{|MIC^{all}_{n_1+n_2}(g_i,g_j)^{1000} - MIC^D_{n_2}(g_i,g_j)^{1000}|}{MIC^{all}_{n_1+n_2}(g_i,g_j)^{1000}}$
Step3 Estimate the distribution of ΔMIC to get the p-value of epistasis
Empirical distribution of $\Delta\text{MIC}^{\text{i}}$
0.0 0.5 1.0 1.5 2.0 8.0
ΔMIC^i, 1 = 1,2, ..,1000
ΔMIC^0

Figure 1: Illustration of the EpiMIC framework for pairwise epistasis detection.

cases. Let $\text{MIC}_n(g_i, g_j)$ be the sample correlation score between the i^{th} SNP and the j^{th} SNP. First, we calculated the $\text{MIC}_n^{\text{all}}(g_i, g_j)$ for all the samples and $\text{MIC}_{n_2}^{D}(g_i, g_j)$ for case samples. Second, we devised a statistic $\Delta\text{MIC} = |\text{MIC}_n^{\text{all}}(g_i, g_j) - \text{MIC}_{n_2}^{D}(g_i, g_j)|/\text{MIC}_n^{\text{all}}(g_i, g_j)$ to compare the MIC in cases and in all samples. ΔMIC denotes how dissimilar the relationship (g_i, g_j) was in cases and across all samples. The greater the ΔMIC, the more likely it is that g_i and g_j interacted.

We wanted to learn the empirical distribution of ΔMIC^0 under the null hypothesis to derive a p value. In this case, we employed a nonparametric permutation strategy: first, shuffled the y with m times, computed ΔMIC by the same procedure described above, and used the resultant sample distribution as an estimate for the distribution of ΔMIC. If the outcome of these m permutations is $\Delta\text{MIC}^1, \cdots, \Delta\text{MIC}^m$, an estimated p value under the null hypothesis is

$$p = \frac{\left|\left\{i : \Delta\text{MIC}^i \geq \Delta\text{MIC}^0\right\}\right|}{m}. \quad (5)$$

In Algorithm 1, we summarized the EpiMIC procedure and showed the whole workflow in Figure 1.

2.2. Simulation Study. To evaluate EpiMIC's ability to control type-I error and detect marker-based pairwise epistasis, we compared EpiMIC with BEAM [34], MDR [29], BOOST [27], and Epi-GTBN [38].

2.2.1. Simulation with GAMETES. The performance of the EpiMIC to detect marker-based, pairwise epistasis was examined in this simulation study. We assigned 10 SNPs to each simulation data set. There were two functional SNPs and eight nonfunctional SNPs among them. To produce the simulated genotype data, we used the freely accessible program GAMETES [47]. This program was created to produce pure and strict epistasis models, which are the most challenging to discover if all n-loci are included in the disease model. Because of this requirement, these models are a desirable gold standard for simulation research on complex epistasis [47, 48].

(1) Type-I Error Evaluation. Type-I error demonstrates a method's capacity to reject the null hypothesis when it is true. We utilized the GAMETES to create two custom disease models without epistasis (Table 1). The baseline odd was denoted by γ. We conducted the simulation for each model 100 times with the following sample size $n \in \{1k, 2k, 3k, 4k, 5k\}$, $\gamma = 1$, and $\theta = 5$. The threshold of significance was fixed at 0.05.

Table 1: Odds table for two models: (a) no effect model with no epistasis between two SNPs; (b) one marginal recessive model with no epistasis between two SNPs.

(a) No effect model

	AA	Aa	aa
BB	γ	γ	γ
Bb	γ	γ	γ
bb	γ	γ	γ

(b) One marginal recessive model

	AA	Aa	aa
BB	γ	γ	γ
Bb	γ	γ	γ
bb	$\gamma(1+\theta)$	$\gamma(1+\theta)$	$\gamma(1+\theta)$

(2) Evaluation of Test Power. The power of a test reflects the likelihood that the procedure will properly accept the alternative hypothesis when the null hypothesis is false. This simulation study used two experimental setups: epistasis models without marginal effects (NME) and epistasis models with marginal effects (ME).

For each parameter setting in the NME scenario, we created 100 data sets. The power under each parameter value was stated as the frequency at which the approach successfully rejects the null hypothesis at a significance level of $\alpha = 0.05$.

(i) We used $h \in \{0.005, 0.01, 0.025, 0.05, 0.1, 0.2\}$ and two distinct minor allele frequencies $(\text{MAF}) \in \{0.2, 0.4\}$ to analyze the influence of heritability h. Five models were developed for each parameter combination, yielding 60 models following Hardy-Weinberg proportions. For all these models, population prevalence was set to 0.2, and the sample size was set to 4,000. The five models were labeled M1 to M5, and they were sorted in general by rising customized odds ratio (COR) using GAMETES. COR is a detectability statistic derived directly from the genetic model. The higher the value, the simpler it is to identify epistasis

(ii) To assess the effect of sample size, we set heritability to 0.025, $\text{MAF} \in \{0.2, 0.4\}$, and prevalence to 0.2, with a sample size of 10,000. Then, using this big data set, 100 data sets were created at random for each of the sample size $n \in \{1k, 2k, 3k, 4k, 5k\}$. In this case, we had a total of 1,000 data sets

In the ME scenario, we generated six models in accordance with Namkung et al. [49]. For each model, 100 replicated data sets with balanced case subjects and control subjects were constructed with a sample size of 4,000 (Table 2).

For BEAM, MDR, BOOST, and Epi-GTBN, let the number of data sets where they identified the epistasis correctly be m_1, then the power can be determined using the following formula:

$$\text{power} = \frac{m_1}{100}. \quad (6)$$

We ran BEAM and Epi-GTBN with the default parameter setting. MDR and BOOST had no parameters to be specific. In EpiMIC, $n^{0.7} \sim n^{0.8}$ is effective experimentally. We use $n^{0.8}$ as a default parameter.

2.3. Experiments Using Data from Rheumatoid Arthritis. To test EpiMIC's capacity to handle true epistasis in a case-control data set, we examined the susceptibility of a series of pairings of SNPs in rheumatoid arthritis (RA), an inflammatory disease characterized by pannus development in synovial joints and cartilage and bone loss. The detailed data set construction can be found in our previous work [48].

Table 2: The detailed information of the six disease models with marginal effects, which included prevalence, MAF, and penetrance for each combination of genotypes.

Models	Prevalence	MAF	Genotypes								
			AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb
Model 1	0.050	0.1	0.061	0.017	0.017	0.017	0.136	0.136	0.017	0.136	0.136
Model 2	0.050	0.1	0.060	0.021	0.021	0.021	0.116	0.116	0.021	0.116	0.116
Model 3	0.046	0.1	0.030	0.080	0.090	0.090	0.010	0.010	0.070	0.040	0.000
Model 4	0.026	0.1	0.030	0.010	0.020	0.010	0.090	0.050	0.020	0.050	0.070
Model 5	0.017	0.1	0.020	0.005	0.020	0.007	0.070	0.001	0.003	0.080	0.090
Model 6	0.052	0.2	0.044	0.066	0.073	0.069	0.021	0.007	0.042	0.073	0.054

Table 3: Type-I error for methods BEAM, BOOST, MDR, Epi-GTBN, and EpiMIC. Sample sizes varied from 1,000 to 5,000 under two disease models: (a) no effect model with no epistasis between two SNPs and (b) one marginal recessive model with no epistasis between two SNPs.

(a) No effect disease model

Methods	Sample size				
	1,000	2,000	3,000	4,000	5,000
BEAM	0	0	0	0	0
BOOST	0	0	0	0	0
MDR	0.03	0.02	0	0.01	0.04
Epi-GTBN	0.01	0	0.02	0.01	0.05
EpiMIC	0.02	0.03	0.05	0.02	0.04

(b) Marginal disease model

Methods	Sample size				
	1,000	2,000	3,000	4,000	5,000
BEAM	0	0	0	0	0
BOOST	0	0	0	0	0
MDR	0.04	0.06	0.06	0.08	0.06
Epi-GTBN	0.06	0.07	0.06	0.08	0.11
EpiMIC	0.03	0.03	0.06	0.05	0.03

Table 4: The statistical power of simulation studies for BEAM, BOOST, MDR, Epi-GTBN, and EpiMIC with $h \in \{0.005, 0.01, 0.025, 0.05, 0.1, 0.2\}$ and MAF $\in \{0.2, 0.4\}$. There are five models for each heritability-MAF combinations. The best-performing approach for each model is shown with a bold font. The results of some heritability-MAF combinations are not listed in the table because all methods under these parameter combinations are 1. These parameter combinations include MAF = 0.2 with $h \in \{0.025, 0.05, 0.1, 0.2\}$ and MAF = 0.4 with $h \in \{0.01, 0.025, 0.2\}$.

MAF	Heritability	Method	Models				
			M1	M2	M3	M4	M5
0.2	0.005	BEAM	0.53	0.95	0.95	0.98	0.97
		BOOST	0.96	**1**	**1**	**1**	**1**
		MDR	0.14	0.84	**1**	**1**	**1**
		Epi-GTBN	0.94	**1**	**1**	**1**	**1**
		EpiMIC	**0.98**	**1**	**1**	**1**	**1**
	0.01	BEAM	1	1	1	1	1
		BOOST	1	1	1	1	1
		MDR	0.34	0.99	1	1	1
		Epi-GTBN	1	1	1	1	1
		EpiMIC	1	1	1	1	1
0.4	0.005	BEAM	0.87	0.93	0.9	0.93	0.93
		BOOST	**1**	**1**	**1**	**1**	**1**
		MDR	0.99	**1**	**1**	**1**	**1**
		Epi-GTBN	**1**	**1**	**1**	**1**	**1**
		EpiMIC	**1**	**1**	**1**	**1**	**1**
	0.05	BEAM	0.76	1	1	0.98	1
		BOOST	1	1	1	1	1
		MDR	1	1	1	1	1
		Epi-GTBN	1	1	1	1	1
		EpiMIC	1	1	1	1	1

3. Results and Discussion

All results were obtained on a workstation equipped with an Intel Xeon CPU E5-2620 v2 @ 2.10 GHz, 96 GB of DDR3, R 4.0.3, and RStudio programming implementation.

3.1. Simulation Study

3.1.1. Type-I Error Evaluation. For type-I error, we set MAF to 0.2 and population prevalence to 0.2, then ranged sample sizes from 1,000 to 5,000. For the no-effect model without epistasis, all the methods tested had a type-I error comparable to the significance level $\alpha = 0.05$ (Table 3(a)). For the disease model without epistasis, but with one marginal SNP, BEAM and BOOST still controlled type-I error, although MDR, Epi-GTBN, and EpiMIC had little inflation. The result implied that we should choose a lower significance level in practical application to reduce the probability of false positive results.

3.1.2. Evaluation of the Power of EpiMIC

(1) The Influence of Heritability. We investigated two types of epistasis disease models to assess the statistical strength of our EpiMIC and the other four methods: epistasis models without marginal effects (NME) and epistasis models with marginal effects (ME).

In the NME scenario, we examined 12 heritability-MAF combinations, with heritability ranging from 0.005 to 0.2

Table 5: Average power for the methods BEAM, BOOST, MDR, Epi-GTBN, and EpiMIC to detect epistasis under 12 heritability-MAF combinations.

MAF	Heritability	Methods				
		BEAM	BOOST	MDR	Epi-GTBN	EpiMIC
0.2	0.005	0.876	0.992	0.796	0.988	0.998
	0.01	1	1	0.866	1	1
	0.025	1	1	1	1	1
	0.05	1	1	1	1	1
	0.1	1	1	1	1	1
	0.2	1	1	1	1	1
0.4	0.005	0.912	1	0.998	1	1
	0.01	1	1	1	1	1
	0.025	1	1	1	1	1
	0.05	0.948	1	1	1	1
	0.1	1	1	1	1	1
	0.2	1	1	1	1	1

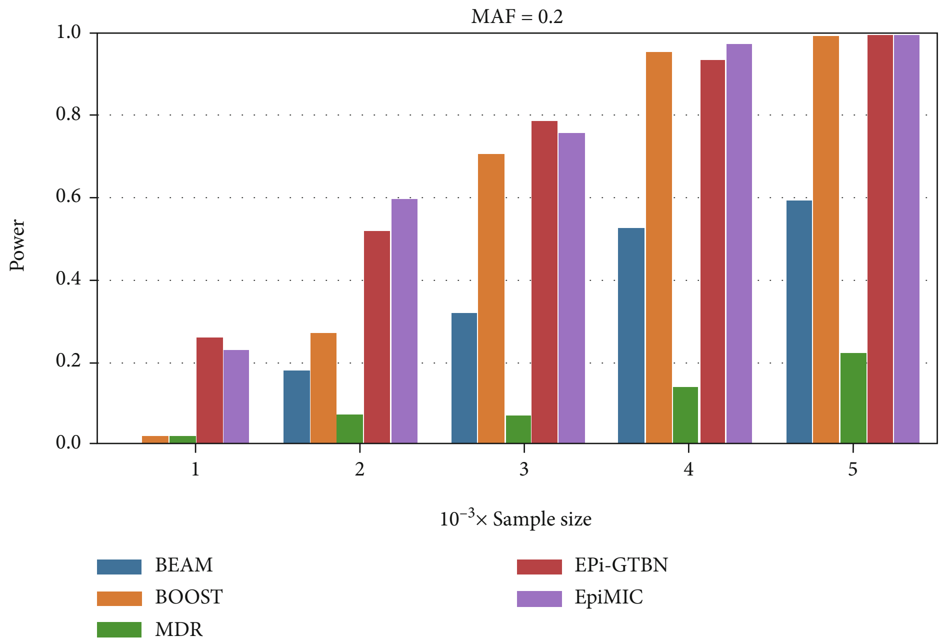

Figure 2: The statistical power of simulation studies for BEAM (blue), BOOST (orange), MDR (green), Epi-GTBN (red), and EpiMIC (purple) under disease model with heritability = 0.005, MAF = 0.2, population prevalence = 0.2, and sample sizes that ranged from 1,000 to 5,000.

(Table 4). Table 4's bold font indicates the best-performing approach in each model for a given heritability-MAF combination. It is worth noting that a higher value suggests better performance. Except for the disease model $M1$ with $h = 0.005$ with MAF = 0.2, EpiMic was slightly better than other methods. For most parameter combinations, EpiMic had the same great performance as BOOST and Epi-GTBN. The statistical power of all the methods achieved 1 when MAF = 0.2 , $h > 0.01$ and MAF = 0.4, $h > 0.005$ except for BEAM.

Heritability had a significant impact on the power of all methods, and the power increased monotonically with an increase in h under a certain MAF (Table 5). Heritability ranged from 0.005 to 0.01, and all methods demonstrated a consistent rising trend (Table 5). The power was also affected by the epistasis SNP pair's minor allele frequencies (MAF). Although BEAM fluctuated under model M1 with MAF = 0.4 and $h = 0.05$, for other methods, the increase in MAF was evident in the improved performance, especially when $h = 0.005$. Heritability is the effect size of epistasis. When it was small, the larger MAF increased the chances for causal genotypic combinations of epistasis SNP pairs to emerge in simulation data sets. For example, for the cases

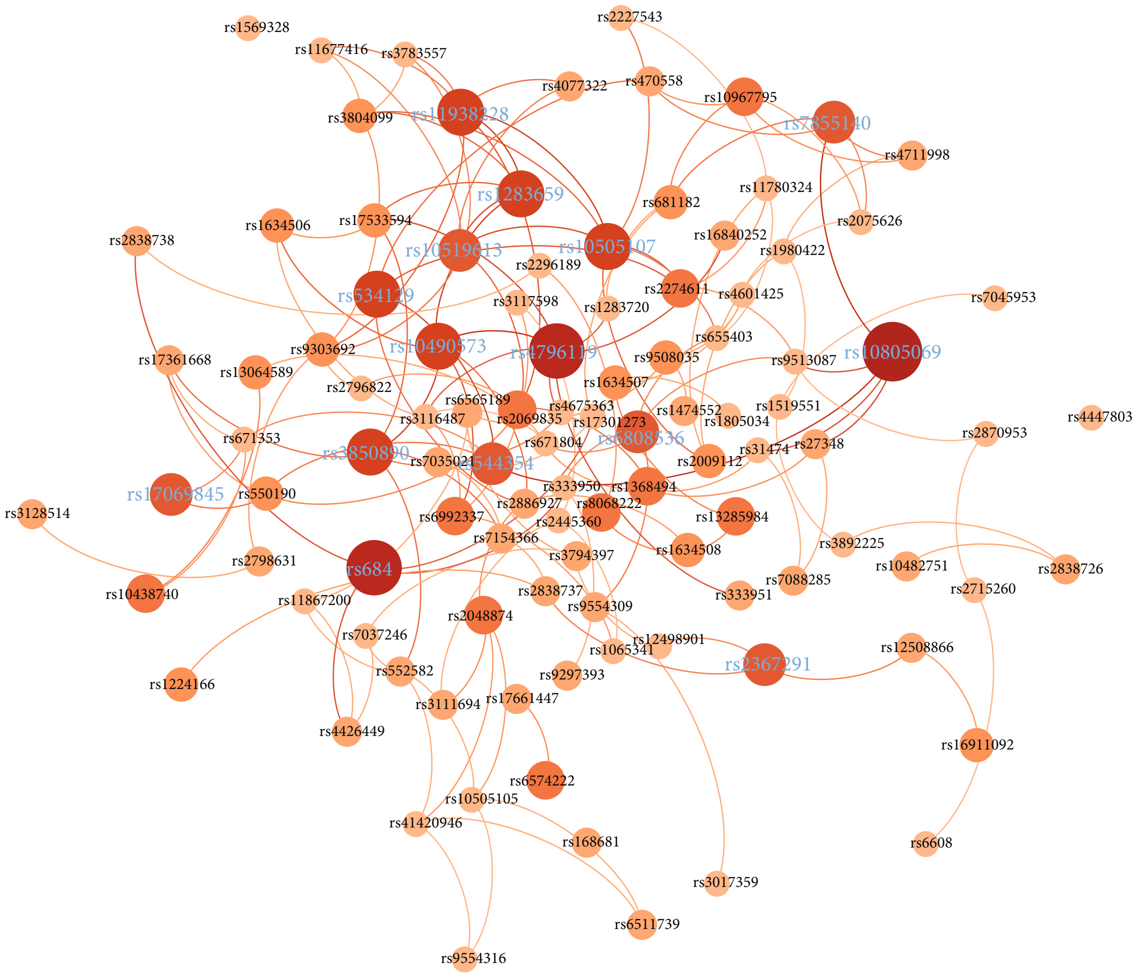

Figure 3: Variant network of rheumatoid arthritis results from the EpiMIC model with identified SNP pairs. The nodes were SNPs, and the edges represented the epistasis relationship. Node size and color reflected the number of epistasis that the node involved in. Edge thickness indicated the maximal information coefficient of SNPs in case samples. The node labels with highlights were the top 15 SNPs ranked by node degree.

of $h = 0.005$, the average power of BEAM was 0.876 with $\mathrm{MAF} = 0.2$, which was lower than 0.912 for $\mathrm{MAF} = 0.4$. Although the performance of the methods under the same model was different, the epistasis detected by the high-power method did not entirely cover epistasis detected by the low-power method. Because these methods were based on different definitions of epistasis, the methods could not simply replace each other; instead, they had a complementary relationship.

It is worth mentioning that, as compared to BEAM or MDR, EpiMIC, BOOST, and Epi-GTBN were more stable for disease models M1 to M5 with varying COR under the same heritability-MAF combination. In the ME scenario, the power to detect epistasis for all methods achieved a 1 that the epistasis disease model with marginal effect was easier to analyze than models without marginal effect.

(2) The Influence of Sample Size. Let the sample size be $n \in \{1k, 2k, 3k, 4k, 5k\}$, with $h = 0.005$, and $\mathrm{MAF} = 0.2$ (Figure 2). As the sample size increased, the power of all methods increased almost monotonically. A larger sample size corresponds to improved performance in all methods.

In conclusion, EpiMIC had superior or comparable performance to detect purely and strictly epistasis in simulation studies, which was the most difficult disease-related patterns with or without marginal effects. EpiMIC benefited from the powerful ability of MIC to capture a wide range of relationships and our designed statistic Δ MIC. If two SNPs interacted under a specific disease model, SNPs showed a more obvious relationship in some cases. Adding control samples decreases the strength of this relationship.

However, the nature of EpiMIC made it likely that it was affected by linkage disequilibrium (LD); because LD is a strong dependency, SNPs in a strong LD block may produce false positives.

3.2. Experiments Using Rheumatoid Arthritis Data. RA is an autoimmune disease in which IL-6, RANK, and $\mathrm{TNF}-\alpha$ are key hereditary risk factors [50]. In the RA investigation, each

Table 6: Detailed information of the top 15 nodes ranked by the node's degree of SNP epistasis network generated using EpiMIC with RA data. The column "corresponding gene" indicates the gene where SNP was located, and the column "gene interaction" shows the genes where the interacting SNPs were located.

rsID	Corresponding gene	Degree	Gene interaction
rs10805069	IL15	12	GM-CSF, Tie2, TLR4, MMP3, FLT1
rs4796119	CCL2	11	M-CSF, CD28, CTLA4, Ang1
rs684	LFA1	10	M-CSF, Ang1, CTSL, RANK, LFA1
rs10490573	CD28	9	CD80, IL15, Ang1, Tie2, CCL2, CCL5, LFA1
rs10505107	Ang1	9	TGFβ, CXCL1, IL15, TLR4, MMP3
rs11938228	CXCL1	9	IL1, IL6, Ang1, APRIL
rs3850890	CD80	9	TLR2, MMP3, IFNγ, AP1, CD28
rs1283659	Ang1	9	CD28, CXCL1, FLT1, CCL2
rs534129	Tie2	9	IL15, Ang1,TLR4, MMP1, MMP3
rs10519613	IL15	8	IL1, IL6, Ang1, APRIL
rs7855140	TLR4	8	IL15, IL17, Tie2, MMP1, IL18
rs544354	IL18	8	IL15, Tie2
rs6808536	CD80	8	M-CSF, Tie2, MMP3, FLT1
rs17069845	RANK	8	TLR2, Tie2
rs2367291	IL15	8	IL8, Tie2, FLT1, RANK, LFA1

unique SNP pair of the hsa05323 pathway was analyzed, yielding $C_{385}^{2} = 73,920$ total pairings for 385 SNPs. We chose 522 results with a significance level $\alpha = 0.005$; after filtering SNP pairs in the same gene, we got 517 epistasis to do the following analysis.

We generated the epistasis network (Figure 3) from 517 epistasis using the network analysis software Gephi, in which the nodes were the SNPs with epistasis and the edge indicated the epistasis relationship. We generated Figure 3 by running the Multigravity Force Atlas algorithm, which prevented nodes from overlapping and controlled the scale of the expansion of the graph while clustering interconnected nodes. The degree of each node represents the number of epistasis that it was involved in, and the edge was weighted by the MIC of pairs of SNPs in case samples. The average degree of the network was 3.009. We filtered the nodes with degree < 5, then ordered the node size and color by its degree. Nodes labeled purple were the top 15 nodes ranked by the node's degree.

Table 6 gives a detailed information of the top 15 nodes, which included their degree, the gene where the SNP was located, and the genes where the interacting SNPs were located. We grouped the interacting SNPs into genes. The higher the node's degree was, the greater the chance that it interacted with more genes. But the number of genes that showed epistasis did not increase monotonically with the degree of a node; due to the different numbers of SNPs that were contained in each gene in the data set and their underlying interval LD pattern, one SNP may be detected as epistatic with multiple SNPs in the LD region from a long gene. From the network, we found some valuable hub genes, such as IL15, CD28, Ang1, Tie2, LFA1, and TLR4, which had at least five interacting genes in the RA pathway. For instance, IL-15 [51], which is a member of the 4 α-helix bundle cytokine family, was detected in the serum of RA patients and synovial fluid and in mouse models of arthritis. In addition, the administration of IL-15 led to the development of severe inflammatory arthritis, indicating that IL-15 may be related to RA treatment. Targeting IL-15 is very critical and valuable.

Table 7: Detailed information of the top 10 epistasis ranked by the MIC in case samples for each pair of SNPs and genes where SNPs were located. The column "Ref" references the literature that showed the regulatory relationship between two genes.

rsID of SNP1	rsID of SNP2	Corresponding gene 1	Corresponding gene 2	Ref
rs4675363	rs1427676	CD28	CTLA4	[52]
rs7537752	rs6574222	M-CSF	FOS	[53]
rs4422395	rs7037246	TLR2	TLR4	
rs13285984	rs1634507	Tie2	CCL4	
rs12089727	rs6808536	MCSF	CD80	
rs2564594	rs1800795	TLR2	IL6	[54]
rs246841	rs266089	GM-CSF	CXCL12	[51]
rs550982	rs1569328	Tie2	AP1	
rs951759	rs266089	Ang1	CXCL12	
rs2256849	rs1474552	FLT1	ITGB2	

We also analyzed the top 10 epistasis ranked by the MIC in case samples for each pair of SNPs (Table 7) and found that five of the top 10 results were supported by prior research [52–55]. For example, the first epistasis was between gene CD80 (rs4675363) and CTLA4 (rs1427676). Costimulatory molecules have a crucial role in the immunoregulatory regulation of T lymphocyte-mediated immunological and inflammatory responses [53]. The best-studied costimulatory signaling pathway was CD28/CTLA4-CD80/CD86. CTLA-4 is a structural homolog of CD28, and it binds the CD80 and CD86 ligands. CTLA-4, on the other hand, has a 20-to-50-fold greater affinity to CD28, which gives a different regulatory function for CTLA to downregulate T cell immunity while allowing CD28 to initiate amplification and to maintain the positive immunity of T cells. T cells were not stimulated abnormally due to the tight and coordinated costimulation signaling pathway of CD28/CTLA4-CD80/CD86.

4. Conclusions

Epistasis between genetic variants is ubiquitous and crucial in uncovering the underlying genetic structure of complex diseases and traits. In this paper, we developed EpiMIC (epistasis detection via maximal information coefficient), which combined maximal information coefficient (MIC) with permutation strategy for case-control studies in GWAS. We transformed the epistasis detection problem by measuring the degree of difference of MIC between pairwise SNPs in cases and in all samples. The method benefits from the powerful ability of MIC to explore various function types equally and to interpret and compare them on the same scale. Because of the weak assumptions about the nature of epistasis and MIC's powerful and practical capacity to capture complicated functional and nonfunctional correlations, our method accurately and effectively recognized additional sorts of interpretable epistasis.

To assess EpiMIC's performance, we conducted simulated and retrospective investigations. For most of the settings tested, EpiMIC's statistical power to detect epistasis was better or equivalent to prior methods, and its power grew monotonically with heritability, MAF, and sample size. Based on a test of type-I error, the method was shown to be stable to sample size. Simulation results also indicated that epistasis detected by the high-power method did not entirely cover epistasis detected by the low-power method. These methods were based on different definitions of epistasis that methods could not simply replace each other but had a complementary relationship. In our analysis of real data, we found several key genes of RA from an epistasis network and significant epistasis that was supported by prior research. We found that local LD inflated EpiMIC's statistical power slightly. SNP pairs in the same gene with a LD pattern were detected as epistasis occasionally. In practice, we suggest filtering out SNP pairs within a local LD structure or select tagSNPs to detect epistasis. Moreover, due to different sequencing coverage, some causal SNPs may be missing in some data sets, and only markers linked to it are found to be epistasis. Therefore, it would be better to locate the interacting genes through marker-based epistasis first and then to combine more biological information to locate the real causal epistasis. In conclusion, EpiMIC is a useful addition to the current toolkit of statistical methods for elucidating epistasis in GWAS case-control studies.

Data Availability

Publicly available data sets were analyzed in this study. This data can be found here: https://www.wtccc.org.uk/info/access_to_data_samples.html.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Conceptualization was performed by Yingjie Guo, Yanpeng Wang, and Lei Xu; data curation was performed by Zhen Liang; formal analysis was performed by Zhian Yuan and Yang Wang; funding acquisition was performed by Yingjie Guo; methodology was performed by Yingjie Guo; project administration was performed by Yanpeng Wang and Lei Xu; resources were secured by Yang Wang; software was secured by Yingjie Guo and Zhian Yuan; supervision was performed by Lei Xu; validation was performed by Yingjie Guo, Zhian Yuan, and Zhen Liang; visualization was performed by Zhen Liang; writing (original draft) was performed by Yingjie Guo; writing (review and editing) was performed by Yang Wang.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, grant number No. 62002243, and the Post-doctoral Foundation Project of Shenzhen Polytechnic China, grant number 6021330009K.

References

[1] A. Buniello, J. A. L. MacArthur, M. Cerezo et al., "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019," *Nucleic Acids Research*, vol. 47, no. D1, pp. D1005–D1012, 2019.

[2] L. A. Hindorff, P. Sethupathy, H. A. Junkins et al., "Potential etiologic and functional implications of genome-wide association loci for human diseases and traits," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 23, pp. 9362–9367, 2009.

[3] R. J. F. Loos, "15 years of genome-wide association studies and no signs of slowing down," *Nature Communications*, vol. 11, no. 1, p. 5900, 2020.

[4] R. J. Klein, C. Zeiss, E. Y. Chew et al., "Complement factor H polymorphism in age-related macular degeneration," *Science*, vol. 308, no. 5720, pp. 385–389, 2005.

[5] T. A. Manolio, F. S. Collins, N. J. Cox et al., "Finding the missing heritability of complex diseases," *Nature*, vol. 461, no. 7265, pp. 747–753, 2009.

[6] A. I. Young, "Solving the missing heritability problem," *PLoS Genetics*, vol. 15, no. 6, article e1008222, 2019.

[7] G. Fang, W. Wang, V. Paunic et al., "Discovering genetic interactions bridging pathways in genome-wide association studies," *Nature Communications*, vol. 10, no. 1, p. 4274, 2019.

[8] M. T. W. Ebbert, P. G. Ridge, and J. S. K. Kauwe, "Bridging the gap between statistical and biological epistasis in Alzheimer's disease," *BioMed Research International*, vol. 2015, 7 pages, 2015.

[9] H. J. Cordell, "Detecting gene-gene interactions that underlie human diseases," *Nature Reviews Genetics*, vol. 10, no. 6, pp. 392–404, 2009.

[10] J. H. Moore, F. W. Asselbergs, and S. M. Williams, "Bioinformatics challenges for genome-wide association studies," *Bioinformatics*, vol. 26, no. 4, pp. 445–455, 2010.

[11] J. Wang, X. Qi, B. Cui, and M. Guo, "A survey of metrics measuring difference for rooted phylogenetic trees," *Current Bioinformatics*, vol. 15, no. 7, pp. 697–702, 2020.

[12] W. Bateson and G. Mendel, *Mendel's Principles of Heredity*, Courier Corporation, 2013.

[13] R. L. Collins, T. Hu, C. Wejse, G. Sirugo, S. M. Williams, and J. H. Moore, "Multifactor dimensionality reduction reveals a three-locus epistatic interaction associated with susceptibility to pulmonary tuberculosis," *BioData Mining*, vol. 6, no. 1, p. 4, 2013.

[14] M. K. Tripathi, M. Yasir, P. Singh, and R. Shrivastava, "A comparative study to explore the effect of different compounds in immune proteins of human beings against tuberculosis: an in-silico approach," *Current Bioinformatics*, vol. 15, no. 2, pp. 155–164, 2020.

[15] O. B. Christiansen, R. Steffensen, H. S. Nielsen, and K. Varming, "Multifactorial etiology of recurrent miscarriage and its scientific and clinical implications," *Gynecologic and Obstetric Investigation*, vol. 66, no. 4, pp. 257–267, 2008.

[16] S. Dasgupta and B. M. Reddy, "The role of epistasis in the etiology of polycystic ovary syndrome among Indian women: SNP-SNP and SNP-environment interactions," *Annals of Human Genetics*, vol. 77, no. 4, pp. 288–298, 2013.

[17] M. B. Taylor and I. M. Ehrenreich, "Higher-order genetic interactions and their contribution to complex traits," *Trends in Genetics*, vol. 31, no. 1, pp. 34–40, 2015.

[18] W. H. Wei, G. Hemani, and C. S. Haley, "Detecting epistasis in human complex traits," *Nature Reviews Genetics*, vol. 15, no. 11, pp. 722–733, 2014.

[19] L. S. Yu, Y. Shi, Q. Zou, S. Wang, L. Zheng, and L. Gao, "Exploring drug treatment patterns based on the action of drug and multilayer network model," *International Journal of Molecular Sciences*, vol. 21, no. 14, p. 5014, 2020.

[20] Y.-J. Tang, Y.-H. Pang, and B. Liu, "IDP-Seq2Seq: identification of intrinsically disordered regions based on sequence to sequence learning," *Bioinformaitcs*, vol. 36, no. 21, pp. 5177–5186, 2021.

[21] R. A. Fisher, "XV.—the correlation between relatives on the supposition of Mendelian inheritance," *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, vol. 52, no. 2, pp. 399–433, 1919.

[22] T. F. Mackay, "Epistasis for quantitative traits in *drosophila*," *Methods in Molecular Biology*, vol. 1253, pp. 47–70, 2015.

[23] T. F. Mackay, "Epistasis and quantitative traits: using model organisms to study gene–gene interactions," *Nature Reviews Genetics*, vol. 15, no. 1, pp. 22–33, 2014.

[24] J. Shao and B. Liu, "ProtFold-DFG: protein fold recognition by combining Directed Fusion Graph and PageRank algorithm," *Briefings in Bioinformatics*, vol. 22, no. 3, article bbaa192, 2021.

[25] L. Sun, G. Liu, L. Su, and R. Wang, "HS-MMGKG: a fast multi-objective harmony search algorithm for two-locus model detection in GWAS," *Current Bioinformatics*, vol. 14, no. 8, pp. 749–761, 2019.

[26] S. Jin, X. Zeng, F. Xia, W. Huang, and X. Liu, "Application of deep learning methods in biological networks," *Briefings in Bioinformatics*, vol. 22, no. 2, pp. 1902–1917, 2021.

[27] X. Wan, C. Yang, Q. Yang et al., "BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies," *The American Journal of Human Genetics*, vol. 87, no. 3, pp. 325–340, 2010.

[28] X. Zhang, S. Huang, F. Zou, and W. Wang, "TEAM: efficient two-locus epistasis tests in human genome-wide association study," *Bioinformatics*, vol. 26, no. 12, pp. i217–i227, 2010.

[29] M. D. Ritchie, L. W. Hahn, N. Roodi et al., "Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer," *American Journal of Human Genetics*, vol. 69, no. 1, pp. 138–147, 2001.

[30] C. H. Yang, H. S. Yang, and L. Y. Chuang, "PBMDR: a particle swarm optimization-based multifactor dimensionality reduction for the detection of multilocus interactions," *Journal of Theoretical Biology*, vol. 461, pp. 68–75, 2019.

[31] F. Abegaz, F. Van Lishout, J. M. Mahachie John et al., "Epistasis detection in genome-wide screening for complex human diseases in structured populations," *Systems Medicine*, vol. 2, no. 1, pp. 19–27, 2019.

[32] X. Fu, L. Cai, X. Zeng, and Q. Zou, "StackCPPred: a stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency," *Bioinformatics*, vol. 36, no. 10, pp. 3028–3034, 2020.

[33] L. Cai, L. Wang, X. Fu, C. Xia, X. Zeng, and Q. Zou, "ITP-Pred: an interpretable method for predicting, therapeutic peptides with fused features low-dimension representation," *Briefings in Bioinformatics*, vol. 22, no. 4, 2021.

[34] Y. Zhang and J. S. Liu, "Bayesian inference of epistatic interactions in case-control studies," *Nature Genetics*, vol. 39, no. 9, pp. 1167–1173, 2007.

[35] N. J. Schork, W. Tang, X. Wu, R. Jiang, and Y. Li, "Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy," *PLoS Genetics*, vol. 5, no. 5, 2009.

[36] Y. Sun, J. Shang, J. X. Liu, S. Li, and C. H. Zheng, "epiACO - a method for identifying epistasis based on ant colony optimization algorithm," *BioData Mining*, vol. 10, p. 23, 2017.

[37] Y. Wang, X. Liu, K. Robbins, and R. Rekaya, "AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm," *BMC Research Notes*, vol. 3, no. 1, p. 117, 2010.

[38] Y. Guo, Z. Zhong, C. Yang et al., "Epi-GTBN: an approach of epistasis mining based on genetic Tabu algorithm and Bayesian network," *BMC Bioinformatics*, vol. 20, no. 1, p. 444, 2019.

[39] D. Wang, Z. Zhang, Y. Jiang et al., "DM3Loc: multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism," *Nucleic Acids Research*, vol. 49, no. 8, article e46, 2021.

[40] L. Chen, J. Li, and M. Chang, "Cancer diagnosis and disease gene identification via statistical machine learning," *Current Bioinformatics*, vol. 15, no. 9, pp. 956–962, 2020.

[41] Y. Shang, L. Gao, Q. Zou, and L. Yu, "Prediction of drug-target interactions based on multi-layer network representation learning," *Neurocomputing*, vol. 434, pp. 80–89, 2021.

[42] X. Wan, C. Yang, Q. Yang, H. Xue, N. L. S. Tang, and W. Yu, "Predictive rule inference for epistatic interaction detection in genome-wide association studies," *Bioinformatics*, vol. 26, no. 1, pp. 30–37, 2010.

[43] R. Jiang, W. Tang, X. Wu, and W. Fu, "A random forest approach to the detection of epistatic interactions in case-control studies," *BMC Bioinformatics*, vol. 10, Suppl 1, p. S65, 2009.

[44] L. De Lobel, P. Geurts, G. Baele, F. Castro-Giner, M. Kogevinas, and K. V. Steen, "A screening methodology based on random forests to improve the detection of gene-gene interactions," *European Journal of Human Genetics*, vol. 18, no. 10, pp. 1127–1132, 2010.

[45] M. Yoshida and A. Koike, "SNPInterForest: a new method for detecting epistatic interactions," *BMC Bioinformatics*, vol. 12, no. 1, p. 469, 2011.

[46] D. N. Reshef, Y. A. Reshef, H. K. Finucane et al., "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011.

[47] R. J. Urbanowicz, J. Kiralis, N. A. Sinnott-Armstrong, T. Heberling, J. M. Fisher, and J. H. Moore, "GAMETES: a fast, direct algorithm for generating pure, strict, epistatic models with random architectures," *BioData Mining*, vol. 5, no. 1, p. 16, 2012.

[48] Y. Guo, H. Cheng, Z. Yuan, Z. Liang, Y. Wang, and D. du, "Testing gene-gene interactions based on a neighborhood perspective in genome-wide association studies," *Frontiers in Genetics*, vol. 12, article 801261, 2021.

[49] J. Namkung, K. Kim, S. Yi, W. Chung, M.-S. Kwon, and T. Park, "New evaluation measures for multifactor dimensionality reduction classifiers in gene–gene interaction analysis," *Bioinformatics*, vol. 25, no. 3, pp. 338–345, 2009.

[50] V. Majithia and S. A. Geraci, "Rheumatoid arthritis: diagnosis and management," *The American Journal of Medicine*, vol. 120, no. 11, pp. 936–939, 2007.

[51] X. K. Yang, W. D. Xu, R. X. Leng et al., "Therapeutic potential of IL-15 in rheumatoid arthritis," *Human Immunology*, vol. 76, no. 11, pp. 812–818, 2015.

[52] L. Sánchez-Martín, A. Estecha, R. Samaniego, S. Sánchez-Ramón, M. Á. Vega, and P. Sánchez-Mateos, "The chemokine CXCL12 regulates monocyte-macrophage differentiation and RUNX3 expression," *Blood*, vol. 117, no. 1, pp. 88–97, 2011.

[53] W. Liu, Z. Yang, Y. Chen et al., "The association between CTLA-4, CD80/86, and CD28 gene polymorphisms and rheumatoid arthritis: an original study and meta-analysis," *Frontiers in Medicine*, vol. 8, 2021.

[54] S. H. Mun, P. S. U. Park, and K. H. Park-Min, "The M-CSF receptor in osteoclasts and beyond," *Experimental & Molecular Medicine*, vol. 52, no. 8, pp. 1239–1254, 2020.

[55] B. L. Diaz, R. M. Sommerfelt, A. J. Feuerherm, T. Skuland, and B. Johansen, "Cytosolic phospholipase A2 modulates TLR2 signaling in s," *PLoS One*, vol. 10, no. 4, 2015.