

Transferring entropy to the realm of GxG interactions

Paola G. Ferrario and Inke R. König

Corresponding author. Inke R. König, Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany. Tel.: ++49 451 500 50610; Fax: ++49 451 500 50604; E-Mail: inke.koenig@imbs.uni-luebeck.de

Abstract

Genome-wide association studies are moving to genome-wide interaction studies, as the genetic background of many diseases appears to be more complex than previously supposed. Thus, many statistical approaches have been proposed to detect gene–gene (GxG) interactions, among them numerous information theory-based methods, inspired by the concept of entropy. These are suggested as particularly powerful and, because of their nonlinearity, as better able to capture nonlinear relationships between genetic variants and/or variables. However, the introduced entropy-based estimators differ to a surprising extent in their construction and even with respect to the basic definition of interactions. Also, not every entropy-based measure for interaction is accompanied by a proper statistical test. To shed light on this, a systematic review of the literature is presented answering the following questions: (1) How are GxG interactions defined within the framework of information theory? (2) Which entropy-based test statistics are available? (3) Which underlying distribution do the test statistics follow? (4) What are the given strengths and limitations of these test statistics?

Key words: information theory; entropy; genetic interactions; estimation

Introduction

Genome-wide association studies are moving to genome-wide interaction studies, as the genetic background of many diseases appears to be more complex than previously supposed. Specifically, complex diseases are not caused by one single but by many genetic variants that potentially interact with each other. Unless these variants also have strong main effects, they are unlikely to be identified using standard single-locus tests. Thus, many statistical approaches have been proposed to detect gene–gene (GxG) interactions, or, more precisely, interactions between genetic variants.

These include traditional approaches, such as regression methods [1], and more novel approaches, such as multifactor dimensionality reduction (MDR) [2, 3, 4] or random forest [5]. Moreover, information theory-based methods are emerging, inspired by the concept of entropy and other therefrom derived measures. While entropy and related measures have already been introduced in the middle of the past century [6, 7], Jakulin and

coauthors applied these quantities to interactions studies in 2003 [8, 9], and Moore *et al.* [10] were the first to introduce entropy to the realm of GxG interactions in 2006. The information theory-based methods are suggested as particularly powerful and, because of their nonlinearity, as better able to capture nonlinear relationships between quantities. Generally, these approaches enjoy a good reputation owing to their model freedom and their capability to quantify or even amplify nonlinear relationships through the basic function of entropy. However, the proposed entropy-based estimators differ to a surprising extent—they even disagree on the basic definitions of interactions from the information theory point of view. Also, although entropy-based measures for interactions are available, it is not always clear how to construct a proper statistical test based on these measures, i.e. how a test statistic can be defined and which distributions it follows under the null and/or alternative hypothesis, respectively.

In summary, the following questions are open: (1) How are GxG interactions defined within the framework of information theory? (2) Which entropy-based test statistics are available?

Paola G. Ferrario is Researcher at the Max Rubner-Institut, Karlsruhe, Germany. She is interested in nonparametric estimation and genetics of complex diseases.

Inke R. König is Professor for Medical Biometry and Statistics at the Universität zu Lübeck, Germany. She is interested in genetic and clinical epidemiology and published over 190 refereed papers.

Submitted: 28 June 2016; Received (in revised form): 16 August 2016

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

(3) Which underlying distributions do the test statistics follow?
 (4) What are the given strengths and limitations of these test statistics? In addition, for practical applications, it is important to understand for which study designs a given entropy-based estimator can be used, and for which test statistics computationally reasonable implementations are available.

To answer these questions, we performed a systematic review of the literature. The results of the search motivate the following article structure: First, in the ‘Definitions and methods’ section, we summarize some fundamental definitions of genetic interaction and information theory, and describe in details our systematic search of the literature. The main results are listed in Table 2, which tabulates a visual summary of all the results, grouped by the underlying information theory-based quantity. The ‘Results for studies on binary traits in unrelated individuals’ and ‘Results for specific study designs’ sections describe the main results in detail. In particular, the ‘Results for studies on binary traits in unrelated individuals’ section is devoted to information theory estimators for the study of a binary trait in unrelated individuals. Here, entropy-based estimators are presented with their strengths and limitations, with information about underlying distributions of test statistics and implemented software, if available. The ‘Results for specific study designs’ section introduces further entropy-based estimators that were proposed for specific study designs, such as family studies. Finally, the ‘Conclusions’ section gives a final evaluation of the methods and some general suggestions on how to choose between the presented estimators when searching for genetic interactions.

Definitions and methods

GxG interactions

We begin by specifying how GxG interactions may be defined. For the general notation, we consider a diallelic genetic variant G such as a single nucleotide polymorphism (SNP) coded as 0, 1 or 2 for the number of minor alleles. Throughout the following, genetic variants are denoted by G_i , G_j , G_1, \dots , where i , j , l take values in the total sample of the genetic variants.

Moreover, we mostly consider for simplicity a binary phenotype P coded as 0, 1 for controls and cases, respectively, so that ‘phenotype’ refers in the first instance to the presence or absence of disease. However, some results will also be shown for the situation of quantitative traits (in the ‘Association with quantitative traits’ section).

Throughout the literature, heterogeneous definitions of interactions exist in the context of genetics in the fields of biology, medicine, biochemistry and biostatistics. It is beyond the scope of this work to detail these, for more information we refer the reader to [1, 11]. For our aims, let a GxG interaction be present as soon as a genetic variant influences the ‘effect’ that another genetic variant has on a trait of interest. This approach has not the ambition to be a precise definition of GxG interaction but to provide a general framework to comprehend many different situations. In specific settings, the magnitude of an interaction directly depends on how the effect is modeled, e.g. using an additive or a multiplicative model. Moreover, to distinguish this from haplotype effects, we assume for the sake of convenience that the two variants are located on different chromosomes or at least are physically distant from each other. Finally, let us assume throughout the whole article that Hardy–Weinberg Equilibrium holds for the control population. Indeed, in the searched literature this is not always specified.

Different study designs have been suggested to be used to detect GxG interactions, which are reviewed in detail in the literature [11]. In a simplified way, we distinguish between the use of data from family members and from independent individuals such as in the classical case-control or cohort design. For interactions, in some situations it may be possible to use a case-only design in which the association between the two variants in only the cases indicates GxG interaction.

Information theory definitions

In this section we summarize the fundamental definitions in information theory to prepare the subsequent definition of entropy-based estimators. For this, we consider two discrete random variables X_1 and X_2 , with potential states $i = 1, \dots, m$ and $j = 1, \dots, m'$, respectively. The probability mass function of one variable X is given by p , where $p(x_i) = p_i = P\{X = x_i\}$, and the joint probability of two variables is denoted by $p_{ij} = P\{X_1 = x_i, X_2 = x_j\}$. The marginal probabilities are $p_i = P\{X_1 = x_i\}$ and $p_j = P\{X_2 = x_j\}$, for X_1 and X_2 , respectively. Figures 1–6 illustrate six fundamental definitions giving their formal mathematical expressions and visualizing them by Euler–Venn diagrams as established procedure for illustration, see for instance [12].

The most fundamental concept is the ‘Shannon entropy’, first introduced by Shannon [6], which aims to quantify the uncertainty within a random variable (see Figure 1). Formally, it is defined as minus the logarithm of the probability distribution of this variable. This leads to the following properties (cf. [13]):

- i. The entropy is zero when one outcome is certain.
- ii. The larger the uncertainty about a variable, the larger the entropy.
- iii. Given two probability distributions defined over the same range, the larger entropy will be assigned to the wider and flatter distribution.
- iv. For independent variants the definition of entropy by the logarithm is especially convenient because in this case it is additive.

While the ‘joint entropy’ (Figure 2) gives the uncertainty of two random variables simultaneously, the ‘conditional entropy’ (Figure 3) of a random variable given another variable quantifies the uncertainty of a random variable when the other variable is known. Finally, the ‘mutual information’ (Figure 4) of two variables represents the reduction of uncertainty of one variable, owing to knowledge of the other one. Similar as for entropy, the mutual information can be conditioned on a third variable yielding the conditional mutual information (CMI).

A slight modification of the Shannon entropy is given by the so-called ‘Rényi entropy’ defined as

$$H_\lambda(X_1) = 1/(1 - \lambda) \log_2 \left(\sum_{i=1}^n p_i^\lambda \right),$$

introduced first by Rényi [14]. It can be considered a generalization of the Shannon entropy because they are equivalent in the special case that λ tends to 1. Of particular interest are the cases of $\lambda \rightarrow 0, 1, \infty$, with an explicit interpretation from the information theory point of view, as de Andrade and Wang [15] point out. Specifically, when $\lambda \rightarrow 1$ the Rényi entropy coincides with the Shannon entropy; when $\lambda \rightarrow 0$ the Rényi entropy is the logarithm of the size of the support of X_1 ; when $\lambda \rightarrow \infty$ it is called min-entropy, with the property that it is never larger than the

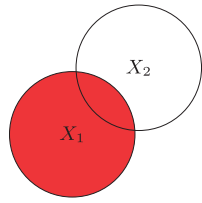


Figure 1. $H(X_1)$: Entropy of a random variable $X_1 := \mathbb{E} \left[\log_2 \left(\frac{1}{p(X_1)} \right) \right] = - \sum_{i=1}^m p_i \log_2(p_i)$

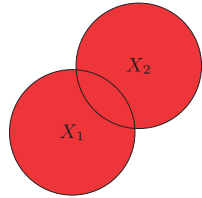


Figure 2. $H(X_1, X_2)$: Joint entropy of the two random variables X_1 and $X_2 := \mathbb{E} \left[\log_2 \left(\frac{1}{p(X_1, X_2)} \right) \right] = - \sum_{i=1}^m \sum_{j=1}^{m'} p_{ij} \log_2(p_{ij})$

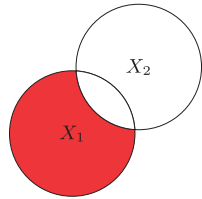


Figure 3. $H(X_1|X_2)$: Conditional entropy of the variable X_1 given the variable $X_2 := \mathbb{E} \left[\log_2 \left(\frac{p(X_1, X_2)}{p(X_2)} \right) \right] = - \sum_{i=1}^m \sum_{j=1}^{m'} p_{ij} \log_2 \left(\frac{p_{ij}}{p_j} \right) = H(X_1, X_2) - H(X_2)$

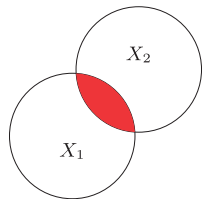


Figure 4. $I(X_1, X_2)$: Mutual information of the variables X_1 and $X_2 := \mathbb{E} \left[\log_2 \left(\frac{p(X_1, X_2)}{p(X_1)p(X_2)} \right) \right] = - \sum_{i=1}^m \sum_{j=1}^{m'} p_{ij} \log_2 \left(\frac{p_{ij}}{p_i p_j} \right) = H(X_2)H(X_2|X_1) = H(X_1)H(X_1|X_2)$

Shannon one. For these special cases, many properties have been derived in the information theory literature.

For three variables, it is necessary to distinguish between the total extent of dependence among the three variables, which is the so called ‘total correlation information’ (TCI; Figure 5), and the amount of information common to all variables but not present in any subset alone, which is the ‘three-way interaction information’ (3WII; Figure 6).

Both the TCI and the 3WII are generalizable to k-way interaction information (KWII) and the k-way TCI, respectively, and these quantities were introduced by McGill [7]. The KWII represents the gain or loss of information owing to the inclusion of additional variables in the model. It quantifies interactions by representing the information that cannot be obtained without observing all k variables at the same time. Different from the bivariate case in which the mutual information can be at most nonnegative, the KWII can also become negative. A positive value of the KWII is termed synergy between variables, while a negative value is named redundancy. In this sense, a synergy

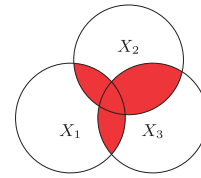


Figure 5. $TCI(X_1, X_2, X_3)$: Total correlation information of three random variables $H(X_1) + H(X_2) + H(X_3) - H(X_1, X_2, X_3)$

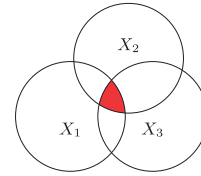


Figure 6. $3WII(X_1, X_2, X_3)$: Three-way interaction information of three random variables $H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3) - H(X_1) - H(X_2) - H(X_3) - H(X_1, X_2, X_3)$

quantifies the positive gain in information of k – 1 variables owing to the knowledge of the kth one, while a redundancy means that the kth variable did not add information to the previous ones. However, synergy and redundancy do not have unique definitions, as we specify again later in the ‘Synergy and redundancy’ section.

Systematic search of the literature

In May 2015 different bibliographic databases were systematically reviewed drawing on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement [16] to identify information theory-based quantities suggested for detecting GxG interactions. Articles were selected that describe the development or methodologically discuss entropy-based estimators to detect GxG interactions. Thus, we did not consider articles

- in which entropy-based estimators were used to detect genetic association, i.e. main effects of association between a single genetic variant and a phenotype,
- or if they presented only an application of entropy-based estimators on real data.

Moreover, because of the focus of this work we did not include entropy-based strategies for other genetics topics, such as feature selection, gene clustering, gene regulatory network construction or visualization. Also, publications were excluded if they were not written in English.

The complete search process was documented, and Table 1 details the search strategy. Owing to the plurality of terms addressing GxG interactions, a number of searches with different keywords were carried out in PubMed. On the other side, for the search in Google Scholar we tried to limit the number of false-positive findings by using only the most precise search terms. In this case, the results were sorted by their relevance, and we considered only the first 365 findings.

Results

In the search, 29 articles were identified as shown in Figure 7. These may be grouped into the following sections with respect to entropy-based definitions of interactions: Interactions

Table 1. Hits from the four systematic searches

Search	Keywords combination	Database	Date	Hits
1	(entropy AND genetic) AND interaction (with activated filter limited to humans)	PubMed (www.ncbi.nlm.nih.gov/pubmed)	23 June 2015	51
2	entropy 'gene-gene interactions' (excluding patents and citations)	Google Scholar (https://scholar.google.de/)	23 June 2015	680
3	epistasis entropy (limited to humans)	PubMed	2 June 2015	18
4	(entropy AND gene) AND interaction (limited to humans)	PubMed	28 May 2015	67

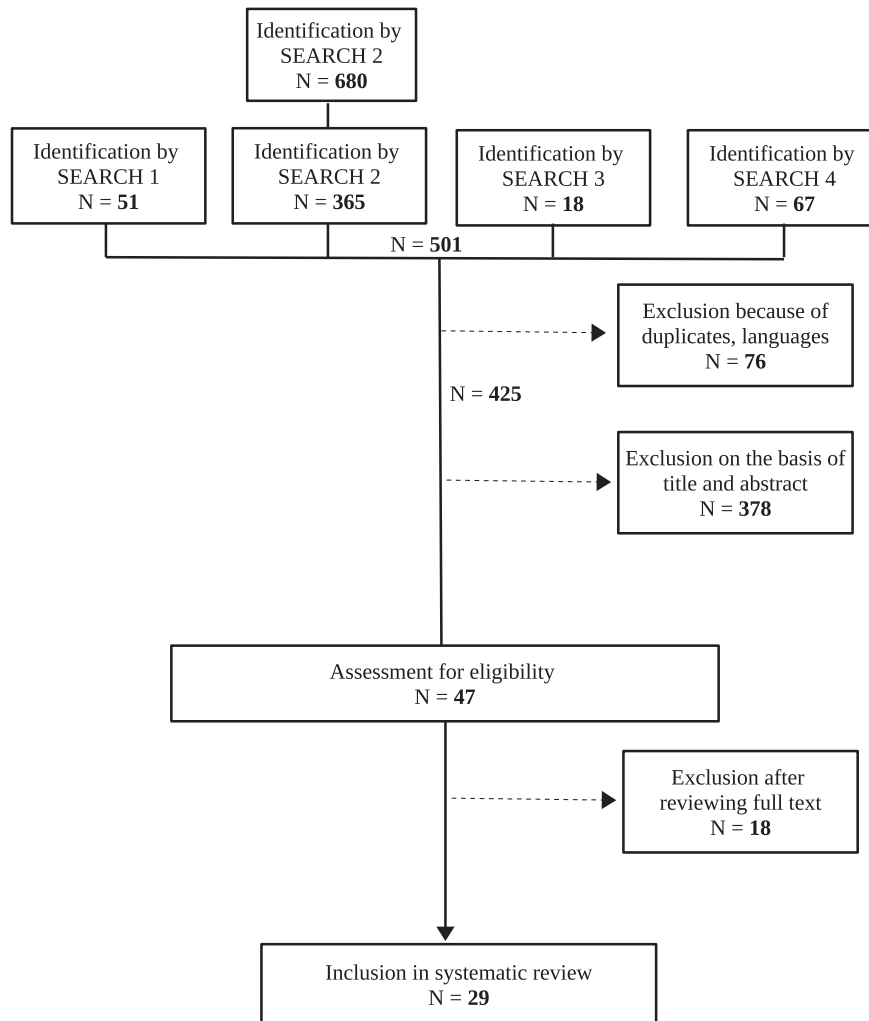


Figure 7. Flow diagram of the search process.

defined with the help of mutual information (Figure 4) are illustrated in the 'Pairwise interactions: Information gain', 'Pairwise interactions: Relative information gain' and 'Synergy and redundancy' sections; interactions defined with the help of three/ k -way interactions (Figure 6) are described in the 'Third order interactions: 3WII, TCI and PAI' and 'Interactions of higher order' sections; finally, interactions defined with the help of other approaches will be given in the 'Rényi entropy' and 'Maximum entropy conditional probability modeling' sections. For an overview, Table 2 lists all findings giving the entropy-based quantities with literature references, the entropy-based definitions of interactions, the availability of a test statistic, of simulation results as well as of an implementation. Specifically,

references are ordered in the table by appearance; estimators used in more than one reference are listed repeatedly if availability of a test statistic, simulations or implementation differ in the corresponding publications.

Results for studies on binary traits in unrelated individuals

Pairwise interactions: information gain

Four of the 29 articles of the systematic search deal with estimators to detect pairwise interactions that are based on a so-called information gain (IG).

Table 2. Results from the systematic literature search

Entropy-based quantity	Reference	Quantity to estimate	Test statistics	Simulations	Implementation
Information gain	[17]	$IG_{Fan} := I_{cases}(G_i, G_j) - I(G_i, G_j)$	Yes	Yes	Yes
	[18, 19]	$IG_{Chen} := I(G_i, G_j P) - I(G_i, G_j)$	Yes	Yes	No
	[20]	$IG_{IGENT} := H(P) - H(P (G_i, G_j))$	Yes	Yes	Yes
Conditional mutual information	[21]	$GenoCMI := \sum_{i=1}^m \sum_{j=1}^{m'} \sum_{p \in \{0,1\}} P(G_i, G_j, P) \log \left(\frac{P(G_i, G_j, P)}{P(G_i P)P(G_j P)} \right)$	Yes	Yes	No
Relative information gain	[22]	$RI_{G_{Dong}} := \frac{H(P) - H(P G_i, G_j)}{H(P)}$	No	Yes	Yes
	[23]	$RI_{G_{Yee}} := \frac{H(P) - H(P G_i, G_j)}{H(P)}$	Yes	Yes	No
Three-way (k-way) Total correlation Information	[24]	$ES(G_i, G_j) := \frac{\min(H(G_i, H(G_j)) - H(G_i, G_j))}{\min(H(G_i), H(G_j))}$	Yes	Yes	Yes
	[20]	$RI_{IGENT} := \frac{H(P) - H(P G_i, G_j)}{H(P)}$	Yes	Yes	Yes
	[17]	$3WII(G_i, G_j, G_1)_{cases} - 3WII(G_i, G_j, G_1)$	Yes	Yes	Yes
Strict information gain	[25, 26, 27]	$TCI(G_i, G_j, G_1)_{cases} - TCI(G_i, G_j, G_1)$	Yes	Yes	Yes
	[28, 29, 30, 31]	$3WII_{Chanda} := 3WII(G_i, G_j, P)$	No	Yes	No
	[32]	$TCI_{Chanda} := TCI(G_i, G_j, P)$	No	Yes	No
phenotype-associated information	[25, 26, 27]	$IG_{strict}(G_i, G_j, G_1, P) := I(G_i, G_j, G_1, P)$	Yes	Yes	Yes
	[30, 31]	$-\max\{IG(G_i, G_1, P), 0\} - \max\{IG(G_j, G_1, P), 0\} - \max\{IG(G_i, G_j, P), 0\} - I(G_i, P) - I(G_j, P) - I(G_1, P)$	Yes	Yes	Yes
Synergy	[33]	$PAI(G_i, G_j, G_1, P) = TCI(G_i, G_j, G_1, P) - TCI(G_i, G_j, G_1)$	No	Yes	No
	[34, 35]	$Sym(G_i, G_j, P) := I(G_i, G_j, P) - [I(G_i, P) + I(G_j, P)]$	No	No	Yes
Rényi entropy	[15]	$Sym(G_i, G_j, P) := I(G_i, G_j, P) - [I(G_i, P) + I(G_j, P)]$	No	Yes	Yes
		$S_{\frac{S_i(G_i, G_j)_{cases}}{S_i(G_i, G_j)_{controls}}}$	Yes	Yes	No
Maximum entropy conditional	[36, 18]	where $S_i(G_i, G_j) := (H_x(G_i) + H_x(G_j)) - H_x(G_i, G_j)$	Yes	Yes	No
Probability models	[37, 15]	$H(P G_i, G_j)$	Yes	Yes	No
Case-only design	[31]	$H(P)$	Yes	Yes	No
Quantitative trait locus studies	[38]	$3WII_{Chanda}$	No	Yes	Yes
Family studies	[39]	$ID_{Ignac} := \frac{I(G_i, G_j, P)}{\max\{H(G_i P), H(G_j P)\} - \max\{H(G_i), H(G_j)\}}$	No	Yes	Yes
	[40]	IG_{IGENT}	Yes	Yes	No
		$I(G_1, G_2)$	No	Yes	Yes

Consider first the mutual information of two arbitrary genetic variants, i.e. $I(G_i, G_j)$ (the general definition was given in the ‘Information theory definitions’ section and Figure 4). In a case-control setting, Fan et al. [17] subtracts the mutual information of two genetic variants estimated in the cases from the same quantities estimated in the controls. If the disease prevalence is small, the mutual information of the controls will be a good approximation of the mutual information of the general population. In this way, the IG of the two markers in the presence of a disease is defined as

$$IG_{\text{Fan}} := I_{\text{cases}}(G_i, G_j) - I(G_i, G_j).$$

By estimating all probabilities emerging in IG_{Fan} by counts, the corresponding final test statistic for testing for interactions is defined as the estimated IG_{Fan} , normalized by a specific quantity Λ , of variance type. The closed expression of the estimator can be found in the Appendix. The test statistic is asymptotically centrally or non-centrally chi-square distributed under the null or the alternative hypothesis, respectively, with one degree of freedom.

The merit of this article lies not only in the clear definition of second-order genetic interactions from the information theory point of view, but also in the construction of a statistical test complete with formulations of null and alternative hypotheses and a proper test statistic with a corresponding distribution. Since 2015, also an implementation of the estimator is available, and the corresponding R code can be found on <http://www.nichd.nih.gov/about/org/diphr/bbb/software/fan/Pages/default.aspx>

As a reviewer pointed out, the IG_{Fan} can be interpreted as contrasting correlations between genetic predictors in cases with those in controls. In this sense, it has a strong link with a ‘case-only’ design (compare also the ‘Case-only design’ section).

In an alternative approach, the information gain IG_{IGENT} suggested by Kwon et al. [20] for detecting interactions is defined as the entropy of the phenotype minus the conditional entropy of the phenotype, given two genetic variants, i.e.

$$IG_{\text{IGENT}} := H(P) - H(P|(G_i, G_j)).$$

A corresponding estimator with its asymptotic distribution is given in Kwon et al. [20] citing results by Goebel et al. [41]. The estimated IG_{IGENT} asymptotically follows a gamma distribution under the null hypothesis of independent variants. The most important contribution of Kwon et al. [20] lies in the freely available and fast implementation called IGENT, which is written in C++ (<http://statgen.snu.ac.kr/software/igent/>). IGENT can be used for an exhaustive as well as for a stepwise search of interacting pairs, depending on whether every possible pair is systematically tested or whether a genetic variant is admitted for pair building only if it shows a main effect. IGENT can be seen as an association test while allowing for interactions, as it calculates the entropy of a phenotype twice, first per se and second given an interacting pair.

The IG introduced by Su et al. [19] is inspired by Moore et al. [10], and it is defined as the CMI of a pair of variants given the phenotype minus the mutual information of this pair, i.e.

$$IG_{\text{Su}} = I(G_i, G_j|P) - I(G_i, G_j).$$

This quantity there is called interaction gain rather than IG. It has a similar structure as IG_{Fan} , where the idea is to evaluate the correlation between genetic variants given (or not) the disease information. The authors emphasize that their estimate requires neither main effects nor any specific genetic model (i.e.

additive, recessive, etc) to identify an effect. Moreover, for detecting interactions, they follow neither an exhaustive nor a stepwise approach, but rather they introduce a strategy for parallelization: First, genetic variants located on the same chromosome are divided into two groups depending on whether they are in a gene or between genes. Then, chunks of SNP pairs are formed by pairing SNPs within the same gene, pairing SNPs in different genes, pairing SNPs in a gene with intragenic SNPs and pairing intragenic SNPs on different chromosomes. These chunks are then tested for interaction at the same time (parallelization), estimating their interaction gain IG_{Su} (replacing probabilities by counts). After this, variants in linkage disequilibrium are discarded, and the remaining pairs are reduced by cutting at a given threshold value. By synchronous calculations of many interaction gains in parallel, the entire test procedure is speeded up, while maintaining accuracy.

Chen et al. [18] introduce the same IG as mentioned above, i.e. $IG_{\text{Chen}} = IG_{\text{Su}}$. However, they do not explicitly describe the estimation of this quantity. Instead, the main objective was a comparison of the performance of different methods for detecting interactions, including MDR, logistic regression (LR) and an estimation based on IG_{Chen} . These competitors were evaluated with regard to type I error rate, power and computational complexity. For the IG_{Chen} -based estimates, the authors conclude that they successfully detect interactions with strong main effects but miss many interacting variants at an acceptable rate of false positives. However, this behavior is not significantly worse than that of the other analyzed methods. Moreover, as expected, the power of the IG_{Chen} tests varies under different genetic models as a function of penetrance, minor allele frequency, linkage disequilibrium and marginal effects; again, the other methods did not significantly differ in their behavior. In particular, the magnitude of the main effect influences the power of the tests. In summary, the authors emphasize that the IG_{Chen} estimate can detect some ground-truth SNPs but has only modest power to detect the entire set of interacting SNPs.

With a similar idea, Zuo et al. [21] show that an estimator simply based on $I(G_i, G_j, P)$ is not able to recognize interaction effects. In particular, when main effects at both markers are large, then the inflation of type I error is unacceptably large with too many false-positive results. To solve these problems, the authors introduce a modification replacing mutual information by CMI that yields the following quantity:

$$\text{GenoCMI} = \sum_{i=1}^m \sum_{j=1}^{m'} \sum_{P \in \{0,1\}} P(G_i, G_j, P) \log \left(\frac{P(G_i, G_j|P)}{P(G_i|P)P(G_j|P)} \right),$$

which will be estimated with the help of the genotype frequencies, penetrances and disease prevalence.

The authors argue that the inflation of the type I error was considerably reduced by their strategy. Finally, the authors analyze the type I error depending on the disease prevalence or the case/control ratio, coming to the conclusion that neither the prevalence nor the case/control ratio influence the type I error considerably.

Finally, we remark that many articles propose the calculation of information-gain-type quantities for feature selection and other aspects in the context of genetic analyses; however, these are beyond the topic of this review.

Pairwise interactions: relative information gain

Our literature search identified four articles that deal with pairwise interactions based on the so-called relative information gain (RIG).

For this, Yee et al. [23] proposed to normalize the IG (of type IG_{IGENT}) with regard to the overall entropy of the phenotype:

$$RIG = \frac{H(P) - H(P|G_i, G_j)}{H(P)}.$$

This RIG, sometimes also called normalized mutual information (NMI, see also [23]), quantifies the proportion of information contained in the interacting variables G_i and G_j that influences the phenotype P . Again, probabilities will be estimated by counts leading to a test statistic of a log-likelihood ratio type with asymptotical χ^2 distribution. Moreover, the authors generate new data sets by repeated shuffling of the phenotypes by fixed genotypes. Using mean and sample standard deviation from the permuted data sets, they standardize the RIG, thus enabling a comparison of the results. The RIG is implemented in the freely available software IGENT as well [20].

The RIG by Dong et al. [22] has a similar structure as the RIG above. In particular, the ‘Entropy-based SNP-SNP interaction method’ (ESNP2) is developed here to detect GxG interactions with the help of the RIG, while the extension ESNP2-Mx enables in addition a best fit to a genetic model. The program, implemented in Java, is free for download (<http://www.biosino.org/papers/esnp2/>).

An extension of the ESNP2 method can be found in Li et al. [42], which describes the ‘Gene-based Information Gain Method’. This extension considers interactions between grouped variants, typically grouped by genes, covering the case that the two genes have different length.

A slightly different approach is taken by Chattopadhyay et al. [24], where the strategy is a combination of different methods for quantifying interactions, to take the advantages of every one of them. The used methods are a Gini score (typically used in Classification and Regression Trees [43]), the Absolute Difference of genotype Probabilities from cases and controls (APD) score (used in MDR, [44]) and an Entropy Score (ES). Focusing here in particular on the entropy score, it is defined as

$$ES(G_i, G_j) := \frac{\min\{H(G_i), H(G_j)\} - H(G_i, G_j)}{\min\{H(G_i), H(G_j)\}}$$

and will be estimated as usual by replacing probabilities by counts. A standardized version is expressed by

$$\frac{ES(G_i, G_j) - \bar{ES}}{\sigma(ES)},$$

where \bar{ES} and $\sigma(ES)$ are the mean and standard deviation of the entropy score calculated over all $\binom{m}{2}$ possible pairwise interactions for m genetic markers, respectively. Finally, the three standardized scores (Gini, APD, entropy) are added to a so-called Z-sum score, for that also the principal component is calculated (the necessity of which is unfortunately not being made clear). The Z-sum score as well as its principal component yield the quantity for detecting disease-associated GxG interactions. The authors evaluate the performance of this estimator under different genetic scenarios and provide a user-friendly program named RASSUN (RAnked Summarized Scores Using Nonparametric-methods). RASSUN is written in R and free for download (<http://www.csjfan.ibtms.sinica.edu.tw/eag/programlist/rassun/rassun.html>).

Third-order interactions: 3WII, TCI and PAI

Estimators for third-order interactions are the topic of 9 of the 29 results from the systematic literature search.

As already described in the ‘Information theory definitions’ section (compare Figures 5 and 6), for up to three interacting variants it is possible to distinguish between the amount of information common to all three attributes (TCI) and the total amount of information common to all three attributes but not present in any subset (3WII). The use of both of these quantities to detect third-order interactions is therefore the natural extension of the use of the IG to detect pairwise interactions. In particular, for three genetic variants, one can express 3WII by

$$3WII := -H(G_i) - H(G_j) - H(G_l) + H(G_i, G_j) + H(G_i, G_l) + H(G_j, G_l) - H(G_i, G_j, G_l)$$

and TCI by

$$TCI := H(G_i) + H(G_j) + H(G_l) - H(G_i, G_j, G_l).$$

Fan et al. [17] consider the interaction or the total correlation between three genetic variants, i.e. $3WII_{Fan} = 3WII(G_i, G_j, G_l)$ or $TCI_{Fan} = TCI(G_i, G_j, G_l)$, respectively, and then express their effects on a phenotype by the differences

$$3WII(G_i, G_j, G_l)_{cases} - 3WII(G_i, G_j, G_l)$$

or

$$TCI(G_i, G_j, G_l)_{cases} - TCI(G_i, G_j, G_l)$$

respectively. Similar to the situation of pairwise interactions, these differences will be estimated in the cases and in the controls and normalized by specific matrices of variance type. The closed expression of the test statistics has a similar construction as T_{IG} (see Appendix) and is therefore not shown. For details, we refer the reader to [17], pp. 12, 13.

Although most authors agree on the general structure of 3WII and TCI, but they disagree with respect to the variables to choose in the general definition. In fact, 3WII and TCI as described in [25, 26, 27, 28, 29, 30, 31], other than in [17], are given by

$$3WII_{Chanda} := 3WII(G_i, G_j, P)$$

and

$$TCI_{Chanda} := TCI(G_i, G_j, P),$$

where Chanda et al. [27] explain the TCI_{Chanda} as ‘the information that cannot be obtained without observing all variables and the phenotype at the same time’. Because the latter group treats $3WII_{Chanda}$ and TCI_{Chanda} as measures for third-order interactions, we list them in this section. However, although they involve three variables, they quantify pairwise interactions with respect to the genetic variants because the third variable is the phenotype, and phenotype and the genetic variants are treated interchangeably. Therefore, a more complex situation is assumed, where not only a number of genetic variants but also the genetic variants and the phenotype are interacting with each other; this represents a crucial difference to the more classical model of genetic variants interacting in their effect on a phenotype. If the phenotype represents the presence or absence of a disease, the model would imply that the presence or absence of disease could affect the status of a genetic variant.

A further expansion of the concept of TCI was given by the same groups [25, 26, 27, 30, 31] in the form of the phenotype-associated interaction information (PAI). Specifically, PAI is given by the difference between the TCI including the phenotype as a variable and the TCI excluding the phenotype, i.e.

$$\text{PAI}(G_i, G_j, G_l, P) = \text{TCI}(G_i, G_j, G_l, P) - \text{TCI}(G_i, G_j, G_l).$$

Some theoretical properties of TCI and PAI are given by Tritchler et al. [30]. Moreover, Chanda et al. [27] present an algorithm, called CHORUS, based on PAI to detect the GxG interactions on quantitative traits. This will be described in detail in the 'Association with quantitative traits' section.

To guard against false-positive results, Hu et al. [32] suggest a novel approach for measuring pure three-way interactions after removing the one-way and two-way effects. All lower-order effects are subtracted from the total IG, including the main effects of the three attributes and all pairwise synergies between them, leading to

$$\begin{aligned} \text{IG}_{\text{strict}}(G_i, G_j, G_l, P) &:= I(G_i, G_j, G_l, P) - \max\{IG(G_i, G_l, P), 0\} \\ &\quad - \max\{IG(G_j, G_l, P), 0\} - \max\{IG(G_i, G_j, P), 0\} \\ &\quad - I(G_i, P) - I(G_j, P) - I(G_l, P), \end{aligned}$$

where their $IG(G_i, G_l, P)$ is defined as $I(G_i, G_l, P) - I(G_i, P) - I(G_l, P)$.

Interactions of higher order

Both 3WII and TCI can be extended to the KWII and total correlation, respectively, which can be used to detect interactions of higher than third order. Similar to the lower-order situations, Fan et al. [17] compare the estimated KWII and total correlation in the cases and in the general population, respectively, to detect k interacting variants associated with a disease.

Furthermore, Chanda et al. [45] introduced a KWII initially as a metric for visualizing GxG interactions. Owing to its specific aim of visualization, that article is not directly relevant for this review, but Chanda et al. [25] subsequently construct an algorithm called AMBIENCE to detect higher order interactions by using KWII and PAI. In particular, AMBIENCE requires as input the number θ of combinations retained in each search iteration and the number of total iterations, τ . To explain the algorithm in a simplified example, suppose we have $n=10$ genetic variants, and we fix θ equal to 2 and τ equal to 5. This means that we are interested in interactions of fifth order, and we want to retain maximally the two 'relative best results'. AMBIENCE will start by calculating the PAI for each of the 10 variants. Then, the $\theta=2$ best results (i.e. the two highest PAIs) will be retained, and for these variants interactions of third, fourth until τ =fifth order will be calculated. AMBIENCE hence delivers $\theta \cdot \tau = 2 \cdot 5$ combinations ranked by PAI.

Another algorithm, AMBROSIA ([29]), reuses the results from AMBIENCE, i.e. the combinations with highest PAI, and tries to decide which of these combinations essentially explain the phenotype, discarding the redundant ones. Moreover, Sucheston et al. [28] compare these results with other common methods such as MDR, concluding that the information theory-based methods have a considerably higher power. However, Lee et al. [46] show recently that higher power often comes at the cost of lower specificity, so that some signal is erroneously identified. We address this problem again in the conclusions.

Based on KWII, Shang et al. [47] developed a software named EpiMiner, which uses a three-stage approach for detecting and also visualizing GxG interactions. This software is available on <https://sourceforge.net/projects/epiminer/files/>. In its first stage, KWII is calculated by supposedly replacing probabilities by frequencies. A previously fixed number of variants is then passed on to the second stage, either user-specified or based on classification by support vector machines. In stage two, permutation tests are conducted on the selected variants sequentially to search for GxG interactions, and the results are ranked by the P -values. The third stage is then reserved for a visualization of the results.

Knights and Ramanathan [48] address the problem of over-dispersed count data, translated in a Poisson-distributed phenotype. The authors use an estimator of KWII type and compare its results with those from a Poisson regression. A Web site with software written in Java is available (http://pharmsci.buffalo.edu/computational_software/murali_1/download/). An estimator is not given explicitly but only the definition of entropy for a Poisson distributed phenotype.

Finally, Brunel et al. [40] consider interactions of higher order by calculating the mutual information between a set of markers with the phenotype, where the set of markers is determined by the following algorithm. First, the set contains just one marker, which is significantly associated with the phenotype, and then a new marker is added to the set; the mutual information calculates whether the new marker adds new information. If yes, a further marker is added to the set; if not, the marker is removed. In this sense, the authors speak about forward and backward steps, depending on whether any further marker is left in the set of is removed.

Synergy and redundancy

Four of the 29 findings from the systematic search deal with estimators based on synergy and redundancy. As already stressed in the 'Information theory definitions' section, other than the mutual information the KWII can also become negative, so that up to three variables one can consider synergy and redundancy between variables. However, these concepts do not have a unique definition and interpretation, which is described in the following.

Anastassiou [33] introduces a synergy defined as

$$\text{Syn}(G_i, G_j, P) = I(G_i, G_j, P) - [I(G_i, P) + I(G_j, P)].$$

Here, Syn quantifies the positive or negative 'gain' of two variables owing to the knowledge of a third one. In case of a positive gain of information, there is said to be a synergy between the variants; in case of a loss of information, a redundancy. Based on this definition, Anastassiou [33], Hu et al. [32] and Moore and Hu [35] propose a generalization of the synergy to three genetic variants. However, this generalization is neither trivial nor unique, and various alternatives are proposed by these authors.

Curk et al. [34] provide an algorithm and a software for detection of interactions by synergy based on $\text{Syn}(G_i, G_j, P)$. To save computation time, an exhaustive approach is avoided in that a heuristic is introduced that involves a threshold for identification of the 'best' low-order interactions before searching for interactions of higher order. The software tool can be found on <http://snpsyn.biolaab.si>.

Moore and Hu [35] use synergy and redundancy for summarizing and visualizing interactions. The synergy of many genetic

variants is used for building epistasis networks, and these networks are visualized by the open-source software ViSEN (Visualization of Statistical Epistasis Networks) [49], following Hu et al. [50], who were the first to introduce networks to visualize entropy-based measures of GxG interactions.

Rényi entropy

A single article deals with estimators based on the Rényi entropy [15]. In particular, a joint Rényi entropy of two genetic variants taken together and the sum of two Rényi entropies of the single genetic variants is considered, calculated on the basis of their marginal probabilities:

$$S_{\lambda}(G_i, G_j) := (H_{\lambda}(G_i) + H_{\lambda}(G_j)) - H_{\lambda}(G_i, G_j).$$

Under the null hypothesis of no interaction effect, the two loci are independent and $S_{\lambda}(G_i, G_j)$ equals zero. To include the phenotype, $S_{\lambda}(G_i, G_j)$ is calculated separately in the cases and in the controls. In particular, the ratio

$$\frac{S_{\lambda}(G_i, G_j)_{\text{cases}}}{S_{\lambda}(G_i, G_j)_{\text{controls}}}$$

is expected to be close to zero if the interacting pair (G_i, G_j) and the phenotype are independent. In this way, estimating the Rényi entropies by replacing probabilities by frequencies, a ratio test is used to test for association of the interacting genetic variants with the disease. The power of the ratio test is said to be low, especially when the marginal effects are strong. For the specific case-only study design, a similarly structured procedure is given to obtain greater power, and this will be addressed in the 'Case-only design' section. Another possibility to increase power considered by the authors is to tune the Rényi entropy by the parameter λ . In fact, they show by simulations that an appropriate choice of λ could decidedly improve the power. Because different choices of λ are what distinguish the Rényi from the Shannon entropy, it can be concluded that it is reasonable not to restrict the investigations to the Shannon entropy. In particular, the tuning of λ should amplify the true difference between two populations to make this true difference more detectable. However, because the true difference between the populations may depend on allele frequencies and other possibly unknown factors, this implies that the optimal λ depends also on partly unavailable information; it is therefore necessary to test and identify the best λ in every new situation.

Maximum entropy conditional probability modeling

Based on the principle of maximum entropy by Jaynes [51], Maximum Entropy Conditional Probability Models (MECPM) is described in its general form by Miller et al. [36], p. 2479: 'When building a probability model, one should agree with all known information while remaining maximally uncertain with respect to everything else'. Miller et al. [36] transfer this idea to genetics replacing the 'known information' by the pairwise probability of disease presence and a particular genotype for a given genetic variant. Maximum uncertainty is obtained by selecting the probability distribution of the phenotype that maximizes Shannon's entropy function but still ensures agreement with the known information ([36], p. 2479).

Chen et al. [18] implemented MECPM as well as seven other methods (including IG [10]) comparing their performances by the number of truly and falsely discovered markers. They found that MECPM performs well in particular by detecting

interactions with moderate effects and at an acceptable rate of false positives. Moreover, the authors generally conclude that the power of all the tests varies as a function of the penetrance, minor allele frequency, linkage disequilibrium and marginal effects.

Results for specific study designs

As described above, the majority of entropy-based methods for detecting interactions pertains to a specific study design, namely, the analysis of unrelated individuals with regard to a binary phenotype, and the phenotype is associated with genotypes of diallelic genetic variants. In addition to that, a few articles were concerned with deviations from this design, and these will be described in the following.

Case-only design

Two of the 29 results of the systematic review deal explicitly with estimators conceived for a case-only design. In this design, an association between the two genetic variants within the cases can be interpreted as a GxG interaction if the prevalence of the disease is low and if the investigated genetic variants are independent from each other in the general population. The recognized advantages of a case-only design compared with the more common case-control design are that (i) a smaller sample size is required and (ii) it might be possible to eliminate selection bias by avoiding to select controls in the first place.

For this constellation, Kang et al. [37] generally express genetic interactions by the difference between joint entropy and entropy by the marginal probabilities under the assumption of no interaction. In addition, they also present a generalization for interactions of higher order as well as a modification to express the magnitude of interaction on a normalized scale, using the ratio rather than the difference between entropies.

The main test for association between genetic variants and phenotype they introduce is defined as follows. Let N be the total sample size of cases, n the number of cases carrying a specific genotype combination h on the w loci of interest. W is the total number of observed genotype combinations on these loci, i.e. $W \leq 3^w$. The entropy of the phenotype, $\hat{H}(P)$, is calculated by

$$\hat{H}(P) = \frac{-\sum_{i=1}^W \frac{n}{N} \log \frac{n}{N}}{-\sum_{i=1}^W \frac{1}{W} \log \frac{1}{W}}$$

if $W > 1$, or $\hat{H}(P) = 0$ if $W = 1$.

As a simplified calculation example, consider that on three loci of interest, we have a total sample size of 100 cases, of whom 97 carry the genotype combination 2, 2, 2, and we have three cases with the genotype combination 1, 2, 2. In this situation $N = 100$ and $W = 2$. For $w = 1$, $n = 97$, and for $w = 2$, $n = 3$. The entropy of the phenotype is therefore

$$\hat{H}(P) = \frac{-\frac{97}{100} \log \frac{97}{100} - \frac{3}{100} \log \frac{3}{100}}{-2 \frac{1}{2} \log \frac{1}{2}}$$

The corresponding test statistic is

$$2N(1 - \hat{H}(P)) \log(W),$$

which is asymptotically central χ^2 distributed under the null hypothesis of no association with $W - 1$ degree of freedom. No proof or citations of this asymptotic result is given, but extensive simulations and real-data examples are presented.

Generally, this approach is complete with a quantity to measure interaction, hypotheses to be tested, corresponding test statistics and asymptotic results. Furthermore, tests for interactions and tests for association when allowing for interactions are presented. However, the approach relies on strong assumptions that may or may not be fulfilled. Moreover, the case-only design can be highly sensitive in case of departure from the independence assumption between genetic variants, as e.g. pointed out in [52].

As mentioned in the 'Rényi entropy' section, de Andrade and Wang [15] also propose a case-only design. In fact, the authors describe that their test statistic for the case-only design is the same as in [37] as soon as $\lambda = 1$, i.e. when the Rényi entropy becomes equal to the Shannon entropy. For $\lambda \neq 1$ the asymptotical distribution of the test statistic is not known and will be approximated by de Andrade and Wang [15] with the help of Monte Carlo simulations.

Association with quantitative traits

We identified three articles that explicitly deal with estimators conceived for association with quantitative traits instead of a dichotomous phenotype. First, Chanda *et al.* [27] extended the previously published algorithm AMBIENCE [25] to a new algorithm called CHORUS, designed specifically for quantitative traits that are normally distributed. Subsequently, CHORUS was extended to the algorithm SYMPHONY in [31], which is designed for vectors of quantitative traits. The principle of both algorithms is the same: Calculate the first order PAI and retain the θ combinations with the highest PAI. For these θ combinations, then calculate the PAI of next order and repeat this procedure up to order τ . Finally, calculate the KWII for all these $\theta \cdot \tau$ winning combinations, which render the test statistic. Thus, CHORUS performs a stepwise search for interacting pairs associated with a multivariate normally distributed phenotype in the presence of main effects.

In a more general approach, Yee *et al.* [39] introduce non-parametric entropy estimates that renounce any assumptions about the distribution of the quantitative phenotype or about its regularity. In particular, the entropy estimators are inspired by sample-spacing techniques of Miller and Fisher [53], and they were modified for specific challenges with small samples in specific combinations, which may occur in the situation of combining loci with small minor allele frequencies.

Analogously to the approach for a binary phenotype [23] that was described above (section 'Pairwise interactions: Relative information gain'), the difference between the entropy and the conditional entropy yields an IG of type IG_{IGENT} , which is then modified and standardized to a RIG. Finally, replacing the entropy and the conditional entropy by their sample-spacing estimates, a test statistic is given for interacting variants associated with a quantitative phenotype. The plausibility of the proposed estimator was examined by simulations, where the phenotype was simulated to follow nine different distributions. Moreover, its performance was compared with two MDR variants, and it was applied to real data. The results of MDR and of spacing entropy were comparable.

Finally, in an approach that is applicable to both binary and quantitative traits, Ignac *et al.* [38] propose an 'information distance' that is defined as

$$ID_{\text{Ignac}} := \frac{I(G_i, G_j|P)}{\max\{H(G_i|P); H(G_j|P)\}} - \frac{I(G_i, G_j)}{\max\{H(G_i); H(G_j)\}}$$

using definition (4) and Equations (2) and (3) there, p. 2.

For this, first the conditional mutual information of two genetic variants taken together given the phenotype is normalized by the maximum between the two conditional entropies of each genetic variant alone given the phenotype. Then a corresponding normalized quantity is calculated for the mutual information (without conditioning by the phenotype), and these two NMI values are subtracted from each other. Thus, ID_{Ignac} takes on values between -1 and 1 , where a negative value indicates that the variants are redundant, whereas a positive value indicates synergy. Three types of permutation tests are presented that test different hypotheses on the presence or absence of main effects together with interaction effects. Finally, the optimal ratio of cases to controls for planning new studies may be quantified based on simulation results, and commented Matlab code is available.

Family studies

In the literature search, no article was identified that presents a method to explicitly deals with family data. However, for the study of sib-pairs, Brunel *et al.* [40] propose to associate the phenotypic and genetic similarities in sib-pairs instead of phenotypes and genotypes in unrelated individuals. While they suggest this in the context of mutual information estimation in the algorithm Mutual Information Statistical Significance, this idea can be adopted in other methods readily.

Conclusion

The systematic literature search identified 29 relevant articles that present mostly different information theory-based estimators and tests. Given this large number, it is obvious that they cannot be treated as a single method but fundamentally differ from each other. In fact, in many cases the estimators are different or even contradictory regarding the basic definition of genetic interactions from the information theory point of view. This raises the following aspects that should be tackled in near future.

First, it would be desirable to find a harmonic definition of interactions based on entropy leading to 'sufficient' estimators. This might be possible by systematically identifying which definitions are redundant, overlapping or contradicting. The first steps in this direction have been made by Lee *et al.* [46] who theoretically compared the relationship between four IG quantities and the interactions detected by a LR model.

Second, the systematic literature search showed that in many cases it is not clear how to construct a proper statistical test based on the proposed measures. A simple replacement of probabilities by frequencies often yields biased or even inconsistent estimators, or extremely slowly converging ones. Therefore, estimators should be sought that are both 'consistent' and if possible 'unbiased' (see also [54]).

Third, the systematic literature search showed that often the underlying distributions of the test statistics under the null or the alternative hypothesis are unknown. Thus, estimators whose 'asymptotic behaviors' are investigated are required.

Finally, the estimators have in many instances still to be adapted to the practical situation of genetic studies. Specifically, extensions remain necessary to account for genotyping errors, missing genotypes, phenocopies or genetic heterogeneity, as already pointed out in 2011 by Fan *et al.* [17].

Thus, we still agree with Fan *et al.* [17], even after 5 years, that a lot of work still needs to be done to understand high order interactions. Bearing these aspects in mind, we tentatively give

the following recommendations, when working with an information theory-based estimator.

First, clarify which types of interactions are defined by the chosen estimators. Consider that stepwise searches (i.e. search for interactions only when main effects have been identified) are computationally much simpler but leave open the question about the genuinity of the signal as well as the possibility of interactions without main effects. Be aware that some interactions could be described completely also by a classical regression model or identified by a classical χ^2 -independent test.

Second, because replacement of probabilities by frequencies represents often just a first 'naive' approach, choose estimators for which a test statistic is already derived or at least simulated (compare Table 2).

Third, consider that some estimators require a new implementation if they are used for genome-wide data or for interactions of higher order, to obtain an estimation in reasonable time. Also, consider that not every implemented estimator is usable for working with high-dimensional data.

Fourth, because of all these previous considerations, it seems pragmatic to first obtain estimators for interactions of second order, before trying to estimate interactions of third or even of higher order.

Funding

German Federal Ministry of Education and Research (BMBF, grant # 01ZX1313J to I.R.K.). German Research Foundation (DFG, grant # KO2240/-1 to I.R.K.).

Key Points

- For detection of GxG interactions, numerous information theory-based methods have been proposed.
- However, these do not agree on the basic definitions of interactions and differ to a surprising extent in their construction or aims.
- Hence, different estimators may serve different purposes, and the selection of a suitable method is supported by this review.

References

1. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* 2009;**10**:392–404.
2. Gola D, Mahachie John JM, van Steen K, et al. A roadmap to multifactor dimensionality reduction methods. *Brief Bioinform* 2016;**17**(2):293–308.
3. Ritchie MD, Hahn LW, Roodi N, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 2001;**69**:138–47.
4. Calle ML, Urrea V, Vellalta G, et al. Improving strategies for detecting genetic patterns of disease susceptibility in association studies. *Stat Med* 2008;**27**:6532–46.
5. Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.
6. Shannon CE. A mathematical theory of communication. *Bell Labs Tech J* 1948;**27**:379–423. 623–656,
7. McGill WJ. Multivariate information transmission. *Psychometrika* 1954;**19**(2):97–116.
8. Jakulin A, Bratko I. Analyzing attribute dependencies. *Lect Notes Comput Sci* 2003;**2838**:229–40.
9. Jakulin A, Bratko I, Smrke D, et al. Attribute interactions in medical data analysis. *Lect Notes Comput Sci* 2003;**2780**:229–38.
10. Moore JH, Gilbert JC, Tsai C-T, et al. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol* 2006;**241**(2):252–61.
11. Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. *Annu Rev Public Health* 2010;**31**:21–36.
12. Jakulin A. Machine learning based on attribute interactions. PhD Dissertation, 2005.
13. Michalowicz JV, Nichols JM, Bucholtz F. *Handbook of Differential Entropy*. Boca Raton: Chapman & Hall Book, 2014.
14. Rényi A. On measures of information and entropy. *PBerkSymp* 1961;**1**:547–61.
15. de Andrade M, Wang X. Entropy based genetic association tests and gene-gene interaction tests. *Stat Appl Genet Mol Biol* 2011;**10**(1):38.
16. Moher D, Liberati A, Tetzlaff J., Altman DG. The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(6): e1000097.
17. Fan R, Zhong M, Wang S, et al. Entropy-based information gain approaches to detect and to characterize gene-gene and gene-environment interactions/correlations of complex diseases. *Genet Epidemiol* 2011;**35**(7):706–21.
18. Chen L, Yu G, Langefeld CD, et al. Comparative analysis of methods for detecting interacting loci. *BMC Genomics* 2011;**12**:344.
19. Su L, Liu G, Wang H, et al. Research on single nucleotide polymorphisms interaction detection from network perspective. *PLoS One* 2015;**10**(3):e0119146.
20. Kwon M-S, Park M, Park T. IGENE: efficient entropy based algorithm for genome-wide gene-gene interaction analysis. *BMC Med Genomics* 2014;**7**(Suppl 1):S6.
21. Zuo X, Rao S, Fan A, et al. To control false positives in gene-gene interaction analysis: two novel conditional entropy-based approaches. *Plos One* 2013;**8**(12):e81984.
22. Dong C, Chu X, Wang Y, et al. Exploration of gene-gene interaction effects using entropy-based methods. *Eur J Hum Genet* 2008;**16**:229–35.
23. Yee J, Kwon M-S, Park T, et al. A modified entropy-based approach for identifying gene-gene interactions in case-control study. *PloS One* 2013;**8**(7):e69321.
24. Chattopadhyay AS, Hsiao C-L, Chang CC, et al. Summarizing techniques that combine three non-parametric scores to detect disease-associated 2-way SNP-SNP interactions. *Gene* 2014;**533**(1):304–12.
25. Chanda P, Sucheston L, Zhang A, et al. AMBIENCE: a novel approach and efficient algorithm for identifying informative genetic and environmental associations with complex phenotypes. *Genetics* 2008;**180**(2):1191–210.
26. Chanda P, Sucheston L, Zhang A, et al. The interaction index, a novel information-theoretic metric for prioritizing interacting genetic variations and environmental factors. *Eur J Hum Genet* 2009;**17**:1274–86.
27. Chanda P, Sucheston L, Liu S, et al. Information-theoretic gene-gene and gene-environment interaction analysis of quantitative traits. *BMC Genomics* 2009;**10**:509.
28. Sucheston L, Chanda P, Zhang A, et al. Comparison of information-theoretic to statistical methods for gene-gene interactions in the presence of genetic heterogeneity. *BMC Genomics* 2010;**11**(1):487.

29. Chanda P, Zhang A, Ramanathan M. Modeling of environmental and genetic interactions with AMBROSIA, an information-theoretic model synthesis method. *Heredity* 2011;107:320–7.
30. Tritchler DL, Sucheston L, Chanda P, et al. Information metrics in genetic epidemiology. *Stat App Genet and Mol Biol* 2011;10(1):20.
31. Knights J, Yang J, Chanda P, et al. SYMPHONY, an information-theoretic method for gene-gene and gene-environment interaction analysis of disease syndromes. *Heredity* 2013;110:548–59.
32. Hu T, Chen Y, Kiralis JW, et al. An information-gain approach to detecting three-way epistatic interactions in genetic association studies. *J Am Med Inform Assoc* 2013;20: 630–6.
33. Anastassiou D. Computational analysis of the synergy among multiple interacting genes. *Mol Syst Biol* 2007;3:83.
34. Curk T, Rot G, Zupan B. SNPsyn: detection and exploration of SNP-SNP interactions. *Nucleic Acids Res* 2011;39:444–9.
35. Moore JH, Hu T. Epistasis analysis using information theory. In: *Epistasis*, vol. 1253 of the *series Methods in Molecular Biology* Springer, New York, 2015, 257–268.
36. Miller DJ, Zhang Y, Yu G, et al. An algorithm for learning maximum entropy probability models of disease risk that efficiently searches and sparingly encodes multilocus genomic interactions. *Bioinformatics* 2009;25(19):2478–85.
37. Kang G, Yue W, Zhang J, et al. An entropy-based approach for testing genetic epistasis underlying complex diseases. *J Theor Biol* 2008;250:362–74.
38. Ignac TM, Skupin A, Sakhanenko NA, et al. Discovering pairwise genetic interactions: an information theory-based approach. *PLoS One* 2014;9(3):e92310.
39. Yee J, Kwon M-S, Jin S, et al. Detecting genetic interactions for quantitative traits using m -spacing entropy measure. *BioMed Res Int* 2015, in press. ID 523641.
40. Brunel H, Gallardo-Chacón J-J, Buil A, et al. MISS: a non-linear methodology based on mutual information for genetic association studies in both population and sib-pairs analysis. *Bioinformatics* 2010;26(15):1811–8.
41. Goebel B, Dawy Z, Hagenauer J, et al. An approximation to the distribution of finite sample size mutual information estimates. *IEEE Int Conf Commun* 2005;2:1102–6.
42. Li J, Huang D, Guo M, et al. A gene-based information gain method for detecting gene-gene interactions in case-control studies. *Eur J Hum Genet* 2015;23:1566–72.
43. Breiman L, Friedman J, Ohlsen RA, et al. *Classification and Regression Trees*. Belmont, California: Wadsworth, 1984.
44. Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* 2003;19(3):376–82.
45. Chanda P, Zhang A, Brazeau D, et al. Information-theoretic metrics for visualizing gene-environment interactions. *Am J Hum Genet* 2007;81(5):939–63.
46. Lee W, Sjölander A, Pawitan Y. A critical look at entropy-based gene-gene interaction measures. *Genet Epidemiol* 2016;40(5):416–24.
47. Shang J, Zhang J, Sun Y, et al. Epiminer: a three-stage co-information based method for detecting and visualizing epistatic interactions. *Digit Signal Process* 2014;24:1–13.
48. Knights J, Ramanathan M. An information theory analysis of gene-environmental interactions in count/rate data. *Hum Hered* 2012;73:123–38.
49. Hu T, Chen Y, Kiralis JW, Moore JH. ViSEN: methodology and software for visualization of statistical epistasis networks. *Genet Epidemiol* 2013;37(3):283–5.
50. Hu T, Sinnott-Armstrong NA, Kiralis JW, et al. Characterizing genetic interactions in human disease association studies using statistical epistasis networks. *BMC Bioinform* 2011;12(364):1–13.
51. Jaynes ET. In RD Rosenkrantz (ed). *Papers on Probability, Statistics and Statistical Physics*. Kluwer Academic Publishers, Dordrecht, 1989.
52. Albert PS, Ratnasinghe D, Tangrea J, et al. Limitations of the case-only design for identifying gene-environment interactions. *Am J Epidemiol* 2001;154(8):687–93.
53. Learned-Miller EG, Fisher JW, III. ICA using spacings estimates of entropy. *J Mach Learn* 2004;4:1271–95.
54. Hong CS, Kim BJ. Mutual information and redundancy for categorical data. *Stat Papers* 2011;52:17–31.

Appendix

Estimation of IG_{Fan} proposed by Fan et al. [17]

In a case-control design with M controls and N cases, X_{ij} and Y_{ij} denote the count of controls and of cases whose genotypes are ($G_A = i, G_B = j$), respectively, where $i, j = 0, 1, 2$.

Define $\hat{p}_{ij} = \frac{X_{ij}}{M}$ and $\hat{q}_{ij} = \frac{Y_{ij}}{N}$. Then

$$f_{ij} = p_{ij} \log \left(\frac{p_{ij}}{p_i p_j} \right), \quad \hat{f}_{ij} = \hat{p}_{ij} \log \left(\frac{\hat{p}_{ij}}{\hat{p}_i \hat{p}_j} \right)$$

$$g_{ij} = q_{ij} \log \left(\frac{q_{ij}}{q_i q_j} \right) \quad \text{and} \quad \hat{g}_{ij} = \hat{q}_{ij} \log \left(\frac{\hat{q}_{ij}}{\hat{q}_i \hat{q}_j} \right)$$

with obvious meaning of q_i, q_j , and with M and N large enough to enable large sample theory and application of central limit theorem. With this,

$$IG_{\text{Fan}} := I_{\text{cases}}(G_i, G_j) - I(G_i, G_j) = g - f = \sum_{i=0}^2 \sum_{j=0}^2 (g_{ij} - f_{ij}) \quad \text{and}$$

$$\hat{g} - \hat{f} = \sum_{i=0}^2 \sum_{j=0}^2 (\hat{g}_{ij} - \hat{f}_{ij}).$$

Moreover, define

$$\hat{\Lambda} := \left(\frac{1}{M} \widehat{\text{var}}_{\text{controls}} \right) + \left(\frac{1}{N} \widehat{\text{var}}_{\text{cases}} \right)$$

where $\widehat{\text{var}}_{\text{cases}}$ and $\widehat{\text{var}}_{\text{controls}}$ are estimators of the asymptotic variances of the random variables

$$\sqrt{M}(\hat{f} - f)$$

and

$$\sqrt{N}(\hat{g} - g),$$

respectively. With this, the test statistic for estimating IG_{Fan} proposed by Fan et al. [17] is defined as:

$$T_{IG} := (\hat{g} - \hat{f})^2 / \hat{\Lambda}$$

T_{IG} is an overall test statistic to test the association between the markers A and B and the disease. Therefore, the null hypothesis is that the two markers are independent of the disease. It is asymptotically centrally or non-centrally chi-square distributed under the null or the alternative hypothesis, respectively, with one degree of freedom. The non-centrality parameter is $\lambda_{IG} = (g - f)^2 / \Lambda$.