PLOS ONE

RESEARCH ARTICLE

Epistasis Test in Meta-Analysis: A Multi-Parameter Markov Chain Monte Carlo Model for Consistency of Evidence

Chin Lin[1], **Chi-Ming Chu**[1,2]*, **Sui-Lung Su**[1,2]*

1 Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan, ROC, **2** School of Public Health, National Defense Medical Center, Taipei, Taiwan, ROC

* chuchiming@web.de (CMC); a131419@gmail.com (SLS)

Abstract

Conventional genome-wide association studies (GWAS) have been proven to be a successful strategy for identifying genetic variants associated with complex human traits. However, there is still a large heritability gap between GWAS and transitional family studies. The "missing heritability" has been suggested to be due to lack of studies focused on epistasis, also called gene–gene interactions, because individual trials have often had insufficient sample size. Meta-analysis is a common method for increasing statistical power. However, sufficient detailed information is difficult to obtain. A previous study employed a meta-regression-based method to detect epistasis, but it faced the challenge of inconsistent estimates. Here, we describe a Markov chain Monte Carlo-based method, called "Epistasis Test in Meta-Analysis" (ETMA), which uses genotype summary data to obtain consistent estimates of epistasis effects in meta-analysis. We defined a series of conditions to generate simulation data and tested the power and type I error rates in ETMA, individual data analysis and conventional meta-regression-based method. ETMA not only successfully facilitated consistency of evidence but also yielded acceptable type I error and higher power than conventional meta-regression. We applied ETMA to three real meta-analysis data sets. We found significant gene–gene interactions in the renin–angiotensin system and the polycyclic aromatic hydrocarbon metabolism pathway, with strong supporting evidence. In addition, glutathione *S*-transferase (GST) mu 1 and theta 1 were confirmed to exert independent effects on cancer. We concluded that the application of ETMA to real meta-analysis data was successful. Finally, we developed an R package, etma, for the detection of epistasis in meta-analysis [etma is available via the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/web/packages/etma/index.html].

Citation: Lin C, Chu C-M, Su S-L (2016) Epistasis Test in Meta-Analysis: A Multi-Parameter Markov Chain Monte Carlo Model for Consistency of Evidence. PLoS ONE 11(4): e0152891. doi:10.1371/journal.pone.0152891

Editor: Sriharsa Pradhan, Inc, UNITED STATES

Received: December 21, 2015

Accepted: March 21, 2016

Published: April 5, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Many complex human traits are considered to be associated with genetic factors, and previous genetic studies have identified a large number of causal variants [1]. However, the sum of the estimated genetic effects has often been much less than the heritability of the trait, a

phenomenon called 'missing heritability' [2]. This 'missing heritability' is often attributed to the technical limitations of epistasis estimation [3–5]. Generally, the most important limitation is sample size. A single study is often ineffective for detecting epistasis [3,6].

Meta-analysis has become a popular method for discovering genetic risk variants, because it can increase detection power [7,8]. However, few studies have sought to detect epistasis [9], because sufficient detailed information is difficult to obtain [10]. The frequencies of genotype combinations in case and control groups are needed for analysis of epistasis by current technology, but most published articles report only genotype frequencies. Thus, reported meta-analysis studies aiming at epistasis detection have been able to use only 20% of the reported studies [11–13]. The largest challenge of epistasis assessment in meta-analysis is the incompleteness of information.

Meta-regression is a common approach to assessing interaction effects in meta-analysis of randomised controlled trials [14,15], and a previous study popularised this method in meta-analysis of genetic association studies [16]. However, the inherent limitations of meta-regression have caused some problems in application of epistasis detection. The most important problem is attenuation bias. The average summary values in each included study are calculated from a small sample size and may thus include large random error [17–19]. Moreover, previous studies considered that two assumptions, rare disease and independence between SNPs, are necessary conditions for a linear relationship [16]. The rare-disease assumption is sometimes difficult to justify, and a previous study found slight error when this assumption was violated [16]. These random errors will lead to inconsistent estimates of interaction effects (see Fig 1), but this phenomenon does not occur in individual data analysis. Inconsistent evidence leads to difficulties in interpretation.

In summary, a single trial often has insufficient sample size, but meta-analysis lacks sufficient detailed individual information. The current method using averaged summary data for detecting interaction effects faces the challenge of inconsistent estimates. We propose a Markov chain Monte Carlo (MCMC)-based method, called 'Epistasis Test in Meta-Analysis (ETMA)', using genotype summary data for obtaining consistent estimates of epistasis in meta-analyses.

Materials and Methods

Derivations and description of ETMA

We assume that SNP1 (x_1) and SNP2 (x_2) are binary variables encoded as 0 and 1 (wild type and mutation, respectively), and that the dependent variable (y) is an outcome event encoded as 0 and 1 (health and disease, respectively). Under the above assumptions, we defined p_1, p_2, p_3, p_4, p_5 and p_6 as follows:

1. Disease risk in subjects with wild-type alleles of SNP1 and SNP2 (p_1):

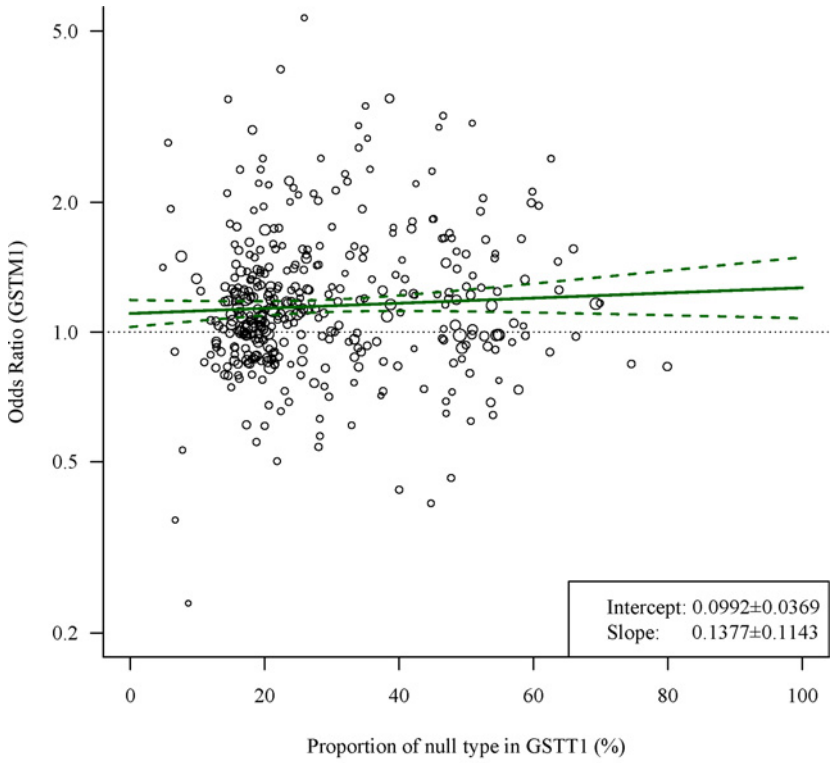
$$p_1 = p(y = 1|x_1 = 0 \cap x_2 = 0)$$

2. Disease risk in subjects with wild-type alleles of SNP1 and mutation of SNP2 (p_2):

$$p_2 = p(y = 1|x_1 = 0 \cap x_2 = 1)$$

3. Disease risk in subjects with mutations of SNP1 and wild type of SNP2 (p_3):

$$p_3 = p(y = 1|x_1 = 1 \cap x_2 = 0)$$


5.0
2.0
1.0
0.5
0.2
Odds Ratio (GSTM1)
Intercept: 0.0992±0.0369
Slope: 0.1377±0.1143
0
20
40
60
80
100
Proportion of null type in GSTT1 (%)

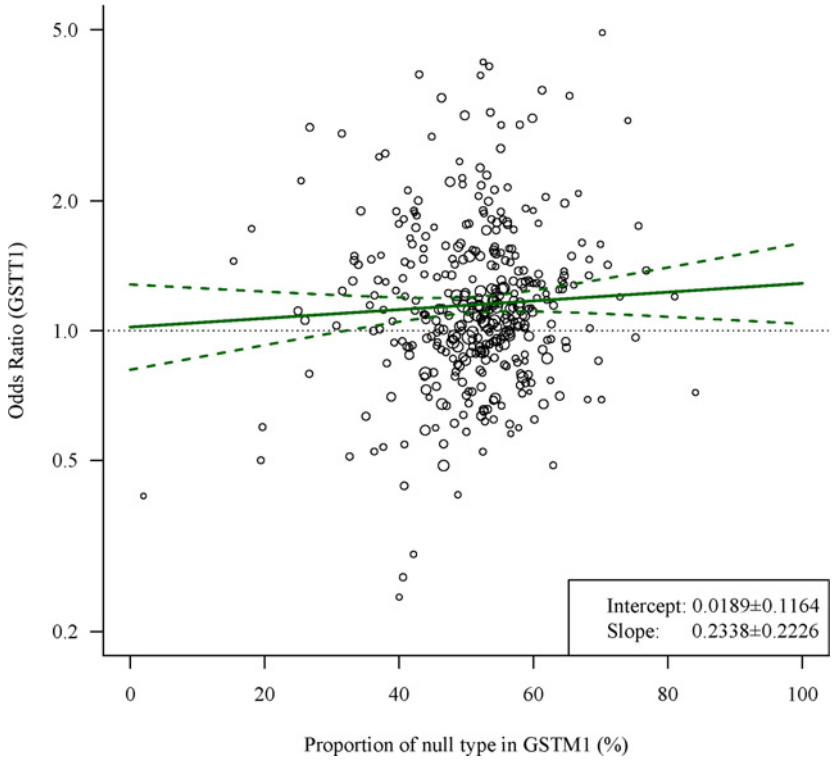

5.0
2.0
1.0
0.5
0.2
Odds Ratio (GSTT1)
Intercept: 0.0189±0.1164
Slope: 0.2338±0.2226
0
20
40
60
80
100
Proportion of null type in GSTM1 (%)

Fig 1. Inconsistent estimates of interaction effects in the same data. This figure describes a meta-regression analysis based on the data from Fang et al. [27] (detailed data are shown in S1 Table). The upper plot describes an investigation of the association between proportions of null/null GSTT1 in cases and the odds ratios of GSTM1 in cancer, and the lower plot describes an investigation of the association between proportions of null/null GSTM1 in cases and the odds ratios of GSTT1 in cancer. The solid lines denote unbiased estimators of odds ratios, and the dashed lines show 95% confidence intervals of odds ratios. According to a previous article, the slopes in meta regression approximate interaction effects [16]. However, the estimates of interaction effect were inconsistent when we exchanged the independent and moderator variables (0.1377 and 0.2338, respectively). This phenomenon does not occur in individual data analysis and leads to problems in interpretation.

doi:10.1371/journal.pone.0152891.g001

4. Disease risk in subjects with mutations of SNP1 and SNP2 (p_4):

$$p_4 = p(y = 1 | x_1 = 1 \cap x_2 = 1)$$

5. Mutation frequency of SNP1 (p_5):

$$p_5 = p(x_1 = 1)$$

6. Mutation frequency of SNP2 (p_6):

$$p_6 = p(x_2 = 1).$$

If x_1 and x_2 are independent, the above six parameters determine the distribution of x_1, x_2 and y in any population. However, we consider p_1, p_5 and p_6 as population-specific parameters and define three constant parameters as follows:

1. Main effect of SNP1 on y ($OR_{y,SNP1}$):

$$\mathrm{OR}_{y,\mathrm{SNP1}} = \frac{p_3(1-p_1)}{p_1(1-p_3)}$$

2. Main effect of SNP2 on y ($OR_{y,SNP2}$):

$$\mathrm{OR}_{y,\mathrm{SNP2}} = \frac{p_2(1-p_1)}{p_1(1-p_2)}$$

3. Gene–gene interaction effect between SNP1 and SNP2 on y ($OR_{interaction}$):

$$\mathrm{OR}_{\mathrm{interaction}} = \frac{p_1 p_4 (1-p_2)(1-p_3)}{p_2 p_3 (1-p_1)(1-p_4)}$$

Thus, p_2, p_3 and p_4 can be calculated by the following equations:

$$p_2 = \frac{p_1 \mathrm{OR}_{y,\mathrm{SNP2}}}{p_1 \mathrm{OR}_{y,\mathrm{SNP2}} + (1 - p_1)} \quad (2.1-1)$$

$$p_3 = \frac{p_1 \mathrm{OR}_{y,\mathrm{SNP1}}}{p_1 \mathrm{OR}_{y,\mathrm{SNP1}} + (1 - p_1)} \quad (2.1-2)$$

$$p_4 = \frac{p_1 \mathrm{OR}_{y,\mathrm{SNP1}} \mathrm{OR}_{y,\mathrm{SNP2}} \mathrm{OR}_{\mathrm{interaction}}}{p_1 \mathrm{OR}_{y,\mathrm{SNP1}} \mathrm{OR}_{y,\mathrm{SNP2}} \mathrm{OR}_{\mathrm{interaction}} + (1 - p_1)} \quad (2.1-3)$$

A case–control study including two loci often provides fourexposure rates: (1) of the x_1 mutation inthe case group ($e_{\mathrm{case,x1}}$), (2) of thex_1 mutation in the control group($e_{\mathrm{ctrl,x1}}$), (3) of the x_2 mutation in the case group ($e_{\mathrm{case,x2}}$) and (4) of the x_2 mutation in the control group ($e_{\mathrm{ctrl,x2}}$). These four exposure rates can be represented as combinations of p_1, p_2, p_3, p_4, p_5 and p_6. Their relationships are shown as follows (detailed calculations are shown in S1 Text):

$$e_{\mathrm{case},x1} = \frac{[p_3(1-p_6) + p_4 p_6]p_5}{p_1(1-p_6)(1-p_5) + p_2 p_6(1-p_5) + p_3(1-p_6)p_5 + p_4 p_6 p_5}$$

$$e_{\mathrm{ctrl},x1} = \frac{[(1-p_3)(1-p_6) + (1-p_4)p_6]p_5}{(1-p_1)(1-p_6)(1-p_5) + (1-p_2)p_6(1-p_5) + (1-p_3)(1-p_6)p_5 + (1-p_4)p_6 p_5}$$

$$e_{\mathrm{case},x2} = \frac{[p_2(1-p_5) + p_4 p_5]p_6}{p_1(1-p_6)(1-p_5) + p_2 p_6(1-p_5) + p_3(1-p_6)p_5 + p_4 p_6 p_5}$$

$$e_{\mathrm{ctrl},x2} = \frac{[(1-p_2)(1-p_5) + (1-p_4)p_5]p_6}{(1-p_1)(1-p_6)(1-p_5) + (1-p_2)p_6(1-p_5) + (1-p_3)(1-p_6)p_5 + (1-p_4)p_6 p_5}$$

According to the above relationship, we can calculate the likelihood of the sample using binomial distribution and execute the MCMC algorithm as follows:

MCMC algorithm. X is an $n \times 8$ matrix including the numbers of variants of SNP1 and SNP2 in case and control in each study (n is the number of studies). P is an $n \times 3$ matrix describing p_1, p_5 and p_6 in each included study, and OR is a 1×3 vector containing $\mathrm{OR}_{y,\mathrm{SNP1}}$, $\mathrm{OR}_{y,\mathrm{SNP2}}$ and $\mathrm{OR}_{\mathrm{interaction}}$. X is a known matrix, and P and OR are unknown matrices. P and OR can be expressed as follows:

$$P = \begin{bmatrix} p_{1,1} & p_{5,1} & p_{6,1} \\ p_{1,2} & p_{5,2} & p_{6,2} \\ p_{1,3} & p_{5,3} & p_{6,3} \\ \cdots & \cdots & \cdots \\ p_{1,n} & p_{5,n} & p_{6,n} \end{bmatrix}$$

$$\mathrm{OR} = \begin{bmatrix} \mathrm{OR}_{y,\mathrm{SNP1}} & \mathrm{OR}_{y,\mathrm{SNP2}} & \mathrm{OR}_{\mathrm{interaction}} \end{bmatrix}$$

We can use the approach outlined in the following iteration process to construct a Markov chain stationary distribution $\Pr(P, \mathrm{OR} | X)$ as follows:

Iteration process. Starting with initial values $OR^{(0)}$ for OR ($OR^{(0)}$ = [1 1 1]), we iterate the following steps for $m = 1, 2, \ldots$

Step 1: Sample $P^{(m)}$ from $\Pr(P^{(m)} | X, OR^{(m-1)})$

Step 2: Sample $OR^{(m)}$ from $\Pr(OR^{(m)} | X, P^{(m)})$

In simple terms, *Step 1* is to assume that $OR_{y,SNP1}$, $OR_{y,SNP2}$ and $OR_{interaction}$ are known parameters and to estimate p_1, p_5 and p_6 in each included study using the Metropolis–Hastings algorithm. This algorithm will find the p_1, p_5 and p_6 that maximise the likelihood of a given sample. Finally in this step, we can obtain the p_1, p_5 and p_6 of each included study. *Step 2* is to assume that p_1, p_5 and p_6 are known parameters and to estimate $OR_{y,SNP1}$, $OR_{y,SNP2}$ and $OR_{interaction}$. We assume that each cell of P or OR is described by a random walk in the logistic or logarithmic normal distribution, respectively. The above two steps are repeated until convergence of the log likelihood.

Implementation in 'etma' package by R language

An R package, etma, is developed for carrying out the epistasis detection in meta-analysis [etma is available via the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/web/packages/etma/index.html]. The main function of etma package is 'ETMA', and ETMA use an $n \times 8$ matrix including the numbers of variants of SNP1 and SNP2 in case and control in each study (n is the number of studies) to analyse gene-gene interaction. Thus, the inputs of ETMA function include: (1) the number of wild type of SNP1 in case group, (2) the number of mutation type of SNP1 in case group, (3) the number of wild type of SNP1 in control group, (4) the number of mutation type of SNP1 in control group, (5) the number of wild type of SNP2 in case group, (6) the number of mutation type of SNP2 in case group, (7) the number of wild type of SNP2 in control group, and (8) the number of mutation type of SNP1 in control group.

Because ETMA is based on MCMC and a 2-steps iteration process (details are shown in 2.1 Derivations and description of ETMA). The main options of ETMA function include: (1) the maximum number of iterations (default is 20), (2) the length of chain to obtain the study-level parameters in step 1 (default is 20,000), (3) the length of chain to obtain the global-level parameters in step 2 (default is 200,000), and (4) the start seed of this algorithm (default is a random seed). Moreover, user also can choose whether want to export MCMC plots in each iterations.

The main outputs include: (1) the beta values (logarithmic ORs) of each SNP and interaction term, (2) the variance covariance matrix of beta value, and (3) the p matrix in iterations process. According these outputs, we can calculate ORs, their confidence intervals, and p values. Fig 2 summarized the pipeline of ETMA function. Finally, a tutorial on epistasis detection using ETMA via 'etma' package is shown in S2 Text.

Simulations

In this subsection, we simulated a meta-analysis of genetic association studies. In summary, we wanted to generate a data including population with different baseline disease risk and minor allele frequency. Moreover, ETMA is a method for analysing the meta-analysis of candidate genetic association studies, so we just need to generate 2 unlinkage SNPs (because the limit of summary data) and disease status. Follow above concept, we generated 20 large populations in each simulation, with three population-specific parameters: (1) the disease risk in subjects with wild-type alleles of SNP1 and SNP2 ($p_{baseline}$), (2) the minor allele frequency of SNP1 (MAF_1) and (3) the minor allele frequency of SNP2 (MAF_2). We defined a series of $p_{baseline}$ in our simulations, summarised in Table 1. The MAF_1 and MAF_2 were generated by the Balding–Nichols

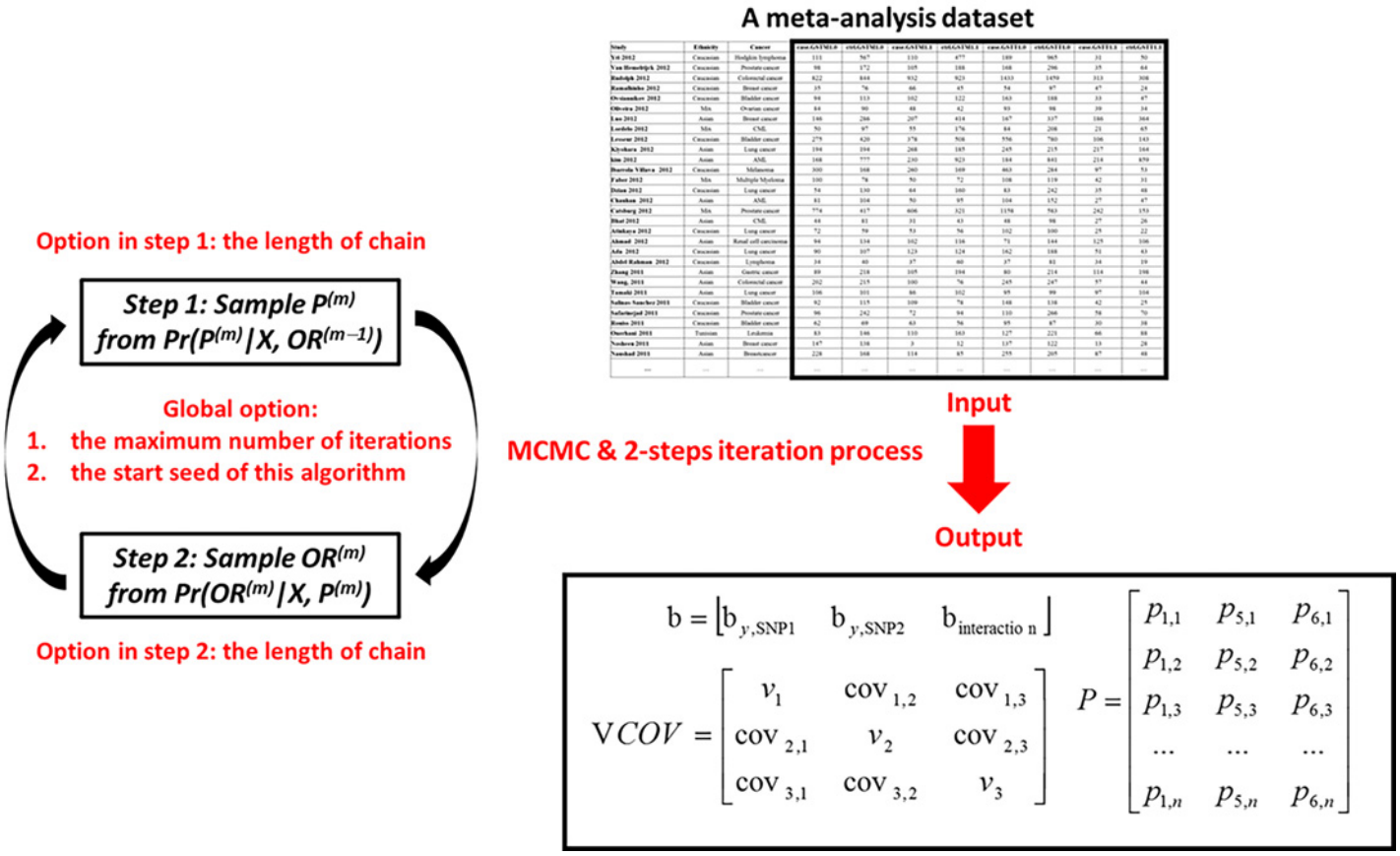


Fig 2. A typical analysis pipeline of ETMA function in 'etma' package. This figure summarized the pipeline of ETMA function. The main input is a meta-analysis dataset, which including the number of wild/mutation type of SNP1/SNP2 in case/control group. The main options include the length of chains in step 1/2, the maximum number of iterations, and the start seed. Main outputs include three matrixes. Matrix b includes the beta values (logarithmic ORs) of each SNP and interaction term, and VCOV is the variance covariance matrix of beta value. *P* is an n by 3 matrix describing three study-specific parameters (p1 = Disease risk in subjects with wild-type alleles of SNP1 and SNP2; p5 = Mutation frequency of SNP1; p6 = Mutation frequency of SNP2)

doi:10.1371/journal.pone.0152891.g002

model [20]. We set the mean mutation frequency ($\bar{\pi}$) at 50% and fixed F_{st} at 0.1 in all simulations, and SNP1/SNP2 are independence and follow Hardy-Weinberg equilibrium. The minor allele frequency (π_i) in each population was randomly generated from a beta distribution ($\alpha = \bar{\pi}(1-F_{st})/F_{st}; \beta = (1-\bar{\pi})(1-F_{st})/F_{st}$). We defined three parameters descripting the effects of SNP1, SNP2 and their integration as $OR_{y,SNP1}$, $OR_{y,SNP2}$ and $OR_{interaction}$, respectively,

Table 1. Summary of simulation conditions.

$p_{baseline}$	$OR_{y,SNP1}$	$OR_{y,SNP2}$	$OR_{interaction}$
~Uniform (0.001, 0.002)	1.0	1.0	1.0
~Uniform (0.01, 0.02)	1.2	1.2	1.2
~Uniform (0.1, 0.2)			1.5
			2.0

$p_{baseline}$: the disease risk in subjects with major homozygous genotype of SNP1 and SNP2 in each simulated population.

$OR_{y,SNP1}$: the main effect of SNP1.

$OR_{y,SNP2}$: the main effect of SNP2.

$OR_{interaction}$: gene–gene interaction effect between SNP1 and SNP2.

doi:10.1371/journal.pone.0152891.t001

and the disease prevalence of individuals with different genotype of SNP1/SNP2 were following logistic regression. The values of $OR_{y,SNP1}$, $OR_{y,SNP2}$ and $OR_{interaction}$ are summarised in Table 1. After we obtained $p_{baseline}$, MAF_1, MAF_2, $OR_{y,SNP1}$, $OR_{y,SNP2}$ and $OR_{interaction}$, the proportion of individual with different type of disease/SNP1/SNP2 could be calculated by Table 2. To use the information of Table 2, we randomly sampled a case–control study with a sample size randomly generated from a uniform (300, 1000) distribution. The proportion of cases was set to 50%.

In the subsequent analysis, we compared three methods: ETMA, individual data analysis and conventional meta-analysis. The detailed calculation method of ETMA is described in section

Table 2. The proportion of individual with different status of disease/SNP1/SNP2 could be calculated by $p_{baseline}$, MAF_1, MAF_2, $OR_{y,SNP1}$, $OR_{y,SNP2}$ and $OR_{interaction}$.

SNP1	SNP2	Disease	Proportion in total population
Major homozygous	Major homozygous	Control	$(1\text{-}MAF_1)^2(1\text{-}MAF_2)^2(1\text{-}q_1)$
Major homozygous	Major homozygous	Case	$(1\text{-}MAF_1)^2(1\text{-}MAF_2)^2q_1$
Major homozygous	Heterogeneous	Control	$2(1\text{-}MAF_1)^2(1\text{-}MAF_2)p_6(1\text{-}q_2)$
Major homozygous	Heterogeneous	Case	$2(1\text{-}MAF_1)^2(1\text{-}MAF_2)p_6q_2$
Major homozygous	Minor homozygous	Control	$(1\text{-}MAF_1)^2(1\text{-}MAF_2)^2(1\text{-}q_3)$
Major homozygous	Minor homozygous	Case	$(1\text{-}MAF_1)^2(1\text{-}MAF_2)^2q_3$
Heterogeneous	Major homozygous	Control	$2MAF_1(1\text{-}MAF_1)(1\text{-}MAF_2)^2(1\text{-}q_4)$
Heterogeneous	Major homozygous	Case	$2MAF_1(1\text{-}MAF_1)(1\text{-}MAF_2)^2q_4$
Heterogeneous	Heterogeneous	Control	$4MAF_1(1\text{-}MAF_1)(1\text{-}MAF_2)p_6(1\text{-}q_5)$
Heterogeneous	Heterogeneous	Case	$4MAF_1(1\text{-}MAF_1)(1\text{-}MAF_2)p_6q_5$
Heterogeneous	Minor homozygous	Control	$2MAF_1(1\text{-}MAF_1)(1\text{-}MAF_2)^2(1\text{-}q_6)$
Heterogeneous	Minor homozygous	Case	$2MAF_1(1\text{-}MAF_1)(1\text{-}MAF_2)^2q_6$
Minor homozygous	Major homozygous	Control	$MAF_1^2(1\text{-}MAF_2)^2(1\text{-}q_7)$
Minor homozygous	Major homozygous	Case	$MAF_1^2(1\text{-}MAF_2)^2q_7$
Minor homozygous	Heterogeneous	Control	$2MAF_1^2(1\text{-}MAF_2)p_6(1\text{-}q_8)$
Minor homozygous	Heterogeneous	Case	$2MAF_1^2(1\text{-}MAF_2)p_6q_8$
Minor homozygous	Minor homozygous	Control	$MAF_1^2(1\text{-}MAF_2)^2(1\text{-}q_9)$
Minor homozygous	Minor homozygous	Case	$MAF_1^2(1\text{-}MAF_2)^2q_9$

$p_{baseline}$: the disease risk in subjects with major homozygous genotype of SNP1 and SNP2 in each simulated population; MAF_1: the minor allele frequency of SNP1; MAF_2: the minor allele frequency of SNP2; $OR_{y,SNP1}$: the main effect of SNP1; $OR_{y,SNP2}$: the main effect of SNP2; $OR_{interaction}$: gene–gene interaction effect between SNP1 and SNP2.

q_1 to q_9: the disease prevalence of individuals with different genotype.

$$q_1 = p_{baseline} = (1 + \exp(-\ln(p_{baseline}/(1 - p_{baseline}))))^{-1}$$

$$q_2 = \left(1 + \exp\left(-\ln\left(p_{baseline}OR_{y,SNP_2}/(1 - p_{baseline})\right)\right)\right)^{-1}$$

$$q_3 = \left(1 + \exp\left(-\ln\left(p_{baseline}OR_{y,SNP_2}{}^2/(1 - p_{baseline})\right)\right)\right)^{-1}$$

$$q_4 = \left(1 + \exp\left(-\ln\left(p_{baseline}OR_{y,SNP_1}/(1 - p_{baseline})\right)\right)\right)^{-1}$$

$$q_5 = \left(1 + \exp\left(-\ln\left(p_{baseline}OR_{y,SNP_1}OR_{y,SNP_2}/(1 - p_{baseline})\right)\right)\right)^{-1}$$

$$q_6 = \left(1 + \exp\left(-\ln\left(p_{baseline}OR_{y,SNP_1}OR_{y,SNP_2}{}^2/(1 - p_{baseline})\right)\right)\right)^{-1}$$

$$q_7 = \left(1 + \exp\left(-\ln\left(p_{baseline}OR_{y,SNP_1}{}^2/(1 - p_{baseline})\right)\right)\right)^{-1}$$

$$q_8 = \left(1 + \exp\left(-\ln\left(p_{baseline}OR_{y,SNP_1}{}^2OR_{y,SNP_2}/(1 - p_{baseline})\right)\right)\right)^{-1}$$

$$q_9 = \left(1 + \exp\left(-\ln\left(p_{baseline}OR_{y,SNP_1}{}^2OR_{y,SNP_2}{}^2/(1 - p_{baseline})\right)\right)\right)^{-1}$$

doi:10.1371/journal.pone.0152891.t002

'Derivations and description of ETMA', and this program used the summary data from each study. Individual data analysis is considered the gold standard for investigating the moderator effect [16,18], and we used a hierarchical generalised linear model based on the lme4 R package [21] with pooled data to estimate the interaction effect. Conventional meta-analysis was calculated based on a previous study [16]. Owing to the inconsistent estimates of interaction effects (refer to Fig 1), we used only the analysis fitting SNP1 as the independent variable and SNP2 as the moderator. Data under each condition were generated from 1,000 simulations.

Application to real data

ETMA is a method for analysing the meta-analysis of candidate genetic association studies. Because the limit of multi-loci analysis technology, previous meta-analysis often focus on the association between a specific disease and a SNP but not on the epistasis. Thus, the existing meta-analysis including more than 1 SNP are rare. Moreover, only few papers completely provided their data, so such data is difficult to obtain. According to above reasons, we only can find 3 independent paper providing sufficient information for ETMA. It does not represent the practicability of ETMA is bad, but represent we need more meta-analysis investigating the epistasis.

Glutathione S-transferase (GST) family and cancer. The GST family detoxifies oxidative stress products, environmental toxins and carcinogens [22,23]. GST mu 1 (GSTM1) and GST theta 1 (GSTT1) are two critical GST family genes located in human chromosome regions 1p13.3 and 22q11.23, respectively. Generally, the variants in GSTM1 and GSTT1 are summarised as two types: (1) functional type and (2) null type [24–26]. Because of lack of detoxification mechanism, investigation of the associations between GSTM1/GSTT1 null type and cancer is popular. We used the data from a meta-analysis of approximately 500 studies investigating the association between GSTM1/GSTT1 and cancer [27] and selected the studies describing the genotypes of both GSTM1 and GSTT1. This filter left 360 studies (375 populations) in our real data analysis (the detailed data are shown in S1 Table).

Polycyclic aromatic hydrocarbons (PAHs) metabolism pathway and oral cancer. PAHs are strong carcinogens [28] found in coal tar, automobile exhaust fumes, charbroiled food and cigarette smoke. Cytochrome P450 1A1 (CYP1A1), located on chromosome 15, had been confirmed to be a component of the PAH metabolism pathway [29]. This pathway also involves the GST family. We used the data from a meta-analysis of approximately 50 studies investigating the association between CYP1A1/GSTM1 and oral cancer [30] and selected the studies describing the genotypes of both GSTM1 and CYP1A1 rs4646903. This filter left 13 studies in our real data analysis (the detailed data are shown in S2 Table).

Renin–angiotensin system (RAS) and chronic kidney disease. The RAS is a system-balancing electrolyte that regulates blood pressure, and a dysfunction of RAS increases the risk of kidney failure [31–33]. Angiotensinogen (AGT) is the initial protein in the RAS and is converted to angiotensin II, a terminal active product in the RAS [34]. This conversion is through renin and angiotensin-converting enzyme (ACE) [34]. We used the data from our earlier meta-analysis of approximately 100 studies investigating the association between ACE insertion/deletion (I/D) and chronic kidney disease [35] and selected the studies including AGT M235T information. We added four related articles published in 2014 [36–39]. There were then 34 studies in our real data analysis (the detailed data are shown in S3 Table).

Results

Simulation analysis

Table 3 shows the type I errors yielded by individual data analysis, ETMA and conventional meta-regression under each simulation condition. The type I errors of ETMA are between

Table 3. Type I error of individual data analysis, ETMA and conventional meta-regression.

Simulation conditions			Individual data analysis	ETMA	Conventional meta-regression
$p_{baseline}$	$OR_{y,SNP1}$	$OR_{y,SNP2}$			
~Uniform (0.001, 0.002)	1.0	1.0	0.047	0.037	0.050
~Uniform (0.001, 0.002)	1.2	1.0	0.039	0.039	0.054
~Uniform (0.001, 0.002)	1.2	1.2	0.039	**0.034**	0.052
~Uniform (0.01, 0.02)	1.0	1.0	0.047	0.037	0.050
~Uniform (0.01, 0.02)	1.2	1.0	0.059	**0.033**	0.048
~Uniform (0.01, 0.02)	1.2	1.2	0.047	**0.034**	0.047
~Uniform (0.1, 0.2)	1.0	1.0	0.047	0.037	0.050
~Uniform (0.1, 0.2)	1.2	1.0	0.055	0.052	0.059
~Uniform (0.1, 0.2)	1.2	1.2	0.043	**0.033**	0.047

$p_{baseline}$: the disease risk in subjects with major homozygous genotype of SNP1 and SNP2 in each simulated population.$OR_{y,SNP1}$: the main effect of SNP1.

$OR_{y,SNP2}$: the main effect of SNP2.

The bold value denotes a significant difference compared with 0.05 (the 95% confidence interval of type I error is between 0.036 and 0.064). Each data point was based on 1,000 simulations.

doi:10.1371/journal.pone.0152891.t003

0.033 and 0.052. In comparison with 0.05, ETMA was more conservative. The range of type I errors in individual data analysis and conventional meta-regression is 0.039–0.059 and 0.047–0.059, respectively. Thus, we judged all methods to have acceptable type I error. However, the meta-regression may have slight bias when the baseline disease risk is set to 0.1–0.2. This bias may be due to violation of the rare-disease assumption. A previous study showed a slight bias at a baseline disease risk equal to 0.1 [16].

Fig 3 shows the power of individual data analysis, ETMA and conventional meta-regression under each simulation condition. Overall, the performances of these three methods were not affected by the simulation conditions (p_1, $OR_{y,SNP1}$ and $OR_{y,SNP2}$). In the power analysis, individual data analysis showed higher power than ETMA, followed by conventional meta-regression. The power of conventional meta-regression was slightly smaller when $OR_{y,SNP1}$ and $OR_{y,SNP2}$ were not equal to 1.0. This result may be due to damage of nonlinear relationship [16]. However, the power curves of ETMA were similar under all simulation conditions.

ETMA gave the higher statistical power compared with conventional meta-regression, and it also solved the challenge of inconsistent estimates (see Fig 1). Although individual data analysis gave the highest statistical power in our results, and previous evidence shows that individual data analysis is the gold standard [16,18,40]. The summary statistics are widely available [8,41], and individual information is difficult to obtain [10,42]. Thus, the practicability of ETMA is better than individual data analysis. In our simulation, the power of ETMA was higher than that of conventional meta-regression, and we considered the reason of higher power in ETMA as below: The first step of calculation in conventional meta-regression is to calculate OR from exposure rate [16]. We considered this step to represent a loss of information compared with ETMA. Moreover, given that our study showed a non-linear relationship between OR and mutation frequency, the linear relationship-based meta-regression was expected to give lower power.

Besides lower statistical power, conventional meta-regression must also face the challenge of inconsistent estimates. Although we ignored the second direction analysis in simulation, researchers will still be confused in real meta-analysis because inconsistent results will lead to difficulties of interpretation. In short, ETMA not only integrates the inconsistent information but also is more sensitive.

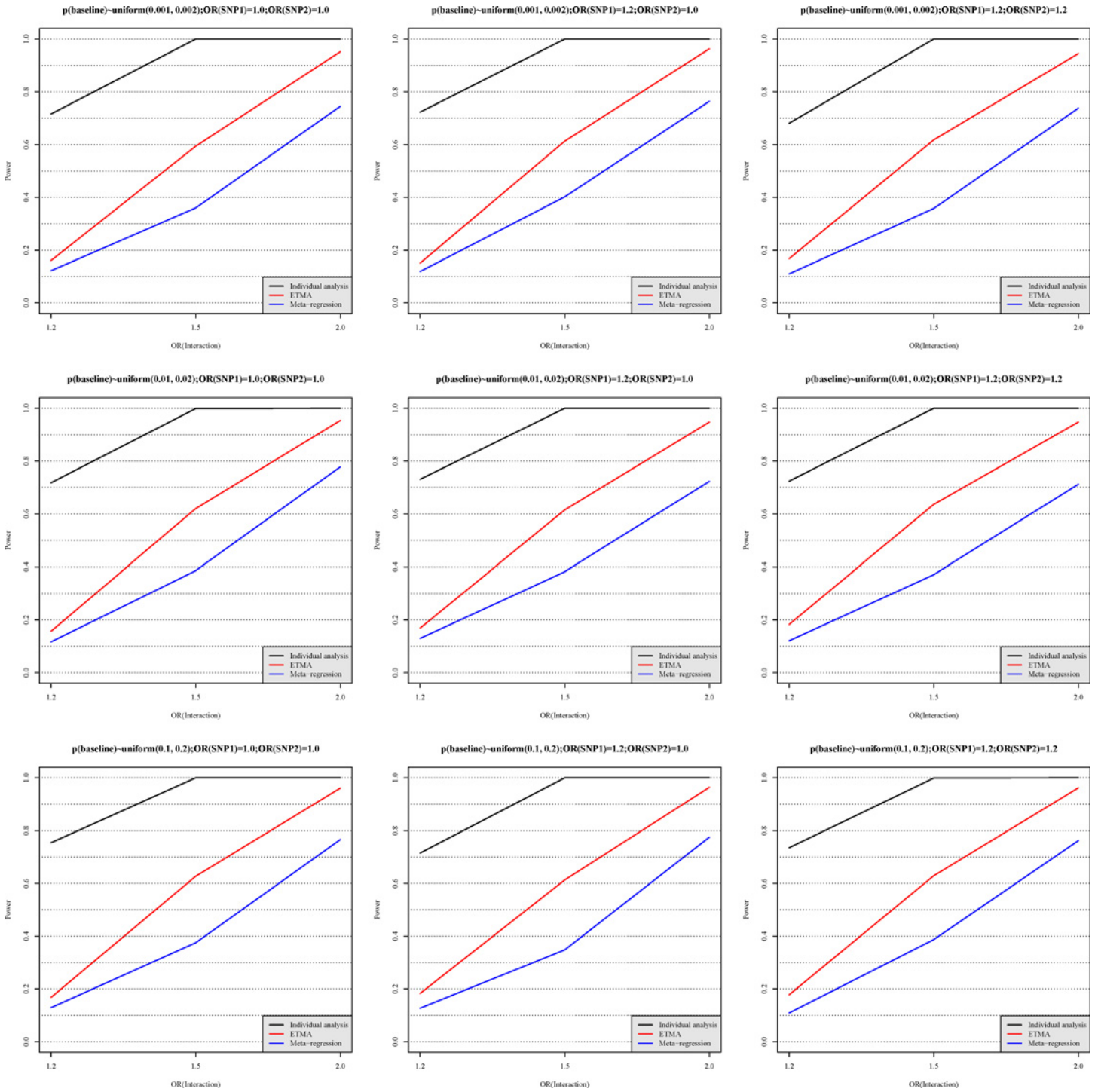


Fig 3. The statistical power of individual data analysis, ETMA and conventional meta-regression. The *x*-axis describes three levels of interaction effect ($OR_{interaction}$ = 1.2, 1.5 or 2.0), and the *y*-axis indicates the statistical power provided by individual data analysis (black), ETMA (red) and conventional meta-regression (blue), respectively. The details of these methods are described in the Method. The different subplots present comparisons using different simulation parameters, and the titles of these subplots show their detailed settings. Each data point was based on 1,000 simulations.

doi:10.1371/journal.pone.0152891.g003

Real data analysis

We applied ETMA to summary statistics from previous meta-analysis [27,30,35] (detailed information is presented in Methods). Table 4 shows the summary results of real data analysis (the detailed calculation process using the etma package is shown in S2 Text). For all studies, the logarithmic OR of SNP1, SNP2 and their interaction in the MCMC plot shows that normal distribution after burn-in time was deleted (the MCMC plots of the data sets are shown in S1–S3 Figs, respectively). Moreover, the marginal density plots show good convergence at each iteration. These results show that ETMA remains robust in analysis of real data.

The result of analysis of the GST family and cancer shows significant ORs of GSTM1 and GSTM2 on cancer [1.110 (95% CI: 1.080–1.141) and 1.125 (95% CI: 1.073–1.180), respectively]. However, the interaction term of GSTM1 and GSTT1 is not significant (p = 0.2525). Although these genes belong to the same family, we also considered this to be a reasonable result. The GST family has many overlapping functions, and GSTM2 can perform more functions in subjects with a GSTM1 null genotype [43]. Moreover, the GSTM1/GSTT1 null genotype has been reported to confer a slight increase in risk [OR: 1.33 (95% CI: 1.10–1.61)] of lung cancer in a small-scale meta-analysis [11]. The result of our analysis was similar [OR: 1.176 (95% CI: 1.142–1.211); data are shown in S2 Text].

The analysis of the metabolism pathway of PAHs and oral cancer shows a significant gene–gene interaction effect (OR: 2.220 (95% CI: 1.166–4.225), $p = 0.0201$), and the main effect of each SNP is not significant (p = 0.2008 and 0.8915 for CYP1A1 and GSTM1, respectively). CYP1A1 and GSTM1 are two important members in the PAH metabolism pathway [29], and PAHs are strong carcinogens [28]. Moreover, a pooled analysis of lung cancer also reported a strong gene–gene interaction between them [44].

The analysis of the RAS and chronic kidney disease also shows a significant gene–gene interaction (OR: 1.305 (95% CI: 1.048–1.624), p = 0.0188). This result indicates an interaction effect between AGT M235T (rs699) and ACE I/D (rs4340) on chronic kidney disease, but that neither alone increases the risk of chronic kidney disease, because its main effect is not significant (p = 0.2073 and 0.9277 in ACE I/D and AGT M235T, respectively). The detailed mechanisms and possible reasons are described in the Discussion. We judged these results to be consistent with expectations. The AGT M235T polymorphism has been confirmed to affect blood AGT concentration [45], and excess AGT leads to a high concentration of angiotensin I in blood [46]. Moreover, the DD genotype of ACE I/D showed higher gene expression and serum ACE levels than the ID genotype, followed by the II genotype [47,48]. Thus, subjects carrying the T allele in AGT M235T and the D allele in ACE I/D may have especially high

Table 4. The result of real data analysis using ETMA.

Real data set		OR (95% CI)	p value
GSTs family and cancer			
	GSTM1 (null type vs. functional type)	1.110 (1.080–1.141)	<0.0001
	GSTT1 (null type vs. functional type)	1.125 (1.073–1.180)	<0.0001
	GSTM1×GSTT1 (interaction term)	0.942 (0.862–1.029)	0.1814
Metabolism pathway of PAH and oral cancer			
	CYP1A1 (AC/CC vs. AA)	0.819 (0.592–1.133)	0.2008
	GSTM1 (null type vs. functional type)	0.981 (0.717–1.340)	0.8915
	CYP1A1×GSTM1 (interaction term)	2.220 (1.166–4.225)	0.0201
RAS and chronic kidney disease			
	ACE (D allele vs. I allele)	0.921 (0.809–1.049)	0.2073
	AGT (T allele vs. M allele)	0.995 (0.884–1.120)	0.9277
	ACE ×AGT (interaction term)	1.305 (1.048–1.624)	0.0188

doi:10.1371/journal.pone.0152891.t004

angiotensin II, based on the RAS pathway [34], and increased risk of chronic kidney disease [49]. In short, we propose that results of our real data analysis are consistent with current evidence.

Discussion

Because the technological limitation of multi-loci analysis, previous meta-analysis often focus on the association between a specific disease and a SNP but not on the epistasis. Thus, the existing meta-analysis including more than 1 SNP are rare. However, epistasis is important in genetic association study. Previous studies considered that 'missing heritability' is often attributed to the technical limitations of epistasis estimation [3–5]. The summary statistics are widely available [8,41], and individual information is difficult to obtain [10,42]. ETMA have solved this technological limitation, and researchers can analyse gene-gene interaction using summary data. In this paper, we re-analysed few previous meta-analysis data [27,30,35], and found significant gene-gene interaction in PAHs metabolism pathway/RAS on oral cancer/chronic kidney disease. These findings may explain a part of 'missing heritability' in oral cancer/chronic kidney disease, and improve our biological knowledge. We believe the multi-locus meta-analysis will be more popular in the future because this technological breakthroughs.

ETMA may lack the ability to detect gene–environment interactions because of issues related to degrees of freedom. ETMA is based on four exposure rates (of the x_1 mutation in the case group, of the x_1 mutation in the control group, of the x_2 mutation in the case group and of the x_2 mutation in the control group) in each included study. Some studies matched the environmental factors to reduce the confounding bias, sacrificing 1 degree of freedom. Thus, fitting of gene–environment interactions using ETMA will constitute overfitting. However, although this defect causes a problem in ETMA, it solves the problem of inconsistent estimates in meta-regression analysis [16]. Owing to matching, the odds ratios of environment factors are unavailable, so that gene–environment interaction analysis using meta-regression will yield a result for only one direction. Thus, we suggest that researchers use conventional meta-regression to detect gene–environment interaction [16] and ETMA to detect gene–gene interaction.

In conclusion, ETMA has acceptable type I error rates under all simulation condition. Moreover, it not only successfully facilitates consistency of evidence but also increases power. Although our results also show that individual data analysis is the most powerful analysis, sufficient detailed information is difficult to obtain, so that the practical value of ETMA for meta-analysis is higher. Because ETMA assumes independence between two loci, analysis of loci on different chromosomes is a better option (at least on different genes). For gene–environment interactions, we suggest that the researcher use conventional meta-regression unless it is verified that the distribution of environmental factors has not been artificially changed (such as by matching). Finally, a package (etma, readers can download it form https://cran.r-project.org/web/packages/etma/index.html) was developed in the R language and may be extensively applied to detect epistasis in meta-analyses.

Supporting Information

S1 Fig. ETMA of GSTM1/GSTT1 and cancer. Page 1 shows the MCMC plot for the first iteration, page 2 the second and so on. The final page shows the final iteration result, and the analysis results are based on this chain value.
(PDF)

S2 Fig. ETMA of CYP1A1/GSTM1 and oral cancer. Page 1 shows the MCMC plot for the first iteration, page 2 the second and so on. The final page shows the final iteration result, and

the analysis results are based on this chain value.
(PDF)

S3 Fig. The ETMA of ACE/AGT and chronic kidney disease. Page 1 shows the MCMC plot for the first iteration, page 2 the second and so on. The final page shows the final iteration result, and the analysis results are based on this chain value.
(PDF)

S1 Table. Meta-analysis data of GSTM1/GSTT1 on cancer from Fang et al. [27]. Variable definitions:
Study: the first author and published year in each included study
Ethnicity: the ethnicity of included population.
Country: the country of study.
Cancer: the cancer type.
case.GSTM1.0: the number of functional GSTM1 carriers in cases (including heterozygous).
ctrl.GSTM1.0: the number of functional GSTM1 carriers in controls (including heterozygous).
case.GSTM1.1: the number of null/null GSTM1 carriers in cases (risk type).
ctrl.GSTM1.1: the number of null/null GSTM1 carriers in controls (risk type).
case.GSTT1.0: the number of functional GSTT1 carriers in cases (including heterozygous).
ctrl.GSTT1.0: the number of functional GSTT1 carriers in controls (including heterozygous).
case.GSTT1.1: the number of null/null GSTT1 carriers in cases (risk type).
ctrl.GSTT1.1: the number of null/null GSTT1 carriers in controls (risk type).
(XLSX)

S2 Table. Meta-analysis data of CYP1A1/GSTM1 on oral cancer from Liu et al. [30]. Variable definitions:
Author: the first author in each included article.
Year: the year of publication.
Country: the country of study location.
case.CYP1A1.0: the number of subjects with AA genotype in rs4646903 in cases.
case.CYP1A1.1: the number of subjects with AC/CC genotype in rs4646903 in cases (risk type).
ctrl.CYP1A1.0: the number of subjects with AA genotype in rs4646903 in controls.
ctrl.CYP1A1.1: the number of subjects with AC/CC genotype in rs4646903 in controls (risk type).
case.GSTM1.0: the number of functional GSTM1 carriers in cases (including heterozygous).
case.GSTM1.1: the number of null/null GSTM1 carriers in cases (risk type).
ctrl.GSTM1.0: the number of functional GSTM1 carriers in controls (including heterozygous).
ctrl.GSTM1.1: the number of null/null GSTM1 carriers in controls (risk type).
(XLSX)

S3 Table. Meta-analysis data of ACE/AGT and chronic kidney disease from Lin et al. [35]. Variable definitions:
Author: the first author in each included article.
Year: the year of publication.
Race: the race of the study population.
Type: the subtype of chronic kidney disease in each study.
case.ACE.0: the number I allele in rs4340 in cases.
case.ACE.1: the number D allele in rs4340 in cases (risk type).
ctrl.ACE.0: the number I allele in rs4340 in controls.
ctrl.ACE.1: the number D allele in rs4340 in controls (risk type).
case.AGT.0: the number M allele in rs699 in cases.

case.AGT.1: the number T allele in rs699 in cases (risk allele).
ctrl.AGT.0: the number M allele in rs699 in controls.
ctrl.AGT.1: the number T allele in rs699 in controls (risk allele).
(XLSX)

S1 Text. Detailed derivations of the relationships between $e_{case,x1}$, $e_{ctrl,x1}$, $e_{case,x2}$ and $e_{ctrl,x2}$ and p_1, p_2, p_3, p_4, p_5 and p_6.
(DOCX)

S2 Text. A tutorial on epistasis detection using ETMA.
(DOCX)

Author Contributions

Conceived and designed the experiments: CL CMC SLS. Performed the experiments: CL CMC SLS. Analyzed the data: CL. Contributed reagents/materials/analysis tools: CL CMC SLS. Wrote the paper: CL.

References

1. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research. 2014; 42(Database issue):D1001–6. Epub 2013/12/10. doi: 10.1093/nar/gkt1229 PMID: 24316577
2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–53. Epub 2009/10/09. doi: 10.1038/nature08494 PMID: 19812666
3. Wei WH, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nature reviews Genetics. 2014; 15(11):722–33. Epub 2014/09/10. doi: 10.1038/nrg3747 PMID: 25200660
4. Mackay TF, Moore JH. Why epistasis is important for tackling complex human disease genetics. Genome medicine. 2014; 6(6):42. Epub 2014/07/18. doi: 10.1186/gm561
5. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(4):1193–8. Epub 2012/01/10. doi: 10.1073/pnas.1119675109 PMID: 22223662
6. McClelland GH, Judd CM. Statistical difficulties of detecting interactions and moderator effects. Psychological bulletin. 1993; 114(2):376–90. Epub 1993/09/01. PMID: 8416037
7. Munafo MR, Flint J. Meta-analysis of genetic association studies. Trends in genetics: TIG. 2004; 20 (9):439–44. Epub 2004/08/18. doi: 10.1016/j.tig.2004.06.014 PMID: 15313553
8. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nature reviews Genetics. 2013; 14(6):379–89. Epub 2013/05/10. doi: 10.1038/nrg3472 PMID: 23657481
9. Salanti G, Sanderson S, Higgins JP. Obstacles and opportunities in meta-analysis of genetic association studies. Genetics in medicine: official journal of the American College of Medical Genetics. 2005; 7 (1):13–20. Epub 2005/01/18. PMID: 15654223
10. Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR. Commentary: meta-analysis of individual participants' data in genetic epidemiology. American journal of epidemiology. 2002; 156(3):204–10. Epub 2002/07/27. PMID: 12142254
11. Liu K, Lin X, Zhou Q, Ma T, Han L, Mao G, et al. The associations between two vital GSTs genetic polymorphisms and lung cancer risk in the Chinese population: evidence from 71 studies. PloS one. 2014; 9(7):e102372. Epub 2014/07/19. doi: 10.1371/journal.pone.0102372 PMID: 25036724
12. Shen YH, Chen S, Peng YF, Shi YH, Huang XW, Yang GH, et al. Quantitative assessment of the effect of glutathione S-transferase genes GSTM1 and GSTT1 on hepatocellular carcinoma risk. Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine. 2014; 35 (5):4007–15. Epub 2014/01/09. doi: 10.1007/s13277-013-1524-2
13. Zhu H, Bao J, Liu S, Chen Q, Shen H. Null genotypes of GSTM1 and GSTT1 and endometriosis risk: a meta-analysis of 25 case-control studies. PloS one. 2014; 9(9):e106761. Epub 2014/09/11. doi: 10.1371/journal.pone.0106761 PMID: 25208225

14. Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clinical trials (London, England). 2005; 2(3):209–17. Epub 2005/11/11.

15. Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. BMC medical research methodology. 2005; 5:14. Epub 2005/04/27. doi: 10.1186/1471-2288-5-14 PMID: 15850485

16. Lin C, Chu CM, Lin J, Yang HY, Su SL. Gene-gene and gene-environment interactions in meta-analysis of genetic association studies. PloS one. 2015; 10(4):e0124967. Epub 2015/04/30. doi: 10.1371/journal.pone.0124967 PMID: 25923960

17. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. Statistics in medicine. 1997; 16(17):1965–82. Epub 1997/09/26. PMID: 9304767

18. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Statistics in medicine. 2002; 21(11):1559–73. Epub 2002/07/12. doi: 10.1002/sim.1187 PMID: 12111920

19. Baker WL, White CM, Cappelleri JC, Kluger J, Coleman CI. Understanding heterogeneity in meta-analysis: the role of meta-regression. International journal of clinical practice. 2009; 63(10):1426–34. Epub 2009/09/23. doi: 10.1111/j.1742-1241.2009.02168.x PMID: 19769699

20. Balding DJ, Nichols RA. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica. 1995; 96(1–2):3–12. Epub 1995/01/01. PMID: 7607457

21. Doran H, Bates D, Bliese P, Dowling M. Estimating the multilevel Rasch model: With the lme4 package. Journal of Statistical Software. 2007; 20(2):1–18.

22. Udomsinprasert R, Pongjaroenkit S, Wongsantichon J, Oakley AJ, Prapanthadara LA, Wilce MC, et al. Identification, characterization and structure of a new Delta class glutathione transferase isoenzyme. The Biochemical journal. 2005; 388(Pt 3):763–71. Epub 2005/02/19. doi: 10.1042/bj20042015 PMID: 15717864

23. Frova C. Glutathione transferases in the genomics era: new insights and perspectives. Biomolecular engineering. 2006; 23(4):149–69. Epub 2006/07/15. doi: 10.1016/j.bioeng.2006.05.020 PMID: 16839810

24. DeJong JL, Mohandas T, Tu CP. The human Hb (mu) class glutathione S-transferases are encoded by a dispersed gene family. Biochemical and biophysical research communications. 1991; 180(1):15–22. Epub 1991/10/15. PMID: 1930212

25. Seidegard J, Vorachek WR, Pero RW, Pearson WR. Hereditary differences in the expression of the human glutathione transferase active on trans-stilbene oxide are due to a gene deletion. Proceedings of the National Academy of Sciences of the United States of America. 1988; 85(19):7293–7. Epub 1988/10/01. PMID: 3174634

26. Bolt HM, Thier R. Relevance of the deletion polymorphisms of the glutathione S-transferases GSTT1 and GSTM1 in pharmacology and toxicology. Current drug metabolism. 2006; 7(6):613–28. Epub 2006/08/22. PMID: 16918316

27. Fang J, Wang S, Zhang S, Su S, Song Z, Deng Y, et al. Association of the glutathione s-transferase m1, t1 polymorphisms with cancer: evidence from a meta-analysis. PloS one. 2013; 8(11):e78707. Epub 2013/11/20. doi: 10.1371/journal.pone.0078707 PMID: 24250808

28. Volk DE, Thiviyanathan V, Rice JS, Luxon BA, Shah JH, Yagi H, et al. Solution structure of a cis-opened (10R)-N6-deoxyadenosine adduct of (9S,10R)-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene in a DNA duplex. Biochemistry. 2003; 42(6):1410–20. Epub 2003/02/13. doi: 10.1021/bi026745u PMID: 12578353

29. Lodovici M, Luceri C, Guglielmi F, Bacci C, Akpan V, Fonnesu ML, et al. Benzo(a)pyrene diolepoxide (BPDE)-DNA adduct levels in leukocytes of smokers in relation to polymorphism of CYP1A1, GSTM1, GSTP1, GSTT1, and mEH. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2004; 13(8):1342–8. Epub 2004/08/10.

30. Liu H, Jia J, Mao X, Lin Z. Association of CYP1A1 and GSTM1 Polymorphisms With Oral Cancer Susceptibility: A Meta-Analysis. Medicine. 2015; 94(27):e895. Epub 2015/07/15. doi: 10.1097/md.0000000000000895 PMID: 26166128

31. Aros C, Remuzzi G. The renin-angiotensin system in progression, remission and regression of chronic nephropathies. Journal of hypertension Supplement: official journal of the International Society of Hypertension. 2002; 20(3):S45–53. Epub 2002/08/20.

32. Hollenberg NK. Aldosterone in the development and progression of renal injury. Kidney international. 2004; 66(1):1–9. Epub 2004/06/18. doi: 10.1111/j.1523-1755.2004.00701.x PMID: 15200407

33. Remuzzi G, Bertani T. Pathophysiology of progressive nephropathies. The New England journal of medicine. 1998; 339(20):1448–56. Epub 1998/11/13. doi: 10.1056/nejm199811123392007 PMID: 9811921

34. Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, et al. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. Circulation research. 2000; 87(5):E1–9. Epub 2000/09/02. PMID: 10969042

35. Lin C, Yang HY, Wu CC, Lee HS, Lin YF, Lu KC, et al. Angiotensin-converting enzyme insertion/deletion polymorphism contributes high risk for chronic kidney disease in Asian male with hypertension—a meta-regression analysis of 98 observational studies. PloS one. 2014; 9(1):e87604. Epub 2014/02/06. doi: 10.1371/journal.pone.0087604 PMID: 24498151

36. Chen WJ, Huang YL, Shiue HS, Chen TW, Lin YF, Huang CY, et al. Renin-angiotensin-aldosterone system related gene polymorphisms and urinary total arsenic is related to chronic kidney disease. Toxicology and applied pharmacology. 2014; 279(2):95–102. Epub 2014/06/08. doi: 10.1016/j.taap.2014.05.011 PMID: 24907556

37. Shaikh R, Shahid SM, Mansoor Q, Ismail M, Azhar A. Genetic variants of ACE (Insertion/Deletion) and AGT (M268T) genes in patients with diabetes and nephropathy. Journal of the renin-angiotensin-aldosterone system: JRAAS. 2014; 15(2):124–30. Epub 2014/04/17. doi: 10.1177/1470320313512390 PMID: 24737640

38. Pawlik M, Mostowska A, Lianeri M, Oko A, Jagodzinski PP. Association of aldosterone synthase (CYP11B2) gene -344T/C polymorphism with the risk of primary chronic glomerulonephritis in the Polish population. Journal of the renin-angiotensin-aldosterone system: JRAAS. 2014; 15(4):553–8. Epub 2013/05/18. doi: 10.1177/1470320313489588 PMID: 23681285

39. Su SL, Yang HY, Wu CC, Lee HS, Lin YF, Hsu CA, et al. Gene-gene interactions in renin-angiotensin-aldosterone system contributes to end-stage renal disease susceptibility in a Han Chinese population. Scientific World Journal. 2014; 2014:169798. doi: 10.1155/2014/169798 PMID: 24977181

40. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. Journal of clinical epidemiology. 2002; 55 (1):86–94. Epub 2002/01/10. PMID: 11781126

41. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nature genetics. 2012; 44(4):369–75, s1-3. Epub 2012/03/20. doi: 10.1038/ng.2213 PMID: 22426310

42. Palla L, Higgins JP, Wareham NJ, Sharp SJ. Challenges in the use of literature-based meta-analysis to examine gene-environment interactions. American journal of epidemiology. 2010; 171(11):1225–32. Epub 2010/04/22. doi: 10.1093/aje/kwq051 PMID: 20406760

43. Bhattacharjee P, Paul S, Banerjee M, Patra D, Banerjee P, Ghoshal N, et al. Functional compensation of glutathione S-transferase M1 (GSTM1) null by another GST superfamily member, GSTM2. Scientific reports. 2013; 3:2704. Epub 2013/09/21. doi: 10.1038/srep02704 PMID: 24048194

44. Hung RJ, Boffetta P, Brockmoller J, Butkiewicz D, Cascorbi I, Clapper ML, et al. CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. Carcinogenesis. 2003; 24(5):875–82. Epub 2003/05/29. PMID: 12771031

45. Jeunemaitre X, Soubrier F, Kotelevtsev YV, Lifton RP, Williams CS, Charru A, et al. Molecular basis of human hypertension: role of angiotensinogen. Cell. 1992; 71(1):169–80. Epub 1992/10/02. PMID: 1394429

46. Caulfield M, Lavender P, Newell-Price J, Kamdar S, Farrall M, Clark AJ. Angiotensinogen in human essential hypertension. Hypertension. 1996; 28(6):1123–5. Epub 1996/12/01. PMID: 8952609

47. Mizuiri S, Hemmi H, Kumanomidou H, Iwamoto M, Miyagi M, Sakai K, et al. Angiotensin-converting enzyme (ACE) I/D genotype and renal ACE gene expression. Kidney international. 2001; 60(3):1124–30. Epub 2001/09/05. doi: 10.1046/j.1523-1755.2001.0600031124.x PMID: 11532108

48. Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. The Journal of clinical investigation. 1990; 86(4):1343–6. Epub 1990/10/01. doi: 10.1172/jci114844 PMID: 1976655

49. Wolf G, Neilson EG. Angiotensin II as a renal growth factor. Journal of the American Society of Nephrology: JASN. 1993; 3(9):1531–40. Epub 1993/03/01. PMID: 8507808