Detecting Disease Association With Rare Variants Using Weighted Entropy

Yumei Li (✉ lymmail@126.com)

Huaihua University

Xinrong Xiang

University of Science and Technology of China

Yang Xiang

Huaihua University

Research Article

Keywords: Rare variant, Association analysis, Common variant, Entropy, Weighted entropy

Posted Date: November 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-956366/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Detecting disease association with rare variants using weighted entropy

Yumei Li ^{1, 2, 4*}, Xinrong Xiang ³, Yang Xiang ^{1, 2, 4}.

1 School of Mathematics and Computational Science, Huaihua University,

Huaihua, Hunan 418008, P. R. China

2 Key Laboratory of Research and Utilization of Ethnomedicinal Plant Resources of

Hunan Province, Huaihua University, Huaihua 418008, China;

3 School of Management, University of Science and Technology of China, HeFei, Anhui,

410081, P. R. China

4 Key Laboratory of Hunan Higher Education for Western Hunan Medicinal Plant

and Ethnobotany, Huaihua University, Huaihua 418008, China

Running title: Detecting rare variants using weighted entropy

* Corresponding author: Dr. Yumei Li

School of Mathematics and Computational Science

Huaihua University, Huaihua, Hunan 418000, China

Tel: +86 0745-2864591

Email: lymmail@126.com; liym74@yahoo.com

Abstract

Background: The rapid development of sequencing technology and simultaneously the availability of large quantities of sequence data provides an unprecedented opportunity for researchers to conduct studies to detect rare variants associated with the disease. However, none of current existing statistical methods has uniform power in all scenarios because they more or less are affected by nonfunctional variants and variants with opposite effect. The present study focuses on identifying rare variant associated with the disease.

Results: we present a robust approach to identify rare variant using weighted entropy theory. This approach here takes the proportion of the minor allele among all k variants as its probability distribution, which reduces the noise incurred by non-causal variants, and uses a weight to strike a balance between deleterious rare variants and protective rare variants, which makes our method impacted less by variants with opposite effect. Through simulation studies, we investigate the performance of our method for rare variant association analyses as well as for common variant association analyses and compared it with Burden test and the SKAT-O test. Simulation study show that the proposed method is valid and outperform two existing methods. Meanwhile, the proposed method is affected slightly by non-causal variants and opposite effect variants with high and stable power for various parameters set.

Conclusions: We conclude that the proposed method here can be used effectively to detect rare variant associated with the disease.

Key words: Rare variant, Association analysis, Common variant, Entropy, Weighted entropy

Introduction

With the common disease/common variants assumption, numerous genome-wide association studies (GWAS) have successfully identified susceptibility loci for complex traits. However, common variants identified by current GWAS usually have a relatively small to modest genetic effect and only explain a small fraction of the trait variability [1]. The rapid development of sequencing technology and the lowering of sequencing costs in the last twenty years made large quantities of sequence data available and thus provides an unprecedented opportunity for researchers to conduct studies to detect rare variants in complex traits [2, 3, 4, 5]. Due to low frequency of rare variants, statistical method developed for common variant generally has low power in rare variant association analysis. Many statistical methods for rare variant have been proposed in the last decade and they can be grouped into two classes. The first class of methods, we call them Burden tests [3, 6, 7], is early developed for rare variant association analysis. The main characteristic of them is to seek to assess the overall genetic burden due to rare variants by collapsing cumulative effects of multiple rare variants in a gene region. When the gene region exists a large proportion of the causal rare variants in same direction with equal effect sizes, these tests can dramatically improve statistical power than single-marker based tests. However, Burden tests will suffer from non-causal variants and variants with opposite effect. Their power will loss seriously if misclassify nonfunctional variants or ignore the impact of variants with opposite effect. Another class of methods are variance component tests, such as the C-alpha test [8], the sum of squared score (SSU) test [9], and the sequence kernel association test (SKAT) [10, 11]. These methods assess the distribution of the aggregated score

test statistics using a kernel function. Although they are robust to causal variants with opposite effects, variance component tests can be less powerful than Burden tests when there exist a large proportion of rare variants in the same direction. Lee et al. (2012) proposed an optimal test, which is a linear combination of combine SKAT and the burden tests (SKAT-O), to overcome the influence of the directions of variants effects. Although SKAT-O can achieve robust power, it will lose power when there exist a large proportion of nonfunction rare variants. Because in reality we rarely have much prior information on the true association pattern with causal rare variants and are hard to choose disease-association mutations [13, 14], developing statistical method that can reduce the impacts of nonfunctional variants and variants with opposite effect always becomes important in rare variant association analysis.

The Shannon entropy [15], as an important metrics in information theory, is usually used to measure uncertainty associated with a of a random variable. The mutual information (MI) of two random variables is the difference between the entropy of a variable and the conditional entropy of this variable given the knowledge of another variable [16]. The MI is nonnegative and equal to zero when the two variables are independent. We can apply the entropy theory to construct the mutual information between the entropy of a variant and its conditional entropy given the phenotype (affected or unaffected). At the same time, considering that each rare variant in a gene region may has different contribution to a disease, we can use weighted entropy [17] to construct a measure to quantify the magnitude of association between the variant and the trait. So, in this paper, we will propose a method to detect rare variants using weighted entropy. Through simulation studies, we will assess and compare the performance of

our method with the existing methods of Burden test and the SKAT-O test for rare variant as well as common variant.

Methods

In this study, all datasets were publically available and no research requiring ethics approval was conducted.

We consider an interesting gene region where there is k common and/or rare variants, each variant having a minor allele a and a normal allele A . Let S_i be the number of copy “ a ” at i th variant ($i = 1, L, k$) carried by a subject. We consider a data set that has N^C affected individuals and N^U unaffected individuals ($N^C + N^U = 2N$). For simplicity, the following measure with a superscript “C” indicates a measure in affected individuals, and a measure with a superscript “U” indicates a measure in unaffected individuals. Let S_{ij}^C, S_{ij}^U , and S_{ij} be the number of copies “ a ” at i th variant for j th subject ($i = 1, L, k; j = 1, \dots, N^C/N^U/N$) in affected individuals, in unaffected individuals, and in total individuals admixed by affected individuals and unaffected individuals respectively. Let $p_i = \frac{N}{S_{ij}^C} \left/ \left(\frac{k}{S_{ij}^C} \frac{N}{S_{ij}^C} \right) \right.$, $p_i^C = \frac{N^C}{S_{ij}^C} \left/ \left(\frac{k}{S_{ij}^C} \frac{N^C}{S_{ij}^C} \right) \right.$, and $p_i^U = \frac{N^U}{S_{ij}^U} \left/ \left(\frac{k}{S_{ij}^U} \frac{N^U}{S_{ij}^U} \right) \right.$. p_i, p_i^C , and p_i^U ($i = 1, L, k$) indicate the proportion of the minor allele a for i th variant among all k variants in total individuals admixed by affected individuals and unaffected individuals, in affected individuals, and in unaffected individuals, respectively. Then the proportion of the normal allele A for i th variant among all k variants is $q_i = 1 - p_i$, $q_i^C = 1 - p_i^C$, and $q_i^U = 1 - p_i^U$. Define a variable X as the state (a or A) for i th variant. In total individuals admixed by affected individuals and unaffected

individuals, X has the probability distribution $\Pr(X = a) = p_i$, $\Pr(X = A) = q_i$. In affected individuals, X has the probability distribution $\Pr(X = a) = p_i^C$, $\Pr(X = A) = q_i^C$. In unaffected individuals, X has the probability distribution $\Pr(X = a) = p_i^U$, $\Pr(X = A) = q_i^U$. Let Y be an individual's phenotype state, $Y = 1$ if the individual is affected, $Y = 0$ if the individual is unaffected. Then we can obtain the entropy X , the entropy of X given $Y = y$, the conditional entropy of X given Y , and MI of X and Y for i th variant,

$$H_i(X) = - p_i \times \log p_i - q_i \times \log q_i \quad (1)$$

$$\begin{aligned} H_i(X | Y = 1) &= - p_i^C \times \log p_i^C - q_i^C \times \log q_i^C \\ H_i(X | Y = 0) &= - p_i^U \times \log p_i^U - q_i^U \times \log q_i^U \end{aligned} \quad (2)$$

$$H_i(X | Y) = \sum_{y=0,1} p(Y = y) \times H_i(X | Y = y) \quad (3)$$

$$MI_i = H_i(X) - H_i(X | Y) \quad (4)$$

It can be seen that MI_i is a non-negative measure quantifying the magnitude of association between i th variant and disease: the larger the value, the stronger the association and $MI_i = 0$ holding only if i th variant is independent with disease.

When the state x of X has the weight ω_x , the weighted entropy of X , the weighted entropy of X given $Y = y$, and the weighted conditional entropy of X given Y are given by the following equation (5) ~ (7), respectively.

$$H_{wi}(X) = - w_a \times p_i \times \log p_i - w_A \times q_i \times \log q_i \quad (5)$$

$$\begin{aligned} H_{wi}(X | Y = 1) &= - w_a \times p_i^C \times \log p_i^C - w_A \times q_i^C \times \log q_i^C \\ H_{wi}(X | Y = 0) &= - w_a \times p_i^U \times \log p_i^U - w_A \times q_i^U \times \log q_i^U \end{aligned} \quad (6)$$

$$H_{wi}(X | Y) = \sum_{y=0,1} p(Y = y) \times H_{wi}(X | Y = y) \quad (7)$$

It is noted that the two states (a and A) of X belong to a variant. While we focus on the

different effect among variants, we naturally let $w_a = w_A$ and denote it w_i for i th variant. Let

$$w_i^U = 1/\sqrt{n_i f_i^U (1 - f_i^U)} \text{ and } w_i^{CU} = 1/\sqrt{n_i l_i (1 - l_i)}, \text{ where } f_i^U = (\sum_{j=1}^{N^U} S_{ij}^U + 1)/(2N^U + 2),$$

$$l_i = (\sum_{j=1}^{N^U} S_{ij}^U + \sum_{j=1}^{N^C} S_{ij}^C + 1)/(2N + 2), \text{ and } n_i = N. \quad w_i^U \text{ is the commonly used weight in}$$

association analysis involved rare variant [7]. Dai et al [2012] has point out that the weight

w_i^U tends to favor those deleterious rare variants and ignore the protective rare variants. They

showed that, when rare variants have protective effect for the disease, $w_i^U < w_i^{CU}$. So, we adopt

the maximum value of w_i^U and w_i^{CU} as the weight, that is, $w_i = \max\{w_i^U, w_i^{CU}\}$. It is noted that

$H_{wi}(X) = w_i \times H_i(X)$ and $H_{wi}(X|Y) = w_i \times H_i(X|Y)$. Then we define a statistic Tw_i as following:

$$Tw_i = H_{wi}(X) - H_{wi}(X|Y) = w_i \times \{H_i(X) - H_i(X|Y)\} \quad (8)$$

According to the performance of MI, we can easily see that Tw_i quantifies the magnitude of

association between i th variant and disease: the larger the value, the stronger the association

and $Tw_i = 0$ holding only if i th variant is independent with disease. Now we define Tw as the

maximum of Tw_i among all k variants, i.e.

$$Tw = \max_i \{Tw_i\} \quad (9)$$

We use Tw as a test statistic for association analysis of variant with the disease. The statistical

significance of Tw can be assessed by a permutation procedure.

Results

Simulation studies

In order to assess the performance of the weighted statistic Tw and compare it with the

existing methods of Burden and SKAT-O tests, we conduct a series of simulation studies under

a wide range of parameter values for three types of variants with only rare variants, only

common variants and both rare variants and common variants. The simulation parameter includes the number of variants (k), the minor allele frequency (MAF) of each variant, the effect size (measured by the odds ratio (OR)) of causal variants, and the sample size ($2N$). We consider 50 ($k=50$) variants with MAF values of variants determined by a uniform distribution $U(a, b)$, where, $a=0.001$ and $b=0.05$ for rare variants and $a=0.2$ and $b=0.3$ for common variants. The genotype data for each individual are simulated using the following method [19]: We first generate haplotypes for 50 variants based on a latent variable $Z=(Z_1, L, Z_k)$ from a multivariate normal distribution with covariance structure $\text{cov}(Z_i, Z_j) = \rho^{|i-j|}$ ($\rho=0$ and $\rho=0.6$) between any two latent components. Then we combine two haplotypes to obtain the genotype value for each individual $X_i=(X_{i1}, L, X_{ik})$. The disease status for each individual is determined by the following logistic model [20]:

$$P(\text{Affected} | X_i) = \frac{1}{1 + \exp(-g)}, \quad g = \ln\left(\frac{c}{1-c}\right) + \ln(\text{OR}) \times X_i$$

where c is the background prevalence of being affected for a subject with no minor alleles (here, we let $c=0.01$).

Under the null hypothesis of no association, we let $\text{OR}=1$ for each variant. Under the alternative hypothesis of association, we let $\text{OR}=1$ for non-causal variants, $\text{OR}=2.5$ for causal common variants, and OR for causal rare variants has a uniform distribution $U(a, b)$, where, $a=1.2$ and $b=3$ for rare variants with positive effect, and $a=0.65$ and $b=0.95$ for rare variants with negative effect. We consider 19 scenarios in the simulation study with the parameter values detailed in Table 1. The sample size $2N$ (N affected individuals and N unaffected

Table 1. Summary of various types of variants in the 19 scenarios.

Scenario	Number of variants					Frequency
	RC ⁺	RC ⁻	RC ⁰	CC	CC ⁰	
I: Rare						
1	50	0	0	-	-	0.001-0.05
2	40	0	10	-	-	
3	30	0	20	-	-	
4	20	0	30	-	-	
5	10	0	40	-	-	
6	40	10	0	-	-	
7	35	15	0	-	-	
8	30	20	0	-	-	
9	25	25	0	-	-	
10	25	5	20	-	-	
11	20	10	20	-	-	
12	15	15	20	-	-	
13	15	5	30	-	-	
14	10	10	30	-	-	
II: Rare+Common						
15	30	5	5	2	8	I+III
16	20	5	15	2	8	
17	15	5	20	2	8	
18	10	5	25	2	8	
III: Common						
19	-	-	-	2	48	0.2-0.3

Note: RC⁺= rare causal variant with positive effect; RC⁻= rare causal variant with negative effect; RC⁰= rare non-causal variant; CC= common causal variant; CC⁰= common non-causal variant.

individuals) is chosen as 1000 or 1600. We conduct 500 simulations for each scenario. In each simulation, we first calculate the value of the weighted statistic T_w and then recalculate it by using the permutation procedure 500 times by permuting the “affected” and “unaffected” labels randomly with equal probability. The estimated p-value in each simulation is the proportion of permutation-based statistics that are larger than the data-based statistic in 500 permutations. For a given significance level α , the power/type I error rate is estimated as the proportion of rejecting the null hypothesis when $\text{p-value} \leq \alpha$ in 500 simulations.

Type I error rates and power

Table 2 shows the estimated type I error rates and Figure S1 exhibits QQ-plots of p-values for our proposed weighted statistic T_w under the null hypothesis of no association. As displayed in Table 2 and Figure S1, estimated type I error rates are not appreciably different from the nominal alpha levels under various genetic parameters, suggesting that type I error is well controlled for T_w .

Table 2. Estimated type I error rates of T_w .

The sample size		Estimated Type I error rate					
		Rare		Rare and common		common	
		$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
2N=1000	$\rho = 0$	.048	.013	.052	.009	.049	.013
	$\rho = 0.6$	.049	.012	.051	.013	.052	.008
2N=1600	$\rho = 0$	.052	.009	.049	.011	.049	.012
	$\rho = 0.6$	.049	.012	.051	.012	.052	.013

Figure 1 shows the estimated power for 19 scenarios when sample size is 1600. We can find that powers for $\rho=0$ are similar to those for $\rho=0.6$. Scenario 1 to scenario 14 in Figure 1 show results for only rare variants existing in the gene region. It can be seen that, when rare variant effects are in the same direction (scenarios {1, 2, 3, 4, 5}), three statistics' power decrease with the increasing of the number of non-causal variants, but with the increasing of the number of non-causal variants from zero to 40, power of Burden test and the SKAT-O decreases from 0.98 to 0.36 and 1.0 to 0.54, respectively, yet power of T_w just decreases from 1.0 to 0.76. These show T_w compared with the other two statistics is the least affected by non-causal variants. We can find that, when there are the opposite effect variants (scenarios {6, 7, 8, 9}), although power of T_w is affected by variants with negative effect, the impact is slight compared with those of Burden test and the SKAT-O. For example, power of T_w reach 86% while the SKAT-O has power 73% and Burden test has just 29% when there are 40 variants with positive effect and 10 variants with negative effect. When the number of variants with opposite effect increases, power of three statistics decrease. Even though there is the same number of variants with positive effects and negative effects (scenario 9), T_w has 68% power while the SKAT-O has 52% power and Burden test has only 7%. By comparing the powers under scenarios {3, 10, 11, 12} and scenarios {4, 13, 14}, we can observe that, when there exist non-causal variants and variants with negative effect, three statistics have lower power than those when there only exist non-causal variants, but T_w still has 70~80% power, far higher than those of the SKAT-O (40~70%) and Burden test (9~30%). Scenarios {15, 16, 17, 18} in Figure 1 display the estimated power when both rare variants and common variants exist in the

gene region and scenario19 in Figure 1 is for common variants existing in the gene region. It can be seen that, when there are both rare variants and common variants, Burden test and the SKAT-O are affected by the number of non-causal rare variants and rare variants with opposite

Figure 1. Empirical power of tree statistics for $\rho=0$ (A) and $\rho=0.6$ (B) when sample size is 1600 at a 0.05 significance level.

effect. SKAT-O has lowest power when there are only common variants. These impacts on T_w are slight, because T_w has stable power of above 90%. Meanwhile, T_w achieve high power of 98% for common variants (scenario19).

Figure 2. Empirical power of tree statistics for $\rho=0$ (A) and $\rho=0.6$ (B) when sample size is 1000 at a 0.05 significance level.

Figure 2 shows the estimated power for 19 scenarios when the sample size is 1000. The similar above conclusions are observed as in Figure 1. For example, three statistics' power decrease with the increasing of the number of non-causal variants and negative effect variants. but impacts of non-causal variants and negative effect variants on T_w are slightest among three statistics. T_w under the sample size of 1000 has still 60% to 80% power, though ~ 0.15 lower than those for sample size of 1600.

Discussion

In this paper, we propose a method to identify rare variants associated with complex disease using weighted entropy. Entropy as an important metrics in information theory can be used to measure uncertainty of a random variable. We quantify the magnitude of association between the variant and the trait by constructing the mutual information between the entropy of a variant and its conditional entropy given the phenotype (affected or unaffected). We ascribe each variant a weight to upweight rarer variants and downweight the more common variants. This weight favors those deleterious rare variants as well as those protective rare variants. The performance of our method is assessed by a series of simulation studies. The results show that this approach here achieves over 90% power when variants has same direct effect. The non-causal variants and variants with opposite effect have a minor effect on it compared with those for the Burden test and the SKAT-O test. The power of our method can reach $\sim 70\%$ even though there exists the same number of variants with positive effect and variants with negative effect. Meanwhile, our propose method has high power for detecting common variants.

In rare variant association study, the number of non-causal variants and the effects of

causal variants are normally unknown. How to reduce the influence of them always is the focus work in developing statistical method for statistic geneticist. Most of existing methods are affected by non-causal variants and variants with opposite effect and their power will loss if misclassify nonfunctional variants or ignore the impact of variants with opposite effect in rare variant association analysis. But non-causal variants and variants with opposite effect have a minor effect on our proposed method. Our method having stable power for various paraments set come from the following two reasons: the first is that we consider the proportion of the minor allele a among all k variants as the probability distribution of a , which reduces the noise incurred by non-causal variants; the second is that we use the maximum value of w_i^U and w_i^{CU} as the weight not w_i^U . w_i^U is the commonly used weight in association analysis involved rare variant, but it tends to favor those deleterious rare variants and ignore the protective rare variants [18]. Selecting the weight w_i can strike a balance between deleterious rare variants and protective rare variants, which makes our method impacted less by variants with opposite effect. It is noted that, in reality rare variants and common variants may both contribute to the heritable variability. Due to low power, many statistical approaches developed for common variants cannot be used to detect rare variants. At the same time, many existing statistical approaches developed for rare variants remain unknown to identify common variants. However, our method can be used not only for rare variants but also for common variants. The main reason is that our proposed method combines all variants to construct the proportion of minor allele and thus both considers genetic information of rare variants and common variants.

Conclusions

The statistic T_w is affected slightly by non-causal variants and the opposite effect variants. It outperforms Burden test and the SKAT-O test and has high and relatively stable power for various parameters set.

List of abbreviations

SKAT sequence kernel association test

SKAT-O the optimal test that combines SKAT and the burden test

MI mutual information

MAF minor allele frequency

OR odds ratio

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Authors' contributions

LYM developed the statistical method and wrote the manuscript. XXR developed the statistical method and revised the manuscript. XY did simulation analysis and revised the manuscript. All authors have read and approved the final version of the manuscript.

Acknowledgments

This work was supported by the Foundation of Hunan Double First-rate Discipline Construction Projects of Bioengineering.

References

1. Asimit J, Zeggini E. Rare Variant Association Analysis Methods for Complex Traits. *Annu Rev Genet* 2010, 44 (1): 293–308.
2. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet.* 2010; 11 (6): 415–425.
3. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet.* 2012; 90 (1): 7–24.
4. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hin-dorff LA, Hunter DJ et al. Finding the missing heritability of complex diseases. *Nature.* 2009; 461 (7265): 747–753.
5. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature.* 2013; 493 (7431): 216–220.
6. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *Am J Hum Genet.* 2008; 83 (3): 311–321.
7. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 2009; 5 (2): e1000384.
8. Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melanders M et al. Testing for an unusual distribution of rare variants. *PLoS Genetics.* 2011; 7 (3): e1001322.
9. Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet. Epidemiol.* 2009; 33 (6): 497–507.
10. Wu MC., Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X. Powerful SNP-

- set analysis for case-control genome-wide association studies. *Am J Hum Genet.* 2010; 86 (6): 929–942.
11. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet.* 2011; 89 (1): 82–93.
12. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics.* 2012; 13 (4): 762–775.
13. Pan W, Shen XT. Adaptive tests for association analysis of rare variants. *Genetic Epidemiology.* 2011; 35 (5): 381–388.
14. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 2014; 95 (1): 5–23.
15. Shannon CE. A mathematical theory of communication. *Bell Systems Tech J.* 1948; 27 (3): 379–423.
16. Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. *Patter recognition.* 1999; 32 (1): 71–86.
17. Guiasu S. Weighted entropy. *Reports on mathematical physic.* 1971; 2 (3): 165–179.
18. Dai YL, Jiang RF, Dong JP. Weighted selective collapsing strategy for detecting rare and common variants in genetic association study. *BMC Genetics,* 2012; 13:7.
19. Wang T, Elston RC. Improved power by use of a weighted score test for linkage disequilibrium mapping. *Am J Hum Genet.* 2007; 80 (2): 353–60.
20. Preston MD, Dudbridge F. Utilising family-based designs for detecting rare variant disease

associations. *Ann Hum Genet.* 2014; 78 (2): 129–140.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [figureS1AppendixQQplots.docx](#)