

## Literature Reviews on Methods for Rare Variant Association Studies

Shurong Fang<sup>1\*</sup>, Shuanglin Zhang<sup>2</sup> and Qiuying Sha<sup>2</sup>

<sup>1</sup>Department of Mathematics, Fairfield University, Fairfield, CT 06824, USA

<sup>2</sup>Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, USA

### Abstract

The widespread availability of genome sequencing data has yielded different rare variant association methods in population-based or family-based designs. However, it is challenging to know which method is appropriate in practice. Our purpose of this paper is to provide a general review of the literature for rare variant association studies and suggestions on future research directions. This paper discusses methods for recent rare variant association studies in three categories. The first two categories are for population-based designs, with/without considering the direction of the effects of causal rare variants. In the third category, methods for family-based designs are concluded.

**Keywords:** Family-based design; Population-based design; Quantitative traits; Qualitative traits; Rare variants

### Introduction

Genome-wide association studies (GWAS) have successfully identified a large number of common variants underlying various complex diseases [1,2]. However, current studies suggest that these common variants identified by GWAS only account for a small fraction of disease heritability [2,3]. It is widely recognized that rare variants are considered to be responsible for the missing heritability [2-4]. Next-generation DNA sequencing technologies allow sequencing of parts of the genome for a large number of individuals and the whole genome for a set of individuals, and thus make directly testing rare variants feasible [3,5-9]. Existing single-marker tests used to detect common variants on complex diseases may not be suitable for detecting rare variants due to the allelic heterogeneity and low frequency of rare variants [10]. Several methods using the strategy of collapsing a group of rare variants in a gene or a pathway have been proposed recently. These methods include the cohort allelic sums test (CAST) method [11], the combined multivariate and collapsing (CMC) method [10], the weighted sum (WS) method [12], the variable minor allele frequency threshold (VT) method [13], and the cumulative minor-allele test (CMAT) method [14], among others. All the above methods assume that all causal variants are risk to diseases, while some causal variants are protective to some diseases [15,16]. More recently, several methods considering the direction of the effects of association have been proposed, including the C-alpha test [17], the sequence kernel association test (SKAT) [18], the adaptive sum (aSum) test [19], and the step-up method [16], etc. All aforementioned methods are only applicable to population-based designs with unrelated samples, whereas family-based designs have been shown to improve power to detect causal rare variants [4,20]. In addition, population-based tests can be seriously confounded by population stratification in rare variant association studies while family-based tests are robust to population stratification. So far, only a few family-based rare variant association methods have been developed, including the sib-pair and odds ratio weighted sum statistics (SPWSS, ORWSS) [20], two adaptive weighting methods (AW-FBAT, AW-Joint) [21], the FBAT-based test (FBAT-T) [22], the TOW for a family-based design (TOW-F) [23], and the TOW for sib-pair designs (TOW-sib) [24], among others. In this article, we will provide a general review of the literature for rare variant association studies. We summarize and discuss the methods for recent rare variant association studies in three categories. In the first two categories, we provide the summary on methods in population-based designs, with or without considering the direction of effects of causal rare variants. In the third category, methods in family-based designs are concluded.

### Materials and Methods

#### Category 1: Methods not robust to the direction of effects of causal variants in population-based designs

Recently, the strategy of collapsing a group of rare variants, in a gene or a pathway, has been proposed. Morgenthaler and Thilly [11] developed the cohort allelic sums test (CAST) which collapses rare variants and then compares collapsed allele frequencies in cases and controls. CAST was a milestone of rare variant association studies and started a sequence of collapsing methods in later research. Li and Leal [10] extended the CAST to come up with the combined multivariate and collapsing (CMC) method in which rare variants are collapsed within different subgroups and the information of both collapsed rare variants and common variants is used in the association test. Both CAST and CMC need to choose a fixed minor allele frequency (MAF) threshold to define common and rare variants. Madsen and Browning [12] proposed the weighted sum (WS) method where both common and rare variants can be included, but the variants are weighted according to their allele frequencies. Thus, common variants are giving small weights while rare variants are given large weights. Price et al. [13] proposed the variable minor allele frequency threshold (VT) method which tests the association using the 'optimal' MAF threshold. Zawistowski et al. [14] developed the cumulative minor-allele test (CMAT) which is based on the summation of minor allele counts across all sites for cases and controls. All these methods are burden tests and essentially test the effect of a weighted combination of variants in a genomic region. Let  $X_{ik}$  denote the genotype (number of minor alleles) of the  $i^{th}$  individual at the  $k^{th}$  variant, and  $W_k$  denote the weight for the  $k^{th}$  variant. The aforementioned methods are essentially testing the effect of a weighted combination of variants,

\*Corresponding author: Shurong Fang, Department of Mathematics, Fairfield University, Fairfield, CT 06824, USA, Tel: (203) 254-4000; E-mail: [sfang@fairfield.edu](mailto:sfang@fairfield.edu)

Received February 01, 2016; Accepted February 16, 2016; Published February 18, 2016

Citation: Fang S, Zhang S, Sha Q (2016) Literature Reviews on Methods for Rare Variant Association Studies. Human Genet Embryol 6: 133. doi:10.4172/2161-0436.1000133

Copyright: © 2015 Fang S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

$X_i = \sum_k w_k x_{ik}$  or its function with different definitions of  $w_k$  [25]. More specifically, CAST, CMC, VT, and CMAT set  $w_k = 1$ . CMAT tests the effect of  $X_i$  while CAST, CMC, and VT test the effect of  $I\{X_i \geq 1\}$ , where  $I\{\bullet\}$  is an indicator function. WS tests the effect of  $X_i$  with  $w_k$  to be the inverse square root of the expected variance based on allele frequencies in the controls. These collapsing methods are more powerful than single-variant tests. However, they assume that all causal variants are risk to diseases, while causal variants may be protective to some diseases [15,16]. When both risk and protective variants are present, the above mentioned methods are underpowered because the opposite association effects will counteract each other [26].

## Category 2: Methods robust to the direction of the effects of causal variants for population-based designs

More recently, several methods that are robust to the direction of the effects of causal variants have been proposed. Neale et al. [17] developed a C-alpha test, comparing the expected variance to the actual variance of the distribution of rare variants in cases versus controls. Wu et al. [18] introduced the sequence kernel association test (SKAT), a variance-component score test, testing for association between variants in a region (both common and rare) while adjusting for covariates. Both C-alpha test and SKAT are essentially testing the variance of the effects rather than the mean. Han and Pan [19] proposed the data-adaptive Sum (aSum) test, which incorporates the signs of the observed effects of causal variants into a burden test. It sets  $w_k = \text{sign}(\hat{\beta}_k)$ , where  $\hat{\beta}_k$  is an estimated coefficient of the  $k^{\text{th}}$  variant based on the marginal logistic linear model for qualitative traits. Hoffmann et al. [16] developed the step-up method, where weights can incorporate MAF, the direction of the effects and the threshold all in a single analysis. It sets  $w_k = a_k s_k v_k$ , where  $a_k$  is a continuous weight (e.g., to incorporate allele frequencies);  $s_k$  determines the direction of the variant effect;  $v_k$  is an indicator variable determining whether the  $k^{\text{th}}$  variant should even be in the model at all. Zhang et al. [27] proposed two grouping strategies (GS) based on WS and used data to decide the direction of the effects of causal variant. Also based on WS, Ionita-Laza et al. [28] proposed the replication-based (R) method, where two one-sided replication-based statistics are applied to risk variants and protective variants, respectively. Sha et al. [29] proposed an adaptive clustering method and adaptive weighting method (AC/AW) to detect rare variant association in the presence of neutral and/or protective variants. Both AC/AW methods are applicable to quantitative and qualitative traits, and have clear advantages from power to computational efficiency comparing with existing collapsing methods and data-driven methods that allow neutral and protective variants. The above methods in Category 2 are also essentially testing the effect of a weighted combination of variants. Thus, how to choose appropriate weights is critical to the performance of these methods. Lin and Tang [30] derived theoretically optimal procedures for combining rare mutations and applied a general score-based test (GS) to population-based samples based on regression models. The proposed test statistic is optimal if  $w_k$  is proportional to the set of regression coefficients. Sha et al. [25] proposed a Test for testing the effect of an Optimally Weighted combination of rare variants (TOW). The optimal weights are analytically derived and calculated from

genotypes and traits in a population-based design. Furthermore, TOW can be extended to a Variable Weight TOW (VW-TOW) [25] to include both rare and common variants. Liu et al. [31] applied meta-analysis of single variant association tests, burden tests, and variable threshold tests and developed RAREMETAL, which could include covariates for both quantitative and qualitative traits. Zeng et al. [32] proposed the likelihood ratio test (LRT) and restricted likelihood ratio test (ReLRT) to test the association of rare variants based on the linear mixed effects model. Like SKAT, LRT and ReLRT examine variance component in the mixed model. However, LRT and ReLRT estimate both the null and alternative models, and provide an indirect estimate of heritability explained by rare variants. The disadvantage is that they are computationally time-consuming. Ladouceur et al. [33] suggested that the power of currently proposed statistical methods depends strongly on the underlying hypotheses of the relationship of traits with proportions of causal variants or/and the direction of the associations. The methods in the first category are more powerful than most of the methods in the second category when all or almost all of rare variants in a region are causal and in the same direction of association, while methods in the second category outperform those in the first category when there are both risk and protective variants, and more generally, when a substantial portion of the variants is neutral. How to select an appropriate method when there is limited biological knowledge in practice? It was recommended that both tests in the first and second categories should be used in the settings where prior biological knowledge is limited [34]. Lee et al. [35] combined a burden test and SKAT into the optimal sequence kernel association test (SKAT-O). Specifically, SKAT-O automatically behaves like a burden test when the burden test is more powerful than SKAT, and works as SKAT when SKAT is more powerful. Derkach et al. [36] developed the Fisher's method to combine p-values from two or more complementary tests (Fisher-CT). When most causal variants have the same direction of association, Fisher-CT consistently outperforms SKAT-O, and is often considerably better than burden tests in the first category and non-burden tests in the second category. Sha and Zhang [37] proposed an optimal combination of single-variant tests (OCST) by combining information from the tests of the three classes: only risk variants, both risk and protective variants and only protective variants. Under some scenarios, OCST is consistently more powerful than Fisher-CT and Fisher-CT is consistently more powerful than SKAT-O. Besides studies for qualitative and quantitative traits, Lin and Tang [30] introduced a score test for potentially censored age-at-onset traits. Wang et al. [38] conducted longitudinal data analysis based on TOW and proposed L-TOW to detect rare variant association in population-based designs, since incorporating traits at multiple-time points may increase the statistical power by providing more information than only using the trait at a single-time point.

## Category 3: Methods for family-based designs

For any type of study design, the statistical power will be improved when rare variants are enriched in samples. If one parent carries a copy of a rare allele, half of the offspring are expected to carry it, thus, variants that are rare in the general population could be common in certain families [39]. Therefore, family-based methods may improve power to detect causal rare variants [20,40]. Moreover, for rare variant associations, population-based association tests can be seriously confounded by population stratification while family-based association tests are robust to population stratification. However, there are only a few methods for family-based designs available so far.

**Qualitative traits:** Zhu et al. [4] proposed a two-stage haplotype-

based (HB) method to identify casual rare haplotypes for qualitative traits. First, a set of risk haplotypes is detected using a small proportion of the sample. Then association with that set of identified haplotypes is tested in a larger case-control sample. Feng et al. [20] introduced a sib-pair weighted sum statistic (SPWSS) to detect both rare and common causal variants in a gene or a genomic region. SPWSS uses either affected or discordant sib-pairs in sequencing or genome-wide association data, is not affected by the directionality of the effect of causal variants, and does not require choosing a MAF threshold. Zhu and Xiong [40] transformed a population-based test to the family-based test by calculating the covariance matrix of the functional principal-component scores, and developed the family-based functional principal-component analysis (FPCA) for qualitative traits. De et al. [22] proposed a method (FBAT-T) in a family-based design by extending the traditional single-SNP Family-Based Association Test (FBAT). Sha et al. [26] proposed a test for Testing the Optimally Weighted combination of variants based on data of Parents and Affected Children (TOW-PAC). TOW-PAC tests the combined effect of rare and common variants in a genomic region by using optimal data-driven weights. Later, He et al. [41] incorporated rare-variant association analysis into the transmission disequilibrium test (TDT) framework [42] to analyze trio sequence data and proposed Rare Variant Extensions of the Transmission Disequilibrium Test (RV-TDT). Choi et al. [43] proposed a Family-based Rare Variant Association Test (FARVAT) for extended families. It is based on the quasi-likelihood of whole families. Epstein et al. [44] proposed a framework for rare variants in affected sib-ships (SIB) based on the logic that rare susceptibility variants should be found more often on regions shared identical by descent by affected siblings. They derived both burden and variance-component tests under the SIB framework. SIB does not require variant information from unaffected relatives. Sha and Zhang [24] tested association of an optimally weighted combination of variants for affected sib-pairs (TOW-sib), based on either unrelated individuals or affected sib-pairs.

**Quantitative traits:** Lin and Tang [30] used generalized linear mixed models to capture the dependence of trait and a score test statistic in a family-based design (GS-F). GS-F is applicable to both quantitative and qualitative traits and allows for covariates. Liu and Leal [45] proposed a unified framework of modeling extreme trait genetic associations (MEGA) for direct quantitative trait loci (QTL) mapping. The framework is based on the mixed effect model, which generalizes the Fisher's biometrical model [46], coupled with a likelihood method. The QTL effects are modeled as fixed effect to facilitate joint analysis of multiple rare variants. Using MEGA and appropriate permutation algorithms, many rare variant tests for unrelated individuals can be extended to the tests for family data. Fang et al. [21] proposed two adaptive weighting methods, Adaptive Weighting Family-Based Association Test (AW-FBAT) and Adaptive Weighting Joint Test (AW-Joint). AW-FBAT uses between-family information to calculate adaptive weights and uses within-family information to test for association, while AW-Joint uses joint information of between-family and within-family components to calculate the adaptive weights and to test for association. Fang et al. [23] extended TOW for unrelated individuals to TOW-F, TOW for Family-based data. TOW-F is robust to population stratifications in a wide range of population structures. Feng et al. [47] provided meta-analysis of rare variants in families (META-F) which applied to both single-variant and gene-level association tests.

## Discussion and Future Directions

As we can see in the previous section, many statistical methods in population-based designs for testing associations of rare variants have been developed. The summary of the currently popular methods is shown in Table 1. Based on the objective and prior biological knowledge of a project in practice, researchers could narrow their choices and select appropriate methods. For rare variant associations, population-based association tests can be seriously confounded by population stratification, since the spectrum of rare variation can be very different in diverse populations. Current studies show that family-

Method	Direction	Quantitative traits	Qualitative traits	Include common variants	Proportion of neutral variants	Covariates	Population stratification
CAST	N	N	Y	N	Y	N	N
CMC	N	N	Y	Y	N	N	N
WS	N	N	Y	Y	N	N	N
VT	N	Y	Y	N	N	N	N
CMAT	N	N	Y	N	Y	Y	Y
C- $\chi^2$	Y	N	Y	N	Y	N	N
SKAT	Y	Y	Y	Y	N	Y	Y
aSum	Y	N	Y	Y	N	N	N
step-up	Y	Y	Y	Y	N	N	N
AC/AW	Y	Y	Y	Y	Y	N	Y
TOW	Y	Y	Y	N	Y	Y	Y
VW-TOW	Y	Y	Y	Y	Y	Y	Y
R	Y	N	Y	N	Y	N	N
GS	Y	Y	Y	Y	N	Y	Y
SKAT-O	Y	Y	Y	N	Y	Y	Y
Fisher-CT	Y	Y	Y	N	Y	Y	N
OCST	Y	Y	Y	N	Y	Y	N
LRT	Y	Y	N	N	N	Y	Y
ReLRT	Y	Y	N	N	N	Y	Y
RAREMETAL	Y	Y	Y	Y	N	Y	Y

Table 1: Summary of statistical methods for population-based designs.

Method	Direction	Quantitative traits	Qualitative traits	Include common variants	Proportion of neutral variants	Covariates	Population stratification
AW-FBAT; AW-Joint	Y	Y	N	N	Y	N	Y
FBAT-T	N	N	Y	Y	N	N	Y
TOW-F	Y	Y	N	N	Y	N	Y
MEGA	N	Y	N	N	N	N	Y
SPWSS	Y	Y	Y	Y	Y	Y	Y
FPCA	Y	N	Y	Y	N	N	Y
GS-F	Y	Y	Y	Y	N	Y	Y
HB	N	N	Y	N	N	N	N
TOW-PAC	Y	N	Y	Y	Y	N	Y
RV-TDT	Y	N	Y	N	N	N	Y
FARVAT	Y	N	Y	N	N	N	Y
META-F	Y	Y	N	N	Y	Y	Y
SIB	Y	N	Y	N	Y	N	Y
TOW-sib	Y	N	Y	Y	Y	N	N

**Table 2:** Summary of statistical methods for family-based designs.

based association tests can be robust to population stratification. In family-based data, association information can be partitioned into between-family and within-family information [21]. Within-family information is robust to population stratification while between-family information can be confounded by population stratification. In addition, from the statistical power point of view, the power under any type of study design can be improved when rare variants are enriched in the samples [20,40]. Several popular statistical methods in family-based designs are summarized in Table 2 for researchers to compare. With the rapid advance of biotechnology, new biological knowledge will become available, and next-generation DNA sequencing technologies will allow sequencing the whole genome. It is significantly important to incorporate this new information to improve statistical power to detect rare variants associated with complex diseases. Continued development of novel statistical methods for identifying rare disease susceptibility variants is needed for population-based designs, and especially for family-based designs. We hope this paper can help researchers with practical problems on rare variants. Most of the methods summarized in this paper considered a single trait. However, a gene often affects multiple traits. Thus, analysis of multiple traits simultaneously will increase power to detect rare variant association. When the same variants affect multiple traits, trait values for an individual will tend to be correlated. Very few methods for common variants association studies have been proposed [48,49]. However, this field is still under way and challenging, and needs our special attention. Meta-analysis has facilitated many discoveries in common variant association studies. It is essential for detecting associations with rare variants too, because meta-analysis can increase the sample size, especially for rare variant association studies. To better explore the relationship between rare variants and complex diseases, it is urgent and essential to develop efficient multiple-trait methods as well as meta-analysis for rare variants studies.

#### Acknowledgement

There are no sources of financial support.

#### References

- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316: 1336-1341.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty

and challenges. *Nat Rev Genet* 9: 356-369.

- Bansal V, Libiger O, Torkamani A, Schork NJ (2010) Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet* 11: 773-785.
- Zhu X, Feng T, Li Y, Lu Q, Elston RC (2010) Detecting rare variants for complex traits using family and unrelated data. *Genet Epidemiol* 34: 171-187.
- Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, et al. (2006) Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci USA* 103: 1810-1815.
- Andre's A, Clark A, Shimmin L, Boerwinkle E, Sing C, Hixson JE (2007) Understanding the accuracy of statistical haplotype inference with sequence data of known phase. *Genetic Epi* 31: 659-671.
- Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, et al. (2007) Medical sequencing at the extremes of human body mass. *Am J Hum Genet* 80: 779-791.
- Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, et al. (2008) Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 40: 592-599.
- Metzker ML (2010) Sequencing technologies-the next generation. *Nature Reviews Genetics* 11: 31-46.
- Li B, Leal SM (2008) Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 83: 311-321.
- Morgenthaler S, Thilly WG (2007) A strategy to discover genes that carry multiallelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res* 615: 28-56.
- Madsen BE, Browning SR (2009) A group-wise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 5: e1000384.
- Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, et al. (2010) Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* 86: 832-838.
- Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zollner S (2010) Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. *Am J Hum Genet* 87: 604-617.
- Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305: 869-872.
- Hoffmann TJ, Marini NJ, Witte JS (2010) Comprehensive approach to analyzing rare genetic variants. *PLoS ONE* 5: e13584.
- Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, et al. (2011) Testing for an Unusual Distribution of Rare Variants. *PLoS Genetics* 7: e1001322.
- Wu M, Lee S, Cai T, Li Y, Boehnke M (2011) Rare variant association testing

- for sequencing data using the sequence kernel association test (SKAT). *Am J Hum Genet* 89: 82-93.
19. Han F, Pan W (2010) A data-adaptive sum test for disease association with multiple common or rare variants. *Hum Hered* 70: 42-54.
20. Feng T, Elston R, Zhu X (2011) Detecting rare and common variants for complex traits: sibpair and odds ratio weighted sum statistics (SPWSS, ORWSS). *Genetic Epidemiol* 35: 398-409.
21. Fang S, Sha Q, Zhang S (2012) Two adaptive weighting methods to test for rare variant associations in family-based designs. *Genet Epidemiol* 36: 499-507.
22. De G, Yip W-K, Ionita-Laza I, Laird N (2013) Rare Variant Analysis for Family-Based Design. *PLoS ONE* 8: e48495.
23. Fang S, Zhang S, Sha Q (2013) Detecting association of rare variants by testing an optimally weighted combination of variants in family-based designs. *Annals of Human Genetics* 77: 524-34.
24. Sha Q, Zhang S (2015) Test of rare variant association based on affected sib-pairs. *Eur J Hum Genet* 23: 229-237.
25. Sha Q, Wang X, Wang X, Zhang S (2012) Detecting association of both rare and common variants by testing an optimally weighted combination of variants. *Genetic Epidemiol* 36: 561-571.
26. Sha Q, Wang S, Zhang S (2013) Adaptive clustering and adaptive weighting methods to detect disease associated rare variants. *Eu J Hum Genet* 21: 332-7.
27. Zhang L, Pei Y-F, Li J, Papiasian CJ, Deng H-W (2010) Efficient utilization of rare variants for detection of disease-related genomic regions. *PLoS ONE* 5: e14288.
28. Ionita-Laza I, Buxbaum JD, Laird NM, Lange C (2011) A new testing strategy to identify rare variants with either risk or protective effect on disease. *PLoS Genet* 7: e1001289.
29. Sha Q, Zhang S (2013) A novel test for testing the optimally weighted combination of rare and common variants based on data of parents and affected children. *Genetic Epidemiol* 38: 135-43.
30. Lin D, Tang Z (2011) A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* 89: 354-367.
31. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M (2014) Meta-analysis of gene level tests for rare variant association. *Nature genetics* 46: 200-204.
32. Zeng P, Zhao Y, Liu J, Liu L, Zhang L, et al. (2014) Likelihood ratio tests in rare variant detection for continuous phenotypes. *Ann Hum Genet*. 78: 320-32.
33. Ladouceur M, Dastani Z, Aulchenko YS, Greenwood CMT, Richards JB (2012) The empirical power of rare variant association methods: results from sanger sequencing in 1,998 individuals. *PLoS Genet* 8: e1002496.
34. Basu S, Pan W (2011) Comparison of statistical tests for disease association with rare variants. *Genet Epidemiol* 35: 606-619.
35. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ (2012) Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 91: 224-237.
36. Derkach A, Lawless J, Sun L (2013) Robust and powerful tests for rare variants using fisher's method to combine evidence of association from two or more complementary tests. *Genetic Epi* 37: 110-121.
37. Sha Q, Zhang S (2014) A rare variant association test based on combinations of single-variant tests. *Genetic Epidemiol* 38: 494-501.
38. Wang S, Fang S, Sha Q, Zhang S (2014) Detecting association of rare and common variants by testing an optimally weighted combination of variants with longitudinal data. *BMC Proceedings* 8: S91.
39. Shi G, Rao D (2011) Optimum designs for next-generation sequencing to discover rare variants for common complex disease. *Genetic Epidemiol* 35: 572-579.
40. Zhu Y, Xiong M (2012) Family-based association studies for next-generation sequencing. *The Am J Hum Genet* 90: 1028-1045.
41. He Z, O'Roak B, Smith J, Wang G, Hooker S (2014) Rare-variant extensions of the transmission disequilibrium test: application to autism exome sequence data. *Am J Hum Genet* 94: 33-46.
42. Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52: 506-516.
43. Choi S, Lee S, Cichon S, Nothen M, Lange C (2014) FARVAT: a family-based rare variant association test. *Bioinformatics*. 15: 197-205
44. Epstein M, Duncan R, Ware E, Jhun M, Bielak L (2015) A statistical approach for rare-variant association testing in affected sibships. *Am J Hum Genet* 96: 543-554.
45. Liu D, Leal S (2012) A unified framework for detecting rare variant quantitative trait associations in pedigree and unrelated individuals via sequence data. *Hum Hered* 73: 105-122.
46. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*, ed 4. Essex, Longman.
47. Fang S, Pistic G, Zhang H, Zawistowski M, Mulas A (2015) Methods for association analysis and meta-analysis of rare variants in families. *Genet Epidemiol* 39: 227-38.
48. Ferreira MAR, Purcell SM (2009) A multivariate test of association. *Bioinformatics* 25: 132-133.
49. Furlotte N, Eskin E (2015) Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model. *Genetics* 200: 59-68.