·综述·

基于家系数据的罕见变异关联分析方法研究进展

陈曦[1] 王斯悦[1] 薛恩慈[1] 王雪珩[1] 彭和香[1] 范梦[1] 王梦莹[1] 武轶群[1] 秦雪英[1] 李劲[1] 吴涛[1] 朱洪平[2] 李静[3] 周治波[2] 陈大方[1] 胡永华[1]

[1]北京大学公共卫生学院流行病与卫生统计学系，北京 100191；[2]北京大学口腔医学院口腔颌面外科，北京 100081；[3]北京大学口腔医学院儿童口腔科，北京 100081

通信作者：吴涛，Email:twu@bjmu.edu.cn

【摘要】 二代测序技术的发展促进了复杂疾病致病性罕见遗传变异的研究。罕见变异的低频性使得单位点关联分析功效不足，因此负荷检验、方差成分检验等整合多个位点信息的关联分析方法得到了广泛应用。但这些方法大多基于人群研究设计，针对家系数据的分析方法较为少见。本文综述了基于家系数据的常用罕见变异关联分析方法，介绍基本原理和特点、适用条件等，并讨论了当前分析方法存在的不足和未来发展的方向。

【关键词】 罕见变异； 家系数据； 多位点关联分析

基金项目：国家自然科学基金（81230066，81872695，81573225）

Family-based association tests for rare variants

Chen Xi[1], Wang Siyue[1], Xue Enci[1], Wang Xueheng[1], Peng Hexiang[1], Fan Meng[1], Wang Mengying[1], Wu Yiqun[1], Qin Xueying[1], Li Jing[1], Wu Tao[1], Zhu Hongping[2], Li Jing[3], Zhou Zhibo[2], Chen Dafang[1], Hu Yonghua[1]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; [2]Department of Oral and Maxillofacial Surgery, Peking University School and Hospital of Stomatology, Beijing 100081, China; [3]Department of Pediatric Dentistry, Peking University School and Hospital of Stomatology, Beijing 100081, China

Corresponding author: Wu Tao, Email: twu@bjmu. edu.cn

【Abstract】 Next-generation sequencing has revolutionized family-based association tests for rare variants. As the lower power of genome wide association study for detecting casual rare variants, methods aggregating effects of multiple variants have been proposed, such as burden tests and variance component tests. This paper summarizes the methods of rare variants association test that can be applied for family data, introduces their principles, characteristics and applicable conditions and discusses the shortcomings and the improvement of the present methods.

【Key words】 Rare variant; Family data; Multi-locus association study

Fund programs: National Natural Science Foundation (81230066, 81872695, 81573225)

复杂疾病（complex diseases）严重威胁着人类健康，其发病通常受多个基因及环境因素的影响[1]，定位致病性遗传因素始终是当前研究的难点和热点。现代分子生物技术的发展为复杂疾病的遗传病因探索提供了重要的技术支持。至2021年6月，GWAS Catalog数据库已收录了2万多项独立的全基因组关联研究（genome-wide association study，GWAS），包含超过25万对基因-表型关联（$P<1\times10^{-5}$）[2]，研究结果促进了复杂疾病遗传病因学、发病机制、靶向药物和风险预测等研究的发展[3]。目前由GWAS得到的单核苷酸多态性（single-nucleotide polymorphism，SNP）阳性位点效应普遍较低[4-5]，阳性发现仅能解释有限的遗传度，存在“缺失的遗传度（missing heritability）”现象[6]，即尚有大量遗传致

DOI：10.3760/cma.j.cn112338-20211224-01013

收稿日期 2021-12-24 **本文编辑** 李银鸽

引用格式：陈曦, 王斯悦, 薛恩慈, 等. 基于家系数据的罕见变异关联分析方法研究进展[J]. 中华流行病学杂志, 2022, 43(9): 1497-1502. DOI: 10.3760/cma.j.cn112338-20211224-01013.

Chen X, Wang SY, Xue EC, et al. Family-based association tests for rare variants[J]. Chin J Epidemiol, 2022, 43(9): 1497-1502. DOI: 10.3760/cma.j.cn112338-20211224-01013.

病因素未被发现。且这些阳性位点在基因组定位分散，存在于不同基因的阳性位点使得GWAS结果的生物学解释和结果利用成为一项难题[7]。为了寻找缺失的遗传度，研究者们提出了其他类型的研究策略，罕见变异（rare variants）研究便是其中之一[3]。

罕见变异通常指人类基因组中弱势等位基因频率（minor allele frequency，MAF）<1%的位点，而GWAS研究检测的遗传位点通常为MAF>5%的常见变异（common variants）[8]，因此难以发现罕见遗传变异位点与疾病之间的关联。既往研究表明，许多孟德尔疾病由高度罕见的遗传突变引起[9]，因此研究者提出复杂疾病的风险也可能受到罕见遗传变异的影响。相对常见变异而言，这类位点频率更低（MAF<1%）但遗传效应可能更高[10]。二代测序技术为定位致病性罕见遗传变异提供了检测手段基础[11]，基因填补算法[12]和高质量单体型参考序列[13]使得低频位点也能够实现高精度的填补，因而基于大样本人群探索罕见变异致病效应变得可行[14]。这类关联研究也取得了一系列成果，癌症、精神病和慢性病等多种疾病的疾病机制探索中发现了能够引起功能改变的罕见位点，也为复杂疾病的药物利用提供了依据[15]。如前列腺素合成酶（PTGIS）的罕见变异能够解释6.1%的肺动脉高压患者病因，且携带这一突变的患者对前列环素类药物治疗更为敏感，或可为精准用药提供指导[16]。

罕见变异的单位点关联研究通常功效过低，因此研究者们开发了适合罕见变异的基于基因/区域的关联研究，即整合某一基因/区域内的位点的关联信号，从而提高统计功效[17]，如负荷检验[18-19]、方差成分检验[20-22]等，这些方法主要适用于研究对象为无亲缘关系人群的研究设计。相对而言，针对家系设计的罕见变异关联研究统计方法较为有限。然而，家系设计在罕见变异关联研究中存在诸多优势。由于罕见变异的基因频率较低，因此在一般人群中较难观察到，但在具有相似遗传背景的家系内部则有可能在多个个体中出现，甚至观察到纯合子个体[23-24]。此外，由于罕见变异的等位基因频率在不同地域、人群间差异较大[25-26]，常用的人群分层调整方法如主成分分析并不适用[8]，而家系设计则能够较好地控制人口分层的问题，在先天性疾病、慢性病的病因学研究中应用广泛。因此，基于家系设计探索罕见变异与复杂疾病的关系具有一定的优势，发展家系设计的罕见变异关联研究分析方法十分必要。

家系数据的罕见变异关联分析方法主要面临两个问题，控制家系内成员的遗传相关性和综合多位点的效应。控制遗传相关性常采用的方法包括：①直接考虑等位基因在亲代和子代之间的传递过程，如传递不平衡检验（transmission disequilibrium test，TDT）和基于家系关联性检验（family-based association test，FBAT）[27-46]，计算已知亲代基因型情况下子代的基因型的条件概率；②采用亲缘关系矩阵（kinship matrix）描述不同个体间的关系，将这类矩阵内置于人群数据的分析方法中[47-65]，如famSKAT-RC等。综合多位点效应的方法则与人群研究中类似，通常采用负荷检验或方差成分检验等，进行以基因/区域为基础的关联研究，而既往为探索常见位点而发展的多位点关联研究分析方法常也可用于罕见位点。部分常见的可用于家系数据的罕见位点关联研究方法见表1。

一、TDT

由Spielman等[28]提出的基于连锁和连锁不平衡的分析方法，TDT假设存在与疾病相关的致病位点，检验标记位点与致病位点是否存在关联及连锁。

Chen等[29]在TDT基础上开发了基于单体型的gTDT（group-wise TDT），将单体型内的多个罕见位点综合为一个单元进行关联检验，并可检验不同遗传模型，如加性模型、显性模型、隐性模型，以及复杂交互作用等。研究者利用gTDT对116个自闭症核心家系的神经递质相关基因外显子测序数据进行了分析，在非5-羟色胺受体基因中发现了部分功能性罕见位点的过度传递，而在5-羟色胺受体基因中则发现了功能性罕见位点的低传递，提示神经递质通路基因的表达在自闭症发病过程中可能起到重要作用[30]。

除了利用单个患者及其双亲的基因及表型数据的情况，TDT还有多种延伸方法，包括利用同胞对、处理父母表型缺失数据的情况及复等位基因位点（同源染色体的相同位点上，存在>2种的等位基因）等[31-34]。FBAT亦可同时检验连锁和关联，并可以用于检验数量性状，还可以处理不完整核心家系的数据[35]。其基本原理为构建每个核心家系中子代的表型与基因型的线性组合统计量，汇总各个家系的数据后采用χ^2检验进行关联分析。FBAT分为两个步骤：第

表1 家系数据罕见位点关联分析方法

方法	人群内相关性	多位点检验	性状	软件包	参考文献
rvTDT	条件	负荷检验，SKAT	二分类性状	R：rvTDT	[66]
gTDT	条件	负荷检验	二分类性状	C++	[29]
RVGDT	条件	负荷检验	二分类性状	Python：RV-GDT	[67]
FBAT	条件	负荷检验，SKAT	二分类和连续型性状	FBAT package	[36,43]
RareIBD	条件	负荷检验	二分类和连续型性状	RareIBD software	[68]
famSKAT	非条件	SKAT	连续型性状	R：famSKATRC	[48,69]
FARVAT	非条件	负荷检验，SKAT	二分类性状	C++	[57]
gskat	非条件	KM test	二分类和连续型性状	R：gskat	[52]
SMMAT	非条件	负荷检验，SKAT	二分类和连续型性状	R：GMMAT	[62]

一步检验标记位点和致病位点间是否存在连锁，第二步通过将子代基因型数据随机化来估计标记位点基因型的分布情况[27]。

在每个核心家系仅有一个患病子代且只考虑一个标记位点的情况下，X_i和Y_i分别代表第i个子代（共n个家系，n个子代）的基因型和表型，在加性模型中，X_i是弱势等位基因的总数，定义：

$$U = \sum_{i=1}^{n}\left(Y_i - \mu\right)\left(X_i - E\left(X_i|P_i\right)\right)$$

其中$E\left(X_i|P_i\right)$为在已知表型和亲代基因型（P_i）的条件下，标记位点和致病位点遵从孟德尔定律不存在连锁情况下子代基因型的期望。在同样的条件分布下可计算$\mathrm{Var}\left(X_i|P_i\right)$，则标化FBAT统计量：

$$Z = U/\sqrt{V}$$

其中，$V = \mathrm{Var}(U) = \sum_{i=1}^{n}\left(Y_i - \mu\right)^2\mathrm{Var}\left(X_i|P_i\right)$，$Z$服从$N(0, 1)$。而当考虑$M$个位点的情况时，则对第$s$个位点有$U_s$和$Z_s$，罕见位点的FBAT统计量定义：

$$W = \sum_{s=1}^{M}U_s = \sum_{s=1}^{M}\sum_{i=1}^{n}\left(Y_i - \mu\right)\left(X_{is} - E\left(X_{is}|P_{is}\right)\right) = \sum_{i=1}^{n}\left(Y_i - \mu\right)\left[\sum_{s=1}^{M}\left(X_{is} - E\left(X_{is}|P_{is}\right)\right)\right]$$

标化的FBAT统计量为$Z = W/\sqrt{\mathrm{Var}(W)}$[36]。这种通过将$M$个位点的信息汇总为一个可以用于关联分析的遗传负荷得分的方法也叫作负荷检验，它假设同时进行分析的一组位点的效应方向和大小相同[37]，但实际情况却更为复杂，同一基因内的位点很可能效应方向不同，部分位点对研究的表型并无影响，部分位点起保护作用，另一些位点为风险位点[20]；而效应大小更难估计，尤其是在同时纳入常见和罕见位点的情况下。针对效应大小可能存在的不一致问题，研究者们提出对不同位点赋予不同的权重；而对于效应方向不一致的问题，则可在FBAT中引入方差成分检验的方法。

加权方法主要有两大类，分别为根据数据本身包含的信息进行加权和根据位点已有的生物学信息进行加权。Guo和Shugart[38]比较了几种不同权重设置方法下的FBAT模型的表现。第一种（FBAT-v1）权重基于频率：

$$w_j = 1/\sqrt{Np_j(1 - p_j)}$$ [39]

式中，p_j为样本人群中第j个位点的等位基因频率。第二（mFBAT-LC）、三（S_β）种方法则均基于回归模型：

$$E\left(T_{il}\right) = \alpha + \beta E(X_{ilj})$$ [40]　（1）

$$E\left(T_{il}\right) = \alpha + \sum_{j=1}^{M}\beta_j E(X_{ilj})$$ [38]　（2）

则$w_j = \widehat{\beta_j}$。其中式（2）为多重遗传标记回归模型，另外两种模型与其相似，但分别引入了LASSO惩罚因子（$S_{\beta.lasso}$）和LASSO及岭回归的混合惩罚因子（$S_{\beta.elasticnet}$）。4种基于回归的方法均不需要亲代的表型数据。模拟结果显示，当基因中大多数罕见变异有相同效应方向时，不加权和频率加权的方法有更好的功效；但综合不同模拟情况，mFBAT-LC和$S_{\beta.elasticnet}$表现更好；4种回归的方法能够克服效应方向不同以及样本量不足的问题而达到较为可观的功效[38]。

Hooli等[41]在美国国立精神卫生研究所（National Institute of Mental Health）的家系队列人群中对470个先证者的*TREM2*基因第2号外显子区域进行测序寻找罕见功能突变位点，然后对携带罕见功能突变的25个先证者的核心家系成员（65个病例和17个对照）进行测序，采用不加权（即假设各位点效应相同）FBAT和MAF加权（假设各位点效应与MAF成反比）FBAT对该区域的4个罕见错义突变位点进行负荷检验，发现不加权情况下4个位点的总体关联效应更强（加权：p=1.8×10^{-3}，不加权：p=4.8×10^{-4}），其中最常见的位点rs75932628对总效应贡献度最大，侧面佐证了既往病例对照研究中发现的阳性位点rs75932628与阿尔茨海默病的高发病风险有关。

方差成分检验是一类在随机效应模型下通过估计一组变异位点的遗传效应的分布情况进行关联分析的方法[8]。不同于负荷检验直接估计一组位点的总体效应，方差成分检验首先估计每一个位点的效应，然后采用多元回归等模型对代表各位点效应的统计量进行合并，估计各位点效应方差对总体方差的贡献，因此允许不同的位点效应方向和效应值不同[8]。在人群研究中，研究者们提出了C-alpha检验[21]、序列核关联检验（sequence kernel association test，SKAT）[20]和方差得分合计检验（sum of squared score test，SSU）[42]等方差成分检验的方法。其中SKAT是关联分析中最常用的方差成分检验方法之一，其优点在于其统计量近似服从χ^2分布，因此不需要通过置换的方法获得P值，计算上更加简洁。此外，模型中引入了协变量，能够纳入遗传位点以外的变量、估计交互作用，数量性状和二分类性状均适用。C-alpha检验和SSU检验可被看作是SKAT的特殊情况[20]。Ionita-Laza等[43]将FBAT与SKAT结合，实现了对家系数据进行多位点的方差成分检验，其统计量Q有如下形式：

$$Q = \sum_{s=1}^{M}w_s^2\left[\sum_{i=1}^{n}(Y_i - \mu)(X_{is} - E\left(X_{is}|P_{is}\right))\right]^2$$

此外，通过设置不同的参数，该模型能够适用于不同研究设计或采用不同检验方法，因而可以比较不同情况下、不同流行病学设计和统计学方法的优劣。模拟数据集分析的结果显示，当目标区域内疾病易感性位点的比例升高时，负荷检验的功效高于SKAT，与既往研究中结果相符。若纳入的位点既有保护性位点也有风险位点且与表型相关的位点所占比例较少时，方差成分检验功效更高。而当位点效应方向相同且效应大小相近时，负荷检验的功效更高[44]。对于二分类性状而言，使用家系设计与人群设计有相似的功效，但对于连续性状，人群设计功效更高。作者还提出，FBAT目前只利用了家系内部的信息，如果能有效利用家系间的信息，其功效可能会更高[45-46]。

二、基于亲缘关系矩阵

基于FBAT的方法由于需要计算在父母/同胞基因型/表

型条件下患者基因型的期望值，因此通常应用于核心家系或以核心家系为单位的大家系，且其目前能结合的多位点模型有限，通常需要满足各位点效应方向相同的假设。因此研究者们提出引入亲缘关系矩阵的方法，这类模型在实际应用中更加灵活，能够处理一般家系结构的数据，同时考虑家系内和家系间的差异，并且能够处理效应不一致的一组位点，因此应用也更为广泛[47]。借助亲缘关系矩阵，负荷检验、方差成分检验等既往基于一般人群所构建的罕见位点关联分析方法能够更加广泛地运用于家系数据中。利用亲缘关系矩阵的方法通常包括线性混合模型（linear mixed model，LMM）和广义估计方程（generalized estimating equations，GEE）。

famSKAT是较为常用的罕见变异关联分析方法之一[48]，其利用LMM在SKAT中引入了家系内部相关性的变量[20]，并且能够同时纳入家系样本和人群样本并计算遗传度。Haller等[49]首先对287名非洲裔美国人和1 028名欧洲裔美国人的尼古丁受体基因CHRNA5、CHRNA3、CHRNA4、CHRNA6和CHRNB3进行测序，选出其中的非同义突变位点，随后检测这些位点在上述人群的亲属（2 504名非裔美国人和7 318名欧洲裔美国人）中的基因型，然后利用famSKAT检验这些位点的变异是否与尼古丁依赖以外的物质依赖（乙醇和可卡因）存在关联，首次发现了CHRNB3和CHRNA3中的罕见位点变异与乙醇或可卡因依赖性风险升高相关。Schifano等[50]也采用LMM处理研究对象间的相关性，但选择使用核函数（kernel machine，KM）对位点效应进行综合[51]，基于KM法的优势在于能够通过指定不同的核函数灵活地进行建模，可以采用负荷检验也可采用方差成分分析[50]。但上述方法仍存在一定的缺陷，均仅适用于连续型变量的表型而不能用于二分类性状。

为解决这一问题，提出了GEE-KM模型[47,52]，这一模型同样采用KM法进行建模，但利用GEE处理研究对象间的相关性，该模型可应用于连续形状和二分类性状，也可通过选择不同的核函数实现负荷检验、方差成分检验或其他检验方法，实现了多种数据情景下的灵活建模。一项对砷代谢效率的研究在一般人群样本中采用亲缘关系矩阵调整人群结构后，采用该模型分别观察了在负荷检验和SKAT两种检验方法下的结果，结果均显示*AS3MT*基因上的罕见位点与二甲基胂酸（砷在哺乳动物体内的主要甲基化代谢产物）百分比存在显著的关联[53]。

基于TDT和亲缘关系矩阵的方法各有其特点和适用的条件。TDT本质为根据患者和父母的基因型推断传递与未传递的等位基因，运用χ^2检验探索等位基因与表型是否存在关联，因此标准的TDT需要获得患者及其父母的基因型数据，通常来说较为适用于早发疾病。当双亲基因型数据难以获得时，研究者们提出了改进方法，如通过患者同胞基因型推断亲本基因型等[54]。而亲缘关系矩阵则是通过相关系数来描述两研究对象间的亲缘关系远近，因此适用于复杂家系结构的数据，在实际操作中更为便利，能够更为充分地运用每个样本的数据。然而，由于家系研究通过纳入患者及其亲属募集样本，因此病例和对照纳入的概率与实际在人群中的分布存在一定的偏差，TDT通过虚拟对照较好地解决了这一问题。但复杂家系中，通常亲属中的患者更倾向于参与研究，缺乏完整的患者-双亲样本使得无法构造虚拟对照，因而通常会导致假阳性率偏高，针对这一问题，研究者们通过重抽样、配对等方法进行校正[47]。

除在解决家系结构、多位点关联检验方法等方面有所进展，研究者们还致力于解决其他可能影响表型-位点关联估计的因素。Oualkacha等[55]构建FaST-LMM，考虑了目标区域以外的位点对表型的影响，并将这部分效应作为多基因随机效应纳入模型。其调整家系内相关性的方法为估计家系内任意两成员间的同源基因（identical by descent，IBD）的比例，IBD的比例可以用基于家族关系的期望值来计算，也可以根据成对个体之间全基因组等位基因计数的协方差来估计，后者被称为RRM（realized relationship matrix）[56]。该方法的另一特点为采用限制性极大似然估计计算统计量，因此统计量不依赖于固定效应，只与零假设下的方差成分估计有关，大大降低了计算难度。FARVAT基于拟似然估计[57]，它能适用于存在群体子结构的情况，并且推广到SKAT-O。而MONSTER则是将SKAT-O应用于家系数据的情况[58]，能够自适应地调整未知的罕见变异效应结构，它还提供了一种评估*P*值的方法，起到类似于置换检验的效果。此外，还有许多基于不同线性模型的方法[52,59-64]，基因-基因交互作用[50]和基因-环境交互作用[65]的分析方法也不断涌现。

综上所述，家系数据罕见变异关联研究主要需要解决样本间不独立和罕见位点频率低难以发现这两个问题，前者解决方法主要有TDT和引入亲缘关系矩阵，后者主要采用能够综合多个位点效应的统计模型放大效应。与人群研究类似，目前应用较多的方法为负荷检验和方差成分检验。其中，基于SKAT的家系关联分析发展最快，不同研究者提供了多种统计方法，不同方法间的主要区别在于采用了不同的线性模型和方差估计方法，即对遗传结构的假设不同，因此各有其适用的最佳范围。虽然已有的方法众多，但仍存在至今未解决的问题，如二分类表型基因-环境交互作用的检验、小样本数据的检验功效等，仍待进一步探索。此外，关联研究的结果仅能代表位点或区域与疾病间存在的统计学的联系，关联研究结果的合理解释及是否存在生物学关联则需进一步通过实验探索。探索基因的生物学功能，为疾病预防、药物开发等治疗措施研发提供线索或依据仍是亟待解决的问题。

利益冲突 所有作者声明无利益冲突

参 考 文 献

[1] 胡永华. 基于群体的复杂性疾病的病因研究[J]. 北京大学学报：医学版, 2007, 39(2): 113-115. DOI: 10.3321/j.issn:1671-167X.2007.02.001.
Hu YH. Population-based etiology studies on complex diseases[J]. J Peking Univ: Health Sci, 2007, 39(2):

113-115. DOI:10.3321/j.issn:1671-167X.2007.02.001.
[2] MacArthur JAL, Buniello A, Harris LW, et al. Workshop proceedings: GWAS summary statistics standards and sharing[J]. Cell Genomics, 2021, 1(1): 100004. DOI: 10.1016/j.xgen.2021.100004.
[3] Tam V, Patel N, Turcotte M, et al. Benefits and limitations of genome-wide association studies[J]. Nat Rev Genet, 2019, 20(8):467-484. DOI:10.1038/s41576-019-0127-1.
[4] Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases[J]. Nature, 2009, 461(7265):747-753. DOI:10.1038/nature08494.
[5] Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic[J]. Cell, 2017, 169(7):1177-1186. DOI:10.1016/j.cell.2017.05.038.
[6] Speed D, Cai N, Johnson MR, et al. Reevaluation of SNP heritability in complex human traits[J]. Nat Genet, 2017, 49(7):986-992. DOI:10.1038/ng.3865.
[7] Goldstein DB. Common genetic variation and human traits [J]. N Engl J Med, 2009, 360(17):1696-1698. DOI:10.1056/NEJMp0806284.
[8] Lee S, Abecasis GR, Boehnke M, et al. Rare-variant association analysis: study designs and statistical tests[J]. Am J Hum Genet, 2014, 95(1): 5-23. DOI: 10.1016/j.ajhg.2014.06.009.
[9] Gibson G. Rare and common variants: twenty arguments [J]. Nat Rev Genet, 2012, 13(2): 135-145. DOI: 10.1038/nrg3118.
[10] Agarwala V, Flannick J, Sunyaev S, et al. Evaluating empirical bounds on complex disease genetic architecture [J]. Nat Genet, 2013, 45(12): 1418-1427. DOI: 10.1038/ng.2804.
[11] Qi WJ, Allen AS, Li YJ. Family-based association tests for rare variants with censored traits[J]. PLoS One, 2019, 14(1):e0210870. DOI:10.1371/journal.pone.0210870.
[12] Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods[J]. Nat Genet, 2016, 48(10):1284-1287. DOI:10.1038/ng.3656.
[13] The Haplotype Reference Consortium. A reference panel of 64 976 haplotypes for genotype imputation[J]. Nat Genet, 2016, 48(10):1279-1283. DOI:10.1038/ng.3643.
[14] Weissenkampen JD, Jiang Y, Eckert S, et al. Methods for the analysis and interpretation for rare variants associated with complex traits[J]. Curr Protoc Hum Genet, 2019, 101(1):e83. DOI:10.1002/cphg.83.
[15] Momozawa Y, Mizukami K. Unique roles of rare variants in the genetics of complex diseases in humans[J]. J Hum Genet, 2021, 66(1): 11-23. DOI: 10.1038/s10038-020-00845-2.
[16] Wang XJ, Xu XQ, Sun K, et al. Association of rare *PTGIS* variants with susceptibility and pulmonary vascular response in patients with idiopathic pulmonary arterial hypertension[J]. JAMA Cardiol, 2020, 5(6): 677-684. DOI: 10.1001/jamacardio.2020.0479.
[17] Hecker J, Townes FW, Kachroo P, et al. A unifying framework for rare variant association testing in family-based designs, including higher criticism approaches, SKATs, and burden tests[J]. Bioinformatics, 2020, 36(22/23): 5432-5438. DOI: 10.1093/bioinformatics/btaa1055.
[18] Li BS, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data[J]. Am J Hum Genet, 2008, 83(3): 311-321. DOI:10.1016/j.ajhg.2008.06.024.
[19] Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies[J]. Genet Epidemiol, 2010, 34(2): 188-193. DOI: 10.1002/gepi.20450.
[20] Wu MC, Lee S, Cai TX, et al. Rare-variant association testing for sequencing data with the sequence kernel association test[J]. Am J Hum Genet, 2011, 89(1): 82-93. DOI:10.1016/j.ajhg.2011.05.029.
[21] Neale BM, Rivas MA, Voight BF, et al. Testing for an unusual distribution of rare variants[J]. PLoS Genet, 2011, 7(3):e1001322. DOI:10.1371/journal.pgen.1001322.
[22] Lee S, Wu MC, Lin XH. Optimal tests for rare variant effects in sequencing association studies[J]. Biostatistics, 2012, 13(4):762-775. DOI:10.1093/biostatistics/kxs014.
[23] Zöllner S. Sampling strategies for rare variant tests in case-control studies[J]. Eur J Hum Genet, 2012, 20(10): 1085-1091. DOI:10.1038/ejhg.2012.58.
[24] Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies[J]. Hum Mol Genet, 2012, 21(R1): R1-9. DOI:10.1093/hmg/dds387.
[25] Gravel S, Henn BM, Gutenkunst RN, et al. Demographic history and rare allele sharing among human populations [J]. Proc Natl Acad Sci USA, 2011, 108(29): 11983-11988. DOI:10.1073/pnas.1019276108.
[26] Zawistowski M, Reppell M, Wegmann D, et al. Analysis of rare variant population structure in Europeans explains differential stratification of gene-based tests[J]. Eur J Hum Genet, 2014, 22(9):1137-1144. DOI:10.1038/ejhg.2013.297.
[27] Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying general genotype-phenotype associations[J]. Eur J Hum Genet, 2001, 9(4): 301-306. DOI:10.1038/sj.ejhg.5200625.
[28] Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM) [J]. Am J Hum Genet, 1993, 52(3):506-516.
[29] Chen R, Wei Q, Zhan XW, et al. A haplotype-based framework for group-wise transmission/disequilibrium tests for rare variant association analysis[J]. Bioinformatics, 2015, 31(9): 1452-1459. DOI: 10.1093/bioinformatics/btu860.
[30] Chen R, Davis LK, Guter S, et al. Leveraging blood serotonin as an endophenotype to identify de novo and rare variants involved in autism[J]. Mol Autism, 2017, 8: 14. DOI:10.1186/s13229-017-0130-3.
[31] Bickeböller H, Clerget-Darpoux F. Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers[J]. Genet Epidemiol, 1995, 12(6):865-870. DOI:10.1002/gepi.1370120656.
[32] Curtis D. Use of siblings as controls in case-control association studies[J]. Ann Hum Genet, 1997, 61(Pt 4): 319-333. DOI:10.1046/j.1469-1809.1998.6210089.x.
[33] Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test[J]. Am J Hum Genet, 1998, 62(2): 450-458. DOI:10.1086/301714.
[34] Knapp M. The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/disequilibrium test[J]. Am J Hum Genet, 1999, 64(3):861-870. DOI:10.1086/302285.
[35] Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information[J]. Hum Hered, 2000, 50(4): 211-223. DOI: 10.1159/000022918.
[36] de Gourab, Yip WK, Ionita-Laza I, et al. Rare variant analysis for family-based design[J]. PLoS One, 2013, 8(1): e48495. DOI:10.1371/journal.pone.0048495.
[37] Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST) [J]. Mutat Res, 2007, 615(1/2): 28-56. DOI: 10.1016/j. mrfmmm. 2006.

09.003.
[38] Guo W, Shugart YY. Detecting rare variants for quantitative traits using nuclear families[J]. Hum Hered, 2012, 73(3):148-158. DOI:10.1159/000338439.
[39] Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic[J]. PLoS Genet, 2009, 5(2): e1000384. DOI: 10.1371/journal. pgen. 1000384.
[40] Xu X, Rakovski C, Xu XP, et al. An efficient family-based association test using multiple markers[J]. Genet Epidemiol, 2006, 30(7): 620-626. DOI: 10.1002/gepi. 20174.
[41] Hooli BV, Parrado AR, Mullin K, et al. The rare *TREM2* R47H variant exerts only a modest effect on Alzheimer disease risk[J]. Neurology, 2014, 83(15): 1353-1358. DOI: 10.1212/WNL.0000000000000855.
[42] Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium[J]. Genet Epidemiol, 2009, 33(6):497-507. DOI:10.1002/gepi.20402.
[43] Ionita-Laza I, Lee S, Makarov V, et al. Family-based association tests for sequence data, and comparisons with population-based association tests[J]. Eur J Hum Genet, 2013, 21(10): 1158-1162. DOI: 10.1038/ejhg. 2012.308.
[44] Ionita-Laza I, Capanu M, de Rubeis S, et al. Identification of rare causal variants in sequence-based studies: methods and applications to *VPS13B*, a gene involved in Cohen syndrome and autism[J]. PLoS Genet, 2014, 10(12): e1004729. DOI:10.1371/journal.pgen.1004729.
[45] Won S, Wilk JB, Mathias RA, et al. On the analysis of genome-wide association studies in family-based designs: a universal, robust analysis approach and an application to four genome-wide association studies[J]. PLoS Genet, 2009, 5(11):e1000741. DOI:10.1371/journal.pgen.1000741.
[46] Ionita-Laza I, McQueen MB, Laird NM, et al. Genomewide weighted hypothesis testing in family-based association studies, with an application to a 100K scan[J]. Am J Hum Genet, 2007, 81(3):607-614. DOI:10.1086/519748.
[47] Wang XF, Lee S, Zhu XF, et al. GEE-based SNP set association test for continuous and discrete traits in family-based association studies[J]. Genet Epidemiol, 2013, 37(8):778-786. DOI:10.1002/gepi.21763.
[48] Chen H, Meigs JB, Dupuis J. Sequence kernel association test for quantitative traits in family samples[J]. Genet Epidemiol, 2013, 37(2): 196-204. DOI: 10.1002/gepi. 21703.
[49] Haller G, Kapoor M, Budde J, et al. Rare missense variants in *CHRNB3* and *CHRNA3* are associated with risk of alcohol and cocaine dependence[J]. Hum Mol Genet, 2014, 23(3):810-819. DOI:10.1093/hmg/ddt463.
[50] Schifano ED, Epstein MP, Bielak LF, et al. SNP set association analysis for familial data[J]. Genet Epidemiol, 2012, 36(8):797-810. DOI:10.1002/gepi.21676.
[51] Wu MC, Kraft P, Epstein MP, et al. Powerful SNP-set analysis for case-control genome-wide association studies [J]. Am J Hum Genet, 2010, 86(6):929-942. DOI:10.1016/j. ajhg.2010.05.002.
[52] Wang XF, Zhang ZY, Morris N, et al. Rare variant association test in family-based sequencing studies[J]. Brief Bioinform, 2017, 18(6): 954-961. DOI: 10.1093/bib/ bbw083.
[53] Delgado DA, Chernoff M, Huang L, et al. Rare, protein-altering variants in *AS3MT* and arsenic Metabolism efficiency: a multi-population association study[J]. Environ Health Perspect, 2021, 129(4): 047007. DOI:10.1289/EHP8152.
[54] 倪鹏生，崔静，沈福民．传递/不平衡检验 TDT (Transmission/Disequilibrium Test)[J]. 国外医学：遗传学分册, 2000, 23(2):82-86.
Ni PS, Cui J, Shen FM. Transmission/disequilibrium test [J]. For Med:Genet, 2000, 23(2):82-86.
[55] Oualkacha K, Dastani Z, Li R, et al. Adjusted sequence kernel association test for rare variants controlling for cryptic and family relatedness[J]. Genet Epidemiol, 2013, 37(4):366-376. DOI:10.1002/gepi.21725.
[56] Amin N, van Duijn CM, Aulchenko YS. A genomic background based method for association analysis in related individuals[J]. PLoS One, 2007, 2(12): e1274. DOI: 10.1371/journal.pone.0001274.
[57] Choi S, Lee S, Cichon S, et al. FARVAT: a family-based rare variant association test[J]. Bioinformatics, 2014, 30(22): 3197-3205. DOI:10.1093/bioinformatics/btu496.
[58] Jiang D, McPeek MS. Robust rare variant association testing for quantitative traits in samples with related individuals[J]. Genet Epidemiol, 2014, 38(1): 10-20. DOI: 10.1002/gepi.21775.
[59] Jiang YX, Conneely KN, Epstein MP. Flexible and robust methods for rare-variant testing of quantitative traits in trios and nuclear families[J]. Genet Epidemiol, 2014, 38(6):542-551. DOI:10.1002/gepi.21839.
[60] Yan Q, Tiwari HK, Yi NJ, et al. A Sequence kernel association test for dichotomous traits in family samples under a generalized linear mixed model[J]. Hum Hered, 2015, 79(2):60-68. DOI:10.1159/000375409.
[61] Wang XX, Zhao XW, Zhou J. Testing rare variants for hypertension using family-based tests with different weighting schemes[J]. BMC Proc, 2016, 10 Suppl 7: 233-237. DOI:10.1186/s12919-016-0036-7.
[62] Chen H, Huffman JE, Brody JA, et al. Efficient variant set mixed model association tests for continuous and binary traits in large-scale whole-genome sequencing studies[J]. Am J Hum Genet, 2019, 104(2): 260-274. DOI: 10.1016/j. ajhg.2018.12.012.
[63] Park JY, Wu C, Basu S, et al. Adaptive SNP-set association testing in generalized linear mixed models with application to family studies[J]. Behav Genet, 2018, 48(1): 55-66. DOI:10.1007/s10519-017-9883-x.
[64] Jiang Y, Ji YQ, Sibley AB, et al. Leveraging population information in family-based rare variant association analyses of quantitative traits[J]. Genet Epidemiol, 2017, 41(2):98-107. DOI:10.1002/gepi.22022.
[65] Guo CY, Wang RH, Yang HC. Family-based gene-environment interaction using sequence kernel association test (FGE-SKAT) for complex quantitative traits[J]. Sci Rep, 2021, 11(1): 7431. DOI: 10.1038/s41598-021-86871-2.
[66] Jiang Y, Satten GA, Han YJ, et al. Utilizing population controls in rare-variant case-parent association tests[J]. Am J Hum Genet, 2014, 94(6): 845-853. DOI: 10.1016/j. ajhg.2014.04.014.
[67] He ZX, Zhang D, Renton AE, et al. The rare-variant generalized disequilibrium test for association analysis of nuclear and extended pedigrees with application to Alzheimer disease WGS data[J]. Am J Hum Genet, 2017, 100(2):193-204. DOI:10.1016/j.ajhg.2016.12.001.
[68] Sul JH, Cade BE, Cho MH, et al. Increasing generality and power of rare-variant tests by utilizing extended pedigrees[J]. Am J Hum Genet, 2016, 99(4):846-859. DOI: 10.1016/j.ajhg.2016.08.015.
[69] Saad M, Wijsman EM. Combining family- and population-based imputation data for association analysis of rare and common variants in large pedigrees [J]. Genet Epidemiol, 2014, 38(7): 579-590. DOI: 10.1002/ gepi.21844.