

猪重要性状全基因组关联分析的研究进展

赵 谦, 浦亚斌, 关伟军, 赵倩君*, 马月辉*

(中国农业科学院北京畜牧兽医研究所, 北京 100193)

摘 要: 全基因组关联分析(Genome-wide association study, GWAS)是在关联分析的基础上对全基因组范围内的遗传标记进行检测,以研究复杂疾病和性状遗传的一种有效方法。国内外许多研究人员针对猪重要性状展开 GWAS,鉴定出大量单核苷酸多态性(Single nucleotide polymorphism, SNP)标记、数量性状基因座(Quantitative trait locus, QTL)和候选基因为猪育种提供必要的分子基础,同时有助于后续相关功能基因的研究和数量性状核苷酸(Quantitative trait nucleotide, QTN)的鉴定。本文主要对 GWAS 在猪重要性状上的研究进展进行综述,同时简单介绍猪重要性状 QTN 的研究进展。

关键词: 猪;全基因组关联分析(GWAS);单核苷酸多态性(SNP);数量性状基因座(QTL);数量性状核苷酸(QTN)

中图分类号: S813.1; S828

文献标志码: A

文章编号: 0366-6964(2015)06-0873-09

Research Progress of Genome-wide Association Studies for Important Traits in Pig

ZHAO Qian, PU Ya-bin, GUAN Wei-jun, ZHAO Qian-jun*, MA Yue-hui*

(Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China)

Abstract: Genome-wide association study (GWAS) is an effective method to detect complex diseases and traits genetics by the genetic markers in the range of the genome, based on association study. To find new SNPs, QTLs and candidate genes, researchers employ GWAS to study main traits in pigs. It is important to research on molecular genetics of pig breeding and study on related functional genes and QTNs. This review will provide reference for research progress of GWAS about important traits in pig. Furthermore, the paper introduces the research progress of QTN briefly.

Key words: pig; GWAS; SNP; QTL; QTN

中国是猪生产大国,猪肉也是人们日常消费的主要肉类。2013年末和2014年6月我国生猪存栏量分别为4.74亿和4.44亿头。面对如此大的生猪饲养规模,只要对猪某一个重要性状进行改良就可能带来巨大的经济效益。鉴定重要性状的主效基因或遗传标记并应用于育种中,可以加快遗传选育进度,提高经济效益。QTN是QTL中的某个特定核苷酸,这个核苷酸的改变会使个体的表型发生改变。QTN是QTL影响性状的根本原因,是性状出现差

异的分子遗传基础。GWAS的根本目的是寻找影响性状的QTN,它通过对全基因组范围内的遗传标记(CNV或SNP)进行统计分析,找到与目的性状相关的DNA标记或基因组区域,为最终QTN的研究打下基础。

N. Risch等^[1]最早于1996年提出关联分析,预测未来复杂疾病遗传学的研究很可能会通过关联分析进行大规模的检测,证明在研究复杂疾病时关联分析比连锁分析有更强的检测力。在GWAS发展

收稿日期: 2014-08-12

项目资助: 国家自然科学基金项目(31201765); 国家科技基础条件平台家养动物物种质资源共享平台

作者简介: 赵 谦(1990-),男,四川绵阳人,硕士生,主要从事功能基因组学研究, E-mail: zqian90@163.com

* 通信作者: 赵倩君, 副研究员, E-mail: zhaoqianjun@caas.cn; 马月辉, 研究员, E-mail: yuehui.ma@263.net

的初始阶段,其大多被应用于医学上复杂疾病的研究。2005年,《Science》报道了第1例与年龄相关的黄斑变性 GWAS 研究^[2],这是 GWAS 首次发现复杂疾病的分子标记。

2005年家猪基因组序列公布^[3],2012年《Nature》又报道了作为重要肉类来源的杜洛克猪全基因组测序结果^[4]。在此基础上,国内外科研人员利用猪的高密度芯片对猪重要性状进行 GWAS,发现了许多重要的 SNPs、QTLs 和候选基因。本文就 GWAS 主要的设计、分析方法以及在猪重要性状上的研究进展进行综述,并对猪重要性状 QTN 的研究进展进行简单介绍。

1 GWAS 的研究过程

1.1 GWAS 试验样本的选择和质量控制

GWAS 需要常见的遗传变异信息,这些信息需是基于足够大样本中一系列明显且全面的变异^[5]。因此,GWAS 研究应尽可能在大样本群体中进行,这样才能获得足够多的变异信息,而样本群体小会使统计结果可信度降低。在 GWAS 中,首先要对试验样本的性状进行选择。目前已研究过的性状以遗传力较高的数量性状居多。因为低遗传力性状会降低 GWAS 结果的准确度;相对于质量性状,数量性状更易测量,衡量性状变化的精确度高。接着,就要对样本和芯片得到的 SNPs 进行质量控制,样本的质量控制就是芯片对样本的检出率是否符合进一步试验的要求;SNPs 的质量控制主要包括 SNP 分型率、最小等位基因频率(Minor allele frequency, MAF)、Hardy-Weinberger 平衡等。

1.2 GWAS 的试验设计

目前,GWAS 的试验设计可分为单阶段和两阶段或多阶段研究设计。①单阶段研究设计的大体过程:在大样本群中选择足够大的目标性状和对照样本,然后一次性分型出目标性状的 SNPs,再用统计学方法分析 SNPs,确定每个 SNP 与目的性状的关联性。这种试验设计需基于大量的试验样本,主要应用于早期的 GWAS。②两阶段或多阶段研究设计的大体过程:第1阶段,基因分型出较小样本中的所有 SNPs 并筛选出显著的 SNPs;然后在第2阶段或第3阶段的更大样本群中对第1阶段筛选出的显著 SNPs 进行基因分型验证,最后进行综合分析。

1.3 GWAS 的统计分析方法

GWAS 数据的分析方法有两种:基于无关个体

的关联分析和基于家系的关联分析(Family-based association study)。

基于无关个体的关联分析包括病例-对照研究(Case-control study)和基于随机群体的关联分析(Population-based association analysis)。前者先对病例组和对照组全基因组中的 SNPs 进行基因分型,再比较病例组和对照组中每个 SNP 等位基因频率的差异,这种方法主要用于质量性状的研究。后者主要用于无关个体数量性状的研究,可利用协方差或者线性回归方程研究 SNPs 在某个性状的关联性。

基于家系的研究一般采用传递不平衡检验(Transmission disequilibrium test, TDT)来分析遗传分子标记与目的性状的关联,其原理:分析某个等位基因从杂合子的父代传递给具有该性状子代的机率是否高于预期值(50%)。TDT 分析的优势在于可以排除群体混杂对于关联分析的影响,其缺点在于其发现阳性关联的检验效能低于相同样本量的病例-对照研究^[6]。基于家系的关联分析可以避免群体分层对 GWAS 结果的影响,提高分析的可靠性。研究人员应根据所研究的性状、可用样本量大小以及样本组成等具体情况,选择合适的试验设计和分析方法。

2 猪重要性状 GWAS 的研究

随着 2009 年 Illumina Porcine SNP60 芯片的推出,国内外研究人员相继对猪重要性状进行 GWAS,发现了许多显著 SNPs 和 QTLs 与重要性状关联,鉴定出大量相关候选基因。这些工作为 QTN 的鉴定及猪的选择性分子育种提供了重要保证。下文主要介绍近几年 GWAS 在猪肥育、肉质、健康等性状上的研究进展。

2.1 肥育性状 GWAS

肥育性状是中等遗传力性状,也是猪生产中复杂且重要的经济性状,主要衡量指标包括生长速度、饲料利用率等。

剩余采食量(Residual feed intake, RFI)是畜禽实际采食量与预期采食量的差值,也是衡量饲料转化率的指标之一。RFI 又可分为 RFI1 和 RFI2。RFI1 是个体日采食量在初测体重和平均日增重上的回归;RFI2 是 RFI1 除去回归中的背膘部分。丹麦哥本哈根大学的研究团队首先对丹麦杜洛克公猪采食行为进行 GWAS,涉及日采食量、日饲喂总时

间、日饲喂次数等,基因分型后利用 R 语言中的 GenABEL 程序包进行关联分析,发现基因组水平上 92 个 SNPs 与采食行为相关,其中 36 个 SNPs 位于先前报道过的 QTLs 内,鉴定出 14 号染色体(*Sus scrofa* chromosome 14, SSC14)上 *MSI2* 基因与日饲喂次数有很强关联。进一步研究发现突触相关基因、去磷酸化基因和促进多肽分泌的基因与采食行为极显著相关^[7]。而后,该团队又鉴定出 15 和 12 个基因座分别与 RFI1 和 RFI2 显著相关,这些基因座中有 10 个 SNPs 与 RFI1 和 RFI2 都显著相关^[8]。最近,该团队还以约克夏公猪为试验群体,利用单变量模型进行 GWAS,发现 12 和 7 个 SNPs 分别与 RFI1 和 RFI2 显著的关联,鉴定出影响 RFI 的一些候选基因^[9]。S. K. Onteru 等^[10]对 ISU-RFI 猪^[11-12]进行 GWAS,发现参与胰岛素释放的基因与 RFI 和平均日采食量相关;参与能量稳态和肌肉生长的基因与平均日增重相关;参与脂肪代谢的基因与背膘厚相关。最重要的是,发现 SSC7 上一个参与骨骼肌发生的 QTL 与眼肌面积显著相关。在猪生长速度方面,E. J. Jung 等^[13]对纯种长白猪利用混合效应模型和线性回归方法进行 GWAS,发现 SSC16 上 1 个极显著的 SNP 影响长白猪 71 日龄体重。

2.2 肉质性状 GWAS

影响猪肉肉质性状的因素主要有公猪膻味、脂肪含量和猪肉的一些理化性质等。

2.2.1 公猪膻味 未阉割公猪的猪肉中因含较高水平的雄甾烯酮和甲基吲哚会使猪肉具有膻味,影响猪肉口感。国内外研究人员对公猪膻味进行 GWAS,并取得了一定成果。N. Duijvesteijn 等^[14]对杜洛克父系公猪进行 GWAS,找到 SSC1 和 SSC6 上影响杜洛克公猪脂肪组织中雄甾烯酮水平的 37 个显著 SNPs。S. J. Rowe 等^[15]利用 GWAS 对丹麦长白公猪甲基吲哚水平进行研究,发现位于 SSC14 上 *CYP2E1* 基因内的 1 个 SNP 极显著与甲基吲哚水平关联,其中 *CYP2E1* 基因编码一种参与甲基吲哚降解的酶。E. Grindflek 等^[16]对长白和杜洛克公猪膻味成分及其相关的性类固醇进行 GWAS,发现 14 个 QTLs 与长白猪皮下脂肪雄甾烯酮含量显著关联,另外 14 个 QTLs 与杜洛克猪皮下脂肪雄甾烯酮含量显著关联,这些 QTLs 中的 7 个与长白和杜洛克猪都显著关联。

2.2.2 理化性质及肌间脂肪 为了精确定位基

因组中影响肉质性状的 QTLs,江西农业大学的研究团队基于苏太猪和 F₂ 代(杜洛克×二花脸)杂交猪对 pH、颜色、滴水损失等肉质性状进行 GWAS,鉴定出染色体水平显著的 127 个 SNPs 与肉质性状相关,其中的 11 个 SNPs 在全基因组水平显著,还找到一些与肉质性状相关的 QTLs 和候选基因^[17]。通过对猪背最长肌和腹部脂肪的脂肪酸组成进行 GWAS,发现在全基因组水平显著的 15 个 QTLs 与 12 种脂肪酸含量相关,其中最显著的 2 个 SNPs 分别与腹部脂肪 C20:0 脂肪酸含量和肌肉 C18:0 脂肪酸含量相关^[18]。

W. Luo 等^[19]以杂交猪(大白×民猪)为试验群体,涉及肌间脂肪含量、大理石纹、水分含量等,利用混合模型、回归和基因组控制的三步分析方法(GRAMMAR-GC)^[20]进行 GWAS,发现 45 个显著 SNPs 与肉质性状相关,其中的 36 个 SNPs 位于 SSC12 上。T. Lee 等^[21]对杜洛克猪的 14 个肉质性状(包括宰后 24 h pH、肉色、滴水损失等)进行 GWAS,发现 26 个 SNPs 与肉质性状显著相关。S. Ponsuksili 等^[22]基于皮特兰×(德国长白×大白)和纯种德国长白猪,利用 GWAS 发现 51 和 200 个 SNPs 分别与皮特兰×(德国长白×大白)猪和纯种德国长白猪肉质性状显著相关。

2.3 健康性状 GWAS

2.3.1 血液 血液性状在动物免疫力和抗病性方面起着重要作用。中国农业大学的研究团队基于混合模型的单位点回归分析对 18 个血液性状进行 GWAS,经过 permutation 法对多重检验进行校正后,发现 10、24 和 77 个染色体水平上显著的 SNPs 分别与白细胞、红细胞和血小板性状相关^[23]。随后,该团队采用单位点回归模型对猪 T 淋巴细胞性状进行 GWAS,鉴定出染色体水平 61 个显著的 SNPs 和基因组水平显著的 3 个 SNPs 与猪 T 淋巴细胞性状相关,还发现 27 个显著的 SNPs 位于免疫相关的 QTLs 区域内^[24]。该团队还对细胞因子和免疫球蛋白 G 进行了 GWAS,鉴定出染色体水平上 32 个显著的 SNPs,其中 4 个 SNPs 在基因组水平显著^[25]。此外,E. J. Jung 等^[26]在 F₂ 代(长白×韩国本土猪)猪的基础上,对 8 个血液性状进行 GWAS,发现位于 SSC3、6、8、13 和 17 上的 257 个显著 SNPs 与血液性状相关,SSC8 上 17.9~130 Mb 的区域与红细胞平均容积和红细胞平均血红蛋白量显著相关。鉴定出 5 个显著 SNPs 位于造血相

关的基因(*KIT*、*IL15*、*TXK*、*ARAP2* 和 *ERG*)内。

血液中一些生理指标与机体免疫力和疾病有直接关系。江西农业大学的研究团队对 F₂ 代猪(白杜洛克×二花脸)18、46、240 日龄 3 个阶段的 18 个血液生理指标进行单标记 GWAS^[27] 后发现基因组水平上 185 个显著的 SNPs 与血液生理指标相关^[28]。血脂含量已经成为临床上评估患心血管疾病风险的重要指标,且具有较高的遗传力。该团队利用 GWAS 鉴定出影响猪血脂含量的 109 个显著 SNPs^[29]。为了鉴定影响血液性状的 QTL,该团队又对 495 头苏太猪进行单标记 GWAS^[27] 和单体型分析,通过单标记 GWAS 发现 161 个显著 SNPs 与 11 个血液性状相关,其中的 44 个 SNPs 在全基因组水平显著;通过单体型分析发现 499 个显著 SNPs 与 9 个血液性状相关,其中的 154 个 SNPs 在全基因组水平显著^[30]。

血液中红细胞相关参数是衡量机体免疫状态的重要指标,W. Luo 等^[31] 采用 GRAMMAR-GC 法^[20] 对 F₂ 代杂交猪(大白×民猪)的 7 项红细胞性状进行 GWAS,找到 62 个基因组水平和 3 个染色体水平显著相关的 SNPs,其中 7 和 5 个 SNPs 分别与血细胞比容和血红蛋白相关。A. Manunza 等^[32] 通过对杜洛克猪第 45 和 190 天时血液中的胆固醇、甘油三酯、低密度和高密度脂蛋白浓度进行 GWAS,揭示了衰老特异性遗传决定因子的存在。

2.3.2 疾病 GWAS 也能对猪疾病的易感基因区域进行定位。利用 GWAS 对母猪产后泌乳障碍综合征(Postpartum dysgalactia syndrome, PDS) 进行研究,在 SSC17 上发现 1 个显著的 SNP 影响 PDS^[33]。N. Rangkasenee 等^[34] 对大白猪软骨病进行 GWAS,找到位于 SSC3、5、8、10、14 和 18 上 19 个显著 SNPs,并鉴定出与软骨病发生相关的候选基因 *TBX5*。拥有 XX 染色体的猪有出现性反转的现象,表现中性或雄性的行为,并常伴有不育、泌尿生殖器感染等疾病。S. Rousseau 等^[35] 对性反转进行 GWAS,发现 SSC12 上一个包含 *SOX9* 基因(参与睾丸分化)的区域极显著与性反转表型关联。W. X. Fu 等^[36] 利用 ROADTRIPS^[37] 软件对仔猪的大肠杆菌 F4ab/F4ac 菌株小肠黏膜体外黏附表型进行 GWAS,检测到 28 和 18 个显著 SNPs 分别与易感 F4ab 和 F4ac 菌株表型关联,并发现 *HEG1* 和 *ITGB5* 两个潜在的候选基因。J. Kogelman 等^[38] 对 538 头猪进行 GWAS 后鉴定出基因组水平显著

的 366 个 SNPs 与肥胖关联。

2.4 繁殖性状 GWAS

S. K. Onteru 等^[39-40] 对母猪终生繁殖力和母猪 1~3 胎繁殖性能进行 GWAS,检测出 14、14、19 和 12 个显著 SNPs 分别与终生总产仔数、终生产活仔数、淘汰胎次和非繁殖天数比率相关;对于母猪 1~3 胎的总产仔数、产活仔数、死胎等 15 个繁殖性状的研究,共检测到 337 个 QTLs 和一些候选基因。J. F. Schneider 等^[41] 对初产母猪的产仔总数、产活仔数、死胎数等 7 个繁殖性状进行 GWAS,发现 124 个显著的 QTLs,其中 11、14、1、33 和 65 个 QTLs 分别与产仔总数、产活仔数、死胎数、窝总仔重和窝平均仔重显著关联。M. S. Lopes 等^[42] 利用 GWAS 对猪乳头数量的加性效应和显性效应进行研究,发现 21 个 SNPs 与乳头数量关联,其中 SSC6、7 和 12 上的 SNPs 与乳头数量的加性效应关联,SSC4 上的 SNPs 与乳头数量的显性效应关联。

2.5 体型、毛色性状 GWAS

L. G. Wang 等^[43] 针对体高、体长、管围等体型性状,对 F₂ 代杂交猪(大白猪×民猪)进行 GWAS,找到 SSC7 上 138 个 SNPs 与体高、体长、管围和臀围关联,SSC1 上 1 个基因组水平显著 SNP 与胸深关联。J. Ren 等^[44] 对藏猪、大河等中国地方猪进行 GWAS,发现了一些和猪棕色毛色显著相关的 SNP,还发现棕色毛色变异的原因是 *TYRP1* 基因中 6 个碱基的缺失。I. C. Cho 等^[45] 利用 GWAS 对杂交猪(长白×朝鲜本地猪)的杂色毛色(白毛和有色毛混杂)进行研究,在 SSC8 上检测到和杂色毛色高度关联或连锁的基因组区域。

2.6 胴体性状 GWAS

胴体性状是一种高遗传力性状,主要衡量指标包括:屠宰率、背膘厚、眼肌面积和瘦肉率等。意大利博洛尼亚大学的科研人员 L. Fontanesi 等^[46] 对意大利大白猪背膘厚进行 GWAS,鉴定出 119 个潜在 SNPs 和 4 个显著 SNPs 与意大利大白猪背膘厚相关。GO 分析结果显示这些 SNPs 涉及的大部分基因参与神经系统的发育和调控,这与人类肥胖 GWAS 研究结果一致。B. Fan 等^[47] 对母猪背膘厚、眼肌面积、体型和四肢的稳健度进行 GWAS,鉴定出许多候选染色体区域,并在这些区域内发现一些新的候选基因。N. Okumura 等^[48] 对背膘厚和管围进行 GWAS,发现背膘厚与 SSC6 上的 3 个显著 SNPs 关联,管围与 SSC7 上的 45 个显著 SNPs 关

联。

2.7 多个性状 GWAS

GWAS 不仅能针对单个性状进行研究,还可以对同一试验样本群的多个性状进行研究。张哲等^[49]对体重达 100 kg 日龄(D100)、活体背膘厚和活体眼肌面积 3 个性状的表型分别进行 GWAS,在 D100 和活体眼肌面积 2 个性状中分别检测到 1 个基因组水平和 6 个染色体水平显著关联的 SNPs;没有检测到与背膘厚显著相关的 SNP。生物信息学分析表明,*BTG1* 和 *EFCAB6* 可能是影响生长性状的重要候选基因。D. Becker 等^[50]对瑞士大白公猪进行 GWAS,发现可能影响猪的外观、肉质、繁殖力和生长性状的 4 个 QTLs。M. P. Sanchez 等^[51]对大白猪的 19 个性状进行 GWAS,鉴定出 4、12 和 7 个 QTLs 分别影响饲料利用率、胴体和肉质性状。

3 猪重要性状 QTN 研究进展

GWAS 能鉴定出与性状关联的基因组区域和 DNA 标记,但其根本目的是寻找影响性状的 QTN,从而揭示引起性状差异的根本原因。在早期 QTN 的研究中,科研人员已鉴定出 3 个与猪重要性状相关的 QTN:兰尼定 1 型受体基因(Ryanodine receptor 1,*RYR1*)^[52]、胰岛素样生长因子 2(Insulin-like growth factor 2,*IGF2*)基因^[53]和黑皮质素受体 4(Melanocortin receptor 4,*MC4R*)基因^[54]。最近,在 P. Stratz 等^[55]进行的 GWAS 中验证了 *RYR1* 基因中的 QTN 与肉质性状相关,同时还发现位于 SSC6、SSC10 和 SSC15 上的 SNPs 与日增重、胴体瘦肉含量和肌间脂肪含量等相关。

近几年,猪重要性状 GWAS 不断开展的同时,关于猪重要性状 QTN 的研究也在陆续进行,并鉴定出 2 个新的 QTNs。*PHKG1* 基因参与编码磷酸化酶激酶的催化亚单位,影响猪肉质性状。J. Ma 等^[56]首先利用 GWAS 将糖酵解潜能主效 QTL 定位于 SSC3 上一段区域内,随后发现,该 QTL 内的 *PHKG1* 基因中内含子 9 的拼接受体位点发生 C/A 点突变,造成开放阅读框中 32 bp 碱基的缺失并且产生一个终止密码子,导致蛋白水平的表达降低和酶活性的减弱。在不同的猪群体中,突变体表现出糖原含量高和肉质差,证实 *PHKG1* 的 C/A 突变是影响猪肉质性状的 QTN。在此之前,J. Ren 等^[57]还将影响猪耳大小的主效 QTL 定位于 SSC7 上一段区域内,进一步研究发现该 QTL 内的 *PPARD*

基因中一个错意突变 G/A 导致编码的蛋白序列中甘氨酸变为谷氨酸,使得 *PPARD* 的功能受到影响并导致猪耳大小性状的差异。这一研究解释了主效 QTL 影响猪耳大小的根本原因,找到影响猪耳大小的 QTN^[57]。上述研究表明,可先利用 GWAS 定位候选 QTL,再辅以其他研究从而精确定位 QTN,最终阐明性状变异的根本原因。

此外,研究人员还对其它与猪重要性状相关的基因进行了研究,找到相关潜在的 QTN。*AN-GPTL4* 基因编码一种在肝和白色脂肪组织中分泌的蛋白,该蛋白抑制脂蛋白脂肪酶的活性并促进白色脂肪组织的分解。Z. Q. Ren 等^[58]发现,*AN-GPTL4* 基因第 3 个内含子内 G/A 的突变与猪肉质性状显著相关,推测该位点可能是影响猪肉质的 QTN。*BMPER* 基因是血管再生和血管发生过程的一个重要基因。Z. Liu 等^[59]发现,*BMPER* 基因启动子内 A/C 突变改变了启动子与 GATA 的亲和力,从而影响 *BMPER* 基因的转录。这一突变与背最长肌肌间脂肪含量相关,推测该位点可能是影响猪肌间脂肪含量的 QTN。*LEF-1* 基因在乳房内皮细胞的发育过程中起着重要作用。R. X. Xu 等^[60]报道,*LEF-1* 基因内 A/G 和 C/T 两个点突变与猪乳头数量显著相关,推测这两个位点可能是影响猪乳头数量的 QTN。

4 GWAS 存在的主要问题

4.1 群体分层

群体分层指由于祖先或遗传血统的不同导致试验组与对照组中等位基因频率存在差异。群体分层是导致病例-对照设计下的关联研究中出现假阳性结果的一种最重要的混杂因素^[61]。在进行 GWAS 时,为了降低假阳性位点的概率,就要对样本进行群体分层检验。Q-Q plot 诊断图是最常用的群体分层检测方法,从图中可以判断样本群体是否有群体分层现象。目前,处理群体分层的主要方法:基因组控制法(Genomic control,GC),结构关联法(Structured association,SA)和主成分分析法(Principal components analysis,PCA)等。

4.2 多重检验

多重假设检验是 GWAS 需要面对和处理的一个重要问题。由于 GWAS 通常是对基因组上的大量 SNPs 同时进行假设检验,即对每一个 SNP 位点与性状的关联性都进行一次假设检验,若以普通的

假设检验来判断 SNP 与性状的关联性将会带来许多假阳性的关联结果。因此需要对多重假设检验进行校正,这是得到可信 GWAS 结果的必要前提。常用的方法:①Bonferroni 校正法,它是目前常用的一种多重检验校正方法。但它没有考虑到 SNP 位点之间的关联性,被认为是多重比较 P 值校正中最为保守的一种,过分严苛,会造成假阴性升高^[62]。②控制错误发现率(False discovery rate, FDR)法,它是现有的校正方法中最为宽松的一种,允许更多的假阳性存在,但同时也降低了假阴性。③Permutation 法,其统计效力较高,对于小样本数据,也可有效降低伪关联结果,但是该法计算量大,耗时长^[63]。

5 GWAS 的注意事项

要对某一性状成功的进行 GWAS 分析,需要考虑多方面的因素。首先应该注意的是性状的选择,应尽量选择遗传力较高的数量性状。其次,应尽可能在大样本中对性状进行研究。因为不管是对于单阶段还是两阶段或多阶段研究设计,在尽可能大的样本中得到的 GWAS 结果会更加可信。最后,根据具体试验情况采用最佳的分析方法,对群体分层进行处理以及多重检验校正。GWAS 群体分层和多重检验导致的假阳性或假阴性问题仍然难以解决。GWAS 不能仅凭 P 值判断某个 SNP 是否与性状或疾病真正关联,研究人员要想发现真正与目的性状关联的变异位点,就必须要多群体、大样本的重复验证研究^[64]。

6 展望

目前,研究人员已经对人和动物的许多复杂性状和疾病进行了 GWAS,发现了大量关联的分子标记(主要是 SNP),筛选了许多候选基因,鉴定了一些重要的 QTLs。但 GWAS 并没有确定真正的功能基因,通过 GWAS 发现的很多显著 SNPs 并不能确定是否影响蛋白质的形成,只能作出推测,很难进行精确的致因变异研究。GWAS 的根本目的是寻找影响性状的 QTN。因此,为了鉴定出影响性状的 QTN,还需要对 GWAS 得到的 DNA 标记和 QTLs 进行进一步的研究,以阐明性状变异的根本原因。另外,通过 GWAS 得到相关变异数据后,也可对相关性状进行深度挖掘,例如进行多种群大样本量验证、易感位点的精细定位和基于通路的 GWAS

等^[65]。

参考文献(References):

- [1] RISCH N, MERIKANGAS K. The future of genetic studies of complex human diseases[J]. *Science*, 1996, 273(5281):1516-1517.
- [2] KLEIN R J, ZEISS C, CHEW E Y, et al. Complement factor H polymorphism in age-related macular degeneration[J]. *Science*, 2005, 308(5720):385-389.
- [3] WERNERSSON R, SCHIERUP M H, JØRGENSEN F G, et al. Pigs in sequence space: a 0.66X coverage pig genome survey based on shotgun sequencing[J]. *BMC Genomics*, 2005, 6:70.
- [4] GROENEN M A, ARCHIBALD A L, UENISHI H, et al. Analyses of pig genomes provide insight into porcine demography and evolution[J]. *Nature*, 2012, 491(7424):393-398.
- [5] HIRSCHHORN J N, DALY M J. Genome-wide association studies for common diseases and complex traits[J]. *Nat Rev Genet*, 2005, 6(2):95-108.
- [6] 严卫丽. 复杂疾病全基因组关联研究进展—遗传统计分析[J]. *遗传*, 2008, 30(5):543-549.
YAN W L. Genome-wide association study on complex diseases: genetic statistical issues[J]. *Hereditas*, 2008, 30(5):543-549. (in Chinese)
- [7] DO D N, STRATHE A B, OSTERSEN T, et al. Genome-wide association study reveals genetic architecture of eating behavior in pigs and its implications for humans obesity by comparative mapping[J]. *PLoS ONE*, 2013, 8(8):e71509.
- [8] DO D N, OSTERSEN T, STRATHE A B, et al. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs[J]. *BMC Genet*, 2014, 15(1):27.
- [9] DO D N, STRATHE A B, OSTERSEN T, et al. Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake[J]. *Front Genet*, 2014, 5:307.
- [10] ONTERU S K, GORBACH D M, YOUNG J M, et al. Whole genome association studies of residual feed intake and related traits in the pig[J]. *PLoS ONE*, 2013, 8(6):e61756.
- [11] CAI W, CASEY D S, DEKKERS J C M, et al. Selection response and genetic parameters for residual feed intake in Yorkshire swine[J]. *J Anim Sci*, 2008, 86(2):287-298.

- [12] YOUNG J M, CAI W, DEKKERS J C M, et al. Effect of selection for residual feed intake on feeding behavior and daily feed intake patterns in Yorkshire swine [J]. *J Anim Sci*, 2011, 89(3): 639-647.
- [13] JUNG E J, PARK H B, LEE J B, et al. Genome-wide association analysis identifies quantitative trait loci for growth in a Landrace purebred population [J]. *Anim Genet*, 2014, 45(3): 442-444.
- [14] DUIJVESTEIJN N, KNOL E F, MERKS J W M, et al. A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6 [J]. *BMC Genet*, 2010, 11(1): 42.
- [15] ROWE S J, KARACAÖREN B, KONINGD D, et al. Analysis of the genetics of boar taint reveals both single SNPs and regional effects [J]. *BMC Genomics*, 2014, 15: 424.
- [16] GRINDFLEK E, LIEN S, HAMLAND H, et al. Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids [J]. *BMC Genomics*, 2011, 12: 362.
- [17] MA J, YANG J, ZHOU L, et al. Genome-wide association study of meat quality traits in a White Duroc × Erhualian F₂ intercross and Chinese Sutai pigs [J]. *PLoS ONE*, 2013, 8(5): e64047.
- [18] YANG B, ZHANG W, ZHANG Z, et al. Genome-wide association analyses for fatty acid composition in porcine muscle and abdominal fat tissues [J]. *PLoS ONE*, 2013, 8(6): e65554.
- [19] LUO W, CHENG D, CHEN S, et al. Genome-wide association analysis of meat quality traits in a porcine Large White × Minzhu intercross population [J]. *Int J Biol Sci*, 2012, 8(4): 580-595.
- [20] AMIN N, DUIJN C M, AULCHENKO Y S, et al. A genomic background based method for association analysis in related individuals [J]. *PLoS ONE*, 2007, 2(12): e1274.
- [21] LEE T, SHIN D, CHO S, et al. Genome-wide association study of integrated meat quality-related traits of the Duroc pig breed [J]. *Asian-Austra J Anim Sci*, 2014, 27(3): 303-309.
- [22] PONSUKSILI S, MURANI E, TRAKOOLJUL N, et al. Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis [J]. *Int J Biol Sci*, 2014, 10(3): 327-337.
- [23] WANG J Y, LUO Y R, FU W X, et al. Genome-wide association studies for hematological traits in swine [J]. *Anim Genet*, 2013, 44(1): 34-43.
- [24] LU X, FU W X, LUO Y R, et al. Genome-wide association study for T lymphocyte subpopulations in swine [J]. *BMC Genomics*, 2012, 13(1): 488.
- [25] LU X, LIU J F, FU W X, et al. Genome-wide association study for Cytokines and immunoglobulin G in swine [J]. *PLoS ONE*, 2013, 8(10): e74846.
- [26] JUNG E J, PARK H B, LEE J B, et al. Genome-wide association study identifies quantitative trait loci affecting hematological traits in an F₂ intercross between Landrace and Korean native pigs [J]. *Anim Genet*, 2014, 45(4): 534-541.
- [27] FURLOTTE N A, ESKIN E, EYHERAMENDY S. Genome-wide association mapping with longitudinal data [J]. *Genet Epidemiol*, 2012, 36(5): 463-471.
- [28] ZHANG Z, HONG Y, GAO J, et al. Genome-wide association study reveals constant and specific loci for hematological traits at three time stages in a White Duroc × Erhualian F₂ resource population [J]. *PLoS ONE*, 2013, 8(5): e63665.
- [29] CHEN C, YANG B, ZENG Z, et al. Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model [J]. *BMC Genomics*, 2013, 14(1): 848.
- [30] ZHANG F, ZHANG Z Y, YANX M, et al. Genome-wide association studies for hematological traits in Chinese Sutai pigs [J]. *BMC Genet*, 2014, 15: 41.
- [31] LUO W, CHEN S, CHENG D, et al. Genome-wide association study of porcine hematological parameters in a large White × Minzhu F₂ resource population [J]. *Int J Biol Sci*, 2012, 8(6): 870-881.
- [32] MANUNZA A, CASELLAS J, QUINTANILLA R, et al. A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants [J]. *BMC Genomics*, 2014, 15: 758.
- [33] PREISSLER R, TETENS J, REINERS K, et al. A genome-wide association study to detect genetic variation for postpartum dysgalactia syndrome in five commercial pig breeding lines [J]. *Anim Genet*, 2013, 44(5): 502-508.
- [34] RANGKASENEE N, MURANI E, BRUNNER R M, et al. Genome-wide association identifies *TBX5* as candidate gene for osteochondrosis providing a functional link to cartilage perfusion as initial factor [J]. *Front Genet*, 2013, 4: 78.
- [35] ROUSSEAU S, IANNUCELLI N, MERCAT M, et al. A genome-wide Association study points out the

- causal implication of *SOX9* in the sex-reversal phenotype in XX pigs[J]. *PLoS ONE*, 2013, 8(11): e79882.
- [36] FU W X, LIU Y, LU X, et al. A genome-wide association study identifies two novel promising candidate genes affecting *Escherichia coli* F4ab/F4ac susceptibility in swine[J]. *PLoS ONE*, 2012, 7(3): e32127.
- [37] THORNTON T, MCPHEEK M S. ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure[J]. *Am J Hum Genet*, 2010, 86(2): 172-184.
- [38] KOGELMAN L J, PANT S D, FREDHOLM M M, et al. Systems genetics of obesity in an F2 pig model by genome-wide association, genetic network, and pathway analyses[J]. *Front Genet*, 2014, 5: 214.
- [39] ONTERU S K, FAN B, NIKKILA M T, et al. Whole-genome association analyses for lifetime reproductive traits in the pig[J]. *J Anim Sci*, 2011, 89(4): 988-995.
- [40] ONTERU S K, FAN B, DU Z Q, et al. A whole-genome association study for pig reproductive traits[J]. *Anim Genet*, 2012, 43(1): 18-26.
- [41] SCHNEIDER J F, REMPEL L A, SNELLING W M, et al. Genome-wide association study of swine farrowing traits. Part II: Bayesian analysis of marker data [J]. *J Anim Sci*, 2012, 90(10): 3360-3367.
- [42] LOPES M S, BASTIAANSEN J, HARLIZIUS B, et al. A genome-wide association study reveals dominance effects on number of teats in pigs[J]. *PLoS ONE*, 2014, 9(8): e105867.
- [43] WANG L G, ZHANG L C, YAN H, et al. Genome-wide association studies identify the loci for 5 exterior traits in a Large White \times Minzhu pig population[J]. *PLoS ONE*, 2014, 9(8): e103766.
- [44] REN J, MAO H, ZHANG Z, et al. A 6-bp deletion in the *TYRP1* gene causes the brown colouration phenotype in Chinese indigenous pigs[J]. *Heredity*, 2011, 106(5): 862-868.
- [45] CHO I C, ZHONG T, SEO B Y, et al. Whole-genome association study for the roan coat color in an intercrossed pig population between Landrace and Korean native pig[J]. *Genes Genom*, 2011, 33(1): 17-23.
- [46] FONTANESI L, SCHIAVO G, GALIMBERTI G, et al. A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes [J]. *BMC Genomics*, 2012, 13: 583.
- [47] FAN B, ONTERU S K, DU Z Q, et al. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs [J]. *PLoS ONE*, 2011, 6(2): e14726.
- [48] OKUMURA N, MATSUMOTO T, HAYASHI T, et al. Genomic regions affecting backfat thickness and cannon bone circumference identified by genome-wide association study in a Duroc pig population[J]. *Anim Genet*, 2013, 44(4): 454-457.
- [49] 张 哲, 贺金龙, 邓 熙, 等. 杜洛克猪生长性状全基因组关联分析[J]. 广东农业科学, 2014, 14: 139-143. ZHANG Z, HE J L, DENG X, et al. Genome-wide association study for production traits in Duroc pig herd [J]. *Guangdong Agricultural Sciences*, 2014, 14: 139-143. (in Chinese)
- [50] BECKER D, WIMMERS K, LUTHER H, et al. A genome-wide association study to detect QTL for commercially important traits in Swiss Large White boars [J]. *PLoS ONE*, 2013, 8(2): e55951.
- [51] SANCHEZ M P, TRIBOUT T, IANNUCELLI N, et al. A genome-wide association study of production traits in a commercial population of Large White pigs: evidence of haplotypes affecting meat quality [J]. *Genet Sel Evol*, 2014, 46(1): 12.
- [52] FUJII J, OTSU K, ZORZATO F et al. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia [J]. *Science*, 1991, 253(5018): 448-451.
- [53] VANLAERE A S, NGUYEN M, BRAUNSCHWEIG M, et al. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig [J]. *Nature*, 2003, 425(6960): 832-836.
- [54] KIM K S, LARSEN N, SHORT T, et al. A missense variant of the porcine melanocortin-4 receptor (*MC4R*) gene is associated with fatness, growth, and feed intake traits [J]. *Mamm Genome*, 2000, 11(2): 131-135.
- [55] STRATZ P, WELLMAN N R, PREUS S S, et al. Genome-wide association analysis for growth, muscularity and meat quality in Pietrain pigs [J]. *Anim Genet*, 2014, 45(3): 350-356.
- [56] MA J, YANG J, ZHOU L, et al. A splice mutation in the *PHKG1* gene causes high glycogen content and low meat quality in pig skeletal muscle [J]. *PLoS Genet*, 2014, 10(10): e1004710.
- [57] REN J, DUAN Y, QIAO R, et al. A missense mutation in *PPARD* causes a major QTL effect on ear size in pigs [J]. *PLoS Genet*, 2011, 7(5): e1002043.

- [58] REN Z Q, WU W J, LIU W H, et al. Differential expression and effect of the porcine ANGPTL4 gene on intramuscular fat[J]. *Genet Mol Res*, 2014, 13(2): 2949-2958.
- [59] LIU Z, SUN W, ZHAO Y, et al. The effect of variants in the promoter of BMPER on the intramuscular fat deposition in *Longissimus dorsi* muscle of pigs[J]. *Gene*, 2014, 542(2): 168-172.
- [60] XU R X, WEI N, WANG Y, et al. Association of novel polymorphisms in lymphoid enhancer binding factor 1 (LEF-1) gene with number of teats in different breeds of pig[J]. *Asian-Austra J Anim Sci*, 2014, 27(9): 1254-1262.
- [61] 潘东东, 李正帮, 张 维, 等. 全基因组关联研究综述[J]. *应用概率统计*, 2014, 30(1): 84-103.
PAN D D, LI Z B, ZHANG W, et al. The review of genome-wide association studies[J]. *Chinese Journal of Applied Probability and Statistics*, 2014, 30(1): 84-103. (in Chinese)
- [62] 王继英, 王海霞, 迟瑞宾, 等. 全基因组关联分析在畜禽中的研究进展[J]. *中国农业科学*, 2013, 46(4): 819-829.
- WANG J Y, WANG H X, CHI R B, et al. Progresses in research of genome-wide association studies in livestock and poultry[J]. *Scientia Agricultura Sinica*, 2013, 46(4): 819-829. (in Chinese)
- [63] NICODEMUS K K, LIU W, CHASE G A, et al. Comparison of type I error for multiple test corrections in large single-nucleotide polymorphism studies using principal components versus haplotype blocking algorithms[J]. *BMC Genet*, 2005, 6(Suppl 1): S78.
- [64] 涂 欣, 石立松, 汪 樊, 等. 全基因组关联分析的进展与反思[J]. *生理科学进展*, 2010, 41(2): 87-94.
TU X, SHI L S, WANG F, et al. Genome-wide association study: advances, challenges and deliberation [J]. *Progress in Physiological Sciences*, 2010, 41(2): 87-94. (in Chinese)
- [65] 权 晟, 张学军. 全基因组关联研究的深度分析策略[J]. *遗传*, 2011, 33(2): 100-108.
QUAN C, ZHANG X J. Research strategies for the next step of genome-wide association study[J]. *Hereditas*, 2011, 33(2): 100-108. (in Chinese)

(编辑 郭云雁)

本刊相关报道(以下文章由本刊网站 www.xmsyxb.com 可以免费下载查阅):

1. 兰 蓉, 朱 兰, 姚新荣, 等. 山羊产羔数全基因组关联分析[J]. *畜牧兽医学报*, 2015, 46(4): 549-554.
LAN R, ZHU L, YAO X R, et al. Genome-wide association study of lambing number in goat[J]. *Acta Veterinaria et Zootechnica Sinica*, 2015, 46(4): 549-554.
2. 刘澳星, 郭 刚, 王雅春, 等. 中国荷斯坦牛初产日龄遗传评估及全基因组关联分析[J]. *畜牧兽医学报*, 2015, 46(3): 373-381.
LIU A X, GUO G, WANG Y C, et al. Genetic analysis and genome wide association studies for age at first calving in Chinese Holsteins[J]. *Acta Veterinaria et Zootechnica Sinica*, 2015, 46(3): 373-381.
3. 樊庆灿, 王金玉, 张跟喜, 等. 运用四种线性模型对京海黄鸡上市体重进行全基因组关联分析[J]. *畜牧兽医学报*, 2014, 45(7): 1053-1059.
FAN Q C, WANG J Y, ZHANG G X, et al. A Genome-wide association study of market weight using four statistic models in Jinghai Yellow chicken[J]. *Acta Veterinaria et Zootechnica Sinica*, 2014, 45(7): 1053-1059.
4. 朱 波, 吴 洋, 齐 欣, 等. 利用混合线性模型和 BayesCPi 进行 QTL-MAS2011 公共数据集的全基因组关联分析[J]. *畜牧兽医学报*, 2014, 45(5): 692-698.
ZHU B, WU Y, QI X, et al. Genome-wide association analysis of the 15th QTL-MAS workshop data using mixed linear model and BayesCPi method[J]. *Acta Veterinaria et Zootechnica Sinica*, 2014, 45(5): 692-698.