

DOI: 10.3724/SP.J.1005.2012.01261

利用紧缩线性模型和贝叶斯模型对猪总产仔数和产活仔数性状的全基因组关联研究

刘小磊¹, 杨松柏¹, Max F Rothschild², ZHANG Zhi-Wu³, 樊斌¹

1. 华中农业大学, 农业动物遗传育种与繁殖教育部重点实验室, 武汉 430070;
2. Department of Animal Science, Iowa State University, Ames IA 50011;
3. Institute for Genomic Diversity, Cornell University, Ithaca NY 14853

摘要: 全基因组关联分析策略已逐渐成为家畜重要经济性状研究的强有力工具。文章使用猪 60K SNP 芯片对一个具多胎繁殖性状记录的商业母猪群(n=820)进行分型检测, 共计 57 814 个 SNP 通过设定质控标准。主成分分析显示群体内不存在显著的群体分层现象, 而后分别运用两种统计模型 Compressed Mixed Linear Model (GAPIT 程序包)、Bayes CPi(GenSel 软件)进行第 1 和第 2 胎次总产仔数和产活仔数性状的全基因组关联分析。从两种分析方法所得结果中各取最显著的 50 个 SNP 位点进行比较: 对于第 1 胎次总产仔数, 两种方法分析结果存在 31 个重合 SNP 位点, 对于第 1 胎次产活仔数, 有 20 个重合 SNP 位点; 且两种统计分析结果中最显著的 SNP 位点都在另一方法中得到验证。与第 1 胎次总产仔数显著关联的 SNP 位于 1、2、3、7、13、16 和 18 号染色体, 与第 1 胎次产活仔数显著关联的 SNP 位于 1、3、4、13 和 16 号染色体上的 11 个区域内。在 1、3、13 和 16 号染色体上共有 5 个区域同时与这两个性状显著关联。与第 2 胎次总产仔数和产活仔数显著关联的区域主要位于 7、10、12、13、14 和 16 号染色体的 6 个重叠区域内。

关键词: 全基因组关联; 总产仔数; 产活仔数; SNP 芯片; 猪

Genome-wide association study of total number born and number born alive in pigs using both compressed mixed linear model and Bayes model

LIU Xiao-Lei¹, YANG Song-Bai¹, Max F Rothschild², ZHANG Zhi-Wu³, FAN Bin¹

1. Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education, Wuhan 430070, China;
2. Department of Animal Science, Iowa State University, Ames IA 50011, USA;
3. Institute for Genomic Diversity, Cornell University, Ithaca NY 14853, USA

Abstract: GWAS (Genome-wide association study) strategy has been extensively used for identification of economical

收稿日期: 2012-06-18; 修回日期: 2012-07-27

基金项目: 国家自然科学基金项目(编号: 31072009), 教育部新世纪人才支持计划项目(编号: NCET-11-0646)和中央高校基本科研业务费专项资金(编号: 2010PY008)资助

作者简介: 刘小磊, 硕士研究生, 专业方向: 动物遗传育种。E-mail: xll198708@gmail.com

通讯作者: 樊斌, 教授, 研究方向: 分子生物学与动物育种。E-mail: fanbin@mail.hzau.edu.cn

网络出版时间: 2012-9-10 03:17:05

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20120910.1517.007.html>

trait loci in livestock animals. Using Illumina's PorcineSNP60 BeadChip, a GWA study of 820 commercial pigs with reproductive traits recorded was performed. The PCA analysis showed that there was no significant population stratification. Two different statistical models Compressed Mixed Linear Model(GAPIT program package)and Bayes CPi (GenSel software) were used to implement GWAS on total number born and number born alive of the first and second parity. To compare the most significant 50 SNPs from each method, a total of 31 and 20 coincided SNPs for total number born in the first parity were identified, and there were 20 coincided SNPs for number born alive in the first parity in the results of both methods. The most significant SNPs were also significant in the results of the other method. The most significantly associated regions for total number born in the first parity were located on SSC1, 2, 3, 7, 13, 16, and 18. The most significantly associated regions for number born alive in the first parity were located on SSC1, 3, 4, 13, and 16. There were 5 common regions significantly associated with both traits on SSC1, 3, 13, and 16. The most significantly associated regions for both total number born and number born alive for the second parity were mainly located on six common regions on SSC7, 10, 12, 13, 14, and 16.

Keywords: GWAS; total number born; number born alive; SNP chip; pig

猪的繁殖力是最为重要经济性状之一，提高猪繁殖力性状能够显著提高商业猪场的经济效益。产仔数和产活仔数是两个最为主要的繁殖性状，但由于其属于低遗传力性状，运用常规育种方法很难提高猪的产仔数目。寻找控制猪产仔数和产活仔数的主效基因和遗传标记，用于分子标记辅助育种具有重要意义。传统候选基因和数量性状定位(QTL)取得的成果主要有鉴定出雌激素受体基因(Estrogen receptor, *ESR*)^[1]、促卵泡素β亚基(Follicle stimulating hormone β subunit, *FSHβ*)基因^[2,3]等。

随着家畜全基因组测序项目完成及全基因组SNP芯片成功研制，全基因组关联研究(Genome-wide association study, GWAS)在经济性状候选基因或基因组区域识别中得到广泛应用。全基因组关联研究主要利用分布于整个基因组中的SNP，基于它们与分析性状的连锁不平衡(Linkage disequilibrium, LD)关系，通过各种统计分析方法，以获得与这些性状关联的候选基因或基因组区域。相对于传统候选基因定位方法及QTL定位方法，全基因组关联分析在精确度和新基因识别方面具有明显优势。Fan等^[4]采用猪全基因组SNP芯片对背膘厚度和眼肌面积进行GWA分析，获得一批具有明显效应的候选基因及基因组区域，包括*MC4R*、*IGF2*等已识别的候选基因以及一些在猪中功能不明确的新基因。Gorbach等^[5]采用GWA策略对残留采食量进行分析，得到一些与该重要性状显著关联的候选基因。Onteru等^[6,7]

利用GWAS策略对母猪繁殖性状和利用年限性状等进行分析，获得候选基因组区域，发现不同胎次之间与产仔数等繁殖性状显著关联的区域重叠较少，提示不同胎次繁殖性状的遗传机理有可能不同。目前用于GWAS的统计分析方法尚无统一，尤其对于低遗传力性状，单一分析方法可能有失偏颇。本研究拟尝试两种不同的全基因组关联分析方法，对猪第1胎次和第2胎次的产仔数和产活仔数性状进行分析，比较不同方法所得结果，以获得一批与上述繁殖性状相关的SNP标记或候选基因。

1 材料和方法

1.1 材料

本研究所用820头母猪资源群体包括408头大白纯种猪和412头大白×长白二元母猪，来自美国Newshaw Choice Genetics公司^[7]。为便于大规模群体比较分析，本研究主要选取性状记录较为完整的第1胎次($n = 687$)和第2胎次($n = 562$)的总产仔数和产活仔数进行GWA研究。

使用Illumina porcine 60K BeadChip进行分型获得SNP数据，使用gPlink软件^[8]进行质控。质控标准如下：平均检出率(Call Rate) > 0.8 ，平均最小等位基因频率(MAF) > 0.01 ，哈代-温伯格平衡检验(H-W equilibrium) χ^2 值 $10e-6$ 。共计57 814个SNPs被筛选出来用于关联分析。

1.2 统计分析

运用两种统计模型(1)Compressed Mixed Linear Model (GAPIT 软件)^[9]和(2)Bayes CPi(GenSel 程序包)^[10]进行第 1 和第 2 胎次总产仔数和产活仔数性状的全基因组关联分析。

Bayes CPi是在 Bayes B、Bayes C 和 GBLUP 方法的基础上发展而来, Bayes B 对使用者给出的先验遗传方差和残差值具有较大依赖性, 但若先验值不准确, 将会明显影响结果的准确性^[11]。Bayes C 和 Bayes CPi 对遗传方差和残差先验值的依赖性明显降低^[12], 但 Bayes C 的 pi 值一般多通过经验值来设定(对于多数分析性状, pi 值设为 0.995 左右)。Bayes CPi 会对先验 pi 值不断重新估计直到稳定。Gensel 程序包的 Bayes CPi 模型为:

$$y = Xb + \sum_{j=1}^K Z_j \mu_j \delta_j + e$$

其中, y 是表型值构成的向量; X 是一个表示固定效应的影响矩阵; b 是固定效应向量; K 是 SNPs 的个数; Z_j 代表在点 j 上 SNPs 的协变量; μ_j 是对点 j 的随机替代效应, 在假设方差为 σ_u^2 的条件下, 当 $\delta_j=1$ 时, μ_j 符合 $N(0, \sigma_u^2)$ 的正态分布, 当 $\delta_j=0$ 时, $\mu_j=0$; δ_j 是一个随机 0/1 随机变量, 表示模型中点 j 缺失(概率为 π)或出现(概率为 $1-\pi$); e 是一个向量表示随机残差效应, 假设 e 服从正态分布 $N(0, \sigma_e^2)$ 。

贝叶斯方法显著性检验不同于传统概率学派, 它是通过比较后验概率的比值大小来决定接受还是拒绝假设。假设检验假说 H_0 和 H_1 , 可根据其后验概率 p_0 和 p_1 来决定, 若 $p_0/p_1 > 1$, 则接受 H_0 , 若 $p_0/p_1 < 1$ 则接受 H_1 , 若两值相等则需进一步分析。

在该统计分析中 $p_0 = \text{model frequency}/(1-pi)$; $p_1 = (1-\text{model frequency})/pi$ 。 pi 值和 model frequency 为 GenSel 程序包计算所得得后验结果。为若 $p_0/p_1 > 1$, 则说明该 SNP 位点对性状有显著影响, 结果显示 Genetic Variance 最大的 50 个位点经显著性检验均为显著。

群体分层和个体亲缘关系存在是造成 GWAS 假阳性的主要原因之一^[13], 群体结构及不对等的亲缘关系可以由混合线性模型解释。在混合模型中, Subpopulation memberships(Q 矩阵)或 PCA(Principle

components analysis)被视为固定效应^[14,15], kinship 用来定义混合线性模型(Q+K)中随机个体效应的方差和协方差结构^[16]。紧缩混合线性模型(Compressed mixed linear model, CMLM)主要将个体分为不同组, 以更好控制群体分层和个体间的亲缘关系。通过最佳分组设置(选择最优组数和用于分组聚类方法)来提高统计效力。CMLM 统计模型为:

$$y = X\alpha + P\beta + K\mu + e$$

其中, y 为表型值构成的向量, X 为基因型, P 为 PCA 矩阵, K 为亲缘关系矩阵, $X\alpha$ 和 $P\beta$ 是固定效应, 即包括 PCA 矩阵的前 3 列和亲缘关系矩阵, $K\mu$ 和 e 是随机效应。CMLM 主要根据 kinship 应用分级群聚将个体分组, 当个体分配到不同组之后, 组内和组外的 kinship 将被当作 reduced kinship 矩阵的元素。初步设定 group.from=1, group.to=1000, group.by=10, 意味着一系列的 mixed model 会被用来获得一个最优 compression 水平, 每个模型使用不同 compression 水平的 reduced kinship 矩阵作为不同组之间的方差-协方差矩阵。产活仔数最佳 compression 水平为 30 组, 总产仔数的最佳 compression 水平为 820 组, 如图 1 所示。

2 结果与分析

选取 GAPIT 计算结果中 P 值(P -value)最显著的前 50 个 SNP 及 GenSel 结果中遗传方差(Genetic variation)最大前 50 个 SNP。两种统计方法交叉结果显示, 与第 1 胎次总产仔数性状显著关联的 SNP 位点有 31 个, 与第 1 胎次产活仔数性状显著关联的 SNP 位点有 20 个, 如表 1 和表 2 所示。与第 2 胎次总产仔数和产活仔数显著关联的位点数分别为 28 个和 27 个, 如表 3 和表 4 所示。

由两种统计方法对第 1 胎次的数据进行分析后, 获得 41 个重合 SNP 位点共分布在 13 个 QTL 区域, 其中在 1、3、13 和 16 染色体上有 5 个 SNP 位点或基因组区域同时与两性状显著关联。第 2 胎次的分析结果显示, 位于 7、10、12、13、14 和 16 号染色体的 6 个区域与两性状显著关联。图 2 至图 5 分别是第 1 胎次和第 2 胎次总产仔数和产活仔数的 Bayes CPi 方法分析结果。图 6 至图 9 分别是第 1 胎次和第 2 胎次总产仔数和产活仔数的 CMLM 方法分析结果。

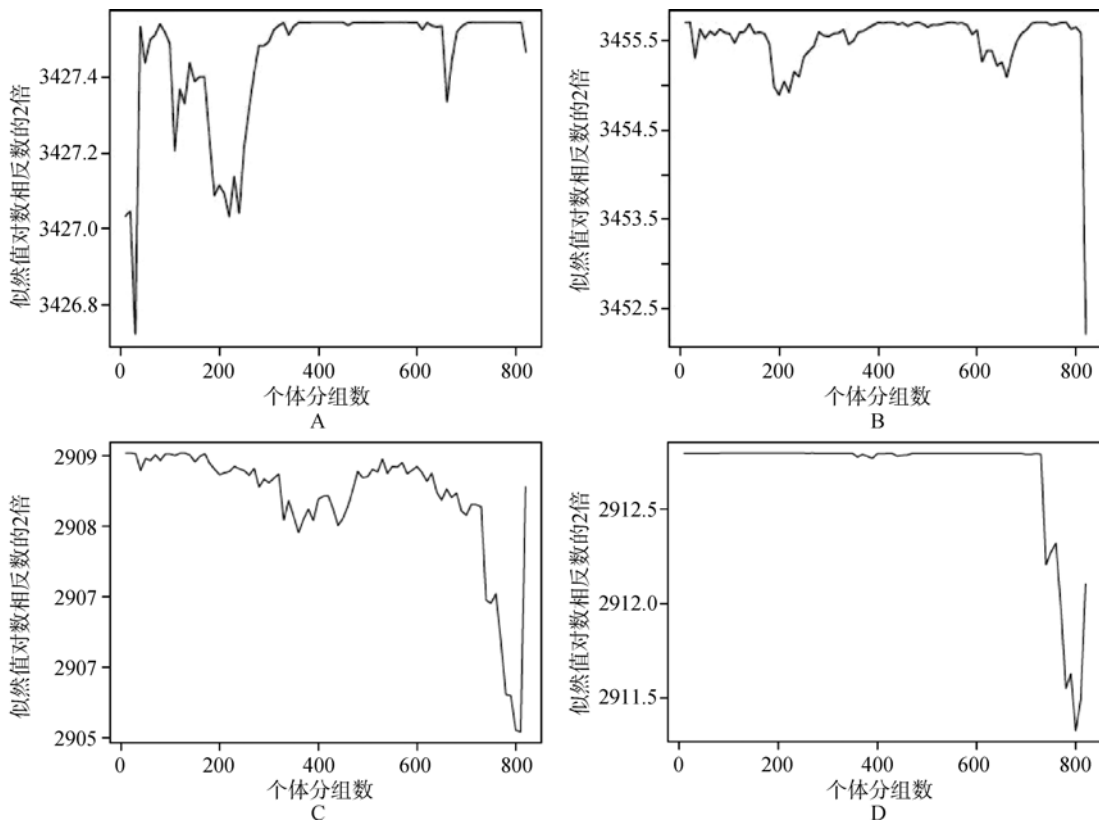


图 1 产仔数和产活仔数的紧凑水平

A: 第 1 胎次产活仔数; B: 第 1 胎次总产仔数; C: 第 2 胎次产活仔数; D: 第 2 胎次总产仔数。横坐标代表个体分组数, 纵坐标代表似然值对数相反数的 2 倍。

3 讨论

Fan 等^[4]利用 Bayes C 法对背膘厚度和眼肌面积两个重要的产肉性状进行全基因组分析, 识别出一批具有明显效应的候选基因及候选区域, 包括以前已识别的候选基因(如位于 1 号染色体与背膘厚度相关的 *MC4R* 基因和位于 2 号染色体与肌肉发育相关的 *IGF2* 基因)、QTL 区域以及一些目前功能尚不明瞭的新基因。Onteru 等^[6,7]亦采用 Bayes C 方法获得了一些与猪繁殖性状候选基因和基因组区域。上述结果表明用于基因组选择的 Bayes 统计方法能够较好地应用于 GWA 研究。

本研究在以往研究基础上, 采用两种新颖方法 CMLM 和 Bayes CPi 对猪第 1 胎次的产仔数和产活仔数进行分析。CMLM 方法可以认为是公畜 BLUP 模型的扩展, 主要通过分子标记或家系谱计算得来的 kinship 为基础, 通过分级群聚将个体分为很多组, 即为不同 compression 水平, 可以从 1 到 n (总个

体数), 计算出一系列混合线性模型的 REML 值来选取最优 compression 水平, 即获得最佳分组数。CMLM 能够有效降低群体分层造成的假阳性率^[9]。Bayes CPi 方法分析中, 我们设定 pi 值为 0.995, 即每次马尔科夫链迭代只有 250~300 个 SNP ($57814 \times 0.5\%$) 位点进入模型。经过 100000 次的迭代循环, 去掉前 4000 次结果, 得到 pi 的后验均值分别为 0.9998(总产仔数)和 0.9999(产活仔数), 后验 pi 值显示控制产仔数和产活仔数的基因为 5~6 个。

在所得显著 SNP 位点中, 位于 1 号染色体 248.5 Mb 区域的 M1GA0001337 位点同时与第 1 胎次产仔数和产活仔数显著关联, 其邻近下游基因 *MELK* (Maternal embryonic leucine zipper kinase, 母缘性胚胎拉链状蛋白激酶基因) 在猪中尚未有详细报道。桑葚胚向囊胚转换为早期猪胚胎发育重要时期, 许志峰^[17]利用 cDNA 微阵列及定量 PCR 技术验证了在此发育时期包括 *MELK* 基因在内的 13 个差异表达基

表 1 第 1 胎次总产仔数在 CMLM 和 Bayes CPi 两种统计结果中同时显著关联的 SNP 位点信息

SNP 位点	染色体	物理位置(bp)	上游基因	基因内	下游基因	P-value(CMLM)
M1GA0001337	1	248540131	<i>RNF38</i>	-	<i>MELK</i>	1.2401E-04
ALGA0008605	1	249574072	<i>POLR1E</i>	<i>GRHPR</i>	<i>FBXO10</i>	4.3611E-04
ALGA0008606	1	249704788	<i>FBXO10</i>	-	<i>TOMM5</i>	5.1823E-04
ASGA0006250	1	250294464	<i>SHB</i>	-	-	3.4982E-04
DIAS0002810	1	250632249	-	<i>IGFBPL1</i>	-	4.3597E-04
ASGA0006294	1	250776969	-	-	<i>TDRD7</i>	1.6078E-04
MARC0059534	1	251002929	-	<i>NCBP1</i>	-	7.6200E-05
ALGA0008726	1	251311226	<i>FOXE1</i>	<i>C9orf156</i>	<i>HEMGN</i>	2.1033E-04
ALGA0008728	1	251358130	<i>C9orf156</i>	<i>HEMGN</i>	-	1.3674E-04
MARC0027678	1	251877223	<i>TBC1D2</i>	<i>GABBR2</i>	<i>ANKS6</i>	4.4414E-04
H3GA0004085	1	251907987	<i>TBC1D2</i>	<i>GABBR2</i>	<i>ANKS6</i>	2.5513E-04
ALGA0015164	2	103964326	<i>MAN2A1</i>	-	<i>SLC25A46</i>	3.4061E-04
ALGA0015201	2	104984983	<i>B8XSJ3_PIG</i>	-	<i>MCC</i>	6.2923E-04
ASGA0015007	3	62180227	<i>LRRTM4</i>	-	<i>FAM176A</i>	6.8200E-05
ASGA0031621	7	20191499	<i>NRSN1</i>	<i>DCDC2</i>	<i>KAAG1</i>	7.3800E-05
ALGA0040620	7	40131728	<i>KIF6</i>	-	<i>DAAM2</i>	5.9712E-04
DRGA0012118	13	13886932	-	<i>RBMS3</i>	-	6.1890E-04
ASGA0072851	16	26926765	-	-	-	3.7290E-04
ALGA0090083	16	28598285	<i>ISL1</i>	-	<i>PELO</i>	1.0223E-04
ALGA0090341	16	33442575	<i>U6</i>	-	<i>MAP3K1</i>	6.0390E-04
ALGA0090494	16	39127560	-	-	<i>RNaseP_nuc</i>	9.2500E-06
ALGA0090520	16	40274210	<i>HTR1A</i>	-	<i>RNF180</i>	3.2968E-04
ALGA0090527	16	40359641	<i>HTR1A</i>	-	<i>RNF180</i>	5.4570E-04
ASGA0073238	16	40538895	<i>HTR1A</i>	<i>RNF180</i>	-	1.8102E-04
ALGA0098337	18	41783443	<i>CHN2</i>	-	<i>CPVL</i>	1.8708E-04
MARC0103339	Unknown	N/A	-	-	-	2.5200E-05
ALGA0119748	Unknown	N/A	-	-	-	4.6600E-05
ALGA0110739	Unknown	N/A	-	-	-	5.1586E-04
ASGA0099022	Unknown	N/A	-	-	-	5.1586E-04
MARC0073104	Unknown	N/A	-	-	-	3.9009E-04
ALGA0102937	Unknown	N/A	-	-	-	5.1586E-04

表 2 第 1 胎次产活仔数在 CMLM 和 Bayes CPi 两种统计结果中同时显著关联的 SNP 位点信息

SNP 位点	染色体	物理位置(bp)	上游基因	基因内	下游基因	P-value(CMLM)
M1GA0001337	1	248540131	<i>RNF38</i>	-	<i>MELK</i>	9.1313E-04
DIAS0002810	1	250632249	-	<i>IGFBPL1</i>	-	9.7933E-04
MARC0059534	1	251002929	-	<i>NCBP1</i>	-	1.0344E-03
ALGA0008728	1	251358130	<i>C9orf156</i>	<i>HEMGN</i>	-	7.7853E-04
MARC0027678	1	251877223	<i>TBC1D2</i>	<i>GABBR2</i>	<i>ANKS6</i>	1.0866E-03
H3GA0004085	1	251907987	<i>TBC1D2</i>	<i>GABBR2</i>	<i>ANKS6</i>	8.4422E-04
MARC0101263	3	23273914	<i>Q29209_PIG</i>	-	<i>GPR139</i>	1.4584E-04
ASGA0015007	3	62180227	<i>LRRTM4</i>	-	<i>FAM176A</i>	3.7190E-04
ASGA0015008	3	62192728	<i>LRRTM4</i>	-	<i>FAM176A</i>	1.0381E-03
MARC0011187	3	74886708	-	-	<i>BCL11A</i>	2.1206E-04
ALGA0020181	3	87692385	<i>A3FG76_PIG</i>	-	<i>PRKCE</i>	1.0200E-05
DIAS0004420	4	84268748	<i>SCYL3</i>	<i>C1orf112</i>	<i>C1orf156</i>	3.4547E-04
MARC0056621	4	127606083	<i>ARHGAP29</i>	<i>ABCA4</i>	<i>GCLM</i>	5.1434E-04
DRGA0012118	13	13886932	-	<i>RBMS3</i>	-	1.0589E-03
DRGA0012544	13	53008002	-	-	<i>Q6WL01_PIG</i>	1.1248E-03
ALGA0090494	16	39127560	-	-	<i>RNaseP_nuc</i>	1.1053E-03
H3GA0047209	16	70206984	<i>SNORD123</i>	<i>SEMA5A</i>	<i>U6</i>	1.7112E-04
ALGA0118231	Unknown	N/A	-	-	-	8.9518E-04
ASGA0085170	Unknown	N/A	-	-	-	1.0516E-03
MARC0103339	Unknown	N/A	-	-	-	8.0500E-05

表 3 第 2 胎次总产仔数在 CMLM 和 Bayes CPi 两种统计结果中同时显著关联的 SNP 位点信息

SNP 位点	染色体	物理位置(bp)	上游基因	基因内	下游基因	P-value(CMLM)
MARC0091344	3	22953207	-	-	-	1.0988E-03
ALGA0021541	3	115275637	TRIB2	-	B3VN77_PIG	1.3092E-03
ALGA0039211	7	20651724	-	FAM65B	-	1.1177E-03
ASGA0034307	7	65873887	NEO1	-	ARIH1	1.1548E-03
DRGA0008567	8	32814044	SLAIN2	-	SLC10A4	5.6112E-04
ASGA0046602	10	12734812	SNORA70	-	U6	9.1500E-06
H3GA0033068	12	4521970	NT5C	-	OTOP2	1.0023E-03
ALGA0111478	12	4974969	RPL38	-	-	3.6086E-04
ASGA0052791	12	5683623	SDK2	-	CDC42EP4	5.9072E-04
MARC0094394	12	7073879	SOX9_PIG	-	IRK2_PIG	5.7351E-04
H3GA0034087	12	30411738	PCTP	ANKFN1	-	1.5070E-03
ALGA0068757	13	17320225	-	-	ARPP-21	7.7799E-04
DRGA0012206	13	17715537	-	ARPP-21	-	2.6951E-04
H3GA0038673	14	7317353	-	-	-	1.4579E-03
ALGA0074792	14	7664330	-	-	ENTPD4	1.0606E-03
ASGA0063465	14	55822247	Q9MYZ4_PIG	-	ACTN2	1.4188E-04
INRA0044133	14	56054557	ACTN2	HEATR1	B7U2G8_PIG	1.1994E-04
MARC0024510	14	83039786	KCMA1_PIG	-	Q6EEI2_PIG	6.0110E-04
ALGA0079169	14	83060718	KCMA1_PIG	-	Q6EEI2_PIG	8.3488E-04
ALGA0080999	14	114317172	LOXL4	-	PYROXD2	7.3851E-04
ASGA0103451	16	17423627	-	-	RXFP3	8.5630E-04
DIAS0001721	16	17530818	S45A2_PIG	AMACR	B9UJD6_PIG	3.0633E-04
ASGA0102515	Unknown	N/A	-	-	-	3.2200E-05
ALGA0120747	Unknown	N/A	-	-	-	7.6754E-04
ASGA0089599	Unknown	N/A	-	-	-	1.1018E-03
MARC0034050	Unknown	N/A	-	-	-	1.1248E-03
ASGA0096565	Unknown	N/A	-	-	-	1.1637E-03
ASGA0100324	Unknown	N/A	-	-	-	1.3140E-03

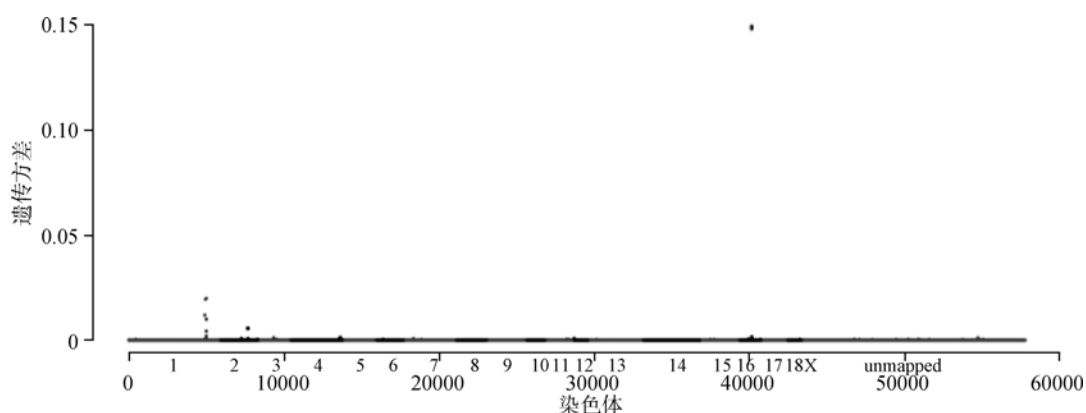
因, 提出 *MELK* 基因可能会促进猪胚胎的成功发育。与第 1 胎次产仔数显著关联的 ALGA0090341 位点位于 16 号染色体 33.4 Mb 区域, 其下游基因 *MAP3K1* (Mitogen-activated protein kinase kinase kinase 1) 被证明在小鼠早期胚胎形成中具有重要作用, 与上述 *MELK* 基因有类似的作用^[18]。此外, 位于 16 号染色体 28.6 Mb 区域 ALGA0090083 位点与第 1 胎次产仔数显著关联, 其附近下游基因 *PELO* (Protein pelota

homolog)) 在精子生成、细胞周期调控等方面起到重要作用^[19]。此外, 两种统计方法结果亦显示, 对于第 1 胎次和第 2 胎次, 影响总产仔数和产活仔数的基因和染色体区域有着明显差异, 这与以往结论相符合^[7]。

假阳性影响着 GWAS 结果可信度, 提高统计效力和降低假阳性率是全基因组关联分析所需考虑重点。CMLM 和 Bayes CPi 方法分别以不同方式来降

表 4 第 2 胎次产活仔数在 CMLM 和 Bayes C_{PI} 两种统计结果中同时显著关联的 SNP 位点信息

SNP 位点	染色体	物理位置(bp)	上游基因	基因内	下游基因	P-value(CMLM)
ASGA0034307	7	65873887	NEO1	-	ARIH1	4.0054E-04
MARC0033686	7	70638297	-	SRP54	-	9.7379E-04
DRGA0007728	7	70679935	-	-	U2	9.7379E-04
MARC0085735	9	10825983	-	-	ACER3	1.3631E-03
ALGA0051344	9	10844306	-	-	ACER3	1.3631E-03
ALGA0051348	9	10994864	ACER3	-	B3GNT6	1.3631E-03
ASGA0043787	9	69984121	PPP1R9A	-	U2	4.1640E-04
ASGA0046602	10	12734812	SNORA70	-	U6	1.7600E-05
H3GA0034087	12	30411738	PCTP	ANKFN1	-	4.8397E-04
ALGA0068757	13	17320225	-	-	ARPP-21	4.7858E-04
DRGA0012206	13	17715537	-	ARPP-21	-	9.4700E-05
MARC0000560	13	22241837	ZNF662	-	C3orf39	1.4165E-04
ASGA0057071	13	25131342	-	-	SCAP_PIG	6.9400E-05
ASGA0057076	13	25357370	PTPN23	CSPG5	SMARCC1	3.7200E-05
ASGA0062747	14	37049852	MED13L	-	TBX3	8.5984E-04
ALGA0076950	14	37285957	MED13L	-	TBX3	7.7036E-04
ALGA0076953	14	37310040	MED13L	-	TBX3	1.2134E-03
H3GA0039826	14	38329332	B0JF98_PIG	-	U6	1.3729E-03
MARC0082918	14	38461143	U6	-	RBM19	1.4110E-03
ASGA0063465	14	55822247	Q9MYZA_PIG	-	ACTN2	2.2266E-04
INRA0044133	14	56054557	ACTN2	HEATR1	B7U2G8_PIG	1.4412E-04
ASGA0103451	16	17448839	-	-	RXFP3	4.5430E-04
DIAS0001721	16	17530818	S45A2_PIG	AMACR	B9UJD6_PIG	8.2004E-04
H3GA0049138	17	47939453	SFRS6	-	IFT52	7.2162E-04
ASGA0102515	Unknown	N/A	-	-	-	1.8000E-05
ASGA0085906	Unknown	N/A	-	-	-	4.2764E-04
ASGA0094388	Unknown	N/A	-	-	-	7.2648E-04

图 2 第 1 胎次总产仔数性状的 Bayes C_{PI} 法全基因组关联分析结果

横坐标表示 SNP 所在基因组位置, 其中 unmapped 表示此 SNP 所在染色体位置未知; 纵坐标为 5 个连续 SNP 位点的遗传方差。

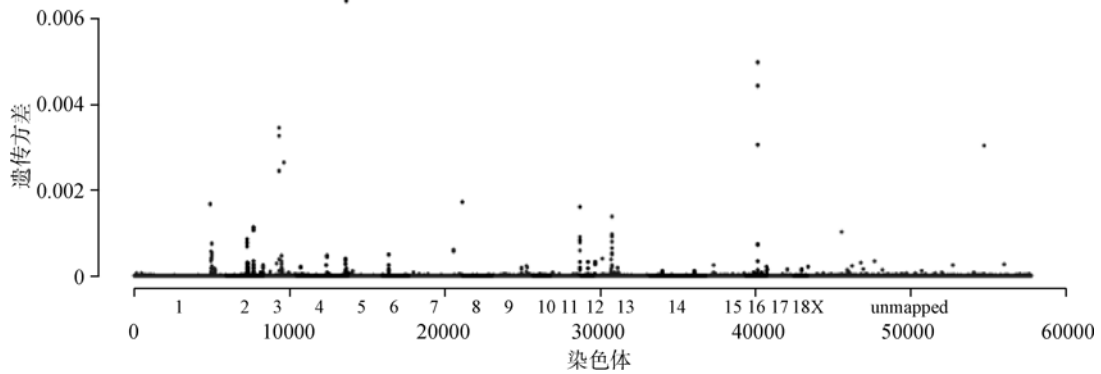


图 3 第一胎次产活仔数性状的 Bayes Cpi 法全基因组关联分析结果
横坐标表示 SNP 所在基因组位置, 其中 unmapped 表示此 SNP 所在染色体位置未知; 纵坐标为 5 个连续 SNP 位点的遗传方差。

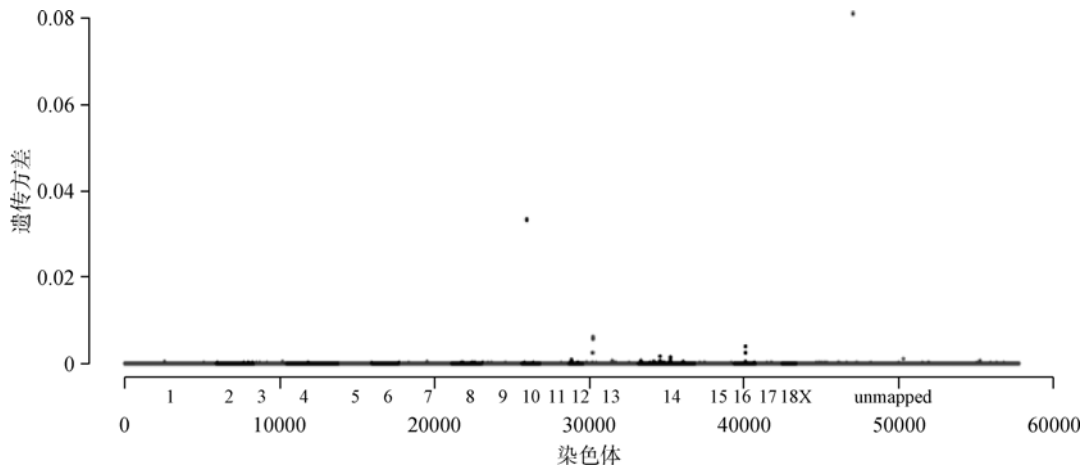


图 4 第 2 胎次总产仔数性状的 Bayes Cpi 法全基因组关联分析结果
横坐标表示 SNP 所在基因组位置, 其中 unmapped 表示此 SNP 所在染色体位置未知; 纵坐标为 5 个连续 SNP 位点的遗传方差。

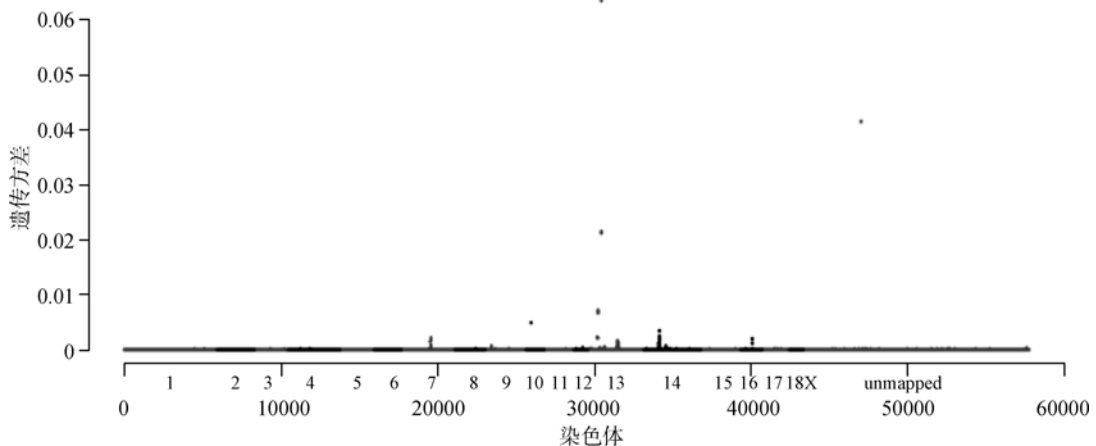


图 5 第 2 胎次产活仔数性状的 Bayes Cpi 法全基因组关联分析结果
横坐标表示 SNP 所在基因组位置, 其中 unmapped 表示此 SNP 所在染色体位置未知; 纵坐标为 5 个连续 SNP 位点的遗传方差。

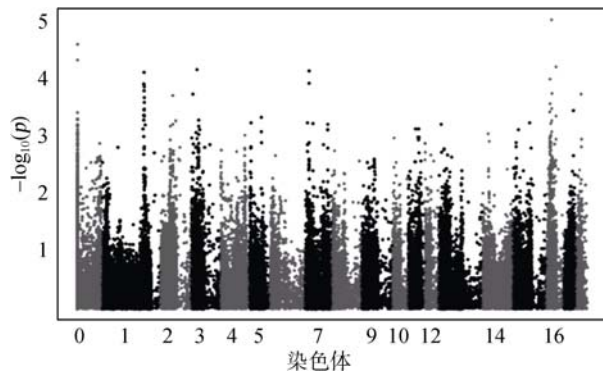


图 6 第 1 胎次总产仔数性状的 CMLM 法全基因组关联分析结果

横坐标表示 SNP 所在基因组位置, 其中染色体 0 表示此 SNP 所在染色体位置未知; 纵坐标为关联分析 P 值的 $-\log_{10}$ 值。

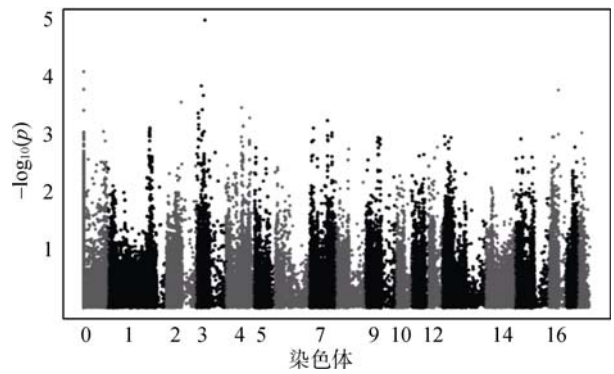


图 7 第 1 胎次产活仔数性状的 CMLM 法全基因组关联分析结果

横坐标表示 SNP 所在基因组位置, 其中染色体 0 表示此 SNP 所在染色体位置未知; 纵坐标为关联分析 P 值的 $-\log_{10}$ 值。

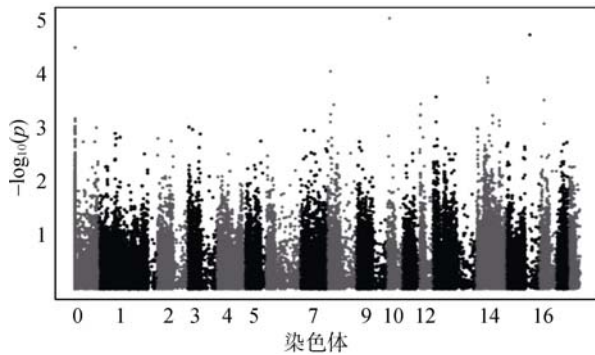


图 8 第 2 胎次总产仔数性状的 CMLM 法全基因组关联分析结果

横坐标表示 SNP 所在基因组位置, 其中染色体 0 表示此 SNP 所在染色体位置未知; 纵坐标为关联分析 P 值的 $-\log_{10}$ 值。

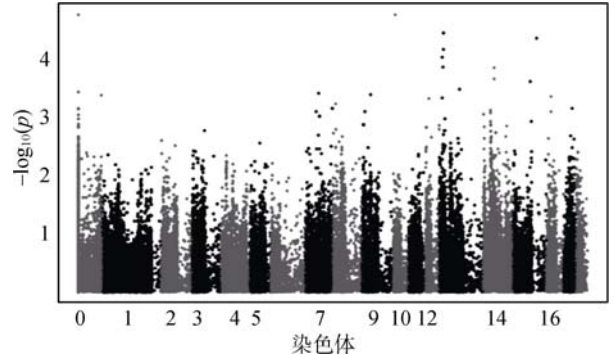


图 9 第 2 胎次产活仔数性状的 CMLM 法全基因组关联分析结果

横坐标表示 SNP 所在基因组位置, 其中染色体 0 表示此 SNP 所在染色体位置未知; 纵坐标为关联分析 P 值的 $-\log_{10}$ 值。

低全基因组关联分析中的假阳性率, 选取两种统计结果中的重合位点会显著降低 II 类错误, 但也会造成 I 类错误的增加, 即漏掉部分候选基因, 这需要进一步改进统计方法, 在降低假阳性率同时保证统计效力不会减弱。

参考文献(References):

- [1] Rothschild M, Jacobson C, Vaske D, Tuggle C, Wang LZ, Short T, Eckardt G, Sasaki S, Vincent A, McLaren D, Southwood O, van der Steen H, Mileham A, Plastow G. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proc Natl Acad Sci USA*, 1996, 93(1): 201–205. [DOI](#)
- [2] Li N, Zhao YF, Xiao L, Zhang FJ, Chen YZ. Candidate gene approach for identification of genetic loci controlling litter size in swine. In: Proceedings of the 6th World Congress Quantitative Genetics of Livestock. 1998, 26: 183–188.
- [3] 赵要风, 李宁, 肖璐, 曹更生, 陈怡真, 张顺, 陈永福, 吴常信, 张建生, 孙士铨, 徐学清. 猪 *FSHβ* 亚基因结构区反转座子插入突变及其与猪产仔数关系的研究. *中国科学c辑*, 1999, 29(1): 81–86. [DOI](#)
- [4] Fan B, Onteru SK, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PLoS One*, 2011, 6(2): e14726. [DOI](#)
- [5] Gorbach DM, Cai W, Dekker JCM, Young JM. Whole-genome analyses for genes associated with residual feed intake and related traits utilizing the PorcineSNP60 BeadChip. In: Pig Genome III Conference. Hinxtton, Cambridge, UK, 2009.
- [6] Onteru SK, Fan B, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF. A whole-genome association study for pig

- reproductive traits. *Anim Genet*, 2012, 43(1): 18–26. [DOI](#)
- [7] Onteru SK, Fan B, Nikkilä MT, Garrick DJ, Stalder KJ, Rothschild MF. Whole-genome association analyses for lifetime reproductive traits in the pig. *J Anim Sci*, 2011, 89(4): 988–995. [DOI](#)
- [8] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am J Hum Genet*, 2007, 81(3): 559–575. [DOI](#)
- [9] Zhang ZW, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu JM, Arnett DK, Ordovas JM, Buckler ES. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet*, 2010, 42(4): 355–360. [DOI](#)
- [10] Fernando RL, Garrick DJ. GenSel - User manual for a Portfolio of genomic selection related analyses. *Anim Breed Genet*, Iowa State University, Ames. <http://bigsgui/login.html>, 2008.
- [11] Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 2001, 157(4): 1819–1829. [DOI](#)
- [12] Kizilkaya K, Fernando RL, Garrick DJ. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J Anim Sci*, 2010, 88(2): 544–551. [DOI](#)
- [13] Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Human Genet*, 2000, 67(1): 170–181. [DOI](#)
- [14] Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*, 2006, 2(12): e190. [DOI](#)
- [15] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, 2006, 38(8): 904–909. [DOI](#)
- [16] Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley J, McMullen M, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*, 2005, 38(2): 203–208. [DOI](#)
- [17] 许志峰. 选殖猪母源性胚胎拉链状白胺酸蛋白质激酶之调控区域及其特性分析[学位论文]. 台湾大学动物科学技术学研究所, 2010.
- [18] Hickson JA, Fong B, Watson PH, Watson AJ. *PP2Cδ* (*Ppm1d*, *WIP1*), an endogenous inhibitor of p38 *MAPK*, is regulated along with *Trp53* and *Cdkn2a* following p38 *MAPK* inhibition during mouse preimplantation development. *Mol Reproduct Dev*, 2007, 74(7): 821–834. [DOI](#)
- [19] Ewing RM, Chu P, Elisma F, Li HY, Taylor P, Climine S, McBroon-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang SD, Ornatsky O, Bukhman Y, Ethier M, Sheng YL, Vasilescu J, Abu-Farha M, Lambert JP, Duewel HS, Stewart I, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D. Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol*, 2007, 3: 89. [DOI](#)

更正

《遗传》2012年34卷第8期978页题为“X染色体的剂量补偿机制”，作者王艳允、陈梅、李斌的文章表1应该为如下表格：

	雌性	雄性
黑腹果蝇	XX	<u>X</u> Y
哺乳类	X _x	XY
秀丽隐杆线虫	<u>xx</u>	X

特此更正！

王艳允
2012.9.27