

# 京海黄鸡体重性状全基因组关联分析

张涛<sup>1,2</sup>, 王文浩<sup>1,2</sup>, 张跟喜<sup>1,2</sup>, 王金玉<sup>1,2</sup>, 薛倩<sup>1,2</sup>, 顾玉萍<sup>3</sup>

1. 扬州大学动物科学与技术学院, 扬州 225009;
2. 江苏省动物遗传繁育与分子设计重点实验室, 扬州 225009;
3. 江苏京海集团, 南通 226103

**摘要:** 体重性状是肉鸡重要的经济性状。为了寻找可用于京海黄鸡体重性状遗传改良的分子标记及候选基因, 本文以 400 只京海黄鸡核心群母鸡为基础, 测定了 0~14 周龄体重, 利用简化基因组测序技术(Specific-locus amplified fragment sequencing, SLAF-seq)对京海黄鸡体重性状进行全基因组关联研究(Genome-wide association study, GWAS), 筛选与京海黄鸡体重性状相关的 SNPs 位点。结果共检测到 100 个与京海黄鸡体重相关的 SNPs 位点, 其中 15 个位点效应达到全基因组显著水平( $P < 1.87 \times 10^{-6}$ ), 85 个位点效应达到全基因组潜在显著水平( $P < 3.73 \times 10^{-5}$ )。通过筛选每个显著 SNP 周围 1 Mb 区域内的基因, 共找到 9 个可能的候选基因, 其中 *FAM124A*(Family with sequence similarity 124A)、*QDPR*(Quinoid dihydropteridine reductase)、*WDR1*(WD repeat domain 1)和 *SLC2A9*(Solute carrier family 2 (facilitated glucose transporter), member 9) 4 个基因可能是影响体重性状的重要候选基因。同时还发现, 4 号染色体 75.6~80.7 Mb 区域集中了大部分与京海黄鸡中后期体重性状显著相关的 SNPs 位点, 该区域可能是影响京海黄鸡中后期生长体重的重要候选区域。

**关键词:** 京海黄鸡; 全基因组关联分析; 体重性状; 简化基因组测序

## A genome-wide association study on body weight traits of Jinghai yellow chicken

Tao Zhang<sup>1,2</sup>, Wenhao Wang<sup>1,2</sup>, Genxi Zhang<sup>1,2</sup>, Jinyu Wang<sup>1,2</sup>, Qian Xue<sup>1,2</sup>, Yuping Gu<sup>3</sup>

1. College of Animal Science and Technology, Yangzhou University, Yangzhou 225009, China;
2. Key Laboratory of Animal Genetics, Breeding, Reproduction and Molecular Design of Jiangsu Province, Yangzhou 225009, China;
3. Jiangsu Jinghai Poultry Group Co., Ltd., Nantong 226103, China

**Abstract:** Body weight traits are important economic characters of broilers. This study was carried out to screen for molecular markers and candidate genes that can be used to improve the body weight traits. A herd of 400 female Jinghai yellow chickens were measured for body weights from 0 to 14 weeks of age. Genome-wide association study (GWAS) was carried out using specific-locus amplified fragment sequencing (SLAF-seq) technology to detect SNPs associated with body weight traits of Jinghai yellow chicken. Finally, 100 SNPs that associated with body weight traits were detected. The results showed that effects of 15 SNPs reached 5% Bonferroni genome-wide significance

收稿日期: 2015-02-11; 修回日期: 2015-04-13

基金项目: 国家肉鸡产业技术体系项目(编号: nycytx-42-G1-05), 江苏高校优势学科建设工程项目和江苏省动物遗传繁育与分子设计重点实验室项目资助

作者简介: 张涛, 博士研究生, 研究方向: 动物遗传育种与繁殖。E-mail: zt991279320@126.com

通讯作者: 王金玉, 教授, 博士生导师, 研究方向: 动物遗传育种与繁殖。E-mail: jywang@yzu.edu.cn

DOI: 10.16288/j.ycz.15-080

网络出版时间: 2015-7-7 13:42:51

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20150707.1342.001.html>

and 85 SNPs reached potential genome-wide significance. Genes in the candidate regions with 1 Mb windows (SNP position  $\pm 0.5$  Mb) surrounding each significant SNP were screened. Finally, nine candidate genes were obtained, among which four genes of *FAM124A* (Family with sequence similarity 124A), *QDPR* (Quinoid dihydropteridine reductase), *WDR1* (WD repeat domain 1) and *SLC2A9* (Solute carrier family 2 (facilitated glucose transporter), member 9) might be important candidate genes influencing body weight traits of Jinghai yellow chicken. Furthermore, it was also found that most SNPs associated with mid and late growth and body weights were intensively located in the region of 75.6–80.7 Mb on chromosome 4. Our study thus provides a basis for genetic understanding of the Jinghai yellow chicken body weight traits.

**Keywords:** Jinghai yellow chicken; GWAS; body weight traits; SLAF-seq

体重性状是肉鸡最为重要的经济性状之一，体重性状的优劣直接影响肉鸡产业。多年来，通过传统的育种方法对肉鸡体重性状进行的遗传改良取得了良好效果，然而，体重性状与体组成等性状一样，为复杂经济性状，受诸多基因的控制。近年来，随着科技的飞速发展，分子标记辅助育种已成为改良遗传性状的新方法。目前，鸡的 QTL 数据库中已经有超过 1500 个 QTLs 被认为和鸡的体重性状相关<sup>[1,2]</sup>。

QTLs 以及 SNPs 等标记的发现，大多采用单链构象多态性 (Single stranded conformational polymorphism, SSCP)、限制性片段长度多态性 (Restricted fragment length polymorphisms, RFLP) 等传统方法，但是传统的方法存在检出率低以及易受环境影响的缺点，结果准确度低<sup>[1]</sup>。目前，全基因组关联分析 (Genome-wide association study, GWAS) 作为一种新的方法在动物育种领域得到广泛应用<sup>[3]</sup>。全基因组关联分析旨在从全基因组范围内寻找与性状关联的 SNPs，得到的结果更具可靠性。近几年，GWAS 在影响鸡体重性状的 QTL 定位方面得到了广泛应用，筛选到了一些影响鸡生长、繁殖、体组成、肉质、抗马立克以及抗新城疫等性状的 SNP 位点和候选基因<sup>[4-11]</sup>。

目前，所有 GWAS 研究都是基于 SNPs 芯片技术，但芯片技术只能检测已知 SNP 位点，无法发现新的位点。最近，一项新的以高通量测序为基础的技术被开发出来，即简化基因组测序技术 (Specific-locus amplified fragment sequencing, SLAF-seq)。相比其他技术，简化基因组测序具有多个优点：(1) 一次测序获得数以百万计的 SNP，密度高；(2) 与芯片相比，可以检测基因组上未知变异位点中新的 SNP；(3) 有无参考基因组物种均适合；(4) SLAF-seq 升级技术获得完整的 2×100 bp 测序片段，SNP 标记转化成功率高<sup>[12]</sup>。

京海黄鸡是本课题组在长期开展我国地方鸡种质资源调查、评价与保护的基础上，经多年连续攻关成功培育而成的我国目前唯一通过国家畜禽遗传资源委员会审定的鸡新品种，具有小型、优质、早熟和抗逆 4 大特点。京海黄鸡的育成，即采用了常规育种方法，又采用了分子辅助选择方法，是鸡育种中理论与实践相结合的一个范例。本研究首次采用 SLAF-seq 技术在全基因组范围内寻找与京海黄鸡体重相关的 SNP 位点，旨在为京海黄鸡体重性状的标记辅助选择寻找新的遗传标记，为京海黄鸡体重性状的改良提供理论素材。

## 1 材料和方法

### 1.1 材料

实验群体为 400 只 11 世代京海黄鸡母鸡，饲养于江苏京海集团江边种鸡场，所有鸡只同一天孵化，自由采食饮水，饲养条件相同。60 日龄采血，0 周龄开始每隔两周称重，记录体重。

### 1.2 方法

#### 1.2.1 基因组 DNA 提取

采用翅静脉采血的方法采集 400 只京海黄鸡母鸡的血液，血液 DNA 采用常规的饱和酚/氯仿法抽提<sup>[13]</sup>，浓度稀释至 50~100 ng/μL，NanoDrop 2000 紫外可见光度计检测 DNA 浓度和质量，确保符合测序分型的要求。

#### 1.2.2 基因组修饰

首先，利用限制性内切酶 *Hae* 对基因组 DNA 进行酶切处理。酶切体系：基因组 DNA 500 ng，NEB Buffer4 1 μL，限制性内切酶 0.12 μL，补水至 50 μL，37℃ 酶切反应 15 h。反应结束后用 QIAGEN 试剂盒纯

化, 50  $\mu\text{L}$  EB 回溶。将酶切产生的不同类型的末端进行平端化修复, 同时对 5'末端进行磷酸化修饰, 具体方法为: 纯化后的 DNA 样品 30  $\mu\text{L}$ , ATP 终浓度为 10 mmol/L 的 T4 DNA Ligase Buffer 10  $\mu\text{L}$ , 10 mmol/L dNTP Mix 4  $\mu\text{L}$ , T4 DNA 聚合酶 5  $\mu\text{L}$ , Klenow 酶 1  $\mu\text{L}$ , T4 多核苷酸激酶 5  $\mu\text{L}$ , 补水至 100  $\mu\text{L}$  混匀, PCR 仪 20 恒温反应 30 min。磷酸化修饰反应结束后用 QIAGEN 试剂盒纯化, 33  $\mu\text{L}$  EB 回溶。在 5'磷酸化的平端 DNA 片段 3'末端添加一个“A”, 以便后续与 5'端有一个突出“T”的 Solexa 接头进行互补连接, 从而提高接头连接效率, 同时阻止 Solexa 接头自连。具体方法为: 纯化的 DNA 样品 32  $\mu\text{L}$ , Klenow Buffer 5  $\mu\text{L}$ , 1 mmol/L dATP 10  $\mu\text{L}$ , Klenow Exo 3  $\mu\text{L}$  混匀, 37 水浴 30 min。反应结束后用 QIAGEN 试剂盒纯化, 10  $\mu\text{L}$  EB 回溶。最后连接 Solexa 接头, 使不同的 DNA 片段具有相同的末端序列, 从而可以在后续实验中能够顺利的杂交到流路池(Flow cell)上, 进行扩增、测序。具体方法为: 纯化的 DNA 样品 10  $\mu\text{L}$ , 2 $\times$ DNA Ligase Buffer 25  $\mu\text{L}$ , Adapter 10  $\mu\text{L}$ , DNA Ligase 5  $\mu\text{L}$  混匀, PCR 仪 20 恒温反应 15 min, 反应结束后用 QIAGEN 试剂盒纯化, 30  $\mu\text{L}$  EB 回溶。

### 1.2.3 PCR 扩增及测序

根据酶切预测软件的分析结果, 切胶回收 500~580 bp 的 DNA 片段, 进行 PCR 扩增。PCR 上游引物为 5'-AATGATACGCGCACCACCGA-3', 下游引物为 5'-CAAGCAGAAGACGGCATAACG-3'。扩增体系为: 纯化的 DNA 产物 8  $\mu\text{L}$ , PCR 上游引物 1.5  $\mu\text{L}$ , PCR 下游引物 1.5  $\mu\text{L}$ , Phusion DNA 聚合酶 20  $\mu\text{L}$ , 补水至 40  $\mu\text{L}$ 。扩增程序为: 98 预变性 30 s; 98 变性 10 s, 65 复性 30 s, 72 延伸 30 s, 18 个循环; 最后再 72 延伸 5 min。反应结束后使用 QIAGEN 试剂盒纯化, 30  $\mu\text{L}$  EB 回溶。将纯化的 PCR 扩增产物进行精确定量(Qubit), 然后在芯片(Flow cell)表面进行桥式 PCR, 使 DNA 片段扩增为单分子 DNA 簇(此过程在 Cluster station 中进行)。单分子 DNA 簇扩增好后将流路池移入 IlluminaHiSeq2000 中, 进行测序。

### 1.2.4 统计分析

根据 DNA 条形码序列, 将原始 reads 按照个体

分类, 剔除低质量的 reads(质量得分<20), 然后使用 SOAP 2.20<sup>[14]</sup>软件将合格的 reads 与获取自 GenBank 的原鸡参照基因组(GenBank 登录号: GCA\_000002315.2)进行比对, 双端均比对到基因组上的 reads 为可靠的, 可以用来定义 SLAF 标签。根据比对和错误校正结果, 选取样本平均深度在 4 以上的组来定义 SLAF 标记。统计基因组上每 100K 范围内 SLAF 标记个数, 得到 SLAF 标记在染色体上的分布(图略), 然后根据定义好的 SLAF 标记, 进行基于 400 个样本的群体内部 SNP 检测, 统计每 100K 范围内 SNP 个数, 绘制 SNP 在每条染色体上的分布图。

使用 Plink(v1.07)<sup>[15]</sup>对得到的 SNP 进行质量控制, 剔除最小检出率低于 85%和最小等位基因频率低于 5%的 SNP。使用 SPSS19.0 对表型数据进行统计分析, 使用 Admixture<sup>[16]</sup>分析群体结构, 使用 TASSEL<sup>[17]</sup>一般线性模型(GLM I)进行全基因组关联分析, 所用模型为:

$$y = \mu + G\alpha + Q\beta + e$$

其中,  $y$  代表表型值;  $\mu$  为固定效应值向量;  $G$  为 SNP 的效应值向量,  $\alpha$  为每个 SNP 的权重向量;  $Q$  为群体结构, 使用 admixture 软件计算<sup>[16]</sup>, 不同群体的百分比作为协变量代入公式;  $\beta$  为每个群体的权重向量;  $e$  为随机误差, 最终, 每个 SNP 位点都能得到一个关联值。独立 SNP 标记具体计算方法参照 Wang 等<sup>[18]</sup>的方法, 最终得到 15 719 个独立的标记。一系列相邻的 SNP 之间如果  $r^2$  值大于 0.4, 则定义为连锁不平衡模块。使用此方法, 最终估计得到独立 SNPs 标记和 Block 模块共 26 767 个, 根据常染色体上的连锁不平衡(LD)模块以及估计的独立 SNPs 标记数量, 计算得到 5% Bonferroni 校正的全基因组显著  $P$  值为  $1.87\text{E}-06(0.05/26767)$  5% Bonferroni 校正的全基因组潜在显著  $P$  值为  $3.73\text{E}-05(1/26767)$ 。使用 TASSEL 3.0 软件生成各时期体重全基因组关联分析的 Quantile-Quantile(QQ)图和曼哈顿图<sup>[19]</sup>。

## 2 结果与分析

### 2.1 表型值的描述性统计

京海黄鸡体重性状的描述性统计结果见表 1, 数据均符合或近似符合正态分布。

### 2.2 测序结果

测序结果显示, SLAF 标记在染色体上分布均匀,

表 1 京海黄鸡体重性状的描述性统计

性状	统计量	极小值	极大值	平均值	标准误	变异系数
0 周龄体重(g)	394	20	45	35.30457	0.19209	0.245264
2 周龄体重(g)	378	85	160	125.2249	0.695615	0.250335
4 周龄体重(g)	393	185	385	267.2697	1.709913	0.245572
6 周龄体重(g)	384	260	570	418.6094	2.612687	0.248397
8 周龄体重(g)	362	405	850	567.2431	3.475832	0.255734
10 周龄体重(g)	391	550	1025	749.3862	4.427484	0.246191
12 周龄体重(g)	377	640	1160	880.6658	5.119986	0.250662
14 周龄体重(g)	382	795	1455	1079.385	5.909753	0.249038

根据定义好的 SLAF 标记,进行基于 400 个样本的群体内部 SNP 检测。结果共检测到 1 122 263 个 SNP,通过统计基因组每 100K 范围内的 SNP 数,得到 SNP 在基因组各染色体上的分布情况(图 1)。结果显示,基于 SLAF 标记检测到的 SNP 在染色体上分布均匀,SNP 数据可靠。结合 SLAF 和 SNP 数据,有 SNP 存在的 SLAF 标记定义为多态性 SLAF 标记,对 SLAF 多态性进行统计,共有 88 135 个多态性 SLAF,平均深度为 5.46。对数据进行质量控制后,最终共有 90 890 个 SNPs 可用于后续的 GWAS 分析。独立 SNP 计算后,共得到独立 SNPs 标记和 Block 模块 26 767 个。

### 2.3 群体结构分析

利用 Admixture 软件分析京海黄鸡的群体结构(图 2),结果显示,样本存在分群现象,当 K 为 10 时峰值最低,说明当把实验群体分为 10 个群来计算群体结构为最优。鉴于群体分层可能影响关联分析的准确性,本文绘制了各个时期体重的 Quantile-Quantile 图(图 3)。结果显示,观测值(纵坐标)与期望值(横坐标)基本相符,说明关联分析并没有因为群体分层而产生假阴性,关联分析结果可靠。

### 2.4 全基因组关联分析

全基因组关联分析结果显示(表 2),共筛选到 15 个与京海黄鸡生长性状达到 5% Bonferroni 全基因组关联显著的位点。其中,1 个位点(rs4256627)与初生重显著相关,位于 4 号染色体的 25.7 kb 处;2 个位点(rs477915563 和 rs480716328)与 6 周龄体重显著相关,均位于 4 号染色体上;2 个位点(rs477915563 和 rs523339314)与 8 周龄体重显著相关,分别位于 4 号和 5 号染色体上;5 个位点(rs1169337928、rs476249085、rs476878211、rs477673186 和 rs477915563)与 10 周龄体重显著相关,其中 4 个位于 4 号染色体,1 个位于 1 号染色体;3 个位点(rs1169337928、rs475641139

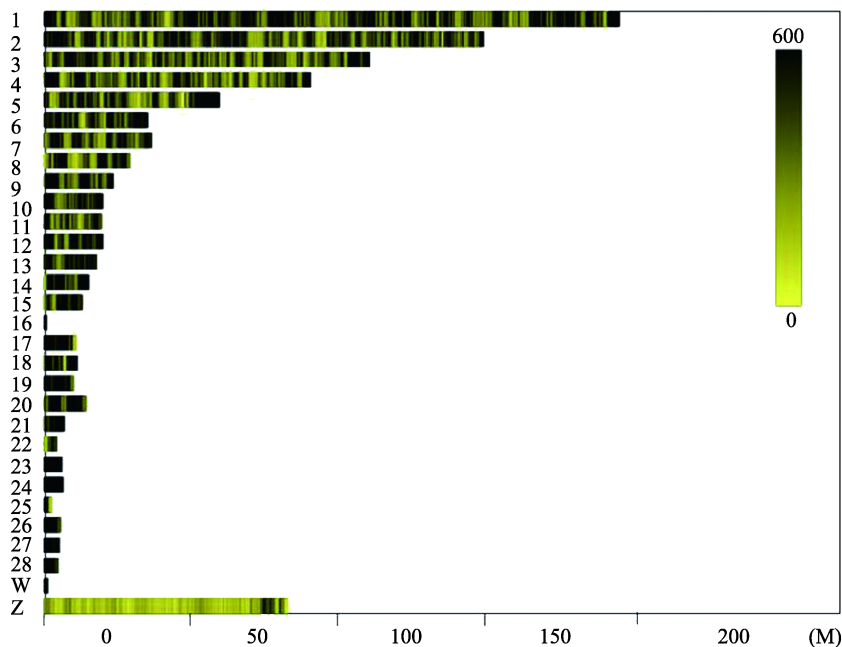


图 1 SNP 在染色体上的分布情况

纵坐标为染色体号,横坐标为染色体长度(M),图右上角标尺表示 SNP 个数,颜色越深,SNP 越多,纯黑色表示 SNP 个数超过 600。

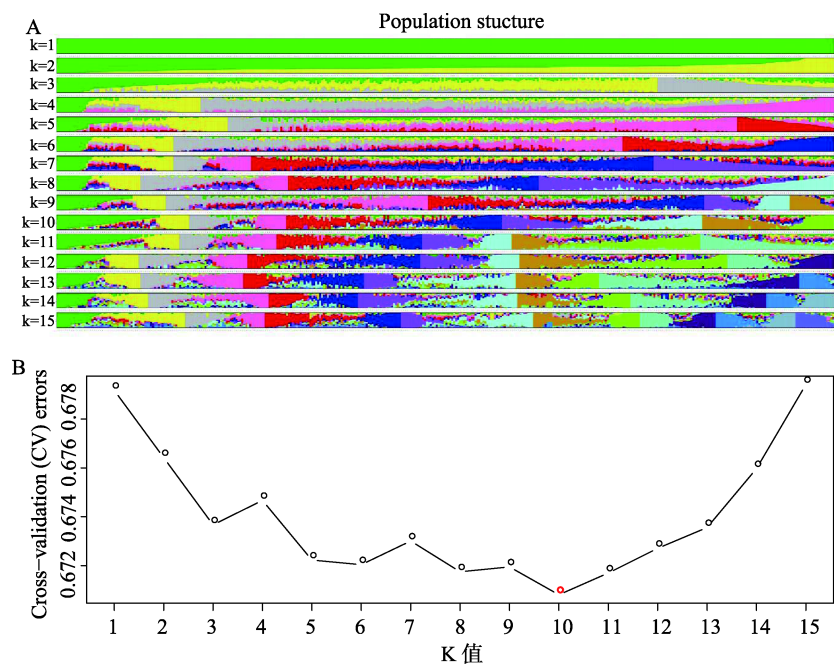


图 2 鸡的群体结构与聚类分析结果

A: 群体聚类图(每种颜色代表一个群, 每行代表一个分群值的情况, 图中展示了样品分群值从 1~15 大聚类情况); B: 群体结构图(每个 K 值对应的 CV 值)。

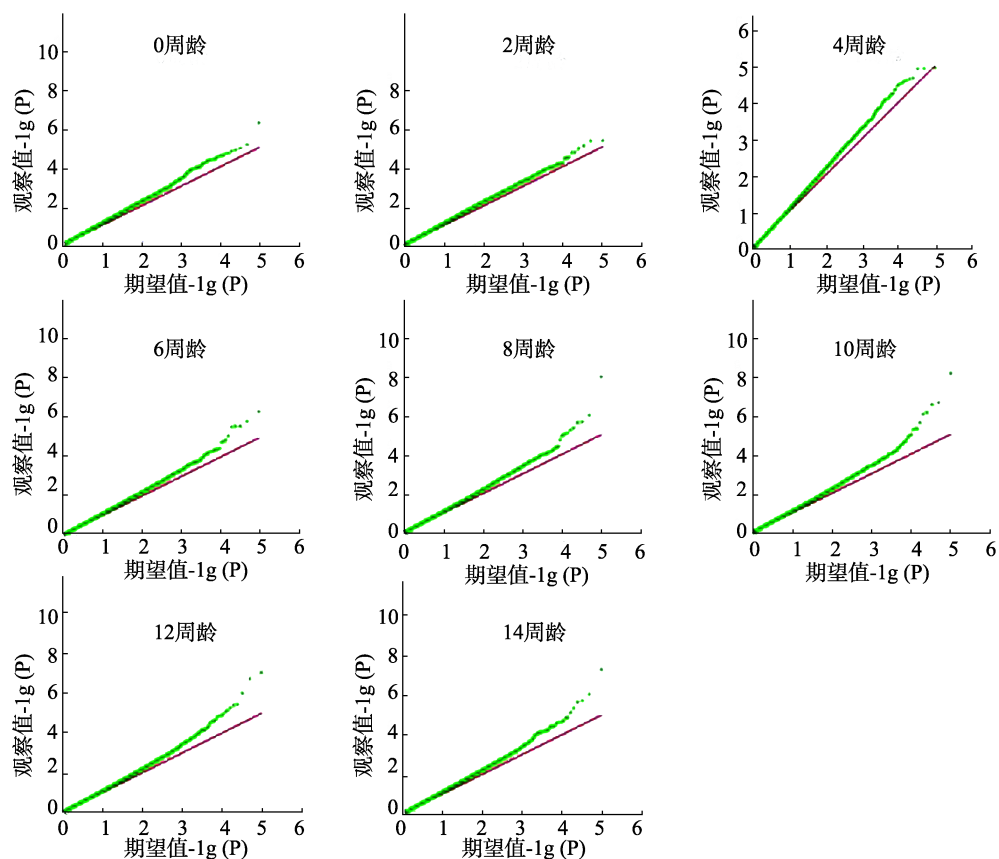


图 3 不同时期体重性状的 Quantile-Quantile 图

红线为假设期望值与观测值相等时的中间线, 绿线为实际每个 SNP 位点根据观测值与期望值大小所定位的位置构成的曲线。



表 2 与生长性状显著关联的 SNPs 位点

性状	SNP 位点	染色体	位置	P 值	最近基因	距离
0 周龄体重	rs4256627	4	256627	6.76E-07	<i>VSIG4</i>	上游 2.0 kb
6 周龄体重	rs477915563	4	77915563	5.54E-07	<i>WDR1</i>	上游 29.3 kb
	rs480716328	4	80716328	1.77E-06	<i>GPR78</i>	下游 9.0 kb
8 周龄体重	rs477915563	4	77915563	1.45E-08	<i>WDR1</i>	上游 29.3 kb
	rs523339314	5	23339314	1.27E-06	未知	
10 周龄体重	rs1169337928	1	169337928	3.94E-07	<i>FAM124A</i>	下游 3.7 kb
	rs476249085	4	76249085	1.01E-06	<i>FBXL5</i>	内部
	rs476878211	4	76878211	1.21E-06	<i>BODIL1</i>	上游 78.6 kb
	rs477673186	4	77673186	3.20E-07	<i>HS3ST1</i>	下游 47.8 kb
	rs477915563	4	77915563	1.06E-08	<i>WDR1</i>	上游 29.3 kb
12 周龄体重	rs1169337928	1	169337928	1.21E-06	<i>FAM124A</i>	下游 3.7 kb
	rs475641139	4	75641139	1.12E-07	<i>QDPR</i>	下游 57.8 kb
	rs477915563	4	77915563	2.23E-07	<i>WDR1</i>	上游 29.3 kb
14 周龄体重	rs1169337928	1	169337928	1.08E-06	<i>FAM124A</i>	下游 3.7 kb
	rs477915563	4	77915563	5.90E-08	<i>WDR1</i>	上游 29.3 kb

和 rs477915563)与 12 周龄体重显著相关, 其中, 2 个位于 4 号染色体, 1 个位于 1 号染色体; 2 个位点(rs1169337928 和 rs477915563)与 14 周龄体重显著相关, 分别位于 1 号和 4 号染色体。没有筛选到与 2 周龄和 4 周龄体重显著相关的 SNPs。此外共筛选出 85 个与京海黄鸡生长性状潜在关联显著的位点(附表 1)。在显著关联的位点中, rs477915563 与京海黄鸡的生长中后期 5 个时期体重(6~14 周龄体重)均关联显著, rs480716328 与生长中后期 4 个时期体重(6、10、12、14 周龄体重)显著相关。各个性状的曼哈顿图见图 4。

本研究一共筛选到 15 个显著的 SNPs 位点和 85 个潜在显著的 SNPs 位点。从 SNPs 的分布情况(图 1)和曼哈顿图(图 4)来看, 与生长前期相关的位点分布较为分散, 没有发现聚集区域, 而与生长中后期相关的 SNPs 则分布较为集中, 大多数位于 4 号染色体 75.6~80.7 Mb 区域内。在每个显著的 SNPs 周围 1 Mb 区域内寻找可能的候选基因, 结果共找到 8 个可能的候选基因(表 2), 分别为 *VSIG4* (V-set and immunoglobulin domain containing 4)、*WDR1*(WD repeat domain 1)、*GPR78*(G protein-coupled receptor 78)、*FAM124A*(Family with sequence similarity 124A)、*FBXL5*(F-box and leucine-rich repeat pro-

tein 5)、*BODIL1*(Biorientation of chromosomes in cell division 1-like 1)、*HS3ST1*(Heparan sulfate (glucosamine) 3-O-sulfotransferase 1) 和 *QDPR*(Quinoid dihydropteridine reductase)基因。其中 *WDR1* 基因和 *FAM124A* 基因分别位于 rs477915563 和 rs1169337928 两个 SNPs 附近, 这两个位点与多个时期的体重显著相关。同时还发现, 4 号染色体 77954046~77954086 bp 之间一段长 40 bp 的区域内存在多个与京海黄鸡生长性状相关的位点, 且均位于 *SLC2A9*(Solute carrier family 2 (facilitated glucose transporter), member 9)基因内部。对以上 9 个可能候选基因进行 GO annotation 分析(表 3), 结果显示: 有 6 个基因在鸡中查找到相关注释, 细胞组分分析表明, 其中 5 个基因均位于细胞重要的组分中; 分子功能分析显示, 6 个基因均具有酶活性或者受体功能; 生物学进程分析显示, 其中 5 个基因参与了细胞包括代谢、信号转导等在内的重要生物学进程。*VSIG4*、*FAM124A* 和 *BODIL1* 3 个基因在鸡中未有相关注释信息。

### 3 讨论

全基因组关联分析以遍布于整个基因组的单核苷酸多态性分子标记为基础, 对复杂性状进行直接的关联分析, 被认为是一种确定影响重要经济性状

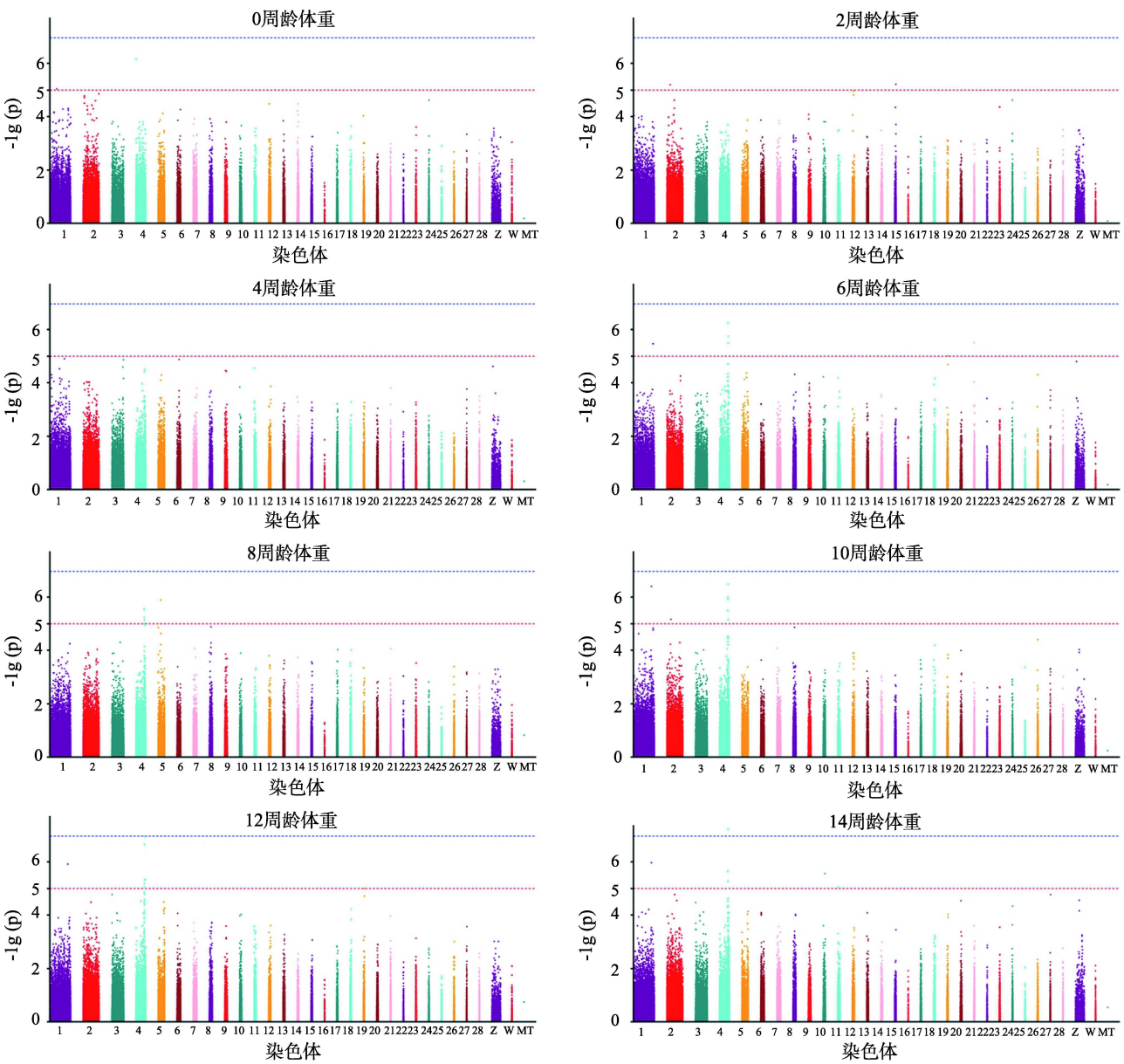


图 4 不同时期体重性状的曼哈顿图  
横坐标为染色体号，纵坐标为  $P$  值，红色虚线为阈值  $1E-05$ ，蓝色虚线为阈值  $1E-07$ ，Z、W 和 MT 分别代表 Z、W 染色体和线粒体染色体。

表 3 候选基因的 GO 注释分析

基因	GO 注释分析		
	细胞组分	分子功能	生物学进程
WDR1	细胞质、细胞骨架	肌动蛋白结合	未知
GPR78	膜整体组分	G 蛋白偶联受体活性	腺苷酸环化酶激活的 G 蛋白偶联受体信号通路
FBXL5	细胞质核周区	铁离子结合	铁离子平衡、正调控细胞蛋白质分解代谢
HS3ST1	未知	磺基转移酶活性	细胞生长代谢过程
QDPR	线粒体、细胞外外来体	6,7-二氢喋呤还原酶活性	氧化还原进程
SLC2A9	核膜	葡萄糖跨膜转运蛋白活性	尿酸盐代谢、己糖跨膜转运

的分子标记的有效方法<sup>[5,9]</sup>。目前, GWAS 已经在鸡生长性状遗传改良方面得到应用。Gu 等<sup>[5]</sup>利用 60 K SNP 芯片对乌骨鸡和白洛克鸡杂交 F<sub>2</sub> 代体重性状进行全基因组关联分析, 结果筛选到 9 个 SNPs 以及位于 4 号染色体上的一个 8.6 Mb(71.6~80.2)的区域对体重性状具有显著影响。Johansson 等<sup>[20]</sup>利用 60K SNP 芯片对两个品系多达 50 代鸡生长性状进行不间断的全基因组关联分析, 发现 4 号染色体 60~80 Mb 区域对生长性状具有重要影响。虽然已有关于鸡生长性状全基因组关联分析的诸多报道, 但是全部研究均是基于芯片技术, 然而芯片技术与飞速发展的测序技术相比, 存在成本高以和通量低等缺点。本研究首次利用下一代测序技术/简化基因组测序在全基因组范围内筛选与京海黄鸡生长性状相关的 SNP 位点, 该技术是一种新开发的通过 *de novo* 测序来寻找 SNPs 的策略, 使用扩大 RRL(Reduced representation library)文库的测序方法, 对大群体进行基因分型, 具有很多优点<sup>[12]</sup>。本研究旨在为京海黄鸡生长性状的遗传改良提供理论基础。

群体分层是影响无关个体实验设计结果准确性的最重要因素之一, 本研究利用 Admixture 软件分析京海黄鸡的群体结构, 结果显示将京海黄鸡分为 10 个群进行计算最为合理, 说明京海黄鸡存在群体分层现象, 若不考虑群体结构的影响, 则分层效应会被误认为是基因效应, 造成关联分析的假阴性。因此本研究将 10 个群体中每个群体的百分比作为协变量代入公式, 进行群体结构校正, 以消除群体分层的影响, 同时绘制每个时期体重的 QQ 图以检验群体校正的准确性, QQ 图结果显示, 观测值与期望值基本相符, 仅在末端存在少许偏离位点, 说明群体结构校正的结果准确可靠, 关联分析并没有因为群体分层而产生假阴性。

关联分析结果显示, 本文共筛选到 99 个效应位点, 其中 15 个效应位点(含 8 个不同的 SNPs 位点)与京海黄鸡 6 个不同生长时期体重显著相关( $P < 1.87E-06$ ), 85 个效应位点与京海黄鸡生长性状潜在显著相关( $P < 3.73E-05$ )。显著关联的位点中, rs477915563 与京海黄鸡的生长中后期 5 个周龄体重(6~14 周龄体重)均显著相关, rs480716328 与生长中后期 4 个周龄体重(6、10、12、14 周龄体重)显著相关, 说明这两个 SNPs 位点对京海黄鸡中后期体重具有重要

影响, 可作为影响京海黄鸡体重性状的重要候选遗传标记。剩余 6 个 SNPs 仅与单个时期体重显著相关, 可靠性尚待进一步研究。

本研究筛选到的 99 个效应位点中, 与生长前期相关的位点分布较为分散, 没有发现聚集区域, 而与生长中后期相关的 SNPs 则分布较为集中, 大多数位于 4 号染色体上 75.6~80.7Mb 区域内, 说明该区域可能为影响京海黄鸡生长性状的重要候选区域, 这与 Gu 等<sup>[5]</sup>和 Johansson 等<sup>[20]</sup>的研究结果相符, 进一步说明了本研究结果的可靠性。然而, 本研究结果与 Xie 等<sup>[4]</sup>和吴丹等<sup>[21]</sup>的研究结果存在差异: Xie 等利用 60K SNP 芯片对白洛克和杏花鸡 F<sub>2</sub> 代杂交群体生长性状和日增重进行全基因组关联分析, 发现位于 1 号染色体的 173.5~175 Mb 区域是影响生长和日增重性状的重要候选区域; 吴丹等研究显示影响北京油鸡生长性状的 SNPs 位点分布较为分散, 没有明显集中区域。这种差异可能由品种差异造成的。

在每个显著的 SNPs 周围 1 Mb 区域内寻找可能的候选基因, 结果共找到 8 个可能的候选基因, 分别为 *VSIG4*、*WDR1*、*GPR78*、*FAM124A*、*FBXL5*、*BODIL1*、*HS3ST1* 和 *QDPR* 基因。其中 *WDR1* 基因和 *FAM124A* 基因分别位于 2 个与多个时期体重显著相关的 SNPs 附近, 可能为重要的候选功能基因。相关研究表明, *WDR1* 基因编码激动蛋白相互作用蛋白 1, 其通过与肌动蛋白解聚因子相互作用来增强肌动蛋白纤维分裂的比率<sup>[22]</sup>, 杜晓燕<sup>[23]</sup>研究表明, *WDR1* 基因可能通过 MAPK(Mitogen-activated protein kinase)信号途径参与对骨骼肌细胞生命活动的调节, 由此可见, 该基因可能参与了京海黄鸡肌肉生长发育的调节, 从而间接影响京海黄鸡体重性状。*FAM124A* 基因在诸多人类疾病相关的 GWAS 分析中被报道<sup>[24~26]</sup>, 但是尚未发现其相关的功能报道, 其同一家族的 FAM184B 为重要的功能基因, 影响牛的采食日增重和屠体重<sup>[27]</sup>。同时本文还发现, 4 号染色体 77954046~77954086 bp 一段长 40 bp 的区域内存在多个与京海黄鸡生长性状相关的位点, 且均位于同一个基因 *SLC2A9*(Solute carrier family 2 (facilitated glucose transporter), member 9)内部, 说明 *SLC2A9* 基因也可能为影响生长性状的重要候选基因。研究表明, 该基因为 1 个新的尿酸转运蛋白, 影响血浆及血清尿酸的沉积、排泄以及痛风<sup>[28~30]</sup>。



本课题组前期对京海黄鸡体组成性状进行了 GWAS 研究, 发现 *QDPR* 为影响体组成性状的重要候选基因, 而北京油鸡体重和体组成性状的 GWAS 研究也显示, *QDPR* 基因是影响北京油鸡体重和体组成性状的重要候选基因<sup>[5,9,21]</sup>, 这说明该基因为影响鸡生长性状的重要候选功能基因。*VSIG4* 是 B7 家族相关蛋白的一员, 该基因是 T 细胞激活的负调控因子, 能够抑制小鼠和人 T 细胞的增殖<sup>[31~33]</sup>; *FBXL5* 基因在维持小鼠细胞和系统铁离子平衡方面发挥重要作用<sup>[34]</sup>; *HS3ST1* 基因可能与正常的血液凝固有关<sup>[35]</sup>; *GPR78* 基因可能与两极性情绪紊乱及精神分裂有关<sup>[36]</sup>; 未发现 *BOD1L1* 相关的功能研究。以上结果显示, *VSIG4*、*GPR78*、*FBXL5*、*BOD1L1* 和 *HS3ST1* 等 5 个基因的相关研究多集中在人类疾病及免疫方面, 在动物尤其是家禽方面的研究尚未见报道, 因此其对京海黄鸡生长性状的影响尚待进一步验证, 而 *WDR1*、*FAM124A*、*QDPR* 和 *SLC2A9* 4 个基因已有相关研究, 可能为影响京海黄鸡生长性状的重要候选功能基因。

附录 附表 1 见文章电子版(www.chinagene.cn)。

## 参考文献

- [1] 樊庆灿, 王金玉, 张跟喜, 唐莹, 张涛, 顾玉萍, 施会强. 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. 畜牧兽医学报, 2014, 45(7): 1053–1059. [DOI]
- [2] 葛玉洋, 李琦华, 樊月圆, 初晓辉, 徐志强, 陈小波, 葛长荣, 贾俊静, 荣华. 单核苷酸多态性在鸡生长发育中的应用. 安徽农业科学, 2013, 41(8): 3395–3396. [DOI]
- [3] 冯春刚, 胡晓湘, 赵要风, 李宁. 全基因组选择及其在动物育种中的应用. 中国家禽, 2008, 30(22): 5–8. [DOI]
- [4] Xie L, Luo CL, Zhang CG, Zhang R, Tang J, Nie QH, Ma L, Hu XX, Li N, Da Y, Zhang XQ. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. *PLoS One*, 2012, 7(2): e30910. [DOI]
- [5] Gu XR, Feng CG, Ma L, Song C, Wang YQ, Da Y, Li HF, Chen KW, Ye SH, Ge CR, Hu XX, Li N. Genome-wide association study of body weight in chicken F<sub>2</sub> resource population. *PLoS One*, 2011, 6(7): e21872. [DOI]
- [6] Liu WB, Li DF, Liu JF, Chen SR, Qu LJ, Zheng JX, Xu GY, Yang N. A genome-wide SNP scan reveals novel loci for egg production and quality traits in white leghorn and brown-egg dwarf layers. *PLoS One*, 2011, 6(12): e28600. [DOI]
- [7] Wolc A, Arango J, Settari P, Fulton JE, O'Sullivan NP, Preisinger R, Habier D, Fernando R, Garrick DJ, Hill WG, Dekkers JCM. Genom-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Anim Genet*, 2012, 43(Suppl. 1): 87–96. [DOI]
- [8] Noorai RE, Freese NH, Wright LM, Chapman SC, Clark LA. Genome-wide association mapping and identification of candidate genes for the rumpless and ear-tufted traits of the araucana chicken. *PLoS One*, 2012, 7(7): e40974. [DOI]
- [9] Liu RR, Sun YF, Zhao GP, Wang FJ, Wu D, Zheng MQ, Chen JL, Zhang L, Hu YD, Wen J. Genome-wide association study identifies loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. *PLoS One*, 2013, 8(4): e61172. [DOI]
- [10] Li DF, Lian L, Qu LJ, Chen YM, Liu WB, Chen SR, Zheng JX, Xu GY, Yang N. A genome-wide SNP scan reveals two loci associated with the chicken resistance to Marek's disease. *Anim Genet*, 2013, 44(2): 217–222. [DOI]
- [11] Luo CL, Qu H, Ma J, Wang J, Li CY, Yang CF, Hu XX, Li N, Shu DM. Genome-wide association study of antibody response to Newcastle disease virus in chicken. *BMC Genet*, 2013, 14(1): 42. [DOI]
- [12] Sun XW, Liu DY, Zhang XF, Li WB, Liu H, Hong WG, Jiang CB, Guan N, Ma CX, Zeng HP, Xu CH, Song J, Huang L, Wang CM, Shi JJ, Wang R, Zheng XH, Lu CY, Wang XW, Zheng HK. SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS One*, 2013, 8(3): e58700. [DOI]
- [13] 马新红, 亢娟娟, 康相涛, 韩瑞丽, 孙桂荣, 潘军, 黄煌, 刘凯, 张立恒. 鸡基因组 DNA 不同提取方法的比较研究. 江西农业大学学报, 2010, 32(1): 181–184. [DOI]
- [14] Li RQ, Yu C, Li YR, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 2009, 25(15): 1966–1967. [DOI]
- [15] 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 analyses. *Am J Hum Genet*, 2007, 81(3): 559–575. [DOI]
- [16] Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individual. *Genome Res*, 2009, 19(9): 1655–1664. [DOI]
- [17] Bradbury PJ, Zhang ZW, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 2007, 23(19): 2633–2635. [DOI]
- [18] Wang D, Sun Y, Stang P, Berlin JA, Wilcox MA, Li QQ. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. *BMC Proc*, 2009, 3(Suppl. 7): S109. [DOI]

- [19] 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]
- [20] Johansson AM, Pettersson ME, Siegel PB, Carlborg Ö. Genome-wide effects of long-term divergent selection. *PLoS Genet*, 2010, 6(11): e1001188. [DOI]
- [21] 吴丹, 刘冉冉, 赵桂苹, 郑麦青, 张磊, 胡耀东, 文杰. 鸡体质量性状基因的全基因组关联研究. *畜牧兽医学报*, 2012, 43(12): 1887–1896. [DOI]
- [22] Chen H, Bernstein BW, Bamburg JR. Regulating actin-filament dynamics *in vivo*. *Trends Biochem Sci*, 2000, 25(1): 19–23. [DOI]
- [23] 杜晓燕. 真核生物高度保守基因 WDR1 的表达和功能初步研究[学位论文]. 长沙: 湖南师范大学, 2006. [DOI]
- [24] Gervasini C, Pfundt R, Castronovo P, Russo S, Roversi G, Masciadri M, Milani D, Zampino G, Selicorni A, Schoenmakers EF, Larizza L. Search for genomic imbalances in a cohort of 24 Cornelia de Lange patients negative for mutations in the *NIPBL* and *SMC1L1* genes. *Clin Genet*, 2008, 74(6): 531–538. [DOI]
- [25] Mick E, McGough JJ, Middleton FA, Neale B, Faraone SV. Genome-wide association study of blood pressure response to methylphenidate treatment of attention-deficit/hyperactivity disorder. *Prog Neuro-psychopharmacol Biol Psychiatry*, 2011, 35(2): 466–472. [DOI]
- [26] Bossini-Castillo L, de Kovel C, Kallberg H, van't Slot R, Italiaander A, Coenen M, Tak PP, Posthumus MD, Wijmenga C, Huizinga T, van der Helm-van Mil AH, Stoeken-Rijsbergen G, Rodriguez-Rodriguez L, Balsa A, González-Álvarez I, González-Gay MÁ, Gómez-Vaquero C, Franke B, LifeLines Cohort Study, Vermeulen S, van der Horst-Bruinsma Ie, Dijkmans BA, Wolbink GJ, Ophoff RA, Maehlen MT, van Riel P, Merriman M, Klareskog L, Lie BA, Merriman T, Crusius JB, Brouwer E, Martin J, de Vries N, Toes R, Padyukov L, Koeleman BP. A genome-wide association study of rheumatoid arthritis without antibodies against citrullinated peptides. *Ann Rheum Dis*, 2015, 74(3): e15. [DOI]
- [27] Lindholm-Perry AK, Sexten AK, Kuehn LA, Smith TP, King DA, Shackelford SD, Wheeler TL, Ferrell CL, Jenkins TG, Snelling WM, Freetly HC. Association, effects and validation of polymorphisms within the *NCAPG-LCORN* locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *BMC Genet*, 2011, 12(1): 103. [DOI]
- [28] Vitart V, Rudan I, Hayward C, Gray NK, Floyd J, Palmer CN, Knott SA, Kolcic I, Polasek O, Graessler J, Wilson JF, Marinaki A, Riches PL, Shu X, Janicijevic B, Smolej-Narancic N, Gorgoni B, Morgan J, Campbell S, Biloglav Z, Barac-Lauc L, Pericic M, Klaric IM, Zgaga L, Skaric-Juric T, Wild SH, Richardson WA, Hohenstein P, Kimber CH, Tenesa A, Donnelly LA, Fairbanks LD, Aringer M, McKeigue PM, Ralston SH, Morris AD, Rudan P, Hastie ND, Campbell H, Wright AF. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet*, 2008, 40(4): 437–442. [DOI]
- [29] Döring A, Gieger C, Mehta D, Gohlke H, Prokisch H, Coassin S, Fischer G, Henke K, Klopp N, Kronenberg F, Paulweber B, Pfeufer A, Rosskopf D, Völzke H, Illig T, Meitinger T, Wichmann HE, Meisinger C. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat Genet*, 2008, 40(4): 430–436. [DOI]
- [30] Caulfield MJ, Munroe PB, O'Neill D, Witkowska K, Charchar FJ, Doblado M, Evans S, Eyheramendy S, Onipinla A, Howard P, Shaw-Hawkins S, Dobson RJ, Wallace C, Newhouse SJ, Brown M, Connell JM, Dominiczak A, Farrall M, Lathrop GM, Samani NJ, Kumari M, Marmot M, Brunner E, Chambers J, Elliott P, Kooner J, Laan M, Org E, Veldre G, Viigimaa M, Cappuccio FP, Ji C, Iacone R, Strazzullo P, Moley KH, Cheeseman C. SLC2A9 is a high-capacity urate transporter in humans. *PLoS Med*, 2008, 5(10): e197. [DOI]
- [31] Vogt L, Schmitz N, Kurrer M O, Bauer M, Hinton HI, Behnke S, Gatto D, Sebbel P, Beerli RR, Sonderegger I, Kopf M, Saudan P, Bachmann MF. VSIG4, a B7 family-related protein, is a negative regulator of T cell activation. *J Clin Invest*, 2006, 116(10): 2817–2826. [DOI]
- [32] 张世杰, 王蕾, 王月颖, 蒋靓, 蒋玉平, 高超, 张学光, 顾宗江. 人 VSIG4 基因转染细胞的构建及其对 T 细胞的共刺激效应的初步研究. *中国免疫学杂志*, 2009, 25(5): 397–401. [DOI]
- [33] 李鸣, 王沂芹, 李燕, 陶光利, 庞琪, 袁发焕. B7 家族共刺激分子 VSIG4 在巨噬细胞/T 细胞共培养模型中的作用. *第三军医大学学报*, 2011, 33(22): 2335–2339. [DOI]
- [34] Ruiz JC, Walker SD, Anderson SA, Eisenstein RS, Bruick RK. F-box and leucine-rich repeat protein 5 (FBXL5) is required for maintenance of cellular and systemic iron homeostasis. *J Biol Chem*, 2013, 288(1): 552–560. [DOI]
- [35] Shworak NW, HajMohammadi S, de Agostini AI, Rosenberg RD. Mice deficient in heparan sulfate 3-O-sulfotransferase-1: normal hemostasis with unexpected perinatal phenotypes. *Glycoconj J*, 2002, 19(4–5): 355–361. [DOI]
- [36] Underwood SL, Christoforou A, Thomson PA, Wray NR, Tenesa A, Whittaker J, Adams RA, Le Hellard S, Morris SW, Blackwood DH, Muir WJ, Porteous DJ, Evans KL. Association analysis of the chromosome 4p-located *G protein-coupled receptor 78 (GPR78)* gene in bipolar affective disorder and schizophrenia. *Mol Psychiatry*, 2006, 11(4): 384–394. [DOI]

附表 1 与生长性状潜在显著关联的 SNPs

性状	SNP 位点	染色体	位置	P 值
0 周龄体重	rs164873473	1	64873473	9.01E-06
	rs29596783	2	9596783	1.66E-05
	rs29596830	2	9596830	3.11E-05
	rs29596842	2	9596842	1.92E-05
	rs280676113	2	80676113	3.68E-05
	rs2113814977	2	113814977	2.52E-05
	rs2145862047	2	145862047	1.39E-05
	rs124390292	12	4390292	3.28E-05
	rs148063686	14	8063686	3.27E-05
	rs243369253	24	3369253	2.41E-05
2 周龄体重	rs230222993	2	30222993	6.27E-06
	rs269203363	2	69203363	2.40E-05
	rs1212221606	12	12221606	1.52E-05
	rs1212221635	12	12221635	1.10E-05
	rs1511249826	15	11249826	6.03E-06
	rs242232472	24	2232472	2.38E-05
4 周龄体重	rs190895746	1	90895746	2.97E-05
	rs1137349933	1	137349933	1.25E-05
	rs3103297408	3	103297408	2.57E-05
	rs3109538936	3	109538936	1.35E-05
	rs481321678	4	81321678	3.23E-05
	rs618225098	6	18225098	1.32E-05
	rs97769271	9	7769271	3.45E-05
	rs914889437	9	14889437	3.65E-05
	rs112127805	11	2127805	2.81E-05
	rsz10657893	Z	10657893	2.41E-05
6 周龄体重	rs1184918155	1	184918155	3.40E-06
	rs1184918156	1	184918156	3.40E-06
	rs477673186	4	77673186	1.94E-05
	rs477954086	4	77954086	3.17E-06
	rs198767371	19	8767371	2.06E-05
	rs198767625	19	8767625	9.74E-06
	rs211170592	21	1170592	3.10E-06
	rsz10657893	Z	10657893	1.57E-05
8 周龄体重	rs475641139	4	75641139	3.01E-06
	rs475765686	4	75765686	1.01E-05
	rs476249085	4	76249085	2.80E-06
	rs476387533	4	76387533	5.75E-06
	rs477673186	4	77673186	7.22E-06
	rs480716328	4	80716328	1.12E-05
	rs51773778	5	1773778	1.39E-05
	rs525471381	5	25471381	2.33E-05
	rs815544609	8	15544609	1.29E-05

续表

性状	SNP 位点	染色体	位置	P 值
10 周龄体重	rs148078219	1	48078219	2.36E-05
	rs1184918155	1	184918155	1.46E-05
	rs1184918156	1	184918156	1.71E-05
	rs37848921	2	37848921	6.81E-06
	rs475641139	4	75641139	3.16E-06
	rs475765686	4	75765686	2.97E-05
	rs477954046	4	77954046	3.53E-05
	rs477954058	4	77954058	3.53E-05
	rs477954067	4	77954067	3.11E-05
	rs477954086	4	77954086	7.10E-06
	rs480716328	4	80716328	6.76E-06
	rs814726973	8	14726973	1.35E-05
12 周龄体重	rs198767371	19	8767371	1.94E-05
	rs198767625	19	8767625	1.00E-05
	rs272268660	2	72268660	3.27E-05
	rs31633105	3	1633105	1.66E-05
	rs445501112	4	45501112	2.99E-05
	rs475765686	4	75765686	8.90E-06
	rs475856029	4	75856029	3.51E-05
	rs475856030	4	75856030	3.04E-05
	rs476878211	4	76878211	1.37E-05
	rs477673186	4	77673186	6.54E-06
	rs477954045	4	77954045	2.60E-05
	rs477954046	4	77954046	1.62E-05
	rs477954058	4	77954058	1.62E-05
	rs477954067	4	77954067	1.61E-05
	rs477954086	4	77954086	4.77E-06
	rs488636962	4	88636962	4.51E-06
	rs552414945	5	52414945	3.20E-05
14 周龄体重	rs1014266685	10	14266685	2.73E-06
	rs113727958	11	3727958	9.05E-06
	rs272268660	2	72268660	1.68E-05
	rs292571639	2	92571639	2.82E-05
	rs292571659	2	92571659	2.82E-05
	rs205938814	20	5938814	2.88E-05
	rs274156667	27	4156667	1.69E-05
	rs31633105	3	1633105	3.35E-05
	rs475641139	4	75641139	2.28E-06
	rs475856030	4	75856030	2.96E-05
	rs476878211	4	76878211	5.40E-06
	rs477954086	4	77954086	2.41E-05
	rsz35421630	Z	35421630	2.77E-05