

DOI: 10.3724/SP.J.1005.2012.01271

## 大白猪和通城猪全基因组选择性清扫分析

李秀领<sup>1</sup>, 杨松柏<sup>1</sup>, 唐中林<sup>2</sup>, 李奎<sup>2</sup>, 刘榜<sup>1</sup>, 樊斌<sup>1</sup>

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. 中国农业科学院北京畜牧兽医研究所, 北京 100193

**摘要:** 长期的人工选择使猪的生产性能得到显著提高, 与选择相关的基因组区域也随之发生特定遗传变异表征(选择信号)。不同类型品种所受到选择强度不一, 选择信号亦不相同, 选择性清扫分析已逐渐成为选择信号的主要检测手段。文章基于商用型大白猪(n=45)和地方猪品种通城猪(n=45)的猪 60K SNP 芯片分型数据, 借助遗传分化系数  $F_{st}$  法进行选择信号检测分析。利用 gPLINK 软件设定质控标准, 共计 34 304 个 SNPs 被筛选出用于统计分析。使用 Genepop 软件包计算两个猪品种之间的遗传分化参数  $F_{st}$ , 所得  $F_{st}$  平均值为 0.3209。选取  $F_{st} > 0.7036$ (即占总  $F_{st}$  值数目的 1%), 共计 344 个 SNPs 被选择出来。SNP 位置注释显示这些位点涉及到 79 个候选基因(*Sus scrofa* Build 9)。利用在线软件 Ingenuity Pathway Analysis 对候选基因的生物通路进行网络分析, 发现它们多与生长繁殖及免疫应答有关, 如 *NCOA6*, *ERBB4*, *RUNX2* 和 *APOB* 等基因。研究结果为进行猪产肉、抗病等性状候选基因和致因突变深入挖掘提供了有益参考。

**关键词:** 选择性清扫; 选择信号; 遗传分化系数  $F_{st}$ ; 网络分析; 猪

## Genome-wide selective sweep analysis on Large White and Tongcheng pigs

LI Xiu-Ling<sup>1</sup>, YANG Song-Bai<sup>1</sup>, TANG Zhong-Lin<sup>2</sup>, LI Kui<sup>2</sup>, LIU Bang<sup>1</sup>, FAN Bin<sup>1</sup>

1. Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China;
2. Institute of Animal Science and Veterinary Medicine, Chinese Academy of Agricultural Sciences, Beijing 100193, China

**Abstract:** The production performance of pigs has been significantly improved due to long-term artificial selection, and the specific variation characterizations (selection signatures) emerged from the selected genome regions. Different types of breeds are subjected to different selection intensities and had different selection signatures. Selective sweep analysis is one of major methods to detect the selection signatures. In this study, based on the 60K BeadChip genotyping data of both commercial Large White (n=45) and local Tongcheng pigs (n=45), genetic differentiation coefficient  $F_{st}$  was applied to detect the selection signatures. Using gPLINK software to set quality control standards, a total of 34 304 SNPs were se-

收稿日期: 2012-06-18; 修回日期: 2012-08-17

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(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: 李秀领, 硕士研究生, 专业方向: 动物遗传育种。E-mail: lixiuling\_87@163.com

通讯作者: 樊斌, 博士, 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 03:16:49

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20120910.1516.006.html>

lected for statistical analysis.  $F_{st}$  values between two breeds were estimated with Genepop package and the average  $F_{st}$  value was 0.3209. Setting  $F_{st} > 0.7036$  (1% of total number of  $F_{st}$  values) as selection threshold, 344 SNPs were obtained and SNP location annotation indicated that there were 79 candidate genes (*Sus scrofa* Build 9). Furthermore, network analysis was performed using Ingenuity Pathway Analysis and the preliminary results suggested that most genes were involved in growth, reproduction, and immune response, such as *NCOA6*, *ERBB4*, *RUNX2*, and *APOB* genes. The findings from this study will contribute to further identification of candidate genes and causal mutations implying for meat production and disease resistance in pig.

**Keywords:** selective sweeps; selection signature; genetic differentiation coefficient  $F_{st}$ ; network analysis; pig

猪作为六畜之首,在人们生活当中占有举足轻重地位。人类早在新石器时期就驯化了猪,距今已有 9000 多年历史<sup>[1]</sup>。家猪由欧亚野猪驯化而来,但与野猪相比,其无论从外表特征、生活习性或是生产性能等各方面均表现出较大差异。由于环境差异和人类有目的对其进行持续定向强烈选择,猪的进化过程更为符合人类意愿,使得满足人类需求和喜好的表型或性能在较短时间内固定,特别是繁殖力高、生长速度快等优势性状。通过不断地改良和选育,全世界已经培育出了 300 多个具有各自特色的品种。不同品种在受到人工选择过程中,表型改变是对目标基因强烈定向选择的结果,在分子水平上表现为群体优势基因频率增加。当种群的某一等位基因受到强烈正选择作用时,其附近与之紧密连锁座位上的中性甚至轻微有害变异,将伴随这条等位基因比例的升高而相应提升在种群中比例,这种现象称为选择性清扫(Selective sweep)或搭载效应(Hitchhiking)<sup>[2]</sup>。受到选择的位点或基因组区域,通常具有选择性清扫信号,如核苷酸多态性下降、连锁不平衡值升高等。检验正向选择的检测方法有多种,包括基于种内多态性和种间分歧HKA检验<sup>[3]</sup>;基于突变频率分布的Tajima's D检验<sup>[4]</sup>, Fu & Li's D检验<sup>[5]</sup>和Fay & Wu's H 检验<sup>[6]</sup>;基于非同义突变和同义突变比例的Ka/Ks检验<sup>[7]</sup>和M.K检验<sup>[8]</sup>;基于连锁不平衡和单倍体结构的LRH检测<sup>[9]</sup>;基于群体结构的 $F_{st}$ 检验<sup>[10]</sup>等。

当某一位点或区域受到正向选择时将会偏离中性分布的座位,出现分化现象。群体遗传分化是指物种群体间存在明显等位基因频率差异,涉及到等位基因的空间分布。随机遗传漂变(Genetic drift)和

选择均可导致群体间的遗传分化。遗传分化系数 $F_{st}$ 是检测品种间遗传分化程度的一个重要指标。 $F_{st}$ 统计量由Wright于1943年首次提出<sup>[11]</sup>,被应用于计算物种亚群之间的分化程度,后经过不断补充和校正,在不同物种间得到了很好验证<sup>[12~16]</sup>。本试验尝试采用 $F_{st}$ 方法对商用品种大白猪和地方品种通城猪的全基因组SNP芯片数据进行分析,检测两个品种间的遗传分化程度,初步筛选受选择的位点或基因组区域,并对候选基因的生物化学通路进行网络分析,研究结果将为猪产肉、抗病等性状候选基因和致因突变深入挖掘提供参考材料。

## 1 材料和方法

### 1.1 材料

选取商用纯种大白猪( $n=45$ )和湖北省地方品种通城猪( $n=45$ ),提取DNA,使用Illumina Porcine 60K BeadChip进行分析获得SNP数据,剔除未定位SNPs位点,剩下46 359个SNPs用gPlink<sup>[17]</sup>软件进行质控。质控标准为:平均检出率(Call rate) $> 0.9$ ,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) $\chi^2$ 值 $10e-6$ 。最终共计34 304个SNPs被筛选出来用于分析。

### 1.2 检测方法

利用Genepop v4.1<sup>[18]</sup>软件包对筛选出来的34 304个SNPs进行处理,计算得到两个群体的每一个SNP位点遗传分化系数 $F_{st}$ 值。无偏 $F_{st}$ 的计算公式为:

$$F_{st} = \frac{MSP - MSG}{MSP + (n_c - 1)MSG}$$

其中MSG为群体内观测均方差,

$$MSG = \frac{1}{\sum_{i=1}^s n_i - 1} \sum_{i=1}^s n_i P_{Ai} (1 - P_{Ai})$$

$MSP$  为群体间观测均方差,

$$MSP = \frac{1}{s-1} \sum_{i=1}^s n_i (P_{Ai} - \bar{P}_A)^2$$

$n_c$  是指校正后的群体间平均样本大小,

$$n_c = \frac{1}{s-1} \sum_{i=1}^s n_i - \frac{\sum_{i=1}^s n_i^2}{\sum_{i=1}^s n_i}$$

上述各式中,  $i$  是总亚群数  $S$  的一个群体,  $i=1, 2, \dots, S$ ;  $p_{Ai}$  是第  $i$  个亚群中 SNP 等位基因 A 的频率;  $n_i$  是亚群  $i$  的样本大小;  $\bar{P}_A$  是各群体中  $P_{Ai}$  的加权平均值, 即  $\bar{P}_A = n_i P_{Ai} / \sum_{i=1}^s n_i$  [19]。

一般情况下,  $F_{st}$  的范围在 0~1 之间, 但无偏估计所得  $F_{st}$  值中会出现负值, 这些负值并无生物学解释, 只是抽样误差造成的结果。

### 1.3 候选基因检测与注释

选择 344 个  $F_{st}$  值较高的 SNPs (占总 SNPs 的 1%), 参照 *Sus scrofa* Build 9 基因组注释版本, 并在 Ensembl Genome Browser 数据库 (<http://asia.ensembl.org/index.html>) 和 NCBI 数据库 (<http://www.ncbi.nlm.nih.gov>) 中搜寻 SNPs 所在的基因区域。若候选基因名称在两个数据库中有歧义, 则按照 Ensembl Genome Browser 中所注释名称。

### 1.4 候选基因的生物学术路网络分析

利用 Ingenuity Pathway Analysis (IPA) v7.0 (<http://www.ingenuity.com>) 在线分析工具, 对所得候选基因的生物学术路进行网络分析。IPA 以分子 (基因、蛋白质、代谢物、药物) 间机制为环境来分析资料, 并将分子现象连接到更上层的细胞性或疾病过程。IPA 每一个网络图都含有多个中心基因或分子, 与中心基因/分子联系紧密的基因/分子也会呈现出来。每一个基因均以不同形状表现出来, 不同形状代表不同的基因功能。

## 2 结果与分析

### 2.1 $F_{st}$ 值分布

34 034 个 SNPs 计算得到的  $F_{st}$  值, 其分布如图 1

所示。Wright [20] 认为, 若群体  $F_{st}$  值在 0~0.05 之间, 说明各亚群间不存在分化;  $F_{st}$  值在 0.05~0.15 之间, 为中度分化;  $F_{st}$  值在 0.15~0.25 之间, 则为高度分化。本研究中, 平均  $F_{st}$  值为 0.3209, 说明两个群体已出现明显分化。

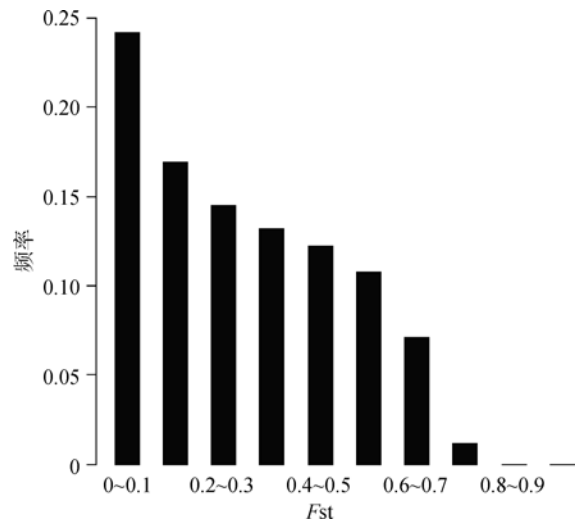


图 1 大白猪和通城猪群体间  $F_{st}$  值分布图

### 2.2 检测候选基因

在 34 034 个  $F_{st}$  值中选取 344 个较大的  $F_{st}$  值, 即  $F_{st} > 0.7036$ , 占总数目的 1%。SNP 注释显示这些位点涉及到 79 个基因 (*Sus scrofa* Build 9), 如表 1 所示。

### 2.3 基因网络分析

为探究候选基因的生物学术功能及与其它基因或分子之间的联系, 对 79 个候选基因做通路分析。选择其中 3 个包含较多数目基因的网络通路图, 共计 61 个基因, 如图 2 所示。

图 2A 中, 该生物学术路包括 26 个候选基因 (*PHF14*, *BPIFB1*, *ITGBL1*, *COL22A1*, *EFHC1*, *CPAMD8*, *KIAA0922*, *NELL1*, *SARNP*, *GPR98*, *PSTPIP1*, *ASRGL1*, *MRPL22*, *ALDH1L2*, *ASH1L*, *AGTPBP1*, *CDK5RAP1*, *PPIL4*, *SETD5*, *BNC2*, *EYA3*, *CKAP4*, *GEMIN5*, *DPYD*, *TTC7A*, *C11orf49*)。以表皮细胞生长因子 (Epidermal growth factor, *EGF*)、雌二醇 ( $\beta$ -estradiol)、糖原合酶激酶 3 $\beta$  (Glycogen synthase kinase 3 beta, *GSK3B*) 为网络中心。*EGF* 促进细胞增殖、分化和生存 [21]。 $\beta$ -estradiol 促进黄体生成素生成, 具有维持妊娠作用。雌激素

表 1 由显著 *Fst* 位点定位到的候选基因

染色体	基因组位置(bp)	SNP 名称	<i>Fst</i> 值	候选基因
1	3953692-4177717	M1GA0000482	0.7155	<i>PDE10A</i>
	17312194-17396895	ALGA0001370	0.7038	<i>PPIL4</i>
	116847457-116898709	H3GA0002594	0.7209	<i>RNF111</i>
	121861605-122421065	ALGA0005718	0.7041	<i>UNC13C</i>
	140989407-141052663	H3GA0002801	0.7404	<i>MEIS2</i>
	214816510-215257042	M1GA0001287	0.7215	<i>BNC2</i>
	288513844-288532154	ASGA0008215	0.7271	<i>DBH</i>
2	7664992-7709096	DBWU0000818	0.7037	<i>ASRGL1</i>
	13427096-13636928	MARC0086081	0.7038	<i>C11orf49</i>
	35177819-35952290	ALGA0013116	0.7246	<i>NELL1</i>
	48159889-48379376	ALGA0013721	0.7065	<i>C19orf29OS</i>
	48179661-48362695	ASGA0010395	0.7065	<i>PIP5K1C</i>
	61413944-61520376	H3GA0006877	0.7176	<i>CPAMD8</i>
	87302674-87461454	DRGA0003183	0.716	<i>GPR98</i>
3	23577500-23625945	M1GA0004189	0.7246	<i>IQCK</i>
	86841621-86978335	MARC0053226	0.718	<i>TTC7A</i>
	93694698-93734479	DBWU0001057	0.7057	<i>THUMPD2</i>
	105344795-105361542	ASGA0016206	0.7113	<i>FAM59B</i>
	109052838-109076900	DIAS0000055	0.74	<i>APOB</i>
4	1712609-2190793	M1GA0005230	0.7118	<i>TRAPPC9</i>
	2882316-3017566	ALGA0022224	0.7209	<i>COL22A1</i>
	93004888-93010516	DBWU0000459	0.7061	<i>USF1</i>
	98260518-98532469	ALGA0026925	0.7061	<i>ASHIL</i>
	98826333-98832103	M1GA0006139	0.7195	<i>DCST1</i>
	115385396-115455624	M1GA0006444	0.7057	<i>KIAA1324</i>
	125046399-125483759	ASGA0022654	0.7288	<i>DPYD</i>
5	12381941-12389594	ALGA0030701	0.7173	<i>CKAP4</i>
	20157789-20203838	ASGA0025069	0.7343	<i>SARNP</i>
	33183472-33417601	ASGA0025405	0.7615	<i>CNOT2</i>
	33526871-33623177	ALGA0031709	0.7269	<i>PTPRB</i>
	67246988-67632770	ALGA0032799	0.7288	<i>SLC2A13</i>
	74641792-74753405	ALGA0033122	0.7243	<i>SLC41A2</i>
	74831523-74862608	ALGA0033127	0.7519	<i>ALDH1L2</i>
6	58467067-58511135	MARC0006446	0.7173	<i>EYA3</i>
	110066302-110217991	ALGA0037458	0.7292	<i>DAB1</i>
7	3930852-4101357	M1GA0009384	0.7133	<i>F13A1</i>
	46909057-47151844	ALGA0040956	0.712	<i>RUNX2</i>
	48483301-48580425	ALGA0041046	0.74	<i>CYP39A1</i>
	62397796-62440268	INRA0026126	0.7057	<i>PSTPIP1</i>
	107168980-107264764	ALGA0043999	0.7412	<i>IFT43</i>
135375349-135450194	ALGA0045948	0.7246	<i>EFHC1</i>	
8	24150956-24416611	MARC0058450	0.7057	<i>KIAA1239</i>
	64733806-64906666	ALGA0113773	0.7176	<i>KIAA0922</i>
	93935642-94260375	ALGA0049141	0.7628	<i>ANK2</i>
	96522141-96538078	ALGA0049219	0.7057	<i>LRIT3</i>

续表 1

染色体	基因组位置(bp)	SNP 名称	Fst 值	候选基因
9	75690930-75793821	DRGA0009538	0.7173	<i>PHF14</i>
	107078630-107089440	ALGA0054746	0.7285	<i>Q8HZW2</i>
10	28377205-28489262	ASGA0047391	0.7417	<i>AGTPBP1</i>
	47425951-47671171	ALGA0059187	0.7061	<i>ARMC4</i>
	51791279-52121327	H3GA0030406	0.7285	<i>PLXDC2</i>
	57573759-57728826	MARC0073944	0.7397	<i>CELF2</i>
11	66145301-66270474	ALGA0063203	0.7061	<i>UGGT2</i>
	70935096-71046518	H3GA0032448	0.7234	<i>ITGBL1</i>
12	5949691-6369897	ALGA0064647	0.7061	<i>SLC39A11</i>
	20579679-20587050	INRA0038788	0.7042	<i>PPP1R1B</i>
13	54816916-55014949	H3GA0036693	0.7111	<i>SETD5</i>
	67052013-67145729	INRA0040659	0.7155	<i>RASA2</i>
14	55696635-55813207	ASGA0063462	0.7285	<i>Q9MYZ4</i>
	72322151-73160656	H3GA0040804	0.7173	<i>CTNNA3</i>
	114230784-114269355	H3GA0042007	0.7343	<i>C10orf28</i>
	135097278-135330360	INRA0047808	0.7628	<i>GRK5</i>
15	15543730-15915160	H3GA0043873	0.7358	<i>TMEM163</i>
	29426065-29427000	M1GA0025948	0.74	<i>TMEM177</i>
	88282774-88325496	H3GA0044724	0.7065	<i>WDR75</i>
	107147960-107599764	ASGA0070420	0.7173	<i>ERBB4</i>
	123644233-123709027	MARC0086651	0.7404	<i>SP140</i>
16	64147594-64250935	MARC0033850	0.7081	<i>B8Y4S5</i>
	65579059-65598814	DBWU0000826	0.7285	<i>MRPL22</i>
	65637588-65684782	SIRI0001384	0.7285	<i>GEMIN5</i>
17	26293939-26554239	ALGA0093960	0.7285	<i>PCSK2</i>
	36504785-36557774	ASGA0076651	0.74	<i>CSNK2A1</i>
	38702396-38729891	H3GA0048846	0.7628	<i>BPIFB1</i>
	38733274-38756096	ALGA0094834	0.7628	<i>A7J153</i>
	38797699-38831219	ASGA0076804	0.7628	<i>CDK5RAP1</i>
	40051472-40146920	ALGA0094899	0.7628	<i>PIGU</i>
	40180021-40233330	ALGA0094904	0.7628	<i>NCOA6</i>
	41896252-42023128	MARC0051907	0.7378	<i>DLGAP4</i>
18	15657789-15820766	DIAS0003365	0.7404	<i>Q6Q7J3</i>
X	120549903-120608925	ASGA0081577	0.7288	<i>MTMR1</i>

作用须经过雌激素受体(*ESR*)介导, 而*ESR*是控制猪产仔性状的一个主效基因, 其基因位点造成的遗传方差大约是总产仔数方差的 10%<sup>[22]</sup>。*GSK3B*功能涉及到能量代谢、神经细胞发育和胚胎模式形成<sup>[23]</sup>。

图 2B 中, 生物学通路包括 21 个候选基因(*DBH*、*NCOA6*、*PCSK2*、*RASA2*、*PIP5K1C*、*ANK2*、

*RUNX2*、*APOB*、*ERBB4*、*GRK5*、*DAB1*、*TRAPPC9*、*KIAA1239*、*UNC13C*、*RNF111*、*CSNK2A1*、*DLGAP4*、*USF1*、*PDE10A*、*PPP1R1B*、*PTPRB*)。以核因子  $\kappa$ -轻链增强激活的 B 细胞 (Nuclear factor kappa-light-chain-enhancer of activated B cells, *NF- $\kappa$ B*), 促分裂素原活化蛋白激酶(Mitogen- activated protein kinase,

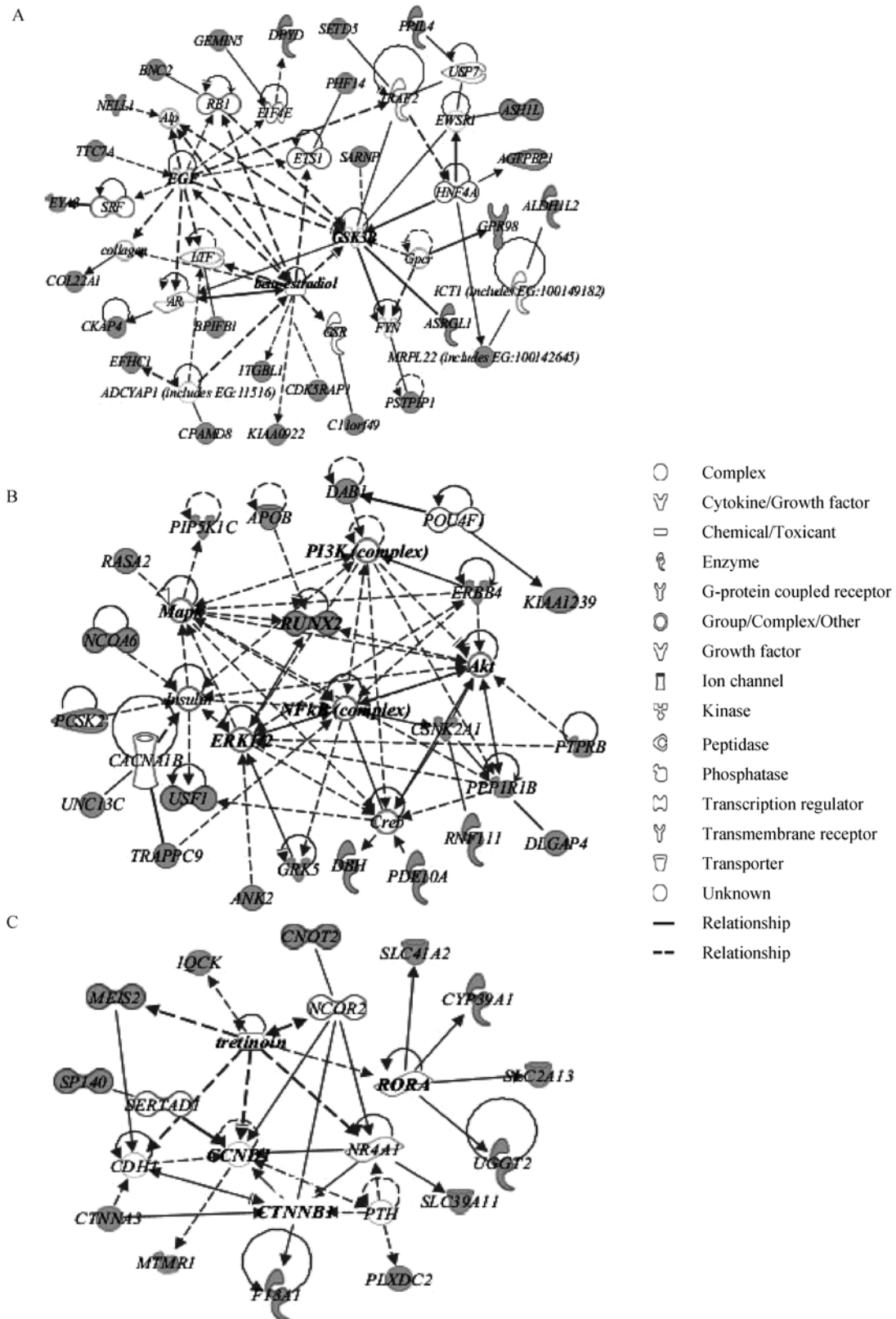


图 2 候选基因生物学通路网络分析图

灰色代表候选基因, 白色代表与之相关的基因。不同图标表示不同生物学功能, 实线表示直接相关, 虚线表示间接相关。关联的网络功能分别为: A: 代谢病, 细胞的功能和维持, 生长和增殖; B: 细胞周期, 细胞发育, 胚胎发育; C: 细胞信号转导, 神经系统发育。

*MAPK*), 细胞外调节蛋白激酶 1/2 (Extracellular signal-regulated kinases 1/2, *ERK1/2*), 蛋白酶 B (Protein Kinase B, *PKB*或*Akt*), 成骨特异性转录因子(Runt-related transcription factor 2, *RUNX2*), 磷脂酰肌醇 3-激酶(Phosphatidylinositol 3-kinases, *PI3K*)为中心基因。*NF- $\kappa$ B*是一个重要转录因子, 调节先天免疫和适应性免疫反应, 还可以通过级联磷酸化, 进入细胞核, 上调基因参与T细胞发育、成熟和增殖, 在调节免疫反应中起着关键作用<sup>[24]</sup>。*MAPK*调节细胞增殖、基因表达、分化、有丝分裂、细胞生存与死亡等生理过程<sup>[25]</sup>。*ERK1/2*(*ERK1* 也称*MAPK3*, *ERK2* 也叫做*MPKAI*)是将信号从细胞表面受体传导至细胞核的关键, 磷酸化的*ERK1/2* 由胞质转位到核内, 进而介导*NF-AT*、*AP-1*、*NF- $\kappa$ B*等转录因子的活化, 且*ERK1/2* 信号通路的激活能促进T淋巴细胞的活化、增殖并抑制其凋亡<sup>[26]</sup>。*Akt*是一种丝氨酸/苏氨酸特异性蛋白激酶, 通过结合和调控下游因子(*NF- $\kappa$ B*、*Bcl-2* 等), 调节细胞生存和代谢, 因此在糖代谢、细胞增殖、迁移和凋亡等过程中起关键作用<sup>[27]</sup>。*RUNX2*是成骨细胞分化相关的一个关键转录因子, 并对*MAPK*通路表达起上调作用<sup>[28]</sup>。*PI3K*参与到细胞增殖、生长、分化和葡萄糖转移等功能的调节, 调节胸腺细胞分化、T细胞活性等<sup>[29]</sup>。

图 2C 中, 生物学通路包括 13 个候选基因(*CTNNA3*、*MEIS2*、*SLC41A2*、*SLC2A13*、*UGGT2*、*CYP39A1*、*SPI40*、*IQCK*、*CNOT2*、*SLC39A11*、*MTMR1*、*PLXDC2*、*F13A1*)。以连环蛋白-1(Catenin beta-1, *CTNNA1*)、视黄酸相关孤立受体 (Retinoic acid-related orphan receptor alpha, *RORA*)、细胞周期素D1(Cyclin D1, *CCND1*)、维生素A酸 (Tretinoin)为中心。*CTNNA1* 基因编码 $\beta$ -连环蛋白<sup>[30]</sup>,  $\beta$ -连环蛋白是钙粘蛋白复合体的一个亚族, 参与到*Wnt*信号通路, 主要功能为介导细胞粘附和参与基因表达。*RORA*参与到细胞应激反应, 还能调节哺乳动物生物钟, 最近研究表明, *RORA*与人的自闭症有关, 雌激素能促进*RORA*基因表达, 雄激素则抑制其表达<sup>[31]</sup>。*CTNND1* 调节细胞新陈代谢、脂肪细胞分化、细胞迁移等生命活动, 其突变、扩增或者超表达都会影响细胞周期进程, 进而增加引发各种肿瘤的发生<sup>[32]</sup>。Tretinoin是维生素A衍生物, 它可以降低表皮细胞凝聚力, 加速表皮新陈代谢, 改善皮肤等功能<sup>[33]</sup>。维生

素A酸也通过促进基因转录来调节胚胎发育, 维生素A可改善胚胎大小整齐度, 提高胚胎同步性, 进而降低胎儿死亡率<sup>[34]</sup>。

19 个基因(*C19orf29OS*、*CBLF2*、*THUMP2*、*FAM59B*、*DCST1*、*KIAAB24*、*IFT43*、*LRIT3*、*Q8HZW2*、*ARMC4*、*Q9MYZ4*、*C10orf28*、*TMEM163*、*TMEM177*、*WDR75*、*B8Y4S5*、*A7J153*、*PIGU*、*Q6Q7J3*)并未包含在上述 3 个通路图中, 一是因为通路图包含所候选基因少, 文中没有选取; 二是因为 IPA 数据库中未录入部分候选基因, 其通路尚未知。

将上述 3 个图合并成为一个基因通路, 除去基本生理信号转导、调节细胞增殖分化等中心基因, 如图 3 所示。发现候选基因中除上述中心基因外, 有些与生长激素受体(Growth hormone receptor, *GHR*)及白细胞介素 1 (Interleukin-1 beta, *IL-1B*)有直接或间接关系。*GHR*是生长激素 (Growth hormone, *GH*)的信号传递者, *GH*参与动物生长过程, 其只受生长激素受体-1 (Growth hormone factor-1, *GHF-1*)的调控, 并通过胰岛素样生长因子(Insulin-like growth factors, *IGFs*)介导而发生作用, 促进动物生长发育, 还影响到肌肉大小和强度及脂肪沉积<sup>[35]</sup>。*IL-1B*是一种促炎症蛋白因子, 具有广泛的免疫调节作用, 促进B细胞功能和T细胞增殖分化。它与猪急性呼吸道疾病—胸膜肺炎和断奶仔猪多系统综合症有关<sup>[36,37]</sup>。此外, 热休克蛋白 70 (Heat shock protein 70, *Hsp70*)可提高细胞应激能力, 主要表现为耐热。*Hsp70* 基因突变还可导致猪肺炎支原体病(也称猪地方性流行性肺炎), 猪肺炎支原体和蓝耳病病毒有协同作用, 增强彼此致病力, 一旦感染肺炎支原体病, 会进一步引发猪的繁殖与呼吸障碍综合征(亦称“蓝耳病”)。

### 3 讨论

大白猪作为欧洲型猪品种, 与中国地方品种通城猪相比, 因所受选择强度和选择方向不同, 两个群体在外貌特征和生产性能等各方面均表现出明显差异, 主要表现为生长速度、繁殖性能及抗病力等不同。本研究中, 两个品种的平均 *F<sub>st</sub>* 值为 0.3209, 说明大白猪和通城猪之间遗传分化度已经很高。在 79 个候选基因中, 有 31 个与猪生长繁殖及抗病性状有关, 暗示在品种驯化和选育过程中所研究两个品

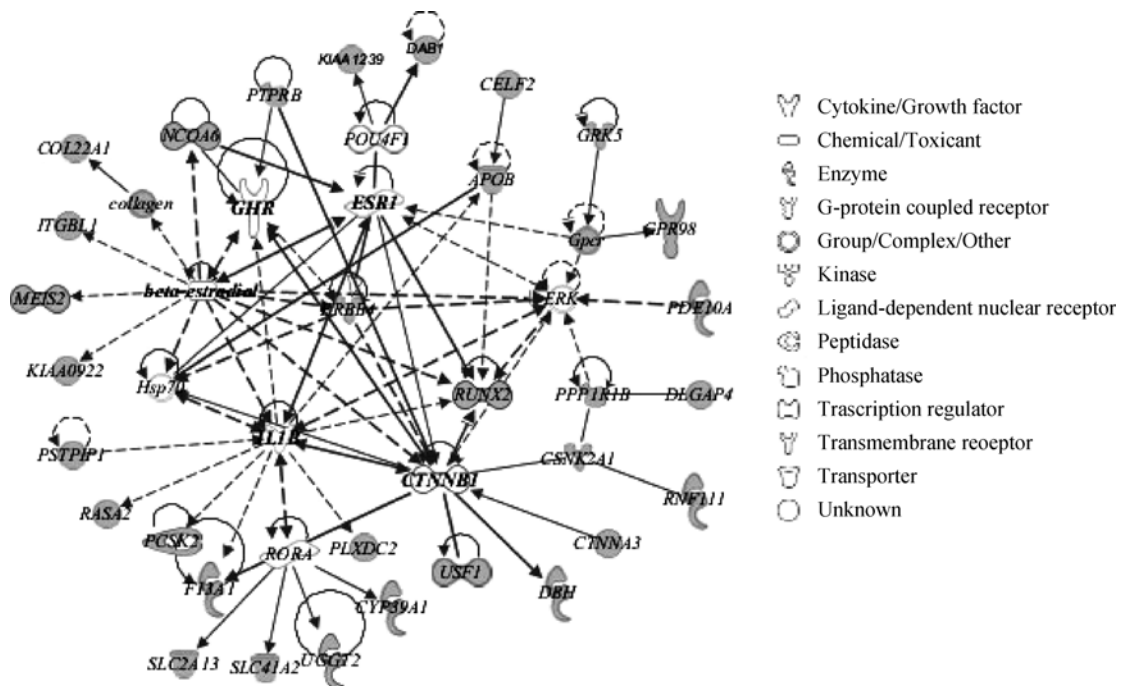


图 3 基因网络图

灰色代表研究中检测到的候选基因，实线表示直接相关，虚线表示间接相关。

种在这些性状上出现差异。

在所检测到基因中，与 $\beta$ -estradiol、*ESR1* 和 *GHR* 等生长繁殖相关的基因有 *ITGBL1*、*KIAA0922*、*NCOA6*、*RUNX2*、*ERBB4*、*MEIS2*、*GPR98*、*GRK5*、*PTRRB*。整合素 $\beta$ 样 1 (Intergrin  $\beta$ - like 1, *ITGBL1*) 调节细胞间粘附和细胞通讯，可能参与母胚相互作用，是子宫内膜接受性的标记分子<sup>[38]</sup>。核受体共激活因子 (Nuclear receptor coactivator, *NCOA6*) 基因编码的蛋白是一种多功能蛋白，涉及到胚胎发育，细胞存活和维持能量平衡，促进生长和发育等生理功能<sup>[39]</sup>。*ERBB4* 是表皮生长因子受体 (Epidermal growth factor receptor, *EGFR*) 亚族中成员，是一种受体酪氨酸激酶。*ERBB4* 与 *EGFR* 结合引发黄体生成素 (Luteinizing hormone, *LH*) 合成，进而促进排卵和雌体生殖器官的发育、维持及重建<sup>[40]</sup>，它还作为神经调节因子 2 (Neuregulin-2) 的配体参与到个体生长发育、细胞分化等生理功能<sup>[41]</sup>，此外还能调节磷脂酰肌醇激酶活性，*PI3K* 最近证明受到雌激素的调控<sup>[42]</sup>。*MEIS2* 基因属于同源结构域包含转录因子亚族的一员，调控胚胎发育和出生后的视网膜<sup>[43]</sup>。

以 *IL1B* 为中心，参与到免疫应答有 *RUNX2*、*APOB*、*PLXDC2*、*PSTPIP1*、*RASA2*、*F13A1*、*PCSK2*

等基因。载脂蛋白 B (Apolipoprotein B, *APOB*) 由肝脏合成，是低密度脂蛋白胆固醇 (Low-density lipoprotein cholesterol, *LDL-C*) 主要结构蛋白，刺激巨噬细胞活性，增强免疫反应，但 *LDL* 的氧化容易产生动脉粥样硬化疾病，产生各种炎症<sup>[44]</sup>。*APOB* 基因突变可引起多种血浆脂蛋白代谢异常，如家族性低 $\beta$ -脂蛋白血症 (Familial hypobetalipoproteinemia, *FHL*)，缺失则容易引发细菌感染<sup>[45]</sup>。脯氨酸-丝氨酸-苏氨酸磷酸酶相互作用蛋白 1 (Proline-serine-threonine phosphatase-interacting protein 1, *PSTPIP1*) 表达主要是造血系统，其突变导致化脓性不育关节炎，脓皮病坏疽和痤疮 (Pyogenic sterile arthritis, pyoderma gangrenosum, and acne, *PAPA*) 综合征，另外它还具有调节 T 细胞活性的功能，在巨噬细胞中也有表达<sup>[46]</sup>。

在两个群体中，本研究检测到与调控上述性状相关的基因，很多基因功能尚待深入研究，尤其是在猪中，这些候选基因的具体生物学功能鲜有报道。生长、繁殖等性状大多受数量性状基因控制，根据 PigQTLdb ([www.animalgenome.org/QTLdb/](http://www.animalgenome.org/QTLdb/)) 数据库信息，结合本文选择出来的分化基因所在染色体位置，可更好了解这些基因所在区域涉及到哪些性状。由结果可以看出，16 号染色体所选基因在 64~66



Mb 之间, 处于 QTL(46.9~92.6 Mb)区域内, 该 QTL 峰值在 64.4 Mb, 涉及性状主要为屠宰率; 17 号染色体所选基因大部分在 36~42 Mb 内, 该区域与 QTL(32~56.4 Mb)重叠, 含有 16 个 QTLs, 包括第 10 肋骨处背膘厚、平均日增重、血液 K 含量和嗜酸性细胞数量等性状。

通城猪毛色为“两头乌, 中间白”, 即: 头颈和臀尾部为黑色, 躯干、四肢为白色, 而大白猪则为全白, 且控制猪毛色基因 *MC1R* 基因已被证明受到人工选择<sup>[47]</sup>, 但在本研究中并未检测出 *MC1R* 在两个品种中出现分化。推测可能原因是, *MC1R* 基因位于猪 6 号染色体 0.256 Mb 处, 但在 *Sus scrofa* Build 9 中, 6 号染色体第一个 SNP 标记则位于 1.06 Mb, 不能检测出该基因。进一步分析发现, 6 号染色体前两个 SNPs 标记的  $F_{st}$  值(0.54, 0.67)远高于其后几个 SNPs 标记  $F_{st}$  值(如图 4 所示), 说明该染色体前端有高度分化区域, 可能包含 *MC1R* 基因。但鉴于本文选择标准, 该基因附近的 SNP 并没有被筛选出。

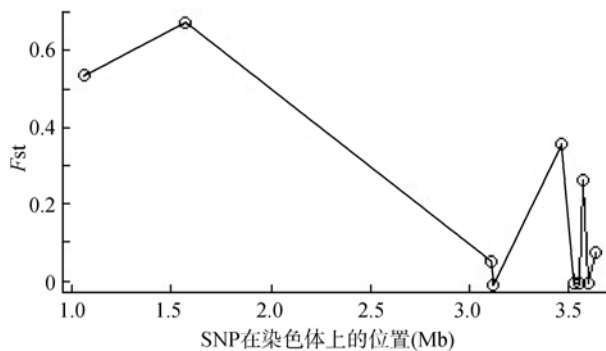


图 4 猪染色体 6 前 10 个 SNPs 的  $F_{st}$  分布图

Akey 等<sup>[48]</sup>基于基因组水平 SNPs 的  $F_{st}$  经验分布, 检测到 174 个候选正选择基因, 这是第一张基于 SNP 数据的人类正选择基因图谱。随着其它物种全基因组 SNP 芯片广泛应用, 探究群体内或亚群间受选择基因或基因组区域已成为一个重要研究热点。在家畜中, 一些已证实了受到选择的基因却没有被检测出来。大部分家养动物的驯化性状是受多基因影响的复杂性状, 受多个基因的共同作用, 且基因互作(上位性)造成基因表达方向和表达程度的不确定, 再加上受到从分子到细胞、组织和个体, 直至生态环境多层次多因素的复杂调控, 难以检测到选择信号。同时, 检测受选择基因的准确性还受很多其

它客观因素的影响。譬如, 如果样本很小, 遗传多样性较低, 则难以区分出正选择或是中性选择; 若突变型在动物驯化之前为中性位点, 只是因人工选择使其固定, 则难以检测到选择<sup>[49]</sup>; 若所选个体之间具有较近亲缘关系, 则所选样本不能代表该群体, 对研究结果亦会造成较大偏差。此外, 基因重组热点区域(hot-spot)的正选择检测具有很高假阳性, 如何消减假阳性造成检测偏差仍待解决; SNP 检测中会遗漏很多低频率位点, 而产生大量中等频率的 SNP, 造成数据测量偏倚(ascertainment bias), 尽管已有很多校正理论和方法, 但误差仍很大; 不同检测方法所得结果也会有差异。

基因组水平上检测选择信号尚处于起始阶段, 已经检测到的基因只是受选择基因/基因组区域中极小部分。随着基因组多态分型技术发展、高密度芯片研制及统计分析方法改进, 更多的选择信号将会被检测出来。本研究利用  $F_{st}$  方法, 对商用型大白猪和地方品种通城猪的分化基因进行检测探索, 尝试检测出影响两品种表型和生产性能差异的基因位点或基因组区域, 为猪产肉、抗病等性状候选基因和致因突变深入挖掘提供基础, 但所得结果有待进一步验证以及对检测出的候选基因功能还须深入研究。

#### 参考文献(References):

- [1] Giuffra E, Kijas JMH, Amarger V, Carlborg Ö, Jeon JT, Andersson L. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, 2000, 154(4): 1785–1791. [DOI](#)
- [2] Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. *Nat Rev Genet*, 2007, 8(11): 857–868. [DOI](#)
- [3] Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data. *Genetics*, 1987, 116(1): 153–159. [DOI](#)
- [4] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 1989, 123(3): 585–595. [DOI](#)
- [5] Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*, 1993, 133(3): 693–709. [DOI](#)
- [6] Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics*, 2000, 155(3): 1405–1413. [DOI](#)
- [7] Li WH, Wu CI, Luo CC. A new method for estimating synonymous and nonsynonymous rates of nucleotide sub-

- stitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol*, 1985, 2(2): 150–174. [DOI](#)
- [8] McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, 1991, 351(6328): 652–654. [DOI](#)
- [9] Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES. Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 2002, 419(6909): 832–837. [DOI](#)
- [10] Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 1973, 74(1): 175–195. [DOI](#)
- [11] Wright S. Isolation by distance. *Genetics*, 1943, 28(2): 114–138. [DOI](#)
- [12] Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ 3rd, Comstock KE, Keller ET, Meisirov JP, von Euler H, Kämpfe O, Hedhammar Å, Lander ES, Andersson G, Andersson L, Lindblad-Toh K. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet*, 2007, 39(11): 1321–1328. [DOI](#)
- [13] Pariset L, Joost S, Marsan PA, Valentini A, Econogene Consortium (EC). Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genet*, 2009, 10: 7. [DOI](#)
- [14] Wahlberg P, Carlborg Ö, Foglio M, Tordoir X, Syvänen AC, Lathrop M, Gut IG, Siegel PB, Andersson L. Genetic analysis of an F<sub>2</sub> intercross between two chicken lines divergently selected for body-weight. *BMC Genomics*, 2009, 10(1): 248. [DOI](#)
- [15] Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, Foulley JL, Gautier M. The genome response to artificial selection: a case study in dairy cattle. *PLoS One*, 2009, 4(8): e6595. [DOI](#)
- [16] Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B; International Sheep Genomics Consortium Members. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*, 2012, 10(2): e1001258. [DOI](#)
- [17] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, 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](#)
- [18] Rousset F. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour*, 2008, 8(1): 103–106. [DOI](#)
- [19] Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*, 1984, 38(6): 1358–1370. [DOI](#)
- [20] Wright S. *Evolution and the Genetics of Populations. Volume 4. Variability Within and Among Natural Populations*. Chicago: University of Chicago Press, 1978. [DOI](#)
- [21] Herbst RS. Review of epidermal growth factor receptor biology. *Int J Radiat Oncol Biol Phys*, 2004, 59(2): 21–26. [DOI](#)
- [22] Short TH, Rothschild MF, Southwood OI, McLaren DG, de Vries A, van der Steen H, Eckardt GR, Tuggle CK, Helm J, Vaske DA, Mileham AJ, Plastow GS. Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. *J Anim Sci*, 1997, 75(12): 3138–3142. [DOI](#)
- [23] Davies G, Jiang WG, Mason MD. The interaction between beta-catenin, *GSK3beta* and *APC* after mitogen induced cell-cell dissociation, and their involvement in signal transduction pathways in prostate cancer. *Int J Oncol*, 2001, 18(4): 843–847. [DOI](#)
- [24] Livolsi A, Busuttill V, Imbert V, Abraham RT, Peyron JF. Tyrosine phosphorylation-dependent activation of *NF-kappa B*. Requirement for p56 *LCK* and *ZAP-70* protein tyrosine kinases. *Eur J Biochem*, 2001, 268(5): 1508–1515. [DOI](#)
- [25] Pearson G, Robinson F, Beers-Gibson T, Xu BE, Karandikar M, Berman K, Cobb MH. Mitogen-activated protein (*MAP*) kinase pathways: regulation and physiological functions. *Endocr Rev*, 2001, 22(2): 153–183. [DOI](#)
- [26] Dong C, Davis RJ, Flavell RA. *Map* kinases in the immune response. *Annu Rev Immunol*, 2002, 20: 55–72. [DOI](#)
- [27] Song G, Ouyang GL, Bao SD. The activation of *Akt/PKB* signaling pathway and cell survival. *J Cell Mol Med*, 2005, 9(1): 59–71. [DOI](#)
- [28] Lee KS, Hong SH, Bae SC. Both the Smad and p38 *MAPK* pathways play a crucial role in *RUNX2* expression following induction by transforming growth factor  $\beta$  and bone morphogenetic protein. *Oncogene*, 2002, 21(47): 7156–7163. [DOI](#)
- [29] Sasaki T, Irie-Sasaki J, Jones RG, Oliveira-dos-Santos AJ, Stanford WL, Bolon B, Wakeham A, Itie A, Bouchard D, Kozieradzki I, Joza N, Mak TW, Ohashi PS, Suzuki A,

- Pernninger JM. Function of *PI3K $\gamma$*  in thymocyte development, T cell activation, and neutrophil migration. *Science*, 2000, 287(5455): 1040–1046. [DOI](#)
- [30] Kraus C, Liehr T, Hülken J, Behrens J, Birchmeier W, Grzeschik KH, Ballhausen WG. Localization of the human  $\beta$ -catenin gene (*CTNNB1*) to 3p21: a region implicated in tumor development. *Genomics*, 1994, 23(1): 272–274. [DOI](#)
- [31] Nguyen A, Rauch TA, Pfeifer GP, Hu VW. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, *RORA*, whose protein product is reduced in autistic brain. *FASEB J*, 2010, 24(8): 3036–3051. [DOI](#)
- [32] Lin HM, Zhao L, Cheng SY. *Cyclin D1* is a ligand-independent corepressor for thyroid hormone receptors. *J Biol Chem*, 2002, 277(32): 28733–28741. [DOI](#)
- [33] Bhawan J, Gonzalez-Serva A, Nehal K, Labadie R, Lufano L, Thorne EG, Gilchrist BA. Effects of tretinoin on photodamaged skin: a histologic study. *Arch Dermatol*, 1991, 127(5): 666–672. [DOI](#)
- [34] Huang FJ, Lin YL. Effects of retinoic acid on pre-implantation embryo development in mice. *Chang Gung Med J*, 2001, 24(11): 681–687. [DOI](#)
- [35] Florini JR, Ewton DZ, Coolican SA. Growth hormone and the Insulin-like growth factor system in myogenesis. *Endocr Rev*, 1996, 17(5): 481–517. [DOI](#)
- [36] Baarsch MJ, Scamurra RW, Burger K, Foss DL, Maheswaran SK, Murtaugh MP. Inflammatory cytokine expression in swine experimentally infected with *Actinobacillus pleuropneumoniae*. *Infect Immun*, 1995, 63(9): 3587–3594. [DOI](#)
- [37] Darwich L, Balasch M, Plnan-Durán J, Segalés J, Domingo M, Mateu E. Cytokine profiles of peripheral blood mononuclear cells from pigs with postweaning multisystemic wasting syndrome in response to mitogen, superantigen or recall viral antigens. *J Gen Virol*, 2003, 84(12): 3453–3457. [DOI](#)
- [38] Schlaepfer DD, Hunter T. Integrin signalling and tyrosine phosphorylation: just the *FAKs*? *Trends Cell Biol*, 1998, 8(4): 151–157. [DOI](#)
- [39] Wang WL, Li Q, Xu J, Cvekl A. Lens fiber cell differentiation and denucleation are disrupted through expression of the N-terminal nuclear receptor box of *Ncoa6* and result in p53-dependent and p53-independent apoptosis. *Mol Biol Cell*, 2010, 21(14): 2453–2468. [DOI](#)
- [40] Park JY, Su YQ, Ariga M, Law E, Jin SLC, Conti M. *EGF*-like growth factors as mediators of *LH* action in the ovulatory follicle. *Science*, 2004, 303(5658): 682–684. [DOI](#)
- [41] Carraway KL, Weber JL, Unger MJ, Ledesma J, Yu N, Gassmann M, Lai C. *Neuregulin-2*, a new ligand of *ErbB3/ErbB4*-receptor tyrosine kinases. *Nature*, 1997, 387(6632): 512–516. [DOI](#)
- [42] Li RW, Capuco AV. Canonical pathways and networks regulated by estrogen in the bovine mammary gland. *Funct Integr Genomics*, 2008, 8(1): 55–68. [DOI](#)
- [43] Bumsted-O'Brien KM, Hendrickson A, Haverkamp S, Ashery-Padan R, Schulte D. Expression of the homeodomain transcription factor *Meis2* in the embryonic and postnatal retina. *J Comp Neurol*, 2007, 505(1): 58–72. [DOI](#)
- [44] Tsimikas S. Oxidized low-density lipoprotein biomarkers in atherosclerosis. *Curr Atheroscler Rep*, 2006, 8(1): 55–61. [DOI](#)
- [45] Peterson MM, Mack JL, Hall PR, Alsup AA, Alexander SM, Sully EK, Sawires YS, Cheung AL, Otto M, Gresham HD. Apolipoprotein B is an innate barrier against invasive *Staphylococcus aureus* infection. *Cell Host Microbe*, 2008, 4(6): 555–566. [DOI](#)
- [46] Cortesio CL, Wernimont SA, Kastner DL, Cooper KM, Huttenlocher A. Impaired podosome formation and invasive migration of macrophages from patients with a *PSTPIP1* mutation and *PAPA* syndrome. *Arthritis Rheum*, 2010, 62(8): 2556–2558. [DOI](#)
- [47] Fang MY, Larson G, Ribeiro HS, Li N, Andersson L. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet*, 2009, 5(1): e1000341. [DOI](#)
- [48] Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res*, 2002, 12(12): 1805–1814. [DOI](#)
- [49] Innan H, Kim Y. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci USA*, 2004, 101(29): 10667–10672. [DOI](#)