

* 学术论文 *

脂肪型和瘦肉型猪肌肉生长和脂肪沉积相关基因的差异表达分析和调控网络构建^{*}李明洲¹ 李学伟^{1**} 朱 砾¹ 滕晓坤² 肖华胜²
帅素容¹ 陈 磊¹ 李 强¹ 郭玉姣¹

1. 四川农业大学动物科技学院肉品分析实验室, 雅安 625014; 2. 生物芯片上海国家工程研究中心, 上海 201203

摘要 骨骼肌细胞和脂肪细胞在生长发育过程中受到的微效多基因间的互作调控机制是决定猪胴体和肉质等相关数量性状的分子基础. 利用包含 140 个与猪肌肉生长和脂肪沉积密切相关基因的 Oligo 功能分类芯片检测了脂肪型的太湖猪和瘦肉型的长白猪在初生及生后 5 月龄间背最长肌中这些基因的动态表达变化, 发现长白猪分别有 18 和 22 个基因, 太湖猪分别有 3 和 7 个基因在月龄间的表达差异达极显著 ($P < 0.01$) 和显著水平 ($P < 0.05$). 聚类结果显示先降后升、逐渐上升和逐渐下降、逐渐上升分别是长白猪和太湖猪在初生至 5 月龄间最具代表性的基因表达模式 ($P < 0.01$), 其中正调控肌纤维生长和脂肪酸合成的基因主要表现为逐渐上调, 而抑制细胞增殖和正调控脂肪酸 β 氧化的基因则主要表现为逐渐下降. 基于动态 Bayes 模型构建的基因调控网络从一定程度上揭示了两品种在肌肉生长和脂肪沉积等生理生化活动方面的明显差异, 可从中挖掘相关性状潜在的关键特征基因. 此外, 对 5 个差异表达基因的荧光定量 RT-PCR 验证显示两种检测方法所获结果的相关系数平均高达 0.876 ± 0.095 . 以上结果筛选出了对于猪胴体和肉质性状可能具有重要影响和具有深入研究价值的基因, 为阐明肌肉生长和脂肪沉积的分子互作机制奠定了基础.

关键词 猪 肌肉生长 脂肪沉积 功能分类芯片 基因差异表达 基因调控网络

肌肉和脂肪的生长发育是一个多基因参与、紧密联系和相互制约的复杂生理过程, 其特异表达基因间相互调控的信息传递网络是决定相关性状的分子基础. 作为重要肉食品来源和模式动物的猪是沉积脂肪能力最强的家畜之一, 近年来对控制猪肌肉生长和脂肪沉积的基因, 特别是与猪肉品质相关基因的研究一直是国内外此类研究的热点. 从动态的观点来看, 猪肉品质主要取决于活体内骨骼肌细胞和脂肪细胞在分化生长相对速度上的竞争, 其平衡

点最终反应了猪的生产类型^[1]. 目前, 猪肉(包括骨骼肌和脂肪)生长发育研究中尚有很多亟待解决的问题, 特别是猪肉性状中具有重要育种意义的某些具体生物学指标的分子遗传基础研究尤为欠缺. 随着“后基因组时代”的到来, 采用系统的观点和方法解析相关基因表达模式的整体特征, 以揭示猪肉生长发育过程中基因表达调控的分子互作机制将是今后猪肉基因组研究中的重要方向^[2]. 我国地方品种太湖猪是典型的脂肪型猪种, 而瘦肉型的长白

2007-08-02 收稿, 2007-08-27 收修改稿

* 教育部长江学者和创新团队发展计划(批准号: IRT0555-6)、教育部高等学校博士学科点专项基金(批准号: 20060626003)、四川省教育厅自然科学重点科研项目(批准号: 2006D004)和四川省教育厅青年基金(批准号: 2006B004)资助项目

** 通信作者: E-mail: lixuewei9125@126.com

猪 (Landrace) 是西方现代猪种, 两品种在胴体和肉质性状的表型上分属两个极端^[3], 是研究肌肉生长和脂肪沉积差异的理想动物模型。

本研究利用 Oligo 基因芯片技术检测猪背最长肌中肌肉生长和脂肪沉积相关基因在不同品种和时间的差异表达规律, 并从多基因互作角度建立了基因调控网络, 其结果有助于利用基因工程技术进行猪胴体和肉质性状的同步改良。

1 材料与方法

1.1 实验动物

分别选用长白猪和太湖猪作为瘦肉型和脂肪型猪种的代表, 设置 6 个等时间点进程(初生, 1, 2, 3, 4 和 5 月龄), 随机分配 4 头(2 ♂, 2 ♀)相同品种的全同胞仔猪到每个月龄内用于取样和性状测定, 共计 48 头。所有试验猪只于同一时间开始组建群体, 在相同条件下饲养, 28 日龄断奶, 自由采食。试验日粮参照美国国家科学研究委员会 (NRC) 猪的营养需要 (1998) 配制, 各阶段日粮消化能均为 $3.40 \text{Mcal} \cdot \text{kg}^{-1}$ ($1 \text{cal} = 4.1868 \text{J}$), 1—2 月龄、2—4 月龄和 4—5 月龄时日粮粗蛋白水平分别为 20.00%, 17.90% 和 15.00%, 可消化赖氨酸含量分别为 1.15%, 0.83% 和 0.66%。实验猪只在达到相应时间点当天进行屠宰并采集倒数 3—4 肋骨间背最长肌样品, 浸泡于 RNAlater (Qiagen 公司) 中, -70°C 保存。

1.2 功能分类芯片的设计制作

结合 Onto-Design 工具^[4] 检索的基因功能分类体系 (gene ontology, GO) 注释信息和 GoPubMed 工具^[5] 查阅的文献资料, 选择 140 个与猪肌肉生长和脂肪沉积密切相关的基因以及阳性和阴性对照各 6 个, 合计 152 个 70 mer 长, 5' 端氨基修饰的探针购自 Qiagen-Operon 公司, 其探针序列来自 “Pig Genome Oligo Microarray Database” (<http://omad.operon.com/pig/query.php>) 数据库。采用 OmniGrid 100 点样仪 (Gene-Machine 公司) 将稀释后的探针点制在 GAPS II Slides 基片 (Corning 公司) 上, 制成 4×164 点的矩阵芯片, 后处理参照其说明。该功能分类芯片的探针序列、基因注解和矩阵信息均已提交到 NCBI 的 GEO 数据库, 登录号为 GPL5171

1.3 RNA 的提取、线性扩增和荧光标记

采用 RNeasy Fibrous Tissue Mini 试剂盒 (Qiagen 公司) 提取和纯化肌肉中的总 RNA, 1% 琼脂糖凝胶电泳检测合格后, 采用 Agilent 2100 分析仪测定总 RNA 浓度并等量混合以构建代表某品种和月龄的 RNA 池。采用 Low RNA Input Linear Amplification Plus-Two Color 试剂盒 (Agilent 公司) 分别进行 Cy3 和 Cy5 通道杂交样本的制备。

1.4 芯片杂交

结合直接环形和配对设计方案^[6,7] 以比较基因在不同品种和月龄间的表达差异, 合计使用 18 张芯片 (图 1)。

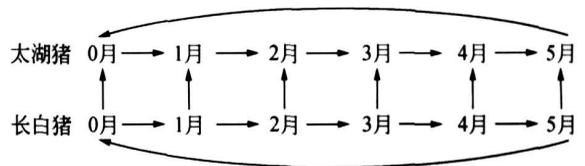


图 1 时间进程和品种比较的实验设计
箭头方向表明荧光染料类型 “Cy5→Cy3”

采用 Pronto Hybridization 试剂盒 (Corning 公司) 进行芯片的杂交和洗涤, 离心干燥后采用 ScanArray 4000 扫描仪 (Perkin-Elmer 公司) 获取荧光图像, 通过 Imagen 7.5 软件 (Bio-Discovery 公司) 进行图像到信号值的转换。

1.5 肌纤维面积和肌内脂肪 (IMF) 含量的测定

采用 TP 1020 型组织脱水机和 RM 2135 型切片 (Leica 公司) 制作背最长肌样品的石蜡切片, H. E 染色后用显微成相系统 (Nikon 公司) 采集代表性区域图像, 通过 Image Pro-Plus 6.0 显微图像分析软件 (Media-Cybernetics 公司) 测定每个样品中 100 根肌纤维的面积。采用 FOSS 2055 型脂肪测定仪 (Foss 公司) 测定 IMF 含量。

1.6 数据分析

原始荧光信号强度值在扣除背景信号值和阴性对照产生的非特异性杂交信号值后同时采用 Composite Loess 和 Print-tip Loess 方法进行片内标准化^[8], 并采用 Pooled Correlation 方法将 4 个重复数据合并^[9], 以上分析由 Webarray 在线分析程序^[10] (<http://bioinformatics.skcc.org/webarray>) 完成。标准化后结果以 M 和 A 值表示, 其中 $M = \ln \text{Cy3} -$

2 结果

2.1 体重、肌纤维面积和 IMF 含量的发育变化

H. E 染色后肌纤维的组织学结构清晰可见, 细胞间光学差异较为明显, 符合图像定量分析的要求

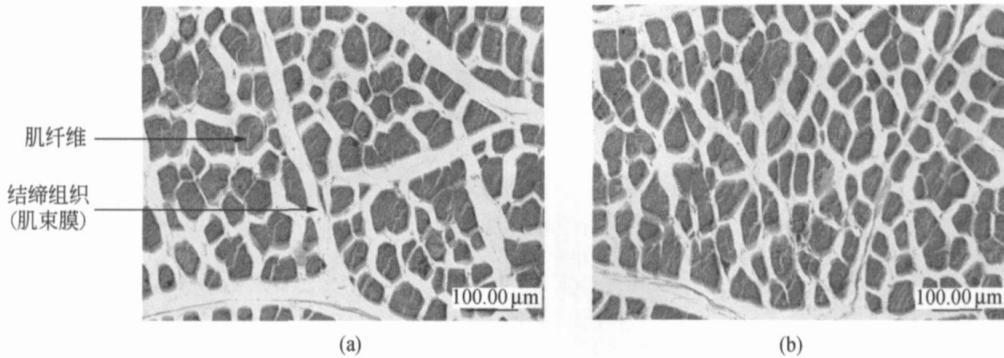


图 2 背最长肌肌纤维切片图(横切)

(a) 长白猪(5月龄); (b) 太湖猪(5月龄)

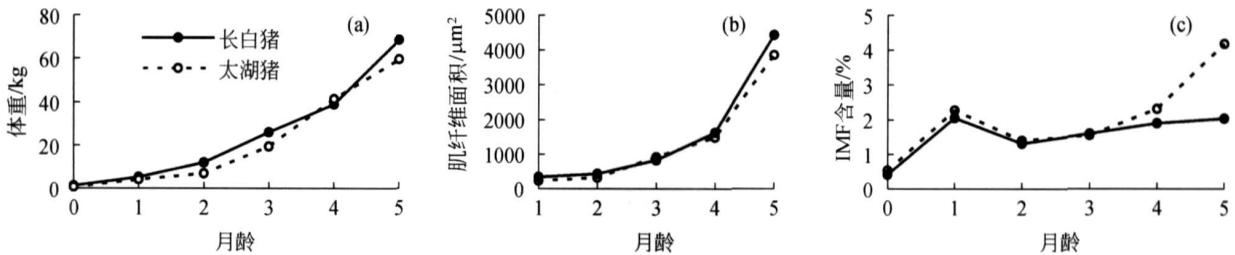


图 3 体重(a)、肌纤维面积(b)和 IMF 含量(c)的发育变化

2.2 基因表达差异的方差和 GCT 分析

综合时间因素的分析结果显示单个基因在品种间的表达差异均没有达到显著水平($P > 0.05$), 分组基因在品种间也均不具有显著性意义($P > 0.05$). 综合品种因素分析结果显示分别有 22 和 21 个基因在月龄间的表达差异为极显著($P < 0.01$)和显著意义($P < 0.05$), “参与糖代谢的酶”和“参与脂肪酸生物合成的酶”这两个基因分组分别具有极显著($P < 0.01$)和显著性意义($P < 0.05$). 品种内的时间效应分析结果显示长白猪分别有 18 和 22 个基因, 太湖猪分别有 3 和 7 个基因在月龄间的表达差异为极显著($P < 0.01$)和显著意义($P < 0.05$), 长白猪中“参与糖代谢的酶”和“参与脂肪酸生物合成的酶”这两个基因分组均具有极显著性意义($P < 0.01$), 而太湖猪的基因分组均不具有显著性意义($P >$

(图 2). 长白猪的体重和肌纤维面积在绝大多数月龄上都略高于太湖猪, 而 IMF 含量的变化趋势却相反. 从 3 月龄开始太湖猪的 IMF 含量明显高于长白猪, 到 5 月龄时太湖猪的 IMF 含量为 4.17% 是长白猪(2.03%)的 2.05 倍(图 3).

0.05)(表 2).

2.3 STEM 聚类分析

长白猪表达谱的第 3, 7 表达模式和太湖猪表达谱的第 2, 7 表达模式分别为其品种内最具代表性的表达模式, 其预期分配基因数(7.7, 7.8 个和 7.7, 8.1 个)与实际分配基因数(18, 18 个和 20, 17 个)之间的差异均达极显著水平($P < 0.01$). 具体讲, 长白猪第 3 表达模式的趋势线为(0, -1, -2, -2, -1, 0)折线, 即从出生后开始下降至 2—3 月间为最低平台期, 然后逐渐升高至初始水平, 第 7 表达模式的趋势线为(0, 1, 2, 3, 4, 5)直线, 即各基因表达量随时间逐渐增加(图 4(a)); 太湖猪第 2 表达模式的趋势线为(0, -1, -2, -3, -4, -5)直线, 即各基因表达量随时间逐渐下降, 第 7 表达模式与长白猪的相同(图 4(b)).

表 2 基因表达差异的方差分析、GCT 分析和表达模式识别^{a)}

序号	基因名称	基因简称	P_1	P_2	P_3	P_4	表达模式	序号	基因名称	基因简称	P_1	P_2	P_3	P_4	表达模式
a 参与脂肪或类固醇代谢的酶和调节蛋白								c 参与糖代谢的酶							
	$P_{E_{\text{minej}}}(n_a=25)$		1.000	0.460	0.730	1.000			$P_{E_{\text{minej}}}(n_c=5)$		0.730	0.002	0.000	0.096	
1	乙酰辅酶 A 酰基转移酶 1	ACAA	0.072	0.035	0.269	0.088	10 10	32	β -烯醇化酶 3	ENO3	0.399	0.331	0.445	0.521	\
2	酰基辅酶 A 脱氢酶, 长链	ACADL	0.409	0.038	0.219	0.147	1 5	33	甘油 3 磷酸脱氢酶 1 (可溶)	GPDI	0.119	0.000	0.000	0.044	7 7
3	酰基辅酶 A 脱氢酶, C-4 至 C-12 直链	ACADM	0.686	0.060	0.086	0.564	3 3	34	乳酸脱氢酶 A	LDHA	0.259	0.000	0.002	0.004	7 7
4	酰基辅酶 A 脱氢酶, C-2 至 C-3 短链	ACADS	0.598	0.346	0.401	0.613	\	35	乳酸脱氢酶 B	LDHB	0.338	0.006	0.041	0.147	2 2
5	酰基辅酶 A 脱氢酶, 短/分支链	ACADSB	0.439	0.174	0.012	0.632	3 2	36	UDP 葡萄糖焦磷酸酶 2	UGP2	0.244	0.003	0.009	0.086	2 2
6	血管生成素样蛋白 4	ANGPTL4	0.119	0.084	0.086	0.086	0 10	d 参与糖生物合成的酶							
7	3 羟基丁酸脱氢酶 2	BDH2	0.066	0.105	0.215	0.512	\		$P_{E_{\text{minej}}}(n_d=2)$		0.890	0.310	0.859	0.875	
8	肉碱脂酰转移酶	CPT1B	0.598	0.290	0.120	0.564	\	37	糖原合成酶 1(肌)	GYS1	0.598	0.142	0.376	0.290	0 2
9	晶状体蛋白 $\lambda 1$	CRY	0.356	0.250	0.471	0.165	1 9	38	磷酸葡萄糖变位酶 1	PGMI	0.478	0.000	0.007	0.028	10 10
10	十二烷烷辅酶 A, δ 异构酶	DCI	0.244	0.006	0.018	0.280	3 3	e 参与脂肪酸生物合成的酶							
11	二酰基甘油酰基转移酶	DGAT	0.598	0.004	0.147	0.089	7 7		$P_{E_{\text{minej}}}(n_e=10)$		0.510	0.040	0.001	1.000	
12	脱氢酶/还原酶(SDR 家族)成员 3	DHRS	0.478	0.258	0.046	0.104	4 10	39	蛋白激酶, AMP 活化, $\alpha 2$ 催化亚基	AMPKA2	0.598	0.003	0.007	0.120	7 7
13	烯酰辅酶 A 水合酶 1, 过氧化物酶体	ECHI	0.598	0.011	0.005	0.150	2 2	40	动力蛋白激活蛋白 6	DCTN6	0.148	0.208	0.246	0.564	\
14	烯酰辅酶 A 水合酶, 短链 1, 线粒体	ECHS1	0.075	0.034	0.046	0.548	2 2	41	异柠檬酸脱氢酶	IDH2	0.075	0.107	0.162	0.564	3 3
15	羟烷基辅酶 A 脱氢酶	HAD	0.598	0.188	0.194	0.564	\	42	脂蛋白脂酶	LPL	0.072	0.042	0.248	0.044	3 6
16	羟烷基辅酶 A 脱氢酶/3 酮乙基辅酶 A 硫解酶/烯酰辅酶 A 水合酶(三功能蛋白), α 亚基	HADHA	0.119	0.022	0.029	0.564	3 2	43	细胞质苹果酸脱氢酶	MDH1	0.119	0.084	0.164	0.541	3 2
17	羟基类固醇(17 β)脱氢酶 4	HSD17B4	0.385	0.059	0.070	0.541	\	44	线粒体苹果酸脱氢酶	MDH2	0.627	0.090	0.345	0.281	0 2
18	羟烷基辅酶 A 脱氢酶/3 酮乙基辅酶 A 硫解酶/烯酰辅酶 A 水合酶(三功能蛋白), β 亚基	LCTH10	0.461	0.015	0.029	0.337	2 2	45	苹果酸酶	MEI	0.437	0.138	0.043	0.078	7 6
19	激素敏感脂肪酶	LIPE	0.232	0.188	0.326	0.281	9 2	46	NADH 脱氢酶 1, α/β 复合体亚基, 1, 8kDa	NDUFA1	0.598	0.133	0.332	0.487	3 3
20	丙酰辅酶 A 羧化酶 β 肽	PCCB	0.457	0.003	0.024	0.086	7 7	47	6 磷酸葡萄糖酸脱氢酶	PGD	0.686	0.034	0.007	0.104	7 5
21	磷脂酶 D2	PLD2	0.059	0.003	0.012	0.127	10 10	48	丙糖磷酸异构酶 1	TP11	0.338	0.000	0.000	0.004	7 7
22	磷脂酸性磷酸酶 2A	PPAP2A	0.598	0.175	0.317	0.341	3 7	f 参与脂肪酸运输的结合蛋白和转运蛋白							
23	过氧化物酶体增殖活化受体 γ 辅激活蛋白 1 α	PPARGC-1	0.119	0.001	0.003	0.078	2 2		$P_{E_{\text{minej}}}(n_f=8)$		0.600	0.690	0.942	0.836	
24	过氧化还原酶 6	PRDX6	0.449	0.042	0.219	0.212	3 3	49	脂肪分化相关蛋白	ADRP	0.598	0.018	0.005	0.044	2 3
25	固醇 C4 甲基氧化酶样	SCMOL	0.598	0.306	0.207	0.537	0 7	50	载脂蛋白 A-II	APOA2	0.072	0.367	0.187	0.564	10 3
b 参与蛋白水解的酶和调节蛋白								51	小窝蛋白-1	CAV1	0.075	0.317	0.437	0.548	\
	$P_{E_{\text{minej}}}(n_b=6)$		0.640	0.880	0.800	0.957		52	心脏型脂肪酸结合蛋白	FABP3	0.153	0.003	0.018	0.078	0 2
26	钙蛋白酶-II	CAPN2	0.461	0.185	0.332	0.537	3 9	53	脂肪细胞型脂肪酸结合蛋白	FABP4	0.598	0.385	0.065	0.104	2 6
27	骨骼肌特异性钙蛋白酶	CAPN3	0.612	0.400	0.437	0.564	\	54	热休克蛋白 90kDa $\beta 1$	HSP90B1	0.598	0.142	0.274	0.352	9 7
28	钙蛋白酶, 小亚基 1	CAPNS1	0.570	0.162	0.187	0.618	10 5	55	解藕联蛋白 2	UCP2	0.119	0.035	0.322	0.150	8 8
29	胱抑素 B	CSTB	0.059	0.119	0.048	0.602	2 1	56	解藕联蛋白 3	UCP3	0.439	0.004	0.029	0.083	8 2
30	组织蛋白酶 D	CTSD	0.600	0.046	0.079	0.281	10 10	g 调控脂肪细胞分化的转录因子							
31	组织蛋白酶 L	CTSL2	0.598	0.022	0.082	0.137	3 3		$P_{E_{\text{minej}}}(n_g=2)$		0.890	0.840	0.873	0.888	
								57	固醇调节元件结合蛋白-1c	ADD1	0.075	0.148	0.067	0.637	5 1
								58	CAAT 增强子结合蛋白 δ	CEBPD	0.075	0.017	0.067	0.104	7 10
								h 生长激素作用的靶基因							
	$P_{E_{\text{minej}}}(n_h=5)$		1.000	0.330	0.677	1.000		59	生长激素受体	GHR	0.119	0.023	0.099	0.150	9 3
60	胰岛素样生长因子受体-1	IGF1R	0.072	0.138	0.186	0.564	10 10	60	胰岛素样生长因子受体-1	IGF1R	0.072	0.138	0.186	0.564	10 10
61	胰岛素样生长因子-II	IGF2	0.463	0.003	0.018	0.104	2 2	61	胰岛素样生长因子-II	IGF2	0.463	0.003	0.018	0.104	2 2
62	胰岛素样生长因子结合蛋白 2	IGFBP2	0.259	0.036	0.099	0.137	2 2	62	胰岛素样生长因子结合蛋白 2	IGFBP2	0.259	0.036	0.099	0.137	2 2
63	胰岛素样生长因子结合蛋白 5	IGFBP5	0.479	0.003	0.023	0.044	6 7	63	胰岛素样生长因子结合蛋白 5	IGFBP5	0.479	0.003	0.023	0.044	6 7
								i 生肌决定因子							
	$P_{E_{\text{minej}}}(n_i=10)$		0.570	0.670	0.674	0.757									

续表

序号	基因名称	基因简称	P_1	P_2	P_3	P_4	表达模式	序号	基因名称	基因简称	P_1	P_2	P_3	P_4	表达模式
64	肌肉特异性 LIM 蛋白	<i>CSRP3</i>	0.600	0.003	0.000	0.002	3 2	81	肌红蛋白	<i>MB</i>	0.119	0.000	0.007	0.044	7 7
65	四和 a 半 LIM 域 1	<i>FHLIC</i>	0.598	0.022	0.033	0.324	3 2	82	肌侵蛋白	<i>MTPN</i>	0.148	0.104	0.029	0.644	10 8
66	四和 a 半 LIM 域 2	<i>FHL3</i>	0.120	0.099	0.055	0.281	7 3	83	肌球蛋白重链 2a	<i>MYH2A</i>	0.751	0.185	0.001	0.044	1 2
67	肌肉生长抑制素基因	<i>GDF8</i>	0.409	0.026	0.007	0.337	6 7	84	肌球蛋白重链 2b	<i>MYH2B</i>	0.702	0.022	0.131	0.137	7 10
68	肌细胞特异性增强子结合因子 2A 型	<i>MEF2A</i>	0.237	0.051	0.074	0.564	0 4	85	肌球蛋白重链(慢型)	<i>MYH7</i>	0.598	0.142	0.479	0.280	\
69	肌细胞特异性增强子结合因子 2B 型	<i>MEF2B</i>	0.598	0.189	0.497	0.212	8 7	86	肌球蛋白轻链	<i>MYLI</i>	0.724	0.201	0.493	0.521	\
70	肌细胞特异性增强子结合因子 2C 型	<i>MEF2C</i>	0.598	0.156	0.042	0.147	0 1	87	兰尼定受体 1	<i>RYR1</i>	0.153	0.003	0.015	0.078	7 10
71	肌细胞生成素	<i>MYOG</i>	0.072	0.317	0.276	0.645	\	88	沉默调节蛋白	<i>SIRT2</i>	0.099	0.135	0.284	0.564	\
72	转化生长因子 β_2	<i>TGFB2</i>	0.439	0.006	0.018	0.137	7 1	89	肌联蛋白帽	<i>TCAP</i>	0.512	0.059	0.269	0.318	3 3
73	转化生长因子 β_3	<i>TGFB3</i>	0.153	0.119	0.346	0.212	6 0	90	肌钙蛋白-II	<i>TNNI1</i>	0.399	0.133	0.090	0.613	9 8
j	影响骨骼肌纤维生长发育的酶、调节蛋白和结合蛋白							91	肌联蛋白	<i>TTN</i>	0.439	0.107	0.015	0.637	7 10
	$P_{E_{\text{min}i,j}}(n_j=18)$		0.560	0.670	0.698	0.728		k	脂肪细胞因子及受体		0.590	0.990	0.869	0.975	
74	肌动蛋白 $\alpha 1$	<i>ACTA1</i>	0.259	0.358	0.284	0.341	\	92	脂联素受体 2	<i>ADIPOR2</i>	0.598	0.424	0.298	0.613	10 4
75	肌动蛋白相关蛋白 2/3 复合体, 亚基 3	<i>ARPC3</i>	0.461	0.074	0.034	0.548	\	93	抵抗素	<i>RETN</i>	0.072	0.239	0.086	0.644	10 0
76	肌动蛋白相关蛋白 2/3 复合体, 亚基 5	<i>ARPC5</i>	0.755	0.003	0.007	0.092	2 1	94	肿瘤坏死因子受体超家族, 成员 1A	<i>TNFRSF1A</i>	0.749	0.311	0.346	0.564	3 4
77	铁蛋白, 重肽 1	<i>FTH1</i>	0.570	0.078	0.344	0.341	3 2	1	其他		0.870	0.660	0.530	1.000	
78	肝素结合 EGF 样生长因子	<i>HBEGF</i>	0.623	0.231	0.007	0.089	7 5		$P_{E_{\text{min}i,j}}(n_i=5)$		0.870	0.660	0.530	1.000	
79	组蛋白脱乙酰酶 5	<i>HDAC5</i>	0.119	0.207	0.390	0.132	6 7	95	腺苷一磷酸脱氢酶 1	<i>AMPD1</i>	0.119	0.003	0.000	0.089	7 7
80	凝溶胶蛋白(芬兰型)	<i>LOC396874</i>	0.119	0.034	0.065	0.089	7 10	96	结缔组织生长因子	<i>CTGF</i>	0.598	0.082	0.009	0.058	3 7
								97	细胞色素 b5A(微粒体)	<i>CYB5</i>	0.075	0.011	0.062	0.165	10 7
								98	细胞色素 P450 家族 2, 亚家族 E, 肽 1	<i>CYP2E1</i>	0.072	0.148	0.180	0.564	2 3
								99	翻译控制肿瘤蛋白	<i>TCTP</i>	0.244	0.280	0.469	0.564	\

a) P_1, P_2, P_3 和 P_4 分别表示综合时间因素的品种效应, 综合品种因素的时间效应, 长白猪的时间效应和太湖猪的时间效应; $P_{E_{\text{min}i,j}}$ 表示该组基因在相应 GCT 分析中得到的 Benjamini-Hochberg FDR 校正后的 P 值, n 表示该分组所包括的基因数目; 表达模式为该基因在品种内不同月龄间的 STEM 聚类中被分配到的表达模式序号(图 4), 按品种顺序表示为(长白|太湖), 其中“\”表示该基因在聚类前已被过滤

2.4 基于 DBN 模型构建基因调控网络

长白猪和太湖猪表达谱数据构建的基因调控网络分别包含 66 和 65 个节点(图 5)。长白猪的网络中 *PPARGC-1* 基因作为主要的父节点直接调控 *ADD1*, *IGF2*, *UCP2* 和 *UCP3* 基因, 而 *UCP3* 和 *UGP2* 作为最为明显的子节点分别受到 19 和 23 个基因的直接调控(图 5(a))。在太湖猪的网络中 *CSRP3* 和 *DCI* 基因作为主要的父节点, *CSRP3* 基因直接调控 *LPL*, *TTN* 和 *UCP3* 基因, *DCI* 基因直接调控 *ARPC5*, *CYB5* 和 *IGF2* 基因, *UGP2*, *TTN* 和 *UCP3* 作为最为明显的子节点, 分别受到 9, 15 和 30 个基因的直接调控(图 5(b))。

2.5 荧光定量 RT-PCR 验证

所选 5 个差异表达基因的荧光定量 RT-PCR 结果与芯片检测结果在品种内不同月龄间表达变化的相关系数平均为 0.876 ± 0.095 (表 3)。

表 3 基因芯片和荧光定量 RT-PCR 结果的相关分析

基因简称	品种	Pearson 相关系数	P 值
<i>IGF-II</i>	长白	0.895	0.016
	太湖	0.976	< 0.001
<i>IGFRI</i>	长白	0.851	0.032
	太湖	0.929	0.007
<i>IGFBP5</i>	长白	0.895	0.016
	太湖	0.965	0.002
<i>PPARGC-1</i>	长白	0.754	0.083
	太湖	0.980	< 0.001
<i>RYR1</i>	长白	0.708	0.116
	太湖	0.804	0.054

3 讨论

研究显示在体重和肌纤维面积指标上长白猪均较同月龄的太湖猪高, 但太湖猪的 IMF 含量却明显高于长白猪。这与太湖猪早熟早肥, IMF 含量较高, 肉质优良, 但生长速度慢, 胴体瘦肉率低

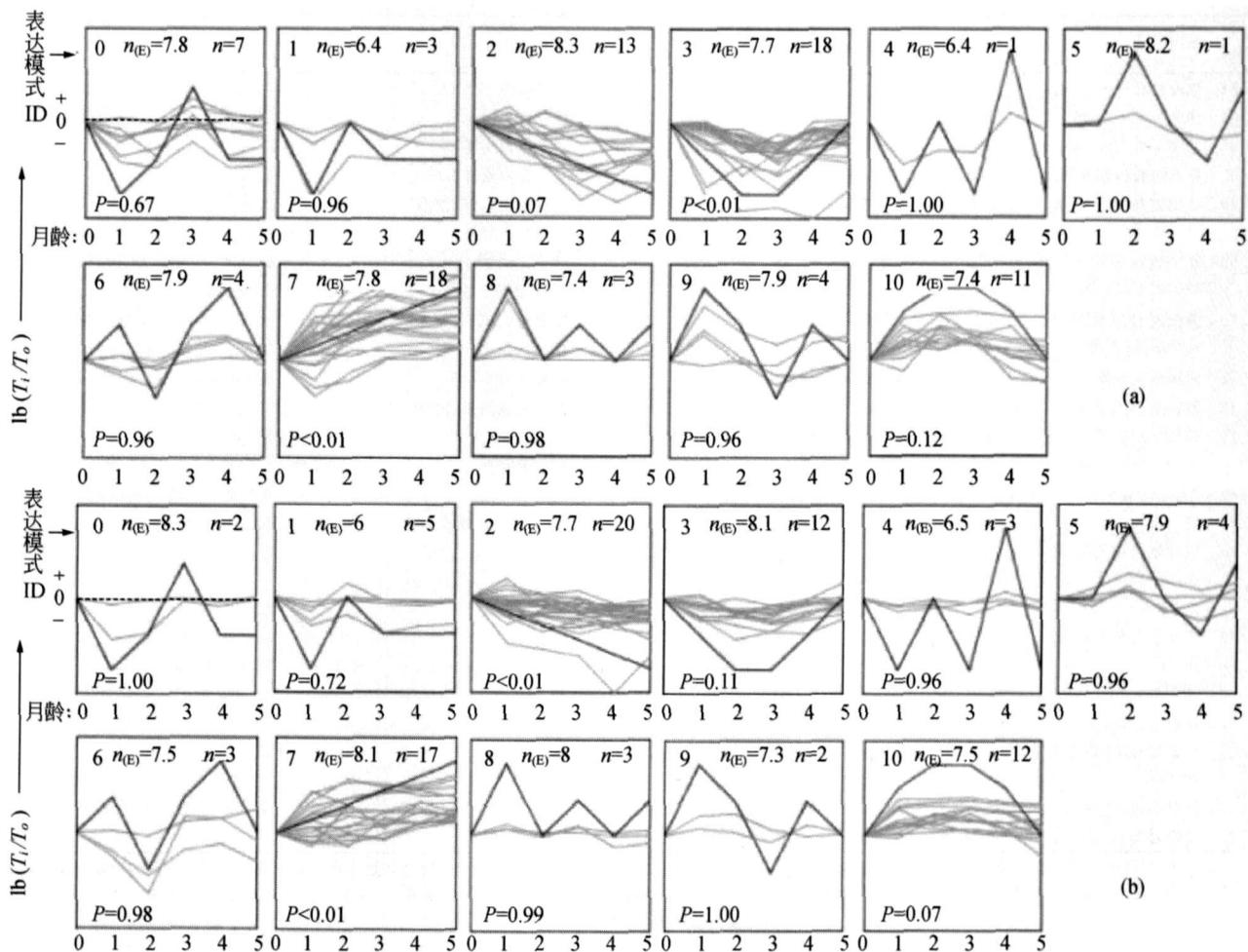
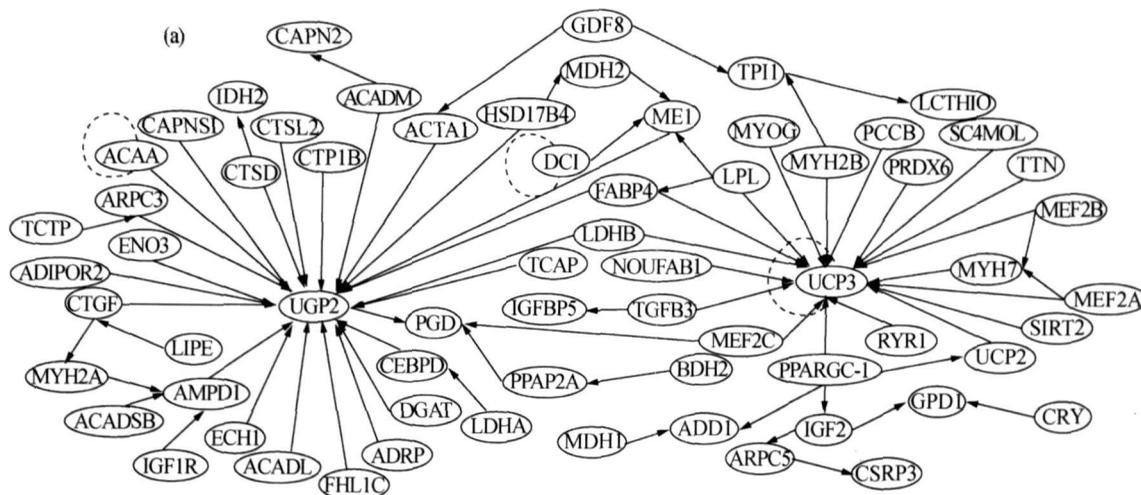


图 4 品种内不同月龄间基因的表达模式

(a) 长白猪; (b) 太湖猪; 方块内黑色粗线条为表达模式的趋势线, 灰色线条为基因的表达变化趋势, $n_{(E)}$ 和 n 分别表示表达模式的预期和实际分配基因数, P 值为 $n_{(E)}$ 与 n 经 MHT 分析得到的 FDR 校正 P 值



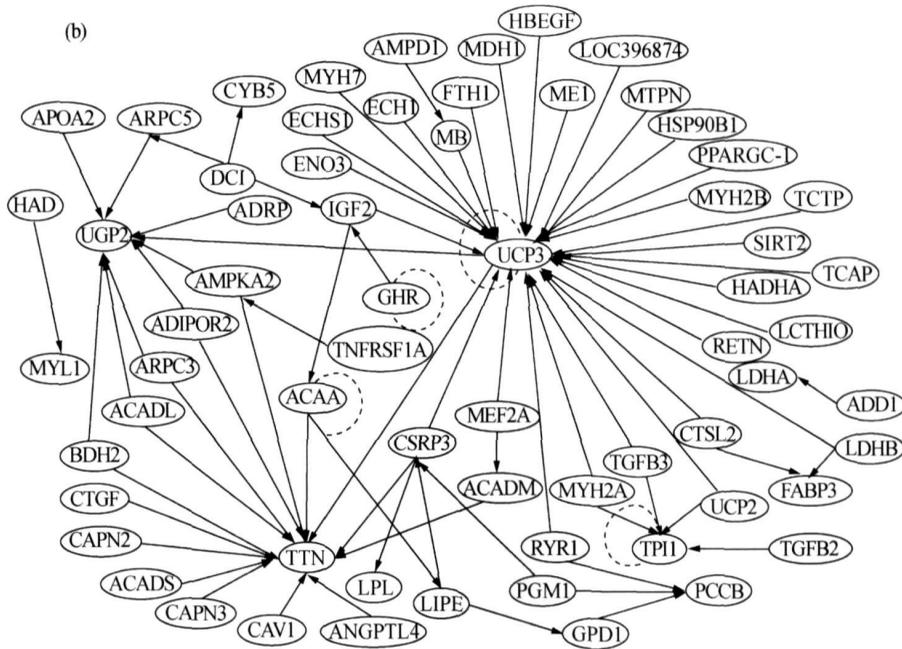


图 5 肌肉生长和脂肪沉积关键基因的调控网络

(a) 长白猪; (b) 太湖猪. 直线箭头方向为从父节点到子节点, 环形箭头表示自我调控

(38.8%—45%), 而长白猪 IMF 含量较低, 肉质较差, 但生长发育快, 饲料利用率高, 胴体瘦肉率高(可高达 65%)的种质特性相一致^[3].

基于 EB 算法的显著性检验会产生 5 种 P 值: (i) F 分布的 P 值; (ii) FDR 校正后的 P 值; (iii) 置换 P 值; (iv) FDR 校正后的置换 P 值和 (v) 族误差率(FWER)校正后的置换 P 值^[23]. Allison 等^[24]和 Yang 等^[25]均建议可主要参考“FDR 校正后的置换 P 值”, 但由于该 P 值尺度较为严格, 故应根据实际情况结合其他 P 值进行差异表达基因的选取. 研究结果中“FDR 校正后的置换 P 值”没有筛选出品种间表达差异达显著水平的基因, 而“置换 P 值”却显示品种间分别有 10 和 19 个基因的表达差异达极显著 ($P < 0.01$) 和显著水平 ($P < 0.05$). 但从差异表达的程度上看, 不同类型 P 值下基因的先后次序是几乎一致的, 所以本文仍采用“FDR 校正后的置换 P 值”来解释试验结果. 另外, GCT 分析的实质是以单个基因在方差分析中的 P 值作为衡量该基因表达量在不同效应检验中变异程度的指标, 从而对分组内各基因间表达变异程度的差异性进行显著性检验, $P_{\text{Ermine}} < 0.05$ 或 0.01 并不代表该分组内各基因的表达差异达到或大部分

达到显著或极显著水平, 表示该分组内各基因表达量变异程度间的差异达显著或极显著水平.

在初生至 5 月龄这个重要的生长发育阶段中, 长白猪分别有 18 和 22 个基因, 而太湖猪只有 3 和 7 个基因的表达差异为极显著 ($P < 0.01$) 和显著水平 ($P < 0.05$) (表 2). 这与长白猪生长发育速度较快, 相关生理生化代谢活动较为旺盛, 而太湖猪生长发育缓慢, 代谢相对较为平缓的种质特性相符. 提示这些基因受到了特殊的调控, 在肌肉中的表达水平伴随各自特异的功能在不同时期表现出明显的起伏变化. 长白猪表达差异极显著的 18 个基因中包含了太湖猪表达差异极显著的 3 个基因 ($CSR P3$, $LDHA$ 和 $TPII$), 尚已有研究证明对于猪的胴体和肉质性状有重要影响的基因主要有 $ADRP$, $AMPK A2$, $ARPC5$, $CSR P3$, $CTGF$, $GDF8$, MB , $MYH2A$ 和 $PPARGC-1$ 这 9 个, 而 $AMPD1$, $ECHI$, $GPD1$, $HBEGF$, $LDHA$, PGD , $PGMI$, $TPII$ 和 $UGP2$ 这 9 个基因尚未见相关报道, 但已知其功能均与调控细胞生长和脂肪酸生物合成及氧化相关, 提示可作为候选基因进行深入研究. 值得注意的是其中的 $AMPD1$ 基因, 其主要作用是催化腺苷一磷酸 (AMP) 脱氨生成肌苷酸 (IMP)^[26], 而

IMP 在水和脂肪中加热能产生明显的肉鲜味, 其在肉品中的含量已被作为衡量肉质的一个重要指标. 但 Tikka 等^[27]发现 IMP 在去磷酸化和水解后产生的次黄嘌呤(HX)却能直接导致猪肉(汤)的苦味和咸涩味, 虽然高 pH 值(> 5.7)的猪肉 IMP 含量较高, 但其 HX 的含量仍比较低 pH 值的猪肉高, 因此认为一味强调提高 IMP 含量是不合适的. Stratil 等^[28]克隆测序得到猪 *AMPDI* 基因 1323 bp 的 EST 序列并将其定位于 4 号染色体区域, 而该区域的其他基因均与 IMF 含量、背膘厚和眼肌面积等重要经济性状相关. 目前 *AMPDI* 基因的相关研究主要集中在鱼和家禽(鸡)上, 建议应给以更多关注.

“参与脂肪酸生物合成的酶”、“生肌决定因子”和“影响骨骼肌纤维生长发育的酶、调节蛋白和结合蛋白”这 3 个基因分组在品种效应的 GCT 分析中虽未达显著水平($P > 0.05$), 但其基因表达变异程度的差异在品种间要明显, 位居前 3 位且包含较多基因(共 38 个). 提示两猪种在脂肪沉积能力、肌纤维数量和生长速度上明显的遗传种质特性差异与这些基因的差异表达规律密切相关.

研究聚类分析所采用的 STEM 算法由 Ernst 等^[29]提出, 是目前能够有效聚类时间序列试验数据的 3 种算法(动态 Bayes^[30]、隐 Markov 模型^[31]和 STEM 聚类)中唯一针对小于 8 个时间点的短时间序列试验设计的, 其结果显示品种间的主要差别在于先降后升的第 3 表达模式中长白猪包含 18 个基因($P < 0.01$), 而太湖猪只包含 12 个基因($P = 0.11$). 提示可能由于出生后环境和营养的应激, 许多参与肌肉生长和脂肪沉积基因的表达量逐渐下降, 但伴随动物个体的逐渐适应和生长发育的需要, 其表达量又逐渐上调, 而长白猪生长发育速度较太湖猪快, 故相关基因表达谱更为活跃. 长白猪和太湖猪的第 2 表达模式分别包含 13 个($P = 0.07$)和 20 个($P < 0.01$)基因, 第 7 表达模式分别包含 18 个($P < 0.01$)和 17 个($P < 0.01$)基因, 均可视为品种内最具代表性的表达模式. 这两个表达模式不仅包含了较多的基因, 且第 2 表达模式为直线下降与第 7 为直线上升恰好相反(图 4). 通过对表达模式内的基因功能进行分类, 发现正调控肌纤维生长和脂肪酸合成的基因大多属于第 7 表达模式, 而抑制细胞增殖和正调控脂肪酸 β 氧化的基因大多属于第 2 表达

模式. 提示伴随动物个体的不断发育成熟, 肌肉生长速度和脂肪沉积强度的不断增加, 促进肌纤维增殖和脂肪酸合成基因的表达水平也在逐渐上升, 而抑制肌纤维增殖和促进脂肪消耗代谢基因的表达水平则逐渐下降.

虽然未知功能的基因与其相同表达模式中已知功能的基因在功能上可能具有相似之处, 但聚类作为一种探索性的知识发现方法其结果本身带有预测性, 加上目前对于许多具体基因生物学功能的认识, 特别是其在整个复杂的调控网络中所处状况的了解还很有限, 即使目前获取到的 GO 注解信息其来源也往往是片面的^[24]. 所以本文没有详细推测某个基因可能的具体功能, 而重点揭示基因在对照品种间和生长发育过程中存在的差异表达规律, 因而最有可能受到重要的调控, 建议通过后续深入可靠的研究对其有更详细的了解, 以期为其他候选基因的定位和发现提供新的思路和线索.

两品种间基因调控网络的差异较大(图 5), 在一定程度上反映了两品种在肌肉生长和脂肪沉积相关的生理生化活动方面存在明显差异的分子基础. 值得注意, 在长白猪网络中起重要调控作用的 *PPARGC-1* 基因由于其在糖异生^[32]、线粒体呼吸^[33]、调节能量代谢以维持体温^[34]、脂肪细胞分化和脂肪酸氧化^[35]过程中具有重要作用, 并直接参与调控过氧化物酶体增殖剂激活受体家族(*PPARs*)等转录因子^[36]和葡萄糖转运蛋白 4 (*GLUT4*)^[37]等重要肉质性状候选基因, 已有研究表明 *PPARGC-1* 基因还具有增殖氧化型肌纤维的作用^[38], 成为目前研究的最新热点. Jacobs 等^[39]报道了猪 *PPARGC-1* 基因编码区的序列, 并在梅山猪-大白猪合成资源群体中发现该基因第 9 外显子中的一个多态位点显著性影响猪的板油重. 研究结果表明其不仅在发育过程中的表达差异显著(表 2), 与 Erkens 等^[22]的研究结果一致, 且模拟的调控网络也显示其为重要的中枢调控基因, 建议可作为重要候选基因研究. 两品种的调控网络均显示 *UCP3* 和 *UGP2* 基因被明显调控, 其中 *UCP3* 作为线粒体内膜的一种载体蛋白, 可将呼吸链与三磷酸腺苷(ATP)产生过程解偶联, 使能量以产热形式散发, 从而增加能量的消耗^[40]. 大量研究已表明 *UCP3* 对机体能量平衡涉及的体重(肥胖)、静止代谢率和食

物(饲料)转化率等性状具有显著效应,是影响肌肉生长和品质的重要功能基因^[2]. *UGP2* 则主要参与葡萄糖代谢^[41], 但涉及其具体功能的研究还尚未见报道, 建议可深入研究. 需要特别指出的是, 由于基因筛选和数据抽样引入的随机性以及算法本身存在的固有局限, 从表达谱数据出发通过反向工程(reverse engineering)建立的调控网络具有一定程度的模糊性, 但作为一种有益的尝试性探索可为进一步深入研究提供一些有价值的参考. 此外, 荧光定量 RT-PCR 验证的结果表明本次芯片试验的结果准确可靠^[42].

随着基因芯片技术的不断发展普及, PubMed 和 GO 这类搜索引擎的不断完善, 人们逐渐认识到, 与包罗万象和大海捞针式的基因组表达谱芯片相比, 基于对某一明确的生物学信号通路的现有知识而设计的功能分类芯片更能有效排除冗余信息, 易于数据挖掘并降低了研究成本. 这类芯片通常只包括几百个与研究对象有确定关系或至少存在理论关系的基因, 具有更强的目的性和针对性. 正是基于这一思路, 现针对影响肌肉生长和脂肪沉积的生物学通路设计并制作了相关功能分类芯片, 以研究生长发育过程中不同猪种相关基因的差异表达规律, 其结果验证并完善了此类研究已有的一些认识, 筛选出了对于猪胴体和肉质性状可能具有重要影响并值得深入研究的一些候选基因.

参 考 文 献

- 1 Suzuki K, Ishida M, Kadowaki H, et al. Genetic correlations among fatty acid compositions in different sites of fat tissues, meat production, and meat quality traits in Duroc pigs. *J Anim Sci*. 2006, 84(8): 2026–2034
- 2 Plastow GS, Carrón D, Gil M, et al. Quality pork genes and meat production. *Meat Sci*. 2005, 7(3): 409–421
- 3 中国猪品种志编委会. 中国猪品种志. 上海: 上海科学技术出版社, 1986, 155–160
- 4 Khatri P, Bhavsar P, Bawa G, et al. Onto-tools: An ensemble of web-accessible ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments. *Nucleic Acids Res*. 2004, 32: W449–456
- 5 Doms A, Schroeder M. GoPubMed: Exploring PubMed with the gene ontology. *Nucleic Acids Res*. 2005, 33: W783–786
- 6 Yang YH, Speed T. Design issues for cDNA microarray experiments. *Nat Rev Genet*. 2002, 3(8): 579–588

- 7 Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet*. 2002, 32: 490–495
- 8 Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*. 2002, 30(4): e15–24
- 9 Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*. 2005, 21(9): 2067–2075
- 10 Xia X, McClelland M, Wang Y. WebArray: An online platform for microarray data analysis. *BMC Bioinformatics*. 2005, 6: 306–311
- 11 Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics*. 2001, 17(6): 520–525
- 12 Wolfinger RD, Gibson G, Wolfinger ED, et al. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol*. 2001, 8(6): 625–637
- 13 Cui X, Hwang JT, Qiu J, et al. Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics*. 2005, 6(1): 59–75
- 14 Lee HK, Braynen W, Keshav K, et al. ErmineJ: Tool for functional analysis of gene expression data sets. *BMC Bioinformatics*. 2005, 6: 269–276
- 15 Ernst J, Bar-Joseph Z. STEM: A tool for the analysis of short time series gene expression data. *BMC Bioinformatics*. 2006, 7: 191–121
- 16 Wu CC, Huang HC, Juan HF, et al. GeneNetwork: An interactive tool for reconstruction of genetic networks using microarray data. *Bioinformatics*. 2004, 20(18): 3691–3693
- 17 Levine M, Davidson EH. Gene regulatory networks for development. *Proc Natl Acad Sci USA*. 2005, 102(14): 4936–4942
- 18 Geier F, Timmer J, Fleck C. Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge. *BMC Syst Biol*. 2007, 1: 11–26
- 19 Dojer N, Gambin A, Mizera A, et al. Applying dynamic Bayesian networks to perturbed gene expression data. *BMC Bioinformatics*. 2006, 7: 249–259
- 20 Chuaqui RF, Bonner RF, Best CJ, et al. Post-analysis follow-up and validation of microarray experiments. *Nat Genet*. 2002, 32: 509–514
- 21 Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002, 3: 34–45
- 22 Erkens T, Van Poucke M, Vandesompele J, et al. Development of a new set of reference genes for normalization of real-time RT-PCR data of porcine backfat and longissimus dorsi muscle, and

- evaluation with PPARGC1A. *BMC Biotechnol*, 2006, 6: 41—48
- 23 Datta S, Datta S. Empirical Bayes screening of many P -values with applications to microarray studies. *Bioinformatics*, 2005, 21(9): 1987—1994
- 24 Allison DB, Cui X, Page GP, et al. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet*, 2006, 7: 55—65
- 25 Yang H, Churchill G. Estimating P -values in small microarray experiments. *Bioinformatics*, 2007, 23(1): 38—43
- 26 Hancock CR, Brault JJ, Terjung RL. Protecting the cellular energy state during contractions: Role of AMP deaminase. *Physiol Pharmacol*, 2006, 57(Suppl 10): 17—29
- 27 Tikik M, Tikik K, Tornngren MA, et al. Development of inosine monophosphate and its degradation products during aging of pork of different qualities in relation to basic taste and retronasal flavor perception of the meat. *J Agric Food Chem*, 2006, 54(20): 7769—7777
- 28 Stratil A, Knoll A, Moser G, et al. The porcine adenosine monophosphate deaminase 1 (AMPD1) gene maps to chromosome 4. *Anim Genet*, 2000, 31(2): 147—148
- 29 Ernst J, Nau GJ, Bar-Joseph Z. Clustering short time series gene expression data. *Bioinformatics*, 2005, 21(Suppl 1): 159—168
- 30 Ramoni MF, Sebastiani P, Kohane IS. Cluster analysis of gene expression dynamics. *Proc Natl Acad Sci USA*, 2002, 99(14): 9121—9126
- 31 Costa IG, Schönthuth A, Schliep A. The Graphical query language: A tool for analysis of gene expression time-courses. *Bioinformatics*, 2005, 21(10): 2544—2545
- 32 Yoon JC, Puigserver P, Chen G, et al. Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature*, 2001, 413: 131—138
- 33 Wu Z, Puigserver P, Andersson U, et al. Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell*, 1999, 98(1): 115—124
- 34 Puigserver P, Wu Z, Park CW, et al. A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell*, 1998, 92(6): 829—839
- 35 Spiegelman BM, Puigserver P, Wu Z. Regulation of adipogenesis and energy balance by PPAR γ and PGC-1. *Int J Obes*, 2000, 24(Suppl 4): 8—10
- 36 Oberkofler H, Esterbauer H, Linnemayr V, et al. Peroxisome proliferator-activated receptor (PPAR) gamma coactivator-1 recruitment regulates PPAR subtype specificity. *J Biol Chem*, 2002, 277: 16750—16757
- 37 Michael LF, Wu Z, Cheatham RB, et al. Restoration of insulin-sensitive glucose transporter (GLUT4) gene expression in muscle cells by the transcriptional coactivator PGC-1. *Proc Natl Acad Sci USA*, 2001, 98(7): 3820—3825
- 38 Lin J, Wu H, Tarr PT, et al. Transcriptional co-activator PGC-1 alpha drives the formation of slow-twitch muscle fibres. *Nature*, 2002, 418: 797—801
- 39 Jacobs K, Rohrer G, Van Poucke M, et al. Porcine PPARGC1A (peroxisome proliferative activated receptor gamma coactivator 1A): Coding sequence, genomic organization, polymorphisms and mapping. *Cytogenet Genome Res*, 2006, 112: 106—113
- 40 Bézère V, Seifert EL, Harper ME. Uncoupling protein-3: Clues in an ongoing mitochondrial mystery. *FASEB J*, 2007, 21(2): 312—324
- 41 Looft C, Paul S, Thomsen PD, et al. Isolation and assignment of the UDP-glucose pyrophosphorylase gene (UGP2) to porcine chromosome 3q21→q22 by FISH and by analysis of somatic cell and radiation hybrid panels. *Cytogenet Cell Genet*, 2000, 89(3—4): 154—155
- 42 Qin LX, Beyer RP, Hudson FN, et al. Evaluation of methods for oligonucleotide array data via quantitative real-time PCR. *BMC Bioinformatics*, 2006, 7: 23—24