论文

基于 RNA-seq 测序数据鉴定和分析猪基因组可变剪接事件

冉茂良, 陈斌*, 李智, 董莲花, 贺长青, 柳小春

湖南农业大学动物科学技术学院, 湖南省畜禽遗传改良重点实验室, 长沙 410128

* 联系人, E-mail: chenbin7586@126.com

收稿日期: 2015-06-23; 接受日期: 2015-11-05; 网络版发表日期: 2016-03-04

国家现代农业产业技术体系建设专项资金(批准号: CARS-36)和湖南省研究生科研创新项目(批准号: CX2015B251)资助

摘要 可变剪接作为一种增加基因组和蛋白质多样性的机制, 广泛发生于真核生物基因转录的过程中. 本文采用 Illumina Hiseq 2500 测序平台对猪睾丸组织 4 个不同发育时期(60 胎龄、90 胎龄、30 日龄和 180 日龄)进行转录组测序, 以猪基因组数据为参考, 鉴定和分析了猪基因组可变剪接事件. 从猪基因组中鉴定出 20398 个基因(80.6%)对应的 92738 个可变剪接事件. 在不同的可变剪接类型中, 以第一个外显子可变剪切(alternative 5' first exon, TSS)、最后一个外显子可变剪切(alternative 3' last exon, TTS)、单外显子跳跃(skipped exon, SKIP)和可变 5'或 3'端剪切(alternative exon ends, AE) 4 种类型为主, 分别占有可变剪接事件的 41.0%, 32.9%, 7.6%和 7.4%. GO 功能富集分析显示, 发生可变剪接的基因主要富集于物质合成、物质结合及酶活性相关的 GO 项中, 而各发育时期特异的可变剪接基因与发育时期的生理状态密切相关, 60 胎龄时主要与酶活性和组织形成相关, 30 日龄时主要与抗环境应激和离子通道活性相关, 180 日龄时则主要与循环系统相关. 此外, 在筛选出 64 个与睾丸素代谢相关的基因中, 63 个基因发生可变剪接, 且以 TSS 和 TTS 为主, 表明这两种可变剪接类型与睾丸素合成和分泌密切相关. 通过对猪基因组可变剪接的分析, 为深入研究可变剪接生物学功能及进一步开展分子育种工作提供理论依据.

关键词 猪, 可变剪接, 睾丸素, RNA-seq

在腺病毒中证实外显子和内含子间隔组成真核基因编码序列后不久^[1], Gilbert 于 1978 年首次发现基因的可变剪接(alternative splicing, AS)现象, 即指基因的 mRNA 前体在剪切内含子过程中以不同的方式剪接形成超过一种 mRNA 成熟体的过程^[2]. 可变剪

接现象广泛发生于真核生物基因的转录过程中, 尽管最初认为高等生物仅 5%左右的基因发生可变剪接^[3], 但随后在人(*Homo sapiens*)和其他几种重要模式生物中的研究证实基因发生可变剪接的比例远不止于此, 人、线虫(*Caenorhabditis elegans*)、果蝇

引用格式: 冉茂良, 陈斌, 李智, 等. 基于 RNA-seq 测序数据鉴定和分析猪基因组可变剪接事件. 中国科学: 生命科学, 2016, 46: 274–284
Ran M L, Chen B, Li Z, et al. Identification and analysis of alternative splicing events in *Sus scrofa* using RNA-Seq data. Sci Sin Vitae, 2016, 46: 274–284, doi: 10.1360/N052015-00174

(*Drosophila melanogaster*) 和拟南芥 (*Arabidopsis thaliana*) 中发生可变剪接的比例分别高达 95%, 71%, 60.7% 和 61%^[4-7]. 可变剪接被认为是增加转录组和蛋白质结构与功能多样性的一种重要机制, 例如, 由可变剪接形成的一个基因编码产生多种蛋白质产物即是人类基因组(少于 25000 个基因)编码产生约 90000 种蛋白质的原因之一^[8,9], 且它们在细胞中具有不同的生理功能^[10].

可变剪接通过调节信号通路中一些关键基因的表达, 在动物生长发育和生理代谢等过程中发挥重要的调控作用. 例如, 在癌细胞中, Ras-MAPK 和 PI3K-mTOR 信号通路中的多个基因发生可变剪接, 如 *mTOR*, *A/B-Raf*, *PTEN*, *PI3K* 和 *Ras* 等基因, 从而改变这两个信号通路的活性以影响细胞转化、癌细胞形成和发育以及肿瘤形成^[11]. 动物神经系统中 *Nf1*(neurofibromatosis type I) 基因在 23a 外显子处发生可变剪接形成两种异构体影响其自身和下游基因 *ERK1/2* 的活性, 进而调节细胞中 Ras 信号通路以调控神经元细胞功能^[12]. 肿瘤细胞中 MAPK 信号通路中的重要基因 *Mnk2* 具有激活肿瘤细胞的作用, 但其最后一个外显子处发生可变剪接形成 *Mnk2a* 和 *Mnk2b* 两种异构体, 其中 *Mnk2a* 蛋白含有一个 MAPK 结合域使其功能发生改变, *Mnk2a* 与 p38-MAPK 通路共同抑制肿瘤细胞的活性^[13]. 由此可见, 可变剪接是产生蛋白质结构和功能多样性的一个重要机制, 因此鉴定出基因可变剪接事件有利于全面研究其功能.

目前, 已有不同的实验设计和分析方法用以鉴定可变剪接事件^[14], 包含多种以基因芯片为基础的方法用于全面鉴定和分析动物基因组中的可变剪接事件, 尽管这些方法是鉴定已知剪接变异体的重要方法, 但是依赖于注释连接和现有基因的探针建立起来的微阵列基因芯片却不能鉴定出新的可变剪接位点^[15-17]. RNA-seq 技术是近几年发展起来研究转录图谱的一种高通量测序技术, 同时也被证实是一种可靠的鉴定新可变剪接事件的方法^[18]. 运用 RNA-seq 技术, 多个物种基因组的可变剪接事件得到全面鉴定和分析, 例如, 人^[4]、小鼠 (*Mus musculus*)^[19]、牛 (*Bovini*)^[20]、果蝇^[6]、线虫^[5]、拟南芥^[7]和中华猕猴桃 (*Actinidia chinensis*)^[21]等. 本研究采用 Illumina Hseq 2500 测序平台对 4 个发育阶段的

沙子岭猪 (*Sus scrofa*) 睾丸组织(60 胚龄、90 胚龄、30 日龄和 180 日龄)建库测序, 以猪全基因组序列为参考, 鉴定各样品中的可变剪接事件, 分析各样品中发生可变剪接事件的共性和特异性, 对发生可变剪接的基因进行 GO(gene ontology)功能富集, 并着重分析了与睾丸素代谢相关基因的可变剪接事件, 以期全面掌握猪基因组可变剪接事件及分子育种提供理论依据.

1 材料与方法

1.1 伦理申明

本研究的动物实验方法经湖南农业大学动物科学技术学院教授委员会和动物福利委员会论证并批准(批准号: 2013-06). 本研究使用的公猪、母猪和仔猪在饲养管理和取样过程中均未遭受任何不必要的伤害.

1.2 样品收集

沙子岭公猪由湖南省沙子岭猪资源场提供, 包括 60 胚龄(E60)、90 胚龄(E90)、30 日龄(D30)和 180 日龄(DM)各 3 头. 妊娠 60 和 90 天的母猪以颈部放血法处死后, 分别剖腹取其子宫内仔猪两侧睾丸组织, 30 和 180 日龄的公猪以人工阉割法取其两侧睾丸组织. 睾丸组织样品经预冷的磷酸盐缓冲液(0.01 mol/L)快速漂洗并沥去溶液后转入 1.5 mL RNase-free 离心管中, 标记后迅速置于液氮中速冷 5 h, 样品置于 -80℃ 保存备用.

1.3 总 RNA 提取

使用 TRIzol 试剂(Invitrogen, Carlsbad, 美国)提取猪睾丸组织样品总 RNA. 采用 1% 的琼脂糖凝胶电泳检测总 RNA 是否降解和受到污染, 并通过 Agilent 2100 生物分析仪使用 RNA Nano 6000 试剂盒(Agilent Technologies, 美国)检测总 RNA 完整性. 采用 Qubit® 2.0 荧光仪和 Qubit® RNA Assay Kit (Life Technologies, USA)检测总 RNA 浓度. 检测方法完全按照试剂盒使用说明书执行. 用于构建 RNA-seq 文库和测序的总 RNA 均满足以下条件: $A_{260/280}$ 介于 1.8~2.1 之间, 总 RNA 量 $\geq 6 \mu\text{g}$, rRNA Ratio(28S:18S) $\geq 1.8:1$, RNA 完整度(RNA integrity number, RIN) ≥ 8.0 .

1.4 RNA-seq 文库构建和测序

RNA-seq 文库构建和测序均由北京诺禾致源生物信息科技有限公司完成。分别取 3 份样品的总 RNA 等量混合后构建 4 个文库, 即 E60, E90, D30 和 DM, 使用 Illumina Hiseq 2500 测序平台进行测序, 测序获得长度为 125 bp 的序列(reads)。去除带接头的 reads、N 比例大于 10% 的 reads(N 表示无法确定碱基信息)和低质量 reads(质量值 $sQ \leq 5$ 的碱基数占整个 read 的 50% 以上的 reads)后, 分别获得 7.04, 6.14, 6.10 和 5.20 G 可用于注释分析的序列片段(clean reads)。本文仅使用得到的 RNA-seq 测序数据鉴定猪基因组的可变剪接事件, 分析各样品中发生可变剪接事件的共性和特异性, 研究猪睾丸发育过程中可变剪接事件的特征及规律。

1.5 测序数据过滤和可变剪接事件鉴定

从猪基因组数据库下载猪全基因组序列(10.2 版)及相关注释文件(ftp://ftp.ensembl.org/pub/release-74/fasta/sus_scrofa/dna/)。采用 TopHat 2 软件^[22]将 RNA-seq 测序得到的 clean reads 与猪基因组进行序列比对, 其中最大比错配数(mismatch)设置为 2, 其余均选用默认参数, 再使用 Cufflinks 软件^[23,24]预测出基因模型。在此基础上, 使用 ASprofile 软件对每个样品的可变剪接事件分别进行分类和统计(图 1)。ASprofile 软件^[25]中包含 12 种可变剪接事件类型(图 1), 即第一个外显子可变剪切(alternative 5' first exon, TSS)、最后一个外显子可变剪切(alternative 3' last exon, TTS)、单外显子跳跃(skipped exon, SKIP)、单

外显子跳跃(模糊边界)(approximate SKIP, XSKIP)、多外显子跳跃(multi-exon SKIP, MSKIP)、多外显子跳跃(模糊边界)(approximate MSKIP, XMSKIP)、单内含子滞留(intron retention, IR)、单内含子滞留(模糊边界)(approximate IR, XIR)、多内含子滞留(multi-IR, MIR)、多内含子滞留(模糊边界)(approximate MIR, XMIR)、可变 5' 或 3' 端剪切(alternative exon ends, AE)和可变 5' 或 3' 端剪切(模糊边界)(approximate AE, XAE)。

1.6 GO 功能富集分析和睾丸素代谢相关基因分析

对于各样本中发生可变剪接的基因, 本研究采用 Venny 在线工具(<http://bioinfogp.cnb.csic.es/tools/venny/index.html>)统计各样本中共有和特异的发生可变剪接的基因数量。将所有发生可变剪接的基因和样本特异基因 ID 比对到 DAVID 数据库, 获得根据 GO(gene ontology)分类的分子功能(molecular function)、生物过程(biological process)和细胞组成(cellular component)的 GO 编号。使用 GOseq 软件^[26]以基因组为背景进行功能富集分析, 并统计每个 GO 类别中的基因与背景基因之间的差异, 对 P 值进行校正, 定义校正后的 $P < 0.05$ 为这些基因的功能在该 GO 类别中显著富集。

为筛选出与睾丸素代谢相关的基因, 参考文献[21]中的方法, 本研究利用猪基因的 GO 注释信息、猪基因组蛋白注释信息和 KEGG 数据库中动物基因功能的信息筛选与睾丸素代谢相关的基因, 并统计这些基因在各样本中发生的可变剪接事件。对筛选

图 1 可变剪切分析流程及 ASprofile 中的可变剪切事件分类

出与睾丸素代谢相关的基因, 采用 HTSeq v0.6.1 计算在各样本中与每个基因匹配上的序列片段数量, 然后使用 Cufflinks 软件^[24]计算出每个基因表达量的 FPKM 值(fragments per kilobase of exon model per million mapped reads), 采用 R 软件包绘制它们在各样本中的表达量热图(heatmap).

2 结果与分析

2.1 猪基因组可变剪接事件鉴定

将 4 个发育阶段的猪睾丸组织的转录组数据与猪基因组进行比对后, 共计鉴定出 92738 个可变剪接事件, 这些可变剪接事件发生于 25237 个基因, 其中包含 20398 个猪基因组中已知基因(占猪基因总数的 80.6%)和 4839 个新预测的基因. 在不同的可变剪接事件类型中, 以 TSS, TTS, SKIP 和 AE 4 种类型为主, 数量分别为 38042, 30478, 7002 和 6870 个, 占有所有可变剪接事件的 41.0%, 32.9%, 7.6%和 7.4%; 余下 8 种可变剪接事件占有所有可变剪接事件的 0.3%~2.3%. 在发生可变剪接的基因中, 以发生 TTS, TSS, SKIP 和 AE 4 种可变剪接类型为主, 基因数量分别为 20098, 20057, 4380 和 2693 个, 占发生可变剪接的基因总数(25237 个)的比例为 79.6%, 75.5%, 17.4%和 10.7%; 发生其余 8 种可变剪接类型的基因的比例介于 0.7%~6.3%之间. 整体来讲, 平均每个可变剪接基因发生 3.67 个可变剪接事件, 其中 AE 和 XAE 的平均分值最高, 分别为 2.55 和 2.97, 其余 10 种可变剪接类型的平均分为 1.29~1.82.

通过对猪睾丸组织 4 个发育时期发生可变剪接的基因进行统计和比较, 得到可变剪接事件在 4 个发育时期的共有性和特异性(图 2). 在各发育时期的猪睾丸组织中均发生可变剪接事件的基因数量达 17565 个, 分别占猪全部基因和发生可变剪接基因总数的 69.4%和 86.1%. 在 180 日龄的猪睾丸组织中特异性发生可变剪接的基因数量远多于其他几个时期, 为 2978 个, 而 60 胚龄、90 胚龄和 30 日龄的分别为 582, 366 和 333 个. 由此可见, 在不同发育时期的睾丸组织中发生可变剪接的基因具有特异性.

可变剪接事件的数量在不同发育时期的猪睾丸组织之间存在较大的差异, 其中以 180 日龄的最多,

图2 猪睾丸组织 4 个发育阶段共有和特有发生可变剪接基因数量

在 60 胚龄、90 胚龄和 30 日龄中随发育时间的进展而逐渐降低, 即 60 胚龄>90 胚龄>30 日龄. 但是, 不同时期的猪睾丸组织发生可变剪接事件类型的比例并无明显差异(图 3). 其中, SKIP, XSHIP, MSKIP, MIR, XMIR 和 XAE 在各发育时期的比例均一致, 分别为 10%, 3%, 3%, 0%, 0%和 1%; TSS, TTS, IR 和 AE 在前 3 个发育时期的比例一致, 分别为 38%, 33%, 3%和 5%, 与 180 日龄时期存在一定的差异; 而 XIR 在后 3 个发育时期的比例均为 2%, 稍高于 60 胚龄时期.

2.2 猪基因组可变剪接基因 GO 注释及功能富集分析

为了解发生可变剪接基因的功能, 对所有发生可变剪接的基因进行了 GO 注释和功能分析, 选取 $P < 1 \times 10^{-10}$ 的 GO 项绘制了图 4. 在生物学过程、分子功能和细胞构成中显著(校正后的 $P < 0.05$)富集的 GO 项分别有 151, 84 和 20 个. 在生物学过程中, 发生可变剪接的基因主要集中于代谢过程(GO:0008152)和生物合成(GO:0009058)等相关 GO 项中. 在代谢过程的子类中, 有较多的可变剪接基因被富集于有机物代谢过程(GO:0071704)、大分子代谢过程(GO:0043170, GO:0044260)、氮化合物代谢过程(GO:0006807)、RNA 代谢过程(GO:0016070)和杂环化合物代谢过程(GO:0046483)等 GO 项中. 在生物合成的子类中, 细胞氮化合物的生物合成(GO:0044271)、碱基化合物的生物合成(GO:0034654)和有机环状化合物的生物合成(GO:1901362)等 GO 项中也有较多的基因富集. 在分子功能中, 物质结合和催化活性调节的 GO

图 3 猪睾丸组织 4 个发育时期中各类型可变剪接事件的比例

表 1 不同发育时期的猪睾丸组织中鉴定的可变剪接事件^{a)}

可变剪接事件	E60	E90	D30	DM	共计
TSS	18483(27985, 38.17%)	18543(26805, 38.17%)	18556(26517, 37.93%)	18806(31331, 39.50%)	20057(38042, 41.02%)
TTS	18870(23903, 32.59%)	18902(23115, 32.91%)	18861(23331, 33.37%)	18924(26638, 33.58%)	20098(30478, 32.86%)
SKIP	2315(7100, 9.68%)	2402(7328, 10.43%)	2342(7150, 10.23%)	2465(8130, 10.25%)	4380(7002, 7.55%)
XSKIP	800(2332, 3.18%)	775(2264, 3.22%)	782(2276, 3.26%)	813(2438, 3.07%)	1167(1785, 1.92%)
MSKIP	592(1840, 2.51%)	627(1908, 2.72%)	600(1814, 2.59%)	577(1936, 2.44%)	1091(1693, 1.83%)
XMSKIP	350(1030, 1.40%)	364(1034, 1.47%)	348(1016, 1.45%)	343(1010, 1.27%)	489(766, 0.83%)
IR	816(2172, 2.96%)	685(1730, 2.46%)	665(1678, 2.40%)	648(1624, 2.05%)	1597(2098, 2.26%)
XIR	590(1900, 2.59%)	494(1516, 2.16%)	474(1484, 2.12%)	504(1584, 2.00%)	1069(1943, 2.10%)
MIR	120(306, 0.42%)	65(172, 0.24%)	68(174, 0.25%)	53(122, 0.15%)	206(266, 0.29%)
XMIR	110(308, 0.42%)	65(180, 0.26%)	78(220, 0.31%)	63(170, 0.21%)	188(269, 0.29%)
AE	1438(3542, 4.83%)	1382(3361, 4.79%)	1379(3381, 4.84%)	1339(3463, 4.37%)	2693(6870, 7.41%)
XAE	321(925, 1.26%)	290(818, 1.16%)	296(823, 1.18%)	308(882, 1.11%)	514(1526, 1.65%)
共计	20529(73343)	20156(70231)	20176(69918)	22552(79328)	25237(92738)

a) 数据表示为: 可变剪接的基因(包含新预测基因)数量(鉴定出的可变剪接事件数量及比例); 其中 1 个基因可对应多个可变剪接事件

项中存在大量发生可变剪接的基因, 例如, 物质结合中的蛋白结合 (GO:0005515)、离子结合 (GO:0043167)、有机环状化合物结合 (GO:0097159) 和杂环化合物结合 (GO:1901363) 等, 催化活性中的水解酶活性 (GO:0016787)、转移酶活性 (GO:0016740) 和激酶

活性 (GO:0016301) 等。在细胞构成中, 则仅细胞组成 (GO:0044464)、细胞内 (GO:0005622)、胞外区 (GO:0005576) 和膜状胞器 (GO:0043227) 等发生可变剪接的基因数量较多。

通过对各时期猪睾丸组织特异性发生可变剪接

图 4 猪发生可变剪接事件基因的最显著 GO 富集结果

的基因进行 GO 功能富集发现: 各时期最显著富集的 GO 项存在明显的差异(图 5), 其中 90 胚龄的特异性基因未注释出 GO 项. 60 胚龄的睾丸组织中最显著富集的 GO 项主要与酶活性和组织形成相关, 例如, 蛋白络氨酸激酶活性(GO:0004713)、 β -呋喃果糖苷酶活性(GO:0004564)、 α -葡萄糖苷酶的活性(GO:0004558)、营养库活性(GO:0045735)、蛋白质合成(GO:0051604)和表皮发育(GO:0008544)等. 30 日龄的睾丸组织中特异性发生可变剪接的基因显著富集于与抵抗环境影响和离子通道相关的 GO 项中, 包括抵抗细菌侵袭(GO:0009617, GO:0042742)、抵抗应激(GO:0006950)、电压门控离子通道活性(GO:0022843)和电压门控钾离子通道活性(GO:0005249)等. 而 180 日龄的睾丸组织中的特异性基因则与循环系统相关, 包括循环系统过程(GO:0003013)、循环系统中的血管过程(GO:0003018)、血液循环(GO:0008015)及管道大小的调节(GO:0035150, GO:0050880)等. 表明这些发育时期特异性的可变剪接对于基因在该时期的功能具有重要调节作用.

2.3 睾丸素相关基因表达差异分析

本研究共鉴定出 64 个与猪睾丸素代谢相关的基因, 其中 63 个基因发生共计 273 个可变剪接事件(网络版附表 1), 发生可变剪接的频率为 98.4%, 远高于猪基因组中已知基因发生可变剪接的频率(80.6%), 且在各可变剪接事件类型中, 以第一个外显子可变剪切(TSS)和最后一个外显子可变剪切(TTS)发生的频率最高, 表明猪睾丸素代谢相关基因的表达与这两种可变剪接类型之间存在密切的联系. 63 个发生可变剪接的基因在猪睾丸组织各发育时期的表达量存在显著的差异(图 6), 其中在 60 胚龄和 90 胚龄中的表达量较为一致, 仅有少量几个基因的表达量显著差异(如 ENSSSCG00000011702, ENSSSCG-00000002686 和 ENSSSCG00000010488), 而 30 日龄与 90 胚龄和 180 日龄之间有超过 80%左右的基因表达量显著差异, 呈现出相反的表达模式. 总体而言, 本研究鉴定出的 63 个与猪睾丸素代谢相关且发生可变剪接的基因表达量在 30 日龄中的表达量相对较高, 而在 60 胚龄、90 胚龄和 180 日龄的表达量相对较低.

图 5 各时期猪睾丸组织特异性发生可变剪接基因的最显著 GO 富集结果

此外, 仅有 2 个基因(ENSSSCG00000028753, ENSSSCG00000001991)在 30 日龄时不发生可变剪接, 有 1 个基因(ENSSSCG000000020706)在 30 日龄时才开始发生可变剪接, 且 180 日龄中睾丸素相关基因的表达量仅 18 个相对较高, 表明睾丸素的代谢在猪 30 日龄时最为旺盛, 这可能与促进包括睾丸组织在内的性器官和骨骼等组织器官快速生长相关.

3 讨论与结论

同一个基因通过可变剪接后形成多种 mRNA 成

熟体, 经翻译产生结构和功能不同的蛋白质, 被认为是形成基因组和蛋白组多样性的一种重要机制^[27,28]. 目前的研究表明, 可变剪接广泛发生于动物、植物和微生物的基因组中. RNA-seq 作为研究基因组的一种被普遍采用的技术, 不仅能够准确定量出基因的表达量^[29,30], 还能精确地识别出基因序列的变化^[31], 从而被广泛应用于鉴定动植物和微生物基因组中发生的可变剪接事件^[6,7,32]. 本研究采用 Illumina Hseq 2500 测序平台对 4 个发育时期的沙子岭猪睾丸组织(60 胚龄、90 胚龄、30 日龄和 180 日龄)建库测序, 以猪基因组为参考, 共鉴定出 92738 个可变剪接事

图6 猪睾丸组织睾丸素相关基因在不同时期中的表达量差异

从上至下的基因 ID 同网络版附表 1

件, 发生可变剪接的基因比例达 80.6%, 证实可变剪接在猪基因中普遍发生。此外, 这一比例高于小鼠的 57% 和牛的 21%^[20], 这可能是本研究对 ASprofile 软件中定义的 12 种类型(图 1)的可变剪接事件进行了鉴定和统计, 但远低于人的 95%^[4], 这可能与本研究的 RNA-seq 的测序深度不够和猪基因组中的基因并未全部表达于这 4 个发育时期的睾丸组织有关, 相信随着猪转录组数据的不断增加, 80.6% 的基因发生可变

剪接这一比例将会有所增加, 本研究同时鉴定出 4839 个新预测的基因发生了可变剪接即可证明这一观点。

本研究通过对所有发生可变剪接的基因进行 GO 功能富集分析发现, 大量的基因富集于物质合成、物质结合及酶活性相关的 GO 项中, 且在其他研究中证实这 3 种 GO 大类与组织器官发育密切相关^[30,33]。这表明, 可变剪接通过调节这些基因的蛋白质产物结构与功能而在睾丸组织生长和发育过程发挥重要作用。此外, 各发育时期均存在基因特异性发生可变剪接, 这表明, 可变剪接在不同发育时期的组织中对基因蛋白质产物的结构和功能发挥着调控作用^[34,35]。对各发育时期的睾丸组织特异性发生可变剪接的基因进行 GO 功能富集分析发现, 各发育时期中显著富集的功能具有较为明显的特征。60 胚龄时主要与酶活性和组织形成相关, 30 日龄时主要与抗环境应激和离子通道先关, 180 日龄时则主要与循环系统相关。猪睾丸组织形成于 59 胚龄左右, 此时则需要各种酶催化产生形成睾丸组织的物质, 如蛋白质和大分子化合物等。在仔猪出生后的一段时间内, 生活环境发生巨大的改变, 受到外界环境的刺激和影响, 且组织器官快速生长发育, 机能逐渐完善, 因此需要在逐渐适应环境的同时快速的生长发育。公猪在 180 日龄时已经达到性成熟且能产生成熟的精子, 需要循环系统维持睾丸内部环境的动态平衡。这表明, 不同发育时期特异可变剪接基因的功能与该时期组织的生理状态密切相关, 且这些基因的可变剪接有助于它们发挥特定的功能。

由睾丸组织分泌的睾丸素, 不仅对于促进机体性成熟具有重要作用^[36,37], 还影响着机体的许多系统, 如血生成、脂代谢^[38]、糖代谢^[39]、体内钙平衡^[40]和骨矿化作用^[41]。本研究共筛选出 64 个与睾丸素代谢相关的基因, 其中 63 个基因发生可变剪接, 表明可变剪接对于这些基因在睾丸素代谢过程中行使正常的功能具有重要作用。在所有的可变剪接事件中, 以第一个外显子可变剪切和最后一个外显子可变剪切发生频率最高, 表明这两种可变剪接类型与睾丸素合成和分泌密切相关。此外, 63 个发生可变剪接的基因在 30 日龄的睾丸组织中的表达量相对于其他几个时期显著增加, 且仅有 2 个基因发生可变剪接, 表明可变剪接对于调控这些基因的表达具有重要作用。

参考文献

- 1 Berget S M, Moore C, Sharp P A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Rev Med Virol*, 2000, 10: 356–362
- 2 Gilbert W. Why genes in pieces? *Nature*, 1978, 271: 501
- 3 Sharp P A. Split genes and RNA splicing. *Cell*, 1994, 77: 805–815
- 4 Pan Q, Shai O, Lee L J, Frey B J, et al. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 2008, 40: 1413–1415
- 5 Ramani A K, Calarco J A, Pan Q, et al. Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res*, 2011, 21: 342–348
- 6 Graveley B R, Brooks A N, Carlson J W, et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature*, 2011, 471: 473–479
- 7 Marquez Y, Brown J W, Simpson C, et al. Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res*, 2012, 22: 1184–1195
- 8 Brett D, Pospisil H, Valcarcel J, et al. Alternative splicing and genome complexity. *Nat Genet*, 2002, 30: 29–30
- 9 Valdivia H H. One gene, many proteins: alternative splicing of the ryanodine receptor gene adds novel functions to an already complex channel protein. *Circ Res*, 2007, 100: 761–763
- 10 Hughes A C, Mort M, Elliston L, et al. Identification of novel alternative splicing events in the huntingtin gene and assessment of the functional consequences using structural protein homology modelling. *J Mol Biol*, 2014, 426: 1428–1438
- 11 Siegfried Z, Bonomi S, Ghigna C, et al. Regulation of the Ras-MAPK and PI3K-mTOR signalling pathways by alternative splicing in cancer. *Int J Cell Biol*, 2013: e568931
- 12 Hinman M N, Sharma A, Luo G, et al. Neurofibromatosis type 1 alternative splicing is a key regulator of Ras signaling in neurons. *Mol Cell Biol*, 2014, 34: 2188–2197
- 13 Maimon A, Mogilevsky M, Shilo A, et al. Mnk2 alternative splicing modulates the p38-MAPK pathway and impacts Ras-induced transformation. *Cell Rep*, 2014, 7: 501–513
- 14 Roy B, Haupt L M, Griffiths L R. Review: alternative splicing (AS) of genes as an approach for generating protein complexity. *Curr Genomics*, 2013, 14: 182–194
- 15 Johnson J M, Castle J, Garrett-Engle P, et al. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, 2003, 302: 2141–2144
- 16 Pan Q, Saltzman A L, Kim Y K, et al. Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev*, 2006, 20: 153–158
- 17 Shen S, Warzecha C C, Carstens R P, et al. MADS+: discovery of differential splicing events from Affymetrix exon junction array data. *Bioinformatics*, 2010, 26: 268–269
- 18 何涛, 王端青, 胡亚欧, 等. 基于 RNA-seq 数据识别果蝇可变剪接位点和可变剪接事件. *中国科学: 生命科学*, 2011, 41: 1016–1023
- 19 Mortazavi A, Williams B A, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 2008, 5: 621–628
- 20 Chacko E, Ranganathan S. Genome-wide analysis of alternative splicing in cow: implications in bovine as a model for human diseases. *BMC Genomics*, 2009, 10 Suppl 3: S11
- 21 周宴秋, 覃瑞, 刘胜毅, 等. 中华猕猴桃基因组可变剪接事件鉴定及分析. *中国科学: 生命科学*, 2015, 45: 289–300
- 22 Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 2013, 14: R36
- 23 Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 2012, 7: 562–578
- 24 Trapnell C, Williams B A, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, 2010, 28: 511–515
- 25 Florea L, Song L, Salzberg S L. Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Res*, 2013, 2: 188
- 26 Young M D, Wakefield M J, Smyth G K, et al. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*, 2010, 11: R14
- 27 Excoffon K J, Bowers J R, Sharma P. Alternative splicing of viral receptors: a review of the diverse morphologies and physiologies of

- adenoviral receptors. *Recent Res Dev Virol*, 2014, 9: 1–24
- 28 Hamid F M, Makeyev E V. Emerging functions of alternative splicing coupled with nonsense-mediated decay. *Biochem Soc Trans*, 2014, 42: 1168–1173
- 29 Li X J, Yang H, Li G X, et al. Transcriptome profile analysis of porcine adipose tissue by high-throughput sequencing. *Anim Genet*, 2012, 43: 144–152
- 30 Shan L, Wu Q, Li Y, et al. Transcriptome profiling identifies differentially expressed genes in postnatal developing pituitary gland of miniature pig. *DNA Res*, 2014, 21: 207–216
- 31 Sultan M, Schulz M H, Richard H, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, 2008, 321: 956–960
- 32 Klerk E D, Hoen P A C. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet*, 2015, 31: 128–139
- 33 Blythe M J, Malla S, Overall R, et al. High through-put sequencing of the *Parhyale hawaiiensis* mRNAs and microRNAs to aid comparative developmental studies. *PLoS One*, 2012, 7: e33784
- 34 Montelli S, Mazzotta G, Vanin S, et al. Period and timeless mRNA splicing profiles under natural conditions in *Drosophila melanogaster*. *J Biol Rhythms*, 2015, 30: 217–227
- 35 Wang E T, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*, 2008, 456: 470–476
- 36 Mangiamale L A, Thompson R R. Testosterone rapidly increases ejaculate volume and sperm density in competitively breeding goldfish through an estrogenic membrane receptor mechanism. *Horm Behav*, 2012, 62: 107–112
- 37 Jordan K, Fromberger P, Stolpmann G, et al. The role of testosterone in sexuality and paraphilia—a neurobiological approach. Part II: testosterone and paraphilia. *J Sex Med*, 2011, 8: 3008–3029
- 38 Cai Z, Xi H, Pan Y, et al. Effect of testosterone deficiency on cholesterol metabolism in pigs fed a high-fat and high-cholesterol diet. *Lipids Health Dis*, 2015, 14: 18
- 39 Grossmann M. Testosterone and glucose metabolism in men: current concepts and controversies. *J Endocrinol*, 2014, 220: R37–R55
- 40 Mandel P, Ebel A, Fontaine R. Effect of testosterone on calcium balance in the castrated rat. *C R Seances Soc Biol Fil*, 1954, 148: 713–715
- 41 Borst S E, Yarrow J F. Injection of testosterone may be safer and more effective than transdermal administration for combating loss of muscle and bone in older men. *Am J Physiol Endocrinol Metab*, 2015, 308: E1035–E1042

Identification and Analysis of Alternative Splicing Events in *Sus Scrofa* Using RNA-Seq Data

RAN MaoLiang, CHEN Bin, LI Zhi, DONG LianHua, HE ChangQing & LIU XiaoChun

Hunan Provincial Key Laboratory for Genetic Improvement of Domestic Animal, College of Animal Science & Technology, Hunan Agricultural University, Changsha 410128, China

Alternative splicing has been widely found in the eukaryotic gene transcription and serves as an important mechanism to enrich the genomic and proteomic diversity. In the current study, we sequenced the transcriptomic profiling of pig testes at four representative stages: 60 dpc (days post coitus, E60), 90 dpc (E90), 30-day-old (D30) and 180-day-old (DM) using Illumina Hiseq 2500 technology. Then, RNA-seq data were further used to identify the alternative splicing events referring the *Sus Scrofa* genome. The results showed that 92738 alternative splicing events were observed from *Sus Scrofa* genome corresponding to 20398 genes (80.6%). Alternative 5' first exon (TSS), alternative 3' last exon (TTS), skipped exon (SKIP) and alternative exon ends (AE) were the highest frequency events with 41.0%, 32.9%, 7.6% and 7.4% of total alternative splicing events, respectively. GO enrichment analysis showed that the alternative splicing genes were mainly enriched in metabolic process, molecular binding and enzyme activity. The alternative splicing events in different developmental stage seemed to preferentially occur in genes with physiological state. Genes in E60 were associated with enzyme activity and tissue formation. Genes in D30 were connected in response to environmental stress and ion channel activity. The genes of DM were involved in circulatory system. In addition, 64 testosterone related genes were selected and 63 genes of them were detected alternative splicing events. Alternative 5' first exon (TSS) and alternative 3' last exon (TTS) were the most frequency events, which revealed that these two alternative splicing events had a crucial role in the metabolic process of testosterone. Through the genome-wide analysis of alternative splicing in *Sus scrofa*, our results will provide insight into further research on the biological function and the molecular breeding in pigs.

***Sus Scrofa*, alternative splicing, testosterone, RNA-seq**

doi: 10.1360/N052015-00174