

夏晗,李凡,彭玲伟,等.基于全长转录组测序的牛脂肪组织可变剪接事件分析[J].华中农业大学学报,2022,41(6):176-183.
DOI:10.13300/j.cnki.hnlkxb.2022.06.020

基于全长转录组测序的牛脂肪组织 可变剪接事件分析

夏晗,李凡,彭玲伟,杜玉琴,郭爱珍,杨利国,周扬

华中农业大学动物科学技术学院、动物医学院/农业动物遗传育种与繁殖教育部重点实验室,武汉 430070

摘要 为探究可变剪接影响下更为复杂的脂肪沉积调控网络,明晰可变剪接事件的发生对脂肪沉积调控机制的影响,本研究基于PacBio测序平台的第三代测序技术,对夷陵黄牛腹腔脂肪、皮下脂肪、肌间脂肪进行全长转录组测序分析。共发现15 445个基因检测到50 520个可变剪接,占牛全部基因的33.4%。对这些基因进行富集分析发现,83个GO条目显著富集,这些显著富集的通路可能参与脂肪沉积网络的调控,其中16个与脂质合成及代谢相关,27条KEGG通路显著富集,其中15条与脂质合成及代谢相关。使用KOG、KEGG、NR、Swiss Prot、GO数据库对测序序列进行注释,80 756条序列共注释到69 259个基因,94 458条CDS序列及对应的pep序列。其中GO分析中15 039条序列富集到507个生物学过程条目,7 907条序列富集到214个细胞组分类目,30 672条序列富集到559个分子功能条目。KEGG数据分析富集到34条通路共49 710条序列,其中7 002条序列参与信号转导途径。以上结果表明,牛脂肪组织存在大量可变剪接事件,并且这些发生可变剪接事件的基因对脂肪沉积的调控发挥重要作用,这可从基因可变剪接角度为解析脂肪沉积的调控网络提供必要的资源和理论依据。

关键词 脂肪沉积; 功能注释; 可变剪接; 全长转录组; 夷陵黄牛

中图分类号 S814.8 **文献标识码** A **文章编号** 1000-2421(2022)06-0176-08

肌内脂肪的含量以及沉积分布影响牛肉的嫩度、系水力、剪切力值、风味和多汁性,适当提高牛肉肌内脂肪含量可改变牛肉品质^[1-2]。但是,当皮下脂肪的积累较多时会降低饲料报酬,造成资源损耗^[3]。脂肪沉积受复杂的调控网络影响,Utrera等^[4]针对包括脂肪沉积能力的牛胴体性状遗传力进行了72项研究,也未能完全了解其遗传机制。

可变剪接是剪接体对前体mRNA选择性的修饰,使同一基因可以表达不同成熟mRNA剪接异构体的现象^[5]。可变剪接被认为在基因中广泛存在,有研究显示,在高通量测序结果中95%以上的人类基因存在可变剪接,由此产生表达多种不同的剪接异构体^[6]。可变剪接使基因表达具有更多可能性,展示了基因功能的多样性^[7]。研究显示可变剪接的发生会导致蛋白结构的变化,导致同一基因的不同蛋白亚型在功能上出现差异^[8];且有研究表明,可变剪接

会对牛的脂肪相关基因表达产生重要影响^[9]。例如,NOVA依赖性的可变剪接会调节白色脂肪组织褐变^[10]、质量减轻导致的皮下脂肪中TCF7L2可变剪接调节可导致脂肪组织中的高血糖和胰岛素作用受损^[11]等。所以了解可变剪接事件的发生对于明晰脂肪沉积调控网络有着重要的作用。

基于PacBio平台的全长转录组测序是第三代测序技术的代表,能直接测序得到RNA转录本的全长片段,可以规避短片段带来的缺失信息的影响,提供更完整的转录组信息,更好地进行差异表达基因分析及功能注释^[12]。同时,长读技术有利于更精确、更准确、更多地发现可变剪接事件,有助于提高对脂肪沉积调控网络的认识^[13]。目前,第三代测序技术已经被广泛应用于植物的遗传信息挖掘,并取得了较好的结果,例如:赭桐的种质资源挖掘^[14]、紫娟茶叶叶片呈色机理及物质合成途径^[15]、掌叶鱼黄草的遗

收稿日期:2022-04-18

基金项目:国家自然科学基金项目(31902148)

夏晗,E-mail:1134181483@qq.com

通信作者:周扬,E-mail:yangzhou@mail.hzau.edu.cn

传信息挖掘^[16]等。但是,目前基于全长转录组的研究较少,尤其是牛类动物繁殖周期长、生长缓慢更是限制了其研究的进展。本研究利用基于PacBio测序平台的第三代测序技术对夷陵黄牛脂肪组织遗传信息进行挖掘,以期为正确认识可变剪接对脂肪沉积调控网络的机制提供理论参考。

1 材料与方法

1.1 试验材料

供试牛为湖北省宜昌地区、年龄在3岁、发育正常、体态良好的3头健康夷陵黄牛。屠宰后,使用消毒手术刀,分别在每头牛相同部位采集5 cm³大小的皮下脂肪、腹腔脂肪、肌间脂肪样品。使用75%乙醇彻底清洗组织,PBS清洗残留在组织上的乙醇,吸干水分后于-80℃液氮保存,用于总RNA提取。

1.2 总RNA提取

使用TRIZOL法提取夷陵黄牛脂肪组织总RNA。提取后的RNA需使用1%凝胶电泳检测样品纯净度,并经过Nano Drop仪器检测,确定总RNA的OD₂₆₀/OD₂₈₀值在1.9~2.1、OD₂₆₀/OD₂₃₀值在1.4~1.8。并使用Qubit对RNA进行定量,Agilent 2100、Bioanalyzer等进一步进行质量检测,提取的总RNA质量满足要求后,低温保存,待后续进行文库构建。

1.3 文库构建及测序

构建文库所使用的RNA需要使用Qiagen试剂盒进行回收纯化预处理,以获得更高质量的RNA。随后使用SMARTer[®] PCR cDNA Synthesis Kit试剂盒反转得到cDNA。依据cDNA质量优化扩增程式,使用KAPA HiFi PCR Kits试剂盒进行扩增,对获得的全长cDNA片段使用SMRT bell template prep kit 1.0试剂盒构建文库。同时,对cDNA进行末端修复、损伤修复处理,并在全长cDNA的双端添加stem loop sequencing adapter进行筛选,文库构建成功后,利用PacBio Sequel测序平台开展全长转录组测序。

1.4 全长转录组数据组装与筛选

对建库测序得到的原始数据,首先使用SMRTLINK 5.1软件对其进行质控,然后使用SNR(Signal Noise Ratio)筛选剔除低质量数据及接头序列,得到环形一致序列(circular consensus sequence, CCS)。去除含5' primer、3' primer、poly A片段,获得的全长转录组序列经过校正和去冗余后,可用于

后续分析。

1.5 可变剪接检测

试验使用ICE(iterative isoform-clustering)中的algorithm模块进行聚类,使用DAGCon得到一致性序列,并通过arrow进行校正。将二代数据校正后的全长序列与参考基因组比对,去除融合基因及非注释数据后,将比对到参考基因组同一基因上的不同isoform进行提取,即为可变剪接。根据可变剪接不同的形成方式进行分类。

1.6 全长转录组功能注释与分类

获得高质量的全长转录组序列后,使用KOG蛋白质真核同源数据库、GO基因本体数据库、KEGG代谢通路数据库、NR非冗余蛋白数据库、Swiss Prot精准蛋白数据库对预测的Isoforms进行注释,用以评估脂肪组织中基因的功能注释信息及其参与的代谢通路及调控网络,并判断基因的物种相似性。

2 结果与分析

2.1 测序质量分析

对采集的夷陵黄牛的皮下脂肪、腹腔脂肪、肌间脂肪样品使用PacBio平台第三代测序技术进行全长转录组测序,共获得19.81 G的原始数据,剔除接头和长度小于50 bp的原始离线数据后,得到13 014 954条Subreads,平均长度为1 474 bp(Full Passes \geq 1、最小预测准确度0.8),核验下机数据中的CCS(circular consensus sequence)的质量,共获得779 345个CCS reads,CCS总碱基数为1 812 898 181 bp,CCS Read平均长度2 320.43 bp,平均Full Pass数量为11。通过验证5'引物、3'引物和poly-A尾的保留,获得高质量的FLNC(full-length nonchimeric reads)640 205条,占比82.15%,平均长度为2 091 bp,以上结果说明全长转录组测序数据质量较好,可以用于后续试验。

2.2 全长转录组 Reads 聚类与校正

利用来源于同一样品的二代测序数据对PacBio平台的第三代测序reads进行校正^[17]。校正数据后,我们获得了779 345个consensus reads。使用ICE去除缺失序列和冗余序列后,进一步使用Lordec软件,以二代数据作为参照对356 772条一致性序列结果进行校正,获得可用于进一步分析的一致性序列总长为790 043 023 bp。

2.3 可变剪接分析

校正后的数据与参考基因组进行比较,得到与2 095个融合基因相关的2 768条reads。将剩余数据进一步过滤去冗余。使用Match Annot软件将比对后结果与注释信息进行比较,并将三代注释结果和原基因组结果进行合并,得到基因注释的PB isoform 63 400条,非基因区序列有36 480条。将比对到参考基因组cDNA区域的序列进行检测,共识别到可变剪接事件50 520个。其中2 406个基因发生外显子跳跃、1 184个基因发生外显子替换、1 262个基因发生外显子接受、4 085个基因发生内含子保留、2 365个基因发生外显子位置替换、4 143个基因发生其他可变剪接事件(表1)。

将检测到可变剪接事件的基因进行GO富集分析,显示与83个条目显著相关($P < 0.05$),其中富集到分子功能类的GO条目有22个,细胞组分类的GO条目14个,生物过程类的GO条目47个。在细胞组分类别中最显著是“薄膜涂层”,其次是“涂层膜”“黏着斑”条目。在分子功能类别中最显著的是“分子功能”,其次是“连接”“蛋白质结合”条目。在生物学过程类别中最显著的是“生物调节”,其次是“细胞过程的调节”“生物过程的调节”,另外,我们还检测到与脂质合成及代谢相关过程的显著富集,包括“糖异

生”“葡萄糖代谢过程”“己糖生物合成过程”“单糖生物合成过程”“糖酵解过程”“丙酮酸代谢过程”“己糖代谢过程”等。这说明可变剪接对于脂肪沉积调控网络可能有着重要的影响(图1)。

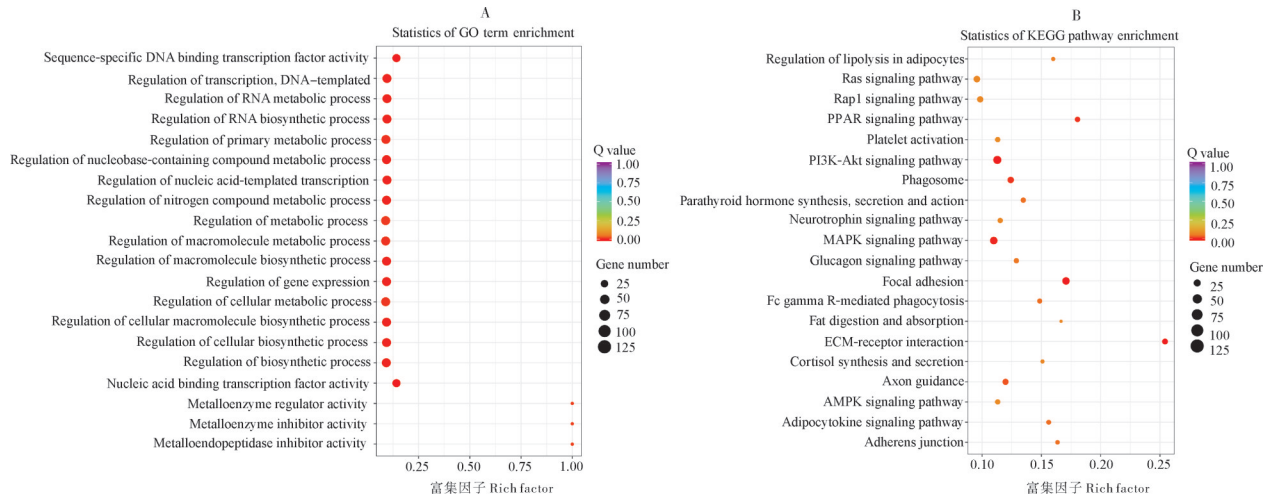
对检测到可变剪接事件的基因进行KEGG富集分析,以期研究可变剪接对基因之间互作关系的影响,结果显示有27条通路显著富集,其中直接与脂质合成及代谢相关的有15条通路,例如“脂肪细胞中脂解的调节”“脂肪消化吸收”“脂肪细胞因子信号通路”“PPAR信号通路”“AMPK信号通路”“PI3K-Akt信号通路”等。这说明可变剪接的参与导致脂肪沉积调控网络的复杂化。

表1 可变剪接类型统计表

Table 1 Statistical table of alternative splicing types

类型 Type	剪接数目 Number of AS	基因数 Number of gene
外显子跳跃 Exon S	4 677	2 406
外显子替换 Alt D	1 939	1 184
外显子接受 Alt A	2 138	1 262
内含子保留 Intron R	14 606	4 085
外显子位置替换 Alt P	6 490	2 365
其他 Others	20 670	4 143

注: AS统计根据测序注释结果与参考基因转录本比较得到。
Note: AS statistics are based on the comparison of sequencing annotation results with reference gene transcripts.



A:GO富集分析散点图 Scatter plot of GO enrichment analysis; B:KEGG富集分析散点图 Scatter plot of KEGG enrichment analysis.

图1 GO及KEGG功能富集图

Fig.1 Functional enrichment of GO and KEGG

2.4 新基因功能注释分析

为了获得更直观的注释结果,使用NR、KOG、GO、KEGG和Swissport 5个数据库进行注释,在全部80 756条序列中有69 259条被注释,如表2所示。

如图2显示至少有16 572个基因被5个数据库所注释,有26 653个基因至少在1个数据库中被注释。其中占全部基因数量81.61%的65 902个基因被NR数据库所注释,根据序列同源性比对结果,其中28 530

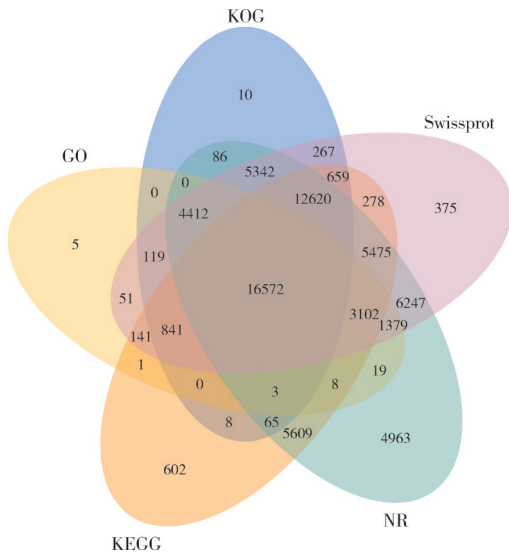


图 2 功能注释韦恩图

Fig.2 Venn diagram of functional annotation

条与普通牛同源、10 028 条与牦牛同源、3 265 条与平原野牛同源、2 970 条与绵羊同源、2 431 条与亚洲水牛同源以及与其他物种同源的 18 678 条,如图 3 所示。这说明了夷陵黄牛在品种形成过程中主要受到普通牛和瘤牛的影响。

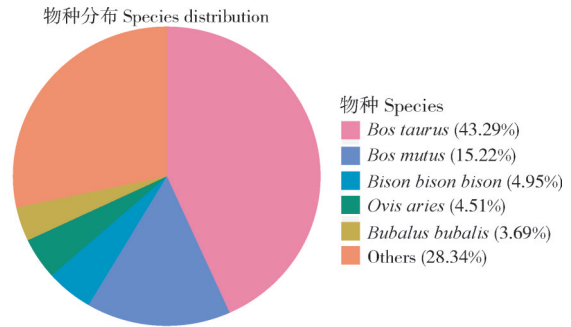


图 3 NR 分类注释统计图

Fig.3 NR classification annotation statistics

表 2 功能注释统计表

Table 2 Function note statistics table

项目 Item	数据库 Database					总基因数 Total gene number	总注释 General notes
	KOG	KEGG	NR	Swiss Port	GO		
基因数量 Gene number	41 004	45 984	65 902	57 880	26 653	80 756	69 256
注释比例/% Ratio of notes	0.78	56.94	81.61	71.67	33.00	—	85.76

KOG 数据库注释到 41 004 条序列在基因区域, 占全部序列的 50.78%。根据序列信息可以分为 26 组, 其中一般功能预测条目注释到 8 406 条序列, 信号转导机制条目 7 221 条, 注释最少的为未知蛋白组 (仅为 52 条)。从分组可以发现, 注释到细胞代谢发育等通用途径的序列占绝大多数。例如: 胞内运输、

分泌和囊泡运输、脂质运输和代谢、能量生产和转化分别注释到 2 822、1 620、920 条序列, 这些途径证明了在夷陵黄牛脂肪组织中进行着高频率的能量代谢以及物质跨膜运输, 这可能与脂肪细胞中脂质积累相关(图 4)。

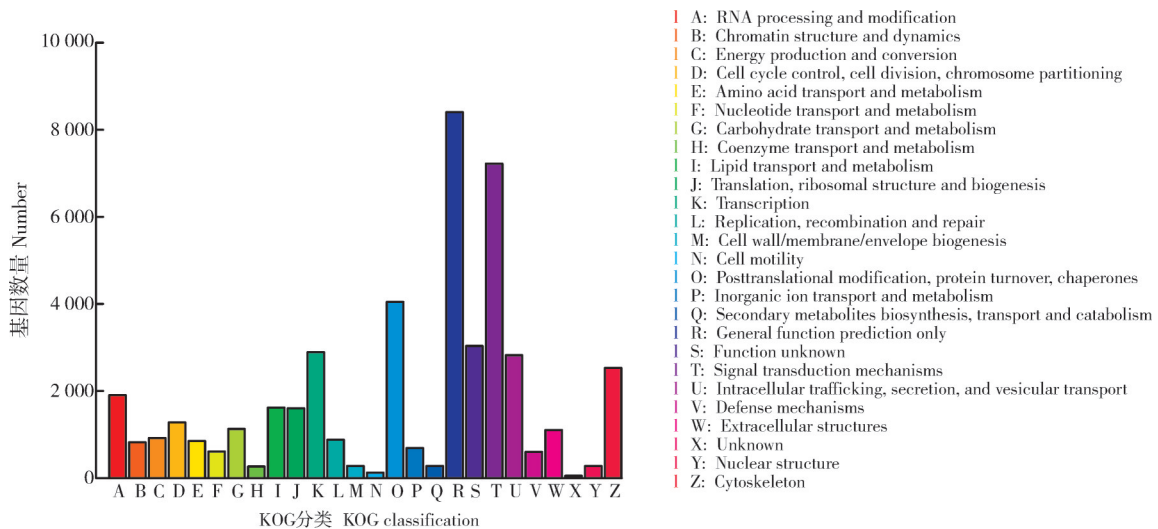


图 4 KOG 分类注释统计图

Fig.4 KOG classification annotation statistics

GO数据库注释到的基因有 26 653 个,占全部基因数量的 33%,富集到 GO 条目共 1 280 个,在分子功能条目上富集到 559 个条目,30 672 条序列,其中以蛋白质结合、核酸结合条目聚集到最多的序列,分别为 5 123 条和 2 706 条;生物过程富集到 507 个条目 15 039 条序列,其中转录调控及蛋白磷酸化聚集序列

最多,分别为 1 495 条和 1 423 条;细胞组分富集到 214 个条目 7 907 条序列,其中膜和膜的组成部分条目聚集的序列最多,分别为 1 142 条和 1 721 条。其中本研究检测到与细胞能量储存与代谢等相关条目的显著富集,可能与脂肪组织复杂的生物学过程及活跃的细胞代谢有关(图 5)。

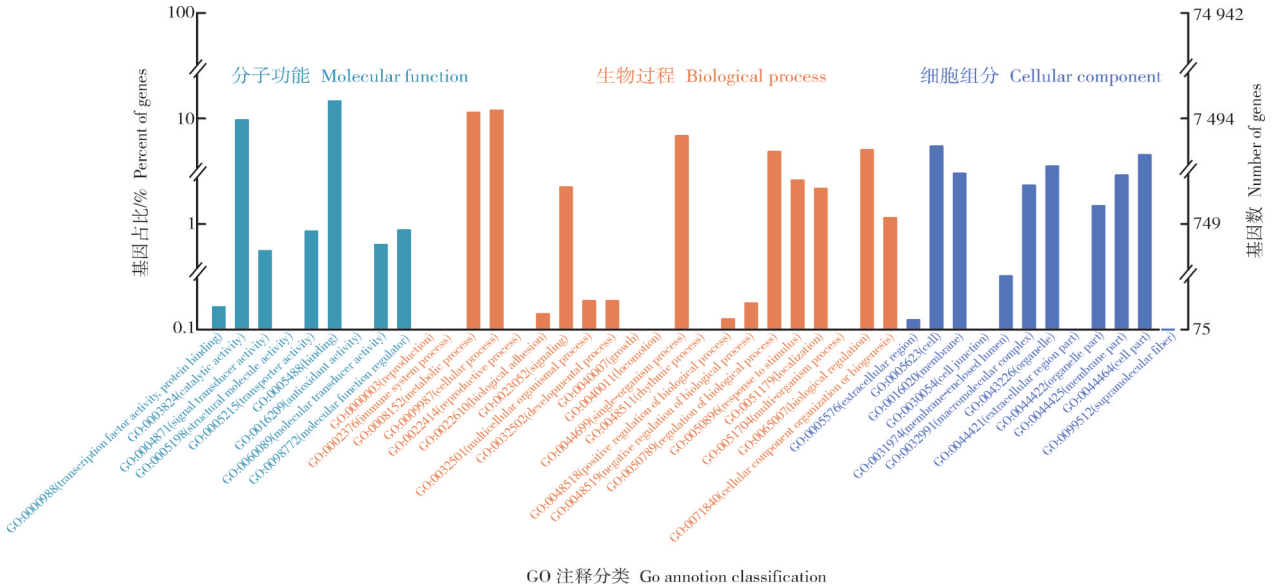


图 5 GO 分类注释统计图

Fig. 5 GO classification annotation statistics

KEGG 的功能注释结果发现,夷陵黄牛脂肪组织全长转录组数据中共注释到基因 45 984 个,占全部基因的 56.94%,KEGG 将注释得到的基因信息分为 5 个大类:细胞过程、环境信息处理、遗传信息处理、代谢、有机系统,这些类别分别富集到 10 745、9 516、5 789、7 005、1 655 个基因,如图 6 所示。KEGG 除富集到“细胞生长和死亡”“细胞运动”“免疫系统”“神经系统”等通路外,还富集到如“能量代谢”“聚糖生物合成和代谢”“脂质代谢”等与脂肪合

成、能量代谢高度相关的路径,这些显著富集的通路可能参与脂肪沉积网络的调控。

将 80 756 条序列与 Swiss-Prot 数据库蛋白库进行比对,共注释得到 69 259 个基因,占全部基因数量的 85.76%。采用 angel 软件,对序列做 CDS 和 Protein 预测,共发现 94 458 条 CDS 序列及其对应的 pep 序列,并且比对到 UTR 非编码区域的 155 214 条序列。将数据与基因注释结果对比优化后,得到 3' 或 5' 延长区域共 29 448 处,分别对应 21 145 条 reads,

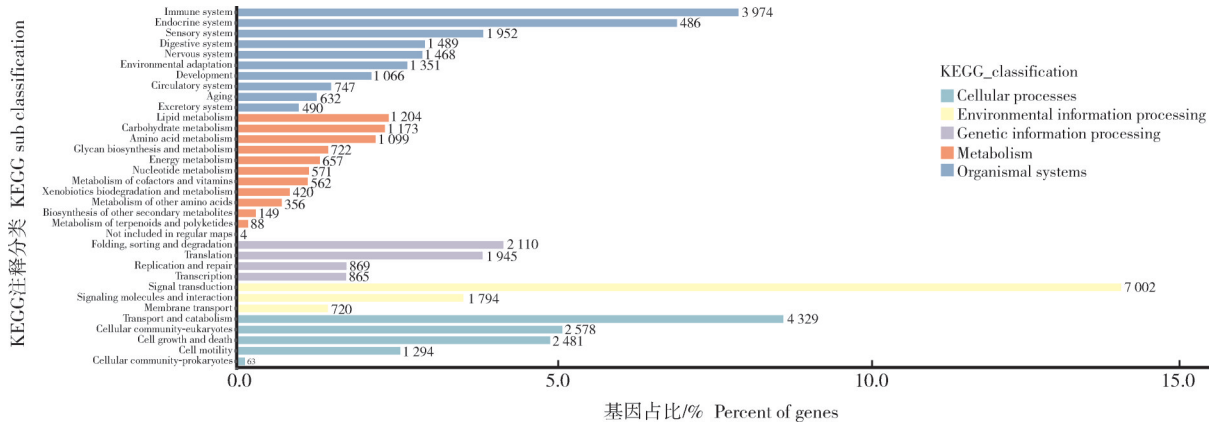


图 6 KEGG 分类注释统计图

Fig.6 KEGG classification annotation statistics

7 176 个基因。

3 讨论

可变剪接已经被证实在 95% 以上的人类基因中存在, 可变剪接的存在使得基因的功能存在不确定性, 脂肪调控网络变得更为复杂^[7]。三代测序技术可以弥补二代测序序列短、完整性差的缺陷, 直接获得全长序列, 从而准确地表征基因结构特点, 同时使用 RNA-seq 对全长转录组数据进行校正, 在进一步提高 PacBio 循环共有序列 (CCS) 质量的同时, 降低长序列的高碱基错误率, 提升可变剪接检出率和准确率^[18]。本研究利用三代测序技术对夷陵黄牛脂肪组织 (皮下脂肪、腹腔脂肪、肌间脂肪) 的总 RNA 进行全长转录组测序。通过数据质控和分析, 一共确定了 779 345 个 CCS reads, 平均长度为 2 320.43 bp, 共识别可变剪接事件 50 520 个, 超过在牛上发现全部可变剪接事件的 33.4%, 说明可变剪接可能与脂肪沉积复杂的调控网络紧密联系^[19]。前 5% 的发生可变剪接基因 GO 和 KEGG 富集分析结果显示, 83 个条目显著富集, 其中与脂质代谢相关过程的有“糖异生”“葡萄糖代谢过程”“糖酵解过程”“丙酮酸代谢过程”“己糖代谢过程”等。KEGG 结果显示, 这些可变剪接相关基因涉及到的通路共有 27 条, 与脂质合成及代谢相关的 15 条, 其中“PPAR 信号通路^[20]”“AMPK 信号通路^[21]”“PI3K-Akt 信号通路^[22]”“MAPK 信号通路^[23]”“Ras 信号通路^[24]”“Rap1 信号通路^[25]”“ECM-受体相互作用通路^[26]”“cGMP-PKG 信号通路^[27]”“Apelin 信号通路^[28]”等均有文献报道。这些可变剪接事件的发现, 是对现有脂肪可变剪接数据库的补充, 也为充分正确认识脂肪沉积调控网络提供了新的理论依据补充。

本研究共发现 80 756 个新基因序列, KOG、KEGG、NR、Swiss Prot、GO 数据库分别注释到其中 41 004、45 984、65 902、57 880、26 653 条序列, 在物种同源性比对中发现夷陵黄牛主要与普通牛和瘤牛同源, 属于南方黄牛的血统组成, 与夏小婷等^[29]的研究结果一致。但是小众动植物普遍存在基因组信息研究匮乏等现象, 在进行同源比对时可参考的基因组序列信息有限, 容易导致比对结果的偏离^[30]。尽管已经证明了全长转录组在缺少基因组信息时有获取数据的优势, 但长读技术依旧受到原始样品质量的

影响^[31]。本研究中, KOG 数据库注释到的序列占全部序列的 50.78%, 其中聚集到一般性功能预测分组的序列数量最多, 说明维持细胞一般性功能的序列占据主要部分。GO 数据库注释的序列占比 33%, 主要与维持细胞正常形态及功能相关, 其中分子功能组以蛋白质结合、核酸结合条目聚集到最多的序列, 生物过程中转录调控及蛋白磷酸化聚集序列最多, 细胞组分中膜的组成部分聚集的序列最多。KEGG 将注释得到的基因信息分为 5 个大类: 细胞过程 10 745 条、环境信息处理 9 516 条、遗传信息处理 5 789 条、代谢 7 005 条、有机系统 1 655 条, 除富集到“细胞生长和死亡”“细胞运动”“免疫系统”“神经系统”等主要通路外, 还富集到如“能量代谢”“聚糖生物合成和代谢”“脂质代谢”等与脂质合成及代谢高度相关的路径。除此之外, 我们还使用 Swiss-Prot 数据库蛋白库注释到 69 259 条序列, 发现 94 458 条可能的 CDS 序列及其对应的 pep 序列。这些数据可以为现有脂肪沉积相关研究进行数据补充, 并为后续研究提供参考信息, 为后续的研究提供宝贵资源。

参考文献 References

- [1] 孙亚红, 李俊霞, 胥福勋, 等. 影响肉牛肌肉脂肪沉积的遗传、管理和营养因素[J]. 中国饲料, 2020(2): 102-106. SUN Y H, LI J X, XU F X, et al. Factors related to genetic, management and nutrition affecting fat deposition in beef muscle[J]. China feed, 2020(2): 102-106 (in Chinese with English abstract).
- [2] 万发春, 张幸开, 张丽萍, 等. 牛肉品质评定的主要指标[J]. 中国畜牧兽医, 2004, 31(12): 17-19. WAN F C, ZHANG X K, ZHANG L P, et al. Main indexes of beef quality evaluation[J]. Animal science abroad, 2004, 31(12): 17-19 (in Chinese).
- [3] 李军涛. 日粮 VA 添加剂量对秦安杂交阉牛牛肉品质及血清生化指标的影响[D]. 郑州: 河南农业大学, 2012. LI J T. Effects of vitamin A on beef quality and serum biochemical indicators of Angus × Qinchuan crossbred steers[D]. Zhengzhou: Henan Agricultural University, 2012 (in Chinese with English abstract).
- [4] UTRERA A R, VAN VLECK L D. Heritability estimates for carcass traits of cattle: a review[J]. Genetics and molecular research: GMR, 2004, 3(3): 380-394.
- [5] MODREK B, LEE C. A genomic view of alternative splicing[J]. Nature genetics, 2002, 30(1): 13-19.
- [6] GALLEGO-PAEZ L M, BORDONE M C, LEOTE A C, et al. Alternative splicing: the pledge, the turn, and the prestige[J]. Human genetics, 2017, 136(9): 1015-1042.
- [7] BLENCOWE B J. The relationship between alternative splicing and proteomic complexity[J]. Trends in biochemical sciences,

- 2017, 42(6):407-408.
- [8] YANG X P, COULOMBE-HUNTINGTON J, KANG S L, et al. Widespread expansion of protein interaction capabilities by alternative splicing[J]. *Cell*, 2016, 164(4):805-817.
- [9] ZHOU Y, SUN J J, LI C J, et al. Characterization of transcriptional complexity during adipose tissue development in bovines of different ages and sexes[J/OL]. *PLoS One*, 2014, 9(7): e101261 [2022-04-18]. <https://doi.org/10.1371/journal.pone.0101261>.
- [10] VERNIA S, EDWARDS Y J, HAN M S, et al. An alternative splicing program promotes adipose tissue thermogenesis[J/OL]. *eLife*, 2016, 5: e17672 [2022-04-18]. <https://doi.org/10.7554/eLife.17672>.
- [11] KAMINSKA D, KUULASMAA T, VENESMAA S, et al. Adipose tissue TCF7L2 splicing is regulated by weight loss and associates with glucose and fatty acid metabolism[J]. *Diabetes*, 2012, 61(11):2807-2813.
- [12] CHAO Y H, YUAN J B, LI S F, et al. Analysis of transcripts and splice isoforms in red clover (*Trifolium pratense* L.) by single-molecule long-read sequencing[J/OL]. *BMC plant biology*, 2018, 18(1): 300 [2022-04-18]. <https://pubmed.ncbi.nlm.nih.gov/30477428/>. DOI: 10.1186/s12870-018-1534-8.
- [13] SHIN S C, AHN D H, KIM S J, et al. Advantages of single-molecule real-time sequencing in high-GC content genomes[J/OL]. *PLoS One*, 2013, 8(7): e68824 [2022-04-18]. <https://doi.org/10.1371/journal.pone.0068824>.
- [14] 章寒, 吴宜静, 张一鸣, 等. 尾叶紫薇全长转录组测序及微卫星标记开发[J]. *江西农业大学学报*, 2021, 43(4): 919-930. ZHANG H, WU Y J, ZHANG Y M, et al. Full-length transcriptome sequencing and microsatellite marker development in *Lagerstroemia caudata* [J]. *Acta Agriculturae Universitatis jiangxiensis*, 2021, 43(4): 919-930 (in Chinese with English abstract).
- [15] 夏丽飞, 孙云南, 宋维希, 等. 基于PacBio平台的紫娟茶树全长转录组分析[J]. *基因组学与应用生物学*, 2020, 39(6): 2646-2658. XIA L F, SUN Y N, SONG W X, et al. Full-length transcriptome analysis of Zijuan tea (*Camellia sinensis* var. *assamica* (Masters) Kitamura) based on PacBio platform[J]. *Genomics and applied biology*, 2020, 39(6): 2646-2658 (in Chinese with English abstract).
- [16] 张守信, 陈英转, 郑奋, 等. 基于PacBio平台的掌叶鱼黄草全长转录组分析[J]. *分子植物育种*, 2022, 20(9): 2939-2948. ZHANG S X, CHEN Y Z, ZHENG F, et al. Full-length transcriptome analysis of *Merremia vitifolia* based on PacBio platform[J]. *Molecular plant breeding*, 2022, 20(9): 2939-2948 (in Chinese with English abstract).
- [17] SUN M, HUANG D J, ZHANG A L, et al. Transcriptome analysis of heat stress and drought stress in pearl millet based on Pacbio full-length transcriptome sequencing[J/OL]. *BMC plant biology*, 2020, 20(1): 323 [2022-04-18]. <https://pubmed.ncbi.nlm.nih.gov/32640987/>. DOI: 10.1186/s12870-020-02530-0.
- [18] ZHANG G Q, SUN M, WANG J F, et al. PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice[J]. *The plant journal: for cell and molecular biology*, 2019, 97(2): 296-305.
- [19] LIU J D, TAN S X, HUANG S Q, et al. ASlive: a database for alternative splicing atlas in livestock animals[J/OL]. *BMC genomics*, 2020, 21(1): 97 [2022-04-18]. <https://doi.org/10.1186/s12864-020-6472-9>.
- [20] MAO Z, FENG M J, LI Z R, et al. ETV5 regulates hepatic fatty acid metabolism through PPAR signaling pathway[J]. *Diabetes*, 2021, 70(1): 214-226.
- [21] ZHAO Y, HU X B, LIU Y J, et al. ROS signaling under metabolic stress: cross-talk between AMPK and AKT pathway[J/OL]. *Molecular cancer*, 2017, 16(1): 79 [2022-04-18]. <https://pubmed.ncbi.nlm.nih.gov/28407774/>. DOI: 10.1186/s12943-017-0648-1.
- [22] HUANG X J, LIU G H, GUO J, et al. The PI3K/AKT pathway in obesity and type 2 diabetes[J]. *International journal of biological sciences*, 2018, 14(11): 1483-1496.
- [23] BARTEL D P. microRNAs: genomics, biogenesis, mechanism, and function[J]. *Cell*, 2004, 116(2): 281-297.
- [24] MAKDISSY N, HADDAD K, MOUAWAD C, et al. Regulation of SREBPs by sphingomyelin in adipocytes via a caveolin and ras-ERK-MAPK-CREB signaling pathway[J/OL]. *PLoS One*, 2015, 10(7): e0133181 [2022-04-18]. <https://doi.org/10.1371/journal.pone.0133181>.
- [25] YEUNG F, RAMÍREZ C M, MATEOS-GÓMEZ P A, et al. Nontelomeric role for Rap1 in regulating metabolism and protecting against obesity[J]. *Cell reports*, 2013, 3(6): 1847-1856.
- [26] MARIMAN E C M, WANG P. Adipocyte extracellular matrix composition, dynamics and role in obesity[J]. *Cellular and molecular life sciences*, 2010, 67(8): 1277-1292.
- [27] AI X, HOU X M, GUO T T. C-type natriuretic peptide promotes adipogenic differentiation of goat adipose-derived stem cells via cGMP/PKG/p38 MAPK signal pathway[J]. *In vitro cellular & developmental biology-animal*, 2021, 57(9): 865-877.
- [28] GUO M, CHEN F, LIN T F, et al. Apelin-13 decreases lipid storage in hypertrophic adipocytes in vitro through the upregulation of AQP7 expression by the PI3K signaling pathway[J]. *Medical science monitor: international medical journal of experimental and clinical research*, 2014, 20: 1345-1352.
- [29] 夏小婷, 贾媛, 赵明, 等. 夷陵黄牛 mtDNAD-loop区遗传多样性研究[J]. *中国牛业科学*, 2017, 43(1): 4-7. XIA X T, JIA Y, ZHAO M, et al. Genetic diversity of mitochondrial DNA D-loop region in Yiling cattle[J]. *China cattle science*, 2017, 43(1): 4-7 (in Chinese with English abstract).
- [30] YANG J, WARISS H M, TAO L D, et al. *De novo* genome assembly of the endangered *Acer yangbiense*, a plant species with extremely small populations endemic to Yunnan Province, China[J/OL]. *GigaScience*, 2019, 8(7): giz085 [2022-04-18]. <https://doi.org/10.1093/gigascience/giz085>.

doi.org/10.1093/gigascience/giz085.

[31] BYRNE A, COLE C, VOLDEN R, et al. Realizing the potential of full-length transcriptome sequencing [J/OL]. *Philosophical*

transactions of the Royal Society of London. Series B, Biological sciences, 2019, 374(1786): 20190097[2022-04-18]. <https://doi.org/10.1098/rstb.2019.0097>.

Identification of alternative splicing events in bovine adipose tissue based on full-length transcriptome sequencing

XIA Han, LI Fan, PENG Lingwei, DU Yuqin, GUO Aizhen, YANG Liguo, ZHOU Yang

College of Animal Science and Technology & College of Animal Medicine, Huazhong Agricultural University/Key Laboratory of Agriculture Animal Genetics, Breeding and Reproduction of Ministry of Education, Wuhan 430070, China

Abstract In order to investigate the more complex regulatory network of fat deposition under the influence of alternative splicing, and to clarify the effect of the occurrence of alternative splicing events on the regulatory mechanism of fat deposition, in this study, full-length transcriptome sequencing was performed on abdominal fat, subcutaneous fat and intermuscular fat of Yiling yellow cattle based on the third-generation sequencing technology of PacBio sequencing platform. A total of 50 520 alternative splicing were detected in 15 445 genes, accounting for 33.4% of all bovine genes. Enrichment analysis of these genes revealed that 83 GO items were significantly enriched, among which 16 were associated with lipid synthesis and metabolism, and 27 KEGG pathways were significantly enriched, among which 15 were associated with lipid synthesis and metabolism. The sequenced sequences were annotated using KOG, KEGG, NR, Swiss Prot, and GO databases, and a total of 80 756 sequences were annotated to 69 259 genes, 94 458 CDS sequences and corresponding pep sequences. In GO analysis, 15 039 sequences were enriched to 507 biological process items, 7 907 sequences were enriched to 214 cellular component items, and 30 672 sequences were enriched to 559 molecular function items. A total of 49 710 sequences were enriched in 34 pathways by KEGG data analysis, of which 7 002 sequences were involved in signal transduction pathways. The above results indicate that there are a lot of alternative splicing events in bovine adipose tissue, and these genes with alternative splicing events play an important role in the regulation of fat deposition, which may provide necessary resources and theoretical basis for analyzing the regulatory network of fat deposition from the perspective of gene alternative splicing.

Keywords lipid deposition; functional annotation; alternative splicing; full-length transcriptome; Yiling cattle

(责任编辑:边书京)