

全转录组学在畜牧业中的应用

石田培, 张莉

中国农业科学院北京畜牧兽医研究所, 北京 100193

摘要: RNA 作为一种大分子参与基因编码、解码、调控、表达等多种生物学过程。目前, 对 RNA 的功能研究主要通过全转录组测序方法来完成。全转录组研究可以对基因结构与功能进行更深层次地分析和探究, 揭示基因表达与生命现象之间的内在联系。现阶段, 基于高通量测序技术的转录本结构研究、基因表达水平研究及非编码区域功能研究在模式动物、猪、禽类中已大量开展, 但在羊上却鲜有报道。本文介绍了利用 RNA-seq 及 Small RNA-seq 技术研究全转录组的一般流程及常用策略, 综述了全转录组学技术在畜牧业领域中的研究进展。

关键词: 基因表达; 全转录组; RNA 测序技术; 高通量测序

Application of whole transcriptomics in animal husbandry

Tianpei Shi, Li Zhang

Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China

Abstract: RNA is a polymeric molecule which is involved in various biological processes including the coding, decoding, regulation, and expression of genes. Whole transcriptome sequencing is the dominant method for studying RNA functions which assists researchers to deepen the exploration and analysis of gene structure and function and to reveal intrinsic links between gene expression and life phenomena. To date, extensive research has been done in animal husbandry models including swine and poultry by using high-throughput RNA sequencing technology. These studies included transcript structure, gene expression level and non-coding region function. In this review, we briefly introduce the general processes and strategies of RNA-seq and small RNA-seq technologies, and summarize the various achievements of the application of whole transcriptomics in animal husbandry.

Keywords: gene expression; whole transcriptome; RNA-seq; high-throughput RNA sequencing

在后基因组学研究中, 转录组学是解读基因组功能元件和揭示细胞及组织分子机制的基础, 在生物表型和基因表达研究中占据了重要地位^[1]。转录

组一般指狭义转录组, 即能够编码蛋白的 RNA, 而生物学中的转录本概念则为生物体特定细胞或组织在某一特定条件下的所有转录产物, 涵盖了编

收稿日期: 2018-07-31; 修回日期: 2018-10-22

基金项目: 国家自然科学基金项目(编号: U1503285) 和中国农业科学院基本科研业务费重大项目(编号: Y2017XM02)资助[Supported by the National

Natural Science Foundation of China(No. U1503285) and Chinese Academy of Agricultural Sciences(No.Y2017XM02)]

作者简介: 石田培, 在读硕士研究生, 专业方向: 动物遗传育种与繁殖。E-mail: 1337684764@qq.com

通讯作者: 张莉, 博士, 研究员, 博士生导师, 研究方向: 动物遗传育种与繁殖。E-mail: zhangli07@caas.cn

DOI: 10.16288/j.ycz.18-218

网络出版时间: 2018/11/15 9:28:51

URI: <http://kns.cnki.net/kcms/detail/11.1913.R.20181115.0928.002.html>

码 RNA (mRNA) 和非编码 RNA (non-coding RNA, ncRNA)^[2]。同相对稳定的基因组比较, 转录组会根据生理状态、生长阶段、生长环境的变化而变化^[3], 具有高度动态可变的特性。除此之外, 转录组图谱中含有丰富的生物学信息, 包括基因表达丰度及差异、基因结构、反义转录本、选择性剪切、单核苷酸多态性和基因融合等。

生物体是一个错综复杂的网络, 构成复杂有机体的任何简单分子都不能孤立存在或行使功能, 如转录组中的各类 RNA。这些不同类型的 RNA 生物学功能各异, 参与众多信号通路的调控。单一的 mRNA 或 ncRNA 研究缺乏关联性, 对多种 RNA 信息进行整合分析, 探索潜在的网络调控机制则成为转录组学研究的趋势。在传统转录组学(以 mRNA^[4]为主)研究中, 样品总 RNA 的提取通常利用 Oligo (dT) 磁珠富集法, 该方法能有效发掘富含 poly (A) 尾巴的 RNA^[5], 包括 mRNA 和一小部分长链非编码 RNA (long non-coding RNA, lncRNA), 但过滤掉了其他 poly (A) minus RNA 组分, 即无 poly (A) 尾 RNAs, 因此导致 circRNA (circular RNA) 和部分 lncRNA 的丢失, 不能全面地反映转录本的真实情况。Yang 等^[6]首次提出了基于 poly (A) minus 富集的 RNA-seq 方法, 可将 sno-lncRNAs (small nucleolar RNAs)、ciRNAs (circular intronic RNA) 和 circRNAs 等这类非 poly (A) 结尾的新型 RNA 分子富集并保留下来。至此, 包括 miRNA^[7]、lncRNA^[8,9] 和 circRNA^[10,11] 等非编码 RNA 的全转录组测序研究应运而生。本文综述了基于 RNA-seq 和 Small RNA-seq 技术的全转录组研究概况, 总结了全转录组测序的一般流程和常用策略, 最后对畜牧业未来的发展趋势进行展望。

1 全转录组测序概述

全转录组是指在一定时空状态下有机体特定细胞或组织能够转录出的所有转录本^[12], 其中蕴含着重要的生物学调控规律。丰富的 RNA 组学信息可用于开展差异表达 ncRNA 的靶基因预测、差异表达 ncRNA 和 mRNA 的正负相关性分析、ceRNA (competing endogenous RNAs) 对靶向 mRNA 的功能富集分析及关键基因挖掘等研究。全转录测序是通过高

通量 RNA-seq 测序技术对 RNA 序列进行测序, 并通过一些定量平台和技术反应其表达量和表达差异, 从而形成表达图谱。全转录组学分析就是以表达谱为基础, 对 RNA 进行鉴定和注释, 预测相应靶细胞或编码潜能, 并基于 GO (Gene Ontology)、COG (Clusters of Orthologous Groups)、KEGG (Kyoto Encyclopedia of Genes and Genomes) 等数据库进行功能富集、聚类分析、信息挖掘和通路探究等。不同于传统转录组学, 全转录组学在 mRNA 研究基础上, 涵盖了 lncRNA、circRNA、miRNA、siRNA (small interfering RNA) 和 piRNA (piwi-interacting RNA) 等多种非编码 RNA 及其之间的调控网络分析。目前, 转录组测序主要集中在单一 RNA 的定量表达和注释分析上, 而全转录组学测序可同时对多种 RNA 进行鉴定及关联分析(图 1)。

2 全转录组测序研究的一般技术流程和方法

2.1 测序样品的准备

在样品采集过程中, 组织中的 RNA 极易受内源或外源 RNA 酶作用而降解, 同样也容易受到蛋白质、DNA、同源和外源酚类等物质的污染, 因此, 样品质量和保存条件是决定试验结果质量的关键。不同物种不同组织部位的 RNA 得率大不相同, 在动物中, 肝脏、脾脏和胰脏组织的得率较大, 而肌肉、脑和心脏等组织得率较小。在皮肤、骨骼和毛发中 RNA 的提取难度较大, 效果也较差。在准备样品时, 首选新鲜组织, 剔除非研究所需的组织类型。若组织体积较大, 多管分组置于 -80 °C 或液氮中长期保存。

2.2 建库测序

2.2.1 cDNA 文库构建和 RNA 测序

总 RNA 质检后便可建库, 由于 miRNA 与其他 3 类 RNA 序列长度差异较大, 需要使用 SE (single-end) 和 PE (paired-end) 两种测序策略同时建立两个文库。链特异性建库的具体步骤如下(图 2): (1) 通过试剂盒在样品总 RNA 中去除 rRNA; (2) 利用六碱基随机引物合成第一条 cDNA 链; (3) 第二条 cDNA 链

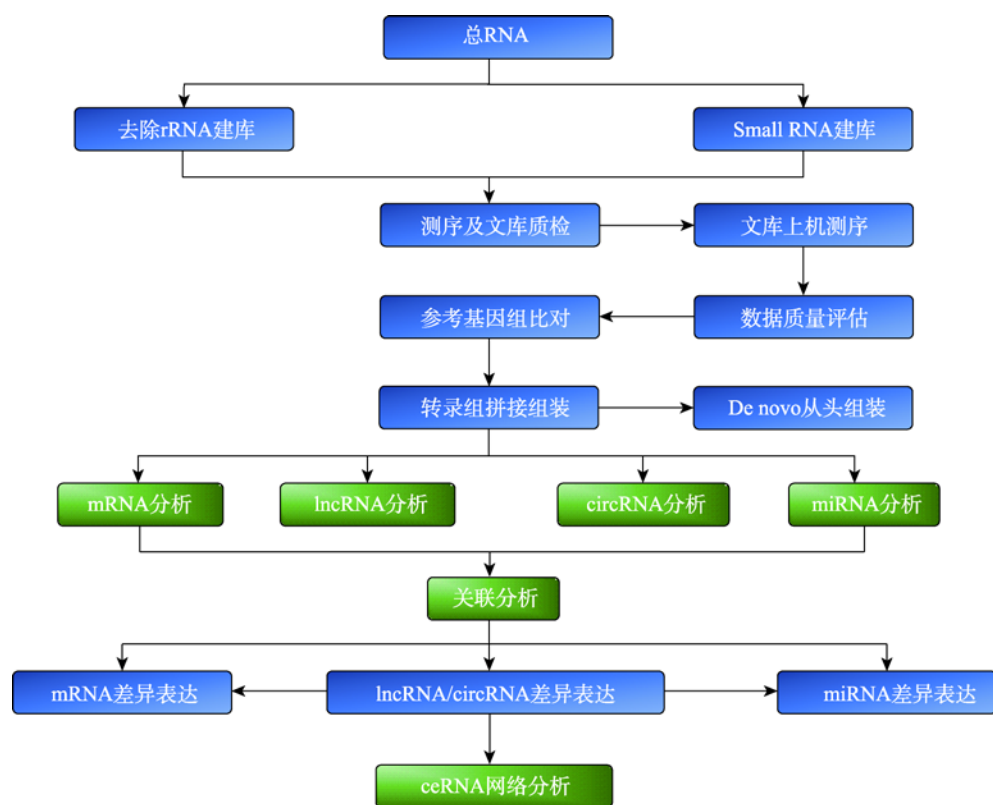


图1 全转录组测序一般流程

Fig. 1 General process for whole transcriptome sequencing

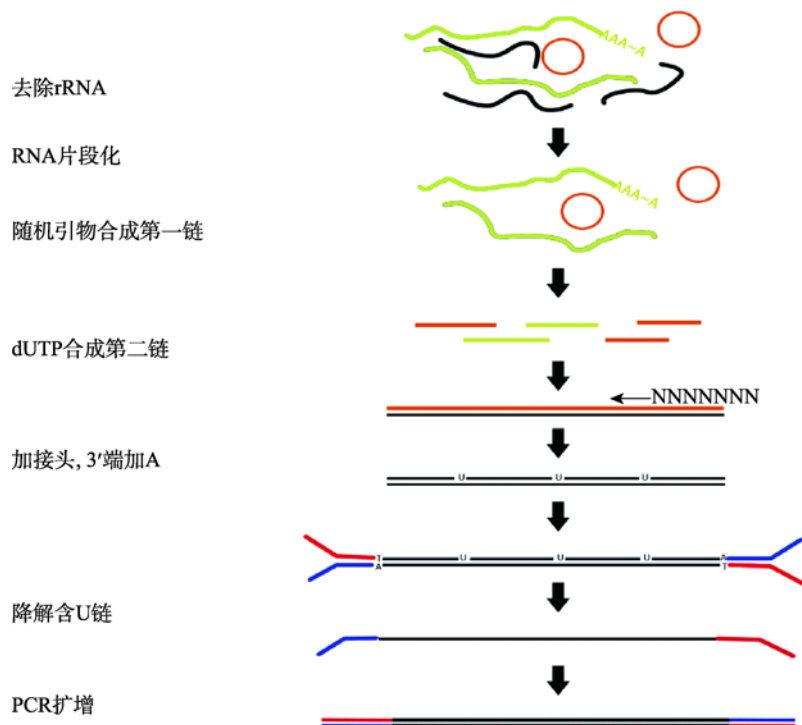


图2 链特异性建库流程

Fig. 2 The construction process for chain-specific library

合成时，用 dUTP 代替 dTTP，使链上布满 U 位点；(4)在 3'端加 A，加接头；(5)使用 USER 酶在尿嘧啶位置上产生一个单核苷酸缺口，借此消化掉第二条链，只保留第一条链；(6) PCR 扩增。小 RNA 建库的具体步骤如下(图 3)：(1)在 3'端加接头；(2)在 5'端加接头；(3)反转录扩增；(4)用凝胶纯化富集扩增的 cDNA 后测序。在测序时，理论上数据量越大越利于后续低丰度基因的完整组装，但实际上并非数据量越大越好，需要根据物种情况及相关研究决定数据量的大小。

2.2.2 原始数据预处理

测序产生的数据是一系列不能直接使用的原始数据，主要是 FASTQ 格式的读段(reads)。其中除实验所需的碱基质量信息之外，还包括测序仪器名称、上机次数、试剂型号等信息。原始数据在建库过程或者测序过程中会产生大量低质量读段、较短的序列、含 N 序列甚至是一些污染序列(如细菌基因序列)^[13]、接头序列、载体序列等，因此在数据分析之前，应对原始数据进行过滤、剪切和校正，以确保后期读段定位、转录本组装、基因定量等流程的顺利

开展。目前可用的质控软件是 FastQC 和 NGSQC^[14]、Qualimap2^[15]、HTQC^[16]、QCchain^[17]、almostSignificant^[18]、fastq-clean^[19]及 FaQCs^[20]等，最常用的是 FastQC 和 HTQC。FastQC 可作为一个单独的 JAVA 程序，在速度上远超其他工具，极短时间内就可以运行数千万的读段，输入文件的选择性也较大，如压缩或未压缩的 EASTQ 或 SAM/BAM 文件。FastQC 除能列出读段的数目及质量编码以外，还能可视化报告碱基内容及质量、读段长度和重复序列等信息。在实际操作中，可联合使用上述质控软件以达最佳过滤效果。

2.2.3 读段定位及转录本重组装

由原始数据进行深度清理及质量控制后获得待分析数据(clean reads)，需要通过一系列软件将读段比对到参考基因组或者转录本上，并根据实际定位情况进行转录本组装^[21]。在全转录组数据分析时，选择比对程序时还需考虑剪接比对(unsplliced aligners)情况^[22]。当生物体没有内含子或进行 miRNA 测序时，可以使用 Quality (MAQ)^[23]、Burrow-Wheeler Aligner (BWA)^[24]和 Bowtie^[25]等连续比对软件，这些

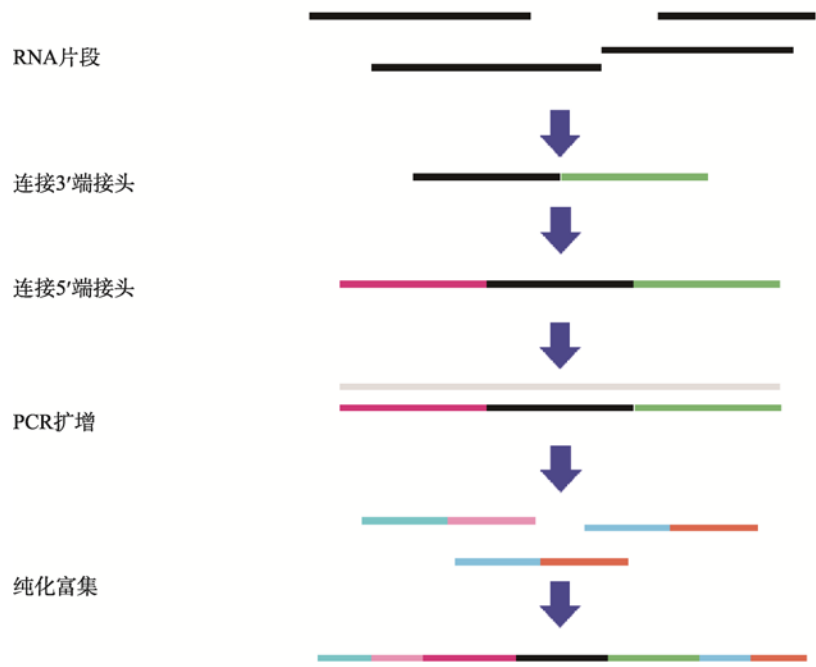


图 3 小 RNA 建库流程
Fig. 3 The construction process for small RNA library

方法用于识别已知外显子和接头,不能识别由可变剪切所产生的新型外显子。但当读段定位至外显子—外显子的交界接头处或作图到有内含子的基因组时,则必须使用剪接比对程序,如 TopHat^[26]、Map-Splice^[27]、STAR^[28]和 GSNAP^[29]等。可先将读段分成两个短片段再参与定位,同时还可记录分割数据以供查找后续转录本,该方法能够识别由可变剪切产生的新转录本。一般情况下,先利用 Bowtie 进行初步匹配,用匹配成功的 reads 来获取基因组覆盖区域,预测剪接点(splice junction)局域。然后利用 Tophat 将未成功匹配的 reads 划分到 splice junction 序列上。如果研究物种没有参考基因组或转录组时,需要将读段自行组装成可用的参考序列,然后将所有测序读段通过从头组装生成重叠群和单一序列,再进行后续数据分析。目前,有多种组装程序可供分析使用^[30,31],如 Newbler^[32]、MIRA^[33]、CAP3^[34]、Seq-Man^[35]、TGICL^[36]、stackPACK^[37]、Velvet^[38]、AbySS^[39]、ALLPATHS2^[40]、Oases^[41]、SOAPdenovo^[42]、Multiple-k method^[43]、Scaffolding using translation mapping (STM)、Trinity^[44]和 PCAP^[45]等。

2.2.4 表达量化及差异分析

经过读段定位后,可根据读段在转录本上的分布情况预测基因丰度。一般来说,通过软件分析获得的注释文件中会含有转录本分布信息,能够通过分析读段的匹配情况来识别新的转录本。在高等动物的生命活动调控过程中,存在可变剪切(alternatively splice, AS)等一系列复杂的调控系统,致使全部转录物不能直接对应基因。此外,还存在不同转

录本外显子共享、读段不能跨越多个剪接位点等问题。因此,需要先进行可变剪接识别^[46]。现阶段,常用转录本识别软件包括 Stringtie^[47]、cufflinks^[48]、CIDANE^[49]、GRIT^[50]、TransComb^[51]、iReckon^[52]、SLIDE^[53]、Montebello^[54]、Augustus^[55]、IsoLasso^[56]、Scripture^[57]、Traph^[58]和 MITIE 等。人(*Homo sapiens*)和斑马鱼(*Danio rerio*)等基因信息注释完整的物种可直接进行基因预测定量分析,但对于一些基因注释信息不完整的物种需要先进行转录本的预测。目前已完成的大量转录组测序数据快速完善了遗传数据库,反复检验校对了基因注释信息,推动了分子生物学的研究进展。

测序过程中测序深度、基因片段大小、运用算法、实验批次等因素极易造成误差,所以在定量时应使用标准化的方法消除差异,最常用的样本内标准化方法包括 RPKM^[59]、FPKM、TMP^[60]和 KPKM 等。根据比对到基因上的 reads 或者 fragments 数目,进行基因长度和测序总量归一化后,即可统计表达量。常用的转录本定量软件包括 HTSeq^[61]、featureCounts^[62]、StringTie、Cufflinks、RSEM^[63]、Sailfish^[64]、kallisto^[65]、IsoLasso 和 NURD^[66]等。为确定在不同条件或平台的表达差异情况,经过量化和标准化表达水平的基因仍需利用测序覆盖度估测表达丰度分布^[67,68]。

RNA-seq 数据的差异表达(differential expression, DE)分析往往以单变量的方式进行,很难拟合出一个普遍适用的统计模型,因此根据生物统计学原理将 DE 分析软件进行下列分析对比(表 1)。为使表达分布符合统计方法的假设, RNA-seq 数据会被某种方

表 1 差异表达分析软件

Table 1 Softwares for differential expression analysis

| 软件名称 | 所属类型 | 原理方法 | 特点 |
|------------------------|----------------|-----------|----------------|
| edgeR | R-Bioconductor | 负二项计数 | 高保守性 |
| DESeq | R-Bioconductor | 负二项计数 | 高保守性,低假阳性 |
| DESeq2 | R-Bioconductor | 负二项计数 | 保守性较高 |
| tweedDESeq | R-Bioconductor | 特威迪分布族计数 | 普适性差 |
| ebSeq | R-Bioconductor | 异构体反褶积 | 可反应差异表达的异构体 |
| Limma ^[73] | R-Bioconductor | 连续数据的线性模型 | 需要将计数转换成连续的值 |
| SAMSeq ^[74] | R 包 | 非参数检验 | 需要多次重复 |
| Cuffdiff | Linux 命令行 | 异构体反褶积 | 可反应差异表达的异构体及基因 |

式归一化。按照归一化在 DE 分析中的前后顺序可将常用软件分为两类：第一类是先进行标准化处理再通过统计学原理计算表达差异，包括 TMM^[69]、DESeq^[70]、PoissonSeq^[71]和 UpperQuartile^[72]等；第二类是将归一化作为 DE 分析的步骤，在处理时直接执行归一化，如 FPKM、RPKM 和 TPM，但可信度较差。

2.3 基因功能分析

生物信息学主要用数学及统计学的方法对生物信息进行存储、分配、检索及分析，是一门集合生命科学、自然科学与信息科学的新学科。目前，生物信息学在生命科学研究中应用十分广泛，在全转录组研究中也占据重要地位。在全转录组学分析中主要以差异基因的功能富集、候选基因的功能预测、调控网络的构建等分析为主。当前以 GO、COG 及 KEGG 数据库为基础的分析工具包、软件和网站众多，如基于 R 语言的 GSA、PADOG、SAFE、Globaltest、Sigpathway、GAGE、GSVA、PLAGE、ZSCORE、SSGSEA、MRGSE、ANCOVA、CAMERA、SPIA、TopoGSA、ToPASeq、NetGSA、NEA 和 GOGANPA 等工具包；DAVID、GOstat、GenMAPP、GOMiner、Onto-Express、EnrichNet 和 NOA 等网站；MetaCore、Cytoscape 和 GSEA 等软件。以上分析软件各有千秋，功能也不尽相同，在数据处理与分析时，应根据实验目选择合适的方法。如需更系统地反映分子调控机制，还可结合其他基因组学手段如全基因组测序、RNA 甲基化、表观修饰等数据进行联合分析。

3 全转录组学技术在畜牧业中的应用

3.1 全转录组测序在猪中的应用

猪(*Sus scrofa*)的饲养对畜牧业发展乃至民生都至关重要。不同品种猪的肌纤维类型对猪肉品质有很大影响，但潜在的分子机制仍不清楚。近年来，有关猪非编码 RNA 的功能研究开展得很多。为探究 miRNA 在猪骨骼肌中的作用，Mcdaneld 等^[75]分别选取了增殖中的卫星细胞(4~代)、胚胎(60、90、105 日龄)、出生胎儿和成年猪组织进行了全转录组研究，

发现了 12 个新型 miRNA 与肌肉生长发育密切相关，其中肌肉特异性 miR-206 在卫星细胞中近乎缺失，但在其他发育阶段表达量却很高；miR-1 在成年猪中的表达丰度最高；miR-133 在胚胎期和初生时的丰度极低甚至检测不到，但 miR-368、miR-376 和 miR-423-5p 在初生猪中极高；miR-432 表达量在胚胎发育早期最高，随后逐渐降低。该研究是对猪胚胎发育期骨骼肌 miRNA 较为全面的转录组分析，为深入探究猪骨骼肌 miRNA 调控机制提供了理论基础。沈一飞^[76]利用 RNA-seq 和 Small RNA-seq 技术对瘦肉型约克夏猪和脂肪型金华猪进行了甲状腺组织的 mRNA、lncRNA 和 miRNA 共表达鉴定与比较，结果发现差异表达 mRNA 492 个，差异表达 lncRNA 48 个和 miRNA 18 个。通过功能分析和聚类，共有 256 个 RNA(其中 18 个 miRNA、1 个 lncRNA)参与到同一个调控网络中。在整个调控网络中，发现 ssc-miRNA-221-5p、ssc-miRNA-708-5p、ssc-miRNA-532-3p 和 novel_12 等发挥重要调控作用。Li 等^[77]对从胎儿期到成年期间获得的组织混合物制备的 10 个小 RNA 测序文库中的荣昌猪(Rongchang pigs) miRNA 进行全面检测，通过哺乳动物 miRNA、前体发夹(pre-miRNA)、高覆盖率猪基因组装配(2009 年 4 月)和表达序列标签(EST)的分析，将猪 miRNAome 的所有组成部分扩展到 867 个 pre-miRNAs (623 个基因组坐标) 编码 1004 个 miRNA，其中 777 个是独特的。对 47 个组织特异性样品中选定的 30 个 miRNA 进行 qRT-PCR 实验，发现测序数据和试验结果一致。Sun 等^[78]对长白猪(Landrace)和兰塘猪(Lantang pigs)背最长肌进行了全转录组测序研究，从 22 469 个编码转录物中筛选出 547 个差异表达 mRNA，通过生物信息学分析挖掘出与肌肉生长发育相关的 17 个基因。此外还发现差异表达的 5566 个 lncRNA 和 4360 个 circRNA。其中，3376 个 lncRNA 和 1401 个 circRNA 在 Lantang 文库中上调表达，而 1590 个 lncRNA 和 2959 个 circRNA 下调表达。通过结合匹配的 miRNA 谱分析测序数据，鉴定出 26 种参与 ceRNA 网络的海绵载体，包括 19 种 lncRNA、40 种 circRNA 和 9 种 mRNA。全转录组研究提供了一种全新的分析方法，对解析猪肌肉生长发育规律和疾病发生机制具有重要意义。

3.2 全转录组测序在禽类中的应用

我国禽类遗传资源丰富, 地方品种各具特色, 但由于总体生产水平较低, 许多优良性状利用效率不高。随着分子生物学和生物信息学的发展与融合, 通过各种技术和手段, 已鉴定出一批与生长、繁殖等重要经济性状相关的分子标记和候选基因。Yu 等^[79,80]选取了 6 和 10 日龄的鸡胚左(L)、右(R)卵巢作为样品, 通过全转录组测序技术对卵巢的退化进行了研究。在 6R 样品中产生 31 066 414 个序列读数, 在 6L 样品中产生 31 900 200 个序列读数, 在 10R 样品中产生 31 400 070 个序列读数, 在 10L 样品中读取 35 504 746 个序列。这些数据中, 大约 73.33% 的序列可以定位到鸡胚卵巢的参考基因组序列上。鸡胚卵巢的发育受许多基因和信号通路的调控, 通过对差异表达基因比对和通路功能聚类发现了 22 个与卵巢发育和退化相关的基因。其中转录水平排名前 20 的卵巢基因可能与碳代谢、吞噬体及钙信号传导密切相关。Glazov 等^[81]在已发现的 miRNA 基础上进行了深度挖掘, 将 3 个小 RNA 文库同时比较并严格区分真正 miRNA 前体与结构相似的 RNA, 共鉴定到 361 个新的 miRNA、88 个新的 miRNA 候选物、18 个 mirtrons (包括 6 个新的非典型 mirtron 候选物) 和 21 个 mirtron 候选物。为识别潜在的禽类特异性 miRNA, 同时与人、狗(*Canis lupus familiaris*)、斑马鱼、爪蟾(*Xenopus laevis*) 和蜥蜴(*Bachia Oxyrhinas*) 进行保守进化分析, 结果表明只有 6 种 miRNA 在非禽类脊椎动物中具有保守性, 剩余 miRNA 可能具有鸟类和/或鸡系特异性。Li 等^[80]分别从孵育 10 天、12 天、14 天和 18 天的白来航鸡(White Leghorns)蛋胚中收集骨骼肌(胸大肌), 经过 RNA-seq 技术测序并与已发布的数据库进行比对, 筛选到 281 个新型基因间 lncRNA, 对这些 lncRNA 进行保守性分析, 并利用 UCSC 数据库评分, 结果表明以上 lncRNA 的保守性均高于随机非编码序列, 但远低于蛋白质编码基因。该研究是首例有关白来航鸡骨骼肌 lncRNA 的分析, 鉴定出的新型 lncRNA 极大丰富了鸡 ncRNA 数据库。

近年来, 鹅(*Anser cygnoides orientalis*)以其适应性强、生长快、营养成分丰富和投入要求低而备受

关注。此前, Kang 等^[82]通过抑制性消减杂交(suppression subtractive hybridization, SSH)方法鉴定了一些与鹅从产卵阶段到产蛋阶段繁殖相关的差异表达基因, Guo 等^[83]也使用相同的方法在产蛋阶段和育雏阶段发现了若干差异表达的基因。Xu 等^[84]采用短读序列技术(Illumina)对 10 只 380 日龄的雌性浙东白鹅(Goose)的卵巢样本进行了从头转录组装配, 使用 Illumina RNA-seq 和 DGE 深度测序并绘制出鹅卵巢组织的转录组图谱, 得到 67 315 996 个 100 bp 的短读数, 组装成 130 514 个独特序列。基于已知蛋白质的 BLAST 结果, 分析鉴定到 52 642 个目标序列。该研究分析了鹅产蛋、育雏期间的转录变化情况, 鉴定出大量与卵泡发育和生殖相关的基因。

病理性肥胖是鸭养殖业所面临的重要问题之一, 其分子机制仍然未知。Chen 等^[85]对家养鸭与野鸭两个品种腹部脂肪进行转录组测序分析, 预测了 23 699 个未注释基因, 确定了 753 个差异表达基因。在北京鸭(*Anatinae*)中, 一些与脂质代谢的相关基因(*IGF2*、*FABP5* 和 *BMP7* 等)和致癌基因(*RRM2*、*AURKA* 和 *CYR61* 等)上调表达, 而与肿瘤抑制和免疫相关的基因(*TNFRSF19*、*TNFAIP6*、*IGSF21* 和 *NCF1* 等)被下调, 这些数据表明鸭的肿瘤发生可能与病理性肥胖密切相关。此外, 发现 280 576 个单核苷酸变异在两个品种之间存在差异, 包括 8641 个异构变异, 富含参与脂质和免疫相关途径的基因, 表明与鸭的代谢功能和免疫相关功能密切相关。

3.3 全转录组测序在反刍动物中的应用

近年来国内牛羊肉市场需求不断增加, 这对牛羊的育种和养殖工作提出了更高的要求, 只有充分了解牛羊生长发育与繁殖等性状的发生机制, 才能提高生产效率, 全转录组学则为其提供了全新的研究手段。Di 等^[86]在滩羊(Tan)和小尾寒羊(STH)的卵巢中鉴定出 483 个 miRNA (包括 97 个已知的、369 个保守的和 17 个预测的新 miRNA)。基于 KEGG 分析, 一些差异表达 miRNA 的靶基因参与生殖激素相关途径(如类固醇生物合成、雄激素和雌激素代谢和 GnRH 信号传导途径)以及卵泡和黄体发育相关途径, 这对绵羊的繁育具有重要意义。Chang 等^[87]使用新一代测序技术(Solexa 高通量测序技术)研究了

绵羊黄体期卵巢组织, 鉴定出 267 种新的 miRNA, 并利用 qRT-PCR 和 Northern 印迹证实了在绵羊卵巢和睾丸中表达的一种新型 miRNA (ovis_aries_ovary_m0033_3p)。根据序列和结构的一致性, 推测 ovis_aries_ovary_m0033_3p 具有类似于 hsa-miR-214-3p 的功能, 能够参与细胞存活、胚胎发育、繁育生殖和卵巢癌抗性的精细调节。张世芳等^[88]采用 Solexa 技术对 5 头特克赛尔羊(Texel)进行 miRNA 深度测序, 获得了 16 532 850 条原始序列读数。通过与哺乳动物成熟 miRNA 数据库、miRNA 前体序列、绵羊基因组数据库的比对分析, 更新 miRNA 前体序列库至 1529 条, 编码的 miRNA 成熟体序列增至 1999 条。Miao 等^[89]对道赛特绵羊(Dorset)和小尾寒羊(Han)卵巢组织测序, 鉴定出可能参与繁殖力调节的候选基因, 这些候选基因参与各种细胞活动, 如代谢级联、催化功能和信号转导。此外, 通过 miRNA 谱分析鉴定了每组绵羊特有的特异性 miRNA, 发现若干与生殖力调控相关的 miRNA。

Billerey 等^[90]检测了 9 头利木赞牛犊(Limousin)的胸肌样本, 每个文库约获得 14~45 百万个配对末端读数, 发现 418 种 lincRNA (large intergenic non-coding RNAs), 与已知的 10 775 种蛋白编码基因存在显著差异。Sun 等^[91]利用 Ribo-Zero RNA-seq 技术深度剖析了胚胎、犊牛和成年牛骨骼肌的转录组谱, 发现犊牛和成年牛之间的表达水平高度相关。在胚胎期有数百个基因显著表达, 但在出生后至少减少了 10 倍, 表明这些基因在肌肉发育中具有潜在作用。此外, 该研究首次分析了牛骨骼肌中全部转录异构体, 发掘出 36 694 个新型异构体, 检测到 185 036 个 SNP 和 12 428 个短插入缺失(InDel)位点。研究发现可变剪接事件、SNP 和 InDel 的数量在胚胎中比在犊牛和成年牛中更多, 这表明基因表达在胚胎中最活跃。Cánovas 等^[92]通过转录组学测序对 7 头荷斯坦奶牛(Holstein cow)的乳样品进行了 SNP 筛选, 共检测到 19 175 个差异表达基因, 100 734 个 SNP 位点, 其中 33 045 个与荷斯坦奶牛 SNP 位点重合, 这些 SNP 位于泌乳期间表达基因的编码区中, 可用于荷斯坦奶牛乳用性状的基因变异分析和关联研究。

在梅花鹿(*Cervus nippon taiouanus*)的育种研究中, Yao 等^[93]对鹿茸进行了转录组测序, 组装出

89 001 个独特序列(平均大小 450 bp), 发现了一些高度表达的基因参与鹿茸快速生长的调节, 包括转录因子、信号分子和细胞外基质蛋白。这些数据是当前梅花鹿最全面的序列资源, 为鹿的分子遗传学和功能基因组学的研究提供了理论基础。

3.4 全转录组测序在马属动物中的应用

全转录组研究在马属动物中开展得较少, 目前主要是通过转录组学测序技术进行基因挖掘、注释和功能预测。Xie 等^[94]从头组装了驴(*Equus africanus asinus*)白细胞的转录组, 鉴定出 264 714 个不同序列, 预测了 38 949 个蛋白质片段。通过比较驴、马(*E. caballus*)和野马(*E. przewalskii*)的蛋白质序列, 将驴蛋白片段与哺乳动物表型相关联。通过比较驴和马的外耳性状相关蛋白, 鉴定出 3 种与耳形大小相关的蛋白质 HIC1、PRKRA 和 KMT2A。Scott 等^[95]通过转录组测序对马的 lncRNA 进行注释, 发现了 20 800 新型转录本, 证明了 lncRNA 独有的特征, 包括低表达、低外显子多样性和低水平的序列保守性。该研究结果所提供的候选基因可作为日后 lncRNA 注释的基准。

近年来发表了大量关于马运动机能、骨骼发育的文献报道, 一部分是通过马组织的 RNA-seq 数据改进蛋白质编码基因的结构注释, 另一部分是对 RNA 序列的分析。如 Park 等^[96]对 6 匹纯种马运动前后的血液和肌肉进行全转录测序, 通过与前人的研究对比, 发现超过 19 417 个新型单基因簇, 鉴定出 189 973 个单核苷酸位点变异(single nucleotide variants, SNV)。使用差异表达分析, 确定了多个运动调节基因: 血液中有 62 个上调基因和 80 个下调基因, 肌肉中有 878 个上调基因和 285 个下调基因。结果表明, 在差异表达的基因中有 91 个转录因子编码基因, 其中包括 56 个功能未知的转录因子候选物可能与早期调节运动机制相关; 此外, 还发现了新型 RNA 表达模式: 同一基因的不同可变剪接形式在运动前后表现出反向表达模式。该研究首次提供了马转录组数据和较为全面的分析结果, 包括运动前后表达的基因, 以及与运动相关的候选基因: 6 个运动相关基因和 91 个早期调节转录因子, 3 个高 SNV 密度的基因, 以及 4 个交替表达的剪接体。

3.5 全转录组测序在其他特种经济动物中的应用

特种经济动物养殖已成为调整农村产业结构、发展特色经济的新亮点,为了发挥皮毛的最大经济价值,科研人员开始研究关于被毛颜色的调控机制。牛晓艳等^[97]首先对不同毛色的獭兔(*Rex rabbit*)进行测序和差异基因分析,找到12408个差异表达基因,然后通过KEGG分析将得到的差异基因聚类富集到相关代谢通路上,结果发现8个与黑素细胞分化相关的差异基因。宋兴超^[98]利用RNA-seq技术对水貂(*Lutreola*)被毛色素沉积机理进行了研究,鉴定出不同时期被毛黑素含量的变化,并根据水貂皮肤组织中成熟黑素细胞的分布特点,开展SNPs检测,将不同基因突变体与水貂毛色表型进行关联分析。mRNA定量表达验证结果表明,*KITLG*、*LEF1*、*DCT*、*TYRP1*、*PMEL*、*TYR*、*Myo5a*、*Rab27a*和*SLC7A11*等基因参与了黑素细胞发育、黑素小体前体形成、黑素小体转运和真黑和褐黑色素合成等生物学过程。

4 结语与展望

全转录组学以其精准、系统、直观的技术优势为畜禽重要经济性状功能基因的挖掘、鉴定与验证提供了新的技术平台和手段,并已广泛地运用在临床医学、药学、生物学、水产学和农林学等多个领域,为人类疾病研究、新药研发和动植物育种等开辟了新的思路。但是,全转录组学在畜牧领域的研究较其他领域而言起步较晚,研究还不够深入,尤其在羊上,转录组的研究目前还主要集中在小RNA测序和基因注释上。本研究团队将开展绵羊全转录组研究,并对其生长性状、肉用性状等重要经济性状进行解析和应用。

参考文献(References):

- [1] Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. *Nature*, 2000, 405(6788): 827–836. [DOI]
- [2] Lindberg J, Lundberg J. The plasticity of the mammalian transcriptome. *Genomics*, 2010, 95(1): 1–6. [DOI]
- [3] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 2009, 10(1): 57–63. [DOI]
- [4] Wong ML, Medrano JF. Real-time PCR for mRNA quantitation. *Biotechniques*, 2005, 39(1): 75–85. [DOI]
- [5] Liu C, Li X, Chen LL. Methods for genome-wide characterization of long noncoding RNAs. *Chem Life*, 2016, (6): 745–752. [DOI]
- [6] Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. *Genome Biology*, 2011, 12(2): R16. [DOI]
- [7] Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. *Cell*, 2009, 136(4): 642–655. [DOI]
- [8] Chen LL. Linking long noncoding RNA localization and function. *Trends Biochem Sci*, 2016, 41(9): 761–772. [DOI]
- [9] Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael G, Chen LL. Long noncoding RNAs with snoRNA ends. *Mol Cell*, 2012, 48(2): 219–230. [DOI]
- [10] Zhang Y, Zhang X, Chen T, Xiang J, Yin Q, Xing Y, Zhu S, Yang L, Chen L. Circular intronic long noncoding RNAs. *Mol Cell*, 2013, 51(6): 792–806. [DOI]
- [11] Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*, 2014, 159(1): 134–147. [DOI]
- [12] Qi YX, Liu YB, Rong WH. RNA-Seq and its applications: a new technology for transcriptomics. *Hereditas(Beijing)*, 2011, 33(11): 1191–1201.
祁云霞, 刘永斌, 荣威恒. 转录组研究新技术: RNA-Seq 及其应用. 遗传, 2011, 33(11): 1191–1201. [DOI]
- [13] Hong QY, Bi XJ, Wang DN, Li ZZ, Yu H, Xia NS, Li SW. Research progress on RNA-Seq technology. *Chin J Biochem Pharm*, 2017, 37(6): 443–448.
洪奇阳, 毕行建, 王大宁, 李子真, 俞海, 夏宁邵, 李少伟. 转录组测序技术研究进展. 中国生化药物杂志, 2017, 37(6): 443–448. [DOI]
- [14] Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, 2012, 7(2): e30619. [DOI]
- [15] Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 2016, 32(2): 292–294. [DOI]
- [16] Yang X, Liu D, Liu F, Wu J, Zou J, Xiao X, Zhao F, Zhu B. HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics*, 2013, 14: 33. [DOI]
- [17] Zhou Q, Su X, Wang A, Xu J, Ning K. QC-Chain: fast and

- holistic quality control method for next-generation sequencing data. *PLoS One*, 2013, 8(4): e60234. [DOI]
- [18] Ward J, Cole C, Febrer M, Barton GJ. Almost significant: simplifying quality control of high-throughput sequencing data. *Bioinformatics*, 2016, 32(24): 3850–3851. [DOI]
- [19] Zhang M, Sun H, Fei Z, Zhan F, Gong X, Gao S. Fastq_clean: an optimized pipeline to clean the Illumina sequencing data with quality control. In: IEEE International Conference on Bioinformatics and Biomedicine. 2015, 44–48. [DOI]
- [20] Lo CC, Chain PSG. Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*, 2014, 15(1): 366. [DOI]
- [21] Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods*, 2011, 8(6): 469–477. [DOI]
- [22] Yang IS, Kim S. Analysis of whole transcriptome sequencing data: workflow and software. *Genomics Inform*, 2015, 13(4): 119–125. [DOI]
- [23] Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*, 2008, 18(11): 1851–1858. [DOI]
- [24] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 2009, 25(4): 1754–1760. [DOI]
- [25] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 2009, 10(3): R25. [DOI]
- [26] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 2009, 25(9): 1105–1111. [DOI]
- [27] Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM, MacLeod JN, Chiang DY, Prins JF, Liu J. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res*, 2010, 38(18): e178. [DOI]
- [28] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 2013, 29(1): 15–21. [DOI]
- [29] Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, 2010, 26(7): 873–881. [DOI]
- [30] Kumar S, Blaxter ML. Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics*, 2010, 11: 571. [DOI]
- [31] Garg R, Patel RK, Tyagi AK, Jain M. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res*, 2011, 18(1): 53–63. [DOI]
- [32] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437(7057): 376–380. [DOI]
- [33] Burlibasa C, Vasiliu D, Vasiliu M. Genome sequence assembly using trace signals and additional sequence information. In: German Conference on Bioinformatics. 1999, 45–56. [DOI]
- [34] Huang X, Madan A. CAP3: a DNA sequence assembly program. *Genome Res*, 1999, 9(9): 868–877. [DOI]
- [35] Swindell SR, Plasterer TN. SEQMAN. Contig assembly. *Methods Mol Biol*, 1997, 70: 75–89. [DOI]
- [36] Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 2003, 19(5): 651–652. [DOI]
- [37] Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Pitsyn AA. A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. *Genome Res*, 1999, 9(11): 1143–1155. [DOI]
- [38] Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*, 2008, 18(5): 821–829. [DOI]
- [39] Jackman SD, Birol İ. Assembling genomes using short-read sequencing technology. *Genome Biol*, 2010, 11(1): 202–202. [DOI]
- [40] MacCallum I, Przybylski D, Gnerre S, Burton J, Shlyakhter I, Gnirke A, Malek J, McKernan K, Ranade S, Shea TP, Williams L, Young S, Nusbaum C, Jaffe DB. ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. *Genome Biol*, 2009, 10(10): R103–R103. [DOI]
- [41] Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 2012, 28(8): 1086–1092. [DOI]
- [42] Li R, Li Y, Kristiansen K, Wang J, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics*, 2008,

- 24(5): 713–714. [DOI]
- [43] Surget-Groba Y, Montoya-Burgos JI. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res*, 2010, 20(10): 1432–1440. [DOI]
- [44] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, 2011, 29(7): 644–652. [DOI]
- [45] Huang X, Wang J, Aluru S, Yang SP, Hillier L. PCAP: a whole-genome assembly program. *Genome Res*, 2003, 13(9): 2164–2170. [DOI]
- [46] Huh JW, Kim YH, Park SJ, Kim DS, Lee SR, Kim KM, Jeong KJ, Kim JS, Song BS, Sim BW, Kim SU, Kim SH, Chang KT. Large-scale transcriptome sequencing and gene analyses in the crab-eating macaque (*Macaca fascicularis*) for biomedical research. *BMC Genomics*, 2012, 13: 163. [DOI]
- [47] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*, 2015, 33(3): 290–295. [DOI]
- [48] Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*, 2011, 27(17): 2325–2329. [DOI]
- [49] Canzar S, Andreotti S, Weese D, Reinert K, Klau GW. CIDANE: comprehensive isoform discovery and abundance estimation. *Genome Biol*, 2016, 17: 16. [DOI]
- [50] Boley N, Stoiber MH, Booth BW, Wan KH, Hoskins RA, Bickel PJ, Celniker SE, Brown JB. Genome-guided transcript assembly by integrative analysis of RNA sequence data. *Nat Biotechnol*, 2014, 32(4): 341–346. [DOI]
- [51] Liu JT, Yu T, Tao J, Li GJ. TransComb: genome-guided transcriptome assembly via combining junctions in splicing graphs. *Genome Biol*, 2016, 17: 213. [DOI]
- [52] Mezlini AM, Smith EJ, Fiume M, Buske O, Savich GL, Shah S, Aparicio S, Chiang DY, Goldenberg A, Brudno M. iReckon: simultaneous isoform discovery and abundance estimation from RNA-seq data. *Genome Res*, 2013, 23(3): 519–529. [DOI]
- [53] Li JJ, Jiang CR, Brown JB, Huang H, Bickel PJ. Sparse linear modeling of next-generation mRNA sequencing (RNA-Seq) data for isoform discovery and abundance estimation. *Proc Natl Acad Sci USA*, 2011, 108(50): 19867–19872. [DOI]
- [54] Hiller D, Wong WH. Simultaneous isoform discovery and quantification from RNA-seq. *Stat Biosci*, 2013, 5(1): 100–118. [DOI]
- [55] Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*, 2006, 34(web server issue): 435–439. [DOI]
- [56] Li W, Feng J, Jiang T. IsoLasso: A LASSO regression approach to RNA-seq based transcriptome assembly. *J Comput Biol*, 2011, 18(11): 1693–1707. [DOI]
- [57] Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Xian A, Fan L, Koziol MJ, Gnirke A, Nusbaum C. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol*, 2010, 28(5): 503–510. [DOI]
- [58] Tomescu, Alexandru I, Kuosmanen, Anna, Makinen, Veli, Rizzi R. A novel min-cost flow method for estimating transcript expression with RNA-Seq. *BMC Bioinformatics*, 2013, 14(Suppl.5): S15. [DOI]
- [59] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 2008, 5(7): 621–628. [DOI]
- [60] Pachter L. Models for transcript quantification from RNA-Seq. 2013. [DOI]
- [61] Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 2015, 31(2): 166–169. [DOI]
- [62] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 2014, 30(7): 923–930. [DOI]
- [63] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 2011, 12: 323. [DOI]
- [64] Rob P, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol*, 2013, 32(5): 462–464. [DOI]
- [65] Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, 2016, 34(5): 525–527. [DOI]
- [66] Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*,

- 2010, 11: 94. [DOI]
- [67] Ma X, Zhang X. NURD: an implementation of a new method to estimate isoform expression from non-uniform RNA-seq data. *BMC Bioinformatics*, 2013, 14: 220. [DOI]
- [68] Oshlack A, Wakefield MJ. Transcript length bias in RNA-seq data confounds systems biology. *Biol Direct* 2009, 4: 14. [DOI]
- [69] Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 2010, 11(3): 1–9. [DOI]
- [70] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*, 2010, 11: R106. [DOI]
- [71] Li J, Witten DM, Johnstone IM, Tibshirani R. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics*, 2012, 13(3): 523–538. [DOI]
- [72] Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Amer Stat Assoc*, 1984, 79(387): 516–524. [DOI]
- [73] Law CW, Chen Y, Wei S, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*, 2014, 15(2): R29. [DOI]
- [74] Li J, Tibshirani R. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. *Stat Methods Med Res*, 2011, 22(5): 519–536. [DOI]
- [75] Mcdaneld TG, Smith TP, Doumit ME, Miles JR, Coutinho LL, Sonstegard TS, Matukumalli LK, Dan JN, Wiedmann RT. MicroRNA transcriptome profiles during swine skeletal muscle development. *BMC Genomics*, 2009, 10: 77. [DOI]
- [76] Shen YF. Sequencing and characterization of mRNA, lncRNA, and miRNA in thyoid gland of Yorkshire and Jinhua Pigs[D]. Zhejiang University, 2016.
沈一飞. 约克夏猪和金华猪甲状腺组织 mRNA, lncRNA 和 miRNA 测序及功能分析[学位论文]. 浙江大学, 2016. [DOI]
- [77] Li M, Xia Y, Gu Y, Zhang K, Lang Q, Chen L, Guan J, Luo Z, Chen H, Li Y, Li Q, Li X, Jiang AA, Shuai S, Wang J, Zhu Q, Zhou X, Gao X, Li X. MicroRNAome of porcine pre- and postnatal development. *PLoS One*, 2010, 5(7): e11541. [DOI]
- [78] Sun J, Xie M, Huang Z, Li H, Chen T, Sun R, Wang J, Xi Q, Wu T, Zhang Y. Integrated analysis of non-coding RNA and mRNA expression profiles of 2 pig breeds differing in muscle traits. *J Anim Sci*, 2017, 95(3): 1092–1103. [DOI]
- [79] Jianning YU, Yan L, Chen Z, Hui LI, Ying S, Zhu H, Shi Z. Investigating right ovary degeneration in chick embryos by transcriptome sequencing. *J Reprod Dev*, 2017, 63(3): 295–303. [DOI]
- [80] Li T, Wang S, Wu R, Zhou X, Zhu D, Zhang Y. Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. *Genomics*, 2012, 99(5): 292–298. [DOI]
- [81] Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Res*, 18(6): 957–964. [DOI]
- [82] Kang B, Guo JR, Yang HM, Zhou RJ, Liu JX, Li SZ, Dong CY. Differential expression profiling of ovarian genes in prelaying and laying geese. *Poult Sci*, 2009, 88(9): 1975–1983. [DOI]
- [83] Guo J, Tang QP, Zhang SJ, Ma YH, Lu HL. Identification of broodiness-related geese genes by suppression subtractive hybridization. *Acta Veter Zootechn Sin*, 2011, 42(10): 1477–1784.
郭军, 汤青萍, 章双杰, 马月辉, 陆火林, 苏建东, 邹剑敏, 陈宽维, 李慧芳. 利用抑制消减杂交技术筛选鹅就巢行为相关基因. 畜牧兽医学报, 2011, 42(10): 1477–1484. [DOI]
- [84] Xu Q, Zhao WM, Chen Y, Tong YY, Rong GH, Huang ZY, Zhang Y, Chang GB, Wu XS, Chen GH. Transcriptome profiling of the goose (*Anser cygnoides*) ovaries identify laying and broodiness phenotypes. *PLoS One*, 2013, 8(2): e55496. [DOI]
- [85] Chen L, Luo J, Li JX, Li JJ, Wang DQ, Tian Y, Lu LZ. Transcriptome analysis of adiposity in domestic ducks by transcriptomic comparison with their wild counterparts. *Anim Genetics*, 2015, 46(3): 299–307. [DOI]
- [86] Di R, He J, Song S, Tian D, Liu Q, Liang X, Ma Q, Sun M, Wang J, Zhao W, Cao G, Wang J, Yang Z, Ge Y, Chu M. Characterization and comparative profiling of ovarian microRNAs during ovine anestrus and the breeding season. *BMC Genomics*, 2014, 15: 899. [DOI]
- [87] Chang W, Wang J, Tao D, Zhang Y, Jianzhong HE, Shi C. Identification of a novel miRNA from the ovine ovary by a combinatorial approach of bioinformatics and experiments. *J Vet Med Sci*, 2015, 77(12): 1617–1624. [DOI]
- [88] Zhang SF, Wei CH, Lu J, Zhang XN, Zhou XL, Zhang SZ, Wang GK, Cao JX, Zhao FP, Zhang L, Du LX. Identification of the microRNAome in texel sheep by deep sequencing. *Chin Anim Husb Veter Med*, 2013, 40(9):

- 19–22.
张世芳, 魏彩虹, 陆健, 张小宁, 周鑫磊, 张淑珍, 王光凯, 曹家雪, 赵福平, 张莉, 杜立新. 深度测序鉴定绵羊 microRNA 转录组. *中国畜牧兽医*, 2013, 40(9): 19–22. [DOI]
- [89] Miao X, Qin QL. Genome-wide transcriptome analysis of mRNAs and microRNAs in Dorset and Small Tail Han sheep to explore the regulation of fecundity. *Mol Cell Endocrinol*, 2015, 402: 32–42. [DOI]
- [90] Billerey C, Boussaha M, Esquerré D, Rebours E, Djari A, Meersseman C, Klopp C, Gautheret D, Rocha D. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics*, 2014, 15: 499. [DOI]
- [91] Sun X, Li M, Sun Y, Cai H, Li R, Wei X, Lan X, Huang Y, Lei C, Chen H. The developmental transcriptome landscape of bovine skeletal muscle defined by Ribo-Zero ribonucleic acid sequencing. *J Anim Sci*, 2015, 93(12): 5648–5658. [DOI]
- [92] Cánovas A, Rincon G, IslasTrejo A, Wickramasinghe S, Medrano JF. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm Genome*, 2010, 21(11–12): 592–598. [DOI]
- [93] Yao B, Yu Z, Mei Z, Liu M, Liu H, Li J. De novo characterization of the antler tip of Chinese Sika deer transcriptome and analysis of gene expression related to rapid growth. *Mol Cell Biochem*, 2012, 364(1–2): 93–100. [DOI]
- [94] Xie FY, Feng YL, Wang HH, Ma YF, Yang Y, Wang YC, Shen W, Pan QJ, Yin S, Sun YJ, Ma JY. *De novo* assembly of the donkey white blood cell transcriptome and a comparative analysis of phenotype-associated genes between donkeys and horses. *PLoS One*, 2015, 10(7): e0133258. [DOI]
- [95] Scott EY, Mansour T, Bellone RR, Brown CT, Mienaltowski MJ, Penedo MC, Ross PJ, Valberg SJ, Murray JD, Finno CJ. Identification of long non-coding RNA in the horse transcriptome. *BMC Genomics*, 2017, 18(1): 511. [DOI]
- [96] Park KD, Park J, Ko J, Kim BC, Kim HS, Ahn K, Do KT, Choi H, Kim HM, Song S, Lee S, Jho S, Kong HS, Yang YM, Jhun BH, Kim C, Kim TH, Hwang S, Bhak J, Lee HK, Cho BW. Whole transcriptome analyses of six thoroughbred horses before and after exercise using RNA-Seq. *BMC Genomics*, 2012, 13: 473. [DOI]
- [97] Niu XY, Ren KL, Cao L, Li YP, Zhen JJ, Feng GL, Huang SF. Transcriptome analysis on coat color related genes in rex rabbits. *China Herb Sci*, 2016, 36(2): 1–6.
牛晓艳, 任克良, 曹亮, 李燕平, 郑建婷, 冯国亮, 黄淑芳. 利用转录组测序方法研究獭兔毛色相关基因. *中国草食动物科学*, 2016, 36(2): 1–6. [DOI]
- [98] Song XC. Study on the mechanisms of hair pigmentation and skin transcriptome annotation in mink (neovison vison) based on high throughput RNA-sequencing[D]. Chinese Academy of Agricultural Sciences, 2016.
宋兴超. 水貂被毛色素沉积机理及基于高通量 RNA-seq 皮肤转录组注释研究[学位论文]. 中国农业科学院, 2016. [DOI]

(责任编辑: 赵要风)