



تولیات دامی

دوره ۲۲ ■ شماره ۳ ■ پاییز ۱۳۹۹

صفحه‌های ۳۴۸-۳۳۷

مسیرهای بیولوژیکی مرتبط با lncRNAهای شناخته‌شده بین ژنی در بافت شکمبه گوساله‌های مبتلا به اسیدوز

بیژن محمودی^۱، جمال فیاضی^{۲*}، هدایت‌الله روشنفکر^۳، محسن ساری^۴، محمدرضا بختیاری‌زاده^۴

۱. دانش‌آموخته دکتری، گروه علوم دامی، دانشکده علوم دامی و صنایع غذایی، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان، ملاثانی، ایران.

۲. دانشیار، گروه علوم دامی، دانشکده علوم دامی و صنایع غذایی، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان، ملاثانی، ایران.

۳. استاد، گروه علوم دامی، دانشکده علوم دامی و صنایع غذایی، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان، ملاثانی، ایران.

۴. دانشیار، گروه علوم دام و طیور، پردیس ابوریحان، دانشگاه تهران، پاکدشت، ایران.

تاریخ پذیرش مقاله: ۱۳۹۹/۰۲/۱۷

تاریخ دریافت مقاله: ۱۳۹۸/۰۹/۰۸

چکیده

هدف این پژوهش شناسایی lncRNAهای دخیل در کنترل فعالیت مسیرهای بیولوژیکی مؤثر در بروز اسیدوز بود. به این منظور دو گروه گوساله شامل گروه شاهد (۳ گوساله نر سالم) و گروه بیمار (۳ گوساله نر مبتلا به اسیدوز) به صورت مقایسه‌ای مورد مطالعه قرار گرفتند. توالی‌یابی جفتی با استفاده از پلاتفرم illumine Hiseq2500 انجام شد. از نرم‌افزار Hisat2 برای هم‌ترازی خوانش‌ها با ژنوم مرجع گاو و بسته نرم‌افزاری StringTie جهت سرهم‌بندی رونوشت‌ها استفاده شد. با استفاده از توالی‌یابی نسل بعد، ۱۶۳۶ ژن متعلق به lncRNAهای شناخته‌شده بین ژنی شناسایی شد که تغییرات بیان ۵۶ ژن معنی‌دار بود ($P \leq 0.05$). ژن‌های هم‌جوار lncRNAهای شناخته‌شده بین ژنی روی ژنوم گاو هشتاد و یک مورد شناسایی شدند. نتایج نشان داد با سطح احتمال $P \leq 0.05$ پنج مسیر بیولوژیکی Apelin signaling pathway، Renin secretion، Glucagon signaling pathway، Gap junction signaling pathway و AGE-RAGE signaling pathway در دیابتیک‌های دیابتی و activity می‌شوند. آنالیز عملکرد مولکولی این ژن‌ها نشان داد دو عملکرد مولکولی شامل gap junction channel activity و phosphatidylinositol phospholipase C activity به‌طور معنی‌دار غنی می‌شوند. برخی lncRNAها در نمونه‌های سالم و اسیدوزی بیان متفاوتی داشتند و کاهش pH به‌عنوان محرک برای فعال شدن برخی مسیرهای بیولوژیکی ترانسکریپشنی پیام عمل کرد. براساس نتایج حاصل، lncRNAهایی که تفاوت بیان معنی‌دار در گروه کنترل و اسیدوز دارند با مسیرهای مرتبط با سوخت‌وساز انرژی شکمبه و ترانسکریپشنی پیام همراه می‌باشند. از lncRNAها می‌توان به‌عنوان عامل پیش‌آگاهی‌دهنده اسیدوز و بیومارکر در اصلاح دام استفاده نمود.

کلیدواژه‌ها: الگوی بیان ژن، ایلومینیا، ترانسکریپشنی پیام، توالی‌یابی، lncRNA.

Biological pathways related to known intergenic lncRNAs in calf ruminal samples affected with acidosis

Bizhan Mahmoudi¹, Jamal Fayazi^{2*}, Hedayatollah Roshanfekr³, Mohsen Sari² and Mohammad Reza Bakhtiarzadeh⁴

1. Former Ph.D. Student, Department of Animal Science, Animal Science and Food Technology Faculty, Agricultural Sciences and Natural Resources, University of Khuzestan, Mollasani, Iran.

2. Associate Professor, Department of Animal Science, Animal Science and Food Technology Faculty, Agricultural Sciences and Natural Resources, University of Khuzestan, Mollasani, Iran.

3. Professor, Department of Animal Science, Animal Science and Food Technology Faculty, Agricultural Sciences and Natural Resources, University of Khuzestan, Mollasani, Iran.

4. Associate Professor, Department of Animal and Poultry Science, College of Aburairhan, University of Tehran, Pakdasht, Iran.

Received: November 29, 2019

Accepted: May 6, 2020

Abstract

The objective of this study was to identify known intergenic lncRNAs related to biological pathways of acidosis in Holstein calves using ruminal tissue. Two groups of healthy calves (N=3) and affected by acidosis (N=3) were compared. Paired-end sequencing method was performed using the Hiseq2500 illumine platform. Hisat2 software was used to align reads to the bovine reference genome and StringTie software package was used to assemble read files into transcripts. Using next generation sequencing, 1636 genes belonging to known intergenic lncRNAs were identified, of which 56 genes showed significant differential expression ($P \leq 0.05$). Neighbor genes of known intergenic lncRNAs were determined on bovine genome. Analysis of biological pathways and molecular function showed that five biological pathways were significantly ($P \leq 0.05$) enriched. These pathways were Apelin signaling pathway, Gap junction, Glucagon signaling pathway, Renin secretion, and AGE-RAGE signaling pathway. Moreover, two molecular functions including gap junction channel activity, and phosphatidylinositol phospholipase C activity were significantly ($P \leq 0.05$) enriched. Some lncRNAs have different expression in healthy and acidosis samples, and the decreased pH acts as a stimulus to activate some biological signaling pathways. In conclusion, it was indicated that lncRNAs with differential expression between the control group and the group affected by acidosis are associated with pathways related to rumen energy metabolism and signaling. Identified differentially expressed lncRNAs could be used as prognostic in acidosis and biomarkers or promising candidates in animal breeding.

Keywords: Gene expression pattern, illumine, lncRNA, Sequencing, Signaling.

مقدمه

در بین عوامل مختلفی که سلامت و در نتیجه عملکرد دام‌ها را تحت تأثیر قرار می‌دهند، بیماری‌های متابولیکی دارای نقش مهمی هستند. بیماری‌های متابولیکی هزینه‌های بسیار زیادی به دامداران تحمیل می‌کنند، به‌طور متوسط ۲۵ درصد از دام‌هایی که در ۶۵ روز اول دوره شیردهی حذف می‌شوند، نتیجه یک یا چند بیماری متابولیکی است [۱۵]. اسیدوز یکی از مهم‌ترین بیماری‌های متابولیکی است. مطالعات حاکی از آن است که اسیدوز و به‌ویژه اسیدوز شکمبه در گاوهای پرواری و شیری یک مشکل جدی صنعت دامپروری است و حتی در مزارع با مدیریت خوب و مکانیزه نیز به‌وفور دیده می‌شود. به‌طور متوسط ۱۱ الی ۲۹ درصد گاوهای شیری و پرواری در معرض اسیدوز قرار دارند [۱۵]. با بررسی ۱۵ گله گاو هلشتاین، گزارش شده است که اسیدوز شکمبه در ۱۹ درصد از گاوهایی که در دوران اولیه شیردهی هستند مشاهده می‌شود. این مقدار در گاوهایی که در اواسط شیردهی هستند به میزان ۲۶ درصد افزایش می‌یابد [۱]. تمایل تولیدکنندگان به استفاده حداکثری از عملکرد تولیدی دام‌های شیری و گوشتی معمولاً با افزایش مصرف مواد متراکم (دانه غلات و کنجاله‌ها و ...) اتفاق می‌افتد و در این خصوص ابتلا به بیماری اسیدوز عامل محدودکننده‌ای می‌باشد. یافتن راه‌های پیشگیری و درمان مؤثر اسیدوز برای رسیدن به تولید پایدار و تضمین سوددهی صنایع دامپروری یک ضرورت حیاتی است. رسیدن به این هدف به میزان زیادی به دانش و آگاهی از سازوکارهای مولکولی و سلولی ایجادکننده این بیماری و عوامل دخیل در آن وابسته است.

در سال‌های اخیر تکنیک‌های توالی‌یابی نقش بسیار مهمی در توضیح بسیاری از جنبه‌های پنهان بیماری‌های مختلف ایفا کرده است. توالی‌یابی ژنوم موجودات مختلف

مثل انسان، موش، برنج، کلزا و برخی حشرات و باکتری‌ها این امکان را به پژوهش‌گران داده بود تا بتوانند جهش‌های مسئول ایجاد بیماری، نشانگرهای ژنتیکی مرتبط با بیماری خاص و جایگاه‌های کنترل‌کننده صفات کمی مرتبط با بیماری‌های پیچیده را شناسایی کنند [۱۰]. با وجود این‌که این مطالعات بسیاری از سازوکارهای مولکولی دخیل در بیماری‌ها را آشکار نموده است، اما هنوز بسیاری از جنبه‌های مولکولی ایجاد بیماری‌ها ناشناخته باقی مانده است.

یکی از این زوایای پنهان سازوکارهای مولکولی دخیل در بیماری‌ها، نقش و عملکرد انواع مختلف RNAهای غیر کدکننده در تنظیم بیان ژن‌ها در سطوح مختلف نسخه‌برداری، بعد از نسخه‌برداری و در مرحله ترجمه است. RNAهای غیرکدکننده به‌طورکلی به دو گروه RNAهای غیرکدکننده کوتاه و RNAهای غیرکدکننده بلند (LncRNAs) تقسیم می‌شوند و به‌ترتیب دارای طول کوچک‌تر از ۲۰۰ و بزرگ‌تر از ۲۰۰ نوکلئوتید هستند. گروه RNAهای کوتاه شامل RNAهای کوچک مداخله‌گر (siRNAs)، میکرو RNAها (miRNAs) و piRNAها (Piwi-interacting RNA) هستند که به‌طور عمده در تنظیم بیان ژن نقش دارند [۱۰]. گروه RNAهای غیرکدکننده بلند دارای برخی از ویژگی‌های مشابه با RNAهای کدکننده، مثل وجود اثرات اپی‌ژنتیکی در بیان آن‌ها، وجود ایترون و ویرایش متناوب در ژن‌های کدکننده آن‌ها هستند. بسیاری از آن‌ها به شکل پلی‌آدنیل‌شده ملاحظه شده‌اند و در ژنوم ممکن است با ژن‌های کدکننده هم‌پوشانی داشته باشند ولی اکثر آن‌ها در نواحی بین ژنی قرار دارند [۲]. هرچند هنوز نقش RNAهای غیرکدکننده بلند به‌خوبی شناخته نشده است، ولی مطالعات صورت گرفته نشان داده‌اند که این نوع از RNAها در سطوح مختلف تنظیم بیان ژن، یعنی هم در مرحله نسخه‌برداری و هم در مرحله بعد از نسخه‌برداری

تولیدات دامی

مسیرهای بیولوژیکی مرتبط با lncRNA های شناخته شده بین ژنی در بافت شکمبه گوساله های مبتلا به اسیدوز

در مزرعه پژوهشی محمدشهر کرج، پردیس کشاورزی و منابع طبیعی دانشگاه تهران به مدت ۱۳۰ روز انجام شد. وزن گوساله ها (۵۵۰ ± ۴۰) کیلوگرم بود. شش گوساله مورد مطالعه به طور تصادفی به دو گروه سه تایی شاهد (سالم) و بیمار (مبتلا به اسیدوز تحت حاد) تقسیم شدند. برای القای اسیدوز تحت حاد، بعد از یک هفته عادت دهی، رژیم غذایی برای ۱۲ ساعت متوقف شد و سپس جیره با ترکیب ۸۰ درصد کنسانتره و ۲۰ درصد علوفه جایگزین شد. برای تأیید ایجاد اسیدوز تحت حاد در گوساله های مورد آزمایش، pH شکمبه با استفاده از pH متر کالیبره شده، اندازه گیری شد. کشتار دامها در کشتارگاه کرج راک انجام شد. سپس از بافت شکمبه نمونه هایی با ابعاد ۱×۱ سانتی متر جدا و بلافاصله بعد از جداسازی، با بافر PBS (pH= ۶/۸) شست و شو داده شده و در تیوب های ۵۰ ml به داخل تانک ازت منتقل شدند. نمونه ها به آزمایشگاه منتقل و تا زمان استخراج RNA کل در یخچال تحت دمای -۸۰ درجه سانتی گراد نگهداری شدند. از هر گوساله دو نمونه گرفته شد (به منظور اندازه گیری خطای کارکردی)، در مجموع ۱۲ نمونه (شش نمونه شاهد و شش نمونه اسیدوزی) برای مطالعه مورد استفاده قرار گرفتند. برای استخراج RNA کل از بافت شکمبه، نمونه های فریز شده تحت دمای -۸۰ درجه سانتی گراد، پودر شدند و میزان ۵۰ الی ۱۰۰ میلی گرم از نمونه های پودر شده به طور جداگانه برای گروه شاهد و بیمار وزن شده و RNA کل با استفاده از کیت ترایزول شرکت اینویترورژن (Invitrogen, CA, USA) مطابق پروتکل تولیدکننده، از نمونه های بافتی استخراج شد.

ارزیابی کیفیت شامل روش استفاده از دستگاه اسپکتوفوتومتر نانو دراپ (مدل ۲۰۰۰، ایالات متحده) و ژل آگارز ۱/۵ درصد برای تأیید کیفیت و کمیت RNA کل استخراج شده، بود. براساس نتایج حاصل از دستگاه نانو

نقش دارند [۹]. به طور کلی مطالعات مولکولی روی اسیدوز در دامها محدود است. در مطالعه ای روی تأثیر اسیدوز شکمبه بر تغییرات ترانسکریپتومی گزارش شده که در اثر القای اسیدوز تعداد ۱۷۲ ژن در بافت شکمبه تغییر بیان پیدا می کنند. مسیر متابولیکی Hemophilic Cell Adhesion Pathway در گاوهای دچار اسیدوز افزایش بیان پیدا می کند. ژن های خاصی که در این مسیر تنظیم شده اند پروتوکادین بتا چهار، هفت، ۱۴ و ۱۵ (PCDHB14, PCDHB15, PCDHB4, PCDHB7) هستند و سطح بیان چهار ژن مذکور که در چسبندگی سلول هموفیلی (hemophilic cell adhesion) نقش دارند در گاوهای مبتلا شده به اسیدوز از گاوهای شاهد بالاتر بود [۱۸]. در مطالعه ای دیگر روی تغییرات میزان mRNA سلول های اپی تلیال شکمبه در گاوهای شیری گزارش شده است که القای اسیدوز باعث افزایش بیان ژن های desmoglein 1 and IGF binding proteins می شود. افزایش بیان این ژن ها در ارتباط با رشد و تغییر ساختار اپی تلیال دانسته شده است [۹]. اما تاکنون نقش lncRNA های بین ژنی در بافت شکمبه گاو گزارش نشده است. با توجه به اهمیت نقش این دسته از ژن ها در کنترل فرایندهای سلولی و به منظور تکمیل مطالعات سلولی، مولکولی و ژنتیکی انجام شده روی اسیدوز، بررسی نقش و عملکرد RNA های غیرکدکننده بلند دخیل در این بیماری می تواند بسیار کمک کننده باشد. هدف از این پژوهش، شناسایی lncRNA های دخیل در پاسخ به اسیدوز در بافت شکمبه گوساله هلشتاین است. یافته های این پژوهش به درک بیشتر ما در مورد عملکردهای مولکولی lncRNA ها در بیماری اسیدوز شکمبه گاو کمک شایانی می کند.

مواد و روش ها

پژوهش حاضر روی شش گوساله نر هلشتاین هشت ماهه

تولیدات دامی

دوره ۲۲ ■ شماره ۳ ■ پاییز ۱۳۹۹

تحت لینوکس استفاده شد (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). بعد از شناسایی مشکلات و معایب موجود در توالی‌ها، از نرم‌افزار Trimmomatic (نسخه ۰/۳۶) برای پیرایش و فیلتر خوانش‌های نامطلوب و حذف توالی‌های با طول کم‌تر از ۴۰ نوکلئوتید استفاده شد. خوانش‌های موجود در سطریهای متناظر دو فایل fastq نمایانگر خوانش‌های جفت انتهایی مربوط به یک قطعه توالی‌یابی شده می‌باشند. رویه پیرایش و فیلترکردن خوانش‌های نامطلوب به این ترتیب بود که در ابتدا آلودگی آداپتور با استفاده از نرم‌افزار Trimmomatic (نسخه ۰/۳۶) بررسی شد و خوانش‌های آلوده حذف شدند. توالی‌های باقی‌مانده با استفاده از نرم‌افزار HISAT2 (نسخه ۲/۱/۰) روی توالی ژنومی گاو هلستاین (ARS-UCD1.2) موجود در داده پایگاه Ensemble (<https://asia.ensembl.org/index.html>) هم‌ردیف شدند و سپس با استفاده از StringTie (نسخه ۲/۱) به رونوشت‌ها سرهم‌بندی گشتند. برای شناسایی و تفکیک lncRNAها از سایر mRNAها، معیارهایی شامل حذف تمامی رونوشت‌های mRNA، حذف رونوشت‌های با طول کم‌تر از ۲۰۰ نوکلئوتید، حذف رونوشت‌های ناشناخته (Novel lncRNA)، حذف توالی با پتانسیل کدکنندگی با استفاده از نرم‌افزارهای CPC (نسخه ۲) [۱۳]، CNCI (نسخه ۲) [۲۳] و PLEK (نسخه ۱/۲) [۱۶]، در نظر گرفته شدند [۳]. توالی‌های حاصل بعد از گذراندن فیلترهای بالا، با استفاده از پایگاه pfam (<http://pfam.xfam.org>) بررسی شدند تا توالی‌های دارای دومین پروتئینی نیز حذف شوند و در نهایت lncRNAهایی که در محدوده ۱۰ هزار بازی ژن‌های کدکننده پروتئین قرار گرفته بودند، به‌عنوان lncRNAهای بین ژنی مورد مطالعه بیش‌تر قرار گرفتند [۳].

برای شناسایی lncRNAهای بین ژنی شناخته‌شده با

درآپ تنها نمونه‌هایی که نسبت ۱/۵ الی ۲/۱ (جذب نوری در طول موج ۲۶۰ نسبت به ۲۸۰) و براساس نتایج ژل آگارز بین باندهای ۲۸S و ۱۸S قرار گرفتند، برای توالی‌یابی انتخاب شدند. نمونه‌ها برای سنتز cDNA و توالی‌یابی با استفاده از فناوری ایلومینا به شرکت BGI هنگ‌کنگ ارسال و عملیات مذکور با پروسه زیر انجام شد.

نمونه‌های ارسال‌شده بعد از کنترل کیفی با دستگاه بایوجیلنت ۲۱۰۰، استفاده شدند. نمونه‌های با کیفیت مناسب تحت فرایند انتخاب mRNAها قرار گرفتند که برای این کار از Oligo dT beads استفاده شد که با پروب شدن به دم پلی A انتهای mRNA باعث حفظ آن‌ها و حذف RNAهای کوچک‌تر، از قبیل RNAهای ریبوزومی، tRNA و غیره می‌شوند. بعد از حذف RNAهای کوچک، mRNAها توسط کیت‌های خاصی شکسته شده و به قطعات کوتاه‌تری تبدیل شدند. با استفاده از پرایمرهای هگزامری، از روی قطعات شکسته شده، cDNA ساخته شد و cDNAهای با طول مشخص (حدود ۲۰۰ باز) انتخاب، انتهای آن‌ها ترمیم شده و باز آدین و سپس آداپتور به آن‌ها اضافه شد و در نهایت با استفاده از دستگاه‌های توالی‌یابی ایلومینا هایسک ۲۵۰۰ مورد توالی‌یابی قرار گرفتند. تعداد بالغ بر ۷۰ میلیون خوانش جفت انتهایی به‌ازای هر نمونه تولید شد، میانگین خوانش‌ها ۷۱۰۵۸۸۲۵ و حداقل و حداکثر خوانش‌ها به‌ترتیب ۷۰۶۱۴۵۳۶ و ۷۱۵۶۲۵۸۲ بود. برای انجام آنالیزهای بیوانفورماتیکی از سوپر کامپیوتر (80 core cpu, 256 RAM, HP) پردیس ابوریحان دانشگاه تهران استفاده شد. برای رسم نمودارها از نرم‌افزار Microsoft Excel 2013 و برای رسم هیتمپ از بسته heatmap در نرم‌افزار RStudio (نسخه ۳/۰) استفاده شد. برای کنترل کیفیت و شناسایی lncRNAها از نرم‌افزارهای مختلف بهره گرفته شد.

برای کنترل کیفی خوانش‌ها قبل و بعد از پیرایش و فیلترکردن خوانش‌ها از نرم‌افزار fastQC (نسخه ۰/۱۱/۵)

تولیدات دومی

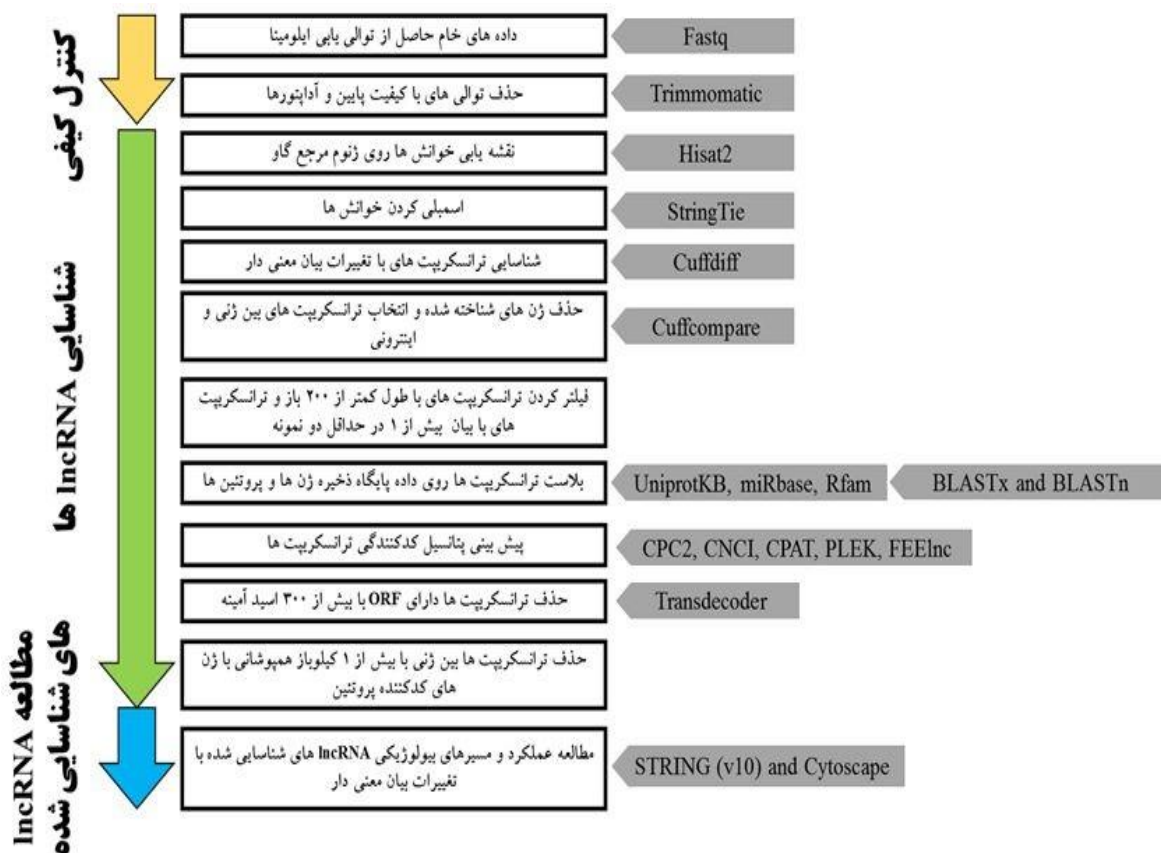
مسیرهای بیولوژیکی مرتبط با lncRNA های شناخته شده بین ژنی در بافت شکمبه گوساله های مبتلا به اسیدوز

آنالیز شده (<http://www.genome.jp/kegg/kegg1.html>) و نتایج بررسی عملکردی برای فرایند زیستی با سطح معنی دار تصحیح شده برای خطای چند آزمونی با روش FDR ($P\text{-adjusted} \leq 0.05$) بررسی شد. مراحل مختلف آنالیز بیوانفورماتیکی داده های RNA-seq به صورت شماتیک در شکل ۱ نشان داده شده است.

نتایج

نتایج حاصل از توالی یابی برای هر نمونه، بیش از ۳۵/۵ میلیون خوانش را نشان داد. بررسی کیفیت خوانش ها نشان داد که تمامی آن ها دارای میانگین کیفی بالاتر از ۳۰ و توزیع GC نرمال هستند. این دو فاکتور نشان دهنده کیفیت بالای خوانش ها و مناسب بودن آن ها برای آنالیزهای بعدی است.

تغییرات بیان معنی دار، از شاخص FPKM (Fragments Per Kb for a Million reads) استفاده شد (رونوشت های با FPKM برابر یا بیش از یک که حداقل در دو نمونه بیان شده بودند). Cuffdiff (نسخه ۲/۱/۱) [۱۳] با استفاده از شاخص FPKM میزان تغییرات بیان (-Log Fold Change) را محاسبه می کند. با استفاده از این نرم افزار ژن های دارای تغییرات بیان معنی دار مشخص شدند. در پژوهش حاضر، رونوشت های با $\log_2(\text{fold change}) \geq 2$ و $P\text{-adjust} \leq 0.05$ به عنوان رونوشت های با تغییرات بیان معنی دار در نظر گرفته شدند [۳]. برای مطالعه هستی شناسی و مسیرهای بیولوژیکی، ژن های هدف lncRNA های شناخته شده بین ژنی، با استفاده از داده پایگاه (<http://www.geneontology.org>) و KEGG



شکل ۱. مراحل مختلف آنالیز بیوانفورماتیکی داده های RNA-Seq

تولیدات دامی

دوره ۲۲ ■ شماره ۳ ■ پاییز ۱۳۹۹

کروموزم‌های ۲۶ و ۲۷ کم‌ترین تعداد ژن‌های lncRNA های شناخته‌شده بین ژنی را دارا بودند (شکل ۲).

تجزیه و تحلیل میزان بیان ژن‌های lncRNA های شناخته‌شده بین ژنی در بافت شکمبه سالم و مبتلا به اسیدوز نشان داد که ۵۶ ژن دارای تغییرات بیان معنی‌دار هستند. خصوصیات ژنومیکی تعدادی از lncRNA های شناخته‌شده با اثرات بزرگتر در نمونه‌های اسیدوزی و سالم و فاصله آن‌ها با مناطق کدکننده پروتئین در جدول (۲) نشان داده شده است. lncRNA های MSTRG.2585 و MSTRG.20535 در نمونه‌های اسیدوزی نسبت به سالم دارای بیان بیشتر و lncRNA های MSTRG.15936 و MSTRG.6750 در نمونه‌های اسیدوزی نسبت به سالم، بیان کم‌تری داشتند (جدول ۲).

در مجموع بعد از اعمال فیلترها و آنالیز توالی‌ها، ۴۳۳۰ ژن به‌عنوان lncRNA ها تعیین شدند. این ژن‌ها در چهار گروه، شامل گروه بین ژنی شناخته‌شده (known lincRNA)، ایترونی شناخته‌شده (intronic lincRNA)، بین ژنی جدید (novel lincRNAs) و ایترونی جدید (intronic lincRNA) قرار گرفتند (جدول ۱). برای lncRNA های شناخته‌شده بین ژنی تعداد ۱۶۳۶ ژن شناسایی شد. برای این گروه از lncRNA ها میانگین طول رونوشت، میانگین تعداد اگزون و میانگین درصد گوانین-سیتوزین به ترتیب ۲۱۸۳/۸۷ نوکلئوتید، ۱/۲۹ اگزون و ۴۴ درصد مشاهده شد. همان‌طورکه در شکل (۲) نیز نشان داده شده است، بررسی توزیع کروموزمی lncRNA های شناخته‌شده بین ژنی نشان داد که این نوع از lncRNA ها بیشتر در کروموزم شماره ۱ و X توزیع شده‌اند.

جدول ۱. خصوصیات رونوشت‌ها و ژن‌های lncRNA ها در بافت شکمبه گوساله هلشتاین مبتلا به اسیدوز

خصوصیات	بین ژنی شناخته‌شده	ایترونی شناخته‌شده	بین ژنی جدید	ایترونی جدید
تعداد ژن	۱۶۳۶	۱۲۱۷	۶۳۹	۸۳۸
تعداد رونوشت	۱۷۳۲	۱۲۴۴	۶۴۰	۸۴۰
میانگین طول رونوشت	۲۱۸۳/۸۷	۳۵۹۰/۹۳	۵۲۴/۶۵	۶۱۹/۶۵
میانگین تعداد اگزون	۱/۲۹	۱/۲۴	۱/۰۷	۱/۰۲
میانگین میزان گوانین-سیتوزین (GC)	۰/۴۴	۰/۴۳	۰/۴۴	۰/۴۳



شکل ۲. توزیع کروموزمی lncRNA های شناخته‌شده بین ژنی در بافت شکمبه گوساله هلشتاین با استفاده از نرم‌افزار Microsoft Excel 2013

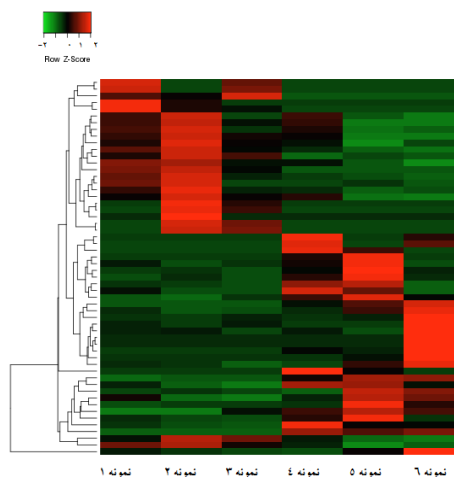
مسیرهای بیولوژیکی مرتبط با lncRNA های شناخته شده بین ژنی در بافت شکمبه گوساله های مبتلا به اسیدوز

جدول ۲. خصوصیات ژنومیکی lncRNA های شناخته شده و فاصله آن ها با مناطق کدکننده پروتئین

فاصله	نزدیکترین ژن بالادست	فاصله	نزدیکترین ژن پایین دست	درصد GC	شماره و موقعیت در کروموزوم	طول	نام lncRNA
۲۵۷۴	SMYD1	-۵۷۸۲	FABP1	۴۵	۱۱	۱۵۲۵	MSTRG.2585
۵۸۱۶	NMRK2	-۲۳۰۳	DAPK3	۵۵	۷	۳۰۵۳	MSTRG.20535
۷۵۵۵	TMEM45B	-۱۰۶۳	NTM	۵۴	۲۹	۱۰۴۱	MSTRG.15936
۷۸۵۸۶۱	PCDH18	-۱۲۱۴۴	SLC7A11	۳۵	۱۷	۴۹۹۸	MSTRG.6750

برای شناسایی مسیرهای بیولوژیکی مرتبط با lncRNA های دخیل در اسیدوز بافت شکمبه، ژن های هم جوار lncRNA های با تغییرات بیان معنی دار آنالیز شدند. نتایج نشان داد که با سطح احتمال $P \leq 0.05$ ، پنج مسیر بیولوژیکی غنی شدند. این مسیرها شامل Apelin signaling pathway، Gap junction، Glucagon، Renin secretion، AGE-RAGE signaling pathway و signaling pathway in diabetic complications بودند. همچنین آنالیز عملکرد مولکولی این ژن ها نشان داد که دو عملکرد مولکولی شامل gap junction channel activity و phosphatidylinositol phospholipase C activity و به طور معنی دار غنی شدند.

آنالیز خوشه ای، lncRNA های با تغییرات بیان معنی دار را در سه خوشه اصلی قرار داد (شکل ۳). بیشترین تعداد lncRNA ها در خوشه یک قرار گرفتند. در این خوشه، ژن های با افزایش بیان (قرمز رنگ) و کاهش بیان (سبز) در کنار همدیگر طبقه بندی شدند. بررسی الگوی بیان این ژن ها نشان داد که ۱۴ ژن افزایش بیان و ۲۰ ژن کاهش بیان معنی دار دارند ($P \leq 0.05$). همچنین ۲۲ ژن در بافت شکمبه مبتلا به اسیدوز، القا یا خاموش شدند (شکل ۴). بیشترین میزان بیان در lncRNA های شناخته شده بین ژنی MSTRG.2585 و MSTRG.20535 مشاهده شدند، در حالی که ژن MSTRG.15936 کمترین میزان بیان را داشت.



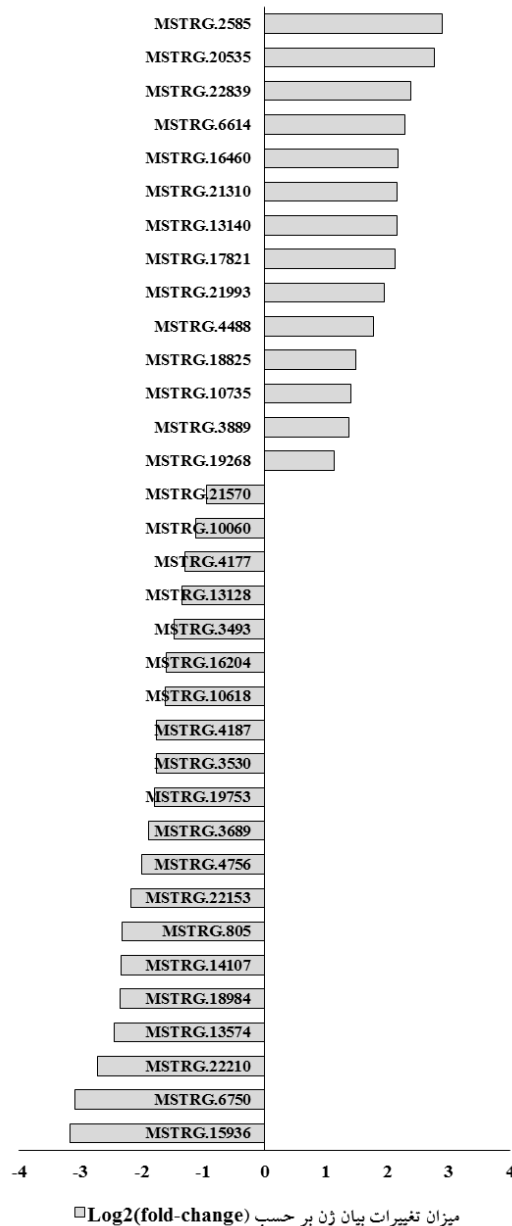
شکل ۳. آنالیز خوشه ای و میزان بیان lncRNA های شناخته شده بین ژنی در نمونه های بافت شکمبه مورد آزمایش با استفاده از بسته

Heatmap در نرم افزار Rstudio

تولیدات دامی

دوره ۲۲ ■ شماره ۳ ■ پاییز ۱۳۹۹

۳۴۳



شکل ۴. الگوی بیان lincRNAهای شناخته‌شده بین ژنی با تغییرات بیان معنی‌دار و میزان بیان lincRNAهای شناخته‌شده بین ژنی با افزایش و کاهش بیان معنی‌دار ($P \leq 0.05$). میزان تغییرات بیان براساس Log2 (fold change) نشان داده شده است. نمودارهای سمت راست محور عمودی نشان‌دهنده افزایش بیان و سمت چپ نشان‌دهنده کاهش بیان معنی‌دار در مقایسه با گروه کنترل است.

است و هیچ هم‌پوشانی با ژن‌های کدکننده پروتئین ندارند. نقش lincRNAهای بین‌ژنی در بیماری‌های مختلف به‌ویژه در انسان و موش مورد مطالعه قرار گرفته است [۲۵]، اما

بحث

lincRNAهای بین ژنی (lincRNAs) گروهی از lincRNAها هستند که طول‌ها آن‌ها بیش از ۲۰۰ نوکلئوتید

تولیدات دائمی

دوره ۲۲ ■ شماره ۳ ■ پاییز ۱۳۹۹

مسیرهای بیولوژیکی مرتبط با lncRNA های شناخته شده بین ژنی در بافت شکمبه گوساله های مبتلا به اسیدوز

میکروب های شکمبه در محیط اسیدی خوب کار نمی کنند. فعالیت آنزیم های بدنی نیز تحت تأثیر pH پایین خون قرار می گیرند. خون اسیدی نمی تواند اکسیژن زیادی حمل کند در نتیجه پاهای گاو که دورترین نقاط بدن گاو به حساب می آیند، کمترین اکسیژن را دریافت خواهند کرد. این مسأله منجر به ایجاد آماس و ورم در پاها می شود. از طرف دیگر توکسین های داخلی تولید شده در شکمبه، آزاد شدن هیستامین را افزایش می دهند که این ماده به آماس و التهاب پاها بیش تر کمک می کند [۱۸].

با عنایت به این که lncRNA ها بیان ژن های هم جوار خود را تنظیم می نمایند، برای مطالعه نقش lncRNA های شناخته شده بین ژنی با تغییرات بیان معنی دار در بافت شکمبه در پاسخ به رخداد اسیدوز، ژن های هم جوار آن ها در محدوده ۱۰ کیلوبازی شناسایی شدند و مسیرهای بیولوژیکی و عملکرد مولکولی آن ها مطالعه شد. برای مشخص کردن عملکرد lncRNA ها از یک رویکرد زیست شناسی سیستم استفاده شد که منجر به شناسایی پنج مسیر بیولوژیکی شد. از پنج مسیر بیولوژیکی غنی شده، سه مسیر متعلق به ترانس سانی پیام در سلول بودند. این مسیرها می توانند در ارتباط با نقش تنظیمی lncRNA های شناخته شده بین ژنی باشند. مسیر بیولوژیکی Apelin signaling یکی از مسیرهای مهم سلولی است و نقش آن در فرایندهای مختلف سلولی از جمله رگ زایی، عملکرد عروق، تکثیر سلولی و متابولیسم انرژی نشان داده شده است. پپتید Apelin با اتصال به گیرنده خود (APJ) که یک نوع G-پروتئین است، باعث فعال شدن آنزیم های زیادی می شود که در نهایت به تغییر بیان ژن های مختلفی می انجامد [۱۴]. دومین مسیر بیولوژیکی، مسیر ترانس سانی Glucagon signaling بود که به طور معنی دار غنی شد. این مسیر بیولوژیکی با اتصال هورمون گلوکاگون به گیرنده های خود در سطح سلول

نقش و عملکرد آن ها در دام و بیماری اسیدوز کم تر مورد توجه قرار گرفته است. در مطالعه حاضر، ۱۶۳۶ lncRNA ی بین ژنی شناسایی شد که دارای متوسط طول ۲۱۸۳/۸۷ نوکلئوتید بودند. این مقدار در مقایسه با متوسط طول lncRNA ی بین ژنی در انسان (۱۰۰۰ نوکلئوتید) و موش (۵۵۰ نوکلئوتید) بیش تر بود [۲۰]. هم چنین میانگین میزان گوانین-سیتوزین در این مطالعه ۴۴ درصد مشاهده شد که در مقایسه با میانگین گوانین-سیتوزین mRNA ها کم تر است [۲۱]. میزان این بازها (بیش از ۵۵ درصد) در نواحی پروموتری و بالادستی ژن همبستگی مثبتی با پتانسیل بیان ژن دارد [۱۷ و ۲۲]. این در حالی است که میزان گوانین-سیتوزین در نواحی اگزونی و ایترونی همبستگی منفی با میزان بیان ژن دارد [۱۷]. میزان بالای گوانین-سیتوزین (بیش از ۹۰ درصد) باعث می شود تا ژن مورد نظر در مقایسه با ژن های دارای میزان گوانین-سیتوزین کم تر (حدود ۳۰ درصد) بین ۷ الی ۱۰۰ برابر بیان بیش تری داشته باشند [۲۲]. pH شکمبه گاو در طول روز در نوسان است، اما باید بالاتر از ۵/۸ باشد تا محیط شکمبه برای فعالیت میکروب ها حفظ شود، اسیدوز شکمبه ای تحت حاد به عنوان دوره ای از کاهش pH شکمبه تعریف شده است و ممکن است با مشکلاتی از قبیل لنگش و کاهش تولید در ارتباط باشد. هرچند pH شکمبه در طول شبانه روز به وفور تغییر می یابد، ولی گاوها دارای سیستمی می باشند تا pH شکمبه آن ها در یک محدوده فیزیولوژیکی محافظت و نگهداری شود. به هر حال اگر تولید اسید ناشی از تخمیر زیاد، بیش تر از بافر (ماده ای که از تغییرات شدید pH جلوگیری می کند) باشد، تعادل pH شکمبه به هم می خورد و ممکن است تا حد زیادی از تعادل خارج شود. اسیدوز متابولیکی وقتی اتفاق می افتد که pH شکمبه به حد پایین تر از پنج تا ۵/۵ برسد و خون گاو اسیدی شود. سلول های بدن گاو همانند

تولیدات دامی

دوره ۲۲ ■ شماره ۳ ■ پاییز ۱۳۹۹

در راستای جبران افزایش نفوذپذیری ناشی از صدمه به لایه سلولی بافت در معرض اسیدوز، ارتباطات سلولی و نفوذپذیری لایه‌های سلولی را کاهش می‌دهد تا خود را در مقابل صدمات ناشی از آن محافظت نماید. یافته‌های این پژوهش نشان می‌دهد lncRNAهای بین‌ژنی، ژن‌های دخیل در مسیر بیولوژیکی Gap junction را هدف قرار می‌دهند. به نظر می‌رسد تنظیم این ژن‌ها در پاسخ به شرایط اسیدی پیش‌آمده در بافت شکمبه باشد تا نفوذپذیری لایه اپیتلیال بافت شکمبه را کاهش دهد [۲۴].

در مطالعه حاضر دخالت مسیر بیولوژیکی Renin secretion در بافت شکمبه تحت شرایط اسیدوز نیز نشان داده شد. این مسیر بیولوژیکی در نهایت به ترشح رنین می‌انجامد که در پاسخ به محرک‌های مختلف به‌ویژه فشار خون اتفاق می‌افتد [۴]. کاهش pH خون باعث افزایش فشار خون می‌شود. بدن در پاسخ به افزایش فشار خون و برای کنترل فشار از سازوکارهای مختلفی استفاده می‌کند که یکی از آنها فعال‌کردن مسیرهای بیولوژیکی مرتبط با ترشح رنین است [۴]. یافته‌های این مطالعه نشان می‌دهند که lncRNAهای بین‌ژنی در تنظیم بیان ژن‌های دخیل در ترشح رنین در پاسخ به pH پایین بافت شکمبه مبتلا به اسیدوز نقش دارند.

آنالیز مولکولی ژن‌های هدف lncRNAهای شناخته‌شده بین ژنی دو عملکرد مولکولی شامل gap junction channel activity و phosphatidylinositol phospholipase C activity را نشان داد. عملکرد اول همگام با مسیر بیولوژیکی Gap junction است که پیش از این مورد بحث قرار گرفت. عملکرد مولکولی غنی‌شده phosphatidylinositol phospholipase C activity در مسیر ترانسانی پیام داخل سلولی نقش دارد [۶]. در اثر رخداد این عملکرد مولکولی، که به‌واسطه وقوع بیماری اسیدوز اتفاق می‌افتد، فسفاتیدیل اینوزیتول شکسته شده و

فعال می‌شود و رابطه مستقیمی با فعالیت‌های لیپولیز دارد که طی آن، لیپیدها برای تأمین انرژی سلول تجزیه‌شده و برای تولید انرژی موردنیاز سلول به‌کار برده می‌شوند [۵]. سومین مسیر ترانسانی پیام، AGE-RAGE signaling pathway بود. این مسیر ترانسانی در بسیاری از بیماری‌ها به‌ویژه دیابت مورد مطالعه قرار گرفته است و نشان داده شده است که این مسیر در پاسخ به غلظت بالای قند و تجمع کلسیم در بدن فعال می‌شود [۷ و ۱۲]. این مسیرهای بیولوژیکی می‌توانند در ارتباط با کاهش pH و افزایش غلظت لیپوپروتئین‌های خون در بافت شکمبه مبتلا به اسیدوز باشند [۱۲].

مسیر بیولوژیکی Gap junction شامل مجموعه‌ای از پروتئین‌های غشایی به نام کانکسین است. این پروتئین‌ها با ایجاد کانال‌هایی در غشای سلول باعث ارتباط بین دو سلول مجاور می‌شوند و مناطقی از غشاهای سلولی است که در آن‌ها پروتئین‌های غشا متعلق به سلول‌های مجاور در تماس نزدیک قرار دارند و از این طریق کانال‌های غشایی دوگانه آبرگیز تشکیل می‌شوند، که در بسیاری از فرایندها از جمله تمایز، فعالیت عصبی، توسعه، پاسخ ایمنی و هماهنگ‌سازی سلولی نقش اساسی دارند. علاوه بر این، در اثر جهش در کانکسین‌ها چندین بیماری انسانی از جمله بیماری‌های عصبی، بیماری‌های پوستی، ناشنوایی و ناهنجاری‌های رشدی ایجاد می‌شود. نشان داده شده است که Gap junction در کنار مسیر بیولوژیکی نیتریک اسید در پاسخ به میزان pH خارج سلولی دخیل هستند و در تنظیم آن نقش دارند [۸]. میزان بیان ژن‌های دخیل در مسیر بیولوژیکی Gap junction تحت شرایط اسیدوز کاهش پیدا می‌کند تا میزان نفوذپذیری لایه اپیتلیال کاهش پیدا کند [۱۹]. تحت شرایط pH پایین ممکن است لایه اپیتلیال بافت‌های مختلف صدمه ببیند و در نتیجه نفوذپذیری آن افزایش پیدا کند [۲۴]. به نظر می‌رسد سلول

مسیرهای بیولوژیکی مرتبط با lncRNA های شناخته شده بین ژنی در بافت شکمبه گوساله های مبتلا به اسیدوز

ممکن است نقش نظارتی بالقوه در تعدیل فرایندهای بیولوژیکی مرتبط با مسیرهای سوخت و ساز انرژی و فرایندهای ترانسسانی پیام سلولی مؤثر بر عملکرد شکمبه گوساله ها در پاسخ به بیماری اسیدوز ایفا کنند. مولکول های lncRNA ضمن این که می توانند به عنوان یک عامل پیش آگاهی دهنده برای تشخیص اسیدوز باشند، در اصلاح نژاد دام نیز به عنوان بیومارکر می توان از آن ها بهره برد. پیش بینی می شود lncRNA هایی که دارای تفاوت معنی دار در دو گروه سالم و اسیدوزی هستند با احتمال زیاد با مسیرهای ضروری برای متابولیسم انرژی در شکمبه مرتبط بوده و با مسیرهای ترانسسانی پیام با تمرکز بر عملکرد سلول های اپیتلیال شکمبه همراه باشند. امید است با انجام پژوهش های گسترده در آینده و با شناخته شدن هرچه پیش تر نقش lncRNA ها بتوان گام بزرگی در اصلاح دام برداشت.

تشکر و قدردانی

از معاونت محترم پژوهشی و فناوری دانشگاه علوم کشاورزی و منابع طبیعی خوزستان به خاطر حمایت های مالی و معنوی، تشکر و قدردانی می گردد.

تعارض منافع

هیچ گونه تعارض منافع توسط نویسندگان وجود ندارد.

منابع

1. Adeva MM and Souto G (2011) Diet-induced metabolic acidosis. *Clinical nutrition* 30 (4): 416-421.
2. Bakhtiarizadeh MR and Salami SA (2019) Identification and expression analysis of long noncoding RNAs in fat-tail of sheep breeds. *G3: Genes, Genomes, Genetics* 9 (4): 1263-1276.
3. Bakhtiarizadeh MR, Hosseinpour B, Arefnezhad B, Shamabadi N and Salami SA (2016) In silico prediction of long intergenic non-coding RNAs in sheep. *Genome* 59 (4): 263-275.

تولید دی آسیل گلیسرول (DAG) و اینوزیتول تری فسفات (IP3) می کند. مولکول های آزاد شده در ادامه باعث فعال شدن پروتئین های کینازی شده و در نهایت با آزاد شدن یون های کلسیم داخل سلولی، برخی از مسیرهای ترانسسانی فعال شده و بیان ژن ها در پاسخ به این سیگنال تغییر پیدا می کند [۱۱].

به عنوان نتیجه گیری کلی، پژوهش حاضر ضمن شناسایی lncRNA های بین ژنی شناخته شده در بافت شکمبه، مسیرهای بیولوژیکی غنی شده و عملکردهای مهم lncRNA های شناخته شده در گوساله های مبتلا به اسیدوز را شناسایی کرد. توالی یابی RNA-seq به خوبی قادر به شناسایی lncRNA ها بوده و آن ها نیز نقش قابل توجهی در تنظیم بیان ژن های هم جوارشان دارند. شناسایی lncRNA های با بیان متفاوت در گروه سالم و اسیدوزی بینش جدیدی در مورد سازوکارهای تنظیمی lncRNA ها به دست آورد. به نظر می رسد کاهش pH به عنوان محرکی عمل می کند و باعث فعال شدن برخی مسیرهای بیولوژیکی ترانسسانی پیام می شود. پنج مسیر بیولوژیکی مهم مربوط به متابولیسم انرژی، لیپید و لیپوپروتئین یافت شد. با یافتن ژن های هدف با عملکردهای شناخته شده در متابولیسم چربی و مسیر پیام رسانی کلسیم، حدس زده می شود lncRNA ها از طریق ژن های هدف هم جوار یا هم بیان شده عمل کنند که با تنظیم این ژن های کدکننده پروتئین می توانند در ایجاد اسیدوز نقش داشته باشند.

این نتایج، به عنوان اولین گزارش از تفاوت بیان lncRNA های شناخته شده در نمونه های سالم و اسیدوزی، ضمن این که حاشیه نویسی ژنوم گاو را بهبود می بخشد، می تواند بینش جدیدی در مورد عملکرد lncRNA ها را در توسعه اسیدوز ارائه نماید. یافته های این پژوهش، نشان می دهد که lncRNA هایی که دارای تفاوت بیان معنی دار در دو گروه سالم و اسیدوزی هستند،

تولیدات دامی

دوره ۲۲ ■ شماره ۳ ■ پاییز ۱۳۹۹

4. Catanzaro DF (2018) Molecular biology of renin and regulation of its gene, in Textbook of Nephro-Endocrinology. Elsevier. p. 389-400.
5. Chapman NA, Dupré DJ and Rainey JK (2014) Theapelin receptor: physiology, pathology, cell signalling, and ligand modulation of a peptide-activated class A GPCR. *Biochemistry and cell biology* 92 (6): 431-440.
6. Cocco L, FolloMY, Manzoli L and Suh PG (2015) Phosphoinositide-specific phospholipase C in health and disease. *Journal of lipid research* 56 (10): 1853-1860.
7. Galsgaard KD, Pedersen J, Knop FK, Holst JJ and Albrechtsen NJW (2019) Glucagon Receptor Signaling and Lipid Metabolism. *Frontiers in physiology* 10:413.
8. Giepmans BN (2006) Role of connexin43-interacting proteins at gap junctions, in *Cardiovascular Gap Junctions*. Karger Publishers. p. 41-56.
9. Hangauer MJ, Vaughn IW and McManus MT (2013) Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS genetics* 9 (6): e1003569 .
10. Ibeagha-Awemu EM and Zhao X (2015) Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Frontiers in genetics* 6 302.
11. Kawakami T and Xiao W (2013) Phospholipase C-β in immune cells. *Advances in biological regulation* 53 (3): 249-257.
12. Kay AM, Simpson CL and Stewart JA (2016) The role of AGE/RAGE signaling in diabetes-mediated vascular calcification. *Journal of diabetes research* 2016: 1-8.
13. Kong L, Zhang Y, Ye ZQ, Lio XQ, Zhao SQ, Wei L and Gao G (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research* 35 (suppl_2): W345-W349.
14. Kudla G, Lipinski L, Caffin F, Helwak A and Zylisz M (2006) High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS biology* 4 (6): e180.
15. Leblanc S (2010) Monitoring metabolic health of dairy cattle in the transition period. *Journal of reproduction and Development*. 56(S): S29-S35.
16. Li A, Zhang J, and Zhou Z (2014) PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC bioinformatics* 15 (1): 311.
17. Li T, Wang S, Wu R, Zhou X, Zhu D and Zhang Y (2012) Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. *Genomics* 99 (5): 292-298.
18. Mackey E (2013) Effects of ruminal acidosis on rumen papillae transcriptome. University of Delaware, Newark, Delaware, thesis.
19. Mohanty I, Parija SC, Suklabaidya S and Rattan S (2018) Acidosis potentiates endothelium-dependent vasorelaxation and gap junction communication in the superior mesenteric artery. *European journal of pharmacology* 827: 22-31.
20. Quinn JJ and Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics* 17 (1): 47.
21. Ransohoff JD, Wei Y and Khavari PA (2018) The functions and unique features of long intergenic non-coding RNA. *Nature reviews Molecular cell biology* 19 (3): 143.
22. Rao YS, Chai XW, Wang ZF, Nie QH and Zhang XQ (2013) Impact of GC content on gene expression pattern in chicken. *Genetics Selection Evolution* 45 (1): 9.
23. Sun L, Luo H, Bu D, Zhao C, Liu Y, Chen R and Zhao Y (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic acids research* 41 (17): e166.
24. Sun YY, Cheng M, Xu M, Song LW, Gao M and Hu HL (2018) The effects of subacute ruminal acidosis on rumen epithelium barrier function in dairy goats. *Small ruminant research* 169: 1-7.
25. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL and Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* 31: 46-53.