Iranian Journal of Animal Science
Vol 50, No 1, Spring 2019 (47-55)
DOI: 10.22059/ijas.2018.246004.653585

علوم دامی ایران
دورۀ ۵۰، شمارۀ ۱، بهار ۱۳۹۸ (ص ۵۵-۴۷)

بیان افتراقی ژن گاوهای بوس تاروس (هلشتاین) و بوس ایندیکوس (کلیستانی) با استفاده از داده‌های RNA-Seq

مینا سلیم‌پور[۱]، سید رضا میرائی آشتیانی[۲*] و محمد حسین بناءبازی[۳]
۱ و ۲. دانشجوی سابق کارشناسی ارشد و استاد، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران
۳. استادیار پژوهشی، مؤسسه تحقیقات علوم دامی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران
(تاریخ دریافت: ۱۳۹۶/۹/۷ - تاریخ پذیرش: ۱۳۹۷/۸/۱۳)

چکیده

این پژوهش با هدف تعیین مقایسه افتراقی پروفایل بیان ژن در دو زیرگونه گاو بوس تاروس (هلشتاین) و بوس ایندیکوس (کلیستانی) انجام شد. برای این منظور، ترانسکریپتوم نمونه‌ای از هریک از جمعیت‌های گاو ماده هلشتاین آمریکا و گاو ماده کلیستانی پاکستان از طریق همردیفی و مکان‌یابی خوانش‌های کوتاه mRNA تشکیل شد. این خوانش‌ها قبلاً با استفاده از فناوری توالی‌یابی نسل جدید به‌صورت داده‌های RNA-Seq تولید شده بودند. نتایج آنالیز بیان افتراقی نشان داد از مجموع ۲۴۶۱۶ ژن و ۲۶۷۱۶ رونوشت شناخته‌شده بر روی ژنوم مرجع گاو، ۴۱ ژن بین این دو زیرگونه بیان متفاوت داشتند ($p< ۰/۰۰۰۰۱۵$). بالاترین شاخص بیان دیجیتال مربوط به یک ژن میتوکندریایی (ENSBTAG00000043545) بودکه تنها در جمعیت کلیستانی بیان شد. یکی از ژن‌های متفاوت بیان‌شده (ENSBTAG00000014332) دارای دو ایزوفرم متفاوت بیان شده بود. آنالیز ماهیت و مسیر ژن‌های متفاوت بیان شده نشان داد که آنها در ۲۰ مسیر به‌ویژه مسیرهای مرتبط با ایمنی، پاسخ به تنش و تشکیل رگ‌های خونی جدید درگیر بودند. یعنی مسیرهایی که در طول زمان سازگاری دو زیرگونه مورد مطالعه با اقلیم‌های خاص و بروز تفاوتهای فنوتیپی بارز برای این صفات را بین آن‌ها موجب شده‌اند.

واژه‌های کلیدی: ایمنی، ترانسکریپتوم، گاو، ماهیت ژن، mRNA.

Differential gene expression of two bovine *Bos taurus* (Holstein) and *Bos indicus* (Cholistani) sub-species using RNA-Seq data

Mina Salimpour[1], Seyed Reza Miraei-Ashtiani[2*] and Mohammad Hossein Banabazi[3]
1, 2. Former M.Sc. Student and Professor, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran
3. Assistant Professor, Animal Science Research Institute of Iran (ASRI), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran
(Received: Nov. 28, 2017 - Accepted: Nov. 4, 2018)

ABSTRACT

The aim of this research was to study gene expression profiling and differential analysis between *Bos taurus* (Holstein) and *Bos indicus* (Cholistani) subspecies. The transcriptome was assembled through aligning and mapping the RNA-Seq reads that have already been sequenced by next generation sequencing technology. Among 24616 genes and 26717 transcripts, only 41 genes were differently expressed. The highest digital gene expression was measured for a mitochondrial gene (ENSBTAG00000043545), and was only expressed in the Cholistani population. One gene had two differentially expressed isoforms. Gene pathway analysis indicated that the differential expressed genes included in pathways, are particularly related to immunity, response to stress and angiogenesis. These pathways have probably resulted in adoption to various climatological conditions and perceptible phenotypes in the studied subspecies during their evolution.

Keywords: Bovine, Gene ontology, immunity, mRNA, Transcriptome.

* Corresponding author E-mail:

مقدمه

بیان ژن مکانیسمی حساب شده، پیچیده و تنظیم‌گر است که باعث می شود ژنوتیپ به‌صورت فنوتیپ ظاهر شود (Banabazi *et al*., 2012a). در واقع رمزهای ژنتیکی که در مولکول DNA ذخیره شده‌اند به‌وسیله بیان ژن‌ها تفسیر می‌شوند و خصوصیات و نحوه بیان ژن باعث به‌وجودآمدن فنوتیپ در ارگانیسم خواهد شد (Banabazi *et al*., 2012b). مطالعات بیان ژن به تولید حجم زیادی از داده‌ها در این زمینه منجر شده است. در میان انواع متعدد داده‌های بیان ژن داده‌های RNA-seq به‌سبب پربرونداد بودن و شامل شدن توالی نواحی رونویسی‌شده و پوشش‌دادن همه این نواحی دارای مزیت زیادی نسبت به سایر انواع داده‌ها هستند در سال‌های اخیر امکانات رایانه‌ای آنالیز این حجم از داده‌ها را در زمان و با هزینه کم امکان‌پذیر ساخته است. از مزایای استفاده از داده‌های RNA-seq شناسایی ژن‌های جدید می‌باشد (Gan *et al*., 2010; Sultan *et al*., 2008). اساس این روش براساس برش کل mRNA (ترانسکریپتوم) به قطعات کوچک و سپس توالی‌یابی همه این قطعات می‌باشد (Marguerat & Bahler, 2010; Haas & Zody, 2010). از این رو از روی تعداد خوانش‌های مکان یابی شده برای یک ناحیه خاص می‌توان به یک معیار دقیق (دیجیتال) از میزان رونویسی آن ناحیه رسید (Ekbom & Galindo, 2011; Wilhelm & Landry, 2009). شاخص بیان ژن یا به‌عبارت دیگر معیار بیان ژن در این روش محاسبه تعداد خوانش یا قطع در حد کیلوباز به‌ازای هر یک میلیون خوانش مکان‌یابی‌شده (FPKM/RPKM) می‌باشد (Mortazavi *et al*., 2008) که این شاخص برای طول ژن و تعداد کل خوانش‌های مکانیابی‌شده در هر نمونه تصحیح شده و مقادیر برآوردشده برای ژن‌های مختلف با هر طول و تعداد خوانش قابل مقایسه و آزمون می‌باشد.

براساس آخرین اطلاعات ژنوم مرجع گاو، تاکنون بیش از ۲۴۰۰۰ ژن روی ۲/۸۷Gb طول ژنوم گاو شناسایی شده است (Bae *et al*., 2010). دو زیر گونه از گاو وجود دارد: زیرگونه *Bos taurus* که از اروپا و شمال امریکا منشأ یافته و بیشتر ویژه مناطق معتدل است و زیرگونه *Bos indicus* که کوهان‌دار (زبو) بوده و در مناطق حاره‌ای پراکنده است. این گونه از جنوب آسیا منشأ یافته است (Fries & Ruvinsky, 1999). تفاوت فنوتیپی قابل‌توجهی بین این دو زیرگونه وجود دارد، از جمله این که گاوهای زبو تنش‌های محیطی را بهتر تحمل می‌کنند و در مقابل انگل‌ها و بیماری‌ها مقاوم‌تر هستند (Dorak *et al*., 1978; Hansen, 2004). Huang *et al*. (2012) داده‌های RNA-seq را برای سه جمعیت گاو هلشتاین، کلیستانی و جرسی تولید کردند و پروفایل بیان ژن این سه جمعیت را مقایسه کردند. مطالعه حاضر با هدف بررسی تأثیر تفاوت فنوتیپی بر میزان بیان ژن‌ها و آنالیز بیان ژن افتراقی بین دو جمعیت گاو هلشتاین (بوس تاروس) و کلیستانی (بوس ایندیکوس) با استفاده از داده‌های Huang *et al*. (2012) انجام شده است.

مواد و روش‌ها

داده‌های RNA-Seq از ادغام ۴۰ نمونه خون کامل گاوهای ماده هلشتاین از مرکز گاو شیری دانشگاه ویسکانسین آمریکا (*Bos taurus*) و ۴۵ گاو ماده کلیستانی (*Bos indicus*) مزرعه گوجایتپیر شهر باهاوالپور واقع در ایالت پنجاب پاکستان تولید شد (Huang *et al*., 2012). برای این منظور نمونه‌های خون بلافاصله پس از اخذ از سیاهرگ دمی برای هر گروه با هم مخلوط شد. سپس گلبول‌های سفید خون (لکوسیت) جدا شده و RNA از آن‌ها استخراج شد. سپس نمونه‌های RNA پلی آدنینه شده و کتابخانه mRNA از آن‌ها تهیه شد. این کتابخانه‌ها با استفاده از Illumina Genome Analyzer IIx. توالی‌یابی شدند. خوانش‌های حاصل از توالی به‌صورت جفتی (paired) یا دوسویه و با طول ۷۵ جفت باز بودند. داده‌های خام در فرمت sra از سایت NCBI (http://www.ncbi.nlm.nih.gov/sra/) دریافت شد و کلیه آنالیزها در محیط سیستم عامل لینوکس اوبونتو ۱۴۰۴ و بر روی کلاستر گروه علوم دامی دانشگاه تهران آنالیز شدند مراحل آنالیز شامل: تبدیل فرمت sra به fastq با استفاده از نرم‌افزار sratoolkits نسخه ۲،۵،۲ و سنجش کیفیت داده‌ها با استفاده از FastQC (Andrews, 2010) نسخه ۳،۱۱،۰

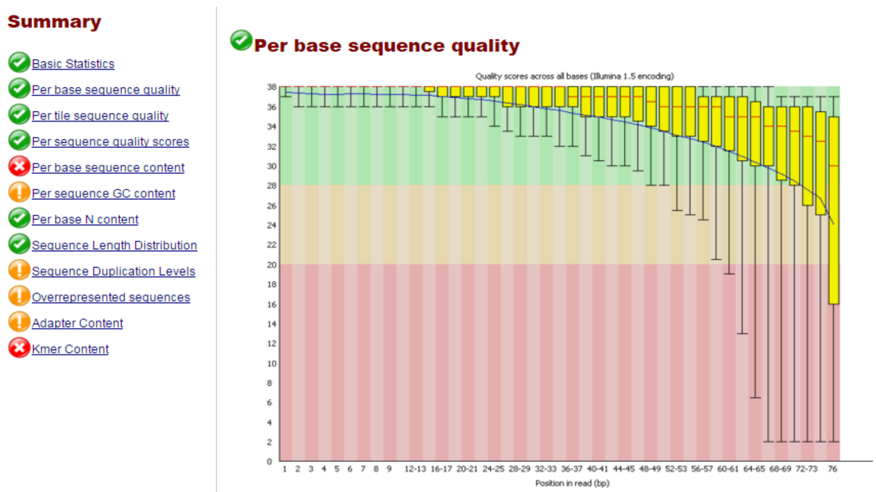
بود. استاندارد کدگذاری کیفیت برای سیستم ایلومینای مورد استفاده در این مطالعه Illumina 1.5 بود که استاندارد سنجش کیفیت به‌روش کلاسیک توالی‌یابی سانجر بر مبنای از صفر تا ۴۰ بود. سپس براساس نتایج سنجش کنترل کیفیت ویرایش خوانش‌ها با استفاده از نرم‌افزار Trimmomatic نسخه ۳۳,۰ (Bolger *et al.*, 2014) انجام شد. این ویرایش شامل حذف آداپتورها و خوانش‌های بی‌کیفیت بود. ترانسکریپتوم برای هریک از دو زیر گونه براساس همردیفی خوانش‌ها و مکان‌یابی آن‌ها روی ژنوم مرجع به‌ترتیب با نرم‌افزار Bowtie2 و TopHat2 و با استفاده دستورات به‌کار برده‌شده در محیط لینوکس انجام شد (Langmead & Salzberg, 2012). ژنوم رفرنس مورد استفاده برای تشکیل ترانسکریپتوم UMD3.1 (ویرایش ۸۲) بود.

اطلاعات حاشیه نویسی به‌کاررفته نیز مربوط به این ویرایش ذکر شده بود. ژنوم مرجع و حاشیه‌نویسی آن از بانک اطلاعاتی Ensemble دریافت شد (ftp://ftp.ensembl.org/pub/release82/fasta/bos_ taurus/dna/Bos_taurus.UMD3.1.dna.toplevel.fa. gzftp://ftp.ensemble.org/pub/release-84/gtf/bos_taurus/Bos_taurus.UMD3,1,84gtf.). اندازه فاصله درون کتابخانه‌ای و انحراف معیار آن برای استفاده از کد TopHat2 با استفاده از bwa (Li & Durbin, 2009) و به روش نمونه‌گیری از داده‌ها انجام شد که به‌ترتیب برای هلشتاین ۱۰۰ و ۱۵ برای کلیستانی ۱۸۰ و ۴۵ برآورد شد. خروجی فایل BAM حاصل از TopHat2 با استفاده از بسته SAMtools مرتب و Index شد.

تشکیل پروفایل بیان ژن برای هریک از دو زیرگونه با استفاده از بسته نرم‌افزاری Cufflinks انجام شد. در این مجموعه نرم‌افزاری تلفیق دو ترانسکریپتوم با استفاده از Cuffmerg و آنالیز بیان افتراقی ژن با استفاده از Cuffdiff به‌عمل آمد. مصورسازی نتایج در قالب گراف با استفاده از پکیج CummeRbund در نرم‌افزاری R (http://www.bioconductor.org/) انجام شد. آنالیز ماهیت و مسیر های ژنی به‌صورت آنلاین بر روی بانک اطلاعاتی Panther (http://pantherdb.org/) انجام شدند.

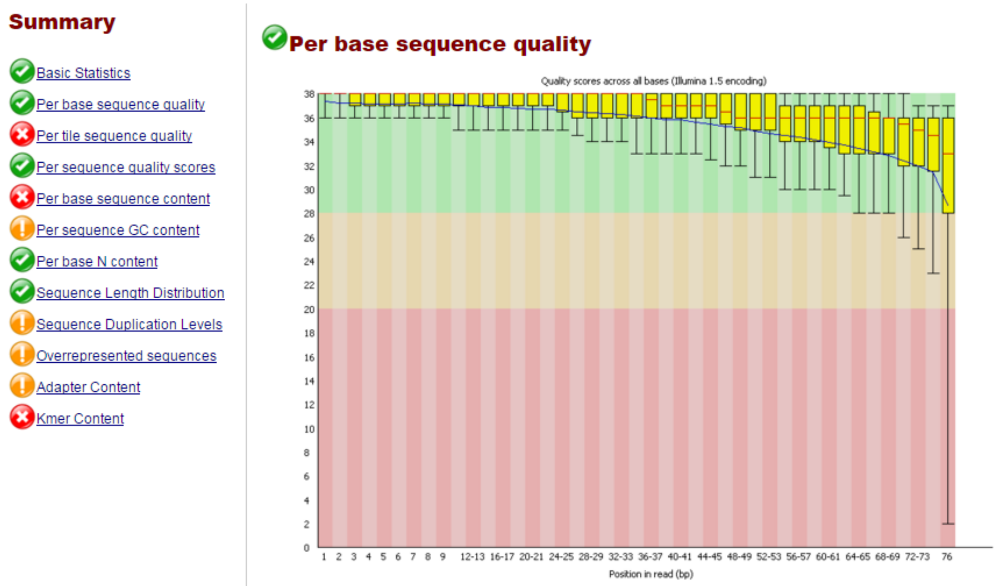
نتایج و بحث

پس از تبدیل فرمت به‌ترتیب ۲۱,۰۷۸,۴۷۷ و ۲۰,۹۴۰,۰۶۳ خوانش کوتاه جفتی (از هر دو انتها توالی‌یابی شده) به طول ۷۵ جفت باز، برای جمعیت‌های هلشتاین و کلیستانی بدست آمد. نتایج سنجش کیفیت برای داده‌های هلشتاین و کلیستانی در شکل‌های ۱ و ۲ نشان داده شده است. محور عمودی و افقی به‌ترتیب امتیاز سنجش کیفیت بر مبنای ۰ تا ۴۰ و موقعیت هر خوانش را نشان می‌دهد.



شکل ۱. خلاصه نتایج آزمون‌های ده‌گانه سنجش کیفیت و آزمون کیفیت انفرادی نوکلئوتیدهای خوانش‌های پیش رو برای جمعیت کلیستانی

Figure 1. Summary of the results quality of control tests for forwrad reads in Cholistani population



شکل ۲. خلاصه نتایج آزمون‌های ده‌گانه سنجش کیفیت و آزمون کیفیت انفرادی نوکلئوتیدهای خوانش‌های پیش رو برای جمعیت هلشتاین

Figure 2. Summary of the results quality control tests for forward reads in Holstein population

این نرم‌افزار سنجش کیفیت داده‌ها برای خوانش‌های جفتی یا دوسویه (رفت و برگشت) را به‌طور جداگانه با ده آزمون مختلف انجام می‌دهد .منطقه سبز دامنه کیفیت مطلوب، منطقه نارنجی قابل قبول و منطقه قرمز دامنه کیفیت نامطلوب را نشان می‌دهد. شاخص ده گانه کنترل کیفیت حاکی از کیفیت مناسب داده‌ها به‌ازای هر باز و خوانش صحیح همه آن‌ها براساس میزان، تعداد N به‌ازای هر باز بود. با در نظرگرفتن حداقل طول خوانش ۵۰ جفت باز در هنگام ویرایش، ۱۹،۳۷۹،۴۸۷ و ۱۹،۳۴۶،۱۷۲ خوانش با کیفیت مناسب برای جمعیت‌های کلیستانی و هلشتاین باقی ماند و حدود ۸ درصد از کل خوانش‌ها کنار گذاشته شد. درصد مکان‌یابی برای خوانش‌ها به‌طور میانگین به‌ترتیب برای کلیستانی و هلشتاین ۸۰،۷ و ۷۹،۳ بود. در جمعیت مورد مطالعه بین ۴ تا ۶ درصد خوانش‌ها همردیفی چندگانه داشتند. نتایج آنالیز بیان ژن افتراقی نشان داد از مجموع ۲۴۶۱۶ ژن و ۲۶۷۱۶ ایزوفرم، تعداد ۴۱ ژن بین دو زیرگونه بیان متفاوت داشتند (P<۰/۰۰۰۱۵). جدول ۱ فهرست این ژن‌ها به‌همراه شناسه آن‌ها در بانک اطلاعاتی Ensemble، موقعیت کروموزومی و نوکلئوتیدی آنها و ارزش P مربوطه را نشان می‌دهد.

از میان این ۴۱ ژن تنها ۱ ژن ۲ ایزوفرم داشت. این ژن بروی کروموزوم ۵ از نوکلئوتید ۹۹۶۴۷۱۲۲ تا نوکلئوتید ۹۹۷۰۱۶۳۰ از نظر عملکرد مشابه ژن‌های NKG2A و NKG2B (پروتئین پیوسته غشایی نوع دوم) است که نام دو ایزوفرم آن ENSBTAT00000056514 و ENSBTAT00000039039 (۷۱۱ جفت باز، ۶ اگزون) ثبت شده است. ایزوفرم ENSBTAT00000056514 یک اگزون بیشتر (۷۲۹ جفت باز) از ایزوفرم دیگری دارد. شکل نقطه‌ای (شکل ۳) میزان انحراف هر ژن را از میانگین بیان ژن در هر دو جمعیت نشان می‌دهد. فاصله بیشتر از خط میانگین نشان‌دهنده بیان متفاوت‌تر می‌باشد. ژن‌های کاملاً متفاوت بیان‌شده بر روی محورهای X و Y آن نمودار و داخل کادر مشخص شده‌اند.

تفاوت بیان ژن بین دو گونه به‌صورت هیستوگرام در شکل ۴ آورده شده است. همچنان‌که در شکل نشان داده شده است بخشی از تفاوت بیان ژن ناشی از آن است که برخی از ژن‌ها فقط در جمعیت کلیستانی بیان شده‌اند و بیان آنها در جمعیت هلشتاین صفر بوده است. از جمله بیشترین میزان تفاوت بیان ژن مربوط به یک ژن میتوکندریایی به نام ENSBTAG00000043545 بود که با مقدار FPKM،

۱۳۵۳۱ فقط در جمعیت کلیستانی بیان شده است. این ژن در موش به عنوان ژن کدکننده tRNA و در سنتز اسیدآمینه فنیل‌آلانین یکی از بیست اسید آمینه‌ی اصلی یاخته‌های زنده، نقش دارد. این ژن درگاو تاکنون حاشیه‌نویسی‌نشده و عملکردی برای آن گزارش نشده است.

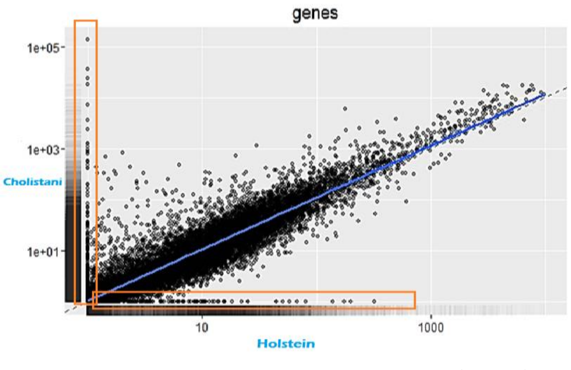
جدول ۱. کد، نام اختصاری و جایگاه ژن‌های متفاوت بیان‌شده در بین دو جمعیت هلشتاین و کلیستانی

Table 1. The code, the abbreviated name and the position of the generals indicate them between the two Holstein and Colonial populations

Annotation of the genes code	Gene name	Gene locus*	p.value
ENSBTAG00000001321	IL1B	11:46410276-46418787	0.00005
ENSBTAG00000001785	TGM3	13:53261634-53304213	0.00005
ENSBTAG00000004322	FOS	10:86883738-86887169	0.00005
ENSBTAG00000007101	F3	3:49110155-49121034	0.0001
ENSBTAG00000007296	-	1:82549641-82550013	0.00005
ENSBTAG00000010069	EGR1	7:51438726-51442500	0.0001
ENSBTAG00000011161	-	8:38683732-38684116	0.0001
ENSBTAG00000012046	JUNB	7:13851130-13852933	0.00005
ENSBTAG00000014332	-	5:99647122-99701630	0.00005
ENSBTAG00000016688	-	X:66722398-66723397	0.00005
ENSBTAG00000017448	EFEMP1	11:38338743-38408288	0.0001
ENSBTAG00000018506	-	5:58013469-58014035	0.00005
ENSBTAG00000020350	DUSP2	11:2250454-2252492	0.00005
ENSBTAG00000021166	FAM71A	16:72811910-72814133	0.0001
ENSBTAG00000027787	-	18:65621141-65621509	0.00005
ENSBTAG00000024311	-	21:60886044-60886971	0.00005
ENSBTAG00000030735	COX7B	7:42303402-42303827	0.00005
ENSBTAG00000031458	-	4:101019787-101021837	0.00015
ENSBTAG00000031814	SDS	17:63302946-63311098	0.00015
ENSBTAG00000032308	U1	3:21433393-21433557	0.00005
ENSBTAG00000034170	DYNLL1	20:11746000-11746270	0.00005
ENSBTAG00000034824	-	20:2680573-3054892	0.0001
ENSBTAG00000037147	U3	20:64476312-64476463	0.00005
ENSBTAG00000037600	RPS15A	8:84843153-84843546	0.00005
ENSBTAG00000037778	CXCL3	6:90811061-90813079	0.00005
ENSBTAG00000038064	-	2:11700813-11701755	0.00005
ENSBTAG00000038411	-	8:62722029-62722587	0.00005
ENSBTAG00000039813	GZMB	21:35135770-35137852	0.00005
ENSBTAG00000043545	-	MT:5686-7299	0.00005
ENSBTAG00000044208	DUSP4	27:24815380-24828237	0.00005
ENSBTAG00000045497	-	21:66764060-66764393	0.0001
ENSBTAG00000045544	EIF2S3Y	X:144159055-144234222	0.00005
ENSBTAG00000045750	DDX3Y	X:143364825-143375658	0.00005
ENSBTAG00000045914	-	7:21247238-21247649	0.00005
ENSBTAG00000047277	PNKD	18:46516556-46516985	0.00005
ENSBTAG00000047354	-	5:32529438-32529956	0.00015
ENSBTAG00000047766	G0S2	16:75539495-75540404	0.00015
ENSBTAG00000047966	-	16:6155902-6156793	0.00005
ENSBTAG00000048102	-	X:143736737-143800954	0.00015
ENSBTAG00000048172	ZRSR2Y	X:142664999-142690152	0.00005
ENSBTAG00000048229	TPT1	25:1394200-1436118	0.00005

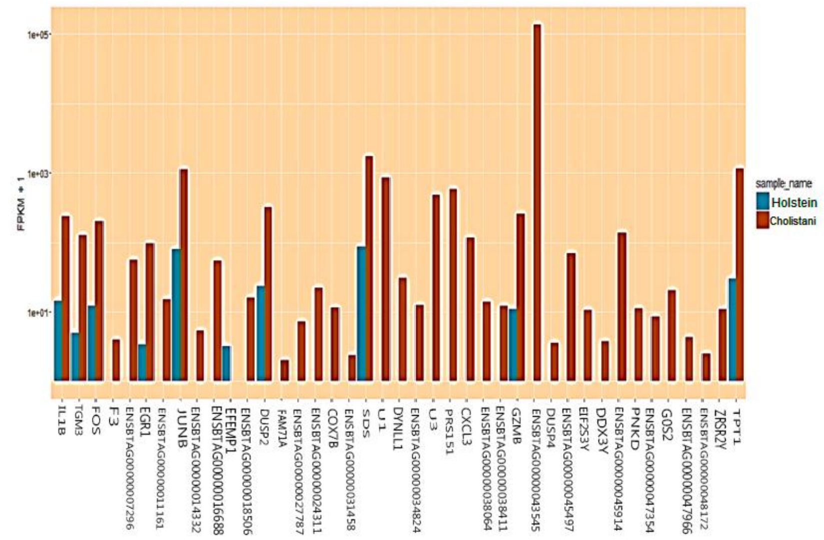
در ستون جایگاه ژن، عدد سمت چپ شماره کروموزوم و اعداد بعد از آن شماره نوکلئوتید شروع و پایان ژن را نشان می‌دهد.

In the gene locus column, the left number represents the chromosome number and the subsequent nucleotide number indicates the start and end of the gene.



شکل ۳. نمودار نقطه‌ای تفاوت بیان ژن بین دو جمعیت هلشتاین و کلیستانی

Figure 3. Scatter plot of differential gene expression between Holstein and Cholistani population



شکل ۴. نمودار تفاوت بیان ژن بین دو جمعیت هلشتاین و کلیستانی

Figure 4. Plot of differential gene expression between Holstein and Cholistani population

نمودارهای حرارتی[1] و آتشفشانی[2] در شکل‌های ۵ و ۶ بیان تصویری آنالیز بیان افتراقی را نمایش می‌دهند. نمودار حرارتی تفاوت بیان را با تغییر رنگ نشان می‌دهد به‌گونه‌ای که تغییر در شدت رنگ میزان تفاوت بیان ژن‌ها می‌باشد و در نمودار آتشفشانی هرچه از مرکزآتشفشان دور می‌شویم بیان متفاوت‌تر می‌شود. در پژوهشی که McCabe *et al.* (2012) با استفاده از فناوری RNA-seq، بیان ژن بافت کبد را در ۵ گاو هلشتاین در تعادل منفی انرژی خفیف (MNEB) و ۶ گاو در تعادل منفی شدید (SNEB) مقایسه نمودند و گزارش کردند از ۱۲۸۳۳ ژن، ۴۱۳ ژن بافت کبدی به‌طور معنی‌داری متفاوت بیان شدند وتغییرات زیادی در سطح ترانسکریپتوم گاوهای در تعادل منفی شدید (SNEB) مشاهده شد که این مربوط به متابولیسم اسیدهای چرب کبدی بود. در پژوهشی دیگر Scholey *et al.* (2015) نیز با استفاده از این فناوری، بیان ژن افتراقی و آنالیز شبکه ژنی را در نمونه‌های پوست بیوپسی شده از جراحت‌های تورم پوست بین انگشتان پنج گاو را با پنج گاو سالم مقایسه نمودند و افزایش ژن‌های keratin6A و interleukin1b و کاهش بیان ژن‌های مرتبط با کراتین را در جراحت‌های ناشی از این بیماری گزارش کردند. Huang *et al.* (2012) نیز

1. heatmap plot
2. Volcano plot

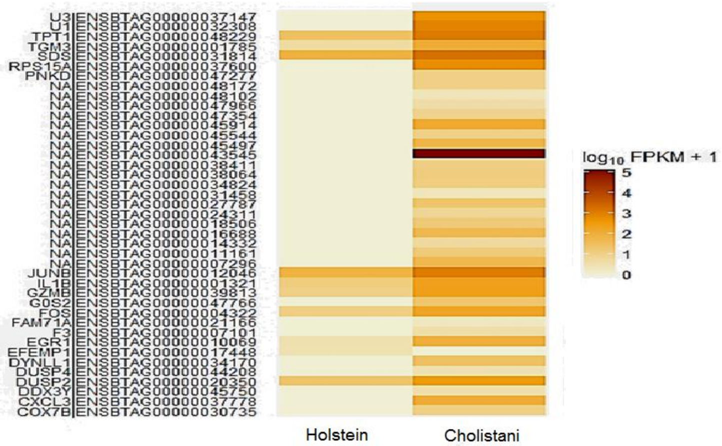
در پژوهشی ویژگی‌های ترانسکریپتوم گلبول‌های سفید خون (لوکوسیت) را در دو نژاد گاو تائورین (هلشتاین و جرزی) و یک نژاد ایندیسین (کُلیستانی)، تعیین و مقایسه نمودند. در سطح ژنومی، تغییرات بازی اختصاصی نژادی در نواحی کدکننده پروتئین شناسایی شدند. از میان ۷،۷۹۳،۴۲۵ باز کدکننده، تنها ۱۶۵ باز بین هلشتاین و جرزی و ۳،۳۸۳ باز بین هلشتاین و کلیستانی متفاوت بودند که از این میان ۸۱۷ باز (۲۵٪) به تغییر اسیدآمینه‌ای در ۶۲۷ ژن منجر می‌شدند. آنالیز بیان ژن تفریقی حاکی از شباهت زیاد بین هلشتاین و جرزی و تفاوت خیلی بیشتری بین نژادهای تائورین و ایندیسین بود.

آنالیز ماهیت ژنی براساس ۳ دسته‌بندی عمومی فرآیندهای زیستی[3]، عملکرد مولکولی[4] و اجزای سلولی[5] انجام شد. شکل‌های ۷، ۸ و ۹ به‌ترتیب نتایج آنالیز ماهیت ژن را برای ۳ دسته‌بندی عمومی فرآیندهای زیستی، عملکرد مولکولی و اجزای سلولی نشان می‌دهند.

آنالیز مسیر (Pathway analysis) (شکل ۱۰) نیز نشان داد ژن‌های متفاوت بیان شده بیشتر در مسیرهای مرتبط با ایمنی با واسطه سلولی (فعال شدن سلول‌های T)، پاسخ به تنش (تنش اکسیداتیو)

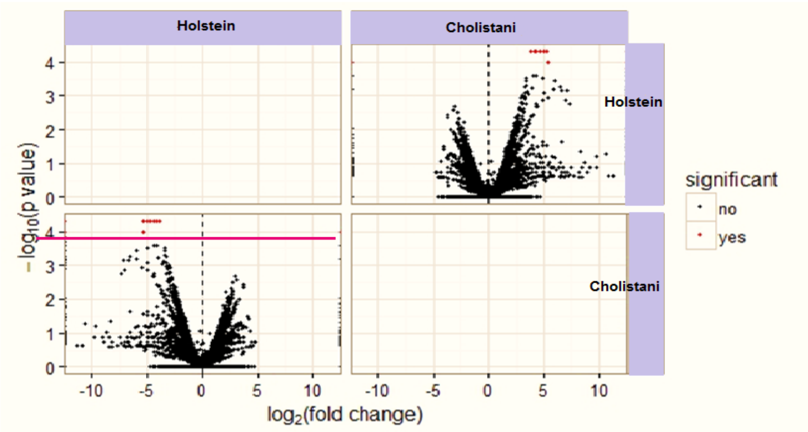
3. Biological process
4. Molecular function
5. Cellular component

وآنژیوژنزیز (تشکیل رگ‌های خونی جدید) درگیر هستند. به عبارت دیگر، این ژن‌ها در مسیرهای منجر به تحمل تنش های محیطی بویژه تنش حرارتی، ایمنی و مقاومت به بیماری نقش دارند. این مسیرها همان‌هایی هستند که در طول زمان موجب سازگاری دو زیرگونه مورد مطالعه با اقلیم‌های خاص آن‌ها شده و همچنین باعث بروز تفاوت‌های فنوتیپی بارز در این صفات بین آن‌ها گردیده است.



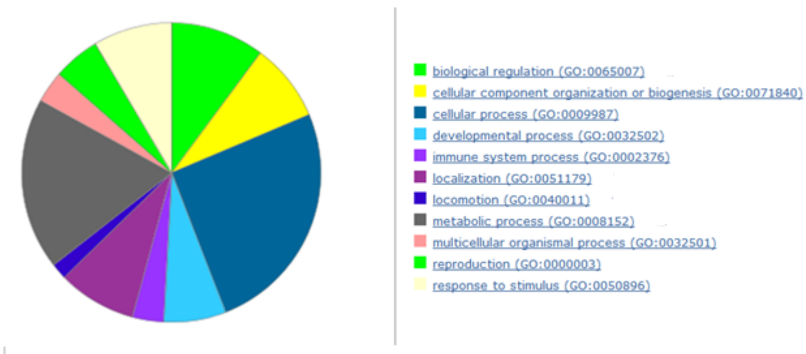
شکل ۵. نقشه حرارتی تفاوت بیان ژن بین دو جمعیت هلشتاین و کلیستانی

Figure 5. Heatmap Plot of differential gene expression between Holstein and Cholistani population



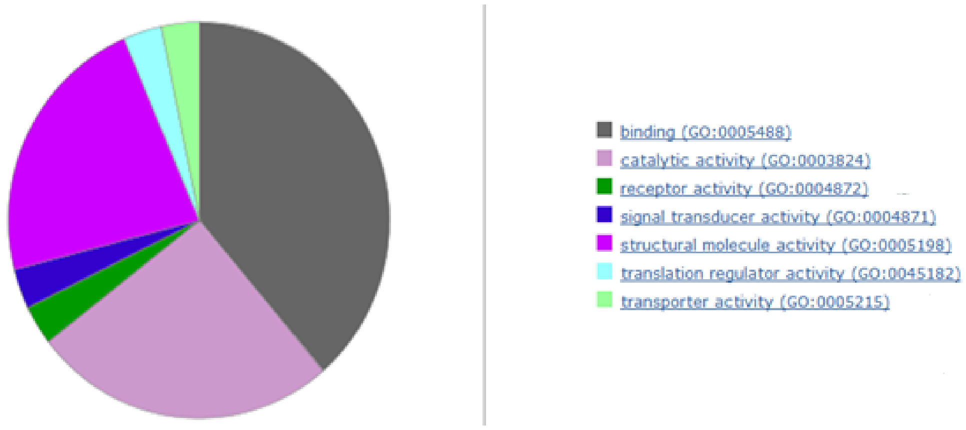
شکل۶. نمودار آتشفشانی آنالیز بیان ژن تفریقی بین دو جمعیت هلشتاین و کلیستانی

Figure 6. Volcano plot of differential gene expression between Holstein and Cholistani population



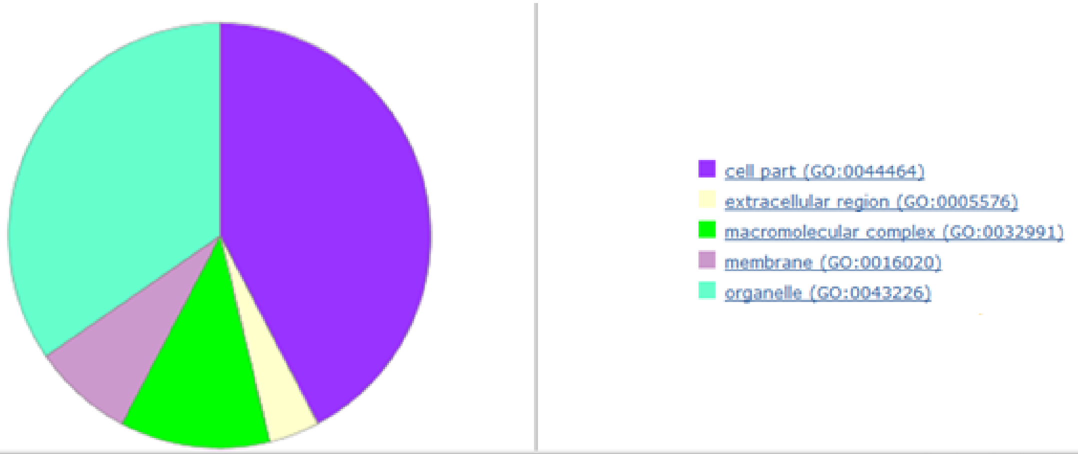
شکل ۷. دسته‌بندی ماهیت ژنی براساس فرآیند زیستی، راهنمای رنگی کنار نمودار جزئیات انواع فرآیندهای زیستی را نشان می‌دهد.

Figure 7. The Gene Ontology classification of the biological process, the color guide along the graph shows the details of the types of biological processes



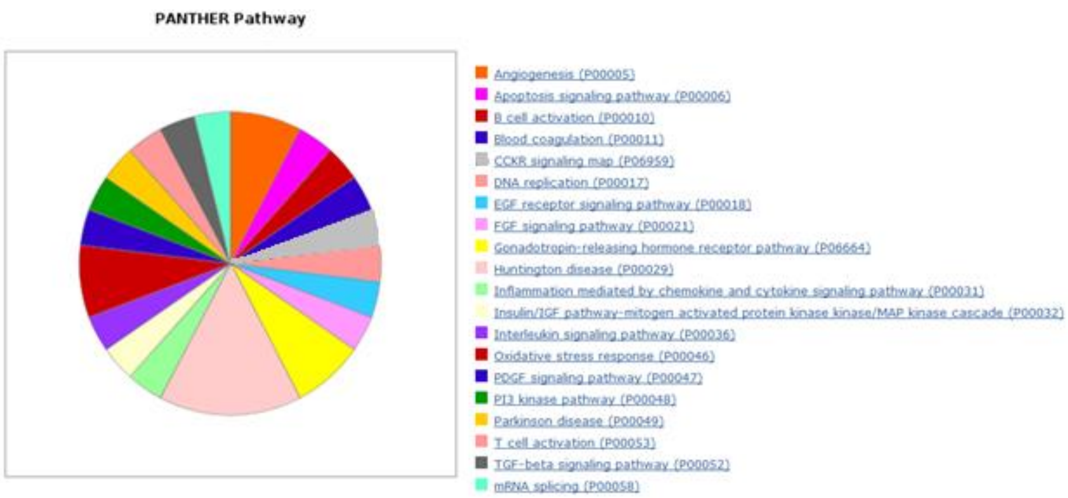
شکل ۸. دسته‌بندی ماهیت ژنی براساس عملکرد مولکولی، راهنمای رنگی کنار نمودار جزئیات انواع عملکردهای مولکولی را نشان می‌دهد.

Figture 8. The Gene Ontology classification of the Molecular function, the color guide along the graph shows the details of the types of Molecular function



شکل ۹. دسته‌بندی ماهیت ژنی براساس اجزای سلولی، راهنمای رنگی کنار نمودار جزئیات انواع اجزای سلولی را نشان می‌دهد.

Figure 9. The Gene Ontology classification of the Cellular component, the color guide along the graph shows the details of the types of Cellular component



شکل ۱۰. نمودار آنالیز مسیر در ژن‌های متفاوت بیان‌شده. راهنمای رنگی کنار نمودار جزئیات انواع مسیرها را نشان می‌دهد.

Figure 10. Pathway analysis for the differential gene expression. The color guide along the graph shows the details of the type’s pathway analysis

از این‌رو دو بخش از تفاوت‌های آشکار فنوتیپی مرتبط با تحمل تنش‌های محیطی به‌ویژه تنش حرارتی و مقاومت به بیماری بین این دو گونه در پروفایل بیان ژن‌های آنها نیز بروز یافته است. این امر مقاومت و تحمل بیشتر جمعیت‌های بومی را در مقایسه با جمعیت هلشتاین تائید کرده و بهره‌برداری از این پتانسیل ژنتیکی منحصربه‌فرد را توصیه می‌نماید.

در نتیجه‌گیری کلی می‌توان گفت که نتایج حاصل از آنالیز بیان ژن افتراقی بین این دو زیرگونه با تفاوت‌های فنوتیپی آن‌ها منطبق بود. از این‌رو انجام چنین مطالعاتی در روشن‌نمودن پس‌زمینه ژنتیکی تفاوت‌های فنوتیپی می‌تواند مؤثر باشد. همچنین با توجه به شباهت زیاد و نزدیکی جغرافیایی جمعیت گاو کلیستانی پاکستان و گاو سیستانی ایران امکان تعمیم نتایج به این گاو بومی کشور دور از ذهن نیست. در نهایت، انجام چنین مطالعاتی به‌منظور مقایسه جمعیت‌های خالص و آمیخته نیز توصیه می‌شود که می‌تواند از هزینه‌های زیاد ارزیابی عملکرد برای انتخاب بهترین آمیخته‌ها جلوگیری نماید.

REFERENCES

1. Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
2. Bae, J. S., Cheong, H. S., Kim, L. H., NamGung, S., Park, T. J., Chun, J. Y. & Shin, H. D. (2010). Identification of copy number variations and common deletion polymorphisms in cattle. *BMC genomics*, 11(1), 1.
3. Banabazi, M. H., Naserkheil, M. & Miraei-Ashtiani, S. R. (2012a). Gene regulatory network for cell cycle of *Saccharomyces cerevisiae* based on weighted correlation. *The third Iranian conference on agricultural biotechnology of Iran*. Ferdowsi University of Mashhad. (in Farsi)
4. Banabazi, M. H., Naserkheil, M. Miraei-Ashtiani, S. R. (2012[b]). An algorithm for identifying differential expressed genes in the *Saccharomyces cerevisiae* using microarray data by R packages. *The third Iranian conference on agricultural biotechnology of Iran.* Ferdowsi University of Mashhad (in farsi)
5. Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
6. Dorak, M. T. (Ed.). (2007). Real-time PCR. Taylor & Francis. Utech, K. B. W., Wharton, R. H., & Kerr, J. D. (1978). Resistance to *Boophilus microplus* (Canestrini) in different breeds of cattle. *Crop and Pasture Science*, 29(4), 885-895.
7. Ekblom, R. & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1-15.
8. Fries, R. & Ruvinsky, A. (1999). The Genetics of Cattle. New York: CABI Publising.
9. Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K. & Chen, X. (2010). Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. *Cell Research*, 20(7),763-783.
10. Haas, B. J. & Zody, M. C. (2010). Advancing RNA-seq analysis. *Nature Biotechnology*, 28(5), 421.
11. Hansen, P. J. (2004). Physiological and cellular adaptations of zebu cattle to thermal stress. *Animal Reproduction Science*, 82, 349-360.
12. Huang, W., Nadeem, A., Zhang, B., Babar, M., Soller, M. & Khatib, H. (2012). Characterization and comparison of the leukocyte transcriptomes of three cattle breeds. *PLoS One*, 7(1), e30244.
13. Langmead, B. & Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357-359.
14. Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.
15. Marguerat, S. & Bähler, J. (2010). RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences*, 67(4), 569-579.
16. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. (2008). Mapping and quantifying mammalian transcriptoms by RNA-seq. *Nature Methods*, 5(7), 621- 628.
17. Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M. & Schmidt, D. (2008). *A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science*, 321, 956-960.
18. Wilhelm, B. T. & Landry, J. R. (2009). RNA-Seq, quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, 48(3), 249-257.