

کشف و فراخوانی چندشکلی های تک نوکلئوتیدی بر روی توالی ژن های متفاوت بیان شده بین دو جمعیت گاو هلشتاین (بوس تاروس) و کلیستانی (بوس ایندیکوس)

- مینا سلیم پور
دانشجوی دکتری گروه علوم دامی، دانشکده علوم کشاورزی، دانشگاه گیلان.
- محمد حسین بنابازی (نویسنده مسئول)
استادیار، موسسه تحقیقات علوم دامی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران

تاریخ دریافت: مهر ۱۳۹۸ تاریخ پذیرش: تیر ۱۳۹۹

شماره تماس نویسنده مسئول: ۰۲۶۳۴۲۵۶۱۲۸

Email: m.banabazi@areeo.ac.ir

شناسه دیجیتال (DOI): 10.22092/asj.2020.123350.1879

چکیده

در مطالعه‌ی حاضر چندشکلی‌های تک نوکلئوتیدی بر روی توالی ژن های متفاوت بیان شده بین دو جمعیت گاو هلشتاین و کلیستانی بررسی شد. برای این منظور ترانسکریپتوم (توالی کل mRNA) از پایگاه SRA استخراج و سپس از طریق همردیفی و مکان یابی خوانش های RNA-seq بروی ژنوم مرجع گاو، آنالیز بیان ژن افتراقی میان دو جمعیت انجام شد. در نهایت از میان ۲۴۶۱۶ ژن و ۲۶۷۱۶ ایزوفرم شناخته شده در گونه گاو، ۴۱ ژن بین دو جمعیت مورد مطالعه بطور معنی داری متفاوت بیان داشتند ($p < 0.000015$). آنالیز ماهیت شناسی ژن و مسیرهای بیولوژیکی نشان داد این ژن‌ها در ۲۰ مسیر درگیر می باشند. اکثر این ژن‌ها در مسیرهای پاسخ ایمنی، زنجیره انتقال الکترون، تحمل تنش حرارتی و مقاومت به بیماری درگیر هستند. کشف و فراخوانی چندشکلی‌های تک نوکلئوتیدی بر روی توالی کل ترانسکریپتوم این دو جمعیت به فهرستی شامل به ترتیب ۵۳۴۷۸ و ۱۴۵۴۴۳ مورد SNP بر روی ترانسکریپتوم آنها انجامید. از این میان، تعداد ۲۴ مورد SNP در هلشتاین و ۱۵۴ مورد در کلیستانی بر روی توالی ۲۳ ژن از مجموع ۴۱ ژن متفاوت بیان شده قرار داشتند ۱۵۷ مورد SNP جدید و ۲۱ مورد دارای شماره rs در پایگاه های اطلاعاتی بودند. احتمالاً تفاوت بیان ژنهای حامل SNP را می توان به SNP های واقع بر توالی آنها نسبت داد. تعیین ماهیت این SNP ها و مشخص نمودن عامل یا غیرعامل بودن آنها، مطالعه بیان اختصاصی آلی و محاسبه پوشش ترانسکریپتومی هر یک از آنها به روشتر شدن صحت و ساز و کار این موضوع کمک خواهد نمود.

واژه‌های کلیدی: کشف و فراخوانی SNP، گاو، ترانسکریپتوم، RNA-Seq، بیان ژن افتراقی.

Animal Science Journal (Pajouhesh & Sazandegi) No 130 pp: 203-214

The Single Nucleotide Polymorphisms (SNP) discovery and calling on genes differentially expressed between Holstein (*Bos taurus*) and Cholistani (*Bos indicus*) cattle populations.

By: Mina Salimpour¹, Mohammad Hossein Banabazi²

¹ Ph.D candidate, Department of Animal Sciences, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran

² Assistant Professor, Department of Biotechnology, Animal Science Research Institute of Iran (ASRI), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran.

Received: October 2019

Accepted: July 2020

In the present study, the single-nucleotide polymorphisms and their transcriptome coverage on differential gene expressions were investigated between Holstein and Cholistani population. Thus, the transcriptomes (mRNA Sequence) for a US Holstein and a Pakistanian Colistani cow population were assembled by aligning and mapping the RNA-Seq reads on the bovine reference genome. The differential gene expression analysis was performed. Finally, 24616 genes and 26 716 isoforms on the transcriptome of these two populations were found among 41 genes that were showed significantly different expression. (P.value < 0,000015). Analysis of the gene ontology (GO) and routes indicated that they are involved in, 20 pathways A large number of genes in the pathways leading to immune response, leading to translate the electron transfer resistance to thermal stress and disease. A total of 53 478 and 145443 have SNP discovered in the genome of which 154 SNP in Cholistani population and 24 SNP in Holstein population were among 41 genes were significant In 18 genes, no SNP was observed. 23 genes had at least one SNP in each of the populations. The number of SNP in the Cholistani breed was almost 5 times expression higher than the Holstein breed. These genes also had a high expression in the Cholistani population breed. This can be one of the causes of differences in gene expression between two populations in this study.

Key words: SNP discovery and calling, Cattle, Transcriptome, RNA-Seq, Differential Gene Expression.

مقدمه

اصولی مطالعه حاضر می باشد، از توالی (RNA-Seq) RNA استفاده شده است. این داده ها واجد تمامی ویژگیهای مورد هدف این مطالعه را است. این توالی ها در سطح ترانسکریپتوم، امکان بررسی بیان دقیق و کمی همه ژنها را همزمان با در اختیار داشتن توالی هر ژن و مطالعه SNP های واقع بر آنها در اختیار قرار می دهد. امروزه استفاده اطلاعات ژنتیکی مبتنی بر SNP ها در برنامه های اصلاح نژاد گاو به عنوان ابزاری برای بهبود انتخاب های فنوتیپی مرسوم است. همچنین پیشرفت در ژنتیک مولکولی و ظهور فناوری های جدید منجر به تشخیص SNP های مرتبط با ژن های موثر بر صفات اقتصادی و شناسایی واریانت های ژنتیکی

متداول ترین و فراوانترین شکل تفاوت ژنتیکی در سطوح مختلف و متنوع ژنتیکی از فرد تا گونه و بالاتر از آن، چندشکلی های تک نوکلئوتیدی هستند که فراوانی بیش از ۱ درصد در یک جمعیت داشته باشد.

<http://www.nature.com/scitable/definition/single-nucleotide-polymorphism-snp-295> وقتی این تفاوتها در توالی های رمزگر و یا تنظیمی واقع شوند تاثیرشان در بیان ژن و محصولات آنها نیز بروز خواهد نمود. به منظور بررسی چندشکلی های تک نوکلئوتیدی در توالی های رمزگر و انطباق آنها با میزان تفاوت بیان ژنها در بین دو زیرگونه گاو که هدف

ایندیکوس (کلیستانی) با استفاده از داده های هوانگ و همکاران (۲۰۱۲) انجام شده است.

مواد و روش ها

داده های مورد استفاده

داده های 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 و همکاران، ۲۰۱۲). به این منظور نمونه های خون بلافاصله پس از اخذ از سیاهرگ دمی برای هر گروه با هم مخلوط شد. سپس گلبول های سفید خون (لکوسیت) جدا شده و RNA از آنها استخراج گردید. سپس نمونه های RNA پلی آدلینه شده و کتابخانه ی mRNA از آنها تهیه شد. این کتابخانه ها با استفاده از Illumina Genome Analyzer IIx. توالی یابی گردیدند. خوانش های حاصل از توالی به صورت جفتی (paired) یا دوسویه و با طول ۷۵ جفت باز بود.

داده های خام در فرمت sra ، <http://www.ncbi.nlm.nih.gov/sra/SRS454433> و <http://www.ncbi.nlm.nih.gov/sra/SRS454432> به ترتیب برای نژاد کلیستانی و هلشتاین از سایت NCBI دریافت شد و کلیه آنالیزها در محیط سیستم عامل لینوکس اوبونتو ۱۴۰۴ و بر روی کلاستر گروه علوم دامی دانشگاه تهران مورد آنالیز قرار گرفت. مراحل آنالیز شامل: تبدیل فرمت sra به fastq با استفاده از نرم افزار sratoolkits نسخه ۲.۵.۲ و سنجش کیفیت داده ها 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 به روش کلاسیک توالی یابی سانجر بر مبنای از صفر تا ۴۰ بود. سپس براساس نتایج سنجش کنترل کیفیت ویرایش خوانش ها با استفاده از نرم افزار Trimmomatic (Bolger و همکاران، 2014) نسخه ۰.۳۳ انجام شد. این ویرایش شامل حذف آداپتورها و خوانش های بی کیفیت بود. ترانسکرپتوم برای هر یک از دو زیر

عامل دخیل در فنوتیپ مورد نظر گردیده است (Djari). و همکاران، 2013). فناوری تعیین توالی RNA^۲، داده هایی دقیق-تر و کمی تر از انواع دیگر داده های مطالعه بیان ژن می-باشند (Wang و همکاران، 2009; Flintoft, 2008). اساس روش RNA-seq بر برش mRNA های رونویسی شده (یا به عبارت دیگر کل ترانسکرپتوم) به خوانش های بسیار کوتاه و توالی یابی این قطعات با استفاده از چارچوب های فنی پربرونداد متنوعی از جمله illumina /Solexa و roch454 و ABI Solid است (Marguerat and Bahler, 2010; Hass and zody, 2010). بدین ترتیب، تفاوت میزان رونویسی در بین قطعات کوتاه RNA بصورت تفاوت در تعداد دفعات خوانش آنها به هنگام توالی یابی بروز می نماید (Eklom and galindo, 2010; Wilhelm and landry, 2009). این فناوری یک روش کارآمد و به صرفه برای شناسایی SNPها در گونه های اهلی و با هزینه ی کم در نواحی رونویسی شده است (Canovas و همکاران، 2010). بر اساس آخرین اطلاعات ژنوم مرجع گاو، تاکنون بیش از ۲۴۰۰۰ ژن و ۲/۲ میلیون SNP بروی ۲/۸۷ گیگاباز طول ژنوم گاو شناسایی شده است (Bae و همکاران، 2010). دو زیر گونه ی گاو وجود دارد. زیرگونه ی *Bos taurus* که از اروپا و شمال آمریکا منشا یافته و بیشتر ویژه مناطق معتدل است و زیرگونه *Bos indicus* که کوهان دار (زبو) بوده و در مناطق حاره ای پراکنده است. این گونه از جنوب آسیا منشا یافته است (Fries and Ruvinsky, 1999). تفاوت فنوتیپی قابل توجهی بین این دو زیرگونه وجود دارد، از جمله این که گاو های زبو تنش های محیطی را بهتر تحمل می کنند و در مقابل انگل ها و بیماری ها مقاوم تر هستند (Dorak و همکاران، 1978; Hansen, 2004). هوانگ و همکاران (۲۰۱۲) داده های RNA-seq را برای سه جمعیت گاو هلشتاین، کلیستانی و جرسی تولید کردند و پروفایل بیان ژن این سه جمعیت را مقایسه کردند. مطالعه ی حاضر با هدف شناسایی چندشکلی های تک نوکلئوتیدی در ژن های متفاوت بیان شده در بین دو زیر گونه ی گاو بوس تاروس (هلشتاین) و بوس

² - RNA-seq

آنالیز ماهیت و مسیر های ژنی به صورت آنالیز بروی بانک اطلاعاتی (<http://pantherdb.org/>) انجام شدند.

کشف SNP

کشف SNP بروی ترانسکریپتوم تشکیل شده و فیلتر نمودن آن ها با استفاده از نرم افزارهای نسخه ی 0.1.19 samtools و Bedtools انجام شد. همچنین تعداد SNPهایی از آخرین به روز رسانی اطلاعات مربوط به واریانت های ژنوم گاو (فایل vcf) که در توالی ترانسکریپتوم تشکیل شده حضور داشتند نیز محاسبه گردید. فهرست SNPهای فیلتر شده با فهرست ژن های متفاوت بیان شده تطبیق داده شد و SNPهای واقع بر توالی این ژن ها استخراج گردید در آخر SNP ها در پایگاه اطلاعاتی مربوطه (<https://asia.ensembl.org/index.html>) مورد

بررسی قرار گرفتند. بحث و نتایج

پس از تبدیل فرمت به ترتیب ۲۱,۰۷۸,۴۷۷ و ۲۰,۹۴۰,۰۶۳ خوانش کوتاه جفتی (از هر دو انتها توالی یابی شده) به طول ۷۵ جفت باز، برای جمعیت های هلشتاین و کلیستانی بدست آمد. شاخص ده گانه کنترل کیفیت حاکی از کیفیت مناسب داده ها به ازای هر باز و خوانش صحیح همه ی آن ها براساس میزان، تعداد N به ازای هر باز بود (سلیم پور، ۱۳۹۵). درصد مکان یابی برای خوانش ها به طور میانگین به ترتیب برای کلیستانی و هلشتاین ۸۰.۷ و ۷۹.۳ بود.

شناسایی ژن ها

خروجی تجزیه و تحلیل داده ها و نتایج آنالیز بیان ژن افتراقی ۲۴۶۱۶ ژن و ۲۶۷۱۶ ایزوفرم بر روی ترانسکریپتوم این دو جمعیت را نشان داد که در این میان ۴۱ ژن شناسایی شدند که بیان متفاوت و معنی داری را داشتند (p.value ۰/۰۰۰۰۱۵) (جدول ۱).

گونه براساس همردیفی خوانش ها و مکان یابی آن ها بروی ژنوم مرجع به ترتیب با نرم افزار Bowtie2 و TopHat2 نسخه ۲,۰,۸ انجام گردید (Langmead and Salzberg, 2012). ژنوم رفرنس مورد استفاده برای تشکیل ترانسکریپتوم UMD3.1 (ویرایش ۸۲) بود.

اطلاعات حاشیه نویسی به کار رفته نیز مربوط به این ویرایش ذکر شده بود. ژنوم مرجع و حاشیه نویسی آن از بانک اطلاعاتی Ensemble دریافت شد

ftp://ftp.ensembl.org/pub/release82/fasta/bos_taurus/dna/Bos_taurus.UMD3.1.dna.toplevel.fa.gtf
ftp://ftp.ensembl.org/pub/release-z-84/gtf/bos_taurus/Bos_taurus.UMD3.1,84gtf.

اندازه ی فاصله ی درون کتابخانه ای و انحراف معیار آن برای استفاده از کد TopHat2 با استفاده از (Li & Durbin, 2009) bwa و به روش نمونه گیری از داده ها انجام شد که به ترتیب برای هلشتاین ۱۰۰ و ۱۵ برای کلیستانی ۱۸۰ و ۴۵ برآورد گردید.

آنالیز بیان افتراقی ژن

تشکیل پروفایل بیان ژن برای هر یک از دو زیرگونه با استفاده از پکیج Tuxedo (Trapnell و همکاران، ۲۰۱۲) انجام شد. در این مجموعه نرم افزاری تلفیق دو ترانسکریپتوم با استفاده از Cuffdiff و آنالیز بیان افتراقی ژن با استفاده از Cuffdiff به عمل آمد. مصورسازی نتایج در قالب گراف با استفاده از پکیج CummeRbund تحت نرم افزاری (<http://www.bioconductor.org/>) انجام شد.

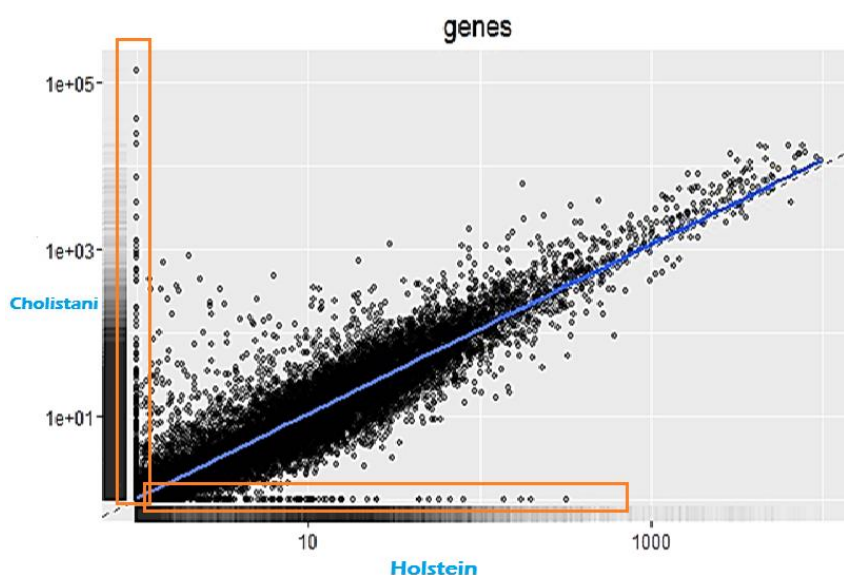
جدول ۱. نتایج مربوط به بیان افتراقی ژن ها در هر یک از دو جمعیت

حاشیه نویسی کد ژن های	نام ژن	سطح معنی داری	میزان بیان در کلیستانی (FPKM)	میزان بیان در هلشتاین (FPKM)
ENSBTAG00000001321	IL1B	0.00005	237.87	13.3081
ENSBTAG00000001785	TGM3	0.00005	123.648	3.83998
ENSBTAG00000004322	FOS	0.00005	197.806	11.0105
ENSBTAG00000007101	F3	0.0001	2.94292	0
ENSBTAG00000007296	-	0.00005	54.0777	0
ENSBTAG00000010069	EGR1	0.0001	94.2759	2.32824
ENSBTAG00000011161	-	0.0001	14.0117	0
ENSBTAG00000012046	JUNB	0.00005	1117.25	78.4322
ENSBTAG00000014332	-	0.00005	4.27815	0
ENSBTAG00000016688	-	0.00005	52.883	0
ENSBTAG00000017448	EFEMP1	0.0001	0	2.16751
ENSBTAG00000018506	-	0.00005	14.9134	0
ENSBTAG00000020350	DUSP2	0.00005	320.062	22.5135
ENSBTAG00000021166	FAM71A	0.0001	0.986594	0
ENSBTAG00000024311	-	0.00005	6.08931	0
ENSBTAG00000027787	-	0.00005	21.1625	0
ENSBTAG00000030735	COX7B	0.00005	10.2812	0
ENSBTAG00000031458	-	0.00015	1.37063	0
ENSBTAG00000031814	SDS	0.00015	1716.15	85.5355
ENSBTAG00000032308	U1	0.00005	854.165	0
ENSBTAG00000034170	DYNLL1	0.00005	29.1696	0
ENSBTAG00000034824	-	0.0001	11.2761	0
ENSBTAG00000037147	U3	0.00005	474.901	0
ENSBTAG00000037600	RPS15A	0.00005	582.344	0
ENSBTAG00000037778	CXCL3	0.00005	116.812	0
ENSBTAG00000038064	-	0.00005	12.9728	0
ENSBTAG00000038411	-	0.00005	11.074	0
ENSBTAG00000039813	GZMB	0.00005	254.454	9.68099
ENSBTAG00000043545	-	0.00005	135031	0
ENSBTAG00000044208	DUSP4	0.00005	2.48719	0
ENSBTAG00000045497	-	0.0001	68.6657	0
ENSBTAG00000045544	EIF2S3Y	0.00005	9.41077	0
ENSBTAG00000045750	DDX3Y	0.00005	2.74239	0
ENSBTAG00000045914	-	0.00005	134.128	0
ENSBTAG00000047277	PNKD	0.00005	10.1181	0
ENSBTAG00000047354	-	0.00015	7.47594	0
ENSBTAG00000047766	G0S2	0.00015	19.3221	0
ENSBTAG00000047966	-	0.00005	3.32136	0
ENSBTAG00000048102	-	0.00005	1.45347	0
ENSBTAG00000048172	ZRSR2Y	0.00005	9.83664	0
ENSBTAG00000048229	TPT1	0.00005	1145.45	28.909

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

ENSBTAT00000039039 (۷۱۱ جفت باز، ۱۶ اگزون) ثبت شده است. ایزوفرم ENSBTAT00000056514 دارای یک اگزون بیشتر (۷۲۹ جفت باز) نسبت به دیگری می باشد. نمودار نقطه ای (نمودار ۳) میزان انحراف هر ژن را از میانگین بیان ژن در هر دو جمعیت نشان می دهد. فاصله ی بیشتر از خط میانگین نشان دهنده ی بیان متفاوت تر می باشد. ژن های کاملا متفاوت بیان شده بروی محورهای X و Y آن نمودار و داخل کادر مشخص شده اند.

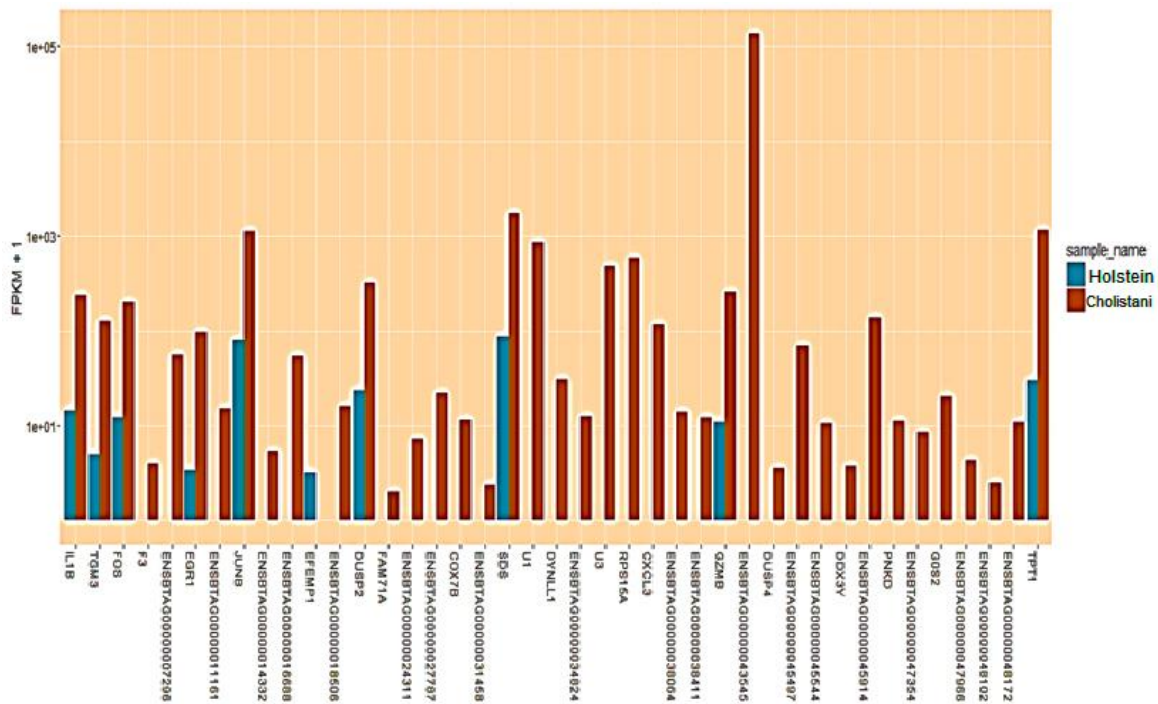
بیشترین میزان بیان در ژن های کروموزوم X به این دلیل قرارگیری ژن های مربوط به ایمنی بروی این کروموزوم و نیز به دلیل طولانی بودن این کروموزوم در بین ژنوم گاو بود. از میان ۴۱ ژن تنها ۱ ژن ۲ ایزوفرم داشت. این ژن بروی کروموزوم ۵ از نوکلئوتید ۹۹۶۴۷۱۲۲ تا نوکلئوتید ۹۹۷۰۱۶۳۰ از نظر عملکرد مشابه ژن های NKG2A و NKG2B (پروتئین پیوسته غشایی نوع ۲) است که نام دو ایزوفرم آن ENSBTAT00000056514 و



نمودار ۳. نمودار نقطه ای تفاوت بیان ژن بین دو جمعیت هلستاین و کلیستانی

بیان شده است. این ژن در موش به عنوان ژن کد کننده tRNA و در سنتز اسید آمینه فنیل آلانین یکی از بیست اسید آمینه ی اصلی ساخته های زنده، نقش دارد.
(<https://www.ncbi.nlm.nih.gov/search/all/term>)
این ژن در گاو تاکنون حاشیه نویسی نشده و عملکردی برای آن گزارش نشده است.

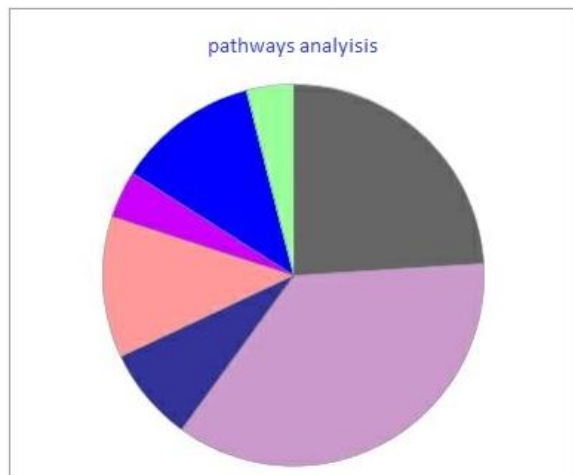
تفاوت بیان ژن بین دو گونه به صورت هیستوگرام در نمودار ۴ آورده شده است. همچنان که از این نمودار برمی آید بخشی از تفاوت بیان ژن ناشی از آن است که برخی از ژن ها فقط در جمعیت کلیستانی بیان شده اند و بیان آنها در جمعیت هلستاین صفر بوده است. از جمله بیشترین میزان تفاوت بیان ژن مربوط به یک ژن میتوکندریایی به نام ENSBTAG00000043545، FPKM ۱۳۵۳۱ فقط در جمعیت کلیستانی



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

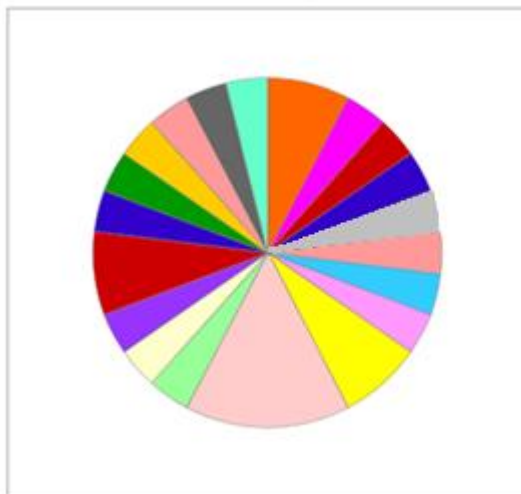
آمیلوئید متصل و عملکرد آن را متعادل کند. (Ohsawa و همکاران، ۲۰۰۱) همچنین با اتصال به فیبرونکتین علاوه بر چسبیدن سلول ها با تاثیر بروی برخی مسیر های درون سلولی، باعث مهار مهاجرت سلولی در تومور ها می شود. (Zheng و همکاران، ۲۰۱۵،

۴۰ ژن در جمعیت کلیستانی نسبت به جمعیت هلشتاین مقدار بیان (عدد FPKM) بیشتری داشت. ژن EFEMP1 تنها ژن موجود در بین ژن های معنی دار بود که در جمعیت هلشتاین بیان بیشتری را نشان می داد. این ژن فیولین را کد میکند. فیولین پروتئین ماتریکس خارج سلولی است و می تواند به پروتئین پیش ساز



- binding (GO:0005488)
- catalytic activity (GO:0003824)
- molecular function regulator (GO:0098772)
- molecular transducer activity (GO:0060089)
- structural molecule activity (GO:0005198)
- transcription regulator activity (GO:0140110)
- transporter activity (GO:0005215)

PANTHER Pathway



- Angiogenesis (P00005)
- Apoptosis signaling pathway (P00006)
- B cell activation (P00010)
- Blood coagulation (P00011)
- CCKR signaling map (P06959)
- DNA replication (P00017)
- EGF receptor signaling pathway (P00018)
- EGF signaling pathway (P00021)
- Gonadotropin-releasing hormone receptor pathway (P06664)
- Huntington disease (P00029)
- Inflammation mediated by chemokine and cytokine signaling pathway (P00031)
- Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032)
- Interleukin signaling pathway (P00036)
- Oxidative stress response (P00046)
- PDGF signaling pathway (P00047)
- PI3 kinase pathway (P00048)
- Parkinson disease (P00049)
- T cell activation (P00051)
- TGF-beta signaling pathway (P00052)
- mRNA splicing (P00058)

نمودار ۱۰. نمودار آنالیز مسیر در ژن های متفاوت بیان شده در جمعیت کلیستانی. راهنمای رنگی کنار نمودار جزئیات انواع مسیرها را نشان می دهد.

کشف SNP های موجود در ژن های با بیان متفاوت

در توالی ترانسکریپتوم تشکیل شده در جمعیت گاو هلشتاین و کلیستانی به ترتیب ۵۳۴۷۸ و ۱۴۵۴۴۳ مورد SNP در سراسر ژنوم کشف شد (بنابازی، ۱۳۹۵). برخی از این SNP ها جدید بوده و با تراشه های تجاری متراکم مورد استفاده برای تعیین ژنوتیپ گاو مشترک نبودند (Banabazi و همکاران، ۲۰۱۵). سپس این محققین کاربرد استفاده از فهرست SNP های مشترک با تراشه های تجاری را در پیش بینی های مبتنی بر ژنوم یک صفت مهم اقتصادی یعنی باقیمانده مصرف خوراک (RFI) در گاو هلشتاین

آنالیز مسیر (Pathway analysis) در جمعیت کلیستانی نیز نشان داد (نمودار ۱۰)، ژن های متفاوت بیان شده بیشتر در مسیرهای مرتبط با ایمنی با واسطه سلولی (فعال شدن سلول های T)، پاسخ به تنش (تنش اکسیداتیو) و آنژیوژنز (تشکیل رگ های 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 که با نتایج Huang و همکاران (۲۰۱۲) مطابقت داشت.

گاما گلوتامیل ترانسفراز E را کد گذاری می کند. Aeschlimann و همکاران، Wang ; 1998 و همکاران، (1995) که در تشکیل پلاکت سلولی در اپیدرم و فولیکول مو دخیل است. ژن SDS نیز در بیوسنتز و سوخت و ساز سرین و ترئونین و گلیسین نقش دارد و بیشترین دخالت در بیوسنتز سرین و آنزیم سرین دهیدراتاز را دارد که باعث تبدیل سرین به پیرووات میشود.

تا کنون بررسی های کمی با هدف کاوش SNP با استفاده از داده های RNA-Seq در دامهای اهلی انجام شده است در پژوهشی که Djari و همکاران، 2013 با استفاده از RNA-Seq انجام دادند، ۳۴۳۷۶ جایگاه SNP را در عضله سه گوساله نر لیموزین کشف کردند. ۵۵ درصد SNP ها در نواحی رمزگر یافت شدند و حدود ۲۲ درصد منجر به یک تغییر اسید آمینه ای شدند. در پژوهش های دیگر Park و همکاران، 2012 بر روی توالی ترانسکریپتوم حاصل از دو بافت خون و عضله شش اسب تارو برد مسابقه ، تعداد ۱۸۳۹۷۳ جایگاه SNP پیدا نمودند که اکثر آنها (۱۷۱۵۵۸ مورد، ۹۰/۳۱٪) در مقایسه با ۱/۱ میلیون SNP گزارش شده قبلی در دو بانک اطلاعاتی برای اسب جدید بودند. (Zhang و همکاران، 2013) نیز تعداد ۴۰۴۸۱ و ۳۸۸۵۱ جایگاه SNP را با استفاده از داده های RNA-seq در دو جمعیت گوسفند یافت نمودند که در مجموع تعداد ۵۹۱۳۹ جایگاه در دو جمعیت متفاوت گزارش کردند.

مطالعه نمودند. این فهرست کاهش یافته شامل حدود ۱۰۰ هزار SNP در مقایسه با حدود ۷۰۰ هزار مورد موجود در تراشه های متراکم تجاری رایج منجر به بهبود حدود ۲ درصدی در برآورد ارزشهای اصلاحی ژنومی گردید (بنابازی و همکاران ۱۳۹۵). نتایج بررسی ژن های گزارش شده در جدول نشان داد تعداد ۱۵۴ مورد SNP در نمونه جمعیت کلیستانی و ۲۴ مورد در نمونه جمعیت هلشتاین در بین ۴۱ ژن بودند. در ۱۸ ژن SNP مشاهده نشد. ۲۳ ژن دارای حداقل یک مورد SNP در هر یک از جمعیت ها در توالی ژن های مورد نظر بودند. (جدول ۲). ۱۵۷ مورد SNP جدید و ۲۱ مورد دارای شماره rs در پایگاه های اطلاعاتی بودند. این می تواند از علت های تفاوت بیان این ژن ها باشد که موجب سازگاری دو زیرگونه مورد مطالعه با اقلیم های خاص آن ها شده و همچنین باعث بروز تفاوت های فنوتیپی بارز در این صفات بین آن ها در طول زمان گردیده است.

تعداد SNP ها در نژاد کلیستانی تقریباً ۵ برابر نژاد هلشتاین بدست آمد و علت آن احتمالاً انتخاب مصنوعی در نژاد هلشتاین است که باعث تشکیل SNP هایی درون این ژن ها شدند و این دو نژاد را در دو گروه متفاوت ژنتیکی قرار دادند. ژن های TGM3 و SDS و EGR1 در دسته ژن های دارای SNP زیاد قرار گرفت. این ژن در پاسخ ایمنی و التهابی، چرخه ی استرس و همچنین در پاسخ به مایکوفنولیک اسید، داروی سرکوب کننده دستگاه ایمنی بیان می شود. همچنین ژن TGM3 در انسان پروتئین گلوتامین

حاشیه نویسی کد زن مادی	موقعیت کروموزومی زن	تعداد SNP کشف شده در نژاد کلستانی	تعداد SNP کشف شده در نژاد هاشتاین	موقعیت نوکلئیدی SNP کلستانی	موقعیت نوکلئیدی SNP در هاشتاین
ENSBTAG00000001321	11:46410276-464118787	1	5	46411022	46410539,46413011,4 6411100,46416361,46 417272
ENSBTAG00000001785	13:53261634-53304213	18	5	53262099,5326620,53268538,532721189,53271407,53272355 ,532744316,532744385,53285734,53286509,53288998,53289 078,53291114,532911896,53292433,53298174,53303588,533 03597,	53266020,53266422,5 3272331,53286509,53 291896
ENSBTAG00000004322	10:86883738-86887169	4	0	86883777,86884139,86884739,86886415	-
ENSBTAG00000007101	3:49110155-49121034	5	0	49110693,4911095,49110770,49115665,49120714	-
ENSBTAG00000010069	7:51438726-51442500	10	2	51439166,51439362,51439906,51440879,51440955,51441122 ,51441698,51441842,51441942,51442351	51440964,51441523
ENSBTAG00000012046	7:13851130-13852933	3	0	99660430,99675955,99679525	-
ENSBTAG00000014332	5:99647122-99701630	17	2	99660430,99675955,99679525,99679741,9967841,99679944, 99687260,99687400,99695508,99695509,99695722,99695900 ,99696126,99696311,99696530,996965561,99696740,	9968021, 99695508
ENSBTAG00000016688	X:66722398-66723397	1	0	667222693	-
ENSBTAG00000020350	11:2250454-2252492	7	0	2251039,2251487,2251542,2251591,225284,2252143,225235 0	-
ENSBTAG00000021166	16:72811910-72814133	3	0	72812017,7212049,72813023	-
ENSBTAG00000031814	17:63302946-63311098	23	5	63306065,63306965,63306972,63307454,63307548,63307635 ,63307779,63308743,63308871,63309548,63309573,6330981 2,63309812,63309820,63309932,63309997,63300106,133001 53,63310399,63310480,63310582,63310620,63310771,63310 853	63303514,63306513,6 3307454,63308458,84 308427
ENSBTAG00000034824	20:2680573-3054892	2	0	2984940,2986058	-
ENSBTAG00000037778	6:90811061-90813079	9	0	90811129,90811383,90811419,90811464,90812431,90812253 4,90812724,90813024,90813059,	-
ENSBTAG00000039813	21:35135770-35137852	5	3	35136142,35136715,35136897,35137656,35137781	35137019,35137067,3 5137076
ENSBTAG00000043545	MT:5686-7299	17	1	5718,5743,5754,5783,5753,5890,5917,6115,6235,6340,6379,6 436,6460,6727,6772,6881,6922,	-
ENSBTAG00000044208	27:24815380-24828237	6	0	24815597,24816375,24817718,24820015,24827881,24828277	-
ENSBTAG00000045497	21:66764060-66764393	2	0	66764091,6674093	-
ENSBTAG00000045750	X:143364825-143375658	5	0	143364919,143365129,143375571,143736948,143737387	-
ENSBTAG00000047766	16:75539495-75540404	6	0	75539625,75539651,75539784,75539804,7554055,75540396	-
ENSBTAG00000047966	16:6155902-6156793	2	0	6156311,6156325	-
ENSBTAG00000048102	X:143736737-143800954	3	0	143736948,143737387,143768373	-
ENSBTAG00000048172	X:142664999-142690152	1	0	142686441	-
ENSBTAG00000048229	25:1394200-1436118	4	1	1394221,1394363,1417133,1417713	1394218

منابع

- Ghaderi Zefrei M., S.R Miraei Ashtiani, M.H. Banabazi, A., Nejati Javaremi, I.G. Imumorin.(2016).Single nucleotide polymorphisms (SNP) on transcriptome of Holstein cows shared with Illumina bovine SNP arrays. Online Journal of Veterinary Research (OJVR).20(3):177-182.
- Canovas, A., G. Rincon, A. Islas-Trejo, S. Wickramasinghe, and J. Medrano. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mammalian Genome 21(11-12):592 - 598.
- Djari, A., D. Esquerre, B. Weiss, F. Martins, C. Meersseman, M. Boussaha, C. Klopp, and D. Rocha. 2013. Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. BMC Genomics 14(1):307.
- 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.
- Eklblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107(1), 1-15.
- Flintoft, L. (2008). Transcriptomics: digging deep with RNA-Seq. Nature Reviews Genetics, 9(8), 568-568.
- Fries, R., Ruvinsky, A. 1999 The Genetics of Cattle. New York: CABI Publishing.
- Haas, B. J., & Zody, M. C. (2010). Advancing RNA-seq analysis. Nature biotechnology, 28(5), 421
- Hansen, P. J. 2004. Physiological and cellular adaptations of zebu cattle to thermal stress. Animal Reproduction Science, 82, 349-360.
- 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.
- بنابازی، م. ۱۳۹۵. تدوین استراتژی های محاسباتی به منظور تلفیق داده های ترانسکریپتومی (RNA-Seq) در پیش بینی های مبتنی بر ژنوم گاو شیری. رساله دکتری. دانشگاه تهران. ایران
- بنابازی، م.، م. ناصر خلیل و س. ر. میرایی آشتیانی. ۱۳۹۲. شبکه تنظیم بیان ژن های چرخه سلولی ساکارومیسیس سرویسیه به روش همبستگی وزن. سومین همایش ملی بیوتکنولوژی کشاورزی ایران. دانشگاه فردوسی مشهد.
- بنابازی، م.، م. ناصر خلیل و س. ر. میرایی آشتیانی. ۱۳۹۱. الگوریتم شناسایی ژن های متفاوت بیان شده در داده های حاصل از ریزآرایه مخمر ساکارومیسیس سرویسیه با استفاده از بسته های نرم افزار R. سومین همایش ملی بیوتکنولوژی کشاورزی ایران. دانشگاه فردوسی مشهد.
- سلیم پور، م. ۱۳۹۵. بررسی بیان افتراقی ژن بین دو جمعیت گاو هلستاین و کلیستانی (یک نژاد پاکستانی) با استفاده از توالی یابی RNA (RNA-seq). پایان نامه کارشناسی ارشد. دانشگاه تهران. ایران
- Aeschlimann D., MK. Koeller, BL. Allen-Hoffman and DF. Mosher.1998. Isolation of a cDNA encoding a novel member of the transglutaminase gene family from human keratinocytes. Detection and identification of transglutaminase gene products based on reverse transcription-polymerase chain reaction with degenerate primers. J Biol Chem. 273 (6): 3452-60.
- Andrews, S.(2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- 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.

- Langmead, B. and S. Salzberg. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357-359.
- Li, H. and R. Durbin. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754 - 1760.
- Marguerat, S., & Bähler, J. (2010). RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences*, 67(4), 569-579.
- Ohsawa I, Takamura C, Kohsaka S. Fibulin-1 binds the amino-terminal head of β -amyloid precursor protein and modulates its physiological function. *J Neurochem*. 2001; 76(5): 1411-20.
- Park, K.-D., J. Park, J. Ko, B. Kim, H.-S. Kim, K. Ahn, K.-T. Do, H. Choi, H.-M. Kim, S. Song, S. Lee, S. Jho, H.-S. Kong, Y. Yang, B.-H. Jhun, C. Kim, T.-H. Kim, S. Hwang, J. Bhak, H.-K. Lee, and B.-W. Cho. 2012. Whole transcriptome analyses of six thoroughbred horses before and after exercise using RNA-Seq. *BMC Genomics* 13(1):473.
- Pennisi, E. (2012). ENCODE project writes eulogy for junk DNA. *Genomics Science*, 337(1159), 61.
- Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10:57 - 63.
- Wang, M., IG. Kim, PM. Steinert and OW. McBride. 1995. Assignment of the human transglutaminase 2 (TGM2) and transglutaminase 3 (TGM3) genes to chromosome 20q11.2. *Genomics*. 23 (3): 721–2.
- Wilhelm, B. T., & Landry, J. R. (2009). RNA-Seq , quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, 48(3), 249-257.
- Zhang, C., G. Wang, J. Wang, Z. Ji, Z. Liu, X. Pi, and C. Chen. 2013. Characterization and Comparative Analyses of Muscle Transcriptomes in Dorper and Small-Tailed Han Sheep Using RNA-Seq Technique. *PLoS ONE* 8(8):e72686.
- Zheng P, Wang Q, Teng J, Chen J. Calumenin and fibulin-1 on tumor metastasis: Implications for pharmacology. *Pharmacol Res*. 2015; 99: 11-5.