نشریه علوم دامی
(پژوهش و سازندگی)

شماره ۱۱٤، بهـار ۱۳۹٦
ص‌ص: ۱٦۹~۱۸۲

توجیه واریانس ژنتیکی صفت باقیمانده مصرف خوراک با چندشکلی‌های تک نوکلئوتیدی (SNP) کشف شده بر روی ترانسکریپتوم گاو هلشتاین

- **محمدحسین بناءبازی**
گروه علوم دامی، دانشکده علوم و مهندسی کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران.
- **اردشیر نجاتی جوارمی** (نویسنده مسئول)
گروه علوم دامی، دانشکده علوم و مهندسی کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران
- **Ikhide G. Immumorin**
گروه علوم دامی، دانشکده کشاورزی، دانشگاه کرنل، نیویورک، آمریکا.
- **مصطفی قادری زفره ای**
گروه علوم دامی، دانشکده کشاورزی، دانشگاه یاسوج
- **سیدرضا میرایی آشتیانی**
گروه علوم دامی، دانشکده علوم و مهندسی کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران

تاریخ دریافت : اسفند ۱۳۹۴ تاریخ پذیرش : اردیبهشت ۱۳۹۵

شماره تماس نویسنده مسئول: ۰۹۱۲۱۰۳٦۴۴۸
Email: javaremi@ut.ac.ir

چکیده

چندشکلی‌های تک نوکلئوتیدی (SNP) امروزه به مهم‌ترین و کارآمدترین ابزار به نژادی تبدیل شده‌اند. مطالعات پویش کل ژنوم و انتخاب ژنومی دو کاربرد مهم در اصلاح نژاد دام هستند که هر دو مبتنی بر تعیین ژنوتیپ تعداد بسیار زیادی از این جایگاه‌ها هستند. بدین منظور آرایه‌های با تراکم بالا از این نشانگرها در صنعت گاو شیری ارائه شده‌اند. از آن‌جایی‌که ترانسکریپتوم تنها حاصل رونویسی بخش‌هایی از ژنوم است که رمزگر بوده و در نهایت به یک فرآورده پروتئینی ترجمه و بیان می شوند، احتمالا SNP های موجود در این مناطق نیز بهتر می توانند ارزش اصلاحی ژنومی را برآورد کنند. بر اساس این فرضیه، با استفاده از بسته نرم افزاری samtools، فهرستی از SNPها بر روی توالی ترانسکریپتوم نمونه‌ای از جمعیت گاو ماده هلشتاین آمریکا کشف و سهم آن‌ها در توجیه واریانس ژنتیکی صفت باقیمانده مصرف خوراک (RFI) در جمعیتی از تلیسه‌های هلشتاین استرالیایی ارزیابی شد. این ترانسکریپتوم با همردیفی و مکان یابی خوانش‌های RNA-Seq بر روی ژنوم مرجع گاو تشکیل شد. در مجموع، یک فهرست شامل ۵۳٤۷۸ نشانگر SNP بر روی توالی ترانسکریپتوم جمعیت مورد مطالعه کشف شد که بیشتر آن‌ها در طراحی آرایه‌های SNP ایلومینا در نظر گرفته نشده‌اند. فهرست یافت شده با آرایه با تراکم بالا ایلومینا تنها ٦۳۳٦ نشانگر مشترک داشت. این فهرست مشترک نتوانست سهم بیشتری از واریانس ژنتیکی صفت باقیمانده مصرف خوراک را توجیه نماید. پیشنهاد می شود جایگاه‌های کشف شده در پیش بینی‌های مبتنی بر توالی کل ژنوم به کار برده شوند تا بهتر ارزیابی گردند.

واژه‌های کلیدی: چندشکلی تک نوکلئوتیدی (SNP)، توجیه واریانس ژنتیکی، باقیمانده مصرف خوراک (RFI)، ترانسکریپتوم، گاو هلشتاین.

Animal Science Journal (Pajouhesh & Sazandegi) No 114 pp: 169-182

Genetic variance explanation of Residual Feed Intake (RFI) by SNPs discovered on transcriptome of Holstein cows

By: Mohammad Hossein Banabazi[1], Ardeshir Nejati Javaremi[1]*, Ikhide G. Imumorin[2],Mostafa Ghaderi Zefrei[3],Seyed Reza Miraei Ashtani[1]

1: Department of Animal Science, Faculty of Agricultural Science and Engineering, College of Agriculture & Natural Resources, University of Tehran, Karaj, IRAN.

2: Department of Animal Science, Cornell University, College of Agriculture and Life Sciences, Ithaca, New York, USA.

3: Department of Animal Science, University of Yasouj, Faculty of Agriculture, 74934, Yasouj, IRAN.

Received: March 2016 **Accepted: May 2016**

In recent years, Single Nucleotide Polymorphisms (SNPs) has been the most important and efficient tool in animal breeding. Genome-wide Association Studies (GWAS) and Genome-enabled predictions (genomic selection) are two major applications of SNPs in animal genetics and breeding that both relay on genotyping a lot of SNPs. Some high dense (HD) SNP arrays have been applied particularly in dairy cattle for this purpose. Since transcriptome is resulted only from transcription of coding genomic regions and finally expressed into a protein, the SNPs located on transcriptome may estimate breeding values well. Based on this hypothesis, SNP discovery was done on a transcriptome assembled from aligning and mapping of RNA-Seq reads on bovine reference genome for an US Holstein cow population by samtools package. Then, the contribution of the discovered SNPs in explanation of genetic variance for Residual Feed Intake (RFI) trait in Australian Holstein was evaluated based on variance components analysis. It was discovered 53478 SNPs that the most weren't on the current Illumina bovine SNP arrays. Only 6336 discovered SNPs were shared with Illumina bovine HD array. This reduced SNP panel was low dense and therefore could not capture genetic variance of RFI larger than Illumina bovine high dense SNP Chip. It's suggested to applying the discovered SNPs on transcritome in genomic predictions based on whole genome sequencing with no limitations for the number of shred loci with the current Illumina SNP arrays.

Key words: Single Nucleotide Polymorphisms (SNP), Genetic variance explanation, Residual Feed Intake (RFI), Transcriptome, Holstein cow

مقدمه

ترانسکریپتوم می‌تواند به درک بهتر ساز وکارهای حد واسط میان ژنوتیپ و فنوتیپ کمک نماید و صحت و دقت پیش‌بینی‌های ژنومی را که امروزه مهمترین ابزار ارزیابی و انتخاب برای به نژادگران دامی می باشد، بهبود بخشیده و پیشرفت ژنتیکی را سریع تر نماید. علاوه بر این، استفاده نکردن از حجم بالایی از داده‌های حاصل از فناوری توالی‌یابی نسل جدید و به‌ویژه داده‌های ترانسکریپتومی در اصلاح نژاد حیوانات اهلی، به معنای از دست دادن یک منبع جدید از داده‌های دقیق و با کیفیت است. نتایج پروژه دائره‌المعارف اجزای DNA[1] نشان می‌دهد که برخلاف دانسته‌های قبلی حدود ۸۰ درصد ژنوم رونویسی می گردد ولی در نهایت ، تنها حدود ۳ تا ۵ درصد از این رونویسی‌ها به صورت mRNA باقی مانده و به پروتئین ترجمه (بیان) می گردند (Pennisi، ۲۰۱۲). بخش زیادی از RNA رونویسی شده به صورت نارمزگر[2] باقی می‌ماند و احتمال می‌رود نقش تنظیمی بر سایر بخش‌های ژنوم داشته باشند (Ponting و همکاران، ۲۰۰۹؛ Wang و Chang، ۲۰۱۱). جامعیت ترانسکریپتوم به حدی است که محققین درگیر در پروژه ENCODE معتقدند می توان به جای ژن، ترانسکریپت (رونوشت)، یعنی قطعه ای از RNA که از روی DNA کد می شود را به عنوان واحد بنیادی ژنوم و واحد اساسی توارث پیشنهاد کرد. داده‌های ترانسکریپتومی به اشکال مختلف تولید می گردند که آخرین نسل از آن‌ها حاصل توالی‌یابی RNA (RNA-Seq)[3] می باشند. این داده ها دقیق‌تر و کمی‌تر از

[1] Encyclopedia of DNA Elements (ENCODE)

[2] noncoding RNA (ncRNA)

[3] RNA Sequencing (RNA-Seq)

انـواع دیگـر مـی‌باشـند (Flintoft، ۲۰۰۸؛ Wang و همکـاران، ۲۰۰۹). اساس این روش بر برش کل mRNA رونویسی‌شده (یـا به‌عبـارت دیگـر کـل ترانسـکریپتوم[۴] یـا کـل اگـزوم[۵]) بـه قطعـات کوچک‌تر[۶] و توالی‌یابی این قطعات بـا اسـتفاده از چـارچوب‌هـای فنی[۷] و پربرونـداد NGS اسـتوار اسـت (Haas و Zody، ۲۰۱۰؛ Marguerat و Bahler، ۲۰۱۰). تفاوت میزان رونویسی در بین خوانش‌های RNA-Seq، بـه صـورت تفـاوت در تعـداد دفعـات خـوانش[۸] قطعـات بـه هنگـام تـوالی‌یـابی بـه روز مـی‌نمایـد (Wilhelmو همکـاران، ۲۰۰۹). بـا بـه دسـت آمـدن یـک معیـار عددی بسیار دقیـق انـدازه‌گیـری بیـان ژن، ایـن روش را مـی تـوان آنـالیز بیـان ژن دیجیتـالی[۹] نیـز نامیـد (Mortazavi و همکـاران، ۲۰۰۸).

استفاده از اطلاعـات ژنتیکـی مبتنـی بـر نشـانگرهای مولکـولی در برنامه‌های اصلاح نژاد گاو به عنوان ابزاری برای بهبود انتخاب‌های فنوتیپی مرسوم رو به افزایش بوده است. ظهور فناوری‌هـای تـوالی یابی جدید فرصتی تازه برای شناسایی واریانت‌های مـرتبط بـا ژن-های موثر بر صفات اقتصادی پدید آورده‌اند. علیـرغم سیر نزولـی هزینه توالی یابی، توالی یابی کل ژنوم یا هـدف قـرار دادن منـاطق بزرگ هنوز برای اکثر آزمایشگاه‌ها بسیار گران اسـت. تـوالی‌یـابی مبتنی بر ترانسکریپتوم، گزینه‌ای ارزان‌تر برای شناسایی تعداد بسیار زیادی چندشکلی و امکان کشف واریانت‌های سببی ارائه می‌نماید (Djari و همکاران، ۲۰۱۳). ارزیابی[۱۰] و انتخاب ژنومی[۱۱] در دام-ها بر اساس پیش‌بینی‌هـای مبتنـی بـر ژنـوم از روی ژنوتیـپ تعـداد زیادی نشانگر (چندشکلی های تک نوکلئوتیدی[۱۲] یـا بـه اختصـار SNP) که پیش از آن میزان تاثیر هر یک با صـفت مـورد بررسـی در یک جمعیت مرجع بـرآورد شـده اسـت، انجـام مـی‌شـود. ایـن پـیش‌بینـی‌هـا بـه‌طـور سـاده مبتنـی بـر جمـع جبـری اثـرات ایـن نشانگرهاست که ارزش اصلاحی ژنومی پیش‌بینی‌شده (GEBV) بـرای هـر فـرد خواهـد بـود (Meuwissen و همکـاران، ۲۰۰۱). برهم خوردن ارتباط میـان نشـانگر و صـفت مـورد مطالعـه پـس از گذشت چنـد نسـل و نیـاز بـه بـرآورد مجـدد اثـرات نشـانگری در جمعیت مرجع، افزایش خطای محاسبات همراه بـا افـزایش تـراکم نشانگرهای موجود بـر روی تراشـه‌هـا، پیچیـدگی‌هـای محاسـباتی ناشی از عدم توازن بین تعداد مشاهدات فنوتیپی (رکوردها) اندک و تعداد بسیار زیاد ژنوتیپ نشانگری به دسـت آمـده بـه ازای یـک فرد (که حتی بالغ بر ۷۰۰ هزار جایگاه هم می‌شود)، نبـودن امکـان گسـترش اثـرات نشـانگری حاصـل از یـک جمعیـت مرجـع بـه جمعیت‌های هدفی که با آن تفاوت نژادی دارند، در نظـر نگـرفتن اثرات محیطـی مـوثر بـر تعـداد رونوشـت از هـر جایگـاه، در نظـر نگرفتن اثرات غیرافزایشی میان جایگاه‌های نشانگر مورد اسـتفاده و به ویژه اثرات اپی‌ستاتیک جایگاه‌ها و ... از جمله چالش‌های پـیش روی این نوع پیش بینی‌هاسـت (Schaeffer، ۲۰۰۶؛ Hayes و همکاران، ۲۰۰۹).

تلفیق داده‌های RNA-Seq با داده‌های ژنومی و به‌کارگیری آن-ها در ارزیابی‌ها و انتخاب‌های ژنومی رایـج ممکـن اسـت بـه رفـع برخی از محدودیت‌های گفته شده بیانجامد و زمینـه سـاز ورود بـه عصر پساژنومی در اصلاح نـژاد حیوانـات اهلـی گـردد. از ایـن‌رو، هدف از انجام مطالعه حاضـر، کشـف نشـانگرهای SNP بـر روی توالی mRNA رونویسـی شـده (ترانسـکریپتوم) و بررسـی میـزان اشتراک فهرست به دسـت آمـده بـا تراشـه هـای رایـج، و کـاربرد فهرست مشترک به‌دسـت آمـده در پـیش بینـی‌هـای ژنـومی بـرای صفت باقیمانده مصرف خوراک[۱۳] (RFI) است. این هدف با ایـن پـیش فـرض تعریـف شـده اسـت کـه احتمـالا از آن‌جـایی‌کـه ترانسکریپتوم تنها حاصل رونویسی بخش‌هـایی از ژنـوم اسـت کـه رمزگر بوده و در نهایت به یک فـرآورده پروتئینـی ترجمـه و بیـان می‌شوند، بنابراین SNP های موجـود در ایـن منـاطق نیـز احتمـالا بهتر می توانند ارزش اصـلاحی ژنـومی را بـرآورد کننـد. از سـوی دیگر از آن‌جایی که این SNP ها در مناطقی واقع شده‌انـد کـه در یک شرایط محیطی خاص رونویسی و بیان شده‌انـد احتمـالا بـرای

[4] Whole Transcriptome
[5] Whole Exome
[6] Short Reads
[7] Platform
[8] Read Counts
[9] Digital Gene Expression
[10] Evaluation
[11] Animal Selection
[12] Single Nucleotide Polymorphisms (SNP)
[13] Residual feed intake (RFI)

ارزیابی‌های ژنومی در آن محیط مناسب ترند. در نهایت، فهرستی از SNP های موجود بر روی توالی ترانسکریپتوم نمونه‌ای از جمعیت گاو ماده هلشتاین آمریکا با استفاده از داده‌های RNA-Seq ارائه شده و عملکرد آن بر روی داده های ژنومی واقعی ارزیابی می گردد.

مواد و روش‌ها

در این مطالعه از داده‌های RNA-Seq مربوط به ادغام ۴۰ نمونه از گاو ماده هلشتاین مزرعه دانشگاه ویسکانسین آمریکا استفاده شد. این داده‌ها در مطالعه Huang و همکاران (۲۰۱۲) تولید شد. به منظور تولید داده‌های مذکور، حجم برابر از نمونه‌های خون انفرادی بلافاصله پس از اخذ از سیاهرگ دمی با هم مخلوط و سپس استخراج کل RNA از مخلوط مذکور انجام شد. توالی‌یابی mRNA با دستگاه Genome Analyzer *IIx* شرکت Illumina (آمریکا) انجام شد. داده های مربوطه از لینک http://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP026487) در بخش آرشیو خوانش‌های کوتاه (SRA) بانک اطلاعاتی مرکز ملی اطلاعات بیوتکنولوژی یا NCBI دریافت شد. داده مورد آنالیز شامل ۲۱۰۷۸۴۷۷ خوانش کوتاه جفتی[14] به طول ۷۵ جفت باز بود.

تمام آنالیزها در محیط سیستم عامل لینوکس اوبونتو (Ubuntu) نسخه ۱۴.۰۴ انجام شدند. فرمت sra به خاطرحجم کم صرفا برای بارگذاری بر روی بانک‌های اطلاعاتی مناسب است و برای آنالیز ابتدا باید به فرمت‌های دیگر تبدیل گردد که یکی از مرسوم‌ترین آن‌ها فرمت fastq است. این تبدیل با استفاده از دستور fastq-dump از مجموعه نرم افزار sratoolkit نسخه ubuntu64 2.5.2 انجام شد. کنترل کیفیت داده ها با استفاده از نرم افزار FastQC v0.11.3 نسخه تحت لینوکس انجام شد. ۱۲ پارامتر از جمله محتوی آداپتورهای مورد استفاده در توالی یابی، محتوی GC در هر خوانش، تعداد نوکلئوتیدهای خوانده نشده و کیفیت خوانش هر باز ، خوانش‌های با وفور بیش از حد[15]، توزیع طول خوانش، خوانش‌های تکراری[16] و ... بر روی داده ها بررسی و کیفیت با توجه به مجموع آن‌ها سنجیده شد. پس از ویرایش داده ها نیز یک‌بار دیگر برای اطمینان از بهبود کیفیت داده ها، کنترل کیفیت انجام شد. ویرایش داده ها با استفاده از نرم افزار Trimmomatic 0.33 انجام گردید. این مرحله شامل حذف آداپتورها و حذف یا ویرایش (کوتاه نمودن) خوانش‌های بی‌کیفیت می‌باشد. فهرست آداپتورهای به کار رفته در هنگام انجام توالی یابی و تشکیل کتابخانه‌های[17] مورد نیاز و با توجه به دستگاه مورد استفاده برای توالی یابی به صورت پیش فرض در برنامه قرار داده شده است (TruSeq2-PE.fa). حداقل طول خوانش مورد قبول پس از ویرایش ۵۰ در نظر گرفته شد (طول همه خوانش ها پیش از ویرایش ۷۵ جفت باز بوده است). هم-ردیفی[18] و مکان یابی[19] خوانش های کوتاه بر روی ژنوم مرجع با استفاده از نرم افزار Tophat2 ، که از bowtie2 به عنوان همردیف کننده استفاده می‌کند، انجام شد. هم‌ردیفی و مکان یابی خوانش‌ها به تشکیل (اسمبلی) ترانسکریپتوم[20] می انجامد. در این مرحله به ژنوم مرجع و نیز اطلاعات حاشیه نویسی آن نیاز بود. آخرین نسخه به روز شده این اطلاعات از بانک اطلاعات ensembl اخذ گردید. در این مطالعه از ژنوم مرجع UMD3.1 (ویرایش ۸۱) در فرمت fasta و فایل حاشیه نویسی آن در فرمت gtf استفاده شد. کشف SNP[21] بر روی ترانسکریپتوم تشکیل شده و فیلتر نمودن آن‌ها با استفاده از نرم افزارهای samtools 0.1.19 و Bedtools انجام شد. همچنین تعداد SNPهایی از آخرین به روزرسانی اطلاعات مربوط به واریانت های ژنوم رفرنس گاو (فایل vcf) که در توالی ترانسکریپتوم تشکیل شده حضور داشتند نیز محاسبه گردید. سپس به منظور بررسی میزان اشتراک بین فهرست SNP های یافت شده و SNP های موجود در تراشه‌های مرسوم طراحی شده توسط شرکت ایلومینا (آمریکا) شامل فهرست‌های با تراکم های متفاوت شامل Illumina Bovine 3K BeadChip، 50K-v2 و HD (به‌ترتیب شامل ۲۹۰۰، ۵۴۶۰۹ و ۷۷۷۹۶۲ نشانگر در سرتاسر ژنوم)، اطلاعات تراشه‌های مذکور از بانک اطلاعاتی v.3

[14] Paired
[15] Overrepresented Sequences
[16] Duplicate Sequences
[17] Library
[18] Alignment
[19] Mapping
[20] Transcriptome Assembly
[21] SNP Discovery

SNPchiMp (Nicolazzi و همکاران، ۲۰۱۵)، که توسط بانک اطلاعات ensembl پشتیبانی می گردد (http://bioinformatics.tecnoparco.org/SNPchimp)، دریافت گردید. علاوه بر این، میزان اشتراک با آخرین فهرست SNP های حاشیه نویسی شده که در پروژه ۱۰۰۰ ژنوم گاو نر منتشر شده (http://www.1000bullgenomes.com/doco/all_snps_annotated_2013_04_25.tab.gz) نیز مقایسه گردید. تا ژانویه ۲۰۱۳، اطلاعات ۲۳۸ گاو نر پردازش و منتشر گردیده است که شامل حدود ۲۸ میلیون InDel و SNP می باشد. در مجموع، تاکنون تعداد۲۶۷۱۶۸۹۶ نشانگر SNP حاشیه نویسی شده در این پروژه گزارش شده است. از آن‌جایی‌که تمام SNP های یافت شده در فهرست SNP های تراشه‌های مرسوم برای پیش بینی‌های مبتنی بر ژنوم حضور ندارند، تنها SNP های مشترک و از بین آن‌ها تنها مواردی که پس از ارزیابی کیفیت ژنوتیپ ها و اعمال برخی شاخص‌های دیگر برای فیلتر نمودن در لیست نهایی باقی می‌مانند را می توان بر روی داده‌های ژنومی واقعی استفاده نمود و عملکرد آن‌ها را با تراشه‌های اصلی مقایسه نمود. برای این منظور، رکوردهای باقیمانده مصرف خوراک (RFI) برای ۸۴۲ تلیسه هلشتاین از مرکز تحقیقاتی Ellinbank دانشگاه ملبورن استرالیا به دست آمد. جزییات مربوط به محاسبه باقیمانده خوراک مصرفی و مشخصات حیوانات رکوردبرداری شده در مطالعه Pryce و همکاران (۲۰۱۲) آمده است. تمامی تلیسه های مذکور با تراشه Illumina HD Bovine ۷۷۷K SNP chip (شرکت Illumina، سان دیه گو، کالیفرنیا، آمریکا) تعیین ژنوتیپ شده اند. پس از کنترل کیفیت که عمدتا بر اساس فراوانی آلل مایتور (MAF)[22] و سطح احتمال انحراف از تعادل هاردی- واینبرگ انجام شد، تعداد SNP۶۰۶۰۹۶ از تراشه مذکور باقی ماند. ماتریس روابط ژنومی(GRM)[23] با استفاده از این تعداد SNP و نیز فهرست SNP های مشترک تشکیل شد(Khansefid و همکاران، ۲۰۱۴). همچنین ماتریس روابط شجره‌ای (NRM)[24] براساس ساختار شجره گفته شده محاسبه شد. سپس از ماتریس‌های GRM و NRM برای برآورد وراثت پذیری صفت RFI (h^2) و انحراف معیار آن (SE)، سهم هر یک از اجزای واریانس ژنتیکی کل (σ^2_g) شامل مولفه پلی ژنیک ($\sigma^2_{gpolygenic}$)، کل SNP ها ($\sigma^2_{gAll_SNP}$)، SNP های مشترک ($\sigma^2_{gCommon_SNP}$) و واریانس باقیمانده (σ^2_e) در قالب سه مدل (۱) مبتنی بر شجره (NRM)، (۲) مبتنی بر ماتریس روابط ژنومی (GRM) و (۳) مبتنی بر شجره بعلاوه ماتریس روابط ژنومی (GRM+NRM) به‌طور جداگانه برای فهرست های شامل تمام SNPها، SNPهای مشترک و مجموع کل SNP ها بعلاوه SNP های مشترک استفاده شد و برآوردهای به دست آمده با هم مقایسه گردید. ماتریس NRM به منظور کاهش اریب پیش بینی-ها برآورد و در مدل گنجانده شد. مقایسه بین مدل‌ها، بر اساس لگاریتم درست نمایی (LogL)[25] برآورد شده برای هر مدل انجام شد. به عبارت دیگر هر مدلی که LogL بزرگ‌تری داشته باشد بهتر عمل کرده است (Khansefid و همکاران، ۲۰۱۴).

نتایج و بحث

در مجموع تعداد جایگاه ۵۳۴۷۸ SNP بر روی ترانسکریپتوم گاو هلشتاین آمریکا کشف شد. از مجموع کل SNP های موجود در بانک اطلاعات مر بوط به SNP های گاو که شامل ۷۶۵۲۵۷۷۶ میلیون عدد می باشد (بر طبق آخرین به روزرسانی فایل vcf ژنوم رفرنس گاو)، تعداد ۳۵۷۴۳۲۱ عدد در ترانسکریپتوم گاو هلشتاین حضور داشتند. نسبت این‌دو به هم معادل ۴/۷ درصد بود که با برآورد ۳ تا ۵ درصد میزان رونویسی از کل ژنوم مطابق دارد. همچنین، ۴۲۰۸۲ نشانگر SNP موجود در این فهرست کشف شده، در پروژه ۱۰۰۰ ژنوم گاو نر نیز حضور دارند. البته نتایج منتشر شده مربوط به این پروژه تاکنون نتایج توالی یابی کامل ژنوم ۲۳۴ گاو نر را شامل می شود و با اضافه شدن اطلاعات مربوط به سایر گاوهای نر ممکن است تعداد بیشتری SNP مشترک یافت گردد. مشابهت حدود ۸۰ درصدی SNP های کشف شده مطالعه حاضر و پروژه ۱۰۰۰ ژنوم گاو نر، صحت و دقت روند کشف SNP بر روی توالی ترانسکریپتوم با استفاده از داده‌های RNA-Seq را تایید می کند. Canovas و همکاران (۲۰۱۰) با استفاده

[22] Minor Allele Frequency (MAF)
[23] Genomic Relationship Matrix (GRM)
[24] Pedigree Relationship Matrix (NRM)
[25] Log of Likelihood (LogL)

از RNA-Seq بیش از ۳۳۰۰۰ جایگاه SNP را در ترانسکریپتوم شیر گاو و نواحی کدکننده ژن‌های بیان شده در طی شیردهی معرفی نمودندکه به ادعای این محققین می‌تواند برای ژنوتیپ یابی و انجام مطالعات مربوط به بررسی ارتباط نشانگر- صفت درگاو هلشتاین به کار رود. آن‌ها نتایج این مطالعه را تاییدی بر این مطلب دانستند که آنالیز ترانسکریپتوم با استفاده از فناوری RNA-Seq روشی کارآمد و به صرفه برای شناسایی SNP ها در نواحی نسخه برداری شده است. این مطالعه راه‌کارهایی برای حداکثر نمودن کشف SNP و پرهیز از تشخیص اشتباه فراهم آورده است.

آنالیز کشف SNP مبتنی بر ژن در *Longissimus thoraci* گاوی نیز با استفاده از RNA-Seq انجام شد. mRNA حاصل از *Longissimus thoraci* سه گوساله نر لیموزین مورد توالی یابی پربروندداد قرار گرفت. جمعا ۱۹۷۵۲ رونوشت شناسایی و ۳۴۳۷۶ SNP تشخیص داده شدند. ۵۵ درصد SNPها در نواحی رمزگر یافت شدند و حدود ۲۲ درصد منجر به یک تغییر اسیدآمینه‌ای شدند. با لحاظ کردن یک آستانه کیفیت بسیار سختگیرانه، تعداد ۸۴۰۷ نشانگر SNP با دامنه اطمینان بالا، که ۱۸ درصد از آن‌ها SNP رمزگر نامترادف بودند، شناسایی شدند. در مجموع، داده‌های RNA-Seq و مجموعه SNPهای رمزگر تازه کشف شده، منابع ژنومی در دسترس برای گاو و به ویژه گاو گوشتی را بهبود می بخشند. تنوع بسیار زیادی در ژن‌های بیان شده در *Longissimus thoracis* لیموزین، در قالب تعداد بسیار زیادی SNP رمزگر نامترادف، وجود دارند که ممکن است برای مطالعه سازوکارهای مسئول تغییرپذیری ژنتیکی صفات مربوط به کیفیت گوشت مفید باشند (Djari و همکاران، ۲۰۱۳). Sharma و همکاران (۲۰۱۲) به منظور یافتن تعداد زیادی SNP مرتبط با صفت، ترانسکریپتوم دو بافت کبد و کلیه را در ۵ نژاد بز توالی یابی نمودند. آن‌ها نوع و ماهیت ۶۸۵۹۷ و ۷۲۰۴۷ جایگاه SNP یافت شده به ترتیب در کبد و کلیه را بررسی کرده و پیشنهاد نمودند SNP های یافت شده در این مطالعه به عنوان منبعی برای مطالعات GWAS در بز و نیز جهت طراحی آرایه-های SNP با تراکم بالا برای این گونه مورد استفاده قرار گیرند.

Park و همکاران (۲۰۱۲) بر روی توالی ترانسکریپتوم حاصل از دو بافت خون و عضله ۶ اسب تاروبرد پیش و پس از مسابقه، تعداد ۱۸۳۹۷۳ نشانگر SNP پیدا نمودند که تقریبا ۹۰ درصد آن‌ها، در مقایسه با ۱.۱ میلیون SNP گزارش شده قبلی در دو بانک اطلاعاتی برای اسب، جدید بودند. Zhang و همکاران (۲۰۱۳)، تعداد ۴۰۴۸۱ و ۳۸۸۵۱ جایگاه SNP را با استفاده از داده‌های RNA-Seq در دو جمعیت گوسفند یافت نمودند که در مجموع تعداد ۵۹۱۳۹ جایگاه در دو جمعیت متفاوت بودند. تفاوت نتایج مطالعه حاضر با مطالعات مذکور، احتمالا ناشی از تفاوت گونه مورد مطالعه، نوع بافت مورد آنالیز و سایر فاکتورهای در نظر گرفته شده در یک آزمایش RNA-Seq از جمله تعداد نمونه ادغام شده است.

توزیع SNP های کشف شده در کل ژنوم و پراکنش آن‌ها بر روی SNP های موجود در یک آرایه با تراکم بالا (جدول ۱) و مقایسه نسبت تعداد SNP به طول کروموزوم های مختلف نشان می‌دهد ارتباط مستقیمی بین تعداد SNP های کشف شده و طول کروموزوم ها وجود ندارد. برای مثال کروموزوم شماره ۲۵ که کوتاه‌ترین کروموزوم است به طول ۴۲۹۰۴۱۷۰ جفت باز تعداد قابل توجهی SNP یعنی ۱۸۹۴ نشانگر دارد. این در حالی است که کروموزوم های شماره ۱ و X که به‌ترتیب با طول ۱۵۸۳۳۷۰۶۷ و ۱۴۸۸۲۳۸۹۹ جفت باز طولانی ترین کروموزوم ها در ژنوم گاو می باشند و تقریبا سه برابر کرموزوم ۲۵ طول دارند، به‌ترتیب ۲۱۹۲ و ۹۰۲ SNP را شامل می‌شوند. توالی‌های غیرکروموزومی (توالی‌هایی که هنوز به هیچ کروموزوم مشخصی منتسب نشده‌اند) و توالی ژنوم میتوکندریایی به‌ترتیب با طول ۹۱۲۱۷۵۷ و ۱۶۳۳۸ جفت باز نیز به‌ترتیب ۱۵۵ و ۱۴ SNP داشتند. نتایج فوق نشان می دهند که در مجموع رونویسی از تمام طول ژنوم با یک توزیع همگن و با پوششی یکسان صورت نمی‌گیرد. به عبارت دیگر، احتمالا برخی نواحی حامل ژن‌های کاندیدای بیشتر ویا ژن‌هایی هستند که برای جمعیت فوق اهمیت بیشتری داشته و رونویسی (بیان) در آن مناطق شدیدتر و با عمق بیشتری صورت گرفته است و در نتیجه این نواحی سهم بیشتری از

کل ترانسکریپتوم تشکیل شده را به خود اختصاص می دهند و در نتیجهSNP ها در این نواحی فراوانی بالاتر داشته و بعد از گذراندن فیلترهای مختلف همچنان در فهرست نهایی باقی مانده-اند. بررسی دقیق‌تر نواحی رونویسی شده و پوشش[۲۶] ترانسکریپتومی این نواحی احتمالا درک ما را از رونویسی (بیان) ژن‌ها و تنظیم آن در جمعیت مورد مطالعه افزایش خواهد داد. این تعداد بسیار کم عملا امکان استفاده و آزمون SNP های مشترک در پیش‌بینی‌های مبتنی بر ژنوم و آزمون و مقایسه عملکرد آن‌ها را با تراشه های اصلی برای دو فهرست 3K و 50K-v2 نمی‌دهد. لذا تنها فهرست مشترک با تراشه بسیار متراکم بر روی داده‌های ژنومی واقعی استفاده و ارزیابی گردید. پس از کنترل کیفیت، از مجموع ۷۱۷۱ نشانگر کشف شده مشترک با آرایه با تراکم بالا، ۶۳۳۶ نشانگر SNP باقی ماندند (شکل ۱).

بررسی میزان اشتراک فهرست کشف شده با تراشه‌های رایج شرکت ایلومینا با تراکم های متفاوت شامل 3K، 50K-v2 و Illumina Bovine HD (به‌ترتیب شامل ۲۹۰۰، ۵۴۶۰۹ و ۷۸۶۷۹۹ نشانگر SNP) نشان داد که به‌ترتیب ۵۰، ۶۱۹ و ۹۶۷۷ جایگاه SNP مشترک بین آن‌ها وجود دارد (جدول ۲ و دو ستون آخر جدول ۱ به تفکیک هر کروموزوم). مقایسه مدل‌های مختلف برازش شده برای برآورد مولفه‌های واریانس و تعیین سهم فهرست های کامل (شامل ۶۰۶۰۹۶ نشانگر SNP) و فهرست مشترک کاهش یافته (شامل ۶۳۳۶ نشانگر SNP) در توجیه کل واریانس ژنتیکی صفت باقیمانده مصرف خوراک نیز در جدول ۳ خلاصه شده است.

[26] Coverage

جدول ۱ - تعداد SNP های در آرایه با تراکم بالا (HD) ، فهرست کشف شده بر روی ترانسکریپتوم گاو هلشتاین و فهرست مشترک با آرایه HD پیش و پس از کنترل کیفیت (QC) به تفکیک هر کروموزوم

کروموزوم	طول کروموزوم (bp)	آرایه HD	آرایه HD پس از QC	کشف شده	مشترک پیش از QC	مشترک پس از QC
۱	۱۵۸۳۳۷۰۶۷	۴۶۶۰۰	۳۷۵۸۰	۲۱۹۱	۳۰۵	۲۷۳
۱۰	۱۰۴۳۰۵۰۱۶	۳۰۵۸۰	۲۵۰۷۸	۲۲۳۴	۳۲۷	۲۹۱
۱۱	۱۰۷۳۱۰۷۶۳	۳۲۱۶۶	۲۶۸۱۷	۲۶۵۲	۳۹۰	۳۶۰
۱۲	۹۱۱۶۳۱۲۵	۲۶۱۷۷	۲۱۰۷۸	۱۲۷۵	۱۷۰	۱۵۷
۱۳	۸۴۲۴۰۳۵۰	۲۳۷۲۱	۱۶۷۴۵	۲۱۷۱	۲۶۸	۲۳۷
۱۴	۸۴۶۴۸۳۹۰	۲۴۸۴۲	۱۷۵۹۳	۱۰۱۹	۱۵۱	۱۳۸
۱۵	۸۵۲۹۶۶۷۶	۲۴۸۴۳	۲۰۰۹۱	۱۶۸۹	۲۰۹	۱۷۷
۱۶	۸۱۷۲۴۶۸۷	۲۴۲۵۴	۱۹۵۴۲	۱۵۹۳	۲۵۵	۲۲۶
۱۷	۷۵۱۵۸۵۹۶	۲۲۵۷۷	۱۸۷۴۳	۱۶۵۷	۲۳۲	۱۹۶
۱۸	۶۶۰۰۴۰۲۳	۱۹۴۸۹	۱۶۳۷۲	۲۸۶۶	۳۳۲	۲۹۰
۱۹	۶۴۰۵۷۴۵۷	۱۹۰۱۶	۱۶۰۷۳	۳۰۸۱	۴۱۹	۳۸۳
۲	۱۳۷۰۶۰۴۲۴	۴۰۱۷۲	۳۱۶۱۳	۲۴۴۰	۳۳۱	۲۸۹
۲۰	۷۲۰۴۲۶۵۵	۲۱۵۶۲	۱۸۲۶۸	۸۲۳	۱۱۹	۱۰۸
۲۱	۷۱۵۹۹۰۹۶	۲۱۲۴۱	۱۷۱۰۶	۱۳۹۲	۱۹۳	۱۷۱
۲۲	۶۱۴۳۵۸۷۴	۱۸۱۱۶	۱۵۶۷۱	۱۳۳۱	۲۳۰	۲۰۷
۲۳	۵۲۵۳۰۰۶۲	۱۵۲۹۷	۱۲۷۲۵	۲۷۷۹	۲۷۶	۲۳۴
۲۴	۶۲۷۱۴۹۳۰	۱۸۷۰۳	۱۵۰۹۵	۸۰۸	۱۱۹	۱۰۳
۲۵	۴۲۹۰۴۱۷۰	۱۲۹۷۷	۱۱۰۲۹	۱۸۹۴	۲۴۲	۲۱۹
۲۶	۵۱۶۸۱۴۶۴	۱۵۳۲۱	۱۳۰۵۵	۹۲۸	۱۱۹	۹۶
۲۷	۴۵۴۰۷۹۰۲	۱۳۲۰۰	۱۱۴۰۹	۵۷۰	۸۶	۷۷
۲۸	۴۶۳۱۲۵۴۶	۱۳۱۰۸	۱۱۰۴۹	۷۶۷	۱۲۵	۱۱۲
۲۹	۵۱۵۰۵۲۲۴	۱۴۷۶۲	۱۲۱۸۶	۱۲۱۷	۱۵۴	۱۳۳
۳	۱۲۱۴۳۰۴۰۵	۳۵۷۳۳	۲۸۸۵۱	۲۹۲۹	۴۱۹	۳۶۲
۴	۱۲۰۸۲۹۶۹۹	۳۵۱۰۵	۲۸۲۶۷	۲۰۰۹	۲۲۱	۱۹۸
۵	۱۲۱۱۹۱۴۲۴	۳۴۹۵۲	۲۷۵۵۹	۳۱۲۰	۴۲۱	۳۷۹
۶	۱۱۹۴۵۸۷۳۶	۳۵۵۸۶	۲۸۸۸۳	۱۶۴۸	۲۳۸	۲۲۰
۷	۱۱۲۶۳۸۶۵۹	۳۳۲۹۰	۲۶۲۸۹	۲۵۷۸	۳۴۴	۳۰۵
۸	۱۱۳۳۸۴۸۳۶	۳۳۶۴۹	۲۲۷۲۴	۱۶۱۰	۲۱۸	۱۸۹
۹	۱۰۵۷۰۸۲۵۰	۳۱۱۱۵	۲۵۱۱۲	۱۱۳۵	۱۶۴	۱۵۶
ژنوم میتوکندریایی	۱۶۳۳۸	۳۴۲	۰	۱۴	۴	۰
X	۱۴۸۸۲۳۸۹۹	۳۹۴۵۶	۱۳۴۹۳	۹۰۲	۹۰	۵۷
توالی های غیر کروموزومی	۹۴۹۹۵۵۶	۰	۰	۱۵۶	۰	۰
مجموع	۲۶۷۰۴۲۲۲۹۹	۷۷۷۹۵۲	۶۰۶۰۹۶	۵۳۴۷۸	۷۱۷۱	۶۳۴۳

جدول ۲ - اشتراک نشانگرهای SNP از منابع مختلف

کشف شده	3K	50K - v2	HD	پروژه ۱۰۰۰ ژنوم گاو نر
کشف شده (53,478)	–			
3K* (2,900)	۵۰	–		
50K-v2**(54,609*)	۶۱۹	۲۸۴۶	–	
HD***(777,962)	۷۱۷۱	۲۸۲۲	۵۲۴۳۸	–
پروژه ۱۰۰۰ ژنوم گاو نر(26,761,896)	۴۲۰۸۲	۲۶۹۰	۴۶۹۵۴	۶۹۵۳۰۶	–

اعداد داخل پرانتز تعداد SNP موجود در هر منبع را نشان می دهد. * آرایه با تراکم پایین (Illumina Bovine3k BeadChip)، ** نسخه دوم آرایه 50K (Illumina BovineSNP50v2 BeadChip)، *** آرایه با تراکم بالا (Illumina BovineHD BeadChip)

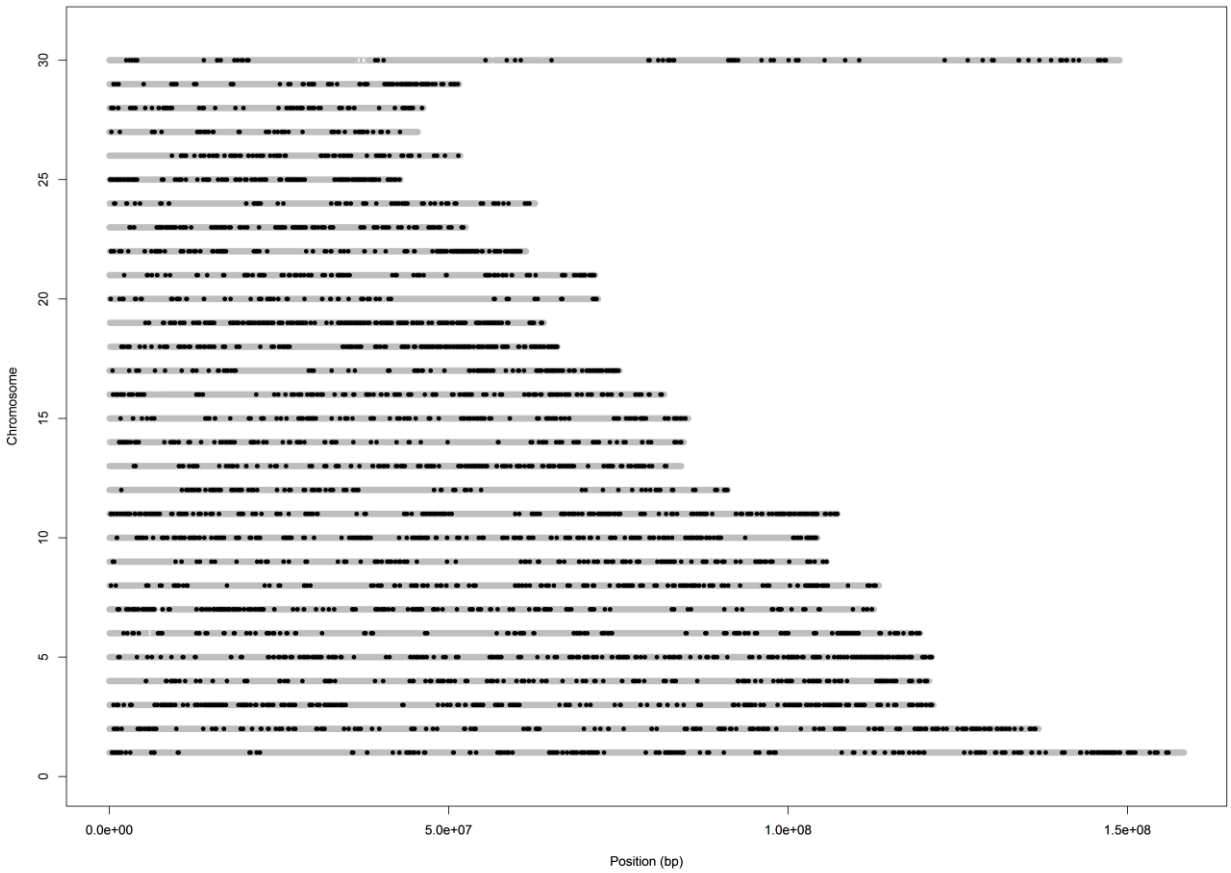
همچنان‌که از جدول مذکور برمی‌آید، فهرست مشترک نتوانست سهمی معادل فهرست کامل یا حتی بیشتر از آن را از کل واریانس ژنتیکی این صفت به خود اختصاص داده و توجیه نماید. بهترین عملکرد (بالاترین LogL) مربوط به مدلی است که همه مولفه ها از جمله ماتریس روابط ژنومی با استفاده از کل SNP ها، فهرست مشترک و ماتریس روابط شجره ای را در بر دارد (۴۴۱/۱۴۹-). کل این اطلاعات حدود ۳۴ درصد از واریانس ژنتیکی را برای صفت مورد مطالعه در جمعیت مورد آزمون توجیه نموده است در حالی‌که بیشترین سهم از واریانس ژنتیکی (۴۰ درصد) را مدلی که تنها شامل مولفه کل SNP ها می باشد توجیه نموده و سهم مدلی که تنها دربرگیرنده فهرست مشترک است معادل ۲۸ درصد است. وراثت پذیری صفت باقیمانده مصرف خوراک با مدل‌های در برگیرنده کل SNP ها و SNP های مشترک به‌ترتیب ۰/۳۹۷۳±۰/۰۹۲۰ و ۰/۲۸۴۸±۰/۰۷۳۸ برآورد گردید. همان-طورکه از شکل ۱ نیز مشخص است تعداد حدود ۶۰۰۰ جایگاه، پوشش کامل و یکنواختی در سرتاسر ژنوم فراهم ننموده است. از سوی دیگر، امکان استفاده و ارزیابی کل فهرست ۵۰ هزارتایی یافت شده بر روی ترانسکریپتوم نیز وجود ندارد چرا که اکثر این جایگاه در فهرست‌های کامل مرسوم (50K-v2 و HD) وجود ندارند. به عبارت دیگر، آن‌چه یافت شده و انتظار عملکرد بهتر از آن می رفت، قابل ارزیابی نبود.

این امر نشان می‌دهد که بازنگری در فهرست آرایه‌های مرسوم به کار رفته برای پیش بینی‌های مبتنی بر ژنوم در گاو و اضافه نمودن تعداد بیشتری از SNPهای رمزگر به آن‌ها، احتمالا می تواند به بهبود چنین پیش بینی‌هایی بیانجامد. این SNPهای رمزگر ممکن است آن دسته از نشانگرهای SNP را شامل شود که در تعداد زیادی ترانسکریپتوم تشکیل شده واقعی مشترک بوده و/یا در یک ترانسکریپتوم واقعی تشکیل شده از داده RNA-Seq تعداد زیادی نمونه ادغام شده کشف و استخراج شده باشند. فهرست ۵۰ هزارتایی حاصل از این مطالعه را تنها می توان بر روی داده‌های توالی‌یابی کل ژنوم و در صورتی که رکوردهای فنوتیپی برای افراد توالی یابی شده وجود داشته باشد (به عنوان مثال، اطلاعات حاصل از پروژه توالی یابی ژنوم کامل ۱۰۰۰ گاو نر) ارزیابی نمود. از آنجایی‌که در پیش‌بینی‌های مبتنی بر توالی کل ژنوم نیز همچنان نشانگرهای SNP به کار خواهند رفت، می توان در صورت عملکرد بهتر فهرست‌های کشف شده بر روی ترانسکریپتوم، آن‌ها را برای پیش بینی‌های مبتنی بر توالی کامل ژنوم پیشنهاد نمود. به هر حال، به‌تدریج پیچیدگی های فنی و محاسباتی و نیز مسائل هزینه‌ای این نوع پیش بینی‌ها کمتر شده و راه‌کار نوینی را پیش روی به‌نژادگران دامی قرار گرفته است (Druet و همکاران، ۲۰۱۴).

جدول ۳ - برآورد وراثت پذیری صفت RFI (h^2) ، انحراف معیار آن (SE) و سهم هر یک از مولفه های واریانس ژنتیکی کل (σ^2_g) برای سه مدل مختلف

مدل	$LogL$	$\sigma^2_{gAll_SNP}$	$\sigma^2_{gCommon_SNP}$	$\sigma^2_{gpolygenic}$	σ^2_e	h^2	SE
NRM	-٤٥١.٢٦٦	-	-	٠.٢٤٦٩	٠.٩٩٨٩	٠.٧٥٢	٠.٢٤٧٢
GRM_allSNPs	-٤٤١.٣٢٩	٠.٤	-	-	١.٠٠٦٨	٠.٦٠٦٨	٠.٣٩٧٣
GRM_allSNPs + NRM	-٤٤٦.٤٤٦	٠.٣٩٦٢	-	٠.٠١٧٩	٠.٩٨٧٤	٠.٧٤٥٣	٠.٢٤٥٢
GRM_6336SNPs	-٤٤٤.١٠٨	-	٠.٢٨٠٩	-	٠.٩٨٧١	٠.٧٠٦٢	٠.٢٨٤٦
GRM_6336SNPs + NRM	-٤٤٤.٨٠٣	-	٠.٢٧٣٥	٠.٠٢٧٧	٠.٩٨٨٤	٠.٦٨٧١	٠.٣٠٤٨
GRM_allSNPs + GRM_6336SNPs	-٤٤١.١٥٠	٠.٣٣٦٨	٠.٠٦٦	-	١.٠٠٤٦	٠.٦٠١٨	٠.٤٠١
GRM_allSNPs + GRM_6336SNPs + NRM	-٤٤٤.١٤٩	٠.٣٣٦٥	٠.٠٦٤٩	٠.٠٠٦٨	١.٠٠٥١	٠.٥٩٦٩	٠.٤٠٦١

مولفه های واریانس شامل اجزاء پلی ژنیک ($\sigma^2_{gpolygenic}$)، کل SNP ها ($\sigma^2_{gAll_SNP}$)، SNP های مشترک ($\sigma^2_{gCommon_SNP}$) و واریانس باقیمانده (σ^2_e) است. مدل‌های مورد بررسی شامل (۱) استفاده از اطلاعات شجره‌ای (NRM)، (۲) برآورد ماتریس روابط ژنومی با استفاده از کل SNP ها و SNPهای مشترک (GRM_allSNPs و GRM_6336SNPs) و (۳) ترکیب برآورد ماتریس روابط ژنومی و اطلاعات شجره ای با استفاده از کل SNP ها، SNPهای مشترک و ترکیب دو فهرست باهم (GRM_allSNPs + NRM ، GRM_6336SNPs + NRM و GRM_allSNPs + GRM_6336SNPs + NRM) می



شکل ۱ – پراکنش موقعیت SNP های مشترک (نقاط تیره) بر روی کل SNP های موجود در آرایه با تراکم بالا (نواحی روشن). محور عمودی شماره کروموزوم و محور افقی طول کروموزوم ها را نشان می‌دهد. کروموزوم شماره ۳۰ همان کروموزوم X است.

پیشنهاد می‌شود که به‌جای کشف SNP بر روی توالی ترانسکریپتوم و استفاده و ارزیابی SNP های مشترک با آرایه های مرسوم، تنها حضور موقعیت نوکلئوتیدی یک SNP از آرایه‌های مرسوم بر روی توالی ترانسکریپتوم های واقعی ملاک انتخاب آن برای قرار گرفتن در فهرست جایگاه‌های یک پانل SNP کاهش یافته[۲۷] باشد. برای تمایز بین این استراتژی پیشنهادی و استراتژی کشف SNP، این استراتژی جدید را می توان فراخوانی SNP[۲۸] نامید. بدیهی است SNPهای فراخوانده شده الزاما بر روی توالی ترانسکریپتوم های واقعی تشکیل شده چندشکل (SNP واقعی) نیستند بلکه تنها به این دلیل که موقعیت ژنومی آن‌ها در نواحی رونویسی شده قرار داشته و از این‌رو متناظر با یک موقعیت نوکلئوتیدی در توالی ترانسکریپتوم های تشکیل شده نیز هست، در پانل کاهش یافته قرار گرفته‌اند. در نتیجه اجرای این استراتژی، احتمالا تعداد بیشتری SNP و با پراکنش همگن‌تر در سرتاسر ژنوم خواهیم داشت که ممکن است عملکرد بهتری نیز داشته باشند. Koufariotis و همکاران (۲۰۱۴)، واریانت‌های موجود در آرایه متراکم Illumina Bovine HD را در قالب ۱۲ گروه دسته بندی نمودند. نتایج آن‌ها نشان داد همان‌طور که انتظار می رفت، واریانت‌های بین ژنی (که بر روی DNA نارمزگر واقع هستند) معمول‌ترین SNP ها بودند (حدود ۶۷ درصد) و درون ژنی‌ها ۳۲ درصد از کل را شامل می‌شدند. البته به نسبت سهم ۳ تا ۵ درصدی از کل ژنوم که به‌صورت mRNA رونویسی می‌شود، سهم حدود ۳۰ درصدی از کل SNP های یک آرایه با تراکم بالا قابل ملاحظه است. اما برخی مطالعات نشان داده‌اند که تمامی SNP هایی از یک آرایه با تراکم بالا که درمناطق کد کننده قرار

[27] Reduced SNP Panel
[28] SNP Calling

دارند (حداکثر حدود ۳۰ درصد از کل آرایه) نیز نمی توانند در مقایسه با کل SNP های موجود در این آرایه‌ها سهم بالاتری از توجیه واریانس ژنتیکی صفات مورد مطالعه را در گونه‌های دامی مختلف به خود اختصاص دهند ویا به پیش بینی‌های مبتنی بر ژنوم صحیح تری بیانجامند.

Koufariotis و همکاران (۲۰۱۴) برای یافتن واریانت‌های مرتبط با ۱۰ صفت در گاو گوشتی و ۱۱ صفت در گاو شیری از میان ۱۲ دسته گفته شده از دو روش افزایش یا افت ارتباطات معنی‌دار و روش برآورد مولفه‌های واریانس استفاده نمودند. افزایش معنی‌دار واریانت‌های مرتبط با صفت (SNPهای معنی دار در GWAS) در دسته واریانت‌های عامل (واریانت‌های نامترادف)[29] و نواحی ۵ کیلوجفت بازی بالا دست و پایین دست ژن‌های کد کننده یافت شد. این نتایج نشان دادند که حاشیه نویسی عملکردی می تواند در تعیین اولویت واریانت‌ها برای قرارگرفتن در یک زیرمجموعه، که با احتمال زیادتری با صفات پیچیده مرتبط خواهد بود، کمک کند. همچنین، روش برآورد مولفه های واریانس که در مطالعه حاضر نیز به کار گرفته شده است روش مناسبی برای ارزیابی زیرمجموعه‌های به دست آمده از آرایه اصلی با تراکم بالا می باشد. Morota و همکاران (۲۰۱۴) نیز SNP های موجود در یک تراشه با تراکم بالا (600K) و رایج برای پیش بینی‌های مبتنی بر ژنوم جوجه‌های گوشتی را بر اساس اطلاعات حاشیه نویسی آن‌ها به دو دسته ژنی و غیر ژنی دسته بندی نموده و توان پیش بینی آن‌ها را برای سه صفت بررسی نمودند. نتایج با توجه به نوع صفت متفاوت بود. به‌طوری‌که عملکرد پیش بینی برای صفت وزن بدن و ناحیه اولتراسونیک گوشت سینه با استفاده از نواحی ژنی بهتر از نواحی غیر ژنی بود ولی در صفت تولید تخم مرغ، این روند برعکس بود. آن‌ها قابلیت پیش بینی توسط کل نشانگرها در سرتاسر ژنوم را نزدیک به بهترین پیش بینی حاصل از یک ناحیه ژنومی ذکر کردند و همچنان روش استفاده از کل SNPها را برای پیش بینی صفات پیچیده مناسب تر دانستند. Do و همکاران (۲۰۱۵) نیز SNP ها را برای صفت باقیمانده مصرف خوراک و اجزای آن حاشیه نویسی نمودند. جمعا ۱۲۷۲ خوک Duroc که هم رکوردهای فنوتیپی و هم ژنوتیپی داشتند را به دو جمعیت مرجع (۹۶۸ خوک) و جمعیت تایید (۳۰۴ خوک) تقسیم نمودند. سپس SNP ها با استفاده از اطلاعات حاشیه نویسی اخذ شده از بانک اطلاعات Ensembl به ۱۴ گروه مختلف تقسیم و صحت و اریب پیش بینی هر دسته محاسبه شد. عملکرد پیش بینی کلاس‌های SNP تفاوت معنی-داری نسبت به یک گروه تصادفی از SNP نشان نداد. در مجموع، حاشیه نویسی ژنومی تاثیر کمی بر صحت پیش بینی صفات پیچیده‌ای که در این تحقیق مطالعه شدند داشت؛ اما ممکن است در مورد صفات دیگر این امر متفاوت باشد.

در مقابل Erbe و همکاران (۲۰۱۲)، نتایج موفقیت آمیزی با یک پانل جایگزین شامل ۵۸۵۳۲ نشانگر به دست آوردند. این پانل کاهش یافته بعد از فیلتر کردن تراشه ایلومینا با تراکم بالا و از میان ۶۲۴۲۱۳ جایگاه SNP کنترل کیفیت شده آن استخراج شده و مبنای انتخاب آن‌ها قرار داشتن در بخش رونویسی شده ژنوم گاو بود. این پانل به‌ازای تمام صفات حدود ۰/۰۳ افزایش در صحت را برای گاو جرسی در پی داشت. ولی از سوی دیگر، مزیت کاهش تعداد قابل ملاحظه SNP ها را نیز داراست. در صورت استفاده از این پانل همراه با یک جمعیت مرجع شامل ترکیبی از نژادها و نیز استفاده از روش BayesR به‌جای GBLUP، افزایش صحت از ۰/۴۳به ۰/۵۲ به‌ازاء ۳ صفت در گاو جرسی مشاهده شد.

لازم به ذکر است که در تمامی مطالعات گفته شده، جایگاه‌های واقع بر روی تراشه ایلومینای با تراکم بالا برای گاو از روی اطلاعات منتشر شده مربوط به حاشیه نویسی ژنوم گاو دسته بندی و ارزیابی شده‌اند ولی در مطالعه حاضر یک ترانسکریپتوم واقعی تشکیل و برای اولین بار از اطلاعات حاصل از آن برای این‌کار استفاده شد.

نتیجه گیری

در مجموع، نتایج مطالعه حاضر صحت و دقت بالای شناسایی SNP های به دست آمده و نیز کارایی داده‌های RNA-Seq را در کشف SNP تایید می‌کند. این مطالعه، فهرستی از تعداد قابل ملاحظه‌ای SNP بر روی ترانسکریپتوم جمعیتی از گاو هلشتاین

29 missense variants (non-synonymous variants)

ارائه می دهد که بسیاری از آن‌ها در مقایسه با SNPهای موجود در آرایه‌های مرسوم، جدید هستند. اما تمامی آن‌ها در بانک اطلاعات SNP های گزارش شده برای گاو و درصد قابل توجهی از آن‌ها در بین SNP هایی که تاکنون در پروژه توالی یابی ژنوم کامل ۱۰۰۰ گاو نر حاشیه نویسی شده اند حضور دارند. این امر صحت روند توالی‌یابی و آنالیز کشف SNP را در داده‌های RNA-Seq مورد آنالیز تایید می کند. ولی به‌هرحال، عملکرد و قابلیت کاربرد فهرست به دست آمده باید با استفاده از داده‌های ژنومی واقعی ویا شبیه سازی شده برای صفات مختلف مورد آزمون قرار گیرد. از سوی دیگر، این جایگاه‌ها صرفا بر روی نواحی رونویسی شده (توالی mRNA) و کد کننده قرار دارند. در صورت تایید کارآمدی، این فهرست می‌تواند در مطالعات بعدی و نیز برای اهداف اصلاح نژادی این گاو با استفاده از نشانگر توانمند SNP مورد استفاده قرار گیرد. مطالعه حاضر، سرآغازی بر ورود به عصر جدیدی از ارزیابی و انتخاب در حیوانات اهلی تحت عنوان "عصرپساژنومی"[30] است که می‌تواند با بهره‌گیری از سطوح مختلف داده‌های اومیکس ادامه یابد و روند پیشرفت ژنتیکی را در جمعیت‌های دامی سرعت بخشد.

تقدیر و تشکر

مولفین بر خود لازم می‌دانند صمیمانه از پروفسور حسن خطیب و تیم تحقیقاتی ایشان در دانشگاه ویسکانسین (آمریکا) که داده‌های RNA-Seq مورد استفاده در این تحقیق را فراهم نمودند و نیز به‌ویژه از جناب آقای مجید خان سفید در دانشگاه ملبورن (استرالیا) که زحمت ارزیابی فهرست ارائه شده در این مطالعه را بر روی داده های ژنومی متقبل شدند و علاوه بر آن، در تمام طول انجام این تحقیق از نظرات سازنده و مشاوره فنی ایشان در آنالیزها بهره بردیم، تشکر و قدردانی نمایند.

منابع

Canovas, A., Rincon, G., Islas-Trejo, A., Wickramasinghe, S., and Medrano, J. (2010). SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mammalian Genome*. 21: 592-598.

Djari, A., Esquerre, D., Weiss, B., Martins, F., Meersseman, C., Boussaha, M., et al. (2013). Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics*. 14: 307.

Do, D.N., Janss, L.L. G., Jensen, J., and Kadarmideen, H.N. (2015). SNP annotation-based whole genomic prediction and selection: An application to feed efficiency and its component traits in pigs. *Journal of Animal Science*. 93: 2056-2063.

Druet, T., Macleod, I.M., and Hayes, B.J. (2014). Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*. 112: 39-47.

Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., Reich, C.M., et al. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*. 95: 4114-4129.

Flintoft, L. (2008). Transcriptomics: Digging deep with RNA-Seq. *Nature Reviews Genetics*, 9: 568-568.

Haas, B.J. and Zody, M.C. (2010). Advancing RNA-seq analysis. *Nature Biotechnolog*, 28: 421-423.

Hayes, B., Bowman, P., Chamberlain, A., and Goddard, M. (2009). Invited review: genomic selection in dairy cattle: progress and challenge. *Journal of Dairy Science*, 92: 433 - 443.

Huang, W., Nadeem, A., Zhang, B., Babar, M., Soller, M., and Khatib, H. (2012). Characterization and Comparison of the Leukocyte Transcriptomes of Three Cattle Breeds. *PLoS ONE*, 7: e30244.

Khansefid, M., Pryce, J.E., Bolormaa, S., Miller, S.P., Wang, Z., Li, C., and Goddard, M. E. (2014). Estimation of genomic breeding values for residual feed intake in a multibreed cattle population. *Journal of Animal Science*. 92: 3270-3283.

[30] Post-Genomic Era

Koufariotis, L., Chen, Y.-P., Bolormaa, S., and Hayes, B. (2014). Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. *BMC Genomics*. 15: 436.

Levin, J., Adiconis, X., Yassour, M., Thompson, D., Guttman, M., Berger, M., et al. (2010). Development and evaluation of RNA-seq methods. *Genome Biology*. 11(Suppl 1): P26.

Marguerat, S. and Bahler, J. (2010). RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences*. 67: 569-579.

Meuwissen, T., Hayes, B. and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157: 1819 - 1829.

Morota, G., Abdollahi-Arpanahi, R., Kranis, A., and Gianola, D. (2014). Genome-enabled prediction of quantitative traits in chickens using genomic annotation. *BMC Genomics*. 15: 109.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods*. 5: 621-628.

Nicolazzi, E.L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., et al. (2015). SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC genomics*, 16: 283.

Park, K.-D., Park, J., Ko, J., Kim, B., Kim, H.-S., Ahn, K., et al. (2012). Whole transcriptome analyses of six thoroughbred horses before and after exercise using RNA-Seq. *BMC Genomics*. 13: 473.

Pennisi, E. (2012). ENCODE Project Writes Eulogy for Junk DNA. *Science*. 337: 1159-1161.

Ponting, C.P., Oliver, P.L. and Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell*. 136: 629-641.

Pryce, J., Arias, J., Bowman, P., Davis, S., Macdonald, K., Waghorn, G. et al. (2012). Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. *Journal of Dairy Science*. 95: 2108 - 2119.

Schaeffer, L. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 123: 218-223.

Sharma, U., Banerjee, P., Joshi, J. and Vijh R.K. (2012). Identification of SNPs in Goats (Capra hircus) using RNA-Seq Analysis. *International Journal of Animal and Veterinary Advances*, 4: 272-283.

Wang, K.C. and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. *Molecular Cell*. 43: 904-914.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*. 10: 57- 63.

Wilhelm, B.T. and Landry, J.R. (2009). RNA-seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*. 48: 249-257.

Zhang, C., Wang, G., Wang, J., Ji, Z., Liu, Z., Pi, X., and Chen, C. (2013). Characterization and Comparative Analyses of Muscle Transcriptomes in Dorper and Small-Tailed Han Sheep Using RNA-Seq Technique. *PLoS ONE*, 8: e72686.