

توجیه واریانس ژنتیکی صفت باقیمانده مصرف خوراک با چندشکلی های تک نوکلئوتیدی (SNP) کشف شده بر روی ترانسکرپتوم 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

تاریخ دریافت: اسفند ۱۳۹۴ تاریخ پذیرش: اردیبهشت ۱۳۹۵

شماره تماس نویسنده مسئول: ۰۹۱۲۱۰۳۶۴۴۸

• سیدرضا میرایی آشتیانی

گروه علوم دامی، دانشکده علوم و مهندسی کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران

Email: javaremi@ut.ac.ir

چکیده

چندشکلی های تک نوکلئوتیدی (SNP) امروزه به مهم ترین و کارآمدترین ابزار به نژادی تبدیل شده اند. مطالعات پوش 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 جایگاه ها هستند. بدین منظور آرایه های با تراکم بالا از این نشانگرها در صنعت گاو شیری ارائه شده اند. از آنجایی که ترانسکرپتوم تنها حاصل رونویسی بخش هایی از ژنوم است که رمزگر بوده و در نهایت به یک فرآورده پروتئینی ترجمه و بیان می شوند، احتمالاً 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 SNP ها بر روی توالی ترانسکرپتوم نمونه ای از جمعیت گاو ماده هلشتاین آمریکا کشف و سهم آن ها در توجیه واریانس ژنتیکی صفت باقیمانده مصرف خوراک (RFI) در جمعیتی از تلیسه های هلشتاین استرالیایی ارزیابی شد. این ترانسکرپتوم با همردیفی و مکان یابی خوانش های RNA-Seq بر روی ژنوم مرجع گاو تشکیل شد. در مجموع، یک فهرست شامل ۵۳۴۷۸ نشانگر SNP بر روی توالی ترانسکرپتوم جمعیت مورد مطالعه کشف شد که بیشتر آن ها در طراحی آرایه های SNP ایلومینا در نظر گرفته نشده اند. فهرست یافت شده با آرایه با تراکم بالا ایلومینا تنها ۶۳۳۶ نشانگر مشترک داشت. این فهرست مشترک نتوانست سهم بیشتری از واریانس ژنتیکی صفت باقیمانده مصرف خوراک را توجیه نماید. پیشنهاد می شود جایگاه های کشف شده در پیش بینی های مبتنی بر توالی کل ژنوم به کار برده شوند تا بهتر ارزیابی گردند.

واژه های کلیدی: چندشکلی تک نوکلئوتیدی (SNP)، توجیه واریانس ژنتیکی، باقیمانده مصرف خوراک (RFI)، ترانسکرپتوم، 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 Banabazi¹, Ardeshir Nejati Javaremi^{1*}, Ikhide G. Imumorin², Mostafa Ghaderi Zefrei³, Seyed Reza Miraei Ashtani¹

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 rely 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 transcriptome 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

mRNA باقی مانده و به پروتئین ترجمه (بیان) می‌گردند (Pennisi, ۲۰۱۲). بخش زیادی از RNA رونویسی شده به صورت نارمزگر^۲ باقی می‌ماند و احتمال می‌رود نقش تنظیمی بر سایر بخش‌های ژنوم داشته باشند (Ponting و همکاران، ۲۰۰۹؛ Wang و Chang، ۲۰۱۱). جامعیت ترانسکرپتوم به حدی است که محققین درگیر در پروژه ENCODE معتقدند می‌توان به جای ژن، ترانسکرپت (رونوشت)، یعنی قطعه ای از RNA که از روی DNA کد می‌شود را به عنوان واحد بنیادی ژنوم و واحد اساسی توارث پیشنهاد کرد. داده‌های ترانسکرپتومی به اشکال مختلف تولید می‌گردند که آخرین نسل از آن‌ها حاصل توالی‌یابی RNA (RNA-Seq)^۳ می‌باشند. این داده‌ها دقیق‌تر و کمی‌تر از

ترانسکرپتوم می‌تواند به درک بهتر ساز و کارهای حد واسط میان ژنوتیپ و فنوتیپ کمک نماید و دقت پیش‌بینی‌های ژنومی را که امروزه مهمترین ابزار ارزیابی و انتخاب برای به نژادگران دامی می‌باشد، بهبود بخشیده و پیشرفت ژنتیکی را سریع‌تر نماید. علاوه بر این، استفاده نکردن از حجم بالایی از داده‌های حاصل از فناوری توالی‌یابی نسل جدید و به‌ویژه داده‌های ترانسکرپتومی 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^۱ نشان می‌دهد که برخلاف دانسته‌های قبلی حدود ۸۰ درصد ژنوم رونویسی می‌گردد ولی در نهایت، تنها حدود ۳ تا ۵ درصد از این رونویسی‌ها به صورت

¹ Encyclopedia of DNA Elements (ENCODE)

² noncoding RNA (ncRNA)

³ RNA Sequencing (RNA-Seq)

نشانگرهاست که ارزش اصلاحی ژنومی پیش‌بینی شده (GEBV) برای هر فرد خواهد بود (Meuwissen و همکاران، ۲۰۰۱). برهم خوردن ارتباط میان نشانگر و صفت مورد مطالعه پس از گذشت چند نسل و نیاز به برآورد مجدد اثرات نشانگری در جمعیت مرجع، افزایش خطای محاسبات همراه با افزایش تراکم نشانگرهای موجود بر روی تراشه‌ها، پیچیدگی‌های محاسباتی ناشی از عدم توازن بین تعداد مشاهدات فنوتیپی (رکوردها) اندک و تعداد بسیار زیاد ژنوتیپ نشانگری به دست آمده به ازای یک فرد (که حتی بالغ بر ۷۰۰ هزار جایگاه هم می‌شود)، نبودن امکان گسترش اثرات نشانگری حاصل از یک جمعیت مرجع به جمعیت‌های هدفی که با آن تفاوت نژادی دارند، در نظر نگرفتن اثرات محیطی موثر بر تعداد رونوشت از هر جایگاه، در نظر نگرفتن اثرات غیرافزایشی میان جایگاه‌های نشانگر مورد استفاده و به ویژه اثرات اپی‌ستاتیک جایگاه‌ها و ... از جمله چالش‌های پیش روی این نوع پیش‌بینی‌هاست (Schaeffer، ۲۰۰۶؛ Hayes و همکاران، ۲۰۰۹).

تلفیق داده‌های RNA-Seq با داده‌های ژنومی و به‌کارگیری آن-ها در ارزیابی‌ها و انتخاب‌های ژنومی رایج ممکن است به رفع برخی از محدودیت‌های گفته شده بیانجامد و زمینه ساز ورود به عصر پساژنومی در اصلاح نژاد حیوانات اهلی گردد. از این رو، هدف از انجام مطالعه حاضر، کشف نشانگرهای SNP بر روی توالی mRNA رونویسی شده (ترانسکریپتوم) و بررسی میزان اشتراک فهرست به دست آمده با تراشه‌های رایج، و کاربرد فهرست مشترک به‌دست آمده در پیش‌بینی‌های ژنومی برای صفت باقیمانده مصرف خوراک^۴ (RFI) است. این هدف با این پیش‌فرض تعریف شده است که احتمالاً از آن‌جایی که ترانسکریپتوم تنها حاصل رونویسی بخش‌هایی از ژنوم است که رمزگر بوده و در نهایت به یک فرآورده پروتئینی ترجمه و بیان می‌شوند، بنابراین SNP‌های موجود در این مناطق نیز احتمالاً بهتر می‌توانند ارزش اصلاحی ژنومی را برآورد کنند. از سوی دیگر از آن‌جایی که این SNP‌ها در مناطقی واقع شده‌اند که در یک شرایط محیطی خاص رونویسی و بیان شده‌اند احتمالاً برای

انواع دیگر می‌باشند (Flintoft، ۲۰۰۸؛ Wang و همکاران، ۲۰۰۹). اساس این روش بر برش کل mRNA رونویسی شده (یا به عبارت دیگر کل ترانسکریپتوم^۴ یا کل اگزوم^۵) به قطعات کوچک‌تر^۶ و توالی‌یابی این قطعات با استفاده از چارچوب‌های فنی^۷ و پربرونداد NGS استوار است (Haas و Zody، ۲۰۱۰؛ Marguerat و Bahler، ۲۰۱۰). تفاوت میزان رونویسی در بین خوانش‌های RNA-Seq، به صورت تفاوت در تعداد دفعات خوانش^۸ قطعات به هنگام توالی‌یابی به روز می‌نماید (Wilhelm و همکاران، ۲۰۰۹). با به دست آمدن یک معیار عددی بسیار دقیق اندازه‌گیری بیان ژن، این روش را می‌توان آنالیز بیان ژن دیجیتالی^۹ نیز نامید (Mortazavi و همکاران، ۲۰۰۸).

استفاده از اطلاعات ژنتیکی مبتنی بر نشانگرهای مولکولی در برنامه‌های اصلاح نژاد گاو به عنوان ابزاری برای بهبود انتخاب‌های فنوتیپی مرسوم رو به افزایش بوده است. ظهور فناوری‌های توالی‌یابی جدید فرصتی تازه برای شناسایی واریانت‌های مرتبط با ژن-های موثر بر صفات اقتصادی پدید آورده‌اند. علیرغم سیر نزولی هزینه توالی‌یابی، توالی‌یابی کل ژنوم یا هدف قرار دادن مناطق بزرگ هنوز برای اکثر آزمایشگاه‌ها بسیار گران است. توالی‌یابی 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 پیش‌بینی‌های مبتنی بر ژنوم از روی ژنوتیپ تعداد زیادی نشانگر (چندشکلی‌های تک نوکلئوتیدی^{۱۲} یا به اختصار SNP) که پیش از آن میزان تاثیر هر یک با صفت مورد بررسی در یک جمعیت مرجع برآورد شده است، انجام می‌شود. این پیش‌بینی‌ها به‌طور ساده مبتنی بر جمع‌گیری اثرات این

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)

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) خوانش های بی کیفیت می باشد. فهرست آداپتورهای به کار رفته در هنگام انجام توالی یابی و تشکیل کتابخانه های^{۱۷} 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^{۱۸} و مکان یابی^{۱۹} خوانش های کوتاه بر روی ژنوم مرجع با استفاده از نرم افزار Tophat2، که از bowtie2 به عنوان همردیف کننده استفاده می کند، انجام شد. همردیفی و مکان یابی خوانش ها به تشکیل (اسمبلی) ترانسکریپتوم^{۲۰} 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^{۲۱} بر روی ترانسکریپتوم تشکیل شده و فیلتر نمودن آن ها 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 تراشه های مرسوم طراحی شده توسط شرکت ایلومینا (آمریکا) شامل فهرست های با تراکم های متفاوت شامل Illumina Bovine 50K-v2, 3K BeadChip و HD (به ترتیب شامل ۲۹۰۰، ۵۴۶۰۹ و ۷۷۹۶۲ نشانگر در سرتاسر ژنوم)، اطلاعات تراشه های مذکور از بانک اطلاعاتی V.3

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

مواد و روش ها

در این مطالعه از داده های RNA-Seq مربوط به ادغام ۴۰ نمونه از گاو ماده هلشتاین مزرعه دانشگاه ویسکانسین آمریکا استفاده شد. این داده ها در مطالعه Huang و همکاران (۲۰۱۲) تولید شد. به منظور تولید داده های مذکور، حجم برابر از نمونه های خون انفرادی بلافاصله پس از اخذ از سیاهرگ دمی با هم مخلوط و سپس استخراج کل RNA از مخلوط مذکور انجام شد. توالی یابی 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 خوانش های کوتاه (SRA) بانک اطلاعاتی مرکز ملی اطلاعات بیوتکنولوژی یا NCBI دریافت شد. داده مورد آنالیز شامل ۲۱۰۷۸۴۷۷ خوانش کوتاه جفتی^{۱۴} به طول ۷۵ جفت باز بود.

تمام آنالیزها در محیط سیستم عامل لینوکس اوبونتو (Ubuntu) نسخه ۱۴.۰۴ انجام شدند. فرمت sra به خاطر حجم کم صرفا برای بارگذاری بر روی بانک های اطلاعاتی مناسب است و برای آنالیز ابتدا باید به فرمت های دیگر تبدیل گردد که یکی از مرسوم ترین آن ها 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، خوانش های با وفور بیش از حد^{۱۵}، توزیع طول خوانش، خوانش های تکراری^{۱۶} و ... بر روی داده ها بررسی و کیفیت با توجه به مجموع آن ها سنجیده شد. پس از ویرایش

¹⁶ Duplicate Sequences

¹⁷ Library

¹⁸ Alignment

¹⁹ Mapping

²⁰ Transcriptome Assembly

²¹ SNP Discovery

¹⁴ Paired

¹⁵ Overrepresented Sequences

شد. سپس از ماتریس های 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_g^2) شامل مولفه پلی ژنیک ($\sigma_{polygenic}^2$)، کل SNP ها ($\sigma_{gAll_SNP}^2$)، SNP های مشترک ($\sigma_{gCommon_SNP}^2$) و واریانس باقیمانده (σ_e^2) 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) به طور جداگانه برای فهرست های شامل تمام SNP ها، SNP های 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 مدل ها، بر اساس لگاریتم درست نمایی ($LogL$)^{۲۵} برآورد شده برای هر مدل انجام شد. به عبارت دیگر هر مدلی که $LogL$ بزرگ تری داشته باشد بهتر عمل کرده است (Khansefid و همکاران، ۲۰۱۴).

نتایج و بحث

در مجموع تعداد جایگاه SNP ۵۳۴۷۸، بر روی ترانسکریپتوم گاو هلشتاین آمریکا کشف شد. از مجموع کل SNP های موجود در بانک اطلاعات مربوط 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 این دو 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 نیز حضور دارند. البته نتایج منتشر شده مربوط به این پروژه تاکنون نتایج توالی یابی کامل ژنوم ۲۳۴ گاو 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 داده های RNA-Seq را تایید می کند. Canovas و همکاران (۲۰۱۰) با استفاده

Nicolazzi SNPchiMp 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_sn_ps_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 تعداد ۲۶۷۱۶۸۹۶ نشانگر SNP حاشیه نویسی شده در این پروژه گزارش شده است. از آنجایی که تمام SNP های یافت شده در فهرست SNP های تراشه های مرسوم برای پیش بینی های 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 شاخص های دیگر برای فیلتر نمودن در لیست نهایی باقی می ماندند را می توان بر روی داده های ژنومی واقعی استفاده نمود و عملکرد آنها را با تراشه های 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. جزئیات مربوط به محاسبه باقیمانده مصرفی و مشخصات حیوانات رکوردبرداری شده در مطالعه Pryce و همکاران (۲۰۱۲) آمده است. تمامی تلیسه های مذکور با تراشه Illumina HD Bovine vWk SNP chip (شرکت Illumina، سان دیه گو، کالیفرنیا، آمریکا) تعیین ژنوتیپ شده اند. پس از کنترل کیفیت که عمدتاً بر اساس فراوانی آلل ماینور (MAF)^{۲۲} و سطح احتمال انحراف از تعادل هاردی-واینبرگ انجام شد، تعداد ۶۰۶۰۹۶ SNP از تراشه مذکور باقی ماند. ماتریس روابط ژنومی (GRM)^{۲۳} با استفاده از این تعداد SNP و نیز فهرست SNP های مشترک تشکیل شد (Khansefid و همکاران، ۲۰۱۴). همچنین ماتریس روابط شجره ای (NRM)^{۲۴} بر اساس ساختار شجره گفته شده محاسبه

²² Minor Allele Frequency (MAF)

²³ Genomic Relationship Matrix (GRM)

²⁴ Pedigree Relationship Matrix (NRM)

²⁵ Log of Likelihood (LogL)

Park و همکاران (۲۰۱۲) بر روی توالی ترانسکریپتوم حاصل از دو بافت خون و عضله ۶ اسب تارو برد پیش و پس از مسابقه، تعداد ۱۸۳۹۷۳ نشانگر SNP پیدا نمودند که تقریباً 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، احتمالاً 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 به ترتیب با طول ۱۵۸۳۳۷۰۶۷ و ۱۴۸۸۲۳۸۹۹ جفت باز طولانی‌ترین کروموزوم‌ها در ژنوم گاو می‌باشند و تقریباً سه برابر کروموزوم ۲۵ طول دارند، به ترتیب ۲۱۹۲ و ۹۰۲ SNP را شامل می‌شوند. توالی‌های غیر کروموزومی (توالی‌هایی که هنوز به هیچ کروموزوم مشخصی منتسب نشده‌اند) و توالی ژنوم میتوکندریایی به ترتیب با طول ۹۱۲۱۷۵۷ و ۱۶۳۳۸ جفت باز نیز به ترتیب ۱۵۵ و ۱۴ SNP داشتند. نتایج فوق نشان می‌دهند که در مجموع رونویسی از تمام طول ژنوم با یک توزیع همگن و با پوششی یکسان صورت نمی‌گیرد. به عبارت دیگر، احتمالاً برخی نواحی حامل ژن‌های کاندیدای بیشتر و یا 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

از RNA-Seq بیش از ۳۳۰۰۰ جایگاه SNP را در ترانسکریپتوم شیر گاو و نواحی کدکننده ژن‌های بیان شده در طی شیردهی معرفی نمودند که به ادعای این محققین می‌تواند برای ژنوتیپ‌یابی و انجام مطالعات مربوط به بررسی ارتباط نشانگر-صفت در گاو 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 راه کارهایی 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 توالی‌یابی پربروناد قرار گرفت. جمعاً ۱۹۷۵۲ رونوشت شناسایی و ۳۴۳۷۶ SNP تشخیص داده شدند. ۵۵ درصد SNP ها در نواحی رمزگر یافت شدند و حدود ۲۲ درصد منجر به یک تغییر اسید آمینه‌ای 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 SNP های رمزگر تازه کشف شده، منابع ژنومی در دسترس برای گاو و به ویژه گاو گوشتی را بهبود می‌بخشند. 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 توالی‌یابی 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.

کل ترانسکریپتوم تشکیل شده را به خود اختصاص می دهند و در نتیجه SNP ها در این نواحی فراوانی بالاتر داشته و بعد از گذراندن فیلترهای مختلف همچنان در فهرست نهایی باقی مانده- اند. بررسی دقیق تر نواحی رونویسی شده و پوشش^{۲۶} ترانسکریپتومی این نواحی احتمالاً درک ما را از رونویسی (بیان) ژن ها و تنظیم آن در جمعیت مورد مطالعه افزایش خواهد داد. این تعداد بسیار کم عملاً امکان استفاده و آزمون SNP های مشترک در پیش بینی های مبتنی بر ژنوم و آزمون و مقایسه عملکرد آن ها را با تراشه های اصلی برای دو فهرست 3K و 50K-v2 نمی دهد. لذا تنها فهرست مشترک با تراشه بسیار متراکم بر روی داده های 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 تراشه های رایج شرکت ایلومینا با تراکم های متفاوت شامل 3K، 50K-v2 و Illumina Bovine HD (به ترتیب شامل ۲۹۰۰، ۵۴۶۰۹ و ۷۸۶۷۹۹ نشانگر SNP) نشان داد که به ترتیب ۵۰، ۶۱۹ و ۹۶۷۷ جایگاه SNP مشترک بین آن ها وجود دارد (جدول ۲ و دو ستون آخر جدول ۱ به تفکیک هر کروموزوم). مقایسه مدل های مختلف برازش شده برای برآورد مولفه های 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.

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

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

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 BovineHD BeadChip)، *** آرایه با تراکم بالا (BovineSNP50v2 BeadChip)

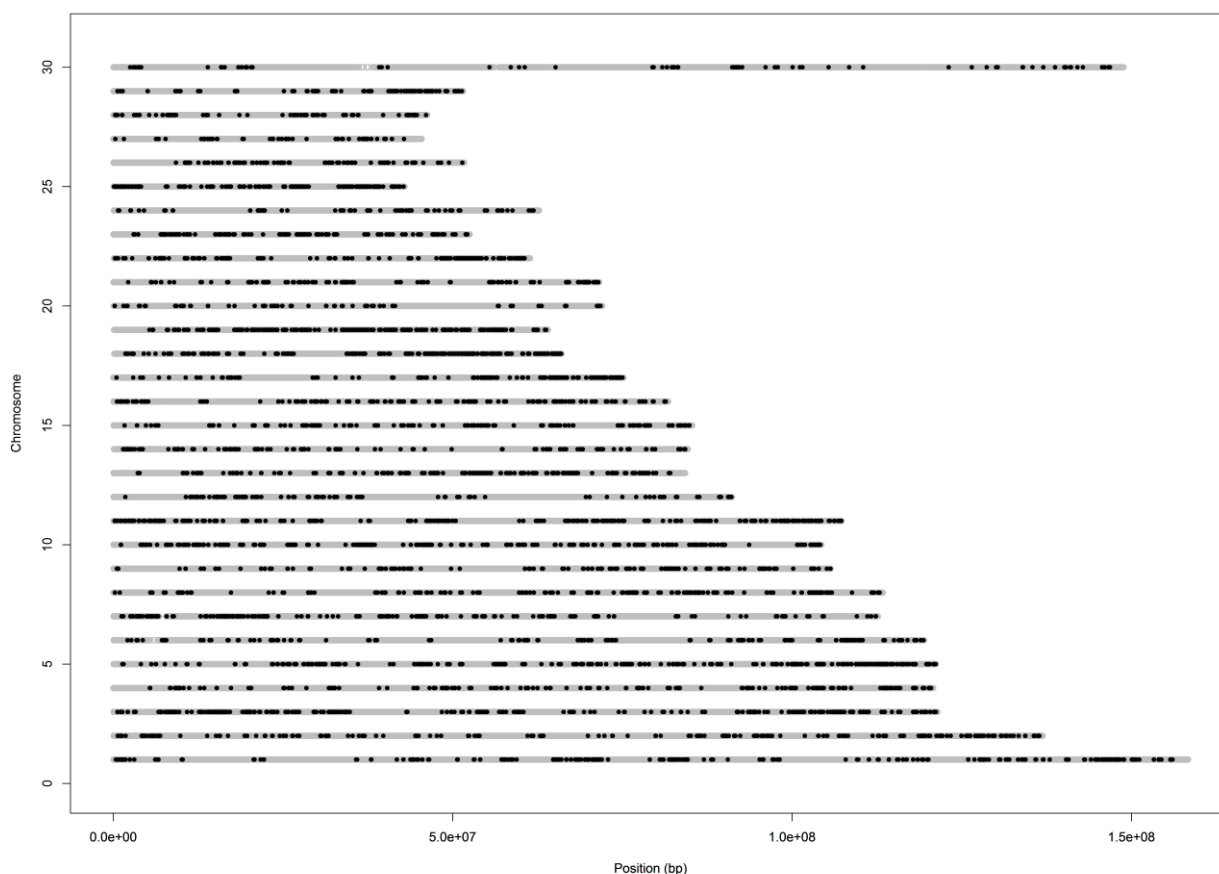
این امر نشان می دهد که بازنگری در فهرست آرایه های مرسوم به کار رفته برای پیش بینی های مبتنی بر ژنوم در گاو و اضافه نمودن تعداد بیشتری از SNP های رمزگر به آنها، احتمالاً می تواند به بهبود چنین پیش بینی هایی بیانجامد. این SNP های رمزگر ممکن است آن دسته از نشانگرهای SNP را شامل شود که در تعداد زیادی ترانسکریپتوم 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 داده های توالی یابی 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 گاونر) ارزیابی نمود. از آنجایی که در پیش بینی های مبتنی بر توالی کل ژنوم نیز همچنان نشانگرهای SNP به کار خواهند رفت، می توان در صورت عملکرد بهتر فهرست های کشف شده بر روی ترانسکریپتوم، آنها را برای پیش بینی های مبتنی بر توالی کامل ژنوم پیشنهاد نمود. به هر حال، به تدریج پیچیدگی های فنی و محاسباتی و نیز مسائل هزینه ای این نوع پیش بینی ها کمتر شده و راه کار نوینی را پیش روی به نژادگران دامی قرار گرفته است (Druet و همکاران، ۲۰۱۴).

همچنان که از جدول مذکور برمی آید، 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 (بالا ترین 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 حالی که 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 مدل های در برگیرنده کل SNP ها و SNP های مشترک به ترتیب ۰/۳۹۷۳±۰/۰۹۲۰ و ۰/۲۸۴۸±۰/۰۷۳۸ برآورد گردید. همان طور که 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 فهرست های 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.

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	-۴۵۱.۳۶۶	-	-	۰.۳۴۶۹	۰.۹۹۸۹	۰.۷۵۲	۰.۲۴۷۲
GRM_allSNPs	-۴۴۱.۳۳۹	۰.۴	-	-	۱.۰۰۶۸	۰.۶۰۶۸	۰.۳۹۷۳
GRM_allSNPs + NRM	-۴۴۶.۴۴۶	۰.۳۹۶۲	-	۰.۰۱۷۹	۰.۹۸۷۴	۰.۷۴۵۳	۰.۲۴۵۲
GRM_6336SNPs	-۴۴۴.۱۰۸	-	۰.۲۸۰۹	-	۰.۹۸۷۱	۰.۷۰۶۲	۰.۲۸۴۶
GRM_6336SNPs + NRM	-۴۴۴.۸۰۳	-	۰.۲۷۳۵	۰.۰۳۷۷	۰.۹۸۸۴	۰.۶۸۷۱	۰.۳۰۴۸
GRM_allSNPs + GRM_6336SNPs	-۴۴۱.۱۵۰	۰.۳۳۶۸	۰.۰۶۶۶	-	۱.۰۰۴۶	۰.۶۰۱۸	۰.۴۰۱
GRM_allSNPs + GRM_6336SNPs + NRM	-۴۴۴.۱۴۹	۰.۳۳۶۵	۰.۰۶۴۹	۰.۰۰۶۸	۱.۰۰۰۵۱	۰.۵۹۶۹	۰.۴۰۶۱

مولفه های واریانس شامل اجزاء پلی ژنتیک ($\sigma^2_{gpolygenic}$)، کل SNP ها ($\sigma^2_{gAll_SNP}$)، SNP های مشترک ($\sigma^2_{gCommon_SNP}$) و واریانس باقیمانده (σ^2_e) است. مدل های مورد بررسی شامل (۱) استفاده از اطلاعات شجره های (NRM)، (۲) برآورد ماتریس روابط ژنومی با استفاده از کل SNP ها و SNP های مشترک (GRM_allSNPs) و (GRM_6336SNPs) و ترکیب برآورد ماتریس روابط ژنومی و اطلاعات شجره ای با استفاده از کل SNP ها، SNP های مشترک و ترکیب دو فهرست باهم (GRM_allSNPs + NRM)، GRM_allSNPs + NRM، GRM_6336SNPs + NRM و 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 می دهد. کروموزوم شماره ۳۰ همان کروموزوم X است.

پانل کاهش یافته قرار گرفته اند. در نتیجه اجرای این استراتژی، احتمالاً تعداد بیشتری SNP و با پراکنش همگن تر در سرتاسر ژنوم خواهیم داشت که ممکن است عملکرد بهتری نیز داشته باشند. Koufariotis و همکاران (۲۰۱۴)، واریانت های موجود در آرایه متراکم Illumina Bovine HD را در قالب ۱۲ گروه دسته بندی نمودند. نتایج آن ها نشان داد همان طور که انتظار می رفت، واریانت های بین ژنی (که بر روی DNA نامرنگ واقع هستند) معمول ترین SNP ها بودند (حدود ۶۷ درصد) و درون ژنی ها ۳۲ درصد از کل را شامل می شدند. البته به نسبت سهم ۳ تا ۵ درصدی از کل ژنوم که به صورت mRNA رونویسی می شود، سهم حدود ۳۰ درصدی از کل SNP های یک آرایه با تراکم بالا قابل ملاحظه است. اما برخی مطالعات نشان داده اند که تمامی SNP هایی از یک آرایه با تراکم بالا که در مناطق کد کننده قرار

پیشنهاد می شود که به جای کشف SNP بر روی توالی ترانسکریپتوم و استفاده و ارزیابی SNP های مشترک با آرایه های مرسوم، تنها حضور موقعیت نوکلئوتیدی یک SNP از آرایه های مرسوم بر روی توالی ترانسکریپتوم های واقعی ملاک انتخاب آن برای قرار گرفتن در فهرست جایگاه های 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 SNP های فراخوانده شده الزاماً بر روی توالی ترانسکریپتوم های واقعی تشکیل شده چندشکل (SNP واقعی) نیستند بلکه تنها به این دلیل که موقعیت ژنومی آن ها در نواحی رونویسی شده قرار داشته و از این رو متناظر با یک موقعیت نوکلئوتیدی در توالی ترانسکریپتوم های تشکیل شده نیز هست، در

²⁷ Reduced SNP Panel

²⁸ SNP Calling

نمودند. جمعا ۱۲۷۲ خوگ Duroc که هم رکوردهای فوتویی و هم ژنوتیپی داشتند را به دو جمعیت مرجع (۹۶۸ خوگ) و جمعیت تایید (۳۰۴ خوگ) 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 ۰/۴۳ به 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 این کار 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

دارند (حداکثر حدود ۳۰ درصد از کل آرایه) نیز نمی‌توانند 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 و یا 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 (SNP های معنی‌دار در GWAS) در دسته واریانت‌های عامل (واریانت‌های نامترادف)^{۲۹} و نواحی ۵ کیلوگفت بازی بالا دست و پایین دست ژن‌های کد کننده یافت شد. این نتایج نشان دادند که حاشیه نویسی عملکردی می‌تواند در تعیین اولویت واریانت‌ها برای قرارگرفتن در یک زیرمجموعه، که با احتمال زیادتری با صفات پیچیده مرتبط خواهد بود، کمک کند. همچنین، روش برآورد مولفه‌های 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 می‌باشد. 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 SNP ها 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

- 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 Biotechnology*, 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.

ارائه می دهد که بسیاری از آن ها در مقایسه با SNP های موجود در آرایه های مرسوم، جدید هستند. اما تمامی آن ها در بانک اطلاعات SNP های گزارش شده برای گاو و درصد قابل توجهی از آن ها 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 توالی یابی و آنالیز کشف SNP را در داده های RNA-Seq مورد آنالیز تایید می کند. ولی به هر حال، عملکرد و قابلیت کاربرد فهرست به دست آمده باید با استفاده از داده های ژنومی واقعی و یا شبیه سازی شده برای صفات مختلف مورد آزمون قرار گیرد. از سوی دیگر، این جایگاه ها صرفاً بر روی نواحی رونویسی شده (توالی mRNA) و کد کننده قرار دارند. در صورت تایید کارآمدی، این فهرست می تواند 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 "عصر پساژنومی"³⁰ است که می تواند با بهره گیری از سطوح مختلف داده های اومیکس ادامه یابد و روند پیشرفت ژنتیکی را در جمعیت های دامی سرعت بخشد.

تقدیر و تشکر

مولفین بر خود لازم می دانند صمیمانه از پروفیسور حسن خطیب و تیم تحقیقاتی ایشان در دانشگاه ویسکانسین (آمریکا) که داده های RNA-Seq مورد استفاده در این تحقیق را فراهم نمودند و نیز به ویژه از جناب آقای مجید خان سفید در دانشگاه ملیورن (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.,

