ژنتیک نوین
دوره نوزدهم، شماره ٤، زمستان ١٤٠٣
صفحه ٣٢٣ - ٣١٧

مقایسه تکنیکال قابلیت‌های دو نرم‌افزار کاربردی TopHat2 و HISAT2 در تجزیه و تحلیل بیوانفورماتیکی داده‌های ترانسکریپتومی مرغ

Technical comparison of the capabilities of two software applications TopHat2 and HISAT2 in bioinformatics analysis of chicken transcriptomic data

سپیده رستمی[1]، محمدتقی بیگی نصیری*[2]، محمود نظری[2]، محسن چراغی زاده[3]

۱- دانشجوی دکترای ژنتیک و اصلاح دام و طیور، گروه علوم دامی، دانشکده علوم دامی و صنایع غذایی، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان، ملاثانی، ایران
۲- به‌ترتیب استاد، دانشیار گروه علوم دامی، دانشکده علوم دامی و صنایع غذایی، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان، ملاثانی، ایران
۳- استادیار مرکز تحقیقات مهندسی سطح پیشرفته و نانومواد، گروه مهندسی برق، دانشگاه آزاد واحد اهواز، ایران

Rostami S[1], Beigi Nassiri MT*[2], Nazari M[2], Cheraghizadeh M[3]

1- PhD Student of Animal genetic and Breeding, Faculty of Animal Science and Food Technology, Agricultural Science and Natural Resources University of Khuzestan, Mollasani, Iran
2- Professor, Associate Professor, Department of Animal Science, Faculty of Animal science and Food Technology, Agricultural Science and Natural Resources University of Khuzestan, Mollasani, Iran
3- Associate Professor, Advanced Surface Engineering and Nano Materials Research Center Department of Electrical Engineering Ahvaz Branch, IAU, Ahvaz, Iran

* نویسنده مسئول مکاتبات، پست الکترونیکی: Mt_nassiri@yahoo.com

(تاریخ دریافت: ١٤٠٣/٠٩/٢٠ - تاریخ پذیرش: ١٤٠٣/١٢/٢٨)

چکیده

در روند تجزیه و تحلیل بیوانفورماتیکی داده‌های اومیکس، پس از فرآیند کنترل کیفیت، اولین گام تجزیه و تحلیل اینست که خوانش‌های هر نمونه مبتنی بر ژنوم مرجع همتراز شوند. در این راستا، نرم‌افزارهای متعددی توسط محققان برای این مرحله همترازی پیشنهاد شده است. با این انگیزه تحقیقاتی، هدف از مطالعه حاضر، مقایسه قابلیت‌های دو نرم‌افزار همترازکننده HISAT2 و TopHat2 است که به بررسی عملکرد و دقت این دو ابزار در همترازی خوانش‌های توالی‌یابی‌شده در تحلیل داده‌های ترانسکریپتومی (RNA-Seq) می‌پردازد. HISAT2، برنامه سریع و حساس برای همترازی خوانش‌های تولید شده توسط توالی‌یابی نسل جدید، به‌دلیل سرعت بالا و نیاز به حافظه کمتر مورد توجه قرار گرفته است. در مقابل، TopHat2 توانایی شناسایی مکان‌های اتصال جدید را با نقشه‌یابی مستقیم به رونوشت‌های شناخته شده ترکیب می‌کند، که هم‌ترازی‌های حساس و دقیقی را حتی برای ژنوم‌های بسیار تکراری ایجاد می‌کند. بدین منظور، از مجموع داده‌های RNA-seq ٤ نمونه جنین جوجه استفاده شد. بعد از کنترل کیفیت داده‌ها با استفاده از نرم‌افزار FastQC، ویرایش توالی مبتنی بر نرم‌افزار Trimmomatic انجام گرفت. نتایج این پژوهش نشان داد که نرم‌افزار HISAT2 می‌تواند ٩٣/٦٨٧٪ از خوانش‌ها را با ژنوم رفرنس با موفقیت همتراز کند. درحالیکه، این مقدار توسط نرم‌افزار TopHat2، ٨٥/٨٥ درصد محاسبه شد. همچنین، HISAT2 با ٨٨/٦٨ درصد از داده‌ها به یک جایگاه خاص در ژنوم اختصاص پیدا کرد، در حالی‌که، این ارزش عددی برای TopHat2 ٧٩/٣٥ درصد به‌دست آمد. در مجموع پیام این پژوهش می‌تواند این عبارت باشد که نرم‌افزار HISAT2 در اکثر موارد عملکرد بهتری نسبت به نرم‌افزار TopHat2 دارد. این ویژگی‌ها باعث می‌شود برنامه HISAT2 برای پروژه‌های بزرگ و داده‌های پیچیده ترانسکریپتومی انتخاب مناسب‌تری باشد. با این حال، TopHat2 هنوز در برخی موارد خاص، مانند تحلیل‌های مربوط به ادغام ژن‌ها، می‌تواند نتایج مفیدی ارائه دهد. بنابراین، ملاک‌های انتخاب بین این دو نرم‌افزار باید بر اساس نیازهای خاص تحقیق و نوع داده‌های مورد استفاده انجام شود.

واژه‌های کلیدی

همترازی خوانش
ترانسکریپتوم
HISAT2
TopHat2

[Downloaded from mg.genetics.ir on 2025-07-17]

مقدمه

فناوری توالی‌یابی RNA (RNA-Seq) با قابلیت عملیاتی بالا، شناسایی و تعیین کمیت همه RNAهای بیان شده از یک نمونه را ممکن می‌سازد و قابلیت کاوش در مناظر پیچیده رونویسی را افزایش می‌دهد (Marguerat and Bähler 2010)، همچنین، بعنوان فناوری که در مقیاس گسترده‌ای مورد پذیرش قرار گرفته، توسعه یافته و در ارائه پوشش بیشتر ترانسکریپت و شناسایی ایزوفرم‌های جدید آن با امکان یافتن مکانیزم‌های اتصال جایگزین، بر میکروآرایه‌ها پیشی گرفته است. با وجود پیشرفت سریع و مقرون به صرفه بودن ترانسکریپتومیکس در تولید داده‌های کلان، تحلیل مؤثر بیوانفورماتیکی برای تفسیر معنادار داده‌ها با اهمیت زیستی ضروری است (Raplee et al. 2019).

در این سناریوی مورد بحث، به دلیل فرآیند اسپلایسینگ که در RNA یوکاریوتی رخ می‌دهد، انتخاب ابزار مناسب ترازکننده از اهمیت زیادی برخوردار است. بنابراین، نرم‌افزار مورد استفاده باید توانایی شناسایی سریع جایگاه‌های هضم آنزیمی را داشته باشد. در این پازل محاسباتی، مرحله هم‌ترازی (یا مکان‌یابی) تجزیه و تحلیل داده‌های RNA-Seq فشرده‌ترین و زمان‌برترین فرآیند محاسباتی است، زیرا، شامل همترازی خوانش‌های تولید شده از یک آزمایش توالی‌یابی با ژنوم مرجع است. زمانیکه، کارایی محاسباتی و دقت بیولوژیکی نتایج از جنبه‌های مرتبط باشند، انتخاب ابزار همترازی مناسب برای این کار، اساسی است (Bianchi et al. 2023).

همترازی اولین مرحله در اکثر مسیرهای آنالیز RNA-seq است و دقت تجزیه و تحلیل‌های مراحل بعد به‌شدت به آن بستگی دارد. الگوریتم‌های زیادی برای این مرحله همترازی ایجاد شده است. با توجه به رشد روزافزون استفاده از نرم‌افزارهای همترازی، این نرم‌افزار از اهمیت ویژه‌ای برخوردار شده است. این موضوع، به ظاهر بی‌ارزش است اما مقایسه نتایج حاصل از رویکردهای مختلف گیج‌کننده و دشوار است (Bahrami 2020). در این راستا، چندین پلتفرم نرم‌افزاری برای همترازی به یک ژنوم مرجع، از جمله Tophat2 (Kim et al 2013) و Hisat2 ارائه شده است (Kim et al. 2019). TopHat2 یک انتخاب محبوب بود اما، به‌دلیل ناکارآمدی محاسباتی آن توسط HISAT2 جایگزین شده است. TopHat2 و HISAT2 بر روی ابزار محبوب همترازی کوتاه Bowtie2 ساخته شده‌اند (Langmead and Salzberg 2012). در حالی‌که، هر دو ابزار بیوانفورماتیکی تراز کننده سریع در نظر گرفته می‌شوند، انتخاب تراز بهینه می‌تواند به‌طور قابل توجهی بر تحلیل پایین دست تأثیر بگذارد. بنابراین، ارزیابی عملکرد ترازکننده‌های مختلف برای شناسایی ابزار بهینه برای کار بسیار مهم است (Westermann and Vogel 2021). TopHat2 از Bowtie برای هم‌راستا کردن خوانش‌ها به مناطق اگزون‌ها با استفاده از همترازی خوانش‌های تقسیم‌شده استفاده می‌کند، HISAT2 از یک رویکرد هم‌راستایی مبتنی بر گراف بهره می‌برد (Kim et al. 2013). این رویکردها، ممکن است هم‌راستایی‌های حساس‌تر و دقیق‌تری به ژنوم ارائه دهند (Dobin et al. 2013).

چندین مطالعه وجود دارد که ابزارهای مختلف را برای تجزیه و تحلیل داده‌های توالی‌یابی RNA مقایسه کرده است. نتایج مطالعات مختلف نشان داده که برنامه Star2 از نظر درصد خوانش‌های منحصر به فرد همتراز شده در مقایسه با HISAT2، عملکرد بهتری دارد و درصد خوانش همتراز نشده در HISAT2 و TopHat2 در هر دو مجموعه ژنومی بیشتر است (Dharshini et al. 2020). برنامه HISAT2 تعداد خوانش‌های کمتری را تراز می‌کند و همسویی بیشتری با توالی‌های مشابه دارد، که دقت هم‌ترازی را به خطر می‌اندازد و به‌طور بالقوه منجر به خروجی‌های اشتباه می‌شود (Raplee et al. 2019). محققین نشان دادند که برنامه TopHat2 در مقایسه با HISAT2 و STAR برای mRNA اسپرم به‌دلیل درصد همترازی اگزونیک بالاتر و کمی‌سازی ژن‌های کم بیان مؤثر بوده است (Ramya et al. 2021). Elyasi Zarringhabaie و همکاران (۲۰۲۳)، به مقایسه نرم‌افزارهای TopHat2, HISAT2 و Bowtie2 جهت تجزیه و تحلیل داده‌های ترانسکریپتومی گاو پرداختند. اما، تاکنون مطالعه‌ای جهت انتخاب نرم‌افزار مناسب برای تجزیه و تحلیل داده‌های ترانسکریپتومی مرغ انجام نشده است. بدین منظور، هدف از مطالعه حاضر مقایسه دو مسیر تجزیه و تحلیل بیوانفورماتیک مختلف با استفاده از HISAT2 و TopHat2 به‌منظور ارزیابی عملکرد و کیفیت خروجی آن‌ها و شناسایی ابزار بهینه برای همترازی دقیق و تجزیه و تحلیل داده‌های ترانسکریپتومی (RNA-seq) مرغ بود.

[Downloaded from mg.genetics.ir on 2025-07-17]

مواد و روش‌ها

برای انجام این پژوهش از داده‌های مجموع چهار نمونه جنین جوجه استفاده شد. در ابتدای کار، RNA بافت جنینی با استفاده از کیت جداسازی RNA از شرکت دنازیست و مطابق دستورالعمل شرکت سازنده استخراج شد. برای بررسی خلوص RNA از الکتروفورز و اسپکتروفتومتر NanoDrop (USA) استفاده شد و هرگونه تخریب و آلودگی بر روی ژل آگارز ۱ درصد مورد بررسی قرار گرفت (Rostami et al, 2019). پس از استخراج RNA کل و اطمینان از کمیت و کیفیت نمونه‌های موردنظر، نمونه‌ها برای توالی‌یابی ابتدا به شرکت توپاز ژن کاوش و از آن‌جا به شرکت Novogene کشور چین ارسال شدند. توالی‌یابی نمونه‌ها با استفاده از فناوری ایلومینا Novaseq 6000 (Illumina, Truseq stranded Total RNA kit) صورت پذیرفت. سنجش کیفیت اولیه داده‌های خام با FastQC (v0.11.5) بررسی شد (Andrews 2010). حذف خوانش‌های با کیفیت پایین و حذف آداپتورها با استفاده از نرم‌افزار Trimmomatic (نسخه 0.39) با پارامترهای: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:25 TRAILING:25 MAXINFO:90:0.90 AVGQUAL:20 MINLEN:50 انجام شد. از "ILLUMINACLIP" برای شناسایی آلودگی آداپتوری با استفاده از TruSeq3-PE.fa استفاده شد، که شامل آلودگی آداپتوری با ۲ باز غیر منطبق، با حد آستانه روش پالیندرومی شناسایی آداپتور ۳۰ و با حد آستانه روش ساده شناسایی آداپتور ۱۰ بود. همچنین، از "LEADING:25" و "TRAILING:25" برای حذف خوانش‌های با کیفیت کمتر از ۲۵ از ابتدا و انتهای خوانش‌ها استفاده شد. "MAXINFO:90:0.90" برای برش خوانش‌ها بر اساس کیفیت استفاده شد، که شامل حداکثر طول ۹۰ و حداقل نسبت کیفیت ۰/۹۰ بود. تنها خوانش‌هایی با طول بیشتر از ۵۰ نوکلئوتید با استفاده از "MINLEN:50" حفظ شدند. در نهایت، خوانش‌هایی با میانگین کیفیت کمتر از ۲۰ با استفاده از "AVGQUAL:20" برش خورده‌اند (Bolger et al. 2014).

جهت همترازی خوانش‌های ترانسکریپتوم با ژنوم مرجع جوجه از دو نرم‌افزار HISAT2 (v2-2.2.1) (Bankizadeh et al. 2021، Bankizadeh et al. 2022) و TopHat2 (v2-1.1) (Kim et al. 2015) استفاده شد. ژنوم مرجع و فایل GTF جوجه از Ensembl (https://ftp.ensembl.org/pub/release-110/fasta/gallus_gallus/dna/) دریافت شدند. نتایج حاصل از همترازی به صورت درصد کل توالی همتراز شده، درصد توالی‌های همتراز شده مختص به یک جایگاه و درصد توالی‌های همتراز شده به بیش از یک جایگاه برای هر نرم‌افزار گزارش شده است.

نتایج

کنترل کیفیت داده‌ها با استفاده از نرم‌افزار FastQC انجام شد. خلاصه نتایج سنجش کیفیت در شکل ۱ ارائه شده است. معمولاً، هنگام ویرایش، نوکلئوتیدهای بی‌کیفیت حذف می‌شوند. نمونه‌ها در مطالعه حاضر از کیفیت خوبی برخوردار بودند. اطلاعات آماری خوانش‌های مربوط به نمونه‌های مختلف در جدول (۱) آورده شده‌اند. میانگین محتوای GC خوانش‌ها برای دو گروه کنترل و تیمار به‌ترتیب ۵۰/۵ و ٤۹ درصد به‌دست آمد که با میانگین گزارش‌های قبلی مطابقت دارد. طول همه خوانش‌ها در همه نمونه‌ها ۱۵۰ جفت بازنوکلئوتیدی بود و میانگین تعداد نوکلئوتیدهای خوانده نشده (N) نیز برای همه خوانش‌ها صفر بود آداپتورهای مورد استفاده برای همه خوانش‌ها از نوع آداپتورهای عمومی ایلومینا تشخیص داده شد. آداپتورها در مرحله ویرایش در خوانش‌ها شناسایی و حذف می‌شوند. در این مطالعه آداپتورها از خوانش‌ها به‌طور کامل حذف شده بودند. پس از ویرایش نیز یکبار دیگر سنجش کیفیت برای اطمینان از بهبود کیفیت داده‌ها انجام شد. جدول ۲ نتایج حاصل از همترازی توالی‌های ترنسکریپتومی بر روی ژنوم مرجع با استفاده از دو نرم‌افزار TopHat2 و HISAT2 را نشان می‌دهد. با توجه به نتایج ارائه شده، بیشترین میزان همترازی بر روی ژنوم مرجع مربوط به نرم‌افزار HISAT2 بوده (۹۳/٦۸۷) که اختلاف معنی‌داری نسبت به نرم‌افزار TopHat2 دارد (P<۰/۰۵) که می‌تواند در آنالیز نواحی غیر کدکننده miRNA مفید باشد. سهم نرم‌افزار TopHat2 در همترازی بر روی ژنوم مرجع ۸۵/۸۵ است که میزان قابل توجهی است. همچنین، بیشترین میزان همترازی به یک جایگاه نیز مربوط به HISAT2 می‌باشد (P<۰/۰۵) که نشان می‌دهد توانسته است عمل همترازی بر روی ژنوم مرجع را با اختصاصیت بیشتری انجام دهد که می‌تواند بر نتایج حاصل از

[Downloaded from mg.genetics.ir on 2025-07-17]

آنالیزهای بعدی تأثیرگذار باشد. نتایج این پژوهش نشان داد، درصد همترازی به چند جایگاه در HISAT2 پایین‌تر از نتایج حاصل از نرم‌افزار TopHat2 بوده و این اختلاف معنی‌دار نبوده است (P>۰/۰۵).

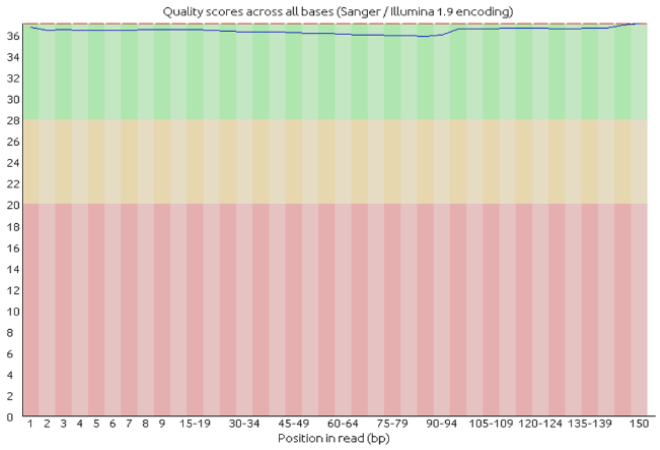
با توجه به نتایج ارائه شده در شکل ۲، تنها ۵ درصد از کل خوانش‌ها در نرم‌افزار HISAT2 تراز چندگانه داشته و به بیش از یک جایگاه روی نقشه ژنوم مرجع متصل شده است که برای نرم‌افزار TopHat2 این میزان ٦/٥ درصد بوده است که نشان‌دهنده اختصاصیت بالای نرم‌افزار HISAT2 در ارتباط با همترازی خوانش‌های حاصل از داده‌های RNA-Seq با ژنوم مرجع است.

بحث

در این پژوهش، ما عملکرد دو ابزار محبوب همترازی RNA-Seq را مقایسه کردیم. از آنجایی‌که، همترازی اولین مرحله در مسیر RNA-Seq است، پس از پیش‌پردازش داده‌ها (بررسی کیفیت، حذف توالی آداپتور، برش خواندن و ...)، بسیار مهم است که یک ابزار همترازی در شرایط مختلف به خوبی عمل کند. زیرا، صحت تحلیل و استنتاج مراحل بعدی به شدت به همترازی دقیق بستگی دارد. تاکنون مطالعات زیادی جهت مقایسه نرم‌افزارهای مختلف همترازی انجام شده که نتایج متنوعی حاصل شده است. نتایج یک مطالعه نشان داد، HISAT2 مستعد خوانش نامناسب در مکان‌های ژنومی حاوی توالی‌های تکراری بود و STAR همترازی‌های دقیق‌تری ایجاد کرد (Raplee et al. 2019).

جدول۱- اطلاعات آماری خوانش‌ها در نمونه‌های مورد مطالعه

نمونه جنین مرغ بومی				
خوانش‌ها	کنترل		تیمار	
	نمونه ۱	نمونه ۲	نمونه ۱	نمونه ۲
تعداد کل خوانش‌ها	۲۵۲۷۱٦۷٦	۲۵۳۳٤٤۹۹	۱۷۱۵۵۰۳۹	۲۲۳۲۵۷۷۰
طول خوانش	۱۵۰	۱۵۰	۱۵۰	۱۵۰
درصد GC	۵۱	۵۰	۵۰	٤۸



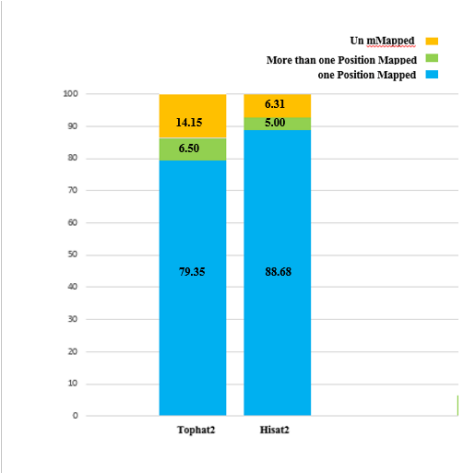
شکل ۱- نمودار گرافیکی کنترل کیفیت به‌وسیله نرم‌افزار FastQC

جدول ۲- درصد همترازی کل، همترازی به یک جایگاه اختصاصی و همترازی به بیش از یک جایگاه در نرم‌افزار TopHat2 و HISAT2

نرم‌افزار	تعداد نمونه	درصد همترازی کل	درصد همترازی به یک جایگاه	درصد همترازی به چند جایگاه
TopHat2	٤	۸۵/۸۵ [b]	۷۹/۳۵ [b]	٦/٥
HISAT2	٤	۹۳/٦۸۷ [a]	۸۸/٦۸ [a]	۵/۰۰۷
SEM		۰/٤٤	۰/۸۷	۰/۸

اعداد هر ستون که دارای حروف غیر مشابه می‌باشند تفاوت معنی‌داری باهم دارند (P<۰/۰۵).

[Downloaded from mg.genetics.ir on 2025-07-17]



شکل ۲- مقایسه دقت همترازی خوانش‌ها در نرم‌افزارهای HISAT2 و TopHat2.

پژوهشی با هدف مقایسه دو ابزار همترازی محبوب HISAT2 و STAR2 انجام شده است، ابزارهای همترازی مختلف به نتایج متفاوتی از نظر زمان محاسباتی، تعداد خوانش‌های تراز شده و تعداد ژن‌های بیان شده متفاوت منجر شدند. میزان کلی همترازی بین دو نرم‌افزار مشابه بود، به‌طوری‌که برای HISAT2، ۹۸/۰۳ و برای STAR2 مقدار آن ۹۸/۷۸ گزارش شده است. اگرچه تفاوت در نرخ همترازی کلی نسبتاً کم بوده، تفاوت در تعداد خوانش‌های همترازی منحصربه‌فرد قابل توجه‌تر است. HISAT2 و STAR2 به‌ترتیب میزان ۸۰/٤۷ و ۸۱/٦٦ درصد از خوانش‌ها را به‌صورت همترازی به یک جایگاه تراز کردند، توجه به این نکته مهم است که این تفاوت از نظر تعداد خوانش‌های همترازی منحصربه‌فرد می‌تواند پیامدهایی برای مراحل بعدی در مسیر یا در نتایج بیولوژیکی نهایی داشته باشد و باید هنگام تفسیر آن‌ها در نظر گرفته شود. می‌خواهیم تأکید کنیم که برای HISAT2 میزان بالاتر همترازی به چند جایگاه گزارش شده است (۱۷/۱۲ درصد برای HISAT2 در مقایسه با ۱۵/۹ درصد برای HISAT2) (Bianchi and Marco 2023). از آنجاییکه، همترازی به چند جایگاه می‌تواند منجر به گزارش نادرست شود و بر تحلیل‌های مراحل بعدی مانند تجزیه و تحلیل تفاوت بیان تأثیر بگذارد (Deschamps-Francoeur et al. 2020)، تصحیح این موضوع به‌طور بالقوه می‌تواند منجر به شناسایی ژن‌های بیشتری توسط HISAT2 شود. این می‌تواند توضیحی ممکن برای این باشد که چرا HISAT2 ژن‌های با بیان متفاوت‌تری نسبت به STAR2 را شناسایی می‌کند. پژوهشگران نتایج مشابهی را با استفاده از داده‌های شبیه‌سازی شده ژنوم انسان حاصل نمودند، به‌طوری که، TopHat2 درصد بیشتری از خوانش‌ها را روی ژنوم مرجع ترسیم می‌کند. همچنین، درصد همترازی به‌درستی بالاتر از سایر نرم‌افزارها است، که می‌تواند در گرفتن مناطق غیر کد کننده مانند miRNA ها و سایر RNA های غیر کد کننده مفید باشد، همچنین، گزارش کردند که نرم‌افزار HISAT2 دارای حساسیت بیشتری در خصوص همترازی خوانش‌ها بوده و از حافظه کمتری نسبت به نرم‌افزارهای دیگر استفاده می‌کند. علاوه بر این دارای دقت و صحت بالاتری برای همترازی به ژنوم مرجع می‌باشد. لذا، پیشنهاد کرد که برای همترازی آنالیزهای RNA-Seq بر روی ژنوم مرجع از HISAT2 استفاده شود (Bahrami et al. 2020). از دیگر مزایای HISAT2 می‌توان به توانایی آن در مدیریت خوانش‌های پیچیده و چندگانه اشاره کرد. با این حال، برخی مطالعات نشان داده‌اند که TopHat2 ممکن است در تحلیل‌های خاصی مانند تجزیه و تحلیل‌های مربوط به ادغام ژن‌ها بهتر عمل کند. علاوه بر این، عملکرد کلی تراز کننده STAR نسبت به سایر ترازها برتر بود که با دقت کلی در شرایط مختلف آزمایش به بیش از ۹۰ درصد رسید (Cox et al. 2024). اندازه‌گیری دقت سطح خوانش باز نوکلئوتیدی، ارزشمندی را در مورد کارایی هر تراز کننده در همترازی خوانش‌های RNA-Seq بر روی ژنوم مرجع به شیوه‌ای آگاهانه ارائه می‌کند. تشخیص اتصال بین یک اگزون و یک اینترون اجازه می‌دهد تا همترازی اتصالات کارآمد و از طریق دقت اتصال برای هر تراز کننده اندازه‌گیری شود. دقت سطح پایه برای تطابق

[Downloaded from mg.genetics.ir on 2025-07-17]

خوانش‌ها با مکان‌های مبدا آن‌ها در ژنوم به حساب می‌آید (Coxe et al. 2024). مقایسه برخی نرم‌افزارهای مختلف همترازی داده‌های حاصل از تعیین توالی RNA از جمله: TopHat2, HISAT2 و Bowtie2 روی ژنوم مرجع نشان داد که بیشترین همترازی صورت گرفته روی ژنوم مرجع گاو با استفاده از نرم‌افزار TopHat2 بوده است. از کل خوانش‌های موجود، نرم‌افزارهای TopHat2 و HISAT2 به‌ترتیب ۹٤/۱۹۷ و ۹۲/۵۲٦ درصد را روی ژنوم مرجع همتراز نمودند. نرم‌افزار HISAT2 عملکرد اختصاصی بیشتری داشت و ۸۹/۲۰۲ درصد از داده‌ها را به یک جایگاه اختصاصی همتراز کرد درصورتی که این پارامتر در خصوص نرم‌افزار TopHat2 به میزان ۸۷/۸۱۲ درصد از کل توالی‌ها بود. از کل توالی‌های به کار گرفته شده تنها ۳/۳۲٤ درصد توسط HISAT2 و ٦/۳۸۵ درصد توسط TopHat2 با بیش از یک جایگاه ژنوم مرجع همتراز شدند. نرم‌افزار Bowtie2 در مقایسه با دو نرم‌افزار دیگر عملکرد پایینی داشت. نتایج حاصل از تحقیق آن‌ها به وضوح نشان داد که همترازی صحیح با ژنوم مرجع نقش بزرگی در نتایج تجزیه و تحلیل بیوانفورماتیکی، به‌ویژه برای شناسایی بیان ژن‌های افتراقی دارد. این نتایج نشان می‌دهد که یک برنامه همترازی مشخص را نمی‌توان به‌طور عمومی برای مجموعه داده‌های RNA-seq اعمال کرد. در حالی‌که، پیشرفت‌های قابل توجه در فناوری توالی‌یابی، طول خوانش‌های نوکلئوتید خروجی را به بیش از ۳۰۰ نوکلئوتید افزایش داده است، ممکن است دقت همترازی افزایش پیدا کند. اما، نتایج حاصل از این تحقیق بر اهمیت تأثیر ابزارهای بیوانفورماتیک برای همترازی توالی‌ها که هدف این مطالعه بود، تأکید می‌کند. پس، پیشنهاد می‌شود که جهت انجام آنالیزهای پایین دستی صحیح، داده‌های ترانسکریپتوم توسط نرم‌افزار HISAT2 با ژنوم مرجع همتراز شوند (Elyasi Zarringhabaie et al. 2023). همچنین، نتایج مقایسه تراز کننده‌های TopHat2، STAR و HISAT2 نشان داد، TopHat2 پایین‌ترین عملکرد را داشته، در حالیکه، سایر ترازکننده‌ها به پوشش رونوشت ۹۰ درصد یا بیشتر دست یافتند. علاوه بر این، بیشترین تعداد خوانش‌های همتراز نشده، به خروجی TopHat2 تعلق داشتند. با توجه به اینکه پشتیبانی TopHat2 قطع شده و HISAT2 جایگزین آن شده، جای تعجب نیست که در مطالعات قبلی و همچنین معیارهای ما کمترین عملکرد را داشته باشد. HISAT2 از یک طرح نمایه‌سازی سلسله مراتبی استفاده می‌کند و از هم‌ترازی محلی برای افزایش تشخیص محل اتصال استفاده می‌کند. این رویکرد ممکن است درجه بالاتری از حساسیت و دقت را در تشخیص اتصالات با هم در حضور تغییرات ژنومی ارائه دهد (Musich et al. 2021).

نتیجه‌گیری کلی

نتایج حاصل از این پژوهش به وضوح نشان داد که همترازی صحیح با ژنوم مرجع نقش مهمی در نتایج تجزیه تحلیل بیوانفورماتیکی، به‌ویژه برای شناسایی بیان ژن‌های افتراقی دارد. اکثر محققان احتمالاً از نرم‌افزارهای همترازی در تنظیمات پارامتر پیش فرض خود برای هم‌ترازی داده‌های RNA-Seq استفاده می‌کنند. نتایج ما نقاط قوت و ضعف الگوریتم‌های هم‌ترازی مختلف را در تنظیمات پیش‌فرض و سایر تنظیمات نشان می‌دهد. این مهم است زیرا، نتایج مراحل بعدی به همترازی قوی بستگی دارد. مطالعات مقایسه‌ای، به کاربران درباره ابزارهای هم‌ترازی که برای همترازی در شرایط خاص قوی و بهینه هستند، اطلاع می‌دهد.در نهایت، در بررسی نتایج این پژوهش، انتخاب بین این دو نرم‌افزار بستگی به نیازهای خاص پروژه و نوع داده‌های مورد استفاده دارد. اگرچه HISAT2 به‌عنوان گزینه‌ای پیشرفته‌تر شناخته می‌شود، TopHat2 نیز هنوز می‌تواند در شرایط خاص مفید باشد. محققان با درک توانایی آن در تجزیه و تحلیل مجموعه داده‌ها، تفسیر بهتری از نتایج خود خواهند داشت. نتایج تجزیه و تحلیل آماری این تحقیق می‌تواند راهنمای خوبی برای محققین استفاده کننده از این نرم‌افزارها باشد.

منابع

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online: http://bioinformatics.Babraham.ac. uk/projects/fastqc.

Bahrami A (2020) Which aligner software is the best for our study. Genetics Research. 7:048. DOI: 10.23937/2378-3648/1410048.

[Downloaded from mg.genetics.ir on 2025-07-17]

Bankizadeh F, Roshanfekr H, Banabazi M, Nazari M, Safari, R (2022). Effect of spliceosome gene complex underlying the transition between the laying and brooding phases in pituitary transcriptome profile of breeder turkey. Veterinary Research & Biological Products 35:40-47. doi: 10.22092/vj.2021.353405.1814.

bankizadeh, F, roshanfekr H, Banabazi MH, Nazari M (2021). Identify protein processing in endoplasmic reticulum pathway as an effective factor in regulating broodiness of. Research Journal of Livestock Science 34:217-228. doi: 10.22092/asj.2020.352593.2113.

Bianchi A, Di Marco A, Pellegrini C (2023) Comparing HISAT and STAR-based pipelines for RNA-Seq Data Analysis: a real experience. In 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS). (pp. 218-224). IEEE. DOI: 10.1109/CBMS58004.2023.00220

Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120. https://doi.org/10.1093/bioinformatics/btu170.

Coxe T, Burks, D J, Singh U, Mittler R, Azad RK (2024) Benchmarking RNA-Seq Aligners at Base-Level and Junction Base-Level Resolution Using the Arabidopsis thaliana Genome. Plants 13:582. https://doi.org/10.3390/plants13050582.

Deschamps-Francoeur G, Simoneau J, Scott MS (2020) Handling multi-mapped reads in RNA-seq. Computational and Structural Biotechnology Journal. 18, 1569-1576. https://doi.org/10.1016/j.csbj.2020.06.014.

Dharshini SA P, Taguchi YH, Gromiha M M (2020) Identifying suitable tools for variant detection and differential gene expression using RNA-seq data. Genomics 112:2166-2172. https://doi.org/10.1016/j.ygeno.2019.12.011.

Dobin A, Davis C A, Schlesinger F, Drenkow J, Zaleski C, Jha S, Gingeras T R (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21. https://doi.org/10.1093/bioinformatics/bts635.

Elyasi Zarringhabaie G, Sadeghi, M, Miraie Ashtiani S R (2023) Comparison of some Alignment Software in the Analysis of Dairy Cows RNA-Seq Data. Research on Animal Production 14:131-138. doi:10.61186/rap.14.39.131. (In Farsi)

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast-spliced aligner with low memory requirements. Nature Methods. 2015; 12(4), 357-360. doi:10.1038/nmeth.3317.

Kim D, Paggi JM, Park C, Bennett C, Salzberg S L (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology 37:907-915. https://doi.org/10.1038/s41587-019-0201-4.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14:1-13. http://genomebiology.com/2013/14/4/R36.

Langmead B, Salzberg S L (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods. 9:357-359. doi:10.1038/nmeth.1923.

Marguerat S, Bähler J (2010) RNA-seq: from technology to biology. Cellular and Molecular Life Sciences 67:569-579. DOI 10.1007/s00018-009-0180-6.

Musich R, Cadle-Davidson L, Osier M V (2021) Comparison of short-read sequence aligners indicates strengths and weaknesses for biologists to consider. Frontiers in Plant Science 12:657240. https://doi.org/10.3389/fpls.2021.657240.

Ramya L, Swathi D, Archana SS, Lavanya M, Parthipan S, Selvaraju S (2021) Establishment of bioinformatics pipeline for deciphering the biological complexities of fragmented sperm transcriptome. Analytical Biochemistry 620:114141. https://doi.org/10.1016/j.ab.2021.114141.

Raplee ID, Evsikov AV, Marín de Evsikova C (2019) Aligning the aligners: Comparison of RNA sequencing data alignment and gene expression quantification tools for clinical breast cancer research. Journal of Personalized Medicine. 2019 9:18. https://doi.org/10.3390/jpm9020018.

Rostami S, Beigi Nassiri MT, Nazari M, Tabatabaei Vakili S (2019) The effect of different levels of soybean lecithin on semen quality and sex ratio of Holstien bull sperm by Real-time qPCR technique. Cell and Tissue Journal 10:133-142. doi: 10.52547/JCT.10.3.133 (in farsi).

Westermann AJ, Vogel J (2021) Cross-species RNA-seq for deciphering host–microbe interactions. Nature Reviews Genetics 22:361-378. DOI 10.1038/s41576-021-00326-y.

[Downloaded from mg.genetics.ir on 2025-07-17]