ترجمه فا

TarjomeFa.Com

ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

بازسازی توالی رونوشت نوپدید (دینوو) از توالی RNA با استفاده از پلتفرم ترینیتی برای تولید و تحلیل مرجع

چکیده :

ترکیب جدید داده های توالی یابی دی ان ای به ما امکان مطالعه ترانسکریپتوم ها را بدون نیاز به یک توالی ژنومی نظیر ارگانیسم های غیر مدل با اهمیت تکاملیو اکولوژیکی، نمونه های سرطانی یا میکروبی می دهد. در این پروتوکل، ما به توصیف کاربرد پلتفرم ترینیتی برای ترکیب ترانسکریپتوم جدید از داده های توالی یابی دی ان ای در موجودات غیر مدل می پردازیم. هم چنین ما اجزای دیگر مربوط به ترینیتی را برای کاربرد های پایین دست از جمله RSEM و براورد فراوانی رونوشت، و بسته های بیوکنداکتور/R برای شناسایی رونوشت های بیان شده در نمونه ها و رویکرد ها برای شناسایی ژن های کد گذاری پروتین ارایه می کنیم. در یک راهنمای اموزشی، یک جریان کاری برای تحلیل ترانسکریپت وابسته به ژنوم برای پلتفرم ترینیتی ارایه شده است. نرم افزار، اسناد و مثال ها نیز از trinityrnaseq.sf.net قابل دسترس هستند.

مقدمه

توالی یابی پر بازده ژنوم ها(توالی یابی-DNA) و ترانسکریپتوم ها(توالی یابی –RNA)زمینه را برای مطالعه اطلاعات ژنتیکی و کارکردی ذخیره شده درون هر موجود با یک مقیاس و سرعت بی سابقه هموار کرده است. برای مثال توالی یابی –RNA در اصل امکان مطالعه همزمان ساختار رونوشت(نظیر اتصال جایگزین)، اطلاعات آللی(برای مثال SNP ها) و بیان با وضوح و دامنه پویایی بالا را در اختیار گذاشته است. این پیشرفت ها به نوبه خود به طور قابل توجهی منجر به تسهیل تحقیقات ژنومیک کارکردی در گونه هایی شده است که منابع ژنتیکی یا مالی آن ها محدود است از جمله بسیاری از موجودات غیر مدل- موجوداتی که اگرچه به طور گسترده ای در یک شرایط تحقیقاتی مطالعه نشده اند، با این وجود از اهمیت تکاملی یا بوم شناسی قابل توجهی برخوردارند. اگرچه بسیاری از برنامه های ژنومی از دیرباز به قابلیت دسترسی به توالی های ژنومی با کیفیت بالا متکی بوده اند، این توالی ها تنها برای بخش کوچکی از موجودات شناخته شده، تعیین شده اند. به علاوه، توالی یابی و گردایش ژنوم هنوز در بسیاری از موارد یک فرایند پر هزینه به دلیل اندازه ژنوم و محتوی تکرار است. بر عکس، از آن جا که تنها بخشی از ژنوم رونویسی می شود، داده

های توالی یابی RNA می توانند یک مسیر سریع و ارزان تر را- در مقیاس هر آزمایشگاه- برای تعیین یک ترانسکریپتوم مرجع برای کاربرد های داون استریم نظیر هم تراز سازی، فیلوژنتیک(تبارزایی) یا ساخت نشانگر در اختیار بگذارند. در واقع حتی در چارچوب یک پروژه توالی یابی کل ژنوم، توالی یابی RNA به یک منبع مهم از شواهد و اطلاعات برای شناسایی ژن های رونویسی شده و تفسیر ساختار اگزون تبدیل شده است. درک پتانسیل کامل توالی یابی RNA مستلزم روش های محاسباتی ای است که می توانند یک ترانسکریپتوم را گردایش کنند حتی زمانی که توالی یک ژنوم قابل دسترس نباشد. دو روش اصلی برای تبدیل داده های خام توالی یابی RNA به توالی های رونوشت وجود دارد: از طریق هدایت توالی های ژنومی گرادیش یافته یا از طریق تجمع نوپدید. رویکرد ژنومی برای مطالعات ترانسکریپتوم سریعا به یک رویکرد استاندارد برای تحلیل توالی یابی RNA برای موجودات مدل تبدیل شده اند و چندین بسته نرم افزاری برای این منظور وجود دارد. با این حال این رویکرد را نمی توان به موجوداتی که ژنوم کامل و گرداوری شده آن ها وجود ندارد به کار برد و حتی برای موجوداتی که دارای ژنوم های به خوبی گردایش یافته(تجمع یافته) می باشند ممکن است نتایج در نسخه های گرادیش ژنومی، متغیر و متفاوت می باشند. در چنین مواردی، یک همگذار ترانسکریپتوم نوپدید نیاز است. با این حال، فرایند گردایش و تجمع یک ترانسکریپتوم، بسیاری از مفروضاتی که بر اساس آن ها هم گذار هایی برای شرایطی که داده های DNA ژنومی به آن متکی هستند را نقض کرده است. برای مثال، پوشش یکنواخت وپاردایم " یک لوکوس(مکان ژنی)- یک کن تیگ" برای RNA معتبر نیست: یک همگذار ترانسکریپتوم دقیق تولید یک کن تیگ به ازای هر رونوشت مجزا(ایزوفرم) به جای تولید یک کن تیگ به ازای هر مکان ژنی(لوکوس) کرده و رونوشت های مختلف دارای پوشش های متفاوتی است که منعکس کننده سطوح بیان متفاوت آن هاست.

چندین ابزار در حال حاضر برای گردایش و همگذاری نوپدید توالی RNA قابل دسترس می باشد.,Trans-ABySS6 (ترانس-ABySS6)، Velvet-Oases (واحه های مخملی) و ترانس نوپدید SOAP همگی نسخه هایی از همگذار های ژنومی اولیه می باشند. ما قبلا، یک رویکرد جایگزین جدید را برای همگذاری ترانسکریپتوم موسوم به ترینیتی توصیف کردیم. ترینیتی، داده های توالی یابی RNA را به بسیاری از گراف های دی براین(در واقع، یک گراف به ازای هر ژن بیان شده) تقسیم کرده و از محاسبه موازی برای باز سازی رونوشت ها از این گراف ها استفاده می کند از جمله ایزوفرم های اتصال جایگزین. ترینیتی می تواند از کتاب خانه های دو طرفه ایلومینا خاص رشته ای استفاده

کند ولی در عین حال می تواند داده های قرائت تک رشته ای و غیر مرتبط با رشته را نیز شامل شود.ترینیتی رونوشت ها را دقیقا با یک رابط ساده و شهودی که نیازی به تعدیل پارامتر ندارد یا نیازمند تعدیل پارامتر اندکی است باز سازی می کند. چندین مطالعه مستقل اثبات کرده اند که ترینیتی در مقایسه با روش های جایگزین بسیار موثر است(برای مثال منابع 9-11، چالش اتصال جایگزین پروژه DREAM، http://www.thedream-project.org/result/alternative-splicing). تعداد زیادی از استنادهایی که گرابهر و همکاران در یک مدت زمان نسبتا کم جمع آوری کرده اند(از زمان انتشار آنلاین آن در می 2011) موید عملکرد بهتر و مطلوبیت ترینیتی می باشند. کاربران ترینیتی طیف وسیعی از موجودات مدل و غیر مدل را از همه فرمانرو ها مطالعه کرده و متعلق به آزمایشگاه های کوچک و پروژه های ژنومی بزرگ(برای مثال تفسیر ژنوم شته نخود نسخه Fabrice Legeai:2 (موسسه ملی تحقیقات زراعی(INRA) و ترنس مورفی(مرکز ملی اطلاعات بیوتکنولوژیRefSeq (NCBI)، ارتباطات شخصی با A.P). هم چنین ترینیتی دارای یک انجمن توسعه دهنده فعال می باشد که به شدت عملکرد و مطلوبیت آن را بهبود داده است(به http://trinityrnaseq. sourceforge.net مراجعه کنید). برای مثال، اگرچه عملکرد زمان اجرای اولین نسخه از نظر محاسباتی کارامد نبود، با این حال جامعه توسعه دهنده ترینیتی از آن زمان به بعد کارایی خود را توسعه داده، نیاز به حافظه را تا نصف کاهش داده و سرعت پردازش را از طریق افزایش موازی موازی سازی و الگوریتم های پیشرفته افزایش داده اند. به علاوه، ترینیتی به یک پلتفرم مدولار تبدیل شده است که به طور یکپارچه از ابزار های شخص ثالث نظیر Jellyfish برای ایجاد کاتالوگ*k*-mer اولیه استفاده می کنند. سایر ابزار های شخص ثالث تلفیق شده در ترینیتی، مطلوبیت ترانسکریپتوم های باز سازی شده آن را بهبود داده است. برای مثال، ترینیتی در حاصل حاضر از ابزار هایی نظیر RSEM, edgeR وDESeq پشتیبانی می کند که این ابزار ها رونوشت های خروجی را گرفته و آن ها را برای آزمون فرضیات مختلف تصحیح می کنند. با توجه به محبوبیت ترینیتی و پیشرفت های قابل توجه از زمان انتشار آن، ارایه یک سری شیوه های دقیق که ویژگی های مختلف آن ها را پوشش دهد بسیار مهم است. شیوه های ارایه شده در این جا کاربرد ترینیتی را برای مطالعات در موجودات غیر مدل توسعه می دهند.

مروری اجمالی بر همگذار توالی یابی RNA ترینیتی

خط لوله همگذار ترینیتی متشکل از سه ماژول متوالی می باشد:Inchworm, Chrysalis and Butterfly (شکل 1). ما به شدت کاربران را تشویق می کنیم تا ابتدا اولین انتشارات ترینیتی را برای توصیف گسترده روشی که اکنون به طور مختصر در مورد آن صحبت می کنیم بخوانند. اولا، همه k-mers هم پوشان از قرائت های توالی یابی DNA استخراج می شوند. سپس Inchworm هر *k*-mer منحصر به فرد را به ترتیب فراوانی نزولی بررسی کرده و سپس کن تیگ های رونوشت را با استفاده از یک اکستنشن حریص بر اساس هم پوشان های (*k*-1)-mer بررسی می کند.Inchworm اغلب تولید رونوشت های طول کامل برای یک ایزوفرم غالب می کنند با این حال تنها بخش های منحصر به فردی از رونوشت های اتصال یافته جایگزین را گزارش می کند. این روش عملیاتی برای مجموعه داده هایی که به طور کلی فاقد توالی های تکراری می باشند نظیر ترانسکریپتوم ها عملکرد هوبی دارد. سپس، Chrysalis ابتدا کن تیگ های Inchworm مربوطه را به اجزای آن با استفاده از قرائت های خاک برای دسته بندی رونوشت ها بر اساس پشتیبانی قرائت مشترک و لینک های قرائت جفتی در صورت وجود خوشه بندی کنند. این فرایند، مناطقی که احتمالا از رونوشت های اتصالی جایگزین یا خانواده های ژنی نزدیک یا خویشاوند ناشی می شوند را خوشه بندی می کند. سپس Chrysalis پیچیدگی ساختاری کن تیگ های Inchworm خوشه بندی شده را با ایجاد یک گراف دی براین برای هر خوشه کد گذاری کرده و سپس قرائت ها را در میان خوشه ها، تقسیم بندی می کند. پارتیشن بندی کن تیگ های Inchworm و قرائت های توالی یابی RNA به خوشه های مجزا(اجزا) امکان پردازش موازی انبوه محاسبات بعدی را می دهد. در نهایت،Butterfly گراف های فردی را به طور موازی پردازش کرده و در نهایت رونوشت های طول کامل را برای ایزوفرم های متصل شده جایگزین گزارش کرده و رونوشت هایی که متناظر با ژن های پراساخت هستند را از هم جدا می کند. باترفلای قرائت های توالی یابی RNA را از طریق گراف رد یابی کرده و ارتباط را بر اساس توالی قرائت و نیز شواهد مربوط به داده های دو طرفه موجود تعیین می کند. وقتی که نتوان ارتباطات را با قرائت های ردیابی شده تایید کرد، باترفلای، گراف را به چندین زیر گراف منفصل تقسیم کرده و هر یک را به طورجداگانه پردازش می کند. سپس باترفلای مسیر های گراف پشتیبانی شده را طی کرده و توالی های رونویسی را طوری باز سازی می کند که منعکس کننده مولکول های cDNA اصلی باشند.

بسته تجزیه تحلیل ترانسکریپتوم برای موجودات غیر مدل

تولید یک مجموعه توالی یابی RNA نوپدید تنها نخستین گام به سمت تحلیل ترانسکریپتوم می باشد. اهداف مشترک برای مطالعه ترانسکریپتوم در هر دو موجودات مدل و غیر مدل شامل شناسایی رونوشت ها، مشخص کردن پیچیدگی ساختاری رونوشت و کد گذاری محتوا و درک این که کدام ژن ها و ایزوفرم ها در نمونه های مختلف(بافت ها، شرایط محیطی و از این قبیل موارد) بیان می شوند می باشد. ترینیتی از دست یابی به این اهداف با در نظر گرفتن نرم افزار محبوب دیگری که احتمال در محیط بیو انفورماتیک نصب شده است، با استفاده از نرم افزار متن باز دیگر به عنوان اجزای پلاگین که مستقیما در محیط نرم افزاری ترینیتی قرار گرفته است و با ارایه اسکریپت های ساده که هدف آن ارایه یک رابط آشنا و کاربر دوست برای ماژول های تجزیه تحلیل پیچیده پشتیبانی می کند. نتایج اغلب به صورت فایل های با قالب جدولی ارایه می شوند که کاربران می توانند آن ها را وارد برنامه اکسل مورد نظر خود کرده و یا نمودار ها یا گراف ها را با فرمت PDF بسازند. در زیر ما به توصیف جزییات تک تک مراحل پروتوکل و شناسایی کاربرد های فعلی نرم افزار پشتیبانی شده می پردازیم.

مقایسه ترانسکریپتوم ها در نمونه ها

در بسیاری از موارد، یک کاربر خواهان مقایسه یک نوع یا سطحی از رونوشت ها بین نمونه ها برای مثال برای ژن هایی که به طور متفاوت بیان می شوند می باشد. دو مسیر احتمالی برای دست یابی به این هدف وجود دارد. دو مسیر محتمل برای دست یابی به این هدف وجود دارد. یک گزینه، تجمیع و گردایش جداگانه قرائت های متناظر با هر یک از انواع نمونه و سپس مقایسه نتایج از هر یک از مجموعه ها و گردایش ها. با این حال محدودیت این رویکرد، نیاز به تطبیق رونوشت های" یکسان و مشابه" حاصل از مجموعه های مستقل است. یک رویکرد جایگزین ساده تر و یا رویکردی که ما آن را توصیه می کنیم، ابتدا ترکیب همه قرائت ها در همه نمونه ها و تکرار های بیولوژیکی به یک مجموعه داده توالی RNA، تجمیع قرائت ها برای تولید یک مجموعه ترینیتی مرجع (شکل 7) و سپس شناسایی سطح هر یک از این رونوشت ها در هر نمونه با هم تراز سازی هر یک از قرائت های نمونه(قرائت های نرمال نشده) به مجموعه ترانسکریپتوم مرجع و شمارش تعداد قرائت هایی که با هر رونوشت هم تراز می باشد است. در نهایت، تست های آماری برای مقایسه تعداد قرائت های مشاهده شده برای هر رونوشت در نمونه های مختلف و رونوشت های مشاهده شده ای که دارای شکل کاملا متفاوتی از قرائت ها در نمونه ها می باشند گزارش

می شوند. تجزیه تحلیل بیشتر رونوشت های بیان شده متفاوت قادر به آشکار سازی الگوهای بیان ژن و ارایه اطلاعات در خصوص روابط میان نمونه های بررسی شده می باشد.

برآورد فراوانی رونوشت

اندازه گیری کمی رونوشت، پیش نیاز و لازمه بسیاری از تحقیقات و پژوهش های فرودست یا پایین دست می باشد. چندین شاخص برای اندازه گیری سطوح فراوانی رونوشت بر اساس داده های توالی یابی RNA پیشنهاد شده اند که عمق توالی یابی و طول رونوشت ها را نرمال سازی می کنند. این شاخص ها شامل قرائت ها در هر کیلوباز از طول رونوشت هدف در میلیون قرائت نقشه یابی شده(RPKM) برای توالی های یک طرفه و یک محاسبه آنالوگ بر اساس شمارش کل قطعات(EPKM) برای داده های توالی یابی RNA دو طرفه می باشند. به منظور محاسبه تعداد قرائت های توالی یابی RNA یا قطعاتی که از رونوشت ها گرفته شده اند، قرائت ها بایستی ابتدا با رونوشت ها هم تراز سازی شوند. وقتی که شما با یک ژنون مرجع و یک ترانسکریپتوم تفسیر شده کار می کنید، قرائت ها معمولا با یک یا هر دوی آن ها هم تراز سازی می شوند. در شرایط گردایش و همگذاری دنیوو(نوپدید)، قرائت ه به رونوشت های هم گذاری شده مجددا هم تراز سازی می شوند. با این حال، ایزوفرم های متصل شده جایگزین و ژن های جدیدا رونویسی شده می توانند توالی های فرعی را به اشتراک بگذارند که طولانی تر از یک جفت قرائت یا تک قرائت باشند و این قرائت ها به طور مساوی به اهداف چندگانه متصل می شوند.چندین روش اخیرا برای براورد شیوه تخصیص صحیح این قرائت ها به رونوشت ها ارایه شده اند به طوری که بهترین تقریب از سطوح بیان واقعی رونوشت ها صورت گیرد. یکی از آن ها نرم افزار RSEM (توالی یابی RNA با بیشینه سازی امید ریاضی) می باشد که از یک فرایند تکراری برای تخصیص قرائت ها به هر رونوشت بر اساس احتمالات مربوط به اینکه قرائت ها متعلق به یک رونوشت باشند(شکل 8) استفاده می کند(شکل 8) و اریبی های مرتبه ای ایجاد شده توسط پروتکل های تولید کتابخانه توالی یابی RNA را در نظر می گیرد.RESM با توزیع نرم افزار ترینیتی همراه است. نرم افزار RESM امروزه نیازمند همسو سازی قرائت های توالی یابی RMA بدون فاصله به رونوشت ها ی باز سازی شده ترینیتی نظیر هم تراز های تولید شده توسط نرم افزار بوتای می باشد. با توجه به رونوشت های تجمیع شده ترینیتی و قرائت های توالی یابی RNA تولید شده از یک نمونه، RSEM مستقیما بوتای را برای هم تراز سازی قرائت ها با رونوشت های ترینیتی اجرا کرده و سپس وفور رونوشت را محاسبه کرده و، تعداد قطعات توالی یابی RNA را متناظر با هر

رونوشت ترینیتی از جمله مقادیر بیان نرمال شده به صورت FPKM هم تراز می کند. علاوه بر براورد سطوح بیان رونوشت های فردی، RESM، براورد های سطح ژن را با استفاده از جزء ترینیتی به عنوان یک پروکسی برای ژن محاسبه می کند. برای مقایسه سطوح بیان رونوشت ها یا ژن های مختلف در نمونه ها، یک ترینیتی شامل یک اسکریپت می باشد که از edgeR برای انجام یک نرمال سازی مقیاس بندی TMM اضافی(میانگین پیراسته از مقادیر M) استفاده می کند که هدف آن پوشش دادن تفاوت ها درتولید RNA سلولی در همه نمونه ها می باشد. هر دو رونوشت های ترینیتی باز سازی شده جزیی و طول کامل می توانند برای براورد بیان ژن در مقایسه با براورد سطوح بیان با استفاده از یک تفسیر رونوشت مبتنی بر ژنوم مرجع و با کیفیت بالا مفید باشند. با این حال، رونوشت های باز سازی شده کامل تر، همبستگی بالاتری با سطوح بیان براورد شده برای رونوشت های مرجع دارد.

تجزیه تحلیل رونوشت های با بیان ژن دیفرانسیل

به منظور براورد بیان ژن دیفرانسیلی بین دو نوع نمونه، حداقل سه تکرار بیولوژیکی از هر نمونه بایستی به طور ایده ال بدست بیاید. جمع آوری داده های تکرار امکان تست این که ایا تفاوت های مشاهده شده در بیان تفاوت معنی داری از تغییرات بیولوژیکی مورد انتظار تحت فرضیه صفر مبنی بر این که رونوشت ها به طور متفاوت بیان نمی شوند(بیان دیفرانسیلی ندارند) را می دهند. در نبود تکرار های بیولوژیکی، هنوز امکان شناسایی رونوشت های بیان شده به طور متفاوت با استفاده از مدل های آماری تغییرات مورد انتظار نظیر تحت توزیع دو جمله ای منفی یا پواسون وجود دارد. توزیع پوآسون به خوبی تغییرات مورد انتظار بین تکرار های فنی را مدل سازی می کند و این در حالی است که توزیع دو جمله ای منفی، عملکرد بسیار بهتری در توجیه و پوشش تغییرات زیاد مشاهده شده را بین تکرار های بیولوژیکی می دهد و از این روی یک مدل مطلوب برای شناسایی رونوشت های با بیان دیفرانسیل با ابزار های نرم افزاری پیشرو است. مجموعه های ترانسکیپتوم ترینیتی می توانند بستر های مفیدی برای ارزیابی تغییرات در بیان ژن بین نمونه ها باشند و نتایج عمدتا با مطالعات بر اساس ترانسکریپتوم های مرجع مطابق است. به همین منظور، ما به ابزار های مربوط به پروژه بیوکنداکتور برای شناسایی رونوشت های بیان شده متفاوت از جمله edgeR و DESeq متکی هستیم.

بیوکنداکتور نیازمند نرم افزار R برای محاسبات آماری است که شامل یک محیط خط فرمان و دستور زبان برنامه نویسی می باشد که می تواند یک مانعی برای کاربران جوان و یا آن دسته از افرادی که فاقد آموزش بیو انفورماتیک

گسترده می باشند باشد. به منظور تسهیل استفاده از ابزار بیوکنداکتور برای مطالعات ترانسکریپتوم، نرم افزار ترینیتی شامل یک سری برنامه ها و اسکریپت های ساده ای است که از نرم افزار R برای شناسایی رونوشت های بیان شده متفاوت استفاده می کنند: تولید فایل های خروجی با قالب جدولی که فهرستی از رونوشت های بیان شده دیفرانسیلی که شامل تغییرات چند برابری و مقادیر معنی دار آماری می باشند، و تولید نمودار هایی نظیر نمودار های MA(که در آن M= نسبت های لگاریتمی و A= مقادیر میانگین است)، نمودار های ولکانو، نمودار های همبستگی و نقشه های حرارتی خوشه بندی شده(هیت مپ) در فرمت PDF(به مرور پروتوکل و روش مراجعه کنید)

پیش بینی منطقه کد کننده پروتین و تفسیر کارکردی رونوشت های ترینیتی: اغلب رونوشت های تجمیع شده از داده های توالی یابی RNA یوکاریوتی بر گرفته از RNA پلی آدنیلات انتظار می رود که کد کننده پروتین ها باشند. یک جست و جوی تشابه توالی مثلا با استفاده از BLASTX در برابر توالی های حاصل از گونه های خویشاوند از نظر تبارزایی و به خوبی تفسیر شده، عملی ترین شیوه برای شناسایی رونوشت های کد کننده و پیش بینی وظایف و عمکلرد آن ها است. متاسفانه، این گونه های خویشاوند با تفسیر خوب اغلب برای ترانسکریپتوم های به تازگی هدف یابی شده قابل دسترس نمی باشند. در چنین شرایطی، استفاده از جدید ترین دیتابیس پروتین با کم ترین تکرار (برای مثال

NCBI's 'nr', ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz

یا

Uniprot ftp://ftp.uniprot.org/pub/databases/uniprot/current_
(release/knowledgebase/complete/uniprot_trembl.fasta.gz

یک جایگزین و راه کار مناسب است.

ترانسکریپتوم های جدیدا هدف یابی شده نیز می توانند پروتین هایی را کد گذاری کنند که به طور ناکافی با تشابه قابل تشخیص با پروتین های شناخته شده نشان داده می شوند. پوشش دادن این مناطق کد کننده مستلزم روش هایی است که مناطق کد گذاری را بر اساس شاهص های مرتبط با ترکیب توالی پیش بینی می کنند. یکی از این روش ها TransDecoder می باشد که ما آن را با ترینیتی با کمک به شناسایی مناطق کد کننده بالقوه در رونوشت های باز سازی شده ایجاد کرده ایم. وقتی که TransDecoder بر روی رونوشت های باز سازی شده از ترینیتی اجرا می شود، مناطق کد کننده پروتین کاندید را بر اساس ترکیب نوکلئوتید، طول قالب خواندن باز و

محتوی دامنه Pfam(اختیاری) شناسایی می کند. اگرچه اجرای TransDecoder و تفسیر کارکردی مناطق کد کننده به صورت بخشی از پروتکل موجود پوشش داده نمی شود، اسناد مربوطه در وب سایت ترینیتی در http://trinityrnaseq.sourceforge.net/analysis/extract_proteins_from_trinity_transcripts.html و http://trinityrnaseq.sourceforge.net/annotation/Trinotate.html دیده می شود. رابط های کاربر گرافیکی پویا نظیر IGV یا GenomeView در مطالعه باز سازی های رونوشت بسیار مفید است. اگرچه آن ها در ابتدا برای ژنوم ها طراحی شده اند این رابط ها را می توان به آسانی برای مشاهده تراز های قرائت های مربوط به رونوشت ها، مناطق کد کننده بالقوه، مناطق تشابه توالی پروتین و تراز های قرائتی کوتاه با ارتباط زوجی مورد استفاده قرار داد(برای مثال، http://trinityrnaseq.sourceforge.net/analysis/read_alignment_visualization_QC.html).

محدودیت های رویکرد ترینیتی برای تحلیل ترانسکریپتوم

اگرچه ترینیتی در باز سازی رونوشت ها و ایزوفرم های اتصال جایگزین بسیار موثر می باشند، در نبود یک ژنوم مرجع ، درک کامل مبانی ساختاری برای تغییرات رونوشتی مشاهده شده، نظیر این که آیا آن ها ناشی از یک یا چند اگزون حذف شده، مناطق متصل شده پذیرنده یا دهنده جایگزین یا اینترون های حفظ شده می باشند، اگر غیر ممکن نباشد سخت است. در حال حاضر ما به بررسی راهبرد های ازمایشی و محاسباتی برای دست یابی بهتر به این اطلاعات از داده های توالی یابی RNA در نبود ژنوم های مرجع می پردازیم. قرائت های توالی یابی RNA و اطلاعات مربوط به جفت سازی امکان حل ایزوفرم ها و پارالوگ ها را توسط ترینیتی می دهد با این حال موفقیت ترینیتی منوط به یافتن تغییرات توالی ای است که می توان آن ها را به طور مناسب با قرائت های جداگانه یا از طریق پیوند های جفتی مرحله بندی کرد. تغییرات توالی که نمی توان آن ها را به طور مناسب مرحله بندی کرد، می توانند منجر به شیمر های هطا بین ایزوفرم ها یا پارالوگ هایی می شود که تشخیص آن ها از داده های قرائت کوتاه به تنهایی غیر ممکن است. پیشرفت در فناوری های قرائت طولانی بایستی به حل این چالش ها کمک کنند. هم چنین، اگرچه ترینیتی امروزه تنها رسما از داده های توالی یابی RNA ایلومینا پشتیبانی می کند، تلاش هایی برای بررسی استفاده و کاربرد قرائت های توالی یابی رونوشت تولید شده از فناوری های جایگزین به خصوص رونوشت

های تولید شده از پاسیفیک بیوساینس و یون تورنت[1] در حال انجام است. در نهایت، همانند توالی یابی ژنومی پر بازده، شواهد مربوط به پلی مورفیسم ها را می توان از داده های توالی یابی RNA ایلومینای نقشه یابی شده به مجموعه های ترینیتی جست و جو کرده و نمودار آن ها را ترسیم کرد. با این حال، محققان بایستی در ارزیابی چند شکلی ها در زمینه توالی یابی RNA و داده های تجمیع ترانسکریپتوم نوپدید به شدت محتاط باشند زیرا مجموعه های رونوشتی ناصحیح یا سوء هم تراز سازی ایزوفرم را می توان به آسانی به صورت شواهد مربوط به پلی مورفیسم سوء تفسیر کرد. به ویژه، وقتی که رونوشت ها به طور قابل توجهی بیان شوند، خطا های توالی یابی می توانند منجر به نسخه های بیان شده بالا شوند. تعیین بهترین شیوه ها برای استفاده از SNP ها در مجموعه های ترانسکریپتوم های نوپدید و بررسی بیان های خاص آلل یک زمینه ای است که نیازمند تحقیقات می باشد. در واقع، چون استفاده از نرم افزار بیو انفورماتیک آسان تر است، توسعه و استفاده از بهترین شیوه های ممکن توسط انجمن تحقیقاتی برای اطمینان از این که نتایج بحث انگیز، ناشی از منابع خطای چندگانه نمی باشند مهم است.

بسته های تجزیه تحلیل جایگزین

ابزار های عالی برای مطالعات ترانسکریپتوم در موجوداتی که توالی ژنوم مرجع با کیفیت بالای آن ها قابل دسترس است از جمله آن دسته از ابزار هایی که در نرم افزار توکسدو (Bowtie, TopHat, Cufflinks, Cuffdiff CummeRbund) یا اسکریپچر ارایه شده اند موجود بوده و به فراوانی مورد استفاده قرار می گیرند. نرم افزار های گردایش و تجمیع توالی یابی RNA نوپدید قابل دسترس شامل Oases , SOAPdenovotrans و TransABySS می باشند. نرم افزار eXpress یک الگوریتم فوق العاده کارامد را برای براورد سطوح بیان رونوشت ژن اجرا کرده و از نرم افزار باوتای[2] برای هم تراز های قرائت کوتاه استفاده کرده و یک جایگزین مناسب برای استفاده از RSEM برای براورد اندازه گیری های رونوشت می باشد. روش های جدید برای تحلیل بیان دیفرانسیل بر اساس داده های توالی یابی RNA نیز در حال ظهور می باشند. با حفظ و بهبود نرم افزار ترینیتی و تجزیه تحلیل های داون استریم مربوط به پشتیبانی، هدف ما بررسی اثرات ابزار های جدید و استفاده از مفید ترین آن ها در تجزیه تحلیل های آینده می باشد به طوری که ما کاربران را تشویق به کشف مستقل روش های جایگزین می کنیم. به علاوه، ما از

[1] Pacific Biosciences30 and Ion Torrent31

[2] Bowtie

کاربران می خواهیم تا به بررسی ابزار های پشتیبانی شده موجود از جمله edgeR و RSEM مستقل از استفاده از نرم افزار های ارایه شده توسط ترینیتی بپردازند زیرا آن ها دارای یک سری قابلیت های اضافی می باشند که ممکن است از طریق رابط های کمکی به طور کامل قابل دسترسی نباشند. پیشرفت های آینده در ترینیتی نه تنها از تجمیع و گردایش ترانسکریپتوم نوپدید عاری از ژنوم پشتیبانی می کنند، بلکه قادر به پشتیبانی توالی های ژنومی مرجع و تفاسیر رونوشتی در صورت امکان هستند. علاوه بر ارایه روش های موثر در راستای کمک به تفسیر ژنومی، این پیشرفت ها بایستی توانایی محققان را برای کشف پیچیدگی رونوشتی موجودات مدل به خصوص در مواردی که در آن ژنوم ها تغییر یافته یا بازآرایی می شوند مثلا تحت شرایط سرطان توسعه دهند.

مرور پروتکل

روش زیر به طور مفصل شیوه اجرای ترینیتی را برای تجمیع و هم گذاری یک مرجع ترانسکریپتوم از توالی یابی RNA از نمونه های چند گانه نشان می دهد: براورد سطوح بیان برای هر رونوشت در هر نمونه و در نهایت، شناسایی رونوشت هایی که به طور متفاوت بین نمونه های مختلف بیان می شوند. این پروتکل یک پیمایش را برای برخی از عملیات استاندارد مورد استفاده برای تولید و تحلیل مجموعه های ترینیتی از جمله مجموعه توالی یابی RNA نوپدید، براورد فراوانی و بیان دیفرانسیل ارایه می کند. برای سهولت، کتابخانه های حاصل از تکرار های بیولوژیکی در پروتکل استفاده نمی شوند بلکه توجه داشته باشید که حداقل سه تکرار بیولوژیکی به ازای هر نمونه یا شرایط به منظور تست معنی داری با توجه به تغییرات فناوری و بیولوژیکی نیاز است. همه روش ها و ابزار ها برای جست و جوی مجموعه در وب سایت نرم افزار ترینیتی توصیف می شوند(http://trinityrnaseq. sf.net) و در این جا ما یک مجموعه از این عملیات را ارایه می کنیم که یک تعادل را بین طول این مقاله و نمایش میزان قابلیت ها ایجاد می کند. این پروتکل بر اساس استفاده از نسخه نرم افزاری Trinityrnaseq_r2013-02-25 بوده و مخاطبان و خواننده ها بایستی در نظر داشته باشند که از ان جا که ترینیتی یک نرم افزار پژوهشی روز به روز در حال تکامل است، برخی از پارامتر ها و نام های فایل ممکن است در نسخه های نرم افزاری آینده تغییر کنند. جدید ترین نسخه این روش(راهنما) را می توان در http://trinityrnaseq.sf.net/trinity_rnaseq_ tutorial.html یافت. یک مورد کاربرد معمول برای ترینیتی، بسیار متفاوت از مورد کاربرد توصیف شده در این پروتکل نیست و ما کاربران را

توصیه به اجرای کامل روش قبل از به کار گیری پروتکل بر روی داده های خاص خود می کنیم. به علاوه، قبل از اجرای مراحل توصیف شده در این روش، ما شما را تشویق می کنیم تا ابتدا کل پروتکل را مطالعه کنید.

مواد

تجهیزات

- داده ها(ملزومات بر طبق اهداف آزمایشی شما متغیر هستند)
- سخت افزار(کامپیوتر 64 بیتی با سیستم عامل لینوکس، رم 1 گیگ به ازای هر 1M PE قرائت)

ترینیتی نسخه trinityrnaseq_r2013-02-25: http://trinityrnaseq.sourceforge.net

نسخه باوتای 0.12.9: http://bowtie-bio.sourceforge.net (RSEM امروزه با باوتای 2 سازگار نیست از این روی اطمینان حاصل کنید که آخرین نسخه برای باوتای 1 را که اکنون v. 0.12.9 می باشد که در 16 دسامبر 2012 منتشر شده است را دریافت می کنید.).

سامتولز نسخه

http://sourceforge.net/projects/samtools/files/ samtools/0.1.18

نسخه

http://www.r-project.org: 2.15 ،R

نسخه بلاست+

ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/: LATEST2.2.27/

اطمینان حاصل کنید که هر یک از ابزار های نرم افزاری نصب شده فوق(به جز ترینیتی) در تنظیمات PATH لینوکس شما وجود دارند. برای مثال اگر شما دارای ابزار هایی هستید که در دایرکتوری /usr/local/ tools ، نصب شده اند، شما می توانید تنظیمات PATH خود را برای استفاده از این دایرکتوری در مسیر جست و جو از طریق

% export PATH=/usr/local/tools:$PATH به روز رسانی کنید.

راه اندازی تجهیزات

به طور دلخواه و تنها برای سهولت، یک متغیر محیطی TRINITY_HOME را جایگزین کرده و '/software/trinityrnaseq' زیر را با مسیر نصب نرم افزار ترینیتی خود جایگزین کنید. در غیر این صورت، شما می توانید مسیر کاملی را بنویسید که $TRINITY_HOME در این راهنما ظاهر می شوند.

```
% export TRINITY_HOME=/software/trinityrnaseq
```

بعد از این که نرم افزار R را نصب کردید، شما بایستی بسته های R زیر را نصب کنید:

1. Bioconductor: http://www.bioconductor.org
2. edgeR: http://www.bioconductor.org/packages/release/bioc/html/edgeR.html
3. gplots

ساده ترین مسیر برای انجام این کار در نرم افزار R به شکل زیر است:

```
% R
> source("http://bioconductor.org/biocLite.R")
> biocLite()
> biocLite("edgeR")
> biocLite("ctc")
> biocLite("Biobase")
> biocLite("ape")
> install.packages("gplots")
```

داده های راهنما و کمکی

برای این داده های راهنما، شما بایستی داده های توالی یابی RNA خاص رشته را از Schizosaccharomyces pombe رشد یافته در چهار شرایط توصیف شده در منبع(45) بدست بیاورد: رشد لگاریتمی، فاز ثابت، رشد دو مرحله ای(ds) و شوک حرارتی(hs) که هر یک با 1 مگابایت داده های توالی یابی rna خاص رشته دو طرفه ایلومینا و 4 مگابایت قرائت دو طرفه همراه است.

قرائت ها برای تجمیع و همگذاری

داده ها را می توان با بازدید از این URL در یک مرور گر وب و یا به طور مستقیم از خط فرمان با استفاده از wget از طریق زیر دانلود کرد:

%wget\http://sourceforge.net/projects/trinityrnaseq/files/misc/TrinityNatureProtocolTutorial.tgz/download

فایل دانلود شده بایستی به صورت 'TrinityNatureProtocolTutorial.tgz' نام گذاری شده و 540 مگابایت اندازه دارد. این فایل را با روش زیر باز کنید:

tar –xvf TrinityNatureProtocolTutorial.tgz

که بایستی تولید فایل های زیر در یک دایرکتوری /TrinityNatureProtocolTutorial با محتویات زیر می کند:

S_pombe_refTrans.fasta # reference transcriptome for S. pombe
1M_READS_sample/Sp.hs.1M.left.fq # PE reads for heatshock
1M_READS_sample/Sp.hs.1M.right.fq
1M_READS_sample/Sp.log.1M.left.fq # PE reads for log phase
1M_READS_sample/Sp.log.1M.right.fq
1M_READS_sample/Sp.ds.1M.right.fq # PE reads for diauxic shock
1M_READS_sample/Sp.ds.1M.left.fq
1M_READS_sample/Sp.plat.1M.left.fq # PE reads for plateau phase
1M_READS_sample/Sp.plat.1M.right.fq
samples_n_reads_described.txt # tab-delimited description
file.

این پروتکل انتظار دارد که داده های خام عاری از هر گونه آداپتور، بارکد ها و سایر نتایج نامطلوب، کیفیت بالایی داشته باشند.

روش

گردایش(همگذاری) توالی یابی RNA نوپدید با استفاده از ترینیتی(زمان:60 تا 90 دقیقه)

1- ایجاد یک پوشه کار و قرار دادن محتویات دایرکتوری'TrinityNatureProtocolTutorial' در آن(بر اساس بخش مواد)

2- به منظور تسهیل تجزیه تحلیل های داون استریم، داده های توالی یابی RNA در کل نمونه ها به یک مجموعه از ورودی ها برای تولید یک مجموعه ترینیتی مرجع منفرد الحاق می شوند. همه قرائت های"چپ" با یک تک فایل ترکیب شده و همه قرائت های" راست" با یک تک فایل از طریق زیر ترکیب می شوند:

```
% cat 1M_READS_sample/*.left.fq > reads.ALL.left.fq
% cat 1M_READS_sample/*.right.fq > reads.ALL.right.fq
```

3- اکنون، قرائت ها در رونوشت ها با استفاده از ترینیتی از طریق مسیر زیر همگذاری و تجمیع می شوند

```
% $TRINITY_HOME/Trinity.pl -seqType fq -JM 10G \-left reads.ALL.left.fq -right
reads.ALL.right.fq - SS_lib_type RF -CPU 6
```

گزینه JM– به کاربر امکان کنترل مقدار RAM مورد استفاده در طی شمارش Jellyfish kmer که در این مورد، 10 گیگابایت رم است را می دهد. گزینه CPU– امکان کنترل تعداد فرایند های موازی را می دهد. بسته به سیستم خود می توانید این تغییرات را اعمال کنید. رونوشت های باز سازی شده رونوشت ها به صورت توالی های با فرمت FASTA در فایل خروجی'trinity_out_dir/ Trinity.fasta' وجود دارند.

4- 'TrinityStats.pl' زیر ، تعدادی از رونوشت ها،اجزا و مقدار N50 کنتیگ رونوشت را بر اساس 'Trinity.fasta' file گزارش می کند. مقدار کنیتگ N50،که به صورت ماکزیمم طولی تعریف می شود که به موجب آن حداقل 50 درصد توالی های تجمیع یافته کل در کن تیگ های با حداقل طول وجود دارند که معمولا یک شاخص رایج برای ارزیابی مجاورت گردایش یا تجمیع ژنوم می باشد. توجه داشته باشید که بر خلاف مجموعه های ژنوم، بیشینه سازی N50 برای ترانسکریپتوم مناسب نیست: استفاده از یک شاخص بر اساس یک مجموعه داده مرجع(از یک گونه یا گونه های خویشاوند) و براورد تعداد ژن های مرجع بازیابی شده و شیوه نام گذاری آن ها به صورت طول کامل(زیر را ببینید) مناسب تر است. با این حال، مقدار N50 برای تایید این که گردایش ادامه دارد(شما یک مقداری را انتظار دارید که طول آن نزدیک به طول رونوشت میانگین S. pombe، میانگین:1397 باز) مفید است. از اسکریپت '$TRINITY_HOME/utilities/TrinityStats.pl' برای بررسی این آماره برای مجموعه های ترینیتی استفاده کنید.

```
% $TRINITY_HOME/util/TrinityStats.pl trinity_out_dir/Trinity.fasta
```

ارزیابی کیفیت(توصیه شده ولی اختیاری)(زمان:90 دقیقه)

5- بررسی عرض ترکیب ژنتیکی و مجاورت رونوشت با اهرم بندی یک مجموعه داده مرجع. ترانسکریپتوم مرجع تفسیر شده از Schizosaccharomyces pombe به صورت S_pombe_refTrans.fasta در نظر گرفته می شود. از مگابلاست و اسکریپت تجزیه تحلیل شده برای تحلیل نمایش آن با مجموعه ترینیتی که به صورت زیر توصیف می شود استفاده کنید:

الف: فایل fasta ترانسکریپتوم مرجع را به صورت دیتابیس بلاست تهیه کنید

```
% makeblastdb -in S_pombe_refTrans.fasta -dbtype nucl
```

ب: مگابلاست را برای هم تراز سازی رونوشت های شناخته شده با مجموعه ترینیتی اجرا کنید

```
% blastn -query trinity_out_dir/Trinity.fasta \-db S_pombe_refTrans.fasta \-out
Trinity_vs_S_pombe_refTrans.blastn \-evalue 1e-20 -dust no -task megablast -
num_threads 2 \-max_target_seqs 1 -outfmt 6
```

پ: وقتی که Blast به پایان رسید، اسکریپت زیر را برای بررسی پوشش طولی دیتابیس فوق اجرا کنید

```
% $TRINITY_HOME/util/analyze_blastPlus_topHit_coverage.pl \
Trinity_vs_S_pombe_genes.blastn \ trinity_out_dir/Trinity.fasta \
trinity_out_dir/Trinity.fasta \
```

6- تعداد قرائت های توالی یابی RNA ورودی را بررسی کنید که با همگذاری و تجمیع ترانسکریپتوم نمایش داده می شود. ترینیتی یک اسکریپت 'alignReads.pl' را ارایه می کند که نرم افزار باوتای را برای هم سو سازی و هم تراز سازی قرائت های قطعات راست و چپ به طور مجزا با کن تیگ های ترینیتی و سپس دسته بندی قرائت ها به جفت ها اجرا می کند ضمن این که تک قرائت ها را که به طور مناسب با جفت خود جفت نمی شوندحفظ می کنند.

alignReads.pl را از طریق زیر اجرا کنید

```
% $TRINITY_HOME/util/alignReads.pl --seqType fq \--left reads.ALL.left.fq--right
reads.ALL.right.fq \--SS_lib_type RF --retain_intermediate_files \--aligner bowtie \--
target trinity_out_dir/Trinity.fasta -- -p 4
```

7- وقتی که "alignReads.pl " با استفاده از داده های خاص رشته اجرا می شود، همان طور که در بالا با تنظیمات پارامتر SS_lib_type RF گفته شد، تراز های با رشته سنس(+) از رشته انتی سنس(-) را جدا می کند. همه فایل

های خروجی از جمله فایل های SAM مرتب شده با نام قرائت و مرتب شده بر اساس مختصات بایستی در دایرکتوری 'bowtie_out/' قرار گیرند. تعداد قرائت های همسو(حداقل یک بار) با رشته سنس رونوشت ها با اجرای برنامه زیر بر روی فایل هم ترازی مرتب شده بر اساس نام قرائت رشته سنس شمارش می شوند:

```
% $TRINITY_HOME/util/SAM_nameSorted_to_uniq_count_stats.pl \
bowtie_out/bowtie_out.nameSorted.sam.+.sam
```

برآورد فراوانی با استفاده از RSEM(زمان:40 تا 60 دقیقه)

براورد های فراوانی رونوشت با اجرای RSEM به طور جداگانه برای هر نمونه به شکلی که در زیر نشان داده شده است بدست می اید. اسکریپت run_RSEM_align_n_estimate.pl،PERL به سادگی یک رابطی را برای نرم افزار RESM ارایه کرده، پارامتر های خط فرمان ترینیتی آشنا را به معادل های RSEM آن ها ترجمه کرده و سپس نرم افزار RSEM را اجرا می کند. هر مرحله مربوطه زیر(مراحل 15-18) فایل های '$ {prefix}.isoforms.results')و('${prefix}.genes.results' حاوی براورد های فراوانی را برای رنوشت های ترینیتی(جدول 1) و اجزا(جدول 2) تولید می کند.${prefix} در نام فایل بر اساس تنظیمات '--prefix' در دستورات زیر است که برای هر نمونه منحصر به فرد است. لطفا توجه داشته باشید که با توجه به پارامتر های گزارش شده در جداول 1 و 2، طول ژن و طول موثر به صورت مجموع وزنی IsoPct از طول های رونوشت و طول های موثر تعریف می شود. تعداد مورد انتظار ژن، TPM و FPKM به صورت مجموع تعداد رونوشت مورد انتظار TOM و FPKM ترعیف می شوند

15- RSEM برای فاز لگاریتمی:

```
% $TRINITY_HOME/util/RSEM_util/run_RSEM_align_n_estimate.pl \--
transcripts trinity_out_dir/Trinity.fasta \--prefix LOG
```

16- RSEM برای فاز تغییر دو مرحله ای

```
% $TRINITY_HOME/util/RSEM_util/run_RSEM_align_n_estimate.pl \
--right Sp.ds.1M.right.fq \--prefix DS
```

17- RSEM برای شوک حرارتی

```
% $TRINITY_HOME/util/RSEM_util/run_RSEM_align_n_estimate.pl \--prefix HS
```

18- RSEM برای فاز ثابت:

```
% $TRINITY_HOME/util/RSEM_util/run_RSEM_align_n_estimate.pl \--
prefix PLAT
```

تجزیه تحلیل بیان دیفرانسیل با استفاده از edgeR، زمان کم تر از 5 دقیقه

توجه داشته باشید که در این بخش، ژن ها و رونوشت ها را می توان به طور جداگانه با استفاده از براورد های فراوانی RSEM متناظر آن ها بررسی کرد. به طور مختصر، ما در این جا تنها رونوشت های زیر را دنبال می کنیم

19- خواهید دید که هر یک از فایل های مربوط به نتایج ایزوفرم RSEM دارای تعدادی از ستون ها می باشد با این حال ما تنها نیاز به یک چیزی مانند 'expected_count داریم. یک ماتریس حاوی تعداد قطعات RNA-seq به ازای هر ویژگی در یک فایل متن جدولی ساده با استفاده از داده های قطعات مورد انتظار تولید شده توسط RSEM را ایجاد کنید.

```
% $TRINITY_HOME/util/RSEM_util/merge_RSEM_frag_counts_single_table.pl\
PLAT.isoforms.results > Sp_isoforms.counts.matrix
```

لطفا به یاد داشته باشید که اولین ستون از ماتریس حاصله، نام رونویسی است. دومین، سومین و به همین ترتیب، تعداد ردیف ها برای هر یک از نمونه های متناظر است. اولین ردیف در بر گیرنده سر عنوان ستون می باشد که شامل یک برچسب برای هر نمونه است.

20- از edgeR برای شناسایی رونوشت های بیان شده دیفرانسیل برای هر جفت از نمونه ها استفاده کنید. اسکریپت زیر، بسیاری از کار های اجرای edgeR یاDESeq را اجرا می کند در این روش، ما تنها از edgeR استفاده می کنیم. از ماتریس ایجاد شده در مرحله 19 به عنوان ورودی استفاده کنید:

```
% $TRINITY_HOME/Analysis/DifferentialExpression/run_DE_analysis.pl \
```

به خاطر داشته باشید که همه edgeR ناشی از مقایسات زوجی ای هستند که اکنون در دایرکتوری خروجی edgeR_dir قرار دارند و نیز شامل فایل های زیر هستند:*.edgeR.DE_results (رونوشت هایی که به طور متفاوت بیان می شوند از جمله تغییرات چند برابر و معنی داری آماری(جدول 3) و *.edgeR.DE_results.MA_n_Volcano.pdf(نمودار های MA و ولکانو از مقایسات زوجی(شکل 9).

21- برای انجام نرمال سازی TMM و برای ایجاد یک ماتریسی از مقادیر بیان اندازه گیری شده در FPKM، ابتدا مقادیر طول رنوشت را از هر یک از فایل های RSEM's *.isoform.results استخراج کنید.

```
% cut -f1,3,4 DS.isoforms.results >  Trinity.trans_lengths.txt
```

22- اکنون، نرمال سازی TMM را انجام دهید

```
% $TRINITY_HOME/Analysis/--lengths Trinity.trans_lengths.txt
```

این دستور تولید فایل های زیر می کند:,Sp_isoforms. counts.matrix.TMM_info.txt' حاوی اندازه موثر کتابخانه برای هر نمونه بعد از نرمال سازی TMM و Sp_isoforms.counts.matrix.TMM_normalized.FPKM' که حاوی مقادیر بیان رونوشت مرمال بر طبق رونوشت و نمونه می باشد که به صورت FPKM اندازه گیری می شود. این فایل ماتریس برای خوشه بندی پروفیل های بیان برای رونوشت ها در نمونه ها استفاده شده و تولید نمودار های نقشه حرارتی یا هیت مپ به شکل زیر می کند.

23- به منظور مطالعه الگو های بیان رونوشت ها یا ژن ها در نمونه ها، محدود سازی تجزیه تحلیل به رونوشت هایی که به طور معنی داری حداقل در یک مقایسه نمونه جفتی بیان می شوند اغلب مفید است. با توجه به مجموعه ای از رونوشت های بیان شده دیفرانسیلی، مقادیر بیان نرمال را استخراج کرده و یک خوشه بندی سلسله مراتبی را برای دسه بندی رونوشت ها با الگو های بیان مشابه در نمونه های مختلف و طبقه بندی نمونه های با پروفیل های بیان مشابه بر طبق رونوشت ها انجام دهید. برای مثال، به دایرکتوری خروجی 'edgeR_dir/' وارد شده و رونوشت هایی که حداقل به طور چهار برابر با معنی داری آماری تصحیح شده کاذب حداکثر 0.001 با استفاده از دستورات زیر بیان می شوند را استخراج کنید:

% cd edgeR_dir/-C 2 -P 0.001

لطفا توجه داشته باشید که پارامتر -C، (تغییرات چند برابر) لگاریتمی را اختیار می کند که در این مورد log2(4) = 2 می باشد. تعدادی از فایل ها تولید می شوند و فایل های با پیش وند 'diffExpr.P0.001_C2 نشان دهنده گزینه های پارامتری است:'diffExpr.P0.001_C2.matrix' در بر گیرنده زیر مجموعه ای از رونوشت ها از ماتریس کامل 'matrix.TMM_normalized.FPKM' می باشد که به صورت بیان دیفرانسیلی شناسایی شده و با آستانه های تعیین شده تعریف می شوند.diffExpr.P0.001_C2.matrix.heatmap.pdf در بر گیرنده یک تصویر هیت مپ خوشه بندی شده می باشد که روابط میان رونوشت ها و نمونه ها(شکل 10 الف) و یک هیت مپ از همبستگی های اسپیرمن زوجی بین نمونه ها(شکل 10 ب) می باشند. 'diffExpr.P0.001_C2.matrix.R.all.RData یک ذخیره محلی از همه داده های تولید شده در طی این تحلیل است که در قسمت روش(مرحله 25) با ابزار های تجزیه تحلیل افزایشی استفاده می شود.

24- تعداد رونوشت های بیان شده دیفرانسیل(متفاوت) را با شمارش تعداد خطوط در فایل با استفاده از دستور زیر تعیین کنید

% wc -l diffExpr.P0.001_C2.matrix

1 را طوری کم کنید که خط سر ستون را به صورت ورودی رونوشت شمارش نکند. توجه داشته باشید که اسکریپت'analyze_diff_expr.pl' به طور مستقیم تعداد رونوشت های بیان شده متفاوت که در آستانه های مربوطه شناسایی می شوند را گزارش می کند.

25- خوشه های رونوشت با پروفیل های بیان مشترک از خوشه های سلسله مراتب تولید شده اولیه با اجرای اسکریپت زیر استخراج کنید که از R برای برش درختی استفاده می کند که رونوشت های خوشه بندی شده سلسله مراتبی را بر اساس معیار های تعیین شده نظیر تولید تعداد خاصی از خوشه ها یا برش درخت در یک ارتفاع خاص نشان می دهد. برای مثال، اسکریپت زیر را برای تقسیم رونوشت ها با برش درخت در 20 درصد ارتفاع آن اجرا کنید:

% $TRINITY_HOME/Analysis/DifferentialExpression/define_clusters_by_cutting_tr
ePtree 20 -R diffExpr.P0.001_C2.matrix.R.all.Rdata

دستور فوق یک دایرکتوری ('diffExpr.P0.001_C2.matrix.R.all.RData.clusters_fixed_P_20/') را تولید می کند که حاوی 'subcluster_*_log2_medianCentered_fpkm.matrix' می باشد- و هر خوشه خود تعریف از رونوشت ها همراه با مقادیر بیانی که تحت تبدیل لگاریتمی و میانه محور هستند ارایه شده و 'my_cluster_plots.pdf' در بر گیرنده یک نمودار از مقادیر بیان میانه و لوگ ترانسفورم(تبدیل لگاریتمی) برای هر خوشه است(شکل 10 ج). توجه داشته باشید که به دلیل دامنه پویای وسیع در مقادیر بیان رونوشت ها در طی این مرحله، مقادیر بیان ابتدا قبل از ترسیم نمودار نقاط داده ها تحت تبدیل لگاریتمی قرار گرفتند. به علاوه، به منظور بررسی الگوهای بیان مشترکی که بر بیان نسبی رونوشت ها در نمونه های مختلف تاکید دارند، میانه هر مقدار بیان رونوشت متعاقبا تعیین شد. این عملیات با تفریق هر مقدار لگاریتم 2(FPKM) میانه رونوشت از مقدار log2(FPKM) در هر نمونه انجام شدند. این داده های حاصله موسوم به مقادیر بیان میانه محور و تبدیل لگاریتمی است که در این مرحله تولید می شود.

26-(اختیاری): اسکریپت مربوط به مرحله 25 را چندین بار با مقادیر متفاوت '--Ptree' به منظور افزایش یا کاهش تعداد خوشه های تولید شده اجرا کنید.

خودکار سازی بخش های مورد نیاز روش: زمان 2-3 ساعت

27-(اختیاری): در صورتی که شما علاقه مند به اجرای بخش هایی از روش بدون تایپ دستی در هر دستور می باشید، اسکریپت زیر را اجرا کنید که بخش های مورد نیاز روشرا اجرا می کند. این بخش ها شامل الحاق قرائت های همه نمونه ها به یک مجموعه داده ورودی، تجمیع قرائت ها با استفاده از ترینیتی، براورد فراوانی به طور جداگانه برای هر نمونه و اجرای edgeR برای شناسایی رونوشت های بیان شده دیفرانسیل. روش خودکار را با دستور زیر از جمله پارامتر اختیاری برای یک تجربه تعاملی اجرا کنید که در آن سیستم متوقف شده و منتظر پاسخ کاربر قیل از ورود به مرحله بعد می ماند

```
% $TRINITY_HOME/util/run_Trinity_edgeR_pipeline.pl \
```

عیب یابی

مشاوره های مربوط به عیب یابی را می توان در جدول 4 یابت

زمان بندی

مراحل 1-3، جمع آوری داده های توالی یابی RNA:10دقیقه

مراحل 4-7: تجمیع و همگذاری ترانسکریپتوم های نوپدید ترینیتی 4 میلیون قرائت ایلومینای دو طرفه:60-90 دقیقه

مراحل 8-13: ارزیابی کیفیت(اختیاری):90 دقیقه

مراحل 14-18: استفاده از RESM برای براورد فراوانی:40-60 دقیقه

مراحل19-25: تحلیل بیان دیفرانسیل با استفاده از EdgeR کم تر از 5 دقیقه

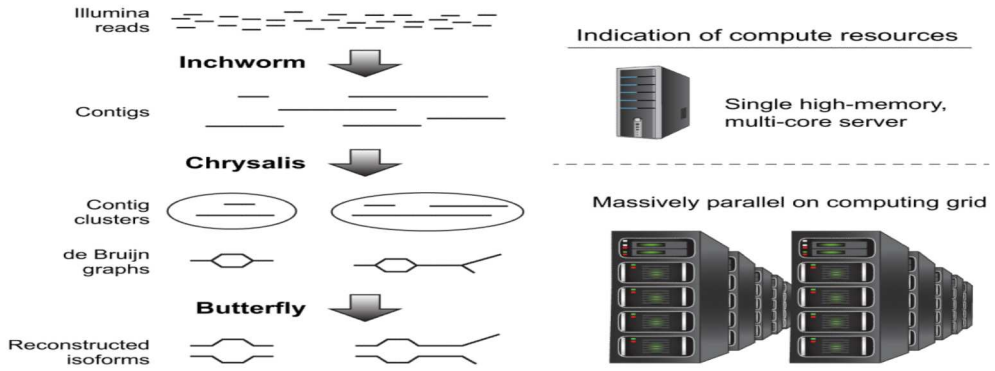
مرحله 26(اختیاری): متغیر

مرحله 27:اتوماسیون بخش های مورد نیاز:2-3 ساعت

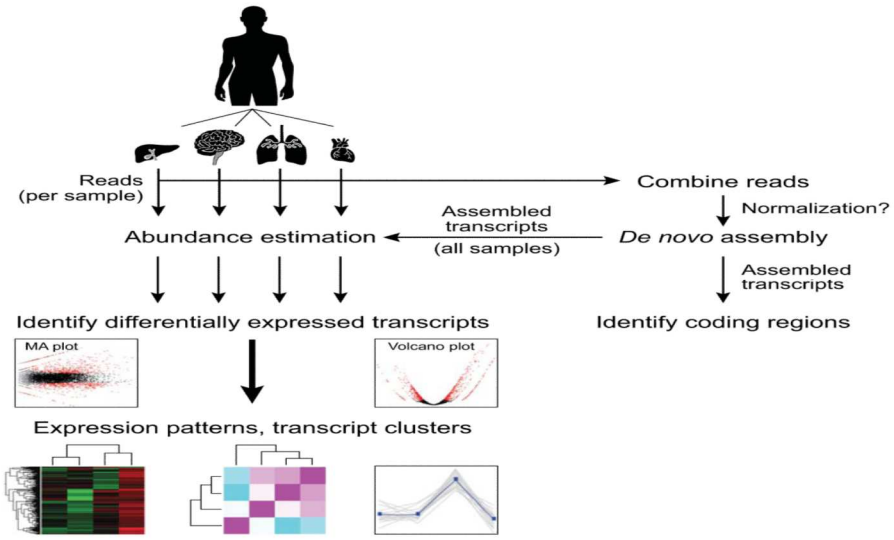
زمان تخصیص داده شده برای هر یک از دستورات اجرا شده برای بخش های مورد نیاز راهنما،که در طی اجرای خودکار از طریق'run_Trinity_edgeR_pipeline.pl فوق الذکر گزارش شده است و در زمان اجرای سرور با عملکرد بالا در موسسه بورد(مشخصات سخت افزاری مربوطه)، در بخش یادداشت های مکمل ارایه شده است.

نتایج پیش بینی شده

از آن جا که خروجی ترینیتی کاملا قطعی نیست، تغییرات بسیار خفیف در خروجی(تعداد یا طول رونوشت ها) ممکن است ناشی از اجرای ترینیتی در زمان های مختلف یا بر روی سخت افزار های متفاوت باشد. آماره تجمیع و هم گذاری ترینیتی در جدول 5 ارایه شده است. نتایج نقشه یابی BLASTIN رونوشت مرجع از مرحله 11 در جدول 6 نشان داده شده است. در همه موارد، 4765 رونوشت های مرجع S. pombe دارای یک BLAST با مقدار E کم تر از 1e-20، و 3401 از کل 5163 رونوشت های مرجع کل می باشد که تقریبا به صورت طول کامل در نظر گرفته شده و هم ترازی کن تیگ ترینیتی تا بیش از 90 درصد تطبیق طول رونوشت مرجع است. تعداد قرائت های هم تراز شده با مجموعه ترینیتی از طریق alignReads.pl و استفاده از هم تراز کننده باوتای که در مرحله 13 بدست آمد، در جدول 7 نشان داده شده اند. تعداد رونوشت های بیان شده دیفرانسیلی شناسایی شده با نرخ کشف کاذب(FDR) حداکثر 0.001 و حداقل تفاوت 4 برابری در مقادیر بیان که در مرحله 24 برررسی شد ، 659 می باشد.

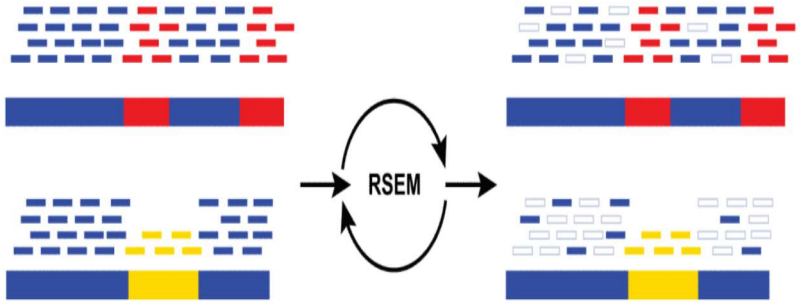


شکل 1: مروری بر همگذاری ترینیتی و خط لوله تجزیه تحلیل. مراحل متوالی کلیدی در ترینیتی(چپ) و منابع محاسباتی مربوطه(راست). ترینیتی قرائت کوتاه ورودی(سمت چپ بالا) را اختیار کرده و ابتدا از ماژول اینچ ورم برای ساخت کن تیک استفاده می کند. این مرحله نیازمند یک سرور با حافظه زیاد(رم 1 گیگابایت به ازای 1 میلیون قرائت جفت شده، ولی بر طبق پیچیدگی قرائت متغیر است: بالا سمت راست) می باشد. Chrysalis (وسط سمت چپ)، کن تیگ های اینچورم مربوطه را خوشه بندی کرده و اغلب تولید ده تا صد ها هزار خوشه کن تیگ اینچورم می کند که هر یک از آن ها به یک جزء گراف دی براین به طور مستقل و به طور موازی بر روی شبکه محاسباتی(پایین سمت راست) پردازش می شود. Butterfly (سمت چپ پایین) سپس، همه توالی های محتمل را از هر جزء گراف استخراج می کند که می تواند موازی سازی نیز شود.

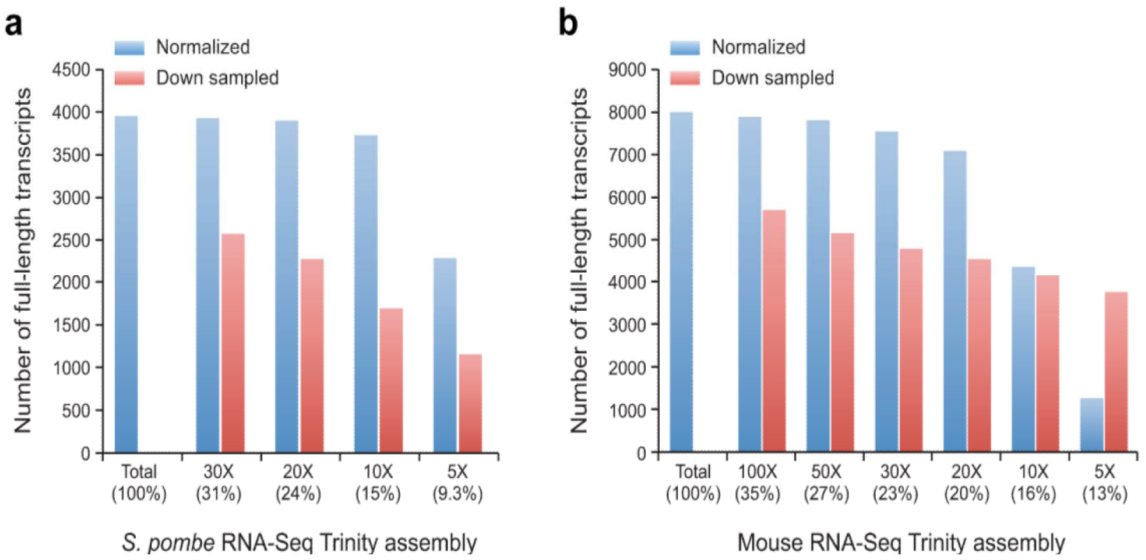


شکل 2: اثرات نرمال سازی داده های مربوط به قطعات کامپیوتری توالی یابی RNA بر روی باز سازی رونوشت طول کامل ترینیتی(a-b). محور y تعداد رونوشت های طول کامل باز سازی شده از مجموعه داده های حاصل از

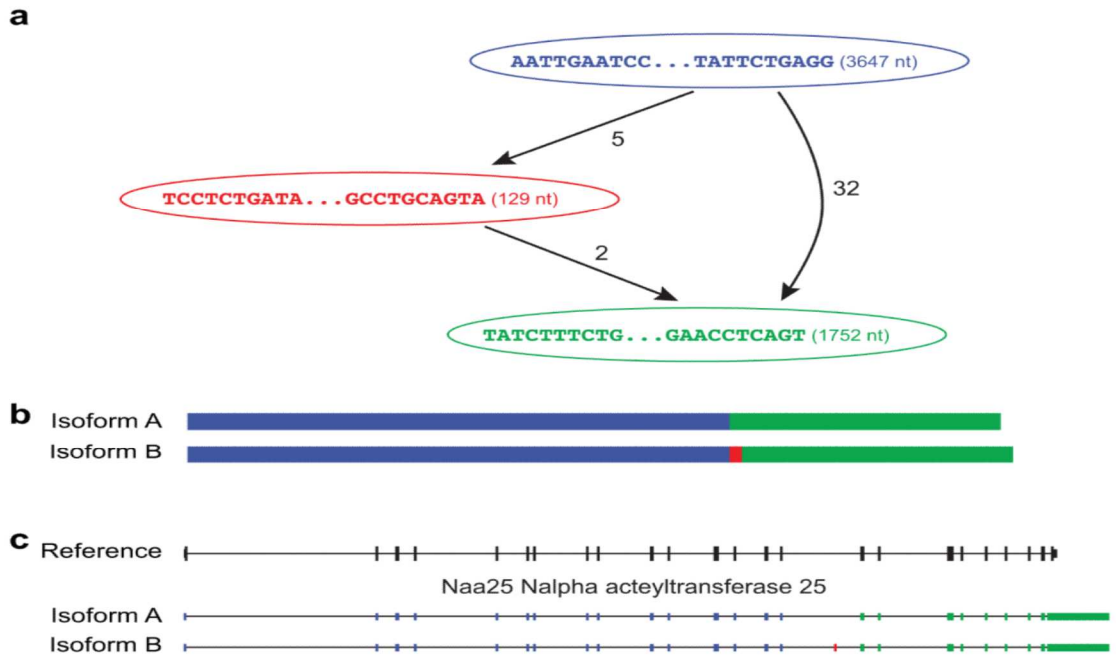
توالی یابی RNA خاص رشته دو طرفه در S. pombe(10 میلیون قرائت دو طرفه)، a: و موش(100 میلیون قرائت دو طرفه)،b: استفاده از مجموعه داده های کامل(کل 100 درصد) یا نمونه های مختلف(محورx) یا از طریق روش نرمال سازی کامپیوتری(با پوشش 5 تا بیش از 100* ماکزیمم *k*-mer هدف(k=25) یا کاهش تصادفی اندازه نمونه، تعداد یکسانی از قرائت ها(خطوط قرمز)



شکل 3: مدل های ترانسکریپتوم و ژنوم از ترانسکریپت های متصل شده جایگزین(a-c). یک مثال از مدل کرافیکی تولید شده توسط نرم افزار Trinity's Butterfly، الف: همراه با رونوشت های باز سازی شده متناظر ب: و ساختار اگزونی ان ها بر اساس تراز سازی با ژنوم موش(پ). هر گره در a با یک توالی ارتباط دارد و لبه های جهت دار توالی های متوالی را از 5 تا 3 در یک رونوشت متصل می کنند. برامدگی ها و شاخه ها، تفاوت های توالی را بین رونوشت های باز سازی شده جایگزین نشان می دهد از جمله اگزون های کاست متصل شده جایگزین: تنها یک برامدگی منفرد در این گراف رونوشت نشان داده شده است که به یک گره قرمز منتهی می شود. یال ها یا لبه ها با تعداد قطعات توالی یابی RNA تفسیر می شوند که رونوشت را از توالی 5 تا توالی 3 پشتیبانی می کند. در این مثال، دو مسیر پشتیبانی شده وجود دارد: یکی از گروه آبی به سبز(که با 32 قطعه پشتیبانی می شود) و تولید ایزوفرم A می کند(b بالا) و دیگری از آبی به قرمز به سبز، که توسط حداکثر 5 قطعه پشتیبانی شده و تولید ایزوفرم B می کند(b، پایین). گره قرمز نتیجه یک اگزون حذف شده جایگزین می باشد که در ساختار ژنی مشهود است(C خط قرمز نشان داده شده در ایزفرم B). گراف های رونوشت هدایت پذیر به طور اختیاری توسط باترفلای تولید می شود و با فرمت نقطه ارایه می شود و می توان با استفاده از graphviz نمودار آن را ترسیم کرد(graphviz).این جزییات در وب سایت ترینیتی دیده می شود(http://trinityrnaseq.sourceforge.net/advanced_trinity_guide.html).

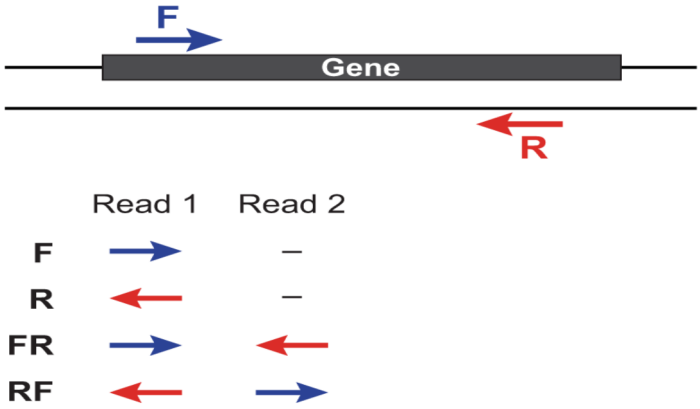


شکل 4: انواع کتابخانه های خاص رشته. نمودار قرائت های توالی یابی سمت چپ (/1) و سمت راست(/2) بر طبق جهات ان ها نسبت به رشته سنس یک توالی رونوشت ترسیم می شود. نوع کتابخانه خاص رشته(F-R,FR, یا FR) بستگی به پروتکل ساخت کتابخانه داشته و برای ترینیتی از طریق پارامتر '--SS_lib_type' خاص کاربر می باشد.



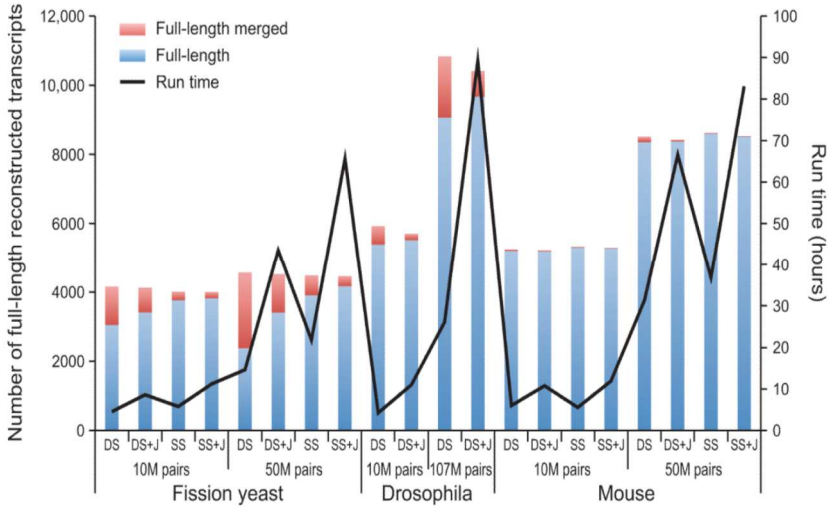
شکل5: باز سازی رونوشت طول کامل با ترینیتی در موجودات مختلف، اعماق توالی یابی و پارامتر ها. محور y در سمت چپ، تعداد رونوشت های کاملا باز سازی شده را برای مجموعه های ترینیتی داده های توالی یابی RNA بر گرفته از مخمر شکافت(*S. pombe*8,41)،*Drosophila melanogaster*11 ، و موش با ترکیبات متفاوت

پارامتر ها نشان می دهد: حالت دو رشته ای DS، حالت خاص رشته ای،+J، استفاده از پارامتر '*jaccard_clip*--' برای تقسیم رونوشت های ترکیب شده کاذب. هر دو نتایج SS و DS برای *S. pombe* و موش ارایه شده ولی تنها نتایج DS برای *Drosophila* ارایه می شود زیرا داده های توالی یابی RNA آن خاص رشته نیست. رنگ آبی نشان دهنده رونوشت های طول کامل است، رنگ قرمز نشان دهنده رونوشت های ترکیبی طول کامل است(یعنی رونوشت هایی که به اشتباه با سایر رونوشت های(معمولا همسایه) ترکیب می شوند(چند وجهی). ستاره های مشکی(مقادیر نشان داده شده در محور Y در سمت راست نمودار) تعداد دفعاتی را با یک سرور با حافظه بالای معاصر(256-412 گیگ رم) نشان می دهد که از ماکزیمم 4 رشته استفاده می کند(CPU4-، به مرحله 6 روش مراجعه کنید).

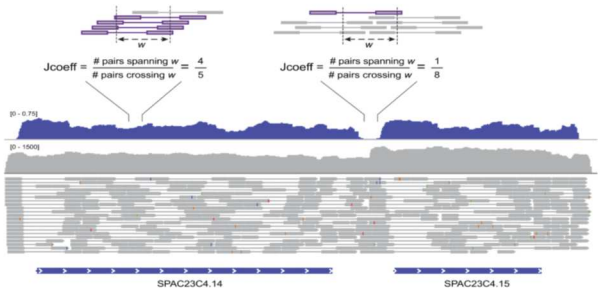


شکل 6: ارزیابی پشتیبانی قرائت جفتی از طریق ضریب تشابه جاکارد. پشتیبانی جفت قرائت ابتدا با شمارش تعداد قطعات توالی یابی RNA(کران های قرائت های جفتی) محاسبه می شود که شامل هر یک از دو نقطه بیرونی از طول پنجره معین می باشد(پیش فرض: 100 باز) و سپس محاسبه ضریب تشابه جاکارد برای مقایسه قطعاتی که در هر دونقطه هم پوشانی دارد. یک مثال برای جفت رونوشت همسایه S. pombe(SPAC23C4.14 و SPAC23C4.15,) با پوشش قرائت دارای هم پوشانی(مسیر خاکستری) نشان داده شده است که منجر به یک رونوشت(ترکیبی) مجاوری می شود که توسط اینچورم ترکیب می شود. با این حال، ضریب تشابه جاکارد(مسیر سیاه) محاسبه شده از قرائت های جفتی(دمبل های خاکستری) به وضوح موقعیت پشتیبانی جفتی کاهش یافته را نشان می دهد. مثال هایی از پشتیبانی جفت قوی(بالا سمت چپ) و ضعیف(بالا سمت راست) که در بالا ترسیم شده اند. هنگام استفاده از پارامتر '--jaccard_clip'، کن تیگ

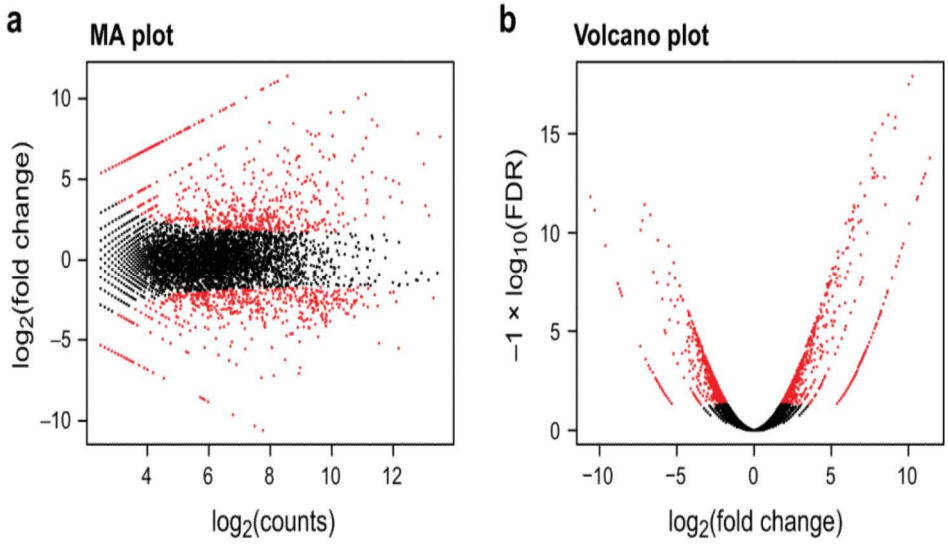
اینچورم به دو رونوشت طول کامل تقسیم می شود و سپس با ChrysalisوButterfly به صورت بخشی از خط لوله ترینتی پردازش می شوند.



شکل 7: همگذاری ترانسکریپتوم نوپدید و جریان تجزیه تحلیل. قرائت ها از چندین نمونه(برای مثال، بافت های مختلف، بالا) به یک مجموعه داده ترکیب می شوند. قرائت ها را می توان برای کاهش تعداد قرائت ها ضمن حفظ تنوع و پیچیدگی نمونه، نرمال سازی کرد. مجموعه قرائت ترکیبی با ترینیتی برای تولید مجموعه ترانسگریپتوم نوپدید مرجع(سمت راست) تجمیع می شود. مناطق کد کننده پروتین را می توان از مجموعه مرجع با استفاده از TransDecoder استخراج کرده و مشخصات آن را بر اساس عملکرد احتمالی بر اساس تشابه توالی یا محتوی حوزه تعیین کرد. به طور جداگانه، تحلیل بیان خاص نمونه با هم تراز سازی قرائت های نمونه اصلی با ترنسکریپتوم مرجع بر اساس نمونه انجام شده و بعد از آن براورد فراوانی با استفاده از RESM صورت می گیرد. رونوشت های بیان شده دیفرانسیلی با استفاده از نرم افزار بیوکنداکتور شناسایی می شوند نظیر edgeR با یک ماتریس حاوی براورد های فراوانی RESM(تعداد قطعات توالی یابی RNA نقشه یابی شده به هر رونوشت از هر نمونه). رونوشت های بیان شده دیفرانسیلی را می توان بر اساس الگوهای بیان آن ها دسته بندی کرد.

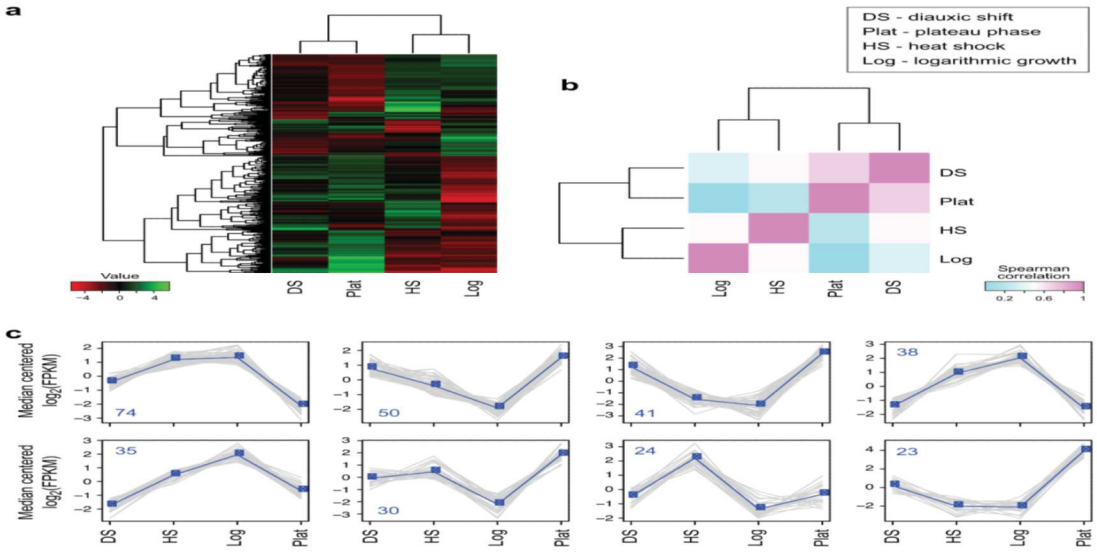


شکل 8: براورد فراوانی از طریق بیشینه سازی امید ریاضی با RESM. یک مثال گویا از براورد فراوانی برای دو رونوشت با توالی های مشترک(آبی) و منحصر به فرد(قرمز و زرد). به منظور براورد فراوانی رونوشت ها، قرائت های توالی یابی RNA((خطوط کوتاه) ابتدا با توالی های رونوشتی (خطوط بلند، پایین) هم تراز می شوند. مناطق منحصر به فرد ایزوفرم ها به طور منحصر به فردی قرائت های توالی یابی RNA (خطوط کوتاه زرد و قرمز) را پوشش داده و توالی های مشترک بین ایزوفرم ها تنها چند قرائت را پوشش می دهند(خطوط آبی کوتاه). یک الگوریتم بیشینه سازی امید ریاضی، که در نرم افزار RESM پیاده سازی می شود،محتمل ترین فراوانی رونوشت ها را براورد کرده و سپس به صورت فراکتال، قرائت ها را به ایزوفرم ها بر اساس فراوانی آن ها تخصیص می دهد. تخصیص قرائت ها به ایزوفم های ناشی از تکرار های بیشینه سازی امید ریاضی به صورت خطوط کوتاه توپر نشان داده شده است(سمت راست) و تخصیصات حذف شده به صورت خطوط توخالی نشان داده شده اند. از این روی، در این مثال، یک بخش زیادی از هر قرائت به ایزوفرم بیشتر بیان شده در بالا تخصیص داده می شود تا به ایزفرم موجود در پایین.



شکل 9: مقایسات زوجی فراوانی رونوشت. دو نمودار مربوط به مقایسه پروفیل های بیان رونوشت بین رشد لگاریتمی و نمونه های رشد ثابت از S. pombe برای شناسایی رونوشت های بیان شده به صورت دیفرانسیلی. A: نمودار MA برای تحلیل بیان دیفرانسیل تولید شده با EdgeR: برای هر ژن، لگاریتم2(تغییر چند برابر)،لگاریتم 2(فاز ثابت/ رشد لگاریتمی) بین دو نمونه ترسیم می شود(A، محور y) که تابعی از لگاریتن 2) تغییر چند برابر)

بین نمونه ها است.(logFC،محور X). رونوشت هایی دارای بیان با تفاوت معنی دار با حداکثر 0.1درصد FDR می باشند به رنگ قرمز هستند.



شکل 10: مقایسات پروفایل های رونویسی در نمونه ها الف: خوشه بندی سلسله مراتبی رونوشت ها و نمونه. هیت مپ یا نقشه حرارتی، سطوح بیان نصبی هر رونوشت(ردیف ها) را در هر نمونه(ستون) نشان می دهد. ردیف ها و ستون ها به شکل سلسله مراتبی خوشه بندی شده اند. مقادیر بیان FPKM، توسط رونوشت تحت تبدیل لگاریتمی قرار گرفته و مبتنی بر میانه هستند.ب: هیت مپی که ماتریس همبستگی اسپیرمن خوشه بندی شده به طور سلسله مراتبی را از مقایسه مقادیر بیان رونوشت (FPKM نرمال شده با TMM) برای هر جفت نمونه نشان می دهد. پ: خوشه های رونوشت استخراج شده از خوشه بندی سلسله راتبی با R, محور ایکس: نمونه ها: محور Y: لگاریتم2(FPKM) مبتنی بر میانه. خطوط خاکستری، رونوشت های فردی: خطوط آبی، مقادیر بیان متوسط در هر خوشه. تعداد رونوشت ها در هر خوشه در گوشه سمت چپ هر نمودار نشان داده شده است، تغییر دو مرحله ای، HS: شوک حرارتی، لگاریتمی: رشد لگاریتمی، Plat: رشد ثابت.

transcript_id	gene_id	length	effective_length	expected_count	TPM	FPKM	IsoPct
comp56_c0_seq1	comp56_c0	3739	3443	637.65	16664.43	7008.23	11.26
comp56_c0_seq2	comp56_c0	3697	3401	4966.34	131393.38	55257.53	88.74
comp62_c0_seq1	comp62_c0	7194	6898	4551.13	59364.09	24965.59	95.54
comp62_c0_seq2	comp62_c0	7076	6778	208.87	2771.95	1165.74	4.46

جدول1

gene_id	transcript_id(s)	length	effective_length	expected_count	TPM	FPKM
comp56_c0	comp56_c0_seq1, comp56_c0_seq2	3701.73	3405.49	5604	148057.81	62265.76
comp62_c0	comp62_c0_seq1, comp62_c0_seq2	7188.74	6892.5	4760	62136.04	26131.33

جدول2

Transcript	logFC	logCPM	PValue	FDR
comp5128_c0_seq1	10.3	11.1	2.13e-22	1.22e-18
comp5231_c0_seq1	10.0	10.9	1.10e-21	3.13e-18
comp5097_c0_seq1	8.7	11.3	5.72e-20	1.10e-16
comp1686_c0_seq1	9.2	10.4	1.01e-19	1.46e-16
comp1012_c0_seq1	8.3	11.5	2.8e-19	3.23e-16

جدول3

Assembly statistic	Value
Total Trinity transcripts	9299
Total Trinity components	8694
Contig N50	1585

جدول4

% Length coverage bin	Count of reference transcripts in bin	Cumulative count of reference transcripts at or above bin level.
100	3401	3401
90	194	3595
80	165	3760
70	197	3957
60	224	4181
50	203	4384
40	158	4542
30	140	4682
20	83	4765
10	0	4765
0	0	4765

جدول5

Read classification	Count of individual reads	Percent of mapped reads
Proper pairing	8102100	93.12
Left only	307933	3.54
Right only	284203	3.27
Improper pairing	6476	0.07

جدول6

ترجمه فا
TarjomeFa.Com

این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

✔ لیست مقالات ترجمه شده

✔ لیست مقالات ترجمه شده رایگان

✔ لیست جدیدترین مقالات انگلیسی ISI

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی