ترجمه فا

TarjomeFa.Com

ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

ابزار های بیوانفورماتیک در تحقیقات lncRNA

چکیده

روش های آزمایشی فعلی برای شناسایی کارکرد ها و نقش تعداد زیادی از کاندید های RNA غیر کد کننده بلند(lncRNA) از نظر بازده محدودیت دارد. از این روی، آگاهی از این که کدام ابزار ها برای درک lncRNA موثر هستند به طوری که سرعت و صحت منطقی حاصل شود، لازم است. در این مقاله، ما به ارزیابی ابزار های بیوانفورماتیک موجود و دیتابیس های مفید برای یافتن RNA غیرکد کننده و تحلیل ساختار ها، حفاظت، اثرات متقابل ، بیان و موقعیت می پردازیم.

کلمات کلیدی: lncRNA، بیان، نگاشت، حفاظت، ساختار ثانویه

1-مقدمه

فناوری های توالی یابی با بازدهی بالا ما را قادر به دست یابی به تعداد زیادی از lncRNA یا RNA های غیر کد کننده کرده است. با این حال، چون روش های شناسایی آزمایشی و تجربی فعلی از نظر توان و بازدهی محدودیت دارند، ابزار های بیوانفورماتیک سریع برای شناسایی lncRNA با صحت منطقی، لازم هستند.

چون lncRNA یک دسته منحصر به فرد از RNA هایی است که قادر به کد کردن پلی پپتید های کارکردی نیست، اولین نقش بیو انفورماتیک، شناسایی lncRNA از طریق فیلتر رونوشت های بلندی است که به نظر نمی رسد پروتین ها را کد گذاری کند. هدف تحقیق lncRNA نه تنها یافتن RNA غیر کد کننده بلند، بلکه شناسایی وظایف و نقش آن هاست. ابزار های بیوانفورماتیک متعددی برای پیش بینی ساختار ها و نقش توالی های RNA از جمله ابزار های مختلفی که از داده های ازمایشی استفاده می کنند وجود دارند.

بدیهی است که ساختار ها، به خصوص ساختار های ثانویه، از عوامل تعیین کننده نقش RNA های غیر کد کننده می باشد. بدیهی است که کارکرد ها و نقش های مشابه بین گونه ها حفظ می شوند. به این ترتیب ساختار های ثانویه و حفاظت آن ها با استفاده از ابزار های بیو انفورماتیک برای تعیین مقوله های کارکردی آن ها بررسی می شود. با این حال پیش بینی ساختار های ثانویه همیشه صحیح نیست. اگرچه استخراج ابعاد ساختاری کارکرد ها اسان نیست، با این حال حوزه های کارکردی دارای ویژگی های ساختاری می باشند.

نشانه های مهم برای کارکرد های lncRNA، از جمله زمان، مکان و روش را می توان از داده های ازمایشی استخراج کرد. الگوی بیان مکانی زمانی توسط توالی RNA نشاندهنده زمان و مکان فعال سازی کارکرد هاست.اثر متقابل با پروتین ها نشان دهنده نوع نقش و کارکرد می باشد. به علاوه، باز های مکمل در دو مولکول RNA اغلب جفت باز ها را تشکیل داده و موجب تخصصی شدن توالی برای RNA هدف می شود. اثر متقابل RNA-RNA را می توان با جست و جوی توالی های مکمل معکوس، فیلتر کرد.

2- یافتن RNA های غیر کد کننده بلند

دو مرحله در شناساییlncRNA وجود دارد.در اولین مرحله، واحد های رونویسی شده lncRNA شناسایی می شوند. قطعات توالی های RNA رونویسی شده که با استفاده از فناوری های توالی یابی نسل اینده مشاهده می شود، بر روی ژنوم مرجع نگاشته شده و برای بدست اوردن واحد های رونویسی شده RNA خلاصه سازی می شوند. دومین مرحله، واحد های رونویسی شده رابه صورت کد کننده یا غیر کد کننده طبقه بندی می کند: توالی واحد های رونویسی شده بر اساس تشابه کدون با توالی های پروتین ارزیابی می شود.

قبل از ظهور فناوری های NGS، پیش بینی کاندید های RNA غیر کد کننده بر اساس توالی ها و تایید بیان آن ها از اهمیت زیادی برخوردار است. برای این پیش بینی، ویژگی های حفاظت شده توالی های کاندید در نظر گرفته می شود. این تحلیل ها برای شناسایی کارکرد های lncRNA اهمیت دارند.

2-1 شناسایی واحد های رونویسی شده lncRNA

پیشرفت اخیر در زمینه فناوری NGS امکان تحلیل رونویسی را در انواع مختلف سلول ها داده است. توالی های بدست امده، قطعات جزیی از رونوشت های با طول زیاد هستند. این رید ها بر روی ژنوم مرجع با استفاده از ابزار هایی نظیر Tophat [1], LAST [2], and STAR [3] نگاشته می شوند. واحد های رونویسی شده RNAاز رید های نگاشته شده با استفاده از ابزار هایی نظیر GIFFLINKS و SCRIPTURE نگاشته می شود.

2-2 ارزیابی پتانسیل کد کننده رونوشت ها

رویکرد های تمایز بین توالی های کد کننده و غیر کد کننده مشابه با روشهای کشف ژن در توالی های ژنومیک می باشند. آن ها بر اساس تشابه با توالی های کد کننده آماره فرکانس های کدون می باشند. تشابه با توالی های کد کننده با استفاده از ابزار های جست و جوی همولوژی شناسایی می شود. لازم به ذکر است که وجود حتی

یک قطعه کوتاه از پپتید ها موید شناسایی رونوشت به صورت کد کننده است. برای lncRNA های جدید، ما قادر به حفاظت بین گونه ها نمی باشیم.

در میان ابزار های ارزیابی پتانسیل کد گذاری، CPC وportait از مقایسات زوجی برای ارزیابی پتانسیل کد گذاری استفاده می کند. بر عکس phyLCSF و RNACODE از ارایشات چندگانه برای تعیین ویژگی های فیلوژنتیکی استفاده می کنند.

امکان ارزیابی پتانسیل کد گذاری با استفاده از ویژگی های آماری توالی های رونوشت بدون اسفتاده از اطلاعات مربوط به همسویی با سایر توالی ها وجود دارد. طول ORF و اریبی کدون از رایج ترین ویژگی هاست. در رگرسیون لجستیک، CPAT از طول ORF، پوشش ORF در رونوشتها و فراوانی های هگزامر استفاده می کند.

2-3 حفظ توالی ها و ساختار ها

چندین ابزار برای یافتن RNA های حفاظت شده از چندین توالی ژنوم وجود دارد با این حال ابزار های ساده بر اساس برنامه نویسی پویای استاندارد با توالی های RNA همسویی ندارد. ابزار های پیشرفته شامل RNAz و QRNA می باشند که قادر به پیش بینی ساختار های ثانویه RNA در توالی های متعدد می باشند. اوفولد نیز دارای ساختار های RNAکارکردی در اریشات توالی متعدد با استفاده از مدل احتمالی پیشرفته برای فرایند جایگزینی در مناطق جفت شده و جفت نشده می باشند با این حال، محاسبه اریشات ساختاری چندگانه توالی های RNA از نظر محاسباتی سخت است به خصوص زمانی که حفاظت از ساختاری های ثانویه در نظر گرفته شود. زمان محاسباتی $O(L^6)$ حتی برای دو توالی با طولL لازم است با این حال ابزار های سریع کمی در بخش 5-2 توصیف شده اند که در آن پیش بینی ساختار با استفاده از همسوسازی ساختاری بحث می شود.

پروفیل های بیان lncRNA بر گرفته از داده های توالی RNA

الگوهای بیان مکانی زمانی(بافت ها، بخش های زیر سلولی و مراحل تمایز و رشد) lncRNAt اطلاعات اساسی برای درک کارکرد های بیولوژیکی lncRNA در سلول ها میباشند. چندین مطالعه توالی RNA وجود دارد که پروفیل های بیان ژن های کد کنندهو غیر کد کننده در بافت های مختلف بدست اورده است. با این حال استخراج الگوهای بیان از دادههای توالی RNA نیازمند منابع محاسباتی بزرگ مقیاس برای فیلترینگ با کیفیت و کمی سازی بیان ژن است.

EXPRESSION ATLAS، داده های توالی RNA را با استفاده از خط لوله محاسباتی اصلی پردازش کرده و پروفیل های بیان بافت ها و لاین های سلولی مختلف از 16 گونه را ارایه کرده است. پروفیل های بیان lncRNA برای این سه گونه ارایه شده است. به ویژه، برای انسان، ATLAS پروفیل های بیان lncRNA را بر گرفته از 32 بافت طبیعی و 675 بافت سرطانی ارایه کرده است.پروفیل های بیان ژن های کد کننده پروتین از چندین بافت مهره داران دیگر نظیر مانند اپوسوم، میمون رشت، بابیان زیتون، مرغ، گاو، فلفل، قورباغه و مارمولک جو و برنج نیز در EXPRESSION ATLAS موجود است با این حال پروفیل های بیان lncRNA به دلیل نبود تفسیر، قابل دسترس نیست. اخیرا CTEX CONSTRTIUM مجموعه ای از داده های توالی RNA را از 43 بافت انسان برای 53934 ژن ارایه کرده است که در آن پروفیل های بیان lncRNA واقع شده اند.

3-1 بافت ویژه بودن

بافت ویژه بودن بیان lncRNA از ویژگی های مهم برای شناسایی lncRNA می باشند. تحلیل اخیر داده های توالی RNA برگکفته از 24 بافت انسانی نشان داده است که اکثریت lncRNA الگوهای بیان بافت ویژه را نشان می دهد، در حالی که این الگو.های بیان در بخش کوچکی از ژن های کد کننده پروتین مشاهده می شوند. واستیل و همکاران(30) اقدام به تحلیل پویایی تکاملی الگوهای بیان lncRNA و بافتویژه بودن در نه بافت در شش پستاندار کردند.

برای بررسی ویژگی الگوهای بیان ژن های کد کننده وغیر کد کننده از ریز ارایه داده های توالی RNA چندین شاخص بر اساس انتروپی شانون پیشنهاد شده است. با این حال این شاخص ها به طور مستقیم تعیین کننده تعداد کمی از بافت ها با سطوح بالای بیان نمی باشد زیرا انتروپی شانون هر ژن درجه ای از انحراف را از الگوی بیان برای همه بافت ها نشان می دهد.ROKA (31) یک ابزار مفید برای تشخیص بافت های واقعی با سطوح بیان بالا و پایین به صورت دادههای پرت بوده و هم زمان ویژگی بافت را با محاسبه انتروپی ارزیابی می کند.

3-2 مکان یابی زیر سلولی

مکان یابی زیر سلولی lncRNA یک عامل مهم دیگر برای کنترل اثرات متقابل درشت مولکولی نظیر کروماتین RNA lncRNA و اثر متقابل پروتین-lncRNA در سلول می باشد. به علاوه، داده های توالی RNA بدست امده از سه اجزای زیر هسته در سلول K565 نیز ارایه شده است. تحلیل چندین لاین سلول انسانی نشان داده

است که lncRNA در هسته بیشتر از سیتوسول هستند.EXPRESION ATLAS این داده های توالی RNA را برای ارایه پروفیل های بیان همه ژن ها از جمله ژن هایlncRNA پردازش کرده است. به علاوه، داده های توالی RNA در بروزر ژنوم UCSC موجود است.

فعل و انفعالات درشت مولکولی شامل lncRNA

شناسایی اهداف اثر متقابل RNA یا پروتین ها lncRNA، یک رویکرد رایج برای تعیین نقش آن هاست. چندین روش ازمایشی نظیر توالی RIA، RAP-RNA وpar-clipو HITS CLIP برای بررسی اثر متقابل RNA-پروتینو RNA-RNA پیشنهاد شده است. در این روش ها، رسوب ایمنی براساس انتی بادی یا پروب انتی سنز، از روش های کلیدی برای تخلیص یک پروتین RNA میباشد. با این حال،این گام های ضروری موجب کوچک تر شدن حوزه تحقیقات می شود. از این روی کاربرد این روش ها در هر یک از این lncRNA ها برای شناسایی جامع اینتراکتوم سخت است.

4-1 روش های پیش بینی برای اثر متقابل RNA-lncRNA

پیش بینی های محاسباتی اثر متقابل RNA-RNA بر اساس انرژی تعاملی است که از اثر متقابل جفت باز بین مولکولی دو مولکول RNA می باشند.INTARNA یک ابزاری برای پیش بینی اثر متقابل RNA-RNA با استفاده از توالی های اولیه دو RNA ارایه می کند. بهبود پیش بینی اثر متقابل RNA -RNA از طریق COPRAN صورت گرفت که از رویکرد ژنومیک با استفاده از سه توالی ژنومی گونه های مجزا استفاده می کند. با این حال این روش ها برای پیش بینی اثر متقابل RNA-lncRNA استفاده نشده اند زیرا این روش ها بر پیش بینی اثر متقابل باکتریایی RNA-lncRNA کوچک مقیاس تاکید دارند.

اخیرا، محققان، خط لوله محاسباتی را از جمله ابزار های تحلیل توالی محاسباتی (Raccess [42], TanTan [43],

LAST [2, 44], IntaRNA [40], and RactIP [45]) برای پیش بینی اثر متقابل RNA-lncRNA ارایه کرده اند. دیتابیس همه اثرات متقابل RNA-lncRNA انسانی پیش بینی شده شامل اثرات متقابل lncRNA-RNA و lncRNA-MRNA برای lncRNA 23898 وm-RNA20185 می باشد.

روش های پیش بینی اثرات متقابل پروتین-lncRNA

چندین روش محاسباتی برای پیش بینی اثرات متقابل پروتین-RNA توسعه یافته است(47-48-49-50-51). در این روش ها، رویکرد های یادگیری ماشینی نظیر تحلیل افتراقی خطی فیشر، ماشین بردار پشتیبانی و جنگل تصادفی برای تفکیک جفت پروتین-RNA از جفت باز ها استفاده شده اند. RPI-seq [47], catRAPID [48], and lncPRO [49] از روش های مبتنی بر توالی هستند که نیازمند توالی های اصلی RNA و پروتین برای داده های ورودی هستند. در میان این سه روش، CATRAPID و LNCPRO ویژگی های فیزیوشیمیایی امینو اسید ها ونوکلوتید ها می باشند و ساختار های ثانویه پروتین وrna را به صورت ویژگی های جفت باز های پروتین-RNA بیان می کند. این دو روش بر پیش بینی اثر متقابل lncRNA-پروتین تاکید می کند. چندین روش محاسباتی برای پیش بینی اثرات متقابل پروتین-RNA توسعه یافته است(47-48-49-50-51). در این روش ها، رویکرد های یادگیری ماشینی نظیر تحلیل افتراقی خطی فیشر، ماشین بردار پشتیبانی و جنگل تصادفی برای تفکیک جفت پروتین-RNA از جفت باز ها استفاده شده اند.RPI-seq [47], catRAPID [48], and lncPRO [49] از روش های مبتنی بر توالی هستند که نیازمند توالی های اصلی RNA و پروتین برای داده های ورودی هستند. در میان این سه روش، CATRAPID و LNCPRO ویژگی های فیزیوشیمیایی امینو اسید ها ونوکلوتید ها می باشند و ساختار های ثانویه پروتین وrna را به صورت ویژگی های جفت باز های پروتین-RNA بیان می کند. این دو روش بر پیش بینی اثر متقابل lncRNA-پروتین تاکید می کند.

پانکلادی و باهلر یک روش پیش بینی را توسعه داده اند که از داده های ازمایشی مختلف نظیر محل پروتین، نیمه عمر RNA، پروفیل ریبوزوم وتحلیل PARS برای پیش بینی اثرات متقابل -MRNA استفاده می کند. با این حال کاربرد این روش برای پیش بینی lncRNA-پروتین اسان نیست زیرا داده های ازمایشی تنها برای lncRNA موجود است.

در همه روش های پیش بینی، داده های ورودی شامل چندین ویژگی های RNA بوده وپروتین برای الگوریتم های یادگیری ماشینی لازم هستند. برای ارزیابی عملکرد پیش بینی با استفادهاز روش های اعتبار ستجی متقابل، مقدار مجموعه داده های کامل به مجموعه داده های اموزشی و ازمایشی برای الگوبرداری تقسیم بندی شد. با این حال پارک و مارکوت نشان داده است که عملکرد پیشبینی شده وابسته به ارزیابی متقابل نامناسب

است و کارایی پیش بینی بهتر زمانی حاصل می شود که مجموعه ازمایشی دارای مولفه های یکسانی به صورت مجموعه اموزشی باشند. پیش بینی اثر متقابل lncRNA- پروتین با استفاده از ابزار های بیوانفورماتیک یک چالش محسوب می شود.

ابزار های تحلیل ساختار های lncRNA

بدیهی است که نه تنها توالی های اصلی RNAغیر کد کننده کارکردی، بلکه ساختار های آن ها ارتباط نزدیکی با نقش آن ها ندارد. اگرچه رابطه بین ساختار و عملکرد برای Rna غیر کد کننده کوتاه اثبات شده است، مطالعات کمی نشان داده اند که ساختار های ثانویه برای کارکرد lncRNA لازم هستند. در این بخش، ابزار های پیش بینی ساختار های lncRNA ارزیابی می شوند.ساختار های ثانویه، ساختار های ثانویه جمعی ، ساختار های ثالث و ساختار های ثانویه مشترک در نظر گرفته می شود.

ساختار ثانویه RNA توالی RNA

ساختار های ثانویه RNA یک شکل خلاصه از ساختار RNA می باشد و با مجموعه ای از جفت باز ها نشان داده می شود. در این زیر بخش، ما به مرور ابزار های پیش بینی های ساختار ثانویه RNA می پردازیم.

پیش بینی ساختار RNA

پیش بینی ساختار ثانویه RNA بر اساس انرژی آزاد ساختار ها می باشند که با استفاده از پارامترهایا انرژی تعیین شده به طور ازمایشی محاسبه می شوند. ابزار های بسیاری برای پیش بینی ساختار های ثانویه RNA از توالی RNA وجوددارد ولی شامل CentroidFold [57, 58], Mfold [59], RNAfold [60], RNAstructure می باشند. پیچیدگی های زمانی روش های مورد استفاده توسط این ابزار ها برابر با $O(L^3)$ است که در آن L طول توالی RNA است.RFOLD موجب کاهش این مسئله با محدود سازی طول جفت باز ها شده و منجر به زمان محاسباتی $O(w^2L)$ می شود. چون $O(w^2L)$ با توجه به طول توالی خطی است، RFOLD برای lncRNA قابل کاربرد است. ابزار های فوق قادر به پیش بینی ساختار ها بدون شبه گره ها می باشند. پیش بینی ساختار های ثانویه RNA با شبه گره ها با استفاده از الگوریتم های دقیق نیازمند هزینه محاسباتی بالا است.LncRNA ها در کنترل چرحه سلول نقش دارند و رونوشت آنتی سنز طبیعی

ANRIL یک تنظیم کننده کلیدی سه ژن مهار کننده تومور p16INK4a, p14ARF و p15INK4b می باشد که همگی از خوشه ژنی CDKN 2 A/B بیان می شوند(24). p16INK4a وp15INK4b از باز دارنده های مهم کیناز4 وابسته به سایکلین می باشند، در حالی که p14ARF به تثبیت HDM1 P53 کمک می کند(25).نقش مهمLncRNA در ارتباط آسیب DNA با اپوپتوزیس سلولی نیز گزارش شده است. آسیب DNA موجب القای LncRNA 5 از پروموتور تنظیم کننده چرخه سلول CDKN1A می شود که یکی از پروموتور های تنظیم کننده چرخه سلول CDKN1A موسوم به PANDA، با فاکتور رونویسیNF-YA با تنظیم کاهشی بیان ژن های دخیل در بهبود آپوپتوزیس تعامل دارد(26). نقش کلیدی دیگر LncRNA در فیزیولوژی طبیعی، مشارکت آن ها در جبران دوز کروموزومی است. در زنان، هر سلول دارای دو رونوشت از کروموزوم X می باشد که یکی از آن ها برای اطمینان از این که دوز صحیح یا مقدار صحیحی از ژن در این کروموزوم قرار دارند، غیر فعال می شوند. غیر فعال سازی X با LncRNA موسوم به Xist کنترل می شود.Xist یکی از نخستین ژن های بیان شده پس از لقاح بوده و PRC1 و PRC2 را هدایت کرده و منجر به تغییرات هیستونی شده و در نهایت همه ژن های موجود در کروموزوم هدف را خاموش می کنند. استفاده از Xist در یکی از کروموزوم ها با سری های دیگر LncRNA کنترل می شمند و شامل آنتی سنز lncRNA Tsix است که فعالیت Xist را بر روی x فعال مهار کرده و Jpx که Xist را بر روی x خاموش، فعال سازی می کند(28).lncRNA هم چنین نقشی حیاتی در کنترل نقش پذیری دارد که در آن یکی از آلل های والدی، به طور اپی ژنتیکی خاموش می شود. احتمالا، بهترین مثال از این RNA غیر کد کننده AIR می باشد که نقش پذیری ژنومی یک منطقه کروموزومی حاوی مکان های ژنی IGF2R/SLC122A2/SLC22A3 را تنظیم می کند.AIR یک خاموش کننده دو سویه بوده و دارای اثر مهار کننده بر روی بیان ژن ها در این منطقه است(29). با توجه به ماهیت فراگیر بیان lncRNA و نقش های کلیدی ان ها در بسیاری از فرایند های فیزیولوژیکی از جمله تنظیم عمومی ژن، بدیهی است که آن ها در علت شناسی بیماری های انسانی نقش دارد.

پیش بینی های ساختار RNA پروبینگ

اخیرا، روش های ازمایشی مختلف برای فنون تعیین ساختار RNAپیشنهاد شده است. این فنون شامل de FragSeq [67], SHAPE-seq [68], DMS-seq [69, 70] Mod-seq [71], MAP-seq [72], and

PARS [73]. (See [74] می باشند. متاسفانه این فنون ازمایشی شامل تنها اطلاعات مربوط به ساختار های ثانویه است.

ساختار های ثانویه توالی های RNA

ساختار های ثانویه RNA مربوط به یک کارکرد خاص به طور تکاملی حفظ شده و و شناسایی این ساختار های حفاظتی گامی مهم برای تحلیل کارکردیlncRNA محسوب می شوند. در پیش بینی ساختار های ثانویه، تغییر هم زمان باز ها باری حفظ جفت باز لازم است. همان طور که در مطالعه پویایی تکاملی lncRNA در شش پستاندار نشان داده شده است اطلاعات تکاملی در خصوص lncRNA برای پیش بینی ساختار های lncRNA لازم است. ابزار هایی نظیر CENTROID ALFOLD و RNA ساختار های ثاویه را از ارایشات متعدد توالی RNA پیش بینی می کنند. به علاوه روش های پیش بینی ساختار ثاتویه RNA مبتنی بر پروب برای پیش بینی ساختار اجماع لازم است.PPFOLD از داده های SHAPE برای پیش بینی ساختار های ثانویه استفاده می کند. ورودی این ابزار ها برای پیش بینی ساختار ثانویه شامل اریشات چندگانه توالی های RNA است که با استفاده ازابزار های سنتی نظیر PROBCONS و سایر ابزار های پیشرفته با در نظر گرفتن ساختار های ثانویه نظیر MAFFT [86], LocARNA [87], LARA [88], MXSCARNA [89] CentroidAlign [90] بدست می اید. محاسبات $O(L^3)$ برای این ابزار ها نیازهستند ولی برای lncRNA پرهزینه می باشند. با این حال به روز رسانی اخیر CENTROIDاز TRFOLD در بخش 5-1-1 استفاده کرده و از CENNTROIDALIGN برای توالی های RNA استفاده میشود.

ساختار های ثانویه مشترک بین دو توالی RNA

همان طور که در بخش 4 توصیف شده است، بسیاری از lncRNA با سایر مولکول ها در سلول اثر متقابل داشته و پیش بینی کمپلکس های شامل lncRNA برای تحلیل کارکردی lncRNA مفید است. پیش بینی ساختار های ثانویه مشترک دو توالی RNA متشکل از پیش بینی جفت باز ها است. با این حال پیش بینی همزمان هر دو سخت و غیر ممکن می باشد. با این وجود، چندین ابزار وجود دارد.PACTIP قادر به پیش بینی

ساختار های ثانویه مشترک ازجمله جفت باز ها می باشد در حالی که INTARNA جفت باز های مولکولی را کنار می گذارد.

پیش بینی ساختار ثالث

پیش بینی ثالث توالی های RNA نیازمند منابع محاسباتی بیشتر از پیش بینی ساختار ثانویه RNA است. به 93 برای مرور ابزار های موجود در پیش بینی ساختار ثالث مراجعه کنید. از این روی پیشبینی ساختار ثالث برای کلlncRNA غیر ممکن است.

ابزار ها یا روش های دیگر مربوط به ساختار های ثانویه RNA

مطالعات اخیر بر اهمیت پلی مورفیسم تک نوکلوتیدی تاکید دارند که موجب تغییر ساختار های ثانویه RNA موسوم به RiboSNitches می شود و از عوامل اصلی بیماری است.RCHANGE و ابزار های دیگر برای تحلیل جهشی lncRNA استفاده می شوند.

تحلیل موتیف های ساختاری RNA

به منظور شفاف سازی کارکرد ها و نقش lncRNA با توجه به ساختار های ثانویه، کشف موتیف های ساختاری ثانویه که با کارکرد ها همبستگی دارند اهمیت دارد. CAPR، پروفیل ساختاری دقیق را محاسبه می کند.[99] MEMERIS [98] RNAcontext موتیف های ساختراى ثانویه در میان توالی های RNA دارای وظایف و نقش های مشابه می باشند. به علاوه، دسترسی به توالی های lncRNA با استفاده از RACCESS محاسبه شده و در پیش بینی اهداف MIRNA و مولکول های با lncRNA محاسبه می شود.

خلاصه ای از این بخش

در این بخش، ما به مرور ابزار های مفید برای پیش بینی ساختار lncRNA می پردازیم. به دلیل محدودیت های مکانی، از این روی روش های دقیق برای هر ابزار ارایه نشده است. مطالعه اخیر ادی 101 یک ابزار خوب برای پروب و بررسی پیش بینی های ساختار ثانویه RNA است. به علاوه دو کتاب در زمینه روش های بیولوژی مولکولی برای مطالعه ساختار های lncRNA استفاده شده اند.

دیتابیس های lncRNA

شواهد نوظهور در خصوص lncRNA های شناسایی شده و ویژگی های زیست شناختی آن ها از جمله ویژگی خواص، از جمله ویژگی های ژنومی، پروفایل های بیان، حفاظت از توالی، اثر متقابل ماکرو مولکولی، تغییرات اپی ژنوم و حاشیه نویسی های کاربردی، باید به پایگاه داده های عمومی به عنوان یک منبع بدست امده اند.

[104] LNCipediaوlncRNome تعداد زیادی از ورودی های lncRNA انسانی از جمله توالی های اصلی و ساختار های ثانویه پیش بینی شده را ارایه کرده اند. در LNCipedia، پتانسیل های کد کننده پروتین lncRNA با استفاده از ابزار های بیوانفورماتیک ارزیابی شده است . به علاوه،LNCipedia ، اثرات متقابل پیش بینی شده lncRNA - RNA برای پلی مورفیسم های نوکلوتیدی و اصلاحات اپی زنومیک ژن های در lncRNA در نظر گرفته می شوند.

پروفیل های بیان lncRNA بر گرفته شده از لاین های سلولی و بافت های انسانی در LNCRNADB و LNCRNATOR می باشند. مورد اول سه قطعه منحصر به فرد از اطلاعات را در خصوص داده های بیان lncRNA ارایه می کند. سطوح بیان lncRNA در بافت های سرطان بر گرفته از داده های توالی RNA در اطلس ژنوم سرطان ، تحلیل lncRNA و ژن های کد کننده پروتین و تحلیل غنی سازی هستی شناسی ژن در نظر گرفته شده است. lncRNA می تواند توصیف دقیقی را از lncRNA ها ارایه می کند. اثر متقابل پروتین- lncRNA از مطالعات مختلف توالی CLIP توسط LNCRNOME و LNCRNATOR ارایه شده است. الگوهای بیان مکانی زمانی(بافت ها، بخش های زیر سلولی و مراحل تمایز و رشد) lncRNAt اطلاعات اساسی برای درک کارکرد های بیولوژیکی lncRNA در سلول ها میباشند. چندین مطالعه توالی RNA وجود دارد که پروفیل های بیان ژن های کد کنندهو غیر کد کننده در بافت های مختلف بدست اورده است. با این حال استخراج الگوهای بیان از دادههای توالی RNA نیازمند منابع محاسباتی بزرگ مقیاس برای فیلترینگ با کیفیت و کمی سازی بیان ژن است.EXPRESSION ATLAS ، داده های توالی RNA را با استفاده از خط لوله محاسباتی اصلی پردازش کرده و پروفیل های بیان بافت ها و لاین های سلولی مختلف از 16 گونه را ارایه کرده است. پروفیل های بیان lncRNA برای این سه گونه ارایه شده است. به ویژه، برای انسان، ATLAS پروفیل های بیان lncRNA را بر گرفته از 32 بافت طبیعی و 675 بافت سرطانی ارایه کرده است.پروفیل های بیان ژن های کد کننده پروتین از چندین بافت مهره داران دیگر نظیر مانند اپوسوم، میمون رشت، بابیان زیتون، مرغ، گاو،

فلفل، قورباغه و مارمولک جو و برنج نیز در EXPRESSION ATLAS موجود است با این حال پروفیل های بیان lncRNA به دلیل نبود تفسیر، قابل دسترس نیست. اخیرا CTEX CONSTRTIUM مجموعه ای از داده های توالی RNA را از 43 بافت انسان برای 53934 ژن ارایه کرده است که در آن پروفیل های بیان lncRNA واقع شده اند.

Species	Sample sources	Expression Atlas ID	# of tissues	# of cell lines	GEO/ENA/EGA ID	Reference
Human	Normal tissues	E-MTAB-2836	32	–	ERP006650(ENA)	[17]
		E-MTAB-513	16	–	GSE30611(GEO)	[18]
		E-GEOD-30352	8	–	GSE30352(GEO)	[19]
	Cancer cell lines	E-MTAB-2706	–	675	EGAS00001000610(EGA)	[20]
		E-MTAB-2980	–	39	PRJNA169425(ENA)	–
	Subcellular compartments	E-GEOD-26284	–	15	GSE26284(GEO)	[21]
Mouse	Normal tissues	E-MTAB-2801	9	–	GSE41637(GEO)	[22]
		E-MTAB-599	6	–	GSE30617(GEO)	[23]
		E-GEOD-30352	6	–	GSE30352(GEO)	[19]
		E-GEOD-41338	5	–	GSE41338(GEO)	[24]
Rat	Normal tissues	E-GEOD-53960	11	–	GSE53960(GEO)	[25]
		E-MTAB-2800	9	–	GSE41637(GEO)	[22]

جدول1

Table 2: Summary of available public data for structural probing techniques in combination with high-throughput sequencing

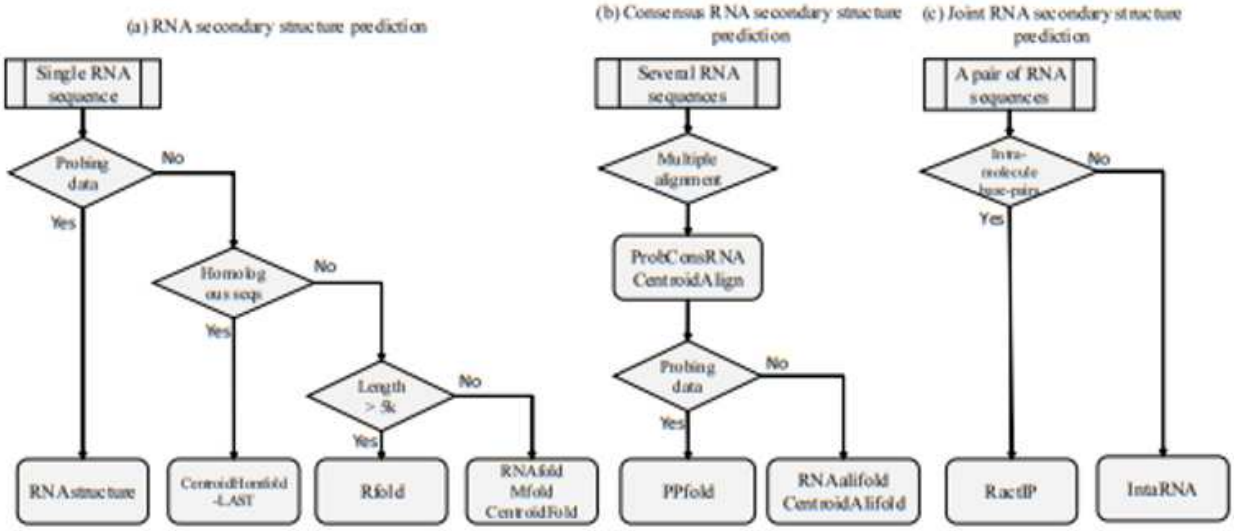
Chemical	Method	Species	Target	GEO	SRA	Platform	read type	length	Ref
DMS	structure-seq	A.thaliana	total RNA (polyA)	–	SRP027216	HiSeq2000	single	40	[69]
DMS	DMS-seq	S.cerevisiae H.sapiens	total RNA (polyA)	GSE45803	SRP020556	HiSeq2000	single	30-50	[70]
DMS	Mod-seq	S.cerevisiae	total RNA	–	SRP029192	HiSeq2000	single	50	[71]
DMS/CMCT	CIRS-seq	M.musculus	total RNA	GSE54106	SRP035427	HiScanSQ	single	50	[112]
RNase P1	Frag-seq	M.musculus	total RNA	GSE24622	SRP004883	SOLiD 3.0	single	50	[67]
RNase V1,S1	PARS-seq	S.cerevisiae	total RNA (polyA)	GSE22393	SRP003175	SOLiD	single	50	[113]
RNase V1,S1	PARS-seq	H.sapiens	total RNA (polyA)	GSE50676	SRP029656	HiSeq	pair	101(*)	[73]
1M7	SHAPE-seq	B.subtilis	total RNA	GSE31573	SRP007955	GAIIx	pair	50	[114]
1M6,1M7,NMIA	SHAPE-seq	HIV-1	genome (RNA)	–	SRP042347	HiSeq MiSeq	pair/single		[115]

جدول2

Table 3: Summary of lncRNA databases

Database	LNCIPEDIA	LNCRNOME	LNCRNADB	LNCRNATOR
# of lncRNA	111,685	113,513	287	34,605
Species	human	human	human, mouse, rat, zebrafish, fly, yeast, chicken, etc.	human, mouse, zebrafish, fly, worm, yeast
Sequence	Y	Y	Y	N
Secondary structure	Y	Y	N	N
Genomic position	Y	Y	Y	Y
Coding potential	CPC, HMMER, PRIDE, PhyloCSF, Ribosome-profiling	PhyloCSF	-	CPC
Conservation	Ensembl Compara	Phastcons	literature	Phastcons
Orthologue	N	N	Y	Y
Expression profile	N	N	Y (Human Body Atlas)	Y (GEO, ENCODE, modENCODE, TCGA)
# of human tissues[1]	-	-	16	10 (cancer)
RNA–protein[2]	N	Y (exp., pred.)	Y (literature)	Y (exp.)
RNA–RNA[2]	Y (pred.)	N	Y (literature)	N
SNP	N	Y	N	N
Epigenome[3]	N	Y	N	N
GO term	N	N	N	Y
Link to references	Y	Y	Y	N
External links	UCSC genome, PubMed, Ensembl	HGNC, OMIM, NCBI, PubMed, Ensembl	PubMed, NCBI	UCSC genome, Ensembl
Reference	[104]	[105]	[110]	[111]

جدول3



شکل1

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