The 4th Iranian Conference on
Systems Biology

شناسایی نواحی ویرایش RNA در بیماران مبتلا به سرطان معده با استفاده از داده های توالی‌یابی RNA

جواد بهروزی[۱]*، شیرین شهبازی[۱]، محمدرضا بختیاری‌زاده[۲]، مهرداد نصرالله زاده ثابت[۳]، حبیب‌الله محمودزاده[۴]

۱-گروه ژنتیک پزشکی، دانشکده پزشکی، دانشگاه تربیت مدرس، تهران، ایران
۲-گروه علوم دام و طیور، پردیس ابوریحان، دانشگاه تهران، تهران، ایران
۳- گروه فناوری های نوین و ژنتیک، دانشکده پزشکی، دانشگاه علوم پزشکی ارتش، تهران، ایران
۴-گروه جراحی سرطان، انستیتو کانسر، مجتمع بیمارستانی امام خمینی، دانشگاه علوم پزشکی تهران، تهران، ایران
*آدرس پست الکترونیک نویسنده مسئول: j.behroozi@modares.ac.ir

چکیده

ویرایش RNA به مفهوم تغییر در توالی RNA پس از رونویسی می باشد، که توسط اعضای خانواده ADAR انجام می شود. این تغییر توالی در اغلب موارد به صورت تبدیل نوکلئوتید A به I از طریق دآمینه شدن آدنوزین به اینوزین بر روی RNA دو رشته ای است. ویرایش RNA از طریق چندین مکانیسم می تواند در سرطان نقش داشته باشد. این فرآیند با تغییر کد ژنتیکی می تواند در ژن های سرگوبگر تومور و یا انکوژن ها اتفاق افتاده و باعث ایجاد جهش در سطح RNA گردد. همچنین افزایش یا کاهش ویرایش RNA در میکروRNAها و یا محل های اتصال آن ها، می تواند منجر به کاهش یا افزایش بیان انکوژن ها و ژن های سرکوبگر تومور شود. شناسایی جایگاه های ویرایش RNA می تواند از طریق مقایسه مستقیم توالی cDNA و DNA ژنومی مربوط به آن انجام شود. روش دیگر برای شناسایی کل جایگاه های ویرایش (ادیتوم) ، استفاده از توالی یابی RNA به صورت تنها می باشد. چندین چالش در رابطه با استفاده از داده های توالی یابی RNA جهت آنالیز ادیتوم وجود دارد که شامل؛ تمایز جایگاه های ویرایش واقعی از SNPها، جهش های سوماتیکی، خطاهای سیستماتیک توالی یابی و خطاهای نقشه یابی می شوند. هدف از انجام مطالعه حاضر استفاده از داده‌های توالی یابی RNA جهت شناسایی نواحی ویرایش شده ژنوم در بیماران مبتلا به سرطان معده می باشد.

کلمات کلیدی: بیوانفورماتیک، داده‌کاوی، ویرایش RNA، توالی یابی RNA، سرطان معده

مواد و روش ها

به منظور بررسی جایگاه های ویرایش در سرطان معده از داده های توالی یابی RNA موجود در بخش GEO بانک اطلاعاتی مرکز ملی بیوتکنولوژی استفاده شد. داده‌های مرتبط با مطالعات بیان ژن در این پایگاه داده ذخیره شده است و بخش اعظم آن‌ها در دسترس همگان قرار دارند. برای اینکار داده‌های توالی‌یابی RNA مربوط به ۸ جفت نمونه آدنوکارسینوم معده به همراه بافت سالم مجاور بازیابی گردید. توالی‌یابی این داده‌ها با استفاده از تکنولوژی Illumina Hiseq 2000 انجام شده است و خوانش‌ها از دو انتها و با طول ۱۰۱ جفت باز می باشند.

پس از بدست آوردن داده های توالی یابی در ابتدا از نرم‌افزار FastQC جهت کنترل کیفیت بازها و خوانش‌ها، همچنین بررسی آلودگی به آداپتور استفاده شد. در ادامه بازها و خوانش‌های با کیفیت پایین و همچنین آلودگی‌های مربوط به آداپتور در خوانش‌ها با استفاده از نرم‌افزار Trimmomatic حذف گردید. جهت مکان یابی خوانش‌های پیرایش شده بر روی ژنوم مرجع از نرم‌افزار Hisat2 استفاده شد. پس از همردیف کردن خوانش‌ها با ژنوم مرجع، خوانش‌های دو نسخه ای که در نتیجه PCRبوجود آمده اند، با استفاده از نرم‌افزار Picard حذف شدند. جایگاه های ویرایش RNA را می توان به عنوان ورایانت های تک نوکلئوتیدی که حاصل تفاوت در توالی RNA با ژنوم مرجع هستند در نظر گرفت، البته این SNVها می توانند SNPهای غیرشایع و مخصوص به فرد هم باشند که باید فیلتر شوند. برای شناسایی SNVها از نرم افزار HaplotypeCaller استفاده شد. واریانت های بدست آمده با SNPهای موجود در آخرین نسخه از پایگاه داده dbSNP مقایسه شده و تمامی SNPها از میان SNVهای شناخته شده حذف گردید تا نواحی احتمالی ویرایش RNA شناسایی شود. در ادامه چندین فیلتر مختلف با نرم افزار GATK بر روی جایگاه های احتمالی ویرایش انجام شد تا دقت شناسایی افزایش پیدا کند. جهت بررسی و مقایسه میزان ویرایش در بافت های سالم و سرطانی از آزمون آماری T جفت‌شده استفاده شد و P-value کمتر از ۰/۰۵ به عنوان سطح معنی‌داری در نظر گرفته شد.

نتایج

در مطالعه حاضر بررسی داده های توالی یابی RNA مربوط به بیماران مبتلا به سرطان معده منجر به شناسایی جایگاه های ویرایش در ۱۲،۳۲۸نقطه از ژنوم شد. جدول ۱ توزیع ویرایش های مختلف را در نواحی مختلف ژنوم نشان می دهد. همانطور که مشاهده می‌شود بیشترین تعداد ویرایش های شناسایی شده در ناحیه ترجمه نشونده انتهای '۳ و کمترین میزان ویرایش در ناحیه ترجمه نشونده انتهای '۵ قرار دارند. نقطه قابل توجه وجود ۱۵۸ ناحیه ویرایش در نواحی اگزونی می باشد که ممکن است منجر به ایجاد جهش های نامترداف در این نواحی گردند. بررسی مقایسه ای ویرایش های موجود در بافت های سرطانی و سالم نیز نتایج قابل توجهی داشت. این بررسی ها نشان داد که از بین کل ویرایش های شناسایی شده، ۳۵۰۹ مورد مختص به بافت تومور و ۳۹۸۵ مورد مختص به بافت سالم می باشد. همچنین ۴۸۶۸ مورد از ویرایش ها به صورت مشترک در هر دو بافت وجود داشتند. آنالیز آماری ویرایش های مختص به بافت سرطانی نشان داده که میزان ویرایش در ۱۲۹ جایگاه دارای تفاوت معنی دار می باشد.

نواحی ژنومی	تعداد ویرایش های شناسایی شده
نواحی بالا دست ژن ها	۲۸۹۹
ناحیه ترجمه نشونده انتهای '۵	۵۷
ناحیه ترجمه نشونده انتهای '۳	۵۸۷۰
نواحی اینترون	۶۷۵
نواحی اگزونی	۱۵۸
نواحی پایین دست ژن ها	۲۵۴۷
نواحی بین ژنی	۱۲۲

بحث، نتیجه گیری و پیشنهادات

با توجه به وجود منابع عظیمی از داده های مربوط به توالی یابی نسل جدید در پایگاه های داده مختلف، گروه های زیادی در صدد استفاده از این داده ها و استخراج اطلاعات مناسب از آن هستند. از جمله Zhu و همکاران از داده های RNA-seq موجود در پایگاه داده GEO استفاده کرده و تعداد ۶۹۵ جایگاه ویرایشA به I را شناسایی نمودند (Zhu et al., 2013). در روشی مشابه Wang و همکارانش از داده های RNA-seq موجود در پایگاه TCGA جهت شناسایی جایگاه های ویرایش استفاده نمودند و از رده های سلولی جهت تایید نتایج خود استفاده کردند (Wang et al., 2017). در سال های اخیر پیشرفت تکنولوژی های توالی یابی منجر به کشف این نکته گردید، که بیشتر وقایع ویرایش در نواحی غیر کد کننده و به صورت خوشه ای انجام می شود. در واقع بیش از نصف ترانسکریپتوم انسانی تحت تاثیر ویرایش قرار می گیرد (Bazak et al., 2014). مطالعه ما نیز هم راستا با مطالعات قبلی بود و نشان داد که در حدود ۵۰ درصد از ویرایش ها در ناحیه ترجمه نشونده انتهای '۳ اتفاق می افتد. این یافته از این جهت مهم است که توالی‌های هدف میکرو RNAها نیز در ناحیه ترجمه نشونده انتهای '۳ قرار دارد و هر گونه تغییر در این ناحیه از mRNA ممکن باعث تغییر در اتصال میکرو RNAها و در نتیجه تغییر در بیان ژن هدف شود. به طوری کلی مطالعه حاضر نشان داد می توان با تکیه بر ابزارهای بیوانفورماتیکی اطلاعات جدیدی را در رابطه با زیست شناسی سرطان بدست آورد.

منابع

1) Bazak, Lily, Ami Haviv, Michal Barak, Jasmine Jacob-Hirsch, Patricia Deng, Rui Zhang, Farren J Isaacs, et al. "A-to-I Rna Editing Occurs at over a Hundred Million Genomic Sites, Located in a Majority of Human Genes." Genome research 24, no. 3 (2014): 365-76.
2) Wang, Yumeng, Xiaoyan Xu, Shuangxing Yu, Kang Jin Jeong, Zhicheng Zhou, Leng Han, Yiu Huen Tsang, et al. "Systematic Characterization of a-to-I Rna Editing Hotspots in Micrornas across Human Cancers." Genome research 27, no. 7 (2017): 1112-25.
3) Zhu, Shanshan, Jian-Feng Xiang, Tian Chen, Ling-Ling Chen, and Li Yang. "Prediction of Constitutive a-to-I Editing Sites from Human Transcriptomes in the Absence of Genomic Sequences." BMC genomics 14, no. 1 (2013): 206