دانشگاه علوم کشاورزی و منابع طبیعی ساری
پژوهشهای تولیدات دامی

شناسایی RNAهای غیرکدکننده کوتاه عملکردی با استفاده از روش‌های بیوانفورماتیکی در گوسفند و بز

عاطفه سیددخت[۱]، حامد احمدی[۲]، علی‌اصغر اسلمی‌نژاد[۳]، علی مسعودی‌نژاد[۴]، محمدرضا نصیری[۵] و بلال صادقی[۶]

۱ و ۵- دکتری و استاد، دانشگاه فردوسی مشهد
۲ و ۴- دکتری و استاد، دانشگاه تهران
۳- دانشیار، دانشگاه فردوسی مشهد، (نویسنده مسوول: agr764@yahoo.co.uk)
۶- استادیار، دانشگاه شهید باهنر کرمان

تاریخ دریافت: ۹۳/۱۰/۵ تاریخ پذیرش: ۹۴/۳/۳۰

چکیده

microRNAها، RNAهای غیرکدکننده‌ای هستند که در تنظیمات پس از رونویسی نقش‌های مهمی را بازی می‌کنند. آن‌ها بیان ژن‌ها را به‌وسیله مسیرهای تداخلی RNA، از طریق برش یا مهار ترجمه mRNA هدف تنظیم می‌کنند. نقش‌های مهمی برای miRNAها در روند تکاملی حیوانات مختلف گزارش شده است، اما اطلاعات محدودی در مورد miRNAهای گوسفند و بز وجود دارد. گوسفند و بز مدلی مناسب برای مطالعات ژنومیک مقایسه‌ای و بیولوژیکی در نشخوارکنندگان هستند. شناسایی miRNAها برای درک مکانیسم بیولوژیکی آن‌ها بسیار حیاتی است. روش‌های شناسایی محاسباتی، روش‌های آزمایشگاهی را برای شناسایی سریع‌تر RNAهای غیرکدکننده در ژنوم‌های جدید، که اغلب تحت شرایط ویژه در انواع مختلفی از سلول‌ها نسخه برداری می‌شوند را تکمیل می‌کنند. اخیراً روش‌های یادگیری ماشین برای پیش‌بینی ژن‌های miRNA جدید استفاده می‌شود. در این پژوهش یک طبقه‌بندی‌کننده جدید بر اساس SVM معرفی شد. این طبقه‌بندی‌کننده دقت بالا، حساسیت متعادل و اختصاصیت برای مجموعه داده miRNA را نشان می‌دهد که ابزاری ایده‌آل، برای شناسایی miRNAها در داده‌های ترانسکریپتوم محسوب می‌شود. ما در این پژوهش یک زیرمجموعه از ویژگی‌های بهینه شامل ۲۰ ویژگی را با استفاده از ماشین بردار پشتیبان (SVM) ایجاد کردیم. سپس با استفاده از برنامه #C، ویژگی‌های استخراج شده برای داده‌های آموزشی محاسبه شد. در این پژوهش، مدل هوشمند ماشین بردار پشتیبان با کرنل RBF و الگوریتم یادگیری SMO بهترین دسته‌بندی کننده برای پیش‌بینی ژن‌های microRNA در گوسفند و بز معرفی شد. حساسیت و اختصاصیت این مدل به ترتیب ۸۸ و ۸۵ درصد به‌دست آمد. سپس یک روش محاسباتی بر اساس آنالیز EST برای تشخیص miRNAهای بالغ گوسفند و بز مورد استفاده قرار گرفت. در کروموزوم یک گوسفند، ۲۳ کاندید miRNA و در بز ۱۵ miRNA از جستجوی هم‌سانی شناسایی شد. نتایج نشان داد که مدل مورد استفاده ما می‌تواند دقت بالایی در شناسایی miRNAهای گوسفند و بز ارائه دهد و نیز اطلاعات مفیدی را در زمینه عملکرد بیولوژیکی آن-ها در گونه‌های نشخوارکنندگان فراهم آورد.

واژه‌های کلیدی: نشخوارکنندگان، microRNA پیش‌ساز، ماشین بردار پشتیبان، گوسفند و بز

مقدمه

microRNAها، RNAهای غیرکدکننده‌ای هستند، که حدود ۲۲ نوکلئوتید طول دارند و می‌توانند بیان ژن‌های هدف را به وسیله اتصال به سایت‌های مکمل تنظیم کنند. نشان داده شده است که miRNAها برای کاهش بیان ژن‌های هدف با اتصال به سایت‌های مکمل‌شان در رونوشت‌ها باعث تخریب رونوشت‌ها یا سرکوب ترجمه می‌شوند (۴۳،۶). با این حال، مطالعات اخیر نشان داده است که miRNAها می‌توانند با اتصال به توالی‌های پروموتور مکمل، ترجمه پروتئین را افزایش دهند (۴۷). miRNAها در موجودات مختلف در بسیاری از فرآیندهای تنظیم ژنی مانند رشد و نمو، غیرفعال‌سازی ژن‌های انتقال یافته، مسیرهای سیگنالینگ سلولی، تنش‌های خارجی، پروتئین‌های مرتبط با سرطان و دفاع در برابر حمله ویروس‌ها درگیر هستند (۲۶).

اولین microRNA در سال ۱۹۹۳ توسط ویکتور اِم بروس در کرم نماتود سی الگانس شناسایی شد (۳۵). به طور هم‌زمان وایتمن در آزمایشگاه گری روکان نیز ژنی را که تحت تأثیر این microRNA قرار داشت، شناسایی نمود (۵۲). البته نتایج این دو مطالعه هر چند که سرآغازی برای شروع یک تحول بزرگ در یافته‌های علم ژنتیک بود اما تا هفت سال بعد مورد توجه جدی قرار نگرفتند، زیرا که انتظار می‌رفت microRNA کشف شده تنها یک مولکول استثنائی باشد که فقط در گونه سی الگانس یافت شده و هیچ ایده و شاهدی دال بر وجود این مکانیسم تنظیم بیان ژن در بین موجودات مختلف وجود نداشت. بعدها در سال ۲۰۰۴ تاریخچه کشف این مولکول‌ها توسط محققین اولیه آن به صورت مقالاتی به چاپ رسید. نکته جالب توجه این است که حدود چهل سال قبل دو دانشمند به نام‌های جاکوب و موناد، وجود microRNAها را پیش‌بینی کرده بودند. این دو برای پدیده مهار لک[۱]، مکانیسمی مشابه مکانیسم microRNA را پیشنهاد داده بودند (۳۴،۱۶،۱۴،۱). موج گسترده شناسایی microRNAها در سال ۲۰۰۱ اتفاق افتاد که تعداد بسیار زیادی microRNA جدید در مگس سرکه، سی الگانس و سلول‌های پستانداران معرفی شدند (۳۳،۳۲،۳۱). امروزه حتی در ویروس‌هایی مانند اپستین‌بار و چندین هرپس ویروس نیز microRNAهایی شناسایی شده‌اند (۵۵). microRNA اولیه

1- Lac

دارای چند بخش ویژه است که برای پردازش هسته‌ای مهم می‌باشد. این بخش‌ها شامل ساقه، نواحی جانبی یا بازوها و حلقه انتهایی می‌باشد. ساقه و بازوها برای اتصال DGCR8 به microRNA و برش آن توسط دروشا ضروری هستند (۵۷،۵۸،۵۹). دروشا در فاصله ۱۱ جفت بازی از محل اتصال بخش تک‌رشته‌ای microRNA اولیه با بخش دورشته‌ای آن و در فاصله‌ای معادل دو چرخش مارپیچ از حلقه انتهایی، microRNA اولیه را برش می‌دهد (۲۴،۱۳). به طور کلی مکانیسم‌های کاهش بیان ژن توسط microRNA را می‌توان به چهار گروه حذف دم پلی‌آدنین، ایجاد اختلال در فاکتورهای آغاز ترجمه، قطع تداوم ترجمه با ایجاد اختلال در پلی ریبوزوم‌ها و تجزیه پروتئین در حال ساخت تقسیم‌بندی نمود (۵۶،۲۹،۱۷). اساساً دو نوع روش برای شناسایی miRNAها وجود دارد. روش اول استفاده از کتابخانه cDNA برای توالی‌هایی با اندازه متفاوت می‌باشد. با این روش می‌توان انوع miRNAهای حفاظت شده و غیرحفاظت شده را شناسایی کرد. اما این روش برای miRNAهایی که به میزان بسیار کمی بیان می‌شوند و در یک مرحله خاصی از رشد سلول یا در سلول بافت خاصی بیان می‌شوند، مناسب نیست. روش دیگر تکنولوژی توالی‌یابی با توان بالا است که یک ابزار قدرتمند برای کشف RNAهای کوتاه و miRNAها محسوب می‌شود (۳۹).

برای حل این مسئله، اولین مرحله در پژوهش روی microRNAها، شناسایی ژن‌های microRNA است. این مسئله شامل پیمایش روی ژنوم گوسفند و بز و شناسایی نواحی از آن است که به‌منظور تولید miRNAها نسخه‌برداری می‌شوند. با در نظر گرفتن اهمیت ژن‌های microRNA در مکانیسم‌های مختلف سلولی که به آن‌ها اشاره شد، مسلماً شناسایی و تعیین جایگاه آن‌ها در دام‌های اهلی نظیر گوسفند و بز به منظور درمان بیماری‌های ژنتیکی ویژه، بهبود صفات پرواربندی و غیره اهمیت می‌یابد (۲۳،۲۵،۲۷،۴۲،۴۴).

به منظور جلوگیری از انجام آزمایشات تعیین سطوح بیان برای پیدا کردن miRNAهای جدید، استفاده از روش‌های محاسباتی با استفاده از یادگیری ماشین به‌دلیل دارا بودن دقت پیش‌بینی بالا به طور موفقیت‌آمیزی برای تشخیص miRNAها بر اساس ویژگی‌های توالی و ساختاری آن‌ها ضروری به نظر می‌رسد.

برای شناسایی miRNAها توسط مدل‌های هوشمند ابتدا لازم است نمونه‌های مثبت و منفی تأیید شده در آزمایشگاه جمع‌آوری یا تولید شوند. مرحله بعد استخراج ویژگی‌ها می‌باشد. پس از استخراج ویژگی‌ها، نوبت آموزش مدل‌ها و یافتن بهینه‌ترین پارامترهای آزاد آن‌ها می‌باشد. در نهایت نتایج باید با یک‌دیگر مقایسه و تحلیل شوند. این پژوهش به منظور دسترسی به اهداف زیر طرح‌ریزی شد: ۱- شناسایی ژن‌های microRNA در گوسفند و بز. ۲- ارائه یک روش بیوانفورماتیکی برای پیش‌بینی اولیه ژن‌های microRNA و شناسایی ژن‌های مذکور. ۳- دسترسی به توالی‌های بالغ ژن‌های miRNA.

مواد و روش‌ها

به‌طورکلی، فعالیت‌های انجام شده در این پروژه را می‌توان به دو قسمت مختلف تقسیم‌بندی کرد. بخش اول: در این قسمت، توالی‌های miRNA برای دو گونه گوسفند و بز از پایگاه داده mirBase (به‌عنوان داده‌های مثبت) و توالی‌های غیر miRNA (به‌عنوان داده‌های منفی) از بانک جهانی ژنی NCBI دانلود شدند. این دو گروه داده برای آموزش مدل ماشین بردار پشتیبان[1] (SVM) برای پیش‌بینی ژن‌های miRNA استفاده شد. در مرحله بعد، ویژگی‌های انتخاب شده (۵۰) برای توالی‌های miRNA جمع‌آوری شد (جدول ۱). سپس به‌منظور استخراج ویژگی‌ها، نرم‌افزار Feature Extraction با زبان C# از سوی نویسندگان این مقاله تولید شد. ویژگی‌های استخراج شده از داده‌های مثبت و منفی، توسط مدل هوشمند ماشین بردار پشتیبان (SVM) با کرنل RBF و الگوریتم یادگیری SMO در نرم‌افزار MATLAB مورد آموزش قرار گرفت. بخش دوم: در این مرحله چگونگی آنالیز microRNA با استفاده از پایگاه‌های آنلاین بیان شده است. این دو بخش به ترتیب برای دو هدف دسترسی به ژن‌های pre-miRNA (بخش اول) و دسترسی به توالی‌های بالغ ژن‌های miRNA (بخش دوم) پایه‌ریزی شدند. مرحله اول قسمت برنامه‌نویسی کار و تولید نرم‌افزار بود. برای انجام این بخش نکات اصلی زیر مدنظر قرار گرفت:

بر اساس پایگاه پژوهشگاه علوم و فناوری اطلاعات ایران (http://www.irandoc.ac.ir)، این پروژه اولین پژوهش در زمینه شناسایی ژن‌های miRNA در گوسفند و بز می‌باشد. در این پروژه برای اولین بار به منظور شناسایی ژن‌های miRNA، مدل ماشین بردار پشتیبان[1] (SVM) در نرم‌افزار MATLAB آموزش داده شد و برای این مدل، ساختارها و پارامترهای بهینه آن استخراج شد.

نرم‌افزار

برای استخراج ویژگی‌ها یک نرم‌افزار کاربرپسند با زبان C# (www.msdn.microsoft.com) در محیط .Net به نام Feature Extraction توسط نویسندگان این مقاله نوشته شد. این نرم‌افزار به کاربر اجازه می‌دهد توالی‌های خود را به دلخواه تنظیم کند. با استفاده از این نرم‌افزار کاربر قادر است رشته‌های خود را وارد سیستم کرده و نواحی دارای پتانسیل را شناسایی نماید. این قابلیت برای افرادی که روی سنتز miRNAها پژوهش می‌کنند، بسیار مفید است.

پایگاه اطلاعاتی

با توجه به ماهیت راه حل ارائه شده برای پیش‌بینی ژن‌های miRNA که در واقع استفاده از یک مدل با نظارت می‌باشد، لازم بود داده‌های آموزشی، در آزمایشگاه تأیید شده باشند. علاوه بر این داده‌های تأیید شده آزمایشگاهی نیز باید مورد بررسی قرار می‌گرفت تا در نهایت مجموعه داده قابل اطمینانی به‌دست آید. اکنون با تهیه این مجموعه داده راه برای پژوهش‌های بعدی در این زمینه هموارتر شده و پژوهش‌گران می‌توانند با سرعت بیشتری در این زمینه تحقیق کنند. مسئله دیگر روند سریع کشف miRNAهای جدید

1- Support Vector Machine (SVM)

است. در این پروژه نقاط با پتانسیل بر اساس آخرین نسخه پایگاه اطلاعاتی mirBase (نسخه بیست و یک) تهیه و برای استفاده دیگران ذخیره شده است.

جمع‌آوری داده‌ها

برای آموزش مدل با نظارت ماشین بردار پشتیبان که در این پژوهش از آن استفاده شده است، نیاز به یک مجموعه داده آموزشی می‌باشد. برای تهیه داده‌ها از پایگاه داده miBase (http://www.mirbase.org) استفاده شد.

نکته دیگر، گونه داده‌های استفاده شده می‌باشد. این پژوهش با توجه به جهت‌گیری گونه‌های نشخوارکنندگان، روی گونه گوسفند و بز تمرکز کرده است.

تولید نمونه‌های مثبت و منفی

برای تولید داده‌های مثبت از ۳۷۳ توالی ثبت شده در پایگاه miRBase استفاده شد. این توالی‌ها با استفاده از نرم‌افزار RNAfold (۲۱) برای ساختار ثانویه، ساختار سنجاق سری و لوپ‌ها مورد بررسی قرار گرفت. برای تولید داده‌های منفی از چند مجموعه توالی (tRNAs, piRNAs snRNAs, snoRNAs) استفاده شد. این توالی‌ها از بانک جهانی ژنی NCBI (www.ncbi.nlm.nih.gov) استخراج شدند. این ۸۲۷ توالی هم با استفاده از نرم‌افزار RNAfold (۲۱) برای ساختار ثانویه، ساختار سنجاق سری و لوپ‌ها مورد بررسی قرار گرفت. در کل مجموعه داده آموزشی استفاده شده در این پروژه شامل ۱۲۰۰ توالی بود. از این نمونه‌ها ۳۷۳ مورد مثبت (miRNA) و ۸۲۷ مورد منفی (non-miRNA) بودند.

استخراج ویژگی‌ها

در این پژوهش ۲۰ ویژگی که باعث تمایز دقیق‌تر و بهتر ژن‌های miRNA با سایر قسمت‌های ژنوم می‌شد، انتخاب شدند (۵۰). سپس برای این ویژگی‌ها در محیط Visual Studio اسکریپت‌های لازم نوشته شد. برای بخش ترمودینامیکی این نرم‌افزار بعد از فراهم کردن ورودی‌های لازم، برنامه RNAFold را فراخوانی و از نتیجه آن استفاده شد.

بررسی ویژگی‌ها و پیاده سازی نرم‌افزار

در این پژوهش از ۲۰ ویژگی (نه ویژگی وابسته به ساختار، هفت ویژگی وابسته به جفت باز و چهار ویژگی توالی نوکلئوتیدی) استفاده شد. لیست ویژگی‌های به‌کار رفته به‌همراه توضیحات آن‌ها در جدول ۱ آمده است. این ویژگی‌ها با نرم‌افزار Feature Extraction، استخراج شد.

جدول ۱- لیست بیست ویژگی مورد استفاده در شناسایی pre- miRNA

Table 1.Detailed list of 20 features used for pre- miRNA identification

	توضیح	ویژگی
۱	حداقل انرژی آزاد با شاخص ۱	$MFEI_1$
۲	حداقل انرژی آزاد با شاخص ۲	$MFEI_2$
۳	حداقل انرژی آزاد با شاخص ۳	$MFEI_3$
۴	حداقل انرژی آزاد با شاخص ۴	$MFEI_4$
۵	حداقل انرژی آزاد نرمال شده	dG
۶	گرایش جفت باز به تشکیل ساختار دوم	dP
۷	حداقل انرژی نرمال جور شدگی	EAFE
۸	فراوانی ساختارهای با حداقل انرژی	Freq
۹	فاصله جفت‌بازهای موثر از یکدیگر	Diff
۱۰	تعداد جفت بازهای نرمال شده	A-U/L⌋
۱۱	تعداد جفت بازهای نرمال شده	G-C/L⌋
۱۲	تعداد جفت بازهای نرمال شده	G-U/L⌋
۱۳	تعداد جفت بازهای نرمال در هر ساقه	ABS
۱۴	میانگین جفت باز در هر ساقه	(A-U)/s٪
۱۵	میانگین جفت باز در هر ساقه	(G-C)/s٪
۱۶	میانگین جفت باز در هر ساقه	(G-U)/s٪
۱۷	درصد بازهای گوانین و سیتوزین	G+C٪
۱۸	فراوانی توالی سیتوزین – گوانین	CG٪
۱۹	فراوانی توالی اوراسیل – اوراسیل	UU٪
۲۰	فراوانی توالی سیتوزین – اوراسیل	CU٪

ویژگی‌های استخراج شده برای داده‌های مثبت و منفی در نرم‌افزار MATLAB با استفاده از یک مدل هوشمند ماشین بردار پشتیبان (SVM) با کرنل RBF و الگوریتم یادگیری SMO با روش اعتبارسنجی متقاطع ده تایی[1] به‌عنوان بهترین دسته‌بندی کننده برای پیش‌بینی ژن‌های microRNA در گوسفند و بز استفاده شد.

روش‌های محاسباتی برای شناسایی ژن‌های microRNA در گوسفند

به‌منظور شناسایی ژن‌های microRNA در گوسفند تمام miRNAهای چهار گونه انسان، موش، بز و گوسفند از miRBase نسخه ۲۱ دانلود شدند (http://www.mirbase.org). سپس توالی‌های ژنومی گوسفند شامل mRNA، cDNA و ESTها که منشاء عمومی ژن‌های microRNA هستند، از بانک جهانی ژن استخراج شدند (www.ncbi.nlm.nih.gov). بعد از حذف توالی‌های تکراری، باقی‌مانده miRNAها روی ژنوم گوسفند از لحاظ هم‌پوشانی مورد بررسی قرار گرفتند. در ادامه توالی‌هایی که برای کد شدن پروتئین مورد استفاده قرار می‌گرفتند، حذف شدند. ساختار ثانویه و پایداری توالی‌های باقی‌مانده با طول ۵۰ تا ۶۰ نوکلئوتید با استفاده از برنامه Mfold (version 3.2) (۶۱) مشخص شد (www.bioinfo.rpi.edu/applications/mfold/rna).

1- 10- Fold Cross Validation

در نهایت توالی‌هایی که ویژگی‌های زیر را به‌دست آوردند به عنوان microRNA بالغ در نظر گرفته شدند: ۱) توانایی تشکیل ساختار سنجاق سری ۲) عدم وجود حباب بزرگ در ساختار ۳) تولید ساختار ثانویه با حداقل انرژی آزاد ۱۸/۸- (۴۷).

روش‌های محاسباتی برای شناسایی ژن‌های microRNA در بز

برای گونه بز نیز در ابتدا توالی‌های miRNA گونه بز استخراج شدند (http://www.mirbase.org). سپس توالی‌های ESTهای بز از بانک جهانی ژن دانلود شد (www.ncbi.nlm.nih.gov/dbEST). در ادامه برای شناسایی miRNAها در بز پنج مرحله زیر طی شد: ۱) بررسی هم‌پوشانی miRNAها با استفاده از نرم‌افزار BioEdit (۲۰) برای حذف توالی‌های زائد و تکراری ۲) استفاده از miRNAهای باقی‌مانده برای BLASTn بر روی ESTهای بز ۳) انتخاب ESTهای با عدم تطبیق کم‌تر از دو جفت بازی برای BLAST بر روی miRNA تأیید شده ۴) انتخاب ESTهای با شباهت بالای ۹۰ درصد برای بررسی و حذف توالی‌های کدکننده پروتئین ۵) استفاده از نرم‌افزار Mfold (۶۱) برای بررسی ساختار ثانویه و بررسی ویژگی‌های توالی‌های انتخاب شده مانند مقدار A-U و حداقل انرژی آزاد (۴۷).

بررسی تنوع نوکلئوتیدی microRNA

برای بررسی تنوع نوکلئوتیدی، mir-134 به‌دلیل مشخص شدن حفظ شدگی بیشتر آن نسبت به سایر miRNAها به‌عنوان شاخص انتخاب شد (۱۰). به‌همین منظور این microRNA در گونه‌های گوسفند، موش و بز دانلود شد (http://www.mirbase.org). سپس برای بررسی تنوع نوکلئوتیدی از نرم‌افزار weblogo (۱۲) و برای بررسی فیلوژنتیک بر اساس رویه neighbor joining از نرم‌افزار ClustalW (۳۲) استفاده شد.

نتایج و بحث

نتایج و بحث بخش اول: نتیجه آنالیز ماشین بردار پشتیبان (SVM)

برای داده‌های آموزشی شرح داده شده در بخش مواد و روش‌ها، مدل هوشمند ماشین بردار پشتیبان (SVM) با کرنل RBF و الگوریتم یادگیری SMO در نرم‌افزار MATLAB، که به‌عنوان بهترین دسته‌بندی کننده هوشمند برای پیش‌بینی ژن‌های microRNA شناخته می‌شود، اجرا شد. در پژوهش حاضر حساسیت و اختصاصیت مدل ماشین بردار پشتیبان به ترتیب ۸۸ و ۸۵ درصد برای شناسایی ژن‌های miRNA در گوسفند و بز به‌دست آمد که نشان می‌دهد که مدل برازش یافته به داده‌های آموزشی از دقت بالایی در پیش‌بینی ژن‌های miRNA برخودار است.

بیشترین مدلی که تاکنون برای شناسایی ژن‌های miRNA مورد استفاده قرار گرفته است، مدل ماشین بردار پشتیبان می‌باشد. جدول ۲ نتایج استفاده از مدل SVM برای شناسایی ژن‌های miRNA انسان را از سوی سایر محققان (۵۰) نشان می‌دهد. با مقایسه سایر نتایج با نتایج پژوهش حاضر می‌توان نتیجه گرفت که پژوهش ما اولین پژوهش صورت گرفته بر روی این گونه از نشخوارکنندگان با استفاده از مدل ماشین بردار پشتیبان است.

جدول ۲- مقایسه نتیجه روش‌های مختلف پیش‌بینی ژن‌های miRNA

Table 2. Comparison of performance of different miRNA prediction tools

منبع	گونه	الگوریتم	حساسیت (٪)	اختصاصیت (٪)
سوور و همکاران (۲۰۰۵)	انسان	SVM	۷۱/۳۲	۹۷/۳۳
زو و همکاران (۲۰۰۵)	انسان	SVM	۹۳/۳۰	۸۸/۰۱
نگ و همکاران (۲۰۰۷)	انسان	SVM	۸۴/۵۵	۹۷/۹۷
هلویک و همکاران (۲۰۰۶)	انسان	SVM	۹۵/۰۱	۹۰/۰۲

با توجه به نتایج گزارش شده از سوی سایر محققان (جدول ۲) می‌توان نتیجه گرفت که از آن‌جایی‌که مقادیر به‌دست آمده برای حساسیت و اختصاصیت مدل‌های یادگیری ماشین، معیاری برای شناسایی دقت مدل در پیش‌بینی miRNAها می‌باشد که کارایی و سودمندی مدل را نشان می‌دهد، مقادیر بالای به‌دست آمده برای حساسیت و اختصاصیت در این پژوهش نشان می‌دهد که مدل یادگیری ماشین (SVM) به‌کار رفته در پژوهش ما، دارای مقادیر حساسیت و اختصاصیتی بالا و تقریباً مشابه با سایر پژوهش‌های انجام گرفته بر روی گونه انسان می‌باشد. در نتیجه دقت بالایی برای پیش‌بینی ژن‌های miRNA دارد. هم‌چنین به‌دلیل متفاوت بودن ماهیت ویژگی‌ها، توالی‌های مورد استفاده برای آموزش سیستم و گونه‌های مورد مطالعه نمی‌توان این روش‌ها را با روش به کار گرفته شده در پژوهش حاضر مقایسه کرد. هر چند رویه به کار گرفته شده در این تحقیق حساسیت و اختصاصیت بالایی (به ترتیب ۸۸ و ۸۵ درصد) را نشان داد.

نتایج و بحث بخش دوم: تجزیه و تحلیل microRNAها با استفاده از بانک‌های اطلاعاتی موجود

به‌منظور دسترسی به miRNAهای بالغ در گونه‌های گوسفند و بز از مجموعه داده‌ها و بانک‌های اطلاعاتی موجود استفاده شد. برای کروموزوم یک گونه بز در نهایت ۱۵ توالی microRNA بالغ به‌دست آمد. پس از حذف miRNAهای تکراری، ۲۳۰ microRNA باقی‌مانده به همراه miRNA تولید شده توسط مدل ماشین بردار پشتیبان به عنوان منبع برای بررسی هم‌پوشانی روی ژنوم بز مورد استفاده قرار گرفتند. از ۵۴۰۰ هدف با عدم جورشدگی حداکثر با چهار جفت باز، تنها ۵۲۰ توالی با طول ۱۶ نوکلئوتید و هم‌پوشانی بالای ۹۰٪ باقی ماندند. سپس توالی‌هایی که به‌عنوان کدکننده پروتئین و یا سایر انواع RNA های غیرکد

شونده شناخته شده مورد استفاده قرار می‌گرفتند، حذف شدند. بعد از بررسی ساختار دوم توالی‌های باقی‌مانده، ۶۲ توالی باقی ماند. این ۶۲ توالی بر روی EST های بز از لحاظ هم‌پوشانی مورد بررسی قرار گرفتند و در نهایت ۱۵ توالی باقی‌مانده به عنوان microRNA بالغ در ژنوم بز تعیین شدند.

محققان دیگری نیز با استفاده از روش‌های متنوع miRNAهای بسیاری را در گونه‌های مختلف شناسایی کردند. به‌طور مثال، کوتینهو و همکاران ۱۲۹ microRNA بالغ را با استفاده از جستجوی شباهت و تولید کتابخانه RNAهای کوچک شناسایی نمودند. هرچند این محققین تمرکز خود را روی بررسی ساختارهای سنجاق سری miRNA انسان گذاشتند (۱۱). هم‌چنین ژو و همکاران ۵۹ microRNA بالغ را در بافت‌های چربی و پستان گاو بر اساس تولید کتابخانه RNAهای کوچک گزارش کردند. لانگ و همکاران miRNAهای چهار بافت گاوی شامل مغز، شش، کبد و قلب را هم‌سانه‌سازی نمودند (۴۰).

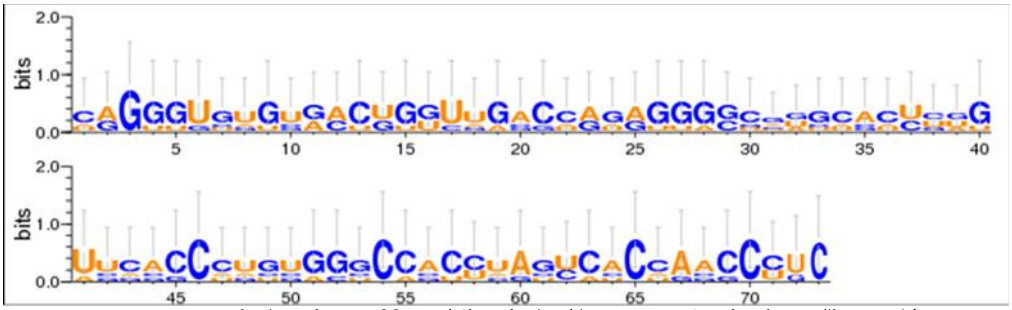
با توجه به نتیجه پژوهش حاضر و هم‌چنین پژوهش‌های متعدد انجام گرفته در زمینه شناسایی miRNAها، می‌توان نتیجه گرفت که با توجه به اهمیت نقش تنظیمی این نوع از RNAهای غیرکدکننده، لزوم انجام تحقیقات بیشتر برای شناسایی روز افزون ژن‌های miRNAهای جدید در نشخوارکنندگان ضروری به‌نظر می‌رسد.

همچنین در این پژوهش، ۲۳ microRNA بالغ را برای کروموزوم یک در گونه گوسفند با استفاده از توالی های EST شناسایی شد. از این تعداد ده مورد با توالی‌های قبلی دارای شباهت بسیار بالایی بود (یک یا دو عدم جور شدگی)، که این خود نشان‌دهنده حفظ شدگی بالای ژن‌های microRNA می‌باشد. تفاوت قابل ذکر در گوسفند با سایر گونه‌های مورد مطالعه حداقل انرژی آزاد آن‌ها می‌باشد. در گوسفند این مقدار برابر با ۲۰- کیلوکالری بر مول بوده در صورتی که در سایر گونه‌ها و به خصوص انسان مقداری بالاتر و حدود ۲۶/۷ کیلوکالری بر مول می‌باشد (۱۶،۱۵). بررسی ساختار دوم microRNA در گوسفند نشان داد حداقل ۱۶ نوکلئوتید در G/U و یا جورشدگی واتسون-کریک درگیر می‌باشند، هر چند این نوکلئوتیدها در تشکیل بازوها و یا حباب‌های بزرگ دخالتی ندارند (۵۱،۲۲،۳،۵،۸). جدول ۳ تعدادی از این توالی‌های پیش‌بینی شده با روش به‌کار رفته در این پژوهش را نشان می‌دهد.

جدول ۳- تعدادی از توالی‌های پیش‌بینی شده توسط مدل ماشین بردارپشتیبان

Table 3. Some of predicted sequences using SVM model

توالی‌ها	حداقل انرژی آزاد
CCAAGACUUGCACUGAUGUUU	-۲۵
AUGUAUATAUGUAUACACAC	-۴۰
UGUGUGUGUGUGCGUGUGCGUG	-۲۸
UCGUUUGCCUUUUUCUGCUU	-۴۰
GUGGUUUUUUUUGUGUUUUU	-۴۷
UCGCCUCCUCCUCUCCA	-۳۷



شکل ۱- مطالعه حفظ شدگی pre-miRNA با استفاده از نرم‌افزار WEblogo در گونه‌های گوسفند، بز و موش

Figure 1. Investigating of pre-miRNA sequence conservation using WEblogo across sheep, goat, and mouse species.

به‌منظور بررسی حفظ شدگی و بررسی فاصله miRNAها در گونه نشخوارکنندگان (گوسفند و بز) از oar-mir-134 به‌عنوان شاخص استفاده شد. ابتدا این microRNA در گونه‌های گوسفند، بز و موش دانلود شد. سپس برای نشان دادن حفظ‌شدگی microRNA از نرم‌افزار Weblogo استفاده شد. همان‌طور که در شکل ۱ مشخص است، این توالی به صورت بالایی حفظ شده است (بزرگ‌تر بودن قطر حروف نشان‌دهنده حفظ شدگی بالاتر می‌باشد).

نتیجه آنالیز فیلوژنتیکی این microRNA نیز بر این نکته تأکید دارد که گونه‌های گوسفند و بز نسبت به سایر گونه‌های مورد مطالعه به یک‌دیگر نزدیک‌تر می‌باشند. این نتیجه با نتایج سایر تحقیق‌های صورت گرفته و منتشر شده دارای هم‌خوانی می‌باشد (۴۸،۳۴،۲۳). تاکنون محققین مختلفی از روش‌های آزمایشگاهی برای شناسایی و بررسی عملکرد microRNA در گونه‌های مختلف استفاده کرده‌اند (۸،۹،۱۸،۴۶،۱۹،۵۲). استفاده از روش‌های بیوانفورماتیکی در پیش‌بینی ژن‌های miRNA نیز بسیار سودمند است. به‌طوری‌که محققین بسیاری با روش‌های مختلف آنالیز داده‌ها، تعداد متنوعی از microRNAها را در گونه‌های مختلف شناسایی کردند. از جمله بنتویچ و همکاران (۵۰)

برنامه PalGrade را با توجه به روش‌های کامپیوتری و ترکیب آن با آنالیز میکرواری هفت برای شناسایی ژن‌های microRNA ایجاد کردند. لیم و همکاران (۵۰) برنامه MiRScan را برای پیش‌بینی ساختار دوم RNA استفاده کردند. استفاده از این روش نشان داد که پیش‌گویی ساختار ثانویه RNA برای شناسایی microRNA در یوکاریوت‌ها مفید است. برزیکوف و همکاران (۵۰) از روش Phylogenetic Shadowing برای شناسایی ژن‌های microRNA استفاده کردند. این روش می‌تواند به صورت دقیق یک نوکلئوتید را از نظر حفاظت شدن مورد بررسی قرار دهد. لی و همکاران (۵۰) با استفاده از برنامه miRseeker در ژنوم دروزوفیلا ۱۱۰ ژن microRNA را شناسایی و ۴۸ مورد آن‌ها را برای کاندید نهایی معرفی نمودند. نام و همکاران (۵۰) برنامه proMIR را با استفاده از آموزش بر پایه برنامه‌نویسی برای شناسایی microRNA تولید نمودند. آمبروس و همکاران (۵۰) براساس توالی cDNA و استفاده از ژنومیک مقایسه‌ای RNA کوچک را شناسایی کردند. آن‌ها از برنامه MFlod برای بررسی ساختار ثانویه و تعیین حداقل انرژی آزاد استفاده کردند. یکی دیگر از روش‌های شناسایی محاسباتی، روش‌هایی مبتنی بر EST است. برای مثال ویلیامز و همکاران (۵۰) از EST غیرکدشونده برای شناسایی RNA کوچک استفاده کردند. لویس و همکاران (۵۰) از Targetscan و کرک و همکاران (۵۰) از Pictar، انرایت و همکاران (۵۰) و جان و همکاران (۲۰۰۷) از miRanda، بریلر و مکدونالد (۵۰) از moving Targets، رمسمیر و همکاران (۵۰) از RNA hybrid ، کاریکدو و همکاران (۵۰) از DIANA-microT (یک روش بیوانفورماتیکی و آزمایشگاهی براساس اسکن یک محل اتصال برای اهداف miRNA) و ساتروم و همکاران (۵۰) از TargetBoost برای شناسایی اهداف microRNA در حیوانات استفاده نمودند (۵۰).

از جمله مطالعات انجام گرفته روی miRNAهای مؤثر روی صفات اقتصادی نشخوارکنندگان، می‌توان به microRNA مخصوص ماهیچه اشاره کرد که ژنی را تنظیم می‌کند که به‌طور مستقیم ویژگی‌ها و خصوصیات اقتصادی را در نشخوارکنندگان تحت تأثیر قرار می‌دهد. هم‌چنین جهش در ژن myostatin در گوسفند Belgian Texel، یک جایگاه هدف برای miR-1 و miR-206 ایجاد می‌کند که سبب مهار رونویسی پروتئین myostatin می‌شود که به افزایش حجم ماهیچه منجر می‌شود. علاوه بر ماهیچه اسکلتی، بافت‌های چربی نیز بر کیفیت و محصول گوشت تأثیرگذار هستند. امروزه پروفایل ترنسکریپتوم microRNA در بافت چربی برای گاو ایجاد شده است. ژو و همکارانش ۱۵۴ توالی microRNA را شناسایی کردند که ۱۳۳ تای آنها با microRNA پستانداران یا کاملاً منطبق است یا در یک یا دو نوکلئوتید متفاوت است. علاوه بر این، با مقایسه توالی‌های به‌دست آمده از غده پستانی گاو، مشخص شد که ۵۴ توالی microRNA ویژه بافت چربی می‌باشد. نقش microRNA در تولید مثل نشخوارکنندگان بسیار مورد توجه است اما هنوز نیاز به ارزیابی مدل‌های نشخوارکنندگان دارد. علاوه بر پروفایل‌هایی که تاکنون بحث شد، پروفایل بافت‌های دیگری که از نظر اقتصادی دارای اهمیت هستند، مانند بافت‌های مربوط به پستان و ایمنی گاو ایجاد شده‌اند. موتیف‌های microRNA در نزدیکی SNPهایی که مربوط به برداشت غذا هستند، یافت شده اند که نشان می‌دهد miRNAها در تنظیم ژن‌هایی که در کارایی غذا نقش دارند، مؤثر می‌باشند. باروزای و همکاران و کوتینهو و همکارانش نیز در پژوهش‌های خود اهداف miRNAها را فاکتورهای رونویسی معرفی کردند(۳،۱۱). هم‌چنین برخی از اهداف miRNAها پروتئین‌هایی است که نقش‌های حیاتی و مهم در رشد و متابولیسم ایفا می‌کنند. miRNAهای جدید در آینده نزدیک می‌توانند برای بهبود تولید گوشت و سایر ویژگی‌های نشخوارکنندگان با توجه به اهداف خود مورد استفاده قرار بگیرند. تحقیقات صورت گرفته برای شناسایی miRNAهای گوسفند و بز با استفاده از روش‌های بیوانفورماتیکی بسیار محدود می‌باشد. برای مثال شنگ و همکاران (۳۹) با ترکیب یک روش محاسباتی بر اساس آنالیز EST و روش آزمایشگاهی براساس کتابخانه cDNA مربوط به RNA های کوتاه، ۳۱ miRNA را در ۲۴ نژاد گوسفند شناسایی کردند که دو مورد از آن‌ها miRNA هایی بود که تا به حال در هیچ گونه‌ای از آن‌ها شناسایی نشده بود. هم‌چنین آن‌ها ۱۲ miRNA متعلق به ماهیچه اسکلتی گوسفند را که کاندیدهای مناسبی برای مطالعه بر روی فرآیندهای تنظیمی رشد ماهیچه وابسته به miRNA بود، کلون کردند که با این روش چهار جفت miRNA-miRNA و یک جفت miRNA-3p/miRNA-5p را از توالی‌های EST گوسفند شناسایی نمودند. آن‌ها با آنالیز بیان این توالی‌ها نشان دادند که برخی از miRNAها فقط در یک بافت خاص بیان می‌شوند و هم‌چنین با استفاده از روش‌های محاسباتی، ۱۲۰ ژن هدف برای ۳۱ miRNA مؤثر در فرآیندهای متابولیکی را که در ناحیه ۳' UTR ژن‌های گوسفند قرار داشت، پیش بینی کردند (۳۹).

لینگ و همکاران (۳۹) کتابخانه cDNA را برای RNAهای کوتاه بزهای بوئر ساختند. آن‌ها miRNAهای ماهیچه بز را با استفاده از تکنولوژی توالی‌یابی Deep-Solexa تولید کردند و با روش‌های بیوانفورماتیکی ویژگی‌های این miRNAها را آنالیز کردند. آن‌ها بر اساس توالی‌یابی Solexa و آنالیزهای بیوانفورماتیکی ۵۶۲ گونه محافظت شده و ۵ miRNA ویژه ژنوم بز را شناسایی کردند. شناسایی و تعیین ویژگی‌های miRNAهای بز اطلاعات مهمی را در شناسایی نقش تنظیمی miRNAها در رشد و نمو ماهیچه‌های بز آشکار می‌سازد (۳۹).

ویی و همکاران (۵۳) از اطلاعات EST بز به منظور پیش‌بینی miRNAهای جدید استفاده کردند و فقط توانستند ۵ miRNA که به‌طور خاص در بز وجود دارند را پیش‌بینی کنند. هم‌چنین نتایج آن‌ها با نتایج فو و همکاران (۵۳) از این نقطه‌نظر که اطلاعات توالی‌های EST برای بز در GenBank بسیار محدود و کمیاب است، مطابقت داشت (۵۳).

همان‌طور که ملاحظه می‌شود، تاکنون مطالعات بسیار اندکی در زمینه شناسایی ژن‌های microRNA در گوسفند و بز با استفاده از روش‌های محاسباتی صورت گرفته است. به‌دلیل عدم هم‌پوشانی مطالعات صورت گرفته و وجود بررسی اهداف مختلف در آن‌ها امکان مقایسه کارهای صورت گرفته با یکدیگر و نتایج این پژوهش وجود نداشت. با مقایسه روش مورد استفاده در پژوهش ما با سایر پژوهش‌های محدود انجام گرفته در زمینه شناسایی ژن‌های miRNA در گوسفند و بز می‌توان نتیجه گرفت که از آنجایی‌که ژنوم کامل بز به‌تازگی منتشر شده است و بسیاری از ژن‌های آن هنوز ناشناخته هستند، هم‌چنین روی ژنوم گوسفند هم هنوز مطالعاتی در دست انجام است، ما با استفاده از روش محاسباتی ارائه شده در این پژوهش توانستیم تعدادی از miRNA های جدید در گوسفند و بز را پیش‌بینی کنیم. با توجه به این‌که تحقیقات در این زمینه به تازگی شروع شده است، تعداد بسیار زیادی از miRNAهای نشخوارکنندگان وجود دارند که به میزان زیادی هم بیان می‌شوند، اما هنوز شناسایی نشده‌اند که با روش‌های شناسایی محاسباتی می‌توانند با دقت بالا پیش‌بینی شوند. با افزایش اطلاعات توالی‌یابی ژنوم گوسفند و بز می‌توان در آینده تعداد بیشتری از miRNAهای جدید را با استفاده از روش‌های بیوانفورماتیکی محاسباتی شناسایی نمود. روش‌های محاسباتی باعث صرفه‌جویی در هزینه و زمان پژوهش‌های آزمایشگاهی می‌شود و در نتیجه سرعت و دقت شناسایی miRNAها افزایش می‌یابد.

با توجه به نقش مهم miRNAها در فرآیندهای بیولوژیکی، آشکارسازی زودهنگام و کم‌هزینه آن‌ها، اهمیت بسیار زیادی دارد. در این مطالعه نقش دسته‌بندی کننده هوشمند ماشین بردار پشتیبان (SVM) در پیش‌بینی ژن‌های miRNA مورد بررسی قرار گرفت. بر روی سه دسته ویژگی ساختاری، موقعیتی و ترمودینامیکی تمرکز و ویژگی‌های برتر انتخاب شد. در این پژوهش، ماشین بردار پشتیبان با کرنل RBF و الگوریتم یادگیری SMO بهترین دسته‌بندی کننده برای ژن‌های miRNA معرفی شد. در ادامه تحقیق به منظور دست‌یابی به miRNAهای بالغ از آنالیزهای مجازی بهره گرفته شد. در نهایت به ترتیب در گوسفند و بز ۲۳ و ۱۵ microRNA بالغ شناسایی شد که می‌توان در پژوهش‌های آزمایشگاهی از آن‌ها برای مطالعات بیشتر بهره یافت.

تشکر و قدردانی

این پژوهش با همکاری آزمایشگاه بیوانفورماتیک مرکز تحقیقات بیوفیزیک- بیوشیمی دانشگاه تهران انجام گرفته است. بدین وسیله از آقای دکتر حامد احمدی در بخش محاسباتی این پروژه به خاطر کمک‌ها و راهنمایی‌های بی‌دریغشان سپاس‌گزاری می‌نماییم.

منابع

1. Ambros, V. 2004. The functions of animal micrornas. Nature, 431: 350-355.
2. Ambros, V., B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, G. Dreyfuss, S.R. Eddy, S. Griffiths-Jones and M. Marshall. 2003. A uniform system for microrna annotation. Rna, 9: 277-279.
3. Barozai, M.Y.K., I.A. Baloch and M. Din. 2011a. Computational identification of micrornas and their targets in two species of evergreen spruce tree (Picea). Waset, 75: 413-418.
4. Barozai, M.Y.K., I.A. Baloch and M. Din. 2012. Identification of micrornas and their targets in helianthus. Molecular Biology Reports, 39: 2523-2532.
5. Barozai, M.Y.K., M. Din and I.A. Baloch. 2011b. Identification of micrornas in ecological model plant mimulus. Journal of Biophysical Chemistry, 2: 322-331.
6. Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297.
7. Berezikov, E., G. Van Tetering, M. Verheul, L. Van Laake, J. Vos, R. Verloop, M. Van de Wetering, V. Guryev, S. Takada, A.J. Van Zonneveld, H. Mano, R. Plasterk and E. Cuppen. 2006. Many novel mammalian microrna candidates identiwed by extensive clon-ing and rake analysis. Genome Research, 16: 1289-1298.
8. Burnside, J., M. Ouyang, A. Anderson, E. Bernberg, C. Lu, B.C. Meyers, P.J. Green, M. Markis, G. Isaacs and E. Huang. 2008. Deep sequencing of chicken micrornas. BMC Genomics, 9: 185.
9. Chen, C.Z., L. Li, H.F. Lodish and D.P. Bartel. 2004. MicroRNAs modulate hematopoietic lineage differentiation. Science, 303: 83-86.
10. Clop, A., F. Marcq, H. Takeda, D. Pirottin, X. Tordoir, B. Bibé, J. Bouix, F. Caiment, J.M. Elsen and F. Eychenne. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin Gene Affects Muscularity in Sheep. Nature Genetics, 38: 813-818.
11. Coutinho, L.L., L.K. Matukumalli, T.S. Sonstegard, C.P. Van Tassell, L.C. Gasbarre, A.V. Capuco and T.P.L Smith. 2007. Discovery and profiling of bovine micrornas from immune-related and embryonic tissues. Physiological Genomics, 29: 35-43.
12. Crooks, G.E., G. Hon, J.M. Chandonia and S.E. Brenner. 2004. WebLogo: A sequence logo generator. Genome Research, 14: 1188-1190.
13. Eddy, S.R. 2004. How do rna folding algorithms work? nature biotechnology, 22: 1457-1458.
14. Enright, A.J., B. John, U. Gaul, T. Tuschl, C. Sander and D.S Marks. 2004. MicroRNA targets in drosophila. Genome Biology, 5: R1.
15. Filipowicz, W., S.N. Bhattacharyya and N. Sonenberg. 2008. Mechanisms of post-transcriptional regulation by micrornas: are the answers in sight? Nature Reviews Genetics, 9: 102-114.

16. Friedman, R.C., K.K.H. Farh, C.B. Burge and D.P. Bartel. 2009. Most mammalian mrnas are conserved targets of microrna. Genome Research, 19: 92-105.
17. Gaidatzis, D., E. Van Nimwegen, J. Hausser and M. Zavolan. 2007. Inference of mirna targets using evolutionary conservation and pathway analysis. BMC Bioinformatics, 8: 69.
18. Glazov, E.A., P.A. Cottee, W.C. Barris, R.J. Moore, B.P. Dalrymple and M.L. Tizard. 2008. A microrna catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Research, 18: 957-964.
19. Glazov, E.A., K. Kongsuwan, W. Assavalapsakul, P.F. Horwood, N. Mitter and T.J Mahony. 2009. Repertoire of bovine mirna and mirna-like small regulatory rnas expressed upon viral infection. PLoS One, 4: e6349.
20. Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: Nucleic Acids Symposium Series, 41: 95-98.
21. Hofacker, I.L. 2003. Vienna rna secondary structure server. Nucleic Acids Research, 31: 3429-3431.
22. Hossain, M.M., N. Ghanem, M. Hoelker, F. Rings, C. Phatsara, E. Tholen, K. Schellander and D. Tesfaye. 2009. Identification and characterization of mirnas expressed in the bovine ovary. BMC Genomics, 10: 443.
23. Huang, J., Z. Ju, Q. Li, Q. Hou, C. Wang, J. Li, R. Li, L. Wang, T. Sun and S. Hang. 2011. Solexa sequencing of novel and differentially expressed micrornas in testicular and ovarian tissues in holstein cattle. International Journal of Biological Sciences, 7: 1016-1020.
24. Huang, T.H., B. Fan, M.F. Rothschild, Z.L. Hu, K. Li and S.H. Zhao. 2007. MiRFinder: an improved approach and software implementation for genome-wide fast microrna precursor scans. BMC Bioinformatics, 8: 341-350.
25. Huang, Y., Q. Zou, S.M. Tang, L.G. Wang and X.J. Shen. 2010. Computational identification and characteristics of novel MicroRNAs from the Silkworm (*Bombyx mori* L.). Molecular Biology Reports, 37: 3171-3176.
26. Kidner, C.A. and R.A. Martienssen. 2005. The developmental role of microrna in plants. Current Opinion in Plant Biology 8: 38-44.
27. Kloosterman, W.P. and R.H.A. Plasterk. 2006. The diverse functions of micrornas in animal development and disease. Developmental Cell, 11: 441-450.
28. Krek, A., D. Grün, M.N. Poy, R. Wolf, L. Rosenberg, E.J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus and M. Stoffel. 2005. Combinatorial microrna target predictions. Nature Genetics, 37: 495-500.
29. Krützfeldt, J., N. Rajewsky, R. Braich, K.G. Rajeev, T. Tuschl, M. Manoharan and M. Stoffel. 2005. Silencing of MicroRNAs in vivo with 'antagomirs'. Nature, 438: 685-689.
30. Lagos-Quintana, M., R. Rauhut, W. Lendeckel and T. Tuschl. 2001. Identification of novel genes coding for small expressed RNAs. Science Signalling, 294: 853-858.
31. Lagos-Quintana, M., R. Rauhut, A. Yalcin, J. Meyer, W. Lendeckel and T. Tuschl. 2002. Identification of tissue-specific MicroRNAs from mouse. Current Biology, 12: 735-739.
32. Larkin, M., G. Blackshields, N. Brown, R. Chenna, P. McGettigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm and R. Lopez. 2007. Clustal W and Clustal X Version 2.0. Bioinformatics, 23: 2947-2948.
33. Lau, N.C., L.P. Lim, E.G. Weinstein and D.P. Bartel. 2001. An abundant class of tiny rnas with probable regulatory roles in caenorhabditis elegans. Science Signalling, 294: 858-867.
34. Lee, R., R. Feinbaum, and V. Ambros. 2004a. A short history of a short RNA. Cell, 116: 89-100.
35. Lee, R. C., R.L. Feinbaum and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to Lin-14. Cell, 75: 843-854.
36. Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark and S. Kim. 2003. The nuclear RNase III drosha initiates MicroRNA Processing. Nature, 425: 415-419.
37. Lee, Y., M. Kim, J. Han, K.H. Yeom, S. Lee, S.H. Baek and V. N Kim. 2004b. MicroRNA Genes Are Transcribed by RNA Polymerase II. EMBO J, 23: 4051-4060.
38. Lewis, B. 2003. Prediction of Mammalian MicroRNA targets., *et al*. 115, 2003, Cell, 115: 787-798.
39. Ling, Y.H., J.P. Ding, X.D. Zhang, L.J. Wang, Y.H. Zhang, Y.S Li, Z.J. Zhang and X.R. Zhang. 2013. Characterization of MicroRNAs from goat (Capra Hircus) by solexa deep-sequencing technology. Genet Molecular Research, 12: 1951-1961.
40. Long, J.E. and H.X. Chen. 2009. Identification and Characteristics of cattle MicroRNAs by homology searching and small RNA cloning. Biochemical Genetics, 47: 329-343.
41. Mathews, D.H., J. Sabina, M. Zuker and D.H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. Journal of Molecular Biology, 288: 911-940.
42. Nilsen, T.W. 2007. Mechanisms of MicroRNA-Mediated Gene Regulation in Animal Cells. Trends Genet, 23: 243-249.
43. Pillai, R.S., S.N. Bhattacharyya and W. Filipowicz. 2007. Repression of protein synthesis by mirnas: how many mechanisms? Trends in Cell Biology, 17: 118-126.
44. Plasterk, R.H.A. 2006. Micro RNAs in Animal Development. Cell, 124: 877-881.

45. Place R.F., L.C. Li, D. Pookot, E.J. Noonan and R. Dahiya. 2008. MicroRNA-373 induces expression of genes with complementary promoter sequences. Proceedings of the national academy of Sciences of the United States of America, 105: 1608-1613.
46. Ramachandra, R.K., M. Salem, S. Gahr, C.E. Rexroad and J. Yao. 2008. Cloning and characterization of MicroRNAs from rainbow trout (oncorhynchus mykiss): their expression during early embryonic development. BMC Developmental Biology, 8: 41-50.
47. Rehmsmeier, M., P. Steffen, M. Hochsmann and R. Giegerich. 2004. Fast and effective prediction of MicroRNA/Target Duplexes. RNA, 10: 1507-1517.
48. Sheng, X., X. Song, Y. Yu, L. Niu, S. Li, H. Li, C. Wei, T. Liu, L. Zhang and L. Du. 2011. Characterization of MicroRNAs from Sheep (Ovis Aries) using computational and experimental analyses. Molecular Biology Reports, 38: 3161-3171.
49. Singh, J. and J. Nagaraju. 2008. In silico prediction and characterization of MicroRNAs from Red flour beetle (Tribolium Castaneum). Insect Molecular Biology, 17: 427-436.
50. Sinha, S., T. Vasulu and R.K. De. 2009. Performance and evaluation of MicroRNA gene identification tools. Journal of Proteomics & Bioinformatics, 2: 336-343.
51. Strozzi, F., R. Mazza, R. Malinverni and J. Williams. 2009. Annotation of 390 bovine MiRNA genes by sequence similarity with other species. Animal Genetics, 40: 125.
52. Tesfaye, D., D. Worku, F. Rings, C. Phatsara, E. Tholen, K. Schellander and M. Hoelker. 2009. Identification and expression profiling of microrna during bovine oocyte maturation using heterologous approach. Molecular Reproduction and Development, 76: 665-677.
53. Wei, Y., S. Chen, P. Yang, Z. Ma and L. Kang. 2009 Characterization and comparative profiling of the small rna transcriptomes in two phases of locust. Genome Biology, 10: R6.
54. Wightman, B., I. Ha and G. Ruvkun. 1993. Posttranscriptional Regulation of the Heterochronic Gene Lin-14 by Lin-4 Mediates Temporal Pattern Formation in C. Elegans. Cell, 75: 855-862.
55. Winter, J., S. Jung, S. Keller, R.I. Gregory and S. Diederichs. 2009. Many roads to maturity: microrna biogenesis pathways and their regulation. Nature cell Biology, 11: 228-234.
56. Yekta, S., I. Shih and D.P. Bartel. 2004. MicroRNA-Directed cleavage of HOXB8 mRNA. Science Signalling, 304: 594-610.
57. Yoon, S. and G.D. Micheli. 2006. Computational Identification of MicroRNAs and Their Targets. Birth Defects Research Part C: Embryo Today: Reviews, 78: 118-128.
58. Yousef, M., L. Showe and M. Showe. 2009. A study of MicroRNAs in silico and in vivo: bioinformatics approaches to MicroRNA discovery and target identification. FEBS Journal, 276: 2150-2156.
59. Zhang, B., X. Pan, Q. Wang, G.P. Cobb and T.A. Anderson. 2006. Computational identification of MicroRNAs and Their Targets. Computational Biology and Chemistry, 30: 395-407.
60. Zhang, Y., J. Wang, S. Huang, X. Zhu, J. Liu. N. Yang, D. Song, R. Wu and G. Skogerbo. 2009. Systematic identification and characterization of chicken (Gallus Gallus) ncRNAs. Nucleic Acids Research, 37: 6562-6574.
61. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31: 3406-3415.

Detecting of Functional Short Non-Coding RNAs using Bioinformatics Methods in Sheep and Goat

Atefeh Seyeddokht[1], Hamed Ahmadi[2], Ali Asghar Aslaminejad[3], Ali Masoudi-Nejad[4], Mohammadreza Nassiri[5] and Balal Sadeghi[6]

1 and 5- PhD and Professor, Ferdowsi University of Mashhad
2 and 4- PhD and Professor, University of Tehran
3- Associate Professor, Ferdowsi University of Mashhad, (Corresponding author: agr764@yahoo.co.uk)
6- Assistant Professor, Shahid Bahonar University of Kerman
Received: December 26, 2014 Accepted: June 20, 2015

Abstract

MicroRNAs (miRNAs) are small non-coding RNAs that have functional roles in post-transcriptional modification. They regulate gene expression by an RNA interfering pathway through cleavage or inhibition of the translation of target mRNA. Numerous miRNAs have been described for their important functions in developmental processes in numerous animals, but there is limited information about sheep and goat miRNAs. Sheep and goat are ideal model organisms for biological and comparative genomics studies in ruminants. Identification of miRNAs is crucial to understanding their biological mechanism. Computational identification approaches can supplement experimental approaches to quickly identify ncRNAs in novel genomes, chiefly miRNAs that are transcribed under particular conditions in specific cell types. Currently, machine learning approaches have been employed to predict novel miRNAs. In this study, we present a new SVM-based classifier. It demonstrated high accuracy, balanced sensitivity and specificity for the miRNA datasets, thus representing an ideal tool for miRNA identification from transcriptome sequencing data. In this research, we generated an optimized feature subset including 20 features using a support vector machine, and we developed a c # program to compute the features in the training sequences. In this study, an intelligent SVM model with RBF kernel and the SMO learning algorithm was the best classifier for predicting microRNA genes in sheep and goat. Sensitivity and specificity of this model were 88% and 85% respectively. Then, expressed sequence tag (EST) analysis was performed for finding sheep and goat mature miRNAs. Chromosome 1 was scanned for finding miRNA potential region. In sheep 23 miRNA genes and, in goat 15 miRNAs had been discovered by homology searching. Our finding demonstrate that the Sheep and goat miRNA sequences can be supplied useful information for investigating biological roles of miRNAs in ruminants.

Keywords: Pre-miRNA Genes, Ruminants, Sheep and Goat, Support Vector Machine