

مقاله پژوهشی

بررسی و تعیین اثرات متقابل مهم پلی مورفیسم های تک نوکلئوتیدی ژن های سایتوکاین های التهابی و ضد التهابی در ابتلا به بیماری لوپوس آریتما توز سیستمیک با استفاده از رویکرد شناسه گزینی منطقی

سعید مهرآور^۱، کرامت نوری جلیانی^۲، کمال اعظم^{۳*}، عباس رحیمی فروشانی^۴، مهدی محمودی^۴، علی اکبر امیرزرگر^۵، ربابه قدسی قاسم آبادی^۱

^۱ کارشناس ارشد آمار زیستی، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت و انستیتو تحقیقات بهداشتی، دانشگاه علوم پزشکی تهران، تهران، ایران
^۲ دانشیار آمار زیستی، دانشکده بهداشت و انستیتو تحقیقات بهداشتی، دانشگاه علوم پزشکی تهران، تهران، ایران
^۳ استادیار آمار زیستی، دانشکده بهداشت و انستیتو تحقیقات بهداشتی، دانشگاه علوم پزشکی تهران، تهران، ایران
^۴ استادیار ایمونولوژی مرکز تحقیقات روماتولوژی، دانشگاه علوم پزشکی تهران، تهران، ایران
^۵ استاد ایمونولوژی مرکز تحقیقات ایمونولوژی مولکولی، دانشگاه علوم پزشکی تهران، تهران، ایران
^{*} نویسنده مسئول: تهران، دانشگاه علوم پزشکی تهران، دانشکده بهداشت، گروه اپیدمیولوژی و آمار زیستی
 پست الکترونیک: kazam@sina.tums.ac.ir

وصول: ۹۲/۳/۱۱ اصلاح: ۹۲/۵/۲۹ پذیرش: ۹۲/۶/۱۶

چکیده

زمینه و هدف: با وجود اینکه تاکنون مطالعات گسترده‌ای بر روی تاثیر پلی مورفیسم های تک نوکلئوتیدی (SNPها) بر روی بیماری های مختلف صورت گرفته است ولی لزوم بررسی اثرات متقابل این SNPها با هم در بروز بیماریها خصوصاً بیماریهای مرتبط با ژنتیک از یکسو و تعداد زیاد این متغیرها و ضعف مدل های کلاسیک آماری در لحاظ کردن آنها از سویی دیگر لزوم بکارگیری روشهای نوین آماری در تعیین این اثرات متقابل را نشان می دهد. لذا در مطالعه حاضر با استفاده از روش نوین شناسه گزینی منطقی به بررسی و تعیین ارتباط مهمترین اثرات متقابل میان این SNPها با بیماری لوپوس آریتما توز سیستمیک پرداخته شد.

مواد و روش کار: در این مطالعه اطلاعات مربوط به ۱۱ پلی مورفیسم تک نوکلئوتیدی واقع در ژن های *TNF-α*، *IL-10*، *IL-1 cluster* و *IL-4Ra* برای ۳۷۶ نفر (۱۸۸ بیمار و ۱۸۸ سالم) مورد بررسی قرار گرفت، برای یافتن مهمترین SNPها و اثرات متقابل آنها در ابتلا به بیماری لوپوس روش آماری نوین شناسه گزینی منطقی بکار گرفته شد و متعاقباً نتایج حاصل در مدل رگرسیون لجستیک بکار گرفته شد. **یافته ها:** نتایج حاصل از رویکرد شناسه گزینی منطقی منجر به شناسایی ۲ اثر متقابل دو طرفه گردید که تنها اثر متقابل مربوط به پلی مورفیسم های *TNF-α* در مدل رگرسیون لجستیک نهایی معنی دار نشان داده شد که به نظر می رسد در ابتلا به بیماری لوپوس موثر باشد.

نتیجه گیری: با توجه به نتایج بدست آمده به نظر می رسد که استفاده از نتایج روشهای نوین آماری مانند شناسه گزینی منطقی می تواند در یافتن مهمترین اثرات متقابل مهم در بین SNPها در مطالعات با تعداد بالای متغیر بسیار سودمند باشد.

واژه های کلیدی: شناسه گزینی منطقی، اثر متقابل، پلی مورفیسم تک نوکلئوتیدی، لوپوس آریتما توز سیستمیک

مقدمه

مربوط به SNP^۱ (پلی مورفیسم تک نوکلئوتیدی) می باشد

[۱]. یک SNP هنگامی رخ می دهد که یک نوکلئوتید در

اگرچه بیش از ۹۹٪ ژنوم انسانها مشترک می باشد ولی با این حال به طور متوسط در حدود ۳ میلیون تفاوت در نوکلئوتید دو شخص وجود دارد. شایعترین این تفاوت

SNPها و نسبت دادن یک مقدار عددی برای نشان دادن میزان اهمیت اثرات متقابل تعیین شده می‌باشد. رگرسیون منطقی مورد استفاده در رویکرد شناسه‌گزینی منطقی یک رگرسیون تعمیم یافته است که در آن متغیرهای پیشگو به صورت دوحالتی می‌باشند و هدف در آن یافتن بهترین ترکیبات منطقی از متغیرهای پیشگو است که بهترین برازش را برای متغیر پاسخ داشته باشند [۹، ۱۰]. در مطالعه حاضر به بررسی و تعیین مهمترین اثرات متقابل بین پلی‌مورفیسم های تک نوکلئوتیدی ژن های IL-4R¹, IL-10, 1 cluster, TNF- SLE پرداخته شد.

روش کار

در این مقاله از داده‌های یک مطالعه مورد-شاهدی استفاده گردید. تعداد ۱۹۹ بیمار مبتلا به لوپوس اریتماتوز سیستمیک که در طی یکسال از اردیبهشت ۱۳۹۰ تا اردیبهشت ۱۳۹۱ با مراجعه به درمانگاه روماتولوژی بیمارستان شریعتی، بیماری آن‌ها بر اساس معیار بازبینی شده سال ۱۹۹۷ انجمن روماتولوژی آمریکا (ACR11) تشخیص داده شده بودند جمع آوری گردیدند. همچنین گروه کنترل شامل ۲۲۱ فرد سالم از میان کارمندان دانشگاه علوم پزشکی تهران که خود و خویشاوندان درجه اول آن‌ها دارای بیماری روماتیسمی و خود ایمن نبودند انتخاب شدند. گروه بیمار و کنترل از نظر سن، جنس و قومیت با همدیگر همسان‌سازی شدند. از کلیه بیماران و افراد گروه کنترل رضایت‌نامه آگاهانه مبنی بر شرکت در این طرح تحقیقاتی گرفته شد و این مطالعه به تایید کمیته اخلاق دانشگاه علوم پزشکی تهران رسید. پس از حذف مشاهدات با مقادیر گمشده در مجموع ۳۷۶ نفر (۱۸۸ بیمار و ۱۸۸ سالم) وارد تحلیل شدند، جدول ۱ نام و فراوانی هر SNP برای هر دو گروه بیمار و سالم را نشان می‌دهد. جهت تعیین مهمترین ترکیبات منطقی با استفاده از رویکرد شناسه‌گزینی منطقی لازم بود که هر SNP به دو متغیر دو حالتی با مقادیر ۰ و ۱ تبدیل شود. براین اساس کدگذاری هر پلی‌مورفیسم تک نوکلئوتیدی S_i به صورت دو متغیر دو حالتی با مقادیر ۰ و ۱ جهت

یک محل جفت باز منفرد^۱ تغییر کند. همین تفاوت اندک در ژنوم انسان است که باعث بروز تفاوت بین انسان‌ها می‌شود و نیز خطر ابتلا به بیماری، خصوصاً بیماریهای مرتبط با ژنتیک را در افراد مختلف تغییر می‌دهد [۲]. بیماری لوپوس اریتماتوز سیستمیک^۲ (SLE) یک بیماری خود ایمنی و مزمن است که می‌تواند بر هر قسمتی از بدن تاثیر بگذارد. علیرغم اینکه علت بیماری مذکور ناشناخته است، تاثیر عاملهای ژنتیکی و محیطی در بروز بیماری به اثبات رسیده است [۳] طبق مطالعات صورت گرفته نقش سایتوکین‌هایی مانند TNF-^۳, IL-1 cluster^۴, IL-10^۵, IL-6^۶, IFN-^۷ در بروز و درمان این بیماری مشخص شده است [۴-۶]. اما تاکنون بیشتر، اثرات اصلی این SNPها مورد توجه بوده است و این در حالی است که گمان نمی‌رود یک SNP به تنهایی نقش مهمی در پیشرفت بیماریهای پیچیده داشته باشد و بجای آن تعامل این SNPها با هم می‌تواند در بروز بیماریها موثر باشد. بر این اساس لازم است که در مدل‌های آماری نقش اثرات متقابل نیز لحاظ شود، اما مدل‌های آماری کلاسیک نظیر رگرسیون لجستیک تنها می‌توانند اثرات اصلی این متغیرها را لحاظ کنند و اثرات متقابل باید از قبل مشخص شده باشند و این با توجه به بعد بالای متغیرهای مورد بررسی امکان لحاظ کردن و بررسی همه این اثرات ممکن نیست، برای مثال اگر ۳۰ متغیر داشته باشیم برای آنالیز اثرات متقابل تا حداکثر ۴ طرفه نیاز به وارد کردن بیش از ۳۰۰۰۰ متغیر داریم. براین اساس بکارگیری روشهای نوین آماری برای یافتن مهمترین اثرات متقابل ضروری به نظر می‌رسد [۷، ۸]. یکی از این روشهای نوین رویکرد شناسه‌گزینی منطقی^۸ می‌باشد که در آن الگوریتم رگرسیون منطقی^۹ بر چندین زیرمجموعه از داده‌ها برازش می‌یابد، یابد، و هدف در آن تعیین مهمترین اثرات متقابل بین

- 1- single base-pair position
- 2 - Systemic Lupus Erythematosus
- 3 - Tumor Necrosis Factor-
- 4 - Interleukin_1 cluster
- 5 - Interleukin-10
- 6 - Interleukin-6
- 7 - Interferon-
- 8 - logic Feature Selection
- 9 - Logic Regression

می شود و هر ترکیب منطقی که مقدار بالایی را برای هر دو شاخص VIM_{single} و $VIM_{multiple}$ دارا باشد به عنوان یک اثر متقابل تاثیرگذار بر روی متغیر پاسخ (بیماری لوپوس) تعیین می گردد [۷]. در مرحله بعدی به منظور غربالگری متغیرهای دو حالتی در رگرسیون لجستیک از روش پیشنهادی هازمر و لمشو^{۱۱} استفاده می کنیم. در این روش برای آزمون های تک متغیره از سطح معنی داری ۰/۲ و در سایر مراحل سطح معنی داری ۰/۰۵ پیشنهاد شده است [۱۱]. پس از انتخاب مدل با اثرات اصلی، در گام بعد اثرات متقابلی که با روش شناسه گزینی منطقی بر روی بیماری لوپوس موثر تشخیص داده شدند را وارد مدل نموده و مورد بررسی قرار می دهیم. در تمام مراحل بالا برای مقایسه مدلها و انتخاب مدل مناسبتر از معیار AIC^{12} استفاده می کنیم. توضیح اینکه این معیار برای سنجش نیکویی برازش و مقایسه بین چند مدل بکار می رود و نشان می دهد که استفاده از یک مدل آماری به چه میزان باعث از دست رفتن اطلاعات می شود. لازم به ذکر است که مدل با مقدار AIC کمتر مطلوبتر است. برای آنالیز رگرسیون لجستیک از نرم افزار SPSS 20 و برای تعیین مهمترین ترکیبات منطقی با استفاده از روش شناسه گزینی منطقی از نرم افزار R 2.15.2 و بسته نرم افزاری logicFS استفاده گردید.

یافته ها

با توجه به نتایج بدست آمده از روش شناسه گزینی منطقی که در نمودار ۱ مشاهده می شود، ۷ ترکیب منطقی مهم و تاثیرگذار بر بیماری لوپوس بر اساس دو شاخص VIM_{single} و $VIM_{multiple}$ مرتب شده اند. دو اثر متقابل $!rs361525_1$ & $!rs1800629_1$ و $!rs361525_1$ & $!rs1800896_1$ دارای مقادیر بالای VIM_{single} و $VIM_{multiple}$ می باشند. علامت "!" بیانگر مکمل متغیر و "&" مترادف عملگر " (AND) می باشد.

انجام تحلیل شناسه گزینی منطقی بر مبنای زیر تعریف گردید:

S_{i1} : حداقل یکی از بازهای بیان کننده S_i از نوع کمترین فراوانی^۱ در جمعیت باشد.

S_{i2} : هر دو بازهای بیان کننده S_i از نوع کمترین فراوانی در جمعیت باشند.

S_{i1} و S_{i2} به ترتیب نشان دهنده اثر غالب و اثر مغلوب برای S_i SNP می باشند. به بیان دیگر S_{i1} بیان کننده این است که S_i SNP از نوع ژنوتیپ هموزیگوت شایع^۲ نباشد نباشد (هر دو باز بیان کننده S_i از نوع بیشترین فراوانی در جمعیت نباشد) و S_{i2} بیان کننده این است که S_i از نوع ژنوتیپ هموزیگوت واریانت^۳ باشد. (هر دو باز بیان کننده S_i از نوع کمترین فراوانی در جمعیت باشد). با تبدیل هر یک از ۱۱ پلی مورفیسیم تک نوکلئوتیدی به دو متغیر دو حالتی، در مجموع ۲۲ متغیر بدست آمد. از این میان متغیرهایی که مقدار فراوانی مشاهده شده (تعداد ۱ها) در آنها کمتر از ۵ بود از آنالیز کنار گذاشته شدند در نتیجه متغیرهای $rs1800629_2$, $rs361525_2$ از مطالعه حذف شده و تحلیل های آماری با ۲۰ متغیر باقیمانده صورت گرفت. برای تعیین مهمترین اثرات متقابل در روش شناسه گزینی منطقی از رگرسیون منطقی استفاده می شود به این صورت که B نمونه بوت استرپ به حجم n از مشاهدات انتخاب شده و در هر بار یک مدل رگرسیون منطقی به هر یک از این B نمونه برازش می یابد. سپس هر عبارت منطقی تعیین شده به شکل نرمال مجزا^۴ تبدیل میشود، یعنی به صورت ترکیبات "یا"^۵ از ترکیبات "و"^۶. در ادامه ادامه با استفاده از مشاهدات خارج کیسه^۷ (OOB) یعنی داده هایی که در نمونه های بوت استرپ حضور ندارند مقادیر VIM^A برای ترکیباتی که در هر یک از دو رویکرد تک درختی^۹ و چند درختی^{۱۰} بدست می آیند محاسبه

1 - Minor Allele Frequency

2 - Homozygous reference genotype

3 - Homozygous variant genotype

4 - Disjunctive normal form

5- OR

6 - AND

7 - Out-of-bag

8 - Variable Importance Measure

9 - Single tree approach

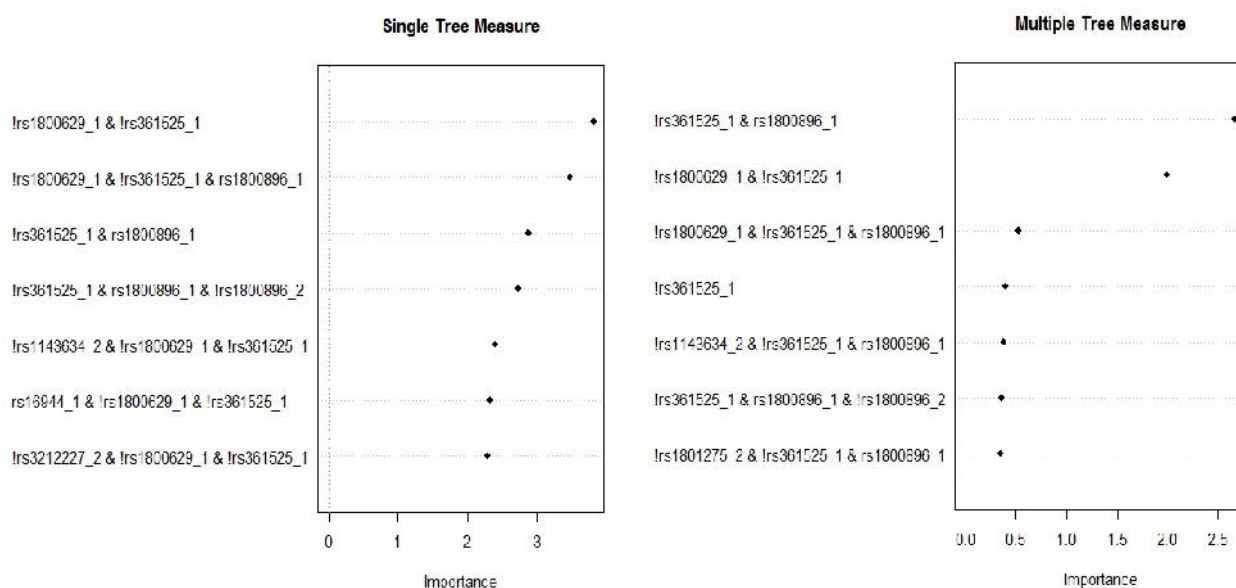
10 - Multiple tree approach

11 - Hosmer and Lemeshow

12 - Akaike information criterion

جدول ۱: فراوانی SNPها در دو گروه بیمار و سالم

سایتوکاین	SNP	ژنوتیپ	تعداد سالم (درصد)	تعداد بیمار (درصد)
IL1A	rs1800587	CC	۸۳ (۴۴/۱)	۸۲ (۴۳/۶)
		CT	۸۵ (۴۵/۲)	۹۱ (۴۸/۴)
		TT	۲۰ (۱۰/۶)	۱۵ (۸/۰)
IL1B	rs16944	CC	۴۳ (۲۲/۹)	۴۶ (۲۴/۵)
		CT	۱۲۱ (۶۴/۴)	۱۱۲ (۵۹/۶)
		TT	۲۴ (۱۲/۸)	۳۰ (۱۶/۰)
IL1R1	rs1143634	CC	۸۵ (۴۵/۲)	۱۰۲ (۵۴/۳)
		CT	۸۹ (۴۷/۳)	۷۶ (۴۰/۴)
		TT	۱۴ (۷/۴)	۱۰ (۵/۳)
IL1RN	rs2234650	CC	۷۴ (۳۹/۴)	۸۶ (۴۵/۷)
		CT	۹۱ (۴۸/۴)	۷۹ (۴۲/۰)
		TT	۲۳ (۱۲/۲)	۲۳ (۱۲/۲)
IL4R	rs315952	TT	۱۰۰ (۵۳/۲)	۱۲۰ (۶۳/۸)
		CT	۸۵ (۴۵/۲)	۶۲ (۳۳/۰)
		CC	۳ (۱/۶)	۶ (۳/۲)
TNF-	rs1801275	AA	۱۴۳ (۷۶/۱)	۱۲۷ (۶۷/۶)
		AG	۳۹ (۲۰/۷)	۵۷ (۳۰/۳)
		GG	۶ (۳/۲)	۴ (۲/۱)
IL10	rs1800629	GG	۱۲۸ (۶۸/۱)	۱۵۴ (۸۱/۹)
		AG	۶۰ (۳۱/۹)	۳۲ (۱۷/۰)
		AA	۰ (۰)	۲ (۱/۱)
IL10	rs361525	GG	۹۴ (۵۰/۰)	۱۷۲ (۹۱/۵)
		AG	۹۳ (۴۹/۵)	۱۶ (۸/۵)
		AA	۱ (۰/۵)	۰ (۰)
IL10	rs1800896	AA	۶۵ (۳۴/۶)	۱۶ (۸/۵)
		AG	۱۱۱ (۵۹/۰)	۱۶۴ (۸۷/۲)
		GG	۱۲ (۶/۴)	۸ (۴/۳)
IL10	rs1800871	CC	۱۰۱ (۵۳/۷)	۷۸ (۴۱/۵)
		TC	۷۵ (۳۹/۹)	۱۰۱ (۵۳/۷)
		TT	۱۲ (۶/۴)	۹ (۴/۸)
IL10	rs1800872	CC	۱۰۱ (۵۳/۷)	۷۰ (۳۷/۲)
		AC	۷۵ (۳۹/۹)	۱۱۰ (۵۸/۵)
		AA	۱۲ (۶/۴)	۸ (۴/۳)



نمودار ۱: ترکیبات منطقی مهم و تاثیر گذار بر بیماری لوپوس آریتما توز سیستمیک که با استفاده از روش شناسه گزینی منطقی و بر اساس دو شاخص VIM_{single} (شکل سمت چپ) و $VIM_{multiple}$ (شکل سمت راست) تعیین شدند.

جدول ۲: ضرایب و نسبت شانس برای مدل رگرسیون نهایی بدون اثرات متقابل

p-value	فاصله اطمینان ۹۵٪ برای نسبت شانس		نسبت شانس	خطای معیار	ضریب رگرسیون	متغیر
	حد بالا	حد پایین				
۰/۰۰۹	۴/۰۵۳	۱/۲۱۵	۲/۲۲۰	۰/۳۰۷	۰/۷۹۷	rs1801275_1
<۰/۰۰۱	۷/۲۹۲	۲/۲۲۱	۴/۰۲۴	۰/۳۰۳	۱/۳۹۲	!
<۰/۰۰۱	۳۵/۲۳۸	۹/۰۱۷	۱۷/۸۲۵	۰/۳۴۸	۲/۸۸۱	! rs361525_1
<۰/۰۰۱	۱۸/۸۵۲	۴/۲۱۰	۸/۹۰۹	۰/۳۸۲	۲/۱۸۷	rs1800896_1
<۰/۰۰۱	۶/۲۸۲	۱/۹۶۴	۳/۵۱۳	۰/۲۹۷	۱/۲۵۶	rs1800871_1
<۰/۰۰۱			۰/۰۰۳	۰/۶۰۹	-۵/۷۴۸	ثابت مدل

شاخص $AIC = ۳۶۳/۸۷۴$

جدول ۳: ضرایب و نسبت شانس برای مدل رگرسیون لجستیک نهایی به همراه اثر متقابل برگزیده

p-value	فاصله اطمینان ۹۵٪ برای نسبت شانس		نسبت شانس	خطای معیار	ضریب رگرسیون	متغیر
	حد بالا	حد پایین				
۰/۰۱۴	۳/۸۹۶	۱/۱۶۳	۲/۱۲۹	۰/۳۰۸	۰/۷۵۶	rs1801275_1
۰/۶۵۹	۳/۸۸۲	۰/۱۱۷	۰/۶۷۴	۰/۸۹۳	-۰/۳۹۴	!
۰/۱۹۷	۱۷/۳۷۹	۰/۵۵۶	۳/۱۰۸	۰/۸۷۸	۱/۱۳۴	rs1800629_1
<۰/۰۰۱	۱۹/۱۸۲	۴/۲۲۲	۸/۹۹۹	۰/۳۸۶	۲/۱۹۷	! rs361525_1
<۰/۰۰۱	۶/۵۰۶	۲/۰۰۸	۳/۶۱۴	۰/۳۰۰	۱/۲۸۵	rs1800896_1
۰/۰۴۱	۴۳/۹۱۶	۱/۰۷۸	۶/۸۸۰	۰/۹۴۶	۱/۹۲۹	*
<۰/۰۰۱			۰/۰۱۷	۰/۹۴۰	-۴/۱۰۲	ثابت مدل

شاخص AIC = ۳۶۲/۵۸۲، * اثر متقابل: rs361525_1 & rs1800629_1

وجود ارتباط بین SNPها و بیماریها صورت گرفته است [۱۲] در این میان مطالعات صورت گرفته بارها بر نقش پلی مورفیسم های ژن TNF- بر روی بروز بیماری لوپوس اشاره شده است [۱۳-۱۵]. همچنین در مطالعات مختلفی نیز نقش سایتوکاین IL-10 در بیماری لوپوس مشخص شده بود [۱۶-۱۸] نتایج حاصل از روش شناسه گزینی منطقی در این مطالعه وجود ۲ اثر متقابل موثر بر بیماری لوپوس را که در واقع مربوط به تعامل SNPها در TNF- و IL-10 می باشد نشان می دهد. از این دو، اثر متقابل rs1800629_1 & rs361525_1! مربوط به تعامل دو SNP در TNF-، و rs361525_1 & rs1800896_1! مربوط به تعامل دو SNP در TNF- و IL-10 می باشد. با توجه به نتایج بدست آمده در جدول ۲ (مدل نهایی و بدون اثر متقابل)، نسبت شانس برای دو متغیر rs1800629_1! و rs361525_1! به ترتیب برابر ۴/۰۲۴ و ۱۷/۸۲۵ می باشد که بیان کننده این است که خطر ابتلا به بیماری در کسانی که پلی مورفیسم

در جدول ۲ متغیرها و ضرایب رگرسیونی آنها به همراه مقدار احتمال و فاصله اطمینان ۹۵٪ برای نسبت شانس در مدل رگرسیون لجستیک نهایی و بدون اثر متقابل آورده شده است. همانطور که دیده می شود ۵ متغیر rs1801275_1، rs1800629_1! و rs361525_1!، rs1800896_1 و rs1800871_1 در مدل باقی می ماند. با وارد کردن این دو اثر متقابل در مدل رگرسیون نهایی تنها اثر متقابل rs1800629_1! & rs361525_1! در مدل رگرسیون لجستیک معنی دار نشان داده شد، جدول ۳ اطلاعات مربوط به مدل نهایی رگرسیون لجستیک با اثرات اصلی و با لحاظ کردن این اثر متقابل را نشان می دهد.

بحث

با توجه به اینکه SNPها در بروز بیماریهای مرتبط به ژنتیک نقش مهمی دارند تاکنون توجه محققین بیشتر به کشف رابطه بین پلی مورفیسمها و بروز بیماریها مورد نظر بوده است و مطالعات بسیاری تاکنون در زمینه بررسی

rs361525_1 به صورت تعامل با سایر SNPها حضور دارد. همچنین می‌بینیم که در رویکرد چند درختی این متغیر به تنهایی و به عنوان ترکیب منطقی چهارم و موثر در بروز بیماری لوپوس شناخته شده است و نیز در تمام ترکیبات منطقی مربوط به رویکرد چند درختی نیز حضور داشته است. نکته دیگر اینکه دو متغیر rs1800629_1 و rs1800896_1 به شکل ترکیب با سایر متغیرها چندین بار در نمودار ۱ دیده می‌شوند که می‌تواند حاکی از اهمیت این دو متغیر به صورت تکی و بصورت تعامل با سایر پلی‌مورفیسم‌ها در بروز بیماری لوپوس باشد.

نتیجه‌گیری

هنگامی که از تعداد زیادی متغیر در آنالیزهای آماری و بخصوص رگرسیون لجستیک استفاده می‌شود، تعیین اثرات متقابلی که از نظر منطقی می‌توانند روی متغیر پاسخ اثر بگذارند ضروری به نظر می‌رسد. بنابراین با توجه به ضعف مدل‌های کلاسیک آماری خصوصاً رگرسیون لجستیک در یافتن اثرات متقابل مراتب بالا، استفاده از روش‌های نوینی مانند شناسه‌گزینی منطقی ضروری به نظر می‌رسد.

تشکر و قدردانی

بدینوسیله از دانشگاه علوم پزشکی تهران و مرکز تحقیقات روماتولوژی بیمارستان شریعتی که ما را در انجام این پژوهش یاری کردند سپاس‌گذاری می‌نماییم.

rs1800629 و rs361525 آنها از نوع هموزیگوت شایع است به ترتیب در حدود ۴ و ۱۷/۸ برابر کسانی است که پلی‌مورفیسم آنها از این نوع نیست. همچنین معیار AIC برای این مدل ۳۶۳/۸۷۴ بدست آمده است. با وارد کردن دو اثر متقابل تعیین شده توسط روش شناسه‌گزینی منطقی در مدل رگرسیون نهایی با اثرات اصلی تنها اثر متقابل rs361525_1 & rs1800629_1! معنی‌دار نشان داده شد ($P\text{-value} < 0.05$). اثر متقابل rs1800629_1 & rs361525_1! بیان می‌دارد که اگر دو پلی‌مورفیسم rs361525 و rs1800629 واقع در TNF- در فردی از نوع هموزیگوت شایع باشد آنگاه خطر ابتلا به بیماری لوپوس در فرد به مقدار اندکی افزایش می‌یابد. همانطور که دیده می‌شود با آمدن این اثر متقابل در جدول ۳، نسبت شانس در اثرات اصلی مربوط به این متغیر تغییر کرده به نحوی که از معناداری خارج شدند و مقدار برآورد پارامترهای این دو متغیر نیز تغییر قابل ملاحظه‌ای داشته‌اند که این بدین معنا می‌باشد که اثرات اصلی به تنهایی قابل استناد نیستند و اثر متقابل موثر بوده است. همچنین مقدار شاخص AIC در این مدل برابر ۳۶۲/۵۸۲ شده است که نسبت به مدلی که اثر متقابل در آن لحاظ نشده بود کاهش یافته است.

با کمی دقت در نمودار ۱ مربوط به رویکرد تک درختی می‌بینیم که در ۷ ترکیب منطقی اول متغیر

antagonist in patients with systemic lupus erythematosus could predict renal manifestation of the disease, Human immunology 2010;71(9):874.

5. Hirankarn N, Avihingsanon Y, Wongpiyabovorn J, Genetic susceptibility to SLE is associated with TNF-alpha gene polymorphism-863, but not-308 and-238, in Thai population, International journal of immunogenetics. 2007;34(6):425-30.

6. Aringer M, Smolen J. Tumour necrosis factor and other proinflammatory cytokines in systemic lupus erythematosus: a rationale for therapeutic intervention, Lupus 2004;13(5):344-7.

7. Schwender H, Ickstadt K, Identification of SNP interactions using logic regression, Biostatistics 2008;9(1):187-98.

References

- Noori S, Nourijelyani K, Mohammad K, "et al", Random Forests Analysis: A modern statistical method for screening in high-dimensional studies and its application in a population-based genetic association study Journal of North Khorasan University of Medical Sciences 2011;3:49-57[Persian]
- Sadegh valian broujeni ea, Personalized medicine: Therapy based on genomic variations, Laboratory Diagnosis 2012;10-8[Persian]
- Lindqvist A, Alarcón-Riquelme ME, The genetics of systemic lupus erythematosus, Scandinavian journal of immunology 1999;50:562-71.
- Brugos B, Kiss E, Dul C, Gubisch W, " et al", Measurement of interleukin-1 receptor

8. Hamid Alavi Majd FP, Yadollah Mehrabi, Maryamossadat Daneshpour, Majid Mirmohammadkhani, Mahdi Hedayati FA, Logic Features Selection in Identification of the Most Important Interactions of Interleukin-6 and Two Important Single Nucleotide Polymorphisms of Its Gene (IL-6-174, IL-6-572) with Some Other Factors in Association with hypertension, *Knowledge & Health* 2012;7(3):93-100[Persian]
9. Moore JH, Dunlap JC. *Computational Methods for Genetics of Complex Traits*: Academic Press; 2010.
10. Ruczinski I, Kooperberg C, LeBlanc M, Logic regression, *Journal of Computational and Graphical Statistics* 2003;12(3):475-511.
11. Hosmer DW, Lemeshow S, *Applied logistic regression*: Wiley-Interscience; 2004.
12. Dinu I, Mahasirimongkol S, Liu Q, Yanai H, Eldin NS, Kreiter E, "et al", SNP-SNP interactions discovered by logic regression explain crohn's disease genetics, *PloS one* 2012;7(10):e43035.
13. Rood M, Van Krugten M, Zanelli E, Van Der Linden M, Keijsers V, Schreuder G, "et al", TNF-308A and HLA-DR3 alleles contribute independently to susceptibility to systemic lupus erythematosus, *Arthritis & Rheumatism* 2000;43(1):129-34.
14. Wilson AG, Gordon C, di Giovine FS, de Vries N, van de Putte L, Emery P, "et al", A genetic association between systemic lupus erythematosus and tumor necrosis factor alpha, *European journal of immunology* 1994;24(1):191-5.
15. Sullivan KE, Wooten C, Schmeckpeper BJ, Goldman D, Petri MA, A promoter polymorphism of tumor necrosis factor associated with systemic lupus erythematosus in African-Americans, *Arthritis & Rheumatism* 1997;40(12):2207-11.
16. Chong W, Ip W, Wong WH, Lau C, Chan T, Lau Y, Association of interleukin-10 promoter polymorphisms with systemic lupus erythematosus, *Genes and immunity* 2004;5(6):484-92.
17. Eskdale J, Wordsworth P, Bowman S, Field M, Gallagher G, Association between polymorphisms at the human IL-10 locus and systemic lupus erythematosus, *Tissue antigens* 1997;49(6):635-9.
18. Khoa PD, Sugiyama T, Yokochi T, Polymorphism of interleukin-10 promoter and tumor necrosis factor receptor II in Vietnamese patients with systemic lupus erythematosus, *Clinical rheumatology*, 2005;24(1):11-3.

Original Article

Identification of SNP Interactions of Inflammatory cytokines and anti-inflammatory genes in Systemic Lupus Erythmatosus's Disease Using Logic Feature Selection

Mehravar S¹, Nourijelyani K², Azam K^{3*}, Rahimi Foroushani A², Mahmoudi M⁴, Amirzargar AA⁵, Ghodsi Ghasem abadi R¹

¹ M.Sc. in Biostatistics, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences

² Associate professor of Biostatistics, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences

³ Assistant Professor of Biostatistics, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences

⁴ Assistant Professor of Immunology, Rheumatology center, Tehran University of Medical Sciences

⁵ Assistant Professor of Immunology/Immunogenetic Laboratory, Department of Immunology, Faculty of Medicine, Tehran University of Medical Sciences

***Corresponding Author:**

Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
Email: kazam@sina.tums.ac.ir

Abstract

Background & objectives: Despite extensive studies conducted on single nucleotide polymorphisms (SNPs) effects on different diseases, but the necessity of investigation into interaction of these SNPs together in diseases in particular genetic based diseases in one hand and existence of a large number of these variables and also weakness of classical statistical models in including such variables in the other hand, underline the necessity of applying new statistical method to determine these interaction effects. In this study we used logic feature selection approach to determine important SNP interactions associated with systemic lupus erythematosus.

Material & Methods: We use information of 11 SNPs located in the genes IL-1 cluster, IL-10, TNF- α , IL-4Ra for 376 subjects (188 control subjects and 188 case) to determine influential SNP interactions in patients affected with systemic lupus erythematosus disease using logic feature selection. Consequently, the results were used in logistic regression model.

Results: the result of the logic feature selection approach led to the identification of two two-way interactions, but in the final logistic regression model only two-way interaction related to polymorphisms in TNF- α was shown to be significant on occurrence of systemic lupus erythematosus disease.

Conclusion: According to our results, it seems that the results of the modern statistical methods like logic feature selection can be helpful to find the important SNP interactions associated with disease in studies with large number of variables.

Keywords: Logic Feature selection, Interactions, Single Nucleotide Polymorphisms, Systemic Lupus Erythematosus

Submitted: 1 June 2013

Revised: 20 Aug 2013

Accepted: 7 Sep 2013