

## به کارگیری آمار غیر خطی (آنالیز مولفه اصلی) در آنالیز داده های تمایز سلول بنیادی به آستروسیت

حکیمه زالی<sup>1</sup>، مصطفی رضایی طایرانی<sup>2\*</sup>، علی صید خانی نهال<sup>3</sup>، شیمیا مرادی<sup>4</sup>

- 1) دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی تهران
- 2) مرکز تحقیقات پروتئومیکس، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی تهران
- 3) گروه بیوشیمی دانشکده پیراپزشکی، دانشگاه علوم پزشکی ایلام
- 4) گروه میکروبیولوژی، دانشکده پزشکی، دانشگاه علوم پزشکی ایلام

تاریخ دریافت:

تاریخ پذیرش:

### چکیده

**مقدمه:** ترکیب آمار تک متغیره و چند متغیره می تواند به شناسایی تغییرات بیولوژیکی واقعی معنی دار در بیان پروتئین ها بین گروههای آزمایشی مختلف کمک نماید. یکی از معمولی ترین روش های آماری که کمک بسیاری به آنالیز ژل های الکتروفورز دو بعدی می کند آنالیز مولفه اصلی (PCA) است. در این مطالعه فرآیند تمایز سلول های بنیادی به آستروسیت ها مورد بررسی پروتئومیکی قرار گرفت و پروتئوم دو نوع سلول توسط نرم افزار آماری مورد آنالیز مولفه اصلی قرار گرفت .

**مواد و روش ها:** آسپیره مغز استخوان از دهنده سالم، تهیه شد و جداسازی مونونوکلئو سل انجام شد. سلول ها در 10 درصد DMEM با گلوکز پایین، گلوتامین، استرپتومایسین و پنی سیلین در شرایط 5 CO<sub>2</sub> درصد و رطوبت 98 درصد در دمای 37 °C انکوبه شدند. برای تمایز این سلول های بنیادی به سلول های آستروسیتی، سلول ها در مجاورت رتینوئیک اسید، cAMP، PDGF، NGF قرار گرفت. از دو رده سلولی سلول های بنیادی و سلول های آستروسیت پروتئین استخراج شد و جداسازی پروتئین ها با الکتروفورز دو بعدی صورت گرفت. به روش رنگ آمیزی نیترات نقره ژل ها رنگ شدند و ژل های اسکن شده با روش های بیوانفورماتیکی و نرم افزار های آماری مورد آنالیز بیوانفورماتیکی و آماری قرار گرفت.

**یافته های پژوهش:** آنالیز بیوانفورماتیکی و آماری ژل های بدست آمده از تکنیک الکتروفورز دو بعدی 774 نقطه پروتئینی در دو گروه مشاهده شد که مقایسه بین گروه ها حکایت از بیان پروتئینهای جدید و خاموشی برخی از پروتئین ها در مسیر های سیگنالینگ تمایز سلولی است. آنالیز خوشه بندی پروتئین ها را از نظر بیان به 3 خوشه اصلی تقسیم نمود که بیانگر وجود پروتئین هایی با بیان مشابه در هر خوشه است که این پروتئین ها می توانند عملکرد مشابهی را در شرایط آزمایش ارائه نمایند یا بیانگر حضور همه آن ها در مسیر بیولوژیکی مشترکی است. آنالیز PCA نتایج حاصل از خوشه بندی را تایید نمود و نشان داد که داده پروتئینی بر طبق شرایط آزمایش خوشه بندی شده است. در نهایت می توان نتیجه گرفت که تمایز باعث تغییر بیان معنی داری در سطح پروتئوم می شود که این تغییرات را می توان با آنالیزهای آماری خوشه بندی و PCA به خوبی مورد بررسی قرار داد و شاخص های تغییرات را مشخص نمود.

واژه های کلیدی: پروتئومیک، آنالیز اجزا اصلی، آستروسیت

\* نویسنده مسئول: مرکز تحقیقات پروتئومیکس، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی تهران

## مقدمه

تجزیه و تحلیل‌های پروتئومیکس امروزه به عنوان یکی از روش های مهم در مطالعات بیان ژن و ژنومیکس عملکردی محسوب می شوند. استفاده از این روش ها به چند دلیل ضرورت دارند. امروزه مشخص شده است که بین سطح m-RNA و پروتئین ها و حتی گاهی توالی آنها رابطه دقیقی برقرار نیست. توالی های DNA و RNA اطلاعات بسیار کمی را در مورد جایگاه پروتئین ها، تغییرات پس از ترجمه و نیمه عمر پروتئین ها به ما می دهند. از طرفی با استفاده از روش های پروتئومیکس تعداد زیادی از پروتئین ها را می توان هم زمان (صدها حتی هزاران پروتئین) در یک ترکیب پیچیده ای مانند محتویات یک سلول لیز شده مورد بررسی قرار داد. با توجه به تعداد بیشتر متغیرها نسبت به تعداد نمونه ها در داده های پروتئومیکی استفاده از روش های آماری چند متغیره جهت آنالیز و مقایسه تغییرات بیان پروتئین ها در نمونه ها ضروری به نظر می رسد. یکی از روش های آنالیز برای این نوع از داده ها استفاده از آنالیز اجزا اصلی (PCA) است که هدف آن کاهش حجم داده ها و تعبیر و تفسیر راحت تر آن ها می باشد. PCA یک تکنیک مفید آماری است که کاربرد آن در زمینه های از قبیل تشخیص چهره، فشرده سازی تصویر و یک تکنیک رایج برای شناسایی یک الگو در داده هایی از بعد بالا است (2و1). این تبدیل که با اسامی دیگری چون تبدیل هتلینگ (Hostelling Transform)، تبدیل کارهانن - لو (Karhunen-Live Transform (KLT)) و بردار های ویژه نیز شناخته می شود، تبدیل بهینه در کارهای فشرده سازی و کاهش بعد است و خطای میانگین مربعات حاصل از فشرده سازی را کمینه می کند. هر چند این تبدیل به علت وابسته بودن به داده ورودی، جای خود را در الگوریتم های کاربردی و عملی، به تبدیل گسسته کسینوسی (Discret Cosine Transform (DCT)) داده است اما در صورت کافی بودن داده ورودی، می تواند تبدیل بهینه را استخراج نماید (1-5).

سلول هایی بنیادی، سلول های تخصص نیافته هستند و به مدت های طولانی از طریق تقسیم سلولی

خود را نوسازی می کنند تحت شرایط تحریک فیزیولوژیکی یا آزمایشی، تبدیل به سلول هایی با عملکردهایی ویژه می شوند. رسیدن سیگنال تمایز به سلول بنیادی باعث می شود بعضی از ژنهایش را روشن کند و پروتئین های جدیدی را بسازد که این پروتئینها کمک می کند که سلول بنیادی مشابه به سلول تمایز یافته شود. بنابراین تغییرات شدیدی در سطح بیان ژن منجر به این تبدیل می شود که منجر به خاموشی یا روشن شدن مسیرهای سیگنالینگ متعددی در سلول می شود که همه این اتفاقات در سطح پروتئوم قابل ردیابی و آشکارسازی است. با توجه به ماهیت داده های پروتئومیکی که دارای متغیرهای متعددی است استفاده از روش آنالیز مولفه اصلی که می تواند همزمان همه نقاط پروتئینی موجود را مورد بررسی قرار دهد و تفاوت های کلی بیان ژن را بین گروه های مختلف درگیر در آنالیز را شناسایی نماید روش موثری است. از طرفی برای انجام آن به سادگی می توان از نرم افزارهای آماده موجود در بازار نیز استفاده نمود (6-10). در این مطالعه فرآیند تمایز سلول های بنیادی به آستروسیت ها توسط پروتئومیکس مورد بررسی قرار داده و نتایج حاصله که شامل پروتئوم دو نوع سلول است توسط نرم افزار آماری مورد آنالیز مولفه اصلی قرار می گیرد تا بتواند به عنوان یکی از بهترین روش های آماری اختلاف و شباهت دو سلول را در سطح پروتئوم بیان نماید.

## مواد و روش ها

### نمونه گیری:

آسپیره مغز استخوان از دهنده سالم، تهیه شد و به دنبال آن، جداسازی مونونوکلئوئر سل ها (سلول های تک هسته ای) با استفاده از ستون فایکول با دانسیته 1.077 انجام گرفت. مونونوکلئوئر سل ها به فلاسک های حاوی DMEM 10 درصد با گلوکز پایین، گلوتامین، استرپتومایسین و پنی سیلین انتقال یافت. انکوباسیون سلول ها در شرایط CO<sub>2</sub> 5 درصد و رطوبت 98 درصد در دمای 37°C انجام شد (11).

اولین پاساژ سلول بنیادی پس از 4 روز با تریپسین 0.25 درصد و EDTA 0.5 درصد صورت گرفت در مرحله بعد برای تمایز این سلول ها به سلول های

بعد دوم شامل ژل پلی آکریل آمید با غلظت 12 درصد حاوی 15 میلی لیتر از محلول ذخیره 30 درصد (آکریل آمید به علاوه بیس آکریل آمید با نسبت 29/2 درصد به 0/8 درصد)، 18/76 میلی لیتر آب دی‌یونیزه و 11/25 میلی لیتر بافر تریس و همراه 30 میکرولیتر TEMED و 30 میکرولیتر آمونیم پرسولفات است. بافر الکتروفورز (Running buffer) به تانک های بالا و پائین افزوده می‌شود. این بافر شامل 0/025 مولار تریس - اسید کلریدریک (یا باز)، 0/192 مولار گلاسیسین و 0/1 درصد SDS می‌باشد. در نهایت پس از برقراری جریان، الکتروفورز در جریان ثابت 30 الی 40 میلی آمپر برای هر پلیت و با استفاده از سیستم خنک شده در دمای 9 درجه سانتیگراد انجام گردید. رنگ آمیزی پروتئین ها به روش رنگ آمیزی نیترا تفره انجام شد و ژل ها اسکن و تصاویر پردازش شد و در نهایت با روش های بیو انفورماتیکی و نرم افزار های آماری مورد آنالیز بیوانفورماتیکی و آماری قرار گرفت(13).

#### آنالیز تصویر ژل های دوبعدی

الکتروفورز دوبعدی قادر است صدها پلی پپتید را از یک محلول پروتئینی بر اساس بار الکتریکی و وزن مولکولی آنها جدا کند. پس از طی مراحل رنگ‌آمیزی، پروتئین‌ها بصورت لکه‌هایی با صفات و ویژگی‌های ظاهری متفاوت مانند شکل، اندازه و شدت رنگ نمایان می‌گردند. پس از اسکن ژل، استخراج داده‌های مربوط به دو گروه با نرم افزار Progenesis Same Spot Non Linear مورد بررسی قرار گرفت. تعداد نقاط پروتئینی و تفاوت آن‌ها (افزایش و یا کاهش بیان) توسط این نرم افزار آنالیز شد و با آمار چند متغیره (آنالیز مولفه اصلی) پروتئین‌ها از نظر بیان خوشه بندی شدند.

#### یافته های پژوهش

نرم افزار های آنالیز ژل های الکتروفورز دو بعدی با بهره گیری از روش های تشخیص الگو و با کمک روش های آمار تک متغیره و چند متغیره به دنبال شناسایی نقاط و تطابق آن‌ها در ژل های مختلف گروه های تحت آنالیز هستند تا تفاوت پروتئومیکی بین حالت های سالم و بیماری یا مراحل مختلف یک

آستروسیتی، سلول های کشت داده شده در مجاورت رتینوئیک اسید، cAMP، PGF، PDGF، NGF قرار گرفت.

#### استخراج پروتئین:

از دو رده سلولی تمایز نیافته (سلول های بنیادی) و سلول های تمایز یافته (سلول های آستروسیت) پروتئین استخراج شد بدین ترتیب که ابتدا فلاسک های حاوی سلول را با PBS دوبار شستشو داده سپس سلول ها را با کمک تریپسین جدا نموده و به آن ها بافر سرد تریس 20 میلی مولار با pH=7/5 اضافه می‌شود. این بافر حاوی 2 میلی مولار اتیلن دی آمین تترا استات 1، 1 میلی مولار فنیل متیل سول فونیل فلوراید 2، 32/3 مولار سوکروز به همراه مخلوطی از آنتی پروتئازها شامل (10 میلی مولار N - اتیل مالامید، 2 میلی مولار ایدواستات، 25 میلی مولار بنز آمیدین، 12/5 میلی مولار 6-آمینو کاپروئیک اسید) اضافه می‌شود. سلولها در درجه حرارت پائین بکمک دستگاه هموژنیز همگون می‌گردد. مخلوط هموژنیزه شده بمدت 10 دقیقه در دمای 10 درجه سانتیگراد در 1000 g سانتریفیوژ شد. مایع روئی حاصله با استفاده از سانتریفیوژ دور بالا بمدت 30 دقیقه در 40000g در 4 درجه سانتیگراد سانتریفیوژ می‌شود.

رسوب حاصل از سانتریفیوژ در حجم یک میلی لیتر بافر استخراج یکنواخت می‌شود. سپس دوباره در 40000g بمدت 30 دقیقه در 4 درجه سانتیگراد سانتریفیوژ می‌گردد. رسوب حاصل در 400 میکرولیتر بافر اولیه و 100 میکرولیتر بافر نمونه (5x) بتدریج و با تکان دادن حل می‌گردد. پس از انجام سانتریفیوژ و تهیه رسوب پروتئینی، غلظت پروتئینهای دو گروه آزمایشی به روش برادفورد (12) مورد سنجش قرار گرفت.

#### الکتروفورز دوبعدی

در بعد اول، IEF (Iso Electric Focusing) ، از نوارهای ژلی IPG (Imobilized PH Gradient) در رنج PH 3 تا 10 استفاده شد و پروتئین ها درون دستگاه IPG بسته به PH ایزوالکتریک خود روی نوار استریپ تفکیک شدند.

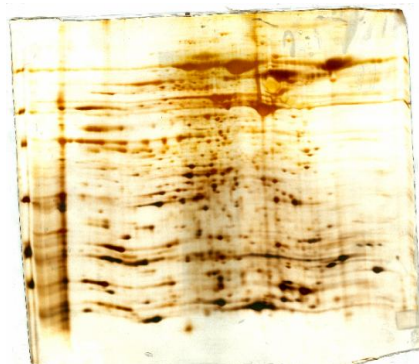
طوری که پروتئین هایی که در یک خوشه ، دارای بیان مشابهی باشند ؛ بدین معنی که تغییر بیان یکسانی را در سلول مورد نظر دارند (با هم ، افزایش و یا کاهش می یابند) که نتایج حاصل از این آنالیز خوشه بندی ، در شکل 3 مشاهده می نماید. حاصل این آنالیز سه خوشه اصلی پروتئینی می باشد.

شکل 4 آنالیز PCA دو بعدی را در دو نمونه مورد آزمایش نشان می دهد که این آنالیز دقیقاً منطبق بر انتخاب داده هایی است که در خوشه بندی شکل 3 صورت گرفته است در حالی که آنالیز PCA در شکل 5 آنالیز هر خوشه پروتئینی در دو گروه را نشان می دهد. شکل 5-الف میزان بیان بالاتر در سلول های بنیادی نسبت به آستروسیت است در حالی که در 5-ب میزان بیان بالاتر در سلول های آستروسیت نسبت به بنیادی را نشان می دهد.

بیماری و حالت های تحت انکوباسیون با دارو و کنترل آن را شناسایی نمایند و پروتئین های مارکر را که می توانند نقش تشخیصی یا هدف های دارویی باشند در گروه های مورد مقایسه کشف نمایند. در این مطالعه از نرم افزار Non Linear Progenesis Same Spot جهت آنالیز آماری دو گروه سلول بنیادی و آستروسیت استفاده شد.

در شکل 1 ، تصویر ژل الکتروفورز دو بعدی از سلول های بنیادی (الف) و آستروسیت (ب) را نشان می دهد. آنالیز ژل ها با کمک نرم افزار Progenesis 774 نقطه پروتئینی را مشخص کرد که تعدادی از آن ها دارای افزایش بیان و تعدادی دارای کاهش بیان بودند . جایگاه نقاط در ژل مرجع در شکل 2 نشان داده شده است .

این نرم افزار خوشه بندی پروتئین ها بر اساس میزان بیان آن ها را مورد آنالیز قرار می دهد ، به

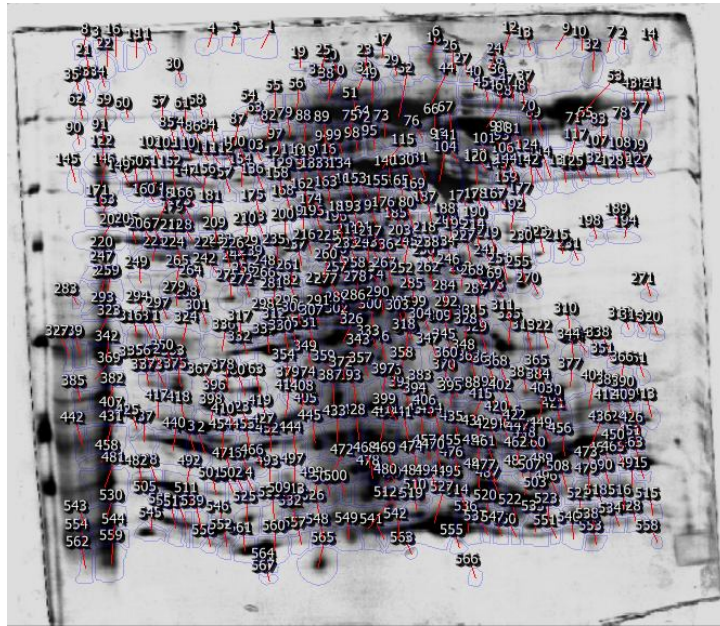


(الف)

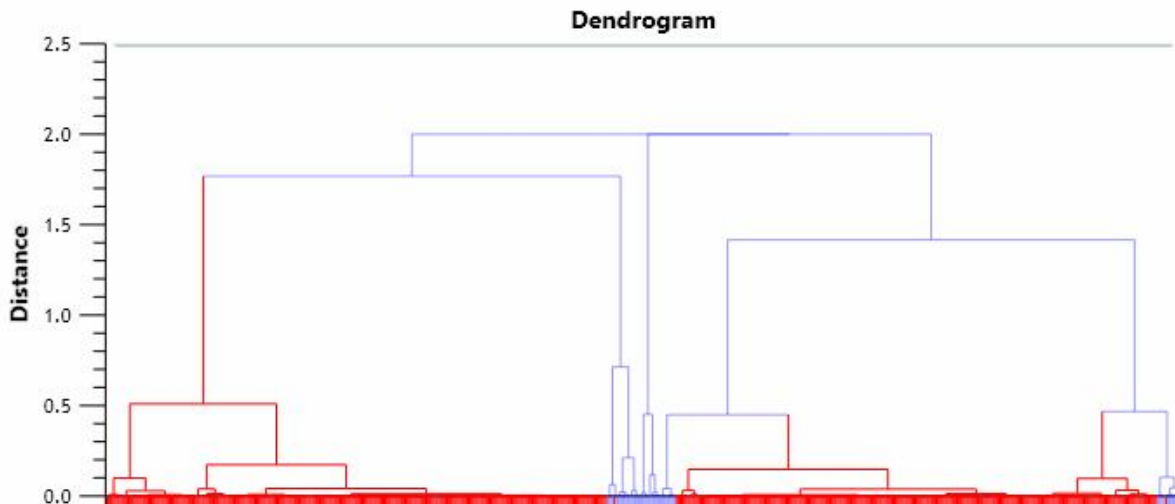


(ب)

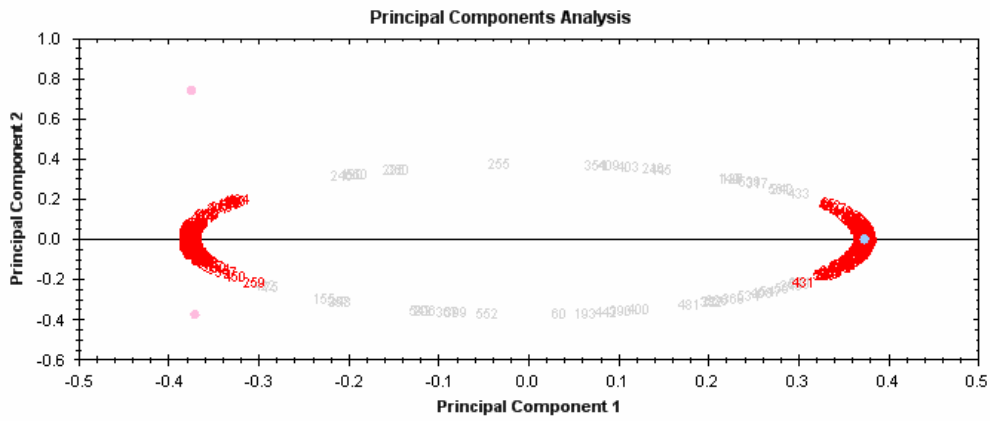
شکل 1: شکل ظاهری ژل های دو بعدی برای سلول های بنیادی (الف) و آستروسیت مشتق از آن (ب)



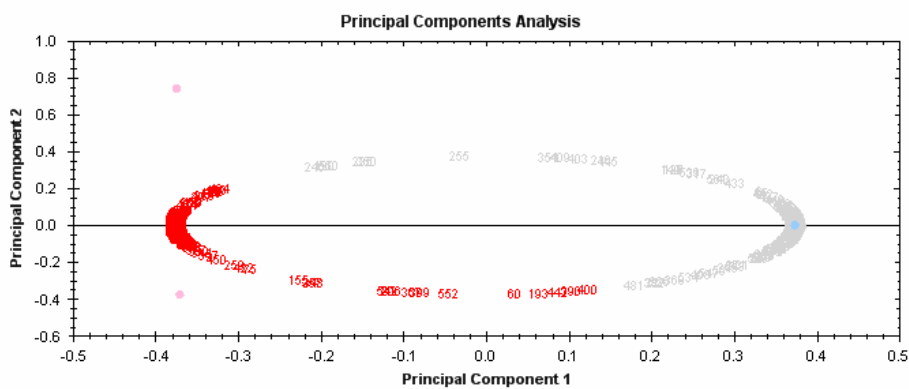
شکل 2: شکل ژل دو بعدی مرجع آنالیز شده توسط نرم افزار Progenesis



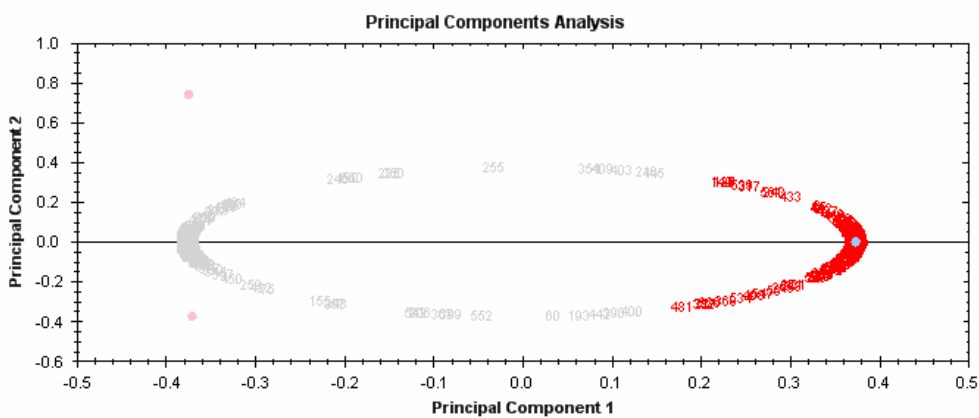
شکل 3: خوشه بندی پروتئین های به دست آمده توسط نرم افزار Progenesis



شکل 4: آنالیز PCA خوشه های پروتئینی انتخاب شده در شکل 3



الف



ب

شکل 5 - آنالیز PCA تفاوت خوشه پروتئینی در دو گروه: میزان بیان بالاتر در سلول های بنیادی نسبت به آستروسیت (الف) و میزان بیان بالاتر در سلول های آستروسیت نسبت به بنیادی (ب) را نشان می دهد

## بحث و نتیجه گیری

آستروسیت ها گروه هتروژنی از انواع سلول های غیر عصبی هستند که دارای عملکردهای متنوعی هستند. در مهاجرت و تمایز سلول های بنیادی عصبی نقش دارند (8و7). آستروسیت ها قادرند از سلول های بنیادی طی فرآیند تمایز تکامل یابند. در واقع، تمایز فرآیندی است که در آن، سلول های غیر تخصصی، به سلول های تخصصی یک بافت خاص تبدیل می شوند و انواع مختلف آن شامل تمایز ریختی، ساختاری، سوخت و سازی و فیزیولوژیکی است (9). یکی از راه های بررسی تغییرات کلی حاصل طی این فرآیند مطالعه دو نوع سلول با روش های پروتئومیک و سپس آنالیز های بیوانفورماتیکی است. پروتئومیک به مطالعه پروتئوم که در یک سلول و در یک زمان مشخص بیان می شود، می پردازد. تغییرات بیان در سطح پروتئوم با کمک تکنیک ژل الکتروفورز دو بعدی مورد آنالیز پروتئومیک قرار گرفت. شکل 1 ژل های دو گروه سلول بنیادی و آستروسیت را نشان می دهد. امروزه استفاده از نرم افزار های آنالیز ژل کمک بیوانفورماتیک را در ساده نمودن آنالیز داده های پروتئومیک امری اجتناب ناپذیر نموده است. پیشرفت هایی که در سطح نرم افزاری صورت گرفته است آنالیز های آماری چند متغیره را در داده های چند بعدی پروتئومیک تسهیل نموده است. در این مطالعه ژل های الکتروفورز دو بعدی با استفاده از نرم افزار *progenesis same spot* مورد آنالیز قرار گرفتند که نتایج یکی از آنالیز ها در شکل 2 نشان داده شده است. مقایسه دو گروه نشان داد که 774 نقطه پروتئینی در دو گروه مشاهده شده است که مقایسه بین ژل های دو گروه در سطح بیان پروتئینی حکایت از بیان نقاط پروتئینی جدید و خاموشی بیان برخی نقاط دیگر است. این نتیجه بیانگر این است که تمایز باعث تغییراتی وسیعی در سطح بیان ژنی می شود.

نرم افزار *progenesis same spot* با استفاده از روش آنالیز همبستگی، خوشه بندی پروتئین ها بر اساس میزان بیان آن ها را مورد آنالیز قرار می دهد، به طوری که پروتئین هایی که در یک خوشه دسته بندی می شوند، دارای بیان مشابهی می باشند؛ بدین

معنی که تغییر بیان یکسانی را در حالت سلولی مورد نظر دارند (با هم، افزایش و یا کاهش می یابند). در شکل 3 آنالیز خوشه بندی، پروتئین ها را از نظر بیان به 3 خوشه اصلی تقسیم نموده است. خوشه سمت چپ دسته پروتئینی را نشان می دهد که دارای بیان بیشتری در گروه سلول بنیادی نسبت به آستروسیت بوده اند و خوشه میانی به همان نسبت که سمت راست می رود این نسبت تغییر می یابد به طوری که در خوشه آخر در سمت راست تغییر بیان در آستروسیت بیشتر از سلول بنیادی است. خوشه بندی در این مطالعه بر اساس میزان بیان بود و نشان دهنده وجود پروتئین هایی است که می توانند وابستگی بیولوژیک نسبت به هم داشته باشند، به طوری که افزایش بیان همزمان چندین پروتئین باید در جبران کاهش بیان یا عدم بیان پروتئین های دیگر باشد. به طوری که تفاوت های بین دو حالت مختلف سلولی مانند پیری و جوانی یا سالم و بیمار و در نهایت سلول تحت انکوباسیون با یک دارو در مقابل کنترل آن بدون حضور دارو می توانند منجر به بروز مسیر سیگنالینگ ویژه ای در سلول شوند که در همراهی با بیان کمپلکس های پروتئینی آن مسیر شوند و به طور ناگهانی با ظهور پروتئین های جدیدی در سلول مواجه شویم. مطالعات پیشین نیز که با کمک این نرم افزار به آنالیز ژل های دو بعدی پرداخته اند نیز نتایج مشابهی را نشان می دهند و بیان می کنند که پروفایل بیانی مشابه اعضای یک خوشه دارای تفسیر بیولوژیکی منطقی است (14-18).

آنالیز آماری چند متغیره دیگری به نام آنالیز مولفه اصلی (PCA) برای تایید آنالیز خوشه بندی توسط نرم افزار انجام می شود که در شکل 4 و 5 نشان داده شده است. آنالیز اجزا اصلی (PCA) یک تکنیک مفید آماری است که برای شناسایی یک نمونه در داده های از بعد بالا است. آنالیز اجزای اصلی یک روش اختیاری چند متغیری است. اگر ما در جایی مجبور هستیم مهم ترین متغیر را یا یک تعداد محدودی از متغیر ها را در یک مجموعه انتخاب کنیم از آنالیز اجزای اصلی کمک می گیریم. این روش یکی از الگوهای تشخیص و شناسایی (تشخیص هویت) در یک مجموعه اطلاعات است. در این روش اطلاعات را بر اساس شباهت ها و

های مختلف خوشه بندی را مورد ارزیابی قرار دهد و outliers موجود در مجموعه داده را تعیین نماید (19). از PCA سه بعدی در تصاویر الگوهای پروتئومیکی جهت شناسایی کلاس نمونه های موجود در یک مجموعه داده استفاده می شود. روش پیشرفته تری جهت دو مجموعه داده مختلف به کار گرفته شد. مجموعه داده ای متشکل از 5 نمونه از سرم موش سالم و 5 نمونه سرم موش هایی که نیکوتین دریافت کرده بودند. مجموعه داده دوم شامل داده های 4 نمونه غده لنفاوی سالم و 4 نمونه غده لنفاوی مبتلا به لنفومای غیر هوچکینی بود. با کمک این روش موفق به شناسایی کلاس های نمونه های موجود در هر دو گروه ژل دو بعدی شدند و اجازه داد تا نواحی از نقشه های دو بعدی مسئول تفاوت های موجود در بین کلاس ها برای هر دو مجموعه داده سرم موشی و لنفومای انسانی مورد شناسایی قرار گیرد (20).

به طور کلی می توان این نتیجه را گرفت تفاوت معنی داری بین دو گروه سلولی بنیادی و آستروسیت در فرآیند تمایز سلولی وجود دارد و این تغییرات در سطح ملکولی با کاهش و افزایش بیان پروتئین ها و حضور یا ناپدید شدن پروتئین ها خود را نشان می دهد. این تغییرات را در سطح ملکولی با آنالیزهای آماری چند متغیره ای نظیر کلاسترینگ یا آنالیز همبستگی و آنالیز مولفه اصلی به خوبی آشکار گردید و گروههای پروتئینی که دارای تغییر بیان بودند خود را به صورت خوشه های مجزایی نشان دادند که برای شناسایی مکانیسم ملکولی تمایز و تعیین مسیرهای بیولوژیک نیاز به مشخص شدن همه پروتئین های هر خوشه است که باید توسط دستگاه اسپکتروفوتومتر جرمی تعیین هویت شوند.

### سپاسگزاری

این مطالعه حاصل طرح تحقیقاتی با کد -1-1391-96-8969 است که توسط کمیته پژوهشی دانشجویان دانشگاه علوم پزشکی شهید بهشتی تصویب گردیده است. از مرکز تحقیقات پروتئومیکس دانشگاه علوم پزشکی شهید بهشتی که در اجرای پروژه همکاری داشتند تشکر می گردد.

تفاوت هایشان بیان می کنند. از آن جا که در اطلاعات از ابعاد بالا، نقشه و طرح خاصی را به سختی می توان در داده ها پیدا کرد در حقیقت آنالیز اجزای اصلی ارتباط بین داده ها را کشف می کند و در جایی که نعمت نمایش گرافیکی در دسترس نیست، آنالیز اجزای اصلی یک ابزار نیرومند برای آنالیز اطلاعات است. مزیت دیگر آن این است که کاهش تعداد ابعاد بدون از دست دادن مقدار زیادی از اطلاعات انجام می پذیرد و هدف اصلی خلاصه کردن داده ها است و به عنوان یک وسیله دسته کننده اطلاعات مورد توجه نیست (1-4).

PCA کاربردهای زیادی در داده های چند بعدی پروتئومیکی دارد. نتایج آنالیز PCA پراکندهی اعضا هر خوشه در اطراف یک بردار را نشان می دهد و بیانگر این است که نقاط نزدیک به یکدیگر متعلق به یک کلاس هستند که در پروتئومیکس تنوع را در بین مجموعه داده می تواند دسته بندی کند و اعضای هر دسته را نیز مشخص نماید. از طرفی اگر گونه های متفاوتی از داده مورد استفاده قرار گیرد می تواند خوشه های هر گونه را نیز مشخص نماید و نشان دهد که همه نمونه های یک گونه در اطراف یک بردار واضح قرار می گیرند و همه خوشه ها به هم نزدیک هستند. در هر دو شکل 4 و 5 جایگاه پروتئین های هر دو گروه را نشان می دهد و بیانگر این است پروتئین هایی که متمایل به یک سمت هستند خوشه های مشترک زیادی با هم دارند. از طرفی PCA بیانگر این است که هیچ outlierی در داده وجود ندارد و داده بر طبق شرایط آزمایش خوشه بندی شده است. در مطالعه پروتئومیکی دیگری نیز از PCA جهت ارزیابی مجموعه داده استفاده شده بود همه نمونه های N.meningitidis به طور واضحی در اطراف یک بردار قرار گرفتند و نمونه های که توسط خوشه بندی به روش شبکه عصبی مصنوعی خوشه بندی نشده بودند و در کلاس ها قرار نگرفته بودند در فاصله ای از آنها که خوشه بندی شده بودند قرار گرفتند و نشان می داد که با توجه به ماهیت داده ها که از دو گونه متفاوت بوده است PCA می تواند خوشه های تشکیل شده از روش



## References

- 1-Abdi H, Williams LJ. Principal component analysis. *Comput Statist* 2010;2: 433-59.
- 2-Jolliffe I.T. *Principal Component Analysis*. 2nd ed. Springer:NY; 2002. P.48-28.
- 3-Miranda AA, Le-Borgne YA, Bontempi G. New routes from minimal approximation error to principal components. *Neural Processing Letters:Springer*; 2008.
- 4-Urfer W, Grzegorzczak M, Jung K. study, a protocol on image acquisition has also been included. *Proteomics* 2006;6: 48-55.
- 5-Harrington BP, Vieira EN, Espinoza J, Nien JK, Romero R, Yergey AL. Analysis of variance–principal component analysis: A soft tool for proteomic discovery. *Analytica Chimica Acta* 2005; 544:118-27.
- 6-Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nature Genet* 2003; 33: 311-23.
- 7- Song H, Stevens CF, Gage FH. Astroglia induce neurogenesis from adult neural stem cells. *Nature* 2002;417:29-32.
- 8- Stevens CF. Astrocytes trigger maturation of neural stem cells. *Howard Hughes Med Inst News* 2002;312:22-27.
- 9- Mckay R. Stem cell in the central nervous system. *Science* 1997;276: 66-71.
- 10-Bishop CM, Tipping M E. A hierarchical latent variable model for data visualization. *IEEE* 1998;20: 281–293.
- 11- Song H, Stevens C. Neural stem cells from adult hippocampus develop essential properties of functional CNS neurons. *Nature Neurosci* 2002;5: 438-45.
- 12-Bradford, MM. A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein. *Analyt Biochem* 1976;72: 248-54.
- 13-Zhou G, Hongmei L, Decamp D. 2D differential in-gel electrophoresis for the identification of esophageal cancer cell cancer-specific protein markers. *Mol Cell Proteomics* 2002; 1: 117-23.
- 14-O’Gorman M, Beauvallet C. An investigation into Crohn’s disease using the proteogenomics same spots analysis platform. *Nonlinear Dynamics, Newcastle-upon-Tyne, UK, Institut National de Recherche Agronomique (INRA), Jouy-en-Josas, France, Hopital Saint-Antoine, Paris, France.*
- 15-Bedon F, Villar E, Vincent D, Dupuy JW, Lomenech AM, Mabialangoma A, et al. Proteomic plasticity of two Eucalyptus genotypes under contrasted water regimes in the field. *Plant Cell Environ* 2012;35:790-805.
- 16-Bièche C, de Lamballerie M, Chevret D, Federighi M, Tresse O. Dynamic proteome changes in *Campylobacter jejuni* 81-176 after high pressure shock and subsequent recovery. *J Proteomics* 2012;75:1144-56.
- 17-Brunelli L, Campagna R, Airolidi L, Cauli O, Llansola M, Boix J, et al. Exploratory investigation on nitro- and phospho-proteome cerebellum changes in hyperammonemia and hepatic encephalopathy rat models. *Metab Brain Dis* 2012;27:37-49.
- 18-Chan K, Casjens S, Parveen N. Detection of Established Virulence Genes and Plasmids To Differentiate *Borrelia burgdorferi* Strains. *Infect Immun* 2012;80:1519-29.
- 19-Lancashire L, Schmid O, Shah H, Ball G. Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. *Bioinformatics* 2005; 21:2191-9.
- 20-Emilio Marengo et al. Application of Three-Way Principal Component Analysis to the Evaluation of Two-Dimensional Maps in Proteomics, *J Proteome Res* 2003;2 : 351-60.

## Application of Non Linear Statistics (Principal Component Analysis) in Data Analysis of Differentiation Stem Cell to Astrocyte

Zali H<sup>1</sup>, Rezaei Tavirani M<sup>2\*</sup>, Sayed Khani Nahal A<sup>3</sup>, Moradi Sh<sup>4</sup>

(Received:                      Accepted: )

### Abstract

*Introduction:* The combination of univariate and multivariate statistics can identify significant biological changes in protein expression between experimental groups. One of the most common statistical methods that help to analyze two-dimensional gel electrophoresis is the principal component analysis (PCA). In this study, the differentiation of stem cells to astrocytes studied by proteomics and cell proteome of two groups were analyzed through PCA method by an appropriate statistical software.

*Materials & Methods:* Bone marrow aspirates was prepared from healthy donors and the mononuclear cells were isolated. Cells were incubated in 10% DMEM with low glucose, glutamine, streptomycin and penicillin in CO<sub>2</sub> 5% and moisture 98% at 37°C. For differentiation of these cells into astrocytes, cells exposed to retinoic acid, cAMP, PGF, PDGF, NGF. Stem cells and astrocyte cell proteome were extracted and separated by two dimensional electrophoresis. The gels were stained using silver staining and the scanned gels were analyzed by using of bioinformatics analysis software.

*Findings:* Bioinformatics and statistical analysis of two-dimensional gel electrophoresis technique revealed that 774 protein spots were detected in the two groups. Comparisons between groups suggested that the expression of new proteins and the silencing of certain proteins in the signaling pathway of cell differentiation have been accomplished. Clustering analysis, divided the expressed proteins into three main clusters indicating that there were clusters of proteins with similar expression could provide similar performance in terms of testing or biological activity in the biochemical pathways.

*Discussion & Conclusion:* PCA analysis confirmed the clustering results showed that the protein has been classified according to the test conditions. Finally, we concluded that the differentiation makes a significant change in the level of expression of the proteome and statistical analysis like clustering and PCA can be considered as a good indicator of the protein expression changes.

*Keywords:* Proteomics, Principal components analysis, Astrocytes

1. Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2. Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

3. Dept of Clinical Biochemistry, Faculty of Medicine, Ilam University of Medical Sciences, Ilam, Iran

4. Dept of Clinical Microbiology, Faculty of Medicine, Ilam University of Medical Sciences, Ilam, Iran

\*(corresponding author)