

# موضوع سمینار:

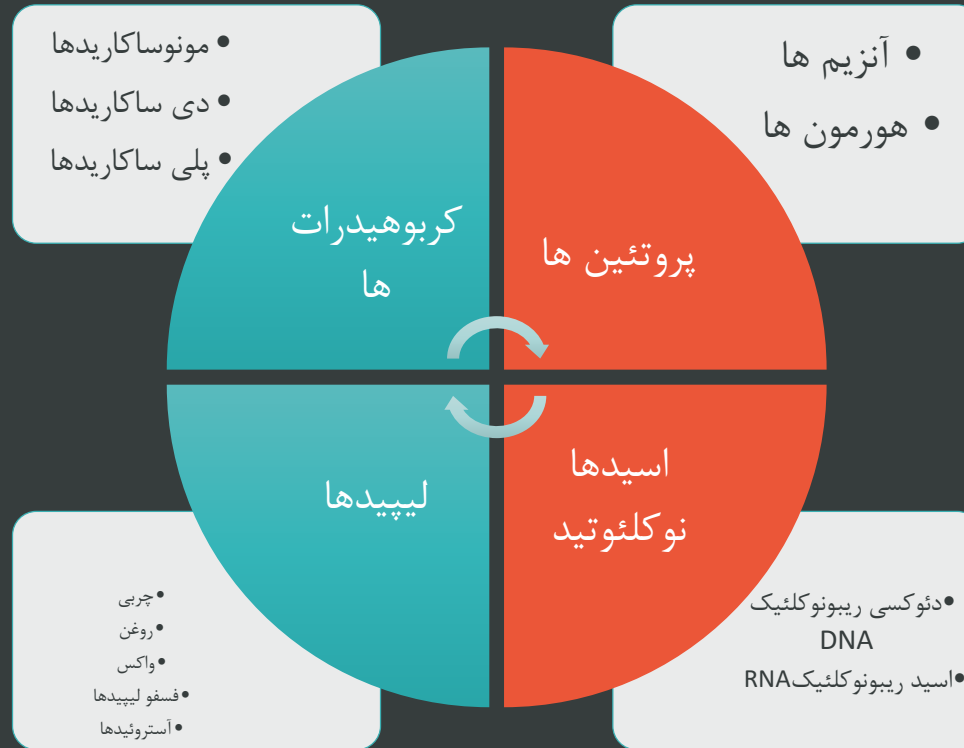
## کاربرد داده کاوی پروتئوم در دسته بندی سرطان و کشف زیست نشانگرها

دانشجو  
رسول نوروزی  
استاد راهنما  
دکتر امیر البدوی

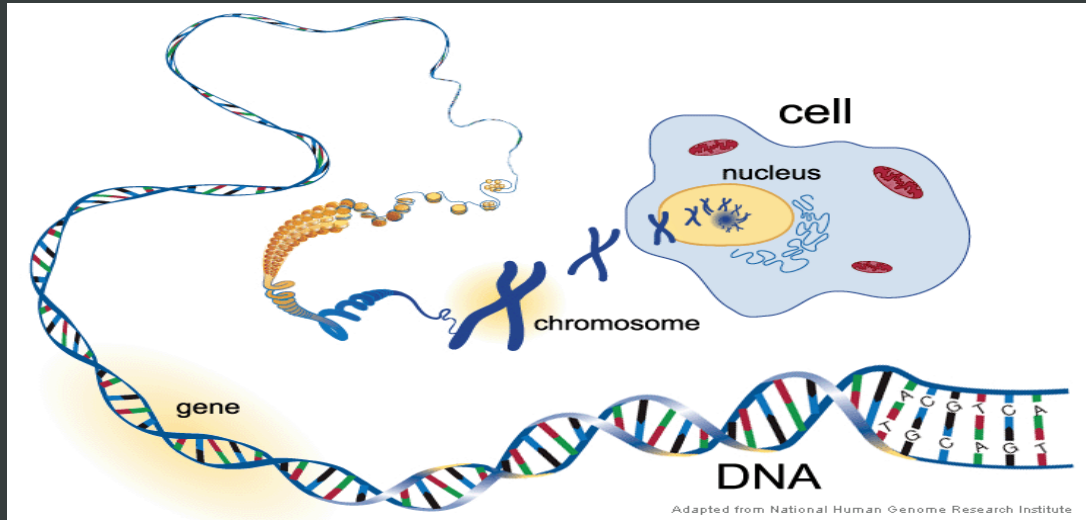
## اهداف سمینار

- ❖ شناخت انسان و بیماری سرطان از دیدگاه مولکولی
- ❖ آشنایی با علم پروتئومیکس: رویکردها و فن آوری ها
- ❖ بررسی و مرور ادبیات داده کاوی پروتئوم و کاربرد آن در دسته بندی سرطان و کشف زیست نشانگرها
- ❖ جمع بندی و نتیجه گیری

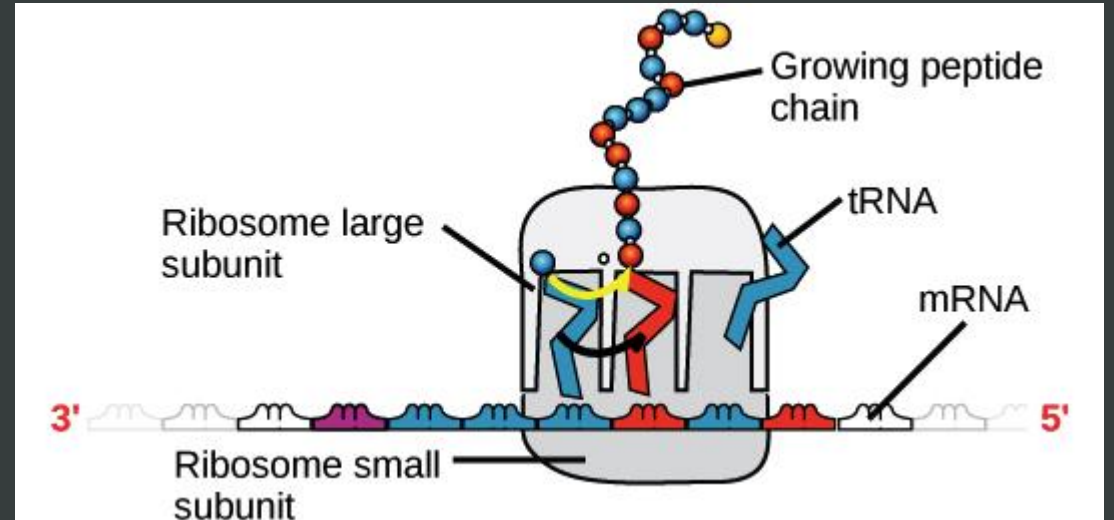
# انسان از منظر زیست شناسی



## از سلول تا دی ان ای و فرآیند ترجمه شدن



( وبسایت سازمان ملی ژنتیک انسانی آمریکا )



( وبسایت سازمان ملی ژنتیک انسانی آمریکا )

# سرطان و اساس مولکولی آن

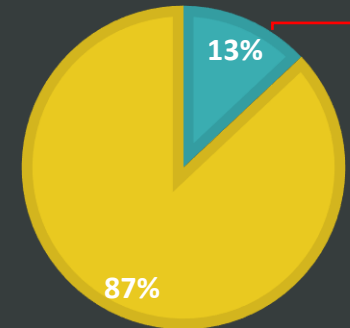
- سرطان یک بیماری ژنتیکی است که ۲۷۷ نوع بیماری را شامل می‌گردد. (پارسا, ۲۰۱۲)
- سرطان را با فعالیت نابجای پروتئین‌های کنترل‌کننده چرخه سلول که موجب ازدیاد در تقسیم سلولی می‌شود تعریف کرد. (Sosa et al., 2014)
- انواع سرطان: کارسینوما، سارکوما، لنفوما، سرطان خون
- از عوامل اساسی ایجاد سرطان می‌توانیم به: الکل، دخانیات، اضافه‌وزن و چاقی، مشکلات در سیستم ایمنی بدن، عفونت و ... اشاره کنیم. (Shukla et al., 2016)
- به‌طور کلی جهش ژنتیکی در دو نوع از تنظیم‌کننده‌های چرخه سلولی موجب توسعه سرطان می‌شود:
  1. تنظیم‌کننده‌های مثبت (آنکو ژن‌ها) oncogenes
  2. تنظیم‌کننده‌های منفی یا سرکوب‌کننده‌ها تومور (Otto and Sicinski, 2017, Joerger and Fersht, 2016)

# ضرورت پرداخت به سرطان

چرا سرطان؟

میزان مرگ میر

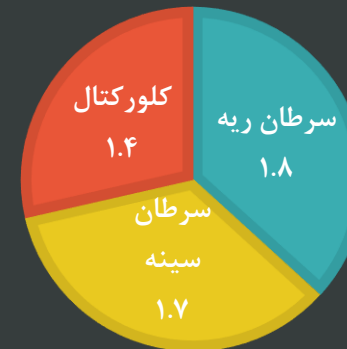
سرطان سایر



پیش‌بینی می‌شود که به این میزان تا ۷۰٪ تا دو دهه آینده اضافه شود

سرطان های شایع

سرطان ریه کلورکتال سرطان سینه



چرا از دیدگاه مولکولی؟

- با کمک از روش‌های مختلف مولکولی قادر هستیم که قدرت بیان ژن‌ها و پروتئین‌های معیوب را تعیین نماییم
- پیدا کردن زیست‌نشانه‌های جدیدی مانند پروتئین که شاخص یک نوع سرطان هستند در تشخیص زودرس و معالجه به‌موقع بیماری سرطان کمک‌های شایان توجهی را می‌نماید که این امر با به‌کارگیری علوم رایانه‌ای به‌خصوص داده‌کاوی ممکن می‌شود
- با تعیین شکل فضایی پروتئین‌های معیوب می‌توان داروهای ضد سرطان جدیدی را ساخت که بتواند سلول‌های در حال سرطانی شدن را مورد هدف قرار بدهند تا از تولید و رشد آن‌ها به سلول‌های سرطانی جلوگیری شود. (پارسا، ۲۰۱۲)

# کاربرد پروتئین در سرطان شناسی

چرا پروتئین ها؟

چگونه پروتئین ها منجر به دسته بندی سرطان و کشف زیست نشانگرها می شوند؟

1. سطح بیان mRNA معمولاً بیانگر میزان دقیق پروتئین-های فعال در سلول نیست
2. توالی ژن ها تغییرات پس از ترجمه اصلاحات پروتئین ها را که ممکن است برای عملکرد درست پروتئین ها ضروری باشد بیان نکند
3. مطالعات ژنتیک قادر به توصیف فرآیندهای پویای سلولی نیست. (Li et al., 2004)

- تغییرات بالینی ممکن است در الگوی پروتئینی یک ارگان یا بافت بازتاب یابد؛ بنابراین امکان پذیر است که پروتئین های موجود در یک نمونه خاص برای تشخیص بیماران سرطانی از غیر سرطانی استفاده شود. (Li et al., 2004)
- هدف اصلی و اساسی استخراج اطلاعاتی است که منجر به کشف الگوهایی برای دسته بندی سرطان و شناسایی زیست نشانگرهایی همچون پروتئین ها که به طور بالقوه ابزاری قدرتمند برای شناسایی، تشخیص و پیشگیری از بیماری ها به خصوص سرطان است.
- به ابزار تحلیلی سطح بالا همچون داده کاوی برای نیل به این اهداف نیاز است. (Thomas et al., 2006)

# پروتئومیکس و رویکردهای آن

پروتئوم به سری کامل پروتئین‌های بیان‌شده در یک لحظه خاص در یک سلول موردنظر اشاره دارد، با این حال امروزه سطح این تعریف، از سلول به بافت، اندام و ارگانیسم نیز تعمیم داده شده است و علم بررسی پروتئوم را پروتئومیکس گویند. (شیردل ET AL., 2013)

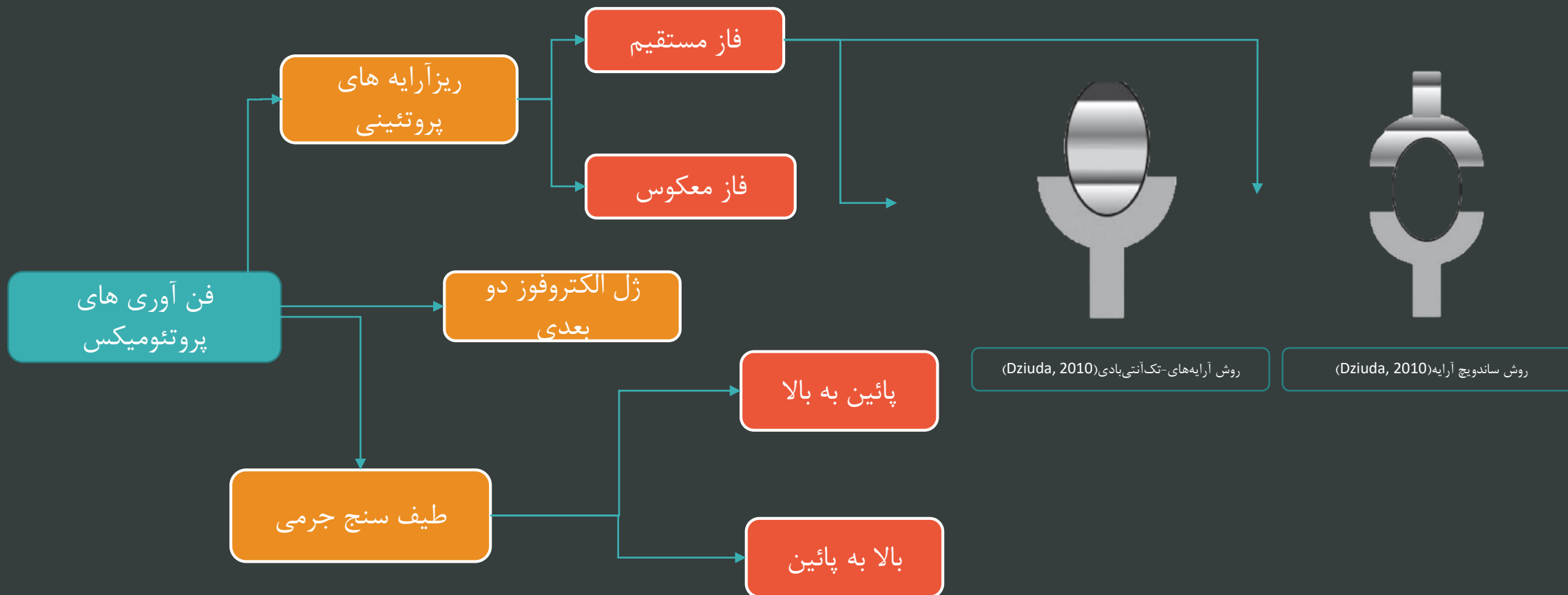




## نمونه های مناسب برای پروتئومیکس

معايب	مزایا	نمونه
مجموعه‌ای پیچیده از پروتئین‌ها است ممکن است موجب عدم دقت در بیان پروتئین‌های مقایسه‌ای بین دو گروه شود، مقادیر زیاد پروتئین‌ها در مایعات زیست پذیر (از قبیل آلبومین و هموگلوبین) ممکن است نقش پروتئین‌های با میزان کم اما دخیل در سرطان را کمرنگ کند.	منبع غنی از زیست‌نشانگرها، سهولت در دستیابی	مایعات زیست پذیر (سروم، پلاسما، اوره و ...)
روشی تهاجمی است، دسترسی به برخی از تومورها و بافت‌ها محدود است، بافت تومور ترکیبی ناهمگون از سلول‌های بدخیم همچنان که از بافت‌های همبند، بافت چربی و سلول‌های التهابی است.	زیست‌نشانگرهای کشف‌شده به‌طور واضح منشأ تومور می‌باشند، به‌هیچ‌عنوان نشانگرهای تومور در سلول‌های تومور پنهان نمی‌شوند؛	بافت
کمبود دسترسی به سل‌لاین‌های مناسب تومورها و بافت‌های مخاطی به‌عنوان گروه کنترل در مطالعات، نبود تعامل بین سلول‌های تومور با سلول‌های استرومال و یا ایمنی که شرایط محیط تومور در بدن را مشخص می‌کند	می‌شود به‌سادگی در آزمایشگاه کشت کرد، می‌تواند در مقدار بسیار زیاد در دسترس باشد ،	سل لاین

# فن آوری های پروتئومیکس

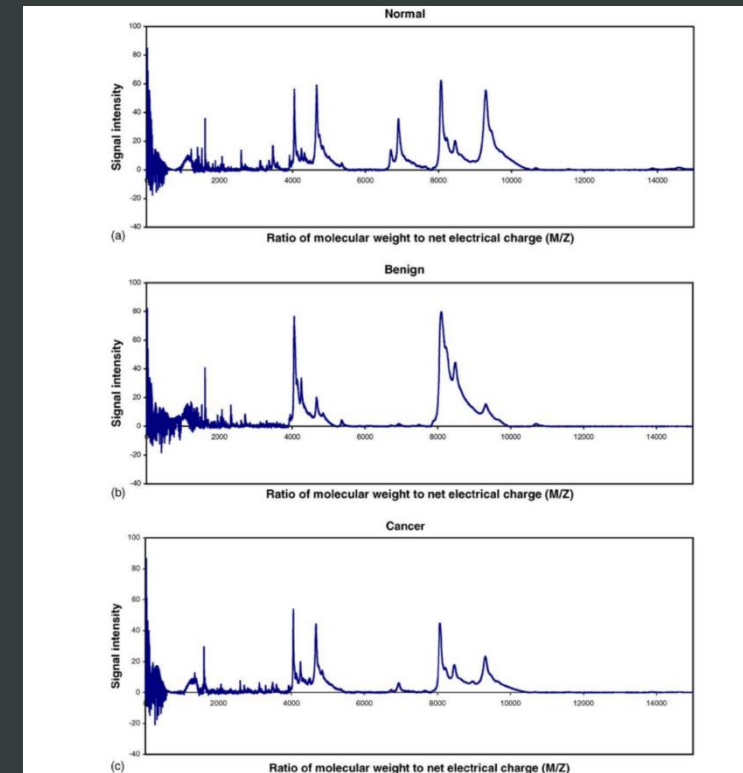


# نمونه ای از داده های حاصل از طیف سنج جرمی

ماتریس شدت M/Z و ماتریس بیان پروتئین (Dziuda, 2010)

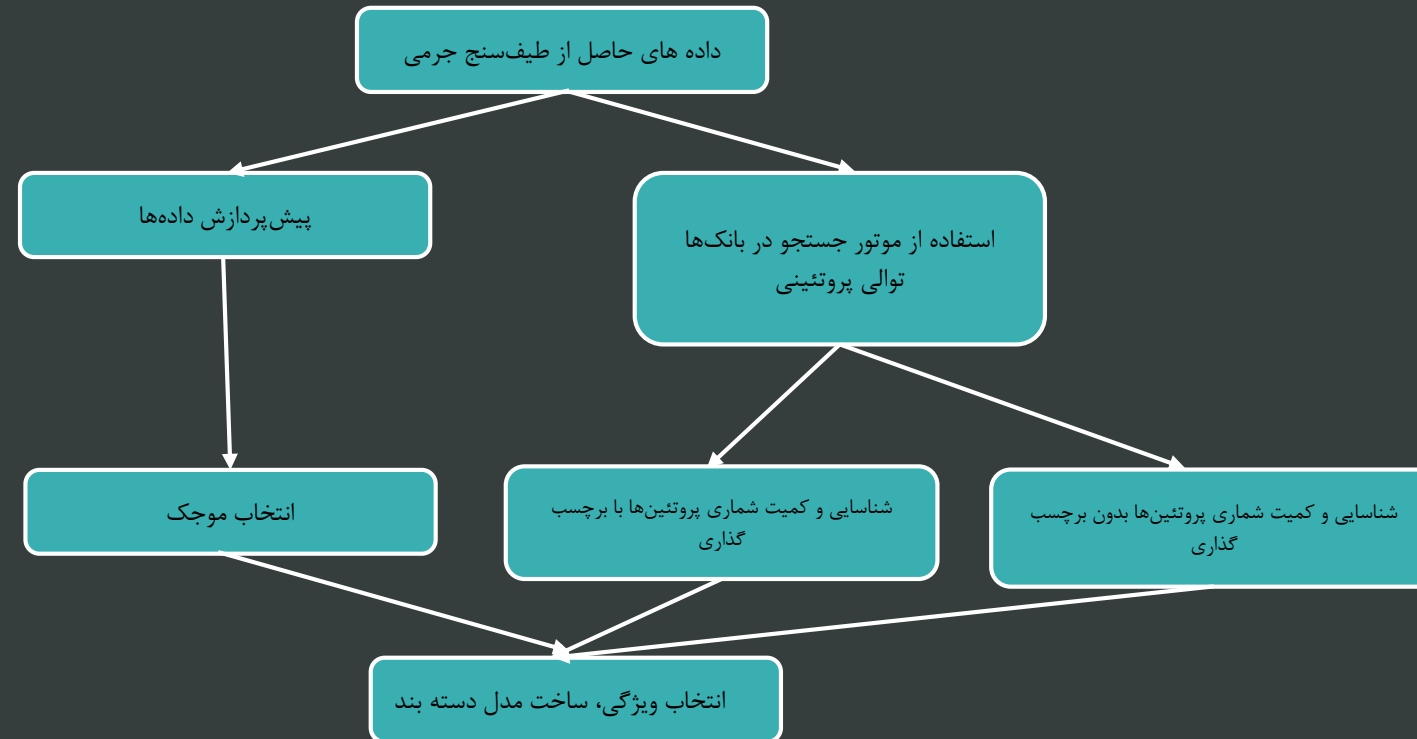
$m/z$	Class 1					***	Class 2			***	Class J	
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5		***	***	***		Sample N-1	Sample N
500.0223	25.9588	26.6526	24.4485	29.7677	26.1579						24.4203	25.9005
500.4307	28.6677	29.4781	26.9261	33.0075	29.8984						27.7164	29.0940
500.8573	33.8643	35.1272	31.9417	38.9422	34.3449						31.4630	33.0437
501.2751	38.0995	40.2232	35.2459	43.2378	36.7521						33.3736	35.1533
501.6931	39.6762	42.3087	35.7506	44.4585	38.0613						34.1921	36.1327
502.1113	39.7707	42.6244	35.3323	44.2172	38.0231						34.1357	36.2976
***												
***												
***												
15985.13	5.1260	4.9562	5.1565	4.4309	5.6946						4.5777	4.5680
15988.53	5.1059	4.9439	5.1494	4.4263	5.6690						4.5746	4.5636
15992.61	5.0613	4.9013	5.1072	4.4050	5.5761						4.5602	4.5554
15995.77	5.0525	4.8931	5.0944	4.4004	5.5560						4.5534	4.5514
15998.93	5.0297	4.8816	5.0752	4.3936	5.5237						4.5499	4.5437

	Class 1					***	Class 2			***	Class J	
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5		***	***	***		Sample N-1	Sample N
Variable 1	50.6083	45.8562	56.6637	66.0191	48.7441						20.3157	22.0191
Variable 2	81.3635	80.8688	86.9904	99.9598	80.8768						35.7969	48.9452
Variable 3	30.7451	21.9715	28.7226	25.6417	25.2696						16.3164	15.8974
Variable 4	25.2859	24.6415	32.1971	30.4539	23.9356						32.3733	65.2203
Variable 5	75.9739	99.6320	95.6099	73.0236	50.5500						86.1237	94.3348
Variable 6	31.3223	61.3896	40.8125	30.7833	25.0216						43.7728	47.9578
***												
***												
***												
Variable p-3	26.2785	36.7642	30.6385	30.8155	24.0845						40.8044	43.6642
Variable p-2	17.9427	19.6682	20.0322	20.0845	16.7964						34.6466	40.4123
Variable p-1	19.5918	20.5732	22.3328	22.2665	18.7692						19.7486	25.0679
Variable p	75.2334	90.1458	89.2247	75.6991	75.5829						87.5451	74.6747



این طیف نمایش دهنده سه دسته سرطان خوشخیم، سرطان بدخیم و سالم را نشان می دهد. (Li et al., 2004)

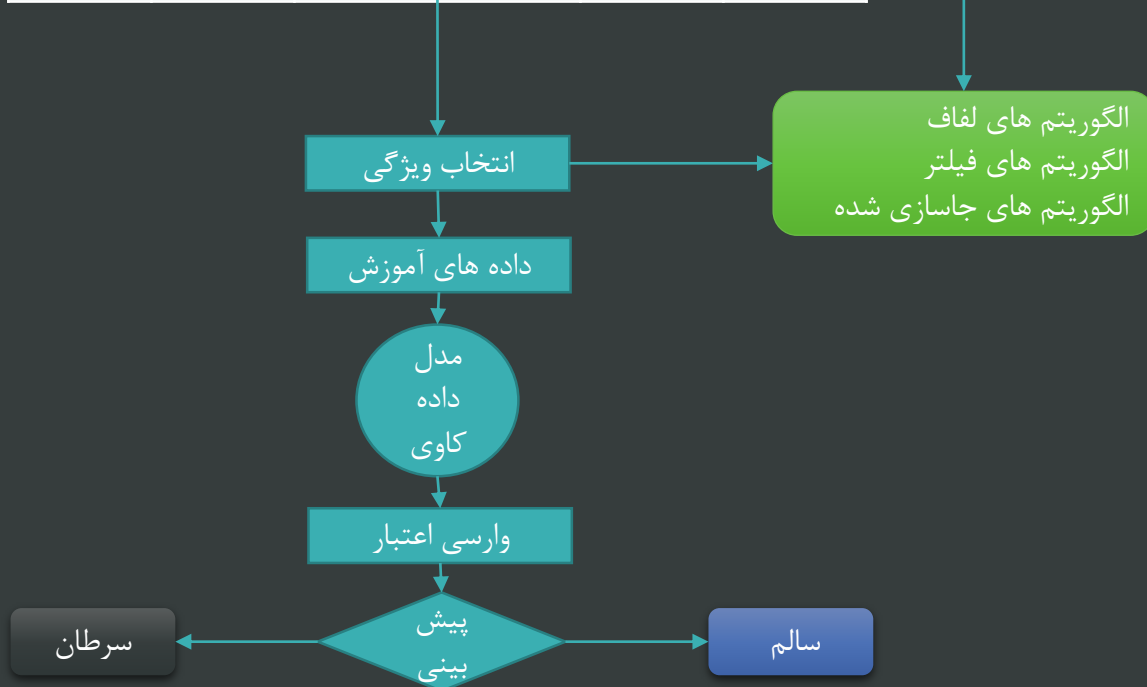
# آنالیز داده‌های طیف‌سنج جرمی



جریان آنالیز داده‌های حاصل از طیف سنج جرمی (Swan, Mobasheri et al. 2013)

# داده کاوی پروتئوم : پیش پردازش داده ها

مجموعه داده					
نمونه ها	ویژگی ۱	ویژگی ۲	...	ویژگی n	برچسب کلاس
بیمار ۱					سرطان
بیمار ۲					سالم
...					...
بیمار n					سالم



الگوریتم های لفاف  
الگوریتم های فیلتر  
الگوریتم های جاسازی شده

پاک سازی داده ها و کاهش ابعاد:

• پاک سازی داده:

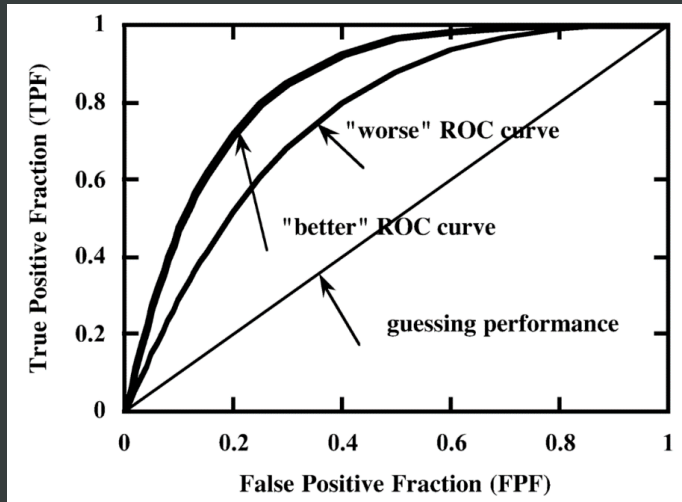
1. کاهش خط مبنا
2. نرمال سازی
3. همگام سازی موجک
4. کاهش نویز

# داده کاوی پروتئوم: مدل های دسته بندی

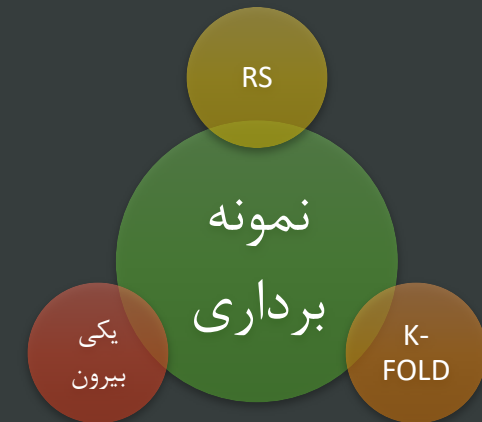
مقایسه مدل های دسته بندی (Swan et al., 2013)

مدل	مزایا و معایب	سرعت یادگیری	سهولت تفسیر پذیری
بیز ساده	مزایا: قابلیت ساده و سریع در بکارگیری، مناسب برای مجموعه داده های با داده های گم شده، معایب: فرض مستقل بودن ویژگی ها از هم	۱	۴
درخت تصمیم	خروجی این الگوریتم به سادگی قابل تفسیر است اما بستگی به نوع الگوریتم مورد استفاده و پیچیدگی درخت ساخته شده دارد، همچنان مناسب برای مجموعه داده ها با داده گم شده	۲	۱
جنگل تصادفی	روشی کارآمد برای مجموعه داده های بزرگ اگرچه در مقابل outlierها حساس نیست	۴	۳
بردارهای ماشین پشتیبان	استفاده از کرنل برای فراگیری توابع پیچیده، با این وجود بسیار کند بوده و چندین پارامتر نیز توسط کاربر باید تعریف شود	۵	۵
شبکه های عصبی مصنوعی	نتایج خروجی قابلیت خواندن ندارد و آموزش مدل ممکن است بسیار آهسته صورت گیرد.	۵	۵

# داده کاوی پروتئوم: نمونه برداری و ارزیابی عملکرد



نمونه‌ای از منحنی ROC (Thomas et al., 2006)



# مروری بر ادبیات کاربرد داده‌کاوی پروتئوم در دسته‌بندی و تشخیص زیست‌نشانگرهای سرطان

مرجع	دقت عملکرد دسته‌بند و زیست‌نشانگرهای شناسایی شده	روش‌های شناسایی، داده‌کاوی و ارزیابی	هدف پژوهش و مجموعه داده‌ها
(Sinues et al., 2015)	دقت دسته‌بند بالا ۹۰٪ و کشف هشت زیست‌نشانگر	استفاده از مدل SVM برای دسته‌بندی و انتخاب ویژگی	دسته‌بندی و کشف زیست‌نشانگرهای سرطان سینه با داده‌های بزاق دهان، ۱۴۱ نمونه بیمار و ۱۱ نمونه سالم، آنالیز داده‌ها ESI-MS
(Htike and Win, 2015)	دقت عملکرد دسته‌بند: ۷۴.۰۳۳۱٪	روش ترکیبی لوجستیک درخت برای مدل‌سازی، الگوریتم RELIEF برای انتخاب ویژگی، استفاده از روش TOP-HAT برای کاهش خط‌مبنا، استفاده از رویکرد یکی بیرون برای اعتبار‌سنجی	هدف دسته‌بندی سرطان لوزالمعده ۸۰ نمونه سرطان لوزالمعده در مقابل ۱۰۱ نمونه سالم،
Guan et al., 2009)	بهترین دقت ۸۳٪ با استفاده از SVM و LOO-CV، رسیدن به دقت ۹۷.۲٪ با ترکیب مدل SVM و انتخاب ویژگی بر مبنای SVM، شناسایی ۳۸ زیست‌نشانگر با چهار رویکرد متفاوت	شناسایی با نرم‌افزار mzMine (v0.60)، SVMs با روش‌های مرتبط انتخاب ویژگی، LOO-CV, 12-FOLD CV, 52-20-split validation	دسته‌بندی و کشف زیست‌نشانگرهای سرطان تخمدان، ۳۷ بیمار بیمارستان مبتلا به سرطان تخمدان پاپیلری سروز و ۳۵ نمونه کنترلی
(Vlahou et al., 2003)	۸۵٪ دقت دسته‌بندی	دسته‌بندی بر اساس درخت-رگرسیون (CART)، اعتبار‌سنجی 10-FOLD، برای دسته‌بندی مویک استفاده از نرم‌افزار Ciphergen Systems، نرم‌افزار الگوهای زیست‌نشانگر (BPS)	دسته‌بندی سرطان بدخیم، خوشخیم و سالم در سرطان تخمدان، ۴۴ سرطان بدخیم، ۶۱ سرطان خوشخیم و ۳۴ نمونه سالم



# مروری بر ادبیات کاربرد داده‌کاوی پروتئوم در دسته‌بندی و تشخیص زیست‌نشانگرهای سرطان

(Le et al., 2005)	۸۵٪ دقت و پیدا کردن چند زیست نشانگر	استفاده از نرم افزار mascot و استفاده از C.SVM با استفاده از اعتبار سنجی یکی-بیرون	دسته بندی نمونه های پروستات از ۱۹ بیمار با متاستاز استخوان و ۱۹ بیمار فاقد آن
(Adam et al., 2002)	۹۰٪ از داده های تست به درستی دسته بندی شدند.	داده تست شامل: ۱۵ کنترل، ۱۵ خوشخیم و ۳۰ پروستات. شناسایی موجک با نرم افزار CIPHERGEN SELDI software و همگام سازی موجک تا با الگوریتم peakminer، استفاده از درخت تصمیم.	دسته بندی سرطان پروستات و کنترل، ۹۷ نمونه کنترلی، ۹۲ نمونه سرطان خوشخیم، ۱۹۷ سرطان پروستات
(Oh et al., 2009)	دقت ۸۷.۹٪، ۲۶ موجک به عنوان زیست نشانگرهای محتمل شناسایی شد.	پیش پردازش: گروه بندی، تصحیح خط مبنا و نرمال سازی با استفاده از نرم افزار TOFWorks	شناسایی زیست نشانگر برای سرطان پروستات، ۱۷۹ سرطان ادرنوکازسینوما و ۷۴ سرطان خوشخیم
(Ralhan et al., 2008)	سه تا از زیست نشانگرها شناخته شد.	استفاده از نرم افزارهای ProteinPilot و iTRQ استفاده از مدل بیز ساده و اعتبار سنجی 3-FOLD	شناسایی زیست نشانگرهای سرطان سر و گردن، پنج مجموعه از چهار نمونه به همراه نمونه کنترل برای هر مجموعه

# مروری بر ادبیات کاربرد داده‌کاوی پروتئوم در دسته‌بندی و تشخیص زیست‌نشانه‌های سرطان

<p>توسعه پنل برای شناسایی زیست نشانگرهای سرطان سینه، ۴۰ نمونه پلاسما خون از سرطان سینه و ۴۰ نمونه از سالم Panel (منظور تعدادی از پروتئین‌ها که با هم کار می‌کنند و ظاهر می‌شوند)</p>	<p>شناسایی پروتئین تا به صورت بدون برچسب گذاری با کمک نرم افزار Eli Lilly، استفاده از مدل شبکه های عصبی مصنوعی</p>	<p>۸۵٪ دقت در داده های تست، دوتا از بهترین پنج پنل پروتئین شناخته شد که شامل هفت پروتئین شد.</p>	<p>(Zhang and Chen, 2009)</p>
<p>کشف زیست نشانگرها، ۶۵ نمونه از بیماران مبتلا به سرطان سینه، سپس نمونه گیری مجدد از آنها بعد از مصرف چهار هفته ای docetaxel 75 mg/m2</p>	<p>استفاده از نرم افزار SpecAlign برای کاهش خط مبنا، همگام سازی موجک، استفاده از مدل AdBoost برای دسته بندی و 5-FOLD CV برای</p>	<p>۶ زیست نشانگر شناخته شد</p>	<p>(Ushijima et al., 2007)</p>
<p>شناسایی زیست نشانگرها، ۱۳۲ بیمار با سرطان لنفوم B-Cell و ۷۵ نمونه کنترلی، داده های آنالیز شده SELDI-TOF-MS از سرم خون</p>	<p>استفاده از درخت تصمیم برای مدل سازی و استفاده از نرم افزار CIPHERGEN برای پیش پردازش داده ها</p>	<p>نه موجک به عنوان زیست نشانگر بالقوه انتخاب در نظر گرفته شد، چهار موجک برای پیش بینی پاسخ بیماران به درمان های استاندارد، حساسیت ۹۴٪، ویژگی ۹۴٪ در در ۸۵ نمونه از مجموعه تست، ۹۴٪ حساسیت و ۹۲٪ ویژگی با ۶۶ نمونه تست.</p>	<p>(Zhang et al., 2007)</p>

## مروری بر ادبیات کاربرد داده‌کاوی پروتئوم در دسته‌بندی و تشخیص زیست‌نشانگرهای سرطان

(Cohen et al., 2011)	۹ نشانگر یافت شده با ۸۹٪ دقت در عملکرد	استفاده از مدل دسته‌بندی مبتنی بر SVM و برای واریسی اعتبار از الگوریتم 10-FOLD استفاده شده است.	کشف زیست‌نشانگرهای سرطان معده، ۷۹ نمونه از بیماران مبتلا به سرطان معده و ۳۳ نمونه از افراد فاقد سرطان که ۱۰ نفر از آنها دارای التهاب معده بودند.
(Washam et al., 2013)	شناسایی ۱۳ زیست‌نشانگر، قدرت دسته‌بندی با حساسیت ۹۱٪ و ویژگی ۹۳٪	استفاده از جنگل تصادفی برای مدل دسته‌بندی، ساخت درخت ۱۰۰۰ مرتبه تکرار شده و هر بار در هر گروه سه ویژگی مورد آزمایش قرار می‌گرفت	شناسایی زیست‌نشانگرها که موجب متاستاز استخوانی در سرطان سینه می‌شود، نمونه از ۱۱۱ زن مبتلا به سرطان سینه به دو گروه تقسیم شده که گروه اول شامل ۴۱ نفر متاستاز استخوان و ۳۶ نفر دیگر فاقد آن، گروه دیگر شامل ۱۷ نفر متاستاز استخوان و ۱۷ نفر دیگر فاقد آن

# داده ها در دسترس برای پژوهش

❖ اطلس پروتئینی انسان (<http://www.proteinatlas.org>)

❖ پایگاه داده مرجع پروتئین انسانی (<http://www.hprd.org>)

❖ مرکز ملی اطلاعات زیست فن آوری (<https://www.ncbi.nlm.nih.gov>)

❖ بانک داده پروتئین (<http://www.rcsb.org/pdb/home/home.do>)

❖ بانک داده شناسایی پروتئین ها (<https://www.ebi.ac.uk>)

❖ Proteopedia (<http://proteopedia.org>)

❖ UniProt (<http://www.uniprot.org>)

	A	B	C	D	E	F	G	H	I	J
1	Gene	Protein	Protein description	Peptide	charge	A1	A2	A9	B2	I
2	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	GWMNDPGLWYDEK	2	63976296	24769764	40046584	5498	
3	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	GWWM(+16)NDPGLW	2	6197049	2523306	12923109	1041	
4	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	VFWYEPESQK	2	8723820	8625410	7146495	496	
5	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	WIMTAAK	2	1086354	1498491	6836681	602	
6	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	WIM(+16)TAAK	2	177862	12706056	851060	777	
7	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	IEIYSSDDLK	2	190592160	118888592	136130736	1684	
8	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	IEIYSSDDLKSWK	3	742601	1987192	2445724	4771	
9	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	SSMSLVRK	2	210993	484660	3281242	54	
10	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	KFSLNTEYQANPETELIN	3	8995346	22224170	23039502	5631	
11	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	FSLNTEYQANPETELINL	3	41705792	41687748	46593780	4379	
12	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	AEPILNISAGPWRSR	2	5289871	6530384	16779556	239	
13	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	GLEDPPEEYLR	2	522095232	784777280	649137216	56294	
14	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	MGFEVSASSFFLDR	2	4875871	5925656	7153578	575	
15	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	MGFEVSASSFFLDR	3	75067	1966584	2582508	198	
16	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	M(+16)GFEVSASSFFLD	2	1367760	1	1		
17	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	M(+16)GFEVSASSFFLD	3	3904164	9913573	19360182	628	
18	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	MGFEVSASSFFLDKRG	3	1511042	5894734	4381281	235	
19	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	ENPYFTNR	2	247100000	292735456	193823904	7914	
20	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	MSVNNQPFK	2	4375857	81305216	2918126	2939	
21	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	M(+16)SVNNQPFK	2	4616198	5216906	1125524	1895	
22	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	MSVNNQPFKSENDLSY	3	3358167	3965633	3892572	429	
23	SUC2	P00724	Invertase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=SUC2 PE=1 SV=1	SENDLSYK	2	153155616	173682416	130623936	6286	
24	<b>SUC2 Total</b>					1274133118	1607309228	1311073297	92276	
25	AKAP13	Q12802	A-kinase anchor protein 13 OS=Homo sapiens GN=AKAP13 PE=1 SV=2	STPSLPC(+57)M(+16)I	3	383171360	396120864	449970880	10815	
26	AKAP13	Q12802	A-kinase anchor protein 13 OS=Homo sapiens GN=AKAP13 PE=1 SV=2	SVSIQNITGVGNDENM	3	25716	45718	14628	21	
27	<b>AKAP13 Total</b>					383197076	396166582	449985508	10818	
28	ECD	O95905	Protein ecdysoneless homolog OS=Homo sapiens GN=ECD PE=1 SV=1	EEKEQNYDLTEVSSEM	4	29705026	8415698	4520073	747	
29	<b>ECD Total</b>					29705026	8415698	4520073	747	

نمونه‌ای از داده‌های آزمایشگاه پروتئومیکس دانشگاه کالیفرنیا جنوبی

## نتیجه گیری و جمع بندی

- پروتئین ها هم عامل بیماری و هم عامل درمانی می توانند باشند.
- شناسایی، توصیف و کمیت شماری پروتئین ها برای مقایسه و دسته بندی دو دسته بیمار و دسته کنترل جهت دسته بندی سرطان ها و کشف زیست نشانگرها استفاده می کند.
- با وجود اینکه بافت ها اطلاعات ارزشمندی را در خود دارند اما به دلیل مشکلات نمونه برداری و تهاجمی بودن آن ها توصیه نمی شود ولی در مقابل مایعات زیست پذیر هم به دلیل گردش در بدن اطلاعات متنوعی در خود دارند و به خصوص نمونه های ادرار به دلیل غیرتهاجمی بودن نمونه های مناسب برای پروتئومیکس است
- طیف سنج جرمی به دلیل سهولت و توان بالا در آنالیز حجم عظیمی از پروتئین ها نسبت به روش ها دیگر بیشتر مورد توجه محققین قرار گرفته است و نکته مهم دیگر اینکه غیر جانب دارانه است.
- لزوم استفاده از انتخاب ویژگی برای کشف زیست نشانگرها و کاهش ابعاد داده و به تبع آن افزایش دقت.
- به دلیل دشواری در نمونه گیری داده های پزشکی تعداد سطرهای داده ها بسیار کم است و نیاز به روش های پیچیده نمونه برداری جهت داده تست و داده آزمایش برای ساخت و ارزیابی مدل داریم
- پرکاربردترین مدل دسته بند الگوریتم بردار ماشین های پشتیبان و درخت تصمیم است به دلیل قدرت و دقت بالای دسته بندی

## چالش ها و فرصت ها

- می توان از روش دسته بندی مبتنی بر قاعده که در بررسی و تحلیل ریزآرایه ها در بیوانفورماتیک استفاده می شود بهره برد.
- همچنان بحث پیش پردازش و کاهش ابعاد در پروتئومیکس جدی است. ضرورت پرداختن به الگوریتم های نوین و مناسب حس می شود.
- همچنین جای خالی ارتباط سایر بیماری ها به طور مثال چاقی و یا دیابت با یک سرطان خاص یا مجموعه ای از سرطان ها با رویکردهای پروتئومیکس به شدت احساس می شود.

با سپاس از توجه شما