

رویکردهای زیست‌شناسی محاسباتی و بیوانفورماتیک برای شناسایی ژن‌های کلیدی در سندرم تخمدان پلی‌کیستیک: یک مرور سیستماتیک

سیده نفیسه طباطبایی^۱، زرین مینوچهر^{۲*}

تاریخ دریافت ۱۴۰۱/۰۸/۲۴ تاریخ پذیرش ۱۴۰۲/۰۱/۱۹

چکیده

پیش‌زمینه و هدف: سندرم تخمدان پلی‌کیستیک (PCOS) شایع‌ترین بیماری غدد درون‌ریز زنانه است و اغلب باعث ناباروری زنان در سنین باروری می‌شود. این سندرم شامل ژن و پروتئین‌های مختلف، مسیرهای متعدد و فرآیندهای پیچیده ترشح هورمون است. از این رو، پاتوژنز PCOS را نمی‌توان با یک عامل توضیح داد. استفاده از زیست‌شناسی محاسباتی و اومیکس‌ها شامل ژنومیکس، ترانسکریپتومیکس، پروتئومیکس و متابولومیکس می‌تواند روش‌های سریع و مؤثرتری برای مطالعه پاتوژنز بیماری‌های پیچیده مثل PCOS ارائه دهند.

روش کار: در این مطالعه جهت یافتن مقالات مرتبط پایگاه‌های اطلاعاتی PubMed، Google Scholar و Science direct بدون محدودیت زمانی و با استفاده از کلیدواژه‌های Polycystic Ovary Syndrome، Computational Biology، Protein-Protein Interaction، Network Biology و Pathways analysis در بازه زمانی ۳ ساله مورد جستجو قرار گرفتند.

یافته‌ها: تاکنون پایگاه‌های داده مختلفی برای ذخیره اطلاعات، میانکنش‌های پروتئینی، شبکه‌ها و مسیرهای زیستی مربوط به انسان طراحی شده و در اختیار عموم قرار گرفته است. سه پایگاه داده PCOSBase، PCOSKB و PCOSDB مخصوص سندرم تخمدان پلی‌کیستیک ابداع شده است.

بحث و نتیجه‌گیری: از آنجا که مکانیسم‌های مولکولی سندرم تخمدان پلی‌کیستیک هنوز به‌طور کامل شناخته نشده است برای درک این سندرم نیاز است علاوه بر نتایج آزمایشگاهی با استفاده از پلتفرم‌های اومیکس، زیست‌شناسی محاسباتی و ابزارهای بیوانفورماتیک میانکنش بین پروتئین‌ها و مسیرهای درگیر در آن شناسایی شود.

کلیدواژه‌ها: زیست‌شناسی محاسباتی، شبکه‌های زیستی، آنالیز مسیر، سندرم تخمدان پلی‌کیستیک، آنالیز میانکنش‌های پروتئینی-پروتئینی

مجله مطالعات علوم پزشکی، دوره سی و سوم، شماره یازدهم، ص ۷۸۵-۷۸۸، بهمن ۱۴۰۱

آدرس مکاتبه: انوبان تهران - کرج، کیلومتر ۱۵، شهرک علم و فناوری پژوهش، بلوار پژوهش، پژوهشگاه ملی مهندسی ژنتیک و زیست‌فناوری، پژوهشگاه زیست‌فناوری صنعت و محیط‌زیست، گروه زیست‌فناوری سامانه‌ای، تلفن: ۰۲۱-۴۴۷۸۷۳۶۲

Email: dminuchehr@gmail.com

مقدمه

متابولیک با سندرم تخمدان پلی‌کیستیک مرتبط هستند که این سندرم علاوه بر اختلال در عملکرد تولیدمثل و غدد درون‌ریز، با مقاومت ذاتی به انسولین^۴ (IR) نیز مشخص می‌شود و منجر به ایجاد سندرم متابولیک^۵ (MetS) و در نتیجه اختلال در متابولیسم کربوهیدرات‌ها و افزایش چربی در ناحیه شکمی می‌شود. همچنین

یکی از دلایل ناباروری در ۲۵ درصد از زوجها اختلالات تخمک‌گذاری است و سندرم تخمدان پلی‌کیستیک^۳ (PCOS) علت اصلی ناباروری و فقدان تخمک‌گذاری است که تقریباً ۷۰ درصد از موارد را شامل می‌شود. مجموعه‌ای از غدد درون‌ریز و ویژگی‌های

^۱ دانشجوی دکتری ژنتیک مولکولی، پژوهشگاه زیست‌فناوری صنعت و محیط‌زیست، پژوهشگاه ملی مهندسی ژنتیک و زیست‌فناوری، تهران، ایران

^۲ دانشیار گروه زیست‌فناوری سامانه‌ای، پژوهشگاه زیست‌فناوری صنعت و محیط‌زیست، پژوهشگاه ملی مهندسی ژنتیک و زیست‌فناوری، تهران، ایران (نویسنده

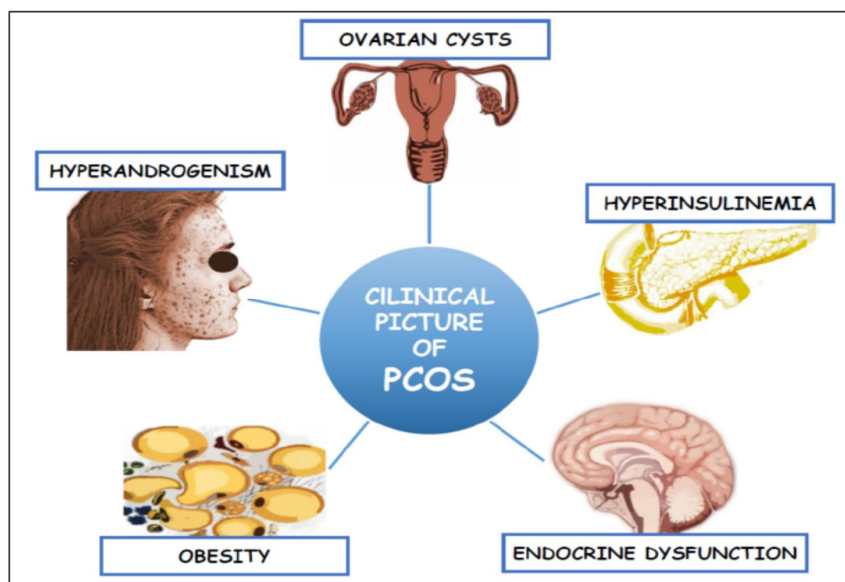
مسئول)

^۳ Polycystic ovary syndrome

^۴ Intrinsic Insulin Resistance

^۵ Metabolic Syndrome

در ایجاد دیس لیپیدمی^۱ (هرگونه اختلال در سطح سرمی چربی‌ها شامل افزایش یا کاهش غیرطبیعی)، فشارخون شریانی^۲ (AH) و بیماری کبد چرب غیرالکلی^۳ (NAFLD) دخالت دارد. این اختلالات به‌نوبه خود منجر به ایجاد بیماری قلبی عروقی^۴ (CVD) می‌شود که هنوز عامل اصلی مرگ‌ومیر در بین زنان است (شکل ۱). زنان مبتلا به PCOS ممکن است خطر سقط‌جنین و عوارض بارداری مانند دیابت بارداری را نیز تجربه کنند (۱، ۲).



شکل (۱): تصویر بالینی سندرم تخمدان پلی‌کیستیک (۳)

بیشتر دانشمندان بر تغییر سبک زندگی و کاهش وزن افراد در معرض خطر، اتفاق نظر دارند و آن را اولین راه پیشگیری و درمان مبتلا به PCOS در نوجوانان می‌دانند (۵). تحقیقات نشان داده است کیفیت زندگی زنان مبتلا به PCOS در مقایسه با افراد سالم و حتی در مقایسه با مبتلایان به دیگر بیماری‌های زنان، پایین‌تر است (۶).

۱- اومیکس و سیستم‌های محاسباتی

بیماری‌های چندعاملی همچون PCOS که در اثر جهش در چندین ژن، عوامل محیطی و سبک زندگی ایجاد می‌شوند، اتیولوژی پیچیده‌ای دارند که برای غلبه بر این پیچیدگی‌ها تاکنون تلاش‌های زیادی انجام گرفته است. مطالعه در سطح مولکولی یکی از آن‌هاست. ظهور فناوری‌های زیستی با تولید مجموعه داده‌های بزرگ اومیکس (ژنومیکس، ترانسکریپتومیکس، پروتئومیکس و متابولومیکس) پیشرفت‌های زیادی در رمزگشایی پاتوبیولوژی بیماری‌ها داشته است. مجموعه داده‌های اومیکس طیف گسترده‌ای از پدیده‌های بیماری از جمله جهش، بیان ژن، بیان پروتئین، پروفایل

PCOS یکی از شایع‌ترین اختلالات غدد درون‌ریز در زنان در سنین باروری است که در ۳ تا ۱۵ درصد از جمعیت زنان در سراسر جهان وجود دارد. علائم اصلی PCOS شامل اختلالات تخمک‌گذاری، افزایش هورمون آندروژن^۵ (HA) و مقاومت به انسولین^۶ (IR) است. اگرچه مطالعات زیادی در مورد PCOS انجام شده است، با این حال، مکانیسم اساسی آن هنوز مشخص نیست. در حال حاضر، PCOS به‌عنوان یک بیماری پیچیده چندژنی شناخته می‌شود که تحت تأثیر عوامل ژنتیکی بسیاری (مانند پلی‌مورفیسم‌های ژنی، اپی‌ژنتیک، نژاد/قومیت) قرار دارد و تظاهرات بالینی آن به دلیل تعامل با بسیاری از عوامل محیطی (به‌عنوان مثال، اجتماعی-اقتصادی، جغرافیایی، سم‌شناسی، سبک زندگی و رژیم غذایی) متفاوت است (۴).

ارتباط وزن زیاد و چاقی با بروز این بیماری در بسیاری از مطالعات به اثبات رسیده و نشان داده شده که ۸۰ درصد از مبتلایان به این سندرم از چاقی و اضافه وزن رنج می‌برند. به همین علت

^۵ Hyperandrogenemia

^۶ Insulin Resistance

^۱ Dyslipidemia

^۲ Arterial Hypertension

^۳ Non-alcoholic fatty liver disease

^۴ Cardiovascular disease

متابولیت‌ها و تعاملات ژنتیکی و فیزیکی بین مولکول‌های زیستی را در بر می‌گیرند که در آن هر مجموعه داده اطلاعات متمایزی را برای درک بیماری‌ها ارائه می‌دهد. بررسی بیماری‌های چندعاملی توسط تنها یک سطح مجموعه داده‌های اومیکس ناکافی است زیرا این بیماری‌ها در سطوح متعدد تنظیم می‌شوند. این بیماری‌ها را می‌توان با ادغام آنالیز داده‌های اومیکس در سطوح مختلف مشخص کرد (۷). آنالیز سطوح مختلف اومیکس چالش جدیدی را برای توسعه روش‌ها یا پایپ لاین آن‌ها، آمار و الگوریتم‌ها و ابزارهای یکپارچه سازی ایجاد کرده است که در کمک به آنالیز سیستم‌های محاسباتی^۱ به شدت مورد نیاز است. ادغام آنالیز داده‌های اومیکس در سطوح مختلف، بهترین راه قابل‌اجرا برای به دست آوردن دیدگاه‌های سیستماتیک و جامع از بیماری‌ها، دستیابی به درک بهتر از مکانیسم‌های بیماری و یافتن درمان‌های پزشکی شخصی است. با کمک آنالیز سیستم‌های محاسباتی، تحقیقات در زمینه بیولوژی و زیست پزشکی در طول چند دهه گذشته فواید فوق‌العاده‌ای داشته است (۸).

آنالیز سیستم‌های محاسباتی، دیدگاه‌های میان‌رشته‌ای را با ریاضیات، الگوریتم‌ها، مدل‌سازی و شبیه‌سازی، پایگاه داده و یا ترسیم شبکه‌ها با استفاده از تکنیک‌های محاسباتی و با بررسی پدیده‌ها و شرایط زیستی خاص، در منظر یک سیستم به هم متصل می‌کند. تاکنون، مطالعات زیادی در ارتباط با داده‌های اومیکس و شبکه‌های زیستی^۲ که یکی از تکنیک‌های اصلی آنالیز سیستم‌های محاسباتی برای رسیدن به یک نمای کلی از پاتولوژی بیماری‌های انسانی است، انجام شده است. شبکه‌های زیستی می‌توانند به‌طور سیستماتیک تمام مولکول‌های تولید شده از مطالعات اومیکس را که مرتبط با بیماری شناسایی شده‌اند، به هم متصل کند. ایجاد و توسعه پایگاه داده یکی دیگر از آنالیزهای سیستم‌های محاسباتی است که اطلاعات کلی را در مورد بیماری‌ها ارائه می‌دهد (۹).

۲- شبکه‌های زیستی

پروتئین‌ها به‌عنوان عوامل اصلی عملکرد بیولوژیکی، فنوتیپ همه موجودات را تعیین می‌کنند. در پیدایش زیست‌شناسی مولکولی، فرض بر این است که پروتئین‌ها به‌طور طبیعی در اشکال ایزوله عملکردی ندارند، ولی در تعامل با پروتئین‌های دیگر و همچنین با سایر ماکرومولکول (DNA، RNA و متابولیت‌ها) واسطه مسیرهای متابولیک، مسیرهای سیگنالینگ و تنظیمی، فرآیندهای سلولی و سیستم‌های ارگانیسمی می‌شوند. بسیاری از خصوصیات بیولوژیکی یا فنوتیپ‌ها از برهمکنش‌های پیچیده بین

ترکیبات بی شمار سلول ناشی می‌شوند. هرگونه اختلال در میانکشی این مولکول‌ها می‌تواند رفتار طبیعی سلول‌ها را مختل کرده و منجر به بیماری یا مشکلات بالینی شود. بنابراین، مطالعه‌ی شبکه‌های زیستی در بیماری‌های چندعاملی ضروری است زیرا می‌توان از آن برای شناسایی رویدادهای بیولوژیکی مختل‌کننده این شبکه‌ها که نقش اساسی در درک عملکرد بیولوژیکی سلول دارند، استفاده کرد (۱۰). در شبکه‌های زیستی، دو نوع آنالیز زیستی ۱- آنالیز میانکشی‌های پروتئینی-پروتئینی^۳، آنالیز مسیرها، وجود دارد که اغلب برای درک پاتوبیولوژی بیماری‌ها استفاده می‌شود.

۲-۱ آنالیز میانکشی‌های پروتئینی-پروتئینی

آنالیز میانکشی‌های پروتئینی-پروتئینی روشی کاربردی است که می‌تواند برای درک بهتر پاتوبیولوژی بیماری‌ها به کار رود و به‌نوبه خود می‌تواند روش‌هایی را برای تشخیص، پیشگیری و درمان بیماری‌ها ارزیابی کند. آنالیز ویژگی‌های شبکه می‌تواند فرصتی برای تفسیر رفتارهای بیولوژیکی طبیعی و تغییر یافته که منجر به بیماری‌ها می‌شود را فراهم کند (۱۱). روش‌های مختلفی برای آنالیز شبکه‌های زیستی وجود دارد که در اینجا سه روش را توضیح می‌دهیم.

(۱) ترکیب میانکشی‌های پروتئین-پروتئین شبکه‌ای را تشکیل می‌دهد که از دو جزء اصلی تشکیل شده است، (۱) گره‌ها^۴ یا دایره‌ها که نشان‌دهنده پروتئین است و (۲) خط‌ها^۵ که ارتباط بین گوی‌ها را نشان می‌دهد (شکل ۲). شبکه میانکشی‌های پروتئین-پروتئین^۵ (PPIN) که حاوی اطلاعاتی در مورد فرآیندهای بیولوژیکی و عملکردهای مولکولی سلول‌ها است، به‌طور گسترده برای توصیف مکانیسم بالقوه بیماری‌ها پیچیده مرتبط با ژن‌ها استفاده می‌شود. بسیاری از بیماری‌ها توسط یک ژن ایجاد نمی‌شوند، بلکه توسط گروهی از مولکول‌ها یا شبکه‌های مرتبط ایجاد می‌شوند. بنابراین، ایجاد و آنالیز شبکه‌های نشانگر زیستی برای شناسایی، پیش‌بینی و درمان بیماری‌ها بسیار مهم است (۱۲). توزیع درجه که کسری از تعداد میانکشی یک گره با تعداد میانکشی‌های یک شبکه است، یکی از اجزای توپولوژی شبکه است که اندازه‌گیری می‌شود. گره‌ای که دارای توزیع درجه بالایی است به‌عنوان پروتئین هاب شناخته می‌شود. تصور می‌شود یک پروتئین هاب، که کدگذاری شده توسط یک ژن ضروری است، نقش مهمی در سلول ایفا می‌کند. هرگونه تغییر فیزیکی یا شیمیایی در این پروتئین هاب، می‌تواند میانکشی با سایر پروتئین‌ها را مختل کرده و در رفتار طبیعی

⁴ Edge

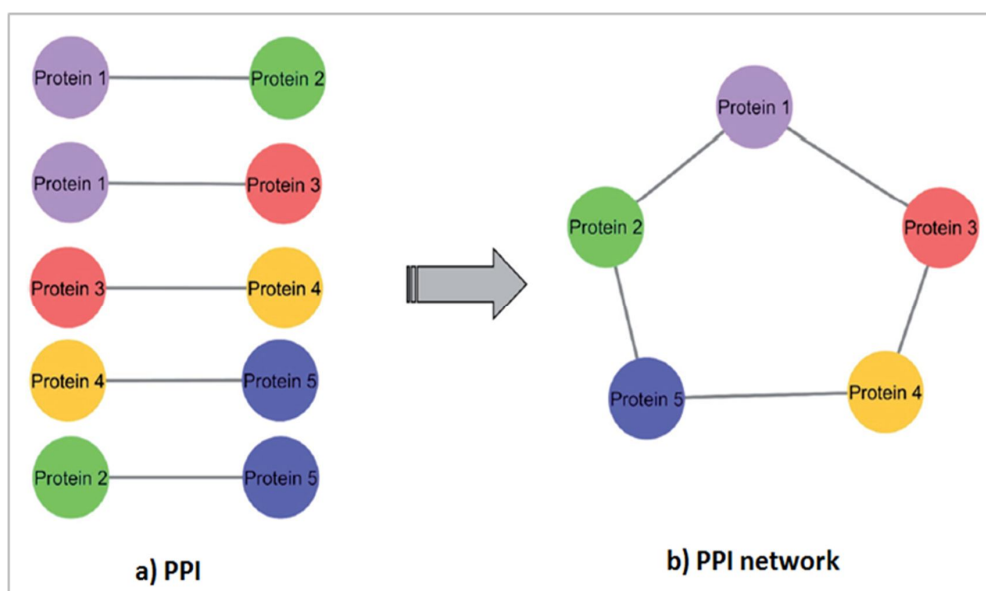
⁵ Protein-protein interaction network

¹ Computational systems analysis

² Network Biology

³ Nodes

سلول‌ها اختلال ایجاد کند و در نهایت منجر به یک بیماری شود (۱۳).



شکل ۲: شکلی شماتیک از (a) میانکنش‌های پروتئینی-پروتئینی، (b) شبکه میانکنش‌های پروتئینی-پروتئینی. شبکه‌ی PPI از چندین میانکنش پروتئینی تشکیل شده است. دایره‌های رنگی اشاره به پروتئین‌های مختلف دارد (۹).

مشابه باشند. در این روش، فرض بر این است که پروتئین‌های یک ماژول^۸ با بیماری‌های مشابهی مرتبط هستند (۱۵، ۱۶). خوشه‌ها با استفاده از الگوریتم‌ها شناسایی می‌شوند. چندین الگوریتم خوشه‌بندی برای ایجاد خوشه‌ها از جمله CFinder (۱۷)، خوشه‌بندی با گسترش همپوشانی پروتئین‌های همسایه (ClusterONE) (۱۸)، خوشه‌بندی بر اساس گروه‌های حداکثر (CMC) (۱۹)، روش نفوذ دسته^۹ (CPM) (۲۰)، خوشه‌بندی مبتنی بر تراکم پیرامونی^{۱۰} (DPCLUS) (۲۱)، خوشه‌بندی مبتنی بر همپوشانی چگالی پیرامون^{۱۱} (DPCLUSO) (۲۲)، الگوریتم پیچیده شناسایی پروتئینی^{۱۲} (IPCA) (۲۳)، الگوریتم ادغام دسته‌های محلی (LCMA) (۲۴)، خوشه‌بندی مبتنی بر جستجوی همسایگی محدود^{۱۳} (RNCS) (۲۵)، خوشه‌بندی مارکوف (MCL) (۲۶)، تشخیص کمپلکس مولکولی (MCOE) (۲۷) و غیره وجود دارند (۲۸).

(۲) روش پیوستگی یکی دیگر از آنالیزهای شبکه‌ای است که می‌تواند برای درک پاتوبیولوژی بیماری‌های انسانی مورد استفاده قرار گیرد. فرضیه اصلی در این روش این است که دو پروتئین (باهم پیوستگی دارند)، با یکدیگر میانکنش می‌دهند و به بیماری‌های یکسانی مربوط می‌شوند. به‌عنوان مثال Oti و همکاران با روش پیوستگی جفتی^۱ پیش‌بینی کردند که پروتئین ژانوس کیناز ۳ (JAK3) به‌عنوان پروتئینی که ممکن است با سندرم نقص ایمنی ترکیبی شدید^۲ (SCID) مرتبط باشد، زیرا JAK3 مستقیماً با پروتئین‌های LCK (لنفوسیت‌های مختص پروتئین-تیروزین^۳، تیروزین فسفاتاز^۴ (PTPRC) و گیرنده اینترلوکین ۲^۵ (IL2RG) میانکنش می‌دهد (۱۴).

(۳) خوشه‌بندی کردن^۷ نیز فن دیگری در آنالیز شبکه‌های زیستی در بیماری‌های انسانی محسوب می‌شود. خوشه به گروه کوچکی اطلاق می‌شود که ویژگی‌های شبکه‌ای آن‌ها از نظر توپولوژیکی

⁸ Module

⁹ Clique Percolation Method

¹⁰ Density-periphery

¹¹ Density-periphery overlapping-based clustering

¹² Identifying Protein Complex Algorithm

¹³ Restricted neighborhood search clustering

¹ Pairwise Linkage

² Janus kinase 3

³ Severe Combined Immunodeficiency Syndrome

⁴ Lymphocyte Specific Protein-Tyrosine

⁵ Protein-Tyrosine Phosphatase

⁶ Interleukin 2 Receptor

⁷ Clustering

۲-۲ آنالیز مسیرهای زیستی

یک مسیر زیستی به گروهی از مولکول‌ها گفته می‌شود که برای انجام عملکرد بیولوژیکی مشابه باهم میانکنش می‌دهند. شبکه میانکنش‌های پروتئینی دارای گره (پروتئین‌ها) و خط‌هاست. درحالی‌که، مسیرها متشکل از تعداد کمی گره است که این گره‌ها ممکن است ژن‌های سیگنالینگ، پروتئین‌ها، کمپلکس‌ها و یا متابولیت‌ها باشند که توسط میانکنش‌های مختلفی از جمله فعال‌سازی، مهار، اتصال و غیره به هم متصل می‌شوند. یک مسیر، مکانیسم انجام یک فعالیت بیولوژیکی خاص را در یک سلول نشان می‌دهد. همانند شبکه میانکنش‌های پروتئینی، ترکیبی از چندین مسیر، شبکه‌ی مسیرها را تشکیل می‌دهد (۲۹).

در شبکه‌های زیستی به‌طور کلی سه نوع مسیر اصلی، سیگنالینگ، تنظیمی و متابولیک وجود دارد. مسیر سیگنالینگ پاسخ سلولی را پس از دریافت سیگنال خارج سلولی پیش‌بینی می‌کند. انتقال سیگنال زمانی شروع می‌شود که در خارج از سلول سیگنالی برای فعال کردن گیرنده‌ای که در سطح سلول قرار دارد پدیدار شود. گیرنده فعال و به سیگنال متصل می‌شود و مولکول‌های داخل سلولی را تغییر می‌دهد تا به هرگونه اختلال در مسیر سیگنال‌دهی پاسخ دهد در نتیجه می‌تواند باعث بیماری شود، زیرا سلول‌ها نمی‌توانند در حالت عادی بازگردند یا زمانی که سیگنال‌ها دریافت می‌کنند به‌درستی پاسخ دهند (۳۰). مسیر تنظیمی بیان ژن یا پروتئین را در یک سلول نشان می‌دهد، خواه افزایش بیان داشته باشد یا کاهش یافته باشد. فعالیت‌های بیولوژیکی مانند رونویسی، ترجمه و اصلاح پس از ترجمه از جمله فعالیت‌هایی هستند که مسیر تنظیمی را در بر می‌گیرند (۳۱). درحالی‌که، در یک مسیر متابولیک، متابولیت پرایمر از طریق یک سری واکنش‌های شیمیایی که توسط آنزیم‌ها کاتالیز می‌شوند، به متابولیت دیگری تبدیل می‌شود (۳۲).

۳- مخزن و پایگاه داده

داده‌ها مهم‌ترین منبع در آنالیز سیستم‌های محاسباتی هستند. اکثر آنالیزها نیاز به ادغام چندین داده برای درک بیماری‌ها، به‌ویژه بیماری‌های پیچیده در یک دیدگاه سیستمیک دارند. برای مثال سطوح مختلف اومیکس (ژنومیکس، ترانسکریپتومیکس، پروتئومیکس، و یا متابولومیکس) با داده‌های میانکنش‌ها (میانکنش‌های پروتئینی یا مسیرها) برای ایجاد شبکه‌های زیستی ادغام می‌شوند. برای درک فرآیندهای سلولی به‌عنوان سیستم‌های وابسته و متصل به هم و در مفهوم یک پدیده بیولوژیکی، نیاز به یک رویکرد یکپارچه است که اطلاعات را تا حد ممکن از منابع مختلف از جمله مقالات، پایگاه‌های داده عمومی، آزمایش‌های بیوشیمیایی و

کنتیکتی، مطالعات فوتوتیپ و آنالیز ژنوم با توان بالا، مطالعات رونویسی، پروتئوم، اینتراکتوم و متابولوم استخراج کند. از این‌رو، مخزن داده^۱ یا توسعه پایگاه‌های داده یکی از رویکردهای اصلی برای تسهیل جستجوی دلخواه از داده‌ها برای انجام آنالیز سیستم‌های محاسباتی است (۳۳).

مخزن داده یکی از تلاش‌هایی است که مجموعه‌های رو به رشد داده‌های آزمایشگاهی را به شیوه‌ای مناسب ادغام می‌کند که می‌توان برای آنالیزهای بیشتر در دسترس عموم قرار گیرد. به‌عنوان مثال، پایگاه‌های داده‌ای مانند ArrayExpress (۳۴)، Gene Expression Omnibus (GEO) (۳۵) و CIBEX (۳۶) وجود دارد که مجموعه‌های داده را برای مطالعات بیان ژن ذخیره می‌کند تا در دسترس عموم قرار گیرد. به غیر از آن، همچنین پایگاه‌های اطلاعاتی مانند PubMed، Scopus Online و Google Scholar برای به دست آوردن مطالعات منتشر شده توسط محققان وجود دارد و چندین مطالعه وجود دارد که مجموعه داده‌های اومیکس تولید شده را در بخش تکمیلی ارائه می‌کند.

پیشینه و هدف مطالعه:

در سال ۲۰۲۱ چن و همکاران با استفاده از آنالیز شبکه‌های میانکنش پروتئینی در مکانیسم مولکولی سندرم تخمدان پلی‌کیستیک به ۵ ژن کلیدی که تاکنون در هیچ مقاله‌ای به‌عنوان ژن درگیر در PCOS مطرح نشده بودند، رسیدند. آن‌ها با استفاده از دیتابیس DisGeNet چیزی حدود ۵۲۲ ژن برای PCOS استخراج کردند و سپس این ژن‌ها را وارد دیتابیس STRING کردند تا میانکنش‌های بین پروتئین‌های این ژن‌ها را آنالیز کنند و با فیلتر کردن به ۴۸۸ ژن و ۲۷۶۷ ارتباط میانکنشی رسیدند. علاوه بر این پس از آنالیز شبکه‌های میانکنش بین پروتئین‌ها با استفاده از الگوریتم MCODE به ۷ ماژول^۲ (ژن‌های مرتبط به هم در واحدها یا ماژول‌های مجزا قرار می‌گیرند) اصلی ژن رسیدند. ماژول‌های اصلی تولید شده در عملکردهای بیولوژیکی خاصی مانند سرطان و تکثیر سلولی و آپوپتوز، تنظیم متابولیسم لیپید و گلوکز، چرخه سلولی و غیره توانمند بودند. آنالیز یکپارچه انجام شده در این مطالعه نشان داد که این ماژول‌های هاب^۳ و ژن‌های مرتبط با آن‌ها ارتباط نزدیکی با پاتوژنز PCOS دارند، که احتمالاً ممکن است بینش جدیدی برای درمان PCOS و مطالعه مکانیسم بیماری‌زایی پنهان آن ارائه دهد. چن و همکاران رابطه بین چندین ژن کلیدی از جمله ALB، TOP2A، PTGER3، NPB و BRD2 را در ماژول‌ها و سندرم تخمدان پلی‌کیستیک که تاکنون بررسی نشده بودند را عنوان کردند که نیازمند اینست که این ژن‌ها با تحقیقات بیشتر در

³ Hub

¹ Data repository

² Module

Polycystic Systematic Review و 5 years با عبارت کلیدی "Ovary Syndrome NOT Treatment" جستجو شد که به ۵۱۹ نتیجه رسید. از میان این تعداد مقالات مرتبط با موضوع انتخاب شدند. همچنین در مقدمه برای قسمت‌های مرتبط با آنالیزهای بیوانفورماتیکی در دیتابیس Google scholar عبارت‌های "Polycystic Ovary Syndrome AND Bioinformatics"، "PCOS AND Computational Biology" و "PCOS AND Network analysis" مورد جستجو قرار گرفتند که در مجموع ۱۶۴۰ مقاله یافت شد. مقالات پس از بررسی عنوان، در مرحله بعد از نظر ارتباط چکیده با هدف مورد نظر ارزیابی شدند. مقالاتی که با استفاده از ابزارهای بیوانفورماتیک درمان یا دارویی برای سندروم تخمدان پلی‌کیستیک پیشنهاد می‌دادند از روند بررسی حذف شدند. ۲- مقالات اصیل پژوهشی برای یافتن مقالاتی که با استفاده از ابزارهای بیوانفورماتیک به ژن‌های کلیدی در این سندروم رسیده‌اند، با عبارت مورد جستجوی "Polycystic ovary syndrome AND Bioinformatics analysis" جستجو شد که ۸ نتیجه یافت شد و از بین آن‌ها مقالات مربوط به RNA ها و میتوکندری حذف گردید. ۳- کتاب‌های چاپ شده با موضوع سندروم تخمدان پلی‌کیستیک به زبان انگلیسی به مطالعه وارد شد.

معیارهای خروج مقالات از مطالعه شامل: عدم ارائه متن کامل مقاله، مقالات با متن کامل غیر از انگلیسی زبان، مقالات با کیفیت پایین و عدم ارتباط مقالات با موضوع مرور سیستماتیک، نتایج به دست آمده از پایان نامه‌ها، مقالات کنفرانسی و وب سایت‌ها از روند بررسی حذف گردید. همچنین مقالات گزارش مورد نیز وارد مطالعه نشدند.

یافته‌ها

سیستم‌های محاسباتی و اومیکس در سندروم تخمدان

پلی‌کیستیک:

PCOS اختلالی ناهمگن است که ممکن است تحت تأثیر عوامل متعددی از جمله ژنتیک، سبک زندگی و محیط قرار گیرد. تعریف مشخصی برای PCOS وجود ندارد زیرا با ترکیبی از ویژگی‌های مختلف که منجر به معیارهای تشخیصی آن می‌شود، توصیف می‌شود. زنان مبتلا به سندرم تخمدان پلی‌کیستیک علائم مختلفی مشاهده می‌کنند و بیماری‌هایی که با سندرم تخمدان پلی‌کیستیک همراه هستند بسیار متنوع است. بسیاری از ژن‌ها، پروتئین‌ها و متابولیت‌ها در پاتوبیولوژی PCOS دخیل هستند بنابراین یک بیماری با پیچیدگی زیاد محسوب می‌شود. در جدول

نمونه‌های بزرگ‌تر، چند مرکزی و قومیت‌های مختلف تأیید شود (۳۷).

زو^۱ و همکاران در سال ۲۰۲۲ با استفاده از آنالیزهای بیوانفورماتیک ژن‌های جدیدی که به‌طور قابل‌توجهی با سندروم تخمدان پلی‌کیستیک و سرطان تخمدان^۲ (OC) مرتبط هستند را شناسایی کردند. آن‌ها دو مجموعه داده GSE140082 را برای سرطان تخمدان و مجموعه داده GSE34526 را برای PCOS از دیتابیس جنو (Geo) استخراج کردند و با استفاده از آنالیزهای متعدد از جمله آنالیز بیان ژن‌ها، آنالیز شبکه میانکنش‌های پروتئینی، آنالیز غنی‌سازی مسیرها و عملکرد ژن‌ها و غیره به ۱۲ ژن کاندید شامل RNF144B، LPAR3، CRISPLD2، JCHAIN، OR7E14P، PTPRD، IL27RA، STAT1، NR4A1، OGN، GALNT6 و CXCL11 مرتبط با این دو بیماری رسیدند. آنالیز حساسیت دارویی نشان داد که ژن OGN می‌تواند بیان بیان FSHR را افزایش دهد، که نشان‌دهنده‌ی این است که OGN می‌تواند پاسخ هورمونی را در PCOS و OC تنظیم کند (۳۸).

هدف از این مطالعه، مروری بر روش‌هایی است که طی سال‌های اخیر بدون نیاز به کار آزمایشگاهی و با استفاده از ابزارهای بیوانفورماتیک و آنالیزهای زیست‌شناسی محاسباتی برای رسیدن به درک صحیحی از لایه‌های مختلف یک بیماری چندعاملی و در نهایت رسیدن به ژن‌های کلیدی در این بیماری‌ها از جمله سندروم تخمدان پلی‌کیستیک ابداع شده است.

مواد و روش کار

در این مطالعه‌ی مروری پایگاه‌های داده لاتین Google Scholar، PubMed، Science direct محدود به زبان انگلیسی، بدون محدودیت زمانی با رعایت معیارهای ورود و خروج با استفاده از کلیدواژه‌های Polycystic Ovarian Syndrome، Protein- Network biology، Computational biology و Pathway analysis و protein interaction analysis در بازه زمانی ۳ ساله از سال ۲۰۲۰ تا ۲۰۲۲ میلادی مورد جستجو قرار گرفتند.

معیارهای ورود در این مطالعه منابع چاپ شده (مقاله یا کتاب) تا سال ۲۰۲۲ شامل: (مقالات مروری، مقالات اصیل پژوهشی به زبان انگلیسی و همچنین کتاب‌های لاتین) با موضوع سندروم تخمدان پلی‌کیستیک و روش‌های بیوانفورماتیک می‌باشد:

۱- مقالات مروری برای تعریف این سندروم، علائم، شیوع و غیره در دیتابیس PubMed با فیلترهای Full text، Review.

² Ovarian Cancer

¹ Zou

درمان، زنان مبتلا به PCOS در معرض خطر بیشتری برای ابتلا به سایر بیماری‌های مزمن (سرطان رحم، دیابت نوع ۲، بیماری‌های قلبی عروقی) قرار دارند، رویکردهای دیگری مانند آنالیز سیستم‌های محاسباتی برای درک بهتر این سندروم باید انجام گیرد.

(۱) تمام پلتفرم‌های اومیکس برای شناسایی مبنای مولکولی PCOS اعمال شده است (۳۹). با وجود تمام مطالعات انجام شده در سطوح مختلف اومیکس، پاتوبیولوژی PCOS هنوز درک نشده است. از آنجایی که شیوع سندرم تخمدان پلی‌کیستیک در زنان افزایش یافته و در صورت عدم

جدول (۱): رویکردهای اومیکس در سندروم تخمدان پلی‌کیستیک

اومیکس	شرح	مثال‌هایی از مطالعات انجام گرفته	منابع
ژنومیکس	شناسایی شواهد ژنتیکی در زنان مبتلا به PCOS	شناسایی ۱۶ جایگاه ژنی مرتبط با PCOS در افراد چینی و اروپایی	(۴۰-۴۳)
ترانس کریپتومیکس	شناسایی ژنهای متفاوت بیان شده (ژن‌های معنی‌دار با بیشترین بیان و کمترین بیان) بین زنان مبتلا به PCOS و نرمال	شناسایی ۲۴۳ ژن بیان شده افتراقی در سلول‌های گرانولوزا بین افراد نرمال و زنان مبتلا به PCOS	(۴۴)
پروتئومیکس	تشخیص پروتئین متفاوت بیان شده بین زنان مبتلا به PCOS و افراد نرمال	۱۸۶ پروتئین بیان شده به‌طور معنی‌داری در مایع فولیکولار بین زنان مبتلا به PCOS و نرمال مشخص شد.	(۴۵)
متابولومیکس	تشخیص متابولیت‌های تغییر یافته بین زنان مبتلا به PCOS و نرمال	متابولیت‌های تغییر یافته در سرم بین زنان نرمال و مبتلا به PCOS اختلال در چندین مسیر متابولیک مانند بیوسنتز هورمون استروئیدی، اسیدهای آمینه، متابولیسم نوکلئوتیدها و گلوکاتینون، همچنین متابولیسم لیپیدها و کربوهیدرات‌ها را نشان دادند.	(۴۶)

شبکه‌های زیستی در پایگاه‌های داده زیر گردآوری و ذخیره شده است (جدول ۲).

در سندروم تخمدان پلی‌کیستیک از آنالیز شبکه‌ها و مسیرها اغلب برای شناسایی ژن‌ها یا پروتئین‌ها، آنتولوژی^۳ آن‌ها و مسیرهایی که ممکن است در این سندروم دخیل باشند استفاده می‌شود. با استفاده از آنالیز شبکه‌ها و مسیرها می‌توان علاوه بر شناسایی اساس مولکولی و عملکردهای بیولوژیکی مرتبط به PCOS، ارتباط مولکولی سندروم تخمدان پلی‌کیستیک را با سایر بیماری‌ها رمزگشایی کرد و با بهبود دانش، روش‌های درمانی برای این سندروم اتخاذ کرد. ترسیم PPIN برای PCOS و آنالیز هسته‌ی اصلی آن نه تنها برای بررسی مکانیسم مولکولی درگیر در ایجاد PCOS مفید است، بلکه مبنایی برای درمان بالینی آن نیز فراهم می‌کند.

در سال ۲۰۱۶ کوری و همکاران از شبکه میانکنش‌های پروتئینی و آنالیز مسیرها استفاده کردند. آن‌ها با ادغام سه مجموعه

آنالیز مسیرها و میانکنش‌های پروتئینی-پروتئینی در سندروم تخمدان پلی‌کیستیک:

شناسایی پروتئین‌هایی که باهم میانکنش می‌دهند، اطلاعات مهم در ارتباط با فعالیت ژن‌ها فراهم می‌کند. این اطلاعات خصوصاً زمانی که یک ژن یا پروتئین جدید شناسایی می‌شود با ارزش می‌باشد، زیرا میانکنش یک پروتئین ناشناخته با پروتئین شناخته شده گاهی می‌تواند نشان‌دهنده‌ی نقش پروتئین ناشناس باشد. چندین روش برای مطالعه‌ی میانکنش بین پروتئین‌ها وجود دارد از جمله سیستم هیبریدی مخمر^۱ و اسپکترومتری جرمی^۲ (۴۷). همه‌ی روش‌هایی که شبکه‌ی کاملی از برهمکنش‌های فیزیکی و کارکردی پروتئین‌ها (اینتراکتوم)^۳ ایجاد کرده‌اند به تدریج در چندین ارگانیزم مانند ویروس (ویروس تیخال)^۴ (۴۸)، پروکاریوت (اشرشیاکلی) (۴۹)، یوکاریوت (مخمر) (۵۰)، نماتد (۵۱)، مگس میوه (۵۲) و انسان (۵۳) شناسایی شده‌اند. مجموعه داده‌های این

³ Interactome

⁴ Ontology

¹ Yeast two-hybrid system

² Mass spectrometry

داده میکرواری مربوط به PCOS و داده‌های میانکنش‌های پروتئینی، آنالیز غنی‌سازی مسیر را انجام دادند و نتایج PCOS را با سرطان تخمدان و اندومتريوز مقایسه کردند. آنالیزها نشان داد که PCOS ارتباط نزدیکی با اندومتريوز و سرطان تخمدان دارد زیرا این سه بیماری دارای چندین مولکول و مسیر مشترک مانند سیگنال دهی MAPK، چرخه سلولی و آپوپتوز بودند (۵۴).
 راملی و همکاران (۲۰۱۹) از میانکنش‌های پروتئین-پروتئین و آنالیز مسیرها برای شناسایی پروتئین و مسیرها برای توضیح روابط بین PCOS و ۱۷ بیماری دیگر مانند میگرن، سرطان تخمدان و اسکیزوفرنی استفاده کردند. آن‌ها از یک رویکرد خوشه‌بندی توسط الگوریتم MCODE برای شناسایی پروتئین‌های مشترک بین

PCOS و سایر بیماری‌ها و آنالیز غنی‌سازی مسیر برای شناسایی مسیرهایی که ممکن است PCOS و بیماری‌های مرتبط با PCOS را به هم متصل کنند، استفاده کردند (۵۵).
 بیماری پیچیده یا چند عاملی مثل PCOS شامل چندین مسیر از جمله مسیرهای سیگنالینگ، تنظیمی و متابولیک می‌شود. ترکیب و ادغام چندین مسیر با انواع دیگر داده‌ها مانند میانکنش‌های پروتئینی یکی از رویکردهای ارزشمندی است که می‌تواند برای بهبود درک مکانیسم‌های پیچیده بیماری مورد استفاده قرار گیرد. داده‌های مسیرها را می‌توان از پایگاه داده‌ی مربوط به مسیرهای بازاری کرد (جدول ۳) (۵۶).

جدول (۲): فهرست پایگاه‌های داده میانکنش‌های پروتئینی و شبکه‌های زیستی مربوط به انسان

ردیف	پایگاه داده	توضیح	نشانی اینترنتی	منابع
۱	Agile Protein Interactomes DataServer (APID)	مجموعه‌ای از اینتراکتوم‌های پروتئینی معتبر شناخته شده یکپارچه، شامل ۴۰۰ گونه از جمله انسان می‌باشد.	http://apid.dep.usal.es/	(۵۷)
۲	BioGRID	مجموعه‌ای از میانکنش‌های ژنتیکی، فیزیکی و شیمیایی در انسان و همه موجودات اصلی	https://thebiogrid.org/	(۵۸)
۳	Database of Interacting Protein (DIP)	به‌طور دستی توسط متخصصان خبره اصلاح می‌شود. نتایج آزمایشگاهی داده‌های PPI را ذخیره می‌کند.	https://dip.doe-mbi.ucla.edu	(۵۹)
۴	GeneMANIA	شامل میانکنش‌های پروتئینی شناخته شده، که از پایگاه‌های داده برگزیده و نتایج آزمایشگاهی PPI مشتق شده است. میانکنش‌های پروتئینی را در ۶ جنس از جمله انسان پیش‌بینی می‌کند.	https://genemania.org/	(۶۰)
۵	Human Integrated Protein-Protein Interaction rEference (HIPPIE)	فقط شامل میانکنش‌های پروتئینی انسانی به دست آمده از نتایج آزمایشگاهی و پایگاه‌های داده PPI منتخب است.	http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/	(۶۱)
۶	Information Hyperlinked over Proteins (iHOP)	میانکنش‌های پروتئینی را که از مجموعه مقالات انتخاب شده‌اند، ارائه می‌دهد.	https://bio.tools/ihop	(۶۲)
۷	Human Protein Reference Database (HPRD)	اطلاعات مربوط به پروتئین‌های انسانی از جمله PPI را، ذخیره می‌کند.	https://www.hprd.org/	(۶۳)
۸	Interologous Interaction Database (I2D)	حاوی میانکنش‌های پروتئینی یکپارچه شناخته شده، تجربی و پیش‌بینی شده برای انسان و پنج گونه دیگر	http://ophid.utoronto.ca/ophidv2.204/	(۶۴)
۹	InnateDB	میانکنش‌های تأیید شده آزمایشگاهی بین ژن‌ها، پروتئین‌ها و مسیرهای سیگنال دهی درگیر در سیستم ایمنی ذاتی و پاسخ به عفونت	https://www.innatedb.com/	(۶۵)

			میکروبی را در انسان، موش و گاو ذخیره می‌کند.
(۶۶)	https://www.ebi.ac.uk/intact/	IntAct Molecular Interaction Database (IntAct)	تمام میانکنش‌ها از مقالات و ۱۱ پایگاه داده PPI بازیابی می‌شوند. این پایگاه‌های داده توسط EMBL-EBI و سایر تیم‌های پایگاه داده PPI اصلاح می‌شوند.
(۶۷)	https://mint.bio.uniroma2.it/	Molecular INTERaction database (MINT)	میانکنش‌های پروتئینی تأیید شده تجربی را که از نظر علمی انتخاب شده‌اند، جمع‌آوری می‌کند.
(۶۸)	https://string-db.org/	STRING	یک پایگاه داده PPI است که شواهد تعاملی را از برهمکنش‌های شناخته شده (پایگاه‌های اطلاعاتی منتخب و آزمایشگاهی)، میانکنش‌های پیش‌بینی شده (در همسایگی، آمیختگی ژن، و هم‌زمانی) و سایر موارد (هم‌بیانی و متن کاوی) ارائه می‌دهد.
(۶۹)	http://matrixdb.univ-lyon1.fr/	The Extracellular Matrix Interaction Database (MatrixDB)	میانکنش‌های پروتئین‌های ماتریکس خارج سلولی، پروتئوگلیکان‌ها و پلی‌ساکاریدها را ذخیره می‌کند.
(۷۰)	http://wilab.inha.ac.kr/hpid/	The Human Protein Interaction Database (HPID)	پروتئین‌ها را از پایگاه داده BIND (این پایگاه داده در دسترس عموم نیست)، DIP و HPRD جمع‌آوری و باقیه را پیش‌بینی می‌کند.
(۷۱)	https://www.imexconsortium.org/	The International Molecular Exchange Consortium (IMEx)	مجموعه پایگاه‌های داده PPI لازم را از پایگاه داده اصلی PPI مانند IntAct، BIND، MINT، DIP و MIPS ارائه می‌کند.
(۷۲)	http://mips.helmholtz-muenchen.de/proj/ppi/	The Mammalian Protein-Protein Interaction Database (MIPS)	PPI‌های با کیفیت بالا را از نتایج آزمایشگاهی جمع‌آوری می‌کند.

جدول (۳): فهرست پایگاه‌های داده مسیرهای زیستی مربوط به انسان

ردیف	پایگاه داده	توضیح	نشانی اینترنتی	منابع
۱	BioCarta	پایگاه داده مسیری است که میانکنش ژن‌ها را در مسیرهای فرآیندهای سلولی انسان فراهم می‌کند.	https://cgap.nci.nih.gov/Pathways/BioCarta-Pathways	(۷۳)
۲	HumanCyc	یکی از پایگاه‌های داده مسیر در BioCyc که از مسیرهای متابولیک انسان تشکیل شده است.	https://humancyc.org/	(۷۴)
۳	Kyoto Encyclopedia of Genes and Genomes (KEGG)	مخزنی برای درک عملکردها و کاربردهای	https://www.kegg.jp/	(۷۵)

		سطح بالای سیستم‌های بیولوژیکی چندین ارگانسیم از جمله انسان که از توالی یابی ژنوم و سایر فناوری‌های تجربی با کارایی بالا به دست آمده‌اند.		
(۷۶)	http://pantherdb.org/	حاوی اطلاعات بیولوژیکی مانند مسیرهای پروتئین است که همراه با ابزارهایی برای آنالیز پروتئین چندین موجود زنده از جمله انسان به کار می‌رود.	PANTHER	۴
(۷۷)	https://reactome.org/	پایگاه داده مسیری برای چندین موجود زنده از جمله انسان که ابزارهایی را برای آنالیز مسیر فراهم می‌کند	Reactome	۵
(۷۸)	https://www.wikipathways.org/ index.php/WikiPathways	پایگاه داده مسیری که حاوی اطلاعات مسیر چند موجود زنده مانند انسان است	Wikipathways	۶

مخزن داده در سندروم تخمدان پلی کیستیک:

مجموعه داده‌های زیادی وجود دارد که توسط پلتفرم‌های اومیکس برای شناسایی پاتوبیولوژی PCOS ایجاد شده است. مجموعه اطلاعات در سرورها به صورت تصادفی توزیع می‌شوند و اگر محققین قصد بازیابی اطلاعات مربوط به PCOS را داشته باشند روندی بسیار خسته کننده دارد. از این رو، داشتن مخزنی که اطلاعات جامع PCOS را ذخیره می‌کند مؤثر است. برای ذخیره و بازیابی اطلاعات مولکولی جمع‌آوری شده در مورد سندروم تخمدان پلی کیستیک تاکنون سه پایگاه داده ایجاد شده است که عبارتند از PCOSBase(www.pcosbase.org)(۷۹)،

PCOSKB(<http://pcoskb.bicnirrh.res.in/>)(۸۰) و PCOSDB(<http://pcosdb.net/>)(۸۱). پایگاه داده PCOSKB و PCOSDB به ترتیب حاوی ۲۴۱ و ۲۰۸ ژن مربوط به PCOS هستند. این پایگاه‌های اطلاعاتی ژن‌های مرتبط به سندرم تخمدان پلی کیستیک را بر اساس متون علمی منتشر شده ارائه می‌دهند. پایگاه داده PCOSBase حدود ۸۱۸۵ پروتئین مرتبط با سندرم تخمدان پلی کیستیک را که از پایگاه‌های داده بیماری‌های قبلی،

مطالعات بیان ژن و پروتئین‌ها به دست آمده بودند را ارائه می‌دهد. هر سه پایگاه داده PCOS شرح مفصلی برای هر ورودی که به PCOS مرتبط می‌شود را در دسترس قرار می‌دهند ولی برای به دست آوردن داده‌های بیشتر و جزئی‌تر از این سندروم، به این دو پایگاه داده اصلی یعنی UniProt (<https://www.uniprot.org/>) و NCBI لینک دارد. پایگاه داده PCOSBase، اطلاعات بیولوژیکی مربوط به PCOS را از جمله مکان کروموزومی، وجود ژن‌ها، مسیرها، دومین‌ها، بیماری‌های مرتبط و مکان بافتی همه پروتئین‌های مرتبط با PCOS را در اختیار عموم قرار می‌دهند.

بحث و نتیجه‌گیری

بعضی از بیماری‌های مزمن پیچیده انسان، مانند چاقی و دیابت نوع دوم، به سرعت نقشه‌برداری شده‌اند که منجر به پیشرفت‌های بهتری از ویژگی‌های این اختلالات می‌شود. بر خلاف پیشرفت مشاهده شده در درک این بیماری‌ها، تاکنون پیچیدگی PCOS به خوبی روشن نشده است که دلایل بالقوه زیادی وجود دارد از جمله: ۱- ناهمگونی

در فنوتیپ ها ۲- ماهیت پیچیده سندرم ۳- معیارهای تشخیصی متغیر ۴- بررسی تعداد محدودی از ژن‌های درگیر در این سندرم ۵- ارتباطات متناقض و غیر تکراری PCOS. جمعیت‌های بزرگ برای هر بررسی اپیدمیولوژیکی حیاتی هستند، با این حال، در مطالعات اپیدمیولوژیکی ژنتیکی، این وضعیت به یک شرط ضروری تبدیل می‌شود. در مورد سندرم تخمدان پلی‌کیستیک، بررسی تعداد کمی ژن ممکن است باعث ایجاد اختلال در جمعیت همگن شود. بر این اساس، مطالعاتی که کل ژن‌های کاندید را بررسی می‌کنند، به بررسی‌های مختلف ژنی ترجیح داده می‌شوند، هرچند ممکن است قابل تکرار نباشند. همه شرط‌های لازم فوق ممکن است با مطالعات ارتباطی گسترده ژنوم، معیارهای تشخیصی منحصر به فرد، متا-آنالیزها و شبکه‌های میانکنش‌های پروتئینی مورد بررسی قرار گیرند (۸۲).

بر همین اساس پیشنهاد می‌شود بیماری‌های پیچیده انسانی مثل سندرم تخمدان پلی‌کیستیک با دید وسیع‌تری مورد توجه قرار گیرند و از آنجاکه ژن‌ها، پروتئین‌های و متابولیت‌های مختلفی در این سندرم درگیر هستند، با استفاده از زیست‌شناسی محاسباتی بررسی لایه‌های مختلف اومیکس از جمله ژنومیکس، ترانسکریپتومیکس، پروتئومیکس و متابولومیکس در این سندرم پیشنهاد می‌شود. همچنین آنالیز ارتباطات ژن‌ها با یکدیگر، میانکنش‌های پروتئینی و ترسیم شبکه این میانکنش‌ها می‌تواند کمک شایانی در درک این سندرم داشته باشد.

امروزه ابزارهای بیوانفورماتیک مختلفی برای دستیابی به لایه‌های مختلف بیماری‌ها ابداع شده است که در کنار نتایج آزمایشگاهی می‌تواند کمک بزرگی در شناخت این بیماری‌ها باشد. برای هر مرحله از اومیکس دیتابیس‌های متفاوتی طراحی شده است. به‌عنوان مثال برای بررسی سندرم تخمدان پلی‌کیستیک در مرحله اول نیاز است تمام ژن‌های درگیر در این بیماری را مطالعه کرد. برای به دست آوردن ژن‌های درگیر در بیماری‌ها می‌توان از دیتابیس‌های بیماری‌های انسان مثل DisGeNet (۸۳)، GWAS Catalog (۸۴)، Online Mendelian Inheritance in Man (OMIM) (۴۵)، Human Gene Mutation Database (HGMD) (۸۵) ژن‌های مورد نیاز برای مطالعه را استخراج کرد. این دیتابیس‌ها علاوه بر ژن‌ها، اطلاعات مربوط به پروتئین‌ها، متابولیت‌ها، داروها، فرایندهای بیولوژیکی، به‌افتها و سایر موارد مرتبط با پاتوبیولوژی، پاتوژن و پاتوفیزیولوژی بیماری‌ها را در خود ذخیره کرده‌اند. علاوه بر این دیتابیس‌هایی مختص به بیماری‌های خاص از جمله دیابت نوع دوم (T2D-Db) (۸۶)، آلزایمر (AlzBase) (۸۷)، سندرم تخمدان پلی‌کیستیک (PCOSBease) و انواع سرطان‌ها (The Cancer Genome Atlas (TCGA) (۸۸)

طراحی شده است. در سال ۲۰۲۱، تقوی و همکاران با استفاده از داده‌های دیتابیس TCGA، پیشنهاد کردند که پروتئین SENP1 می‌تواند به‌عنوان یک بیومارکر در کوهرت Pan-Cancer محسوب شود (۸۹).

برای رسیدن به ژن‌های کلیدی در بیماری‌های مختلف بعد از استخراج ژن‌ها از دیتابیس‌های موردنظر، نیاز است ژن‌هایی که بیان کاهش یافته و یا افزایش یافته دارند جدا شوند و سپس ارتباط آن‌ها را با ترسیم شبکه‌های میانکنشی بین پروتئین‌ها و آنالیز مسیرهای مختلفی که درگیر هستند، بررسی کنیم. نهایتاً با بررسی این ارتباطات ممکن است تظاهرات بالینی مختلف و متناقضی که بیماری‌های پیچیده‌ای مثل سندرم تخمدان پلی‌کیستیک در جمعیت‌های مختلف و شرایط اقلیمی متفاوت از خود بروز می‌دهند برابمان قابل درک باشد و بتوانیم با درک صحیحی از این بیماری‌ها به پیشگیری، تشخیص و با درمان آن‌ها کمک کنیم.

در سال ۲۰۱۷، محمد صابری انوار و همکاران با استفاده از نرم‌افزار Cytoscape (پلتفرمی برای تجسم میانکنش‌های شبکه‌های مولکولی و مسیرهای بیولوژیکی و ادغام این شبکه‌ها با حاشیه نویسی و پروفایل‌های بیان ژن)، پروتئین‌هایی را معرفی کردند که می‌تواند به‌عنوان مارکرهای تشخیصی و اهداف درمانی در سرطان معده استفاده شود. از طرفی از روند ارائه شده در مطالعه، می‌توان برای شناسایی اهداف کلیدی در سایر بیماری‌های پیچیده نیز استفاده نمود (۹۰).

در سال ۲۰۲۰، Islam و همکاران ژن‌های اصلی دخیل در پاتوژن PCOS را با استفاده از آنالیزهای بیوانفورماتیک شناسایی کردند. این گروه داده‌های ماکرواروی DNA مربوط به مبتلایان به PCOS را با کد GSE124226 از دیتابیس GEO استخراج کردند و با استفاده از R package limma ژن‌هایی با بیان متفاوت به دست آوردند. در مجموع ۱۸۰ ژن با بیان متفاوت بازیابی شد که ۷۳ ژن افزایش بیان و ۱۰۷ ژن کاهش بیان را نشان می‌داد. آنالیز عملکرد ژن‌ها با استفاده از پایگاه داده DAVID انجام شد. سپس با استفاده از نرم‌افزار Cytoscape شبکه میانکنش‌های بین پروتئین‌ها را ترسیم کردند و با استفاده از سیستم امتیازدهی cytoHubba، به ۴ ژن کلیدی یا هاب ژن (RARA، KPNB1، REL و MAPIB) دخیل در پاتوژن PCOS رسیدند. آنالیز ماژول‌ها نیز با استفاده از افزونه MCODE انجام شد. در نهایت، آن‌ها از ژن‌های هاب شناسایی‌شده برای نشان دادن علائم دارویی مهم که ممکن است به‌عنوان اهداف درمانی برای PCOS مفید باشد، استفاده کردند (۹۱).

به همین ترتیب و با استفاده از دیتابیس‌ها و پایگاه‌های داده مختلف در سه سال اخیر، مطالعات مختلفی سعی در درک این

این بیماری نیاز است علاوه بر بررسی‌های آزمایشگاهی، با استفاده از ابزارهای بیوانفورماتیک و ترسیم شبکه میانکنش‌های پروتئینی و به‌طور کلی با استفاده از زیست‌شناسی محاسباتی به کشف و درک ارتباطات این بیماری با سایر بیماری‌ها و ژن‌های درگیر در این سندروم، پرداخته شود.

تشکر و قدردانی

بدین وسیله از تمامی پژوهشگرانی که از مطالعات آن‌ها در این مقاله استفاده شده است، تشکر و قدردانی می‌شود.

سندروم و شناخت ارتباطات پروتئینی PCOS با دیگر بیماری‌ها داشته‌اند و ژن‌های متفاوتی در مراحل مختلف این بیماری به‌عنوان هاب ژن معرفی کرده‌اند (۹۵-۹۲).

سندرم تخمدان پلی‌کیستیک در سراسر جهان (PCOS) یکی از شایع‌ترین اختلالات هورمونی در زنان در سنین باروری است. با این حال، فقدان مطالعات ژنتیکی در مکانیسم‌های مولکولی این بیماری وجود دارد. از آنجاکه PCOS عامل اصلی ناباروری در زنان است و زنان مبتلا به PCOS در معرض سایر بیماری‌های مزمن از جمله سرطان رحم، بیماری‌های قلبی و عروقی، دیابت نوع دوم و غیره هستند بنابراین برای پیشگیری، تشخیص و درمان زودهنگام

References:

- Cunha A, Póvoa AM. Infertility management in women with polycystic ovary syndrome: a review. *Porto Biomed J* 2021a;6(1):e116. Available from: <http://dx.doi.org/10.1097/j.pbj.0000000000000116>.
- Pourhoseini SA, Babazadeh R, Abedian Z, Mazloum SR. Frequency of the Phenotypes of Polycystic Ovarian Syndrome in Iranian Adolescents (Mashhad) based on Rotterdam Criteria in 2018. *Iranian J Obstetr Gynecol Infertility* 2021;24(2):14-22.
- Rajska A, Buszewska-Forajta M, Rachoń D, Markuszewski MJ. Metabolomic insight into polycystic ovary syndrome—An overview. *Int J Mol Sci* 2020;21(14):4853.
- Nautiyal H, Imam SS, Alshehri S, Ghoneim MM, Afzal M, Alzarea SI, et al. Polycystic Ovarian Syndrome: A Complex Disease with a Genetics Approach. *Biomedicines* 2022;10(3):540.
- Abdolahian S, Ramezani Tehrani F, Nahidi F, Ghodsi D, Jafari M, Alavi Majd H. Relationship between body mass index and the clinical predictors of polycystic ovary syndrome in adolescent girls. *Iranian J Obstetr Gynecol Infertility* 2021;24(10):40-7.
- Zare Bidaki F, Chaichian S, Arbabi Bidgoli S. The role of oral Curcumin Nanomicell on improvement of menstrual disorders, depression and anxiety in women with Polycystic Ovarian Syndrome. *Iranian J Obstetr Gynecol Infertility* 2021;24(1):55-66.
- Egea RR, Puchalt NG, Escrivá MM, Varghese AC. OMICS: current and future perspectives in reproductive medicine and technology. *J Hum Reprod Sci* 2014;7(2):73.
- Minuchehr Z, Kheitan S. Bioinformatics in a Nutshell. *Bioinf Biocomp Res* 2016;1:34-41.
- Afiqah-Aleng N, Mohamed-Hussein Z-A. Computational systems analysis on polycystic ovarian syndrome (PCOS). *Polycystic Ovarian Syndrome* 2020:125-48.
- Barabasi A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;5(2):101-13.
- Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;12(1):56-68.
- Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature* 2009;461(7261):218-23.
- Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* 2005;21(23):4205-8.
- Oti M, Snel B, Huynen MA, Brunner HG. Predicting disease genes using protein-protein interactions. *J Med Genet* 2006;43(8):691-8.
- Liu W, Wu A, Pellegrini M, Wang X. Integrative

- analysis of human protein, function and disease networks. *Sci Rep* 2015;5(1):1-11.
16. Navlakha S, Kingsford C. The power of protein interaction networks for associating genes with diseases. *Bioinformatics* 2010;26(8):1057-63.
 17. Adamcsek B, Palla G, Farkas IJ, Derényi I, Vicsek T. CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 2006;22(8):1021-3.
 18. Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* 2012;9(5):471-2.
 19. Liu G, Wong L, Chua HN. Complex discovery from weighted PPI networks. *Bioinformatics* 2009;25(15):1891-7.
 20. Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 2005;435(7043):814-8.
 21. Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K, Kanaya S. Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinf* 2006;7(1):1-13.
 22. Bozlul KM, Wakamatsu N, Altaf-Ul-Amin M. DPCLUSOST: a software tool for general purpose graph clustering. *J Compu Aided Chem* 2017;18:76-93.
 23. Li M, Chen J-e, Wang J-x, Hu B, Chen G. Modifying the DPCLUS algorithm for identifying protein complexes based on new topological structures. *BMC Bioinf* 2008;9(1):1-16.
 24. Li X-L, Foo C-S, Tan S-H, Ng S-K. Interaction graph mining for protein complexes using local clique merging. *Genome Inform* 2005;16(2):260-9.
 25. King AD, Pržulj N, Jurisica I. Protein complex prediction via cost-based clustering. *Bioinformatics* 2004;20(17):3013-20.
 26. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 2002;30(7):1575-84.
 27. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf* 2003;4(1):1-27.
 28. Anvar MS, Minuchehr Z, Shahlaei M, Kheitan S. Gastric cancer biomarkers; A systems biology approach. *Biochem Biophys Rep* 2018;13:141-6.
 29. García-Campos MA, Espinal-Enriquez J, Hernández-Lemus E. Pathway analysis: state of the art. *Front Physiol* 2015;6:383.
 30. Torres-Ayuso P, Sahoo S, Ashton G, An E, Simms N, Galvin M, et al. Signaling pathway screening platforms are an efficient approach to identify therapeutic targets in cancers that lack known driver mutations: A case report for a cancer of unknown primary origin. *NPJ Genomic Med* 2018;3(1):1-7.
 31. Varala K, Marshall-Colón A, Cirrone J, Brooks MD, Pasquino AV, Lérán S, et al. Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. *Proc Natl Acad Sci U S A* 2018;115(25):6494-9. Available from: <http://dx.doi.org/10.1073/pnas.1721487115>.
 32. Doong SJ, Gupta A, Prather KL. Layered dynamic regulation for improving metabolic pathway productivity in *Escherichia coli*. *Proc Natl Acad Sci U S A* 2018;115(12):2964-9.
 33. Kitano H. Computational systems biology. *Nature* 2002;420(6912):206-10.
 34. Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, et al. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res* 2015;43(D1):D1113-D6.
 35. Clough E, Barrett T. The gene expression omnibus database. *Statistical genomics*: Springer; 2016. p. 93-110.
 36. Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y.

- CIBEX: center for information biology gene expression database. *C R Biol* 2003;326(10-11):1079-82.
37. Chen Q, Zheng B, Du S, Lin Y. Explore the potential molecular mechanism of polycystic ovarian syndrome by protein-protein interaction network analysis. *Taiwan J Obstet Gynecol* 2021;60(5):807-15.
 38. Zou J, Li Y, Liao N, Liu J, Zhang Q, Luo M, et al. Identification of key genes associated with polycystic ovary syndrome (PCOS) and ovarian cancer using an integrated bioinformatics analysis. *J Ovarian Res* 2022;15(1):1-16.
 39. Pei C-Z, Jin L, Baek K-H. Pathogenetic analysis of polycystic ovary syndrome from the perspective of omics. *Biomed Pharmacother* 2021;142:112031.
 40. Chen Z-J, Zhao H, He L, Shi Y, Qin Y, Shi Y, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16. 3, 2p21 and 9q33. 3. *Nat Genet* 2011;43(1):55-9.
 41. Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet* 2012;44(9):1020-5.
 42. Hayes MG, Urbanek M, Ehrmann DA, Armstrong LL, Lee JY, Sisk R, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nat Comm* 2015;6(1):1-13.
 43. Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat Comm* 2015;6(1):1-7.
 44. Lan C-W, Chen M-J, Tai K-Y, Yu DC, Yang Y-C, Jan P-S, et al. Functional microarray analysis of differentially expressed genes in granulosa cells from women with polycystic ovary syndrome related to MAPK/ERK signaling. *Sci Rep* 2015;5(1):1-10.
 45. Ambekar AS, Kelkar DS, Pinto SM, Sharma R, Hinduja I, Zaveri K, et al. Proteomics of follicular fluid from women with polycystic ovary syndrome suggests molecular defects in follicular development. *J Clin Endocrin Metabol* 2015;100(2):744-53.
 46. Dong F, Deng D, Chen H, Cheng W, Li Q, Luo R, et al. Serum metabolomics study of polycystic ovary syndrome based on UPLC-QTOF-MS coupled with a pattern recognition approach. *Anal Bioanal Chem* 2015;407(16):4683-95.
 47. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Nat Acad Sci* 2001;98(8):4569-74.
 48. Fossum E, Friedel CC, Rajagopala SV, Titz B, Baiker A, Schmidt T, et al. Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathogens* 2009;5(9):e1000570.
 49. Peregrin-Alvarez JM, Xiong X, Su C, Parkinson J. The modular organization of protein interactions in *Escherichia coli*. *PLoS Comput Biol* 2009;5(10):e1000523.
 50. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 2006;440(7084):637-43.
 51. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, et al. A map of the interactome network of the metazoan *C. elegans*. *Science* 2004;303(5657):540-3.
 52. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, et al. A protein interaction map of *Drosophila melanogaster*. *Science* 2003;302(5651):1727-36.
 53. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005;437(7062):1173-

- 8.
54. Kori M, Gov E, Arga KY. Molecular signatures of ovarian diseases: Insights from network medicine perspective. *Syst Biol Reprod Med* 2016;62(4):266-82.
55. Ramly B, Afiqah-Aleng N, Mohamed-Hussein Z-A. Protein-protein interaction network analysis reveals several diseases highly associated with polycystic ovarian syndrome. *Int J Mol Sci* 2019;20(12):2959.
56. Alaimo S, Marceca GP, Ferro A, Pulvirenti A. Detecting disease specific pathway substructures through an integrated systems biology approach. *Noncoding RNA* 2017;3(2):20.
57. Alonso-López D, Campos-Laborie FJ, Gutiérrez MA, Lambourne L, Calderwood MA, Vidal M, et al. APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database* 2019;2019.
58. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 2006;34(suppl_1):D535-D9.
59. Xenarios I, Fernandez E, Salwinski L, Duan XJ, Thompson MJ, Marcotte EM, et al. DIP: the database of interacting proteins: 2001 update. *Nucleic Acids Res* 2001;29(1):239-41.
60. Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, et al. GeneMANIA update 2018. *Nucleic Acids Res* 2018;46(W1):W60-W4.
61. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2. 0: enhancing meaningfulness and reliability of protein-protein interaction networks *Nucleic Acids Res* 2016:gkw985.
62. Hoffmann R, Valencia A. A gene network for navigating the literature. *Nature genetics* 2004;36(7):664-.
63. Keshava Prasad T, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human protein reference database—2009 update. *Nucleic Acids Res* 2009;37(suppl_1):D767-D72.
64. Brown KR, Jurisica I. Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol* 2007;8(5):1-11.
65. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res* 2013;41(D1):D1228-D33.
66. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* 2014;42(D1):D358-D63.
67. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 2012;40(D1):D857-D61.
68. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2016:gkw937.
69. Launay G, Salza R, Multedo D, Thierry-Mieg N, Ricard-Blum S. MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. *Nucleic Acids Res* 2015;43(D1):D321-D7.
70. Han K, Park B, Kim H, Hong J, Park J. HPID: the human protein interaction database. *Bioinformatics* 2004;20(15):2466-70.
71. Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, et al. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods* 2012;9(4):345-50.
72. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, et al. The MIPS mammalian protein-protein interaction database. *Bioinformatics* 2005;21(6):832-4.

73. Nishimura D. BioCarta. Biotech Software & Internet Report. Comput Software J Sci 2001;2(3):117-20.
74. Trupp M, Altman T, Fulcher CA, Caspi R, Krummenacker M, Paley S, et al. Beyond the genome (BTG) is a (PGDB) pathway genome database: HumanCyc. *Genome Biol* 2010;11(1):1-.
75. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;45(D1):D353-D61.
76. Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Protein networks and pathway analysis*: Springer; 2009. p. 123-40.
77. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The reactome pathway knowledgebase. *Nucleic Acids Res* 2018;46(D1):D649-D55.
78. Pico AR, Kelder T, Van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. *PLoS Biol* 2008;6(7):e184.
79. Afiqah-Aleng N, Harun S, Nor Muhammad NA, Mohamed-Hussein Z-A. PCOSBase: a manually curated database of polycystic ovarian syndrome. *Database* 2017;2017.
80. Joseph S, Barai RS, Bhujbalrao R, Idicula-Thomas S. PCOSKB: A KnowledgeBase on genes, diseases, ontology terms and biochemical pathways associated with PolyCystic Ovary Syndrome. *Nucleic Acids Res* 2016;44(D1):D1032-D5.
81. Maniraja J, Vetrivel U, Munuswamy D, Melanathuru V. PCOSDB: PolyCystic ovary syndrome DataBase for manually curated genes associated with the disease. *Bioinformatics* 2016;12:4-8.
82. Geronikou SA, Pavlopoulou A, Cokkinos DV, Bacopoulou F, Chrousos GP. Polycystic ovary syndrome revisited: an interactions network approach. *Eur J Clin Invest* 2021;51(9):e13578.
83. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 2016:gkw943.
84. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42(D1):D1001-D6.
85. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genetics* 2017;136(6):665-77.
86. Agrawal S, Dimitrova N, Nathan P, Udayakumar K, Lakshmi SS, Sriram S, et al. T2D-Db: an integrated platform to study the molecular basis of Type 2 diabetes. *BMC Genomics* 2008;9(1):1-12.
87. Bai Z, Han G, Xie B, Wang J, Song F, Peng X, et al. AlzBase: an integrative database for gene dysregulation in Alzheimer's disease. *Mol Neurobiol* 2016;53(1):310-9.
88. Hutter C, Zenklusen JC. The cancer genome atlas: creating lasting value beyond its data. *Cell* 2018;173(2):283-5.
89. Taghvaei S, Sabouni F, Minuchehr Z. Evidence of Omics, Immune Infiltration, and Pharmacogenomic for SENP1 in the Pan-Cancer Cohort. *Front Pharmacol* 2021;12.
90. Saberi M, Minuchehr Z, Shahlaci M, Kheitan S. An in silico method to identify key proteins involved in the development of gastric cancer. *Res Med* 2017;41(3):199-209.
91. Islam MR, Ahmed ML, Paul BK, Bhuiyan T, Ahmed K, Moni MA. Identification of the core ontologies and signature genes of polycystic ovary syndrome (PCOS): A bioinformatics analysis.

- Inform Med Unlocked 2020;18:100304.
92. Liang Q-Q, Wang D-J. Bioinformatic Analysis Identifies Potential Key Genes in the Pathogenesis of Polycystic Ovary Syndrome. *J Bioinf Syst Biol* 2022;5:78-92.
93. Devarbhavi P, Telang L, Vastrad B, Tengli A, Vastrad C, Kotturshetti I. Identification of key pathways and genes in polycystic ovary syndrome via integrated bioinformatics analysis and prediction of small therapeutic molecules. *Reprod Biol Endocrinol* 2021;19(1):1-39.
94. Yang D, Li N, Ma A, Dai F, Zheng Y, Hu X, et al. Identification of potential biomarkers of polycystic ovary syndrome via integrated bioinformatics analysis. *Reprod Sci* 2021;28(5):1353-61.
95. Hossain MA, Al Amin M, Hasan MI, Sohel M, Ahammed MA, Mahmud SH, et al. Bioinformatics and system biology approaches to identify molecular pathogenesis of polycystic ovarian syndrome, type 2 diabetes, obesity, and cardiovascular disease that are linked to the progression of female infertility. *Inf Med Unlocked* 2022;30:100960.

COMPUTATIONAL BIOLOGY APPROACHES AND BIOINFORMATICS TO IDENTIFY KEY GENES IN POLYCYSTIC OVARY SYNDROME: A SYSTEMATIC REVIEW

Seyede Nafise Tabatabaei¹, Zarrin Minucheher^{*2}

Received: 15 November, 2022; Accepted: 08 April, 2023

Abstract

Background & Aims: Polycystic ovary syndrome (PCOS) is the most common female endocrine disease and often causes infertility in the women of reproductive age. This syndrome includes different genes and proteins, multiple pathways, and complex processes of hormone secretion. Therefore, a single factor cannot explain the pathogenesis of PCOS. Using computational biology and omics including genomics, transcriptomics, proteomics, and metabolomics, can provide faster and more effective methods for studying the pathogenesis of complex diseases such as PCOS.

Materials & Methods: In this study, to find the related studies, PubMed, Google Scholar, and Science Direct databases were searched without time constraints for 3 years using the keywords of "Polycystic Ovary Syndrome, Computational Biology, protein-protein interaction, Network Biology, and Pathways analysis".

Results: Various databases have been designed and made available to the public to repository data, Protein-Protein interactions, networks, and biological pathways related to humans. Three databases PCOSBase, PCOSKB, and PCOSDB have been explicitly developed for polycystic ovary syndrome.

Conclusion: Since the molecular mechanisms of polycystic ovary syndrome are still not completely understood, to realize this syndrome, besides experimental results using omics platforms, computational biology, and bioinformatics tools, it is necessary to identify the interaction between proteins and the pathways involved in it.

Keywords: Computational Biology, Network Biology, Pathway Analysis, Polycystic Ovarian Syndrome, Protein-Protein Interaction Analysis

Address: Department of Systems Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), Shahrak-e Pajoohesh, km 15, Tehran - Karaj Highway, Tehran, Iran,

Tel: 021-44787362

Email: dminucheher@gmail.com

SOURCE: STUD MED SCI 2023; 33(11): 785 ISSN: 2717-008X

Copyright © 2023 Studies in Medical Sciences

This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) which permits copy and redistribute the material just in noncommercial usages, as long as the original work is properly cited.

¹ Ph.D. Student in Molecular Genetics, Industry and Environment Biotechnology Research Institute, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

² Associate Professor of Department of Systems Biotechnology, Industry and Environment Biotechnology Research Institute, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran (Corresponding Author)