

Bioinformatic analysis of SNPs of *CD151* gene in epidermolysis bullosa |

Mahdis Kashani¹, Zahra Rezvani^{1*}

1- Department of Cell and Molecular Biology, Faculty of Chemistry, University of Kashan, Kashan, Isfahan, Iran

*Corresponding Author: Zahra Rezvani, Department of Cell and Molecular Biology, Faculty of Chemistry, University of Kashan, Kashan, Isfahan, Iran; Email: rezvani@kashanu.ac.ir

| Received | 2025/08/14

| Revised | 2026/01/28

| Accepted | 2026/06/22

| Abstract |

Background and Aim: Single Nucleotide Polymorphisms (SNPs) are common genetic variants implicated in human disorders. CD151, a tetraspanin component of the hemidesmosome, is essential for skin integrity, and its dysfunction can cause Epidermolysis Bullosa Simplex (EBS). This study aimed to predict high-risk SNPs in the *CD151* gene and evaluate their impact on the protein's stability, structure, and function using bioinformatics.

Materials and Methods: *CD151* SNPs were collected from dbSNP, and the protein sequence from UniProt. Comprehensive bioinformatics tools were used for effect prediction. Mutation risk and disease association were assessed using SIFT, PROVEAN, PolyPhen-2, PANTHER, and meta-predictors like Meta-SNP. Stability changes were predicted by I-Mutant and MUpro, and sequence conservation by ConSurf. Secondary/tertiary structures were modeled using NetSurfP-2.0, SOPMA, and HOPE, with protein-protein interactions analyzed by STRING and GeneMANIA.

Results: Screening 27 polymorphisms identified 8 high-risk variants: A81T, R87L, D147N, A145D, F153C, R178C, G227S, and G229S. Predictions showed these SNPs cause a reduction in protein stability, are predominantly located in conserved regions, and disrupt alpha-helical structures, thereby disturbing the overall protein fold. Altered surface positioning of mutated residues could also impair critical CD151 interactions.

Conclusion: The computational analyses and screenings demonstrated that 8 key *CD151* mutations potentially destabilize the protein and structure, altering its secondary structure and interactions. These SNPs have significant reference value for future studies. The identification of these high-risk SNPs paves the way for the design of targeted molecular screening panels for EBS patients, which could aid in earlier and more precise diagnosis. Final confirmation of these results necessitates further clinical and laboratory studies.

| Keywords |

Epidermolysis bullosa |
CD151 |
SNP |
Bioinformatic |

| Iau Science |

بررسی بیوانفورماتیکی پلی مورفیسم‌های تک‌نوکلئوتیدی (SNPs) ژن *CD151* در بیماری پروانه‌ای

مهدیس کاشانی^۱، زهرا رضوانی^{*}

۱- گروه زیست‌شناسی سلولی و مولکولی، دانشکده شیمی، دانشگاه کاشان، کاشان، اصفهان، ایران

*نویسنده مسئول: زهرا رضوانی، گروه زیست‌شناسی سلولی و مولکولی، دانشکده شیمی، دانشگاه کاشان، کاشان، اصفهان، ایران؛

پست الکترونیکی: rezvani@kashanu.ac.ir

تاریخ پذیرش | ۱۴۰۵/۰۴/۰۱

تاریخ ویرایش | ۱۴۰۴/۱۱/۰۸

تاریخ دریافت | ۱۴۰۴/۰۵/۲۳

چکیده

زمینه و هدف: آنتی ژن *CD151*، یک تتراسپاین در همی دسموزوم، نقشی کلیدی در حفظ یکپارچگی پوست دارد. اختلال در این پروتئین می‌تواند منجر به بیماری پروانه‌ای سیمپلکس (EBS) شود. هدف این مطالعه، پیش‌بینی SNPs پُرخطر ژن *CD151* و ارزیابی تأثیر آن‌ها بر پایداری، ساختار و عملکرد پروتئین با استفاده از رویکردهای بیوانفورماتیکی بود.

مواد و روش‌ها: SNPs ژن *CD151* از dbSNP و توالی پروتئین از UniProt استخراج شدند. ارزیابی خطر جهش و ارتباط با بیماری از طریق ابزارهای SIFT، PROVEAN، PolyPhen-2، PANTHER و روش‌های تجمعی مانند Meta-SNP انجام گرفت. تأثیر جهش‌ها بر پایداری پروتئین با I-Mutant و MUpro و حفاظت‌شدگی توالی با ConSurf بررسی شد. ساختار ثانویه و سه‌بعدی پروتئین با NetSurfP-2.0، SOPMA و HOPE و میان‌کنش‌های پروتئینی با STRING و GeneMANIA تحلیل شدند.

نتایج: از میان ۲۷ پلی مورفیسم، ۸ جهش به‌عنوان واریانت‌های پُرخطر شناسایی شدند: D147N، R87L، A81T، G229S و G227S، R178C، F153C، A145D. پیش‌بینی‌ها نشان دادند که این جهش‌ها باعث کاهش پایداری پروتئین، قرارگیری در نواحی حفاظت‌شده و اختلال در مارپیچ آلفا می‌شوند که ساختار کلی پروتئین را به هم می‌ریزد. همچنین، موقعیت سطحی یا غیرسطحی آمینواسیدهای جهش‌یافته می‌تواند تعامل‌های حیاتی پروتئین *CD151* با سایر مولکول‌ها را مختل سازد.

نتیجه‌گیری: تحلیل‌های مبتنی بر محاسبه‌ها و غربالگری‌های انجام‌شده نشان دادند که ۸ جهش کلیدی در ژن *CD151* به‌طور بالقوه بر پایداری و ساختار پروتئین *CD151* تأثیر گذاشته و ساختار ثانویه و تعامل‌های آن را دگرگون می‌سازند. این جهش‌ها دارای ارزش مرجعی مهمی برای مطالعه‌های آینده هستند. شناسایی این SNPs پُرخطر، مسیر را برای طراحی پِنل‌های غربالگری مولکولی هدفمند در بیماران EBS باز می‌نماید و می‌تواند به تشخیص زود هنگام و دقیق‌تر کمک کند. تأیید نهایی این نتایج مستلزم مطالعه‌های بالینی و آزمایشگاهی بیشتر است.

واژگان کلیدی

بیماری پروانه‌ای

پلی مورفیسم‌های تک‌نوکلئوتیدی

CD151

بیوانفورماتیک

علوم Iau

۱ | مقدمه

بیماری پروانه‌ای یا اپیدرمولیز بولوزا (EB)^۱ گروهی از بیماری‌های ژنتیکی موروثی مرتبط با شکنندگی پوست است که منجر به ایجاد تاول، فرسایش و زخم بر روی پوست و غشاهای مخاطی، در پاسخ به حداقل آسیب می‌شود (۱). تنوع بالایی ژن‌های مرتبط و جهش‌های بیماری‌زا در این بیماری می‌تواند منجر به انواع فنوتیپ‌های شدید و تظاهرات بالینی شود. این بیماری معمولاً توسط جهش در ژن‌های دخیل در ثبات پوست ایجاد می‌شود (۲). در حال حاضر ۲۰ ژن یافت شده است که رشته‌های میانی کراتین اسکلت سلولی، پروتئین‌های تماس سلولی (دسموزوم‌ها و همی‌دسموزوم‌ها) و مولکول‌های مسئول یکپارچگی مورفولوژیکی پوست به‌واسطه تماس‌های چسبندگی داخل اپیدرمی و ترکیب‌های کمپلکس نگهدارنده در اتصال پوستی اپیدرمی را کد می‌کنند (۳). بیماری پروانه‌ای از نظر ژنتیکی و بالینی بسیار متنوع است و به‌صورت اتوزومال غالب و مغلوب به ارث می‌رسد (۴). بروز این بیماری ۱/۴-۲۵ در هر میلیون تولد زنده و شیوع آن ۵۴-۲/۸۲ در هر میلیون جمعیت است. تقریباً ۵۰۰۰۰ نفر در جهان با بیماری پروانه‌ای زندگی می‌کنند (۵).

جدیدترین ژن دخیل در بیماری پروانه‌ای سیمپلکس *CD151* است که همراه با نفروپاتی است و نقص در آن موجب اختلال در سلول‌ها و ازدست‌دادن چسبندگی سلول به سلول بین کراتینوسیت‌های پایه می‌شود (۶). آنتی‌ژن *CD151* عضوی از خانواده پروتئین‌های سطح سلولی تتراسپانین، به‌عنوان جزئی از همی‌دسموزوم شناسایی شده است (۷). آنتی‌ژن *CD151* به‌عنوان یک تثبیت‌کننده عملکرد اینتگرین کمپلکس‌هایی را با اینتگرین $\alpha 3\beta 1$ و $\alpha 6\beta 4$ در سطح قاعده جانبی کراتینوسیت‌های غشای پایه سلول تشکیل می‌دهد (۸). *CD151* یک تنظیم‌کننده اصلی فرایند سیگنال‌دهی اینتگرین‌های متصل به لامینین است که برای تحرک سلول‌ها ضروری است. تحرک سلولی توسط *CD151* با فعال کردن *Rac* و *Cdc42* از طریق سیگنال‌دهی گیرنده اینتگرین و فاکتور رشد ارتقا می‌یابد (۹). علاوه بر این به‌دلیل ارتباط با فسفاتیدیل‌اینوزیتول-۴-کیناز نقش *CD151* در انتقال سیگنال و پیوند با اسکلت سلولی

پیشنهاد شده است. *CD151* توزیع سلولی و بافتی وسیعی دارد از جمله اپیتلیوم، اندوتلیوم، عضله، گلومرول‌های کلیوی و لوله‌های پیچیده نزدیک و دور، سلول‌های شوان و سلول‌های دندریتیک با یک گونه RNA در بیشتر بافت‌های بالغ انسان مشاهده شده است. این پروتئین یک پروتئین ۲۵۳ اسید آمینه‌ای است که حاوی یک حلقه خارج سلولی کوچک و یک حلقه بزرگ سیتوپلاسمی با دُمین‌های کربوکسی و آمینو کوتاه است. همچنین یک N-گلیکوزیلاسیون تشکیل می‌دهد و روی چندین ریزدوی سیستمین پالمیتویله می‌شود. ژن *CD151* انسانی روی بازوی کوتاه کروموزوم ۱۱ قرار دارد و شامل ۸ اگزون است که اگزون‌های ۲ تا ۸ آن پلی‌پپتید *CD151* را کد می‌کنند (۱۰ و ۱۱).

پلی‌مورفیسم‌های تکنوکلوتیدی (SNPs)^۲ یک تنوع تکنوکلوتیدی در یک مکان خاص در ژنوم هستند که طبق تعریف در بیش از یک درصد جمعیت یافت می‌شوند. SNPs زمانی که در مکان‌های تنظیمی یک ژن وجود داشته باشند می‌توانند بر سرعت رونویسی تأثیر بگذارند و موجب تغییراتی در تولید پروتئین کدگذاری شده شوند که می‌توانند تغییراتی در ساختار پروتئین و در نتیجه عملکرد ایجاد کنند که در نهایت منجر به ایجاد بیماری یا تغییر در پاسخ به دارو یا سم محیطی می‌گردد. از این رو SNPs به‌عنوان نشانگرهای مولکولی در بسیاری از مطالعه‌های ژنتیکی بیماری و فارماکولوژیک مورد استفاده قرار گرفته‌اند (۱۲).

در یک مطالعه ژنومی گسترده، DNA ۹۲ بیمار مشکوک به بیماری پروانه‌ای با استفاده از توالی‌یابی نسل جدید (NGS)^۳ هدفمند بر روی پنبلی شامل ۲۱ ژن مرتبط با شکنندگی پوست (از جمله *CD151*) بررسی شد. تحلیل‌های بیوانفورماتیک جامع منجر به شناسایی جهش‌های بیماری‌زا، از جمله در ژن *CD151*، گردید. این تحلیل‌ها، به‌ویژه در مورد پلی‌مورفیسم‌های تکنوکلوتیدی، برای پیش‌بینی بیماری‌زایی و پیامدهای عملکردی آن‌ها در بیماری پروانه‌ای بسیار حیاتی بودند (۱۳). پژوهش‌های پیشین عمدتاً بر روی جهش‌های بیماری‌زا در ژن *CD151* که منجر به بیماری پروانه‌ای می‌شوند متمرکز بوده‌اند (۱۴-۱۶). با این حال، مطالعه جامع پلی‌مورفیسم‌های غیربیماری‌زا و همچنین پلی‌مورفیسم‌هایی

³ Next Generation Sequencing

¹ Epidermolysis Bullosa

² Single-Nucleotide Polymorphism

با فراوانی آلی نادر کمتر از ۱ درصد و همچنین SNPهایی با داده‌های تکراری یا نامشخص حذف شدند و تنها SNPهایی که منجر به جایگزینی اسیدآمینه^۱ در توالی گذراندند پروتئین می‌شوند، برای تحلیل‌های بعدی گزینش شدند.

۲-۲ | پیش‌بینی اثرهای عملکردی و بیماری‌زایی

اثرهای SNP با استفاده از ابزارها و سرورهای آنلاین خاصی از جمله SIFT^۲، PROVEAN^۳، PANTHER و PolyPhen-2^۴ پیش‌بینی شد. SIFT (<https://sift.bii.a-star.edu.sg/>) با استفاده از هم‌ترازی چندگانه توالی‌های همولوگ، میزان حفاظت تکاملی هر موقعیت اسیدآمینه را ارزیابی کرده و تحمل‌پذیری جایگزینی‌های آمینواسیدی را پیش‌بینی می‌نماید. این ابزار قادر است جهش‌های درگیر در بیماری را از پلی‌مورفیسم‌های خنثی تشخیص دهد. خروجی این ابزار یک امتیاز بین ۰ تا ۱ است؛ به‌طوری‌که جایگزینی‌هایی با امتیاز کمتر از ۰/۰۵ را به‌عنوان مضر^۵ پیش‌بینی می‌کند (۱۹).

PROVEAN (http://provean.jcvi.org/seq_submit.php) نیز می‌تواند پیش‌بینی‌هایی را نه‌تنها برای جایگزینی‌های اسیدآمینه منفرد، بلکه برای چندین جایگزین، درج و حذف اسیدآمینه با استفاده از همان طرح امتیازدهی اساسی ایجاد کند. دقت نمردهی این سرور بر روی ۲/۵- تنظیم شده است. در صورتی‌که امتیاز از ۲/۵- کمتر باشد نشان‌دهنده مضر بودن جهش است (۲۰).

PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2>) نیز تأثیر احتمالی جایگزینی اسیدهای آمینه را بر ثبات و عملکرد پروتئین‌های انسانی با استفاده از ملاحظات تکاملی ساختاری و مقایسه‌ای پیش‌بینی می‌کند. پیش‌بینی این سرور امتیاز عددی از ۰/۰ (خوش‌خیم) تا ۱/۰ (مضر) است (۲۱). این ابزار به‌دلیل استفاده از داده‌های ساختاری و تکاملی به‌طور هم‌زمان، در مطالعه‌های بیماری‌های ژنتیکی انسانی، پزشکی شخصی، سرطان و تحلیل SNPهای نادر کاربرد بسیار گسترده‌ای دارد.

PANTHER (<https://www.pantherdb.org>) یک سیستم نرم‌افزاری جامع و رایگان در دسترس برای ارتباط دادن تکامل توالی پروتئین به تکامل عملکردهای پروتئین خاص و نقش‌های بیولوژیکی

که ممکن است به‌عنوان عوامل مستعد یا تعدیل‌کننده بیماری عمل کنند کمتر مورد توجه قرار گرفته است.

در این مطالعه، از ابزارهای مختلف بیوانفورماتیکی برای غربالگری SNPs بهره برده تا مضرترین آن‌ها را در پروتئین CD151، که نقش مؤثری در بیماری پروانه‌ای سیمپلکس دارد، شناسایی گردد. هدف نهایی این پژوهش نیز بررسی جامع تأثیر این واریانت‌های پُرخطر بر ساختار و عملکرد پروتئین CD151 است تا درک عمیق‌تری از خواص مولکولی آن‌ها به‌دست آید و زمینه‌ای برای ارائه اهداف و گزینه‌های تشخیصی و درمانی مؤثرتر فراهم گردد.

۲ | مواد و روش‌ها

۲-۱ | جمع‌آوری داده‌ها

توالی پروتئین CD151 در قالب FASTA از طریق پایگاه داده UniProtKB (<https://www.uniprot.org>) بارگیری شد. UniProtKB یک مخزن جامع و با کیفیت از توالی‌های پروتئینی است که همراه با حاشیه‌نویسی‌های عملکردی ارائه می‌شود و داده‌های آن به‌صورت مداوم با استفاده از استخراج اطلاعات از منابع علمی، سیستم‌های خودکار و روش‌های یادگیری ماشین به‌روزرسانی می‌گردد. این پایگاه علاوه بر فراهم کردن آخرین نسخه‌های پروتئوم‌های مرجع، دسترسی به ساختارهای پیش‌بینی‌شده AlphaFold و سایر ابزارهای تحلیلی را برای بررسی دقیق‌تر ویژگی‌های ساختاری و عملکردی پروتئین‌ها امکان‌پذیر می‌سازد (۱۷). در این مطالعه، توالی CD151 (شماره شناسه: P48509) از بخش پروتئین‌های بازبینی‌شده UniProtKB استخراج و برای تحلیل‌های بیوانفورماتیکی بعدی مورداستفاده قرار گرفت. SNPهای مرتبط با ژن CD151 از dbSNP (<https://www.ncbi.nlm.nih.gov/snp>) که زیرمجموعه‌ای از پایگاه داده NCBI است جمع‌آوری شد. dbSNP یکی از منابع اصلی و معتبر برای گردآوری و شناسایی انواع مختلف واریانت‌های ژنتیکی است و داده‌های آن به‌طور مداوم با نتایج پروژه‌های بزرگ ژنومی مانند ۱۰۰۰ Genomes و gnomAD به‌روزرسانی می‌شود (۱۸). در این مرحله SNPها

^۴ Polymorphism phenotyping v2

^۵ Deleterious

^۱ Non-Synonymous

^۲ Sorting intolerant from tolerant

^۳ Protein variation effect analyzer

SNVهای بیماری‌زا را با ترکیب برخی از روش‌های پیشرو در پیش‌بینی بیماری (PhD-SNP) و ارتباطات عملکرد (SIFT, PANTHER و SNAP) شناسایی می‌کند و خروجی را به صورت برچسب Disease/Neutral همراه با امتیاز احتمال ارائه می‌دهد (۲۵).

۲-۳ | بررسی پایداری پروتئین

SNPها تمایل دارند با کاهش یا افزایش پایداری پروتئین بر قدرت پروتئین تأثیر بگذارند. برای پیش‌بینی این اثرها از ۲ ابزار آنلاین I-Mutant و MUpro استفاده شد.

I-Mutant (<https://folding.biofold.org/i-mutant/i-mutant2.0.html>) با تکیه بر شبکه‌های عصبی، تغییر در انرژی آزاد گیبس تاشدگی ($\Delta\Delta G$) را پیش‌بینی می‌کند؛ مقدار $\Delta\Delta G$ منفی نشان‌دهنده کاهش پایداری و افزایش انرژی لازم برای تاشدگی پروتئین جهش‌یافته است (۲۶). هم‌زمان (<https://mupro.proteomics.ics.uci.edu>) MUpro که از الگوریتم‌هایی مانند بردارهای پشتیبان (SVM) استفاده می‌کند یک امتیاز پیوسته (بین $-1/0$ تا $+1/0$) ارائه می‌دهد که در آن امتیاز منفی نیز مؤید کاهش پایداری پروتئین است. پیش‌بینی کاهش پایداری توسط هر دو ابزار، قوی‌ترین شواهد محاسباتی برای بیماری‌زا بودن یک SNP است چراکه پروتئین‌های ناپایدار تمایل به تاخوردگی نادرست، تجزیه سریع‌تر توسط پروتئازوم‌ها و یا تجمع را دارند (۲۷). استفاده هم‌زمان از I-Mutant و MUpro، با فراهم کردن تأیید متقابل نتایج بر پایه مدل‌های محاسباتی متفاوت، دقت پیش‌بینی اثر مخرب SNPها بر پایداری ساختاری CD151 را به‌طور چشمگیری افزایش می‌دهد.

۲-۴ | حفاظت‌شدگی و نواحی عملکردی

ConSurf (<http://consurf.tau.ac.il>) ابزاری است که مناطق عملکردی و ساختاری ماکرومولکول‌ها را با ارزیابی تکاملی جایگزینی اسیدهای آمینه/نوکلئیک در میان توالی‌های همولوگ نشان می‌دهد. این ابزار حفاظت‌شدگی را به ۹ امتیاز تقسیم‌بندی می‌کند که امتیاز ۶ تا ۹ نقاط حفاظت‌شده محسوب می‌شوند و یک شاهد قوی تکاملی فراهم می‌کند که نشان می‌دهد آن جهش‌ها در واقع مخرب هستند؛ زیرا مکانیسم‌های طبیعی در طول تکامل، این موقعیت‌ها را از تغییر حفظ کرده‌اند (۲۸).

است. PSEP^۱ (حفظ تکاملی یک موقعیت خاص)، مدت‌زمان حفظ‌شدگی یک اسیدآمینه معین در دودمان منتهی به پروتئین (برحسب میلیون‌ها سال) را اندازه‌گیری می‌کند. هر چه یک موقعیت طولانی‌تر حفظ شود، احتمال تأثیر مخرب آن بیشتر می‌شود. اگر جهش بیشتر از ۴۵۰ میلیون سال باشد احتمال زیادی دارد که اثر مخرب داشته باشد. این ابزار از داده‌های تکاملی و درخت‌های فیلوژنتیک برای محاسبه امتیاز احتمال اثر مضر (Pdel)^۲ برای هر واریانت استفاده می‌کند. امتیاز Pdel یک مقدار پیوسته بین ۰ (احتمال کم) تا ۱ (احتمال زیاد) را نشان می‌دهد که براساس آن احتمال مضر بودن واریانت بر عملکرد پروتئین سنجیده می‌شود (۲۲). واریانت‌هایی که Pdel آن‌ها بزرگ‌تر از ۰/۵ بودند به‌عنوان مضر و واریانت‌هایی که Pdel آن‌ها کم‌تر از این آستانه بودند به‌عنوان احتمالاً خوش‌خیم طبقه‌بندی شدند.

سه ابزار آنلاین PhD-SNP, SNPs&GO و Meta-SNP برای تعیین ارتباط SNPهای گرینش‌شده با بیماری استفاده شد. PhD-SNP (<http://snps.biofold.org/phd-snp/phd-snp.html>) یک ابزار یادگیری ماشین مبتنی بر شبکه عصبی است که برای پیش‌بینی ارتباط بیماری‌ها با SNP غیرهم‌معنی (nsSNP) با دقت ۷۸ درصد استفاده می‌شود. این ابزار تنها براساس اطلاعات توالی پروتئین و ویژگی‌های فیزیکوشیمیایی اسیدهای آمینه تحلیل خود را انجام می‌دهد، بنابراین حتی برای پروتئین‌هایی که ساختار سه‌بعدی آن‌ها در دسترس نیست هم کاربرد دارد. طبقه‌بندی SNPها به بیماری‌زا یا خنثی آن‌ها را در مقیاس ۰ تا ۹ رتبه‌بندی می‌کند (۲۳).

SNPs&GO (<http://snps-and-go.biocomp.unibo.it/snps-and-go>)

روشی برای پیش‌بینی پلی‌مورفیسم‌های اسیدآمینه مضر (SAPs) با استفاده از حاشیه‌نویسی عملکردی پروتئین است. این سرور با ادغام ویژگی‌های فیزیکوشیمیایی و تکاملی توالی با اطلاعات ردمبندی کاربردی زن‌ها و استفاده از طبقه‌بندی بردارهای پشتیبان دقتی بیش از ۸۰ درصد ارائه می‌دهد. خروجی این ابزار شامل برچسب Disease/Neutral، امتیاز احتمال بیمارزایی و شاخص اطمینان است و از ابزارهای معتبر در شناسایی جهش‌های مضر محسوب می‌شود (۲۴).

Meta-SNP (<https://snps.biofold.org/meta-snp>) عملکرد

بهتری نسبت به بهترین پیش‌بینی‌کننده منفرد دارد. این سرور

² Probability of deleterious

¹ Position-specific evolutionary preservation

۲-۵ | پیش‌بینی ساختار ثانویه و حلالیت

NetSurfP-2.0 (<https://services.healthtech.dtu.dk/services/NetSurfP-2.0>)

ابزاری نوین و مبتنی بر شبکه‌های عصبی عمیق است که می‌تواند مهم‌ترین ویژگی‌های ساختاری یک پروتئین را با دقت و زمان اجرای بی‌سابقه پیش‌بینی کند. این ابزار با استفاده از یک مدل یکپارچه واحد می‌تواند محلولیت سطحی، ساختار ثانویه، اختلال ساختاری و زوایای Ψ و Φ (مانند Ψ) را به ازای هر رزیدو از توالی ورودی پیش‌بینی نماید (۲۹).

SOPMA (https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html)

یک ابزار بیوانفورماتیکی کلاسیک و پرکاربرد برای پیش‌بینی ساختار ثانویه پروتئین‌ها براساس خودبهبه‌سازی هلیکس‌ها است. این ابزار با استفاده از الگوریتم‌های آماری و بهینه‌سازی تکراری، صفحات بتا و کویل‌ها را از روی توالی پروتئین پیش‌بینی می‌کند. دقت این روش ۶۳ درصد می‌باشد (۳۰).

۲-۶ | تحلیل ساختار سه‌بعدی

HOPE (<https://www3.cmbi.umcn.nl/hope>) یک ابزار

تجزیه و تحلیل خودکار برای جهش‌های پروتئین می‌باشد. این ابزار اطلاعات گسترده‌ای را که عبارت‌اند از: WHAT IF برای محاسبات ساختاری بر روی فایل PDB یا یک مدل همسانی که توسط YASARA ساخته شده است، HSSP برای امتیازهای حفاظتی، سرورهای DAS برای پیش‌بینی‌های مبتنی بر توالی و Uniprot برای حاشیه‌نویسی‌های توالی جمع‌آوری می‌کند. در نهایت HOPE گزارشی را از تأثیر جهش که بر ساختار سه‌بعدی پروتئین تمرکز دارد ارائه می‌دهد (۳۱).

۲-۷ | میان‌کنش‌های ژن و پروتئین CD151 با دیگر

ژن‌ها و پروتئین‌ها

از دو سرور STRING و GeneMANIA برای بررسی میان‌کنش‌های ژن و پروتئین CD151 با ژن‌ها و پروتئین‌های دیگر استفاده شد. هدف پایگاه داده STRING جمع‌آوری و ادغام این اطلاعات، با ادغام داده‌های ارتباط پروتئین-پروتئین شناخته‌شده و پیش‌بینی‌شده برای تعداد زیادی از موجودات است. ارتباط‌ها در STRING (<https://string-db.org/>) شامل تعامل‌های مستقیم (فیزیکی) و همچنین تعامل‌های غیرمستقیم

(عملکردی) است، تا زمانی که هر دو خاص و از نظر بیولوژیکی معنادار باشند (۳۲). GeneMANIA (<https://genemania.org/>) یک وبسایت کاربرپسند انعطاف‌پذیر برای ایجاد فرضیه در مورد عملکرد ژن، تجزیه و تحلیل فهرست ژن‌ها و اولویت‌بندی آن‌ها برای سنجش عملکردی است. همچنین ژن‌های مشابه عملکردی را با استفاده از انبوهی از داده‌های ژنومیک و پروتئومیکس پیدا می‌نماید؛ علاوه بر این عملکرد ژن را نیز پیش‌بینی می‌کند (۳۳).

۳ | نتایج

۳-۱ | غربالگری چندابزاری و شناسایی واریانت‌های پرخطر

تحلیل ۲۷ SNP غیر مترادف ژن CD151 با چهار ابزار اصلی غربالگری آغاز شد که با استفاده از ابزار SIFT جهش‌های بدخیم P137L, L110V, R87L, K84R, A81T, V71I, A145D, F153C, H154Y, R171C, R178C شناسایی شدند که همگی امتیازی بین ۰ تا ۰/۰۵ داشتند. با استفاده از PROVEAN نیز جهش‌های بدخیم K84R, A81T, R87L, R171C, F153C, D147N, A145D, P137L, R178C, Q217P, G227S, G229S با امتیازی کمتر از ۲/۵- شناسایی شدند. سرور PolyPhen-2 ساختار و عملکرد پروتئین جهش‌یافته را با هم‌سویی ساختار و شباهت مشاهده‌شده در آن‌ها پیش‌بینی نمود. براین اساس ۱۰ جهش با احتمال خیلی زیاد در ایجاد آسیب شناسایی شدند که امتیازی بالاتر از ۰/۹۶۲ داشتند. این جهش‌ها عبارت‌اند از: R171C, F153C, D147N, A145D, R87L, A81T, V71I, R178C, G227S, G229S. علاوه بر این ۱۰ جهش، ۵ جهش E213A, H154Y, L110V, R87Q, S50T, E6K نیز با احتمال آسیب‌زایی کمتر شناسایی شدند. تجزیه و تحلیل بیوانفورماتیکی واریانت‌های پروتئین CD151 با استفاده از ابزار PANTHER نشان داد که بیشتر آن‌ها امتیاز بالای ۰/۵ داشتند. در میان آن‌ها S50T, A81T, L110V, P137L, D147N, F153C, Q217P و G227S امتیازی نزدیک به ۱ داشتند. علی‌رغم تنوع واریانت‌های اولیه در ۴ مجموعه داده مورد بررسی، تجزیه و تحلیل هم‌پوشانی نشان داد که تنها ۲ واریانت جایگزینی اسید آمینه، A81T و F153C به صورت

براساس ConSurf، به جز R178C باقی جهش‌های کلیدی امتیاز حفاظت‌شدگی بالای ۶ داشتند که بیانگر اهمیت تکاملی بالا و تأکید بر تأثیر مخرب آن‌هاست. همچنین آمینواسید آسپارتیک‌اسید در موقعیت ۱۴۷ به‌عنوان یک اسیدآمین به عملکرد کاتالیتیک یا اتصال لیگاند مشخص شد که آسیب‌زایی جهش D147N را از نظر عملکردی تأیید می‌کند (شکل ۱).

همچنین ۵۸/۵۰ درصد ساختار ثانویه پروتئین آنتی‌ژن CD151 طبق SOPMA از مارپیچ آلفا و ۲۵/۶۹ درصد آن از کوئل تشکیل شده بود. طبق NetSurfP-2.0 و SOPMA به جز جهش R178C که در ناحیه کوئل واقع شده بود، باقی جهش‌های پُرخطر در ناحیه هلیکس آلفا قرار داشتند. قرارگیری بیشتر جهش‌ها در ساختارهای ثانویه حیاتی (هلیکس آلفا) و حفظ‌شدگی بالای آن‌ها به‌شدت دلالت بر آن دارد که اختلالات آن‌ها به‌طور مستقیم ثبات ساختار سه‌بعدی و در نتیجه بیان عملکردی پروتئین CD151 در سطح سلول را تحت تأثیر قرار می‌دهد. قرارگیری جهش R178C در ناحیه کوئل/لوپ نشان می‌دهد که تأثیر آن شاید موضعی‌تر و مربوط به انعطاف‌پذیری و تعامل‌های سطحی است (شکل ۲ و ۳).

مداوم در تمام مجموعه‌ها تکرار شده بودند و واریانت‌های R87L، R171C، R178C، P137L، D147N، G227S و A145D در ۳ مجموعه مشترک بودند (جدول ۱). در تأیید بیماری‌زایی، ابزارهای PhD-SNP، SNPs&GO و Meta-SNP از بین ۲۷ جهش، ۴ جهش R87L، A145D، D147N و F153C در هر سه سرور بیماری‌زا اعلام و ۷ جهش A81T، K84N، R178C، Q217P، G227S و G229S در ۲ سرور بیماری‌زا اعلام شدند (جدول ۲).

۳-۲ | تأثیر بر پایداری پروتئین و ساختار ثانویه

برای بررسی پیامدهای ساختاری، پایداری پروتئین CD151 جهش‌یافته با استفاده از سرورهای I-Mutant و MUpro ارزیابی شدند. جهش‌هایی که در هر دو سرور کاهش پایداری (انرژی آزاد گیبس منفی در I-Mutant) را نشان دادند به‌عنوان SNPهای بیماری‌زای پُرخطر با احتمال کاهش ثبات پروتئین و افزایش تخریب در نظر گرفته شدند. با اشتراک‌گیری اثرهای آسیب‌زایی، بیماری‌زایی و پایداری پروتئین، ۸ جهش کلیدی A81T، R87L، A145D، D147N، F153C، R178C، G227S و G229S برای تحلیل ساختاری عمیق‌تر انتخاب شدند (جدول ۳).

جدول ۱- امتیازهای به‌دست‌آمده از SIFT، PROVEAN، PANTHER و PolyPhen-2.

SNP ID	Mutation	SIFT	PROVEAN	PANTHER	PolyPhen-2
rs371206840	E6K	0.147	-1.070	0.570	0.954
rs199816860	K8T	0.123	-2.413	0.570	0.005
rs376496318	S50T	0.439	-1.789	0.740	0.549
rs147298858	V71I	0.034	-0.765	0.570	0.988
rs141569440	A81T	0.028	-2.775	0.740	0/997
rs150909098	K84R	0.018	-2/852	0.570	0.186
rs116211118	K84N	0.928	-0.385	0.570	0.018
rs377680693	R86H	0.413	-0.572	0.500	0.012
rs200217814	R87L	0.026	-5.193	0.570	0.992
rs200217814	R87Q	0.295	-2.397	0.570	0.941
rs370471055	L102V	0.315	-1.977	0.500	0.209
rs146079952	L110V	0.013	-2.113	0.740	0.688
rs34215390	T120M	0.076	-1.648	0.190	0.337
rs111956783	D128G	0.392	-1.612	0.500	0.001
rs374834878	P137L	0.045	-3.961	0.740	0.002
rs11828133	A145D	0.006	-4.360	0.570	0.999
rs372034658	D147N	0.286	-3.828	0.850	0.994
rs375264131	F153C	0.002	-6.141	0.850	1.000
rs148880760	H154Y	0.008	-1.119	0.500	0.723

rs139042921	R171C	0.044	-3.533	0.500	0.962
rs367700291	R178C	0.013	-4.438	0.570	1.000
rs200052049	E213A	0.268	-1.629	0.500	0.877
rs200997452	Q217P	0.216	-2.920	0.740	0.072
rs56125304	R221S	0.391	-0.218	0.500	0.124
rs201597386	G227S	0.076	-4.570	0.850	0.991
rs201597386	G229S	0.071	-3.920	0.570	0.993

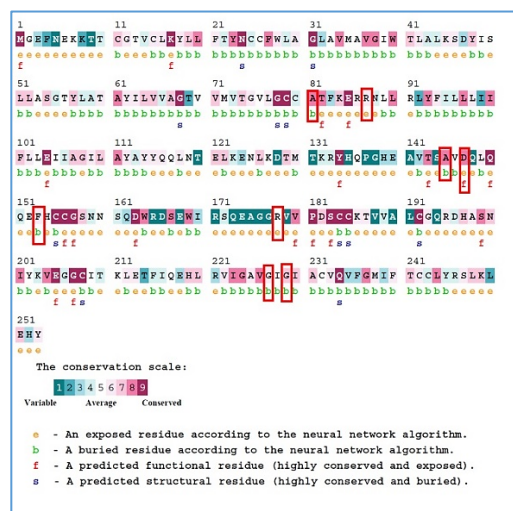
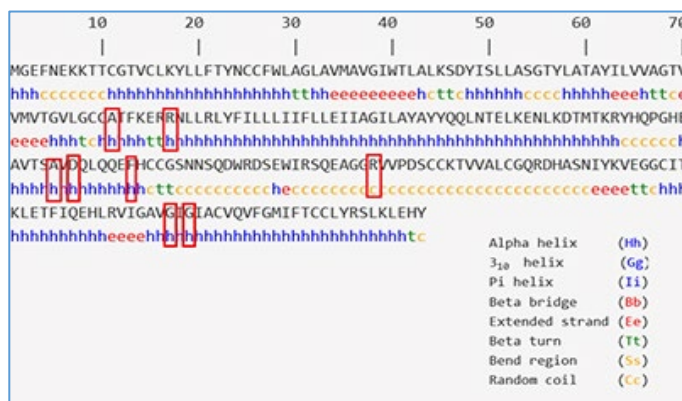
جدول ۲- نتایج بررسی اثر بیماری‌زایی توسط PhD-SNP، SNPs&GO و Meta-SNP.

SNP ID	Mutation	PhD-SNP	SNPs&GO	Meta-SNP
rs371206840	E6K	Neutral	Neutral	Neutral
rs199816860	K8T	Neutral	Neutral	Neutral
rs376496318	S50T	Neutral	Neutral	Neutral
rs147298858	V71I	Disease	Neutral	Neutral
rs141569440	A81T	Disease	Neutral	Disease
rs150909098	K84R	Neutral	Neutral	Neutral
rs116211118	K84N	Disease	Neutral	Disease
rs377680693	R86H	Neutral	Neutral	Neutral
rs200217814	R87L	Disease	Disease	Disease
rs200217814	R87Q	Neutral	Neutral	Neutral
rs370471055	L102V	Neutral	Neutral	Neutral
rs146079952	L110V	Neutral	Neutral	Neutral
rs34215390	T120M	Neutral	Neutral	Neutral
rs111956783	D128G	Neutral	Neutral	Neutral
rs374834878	P137L	Neutral	Neutral	Neutral
rs11828133	A145D	Disease	Disease	Disease
rs372034658	D147N	Disease	Disease	Disease
rs375264131	F153C	Disease	Disease	Disease
rs148880760	H154Y	Neutral	Neutral	Disease
rs139042921	R171C	Neutral	Disease	Neutral
rs367700291	R178C	Neutral	Disease	Disease
rs200052049	E213A	Neutral	Neutral	Neutral
rs200997452	Q217P	Disease	Neutral	Disease
rs56125304	R221S	Neutral	Neutral	Neutral
rs201597386	G227S	Disease	Disease	Neutral
rs201597386	G229S	Disease	Neutral	Disease

جدول ۳- نتایج بررسی اثر جهش بر پایداری پروتئین

SNP ID	Mutation	I-mutant	DDG (Kcal/mol)	Mupro	DDG (Kcal/mol)
rs371206840	E6K	Decrease Stability	-0.56	Decrease Stability	-1.93
rs199816860	K8T	Increase Stability	0.52	Decrease Stability	-0.72
rs376496318	S50T	Decrease Stability	-0.69	Decrease Stability	-1.31
rs147298858	V71I	Decrease Stability	-1.18	Decrease Stability	-0.17
rs141569440	A81T	Decrease Stability	-0.75	Decrease Stability	-0.60
rs150909098	K84R	Decrease Stability	-0.29	Increase Stability	0.14
rs116211118	K84N	Increase Stability	0.20	Increase Stability	0.02
rs377680693	R86H	Decrease Stability	-0.67	Decrease Stability	-1.09
rs200217814	R87L	Decrease Stability	-0.43	Decrease Stability	-0.16

rs200217814	R87Q	Decrease Stability	-0.44	Decrease Stability	-0.97
rs370471055	L102V	Decrease Stability	-0.84	Decrease Stability	-1.15
rs146079952	L110V	Decrease Stability	-1.41	Decrease Stability	-0.40
rs34215390	T120M	Increase Stability	0.02	Increase Stability	0.07
rs111956783	D128G	Decrease Stability	-0.19	Decrease Stability	-1.24
rs374834878	P137L	Increase Stability	0.20	Increase Stability	0.09
rs11828133	A145D	Decrease Stability	-0.38	Decrease Stability	-0.63
rs372034658	D147N	Increase Stability	1.54	Decrease Stability	-0.50
rs375264131	F153C	Decrease Stability	-0.86	Decrease Stability	-1.38
rs148880760	H154Y	Increase Stability	1.72	Decrease Stability	-0.29
rs139042921	R171C	Decrease Stability	-0.53	Decrease Stability	-0.71
rs367700291	R178C	Decrease Stability	-0.58	Decrease Stability	-0.42
rs200052049	E213A	Decrease Stability	-0.73	Decrease Stability	-0.78
rs200997452	Q217P	Decrease Stability	-1.13	Decrease Stability	-0.88
rs56125304	R221S	Decrease Stability	-1.22	Decrease Stability	-1.09
rs201597386	G227S	Decrease Stability	-1.16	Decrease Stability	-0.26
rs201597386	G229S	Decrease Stability	-1.17	Decrease Stability	-0.33



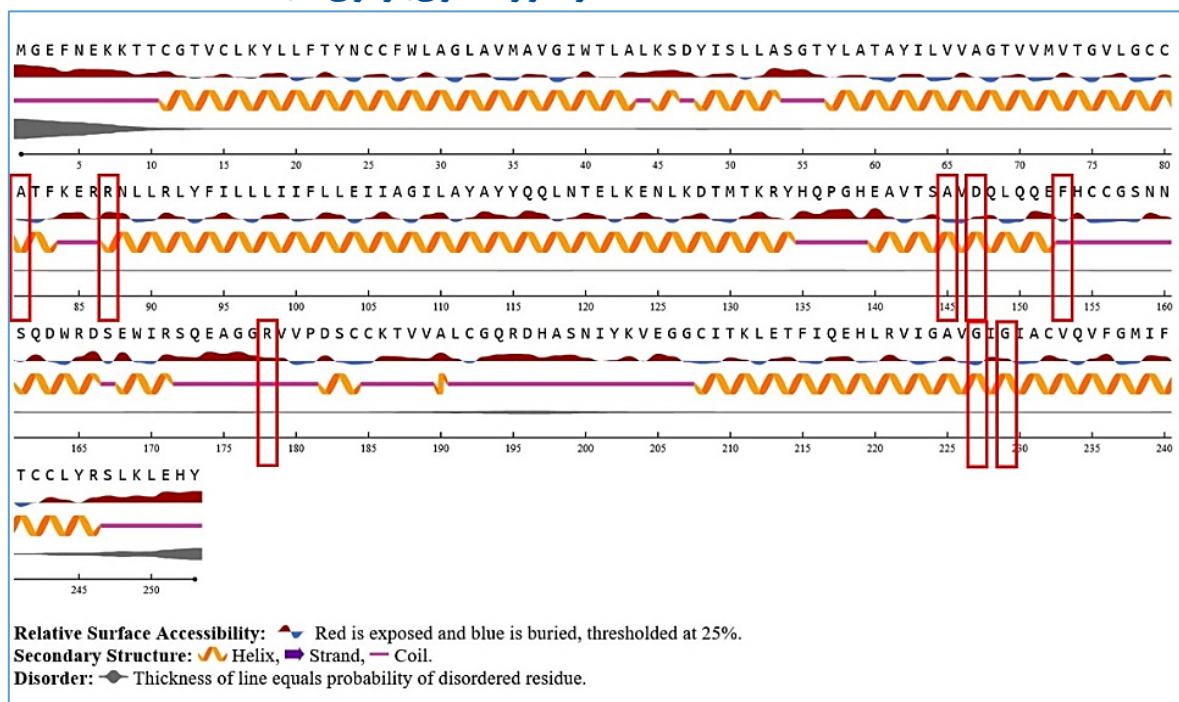
شکل ۲- ساختار ثانویه به دست آمده از SOPMA که در آن هلیکس‌ها به رنگ آبی و با حرف h مشخص شده‌اند و کویل‌ها به رنگ نارنجی با حرف c نشان داده شده‌اند. تنها جهش R178C در ناحیه کویل قرار دارد و باقی جهش‌های مشخص شده در ناحیه هلیکس آلفا قرار دارند.

شکل ۱- تعیین اسیدهای آمینه حفاظت شده در بین گونه‌ها در آنتی ژن CD151 با استفاده از ConSurf.

۳-۳ | تحلیل عملکردی جهش‌ها و اثر آن‌ها بر بیان پروتئین

اسید آمینه کوچک و به شدت آبگریز است، در حالی که ترئونین حجیم‌تر بوده و دارای یک گروه هیدروکسیل قطبی است. این جایگزینی موجب بروز ناسازگاری فضایی در هسته مارپیچ می‌شود و میزان آب‌گریزی در ناحیه‌ای که انتظار می‌رود به سمت مرکز آبگریز پروتئین باشد را کاهش می‌دهد.

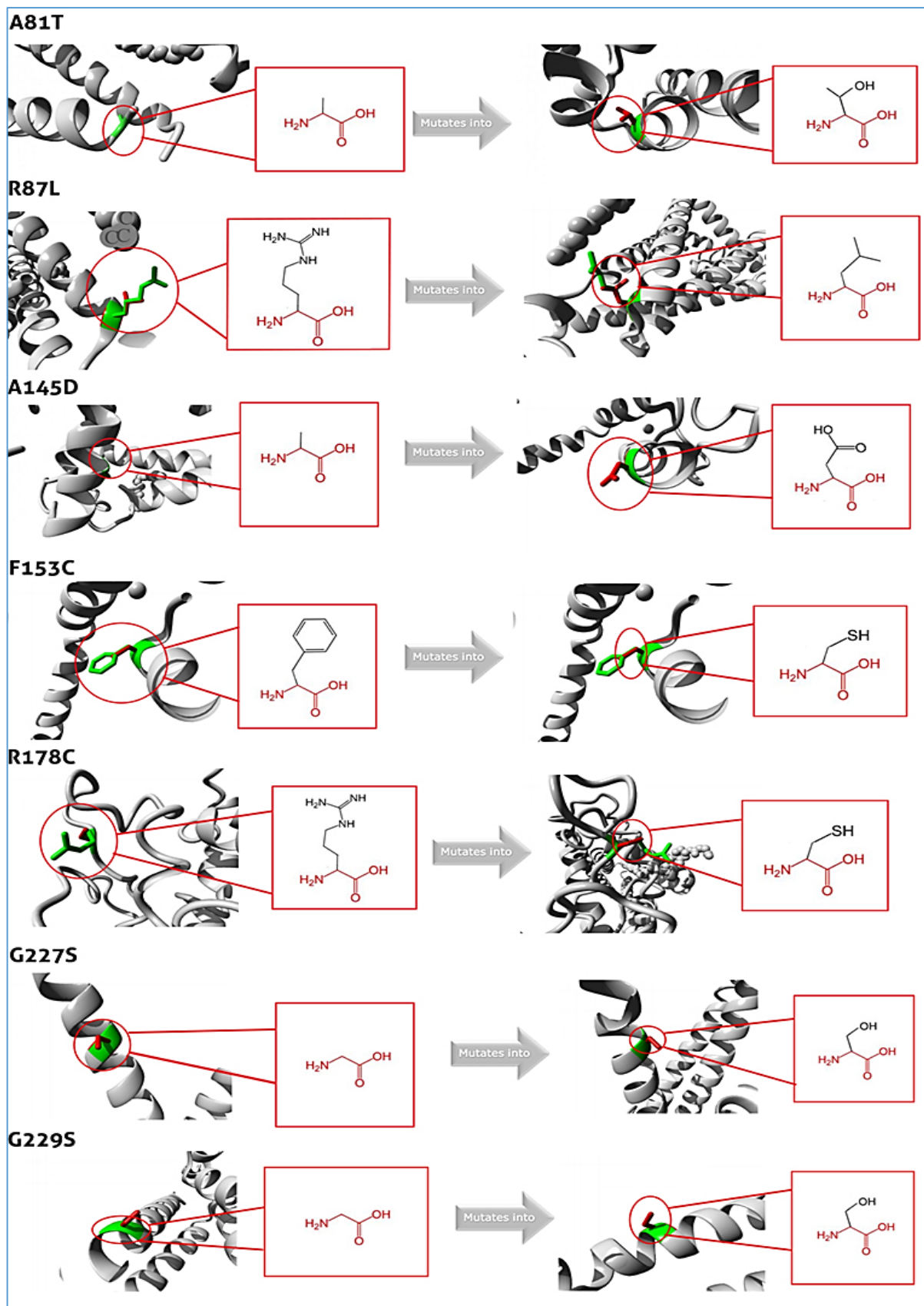
بررسی جزئیات تغییرهای ساختاری توسط ابزار HOPE و سایر پیش‌بینی‌های ساختاری، مکانیسم احتمالی آسیب‌رسانی این ۸ واریانت پُرخطر و در نتیجه کاهش عملکرد و بیان غشایی پروتئین CD151 را روشن ساخت. براساس گزارش‌های بیان شده از HOPE جهش A81T منجر به جایگزینی آلانین با ترئونین در یک مارپیچ آلفای حیاتی می‌شود. آلانین یک



شکل ۳- ساختار ثانویه استخراج شده از NetSurfP-2.0 که علاوه بر مشخص کردن هلیکس‌ها و کوئل‌ها دسترسی سطحی نسبی را نیز مشخص می‌کند.

قوی با لیگاند‌ها یا رزیدوهای هم‌بار شود و نیز پایداری ساختار آلفا را به‌خطر بیندازد که در نتیجه می‌تواند به‌طور مستقیم در فعالیت و اتصال پروتئین تأثیرگذار باشد. جهش D147N نیز جایگزینی آسپارتیک اسید با آسپارژین است. آسپارتیک اسید دارای بار منفی بوده درحالی‌که آسپارژین خنثی است. این جایگزینی باعث از دست رفتن بار منفی در یک موقعیت عملکردی کلیدی می‌شود. از دست رفتن بار در این موقعیت می‌تواند به‌طور مستقیم برهم‌کنش‌های حیاتی با سایر مولکول‌های دارای بار مثبت موردنیاز برای اتصال لیگاند یا فعالیت آن‌تیزن CD151 را تضعیف کرده یا از بین ببرد. این جهش به‌طور مستقیم پتانسیل عملکرد اتصال CD151 را کاهش می‌دهد. به‌علاوه جهش F153C جایگزینی فنیل‌آلانین با سیستئین در یک ناحیه بسیار حفاظت شده است.

این اختلالات، بسته‌بندی هسته پروتئین را تضعیف کرده و پتانسیل تاخوردگی نادرست در شبکه اندوپلاسمی و در نهایت تجزیه وابسته به یوبیکوئیتین را افزایش می‌دهد؛ در نتیجه بیان *CD151* عملکردی در سطح سلول کاهش چشمگیری می‌یابد. جهش R87L شامل جایگزینی آرژنین با لوسین است. آرژنین یک آمینواسید بزرگ با بار مثبت است که اغلب در تعامل‌های الکترواستاتیک یا پل نمکی شرکت می‌کند. لوسین نیز یک آمینواسید آبگریز و خنثی است. این تغییر باعث حذف بار مثبت حیاتی رزیدو و آبگریزتر شدن شدید ناحیه می‌شود. در صورتی‌که این ناحیه در سطح پروتئین یا مرز غشایی قرار داشته باشد، حذف بار مثبت می‌تواند بر تعامل با فسفولیپیدها یا مولکول‌های دارای بار منفی اثر گذاشته و تشکیل کمپلکس‌های CD151/اینترگرین را مختل سازد. جهش A145D جایگزینی آلانین با آسپارتیک اسید در یک ماریچ آلفاست؛ آسپارتیک اسید هم بزرگ‌تر بوده و هم دارای بار منفی است و تمایل کمتری برای حضور در ساختار آلفایی دارد. این تغییر می‌تواند موجب دفع الکترواستاتیک



شکل ۴- نمایشی تصویری از جایگاه اسید آمینه‌های جهش‌یافته در ساختمان سه‌بعدی پروتئین و تغییرهای ایجاد شده بعد از جهش توسط سرور HOPE، رنگ سبز اسید آمینه طبیعی را نشان می‌دهد و رنگ قرمز اسید آمینه جهش‌یافته را مشخص می‌نماید. همچنین برای جهش D147N شکلی یافت نشد.

خارج سلولی حاوی کلاژن نیز با ژن‌های *LAMB3*، *LAMA3*، *LAMC2*، *DST* و *COL7A1* ارتباط دارد. STRING نیز ارتباط‌های عمیقی را با ژن‌های *ITGA6*، *ITGB4*، *ITGA3*، *DST*، *PLEC*، *COL17A1* و *PTGFRN* نشان می‌دهد که با توجه به مطالعه‌های صورت گرفته بیشتر این ژن‌ها در بیماری پروانه‌ای نقش دارند.

۴ | بحث

ژن *CD151* که پروتئینی غشایی از خانوادهٔ تتراسپانین‌ها را کُد می‌کند، نقش گسترده‌ای در رشد، فعال‌سازی و تحریک سلول‌ها به‌ویژه در اپی‌تلیوم، اندوتلیوم، عضلات، گلومرول کلیه و ... ایفا می‌کند (۱۵). نقش کلیدی این ژن و جهش‌های آن در بروز بیماری پروانه‌ای توسط چندین مطالعهٔ آزمایشگاهی به اثبات رسیده است (۱۳-۱۶). باوجود این اهمیت، بررسی جامع و سیستماتیک واریانت‌های غیر مترادف *CD151* و اولویت‌بندی جهش‌های بالقوهٔ پاتوژنیک همچنان ناکافی بوده است؛ بیشتر پژوهش‌های پیشین تنها چند جهش محدود را با روش‌های تک‌ابزاری یا بدون تحلیل ساختاری دقیق ارزیابی کرده‌اند.

نوآوری اصلی این مطالعه در استفاده از یک رویکرد چندابزاری و یکپارچه‌شده برای تحلیل ۲۷ واریانت غیر مترادف *CD151* است. برخلاف مطالعه‌های پیشین که به‌صورت متداول، تنها از یک روش پیش‌بینی (نظیر SIFT یا PolyPhen-2) یا بررسی‌های آزمایشگاهی محدود استفاده کرده‌اند، این پژوهش با ادغام چهار ابزار معتبر پیش‌بینی عملکرد (SIFT، PROVEAN، PolyPhen-2 و PANTHER)، تحلیل پایداری ساختاری، مدل‌سازی سه‌بعدی و بررسی حفاظت تکاملی توانست یک نقشهٔ اولویت‌بندی دقیق از واریانت‌های پُرخطر ارائه دهد. چنین سطحی از ادغام داده‌ها در مطالعه‌های مرتبط با *CD151* تاکنون گزارش نشده و نشان‌دهندهٔ نوآوری روش‌شناختی مهم این پژوهش است.

نتایج تحلیل هم‌پوشانی نیز نشان داد که ۸ جهش *R87L*، *A81T*، *A145D*، *D147N*، *F153C*، *R178C* و *G227S* و *G229S* در مجموع بیشترین نشانه‌های بیماری‌زایی را دارند. از میان آن‌ها، ۳ جهش *R87L*، *A145D* و *F153C* توسط تمام ابزارهای بیوانفورماتیکی به‌عنوان واریانت‌های بسیار آسیب‌زا شناخته شدند؛ یافته‌ای که با گزارش‌های آزمایشگاهی وحیدزاده و همکاران در سال ۲۰۱۸ دربارهٔ حساسیت ساختاری نواحی EC2 پروتئین *CD151* نیز هم‌خوانی دارد (۱۳). تحلیل پایداری پروتئین نیز نشان داد که تقریباً تمام این جهش‌ها موجب کاهش پایداری ساختار پروتئین می‌شوند؛ موضوعی که با نقش

فنیل‌آلنین یک آمینواسید بزرگ و حلقوی با خاصیت آب‌گریزی است. همچنین سیستمین بسیار کوچک‌تر بوده و توانایی تشکیل پیوند دی‌سولفیدی را دارد. کاهش شدید حجم فضایی در یک ناحیهٔ بسیار حفاظت‌شده، بسته‌بندی ساختاری در آن نقطه را مختل می‌کند. رزیدوی F153 به‌احتمال در تماس فضایی حیاتی با مولکول‌ها یا حوزه‌های دیگر برای تثبیت ساختار یا تعامل‌های خارجی قرار دارد. این کاهش حجم می‌تواند منجر به از دست رفتن احتمالی تعامل‌های خارجی با شرکای تعاملی شود و به ساختار پروتئین آسیب برساند. جهش R178C جایگزینی آرژنین با سیستمین در ناحیهٔ کوئل پروتئین است. آرژنین دارای بار مثبت و سیستمین خنثی است. در این جهش، بار و اندازه تغییر پیدا می‌کند. از بین رفتن بار مثبت آرژنین در این موقعیت می‌تواند برهم‌کنش‌های پروتئین با سایر مولکول‌های دارای بار منفی را مختل کند. مهم‌تر از آن R178 در پروتئین *CD151* با آسپارتیک اسید در موقعیت ۱۹۶ (D196) پلی‌نمکی تشکیل می‌دهد. حذف بار مثبت R178، تشکیل این پلی‌نمکی حیاتی را مختل می‌کند و ثبات ساختاری موضعی و کلی پروتئین را به‌شدت کاهش می‌دهد. علاوه بر کاهش پایداری، کوچک‌تر شدن اندازه می‌تواند تعامل‌های خارجی را از بین ببرد. این اختلال ساختاری به‌احتمال منجر به کاهش شدید ثبات و بیان غیرعملکردی پروتئین *CD151* می‌شود. جهش‌های G227S و G229S هر دو با تغییر آمینواسید گلیسین به سرین که انعطاف‌پذیری بالایی دارد و فاقد زنجیرهٔ جانبی است، می‌توانند موجب از بین رفتن عملکرد به‌ویژه در نواحی با ساختار آزاد یا لوپ‌های نیازمند انعطاف (مانند کوئل‌ها) شوند. کاهش انعطاف‌پذیری می‌تواند پویایی پروتئین را مختل کند که برای تغییر شکل‌های انطباقی لازم برای تعامل‌های عملکردی و یا تسهیل لنگر انداختن آن در غشای سلولی حیاتی است. این جهش‌ها از طریق ایجاد ساختاریافتگی در یک ناحیهٔ انعطاف‌پذیر، عملکرد پروتئین را از بین می‌برند. شکل ۴ که برگرفته از HOPE می‌باشد، جهش‌ها را به‌صورت تصویری از ساختار سه‌بعدی پروتئین توصیف می‌کند.

۳-۴ | تحلیل میان‌کنش‌ها و ارتباط با بیماری

با بررسی میان‌کنش ژن *CD151* با سایر ژن‌ها توسط GeneMANIA مشخص شد که با ژن‌های *ITGA6*، *LAMB3*، *LAMA3*، *LAMC2*، *KRT5*، *COL17A1*، *PLEC*، *ITGB4*، *KRT14* و *DST* در اتصال‌های سلولی و سازمان‌دهی آن ارتباط دارد. همچنین در ماتریکس

جهش‌ها می‌توانند در مشاوره ژنتیکی خانوادگی، PGD/PND^۶ و ارزیابی خطر انتقال بیماری مورد استفاده قرار گیرند. از لحاظ پژوهش‌های تجربی نیز این مجموعه از جهش‌ها اهداف مناسبی برای مطالعه‌های عملکردی، طراحی مدل‌های حیوانی با CRISPR-Cas9 و توسعه درمان‌های مبتنی بر جایگزینی یا ویرایش ژن فراهم می‌نمایند. با وجود قوت‌های روش‌شناختی محدودیت‌هایی نیز وجود دارد: پیش‌بینی‌های بیوانفورماتیکی به‌تنهایی قطعی نیستند و نیازمند تأیید آزمایشگاهی هستند؛ تحلیل‌های ساختاری به مدل‌سازی همولوژی وابسته‌اند و پایگاه‌های داده ژنتیکی ممکن است ناقص باشند. باوجود این محدودیت‌ها، رویکرد چندلایه مطالعه تصویر جامعی از واریانت‌های مهم *CDI51* و پیامدهای بالقوه آن‌ها ارائه کرده و یک چارچوب عملی برای اولویت‌بندی جهش‌ها در پژوهش‌های آزمایشگاهی، تشخیصی و درمانی آینده فراهم می‌نماید.

۵ | نتیجه‌گیری

در این مطالعه با تجزیه و تحلیل‌های بیوانفورماتیکی عملکردی و ساختاری، SNP‌های مرتبط با ژن انسانی *CDI51*، ۸ جهش کلیدی شامل A81T، G229S، G227S، R178C، F153C، D147N، A145D، R87L شناسایی شد که به‌عنوان مارکرهای ژنتیکی بالقوه برای غربالگری و تشخیص زود هنگام بیماران مشکوک به انواع EB مرتبط با *CDI51* خصوصاً EBS قابل استفاده هستند. با وجود ارزش تحلیلی این مطالعه، تأیید تجربی اثر واریانت‌ها نیز ضروری است. آزمایش‌هایی مانند ارزیابی چسبندگی سلولی، پایداری پروتئین و ایجاد مدل‌های سلولی یا حیوانی می‌تواند نقش واقعی این جهش‌ها را روشن‌تر نماید. همچنین، بررسی ژن‌های دیگر مرتبط با چسبندگی اپیتلیال و مجموعه اینتگرین-تراسپانین می‌تواند درک جامع‌تری از مکانیزم‌های مولکولی EB فراهم سازد. در مجموع، این نتایج مبنایی برای اولویت‌بندی جهش‌ها و هدایت مطالعه‌های آینده در حوزه‌های عملکردی، ژنتیکی و درمانی ایجاد می‌نماید.

۶ | ملاحظات اخلاقی

این مطالعه نیاز به ملاحظات اخلاقی ندارد.

کلیدی پایداری تراسپانین‌ها در تشکیل کمپلکس‌های چسبندگی هم‌خوان است. هرچند جهش D147N اثر کمتری بر پایداری ایجاد کرد، اما حفاظت تکاملی بالا و تغییر در بار الکتریکی موضع جهش، احتمال اثرگذاری عملکردی آن را تقویت می‌کند.

مقایسه نتایج این پژوهش با مطالعه‌های پیشین نیز نشان می‌دهد که مطالعاتی مانند گزارش کارماتیک کرو و همکاران^۱ در سال ۲۰۰۴، تنها یک یا دو جهش را در ارتباط با بیماران EB شناسایی کردند (۳۴) و مطالعه‌های حیوانی ساچز و همکاران^۲ (۲۰۰۶) و نیلور و همکاران^۳ (۲۰۲۲) که اثر جهش پروتئین *CDI51* و رابط *G* آن با اینتگرین $\alpha 3\beta 1$ را تنها بر روی نروپاتی کلیوی بررسی کرده است (۳۵ و ۳۶)؛ مطالعه حاضر برای نخستین بار مجموعه کمابیش گسترده‌ای از SNP‌های *CDI51* را به‌طور هم‌زمان از منظر عملکردی، ساختاری و تکاملی در ایجاد EBS ارزیابی نموده است. این موضوع باعث شده بتوان مجموعه‌ای از جهش‌های کلیدی را با دقت بیشتر برای بررسی‌های آزمایشگاهی و مدل‌های حیوانی آینده اولویت‌بندی نمود.

به‌علاوه محل قرارگیری جهش‌ها نیز اهمیت قابل توجهی دارد. بیشتر واریانت‌های پُرخطر در نواحی مارپیچ آلفا یا مناطق حفاظت‌شده خارج سلولی قرار داشتند؛ نواحی‌ای که نقش مستقیمی در اتصال *CDI51* به پروتئین‌های لامینین، اینتگرین $\alpha 3\beta 1$ و تشکیل همی‌دسموزوم‌ها دارند. این یافته با گزارش‌های مولدر و همکاران^۴ (۲۰۱۹) و بردیتچوسکی و همکاران^۵ (۲۰۰۹) همسو است که نشان داده‌اند جهش در این نواحی می‌تواند انسجام ساختاری پوست را مختل کرده و باعث تاول‌زایی شود (۳۷ و ۳۸). تحلیل حفاظت تکاملی نیز نشان داد که چند جهش در نواحی بسیار حفاظت‌شده قرار دارند؛ موضوعی که احتمال اثرگذاری عملکردی آن‌ها را دوچندان می‌نماید. یافته‌های این پژوهش از نظر کاربردی نیز اهمیت بالایی دارند. از منظر تشخیص ژنتیکی، جهش‌های اولویت‌بندی‌شده می‌توانند به‌عنوان کاندیدهای قوی برای پانل‌های NGS مورد استفاده قرار گیرند و امکان غربالگری دقیق‌تر بیماران مشکوک به انواع EB مرتبط با *CDI51* را فراهم کنند. از نظر بالینی، شناسایی این جهش‌ها می‌تواند در پیش‌بینی شدت بیماری، احتمال درگیری کلیوی یا بروز عوارض سیستمیک نقش‌آفرین باشد. در حوزه ژنتیک پزشکی نیز این

^۴ Molder et al.

^۵ Berdichevski et al.

^۶ Preimplantation Genetic Diagnosis/Prenatal Diagnosis

^۱ Karamatic Crew et al.

^۲ Sachs et al.

^۳ Naylor et al.

۷ | تشکر و قدردانی

بدین وسیله از تمامی افرادی که با ارائه مشاوره‌های علمی و آسان‌سازی در دسترسی به منابع پژوهشی، در پیشبرد این پژوهش مشارکت داشتند سپاسگزاری می‌شود. همچنین از تمامی پژوهشگرانی که آثار علمی آن‌ها مورد استفاده قرار گرفته است تشکر می‌نمایند.

۸ | تعارض منافع

نویسندگان این مقاله اعلام می‌دارند که هیچ‌گونه وابستگی مالی، تجاری یا منافع شخصی که بتواند بر نتایج یا روند انجام این پژوهش تأثیر بگذارد وجود ندارد.

۹ | سهم نویسندگان

مهدیس کاشانی: مسئولیت جمع‌آوری داده‌ها، تحلیل‌های آماری و نگارش پیش‌نویس اولیه مقاله.

دکتر زهرا رضوانی: مسئولیت ایده‌پردازی، طراحی پژوهش و نظارت بر کل پروژه.

در نهایت تمامی نویسندگان در بازبینی و تأیید نسخه نهایی متن مقاله مشارکت داشته‌اند.

۱۰ | کد اخلاق

این پژوهش که مستخرج از ابزارهای رایانه‌ای زیستی است نیاز به کد اخلاق ندارد.

۱۱ | استفاده از هوش مصنوعی

در فرایند تهیه این مقاله، نویسندگان از هوش مصنوعی Gemini به منظور ویرایش زبانی و بهبود نگارش استفاده کرده‌اند. پس از بهره‌گیری از این ابزار، تمامی محتوای تولیدشده توسط نویسندگان مورد بازبینی و ویرایش قرار گرفته و مسئولیت کامل صحت، اصالت و محتوای نهایی مقاله بر عهده نویسندگان است.

| **Extended Abstract****Bioinformatic analysis of SNPs of CD151 gene in epidermolysis bullosa**

Background and Aim: CD151, a tetraspanin component of the hemidesmosome, is essential for skin integrity, and its dysfunction can cause Epidermolysis Bullosa Simplex (EBS). This study aimed to predict high-risk SNPs in the CD151 gene and evaluate their impact on the protein's stability, structure, and function using bioinformatics.

Materials and Methods: *CD151* SNPs were collected from dbSNP, and the protein sequence from UniProt. Comprehensive bioinformatics tools were used for effect prediction. Mutation risk and disease association were assessed using SIFT, PROVEAN, PolyPhen-2, PANTHER, and meta-predictors like Meta-SNP. Stability changes were predicted by I-Mutant and MUpro, and sequence conservation by ConSurf. Secondary/tertiary structures were modeled using NetSurfP-2.0, SOPMA, and HOPE, with protein-protein interactions analyzed by STRING and GeneMANIA.

Results: Screening 27 polymorphisms identified 8 high-risk variants: A81T, R87L, D147N, A145D, F153C, R178C, G227S, and G229S. Predictions showed these SNPs cause a reduction in protein stability, are predominantly located in conserved regions, and disrupt alpha-helical structures, thereby disturbing the overall protein fold. Altered surface positioning of mutated residues could also impair critical CD151 interactions.

Conclusion: The computational analyses and screenings demonstrated that 8 key *CD151* mutations potentially destabilize the protein and structure, altering its secondary structure and interactions. These SNPs have significant reference value for future studies. The identification of these high-risk SNPs paves the way for the design of targeted molecular screening panels for EBS patients, which could aid in earlier and more precise diagnosis. Final confirmation of these results necessitates further clinical and laboratory studies.

| **Keywords**

Epidermolysis bullosa, *CD151*, SNP, Bioinformatic

1. Mariath LM, Santin JT, Schuler-Faccini L, Kiszewski AE. Inherited epidermolysis bullosa: update on the clinical and genetic aspects. *A Bras Dermatol*. 2020;95(5):551-69. doi: 10.1016/j.abd.2020.05.001
2. Kocher T, Koller U. Advances in gene editing strategies for epidermolysis bullosa. *Prog Mol Biol Transl Sci*. 2021;182:81-109. doi: 10.1016/bs.pmbts.2020.12.007
3. Ryumina I, Goryunov K, Silachev D, Shevtsova YA, Babenko V, Marycheva N, et al. Pathogenetic Therapy of Epidermolysis Bullosa: Current State and Prospects. *Bull Exp Biol Med*. 2021;171(1):109-21. doi: 10.1007/s10517-021-05182-8
4. Kotalevskaya YY, Stepanov V. Molecular genetic basis of epidermolysis bullosa. *Vavilov J Genet Breed*. 2023;27(1):18-27. doi: 10.18699/VJGB-23-04
5. Hou P-C, Wang H-T, Abhee S, Tu W-T, McGrath JA, Hsu C-K. Investigational treatments for epidermolysis bullosa. *Am J Clin Dermatol*. 2021;22:801-17. doi: 10.1007/s40257-021-00626-3
6. Bardhan A, Bruckner-Tuderman L, Chapple IL, Fine J-D, Harper N, Has C, et al. Epidermolysis bullosa. *Nat Rev Dis Primers*. 2020;6(1):78. doi: 10.1038/s41572-020-0210-0
7. Todorović V, Kligys K, Dusek R, Jones J, Green KJ. Desmosomes and Hemidesmosomes. *Encyclopedia of Biological Chemistry: Second Edition*: Elsevier Inc; 2013. p. 636-43. doi: 10.1016/B978-0-12-378630-2.00472-2
8. Bruckner-Tuderman L, Has C. Disorders of the cutaneous basement membrane zone—the paradigm of epidermolysis bullosa. *Matrix Biol*. 2014;33:29-34. doi: 10.1016/j.matbio.2013.07.007
9. Gorain B, Bhattamishra SK, Choudhury H, Nandi U, Pandey M, Kesharwani P. Overexpressed receptors and proteins in lung cancer. *Nanotechnology-Based Targeted Drug Delivery Systems for Lung Cancer*: Elsevier; 2019. p. 39-75. doi: 10.1016/B978-0-12-378630-2.00472-2
10. Karamatic Crew V, Burton N, Kagan A, Green CA, Levene C, Flinter F, et al. CD151, the first member of the tetraspanin (TM4) superfamily detected on erythrocytes, is essential for the correct assembly of human basement membranes in kidney and skin. *Blood*. Academic Press. 2004;104(8):2217-23. doi: 10.1016/B978-0-12-815720-6.00003-4
11. Sterk LMT, Geuijen CA, Oomen LC, Calafat J, Janssen H, Sonnenberg A. The tetraspan molecule CD151, a novel constituent of hemidesmosomes, associates with the integrin $\alpha 6 \beta 4$ and may regulate the spatial organization of hemidesmosomes. *J Cell Biol*. 2000;149(4):969-82. doi: 10.1083/jcb.149.4.969
12. Kim S, Misra A. SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng*. 2007;9:289-320. doi: 10.1146/annurev.bioeng.9.060906.151927
13. Vahidnezhad, H., Youssefian, L., Saeidian, A. H., Mahmoudi, H., Touati, A., Abiri, M., ... & Uitto, J. Recessive mutation in tetraspanin CD151 causes Kindler syndrome-like epidermolysis bullosa with multi-systemic manifestations including nephropathy. *Matrix Biol*. 2018;66, 22-33. doi: 10.1016/j.matbio.2017.11.003
14. Dunn, C., Ambur, A., Foss, M., & Nathoo, R. Expanding the spectrum of epidermolysis bullosa simplex: Syndromic epidermolysis bullosa simplex with nephropathy and epilepsy secondary to CD151 tetraspanin defect—a case report and review of the literature. *JAAD Case Rep*. 2022;23, 136-140. doi: 10.1016/j.jdcr.2022.03.012
15. Almokali, K., Alshalawi, H., Aldriwesh, M. G., & Alotibi, R. S. Nephrotic syndrome: Pretibial epidermolysis bullosa in a patient with CD151 tetraspanin defect: A case report. *Int J Health Sci*. 2024;18(1), 35-40. doi: 10.5330/ijhs.v18n1.35
16. Has, C., & Fischer, J. Inherited epidermolysis bullosa: new diagnostics and new clinical phenotypes. *Exp Dermatol*. 2019;28(10), 1146-1152. doi: 10.1111/exd.13668
17. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. 2023;51(D1):D523-D531. doi: 10.1093/nar/gkac1052
18. Phan L, Zhang H, Wang Q, Villamarin R, Hefferon T, Ramanathan A, Kattman B. The

- evolution of dbSNP: 25 years of impact in genomic research. *Nucleic Acids Res.* 2025;53(D1):D925-D931. doi: 10.1093/nar/gkae977
- 19.Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31(13):3812-4. doi: 10.1093/nar/gkg509
- 20.Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 2015;31(16):2745-7. doi: 10.1093/bioinformatics/btv195
- 21.Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;76(1):7.20.1-7.20.41. doi: 10.1002/0471142905.hg0720s76
- 22.Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Protein networks and pathway analysis.* 2009:123-40. doi: 10.1007/978-1-60761-175-2_7
- 23.Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, De Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008;24(24):2938-9. doi: 10.1093/bioinformatics/btn564
- 24.Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC genomics.* 2013;14(3):1-7. doi: 10.1186/1471-2164-14-S3-S6
- 25.Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC genomics.* 2013;14(3):1-9. doi: 10.1186/1471-2164-14-S3-S2
- 26.Capriotti E, Fariselli P, and Casadio R. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. *Nucl. Acids Res.* 2006;33(suppl_2):W306-W310. doi: 10.1093/nar/gki375
- 27.Cheng J, Randall A, and Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Prot. Struct. Funct. Bioinform.* 2006;64(4): 1125-1132. doi: 10.1002/prot.20810
- 28.Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 2016;44(W1):W344-W50. doi: 10.1093/nar/gkw408
- 29.Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI, Soenderby CK, et al. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins: Struct, Funct, Bioinf.* 2019;87(6):520-7. doi: 10.1002/prot.25674
- 30.Chandra R. In silico study of secondary structure of hemoglobin protein. *Res J Pharm Technol.* 2021;14(12):6245-9. doi: 10.52711/0974-360X.2021.01080
- 31.Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC bioinformatics.* 2010;11(1):548. doi: 10.1186/1471-2105-11-548
- 32.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;45(D1): D362-D368. doi: 10.1093/nar/gkw937
- 33.Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, Morris Q. GeneMANIA update 2018. *Nucleic acids res.* 2018;46(W1):W60-W64. doi: 10.1093/nar/gky311
- 34.Sachs N, Kreft M, van den Bergh Weerman MA, Beynon AJ, Peters TA, Weening JJ, Sonnenberg A. Kidney failure in mice lacking the tetraspanin CD151. *J Cell Biol.* 2006;175(1):33-9. doi: 10.1083/jcb.200603073
- 35.Karamatic Crew V, Burton N, Kagan A, Green CA, Levene C, Flinter F, Brady RL, Daniels G, Anstee DJ. CD151, the first member of the tetraspanin (TM4) superfamily detected on erythrocytes, is essential for the correct assembly of human basement membranes in kidney and skin. *Blood.* 2004;104(8):2217-2223. doi: 10.1182/blood-2004-04-1512
- 36.Naylor RW, Watson E, Williamson S, Preston R, Davenport JB, Thornton N, Lowe M,

Williams M, Lennon R. Basement membrane defects in CD151-associated glomerular disease. *Pediatr Nephrol*. 2022 Dec;37(12):3105-15. doi: 10.1007/s00467-022-05447-y

37. Te Molder L, Juksar J, Harkes R, Wang W, Kreft M, Sonnenberg A. Tetraspanin CD151 and integrin $\alpha 3 \beta 1$ contribute to the stabilization of integrin $\alpha 6 \beta 4$ -containing cell-matrix adhesions. *J Cell Sci*. 2019;132(19):jcs235366. doi: 10.1242/jcs.235366

38. Berditchevski F, Gilbert E, Griffiths MR, Fitter S, Ashman L, Jenner SJ. Analysis of the CD151- $\alpha 3 \beta 1$ integrin and CD151-tetraspanin interactions by mutagenesis. *J Biol Chem*. 2001;276(44):41165-41174. doi:10.1074/jbc.M104041200