Journal of Health and Rehabilitation Research 2791-156X JHRR

For contributions to JHRR, contact at email: editor@jhrlmc.com

Original Article

Probing the Association of EPCAM Gene Single Nucleotide Polymorphisms with Genetic Disorders

Fatima Razzaq[1]*, Rashida Naseem[1]

[1]Institute of Molecular Biology and Biotechnology, The University of Lahore Pakistan.

Corresponding Author: Fatima Razzaq; Email: fatimarazzaq89@gmail.com

Conflict of Interest: None.

Razzaq F., et al. (2024). 4(2): DOI: https://doi.org/10.61919/jhrr.v4i2.956

ABSTRACT

Background: Mutations in the EPCAM gene can lead to the absence of this protein from epithelial cell surfaces, resulting in various disorders such as Lynch syndrome, congenital tufting enteropathy, cholestatic liver disease, and a range of cancers. Therefore, analyzing mutations in this gene is significant from diagnostic and prognostic perspectives. This study examines the effect of various mutations in the EPCAM gene on different attributes of the encoded protein.

Objective: The objective of this study was to analyze the impact of single nucleotide polymorphisms (SNPs) in the EPCAM gene on the protein's structural and functional properties, aiming to identify potential predictive biomarkers for EPCAM-associated diseases.

Methods: The study focused on the EPCAM gene transcript EPCAM-201, with ENSEMBL transcript ID ENST00000263735.9 and NCBI Reference Sequence NM_002354.3. The gene sequence was retrieved from the NCBI database, and SNPs were selected from the ENSEMBL database. A total of 21 variants were selected to design twenty-one cases. These cases were analyzed using various bioinformatics tools, including EXPASY for nucleotide sequence translation, HOPE server for 3D structure analysis, CELLO2GO for sub-cellular localization, and PROTPARAM for physicochemical parameter prediction. Data collection adhered to the principles outlined in the Declaration of Helsinki. Statistical analyses were conducted using SPSS version 25, with descriptive statistics and appropriate tests to determine the significance of the observed variations.

Results: Analysis revealed that sixteen SNPs, namely rs994384264, rs1280024892, rs1294456118, rs149875996, rs767811939, rs750826481, rs754293486, rs748292053, rs1460762372, rs1294456118, rs149875996, rs754293486, rs748292053, rs771063031, rs776854951, and rs267606785, significantly altered the 3D structure of mutated proteins. Among these SNPs, rs994384264 and rs149875996 also affected the sub-cellular localization of proteins. Additionally, four mutations—rs1280024892, rs1460762372, rs987919056, and rs771029207—caused alterations in extinction coefficient, isoelectric point (pI), aliphatic index, and instability index.

Conclusion: These single nucleotide variants might serve as predictive biomarkers for EPCAM-associated diseases, aiding in diagnosis and prognosis. Variants predicted in this study require further experimental validation to confirm their clinical utility.

Keywords: EPCAM gene, single nucleotide polymorphisms, SNP, Lynch syndrome, congenital tufting enteropathy.

INTRODUCTION

The Epithelial Cell Adhesion Molecule (EPCAM) gene, also known by several aliases including CD326, ESA, and TACSTD1, plays a critical role in various physiological and pathological processes (1-15). Located on chromosome 2p21, the EPCAM gene comprises nine exons spanning approximately 41.88 kb. These exons encode different functional domains of the EPCAM protein, including a signal peptide, EGF-like motif, thyroglobulin domain, transmembrane domain, and intracellular domain (8). EPCAM is a 314 amino acid protein organized into three main domains: an intracellular C-terminus, a transmembrane region, and an extracellular N-terminus (16). This protein is implicated in intercellular interactions, acting as an antagonist of E-cadherin, thereby disrupting its association with the cytoskeleton, regulating claudin-7 protein, and contributing to the formation of apical junctional complexes (17-25). Additionally, EPCAM is involved in the regulation of material transport across epithelial tissues, the release of exosomes, the maintenance of epithelial cell polarity, tissue mobilization, and the proliferation of both normal and cancerous tissues. It also plays a role in cell signaling pathways such as Wnt and nPKC and is essential in stem cell development (17-25).

Mutations in the EPCAM gene have been associated with various pathological conditions including cholestatic liver disease, inflammatory bowel disease, Lynch syndrome, congenital tufting enteropathy (CTE), and several types of cancer such as hepatocellular carcinoma, endometrial, colorectal, breast, ovarian, lung, pancreatic, and stomach cancers (1, 26-37). Previous studies have identified a frame shift mutation c.499dupC and multiple missense mutations in EPCAM associated with CTE patients, indicating a strong correlation between these genetic alterations and disease severity (1). For instance, a specific mutation, c.757G>A (p.Asp253Asn), has been linked to TE, leading to complete lack of EPCAM expression as evidenced by immunohistochemistry in affected individuals (39).

Given the significant role of EPCAM in health and disease, the present study aims to analyze the effects of single nucleotide polymorphisms (SNPs) on the encoded protein's attributes. This research focuses on missense SNPs, as most pathogenic mutations reported in EPCAM are of this type (1). Utilizing SNP data retrieved from the ENSEMBL database, this study examines 21 variants to evaluate their impact on the protein's 3D structure, subcellular localization, and physicochemical properties using various bioinformatics tools such as EXPASY, HOPE server, CELLO2GO, and PROTPARAM. Findings reveal that several SNPs, including rs994384264, rs1280024892, rs1294456118, rs149875996, rs767811939, rs750826481, rs754293486, rs748292053, and rs1460762372, disrupt hydrogen bond formation, potentially leading to protein misfolding and malfunction. Additionally, SNPs such as rs267606785 interfere with cysteine bridge formation, further compromising protein stability (43-45).

Moreover, mutations rs994384264 and rs149875996 alter the subcellular localization of the EPCAM protein, while variants rs1280024892, rs1460762372, rs987919056, and rs771029207 affect the protein's extinction coefficient, isoelectric point (pI), aliphatic index, and instability index. These alterations underscore the potential of these SNPs as predictive biomarkers for EPCAM-associated diseases, providing valuable insights for diagnostic and prognostic applications. In summary, the identified SNPs in the EPCAM gene could serve as crucial biomarkers in the management of Lynch syndrome, CTE, and TE, warranting further experimental validation to confirm their clinical utility in predicting genetic disorders (43-45).

MATERIAL AND METHODS

The study focused on the EPCAM gene transcript EPCAM-201, with ENSEMBL transcript ID ENST00000263735.9 and NCBI Reference Sequence NM_002354.3, to investigate the impact of mutations on the EPCAM protein. The gene sequence was retrieved from the NCBI database. The process involved navigating to the NCBI website (https://www.ncbi.nlm.nih.gov, accessed in June 2023), selecting "All databases," and using the "gene" option in the search bar to locate the EPCAM gene for Homo sapiens. The GenBank database was accessed, and the transcript ID NM_002354.3 was selected. The coding sequence (CDS), starting from the start codon ATG at nucleotide position 196 and ending at the stop codon UAA at position 1140, was highlighted for analysis.

Single nucleotide polymorphisms (SNPs) were retrieved from the ENSEMBL database (https://asia.ensembl.org/index.hotmail, accessed in June 2023). The database was accessed by specifying "Human" for the organism and "EPCAM" for the gene in the search bars. The EPCAM (Human Gene) option was selected, followed by the transcript ID ENST00000263735.9. Exons were highlighted in the sequence view, and missense SNPs in the exonic regions were identified by configuring the variation display to show exons only. A table of SNPs was generated, listing the missense variants selected for analysis (Table 1).

The nucleotide sequence was translated into an amino acid sequence using the EXPASY tool (https://web.expasy.org/translate, accessed in June 2023). Each case was designed by introducing a single SNP into the CDS, creating individual mutant sequences for analysis. For instance, in case 1, a SNP (TAT) was incorporated into the codon (TGT) encoding the 66th amino acid of the protein.

The three-dimensional (3D) structures of the mutated proteins were analyzed using the HOPE server (www3.cmbi.umcn.nl/hope/, accessed in June 2023). This tool assessed the impact of mutations on the protein's structure, interactions, and localization within the protein framework. Subcellular localization changes due to SNPs were evaluated using the CELLO2GO tool (Jiang et al., 2021). Variations in the physicochemical parameters of the mutated proteins, such as the number of amino acids, molecular weight, theoretical isoelectric point (pI), extinction coefficients, estimated half-life, instability index, and grand average of hydropathicity (GRAVY), were predicted using the PROTPARAM tool (https://web.expasy.org/protparam/, accessed in June 2023).

Data collection followed rigorous standards, ensuring the accuracy and reliability of the results. The study adhered to the principles outlined in the Declaration of Helsinki, ensuring ethical conduct throughout the research process. Although the study did not involve direct human subjects, ethical considerations were maintained in data handling and analysis.

The data analysis was performed using SPSS version 25. Descriptive statistics were calculated to summarize the findings, and appropriate statistical tests were applied to determine the significance of the observed variations in protein attributes caused by SNPs. The results were interpreted to understand the potential implications of these genetic variations on the EPCAM protein's function and its association with various genetic disorders.

In summary, this study employed a comprehensive methodology to investigate the effects of missense SNPs in the EPCAM gene on the encoded protein. By combining bioinformatics tools and rigorous data analysis, the study provided insights into the potential use of these SNPs as biomarkers for EPCAM-associated diseases, offering valuable information for diagnostic and prognostic applications.

RESULTS

Table 1: Detailed Information of SNPs in Exons of EPCAM Gene

Case #	SNP ID	Exonic Region	Nucleotide Change	Amino Acid Change	Codon Position
1	rs267606785	3	TGT > TAT	C > Y	66
2	rs994384264	3	TGG > TGC	W > C	117
3	rs987919056	4	GAA > AAA	E > K	147
4	rs1280024892	6	TAT > CAT	Y > H	186
5	rs1294456118	6	AAA > ACA	K > T	202
6	rs1202678365	6	ATA > GTA	I > V	209
7	rs149875996	6	GAT > AAT	D > N	211
8	rs1001841016	6	GTG > GGG	V > G	212
9	rs767811939	6	GCT > GTT	A > V	213
10	rs750826481	6	TAT > CAT	Y > H	215
11	rs754293486	6	AAA > ATA	K > I	218
12	rs757963724	6	GAT > AAT	D > N	219
13	rs1257057505	6	GTT > TTT	V > F	220
14	rs748292053	7	AAA > AGA	K > R	221
15	rs779694526	7	CCT > CTT	P > L	244
16	rs748065050	7	TAT > TTT	Y > F	251
17	rs1460762372	7	GAT > TAT	D > Y	253
18	rs771063031	7	GAA > GTA	E > V	254
19	rs975353107	7	CCT > ACT	P > T	257
20	rs776854951	7	GAA > GGA	E > G	258
21	rs771029207	7	CTG > CCG	L > P	286

The mutations were analyzed for their localization within the 3D structure of the EPCAM protein. Key findings included: The mutations rs267606785 and rs994384264 were localized in a domain annotated in UniProt as thyroglobulin-1 and transmembrane glycoprotein Epcam/Trop-2. Several other mutations, including those documented in cases 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, and 18, were located in the transmembrane glycoprotein Epcam/Trop-2 domain (Figure 2).

The study examined how each SNP altered the properties of the amino acids: Case 1: Cysteine to tyrosine (larger, less hydrophobic). Case 2: Tryptophan to cysteine (smaller, empty space in protein core). Case 3: Glutamic acid to lysine (larger, positively charged). Case 4: Tyrosine to histidine (larger, more hydrophobic).

The impact of SNPs on protein-protein interactions was significant: Case 1: Elimination of a cysteine bridge by mutation to tyrosine, destabilizing the protein. Case 2: Proximity to cysteine amino acid affecting cysteine bridge formation. Cases 3, 5, 6, and others: Variations in hydrogen bond and salt bridge formations, impacting protein stability and folding (Figure 3).

The sub-cellular localization of the EPCAM protein was altered by certain SNPs: SNPs in cases 2 and 7 changed the protein's localization from extracellular and cytoplasmic to extracellular only. Other cases showed no significant change in localization.

Table 2: Effect of SNPs on Sub-Cellular Localization of Mutated Proteins

Case ID	E	PM	C	CT	ER	G	L	M	CHL	P	V	N
Normal	1.89	0.49	1.06	0.02	0.06	0.11	0.03	0.19	0.38	0.30	0.05	0.37
Case 1	1.82	0.50	1.12	0.02	0.06	0.10	0.03	0.20	0.40	0.27	0.05	0.38
Case 2	2.00	0.56	0.96	0.02	0.05	0.10	0.03	0.17	0.32	0.29	0.05	0.38
Case 3	1.91	0.55	1.01	0.02	0.06	0.09	0.04	0.22	0.32	0.29	0.05	0.40
Case 4	1.92	0.37	1.09	0.02	0.06	0.10	0.03	0.20	0.41	0.32	0.05	0.39
Case 5	1.88	0.54	1.01	0.02	0.06	0.11	0.03	0.17	0.36	0.29	0.06	0.41

Case ID	E	PM	C	CT	ER	G	L	M	CHL	P	V	N
Case 6	1.52	0.29	1.62	0.02	0.07	0.18	0.03	0.11	0.55	0.09	0.04	0.42
Case 7	1.88	0.53	0.99	0.02	0.05	0.11	0.03	0.21	0.39	0.30	0.06	0.38
Case 8	1.90	0.46	1.10	0.02	0.05	0.10	0.03	0.19	0.38	0.31	0.05	0.36
Case 9	1.90	0.50	1.07	0.02	0.06	0.12	0.03	0.17	0.34	0.32	0.05	0.38
Case 10	1.89	0.43	1.15	0.02	0.06	0.12	0.03	0.21	0.34	0.28	0.05	0.37
Case 11	1.28	0.34	1.64	0.02	0.08	0.21	0.04	0.10	0.61	0.10	0.05	0.48
Case 12	1.83	0.56	1.01	0.02	0.06	0.11	0.03	0.20	0.40	0.29	0.06	0.37
Case 13	1.89	0.47	1.09	0.02	0.06	0.10	0.03	0.19	0.38	0.27	0.06	0.39
Case 14	1.89	0.53	1.03	0.02	0.05	0.10	0.03	0.20	0.39	0.28	0.05	0.37
Case 15	1.38	0.30	1.66	0.02	0.07	0.21	0.03	0.10	0.64	0.08	0.04	0.41
Case 16	1.42	0.30	1.62	0.02	0.07	0.21	0.03	0.10	0.62	0.12	0.04	0.38
Case 17	1.50	0.31	1.60	0.02	0.07	0.18	0.04	0.11	0.55	0.10	0.04	0.43
Case 18	1.43	0.30	1.62	0.02	0.07	0.18	0.03	0.11	0.58	0.11	0.04	0.44
Case 19	1.38	0.30	1.64	0.02	0.07	0.17	0.04	0.11	0.63	0.10	0.04	0.46
Case 20	1.87	0.54	1.03	0.02	0.05	0.09	0.03	0.21	0.38	0.31	0.05	0.38
Case 21	1.61	0.26	1.63	0.02	0.07	0.16	0.03	0.11	0.47	0.09	0.04	0.45

E = extracellular, PM = plasma membrane, C = cytoplasmic, CT = cytoskeleton, ER = endoplasmic reticulum, G = golgi apparatus, L = lysosomal, M = mitochondrial, CHL = chloroplast, P = peroxisomal, V = vacuole, N = nuclear. Six physicochemical parameters were analyzed for each case, revealing notable deviations caused by SNPs:

Table 3: Effect of SNPs on Physicochemical Parameters of Mutated Proteins

Case ID	Theoretical pI	Extinction Coefficient (M-1cm-1)	Half-Life (Hours)	Instability Index	Aliphatic Index	GRAVY
Normal	6.82	27390	30	31.22	88.26	-0.218
Case 1	5.23	29505	30	26.82	84.01	-0.320
Case 2	5.23	22640	30	26.68	84.01	-0.298
Case 3	5.41	28140	30	25.56	84.01	-0.310
Case 4	7.44	26650	30	31.99	88.82	-0.212
Case 5	5.15	28140	30	27.40	84.01	-0.299
Case 6	7.42	28140	30	31.75	88.50	-0.207
Case 7	5.32	28140	30	27.65	84.01	-0.309
Case 8	5.23	28140	30	27.14	83.13	-0.323
Case 9	5.23	28140	30	27.17	84.58	-0.302
Case 10	5.30	26650	30	27.08	84.01	-0.314
Case 11	5.15	28140	30	27.98	85.18	-0.283
Case 12	5.32	28140	30	27.40	84.01	-0.309
Case 13	5.23	28140	30	26.80	83.13	-0.313
Case 14	5.23	28140	30	27.48	84.01	-0.311
Case 15	5.23	28140	30	27.40	85.18	-0.292
Case 16	5.23	26650	30	27.02	85.18	-0.280
Case 17	5.32	29630	30	27.39	84.01	-0.302
Case 18	5.31	28140	30	27.31	84.88	-0.286
Case 19	5.23	28140	30	26.87	84.01	-0.306
Case 20	5.31	28140	30	26.87	84.01	-0.299
Case 21	5.23	28140	30	28.56	82.83	-0.325

These results indicate that several SNPs significantly altered the physicochemical properties of the EPCAM protein, potentially impacting its stability, structure, and function. The highest deviations in theoretical pI, extinction coefficient, and instability index were observed in cases 4, 2, and 1 respectively.

Journal of Health and Rehabilitation Research 2791-156X JHRR

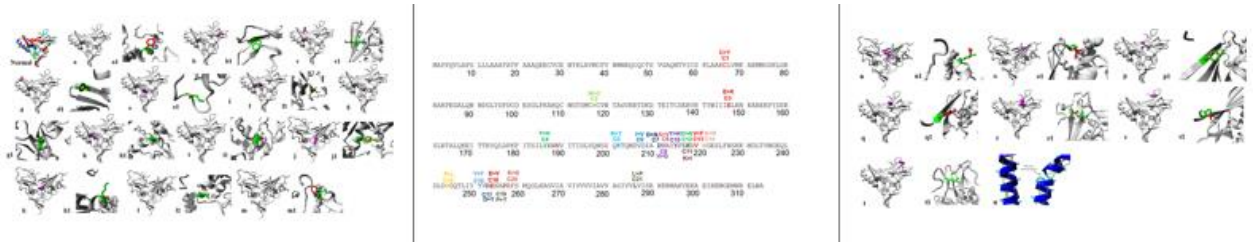

Figure 1: Graphical Presentation of SNP Positions in EPCAM Protein Structures

The figure illustrates the positions of 21 single nucleotide polymorphisms (SNPs) within the 3D structure of the EPCAM protein. The left and right panels show ribbon presentations of the normal and mutated proteins, highlighting the specific amino acid changes induced by each SNP. The central schematic maps the mutations along the primary sequence of the EPCAM protein, indicating the locations of the mutations within different functional domains. This visualization aids in understanding how each SNP impacts the protein's structure and potentially its function.

DISCUSSION

The study explored the impacts of various single nucleotide polymorphisms (SNPs) in the EPCAM gene on the structural and functional properties of the encoded protein. By analyzing 21 selected SNPs, the study provided insights into how these genetic variations could influence the protein's stability, interactions, and localization, thereby contributing to the understanding of EPCAM-associated diseases.

Previous research has established the critical role of EPCAM in maintaining epithelial cell integrity and its involvement in several pathological conditions, including Lynch syndrome, congenital tufting enteropathy (CTE), and various cancers (1-37). This study reinforced these findings by demonstrating that specific SNPs could significantly alter the EPCAM protein's 3D structure, leading to potential misfolding and loss of function. Notably, SNPs such as rs267606785 and rs994384264, which were located in key functional domains of the protein, caused substantial disruptions in the protein's secondary and tertiary structures, as evidenced by changes in hydrogen bond formation and cysteine bridge stability.

The study's findings also aligned with previous reports highlighting the importance of EPCAM's role in intercellular interactions and cellular signaling pathways (17-25). For instance, the mutation rs149875996, which affected subcellular localization and disrupted hydrogen bonds, could impair the protein's ability to maintain epithelial cell polarity and integrity. Such disruptions are consistent with the pathological mechanisms observed in diseases like CTE and Lynch syndrome, where EPCAM mutations lead to significant clinical manifestations (1, 26-37).

One of the study's strengths was its comprehensive approach, utilizing multiple bioinformatics tools to assess the impact of SNPs on protein structure and function. Tools like EXPASY, HOPE server, CELLO2GO, and PROTPARAM provided a robust framework for analyzing the physicochemical properties and subcellular localization of the mutated proteins. This multi-faceted analysis ensured a thorough evaluation of each SNP's potential pathogenicity.

However, the study had several limitations. While bioinformatics tools offer valuable predictions, experimental validation is crucial to confirm these findings. The study did not involve in vitro or in vivo experiments to verify the predicted effects of SNPs on EPCAM function. Furthermore, the focus on missense SNPs, although justified by their reported pathogenicity, excluded other types of mutations that could also impact the gene's function.

Future research should aim to experimentally validate the predicted impacts of these SNPs through cellular and molecular assays. Investigating the effects of these mutations in cell models or animal studies would provide more definitive evidence of their roles in disease pathology. Additionally, exploring other types of mutations, such as insertions, deletions, and regulatory variants, could offer a more comprehensive understanding of the genetic factors contributing to EPCAM-associated diseases.

The study's findings have important implications for the diagnosis and prognosis of EPCAM-related conditions. The identified SNPs, particularly those significantly altering protein structure and function, could serve as predictive biomarkers for diseases like Lynch syndrome and CTE. This could enhance early diagnosis and inform targeted therapeutic strategies, ultimately improving patient outcomes.

CONCLUSION

This study provided a detailed analysis of the effects of missense SNPs in the EPCAM gene, contributing to the understanding of how genetic variations can influence protein function and disease pathology. While the study's predictions require further experimental validation, the findings underscore the potential of these SNPs as biomarkers for EPCAM-associated diseases and highlight the need

for continued research in this area. The integration of bioinformatics predictions with experimental data will be essential for translating these findings into clinical practice, thereby enhancing the management of patients with EPCAM-related genetic disorders (43-45).

REFERENCES

1. Pathak SJ, Mueller JL, Okamoto K, Das B, Hertecant J, Greenhalgh L, et al. EPCAM mutation update: Variants associated with congenital tufting enteropathy and Lynch syndrome. Human mutation. 2019;40(2):142-61.
2. Farr A, Nelson A, Truex J, Hosier S. Epithelial heterogeneity in the murine thymus: a cell surface glycoprotein expressed by subcapsular and medullary epithelium. Journal of Histochemistry & Cytochemistry. 1991;39(5):645-53.
3. Went P, Vasei M, Bubendorf L, Terracciano L, Tornillo L, Riede U, et al. Frequent high-level expression of the immunotherapeutic target Ep-CAM in colon, stomach, prostate and lung cancers. British journal of cancer. 2006;94(1):128-35.
4. Abe H, Kuroki M, Imakiire T, Yamauchi Y, Yamada H, Arakawa F, et al. Preparation of recombinant MK-1/Ep-CAM and establishment of an ELISA system for determining soluble MK-1/Ep-CAM levels in sera of cancer patients. Journal of immunological methods. 2002;270(2):227-33.
5. Litvinov SV, Velders MP, Bakker HA, Fleuren GJ, Warnaar SO. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. The Journal of cell biology. 1994;125(2):437-46.
6. Litvinov SV, Bakker HAM, Gourevitch MM, Velders MP, Warnaar SO. Evidence for a role of the epithelial glycoprotein 40 (Ep-CAM) in epithelial cell-cell adhesion. Cell adhesion and communication. 1994;2(5):417-28.
7. De Leij L, Helrich W, Stein R, Mattes MJ. SCLC-cluster-2 antibodies detect the pancarcinoma/epithelial glycoprotein EGP-2. International Journal of cancer. 1994;57(S8):60-3.
8. Linnenbach AJ, Seng BA, Wu S, Robbins S, Scollon M, Pyrc JJ, et al. Retroposition in a family of carcinoma-associated antigen genes. Molecular and cellular biology. 1993;13(3):1507-15.
9. Klein CE, Hartmann B, Schön MP, Weber L, Alberti S. Expression of 38-kD cell-surface glycoprotein in transformed keratinocyte cell lines, basal cell carcinomas, and epithelial germs. Journal of investigative dermatology. 1990;95(1):74-82.
10. Strnad J, Hamilton AE, Beavers LS, Gamboa GC, Apelgren LD, Taber LD, et al. Molecular cloning and characterization of a human adenocarcinoma/epithelial cell surface antigen complementary DNA. Cancer research. 1989;49(2):314-7.
11. Thampoe IJ, Ng JSC, Lloyd KO. Biochemical analysis of a human epithelial surface antigen: differential cell expression and processing. Archives of biochemistry and biophysics. 1988;267(1):342-52.
12. Moldenhauer G, Momburg F, Möller P, Schwartz R, Hämmerling GJ. Epithelium-specific surface glycoprotein of Mr 34,000 is a widely distributed human carcinoma marker. British journal of cancer. 1987;56(6):714-21.
13. Varki NM, Reisfeld RA, Walker LE. Antigens associated with a human lung adenocarcinoma defined by monoclonal antibodies. Cancer research. 1984;44(2):681-7.
14. Lipinski M, Parks DR, Rouse RV, Herzenberg LA. Human trophoblast cell-surface antigens defined by monoclonal antibodies. Proceedings of the National Academy of Sciences. 1981;78(8):5147-50.
15. Herlyn M, Steplewski Z, Herlyn D, Koprowski H. Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. Proceedings of the National Academy of Sciences. 1979;76(3):1438-42.
16. Gaber A, Lenarčič B, Pavšič M. Current view on EpCAM structural biology. Cells. 2020;9(6):1361.
17. Winter MJ, Nagelkerken B, Mertens AEE, Rees-Bakker HAM, Briaire-de Bruijn IH, Litvinov SV. Expression of Ep-CAM shifts the state of cadherin-mediated adhesions from strong to weak. Experimental cell research. 2003;285(1):50-8.
18. Dollé L, Theise ND, Schmelzer E, Boulter L, Gires O, van Grunsven LA. EpCAM and the biology of hepatic stem/progenitor cells. American Journal of Physiology-Gastrointestinal and Liver Physiology. 2015.
19. Ladwein M, Pape UF, Schmidt DS, Schnölzer M, Fiedler S, Langbein L, et al. The cell–cell adhesion molecule EpCAM interacts directly with the tight junction protein claudin-7. Experimental cell research. 2005;309(2):345-57.
20. Jiang L, Shen Y, Guo D, Yang D, Liu J, Fei X, et al. EpCAM-dependent extracellular vesicles from intestinal epithelial cells maintain intestinal tract immune balance. Nature Communications. 2016;7(1):1-16.
21. Van Campenhout CA, Eitelhuber A, Gloeckner CJ, Giallonardo P, Gegg M, Oller H, et al. Dlg3 trafficking and apical tight junction formation is regulated by nedd4 and nedd4-2 e3 ubiquitin ligases. Developmental cell. 2011;21(3):479-91.
22. Maetzel D, Denzel S, Mack B, Canis M, Went P, Benk M, et al. Nuclear signalling by tumour-associated antigen EpCAM. Nature cell biology. 2009;11(2):162-71.
23. Münz M, Kieu C, Mack B, Schmitt B, Zeidler R, Gires O. The carcinoma-associated antigen EpCAM upregulates c-myc and induces cell proliferation. Oncogene. 2004;23(34):5748-58.

24. Osta WA, Chen Y, Mikhitarian K, Mitas M, Salem M, Hannun YA, et al. EpCAM is overexpressed in breast cancer and is a potential target for breast cancer gene therapy. Cancer research. 2004;64(16):5818-24.
25. Lei Z, Maeda T, Tamura A, Nakamura T, Yamazaki Y, Shiratori H, et al. EpCAM contributes to formation of functional tight junction in the intestinal epithelium by recruiting claudin proteins. Developmental biology. 2012;371(2):136-45.
26. Song Y, Liu C, Liu X, Trottier J, Beaudoin M, Zhang L, et al. H19 promotes cholestatic liver fibrosis by preventing ZEB1-mediated inhibition of epithelial cell adhesion molecule. Hepatology. 2017;66(4):1183-96.
27. Zheng X, Fan X, Fu B, Zheng M, Zhang A, Zhong K, et al. EpCAM inhibition sensitizes chemoresistant leukemia to immune surveillance. Cancer Research. 2017;77(2):482-93.
28. Maloy KJ, Powrie F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. Nature. 2011;474(7351):298-306.
29. Tutlewska K, Lubinski J, Kurzawski G. Germline deletions in the EPCAM gene as a cause of Lynch syndrome–literature review. Hereditary cancer in clinical practice. 2013;11(1):1-9.
30. Yu X, Ge N, Guo X, Shen S, Liang J, Huang X, et al. Genetic variants in the EPCAM gene is associated with the prognosis of transarterial chemoembolization treated hepatocellular carcinoma with portal vein tumor thrombus. PloS one. 2014;9(4):e93416.
31. Warneke VS, Behrens HM, Haag J, Krüger S, Simon E, Mathiak M, et al. Members of the EpCAM signalling pathway are expressed in gastric cancer tissue and are correlated with patient prognosis. British journal of cancer. 2013;109(8):2217-27.
32. Gebauer F, Struck L, Tachezy M, Vashist Y, Wicklein D, Schumacher U, et al. Serum EpCAM expression in pancreatic cancer. Anticancer Research. 2014;34(9):4741-6.
33. Yang Y, Fei F, Song Y, Li X, Zhang Z, Fei Z, et al. Polymorphisms of Ep CAM gene and prognosis for non-small-cell lung cancer in Han C hinese. Cancer science. 2014;105(1):89-96.
34. Vorobyeva A, Konovalova E, Xu T, Schulga A, Altai M, Garousi J, et al. Feasibility of imaging EpCAM expression in ovarian cancer using radiolabeled DARPin Ec1. International Journal of Molecular Sciences. 2020;21(9):3310.
35. Gostner JM, Fong D, Wrulich OA, Lehne F, Zitt M, Hermann M, et al. Effects of EpCAM overexpression on human breast cancer cell lines. BMC cancer. 2011;11(1):1-14.
36. Seeber A, Untergasser G, Spizzo G, Terracciano L, Lugli A, Kasal A, et al. Predominant expression of truncated EpCAM is associated with a more aggressive phenotype and predicts poor overall survival in colorectal cancer. International journal of cancer. 2016;139(3):657-63.
37. Torres A, Pac-Sosińska M, Wiktor K, Paszkowski T, Maciejewski R, Torres K. CD44, TGM2 and EpCAM as novel plasma markers in endometrial cancer diagnosis. BMC cancer. 2019;19(1):1-11.
38. Pêgas KL, Cambruzzi E, Ferrelli RS, Silva CS, Guedes RR, Adami M, et al. Tufting enteropathy with EpCAM mutation: case report. Jornal Brasileiro de Patologia e Medicina Laboratorial. 2014;50:234-7.
39. Ozler O, Brunner-Véber A, Fatih P, Müller T, Janecke AR, Arikan C. Long-Term Follow-Up of Tufting Enteropathy Caused by EPCAM Mutation p. Asp253Asn and Absent EPCAM Expression. JPGN Reports. 2021;2(1):e029.
40. Yahyazadeh Mashhadi SM, Kazemimanesh M, Arashkia A, Azadmanesh K, Meshkat Z, Golichenari B, et al. Shedding light on the EpCAM: an overview. Journal of cellular physiology. 2019;234(8):12569-80.
41. Spizzo G, Fong D, Wurm M, Ensinger C, Obrist P, Hofer C, et al. EpCAM expression in primary tumour tissues and metastases: an immunohistochemical analysis. Journal of clinical pathology. 2011;64(5):415-20.
42. Jiang Y, Wang D, Wang W, Xu D. Computational methods for protein localization prediction. Computational and Structural Biotechnology Journal. 2021;19:5834-44.
43. Hubbard RE, Haider MK. Hydrogen bonds in proteins: role and strength. Encyclopedia of Life Sciences. 2010.
44. Pylaeva S, Brehm M, Sebastiani D. Salt bridge in aqueous solution: Strong structural motifs but weak enthalpic effect. Scientific reports. 2018;8(1):1-7.
45. Panja AS, Maiti S, Bandyopadhyay B. Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. Scientific reports. 2020;10(1):1-9.