

RESEARCH NOTE

Open Access



An investigation into the role of inherited *CEACAM* gene family variants and colorectal cancer risk

Anna L. W. Huskey^{1,2} and Nancy D. Merner^{1*}

Abstract

Objective: This study was designed to determine if *CEACAM* mutations are associated with inherited risk of colorectal cancer. Recently, protein-truncating mutations in the *CEACAM* gene family were associated with inherited breast cancer risk. That discovery, along with aberrant expression of *CEACAM* genes in colorectal cancer tumors and that colorectal cancer and breast cancer share many risk factors, including genetics, inspired our team to search for inherited *CEACAM* mutations in colorectal cancer cases. Specifically utilizing The Cancer Genome Atlas (TCGA) blood-derived whole-exome sequencing data from the colorectal cancer cohort, rare protein-truncating variants and missense variants were investigated through single variant and aggregation analyses in European American and African American cases and compared to ethnic-matched controls.

Results: A total of 34 and 14 different *CEACAM* variants were identified in European American and African American colorectal cancer cases, respectively. Nine missense variants were individually associated with risk, two in African Americans and seven in European Americans. No identified protein-truncating variants were associated with CRC risk in either ethnicity. Gene family and gene-specific aggregation analyses did not yield any significant results.

Keywords: Colorectal cancer, *CEACAM*, TCGA, Inherited, Familial, Genetics, Genetic risk, Risk variant

Introduction

Colorectal cancer (CRC) is the fourth most commonly diagnosed cancer in the US [1], and the lifetime risk of development is 4–5% [1, 2]. However, this risk can increase with many factors, including a family history of CRC [1]. Approximately 30% of CRC cases are familial [2, 3], and of those cases with a known genetic cause, the majority have Lynch syndrome [4]. However, up to 30% of familial cases are estimated to be genetically unsolved [5].

Attempting to discover new CRC genetic risk factors, herein, the *CEACAM* (Carcinoembryonic antigen-related

cell adhesion molecule) gene family was investigated. *CEACAM* genes are a part of the Ig superfamily. These genes have diverse functions, including cell adhesion and signaling, influencing immunity, angiogenesis, and cancer [6–8]. Aberrant expression of *CEACAM* genes has long been associated with tumorigenesis, and atypical expression has been heavily linked to CRC development and progression [6, 8]. In 1965, CEA (more currently known as *CEACAM5*) was first identified as a tumor marker for CRC [9, 10]. Additionally, *CEACAM6* is overexpressed in CRC and has been determined to increase invasiveness [11]. Contrarily, *CEACAM1* [12, 13] and *CEACAM7* [14] have decreased expression in CRC. Furthermore, somatic mutations in *CEACAM1* [13] and *CEACAM5* [15] have been detected in CRC tumors. Nonetheless, the impact of inherited *CEACAM* gene mutations on CRC risk has yet to be determined.

*Correspondence: ndm0011@auburn.edu

¹ Department of Pathobiology, College of Veterinary Medicine, Auburn University, 1130 Wire Road, Auburn, AL 36849, USA

Full list of author information is available at the end of the article



Recently, rare protein-truncating variants (PTVs) in the *CEACAM* gene family were associated with the inherited risk of breast cancer [16]. That discovery, along with aberrant expression of *CEACAM* genes in CRC tumors and that CRC and breast cancer share many risk factors, including genetics [1, 17, 18], inspired our team to determine if *CEACAM* mutations are associated with CRC inherited risk.

Main text

Methods

Blood-derived exomes of CRC cases in The Cancer Genome Atlas (TCGA) were analyzed to investigate if *CEACAM* mutations play a role in inherited risk. Through approved research project #10805, whole-exome binary sequence alignment mapping (BAM) files were downloaded from the Genomic Data Commons (GDC) Data Portal Repository. Samples were acquired by setting specific filters. Filters under the ‘Cases’ category included Project (TCGA-COAD), Samples Sample Type (Blood-Derived Normal), and Race (‘Black or African American’ and ‘White’). The samples were further filtered under the ‘Files’ category, including Experimental Strategy (WXS) and Data Format (BAM). A total of 48 sample files were obtained for African Americans and 199 for European Americans. These files were downloaded using the GDC Data Transfer Tool (version 1.2.0).

The downloaded BAM files, which had previously been aligned to the hg38 human reference genome, were processed using the remaining portions of a pipeline adapted from the Genome Analysis Toolkit’s (GATK’s) best practices pipeline [19]. Base quality scores were recalibrated using BaseRecalibrator. Following base recalibration, the BAM files underwent coverage calculations for the exome and each *CEACAM* gene. Samtools depth function [20, 21] was used to determine the exome coverage using a BED file generated from UCSC Table Browser with the specifications: clade (Mammal), genome (Human), assembly (Dec. 2013 (GRCH38/hg38), group (Genes and Gene Predictions), track (NCBI RefSeq), and table (UCSC RefSeq (refGene)) with genome as the region of interest and “Whole Gene” selected. Samtools coverage function [20, 21] was used to generate coverage values for the *CEACAM* genes from a set of gene-specific intervals; including *CEACAM1* (NM_001184815; chr19:42507306-42528481), *CEACAM3* (NM_001815 at chr19:41796587-41811554), *CEACAM4* (NM_001817; chr19:41618971-41627074), *CEACAM5* (NM_004363; chr19:41708626-41730421), *CEACAM6* (NM_002483; chr19:41755530-41772210), *CEACAM7* (NM_006890; chr19:41673303-41688270), *CEACAM8* (NM_001816 at chr19:42580243-42594924), *CEACAM16* (NM_001039213; chr19:44699151-44710718), *CEACAM18* (NM_001278392;

chr19:51478643-51490605), *CEACAM19* (NM_020219; chr19:44671452-44684355), *CEACAM20* (NM_001102597; chr19:44506159-44529675), and *CEACAM21* (NM_001098506; chr19:41576166-41586844). Furthermore, regarding variant calling, the recalibrated BAM files were converted into genome variant calling format (gVCF) files using HaplotypeCaller (GATK version 4.1.9). GenomicsDBImport was used to generate ethnic-specific *CEACAM* gene family datasets, which were obtained by extracting the *CEACAM* gene intervals listed above. This process was followed by the GenotypeGVCFs function to generate ethnic-specific VCF files (GATK version 4.1.9). The two ethnic-specific VCF files were then annotated using ANNOVAR (version June 2020). Variants were filtered to include rare PTVs (nonsense mutations, frameshifting mutations, or splice-site affecting mutations) and missense variants with ethnic-specific minor allele frequencies (MAFs) of < 1% in Exome Variant Server (EVS; National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project) [22]. Each variant was individually investigated using the Fisher’s exact test [23, 24] in R (v 3.5.1), comparing MAFs of ethnic-specific TCGA CRC cases and EVS controls. Additionally, coverage values for each variant were assessed to determine the cohort’s average coverage at that genomic location. Subsequently, PTVs and missense variants were investigated together and as individual groups in gene-based and gene family-based aggregation analyses using the Fisher method through the ‘sumlog’ command as part of the ‘metap’ package within R [25, 26]. P-values were not corrected for multiple testing. Lastly, missense pathogenicity was predicted using Polyphen2 [27]. For all significant mutations, protein analysis using InterPro [28] and the Eukaryotic Linear Motif (ELM) resource [29] was carried out to identify *CEACAM* domains and binding motifs, respectively.

Results

The whole-exome BAM files downloaded from TCGA had an average exome coverage of 8X, ranging from 2.3X to 21.4X among the samples. Coverage values were also generated for each *CEACAM* gene (Additional file 1: Table S1). The average coverage for the gene family was 22.9X, with 100% of the bases covered at least 1X (Additional file 1: Table S1).

After filtering for rare PTVs and missense variants in the entire *CEACAM* gene family within the TCGA CRC cohort, a total of 14 different variants were identified in African American cases (one frameshift and 13 missense; Additional file 2: Table S2), and 34 different variants were identified in European American cases (one frameshift, two splice, and 31 missense; Additional file 3: Table S3). All identified variants were heterozygous, and there were

no cases of compound heterozygosity. The average coverage for the 14 variants identified in African Americans was 49X, ranging from 19 to 423X. Similarly, the average coverage for the 34 variants detected in European Americans was 42X, ranging from 3 to 923X. No identified PTVs were associated with CRC risk in either ethnicity.

In African American cases, five of the 13 missense variants were classified as probably damaging; however, none of those mutations were associated with CRC risk. Only two variants were determined to be individually associated with African American CRC risk, including *CEACAM3*:p.(Y95N) and *CEACAM8*:p.(T247A), both predicted to be likely benign (Table 1).

In European American cases, 10 of the 31 missense variants were predicted to be probably damaging, but only two of which were found to be associated with CRC risk, *CEACAM1*:p.(Y68C) and *CEACAM18*:p.(C357G). A total of seven variants were determined to be individually associated with CRC in European Americans, all of which were missense variants, including the two aforementioned probably damaging missense variants and five predicted to be benign (Table 2).

Gene family and gene-specific aggregation analyses did not yield any significant results, including a combined assessment of PTVs and missense variants, as well as

group analyses of PTVs, missense mutations, and probably damaging missense mutations.

Discussion

Upon surveying the *CEACAM* gene family for rare PTVs and missense variants in CRC cases from TCGA and controls from the EVS, no gene-based or gene family-based associations with inherited risk of CRC were revealed. These results were unexpected due to the previous association of rare PTVs in the *CEACAM* gene family with inherited breast cancer risk [16], the known similarities between breast cancer and CRC risk [1, 17, 18], and the dis-regulation of *CEACAM* genes in CRC tumors [6, 8–15]. Moreover, it has been demonstrated that *CEACAM* gene function can be affected by even minor genetic changes [27], and specific residues within *CEACAM* proteins are crucial for normal function [12, 30, 31].

Despite the lack of association from aggregation analyses, individual variants were associated with CRC inherited risk (Tables 1 and 2). All associations involved individual missense variants; none involved PTVs, unlike the association of *CEACAM* PTVs with breast cancer risk [16]. Only four different PTVs were detected amongst all CRC cases, none of which overlapped between ethnicities. In European American CRC cases, two splice variants

Table 1 Significant rare mutations identified in TCGA CRC African American (AA) cohort

Gene	Chr 19 position	Mutation type	Functional prediction—polyphen	cDNA change	Protein change	TCGA AA Colon MAF (%)	EVS AA MAF (%)	AA individual P-values
<i>CEACAM3</i> : NM_001815	41797807	missense	benign: 0.159	c.283T>A	p.(Y95N)	5.208	0.894	0.002
<i>CEACAM8</i> : NM_001816	42589003	missense	benign: 0.001	c.739A>G	p.(T247A)	4.167	0.931	0.015

Table 2 Significant rare mutations identified in TCGA CRC European American (EA) cohort

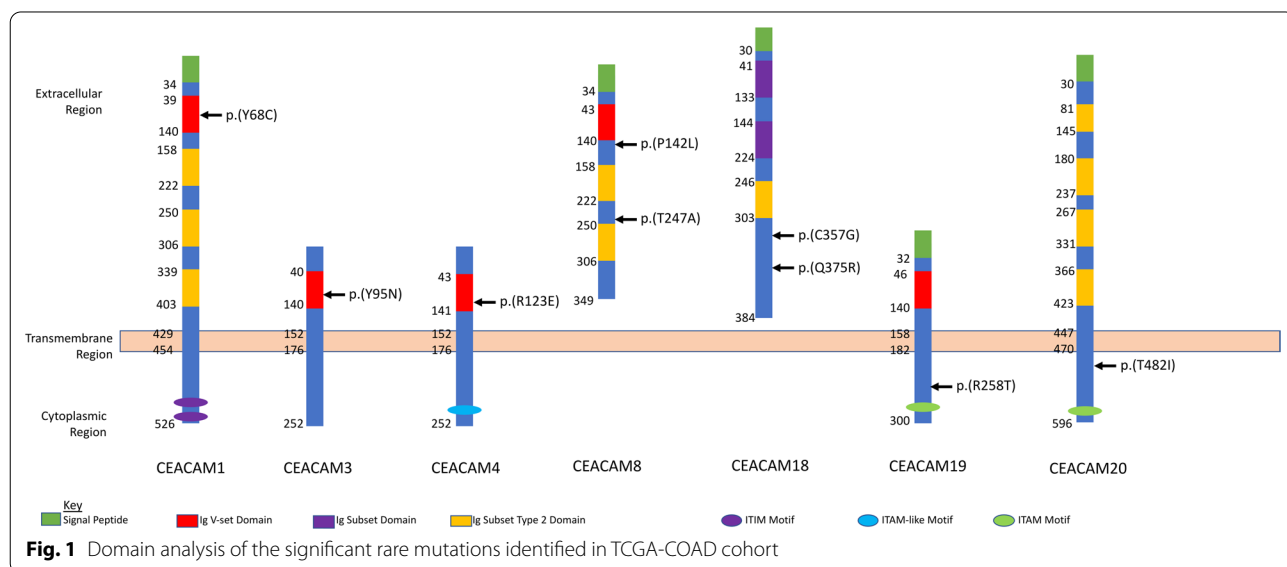
Gene	Chr 19 position	Mutation type	Functional prediction—polyphen	cDNA change	Protein change	TCGA EA colon MAF (%)	EVS EA MAF (%)	EA individual P-values
<i>CEACAM1</i> : NM_001184815	42527262	missense	probably-damaging: 1.0	c.203A>G	p.(Y68C)	0.503	0.070	0.046
<i>CEACAM4</i> : NM_001817	41625657	missense	benign: 0.325	c.368G>A	p.(R123E)	0.503	0.000	0.002
<i>CEACAM8</i> : NM_001816	42589735	missense	benign: 0.005	c.425C>T	p.(P142L)	0.503	0.012	0.006
<i>CEACAM18</i> : NM_001080405	51483229	missense	probably-damaging: 1.0	c.1069T>G	p.(C357G)	0.503	0.059	0.036
	51483284	missense	benign: 0.013	c.1124A>G	p.(Q375R)	0.503	0.059	0.036
<i>CEACAM19</i> : NM_020219	44681293	missense	benign: 0.01	c.773G>C	p.(R258T)	1.005	0.093	0.001
<i>CEACAM20</i> : NM_001102597	44512936	missense	benign: 0.062	c.1445C>T	p.(T482I)	0.503	0.000	0.002

were detected, including *CEACAM7:c.64+1G>T* and *CEACAM21:c.882+1G>A*, and a frameshift mutation was detected, *CEACAM20:p.(F542Sfs*56)*. One frameshift mutation was detected in an AA CRC case, *CEACAM21:p.(T32Pfs*47)*.

Overall, 9 missense variants were determined to be individually associated with risk, two in African Americans and seven in European Americans. Three associated variants were within the Ig V-set (variable) domain (Fig. 1), including *CEACAM1:p.(Y68C)* and *CEACAM4:p.(R123E)*, which were associated with European American CRC risk, and *CEACAM3:p.(Y95N)*, which was associated with African American CRC risk (Fig. 1). The Ig V-set domain is crucial for the dimerization of many CEACAM proteins and their ability to function within normal ranges [31, 32]. In *CEACAM1*, mutating particular residues within the Ig V-set domain can affect the monomer-homodimer exchange and result in the protein staying in a monomeric state [31]. *CEACAM1*'s ability to dimerize is required for proper function [33–36]. Knowing that *CEACAM1* dimerization is crucial and *CEACAM1*'s current role in CRC [12, 13], *CEACAM1:p.(Y68C)* is a probable CRC inherited risk factor. *CEACAM3:p.(Y95N)* has been reported as benign in ClinVar; however, limited information was provided for that clinical classification [37]. Considering *CEACAM3* has potential links to CRC [38, 39], validating the association of *CEACAM3:p.(Y95N)* with AA CRC inherited risk is crucial in identifying possible risk factors. Lastly, *CEACAM4* has been previously associated with thyroid cancer [40], but its role in CRC is unknown. Missense variants within the Ig V-set domain identified in this study could result in repressed dimerization and require further investigation.

Two statistically significant missense variants were identified in both *CEACAM8* and *CEACAM18*. The two variants in *CEACAM8*, *p.(P142L)* and *p.(T247A)*, were associated with CRC risk in European American and African American cases, respectively, and occur between functional domains of the protein (Fig. 1). Even though the role of these variants is unclear, *CEACAM8* forms dimers with *CEACAM6* and *CEACAM1* [32, 35], both of which have previous associations with CRC [11–13]. *CEACAM18* *p.(C357G)* and *p.(Q375R)* were significantly associated in European American CRC, and *p.(C357G)* was predicted to be pathogenic through PolyPhen2 [27]. These mutations occur after known functional domains for *CEACAM18* (Fig. 1) but could influence how the protein interacts with the cell membrane. Beyond these two *CEACAM18* variant associations, there is no known link between *CEACAM18* and CRC.

A single missense mutation in both *CEACAM19* [*p.(R258T)*] and *CEACAM20* [*p.(T482I)*] was associated with European American CRC. Both of these mutations occur within the cytoplasmic region of the protein but before the ITAM binding motifs (Fig. 1). The possible impacts of these mutations are unclear; however, *CEACAM19* and *-20* have previous cancer links [41–45]. Furthermore, *CEACAM20* has been determined to play a role in gut microbiome regulation [46, 47]. The microbiome is known to influence CRC risk and progression [1], which could explain *CEACAM20*'s role in CRC risk. Additionally, *CEACAM* gene expression is altered in Inflammatory Bowel Disease (IBD)[38, 48], another well-established risk factor for CRC [49–51]. Exploring how *CEACAM* mutations and aberrant expression result in both IBD and CRC is extremely important.



Unfortunately, IBD diagnoses were unavailable for TCGA CRC cases to explore that link.

Overall, this study aimed to determine if inherited *CEACAM* variants play a role in CRC risk. No gene- or gene family-based associations were identified, but nine individual missense variants in seven different *CEACAM* genes appear to be associated with inherited CRC risk. Further investigation is warranted.

Limitations

It is important to note that the TCGA CRC cohort is not a hereditary/familial CRC cohort. Though *CEACAM* variants do not appear to play a significant role in this cohort, studying hereditary/familial CRC cohorts could reveal different findings. Such investigations are important considering that a large percentage of inherited CRC is suspected to be influenced by lower penetrant variants compounded with environmental factors [1, 5]. Furthermore, the TCGA CRC cohort was subdivided by ethnicity, and European American cases were represented ~4X more than African American cases. This underrepresentation is a concerning limitation, as African Americans have the highest CRC incidence and mortality rates of all ethnicities in the United States [52]. Both TCGA CRC ethnic groups had a limited number of cases, and with the prevalence of previous research linking the *CEACAM* genes to spontaneous CRC [6, 8, 11–15, 38, 39, 53–55], more genetic and functional investigations of the *CEACAM* gene family should be carried out.

Abbreviations

AA: African American; BAM: Binary sequence alignment mapping; CRC: Colorectal cancer; ELM: Eukaryotic linear motif; EA: European American; EVS: Exome variant server; GATK's: Genome analysis toolkit's; gVCF: Genome variant calling format; GDC: Genomic data commons; MAFs: Minor allele frequencies; NHLBI: National Heart, Lung, and Blood Institute; PTVs: Protein-truncating variants; TCGA: The Cancer Genome Atlas.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13104-022-05907-6>.

Additional file 1: Table S1. Coverage values for the *CEACAM* genes.

Additional file 2: Table S2. Full list of rare (MAF < 1%) *CEACAM* mutations in African American TCGA-COAD cohort and EVS African American cohort. This includes rare stop gain, frameshifting, splice-site and missense mutations identified in the "Black or African American" TCGA-COAD cohort and the EVS African American cohort.

Additional file 3: Table S3. Full list of rare (MAF < 1%) *CEACAM* mutations in European American TCGA-COAD cohort and EVS European American cohort. This includes rare stop gain, frameshifting, splice-site and missense mutations identified in the "White or Caucasian" TCGA-COAD cohort and the EVS European American cohort.

Acknowledgements

We would like to acknowledge the Office of Information Technology at Auburn University Hopper High-Performance Computing Cluster and Easley High-Performance Computing Cluster for compute time.

Authors' contributions

ALWH and NDM wrote the manuscript. ALWH performed bioinformatic processing and statistical analyses. Both authors read and approved the final manuscript.

Funding

This research was supported by the Department of Pathobiology in the Auburn University College of Veterinary Medicine along with the Department of Drug Discovery and Development in the Harrison School of Pharmacy. AURIC Graduate Fellowship Program supported graduate student endeavors (to A.L.W.H.)

Availability of data and materials

The datasets supporting the conclusions of this article are available in The Cancer Genome Atlas GDC data portal TCGA-COAD repository, <https://portal.gdc.cancer.gov/projects/TCGA-COAD>.

Declarations

Ethics approval and consent to participate

All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki and have been approved by the Auburn University Institutional Review Board of the Office of Research Compliance (protocol #19-302 EP 1907). Furthermore, a request (#44682-1) for a Data Use Certification for TCGA data access was submitted and project (#10805) was approved. TCGA study participants provided informed consent through NIH-approved protocols.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Pathobiology, College of Veterinary Medicine, Auburn University, 1130 Wire Road, Auburn, AL 36849, USA. ²Department of Drug Discovery and Development, Harrison School of Pharmacy, Auburn University, 3306 Walker Building, Auburn, AL 36849, USA.

Received: 26 October 2021 Accepted: 11 January 2022

Published online: 03 February 2022

References

- Colorectal Cancer Facts and Figures 2020–2022. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2020-2022.pdf>. Accessed Mar 2021.
- Calvert PM, Frucht H. The genetics of colorectal cancer. *Ann Intern Med*. 2002;137(7):603–12.
- Kastrinos F, Syngal S. Inherited colorectal cancer syndromes. *Cancer J*. 2011;17(6):405–15.
- Haraldsdottir S, Rafnar T, Frankel WL, Einarsdottir S, Sigurdsson A, Hampel H, Snaebjornsson P, Masson G, Weng D, Arngrimsson R, et al. Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nat Commun*. 2017;8:14755.
- Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology*. 2010;138(6):2044–58.

6. Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev.* 2013;32(3–4):643–71.
7. Han ZW, Lyv ZW, Cui B, Wang YY, Cheng JT, Zhang Y, Cai WQ, Zhou Y, Ma ZW, Wang XW, et al. The old CEACAMs find their new role in tumor immunotherapy. *Invest New Drugs.* 2020;38(6):1888–98.
8. Kuespert K, Pils S, Hauck CR. CEACAMs: their role in physiology and pathophysiology. *Curr Opin Cell Biol.* 2006;18(5):565–71.
9. Gold P, Freedman SO. Demonstration of tumor-specific antigens in human colonic carcinomata by immunological tolerance and absorption techniques. *J Exp Med.* 1965;121:439–62.
10. Gold P, Freedman SO. Specific carcinoembryonic antigens of the human digestive system. *J Exp Med.* 1965;122(3):467–81.
11. Kim KS, Kim JT, Lee SJ, Kang MA, Choe IS, Kang YH, Kim SY, Yeom YI, Lee YH, Kim JH, et al. Overexpression and clinical significance of carcinoembryonic antigen-related cell adhesion molecule 6 in colorectal cancer. *Clin Chim Acta.* 2013;415:12–9.
12. Fournes B, Sadekova S, Turbide C, Letourneau S, Beauchemin N. The CEACAM1-L Ser503 residue is crucial for inhibition of colon cancer cell tumorigenicity. *Oncogene.* 2001;20(2):219–30.
13. Song JH, Cao Z, Yoon JH, Nam SW, Kim SY, Lee JY, Park WS. Genetic alterations and expression pattern of CEACAM1 in colorectal adenomas and cancers. *Pathol Oncol Res.* 2011;17(1):67–74.
14. Messick CA, Sanchez J, DeJulius KL, Hammel J, Ishwaran H, Kalady MF. CEACAM-7: a predictive marker for rectal cancer recurrence. *Surgery.* 2010;147(5):713–9.
15. Gu S, Zaidi S, Hassan MI, Mohammad T, Malta TM, Noushmeh H, Nguyen B, Crandall KA, Srivastav J, Obias V, et al. Mutated CEACAMs disrupt transforming growth factor beta signaling and alter the intestinal microbiome to promote colorectal carcinogenesis. *Gastroenterology.* 2020;158(1):238–52.
16. Huskey ALW, McNeely I, Merner ND. CEACAM gene family mutations associated with inherited breast cancer risk—a comparative oncology approach to discovery. *Front Genet.* 2021. <https://doi.org/10.3389/fgene.2021.702889>.
17. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895–2015. *Nat Rev Cancer.* 2015;15(3):181–94.
18. Scott RJ, Ashton KA. Familial breast and bowel cancer: does it exist? *Hered Cancer Clin Pract.* 2004;2(1):25–9.
19. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013. <https://doi.org/10.1002/0471250953.bi1110543>.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
21. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021. <https://doi.org/10.1093/gigascience/giab008>.
22. NHLBI GO Exome Sequencing Project (ESP). <http://evs.gs.washington.edu/EVS/>. Accessed Jan 2021
23. Sprent P. Fisher exact test. In: Lovric M, editor. *International encyclopedia of statistical science*. Berlin: Springer; 2011. p. 524–5.
24. Gilliam D, O'Brien DP, Coates JR, Johnson GS, Johnson GC, Mhlanga-Mutangadura T, Hansen L, Taylor JF, Schnabel RD. A homozygous KCNJ10 mutation in Jack Russell Terriers and related breeds with spinocerebellar ataxia with myokymia, seizures, or both. *J Vet Intern Med.* 2014;28(3):871–7.
25. Fisher RA. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd; 1925.
26. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for meta-analysis in medical research*. Chichester: Wiley; 2000.
27. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013. <https://doi.org/10.1002/0471142905.hg0720s76>.
28. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. InterPro: the integrative protein signature database. *Nucleic Acids Res.* 2009;37(Database issue):D211–215.
29. Kumar M, Gouw M, Michael S, Samano-Sanchez H, Panca R, Glavina J, Diakogianni A, Valverde JA, Bukirova D, Calyseva J, et al. ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.* 2020;48(D1):D296–306.
30. Kuroki M, Abe H, Imakiirei T, Liao S, Uchida H, Yamauchi Y, Oikawa S, Kuroki M. Identification and comparison of residues critical for cell-adhesion activities of two neutrophil CD66 antigens, CEACAM6 and CEACAM8. *J Leukoc Biol.* 2001;70(4):543–50.
31. Gandhi AK, Sun ZJ, Kim WM, Huang YH, Kondo Y, Bonsor DA, Sundberg EJ, Wagner G, Kuchroo VK, Petsko GA, et al. Structural basis of the dynamic human CEACAM1 monomer-dimer equilibrium. *Commun Biol.* 2021;4(1):360.
32. Bonsor DA, Gunther S, Beadenkopf R, Beckett D, Sundberg EJ. Diverse oligomeric states of CEACAM IgV domains. *Proc Natl Acad Sci USA.* 2015;112(44):13561–6.
33. Kim WM, Huang YH, Gandhi A, Blumberg RS. CEACAM1 structure and function in immunity and its therapeutic implications. *Semin Immunol.* 2019;42: 101296.
34. Zhuo Y, Yang JY, Moremen KW, Prestegard JH. Glycosylation alters dimerization properties of a cell-surface signaling protein, carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1). *J Biol Chem.* 2016;291(38):20085–95.
35. Skubitz KM, Skubitz AP. Interdependency of CEACAM-1, -3, -6, and -8 induced human neutrophil adhesion to endothelial cells. *J Transl Med.* 2008;6:78.
36. Rueckschloss U, Kuerten S, Ergun S. The role of CEA-related cell adhesion molecule-1 (CEACAM1) in vascular homeostasis. *Histochem Cell Biol.* 2016;146(6):657–71.
37. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016;44(D1):D862–868.
38. Kelleher M, Singh R, O'Driscoll CM, Melgar S. Carcinoembryonic antigen (CEACAM) family members and inflammatory bowel disease. *Cytokine Growth Factor Rev.* 2019;47:21–31.
39. Hollandsworth HM, Amirfakhri S, Filemoni F, Schmitt V, Wennemuth G, Schmidt A, Hoffman RM, Singer BB, Bouvet M. Anti-carcinoembryonic antigen-related cell adhesion molecule antibody for fluorescence visualization of primary colon cancer and metastases in patient-derived orthotopic xenograft mouse models. *Oncotarget.* 2020;11(4):429–39.
40. Wakabayashi-Nakao K, Hatakeyama K, Ohshima K, Ken Yamaguchi K, Mochizuki T. Carcinoembryonic antigen-related cell adhesion molecule 4 (CEACAM4) is specifically expressed in medullary thyroid carcinoma cells. *Biomed Res.* 2014;35(4):237–42.
41. Zisi Z, Adamopoulos PG, Kontos CK, Scorilas A. Identification and expression analysis of novel splice variants of the human carcinoembryonic antigen-related cell adhesion molecule 19 (CEACAM19) gene using a high-throughput sequencing approach. *Genomics.* 2020;112(6):4268–76.
42. Estiar MA, Esmaili R, Zare AA, Farahmand L, Fazilaty H, Zekri A, Jafarbeik-Iravanji M, Majidzadeh AK. High expression of CEACAM19, a new member of carcinoembryonic antigen gene family, in patients with breast cancer. *Clin Exp Med.* 2017;17(4):547–53.
43. Michaelidou K, Tzovaras A, Missitzis I, Ardavanis A, Scorilas A. The expression of the CEACAM19 gene, a novel member of the CEA family, is associated with breast cancer progression. *Int J Oncol.* 2013;42(5):1770–7.
44. Zhao H, Xu J, Wang Y, Jiang R, Li X, Zhang L, Che Y. Knockdown of CEACAM19 suppresses human gastric cancer through inhibition of PI3K/Akt and NF- κ B. *Surg Oncol.* 2018;27(3):495–502.
45. Zhang H, Eisenried A, Zimmermann W, Shively JE. Role of CEACAM1 and CEACAM20 in an in vitro model of prostate morphogenesis. *PLoS ONE.* 2013;8(1): e53359.
46. Kitamura Y, Murata Y, Park JH, Kotani T, Imada S, Saito Y, Okazawa H, Azuma T, Matozaki T. Regulation by gut commensal bacteria of carcinoembryonic antigen-related cell adhesion molecule expression in the intestinal epithelium. *Genes Cells.* 2015;20(7):578–89.
47. Murata Y, Kotani T, Supriatna Y, Kitamura Y, Imada S, Kawahara K, Nishio M, Daniwijaya EW, Sadakata H, Kusakari S, et al. Protein tyrosine phosphatase SAP-1 protects against colitis through regulation of CEACAM20 in the intestinal epithelium. *Proc Natl Acad Sci USA.* 2015;112(31):E4264–4271.
48. Saiz-Gonzalo G, Hanrahan N, Rossini V, Singh R, Ahern M, Kelleher M, Hill S, O'Sullivan R, Fanning A, Walsh PT, et al. Regulation of CEACAM family members by IBD-associated triggers in intestinal epithelial cells, their

correlation to inflammation and relevance to IBD pathogenesis. *Front Immunol.* 2021;12: 655960.

49. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature.* 2005;438(7069):803–19.
50. Stidham RW, Higgins PDR. Colorectal cancer in inflammatory bowel disease. *Clin Colon Rectal Surg.* 2018;31(3):168–78.
51. Kim ER, Chang DK. Colorectal cancer in inflammatory bowel disease: the risk, pathogenesis, prevention and diagnosis. *World J Gastroenterol.* 2014;20(29):9872–81.
52. Augustus GJ, Ellis NA. Colorectal cancer disparity in African Americans: risk factors and carcinogenic mechanisms. *Am J Pathol.* 2018;188(2):291–303.
53. Scholzel S, Zimmermann W, Schwarzkopf G, Grunert F, Rogaczewski B, Thompson J. Carcinoembryonic antigen family members CEACAM6 and CEACAM7 are differentially expressed in normal tissues and oppositely deregulated in hyperplastic colorectal polyps and early adenomas. *Am J Pathol.* 2000;156(2):595–605.
54. Rizeq B, Zakaria Z, Ouhtit A. Towards understanding the mechanisms of actions of carcinoembryonic antigen-related cell adhesion molecule 6 in cancer progression. *Cancer Sci.* 2018;109(1):33–42.
55. Han ZM, Huang HM, Sun YW. Effect of CEACAM-1 knockdown in human colorectal cancer cells. *Oncol Lett.* 2018;16(2):1622–6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

