Research article | Published: 10 July 2023

Identification of ZMYND19 as a novel biomarker of colorectal cancer: RNA-sequencing and machine learning analysis

Ghazaleh Khalili-Tanha, Reza Mohit, Alireza Asadnia, Majid Khazaei, Mohammad Dashtiahangar, Mina Maftooh, Mohammadreza Nassiri, Seyed Mahdi Hassanian, Majid Ghayour-Mobarhan, Mohammad Ali Kiani, Gordon A. Ferns, Jyotsna Batra, Elham Nazari ✉ & Amir Avan ✉

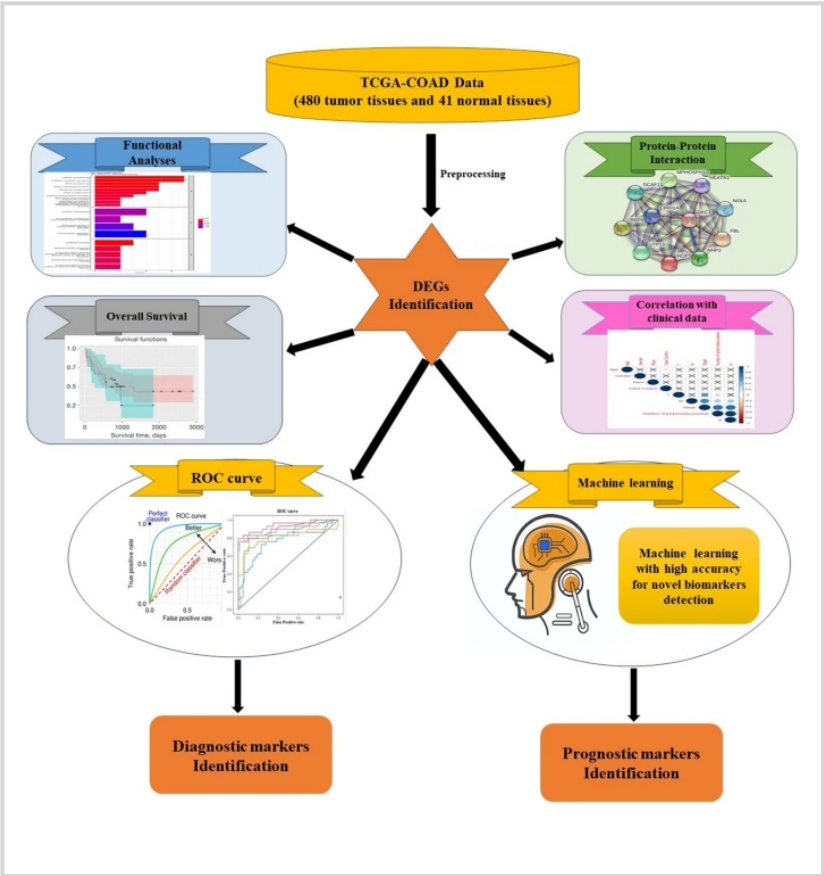
Journal of Cell Communication and Signaling (2023)

Abstract

Colorectal cancer (CRC) is the third most common cause of cancer-related deaths. The five-year relative survival rate for CRC is estimated to be approximately 90% for patients diagnosed with early stages and 14% for those diagnosed at an advanced stages of disease, respectively. Hence, the development of accurate prognostic markers is required. Bioinformatics enables the identification of dysregulated pathways and novel biomarkers. RNA expression profiling was performed in CRC patients from the TCGA database using a Machine

Learning approach to identify differential expression genes (DEGs). Survival curves were assessed using Kaplan–Meier analysis to identify prognostic biomarkers. Furthermore, the molecular pathways, protein–protein interaction, the co-expression of DEGs, and the correlation between DEGs and clinical data have been evaluated. The diagnostic markers were then determined based on machine learning analysis. The results indicated that key upregulated genes are associated with the RNA processing and heterocycle metabolic process, including C10orf2, NOP2, DKC1, BYSL, RRP12, PUS7, MTHFD1L, and PPAT. Furthermore, the survival analysis identified NOP58, OSBPL3, DNAJC2, and ZMYND19 as prognostic markers. The combineROC curve analysis indicated that the combination of C10orf2 -PPAT- ZMYND19 can be considered as diagnostic markers with sensitivity, specificity, and AUC values of 0.98, 1.00, and 0.99, respectively. Eventually, ZMYND19 gene was validated in CRC patients. In conclusion, novel biomarkers of CRC have been identified that may be a promising strategy for early diagnosis, potential treatment, and better prognosis.

Graphical abstract



Abbreviations

CRC: Colorectal cancer

DEGs: Differential expression genes

COAD: Colon adenocarcinomas

READ: Rectum adenocarcinomas

TCGA: The Cancer Genome Atlas

SVM: Support vector machine

KNN: K-nearest neighbors algorithm

DTs: Decision Tree

RF algorithm: Random Forest

ML: Machine learning

ROC curve: Receiver operating characteristic curve

AUC: Area under the *Curve*

GDAC: Global Data Assembly Centres

GEO: Gene expression omnibus

FFPE: Formalin-fixed *Paraffin*-embedded

IOSCA: Infantile-onset spinocerebellar ataxia

PEO: Progressive external ophthalmoplegia

snoRNPs: Small nucleolar RNPs

HCC: Hepatocellular carcinoma

OS: Overall survival

ccRCC: Clear cell renal cell carcinoma

PC: Prostate cancer

CDKs: Cyclin-dependent kinases

OC: Ovarian cancer

DL: Deep learning

ESCC: Esophageal squamous cell carcinoma

NSCLC: Non-small cell lung cancer

HNSCC: Head and neck squamous cell carcinoma

LASSO: Least absolute shrinkage and selection operator

DEMs: Differentially expressed miRNA

References

Agarwal S, Behring M, Hale K, Al Diffalha S, Wang K, Manne U et al (2019) MTHFD1L, a folate cycle

enzyme, is involved in progression of colorectal cancer. Transl Oncol 12(11):1461–1467

Akkoca AN, Yanık S, Özdemir ZT, Cihan FG, Sayar S, Cincin TG et al (2014) TNM and Modified Dukes staging along with the demographic characteristics of patients with colorectal carcinoma. Int J Clin Exp Med 7(9):2828

Alzahrani SM, Al Doghaither HA, Al-Ghafari ABJM, Oncology C (2021) General insight into cancer: an overview of colorectal cancer. Mol Clin Oncol 15(6):1–8

Ayala GE, Dai H, Li R, Ittmann M, Thompson TC, Rowley D et al (2006) Bystin in perineural invasion of prostate cancer. Prostate 66(3):266–272

Bi J, Huang Y, Liu Y (2019) Effect of NOP2 knockdown on colon cancer cell proliferation, migration, and invasion. Transl Cancer Res 8(6):2274

Biau G, Scornet EJT (2016) A random forest guided tour. TEST 25(2):197–227

Dinga R, Penninx BW, Veltman DJ, Schmaal L, Marquand AF (2019) Beyond accuracy: measures

for assessing machine learning models, pitfalls and guidelines

Diray-Arce J, Liu B, Cupp JD, Hunt T, Nielsen BL (2013) The ArabidopsisAt1g30680 gene encodes a homologue to the phage T7 gp4 protein that has both DNA primase and DNA helicase activities. BMC Plant 13(1):1–11

Dong D, Zhang W, Xiao W, Wu Q, Cao Y, Gao X et al (2021) A GRN autocrine-dependent FAM135B/AKT/mTOR feedforward loop promotes esophageal squamous cell carcinoma progression. Cancer Res 81(4):910–922

Du J, Gong A, Zhao X, Wang G (2022) Pseudouridylate synthase 7 promotes cell proliferation and invasion in colon cancer through activating PI3K/AKT/mTOR signaling pathway. Dig Dis Sci 67(4):1260–1270

Duan L, Yang W, Wang X, Zhou W, Zhang Y, Liu J et al (2019) Advances in prognostic markers for colorectal cancer. Expert Rev Mol Diagn 19(4):313–324

Eich M-L, Pena MDCR, Chandrashekar DS, Chaux A, Agarwal S, Gordetsky JB et al (2019) Expression and role of methylenetetrahydrofolate

dehydrogenase 1 like (MTHFD1L) in bladder cancer. Transl Oncol 12(11):1416–1424

Estève D, Roumiguié M, Manceau C, Milhas D, Muller C (2020) Periprostatic adipose tissue: a heavy player in prostate cancer progression. Curr Opin Endocr Metab Res 10:29–35

Fergus P, Chalmers C (2022) Performance evaluation metrics. Applied deep learning. Springer, pp 115–138

Gao S, Sha Z, Zhou J, Wu Y, Song Y, Li C et al (2021) BYSL contributes to tumor growth by cooperating with the mTORC2 complex in gliomas. Cancer Biol Med 18(1):88

Gatalica Z, Torlakovic E (2008) Pathology of the hereditary colorectal carcinoma. Fam Cancer 7(1):15–26

Gaviraghi M, Vivori C, Tonon G (2019) How cancer exploits ribosomal RNA biogenesis: a journey beyond the boundaries of rRNA transcription. Cells 8(9):1098

Godoy H, Mhawech-Fauceglia P, Beck A, Miliotto A, Miller A, Lele S et al (2013) Developmentally restricted differentiation antigens are targets for

immunotherapy in epithelial ovarian carcinoma. Int J Gynecol Pathol 32(6):536

Goswami MT, Chen G, Chakravarthi BV, Pathi SS, Anand SK, Carskadon SL et al (2015) Role and regulation of coordinately expressed de novo purine biosynthetic enzymes PPAT and PAICS in lung cancer. Oncotarget 6(27):23445

Guerrieri AN (2020) Evaluation of the role of DKC1 overexpression in breast cancer

Hammad A, Elshaer M, Tang XJMB (2021) Engineering identification of potential biomarkers with colorectal cancer based on bioinformatics analysis and machine learning. Math Biosci Eng 18(6):8997–9015

He J, Yu J (2019) Long noncoding RNA FAM83A-AS1 facilitates hepatocellular carcinoma progression by binding with NOP58 to enhance the mRNA stability of FAM83A. Bioscience reports. Biosci Rep 39(11)

Hirano M, DiMauro SJN (2001) ANT1, Twinkle, POLG, and TP: new genes open our eyes to ophthalmoplegia. AAN Enterprises, pp 2163–2165

Hossain MJ, Chowdhury UN, Islam MB, Uddin S, Ahmed MB, Quinn JM et al (2021) Machine learning and network-based models to identify genetic risk factors to the progression and survival of colorectal cancer. Comput Biol Med 135:104539

Hou P, Shi P, Jiang T, Yin H, Chu S, Shi M et al (2020) DKC1 enhances angiogenesis by promoting HIF-1α transcription and facilitates metastasis in colorectal cancer. Br J Cancer 122(5):668–679

Hu M, Fu X, Si Z, Li C, Sun J, Du X et al (2019) Identification of differently expressed genes associated with prognosis and growth in colon adenocarcinoma based on integrated bioinformatics analysis. Front Genet 10:1245

Hu Q, Masuda T, Koike K, Sato K, Tobo T, Kuramitsu S et al (2021) Oxysterol binding protein-like 3 (OSBPL3) is a novel driver gene that promotes tumor growth in part through R-Ras/Akt signaling in gastric cancer. Sci Rep 11(1):1–12

Jiao H-l, Weng B-s, Yan S-s, Lin Z-m, Wang S-y, Chen X-p et al (2020) Upregulation of OSBPL3 by HIF1A promotes colorectal cancer progression through activation of RAS signaling pathway. Cell Death Dis 11(7):1–13

Kadowaki S, Kakuta M, Takahashi S, Takahashi A, Arai Y, Nishimura Y et al (2015) Prognostic value of KRAS and BRAF mutations in curatively resected colorectal cancer. World J Gastroenterol WJG 21(4):1275

Katsoulas A, Rachid Z, McNamee JP, Williams C, Jean-Claude BJ (2008) Combi-targeting concept: an optimized single-molecule dual-targeting model for the treatment of chronic myelogenous leukemia. Mol Cancer Ther 7(5):1033–1043

Kelley RK, Venook AP (2011) Prognostic and predictive markers in stage II colon cancer: is there a role for gene expression profiling? Clin Colorectal Cancer 10(2):73–80

Kim GP, Colangelo LH, Wieand HS, Paik S, Kirsch IR, Wolmark N et al (2007) Prognostic and predictive roles of high-degree microsatellite instability in colon cancer: a National Cancer Institute-National Surgical Adjuvant Breast and Bowel Project Collaborative Study. J Clin Oncol 25(7):767–772

Kim JH, Kim SY, Baek JY, Cha YJ, Ahn JB, Kim HS et al (2020) A phase II study of avelumab monotherapy in patients with mismatch repair-deficient/microsatellite instability-high or POLE-

mutated metastatic or unresectable colorectal cancer. Cancer Res Treat 52(4):1135–1144

Kristensen VN, Lingjærde OC, Russnes HG, Vollan HKM, Frigessi A, Børresen-Dale A-L (2014) Principles and methods of integrative genomic analyses in cancer. Nat Rev Cancer 14(5):299–313

Lee JK, Chan AT (2011) Molecular prognostic and predictive markers in colorectal cancer: current status. Curr Colorectal Cancer Rep 7(2):136–144

Lee D, Xu IM-J, Chiu DK-C, Lai RK-H, Tse AP-W, Li LL et al (2017) Folate cycle enzyme MTHFD1L confers metabolic advantages in hepatocellular carcinoma. J Clin Investig 127(5):1856–1872

Li H, Chen L, Han Y, Zhang F, Wang Y, Han Y et al (2021) The identification of RNA modification gene PUS7 as a potential biomarker of ovarian cancer. Biology 10(11):1130

Liang J, Wen J, Huang Z, Chen X-p, Zhang B-x, Chu L (2019) Small nucleolar RNAs: insight into their function in cancer. Front Oncol 9:587

Liao H, Gaur A, McConie H, Shekar A, Wang K, Chang JT et al (2021) hNOP2/NSUN1 regulates ribosome biogenesis through stabilization of

snoRNP complexes and cytosine-5 methylation of 28S rRNA. *bioRxiv*

Liu K, Chen H-L, Wang S, Gu M-M, Chen X-M, Zhang S-L et al (2016) High expression of RIOK2 and NOB1 predict human non-small cell lung cancer outcomes. Sci Rep 6(1):1–8

Liu H, Li J, Zhao H, Liu X, Ye X (2021) DNAJC2 is reversely regulated by miR-627-3p, promoting the proliferation of colorectal cancer. Mol Med Rep 24(2):1–10

Liu B, Zhang J, Huang C, Liu H (2012) Dyskerin overexpression in human hepatocellular carcinoma is associated with advanced clinical stage and poor patient prognosis

Maurya NS, Kushwaha S, Chawade A, Mani A (2021) Transcriptome profiling by combined machine learning and statistical R analysis identifies TMEM236 as a potential novel diagnostic biomarker for colorectal cancer. Sci Rep 11(1):1–11

Micu BV, Andercou O, Vesa SC, Micu CM, Pop TR, Constantea N (2017) The prognostic role of microsatellite instability in colorectal cancer patients. Ann Ital Chir 6:425–432

Nassar ZD, Aref AT, Miladinovic D, Mah CY, Raj GV, Hoy AJ et al (2018) Peri-prostatic adipose tissue: the metabolic microenvironment of prostate cancer. BJU Int 121:9–21

Nazari E, Aghemiri M, Avan A, Mehrabian A, Tabesh HJGR (2021) Machine learning approaches for classification of colorectal cancer with and without feature selection method on microarray data. Gen Rep 25:101419

Nazari E, Chang H-CH, Deldar K, Pour R, Avan A, Tara M et al (2020) A comprehensive overview of decision fusion technique in healthcare: a systematic scoping review. Iran Red Crescent Med J 22(10)

Nazari E, Biviji R, Roshandel D, Pour R, Shahriari MH, Mehrabian A et al (2022) Decision fusion in healthcare and medicine: a narrative review. Mhealth 8.

Nikali K, Suomalainen A, Saharinen J, Kuokkanen M, Spelbrink JN, Lönnqvist T et al (2005)

Infantile onset spinocerebellar ataxia is caused by recessive mutations in mitochondrial proteins twinkle and twinky. Hum Mol Genet 14(20):2981–2990

O'Brien R, Tran SL, Maritz MF, Liu B, Kong CF, Purgato S et al (2016) MYC-driven neuroblastomas are addicted to a telomerase-independent function of dyskerin. Cancer Res 76(12):3604–3617

Ołdak M, Oziębło D, Pollak A, Stępniak I, Lazniewski M, Lechowicz U et al (2017) Novel neuro-audiological findings and further evidence for TWNK involvement in perrault syndrome. J Transl Med 15(1):1–13

Ramos M, Esteva M, Cabeza E, Llobera J, Ruiz A (2008) Lack of association between diagnostic and therapeutic delay and stage of colorectal cancer. Eur J Cancer 44(4):510–521

Siegel RL, Miller KD, Fuchs HE, Jemal A (2021) Cancer statistics, 2021. CA Cancer J Clin 71(1):7–33

Sieron P, Hader C, Hatina J, Engers R, Wlazlinski A, Müller M et al (2009) DKC1 overexpression

associated with prostate cancer progression. Br J Cancer 101(8):1410–1416

Song D, Guo M, Xu S, Song X, Bai B, Li Z et al (2021) HSP90-dependent PUS7 overexpression facilitates the metastasis of colorectal cancer cells by regulating LASP1 abundance. J Exp Clin Cancer Res 40(1):1–24

Spelbrink JN, Li F-Y, Tiranti V, Nikali K, Yuan Q-P, Tariq M et al (2001) Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. Nat Genet 28(3):223–231

Sperandei S (2014) Understanding logistic regression analysis. Biochem Med 24(1):12–18

Su Y, Zhang M, Zhang L, Chen S, Zhang D, Zhang XJA (2019) Construction of an miRNA–mRNA regulatory network in colorectal cancer with bioinformatics methods. Anti Cancer Drugs 30(6):588–595

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J et al (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional

discovery in genome-wide experimental datasets. Nucleic Acids Res 47(D1):D607–D613

Tomczak K, Czerwińska P, Wiznerowicz M (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol 19(1A):A68

Varoquaux G, Colliot O (2022) Evaluating machine learning models and their diagnostic value

Wang H, Xiao W, Zhou Q, Chen Y, Yang S, Sheng J et al (2009) Bystin-like protein is upregulated in hepatocellular carcinoma and required for nucleologenesis in cancer cell proliferation. Cell Res 19(10):1150–1164

Wang F, Yuan JH, Wang SB, Yang F, Yuan SX, Ye C et al (2014) Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. Hepatology 60(4):1278–1290

Wang T-H, Lee C-Y, Lee T-Y, Huang H-D, Hsu JB-K, Chang T-H (2021) Biomarker identification through multiomics data analysis of prostate cancer prognostication using a deep learning

model and similarity network fusion. Cancers 13(11):2528

Wang J, Huang R, Huang Y, Chen Y, Chen F (2021a) Overexpression of NOP58 as a prognostic marker in hepatocellular carcinoma: a TCGA data-based analysis. Adv Ther 38(6):3342–3361

Wang G, Qu F, Liu S, Zhou J, Wang Y (2021b) Nucleolar protein NOP2 could serve as a potential prognostic predictor for clear cell renal cell carcinoma. Bioengineered 12(1):4841–4855

Wei C, Wang B, Chen Z-H, Xiao H, Tang L, Guan J-F et al (2022) Validating RRP12 expression and its prognostic significance in HCC basebd on data mining and bioinformatics methods. Front Oncol 12:812009

Wu H, Qin W, Lu S, Wang X, Zhang J, Sun T et al (2020) Long noncoding RNA ZFAS1 promoting small nucleolar RNA-mediated 2′-O-methylation via NOP58 recruitment in colorectal cancer. Mol Cancer 19(1):1–22

Yang Y-S, Yuan Y, Hu W-P, Shang Q-X, Chen L-Q (2018) The role of mitochondrial folate enzyme MTHFD1L in esophageal squamous cell carcinoma. Scand J Gastroenterol 53(5):533–540

Zhang S, Li X, Zong M, Zhu X, Cheng DJ (2017) Technology learning k for Knn classification. ACM Trans Intell Syst Technol 8(3):1–19

Zhang M, Meng L, Zhang Z, Wu J, Chen X, He J (2022) The respective relations of expression of OSBPL3 with Ki-67 expression and KRAS mutation in CRC for diagnosis and prognosis

Zhou C, Ye M, Ni S, Li Q, Ye D, Li J et al (2018) DNA methylation biomarkers for head and neck squamous cell carcinoma. Epigenetics 13(4):398–409

Acknowledgements

The authors acknowledge the financial support from Mashhad University of Medical Sciences.

Funding

This study was supported by a grant awarded to Amir Avan by the Mashhad University of Medical Sciences.

Author information

Ghazaleh Khalili-Tanha, Alireza Asadnia, Majid Khazaei and Elham Nazari have contributed equally to this work.

Authors and Affiliations

Metabolic Syndrome Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

Ghazaleh Khalili-Tanha, Alireza Asadnia, Majid Khazaei, Mina Maftooh, Seyed Mahdi Hassanian, Majid Ghayour-Mobarhan, Elham Nazari & Amir Avan

Basic Sciences Research Institute, Mashhad University of Medical Sciences, Mashhad, Iran

Seyed Mahdi Hassanian, Mohammad Ali Kiani & Elham Nazari

Medical Genetics Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

Ghazaleh Khalili-Tanha & Alireza Asadnia

Department of Anesthesia, Bushehr University of Medical Sciences, Bushehr, Iran

Reza Mohit

School of Medicine, Gonabad University of Medical Sciences, Gonabad, Iran

Mohammad Dashtiahangar

Recombinant Proteins Research Group, The Research Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran

Mohammadreza Nassiri

Department of Pediatrics, Ghaem Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

Mohammad Ali Kiani

Brighton & Sussex Medical School, Division of Medical Education, Falmer, Brighton, Sussex, BN1 9PH, UK

Gordon A. Ferns

Centre for Genomics and Personalised Health, Queensland University of Technology, Brisbane, 4059, Australia

Jyotsna Batra

Translational Research Institute, Queensland University of Technology, Brisbane, 4102, Australia

Jyotsna Batra

College of Medicine, University of Warith Al-Anbiyaa, Karbala, Iraq

Amir Avan

Faculty of Health, School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia

Jyotsna Batra & Amir Avan

Contributions

GK-T, EN, RM, AA, and MD have gathered study data and written the manuscript. MK, SMH, MM, MG-M, GAF, MAK, MN, and AA have provided critical revision of the final manuscript. EN and AA have contributed to the study design and approved the final version of the manuscript. All the authors read and approved the final version of the manuscript.

Corresponding authors

Correspondence to Elham Nazari or Amir Avan.

Ethics declarations

Conflict of interest

The authors have no conflict of interest to disclose.

Additional information

Publisher's Note

Rights and permissions

About this article

Cite this article

Khalili-Tanha, G., Mohit, R., Asadnia, A. *et al.* Identification of ZMYND19 as a novel biomarker of colorectal cancer: RNA-sequencing and machine learning analysis. *J. Cell Commun. Signal.* (2023). https://doi.org/10.1007/s12079-023-00779-2

Received	Accepted	Published
01 July 2022	29 May 2023	10 July 2023

DOI
https://doi.org/10.1007/s12079-023-00779-2

Keywords

CRC **Machine learning** **Biomarker**

Bioinformatic analysis