

УДК 577.2+577.29+57.021021

Экспрессия изоформ рецептора CD44 в клеточных линиях колоректального рака

© 2021 М.П. РАЙГОРОДСКАЯ^{1†}, В.О. НОВОСАД^{2†}, С.А. ТОНЕВИЦКАЯ^{1,3},
Д.В. МАЛЫЦЕВА^{1,3*}

¹ Научно-технический центр «БиоКлиникум», Москва, 115088

² Федеральное государственное бюджетное учреждение науки Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук, Москва, 117997

³ Факультет биологии и биотехнологии, Национальный исследовательский университет «Высшая школа экономики», Москва, 117312

*e-mail: dmaltseva@gmail.com

†Авторы внесли равный вклад в данную работу.

Поступила в редакцию 30.09.2021 г.

После доработки 05.10.2021 г.

Принята к печати 10.11.2021 г.

На основании данных, представленных в базе Cancer Cell Line Encyclopedia по секвенированным последовательностям мРНК, проведен анализ паттернов экспрессии вариантов мРНК, кодирующих изоформы белка CD44, в 55 клеточных линиях колоректального рака (КРР). Показано, что распределение изоформ мРНК CD44 значительно различается между линиями КРР. В 13 исследованных линиях, включая Caco-2 и RKO, экспрессия CD44 не зарегистрирована. Для большинства остальных линий изоформа 3 мРНК CD44 является наиболее представленной, однако уровень ее экспрессии варьирует, а в линиях MDST8 и SNU503 отсутствует. Самый высокий уровень экспрессии изоформы 3 наблюдается в линиях CW2 и T84. Следующая по частоте встречаемости — изоформа 4, при этом уровень ее экспрессии ниже такового для изоформы 3, кроме линий HCT116, SNU81, NCIN508 и SNUC4. Наиболее высокая экспрессия изоформы 4 обнаружена в линии MDST8 — единственной линии, в которой экспрессируется изоформа 6. Также в клеточных линиях КРР представлена изоформа 2 мРНК CD44, наиболее высокий уровень экспрессии которой обнаружен в линии SNU503. Экспрессия изоформ 1, 5 и 7 не обнаружена ни в одной из исследуемых линий. При выборе клеточной модели для исследования роли CD44 при КРР необходимо учитывать профиль экспрессии мРНК конкретных изоформ этого белка.

Ключевые слова: линии колоректального рака, изоформы CD44, секвенирование мРНК, CCLE

doi: 10.21519/0234-2758-2021-37-6-95-100

Эффективность терапии колоректального рака (КРР) в мире за последние годы значимо возросла, тем не менее этот вид онкологических заболеваний по-прежнему занимает третье место по частоте летальных исходов [1]. Более 70% смертей при КРР связаны с метастазами в печени, который наблюдаются у 25% пациентов уже при постановке первичного диагноза [2]. Примерно у 50% пациентов метастазы развиваются в дальнейшем.

В настоящее время прогнозирование результатов лечения и выбор схемы терапии для пациентов с КРР основаны на классификации TNM [3], которая не позволяет точно спрогнозировать течение КРР у пациентов на ранней стадии заболевания. В связи с этим обнаружение биомаркеров злокачественных опухолей толстой и прямой кишки остается актуальной и важной задачей для диагностики и прогнозирования КРР.

Список сокращений: КРР — колоректальный рак; ОСК — опухолевые стволовые клетки; CCLE (Cancer Cell Line Encyclopedia) — база данных характеристик клеточных линий, используемых в качестве моделей для изучения опухолей.

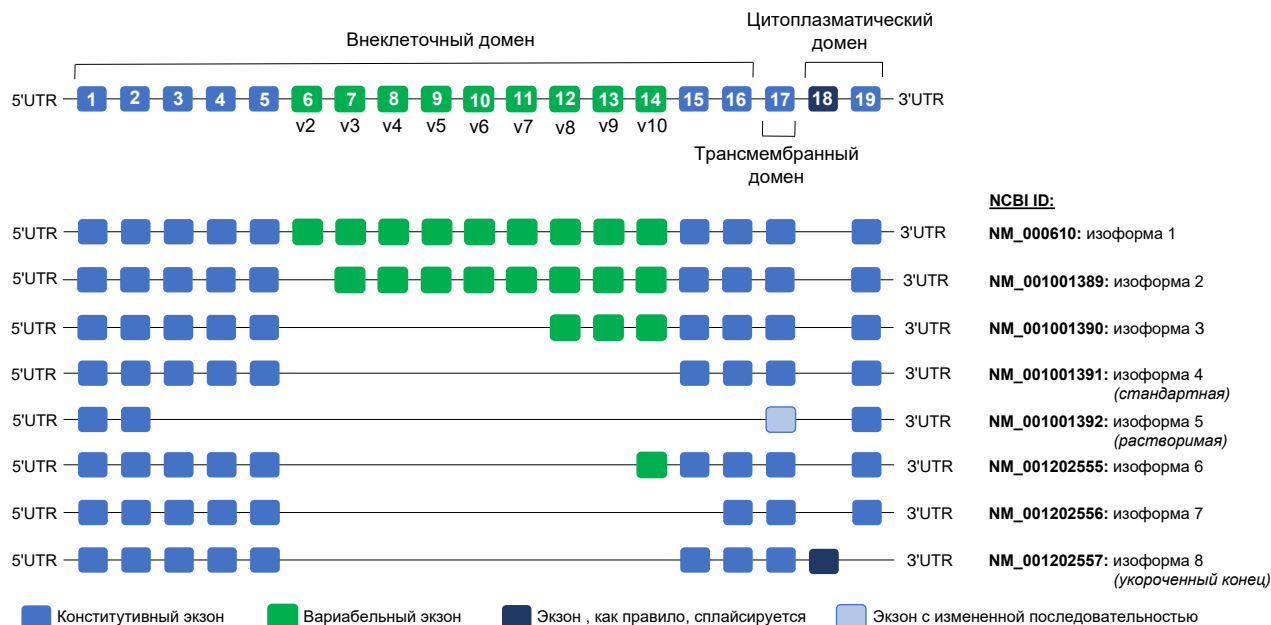


Рис. 1. Схематическое представление изоформ мРНК CD44, существование которых подтверждено экспериментально.

Fig. 1. Schematic representation of CD44 mRNA isoforms, the existence of which has been confirmed experimentally.

Накапливается все больше данных о важной роли опухолевых стволовых клеток (ОСК) в развитии и прогрессировании опухолевых заболеваний [4]. Это дает основания предполагать перспективность изучения маркеров ОСК в контексте диагностики и прогнозирования КРР. Трансмембранный гликопротеин CD44 принадлежит к числу доказанных и наиболее важных маркеров ОСК при КРР [4]. Повышенная экспрессия CD44 выявлена при различных онкологических заболеваниях, включая КРР [5–10].

У человека ген *CD44* включает 19 экзонов, которые подвергаются альтернативному сплайсингу, в результате чего не исключено появление множества изоформ мРНК CD44 [11, 12]. Результаты *in silico* анализа предсказывают существование 27 изоформ, однако только 8 из них подтверждены экспериментально (рис. 1). Самая короткая изоформа, состоящая из пяти первых (1–5) и четырех последних (15–17, 19) экзонов, экспрессируется в большинстве типов тканей и рассматривается как стандартная изоформа CD44 (CD44s). Включение варибельных экзонов v2–v10 (в ДНК соответствуют экзонам 6–14) приводит к появлению изоформ с большей молекулярной массой (CD44v), которые экспрессируются только в некоторых эпителиальных тканях и типах опухолей [13]. Варибельные изоформы CD44 часто обозначают в соответствии с номерами варибельных экзонов, включенных в состав мРНК (например, CD44v2–10). Хотя экспрессия конкретных изоформ CD44 и их функции еще мало изучены,

появляется все больше данных, свидетельствующих о различной функциональной нагрузке изоформ CD44 [14–17]. Стоит отметить, что исследования осложняются отсутствием стандартной номенклатуры изоформ этого белка. В публикуемых статьях авторы часто применяют номенклатуру, основанную либо на коммерческих названиях используемых моноклональных антител, либо на указании конкретного варибельного экзона, к которому специфичны антитела, например: CD44v3, CD44v6, CD44v9. В этих случаях часто не учитывается, что детекция конкретного варибельного экзона может привести к обнаружению всех изоформ, содержащих его, а не одного конкретного белка. Так, экзон v9 может входить в состав разных изоформ: CD44v2–10, CD44v3–10, и CD44v8–10, — которые соответствуют изоформам 1, 2 и 3 в базе NCBI (рис. 1). В этой связи данные тотального секвенирования мРНК представляются мощным инструментом для анализа профиля экспрессии изоформ CD44. Стоит также отметить существующее различие в обозначениях изоформ CD44 в базах данных UniProt и NCBI. Здесь мы будем использовать номенклатуру NCBI.

По данным мета-анализа, недавно опубликованного Wang Z. и др. [5], в опухолях пациентов с КРР экспрессируются варианты мРНК CD44, содержащие варибельные экзоны v2, v3, v6 и v9 [5]. Однако по-прежнему остается непонятной их функциональная значимость. В связи с этим исследования относительно роли

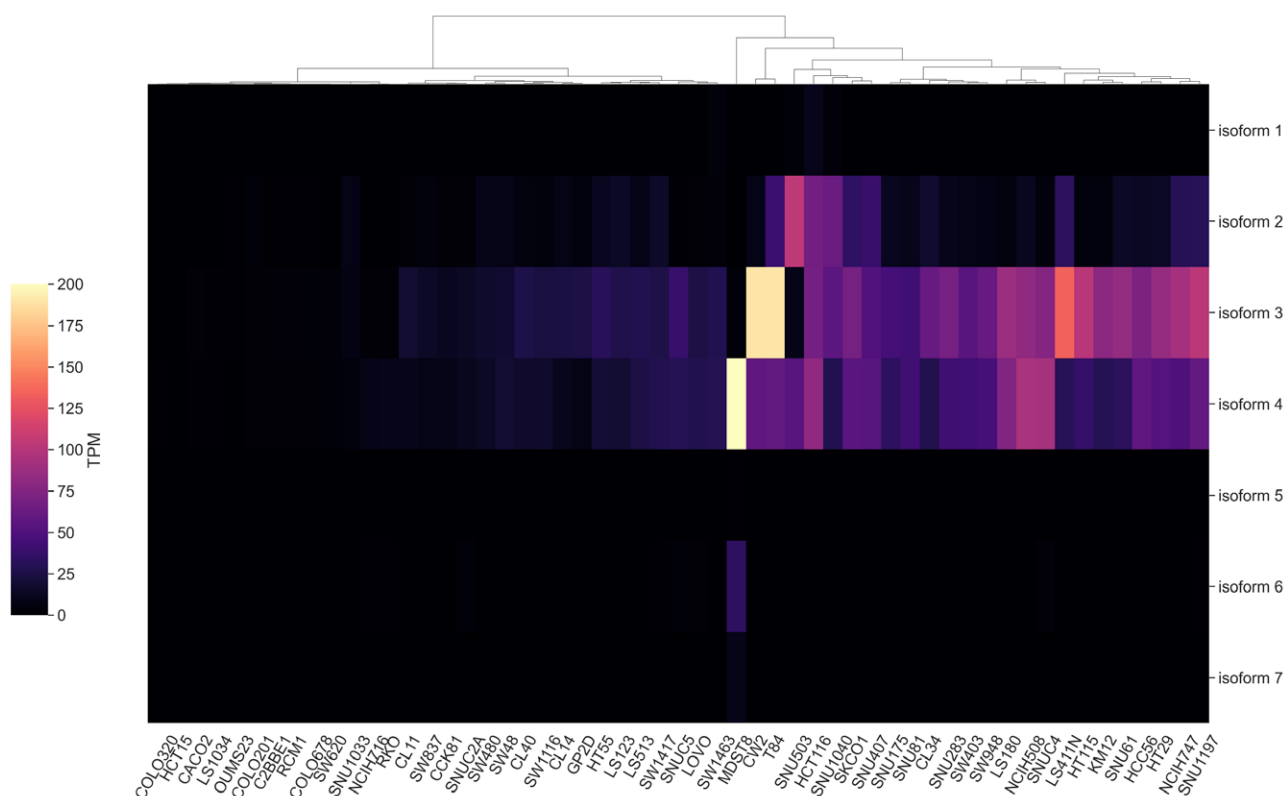


Рис. 2. Тепловая карта с иерархической кластеризацией клеточных линий КРР на основании экспрессии изоформ мРНК CD44. Для построения тепловой карты использованы данные секвенирования мРНК из базы данных CCLE.

Fig. 2. Heat map with hierarchical clustering of CRC cell lines by expression of CD44 mRNA isoforms. Built based on total mRNA sequencing data from the CCLE database.

конкретных изоформ CD44 (или определенного профиля изоформ) при КРР не теряют актуальности. Для решения этой задачи требуются модельные клеточные линии с хорошо аннотированным профилем экспрессии изоформ CD44.

В представленном исследовании на базе общедоступных данных из Cancer Cell Line Encyclopedia (CCLE) [18] по тотальному секвенированию мРНК проанализированы профили экспрессии изоформ мРНК CD44 в клеточных линиях КРР. В результате проведенного анализа идентифицированы линии потенциально наиболее релевантные для исследования роли изоформ CD44 при КРР.

УСЛОВИЯ ЭКСПЕРИМЕНТА

Обработка результатов секвенирования

Общедоступные данные секвенирования мРНК для 55 клеточных линий КРР были загружены с портала CCLE (<https://sites.broadinstitute.org/ccle>) в виде таблиц экспрессии в шкале TPM. Единицы TPM получали нормированием ридов сначала на длину сплайс-варианта (в тысячах нуклеотидов), а потом на глубину библиотеки

(сумма полученных значений, в миллионах ридов). В используемом наборе данных содержалась информация для изоформ с 1 по 7, данные для изоформы 8 (в соответствии с номенклатурой NCBI) отсутствовали.

Для анализа и визуализации распределения уровней экспрессии изоформ мРНК CD44 в клеточных линиях КРР была построена тепловая карта и иерархическая кластеризация методом Варда [19]. Для графического представления полученных данных использован модуль seaborn языка программирования Python [20].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Нами проведен анализ экспрессии изоформ мРНК CD44 в 55 клеточных линиях КРР с использованием данных секвенирования тотальной мРНК из базы CCLE. В результате выявлено, что распределение изоформ значительно различается между линиями. На рис. 2 представлена тепловая карта с иерархической кластеризацией клеточных линий КРР по экспрессии изоформ мРНК CD44.

Исследуемые линии можно разделить условно на пять групп по уровню экспрессии изоформ

2, 3 и 4: 1) 13 линий (с Colo320 по RKO), в которых не экспрессируется или практически не экспрессируется ни одной изоформы CD44; 2) 17 линий (с CL11 по SW1463), которые слабо экспрессируют изоформу 3, и примерно половина из них также слабо экспрессирует изоформу 4; 3) 17 линий (с SNU175 по SNU1197), в которых хорошо выражена экспрессия изоформы 3, в части из них также экспрессируется изоформа 4, но уровень экспрессии этой изоформы значительно ниже такового для изоформы 3 (кроме линий SNU81, NCIH508 и SNUC4); 4) 4 линии (с HCT116 по SNU407), в которых наблюдается близкий средний уровень экспрессии трех изоформ мРНК CD44 — 2, 3 и 4; 5) в эту группу можно включить оставшиеся 4 линии, в которых паттерн экспрессии изоформ мРНК CD44 существенно отличается от всех вышеперечисленных. Так, в линии MDST8 наблюдается очень высокий уровень экспрессии изоформы 4 и низкий — изоформы 6, экспрессия которой не наблюдается ни в одной другой из рассмотренных линий. В линиях CW2 и T84 на очень высоком уровне представлена изоформа 3 и на уровне ниже среднего — изоформа 4 (а также изоформа 2 в линии T84). В линии SNU503 наиболее выражена изоформа 2, а также экспрессируется изоформа 4. Таким образом, изоформы 1, 5 и 7 не представлены ни в одной из исследуемых линий.

Отметим, что изоформа 4 соответствует так называемой стандартной форме CD44 (или CD44s), которая не содержит варибельных экзонов. CD44s — наиболее распространенная изоформа и экспрессируется в большинстве тканей позвоночных [12, 15]. Эта изоформа представлена практически во всех рассмотренных линиях, экспрессирующих CD44 (группы 2–5). Изоформа 3, содержит варибельные экзоны v8, v9 и v10 и соответствует варианту CD44v8–10 (или CD44E). Эта изоформа рассматривалась ранее только как эпителиальный вариант белка CD44, поскольку экспрессируется в нормальных эпителиальных тканях, в том числе кишечника, и в тканях железистого эпителия [13, 21, 22]. Однако в более поздних работах показано, что включение экзона v9 в CD44 может рассматриваться как маркер ОСК [23, 24], а в процессе эпителиально-мезенхимального перехода (EMT) содержание изоформ CD44 с эксоном v9 увеличивается [25].

Изоформа 2 содержит варибельные экзоны v3–v10 и включает все наиболее часто встречаемые в исследованиях экзоны v3, v6 и v9. Линия SNU503 представляется наиболее релевантной моделью для изучения роли изоформы 2 при KPP. Изоформа 2 экспрессируется в линиях,

отнесенных к группам 4, 5, и в некоторых линиях группы 3.

В связи с тем, что ни в одной из клеточных линий KPP не экспрессируется изоформа 1, содержащая экзон v2, в настоящее время отсутствуют модели для исследования роли такого варианта CD44, экспрессия которого была показана в клинических образцах как прогностический маркер низкой выживаемости пациентов при KPP [5].

Обобщая полученные данные, следует сделать вывод, что распределение изоформ CD44 значительно различается между линиями KPP. Так, в таких часто используемых для изучения KPP линиях, как RKO, HCT116 и HT29, паттерны экспрессии изоформ мРНК CD44 отличаются. При выборе модели для исследования роли CD44 при KPP необходимо учитывать профиль экспрессии конкретных изоформ его мРНК.

ФИНАНСИРОВАНИЕ

Работа выполнена при финансовой поддержке Российского Научного Фонда (проект № 17-14-01338).

ЛИТЕРАТУРА

1. Bray F., Ferlay J., Soerjomataram I., et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.*, 2018, 68, 394–424. doi: 10.3322/caac.21492
2. Van Cutsem E., Cervantes A., Nordlinger B., Arnold D., ESMO Guidelines Working Group. Metastatic colorectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.*, 2014, 25(Suppl. 3), iii1–9. doi: 10.1093/annonc/mdu260
3. Weiser M.R. AJCC 8th edition: Colorectal cancer. *Ann. Surg. Oncol.*, 2018, 25, 1454–1455. doi: 10.1245/s10434-018-6462-1
4. Munro M.J., Wickremesekera S.K., Peng L., et al. Cancer stem cells in colorectal cancer: a review. *J. Clin. Pathol.*, 2018, 71, 110–116. doi: 10.1136/jclinpath-2017-204739
5. Wang Z., Tang Y., Xie L., et al. The prognostic and clinical value of CD44 in colorectal cancer: a meta-analysis. *Front. Oncol.*, 2019, 9, 309. doi: 10.3389/fonc.2019.00309
6. Chen J., Zhou J., Lu J., et al. Significance of CD44 expression in head and neck cancer: a systemic review and meta-analysis. *BMC Cancer*, 2014, 14, 15. doi: 10.1186/1471-2407-14-15
7. Luo Y., Tan Y. Prognostic value of CD44 expression in patients with hepatocellular carcinoma: meta-analysis. *Cancer Cell Int.*, 2016, 16, 47. doi: 10.1186/s12935-016-0325-2

8. Li X., Ma X., Chen L., et al. Prognostic value of CD44 expression in renal cell carcinoma: a systematic review and meta-analysis. *Sci. Rep.*, 2015, 5, 13157. doi: 10.1038/srep13157
9. Lin J., Ding D. The prognostic role of the cancer stem cell marker CD44 in ovarian cancer: a meta-analysis. *Cancer Cell Int.*, 2017, 17, 8. doi: 10.1186/s12935-016-0376-4
10. Chen Y., Fu Z., Xu S., et al. The prognostic value of CD44 expression in gastric cancer: a meta-analysis. *Biomed. Pharmacother.*, 2014, 68, 693–697. doi: 10.1016/j.biopha.2014.08.001
11. Screaton G.R., Bell M.V., Jackson D.G., et al. Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proc. Natl. Acad. Sci. USA*, 1992, 89, 12160–12164. doi: 10.1073/pnas.89.24.12160
12. Azevedo R., Gaiteiro C., Peixoto A., et al. CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. *Clin. Proteomics*, 2018, 15, 22. doi: 10.1186/s12014-018-9198-9
13. Ponta H., Sherman L., Herrlich P.A. CD44: from adhesion molecules to signalling regulators. *Nat. Rev. Mol. Cell Biol.*, 2003, 4, 33–45. doi: 10.1038/nrm1004
14. Zöller M. CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat. Rev. Cancer*, 2011, 11, 254–267. doi: 10.1038/nrc3023
15. Xu H., Niu M., Yuan X., et al. CD44 as a tumor biomarker and therapeutic target. *Exp. Hematol. Oncol.*, 2020, 9, 36. doi: 10.1186/s40164-020-00192-0
16. Bhattacharya R., Mitra T., Ray Chaudhuri S., Roy S.S. Mesenchymal splice isoform of CD44 (CD44s) promotes EMT/invasion and imparts stem-like properties to ovarian cancer cells. *J. Cell. Biochem.*, 2018, 119, 3373–3383. doi: 10.1002/jcb.26504
17. Skandalis S.S., Karalis T.T., Chatzopoulos A., Karamanos N.K. Hyaluronan-CD44 axis orchestrates cancer stem cell functions. *Cell. Signal.*, 2019, 63, 109377. doi: 10.1016/j.cellsig.2019.109377
18. Ghandi M., Huang F.W., Jané-Valbuena J., et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, 2019, 569, 503–508. doi: 10.1038/s41586-019-1186-3
19. Ward J.H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.*, 1963, 58, 236–244. doi: 10.2307/2282967
20. Waskom M.L. (n.d.) seaborn: statistical data visualization. *J. Open Source Software*, 2021, 6, 3021. doi: 10.21105/joss.03021
21. Brown T.A., Bouchard T., St John T., et al. Human keratinocytes express a new CD44 core protein (CD44E) as a heparan-sulfate intrinsic membrane proteoglycan with additional exons. *J. Cell Biol.*, 1991, 113, 207–221. doi: 10.1083/jcb.113.1.207
22. Naor D., Nedvetzki S., Golan I., et al. CD44 in cancer. *Crit. Rev. Clin. Lab. Sci.*, 2002, 39, 527–579. doi: 10.1080/10408360290795574
23. Suwannakul N., Ma N., Thanan R., et al. Overexpression of CD44 variant 9: a novel cancer stem cell marker in human cholangiocarcinoma in relation to inflammation. *Mediat. Inflamm.*, 2018, 1–8. doi: 10.1155/2018/4867234
24. Suwannakul N., Ma N., Midorikawa K., et al. CD44v9 induces stem cell-like phenotypes in human cholangiocarcinoma. *Front. Cell Dev. Biol.*, 2020, 8, 417. doi: 10.3389/fcell.2020.00417
25. Taniguchi D., Saeki H., Nakashima Y., et al. CD44v9 is associated with epithelial-mesenchymal transition and poor outcomes in esophageal squamous cell carcinoma. *Cancer Med.*, 2018, 7, 6258–6268. doi: 10.1002/cam4.1874

Expression of CD44 Isoforms in Human Colorectal Tumor Cell Lines

M.P. RAIGORODSKAYA^{1†}, V.O. NOVOSAD^{2†}, S.A. TONEVITSKAYA^{1,3}, D.V. MALTSEVA^{1,3*}

¹ Bioclinicum Scientific Research Center, Moscow, 115088, Russia

² Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997, Russia

³ Faculty of Biology and Biotechnology, National Research University Higher School of Economics, Moscow, 117312, Russia

*e-mail: dmaltseva@gmail.com

[†]These authors have contributed equally to this work.

Received September 30, 2021

Revised October 05, 2021

Accepted November 10, 2021

Abstract—The expression profile of CD44 isoforms in 55 colorectal cancer (CRC) cell lines has been evaluated based on mRNA sequencing data from the Cancer Cell Line Encyclopedia database. The distribution of CD44 mRNA isoforms differs significantly between CRC lines. In 13 studied lines, including Caco-2 and RKO, CD44 expression was not detected. In most other lines CD44 mRNA isoform 3 was the most abundant; however, the level of its expression varied, and it was absent in the MDST8 and SNU503 lines. The highest level of isoform 3 was observed in CW2 and T84 lines. The next most frequent was isoform 4 with its expression level lower than that of isoform 3, except for the HCT116, SNU81, NCIH508 and SNUC4 lines. The highest expression of isoform 4 was detected in the MDST8 line, the only line in which isoform 6 was also expressed. Isoform 2 was also present in CRC cell lines; its highest expression level was found in the SNU503 line. Isoforms 1, 5, and 7 did not show expression in any of the studied lines. It is necessary to take into account the mRNA expression profile of specific CD44 isoforms when choosing a cell model to study its role in CRC.

Key words: colorectal cancer, CD44 isoforms, mRNA sequencing, CCLE

Funding—The study was funded by the Russian Science Foundation (project no. 17-14-01338).

doi: 10.21519/0234-2758-2021-37-6-95-100