

(19)



**Евразийское
патентное
ведомство**

(21) **201900374** (13) **A2**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки
2020.02.28

(51) Int. Cl. **G16H 50/20** (2018.01)

(22) Дата подачи заявки
2019.08.16

(54) **СПОСОБ И СИСТЕМА ДЛЯ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ КОЛОРЕКТАЛЬНОГО РАКА**

(31) **2018129996**

(32) **2018.08.17**

(33) **RU**

(71) Заявитель:
**ФЕДЕРАЛЬНОЕ
ГОСУДАРСТВЕННОЕ
АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО
ОБРАЗОВАНИЯ ПЕРВЫЙ
МОСКОВСКИЙ
ГОСУДАРСТВЕННЫЙ
МЕДИЦИНСКИЙ УНИВЕРСИТЕТ
ИМЕНИ И.М. СЕЧЕНОВА
МИНИСТЕРСТВА
ЗДРАВООХРАНЕНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
(СЕЧЕНОВСКИЙ УНИВЕРСИТЕТ)
(ФГАОУ ВО ПЕРВЫЙ МГМУ ИМ.
И.М. СЕЧЕНОВА МИНЗДРАВА
РОССИИ (СЕЧЕНОВСКИЙ
УНИВЕРСИТЕТ)) (RU)**

(72) Изобретатель:

**Глыбочко Петр Витальевич,
Свистунов Андрей Алексеевич,
Фомин Виктор Викторович, Копылов
Филипп Юрьевич, Секачева
Марина Игоревна, Царьков Петр
Владимирович, Егоров Алексей
Викторович, Гитель Евгений
Павлович, Рагимов Алигейдар
Алекперович (RU)**

(74) Представитель:

Куприянова О.И. (RU)

(57) Изобретение относится к области медицины, а именно онкологии, и может быть использовано для скринингового определения вероятности наличия колоректального рака или выявления данного онкологического заболевания на ранней стадии. Скрининговое определение вероятности наличия колоректального рака основано на измерении уровня биомаркеров в образце биологической жидкости, полученном у субъекта: ApoA2, B2M, CYFRA.21.1, Ddimer, HE4, hsCRP, TTR, CEA, sVCAM.1, ApoA1, CA19.9, CA125, с последующей обработкой совокупности полученных значений биомаркеров с использованием по меньшей мере одной классификационной модели, обученной для определения вероятности наличия колоректального рака.

A2

201900374

201900374

A2

СПОСОБ И СИСТЕМА ДЛЯ СКРИНИНГОВОГО ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ НАЛИЧИЯ КОЛОРЕКТАЛЬНОГО РАКА

Область техники, к которой относится изобретение

Изобретение относится к области медицины, а именно онкологии, и может быть использовано для скринингового определения вероятности наличия колоректального рака или выявления данного онкологического заболевания на ранней стадии.

Уровень техники

Злокачественные опухоли представляют собой одну из самых значимых проблем здравоохранения не только в России, но и во всем мире.

Онкологические заболевания являются второй по частоте причиной смерти в России. Средний показатель заболеваемости злокачественными новообразованиями по России в 2016г. составил 408,6 чел. на 100000 населения. Средний показатель смертности – 201,6 чел. на 100000 населения. Онкологическая заболеваемость растет во всем мире. За последние 10 лет она увеличилась более чем на 20%.

В случае развития злокачественного заболевания, стадия, на которой онкологический процесс будет выявлен, является одним из определяющих факторов, обуславливающих продолжительность жизни пациента.

Колоректальный рак (КРР) стоит на третьем месте по распространенности среди населения индустриальных стран (Claudia Allemani, Hannah K Weir, Helena Carreira et al. Global surveillance of cancer survival 1995- 2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet 2014;(November 26). doi:[http://dx.doi.org/10.1016/S0140-6736\(14\)62038-9](http://dx.doi.org/10.1016/S0140-6736(14)62038-9)).

Статистика развитых стран мира свидетельствует о неуклонном росте впервые выявленных случаев рака толстой и прямой кишки по сравнению со злокачественными опухолями любой другой локализации, кроме рака легкого.

Смертность от КРР также очень высока, например, в США 10% всех онкологических смертей приходится на КРР, для Японии эта цифра 13%. Россия относится к странам с высоким уровнем заболеваемости КРР и смертности от него.

В 2010 г. в нашей стране зарегистрировано более 58 тыс. новых случаев КРР; в структуре онкологической заболеваемости в 2008 г. рак ободочной кишки находился на 4-м месте (6,5% случаев), рак прямой кишки - на 5-м месте (4,9%). Соответственно

увеличивается и смертность. В России КРР занимает 7 место по 1-годовой летальности (2016 г.).

Главными факторами риска для развития КРР являются генетическая предрасположенность, воспалительные заболевания кишечника, такие как болезнь Крона, язвенный колит, дивертикулит.

Онкологический скрининг предусматривает обследование людей, считающих себя здоровыми, с целью выявления онкологического заболевания на ранней стадии. Это позволиткратно снизить риск смерти.

На ранних этапах КРР развивается с минимальными клиническими симптомами. В то же время, смертность от КРР можно снизить с помощью ранней диагностики.

Для обнаружения колоректального рака известно применение следующих инвазивных методов - колоноскопии, гибкой сигмоскопии, ирригоскопии с двойным контрастированием, компьютерно - томографической колонографии («виртуальной колоноскопии»).

Сигмоскопия позволяет обследовать внутреннюю поверхность толстой кишки на расстоянии 60 см от анального отверстия. Однако очевидным недостатком этого метода является возможность обследования только левой части толстой кишки, а правая ее часть остается необследованной. В то время как специфичность сигмоскопии очень высока (98-100%), чувствительность в отношении выявления КРР низка от 35% до 70% из-за наличия большого количества правосторонних аденом. Колоноскопия позволяет выявить и удалить полипы, провести биопсию опухоли, расположенной в толстой кишке. Специфичность и чувствительность выявления полипов и КРР высоки.

По данным математического моделирования отдаленные результаты полиэктомии (United States National Polyp Study) показывают почти 90% снижение случаев заболеваемости КРР и смертельных исходов от него (Regula J., Rupinski M., Kraszewska E., Polkowski M. and other. Colonoscopy in colorectal cancer screening for detection of advanced neoplasia. N. Engl. J. Med. 2006; 355: 1863-1872).

Колоноскопия является «золотым стандартом» выявления КРР, все пациенты с положительным результатом других скрининговых исследований должны быть в последующем направлены на колоноскопию.

Ирригоскопия с двойным контрастированием (ИДК) позволяет исследовать всю толстую кишку, ее чувствительность и специфичность ниже диагностических показателей, получаемых при проведении колоноскопии. Даже при наличии больших полипов и опухолей ИДК обладает существенно более низкой чувствительностью (48%), чем колоноскопия. Кроме того, ИДК дает больше, чем колоноскопия, ложно -

положительных результатов (артефакты, определяемые как полипы) (Winawer S.J., Stewart E.T., Zauber A.G., Bond J.H., and other. A comparison of colonoscopy and doublecontrast barium enema for surveillance after polypectomy. National Polyp Study Work Group. N. Engl. J. Med. 2000; 342: 1766-1772).

Пациентам, у которых при ИДК была выявлена патология, в последующем необходимо провести колоноскопию. Несмотря на эти недостатки ИДК широко распространена и тот факт, что с ее помощью можно выявить до 50% больших полипов, обуславливает ее применение при отсутствии возможности проведения более точных исследований.

Компьютерно-томографическая колонография (КТК). Послойной спиральное компьютерно-томографическое сканирование брюшной полости и таза с последующей цифровой обработкой и анализом изображений может создать как дву -, так и трехмерную реконструкцию просвета толстой кишки («виртуальная колоноскопия»). Проведение этого исследования требует инсuffляции воздуха для раздувания кишки до возможного объема, который может перенести пациент. Мета - анализ исследований, в которых КТК использовали для выявления колоректальных полипов и рака, показал высокую чувствительность (93%) и высокую специфичность (97%) при наличии полипов размером в 10 мм или более. Однако, при комбинации полипов больших и средних размеров (свыше 6 мм) чувствительность метода снижалась до 86% и специфичность также до 86%. (Kim D.H., Pickhardt P.J., Taylor A.J., Leung W.K. and other. CT colonography versus colonoscopy for the detection of advanced neoplasia. The New England journal of medicine. 2007; 357: 1403-1412).

Большим недостатком КТК является и то, что для его повторного проведения пациента необходимо подвергнуть повторному воздействию ионизирующего излучения. Вышеописанные инвазивные методы не подходят для ранней диагностики КРР, т.к. ни одно из перечисленных выше исследований не может охватить полной картины развития и распространения КРР.

К широко распространенным неинвазивным методам определения КРР, используемым в клинике, относятся анализ кала на скрытую кровь, иммунохимическое определение гемоглобина в стуле пациента, определение ДНК маркеров в стуле пациента.

Анализ кала на скрытую кровь основан на определении гемоглобина в стуле пациента (Allison JE, Sakoda LC, Levin TR et al. Screening for colorectal neoplasms with new fecal occult blood tests: update on performance characteristics. J Natl Cancer Inst 2007; 99:1462-1470).

Наиболее распространенными являются тесты Hemoccult II (Smith Kline Diagnostics) и Hemoccult II Sensa (Smith Kline Diagnostics). При наличии крови в кале гваяковый тест показывает синее окрашивание. Изменение цвета связано с пероксидазоподобной активностью гемоглобина в стуле. Для гваякового теста характерна большая доля ложноположительных реакций, низкая специфичность выявления КРР. Ложнопозитивные сигналы могут возникать из-за нарушения диеты при подготовке к анализу, например, употребление красного мяса, растительных пероксидаз (некоторые фрукты и овощи), употребление высоких доз витамина С.

Наличие геморроя и ангиодисплазий, также могут вызывать ложные положительные результаты. Современные рекомендации по Hemoccult тестированию включают в себя пищевые ограничения (Ransohoff D.F., Lang C.A., Part I: Suggested technique for fecal occult blood testing and interpretation in colorectal cancer screening. *Ann Intern Med* 1997 126:808).

Из-за низкого уровня диагностической чувствительности, этот тест не позволял выявлять ранние стадии рака (Heresbach D, Manfredi S, D'Halluin P N, Bretagne JF, Branger B. Review in depth and meta-analysis of controlled trials on colorectal cancer screening by fecal occult blood test. *Eur J Gastroenterol Hepatol* 2006; 18(4):427-433).

В основе иммунохимического выявления гемоглобина в стуле пациента лежит взаимодействие гемоглобина со специфическими антителами, с последующей регистрацией сигнала иммунохимическим методом (ИФА). Эти тесты не реагируют с нечеловеческим гемоглобином или пероксидазами растений, что позволяет не соблюдать диетические ограничения (Cole SR, Young GP. Effect of dietary restriction on participation in fecal occult blood test screening for colorectal cancer. *Med J Aust* 2001; 175(4):195-198). Примерами таких тестов являются HemeSelect (SmithKline Diagnostics, США) и InSure (Enterix Inc., США). Количественное определение гемоглобина в стуле пациента лежит в основе теста FOB Gold (Sentinel Diagnostics, Италия).

К недостаткам метода относится низкая чувствительность выявления КРР.

Метод обнаружения ДНК маркеров в стуле пациента стремительно развивается с начала 2000 г. (Ahlquist DA. Colorectal cancer screening by detection of altered human DNA in stool: Feasibility of a multitarget assay panel. *Gastroenterology* 2000; 119(5):1219-1227). Целесообразность тестирования стула на наличие ДНК для выявления КРР была впервые продемонстрирована в 1992 г. (Sidransky D, Tokino T, Hamilton SR, et al. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* 1992; 256(5053): 102-105). ДНК-тест позволяет обнаруживать до 52% инвазивных опухолей КРР по сравнению с 13%, обнаруживаемыми с помощью обычного гваякового теста ($P=0,003$),

с сопоставимым уровнем специфичности. Однако в настоящее время этот тест сложно применим в реальной клинической практике из-за низкой доступности для подобной методики, что делает ее невозможным для обследования широкой популяции.

Известны также способы выявления KPP *in vitro* на ранних стадиях заболевания, основанные на определении биохимических маркеров в биологических тканях и жидкостях пациента, например, цельной крови, сыворотке или плазме. Возможность использования в клинической практике биохимических маркеров при оценке колоректального рака была рассмотрена Европейской группой по опухолевым маркерам (EGTM) (Duffy MJ et al. Clinical utility of biochemical markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines. *Eur J. Cancer* 39 (2003) 718-727).

В настоящее время в качестве биохимического маркера для диагностики KPP известно использование карциноэмбрионального антигена (CEA). Известно, что CEA увеличивается в 95% образцов тканей, полученных от пациентов с колоректальным, желудочным и поджелудочным раком, а также в большинстве случаев рака молочной железы, легких, головы и шеи (Goldenberg, DM, et al., *J. Natl. Cancer Inst. (Bethesda)* 57 (1976) 11-22). Повышенные уровни CEA в сыворотке крови также отмечались у пациентов с выявленным колоректальным раком на ранней стадии заболевания (Carriquiry, LA и Pineyro, A., *Dis. Colon Rectum* 42 (1999) 921-929, Herrera, MA, et al., *Ann. Surg.* 183 (1976) 5-9, Wanebo, HJ, et al., *N. Engl. J. Med.* 299 (1978) 448-451).

Однако определение CEA в сыворотке крови не обладает достаточными чувствительностью и специфичностью, позволяющими использовать его в качестве скринингового теста на колоректальный рак в бессимптомной популяции (Reynoso G., et al., *JAMA* 220 (1972) 361-365, Sturgeon, C., *Clinical Chemistry* 48 (2002) 1151-1159).

Для ранней диагностики KPP известно также использование панели из биохимических маркеров.

В частности, известен способ выявления на ранних этапах колоректального рака (US7731938 B2), основанный на определении в биологической жидкости пациента биохимических маркеров, включая цитокератин 19 (CYFRA 21.1). В частности, согласно данному изобретению оценку KPP *in vitro* осуществляют посредством измерения в образце крови концентрации CYFRA 21.1 и одного или нескольких других маркеров KPP, в качестве которых могут быть использованы NSE (нейроспецифическая энзолаза, также известная как гликолитическая энзолаза (2-фосфо-D-глицерата гидролаза, EC 4.2.1.11, молекулярная масса около 80 кД)); CA 19-9; CEA (карциноэмбриональный антиген); ASC «Связанный с апоптозом спексоподобный белок, содержащий связанный с каспазой рекрутирующий домен», также известен как «мишень, индуцированная метилированием

молчания 1» (TMS1); OPN (остеопонтин); FERR (ферритин); MASP (предшественник маспина), NNMT (N-метилтрансфераза белка никотинамида).

Из KR20150069056 (A) (прототип) известно определение наличия KPP по измеренным в сыворотке крови маркерам следующих белков: AFP, CEA, CA19-9, CA125, PSA, ApoA1, ApoA2, TTR, B2M, IL-6 и CRP, PAI-1. Результат получают посредством использования известных алгоритмов многостадийной математической обработки и анализа измеренных значений концентрации биомаркеров.

Однако стоит отметить, что ввиду возможных межрасовых различий в молекулярных механизмах карциногенеза (Sylvester BE, Huo D, Khramtsov A, et al. Molecular Analysis of Colorectal Tumors within a Diverse Patient Cohort at a Single Institution. Clinical Cancer Research. 2012;18(2):350-359) целесообразность использования единого набора биомаркеров для разных рас остается под вопросом. Так, результаты опубликованного мета-анализа указывают на более низкий предсказательный потенциал композитных микро-РНК тестов для европеоидной расы по сравнению с азиатской (Yang, X., Zhong, J., Ji, Y. et al. The expression and clinical significance of microRNAs in colorectal cancer detecting. Tumor Biol. (2015) 36: 2675). По данным другого мета-анализа тест на основе комбинации биомаркеров HSP60, CA 19-9 и CEA показал 97% чувствительность и 91% специфичность для китайской популяции, в то время как для европейской популяции эти показатели составили 47% и 90% соответственно (Bhardwaj et al., "Blood-Based Protein Signatures for Early Detection of Colorectal Cancer. Clinical and Translational Gastroenterology volume 8, page e128 (2017)).

Заявляемое изобретение основано на исследовании нового комплекса маркеров, позволяющем повысить точность и достоверность определения наличия заболевания при скрининге колоректального рака у конкретного пациента европейской популяции, формирование на этой основе той или иной группы риска и выявление тех пациентов, которые нуждаются в углубленном дорогостоящем обследовании для обнаружения ранней стадии рака толстой кишки.

Раскрытие изобретения

Технической проблемой, решаемой настоящим изобретением, является создание более точного способа определения вероятности наличия колоректального рака в европейской популяции.

Достижимым техническим результатом является повышение точности скринингового выявления наличия рака у конкретного пациента европейской популяции, причем уже на ранних стадиях его развития, посредством биостатистической обработки

результатов анализа фракции сыворотки и плазмы крови с определением концентрации комплексной группы биомаркеров.

Технический результат достигается посредством реализации способа скринингового определения вероятности наличия колоректального рака, включающего измерение уровня биомаркеров в образце биологической жидкости, полученном у субъекта: ApoA2, B2M, CYFRA.21.1, Ddimer, HE4, hsCRP, TTR, CEA, sVCAM.1, ApoA1, CA19.9, CA125, с последующей обработкой совокупности полученных значений биомаркеров с использованием, по меньшей мере, одной классификационной модели, обученной для определения высокой или низкой вероятности наличия колоректального рака. В качестве биологической жидкости могут быть использованы образцы плазмы или сыворотки крови, а также цельной крови". В качестве классификационных моделей могут быть использованы: метод «случайного леса» (random forest), и/или линейный дискриминантный анализ, и/или метод опорных векторов. При этом обученная классификационная модель может быть получена посредством реализации следующих шагов:

- формирование обучающей и тестовой выборок записей субъектов с измеренными значениями биомаркеров ApoA2, B2M, CYFRA.21.1, Ddimer, HE4, hsCRP, TTR, CEA, sVCAM.1, ApoA1, CA19.9, CA125, включающих записи о пациентах разного пола и возраста;
- обучение классификационной модели для выявления заданной патологии, используя записи обучающей и тестовой выборки;
- сохранение связей и весов обученной классификационной модели для последующего определения вероятности наличия колоректального рака по итогам обработки измеренных данных биомаркеров субъекта.

При формировании обучающей и тестовой выборки, включают записи субъектов с выявленной патологией - наличие колоректального рака и отсутствие колоректального рака.

Технический результат достигается также посредством реализации системы скринингового определения вероятности наличия колоректального рака, включающей:

- модуль ввода измеренных значений биомаркеров субъекта;
- модуль хранения данных, выполненный с возможностью хранения обучающей и тестовой выборки классификационной модели, связей и весов обученной классификационной модели, записей субъектов с измеренными значениями биомаркеров ApoA2, B2M, CYFRA.21.1, Ddimer, HE4, hsCRP, TTR, CEA, sVCAM.1, ApoA1, CA19.9, CA125, включающие записи о пациентах разного пола и возраста;

- модуль обученной классификационной модели, выполненный с возможностью построения и обучения, по меньшей мере, одной классификационной модели для определения наличия заданной патологии по упомянутым маркерам, взятым из модуля хранения данных;
- модуль диагностики, выполненный с возможностью обработки введенных значений биомаркеров субъекта с использованием, по меньшей мере, одной обученной классификационной модели;
- модуль вывода данных, выполненный с возможностью получения данных о высокой или низкой вероятности наличия колоректального рака.

Краткое описание чертежей

Изобретение поясняется чертежами, где:

На фиг. 1А. представлена диаграмма рассеяния «возраст пациента - концентрация биомаркеров». Точки – индивидуальные измерения, линии – предсказания линейной регрессионной модели. На графиках приведены значения корреляционных коэффициентов, рассчитанных по методу Пирсона и Р-значения, рассчитанные по тесту Стьюдента; на фиг.1 Б. представлена диаграмма размаха для оценки значимости гендерных различий в концентрациях биомаркеров. На графиках приведены Р-значения, полученные при помощи критерия Стьюдента, данные для женщин отмечены ,  – для мужчин;

На фиг. 2 - ROC-кривые для оценки предсказательной способности отдельных биомаркеров, где на фиг. 2А -  - AFP,  - ApoB,  - Ddimer,  - t.PSA,  - ApoA1,  - CYFRA.21.1,  - sVCAM.1,  - TIR; на фиг. 2Б -  - ApoA2,  - CA125,  - CA19.9,  - HE4,  - B2M,  - CA15.3,  - CEA,  - hsCRP.

На фиг. 3. - Примеры деревьев решений, полученных в результате обучения многофакторного классификационного алгоритма random forest на экспериментальных данных по 12 биомаркерам;

На фиг. 4 - Визуализация результатов разделения пациентов на 2 класса (здоровые доноры -  и пациенты с КРР - ) при помощи линейного дискриминантного анализа по 12 биомаркерам;

На фиг. 5 - Примеры 3-мерных проекций разделения объединенной популяции пациентов на 2 класса (здоровые доноры -  и пациенты с КРР - ) при помощи метода опорных векторов по 12 биомаркерам;

На фиг.6 - Доля классификаторов стратифицированная по AUROC в зависимости от количества включенных в них биомаркеров. Обучение проводилось при помощи А. Метода опорных векторов Б. Линейного дискриминантного анализа.

На фиг.7 - ROC-кривые для оценки предсказательной способности различных классификационных алгоритмов. А. Весь набор данных был использован как для обучения модели, так и для ее валидации; Б. 80% данных было использовано для обучения модели, 20% - для валидации. На фигуре отмечены ~~xxxx~~ - SVM, ---- - RF, ——— - LDA.

На фиг.8 - Блок-схема системы, предназначенной для оценки вероятности наличия КРР на основе данных пациента.

На фиг.9 - Алгоритм оценки вероятности наличия КРР на основе данных пациента.

Осуществление изобретения

Исходная группа биомаркеров, используемая в диагностическом тесте на оценку определения вероятности наличия колоректального рака (КРР) была получена с использованием многофакторной классификационной модели. Подобные методы позволяют находить комбинации биомаркеров, обладающих наибольшим диагностическим потенциалом. Математическая модель проходит обучение на экспериментальных измерениях заданного набора биомаркеров, полученных на смешанной выборке из здоровых добровольцев и пациентов с КРР. Обученная модель может быть использована для оценки риска наличия заболевания у пациента на основе показателей его биомаркеров.

В рамках проведенной работы на этапе разработки диагностически значимого комплекса показателей были использованы данные измерений 16 биомаркеров (AFP, CEA, CA 19-9, CA 125, HE4, tPSA, CA 15-3, B2M, hsCRP, Ddimer, CYFRA 21-1, ApoA1, ApoA2, Apo B, TTR, sVCAM-1), полученные на выборке здоровых добровольцев европеоидной популяции (n=203, 104 женщины и 99 мужчин 36-80 лет, средний возраст 53 года) и пациентов с КРР (n=102, 56 женщин и 47 мужчин 31-85 лет, средний возраст 63 года).

Статистическая обработка экспериментальных данных и разработка классификационных моделей проводилась в среде R {RDevelopmentCoreTeam (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0}.

Первым этапом являлся статистический анализ и визуализация данных. Для здоровых добровольцев было оценено влияние пола и возраста на показатели биомаркеров (Фиг.1).

По итогам проведения исследования на данном этапе был сделан вывод об отсутствии значимой корреляции между возрастом пациентов и показателями биомаркеров.

На следующем этапе проводилась оценка значимости различий в уровнях отдельных биомаркеров между здоровыми добровольцами и пациентами с КРР при помощи критерия Стьюдента после нормализации экспериментальных данных путем log-трансформации (Таблица 1). Результаты были сопоставлены с опубликованными ранее данными исследования, проведенного в госпитале Хенаня (Hou P, Luo J, Zhang J. A study of optimized model of serum tumor markers of colorectal cancer based on intelligent algorithms Int J Clin Exp Med 2016;9(9):17153-17164).

Таблица 2. Сравнение показателей биомаркеров в плазме крови здоровых доноров и пациентов с КРР

	Текущее исследование			Исследование, проведенное в госпитале Хенаня		
	Среднее значение		Р-значения ¹	Среднее значение		Р-значения ²
	Здоровые доноры	Пациенты с КРР		Здоровые доноры	Пациенты с КРР	
AFP, МЕ/мл	2.84	2.73	0.3	3.69	20.5	<0.05
АроА1, г/л	1.60	1.40	6e-10 ***	-	-	-
АроА2, г/л	0.30	0.23	2e-23 ***	-	-	-
АроВ, г/л	1.03	1.06	0.2	-	-	-
В2М, нг/мл	1477.29	2056.66	2e-19 ***	-	-	-
CA125, МЕ/мл	10.69	16.95	1e-05 ***	15.05	40.14	<0.05
CA15.3, МЕ/мл	15.13	15.27	0.6	15.69	20.96	<0.05
CA19.9, МЕ/мл	6.60	18.77	2e-04 ***	24.31	50.10	<0.05
СЕА, нг/мл	1.86	47.39	7e-12 ***	4.28	29.31	<0.05
CYFRA.21.1, нг/мл	1.37	4.04	3e-15 ***	1.98	8.75	<0.05
Ddimer, нг/мл	119.75	435.54	4e-15 ***	-	-	-
HE4, пмоль/л	51.43	76.75	5e-15 ***	-	-	-
hsCRP, мг/л	1.77	11.27	3e-19 ***	-	-	-
sVCAM.1, нг/мл	658.37	848.68	1e-09 ***	-	-	-
t.PSA, нг/мл	1.13	1.90	0.001 **	-	-	-
TTR, мг/дл	25.64	19.25	1e-12 ***	-	-	-

1 – сравнение по тесту Стьюдента после нормализации экспериментальных данных; ** - Р-значения<0.005, *** - Р-значения<0.001. 2 – метод сравнения в оригинальном исследовании не приведен.

На основании проведенного анализа был сделан вывод об отсутствии значимых различий в концентрациях AFP, АроВ, СА15.3 между здоровыми добровольцами и пациентами с КРР. В тоже время в исследовании, полученном на азиатской популяции, наблюдалось статистически значимое различие в уровнях биомаркеров AFP и СА15.3 между здоровыми добровольцами и пациентами с КРР, что указывает на необходимость учета межрасовых различий при разработке диагностических тестов.

Для оценки диагностической ценности отдельных биомаркеров использовался метод логистических регрессий. В данных статистических моделях рассматривалась взаимосвязь между концентрацией биомаркера и вероятностью наличия заболевания (уравнение 1):

$$P(Y) = 1 / (1 + e^{-(b_0 + b_1 \cdot X)}) \quad (1)$$

где $P(Y)$ – вероятность наличия заболевания, b_0 и b_1 – коэффициенты, определяемые по экспериментальным данным, X – предиктор (концентрация биомаркера).

Предсказательная способность логистических моделей оценивалась при помощи ROC-анализа, предполагающего определение чувствительности, специфичности и точности метода относительно тестового или общего набора данных. Для этого значение пороговой вероятности, определяющей наличие заболевания, варьировалось в пределах от 0 до 1 с заданным шагом, для каждого шага рассчитывалась доля верно диагностированных случаев заболевания (чувствительность) (S_e), правильно определенных случаев отсутствия заболевания (специфичность) (S_p), а так же общая доля правильно диагностированных случаев как наличия так и отсутствия заболевания (точность) (Acc), (уравнения 2-4):

$$S_e = TP / (TP + FN) \cdot 100\% \quad (2)$$

$$S_p = TN / (TN + FP) \cdot 100\% \quad (3)$$

$$Acc = (TN + TP) / (TN + FP + TP + FN) \cdot 100\% \quad (4)$$

где **TP** – верно классифицированный положительный результат (верно диагностированное заболевание), **FP** – ложноположительный результат (ошибочно диагностированное заболевание), **TN** – верно классифицированный отрицательный результат (верно диагностированное отсутствие заболевания), **FN** – ложноотрицательный результат (ошибочно диагностированное отсутствие заболевания).

Полученный набор значений чувствительности и специфичности использовался для построения ROC-кривой. В качестве интегрального показателя качества моделей использовалась площадь под ROC-кривой (AUROC): предикторы с максимальной

предиктивной способностью показывают наибольшие значения AUROC. Результаты ROC-анализа приведены на фиг. 2 и таблице 2.

Таблица 2. Диагностическая ценность отдельных биомаркеров

Биомаркер	Специфичность, %	Чувствительность, %	Точность, %	AUROC
ApoA2	94	65	84	0.87
B2M	80	73	78	0.83
CYFRA.21.1	79	72	77	0.82
Ddimer	64	82	70	0.8
HE4	67	80	72	0.79
hsCRP	85	62	77	0.79
TTR	85	57	76	0.77
CEA	92	55	79	0.75
sVCAM.1	86	50	74	0.72
ApoA1	67	68	67	0.71
t.PSA	80	63	74	0.7
CA19.9	86	44	72	0.66
CA125	58	65	60	0.64
Rantes	68	60	65	0.64
ApoB	64	50	59	0.55
AFP	49	65	54	0.55
CA15.3	76	32	61	0.52
VEGFR1	87	26	67	0.52

На основе результатов статистического анализа данных и оценки предсказательной способности однофакторных логистических моделей были отобраны биомаркеры, которые впоследствии были включены в многофакторные классификационные модели. Критерием включения биомаркеров являлись $pval < 0.001$ (Таблица 1) и $AUROC > 0.6$ (Таблица 2). t.PSA не был включен в многофакторные классификаторные модели, так как данный биомаркер не измеряется у женщин и потому содержащий t.PSA не мог бы быть использован у данной группы обследуемых. Таким образом, для построения классификационных моделей были отобраны экспериментальные измерения следующих биомаркеров: ApoA2, B2M, CYFRA.21.1, Ddimer, HE4, hsCRP, TTR, CEA, sVCAM.1, ApoA1, CA19.9, CA125.

Разработка многофакторных классификационных моделей являлась завершающим этапом исследования. Различные способы машинного обучения (randomforest, линейный дискриминантный анализ, метод опорных векторов) были использованы в рамках текущей задачи. Оценка параметров моделей (обучение), производилась на объединённых данных, полученных на здоровых добровольцах и пациентах с КРР, и была направлена на минимизацию предсказательных ошибок алгоритма.

Метод «random forest» (RF) подразумевает создание совокупности кросс-валидированных решающих деревьев. Каждое из таких деревьев проходит обучение на подвыборке данных, включающей информацию лишь по части биомаркеров и наблюдений, и валидируется на подвыборке, не использованной для его построения (бэггинг). На основании предсказаний каждого из построенных деревьев решений пациент причисляется к одной из групп (здоровые доноры или пациенты с КРР), финальное предсказание классификатора определяется большинством голосов построенных деревьев (см. фиг.3 А, Б).

Использование линейного дискриминантного анализа (LDA) предполагает поиск линейной комбинации биомаркеров - дискриминанты, обеспечивающей наилучшее разделение всей популяции обследуемых на здоровых добровольцев и пациентов с КРР. Данная задача решается за счет нахождения оси, проекция на которую обеспечивает максимальное отношение общей дисперсии линейной комбинации биомаркеров выборки к сумме дисперсий линейной комбинации биомаркеров внутри классов (см.фиг.4).

Таблица 3. Значения линейных коэффициентов при дискриминанте (LDAcomponent 1)

Биомаркер	Коэффициент
SEXM	2.77E-01
CEA	-2.84E-03
CA19.9	6.75E-03
CA125	1.12E-02
HE4	-4.63E-03
B2M	1.43E-03
hsCRP	8.34E-03
Ddimer	6.68E-04
CYFRA.21.1	2.64E-02
ApoA1	1.15E+00
ApoA2	-1.88E+01
TTR	-5.82E-03
sVCAM.1	1.93E-04

Использование метода опорных векторов (SVM) предполагает нахождение $(n-1)$ -мерной гиперплоскости, разделяющей n -мерное пространство значений биомаркеров на два класса. Обучение алгоритма подразумевает поиск такой гиперплоскости, которая обеспечивает наименьшую эмпирическую ошибку классификации и максимизирует расстояние между значениями биомаркеров пациентов, относящихся к разным классам (см. фиг.5).

На первом этапе построения многофакторных моделей проводилось изучение диагностической ценности различных комбинаций биомаркеров из приведенной выше группы. Для этого все возможные комбинации, включающие от 2 до 12 биомаркеров были использованы для построения классификационных моделей (4803 варианта). Для обучения использовались объединённые данные, полученные на здоровых добровольцах и пациентах с КРР, и методы линейного дискриминантного анализа и опорных векторов. Разработанные модели были ранжированы в соответствии с их предсказательным потенциалом, оцененным по показателю AUROC (фиг. 7, таблица 4).

Как видно из фиг. 6, наибольшей предсказательной способностью обладают комплексные тесты, включающие 11-13 биомаркеров, в то время как для относительно небольшой доли классификаторов, включающих комбинации из 2-3 биомаркеров, показатель AUROC составляет более 80%.

Финальной фазой построения классификаторов являлась их валидация.

Объединённые данные, полученные на здоровых добровольцах и пациентах с КРР были случайным образом разделены на обучающую и тестовую выборки. Оценка параметров моделей (обучение), производилась на обучающей выборке и была направлена на минимизацию предсказательных ошибок алгоритма. Валидация обученных моделей заключалась в оценке их предсказательной способности на тестовой выборке. Предсказательная способность многофакторных классификационных моделей оценивалась при помощи ROC-анализа как это было сделано ранее для отдельных биомаркеров (фиг. 7, Таблица 4).

Таблица 4. Диагностическая ценность обученных многофакторных классификационных моделей

Метод	Чувствительность, %	Специфичность, %	Точность, %	Пороговая вероятность, %	AUROC
Весь набор данных был использован как для обучения модели, так и для ее валидации					
RF	100	100	100	55	1
LDA	88	90	89	79	0.95
SVM	86	96	92	55	0.95
80% данных было использовано для обучения модели, 20% - для валидации					
RF	89	91	90	66	0.94
LDA	78	95	90	48	0.91
SVM	88	88	88	53	0.9

Финальные классификационные модели представляют собой обученные алгоритмы, позволяющие предсказать вероятность наличия колоректальной карциномы на основании экспериментальных измерений биомаркеров пациентов.

Финальное решение - определение вероятности наличия колоректального рака, рассчитывается как медиана значений вероятностей KPP, рассчитанных в 3 классификационных моделях (RF, LDA SVM), обученных на всей выборке пациентов, см. например, Kittler J, Hatef M, Duin RPW et al, On Combining Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 20, NO. 3, MARCH 1998 226-39.

Для реализации заявляемого способа и было разработано программное обеспечение (ПО), позволяющее на основе данных конкретного пациента (пол и результаты измерения биомаркеров) рассчитывать вероятность наличия у него KPP. Блок-схема реализации изобретения представлена на фиг. 8.

Компьютерно-реализуемая система состоит из (1) интерфейса, включающего устройство ввода данных пациента (пол и результаты измерений биомаркеров) и вывода результатов расчета (вероятность наличия KPP); (2) блока памяти, содержащего обученные классификаторные модели и программные продукты, необходимые для работы с ними (R portable, Google Chrome Portable) и (3) программного модуля, с помощью которого реализуется программный код, необходимый для обмена данных между интерфейсом и блоком памяти. Для создания графического интерфейса был использован пакет shiny (Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2017). shiny: Web Application Framework for R. R package version 1.0.5. <https://CRAN.R->

[project.org/package=shiny](https://CRAN.R-project.org/package=shiny)) созданный на базе среды R {RDevelopmentCoreTeam (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0}. Для работы с данным пакетом необходимо наличие программных продуктов R portable и Google Chrome portable, хранящихся в блоке памяти. Для работы с предложенными моделями необходимы следующие пакеты: (1) RandomForest (A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18-22); (2) MASS (Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0); (3) e1071 (David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel and Friedrich Leisch (2017). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-8. <https://CRAN.R-project.org/package=e1071>).

Алгоритм оценки вероятности наличия КРР на основе данных пациента представлен на фиг.9.

Данные пациента вводятся через интерфейс и подаются в качестве входных переменных в разработанные модели, в каждой из которых производится расчет вероятности наличия КРР. Далее по результатам модельных предсказанной рассчитывается среднее значение, которое выводится в окно вывода.

Диагностическая мультиплексная панель для оценки риска КРР включает биомаркеры, показавшие максимальный предсказательный потенциал в рамках проводимого исследования (рис. 2, таблица 2): ApoA2, B2M, Ddimer. Кроме того, в заявляемый комплекс включены дополнительные биомаркеры, обладающие меньшим предсказательным потенциалом, однако значимо различные между здоровыми добровольцами и пациентами с КРР (Таблица. 1): HE4, hsCRP, sVCAM1, ApoA1, TTR и СА125 в исследуемой популяции.

Ниже представлены клинические примеры применения способа.

Пример 1 Больной С., 56 лет.

В феврале 2018 года в связи с жалобами на дискомфорт в гипогастрии, слабость и быструю утомляемость обратился в поликлинику по месту жительства.

Был осмотрен терапевтом. Рекомендован прием витаминов, общий анализ крови, в котором клинически значимых отклонений не было выявлено.

Пациенту было предложено принять участие в программе Онкопоиска.

Пациент обследован в рамках программы. Получены следующие результаты: СЕА 69,44 нг/мл, СА 19-9 60,78 МЕ/мл, СА 125 17,75 МЕ/мл, HE4 111,6 пмоль/л, B2M 2800 нг/мл, hsCRP 85 нг/мл, Ddimer 1123 нг/мл, CYFRA 21.1 4,03 нг/мл, Apo A1 1,11 г/л, Apo A2 0,175 г/л, TTR (prealb) 15 мг/дл, sVCAM.1 1125 нг/мл.

При обработке полученных результатов заявляемым способом выявлена высокая вероятность колоректального рака.

Выполнено обследование: при колоноскопии выявлена экзофитная опухоль около 1 см в сигмовидной кишке.

При гистологическом исследовании: умеренно дифференцированная аденокарцинома. Пациент был прооперирован. Выполнена резекция сигмовидной кишки.

При плановом гистологическом исследовании опухоль в сигмовидной кишке имеет строение умеренно дифференцированной аденокарциномы, прорастающая в мышечный слой стенки кишки. В краях резекции опухолевого роста не обнаружено. Отмечается полнокровие сосудов и отек. В одном из 5 удаленных л/узлов брыжейки - метастаз рака аналогичного строения.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ скринингового определения вероятности наличия колоректального рака, включающий измерение уровня биомаркеров в образце биологической жидкости, полученном у субъекта: ApoA2, B2M, CYFRA.21.1, Ddimer, HE4, hsCRP, TTR, CEA, sVCAM.1, ApoA1, CA19.9, CA125, с последующей обработкой совокупности полученных значений биомаркеров с использованием, по меньшей мере, одной классификационной модели, обученной для определения высокой или низкой вероятности наличия колоректального рака.
2. Способ по п.1, характеризующийся тем, что в качестве классификационных моделей используют метод «случайного леса» (random forest), и/или линейный дискриминантный анализ, и/или метод опорных векторов.
3. Способ по п.1, характеризующийся тем, что обученную классификационную модель получают посредством реализации следующих шагов:
 - формируют обучающую и тестовую выборку записей субъектов с измеренными значениями биомаркеров ApoA2, B2M, CYFRA.21.1, Ddimer, HE4, hsCRP, TTR, CEA, sVCAM.1, ApoA1, CA19.9, CA125, включающие записи о пациентах разного пола и возраста;
 - обучают классификационную модель выявлению заданной патологии, используя записи обучающей и тестовой выборки;
 - сохраняют связи и веса обученной классификационной модели, для последующего определения вероятности наличия колоректального рака по итогам обработки измеренных данных биомаркеров субъекта.
4. Способ по п. 3, характеризующийся тем, что при формировании обучающей и тестовой выборки, включают записи субъектов с выявленной патологией - наличие колоректального рака и отсутствие колоректального рака.
5. Система скринингового определения вероятности наличия колоректального рака, включающая
 - модуль ввода измеренных значений биомаркеров субъекта;
 - модуль хранения данных, выполненный с возможностью хранения обучающей и тестовой выборки классификационной модели, связей и весов обученной классификационной модели, записей субъектов с измеренными значениями биомаркеров ApoA2, B2M, CYFRA.21.1, Ddimer, HE4, hsCRP, TTR, CEA, sVCAM.1, ApoA1, CA19.9, CA125, включающие записи о пациентах разного пола и возраста;

- модуль обученной классификационной модели, выполненный с возможностью построения и обучения, по меньшей мере, одной классификационной модели для определения наличия заданной патологии по упомянутым маркерам, взятым из модуля хранения данных;
- модуль диагностики, выполненный с возможностью обработки введенных значений биомаркеров субъекта с использованием, по меньшей мере, одной обученной классификационной модели;
- модуль вывода данных, выполненный с возможностью получения данных о высокой или низкой вероятности наличия колоректального рака.