

СТАТИСТИЧЕСКИ ЗНАЧИМЫЕ ПОПУЛЯЦИИ Т-КЛЕТОК ПРИ ДИАГНОСТИКЕ РАССЕЯННОГО СКЛЕРОЗА

© 2019 г. Я. В. Козичук¹, А. Г. Ильвес², И. В. Кудрявцев^{3*},
Д. А. Москаленко¹, О. М. Новоселова², К. С. Рубаник², М. К. Серебрякова³,
Д. О. Старов¹, Б. А. Тимченко¹, Л. Н. Прахова², И. С. Лобанов¹

*E-mail: igorek1981@yandex.ru

¹Университет ИТМО, Санкт-Петербург, Россия;

²ФГБНУ Институт мозга человека им. Н. П. Бехтерева РАН, Санкт-Петербург, Россия;

³ФГБНУ «Институт экспериментальной медицины», Санкт-Петербург, Россия

Поступила: 12.11.2018. Принята: 21.01.2019

Цель исследования: поиск статистически значимых популяций Т-клеток при постановке диагноза рассеянный склероз (РС). С применением многоцветной проточной цитометрии был проведен анализ абсолютного содержания различных субпопуляций Т-лимфоцитов (Т-хелперов, Тх и цитотоксических Т-клеток, Тцит) в периферической крови 61-го условно здорового добровольца и 47-ми больных РС. На основании экспрессии маркеров дифференцировки (CD45RA, CD62L, CD27 и CD28) и эффекторных молекул (CD56 и CD57) Тх и Тцит были разделены на основные популяции, находящиеся на разных стадиях созревания. Выявлены следующие статистически значимые популяции Т-клеток: CD56⁺CD57⁺ Т-лимфоциты, EM Тх, EM3 Тцит, EM Тцит, CD56⁺CD57⁻ Т-лимфоциты, EM2 Тцит. Значимость также была подтверждена при расчете статистики Хи-квадрат. На основе полученной информации было выбрано три группы популяций Т-клеток. На каждой группе популяций была построена модель для диагностики рассеянного склероза на базе алгоритма k ближайших соседей. Точность предсказания построенных моделей варьируется в пределах 0.69–0.90.

Ключевые слова: рассеянный склероз, машинное обучение, алгоритм k-ближайших соседей, Т-клетки, проточная цитометрия

DOI: 10.31857/S102872210005023-0

Адрес: 197376, Санкт-Петербург, ул. академика Павлова, 12. ФГБНУ «Институт экспериментальной медицины», Кудрявцев Игорь Владимирович. Тел.: +78122342929.

E-mail: igorek1981@yandex.ru

Авторы:

Козичук Я. В., студент магистратуры, факультет Систем Управления и Робототехники, Университет ИТМО, Санкт-Петербург, Россия;

Ильвес А. Г., к.м.н., с.н.с. лаборатории нейроиммунологии ФГБНУ Институт мозга человека им. Н. П. Бехтерева РАН, Санкт-Петербург, Россия;

Кудрявцев И. В., к.б.н., с.н.с. отдела иммунологии ФГБНУ «Институт экспериментальной медицины», Санкт-Петербург, Россия;

Москаленко Д. А., студент магистратуры, факультет технологического менеджмента и инноваций, Университет ИТМО, Санкт-Петербург, Россия;

Новоселова О. М., м.н.с. лаборатории нейрореабилитации ФГБНУ Институт мозга человека им. Н. П. Бехтерева РАН, Санкт-Петербург, Россия;

Рубаник К. С., м.н.с. лаборатории нейрореабилитации ФГБНУ Институт мозга человека им. Н. П. Бехтерева РАН, Санкт-Петербург, Россия;

Серебрякова М. К., н.с. отдела иммунологии ФГБНУ «Институт экспериментальной медицины», Санкт-Петербург, России;

Старов Д. О., студент магистратуры, факультет Систем Управления и Робототехники, Университет ИТМО, Санкт-Петербург, Россия;

Тимченко Б. А., студент магистратуры, факультет Систем Управления и Робототехники, Университет ИТМО, Санкт-Петербург, Россия;

Прахова Л. Н., д.м.н., заведующая лабораторией нейрореабилитации ФГБНУ Институт мозга человека им. Н. П. Бехтерева РАН, Санкт-Петербург, Россия;

Лобанов И. С., к.ф.-м.н. доц. факультета Систем Управления и Робототехники, Университет ИТМО, Санкт-Петербург, Россия.

ВВЕДЕНИЕ

Раннее выявление любого заболевания является залогом успешного лечения и выздоровления пациента, поэтому в современном мире значительные усилия прилагаются к со-

вершенствованию медицинской диагностики. В 2016 году в Российской Федерации было выявлено 519 тысяч случаев злокачественных новообразований, что выше показателя 2016 года на 3% [1]. «Чума XX века» — вирус иммунодефицита человека — всего за 100 лет унес более 20 миллионов жизней. Более 35 миллионов человек в мире в настоящее время стоят на учете с данным диагнозом [2]. В настоящее время существуют ряд заболеваний, тяжело поддающихся диагностике, что частично объясняется недостатком знаний о механизме возникновения заболеваний. К числу таких заболеваний относятся и аутоиммунные заболевания. Для их диагностики нужны новые методы.

Ряд задач, для которых человеку тяжело предложить алгоритм решения, может быть решен методами машинного обучения. Есть примеры успешного использования данных методов в таких областях, как интернет-маркетинг, банковская сфера, сфера связи, а также медицинская диагностика. Современные корпорации, включая такие технологические гиганты, как Google и Amazon, активно проводят исследования в данной области и достигают определенных успехов. Например, сотрудники Google в 2018 году опубликовали работу, в которой они при помощи методов глубокого обучения прогнозируют риски сердечно-сосудистых заболеваний на основе снимков глазного дна [3]. Исследователи из Стэнфорда разработали программный комплекс, который самостоятельно диагностирует пневмонию, используя рентгеновские снимки грудной клетки, причем точность предсказания сравнима с навыками опытного рентгенолога [4]. Группа немецких ученых в 2014 году опубликовала работу, в которой был представлен алгоритм, детектирующий циркулирующие опухолевые клетки в периферической крови [5].

В данной работе мы будем рассматривать рассеянный склероз, являющийся представителем аутоиммунных болезней. Рассеянный склероз (РС) — это широко распространенное нейродегенеративное хроническое заболевание центральной нервной системы. Данному заболеванию, в основном, подвержены лица в возрасте от 10 до 59 лет [6]. В современном мире наблюдается увеличение числа случаев РС, что обусловлено улучшением качества диагностики, повышением возможностей терапии, а также в связи с истинным увеличением больных по неизвестным причинам. По данным Международной Федерации больных рассеянным склерозом в период

с 2008 по 2013 гг глобальная распространенность РС возросла с 30 до 33 случаев на 100 000 населения [7].

Существует ряд научных трудов, в которых рассеянный склероз изучается методами машинного обучения. В 2007 году была опубликована работа, в которой группа ученых из Индии построила симуляцию распространения аутоиммунных очагов демиелинизации в центральной нервной системе [8]. В 2017 году группа ученых из Франкфуртского университета им. Гёте представила работу, в которой они исследовали применимость методов машинного обучения в прогнозировании рассеянного склероза. В качестве исходных данных были использованы демографические и клинические данные, а также результаты МРТ [9]. В этом же году была опубликована работа, где европейские ученые пытались найти статистически значимые различия между формами РС, используя данные МРТ головного мозга [10].

Большинство исследований в данной области сфокусировано на обработке результатов магнитно-резонансной томографии, демографических, клинических, неврологических показателей или их совокупности [11–15]. Но так как РС является системным заболеванием, то при его патогенезе присутствует и изменение иммунологических показателей, таких как размеры популяций Т-клеток.

В данной работе мы рассмотрим различные методы машинного обучения в прогнозировании рассеянного склероза на основе результатов анализа абсолютного содержания различных популяций Т-лимфоцитов у больных с диагнозом рецидивирующего рассеянного склероза.

Целью исследования являлся поиск статистически важных типов Т-лимфоцитов при первичной диагностике рассеянного склероза на основе анализа крови.

МАТЕРИАЛЫ И МЕТОДЫ

Объектом исследования служила венозная кровь 47-ми больных с рецидивирующим РС, полученная путем пункции периферической вены и люмбальной пункции, соответственно, и собранная в вакуумные пробирки с содержанием КЗЭДТА. Диагноз РС устанавливался согласно критериям МакДональда [16]. В исследование вошли пациенты, никогда не получавшие препаратов, изменяющих течение РС (ПИТРС) и системных кортикостероидов в течение последних 90 дней. Всем пациентам проводился стандарт-

ный неврологический осмотр с последующей балльной оценкой степени поражения функциональных систем: зрительной, пирамидной, мозжечковой, чувствительной, функций тазовых органов, ствола мозга. Также подсчитывался балл по интегративной шкале EDSS — расширенной шкале инвалидизации [17]. В группу сравнения вошли образцы периферической крови от 61-го условно здорового добровольца. Обе группы пациентов не различались по половому и возрастному составу. Исследования были проведены с информированного согласия испытуемых и в соответствии с Хельсинкской декларацией Всемирной ассоциации «Этические принципы проведения научных медицинских исследований с участием человека» с поправками 2000 г. и «Правилами клинической практики в Российской Федерации», утвержденными Приказом Минздрава РФ от 19.06.2003 г. № 266. Исследование одобрено Комиссией по этике ФГБУН «Институт мозга человека им. Н. П. Бехтеревой» РАН 23 октября 2014 года.

Для выявления основных популяций Т-лимфоцитов периферической крови использовалась следующая панель моноклональных антител, конъюгированных с различными флуорохромами (все антитела производства Beckman Coulter, США): CD57-FITC (клон NC1, кат. № IM0466U), CD56-PE (клон N901 (NKH-1), кат. № A07788), CD62L-ECD (клон DREG56, кат. № IM2713U), CD28-PC5 (клон CD28.2, кат. № 6607108), CD27-PC7 (клон 1A4CD27, кат. № A54823), CD3-APC (клон UCNT1, кат. № IM2467), CD8-APC-Alexa Fluor 700 (клон B9.11, кат. № A66332), CD45RA-APC-Alexa Fluor 750 (клон 2H4LDH11LDB9 (2H4), кат. № A86050), CD4-Pacific Blue (клон 13B8.2, кат. № A82789), CD45-Krome Orange (клон J.33, кат. № A96416). Указанным коктейлем антител окрашивали 100 мкл периферической крови в соответствии с рекомендациями производителя. Для удаления эритроцитов из образцов периферической крови использовался лизирующий раствор VersaLyse Lysing Solution (кат. № A09777), к 975 мкл которого *ex tempera* добавляли 25 мкл фиксирующего раствора IOtest 3 Fixative Solution (кат. № A07800). Поиск абсолютное содержания клеток исследованных популяций Т-лимфоцитов проводили с помощью реагента FlowCount™ (Beckman Coulter, США).

При построении модели для диагностики РС применялась фенотипическая характеристика основных популяций Т-лимфоцитов, причем

их абсолютное значение (кол-во клеток на моль) из общего пула циркулирующих Т-лимфоцитов. Т-лимфоциты определяли как CD3⁺ лимфоциты. В дальнейшем, на основании коэкспрессии CD3 и CD4 выявляли Т-хелперы (Тх), а на основании коэкспрессии CD3 и CD8 — цитотоксические Т-лимфоциты (Тцит). Тх и Тцит разделяли на «наивные» клетки («N», CD45RA⁺CD62L⁺), клетки центральной («CM», CD45RA⁺CD62L⁺) и эффекторной («EM», CD45RA⁺CD62L⁺) памяти, а также «терминально-дифференцированные» CD45RA-позитивные клетки эффекторной памяти («TEMRA», CD45RA⁺CD62L⁺). При анализе Тцит на основании уровня экспрессии ко-стимулирующих молекул CD27 и CD28 среди EM клеток выделяли популяции EM1, EM2, EM3 и EM4 с фенотипами CD27⁺CD28⁺, CD27⁺CD28⁻, CD27⁻CD28⁻ и CD27⁻CD28⁺, соответственно. Среди TEMRA Тцит также определяли содержание «пре-эффекторов» 1 типа («pE1», CD27⁺CD28⁺), «пре-эффекторов» 2 типа («pE2», CD27⁺CD28⁻) и зрелых эффекторных клеток («E», CD27⁻CD28⁻), как это было детально описано ранее [17]. Более того, на общей популяции Т-лимфоцитов определяли уровни коэкспрессии эффекторных молекул CD56 и CD57, разделяя все CD3⁺ лимфоциты периферической крови на четыре субпопуляции с фенотипами CD56⁻CD57⁺, CD56⁺CD57⁺, CD56⁺CD57⁻ и CD56⁻CD57⁻.

Модель для диагностики рассеянного склероза мы строим на базе алгоритма к ближайших соседей (kNN), который успешно справляется с различными задачами машинного обучения — классификация, регрессия, кластеризация. Задачу, поставленную в данной работе, можно формализовать как задачу бинарной классификации. Классификация — раздел машинного обучения, который занимается поиском наилучшей стратегии присвоения классов объектам на основе прошлых наблюдений пар типа объект-класс [18].

Данный алгоритм относится к классу метрических. Такие алгоритмы базируются на поиске сходства между изучаемыми объектами. Если мера сходства введена успешно, то схожим объектам удастся присвоить одинаковые классы. Для формализации понятия сходства вводится метрика $p(u, x)$ — обобщение понятия расстояния между u — изучаемым объектом и x — объектом из выборки в пространстве объектов X . Данная метрика задается различными способами: Манхэттенское расстояние, Евклидово расстояние, расстояние Левенштейна, и прочие.

При классификации нового объекта, алгоритм kNN работает следующим образом: исследуемый объект относится к наиболее часто встречающемуся классу среди его k ближайших соседей в заданной метрике.

В данной задаче в качестве объекта выступает пациент, а в качестве признаков объекта – абсолютные значения популяций Т-лимфоцитов. Для поиска статистически важных признаков мы будем использовать критерий Хи-квадрат, который позволяет найти значимость связи между двумя переменными. На основе данного метода мы можем найти значимость признаков относительно класса объекта. Использование при построении алгоритмов более значащих признаков повышает его качество. Построение моделей было осуществлено при помощи языка программирования Python 3 и библиотека scikit-learn-0.19.1.

РЕЗУЛЬТАТЫ

После проведения процедуры отбора значимых признаков при помощи критерия Хи-квадрат, обнаружено, что значимыми показателями для построения моделей диагностики рассеянного склероза являются: (см. **рис. 1**). На представленной гистограмме большие значения у следующих субпопуляций: CD56⁺CD57⁺ Т-лимфоциты, EM Tx, EM3 Тцит, EM Тцит, CD56⁺CD57⁺ Т-лимфоциты, EM2 Тцит.

Так, абсолютное содержание Т-лимфоцитов с фенотипом CD56⁺CD57⁺ в группе больных РС снижалось ($p=0,015$) с 117 (69; 196) до 80 (45; 113) клеток в 1 мкл периферической крови. Параллельно отмечалось уменьшение ($p<0,001$) концентрации EM Tx, концентрация которых у больных РС находилось в пределах 138 (100; 198) кл/мкл, а в группе сравнения составляло 207 (157; 248) кл/мкл. Кроме того, третьим по значимости показателем в рамках использованной модели являлось абсолютное содержание цитотоксических Т-лимфоцитов популяции EM3. Уровень этих клеток в периферической крови больных РС был значимо ниже такого условно здоровых добровольцев (22 (11; 33) и 28 (14; 71) клеток в 1 мкл периферической при $p=0,046$, соответственно).

Во вторую группу значимых признаков попали абсолютные значения по содержанию в периферической крови EM Тцит, Т-клеток с фенотипом CD56⁺CD57⁺ и Тцит популяции EM1. Так, уровень EM Тцит у больных РС был ниже значений, полученных для группы сравнения (117 (81; 182) и 163 (119; 204) кл/мкл при $p=0,005$, соответственно). Тогда как абсолютное содержание Т-лимфоцитов, экспрессирующих CD56, но негативных по CD57, при РС возрастало при сравнении с контролем (40 (20; 65) и 26 (17; 40) кл/мкл при $p=0,008$, соответственно). Уровень CD27⁺CD28⁺ EM Тцит при

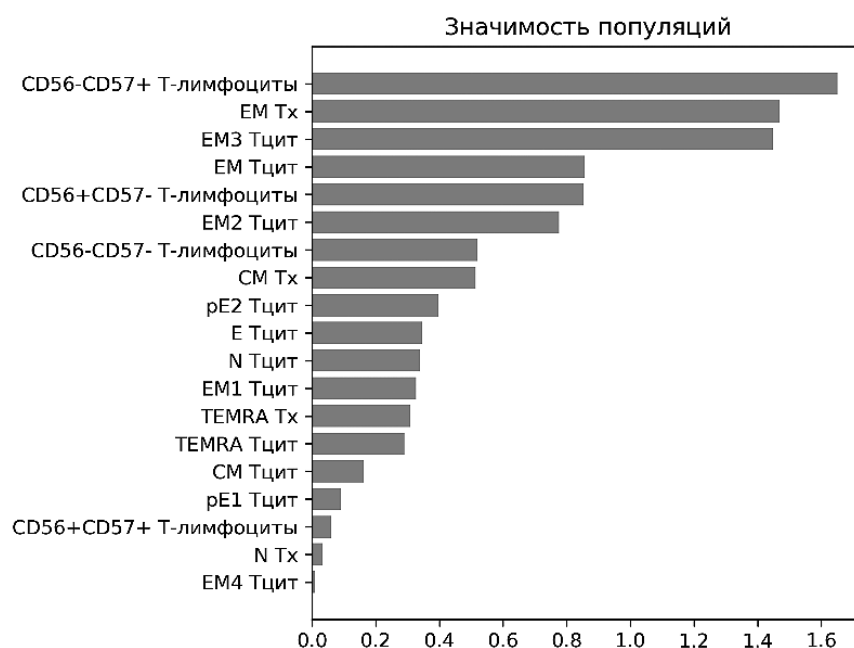


Рис 1. Значимость популяций, рассчитанных на основе статистики Хи-квадрат

РС достоверно ($p=0,002$) понижался с 23 (15; 29) до 15 (9; 25) кл/мкл.

На основе этой информации было построено 3 модели на основе алгоритма kNN для 3, 6 и 20 признаков. Исходные данные были нормированы алгоритмом мини-макс. Необходимость данной нормировки обусловлена тем, что при ее отсутствии признаки с наибольшим разбросом значений будут доминировать в метрике, а остальные практически не будут иметь влияния.

В процессе обучения, для контроля качества получаемой модели был использован метод скользящего контроля [20]. В качестве реализации метода нами был выбран так называемый q-fold, где параметр q мы взяли равным 10. Для оценки качества моделей мы использовали точность, чувствительность и F-меру. Они рассчитываются на основе трех показателей: количество истинно-положительных ответов (ИП), количество истинно-негативных ответов (ИН), количество ложно-позитивных ответов (ЛП). Их формулы расчета выглядят следующим образом:

$$\text{Точность} = \frac{\text{ИП}}{\text{ИП} + \text{ЛП}}$$

$$\text{Чувствительность} = \frac{\text{ИН}}{\text{ИН} + \text{ЛП}}$$

$$F\text{-мера} = \frac{2 * \text{точность} * \text{чувствительность}}{\text{точность} + \text{чувствительность}}$$

Результаты работы алгоритма k-средних зависят от числа соседей k и от метрики $p(u, x)$. Мы оптимизировали параметры алгоритма, методом случайного поиска [21]. В результате мы получили модели со следующими характеристиками (см. таб. 1)

ОБСУЖДЕНИЕ

Приведенные выше результаты указывают на то, что значимыми показателями для выявления больных с РС являются абсолютное содержание в периферической крови лимфоцитов с фенотипами $\text{CD3}^+\text{CD56}^-\text{CD57}^+$, $\text{CD3}^+\text{CD4}^+\text{CD45RA}^-\text{CD62L}^-$ и $\text{CD3}^+\text{CD8}^+\text{CD45RA}^-\text{CD62L}^-\text{CD27}^-$

CD28^- . CD57 , также известный под названием HNK1 (от англ. «human natural killer-1»), являющийся углеводным эпитопом, представленным на цитоплазматической мембраны клеток, обладающих выраженной цитолитической активностью [22]. Уровень CD57 на поверхности Т-лимфоцитов тесно связан со способностью этих клеток накапливать в составе цитоплазматических гранул ключевые эффекторные молекулы – перфорин и гранзимы [23]. Эти результаты, полученные при помощи проточной цитофлуориметрии, подтверждаются данными молекулярно-биологических исследований [24]. Так, CD57^+ цитотоксические Т-клетки активно экспрессировали гранзим В, гранулизин и перфорин при сравнении с CD57 -негативными клетками аналогичной популяции. Таким образом, использование CD57 в качестве «суррогатного» маркера позволяет без использования трудоемких методов окраски на внутриклеточные антигены выявить эффекторные цитотоксические клетки, содержащие в цитоплазме необходимый набор цитолитических молекул. Изменения в содержании зрелых $\text{CD3}^+\text{CD57}^+$ лимфоцитов в периферической крови больных указывают на диагностическую значимость данной популяции клеток в диагностике РС.

Исследования ЕМ Тх показали, что именно в составе данной популяции Тх присутствуют основные клетки-продуценты ключевых цитокинов – $\text{IFN}\gamma$ и IL-4 [25]. Более того, большинством исследователей лимфоциты именно данного фенотипа рассматривают в качестве зрелых или «эффекторных» Тх различных типов [26]. Особый интерес при исследовании РС представляет тот факт, что дифференцировка и «поляризация» в сторону Th17 ЕМ – лимфоцитов, играющих ведущую роль в патогенезе РС – у человека находится под контролем провоспалительных цитокинов IL-1 β и IL-6 [27]. Уровень экспрессии этих цитокинов лимфоцитами, равно как и их содержание в сыворотке крови и спинномозговой жидкости больных РС тесно связаны с тяжестью заболевания и его прогрессией [28].

Таблица 1. Характеристики полученных моделей на основе алгоритма kNN.

Кол-во признаков	Точность	Чувствительность	F1-мера	Кол-во соседей	Метрика
3	0.90	0.90	0.90	21	Манхэттенское расстояние
6	0.77	0.75	0.76	25	Евклидово расстояние
20	0.69	0.71	0.69	25	Евклидово расстояние

С другой стороны, столь же существенную роль в формировании пула эффекторных Th17 играет IL-23 [29], сывороточные концентрации которого у больных РС существенно превосходят таковые условно здоровых доноров [30]. Ключевыми эффекторными цитокинами лимфоцитов данной популяции являются IL-17A, IL-17F и IL-22, роль которых в патогенезе РС (в силу их высокой концентрации у больных) обсуждается уже длительное время [31]. При анализе ЕМ Т-хелперов – лимфоцитов, мигрирующих в периферические ткани и обладающих выраженными эффекторными свойствами – показано, что снижение уровня CCR6⁺ Th17 именно в рамках ЕМ клеток в циркуляции у больных РС коррелировало с увеличением инвалидизации ($r = -0,489$ при $p = 0,002$) [32].

Что же касается цитотоксических Т-клеток популяции ЕМ3, то они рассматриваются в качестве высоко дифференцированной популяции общего пула ЕМ клеток [33]. «Незрелые» Тцит с фенотипом CD27⁺CD28⁺ в ходе дифференцировки сначала теряют CD28, приобретая фенотип CD27⁺CD28⁻, тогда как зрелые клетки обладают фенотипом CD27⁻CD28⁻. Параллельно с потерей костимулирующих молекул Тцит, начинают экспрессировать цитолитические молекулы и обретают способность проявлять *ex vivo* цитотоксические свойства по отношению к клеткам-мишеням [34]. ЕМ3 клетки также способны экспрессировать гены цитолитических факторов и проявлять цитотоксические свойства в условиях *in vitro*, причем уровень активности этих клеток примерно в 10 раз превышал таковой клеток ЕМ1 и не уступал зрелым эффекторным клеткам популяции ТЕМРА [35]. Динамика изменения ЕМ1-ЕМ4 была изучена нами в периферической крови больных РС с длительными сроками заболевания [36]. Так, было показано, что у данной группы больных наблюдается достоверное снижение ($p < 0,001$) относительного содержания ЕМ1 клеток, коэкспрессирующих CD27 и CD28, с 18,96% (14,15; 25,05) до 12,21% (9,69; 20,55). Параллельно наблюдалось накопление ЕМ3 Тцит – в рамках группы условно здоровых доноров ЕМ3 составляли 4,79% (2,69; 14,31), тогда как у больных РС их уровень в циркуляции повышался до 8,46% (4,96; 18,98) от общего числа клеток с фенотипом CD3⁺CD8⁺. По-видимому, на ранних сроках после постановки диагноза РС абсолютное содержание ЕМ3 Тцит может снижаться, тогда как по мере прогрессирования РС уровень этих клеток в крови существенно повы-

шается, что еще раз указывает на диагностическую значимость этой популяции лимфоцитов.

Что же касается общей популяции ЕМ Тцит, то в литературе можно найти весьма противоречивые данные об изменениях этих клеток при РС. Так, некоторые исследователи указывают на то, что в крови больных наблюдалось достоверное увеличение общего пула ЕМ Тцит с фенотипом CD45RA⁻CCR7⁻ [37]. Сходные результаты были получены и другой группой исследователей, которые тоже отметили увеличение уровня ЕМ CD3⁺CD8⁺ лимфоцитов, что, по мнению авторов, указывало на системную активацию данного звена приобретенного иммунитета [38]. С другой стороны, Lui и соавторы отметили, что у больных РС среди цитотоксических Т-клеток существенно увеличивались уровни CM и ТЕМРА Тцит, что могло свидетельствовать о наличии хронической воспалительной реакции у данных пациентов [39]. Результаты собственных предварительных исследований, проведенных на периферической крови больных РС с длительными сроками заболевания, показали отсутствие достоверных различий по относительному содержанию ЕМ Тцит при сравнении с условно здоровыми добровольцами [36]. Однако, другой группой авторов отмечено, что снижение общего числа цитотоксических Т-клеток в периферической крови было связано с резким падением числа циркулирующих CD45RA⁻CD62L⁻ и CD45RA⁺CD62L⁻ среди CD3⁺CD8⁺ лимфоцитов [40]. Последние наблюдения подтверждаются результатами, полученными в ходе проведенного исследования.

Таким образом, анализ литературных данных по динамике различных популяций Тх и Тцит в периферической крови больных РС указывает, что изменения именно в клетках зрелых эффекторных популяций могут рассматриваться в качестве потенциально значимых маркеров для диагностики РС. Разработка и применение методов машинного анализа, учитывающих весь комплекс изменений популяционного состава лимфоцитов и подробную клиническую характеристику пациента, позволит в перспективе разработать эффективные подходы для раннего выявления РС.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Злокачественные новообразования в России в 2016 году (заболеваемость и смертность), под редакцией А. Д. Каприна, В. В. Старинского, Г. В. Петровой, 2018. Malignant neoplasms in Russia in 2016 (morbidity and mortality). Kaprin A. D., Starinsky V. V., Petrov G. V (eds), 2018.

2. ЮНЭЙДС. Информационный бюллетень, 2017. URL http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_ru.pdf
3. Poplin R., Varadarajan A. V., Blumer K., Liu Y., McConnell M. V., Corrado G. S., Peng L., Webster D. R. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering* 2018, 2, 158–164.
4. Rajpurkar P., Irvin J., Zhu K., Yang B., Mehta H., Duan T., Ding D., Bagul A., Langlotz C., Shpanskaya K., Lungren M. P., Ng A. Y. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *Arxiv preprints*, 2017.
5. Krusekopf S., Lücke J., Thilo Figge M. Automated detection of circulating tumor cells with naive Bayesian classifiers. *Cytometry A* 2014, 85, 501–511.
6. Шмидт Т. Е., Яхно Н. Н. Рассеянный склероз: руководство для врачей. МЕДпресс-информ, М., 2016, 272. [Schmidt T. E., Jahno N. N. Multiple sclerosis: a guide for doctors. MEDpress-inform, Moscow, 2016, 272.
7. Atlas of MS2013. URL: <http://www.msif.org/wp-content/uploads/2014/09/Atlas-of-MS.pdf>
8. Raman K. M. T. Simulation of Spread and Control of Lesions in Brain. *Computational and Mathematical Methods in Medicine* 2012, Article ID383546, 6 pages.
9. Zhao Y., Healy B. C., Rotstein D., Guttmann C. R., Bakshi R., Weiner H. L., Brodley C. E., Chitnis T. Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS One* 2017, 12(4), e0174866.
10. Ion-Mărgineanu A., Kocevar G., Stamile C., Sima D. M., Durand-Dubief F., Van Huffel S., Sappey-Mariniere D. Machine Learning Approach for Classifying Multiple Sclerosis Courses by Combining Clinical Data with Lesion Loads and Magnetic Resonance Metabolic Features. *Front Neurosci* 2017, 11, 398.
11. Kiiski H., Jollans L., Donnchadha S. Ó., Nolan H., Lonnegan R., Kelly S., O'Brien M. C., Kinsella K., Bramham J., Burke T., Hutchinson M., Tubridy N., Reilly R. B., Whelan R. Machine Learning EEG to Predict Cognitive Functioning and Processing Speed Over a 2-Year Period in Multiple Sclerosis Patients and Controls. *Brain Topogr* 2018, 31, 346–363.
12. Wotschel V., Alexander D., Kwok P. P., Chard D. T., Stromillo M. L., De Stefano N., Thompson A. J., Miller D. H., Ciccarelli O. Predicting outcome in clinically isolated syndrome using machine learning. *NeuroImage Clinical* 2015, 7, 281–287.
13. Bendfeldt K., Klöppel S., Nichols T. E., Smieskova R., Kuster P., Traud S., Mueller-Lenke N., Naegelin Y., Kappos L., Radue E. W., Borgwardt S. J. Multivariate pattern classification of gray matter pathology in multiple sclerosis. *NeuroImage*, 2012, 60, 400–408.
14. Eshaghi A., Wotschel V., Cortese R., Calabrese M., Sahraian M. A., Thompson A. J., Alexander D. C., Ciccarelli O. Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology* 2016, 87, 2463–2470.
15. Eshaghi A., Wotschel V., Cortese R., Calabrese M., Sahraian M. A., Thompson A. J., Alexander D. C., Ciccarelli O. Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses. *Front Neurosci* 2016, 10, 478.
16. Polman C. H., Reingold S. C., Banwell B., Clanet M., Cohen J. A., Filippi M., Fujihara K., Havrdova E., Hutchinson M., Kappos L., Lublin F. D., Montalban X., O'Connor P., Sandberg-Wollheim M., Thompson A. J., Waubant E., Weinshenker B., Wolinsky J. S. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 2011, 69, 292–302.
17. Kurtzke J. F. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983, 33, 1444–1452, Kappos L. Slightly modified, version 09/08.
18. Кудрявцев И. В., Борисов А. Г., Кроби́нец И. И., Савченко А. А., Серебрякова М. К. Определение основных субпопуляций цитотоксических Т-лимфоцитов методом многоцветной проточной цитометрии. *Медицинская иммунология* 2015, 17(6), 525–538. [Kudryavtsev I. V., Borisov A. G., Krobins I. I., Savchenko A. A., Serebryakova M. K. Multicolor flow cytometric analysis of cytotoxic T cell subsets. *Medical Immunology (Russia)* 2015, 17(6), 525–538.]
19. Machine Learning, Neural and Statistical Classification. Michie D., Spiegelhalter D. J., Taylor C. C. (eds). Overseas Press, 1994, 290.
20. Kohav R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. Kaufmann M. (ed), San Mateo, CA 1995, 1137–1143.
21. Bergstra J., Bengi Y. Random search for hyper-parameter optimization. *The Journal of Machine Learning Research* 2012, 13, 281–305.
22. Focosi D., Bestagno M., Burrone O., Petrini M. CD57⁺ T lymphocytes and functional immune deficiency. *J Leukoc Biol* 2010, 87(1), 107–116.
23. Chattopadhyay P. K., Betts M. R., Price D. A., Gostick E., Horton H., Roederer M., De Rosa S. C. The cytolytic enzymes granzyme A, granzyme B, and perforin: expression patterns, cell distribution, and their relationship to cell maturity and bright CD57 expression. *J Leukoc Biol* 2009, 85(1), 88–97.
24. Le Priol Y., Puthier D., Lecureuil C., Combadiere C., Debre P., Nguyen C., Combadiere B. High cytotoxic and specific migratory potencies of senescent CD8⁺ CD57⁺ cells in HIV-infected and uninfected individuals. *J Immunol* 2006, 177(8), 5145–5154.
25. Okada R., Kondo T., Matsuki F., Takata H., Takiguchi M. Phenotypic classification of human CD4⁺ T cell subsets and their differentiation. *Int Immunol* 2008, 20(9), 1189–1199.
26. Mahnke Y. D., Brodie T. M., Sallusto F., Roederer M., Lugli E. The who's who of T-cell differentiation: human memory T-cell subsets. *Eur J Immunol* 2013, 43(11), 2797–2809.
27. Acosta-Rodriguez E. V., Rivino L., Geginat J., Jarrossay D., Gattorno M., Lanzavecchia A., Sallusto F., Napolitani G. Surface phenotype and antigenic specificity of human interleukin 17-producing T helper memory cells. *Nat Immunol*, 2007, 8(6), 639–646.

28. Rvaris M., Barnes D., Woodrofe N., du Boulay G. H., Thorpe J. W., Thompson A. J., McDonald W. I., Miller D. H. Patterns of disease activity in multiple sclerosis patients: a study with quantitative gadolinium-enhanced brain MRI and cytokine measurement in different clinical subgroups. *J Neurol* 1996, 243(7), 536–542.
29. Kebir H., Ifergan I., Alvarez J. I., Bernard M., Poirier J., Arbour N., Duquette P., Prat A. Preferential recruitment of interferon-gamma-expressing TH17 cells in multiple sclerosis. *Ann Neurol* 2009, 66(3), 390–402.
30. Shajarian M., Alsahebfsoul F., Etemadifar M., Sedaghat N., Shahbazi M., Firouzabadi F. P., Dezhshibi H. M. IL-23 plasma level measurement in relapsing remitting multiple sclerosis (RRMS) patients compared to healthy subjects. *Immunol Invest* 2015, 44(1), 36–44.
31. Babaloo Z., Aliparasti M. R., Babaiea F., Almasi S., Baradaran B., Farhoudi M. The role of Th17 cells in patients with relapsing-remitting multiple sclerosis: interleukin-17A and interleukin-17F serum levels. *Immunol Lett* 2015, 164(2), 76–80.
32. Кудрявцев И. В., Ильвес А. Г., Борисов А. Г., Минеев К. К., Петров А. М., Савченко А. А., Серебрякова М. К., Столяров И. Д. CCR6-позитивные Т-хелперы периферической крови при рассеянном склерозе. Цитокины и воспаление 2016, 15(2), 166–172. [Kudryavtsev I. V., Ilves A. G., Borisov A. G., Mineev K. K., Petrov A. M., Savchenko A. A., Serebriakova M. K., Stoliarov I. D. CCR6-positive T helper subsets of peripheral blood in multiple sclerosis. *Cytokines and inflammation* 2016, 15(2), 166–172.]
33. Romero P., Zippelius A., Kurth I., Pittet M. J., Touvrey C., Iancu E. M., Corthesy P., Devere E., Speiser D. E., Rufer N. Four functionally distinct populations of human effector-memory CD8⁺ T lymphocytes. *J Immunol* 2007, 178(7), 4112–4119.
34. Rufer N., Zippelius A., Batard P., Pittet M. J., Kurth I., Corthesy P., Cerottini J. C., Leyvraz S., Roosnek E., Nabholz M., Romero P. Ex vivo characterization of human CD8⁺ T subsets with distinct replicative history and partial effector functions. *Blood* 2003, 102(5), 1779–1787.
35. Romero P., Zippelius A., Kurth I., Pittet M. J., Touvrey C., Iancu E. M., Corthesy P., Devere E., Speiser D. E., Rufer N. Four functionally distinct populations of human effector-memory CD8⁺ T lymphocytes. *J Immunol* 2007, 178(7), 4112–4119.
36. Кудрявцев И. В., Ильвес А. Г., Кробицец И. И., Минеев К. К., Серебрякова М. К., Петров А. М., Столяров И. Д. Субпопуляционный состав Т-хелперов и цитотоксических Т-лимфоцитов периферической крови при рассеянном склерозе. Цитокины и воспаление 2016, 15, 1. 91–99. [Kudryavtsev I. V., Krobince I. I., Mineev K. K., Serebriakova M. K., Petrov A. M., Stoliarov I. D. Helper and cytotoxic T lymphocyte subsets in patients with multiple sclerosis. *Cytokines and inflammation* 2016, 15(1), 91–99.]
37. Mikulkova Z., Praksova P., Stourac P., Bednarik J., Michalek J. Imbalance in T-cell and cytokine profiles in patients with relapsing-remitting multiple sclerosis. *J Neurol Sci* 2011, 300(1–2), 135–141.
38. Haegele K. F., Stueckle C. A., Malin J. P., Sindern E. Increase of CD8⁺ T-effector memory cells in peripheral blood of patients with relapsing-remitting multiple sclerosis compared to healthy controls. *J Neuroimmunol* 2007, 183(1–2), 168–174.
39. Liu G. Z., Fang L. B., Hjelmström P., Gao X. G. Increased CD8⁺ central memory T cells in patients with multiple sclerosis. *Mult Scler* 2007, 13(2), 149–155.
40. Pender M. P., Csurhes P. A., Pfluger C. M., Burrows S. R. Deficiency of CD8⁺ effector memory T cells is an early and persistent feature of multiple sclerosis. *Mult Scler* 2014, 20(14), 1825–1832.

STATISTICALLY SIGNIFICANT T-CELL POPULATIONS DURING DIAGNOSIS OF SCATTERED SCLEROSIS

© 2019 Y. V. Kozichuk¹, A. G. Ilves², I. V. Kudryavtsev^{3*}, D. A. Moskalenko¹,
O. M. Novoselova², K. S. Rubanik², M. K. Serebryakova³, D. O. Starov¹,
B. A. Timchenko¹, L. N. Prakhova², I. S. Lobanov¹

*E-mail: igorek1981@yandex.ru

¹IFMO University, St. Petersburg, Russia;

²N.P. Bechtereva Institute of the Human Brain of the Russian Academy
of Sciences (IHB RAS), St. Petersburg, Russia;

³Institute of Experimental Medicine (FSBSI "IEM"), St. Petersburg, Russia

Received: 12.11.2018. Accepted: 21.01.2019

To search for statistically significant T-cell populations in the diagnosis of multiple sclerosis (MS). The analysis of the absolute content of various subpopulations of T-lymphocytes (T-helpers (Th) and cytotoxic T-cells (Tcyt)) in the peripheral blood of 61 healthy volunteers and 47 patients with MS was carried out using multi-color flow cytometry. Based on the expression of differentiation markers (CD45RA, CD62L, CD27 and CD28) and effector molecules (CD56 and CD57), Th and Tcyt were divided into main populations at different stages of maturation. The following statistically significant populations of T-cells were identified: CD56-CD57⁺ T-lymphocytes, Em Th, EM3 Tcyt, CD56⁺CD57⁻ T-lymphocytes, EM2 Tcyt. The significance of these populations was also confirmed in the calculation of Chi-square statistics. Based on the information received, three groups of T-cell populations were selected. A model for the diagnosis of multiple sclerosis based on the algorithm of K nearest neighbors was built on each group of populations. The accuracy of prediction of the constructed models varies in the range of 0.69–0.90.

Key words: multiple sclerosis, machine learning, kNN, T-cells, flow cytometry

Authors:

Kozichuk Y. V., graduated student, Faculty of Control Systems and Robotics, IFMO University; St. Petersburg, Russia;

Ilves A. G., PhD, senior researcher, laboratory of neuroimmunology, N. P. Bechtereva Institute of the Human Brain of the Russian Academy of Sciences St. Petersburg, Russia;

Kudryavtsev I. V., ☒ PhD, senior researcher, department of immunology, "Institute of Experimental Medicine", St. Petersburg, Russia.

197376, St. Petersburg, acad. Pavlov str., 12, Scientific Research Institute of Experimental Medicine; Phoe/fax: +7–8122342929,
E-mail: igorek1981@yandex.ru;

Moskalenko D. A., graduated student, Faculty of Technological Management and Innovations, IFMO University; St. Petersburg, Russia;

Novoselova O. M., junior researcher, laboratory of neurorehabilitation, N. P. Bechtereva Institute of the Human Brain of the Russian Academy of Sciences, St. Petersburg, Russia;

Rubanik K. S., junior researcher, laboratory of neurorehabilitation, N. P. Bechtereva Institute of the Human Brain of the Russian Academy of Sciences, St. Petersburg, Russia;

Serebriakova M. K., Research Associate, department of immunology, Federal State Budgetary Scientific Institution "Institute of Experimental Medicine", St. Petersburg, Russia;

Starov D. O., graduated student, Faculty of Control Systems and Robotics, IFMO University; St. Petersburg, Russia;

Timchenko B. A., graduated student, Faculty of Control Systems and Robotics, IFMO University; St. Petersburg, Russia;

Prakhova L. N., MD, Head of the laboratory of neurorehabilitation, N. P. Bechtereva Institute of the Human Brain of the Russian Academy of Sciences, St. Petersburg, Russia;

Lobanov I. S., PhD (Physics and Mathematics), docent, Department of Control Systems and Robotics, IFMO University; St. Petersburg, Russia.