

© Коллектив авторов, 2024

УДК 616-006.4

Сабурина И.Н.², Кузьмина Е.Г.¹, Черкашенко В.Н.², Павлов В.В.¹, Зацаренко С.В.¹, Антонов К.Н.³, Богданенко Е.В.², Устинова Е.Е.², Морозов С.Г.²

Сравнительный анализ структурных особенностей многомерных данных периферической крови при хроническом лимфоцитарном лейкозе и неходжкинской В-лимфоме

¹Медицинский радиологический научный центр им. А.Ф. Цыба – филиал ФГБУ «Национальный медицинский исследовательский центр радиологии» Минздрава России, 249031, Обнинск, Россия, ул. Королева, д. 4;

²ФГБНУ «Научно-исследовательский институт общей патологии и патофизиологии», 125315, Москва, Россия, ул. Балтийская, д. 8;

³ФГБОУ ВО «Московский государственный университет им. М.В. Ломоносова»

119991, Москва, Россия, Ленинские горы, д. 1

Актуальность. Правильная идентификация лимфопролиферативных заболеваний хронического лимфолейкоза (ХЛЛ) и неходжкинских В-клеточных лимфом (НХЛ) является ключевым этапом скрининга, который требует высокой точности и интерпретируемости. Поэтому разработка современных методов дифференциальной диагностики НХЛ и ХЛЛ по-прежнему актуальна. Неинвазивные методы дифференциальной диагностики на основании исследований с последующим математическим анализом при помощи искусственных нейронных сетей (ИНС) позволят дифференцировать эти заболевания с высокой точностью только по данным крови, не используя материал других органов и тканей. Цель исследования – провести многомерный анализ состава субпопуляций лимфоцитов и опухолевых клеток в периферической крови и оценить взаимосвязи в системе «иммунитет – опухолевый рост», внедрив методы искусственных нейронных сетей (ИНС). Это позволяет выявлять структурные особенности, отличающие НХЛ от ХЛЛ.

Методика. Анализ данных осуществлялся при помощи искусственной нейронной сети-самоорганизующихся карт Кохонена (SOM (self organized map)). Эта ИНС позволяет выявлять структуру в многомерных данных, осуществляя проецирование многомерных образов в пространство пониженной размерности (2-х или 3-х мерное), что не требует принятия каких-либо априорных гипотез о структуре данных.

Результаты. Для построения и поиска различий многомерных образов болезни ХЛЛ от НХЛ были использованы искусственные нейронные сети (ИНС). Состояние иммунитета и опухолевых клеток сопоставлено по данным периферической крови пациентов с В-клеточной НХЛ (352 пациента) с данными ХЛЛ (315 пациентов). Выявлены структуры, отражающие различия состояний «НХЛ – ХЛЛ». Характер распределений и значения показателей иммунитета и опухолевого роста в многомерном пространстве позволяет с высокой точностью различать состояния ХЛЛ – НХЛ, определять стадию развития опухоли и в дальнейшем выбрать правильную тактику лечения.

Заключение. Отличия НХЛ и ХЛЛ, по образам многомерного анализа, создают основу для неинвазивного метода автоматической диагностики патологий НХЛ и ХЛЛ.

Ключевые слова: периферическая кровь; многомерный анализ; иммунитет; опухолевые В-клетки; НХЛ; ХЛЛ

Для цитирования: Сабурина И.Н., Кузьмина Е.Г., Черкашенко В.Н., Павлов В.В., Зацаренко С.В., Антонов К.Н., Богданенко Е.В., Устинова Е.Е., Морозов С.Г. Сравнительный анализ структурных особенностей многомерных данных периферической крови при хроническом лимфоцитарном лейкозе и неходжкинской В-лимфоме. *Патологическая физиология и экспериментальная терапия.* 2024; 68(1): 69-80.

DOI: 10.25557/0031-2991.2024.01.69-80

Участие авторов: концепция, анализ данных проточной цитофлуорометрии, создание базы клинических и иммунологических данных, статистическая обработка, клинико-иммунологическая интерпретация данных многомерного математического анализа, подготовка рукописи, редактирование текста статьи, общее руководство исследованием – Кузьмина Е.Г.; концепция, применение методов многомерного анализа, подготовка рукописи и иллюстративного материала, редактирование текста статьи – Черкашенко В.Н.; концепция, подготовка и редактирование рукописи, общее руководство исследованием – Сабурина И.Н.; иммунофенотипирование, подготовка иллюстраций, редактирование рукописи – Зацаренко С.В.; клиническая диагностика заболеваний – Павлов В.В.; работа с литературой – Богданенко Е.В.; применение методов многомерного анализа, подготовка иллюстративного материала – Антонов К.Н.; применение методов многомерного анализа, подготовка иллюстративного материала – Устинова Е.Е.; одобрение концепции и рукописи, общее руководство – Морозов С.Г.. Все соавторы – утверждение окончательного варианта статьи, ответственность за целостность всех частей статьи.

Для корреспонденции: Сабурина Ирина Николаевна, e-mail: saburina@mail.ru

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила 18.01.2024

Принята к печати 25.01.2024

Опубликована 28.03.2024

Saburina I.N.², Kuz'mina E.G.¹, Cherkashenko V.N.², Pavlov V.V.¹, Zacarenko S.V.¹, Antonov K.N.³, Bogdanenko E.V.², Ustinova E.E.², Morozov S.G.²

Comparative analysis of structural features of multidimensional data of peripheral blood in chronic lymphocytic leukemia and non-hodgkin B-lymphoma

¹Tsyb Medical Radiological Research Center, Branch of the National Medical Research Radiological Center the Ministry of Health of the Russian Federation, 4 Koroleva Str., Obninsk, 249031, Russian Federation;

²Institute of General Pathology and Pathophysiology, 8 Baltiyskaya St., Moscow, 125315, Russian Federation;

³Federal State Budget Educational Institution of Higher Education M.V. Lomonosov Moscow State University, 1 Leninskie Gory, Moscow, 119991, Russian Federation

Relevance. Correct identification of the lymphoproliferative diseases chronic lymphocytic leukemia (CLL) and non-Hodgkin B-cell lymphomas (NHL) is a key screening step that requires high accuracy and interpretability. Therefore, the development of modern methods for the differential diagnosis of NHL and CLL is still relevant. Non-invasive methods of differential diagnosis based on research followed by mathematical analysis using artificial neural networks (ANN) will make it possible to differentiate these diseases with high accuracy only based on blood data, without using material from other organs and tissues. **The purpose of the study** is to conduct a multivariate analysis of the composition of subpopulations of lymphocytes and tumor cells in peripheral blood and evaluate the relationships in the «immunity – tumor growth» system by introducing artificial neural network (ANN) methods. This allows us to identify structural features that distinguish NHL from CLL.

Methods. Data analysis was carried out using an artificial neural network – self-organizing Kohonen maps (SOM (self organized map)). This ANN allows you to identify structure in multidimensional data by projecting multidimensional images into a reduced-dimensional space (2- or 3-dimensional), which does not require the adoption of any a priori hypotheses about the structure of the data.

Results. Artificial neural networks (ANN) were used to construct and search for differences between multidimensional images of CLL disease and NHL. The state of immunity and tumor cells was compared according to peripheral blood data of patients with B-cell NHL (352 patients) with data from CLL (315 patients). Structures reflecting the differences between the states «NHL – CLL» have been identified. The nature of the distributions and values of immunity and tumor growth indicators in multidimensional space allows us to distinguish with high accuracy the conditions of CLL – NHL, determine the stage of tumor development and subsequently select the correct treatment tactics.

Conclusion. The differences between NHL and CLL, according to multivariate analysis, create the basis for a non-invasive method for automatic diagnosis of NHL and CLL pathologies.

Keywords: peripheral blood; multivariate analysis; immunity; tumor B cells; NHL; CLL

For citation: Saburina I.N., Kuz'mina E.G., Cherkashenko V.N., Pavlov V.V., Zacarenko S.V., Antonov K.N., Bogdanenko E.V., Ustinova E.E., Morozov S.G. Comparative analysis of structural features of multidimensional data of peripheral blood in chronic lymphocytic leukemia and non-hodgkin B-lymphoma. *Patologicheskaya Fiziologiya i Eksperimental'naya terapiya. (Pathological Physiology and Experimental Therapy, Russian Journal)*. 2024; 68(1): 69–80. (in Russian)

DOI: 10.25557/0031-2991.2024.01.69-80

Author's contribution: concept, analysis of flow cytometry data, creation of a clinical and immunological data base, statistical processing, clinical and immunological interpretation of multivariate mathematical analysis data, preparation of the manuscript, editing the text of the article, general supervision of the study – Kuzmina E.G.; concept, application of multivariate analysis methods, preparation of the manuscript and illustrative material, editing the text of the article – Cherkashenko V.N.; concept, preparation and editing of the manuscript, general supervision of the study – Saburina I.N.; immunophenotyping, preparation of illustrations, manuscript editing – Zatsarenko S.V.; clinical diagnosis of diseases – Pavlov V.V.; work with literature – Bogdanenko E.V.; application of multidimensional analysis methods, preparation of illustrative material – Antonov K.N.; application of multidimensional analysis methods, preparation of illustrative material – Ustinova E.E.; approval of the concept and manuscript, general guidance – Morozov S.G. All co-authors – approval of the final version of the article, responsibility for the integrity of all parts of the article.

For correspondence: *Irina N. Saburina*, chief researcher, Institute of General Pathology and Pathophysiology, Laboratory of Cell Biology and Developmental Pathology, e-mail: saburina@mail.ru

Information about the authors:

Saburina I.N., <https://orcid.org/0000-0003-2014-2535>

Kuzmina E.G., <https://orcid.org/0000-0002-0728-593X>

Cherkashenko V.N., <https://orcid.org/0009-0009-0071-471X>

Zatsarenko S.V., <https://orcid.org/0000-0002-6335-1401>

Antonov K.N., <https://orcid.org/0009-0009-0091-3591>

Ustinova E.E., <https://orcid.org/0009-0008-8976-0212>

Morozov S.G., <https://orcid.org/0000-0001-5822-5729>

Conflict of interests. The authors declare no conflict of interest.

Financing. The study had no sponsorship

Received 18.01.2024

Accepted 25.01.2024

Published 28.03.2024

Введение

Актуальность. Ранее нами было показана возможность методами многомерного анализа выявлять образы патологии В-клеточной НХЛ и ХЛЛ и отличать их от статуса здоровья по результатам структуризации состояния иммунитета и циркулирующих в крови опухолевых В-клеток. При сравнении многомерных данных пациентов с НХЛ и ХЛЛ с данными здоровых людей выявлены различия в их структуре, позволяющие дифференцировать состояние здоровых людей от пациентов с первой и второй патологией. Наличие общности и различий в структуре многомерных данных больных и здоровых людей позволило высказать гипотезу о возможности построения автоматической классификационной процедуры проведения диагностики для пациентов ХЛЛ и НХЛ. Однако очевидно, что после предварительного предположения о наличии у пациента зрелоклеточной лимфопрлиферации (ХЛЛ/НХЛ), следующим шагом должно следовать проведение дифференциальной диагностики этих патологий (НХЛ или ХЛЛ). Проверка возможности проведения дифференциальной диагностики ХЛЛ и НХЛ по результатам многомерного анализа состояния иммунитета и опухолевых клеток в периферической крови составила основу данного исследования.

Цель данного исследования – проведение многомерного анализа структур, выявляемых по клеточному составу иммунной системы и опухолевых В-клеток в периферической крови пациентов с НХЛ и ХЛЛ, для разработки подходов дифференциальной диагностики данных патологий.

Для достижения цели были сформулированы и решены следующие задачи:

Используя методы искусственных нейронных сетей, ИНС, самоорганизующимися картами Кохонена (СОКК) визуализирован многомерный массив данных и сопоставлено состояние иммунитета и опухолевых В-клеток в периферической крови пациентов с НХЛ и ХЛЛ.

При помощи результатов визуализации данных проведена кластеризация и описаны выявленные в многомерных данных структуры для пациентов в двух группах патологии – НХЛ и ХЛЛ.

Проведен анализ признаков структуры НХЛ и ХЛЛ, и определены показатели, формирующие ее различия, которые могут быть использованы для построения алгоритма автоматизированной дифференциальной диагностики этих заболеваний.

Методика

Объекты исследования. В анализ включены данные клинико-иммунологического обследования 352 паци-

ентов с В-клеточными НХЛ и 315 пациентов с ХЛЛ. Для дифференциальной диагностики ХЛЛ и НХЛ использованы критерии классификации ВОЗ опухолей лимфоидной и кроветворной системы: объективный анализ ультразвукового, магнитно-резонансного обследования пациентов, иммуногистохимического исследования материала лимфатических узлов, тканей и костного мозга, иммунофенотипирования опухолевых клеток крови и костного мозга [1-4].

В группу НХЛ включены пациенты с фолликулярной лимфомой, диффузной крупноклеточной лимфомой, лимфомой из малых лимфоцитов, лимфомой из клеток зоны мантии, лимфомой из клеток маргинальной зоны и более редко встречающиеся нозологические формы. Средний возраст пациентов составил: НХЛ – 60 лет, ХЛЛ – 62 года, что соответствует данным литературы [5].

Обследование больных проведено в 1996–2019 гг.

Оценка состояния иммунитета и опухолевых клеток. По гемограмме подсчитывали количество лейкоцитов, относительное и абсолютное число лимфоцитов. Состав ядросодержащих клеток оценивали в цельной крови после предварительного лизиса эритроцитов лизирующим раствором фирмы BD Biosciences. Далее в лимфоцитарном гейте определяли субпопуляции иммунокомпетентных клеток и иммунофенотип В-лимфоцитов методами 2-6-цветной проточной цитометрии на приборе FACSCan или FACSCanto II (BD Biosciences) [1-3].

Для многомерного анализа состояния иммунитета и признаков В-клеточной пролиферации для каждого пациента были использованы 13 показателей, которые включали количество лейкоцитов, Лц абс, $\times 10^9$ кл/л, процент Лф, общих Т-клеток, $CD3^+CD16^-$; Т-хелперов, $CD3^+CD4^+$ %; Т-киллеров, $CD3^+CD8^+$; иммунорегуляторный индекс, $CD4/CD8$; процент НК-клетки, $CD16^+CD3^-$; активированных Т-клеток, $CD3^+HLADR^+$; активированных лимфоцитов, $HLADR^+$, включая нормальные и опухолевые В-лимфоциты и активированные Т-клетки. Идентификация В-лимфоцитов проведена по маркерам $CD19^+$ %, $CD20^+$ %, $CD22^+$ %, $CD23^+$ %.

Методы анализа экспериментальных данных. Визуализация многомерных данных, характеризующих состояние иммунитета и В-клеточной пролиферации, проведена с использованием самоорганизующихся карт СОКК (Self Organizing Maps, SOM), разработанных Т.Кохоненом [6-11]. Этот подход обеспечивает два важных аспекта анализа данных – проецирование многомерных данных в пространство пониженной размерности (плоскость) и визуализацию много-

мерных данных в виде плоской карты и, на ее основе осуществляет кластеризацию, выделяя в построенной двумерной карте кластеры векторов признаков пациентов и здоровых людей с близкими (похожими) характеристиками (рис. 1).

Сети Кохонена относятся к классу сетей, обучение которых производится «без учителя», то есть ре-

зультат обучения зависит только от структуры входных данных. Сеть Кохонена имеет два слоя: входной и выходной, составленный из нейронов упорядоченной структуры (рис. 1, а). Выходной слой называют также слоем топологической карты, или «экраном». Нейроны выходного слоя располагаются в узлах двумерной сетки с прямоугольными или шестиугольными

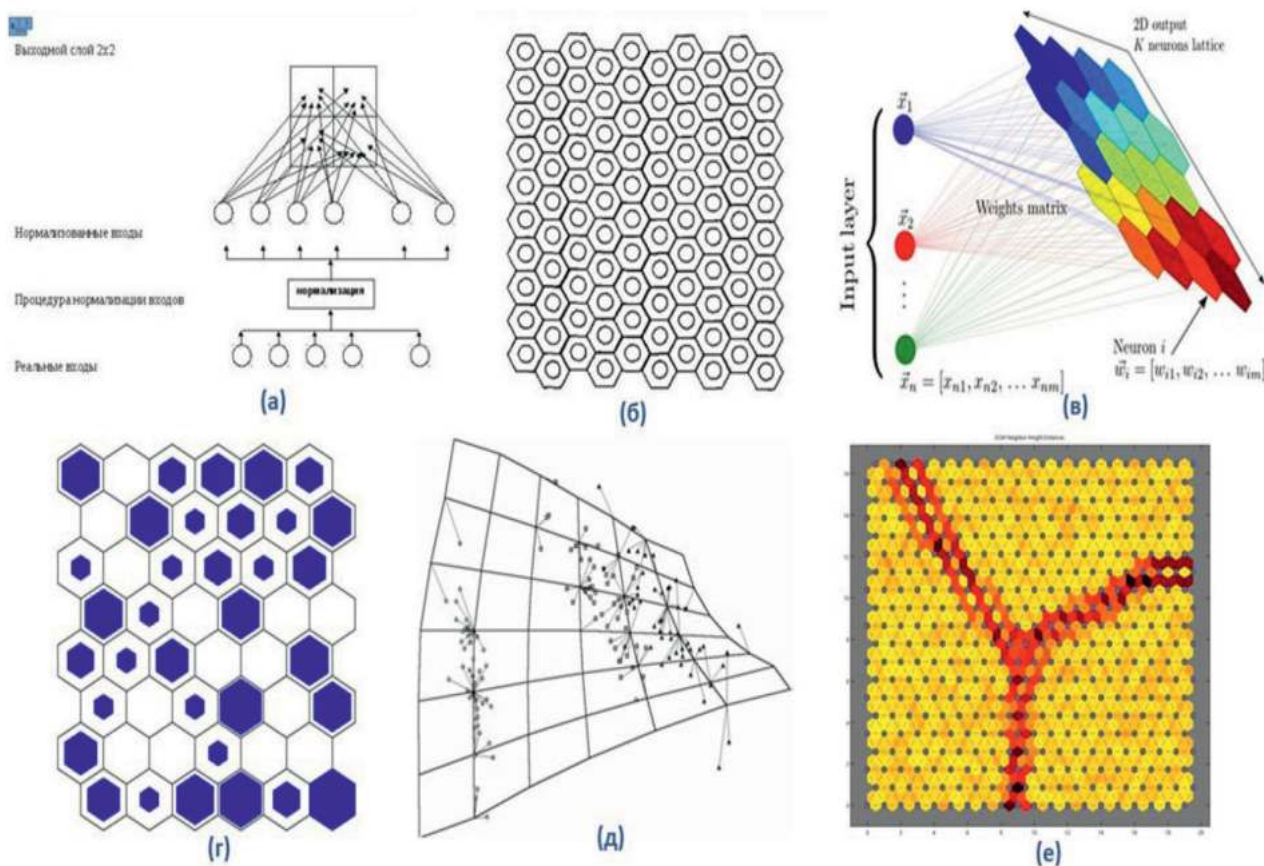


Рис. 1. Иллюстрация алгоритма работы самоорганизующихся сетей Кохонена (SOM).

(а) – архитектура самоорганизующихся сетей Кохонена, (б) – топология элементов (нейронов) СОКК (на рисунке – правильные шестиугольники), (в) – группировка векторов анализируемых данных по степени близости (похожести) – похожие векторы проецируются в одни и те же области карты, (г) – иллюстрация количества анализируемых векторов, попадающих в каждый нейрон сети СОКК пропорционально закрашенной площади каждого нейрона, (д) – алгоритм обучения устроен так, что в местах скопления анализируемых векторов признаков он деформирует сетку и она «прогибается» под «тяжестью» данных в таком нейроне, увлекая за собой и соседние узлы сети, что формирует на карте области с похожими векторами признаков – кластеры данных, (е) – сформированная на сетке структура из 3 кластеров; области, содержащие скопление векторов признаков закрашиваются одним цветом, а области, в которых нет спроецированных векторов признаков – другим, именно они и формируют границы между кластерами.

Fig. 1. Illustration of the algorithm for the operation of self-organizing Kohonen networks (SOM).

(a) – architecture of self-organizing Kohonen networks, (b) – topology of elements (neurons) of the SOCC (regular hexagons in the figure), (c) – grouping of vectors of analyzed data according to the degree of proximity (similarity) – similar vectors are projected into the same areas of the map, (d) illustration of the number of analyzed vectors that fall into each neuron of the SOCC network in proportion to the shaded area of each neuron, (e) – learning algorithm is designed in such a way that in places where the analyzed feature vectors accumulate, it deforms the grid and it «bends» under the «weight» of the data in such a neuron, dragging with it neighboring network nodes, which forms areas on the map with similar feature vectors – data clusters, (e) – structure of 3 clusters formed on the grid; areas containing a cluster of feature vectors are painted with one color, and areas in which there are no projected feature vectors are painted with another color; they form the boundaries between clusters.

ми ячейками (рис. 1, б). Количество нейронов в сетке определяет степень детализации результата работы алгоритма, и зависит от объема данных, в которых пытаются выявить структуру. Самоорганизующиеся карты, в ходе своего обучения, анализируют характер расположения точек входного слоя в m -мерном пространстве и воспроизводят на выходе нейронной сети топологический порядок и определенную степень регулярности исходных данных (т.е. метрическую близость векторов исходных данных). Похожие вектора признаков проецируются в нейроны, расположенные близко на карте. То есть, объекты, близкие в исходном пространстве остаются близкими и на плоскости карты (рис. 1, в). Алгоритм подсчитывает количество анализируемых векторов признаков, попавших в каждый нейрон, и отображает их количество, (рис. 1, г) величиной закрашиваемой площади нейрона. Подгонка SOM заключается в итеративной настройке вектора весовых коэффициентов каждого нейрона в выходном слое, для чего используется модифицированный алгоритм соревновательного обучения Хебба, действующий по принципу «победитель забирает все», который учитывает не только вклад нейрона-победителя, но и ближайших его соседей, расположенных в его окрестностях. При обучении стека нейронов деформируется, увеличивая плотность узлов (нейронов) в местах сгущения данных при обучении (рис. 1, д). Места скопления нейронов, в которых концентрируются вектора анализируемых данных, закрашиваются одним цветом – это выявленные кластеры. В этих местах объем данных «давит» на карту, деформирует и продавливает ее, формирует низменность (овраг, залив, море). Поэтому плоскость, на которую проецируются анализируемые данные (входные вектора), названа картой.

В силу этого, алгоритм формирует на карте области, в которых располагаются похожие друг на друга входные векторы данных, и, тем самым, выявляет и визуализирует кластеры (рис. 1, е).

Процедура анализа многомерных данных состоит из двух шагов:

Сначала сетевой алгоритм выявляет и визуализирует на плоскости структуру многомерных данных (строится так называемая U матрица, раскрашивающая «экран» цветами в соответствии с **расстояниями** между анализируемыми входными векторами-признаками). Синий цвет кодирует малые расстояния (большую похожесть признаков), а красный цвет – границы между выявленными кластерами. Таким образом, области, закрашенные алгоритмом обучения в синий цвет, выделяют и визуализируют кластеры (близкие по свойствам объекты), а области, закрашенные в цвета крас-

но-оранжевой гаммы, характеризуют границы между выявленными кластерами.

На следующем этапе алгоритм раскрашивает «экран» с выявленной кластерной структурой цветами, которые кодируют уже **значения признаков** (красный – высокие значения, а синий цвет – малые значения признаков).

Для численных экспериментов с этим типом сетей нами использована программная среда Матлаб и библиотека SOM ToolBox, разработанная в компании MathWorks [11].

Результаты

Выявление структуры и ее визуализация в многомерных данных НХЛ и ХЛЛ. На рисунке 2 представлена визуализация структуры, выявленной в многомерных данных пациентов с НХЛ и ХЛЛ. В соответствии с алгоритмом, многомерные данные проецируются на плоскость нейронов карты СОКК. В левой части рисунка (рис. 2, а) представлена U-матрица (матрица взаимных расстояний между векторами признаков пациентов в виде трехмерной поверхности). На ней цветом кодируется сходство входных векторов друг с другом.

Далее, с очевидностью, встает вопрос о том, как выделенная и описанная кластерная структура соотносится с групповой структурой данных, т.е. с принадлежностью к группе пациентов ХЛЛ или НХЛ. Для ответа на этот вопрос мы, как и в предыдущей работе, использовали возможность маркировки нейронов сети СОКК текстовыми метками на основании **групповой принадлежности**. Кроме того, метод СОКК позволяет оценивать количество анализируемых входных векторов (данных пациентов), попадающих в каждый из нейронов сети СОКК. На **рисунке 2, (б)** представлена картина распределения векторов признаков, принадлежащих группам пациентов с НХЛ и ХЛЛ. Выделенная кластерная структура в высокой степени соответствует групповой структуре анализируемых данных (поверхность карты разделена границей (кривой), которая повторяет расположение «хребта» на **рисунке 2, (а)**. При этом кластер 1 совпадает с данными группы пациентов НХЛ (верх карты), а кластер 2, в целом, совпадает с данными пациентов ХЛЛ (нижняя часть карты). Кроме того видно, что часть векторов, принадлежащих группе пациентов с ХЛЛ (кластер 2), попадает в кластер 1, что определяет большую вариативность данных группы ХЛЛ из 2 кластера (обозначены овалами на **рисунке 2, (б)**). Также часть векторов пациентов с НХЛ попадает во второй кластер, т.е. расположены ниже границы между кластерами. Они также выделены овалами. Важным моментом, который следует отметить, является

то, что все ошибочные классификации (овалы на рисунке 2, (б)) локализованы в районе границы между кластерами, что свидетельствует о достаточно хорошем представлении кластерной структуры данных.

Далее, для понимания того, какими признаками формируется выявленная кластерная структура, на ри-

сунке 3 показано распределение всех 13 анализируемых показателей по U матрице. Видно, что все анализируемые признаки распадаются на две группы – признаки, контрастирующие кластерную структуру и «шумящие» признаки. В группу сильно контрастирующих кластеры входят 10 признаков: проценты лимфоци-

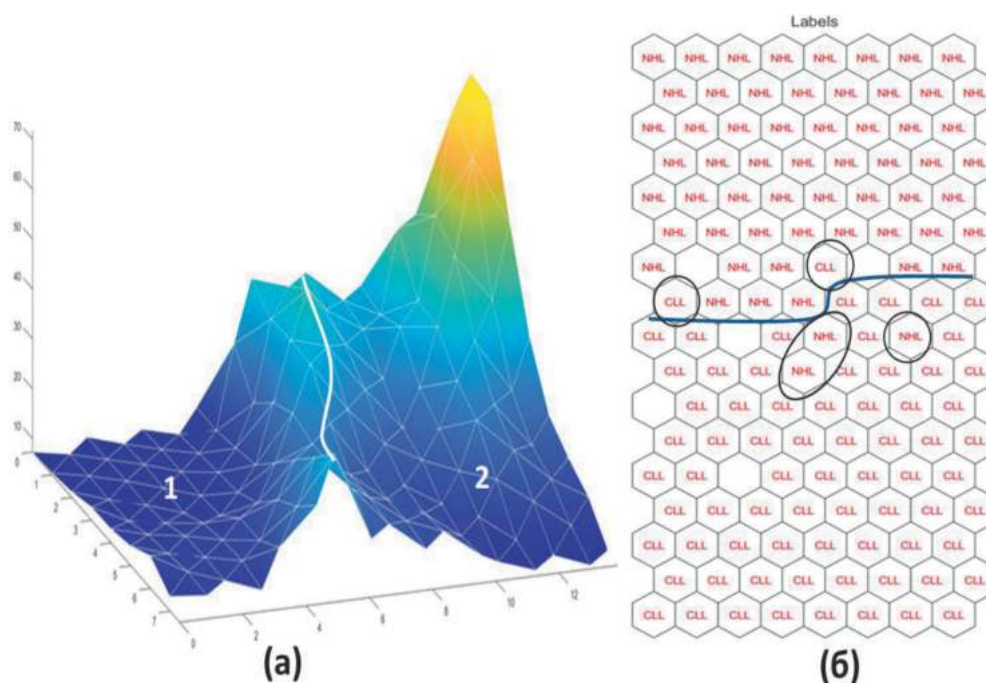


Рис. 2. Выявление структуры в многомерных данных показателей крови для NHL и ХЛЛ по 13 показателям: (а) – U матрица (матрица похожести анализируемых векторов признаков, представленная в виде поверхности) четко выделяет 2 кластера в анализируемых данных (разделены «хребтом», маркированным белой кривой); (б) – наложение на выявленную в многомерных данных структуру групповой принадлежности пациентов (ХЛЛ и NHL). Нейроны на карте СОКК помечены текстовыми метками (CLL – пациенты с ХЛЛ и пациенты с неходжкинской лимфомой – NHL). Овалами обозначены места на карте, в которых в кластеры одной группы попадают пациенты другой группы (ошибки классификации).

Синим цветом на карте маркируются области, в которых содержатся вектора, очень похожие друг на друга. А зелено-желто-красной гаммой представлены области карты, в которых располагаются вектора, менее похожие друг на друга. Как видно из рисунка 2 (а) в многомерных данных пациентов с NHL и ХЛЛ выявлено 2 кластера, которые разделены «хребтом», идущим через всю карту от ее дальней границы к ближней границе. Граница между кластерами обозначена на рисунке линией белого цвета, которая делит плоскость U-матрицы на левую и правую части, обозначенные цифрами 1 и 2 соответственно. Относительно небольшая высота «хребта» определяется тем, что в кластере 2 наблюдается большая неоднородность векторов признаков; в дальнем правом углу карты возвышается «гора» в которой располагаются вектора признаков с очень непохожими свойствами (это определяет большую вариативность признаков пациентов 2 кластера). В многомерных данных пациентов с NHL и ХЛЛ нам удалось четко выделить 2 кластера.

Fig. 2. Identifying structure in multivariate blood data for NHL and CLL according to 13 indicators: (a) – U matrix (similarity matrix of analyzed feature vectors, presented as a surface) clearly identifies 2 clusters in the analyzed data (separated by a «ridge» marked with a white curve); (b) – overlay on the structure of group membership of patients (CLL and NHL) identified in multidimensional data. Neurons on the SOCC map are marked with text labels (CLL patients with CLL and patients with non-Hodgkin lymphoma – NHL). Ovals indicate places on the map where patients from another group fall into clusters of one group (classification errors).

Areas that contain vectors that are very similar to each other are marked in blue on the map. And the green-yellow-red gamma represents areas of the map in which vectors are located that are less similar to each other. As can be seen from Figure 2 (a), in the multidimensional data of patients with NHL and CLL, 2 clusters were identified, which are separated by a «ridge» running across the entire map from its far border to the near border. The boundary between the clusters is indicated in the figure by a white line, which divides the plane of the U-matrix into left and right parts, designated 1 and 2, respectively. The relatively small height of the «ridge» is determined by the fact that in cluster 2 there is a large heterogeneity of feature vectors; in the far right corner of the map there is a «mountain» in which feature vectors with very dissimilar properties are located (this determines the large variability of features of patients in cluster 2). In the multivariate data of patients with NHL and CLL, we were able to clearly identify 2 clusters.

тов и CD3, CD4, CD8, CD16, HLA-DR, CD19, CD20, CD22 и CD23 клеток.

Как видно из **рисунка 3**, по 4-м признакам пациенты с НХЛ обладают большими значениями, чем пациенты с ХЛЛ. Этими признаками являются проценты CD3, CD4, CD8 и CD16 клеток. По остальным 6 контрастирующим признакам наблюдается противоположная картина; по ним пациенты с НХЛ обладают меньшими значениями, чем пациенты с ХЛЛ. К этой группе относятся: проценты лимфоцитов, CD19, CD20, CD22, CD23 и HLA-DR клеток. Все 10 перечисленных ранее признаков, контрастируют кластерную структуру, почти точно повторяя границы выделенных кластеров на **рисунках 2, (а) и 2, (б)**.

К группе «шумящих» признаков относятся 3 признака: абсолютное содержание лейкоцитов, соотношение CD4/CD8 и процент CD3-HLA-DR.

Чтобы наглядно представить, почему показатели демонстрируют контрастность или «шум», представлены гистограммы распределения контрастирующих выявленную кластерную структуру признаков (**рис. 4**).

Все они характеризуются одним общим свойством: имеют противоположную скошенность (третий момент распределения, называемый асимметрией). Например, на **рисунке 4, (а)** значение показателя «% лимфоцитов» имеет правую скошенность для НХЛ и левую скошенность для ХЛЛ. Из статистики известно, что лучшей оценкой средней тенденции для скошенных распределений является мода (величина максимального столбика в гистограмме). Поэтому, на всех четырех рисунках взаимная противоскошенность распределений приводит к тому, что их моды становятся сильно различимыми. Именно по этой причине эти показатели и становятся контрастирующими при многомерном анализе.

Далее рассмотрим гистограммы распределений для трех «шумящих» признаков (**рис. 5**). Видно, что для обеих групп пациентов, и ХЛЛ, и НХЛ, распределения являются право скошенными. Но обращает на себя внимание тот факт, что ушло свойство взаимной скошенности, а моды распределений практически совпадают по величине. Это и приводит к тому, что эти пере-

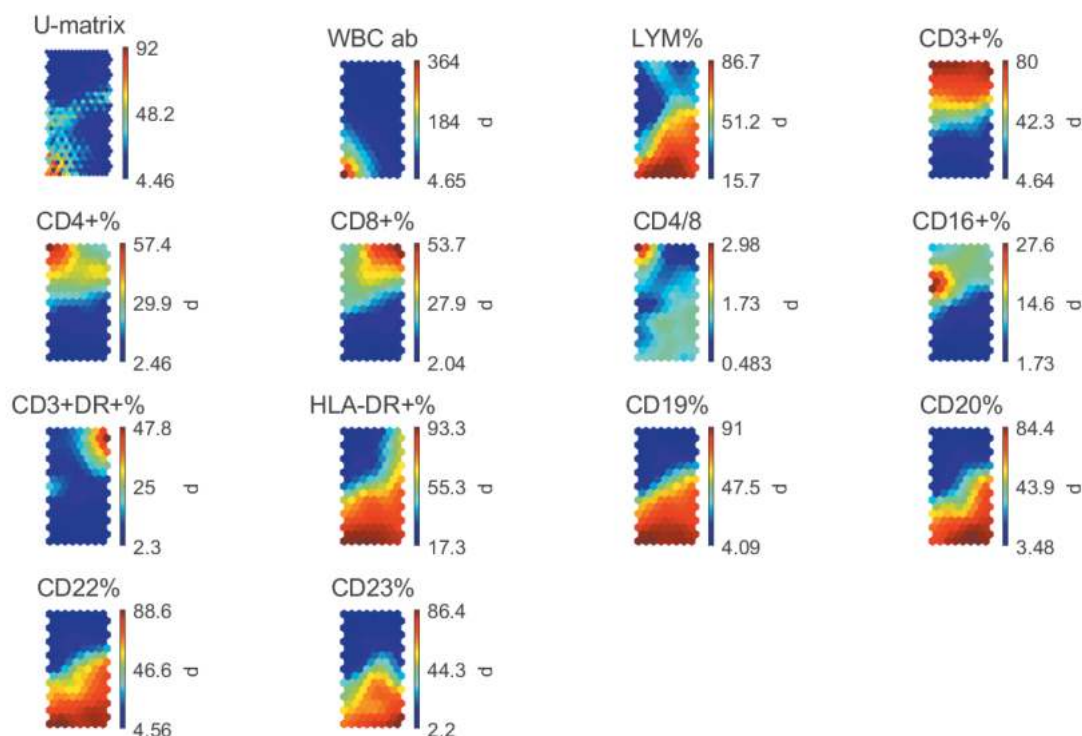


Рис. 3. Распределение значений 13 анализируемых признаков больных НХЛ и ХЛЛ по СОКК. В левом верхнем углу располагается U-матрица с выделенными 2 кластерами и границей между ними (кластер 1 – НХЛ, кластер 2 – ХЛЛ), а остальные рисунки – распределение разных значений 13 признаков по карте (название конкретного признака показано над конкретным элементом рисунка).

Fig. 3. Distribution of values of 13 analyzed signs of patients with NHL and CLL according to SOCC. In the upper left corner there is a U-matrix with 2 highlighted clusters and the boundary between them (cluster 1 – NHL, cluster 2 – CLL), and the rest of the figures are the distribution of values of 13 features on the map (the name of a specific feature is shown above a specific element of the picture).

менные бесполезны для выявления структуры и скорее затрудняют распознавание пациентов разных групп.

Обсуждение

Как известно, выявляемые в крови лимфоциты, находятся в ней недолго, в течение суток многократно, несколько раз покидают ее, мигрируют в лимфоидные органы и ткани и вновь возвращаются. Таким образом, динамичная интегрирующая система крови, как в зеркале, отражает, происходящие в организме изменения, в том числе состояние популяций лимфоцитов и других типов клеток [12]. Мы высказали гипотезу, что имеется возможность разработать алгоритм диагностики вариантов зрелоклеточной лимфопротиферации, НХЛ и ХЛЛ, по различию содержания в периферической крови иммунокомпетентных и опухолевых клеток. При НХЛ наблюдается меньший выход опухолевых клеток в кровь, чем при ХЛЛ [13]. Для нее характерен первичный локальный злокачественный опухолевый рост, исходящий преимущественно из внекостномозговой лимфоидной ткани. Первичное возникновение процесса – лимфатические

узлы (нодальные формы НХЛ) или ткани, богатые лимфоцитами (тканевые экстра nodальные формы НХЛ типа MALT-лимфом). В отличие от лейкозов с первичным поражением костного мозга и лейкемическими нарушениями периферической крови, лимфомы возникают в лимфатических узлах и проникают в окружающие ткани. При этом в костном мозге долгое время опухолевые клетки не обнаруживаются или выявляются в минимальном количестве. Значительно реже и в меньшем количестве наблюдаются лейкемические нарушения периферической крови. Степень их проявления при НХЛ во многом обусловлена особенностями транслокаций, произошедших в геноме В-клеток, зоны лимфоузла, относительно герминального центра (прегерминальные, центрофоликулярные и постгерминальные НХЛ), в которой произошла трансформация клеток [1, 14-16]. Опухолевые клетки из разных зон лимфоузла, характеризуются разным активационным статусом, обладают разной пролиферативной активностью и возможностями выхода в циркуляцию.

В противоположность НХЛ, одним из основных диагностических критериев ХЛЛ считается изначаль-

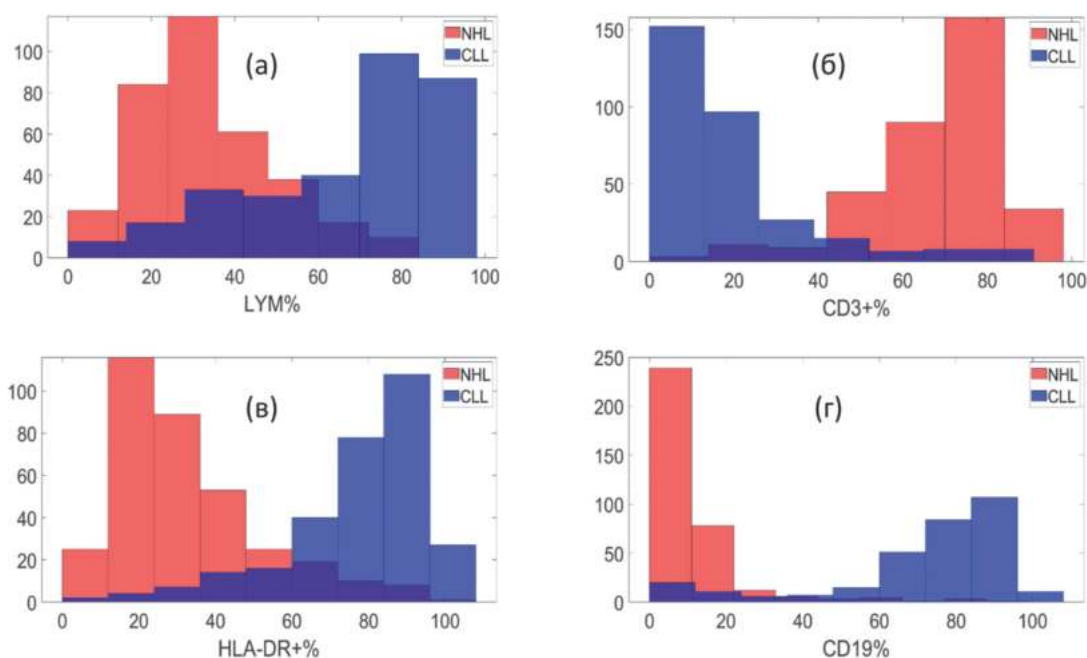


Рис. 4. Гистограммы распределения контрастирующих признаков для выявления структуры в данных пациентов с НХЛ (красный цвет) и ХЛЛ (синий цвет). Область перекрывания данных – фиолетовый цвет.

(а) – % лимфоцитов в крови; (б) – CD3%; (в) – HLA-DR%; (г) – CD19%.

Fig. 4. Histograms of the distribution of contrasting features to identify structure in data from patients with NHL (red) and CLL (blue). The area of data overlap is purple.

(а) – % of lymphocytes in the blood; (б) – CD3%; (в) – HLA-DR%; (г) – CD19%.

ное стойкое обнаружение в крови моноклональных опухолевых лимфоцитов в количестве, превышающем 5×10^9 /л крови. Патологический лимфогенез инициируется онкогенным событием, вызывающим нарушение нормального клеточного цикла. В нем могут действовать два механизма — активация онкогенов либо подавление опухолевых супрессоров (антионкогенов). Транслокации генетического материала в хромосомах при ХЛЛ многочисленны [1, 14-16], опухолевые клетки персистируют, сохраняя жизнеспособность из-за снижения апоптоза, а также за счет изначальной активации клеток и нарушения контроля пролиферации. Злокачественные В-клетки появляются в костном мозге (предполагается, что ХЛЛ может развиваться, помимо других причин, вследствие лейкемогенной трансформации в мультипотентных стволовых клетках) [16], пролиферируют в лимфатических узлах, из-за дефектов апоптоза и снижения супрессорного контроля накапливаются в периферической крови, костном мозге, лимфатических узлах и селезенке [17].

При ХЛЛ у пациентов часто выявляется дисрегуляция иммунного ответа, обусловленная нарушения-

ми противоопухолевого иммунитета. Выраженные дефекты функции Т-клеток описывают как нарушение баланса Т-клеточных субпопуляций, дефективности формирования иммунных синапсов с антигенпрезентирующими клетками, нарушениями киллерной эффекторной функции, повышением уровня регуляторных супрессорных Т-клеток.

Постепенная утрата иммунокомпетентности в поддержании генетического гомеостаза клеточного состава организма в патогенезе зрелоклеточной лимфопротиферации многократно описана [18-20]. Многокомпонентность клеточного состава, осуществляющего иммунный контроль, и многоликость фенотипов опухолевых В-клеток создают представления о зрелоклеточной лимфопротиферации как о многомерном процессе.

Особенности патофизиологии НХЛ и ХЛЛ по признаку различий лейкемизации опухолевого процесса и дисфункции иммунитета, на наш взгляд, можно использовать в качестве критериев, пригодных для дифференциации многомерных структурных образов этих заболеваний.

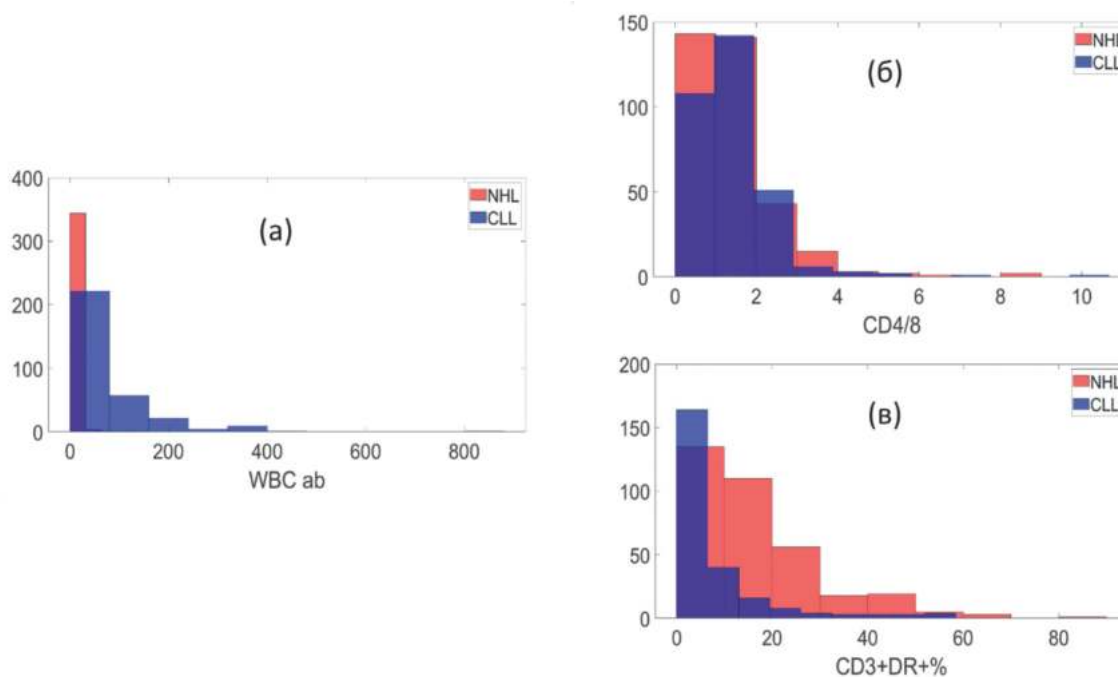


Рис. 5. Гистограммы распределения «шумящих» признаков при анализе обнаруженной кластерной структуры в многомерных данных пациентов с НХЛ (красный цвет) и ХЛЛ (синий цвет). Область перекрытия данных – фиолетовый цвет.

(a) – % лимфоцитов; (б) – иммунорегуляторный индекс, соотношение CD4/CD8; (в) – процент CD3-HLA-DR.

Fig. 5. Histograms of the distribution of «noisy» features when analyzing the detected cluster structure in multidimensional data from patients with NHL (red) and CLL (blue). The area of data overlap is purple.

(a) – % of lymphocytes; (b) – immunoregulatory index, CD4/CD8 ratio; (c) – percentage of CD3-HLA-DR.

На первом этапе применения методов ИНС были построены многомерные образы и выявлены различия при сопоставлении состояния здоровья с состоянием ХЛЛ и с состоянием НХЛ [1].

На втором этапе выявлена структуризация данных и определена возможность различения многомерных образов заболеваний НХЛ и ХЛЛ. Результаты анализа многообещающи. Многомерный анализ находит различия в выраженности нарушений иммунитета (CD3, CD4, CD8 и CD16) и уровня опухолевых клеток (HLA-DR, CD19, CD20, CD22, CD23 и % лимфоцитов), находящихся в крови, т.е. степень противостояния иммунитета росту опухоли при изучаемых патологиях различна: контрастность иммунитета и опухолевого роста при НХЛ и ХЛЛ очевидна. Она обусловлена 10 признаками из 13, контрастно описываемыми противодействием иммунитета и опухолевых клеток: при ХЛЛ процент иммунокомпетентных клеток значительно ниже, а опухолевых клеток выше, чем при НХЛ.

Поэтому процентное содержание лимфоцитов крови и уровень В-клеток (CD19, CD20, CD22, CD23, HLA-DR), выступают при кластеризации как признаки, контрастирующие происхождение, первичную локализацию и особенности опухолевого роста НХЛ и ХЛЛ. При ХЛЛ эти показатели существенно выше, чем при НХЛ. При ХЛЛ нарастают признаки опухолевой пролиферации за счет выхода в кровеносное русло активированных опухолевых со сниженным потенциалом гибели В-клеток, накапливающихся по мере развития процесса. Начальные же варианты ХЛЛ более похожи на моноклональный лимфоцитоз, проявляют низкую лейкомизацию, поэтому их многомерный образ может быть «заброшен» анализом на территорию кластера НХЛ.

В свою очередь, образ НХЛ из малых лимфоцитов в случае выраженного проявления лейкомизации, по своей многомерной структуре может быть схож с ХЛЛ из малых лимфоцитов; т.е. оказаться близким или аналогичным ХЛЛ и метод картирует его на территории, занимаемой кластером ХЛЛ. Это сходство находится в соответствии с классификацией лимфом зрелоклеточного происхождения, объединяющей ХЛЛ и НХЛ, опухолевым субстратом которых являются малые лимфоциты, в общей рубрике [4]. Мы полагаем, что за счет таких вариантов, а также других нозологий со сходными этапами патогенеза, обнаруживаются «забросы» между кластерами на «чужую» территорию, т.е. происходит перекрытие и «смазанность» границ между кластерами в многомерном пространстве.

Выход и накопление в крови опухолевых клеток отрицательно коррелирует с уровнем Т-лимфоцитов (CD3, общих Т-клеток, CD4+Т-хелперов, CD8+Т-кил-

леров) и CD16+НК-клеток, отражая снижение их относительного уровня в циркуляции. Важно, что проценты этих клеток по мере роста популяции опухолевых В-клеток в крови при ХЛЛ снижаются значительно, чем при НХЛ, что делает их величины пригодными для дифференциации состояний в качестве контрастирующих признаков. По ним (CD3, CD4, CD8 и CD16) пациенты с НХЛ обладают большими значениями, чем пациенты с ХЛЛ. Мы не будем в сравнительном аспекте обсуждать степень утраты иммунокомпетентности, т.к. это не входит в задачи данного исследования.

Таким образом, кластерная структура данных, выявленная методами ИНС при НХЛ и ХЛЛ, наглядно отражает особенности течения опухолевого процесса при этих видах зрелоклеточной лимфопротиферации и ее отражение в уровне показателей крови может быть использовано для построения автоматизированного алгоритма различения этих заболеваний.

Заключение и выводы

В настоящем исследовании изучена возможность и определены перспективы применения методов ИНС для различения вариантов зрелоклеточной В-клеточной лимфопротиферации (НХЛ и ХЛЛ) за счет оптимизации использования иммунограмм крови без необходимости использования клеток лимфатических узлов, костного мозга, других органов и тканей. В анализ включены данные клинико-иммунологического обследования 352 пациентов с В-клеточными НХЛ и 315 пациентов с ХЛЛ.

Полученные результаты позволяют сделать следующие выводы:

1. Подтверждена гипотеза о высокой эффективности применения технологии самоорганизующихся карт Кохонена (СОКК) для выявления структур в многомерных массивах экспериментальных данных больных НХЛ и ХЛЛ. Благодаря ее применению выявлена четкая двух кластерная структура в многомерных данных НХЛ и ХЛЛ.

2. Доказано однозначное соответствие кластерной структуры групповой структуре анализируемых данных: два кластера соответствуют распределениям пациентов двух групп анализируемых заболеваний, НХЛ и ХЛЛ.

3. Структуры данных ХЛЛ и НХЛ различаются по степени выхода опухолевых клеток в кровь (большей при ХЛЛ) и степени угнетения иммунитета (меньшей при НХЛ). Полученная интерпретация логична и не противоречит патофизиологии анализируемых состояний ХЛЛ и НХЛ.

4. Выдвинута гипотеза о том, что анализируемое множество данных может быть использовано для по-

строения высокоэффективной автоматической диагностической процедуры распознавания НХЛ и ХЛЛ. Важно, что при этом отпадает необходимость анализа показателей костного мозга, лимфоузлов и других тканей, что значительно облегчит процедуру диагностики анализируемых патологических состояний.

Литература

(п.п. 1-3; 6; 7; 9; 11; 16-18 см. References)

4. Российские клинические рекомендации по диагностике и лечению лимфопролиферативных заболеваний. Под ред. Поддубной И.В., Савченко В.Г. М.: 2018.
5. Злокачественные новообразования в России в 2017 году (заболеваемость и смертность). Под ред. Каприна А.Д., Старинского В.В., Петровой Г.В. МНИОИ им. П.А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России. М.: 2018.
8. Зиновьев А. Визуализация многомерных данных. Красноярск: 2000.
10. Тант Зин Пью, Тин Чжо, Пья Сон Ко Ко и Пайе Тэйн Наинга. Методика системы распознавания образов с помощью самоорганизующихся карт Кохонена нейронных сетей на основе *Matlab*. Интернет-журнал «Науковедение». 2013; 5.
12. Ярилин А.А. Иммунология. Учебник. М.: Гэотар-Медиа; 2010.
13. Неприна Г.С., Кузьмина Е.Г., Павлов В.В. и др. Иммунофенотипический анализ лимфоцитов костного мозга и периферической крови у больных В-клеточными злокачественными лимфомами. Медицинская иммунология. 2006; 8(2-3): 347-8.
14. Демидова И.А., Никитин Е.А. Молекулярные методы в онкогематологии. Гериатрическая гематология. М.: Медиум; 2011; 251-83.
15. Обухова Т.Н. Цитогенетика зрелоклеточных лимфатических опухолей. Гериатрическая гематология. М.: Медиум. 2011; 239-50.
19. Казанский Д.Б. Т-лимфоциты в развитии хронического лимфолейкоза. Клиническая онкогематология. Фундаментальные исследования и клиническая практика. Клиническая онкогематология. 2012; 5 (2): 84-95.
20. Свирновский А. И. Хронический лимфоцитарный лейкоз: парадигмы и парадоксы. Медицинские новости. 2008; 13: 7-19.
- an Research Initiative on CLL (ERIC) & European Society for Clinical Cell Analysis (ESCCA) Harmonisation project. *Cytometry B Clin Cytom.* 2018; 94: 121-8.
4. Poddubnaya I.V., Savchenko V.G. Russian recommendations for diagnosis and individual lymphoproliferative diseases. *Rossiyskiye rekomendatsii po diagnostike i otdel'nym limfoproliferativnym zabolevaniyam.* Moscow: 2018. (In Russian)
5. Kaprina A.D., Starinskiy V.V., Petrova G.V. *Malignant neoplasms in Russia in 2017 (morbidity and mortality). [Zlokachestvennyye novoobrazovaniya v Rossii v 2017 godu (zabolevayemost' i smertnost')]*. Moscow MNIIOI im. P.A. Herzen – branch of the Federal State Budgetary Institution «NMITsology» of the Ministry of Health of Russia. Moscow: 2018. (In Russian)
6. Kohonen T., Essentials of the self-organizing map. *Neural Networks.* 2013; (37): 52–65.
7. Dubravko Miljković. Brief Review of Self-Organizing Maps. *MIPRO.* May Opatija, Croatia. 2017; 22- 6.
8. Zinoviev A. *Preview of multivariate data. [Vizualizatsiya mnogomernykh dannykh]*. Krasnoyarsk: 2000. (In Russian)
9. Ji Zhang and Hai Fang. *Using Self-Organizing Maps to Visualize, Filter and Cluster, in: Applications of Self-Organizing Maps, chapter 9.* Ed., Magnus Johnsson. 2012.
10. Thant Zin Pyo, Tin Zhuo, Pya Son Ko Ko and Paye Thein Nainga. Methodology for a pattern recognition system using self-organizing Kohonen maps of neural networks based on Matlab. *Internet journal «Science».* 2013; 5.
11. Mathworks Inc., Self Organising Maps Toolbox, <https://www.mathworks.com>
12. Yarilin A.A. *Immunology. Textbook. [Immunologiya. Uchebnik]*. «Geotar-Media». 2010. (In Russian)
13. Neprina G.S., Kuzmina E.G., Pavlov V.V., et al. Immunophenotypic analysis of bone marrow and peripheral blood lymphocytes in patients with B-cell malignant lymphomas. *Meditsinskaya immunologiya.* 2006; 8(2-3): 347-8. (In Russian)
14. Demidova I.A., Nikitin E.A. Molecular methods in oncohematology. *Geriatricheskaya gematologiya.* Moscow: 2011: 251-83. (In Russian)
15. Obukhova T.N. Cytogenetics of mature cell lymphatic tumors. *Geriatricheskaya gematologiya.* Moscow: 2011: 239-50. (In Russian)
16. Kikushige Y., Ishikawa F., Miyamoto T., et al. Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer Cell.* 2011; 20: 246-59.
17. Rozman C., Montserrat E. Chronic lymphocytic leukemia. *N Engl J Med.* 1995; 333: 1052-7.
18. Scrivener S., Kaminski E.R., Demaine A., Prentice A.G. Analysis of the expression of critical activation/interaction markers on peripheral blood T cells in B-cell chronic lymphocytic leukaemia: evidence of immune dysregulation. *British Journal of Haematology.* 2001; 4(112): 959–64. DOI: 10.1046/j.1365-2141.2001.02672
19. Kazansky D.B. T-lymphocytes in the development of chronic lymphocytic leukemia. Clinical oncohematology. Basic research and clinical practice. *Klinicheskaya onkogematologiya.* 2012; 5(2): 84-95. (In Russian)
20. Svirnovsky A.I. Chronic lymphocytic leukemia: paradigms and paradoxes. *Meditsinskiye novosti.* 2008; 13: 7-19. (In Russian)

References

1. Hallek M., Al-Sawaf O. Chronic lymphocytic leukemia: 2022 update on diagnostic and therapeutic procedures. *Am J Hematol.* 2021; 96(12): 1679-705. <https://doi.org/10.1002/ajh.26367>
2. Swerdlow S.H., Campo E., Pileri S.A., et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 2016; 19; 127(20): 2375–90. DOI: 10.1182/blood-2016-01-643569
3. Cree I.A., Tumours A.C., Kreuzer K.A., Soosapilla A., et al. The WHO Classification of Haematolymphoid, Reproducible diagnosis of chronic lymphocytic leukemia by flow cytometry: an Europe-

Сведения об авторах:

Сабурина Ирина Николаевна, доктор биол. наук, гл. науч. сотр., лаб. клеточной биологии и патологии развития ФГБНУ НИИОПП, e-mail: saburina@mail.ru;

Кузьмина Евгения Геннадьевна, канд. биол. наук, вед. науч. сотр. отд-ния клинической иммунологии МРНЦ им. А.Ф. Цыба – филиал ФГБУ «НМИЦ радиологии» Минздрава России;

Черкашенко Владимир Николаевич, науч. сотр., лаб. клеточной биологии и патологии развития ФГБНУ НИИОПП;

Павлов Вячеслав Владимирович, канд. мед. наук, врач-гематолог отд-ния лучевой и лекарственной терапии гемобластозов МРНЦ им. А.Ф. Цыба – филиал ФГБУ «НМИЦ радиологии» Минздрава России;

Зацаренко Светлана Валерьевна, науч. сотр. клинической иммунологии МРНЦ им. А.Ф. Цыба – филиал ФГБУ «НМИЦ радиологии» Минздрава России;

Антонов Кирилл Николаевич, студент факультета вычислительной математики и кибернетики (ВМК) ФГБОУ ВО «МГУ им. М.В. Ломоносова»;

Богданенко Елена Валентиновна, ст. науч. сотр. лаб., клеточной биологии и патологии развития ФГБНУ НИИОПП;

Устинова Екатерина Евгеньевна, мл. науч. сотр., лаб. клеточной биологии и патологии развития ФГБНУ НИИОПП;

Морозов Сергей Георгиевич, доктор мед. наук, проф., член-кор. РАН, директор ФГБНУ НИИОПП.