

**Н.С. КАРНАУХОВ,**

к.м.н., Ростовский научно-исследовательский онкологический институт, г. Ростов-на-Дону, Россия,  
e-mail: nick07@bk.ru

**Р.Г. ИЛЬЮХИН,**

Ростовский государственный медицинский университет, г. Ростов-на-Дону, Россия,  
e-mail: ruslan.spark@icloud.com

## ВОЗМОЖНОСТИ ТЕХНОЛОГИЙ «BIG DATA» В МЕДИЦИНЕ

**УДК 614.2**

Карнаухов Н.С., Илюхин Р.Г. *Возможности технологий «Big Data» в медицине* (Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный медицинский университет» Министерства здравоохранения Российской Федерации, Ростовский научно-исследовательский онкологический институт, г. Ростов-на-Дону, Россия)

**Аннотация.** Статья посвящена исследованию особенностей использования технологий анализа больших объемов неструктурированных и многообразных данных («Big Data») в медицине. Рассмотрена история развития, целесообразность применения и накопленный опыт по использованию технологий «Big Data» в Европе и России, определены их возможности и преимущества, а также отдельное внимание уделено перспективным направлениям развития. Детально проанализирована сфера их применения на примере ранней диагностики и лечения онкологических заболеваний, где они нашли применение при составлении и изучении «омиксных» баз данных, которые включают в себя генетическую информацию каждого отдельно взятого пациента и открывают возможности персонализированного подбора лечения. Также отмечен потенциал аналитических технологий в системе здравоохранения и эпиднадзора.

**Ключевые слова:** «Big Data», биоинформатика, онкология, анализ информации, персонализированная медицина

**UDC 614.2**

Karnaukhov N.S., Ilyukhin R.G. *Capabilities of «Big Data» technologies in medicine* (Federal State Budgetary Educational Institution of Higher Education «Rostov State Medical University» of the Ministry of Healthcare of the Russian Federation, Rostov Research Institute of Oncology, Rostov-on-Don, Russia)

**Abstract.** The article is devoted to the study of the peculiarities of using the «Big Data» technology in medicine. The scope of this technology has been analyzed in detail. On the example of the history of development and experience the «Big Data» technology in Europe and Russia and early diagnostics of oncological diseases by compiling genomic databases, that include the genetic information of each individual patient, its advantages, and disclosed opportunities for treatment are determined. Special attention is paid to the prospects of the development of «Big Data» technology in medicine and healthcare. The potential of analytical technologies in the epidemiological surveillance was also noted.

**Keywords:** «Big Data», bioinformatics, oncology, data analysis, personalized medicine.

За последние несколько десятилетий медицинские идеи и технологии, считавшиеся ранее научной фантастикой, стремительно ворвались в фундаментальную и клиническую медицину, что позволяет ей совершить переход от стандартизированных методов лечения к индивидуальному подбору лекарственных средств с учетом генетических особенностей каждого конкретного пациента [1, 2].

Кроме того, генерация идей на стыке медицины и информационных технологий меняет привычные подходы к мониторингу состояния здоровья и лечению. В данном контексте, рассматривая медицину как комплексную и специфическую сферу деятельности, которая характеризуется очень высоким уровнем разнородной лечебной и экономической информации, актуализируются две задачи – оптимизация клинической и экономической деятельности медицинского учреждения. Конечной целью этих задач является оптимизация деятельности системы здравоохранения в целом и любого медицинского учреждения в частности.

Не вызывает сомнения тот факт, что новейшие технологии и создание экспертных систем в медицине, способных на основе симптомов болезней

и лабораторно-инструментальных данных предвидеть и достоверно прогнозировать возможные заболевания, могут значительно облегчить принятие решения специалистами в медицинской сфере. Такие технологии помогут выбрать врачу наиболее важные результаты в огромном количестве информации из истории болезни для назначения адекватного лечения.

Внедрение автоматических систем обработки и анализа больших объемов данных актуальны не только в клинической практике, но и в системе здравоохранения в целом для планирования финансирования и управления медицинскими организациями, осуществления контроля по обороту лекарственных средств и ценообразования в масштабах страны, с учетом особенностей каждого региона, для достижения лучшей эффективности.

Мы полагаем, что для решения вышеуказанных задач можно применять технологии «Big Data». Под данным термином принято понимать технологии, позволяющие обрабатывать огромные объемы структурированной и неструктурированной информации. Использование подобных технологий в некоторых случаях дает возможность анализировать данные историй болезни в совокупности с результатами десятков лабораторно-инструментальных исследований и ставить окончательный диагноз в разы быстрее и эффективнее; помогает врачам, основываясь на многопараметрических моделях, определить риски возникновения заболевания или осложнений заболевания, сократить время проведения научных исследований с года до нескольких недель.

По информации агентства Ovum, рынок инструментов для аналитической обработки данных о пациентах вырастет до \$11 млрд. к 2018 г. [3]. Согласно последним прогнозам, сумма инвестиций в программы по поддержке проектов, основанных на технологии «Big Data» для медицины, может достичь 656 млн. долларов к 2020 г. [4]. Такие разработки стали особо актуальными с появлением преобразующих технологий для медико-биологических исследований, к числу которых относится секвенирование ДНК и визуализация биомедицинских данных.

Таким образом, исследование особенностей и перспектив применения технологий «Big Data» в медицине и здравоохранении является актуальной научно-практической задачей.

Стремительное развитие биоинформатики, а также возможности анализа больших данных открывает для создания принципиально новых методов

диагностики и лечения, вызывает необходимость проведения дальнейших углубленных исследований в направлении применения аналитических технологий в медицине.

Медицина одна из первых обратилась к информационным технологиям в 60-х годах прошлого века с вопросом сбора, хранения и обработки больших объемов различных статистических данных, которых к тому времени накопилось большое количество, причем большая часть на бумажных носителях. На сегодняшний день количество медицинских данных значительно увеличилось в связи с переходом отрасли на цифровой формат. Несмотря на это значительная часть медицинских данных продолжает храниться пока в бумажном виде, но наблюдаемая тенденция к быстрой оцифровке уже способствовала накоплению терабайтов информации [5]. По некоторым расчетам объем медицинских данных в 2014 г. достиг примерно 760 петабайт [6]. Последние исследования показывают, что более 30% всех данных, хранящихся на планете, составляют медицинские данные, и в будущем ожидается быстрое увеличение этой доли за счет создания новой и оцифровки уже имеющейся информации [7]. Прогнозируется, что к 2020 г. количество медицинских данных будет достигать ~25000 петабайт [8]. Немаловажным является и тот факт, что вся эта информация уже не исключительно статистическая, а представляет собой массив неструктурированных и разнородных данных, что определяет различные инструменты и методы для их сбора, хранения и анализа.

Таким образом, «Big Data» в медицине – это непрерывно и быстро пополняемые электронные массивы качественно различающихся между собой медицинских и парамедицинских данных огромного объема, которыми невозможно управлять посредством традиционных инструментов и методов программного и/или аппаратного обеспечения [9].

Особую актуальность технологии «Big Data» приобрели в процессе анализа молекулярных данных, получаемых путем изучения структуры молекул ДНК (геномика и эпигеномика), молекул РНК (транскриптомика), белковых молекул (протеомика), взаимодействии клеточных метаболитов (метабономика) и других типов подобных данных, где скорость генерации информации в разы превышает скорость ее обработки. Это, в свою очередь, обозначило одно из наиболее перспективных направлений использования технологий «Big Data» в медицине, так как решение проблемы эффективного анализа



данных предоставляет значительное преимущество в ранней диагностике онкологических заболеваний.

Очевидно, что смертность в онкологии была бы гораздо ниже, а лечение – эффективнее, если бы все опухоли удавалось обнаружить на ранних стадиях. Многие исследователи работают над поиском надежных биомаркеров диагностики заболеваний на основе различных методов: от секвенирования ДНК и анализа модификации гистонов до масс-спектрометрии и жидкостной хроматографии [10]. Но результаты этих исследований представляют собой большой массив данных, обработка которых невозможна без специальных информационно-аналитических систем, которые были бы способны находить заранее заданные паттерны среди огромного объема «омиксных» баз данных.

Рассматривая в данном контексте накопленный опыт применения технологий «Big Data», следует отметить, что в 2012 году Европейская лаборатория молекулярной биологии Института Биоинформатики в Кэмбридже, пользуясь системой «HiSeq X Ten», секвенировала 18.000 человеческих геномов за год.

А уже в 2014 году Великобритания запустила государственный проект 100.000 геномов, который ведет специально созданная для этого компания «Genomics England». Данный проект предполагает хранение 100 тысяч человеческих геномов, значительная часть этой информации на сегодняшний день уже собрана [11].

В Исландии «deCODE Genetic» продемонстрировали силу популяционной геномики, совмещая обширную генеалогическую информацию и истории болезней с геномной информацией 150.000 людей (включая 15.000 полностью секвенированных геномов). Эти результаты позволили «deCODE» обобщить распространенность известных генетических факторов риска в популяции, включая генные варианты, связанные с раком молочной железы, диабетом и болезнью Альцгеймера [12].

В отечественных медицинских учреждениях, к сожалению, пока нет глобальной программы работы с большими данными. Как отмечает старший научный сотрудник Института вычислительной математики РАН, профессор Сколковского института науки и технологий И. Оселедец: «В России есть проблема, прежде всего, со сбором информации, так что о технологиях ее обработки сейчас говорить преждевременно. В отдельных медучреждениях ведутся пилотные проекты с использованием больших объемов данных, но в целом в отрасли остается очень много учреждений, где даже самое

высокотехнологичное диагностическое оборудование не подключено к Интернету, не собирает, не накапливает и не анализирует информацию обо всем потоке проходящих через него пациентов» [13].

При этом следует акцентировать внимание на том, что в РФ есть единичные стартапы, в основном они находятся на ранней стадии развития и предлагают специализированные решения для обработки медицинских и «омиксных» данных. К числу таких компаний относится, например, «Genotek», «ParseqLab», «Atlas», «Мой ген» и некоторые др.

Кроме того, согласно дорожной карте «Healthnet», посвященной развитию технологий в медицине на период до 2035 г. в сегменте «ИТ в медицине» планируется разработать системы поддержки принятия решений с использованием алгоритмов обработки больших объемов данных для локального использования в медицинских организациях [14].

Стоит отметить, что сегодня «Big Data» – это не только информация, но и инструмент, который имеет неограниченные возможности и способствует получению новых решений, качественной трансформации процессов медицинской помощи и прогрессивному развитию системы здравоохранения в целом. Ученые исследуют возможности использования больших объемов данных для улучшения уровня доказательности и принятия клинических решений. В настоящее время технологии «Big Data» используются для: совершенствования принятия решений в клинической практике; выявления пациентов группы риска, у которых возникли побочные реакции или установлена неэффективность лекарственных средств; повышения качества индивидуального обслуживания пациентов с целью получения лучших результатов; улучшения показателей доказанной эффективности лекарств с учетом данных, полученных во время реальной практики; улучшения планирования ресурсов здравоохранения.

Стоит отметить и перспективные направления применения технологий анализа больших объемов информации в сфере информатизации медицины и практического здравоохранения:

1. Прежде всего, это интеграция геномной информации с электронными историями болезни, что приобретает особую актуальность для европейских стран, которые нацелены на то, чтобы данный формат сделать стандартом национальной медицинской помощи.

2. Также следует обратить внимание на необходимость формирования хорошей платформы для



работы с геномными данными, поскольку при секвенировании генома создается огромный файл, и исследователю необходимо уметь извлекать из него конкретные участки, делать необходимые запросы, применять к их результатам программные инструменты биоинформатики. В таких запросах могут участвовать десятки тысяч файлов и десятки инструментов, передающих друг другу результаты своей работы.

3. Перспективным направлением является совмещение информации о генотипе и фенотипе. Как показывает практика, это направление является очень выгодным с исследовательской точки зрения. Наиболее клинически релевантные варианты генов могут быть идентифицированы с помощью ассоциативных исследований целых геномов, в которых люди с определенными заболеваниями обследовались с целью обнаружения связанных с ними генетических сигнатур. В результате исследователи смогут отталкиваться от историй болезни, чтобы определить, какие клинические проявления наиболее распространены у лиц с конкретным генетическим вариантом.

4. Усовершенствование эпиднадзора за неинфекционными заболеваниями (НИЗ). В сентябре 2013 г. состоялось техническое совещание по использованию больших данных и социальных сетей для эпиднадзора за НИЗ, которое было организовано совместно Европейским региональным бюро ВОЗ и инициативой ООН «Глобальный пульс». Целью совещания было рассмотреть потенциал применения для эпиднадзора за НИЗ, в дополнение к традиционным методам сбора данных, также новых источников, в частности социальных сетей, цифрового следа ежедневных транзакций и массовых анонимных электронных данных о состоянии здоровья.

Данные социальных сетей обладают потенциалом для оценки установок и поведенческих характеристик населения в отношении, например, НИЗ

и их факторов риска, который может быть полезным при изучении тенденций и в ходе анализа взглядов населения по важным темам, что можно использовать в процессе разработки стратегий и кампаний общественного здравоохранения. Цифровой след (данные мобильной телефонии, транзакций из супермаркетов, записи операций по кредитным картам) может давать новые представления относительно уже существующих данных здравоохранения и обеспечивать доступ к информации о тех сегментах населения, которые трудно обследовать с использованием традиционных методов эпиднадзора.

Подводя итоги, не подлежит сомнению, что технологии «Big Data» могут оказать существенное воздействие на медицину и здравоохранение, причем как в плане повышения точности диагностики, эффективности лечения и своевременного предупреждения болезней, так и в контексте улучшения работы самих медицинских учреждений. Так, по оценкам многих аналитических компаний, внедрение технологий анализа «Big Data» в медицину может сэкономить для здравоохранения большую часть бюджета. Экономия произойдет за счет более своевременной и точной постановки диагноза, сокращения пребывания пациента в стационаре, подбора эффективного лечения, сокращения расходов на исследования. Ожидается, что с помощью использования современных аналитических систем эффективность лечения будет повышена благодаря обработке всей доступной информации. В практику врачей войдет широкое использование систем поддержки принятия решений и экспертных систем нового поколения, позволяющих предоставить врачам всеобъемлющий доступ к опыту коллег посредством анализа электронных медицинских паспортов пациентов как в локальном варианте, так и в масштабе страны. Это, в свою очередь, позволит минимизировать субъективный человеческий фактор при принятии врачебных решений о стратегии лечения пациента.

## ЛИТЕРАТУРА



1. Макарова Е.П., Розенцвит С.К., Буллик А.В. Перспективные инновационные направления в сфере здравоохранения в России. Научная дискуссия: инновации в современном мире. – 2016. – № 6-1 (49). – С. 67-73.
2. Кит О.И., Владимиров Л.Ю., Абрамова Н.А., Сторожакова А.Э., Попова И.Л., Тихановская Н.М., Агиева А.А., Рядинская Л.А., Никипелова Е.А., Ежова М.О. Возможности таргетной терапии предлеченных больных с метастатическим колоректальным раком. Современные проблемы науки и образования. – 2017. – № 4. – С. 6.



3. Davis-Dusenbery Brandi Precision Medicine & Big Data. Pharmaceutical executive. – 2017. – № 3. – С. 14.
4. Wu P.Y., Cheng C.W., Kaddi C.D., Venugopalan J., Hoffman R., Wang M.D. Omic and Electronic Health Record Big Data Analytics for Precision Medicine. IEEE transactions on biomedical engineering. – 2017. – № 2. – С. 263–273.
5. Тушенцова К.В. Современные аналитические технологии в медицине. Теория и практика современной науки. – 2016. – № 5(11). – С. 955–957.
6. Lederman R., Dreyfus S. Managing Health Information Delivery Processes for Better Medical Decision Making. Frontiers in artificial intelligence and applications. – 2014. – Jun. – С. 329–340.
7. Spagnuolo D., Lenzini G. Transparent Medical Data Systems. Journal of medical systems. – 2017. – № 1. – С. 1–12.
8. Gorzałczany Marian B.; Rudziński Filip Interpretable and accurate medical data classification – a multi-objective genetic-fuzzy optimization approach. Expert systems with applications. – 2017. – № 71. – С. 26–39.
9. Цветкова Л.А., Черченко О.В. Технология больших данных в медицине и здравоохранении России и мира. Врач и информационные технологии. – 2016. – № 3. – С. 60–73.
10. Ильясова Н.Ю., Куприянов А.В., Попов С.Б., Парингер Р.А. Особенности использования технологий Big Data в задачах медицинской диагностики. Системы высокой доступности. – 2016. – 12(1). – С. 45–52.
11. McGrath J.A. Rare inherited skin diseases and the Genomics England 100 000 Genome Project. British journal of dermatology. – 2016. – № 2. – С. 257–258.
12. Arnar David O.; Andersen Karl; Thorgeirsson, Gudmundur Genetics of cardiovascular diseases: lessons learned from a decade of genomics research in Iceland. Scandinavian cardiovascular journal. – 2016. – № 5. – С. 260–265.
13. Лисица А.В., Пономаренко Е.А., Лохов П.Г., Арчаков А.И. Постгеномная медицина: альтернатива биомаркерам. Вестник Российской академии медицинских наук. – 2016. – № 3. – С. 255–260.
14. Healthnet URL: <http://www.nti2035.ru/markets/healthnet> (accessed 15.12.2018).

## Новости отрасли



### УТВЕРЖДЕН ПРОЕКТ «СОЗДАНИЕ ЕДИНОГО ЦИФРОВОГО КОНТУРА НА ОСНОВЕ ЕГИСЗ»

По итогам заседания президиума Совета при Президенте Российской Федерации по стратегическому развитию и национальным проектам 24 декабря 2018 года был утвержден паспорт национального проекта «Здравоохранение». Одной из программ этого проекта является «Создание единого цифрового контура здравоохранения на основе ЕГИСЗ».

Этот федеральный проект рассчитан на 2019–2024 гг. Цель: повышение эффективности функционирования здравоохранения России путем дальнейшего развития и углубленного внедрения информационных и платформенных решений, что по мнению Министерства здравоохранения, сформирует так называемый «единый цифровой контур».

На реализацию проекта в 2019–2024 гг. планируется выделить порядка 177 млрд. руб. целевого финансирования, причем 60% этой суммы выделит федеральный центр, остальное будет финансироваться за счет региональных бюджетов.

Источник: <http://www.kmis.ru/blog/o-proekte-sozdaniia-edinogo-tsifrovogo-kontura>