

Л.О. МИНУШКИНА¹, О.Ю. КОЛЕСНИЧЕНКО², А.Л. МАЗЕЛИС³, Л.С. МАЗЕЛИС³, Д.А. СОЛДАТОВ³, Ю.Ю. КОЛЕСНИЧЕНКО⁴, А.В. МАРТЫНОВ⁵, В.В. ПУЛИТ⁵, А.Н. ДОЛЖЕНКОВ⁵, И.Н. ГРИГОРЕВСКИЙ⁶

¹ ФГБУ ДПО «Центральная государственная медицинская академия УД ПРФ», Москва

² Институт социологии, психологии и гуманитарных наук, ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова» Минздрава России, Москва

³ ФГБОУ ВО «Владивостокский государственный университет экономики и сервиса», Владивосток

⁴ Uzgraph, Москва

⁵ Компания СП.АРМ, Санкт-Петербург

⁶ Институт программных систем им. А.К. Айламазяна РАН, Национальная суперкомпьютерная технологическая платформа, Переславль-Залесский

10.21518 / 1561-5936-2018-7-8-30-40

Аналитика Больших данных МИС НА ОСНОВЕ КЛАСТЕРНОГО АНАЛИЗА: ПАЦИЕНТЫ С СД

Результаты кластерного анализа записей МИС были использованы для поиска прямых и косвенных указаний на различия в стационарном ведении пациентов с сахарным диабетом 1-го типа (СД 1) – в назначаемых диагностических процедурах и операциях. Такой анализ позволяет углубиться в патогенетические механизмы развития заболевания и служит для прогнозирования потоков пациентов, формирования персонализированного подхода к терапии и планирования стационарной медпомощи. Выявлено 5 кластеров, определены кластерообразующие медуслуги, кластеры подробно описаны. Дополнительно проанализированы открытые международные генетические базы данных для некоторых ключевых с точки зрения патогенеза СД 1 молекулярных комплексов. Все данные сопоставлены. Впервые предложен связанный континуум патогенетического каскада при СД 1, построенный на основе кластерного анализа потока пациентов по данным МИС. Континуум отражает ключевые звенья патогенеза.

Ключевые слова:

кластерный анализ,
СД 1, континуум
патогенетического
каскада, медицинские
информационные системы

ВВЕДЕНИЕ

Информационная трансформация медицины вывела клиническое мышление врача на новый уровень, который можно обозначить как синтетизм. Врач не в состоянии сам справиться со всеми потоками информации, особенно с учетом самых разных комбинаций заболеваний у больного. В этом ему призваны помочь медицинские информационные системы (МИС) и алгоритмы искусственного интеллекта (системы поддержки принятия решений). Но чтобы создать такие алгоритмы, необходимо сначала изучать закономерности, которые до информационного этапа не были доступны для анализа, так как не было таких технологических возможностей. Синтетизм в клиническом мышлении врача – это способность к интеграции разных знаний в единую картину. Чем больше разных данных будет записываться в МИС, тем сложнее их синтез **в единый континуум**. Основной проблемой является наличие технической и финансовой возможности

SUMMARY

Keywords: Cluster analysis, Diabetes Mellitus type 1, pathogenetic cascade continuum, medical information system

Cluster analysis of medical information system's records used to find direct and indirect indications to differences in the hospital management of patients with Diabetes Mellitus type 1, according to the prescribed diagnostic examinations and surgery. This approach allows to investigate disease's pathogenetic mechanisms and provides predictive analytics of patients flow for more personalized treatment. Five clusters have been identified. Open international genetic databases were analyzed additionally in terms of some key pathogenesis molecular complexes. All data was compared and connected continuum of the Diabetes Mellitus type 1 pathogenetic cascade was proposed.

L.O. MINUSHKINA¹, O.YU. KOLESNICHENKO², A.L. MAZELIS³, L.S. MAZELIS³, D.A. SOLDATOV³, YU.YU. KOLESNICHENKO⁴, A.V. MARTYNOV⁵, V.V. PULIT⁵, A.N. DOLZHENKOV⁵, I.N. GRIGOREVSKY⁶

¹Central State Medical Academy at the Department of Presidential Affairs, Moscow

²Institute of sociology, psychology and humanities, I.M. Sechenov First Moscow State Medical University, Moscow

³Vladivostok State University of Economics and Service, Vladivostok

⁴Uzgraph, Moscow

⁵SP.ARM, St. Petersburg

⁶Aylamazyan Program Systems Institute of Russian Academy of Sciences, National Supercomputing Technology Platform, Pereslavl-Zalesky

MEDICAL INFORMATION SYSTEM'S BIG DATA ANALYTICS BASED ON CLUSTER ANALYSIS: PATIENTS WITH DIABETES MELLITUS TYPE 1

для сбора максимального объема всевозможных медицинских данных о пациенте, особенно данных современных генетических и биохимических лабораторных исследований, которые дороги и малодоступны.

Представляется, что сбор Больших данных с потоков пациентов будет настолько критически важным для развития медицины, что возникнет практика приглашения пациентов для бесплатного обследования, чтобы максимально заполнить ту или иную базу данных для ее дальнейшего научно-практического использования. Такая практика уже внедрена в сфере сбора генетических данных, скоро такие же тенденции наметятся и в клинической медицине в целом. Однако не стоит ждать появления масштабных и исчерпывающих по содержанию баз данных по потокам пациентов. Необходимо начать активно анализировать имеющиеся накопленные записи МИС с позиции синтетизма.

Например, в данной статье анализируется кластеризация потока пациентов с сахарным диабетом 1-го типа (СД 1) по записям МИС нескольких больниц за 5 лет. Кластерный анализ записей МИС ранее уже проводился [1–3] и показал свою релевантность для медицинского анализа (более активно кластерный анализ применяется в социологии и молекулярной биологии). В представленной статье кластеризация используется не только для поиска прямых указаний на различия в стационарном ведении пациентов (одной ведущей нозологической группы) по назначаемым диагностическим процедурам и операциям, но и для поиска косвенных указаний на причины выявляемых различий. Результаты такого анализа помогают углубиться в патогенетические механизмы заболеваний и служат основой для разработки алгоритмов систем поддержки принятия решений, а также для прогнозирования потоков пациентов с целью планирования и оптимизации затрат.

МЕТОДЫ

Анализировались данные МИС qMS (СПАРМ) за период с 2013 по 2017 г. (несколько разных больниц), кото-

рые представляют собой коды клинических диагнозов по МКБ-10 (нозологическая группа СД 1, E10), отметки о количестве и видах проведенных обследований, процедур и операций, а также анализы крови (были выбраны результаты исследования уровня гликозилированного гемоглобина (HbA1c, %) и паратгормона (ПТГ, пг/мл). Общая выборка включала 862 пациента (304 мужчины и 558 женщин). Исследовательской группе не передавались персональные данные пациентов, информация о каждом из них кодировалась отдельным кодом по каждому эпизоду госпитализации.

Кластерный анализ проведен по полному набору диагностических процедур и операций (по медуслугам) иерархическим методом и методом k-средних. Иерархическая кластеризация позволила определиться с рациональным числом кластеров. Было принято решение о разделении полной выборки пациентов методом k-средних на 5 кластеров. Для оценки валидности разделения на кластеры использовали индекс Дэвиса – Болдина. Анализы крови (HbA1c и ПТГ) вписывались уже в построенные кластеры, то есть в самой кластеризации не участвовали. Кластеризация основана исключительно на коллективном «решении врача» о назначении той или иной процедуры и операции, что отражают данные МИС. Дополнительно проведен кластерный анализ этой же матрицы данных по самим медуслугам, без пациентов, выявлено 5 кластеров. Фактически это разделение медуслуг на группы по признаку совместного их назначения и по частоте их встречаемости в анализируемой матрице данных.

В качестве инструмента для проведения кластеризации использован язык программирования Python. Для автоматизации расчетов была написана программа в интерактивной среде iPython с использованием библиотек NumPy, Pandas и Sklearn, предназначенных специально для анализа и обработки данных.

Исследование выполнялось под эгидой Национального суперкомпьютерного форума – Национальной

суперкомпьютерной технологической платформы. Результаты докладывались на ежегодном форуме, и аналитические подходы корректировались с учетом замечаний после обсуждения докладов научной аудиторией.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Вспомогательная кластеризация по медуслугам

Кластеризация по медуслугам без пациентов (вспомогательный кластерный анализ) показала, что есть два вида обследования пациентов с СД 1, которые назначаются большинству из них: ЭКГ (n = 719) и УЗИ органов брюшной полости и почек (n = 497). Это первый кластер. В дальнейшем эти два вида обследований не учитывались при анализе результатов основной кластеризации по пациентам, так как конкретно эти два вида записей в МИС не дают различий по кластерам.

Второй кластер объединил три вида обследования больных: эхокардиографию (n = 345), УЗИ щитовидной железы (n = 257) и денситометрию костей скелета (n = 191). Это базовая распространенная триада назначений обследований, которые делали врачи больным с СД 1. Третий кластер содержит только один вид назначений – курс лечебной физкультуры (n = 150).

Четвертый кластер включал восемь видов обследования: суточное мониторирование ЭКГ (n = 115), отдельно УЗИ почек и надпочечников (n = 56), отдельно УЗИ органов брюшной полости (n = 55), УЗИ органов малого таза (n = 76), эзофагогастродуоденоскопию (n = 187), рентгенографию легких (n = 192), рентгенографию суставов и конечностей (n = 69), физиотерапию (n = 60). Это блок часто встречающихся назначений у данной категории больных в том или ином сочетании. Пятый кластер – это все остальные медуслуги, зафиксированные системой в анализируемой матрице данных.

Вспомогательная кластеризация по медуслугам без пациентов учитывалась при детальном анализе результатов кластеризации по пациентам

ТАБЛИЦА 1 Характеристика кластеров пациентов с СД 1 (основной кластерный анализ)

Кластер	Количество пациентов (%)*	Возраст, лет	Дни в стационаре	Название кластера
Кластер Мах-5 1к	n = 127 (14,7%)	50,90 ± 1,56	21,76 ± 1,45	Сердечно-легочный, реактивно-парадоксальный (с реакцией паратгормона)
Кластер Мах-5 2к	n = 205 (23,8%)	42,05 ± 1,20	15,86 ± 0,48	Тиреоидный и высокого риска рака легкого
Кластер Мах-5 3к	n = 387 (44,9%)	45,94 ± 0,86	12,42 ± 0,39	Хирургический, гемодиализный, терминальный по осложнениям
Кластер Мах-5 4к	n = 42 (4,9%)	51,36 ± 3,19	19,31 ± 1,24	Лечебно-физкультурный
Кластер Мах-5 5к	n = 101 (11,7%)	48,31 ± 1,46	18,18 ± 0,79	Денситометрический, с плохим гликемическим контролем

* % от всех 862 пациентов в выборке.

(основной кластерный анализ). Часто назначаемые процедуры как общие скрининговые могут искажать уникальный профиль кластера, особенно если кластер немногочисленный по сравнению со всей выборкой.

Основной кластерный анализ

В описании основного кластерного анализа использовались два подхода. С одной стороны, показательным является отношение числа встречаемости медуслуги определенного вида к количеству пациентов в кластере (в процентах) – *долевое распределение в кластере вида медуслуги*. С другой стороны, надо учитывать и процент встречаемости вида медуслуги по отношению ко всему количеству медуслуг данного вида во всей анализируемой выборке – *абсолютное распределение в кластере вида медуслуги*.

Без учета обоих показателей можно не заметить асимметрии в распределении медуслуг, что исказит понимание данных. Таким образом, кластер должен быть охарактеризован двумя упомянутыми показателями, причем оба показателя являются кластерообразующими.

В *таблице 1* представлены характеристики основной кластеризации по пациентам, она обозначена как Мах-5 (учет максимально всей матрицы данных с разделением на 5 кластеров). Каждый отдельный кластер имеет свое обозначение: 1к, 2к, 3к, 4к, 5к. Используя долевое и абсолютное распределение в кластере вида медуслуги, а также два показателя крови, кластерам дали названия. В *таблице 2* представлены средние значения гликозилированного гемоглобина и паратгормона по кластерам, а также процент пациентов, которым

назначались эти анализы в рамках их пребывания в стационаре (в МИС зафиксированы только те анализы, которые делались непосредственно в стационаре). На *рисунке 1* средние уровни HbA1c и ПТГ для каждого кластера представлены графически.

В *таблице 3* кластеры представлены с позиции распределения зафиксированных в МИС случаев диабетической комы (МКБ E10.0). Коды МКБ в процессе выполнения кластеризации не были задействованы (учитывались только медуслуги). Кластер Мах-5 3к, обозначенный как терминальный, характеризуется наибольшим количеством случаев комы.

Оценка компенсации углеводного обмена и уровня паратгормона

На *рисунке 2* показано взаимоотношение уровней (по каждому случаю) паратгормона и гликозилированного

ТАБЛИЦА 2 Распределение средних уровней гликозилированного гемоглобина и паратгормона по кластерам пациентов с СД 1

Кластер	Средний уровень гликозилированного Hb (HbA1c), %*	% (n) обследованных в кластере на HbA1c	Средний уровень паратгормона, ПТГ, пг/мл**	% (n) обследованных в кластере на ПТГ
Кластер Мах-5 1к	8,36 ± 0,36	67,7% (86)	170,47 ± 27,31	33,8% (43)
Кластер Мах-5 2к	8,89 ± 0,25	84,9% (174)	104,47 ± 27,90	14,6% (30)
Кластер Мах-5 3к	9,42 ± 0,48	43,7% (169)	168,31 ± 27,46	11,4% (44)
Кластер Мах-5 4к	8,51 ± 0,64	78,6% (33)	39,17 ± 6,38	21,4% (9)
Кластер Мах-5 5к	9,33 ± 0,21	88,1% (89)	83,05 ± 15,50	59,4% (60)

*Нормальный диапазон значений HbA1c: 4–5,9% [4].

**Нормальный диапазон значений ПТГ: 15–80 пг/мл; у пациентов с ХПН, находящихся на гемодиализе, содержание ПТГ в крови предпочтительно должно быть 150–300 пг/мл [4].

РИСУНОК 1 Графическое представление распределения средних уровней паратгормона и гликозилированного гемоглобина по кластерам

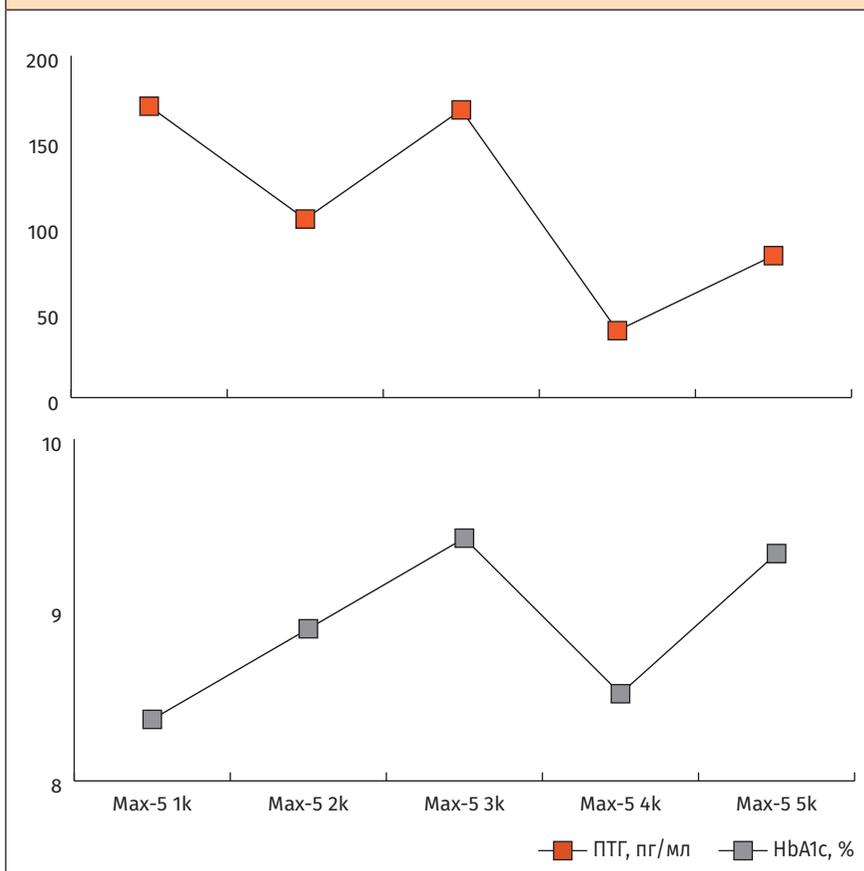


ТАБЛИЦА 3 Распределение случаев комы (E10.0) по кластерам пациентов с СД 1

Кластеры Max-5	1k	2k	3k	4k	5k
Долевое распределение в кластере (процент встречаемости случаев от количества пациентов в кластере)	1,6%	6,8%	7%	0%	0%
Абсолютное распределение в кластере (процент встречаемости случаев от всего количества случаев в анализируемой выборке, n = 44)	4,5%	32%	64%	0%	0%

гемоглобина для всей выборки (у тех пациентов, у кого были сделаны оба анализа, n = 145). Статистически значимой корреляции и регрессии не выявлено, как не выявлено и статистически значимых различий показателей крови между кластерами. Это обусловлено тем, что пациенты не ранжировались по жестким критериям (диагностические, возрастные группы и т. п.), а анализировалась вся выборка записей в МИС с отметкой

МКБ E10, которая имела разнородный состав по возрасту, полу, целям госпитализации, лечению, сопутствующим заболеваниям и осложнениям. Значения HbA1c выше нормально-го уровня свидетельствуют о плохой компенсации углеводного обмена у пациентов. Можно отметить, что прослеживается нелинейная зависимость, которая до уровня HbA1c 9% имеет черты возрастающей функции, а после уровня HbA1c 9%

имеет черты убывающей функции. Всего в выборке пациентов отметка E10.7, с множественными осложнениями, стоит у 56% человек, отметка E10.2, с поражением почек, – у 20%. Во всей выборке пациентов только у четверых был записан в МИС сопутствующий диагноз «гиперпаратиреоз» (два пациента имели отметку E21.0, первичный гиперпаратиреоз; два пациента имели отметку E21.1, вторичный гиперпаратиреоз). Все эти случаи вошли в кластер Max-5 3k. При СД происходит прямое воздействие глюкозотоксичности и оксидативного стресса на ткани (включая стенки сосудов) и неферментативное гликозилирование белков организма, в том числе базальной мембраны капилляров почечных клубочков и сосудов (микроангиопатия, гломерулосклероз, артериосклероз, интерстициальный фиброз). По мере поражения почек (диабетическая нефропатия) возникают нарушения фосфорно-кальциевого обмена, гиперфосфатемия, гипокальциемия и ответная гиперсекреция ПТГ околощитовидными железами (вторичный и третичный гиперпаратиреоз). При утяжелении хронической почечной недостаточности (ХПН) развивается дефицит синтезируемого в почках кальцитриола, что в итоге приводит к снижению всасывания Ca²⁺ в кишечнике (кальцитриол индуцирует синтез кальций-переносящих белков, обеспечивающих всасывание Ca²⁺ из полости кишечника). В результате гипокальциемия усугубляется, и для ее компенсации Ca²⁺ начинает поступать в кровь из костей скелета – происходит активация процессов костной резорбции (остеопения и остеопороз) [4–7].

Есть исследования, в которых показана обратная зависимость между HbA1c и ПТГ, то есть ухудшение гликемического контроля сопровождалось снижением уровня ПТГ (ранее повышенного на начальной стадии развития ХПН), что связывают с угнетающим влиянием глюкозы на паратиреоциты со снижением ими секреции ПТГ [6, 8–10]. Гипокальциемия оказывает негативное влияние на сердце, провоцируя

РИСУНОК 2 Графическое представление взаимоотношения уровней (по каждому случаю) паратгормона и гликозилированного гемоглобина по всей выборке (n = 145)

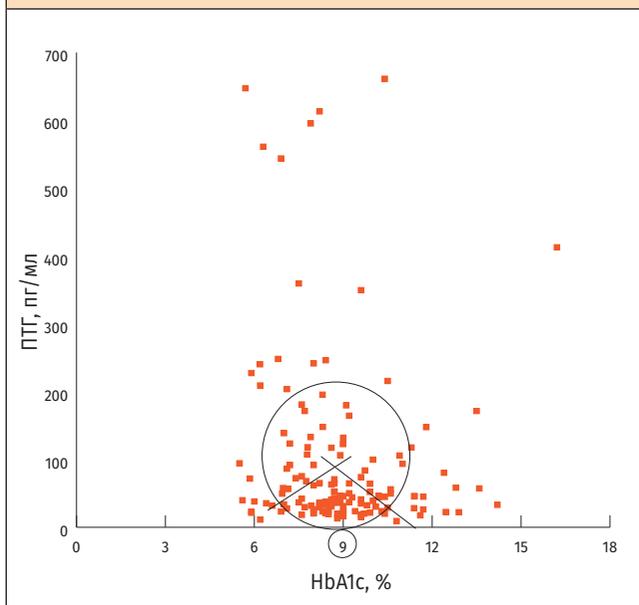
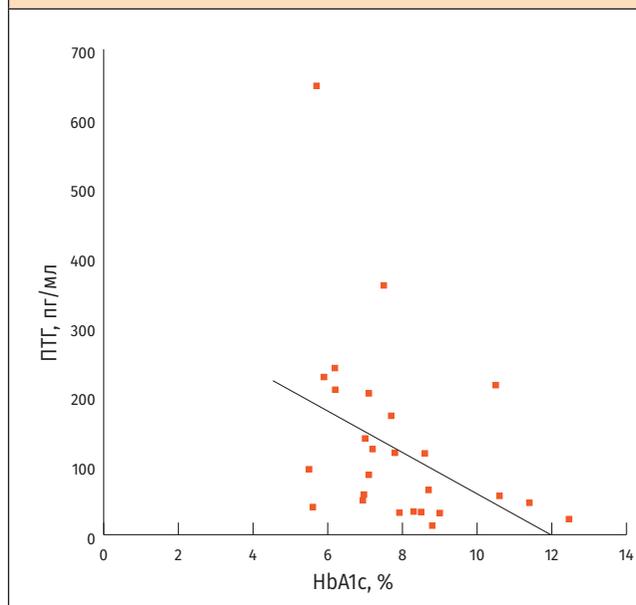


РИСУНОК 3 Графическое представление взаимоотношения уровней (по каждому случаю) паратгормона и гликозилированного гемоглобина в кластере Max-5 1k (n = 26)



нарушение атриовентрикулярной проводимости, удлинение интервала QT, экстрасистолию, пароксизмальную тахикардию, создавая высокий риск желудочковой тахикардии и внезапной смерти. Развивающаяся вследствие микроангиопатии ишемия почки сопровождается повышением синтеза ренина юкстагломерлярным аппаратом почки и активацией ренин-ангиотензин-альдостероновой системы (РААС), что приводит к развитию артериальной гипертензии. Сосуды поражаются атеросклерозом и кальцифицируются [5, 6].

Кластер Max-5 1k

Сердечно-легочный, реактивно-парадоксальный (с реакцией паратгормона)

Первый кластер отличается самым длительным пребыванием пациентов в стационаре, более высоким средним возрастом, парадоксальным соотношением уровней HbA1c (самый низкий в сравнении с другими кластерами, но выше нормы) и паратгормона (самый высокий в сравнении с другими кластерами, выше нормы). Для этого кластера выявлена корреляционная связь (рис. 3), имеющая среднюю

силу и отрицательную направленность: линейный коэффициент корреляции $r = -0,4$ (корреляционный анализ по Пирсону). Значение корреляции и линейной регрессии статистически недостоверно, коэффициент детерминации R^2 имеет низкое значение (16%). Но все же на данную корреляцию стоит обратить внимание как на тенденцию, с учетом того, что анализируется не специально сконструированная по каким-то признакам группа пациентов, а разнородный, без фильтров, поток пациентов с разными коморбидными сочетаниями болезней и сопутствующими заболеваниями. Для первого кластера можно отметить следующие блоки: *рентгенография легких*; эхокардиография, *суточное мониторирование ЭКГ и АД*, тредмил-тест; УЗИ брахиоцефальных артерий (триплексное сканирование); операции на ЖКТ и диагностические обследования ЖКТ (рентгенография, компьютерная томография, магнитно-резонансная томография, сцинтиграфия, эзофагогастродуоденоскопия, биопсия); обследование почек и *почечной артерии* (УЗИ,

компьютерная томография); рентгенография и сцинтиграфия костей скелета (череп, позвоночник, конечности). В этот кластер также попали УЗИ щитовидной железы, компьютерная томография шеи, трепанобиопсия подвздошной кости, электроэнцефалография, аудиометрия. На инфографике отмечены основные кластерообразующие характеристики: *рентгенография легких, суточное мониторирование ЭКГ и АД, УЗИ почечной артерии*. Цвет маркировки кластера – голубой (см. в конце статьи).

Кластер Max-5 2k

Тиреоидный и высокого риска рака легкого

Второй кластер характеризуется более высоким средним уровнем HbA1c и менее высоким средним уровнем паратгормона. Кластер отличается тем, что среди случаев онкозаболеваний *преобладает рак легкого и бронхов* (табл. 4). Основные блоки обследований: *УЗИ и биопсия щитовидной и паращитовидных желез; спирометрия*; операции на легких; эхокардиография и суточное мониторирование ЭКГ; УЗИ сосудов (триплексное

ТАБЛИЦА 4 Распределение онкологических заболеваний (коды по МКБ-10) по кластерам

Кластер	Случаи рака (n = 25)
Мах-5 1к n = 2	C34.0 – Бронхов (1). C64.0 – Почки (1)
Мах-5 2к n = 7	C34.8 – Бронхов или легкого (3). C34.9 – Бронхов или легкого (1). C78.0 – Легкого (1). C78.2 – Плевры (1). C50.2 – Молочной железы (1)
Мах-5 3к n = 12	C34.1 – Бронхов или легкого (1). C34.3 – Бронхов или легкого (1). C34.8 – Бронхов или легкого (1). C44.0 – Губы (1). C32.0 – Гортани (1). C73.0 – Щитовидной железы (1). C50.8 – Молочной железы (1). C53.8 – Шейки матки (1). C16.9 – Желудка (1). C20.0 – Прямой кишки (1). C66.0 – Мочеточника (1). C67.0 – Мочевого пузыря (1)
Мах-5 4к n = 4	C78.0 – Легкого (1). C25.0 – Поджелудочной железы (1). C22.0 – Печени (1). C20.0 – Прямой кишки (1)
Мах-5 5к n = 0	Нет

сканирование); УЗИ и биопсия почек; диагностическое обследование ЖКТ с биопсией. В кластер попали мастэктомия, скintiграфия парасци-товидных желез, аортокоронарное шунтирование, аортобедренное шунтирование, стентирование сонной артерии, ангиография сосудов головного мозга, рентгенография суставов и конечностей, компьютерная томография глаз, физиотерапия.

На инфографике отмечены основные кластерообразующие характеристики: *спирометрия (функция внешнего дыхания), УЗИ и биопсия почек, УЗИ и биопсия щитовидной и парасци-товидных желез*. Цвет маркировки кластера – красный.

Кластер Мах-5 3к

Хирургический, гемодиализный, терминальный по осложнениям

На примере кластера Мах-5 3к можно проиллюстрировать, какую обманчивую картину дает учет только долевого распределения медуслуг в кластере. По этому показателю кластер выглядит как пустой, в нем есть лишь заметное совпадение по гемодиализу. Но при анализе абсолютного распределения медуслуг в кластере открылась совершенно иная картина – кластер оказался насыщенным по проведенным разнообразным операциям и терминальным по разным осложнениям. Основные блоки, характерные для этого кластера: операции на щитовидной,

паращитовидной, молочной железах; офтальмологические операции; *транслюминальная баллонная ангиопластика со стентированием* разных артерий, тромбэндартерэктомии, флебэктомии; ангиография; операции на ЖКТ; удаление и пересадка почки; формирование артериовенозной фистулы и *гемодиализ, перитонеальный диализ; магнитно-резонансная томография головного мозга; рентгенография суставов и костей конечностей*. В этот кластер сгруппировались операция на коже и удаление лимфоузлов, магнитно-резонансная томография суставов и эндопротезирование тазобедренного сустава, санация гнойно-некротического очага стопы.

Третий кластер характеризуется наибольшей встречаемостью комы (отметка в строке диагноза E10.0, СД 1 с комой) и онкологических заболеваний разной локализации (n = 12, табл. 4). Для кластера выявлен самый высокий средний уровень HbA1c $9,42 \pm 0,48\%$ и высокий средний уровень паратгормона в крови $168,31 \pm 27,46$ пг/мл. Этот кластер самый вместительный по количеству пациентов (44,9%), и у него небольшой средний срок пребывания в стационаре в сравнении с другими кластерами.

На инфографике отмечены основные кластерообразующие характеристики: *гемодиализ, транслюминальная баллонная ангиопластика*

со стентированием коронарных артерий, магнитно-резонансная томография головного мозга. Цвет маркировки кластера – фиолетовый.

Кластер Мах-5 4к

Лечебно-физкультурный

Четвертый кластер отличается тем, что в него собрались все пациенты, которым были назначены курсы *лечебной физкультуры*. У кластера самый низкий средний уровень паратгормона и относительно низкий средний уровень HbA1c. Кластер малочисленный, с более высоким средним возрастом. Основные блоки, характерные для этого кластера: *лечебная физкультура; эхокардиография, коронароангиография, КТ-ангиография, аортокоронарное шунтирование; операции и диагностическое обследование ЖКТ, включая рентгенографию разных отделов ЖКТ и компьютерную томографию органов брюшной полости, эзофагогастродуоденоскопию, колоноскопию, биопсию прямой кишки; пункция плевральной полости, однофотонная эмиссионная компьютерная томография легких, бронхоскопия, рентгенография грудной клетки; УЗИ щитовидной железы; маммография; компьютерная томография органов малого таза; денситометрия*.

На инфографике отмечены основные кластерообразующие характеристики: *лечебная физкультура, КТ-ангиография, компьютерная томография*

органов брюшной полости. Цвет маркировки кластера – желтый.

Кластер Max-5 k

Денситометрический, с плохим гликемическим контролем

Пятый кластер состоит из пациентов, каждый из которых прошел рентгеновскую денситометрию костей скелета. Данный кластер характеризуется относительно низким средним уровнем паратгормона и сравнительно высоким средним уровнем HbA1c. Основные блоки, характерные для этого кластера: УЗИ и биопсия щитовидной железы; УЗИ молочных желез; исследование сердца – эхокардиография, суточное мониторирование ЭКГ, велоэргометрия; эзофагогастродуоденоскопия и Хелпил-тест (определение бактерии *Helicobacter pylori*); УЗИ трансплантированной почки, исследование органов малого таза; УЗИ сосудов (триплексное сканирование); пункция костного мозга; аудиометрия.

На инфографике отмечены основные кластерообразующие характеристики: денситометрия костей, УЗИ молочных желез, УЗИ трансплантированной почки. Цвет маркировки кластера – зеленый.

Синтез данных и построение единого континуума

Кластерный анализ, проведенный на медицинских данных по потокам пациентов, подтвердил свою релевантность, но необходимо понимать, что этот метод показывает обобщенные тенденции, а не маркирует каждого конкретного пациента. Не все пациенты, попавшие в один кластер, абсолютно одинаковы, напротив, внутри кластера есть различия между пациентами. Это отличает кластерный анализ от привычного жесткого статистического анализа по группам. Кластеризация имеет гибкость, которая позволяет достичь нового уровня понимания процессов в едином связанном континууме.

В данном исследовании получено пять кластеров пациентов, для каждого кластера определены кластерообразующие характеристики, то есть те

медуслуги, которые наиболее выражено различаются по кластерам (абсолютное и долевое распределение). Суммарно эти характеристики представлены на инфографике. Ценность данного кластерного разбиения заключается в том, что оно основано на коллективном выборе врачей нескольких медучреждений, и даже шире – тех врачей, кто направлял пациентов на госпитализацию из поликлиник. Врачи назначали пациентам то или иное обследование или вмешательство, основываясь на своем профессиональном опыте ведения больных. Именно поэтому не вызывает никаких сомнений, что сложившаяся кластерная картина имеет глубокие патогенетические корни.

Для того чтобы попытаться понять, какие закономерности могут лежать в основе получившегося кластерного разбиения, были проанализированы открытые для пользователей в Интернете генетические базы данных для некоторых ключевых с точки зрения патогенеза СД 1 молекулярных комплексов или маркеров сигнальных путей (табл. 5). Этот уровень изучения патогенеза заболеваний находится сегодня еще на стадии эксперимента, так как многие сигнальные пути остаются неизученными и неясными.

СД 1 как аутоиммунное заболевание связывают с генетической предрасположенностью – полиморфизмом антигенов II класса HLA-DRB1 (а также DQB1 и DQA1) [4, 11]. Происходит аутоиммунная деструкция Т-лимфоцитами β -клеток островков Лангерганса поджелудочной железы. Экспрессия гена в антигенпредставляющих клетках разных тканей и органов различается. Основные органы, интересные в аспекте кластерообразующих целей инструментального обследования пациентов, представлены в таблице 5. В легких экспрессия гена HLA-DRB1 выражена более всего, орган насыщен как РНК, так и самим белком. Антиген HLA-DRB1 связан с риском возникновения *рака легкого*, указание на это есть в генетической базе данных The Human Protein Atlas (www.

proteinatlas.org) и в ряде научных работ [12–14]. В нашей первой статье по исследованию данной когорты больных было показано, что легкие при СД 1 необходимо считать непосредственным органом-мишенью, с высоким риском развития онкологического заболевания [15].

Субстраты инсулиновых рецепторов IRS1 и IRS2 активируют сигнальные пути при взаимодействии рецептора с лигандами – инсулином и инсулиноподобным фактором роста (IGF-I, IGF-II). IRS1 и IRS2 могут участвовать в канцерогенезе (росте, миграции, трансформации опухоли) [16]. Уменьшение числа IRS1 и IRS2 считают одним из базовых механизмов развития сердечной недостаточности [17]. Редукции IRS1 и IRS2 способствует избыточная стимуляция инсулином и паратгормоном. Увеличение концентрации ПТГ приводит к развитию инсулинорезистентности в тканях и повышению уровня глюкозы в крови. При этом ПТГ угнетает сигнальный путь инсулинового рецептора посредством цАМФ и способствует редукции экспрессии IRS1 и транспортера GLUT4, который переносит глюкозу внутрь клетки при активации инсулинового рецептора [18–21]. В свою очередь, нарастание концентрации глюкозы в крови подавляет активность паратириоцитов [10].

ПТГ взаимодействует с рецептором PTH1R, с этим же рецептором связывается и ПТГ-родственный белок (Parathyroid Hormone-Related Protein, PTHrP). Этот молекулярный комплекс относится к паракринным гормонам. Обнаружена защитная роль PTHrP в легочной ткани, при его синтезе в пневмоцитах II типа (в гранулярных клетках плоского эпителия, выстилающего стенки легочных альвеол) происходит восстановление поврежденных альвеол [22]. Низкая концентрация PTHrP приводит к худшим исходам заболеваний легких. Сигнальный путь активации рецептора PTH1R посредством PTHrP останавливает пролиферацию раковых клеток в легких [23, 24]. Помимо этого, PTHrP имеет еще ряд положительных плейотропных эффектов, включая

таблица 5 Различия выраженности экспрессии генов по органам и тканям: субстратов инсулиновых рецепторов (IRS 1, 2), антигена гистосовместимости (HLA-DRB1), рецептора к ПТГ (PTH1R) и ПТГ-родственного белка (PTHrP) (сводная таблица по открытым генетическим базам данных)

Орган/ткани	IRS1 RPKM	IRS2 RPKM	HLA-DRB1 РНК, TPM	HLA-DRB1 белок	PTH1R TPM	PTHrP 100xFPKM
Головной мозг	<5	>10	<50	–	>10	>10
Щитовидная железа	>15	>5	<100	–	<10	>10
Сердце	<5	>10	<50	–	<10	>10
Коронарные артерии	–	–	–	–	>20	>10
Легкие	<5	<5	>800	пневмоциты, макрофаги	>30	>10
Молочные железы	<5	>10	<50	–	>10	>10
Почки	<5	<5	<100	–	>200	<10
Печень	<5	>5	<50	–	>20	<10
Желчный пузырь	<5	<5	>100	–	–	>10
Поджелудочная железа	<5	<5	>10	–	<10	<10
Пищевод	<5	<5	<100	–	>20	>10
Желудок	<5	<5	>200	–	<10	<10
Тонкая кишка	<5	<5	>200	–	<10	<10
Толстая кишка	<5	>5	>100	–	<10	>10
Жировая ткань	>5	>20	<100	–	>20	<10

Сокращения и источники (международные генетические базы данных):

IRS1, IRS2 – Insulin Receptor Substrate 1, 2
<https://www.ncbi.nlm.nih.gov/gene/3667>
<https://www.ncbi.nlm.nih.gov/gene/8660>
 HLA-DRB1 – Human Leukocyte Antigens class II histocompatibility antigen, D Related beta chain
<https://www.proteinatlas.org/ENSG00000196126-HLA-DRB1/tissue>
 PTH1R – Parathyroid Hormone Receptor 1
<https://gtexportal.org/home/gene/PTH1R>
 PTHrP – Parathyroid hormone-related protein /Parathyroid Hormone Like Hormone, PTHLH
<http://www.genecards.org/cgi-bin/carddisp.pl?gene=PTH1R>

Сокращения единиц измерения:

RPKM – Reads Per Kilobase of transcript, per Million mapped reads
 TPM – Transcripts Per Million
 FPKM – Fragments PerKilobase Million

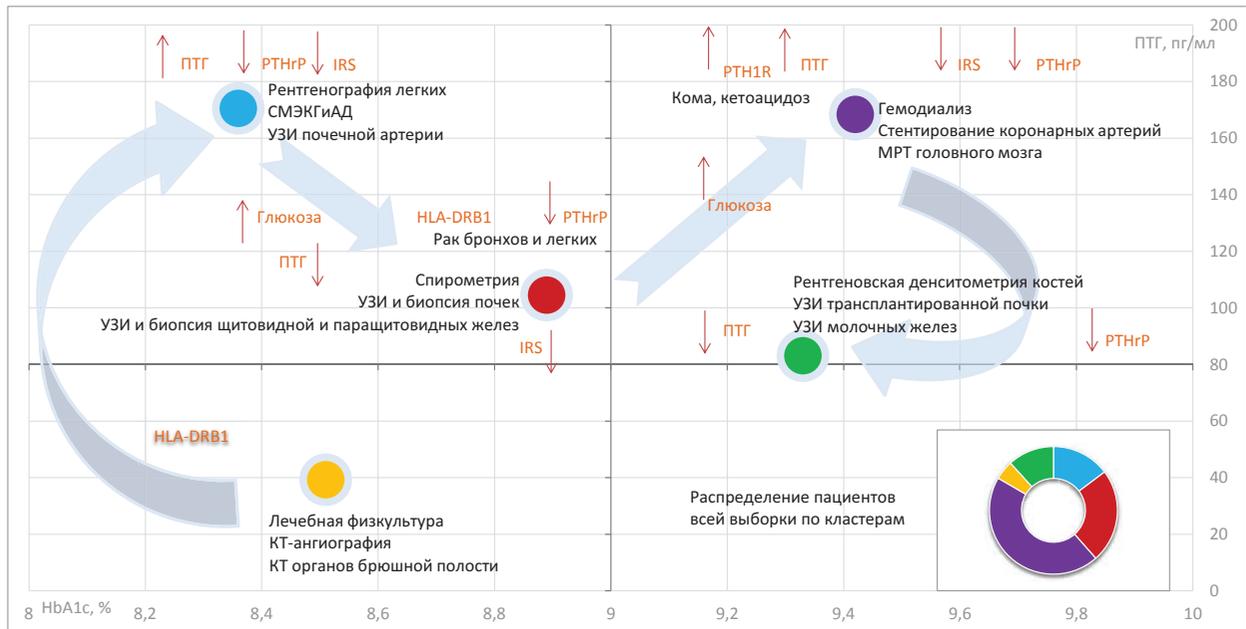
сосудорасширяющий эффект, увеличение почечного кровотока и скорости клубочковой фильтрации через сигнальный путь рецептора PTH1R, но предполагают и наличие еще не открытых дополнительных рецепторов для PTHrP [25]. Следует отметить, что есть указания на более длительное, предпочтительное и

стабильное связывание ПТГ с рецептором PTH1R в сравнении с PTHrP [26, 27]. Можно предположить, что в условиях повышенной концентрации ПТГ возникает конкуренция за рецептор. Но – этот вопрос до конца не изучен – два гормона связываются с разными доменами рецептора PTH1R.

На механизм снижения эффекта PTHrP можно взглянуть и со стороны регуляции самого рецептора. Описана резистентность к ПТГ на фоне нарастания тяжести хронической почечной недостаточности, что объясняется снижением экспрессии гена рецептора PTH1R в условиях уремии [28, 29]. Но есть и обратные указания на то, что метаболический ацидоз (уремический ацидоз, диабетический кетоацидоз) способствует повышению экспрессии гена рецептора PTH1R и большему связыванию ПТГ с этим рецептором [30]. Таким образом, описаны разные варианты влияния на экспрессию гена PTH1R у пациентов с СД 1: снижение или повышение экспрессии. Но следует подчеркнуть, что обе эти ситуации могут редуцировать сигнальный путь для PTHrP – либо напрямую при снижении экспрессии рецептора, либо опосредованно через конкуренцию при повышении экспрессии рецептора с усилением его связывания с ПТГ.

Секреция ПТГ при СД усиливается в ответ на гипокальциемию. Этот гормон способствует разрушению костного матрикса, влияя на остеобласты и остеокласты. ПТГ также опосредованно увеличивает канальцевую реабсорбцию и кишечную абсорбцию Ca²⁺ [4, 31]. Восполнение за счет резорбции костей концентрации Ca²⁺ в крови – это не единственный механизм, разрушающий кости при СД 1. Отсутствие эндогенного инсулина и его недостаток на фоне инсулинотерапии, а также инсулинорезистентность (в том числе из-за ПТГ) угнетают дифференцировку и пролиферацию остеобластов, на которые в норме инсулин оказывает анаболическое действие [32]. Гипергликемия подавляет дифференцировку остеобластов и способствует гликированию белкового матрикса костной ткани с образованием конечных продуктов гликирования (Advanced glycation end-products, AGEs), а кетоацидоз активирует остеокласты [4, 31–33]. Также развивается окислительный стресс в костной ткани, происходит активация воспалительных

РИСУНОК 4 Континуум патогенетического каскада при СД 1, построенный на основе кластерного анализа потока пациентов по данным МИС



цитокинов, нарушение микроциркуляции [33, 34]. PTHrP участвует в физиологическом процессе ремоделирования костной ткани [35]. Среди множества сигнальных путей, задействованных в регуляции остеобластов и остеокластов, особо выделяют RANKL (Receptor activator of nuclear factor kappa B ligand), цитокин семейства факторов некроза опухоли [27, 31]. RANKL участвует в дифференцировке и активации остеокластов при стимулирующем влиянии ПТГ [27, 36]. Изменения в костях при СД называют «диабетопороз», при нем происходят патологические изменения матрикса и снижение минеральной плотности, что приводит к переломам [37]. Еще один фактор, усугубляющий поражение костей, – хроническая почечная недостаточность. Почечные остеодистрофии, особенно адинамическая костная болезнь (редуцированный клеточный состав, снижение интенсивности ремоделирования кости, низкий уровень ПТГ), осложняют состояние больных, находящихся на гемодиализе, а также перенесших трансплантацию

почки (посттрансплантационная остеопатия) [4, 37]. На *рисунке 4* представлен созданный на основе анализа кластеров континуум патогенетических изменений при СД 1. Номера кластеров специально не указаны, чтобы не нарушать восприятие выстроенной цепочки. Континуум начинается с кластера, маркированного желтым цветом, указан антиген HLA-DRB1 как один из базовых вовлеченных генетических факторов заболевания. Поражение почек на фоне гипергликемии приводит к возникновению артериальной гипертензии, повышению уровня ПТГ, который, в свою очередь, вызывает инсулинорезистентность. Нарастание почечной недостаточности и высокий уровень ПТГ приводят к подавлению эффектов PTHrP. Этот этап отражен в кластере, маркированном голубым цветом. Далее уровень глюкозы нарастает, и функция паратиреоцитов угнетается. Кластер, маркированный красным цветом, показывает этап, когда влияние онкогенных факторов, включая HLA-DRB1, в условиях подавления протективного эффекта

PTHrP запускает онкологический процесс – рак легкого. Следующий этап – дальнейшее усугубление почечной недостаточности, развитие кетоацидоза и комы, повышение на фоне метаболического ацидоза ПТГ и экспрессии рецептора PTH1R. Пациенты переходят на гемодиализ, у них повышается риск атеросклеротического поражения артерий, требующего оперативного вмешательства, что было показано в предыдущей статье по данному исследованию [15]. Этот кластер маркирован фиолетовым цветом. И завершающий этап континуума – трансплантация почки и выраженность костных изменений. Кластер маркирован зеленым цветом. Для удобства на рисунке в правом нижнем углу показано процентное распределение пациентов по кластерам, чтобы было понятно, на каком этапе континуума находится больше всего пациентов. Как видно из представленных данных на *рисунке 4*, кластерообразующие медуслуги коснулись тех органов, которые связаны с выраженной экспрессией выбранных молекулярных маркеров

сигнальных путей (табл. 5). Легкие, головной мозг, сердце, молочные железы, щитовидная железа, органы брюшной полости связаны с выбранными маркерами, в них происходят патологические изменения параллельно с изменениями экспрессии этих маркеров.

ЗАКЛЮЧЕНИЕ

Таким образом, кластерный анализ, произведенный в рамках концепции Больших данных, позволяет выстроить континуум, показывающий этапы развития патогенетического процесса. Совпадение кластерообразующих целей инструментального

обследования пациентов с органами, которые при СД 1 наиболее подвержены повреждению на уровне молекулярных сигнальных путей, свидетельствует о высоком уровне клинического мышления врачей, задействованных в ведении пациентов.



ИСТОЧНИКИ

- Колесниченко О.Ю., Мазелис А.Л., Николаев А.Э., Мартынов А.В., Пулит В.В., Смородин Г.Н., Колесниченко Ю.Ю. Intranet-пациент в медицинских информационных системах. Социология медицины, 2017, 16(1): 49-55.
- Мартынов А.В., Пулит В.В., Смородин Г.Н., Колесниченко О.Ю., Мазелис Л.С., Мазелис А.Л., Колесниченко Ю.Ю., Минушкина Л.О., Баландин С.И. Актуальность и опыт освоения новой области Больших данных МИС. Труды ежегодного международного конгресса «Информационные технологии в медицине», 2017. Москва. Web: <http://itmcongress.ru/>.
- Мазелис Л.С., Мазелис А.Л., Смородин Г.Н., Колесниченко О.Ю., Мартынов А.В., Пулит В.В., Колесниченко Ю.Ю., Баландин С.И. Моделирование стационарного ведения пациентов на основе кластерного анализа. Труды ежегодного международного конгресса «Информационные технологии в медицине», 2017. Москва. Web: <http://itmcongress.ru/>.
- Эндокринология. Национальное руководство. Краткое издание. Под ред. И.И. Дедова, Г.А. Мельниченко. М.: ГЭОТАР-Медиа, 2013. 752 с.
- Серов В.В., Пальцев М.А. Почки и артериальная гипертензия. М.: Медицина, 1993. 256 с.
- Хантакова Е.А., Хамнуева Л.Ю., Орлова Г.М. Анализ фосфорно-кальциевого обмена и функции паращитовидных желез у больных сахарным диабетом 1 типа. Сахарный диабет, 2013, 4: 33-37.
- Leidig-Bruckner G, Ziegler R. Diabetes mellitus a risk for osteoporosis? *Experimental and Clinical Endocrinology & Diabetes*, 2001, 109 (2):S493-514.
- Heidbreder E, Götz R, Schafferhans K, Heidland A. Diminished Parathyroid Gland Responsiveness to Hypocalcemia in Diabetic Patients with Uremia. *Nephron*, 1986, 42: 285-289.
- McNair P, Christensen MS, Madsbad S, Christiansen C, Transbøl I. Hypoparathyroidism in diabetes mellitus. *Acta Endocrinologica (Copenhagen)*, 1981, 96 (1): 81-86.
- Sugimoto T, Ritter C, Morrissey J, Hayes C, Slatopolsky E. Effects of high concentrations of glucose on PTH secretion in parathyroid cells. *Kidney International*, 1990, 37: 1522-1527.
- Arango M-T, Perricone C, Kivity S, Cipriano E, Ceccarelli F, Valesini G, Shoenfeld Y. HLA-DRB1 the notorious gene in the mosaic of autoimmunity. *Immunology Research*, 2017, 65: 82-98.
- Davies KD, Ng TL, Estrada-Bernal A, Le AT, Ennever PR, Camidge DR, Doebele RC, Aisner DL. Dramatic Response to Crizotinib in a Patient with Lung Cancer. Positive for an HLA-DRB1-MET Gene Fusion. *JCO Precision Oncology*, 2017, 1: 1-6.
- Nacht M, Dracheva T, Gao Y, Fujii T, Chen Y, Player A, Akmaev V, Cook B, Dufault M, Zhang M, Zhang W, Guo M, Curran J, Han S, Sidransky D, Bucetow K, Madden SL, Jen J. Molecular characteristics of non-small cell lung cancer. *Proceedings of the National Academy of Sciences of the USA*, 2001, 98 (26): 15203-15208.
- Yang L, Wang LJ, Shi GL, Ni L, Song CX, Zhang ZX, Xu SF. Analysis of HLA-A, HLA-B and HLA-DRB1 alleles in Chinese patients with lung cancer. *Genetics and Molecular Research*, 2010, 9 (2): 750-755.
- Колесниченко О.Ю., Колесниченко Ю.Ю., Минушкина Л.О., Смородин Г.Н., Мартынов А.В., Пулит В.В., Долженков А.Н. Возможности применения бинарного кода Грея для аналитики Больших данных МИС: пациенты с СД 1-го типа. Ремедиум, 2017, 10: 38-47.
- Hakuno F, Fukushima T, Yoneyama Y, Kamei H, Ozoe A, Yoshihara H, Yamanaka D, Shibano T, Sone-Yonezawa M, Yu B-C, Chida K, Takahashi S-I. The novel functions of high-molecular-mass complexes containing insulin receptor substrates in mediation and modulation of insulin-like activities: emerging concept of diverse functions by IRS-associated proteins. *Frontiers in Endocrinology*, 2015, 6: 73.
- Qi Y, Xu Z, Zhu Q, Thomas C, Kumar R, Feng H, Dostal DE, White MF, Baker KM, Guo S. Myocardial Loss of IRS1 and IRS2 Causes Heart Failure and Is Controlled by p38 MAPK During Insulin Resistance. *Diabetes*, 2013, 62 (11): 3887-3900.
- Rahimi Z. Parathyroid hormone, glucose metabolism and diabetes mellitus. *Journal of Parathyroid Disease*, 2014, 2 (1): 55-56.
- Chang E, Donkin SS, Teegarden D. Parathyroid Hormone Suppresses Insulin Signaling in Adipocytes. *Molecular and Cellular Endocrinology*, 2009, 307(1-2): 77-82.
- Cheung PS, Thompson NW, Brothers TE, Vinik AI. Effect of hyperparathyroidism on the control of diabetes mellitus. *Surgery*, 1986, 100 (6): 1039-1047.
- Chiu KC, Chuang L-M, Lee NP, Ryu JM, McGullam JL, Tsai GP, Saad MF. Insulin sensitivity is inversely correlated with plasma intact parathyroid hormone level. *Metabolism: Clinical and Experimental*, 2000, 49(11): 1501-1505.
- Hastings RH, Defetos LJ. Parathyroid Hormone-Related Protein and Lung Injury. *Chest J*, 2002, 122(4): 1494-1495.
- Montgrain PR, Phun J, Werff RV, Quintana RA, Davani AJ, Hastings RH. Parathyroid-hormone-related protein signaling mechanisms in lung carcinoma growth inhibition. *Springer Plus*, 2015, 4: 268.
- Hastings RH, Montgrain PR, Quintana R, Rascon Y, Defetos LJ, Healy E. Cell cycle actions of parathyroid hormone-related protein in non-small cell lung carcinoma. *The American Journal of Physiology - Lung Cellular and Molecular Physiology*, 2009, 297 (4): L578-L585.
- Clemens TL, Cormier S, Eichinger A, Endlich K, Fiaschi-Taesch N, Fischer E, Friedman PA, Karaplis AC, Massfelder T, Rossert J, Schlüter K-D, Silve C, Stewart AF, Takane K, Helwig J-J. Parathyroid hormone-related protein and its receptors: nuclear functions and roles in the renal and cardiovascular systems, the placental trophoblasts and the pancreatic islets. *British Journal of Pharmacology*, 2001, 134(6): 1113-1136.
- Suzuki N. Parathyroid Hormone-Related Protein. Subchapter 26B. *Handbook of Hormones: Comparative Endocrinology for Basic and Clinical Research*. Ed. by Y. Takei, H. Ando, K. Tsutsui. Oxford: Academic Press, 2015: 227-229.
- Datta NS, Abou-Samra AB. PTH and PTHrP Signaling in Osteoblasts. *Cellular Signalling*, 2009, 21 (8): 1245-1254.
- Picton ML, Moore PR, Mawer EB, Houghton D, Freemont AJ, Hutchison AJ, Gokal R, Hoyland JA. Down-regulation of human osteoblast PTH/PTHrP receptor mRNA in end-stage renal failure. *Kidney International*, 2000, 58 (4): 1440-1449.
- Ureña P, Mannstadt M, Hruby M, Ferreira A, Segre GV, Drücke T. Down-regulation of the PTH/PTHrP receptor in uremia. *Journal of Bone and Mineral Metabolism*, 1994, 12(1): S87-S90.
- Disthabanchong S, Martin KJ, McConkey CL, Gonzalez EA. Metabolic acidosis up-regulates PTH/PTHrP receptors in UMR 106-01 osteoblast-like cells. *Kidney International*, 2002, 62 (4): 1171-1177.
- Мкртумян А.М. Остеопороз – незамеченная болезнь при сахарном диабете. Открытая лекция. Пленарное заседание Московского городского научного общества терапевтов. Первый МГМУ им. И.М. Сеченова МЗ РФ, 2018.
- Бабенко А.Ю., Никитин В.С., Каронова Т.Л. Диабет и кость. *Медицинский совет*, 2015, 17: 108-113.
- Napoli N, Chandran M, Pierroz DD, Abrahamson B, Schwartz AV, Ferrari SL, International Osteoporosis Foundation (IOF) Bone and Diabetes Working Group. Mechanisms of diabetes mellitus-induced bone fragility. *Nature Reviews Endocrinology*, 2017, 13 (4): 208-219.
- Ferrari S. Diabetes and Bone. *Calcified Tissue International*, 2017, 100: 107-108.
- Dean T, Vilardaga J-P, Potts-JrJT, Gardella TJ. Altered Selectivity of Parathyroid Hormone (PTH) and PTH-Related Protein (PTHrP) for Distinct Conformations of the PTH/PTHrP Receptor. *Molecular Endocrinology*, 2008, 22 (1): 156-166.
- Epstein S, Defeudis G, Manfrini S, Napoli N, Pozzilli P, Scientific Committee of the First International Symposium on Diabetes and Bone. Diabetes and disordered bone metabolism (diabetic osteodystrophy): time for recognition. *Osteoporosis International*, 2016, 27 (6): 1931-1951.
- Рекина И.В., Родионова С.С., Колондаев А.Ф., Морозов А.К. Профилактика потери костной ткани у больных с трансплантацией почки. *Пособие для врачей*. М., 1999.