

РАЗРАБОТКА ЭКСПЕРТНОЙ СИСТЕМЫ ДЛЯ ОПРЕДЕЛЕНИЯ ПРЕДИКТИВНЫХ РИСКОВ ЗАБОЛЕВАНИЙ

Н.Э. Посвалюк, С.А. Погорелов
Вычислительный центр ДВО РАН,
ул. Ким Ю Чена 65, г. Хабаровск, 680000,
e-mail: admvc@ccfebras.ru

Современные экспертные системы находят применение в научных и исследовательских задачах биологии и медицины. В данной работе приводятся результаты исследований по подбору данных для проектирования экспертной системы, которая позволит выделять группы риска на основе данных молекулярно-генетического тестирования населения, проживающего в экологических условиях Дальневосточного региона, с целью оценки рисков развития социально значимых заболеваний.

Ключевые слова: экспертная система, математическое моделирование, прогноз, генетические полиморфизмы, эпигенетические факторы.

В настоящее время экспертные системы (ЭС) и системы поддержки принятия решений находят свое применение в различных областях экономики, медицины и науки. Экспертные системы разрабатываются на основе математических методов, эмпирических данных предметной области и результатов научных исследований специалиста предметной сферы. В зависимости от характера решаемых задач экспертные системы делятся на широконаправленные и узконаправленные.

В контексте исследования основной функцией ЭС является получение результатов, не уступающих по качеству и эффективности решениям, которые получает человек – эксперт-генетик. Еще одна значимая функция ЭС – облегчить жизнь человеку при решении задач, связанных с нарабаткой опыта [1]. ЭС в медицине представляют отдельный класс задач. В качестве примера можно привести ЭС MOLGEN для специалистов по клонированию генов в молекулярной генетике [2], хотя обычно, несмотря на все попытки экспертных систем работать на уровне среднего специалиста в своей области, они не могут заменить действующих профессиональных экспертов. С другой стороны, ЭС оказываются очень полезными в обучении начинающих специалистов [3, 4].

Отметим, что не удалось найти в литературе описание ЭС для изучения результатов генетического тестирования полиморфизмов человека. Попытки автоматизации расшифровки предпринимаются практикующими генетиками.

Настоящее исследование затрагивает ряд проблем: необходимость разрабатывать алгорит-

мы для получения требуемых эффективных решений для ЭС и подготовку данных для разработки экспертной системы, предназначенной для оценки риска развития социально значимых заболеваний у населения на базе молекулярно-генетического исследования с учетом эпигенетических влияний экологических факторов.

Проводимое исследование носит междисциплинарный характер и затрагивает предметные области таких областей знаний, как математика, информационные науки, профилактическая медицина, экология, молекулярная генетика.

Современная медицина, получив беспрецедентные диагностические возможности из ресурсов бурно развивающихся биологических и информационных наук, выводит врачебную деятельность на принципиально новые позиции, становится персонифицированной и предсказательной, основанной на уникальной генетической предрасположенности человека к реагированию на различные внешние и внутренние воздействия.

Генетическая предикция определяет и индивидуальные особенности протекания метаболических, нейродинамических и иммунопатологических процессов в организме человека. Накопленная в научной литературе информация о молекулярно-генетических характеристиках, определяющих подверженность или устойчивость организма человека к воздействию экологических факторов, питания, образа жизни и уровня стресса, способствует формированию у специалистов нового вектора клинического мышления, отражающего системный подход к оценке здоровья

человека. Определенные климато-географические, экологические и социально-экономические условия территории постоянного проживания человека, характеристики индивидуальной зоны его жизнедеятельности («экологической ниши индивидуума») моделируют на эпигенетическом уровне генетически детерминированные процессы, нивелируя или проявляя риски развития заболеваний.

С учетом новых знаний, системного подхода к оценке болезненности населения территории уже недостаточно статистических данных натуральных наблюдений для прогноза здоровья населения территории. Разработка новых методов оценки рисков развития социально значимых заболеваний на основе изучения генетических полиморфизмов человека с учетом эпигенетических влияний окружающей среды позволит осуществлять не только индивидуальные, но и популяционные прогнозы для территории постоянного проживания населения.

В условиях активной внутренней и внешней миграции, межэтнических браков ожидается трансформация адаптационных процессов, появление известных заболеваний с новыми характеристиками, встреча с ранее неизвестными заболеваниями и реакциями организма на воздействия региональных биотических факторов, изменение характера питания и образа жизни [5].

Для разработки ЭС авторами предлагается в качестве социально значимого заболевания рассмотреть сахарный диабет 2 типа (СД), распространенный повсеместно в различных этнических группах. Заболевание в последние десятилетия приобретает эпидемический характер, становится серьезной социально-экономической проблемой для большинства развитых стран. Прогнозируется к 2030 году увеличение числа заболевших до 552 миллионов человек, что на 140 миллионов увеличит мировую когорту больных СД. Рост заболеваемости и распространенности СД происходит преимущественно за счет СД 2 типа (98% от всех случаев заболевания СД) [6].

Предполагается, что число не диагностированного СД также внушительно. Не менее половины пациентов, имеющих латентный СД, получают лечение в связи с сердечно-сосудистыми заболеваниями, нарушениями со стороны периферической и вегетативной нервной системы, иммунной системы и не знают о наличии у них диабета, не включены в статистику. Этот курьез связан не только с дефектами оказания медико-профилактической помощи населению, с недостаточной

медицинской активностью пациентов, но и с индивидуальными клиническими характеристиками заболевания, определяющимися генами предрасположенности.

Высокая смертность среди больных СД трудоспособного возраста актуализирует проблему своевременного выявления рисков заболевания и его осложнений. Доступные современные технологии молекулярно-генетического тестирования рисков СД и развития его осложнений еще не вошли в клинико-диагностические стандарты, что не позволяет врачам государственных медицинских учреждений использовать их на практике. Соответственно диагностика еще неопределенно долгое время будет осуществляться по обращаемости пациента к врачам с жалобами или болезненным состоянием, то есть не в состоянии предболезни, риска, а уже при наличии клинических проявлений расстройства здоровья.

Внедрением в медицинскую практику молекулярно-генетического тестирования риска развития СД 2 типа и его осложнений занимаются преимущественно врачи-исследователи, а также частнопрактикующие врачи, мотивированные на применение альтернативных и инновационных технологий в области медицины, привлечение пациентов, достижение положительного медико-экономического результата. Но и среди специалистов, готовых внедрять молекулярно-генетические технологии, получение сведений о наборе определенных генетических полиморфизмов – маркеров заболевания порой не приводит к ожидаемому эффекту в силу того, что врачи не готовы интегрировать представленные результаты молекулярно-генетического тестирования в классический клинико-диагностический комплекс и получать целевую информацию нового качества. Врачи также не готовы мотивировать пациентов из групп риска на молекулярно-генетическое тестирование и проведение профилактических мероприятий в случае выявления рисков.

Генетическая предрасположенность к СД 2 типа часто сочетается с ожирением. Специалистами обнаружено около 20 генов, полиморфизмы в которых являются факторами риска возникновения СД 2 типа. Среди них наиболее часто тестируемыми являются следующие полиморфизмы:

- ABCC8 – рецептор сульфанилмочевины (SUR1) G/T Ala1369Ser;
- SLC22A1 – транспортер органических катионов (OCT1) G/A Gly465Arg;
- TCF7L2 – транскрипционный фактор C53341T;

- TCF7L2 – транскрипционный фактор 7 G103894T Intron3.

Метаболический синдром, связанный с риском развития инсулинорезистентности и СД 2 типа и ожирение, ассоциированное с СД 2 типа, тестируются в панели следующих полиморфизмов нарушений липидного обмена:

- ApoA1 – аполипопротеин A1 G-75A Promoter;
- ApoA5 – аполипопротеин A5 C56G S19W;
- ApoC3 – аполипопротеин C3 T-482C Promoter;
- ApoE – аполипопротеин E E2/E3/E4 Cys112Arg;
- SEPT – переносчик эфира холестерина G1264A Val422Ile;
- LDLR – рецептор липопротеинов низкой плотности G7250T;
- LPL – липопротеинлипаза C1791G Ser474Stop;
- FTO – 2-оксоглутаратзависимая демителаз нуклеиновых кислот g.87653T>A;
- INSIG2 – инсулин-индуцирующий белок G-10026C Promoter;
- PPARC – рецептор активатора пероксисом C34G Pro12Ala;
- ADRB3 – бета-3-адренергический рецептор T191C Trp64Arg;
- GNB3 – гуанин нуклеотид-связывающий протеин бета-3 (G-белок b3) – C825T splicing defect.

Разнообразие генетических маркеров, характерных для различных популяционных групп, указывает на необходимость учитывать значимость этнической составляющей при выявлении наследственной предрасположенности. Комплексное изучение генетических основ СД 2 типа в русской популяции с учетом молекулярно-генетического тестирования многих аллельных вариантов генов-кандидатов, анализом их сочетаний, а также ассоциаций с клиническими и гормонально-метаболическими проявлениями проведено в группе 192 неродственных испытуемых женского (n=120) и мужского (n=72) пола К.А. Вахромеевой (2015) [7].

Для создания доступных, мобильных программных комплексов для поддержки врача и направления диагностического процесса при принятии решения в многокритериальной среде с данными о генетической предрасположенности требуется проведение анализа больших массивов данных генетических исследований, клиничко-лабораторных сведений, экологических факторов зоны жизнедеятельности пациента. Информационная модель пациента с СД с учетом его генети-

ческой предрасположенности и индивидуальных медико-социальных характеристик в регламентированных экологических условиях региона позволит выявлять контингенты риска СД, формировать прогнозы и управлять лечебно-профилактическим процессом.

Авторами собраны данные о 1250 пациентах с СД и родственниках пациентов с СД. У всех пациентов проведены молекулярно-генетические исследования с тестированием по панелям «СД 2 типа», «Метаболический синдром», «Липидный обмен» с дополнительными исследованиями отдельных генетических полиморфизмов по показаниям. Анализ генетических полиморфизмов проводился в лаборатории генной диагностики в Институте химической биологии и фундаментальной медицины СО РАН при участии специалистов МГЦ «Генетика семейного здоровья» г. Хабаровска.

Проведенная работа позволила получить выборку генов предрасположенности к СД 2 типа (PPARG, TCF7L2 (2 мутации), INSIG2, ABCC8, KCNJ11) и генов-предикторов развития метаболического синдрома с инсулинорезистентностью (INSIG2, FTO, ADRB3, GNB3, IL6 (C174G)).

Разработан алгоритм решения задач по прогнозу развития СД 2 типа для экспертной системы на основе математических методов анализа (сопоставления) по данным медицинской диагностики и тестирования генов – предикторов СД с учетом экологических факторов окружающей среды. Ожидается, что ЭС позволит прогнозировать индивидуальные риски развития СД 2 типа и его осложнений, формировать прогнозы течения и исходов заболевания, определять терапевтический выбор и оценивать ожидаемую эффективность персонифицированных лечебно-реабилитационных мероприятий [7, 8].

Отметим, что наборы данных для анализа в экспертной системе с позиций системной генетики включают не только идентификацию и структурный анализ генов – предикторов, но и механизмы эпигенетической регуляции их функции на разных уровнях реализации генетической информации в различных экологических условиях. Как методы анализа рассматриваются два подхода – статистический и нейросетевого моделирования.

На первом этапе исследований экспертами предметной области производился выбор значимых факторов для создания панели эталонных полиморфизмов, ассоциированных с сахарным диабетом 2 типа, инсулинорезистентностью и осложнениями СД. В основу исследования были

положены популяционные данные о генетических полиморфизмах, регулирующих процессы метаболизма жиров и углеводов, полученные в период с 2012 по 2017 гг. в группе дальневосточного населения. Выявлено, что для получения достоверных результатов необходимо изучать генные сети, регулирующие сложные иммунопатологические и метаболические процессы в организме. При этом учитывалось, что один полиморфизм не играет роли в оценке влияния на реализацию различных патологических состояний. Только оценка генных сетей позволяет рассчитать риски развития заболеваний и прогноз исходов.

Экологическая нагрузка на обследованного человека с определенной генетической предрасположенностью к СД рассчитывается на основе балльной оценки из БД «Характеристики зоны жизнедеятельности», в которой он проживает и работает не менее 80% годового времени. Предполагается оценка экологической нагрузки в баллах от 1 до 10; фактически используется метод экспертной (балльной) оценки, основанный на интегрированном показателе по природным средам (вода, воздух, почва). В проведенном исследовании использовались оценки, представленные экспертом-экологом [9].

Предложен алгоритм, использующий метод сопоставления, основанный на сравнении эталонной специализированной панели генов – регуляторов углеводного и липидного обмена с панелью генов, полученной в процессе тестирования конкретного пациента. В настоящее время эта операция выполняется специалистом вручную, что занимает много времени.

Для модельного эксперимента подготовлены результаты генетического тестирования населения ($n = 1250$), проведенного в объеме разработанной специализированной панели генов.

Испытание экспертной модели в тестовом режиме позволило сделать следующие выводы.

Разработанная панель полиморфизмов позволяет с высокой степенью вероятности отнести пациента к контингенту группы риска.

Достаточно большой процент (12–20%) отнесенных к контингенту группы риска не имеют клинических проявлений заболевания.

Экспертная система нуждается в дальнейшем усовершенствовании в разрезе расчета интегрального показателя по природным средам, характеризующего экологию территории, и проведении дальнейших генетических исследований населения с целью оптимизации эталонной панели полиморфизмов.

ЛИТЕРАТУРА:

1. Нейлор К. Как построить свою экспертную систему: пер. с англ. М.: Энергоатомиздат, 1991. 286 с.
2. Бояркина А. К., Ермолаева В. В. Экспертные системы // Молодой ученый. 2016. № 11. С. 286–289. URL: <https://moluch.ru/archive/115/31247/> (дата обращения: 26.11.2018).
3. Пономарева Н.С., Панич А.Е., Машкина Е.В., Рымашевский А.Н., Шкурат Т.П. Экспертная система поддержки принятия решений в оценке риска репродуктивных потерь // Современные проблемы науки и образования. 2015. № 2–1. URL: <http://www.science-education.ru/ru/article/view?id=20941> (дата обращения: 26.11.2018).
4. Бурцева А.Л., Берестнева Е.В., Степаненко Н.П. Создание Базы Знаний для медицинской экспертной системы // Современные наукоемкие технологии. 2016. № 31. С. 14–17.
5. Генетический паспорт – основа индивидуальной и предиктивной медицины / под ред. В.С. Баранова. СПб., 2009. 528 с.
6. Джаратано Дж., Райли Г. Экспертные системы. Принципы разработки и программирование. М.: Вильямс, 2007. 1152 с.
7. Вахромеева К.А. Полиморфизмы генетических маркеров сахарного диабета 2 типа и их ассоциации с клинико-метаболическими показателями в русской популяции: автореф. дисс. ... канд. мед. наук. Тюмень, 2015. 22 с.
8. Любарский Ю.Я. Интеллектуальные информационные системы. М.: Наука, 2013. 228 с.
9. Государственный доклад «О состоянии и использовании минерально-сырьевых ресурсов Российской Федерации в 2016 и 2017 годах». URL: http://www.mnr.gov.ru/docs/o_sostoyanii_i_ispolzovanii_mineralno_syrevykh_resursov_rossiyskoy_federatsii/2017_doklad_o_sostoyanii_i_ispolzovanii_mineralno_syrevykh_resursov_rossiyskoy_federatsii/ (дата обращения: 26.11.2018).

DEVELOPMENT OF AN EXPERT SYSTEM TO DETERMINE THE PREDICTIVE RISKS OF DISEASES

N.Uh. Posvalyuk, S.A. Pogorelov

Modern expert systems are used to solve scientific and research problems in biology and medicine. This paper presents the research results on the selection of data in order to design a special expert system. The system will be helpful in identifying risk groups. It is based on molecular genetic testing data of the population living in the environmental conditions of the Russian Far, in order to assess the risk of development of socially unfavorable diseases.

Keywords: *expert system, mathematical modeling, prediction, genetic polymorphisms, epigenetic factors.*