

Системный анализ в медицине и биологии

Алгоритмы построения портфолио для получения допуска к клиническим испытаниям в области регенеративной медицины*

Р.Е. Суворов, Н.В. Ярыгин, К.В. Попов, И.В. Холоденко, И.В. Смирнов

Аннотация. В работе описываются принципы построения портфолио для получения допуска к клиническим испытаниям в области регенеративной медицины. Предложенные принципы включают методы: автоматизированного сбора информации о (до)клинических исследованиях; сравнительного анализа предлагаемого метода регенеративной медицины с уже имеющимися в клинической практике методами/препаратами; составления обзора по аналогичным методам регенеративной медицины, уже вошедшим в клиническую практику. Разработанные методы обеспечивают информационно-аналитическую поддержку для ускорения перехода от экспериментальных исследований к клинической практике в области регенеративной медицины.

Ключевые слова: интеллектуальный анализ данных, извлечение информации из текстов, сравнительный анализ клинических испытаний.

1. Введение

Регенеративная медицина — сравнительно новое и бурно развивающееся направление. Начало стремительного роста числа публикаций посвященных регенеративной медицине пришлось на конец 1990-х. В основе методов регенеративной медицины лежит концепция стимулирования восстановления тканей и органов *in vivo* или их замены выращенными вне организма живыми эквивалентами. В значительной мере развитие этого направления обусловлено достижениями в области клеточной и молекулярной биологии.

В последнее время методы регенеративной медицины все больше и больше стали применяться в клинической практике. Известно, что для введения того или иного препарата или метода лечения в клиническую практику необходимо провести целый ряд доклинических и клинических исследова-

ний. Зачастую переход от экспериментальных исследований к клинической практике затягивается на многие годы, а иногда даже оказывается невозможным в силу ряда причин. Подобные трудности в частности могут быть связаны с отсутствием возможности собрать максимально полную информацию, касающуюся эффективности определенного метода лечения и/или препарата, полученных в доклинических исследованиях, правильным образом систематизировать эту информацию, объективно оценить все возможные риски.

Настоящая работа направлена на создание новых подходов и методов сбора и анализа информации о клинических и доклинических испытаниях в области регенеративной медицины. В статье описаны разработанные методы оценивания адекватности использованных в доклинических исследованиях животных моделей заболеваний и достаточности полученных доклинических данных, метод сравнительного анализа предлагаемого

* Работа выполнена при поддержке РФФИ, проект №13-07-12156 «офи_м».

метода регенеративной медицины с уже имеющимися в клинической практике методами/препаратами, метод составления обзора по аналогичным методам регенеративной медицины, уже вошедшим в клиническую практику. Разработанные методы обеспечивают информационно-аналитическую поддержку для ускорения перехода от экспериментальных исследований к клинической практике в области регенеративной медицины.

2. Интеллектуальный анализ информации о (до)клинических испытаниях

Разрабатываемые алгоритмы построения портфолио для получения допуска к клиническим испытаниям основаны на интеллектуальном анализе данных, информационном поиске и методологии выполнения мета-анализов биомедицинских исследований.

С точки зрения анализа данных, можно выделить три области, смежные с нашим исследованием:

1. Автоматический анализ текстов по биомедицинской тематике.
2. Автоматический анализ информации об исследованиях.
3. Аналитические системы поиска и исследования информации в сети Интернет в целом.

В рамках первой группы большая часть исследований посвящена извлечению структурированной информации из статей, описаний клинических исследований и т.п. К этой области в первую очередь можно отнести системы cTAKES [1], ExaCT [2], EDGAR [3], MedEx [4]. В основном выполняется извлечение такой информации, как наименования заболеваний, условия, наименования лекарств и лечебных процедур, дозировки, аббревиатуры генов. Подавляющее большинство систем работает только с английским языком. Одна из основных причин – опора на метатезаурус UMLS [5], – полная версия которого существует только для английского языка. В вышеупомянутых системах в основном используются такие методы классификации как SVM, CRF, дополненные вручную составленными правилами в виде лексико-синтаксических шаблонов или регулярных выражений и словарями (UMLS). К этой группе исследований также можно отнести работу [6], посвященную анализу и структурированию естественно-языковых описаний критериев отбора пациентов для клинических испытаний. В [7] для той же задачи применяется подход, основанный на методах машинного обучения с частичным привлечением учителя: полуавтоматически строится словарь тер-

минологии и синтаксических шаблонов, которые затем применяются к текстам для построения семантической сети, описывающей критерии отбора пациентов. Сложность формализации критериев в виде, подходящем для автоматического сравнения также подтверждает анализ, приведенный в [8]. В [9] приводится более гибкий подход, основанный на извлечении ключевых слов и проверки наличия отрицаний при них и последующей индексации с целью поиска по критериям отбора пациентов.

База клинических испытаний ClinicalTrials.gov часто становится объектом исследований, которые можно разделить на две основные группы: анализ базы в целом с целью получения некоторой статистики [10], анализ отдельных документов.

Значительная часть исследований, связанных с ClinicalTrials.gov заключается в предоставлении помощи в поиске информации, в том числе в поиске по критериям отбора пациентов [6]. В [11] предлагается скомбинировать извлечение ассоциативных правил и фильтрацию терминологии по статьям в метатезаурусе UMLS для определения характерных критериев отбора пациентов для различных заболеваний. В [12] обсуждается вопрос использования ClinicalTrials.gov для поиска участников для клинических испытаний. В списке основных заключений статьи приводится необходимость разработки комплексной системы, интегрирующей разные источники и этапы обработки информации, в том числе более глубокий анализ естественного языка.

В целях расширения возможностей автоматического анализа применяются онтологии и RDF-хранилища для приведения информации о клинических испытаниях в единообразную форму [13]. В [14] с помощью методов нечеткого сравнения текстов агрегируются несколько онтологий, элементы которой привязываются к карточкам клинических испытаний ClinicalTrials.gov. Однако выполняемый автоматический анализ недостаточно глубокий для помощи при сравнении исследований. К тому же, отсутствует привязка к научным публикациям, содержащим информацию о соответствующих доклинических испытаниях, что могло бы быть очень полезным для составления портфолио.

В [15] также описывается подход к систематизации и анализу доступной информации о клинических испытаниях с помощью онтологии. При этом поднимается вопрос о фильтрации найденных испытаний по качеству организации эксперимента, значимости результатов и схожести проблемы. Все характеристики, включая параметры сравнения, добавляются в базу знаний вручную. Авторы также подмечают, что предложенная структура

данных неспособна отразить все критерии отбора пациентов.

В [16] поднимается важный вопрос связывания карт клинических испытаний ClinicalTrials.gov и научных публикаций, а также открытости результатов исследования. Авторы заявляют, что лишь 28% карт клинических испытаний содержат ссылки на статьи, в то время как 44% испытаний имеют опубликованные результаты, хотя они не привязаны к картам ClinicalTrials.gov. Отсутствие открытых результатов испытания, находящегося во второй или более старшей фазе, скорее всего, говорит об успешности и переходе работ в коммерческие организации. Этот фактор необходимо учитывать при подготовке нового клинического испытания.

Всесторонний анализ состояния исследований в области автоматизации составления систематических обзоров [17] говорит о том, что отдельные этапы построения систематического обзора частично автоматизированы, но с каждым этапом связаны достаточно значительные проблемы, в том числе отсутствие как программной, так и методологической интеграции между различными инструментами, недостаточная гибкость при сборе и сравнении характеристик исследований. Это подтверждает необходимость разработки единой методологии и реализующих её программных средств, опирающихся на более общие подходы, например, информационный поиск и анализ текстов на естественном языке.

В [18] описывается попытка полуавтоматического преобразования базы ClinicalTrials.gov в структуру, более удобную для анализа с целью извлечения глобальных закономерностей и сравнения. Однако детализация структуры недостаточна для ее применения при составлении систематических обзоров и мета-анализов.

В [19, 20] и аналогичных работах предлагается подойти к проблеме подготовки новых клинических испытаний с другой стороны, а именно отказаться от автоматического анализа существующих баз данных, как правило, содержащих информацию в полуструктурированном виде (в виде текстов, иногда разбитых на отдельные поля), и административно организовать сбор, хранение и поиск нужной информации сразу в структурированном виде.

В [21] представлен подход, аналогичный предлагаемому в настоящей работе. Авторы описывают программные средства, фактически представляющие собой кластеризующую метапоисковую машину: сначала выполняется метапоиск по научным статьям, потом каждый документ классифицируется по типу (клиническое испытание, обзор и т.п.),

затем выполняется тематическая кластеризация и ранжирование по степени важности для включения в систематический обзор. К основным отличиям описываемой системы от разрабатываемой в настоящей работе, которые также являются и недостатками, можно отнести отсутствие интеграции разнородных данных (полуструктурированных карт из ClinicalTrials.gov и научных публикаций); интерактивности и функций извлечения структурированной информации при отборе и сравнении похожих исследований; отсутствие строгой методологии для оценки адекватности статей с точки зрения включения в аналитический обзор. В [22] также рассматривается применение кластеризации для упрощения поиска информации о клинических испытаниях.

В [23] анализируется проблема поиска информации о клинических испытаниях. Авторы приходят к выводу о том, что в общем случае для не специалиста в работе с поисковыми машинами задача составления полной картины является достаточно непростой, потому что ClinicalTrials.gov неполна, так как регистрация в ней проводимых испытаний обязательна только в США, а также не все карты испытаний содержат результаты, хотя опубликованы в статьях в различных изданиях.

В настоящее время в медицинской практике широко используются различного рода опросники, в частности, для сбора анамнеза пациентов, для подтверждения и/или уточнения диагноза. В последнее время все больше разрабатывается опросников, направленных на оценку качества жизни пациентов после какого-либо лечения. Термин «качество жизни» включает в себе такие аспекты существования индивидуума, как физическое состояние, эмоциональный, психологический и социальный статус. Оценка качества жизни затруднительна вследствие субъективности ощущений пациента. Однако этот критерий является крайне важным при анализе результатов лечения серьезных заболеваний, например, связанных с частичным и/или полным параличом, таких как инсульт или травма спинного/головного мозга и др. Современные опросники позволяют максимально объективно оценить изменения, связанные с качеством жизни [24]. Огромное разнообразие опросников утверждено в области психологии, где они являются основным инструментом при установлении диагноза, и в психиатрии. При проведении различных клинических испытаний также используется целый набор опросников, что позволяет оценивать эффективность того или иного клинического подхода [25, 26].

Таким образом, вышеприведенный обзор подтверждает актуальность проблемы, необходимость применения более гибких инструментов, эффек-

тивно взаимодействующих с экспертом в процессе анализа информации, а также методологической и программной интеграции различных источников и этапов обработки в одной программной системе.

3. Составление портфолио для допуска к клиническим испытаниям

Подходы к составлению портфолио для допуска к клиническим испытаниям основаны на комбинации современных методов интеллектуального анализа данных, анализа текстов на естественном языке и информационного поиска. Ниже приводится более детальное описание общей схемы решения задачи составления портфолио, а также её основных этапов.

Общая схема построения портфолио состоит из следующих этапов:

1. Ввод исходных данных.
2. Сбор материалов по тематике исследований.
3. Автоматизированный анализ материалов.
4. Отбор адекватных исследований.
5. Сравнительный анализ исследований.

Пункты 2-5 могут быть повторены для получения достаточно полной картины предметной области. Пользователь (эксперт) интерактивно взаимодействует с системой на всех этапах работы алгоритма.

Основные исходные данные для описания клинических испытаний вводятся в систему в виде текста на естественном языке и включают следующие поля:

- название нозологической формы;
- описание предлагаемых методов лечения;
- описание критериев отбора пациентов.

После ввода исходных данных выполняется автоматический анализ текста: сегментация, морфологический, синтаксический и семантический анализ, сопоставление слов со статьями метатезауруса, извлечение именованных сущностей. Таким образом, результатом первого этапа алгоритма является полуструктурированное описание объекта и метода испытаний в виде ключевой лексики, а также соответствия ключевых лексических элементов (слов и словосочетаний) и меток (наименований типов извлекаемой информации). Помимо наименований типов извлекаемой информации, лексическим элементам присваиваются метки в соответствии с наименованиями полей в форме ввода исходных данных, в которых встречаются эти лексические элементы (например, слово «подтвержденный» должно получить метку «КритерииОтбораПациентов»).

Следующий этап – сбор материалов по тематике исследований. В настоящей работе предлага-

ется следовать подходу, используемому в метапоисковых системах: отсутствие своего постоянного поискового индекса, отправка запросов к другим поисковым системам и переранжирование полученных документов с построением временного поискового индекса. Сбор материалов выполняется системой автоматически на основе введенных исходных данных и результатов их автоматического анализа. В качестве опорных поисковых машин можно использовать любые предметно-зависимые поисковые машины. В настоящее время в опытном образце системы используется база клинических испытаний ClinicalTrials.gov и электронная медицинская библиотека PubMed. Метод сбора информации состоит из нескольких итераций, каждая из которых в свою очередь содержит несколько этапов:

1. Генерация запроса.
2. Отправка запроса к ClinicalTrials.gov.
3. Анализ карточки клинического испытания с целью получения ссылок на научные статьи или дополнительной ключевой лексики.
4. Загрузка карточки клинического испытания и цитируемых статей.
5. Отправка запроса к PubMed.
6. Скачивание нескольких наиболее релевантных статей, если они не были скачаны ранее.

Пункты 1-6 повторяются до получения достаточного количества информации.

Генерация запроса – стохастический процесс, направленный на генерацию множества различающихся запросов ограниченной длины по определенной тематике. Применение стохастического процесса для генерации запросов лимитированной длины необходимо в виду следующих причин:

- Большинство поисковых машин имеют ограничение на допустимую длину запроса.
- Использование всего текста исходных данных нецелесообразно в виду особенностей работы поисковых машин. Опыт показывает, что длинные запросы с одной стороны приводят к увеличению полноты, а с другой – к сильному падению точности. Другой особенностью является учет порядка слов в запросе: некоторые поисковые машины его учитывают, некоторые – нет.
- Невозможно выбрать один запрос, по которому поисковая машина найдет все действительно релевантные документы.

Тематика задается извлеченными из исходных данных и ранее загруженных документов полуструктурированными описаниями исследований. Каждый запрос формируется путем генерации заданного количества реализаций дискретной случайной величины, значения которой соответствуют

различным лексическим элементам в имеющихся полуструктурированных описаниях. Распределение вероятностей зависит от меток, которые получили соответствующие лексические элементы в результате автоматического анализа текста. Количество реализаций случайной величины (соответственно, длина запроса) является настроечным параметром системы. Используемое распределение вероятностей является нестационарным и формируется, исходя из следующих принципов:

- Лексические элементы, получившие более значимые метки (к которым относятся, например, наименования лечебных процедур и нозологических форм), а также встречающиеся чаще, становятся более вероятными.
- Вероятность значений, соответствующих лексическим элементам, относящимся к разным полям формы ввода исходных данных, изменяется от запроса к запросу с целью поиска документов с разным приоритетом характеристик: похожие в первую очередь по нозологической форме, похожие в первую очередь по методам лечения, похожие в первую очередь по критериям отбора пациентов и т.п.
- После каждой итерации алгоритма сбора информации эксперту предъявляются загруженные карточки клинических испытаний, статьи и использованный запрос. Если эксперт указывает на необходимость продолжения поиска, итерация повторяется. В противном случае, система переходит к следующему пункту общего алгоритма. Таким образом, результатом этого этапа является коллекция документов по тематике предлагаемых клинических испытаний.

Цель автоматизированного анализа материалов заключается в извлечении структурированной информации из статей, которая может быть использована затем для решения следующих задач:

- пополнения ключевой лексики для составления новых поисковых запросов;
- помощи эксперту в отборе адекватных исследований;
- сравнения исследований.

Общий алгоритм анализа материалов заключается в последовательном анализе всех загруженных и еще не проанализированных документов. Анализ документа включает в себя следующие основные этапы:

1. сегментация;
2. морфологический анализ;
3. сопоставление со словарем;
4. синтаксический анализ;
5. общий семантический анализ (semantic role labelling);

6. распознавание именованных сущностей;
7. извлечение структурированной информации с помощью сторонних программных средств;
8. извлечение структурированной информации с помощью встроенных средств;
9. проверка экспертом извлеченной информации в интерактивном режиме.

Пункты 1-7 реализуются с помощью системы cTAKES. Входными данными для каждого этапа являются результаты работы всех предыдущих. Пункт 8 направлен на извлечение специализированной информации, которая не выделяется сторонними программными средствами, и реализуется посредством набора бинарных классификаторов. Каждый классификатор отвечает на вопрос “Принадлежит ли некоторый лексический элемент к заданному типу извлекаемой информации?”. Пункт 9 направлен на коррекцию ошибок классификации, допущенных в пункте 8: эксперт либо помечает дополнительные лексические элементы, пропущенные классификаторами, либо снимает метки с ошибочно помеченных. При достижении количеством правок некоторого заданного порога классификаторы переобучаются. Таким образом достигается последовательное улучшение качества извлечения структурированной информации. По умолчанию, в рамках пункта 8 выделяются фрагменты текста, соответствующие описаниям нозологической формы, методов лечения и критериям отбора пациентов.

Отбор адекватных исследований заключается в выборе тех, которые подходят для сравнения с исходным. Исследование считается адекватным если исследуемые в сравниваемых работах нозологические формы, методы лечения или критерии отбора пациентов являются схожими или родственными. Обзор работ в области автоматического анализа текстов показал, что в настоящее время не существует достаточно надежных методов сравнения объектов и процессов, описываемых в текстах. Поэтому финальное решение о включении или отсеивании найденных исследований в результирующее портфолио принимается экспертом. Разработанный алгоритм создания портфолио направлен на помощь эксперту в принятии объективного решения. Основой предложенного алгоритма является опросник. С целью уменьшения временных затрат на заполнение опросника при переходе к очередному вопросу эксперту предъявляются выдержки из рассматриваемого документа, наиболее вероятно имеющие отношение к текущему вопросу, а также автоматически помечается наиболее вероятный вариант ответа. Выдержки выбираются с учетом соответствия лексических элементов полям фор-

мы ввода исходных данных, полученным во время анализа ранее загруженных и проанализированных документов. Для выбора варианта ответа для каждого вопроса система создает классификатор и периодически дообучает его на основе ответов эксперта. Входные признаки для таких классификаторов соответствуют меткам (наименованиям типов извлекаемой информации). Значение признака вычисляется как количество общих лексических элементов, получивших соответствующую метку в описании предлагаемого испытания и найденном документе, нормированное на суммарное количество слов в обоих документах. В качестве метода классификации используется метод опорных векторов с линейным ядром. Таким образом, опросник определяет достаточные критерии для принятия решения об адекватности найденного исследования с точки зрения включения его в портфолио и не дает эксперту отойти от строгой методологии или «забыть» некоторые важные моменты. С другой стороны, система предоставляет удобные инструменты для ускорения заполнения опросника.

Опросник составляется на основе следующих разделов:

I. Выбор вида животного. Выбранный вид животного для анализа биологической активности и безопасности продукта регенеративной медицины должен демонстрировать биологический ответ на исследуемый продукт, похожий на тот, который ожидается на людях. При этом, для определения релевантности вида необходимо учитывать следующие факторы: а) сопоставимость физиологии и анатомии выбранного вида животного с человеком, б) пермиссивность/восприимчивость к инфицированию и репликации вирусного вектора или микробного вектора для генной терапии, в) иммунологическая толерантность к человеческому продукту клеточной терапии или человеческого трансгену, экспрессируемому за счет продукта генной терапии, г) возможность использования запланированной системы/процедуры клинической доставки продукта.

II. Выбор животных моделей заболевания. Доклинические исследования, проведенные на животных моделях заболеваний, должны давать представление о взаимоотношениях дозы продукта с его активностью и токсичностью. Причем для разработки клинических испытаний продуктов регенеративной медицины потенциально пригодными также могут являться животные модели заболеваний, используемые в базовых экспериментальных исследованиях. Благодаря особенностям продуктов регенеративной медицины (например,

потенциально пролонгированная продолжительность предполагаемого воздействия продукта, устойчивость продукта *in vivo*, сложный механизм действия, включающий взаимодействие между продуктом и окружением очага поражения, инвазивный способ применения), животные модели заболеваний являются более предпочтительными по сравнению со здоровыми животными для анализа активности и безопасности этих продуктов. Такие исследования позволяют лучше определить соотношение рисков и пользы, связанных с исследуемыми продуктами. Кроме того, использование моделей заболеваний предоставляет возможность для потенциальной идентификации биомаркеров активности-риска, которые могут быть применимы для мониторинга в клинических испытаниях. Каждая модель имеет сильные и слабые стороны; таким образом, ни одна модель не предсказывает с абсолютной точностью эффективности и безопасности результатов исследуемого продукта регенеративной медицины в популяции пациентов. На профиль активности и безопасности продукта может влиять время применения относительно начала заболевания, следовательно, должна быть охарактеризована стадия заболевания в момент применения продукта. Возможно, необходимы несколько животных моделей для адекватной оценки функциональных аспектов и потенциальных токсичностей одного продукта.

III. Характеристики исследуемого продукта регенеративной медицины. Продуктами регенеративной медицины являются различные препараты на основе клеточного материала (стволовые/прогениторные и (пред)дифференцированные клетки, индуцированные плюрипотентные клетки, различные скэффолды, заселенные клетками и пр.), генетические конструкции. Каждый из этих типов продуктов обладает собственными особенностями, которые необходимо охарактеризовать (так, например, для клеточных продуктов необходимо охарактеризовать их фенотип, источник получения, потенциальность, генетическую взаимосвязь с пациентом (аллогенные, аутологичные, ксеногенные и прочие параметры). Для всех продуктов регенеративной медицины необходим сбор данных, связанных с их анализом в экспериментах *in vitro*, для оценки аспектов биологической активности исследуемого продукта (например, секреция ростовых факторов, профиль иммунологических ответов, экспрессия нейротрансмиттеров). А также данные по их распределению в организме животных после введения и предполагаемым механизмам действия *in vivo*.

IV. Доказательство концепции (Proof-of-Concept Studies). Первичной целью этих исследований

является установление осуществимости и рациональности использования исследуемого продукта в определенной популяции пациентов. Данные исследования помогают дополнительно оценить риски и пользу продукта, они могут значительно способствовать выбору животной модели. Должны быть исследованы следующие параметры: а) Фармакологически эффективный диапазон доз (т.е. минимальная эффективная доза и биологическая оптимальная доза). б) Оптимизация способа введения и подтверждение, что продукт достигает анатомического сайта/ткани/клетки-мишени. с) Оптимизация времени применения продукта относительно начала заболевания. d) Оптимизация режима дозирования. е) Характеристика предполагаемого механизма действия или предполагаемые биологические активности исследуемого продукта. Данные, полученные в *in vitro* и *in vivo* доклинических тестированиях, могут служить ориентиром при дизайне и проведении доклинических токсикологических исследований, и клинических испытаний ранних фаз.

V. Токсикологические исследования. Оценка безопасности должна быть достаточно широкой, чтобы можно было провести идентификацию, охарактеризовать и количественно оценить потенциальные местные и системные токсичности, их начало (то есть, острые или с задержкой), возможность устранения любых видов токсичности, и эффект уровня дозы продукта на результаты токсичности. При оценке токсикологических исследований необходимо учитывать следующие параметры: 1) клинические показания; 2) количество и качество опубликованной доклинической или клинической информации о безопасности конкретного исследуемого продукта или похожего продукта (т.е., известные токсичности и побочные эффекты); 3) количество и качество существующих фармакологических (*in vitro/in vivo*) или Proof-of-Concept данных для конкретного исследуемого продукта или для похожего продукта; 4) данные об имеющемся доклиническом/клиническом опыте с предлагаемой клинической процедурой доставки или с какой-либо схожей процедурой; 5) биологическая отвечаемость вида животного на исследуемый продукт; 6) предполагаемый механизм действия продукта; 7) собственные характеристики продукта; 8) патофизиология животной модели заболевания; 9) адекватное количество животных каждого пола, которые соответственно рандомизированы в каждую группу; количество животных, требуемых для каждой группы, будет варьировать, исходя из соображений безопасности исследуемого продукта, вида, модели и системы доставки; 10) соответ-

ствующие контрольные группы: примеры включают животных, которые не получили лечения, которым сделали ложную операцию или ввели только транспортировщик препарата (среду), только адъювант, нулевой вектор или только скэффолд; 11) несколько уровней доз исследуемого продукта, которые должны охватывать предполагаемый диапазон клинических доз, для конкретных выбранных уровней доз должны быть представлены обоснования с подтверждающими их данными; 12) режим дозирования, который отражает предполагаемый клинический режим дозирования, в максимально возможной степени; 13) способ введения, который как можно ближе имитирует клинический способ применения продукта; 14) множество временных точек выведения животных из исследования, чтобы определить начало развития потенциальной острой, хронической и/или отсроченной токсичностей; временные интервалы между выведением животных из эксперимента зависят от используемой животной модели, исследуемого продукта, режима дозирования, наблюдаемых фармакодинамических и фармакокинетических ответов и предполагаемой популяции пациентов; 15) конечные точки безопасности, которые охватывают потенциальные токсичности; стандартные параметры, которые оцениваются на данном этапе включают смертность (с установлением причины гибели, если это возможно), клинические наблюдения, масса тела, физические оценки, потребление пищи/аппетит, потребление воды, клиническая патология (химия сыворотки, гематология, коагуляция, общий анализ мочи), масса органов, гистопатология; 16) дополнительные параметры, специфичные либо для исследуемого продукта, либо для предназначенной популяции пациентов.

Примеры параметров исследования, специфичных для продукта, включают гуморальный или клеточный иммунный ответы, биораспределение вектора, судьба продукта клеточных технологий, поведенческие тесты, неврологические оценки, офтальмологические оценки, визуализация (т.е., МРТ, УЗИ, рентгенография), наличие аномального/эктопического роста (т.е. гиперплазия, опухоли), предполагаемые биомаркеры, а также специализированная гистопатология (иммуногистохимия). Эти собранные данные должны включать морфологические и функциональные анализы, чтобы определить, существует ли связь между нетерминальными и конечными результатами.

VI. Способ введения, используемый для доставки исследуемого продукта, должен имитировать в максимально возможной степени способ введения, который планируется использовать в клинических ис-

следованиях. Потенциальные риски, которые могут быть связаны с использованием новых систем и/или процедур доставки исследуемого продукта, должны быть оценены и идентифицированы.

На основании этих данных доклинических исследований разрабатывается руководство для проведения ранних фаз клинических испытаний. Причем, каждый раздел опросника можно детализировать, дополнив его вопросами, которые позволят выделить более специфическую информацию. Например, это касается характеристики продукта регенеративной медицины, где каждому типу продукта присущи свои специфические особенности, которые должны быть охарактеризованы в ходе проведения доклинических исследований и, следовательно, учтены при создании портфолио.

При разработке портфолио для поздних фаз клинических испытаний необходимо учитывать данные по эффективности продукта регенеративной медицины при конкретной нозологии. При этом учитываются результаты доклинических исследований (*in vitro* и *in vivo*), по схеме, описанной выше, а также данные пилотных и ранних фаз клинических испытаний. В этом случае опросник дополняется соответствующим разделом (данные клинических испытаний).

Последний этап составления портфолио – сравнительный анализ предлагаемого и найденных исследований. Сравнение исследований должно отвечать, в том числе, на следующие вопросы:

- Какие методы лечения показали наибольшую результативность?
- Какие методы лечения наиболее часто применяются?
- Как зависит результативность от характеристик пациентов?
- Какая результативность достигается в среднем? и т.п.

Эксперту предоставляется инструментарий, который помог бы ему ответить на вышеприведенные вопросы с помощью интеллектуального поиска и многокритериальной группировки исследований. Алгоритм сравнительного анализа состоит из следующих основных шагов:

1. Перевод информации об исследованиях в векторное пространство.
2. Выбор критериев группировки.
3. Ввод поискового запроса и фильтрация документов (опционально).
4. Кластеризация документов в соответствии с выбранными критериями.

Пункты 2-4 могут выполняться несколько раз до получения представления предметной области необходимой детализации.

Алгоритм использует представление документов в виде точек в многомерном пространстве, в котором каждое измерение соответствует некоторой комбинации нормальной формы лексического элемента (слова или словосочетания) и метки (наименование типа извлекаемой информации). Значения координат документов присваиваются в соответствии с формулой TF-IDF. При кластеризации используется модифицированная формула расстояния Хэмминга, в которой каждое слагаемое умножается на вес, соответствующий важности этого слагаемого, а выбираемый экспертом критерий определяет значения этих весов. Это позволяет группировать и сравнивать исследования по нозологическим формам, методам, результатам, характеристикам пациентов, а также их комбинациям.

Используя вышеописанные функции, эксперт создает подборку информационного материала по заданным параметрам, который сгруппирован в соответствии с требованиями к каждому пункту портфолио. Таким образом, итоговое портфолио должно содержать следующие разделы: 1. Характеристика продукта регенеративной медицины (данные *in vitro* исследований биологической активности, данные первичных *in vivo* исследований). 2. Предпосылки и обоснование использования продукта для лечения конкретной нозологии. 3. Оценка адекватности используемой животной модели (или нескольких моделей). 4. Оценка безопасности исследуемого продукта в выбранной модели заболевания. 5. Установление режима дозирования и способов введения продукта. 6. Данные токсикологических исследований (с учетом вышеуказанных требований). 7. Описание предполагаемых/установленных механизмов действия. 8. Сравнительный анализ предлагаемого и уже имеющихся методов лечения данной нозологии. 9. Разработка критериев отбора пациентов. 10. Дизайн клинического испытания. 11. Методы клинического мониторинга.

4. Составление портфолио для модельных протоколов клеточной терапии

Для проверки разработанной концепции составления портфолио были выбраны два модельных протокола клеточной терапии. Можно выделить несколько ключевых характеристик клеточных препаратов, характеризующих уровень его потенциальной опасности. К таким характеристикам можно отнести: использование аллогенных или аутологичных клеток, тип клеток (степень их дифференцировки), способ введения, количество клеток в дозе.

Протокол 1.

Лечение ишемического инсульта внутривенным введением мезенхимальных стромальных

клеток (МСК), выделенных из пуповины человека. Начальная доза 100 млн. клеток.

Поиск в базе Clinicaltrials.gov показал, что внутривенное введение МСК, выделенных из пуповины человека, в протоколах клинических испытаний предлагается в дозе от 20 до 200 млн. клеток. Обнаружено три аналогичных протокола, направленных на лечение ишемического инсульта. Таким образом, предложенный протокол можно рассматривать, как безопасный и начальная доза выбрана адекватно.

Протокол 2.

Лечение карциномы пищевода внутривенным введением аутологичных генетически модифицированных Т-клеток, несущих химерный рецептор, в дозе 10^6 клеток/кг.

Поиск в базе Clinicaltrials.gov показал, что внутривенное введение генетически модифицированных Т-клеток в протоколах клинических испытаний предлагается в дозе от $0,5 \cdot 10^6$ клеток/кг до $3 \cdot 10^{11}$ клеток, что соответствует $4 \cdot 10^9$ клеток/кг при весе пациента 75 кг. Мы можем оценить предложенный метод как безопасный. Но возникает вопрос целесообразности введения клеток в предложенной дозировке. Она находится в самом начале интервала дозирования существующих протоколов. Это может привести к недостаточной эффективности разрабатываемого метода.

5. Заключение

Разработанные принципы построения и оценки портфолио для допуска к клиническим исследованиям могут быть использованы для разработки протоколов доклинических и клинических исследований методов регенеративной медицины, проводимых в рамках регистрации биомедицинских клеточных продуктов.

В наибольшей мере подход сопоставления с существующими протоколами клинических испытаний полезен для оценки безопасности разрабатываемого метода, поскольку биологические отличия человека и модельных животных не всегда позволяют адекватно перенести оценку безопасности, полученную в доклинических исследованиях.

Литература

1. Savova G. K. et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications //Journal of the American Medical Informatics Association. – 2010. – Т. 17. – №. 5. – С. 507-513.
2. Kiritchenko S. et al. ExaCT: automatic extraction of clinical trial characteristics from journal publications //BMC medical informatics and decision making. – 2010. – Т. 10. – №. 1. – С. 56.
3. Rindflesch T. C. et al. EDGAR: extraction of drugs, genes and relations from the biomedical literature //Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing. – NIH Public Access, 2000. – С. 517.
4. Xu H. et al. MedEx: a medication information extraction system for clinical narratives //Journal of the American Medical Informatics Association. – 2010. – Т. 17. – №. 1. – С. 19-24.
5. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology //Nucleic acids research. – 2004. – Т. 32. – №. suppl 1. – С. D267-D270.
6. Tu S. W. et al. A practical method for transforming free-text eligibility criteria into computable criteria //Journal of biomedical informatics. – 2011. – Т. 44. – №. 2. – С. 239-250.
7. Weng, C., Wu, X., Luo, Z., Boland, M. R., Theodoratos, D., Johnson, S. B. EliXR: an approach to eligibility criteria extraction and representation //Journal of the American Medical Informatics Association. – 2011. – Т. 18. – №. Supplement 1. – С. i116-i124.
8. Ross J. et al. Analysis of eligibility criteria complexity in clinical trials //AMIA Summits on Translational Science Proceedings. – 2010. – Т. 2010. – С. 46.
9. Miotto R., Weng C. Unsupervised mining of frequent tags for clinical eligibility text indexing //Journal of biomedical informatics. – 2013. – Т. 46. – №. 6. – С. 1145-1151.
10. Thiers F. A., Sinskey A. J., Berndt E. R. Trends in the globalization of clinical trials //Nature Reviews Drug Discovery. – 2008. – Т. 7. – №. 1. – С. 13-14.
11. Luo Z., Miotto R., Weng C. A human-computer collaborative approach to identifying common data elements in clinical trial eligibility criteria //Journal of biomedical informatics. – 2013. – Т. 46. – №. 1. – С. 33-39.
12. Pfiffner P. B. et al. ClinicalTrials.gov as a Data Source for Semi-Automated Point-Of-Care Trial Eligibility Screening //PloS one. – 2014. – Т. 9. – №. 10. – С. e111055.
13. Sim I. et al. Ontology-Based Federated Data Access to Human Studies Information //AMIA Annual Symposium Proceedings. – American Medical Informatics Association, 2012. – Т. 2012. – С. 856.
14. Hassanzadeh O. et al. Linkedct: A linked data space for clinical trials //arXiv preprint arXiv:0908.0567. – 2009.

15. *Sim I., Olasov B., Carini S.* An ontology of randomized controlled trials for evidence-based practice: content specification and evaluation using the competency decomposition method // *Journal of biomedical informatics.* – 2004. – Т. 37. – №. 2. – С. 108-119.
16. *Huser V., Cimino J. J.* Linking ClinicalTrials.gov and PubMed to track results of interventional human clinical trials // *PloS one.* – 2013. – Т. 8. – №. 7. – С. e68409.
17. *Tsafnat G.* et al. Systematic review automation technologies // *Syst Rev.* – 2014. – Т. 3. – №. 1. – С. 74.
18. *Tasneem A.* et al. The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty // *PLoS One.* – 2012. – Т. 7. – №. 3. – С. e33677.
19. *Deshpande A. M.* et al. TrialDB: a web-based clinical study data management system AMIA 2003 open source expo. – 2003. (электронная публикация http://pubmedcentralcanada.ca/pmc/articles/PMC1480035/pdf/amia2003_0794.pdf)
20. *Brandt C. A.* et al. Informatics tools to improve clinical research study implementation // *Contemporary clinical trials.* – 2006. – Т. 27. – №. 2. – С. 112-122.
21. *Cohen A. M.* et al. Evidence-based medicine, the essential role of systematic reviews, and the need for automated text mining tools // *Proceedings of the 1st ACM International Health Informatics Symposium.* – ACM, 2010. – С. 376-380.
22. *Hao T.* et al. Clustering clinical trials with similar eligibility criteria features // *Journal of biomedical informatics.* – 2014. – Т. 52. – С. 112-120.
23. *Atkinson N. L.* et al. Using the Internet to search for cancer clinical trials: A comparative audit of clinical trial search tools // *Contemporary clinical trials.* – 2008. – Т. 29. – №. 4. – С. 555-564.
24. *Исследование качества жизни в медицине:* учеб. пособие для вузов/А.А. Новик, Т.И. Ионова ; под ред. Ю.Л. Шевченко; Нац. мед.-хирург. центр им. Н.И. Пирогова Минздрава Рос. - М. : ГЭОТАР-МЕД, 2004.
25. *Hermanns N., Schmitt A.* et al., The Effect of a Diabetes-Specific Cognitive Behavioral Treatment Program (DIAMOS) for Patients With Diabetes and Subclinical Depression: Results of a Randomized Controlled Trial. *Diabetes Care.* 2015 Jan 20. pii: dc141416.
26. *Di Maio M., Gallo C.* et al., Symptomatic Toxicities Experienced During Anticancer Treatment: Agreement Between Patient and Physician Reporting in Three Randomized Trials. *J Clin Oncol.* 2015 Jan 26. pii: JCO.2014.57.9334.

Суворов Роман Евгеньевич. Мл.н.с. ИСА ФИЦ ИУ РАН. Окончил в 2012 г. Рыбинский государственный авиационный технический университет им. П.А. Соловьева. Количество печатных работ: 22. Область научных интересов: машинное обучение, интеллектуальный анализ данных, анализ текстов на естественном языке, извлечение информации, распределенные вычисления. E-mail: rsuvorov@isa.ru.

Ярыгин Николай Владимирович. Директор клиники ГКБ №5 ГБОУ ВПО МГМСУ им А.И. Евдокимова. Окончил в 1994 г. Московский медицинский стоматологический институт им. Н.А. Семашко. Доктор мед. наук. Количество печатных работ: 180. Область научных интересов: травматология, ортопедия, регенеративная медицина. E-mail: nvaryagin@yandex.ru

Попов Константин Васильевич. С.н.с. Федерального государственного бюджетного учреждения науки Институт молекулярной биологии им. В.А. Энгельгардта РАН. Канд. биол. наук. Окончил в 1998 г. МФТИ. Количество печатных работ: 35. Область научных интересов: регенеративная медицина, клеточная биология. E-mail: konstantin.v.popov@gmail.com

Холоденко Ирина Викторовна. С.н.с. Федерального государственного бюджетного научного учреждения «Научно-исследовательский институт биомедицинской химии им. В.Н. Ореховича» (ИБМХ). Канд. биол. наук. Окончила в 2002 г. МГУ им. М.В. Ломоносова. Количество печатных работ: 30. Область научных интересов: клеточная биология, регенеративная медицина, тканевая инженерия, онкология, иммунология. E-mail: irkhol@yandex.ru

Смирнов Иван Валентинович. Зав. лабораторией ИСА ФИЦ ИУ РАН. Окончил в 2003 г. Российский университет дружбы народов. Канд. физ.-мат. наук, доцент. Количество печатных работ: 52. Область научных интересов: обработка естественного языка, интеллектуальный анализ информации. E-mail: ivs@isa.ru