

РАЗРАБОТКА ПРОГРАММНОГО КОМПЛЕКСА ДЛЯ ПОДДЕРЖКИ МЕДИКО-ГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЙ

Ф.А. Колпаков*

*ООО «Институт системной биологии», Новосибирск
Конструкторско-технологический институт вычислительной техники СОРАН,
Новосибирск*

*e-mail: fedor@biouml.org

И.В. Тяжев

ООО «Институт системной биологии», Новосибирск

Н.И. Толстых

ООО «Институт системной биологии», Новосибирск

Е.Н. Кудрявцева

Институт химической биологии и фундаментальной медицины СОРАН, Новосибирск

Р.Н. Шарипов

ООО «Институт системной биологии», Новосибирск

Институт цитологии и генетики СОРАН, Новосибирск

У.А. Боярских

Институт химической биологии и фундаментальной медицины СОРАН, Новосибирск

М.Л. Филипенко

Институт химической биологии и фундаментальной медицины СОРАН, Новосибирск

Г.И. Лифшиц

Центр новых медицинских технологий СО РАН, Новосибирск

Аннотация

Современные медико-генетические исследования в силу использования высокопроизводительных методов и оборудования позволяют получить большой объем качественных и количественных данных. Эти данные необходимо сохранить, защитить от несанкционированного доступа, подвергнуть дальнейшему анализу, а также иметь возможность совместной работы с соответствующими правами доступа для персонала медицинского центра. В данной статье представлен программный комплекс “Genetics” для поддержки медико-генетических центров, обеспечивающий широкий спектр работ с персональными данными пациентов и результатами исследований согласно всем требованиям, предъявляемым Законодательством РФ. Разработанный комплекс успешно внедрен в ЦНМТ СО РАН, что позволило значительно ускорить обработку и анализ данных.

ВВЕДЕНИЕ

Исследование мультифакториальных заболеваний, например, таких как метаболический синдром, сердечно-сосудистые и онкологические заболевания, требует накопления большого количества клинических данных, включающих всестороннюю информацию о пациенте (история болезни, образ жизни, данные лабораторных исследований, генетическая предрасположенность, результаты генотипирования и т.п.). Для проведения таких исследований может потребоваться объединение усилий нескольких медицинских центров, чтобы создать представительные выборки для последующего

статистического анализа. Оптимальным способом ввода данных при мультицентровых исследованиях является веб-интерфейс, когда авторизованный персонал имеет возможность непосредственного просмотра, ввода и редактирования соответствующих данных.

ЦЕЛИ И ЗАДАЧИ

Целью данной работы являлось создание на основе современных веб-технологий программного комплекса для проведения мультицентровых клинико-эпидемиологических исследований мультифакториальных заболеваний с учетом действующего законодательства РФ по защите персональных данных.

В рамках междисциплинарных проектов СО РАН нами был разработан специальный программный комплекс “Genetics”, который позволяет решать следующие задачи:

- ведение реестра обследованных пациентов;
- возможность настройки показателей (добавление, редактирование) используемых в различных клинико-эпидемиологических исследованиях (различного вида опросники, методы лечения, информация о наследственных заболеваниях, результаты генотипирования и т.п.);
- просмотр и редактирование информации о пациенте в виде интерактивной карточки;
- просмотр и ведение списка полиморфизмов исследуемых генов;
- возможность массовой загрузки уже существующих данных путем импорта данных о пациентах и результаты генотипирования;
- формирование выборок по заданным оператором условиям и их сохранение в виде обезличенных персональных данных в отдельную базу данных;
- статистический анализ данных, средствами разработанного комплекса и пакетов BioUML (<http://www.biouml.org>) и R/Bioconductor (<http://www.bioconductor.org>).

АРХИТЕКТУРА ПРОГРАММНОГО КОМПЛЕКСА

Программный комплекс построен по классической трехуровневой архитектуре (Рис. 1):

1 уровень – базы данных (MySQL);

2 уровень – сервер приложений, используемый для реализации основной части бизнес-логики и анализа статистических данных;

3 уровень – пользовательский интерфейс.

Применение порталных технологий предполагает использование в качестве пользовательского интерфейса обычного веб браузера («тонкий» клиент). Для генерации презентационной логики и обеспечения основной части бизнес-логики используется технология BeanExplorer Enterprise Edition (<http://www.beanexplorer.com>), которая представлена набором сервлетов на сервере приложений и использует протокол JDBC для работы с сервером баз данных. Комплекс обеспечивает решение целого ряда задач от ведения реестра пациентов до формирования выборок пациентов и статистического анализа данных средствами модуля BioUML Enterprise Edition (BioUML EE), а также с привлечением внешних программ (например, R/Bioconductor).

Пользовательский интерфейс предоставляет оператору следующие возможности:

- доступ в систему осуществляется только для авторизованных пользователей (идентификация пользователя и пароля);

- интерфейс для ввода данных состоит из карточки, содержащей следующие вкладки:
 - общие данные о пациенте (Рис. 2),
 - диагноз,
 - качество жизни,

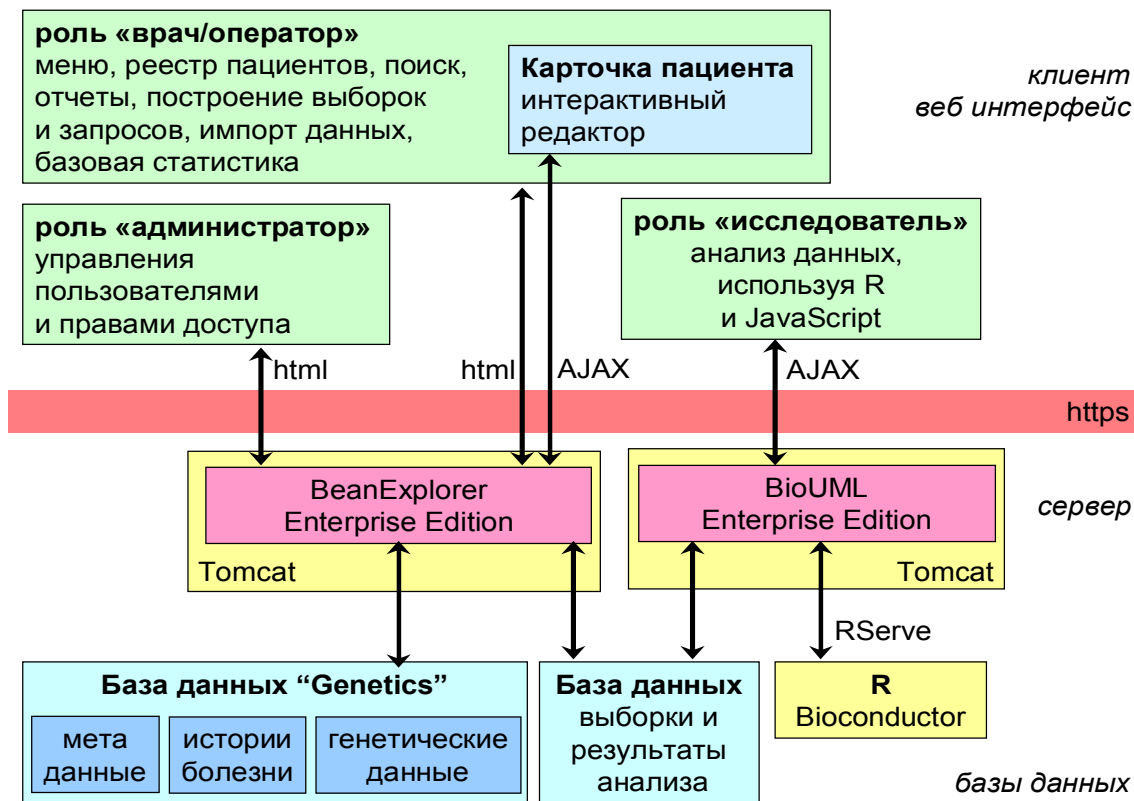


Рис. 1. Архитектура программного комплекса «Genetics».

- факторы риска,
 - питание,
 - данные генетического анализа,
 - данные о лечении (Рис. 2),
 - данные об общем состоянии здоровья.
- при вводе данных, везде, где это возможно, используются справочники;
 - осуществляется контроль заполнения обязательных полей;
 - веб интерфейс обеспечивает возможность формирования выборок пациентов по набору заданных показателей и выгрузку данных для их анализа в различных программах с возможностью статистических расчетов (Matlab, BioUML).

ПЕРСОНАЛЬНЫЕ ДАННЫЕ

Данные, передаваемые оператором в базу данных, содержат информацию о более чем 200 показателей, характеризующих пациента, что позволяет оптимизировать процесс статистического анализа, установления диагноза и назначения дальнейшего курса лечения.

Федеральный закон № 152 от 27.07.2006 года «О персональных данных» [1] относит медицинские данные к 1-й категории персональных данных, предъявляя жесткие требования по обеспечению их защиты. Программное обеспечение, работающее с данными пациентов, обязано соответствовать заявленным для этой категории условиям, поскольку нарушение

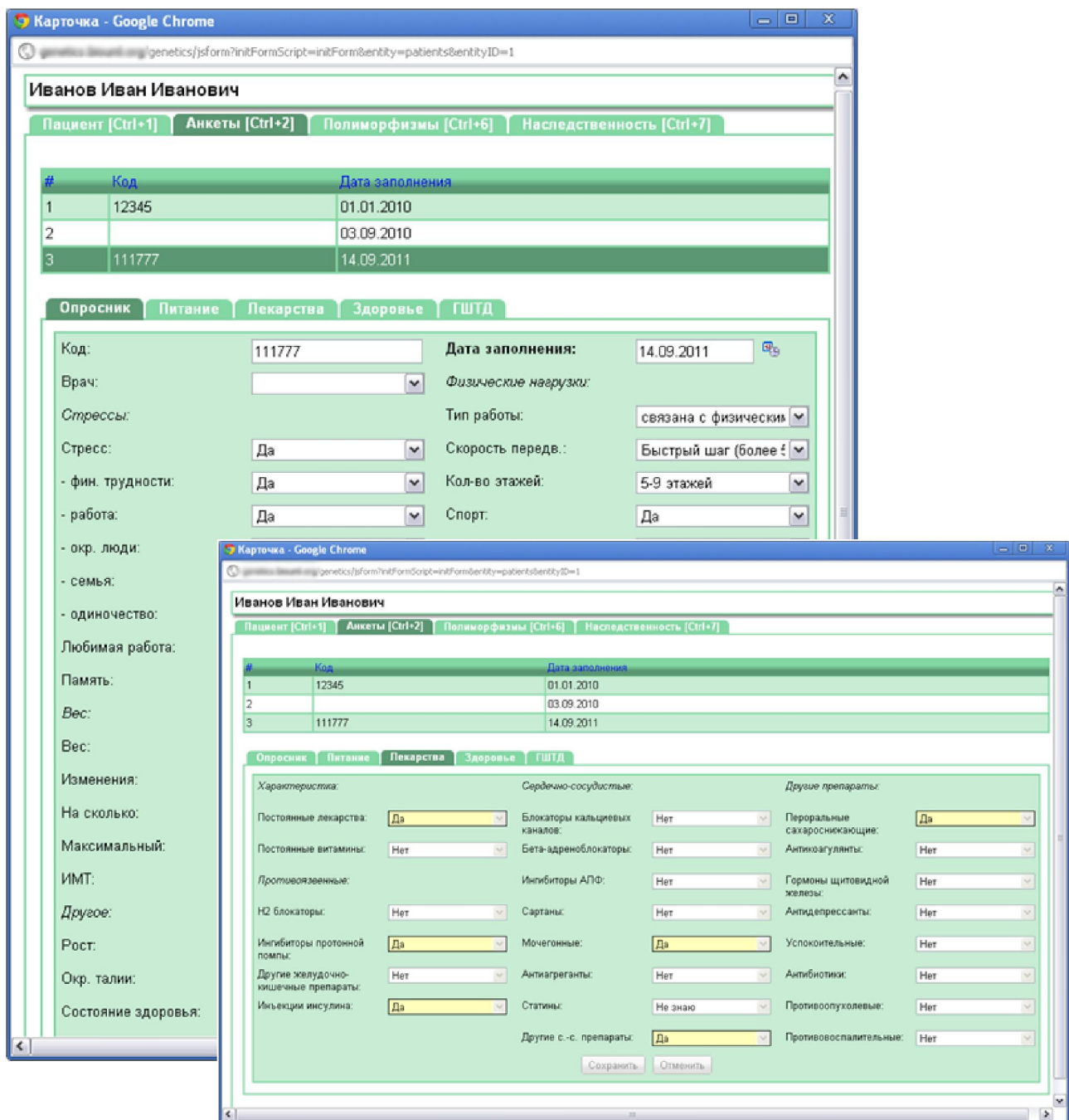


Рис. 2. Формы комплекса “Genetics” для ввода персональных данных пациента и информации о его лечении.

заданной характеристики безопасности персональных данных, обрабатываемых в них, может привести к серьезным негативным последствиям.

Следует отметить, что для проведения статистического анализа нет необходимости использовать полные персональные данные пациентов, поэтому для дальнейшей работы они могут быть обезличены. Такие данные относятся к 4-й категории, и требования к их защите минимальны. Это позволяет привлечь к анализу клинико-эпидемиологических данных исследователей, которые не имеют прав доступа к персональным данным, соответствующим высоким категориям (1-3). Для статистического анализа в разработанном нами программном комплексе используются только обезличенные данные пациентов (4-я категория), созданные средствами комплекса на основе персональных данных (1-я категория), которые хранятся в

отдельной защищенной базе данных, работа с которой осуществляется при помощи BioUML EE и пакета R/Bioconductor.

РАБОТА С ВЫБОРКАМИ И СТАТИСТИЧЕСКИЙ АНАЛИЗ

Одной из важных возможностей “Genetics” является формирование выборок пациентов по определенным параметрам и условиям. Для этого в меню построителя выборок задаются требуемые параметры и дополнительные условия (Рис. 3). В результате из общего списка анкет пациентов отбираются удовлетворяющие всем заданным условиям и создается таблица, содержащая вместо полных ФИО пациентов уникальные идентификаторы, которая сохраняется в базе данных для дальнейшего анализа в BioUML EE.

Наиболее мощным свободно распространяемым пакетом программ для статистического анализа широкого круга биологических и медицинских данных в настоящее время является платформа R/Bioconductor, поэтому при создании модуля статистических

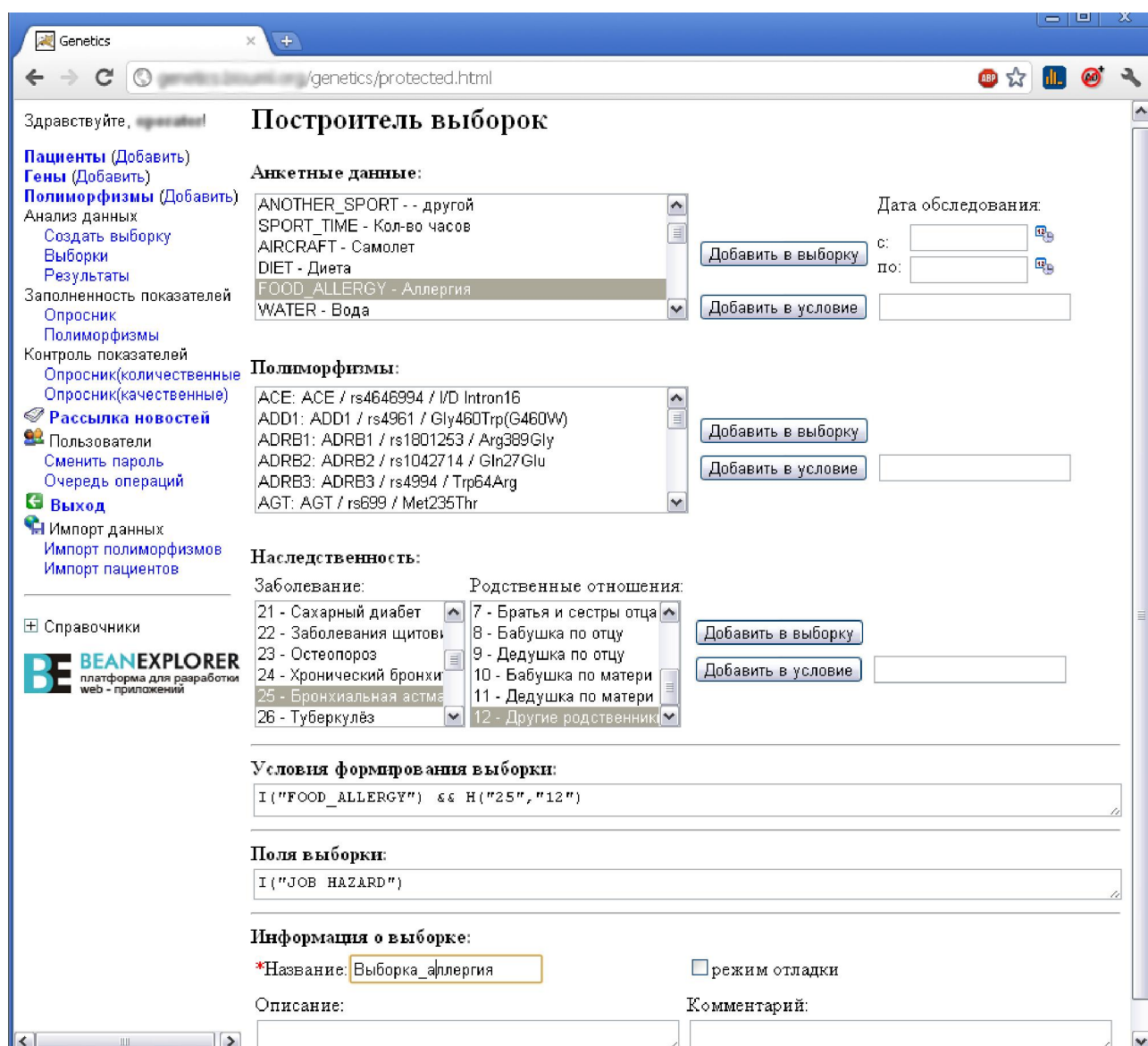


Рис. 3. Меню построителя выборок программного комплекса “Genetics”. Пример создания для исследования выборки пациентов, страдающих пищевой аллергией, работающих на вредном производстве, родственники которых страдали бронхиальной астмой.

исследований в программном комплексе мы ориентировались на его интеграцию с R. Поскольку анализ данных будет проводиться в разных организациях, программный комплекс “Genetics” позволяет проводить удаленную работу с данными пациентов и пакетом R/Bioconductor, развернутом на специальном сервере, через веб-интерфейс.

Внедрение разработанного нами программного комплекса “Genetics” в ЦНМТ СО РАН позволило значительно облегчить, унифицировать и ускорить ввод информации в базу данных, а также обработку и проведение статистического анализа данных пациентов. Накопленная в базе информация о пациентах и результатах исследований в дальнейшем позволит проводить более масштабные исследования, а также, благодаря программной гибкости, заложенной в архитектуре комплекса, интегрировать данные категории 4 с международными базами данных для совместных исследований, если возникнет необходимость в рамках глобальных проектов.

Работа поддержана в рамках междисциплинарного проекта №17 СО РАН и проекта №33 программы «Фундаментальные науки – медицине» Президиума РАН.

ЛИТЕРАТУРА

[1] Федеральный закон № 152 от 27.07.2006г. «О персональных данных»
<http://base.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=117587>