

## Резюме проекта, выполняемого

в рамках ФЦП

### «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014 – 2020 годы»

по этапу № 3

Номер Соглашения о предоставлении субсидии: 14.576.21.0073

Тема: «Разработка Интернет-доступного сервиса поисковой протеомной машины для идентификации белков живых организмов.»

Приоритетное направление: Науки о жизни

Критическая технология: Геномные, протеомные и постгеномные технологии

Период выполнения: 05.11.2014 - 31.12.2015

Плановое финансирование проекта: 7.00 млн. руб.

Бюджетные средства 3.50 млн. руб.,

Внебюджетные средства 3.50 млн. руб.

Получатель: Общество с ограниченной ответственностью "КуБ"

Ключевые слова: протеомика, протеом человека, биоинформатика, хромато-масс-спектрометрия, белки, пептиды

#### 1. Цель проекта

Целью проекта является разработка Интернет-доступного сервиса для хранения и обработки данных хромато-масс-спектрометрических (ХМС) измерений, полученных в рамках высокопроизводительного панорамного протеомного анализа на содержание белков в клинических биопробах, включая экстракты клеток тканей и физиологических жидкостей.

Основой сервиса является протеомная поисковая машина (ППМ) для: (1) автоматизированной обработки результатов ХМС измерений в глубоком панорамном анализе биопроб; (2) идентификации пептидов/белков; (3) отчета результатов поиска в стандартизированных форматах, рекомендованных международным консорциумом Протеом Человека (HPP).

#### 2. Основные результаты проекта

На 1-м этапе разработана библиотека парсеров и функций, предназначенная для разработки специализированного программного обеспечения поисковой протеомной машины (ППМ) с функциональными возможностями, включающими обработку результатов ХМС измерений, поиска пептидных идентификаций в пробах на основе протеомных баз данных, обработку результатов поиска и их представления в стандартизированных форматах. Был разработан протокол работы с данными ХМС измерений для их первичной обработки и конвертации в стандартизированные форматы *MGF*, *mzML* и др. для последующей работы ППМ с данными и получения результатов измерений. Получены экспериментальные данные панорамных ХМС измерений для смесей стандартов и биопроб для проведения сравнения поисковых алгоритмов. Показана необходимость совмещения функции валидации результатов поиска (в аналогах проводится отдельно) для повышения количества числа идентификаций за счет дополнительного поиска по спектрам второго уровня (МС/МС) с индексами достоверности, полученными по результатам первого поиска, ниже порогового. На основе сравнения поисковых алгоритмов обоснована концепция автоматизированной оптимизации входных параметров поиска в разрабатываемой ППМ.

На 2-м этапе разработано программное обеспечение поисковой протеомной машины (ППМ *IdentiPROT*) с функциональностями, соответствующими требованиям ТЗ. ПО ППМ *IdentiPROT* было размещено на разработанном сервере хранения и обработки данных мощностью 6 Тб. К серверу был организован сетевой защищенный доступ по локальной сети и по сети Интернет по протоколу HTTP для работы с ППМ из клиентского места. Для работы пользователя с ППМ был разработан графический веб-интерфейс, размещенный на сервере. Были получены экспериментальные данные для тестирования отдельных модулей и функций ППМ, а также разработаны проекты технической документации ППМ, включая текст и описание ПО. Был осуществлен запуск ППМ *IdentiPROT* и разработана программа и методики исследовательских испытаний.

На 3-м этапе были получены экспериментальные данные панорамного хромато-масс-спектрометрического анализа биологических проб с наработкой более 25000 масс-спектров второго уровня МС/МС для проведения исследовательских испытаний ППМ *IdentiPROT*. Проведены исследовательские испытания ППМ *IdentiPROT*. (п. 3.1).

По результатам исследовательских испытаний доработано ПО ППМ, включая добавление более 50 новых функций с целью

повышения технических характеристик и функциональные возможности ППМ (п.3.2). Разработано руководство пользователя и администратора ППМ IdentiPROT для обработки хромато-масс-спектров (п.3.3-3.4). Проведено обобщение и выводы по результатам ПНИ (п.3.5). Разработаны рекомендации и предложения по использованию результатов ПНИ (п.3.6). Разработан проект технического задания на проведение ОКР по теме: «Создание поисковой протеомной машины для идентификации белков протеомов живых организмов на основе хромато-масс-спектрометрических измерений и Интернет-доступный сервис на ее основе» (п.3.7). Подготовлен Заключительный отчет о ПНИ (п.3.8).

Результаты ПНИ в отчетный период были доложены на 3 мероприятиях, а также опубликованы в 2 реферируемых научных изданиях, зарегистрированных в базах Web of Science.

1) Разработанное ПО ППМ IdentiPROT содержит три основных функциональных элемента, реализующих все базовые функциональности поисковых протеомных машин по обработке экспериментальных данных, а также алгоритмов статистической валидации пептидных и белковых идентификаций. Разработанный сервер хранения и обработки протеомных данных позволяет хранение до 6 Тб пользовательских данных, параллельный доступ к пользовательскому интерфейсу и многопоточную обработку данных.

2) Реализована оптимизация 6 входных параметров поиска при обработке экспериментальных данных с использованием ПО ППМ IdentiPROT, а также программный интерфейс для использования сторонних скоринг-функций. Данные функциональности отсутствуют в аналогичных сервисах и программных решениях.

3) ПО ППМ IdentiPROT полностью отвечает требованиям ТЗ и содержит программную реализацию всех функциональностей, перечисленных в п. 4.2 ТЗ.

4) Предварительные тесты ПО ППМ IdentiPROT на экспериментальных данных, полученных для биологических проб клеточных линий, показали её полное соответствие мировым аналогам.

### **3. Охраноспособные результаты интеллектуальной деятельности (РИД), полученные в рамках прикладного научного исследования и экспериментальной разработки**

Изобретение заявка № 2015104718/20(007386) от 12.02.2015г., тема: «Способ оптимизации вычисления статистических показателей достоверности при идентификации веществ путем масс-спектрометрического анализа с последующим поиском по базам данных».

### **4. Назначение и область применения результатов проекта**

Предлагаемая ППМ предназначена для определения белкового состава биопроб в: (1) лабораторной диагностике: количественное выявление белковых маркеров социально значимых заболеваний; (2) биомедицинских исследований в области протеомики и постгеномных технологий;

(3) развитие методов персонализированной медицины.

Реализация проекта позволит уменьшить зависимость от зарубежных коммерческих продуктов.

### **5. Эффекты от внедрения результатов проекта**

Разрабатываемая ППМ нового поколения должна не просто «дополнять» стандартные решения ППМ, а предоставлять независимую оценку достоверности идентификаций. Предусмотрено повышение достоверности и чувствительности анализа, что позволит проводить направленный поиск белков, количественно характеризующих стадии развития социально-значимых заболеваний, а также раннего развития патологий на клеточном уровне. На сегодня результаты протеомных исследований позволяют сократить на 30% расходы на разработку новых лекарств. Сокращается и время, требуемое на разработку лекарства, от лаборатории до клиники.

### **6. Формы и объемы коммерциализации результатов проекта**

Планируемые формы коммерциализации результатов проекта: (1) продажа услуги кратко- и долгосрочного Интернет-доступа к ППМ на сервере поставщика услуги, ООО «КуБ»; (2) предоставление платных услуг ООО «КуБ» по обработке данных ХМС измерений на основании договоров об оказании услуг; (3) продажа лицензии производителям или дилерам на использование ППМ в составе поставляемого/обслуживаемого масс-спектрометрического оборудования; (4) разработка, в рамках договоров об оказании услуг, модифицированных версий ППМ с набором функциональностей, определяемых Заказчиками для решения специализированных задач; (5) продажа дистрибутива ППМ для установок на ПК Заказчика.

Стоимость лицензии с ПО ППМ оценивается в 400 тыс руб/ПК, что в два раза дешевле коммерческого аналога Mascot (Matrix Science, США). Стоимость лицензионного использования в составе поставляемого в РФ масс-спектрометрического оборудования оценивается в 500 тыс руб, что также в два раза ниже соответствующих ППМ Mascot или SEQUEST. Объем услуг оценивается 3-5 млн. рублей/год и зависит от маркетинговых инвестиций. Планируется реализация платной услуги по Интернет-доступу к ППМ за пределами РФ. Коммерциализация результатов проекта на территории РФ предусматривает использование административных ресурсов ФАНО, госпрограмм развития фарминдустрии для реализации 100% импортозамещения ППМ в составе поставляемого (бюджетное финансирование) масс-спектрометрического оборудования.

**7. Наличие соисполнителей**

ИП Дубровина - 2015 год  
Федеральное государственное бюджетное научное учреждение "Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича" (ИБМХ) - 2015 год  
ООО "Дата Технологии" - 2015 год

Общество с ограниченной ответственностью "Куб"

\_\_\_\_\_  
директор  
(должность)

\_\_\_\_\_  
(подпись)                      Кудрявцев А.М.  
(фамилия, имя, отчество)

**Руководитель работ по проекту**

\_\_\_\_\_  
заведующий лабораторией  
(должность)

\_\_\_\_\_  
(подпись)                      Горшков М.В.  
(фамилия, имя, отчество)

**М.П.**