

## "УТВЕРЖДАЮ"

Директор Федерального государственного  
учреждения "Федеральный исследовательский  
центр Институт прикладной математики им.  
М.В. Келдыша Российской академии наук"  
член-корр РАН



А.И. Аптекарев

« 19 » сентября 2017

## ОТЗЫВ

ведущей организации на диссертацию

**Кулаковского Ивана Владимировича**

«Регуляторные мотивы в геномах высших эукариот и их роль в экспрессии генов», представленную на соискание степени доктора биологических наук по специальности 03.01.09 – «Математическая биология, биоинформатика»

### **Актуальность темы диссертационной работы.**

Актуальность темы диссертации Кулаковского И.В. определяется активным развитием современных высокопроизводительных методов молекулярной биологии (в частности, методов массового прочтения ДНК и определения сайтов связывания ДНК с белками) и применением этих методов к широкому спектру биологических задач. Быстрое накопление большого объема экспериментальных данных требует разработки новых компьютерных биоинформатических технологий для обработки этих данных и получения новых биологических знаний. Представленная диссертация сфокусирована на задачах регуляторной геномики, а именно, идентификации и систематизации характерных

паттернов в некодирующих участках геномов высших эукариот, участвующих в ДНК-белковом узнавании, для установления роли геномной последовательности в регуляции экспрессии генов. Разработка и использование новых методов для компьютерного анализа паттернов-"мотивов" в регуляторных геномных последовательностях с использованием современных экспериментальных данных, безусловно, являются актуальными.

### **Новизна и значимость основных результатов.**

Работа включает в себя и методическую часть, состоящую в разработке комплекса новых алгоритмов и компьютерных программ, и практическую часть по применению авторских методов для анализа разнообразия регуляторных паттернов, распознаваемых белками-факторами транскрипции, и конкретных регуляторных систем, зависящих от паттернов в геномных последовательностях при управлении экспрессией генов у высших эукариот. Представленная работа вносит принципиальный вклад в решение проблем биоинформатики регуляторных мотивов у высших эукариот, что подтверждается высокими импакт-факторами журналов, в которых опубликованы результаты исследований, и активным цитированием авторских статей, судя по данным основных систем учета научного цитирования. Основные результаты, представленные в диссертации, состоят в следующем:

1. В диссертации представлен комплекс новых алгоритмов и программ для анализа паттернов-"мотивов" в объемных выборках геномных последовательностей, получаемых на основе современных методов высокопроизводительного секвенирования. В набор входят и методы идентификации паттернов *de novo*, и методы сравнения паттернов с известными, и методы оценки регуляторного потенциала однонуклеотидных вариантов. Авторские методы предоставлены в открытый доступ в сети Интернет как программы с открытым исходным кодом.
2. Создана полномасштабная коллекция паттернов, узнаваемых в ДНК факторами транскрипции человека и мыши. При построении коллекции использован широкий спектр современных экспериментальных данных по ДНК-белковому узнаванию. Коллекция покрывает все основные известные структурные семейства факторов транскрипции и предоставлена в сети Интернет в виде интерактивной базы данных.

3. С помощью авторских биоинформатических методов получен ряд важных результатов, имеющих значения для молекулярной биологии и, конкретно, для регуляторной геномики высших эукариот: (а) Найден тройственный композитный элемент сайтов связывания ключевых регуляторов плюрипотентности; (б) получены данные о функции гомотипических пар участков, связываемых фактором транскрипции Sp1, в мышинной модели эритролейкемии; (в) Подтверждена локализация пиримидин-богатых паттернов 5'-области мРНК, являющихся мишенями сигнального каскада mTOR, выдвинута гипотеза об альтернативной регуляции альтернативных транскриптов mTOR-мишеней на уровне трансляции; (г) Выявлен отрицательный отбор, действующих на онкологические мутации в сайтах связывания факторов транскрипции, а именно, установлено, что многие участки связывания факторов транскрипции избегают мутационных изменений; (д) Проведена систематизация паттернов, представленных в промоторах, активных в различных типах клеток; установлено, что разнообразие промоторных регуляторных сигналов описывается каталогом известных мотивов; (е) Выявлено систематическое избегание ко-локализации сайтов связывания факторов транскрипции и CG-пар, дифференциальное метилирование которых связано с активностью близлежащих промоторов.

#### **Содержание и структура работы.**

Диссертационная работа построена по традиционной схеме, включающей введение, обзор литературы, разделы «Материалы и методы» и «Результаты», а также заключение, выводы и список процитированных источников из 541 наименования.

Введение посвящено формулировке актуальности темы, цели и задачам, обоснованию научной новизны и значимости работы и описанию личного вклада автора.

Обзор литературы включает в себя три подраздела, сфокусированных на (1) понятии паттерна-"мотива" и роли паттернов, соответствующих участкам связывания белков-регуляторов транскрипции, в структуре регуляторных последовательностей генома; (2) вычислительном представлении мотивов и сопутствующих методах анализа; (3) экспериментальных методах изучения ДНК-белкового узнавания с акцентом на современных методах, использующих техники высокопроизводительного секвенирования.

В разделе «Материалы и методы» представлены авторские методы для идентификации мотивов в нуклеотидных последовательностях, сравнения мотивов, оценки статистической значимости вхождений и важности однонуклеотидных замен.

В разделе «Результаты» представлена авторская систематическая коллекция мотивов связывания факторов транскрипции человека и мыши, а также результаты использования биоинформатических методов и коллекции мотивов для анализа конкретных примеров регуляции экспрессии генов у мыши и человека.

#### **Рекомендации по использованию результатов и выводов диссертации.**

Результаты, представленные в диссертации, и выводы, сделанные в работе, имеют важное значение для биоинформатики регуляторных паттернов высших эукариот; методы, представленные в работе, могут применяться в различных задачах по анализу коротких паттернов в нуклеотидных последовательностях. Таким образом, результаты работы могут использоваться в учебном процессе и научных исследованиях, проводимых в организациях, работающих в области биоинформатики и молекулярной биологии, в том числе в Федеральном государственном бюджетном учреждении науки Институте биологии гена Российской академии наук, на Биологическом факультете и Факультете биоинженерии и биоинформатики МГУ им. М.В. Ломоносова, в Федеральном государственном бюджетном научном учреждении «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», в Федеральном государственном автономном образовательном учреждении высшего образования «Санкт-Петербургский политехнический университет Петра Великого», в Федеральном государственном бюджетном учреждении науки Институте белка Российской академии наук и других научно-исследовательских и образовательных организациях.

#### **Общие замечания и пожелания.**

Текст диссертации не лишен опечаток и стилистических недочетов, которые, тем не менее, не препятствуют пониманию текста. Представленная работа представляет собой законченное исследование, представленные методы и базы данных имеют широкие возможности для практического использования. Так, перспективным выглядит

использование результатов диссертации в исследованиях комбинаторной регуляции экспрессии для идентификации характерных комбинаций и взаимного расположения геномных участков, узнаваемых различными факторами транскрипции. Ключевые биологические результаты, касающиеся ко-локализации сайтов связывания и других функциональных элементов генома, таких как мутационные контексты, также имеют потенциал для дальнейших исследований в области регуляции экспрессии генов высших эукариот.

### **Заключение.**

По объему, значимости выполненной работы, количеству и уровню публикаций в ведущих международных изданиях, соответствующих тематике работы, представленная диссертация безусловно является полноценным научным исследованием, выводы работы достоверны и подтверждены фактическим материалом, автореферат диссертации полно отражает ее содержание. Диссертационная работа Кулаковского Ивана Владимировича «Регуляторные мотивы в геномах высших эукариот и их роль в экспрессии генов» соответствует п.9 Положения «О порядке присуждения ученых степеней», утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, с изменениями Постановления Правительства Российской Федерации от 21 апреля 2016 года №335, а ее автор – **Кулаковский Иван Владимирович заслуживает присуждения ученой степени доктора биологических наук по специальности «03.01.09 - Математическая биология, биоинформатика».**

Диссертационная работа Кулаковского Ивана Владимировича рассмотрена, обсуждена и одобрена на межлабораторном семинаре Института математических проблем биологии РАН - филиала ИПМ им. М.В. Келдыша РАН - 8 сентября 2017 года, протокол № 9.

