Экспертиза диссертации

Автор: Иван Владимирович Кулаковский

Название: Регуляторные мотивы в геномах высших эукариот и их роль в экспрессии генов

Объект исследования: Некодирующие участки геномов высших эукариот, участвующие в регуляции транскрипции, сайты связывания факторов транскрипции и их мотивы.

Цель исследования: идентификация и систематизация мотивов ДНК-белкового узнавания факторов транскрипции в последовательностях не кодирующих районов генома путем вычислительного анализа результатов современных высокопроизводительных экспериментальных методов и практическое использование мотивов для задач регуляторной геномики высших эукариот.

Задачи работы: (1) разработка новых методов биоинформатики для идентификации, поиска и сравнения паттернов-мотивов в больших выборках нуклеотидных последовательностях; (2) создание систематической коллекции мотивов связывания факторов транскрипции мыши и человека на базе современных экспериментальных данных; (3) апробация разработанных методов в разнообразных практических задачах регуляторной геномики по изучению: (а) взаимной локализации мотивов ключевых факторов плюрипотентности OCT4/SOX2/NANOG, (б) связи кластеризации сайтов связывания фактора Sp1 с экспрессией генов в мышинной модели эритролейкемии, (в) действия отбора на соматические мутации в сайтах связывания факторов транскрипции, происходящие в геномах раковых клеток, (г) взаимосвязи регуляции трансляции и транскрипции на примере сигнального пути mTOR; (д) взаимной локализации сайтов связывания факторов транскрипции и CpG-светофоров; (е) разнообразия мотивов в промоторах геномов мыши и человека.

Методы исследования: Вычислительные методы регуляторной геномики, в первую очередь методы биоинформатики для анализа последовательностей. Ключевые методы являются авторскими: разработаны новые алгоритмы и компьютерные программы для идентификации мотивов в больших выборках нуклеотидных последовательностей, для сравнения мотивов и оценки статистической значимости вхождений.

Обзор литературы: Обзор литературы содержит 541 ссылку, включая как классические работы, так и наиболее современные литературные источники по теме. Обзор покрывает все основные вопросы, связанные с темой диссертации.

Основные результаты: разработаны новые методы биоинформатики для анализа мотивов в нуклеотидных последовательностях: методы идентификации мотивов на основе современных экспериментальных данных, полученных с использованием высокопроизводительного секвенирования; метод оценки сходства мотивов на основе меры Жаккара; усовершенствованный метод оценки важности однонуклеотидных вариантов во вхождениях мотивов.

Проведена полномасштабная идентификация и верификация мотивов ДНК-белкового узнавания на основе публично доступных результатов сотен экспериментов по изучению сайтов связывания различных факторов транскрипции *in vivo* и *in vitro*. Создана новая база данных, содержащая мотивы сайтов связывания для нескольких сотен факторов транскрипции мыши и человека. Для многих факторов транскрипции представленные мотивы описывают ДНК-сайты связывания наилучшим образом среди существующих источников.

С помощью новых методов и систематической коллекции мотивов получен ряд новых результатов в регуляторной геномике высших эукариот. (I) Найден композитный элемент сайтов связывания ключевых факторов плюрипотентности OCT4-SOX2/NANOG. (II) Установлено, что действие фактора транскрипции Sp1 на активность генов-мишеней в мышечной модели эритролейкемии зависит от локализации и ориентации пар сайтов связывания: тандемные сайты связывания Sp1 ингибируют экспрессию генов при локализации в энхансерах и активируют экспрессию генов при локализации в промоторах, не пересекающихся с CpG-островками. (III) С помощью совместного анализа данных рибосомного профайлинга и кэп-анализа экспрессии генов для мРНК-мишеней сигнального пути mTOR подтверждена локализация пиримидин-богатых мотивов в непосредственной окрестности 5'-концов и обнаружены протяженные регионы инициации транскрипции, что свидетельствует в пользу альтернативной трансляционной регуляции mTOR-мишеней в зависимости от активности соответствующих промоторов. (IV) Выявлено давление отрицательного отбора на мутации в сайтах связывания факторов транскрипции, возникающие в геномах различных типов

раковых клеток. (V) Проведена идентификация мотивов в полногеномном каталоге промоторов, определенных в проекте FANTOM5 для различных типов клеток мыши и человека, и установлено, что разнообразие регуляторных мотивов практически исчерпано. (VI) Выявлено избегание сайтами связывания факторов транскрипции перекрытия с CpG-светофорами, т.е. такими CpG-парами генома, для которых дифференциальное метилирование ассоциировано с активностью близлежащих промоторов.

Публикация результатов:

По материалам диссертации опубликована **21** статья в рецензируемых международных журналах, **2** приглашенные главы-обзора, **2** статьи в российских журналах, **2** статьи в рецензируемых сборниках конференций. В **7** статьях автор диссертации выступает в качестве первого автора, в **9** статьях в качестве автора, ответственного за переписку. Результаты работы были представлены в виде устных (включая приглашенные и пленарные) и постерных докладов на множестве конференций, как международных, так и российских, в том числе «Биология – наука 21 века» (Пушино, 2017), BGRS (Новосибирск, 2016, 2012, 2010), SocBiN Bioinformatics (Москва, 2016), MCCMB (Москва, 2015, 2013, 2011), ISMB/ECCB (Дублин, 2015; Берлин, 2013; Вена, 2011), BIOSTEC BIOINFORMATICS (Барселона, 2013), POSTGENOME (Казань, 2012), ECCB (Базель, 2012; Гент, 2010), “Albany 2011: The 17th conversation” (Олбани, США), ESF FG&D (Дрезден, 2010).

Заключение:

Исследование выполнено на высоком научном уровне, полно опубликовано в высокорейтинговых журналах. Тема исследования полностью соответствует специальности «Математическая биология, биоинформатика» и профилю Совета.

Председатель комиссии:

д.б.н. Гельфанд М.С.

Члены комиссии:

д.б.н. Миронов А.А.

д.ф.-м.н. Ройтберг М.А.