



В.Г. Левицкий*, Д.Ю. Ощепков*,
Н.И. Ершов*, Л.О. Брызгалов*,
Е.В. Антонцева*, Г.В. Васильев*,
Т.И. Меркулова*

Анализ данных массовой иммунопреципитации хроматина с помощью новых методов распознавания сайтов связывания транскрипционных факторов FoxA

Транскрипционные факторы (ТФ) семейства FoxA играют важную роль в раннем эмбриогенезе, органогенезе и поддержании метаболического гомеостаза организма. Характерной особенностью сайтов связывания (СС) FoxA является их очень высокая вариабельность, в результате один из предложенных вариантов консенсусной последовательности этих сайтов VAWTRTGTKRTY способен описывать меньшую часть известных сайтов. Целью нашей работы являлись: 1) поиск олигонуклеотидных мотивов в вырожденном 15-буквенном коде, наиболее полно описывающих известные FoxA сайты (допускается описание одного сайта сразу несколькими мотивами); 2) построение на их основе оптимизированных динуклеотидных весовых матриц (PWM); верификация предсказанных сайтов с использованием данных по связыванию FoxA2 с хроматином клеток печени мыши, полученных методом иммунопреципитации в сочетании с массовым параллельным секвенированием (ChIP-Seq). Данные ChIP-Seq эксперимента представлены в виде: 1) профиля ChIP-Seq эксперимента, который содержит целые значения ≥ 1 и показывает, сколько раз происходило наложение секвенированных участков ДНК на соответствующий район геномной последовательности; 2) таблицы хромосомных локализаций 11475 районов геномной ДНК мыши (локусов), для которых максимальна высота пика в локусе больше либо равна 10.

Для разработки подхода к распознаванию FoxA сайтов использовали выборку из 37 нуклеотидных последовательностей экспериментально выявленных СС этого фактора длиной 65 п.н. из базы TRRD. В результате поиска мотивов методом на основе генетического алгоритма нами отобраны 3: YRTTTRYBYWDD, BNTSTTTKBHVBW и НТВТТТГВDBH. С использованием этих мотивов было составлено 3 выборки обучения для построения весовых матриц. Фланкирующие последовательности обеспечили дополнительный рост точности распознавания ССТФ FoxA. Для матриц были выбраны пороги, которые соответствуют 50 %-й ошибке недопредсказания, оцениваемой с помощью скользящего контроля (jack-knife). В отдельности мотивы были найдены в 21, 22 и 21 последовательностях исходной выборки, а 33 последовательности из 37 содержали, по крайней мере, один из указанных мотивов. Из данных, приведенных на рис. 1, а, видно, что одновременное использование всех трех методов позволяет идентифицировать от 55 до 63 % локусов в качестве содержащих FoxA сайты в составе ДНК. При этом подавляющее число предсказанных сайтов локализовано с отклонением менее чем в 200 нт от пиков высотой 10 (рис. 1, б). Полученные нами данные дают основание предполагать, что зафиксированное методом ChIP-seq связывание FoxA2 с

районами хроматина клеток печени мыши в подавляющем большинстве случаев произошло за счет взаимодействия этого белка непосредственно с ДНК. Таким образом, тщательное исследование обучающей выборки и выделение трех типов СС FoxA позволило построить реалистичный подход к распознаванию этих сайтов, пригодный для полногеномного анализа.

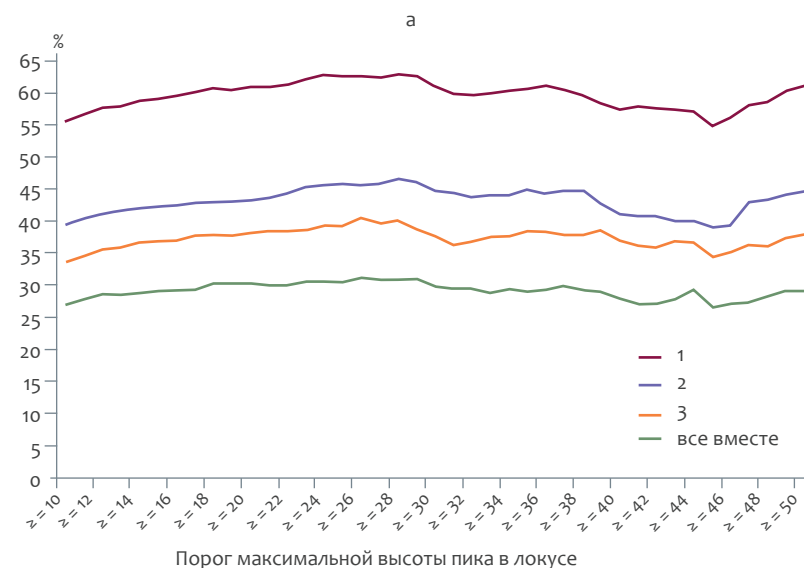


Рис. 1. Анализ выборок локусов, характеризуемых разной максимальной высотой пиков. Каждая точка графика соответствует расчетам для выборки, в которую входили только локусы с максимальной высотой пика не менее заданного значения (ось X). Ось Y – (а) – доля локусов, имеющих потенциальные сайты по отношению к числу всех локусов для заданной выборки; (б) доля локусов, имеющих потенциальные сайты с отклонением менее 200 нт от пика с высотой не менее 10.

