

# КОМПЬЮТЕРНЫЙ АНАЛИЗ КЛАСТЕРОВ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ В ГЕНОМЕ С ПОМОЩЬЮ МЕТОДА CHIP-SEQ

Дергилёв А.И., Орлов Ю.Л.<sup>1</sup>

Новосибирский государственный университет, Россия, 630090, г. Новосибирск, ул.  
Пирогова, д. 2, +7(913)0640743, arturd1993@yandex.ru

<sup>1</sup>Институт Цитологии и Генетики СО РАН, Россия, 630090, г. Новосибирск, пр. ак.  
Лаврентьева, д. 10, orlov@bionet.nsc.ru

Решаемая научная задача – исследование сайтов связывания транскрипционных факторов (ССТФ). ССТФ - это участки ДНК, связанные с белковыми факторами транскрипции, которые не всегда однозначно описываются нуклеотидной последовательностью. Вследствие чего встают задачи анализа полногеномных данных ChIP-seq, выявления координат сайтов ССТФ и сопоставления этой информации с геномной аннотацией (расположением генов, промоторных районов).

Рассмотрены статистические особенности распределения сайтов связывания транскрипционных факторов в геноме мыши, полученные с помощью ChIP-seq экспериментов в эмбриональных стволовых клетках. Определены кластеры сайтов, содержащие четыре и более сайта связывания различных транскрипционных факторов в геноме мыши, описано их расположение относительно регуляторных районов генов. Показано присутствие двух типов совместной локализации сайтов: кластеры, содержащие сайты связывания факторов Oct4, Nanog, Sox2, расположенные в дистальных районах, и кластеры, содержащих сайты связывания n-Мус, с-Мус, расположенные в основном в промоторных районах генов мыши. Анализ новых данных ChIP-seq по связыванию транскрипционных факторов Nr5a2, Tbx3 в том же типе клеток подтвердил разделение кластеров сайтов связывания транскрипционных факторов два типа – содержащие и не содержащие сайты связывания регуляторов плюрипотентности (Oct4, Nanog и др.). Разработана компьютерная программа статистической обработки данных расположения генов и доменов хроматина, анализирующая экспериментальные данные локализации сайтов, полученные методами ChIP-seq в геноме мыши и в геноме человека. С помощью этой программы выявлено наличие предпочтений в положении сайтов связывания транскрипционных факторов различных типов; рассчитаны расстояния между ближайшими сайтами связывания ТФ группы Oct4, Nanog, Sox2 и сайтами связывания ТФ n-Мус и с-Мус. Оценено присутствие нуклеотидных мотивов сайтов связывания транскрипционных факторов в выделенных участках ChIP-seq, уточнены нуклеотидные мотивы. Показана корреляция присутствия мотивов с интенсивностью связывания ChIP-seq. Разработаны компьютерные методы оценки кластеризации сайтов связывания различных транскрипционных факторов для новых данных ChIP-seq. Программы доступны по запросу к авторам.