Секция «Биоинформатика»

Разработка базы данных парных корреляций эпигенетических факторов

Научный руководитель – Миронов Андрей Александрович

Горбачева Дарья Викторовна

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова, Факультет биоинженерии и биоинформатики, Москва, Россия

E-mail: galvaniccell@gmail.com

Под эпигенетическими факторами можно понимать вещества и процессы, влияющие на активность генов, не меняя нуклеотидную последовательность ДНК. Они обуславливают огромное разнообразие структур и функций клеток в составе многоклеточного организма, поэтому их изучение важно для функцинальной геномики. Известно, что регуляция транскрипции осуществляется в результате совокупного эффекта многих эпигенетических факторов и что ДНК-связывающие белки взаимодействуют друг с другом для обеспечения точности регуляции экспрессии генов [1], поэтому для более глубокого понимания этой регуляции стоит изучать взаимодействия эпигенетических факторов.

В данной работе с помощью средстр Python и программы StereoGene (которая, рассматривая каждый из двух треков как функцию от координаты в геноме, рассчитывает коэффициент корреляции Пирсона между ними [2]) были рассчитаны парные корреляции всех ChIP-seq треков различных модификаций гистонов и сайтов связывания транскрипционных факторов для наиболее изученных клеточных линий K562 (157 различных факторов, 12246 парных корреляций), GM12878 (105 различных факторов, 5460 парных корреляций), HepG2 (87 различных факторов, 3741 парная корреляция). Данные получены из публичной базы данных ENCODE.

В результате при прямом подсчете большая часть корреляций оказалась положительной (средний коэффициент корреляции около 0,4). При подсчете с попытками вычесть из корреляций эффекты положения генов, нуклеосом, открытого хроматина этот эффект нивелируется, однако некоторые треки отдельных эпигенетических факторов все же остаются значимо скореллированными. Это можно рассматривать как подтверждение наличия комплексного взаимодействия эпигенетических факторов.

Источники и литература

1) Cheng C., Alexander R., Min R., Leng J., Yip K. Y., Rozowsky J., Yan K.K., Dong X., Djebali S., Ruan Y. et al. (2012). Understanding transcriptional regulation by integrative analysis of transcription factor binding data. // Genome Research 22, 9 (2012), 1658–1667.

2) Stavrovskaya E.D., Niranjan T., Fertig E.J., Wheelan S.J., Favorov A.V., Mironov A.A. StereoGene: rapid estimation of genome-wide correlation of continuous or interval feature data. // Bioinformatics 33, 20 (2017), 3158-3165.