

## Применение непрерывных НММ к разметке хроматина

Научный руководитель – Миронов Андрей Александрович

*Маргасюк Сергей Дмитриевич*

*Студент (специалист)*

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

*E-mail: smargasyuk@airmail.cc*

Посттрансляционная модификация гистонов – один из способов регуляции структуры хроматина. Функции этих модификаций можно упрощенно разделить на две крупные категории: поддержание глобальной структуры хроматина (доменов) и управление связанными с ДНК биологическими механизмами [2]. Изучение модификаций основано, в первую очередь, на анализе данных экспериментов ChIP-Seq. При том, что показана связь отдельных модификаций с участками генов (например, TSS) и биологически значимыми параметрами (например, уровнем экспрессии по данным mRNA-Seq) [4] с помощью анализа корреляции, для описания функционального эффекта совокупности меток (то есть разделения хроматина на функционально значимые классы) используются более сложные статистические подходы, такие как иерархическая кластеризация [3] и скрытая марковская модель [1]. Цель данной работы – улучшить существующие методы кластеризации хроматина, основанные на скрытой марковской модели, исходя из двух следующих соображений.

Во-первых, вычисление корреляции между данными ChIP-Seq для различных модификаций показывает положительную корреляцию между любыми двумя из них, что противоречит предположению об их различной функции. По нашей гипотезе, этот эффект вызван зависимостью экспериментально определяемого уровня всех модификаций от плотности нуклеосом, различной в разных геномных позициях. Для выделения плотности нуклеосом как отдельного показателя применен метод главных компонент: вместо вектора данных ChIP-Seq для каждой геномной позиции используется длина проекции вектора данных на первую главную компоненту и проекция вектора данных на ортогональное ей подпространство. Легко показать, что последовательность длин проекции является последовательностью такой, что сумма квадратов расстояний до исходных последовательностей, полученных из экспериментов, является минимальной; таким образом, полученная последовательность является хорошей оценкой для фактора, влияющего на все эксперименты. Обучение скрытой марковской модели на полученных данных показывает отличный от предыдущих работ характер связи модификаций.

### Источники и литература

- 1) Ernst, J., Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM // Nature Protocols volume 12, pages 2478–2492 (2017)
- 2) Kouzarides, T. Chromatin modifications and their function // Cell. 2007 Feb 23;128(4):693-705
- 3) Noureen, N., Qadir, M. A. et al. ChromClust: A semi-supervised chromatin clustering toolkit for mining histone modifications interplay // Genomics Volume 106, Issue 6, December 2015, Pages 355-359
- 4) Stavrovskaya, E. D., Mironov, A.A. et al. StereoGene: Rapid Estimation of Genomewide Correlation of Continuous or Interval Feature Data // bioRxiv 059584