

## Интерактивная визуализация предсказанных участков связывания факторов транскрипции

Научный руководитель – Кулаковский Иван Владимирович

*Никонов Михаил Андреевич*

*Студент (специалист)*

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

*E-mail: michael.nikonov.01@icloud.com*

Регуляция транскрипции генов у высших эукариот управляется специальными белками - факторами транскрипции, - распознающими характерные паттерны, мотивы в некодирующих регуляторных областях генома: промоторах и энхансерах. Классическая вычислительная модель мотива - позиционно-весовая матрица (ПВМ). ПВМ позволяет найти вхождения мотива в последовательности изучаемого регуляторного района и выдвинуть предположение о белках-регуляторах конкретного гена, либо уточнить положение непосредственно связываемого белком участка ДНК в протяженном регионе генома. Существует масса инструментов для использования ПВМ при сканировании последовательностей, как предназначенных для использования в командной строке, так и имеющих графический интерфейс, в том числе, доступный для использования через сеть интернет (например matrix-scan в пакете RSAT[1], FIMO в пакете MEME[2]). Однако существующие веб-инструменты построены по классической схеме: параметры сканирования задаются пользователем заранее, вычисления производятся на серверной стороне, а выдача результатов является во-многом статической.

Одной из популярных коллекций ПВМ для факторов транскрипции человека и мыши является HOCOMO[3]. HOCOMO включает в себя MoLoTool (Motif Locatoin Tool), новое веб-приложение для поиска вхождений ПВМ в последовательностях, использующий современные возможности веб-разработки для проведения вычислений на стороне клиента, что позволяет интерактивно визуализировать результаты, в том числе динамически изменять подсветку вхождений ПВМ в последовательности в зависимости от порогового значения оценки или статистической значимости вхождения. Помимо подсветки вхождения мотива генерируется таблица, содержащая информацию по каждому вхождению, которая также динамически перерисовывается при изменении порогового значения. Исходная версия MoLoTool хорошо демонстрирует успешность идеи, но содержит множество ошибок и недоработок.

В нашей работе мы создали новую версию MoLoTool. Ее функционал был увеличен, а также был проведен полный редизайн пользовательского интерфейса: инструмент теперь доступен для скачивания в виде самостоятельного веб-приложения, графические элементы были значительно улучшены и компактизованы в целях одновременной демонстрации большего количества информации; поиск моделей в коллекции теперь осуществляется по ID, гену, семейству и подсемейству; добавлена клавиатурная навигация в окне поиска; полностью переделана система выбора моделей для определения вхождений мотивов; упрощен процесс ввода и редактирования исходных последовательностей; добавлена демонстрация обратного лого; добавлена подсветка вхождений мотивов в последовательности по наведению на соответствующую строку в таблице данных; добавлена поддержка буквы 'N' во вводимой последовательности; исправлены ошибки прежней версии; добавлены иные

возможности в обновленной версии. MoLoTool 2.0 будет интегрирован в следующий релиз коллекции HOCOMOCO.

### Источники и литература

- 1) Nguyen NTT, Contreras-Moreira B et al. RSAT 2018: regulatory sequence analysis tools 20th anniversary // Nucleic Acids Research. 2018 Jul;46(W1):W209–W214
- 2) Charles EG, Timothy L et al. FIMO: scanning for occurrences of a given motif // Bioinformatics. 2011 Apr;27(7):1017–1018
- 3) Kulakovskiy IV, Vorontsov IE et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis // Nucleic Acids Research. 2018 Jan;46(D1):D252–D259