

## Домашнее задание по анализу структуры хроматина.

Задание предполагает освоение метода оценки сходства данных Hi-C на примере карт взаимодействий двух линий дрозофилы.

Под анализом структуры хроматина обычно понимают определение таких структур, как петли, ТАДы и компартменты, и их последующую интеграцию с другими типами омиксных данных. С примерами их поиска вы можете ознакомиться в ноутбуке [https://colab.research.google.com/drive/1vcxV9b\\_GD4oYD5LR5m5HTsF\\_1cfGUmr7?usp=sharing](https://colab.research.google.com/drive/1vcxV9b_GD4oYD5LR5m5HTsF_1cfGUmr7?usp=sharing)

Однако не менее важную роль имеет предварительный анализ данных Hi-C, включающий также кластеризацию реплик.

Что такое кластеризация реплик? Предположим, для некоторого типа омиксных данных у вас есть 2 биологических повторности для двух условий (эксперимент-контроль/ здоров-болен/ нейрон-астроцит и т.д. - несложно продолжить эту цепочку). Прежде чем исследовать различия в структуре хроматина (или любом другом типе омиксных данных) в этих двух условиях, важно показать, что образцы, относящиеся к одному и тому же условию, больше похожи друг на друга по сравнению с образцами, относящимися к другому условию. Иначе полученная в ходе анализа разница в структуре хроматина может быть расценена как случайный результат вследствие сильного разброса данных.

Не менее важным шагом при предварительном анализе данных Hi-C является построение графика скейлинга, который позволяет также оценить их качество и в грубом приближении обнаружить разницу в структуре хроматина.

Целью данной работы является

- 1) построение дендрограммы реплик Hi-C для клеточных линий дрозофилы bg3 (нервная ткань) и kс167 (эмбриональная линия) на основе коэффициента SCC
- 2) построение графика скейлинга для одного из образцов в логарифмических координатах (сделать корректные подписи, включая единицы измерений)
- 3) интерпретация графиков, а также комментарии к ячейкам, начинающимся с ###!!.

Более подробное описание можно найти в ноутбуке [https://colab.research.google.com/drive/1IsPiDVskF7DZ3C-MKn3gFXLeBYsVqE3t?usp=share\\_link](https://colab.research.google.com/drive/1IsPiDVskF7DZ3C-MKn3gFXLeBYsVqE3t?usp=share_link)

Дедлайн - 10 мая 10:00.

На сайте github.com нужно создать репозиторий и привести ссылку на этот репозиторий в общей гугл-таблице (вкладка HW4)

[https://docs.google.com/spreadsheets/d/10r73G68KiXa-7Kf\\_u1Nir1clTd-3xy1NbBX7\\_kGk0A/edit#gid=0](https://docs.google.com/spreadsheets/d/10r73G68KiXa-7Kf_u1Nir1clTd-3xy1NbBX7_kGk0A/edit#gid=0)

Результат может быть представлен в виде html или pdf ноутбука с комментариями и графиками.

Данные и презентации можно найти тут: <https://drive.google.com/drive/folders/142vwyCNlcN3tfBeelfohtKnqmjzzVjcD?usp=sharing>