## **Статья:**

Marhabaie, M., Wharton, T. H., Kim, S. Y. & Wharton, R. P. Widespread regulation of the maternal transcriptome by Nanos in Drosophila. *PLOS Biology* vol. 22 e3002840 (2024). <https://doi.org/10.1371/journal.pbio.3002840>

## **Данные**

В качестве исходных данных вы имеете данные секвенирования РНК (bulk RNA-seq) эмбрионов дрозофилы. Вам доступны данные RNA-seq нескольких групп образцов:

* Эмбрионы 9 стадии с разными генотипами (cycle9):
  + дикий тип - w,
  + экспрессирующие химеру Upf1-Nos,
  + экспрессирующие химеру Upf1-NosL7 (мутантный Nos),
  + экспрессирующие Upf1,
  + экспрессирующие Nos,
  + экспрессирующие NosL7 (мутантный Nos)
* Эмбрионы 0-1 и 2-3 часов с разными генотипами (hours):
  + дикий тип – w,
  + делеция Nos – nos.

Данные предобработаны, поэтому вы имеете две таблицы с «сырыми» экспрессиями генов в каждом из образцов: ***07\_marhabaie\_2024\_cycle9\_genes\_counts.tsv*** и ***07\_marhabaie\_2024\_hours\_genes\_counts.tsv***. В таблицах в строках записаны гены (первый столбец – идентификаторы генов по БД Flybase), в столбцах - образцы (имена столбцов – названия образцов с некоторым суффиксом).

Также у вас есть значения TPM для генов (файлы ***07\_marhabaie\_2024\_cycle9\_genes\_TPM.tsv*** и ***07\_marhabaie\_2024\_hours\_genes\_TPM.tsv***), а также значения экспрессии для каждого транскрипта в «сырых» каунтах и TPM (файлы ***07\_marhabaie\_2024\_cycle9\_tx\_counts.tsv****,* ***07\_marhabaie\_2024\_hours\_tx\_counts.tsv****,* ***07\_marhabaie\_2024\_cycle9\_tx\_TPM.tsv*** и ***07\_marhabaie\_2024\_hours\_tx\_TPM.tsv***).

Описания образцов (к какой группе принадлежит образец) содержатся в файле ***07\_marhabaie\_2024\_metadata.csv***.

## **Вопросы к данным**

Вашу групповую работу вы должны разделить на 4 примерно равноценные части. Обязательными элементами защиты проекта являются краткий рассказ о статье и соответственно природе ваших данных, разведывательный анализ данных (изучение качества данных) и выводы.

Знакомясь со статьей, обратите внимание на то, как были полученные данные RNA-seq: как готовили образцы, какой протокол использовали, как избавлялись от рРНК, на чем секвенировали и т.д.

Вы можете задавать к вашим данным любые интересующие вас вопросы. Например, вы можете сравнить экспрессию генов между разными группами образцов. Вы можете взять в анализ все образцы или сосредоточиться только на интересных вам группах образцов. Получив списки дифф экспрессируемых генов, проведите функциональную аннотацию генов.

Также будет здорово, если вы визуализируете распределение дифф экспрессируемых генов по геному и проверите, не наблюдается кластеризации генов в конкретных локусах генома.