

This folder contains data files related to the Huesken dataset and its modeling

ACGUT_bcuts

- The calculated BCUT values for all five natural nucleotides

Huesken_ID-21ntSeq-NormAct_2431

- ID, potency and siRNA sequence

Huesken-train-test-values-9-27-2022

- ID, potency, 252 BCUT values and predicted potencies for the training set, shown in the last 2153 rows, and test set shown in the first 278 rows of this spreadsheet.

rational_split.cpp

- The CPP code for conducting splitting