Supplemental File S6 - Scripts for data analysis

**Global fitness landscapes of the Shine-Dalgarno sequence**

**Syue-Ting Kuo#, Yuan-Ju Cheng, Ruey-Lin Jahn, Yi-Lan Chen, Yun-Ju Lee, Florian Hollfelder, Jin-Der Wen, Hsin-Hung David Chou\***

**#**Scripts organizer. Email: [d08b21005@ntu.edu.tw](mailto:d08b21005@ntu.edu.tw)

\*Corresponding author. Email: [chouhh@ntu.edu.tw](mailto:chouhh@ntu.edu.tw)

Scripts are organized into two major function categories:

1. Sort-seq data processing
2. Data analysis and visualization

**Category I. Sort-seq data processing**

Sequence hard-matching by the Linux grep program removes irrelevant or indel-containing reads from raw sequencing data. In-house Python scripts extract the 9-nt SD sequence from each read and count the occurrence of each genotype in each rank.

Three shell scripts and three python scripts (XXX indicates the library and experimental replicate):

script\_XXX.sh

ReadCountTable\_XXX.py

Input:

Raw sequencing data is available at NCBI (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA516114>, Accession: PRJNA516114).

Output:

One read count table for each library replicate.

**Category II. Data analysis and visualization**

Scripts for data analysis and plotting are listed in the following table. They are organized into six subsets/folders in terms of figure presentation. The function of each script corresponds to the description in the figure legend. README.docx files provide detailed information for script implementation.

|  |  |
| --- | --- |
| **Folder** | **File content** |
| Fig\_1 | README.docx  fepb\_RNAfold.csv  arti\_RNAfold.csv  dmsc\_RNAfold.csv  RNAsubopt\_dG.csv  Fig1B\_LibraryDistribution.py  Fig1B\_Fold\_analysis.py  Fig1B\_Fold\_plot.R  Fig1B\_SDaSD\_analysis.py  Fig1B\_SDaSD\_plot.R  Fig1C\_LibraryCorrelation.py  util.py |
| Fig\_2 | README.docx  Fig2A\_analysis.py  Fig2A\_plot.py  Fig2B\_analysis.py  Fig2B\_plot.py  util.py |
| Fig\_3 | README.docx  Fig3A\_NucleotideComposition.py  Fig3B\_NucleotideContent.R  Fig3C\_ConsensusMotif.py  util.py |
| Fig\_5 | README.docx  fepb\_SDR\_union\_count25\_e1.csv  fepb\_SDR\_union\_count25\_e2.csv  fepb\_SDR\_union\_count25\_o1.csv  fepb\_SDR\_union\_count25\_o2.csv  arti\_SDR\_union\_count25\_e1.csv  arti\_SDR\_union\_count25\_e2.csv  arti\_SDR\_union\_count25\_o1.csv  arti\_SDR\_union\_count25\_o1.csv  dmsc\_SDR\_union\_count25\_e1.csv  dmsc\_SDR\_union\_count25\_e2.csv  dmsc\_SDR\_union\_count25\_o1.csv  dmsc\_SDR\_union\_count25\_o2.csv  Fig5A\_SingleNucleotide.R  Fig5C\_PairwiseEpistasis.R  Fig5E\_LibraryCorrelation.py  util.py |
| Fig\_6 | README.docx  Fig6\_MutationEffect.R |
| Fig\_Supplemental | README.docx  FigS7A\_analysis.py  FigS7A\_plot.R  FigS7B\_analysis.py  FigS7B\_plot.R  FigS7C\_LibraryCorrelation3D.py  FigS14.py  FigS16-27\_MutationEffectTotal.R  util.py |