dms2dfe: Comprehensive Workflow for Analysis of Deep Mutational Scanning Data

Rohan Dandage[1] and Kausik Chakraborty[1]

[1]CSIR-Institute of Genomics and Integrative Biology, New Delhi, India.

04 July 2017

Summary

`dms2dfe` is a python package that integrates steps in the analysis of Deep Mutational Scanning (Fowler, Stephany, and Fields 2014) data. Using this end-to-end workflow, users can implement various processing methods and downstream applications in the deep sequencing based bulk mutational scanning experiments.

Recently, owing to evolution of sequencing and phenotyping technologies, large scale genotype to phenotype data is increasingly being generated. Along this line of research, Deep Mutational Scanning method allows comprehensive assessment of all the substitutions of a given gene. In the analysis of Deep Mutational Scanning data, `dms2dfe` allows end-to-end workflow that addresses issue of noise control and offers variety of crucial downstream analyses. In downstream analyses, `dms2dfe` workflow provides estimation of distributions of effect sizes, identification of potential molecular constraints and generation of data-rich visualizations (Dandage and Chakraborty 2016).

As an input for the workflow, deep sequencing data (whether unaligned or aligned) or list of genotypic varants can be provided. For data structure, `dms2dfe` uses DataFrames from robust Pandas library (McKinney 2010).

Source code and issue tracker is available in `dms2dfe`'s GitHub repository.[1] Documentarion and API[2] are generated using Sphinx.[3]

[1]https://github.com/kc-lab/dms2dfe
[2]https://kc-lab.github.io/dms2dfe
[3]http://www.sphinx-doc.org

References

Dandage, Rohan, and Kausik Chakraborty. 2016. "dms2dfe : Comprehensive Workflow for Analysis of Deep Mutational Scanning Data." *BioRxiv*, 072645. doi:10.1101/072645.

Fowler, Douglas M, Jason J Stephany, and Stanley Fields. 2014. "Measuring the activity of protein variants on a large scale using deep mutational scanning." *Nature Protocols* 9 (9). Nature Publishing Group: 2267–84. doi:10.1038/nprot.2014.153.

McKinney, Wes. 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman, 51–56.