

PEA-m6A User Manual (version 1.0)

- PEA-m6A is an ensemble learning framework for predicting m6A modifications at regional-scale.
- PEA-m6A consists of four modules: **Sample Preparation, Feature Encoding, Model Development and Model Assessment**, each of which contains a comprehensive collection of functions with pre-specified parameters available.
- PEA-m6A was powered with an advanced packaging technology, which enables compatibility and portability.
- PEA-m6A project is hosted on <http://github.com/cma2015/PEA-m6A>
- PEA-m6A docker image is available at <http://hub.docker.com/r/malab/peam6a>
- PEA-m6A server can be accessed via <http://peam6a.omstudio.cloud>

PEA-m6A Features Encoding

This module transfers the sample sequences into a numerical feature matrix, using different encoding strategies.

Functions	Description	Input	Output	Reference
Pretrain WeakRM*	A weakly supervised learning framework, WeakRM*; it is a modified version of WeakRM(Huang et al., 2021) that can be used to capture deep learning-driven features.	Positive and negative samples in FASTA format	A deep learning-driven features extractor	In-house scripts
Feature Matrix Generation	Characterize each sample with numeric features.	Sequences in FASTA format	Feature matrix separated by ','	In-house scripts

Pretrain WeakRM*

This function can be used to pre-train a **WeakRM*** model to capture motifs from long sequences and learn local dependencies. Using a gated attention mechanism, the output of the penultimate layer in **WeakRM*** is extracted as deep learning-driven features.

Input

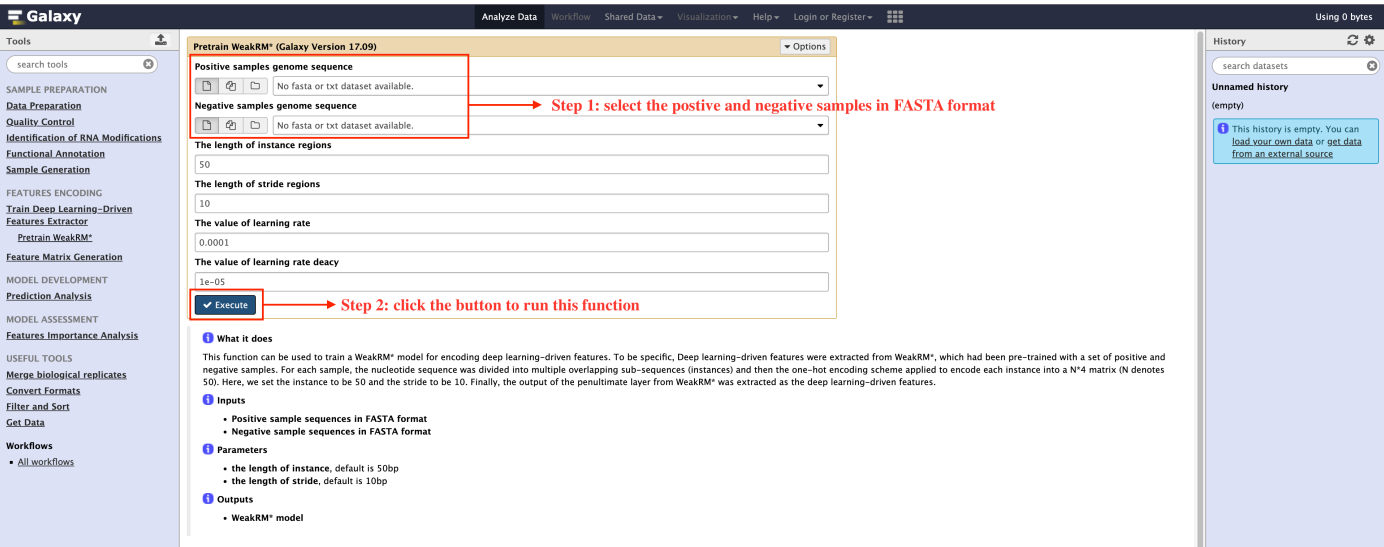
- **Positive samples genome sequence:** sequences in FASTA format
- **Negative samples genome sequence:** sequences. in FASTA format

Output

- A **WeakRM*** model in binary format

How to use this function

- The following screenshot to run this function



Feature Matrix Generation

This function can be used to encode RNA modifications regions sequences into a feature matrix. To be specific, **Statistic-based features** integrated several commonly used feature encoding strategies including **Nucleic acid composition related features** and **Pseudo nucleotide composition**. **Deep learning-driven features** were generated by pre-trained **WeakRM*** models.

Input

- **RNA modification in FASAT format:** which can be generated by function **Sample Generation**

Output

- Feature matrix with rows indicating samples, columns representing feautres

