# A deep learning model to predict Epstein-Barr virus (EBV) integration sites.

Prerequisites：

1. Pyhton（3.7）;
2. Tensorflow（1.13.1）(An open source software library for numerical computation of high performance, <https://pypi.org/project/tensorflow/>);
3. Keras（2.2.4）(An Application Programming Interface, API, for deep learning in Python)（the Python deep learning API <https://keras.io/>）;
4. Scikit-learn（0.24）（An free software library for machine learning in Python, <https://scikit-learn.org/stable/>）;
5. PyCharm（A professional Python Integrated Development Environment, IDE <https://www.jetbrains.com/pycharm/>）;
6. Anaconda（1.9.7）(An open source Python distribution， <https://www.anaconda.com/>);
7. CUDA（10.0.130）（A computing platform launched by graphics card manufacturer NVIDIA <https://developer.nvidia.com/cuda-toolkit-archive>）;
8. Cudnn（7.0）（A GPU accelerating library for deep neural networks <https://developer.nvidia.com/rdp/cudnn-archive>）.

Among them, Tensorflow is equipped with a GPU version, which can be used for accelerated calculations with CUDA and Cudnn.（See for details <https://docs.nvidia.com/cuda/cuda-toolkit-release-notes/index.html>）

Most of the above can be installed by pip with version number, e.g.

pip install tensorflow==1.13.1

File Description:

1. EBVData: The folder to store the data, with two files, dsVIS\_Data and VISDB\_Data.

dsVIS\_Data:

There are the positive data dsVIS\_pos\_Data.mat with sizes of 2663×2000 and the negative data dsVIS\_neg\_Data.mat with sizes of 26630×2000, each row of which is one DNA sequence.

VISDB\_Data:

There are the positive data VISDB\_pos\_Data.mat with sizes of 1104×2000 and the negative data VISDB\_neg\_Data.mat with sizes of 11040×2000, each row of which is one DNA sequence.

1. Model: The folder to store the trained DeepEBV model.
2. Test\_result: The folder to store test results of the model.
3. EBVDataProcessing.py: This is a program used for data processing, which encodes the data to one-hot code.
4. DeepEBV.py: This is the main program, including the creation, loading and testing of the model.

Run:

1. Run EBVDataProcessing.py first for data encoding. Four files (dsVIS\_Test\_Data.npy, dsVIS\_Test\_Label.npy, VISDB\_Test\_Data.npy and VISDB\_Test\_Label.npy) will be generated and stored in the folder EBVData.
2. Then run DeepEBV.py The test results will be generated and stored in the folder test\_ Results.

If you have any questions, please contact me.

Email: liangjiuxing@m.scnu.edu.cn