**ESOCVD V1.0**

ESOCVD is a matlab-based computational framework for estimating circadian phase of samples from untimed transcriptome dataset.

The main function of ESOCVD is the file “MyESOCVD.m”. The output of this model is the optimal of event sequence **S** and its corresponding likelihood value. The model outcome also includes matrix pv, which represent the probability of one subject belongs to each stage.

**Normalization**

A SVD matrix is generated from the original gene expression profile with seed genes. The rows represent samples and the columns represent eigengenes. In our paper, we used 12 eigengenes. After obtained the SVD matrix, normalization will be implemented. We normalized each patient sample by subtracting the mean we calculated in the controls and dividing by the standard deviation we calculated in the controls. Finally, our controls have a mean of 0 and a standard deviation of 1, the patients won’t have a mean of 0 or standard deviation of 1, because they are z-scored relative to the controls and they should have some disease signal.

**Model replicate**

To against the local minimum of model fitting, ESOCVD model can be repeat 200 times on each dataset with parallel computing.

**Circadian phase estimation**:

After obtaining the optimal sequence **S**, we got a matrix , which represents the probability of each subject at each stage. Run the file Test.m, circadian phase of all the samples will be estimated.

**Related paper:**

A novel bioinformatics approach to reveal oscillatory patterns of gene expression, 2020, bioxiv.