Inferring gene regulatory network via fusing gene expression image and RNA-seq data

Tuesday, January 18, 2022   3:30 PM

Abstract

Motivation: Recently, with the development of high-throughput experimental technology, reconstruction of gene regulatory network (GRN) has ushered in new opportunities and challenges. Some previous methods mainly extract gene expression information based on RNA-seq data, but the associated information is very limited. With the establishment of gene expression image database, it is possible to infer GRN from image data with rich spatial information.

Results: Firstly, we propose a new convolutional neural network (called SDINet), which can extract gene expression information from images and identify the interaction between genes. SDINet can obtain the detailed information and high-level semantic information from the images well. And it can achieve satisfying performance on image data(Acc: 0.7196, F1: 0.7374). Secondly, we apply the idea of our SDINet to build an RNA-model, which also achieves good results on RNA-seq data(Acc: 0.8962, F1: 0.8950). Finally, we combine image data and RNA-seq data, and design a new fusion network to explore the potential relationship between them. Experiments show that our proposed network fusing two modalities can obtain satisfying performance(Acc: 0.9116, F1: 0.9118) than any single data.

Conclusion: We propose a new network that can better extract gene expression information from image data. Besides, we combine image data and RNA-seq data to infer GRN and design a new fusion network to learn the joint features of these two data types for the first time.

Availability: Data and code are available from `https://github.com/guofei-tju/Combine-Gene-Expression-images-and-RNA-seq-data-For-infering-GRN`.

Contact: guofei@csu.edu.cn

Problem

Predict regulatory relationship between gene pairs

Claimed Novelty:

- Fusing RNA-seq and gene expression image for regulatory relation prediction
- A CNN based deep neural network model, accordingly
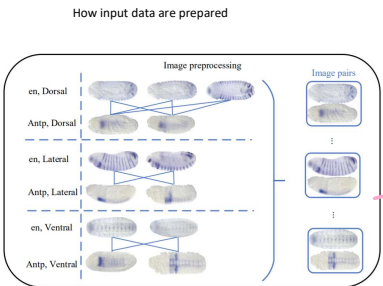

How input data are prepared

Fig. 2: The method of preprocessing for image data. We take Antp as target gene and en as TF as an example. These ISH images come from three different directions: lateral, ventral and dorsal. We concatenate these images from the same direction to form image pairs.
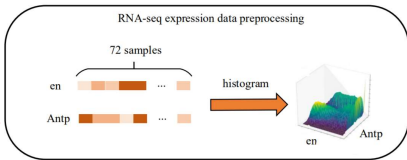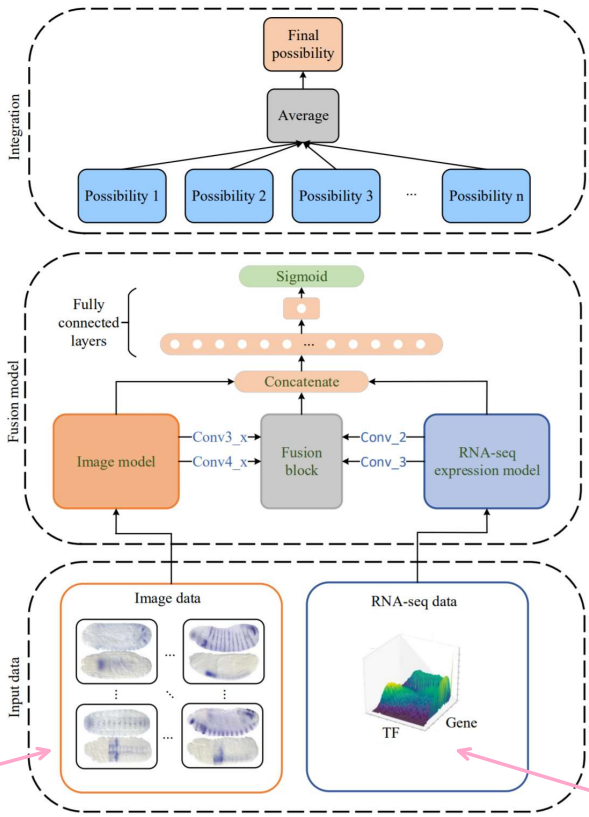

Fig. 4: The method of preprocessing for RNA-seq data. Antp is target gene and en is TF, containing RNA-seq expression information from 72 samples. We calculate two-dimensional histograms as input for our model.