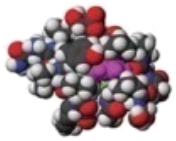
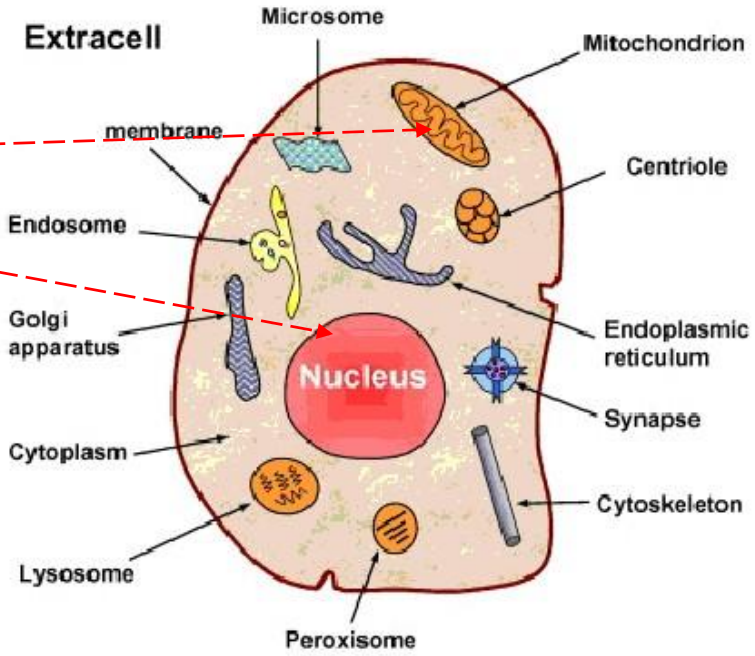
Course Project: Prediction of Protein subcellular localization based on microscopic images

- Human cells are organized into different biochemical cellular compartments. The knowledge of the subcellular localization of those proteins can provide important clues for understanding their functions.



Protein

OPA1



Shen and Chou, *Biochem Biophys Res Commun*, 2007

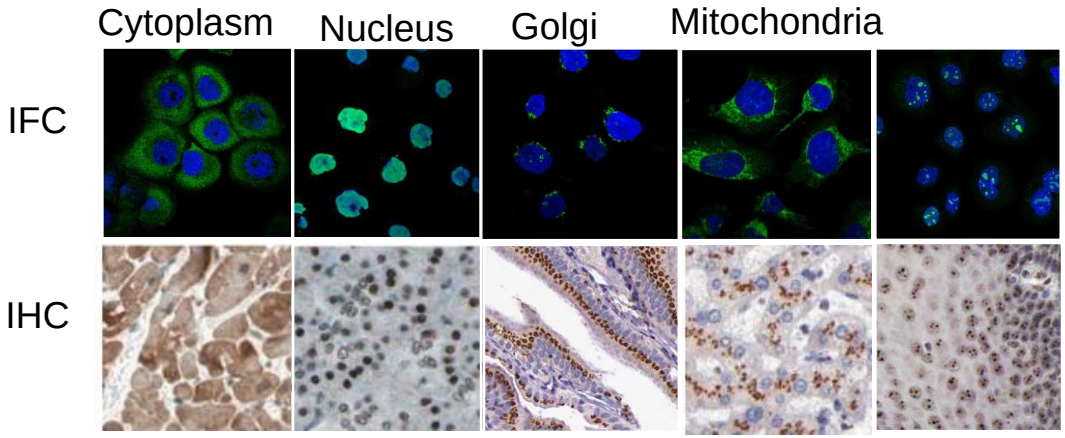
Research Background

Biological expert observation

1D Sequence-based Prediction

```
QIPQEFVKLQVSQEEFLCMKVLLLLNTIPLEGLRSQTQFEEMRSSYIRELIKAIGLRQKG
VVSSSQRFYQLTKLLDNLHDLVKQLHLYCLNTFIQSRALSVEFPEMMSEVIAAQLPKILA
GMVKPLLFH
VINTFDGVADYLQTYHKLPDNYITKSEAQALGWVASKGNLADVAPGKSIGGDIFSNREGK
LPGKSGRTWREADINYTSGFRNSDRILYSSDWLIYKTTDHYQTFTKIR
DIVLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYYTTTLADGVPS
RFSGSGSGTQYSLKINSLQPEDFGSYYCQHFWSTPRTFGGGTKLEIK
TGDNSNVVTMIRAGSYPKVNPTPTWVRAIPFEVSVQSGIAFKVPVGSLFSANFRTDSFTS
VTVMSVRAWTQLTPPVNEYSFVRLKPLFKTGDSTEEFEGRASNINTRASVGYRIPTNLRQ
NTVAADNVCEVRSNCRQVALVISCCFN
```

2D Image-based Prediction



Dataset

The human protein atlas (HPA, http://proteinatlas.org) stores millions of microscopic images of immunohistochemistry (IHC) and immunofluorescence (IF) showing the spatial distribution of proteins in cells.

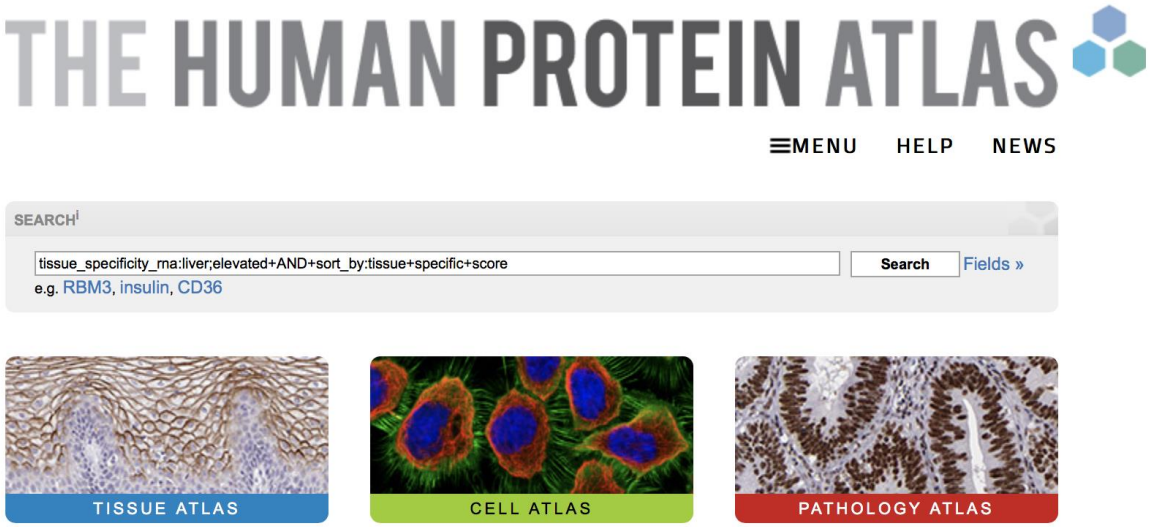
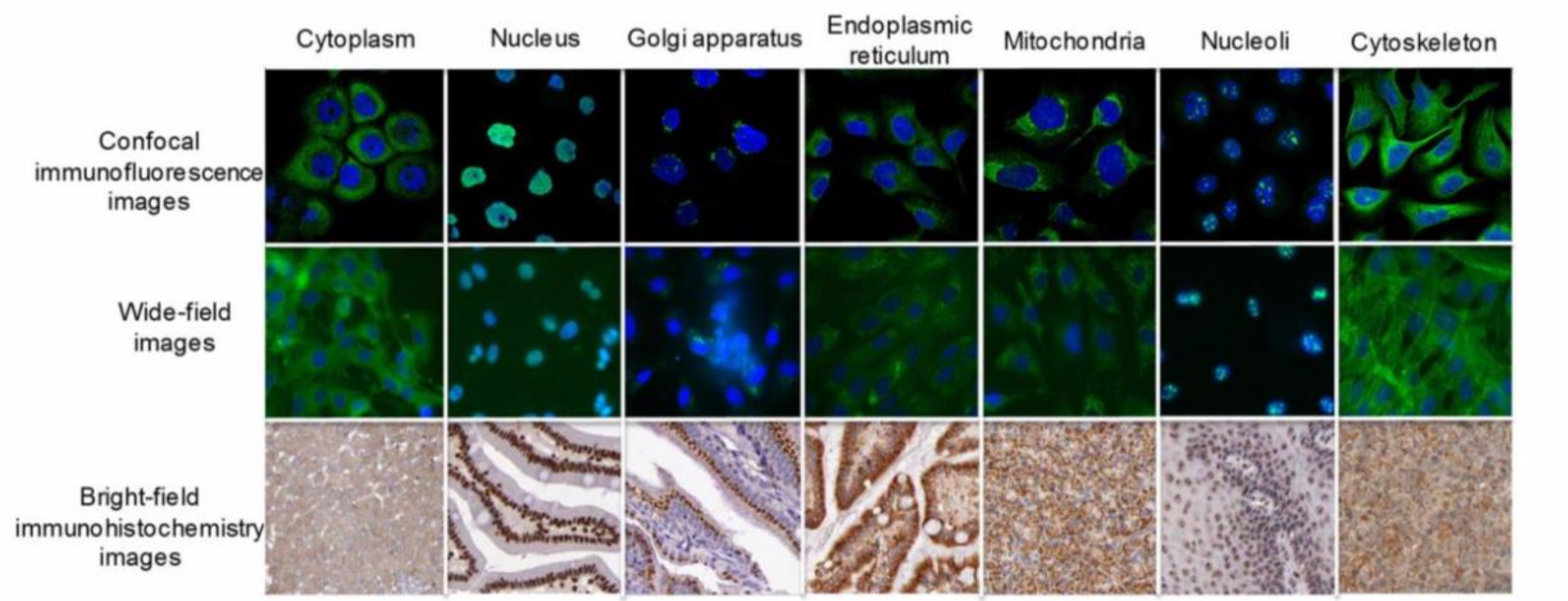


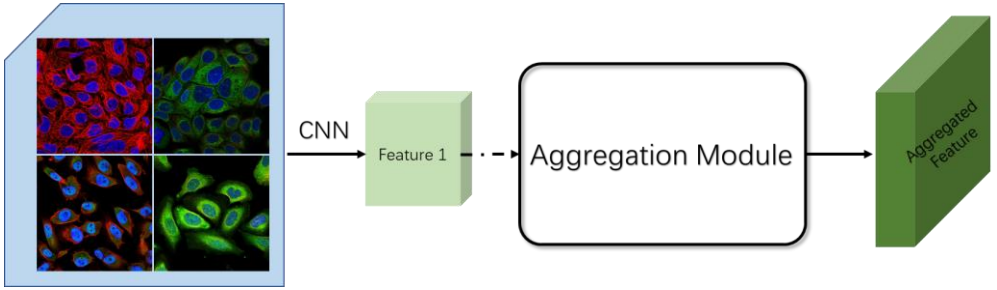
Image-based prediction of protein subcellular localization



七种主要亚细胞位置结构的不同类型图像的图例

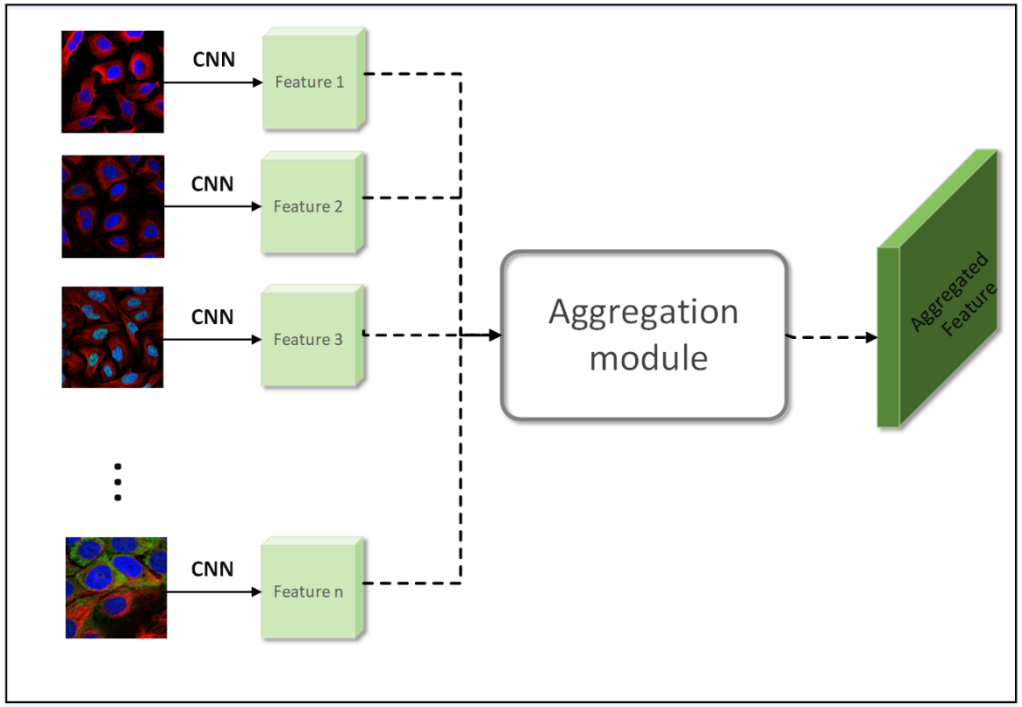
Multi-instance Deep Learning Models

Pre-aggregation



VS

Post-aggregation



Task

- 1. Build a predictor for protein subcellular localization using the given training data (work in team)
- 2. Experiment with Huawei Atlas 200 High performance Computation Platform
  - The teaching Assistants will give demonstration of using GPUs on Huawei Atlas Implementing and also provide Pre-trained model to students so they can import the model and perform inference on the test data.