# 选题三

### 高通量显微镜图像中的蛋白质亚细胞定位评估

### 背景介绍

绿色荧光蛋白GFP是实验室常用的蛋白标记物，通过显微镜观察荧光染色物的位置可以获知被标记蛋白的亚细胞定位。更多信息[..](https://wenku.baidu.com/view/b67b896202768e9951e73882.html)

### 数据来源

数据下载链接：<https://jbox.sjtu.edu.cn/l/SHwJtx>

每张显微图中仅有一个细胞，图中荧光蛋白的亚细胞定位标签从0~11分别代表：

|  |  |
| --- | --- |
| **label** | **label\_idx** |
| cell periphery | 0 |
| cytoplasm | 1 |
| endosome | 2 |
| er | 3 |
| golgi | 4 |
| mitochondrion | 5 |
| nuclear periphery | 6 |
| nucleolus | 7 |
| nucleus | 8 |
| peroxisome | 9 |
| spindle pole | 10 |
| vacuole | 11 |

### 数据介绍

该数据集有两个通道(RGB颜色模式下B通道为0)，红色荧光蛋白(mCherry)与胞质定位从而标记细胞轮廓，绿色荧光蛋白(GFP)在3’端标记内源性基因，表征蛋白的丰度以及蛋白的亚细胞定位。整个数据集包含65000个训练图像，12500个验证图像，以及12500个测试集。

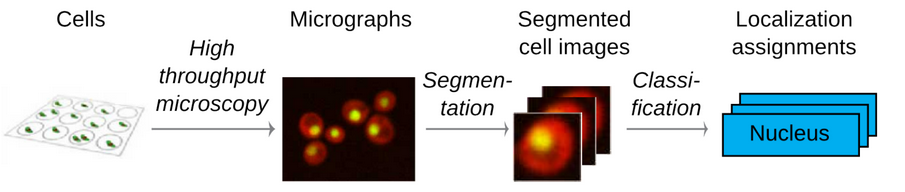


图4 来自文献的亚细胞定位研究方法

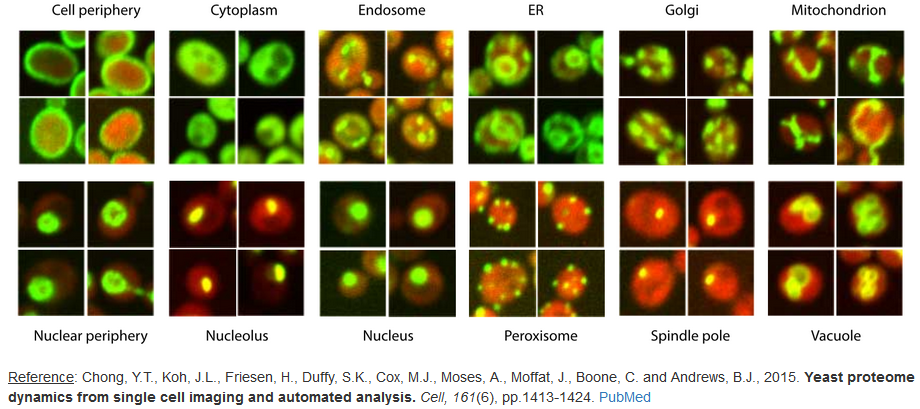


图5 不同类型的蛋白的亚细胞定位示例

### 探索研究以下问题：

1. 建立分类模型，预测GFP的亚细胞定位
2. 考虑一下纳入模型的训练集如果只用一个通道的数据与用2个通道的数据有什么区别?
3. 思考一下，你可以用红色荧光覆盖的区域来估计细胞的大小吗？（可选）