**Iterative method for identification of essential proteins based subcellular localization, orthology and PPI networks**

**A brief introduction of SON**

SON是一种基于融合亚细胞定位信息、蛋白质同源信息及PPI拓扑特征的关键蛋白质识别算法，在关键蛋白质的识别上具有十分良好的性能，也是目前的state-of-the-art method，其构成如下：

其中NOS为蛋白质同源信息测度、NSL为亚细胞定位信息测度、NNC为PPI拓扑特征测度。

**NNC**

NNC为NC的归一化形式：

**NOS**

NOS为蛋白质同源属性，

其中OS(i)为蛋白质i的同源物种的数目（同源属性是指不同物种的蛋白质具有相似的序列或结构， 例如人的蛋白质与酵母菌的蛋白质有一部分属于同源蛋白质，具有相似的功能，而该部分蛋白质则可能具有关键性，参与同源的物种越多，越倾向于为关键蛋白质）。

**NSL**

关键蛋白质倾向于出现在某些亚细胞位置中，所以出现在这些位置的蛋白质可能为关键蛋白质，SON在计算时对每个亚细胞位置进行评分，由于一个蛋白质可能出现在多个亚细胞位置中（如转录蛋白等），因此蛋白质的亚细胞位置得分为该蛋白的所有位置得分的总和，

其中为蛋白质的位置集合，为亚细胞位置得分。

而LCC的计算中采取的方法则是按照计算该位置出现的关键蛋白质的个数来评价该位置得分，该位置出现的关键蛋白质越多，得分越高，反之越低。

SON的计算方法为实现通过NNC对所有蛋白质节点进行排序，然后取排名在前5%的蛋白质为关键蛋白质，倒数5%的作为非关键蛋白质，并统计这些蛋白质中的各个位置出现的频数与，

**Our method**

SON的问题在于在计算NSL时只用了NC即PPI拓扑特征对蛋白质的关键性进行判别而PPI网络存在高假阳性，通过PPI网络得到的结果并不理想，而运用此方法对计算亚细胞位置得分也不准确，因此我们运用如下迭代方法计算：

1. Compute NNC, NOS first
3. **for** i = 1:itr
5. **if** i == 1
6. use NNC to compute NSL, compute pri = (1- α)\* NOS (i)+  α \*((1-β )\* NSL (i)+  β\* NNC (i))
7. **else**
8. compute pri
9. [~,I] = sort(pri,'descend');
10. use I to compute NSL\_i
11. end
13. use NSL\_itr to compute pri
15. end

由于我们并未收集到Orthologous proteins dataset，因此我们只比较SON算法的前半部分

只迭代运行一次，用来计算NSL的可靠排序结果来自另一篇报告SGC，为SON设定参数，为方便表示我们将本文算法命名为ISON（Iterative SON）.