

A new algorithm with adaptive parameter for identification of essential proteins based on subcellular localization, gene expression and protein complex

Our method

我们运用三种生物信息来识别关键蛋白质，分别是基因表达数据、蛋白质模块及亚细胞定位信息，其构成如下：

$$SGC(i) = ((PC(i) + NSL_INNEC(i)) * a + PEC(i) * (1 - a)) * b + NNC(i) * (1 - b);$$

其中 PEC 为基因共表达测度、NSL_INNEC 为亚细胞定位信息测度、NNC 为 PPI 拓扑特征测度、PC 为蛋白质模块测度。

NNC

NNC 为 NC 的归一化形式：

$$NNC = \frac{NC}{Max_NC}$$

NSL_INNEC

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，

$$SL(i) = \sum_{j \in N(i)} LCC(j)$$

其中N(i)为蛋白质的位置集合，LCC(j)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fl与hl，

$$LCC(l) = \begin{cases} 1 - \frac{fl}{hl}, & fl < hl \\ \frac{fl}{hl} - 1, & fl \geq hl \end{cases}$$

计算完 NSL 后我们将 NSL 与 INNEC 进行融合

$$NSL_INNEC(i) = NSL(i) * INNEC(i)$$

PC

$$PC(i) = \frac{pc(i)}{\max_pc}$$

$pc(i)$ 为包含蛋白质 i 的蛋白质模块数，即一个蛋白质出现在的蛋白质模块越多，越关键。

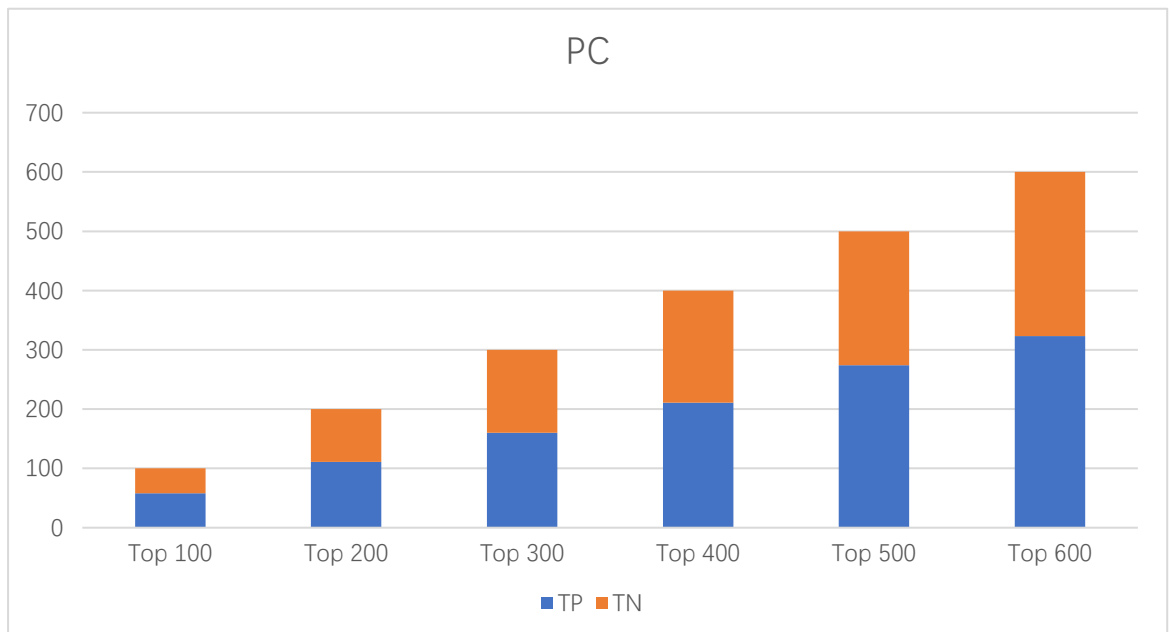
PEC

PEC 为一种融合基因表达数据与蛋白质边聚系数的测度，能有效利用基因表达数据来识别关键蛋白质。

为方便表示我们的将我们的算法命名为 SGC，其构成如下：

$$SGC(i) = \underbrace{\left(\underbrace{(PC(i) + NSL_INNEC(i))}_{\text{Node based information}} * a + \underbrace{PEC(i) * (1 - a)}_{\text{Edge based information}} \right)}_{\text{Biological information}} * b + \underbrace{NNC(i) * (1 - b)}_{\text{PPI information}}$$

大多数方法在组合各信息时都采用参数调节的方法，设置一个最优组合参数，而我们的算法中依据组合信息的不同提出了一种参数自适应方法，思路如下各信息在识别关键蛋白质中的适应域不一样，以 PC 为例我们进行如下说明：



可以看到 PC 在 Top100 的排名中作用更为显著，一般来说，生物信息的关键蛋白质排序作用在排名较前列的关键蛋白质时更可信。

所以在排序过程中权重依据所需的关键蛋白质数目进行自适应调整，参数自适应模型为

$$a = \alpha_1 * \text{input} + \beta_1$$

$$b = \alpha_2 * \text{input} + \beta_2$$

其中 input 为期望得到的关键蛋白质数目，本文中取 $\alpha_1 = 0.49$ ， $\beta_1 = -0.0005$ ； $\alpha_2 = 1$ ， $\beta_2 = -0.0003$ ，即计算排名靠前的关键蛋白质时生物信息的权重更大，基于节点的生物信息权重也更大，随着 input 的增大 PPI 拓扑特征的权重逐渐增大，基于边的生物信息的权重也逐渐增大。

实验结果如下，对比数据来自于 OGN (PLos One 2018, Predicting essential proteins by integrating orthology, gene expressions, and PPI networks)

