

蛋白质互作网络的可控性

小组成员： 秦聪
汪贤猛



OUTLINE

- 一、研究背景
- 二、研究思路
- 三、具体内容
- 四、结论
- 五、评价及思考



一、研究背景

近十多年来，有关蛋白质互作网络的研究成为热点。

2000年，Uetz等将酵母的5 341个ORF作为诱饵和6000个ORF作为猎物，得到817个蛋白质之间的691对相互作用。

(Nature, 2000, 403 (6770) :623 -627.)

2005年，Stelzl等首次采用Y2H法大规模研究人类蛋白质相互作用，从1 705种蛋白质中发现3 186对蛋白质相互作用，第一次构建了较为全面的人类蛋白质相互作用网络。

(Cell, 2005, 122 (6) :957 -968.)



其后，有关蛋白质互作网络的信息以及对于蛋白质互作网络结构功能的研究越来越多。

2011年，Liu等介绍了一种预测网络控制节点的最大匹配方法（Maximum matching approach），但只应用于有向网络。（Nature 473(7346):167 - 173.）

2012年，Nacher 和 Akutsu提出网络的最小支配集（minimum dominating sets, 简称MDSets）节点在无向网络的控制中起到重要作用的观点。

（ New J Phys 14(7):073005 - 073028. ）

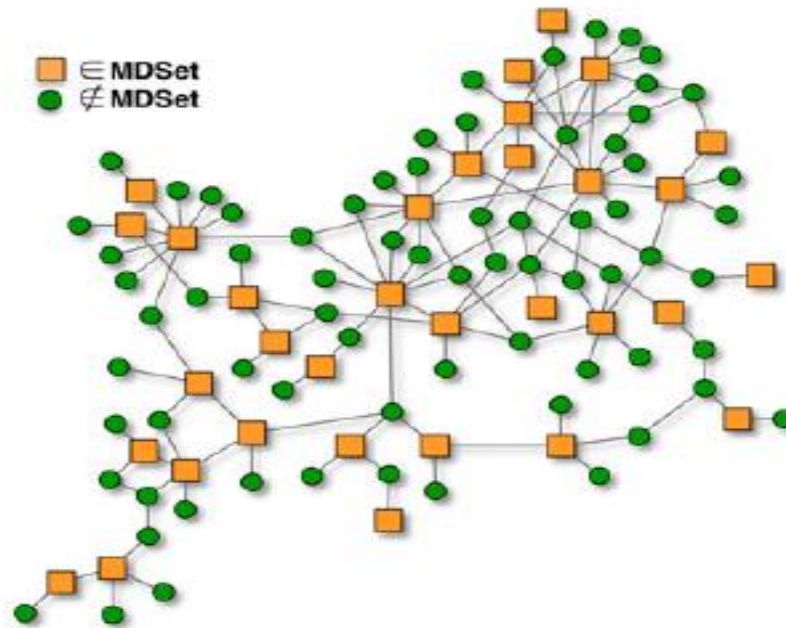


最小支配集

设 $D \subseteq V(G)$, 若对 $\forall u \in V(G)$, 要么 $u \in D$, 要么 u 与 D 中的某些顶点相邻, 则称 D 为图 G 的一个**支配集**。

图 G 的含顶点最少的支配集称为**最小支配集**。最小支配集一般**不唯一**。

A



问题1:

最小支配集蛋白有哪些特征？有什么生物学意义？



一般来说，蛋白质互作网络中连接数多的节点相对比较重要，点度中心性高的蛋白质失活后更容易致死，这些失活致死的蛋白质称为必须蛋白质（essential protein）。

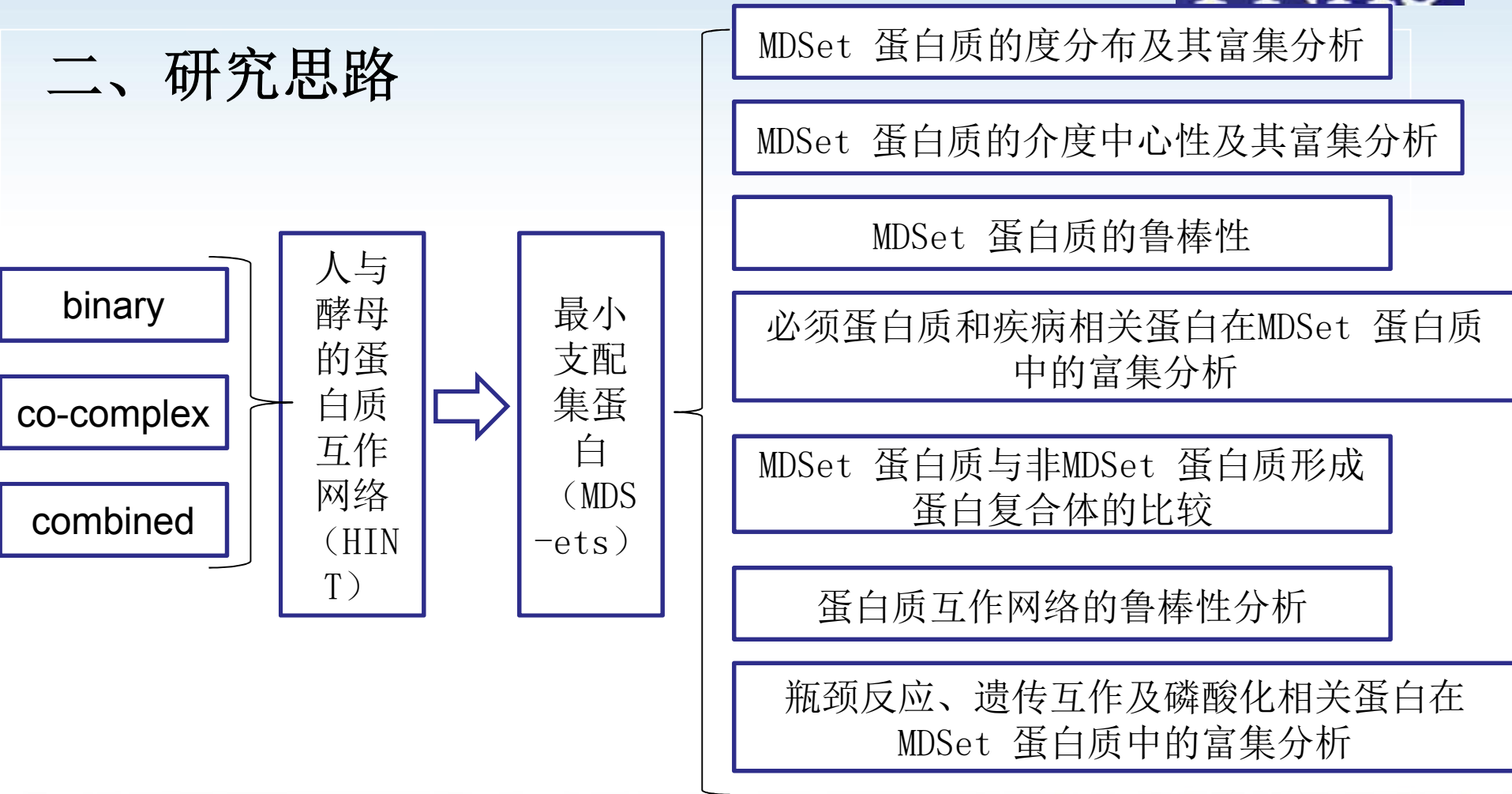
人类的病毒和寄生生物通过控制某些特定的蛋白质来控制宿主细胞，这些蛋白质同时与癌症有关。

问题2:

最小支配集蛋白质与必须蛋白质以及某些疾病（例如癌症）相关蛋白之间有什么关系？



二、研究思路



三、具体内容

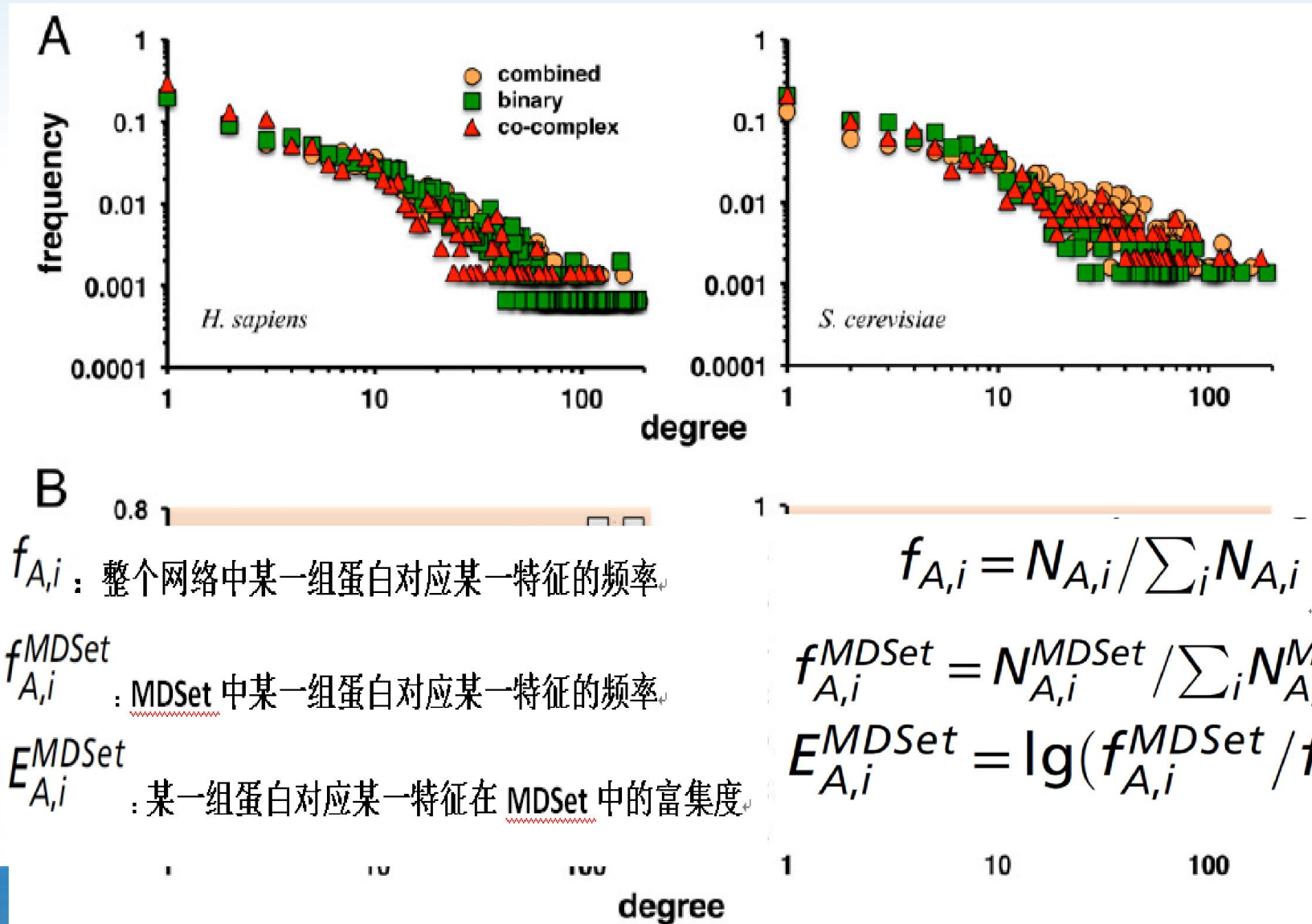
1、人与酵母的蛋白质互作网络基本信息

B

		combined	binary	co-complex
<i>H. sapiens</i>	N_{proteins}	8,495	8,073	3,089
	N_{ppi}	28,627	24,306	6,768
	$\text{size}_{\text{MDSet}}$	1,509	1,517	704
	$\%_{\text{MDSet}}$	17.8%	18.8%	22.8%
	$\langle k \rangle$	6.7	6.0	4.4
	$\langle k_{\text{MDSet}} \rangle$	17.0	14.9	8.7
		combined	binary	co-complex
<i>S. cerevisiae</i>	N_{proteins}	4,467	3,591	3,241
	N_{ppi}	22,243	9,386	14,405
	$\text{size}_{\text{MDSet}}$	629	719	482
	$\%_{\text{MDSet}}$	14.1%	20.0%	14.9%
	$\langle k \rangle$	10.0	5.2	8.9
	$\langle k_{\text{MDSet}} \rangle$	23.8	11.1	18.6



2、MDSet 蛋白质的度分布及其富集分析（单个MDSet 的富集）



$f_{A,i}$: 整个网络中某一组蛋白对应某一特征的频率

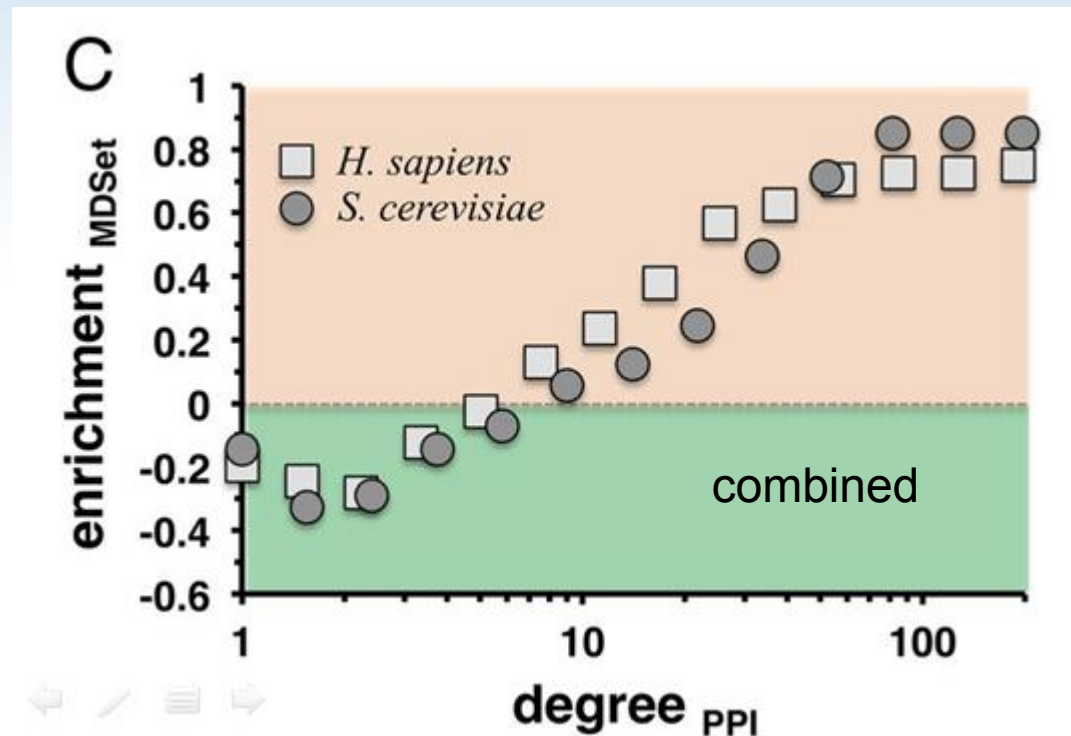
$f_{A,i}^{MDSet}$: MDSet 中某一组蛋白对应某一特征的频率

$E_{A,i}^{MDSet}$: 某一组蛋白对应某一特征在 MDSet 中的富集度

$$f_{A,i} = N_{A,i} / \sum_i N_{A,i}$$

$$f_{A,i}^{MDSet} = N_{A,i}^{MDSet} / \sum_i N_{A,i}^{MDSet}$$

$$E_{A,i}^{MDSet} = \lg(f_{A,i}^{MDSet} / f_{A,i})$$



$f_{A,i}$: 整个网络中某一组蛋白对应某一特征的频率。

$$f_{A,i} = N_{A,i} / \sum_i N_{A,i}$$

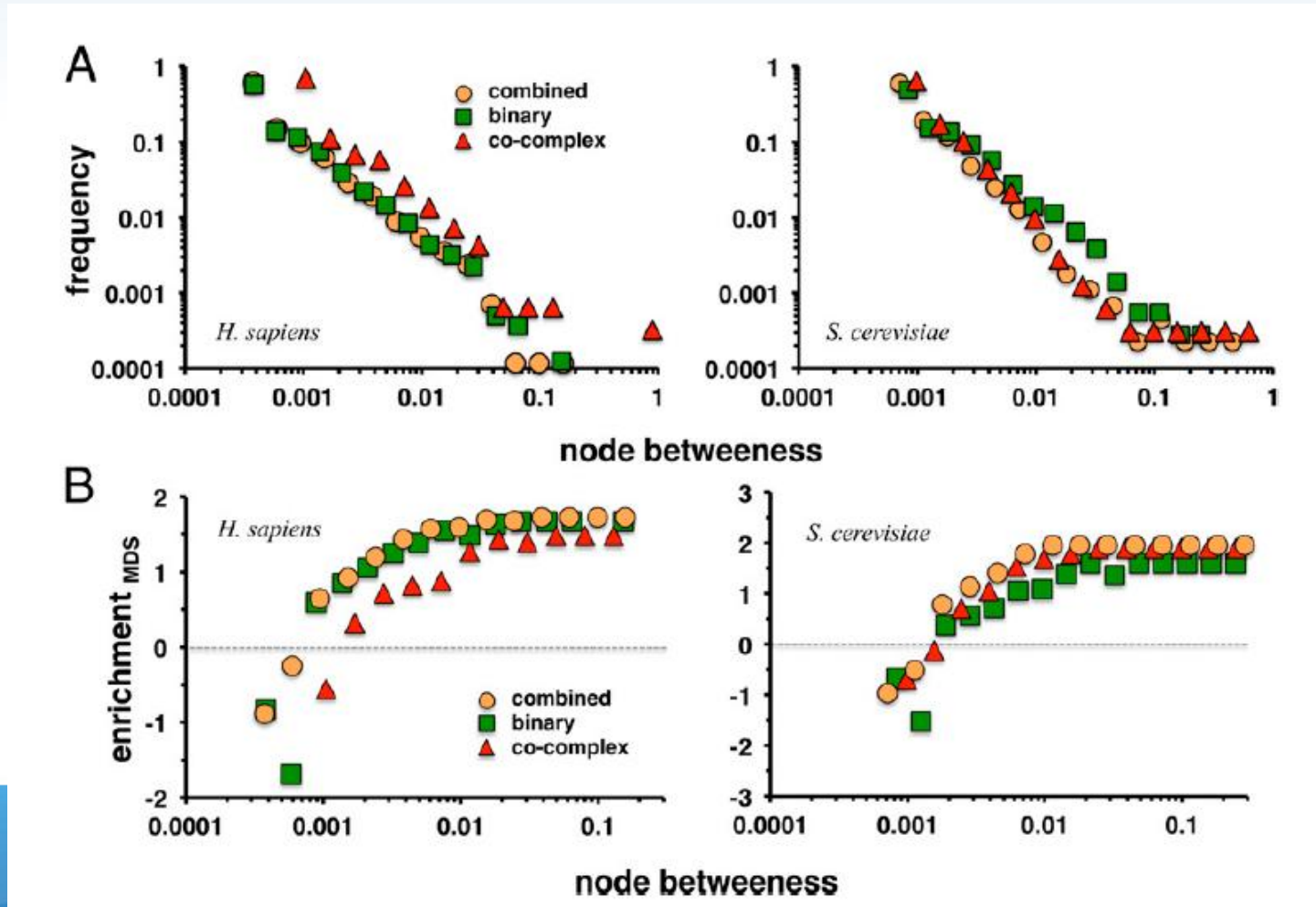
$f_{A,i}^{MDSets}$: MDSets 中某一组蛋白对应某一特征的频率。

$$f_{A,i}^{MDSets} = N_{A,i}^{MDSets} / \sum_i N_{A,i}^{MDSets}$$

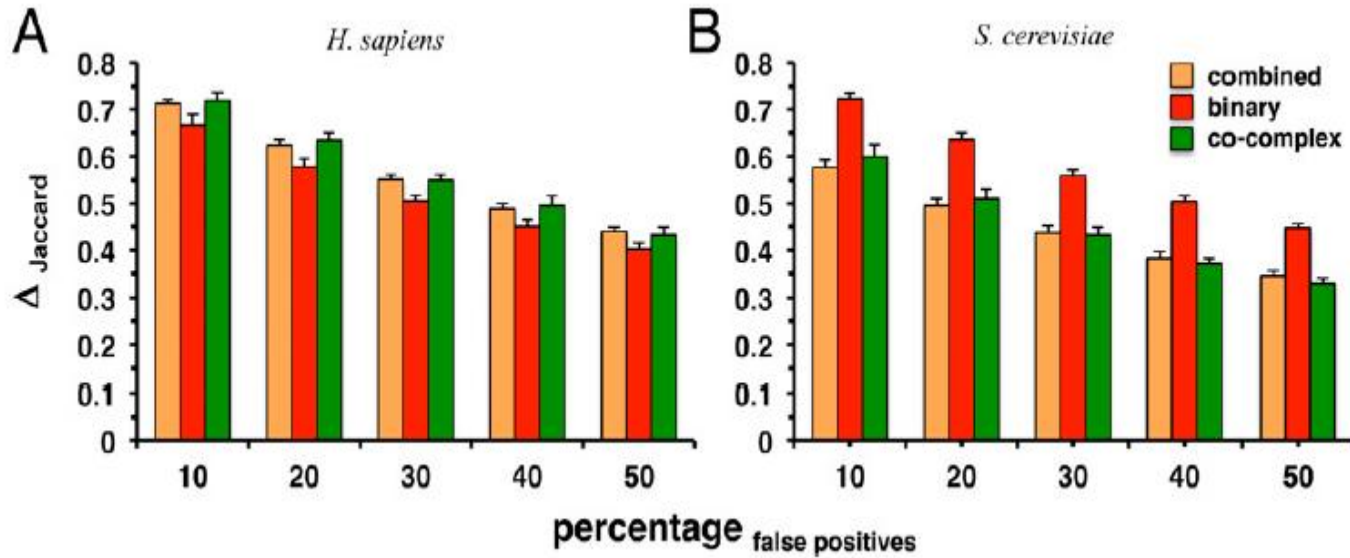
$E_{A,i}^{MDSets}$: 某一组蛋白对应某一特征在 MDSets 中的富集度。

$$E_{A,i}^{MDSets} = \lg(f_{A,i}^{MDSets} / f_{A,i})$$

3、MDSet 蛋白质的介度中心性及其富集分析



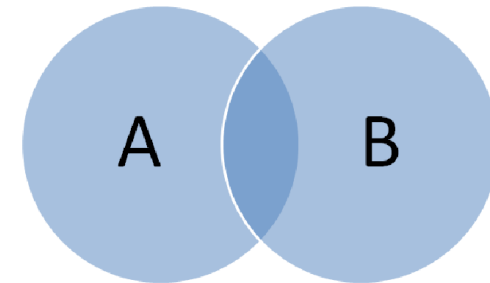
4、MDSets 蛋白质的鲁棒性



实际得到网络 \longrightarrow 假设无假阳性 \longrightarrow MDSetsA

随机删除X%网络 \longrightarrow X%假阳性 \longrightarrow MDSetsB

Jaccard index $\frac{AB}{A+B-AB}$



5、必须蛋白质和疾病相关蛋白在MDSet 蛋白质中的富集分析

D

	<i>H. sapiens</i>		
	combined	binary	co-complex
cancer-related	6.9×10^{-9}	2.5×10^{-9}	5.3×10^{-4}
virus-targeted	3.1×10^{-11}	1.4×10^{-12}	6.2×10^{-3}

	<i>S. cerevisiae</i>		
	combined	binary	co-complex
essential	6.5×10^{-6}	1.8×10^{-8}	5.6×10^{-3}

方法: Fisher 精确概率检验

目的: 检验一个总体中两类元素的频数是否有差异

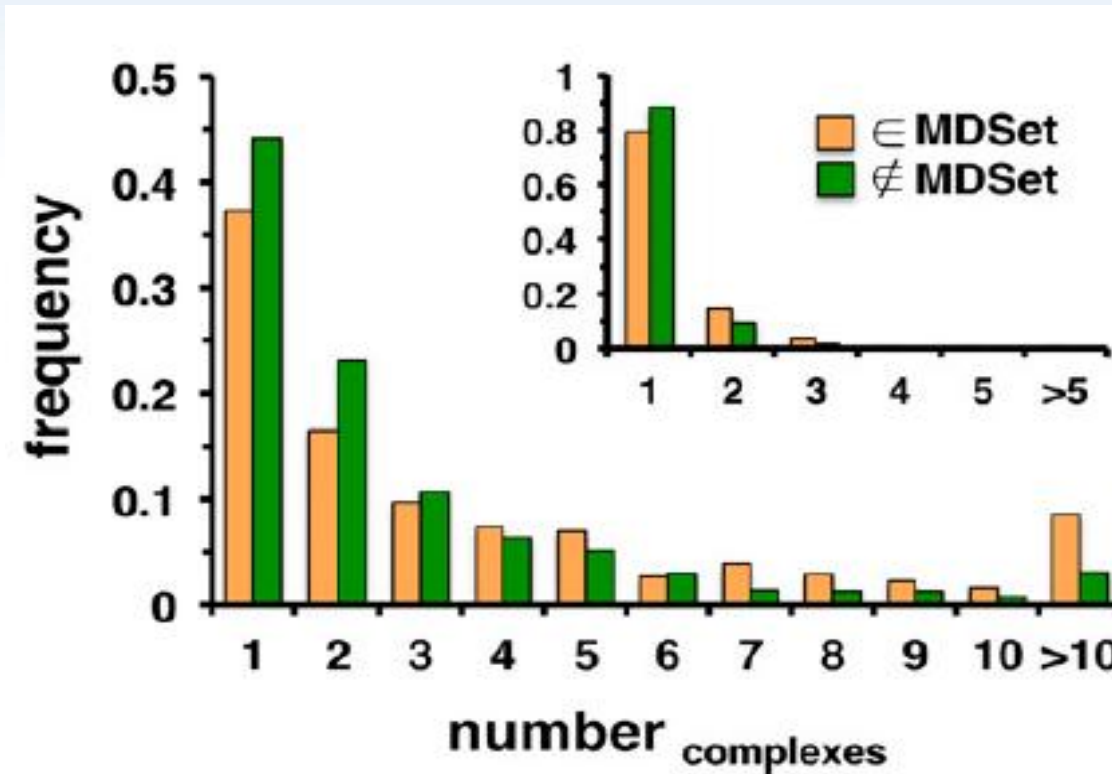
	Cancer-related	Non cancer-related	Total
MDSet	a	1509-a	1509
Non-MDSet	c	27118-c	27118
Total	a+c	28627-a-c	28627

计算P

$P = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a! b! c! d! n!}$
 算出极端情形出现的概率并与P相加, 双侧检验与 0.5α 比较



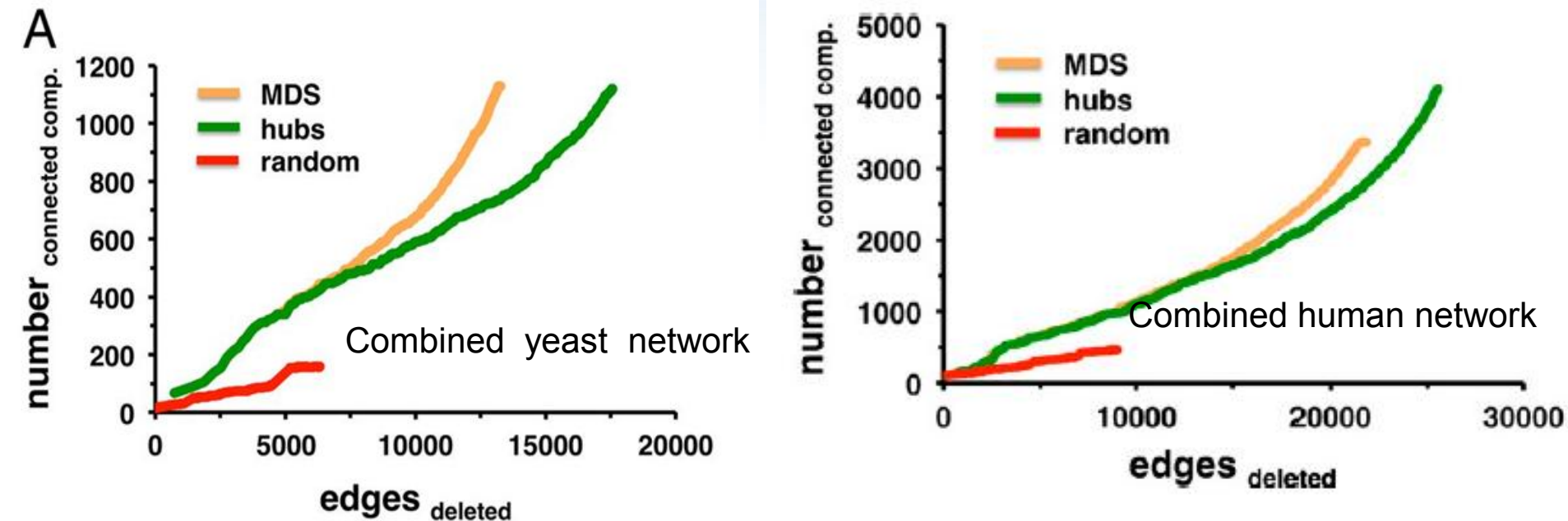
6、MDSset 蛋白质与非MDSset 蛋白质形成蛋白复合体的比较



Wilcoxon秩和检验：检验两个变量总体总体分布是否有差异



7、蛋白质互作网络的鲁棒性分析



方法：三种蛋白质集按度的高低删除边

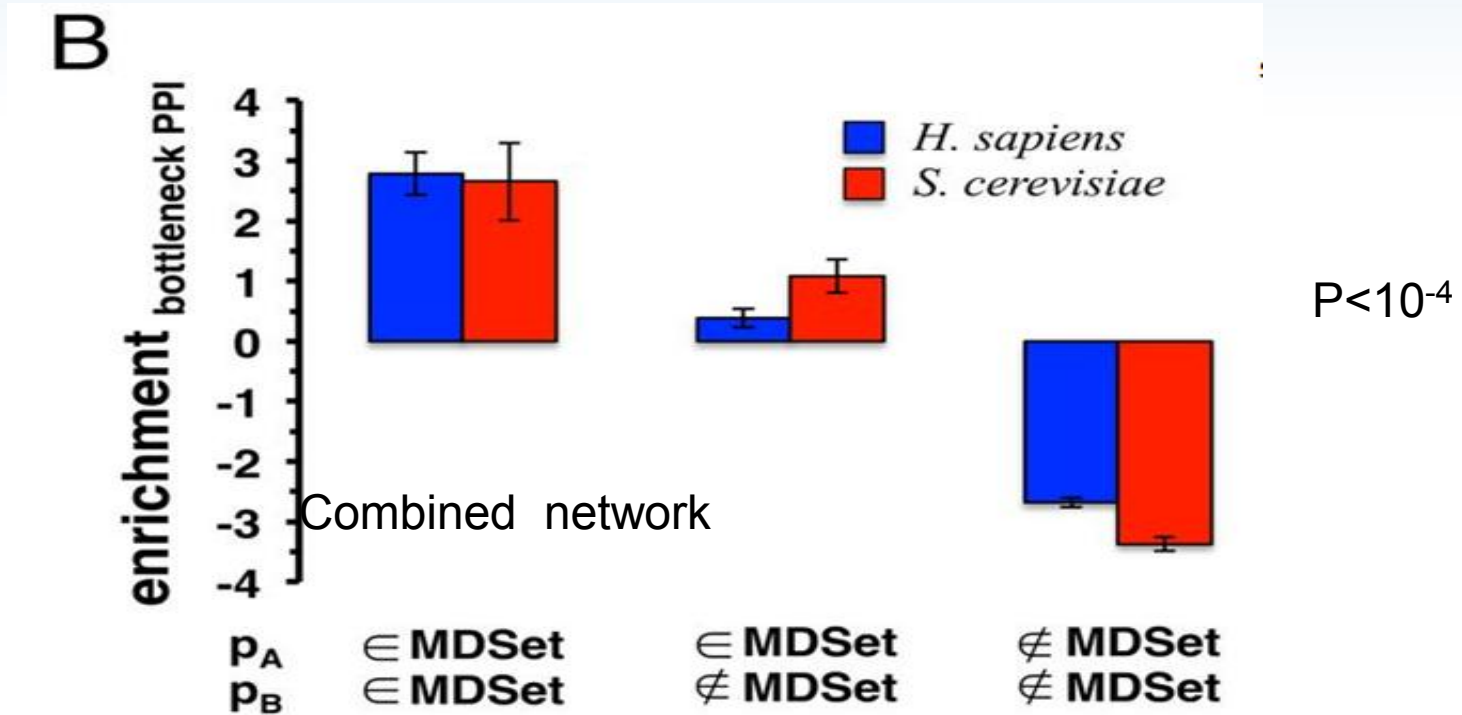
逻辑思路：1、若图G中任意两顶点都连通，称G是连通图。

2、无向图G的连通分支是一个最大的连通子图。

3、连通图G的连通分支数为1，连通分支数越大，图的连通性越差



8、瓶颈互作、遗传互作、转录因子及激酶蛋白在MDSet 蛋白质中的富集分析（多个MDSet 蛋白的富集）



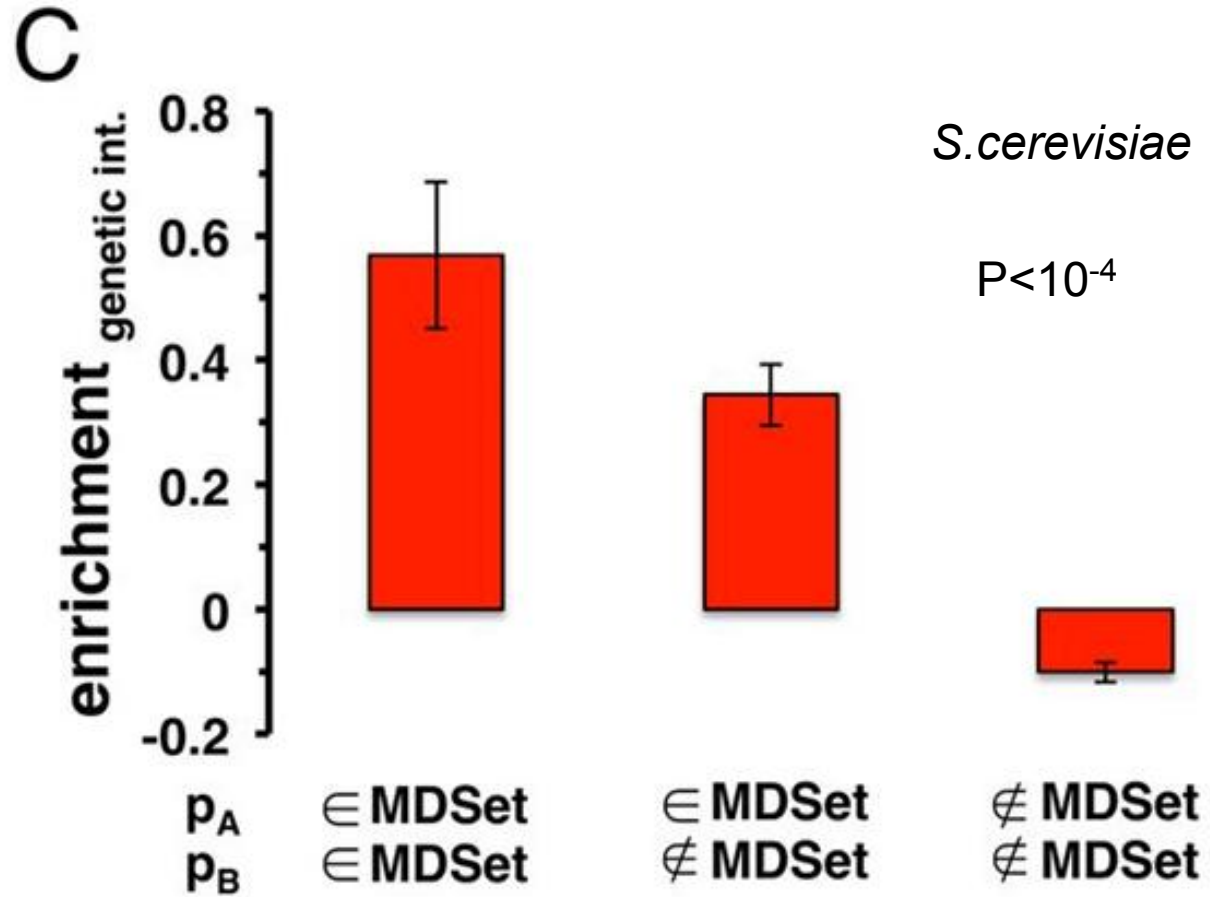
瓶颈互作：前10%最高边介数的互作

方法：在MDSet蛋白中随机抽样10000次，

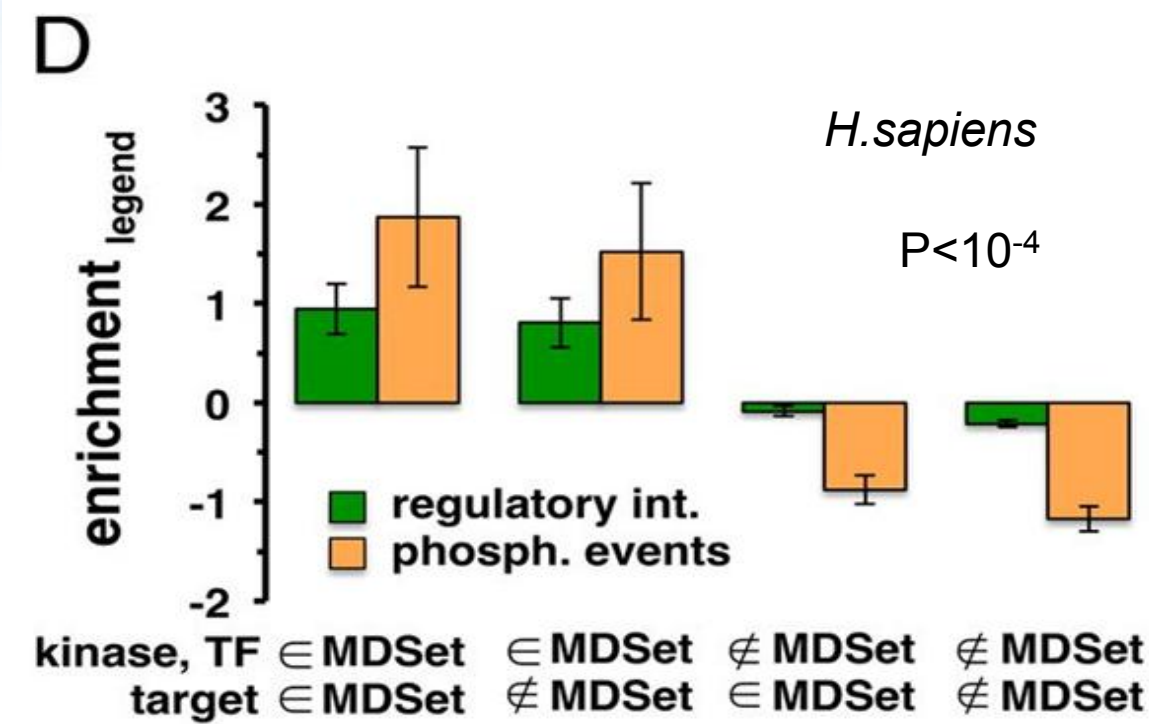
计算富集度，然后进行Fisher精确检验



遗传互作



转录因子及激酶



四、结论

- 1、 高的度数并非是一个蛋白质加入MDSets的标准。
- 2、 与度高的Hub节点相比， MDSets有以下特点：
 - (1) MDSets 考虑了整个网络，无需任何参数。
而Hub节点未考虑整个网络，需要阈值。
 - (2) MDSets更多出现在高介度的节点之中。
 - (3) 对MDSets节点的蓄意攻击比Hub节点对于整个的破坏性（连通性）更大。



五、评价及思考

- 1、本文通过对人和酵母MDSet蛋白的网络结构分析，MDSet蛋白与Hub蛋白的比较，以及MDSet蛋白生物学意义的阐述，证明了MDSet蛋白在蛋白质互作网络控制中的重要地位。
- 2、由于实验技术的原因，蛋白质互作网络有很多的假阳性。我们认为应该增加对于蛋白质网络的可靠性分析。
- 3、作者指出MDSet蛋白更多出现在高介度节点之中，我们认为应该比较一下MDSet蛋白与Hub蛋白的介度中心性，然后看是否可以解释实验7（蛋白质互作网络的鲁棒性）的结果。



4、通过阅读相关资料我们了解到网络结构和基因冗余都可以是生物网络更加稳健，建议增加MDSset蛋白对应的基因重复性的富集分析，从而补充证明MDSset蛋白对生物网络稳健性的重要意义。



Thank you!

