

果蝇SOCS36E蛋白的结构 及信号通路分析

G04A 陈佩宁

G04B 赵瑜涵

G04C 邝沃钿

G04D 杨迪

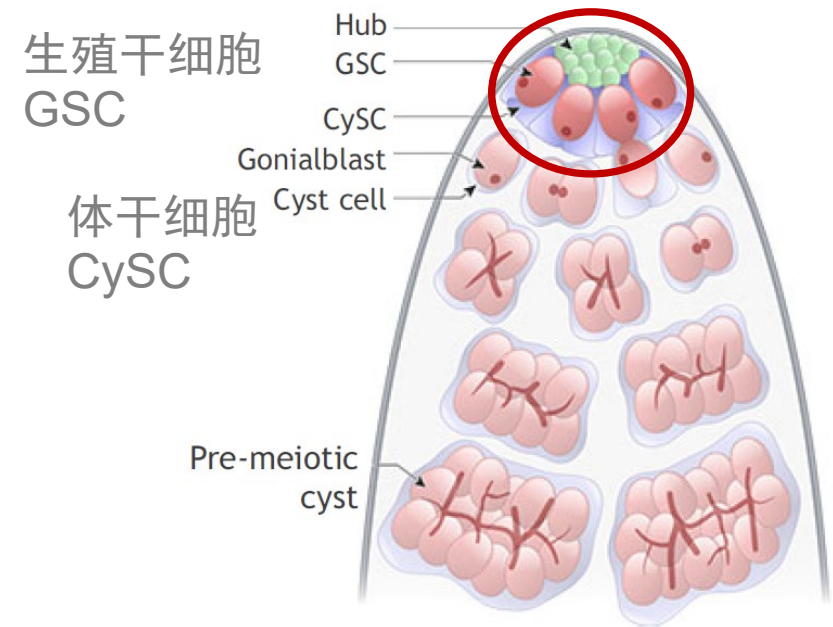
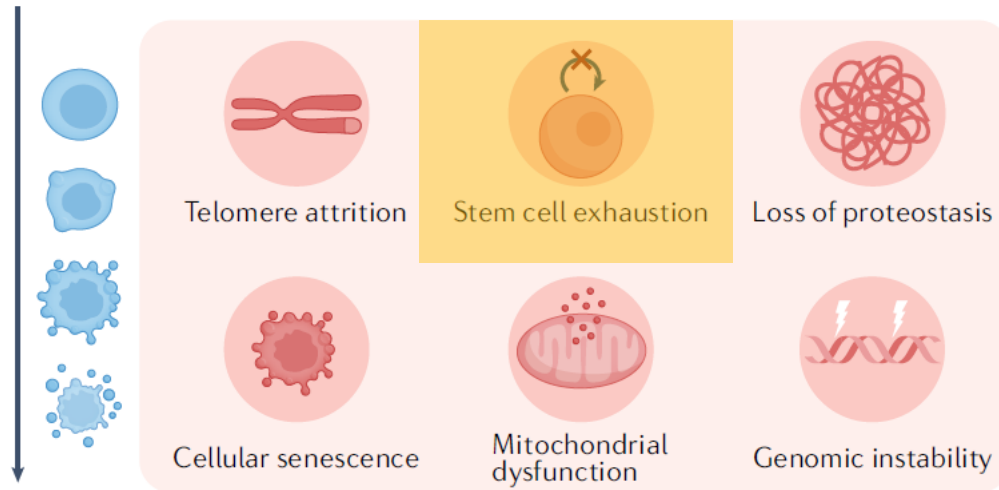
大纲

- 背景介绍
- 文献查找
- 结构分析
- 互作网络
- 总结及收获

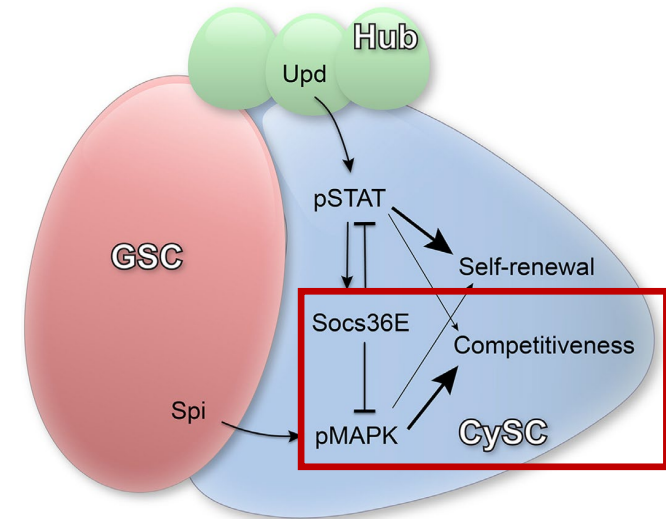
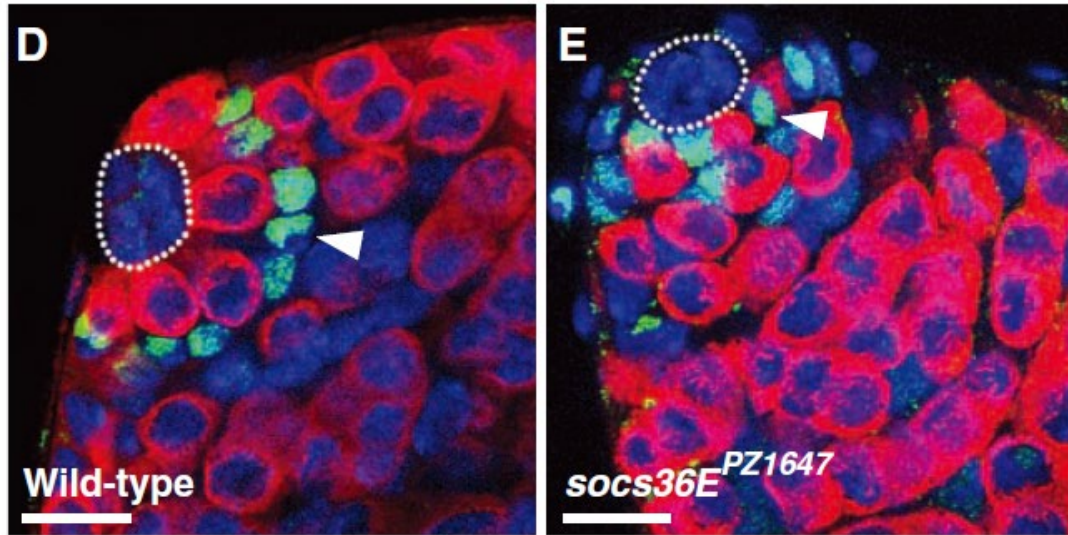
背景介绍

为什么果蝇衰老过程中生殖干细胞会逐渐被耗竭？

Molecular and cellular hallmarks of ageing



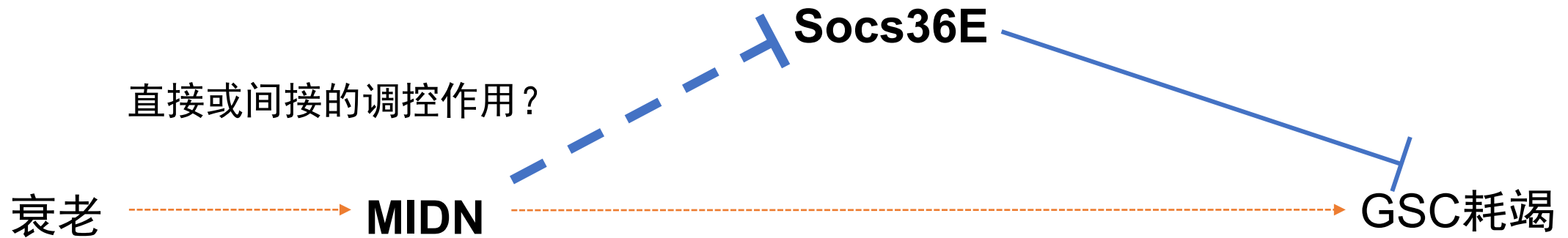
Socs36E会抑制干细胞耗竭



Socs36E参与了GSC和CySC竞争的过程



衰老过程中的干细胞耗竭的具体机制？



科学问题：

MIDN是否能调控socs36E？

一. MIDN可能通过哪种方式调控Socs36E?

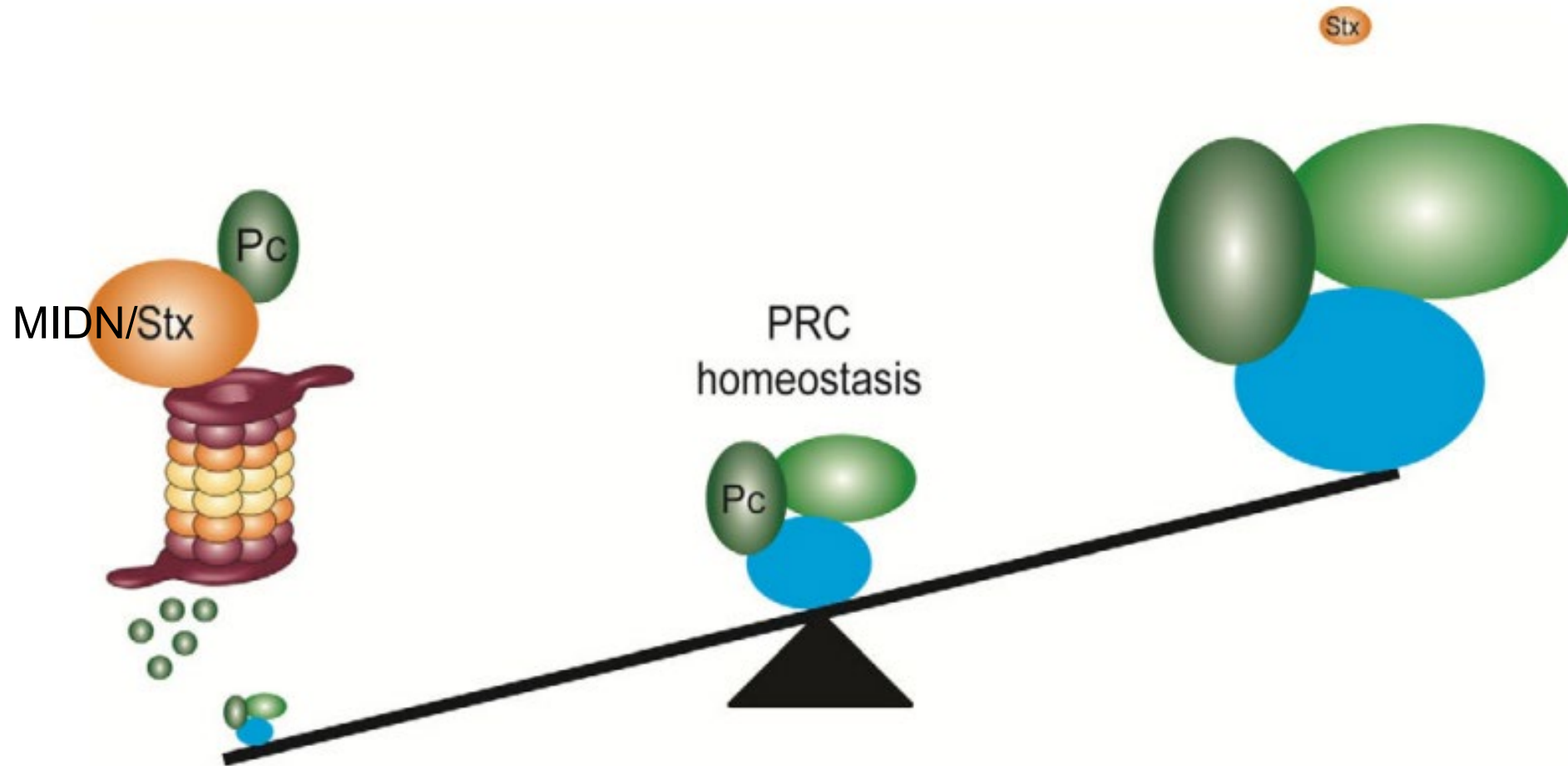
- 分别找MIDN和Socs36E的背景，参与的生物学过程和机制
- 是否有MIDN和Socs36E直接互作的文献
- 没有直接互作时，从两个蛋白的功能和机制做出间接互作的推测

MIDN背景 - Flybase References

1. Zhao, Z., Zhao, X., He, T., Wu, X., Lv, P., Zhu, A.J., and Du, J. (2021). Epigenetic regulator Stuxnet modulates octopamine effect on sleep through a Stuxnet-Polycomb-Oct β 2R cascade. *EMBO Rep* 22, e47910. [10.15252/embr.201947910](https://doi.org/10.15252/embr.201947910).
2. Du, J., Zhang, J., He, T., Li, Y., Su, Y., Tie, F., Liu, M., Harte, P.J., and Zhu, A.J. (2016). Stuxnet Facilitates the Degradation of Polycomb Protein during Development. *Dev Cell* 37, 507–519. [10.1016/j.devcel.2016.05.013](https://doi.org/10.1016/j.devcel.2016.05.013).

没有文献直接表明MIDN可以调控Socs

MIDN背景：MIDN如何调控其他蛋白（如Pc蛋白）的水平？

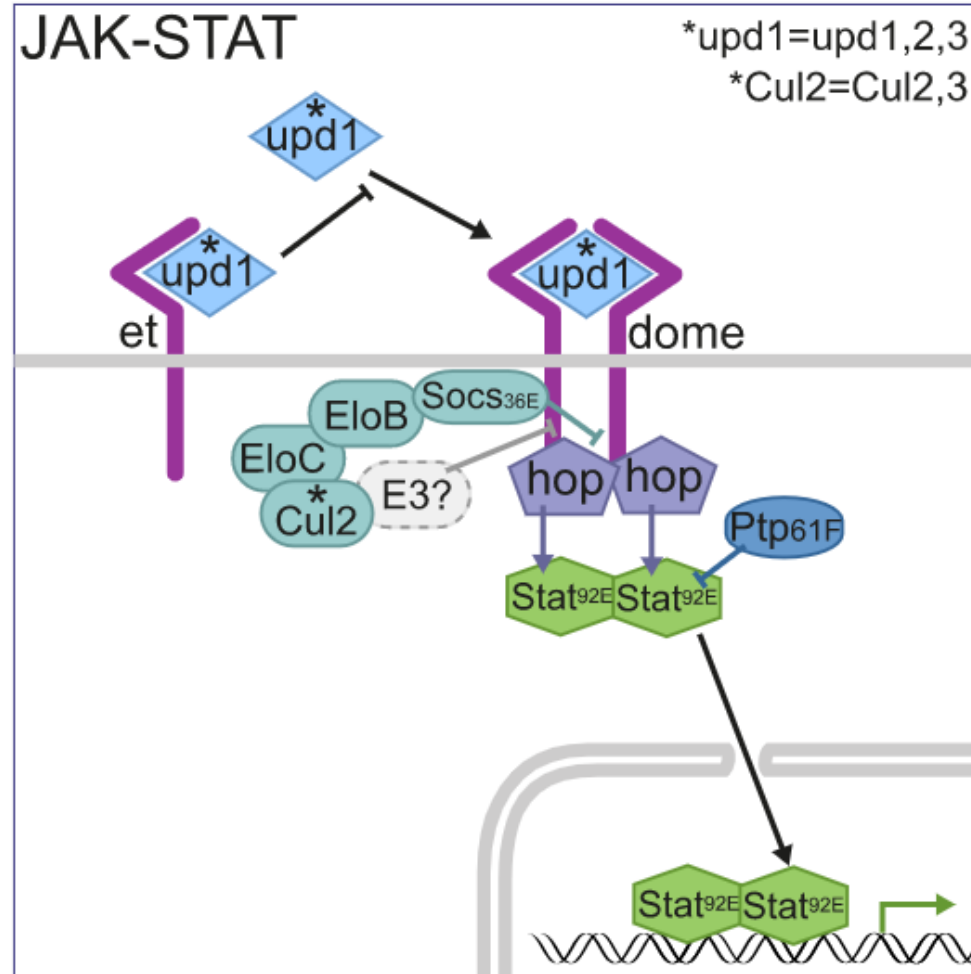


MIDN可能是通过使Socs或Socs调节蛋白降解发挥作用的。

Socs背景：Socs如何被调控的？

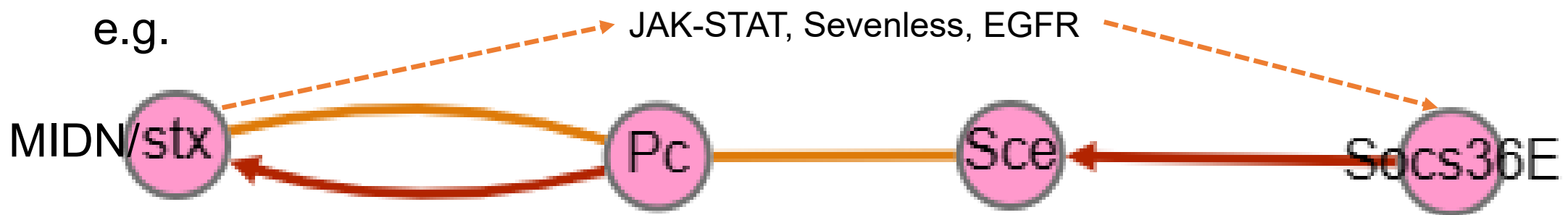
果蝇专用数据库Flybase里收录了SOCS36E可能参与的三个信号通路：

- (1) JAK-STAT信号通路
- (2) Sevenless信号通路
- (3) EGFR信号通路



MIDN可能通过哪种方式调控Socs36E?

- 文献中没有信息直接说明MIDN和Socs36E的关系
- 如果有直接互作，推测可能是MIDN促进了Socs36E的降解。
- 如果是间接互作，可能是MIDN促进了Socs36E调节蛋白的降解。但也没有充分的证据。





二. 从Socs36E结构上探究它的 蛋白水平调控

Phyre, alphafold


MIDN和Socs36E蛋白注释不完善 - Uniprot


Q8INY1 · Q8INY1_DROME

Protein ⁱ	Suppressor of cytokine signaling at 36E, isoform D
Gene ⁱ	Socs36E
Status ⁱ	 UniProtKB unreviewed (TrEMBL)
Organism ⁱ	Drosophila melanogaster (Fruit fly)

Amino acids	737
Protein existence ⁱ	Predicted
Annotation score ⁱ	 5/5

Q8SXD4 · MIDN_DROME

Protein ⁱ	Midnolin homolog
Gene ⁱ	stx
Status ⁱ	 UniProtKB reviewed (Swiss-Prot)
Organism ⁱ	Drosophila melanogaster (Fruit fly)

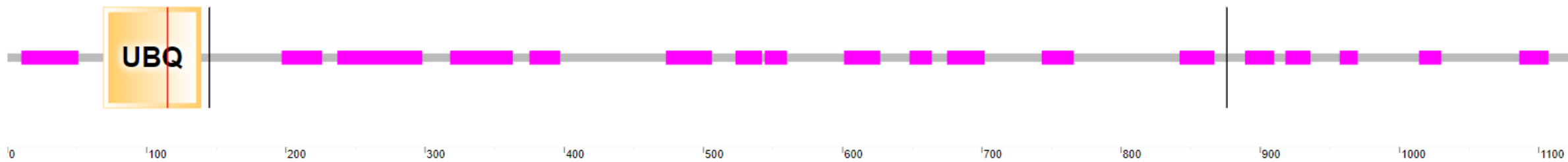
Amino acids	1122
Protein existence ⁱ	Evidence at protein level
Annotation score ⁱ	 4/5

MIDN结构域信息


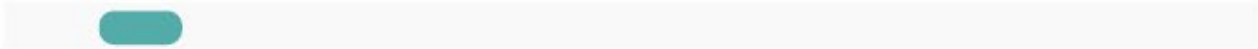
SMART

Domains within *Drosophila melanogaster* protein **MIDN_DROME** (Q8SXD4)

Midnolin homolog



Pfam

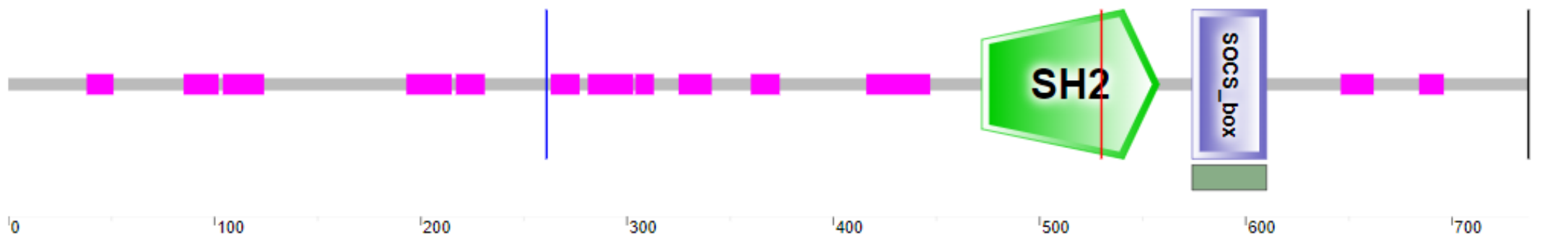
ACCESSION	NAME	SOURCE DATABASE	MATCHES
 IPR000626	Ubiquitin-like domain	InterPro	 500 1000

SOCS36E结构域信息

SMART

Domains within *Drosophila melanogaster* protein Q8INY1_DROME (Q8INY1)

Suppressor of cytokine signaling at 36E, isoform D



Pfam

ACCESSION	NAME	SOURCE DATABASE	MATCHES
PF00017	SH2 domain	Pfam	
PF07525	SOCS box	Pfam	

果蝇SOCS36E和MIDN蛋白结构

- 没有测定的PDB结构
- 通过phyre进行结构预测，并与alphafold预测的结果进行对比
- 利用ZDOCK对预测的结构进行蛋白分子对接预测（实际上应该先通过实验验证二者确实存在直接相互作用）
- 查询同源蛋白的PDB结构

SOCS36E结构预测 - phyre

Top model

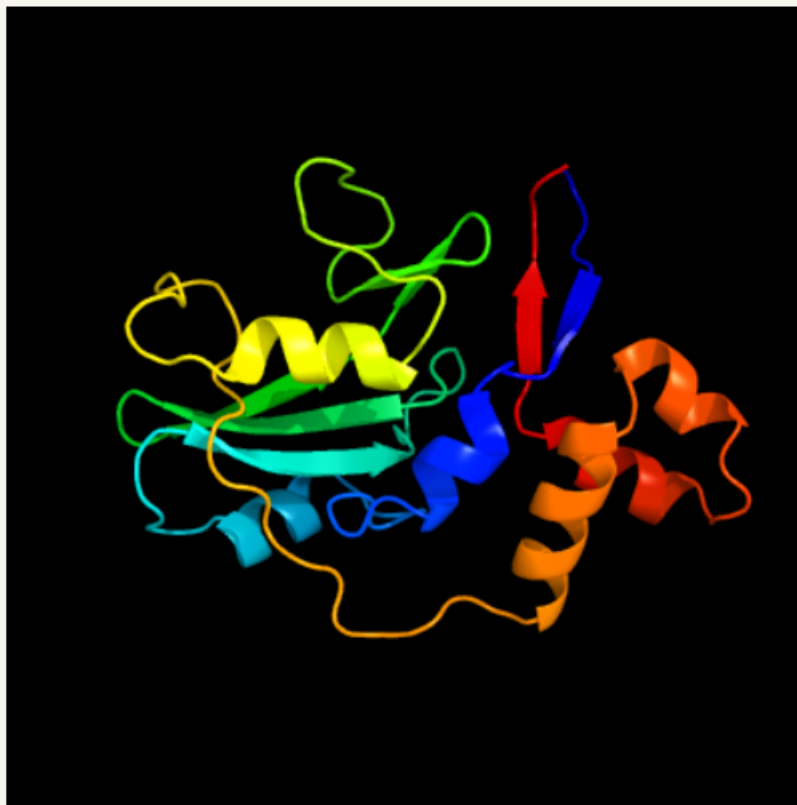


Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:46.969 **Y**:46.362 **Z**:44.589

Model (left) based on template [c2izvA](#)

Top template information

PDB header:transcription

Chain: A: **PDB Molecule:**suppressor of cytokine signaling 4;

PDBTitle: crystal structure of socs-4 in complex with elongin-b and elongin-c at2 2.55a resolution

PDB Entry: [PDBe](#) [RCSB](#) [PDBj](#)

-

Confidence and coverage

Confidence:

100.0%

Coverage: **22%**

163 residues (22% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

Additional confident templates have been detected (see [Domain analysis](#)) which cover other regions of your sequence.

339 residues (46%) could be modelled at >90% confidence using multiple-templates.

You may wish to try resubmitting your sequence in "intensive" mode to model more of your sequence.

Warning: 61% of your sequence is predicted disordered. Disordered regions cannot be meaningfully predicted.

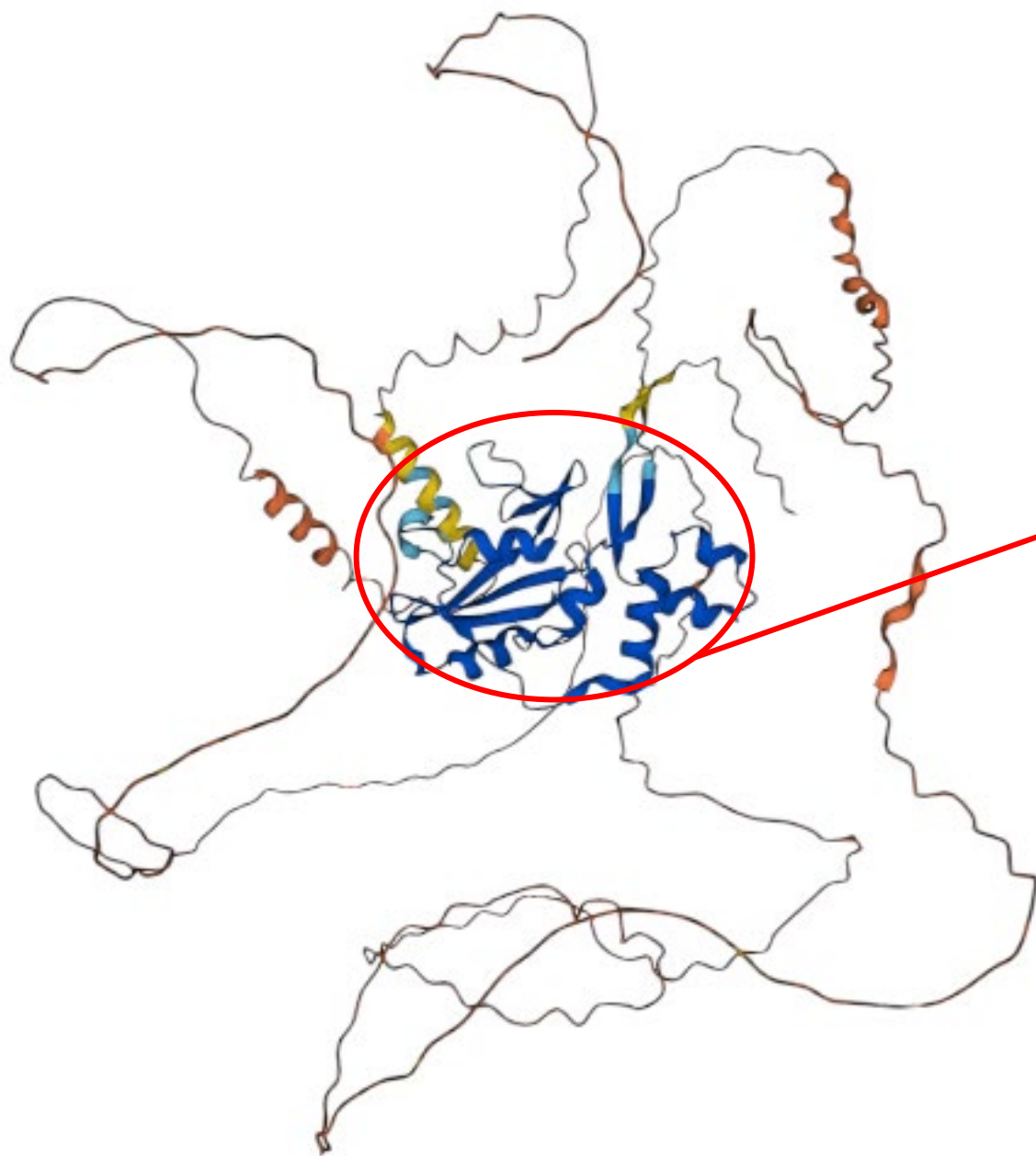
3D viewing

[Interactive 3D view in JSmol](#)

For other options to view your downloaded structure offline see the [FAQ](#)

- 覆盖度只有22%
- 61%的序列被预测为无序区

SOCS36E结构预测 - alphafold



Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

- 大部分序列被预测为无序区，预测效果差

SOCS36E同源蛋白测定的结构

(1) Q8WXH5 · **SOCS4_HUMAN**

- 存在测定的PDB结构2izv，是phyre对SOCS36E建模时使用的模板；只覆盖SOCS4_HUMAN片段中的第274-437位氨基酸（全长440 aa）

(2) O54928 · **SOCS5_MOUSE**

- 存在测定的PDB结构2n34，由NMR方法测定；只覆盖SOCS5_MOUSE片段中的第175-244位氨基酸（全长536 aa）

MIDN结构预测 - phyre



Image coloured by rainbow N → C terminus
Model dimensions (Å): **X**:27.217 **Y**:37.430 **Z**:30.973

Top model

Model (left) based on template [d1c3ta](#)

Top template information

Fold: beta-Grasp (ubiquitin-like)

Superfamily: Ubiquitin-like

Family: Ubiquitin-related

PDB entry: [PDBe](#) [RCSB](#) [PDBj](#)

Confidence and coverage

Confidence: **98.3%** Coverage: **7%**

75 residues (7% of your sequence) have been modelled with 98.3% confidence by the single highest scoring template.

Additional confident templates have been detected (see [Domain analysis](#)) which cover other regions of your sequence.

145 residues (13%) could be modelled at >90% confidence using multiple-templates.

You may wish to try resubmitting your sequence in "intensive" mode to model more of your sequence.

Warning: 78% of your sequence is predicted disordered. Disordered regions cannot be meaningfully predicted.

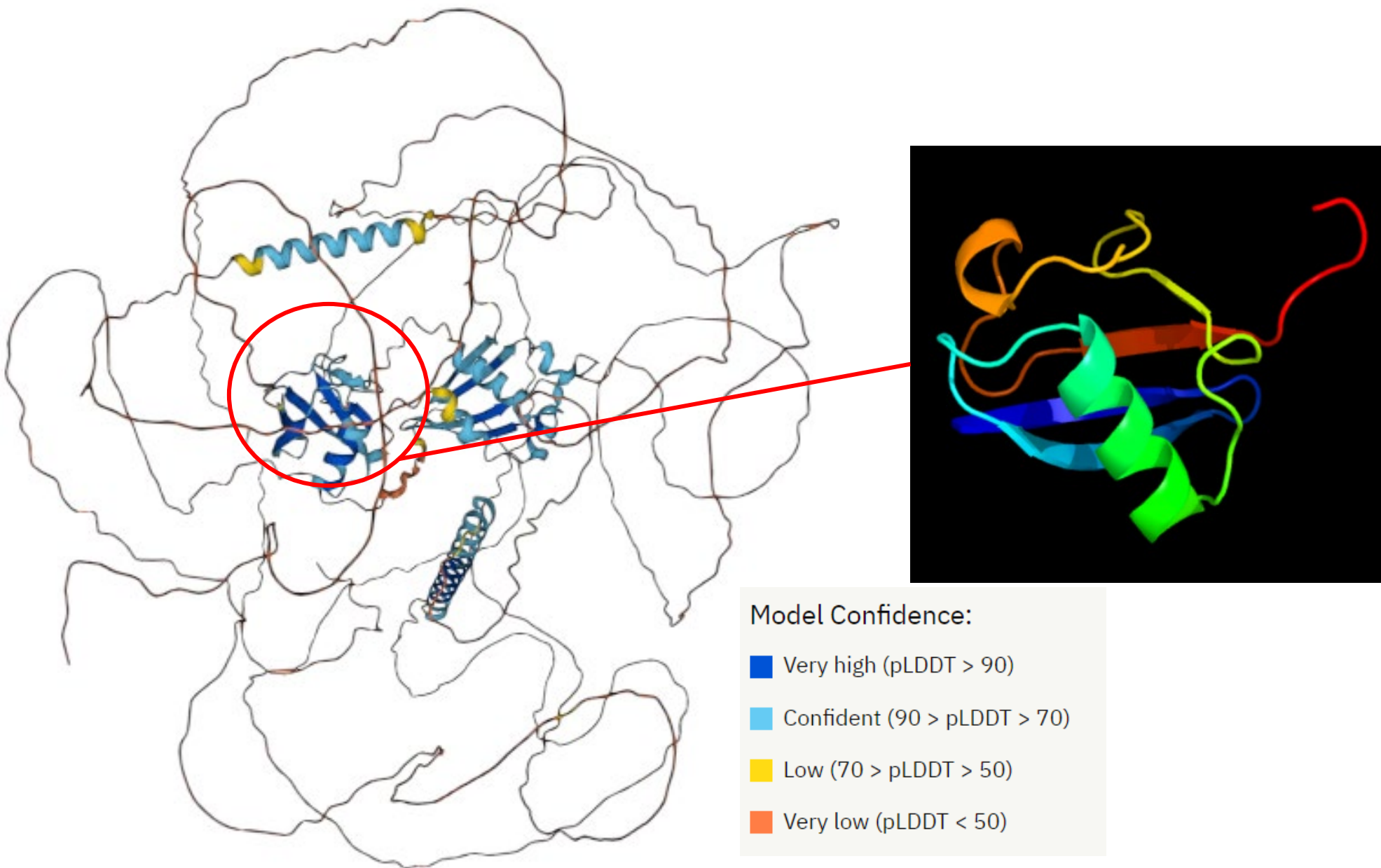
3D viewing

[Interactive 3D view in JSmol](#)

For other options to view your downloaded structure offline see the [FAQ](#)

- 覆盖度只有7%，只预测出 ubiquitin-like 结构域的结构（第 69-143 位氨基酸）
- 78% 的序列被预测为无序区

MIDN结构预测 - alphafold



- 大部分序列被预测为无序区，预测效果差

结构部分结论

ZDOCK SERVER

[ZDOCK](#) [M-ZDOCK](#) [Help](#) [Tools](#) [References](#)

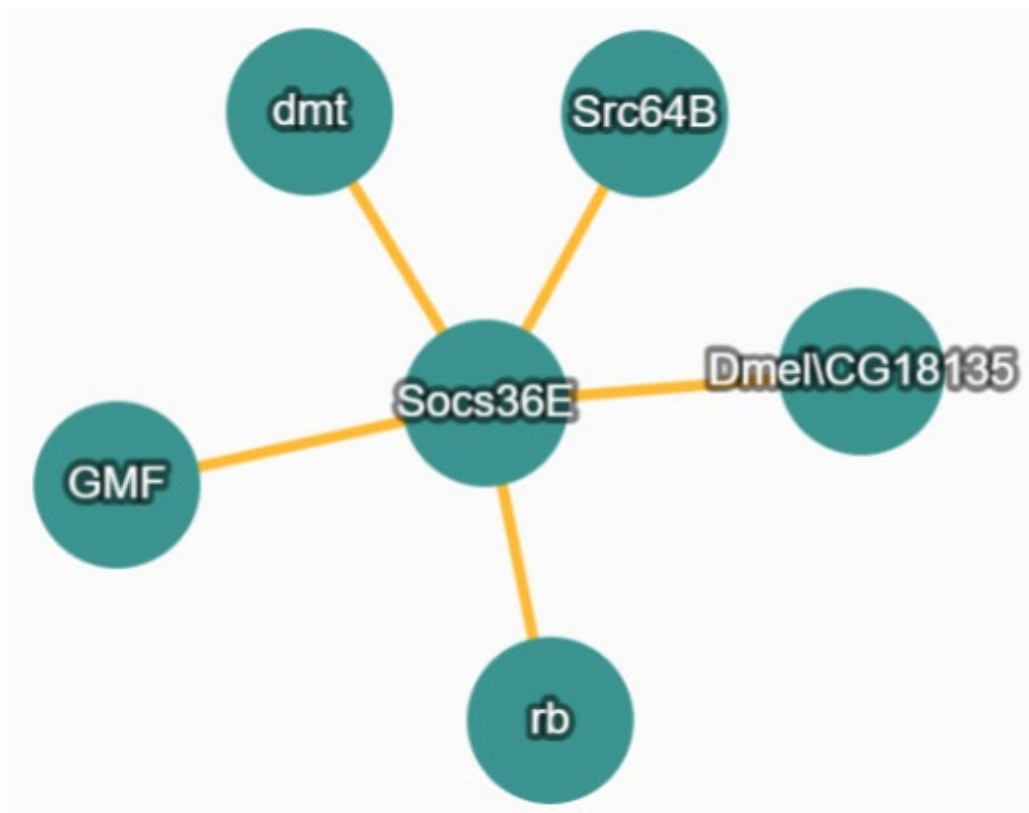
Input Protein 1	<input type="text" value="PDB File"/>	<input type="button" value="Choose File"/>	No file chosen
Input Protein 2	<input type="text" value="PDB File"/>	<input type="button" value="Choose File"/>	No file chosen
Enter your email:	<input type="text"/>		

MIDN和Socs36E的蛋白序列特征及结构数据库中信息太少，且结构预测的准确度都很差，难以进行后续蛋白分子对接预测。所以从现有的序列特征还暂时无法得出二者之间有直接互作。

三. 从互作网络寻找Socs36E和MIDN的互作关系

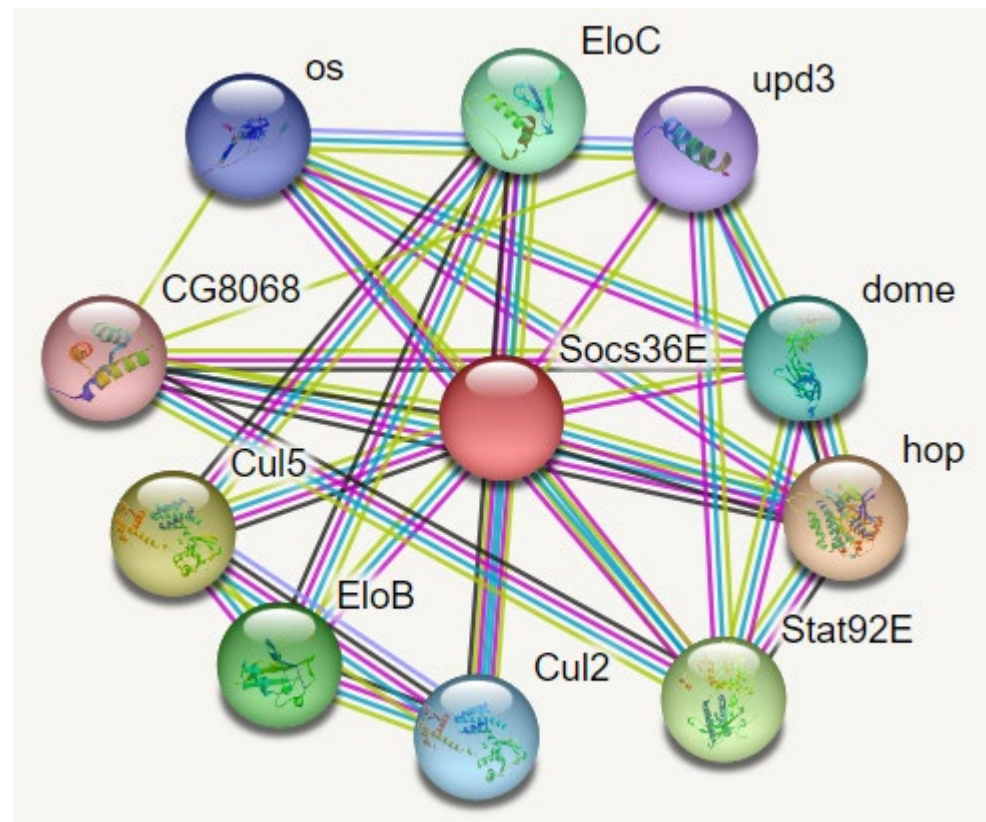
- 互作网络的数据库，EsyN的使用
- Socs36E和MIDN及对应同源蛋白关系的查找
- 可能的调控模型

互作网络数据库



IntAct数据库

在IntAct数据库中收录的SOCS36E的互作蛋白，都是经过酵母双杂验证，可信度高。



STRING数据库

在String数据库中收录的SOCS36E的互作蛋白，其中通过X射线测结构验证的这些互作关系可信度高，其他的需要进一步的验证。

EsyN



- Home
- Tutorial
- Graphs
- Petri Nets
- Trees
- Explore
- My esyN

[Cite](#) [Partners](#) [Contacts](#) [Issues](#)

Dive right in, generate a network and start exploring interactions.

Type a list or upload a file below.

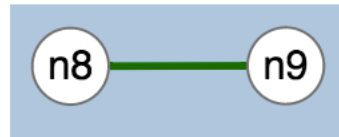
If you only input one gene, select "Include interactors of this gene"

Enter a **comma separated** list of genes:

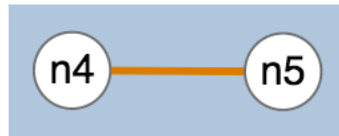
stx, Socs36E



Selected edge



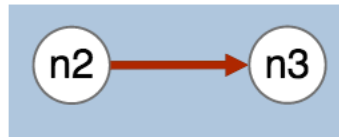
Genetic interaction edge



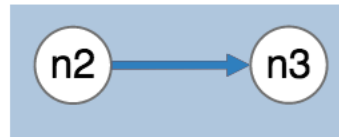
Physical interaction edge



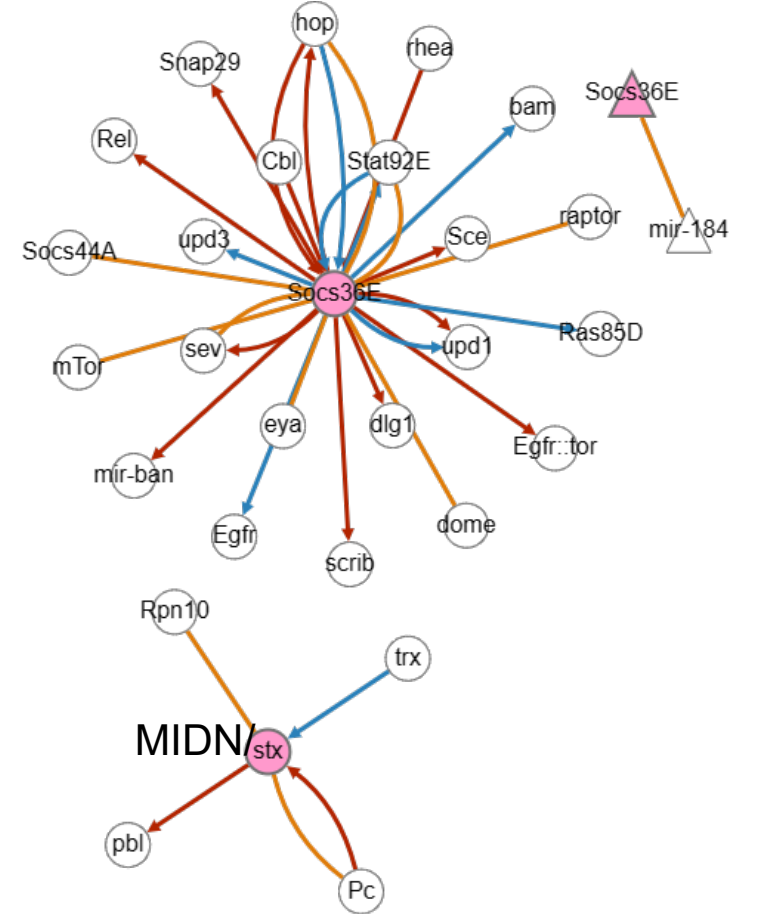
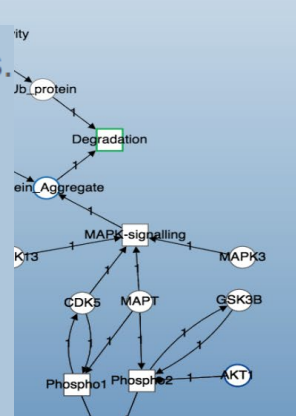
Within complex interaction edge



Suppressing genetic interaction



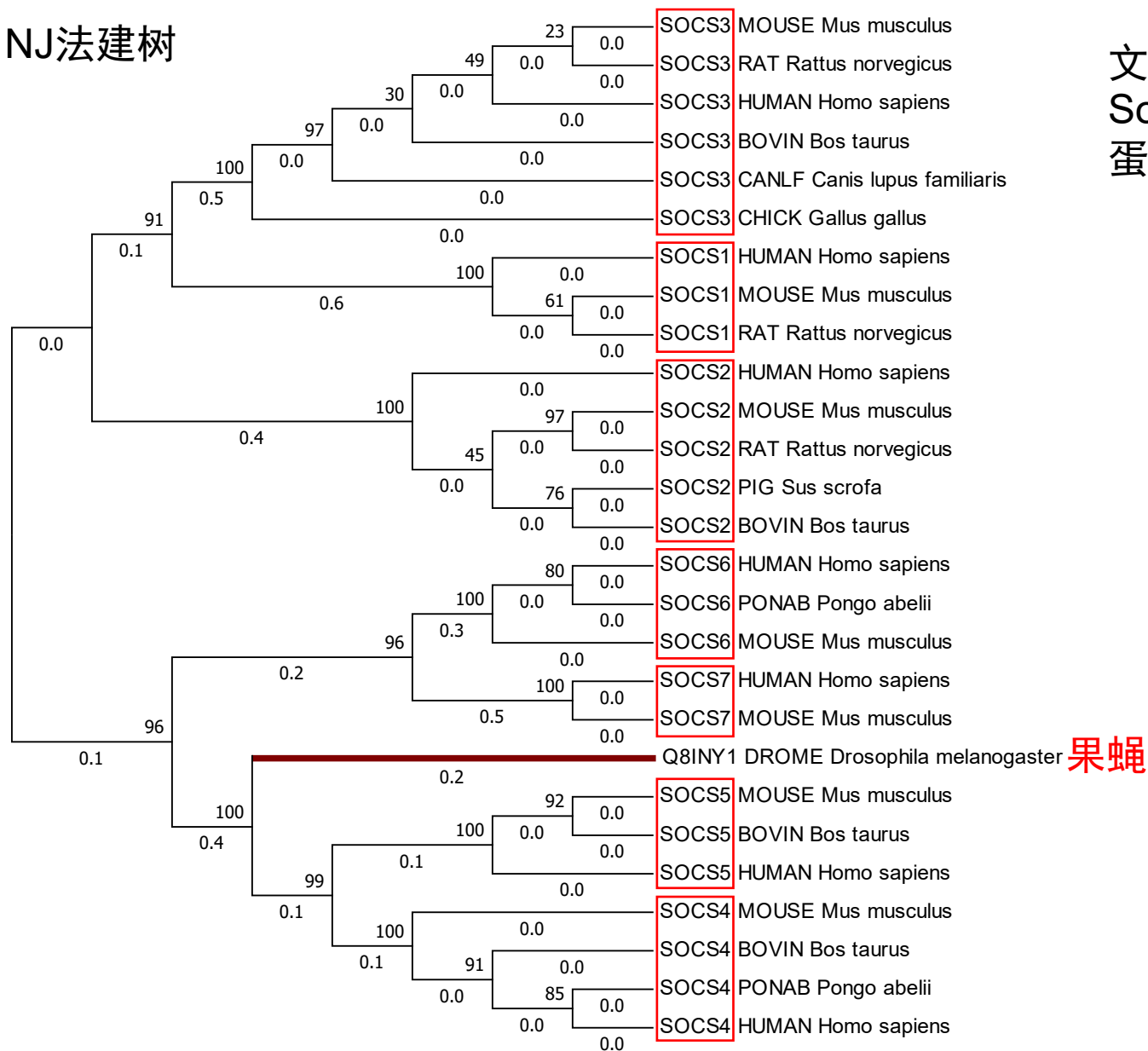
Enhancing genetic interaction



果蝇中MIDN和Socs36E没有直接互动

SOCS36E同源蛋白和系统发生树 - MAGA

MUSCLE进行多序列比对，NJ法建树

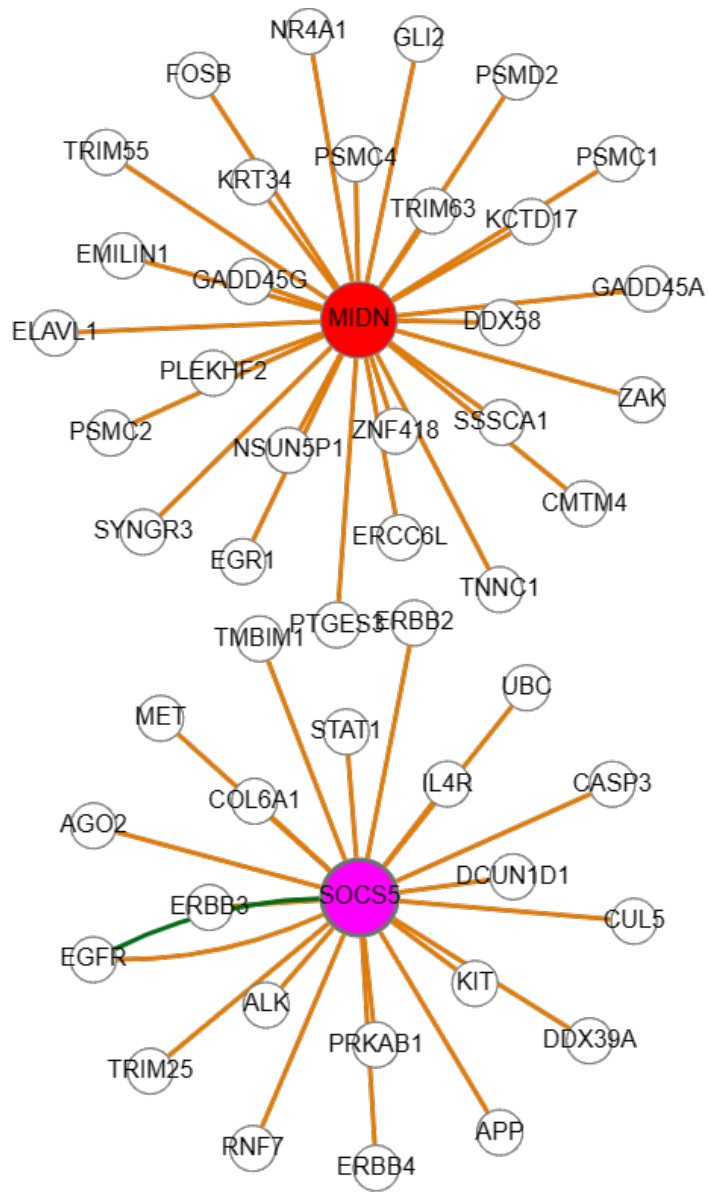


文献中查找到的
Socs36E的人中的同源
蛋白是Socs5

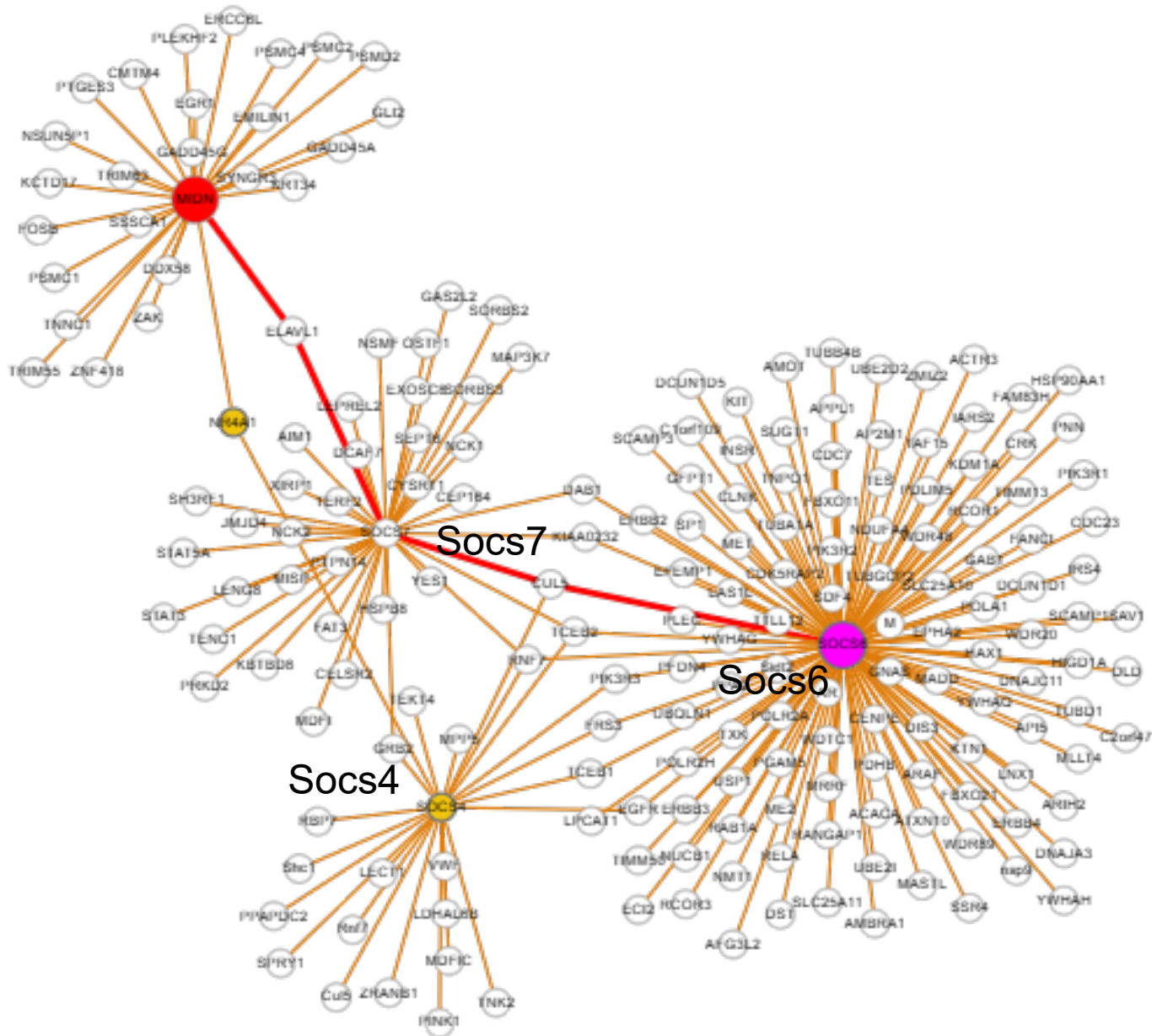
果蝇

Flybase Ortholog

Homo sapiens (Human) (9) +								
Species\Gene Symbol	Score	Best Score	Best Reverse Score	Source	Alignment	Complementation?	Transgene?	
				Compara eggNOG Hieranoid Homologene Inparanoid Isobase OMA OrthoDB OrthoFinder orthoInspector Panther Phylome RoundUp TreeFam ZFIN				
Hsap\SOCS4	7 of 15	Yes	Yes	● ● ○ ○ ○ ○ ○ ○ ○ ● ● ○ ● ○ ● ● ×	≡			1
Hsap\SOCS5	7 of 15	Yes	Yes	● ● ○ ○ ○ ○ ○ ○ ○ ● ● ● ● ○ ○ ● ×	≡			
Hsap\CISH	2 of 15	No	No	● ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×	≡			
Hsap\SOCS1	2 of 15	No	No	● ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×	≡			
Hsap\SOCS2	2 of 15	No	No	● ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×	≡			
Hsap\SOCS3	2 of 15	No	No	● ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×	≡			
Hsap\SOCS6	2 of 15	No	No	● ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×	≡			
Hsap\SOCS7	1 of 15	No	No	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×	≡			
Hsap\TNP2	1 of 15	No	Yes	○ ● ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×	≡			



Human MIDN & SOCS5
无直接互作证据



Human MIDN & Socs4/6/7可能通过其他蛋白介导互作

EsyN “Find a path”

Advanced tools

Interactions Network Statistics **Find a Path** Batch

Select Two Nodes and Find the Shortest Path Between Them ^[?]

Warning: If the network is changed after the calculation, the results shown might not be correct. ×

Advanced Options:

Select & Calculate View Shortest Path Details Network Parameters Reset Style

Results of pathway analysis

Number of nodes: 12; Number of edges: 28;

Source: MIDN

Target: SOCS5

Distance: 3

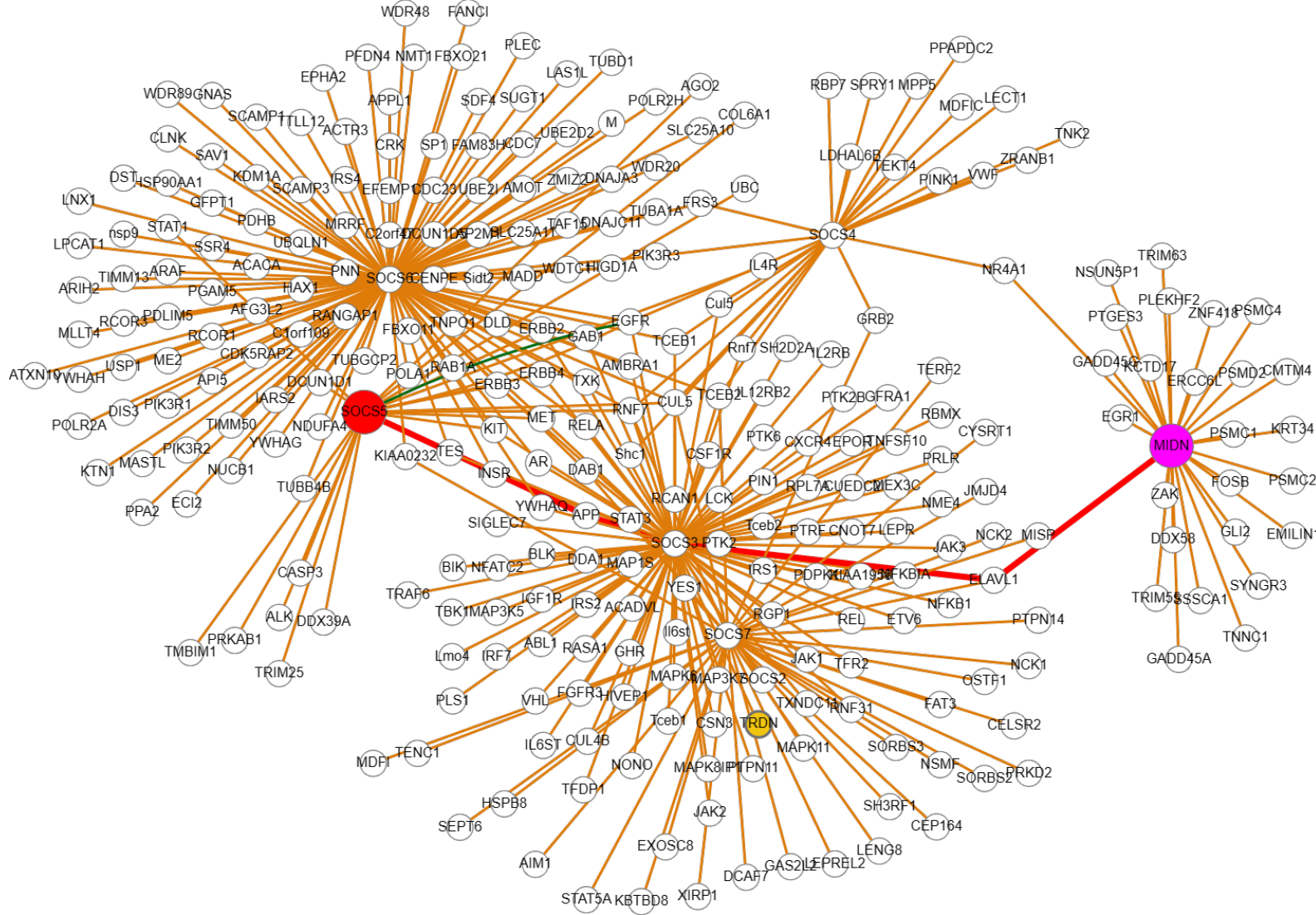
MIDN → DDX58

DDX58 → CUL5

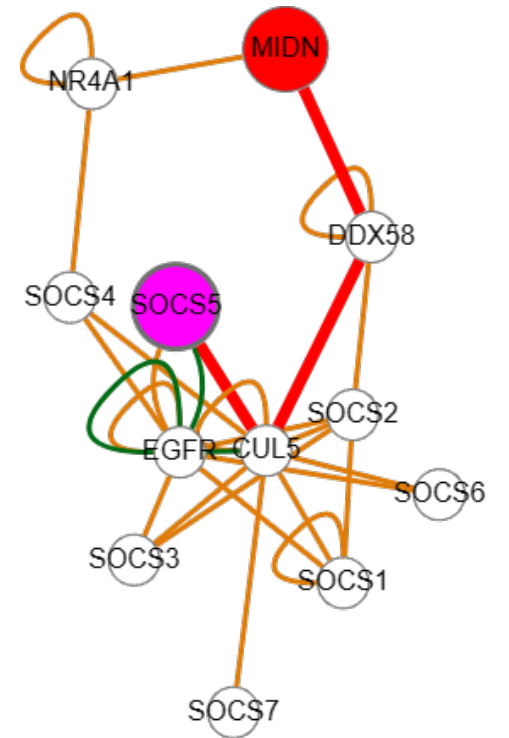
CUL5 → SOCS5

Socs5和MIDN可能存在间接互作

Input: MIDN, SOCS1, SOCS2, SOCS3, SOCS4, SOCS5, SOCS7, SOCS6



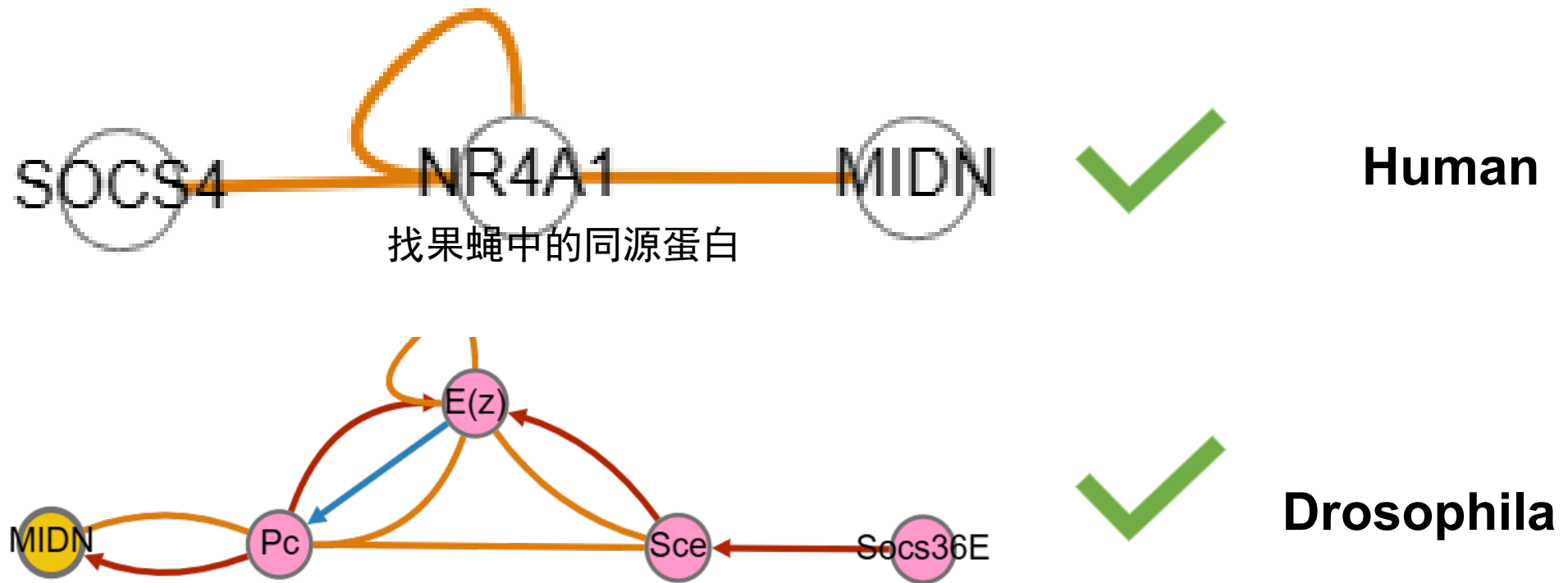
Input: MIDN, ELAVL, SOCS7, CUL5, SOCS6, NR4A1, DDX58, SOCS2, SOCS3, EGFR, SOCS5, SOCS1, SOCS4



但结果不一定可靠

总结：MIDN和Socs之间可能的调控模型

- 互作数据库中也不存在直接互作，但MIDN和Socs36E在果蝇中和人中可能有间接互作。



课程总结与收获

本学期学习了我们学习了门户网站，比如Uniprot, PDB的使用，生物信息工具如blast, SwissPDBViewer蛋白结构分析，MEGA系统发生等等。

将来需要的时候也可以在ABC网站 (<http://abc.gao-lab.org/>) 上找tools, 对我们的课题有一些指导, 给了我们很多启发。比如一般去flybase等专用数据库上查到基因信息, 比较全面, 可信度高; 遇到不懂的可以直接找官方帮助文档等。

谢谢！
请老师同学批评指正。