

計算系統生物學
Computational Systems Biology

電機系控制組
吳謂勝老師實驗室

生醫大數據分析處理

- 本實驗室培養數據科學家(data scientists)來做生醫大數據分析處理
- 數據科學家：運用數據和科學，創造新知識的人
- Ex1. 從生醫大數據mining出有用的生物知識(癌症的致病基因，癌症診斷生物標誌, ...)
- Ex2. 從財經大數據mining出可投資的標的

50 Best Jobs in America for 2020

Best Jobs 2020 United States

Share    

Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1 Front End Engineer	\$105,240	3.9/5	13,122	View Jobs
#2 Java Developer	\$83,589	3.9/5	16,136	View Jobs
#3 Data Scientist	\$107,801	4.0/5	6,542	View Jobs
#4 Product Manager	\$117,713	3.8/5	12,173	View Jobs
#5 DevOps Engineer	\$107,310	3.9/5	6,603	View Jobs
#6 Data Engineer	\$102,472	3.9/5	6,941	View Jobs
#7 Software Engineer	\$105,563	3.6/5	50,438	View Jobs
#22 Electrical Engineer	\$77,035	3.7/5	7,756	View Jobs

https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

50 Best Jobs in America for 2019

Best Jobs | 2019 | United States

Share |    

Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1 Data Scientist	\$108,000	4.3/5	6,510	View Jobs
#24 Electrical Engineer	\$77,000	3.7/5	7,191	View Jobs

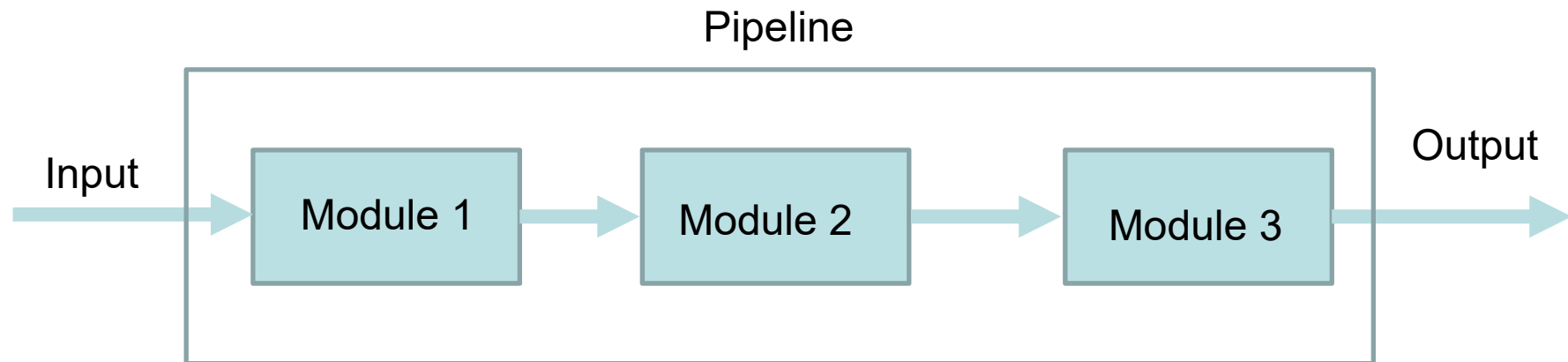


Data scientist is the top job in America for the fourth year in a row, according to Glassdoor's "50 Best

https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

數據科學家的技能

- 熟悉統計及資訊科學，寫程式建立pipeline來處理數據(**data**)以得到有用的知識(**knowledge**)
- 利用現有程式套件+自己寫的程式兜出pipeline



本實驗室訓練之程式技能

- 簡單Linux操作
- 熟悉各種生醫數據處理程式
- 用Python處理生醫數據
- 用Python Django架設網站
- 用Javascript在網頁裡做表格和畫圖

研究方向

- 生物問題演算法開發 (deep learning)
- 生物網頁工具開發, 生物資料庫建立

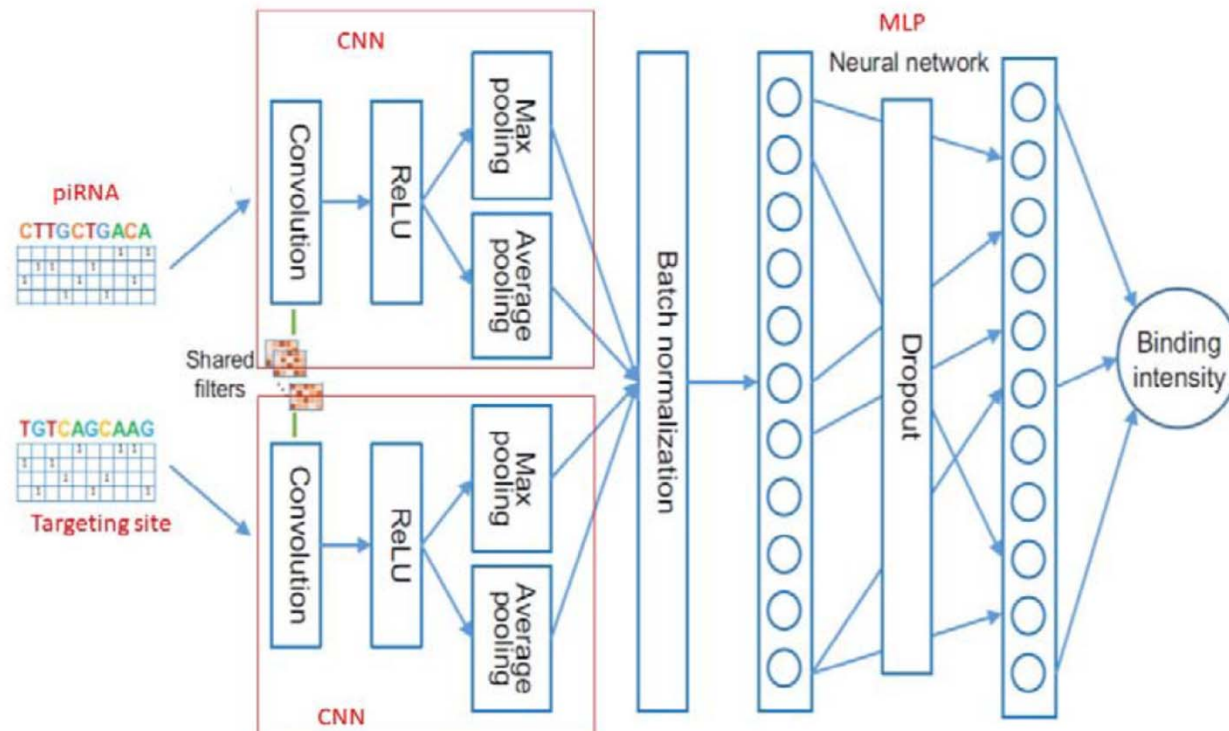
開發演算法來解答生物問題

piRNA

RNA

- 用 Deep learning 演算法來做分類問題


Ex. 預測 piRNA targets (結合 vs. 不會結合)



- 我們和芝加哥大學李亨啟教授合作，開發演算法偵測piRNA targets，在2018年2月刊登在Science期刊上，成為成大電資學院創院以來第一篇能榮登Science的研究！

AAAS ANNUAL MEETING Feb 15-19 2018 ★ Austin, TX REGISTER NOW >

Institution: NATL CHENG KUNG UNIVERSITY
Log in | My account | Contact Us
NATL CHENG KUNG UNIVERSITY



SHARE

REPORT



0



0

The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes

Donglei Zhang^{1,2,*}, Shikui Tu^{3,4,*}, Michael Stubna¹, Wei-Sheng Wu⁵, Wei-Che Huang⁵, Zhiping Weng³, Heng-Chi Lee^{1,†}

+ See all authors and affiliations

Science 02 Feb 2018:
Vol. 359, Issue 6375, pp. 587-592
DOI: 10.1126/science.aao2840



Science

Vol 359, Issue 6375
02 February 2018

Table of Contents
Print Table of Contents
Advertising (PDF)
Classified (PDF)
Masthead (PDF)

新聞報導

成大電機系教授吳謂勝開發利器 助線蟲研究大突破榮登《Science》

捷報！電機系吳謂勝教授與芝加哥大學國際合作團隊 研究成果躍登**Science**期刊

別小看線蟲 戲稱為「諾貝爾獎得主大恩人」



分享



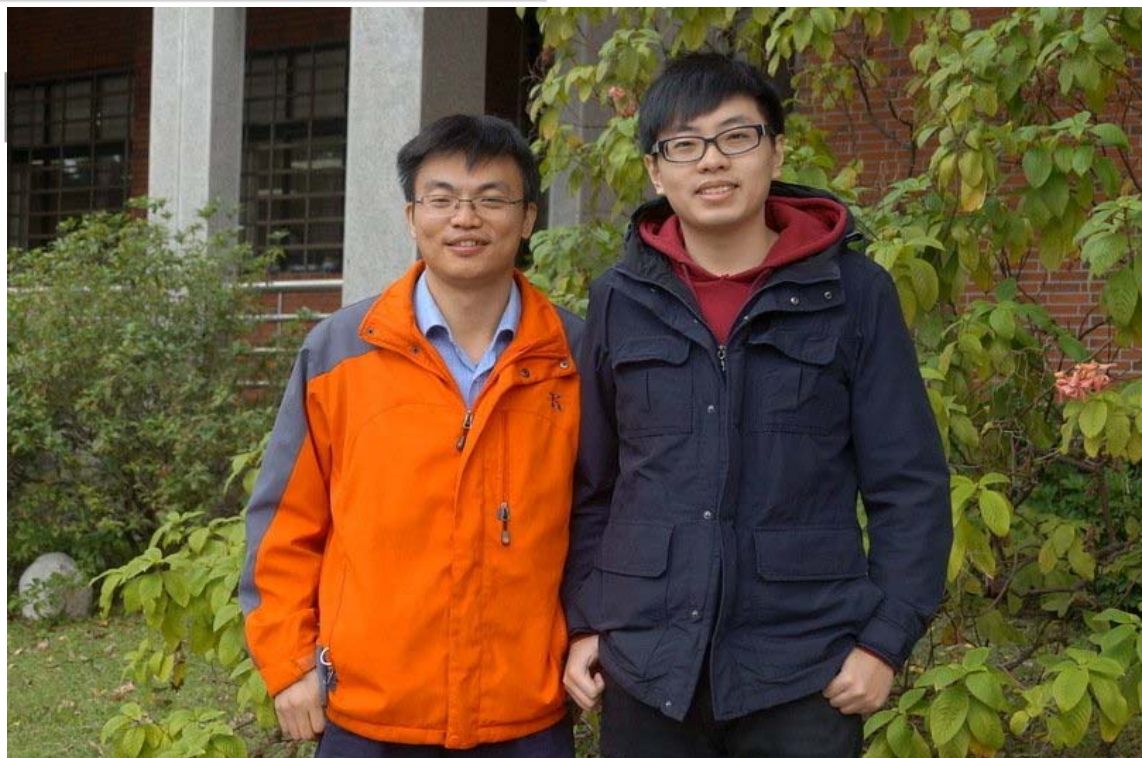
分享



留言



列印



校長頒獎



生物網頁工具及生物資料庫

- pirScan
- <http://cosbi4.ee.ncku.edu.tw/pirScan/>

- pirTarBase
- <http://cosbi6.ee.ncku.edu.tw/piRTarBase/>

Input page of pirScan

pirScan: a web server to predict piRNA targeting sites and to avoid transgene silencing

[Scan](#) [Tutorial](#) [Contact](#)

Scan piRNA target sites in your sequence

Input

Input your DNA (without introns) or spliced RNA sequence:

```
agttttacttttcgcttttcgATGGCACCTCCACAAGTAAGAAGTCCGCTAGGTTAAGCAAGAGA
TGCCAAGAAGAAAAGGTTAAGCTTCAGAAGAAAATGTCGGATTTAAGGCAAAATCTAA
GTCGGCTAAAAAGAGTAATAAGAAATCAAGAAAAGTCCGCTCAAAGACAAAGCCCAA
TTGACATCGTCCCACAACACGTGTGCTGTGACACAGACGTTTGCAAGGCTGATGCCTTGA
ACATTGACTACAAATCAGGTGACTGTGCGATGCTTGTCTCCGAAGGAGGTTTCCTGT
GAATGTCAAGAGAAAATTGTGGCACATTCTTACCGCCAACCATTTACCATCATCAAAATC
GCGTTGGCTCAGTTCATGCTCATTGGGGAAGCAACTCGAAAGAAGGATCCGAGCACTTT
TTGGACGGAAAACAACCTAGCGGAGAGGTTCACTTTGTATTCTGGAACACCAGCTATGAG
TCGTTAATGTGGCACTCAGCAAGCCCGATGGATTGGCGGTTGTGGAGTCTTCTGAAGG
AAGCAAAATGCAATGCAATTAGGATCGGCTGATGGACAGACTGGCCAAAGCCGACCA
```

Specify your sequence name:

Specify coding sequence (CDS) region:
 Whole input sequence 23 - 865 None

piRNA targeting rules

Default Setting

Number of mismatches allowed at seed region:

- number of non-GU pairs \leq
- number of GU pairs \leq

Number of mismatches allowed at non-seed region:

- number of non-GU pairs \leq
- number of GU pairs \leq

Total number of mismatches at seed & non-seed regions \leq

piRNA targeting score \geq

5' **input sequence** 3'

3' **piRNA #1** 5' 3' **piRNA #2** 5'

5' **piRNA** 3'

seed non-seed

1 2 7 8 21

Output page of pirScan

8 Identified piRNA target sites (Graphical View)



8 Identified piRNA target sites (Table View)

[Download table](#)

Show 10 entries

Legend: non-GU mismatch (yellow), GU mismatch (blue), mismatch at the 1st position of piRNA (green), seed region (red)

piRNA	piRNA targeting score	targeted region in input sequence	# mismatches	position in piRNA	# non-GU mismatches in seed region	# GU mismatches in seed region	# non-GU mismatches in non-seed region	# GU mismatches in non-seed region	pairing (top:Input sequence, bottom:piRNA)
21ur-11812	3	141-161	5	1,5,14,15,20	0	1	2	1	5' GUAAUAAGAAAUUC AAGAAA G 3' 3' CGUUAUAAUUUAAG UUUUUU U 5'

8 Identified piRNA target sites (Sequence View)

[Download seqView](#)

Lowercase/Uppercase text indicates UTRs/CDS

Legend: non-GU mismatch (yellow), GU mismatch (blue), mismatch at the 1st position of piRNA (green), seed region (red), Codon (green)

1 aguuuuacuuuuucgcuuuucgAUGGCACCUCCACAAGUAAGAAGGUCCGCUAGGUUAAGCAAGAGAUGCCAAGAAGAAAAGGUUAAGCUUCAGAAAGAA 100
S K R C Q E E K K L Q K K
UUUCAUAUGGUUUUUUUU CGAAUUUUUUUU

歡迎有興趣來本實驗室做
專題的同學來找我聊聊

- 系統生物學是要**定量**研究**整個生物系統**的特性，而不是像傳統生物學家只研究生物系統裡面的某個部分，導致研究落入見樹不見林的窘境。

Ex. 藥物副作用

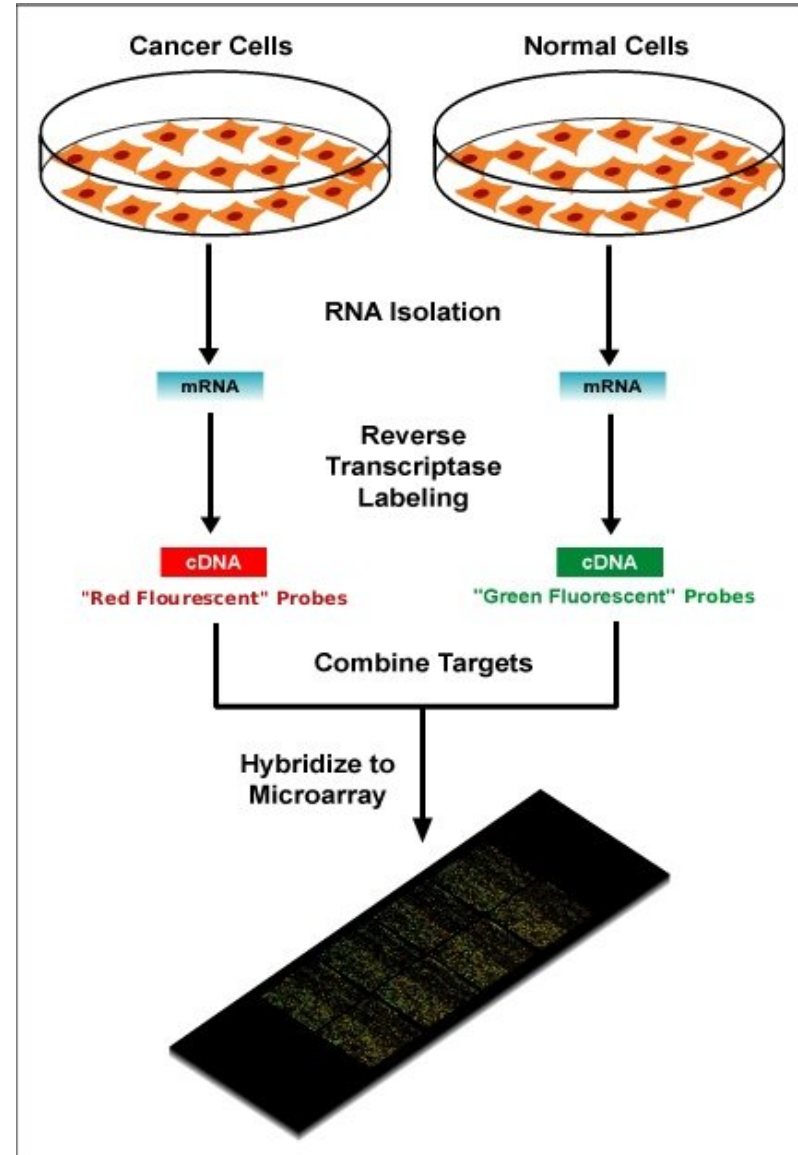
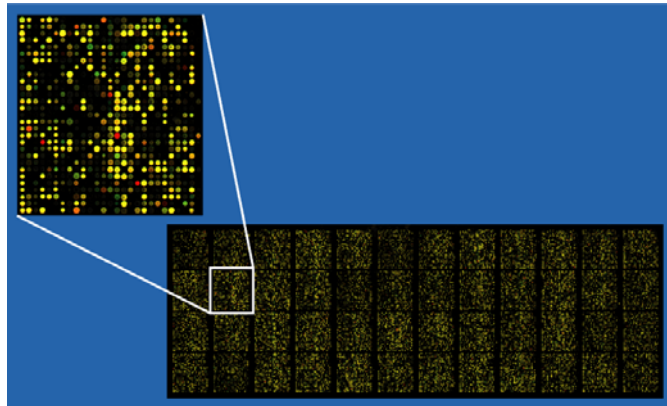
- 計算系統生物學的目標是要對生物系統建立**數學模型**，讓我們可以**解釋及預測生物系統的行為**。這個數學模型還必須通過生物實驗的驗證，才算是一個有實用價值的模型。

- 生物系統是一個非常複雜的**非線性系統**，因此唯有透過研究系統裡面的組成份子之間複雜的**交互作用**，才能瞭解為什麼生物系統**有能力執行各種不同的功能**。
- 各種**高通量生物實驗技術**所產生的數據（如基因體序列資料、基因表現之mRNA資料、蛋白質交互作用資料），都是**對於生物系統的量測結果**，可以當成我們建立數學模型所需的資料。

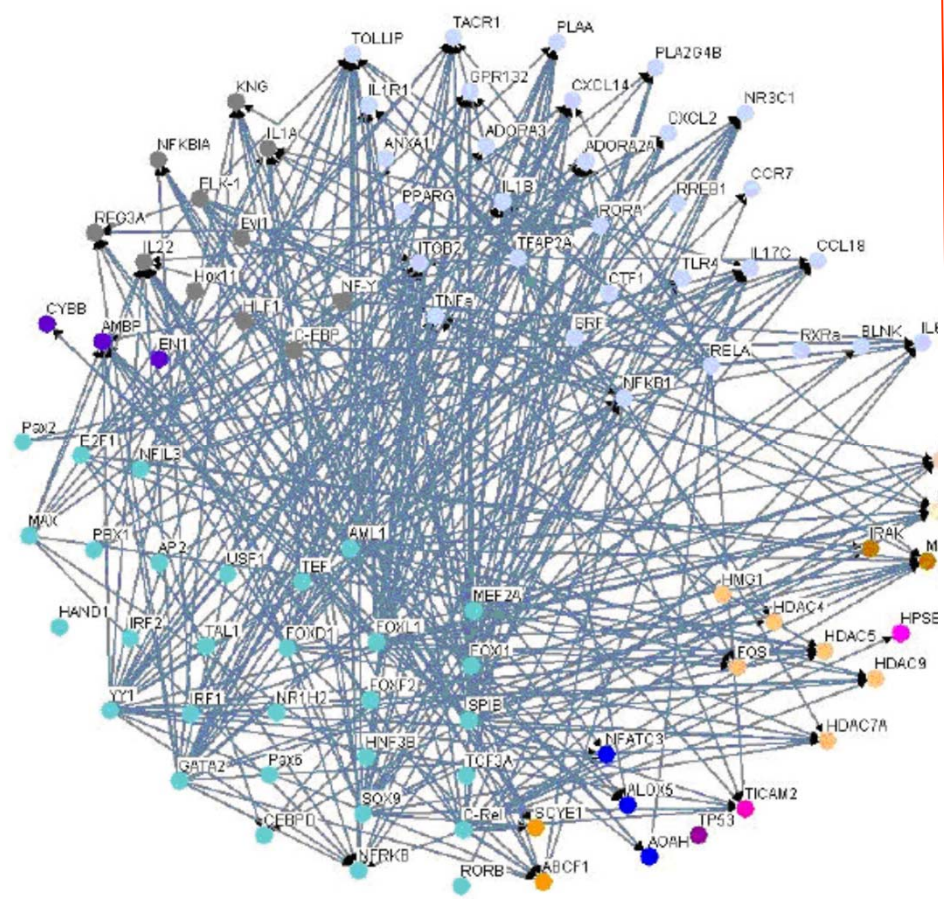
- 電機系控制組所學的非線性系統理論正是最適合用來建立複雜生物系統數學模型的工具。
- 藉由系統理論，我們可以分析生物系統的諸多系統特性（如穩定性、強健性）。當我們可以成功分析生物系統特性之後，再來就是要去控制生物系統，使生物系統能夠擁有我們想要的特性，變成有醫療或是經濟上的價值。
- 希望有學習動機的學生可以加入我們的行列，共同來解開生物系統的運作之謎。

Example 1:

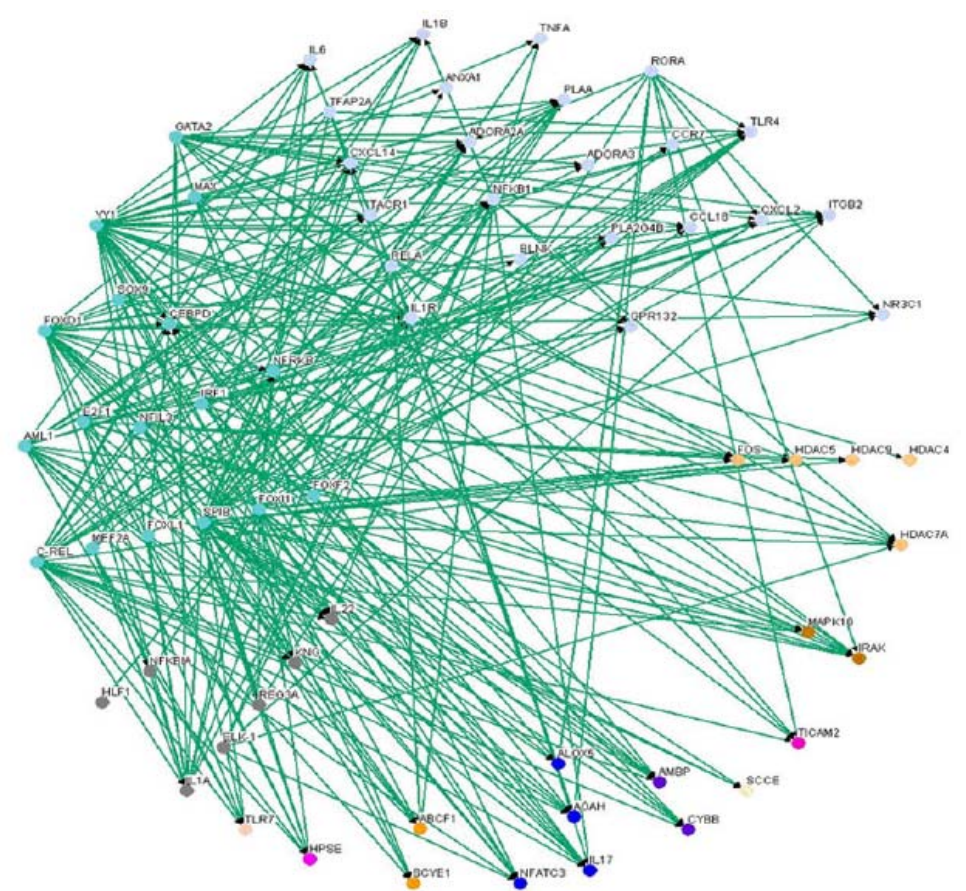
Using microarray to study cancer



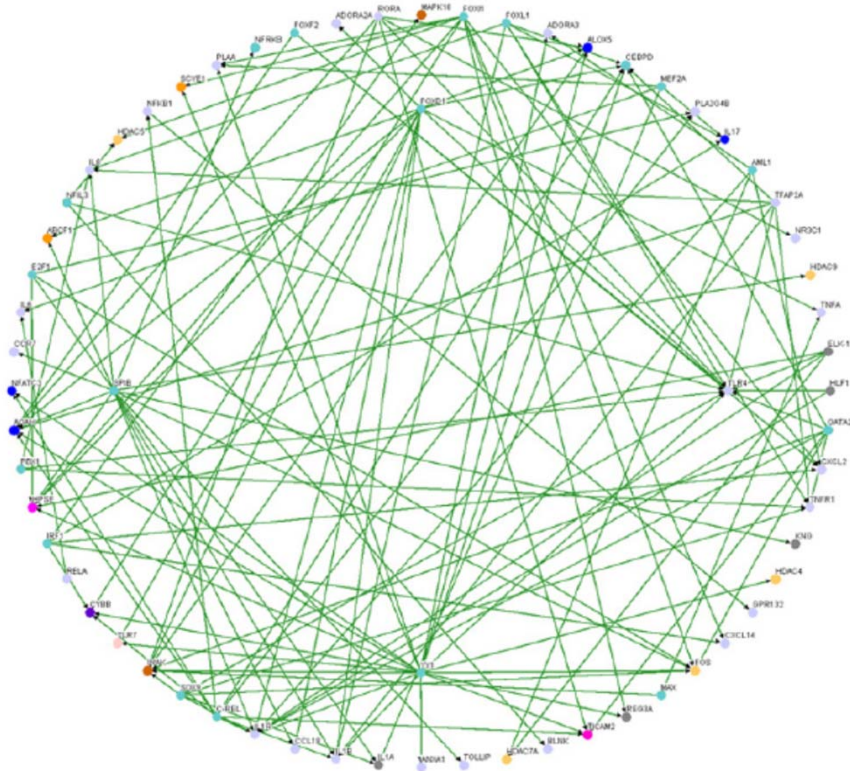
Using mathematical models to reconstruct gene regulatory networks in **normal cell**



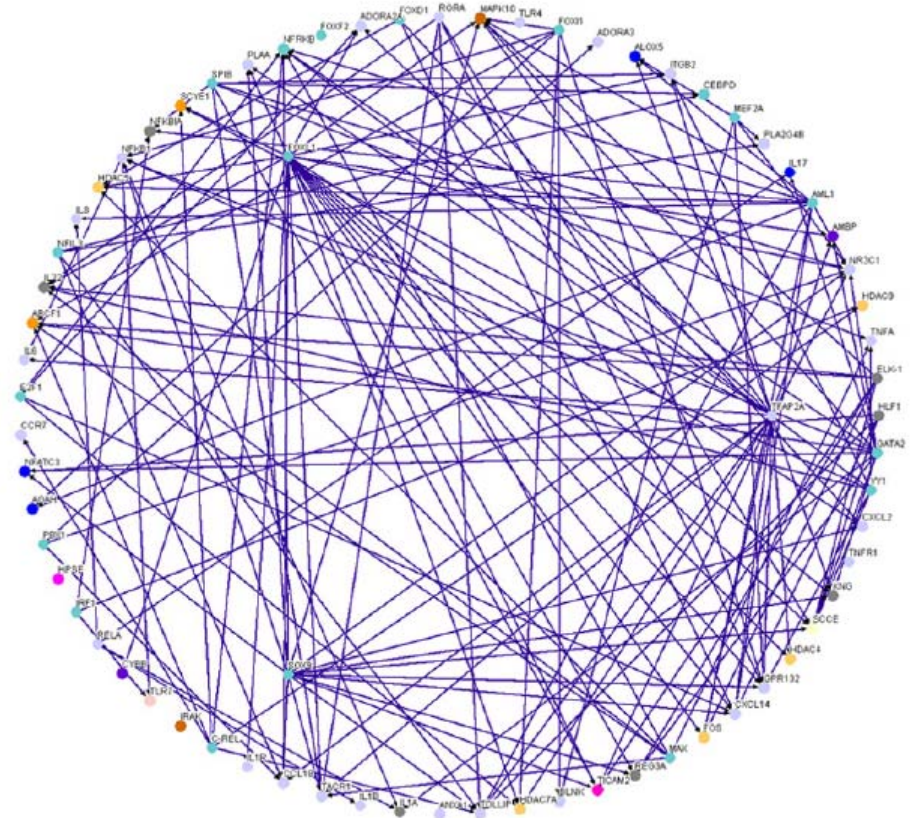
Using mathematical models to reconstruct gene regulatory networks in **cancer cell**



Edges appear in the normal cell but not in the cancer cell

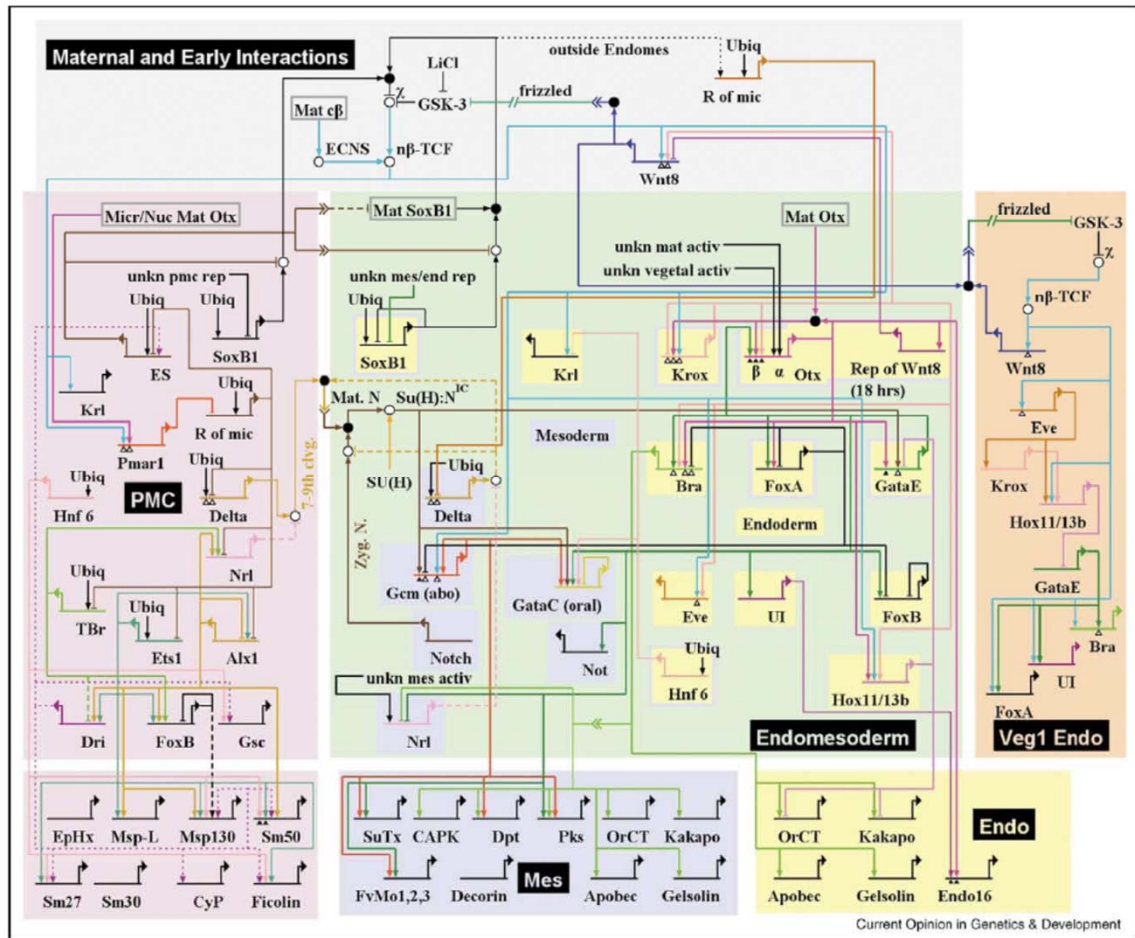
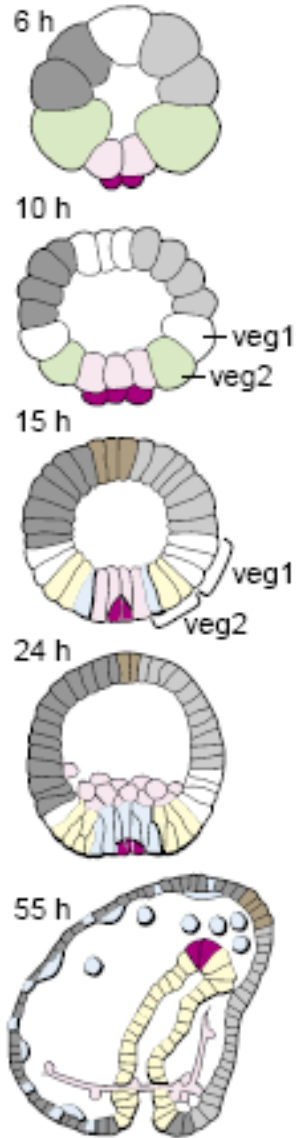


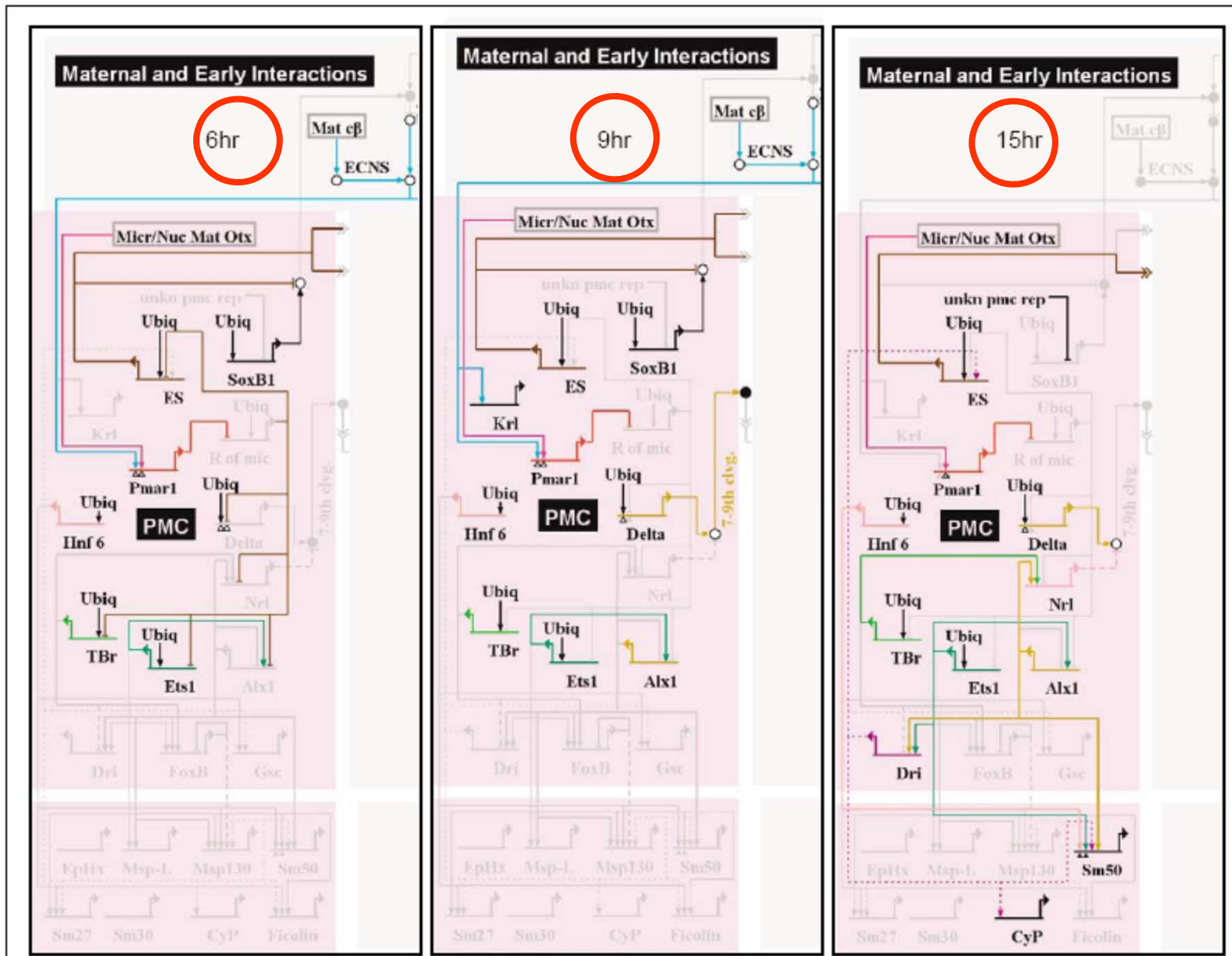
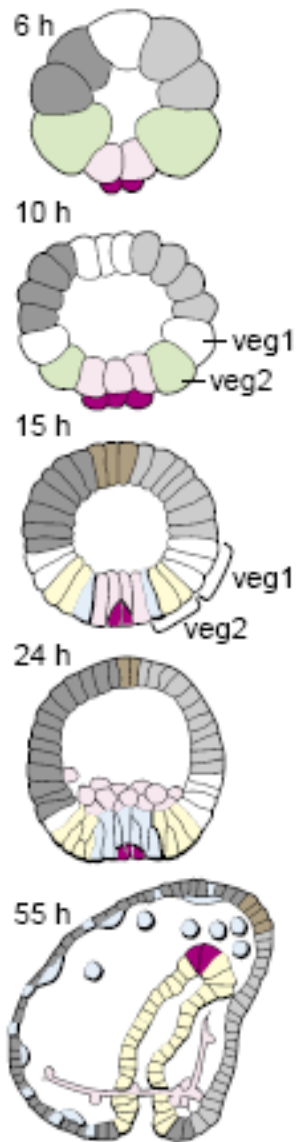
Edges appear in the cancer cell but not in the normal cell



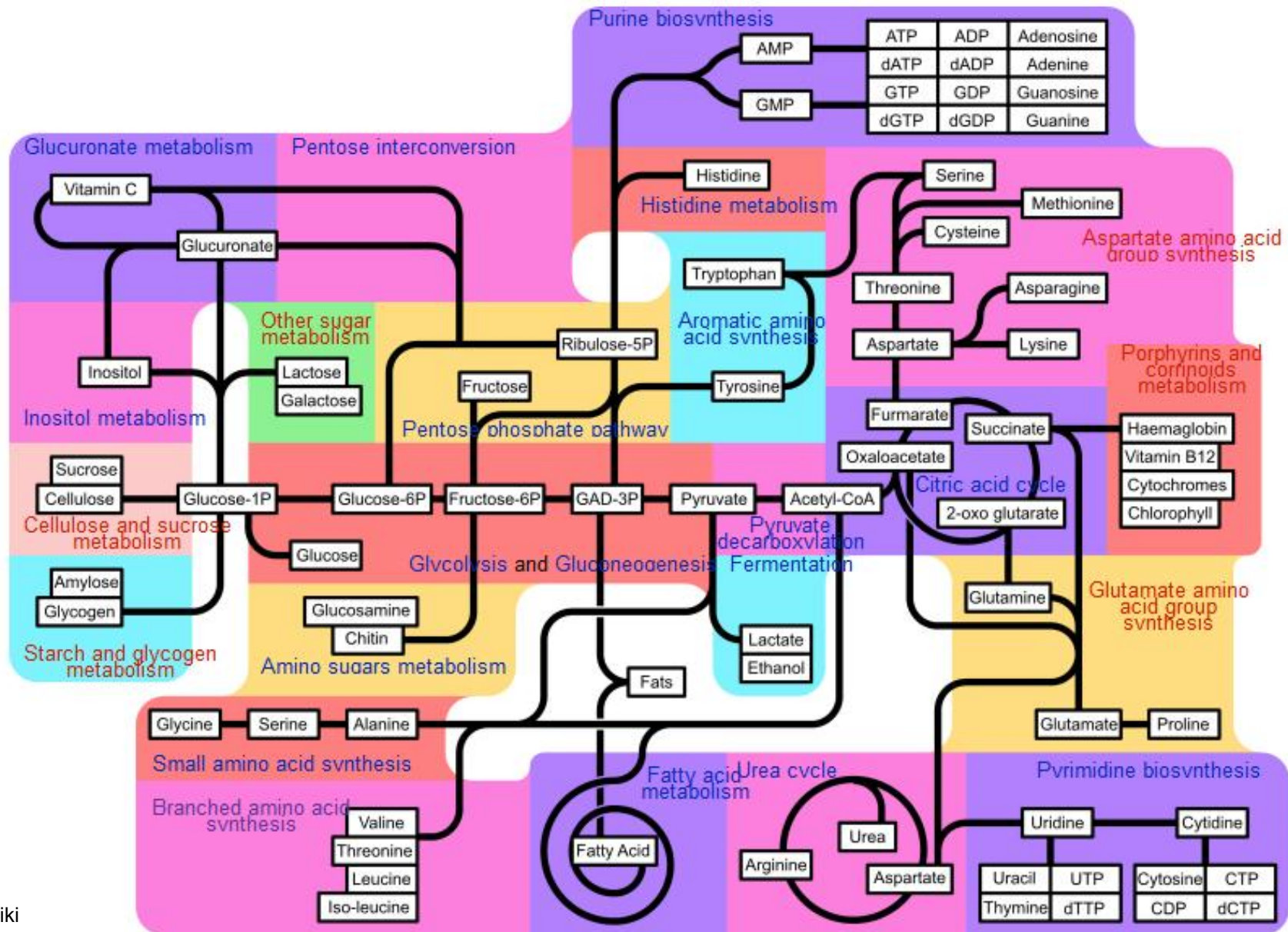
These edges suggest candidate **gene-gene interactions which may be related to cancer formation**

Example 2: Gene regulatory networks in Sea Urchin embryo early development





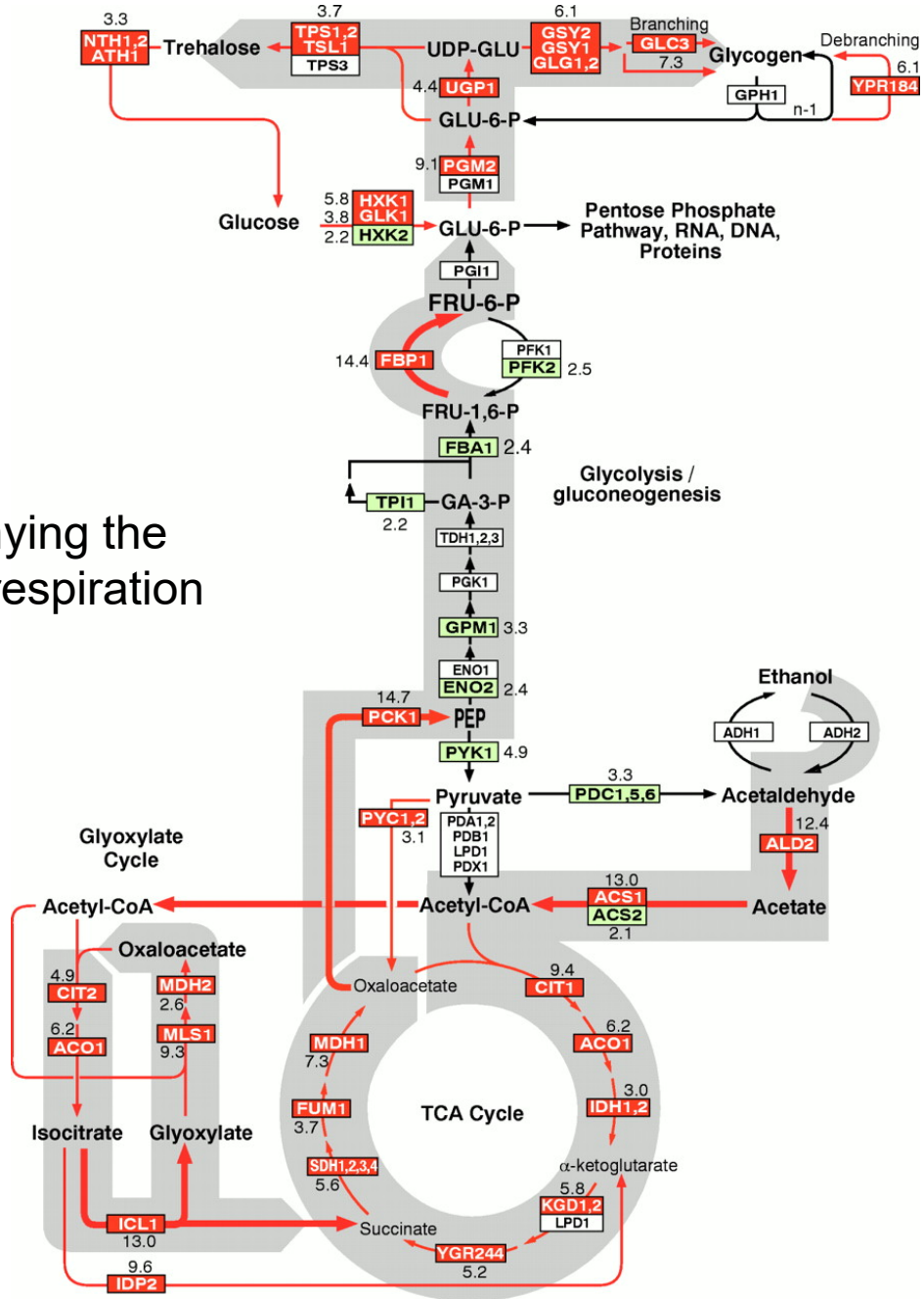
Example 3: Metabolic networks



from Wiki

Carbohydrate Metabolism

gene expression changes accompanying the metabolic shift from fermentation to respiration



from Science 1997, 278: 680-686

Example 4: Bio-database construction

YPA

Home Search Help Statistics About us Contact

YPA (Yeast Promoter Atlas) is a repository of promoter features in *Saccharomyces cerevisiae*. It integrates various resources (including promoter sequences, TSSs, TATA boxes, TFBSs, nucleosome occupancy, DNA bendability, TF-TF interaction, and gene expression data) and provides a comprehensive view of the promoter regions. Please [check the settings](#) before using YPA.

Query YPA

[Go](#) (ex: YBL087C, RPL23A, or RPL*)
[Advanced search](#)

Browse YPA

By genomic coordinate

[Go](#) (ex: II:60k, II:60000, 2:60000, YBL087C or RPL23A)

By chromosome

[I](#) [II](#) [III](#) [IV](#) [V](#) [VI](#) [VII](#) [VIII](#) [IX](#) [X](#) [XI](#) [XII](#) [XIII](#) [XIV](#) [XV](#) [XVI](#)

By regulatory element

[TFBS \(transcription factor binding site\)](#)

Download YPA

[Go](#) (ex: II:60k~II:61k or YBL087C or RPL23A)

How to cite YPA

[Darby Tien-Hao Chang](#), [Cheng-Yi Huang](#), [Chi-Yeh Wu](#), and [Wei-Sheng Wu](#), "[YPA: an integrated repository of promoter features in *Saccharomyces cerevisiae*](#)", *Nucleic Acids Res.* 2011 39: D647-D652.

YPA is selected as a **featured article**—the **top 5%** of papers in terms of originality, significance and scientific excellence—of [NAR](#).

News

YPA version 1.5 is released

This release recovered 485 TATA boxes hidden in the revision of [YPA 1.1](#) and updated interaction data. [more...](#)

YPA version 1.4 is released

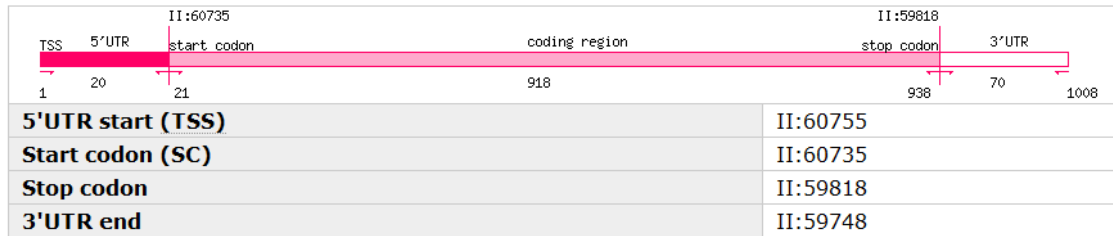
This release updated interaction data and provided information of TFs without TFBS information. [more...](#)

YPA version 1.3 is released

This release updated interaction data. [more...](#)

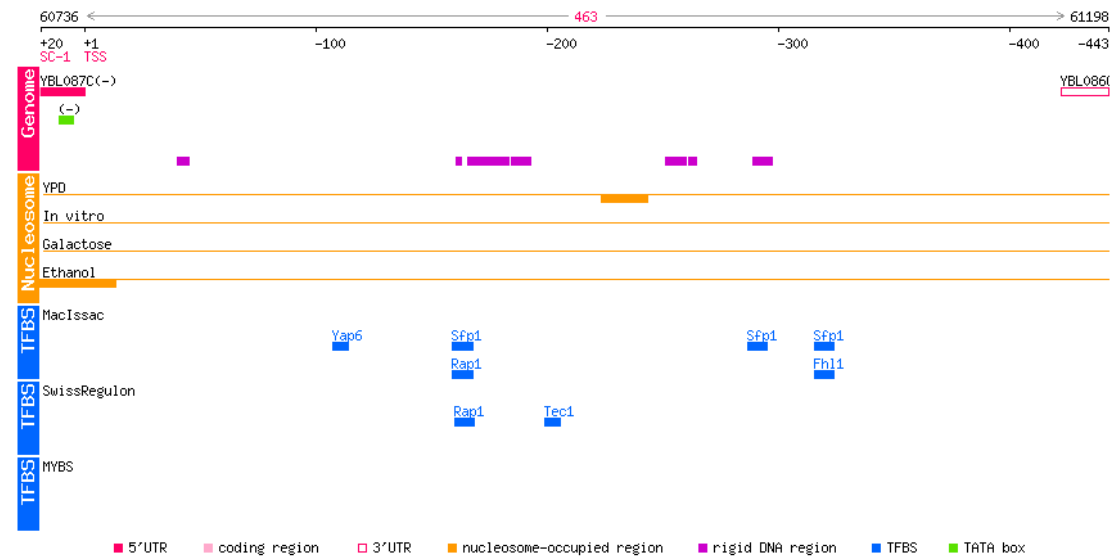
[previous news...](#)

ORF organization



[Back to top](#)

Promoter map



Sequence map

```

-443 AATATTTGTA TTGTTTTATG TAATTATAAT CTTAAGCTTG CAATACCAAT CACTATTACC ATGTTAATGA TCCCCAGAT AAAAAGACAA GACAGAGCAT -344
-343 ATATCAGACAT ACATTAAACAT CTGTACACCC CCAACTGCTT TGA AAAAACCC CATGCAGTTA TTATTTTCAA CAGACCAAGT TTTTTCATCG CTTATCCGGC -244
      Sfp1 MI
      Phl1 MI
-243 TCACAGAACG AGCCGTACAA CCGTCATCAG ACTTGCATGG AAATGTTGAA AAAGAGGTAA AAAAGAAAAT ATCAAAACCA AACACGCTCG GGCTGAAAAA -144
      Tec1 SR
      Rap1 SR
      Sfp1 MI
      Rap1 MI
-143 TTAATATGCG AGGAACGAA CTGTAGCCAT AATGTACCTG TTCCCGCTAA TAGTAATACC ATCGTATTGC TAAGGTTAGT AGCTAAAAT CTCTTACTTT -44
      Yap6 MI
-43 CTTTATTTTG ACTCTTAGGT CTTACAAGCA ATACAAAACC AACACACCTA TATATACTA ATA start codon 20
    
```

Example 5: Bio-server construction

Gene sets for comparison

Gene set I

```
YPR189W YPR048W YPL263C YPL160W
YPL064C YOR224C YNR027W YMR049C
YML074C YLR244C YHR207C YER163C
YER018C YDR197W YDR196C YDR031W
YDL101C YDL074C YCR079W YCR072C
```

Miscellaneous

- FDR
- GO term
- [Load sample gene sets to analyze](#)
- [See the result of iESR genes vs. rESR genes](#)

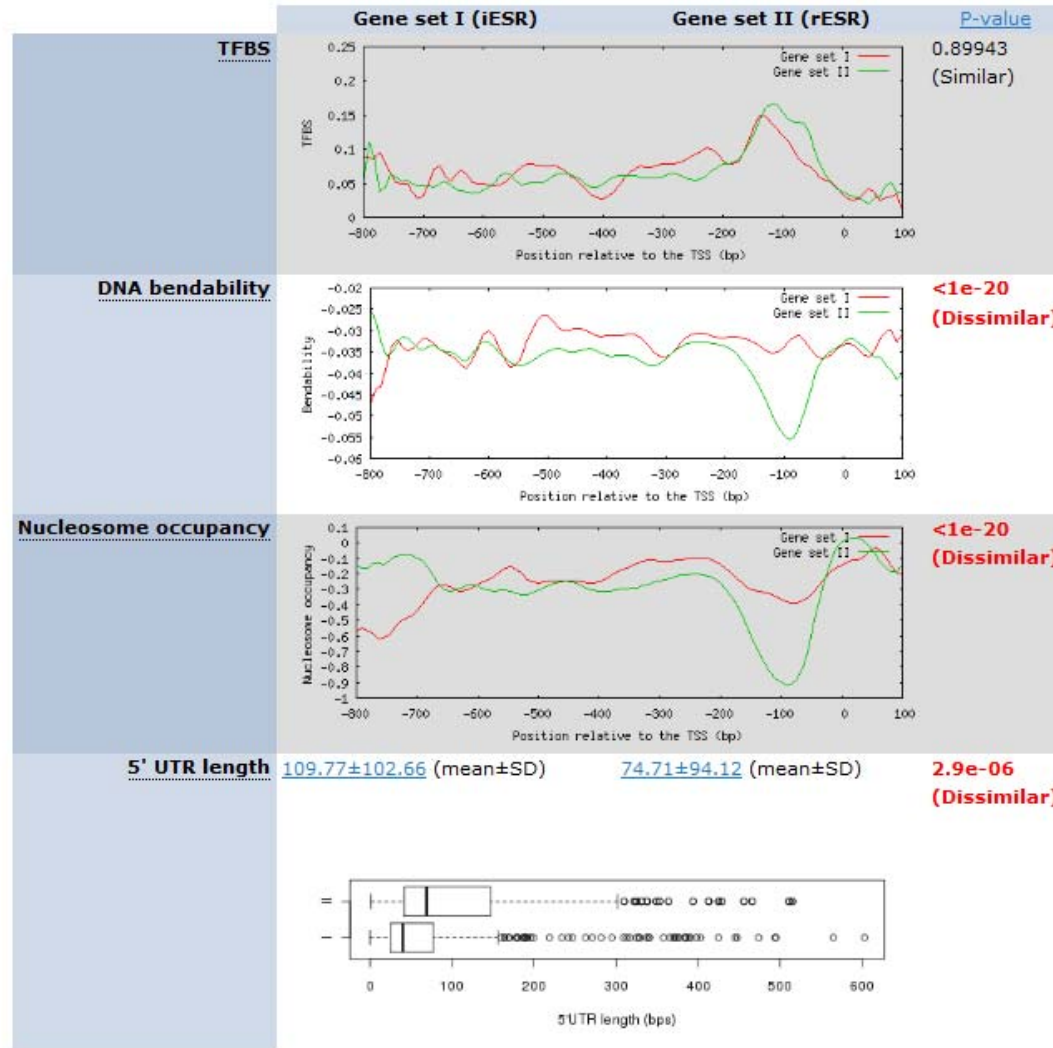
Gene set II

None
 User specified

```
YKL121W YLR001C YNL183C YNL241C
YCR004C YML004C YKL065C YLR290C
YKL067W YGR209C YDR453C YML131W
YNL134C YOL151W YOL150C YKR011C
YKL103C YDL124W YMR173W-A YMR173W
```

TATA ([Basehoar et al. 2004](#))
 TATA-less ([Basehoar et al. 2004](#))
 cell cycle ([Spellman et al. 1998](#))
 essential ([Yeast Genome Deletion Project](#))
 iESR ([Gasch et al. 2000](#))
 rESR ([Gasch et al. 2000](#))
 OPN ([Genome Res. 2008](#))
 DPN ([Genome Res. 2008](#))

Analysis summary



	Gene set I (iESR)	Gene set II (rESR)	Commom	P-value
TFB	<u>16</u>	<u>7</u>	<u>0</u>	4.1e-06
TFR	<u>78</u>	<u>15</u>	<u>10</u>	0.00010
PPI	<u>8</u>	<u>556</u>	<u>1</u>	2.3e-15
Pathway	<u>15</u>	<u>7</u>	<u>0</u>	5.9e-06
Functional group	<u>4</u>	<u>4</u>	<u>0</u>	0.02857
Gene ontology				
BP	<u>55</u>	<u>102</u>	<u>0</u>	1.0e-43
CC	<u>17</u>	<u>43</u>	<u>0</u>	2.6e-15
MF	<u>20</u>	<u>33</u>	<u>0</u>	4.9e-15

Summary

- 本實驗室兩大研究方向
 1. 系統生物學演算法開發：可學到生物知識，系統模型建立，演算法設計，程式寫作 (Python, Matlab)
 2. 網頁資料庫或網頁伺服器開發：可學到生物知識，演算法實現，程式寫作 (Python, GD library, Java script, R, Html, ...)

可獲得之能力

- 生物知識
- 演算法
- 程式語言能力
- 口頭報告技巧
- 英文論文寫作

研究成果

- 本實驗室專精生物網頁資料庫及生物資訊網頁工具開發，每個碩士畢業生皆有能用Python程式語言做出一個專業的網站。本實驗室研究成果豐碩，每年皆可發表三篇以上領域排名前20%的SCI papers。本實驗室的代表作為跟芝加哥大學老師合作的研究，在2018年2月刊登在Science期刊上，成為成大電資學院創院以來第一篇能榮登Science的研究！




AAAS ANNUAL MEETING
Feb 15-19 2018 ★ Austin, TX

REGISTER NOW >

Institution: NATL CHENG KUNG UNIVERSITY
Log in | My account | Contact Us

NATL CHENG KUNG UNIVERSITY



SHARE

REPORT



0



0

The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes

Donglei Zhang^{1,2,*}, Shikui Tu^{3,4,*}, Michael Stubna¹, Wei-Sheng Wu⁵, Wei-Che Huang⁵, Zhiping Weng³, Heng-Chi Lee^{1,†}

+ See all authors and affiliations

Science 02 Feb 2018:
Vol. 359, Issue 6375, pp. 587-592
DOI: 10.1126/science.aao2840



Science

Vol 359, Issue 6375
02 February 2018

Table of Contents
Print Table of Contents
Advertising (PDF)
Classified (PDF)
Masthead (PDF)