

# GO解析

東京科学大学 ILA国府台  
中林 潤

JH人材育成課 バイオインフォマティクソン育成講座



# Gene Ontology解析

## GO term

- 生物種の異なるデータベース間の比較を行うために定義された生物学的な統一用語。
- Biological process, Cellular component, molecular functionの3項目に大別される。
- 階層構造を持っている。

GO解析：遺伝子リストの中に特定のGOタームを持つ遺伝子が高頻度に存在しているか判定して、その機能を推定する



# Gene Ontology Consortium

The screenshot shows the Gene Ontology Consortium website. At the top, there is a navigation menu with links for Home, Documentation, Downloads, Community, Tools, About, and Contact us. The main content area is divided into several sections:

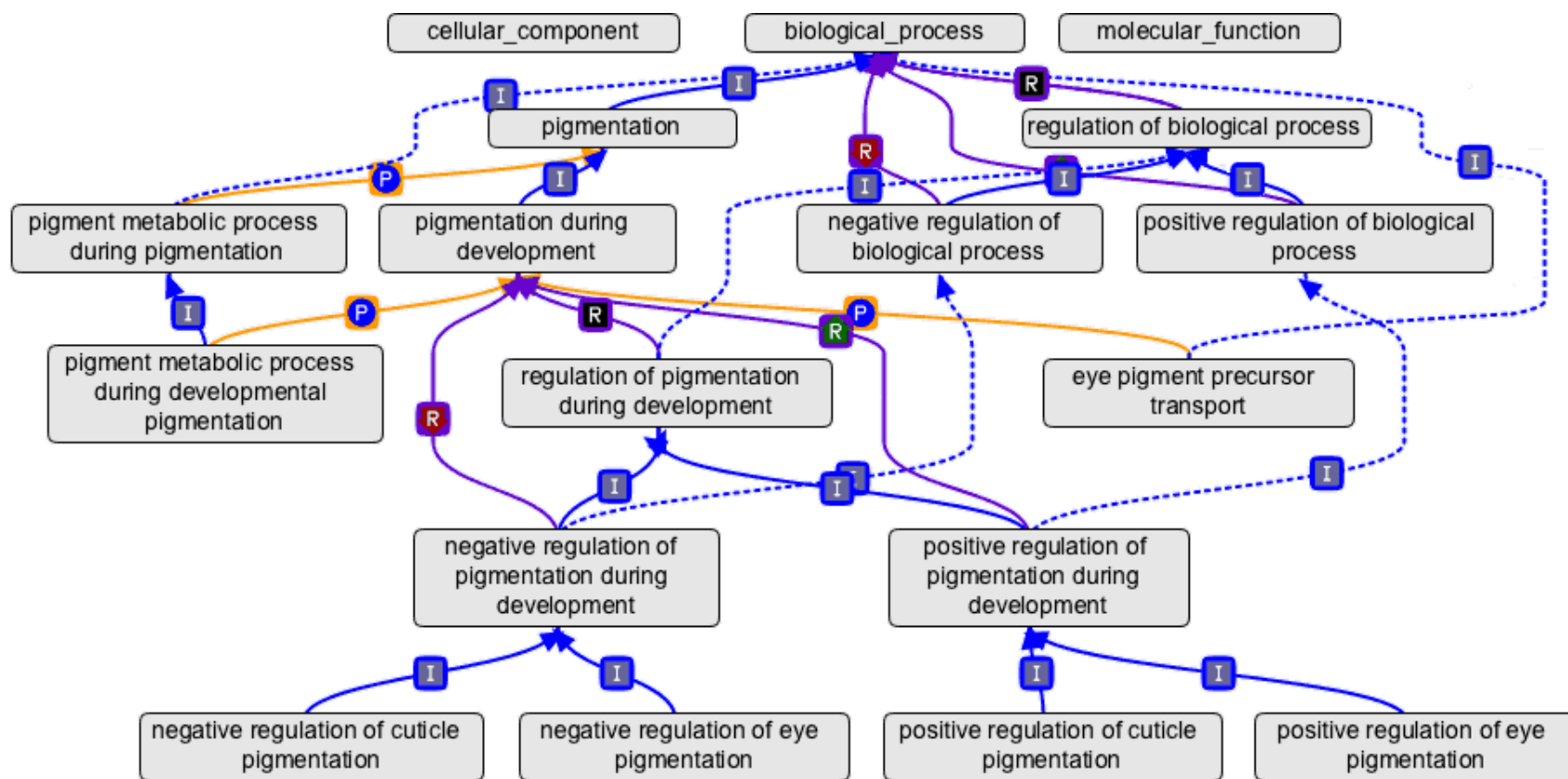
- Enrichment analysis:** A sidebar on the left containing a list of gene symbols (CDH6, MT2A, EGFR, TMCO1, GLI3R1), a dropdown menu for 'biological process', a selection for 'Homo sapiens', and a 'Submit' button. Below this is a 'Help' section powered by PANTHER and a 'Statistics' section with a bar chart.
- Other GOC tools:** A section for exploring other GOC tools in the AmiGO software suite.
- Tweets:** A section for social media updates, including a tweet from @UCLgenomics.
- Gene Ontology Consortium:** The main header section with a search bar for GO data and a 'Search' button.
- Ontology:** A section describing the framework for the model of biology, defining concepts and relationships.
- Annotations:** A section for downloading annotations and filtering them.
- Search documentation:** A search bar for documentation.
- User stories:** A section for exploring documentation related to personal user stories.
- What is the Gene Ontology?:** A section with a list of links: 'An introduction to the Gene Ontology', 'What are annotations?', 'Ten quick tips for using the Gene Ontology', 'Enrichment analysis', and 'Downloads'.
- Recent news:** A section listing recent newsletters and news items with their post dates.
- Tags:** A section for navigation, listing 'navigation' and 'User story: Everybody'.

At the bottom of the page, there is a copyright notice: 'Copyright © 1999-2017 the Gene Ontology (CC-BY 4.0)'. Below this are links for 'Helpdesk', 'Citation/attributions', 'Terms of use', and 'RSS'. A footer note states: 'Member of the Open Biological and Biomedical Ontologies. The Gene Ontology Consortium is supported by a P41 grant from the National Human Genome Research Institute (NHGRI) [grant 5U41HG002273-14]. The Gene Ontology Consortium would like to acknowledge the assistance of many more people than can be listed here. Please visit the acknowledgements page for the full list.'

<http://www.geneontology.org>

# GO tree

## 3つの大分類



大まかなターム



詳細なターム



# Fisherの正確確率検定

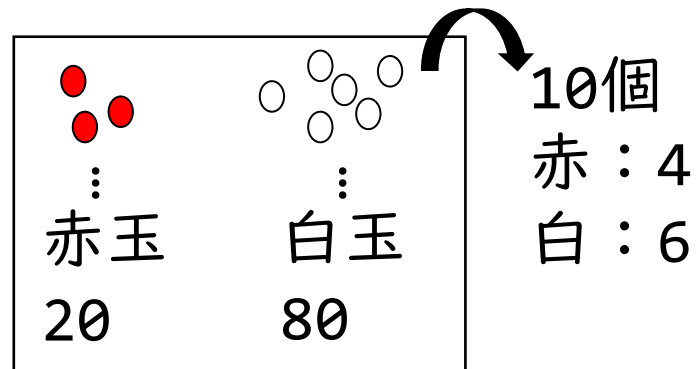
ある集団のある変数に出現頻度の偏りがあるか判定する方法

2×2 クロス集計表

|    | 性質1 | 性質2 |
|----|-----|-----|
| A群 | a   | b   |
| B群 | c   | d   |
|    | g   | h   |

e  
f  
n

$$P_a = \frac{{}^e C_a \times {}^f C_c}{{}^n C_g} = \frac{e! f! g! h!}{n! a! b! c! d!}$$



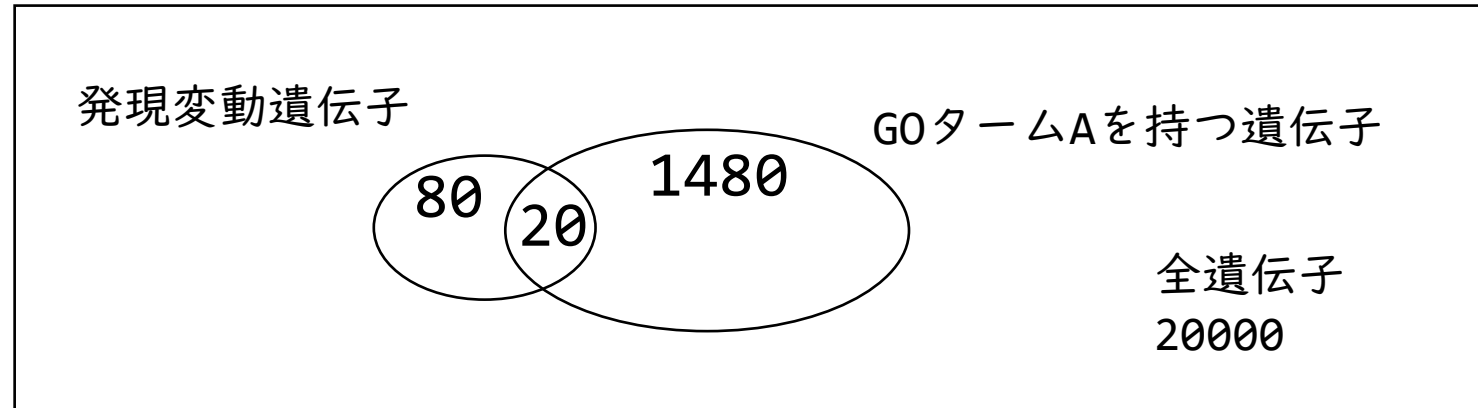
計100個

|   | 取った | 残り |     |
|---|-----|----|-----|
| 赤 | 4   | 16 | 20  |
| 白 | 6   | 74 | 80  |
|   | 10  | 90 | 100 |

$$p = 0.0841073$$



# GO解析のクロス集計表



全遺伝子 : 20000  
GOタームAを持つ遺伝子 : 1500  
発現変動遺伝子 : 100

|              | 発現変動遺伝子 | 残り    |       |
|--------------|---------|-------|-------|
| GOタームAを持つ遺伝子 | 20      | 1480  | 1500  |
| その他の遺伝子      | 80      | 18420 | 19900 |
|              | 100     | 18500 | 20000 |

$p = 0.00004509$



# GO解析のクロス集計表

全遺伝子 : 20000  
GOタームAを持つ遺伝子 : 1500  
発現変動遺伝子 : 100

$$p = 0.3379$$

|                  | 発現変動<br>遺伝子 | その他   |       |
|------------------|-------------|-------|-------|
| GOタームAを<br>持つ遺伝子 | 10          | 1490  | 1500  |
| その他              | 90          | 18410 | 18500 |
|                  | 100         | 19900 | 20000 |

全遺伝子 : 20000  
GOタームBを持つ遺伝子 : 100  
発現変動遺伝子 : 100

$$p = 0.0001461$$

|                  | 発現変動<br>遺伝子 | その他   |       |
|------------------|-------------|-------|-------|
| GOタームBを<br>持つ遺伝子 | 5           | 95    | 100   |
| その他              | 95          | 19805 | 19900 |
|                  | 100         | 19900 | 20000 |



# $\chi^2$ 検定 (独立性の検定)

2×2 分割表

|    | 属性①      | 属性②      |
|----|----------|----------|
| A群 | $N_{11}$ | $N_{12}$ |
| B群 | $N_{21}$ | $N_{22}$ |

$$N_{11} + N_{12} = N_A$$

$$N_{21} + N_{22} = N_B$$

$$N_{11} + N_{21} = N_1$$

$$N_{11} + N_{21} = N_2$$

$$N_{11} + N_{12} + N_{21} + N_{22} = N_T$$

「A群かB群か (行)」と「属性①か属性②か (列)」が関係ない



行と列が独立

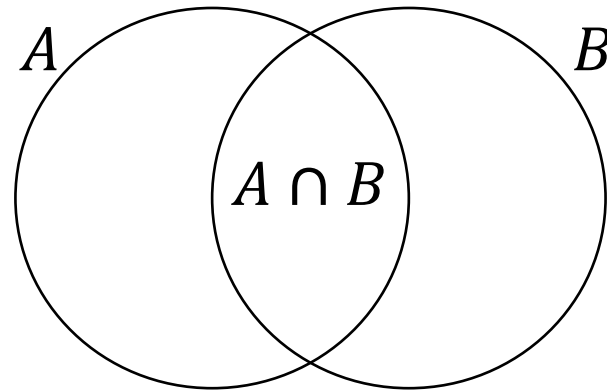
帰無仮説：行と列は独立である

対立仮説：行と列が独立でない



# 独立な事象の同時確率

- 積事象（結合事象） $A \cap B$



- 独立な事象：事象AとBは無関係

$$A \cap B = P_A \times P_B$$

コインとサイコロを投げて、コインは表、サイコロは1が出る確率

$$\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$$



# 独立な2x2分割表

期待度数

|    | 属性①                                                                              | 属性②                                                                              |       |
|----|----------------------------------------------------------------------------------|----------------------------------------------------------------------------------|-------|
| A群 | $N_T \times \frac{N_1}{N_T} \times \frac{N_A}{N_T} = \frac{N_1 \times N_A}{N_T}$ | $N_T \times \frac{N_2}{N_T} \times \frac{N_A}{N_T} = \frac{N_2 \times N_A}{N_T}$ | $N_A$ |
| B群 | $N_T \times \frac{N_1}{N_T} \times \frac{N_B}{N_T} = \frac{N_1 \times N_B}{N_T}$ | $N_T \times \frac{N_2}{N_T} \times \frac{N_B}{N_T} = \frac{N_2 \times N_B}{N_T}$ | $N_B$ |
|    | $N_1$                                                                            | $N_2$                                                                            | $N_T$ |



# $\chi^2$ 検定

帰無仮説：行と列は独立である

対立仮説：行と列が独立でない

実測データ

|    | 属性① | 属性② |   |
|----|-----|-----|---|
| A群 | A   | B   | E |
| B群 | C   | D   | F |
|    | G   | H   | N |

期待度数

|    | 属性①                              | 属性②                              |
|----|----------------------------------|----------------------------------|
| A群 | $\hat{A} = \frac{E \times G}{N}$ | $\hat{B} = \frac{E \times H}{N}$ |
| B群 | $\hat{C} = \frac{F \times G}{N}$ | $\hat{D} = \frac{F \times H}{N}$ |

$$\chi^2 = \frac{(A - \hat{A})^2}{\hat{A}} + \frac{(B - \hat{B})^2}{\hat{B}} + \frac{(C - \hat{C})^2}{\hat{C}} + \frac{(D - \hat{D})^2}{\hat{D}}$$

$\chi^2$ は自由度1の $\chi^2$ 分布に従う



# $\chi^2$ 検定

|    | 有効  | 無効 |     |
|----|-----|----|-----|
| 新薬 | 68  | 22 | 90  |
| 偽薬 | 51  | 39 | 90  |
|    | 119 | 61 | 180 |



## 期待度数

|    | 有効                                 | 無効                                |
|----|------------------------------------|-----------------------------------|
| 新薬 | $\frac{119 \times 90}{180} = 59.5$ | $\frac{61 \times 90}{180} = 30.5$ |
| 偽薬 | $\frac{119 \times 90}{180} = 59.5$ | $\frac{61 \times 90}{180} = 30.5$ |

新薬 : 68+22=90  
偽薬 : 51+39=90  
有効 : 68+51=119  
無効 : 22+39=61  
計 180

有効 : 無効=119 : 61



# $\chi^2$ 検定

- 新薬の有効性の判定

期待度数

|    | 有効  | 無効 |     |
|----|-----|----|-----|
| 新薬 | 68  | 22 | 90  |
| 偽薬 | 51  | 39 | 90  |
|    | 119 | 61 | 180 |

|    | 有効   | 無効   |
|----|------|------|
| 新薬 | 59.5 | 30.5 |
| 偽薬 | 59.5 | 30.5 |

$$\chi^2 = \frac{(68-59.5)^2}{59.5} + \frac{(22-30.5)^2}{30.5} + \frac{(51-59.5)^2}{59.5} + \frac{(39-30.5)^2}{30.5} = 7.17$$

|      |      |      |      |      |      |       |      |       |
|------|------|------|------|------|------|-------|------|-------|
| 0.9  | 0.75 | 0.5  | 0.25 | 0.1  | 0.05 | 0.025 | 0.01 | 0.001 |
| 0.02 | 0.1  | 0.45 | 1.32 | 2.71 | 3.84 | 5.02  | 6.64 | 10.83 |



# リスク比（相対危険度）

2x2分割表

|       | 肺がん（－） | 肺がん（＋） |         |
|-------|--------|--------|---------|
| 喫煙（－） | A      | B      | A+B     |
| 喫煙（＋） | C      | D      | C+D     |
|       | A+C    | B+D    | A+B+C+D |

$$\text{非喫煙者の肺がんのリスク} = \frac{B}{A+B}$$

$$\text{喫煙者の肺がんのリスク} = \frac{D}{C+D}$$

$$\text{リスク比} = \frac{\frac{B}{A+B}}{\frac{D}{C+D}}$$



# オッズとオッズ比

- オッズ：ある事象が起こる確率と起こらない確率の比

$$\frac{p}{1-p} : \text{物事が起こる見込みを表す}$$

- オッズ比：ある条件下のオッズと別の条件下のオッズの比

|        | 肺がん (-) | 肺がん (+) |         |
|--------|---------|---------|---------|
| 喫煙 (-) | A       | B       | A+B     |
| 喫煙 (+) | C       | D       | C+D     |
|        | A+C     | B+D     | A+B+C+D |

$$\text{非喫煙者のオッズ} = \frac{\frac{A}{A+B}}{\frac{B}{A+B}} = \frac{A}{B}$$

$$\text{喫煙者のオッズ} = \frac{\frac{C}{C+D}}{\frac{D}{C+D}} = \frac{C}{D}$$

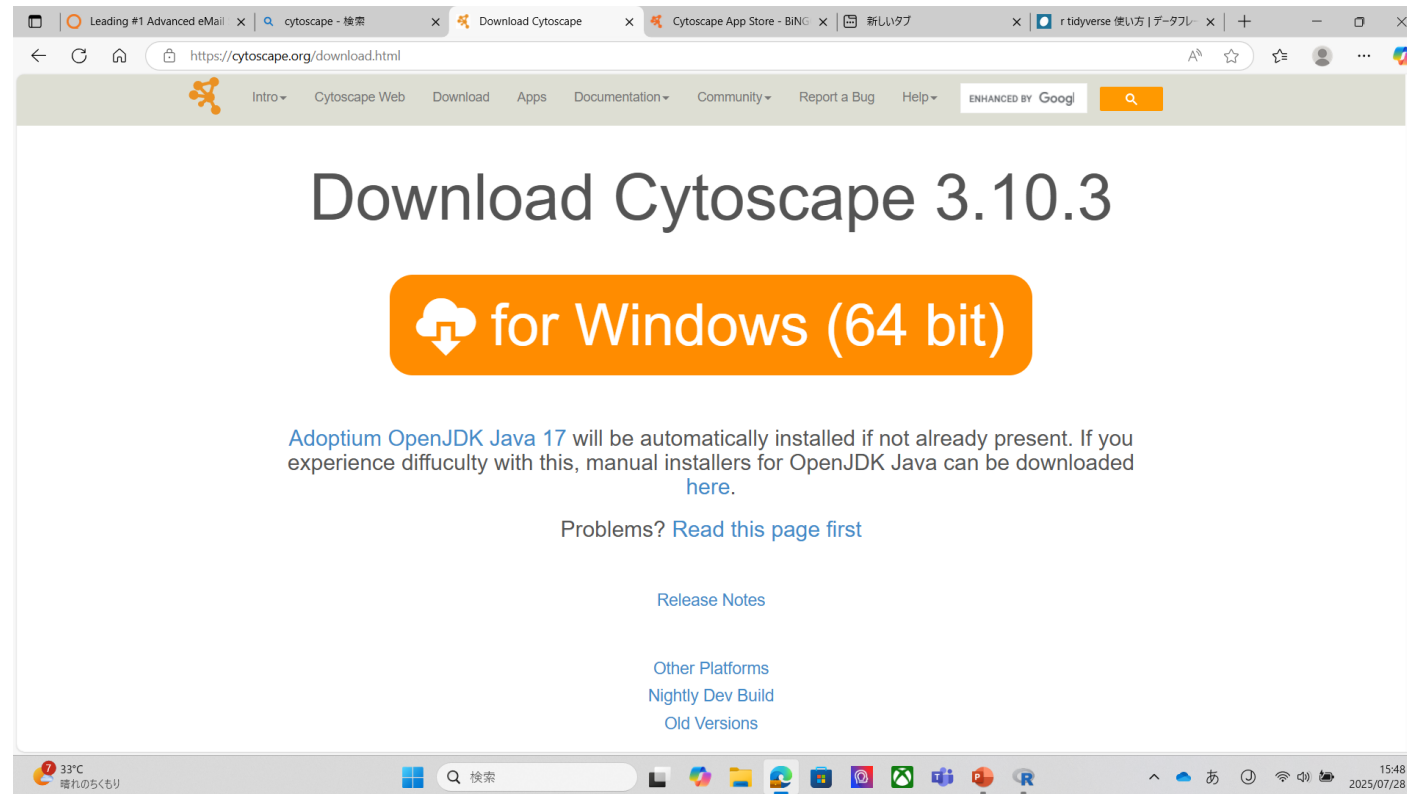
$$\text{オッズ比} = \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{AD}{BC}$$

オッズ比 = 1 : 行と列が独立



# Cytoscape

## Cytoscape : グラフ作成用アプリケーション

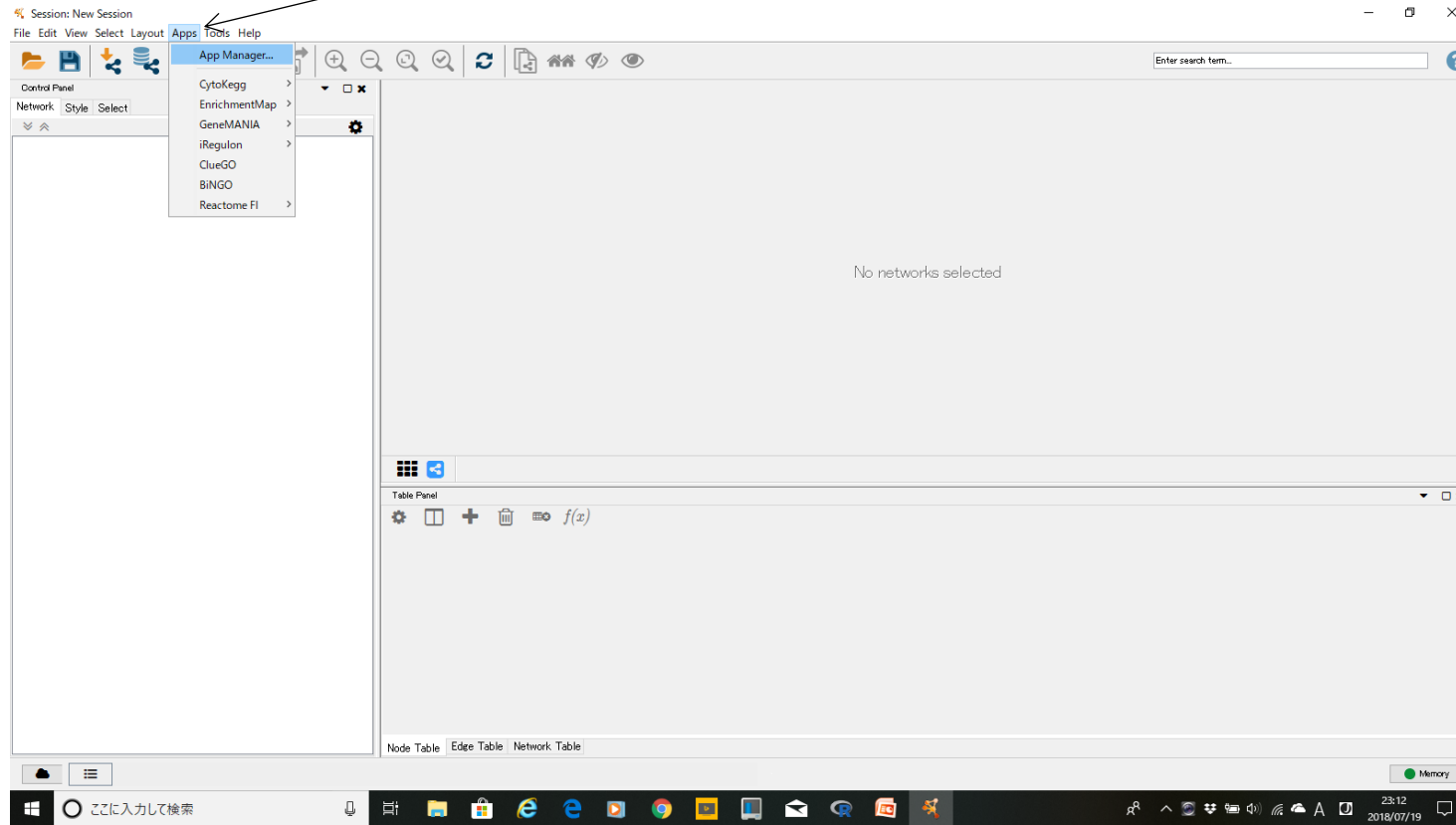


<http://www.cytoscape.org>

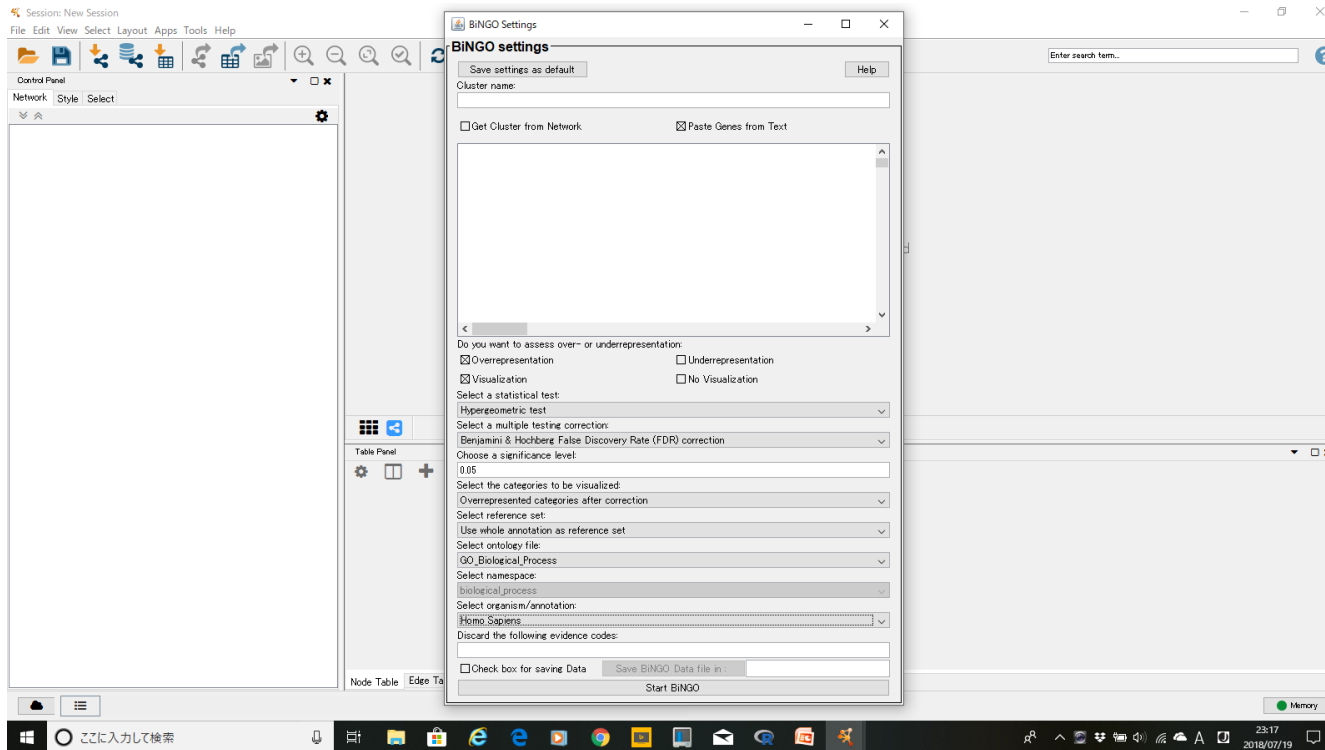


# CytoscapeにAPPをインストール

APPsのApp Managerを選択



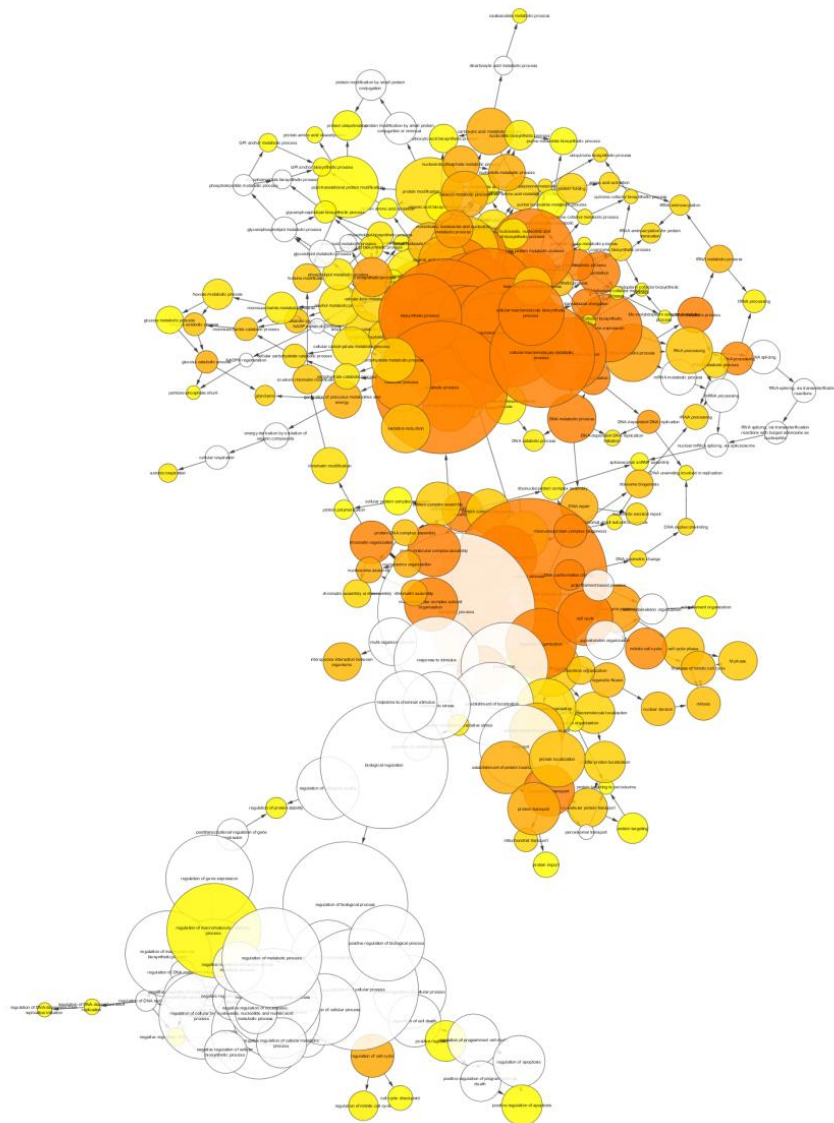
# BiNGOによるGO解析



BiNGO settingsウィンドウの  
Cluster Nameに適切な名前を入力  
Paste Gene from Textにチェック  
入力欄に遺伝子名を入力  
生物種を選択  
Start BiNGOをクリック



# BINGOによるGO解析



| GO term                            | p-value  |
|------------------------------------|----------|
| cellular metabolic process         | 9.52E-58 |
| cellular protein metabolic process | 1.77E-20 |
| nucleic acid metabolic process     | 1.15E-19 |
| DNA metabolic process              | 7.14E-15 |
| protein metabolic process          | 2.20E-14 |
| RNA metabolic process              | 1.11E-07 |
| cellular lipid metabolic process   | 1.85E-05 |
| M phase of mitotic cell cycle      | 1.28E-06 |
| cell cycle phase                   | 3.06E-06 |
| positive regulation of cell death  | 1.40E-03 |
| positive regulation of apoptosis   | 1.49E-03 |



# まとめ

- GOターム：遺伝子の機能や局在を表す統一された用語
- GO解析：遺伝子リスト中に特定のGOタームを持つ遺伝子が高頻度に存在しているかどうか判別する。
- Fisherの正確確率検定：出現頻度の偏りを判定する検定方法

