

遺伝子データベースからの テキストマイニングによる Pathwayにおけるタンパク質の可視化

1815070 武藤 克弥

富山県立大学 電子・情報工学科

November 19, 2021

背景

遺伝子データベースにテキストマイニングを適用し、生命科学分野の新たな知見を得る試みは、依然として需要が高い。その中でもタンパク質間の相互作用を示した Pathway を分析して、タンパク質の機能予測や、化合物の反応経路予測を行うことが盛んである。

目的

- ① Pathway 内のタンパク質に対して共起分析を行い、共起ネットワークを 3D グラフに描画する
- ② 得られたネットワークに対して新たな分析をユーザに促すきっかけを作る

Pathway とは

- 代謝で起こる酵素反応，シグナル（神経）伝達を描いたマップ
→ 遺伝子・タンパク質の相互作用が見られる
- 各分野で得られた実験結果から専門家が手動で作成している
- 四角：タンパク質 (遺伝子)，丸：化合物 (酵素)，矢印+p：リン酸化，矢印+m：メチル化 などを表す

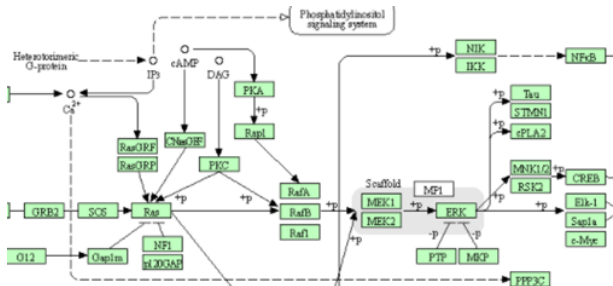


図 1: Pathway の例

反応経路予測

- Pathway データベースを用いてもまだまだ知られていない反応経路は多い + 実験で特定するにも労力と手間がかかる
→既存の Pathway データから未知の経路を予測しようとする流れになっている
- 予測精度を上げるためにどんな特徴量を組み込むか、性能評価する研究が多い

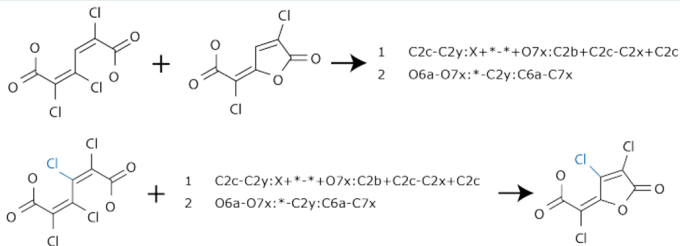


図 2: 抽出された特徴量を用いた反応予測

Pathway 分析の例 (2)

5/10

重要ノードの特定

- Pathway もノード（頂点）とエッジ（辺）を持つグラフ構造
- グラフ理論の媒介中心性（どれくらい他ノードに情報伝達しているか）を用いて、すい臓がんパスウェイの中で重要な上位3つのノードを特定

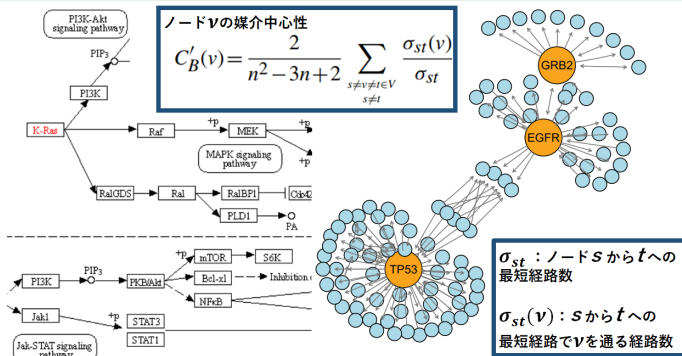


図 3: すい臓がん Pathway と分析の出力結果

Pathway 分析と共起分析の組み合わせ

- ① ヒトの Pathway と遺伝子情報をスクレイピング
- ② 同じ Pathway 内に同時出現するタンパク質 = 互いに共起し合っていると仮定し、共起頻度をカウント
- ③ 共起頻度に応じてエッジの色分け→ 3D グラフ出力

| | hsa:5594 | hsa:5595 | hsa:5291 | hsa:5293 | hsa:5290 |
|-----------|----------|----------|----------|----------|----------|
| hsa:5594 | 0 | 117 | 84 | 84 | 84 |
| hsa:5595 | 117 | 0 | 84 | 84 | 84 |
| hsa:5291 | 84 | 84 | 0 | 106 | 106 |
| hsa:5293 | 84 | 84 | 106 | 0 | 106 |
| hsa:5290 | 84 | 84 | 106 | 106 | 0 |
| hsa:8503 | 83 | 83 | 103 | 103 | 103 |
| hsa:5296 | 83 | 83 | 103 | 103 | 103 |
| hsa:5295 | 83 | 83 | 103 | 103 | 103 |
| hsa:207 | 81 | 81 | 90 | 90 | 90 |
| hsa:10000 | 81 | 81 | 90 | 90 | 90 |
| hsa:208 | 81 | 81 | 90 | 90 | 90 |
| hsa:5604 | 89 | 89 | 71 | 71 | 71 |

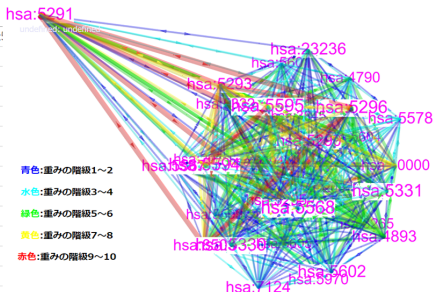


図 4: Pathway 共起分析

3D グラフから分かったこと

- とりあえず赤色のネットワークに着目
- PIK3K と Akt という 2 つのグループネットワークが得られた
- 2 つのネットワークともがんに関連している→なぜこの 2 つが 1 番強く現れたのかを分析することにつながる

今やっていること

- 共起分析の改善
 - 現状は同じ Pathway にあれば共起
→離れすぎていたらあまり共起といえない
 - 同 Pathway かつ 矢印でつながっているとき共起頻度をカウントするように改善
 - Pathway(xml 形式) を解析するライブラリを組み込んだプログラムを制作中
- 生物工学科の先生との対談に向けた勉強

逆合成経路の推定

- ① 欲しい物質（図5左）を作る方法が複数あるとしたとき
- ② 生成するコスト・手数がかからない最適な合成パターン（経路）を推定する

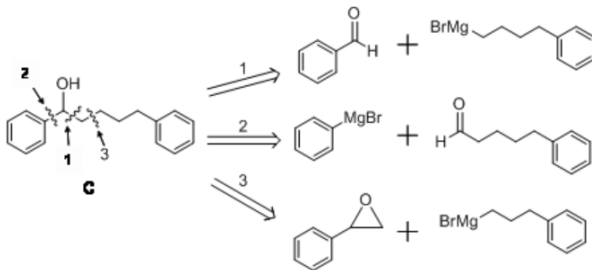


図 5: 3 通りの合成経路がある場合の例

逆合成経路を Pathway で置き換え

- ① 新薬開発の場合、途中の物質すら作り方が未知の場合がある
- ② Pathway 中に、その物質を生成する経路（酵素反応）が既に存在しているなら、見つけ出してそれに置き換える

最適な合成経路が分かったとしても、その途中の物質を実際に作れるかは別

→ Pathway 内の化学反応（分かっている合成方法）に置き換えられれば需要がある

システムの変更

置き換えられるようなシステムに変更する必要がある
（話の内容次第で変わる）

- 欲しい物質をキーワードにして、Pathway 内でその物質を生成しているネットワークを検索
- ネットワークを抽出して置き換える

今後の予定

- 会談までに知識をつける
- 専門家からの提案内容に合わせてシステムを完成させる