

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

December 3, 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

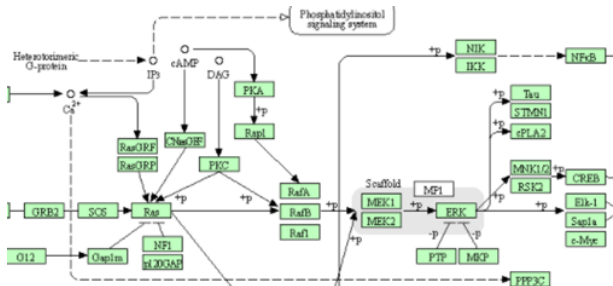


図 1: Pathway の例

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

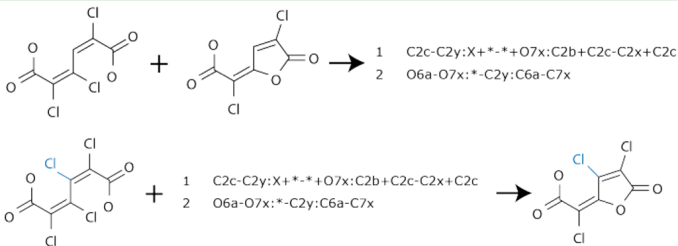


図 2: 抽出された特徴量を用いた反応予測

Pathway 分析の例 (2)

5/15

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

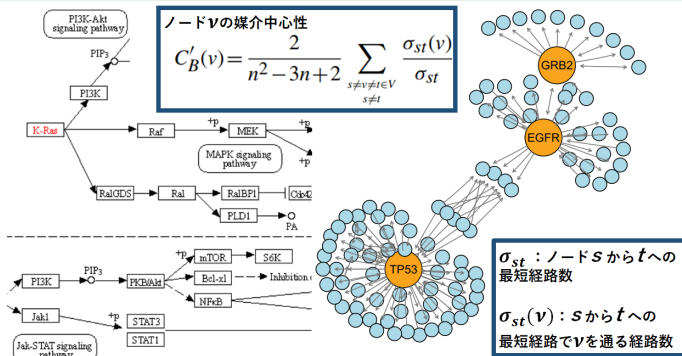


図 3: すい臓がん Pathway と分析の出力結果

背景

研究概要

提案手法

課題

これからの予定

今後の予定

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

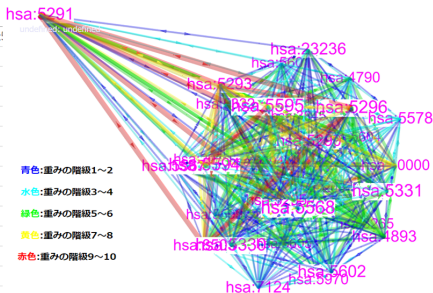


図 4: Pathway 共起分析

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

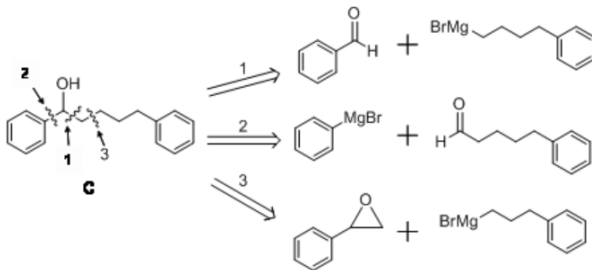


図 5: 3 通りの合成経路がある場合の例

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

これからやっていきたいこと

- ① 3D グラフ共起分析は廃止→逆合成 (レトロシンセシス) 解析の酵素置き換えに
- ② 酵素の反応式データ (約 12,000 パターン収録) を置き換えに使用したい

今やっていること

- 逆合成に関する化学の勉強
- 「生体触媒 (酵素) を用いた逆合成」に関する論文の多読
→ 反応経路を酵素で置き換える手法 & 置き換える際に生じる問題などの勉強

酵素の反応データについて

KEGG REACTION データベースから反応式情報を抽出したもの
→ R 言語を使ってスクレイピングしているソース元があった

- 反応パスウェイとそのパスウェイ内に登録されている反応式を記した対応表を作成
- 反応式の R*****番号から反応式を抽出

map_id	description	
m00010	Glycolysis / Gluconeogenesis	R00014; R00199; R00947; R00959; R02073; R02187; R
m00020	Citrate cycle (TCA cycle)	R00014; R00268; R
m00030	Pentose phosphate pathway	R00305; R01049; R01544; R01621; R02740; R02749; R
	Pentose and	R00264; R00286; R01529; R01540; R01903; R01904; R

ENTRY	NAME	DEFINITION	EQUATION
R00001	polyphosphate polyphosphohydrolase	Polyphosphate + n H ₂ O <=> (n+1) Oligophosphate	C00404 + n C00001 <=> (n+1) C02174
R00002	Reduced ferredoxin:dinitrogen oxidoreductase (ATP-hydrolysing)	16 ATP + 16 H ₂ O + 8 Reduced ferredoxin <=> 8 e ⁻ + 16 Orthophosphate + 16 ADP + 8 Oxidized ferredoxin	16 C00002 + 16 C00001 + 8 C00138 <=> 8 C05359 + 16 C00009 + 16 C00008 + 8 C00139
R00004	diphosphate phosphohydrolase; pyrophosphate phosphohydrolase	Diphosphate + H ₂ O <=> 2 Orthophosphate	C00013 + C00001 <=> 2 C00009
			C01010 +

図 6: 対応表と反応式の表

背景

研究概要

提案手法

課題

これからの予定

今後の予定

酵素と逆合成 (レトロシンセシス)

逆合成では、目的の化学物質を生成するために、自然界のまたは人工的に作られた化学反応を用いて最適な合成経路を見つけていく。

人工的に開発された反応と、自然由来の酵素反応を組み合わせで合成されることもあるが、酵素に関して未知の部分が多く、天然の酵素反応が合成に用いられることは人工のものに比べてかなり少ない。

酵素を触媒として用いることのメリット

- ① 合成ステップ数の減少によるコスト削減
- ② 溶媒の使用頻度減少による良環境性

「生体触媒 (酵素) を用いた逆合成」に関する論文で学んだこと

- この酵素が使える→実際に置き換えられるか検証の流れ (何の溶媒使うか&触媒酵素の濃度など実験結果で設定)
- 合成後は混合物→どうやって目的の化合物だけに分離するかも検討

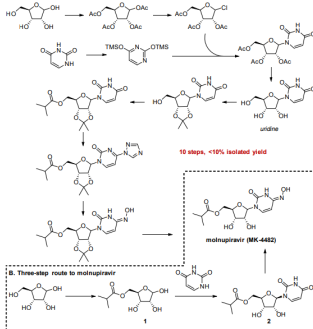


図 7: 酵素反応を用いたステップ数短縮

今のところ考えている方向性

- 約 12000 の反応データから、置き換え候補の見当がついても、実際の実験 (スクリーニング) でそれがふさわしいか分かるのは数年かかることも
→ 徹底的なシミュレーションで補う？
- 逆合成経路の置き換え候補となりそうな酵素経路を列挙
- 様々な観点 (収率 (用いた材料に対して得られる生成物の割合), 手数, 環境面, 低コスト, 大量生産性 など) からスコアリング, 最適な酵素反応を挙げる.

今後の予定

- 論文の深掘り
- 持っているデータとの組み合わせ検討