異なる実験条件で得られたプロテオームデータの統合ネットワーク解析

Integrated network analysis of proteome data obtained under different experimental conditions

西崎愛花[1,2],河野信[1,3]

1.北里大学未来工学部 ,2.北里大学理学部 ,3.情報・システム研究機構 ライフサイエンス統合データベースセンター

1 研究背景

jPOST をはじめとする ProteomeXchange リポジトリから様々な実験で得られたプロテオームデータがオープンデータとして公開されている。プロテオームの解析を行う際には単一プロジェクトのデータだけでなく、異なる施設や異なるサンプルから得られたデータを統合して比較できると有用である。しかしながら、これらの異なる実験条件で得られたデータを単純に比較することは難しい。

メタボロームの分野では、異なる実験条件のデータをネットワークを介して解析する試みがなされており[Matsuta, R., *et al. BMC Bioinformatics* **23**, 508 (2022)]、この手法をプロテオームデータにも適用した。

2 研究目的

今回の研究では異なる実験条件で得られたプロテオームデータを統合して解析するためにネットワークの作成を行う。

構築されたネットワークからプロテオーム実験群間での関係性を発見する。

3 研究方法

データセット

quantMS[Dai, C., *et al. Nature Methods* **21**, 1603–1607 (2024)]で再解析されているラベル化定量のデータで、サンプル間の比率が計算されている 36 プロジェクト、2370 実験条件を用いた。

解析手法

1) 2実験条件の比較（＝実験群）においてタンパク質変化量が 2 倍以上かつ多重検定補正後の p 値が 0.01 以下のタンパク質を取得
2) 2つの実験群からタンパク質変化量のクロス集計表を作成（表1、2）
3) 作成したクロス集計表から、タンパク質の増減挙動についてのオッズ比ならびにカイ二乗検定で p 値を計算
4) カイ二乗検定で有意（有意水準 0.05 を Bonferroni 法で多重検定補正した p 値 $< 1.78 \times 10^{-8}$）であった実験群の組み合わせから、Cytoscape を用いてネットワークを構築（図1）
5) カイ二乗検定で有意であった実験群の組み合わせから、違うプロジェクト由来の実験群の組み合わせを抽出してネットワークを構築（図2）

カイ二乗検定

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

O = 頻度の観測値
E = 帰無仮説の下における頻度の期待値（理論値）

オッズ比計算式

$$odds\ ratio = \frac{m_{1,1} * m_{2,2}}{m_{1,2} * m_{2,1}}$$

ネットワーク上で関連があるデータ、関連がないデータの解析

作成したネットワーク上で関連が見られた実験群間（図3）と、関連が見られなかった実験群間（図4）についてそれぞれタンパク質の発現挙動と相関係数を調べた。

エンリッチメント解析

オッズ比が最大であった実験群の組み合わせに対して、両方の実験群において共に発現量が増加していたタンパク質について、DAVID[https://david.ncifcrf.gov/tools.jsp]を用いてエンリッチメント解析を行った（表3）。

4 結果

①全データから構築したネットワーク

今回作成したネットワークは **2370** ノード、
有意水準を満たしたエッジの本数は **7.2万** で、
7個 のサブネットワークに分類された。

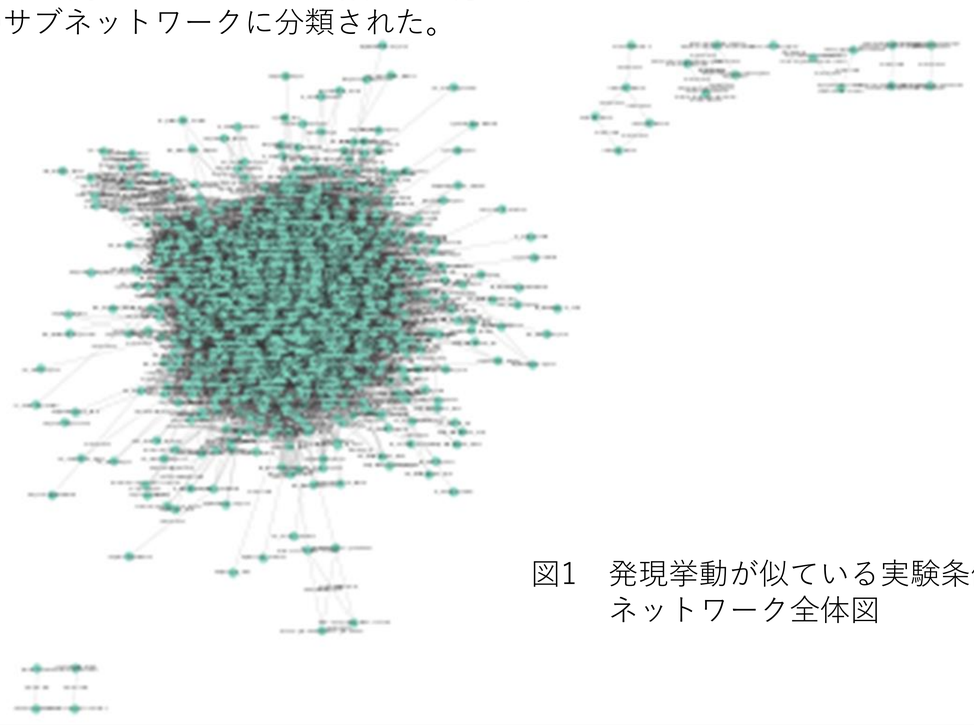

図1　発現挙動が似ている実験条件のネットワーク全体図

② 異なるプロジェクトのみで構築したネットワーク

異なるプロジェクトの実験群の組み合わせから作成したネットワークは **113** ノード、
有意水準を満たしたエッジの本数は **145** で、
5個 のサブネットワークに分類された。

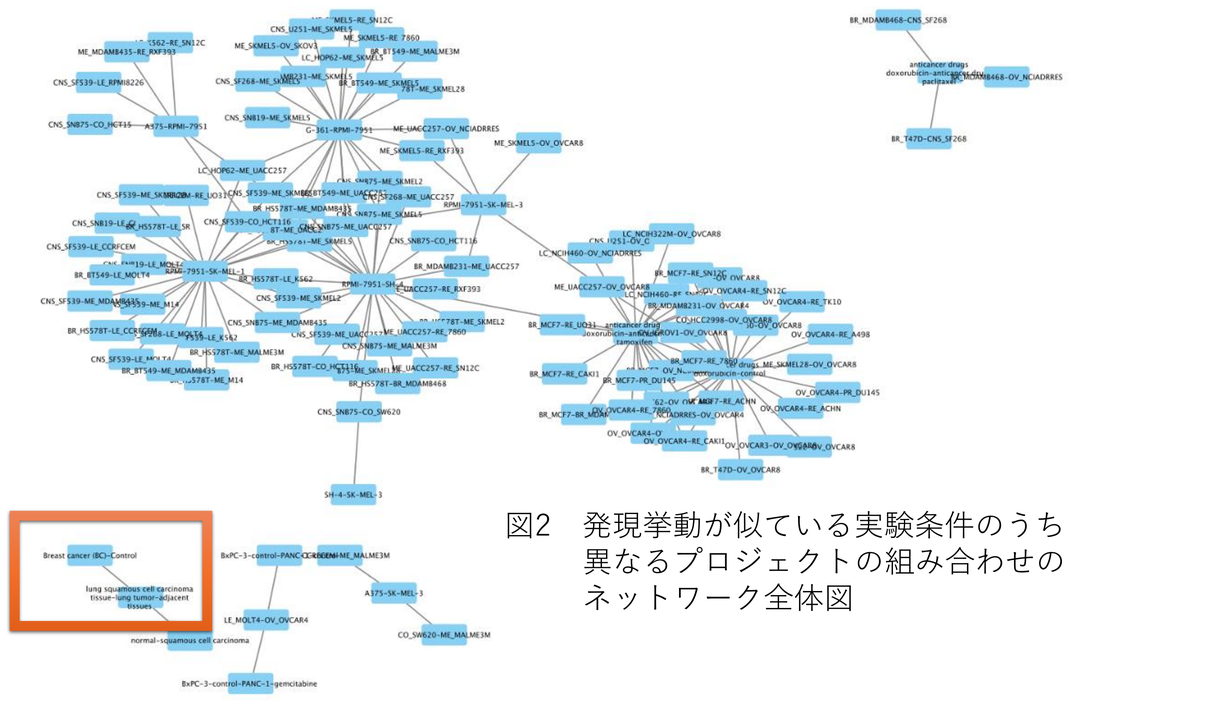

図2　発現挙動が似ている実験条件のうち異なるプロジェクトの組み合わせのネットワーク全体図

③ ネットワーク上で関連があるデータ（オッズ比:1701.0）

表1　ネットワーク上で関連があるデータのクロス集計表

lung squamous cell carcinoma tissue-lung tumor-adjacent tissues	Breast cancer (BC)-Control: up	Breast cancer (BC)-Control: down
up	121	0
down	0	3

ネットワーク上で関連が見られた実験群間では、タンパク質の発現量比に相関関係が見られた

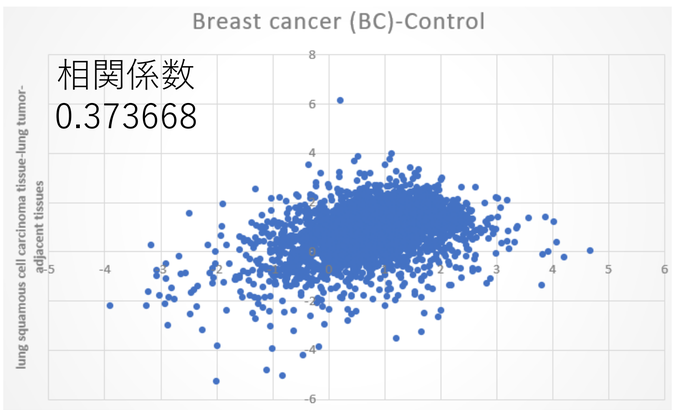

図3　ネットワーク上で関連があった実験群間のタンパク質発現量比の散布図

④ ネットワーク上で関連が見られなかったデータ（オッズ比:1.3548）

表2　ネットワーク上で関連がないデータのクロス集計表

lung squamous cell carcinoma tissue-lung tumor-adjacent tissues	A375-SK-MEL-3: up	A375-SK-MEL-3: down
up	90	93
down	5	7

ネットワーク上で関連が見られたものと比較して、相関関係が見られなかった

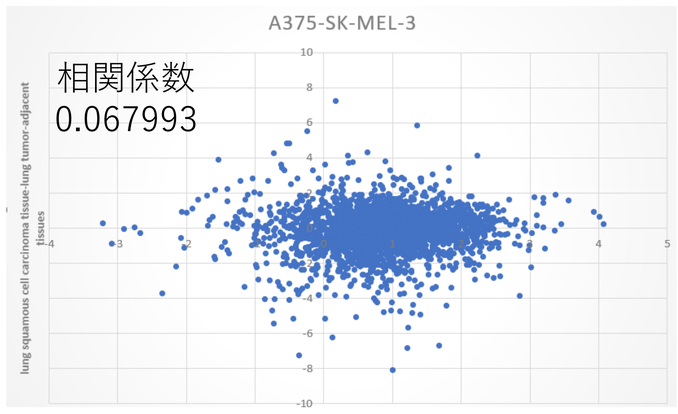

図4　ネットワーク上で関連がなかった実験群間のタンパク質発現量比の散布図

⑤エンリッチメント解析

表3　DAVIDによる解析結果（biological process）

268 chart records　Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	GOTERM_BP_FAT	cytoplasmic translation	RT		26	21.1	3.4E-30	6.6E-27
	GOTERM_BP_FAT	gene expression	RT		73	59.3	1.3E-27	1.3E-24
	GOTERM_BP_FAT	macromolecule biosynthetic process	RT		77	62.6	3.0E-27	2.0E-24
	GOTERM_BP_FAT	ribonucleoprotein complex biogenesis	RT		35	28.5	4.6E-26	2.2E-23
	GOTERM_BP_FAT	translation	RT		32	26.0	2.8E-25	1.1E-22
	GOTERM_BP_FAT	mRNA splicing, via spliceosome	RT		25	20.3	1.3E-20	3.8E-18
	GOTERM_BP_FAT	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	RT		25	20.3	1.3E-20	3.8E-18
	GOTERM_BP_FAT	RNA splicing, via transesterification reactions	RT		25	20.3	1.9E-20	4.6E-18
	GOTERM_BP_FAT	protein-RNA complex assembly	RT		23	18.7	3.5E-20	7.0E-18
	GOTERM_BP_FAT	mRNA processing	RT		30	24.4	3.6E-20	7.0E-18
	GOTERM_BP_FAT	RNA splicing	RT		28	22.8	5.0E-20	9.0E-18
	GOTERM_BP_FAT	mRNA metabolic process	RT		33	26.8	6.2E-20	1.0E-17
	GOTERM_BP_FAT	protein-RNA complex organization	RT		23	18.7	8.8E-20	1.3E-17
	GOTERM_BP_FAT	nucleic acid metabolic process	RT		59	48.0	4.1E-19	5.7E-17
	GOTERM_BP_FAT	RNA processing	RT		43	35.0	1.8E-18	2.3E-16

オッズ比が最大であった実験群の組み合わせについて、共に発現量が増加していたタンパク質は、遺伝子発現に関わるものであることが分かった。

5 まとめ・展望

- 公開プロテオーム定量データから、タンパク質の発現挙動が似ている実験条件のネットワークを作成した。
- ネットワークから関連がある実験群を抽出できたことを確認した。
- 今後の展望として、今回作成されたネットワークから既知の疾患等の関係性について再発見することが可能であるかについて検証を行う。
- 加えてネットワーク上で関連が見られた、未知の関係について推定を行う。
- 今回はラベル化定量のデータセットのみを使用したが、quantMS ではラベルフリー定量のデータも公開されているので、これらを加えてネットワークを構築し、分析を行う予定である。

本研究は、科学技術振興機構（JST）統合化推進プログラム　「創発的再解析のためのメタボローム統合データベース」（JPMJND2403）の支援のもと実施されました