

## 科学研究費助成事業 研究成果報告書

平成 30 年 5 月 25 日現在

機関番号：11301

研究種目：若手研究(B)

研究期間：2015～2017

課題番号：15K18464

研究課題名(和文)組織横断型遺伝子共発現法を用いた細胞間コミュニケーション解析

研究課題名(英文) Analysis of cell-cell communication using comparison of gene co-expression for different tissues

研究代表者

大林 武 (Obayashi, Takeshi)

東北大学・情報科学研究科・准教授

研究者番号：50397048

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：細胞機能は多数の遺伝子の機能が組み合わさって構成されているため、その組み合わせを推定するための遺伝子ネットワーク解析が、遺伝子と細胞機能を繋げる解析として広く用いられている。特に遺伝子発現プロファイルの類似性に基づく遺伝子共発現ネットワーク解析は、タンパク質相互作用ネットワーク解析と並んで遺伝子ネットワーク解析の重要なツールとなっている。本研究では、モデル植物シロイヌナズナを対象として、遺伝子共発現ネットワークを細胞間で比較統合解析するための技術開発を行った。また、その成果を植物の遺伝子共発現データベースATTED-II [ <http://atted.jp> ]にて公開した。

研究成果の概要(英文)：Cell functions are composed of the individual functions of many genes. Gene network analysis is widely used to estimate the functional combination of genes. Similarly with protein-protein interaction network analysis, gene co-expression network analysis, which is a similarity of gene expression profiles, is widely used as network analysis. In this study, I have developed a methodology to compare and integrate multiple gene co-expression networks among multiple tissues of a model plant, Arabidopsis. The results based on the new methodologies in this study were available in a plant co-expression database, ATTED-II [ <http://atted.jp/> ].

研究分野：システムゲノム科学

キーワード：遺伝子共発現 ネットワーク生物学 シロイヌナズナ データベース 共発現データベース ゲノム進化

## 1. 研究開始当初の背景

多細胞生物は多数の細胞が逐次情報交換を行い、協調して働くことによって成立している。例えば、野外における植物は常に複合的な環境変化にさらされており、強光ストレスは葉が、乾燥や塩ストレスは根がその異常に対する初期応答を行い、その情報が他の組織へと伝わることで全身応答が達成される。このような組織間の連携は個体の生存に必須であり、植物ホルモンや代謝産物、もしくは miRNA のような制御因子が維管束系を通じて輸送され、各組織における遺伝子発現を調節し全身応答となることが、近年次々と報告されている。しかしその一方で、細胞間コミュニケーションの全体像は不明なままである。

近年、個々の細胞における研究では、遺伝子ネットワークを利用した解析が大きな成功を収めている。我々が開発を進めている遺伝子共発現法は、遺伝子発現パターンの類似性に注目した方法であり、マイクロアレイ技術の普及による遺伝子発現データの蓄積に伴って、機能未知遺伝子の機能同定に大きな威力を発揮している。しかし、このような遺伝子共発現解析は仮想的な単一細胞における遺伝子ネットワークを見ているに他ならない。異なる組織には異なる遺伝子ネットワークが存在し、それらの間では物質のやりとりを通じて情報が交換され、個々の細胞の遺伝子ネットワークは細胞を超えて繋がっている。すなわち、組織ネットワークを既に実績のある遺伝子ネットワークと接続することで、多細胞システムを一つの巨大な遺伝子ネットワークをして捉えることが可能になる。種々の細胞間コミュニケーション研究を共通の遺伝子ネットワークにマップし統合していくことは、多細胞生物の全細胞モデリングの基盤として有用であろう。

申請者はこれまでに植物および動物の共発現データベースの開発を進めてきた。特に植物の共発現データベース **ATTED-II** (<http://atted.jp>) は盛んに利用され、国内外で高い評価を得ている。またこれまでに条件別共発現を比較する手法を開発し、比較共発現解析により共発現情報の生物学的解釈を可能にした。また細胞内小器官別の遺伝子ネットワークを構築し、光呼吸などの複数のオルガネラにまたがる機能を見出す手法も確立している。これらの手法を細胞間にスケールアップすることで、細胞間の情報交換を含めた、組織横断型の遺伝子共発現ネットワークの作成が可能である。

## 2. 研究の目的

組織横断型共発現ネットワークの構築には、複数の組織サンプルを同時に測定することが必要であるが、近年の網羅的遺伝子発現データの充実により、そのような複数組織の同時測定データが利用できる可能性がでてきた。同時測定データの代表例としては、国

際プロジェクト **AtGenExpress** の非生物学的ストレスデータがあり、様々なストレス処理に対する葉と根のサンプルが時系列で測定されている。

一方でそのような複数の組織を同時測定しているデータの多くは、データ点が少ないといった問題がある。例えば、サイトカニン処理をおこなったイネの葉と根の実験 (**NCBI GEO DataSet ID: 2632**) は、各組織 12 点のデータからなっている。このような小サンプルから共発現ネットワークを構築することはできない。そこで本研究では、植物組織別の遺伝子共発現ネットワークの構築と比較統合解析のため、小サンプルデータに基づく高精度遺伝子共発現情報の導出、ならびに遺伝子共発現の比較法を開発し、組織横断共発現ネットワークの大規模構築の基盤を確立することを目的とした。

## 3. 研究の方法

### (1) 技術開発

遺伝子共発現ネットワークの質は用いるマイクロアレイデータの数に依存する。組織横断型遺伝子発現データは極めて少なく、組織特異的遺伝子発現データも、遺伝子発現データ全体と比較すると一部にしかすぎない。そのため、遺伝子発現データを増やすための収集方法、少数の遺伝子発現データから高精度な共発現情報を導出する方法、少数データに基づく複数の共発現を比較統合する方法の開発を行う。

### (2) 遺伝子発現データ

マイクロアレイデータは **EBI ArrayExpress** から生データをダウンロードし、一括で補正を行うことで補正条件によるバイアスが含まれないようにする。データ数を増やす一環として、マイクロアレイデータのみならず、**RNAseq** のデータも用いる。**RNAseq** による遺伝子発現量データは次世代シーケンサデータのリポジトリである **NCBI SRA** から、サンプルのアノテーションに基づき収集する。**RNAseq** データには **RNA** 分子種の濃縮が行われているエントリも含まれるため、メタデータやデータ自体の特性に注目して、共発現解析に不適切なエントリを取り除く。

### (3) 遺伝子共発現データの評価

遺伝子共発現情報においては、タンパク質相互作用データのような正解セットを構築することができないため、良い遺伝子共発現情報が持ちうる性質を用いて性能評価を行う。「データ再現性」は、異なる発現量データのプラットフォーム (各種マイクロアレイ、各種 **RNAseq** のプラットフォーム) に基づく共発現値が類似しているかを定量化したものである。「ゲノム配列との無矛盾性」は、共発現に関与するゲノム配列の特徴が、所与の遺伝子共発現情報と矛盾していないかを

定量化したものである。「遺伝子機能の同調性」は、これまでに多数の実績があるように遺伝子共発現が遺伝子機能の推定に用いられていることから、遺伝子オントロジーやKEGG パスウェイのアノテーションが共発現遺伝子ペアで共有されて度合いを定量化したものにになる。

#### 4. 研究成果

##### (1) サンプル補正法の高度化

「データ再現性」「ゲノム配列との無矛盾性」「遺伝子機能の同調性」の観点から遺伝子共発現の質を評価すると、遺伝子共発現情報は作成に用いるサンプル数に強く依存することがわかる。そのため一般に多数の実験を組み合わせ、サンプル数を増やすメタ解析が用いられるが、各実験は異なる環境で測定されているため、仮に同一の実験条件であったとしてもシステムエラーが含まれてしまう。このシステムエラーを取り除くバッチ補正法を検討した結果、実験特異的なノイズとして、ノイズの量だけでなくノイズの分散も考慮して補正を行うと、遺伝子発現データとサンプル条件をより正しく復元でき、異なる実験を組み合わせる影響を小さくできることを見出した。

##### (2) 共発現導出法の高度化

遺伝子共発現の指標として広く用いられるピアソンの相関係数は、測定ノイズ存在下においてサンプル数の影響を大きく受けるため、サンプル数が異なる共発現を直接比較することに向かない。この問題に対してピアソンの相関係数を相互順位に変換した MR 値が有効であることを、本研究を始めるまでに見出していたが、「データ再現性」「ゲノム配列との無矛盾性」「遺伝子機能の同調性」の観点から、本研究で対象とする小規模サンプルに対しても MR 値が有効であることを確認した。さらに、MR 値の特性（分散、歪度、尖度）が MR 値の大きさに依存してしまう問題があることがわかった。小サンプル数に基づく共発現の比較では統計検定が必須となるが、多くのパラメトリックな統計検定は等分散を仮定している。この問題について検討を進めた結果、MR 値を割合に変換したのちにロジット変換を行うことでこの問題を回避できることがわかった。

ブートストラップ法を用いて共発現指標の誤差分布を詳細に解析したところ、ロジット変換した MR 値の誤差分布はプラットフォーム特異的な標準偏差をもつ正規分布として近似できることを見出した。これらの性質から、複数のプラットフォームの MR 値を  $MR_i$ 、その重みを  $w_i$  ( $w_1, w_2, \dots, w_k; \sum w_i = 1$ ) とすると統合 MR の点推定値は次のように計算できることになり、異なるプラットフォームのデータを合わせて共発現精度を向上させることが可能になった。

$$MR_{average} = \frac{N \prod MR_i^{w_i}}{\prod (N - MR_i)^{w_i} + \prod MR_i^{w_i}}$$

##### (3) サンプル選択の高度化

各実験にはそれぞれ特有の実験環境や測定環境があり、マニュアルで収集した遺伝子発現データのメタデータだけでは環境を揃えるのに不十分であることがわかった。そこで、任意のサンプル集合間でサンプル条件がどの程度類似しているかを、サンプルカバー率とサンプル特異性の観点から定量化する手法を構築し、個体の生育条件がある程度揃っているサンプル集合を選択できるようにした。さらに、遺伝子発現データならびに共発現強度の差異が、組織の違いに由来するものか、個体の生育条件の違いに由来するものかを区別するために、近縁生物種の遺伝子発現データを比較する手法を開発した。

組織別の遺伝子発現データの選択においては、目的外の組織サンプルを除去しつつ、できるだけサンプル数を確保するというトレードオフを解決する必要があるため、サンプルの類似性に基づいた手動選択を補佐する仕組みを構築した。

さらにバギング法を適用することで、特に小サンプルにおいては、一つ一つのサンプルの寄与が大きく、安定した共発現を得るのが難しい。それに対しては、実験単位でブートストラップを行い、得られた共発現度を平均化する手法により、共発現の質を改善できることを見出した。

##### (3) 大規模データでの検証

上記手法が小規模データのみにも有効なのか、大規模データにも有効なのかを検証するため、植物の共発現データベース ATTED-II における複数プラットフォームが利用できる生物種を対象に手法の検証を行った。バッチ補正法の導入により、共発現精度の向上が確認された。ロジット変換 MR に基づく比較共発現を用いることで、共発現値の広い範囲において系統特異的な共発現を多数検出することに成功した。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 5 件)

- (1) Obayashi T, Aoki Y, Tadaka S, Kagaya Y, Kinoshita K. ATTED-II in 2018: A Plant Coexpression Database based on Investigation of Statistical Property of the Mutual Rank Index. *Plant Cell Physiology* 59, e3 (2018) 査読あり doi: 10.1093/pcp/pcx191
- (2) Narise T, Sakurai N, Obayashi T, Ohta

H, Shibata D. Co-expressed Pathways DataBase for Tomato: a database to predict pathways relevant to a query gene. BMC Genomics, 18, 437 (2017) 査読あり  
doi: 10.1186/s12864-017-3786-3

- (3) Aoki Y, Okamura Y, Ohta H, Kinoshita K, Obayashi T. ALCodb: Gene Coexpression Database for Microalgae. Plant Cell Physiology, 57, e3 (2016) 査読あり  
doi: 10.1093/pcp/pcv190
- (4) Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T. ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. Plant Cell Physiology, 57, e5 (2016) 査読あり  
doi: 10.1093/pcp/pcv165
- (5) Okamura Y, Obayashi T, Kinoshita K. Comparison of gene coexpression profiles and construction of conserved gene networks to find functional modules. PLoS One, 10, e0132039 (2015) 査読あり  
doi: 10.1371/journal.pone.0132039

[学会発表] (計6件)

- (1) Obayashi T, Aoki Y, Kinoshita K. A revised coexpression calculation procedure in ATTED-II version 9 with batch normalization and bagging methods. The 59th Annual Meeting of the Japanese Society of Plant Physiologists. March, 2018.
- (2) Obayashi T, Aoki Y, Okamura Y, Tadaka S, Kinoshita K. Sample Alignment of Different Species toward Species-Specific Coexpression Analyses. The 58th Annual Meeting of the Japanese Society of Plant Physiologists. March, 2017.
- (3) Obayashi T, Aoki Y, Okamura Y, Tadaka S, Kinoshita K. Meta-analysis of gene coexpression in ATTED-II toward species-specific coexpression analyses. The 57th Annual Meeting of the Japanese Society of Plant Physiologists. March, 2016.
- (4) Aoki Y, Obayashi T, Kinoshita K. Identification of Genomic Features Associated with Gene Coexpression in Arabidopsis thaliana. The 57th Annual Meeting of the Japanese Society of Plant Physiologists. March, 2016.
- (5) Aoki Y, Obayashi T, Kinoshita K. Relationship Analysis between 3D Genome Architecture and Gene Coexpression. Informatics in Biology, Medicine and Pharmacology 2015.

October, 2015.

- (6) Obayashi T, Okano Y, Narise T, Aoki Y, Okamura Y, Tadaka S, Kinoshita K. Prediction of functional divergence among paralogous genes based on gene coexpression. The 58th Annual Meeting of the Japanese Society of Plant Physiologists. March, 2015.

[その他]

ホームページ等  
<http://atted.jp>

## 6. 研究組織

### (1) 研究代表者

大林 武 (Obayashi, Takeshi)  
東北大学・大学院情報科学研究科・准教授  
研究者番号: 50397048

### (2) 研究分担者

無し

### (3) 連携研究者

無し

### (4) 研究協力者

木下 賢吾 (Kinoshita, Kengo)  
東北大学・大学院情報科学研究科・教授  
研究者番号: 60332293