

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成25年5月31日現在

機関番号：82401

研究種目：若手研究(B)

研究期間：2011～2012

課題番号：23700355

研究課題名（和文） ポストゲノム機能解析を促進する新規データ統合手法の開発～新規共発現関係の探索～

研究課題名（英文） Development of a novel integrated data mining that facilitates functional analysis in the post-genomic era -- Exploring novel gene coexpression --

研究代表者

福島 敦史 (FUKUSHIMA ATSUSHI)

独立行政法人理化学研究所・メタボローム情報チーム・研究員

研究者番号：80415281

研究成果の概要（和文）：

遺伝子共発現解析による遺伝子機能予測を向上させるために、新規データ統合手法の開発を目指した。また二つの条件間で各々測定された遺伝子群間の共発現パターンが二条件間で有意に異なる遺伝子群を探す手法（共発現差異）を実装した DiffCorr パッケージを開発し、さらなる遺伝子機能予測の向上を推進した。この手法はバイオマーカー候補の探索に役立ち、因果ネットワーク推定の第一歩となる。

研究成果の概要（英文）：

We have investigated an integrated data mining that facilitates functional analysis to improve limits in gene coexpression approach for function prediction. We have developed the DiffCorr package, a simple method for identifying pattern changes between 2 experimental conditions in correlation networks, which builds on a commonly used association measure. Our method is suitable for the first step towards inferring causal relationships and detecting biomarker candidates.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
交付決定額	2,200,000	660,000	2,860,000

研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：バイオインフォマティクス、トランスクリプトミクス、遺伝子共発現、データ統合、データ正規化、differential correlation、マイクロアレイ

1. 研究開始当初の背景

実験条件にわたった発現パターンの類似尺度（例えばピアソン相関係数）を算出し、機能既知遺伝子の発現パターンとの類似性から機能予測を行う手法は遺伝子共発現解析

と呼ばれる。通常この共発現計算には、多様な実験デザインの下、各研究者が自らの興味に基づき測定して得た公共利用可能なマイクロアレイデータセットを使用する。したがって類似尺度計算以前に「データ統合」が必然的に行われている。ところが異なるサンプル

ル数、異なる繰り返し数、環境擾乱や遺伝子改変の有無などの違いによって、トランスクリプトーム変化への寄与はデータセットごとに異なっている。以上の寄与をいかに考慮するかは、最終的な共発現関係の結果に影響するため重要と考えられた。共発現データの信頼性を高めるために類似尺度計算前のデータ統合手法に改善を加える必要があった。

2. 研究の目的

遺伝子共発現解析による遺伝子機能予測を向上させるために、実験データセットごとの相関係数分布を考慮した重み付けを用いる新規データ統合アルゴリズムの開発を目指した。

3. 研究の方法

共発現データを計算する際、異なった実験デザインおよび種々の研究室から得られたデータの統合次第で共発現の結果が変化する。このため、異種のデータセットを統合する際にデータセットの各々が最終的な共発現結果へいかに寄与するかを効果的に測り、データの重み付けを行う合理的な戦略を採用する。この手法により、既存の共発現ツールで見落とされていた可能性のある共発現関係を明らかにし、遺伝子機能の新規発見を目指す。

4. 研究成果

本課題により以下の成果が得られた。当初の想定とは異なる新たな研究の発展と課題が生じたので、以下にまとめて示す。

- (1) 遺伝子共発現解析による遺伝子機能予測を向上させることを目的として、実験データセットごとの相関係数分布を考慮した重み付けを用いる新規データ統合アルゴリズムの開発を進めた。収集したアレイデータの重複を取り除き、メタデータと共に整理を行った。各データの品質評価を行うためにロバスト線形フィッティング法を適用し、共発現推定に使用できない異常データを検出し、以降の解析から取り除いた。収集したシロイヌナズナのマイクロアレイ (Affymetrix GeneChip ATH1) のデータセットの各々に対して、共発現データを計算する際に影響を及ぼすと考えられる実験デザイン、統計的検定による発現差異遺伝子群のデータ分布および相関係数値の分布から、最終的な共発現結果への寄与度を効果的

に測る指標を選定した。これら指標から各データセットの特異性・寄与度と見なすスコアに基づいてデータ統合の際の重み付けを行った。しかし、データ統合の有効性を明白に示す手法および他手法との直接比較が難しいことが浮き彫りになった。

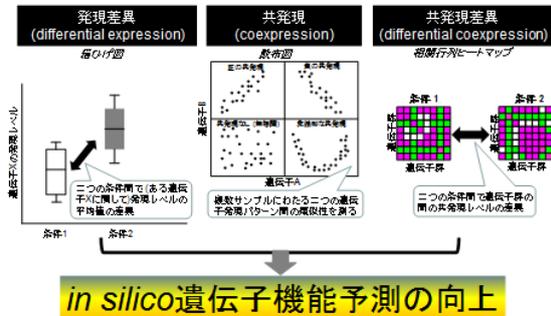


図1. 発現差異と共発現差異の違い
変異体と野生型のような二つの条件間で、ある遺伝子Xの発現レベルに差があるかどうかは検定などで調べられる(左図)。複数サンプルにわたった二つの遺伝子発現パターンは類似尺度を用いて特徴づけられる(共発現、中図)。二つの条件間で各々測定された遺伝子群間の共発現パターンは二条件間で異なる場合がある(右図)。これら共発現差異遺伝子群には鍵酵素遺伝子や転写調節因子などが含まれる可能性が示唆されている。

- (2) 並行して、シロイヌナズナのみならずイネ・トマトといった作物に関するアレイデータについても網羅的なデータおよびメタデータ収集を行った。二つの実験条件間における共発現関係を互いに比較し、統計的に有意に異なる共発現関係を考えることができる(共発現差異)。ふつうマイクロアレイ解析では、二条件間で遺伝子発現レベルの平均値に有意な差異を持つ「発現差異遺伝子」を探す(図1左)。それに対し、共発現差異は、二つの条件間で各々測定された遺伝子群間の共発現パターン(図1中央)が二条件間で有意に異なる遺伝子群を探す(図1右)。このような解析手法は、マイクロアレイ等のトランスクリプトームデータが一定数以上蓄積されていれば、生物種を問わず適用できる。本研究では実用作物であるトマトのマイクロアレイデータを用いて、これら共発現差異遺伝子の意義を調べた。その結果、リコピンやフラボノイド生合成経路の鍵酵素遺伝子が共発現差異遺伝子に含まれる可能性が示唆され、重複遺伝子間でも数は少ないが有意な共発現差異遺伝子が存在することを示した(Fukushima et al. *Plant Physiol.*, 2012)。くわえて、大規模な機能ゲノクスデータから共発現差異遺伝子を同定するためのソフトウェア DiffCorr を開発・提案している(Fukushima *Gene*, 2013)その特徴を使って、さらなる遺伝子機能予測の向上が期待できることがわかった。

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 8 件)

- ① Miyako Kusano, Atsushi Fukushima, "Current challenges and future potential of tomato breeding using omics approaches", *Breed. Sci.*, 査読有, in press.
- ② Atsushi Fukushima, Miyako Kusano, "Recent progress in the development of metabolome databases for plant systems biology", *Front. Plant Sci.*, 査読有, 4, 73 (2013).
- ③ Atsushi Fukushima, "DiffCorr: an R package to analyze and visualize differential correlations in biological networks", *Gene*, 査読有, 518, 209-214 (2013).
- ④ Keiko Yonekura-Sakakibara, Atsushi Fukushima, Kazuki Saito, "Transcriptome data modeling: an approach toward target identification for plant metabolic engineering", *Curr. Opin. Biotechnol.*, 査読有, 24, 285-290 (2013).
- ⑤ Atsushi Fukushima, Tomoko Nishizawa, Mariko Hayakumo, Shoko Hikosaka, Kazuki Saito, Eiji Goto, Miyako Kusano, "Exploring tomato gene functions based on coexpression modules using graph clustering and differential coexpression approaches", *Plant Physiol.*, 査読有, 158, 1487-1502 (2012).
- ⑥ Miyako Kusano, Atsushi Fukushima, Naoko Fujita, Yozo Okazaki, Makoto Kobayashi, Naoko Fujita Oitome, Kaworu Ebana, Kazuki Saito, "Deciphering starch quality of rice kernels using metabolite profiling and pedigree network analysis", *Mol. Plant*, 査読有, 5, 442-451 (2012).
- ⑦ Takayuki Tohge, Miyako Kusano, Atsushi Fukushima, Kazuki Saito, Alisdair R. Fernie, "Transcriptional and metabolic programs following exposure of plants to UV-B irradiation" *Plant Signal. Behav.*, 査読有, 6, 1987-1992

(2011).

- ⑧ Miyako Kusano, Takayuki Tohge, Atsushi Fukushima, Makoto Kobayashi, Naomi Hayashi, Hitomi Otsuki, Youichi Kondou, Hiroto Goto, Mika Kawashima, Fumio Matsuda, Rie Niida, Minami Matsui, Kazuki Saito, Alisdair R. Fernie, "Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to ultraviolet-B light" *Plant J.*, 査読有, 67, 354-369 (2011).

[学会発表] (計 8 件)

- ① Atsushi Fukushima, "DiffCorr: an R package to analyze and visualize differential correlations in biological networks", PRIB 2012, Tokyo, Japan, November 8-10, 2012.
- ② Keiko Sakakibara, Atsushi Fukushima, Ryo Nakabayashi, Kousuke Hanada, Fumio Matsuda, Satoko Sugawara, Eri Inoue, Takashi Kuromori, Takuya Ito, Kazuo Shinozaki, Bunyapa Wangwattana, Mami Yamazaki, Kazuki Saito, "Transcriptome Independent Component Analysis Lead to the Identification of Two Glycosyltransferases Involved in Anthocyanin Modification in Arabidopsis thaliana", 26th International Conference on Polyphenols (ICP2012), Florence, Italy, July 22-26, 2012.
- ③ Atsushi Fukushima, Tomoko Nishizawa, Mariko Hayakumo, Shoko Hikosaka, Kazuki Saito, Eiji Goto, Miyako Kusano, "Characterizing coexpression modules using graph clustering and differential coexpression approaches in Solanum lycopersicum", 8th Solanaceae and 2nd Cucurbitaceae Genome Joint Conference (SOL&ICuGI 2011), Kobe, Japan, November 28-December 2, 2011.
- ④ Atsushi Fukushima, Tomoko Nishizawa, Mariko Hayakumo, Shoko Hikosaka, Kazuki Saito, Eiji Goto, Miyako Kusano, "Exploring tomato gene functions based on coexpression modules using graph clustering approaches", The 2011 Annual Conference of the Japanese Society for Bioinformatics, Kobe, Japan, November 8-10, 2011.
- ⑤ Keiko Sakakibara, Atsushi Fukushima,

Ryo Nakabayashi, Kousuke Hanada, Fumio Matsuda, Satoko Sugawara, Eri Inoue, Takashi Kuromori, Takuya Ito, Kazuo Shinozaki, Bunyapa Wangwattana, Mami Yamazaki, Kazuki Saito, "Two Glycosyltransferases Involved in Anthocyanin Modification Delimited by Transcriptome Independent Component Analysis in Arabidopsis thaliana" Gordon Research Conference on Plant Metabolic Engineering, Waterville, Valley, USA, July 24-29, 2011.

- ⑥ 福島 敦史, "DiffCorr: an R package to analyze and visualize differential correlations in biological networks", 生命医薬情報学連合大会(2012年日本バイオインフォマティクス学会年会情報計算化学生物学会(CBI学会)年次大会オミックス医療研究会年会), 東京, 日本, 2012年10月28-31日.
- ⑦ 福島 敦史, 西澤 具子, 早雲 まり子, 彦坂 晶子, 斉藤 和季, 後藤 英司, 草野 都, "Characterizing coexpression modules using graph clustering and differential coexpression approaches in Solanum lycopersicum", 第53回日本植物生理学会年会, 京都, 日本, 2012年3月16-18日.
- ⑧ 榊原 圭子, 福島 敦史, 中林 亮, 花田 耕介, 松田 史生, 菅原 聡子, 井上 恵理, 黒森 崇, 伊藤 卓也, 篠崎 一雄, Wangwattana Bunyapa, 山崎 真巳, 斉藤 和季, "独立成分分析によるアントシアニン修飾関連配糖化酵素遺伝子の機能同定", 第53回日本植物生理学会年会, 京都, 日本, 2012年3月16-18日.

[その他]

ホームページ等

DiffCorr package

<http://diffcorr.sourceforge.net/>

6. 研究組織

(1) 研究代表者

福島 敦史 (FUKUSHIMA ATSUSHI)

独立行政法人理化学研究所・メタボローム情報チーム・研究員

研究者番号: 80415281

(2) 研究分担者

なし

(3) 連携研究者

なし