

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 13 日現在

機関番号：82626

研究種目：若手研究(A)

研究期間：2012～2015

課題番号：24680032

研究課題名(和文)マイクロRNA機能推定に向けたネットワーク解析手法の開発

研究課題名(英文)Development of network analysis methods for functions of RNAs

研究代表者

瀬々 潤 (Sese, Jun)

国立研究開発法人産業技術総合研究所・創薬基盤研究部門・主任研究員

研究者番号：40361539

交付決定額(研究期間全体)：(直接経費) 16,800,000円

研究成果の概要(和文)：本研究では主として2個の成果がある。一つは与えられた生物学的ネットワークを探索するアルゴリズムの構築である。たんぱく質相互作用等のネットワークとマイクロRNAのターゲットとする遺伝子群を与えた時、複数のマイクロRNAがターゲットとする部分ネットワークを高速に探索する。もうひとつはネットワークを構築する統計的なアルゴリズムである。転写因子やマイクロRNAによる転写制御は、複数の因子が同時に関わることで転写の制御が行われる。その関係を統計的有意性を持って発見できる手法の開発に成功した。

研究成果の概要(英文)：This study contains two major results. One is development of an analysis algorithm for a large biological network such as protein-protein interactions. Given the network and connection between microRNAs and their target genes, the algorithm can efficiently find the subnetworks affected by multiple microRNAs. Because effects of most microRNAs are modest, such overlaps would strongly inhibit the expressions of target genes in an environment. The other is a method to construct networks statistically. Given candidate regulators of genes, such as transcription factors/micro RNAs, and a expression profile, the algorithm can detect combinations of regulators that activate or inhibit expressions with statistical significance.

研究分野：バイオインフォマティクス

キーワード：ネットワーク 転写制御 統計的有意性

1. 研究開始当初の背景

ゲノム関連の技術の進歩は目覚ましい。特にゲノム配列を観測する技術の高速化(並列化)と価格の低下は著しく、特に価格は1年で半額になるペースで下落をしていた。この研究開始当初は、価格の下落により、例えば、遺伝子発現量の網羅的観測法として名を馳せたマイクロアレイ技術は、ゲノム配列を利用した RNA-seq 法へと置き換わろうとしている時期であった。また、RNA-seq 法は、単純なマイクロアレイ法の置き換えだけではなく、「配列を読む」という特徴を活かすことで、遺伝子発現量以上のもの、たとえば、存在や機能が確認されることはあるものの、その全体像は未知であるマイクロ RNA を観測することが可能になったり、人間は父方母方両方のゲノムを有する二倍体であるが、そのどちらのゲノムから発現した遺伝子であるのが理解できたり、更には、選択的スプライシングが今まで以上に観測できたりと、様々な情報の観測が可能となってきた。

2. 研究の目的

本研究の目的は、ネットワーク解析に着目することで、RNA やマイクロ RNA の機能を明らかにする方法を開発することである。RNA の機能は、マイクロアレイ法などを用いることで網羅的に観測され、米国 NCBI の Gene Expression Omnibus などを通じて観測結果が公開されている。マイクロ RNA はガン細胞や神経細胞においてもマイクロ RNA は重要な役割を果たしていることが明らかとなっている一方、その全容は明らかではない。これらの機能やどのように制御されているか、あるいは、これらがどのように転写制御に関わっているかを調べる手法を作ることができれば、生命の理解を遺伝子発現情報のビッグデータをもとに解析が可能になるのではないかと考えた。

3. 研究の方法

本研究は大きく2つに別れる。一つは生物学的実験を通じて構築されているネットワークを解析するネットワーク解析手法の構築。もう一つは、データからネットワークを作る方向に向かう研究である。

前者の研究は、たんぱく質相互作用やパスウェイ情報として構築されているネットワーク情報を利用した。これらのネットワークは、ノードが遺伝子やたんぱく質、エッジ(リンク)がそれらの間に何らかの関係があることを示している。各遺伝子は、ネットワーク構築とは独立した研究で遺伝子発現等の情報が観測されているので、それらを統合することで、いつどこで働いているのかを明確化することができると思った。また、マイクロ RNA

と遺伝子間の制御関係が HITS-CLIP 法等で観測され、データベースに蓄積されているので、その情報の統合を行った。

後者のネットワークを構築に向けた戦略は、ネットワーク一般に2つの要素の間を結びリンク構造を考えているが、例えば iPS 細胞化の山中因子を考えると4つの転写因子が関わっている。このように1対1ではなく、多対1の関係を発見する必要性が出てくる。しかしながら、このような複合的な関係を考え始めると、計算時間の爆発、多重検定による偽陽性を防ぐ必要性の2点から、統計的に有意な結果を得ることが難しい問題点が存在していた。これらの問題の解決を行った。

4. 研究成果

ネットワーク解析の問題は、ネットワークノードの上に複数のラベルが貼られているマルチラベル問題でモデル化し、頻出パターン解析において取れるパターンに制約がある状況を考える、という定式化により、全たんぱく質相互作用を考えたとしても高速に計算可能なアルゴリズムを開発した。更に、その並列計算にも取り組んだ。結果として、臓器特異的に働く部分ネットワークや、複数のマイクロ RNA が共通してターゲットとすることで、遺伝子の働きを止めているネットワークの推定を行うことが可能となった。

ネットワーク構築に向けた戦略は、多対1の関係を見つけるために必要である計算の高速化と多重検定を行った上での統計的有意性の保証を同時に解決するアルゴリズムの構築を行った。内容は、高速な計算のために頻出パターン分析法を用い、統計的な保証に関しては、既存の多重検定法が非常にゆるい上界を利用していたのに対して、タロンの補正を用いることで、今までより厳しい上界を求める事を可能とした。これらを組み合わせることで、高速に全組合せを探索し、かつ、統計的有意性も保証できる手法が構築できた。

また、酵母およびヒトの乳がん細胞から観測した遺伝子発現情報と公共の転写因子情報を統合して解析することで、酵母では既存の補正法では見逃されてきた4遺伝子の組合せを、ヒトからは8遺伝子の組合せを推定することが可能となった。

5. 主な発表論文等

[雑誌論文](計15件)

Masahiro Tokunaga, Chikara Kokubu, Yusuke Maeda, Jun Sese, Kyoji Horie, Nakaba Sugimoto, Taroh Kinoshita, Kosuke Yusa and Junji Takeda. Simulation and

estimation of gene number in a biological pathway using almost complete saturation mutagenesis screening of haploid mouse cells. *BMC Genomics*. 15: 1016, 2014.

Makoto Yamada, Masashi Sugiyama, and Jun Sese. Least-squares independence regression for non-linear causal inference under non-gaussian noise. *Machine Learning*. Vol. 96 (3), pp. 249-267, 2014.

Shin-ichi Minato, Takeaki Uno, Koji Tsuda, Aika Terada and Jun Sese. A fast method of statistical assessment for combinatorial hypotheses based on frequent itemset enumeration. *Machine Learning and Knowledge Discovery in Databases (ECML/PKDD 2014)*. LNCS 8725, pp. 422-436, 2014.

Shingo Okuno, Tasuku Hiraishi, Hiroshi Nakashima, Masahiro Yasugi, and Jun Sese. Parallelization of extracting connected subgraphs with common itemsets. *Information and Media Technologies*. Vol. 9, No. 3, pp. 233-250, 2014.

Satoru Akama, Rie Shimizu-Inatsugi, Kentaro K. Shimizu, and Jun Sese. Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*. *Nucleic Acids Research*. 42 (6), e46, 2014.

Shiori Setoguchi, Hisaki Takamori, Tadashi Aotsuka, Jun Sese, Yukio Ishikawa, Takashi Matsuo. Sexual dimorphism and courtship behavior in *Drosophila prolongata*. *Journal of Ethology*. Vol. 32, Issue 2, pp. 91-102. May, 2014.

Aika Terada, Koji Tsuda, and Jun Sese. Fast Westfall-Young permutation procedure for combinatorial regulation discovery. *IEEE Bioinformatics and Biomedicine (BIBM 2013)*. pp. 153-158, 2013.

Aika Terada and Jun Sese. Bonferroni correction hides significant motif combinations. *13th IEEE Bioinformatics and Bioengineering (BIBE 2013)*, 2013.

Aika Terada, Mariko Okada-Hatakeyama, Koji Tsuda, Jun Sese. Statistical significance of combinatorial regulations. *Proc. Natl. Acad. Sci.*, Vol. 110, No. 32, pp. 12996-13001, 2013.

Akiko Izawa and Jun Sese. RECOT: A Tool for the Coordinate Transformation of Next-Generation Sequencing Reads for comparative genomics and transcriptomics. *Source Code for Biology and Medicine*, 2013, 8:6.

Atsushi Ogura, Konstantinos Sousounis, Takeya Moritaki, Yuki Okuda, Jun Sese, Kentaro Shimizu, Masa-aki Yoshida, Panagiotis A. Tsonis. Loss of the six3/6 controlling pathways might have resulted in pinhole-eye evolution in *Nautilus*. *Scientific Reports*, 3, 1432, 2013.

Chih-Chieh Yu, Mari Furukawa, Kazuhiro Kobayashi, Chizuru Shikishima, Pei-Chiang Cha, Jun Sese, Hiroko Sugawara, Kazuya Iwamoto, Tadafumi Kato, Juko Ando, Tatsushi Toda. Genome-wide DNA methylation and gene expression analyses of monozygotic twins discordant for intelligence levels. *PLoS ONE*, 7(10): e47081, 2012

Mutsumi Fukuzaki, Masa-aki Yoshida, Atsushi Ogura and Jun Sese. Systematic Measurement of Mismatch Effect for Designing Inter-Species Microarray. *IEEE Bioinformatics and Biomedicine 2012 (BIBM 2012)*, pp. 478-481, 2012.

Jun Sese. Data Analysis of Large-Scale Glycan-Sample Interactions. *Trends in Glycoscience and Glycotechnology*, Vol. 24, No. 137, pp. 122-128, 2012.

Rina Nakazawa, Takayuki Itoh, Jun Sese, and Aika Terada. Integrated visualization of gene network on ontology applying hierarchical graph visualization technique. *16th Int. Conf. on Information Visualization (IV) 2012*, pp. 81-86, July 11-13, 2012.

〔学会発表〕(計 25 件)

RECOT: リードの種間座標変換ツール. 2012年5月(ホテル阪急エキスポパーク・大阪)

ビッグデータのもたらす生命科学の新潮流. ITBIO融合プログラム. 2012年7月(琉球大学・沖縄)

Gene expression analysis in polyploid species using NGS. 2012年8月(北海道大学・北海道)

生命科学研究に変化をもたらす計算機屋の飼いかた. 2012年8月(足助里山ユースホステル・愛知県)

バイオインフォマティクス～計算機が先

導する生命現象の解明. 2012年12月(はこだて未来大学・北海道)

計算機を積極的に利用した生命科学研究の推進. 2013年4月(筑波大学菅平・長野県)

ゲノムワイド実験時代の統計的データ解析～偽陽性を防ぐために. 2013年6月(大阪大学・大阪)

組合せで働く転写因子発見に向けた検出力の高い多重検定法. 2013年6月(北海道大学・北海道)

バイオインフォマティクスによる遺伝子発現解析. 2013年7月(理化学研究所・兵庫)

Statistical significance of combinatorial regulations. 2013年7月(理化学研究所・神奈川)

On statistical significance for untangling complex genotype-phenotype connection. 2013年9月(チューリッヒ大学・スイス)

Revisiting statistical significance for untangling complex genotype-phenotype associations. 2013年10月(ソウル・韓国)

複雑な遺伝型-表現型を解くための多重検定補正. 2013年10月(慶応大学・神奈川)

全ゲノム時代の情報・数理・統計. 2013年10月(東京大学・東京)

Statistical significance of combinatorial features with frequent itemset mining. 2013年10月(イリノイ大学・アメリカ)

Revisiting statistical significance for untangling complex genotype-phenotype associations. 2014年1月(九州工業大学・福岡)

ゲノム網羅的な遺伝子発現量の解析～情報解析の観点から、実験計画を考える～. 2014年2月(金沢大学・金沢)

生命科学に革新を与える情報・統計学を目指して. 2014年10月(仙台国際会議場・宮城)

ゲノム網羅的計測情報の解析に向けた情報・統計手法の再考察. 2014年10月(東京大学・東京)

Understanding the complex associations between genes and traits by new statistical approaches. 2014年11月(パシフィコ横浜・神奈川県)

①木を見て森も見ると同時に：相互作用を見つけるためのデータ解析技術. 2014年11月(東京大学・東京)

② On bridging genome informatics innovation to life science industry in Odaiba. 2014年12月(プラザ平成・東京)

③RNA-seqによる変動遺伝子抽出の統計. 2014年12月(産総研臨海センター別館・東京)

④Statistical assessment for untangling relationships among genes, traits and environments. 2014年12月(京都大学・京都)

⑤協調的に働く発現調節因子の発見に向けた統計的データ解析. 2014年12月(京都大学・京都)

〔図書〕(計 1件)

Tetsuo Shibuya, Hisashi Kashima, Shandar Ahmad, and Jun Sese (eds). Pattern Recognition in Bioinformatics. Springer, 2012.

〔産業財産権〕
出願状況(計 0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計 0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

6. 研究組織

(1)研究代表者

瀬々 潤 (SESE, Jun)

産業技術総合研究所・生命情報科学研究センター・研究チーム長(研究終了時)

研究者番号：40361539

(2)研究分担者

なし

(3)連携研究者

なし