

## 課題名 絶対定量オーミックスからの知識発見

代表研究者 東京大学大学院新領域創成科学研究科 教授 伊藤 隆司

### 1. 代表研究者による成果概要報告

#### 1-1. 研究目的

ポストゲノムシーケンス時代を迎え、様々な機能ゲノミクスデータの蓄積が急速に進んでいる。機能ゲノミクスが扱う分子情報、即ち遺伝子発現・蛋白質発現・翻訳後修飾・相互作用・局在等は、ゲノム配列が基本的に不変であるのとは対照的に、生物学的コンテキストに応じて変動する。したがって、これらの情報を生命のシステムの理解に活用するには、定性的記述のみでは不十分で、ダイナミクスの理解につながる定量的記載が不可欠である。特に、相対変化のみならず絶対値を把握すると、分子ネットワークのリアルな像とそれに基づく正確な予測が可能になる。と同時に、データ互換性が格段に向上し、定量性を上手に利用することで新知識の発見も加速されると期待される。こうした認識に基づいて本研究開発課題では、計測を支援するバイオインフォマティクス技術も含めた絶対定量データ取得法の確立と、データの定量性を活用して新知識発見を促す解析手法の構築を目標に開発を進めた。

#### 1-2. 研究開発概要

上記目的を達成する為に、以下の4グループからなる体制で研究開発を進めた。

- 1) グループ A (遺伝子発現および蛋白質-核酸相互作用絶対定量解析グループ)
- 2) グループ B (蛋白質-脂質相互作用定量解析グループ)
- 3) グループ C (定量オーミックス解析からの知識発見手法開発グループ)
- 4) グループ D (定量オーミックスデータ統合解析支援システム開発グループ)

具体的には、グループ A は出芽酵母をモデルに、発現遺伝子絶対定量法 GATC-PCR による計測システムを、プライマー設計やデータ処理のソフトウェア開発も含めて構築し、出芽酵母トランスクリプトームの絶対定量を行なった。更に、転写因子-プロモータ相互作用解析への応用を目指して全長 cDNA 解析に基づくプロモータ同定とプライマー設計を行なった。その過程で 1000 種を優に越える非コーディング RNA を同定し、酵母トランスクリプトームの予想外の複雑性を明らかにした。一方、蛋白質の絶対定量に関しては、標準ペプチドを連結した人工蛋白質を標準物質として質量分析法を用いて計測する技術を開発した。グループ B は蛋白質-脂質相互作用解析法の検討を続け、グループ A と共同で出芽酵母の全 PX ドメインについて各種フォスフォイノシタイド (PI) との相互作用解析を行った。また PI 相互作用に影響を及ぼす PX ドメイン内のオルタナティブスプライシングの例を見出した。グループ C は、文献からの相互作用情報抽出、相互作用の予測と相互作用データからの知識発見、機械学習を用いた DNA 上のヌクレオソーム占有領域およびヒストン修飾領域の予測と解

析について研究を行った。グループDは、グループAおよびCと共同で、定量データを知識発見に活用する仕組みとして、遺伝子発現定量データと既知特徴情報（アノテーション情報）を関連付けることによって、遺伝子発現実験結果の解釈を支援する手法の研究開発を行い、ART (Annotation Rating Tool)としてツール化した。また特異発現パターンを示す特徴遺伝子群（特定の特徴を有する遺伝子グループ）の抽出手法の研究も進めた。これらの手法を中心としたトランスクリプトーム解釈支援システム EAST (Enhanced Annotator for Saccharomyces Transcriptome)を開発した。

### 1-3. 研究成果概要

#### ○遺伝子発現および蛋白質-核酸相互作用絶対定量解析グループ

- 1) 発現遺伝子絶対定量法 GATC-PCR の高度化を進めた。
- 2) GATC-PCR や RACE に要求される高性能プライマーの設計法を考案し、それを実装したプログラム SDSSPrimer を開発した。
- 3) SDSSPrimer でデザインしたプライマーを用いて出芽酵母トランスクリプトームの絶対定量を行った。
- 4) 転写因子-プロモータ相互作用の定量解析の基盤情報として、全長 cDNA 解析により 3,683 遺伝子の転写開始点を決定し、GATC-PCR 実験をデザインした。
- 5) 上記の解析の過程で、1000 種を超える非コーディング RNA を全長 cDNA として同定し、酵母トランスクリプトームの予想外の複雑性を明らかにした。
- 6) GATC-PCR の発想を蛋白質に応用したペプチド連結型標準物質を用いた質量分析による蛋白質絶対定量法を開発した。

#### ○蛋白質-脂質相互作用定量解析グループ

- 1) 出芽酵母の全 PX ドメインについて各種フォスフォイノシタイド (PI) との相互作用解析を行った。
- 2) PX ドメインと PI の相互作用を解析し、結合能を著しく損なうオルタナティブスプライシングの例を見出した。

#### ○定量オーミック解析からの知識発見手法開発グループ

- 1) 大規模な名詞辞書に基づき、相互作用を含む名詞間の関係情報を抽出するためのテンプレートを自動的に発見する手法を開発した。
- 2) 転写因子結合サイトのデータと遺伝子発現データの組み合わせに対して相関ルール発見アルゴリズムを用いることにより、着目すべき転写調節モジュールを発見する手法を開発した。
- 3) DNA 上でヌクレオソームによる占有とヒストン修飾が行われる領域をコンピュータで予測し解析する方法を提案した。

#### ○定量オーミックデータ統合解析支援システム開発グループ

- 1) トランスクリプトームの解釈支援手法 ART を開発し、既報告トランスクリプトームやグループ A のデータを用いた評価において、その有効性を示した。
- 2) 特異な発現パターンを持ち、かつ特定の特徴を共有する遺伝子グループの抽出手法を開発、評

価を実施した。

3) 上述の二つの手法を実装したトランスクリプトーム解釈支援システム EAST を開発した。EAST はこれらの手法の使い勝手を向上される仕組みや、その他の様々なツールも提供している。

## 2. 事後評価結果

### 2-1. 当初計画の達成度

○当初の研究開発計画から見た達成度等はどうか

蛋白質-核酸相互作用絶対定量解析では、GATC-PCR による定量的トランスクリプトーム解析、そのためのプライマー設計法の開発、酵母の転写開始点の網羅的同定など、途中で生じた問題点の克服も含め十分な達成度が得られた。しかし、蛋白質-脂質相互作用の定量解析における技術開発の点では、解析対象とする物質の性質上困難な点も多く十分に達成したとは言えなかった。知識発見手法や支援システムの開発ではインフォマティクスとしての新規性は少ないものの、実験支援技術としては、実用的かつ有効な方法が開発されたといえる。

○当初計画では想定されていなかった新たな展開が生じたか

当初想定されていなかった展開として、ORF 単位とは無関係に多様な部位から、しかも DNA のセンス・アンチセンスの両鎖からの転写が予想外に多く見出され、転写の複雑性を見出したことがあげられる。このことから、トランスクリプトーム自身の考え方や解析法の再検討を迫られる結果となったが、科学的な価値は高い。本研究の大いなる成果と見ることも出来る。非コード RNA の問題への展開は興味深い。

### 2-2. 知的財産権、外部発表（論文等）等研究開発成果の状況

論文や解析ツール、データベースの公開など質的にも量的にも十分な成果が得られている。

### 2-3. 研究開発成果の公開による波及効果

○データベース・プログラムや論文、学会発表など、これまでの成果公開によるバイオインフォマティクス領域およびライフサイエンス分野全体への影響や効果はどうか。

開発された解析技術、プログラムに関しては基本的には酵母以外の生物にも応用できるものであるが、現段階では酵母の知識に留まっており、ライフサイエンス分野全体には及ばない。ゆえに、現時点で目に見える形での波及効果は表われていないが、技術開発として期待できるものが得られていると考えられる。今後はこれらの成果を広める努力が必要であろう。

○国内外の類似研究開発の動向と比較して本課題の意義はどうか。

酵母の転写開始点を網羅的に決定できたことは大きな意義があり、データベースは国際的にも最大のものとして注目されているようである。また、GATC-PCR による遺伝子発現の定量化においても海外の real competitive PCR との比較がなされており、それよりも良好な結果が得られている。

#### 2-4. 成果の実用化の可能性及び成果から予想される波及効果

○今後、成果の生命科学、研究基盤、産業への貢献度はどう見込まれるか。

絶対定量オーミックスは重要な課題であり、今回の成果が基盤的手法あるいは技術になることが必要である。今後高等生物への展開が図られれば、その波及効果は大きく期待できる。

#### 2-5. 総合評価

実験とインフォマティクスの融合という点で、これからの生命科学のパイロット的な研究であったと思う。その結果として予定の計画を十分にクリアし、新しい問題提起にもつながった。本解析技術やデータは基盤となるものであり、今後の発展とその支援をどうしていくかが評価者と被評価者、両者にとっての課題である。また、直接の目的ではなかったが情報科学と生物科学の融合に向けた人材が育ったことは評価される。

### 3. 主な論文発表等

1. Miura, F., Uematsu, C., Sakaki, Y. & Ito, T.: A novel strategy to design highly specific PCR primers based on the stability and uniqueness of 3'-end subsequences. *Bioinformatics* **21**, 4363-4370 (2005).
2. Pham, T.H., Clemente, J.C., Satou, K. & Ho, T.B.: Computational discovery of transcriptional regulatory rules. *Bioinformatics* **21**, Suppl.2, 101-107 (2005).
3. Pham, T.H., Tran, D.H., Ho, T.B., Satou, K. & Valiente, G.: Qualitatively predicting acetylation and methylation areas in DNA sequences. *Genome Inform.* **16**(2), 3-11 (2005).
4. Miura, F., Kawaguchi, N., Sese, J., Toyoda, A., Hattori, M., Morishita, S. & Ito, T.: A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc. Natl. Acad. Sci. USA* **103**, 17846-17851 (2006).
5. Yamaguchi, Y., Ota, K. & Ito, T.: A novel cdc42-interacting domain of the yeast polarity establishment protein Bem1: implications for modulation of mating pheromone signaling. *J. Biol. Chem.* **282**, 29-38 (2007).

#### 参考

論文発表 国内 0 件、海外 23 件

口頭発表 国内 5 件、海外 6 件

ポスター発表 国内 3 件、海外 2 件