

絶対定量オーミックスからの知識発見

東京大学大学院新領域創成科学研究科情報生命科学専攻

伊藤 隆司

How can we learn more from quantitative omics data ?

Takashi Ito

Department of Computational Biology,

Graduate School of Frontier Sciences,

The University of Tokyo

Genome analysis and functional genomics have thus far contributed to the first and second steps of systems-level understanding of the biological systems, or the identification of systems components and the deciphering of networks among them, respectively. However, the current knowledge remains largely qualitative and limited in terms of spatiotemporal resolution. We thus expect quantitative omics to expand and deepen our understanding of the biological systems. Here we report our wet and dry efforts for the absolute quantification of transcriptome and subsequent knowledge discovery by exploiting the quantitative nature of the data.

1. はじめに

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。こうしたオーミックデータの活用によって、システム生物学も現在よりも大きなシステムの理解を射程に収めることが出来ると期待されている。システムの理解の第一歩は構成要素の同定であるが、ゲノム解析による遺伝子の網羅的発見はそのための究極の基盤整備でもあった。これに続く段階は、要素間の関係性の同定（ネットワーク解明）であり、機能ゲノミクスは従来にはない形でこの情報の充実を支えている。

しかし、基本的に不変であるゲノム配列とは対照的に、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・局在等は全て生物学的コンテキストに応じて変動するものである。したがって、システムの理解には、定性的記述のみでは不十分で、定量的計測に基づくダイナミクスの理解が不可欠である。その際に、相対的な変化のみならず絶対値をも把握すると、絶対定量では見えないよりリアルなネットワーク像を得ることが出来る。と同時に、データ互換性は格段に向上し、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ことが期待される。

こうした認識に基づいて本研究開発課題では、支援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方法の構築を目標に開発を進めた。

2. 研究開発の成果

我々は、発現遺伝子および蛋白質結合核酸の絶対定量の為の基盤技術として、全遺伝子を確実に1分子ずつ含むゲノムDNAを天然の理想的標準として計測を行なうアダプター付加競合増幅PCR (GATC-PCR) の開発を進めてきた。本手法の評価として、同一サンプルを3分割しreal time PCRおよび質量分析によるreal competitive PCRという既存手法との比較検討を他機関の協力も得て行い、その信頼性を確認した。

GATC-PCRによる増幅では、その特異性が1本の遺伝子特異的プライマーに依存するので、その設計には特別の配慮が必要とされる。我々は増幅副産物の構造解析に基づいて、プライマーと鋳型の会合率 f という概念を導入し、 f が閾値を越える3'側最小部分配列をSpecificity-Determining Sub-Sequence (SDSS) と定義した。各プライマーのSDSSが対象ゲノム中に出現する頻度によってそのプライマーの特異性を評価する独自のアルゴリズムによるプライマー設計を行なったところ、その有効性が確認された[1]。そこでこれを実装したプログラムを公開した[2]。また大量の電気泳動データからのバンド同定や定量の為のソフトウェア (SizeCaller, AQUOS) もデータ生産現場からのフィードバックを繰り返しながら開発した。これらを自動分注装置、384穴サーマルサイクラー、マルチキャピラリーシークエンサーを組み合わせたデータ生産系と統合してGATC-PCRシステムを構築し、酵母トランスクリプトームの計測を行った。

本技術を転写因子-DNA相互作用の解析に応用するには、プロモータ領域の同定が不可欠であるが、酵母ではそれに必要な転写開始点 (TSS) 情報の欠落が顕著である。そこでベクターキャップ法による全長cDNA解析を行ない、3,600遺伝子のTSS情報を得て世界最大のデータセットを構築した[3]。と同時に、遺伝子間領域の小遺伝子、アンチセンス転写物、ORF内TSSによる転写物等の新規転写物を1,000種以上同定し、トランスクリプトームの予想外の複雑さを明らかにした。この結果は、GATC-PCRのアンプリコン設計のみならず、DNAチップの設計にも重要な変更を迫るものであろう。

なおGATC-PCRはゲノム中の遺伝子コピー数の計測に極めて有効であり、プライマー設計の高度化によりヒトゲノムのコピー数変異の検出にも応用が可能であると思われる。またGATC-PCRのアナロジーから、タンパク質ネットワークの絶対定量技術の開発にも成功した。この技術は、質量分析で定量するペプチドを縦列に連結した人工タンパク質を合成して標準として使用するものであり、タンパク質間相互作用 (複合体) の定量解析が可能であることを示したが (投稿中)、核酸や脂質との相互作用の定量的評価にも応用可能であろう。

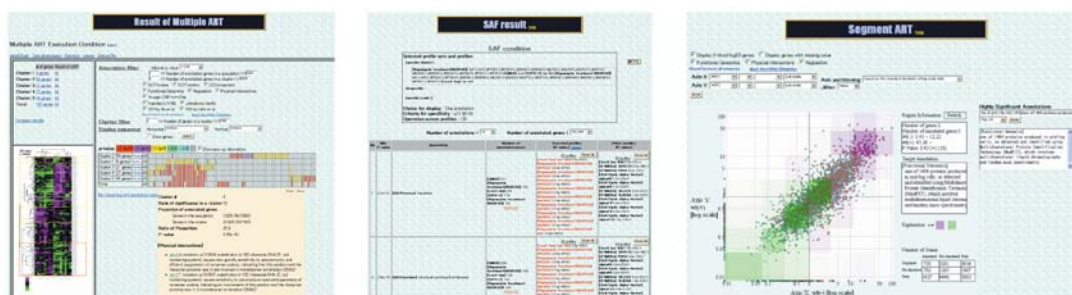
計測系の構築と平行して、絶対定量データの理解や解釈を支援するシステムの開発を進めた。まず、発現量を指標に選択された遺伝子集団が共有する機能や特徴を、各遺伝子に関連付けられている既存知識を整理して提示することで、研究者の理解を促進する手法の研究を進めART/G3を開発した。既存知識としてはGene Ontology (GO) のみならず、Yeast Proteome Database (YPD) の機能ゲノミクス・アノテーション (様々な条件下で計測されたマイクロアレイの結果等) やChIP-Chipデータ、比較ゲノムによる保存上流配列情報等も利用できる。その結果、非専門家がオリエンテーションを得るのに適したGOに基づく基礎的解釈の提供のみならず、専門家

による新知識発見の契機をも提供できるようになった。

次に、データの定量性を知識発見に活用する為に、従来は遺伝子を絞り込むためにのみ使用されてきた発現量を、絞り込んだ遺伝子群に解釈を施す際の「重み」として活用できるように上記手法の拡張を行い、それを実装したART/EXを開発した。ART/EXを用いると、トランスクリプトームデータ全体をそのまま解析して、それを「遺伝子リスト」から構造を持った「機能表現リスト」へと変換することもできる。これによりトランスクリプトームの特徴抽出が極めて容易になり、また遺伝子選択に伴う情報損失を回避できる（投稿中）。

更に、単一のプロファイルの解釈だけでなく、複数のプロファイル間で共有されている機能的特徴を抽出するツール（SAF）や、逆にユーザーが興味を持った特定の機能の組み合わせを共有する発現プロファイルを検索するツール（EPF）も開発した。

こうして開発した手法やツールはWebベースの統合解釈支援システム「EAST: Enhanced Annotator for Saccharomyces Transcriptome」として公開した[4]。EASTはクラスタリング機能や上流配列解析機能も備えており、アップロードしたユーザーデータのクラスタリング、上記ツールを駆使したクラスタの自動解釈（下図左）、更に複数のデータセット間の共通性検索（下図中央）など、これまで研究者が様々なサイトやツールを渡り歩きながら行ってきた一連の解析をシームレスに展開することが出来る。また、絶対定量データの場合には、その特質を活かすために、比率のみならず差分の視点も加えた複眼的解析が可能な環境が提供されている（下図右）。



トランスクリプトームの機能的解釈と並んで、分子機構解析の為に転写制御規則の発見手法の開発も進めた。特に、転写因子とDNAの相互作用データと遺伝子発現データを組み合わせ、転写因子の発現量の変化と転写対象遺伝子の発現量の変化がどの程度関係しているかを量的に表現し、ルール発見手法であるCN2-SDを応用することにより、「転写因子Xと転写因子Yの発現量が増加すれば、対象遺伝子Zの発現量が減少する」というような規則を発見する手法を開発し、その有効性を示すことが出来た[5]。また遺伝子発現の基盤と注目されているヒストンのアセチル化・メチル化に関してもそれらを予測する手法を開発した[6]。

3. まとめ

実験手法・計測技術と関連するバイオインフォマティクス技術を連携しながら研究し、独自の技術群を開発した。トランスクリプトーム計測に関しては、トランスクリプトーム像の変化と最近の技術革新に伴う新しい計測法の必要性を感じているが、定量値を解釈に活用するART/EXなどの仕組みは、計測法によらず今後も有効に利用できると思われる。

これらの一連のツールはトランスクリプトームデータのみならず、遺伝子に数値が付加されたプロファイルであれば、それが何であれ遺伝子機能のプロファイルへと変換することが出来る。したがって、トランスクリプトーム、プロテオーム、フェノームの枠を超えて、実験を機能という観点で相互に比較することが可能になり、そこからの新たな知識発見を促すことが出来る。またGOを用いると種を越えた比較も原理的には可能であると思われ、比較トランスクリプトーム・比較機能ゲノムツールへの展開を予定している。

問題点としては、専門家が求める新知識発見には、最新の文献に基づく詳細で良質なアノテーション情報が欠かせない点があげられる。今回の開発研究では、それらを商用データベースYPDに頼らざるを得なかった。今後こうした情報が、公的なデータベースから入手できるような環境が整備されることを期待したい。

4. 研究開発実施体制

代表研究者 伊藤 隆司（東京大学大学院新領域創成科学研究科）

研究開発題目

- (1) 遺伝子発現および蛋白質-核酸相互作用絶対定量解析

グループリーダー 伊藤 隆司（東京大学大学院新領域創成科学研究科）

- (2) 蛋白質-脂質相互作用定量解析

グループリーダー 住本 英樹（九州大学生体防御医学研究所）

- (3) 定量オーミック解析からの知識発見手法開発

グループリーダー 佐藤 賢二（北陸先端科学技術大学院大学知識科学研究科）

- (4) 定量オーミックデータ統合解析支援システム開発

グループリーダー 吉田 美寸夫（インテックウェブアンドゲノムインフォマティクス(株)）

5. 参考文献

[1] Miura, F. et al. *Bioinformatics* 21 (24), 4363-4370, 2005.

[2] <http://itolab.cb.k.u-tokyo.ac.jp/GATC/SDSSPrimer.html>

[3] Miura, F. et al. *Proc. Natl. Acad. Sci. USA*, in press.

[4] <http://itolab.cb.k.u-tokyo.ac.jp/BIRD/GATC-PCR>

[5] Pham, T. H. et al. *Bioinformatics* 21 (Suppl 2), ii101-ii107, 2005.

[6] Pham, T. H. et al. *Genome Inform.* 16 (2), 3-11, 2005.