

# インタラクトーム解析からの生物知識獲得

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 the interactome?

Takashi Ito

Department of Computational Biology,

Graduate School of Frontier Sciences,

The University of Tokyo

The interactomics should expand its scope to include not only protein-protein interactions but protein-nucleic acid and protein-lipid interactions. It should also aim at quantitative description with spatio-temporal resolution. Toward this goal, novel measurement methods have to be co-developed with bioinformatics techniques that support the experimentation and fully exploit quantitative nature of the data for the interpretation and subsequent knowledge discovery. Here we introduce such efforts in our BIRD project, including those for absolute quantitation of the budding yeast transcriptome to find novel knowledge on gene-gene interactions or gene networks.

### 1. はじめに

出芽酵母蛋白質の網羅的2ハイブリッド解析によって始まったインタラクトーム解析は、今や機能ゲノム科学の最重要分野のひとつとして認識されるに至った。それに伴いインタラクトームを扱うバイオインフォマティクス(BI)も長足の進歩を遂げつつある。この分野の次なる課題は、対象とする相互作用の拡大と解析の深化であろう[1]。その為には、インタラクトームインフォマティクスの更なる発展は勿論、新規の実験手法・計測技術の開発による新しいデータの取得も必要である。とりわけ、これまでのインタラクトームデータにはない特性、すなわち定量性と時間分解能、を備えたデータの取得が重要である。そして BI と連携しない機能ゲノム計測はあり得ないから、新しい大規模解析を支援する BI 技術とデータの特性を知識発見に活用する BI 技術が平行して開発されねばならない。

このような認識に基づいて、本研究開発では、従来の蛋白質間相互作用に加えて、遺伝子間相互作用、蛋白質-DNA 相互作用、および蛋白質-脂質相互作用を解析する技術の開発を試みた。特に従来までの定性的な相互作用カタログ化から、相互作用ネットワークのダイナミクス理解につながる定量的情報の獲得へと、内容を深めることを目指して定量に力点を置いた。そして定量計測を支援する BI 技術の開発も進めて大規模な計測系の構築を行い、データの定量性を知識発見に活用する新しい手法の開発を進めた。

## 2. 研究開発の成果概要

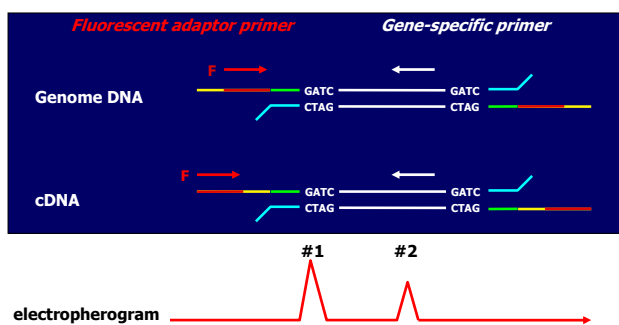
蛋白質間相互作用に関しては、まずデータマイニングの代表的手法の一つである「相関ルール発見」を用いて、相互作用する蛋白質間の関係に関するルールの抽出を試みた。これは先駆的な取り組みであり、既知ルールの再発見で手法の妥当性を示すことが出来た[2]。しかしその一方で、新規ルールの発見には至らず、用いる特徴データの改善の必要性が明らかになった。特に現在のインタラクトームデータが全長蛋白質同士のもので占められており、相互作用領域に関するデータの必要性が浮かび上がった。

そこで、蛋白質複合体の立体構造データに基づくサポートベクターマシンによる学習によって、相互作用部位を予測する研究に取り組んだ。その結果、特徴ベクトルの選択や表面パッチの取り扱いに工夫を凝らすことで、先行研究よりも高い精度での予測を行うことに成功した[3]。

その一方で相互作用部位を迅速に同定する実験手法として、構造を損なわずに特定の相互作用だけを障害する変異を同定する「保証付き逆2ハイブリッド法」を開発し、変異体による相互作用の機能解明や新規相互作用ドメインの発見で有効性を示した。またエラープローンPCRと2ハイブリッド選択によって、進化を加速すると同時に選択圧を特定の相互作用に局限することで、変異のフットプリントとして相互作用部位の予測を容易ならしめる「加速化系統フットプリント法」も考案した。後者は相互作用強度情報も統合することが可能で、BI との融合で今後、有効な方法論に成長し得ると期待される。

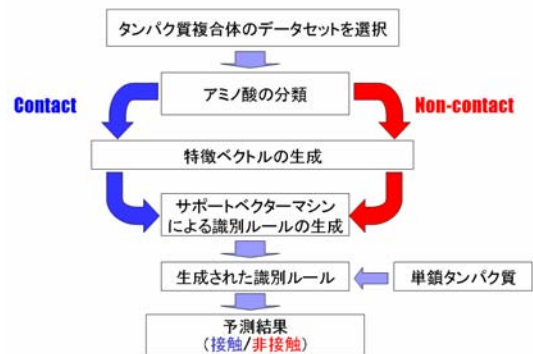
また敢えて挑戦的な課題として取り上げた蛋白質脂質相互作用に関しては、我々が見出した PX/PB2 ドメインを例に相互作用解析手法の開発を進めた。と同時に、このドメインと脂質の相互作用が蛋白質との相互作用によって制御される事を見出した [4]。

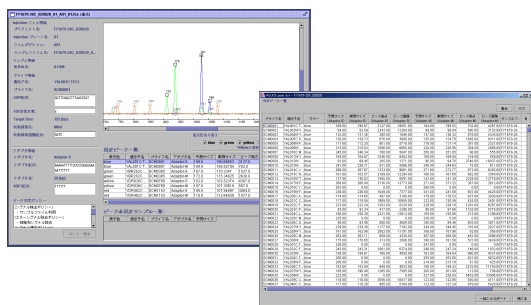
蛋白質-核酸相互作用の解析に関しては、転写因子とプロモータの相互作用およびその結果としての遺伝子発現を絶対定量化することを目指して、高精度の計測が可能な手法として、ゲノム DNA を標準とするアダプタ付加競合 PCR 法 (GATC-PCR) を開発した。



続いてプライマーの3'側部分配列の熱力学的な会合率計算に基づいて、プライマーの実効的な特異性を判定する独自のアルゴリズムを考案して、GATC-PCR に必要な特異性の高いプライマー設計を実現し、出芽酵母の5883遺伝子を計測するプライマーの設計と合成を行った。

更に自動分注装置、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 GATC-PCR 計測系の構築を進めるのと平行して、電気泳動データを管理して自動的にバンドを同定して定量実験を支援するシステム AQUOS の開発を行った。





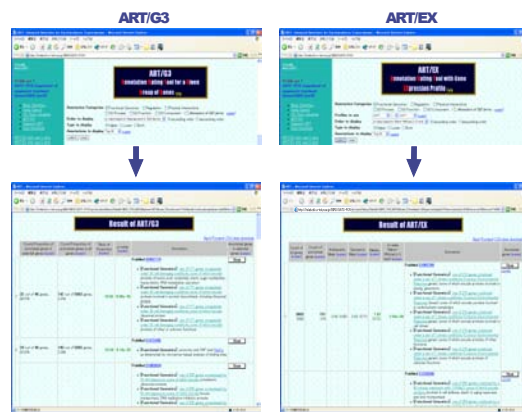
これにより出芽酵母のトランスクリプトームの絶対定量が初めて可能になった。その結果、トランスクリプトームのサイズや発現頻度からみた構造に関して従来の説とは異なる知見が得られつつある。このような計測は地味ではあるが細胞システム理解の基盤データとして貴重なものとなる。

計測系の構築に伴って、発現遺伝子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に着手した。具体的には、発現量を指標に選択された遺伝子集団に共通する機能や特徴を、集団中の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、研究者の理解を支援する手法の研究開発である[5]。既存知識としては GO(Gene Ontology)のみならず、様々な条件下で計測されたマイクロアレイの結果等の機能ゲノム科学実験も用いている。特に後者の利用によって、GO による表層的理解にとどまらず、新知見発見の契機が提示されることが確認できたので、これを実装したツールとして ART/G3を開発した。

更に、従来は遺伝子を絞り込むためにのみ使われてきた発現量の数値を、今度は絞り込んだ遺伝子群に生物学的な解釈を施す際の「重み」として活用できるように上記手法の拡張を行った。これにより、絞り込みに入り込みがちな研究者の先入観や経験則が排除されて客観性が高まると同時に、絞り込みに伴って生じる情報のロスも最小限に抑えることが出来ると期待される。

実際にこれらの手法を、我々が独自に開発した転写因子の活性化戦略に基づく遺伝子間相互作用の解析に応用した。その結果、従来のツールでは見出されず、研究者による長時間の検討の末、ようやく見出された薬剤耐性制御転写因子の予想外の浸透圧制御機能の発見が容易に達成され、この手法の有用性が示された[6]。そこでこの機能を実装したツール ART/EX を開発した。

ART/G3や ART/EXに加えて様々な解釈支援ツールを実装した統合データ解析システム EAST の構築を行い、Web 上から利用可能な形で公開した[7]。同システムは、特定遺伝子上流配列に特異的な転写因子結合パターンを探索する機能や GATC-PCR 法の絶対定量性を活かして、比率と差分の双方の視点からデータを解析するインターフェースなども備えており、チーム内の生物知識発見ワークベンチとして活用されている。また EAST が持つ ART/G3や ART/EX の機能は、遺伝子に計測値が付与されるタイプの機能ゲノムクスデータ全般の解釈に有用なツールであるので、外部データの解析にも利用可能な形で提供している。



### 3. まとめ

インタラクトーム解析を広義に捉えてその更なる発展を促す為に、様々な実験手法や計測技術の開発を BI 技術と連携のもとに進め、それぞれユニークな基盤技術の開発が達成された。とりわけ遺伝子断片の絶対定量法に解析支援的 BI 技術を組み合わせて構築した GATC-PCR システムは、遺伝子発現、転写因子-DNA 相互作用をはじめ様々な応用が可能なものであり、これまでに例のない種類のデータの創出を可能にする。特にその定量性は知識発見に有用なものであると思われ、それを活用する BI 技術の開発にも成果を挙げた。

全遺伝子のリストを網羅的計測値によって構造化するのが機能ゲノム科学の特徴であり、その計測値を構造化のみならず解釈にも活用する方向性は、今後更に追求、発展させるべきものであると考えている。

### 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) 蛋白質間および蛋白質核酸相互作用解析

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) 生物情報からの知識発見ツール開発

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) 生物情報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] Ito, T. et al. *Mol. Cell. Proteomics* 1, 561-566, 2002
- [2] Oyama, T. et al. *Bioinformatics* 18, 705-714, 2002
- [3] Minakuchi, Y. et al. METMBS'03, 22-28, 2003
- [4] Ago, T. et al. *Proc. Natl. Acad. Sci. USA* 100, 4474-4479, 2003.
- [5] Oyama, T. et al. *Genome Informatics* 14, 312-313, 2003
- [6] Onda, M. et al. *Gene* 332, 51-59, 2004.
- [7] <http://itolab.cb.k.u-tokyo.ac.jp/BIRD/GATC-PCR>