

2005年6月30日
独立行政法人 理化学研究所

遺伝子発現情報(タイリングアレイ)で未知の遺伝子構造を推定

- 世界初の情報処理技術でゲノム研究をリード -

◇ポイント◇

- ・ シロイヌナズナで、新たに約 5,200 箇所が発現する遺伝子構造を発見
- ・ 独自に開発したプログラム全情報を発見された遺伝子情報とあわせて公開
- ・ cDNA とは別の手法で遺伝子の構造を決定

独立行政法人理化学研究所（野依良治理事長）は、タイリングアレイ^{※1}（遺伝子発現情報）とゲノム情報を組み合わせて統計的に解析することにより、遺伝子の構造を高い精度で推定する新しい情報処理技術の世界で初めて開発しました。この技術をシロイヌナズナのデータに適用し、これまでに知られている約 2 万 7 千の遺伝子領域のほかにも、約 5 千 2 百箇所の領域で発現している新たな遺伝子構造を見出しました。ゲノム科学総合研究センター（榎佳之センター長）、ゲノム変異機能情報研究チームの豊田哲郎チームリーダーらによる研究成果です。

解析したこれらの遺伝子構造は同センターが運営する統合データベース GPS（Genome-Phenome Superhighway）にて公開し、遺伝子のネットワークや変異体の有無に関する情報もあわせて提供しました。また、この処理を行うために独自に開発したプログラム ARTADE（アートエード）の全ソースコード（プログラム内容）も公開しています。

研究チームはこの技術をもとに、理研植物科学研究センターの篠崎一雄センター長らと協力してタイリングアレイを使った新規遺伝子の探索と機能解析を推進しており、乾燥や低温などのストレスに耐性がある作物や有用物質生の産性を高める作物の創出にもつながるものと期待されます。

本研究成果は、英国の科学雑誌『The Plant Journal』8月号に掲載されます。

1. 背景

これまでに様々な生物種において、ゲノムの全塩基配列が解読されてきましたが、その中から遺伝子として発現するのはごく一部であり、発現する遺伝子の構造を新たに発見して機能をつきとめれば特許を取得することもできるため、国際的な競争になっています。

これまで遺伝子構造を確定する手段として、完全長 cDNA の構造を決定するプロジェクトによる解析が進められてきました。さらに、それ以外にも遺伝子の発現領域が多数存在していることがタイリングアレイを用いた遺伝子発現解析から示唆されていました。

タイリングアレイとは、ゲノムのあらゆる部分で遺伝子発現の有無を検出できるように、検出用プローブがゲノム上で等間隔に対応づけられているもので、あたかもタイルを敷き詰める様子に似ていることからそう呼ばれています。（図 1）

しかし、これまでタイリングアレイのデータから遺伝子構造を推定する作業は、人が目視でチェックしながら個々に判定していたため解析に時間と労力がかかり、その結果得られる遺伝子構造はスプライス部位（DNA内の遺伝子情報とそれ以外の配列=イントロンとエクソンの境界）が曖昧であるため、その遺伝子がコードするタンパク質を推定することもできませんでした。信頼性の高い結果を得るには、曖昧な判断基準に基づくのではなく、統計的に正しい理論に基づいて確率や有意性をきちんと評価する方法が求められていました。

2. 研究手法と成果

情報処理によって生物学的なデータから新しい発見や知識を得る方法論を研究する分野はバイオインフォマティクスと呼ばれています。生物学の分野では技術革新によって次々と新しい種類のデータが生み出されるため、それに対応するように新しいバイオインフォマティクス手法が研究開発されています。これまでに、タイリングアレイデータを処理することで遺伝子構造を高い精度で推定する技術はありませんでしたが、本研究により、高い精度で全ゲノムにわたり網羅的に遺伝子構造を推定するバイオインフォマティクスが開発されることになりました。タイリングアレイデータから塩基配列として遺伝子構造を推定して見せたのは本研究が最初です。（図2）

遺伝子構造の推定の精度を向上させる工夫点として、従来の方法はタイリングアレイデータのみを頼っていましたが、今回はこれに加えてゲノム塩基配列情報も一緒に考慮させるようにしました。これにより、スプライス部位を正確に予測できるようになりました。理論的には、観測情報からベイズ推定により遺伝子構造の事後確率分布関数を求め、その確率比（オッズ）が最大になるように遺伝子構造を推定します。確率モデルには、最尤推定法（さいゆうすいていほう）^{*2}とマルコフモデル^{*3}を組み合わせています。統計的な情報理論に基づいて確率や有意性を算出しているため、得られた結果を評価する際の信頼性が増しました。また、今回の研究により、「遺伝子発現の有意性が増すほど、そのデータから推定される遺伝子構造の信頼性も増す」という法則性が見つかり、今後のバイオインフォマティクス研究に大きく貢献する知見も得られました。

今回開発された技術により全ゲノムにわたり高精度な遺伝子構造を数時間で自動的に決定できるようになりました。この処理を行うプログラムについても全ソースコードを公開したことで、全世界の研究者が改良を加えつつ様々な生物種のデータに適用できるようになりました。このプログラムを使えば、タイリングアレイデータから全自動で網羅的に遺伝子構造が塩基配列として出力されるため、それを配列データベースへ検索すれば、新規な遺伝子を直ぐに見つけ出すことができます。この技術をシロイヌナズナのデータ適応したところ、これまで知られている約2万7千の遺伝子領域のほかにも、約5千2百箇所の領域で発現している新たな遺伝子構造を見出しました。こうして予測した遺伝子構造は、データベースとして公開しており自由に閲覧できます（<http://omicspace.riken.jp/ARTADE>）。

3. 今後の展開

これまでもタイリングアレイデータを使った研究論文が発表されてきましたが、そこから遺伝子構造を推定する技術がこれまで存在しなかったために、十分な解析が行われてきませんでした。今回、世界に先駆けてブレイクスルーとなる情報処理技術を開発し、網羅的に新規遺伝子の存在を塩基配列の構造として推定してみせたことは、同センター、ゲノム科学総合研究センターのバイオインフォマティクスのレベルの高さを示すものとして誇ることができます。

遺伝子の構造情報が塩基配列として得られるため、既知の遺伝子と比較したり、タンパク質構造を推定したり、実験的な確認をしたりするのが容易になり、新規遺伝子の探索研究効率が飛躍的に高まると期待されます。特に配列データとして遺伝子構造が具体的に得られる点は、知的財産権を主張する上で重要なファクターを得られることを意味します。

また、タイリングアレイを処理することで従来あると考えられていた遺伝子数を上回る新規な遺伝子の存在が構造情報として推定されたことは、科学的にも大きなインパクトを与えます。今回開発された技術は、これまでわが国が強力で推進してきた cDNA プロジェクトとは別の側面から遺伝子構造情報を補完します。

(問い合わせ先)

独立行政法人理化学研究所 横浜研究所
ゲノム科学総合研究センター
ゲノム機能情報研究グループ
ゲノム変異機能情報研究チーム

チームリーダー 豊田 哲郎

Tel : 045-503-9610 / Fax : 045-503-9553

独立行政法人理化学研究所
横浜研究所 研究推進部

星野 美和子

Tel : 045-503-9117 / Fax : 045-503-9113

(報道担当)

独立行政法人理化学研究所 広報室

Tel : 048-467-9272 / Fax : 048-462-4715

Mail : koho@riken.jp

<補足説明>

※1 タイリングアレイ

タイリングアレイとは、遺伝子の発現情報で、解読済みのゲノムデータから等間隔に（タイル状に）抜き出した塩基配列を検出用プローブとして搭載した DNA チップ

プのことである。このチップに、生体内で遺伝子として転写発現している RNA（リボ核酸）を鋳型にしてつくった標識サンプルをハイブリダイズ（相補的に会合）させることで観測されるシグナル情報がタイリングアレイデータである。この際、RNA と相補的な塩基配列をもつプローブのみに強いシグナルが現れることから、鋳型にした RNA が未知のものであっても部分的な塩基配列を知ることができる。タイリングアレイではゲノムのあらゆる部分をプローブが網羅しているため、ゲノムから転写された RNA であれば、原理的にはその構造を推定できるはずだが、シグナルにはノイズが大きく、プローブ間隔も広いなどの問題点があり容易ではない。

※2 最尤推定法(さいゆうすいていほう)

最尤推定法とは、確率モデルに含まれる未知のパラメタを測定データに基づいて推定する方法のひとつで、観測データの尤度（ゆうど）が最大になるようにパラメタを決定する方法。尤度とは、確率モデルでパラメタを仮定した場合に、その観測データが得られる確率のことであり、得られた事象の尤もらしさを意味する。今回の確率モデルでは、エクソンやイントロン領域においてシグナル強度がある閾値を越える確率は未知のパラメタとして扱われ、タイリングアレイの観測データが一番もっともらしくなるように決定される。

※3 マルコフモデル

「あるポイントの状態が、直前のポイントの状態のみに依存する」という確率論的構造をとるモデルであり、塩基配列で局所的にあらわれるパターンを学習させるのによく用いられる。遺伝子構造のスプライス部位には典型的なパターンがあるため、その部位が精度よく予測できるようになった。

通常の遺伝子発現アレイ



タイリングアレイ



図 1 通常の遺伝子発現アレイとタイリングアレイの違い

↓タイリングアレイデータ(発現値の高いところは赤、低いところは緑で表示)



タイリングアレイデータはノイズが大きく、このままでは、どこがエクソンでどこがイントロンがわかりにくかった

本研究の成果として、下記のような遺伝子構造を高い精度で推定できるようになった(四角で囲まれたエクソンが9個つながって1つの遺伝子を形成)。

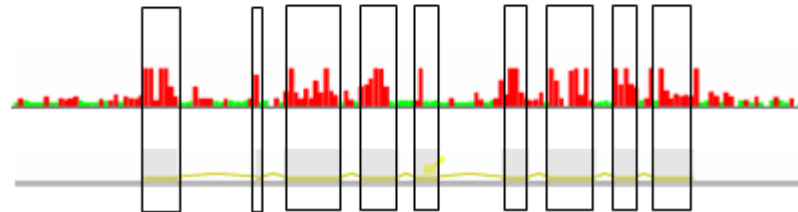


図2 タイリングアレイデータからの遺伝子構造予測における課題