論　文　の　内　容　の　要　旨

論文題目　DNA マイクロアレイのデータ解析 –数理的手法によるアプローチ–

指導教官　加我　君孝教授

東京大学大学院医学系研究科
平成 13 年 4 月入学
　医学博士課程
　外科学専攻
　西村　信一

　ゲノム解読がほぼ完了した結果、生命科学においては遺伝子発現がどのようなからくりによって調節されているのかを調べることが、次の大きな課題となっている。このような状況の中において、1990 年代後半に登場した DNA マイクロアレイは、ほぼ全ゲノムにわたる多数の遺伝子発現を同時に計測することができ、これからの遺伝子発現の研究に欠かせない道具として今後ますますその使用が増加するものと考えられる。DNA マイクロアレイをデータ処理の観点から考えてみると、その実験から得られるデータが大量であることがまず特徴として挙げられる。この大量のデータの中からいかに有用な情報を引き出してくるか、という目的に対してこれまでに多種のデータ処理法が提案されている。DNA マイクロアレイ自体が比較的新しい手法であること、また実験から得られるデータがこれまでの手法に比べて桁違いに大きいこと、から DNA マイクロアレイのデータ処理はいわゆるバイオインフォマティックス（生物情報学）の一分野として大きな比重を占めるにいたっている。

　本論文においての研究は大きく 2 部に分けて行った。

【マイクロアレイシグナル解析手法の比較および QRT-PCR 法との比較】

第一部は、DNA マイクロアレイからデータを得る最も基礎的な部分、シグナル処理法（蛍光強度から、遺伝子の発現量を計算する手法）についての比較検討、および定量的 RT-PCR 法(QRT-PCR 法)という遺伝子発現量を計測するためのほかの手法と DNA マイクロアレイの比較検討、についてである。今後マイクロアレイの実験を行うにあたって、どのような統計・データ解析法を用いるにしてもシグナル解析法は最も基本的な部分として必ず介在する手法であり、この手法の優劣を検討し、各手法の特徴を明らかにすることは意義深いものと考えられる。さらに、DNA マイクロアレイから得られるデータの信頼性を、QRT-PCR 法と比較することにより、QRT-PCR 法による追加実験での検証が必要なのかどうか、について検討をおこなった。

実験データとして、50 人分の死後脳から抽出した mRNA 発現量を、DNA マイクロアレイ（HGU-95A、Affymetrix 社）および QRT-PCR 法により計測したものを用いた。

本論文で提案した数理的モデルにより、一般的に用いられることの多い 3 種のマイクロアレイシグナル処理法（MAS5.0, RMA, MBEI）について、シグナル・ノイズ比を比較検討した結果、RMA および MBEI が MAS5.0 に比べて統計的に有意に優れていることが示された。さらに、ハウスキーピング遺伝子の発現量解析により、QRT-PCR のノイズ量は DNA マイクロアレイよりも多いことが示された。ベイズ統計学的手法を用いて DNA マイクロアレイと QRT-PCR 法の感度･ノイズを比較した結果、感度は DNA マイクロアレイのほうが劣っていることが示された。また、シグナル処理法ごとの感度の違いも明らかとなった。現在では DNA マイクロアレイから得られたデータの信頼性を確認するために QRT-PCR 法と比較することが一般的に推奨されているが、DNA マイクロアレイの感度は遺伝子発現量に依存し、RMA および MBEI をシグナル処理法として用いた場合には、QRT-PCR のシグナル・ノイズ比は DNA マイクロアレイのシグナル・ノイズ比にも劣る場合がありうることが示された。

結論として、3 種の DNA マイクロアレイシグナル処理法の特性、および QRT-PCR 法と比較した場合の DNA マイクロアレイのデータの特性が明らかとなった。

【DNA マイクロアレイからの遺伝子相互作用の推定】

第２部は、DNA マイクロアレイから得られた遺伝子発現データからの遺伝子相互作用の推定に関する新手法の提案である。遺伝子全体の相互作用をネットワークの形で導き出すことは生命科学のひとつの目標となっており、これを DNA マイクロアレイにより得られた遺伝子発現データから導き出そうという試みが広く行われている。現在では 2 次の相互作用を用いた手法が広く用いられているが、これらの手法は遺伝子発現パターンの類似性を発見することができるものの、さらに高次元の関係、すなわちほかの遺伝子の発現に影響を与えるような関係を見出す力はない。3 次以上の高次元の相互作用を検出すための手法もこれまでにいくつか提案されているが、しかし、ネットワークの規模が増大するにつれて計算量が爆発的に増大する、また推定しようとするネットワークの規模が大きくなるにつれてますます大量のデータが必要になる、といった物理的な制限もあり、現時点で決定的な手法はいまだ現れていない。

本論文においては、数理手法として対数線形モデルを用い、これまでに提案されている手法よりも比較的少数の遺伝子で、さらに詳細な遺伝子間の相互作用に踏み込むことをめざした。これまでに提案されている高次元のネットワーク推定手法においては、条件付き独立の強い定義が用いられているが、われわれの手法では条件付き独立の弱い定義による遺伝子の相互作用が推定できることを示した。また、IPIG 法として、ある遺伝子の発現に影響を与える遺伝子、あるいはある 2 つの遺伝子の間の関係に影響を与える第 3 の遺伝子、を導きだす手法を提案した。

実験データとして、DNA マイクロアレイ（HGU-95Av2、Affymetrix 社）を用いて 327 人の急性リンパ性白血病患者における 12,558 種の遺伝子発現を計測したものを用いて検討を行った。このデータは Yeoh らにより報告されたもので、

インターネット上で公開されている。

まず、対数線形モデルを用いることで、他の遺伝子の発現に影響を与える遺伝子の候補を選び出した。われわれの手法により選ばれた遺伝子は、転写因子など実際に他の遺伝子の発現に影響を与える可能性を持つ遺伝子であることが示唆された。また、転写因子 2 種とハウスキーピング遺伝子 2 種を比較した結果、本論文で注目した条件付き独立の弱い定義による関係が、転写因子のほうにより多く見られることを示した。さらに、IPIG 法による 2 つの遺伝子の間の関係に影響を与える遺伝子の探索の適用例として、XBP-1 と IGHM という 2 つの遺伝子の相互関係を次々と分解することにより、相互関係に影響を与える遺伝子の候補を示した。それらの候補遺伝子は、文献的には転写修飾因子や転写因子など、実際に遺伝子相互関係に影響を与えている可能性があることが示唆された。

結論として、本論文で提案した対数線形モデルによる遺伝子相互関係の解析により、実際の生体内での遺伝子の相互関係を、既存手法に比べてより詳細に捉えることができると考えられた。