

|          |        |          |        |
|----------|--------|----------|--------|
| 氏名       | 長嶋 剛史  | 学籍番号     | 060016 |
| 主テーマ指導教員 | 小長谷 明彦 | 副テーマ指導教員 | 杉山 公造  |
| 主指導教員    | 小長谷 明彦 | 副指導教員    | 佐藤 賢二  |

**<研究テーマ>**

遺伝子に関する知識を利用した遺伝子発現情報解析手法に関する研究

**<研究の目的と効果>**

本研究では、遺伝子に関する知識を利用して cDNA マイクロアレイから得られる遺伝子発現情報を解析する手法について提案する。

マイクロアレイによって得られる発現情報の解析は、画像処理とそれに続くクラスティングの 2 つのステップから構成される。最も重要かつ困難なのはクラスティング結果の解釈である。しかしながら、従来型のクラスティング技法により得られるのは、類似した発現パターンの集合とそこに含まれる個々の遺伝子の名称のみであり、機能未知の遺伝子に対して十分な情報が得られているとは言い難い。

そこで本研究では、クラスティング法と遺伝子に関する知識とを組み合わせることで、目的とする遺伝子に対してより詳細な機能情報を与えることを目指す。遺伝子に関する知識は、個々の遺伝子に対し BLAST や FASTA といった配列相同性検索手法や InterProScan のようなモチーフ・ドメイン検索手法を適用して得られた情報に基づいて、MEDLINE からアブストラクトをダウンロードし、ここから抽出する。抽出する遺伝子の知識としては、タンパク質間相互作用や疾病、Gene Ontology 用語等を想定している。これを発現プロファイルのクラスティング結果と組み合わせることによって、配列検索によって得られる遺伝子名や発現プロファイルのクラスティングによって得られる共発現遺伝子に関する情報のみではなく、個々の遺伝子に特化したより詳細な機能情報を与えることを目指す。また、個々の遺伝子に付随する情報間の関係からの機能推定も合わせて行う。これにより、クラスティング結果や個別データベースに対する検索からでは発見が困難であった、疾病とタンパク質間相互作用の関連や、Gene Ontology 用語と疾病の関連といった、種類の異なるデータ間での関係からの機能推定を試みる。

提案した手法に対する評価は、本システムによって提示された個々の遺伝子に対する予測結果と専門家によるマニュアルチェックとを比較することにより行なう。また、必要に応じて実際にウェットの実験を行ない、予測結果が正しいか否かを確認する。