

機関番号：10103

研究種目：若手研究 (B)

研究期間：2009 ~ 2010

課題番号：21700233

研究課題名 (和文) 心の状態にかかわる遺伝子モジュール解析システムの開発

研究課題名 (英文) Gene module database system for human mentality

研究代表者

岡田 吉史 (OKADA YOSHIFUMI)

室蘭工業大学・大学院工学研究科・助教

研究者番号：00443177

研究成果の概要 (和文)：

本研究では、心の状態にかかわる遺伝子発現データの解析支援を目的として、1) 個々の心理状態で特有の発現パターンを示す遺伝子群 (モジュール) を抽出する新しい手法を開発し、2) 抽出されたモジュールとその生物学的機能情報を集積したデータベースを構築した。これにより、複数の公共データベースに分散して存在する発現データが本システム上に整理・組織化され、生物学的に有用なモジュールを容易に発見できるようになった。

研究成果の概要 (英文)：

Aiming at assisting analysis of gene expression data for human mentality, this study proposed a novel method for identifying genetic responses (gene modules) specific for each mentality and built a database storing extracted modules and their biological knowledge. This study hereby enabled us to consolidate gene expression data dispersed in multiple public databases and to discover biological useful modules.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	1,600,000	480,000	2,080,000
2010 年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総計	2,900,000	870,000	3,770,000

研究分野：バイオインフォマティクス、感性工学

科研費の分科・細目：情報学・感性情報学・ソフトウェア工学

キーワード：心理状態、遺伝子発現、モジュール、データベース

## 1. 研究開始当初の背景

心理状態を反映する生体情報の抽出は、感性工学における重要なテーマとなっている。これまで、心理評価指標として、主に脳波、心拍、サーモデータなどの生理データや、唾液中に含まれる免疫物質 (ストレスマーカー) が用いられてきた。心理状態と遺伝子発

現の関係性は主に精神神経免疫学や医学の分野で研究されてきたが、感性工学分野ではほとんど手つかずの状態であった。

最近、DNA マイクロアレイ技術を用いて、人間の様々な心理状態での遺伝子応答を調べる試みが多くの研究グループによってなされている。これにより、大量の遺伝子発現

データがもたらされているが、それらは生データのまま複数の公共データベースに分散して管理されているのが現状となっている。

そのような膨大な分散データを整理・組織化し、さらに、個々の心理状態で特有の遺伝子情報を同定してデータベース化することで、感性工学と遺伝学をつなぐ新しい情報リソースになると期待される。

## 2. 研究の目的

本研究の目的は、感性工学への新しい切り口として、バイオインフォマティクス手法に基づいて「心の状態にかかわる遺伝子」を包括的に解析するための情報基盤を与えることであった。具体的には、以下の2つの課題の解決を目標に研究を行った：(1) 遺伝子の発現データから特定条件下で特異的な応答パターンを示す遺伝子モジュール（以下、単にモジュールと呼ぶ）を発見する新しいアルゴリズムを開発し、これを公的な生物データベースから収集した様々な心理状態に関する発現データに適用する。(2) 個々の心理状態ごとに、抽出モジュールとそれらの機能情報や関連文献情報を併せてデータベース化し、最終的にはキーワード検索によるモジュール解析システムとして Web 上で公開する。

## 3. 研究の方法

本研究は以下の2ステップからなる：

### (1) 新規遺伝子モジュール抽出法の開発

本研究では対象とする遺伝子の発現データは、行と列にそれぞれサンプルと遺伝子（あるいはその逆）、要素に発現値を並べた行列データである。ここでサンプルとは、例えば特定の心理状態にある1人の被験者から採取された細胞を指す。通常、発現データは数十サンプルに関する数万遺伝子の発現値からなり、そのような大規模データからモジュールを見つけ出すにはサンプル群と遺伝子群の膨大な組み合わせをチェックする必要がある。加えて、欠損値や異常値などのノイズも含めてモジュール抽出を行うとなると、探索空間はさらに爆発的に増大する。本研究では、飽和集合マイニング法と疑似クリーク列挙法を組み合わせた新しい方法を考案し、ノイズを許容したモジュールを現実的な時間内で探索することを可能にした。本法は、まず飽和集合マイニングにより列方向で発現値が完全に一致するモジュール（コアモジュールと呼ぶ）を網羅的に探索し、続いて、疑似クリーク列挙法により、それらのコアモジュールをノイズを許容しながら拡張していくというアプローチをとる。評価実験として、ベンチマークデータ（4種の癌細胞に関する発現データ）に本法を適用し、抽出されたモジュールの機能解析を行った。

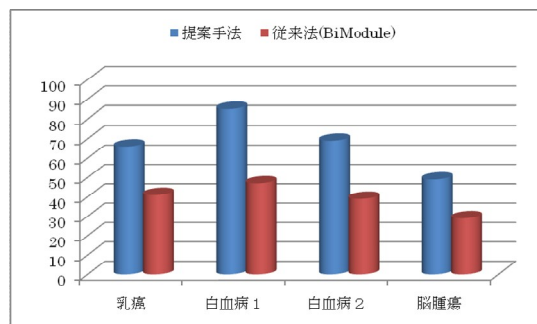


図1. 機能解析ツールで統計的に有意な機能を持つと判定されたモジュールの割合

### (2) モジュールデータベースの構築

公的データベース GEO と ArrayExpress から、心的ストレス（孤独感、笑いなど）や精神病を持つ被験者から採取された遺伝子発現データセットを収集した。現時点（2011年6月）で、様々な心理状態に関する36個のデータセットを取得している。個々のデータセットには複数のサンプルが含まれるが、それらは分断されて提供されるため、行列形式データに変換した。それぞれの行列データに対し、本研究で開発した方法を用いてモジュールを抽出し、データベース化を行った。さらに、検索結果として得られる心理状態のモジュールをヒートマップ形式で表示し、外部の知識ベース、文献データベース、配列解析ツールへのリンクを動的に生成するキーワード検索インターフェースを開発した。

## 4. 研究成果

### (1) 新規遺伝子モジュール抽出法の性能

図1は、抽出されたモジュールのうち、遺伝子機能解析ツールによって生物学的に意味があると判定されたもの ( $p < 0.001$ ) の割合を示している。この図では比較対象として、研究代表者が以前開発した方法 (BiModule と呼ぶ) のスコアも示している。以前、研究代表者らは、BiModule が既存の代表的なモジュール抽出法と比較して、遺伝子機能解析で最も良好なスコアを提示することを示した。しかし、図1からわかるように提案手法は BiModule の結果をはるかに上回るスコアを示している。この理由として以下のことが考えられる。BiModule は発現データ行列の列（遺伝子）方向の発現値が完全に一致したモジュールのみを抽出する。したがって、1つでもノイズを含む遺伝子は、たとえ他の発現値が全て一致していたとしてもモジュールの遺伝子とはみなされない。つまり、これは重要な遺伝子を取りこぼしてしまう可能性があることを示している。一方、提案手法で

はそのような問題は起こらないため、機能が

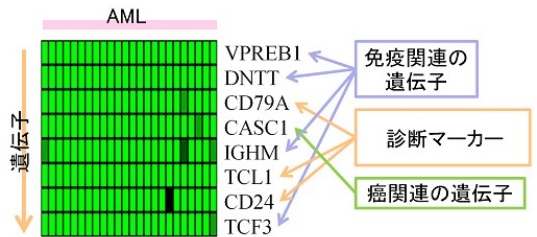


図 2. 白血病データから得られたモジュール。行と列はそれぞれ遺伝子、サンプル、色は発現レベルを表す。

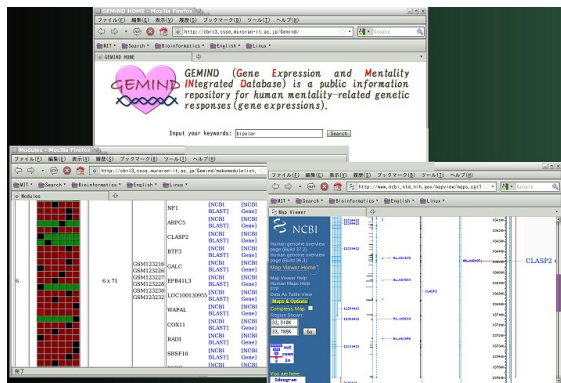


図 3. モジュール解析システム GEMIND のスクリーンショット

共通の遺伝子群が適切にモジュールにとりこまれるのではないかと考えている。

図 2 に、白血病データから抽出されたモジュールの 1 例を示す（説明のためにサンプルと遺伝子を入れ替えてある）。ノイズを許容しなければ CD79A, CASCI, IGHM, CD24, TCF3 のような白血病に関係していると思われる重要な遺伝子はこのモジュールから除外されてしまう。ここでは、全ての遺伝子が白血病と関わりをもつ場合の例を示したが、ほとんどのモジュールは機能が既知の遺伝子と未知の遺伝子が入り混じっている。そのようなモジュールからは、既知の遺伝子機能を手がかりとして未知の遺伝子機能を推定することが可能になるであろう。

以上のように、提案手法は有用なモジュールを発見する方法として有効であることが確認された。

## (2) モジュール解析システムの公開

図 3 は、本研究で開発したモジュール解析システム (GEMIND と呼ぶ) のスクリーンショットである。GEMIND のトップページからキーワード検索を行うと、該当する心理状態のデータセットの説明とモジュールリストへのリンクが表示される。モジュールリストのページでは、各モジュールの全ての遺伝子につ

いて、瞬時に生物学的機能や関連文献を調べたり、配列解析を行えるよう外部ページへのリンクが張られている。このように本システムでは、心理状態にかかわるモジュールの解析に必要な様々な情報が統合されている。これにより、ユーザは必要な情報を得るために複数の生物データベースを右往左往する手間を省くことができる。

## (3) 国内外における研究の位置づけ

「心的ストレス - 遺伝情報系 - 病気」の関連性の理解は、精神神経免疫学や医学に多大な貢献をもたらすと考えられる。現在、心的ストレスがもたらす疾患や症候群に関する発現データが徐々に蓄積されつつある。それらをモジュール解析することで、マーカー遺伝子の候補を得られる可能性があり、心理状態や精神疾患を識別する DNA チップの開発、あるいは創薬研究のための有効な情報になりうる。

このように、遺伝子レベルで見た心の状態を対象とした本研究は、今後の感性工学の発展へ向けた新しい扉を開くものであり、さらに、新産業の創出という点でも有益な開発研究となるものと期待される。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 5 件)

- ① Kentaro Fukuta, Tomomasa Nagashima, Takashi Uozumi and Yoshifumi Okada, A Novel Gene Analysis Method for Biomarker Mining in DNA Microarray Data, International Journal of Computer Information Systems and Industrial Management Applications, 査読有, vol. 3, pp. 256-263, 2011.
- ② Yoshifumi Okada, Takahiro Tada, Kentaro Fukuta and Tomomasa Nagashima, Audio Classification Based on Closed Itemset Mining Algorithm, International Journal of Computer Information Systems and Industrial Management Applications, 査読有, vol. 3, pp. 159-164, 2011.
- ③ Yoshifumi Okada and Masahiro Sawai, "GEMIND: gene expression module database for human mentality", The journal of three dimensional images, 査読有, vol. 23, no. 3, pp. 30-35, 2009.
- ④ Yoshifumi Okada and Terufumi Inoue, "Identification of differentially expressed gene modules between two-class DNA microarray data",

Bioinformation, 査読有, vol.4, no. 4, pp.134-137, 2009

- ⑤ Wataru Fujibuchi, Hyeryung Kim, Yoshifumi Okada, Takeaki Taniguchi, and Hideko Sone, "High-Performance Gene Expression Module Analysis Tool and Its Application to Chemical Toxicity Data", Methods in Molecular Biology, 査読有, vol. 577, pp. 55-65, 2009
- [学会発表] (計 17 件)
- ① Yasuo Kudo and Yoshifumi Okada, Gene Expression Data Analysis Using Heuristic Attribute Reduction in Rough Set Theory, SCIS & ISIS 2010, 2010/12/12, 岡山.
- ② 井上皓文, 岡田吉史, 遺伝子モジュールを用いた病気のクラス識別, 日本感性工学会第3回生命ソフトウェア部会研究会, 2010/11/27, 北海道.
- ③ 大橋克哉, 多田貴博, 井上皓文, 岡田吉史, 飽和集合マイニングに基づく遺伝子発現データ解析, 日本感性工学会第3回生命ソフトウェア部会研究会, 2010/11/27, 北海道.
- ④ 多田貴博, 長島知正, 岡田吉史, 飽和集合列挙法を用いた多クラス識別器の開発, 日本感性工学会第3回生命ソフトウェア部会研究会, 2010/11/27, 北海道.
- ⑤ Yoshifumi Okada, Takahiro Tada, Kentarou Fukuta and Tomomasa Nagashima, Audio classification based on a closed itemset mining algorithm, CISIM 2010, 2010/10/8, ポーランド.
- ⑥ Kentaro Fukuta, Tomomasa Nagashima, Takashi Uozumi and Yoshifumi Okada, Gene analysis Framework for a Biomarker Mining, CISIM 2010, 2010/10/8, ポーランド.
- ⑦ Kentaro Fukuta, Takashi Uozumi and Yoshifumi Okada, A novel biomarker mining method for identifying cancer genetic networks in gene expression data, ICHS 2010, 2010/9/17, 京都
- ⑧ Kentaro Fukuta, Yoshifumi Okada and Tomomasa Nagashima, LEAF: leave-one-out forward selection method for cancer classification using gene expression data, 9th IEEE/ACIS, 2010/8/18, 山形
- ⑨ Takahiro Tada, Nagashima, T. and Okada, Y., Rule-based Classification for Audio Data Based on Closed Itemset Mining, IMECS 2010, 2010/3/17, 香港.
- ⑩ 吉田恵梨香, 多田貴博, 長島知正, 岡田吉史, 飽和集合マイニングに基づくクラ

ス識別法 — 遺伝子発現データへの適用 —, 第 42 回計測自動制御学会北海道支部学術講演会, 2010/2/22, 札幌.

- ⑪ 大道一史, 福多賢太郎, 岡田吉史, 長島知正, F 値を用いた遺伝子発現差解析手法とその評価, 第 42 回計測自動制御学会北海道支部学術講演会, 2010/2/22, 札幌.
- ⑫ Okada, Y., Inoue, T. and Nagashima, T., Discriminative Gene Module Analysis for Two-Class DNA Microarray Data and Its Application to Cancer Datasets, Int. Conf. Humanized Systems, 2009/10/30, 韓国.
- ⑬ Nagashima, T., Okada, Y. and Sawai, M., Application of F-Value to Classification Problems with Numerous Numbers of Explanatory Variables, 2009 International Symposium on Nonlinear Theory and its Applications, 2009/10/18, 札幌.
- ⑭ Inoue, T., Nagashima, T. and Okada, Y., Detection of Differentially Expressed Gene Modules in Human Cancer, 2009 International Symposium on Nonlinear Theory and its Applications, 2009/10/18, 札幌.
- ⑮ 藤渕航, 岡田吉史, Horton Paul, CellMontage/SAMURAI2: Enhanced Gene Expression Similarity Search Tools, CBRC2009, 2009/10/4, 東京.
- ⑯ Okada, Y., Inoue, T. and Nagashima, T., Extraction of Discriminative Gene Modules between Two-Class DNA Microarray Data and Its Application to Chronic Loneliness Patients, Proc. of 2009 International Conference on Biometrics and Kansei Engineering, 2009/6/25, ポーランド.
- ⑰ Nagashima, T., Wang, X., Okada, Y. and Sawai, M., Applicability of F-value to Classification Problems with a Numerous Number of Explanatory Variables: Toward Classification Problems in Biometric and Kansei Engineering, Proc. of 2009 International Conference on Biometrics and Kansei Engineering, 2009/6/25, ポーランド.

[その他]

ホームページ等

<http://cbrl3.csse.muroran-it.ac.jp/Geminid>

6. 研究組織

(1) 研究代表者

岡田 吉史 (OKADA YOSHIFUMI)  
室蘭工業大学・大学院工学研究科・助教  
研究者番号：00443177

(2) 研究分担者

なし

(3) 連携研究者

なし