
基礎系教育講演

マイクロアレイデータ解析法 — DAVID による機能解析を中心に —

堀口 大吾

キーワード：マイクロアレイ，クラスター分析，DAVID，GO 解析

Microarray Data Analysis —Mainly on a Functional Analysis using DAVID—

Taigo HORIGUCHI

Abstract : Microarray technique is one of methods for analyzing gene expression. In modern life science of post-genome era, extraction of biologically meaningful information from microarray data is one of the most important problems. Cluster analysis is a traditional way of analyzing microarray data and can classify genes showing similar expression pattern into one group. However, this method cannot clarify the biological properties of each gene group, and the meaning of gene expression changes of the sample, even though common regulatory mechanism of gene expression is implied. Recently, a new analyzing method, which combines microarray data with Gene Ontology information, has developed. This method classifies genes into groups according to accumulated functional information and selects statistically important groups. Using this method, we can obtain clues to find out biologically important gene groups. In this review, theory and the practical procedure of microarray data analysis using DAVID will be mainly discussed.

I. マイクロアレイ

それぞれの生命現象の場面における遺伝子発現の特徴を捉え、その制御のメカニズムを理解することは、分子生物学の最も中心的な課題の一つである。遺伝子の発現レベルを個別に検討する方法としては、例えばノーザン・ブロット法があり、1970年代の終わり頃より用いられてきた¹⁾。この方法は、細胞や組織より調製した RNA をメンブレンに固定し、標識したオリゴ DNA プローブを用いて特定の遺伝子の発現を検出する。基本的に一回の実験に一種類のプローブしか使えないため、一種類の遺伝子発現しか調べることができない。これに対して、一回の実験で、複数の遺伝子発現を同時に検出で

きるようにしたのがマイクロアレイである。その検出原理は、相補的な核酸の特異的なハイブリダイゼーションに基づく点でノーザン・ブロットと同じであるが、既知の塩基配列のオリゴ DNA をスライドガラス上の特定の位置に多数貼り付ける技術の開発によって可能になった。

1990年頃、Fodor らが固相法による核酸の逐次合成法を発表し²⁾、さらに1995年には、スタンフォード大学の Brown らがインクジェットプリンタの技術を応用して、ガラススライド上に DNA を貼り付ける技術を開発した³⁾。Brown らはその論文で、実際に遺伝子の発現を検出している。当初、高々 48 遺伝子の検出でスタート

したこの技術は、その後、搭載できる遺伝子数を飛躍的に増大させ、現在では、その生物種が持つほぼ全ての遺伝子を網羅的に検出できるまでになっている。

この方法により、例えば、無処理の細胞と薬剤処理をした細胞の遺伝子発現の違い、あるいは健常者と患者の同じ組織における遺伝子発現の違いなどを調べることができる。患者で特徴的な遺伝子発現の違いが見つければ、疾患のメカニズムを理解する手がかりが得られる可能性がある。何らかの処理をした細胞での遺伝子発現の違いを検出できれば、その処理の効果について検討することができる。

マイクロアレイ実験では、まず組織や細胞から RNA を抽出する。その後、RNA の品質チェックを行い、これを鋳型に cDNA、あるいは cRNA を合成する。この際に核酸に標識をする。その後、標識された核酸を DNA チップ上でハイブリダイズさせ、専用スキャナでのシグナルの検出により、データを取得する。最後に得られたデータの解析、データマイニングである。巨大な情報の中から、基本的にコンピュータを利用することで、有用で新しい知見を発掘することを、鉱山から鉱石を発掘することになぞらえて「データマイニング」と呼ぶ。

現在、標準的な実験の場合、ハイブリダイゼーションやシグナルの検出等は、外部機関での受託解析で行われることがほとんどである。したがって、研究者がしなければならないことは、実験全体のデザインを除けば、サンプルからの RNA の抽出とデータマイニングである。RNA の抽出は、よほどの特殊なサンプルでない限り、方法が確立しており、大きな問題はない。最も、問題なのは最後のデータマイニングのステップである。

II. データマイニング

上に述べたようにマイクロアレイデータ自体は、今では比較的容易に得ることができる。問題は、そこから何をどうやって読み取るのかである。

マイクロアレイデータそのものは、遺伝子の発現レベル情報である。そこから興味ある遺伝子の発現レベルが、対照群と実験群とで、どのように変動しているのかを読み取れば、研究者にとって有用な情報が得られるかもしれない。しかし、その場合、調べる遺伝子は、研究者自身が知っている遺伝子に限定されてしまいがちである。その数は、通常、限定的なため、せっかく膨大なゲノムワイドの情報が得られていても、この方法ではわずかな偏った情報しか読み取れない、あるいは始めから予想されたような結果しか得られない可能性が少なくない。また、始めから興味が限定的であるなら、マイクロアレイをしなくとも定量的 PCR 等で簡便に遺伝子発現レベルについての情報を得る方が迅速で、かつ経済的である。

マイクロアレイでは、ゲノムワイドな情報が得られるのだから、それを最大限に有効活用すべきであるが、そ

れには数万の遺伝子発現情報をただ眺めているだけでは困難であり、何らかの方法によって情報を整理する必要がある。

マイクロアレイデータの代表的な解析方法としては、対照群、実験群それぞれ複数サンプルの解析を行い、発現レベルの（平均値の）差が統計学的に有意であるかどうか検討したり（有意差検定）、対照群と実験群での遺伝子発現レベル変化（発現変動倍率）を個々の遺伝子について計算し、変化の大きさ順に並べ変えるといったことも、新しい知見を得る手がかりを与ええる。さらに、比較の対象が3つ以上ある場合には、それらの間で遺伝子発現のパターンの類似性を比較することにより、遺伝子をグループ化すること（クラスター分析）も行われてきた。近年は、個々の遺伝子についての様々な知識が蓄積し、データベースの整備が進んできたため、これら遺伝子情報をマイクロアレイデータの解析に利用する方法（GO 解析、Pathway 解析）も様々に考案されている。

本稿では、クラスター分析と GO 解析について、その原理や考え方について簡単な例を用いて紹介したい。

III. クラスター分析

クラスター分析は、遺伝子発現パターンの類似性を指標に遺伝子をグループ化し、デンドログラム（樹形図）を作成して、データの見通しを良くする方法である。クラスター分析の手法は、いくつかあるが、比較的分かり易いと思われる階層的クラスター分析について、以下に見てみることにする。

例えば、図 1 A のようなデータがあるとする。遺伝子 a の発現レベルが、サンプル 1 で 90、サンプル 2 で 190 である。これを (90, 190) と書くことにする。遺伝子 b の発現レベルについても同様に (280, 370) と書く。これを二次元ベクトルと見なして、グラフ上にプロットすると図 1 B の a , b のようになる。点 a と点 b との間の距離 D は、式 1 で計算できる。

$$D = \sqrt{(280-90)^2 + (370-190)^2} \quad (式 1)$$

遺伝子数をさらに3つ増やし (c , d , e)、それぞれの点間の距離も同様にして求めることができる。

求まった距離を元にデンドログラムを作成してみよう。最も距離が近いのは、点 b と d である。これを短い線で結ぶ。次に近いのは a と e である。これもまた、さきほどより少し長い線で結ぶことができる。 a と e の中点と c との距離が次に近いので、これを、さきほどよりさらに少し長い線で結ぶ。最後に点 b , d と a , c , e を長い線で結ぶとデンドログラムができあがる (図 1 C)。

この場合は、二次元のデータを二次元のデンドログラムに描き直したただけなので、図 1 B と図 1 C では図の形状が異なるだけであるが、サンプル数がさらに多い場合には、 n 次元のデータとなるためデンドログラムに直す

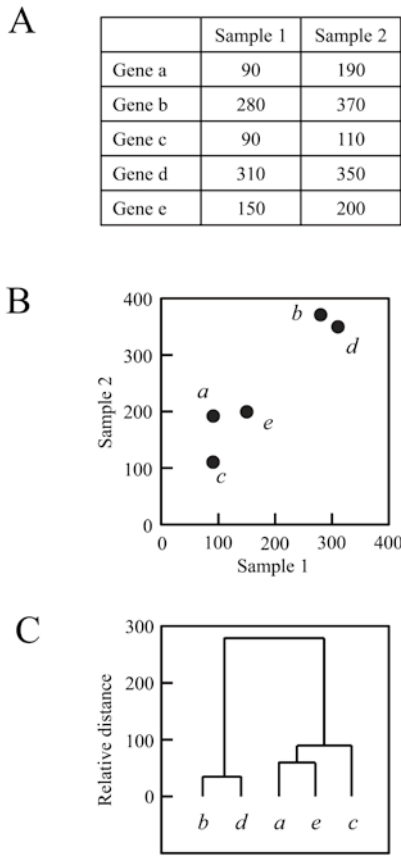


図1 遺伝子発現レベルのプロット例
 A) 遺伝子発現データ例
 B) データ A の散布図
 C) データ A に基づくデンドログラム

ことは、遺伝子発現データの構造を理解する上でそれなりに有用である。

距離の計算方法や考え方には他にも数種類あるが、いずれにしても遺伝子の発現パターンの類似性をこのような距離データに変換し、距離が近いもの同士をグループ化していく。図1のデータでは、発現レベルの高いグループと低いグループに大きく分かれている。

さて、このクラスター分析から、何が読み取れるであろうか？ 例えば、発現パターンが類似している遺伝子群（すなわちデンドログラムで一つの枝の下に所属している、比較的小きな遺伝子集団）に含まれる遺伝子の大部分が機能既知であり、中に少数の未知の遺伝子が混ざっていた場合、しかもその既知の遺伝子の機能が共通であった場合には、未知の遺伝子も機能既知の遺伝子と同様の機能を持ったものと推測できるかもしれない。また、発現パターンが類似した遺伝子同士は、共通の発現制御を受けている可能性があるため、共通の転写因子の存在が示唆されるかもしれない。

しかし、そのグループに含まれる遺伝子の個々の機能は、研究者が知っている範囲のものであれば既知だが、

それ以外は一一つ確認する必要がある。また、同じ枝の下にある一つのまとまった遺伝子グループに共通の性質を見いだせない場合、どうやって解釈すべきか苦勞することもありうる。解析しているサンプルにおける遺伝子発現変動で何に注目すべきなのか、何が重要なのか、そういったことはデンドログラムが直接答えてくれるわけではない。

IV. Gene Ontology (遺伝子オントロジー)

クラスター分析は、マイクロアレイで得られた遺伝子発現情報のみで分析する方法であったが、これまでの生命科学研究で蓄積された膨大な遺伝子情報を援用してマイクロアレイデータを解析する方法もある。その一つに Enrichment 解析があるが、この解析法について見る前に、この解析法で利用する遺伝子情報の代表的なものの一つである Gene Ontology について見てみよう。

Gene ontology (GO) は、多様な遺伝子データベース間の情報の共有を目的に1998年に設立された Gene Ontology Consortium によって管理されている遺伝子機能の表現方法であり、いくつかのキーワード (term) の組み合わせにより、個々の遺伝子機能を表現するものである⁴⁾。

Ontology は大きく3つのカテゴリー (1. Cellular Component, 2. Biological Process, 3. Molecular Function) に分けられており、それぞれのカテゴリーの中に明確に定義された多くのキーワードが含まれている。1. Cellular Component は、細胞内での局在を表す。2. Biological Process は、どのような代謝経路、シグナル経路に関連しているかを表す。3. Molecular Function は、文字通り分子機能を表す。たとえば、Hexokinase 1 (HK1) という酵素の場合、表1に示した複数の GO term が付与されている。

このように、個々の遺伝子についての既知の情報を、定義された GO term の集合に還元することで、コンピュータでの処理に馴染みやすくなっている。異なる遺伝子が、共通の GO term を持っていれば持っているほど、それら遺伝子同士は類縁のもの、あるいは関係の深いものと考えることができる。

V. DAVID

蓄積された遺伝子情報を援用するマイクロアレイデータの分析法として、DAVID (The Database for Annotation, Visualization and Integrated Discovery) の Functional Annotation Tool を用いた GO 解析を紹介する^{5,6)}。これは Enrichment 解析の一種である。

DAVID での GO 解析の概要は、以下の通りである。
 ①まず、既知の遺伝子情報を元に、遺伝子をグループに分ける。
 ②次に分析したい遺伝子リストに含まれている遺伝子が、①の遺伝子グループのどこに属するものが多いか検索する。「分析したい遺伝子リスト」とは研究者

表1 Hexokinase1 に付与されている GO term

Cellular Components	
GO ID	Qualified GO term
GO:0005739	mitochondrion
GO:0005741	mitochondrial outer membrane
GO:0005829	cytosol
GO:0005929	cilium
GO:0045121	membrane raft
GO:0097228	sperm principal piece
Biological Process	
GO ID	Qualified GO term
GO:0001678	cellular glucose homeostasis
GO:0005975	carbohydrate metabolic process
GO:0006006	glucose metabolic process
GO:0006096	glycolytic process
GO:0008645	hexose transport
GO:0015758	glucose transport
GO:0044281	small molecule metabolic process
GO:0046835	carbohydrate phosphorylation
GO:0051156	glucose 6-phosphate metabolic process
GO:0055085	transmembrane transport
GO:0061621	canonical glycolysis
GO:0072655	establishment of protein localization to mitochondrion
GO:0072656	maintenance of protein location in mitochondrion
GO:1903599	positive regulation of mitophagy
Molecular Function	
GO ID	Qualified GO term
GO:0004340	glucokinase activity
GO:0004396	hexokinase activity
GO:0005515	protein binding
GO:0005524	ATP binding
GO:0005536	glucose binding
GO:0008865	fructokinase activity
GO:0016773	phosphotransferase activity, alcohol group as acceptor
GO:0019158	mannokinase activity

がマイクロアレイの結果、着目した遺伝子のリストであり、例えば「発現が2倍以上上昇していた全遺伝子」などがそれにあたる。②の解析により、「分析したい遺伝子リスト」に、どのような生物学的プロセスのものが多く含まれているのかを知ることができ、マイクロアレイデータ解析の手がかりが得られる。

図2の青ないし赤い点は遺伝子を表しており、枠で囲まれた内側にある遺伝子は、全て同じGO termを共通に持つものとする。また、異なる枠は、異なるGO termを意味することとし、それぞれの枠をクラスターと呼ぶこ

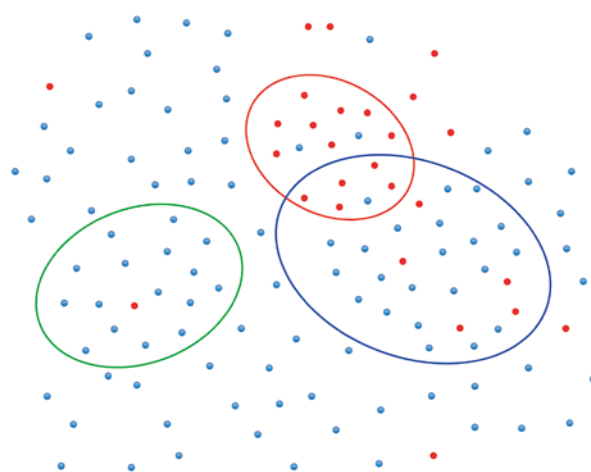


図2 GO term によるクラスターの概念図
各点は遺伝子を表し、このうち赤い点は研究者の遺伝子リストに含まれていたもの、青い点は含まれていなかったものを表す。各楕円は、それぞれ異なるGO termを表す。一つの楕円に含まれている点は、共通に同じGO termを付与されている。

とにする。

これとは別に、研究者がマイクロアレイのデータに基づき「分析したい遺伝子リスト」を作成する。このリストに含まれている遺伝子を図2では、赤い点で示した。この図で、赤枠の中に赤い点が多く、青枠の中に赤い点は少ない。しかも赤枠内の全遺伝子のうちの高い割合が、赤い点になっており、青枠内の赤い点の割合はそれよりずっと少ない。このことは、研究者の遺伝子リストには赤枠のGO termを持つ遺伝子が濃縮(Enrich)されていることを意味している。クラスターに、「分析したい遺伝子リスト」中の遺伝子が高い割合で含まれていれば、そのクラスターが意味する生物学的プロセスは、その遺伝子リストで重要度が高いと判断できる。

このように特定の遺伝子リスト中にどのような性質の遺伝子が濃縮されているかを調べる解析をEnrichment解析と呼ぶ。いくつか異なる考え方による異なる解析法が存在するが、以下ではDAVIDによるGO解析の理論的背景と実際の使用方法を説明する。

まず①の遺伝子のGO termによる分類と、それによるクラスターの生成について。共通したGO termを持つ遺伝子のグループ化には κ 統計量を用いる。 κ 統計量は一致度の指標としてよく用いられるものである。ここでは、各遺伝子に付与されているGO termの一致度を計算する。

今、GO termが15種あるとして(T1~T15)、このうち遺伝子Aに付与されているものがT1~T10、付与されていないものがT11~T15とする。図3Aではそれぞれ付与されているものを1、されていないものを0で表記してある。遺伝子Bについても同様に表記する。

A

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
Gene A	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Gene B	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0

B

		Gene B		Total
		1	0	
Gene A	1	10	0	10
	0	0	5	5
Total		10	5	15

C

		Gene B		Total
		1	0	
Gene A	1	4	3	7
	0	4	4	8
Total		8	7	15

図3 κ 統計量の計算例

- A) 2つの遺伝子とそれらに付与されている15個のGO term。GO term が付与されている場合に1，付与されていない場合を0として表示した。
- B) データAを2×2分割表で表示したもの。
- C) データAとは異なる場合の例を2×2分割表で表示したもの。

つまり、遺伝子Aと遺伝子BはそれぞれGO term T1～T10が共通に付与されており、GO term T11～T15が共通して付与されていない。この状態を2×2分割表で表すと、図3Bのようになる。遺伝子Aと遺伝子Bで共通に1であるGO term数が10、共通に0であるGO term数が5、一方が1で他方が0のGO termは無いのでいずれも0である。この時、遺伝子Aと遺伝子BとはGO termの付与に関して、その特徴が完全に一致している。

この一致の程度を数値化するのにκ統計量を用いられる。κ統計量は式2で計算できる。

$$\kappa = \frac{P_0 - P_e}{1 - P_e} = \frac{1 - 0.556}{1 - 0.556} = 1 \quad (\text{式2})$$

ここで、 P_0 は、一致している項目の合計数を全体の項目数で割った値、 P_e は表の外側の合計値の割合同士のかけ算を足し合わせたものである。

$$P_0 = \frac{10}{15} + \frac{5}{15} = 1$$

$$P_e = \frac{10}{15} \times \frac{10}{15} + \frac{5}{15} \times \frac{5}{15} = 0.556$$

これとは、異なり遺伝子Aと遺伝子Bとで、付与されているGO termが完全には一致していない場合を図3Cに示した。この時のκ統計量は式3で計算できる。

$$\kappa = \frac{P_0 - P_e}{1 - P_e} = \frac{0.533 - 0.498}{1 - 0.498} = 0.070 \quad (\text{式3})$$

$$P_0 = \frac{4}{15} + \frac{4}{15} = 0.533$$

$$P_e = \frac{7}{15} \times \frac{8}{15} + \frac{8}{15} \times \frac{7}{15} = 0.498$$

κ統計量は、遺伝子Aと遺伝子Bとが完全に一致している場合には1に、完全にバラバラの場合には0に近づく。一致したものが全く無い場合には、逆向きに一致していることになり-1に近づく。

図4には8種類の遺伝子に付与されている15種類のGO termの状態例を示した。これらについてκ統計量を用いてその一致度を判断することで、遺伝子A, B, Cを一つのグループに、遺伝子E, F, Gを別のグループに分類できる。また、遺伝子Dはこの両者と少しずつ共通していること、遺伝子Hはいずれとも異なることがわかる。いずれにせよ、各遺伝子に付与されているGO termの一致度を手がかりに、機能等が近いものを計算によってグループ化できる。

一方、マイクロアレイで得られた遺伝子リストに、ある機能グループの遺伝子が、どれだけ濃縮されているのかという問題であるが、これにはフィッシャーの正確確率検定が用いられている。

例えば、ゲノムに30,000個の遺伝子があり、そのうち40個が、ある生物学的プロセス、例えばp53シグナル経路に属するとする。一方、研究者がマイクロアレイの結果として、例えば発現上昇した遺伝子を300個検出し、このうち3個がこのp53シグナル経路に属しているとする(図5A)。この時、研究者の遺伝子リストにp53経路の遺伝子が濃縮されていると言えるか？

直感的に言えば、ゲノム遺伝子群で30,000に対して40の割合でこの経路の遺伝子が含まれているので、研究者の遺伝子群300に対して0.4個、この経路の遺伝子が含まれていれば同じ割合であり、濃縮されているとは言えない。それ以上の数の遺伝子が含まれていれば、この遺伝子群にこの経路の遺伝子が濃縮されていると考えられる。つまり、特定の生物学的プロセスの遺伝子群が

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15
Gene A	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Gene B	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Gene C	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Gene D	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0
Gene E	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Gene F	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Gene G	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Gene H	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0

図4 GO term 付与の様子の例

8つの遺伝子とそれらに付与されている15個のGO term。遺伝子ごとに付与されているGO termが異なる様子、およびGO termの付与の状態によって遺伝子をグループ化できることを表す。

A

	User genes	Genome	Total
In pathway	3	40	43
Not in pathway	297	29960	30257
Total	300	30000	30300

B

	User genes	Genome	Total
In pathway	a	b	43
Not in pathway	c	d	30257
Total	300	30000	30300

図5 フィッシャーの正確確率検定を用いたGO解析計算の概念図

A) 研究者がもつ遺伝子リストと全ゲノムの遺伝子で、ある一つの生物学的プロセスに含まれる遺伝子数と含まれない遺伝子数を2×2分割表で表示。

B) データAについて、フィッシャーの正確確率検定の計算をする際に必要な、より一般的な2×2分割表。

まとまって発現上昇しているということは、対照群と比べて実験群で、そのプロセスが着目すべき重要なものであることを示唆している。

少し統計学的に考えると、研究者の遺伝子群300個とゲノムの遺伝子群30,000個が区別できるとして、合わせて30,300個の遺伝子の中から、43個を選ぶことを考える。選んだ遺伝子のうち3個が研究者のもので、40個がゲノムのものであるような事象が生じる確率は、式4の計算で求められる。

$$P_0 = \frac{300C_3 \times 30000C_{40}}{30300C_{43}} = 0.008 \quad (\text{式4})$$

今、図5Aで図5Bのaの値が3、bの値が40であったが、この表の周辺度数を一定にし、a、b、c、dの値を様々に変化させると、同様にしてそれぞれの生起確率が計算できる。

さて、この生物学的プロセスに属する遺伝子の「割合」が研究者の「遺伝子群」とゲノムの「遺伝子群」で同じであれば、「割合」は「遺伝子群」に依らない、つまり「独立」ということになる。割合が遺伝子群で異なれば「独立ではない」ということになる。ここで「割合」と「遺伝子群」は独立であるという帰無仮説と「独立とは言えない」という対立仮説を立て、この事象が生じる時のp値を計算して、仮説の検定を行う。

図5Aの場合、p値は、それより希な事象の確率を全て加えた値で0.009になる。この場合「遺伝子群」によって「割合」が異なる、つまり「割合」と「遺伝子群」は独立ではないが、統計学的にもp値は0.9%となり、有意水準5%あるいは1%よりも低い。したがって「割合と遺伝子群は独立である」という帰無仮説は棄却される。つまり、研究者の遺伝子群で、この経路の遺伝子が濃縮されていることを意味する。つまり、この経路の機能が研究者サンプルで重要な意味をもつということになる。

別の場合として、研究者の遺伝子群にこの生物学的プロセスの遺伝子が含まれていない場合を考えてみると、この場合の生起確率は0.652。同様の計算をすると、p値は限りなく1に近づく。つまり、この場合、「割合」が研究者の遺伝子リストとゲノム全体でほぼ同じであり、「割合」と「遺伝子群」は独立である。p値も有意水準より、はるかに大きく「独立である」という帰無仮説は棄却できない。つまり、この経路の遺伝子は濃縮されていないということになる。このようにして、マイクロアレイの遺伝子リストにどの機能グループの遺伝子が

濃縮しているかが計算される。

DAVID はウェブ上に公開されており (<https://david.ncifcrf.gov/home.jsp>), 誰でも自由に無料で利用することができる。解析の実際であるが、まず遺伝子リストを用意する。DAVID は様々な遺伝子 ID を認識することができるので、Entrez Gene ID でも、Gene symbol でも適当な ID のリストでよい。その上で、DAVID ホームページにアクセスし、“Start Analysis” をクリックすると、遺伝子リストを貼り付ける画面が現れる。遺伝子リストを貼り付け、適切な条件を指定して“Submit List” をクリックすれば、遺伝子リストのアップロードは完了である。画面が“Analysis Wizard” に替わるので、ここで“Functional Annotation Tool” を選択する。すると“Annotation Summary Results” に切り替わる。ここで、例えば“Functional Annotation Clustering” を選択すれば、Enrichment score の高い順にクラスターのリストが現れる。それぞれのクラスターには、Enrichment Score が付されているが、これが濃縮の度合い、すなわちその生物学的プロセスの重要さの指標とされる。ちなみに DAVID での各クラスターには多くの場合、複数の関連する GO term や Gene Ontology 以外のデータも利用されており、これらのキーワードなども含まれていて、それぞれについて p 値が計算される。この Enrichment score は、そのクラスターに含まれる GO term それぞれについて計算された p 値の相乗平均の対数を取り -1 を掛けたものである。 p 値が小さいほど、重要度が高いので、逆に Enrichment score が大きいほど、その分子機能、局在、生物学的プロセスは注目すべきものであると言える。Functional Annotation Clustering のアウトプットは、この Enrichment score の大きい順になっている。

この分析では、研究者は、適当な遺伝子リストを用意すればよく、研究者が、そのリストに含まれている遺伝子それぞれの性質について何ら予備知識を持っていなくても、これを DAVID で解析することで、どのような生物学的プロセスが重要なのか、その具体的な内容を知ることができる。ただし、得られた生物学的プロセスが実験群にとってどのような意味があるのか解釈するのは、もちろん研究者自身である。

図 6 に示したのは、アデニル酸キナーゼ 2 (AK2) をノックアウトしたショウジョウバエの幼虫における遺伝子発現を調べた結果である。AK2 は、ミトコンドリアの内膜と外膜の膜間にあって、ミトコンドリア・マトリックスと細胞質との間でのアデニル酸のやりとりに関与している。この AK2 が欠失すると幼虫は発生の過程で死ぬが、遺伝子発現をマイクロアレイで調べ、DAVID による GO 解析を行ってみたところプロテアソームのサブユニットが一様に発現低下している様子が、高い Enrichment score と共に検出された⁷⁾。この結果は、ミトコンドリア機能に障害があるとプロテアソーム機能に影響が出ることを示唆しており興味深い。

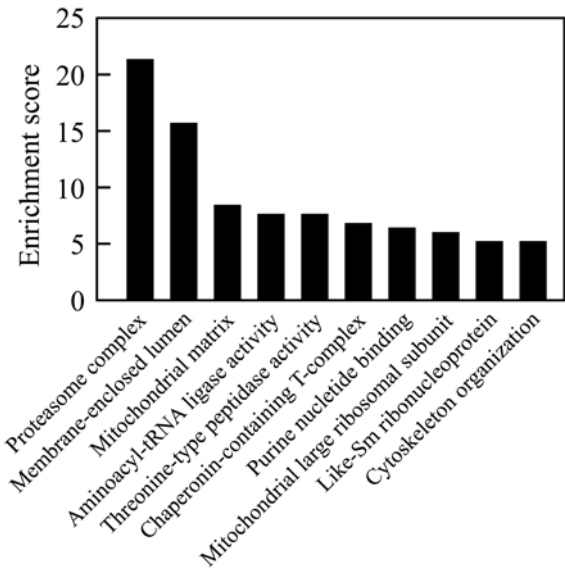


図 6 DAVID を用いた GO 解析の例
アデニル酸キナーゼ 2 をノックアウトしたショウジョウバエの幼虫での遺伝子発現を野生型のそれと比較し、発現が $1/2$ 以下に低下していた遺伝子群について GO 解析を行った。Enrichment score が高い順にその生物学的プロセスのキーワードを示した。

VI. 終わりに

マイクロアレイの開発後、実験技術、コンピュータ技術の発達によって、生命現象を網羅的に解析しようとする傾向はより強まっている。現在では、解析対象は RNA にとどまらずタンパク質や代謝物に広がり、オミックス解析と総称されている。それらは一度に膨大な情報が得られるため、とても有用に思えるのだが、逆に情報が多すぎて、解釈をするために果たしてどこから手をつければいいのか途方に暮れる場合も少なくない。マイクロアレイは、それらの中では歴史がある方なので、実験技術やデータの解釈法について様々な検討がされており、比較的信頼性の高い方法になっていると思われる。この総説が、これからマイクロアレイ解析を始めようとする方にとって多少なりとも参考になれば幸いである。

謝 辞

本稿の準備にあたっては、徳島大学大学院医歯薬学研究部分子医化学分野の野間隆文教授を始めとする教室員の方々に、ご指導、ご協力いただきました。深謝いたします。

参考文献

- 1) Alwine JC, Kemp DJ and Stark GR: Method for detection of specific RNAs in agarose gels by transfer

- to diazobenzoyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* 74, 5350-5354 (1977)
- 2) Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT and Solas D: Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767-773 (1991)
 - 3) Schena M, Shalon D, Davis RW and Brown PO: Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science* 270, 467-470 (1995)
 - 4) Gene Ontology Consortium: The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32, D258-D261 (2004)
 - 5) Huang DW, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44-57 (2009)
 - 6) Huang DW, Sherman BT and Lempicki RA: Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37, 1-13 (2009)
 - 7) Horiguchi T, Fuka M, Fujisawa K, Tanimura A, Miyoshi K, Murakami R and Noma T: Adenylate kinase 2 deficiency limits survival and regulates various genes during larval stages of *Drosophila melanogaster*. *J Med Invest* 61, 137-150 (2014)