

本書の活用の仕方

藤 淵 航

本書は多くの著名な、またマイクロアレイ解析の第一線で活躍中の先生に執筆いただいた。そのため、内容が多岐にわたるだけでなく、初学者から中級の解析を目指す読者にも十分な内容を盛り込んでいる。本書を120%活用するためには、本書の構成と概要を理解することが最も重要な手がかりとなる。各章へと入られる前にまずぜひこの「活用の仕方」を読んで少し頭の準備をしていただくと内容の理解度が格段によくなり、目的とする解析に関係する項目を見つけやすくなるであろう。特に、各章とその項目の構成をまとめた図（図1）は、全体を俯瞰するには好都合である。また、それぞれの解析法の名前がわからずに漠然とした解析のイメージをもたれている場合には、図2のフローチャートを参考にして何を解析したいのかを明確にいただけると幸いである。

本書の構成と各章の概要

本書は4章からなり、それぞれ「遺伝子測定データの標準化」「有意差解析」「クラスタリング」「ネットワーク解析」と、どれもマイクロアレイデータ解析では欠くことのできない内容となっている。各章はさらに【基礎知識編】、【基本解析編】、【応用解析編】に分かれ、ちょうど大学の講義で習うような、概論、各論（基礎）、応用論に対応している。さらに、各所に設けられた「コラム」では、アドバンストな内容の世界にまで触れることができ、向学心旺盛な読者にも満足いく内容となっている。

本書の全体の構成を、各章のシンボルとなる解析法の結果を示したグラフィックスとともにわかりやすく示した図を図1に示す。それぞれのグラフィックスは各項目を担当した著者によって作成されたものであり、目的とする解析法のイメージをつかんでいただきたい。

「遺伝子測定データの標準化」(第1章)では、特に初学者に向けて、平均、標準偏差、はずれ値、正規分布などの統計の基本事項から、一色法、二色法でのシステムの違いに応じて使用される補正法までを実験者の立場から【基本解析編】として詳述してある。さらに【応用解析編】には、Rを使用した高度な正規化や最新の配列決定法による定量装置の紹介までを取り上げている。

「有意差解析」(第2章)では、マーカー遺伝子を見つけ出すための基本的な手法である二群検定や多群検定、相関と回帰についてExcelを用いた手法を【基本解析編】で示してある。また【応用解析編】には、ラボで出くわすことが多い問題であるにもかかわらず欧米で出版されているデータ解析の本ではこれまであまり取り上げられていなかった三群以上のデータに対する発現傾向解析や、最近話題のGSEAによる遺伝子グループ単位での解析を紹介し、またマーカーを用いた予測を行う時に必要となるSVMによる判別解析までを紹介してある。

「クラスタリング」(第3章)では、【基本解析編】にて最もよく使用される階層クラスタリングとTreeViewによるその表示法や自己組織化マップクラスタリングの基本を紹介した後に、【応用解析編】ではモジュール研究に用いられるバイクラスタリング法と最小木によるクラスタリングの表現法を詳述した。

第4章の「ネットワーク解析」には、【基本解析編】としてパスウェイデータ検索システムの紹介や既存データへのマッピング方法、PubMedデータとの照合法について記載した。【応用解析編】では実験研究者の枠にとどまらず、インフォマティクス研究が得意な方を対象としてベイズ法や相関係数に基づいたネットワークの構築法や実験⇄理論の照合法を紹介した。

【基本解析編】、【応用解析編】の構成

【基本解析編】、【応用解析編】の各項目には、おおまかに「はじめに(タイトル下の囲み内の文)」「解析法の原理や概略」「解析に必要な環境・ソフト」「データの準備」「プロトコル」「コラム」という構成になっている。特に初心者は、「はじめに」を読むことでその解析が使用される背景や実際に紹介したプロトコルが何を目的としているかを知ることができる。また、各項目で知る必要がある統計学や情報理論などの基本知識は膨大であるが、その中で最も重要な骨格部分のみを「解析法の原理や概略」に示した。さらに、実習形式で「プロトコル」を実行することで、統計や情報解析になじみのない初学者でも簡単にその解析法をマスターできるようになっている。しかし、実際に論文を書く必要がある場合には、本プロトコルだけでは満足いかないことがある

かもしれない。そのようなすでに中級レベルの読者向けに、「コラム」としてアドバンスな内容も盛り込んでいただいた。どうしても自分の研究内容にフィットした解析法がプロトコールに見つからない場合にはこのコラムにそのヒントがあるかもしれないのでそこを参照していただきたい。例えば、第1章「**遺伝子測定データの標準化**」の【応用解析編】「Rを用いた標準化」では、紙面の都合で主に遺伝子（mRNA）発現解析用アレイとしてAffymetrix社とAgilent社の正規化プロトコールのみ記した。これ以外の、例えばタイリングアレイやその他の会社のアレイの処理、また異なる手法についてはコラムに可能な範囲で記述した。

活用のポイント

特に本書を120%活用してもらうためのポイントを以下に短くまとめた。ぜひ、このポイントを頭に入れて本書を徹底的に活用していただきたい。

●ポイント1：基礎知識編のフローチャートを活用する！

マイクロアレイ解析の初心者はいくつもある解析手法についてその名前を理解するだけでも大変である。本書を構成する4つの章である「**遺伝子測定データの標準化**」「**有意差解析**」「**クラスタリング**」「**ネットワーク解析**」には、冒頭にそれぞれ【基礎知識編】が設けられている。この中には、各章の項目をフローチャート形式にまとめた図がある。これは、各章のまとめ役になっていただいた先生がご自分の経験から、いくつもある解析法の位置づけを行っていただいたもので、利用価値が高い。これを見れば、どの解析法とどの解析法が近い関係にあるかをすぐに理解でき、頭がすっきりするはずである。

●ポイント2：基本理論は「解析法の原理や概略」で習得せよ！

初学者が統計学や情報理論をすべて理解することは困難（ほぼ不可能）である。そのため、余分なことをすべて捨てて最後に残った「解析法のエッセンス」にあたる部分のみ著者の先生に記していただいた。これが、各項目にある「解析法の原理や概略」である。これは言わば、ダイエットをつくしてもうこれ以上は減らすことのできない部分であり、これさえ理解すればプロトコールでやっている意味が理解できる重要な内容となっている。

●ポイント3：とにかく手を動かすことが理解への早道！

いくら「解析法の原理や概略」に解析法のエッセンスが書かれているとしても、初学

者がこれを理解しようとしてもそううまくはいかないこともある。本書のもつプロトコール形式の強みは何と言っても、パソコンとデータさえあれば、理論がわからなくとも解析法をとにかく試してみることができる点である。実際に手を動かし、目で確認することで実際の解析の道筋や問題点がわかる可能性は多いにある。もしかすると、プロトコールのみ覚えることでも、研究室では即席の「アレイ解析学者」になれるかも(?) しない。

●ポイント4：「コラム」でよりくわしい解析法を知る！

探している解析法がプロトコールと完全に一致しない場合がある。またはもう少し踏み込んだ少々高度な解析法を知りたいこともあるだろう。そのような場合には「コラム」を参照してもらいたい。「コラム」には著者が紙面の都合などで書き落としていたり、現実に論文で用いられたりするがプロトコールには載せられなかった、より洗練された解析法を書いている。例えば、「標準化」のところで「はずれ値を探す」方法については、基本となるGrubbsの方法を中心に記した。しかし、コラムにはバイオインフォマティクスや統計の論文で使用されることの多いAIC法（有名な統計学者である赤池弘次先生の手法）を用いた手法を紹介してある。中級以上の読者はぜひこの「コラム」を参考に一味違う方法を試してもらいたい！

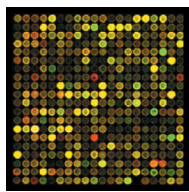
●ポイント5：「解析法発見フローチャート」を使い！

マイクロアレイの実験データが手に入り、漠然と何かやりたいのだが、それが何かわからない、などということが実験室で現実に行き起きていることかもしれない。例えば、ある生物種の発生途中でマイクロアレイデータを取ったが、どのような解析が適当なのか？ などといった問題を実際に耳にしたことがある。図2の「マイクロアレイ解析法発見フローチャート」は、そのような解析法に全くなじみのない実験研究者のために用意したものである。質問形式で各質問に答えていくと何をやればよいのか、またどのような解析法があるのかのヒントをすぐに発見できるはずである。

■ マイクロアレイ解析法発見フローチャート

上述したように、22～23ページの図2に解析法発見フローチャートを記した。まずは実験室で行ったマイクロアレイ実験がどのようなタイプかを考えて質問に答え、さらに解析したい内容に関する質問に答えていくと、本書で取り上げた解析プロトコールを即座に見つけることができる。ぜひ活用してもらいたい。

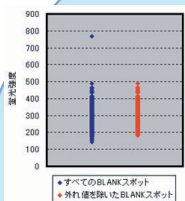
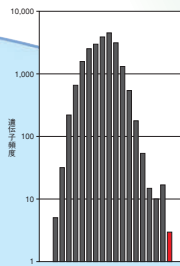
図1 マイクロアレイデータ解析の全体像と関連性



マイクロアレイデータ

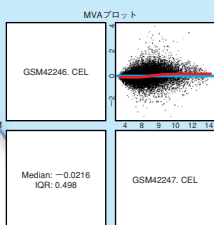
正規分布による標準化
平均値, 標準偏差, はずれ値
→ 【基本解析編-①】

一色法アレイの標準化
バックグラウンド補正
Quantile normalization
→ 【基本解析編-②】

二色法アレイの標準化
Global normalization
→ 【基本解析編-③】

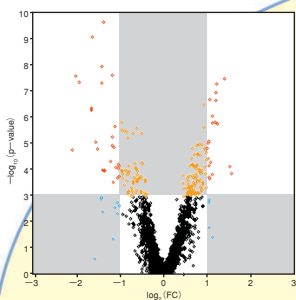
遺伝子測定データの標準化 (第1章)



Rを用いた標準化
Bioconductor
→ 【応用解析編-①】

	A1	tag	
	A	B	C
1	tag	UniGene	Symbol
2	ATAATACATA	Mito	mt-Atp6
3	TTGCTGCCCTT	Mm.332931	1200016E
4	AAGATCAAGA	Mm.686	Actc1
5	TACCGACATC	Mm.40070	8-10441

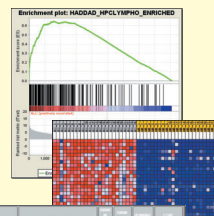
その他の大規模解析データの標準化
次世代シーケンサー
→ 【応用解析編-②】



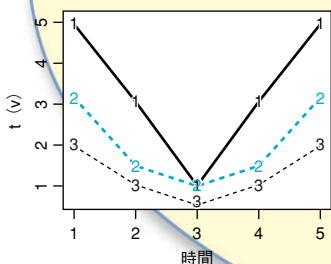
二群からの有意差検定
 t 検定, MW-U 検定
→ 【基本解析編-①】

相関係数と回帰分析
ピアソンの積率相関係数,
単回帰分析
→ 【基本解析編-③】

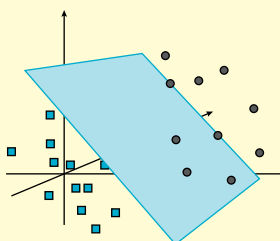
多群からの有意差検定
ANOVA-F 検定, KW 検定
→ 【基本解析編-②】



有意差解析 (第2章)



発現傾向解析
ヨングヒール検定,
累積カイ2乗検定
→ 【応用解析編-①】

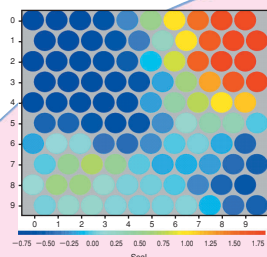


判別分析
Support Vector Machine
→ 【応用解析編-③】

Gene Set	Enrichment Score	Adjusted P-value	Q-value
GO:0003674	0.15	0.0001	0.0001
GO:0003675	0.12	0.0002	0.0002
GO:0003676	0.10	0.0003	0.0003
GO:0003677	0.08	0.0004	0.0004
GO:0003678	0.06	0.0005	0.0005

機能グループ解析
Gene Set Enrichment Analysis
→ 【応用解析編-②】

自分の目的とする解析は どこにあるのかイメージをつかもう



非階層型クラスタリング

k平均法,
自己組織化マップ
→【基本解析編-③】

階層クラスタリング

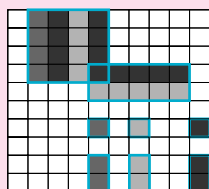
Cluster3.0
→【基本解析編-①】

Java TreeViewによる表示

TreeView1.1.1
→【基本解析編-②】



クラスタリング (第3章)

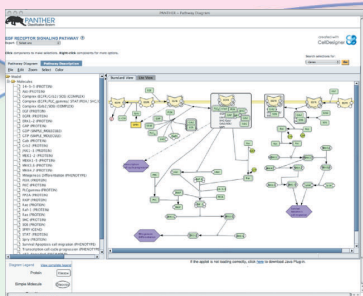
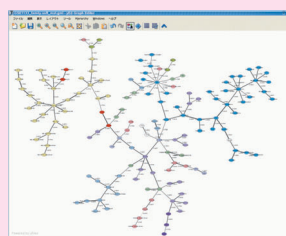


バイクラスタリング

BicAT, BiMax
→【応用解析編-①】

グラフによる類似関係の可視化

最適木, yEd
→【応用解析編-②】



パスウェイデータベースの基本操作

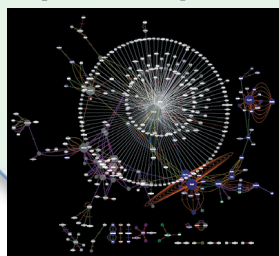
Reactome, PANTHER, cPath
→【基本解析編-①】

文献データとの照合

Info-PubMed, iHOP
→【基本解析編-③】

パスウェイデータへの マッピングと可視化

Cytoscape
→【基本解析編-②】



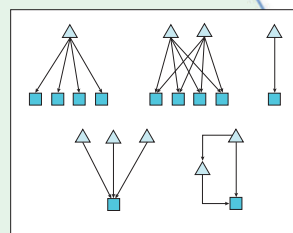
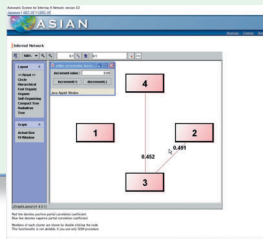
ネットワーク解析 (第4章)

S-systemモデルによるネットワーク推定

sAnalysis
→【応用解析編-②】

偏相関係数によるネットワーク推定

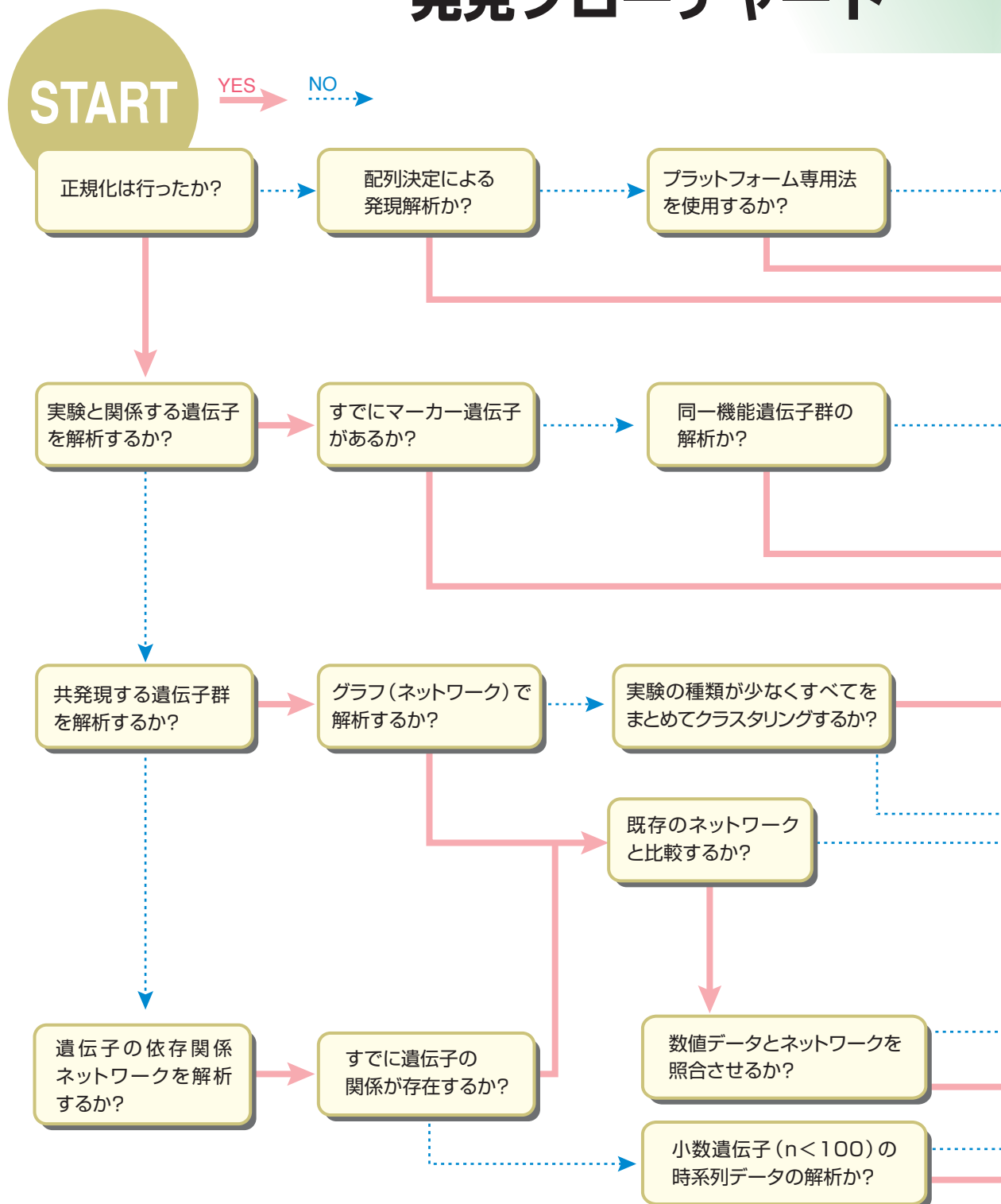
グラフィカル・ガウシアン・モデル法
→【応用解析編-①】



ネットワークと 実験値の適合度評価

ネットワークモチーフ分解法
→【応用解析編-③】

マイクロアレイ解析法 発見フローチャート



質問に答えながら 最適な解析法を見つけ出そう

遺伝子測定データの標準化 (第1章)

二色法アレイか?

- ▶ 正規分布による標準化 (本書31ページ～)
- ▶ 一色法アレイの標準化 (本書39ページ～)
- ▶ 二色法アレイの標準化 (本書47ページ～)
- ▶ Rを用いた標準化 (本書56ページ～)
- ▶ その他の大規模解析データの標準化 (本書66ページ～)

有意差解析 (第2章)

系列実験か?

2実験だけか?

上昇／下降傾向の解析か?

- ▶ 二群からの有意差検定 (本書82ページ～)
- ▶ 多群からの有意差検定 (本書90ページ～)
- ▶ 相関係数と回帰分析 (本書96ページ～)
- ▶ 発現傾向解析 (本書105ページ～)
- ▶ 機能グループ解析 (本書113ページ～)
- ▶ 判別分析 (本書124ページ～)

クラスタリング (第3章)

クラスター間の階層を
必要とするか?

- ▶ 階層クラスタリング (本書139ページ～)
- ▶ Java TreeViewによる表示 (本書149ページ～)
- ▶ 非階層型クラスタリング (本書159ページ～)
- ▶ バイクラスタリング (本書168ページ～)
- ▶ グラフによる類似関係の可視化 (本書178ページ～)

ネットワーク解析 (第4章)

文献で解析するか?

全遺伝子で解析するか?

- ▶ パスウェイデータベースの基本操作 (本書194ページ～)
- ▶ 文献データとの照合 (本書215ページ～)
- ▶ パスウェイデータへのマッピングと可視化 (本書205ページ～)
- ▶ ネットワークと実験値の適合度評価 (本書243ページ～)
- ▶ 偏相関係数によるネットワーク推定 (本書224ページ～)
- ▶ S-systemによるネットワーク推定 (本書234ページ～)