

## 22

AutoGateによるデータ解析  
のさらなる自動化

Leonore A. Herzenberg, Stephen Meehan,  
Guenther Walther, David Parks, Wayne Moore, Connor Meehan,  
Megan Phillips, Eliver Ghosn, Leonard A. Herzenberg

フローサイトメトリーの応用範囲は、開発当初のわれわれの想像をはるかに超えて、医学から海洋学まで多岐にわたり、その利用者は拡大の一途をたどっている。一方、機器の性能も向上の一途をたどり、1回の実験で得られるデータの量・複雑さは指数関数的に増加している。そのため、シンプルかつ論理的な再現性のあるデータ解析手法が必要とされている。本稿では、蛍光漏れ込み補正の自動化に続いてわれわれが開発した、統計学的手法を用いた細胞集団同定の自動化、AutoGateについて概説する。

## はじめに

はるか昔、まだTリンパ球とBリンパ球の区別がなく、造血幹細胞についてほとんどなにも知られていなかったころ、われわれはFluorescence Activated Cell Sorter (FACS) を着想し開発した。ほぼ半世紀後の現在、細胞生物学は隆盛を極め、BD Biosciences、ベックマン・コールター、ペイバイオサイエンス、ソニー、DVS Sciences など数々の会社がフローサイトメーターを販売し、世界中の基礎および臨床の現場でデータを収集し細胞をソーティングしている。フローサイトメーターはTリンパ球、Bリンパ球や造血幹細胞の研究、そしてHIV、白血病、免疫不全をはじめとする数多くの疾患の診断と治療に欠かせないものとなっている。

現在入手可能なフローサイトメーターは初期の物に比べて比較にならないほど高性能である。最初は1つのレーザーに1つの検出器だけであったが、何年にも渡って継続的に改良が重ねられ、現在では18以上の蛍光色素を同時使用できる機種まで登場している。そして質量分析を利用したCyTOFの登場により同時に観察できるパラメーターはさらに倍増しつつある。加え

て、より低機能だが低価格の手頃なフローサイトメーターも続々登場している。これによって、フローサイトメーターの選択肢が増え、ユーザー数の増加とともにその利用分野は多様性を増している。

これは、多くの科学者・臨床家が日々フローサイトメトリーのデータを収集し、それに基づいて判断を下していることを意味する。加えて、多くの科学者・臨床家が科学誌に掲載されたフローサイトメトリーのデータを解釈し判断することを必要とされていることも意味する。かつてフローサイトメトリーはマイナーな分野で、そのデータの分析と解釈は限られた専門家の仕事であったが、いまや誰もが行わなければならないタスクなのである。そのためシンプルでよりよいデータ解析のためのツールの開発が急務である。

そこで本稿では、もうすぐ利用可能になるフローサイトメトリーのデータの解析を自動化するソフトウェアを紹介する。AutoGateとよばれるそのソフトウェアは、現在の手動解析ソフトの多くの部分を踏襲しているだけでなく、新しく導入された統計学的アルゴリズムによって細胞集団を自動的に同定し、サンプル間での比較を可能にする。AutoGateは今までのように

ユーザーが主観的にゲートをかける必要をなくすといっても過言ではない。もちろん自動的に検出された細胞集団のどれを次の解析対象にするか、などはユーザーの判断にゆだねられることになる。

## フローサイトメトリーデータの収集と記録

フローサイトメーターがデータを収集する仕組みについては本稿を含むさまざまな場面で解説されている。細胞は識別可能な蛍光色素でラベルされた抗体や試薬によって染色される。染色された細胞をフローサイトメーターにかけると、フローサイトメーターは細胞による2種類の散乱光（細胞のサイズを反映するFSCと粒度を反映するSSC）と、各蛍光色素から発せられた細胞あたりの蛍光強度を測定する。今日のフローサイトメーターは最大毎秒12,000個程度の細胞を解析しデータを記録することができる。例えば1つのサンプルに含まれる500,000個の細胞ごとの8つの蛍光強度とFSC, SSCが1つのファイルに記録されることになる。

各サンプルのデータファイルは実験ごとにまとめられて各自のパーソナルコンピューター、もしくは専用のデータストレージに収納される。データファイルのサイズは巨大なので、すべての実験データを個人個人が各自のパーソナルコンピューターに収納しておくことは困難になりつつある。解析結果を含めたデータの欠落を防ぐためにも、研究室単位もしくは研究所単位でのデータ管理システムを整備することが望ましい。

スタンフォード大学の共用FACSコア・ファシリティでは、2003年にわれわれの研究室のWayne Mooreが開発したデータ自動保管システムが利用されている。データは各フローサイトメーターからデータ保管サーバに自動的に格納され、ユーザーはFlowJoなどの解析ソフトウェア上から、必要なときに必要なデータを数クリックでダウンロードできる。このような研究所単位でのデータ管理システムの需要は高まっており、

それに呼応してCytobank (Cytobank社)、CytoGenie (Woodside Logic社)などの汎用サービスが登場してきている。

われわれの研究室と共同開発中のCytoGenieは、フローサイトメーターが収集したデータのみならず、実験計画や使用した試薬の情報など、フローサイトメトリーに関するすべての情報を統合して管理するシステムである。後に紹介するAutoGateなどの解析機能も統合され、その解析結果も収納されるしくみとなっている。

## 蛍光漏れ込み補正

フローサイトメトリーのデータ解析の最初のステップは蛍光漏れ込み補正である。このステップは、もし各蛍光色素からの蛍光が他の蛍光色素用の検出器に漏れ込まなければ必要ない。しかしフローサイトメトリーに使用される蛍光色素の蛍光波長は幅広いことが多く、現実には、わずか2種類の蛍光色素しか使用しない実験でも漏れ込みが起こる場合がほとんどである。このため正確な蛍光強度の決定には蛍光漏れ込み補正が必要となる。

当初、蛍光漏れ込み補正はアナログ電気回路としてフローサイトメーターに組み込まれていた（現在でもアナログ電気回路を使用している機種もある）。しかし、現在のソフトウェアを用いたデジタル演算による蛍光漏れ込み補正は、格段に正確な結果をもたらす。その理由については2006年のNature Immunologyの総説<sup>1)</sup>で詳しく解説した。この総説は本稿で登場するその他の問題の背景を理解するためにも大変有用なので、是非一読していただきたい（その内容は本書の**基礎編-1**でもカバーされている）。

われわれの研究室はアナログ電気回路による蛍光漏れ込み補正を最初に実現しただけでなく、ソフトウェアによる蛍光漏れ込み補正理論についても開発の先頭に立ってきた。現在、その理論はさまざまなフローサイトメトリー・データ解析用ソフトウェア・パッケージ

に実装され、利用することができる。例えば、FlowJo (Tree Star社) は、コントロール・サンプルのデータが適切に収集されていれば、ほぼ自動的に蛍光漏れ込み補正を行う機能を提供している。

最新の蛍光漏れ込み補正理論の実装である AutoComp は、CytoGenie や FlowJo から必要なコントロール・サンプルの情報を収集し、完全に自動化された蛍光漏れ込み補正を実現している。

ひとたび必要な情報が得られたら、AutoComp は最新の統計学的手法を用いて蛍光波長の重複を計算し、そこから蛍光漏れ込みの補正係数が求められる。この補正係数は FlowJo などの解析ソフトウェアに渡され、各サンプルのデータに適用される。後述する AutoGate はこれらの蛍光漏れ込み補正の情報を自動的に取り込み利用する。

## 細胞集団の同定と定量化

蛍光漏れ込み補正の次に行う解析が、細胞集団を同定し、その頻度や特徴となるマーカーの発現量を定量化する作業である。このステップのためにさまざまな解析ツールが利用可能である。しかしここで留意が必要となるのが、サンプルからいくつの細胞集団が同定されたとしても、もう1つ染色するマーカーを追加することによって、さらなる細分化や全く違う細胞集団の同定が可能になるということである。そこでわれわれは原則的に、予算が許す限りの多様な試薬 (抗体) を入手して、可能な限り多くのマーカーを同時染色することを推奨している。

この原則に従うと、より多くの蛍光漏れ込み補正を必要とし、細胞集団の同定も複雑になるため、解析の難易度は高くなるが、先に紹介した蛍光漏れ込み補正の自動化と、次に紹介する細胞集団同定の自動化によってユーザーの負担は顕著に軽減される。

蛍光色素のかわりに金属同位体でラベルし質量分析の原理で検出する CyTOF は、単位時間あたりに解析できる細胞数はまだまだ限られているが、さらに多く

のパラメーターの同時測定を可能にする。細胞集団の同定を自動解析する手法は、蛍光色素を測定する従来のフローサイトメトリーのデータにも質量分析によるデータにも利用可能である。

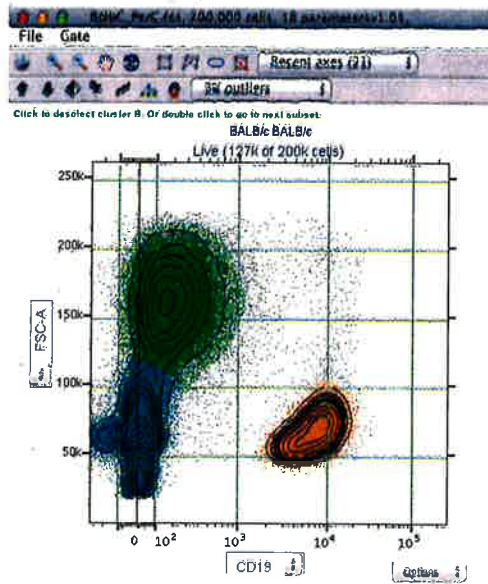
## 細胞集団同定の自動化

データ解析用のソフトウェアは、ユーザーが選択した蛍光色素の組み合わせでデータをプロットし、その分布から視覚的に判断して細胞集団の境界をユーザーが主観的に定義する (ゲートをかける) 機能を提供している。そしてゲートをかけた細胞集団について次の蛍光色素の組み合わせでデータをプロットし、さらに細分化していくことを可能にしている。この作業は3つの蛍光色素を使った実験ならば初心者でも簡単に行うことができるが、同時に使用する蛍光色素数が増えるにしたがって急激に難しくなる。同時に使用する蛍光色素数が6を超えると、もはや芸術とよべる領域になり、専門家のみが可能な作業となる。

しかし、前述のように同時に観察するパラメーター数が多ければ多いほど、より正確な細胞集団の同定が可能になる。実際に、今日の平均的なフローサイトメーターのユーザーは、研究対象とする細胞集団を同定するためには6~12、もしくはそれ以上の蛍光色素を同時に使用する必要があると認識している。

この問題を解決するためにわれわれは AutoGate を開発した。AutoGate は統計学的手法によってデータの解析をさらに自動化し、現在の解析ソフトウェアよりも簡単に、シンプルでわかりやすい結果を手にすることができる。現在の解析ソフトウェアでは、細胞集団を定義する境界 (ゲート) はユーザーが主観的に決定することを繰り返す。例えばマーカーAとマーカーBの二次元プロットからある細胞集団にゲートをかけ、その細胞集団を次にマーカーCとマーカーDの二次元プロットで観察してさらに細分化して、ということを目的細胞集団が見つかるか、選択肢がなくなるまで繰り返す。一方、AutoGate は統計学的に有意な細胞集団

A) AutoGate



B) FlowJo

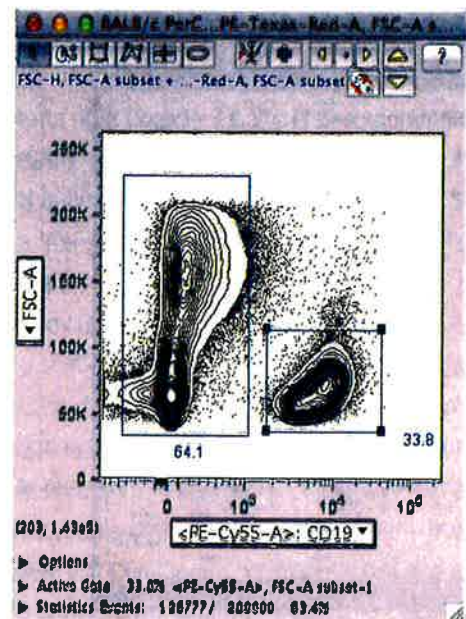
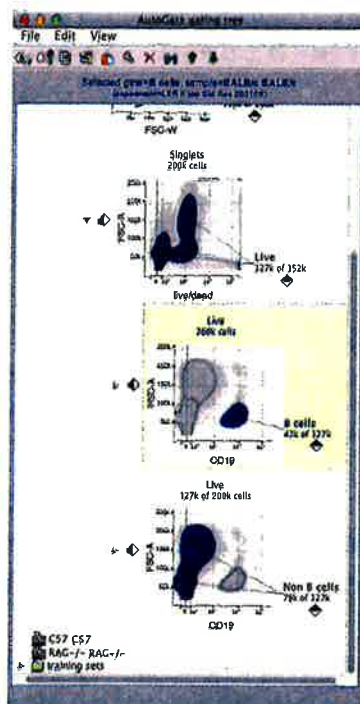


図1 統計的手法による細胞集団の同定

A) AutoGateは表示されたデータのなかから統計的に有意な細胞集団を自動的に同定する。B) 同じデータをFlowJoを用いて主観的にゲートをかかけた状態

A)



B)

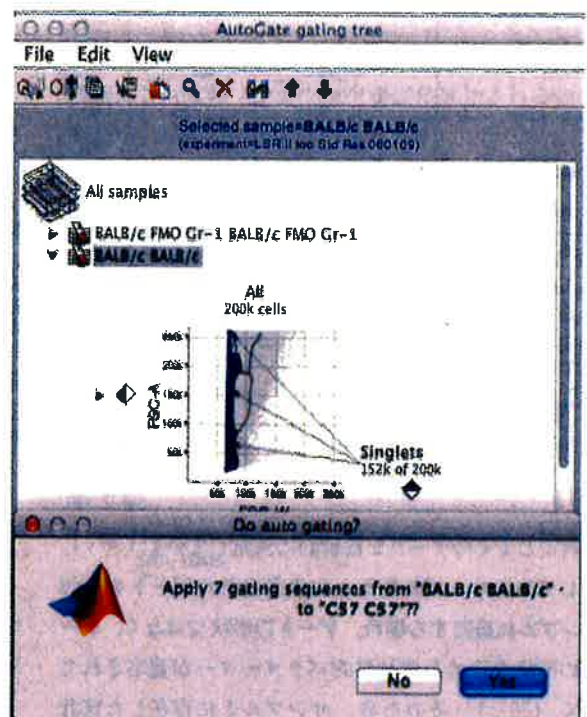


図2 AutoGateによる細胞集団同定の手順

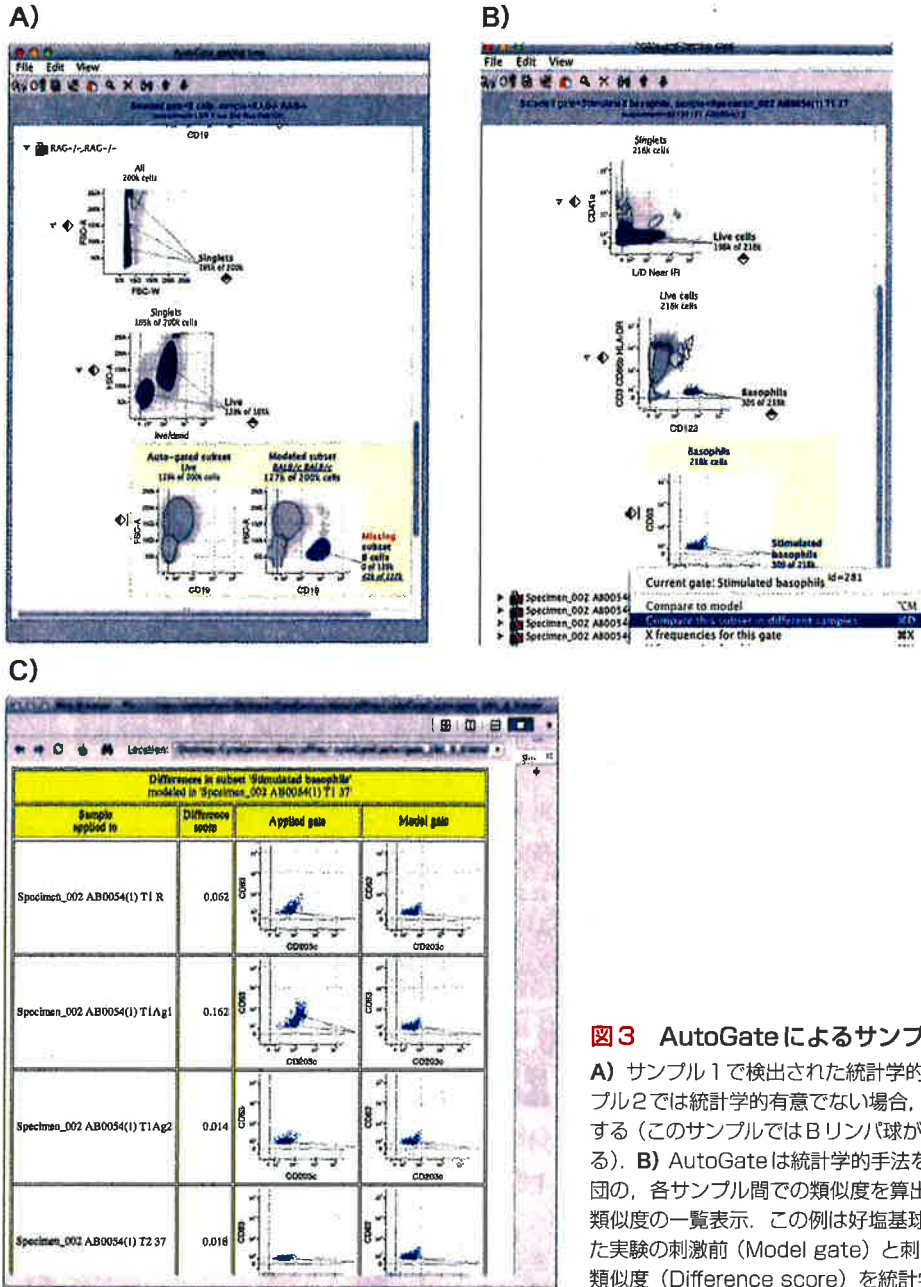


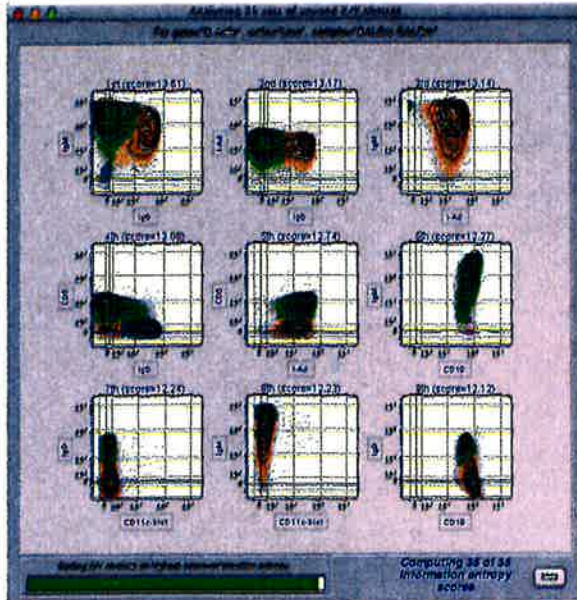
図3 AutoGateによるサンプル間の統計学的比較

A) サンプル1で検出された統計学的有意な細胞集団が、サンプル2では統計学的有意でない場合、AutoGateはそれを検知する（このサンプルではBリンパ球がMissingと判定されている）。B) AutoGateは統計学的手法を用いて、着目した細胞集団の、各サンプル間での類似度を算出できる。C) サンプル間類似度の一覧表示。この例は好塩基球にさまざまな刺激を加えた実験の刺激前（Model gate）と刺激後（Applied gate）の類似度（Difference score）を統計学的に算出している

を同定してそのゲートを自動的に決定していく（図1）。そしてあるサンプルをもとに決定されたゲートを他のサンプルに適応する場合、ゲートの形状ではなく、ゲートの形状を決めた統計学的パラメーターが適応されていく（図2）。そのため、サンプル1に存在した統計

学的有意な細胞集団が、サンプル2にも統計学的有意に存在するの否か、という情報を得ることができる（図3）。

加えて、ゲーティングを繰り返して細胞集団の細分化を進めていく際に、AutoGateは次ほどのマーカー



**図4** AutoGateによる論理的なゲーティングの展開  
AutoGateは次にどのマーカーの組み合わせに注目したらよいかの順位を提供する

の組み合わせに注目するべきか、の優先順位を計算することができる(図4)。これによってゲーティングを深めていく順番も、ユーザによる主観的なものから、より論理的なものとなる。

現在のところ、12の蛍光色素とFSC, SSCのデータの解析まで可能となっているが、CyTOFなどのさらに高次元のデータやフローサイトメトリー以外のデータも解析できるように拡張を予定している。CytoGenie, AutoComp, そしてAutoGateを使うことによって、高次元のフローサイトメトリー・データを誰もが確実に解析することが可能になるのである。

## おわりに： フローサイトメトリーの 過去と未来

この原稿の依頼を受けたとき、われわれと監修者/編者は、われわれが2006年にNature Immunologyに

発表した総論<sup>1)</sup>のアップデート版を書くことを想定していた。しかしながら、2006年版の記事を読み返してみると、その内容は今日でも変わらず通用することに気付いた。なぜLogicle(もしくはBi-exponential)表示を利用すると、対数軸表示よりも正確にデータを解析できるのかといった今日の高次元フローサイトメトリー・データの解析と解釈に必要な理論はすべて含まれている。

そこでわれわれはこの機会に、過去に発表した内容のアップデートではなく最新の成果に基づいた未来の技術について紹介することにした。Logicle表示や蛍光漏れ込み補正の自動化などの技術に、統計学的手法による細胞集団同定の自動化が統合されるのである。

CytoGenieシステムには、実験計画やデータを統合して保管するCytoGenie、蛍光漏れ込み補正を自動的に行うAutoComp、統計学的手法によって細胞集団を自動的に同定するAutoGateが含まれている。われわれはこのシステムを、専門家の手を借りずに誰もがラップトップやPCにダウンロードしてインストールし、使いはじめられるようにデザインした。政府・教育機関などの非営利団体は、このシステムを無料もしくはきわめて低コストで利用できるようにする方針である。詳細はこれらのソフトウェアが提供開始される2014年の早い時期に発表する予定である。

本稿によって、最新のソフトウェア技術により可能となる新しい可能性が読者の皆様に伝わると確信している。これらの技術によって、高次元のフローサイトメトリーデータの解析は、専門家だけの領分から、誰もが行えるものへとなるであろう。

翻訳：清田 純

### ◆文献

1) Herzenberg, L. A. et al. : Nat. Immunol., 7: 681-685, 2006