

京都大学	博士 (医科学)	氏名	岡田 大瑚
論文題目	Decomposition of a set of distributions in extended exponential family form for distinguishing multiple oligo-dimensional marker expression profiles of single-cell populations and visualizing their dynamics (分布セットの拡大指数型分布族形式への分解による、オリゴ次元マーカーを測定した複数の1細胞発現プロファイルの識別とそのダイナミクスの可視化)		
(論文内容の要旨)			
<p>サイトメトリーは個々の細胞の表面マーカーの発現量を調べることができる技術であり、代表的なシングルセル発現解析技術のうちの1つである。このようなシングルセル発現データを多次元データとして扱い情報科学的アプローチを適用する研究が盛んにおこなわれており、多くのアルゴリズムが成果をあげている。しかし、これらの手法のほとんどは単一のデータを対象に細胞集団の同定や可視化を行うものであり、細胞集団プロファイル全体の特徴を定量化するような手法は極めて限定されていた。近年、大規模に取得されたサイトメトリーデータを用いた研究が盛んに実施されているが、これらの研究でも取得されたデータは特定の細胞サブセットの定量化のために用いるにとどまっており、細胞集団プロファイル全体の特徴を対象にはしていない。そこで、複雑な生命現象の全貌を捉えるためには、シングルセル発現データから推定される細胞集団プロファイル全体の異同とその生物学的背景を統合的に評価するための新規のデータ解析手法を確立することが必要であると考えられた。一方で、情報科学分野では、確率分布の幾何学的な構造を研究対象とする情報幾何学が近年注目を集めており、シングルセル発現データを確率分布として捉えることでこれらの知見が利用できると考えられた。</p> <p>本研究では、情報幾何学の知見を用いて、多次元シングルセル発現プロファイルのデータセットを低次元空間に埋め込み、それらの異同を説明する主要な構成分布をデータドリブンに同定する新規のノンパラメトリック統計的手法を提案した。まず、従来の情報幾何学で主な研究対象とされていた指数型分布族の制約を部分的に緩めた拡大指数型分布族を新たに定義し、任意の確率分布セットを拡大指数型分布族の形式で表現するアルゴリズム (拡大指数型分布族分解法: Decomposition into Extended Exponential Family (DEEF)) を開発した。次に、DEEF が理論的な確率分布のセットに対して、適切に潜在的なパラメーター構造を抽出し、低次元座標空間に埋め込むことができることをシミュレーションデータを用いて示した。さらに、成人ヒト乳腺に対する上皮成長因子(EGF)による刺激後のダイナミクスを調べたサイトメトリーの実データセットにDEEF を適用した。その結果、2つのパラメータを用いて EGF 刺激後の細胞集団プロファイルの複雑なダイナミクスが記述され、それらを低次元座標空間上で軌跡として可視化できることが示された。また、DEEF が寄与の大きい座標に基づいて分布を再構成することができるという特徴を用いて、シングルセル発現データセットの次元削減および時系列の再構成手法を示した。</p> <p>本研究より、DEEF が細胞集団プロファイルの異同をデータドリブンに解析する上で有用な手法であると考えられた。DEEF のさらなる応用として、各細胞集団プロファイルに付与された座標軸の値を用いて、サンプルの属性やゲノム・オミックスデータとの関連解析を行い、シングルセル発現データとゲノム・オミックスデータの複合解析へと展開していくことが考えられる。</p>			

(論文審査の結果の要旨)

サイトメトリーデータは代表的なシングルセル発現データのうちの1つである。サイトメトリーデータのこれまでの一般的な利用方法は、ゲーティングや計算生物学的手法を用いたクラスタリングによってマーカー遺伝子の発現分布から細胞サブセットを同定・定量するというものであった。申請者はサイトメトリーデータセットが確率分布からの標本セットと捉えられることに着目し、この解析を可能とする確率分布同士の異同を定量する情報幾何学を用いた手法を開発した。その結果、申請者は細胞集団プロファイル全体を分布として取り扱いノンパラメトリックな方法で指数型分布族様の式表現を得るという理論を構築し、既存の手法が抱えていた結果の解釈性に関する課題を解決する新規のサイトメトリーデータの異同の定量化を実現した。本論文は、これまで実施が困難であった大量のサイトメトリーデータの解釈可能な分析を一定の時間内で実施可能とし、近年大規模化が進むサイトメトリーデータ解析研究において実用的かつ先駆的な研究であり、幅広い生命科学分野での分布状データへの展開も期待できるものである。

したがって、本論文は博士 (医科学) の学位論文として価値あるものと認める。

なお、本学位授与申請者は、令和2年10月20日実施の論文内容とそれに関連した試問を受け、合格と認められたものである。