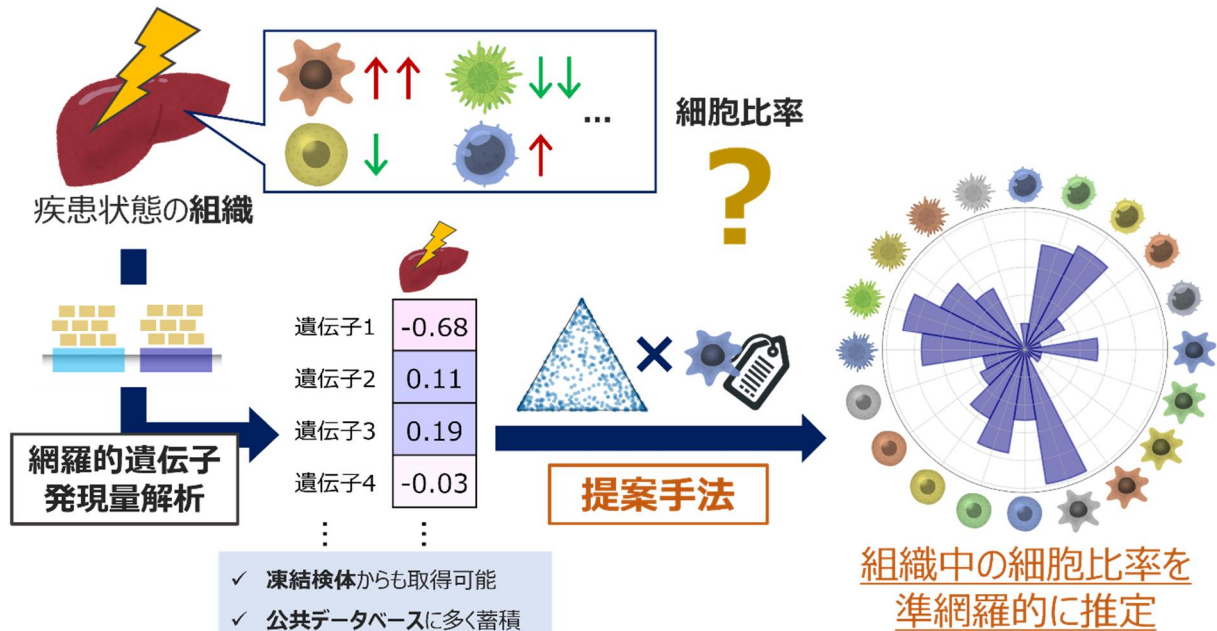


組織中の細胞種の比率を準網羅的に推定するアルゴリズムを開発 ——遺伝子発現量を言葉に見立てた潜在変数モデルの効果的な活用——

発表のポイント

- ◆ガイド付き潜在的ディリクレ配分法により、組織の網羅的遺伝子発現量データから当該組織中の細胞種の比率を準網羅的に推定するアルゴリズムを開発しました。
- ◆マーカー遺伝子が既知である細胞種について、網羅的に組織中の細胞比率の推定が可能となりました。
- ◆組織中の細胞比率に関連する基礎研究の発展や臨床現場での応用の促進に資すると期待されます。



提案手法による組織中の細胞比率推定フロー

概要

東京大学大学院薬学系研究科の東一織大学院生、水野忠快助教らによる研究グループは、組織の網羅的遺伝子発現量（トランスクリプトーム）データから当該組織中の細胞種の比率を準網羅的（注1）に推定するアルゴリズムを開発しました。

組織中の細胞比率は、免疫細胞の浸潤などにより変化するため、疾患の病型など、個体の状態を表現する重要な情報です。一方、組織中の細胞比率を実測する方法論は、多くの場合、新鮮組織を対象としており、ビッグデータの蓄積や臨床現場での取得は困難でした。本研究グループは、遺伝子を言葉として扱い、同じ言葉として与えられるマーカー遺伝子名（注2）を効果的に活用することで、凍結組織などからも容易に取得可能であり、また公共データベース（公共DB、注3）にも豊富に蓄積されているトランスクリプトームデータより、多くの細胞種の比

率を推定する方法論を開発しました（図1）。本研究成果は7月10日、国際科学誌「Briefings in Bioinformatics」に掲載されました。

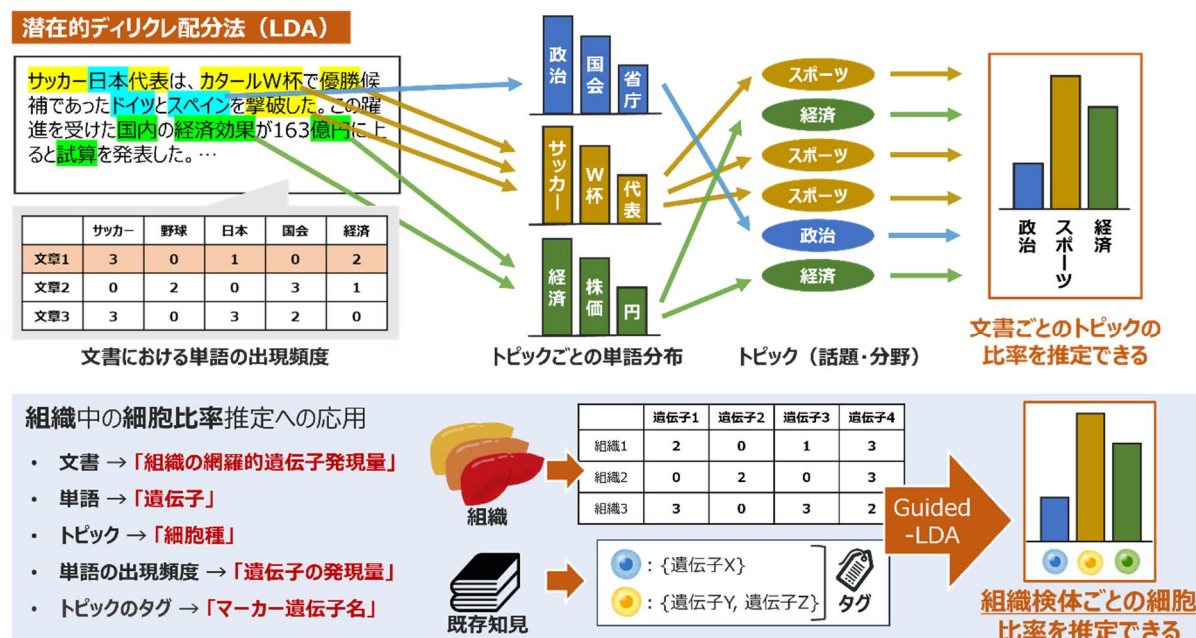


図1：アルゴリズムの概要図

発表内容

<研究の背景>

東京大学大学院薬学系研究科の水野助教らの研究グループは、データサイエンスを駆使して医薬品の認識されていない側面（注4）を理解し、活用する研究を進めています。組織中の細胞比率を推定するために、生物情報学的手法である Deconvolution 法が多く開発されています。Deconvolution 法は、比較的容易に取得でき、公共DBにも豊富に蓄積されているトランスクリプトームデータを入力とし、細胞比率を推定します。しかし、既存の多くの手法は、解析対象データ以外のトランスクリプトームデータ（リファレンスデータ）を必要とするリファレンスベース手法（注5）であり、解析対象データとリファレンスデータの相性によって精度が左右されるという課題がありました。最近では、リファレンスデータを必要としないリファレンスフリー手法も開発されていますが、リファレンスがないため、得られた比率がどの細胞種のものであるかを解釈するのが難しいという課題がありました。

<研究の内容>

本研究では、リファレンスベース手法とリファレンスフリー手法の両方の課題を同時に解決する「Guided LDA Deconvolution (GLDADec)」を開発しました（図1）。提案手法は、トランスクリプトームデータを文書に見立てるリファレンスフリー手法を基盤とし、リファレンスデータへの依存性の問題を解決します。さらに、各細胞種のマーカー遺伝子名をモデルに組み込むことで、解析結果の解釈性を向上させています。この手法は、遺伝子発現量を言葉として扱い、マーカー遺伝子名という細胞種の「タグ」を効果的に活用するモデルです。

Deconvolution 法の性能評価には、構成細胞が明確な血液検体のデータセットが頻用されます。本研究でもまず血液検体での性能評価を行ったところ、複数のデータセットにおいて、提案手法が既存手法を一貫して凌駕することが示されました（図2）。一方、組織は血液に比して

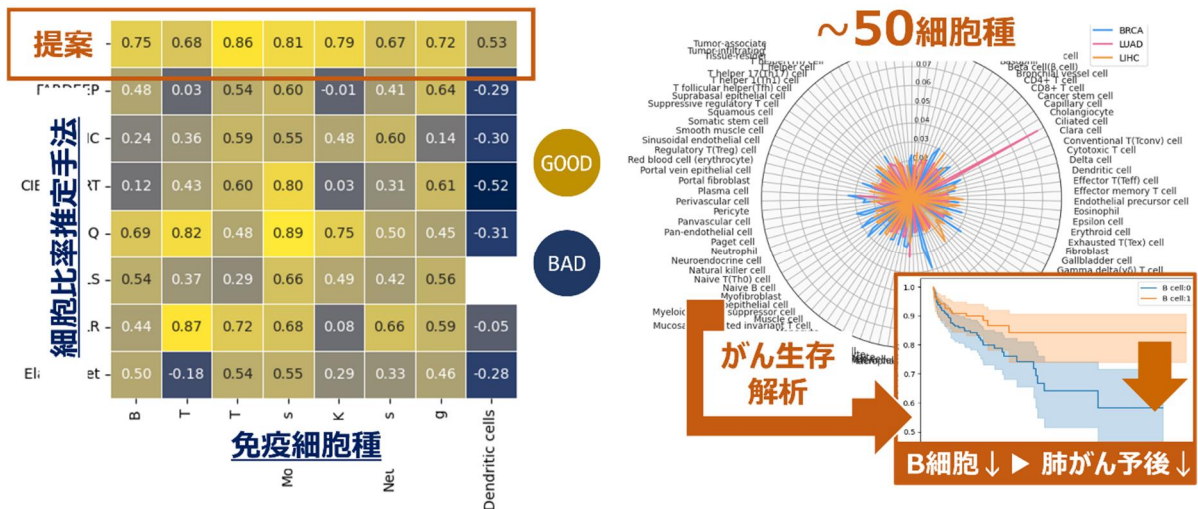


図 2 : 性能評価

多様な細胞種で構成されているため、血液検体での評価が組織でも正しいとは限りません。同研究グループは以前、マウスおよびラット肝臓のトランスクリプトームデータと免疫細胞の実測データが対となった独自の Deconvolution 法性能評価データセット（注6）を構築しています。そこで同データセットでも検証したところ、提案手法は既存手法を凌駕し、組織中の免疫細胞比率を正確に推定できることが実証されました。

本手法の特筆すべき点は、解析対象のトランスクリプトームデータ以外に必要な情報が細胞種のマーカー遺伝子名のみであることです。既存の知見として多くの細胞種のマーカー遺伝子名が蓄積されているため、網羅的な細胞種の推定が実質的に可能となります。そこで、ヒトがんのデータベース（注7）で公開されている3種類のヒト腫瘍サンプルのトランスクリプトームデータに本手法を適用しました。結果、約50種類の細胞種の比率が推定され、その細胞比率に基づく生存解析（注8）は既存の報告されている関係性を確かに検出していることが確認されました（図2）。

<今後の展望>

提案手法は二つの恩恵をもたらします。一つは公共DBの利活用の促進です。トランスクリプトームデータは公共DBに豊富に蓄積されており、人類の財産とも言えます。提案手法を公共DBへ適用することで、様々な状態の組織中の準網羅的な細胞比率データをこれらより大量に取得可能となります。単独の研究者では様々なコストの面で困難であった細胞比率に関するビッグデータが得られるため、パターン認識などのデータ解析に供することで、これまで実験科学的には得られなかった知見にアクセスできるようになると期待されます。もう一つは、凍結検体からの細胞比率の取得が容易となることです。フローサイトメトリー（注9）などの実験科学的手法では、一般に新鮮な組織を対象としており、臨床現場での取得は容易ではありません。一方、トランスクリプトームデータは凍結検体からも取得可能です。提案手法を適用することで、凍結検体からでも容易に、そして多く細胞種の比率を取得可能となります。

以上のように、提案手法は、組織中の細胞比率に関連する基礎研究の発展や臨床現場での応用の促進に資すると期待されます。

発表者・研究者等情報

東京大学 大学院薬学系研究科

東 一織 博士課程

水野 忠快 助教

楠原 洋之 教授

論文情報

雑誌名 : Briefings in Bioinformatics

題 名 : GLDADec: marker-gene guided LDA modelling for bulk gene expression deconvolution

著者名 : Iori Azuma, Tadahaya Mizuno*, Hiroyuki Kusuhara

DOI: 10.1093/bib/bbae315

URL: <https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbae315>

研究助成

本研究は、科研費「基盤研究 (C) (課題番号: 21K06663)」、「武田科学振興財団薬学系研究助成」、および「持田記念医学薬学振興財団研究助成」の支援により実施されました。

用語解説

(注1) 準網羅的

マーカー遺伝子名が判明している細胞種を推定可能となる。細胞種が同定されているということは当該細胞種を規定するマーカーの存在を示唆しているため、概ね網羅的な推定が可能となる。

(注2) マーカー遺伝子名

細胞種 X が細胞種 X であることを判断するためのタグとして用いられる遺伝子の名称。細胞種ごとに複数のマーカー遺伝子が存在することが多く、既存知見としてデータベースなどにまとめられている。

(注3) 公共データベース

GEO dataset など、公開され、誰でもトランスクリプトームデータにアクセス・ダウンロード可能なように整備されたデータベースのこと。

(注4) 医薬品の認識されていない側面

市場撤退につながる有害事象を引き起こす副作用や、ドラッグリポジショニングにつながる作用など、開発時には認識されていないもののその後明らかとなった側面のこと。

(注5) リファレンスベース手法

解析対象の組織トランスクリプトームデータとは別に、比率を知りたい細胞種だけのトランスクリプトームデータを必要とする細胞比率推定手法のこと。

(注6) 独自の Deconvolution 法性能評価データセット

採取した肝臓の一部をトランスクリプトーム解析、残りをフローサイトメトリーに供することで、トランスクリプトームデータと細胞比率の実測値とが対応したデータセットのこと。当研究グループが以前の研究にて公開している。

(注7) ヒトがんのデータベース

アメリカ国立がん研究所 (NCI) が公開している The Cancer Genome Atlas と呼ばれるデータベース。ヒトがんのトランスクリプトームデータベース等を収載している。

(注8) 細胞比率に基づく生存解析

細胞比率より特定の細胞種の存在量の高低を数値化できる。当該数値により患者を分けた際のそれぞれの群の生存分布の比較のこと。

(注9) フローサイトメトリー

細胞種ごとに抗体を用いて蛍光標識し、蛍光をベースに細胞数のカウントを可能とする実験機器。細胞数の解析のデファクトスタンダードの方法論。一般には新鮮組織を解析対象とし、一度の試験に比較的時間がかかる。

問合せ先

〈研究に関する問合せ〉

東京大学大学院薬学系研究科

助教 水野 忠快 (みずの ただはや)

Tel : 03-5841-4771 E-mail : tadahaya@mol.f.u-tokyo.ac.jp

〈報道に関する問合せ〉

東京大学大学院薬学系研究科 庶務チーム

Tel : 03-5841-4702 E-mail : shomu@mol.f.u-tokyo.ac.jp