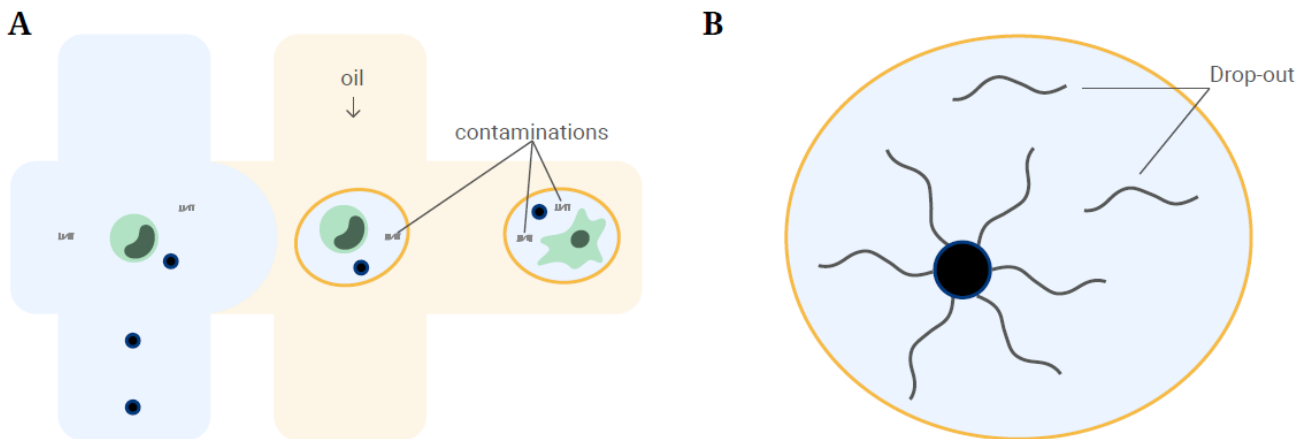


# BioTuring Smooth Binary アルゴリズム の簡単な説明

## A. Introduction

既知のマーカによるセルタイプのラベリングには、通常、細胞がマーカを発現しているか否かを判定します。細胞にゼロ以外のシグナルがある場合は、発現していると見なされ、その逆も同様です。しかし、実験によるコンタミネーションにより、バックグラウンドノイズが生じると、偽シグナルが発生するという問題が生じます。

さらに、特定のシングルセル技術では、細胞が実際にはマーカを発現しているにもかかわらず、カウントを示さないことがあり、これは「dropout」現象として知られています。この問題の多くは、細胞の真のマーカシグナルを正確に捉える技術の限界に起因します。



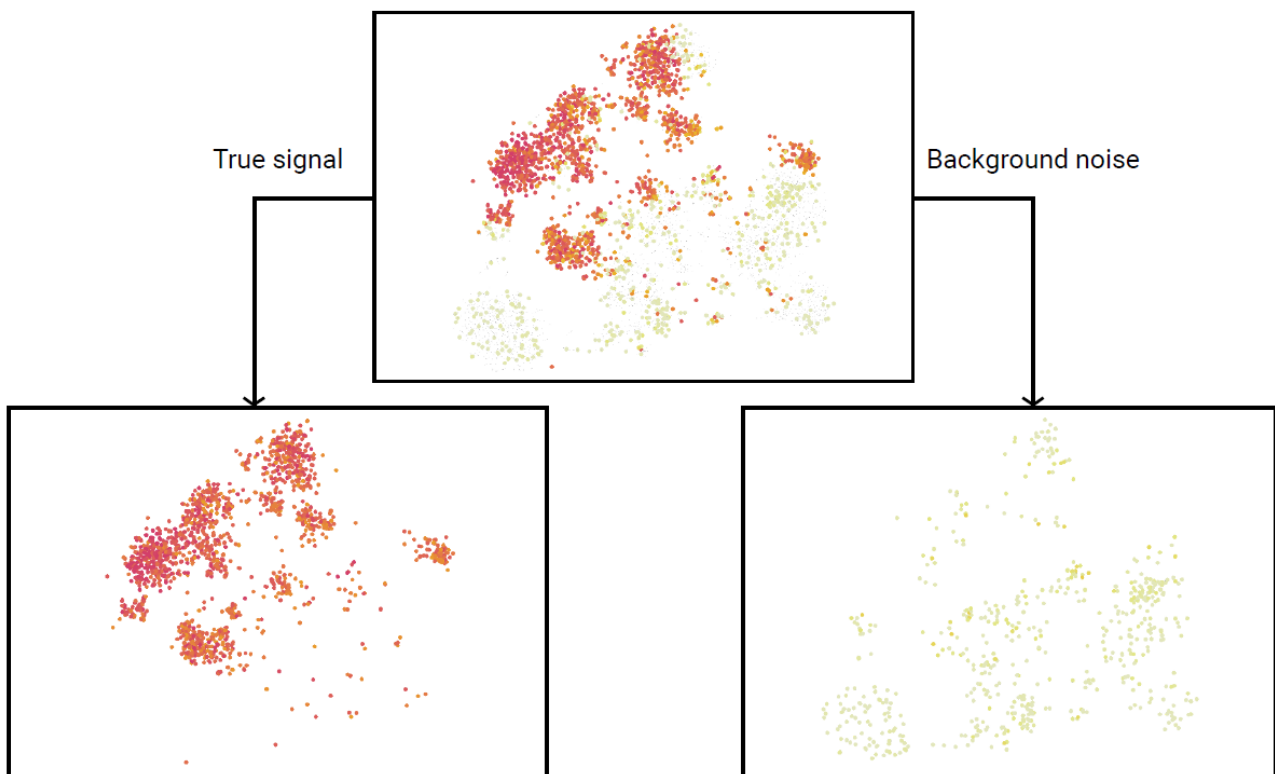
**Figure 1. シングルセル実験におけるバックグラウンドノイズとdropout**

**A.** 実験環境からコンタミネーションされた RNA は、細胞に誤ったシグナルをもたらす可能性がある。**B.** 細胞からの RNA は、技術的に捕捉できない場合がある。

BioTuring アルゴリズムは、細胞があるマーカーを発現しているかどうかを 3 つのステップで分類することで、これらの課題に対応しています。

1. **バックグラウンドシグナルの除去** : 真のマーカー発現と誤解されるようなシグナルを排除することができます。
2. **発現値を2値化する** : 発現レベルは2つのカテゴリーに単純化され、細胞が「発現している」か「発現していない」かが明確になります。
3. **dropoutシグナルを置き換える** : 隣接する細胞の発現状態を考慮することで「dropout」を補正します。

## B. バックグラウンドシグナルの除去



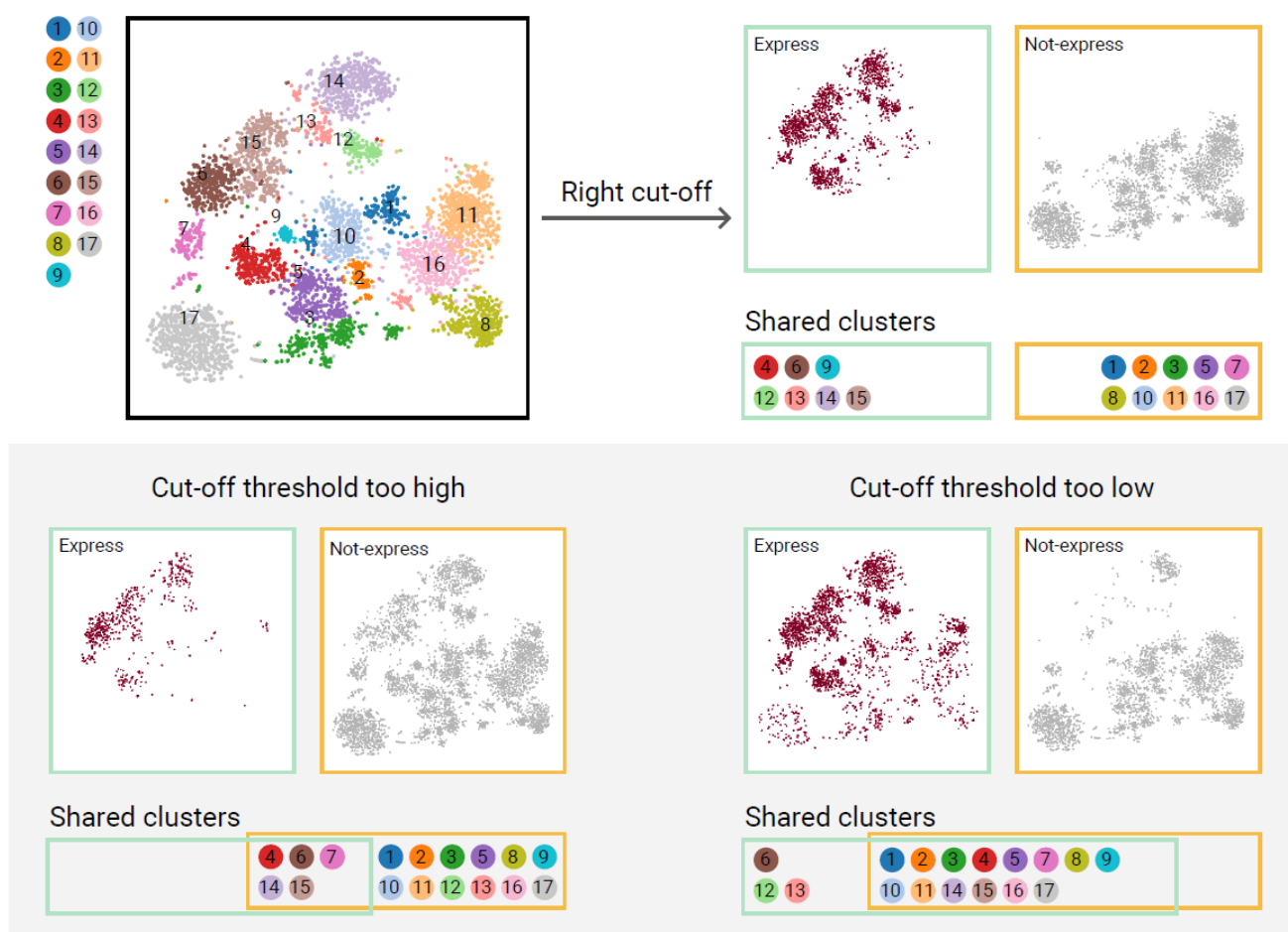
**Figure 2. 2D空間におけるバックグラウンドノイズの分布**

真のシグナルを発現している細胞は密なクラスターを形成し、高い発現値を示す。対照的に、シグナルノイズとなる細胞はランダムに散在し、低い発現値を示す特徴がある。

真のマーカーシグナルを持つ細胞はクラスターを形成しやすい原則に基づいてアルゴリズムが動作します。同様に、シグナルを持たない細胞もクラスターを形成しますが、ノイズシグナルを持つ細胞はランダムに分布します。

バックグラウンドのカットが不十分だと、発現細胞が多数のクラスターにまたがって存在することになり、発現細胞と非発現細胞のクラスターリストが大幅に重複することになります。一方、厳しすぎるカットオフしきい値は、真のシグナルを除去し、発現細胞が非発現細胞に分類されてしまい、同様の重複の問題を引き起こす可能性があります。

適切なカットオフしきい値のみが、発現細胞を真のクラスター内に保ち、非発現細胞を引き離し、両者を正確に区別することができます。そこで、アルゴリズムは発現細胞と非発現細胞のクラスターリストの重なりを最小にするカットオフしきい値を見つけようとします。



**Figure 3. バックグラウンドシグナル除去のためのカットオフしきい値の決定**

適切なカットオフしきい値により、発現細胞と非発現細胞のクラスターが明確に分離され、重複が最小限に抑えられる。逆に、しきい値の設定が低すぎたり高すぎたりすると、2つのクラスターグループ間でかなりの重複が生じる。

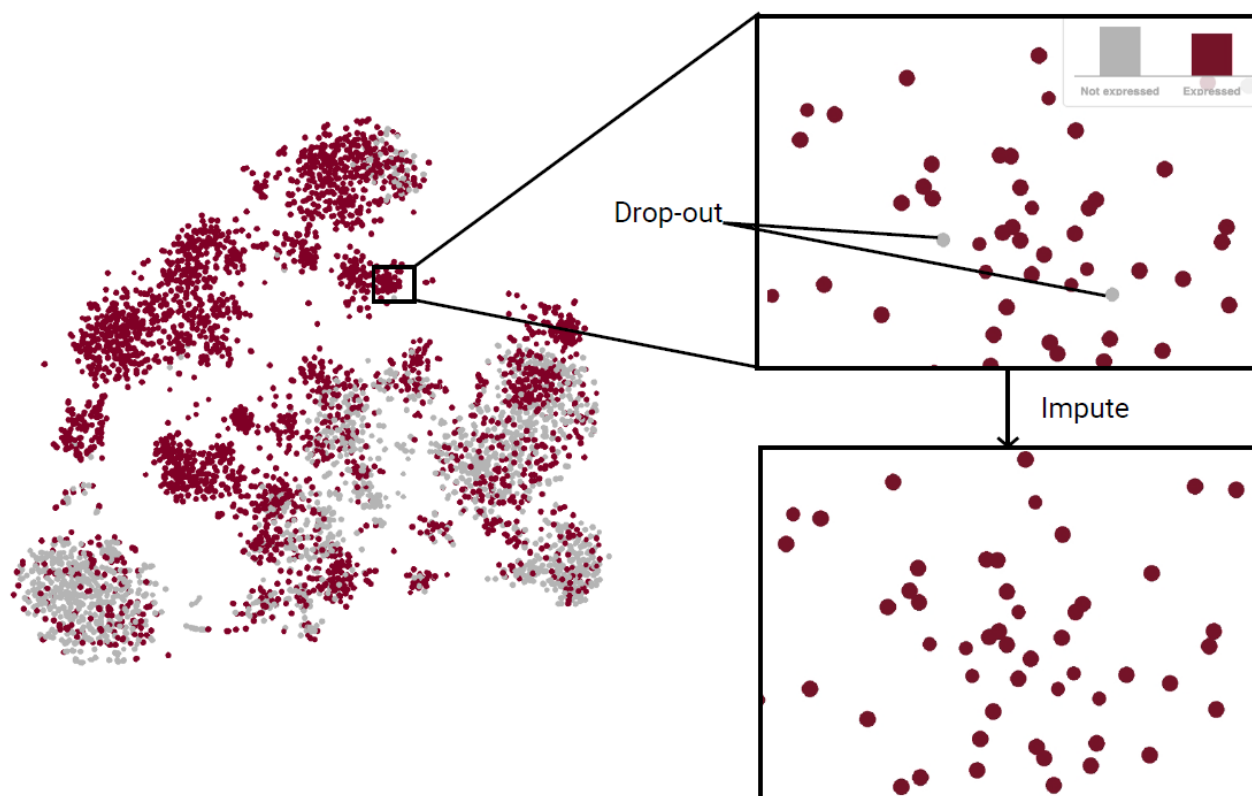
## C. 発現値を2値化する

各遺伝子の適切なカットオフしきい値を決定すると、このしきい値以下の発現値を持つ細胞は非発現、それ以上の発現値を持つ細胞は発現とみなされます。

## D. dropoutシグナルを置き換える

バックグラウンドノイズなどのdropoutはランダムに発生します。全遺伝子プロファイルを考慮すると、1つの遺伝子が存在しなくても、他の同じタイプの細胞との類似性に与える影響はごくわずかであり、セルタイプによる正しいクラスタリングが可能になります。

その結果、ある遺伝子を発現していない細胞でも、隣接する細胞の大多数がその遺伝子を発現していれば、アルゴリズムは集団内の一貫性を保つために発現であると推測します。



### Figure 4. dropoutシグナルを置き換える

示された細胞はマーカーを発現しているはずだが、技術的な制約により発現していないように見える。インプテーションプロセスでは、隣接する細胞の発現プロファイルを使用することでこれを補正する。

## BioTuring Smooth Binary アルゴリズムは以下のツールで使用されます。

- **Marker genes**

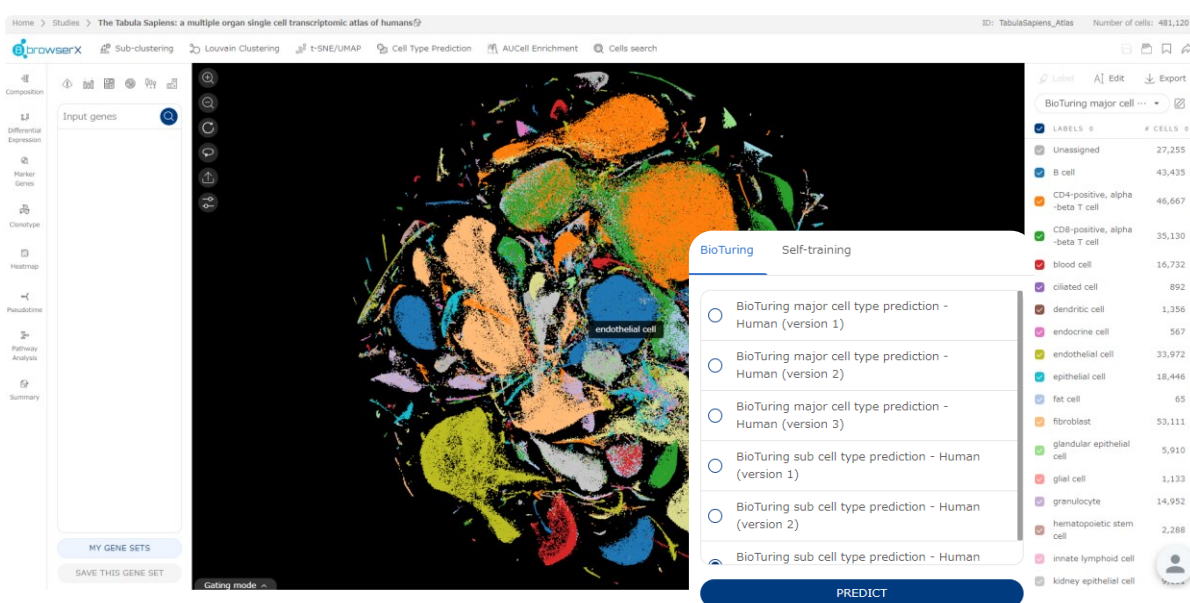
ある細胞集団を他の細胞集団と区別するのに役立つ遺伝子パネル（発現量の異なる遺伝子のセット）を見つけることが出来る機能です。

- **Cell type prediction**

BioTuring社の包括的なデータベースに基づいて構築された自動 Cell Type 予測機能です。クリックだけで Cell Type の予測を行うことが可能です。

- **Cell search**

数百万の細胞のデータベース全体を参照して、選択した集団と類似した特性または生物学的シグネチャーを共有する細胞集団を引き出すことができます。



### Cell type prediction