# scRNA-seq

東京科学大学　ILA国府台
中林　潤

JH人材育成課　バイオインフォマティシャン育成講座　⑩

営利目的でのご利用はご遠慮ください

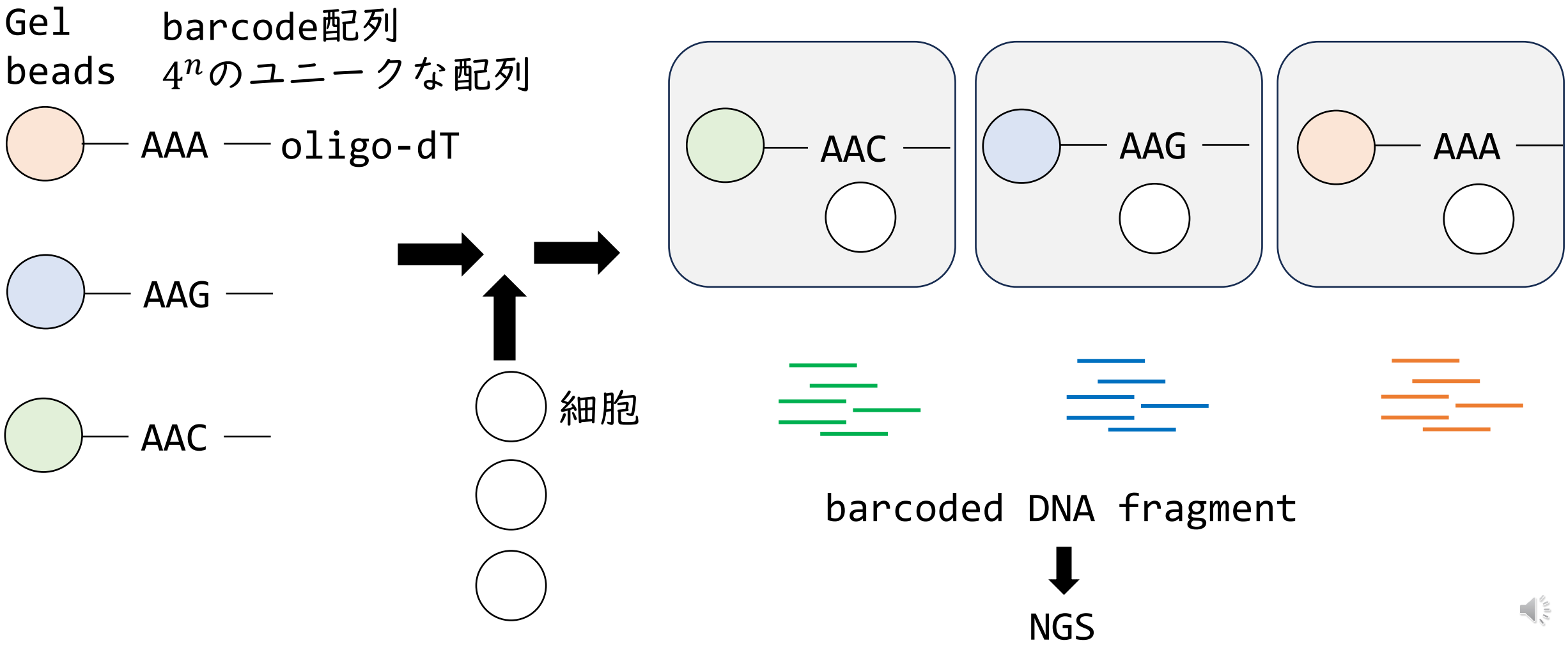
www.icrweb.jp

# scRNA-seq

- 1細胞レベルでRNA-seqを行うもの
- それぞれの細胞に発現している全ての遺伝子の発現量を網羅的に定量する。

営利目的でのご利用はご遠慮ください
www.icrweb.jp

# Barcoding




ユニークな塩基配列で細胞を標識する

営利目的でのご利用はご遠慮ください

www.icrweb.jp

# 10x Genomics GEM-X


https://www.10xgenomics.com/platforms/chromium/technology

営利目的でのご利用はご遠慮ください

www.icrweb.jp

# データ処理

- データのクオリティチェック
  細胞数
  遺伝子数
  ミトコンドリア遺伝子の割合

- 正規化
  トータルカウント数で正規化

- 変動遺伝子の抽出

- スケーリング
  平均0、分散1にスケーリング

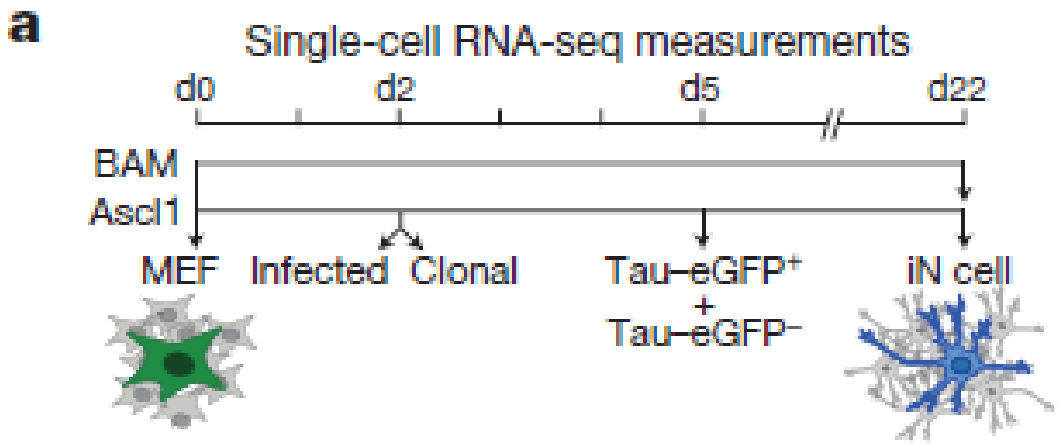
営利目的でのご利用はご遠慮ください
www.icrweb.jp

# 次元縮約

- データの次元数を減らす。
- PCAで線形圧縮
- tSNE、UMAPで非線形圧縮

営利目的でのご利用はご遠慮ください
www.icrweb.jp

# iNeuron細胞の分化




MEF細胞にAscl1遺伝子を強制発現すると
未分化状態に戻る

⬇

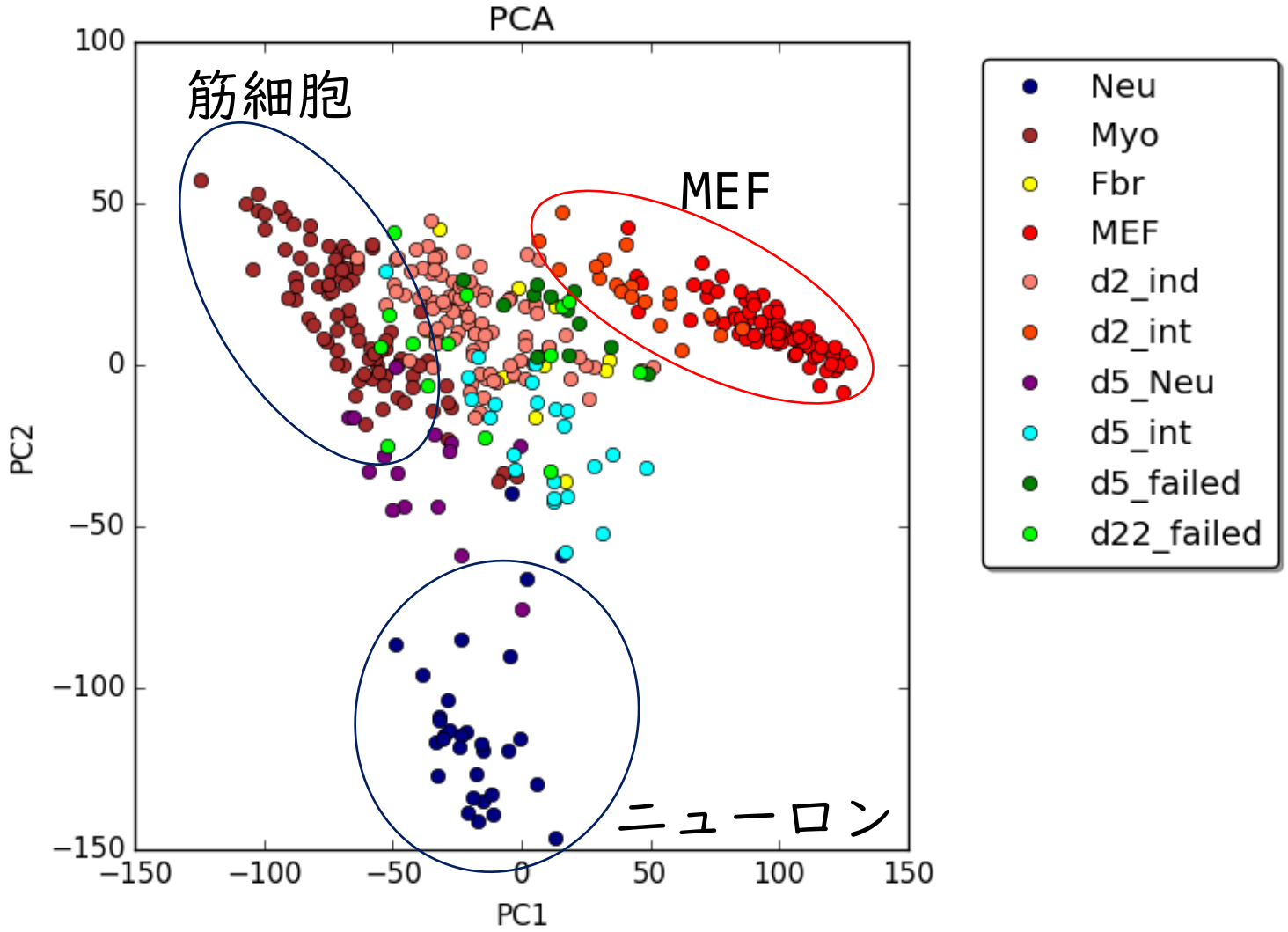
神経細胞や筋細胞へ分化する

Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq

Treutlein B et al. ***Nature*** 2016 8;534(7607):391-5

営利目的でのご利用はご遠慮ください

# PCA



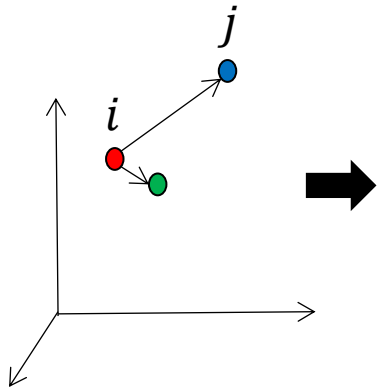

営利目的でのご利用はご遠慮ください

www.icrweb.jp

# tSNE

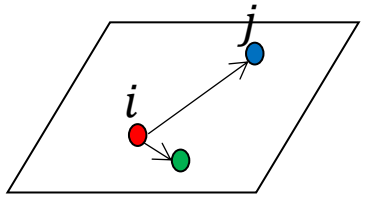
高次元空間上の距離を正規分布で表現

$$p_{j|i} = \frac{e^{\left(-\left\|x_i - x_j\right\|^2 / 2\sigma_i{}^2\right)}}{\sum_{k \neq i} e^{\left(-\left\|x_i - x_k\right\|^2 / 2\sigma_i{}^2\right)}}$$




低次元空間上の距離をt分布で表現

$$q_{j|i} = \frac{e^{\left(-\left\|y_i - y_j\right\|^2\right)}}{\sum_{k \neq i} e^{\left(-\left\|y_i - y_k\right\|^2\right)}}$$




カルバック・ライブラー情報量：$D(p\|q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$

カルバック・ライブラー情報量が最大となるよう、低次元空間の座標を決める。

高次元空間内で距離の近い点同士は、低次元に変換後も近傍にいる確率が高いとする。

営利目的でのご利用はご遠慮ください

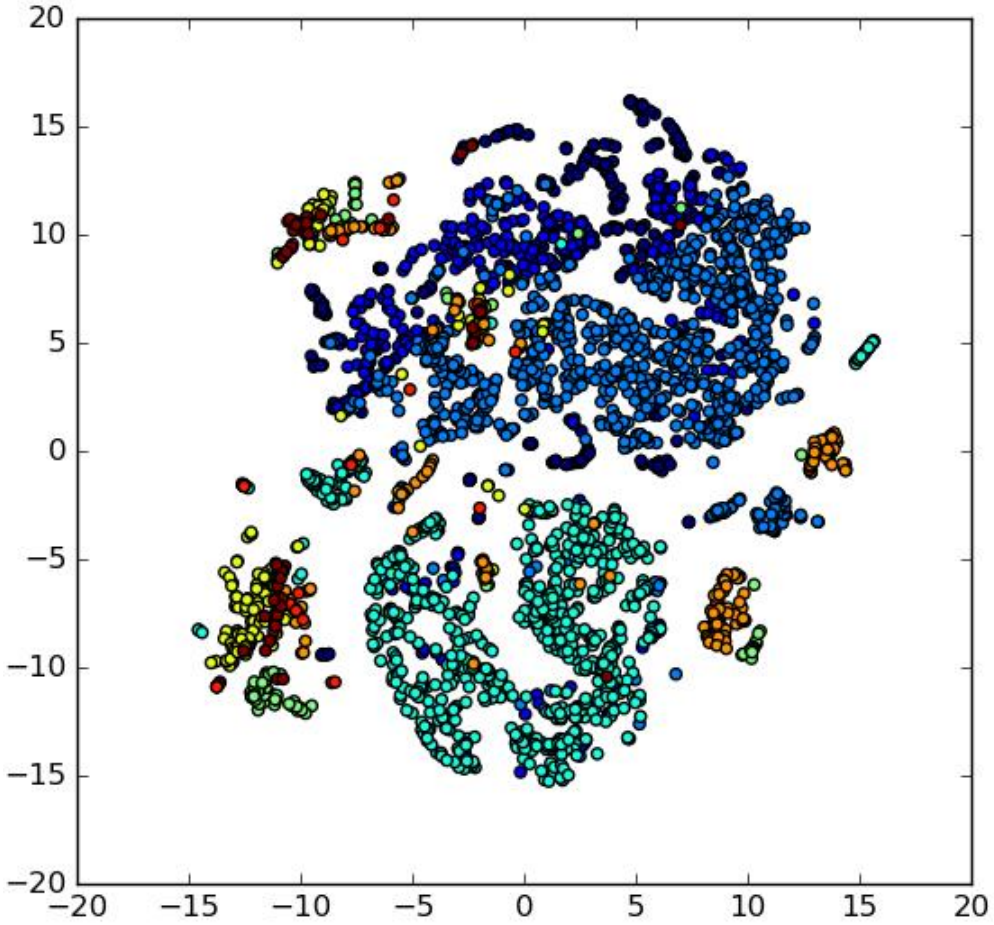
www.icrweb.jp

# tSNEの実例

19946次元のデータ

↓

2次元に縮約




- interneurons
- pyramidal SS
- pyramidal CA1
- oligodendrocytes
- microglia
- endothelial-mural
- astrocytes ependymal

Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.
Zeisel A. et al. ***Science*** 2015 Mar 6;347(6226):1138-1142

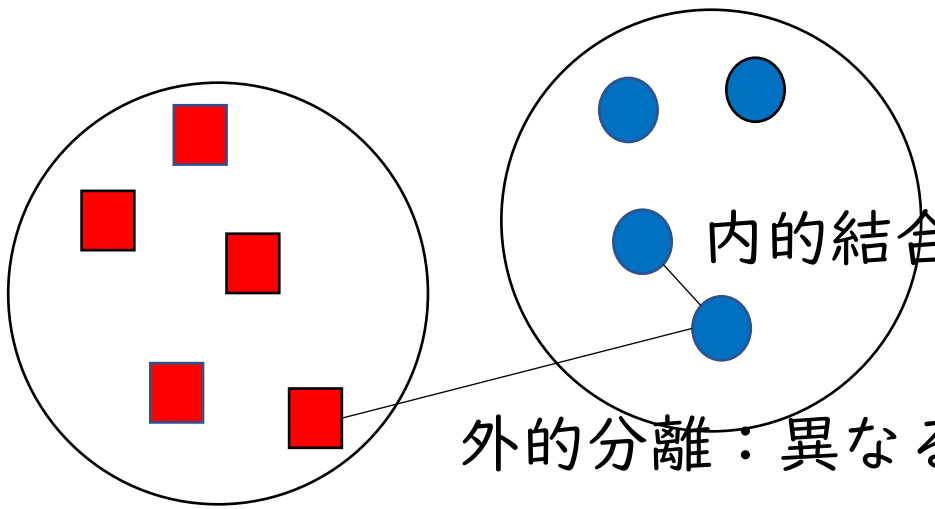
営利目的でのご利用はご遠慮ください
www.icrweb.jp

# クラスタ化

クラスタ化：サンプルを部分集合に分ける

距離（類似度）

ユークリッド距離：$D = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$



内的結合：同じ分類に属する対象は互いに似ている

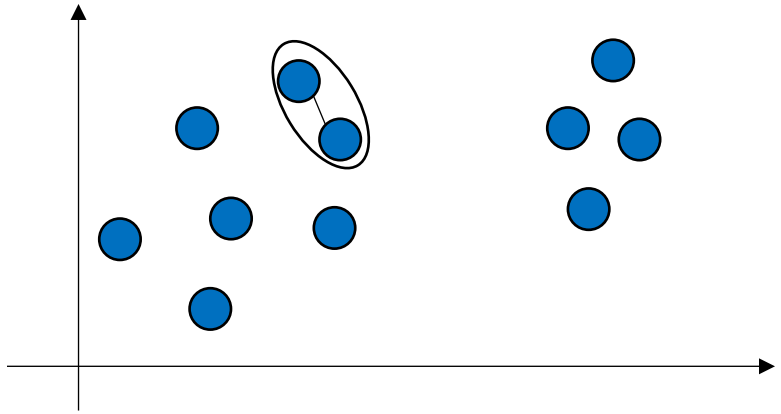
外的分離：異なる分類に属する対象は似ていない

営利目的でのご利用はご遠慮ください

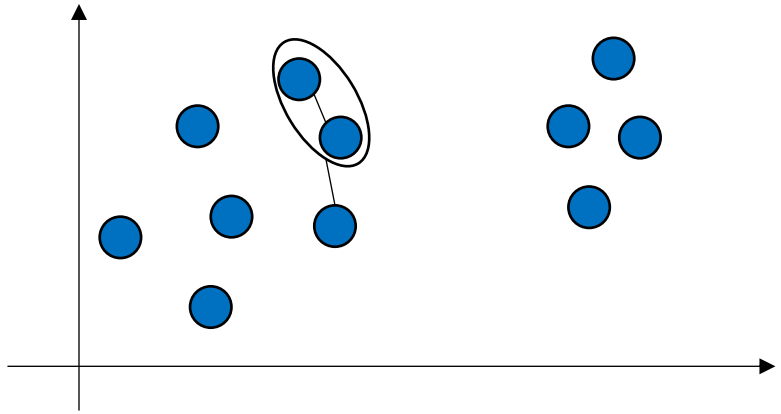
www.icrweb.jp

# 階層的クラスタ化（hierarchical clustering）

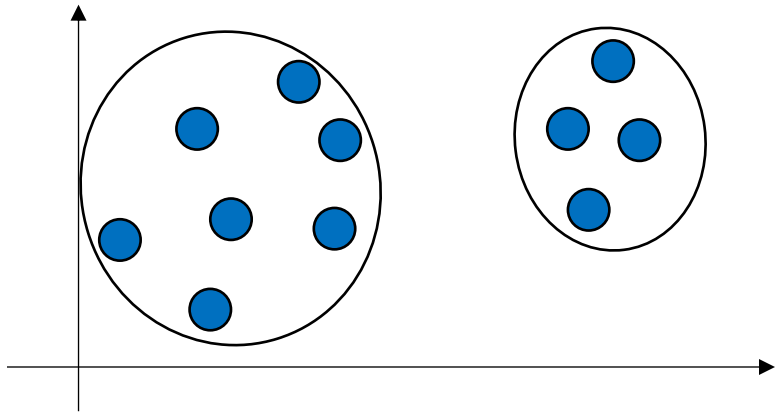
各データポイント間の距離を計算し、
距離の短いものからクラスタ化する



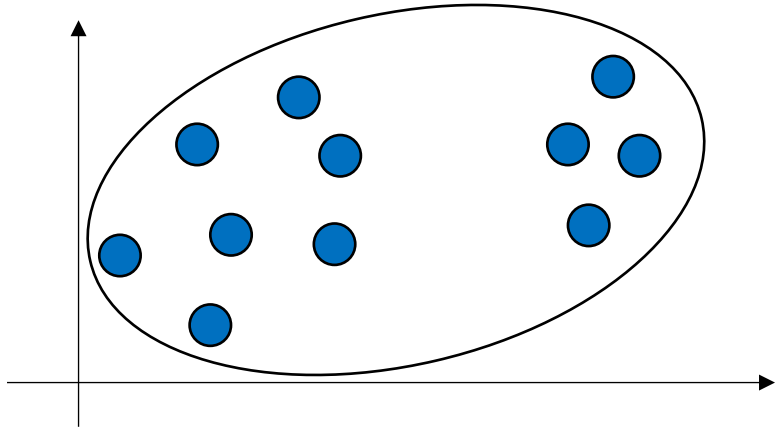
出来たクラスタを一つの点として、
距離を計算する



クラスタを再編していく



全てのデータポイントを含むクラスタが
できるまで繰り返す

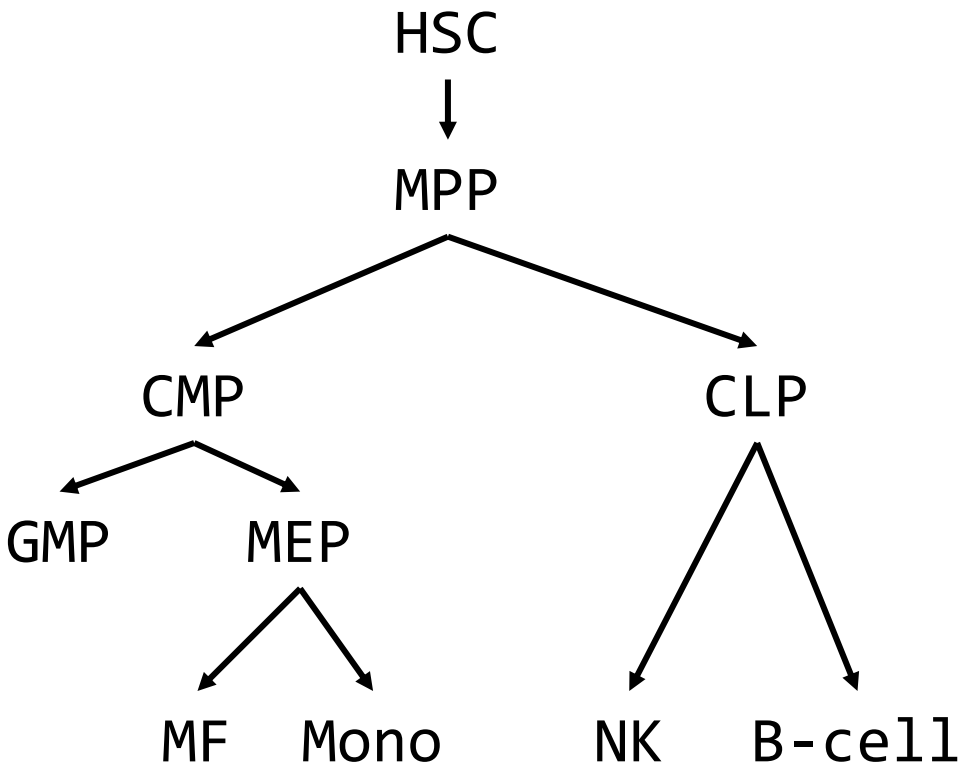


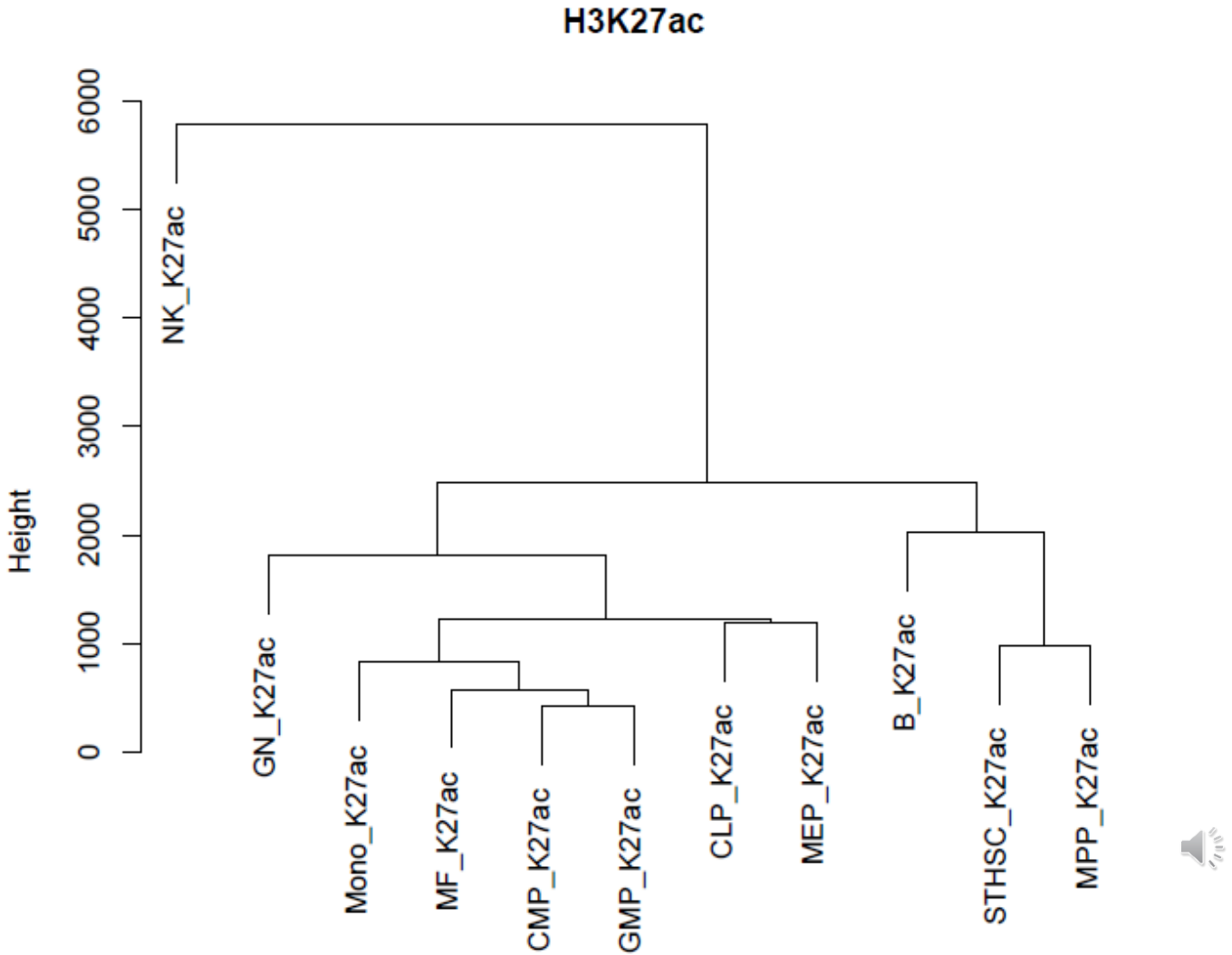
営利目的でのご利用はご遠慮ください

www.icrweb.jp

# 階層的クラスタ化

血球系細胞の細胞系譜




営利目的でのご利用はご遠慮ください

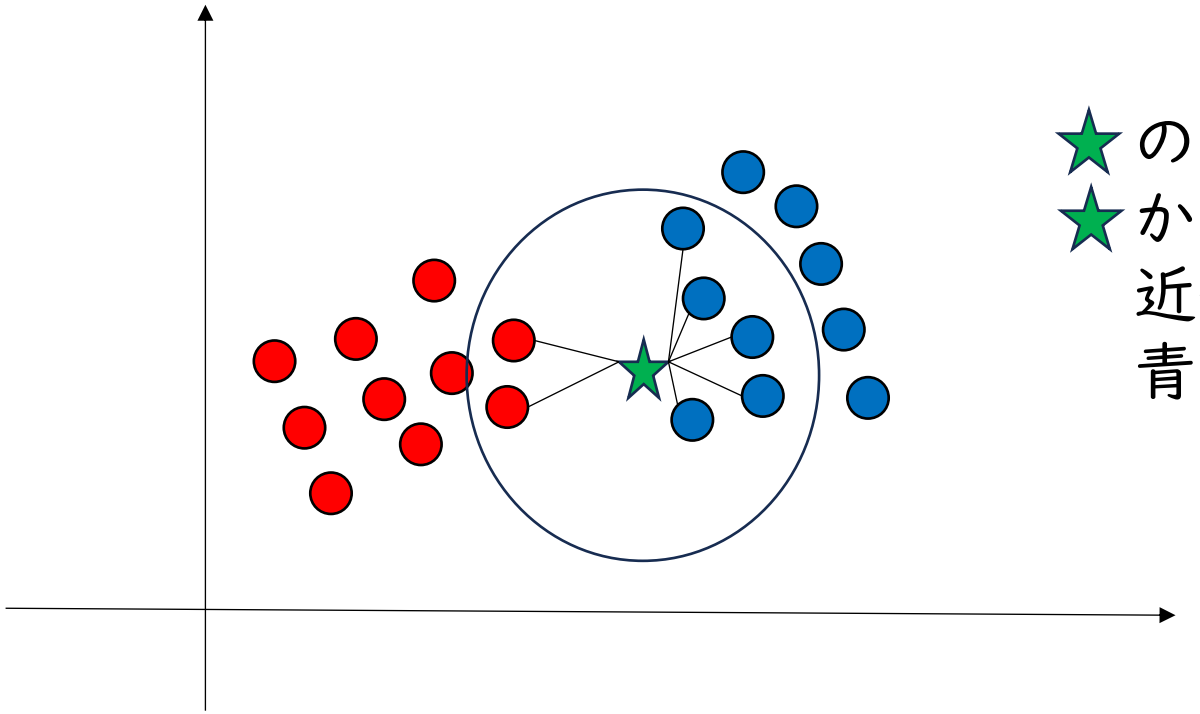
www.icrweb.jp

# K-nearest Neighbors（KNN）

- ランダムにグループ分け
- 1点を選ぶ
- 他の点とのユークリッド距離を計算
- 近傍のk個の点を取る
- 最も多いグループに入れる
- 繰り返し

営利目的でのご利用はご遠慮ください
www.icrweb.jp

# K-nearest Neighbors（KNN）



★の点が青、赤どちらのグループに属するか？
★からの距離を計算
近傍7個の点の中で、青が5個、赤が2個
青に入れる

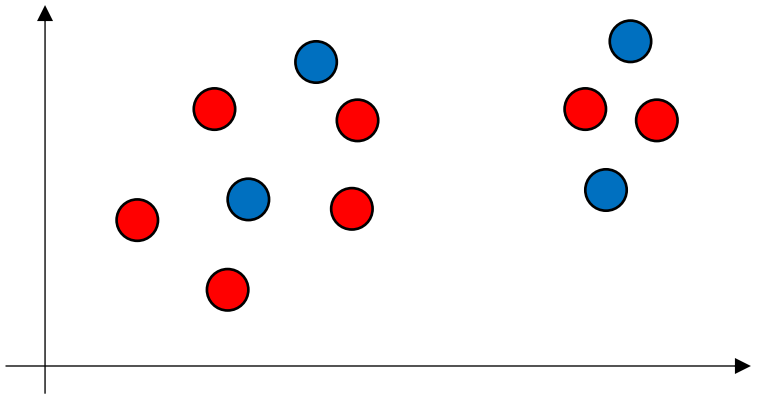
- 簡便に実装できる。
- Kの値を適切に設定する必要がある。
- データサイズが大きくなると計算量が増える。

営利目的でのご利用はご遠慮ください
www.icrweb.jp

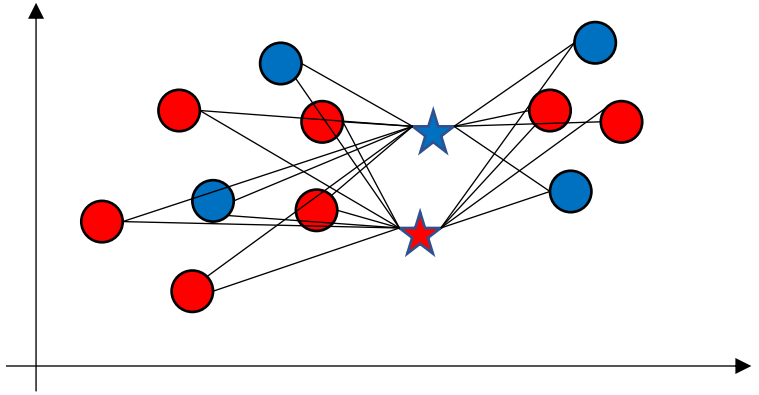
# k-means法

- k-means法によるクラスタ化

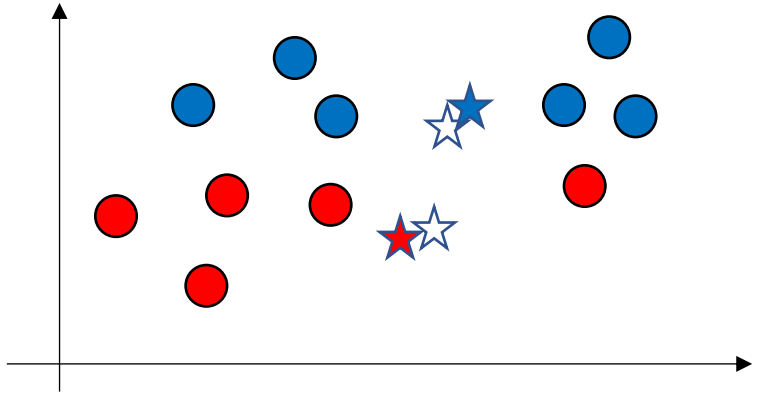
1）ランダムにクラスタを割り振る



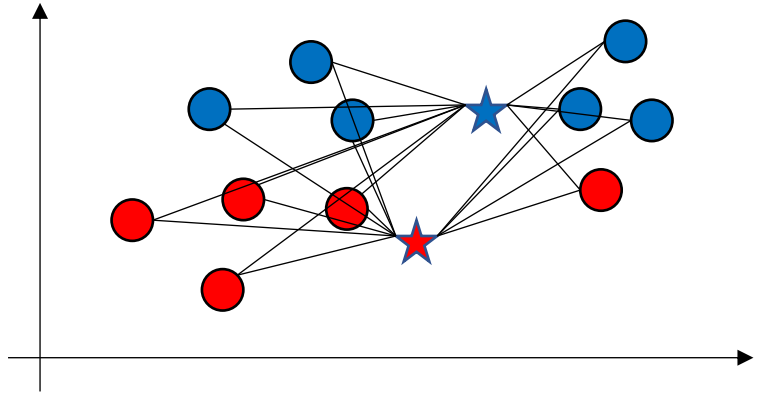
2）重心と重心までの距離を計算する



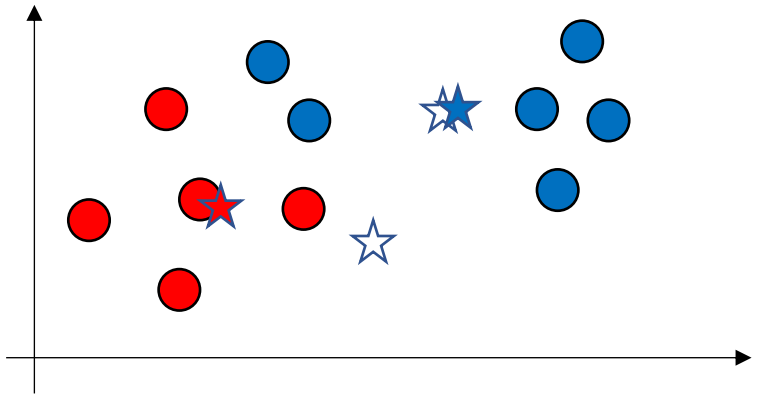
3）重心の近いクラスタに変更する



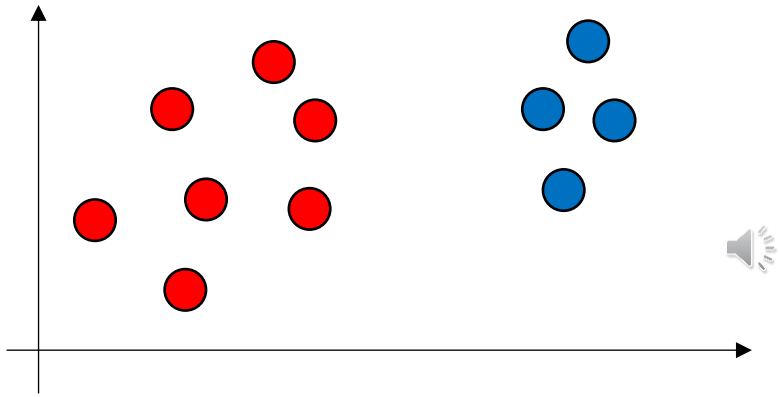
4）重心と重心までの距離を再計算



5）クラスタ変更



6）重心が変わらなくなるまで繰り返す

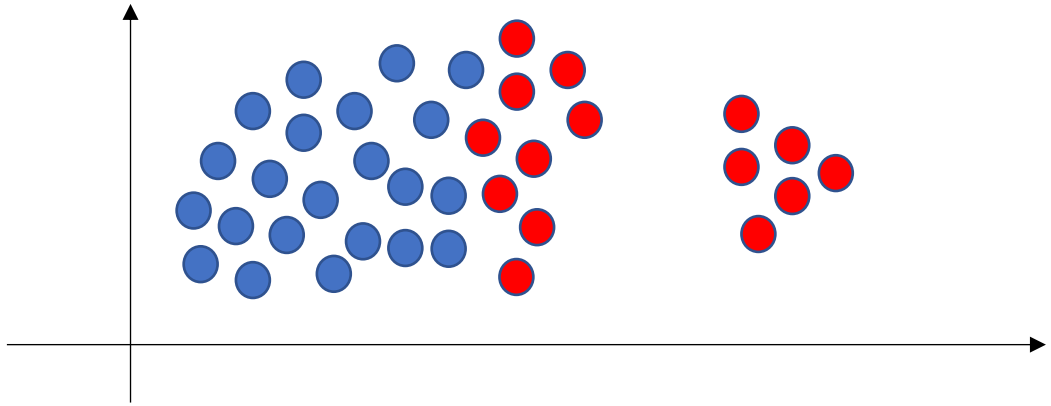


営利目的でのご利用はご遠慮ください

www.icrweb.jp

# k-means法の問題点

- クラスタ数を予め指定する必要がある
- サンプルは各クラスタに均等に分配されていることを暗に仮定している

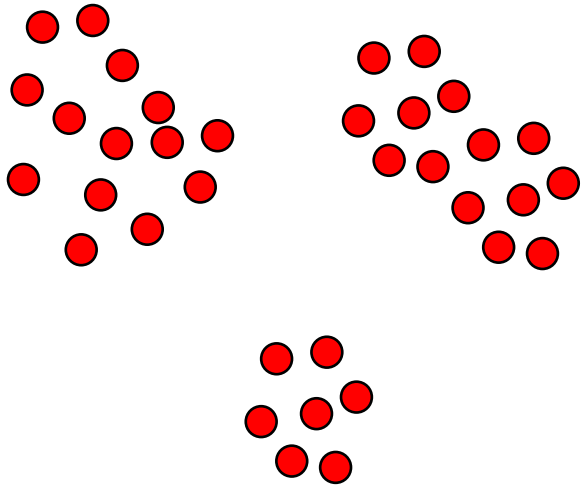


- 収束する保証がない
- 結果が試行ごとに異なる可能性がある

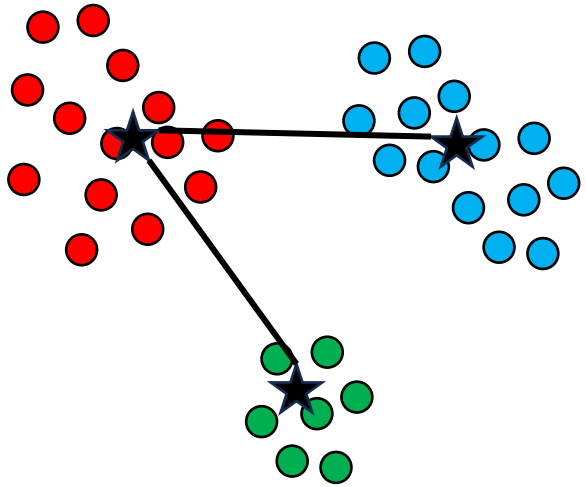
営利目的でのご利用はご遠慮ください
www.icrweb.jp

# Trajectory inference

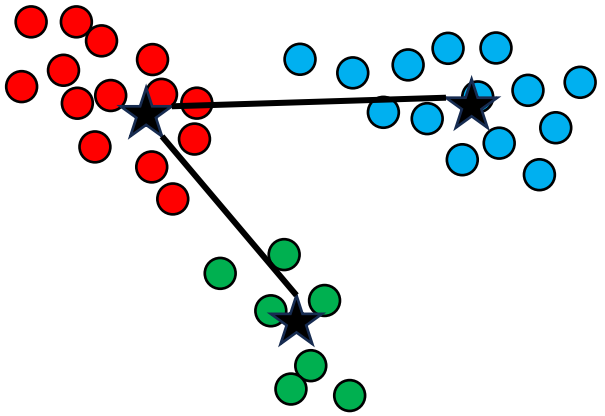
- データ間の距離を連続的に変化する系統関係と見做して、遷移経路を推定する。
- 系統樹の形成



低次元空間にサンプルを配置



クラスタ化
重心を結んでtreeを形成



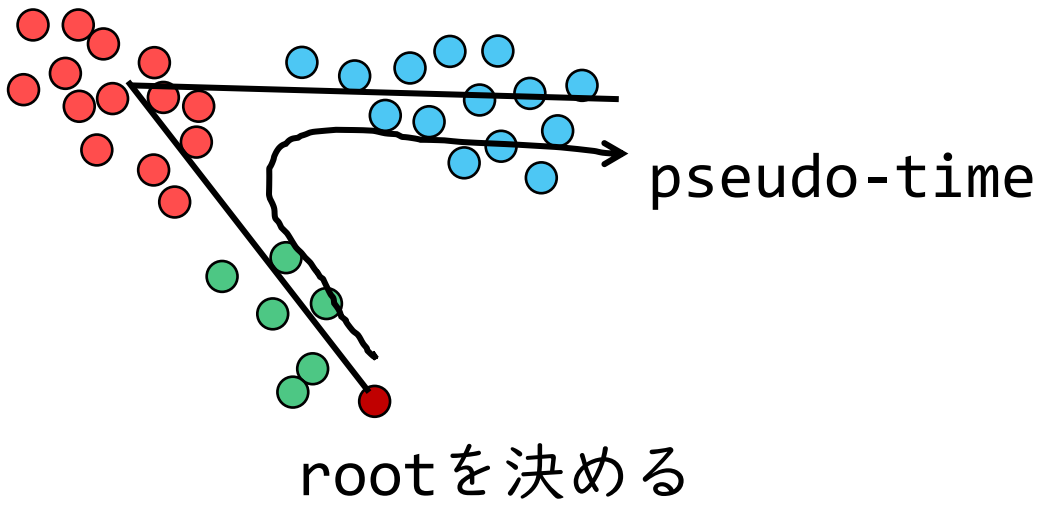
最も近い枝にサンプルの
座標を寄せる

営利目的でのご利用はご遠慮ください

www.icrweb.jp

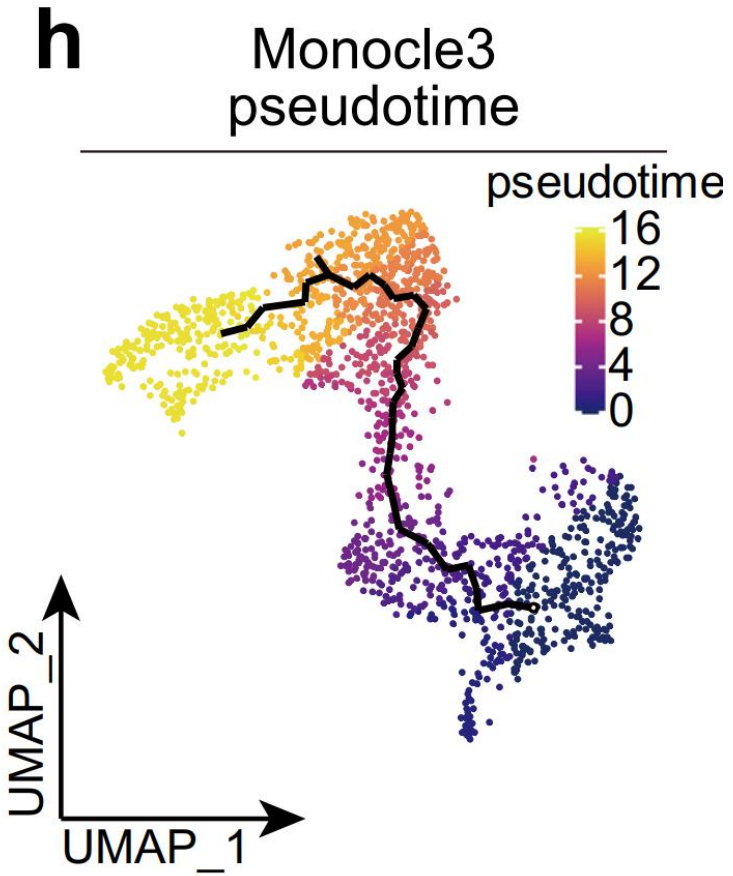
# Pesudo-time

- 遷移経路の推定
- rootを決めて、rootからの距離の近い方から遷移していくとして、pseudo-timeを推定する




営利目的でのご利用はご遠慮ください
www.icrweb.jp

# Trajectory inference




Single cell transcriptomics clarifies the basophil differentiation trajectory and identifies pre-basophils upstream of mature basophil.
Miyake K et al. ***Nat Commun.*** 2023 May 18;14:2694.

営利目的でのご利用はご遠慮ください

# まとめ

- 1細胞レベルで遺伝子の発現量を網羅的に解析する手法
- バーコード配列で細胞を標識する。
- クラスタ化：サンプルをデータの特徴に応じてグループに分ける。
- Trajectory inference：サンプルを低次元空間内でクラスタ化し、クラスタ間の距離から系統を想定して関係を推定する。

営利目的でのご利用はご遠慮ください
www.icrweb.jp