

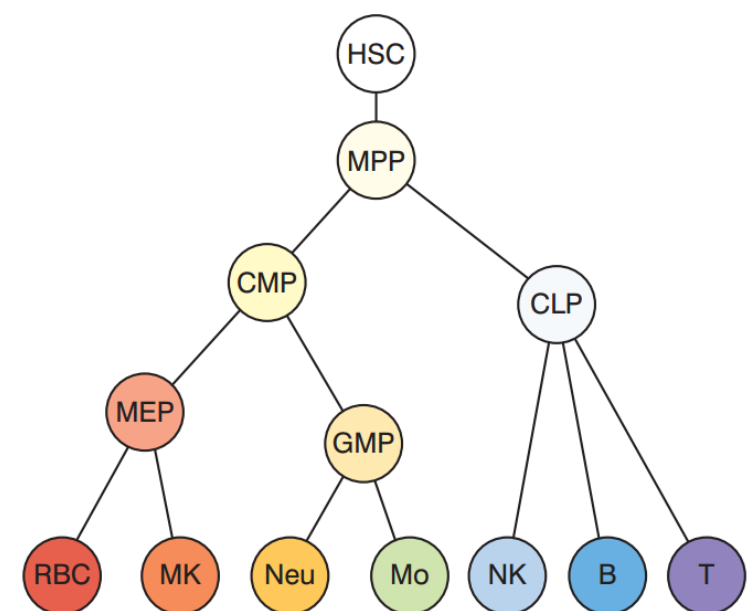
# 細胞の遺伝子発現データ解析と離散数学

早水 桃子（統計数理研究所 モデリング研究系 助教， JSTさきがけ研究者）

## Introduction：細胞生物学とグラフアルゴリズム

### 細胞の分化

細胞の形態や機能が変わり、  
別の種類の細胞に変化する  
不可逆的な運命決定プロセス

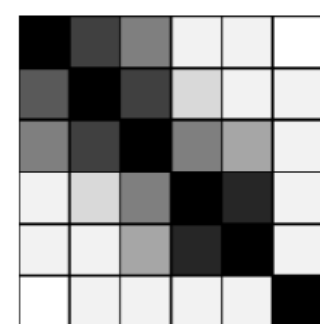


### 細胞生物学者のニーズ

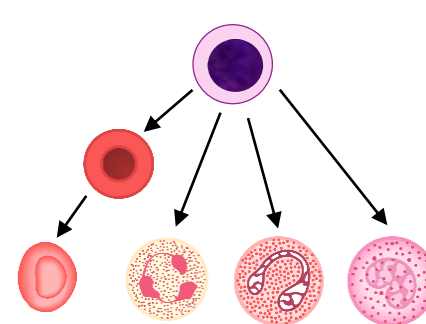
個々の細胞の遺伝子発現量を  
包括的に計測する技術の登場

→ 細胞の詳細な情報を活用して  
細胞分化の木モデルを構築したい  
(→ 運命を制御する遺伝子を予測)

距離行列



細胞分化の木モデル

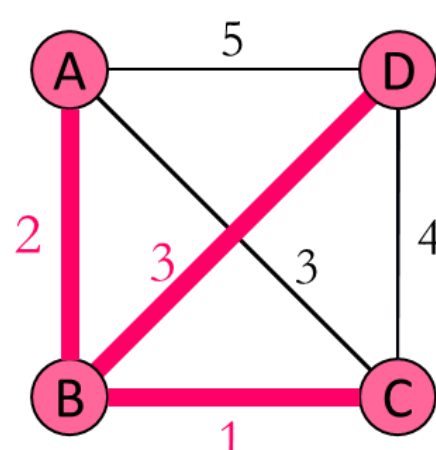


### 最小全域木を求めるアルゴリズム

- 細胞同士の距離（非類似度）から細胞の分化の木モデルを構築するためによく使われる経験則

- ✓ 計算効率 ◎
- ✓ 生物学的な実績 ◎

(Trapnell et al., Nature Biotechnology (2014) など)

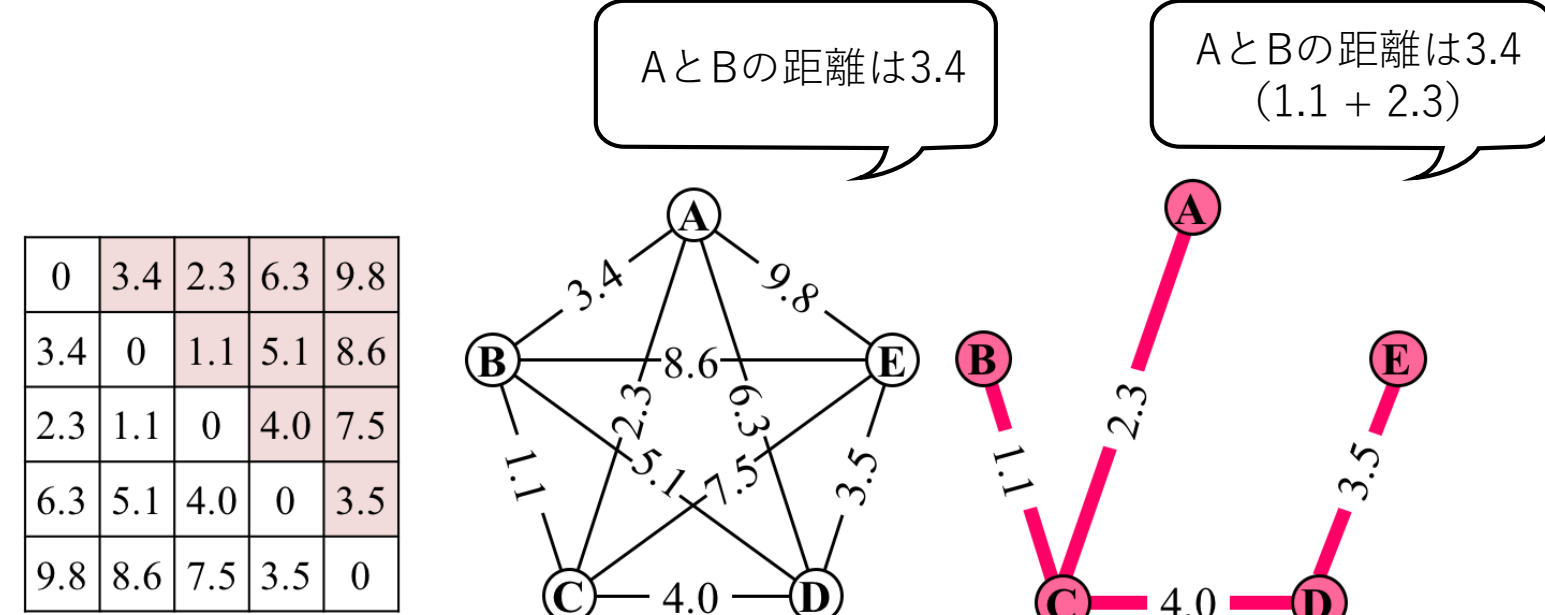


- 既存の経験則の根本的な問題点

最小全域木はもとの距離情報をよく表していることもあるが、  
全然フィットしていないこともある！

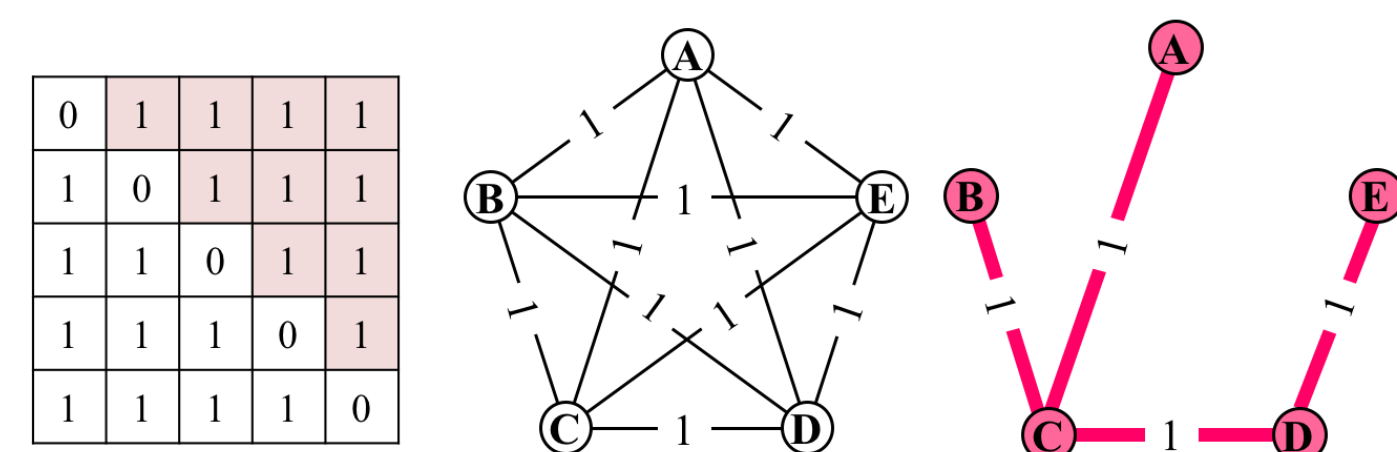
最小全域木が完璧にフィットしている例

- 任意の二点に対して、ペアワイズ距離が保存されている



最小全域木が全くフィットしていない例

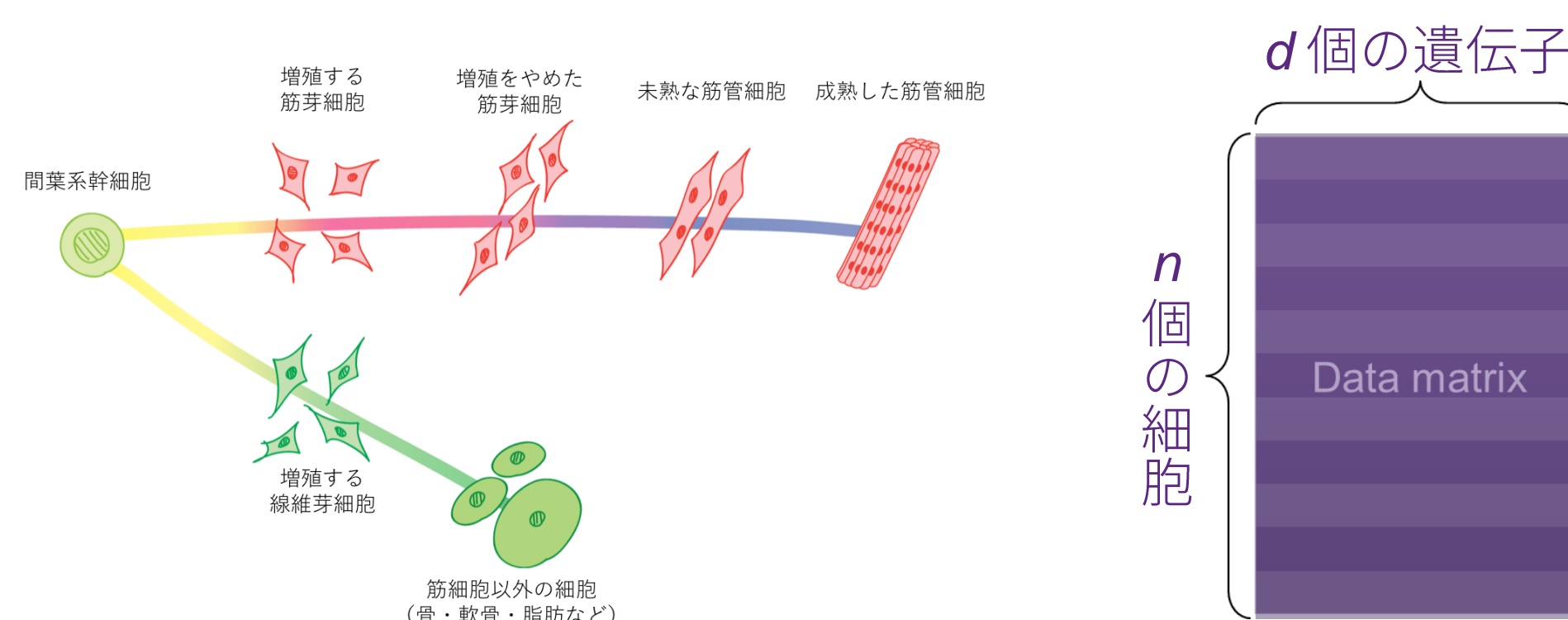
- もとの距離が保存されておらず、一部の距離情報が失われている



### 最小全域木を用いたデータ解析の実例

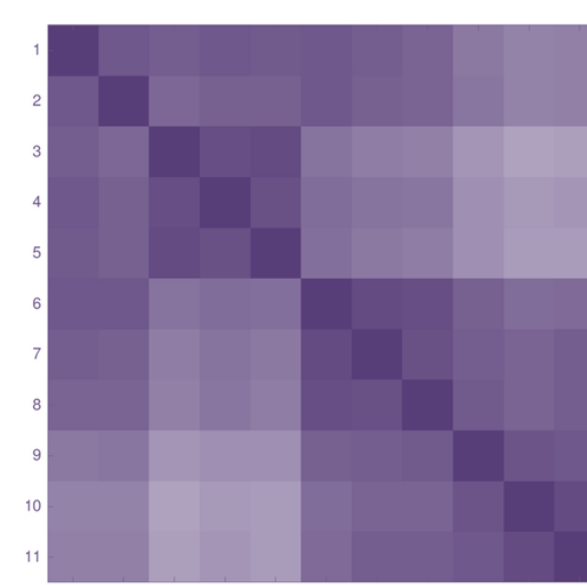
#### データセット [Trapnell et al., Nature Biotechnology (2014)]

骨格筋細胞への分化メカニズムを調べるために様々な細胞から測定  
された遺伝子発現データ（1つの細胞からd次元の実ベクトルが得られる）

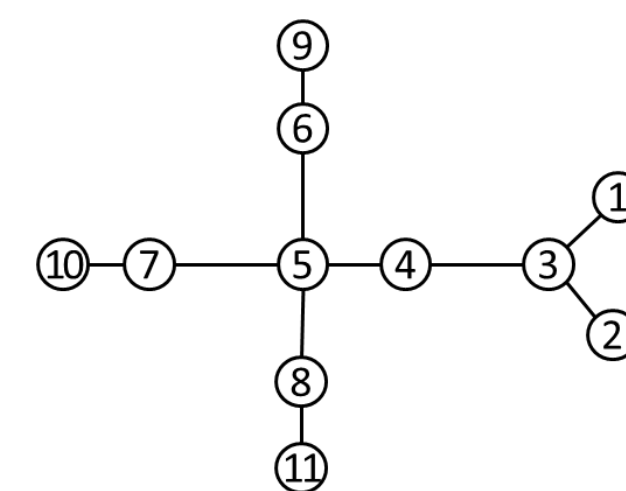


### データ解析の流れ

- データ行列から距離行列Dを計算し、最小全域木Tを計算・描画する

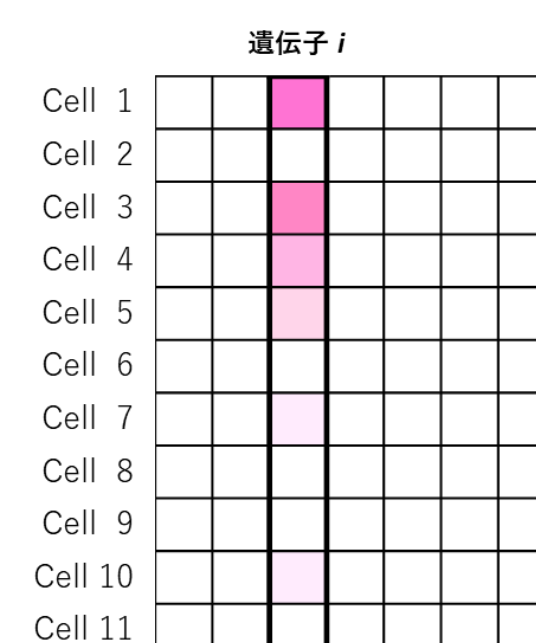
 $n \times n$  の距離行列D

何らかの全域木T（頂点数n）

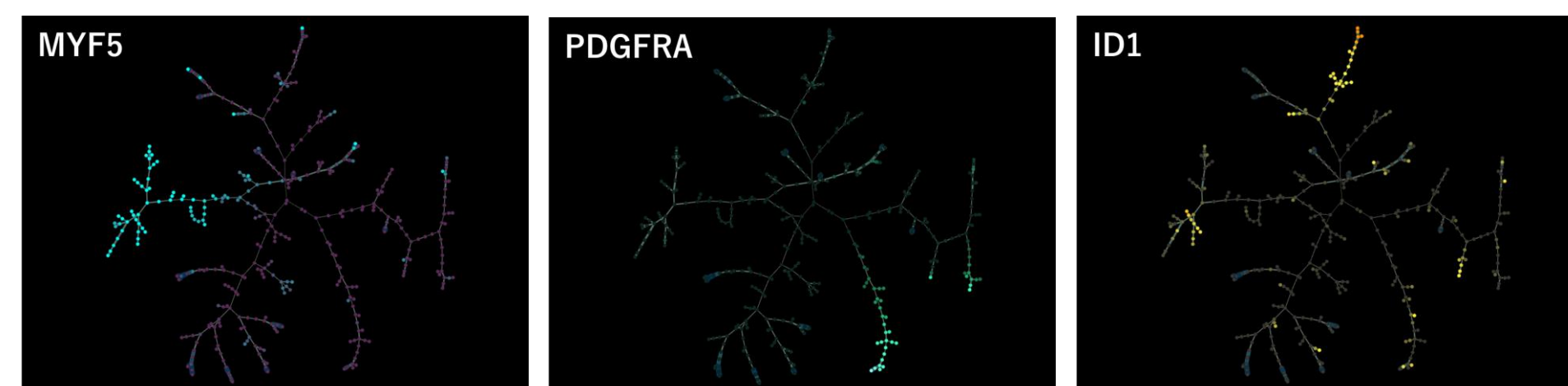
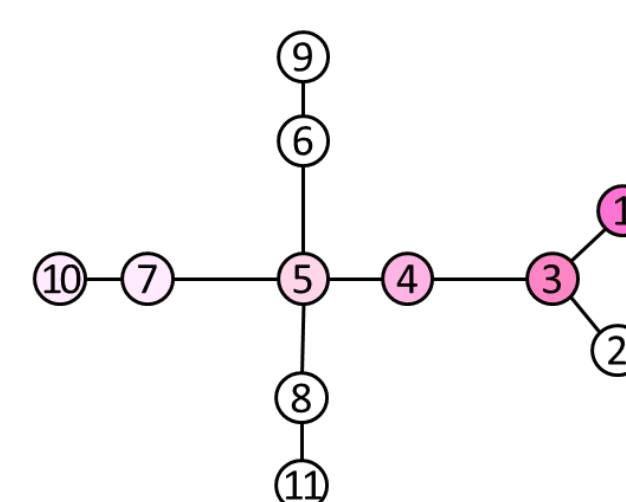


- 各細胞の遺伝子発現量の値に応じて頂点に色を付けることで、  
幹細胞の分化メカニズムに重要な遺伝子を視覚的に探索する

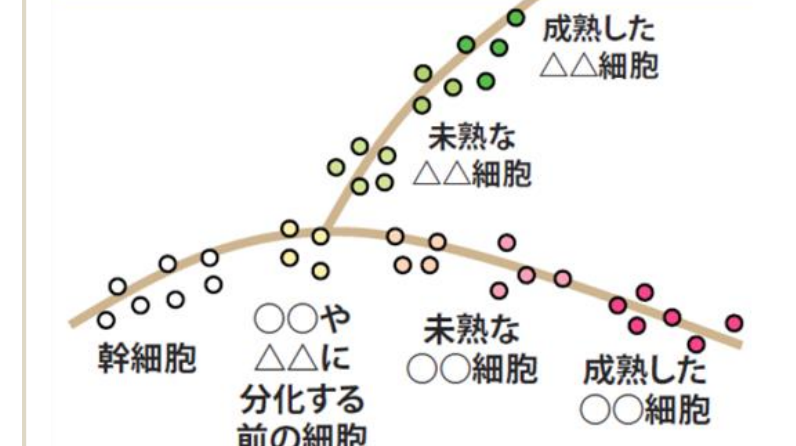
もとのデータ行列



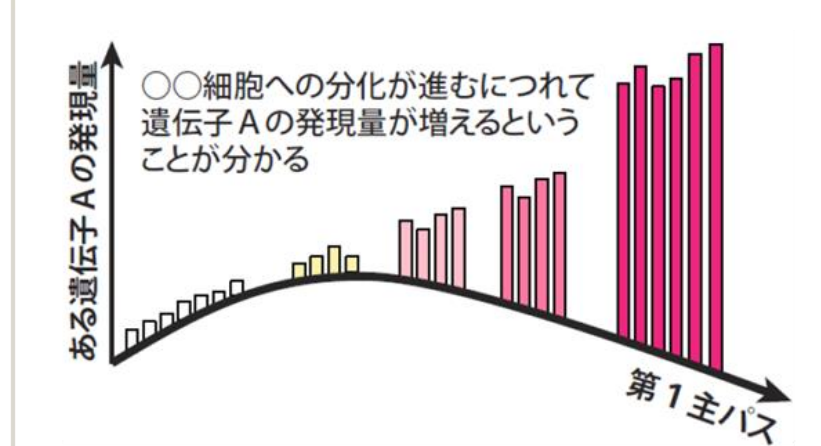
何らかの全域木T（頂点数n）



細胞分化の全体像の描出

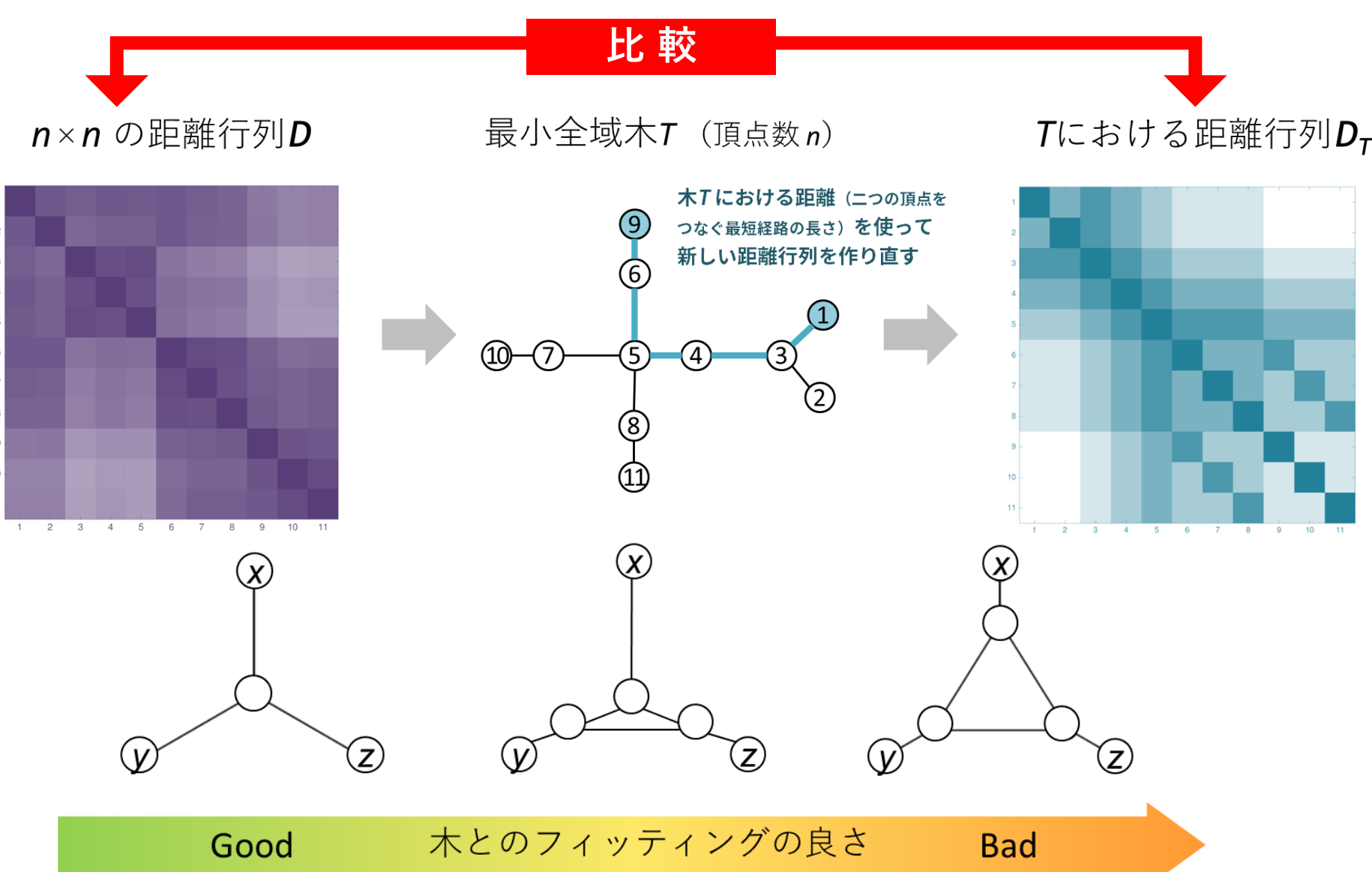


細胞分化に関わる重要遺伝子の探索



### 関連する ongoing work の紹介

#### 距離空間と木の“goodness-of-fit”の評価尺度の構成



#### 論文

M. Hayamizu, H. Endo and K. Fukumizu:  
“A characterization of minimum spanning tree-like metric spaces”,  
IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), vol.  
14, no. 2, pp. 468-471, March-April 2017,  
[doi:10.1109/TCBB.2016.2550431](https://doi.org/10.1109/TCBB.2016.2550431).