

# 非計量多次元尺度構成法を用いた 遺伝子発現実験のノイズリダクション

中大理工 物理 田口善弘  
イリノイ大 物理 大野克嗣

Satwik Rajaram

1. 非計量多次元尺度構成法とは？
2. 遺伝子発現実験の nMDS による解析
3. ノイズリダクション
4. まとめ

# 1. 非計量多次元尺度構成法 (nMDS) とは？

N 個の対象物がある。 ( $i=1, \dots, N$ )  
 $i_1$  番目と  $i_2$  番目の対象物の非類似性が  $s(i_1, i_2)$  とする。このとき、ある空間内に N 個の対象物を

$$\begin{aligned} \dots\dots < s(i_1, i_2) < \dots\dots \\ \dots\dots < r(i_1, i_2) < \dots\dots \end{aligned}$$

となるように埋めこめ、という手法。  
 $r(i_1, i_2)$  は距離であり、つまり「違うもの程遠くに配置」ということになる  
(距離と非類似度の大小関係を合わせよ)。  
大野・田口はアルゴリズムを開発。

例：N 人の人に L 項目の ○ × 式  
アンケートをする。

A : ○ × ○ ○ × ×  
B : ○ × × × × ○  
C : × × ○ ○ ○ ○

一致しない数を  
非類似度：  
A と B の非類似度：3

埋めこみ

距離の大小関係と  
非類似度をなるべく  
合わせる

A

C

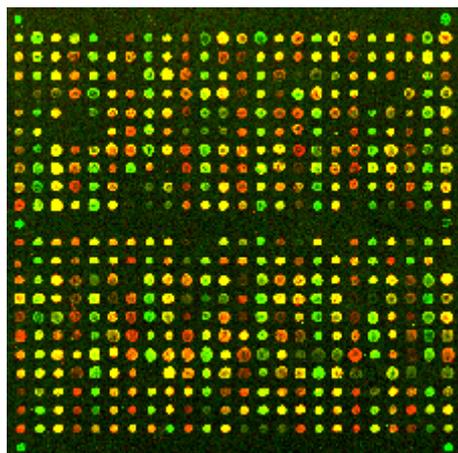
B

## 2. 遺伝子発現実験の nMDS による解析

Iyer, V. R. et al, (1999)  
Science, vol. 283, pp.83-87.

### 細胞周期遺伝子

517 genes, 11 experiments (time)

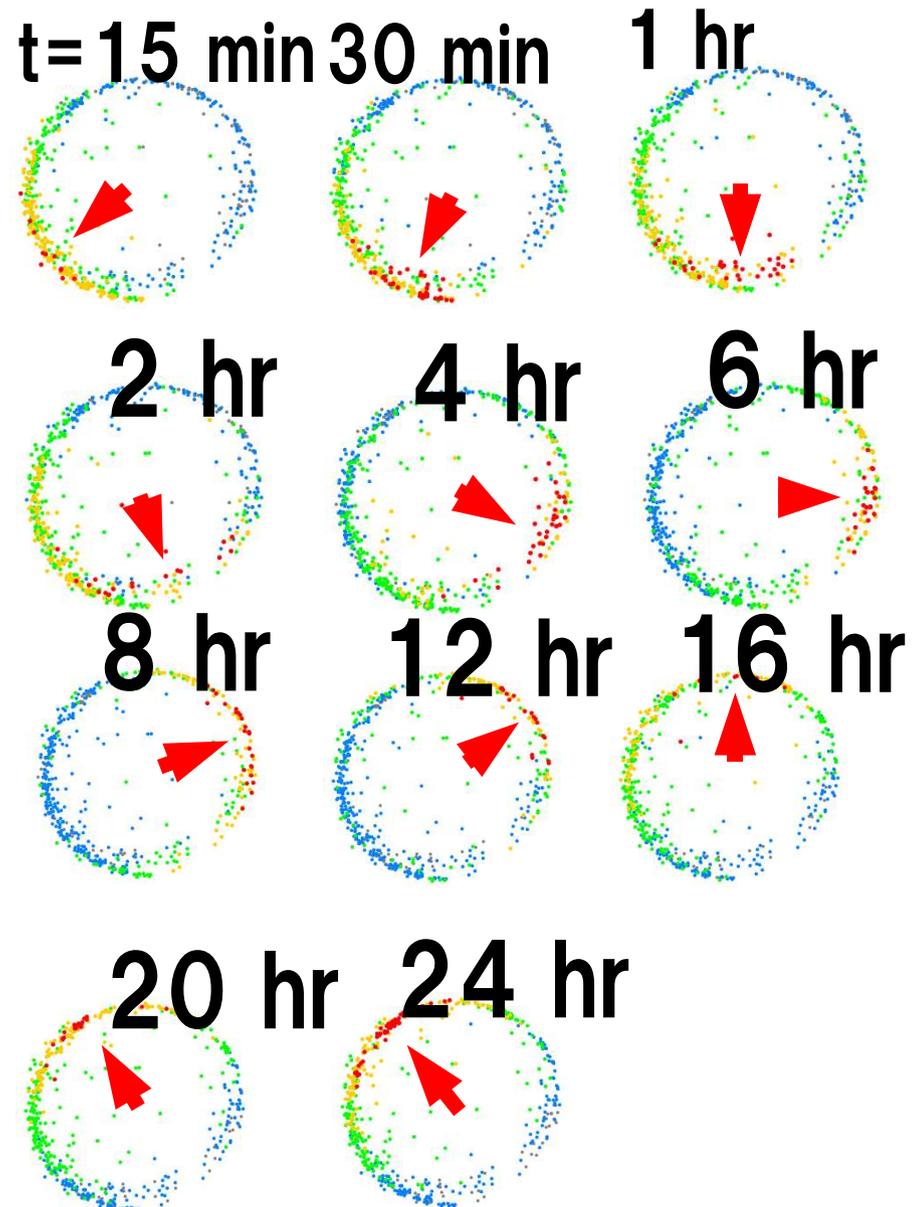
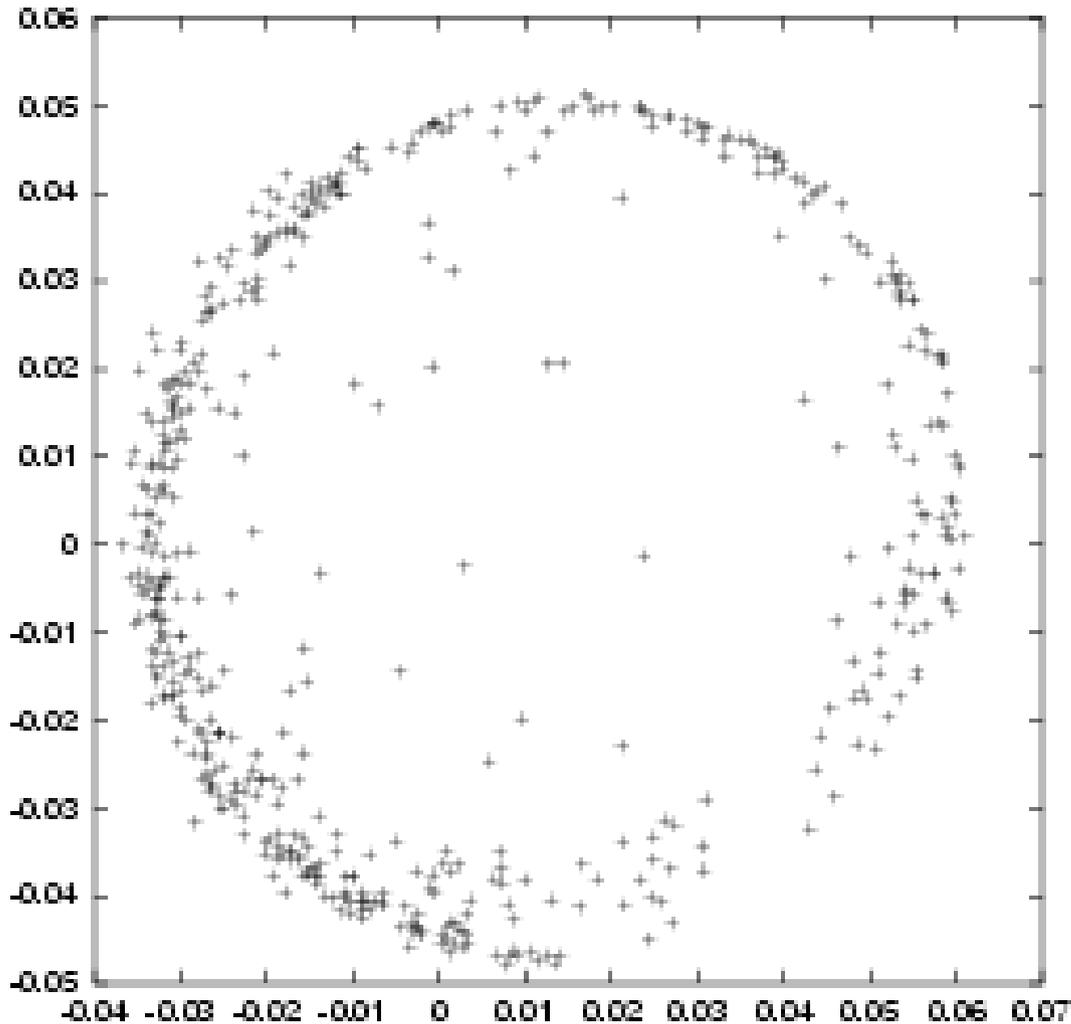


個々の点が遺伝子の発現度合いを表現している。一時刻、一枚のプレート

非類似度  $s(i_1, i_2)$  :

マイクロアレイ実験値の  
配列同士の相関係数 (の逆数など)

# 2次元に埋め込むとリングになる (発現遺伝子が時刻ごとに並ぶ)



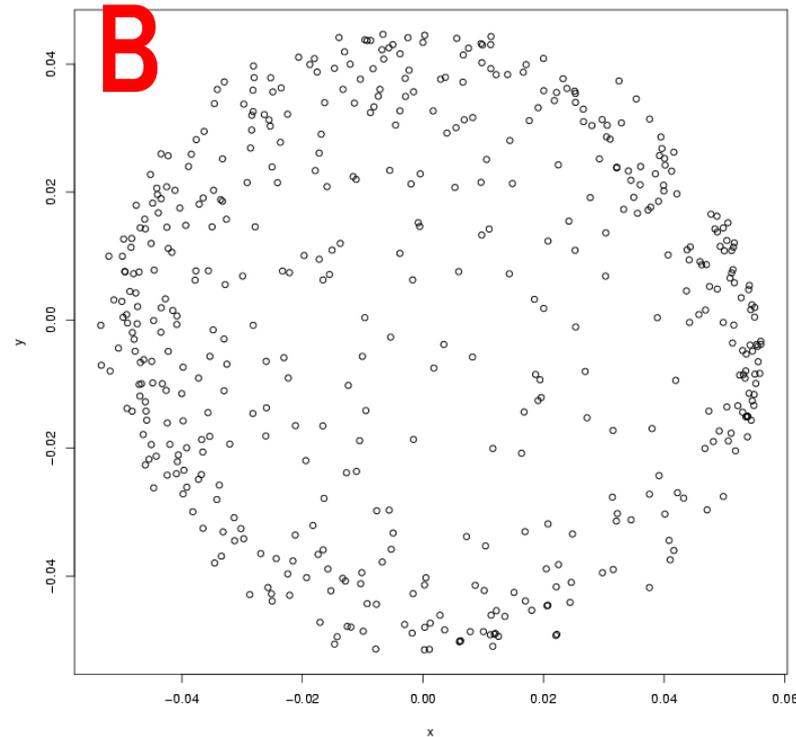
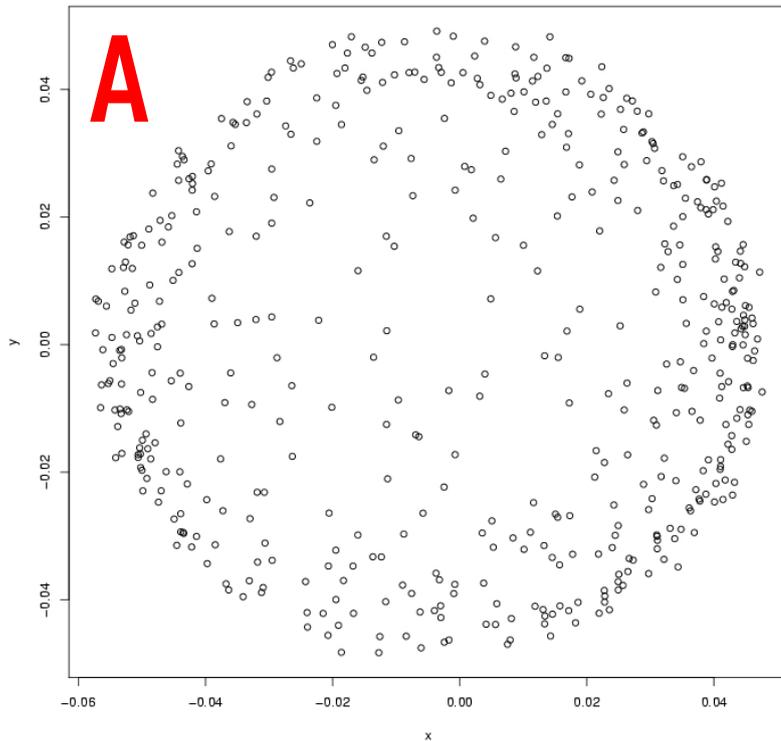
**nMDS は遺伝子発現実験の中の「隠れた秩序」を自動的に可視化する能力がある。**

**→ しかし、これは「選ばれた遺伝子」の場合。  
意味のある遺伝子がどれだか解らない  
場合にはどうするか？**

**当該実験に関係ない遺伝子は発現しない。  
しかし、ハウスキーピング遺伝子もあるので  
発現の多い遺伝子が実験スペシフィック  
とも言えない。**

### 3. ノイズリダクション

遺伝子を選ばないと明確なパターンが出ない  
(500遺伝子をランダムに選んだ)



Cho, R., J., et al (2001)  
Transcriptional regulation  
and function during the  
human cell cycle, Nature  
Genetics vol. 27 p48.

Blackshaw, S., et al (2004)  
Genomic Analysis of Mouse  
Retinal Development, PLoS  
Biology vol.2 e247.

○「実験に関係ない遺伝子」をどう取り除くか？

**実験に関係ない遺伝子の発現は  
時間に無関係のはず(乱数と区別不能)**

→ **縮約できない・うめこめない**

**「埋め込みの悪い遺伝子」を排除。**

# 埋め込みの悪さの定義:

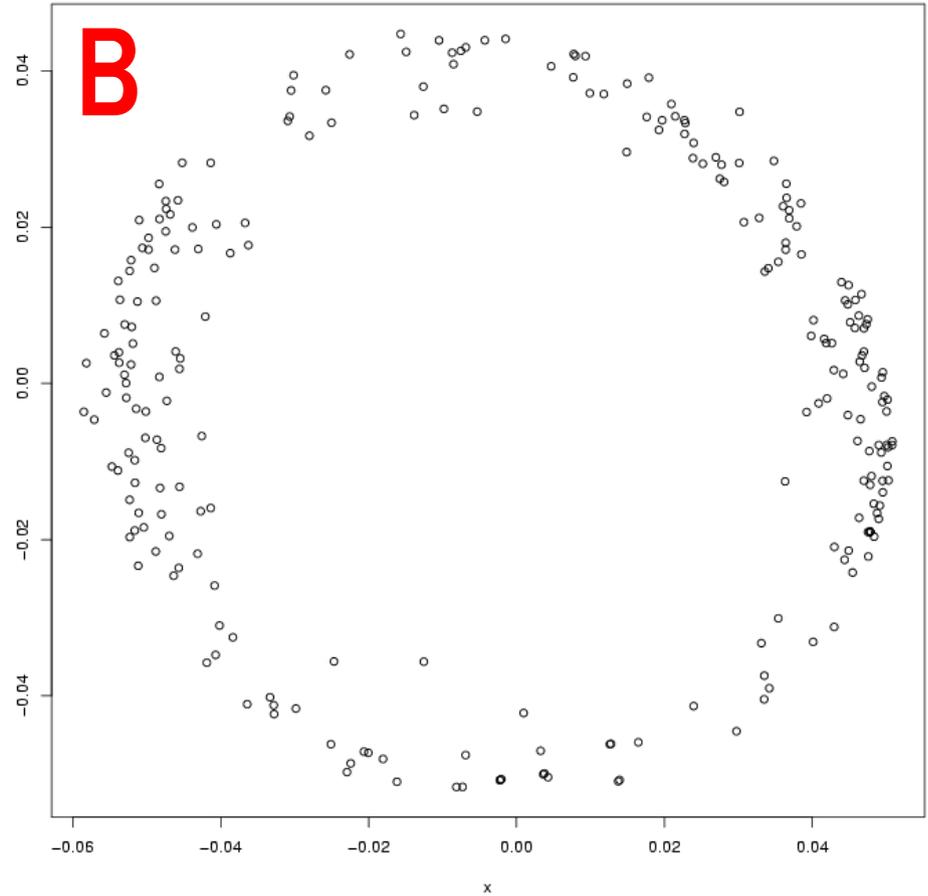
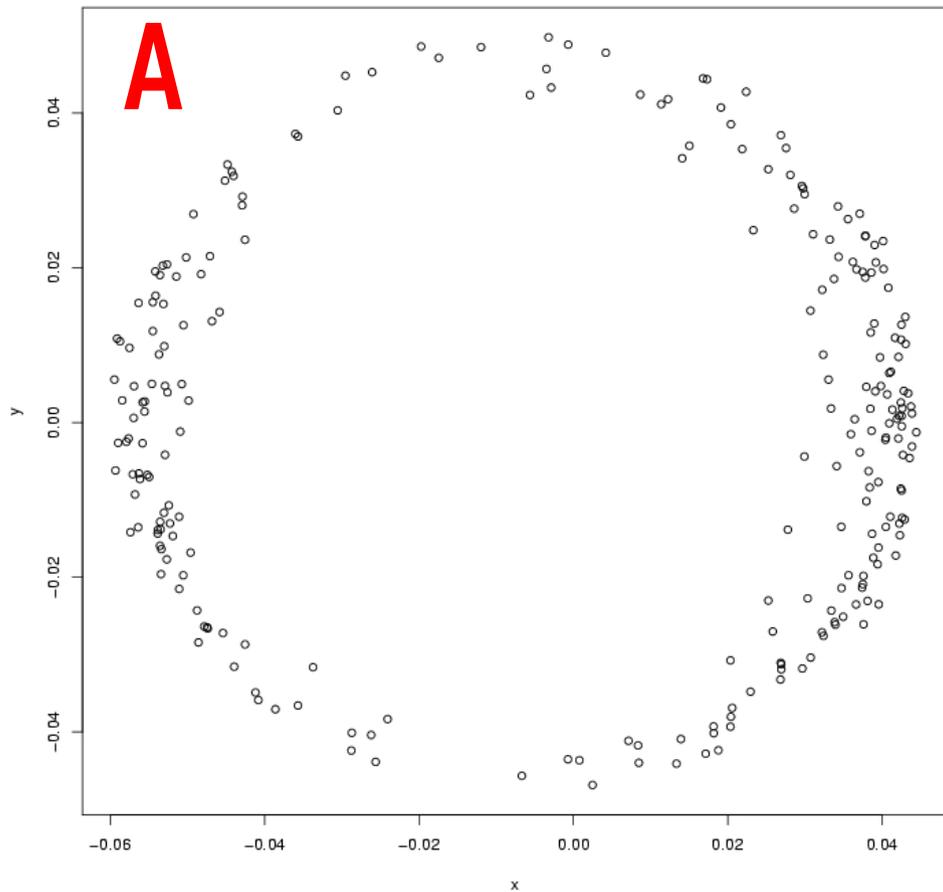
遺伝子  $i_1$  の埋め込みのよさ:

$i_1$  : 固定、 $i_2 = 1, \dots, N$

$$\begin{aligned} & \dots < s(i_1, i_2) < \dots \\ & \dots < r(i_1, i_2) < \dots \end{aligned}$$

$i_1$  と他の遺伝子との非類似度  $s(i_1, i_2)$  と埋め込み空間での距離  $r(i_1, i_2)$  の大小関係が一致していないものを「埋め込みが悪い」とする。

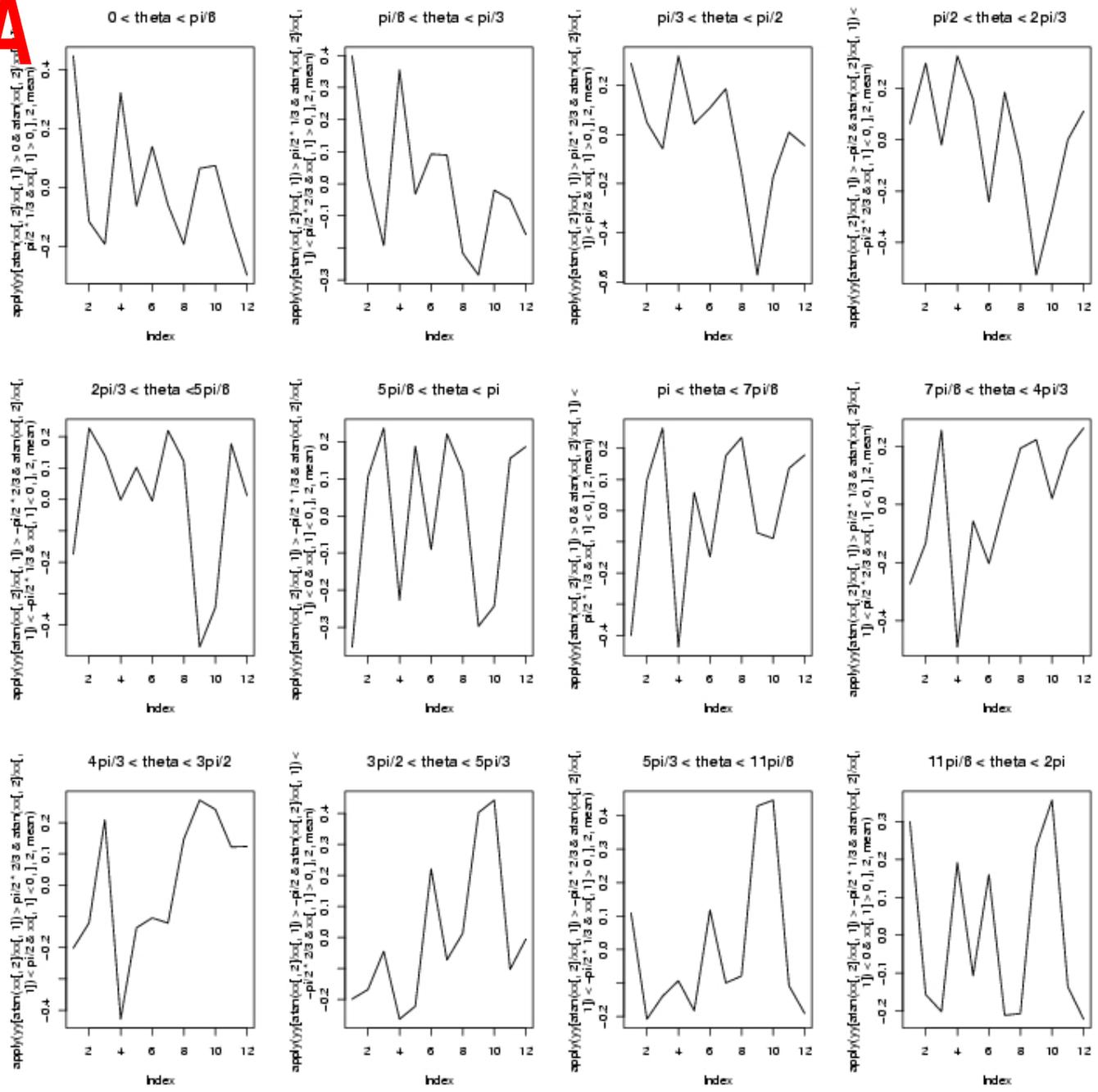
# 埋め込みの悪いものを半分取り除くと ...



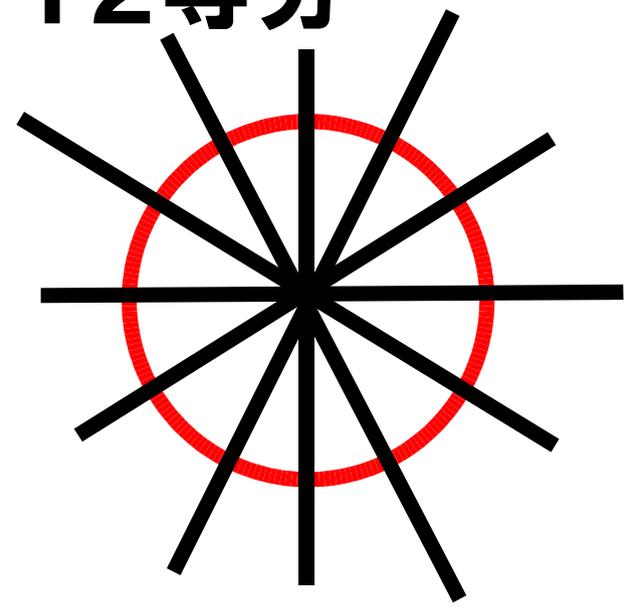
再び、「リング」が出現する。

# プロファイルの意味

A



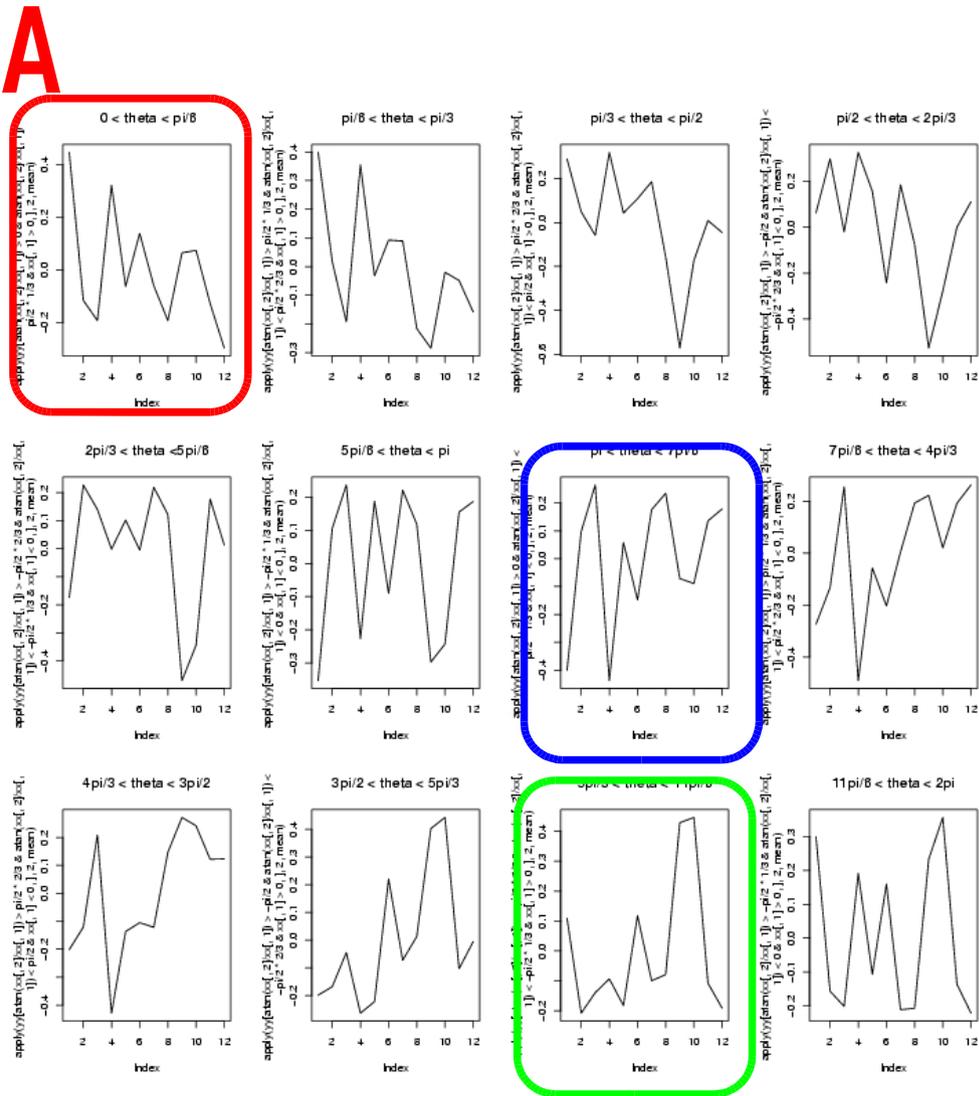
## 12等分



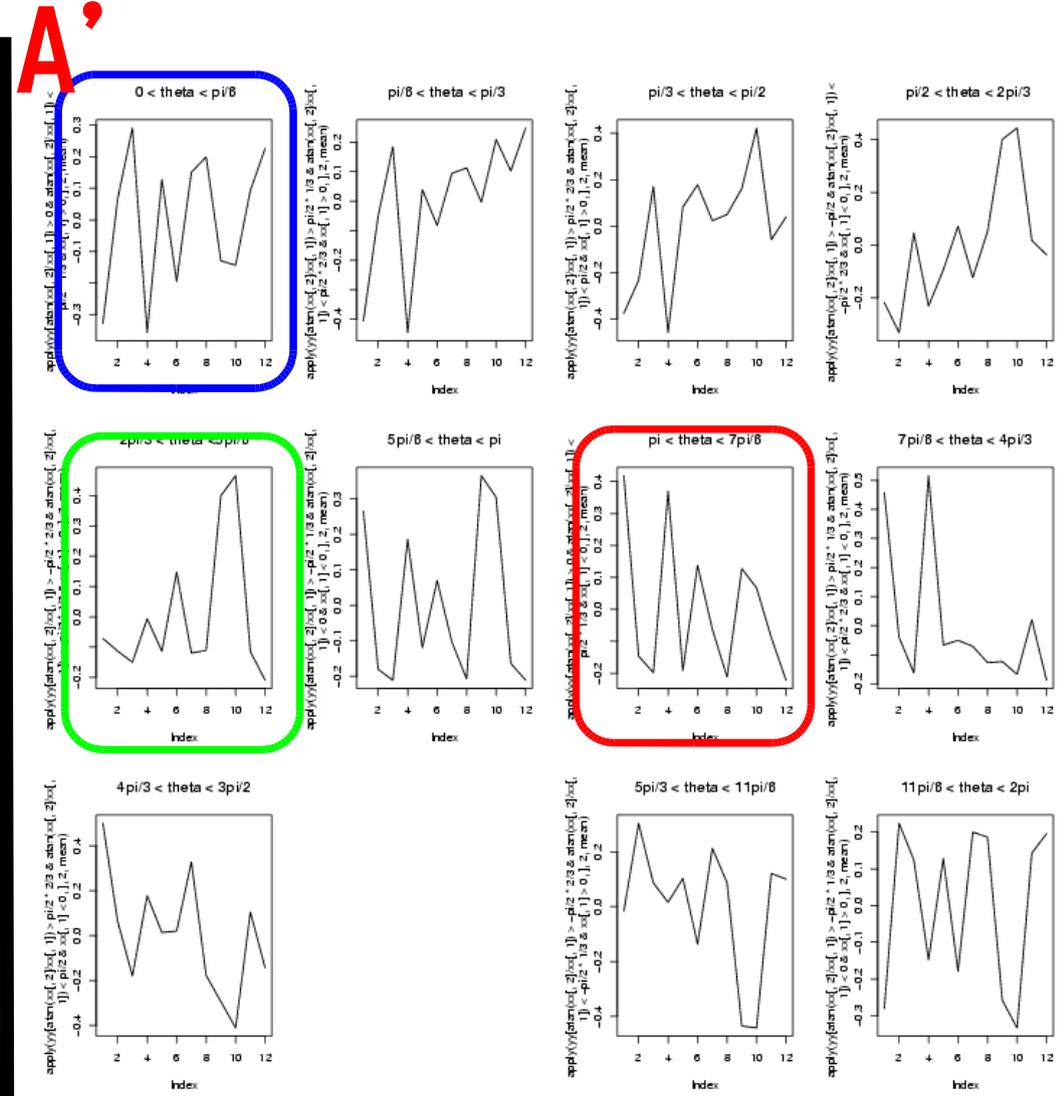
## 各領域ごとの平均プロファイル

乱数に見えるが  
乱数ではない。

# 2つのランダムセットの比較:よく一致している

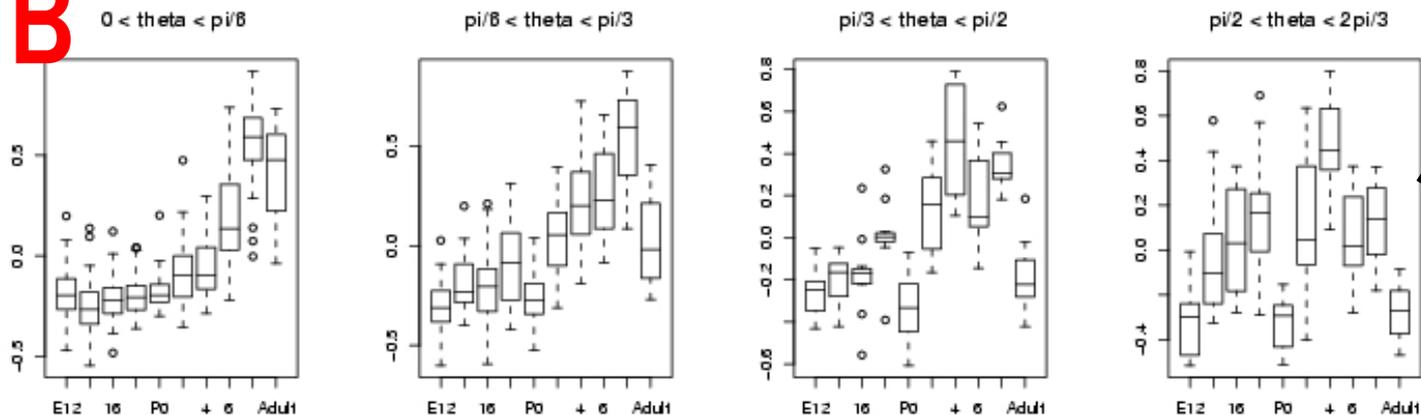


ランダムセット1

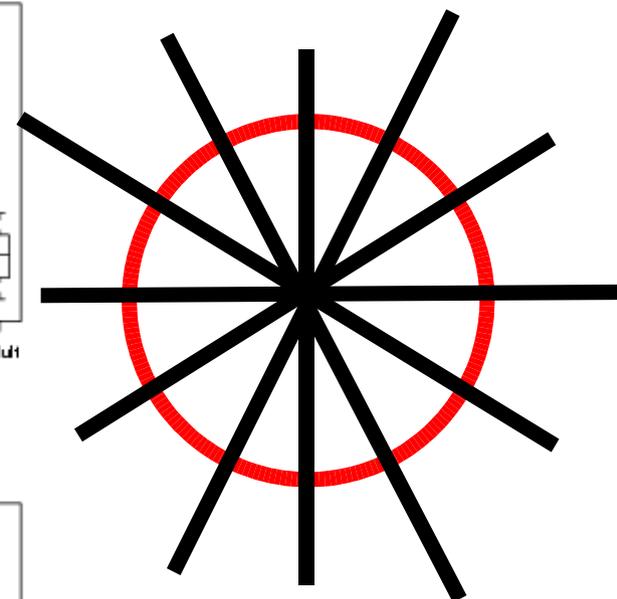


ランダムセット2

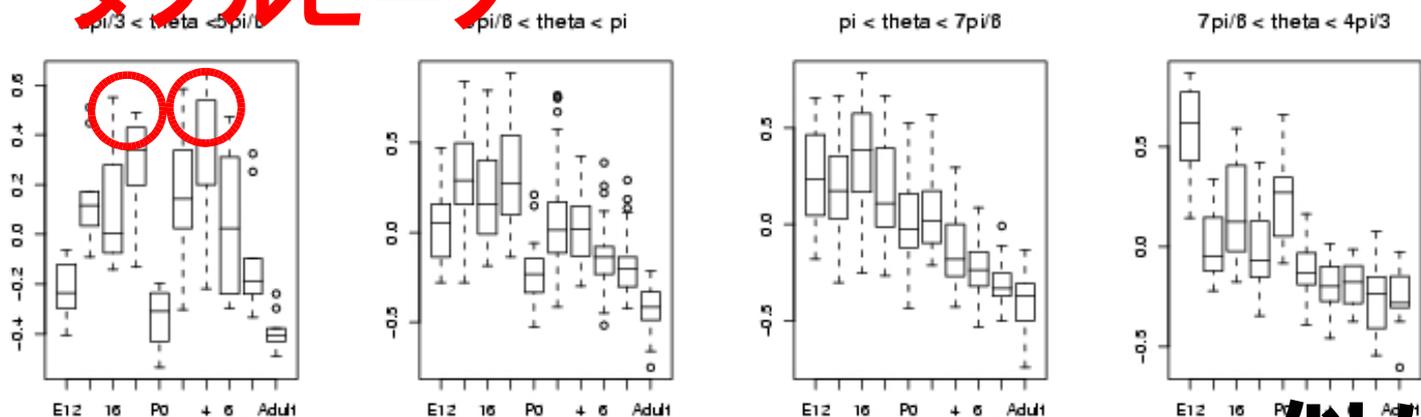
**B**



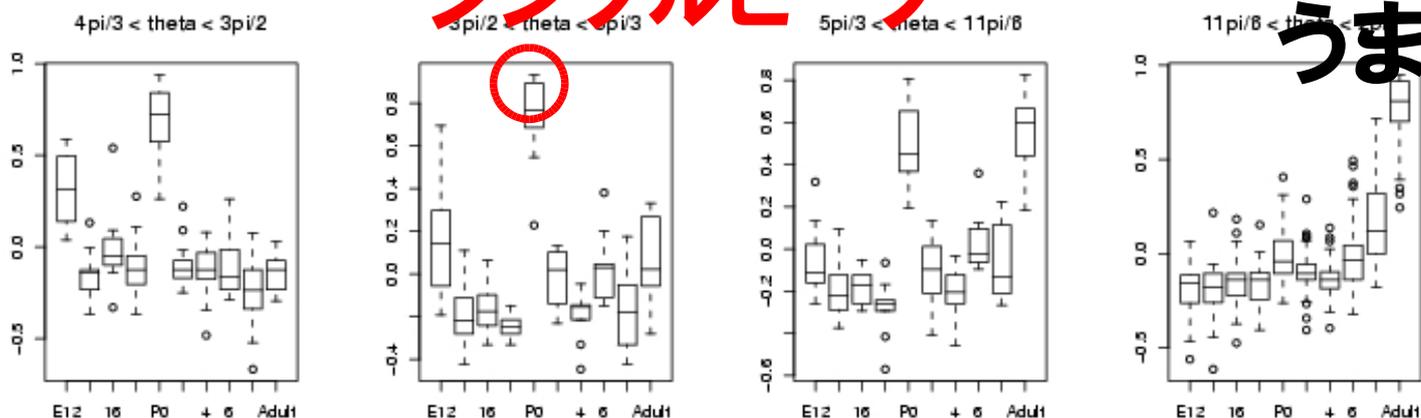
**12等分**



**ダブルピーク**

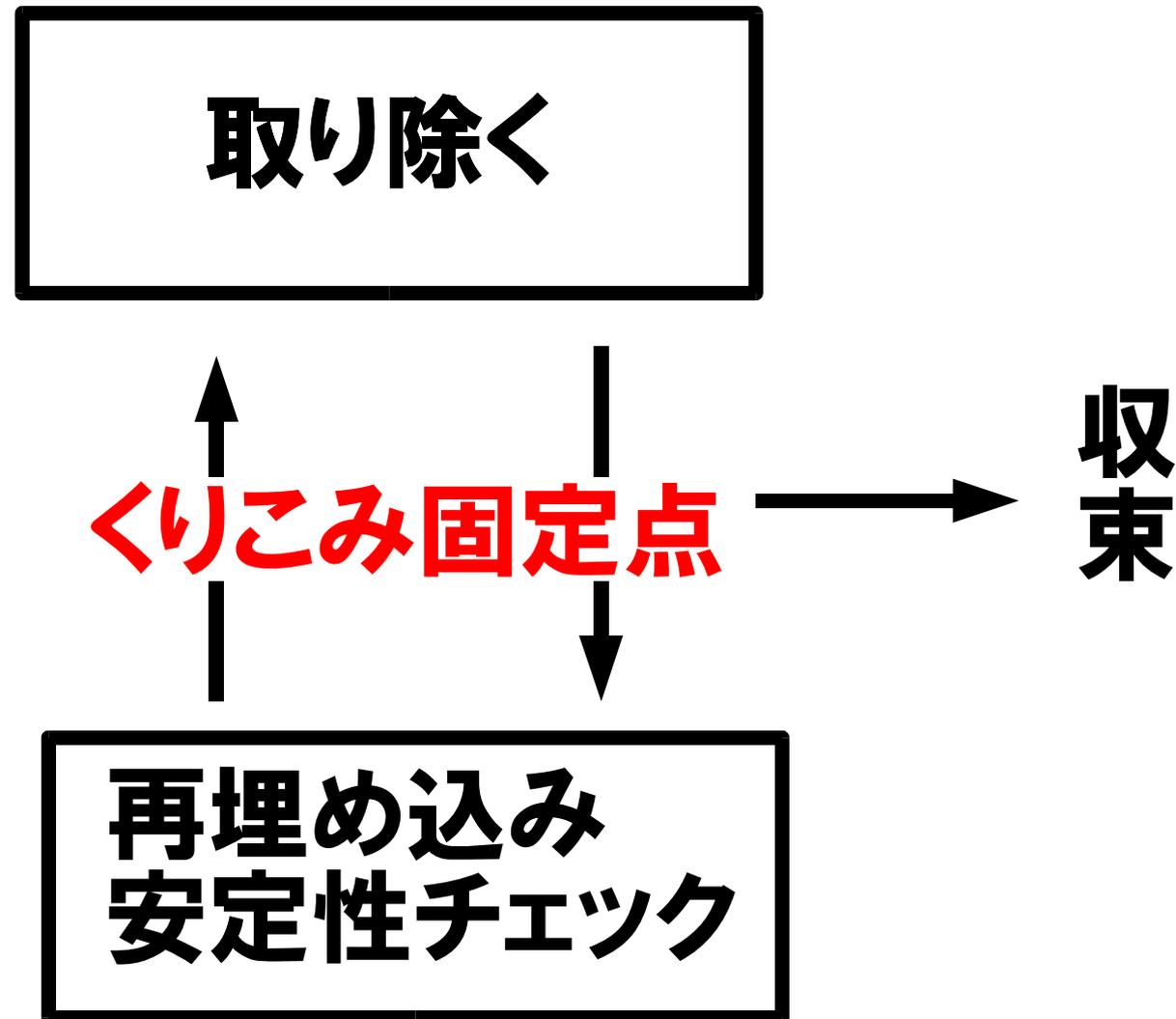


**シングルピーク**



**似たプロファイルを  
うまく集めている**

実際には「いくつ遺伝子を捨てるか？」が問題



## 4. まとめ

- **遺伝子発現実験を nMDS で解析した。**
- **実験と無関係な遺伝子を「埋め込みの悪さ」で除去したところ、リングが残った。**
- **リングに沿った遺伝子プロファイルには乱数とは異なる明確なパターンが見られた。**
- **除去遺伝子数を安定性(くりこみ群)から決めるのが今後の課題である。**