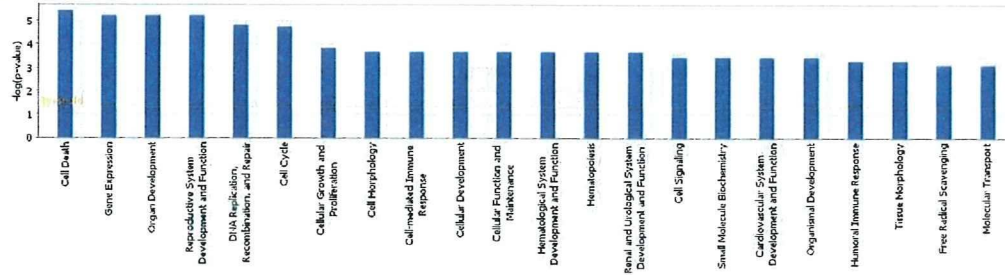


Analysis: Re-Selected 3MC only 0317 - 2010-03-17 05:28 午 巳

■ Re-Selected 3MC only 0317 - 2010-03-17 05:25 午 巳



© 2000-2010 Ingenuity Systems, Inc. All rights reserved.

国立医薬品食品衛生研究所 毒性部 御中



# 委託研究報告書 (STEP8)

## 同期発現インフォマティクス研究



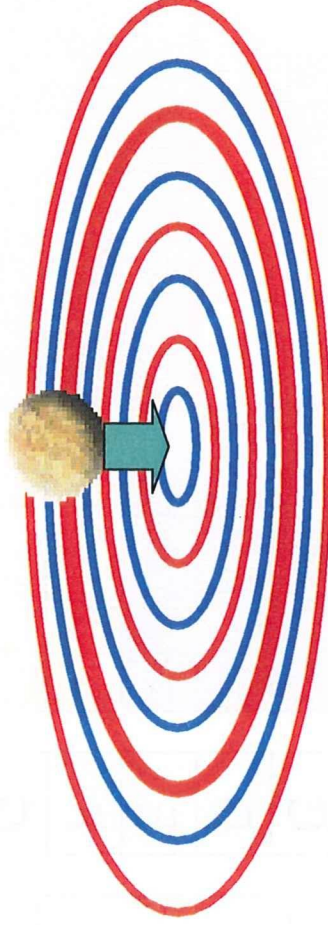
# 1. 目的

平成17年度において、クラスタリング（化合物影響により似たような反応を示す遺伝子群=クラスタを抽出する技術）を用いて、暴露する化合物によって異なる反応パターンを示しているも、高率に同じクラスタを形成している遺伝子群(同期発現遺伝子群)を抽出する研究を行った。

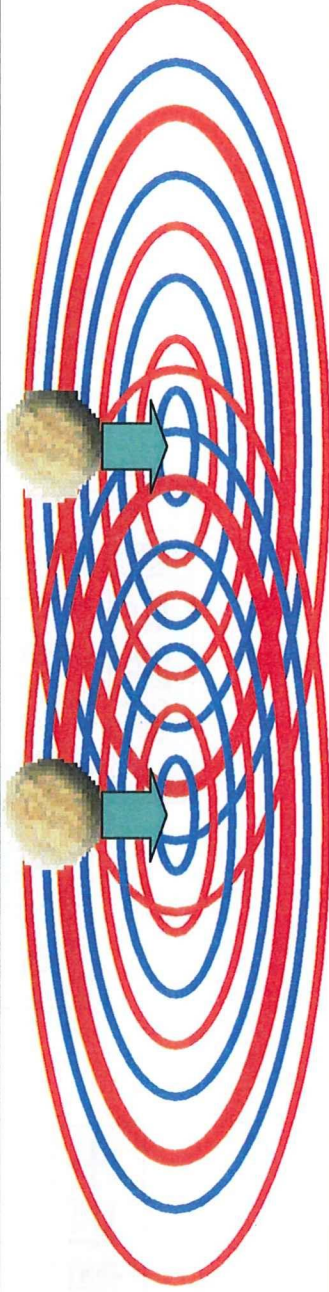
平成20年度において、半特異的結合を考慮したマイクロアレイデータの補正方法（MLANG）の基本部分を開発した。この手法を用いることにより、発現値の推定精度が向上したと期待される。  
この高精度の推定発現値を用いて、同期発現判定がどのように変更されるか調査した。

## 2. 1. 同期率計算の基本概念

池に石を投げ入れると同心円状に波紋が広がる様に、化合物を投与することにより、RNA転写への影響が遺伝子カスケード上に広がっていく。



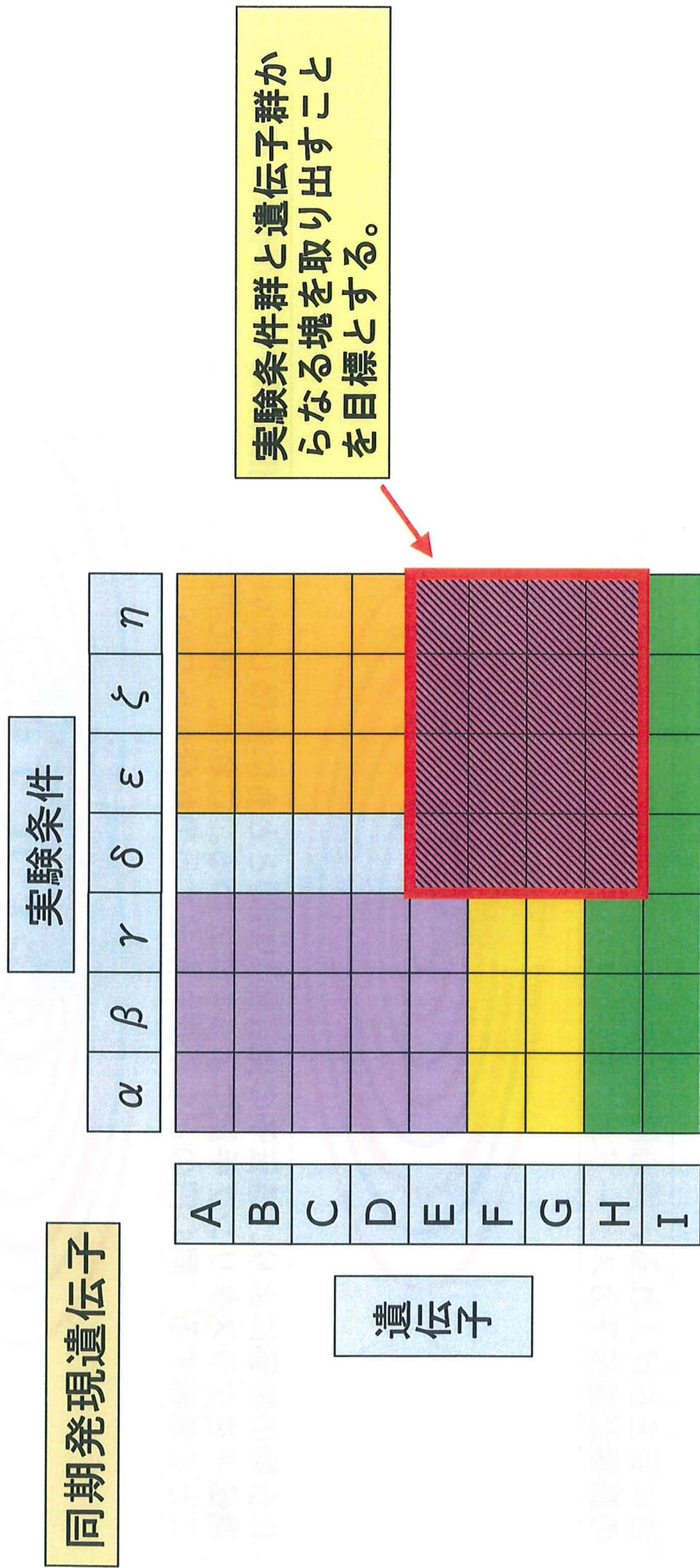
化合物の影響により、遺伝子の発現量が時間とともに変化していく。変化のパターンが似ている遺伝子をクラスタリング手法によってまとめる。これは、同心円状に存在する遺伝子をまとめることを意味する。同心円のどちら側に存在するのか分からない。



複数の化合物に対する反応から、常に同期して発現するような遺伝子を見つけ出すことにより、特定の遺伝子グループの同定が可能であると考えられる。



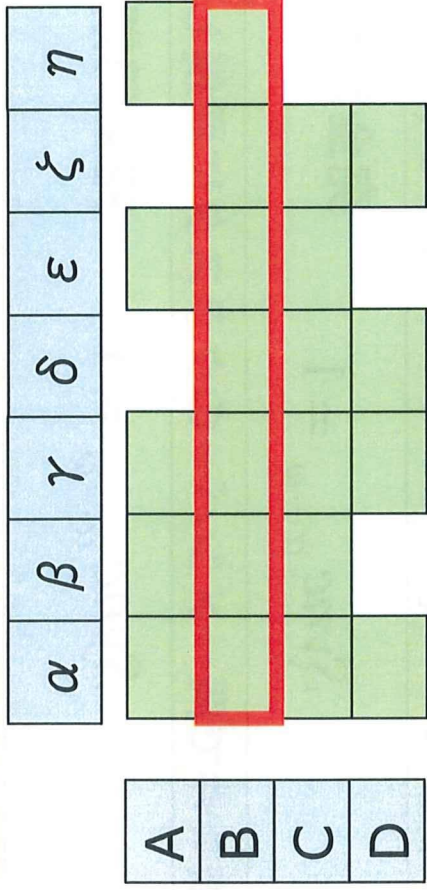
## 2. 1. 同期率計算の基本概念



単一の実験条件で、遺伝子軸方向の指標としてクラスタリングを実施できる。単一の遺伝子で実験条件軸方向でまとまりを示す指標を作成する。

## 2. 1. 同期率計算の基本概念

各遺伝子の周囲の状況を考慮し、実験条件をまたがっても、周囲の状況の変化が少ない場合に、大きな値を示す数値を作成する。



(1) 値域：0以上1以下  
同期しているときみなせる場合に1

計算の取り扱いがしやすい

(2) 性質 実験条件を増やした場合に低下する

$$0 \leq \text{Sync}_{\alpha\beta\gamma\delta}(A) \leq \text{Sync}_{\alpha\beta\gamma}(A) \leq 1$$



## 2. 2. 持っているべき性質①

重複を含まないクラスタ結果で、性質を定義する。

(1) 全実験条件でUNIQUEなクラスタとなる場合

$$\text{Sync}_{\alpha \dots \omega} = 1 \quad \text{完全同期}$$

(2) 全実験条件で同一メンバーのクラスタに所属している場合

$$\text{Sync}_{\alpha \dots \omega} = 1 \quad \text{完全同期}$$

(3) ある実験条件でどのクラスタにも所属していない場合

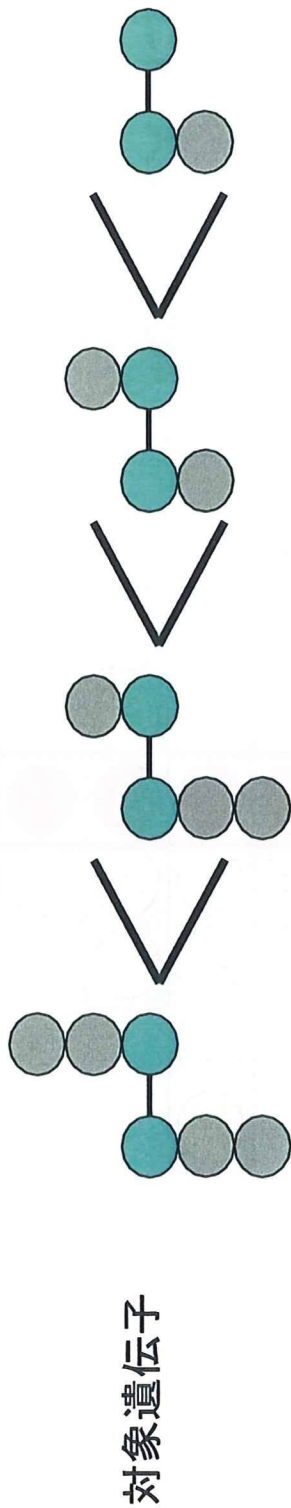
$$\text{Sync}_{\alpha \dots \omega} = 0 \quad \text{完全非同期}$$



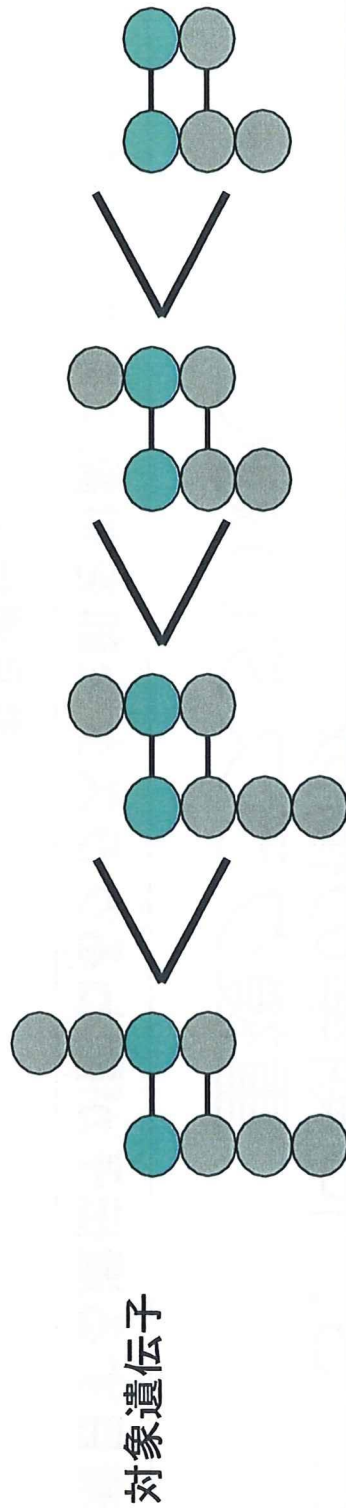
## 2. 2. 持っているべき性質②

2 実験条件間、重複を含まないクラスタの場合

(1) 対象遺伝子だけ重複

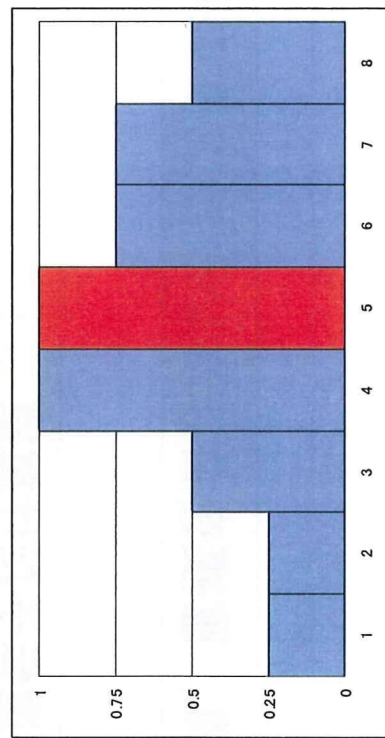
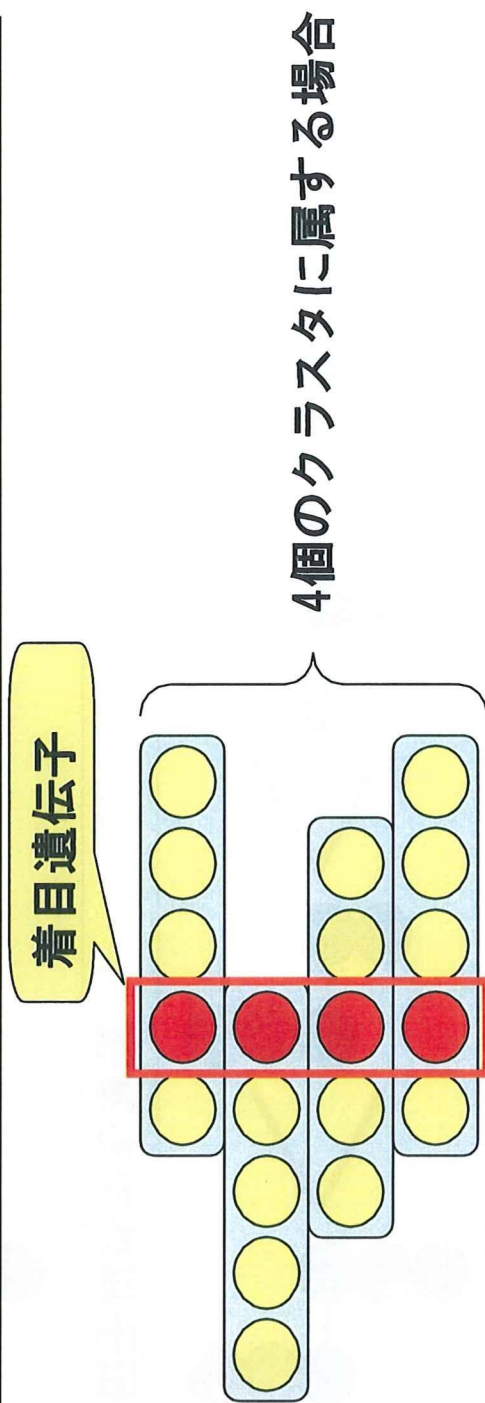


(2) 対象遺伝子以外でも重複



## 2. 3. 同期率の構成 重複クラスタの対処①

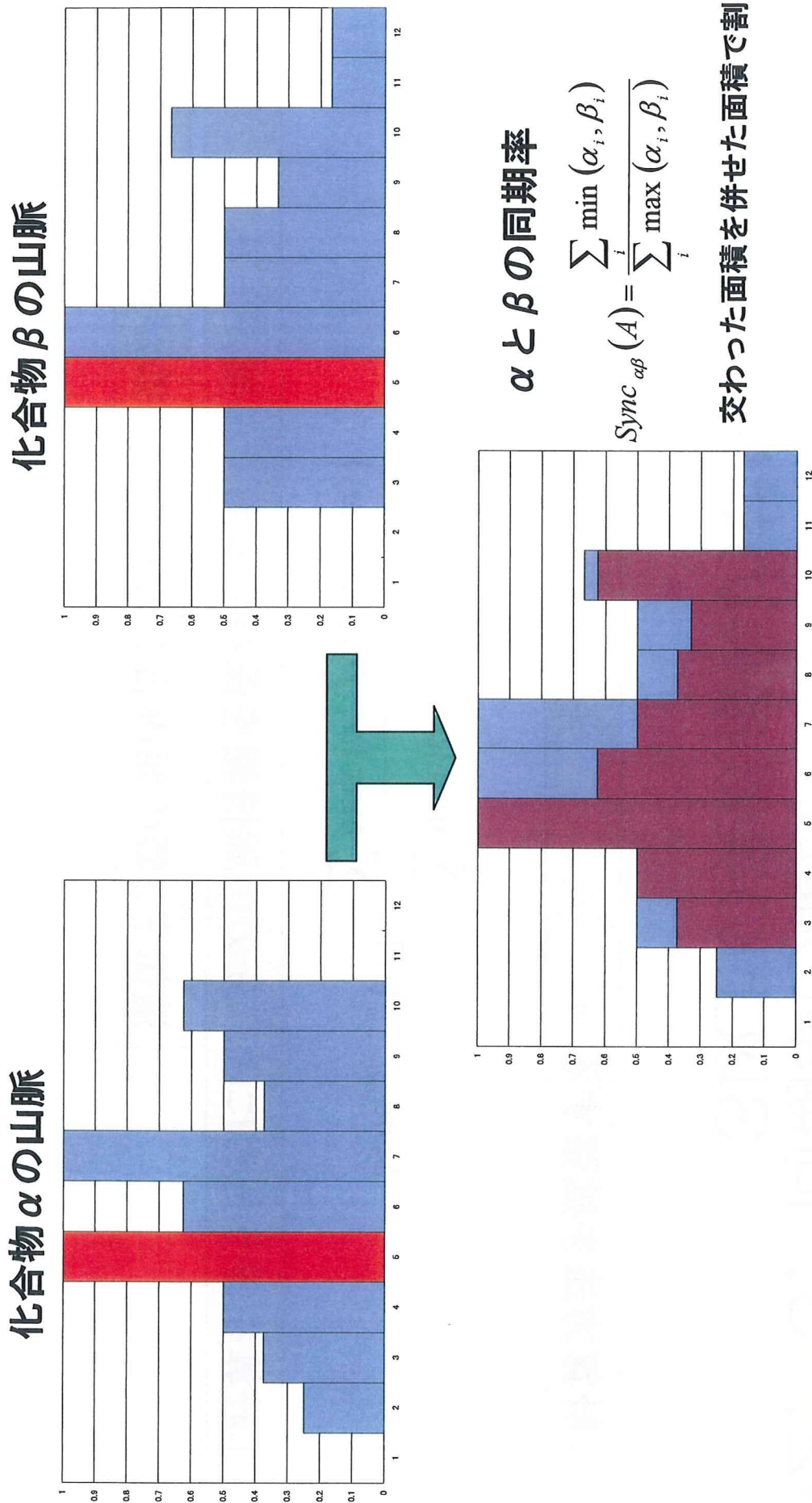
着目する遺伝子が関わるクラスタ全部を対象にして分析を行う。



属する個数の割合から、山脈のような形状が描ける

この山脈の形状が似ていれば、近くの遺伝子の存在状況が似ているといえるのではないか？

## 2. 3. 同期率の構成 重複クワスタの対処②





## 2. 3. 同期率の構成 3以上の実験条件の対処

計算方法を拡張する

$$\text{Sync}_{\alpha\beta\gamma \dots \omega}(A) = \frac{\sum_i \min(\alpha_i, \beta_i, \gamma_i, \dots, \omega_i)}{\sum_i \max(\alpha_i, \beta_i, \gamma_i, \dots, \omega_i)}$$

計算を行うためには、元の山脈情報を保持しておく必要がある

遺伝子2個の組み合わせなので、数が多い

### 3. 1. 同期率計算の改良

- ・ 初期の同期率計算では次のような問題を抱えていた
  - UNIQUEとなった遺伝子に対する同期率が他の遺伝子と比較して飛びぬけて大きな値を示す。
  - クラスタリングの際に、クラスタリング対象から外れていると同期率が計算できず、同期率の比較ができない。

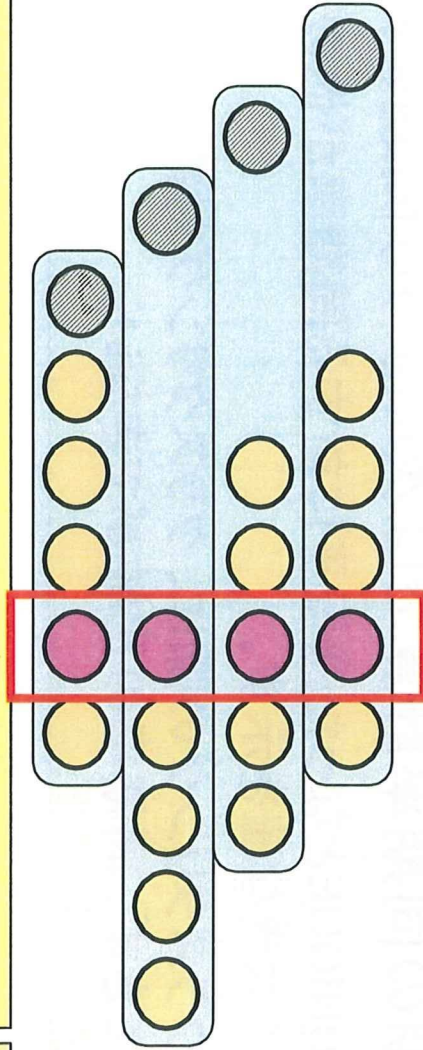


### 3. 1. 同期率計算の改良

基本の計算方法では、ユニーク遺伝子、I SOLATE 遺伝子、および、クラスタリングを実施しなかった遺伝子が突出した結果となる。

対応策

各クラスターにダミーの遺伝子がついたものとみなす。何個のダミー遺伝子を与えるかをパラメータとして指定する。

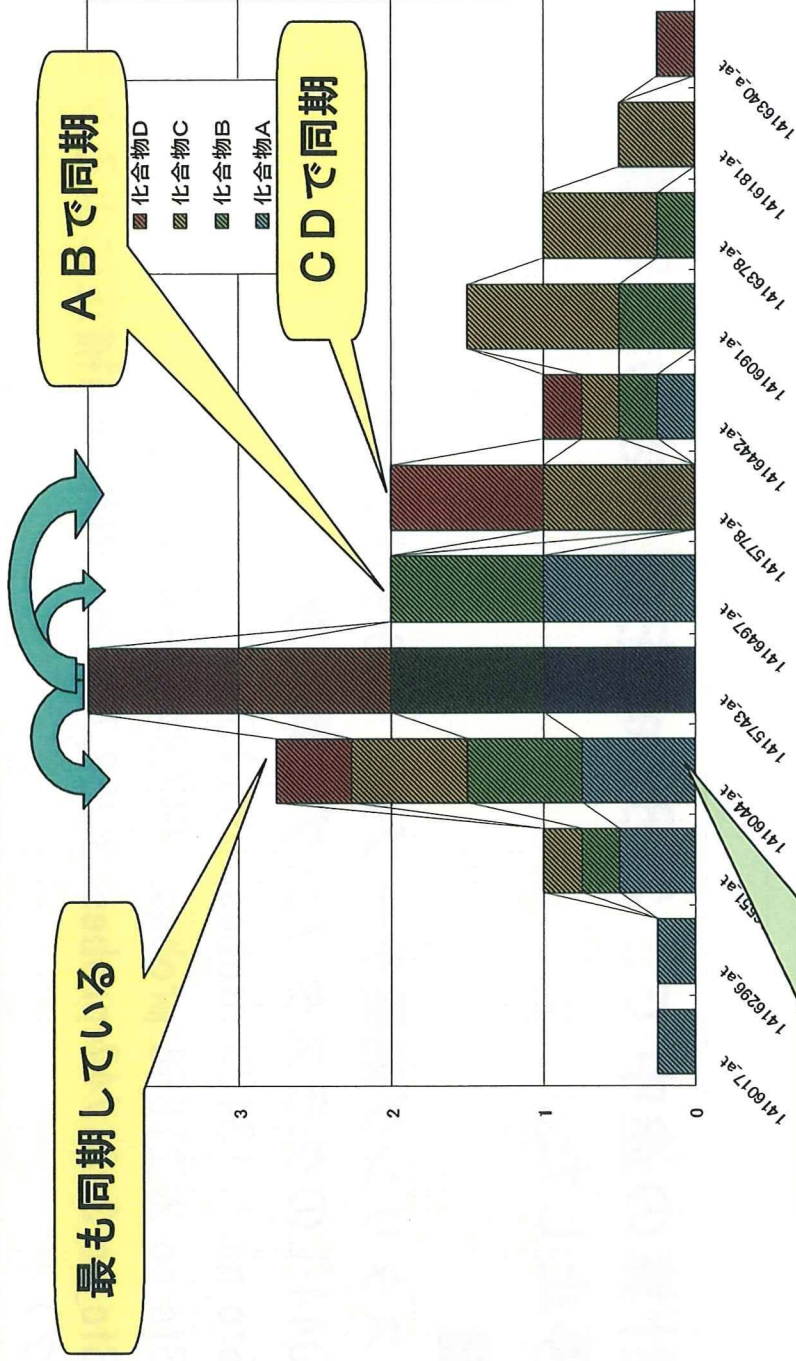


クラスタリング対象外とした化合物が存在した場合

同期率を「0」と定義する

### 3. 2. 同期率計算の改良

同期率そのものは、化合物を貫いて、同じ動きをしている遺伝子が存在するかの指標だが、どの遺伝子と同じ動きをしているか不明であった。



対象プロローブ

本手法で得られた数値は可換ではない

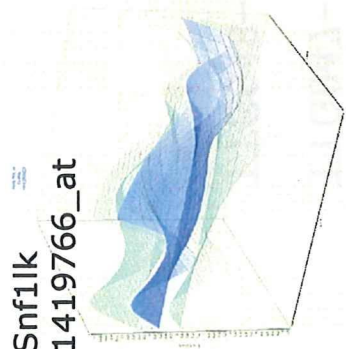
同期している対象化合物の調査は計算量の問題で困難



## 4. クラスタリング計算途中での停止

- ・ 同期率計算の途中で、2TBytesを使いきり、容量不足のため計算が停止した。
- ・ データ量
  - クラスタリング結果テーブル 249,907,396行
  - TTG044-Lのクラスタリング結果行数
    - ・ Sfc\_no 1: 13473 probes, 56,151 行
    - ・ Sfc\_no 2: 21866 probes, 147,843 行
    - ・ Sfc\_no 3: 31747 probes, 3,253,781 行 (MLANG補正によるクラスタリング)
- 特殊な結果
  - ・ RIGOROUSプローブが1315個からなるクラスタが存在する

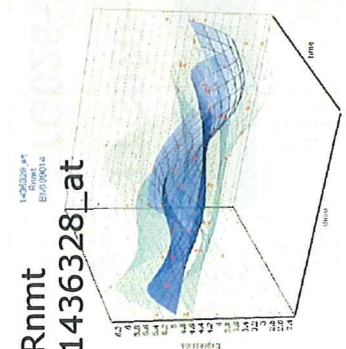
# 4. クラスタリング計算途中での停止 TTGO44-Lのクラスタ状況



Snf11k  
1419766\_at

最も多くのクラスタ(366個)に属するプローブ(1419766\_at)に対する各クラスタのRIGOROUSプローブ(近い10個)を確認する。

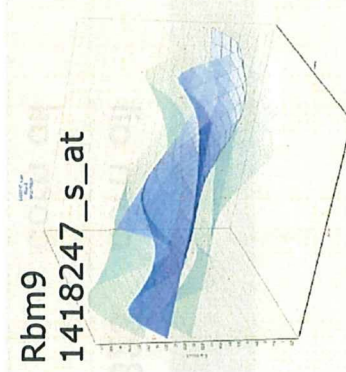
ここに挙げたRIGOROUSプローブは、別々のクラスタに所属するが、ほぼ同じ形となっている。何らかのバイアスがかかり、大きなクラスタが出来上がり、その周辺にも多くのクラスタが出来上がったものと思われる。



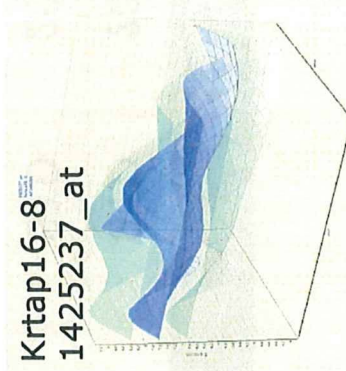
Rnmt  
1436328\_at



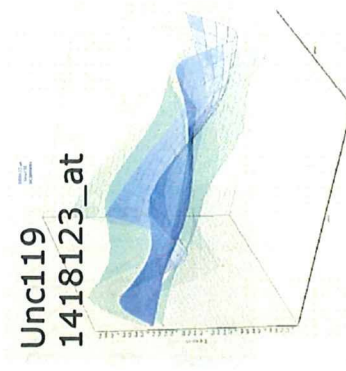
Paqr5  
1456654\_at



Rbm9  
1418247\_s\_at



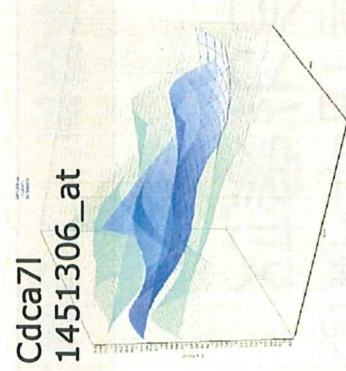
Krtap16-8  
1425237\_at



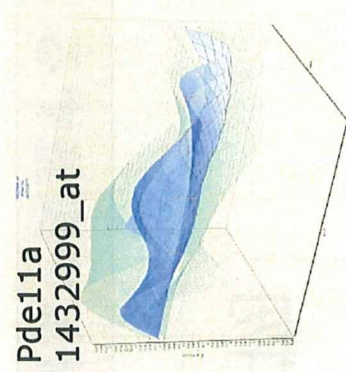
Unc119  
1418123\_at



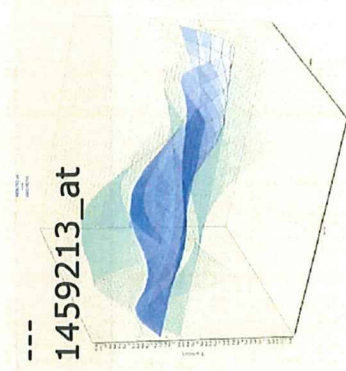
2310057JI8Rik ///  
2310065H11Rik  
1419643\_s\_at



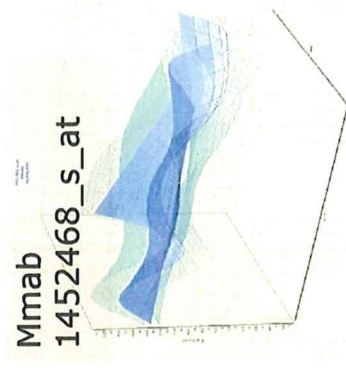
Cdca7l  
1451306\_at



Pde11a  
1432999\_at



---  
1459213\_at



Mmab  
1452468\_s\_at



## 5. 1. 同期率計算 対象化合物

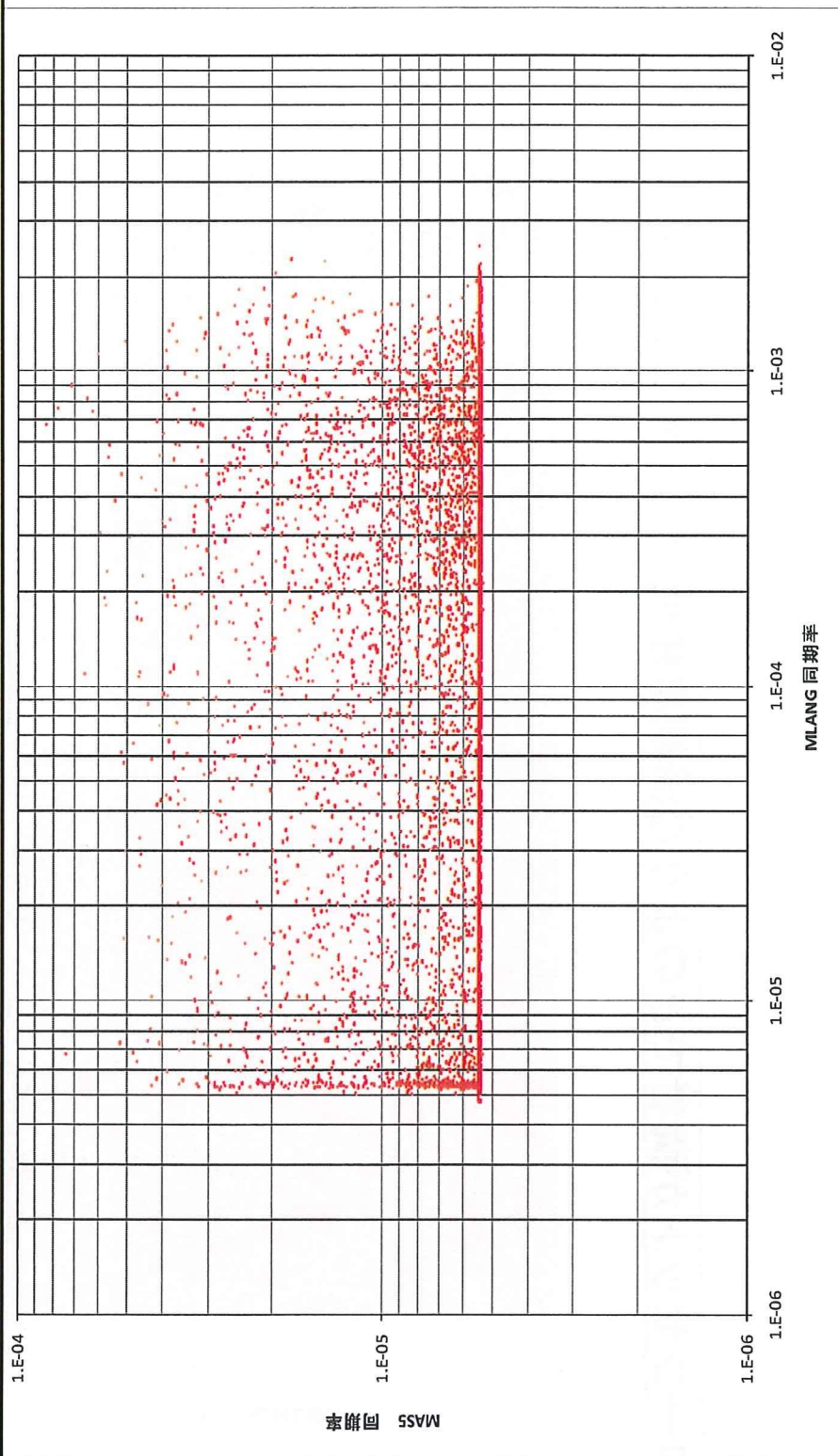
次の化合物に対して、同期率計算を実施する

化合物	溶媒	餌種類	実験日
TTG020-L TCDD_TEST20040323	corn oil	STD	2003/10/16
TTG026-L TCDF	corn oil	STD	2003/11/13
TTG047-L Bisphenol A	MC + 0.1%DMSO	PLD	2004/6/17
TTG048-L Genistein	MC + 0.1%DMSO	PLD	2004/7/1



## 5. 2. 同期率計算 結果比較（同期率値）

MAS5とMLANGを用いて推定された発現値を用いて、同期率を計算した。各プロポーザ  
セットが両者でどのような同期率となったかを示す。



MAS5とMLANGで相関がないように思われる。