

比較トランスクリプトーム向けマイクロアレイの設計手法の提案

福崎 睦美 (指導教員：瀬々 潤)

1 まえがき

本研究では、検証実験に基づく配列距離定義を用いた比較トランスクリプトーム向けマイクロアレイの設計手法を提案する。

生物学の分野では、PCRのプライマー設計など個々の遺伝子に特異的な領域を特定する問題が扱われてきた。プライマーの設計手法は数多くあるが、特異性の判定はプライマー配列が標的でない遺伝子の部分配列でないことを確認する方法が用いられている[1]。この方法は同じく遺伝子の特異的な領域探索を必要とするマイクロアレイのプロブ設計にも受け継がれているが、本研究の検証実験により、この判定方法がマイクロアレイ実験には十分でないことが明らかになった。

本研究では検証実験から得た知見により新たに特異性の定義を行い、さらにその応用として比較トランスクリプトーム用マイクロアレイの設計手法を提案する。また、実際にマウス、ラット、ヒトでの比較トランスクリプトーム用マイクロアレイの設計と実験を行い、本研究の特異性定義と提案手法の妥当性を示した。

2 研究背景

2.1 PCRとマイクロアレイ

PCRは、遺伝子の複製作用を用いて任意の遺伝子配列を増幅する手法であり、プライマー設計問題は複製作用を開始させる為に利用するプライマー(遺伝子断片)を決定する問題である。プライマー設計の条件の一つが、標的とする領域以外はプライマーと結合できないという特異性である。PCRの実験上、プライマーはその配列と遺伝子間に一カ所でも変異が存在すると結合する確率は著しく下がるため、プライマーが他の遺伝子の部分配列になっていなければ、ある程度特異性は保たれると考えられている。

このような方則は、遺伝子が機能を発揮する(発現する)頻度を観測するマイクロアレイにおいても同様に当てはまると考えられてきた。発現過程において、遺伝子はRNAに転写されるため、マイクロアレイ実験は、各遺伝子の特異的な領域(プロブ)を特定し、細胞の中にそのプロブに相当するRNAがどの程度存在するかを観測することで、各遺伝子の発現頻度を見分けている。プロブ設計では、PCRと同様に標的でない遺伝子とプロブの間いくつかの変異が存在すれば、プロブはその遺伝子に結合しないため、特異的であると判断されることが多い。しかし本研究で行った検証実験から、マイクロアレイ実験でのプロブと遺伝子の結合にはPCRと異なる性質があることがわかった。

2.2 マイクロアレイプロブの特異性検証実験

本研究では、マイクロアレイプロブの特異性の定義を検証するため、51個の変異パターンを用意し、変異のないプロブと変異を持ったプロブで観測された発現量を比較する実験を行った。

実験結果によると、同じ変異数の場合プロブの前

半に変異領域を持ったものは、後半に変異領域を持つものよりも大きく発現量が低下した。また、プロブの末尾に大量の変異を持つパターンよりも、前半に少量の変異を持つパターンの発現量への影響が大きかった。よって、単純に変異数に比例して発現量が低下するわけではなく、発現量に大きな影響を与える要因は変異位置であることがわかった。本研究では、この実験結果を考慮して比較トランスクリプトームに応用した。

2.3 比較トランスクリプトームへの応用

比較トランスクリプトーム解析は、種間で類似している遺伝子の発現頻度を網羅的に比較する研究である。この研究は、創薬研究のプロセスの短縮や発生過程の違いを生む原因の発見に役立つと考えられ、注目されている。既に網羅的な遺伝子発現量の観測に市販のマイクロアレイを使用した比較トランスクリプトーム解析[2,3]も行われているが、類似した遺伝子間にも進化の過程で様々な変異が入っており、市販されているマイクロアレイによる実験では十分な比較ができないという問題があった。

しかし、本研究で発見した発現に影響を与えない変異パターンを考慮すれば、十分な数の遺伝子間で発現量の比較が可能なマイクロアレイを設計可能となる。よって、本研究では検証実験から得た知見に則したプロブの特異性と、プロブと遺伝子が結合できる変異の条件の定義を行い、比較トランスクリプトーム解析に適したマイクロアレイの設計手法を提案する。

3 定義

本研究では発現に影響を与えない変異とプロブの特異性の定義に、変異位置を基にした重み付き編集距離を用いている。

プロブを長さ60の配列 $p = p_0p_1\dots p_{59}$ 、遺伝子を長さ n の配列 $g = g_0g_1\dots g_{n-1}$ としたとき、本研究の重み付き編集距離 $D(p, g)$ の定義は以下の通りである。

$$D(p, g) = \min\{d(p_{59}, g_j)\} \text{ where } 59 \leq j < n-1$$
$$d(p_i, g_j) = \min \begin{cases} 0 & (i < 0) \\ \sum_{k=0}^i w(k) & (j < 0) \\ d(p_{i-1}, g_j) + w(i) \\ d(p_i, g_{j-1}) + w(i) \\ d(p_{i-1}, g_{j-1}) + w(i) & (p_i \neq g_j) \\ d(p_{i-1}, g_{j-1}) & (p_i = g_j) \end{cases}$$
$$w(i) = \begin{cases} 10 & (i < 30) \\ 5 & (30 \leq i < 40) \\ 1 & (40 \leq i) \end{cases}$$

さらに、本研究では p が g の発現量観測に利用できる条件を

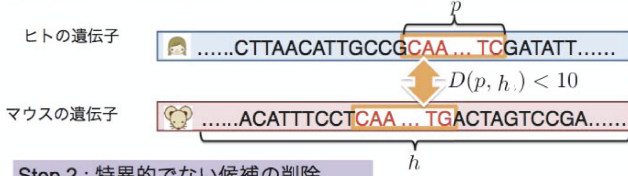
$$D(p, g) < 10$$

とし、逆に p が g に結合できない条件を

$$D(p, g) > 20$$

と定義した。

Step 1 : 類似した遺伝子間で、プローブ候補領域を見つける



Step 2 : 特異的でない候補の削除

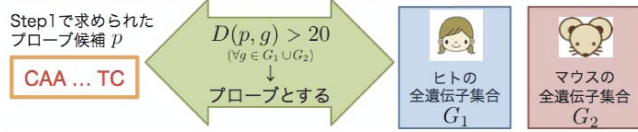


図 1: 提案手法概要

4 提案手法

ある m 種間で共通に使用できるマイクロアレイのプローブ設計手法を以下に示す。

図 1 は提案手法の手順をヒト-マウス間で実行する場合の例を示している。提案手法は 2 つのステップにより構成されており、ステップ 1 では類似した遺伝子のグループ中の全遺伝子に共通して結合できる配列を探索する。 H を互いに類似した遺伝子のグループとし、一つの遺伝子グループに対して 1 プローブを設計する。ステップ 1 では、以下の条件を満たす長さ 60 の配列 p を求める。ただし、 p は、 H の要素のうち 1 遺伝子の部分配列とする。

$$D(p, h) < 10 \text{ for } \forall h \in H$$

この条件付けにより、 p は標的とする遺伝子集合 H に含まれる全遺伝子に結合できる配列となるので、プローブの候補とする。ステップ 2 ではステップ 1 で求めたプローブ候補配列 p について、標的とする遺伝子グループに特異的であるかの判定を行う。 $G = \text{種 1 の全遺伝子集合} \cup \text{種 2 の全遺伝子集合} \cup \dots \cup \text{種 } m \text{ の全遺伝子集合}$ としたとき、プローブが特異的であるとする条件は以下の通りである。

$$D(p, g) > 20 \text{ for } \forall g \in G - H$$

この式が満たされなかった場合、もう一度ステップ 1 に戻り、別のプローブ候補を探索する。以上の 2 ステップの条件を満たす p を、遺伝子グループ H に共通して使用できるプローブとする。

5 実データでの実験と考察

本研究では、提案手法の妥当性を考えるため、実際にマウス-ラット-ヒトの 3 種に共通して使用できる比較トランスクリプトーム用マイクロアレイを設計し、実験を行った。遺伝子配列は NCBI RefSeq から、類似した遺伝子のデータは NCBI HomoloGene から取得した [4]。

まず、変異を許容せずにプローブを設計した場合と、発現に影響を与えない変異を考慮した提案手法で設計した場合とで、設計可能なプローブ数に差が出るかを検証した結果を表 1 に示す。変異を許容しないプローブ設計では、設計できたプローブ数は 2,280 個であり、入力した類似遺伝子のグループ数の 16% 程度だったのに対し、提案手法で発現に影響のない変異を許容したプローブ設計を行った場合、6,683 個のプローブ(全体の約 47%) を作成することができた。さらに、3 種共通でプローブを設計できない場合には 2 種間にのみ

表 1: 設計プローブ数の比較

	使用できる種	プローブ数 (全体からの割合)
変異を許容しない	マウス, ラット, ヒト	2,280(16%)
提案手法	マウス, ラット, ヒト	6,683(47%)
	マウス, ヒト	1,400(9.8%)
	マウス, ラット	4,705(33%)
	ラット, ヒト	109(0.7%)

共通して使用できるプローブを設計し、もう 1 種は種特異的なプローブを設計したところ、提案手法では入力した類似遺伝子グループの 90% 以上に対して、2 種以上に共通して使用できるプローブを設計可能であった。これは、比較トランスクリプトームの目的である網羅的な遺伝子発現量比較に非常に有効である。

さらに、提案手法で設計したアレイを用いて実験を行った結果、プローブが特異性を維持しつつ、類似遺伝子間でのみ発現に影響のない変異を許容しているかを検証した。検証では、2 種間で同じ細胞を使ってマイクロアレイ実験を行い、(1) 2 種で共通して使用できるプローブが、2 種間で同程度の発現になっているか (2) 1 種特異的に設計したプローブの標的とする種での発現量が、標的でない種での発現量より有意に高いかの 2 点を確認した。検証には、ヒト-ラットのアストロサイトでの発現量比較実験とマウス-ラットの神経細胞での発現量比較実験の結果を用いた。結果、種間で共通したプローブの発現差は 0 に近く、種特異的なプローブは標的とする種での発現量が有意に高いことがわかった。よって、本研究のプローブ特異性定義とマイクロアレイ設計手法は比較トランスクリプトーム解析に有効であることが示された。

6 あとがき

本論文では、実験結果から得た検証結果を基にマイクロアレイ実験におけるプローブの特異性定義を行い、それを活かした比較トランスクリプトーム向けマイクロアレイの設計手法を提案した。今後の課題として、プローブ作成時の計算時間の削減や、さらに実験結果に則したプローブの設計を行えるよう定義の改良を行っていきたい。

謝辞

本研究を進めるにあたり、貴重なデータと数多くのご助言を賜りましたお茶の水女子大学アカデミック・プロダクション特任助教小倉淳先生と特任研究員吉田真明さんに感謝致します。

参考文献

- [1] Dan G. *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*; Cambridge University Press, 1997, pp. 178-180
- [2] Su AI *et al.*, A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA.* 2004; 101:6062:6067.
- [3] Wan J *et al.*, A method for cross-species gene expression analysis with high-density oligonucleotide arrays, *Nucleic Acids Research.* 2004; 32:e93.
- [4] K.D. Pruitt and D.R. Maglott. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Research.* 2001, Vol. 29(1), pp. 137-140