

委託研究報告書 (STEP11)
マイクロアレイデータ補正技術におけるハイブリダイゼーション液中のcRNA濃度影響の検討

平成25年3月1日

株式会社NTTデータ

Copyright(C)2012-2013 NTT DATA Corporation

テーマ

マイクロアレイ補正技術(MLANG)は、ターゲット・プローブの全組み合わせにおける結合離脱特性を係数として求め、それらの係数を用いることによりマイクロアレイ測定結果の補正を行う。マイクロアレイ実験においてcRNA濃度により、結合離脱特性が変化する可能性が見いだされた。結合離脱特性に変化を与える要因の検討と、cRNA濃度によりどの程度の影響を受けるかを把握し、補正方法の検討を行う。

Copyright(C)2012-2013 NTT DATA Corporation

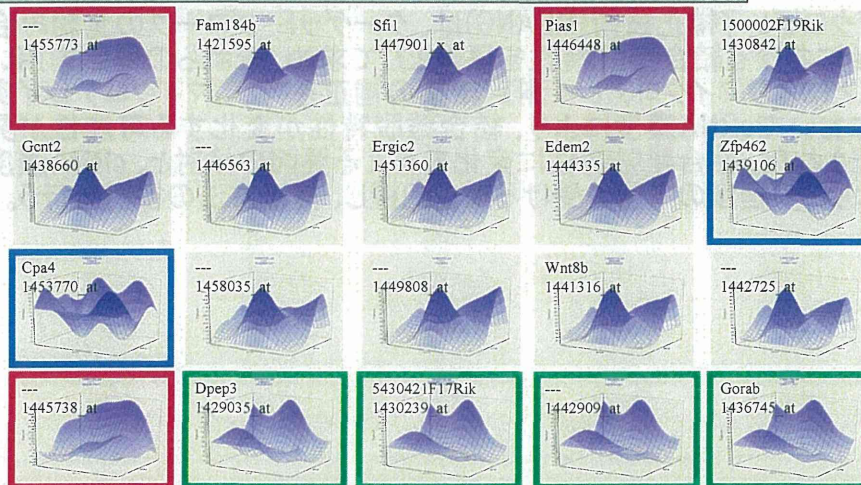
1.1. 先行研究における課題

平成23年度研究において、MLANG補正結果を用いてクラスタリングを実施した。その中で、複雑な形状にも関わらず、多くの遺伝子を含む、異常なクラスタが形成された。

低発現のプローブセットにおいて、高発現のプローブセットの反転形状が多数存在するなど、係数不足と考えられる状況がみつき、MLANG係数の学習過程で、係数を追加することを実施した。係数の追加をランダムで行ったため、望ましくない局所解に陥ったものと考えられる。

1.1. 先行研究における課題

プロジェクトTTG020-L(TCDD)に対して、クラスタリングを実施した例を示す。



同じクラスタに所属したプローブセットを同じ色で囲んである。MAS5のサマライズ結果より高い値を示すものが多い。

1.2.係数学習の変更①

これらのクラスタを構成するプローブセットは、同じクラスタ内の発現量の大きなプローブセットとの係数が、係数学習過程で追加されていた。

このようなプローブセット間の関係を特定し、係数を取り除いた。

プローブ補正信頼性係数2が0.15以下

影響遺伝子の発現量が30コピー以上

このような場合に、大きな影響を受けているとみなして、係数を削除した。

しかし、別の問題が見つかった

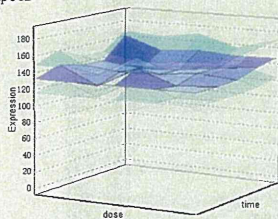
変動の少ない遺伝子の間で関係がある場合には、分離できず、一方に偏る。

1.3.係数学習の変更②

係数学習の変更案の検討

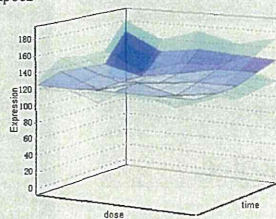
MAS5による補正

TTG020-L_SpNC_019009_at
Apoc2



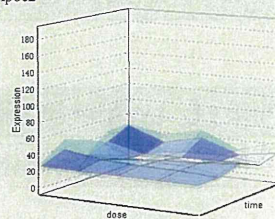
MLANG初期係数による補正

perclin_TTG020-L_29000_t
Apoc2



MLANG学習163週目による補正

perclin_TTG020-L_16300_t
Apoc2



初期係数(連続一致塩基長による係数付与)は、かなり正確で、学習を進めたことによる局所解より優れている。

初期係数から、10回程度学習を行うだけでよいと考えられる。→初期学習において大きく変動する係数を取り除き、再度、学習する。

2.cRNA濃度影響の課題

マイクロアレイのハイブリダイゼーション液の調整の際に、検体(測定対象のマウスRNA)を含めずにGSC(Percellome用スパイクカクテル)とAffymetrixの標準プロトコルの添加RNAだけで、実験を行ったところ、以下のように通常とは異なる結果となった。

- ① GSCの1番濃いRNA(Thr)よりも、2番目に濃いRNA(Lys)用のプローブの方が、明るくなった。

Thrは、Lysの3倍量が存在するにも関わらず、Lysの方が明るい。

- ② 検体が存在しないにもかかわらず、検体用のプローブの一部が、グリッド用プローブよりも明るくなった。

グリッド用RNAは、標準プロトコルの添加物として大量に添加されているにも関わらず、さらに明るいプローブが存在した。

Copyright(C)2012-2013 NTT DATA Corporation

2.1.GSC+添加RNAのみの実験で明るいプローブ

通常の検体RNAを添加せず、GSCとAffymetrix社のプロトコルの添加RNAだけで計測をした際の明るいプローブ上位30個を示す。

Thrよりも、Lysが明るい傾向にある。

プローブセットとしての特性と考えられる。

Lysのプローブ群の間に、MURINE_B2や一般のプローブがみられる。

各プローブの特性と考えられる。

No.	Prb.ID	Prb.No.	tm	Serial
1	AFFX-LysX-3 at	18	pm	7833
2	AFFX-LysX-3 at	10	pm	7793
3	AFFX-LysX-3 at	9	pm	7414
4	AFFX-r2-Bs-lys-3 at	1	pm	7320
5	AFFX-LysX-3 at	15	pm	7321
6	AFFX-MURINE B2 at	11	pm	7157
7	AFFX-LysX-3 at	19	pm	6917
8	AFFX-MURINE B2 at	10	pm	6514
9	AFFX-LysX-3 at	17	pm	6459
10	AFFX-LysX-3 at	8	pm	6272
11	AFFX-MURINE B2 at	9	pm	6284
12	AFFX-LysX-M at	3	pm	5749
13	AFFX-LysX-3 at	16	pm	5693
14	AFFX-LysX-3 at	12	pm	5628
15	AFFX-LysX-M at	19	pm	5339
16	AFFX-LysX-3 at	14	pm	5131
17	AFFX-LysX-5 at	2	pm	5058
18	1441700 at	8	pm	5056
19	AFFX-LysX-3 at	11	pm	4946
20	AFFX-r2-Bs-lys-3 at	0	pm	4788
21	AFFX-LysX-M at	8	pm	4777
22	AFFX-LysX-3 at	7	pm	4655
23	AFFX-r2-Bs-lys-M at	6	pm	4223
24	1453723 x at	1	pm	4157
25	AFFX-Thr X-3 at	9	pm	4084
26	1435643 x at	1	pm	4038
27	AFFX-r2-Bs-thr-3 s at	4	pm	4004
28	AFFX-r2-Bs-lys-M at	8	pm	3997
29	AFFX-LysX-5 at	4	pm	3839
30	AFFX-LysX-M at	16	pm	3814

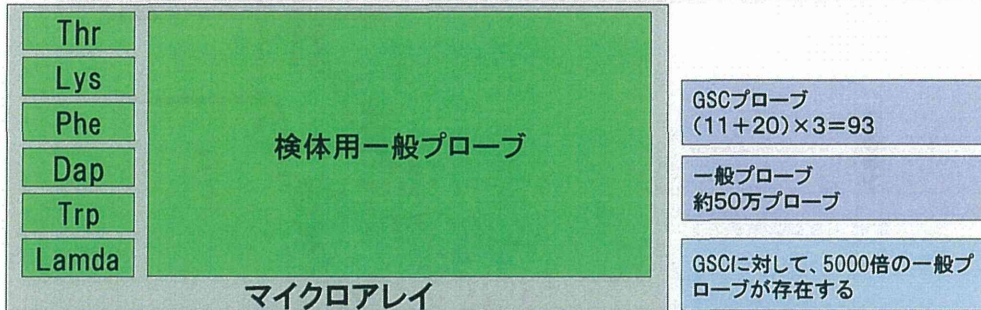
Copyright(C)2012-2013 NTT DATA Corporation

2.2.LysがThrよりも明るくなる逆転現象

基本的な疑問

Lysが明るくなったのか？ Thrが暗くなったのか？

cRNA濃度が通常より極端に低いために、Thrが暗くなる傾向にあるのではないだろうか。

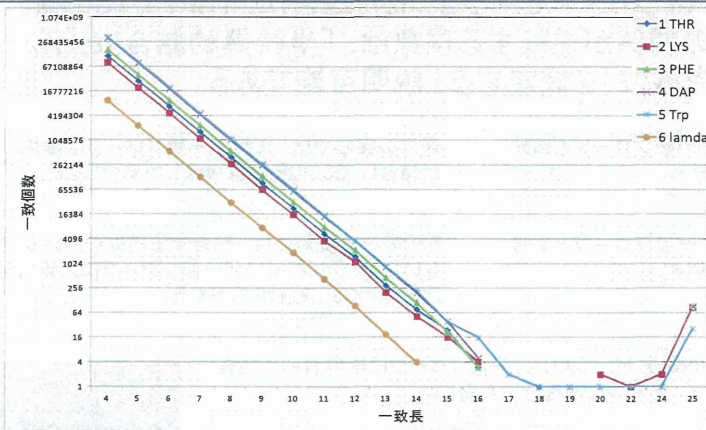


通常濃度では、検体中に多くのRNAが存在する。検体用一般プローブは、非常に短い(3-4塩基)半特異的結合は、検体のRNAとの間で多く発生している。このため、GSCが一般プローブと半特異的結合をする確率はかなり低い。しかし、検体が存在しないと、一般プローブすべてに半特異的結合する可能性が高く、GSCの見かけの濃度は減少する。溶液中のRNA濃度の有限性が問題になった可能性がある。

Copyright(C)2012-2013 NTT DATA Corporation

2.3.GSC配列が一般プローブ(PM)と半特異的に結合し得る箇所の数

GSC配列が一般プローブ(PM)と半特異的に結合し得る箇所を網羅的に検索した



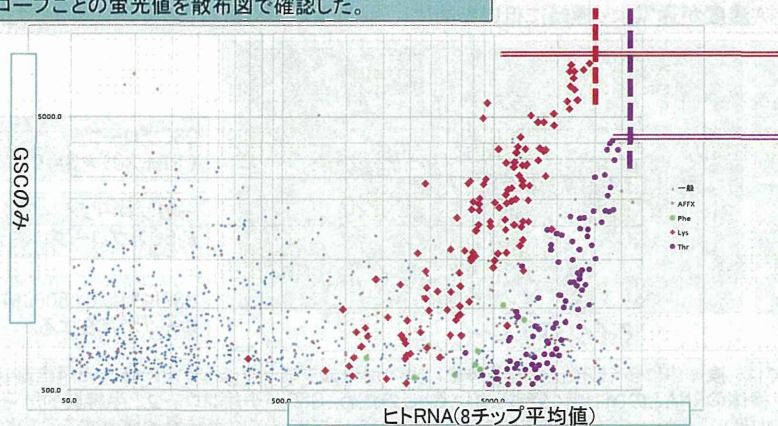
Lamda以外の5個のうち、THRは半特異的結合箇所が最も多い。LYSは半特異的結合箇所が少ない(約4倍)。THRとLYSはターゲットの長さが2倍あるので、相対的に8倍結合しやすい(奪われやすい)と考えられる。

Copyright(C)2012-2013 NTT DATA Corporation

2.4.異種検体での計測結果

マウス用マイクロアレイチップに対して、GSCのみをハイブリダイズした計測と、ヒトRNAを検体として用いた計測の結果を検討する。

プローブごとの蛍光値を散布図で確認した。



ヒトRNAが存在する場合には、THRとLYSの逆転現象が発生しなかった。

Copyright(C)2012-2013 NTT DATA Corporation

2.5.THRとLYSの逆転現象の発生原因

GSCのみをハイブリダイゼーションした場合に、THRとLYSの蛍光値と濃度の関係が逆転する現象は、「半特異的結合」と「溶液中RNAの有限性」を仮定すると説明可能である。

THRはLYSよりも、一般プローブと部分一致塩基配列が多い。

通常の実験では、大半の一般プローブは検体のRNAが結合し、GSCのRNAは少量しか結合しない。

THRはLYSよりも、一般プローブと半特異的結合を起こしやすい。

GSCの実験では、結合する検体のRNAが存在しないため、GSCのRNAが一般プローブに大量に結合し、溶液中のGSC濃度が低下する。

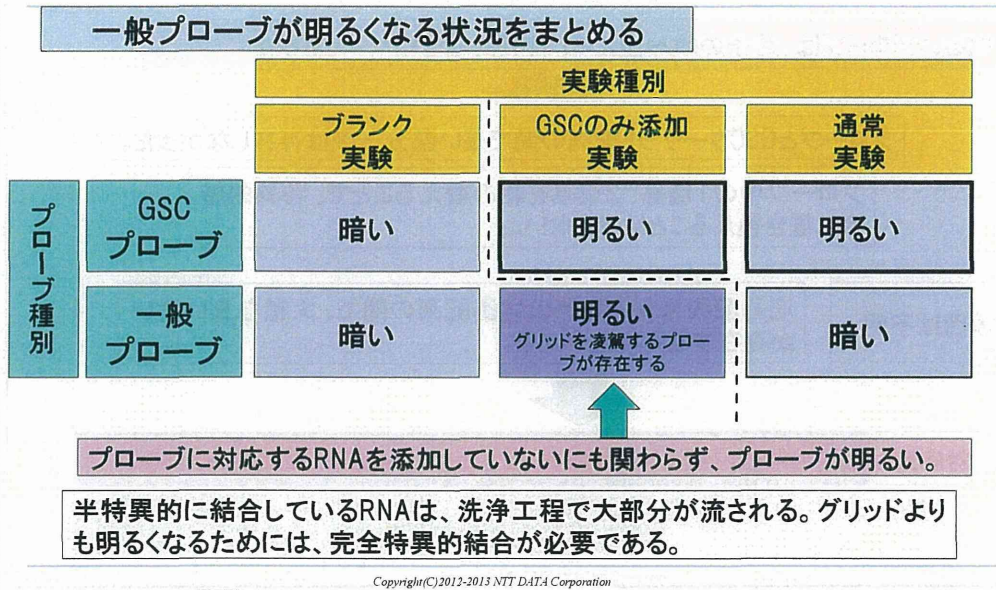
THRの方がLYSよりも、一般プローブに半特異的に結合し易く、溶液中の濃度低下が著しい。

定性的には正しいが、影響を与えるほど大きいかを定量的に調べるには、ハイブリダイゼーションステップ中の半特異的結合の量を計測する必要があり、困難である。

THRの方がLYSよりも、溶液中の濃度が低下し、蛍光値が小さくなる。

Copyright(C)2012-2013 NTT DATA Corporation

2.6.一般プローブが明るくなる可能性



2.7.一般プローブが明るくなる原因の検討

No	Prb_ID	PrbNo	pm	Stad
1	1441700_at	8	pm	5056
2	1453723_x_at	1	pm	4157
3	1435643_x_at	1	pm	4038
4	1435643_x_at	2	pm	3558
5	1453723_x_at	2	pm	3490
6	1420465_s_at	5	pm	3397
7	1415969_s_at	10	pm	3272
8	1420465_s_at	4	pm	3149
9	1415969_s_at	4	pm	3045
10	1449434_at	3	pm	2993
11	1415969_s_at	3	pm	2829
12	1436722_s_at	8	pm	2819
13	1420465_s_at	3	pm	2800
14	1415964_at	9	pm	2762
15	1420465_s_at	6	pm	2763
16	1449106_at	6	pm	2692
17	1415968_s_at	1	pm	2548
18	1425267_a_at	4	pm	2525
19	1435643_x_at	5	pm	2501
20	1420465_s_at	7	pm	2449
21	1415968_s_at	0	pm	2390
22	1417867_at	9	pm	2367
23	1417867_at	5	pm	2360
24	1449434_at	1	pm	2351
25	1420465_s_at	8	pm	2336
26	1418055_at	10	pm	2335
27	1420465_s_at	9	pm	2305
28	1449436_s_at	4	pm	2330
29	1449434_at	8	pm	2208
30	1424695_at	7	pm	2188

一般のミスマッチプローブの明るさを
超えている

大過剰量存在する

グリッド用配列だけが、大過
剰量存在する。

グリッド用配列と短い部分一
致しか存在しない

洗い流されてしまう。

or 特異的結合

中央1塩基置換でも、大部
分は洗い流されてしまう。

プローブ全長で一致するも
のはない

洗い流されてしまう。

一致しているかの判断は、配列情報に基づいて
いる。配列情報が間違っている可能性はあるか？

2.7. 一般プローブが明るくなる原因の検討

明るいプローブは、若干の組み換えで、特異的結合になりえるか？

プローブとGSCターゲット配列の間で長い部分一致は存在しなかった。

プローブ中の1塩基、2塩基を組み替えることで、特異的結合に置き換えることはできない。

原因は不明

塩基配列情報と実際の塩基配列の間で、大幅な食い違いが存在する可能性がある。

対応策

MLANG補正係数から今回見つかったプローブ(ペア)を除去する。

2.8. 低濃度RNA計測結果のまとめ①

低濃度RNAにおける解釈が難しい点 その1

GSCのLysとThrの濃度比と蛍光強度が逆転していた。

LysはThrより、濃度は薄いのに、蛍光値は高かった。

仮説

Thrは、Lysより、一般プローブに半特異的に吸着しやすい。検体RNAが存在しない場合には、Thr RNAは一般プローブに吸着し、検知用プローブへ吸着できる量が限られる。このため、検知用プローブがあまり明るくならない。

事実

GSCと全プローブの結合を確認したところ、ThrはLysより、一般プローブと配列の部分一致が多かった。

チップとは異なるヒトのRNAを検体とし、cRNAの濃度が高い場合には、ThrとLysの逆転現象は発生しなかった。

仮説を支持する事実

2.8.低濃度RNA計測結果のまとめ②

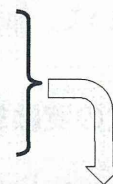
低濃度RNAにおける解釈が難しい点 その2

非常に明るい一般プローブが存在した(THRより明るい)

しかし、GSC及びグリッド用オリゴとの配列とは強い関係がなかった。

GSCの実験は1チップだけであり、偶発的な要素の可能性

Affymetrixの設計・製造工程に起因している可能性



これ以上の追及を行わない

一般的な測定においては、実質的な影響が小さいため

3.MLANG補正プログラム/ラット対応

NIHS/毒性部のPercellomeプロジェクトでは、実験動物としてマウスを用いている。これに対して、基盤研究所のトキシゲノミクスプロジェクト(TGP)では、実験動物としてラットを用いている。

MLANG補正は、マウス用チップ(MOE430v2)に特化している。ラットにも対応し、TGPの実験結果も補正できるようにする。

MLANG補正プログラムは、マウス専用として記述している箇所がある。これを、ラットにも適用可能なように拡張する。その後、ラット用の係数を作成する。

3.1. ラット用係数の作成

初期係数の作成方法

Affymetrix社提供の「ターゲット配列」と「プローブ配列」を用いて、部分一致箇所に対応した初期係数を作成する。

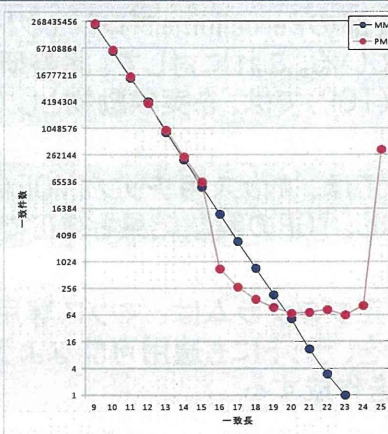
学習: 次の実験結果を用いて係数の学習を実施する。

Liver-Brain-Mixture (5種類×Duplicate)
 肝臓溶媒(4時刻×Triplicate)
 腎臓溶媒(4時刻×Triplicate)

3.2. 部分一致の分布

ターゲット配列とプローブ配列を元に部分一致配列を探索し、一致長とその件数を求めた。

一致長	MM	PM	総計
9	231,880,005	244,357,820	476,237,825
10	57,599,166	61,470,082	119,069,248
11	14,251,123	15,432,158	29,683,281
12	4,221,647	3,854,971	8,076,618
13	847,384	962,836	1,810,220
14	204,972	240,931	445,903
15	49,807	65,092	114,899
16	12,124	711	12,835
17	2,940	276	3,216
18	729	147	876
19	184	96	280
20	53	70	123
21	11	74	85
22	3	85	88
23	1	65	66
24		105	105
25		347,865	347,865



15塩基以下は、塩基長が1短くなると4倍ずつ増えていた。これは、塩基種類数に対応しており、偶然の結果と考えられる。