

3.3.最大値の検討（最大値の計算方法）

チップごとの最大値を次の手順により求める

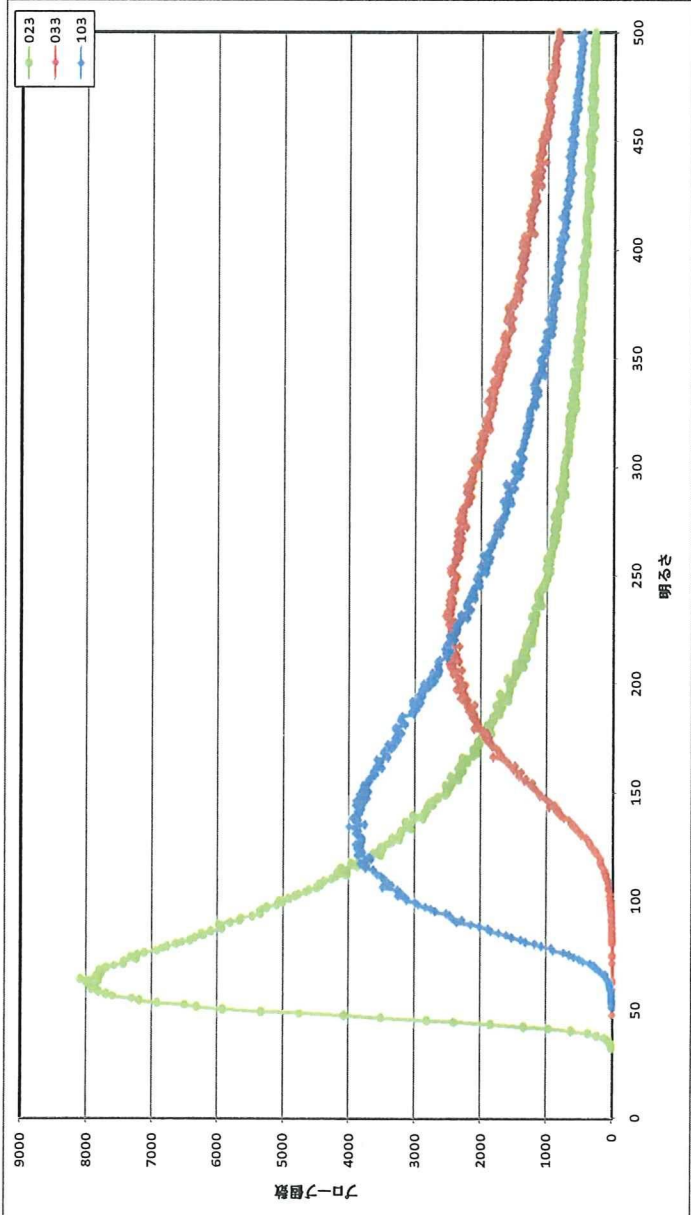
PerfectMatch(PM)を対象とする。(MisMatch(MM)が大きくなるのは、偶然とみなす)
指標を作成し、それらの指標が通常示す値と異なる程度をポイントに直し、ポイントが閾値を超えていない最大の Intensityの5%増をImaxとして設定する

複数の条件を考慮し、単一条件で偶発的な事象を排除する

4.背景しベルの推定

プローブに検体や添加RNAが結合していない状態での、プローブの明るさを、光学的な補正で「0」にできているとは考えにくい。バックグラウンドレベルを推定し、蛍光値から減算することで、各プローブのターゲット由来の真の明るさを推定する。

TTG020の典型的と
思われる分布(3チッ
プ)を示す



最も暗いプローブが、50程度の輝度である。RNAの結合と仮定すると、全体のもっとも暗いという前提なので、本来の目標RNA以外が結合し、洗浄工程でも流されなかったという事であり、考えにくい。もう少し上をバックグラウンドレベルとしてみてみるのが適切と思われる。全体プローブの5percentileをバックグラウンドレベルとみなす

5.濃度初期値の計算方法

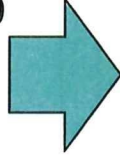
Mlangでは、濃度の初期値を与えて、収束計算を実施している。適切な濃度初期値を与えることにより、高精度で、早く収束する。

- プローブの係数を用いて、初期推定値を計算し、結果の精度高向上させる

6. 細胞あたりコピー数の計算方法

MAS5のサマライゼーション結果では、対数領域で回転することがみられ、Perccellome法では、対数領域での1次式での変換を実施してきた。この変換式は、線形領域において曲線になる。Mlangは、線形領域における曲線となる要因の多くを取り除くことができるので、この変換式は不要と思われる

$$\log S = a \log C + b$$



$$\log S = \log C + b \quad \equiv \quad S = \alpha \cdot C$$

Mlangで、補正可能なこと

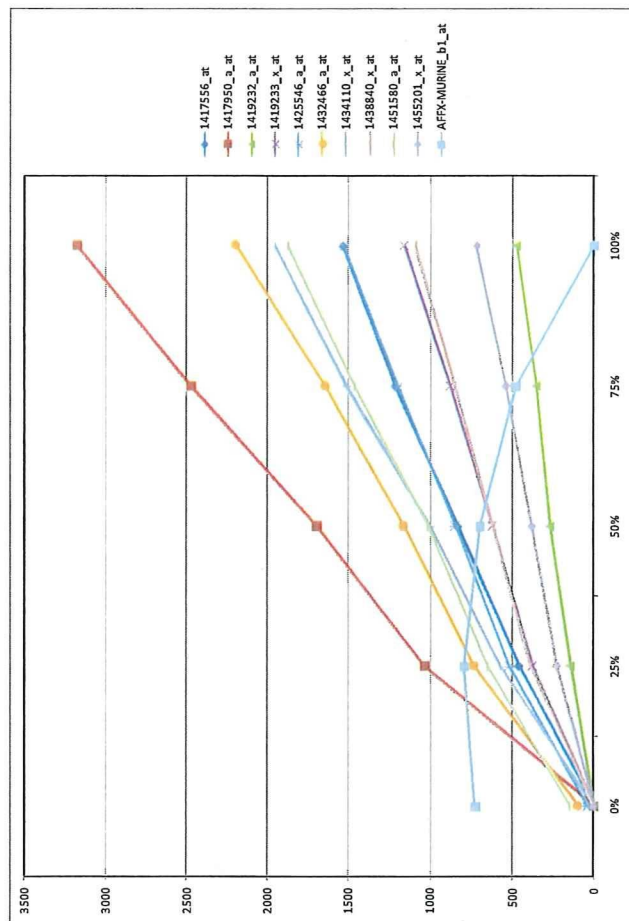
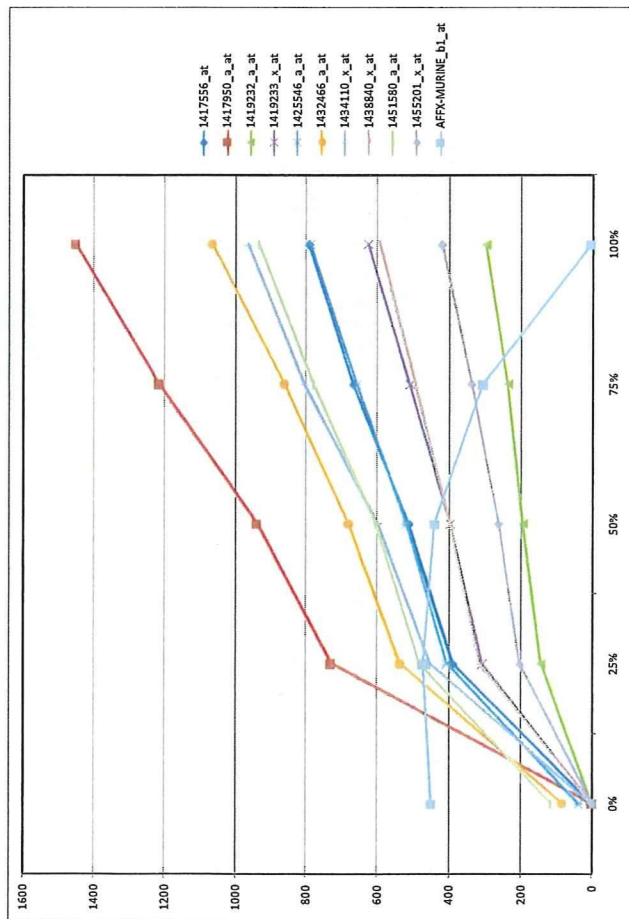
1. 飽和現象
2. 非特異的結合
半特異的な結合により、プローブと結合し、本来のターゲットだけの場合よりも明るくなる
3. 競合的結合
半特異的な結合により、先にプローブが占められ、本来のターゲットが少なくなり、半特異的結合が流されることで、蛍光が小さくなる

6.細胞あたりコピー数の計算方法

高発現遺伝子(50%:50%で、raw10000以上)

SpNC

線形



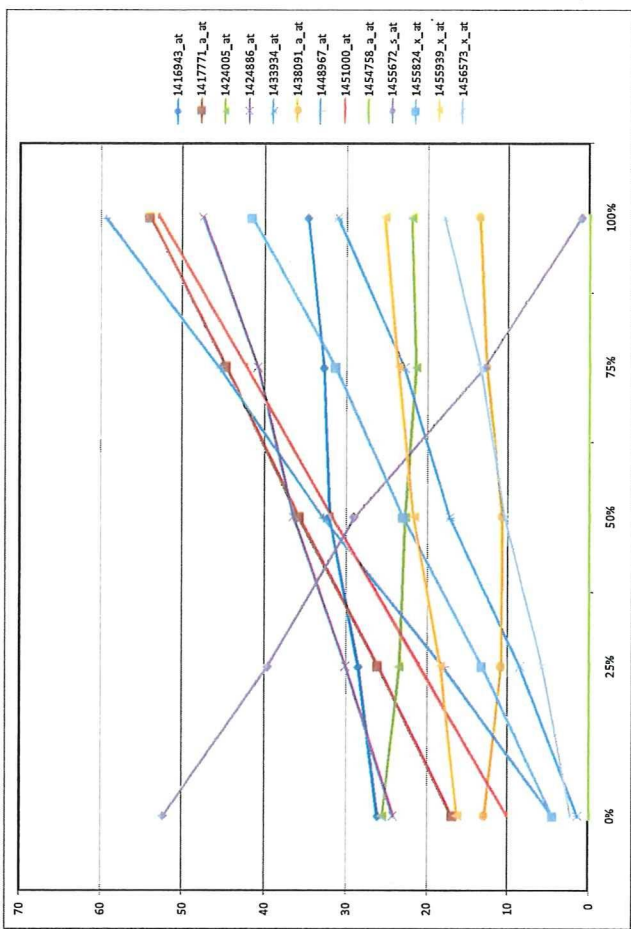
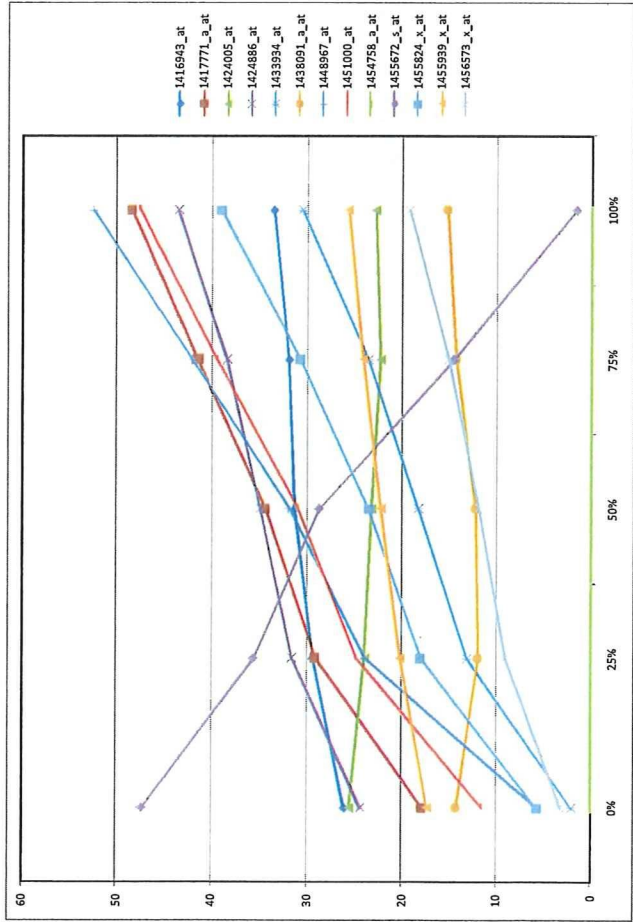
25%において、SpNC変換は、屈曲がみられる

6.細胞あたりコピー数の計算方法

中発現遺伝子(50%:50%で、raw1000程度)

SpNC

線形



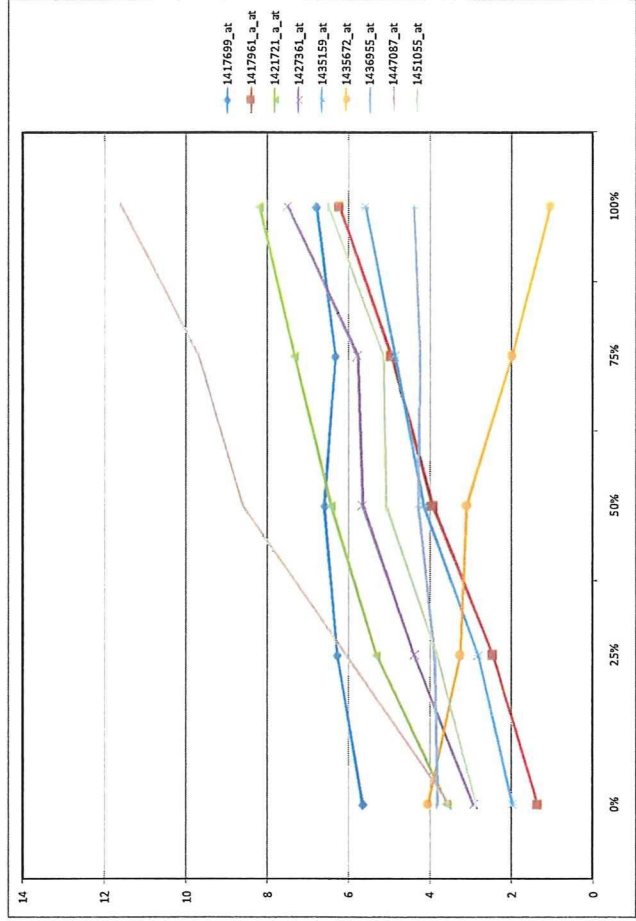
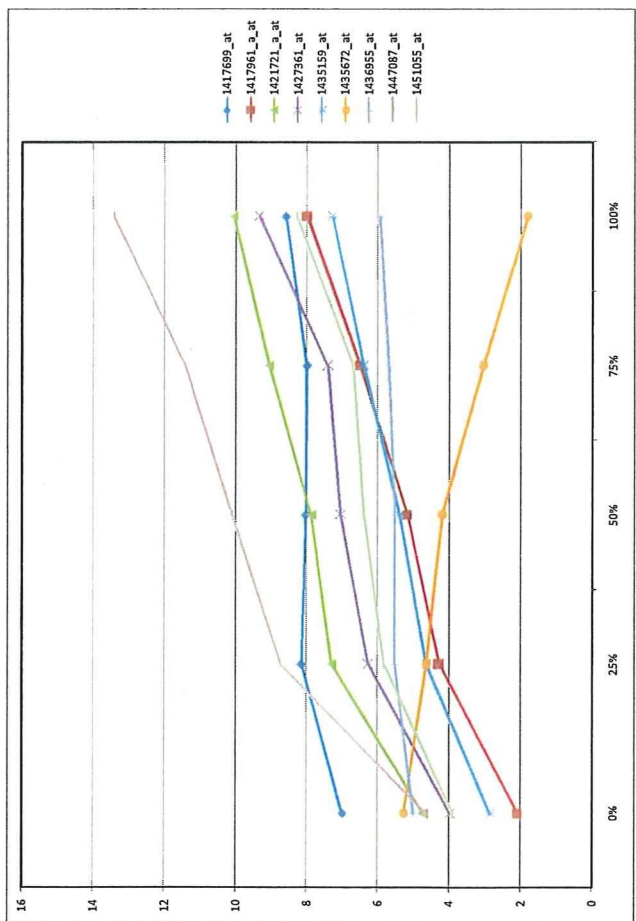
25%において、SpNC変換は、屈曲がみられる

6.細胞あたりコピー数の計算方法

低発現遺伝子(50%:50%で、raw100程度)

SpNC

線形

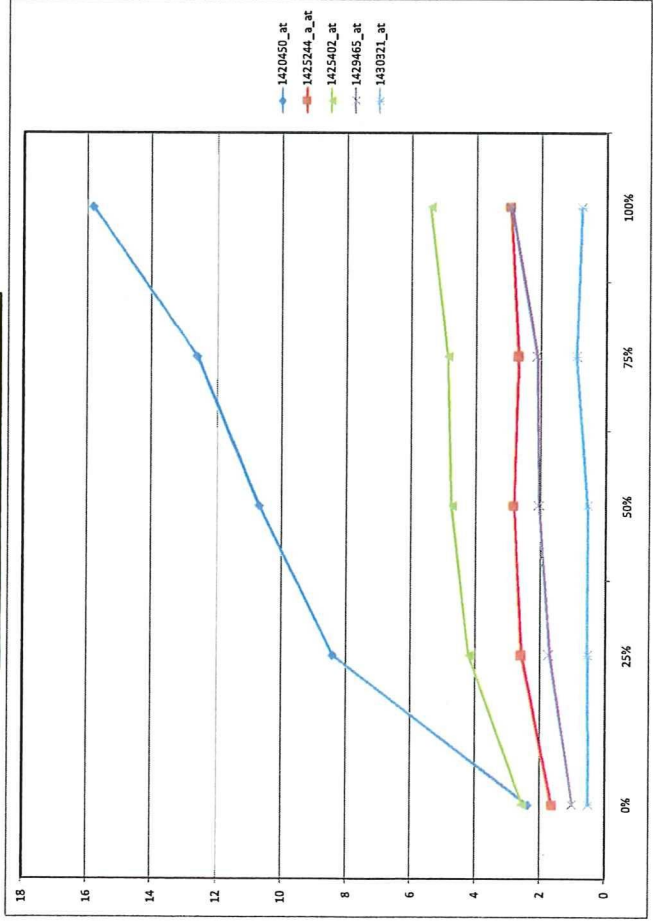


25%において、SpNC変換は、屈曲がみられる

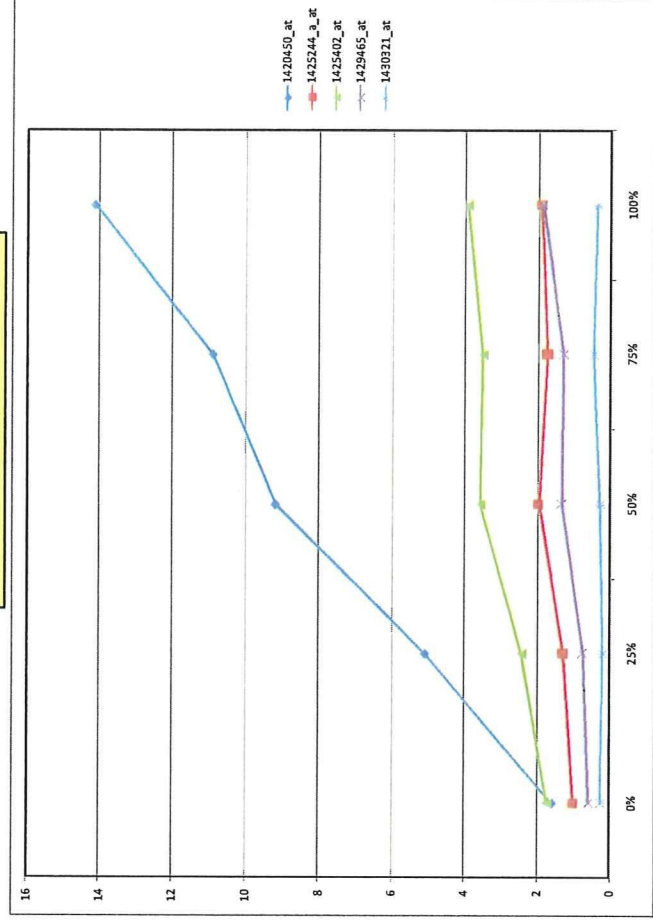
6.細胞あたりコピー数の計算方法

極低発現遺伝子(50%:50%で、raw10程度)

SpNC



線形

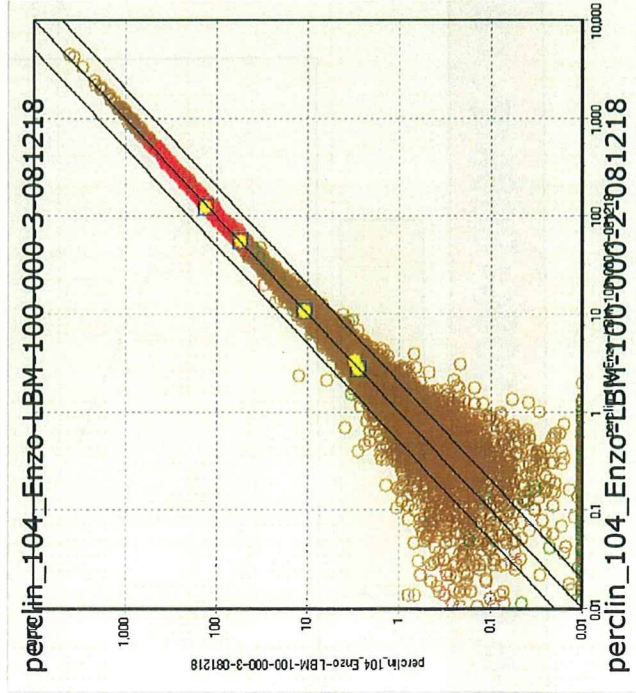


25%において、SpNC変換は、屈曲がみられる

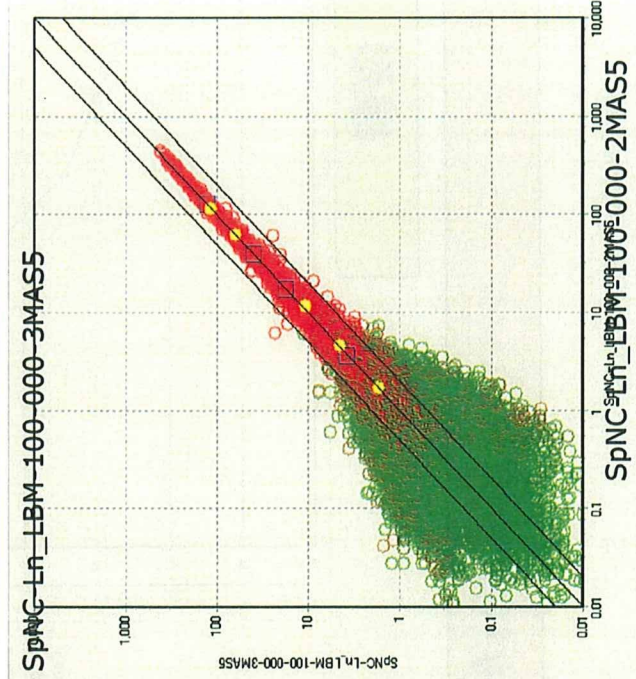
6.細胞あたりコピー数の計算方法

Liver100%の2チップに対する補正結果の分布を示す

Mlang+線形変換



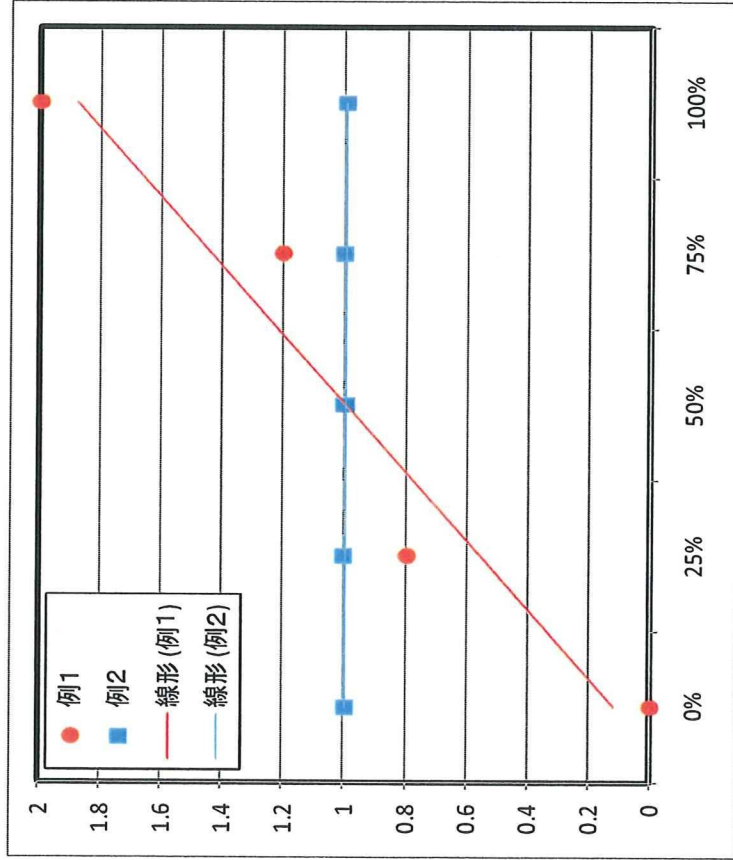
MAS5+ScaI4



Mlangでは、5000コピーから50コピー程度の間はほぼ対角線上に乗り、再現性が高い。その下は徐々に広がっているが、対角線から離れると疎らになっている。

7.直線性の指標

- 直線性の指標として、相関係数の平方を使用してきました。しかし、傾きに依存した値となっています。50%の切片を1に正規化した値での残差平方和(RSS: residual sum of squares)を用いれば、傾きに依存せず、直線性を表現できる

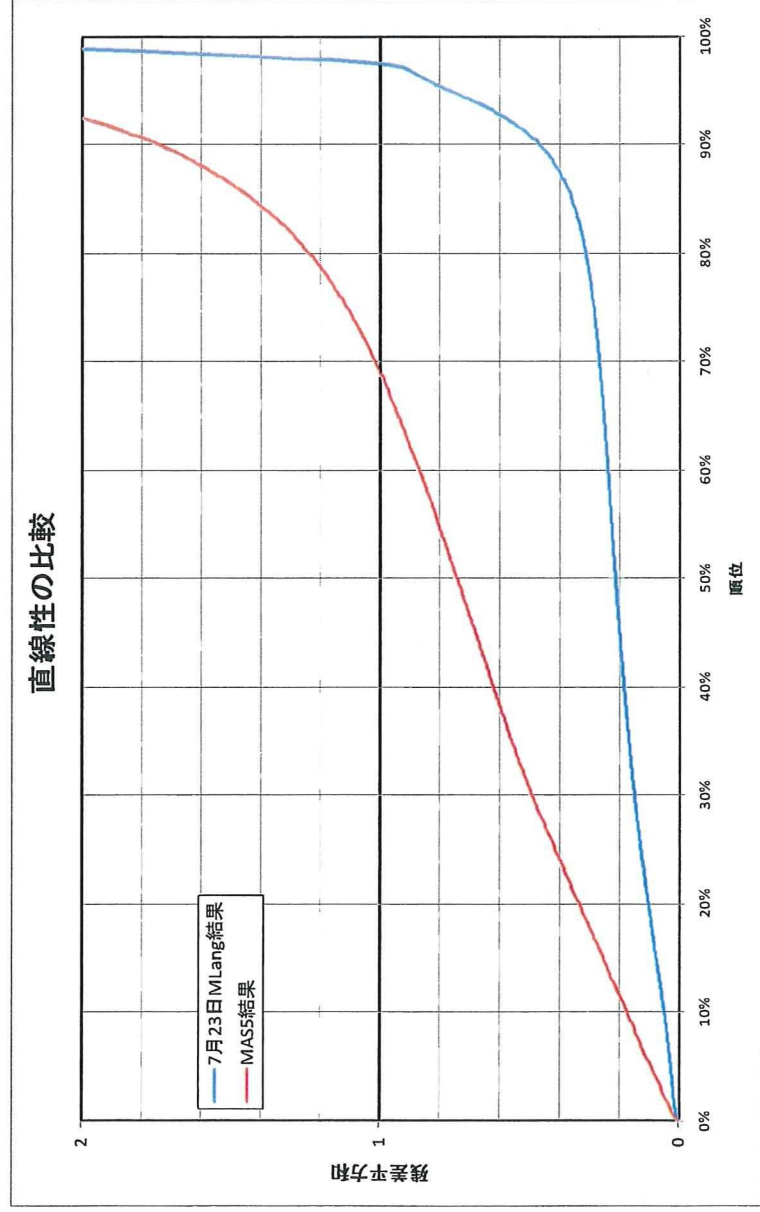


例1	相関係数の平方	残差平方和
0.9308	0.1440	0に近いほうが直線性が高い
0.0000	0.0000	1に近いほど、直線性が高い

例2のほうが、近似直線の上に乗っており、直線性が高い。残差平方和のほうが良い指標と考えられる。

7.直線性の指標

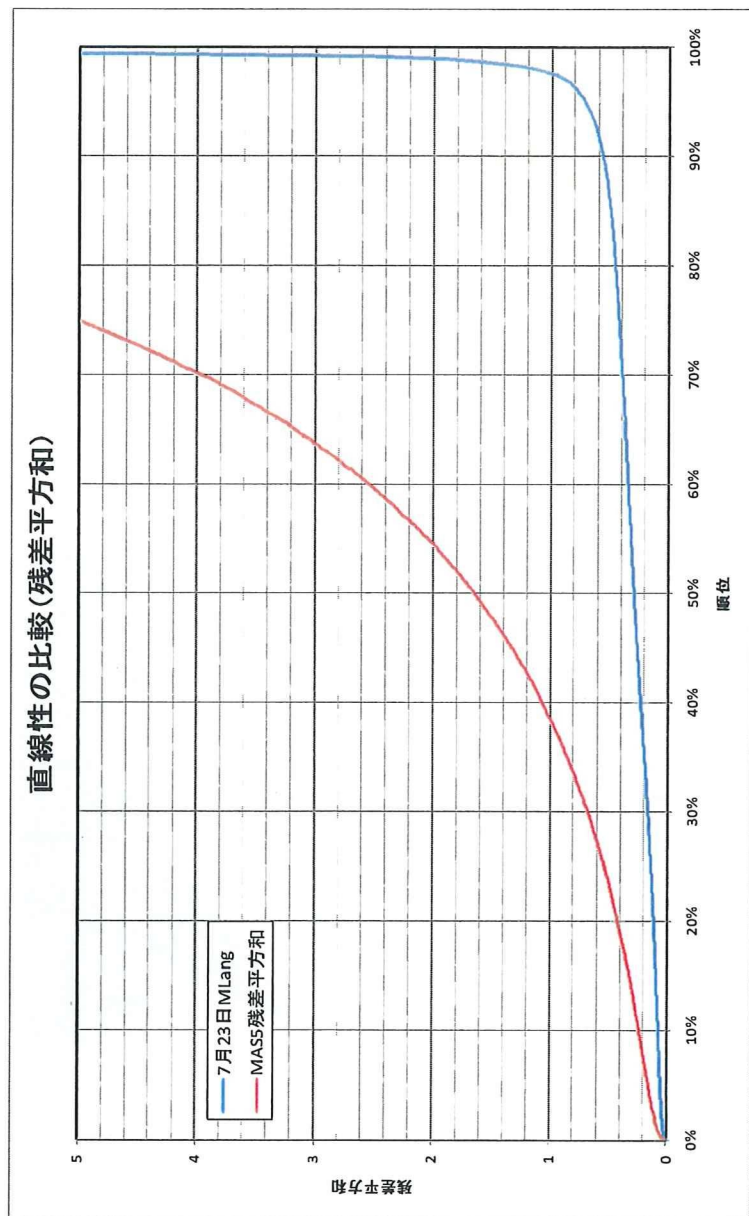
- RSSIによる直線性評価をMAS5とM-Langで比較した。



MAS5の上位20%が示していた直線性を、Mlangでは、80%以上のプローブセットで示している。

7.直線性の指標

- 実験条件ごとの平均を行わずにRSSによる直線性評価を実施した



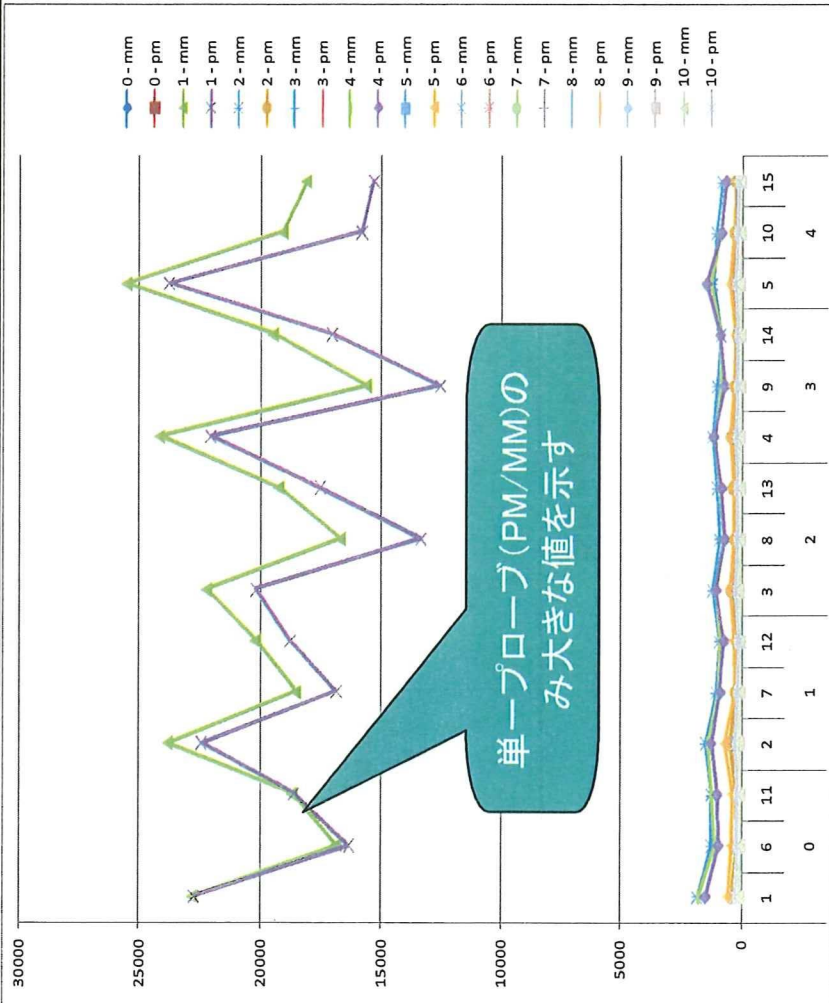
MAS5では、上位30%よりも下のプロポーゼットでは、Triplicateを行うことで、安定性を保っていたと思われ。

7. 異常な値を示すプローブ (LBMでの確認)

プローブセット内で、他のプローブよりかなり高い値を示すプローブが存在する

Prb_no=33808 :
1449508_at

配列が似ているもの(作成した係数)
にも大きな値を示すものはいない



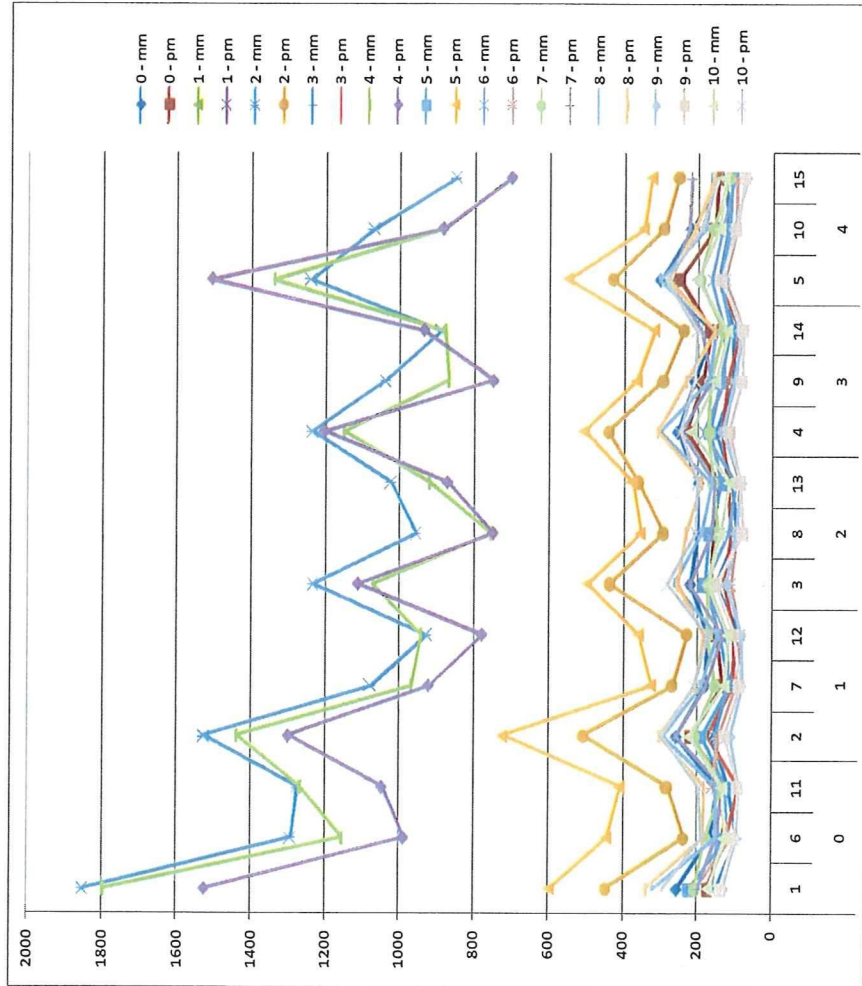
配列情報から関係の強いターゲットを見つけたことが出来なかったと考えられる。このプローブは、存在しないものとして取り扱うことにより、全体補正精度を向上させる

7. 異常な値を示すプロローブ (LBMでの確認)

縦軸を変更し、2番目に大きなプロローブの状況を確認した

縦軸変更

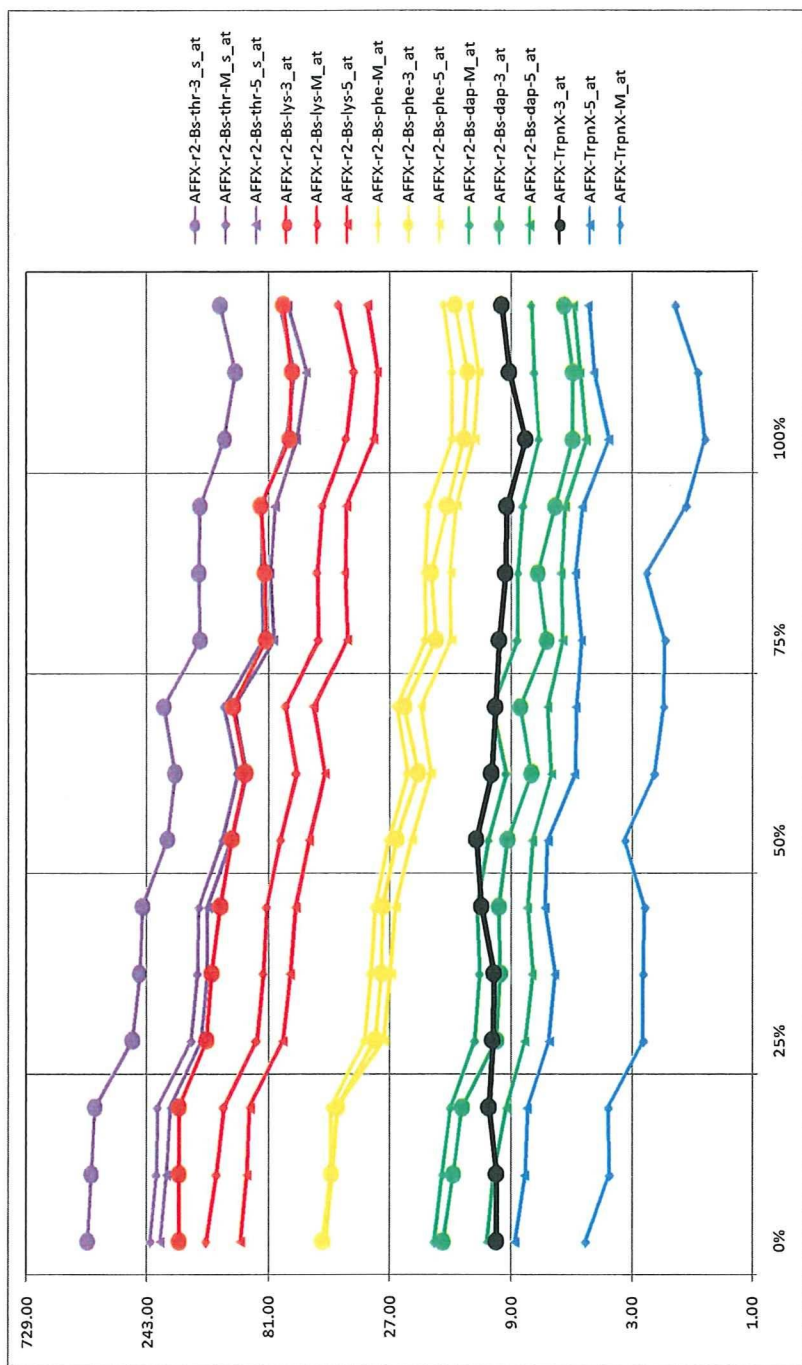
Prb_no=33808 :
1449508_at



MMのプロローブが上に来ており、他の条件での学習が必要と考えられる

7. 異常な値を示すプローブ (LBMでの確認)

GSCのMlang補正後、Perccellome前の値

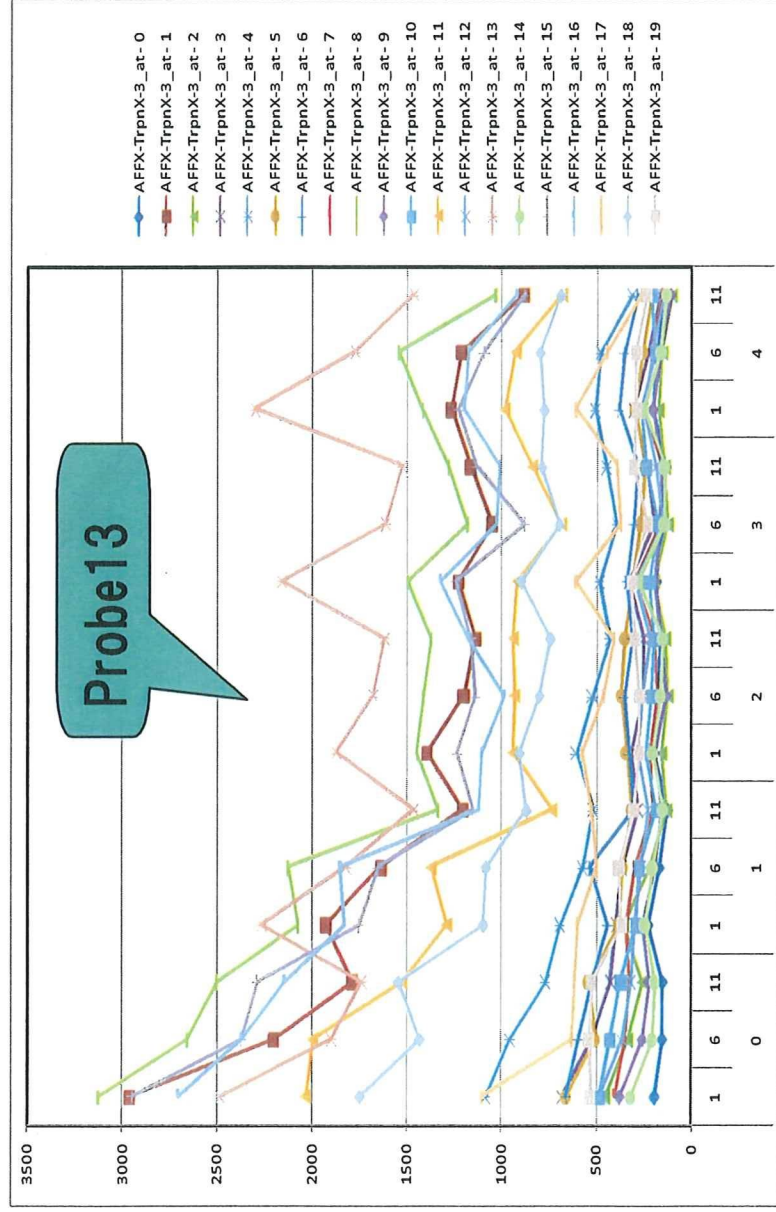


AFFX-TrpnX-3_atの動きが、他のGSCプローブと異なる

他のRNAの影響を強く受けており、分離できていないと思われる

7. 異常な値を示すプローブ (LBMでの確認)

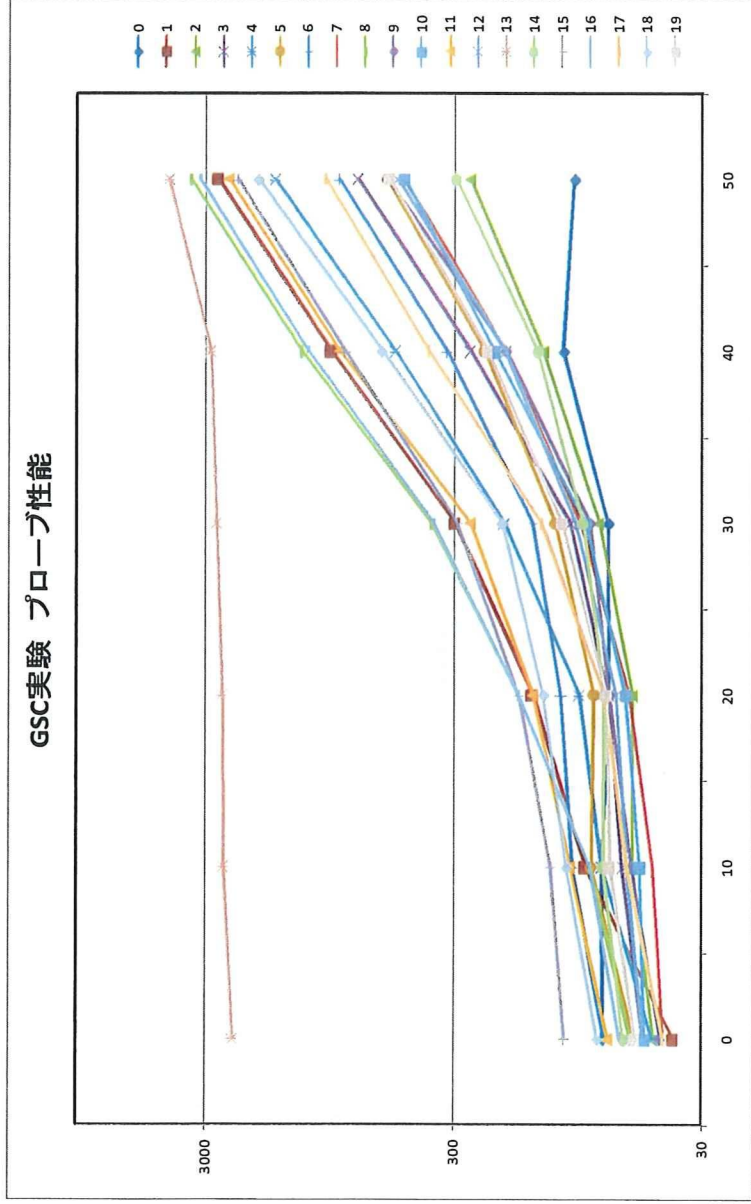
AFFX-TrpnX-3_atのプローブレベルの値(生値)



Probe13は、Trpn以外の影響で明るくなっていると考えられる

7. 異常な値を示すプローブ (LBMでの確認)

- GSC実験による確認TrpnXの各プローブの値(生値)



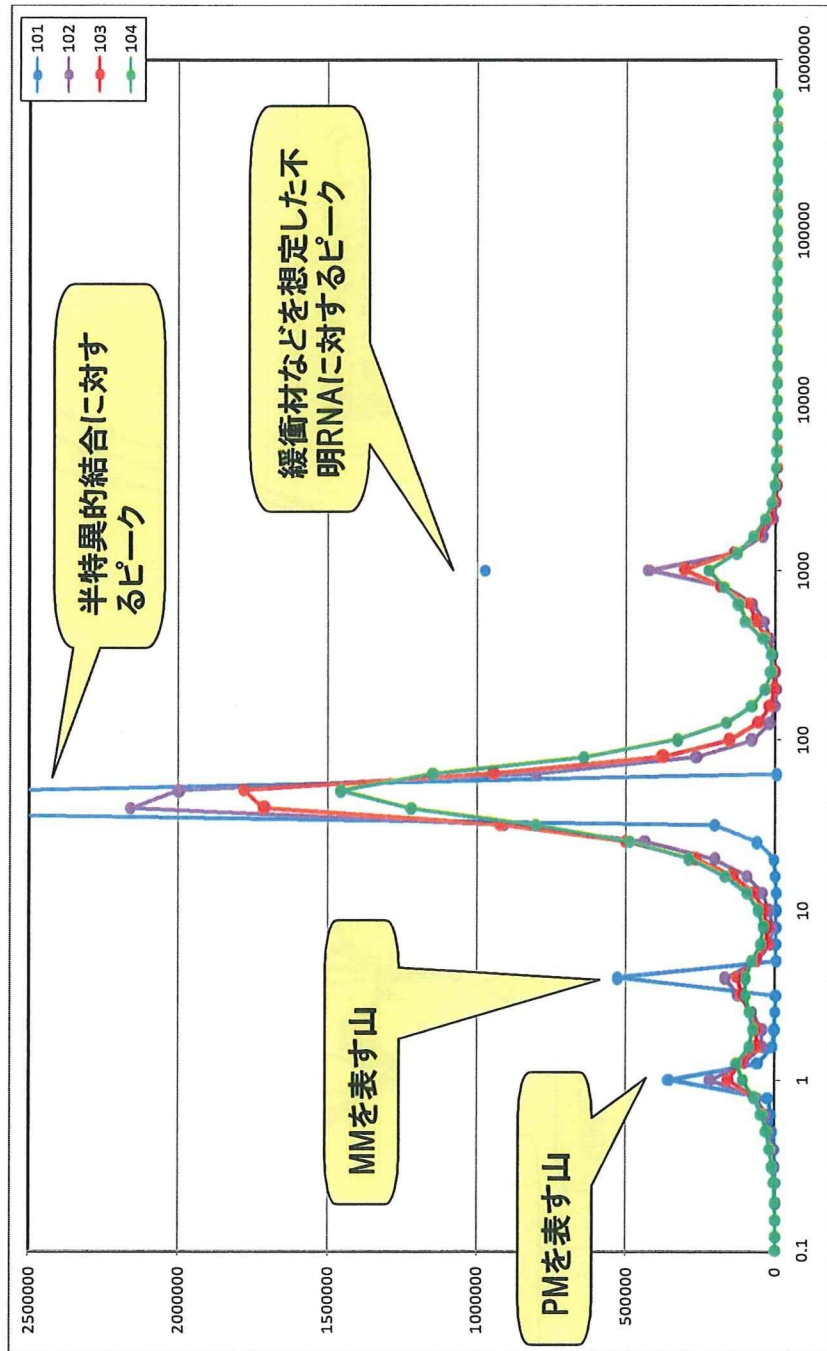
解決策

Mlangで、プローブ13を使用しない。
 係数番号104を103からコピー
 後TrpnXの当該プローブを削除
 する。

プローブ13は、Liverまたは添加物の影響と考えられる
 プローブ0は、TrpnXの値を反映していないと考えられる

8.係数学習の進行

係数学習の進行による係数値の分布の変化(縦軸線形)

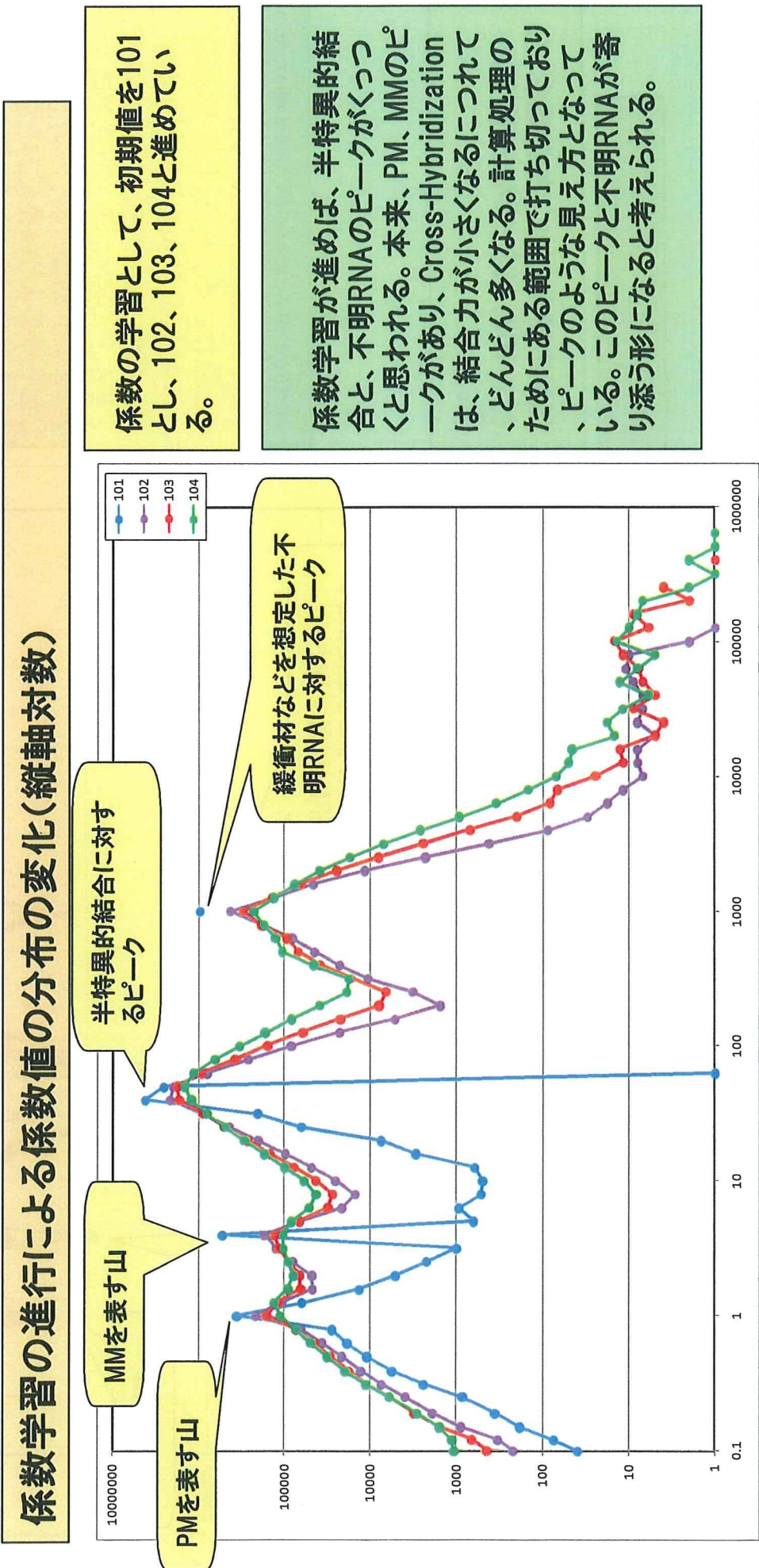


係数の学習として、初期値を101とし、102、103、104と進めている。

係数学習が進めば、半特異的結合と、不明RNAのピークがくっつくと思われ。本来、PM、MMのピークがあり、Cross-Hybridizationは、結合力が小さくなるにつれて、どんどん多くなる。計算処理のためにある範囲で打ち切っており、ピークのような見え方となっている。このピークと不明RNAが寄り添う形になると考えられる。

MMは、プローブに対して特異性が低く、学習の進行により値が変更されやすいと考えられ、初期のピークから大きく下がっている。

8.係数学習の進行



MMIは、プローブに対して特異性が低く、学習の進行により値が変更されやすいと考えられ、初期のピークから大きく下がっている。

8. 係数学習の進行

GSCプローブに対する、全ターゲットの係数値状況(103)

