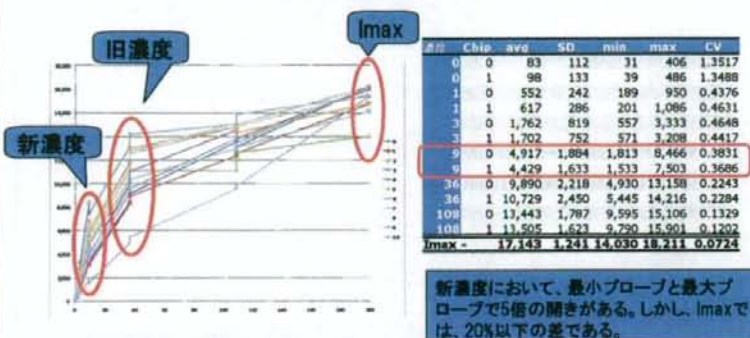


Percellome用GSCの飽和実験 (新旧GSCプローブ)

プローブセット内の発現値差異の原因(Thr)

発現値の差異は存在するが、Imaxの差異がほとんど存在しない。



Percellome用GSCの飽和実験 (新旧GSCプローブ)

HybridizationとWashingにおける結合の数

GSC飽和実験の最大濃度(Thr)の結果

Imaxは同一プローブセット内で、同じような値を示す。

Washing工程における「結合力」は差異がほとんどない。

目標RNAの濃度を極端に高くすれば、Hybridization後に目標のRNAで占められ、吸着したRNA数は同一と考えられる。これらのRNAがWashing工程で、洗い流されても、似たような値になる。ゆえにWashing工程における「結合力」に差異はない。

Washing工程において

「結合力」は差異がほとんどない。

仮説1

WashingとHybridizationでは、結合力の値が違う

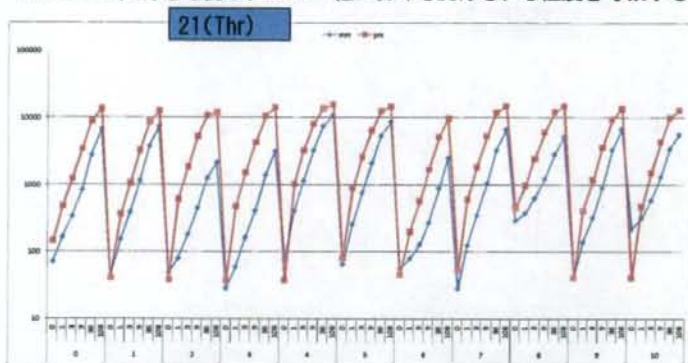
仮説2

Washingでは、差異が現れるほど時間がない。平衡状態になっていない。

両仮説とも、妥当と思われる

Percellome用GSCの飽和実験 (PM-MM関係考察)

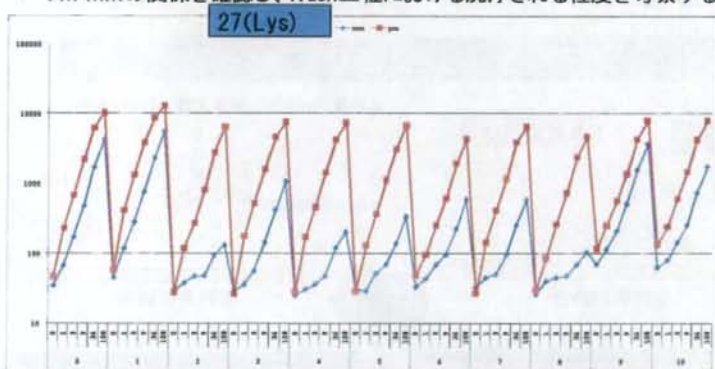
- PM-MMの関係を確認し、Wash工程における洗浄される程度を考察する



高濃度ではPMが飽和し、低濃度では逆転等がみられるが、MMは、PMの1/10~1/3程度の値で推移している。他の変動成分の影響が少ないため、これが本来の動きに近いと考えられる。

Percellome用GSCの飽和実験 (PM-MM関係考察)

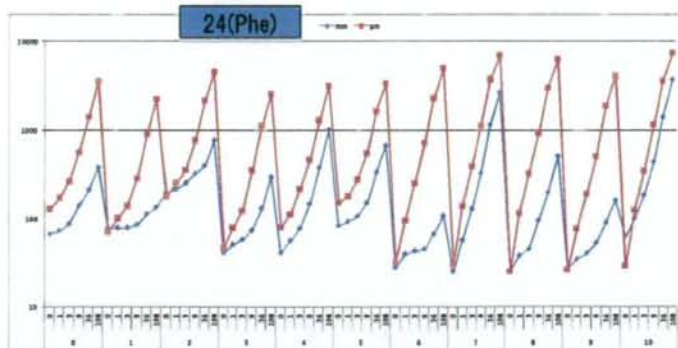
- PM-MMの関係を確認し、Wash工程における洗浄される程度を考察する



高濃度ではPMが飽和し、低濃度では逆転等がみられ、また、いくつかの例外ブロープが存在するが、MMは、PMの1/10~1/3程度の値で推移している。

Percellome用GSCの飽和実験 (PM-MM関係考察)

- PM-MMの関係を確認し、Wash工程における洗浄される程度を考察する

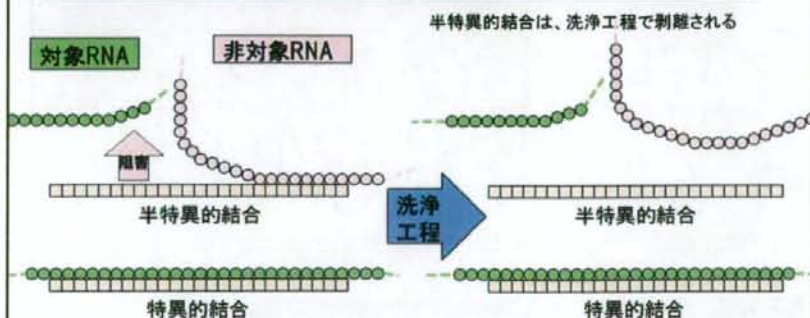


低濃度で背景と考えられる数値が出ている。また、いくつかの例外ブロープが存在するが、MMは、PMの1/10~1/3程度の値で推移している。

Langmuir吸着等温式が多溶質への拡張

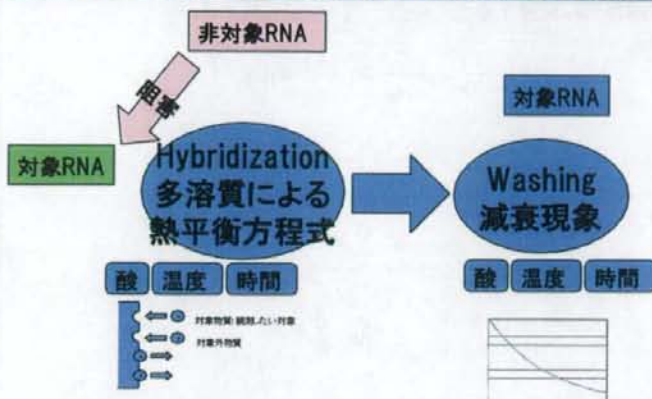
半特異的結合 (Semi-specific Hybridization) という概念を導入し、モデルを構築する

半特異的結合は、プローブと間にある程度の大きさの結合力が働か外れにくく、Hybridization時間内では、特異的結合に置き換えることができない



Langmuir吸着等温式の多溶質への拡張

各RNAの真の濃度がわかっている場合の計測値を導き出すモデルを仮定する



Langmuir吸着等温式の多溶質への拡張

多溶質による吸着のモデル化、Langmuir吸着等温式と同じ仮定

吸着速度は、吸着体の残り個数と溶質の濃度に比例すると仮定する

$$V' = K \cdot \left(N - \sum_{i=1}^n N_i \right) \left(\sum_{i=1}^n C_i \right)$$

個々の溶質の吸着速度は、溶質の濃度の比に比例する

$$V'_i = \frac{C_i}{\sum_{j=1}^n C_j} V'$$

個々の溶質の離脱速度は、吸着した個数に比例する

$$V_i = \alpha_i \cdot N_i$$

式変形

プローブに吸着したターゲットjの個数

$$N_j = \frac{C_j}{\alpha_j} \cdot K \cdot \left(N - \sum_{i=1}^n N_i \right) + \epsilon_j$$

プローブに吸着したターゲットjの離脱速度係数
吸着速度係数は1とし、 α で調整する

$$\alpha_j \geq 0$$

Langmuir吸着等温式の多溶質への拡張

計算式の展開
計算可能な陽形式に変換する

$$N_y = \frac{C_j}{\alpha_y} \cdot K \cdot \left(N - N_y - \sum_{k=1}^m N_{ik} \right) + \varepsilon_y$$

$$N_y + \frac{C_j}{\alpha_y} K N_y = \frac{C_j}{\alpha_y} \cdot K \cdot \left(N - \sum_{k=1}^m N_{ik} \right) + \varepsilon_y$$

$$N_y \left(1 + \frac{C_j}{\alpha_y} \right) = \frac{C_j}{\alpha_y} \cdot \left(N - \sum_{k=1}^m N_{ik} \right) + \varepsilon_y$$

$$N_y = \frac{N - \sum_{k=1}^m N_{ik}}{\alpha_y + C_j} C_j + \varepsilon_y$$

$$N_y = \frac{N + \hat{N}_y - \sum_{k=1}^m \hat{N}_{ik}}{\alpha_y + C_j} C_j + \varepsilon_y$$

Langmuir吸着等温式の多溶質への拡張

各RNAの真の濃度がわかっている場合の計測値を導き出すモデル(数式表現)

① Hybridizationで、プローブ*l*に吸着したターゲット*j*の個数

$$N_{lj} = \frac{N - \sum_{k=1}^m N_{ik}}{\alpha_y + C_j} C_j$$

Washing

② Washing後で、プローブ*l*に吸着したターゲット*j*の個数

$$n_{lj} = N_{lj} \cdot e^{-\gamma \alpha_y}$$

観測

③ Washing後で、プローブ*l*で観測される値

$$n_l = \sum_{j=1}^m n_{lj} + \bar{\varepsilon}_l$$

離脱速度比例定数(結合強度)の算出

- プローブとRNAの配列から離脱速度を算出できるとすれば、真の濃度を観測値から求めることができる。
- 配列から離脱速度が算出可能か試みた。

離脱速度比例定数(結合強度)の算出

- プローブとRNAの配列から離脱速度の計算方法を、次のように仮定する

α_{ij} : プローブ i に対する、RNA j の離脱速度比例定数

小さい値ほど似ている

離脱速度は、結合力の逆数であると仮定する

$$\alpha_{ij} = 1/c_{ij}$$

c_{ij} : プローブ i に対する、RNA j の結合強度

結合は連続した配列の長さにより規定可能である。

長いほど強固である

複数箇所で一一致する場合は、足し算可能である

$$c_{ij} : f(l_{ij})$$

l_{ij} : プローブとRNA j の一致する配列の最大長

離脱速度比例定数(結合強度)の算出

濃度がターゲットと一次の関係で運動していると仮定する

$$C_j = p_{ij} \cdot C_i + q_j$$

代入する

$$\sum_{j=1}^n \frac{P_{ij}}{\alpha_j} = p_i$$

$$\sum_{j=1}^n \frac{q_j}{\alpha_j} = q$$

$$n_i = N \cdot \frac{1}{\alpha_i p_i + 1} \cdot \frac{c_i}{\frac{K}{\alpha_i p_i + 1} + c_i}$$

1溶質の式 $n = N \cdot \frac{c}{\frac{\alpha}{K} + c}$

lmax: 1溶質 → 多溶質

$$N \rightarrow N \cdot \frac{1}{\alpha_i p_i + 1}$$

最大値は、運動している成分の影響のみを受ける。

Langmuir飽和式は、一定量の成分(背景)の影響を受けるが最大値は影響を受けない。

通常の実験では、きれいなLangmuir飽和式(シグモイド曲線)が描かれ、それ以外の状況を見ることができない。
1. Latin Square (Alfymatrix 公開データ)
2. Liver Brain Mixture
3. Graded Spike Cocktails 飽和実験

GSC飽和実験の最大濃度RNAの最大値は理論通り、ほぼ一定している。第2濃度の最大値を用いて、離脱速度の検討が可能ではないだろうか

離脱速度比例定数(結合強度)の算出

GSC飽和実験の第1濃度(Thr)、第2濃度(Lys)に着目し、推定された最大値は、理論値+誤差で示されると仮定する

$$\bar{N} = N \cdot \frac{1}{\alpha_i p_i + 1} + \varepsilon$$

$$\bar{N} = N \cdot \frac{1}{\alpha_i \sum_{j=1}^n \frac{P_{ij}}{\alpha_j} + 1} + \varepsilon$$

$$\bar{N} \left(\alpha_i \sum_{j=1}^n \frac{P_{ij}}{\alpha_j} + 1 \right) - N = \varepsilon$$

- 第1濃度をターゲットとしてみた場合
 - 第1濃度でターゲット箇所(25mer一致)の場合
 - 第1濃度でターゲット箇所以外(24mer以下一致)の場合
 - 第2濃度の場合
 - 第2濃度をターゲットとしてみた場合
 - 第2濃度でターゲット箇所(25mer一致)の場合
 - 第2濃度でターゲット箇所以外(24mer以下一致)の場合
 - 第1濃度の場合
- 第3濃度以下は無視する

α を配列から計算を行ってみて、誤差 ε が小さくなりそうな計算式を探る

離脱速度比例定数(結合強度)の算出

既知の情報を引数とする結合力に関する関数を組み立てる

Hybridization時点とWashing時点の結合力を求める

次の例を考える

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25				
例1	C	T	T	C	T	T	A	A	C	G	G	T	A	C	T	T	A	T	T	A	C	A	G	T	A	C	C	A	A
例2	A	T	A	T	G	G	T	C	G	A	C	C	C	T	G	G	T	T	C	C	A	G	T	A	C	C	A	C	A
例3	C	C	A	T	G	G	T	C	G	A	T	T	A	G	C	C	G	T	T	A	C	A	G	T	A	C	C	C	A
Probe PM		A	T	G	G	T	C	G	A	T	T	A	G	C	C	G	T	T	A	C	A	G	T	A	C	C			

例1	C	T	T	C	T	T	A	A	C	G	G	T	A	C	T	T	A	T	T	A	C	A	G	T	A	C	C	A	A
例2	A	T	A	T	G	G	T	C	G	A	C	C	C	T	G	G	T	T	C	C	A	G	T	A	C	C	A	C	A
例3	C	C	A	T	G	G	T	C	G	A	T	T	A	G	C	C	G	T	T	A	C	A	G	T	A	C	C	C	A
Probe MM		A	T	G	G	T	C	G	A	T	T	A	G	C	C	G	T	T	A	C	A	G	T	A	C	C			

離脱速度比例定数(結合強度)の算出

プローブとターゲットとの結合力は各塩基での結合力の和により表現可能であると仮定する

各塩基での結合力を次のように計算可能であると仮定する

配列が一致した場合に、結合力があり、一致しない場合には結合力0とする

G-C対は、A-T対の α 倍の結合力を有する

プローブ内位置での結合力差異は、位置に関する一次式で表現可能

$$pos(i) = \beta \cdot (i - 13) + 1$$

不一致部分から、n-mer連続して一致した箇所から結合力が発生する

$$Bind = \sum_{i=1}^{25} s_i \cdot \alpha_i \cdot pos(i) \cdot \gamma_i$$

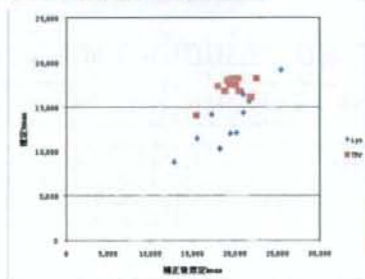
洗浄工程においては、25mer一致以外は洗い流されると仮定する。
検証を迅速に行うために仮定する

離脱速度比例定数(結合強度)の算出 (試算)

不一致と5個以上離れている場合に結合力を発生すると仮定した。
ただし、ThrとLysの間では、9個以上連続で一致することはなく、Probe末端での結合が反映されている

GC比率 0.54 GC塩基がAT塩基に対する結合力
傾き -0.01 3'末端と5'末端の間の結合力差率
 γ_{end} 1 末端における結合力
 γ_{mid} 0.82 中央における結合力

GCペアの方が弱いという結果になった。GSC第1濃度、第2濃度の特殊な関係の可能性もある



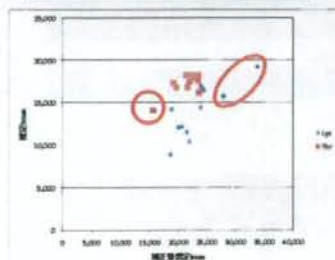
正の相関がみられ、阻害が弱く見積もられているようである。

離脱速度比例定数(結合強度)の算出 (試算)

不一致と4個以上離れている場合に結合力を発生すると仮定した。
ただし、ThrとLysの間では、9個以上連続で一致することはなく、Probe末端での結合が反映されている

GC比率 0.431 GC塩基がAT塩基に対する結合力
傾き 0.042 3'末端と5'末端の間の結合力差率
 γ_{end} 1 末端における結合力
 γ_{mid} 0.825 中央における結合力

GCペアの方が弱いという結果になった。GSC第1濃度、第2濃度の特殊な関係の可能性もある



第3プローブを除き補正後の最大値が一定の範囲に収まっている。

離脱速度比例定数(結合強度)の算出 (試算結果検討)

離脱速度の計算をGSCの飽和実験結果を用いて実施した。

GSCの2RNA間である程度の結果が出た。

改善すべき点が見つかった。

GCペアが、ATペアよりも結合力が弱い

対象RNAを増やし、配列条件を増やす必要がある

範囲に収まらない組み合わせが存在した

結合力の考え方(塩基間の強度の合算)

外れることの可能性を考えるためには、熱力学的偶然性で評価すべきと思われる。

Washing時の離脱速度

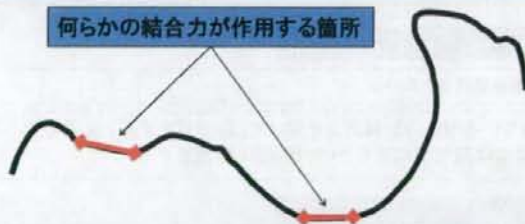
25mer 以外でも残るとすれば、現象を説明しやすいか検討する

離脱速度比例定数(結合強度)の算出 ターゲットの複数個所への合致

課題

単一のプローブがターゲット内の複数個所と合致するケースが存在する

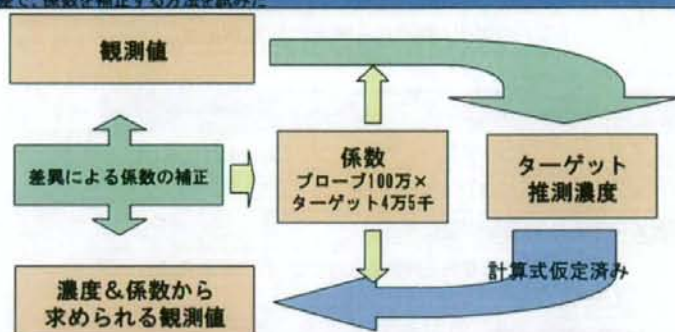
何らかの結合力が作用する箇所



別の結合個所が存在する場合には、別の溶質が存在するとみなせる。しかし、複数溶質に対するLangmuirの式の中で、結合力(離脱速度の逆数)と濃度の積が現れており、結合力の和をそのRNAに対する結合力とみなして取り扱うことが可能である。

係数補正による係数の同定

プローブとターゲットのペアの塩基配列に基づいた結合力の同定は、ターゲット範囲外での結合など、未知要素が多く、困難なので、観測値から推測値を求め、更に観測値を推定し、その誤差で、係数を補正する方法を試みた



複数のチップ、実験条件で、繰り返し実施することで本来あるべき係数に近づいていく

ただし、計算時間がかかる。また、取返が保証できない

係数補正による係数の同定 計算方法(概要)

濃度 c_j 、係数 α_{ij} が決定されている場合

①Hybridization後の吸着個数 N_{ij} を繰り返し計算で求め、そのあと②Washing後の吸着個数 n_{ij} 、③Washing後の全吸着個数 n_i を順次計算する

観測値 n_i 、係数 α_{ij} が決定されている場合

適切な方法で濃度初期値を求める

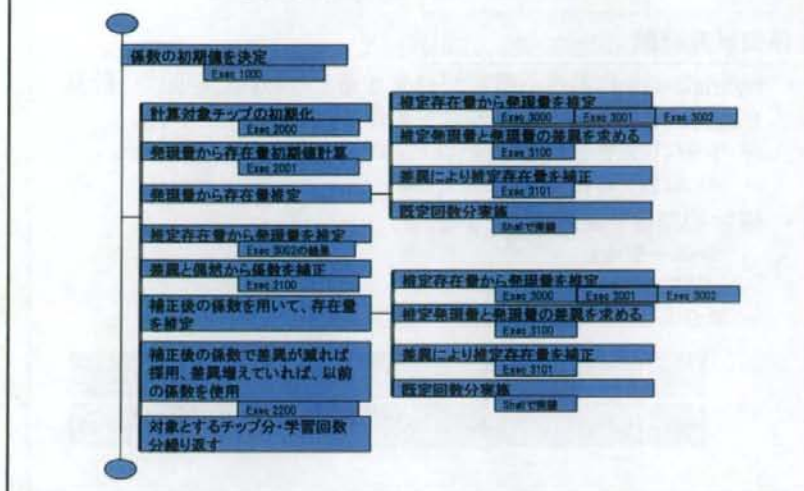
c_j が決定されている場合の計算方法を用いて、観測値を求め、結果の値との差がある場合には濃度の補正を行い、繰り返し計算をする

複数のチップで

プローブ配列から適切な初期値を求める(後述)

観測値から真の値を求める。その値から観測されるべき値を計算する。その差異から係数の変更を行う

係数補正による係数の同定 係数決定アルゴリズム



係数補正による係数の同定

計算前提

次のことを前提として計算を行う

I_{max} を 20000で固定する

GSCの実験で求められた知見に基づく

明らかにエラーと考えられる状況を排除する

発現量が最大値 I_{max} の95%を超過しているプローブ

エラープローブとして計算しない

すべての実験条件でエラープローブとして判定されることは
ないはずなので、適切な推測が可能はずである

係数補正による係数の同定

係数計算範囲

- Hybridizationにおける阻害が発生すると思われる範囲で、計算を実施する範囲を、計算時間を考慮し、次のようにする。
 - 中央に位置する場合、13mer以上連続
 - 端に位置する場合、12mer以上連続
- 係数初期値を次のようにする。
 - 25mer一致→1
 - 一致数が1個減るごとに、1.4倍
 - 端の塩基を含まないと、32倍

例1	第1塩基から12mer連続	1.4の13乗≒100
例2	第2塩基から23mer連続	1.4の2乗×32≒64

係数補正による係数の同定

係数補正方法

- 偶然の補正と評価関数による探索で収束させることもできるが、時間がかかる。次の方法で補正する

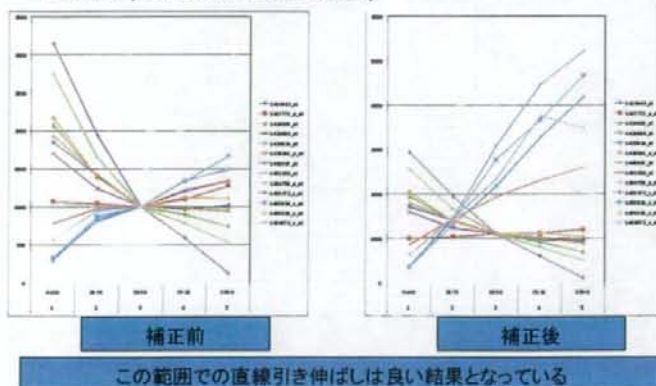


係数補正による係数の同定(試行1)

- 学習による方法を用いて係数が同定可能か、また、発現値飽和の補正の可能性を探る
- 対象 LBM 15チップ
 - 各チップ1回ずつの学習
 - 中央は13mer連続一致、端は12mer連続一致に影響を及ぼしているものとして計算を行う。
 - 対象となる関係: 8,307,1055対

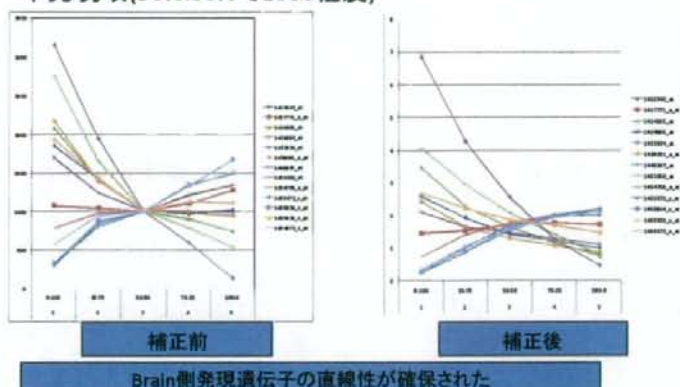
係数補正による係数の同定(試行1)

- 飽和補正結果(LBM:単Langmuir(昨年度実施))
- 中発現域(50%:50%で1000程度)



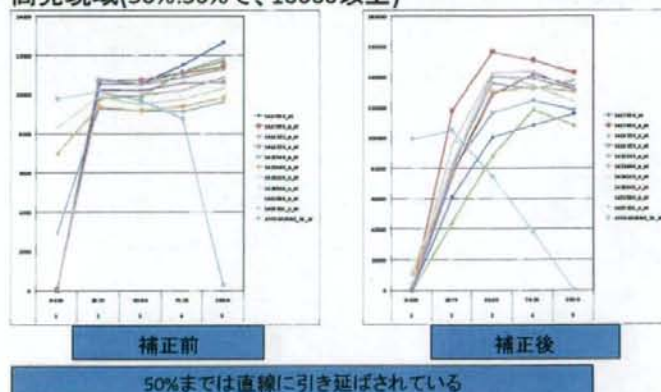
係数補正による係数の同定(試行1)

- 飽和補正結果(LBM: 多Langmuir)
- 中発現域(50%:50%で1000程度)



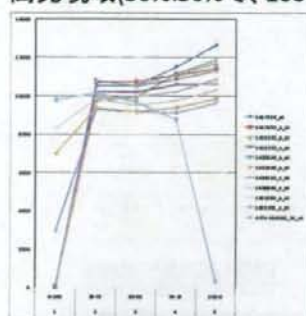
係数補正による係数の同定(試行1)

- 飽和補正結果(LBM: 単Langmuir(昨年度実施))
- 高発現域(50%:50%で、10000以上)

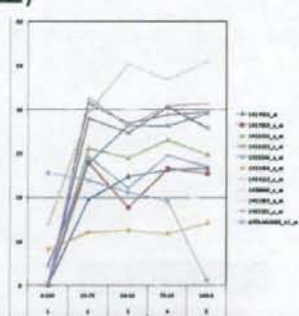


係数補正による係数の同定(試行1)

- 飽和補正結果(LBM:多Langmuir)
- 高発現域(50%:50%で、10000以上)



補正前

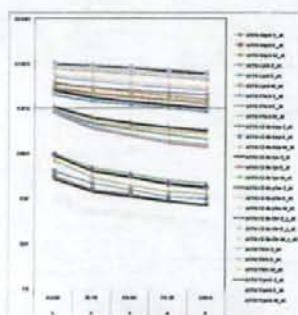


補正後

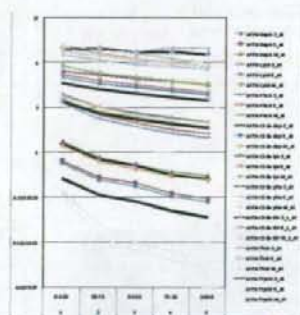
飽和最大値(l_{max})の再考が必要である

係数補正による係数の同定(試行1)

LBM(5C:多Langmuir)



補正前



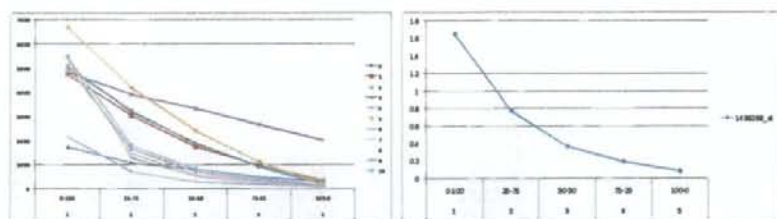
補正後

補正前はPMの平均値

4番目(Dap)がうまく補正できているように思われる

係数補正による係数の同定(試行1)

1844 F1:凸凸Probe1:第1系group1



補正前

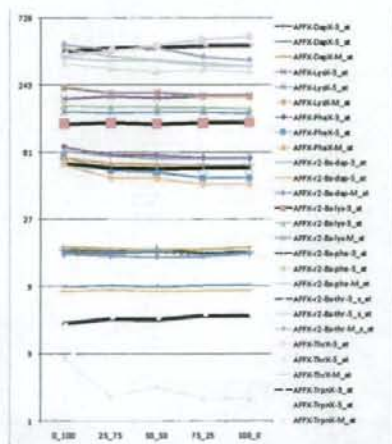
補正後

補正後、下に凸な状況は残っているが、最大の落ち込みProbeよりも緩やかな凸にはなっている

係数補正による係数の同定(試行1)

絶対化計算(Percellome)
絶対化計算をすることで、Liver、Brain
固有の偏差を打ち消せる部分がある
と考えられるので実施してみる

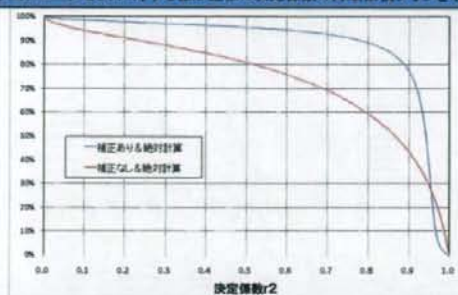
補正 & 絶対化計算を行
うと、GSCは水平になる。



係数補正による係数の同定(試行1)

LBM: 補正と絶対計算後の直線性

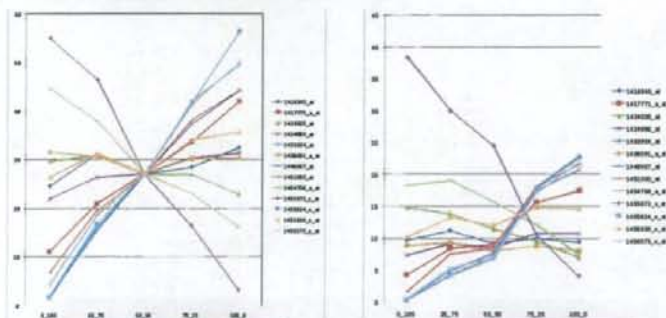
LBMは、配合割合を変更しているだけなので、補定濃度は直線になるはずである。直線に近づいたかの指標として、全プロブセットに対する線形近似の決定係数の降順累積グラフを示す



Percolome絶対計算の直線性と比して格段に改善していると考えられる
決定係数1付近が少ないのは、ControlRNAを用いた補正の結果と思われる

係数補正による係数の同定(試行1)

- 飽和補正結果(LBM: 多Langmuir&絶対化)
- 中発現域(50%:50%で1000程度)



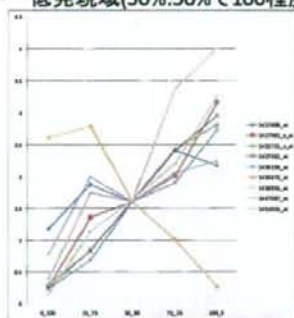
補正無&絶対計算

補正&絶対計算

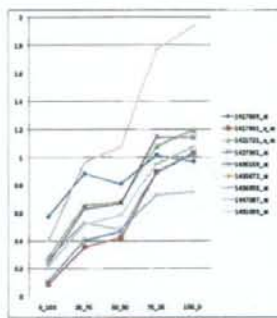
50:50において発現量の差異が大きい、どちらが真値か?

係数補正による係数の同定(試行1)

- 飽和補正結果(LBM: 多Langmuir&絶対化)
- 低発現域(50%:50%で100程度)



補正無&絶対計算

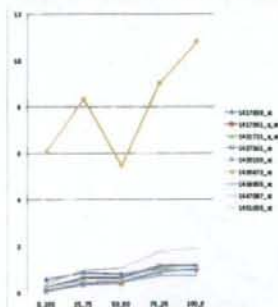


補正&絶対計算

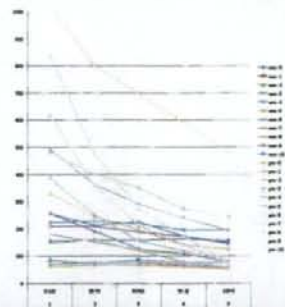
1435672_atは、補正後非常に大きな値となった

係数補正による係数の同定(試行1)

- 飽和補正結果(LBM: 多Langmuir&絶対化)
- 低発現域(50%:50%で100程度)



補正&絶対計算

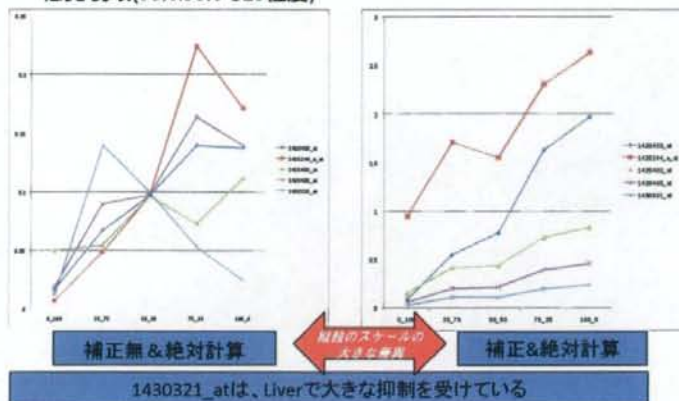


1435672_at Probe発現値

Brain側で多く、Liver側で少ないProbeは典型的な抑制カーブを描いている。このような抑制がなかったものとした場合の濃度を推定している。こちらが真値ではないだろうか？

係数補正による係数の同定(試行1)

- 飽和補正結果(LBM: 多Langmuir&絶対化)
- 低発現域(50%:50%で10程度)



係数補正による係数の同定(試行1)

- まとめ
 - 多くのプローブセットで線形性が確保された。
 - 決定係数累積%グラフ
 - LBMが濃度に関して等間隔の実験であることを与えずに補正を行っている。
 - 特に、下に凸な状況の補正が行われた。
 - 発現量が小さな領域では、今までよりもかなり大きな値が推定された。
- 課題
 - 最大飽和係数(I_{max})の再検討
 - 高発現値のプローブセットで線形性が再現できなかった。飽和最大値は全体で一つの重要なパラメータなので、再度検討する。
 - 係数対象の最少長さの同定
 - 今回は端12mer、中央13merで打ち切った、どこまで行えばよいのか？
 - 係数対象のターゲット外の遺伝子領域の合致に関して
 - RNAの中で、ターゲットされている部分は一部であり、プローブには、ターゲット以外の箇所も結合する。また、添加RNAも結合する。これらがどの程度影響するのか不明である。