

全プローブ計算⑥

Teradataによる線形領域モデル計算を用いた絞込

- Teradataを用いた線形計算を用いて、対象プローブの絞り込みを実施
- 次の条件を用いた
 - 上に凸(線形領域2次モデルの2次係数が負)
 - 線形領域線形モデルの傾きが平均値の5%以上
 - 線形領域線形モデルの傾きが標準偏差の5%以上

終了状態	Brain側飽和	Liver側飽和
勾配ベクトル判定収束	51,278	51,239
移動量判定収束	48,288	18,678
局所的最小値	50,203	50,139
総計	149,769	120,056

全プローブ計算⑥

非線形最適化の課題&解決テクニック

- いくつかのプローブにおいては、適切な解に収束しなかった
 - 繰り返しの範囲量を超えた。または収束回数を超えた。
 - 非線形計算のパラメータを変更してやり直すようにプログラムに組み込み、ほとんどのプローブで収束するようにした
 - 繰り返しの最中に無限の数値が現れ、バッチモードのRが中断した
 - メッセージ「nlm」により有限でない値が与えられました」
 - 手作業でパラメータ変更実施

補正方法	動物種	モデル	ProbesetNo	Pm/Mm	Probe	対処方法
PM	Liver	Langmuir	10709	MM	6	I00=18.5, BP0=16
	Liver	Linear	43161	PM	3	BPOを半分に
	Brain	Linear	1680	PM	10	BPOを半分に
	Brain	Linear	35906	PM	5	BPOを半分に
PMMM	Liver	Langmuir	11913	pm	4	I00=92, k00=1.11, BP0=11.9
	Brain	Linear	15256	pm	4	BPOを半分に

今後も発生する可能性はあるが、既に出来上がったモデルのパラメータを与えることで、回避できる。

全プローブ計算⑥非線形最適化の結果

- Liver側飽和では多くのプローブでLangmuirモデルが採用された。
- Brain側飽和では、Langmuirモデルが採用されたプローブは限られている

終了状態	Brain側飽和	Brain側採用個数	Liver側飽和	Liver側採用個数
勾配ベクトル判定収束	51,278	3	51,239	6,383
移動量判定収束	48,288		18,678	746
局所的最小値	50,203	16	50,139	2,742
総計	149,769	19	120,056	9,871

全プローブ計算⑥ 結果Langmuir Model採用個数

- ProbeSet中で何個のProbeがLangmuirモデルを採用しているか
- Brain側飽和

合計 / Count(*)	列ラベル					
行ラベル	0	1	2	3	4	総計
0		3	3	2		8
1	2					2
5					1	1
総計	2	3	3	2	1	11

Brain側飽和では非常に少ない

全プローブ計算⑥ Langmuir Model採用個数

• Liver側飽和

合計 / Count(*) 行ラベル	列ラベル	0	1	2	3	4	5	6	7	8	9	10	11	総計
0			1089	225	124	81	42	41	17	10	6	3		1638
1		1077	196	62	45	32	20	34	34	21	6	6		1533
2		39	19	41	27	25	17	19	23	12	10	7	1	240
3		11	13	15	9	8	15	17	15	16	21	7	10	157
4		2	3	2	2	14	7	9	17	19	22	13	8	118
5					1	5	4	8	7	16	13	15	13	82
6					3	2	1	3	6	7	10	9	19	60
7		1					2	2	6	12	13	18	17	69
8					1	1	1	4	1	1	10	12	13	44
9							1	1	1	1	3	7	7	20
10								1	1		2	10	18	32
11											1	5	5	11
総計		1130	1320	345	212	168	110	137	127	115	117	112	109	4002

PM側のLangmuirモデル採用個数が多いProbeSetが補正の影響を受けるはずである

全プローブ計算⑥検証①

- LBMでの線形性チェック
 - 50:50を1に正規化したグラフを作成し、線形性のチェック
 - Langmuir変換を含むプローブセットを全てプロットし、目視で線形性の確認する。値の低いもの以外で変な値を示すものがないことを確認する

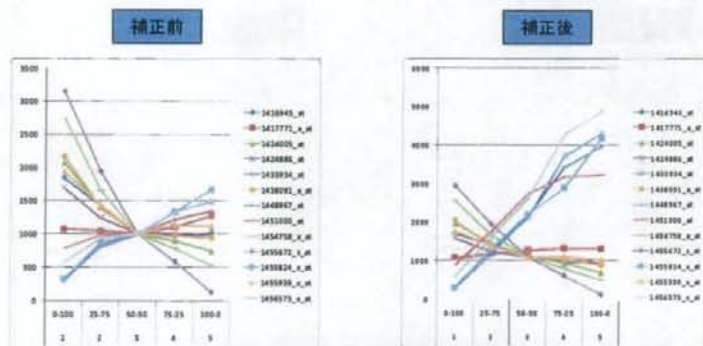
全プローブ計算⑥検証①

- 対象プローブセット: 50%:50%において、1000程度の値を持つ

affy_id	Intensity
1424886_at	1008.53
1433934_at	1004.20
1424005_at	1002.17
1417771_a_at	1005.90
1451000_at	1000.37
1456573_x_at	1005.87
1455824_x_at	1006.07
1416943_at	1007.17
1454758_a_at	1007.60
1438091_a_at	1008.33
1448967_at	1003.97
1455939_x_at	1006.53
1455672_s_at	1004.00

全プローブ計算⑥検証①

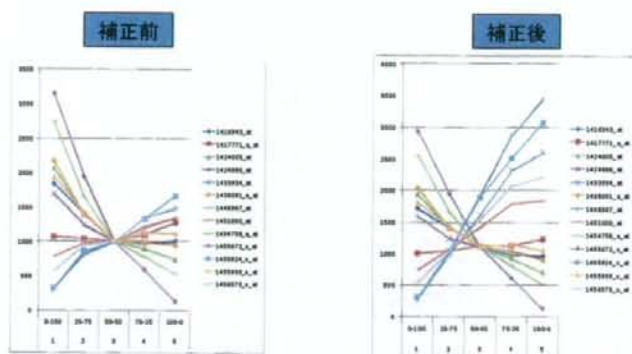
- 対象プローブセット: 50%:50%において、1000程度の値を持つ



Liver 側飽和のプローブは直線に引き延ばされ、高い値で、不安定になっている。

全プローブ計算⑥検証① 高値補正付き

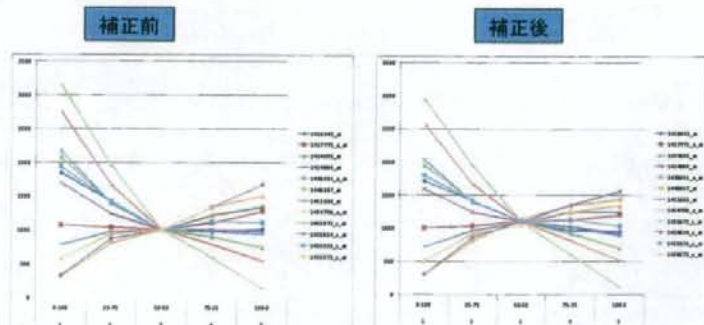
- 対象プローブセット: 50%:50%において、1000程度の値を持つ



高値を補正する式にすると若干安定する

全プローブ計算⑥検証① 飽和値指定

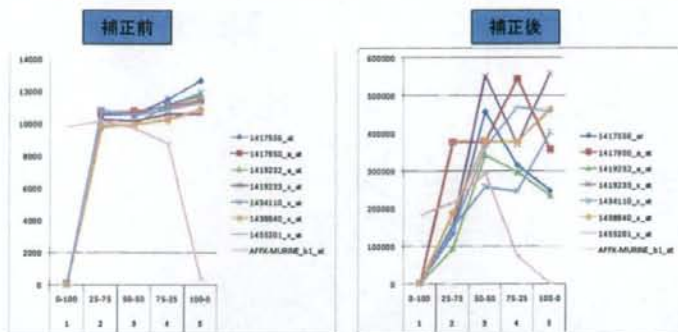
- 対象プローブセット: 50%:50%において、1000程度の値を持つ



飽和値を指定すると安定する

全プローブ計算⑥検証①

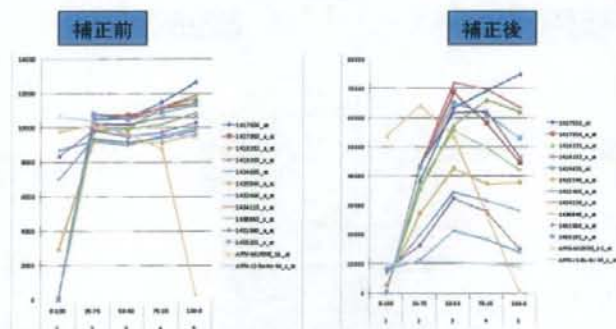
- 対象プローブセット: 50%:50%において、10000以上の値を持つ



0%から50%まではすこし上向きだが直線に見えなくもない。しかし、50%を超えると、崩れている。Logmuirの最大値推定に問題がありそうである

全プローブ計算⑥検証① 高値補正付き

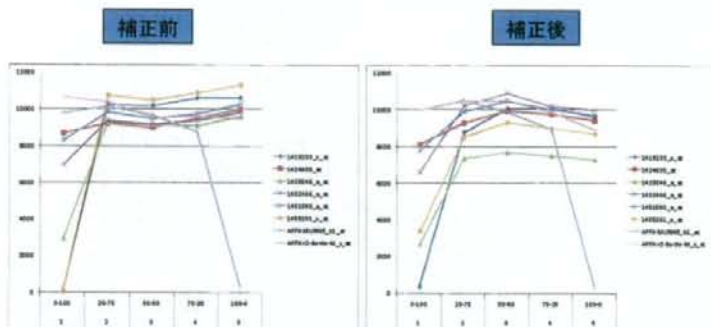
- 対象プローブセット: 50%:50%において、10000以上の値を持つ



高値に対する補正を行う変形をすると、若干安定する

全プローブ計算⑥検証① 高値補正付き

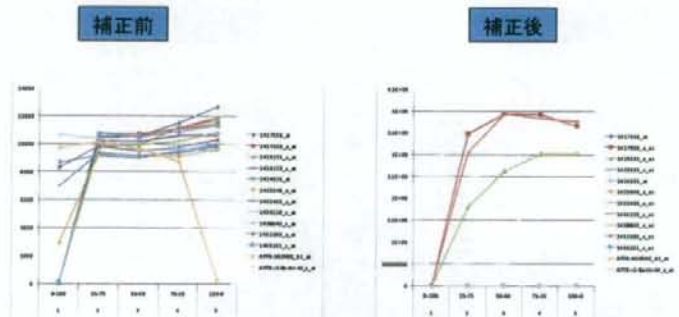
- 対象プローブセット: 50%:50%において、10000以上の値を持つ



高値に対する補正を行う変形をすると、若干安定する

全プローブ計算⑥検証① 飽和値指定

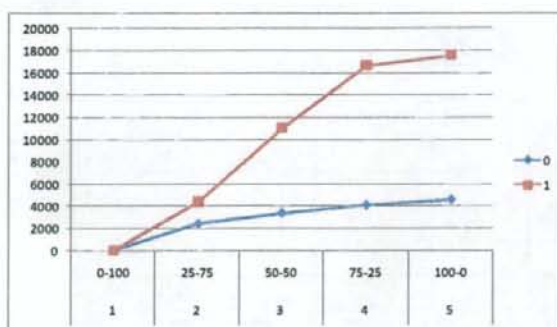
- 対象プローブセット: 50%:50%において、10000以上の値を持つ



非線形最適化で失敗している。丹念に調べる必要がある

全プローブ計算⑥検証①

- Cyp1a2(1450715_at)

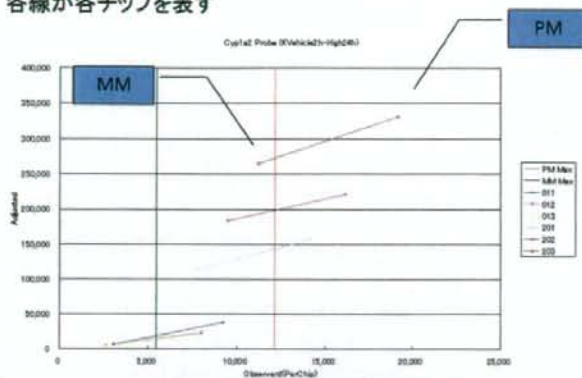


全プローブ計算⑥検証②

- 基本的確認
 - Percellome論文(BMC Genomics 2006, 7:64 doi:10.1186/1471-2164-7-64)において、QPCR計測を行っている遺伝子で飽和していると考えられる遺伝子が、本方法と、結果が合致しているかを確認する。
- 課題
 - QPCR対象のプローブセットがLangmuirモデルになっているか？
 - Cyp7a1は、Langmuirモデル対象となっている。
 - サーフエースまで描いた遺伝子はLangmuir対象は他になし。
 - 針刺済みでは、他にあるか？
- 懸念事項
 - Percellome自身が飽和の影響を受けていて、高発現域でBiasを発生していないか？
- 対象データ
 - QPCRと同じ化合物を用いた実験のCELファイル
 - TTG20を使用

全プローブ計算⑥検証②プローブ単位

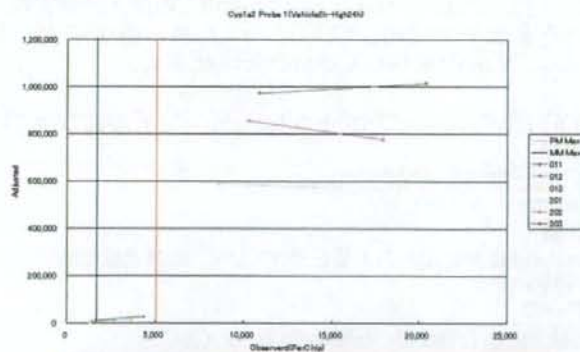
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥検証②プローブ単位

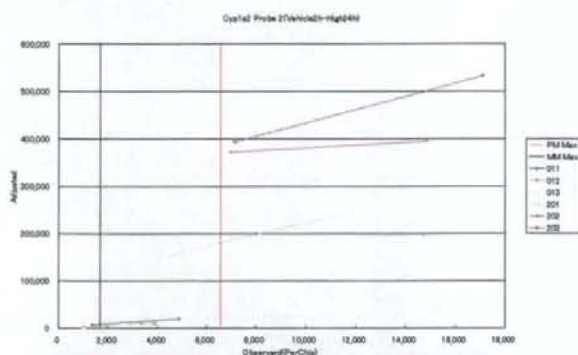
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥検証②プローブ単位

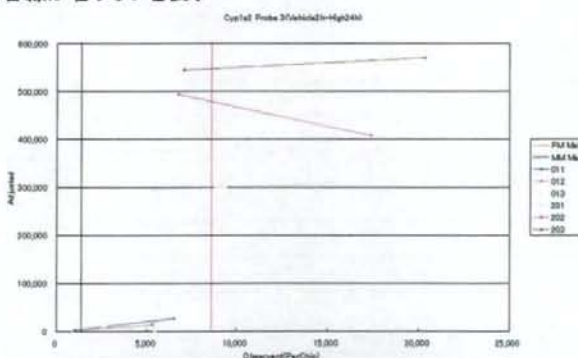
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥検証②プローブ単位

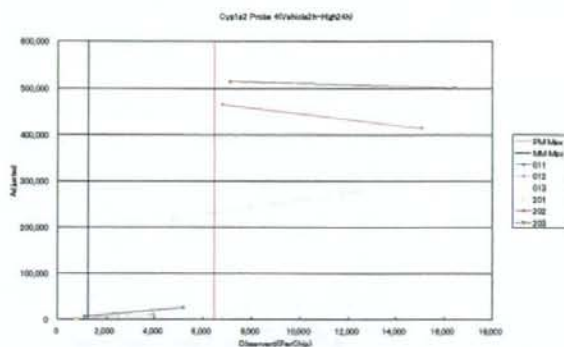
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥検証②プローブ単位

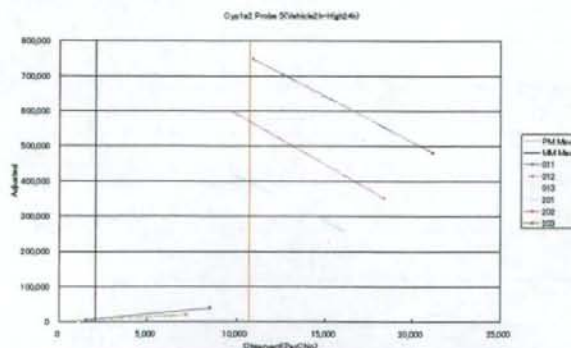
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥検証②プローブ単位

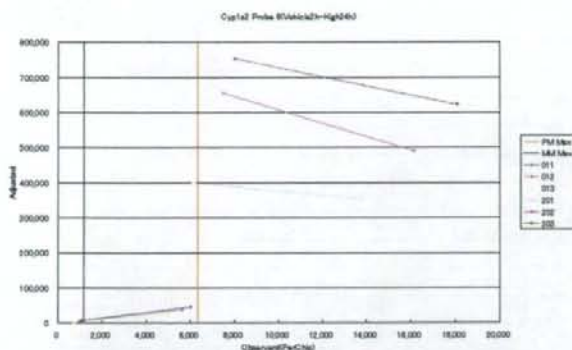
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



PM-MMで値が逆転しており、結果を不安定にしている

全プローブ計算⑥検証②プローブ単位

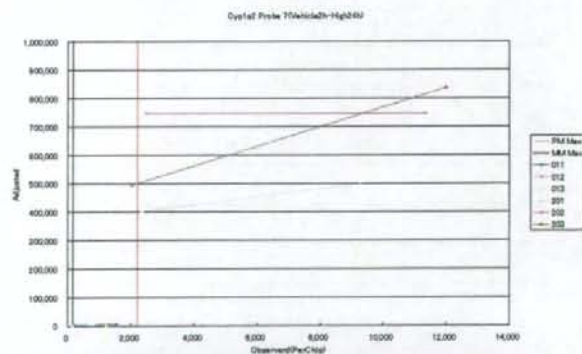
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥検証②プローブ単位

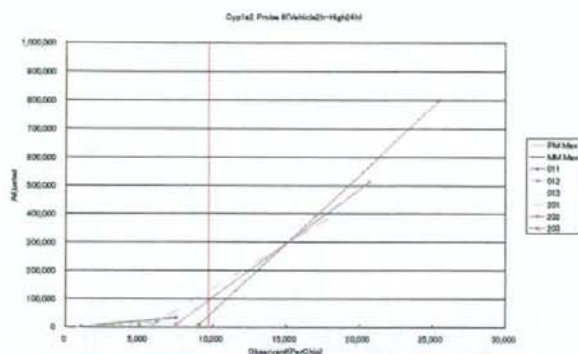
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥検証②プローブ単位

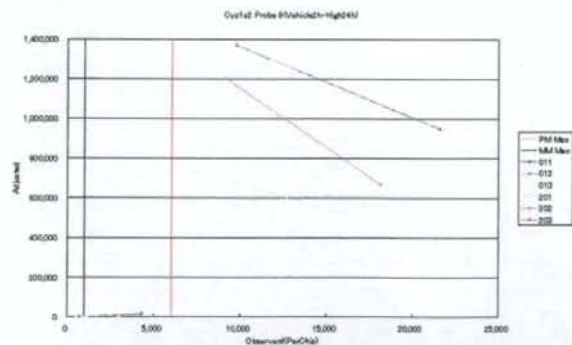
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



MMはLangmuirが選択されていない

全プローブ計算⑥検証②プローブ単位

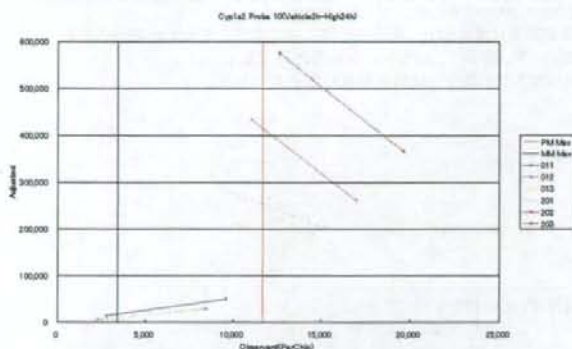
- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥検証②プローブ単位

- Cyp1a2のプローブ単位でチェック
 - 各線が各チップを表す



計測値はLangmuir曲線の最大値を大きく上回っている。

全プローブ計算⑥課題:変換範囲

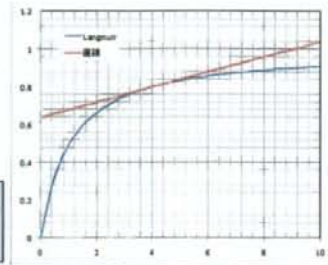
- Cyp1a2において、変換結果として良好な形状を示さなかった。
- 考察
 - Liverサンプルであり、TCDDの影響により通常状態より大きな値を示している
 - 大きな値で、PMとMMで、実測値よりも補正值で大小関係が逆転する場合もみられ、不安定になっている。
- 対策
 - 不安定性を解消するために、高値補正として、計算結果が安定する直線を用いてみる

全プローブ計算⑦

- アイデア
 - Langmuir方程式と直線を組み合わせる
- 数値計算を行う上で、Langmuirの方程式は上に有界な数式である。次の2点で数値計算上問題がある。
 - 濃度に対する係数に鈍感な関数で、完全飽和に近い状態で適切な収束が得られない。
 - 飽和していない場合には、AICにより、排除される
 - 今回は逆関数にするので、逆関数の際に右に有界で定義域が制限される。
- 高濃度において、直線にフィットするものとみなす。
 - この領域に入った場合には詳細な検討が必要である。

$$f(x) = \begin{cases} \frac{x}{x+1} & \text{for } x \leq 4 \\ 0.04x + 0.04 & \text{for } x \geq 4 \end{cases}$$

$$I_{N'} = \begin{cases} I_s \frac{k_s C_i}{k_s C_i + 1} + bG_s & \text{for } k_s C_i \leq 4 \\ 0.04 I_s k_s C_i + bG_s & \text{for } k_s C_i \geq 4 \end{cases}$$



これまで、高濃度域で、放物線を使用してきた。直線を使用することで、安定性が増すことが考えられる

全プローブ計算⑦ 結果

非線形最適化の結果

- 現在までの非線形モデルよりも非線形計算の安定性がよくない。非線形最適化計算中も数値計算上のエラーで多く停止した

終了状態	Brain側飽和	Liver側飽和
勾配ベクトル判定収束	74,275	66,007
移動量判定収束	40,917	26,076
局所的最小値	34,491	27,883
繰り返し回数が上限を超えた	86	90
総計	149,769	120,056

収束した

局所的収束

収束失敗

今までの補正式と比較して収束の安定性が悪い

全プローブ計算⑦ 結果

Langmuir Model採用個数

- ProbeSet中で何個のProbeがLangmuirモデルを採用しているか
- Brain側飽和

PM

PM

合計 / Count(*)	列ラベル	0	1	2	3	4	5	6	7	8	9	10	12	総計
0				3908	1956	1021	421	189	71	9	3			7578
1		4837	3627	2346	1321	703	284	89	32	7	1			13447
2		2453	2256	1591	969	519	246	108	31	11	1			8189
3		1308	1152	888	547	311	144	53	15	6	2		1	4407
4		686	593	409	235	98	53	21	9	4				2108
5		348	270	153	77	32	13	2	2			1		898
6		109	73	45	19	11	5	1						263
7		23	18	16	3	6	1	1						68
8		7	5	1	1	1								15
9		1	2			1								4
総計		9772	12106	7385	4193	2103	937	346	98	31	4	1	1	36977

PM側で11プローブが対象となるものがなく、適切な補正は難しい

全プローブ計算⑦ 結果

Langmuir Model採用個数

- Liver側飽和

PM

PM

合計 / Count(*)	列ラベル	0	1	2	3	4	5	6	7	8	9	10	11	総計
0			4486	1161	321	161	103	55	46	18	5	7	1	6384
1		6341	2872	937	380	194	117	83	49	39	12	4	2	11010
2		1226	861	403	222	156	113	117	71	37	25	6	1	3238
3		162	163	127	100	106	108	91	68	40	24	13	1	1003
4		30	58	58	89	60	122	91	77	56	27	9	3	660
5		18	36	46	66	83	78	76	73	36	23	7		543
6		21	39	58	81	88	63	62	44	29	18	7	4	495
7		13	35	35	47	47	38	50	29	20	17	4		335
8		4	19	37	32	46	47	35	20	17	9			266
9		2	6	21	16	17	23	21	17	8	4	1		136
10		3	3	5	10	3	7	8	9	4				52
11		1	2			2	2	2		1	2			12
総計		7822	8580	2889	1304	963	821	691	503	305	166	58	12	24114

PM側で11プローブが対象となるものが少なく、適切な補正は難しい

全プローブ計算⑦検証①

- LBMでの線形性チェック
 - 50:50を1に正規化したグラフを作成し、線形性のチェック
 - Langmuir変換を含むプローブセットを全てプロットし、目視で線形性の確認する。値の低いもの以外で変な値を示すものがないことを確認する

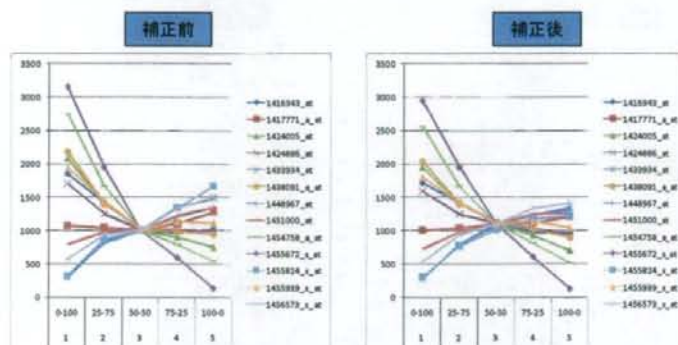
全プローブ計算⑦検証①蝶の羽(1)

- 対象プローブセット:50%:50%において、1000程度の値を持つ

affy_id	Intensity
1424886_at	1008.53
1433934_at	1004.20
1424005_at	1002.17
1417771_a_at	1005.90
1451000_at	1000.37
1456573_x_at	1005.87
1455824_x_at	1006.07
1416943_at	1007.17
1454758_a_at	1007.60
1438091_a_at	1008.33
1448967_at	1003.97
1455939_x_at	1006.53
1455672_s_at	1004.00

全プローブ計算⑦検証①蝶の羽(2)

- 対象プローブセット: 50%:50%において、1000程度の値を持つ



一部のプローブで飽和に近い曲線になっている

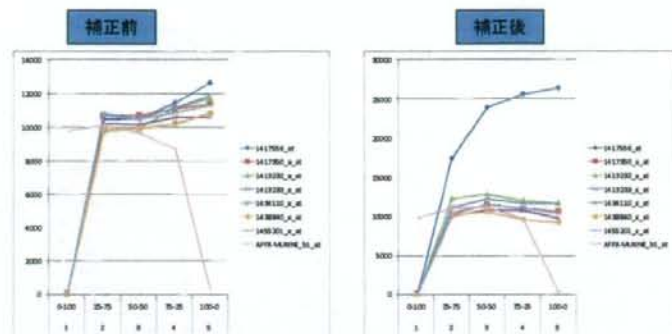
全プローブ計算⑦検証①蝶の羽(1)

- 対象プローブセット: 50%:50%において、10000程度の値を持つ

affy_id	Intensity
1417556_at	10550.2
1417950_a_at	10735.1
1419232_a_at	9940.1
1419233_x_at	10193.0
1434110_x_at	10519.8
1438840_x_at	9940.5
1455201_x_at	10504.2
AFX-MURINE_b1_at	9686.6

全プローブ計算⑦検証①蝶の羽(5)

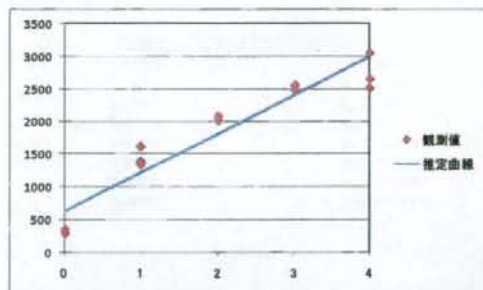
- 対象プローブセット:50%:50%において、10000以上の値を持つ



141756_atは、補正により、引き延ばすことができたと考え

全プローブ計算⑦状況確認&課題

- あるプローブにおいて、収束結果後のLangmuir曲線の状況を図化し把握した。



課題:高値の補正用直線部分しか現われていない。

補正係数を求めるときと、補正計算を実施する際に高値補正の扱いを変更すべきと考えられる

全プローブ計算⑧

数値計算上の技巧 Langmuir方程式と直線を組合せる

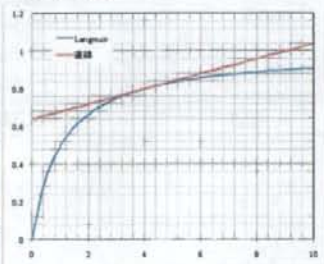
- 数値計算を行う上で、Langmuirの方程式は上に有界な数式である。次の2点で数値計算上問題がある。
 - 濃度に対する係数に鈍感な関数で、完全飽和に近い状態で適切な収束が得られない。
 - 飽和していない場合には、AICにより、排除される
 - 今回は逆関数にするので、逆関数の際に右に有界で定義域が制限される。
- 高濃度において、直線にフィットするものとみなす。
 - この領域に入った場合には詳細な検討が必要である。

$$f(x) = \begin{cases} \frac{x}{x+1} & \text{for } x \leq 4 \\ 0.04x + 0.64 & \text{for } x \geq 4 \end{cases}$$

$$I_p = \begin{cases} I_s \frac{k_p c_j}{1 + k_p c_j} + b g_p & \text{for } k_p c_j \leq 4 \\ I_s (0.04 k_p c_j + 0.64) + b g_p & \text{for } k_p c_j \geq 4 \end{cases}$$

これまで、高濃度域で、放物線を使用してきた。直線を使用することで、安定性が増すことが考えられる。

非線形最適化の収束計算の安定性のため、本式のパラメータ推定では、直線部分がないものとして計算を行う。



全プローブ計算⑧

Langmuir等温式による補正値を求める式

- LBMなどで、プローブ毎に、各係数を求めておき、各実験で、補正値を計算する

$$I = I_s \frac{k_p c_j}{1 + k_p c_j} + b g_p$$

I : Observed Intensity

I_s : Saturation Intensity

k : equilibrium constant

c : concentration

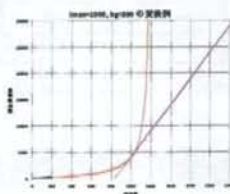
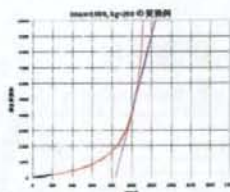
$b g$: background component

$$k_p c_j = \frac{I - b g_p}{I_p - I}$$

と求め、

$$I = I_p k_p c_j + b g_p$$

を補正値とする



高濃度における補正 ($k_p c_j \geq 4$ の場合)

$$I = I_s (0.04 k_p c_j + 0.64) + b g_p$$

$$k_p c_j = \frac{\left(\frac{I - b g_p}{I_p} \right) - 0.64}{0.04}$$

と求め、

$$I = I_p k_p c_j + b g_p$$

を補正値とする