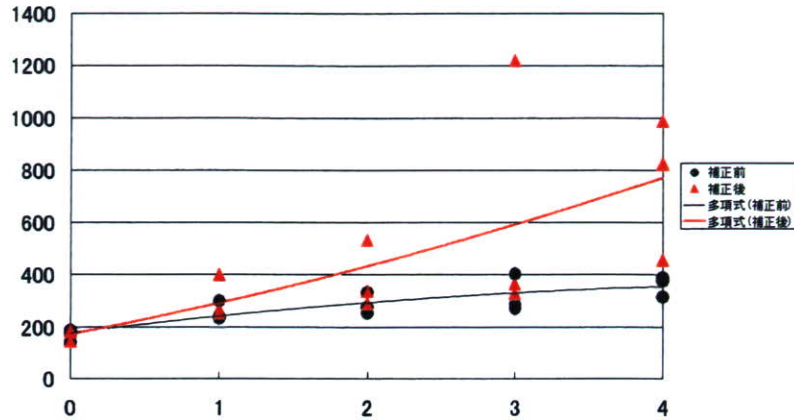


1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe9



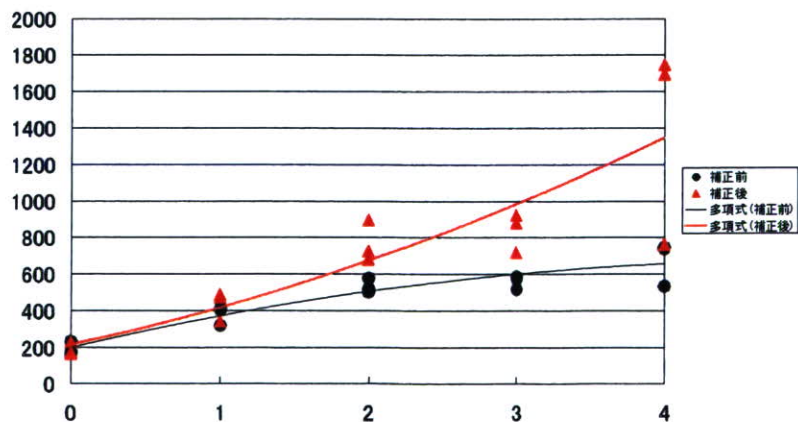
補正がうまくいっているようにみえる

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe10



補正がうまくいっているようにみえる

Copyright © NTT COMWARE 2008

1.6. 全プローブ計算②



- 全プローブに対して計算を一通り可能であるか確認した。

Copyright © NTT COMWARE 2008

1.6.1. 全プローブ②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た。

	Liver側飽和	Brain側飽和
計算	87,288	40,299
Langmuir採用	35,161	246
失敗:収束回数超過	0	1
失敗:σ減少せず	708	1,548

Copyright © NTT COMWARE 2008

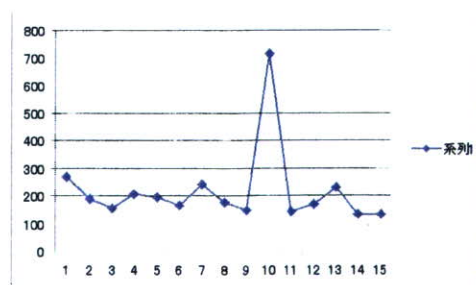
1.6.2. 全プローブ計算②収束失敗(収束回数超過)のプローブ

- Brain側飽和条件のプローブ28083/mm/7はLangmuirモデルで、収束回数が超過するという結果になった。一実験だけ外れた値を持っているために収束失敗したと考えられる。

Copyright © NTT COMWARE 2008

1.6.2. 全プローブ計算②収束失敗(σ 減少不十分)のプローブ

- パラメータ数を増加させれば、標準偏差は減少するはずであるが、はずれ値を含むなどの場合にはうまくいかないことも考えられる。このような場合には、Langmuirモデルが採用できなかったものとみなす。新たなLBMで係数決定を行えばLangmuirモデルが採用される可能性が高い。



失敗プローブ例
4806-12w-9

Copyright © NTT COMWARE 2008

1.6.3. 全プローブ計算② 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

		MM Langmuir Model個数			総計
		1	2	11	
PM Langmuir Model数	3	2			2
	5	2			2
	10		1		1
	19			1	1
総計		4	1	1	6

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
5個以上と考えると、4ProbeSet

1.6.3. 全プローブ計算② Langmuir Model採用モデル

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

		MM Langmuir Model個数											総計			
		1	2	3	4	5	6	7	8	9	10	11				
PM Langmuir Model数	1	855	97	5												957
	2	204	51	4												259
	3	89	28	6	1	1		1								126
	4	74	28	12	2											116
	5	56	39	22	4	1										122
	6	57	40	22	3	1	2									125
	7	52	46	32	10	6	2	1								149
	8	30	37	52	23	13	3	4				1				163
	9	25	33	40	37	23	20	8	3	3						192
	10	12	22	34	32	38	30	37	21	13	9	2				250
	11	4	7	13	21	43	47	65	81	81	83	105				550
	12							1								1
	15			1												1
	総計		1458	428	243	133	126	105	116	105	97	93	107			3011

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
5個以上と考えると、1500個程度

1.7. 全プローブ計算③



- 非線形最適化の課題を解決を試みて全プローブの補正計算結果を求め、サマライズ(あるプローブセットに属するプローブ群の測定値を取りまとめ、プローブセットとしての測定値を算出すること)後の結果に対する考察を試みる

Copyright © NTT COMWARE 2008

1.7.1. 全プローブ計算③非線形最適化の課題 (Langmuirモデル)



- 収束結果判定
 - 結果コード=2の場合、 k_p が大きすぎると失敗
 - $StepTol=1.0e-17$ として、完全不動の状態でしか収束させない
- 初期値
 - I_p は、最大値と最小値の差を与える
 - 増加の場合、100:0の値に100:0と25:75の差を加えて、最大値のおおよその値を見つける
 - B_{gp} は、最小値の半分

Copyright © NTT COMWARE 2008

1.7.1. 全プローブ計算③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	飽和側	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	BP0を半分に
	Brain	Linear	1680	PM	10	BP0を半分に
	Brain	Linear	35906	PM	5	BP0を半分に
PMMM	Liver	Langmuir	11913	pm	4	I00=92, k00=1.11, BP0=11.9
	Brain	Linear	15256	pm	4	BP0を半分に

今後も発生する可能性は高いが、既に出来上がったモデルのパラメータを与えることで、回避可能と考える。

Copyright © NTT COMWARE 2008

1.7.2. 全プローブ計算③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た。
- 標準偏差が減少しないという、理論と外れた結果が存在した。
 - 収束先の存在しないモデルに対して最適化を実施した可能性が高い

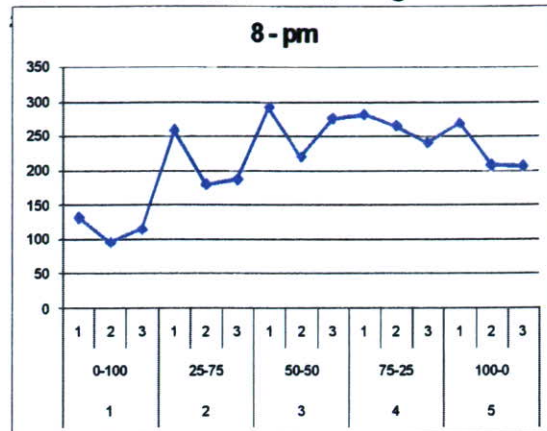
	Liver側飽和	Brain側飽和
計算	109,879	130,706
Langmuir採用	77,684	9,094
失敗: σ 減少せず	127	8

Copyright © NTT COMWARE 2008

1.7.3. 全プローブ計算③収束失敗(σ 減少不十分)のプローブ



- パラメータ数を増加させれば、誤差は減少するはずであるが、はずれ値を含むなどの場合にはうまくいかないことも考えられる。このような場合には、Langmuirモデルが採用できなかった



例: 2811-8-PM

Copyright © NTT COMWARE 2008

1.7.4. 全プローブ計算③ Langmuir Model採用個数 (Brain側飽和)



- 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(*)	In MM				13 ?	総計
n PM	1	2	3			
1	409	45	5		3083	3542
2	73	4	3		314	394
3	14				45	59
4	4	2			9	15
5	2	2			1	5
6		2			1	3
7	1		1			2
8	1	2				3
10		1				1
20					1	1
?	3194	257	16		3453	7492
総計	3698	315	25	1	3453	7492

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

Copyright © NTT COMWARE 2008

1.7.4. 全プローブ計算③ Langmuir Model採用個数 (Liver側飽和)

- Liver側飽和

合計 / Count(*)	n MM													総計	
n PM	1	2	3	4	5	6	7	8	9	10	11	12	13 ?		
1	2639	663	89	13	1									4277	7682
2	780	345	57	11	1									939	2133
3	286	130	41	6	2				1					299	765
4	159	85	41	16	2	2	1							126	432
5	100	95	40	23	10	3	1							84	254
6	93	82	50	22	16	10								53	326
7	78	77	56	47	22	10	4	1						50	345
8	66	67	67	49	40	22	12	5	1					27	356
9	31	57	65	67	47	41	27	14	6	2	1			9	367
10	21	31	50	73	59	73	75	59	37	15	7			7	507
11	3	13	21	42	63	83	117	154	166	190	265			3	1100
15						1								1	1
16								1						1	2
19														1	1
?	6005	925	86	10											7026
総計	10263	2570	672	378	263	225	237	234	211	207	274	1	1	5871	21407

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

Copyright © NTT COMWARE 2008

1.7.5. 全プローブ計算③ Langmuir補正時エラーファイル

- 飽和モデルを用いて次の状態となった場合、エラーファイルに出力する
 - モデル作成LBMの最小最大の範囲を逸脱した
 - モデル作成LBMの3σの範囲を逸脱した
 - Langmuirモデルが採用されているプローブでシグモイド関数の(0,1)の範囲を逸脱した
 - バックグラウンドレベル(b_{gp})を下回る
 - 飽和レベル(I_p)を上回る
- モデル作成LBMの範囲を逸脱している状態を参考に、追加すべき標準データの作成を考える
- Langmuirモデル範囲の逸脱に関して、何らかの対処が必要

Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③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異常な値を示すものがないことを確認する

Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③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

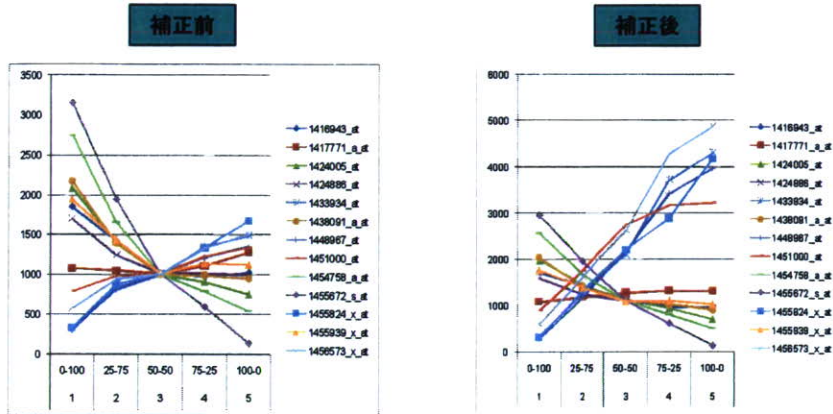
ally_id	Intensity
1424886 at	1008.53
1433934 at	1004.20
1424005 at	1002.17
1417771 a at	1005.90
1451000 at	1000.37
1456573 x at	1005.87
1455824 x at	1006.07
1416943 at	1007.17
1454758 a at	1007.60
1438091 a at	1008.33
1448967 at	1003.97
1455939 x at	1006.53
1455672 s at	1004.00

Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③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



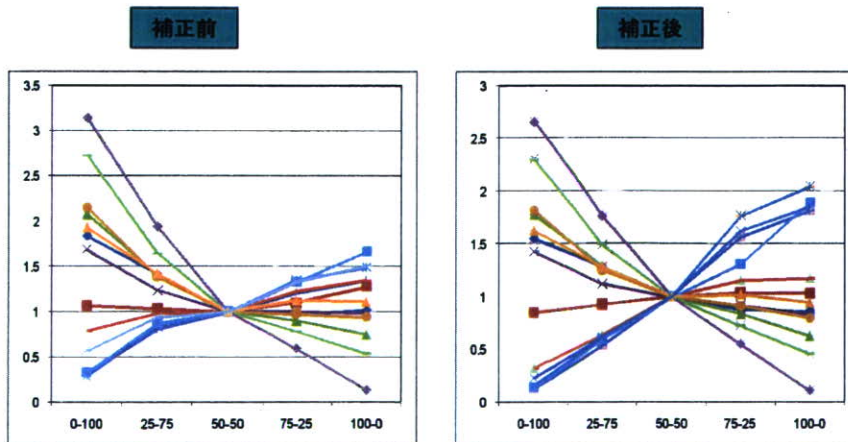
Liver 側値和のプローブは直線に引き延ばされ、高い値で、不安定になっている。

Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③検証①



- 50:50で1に統一

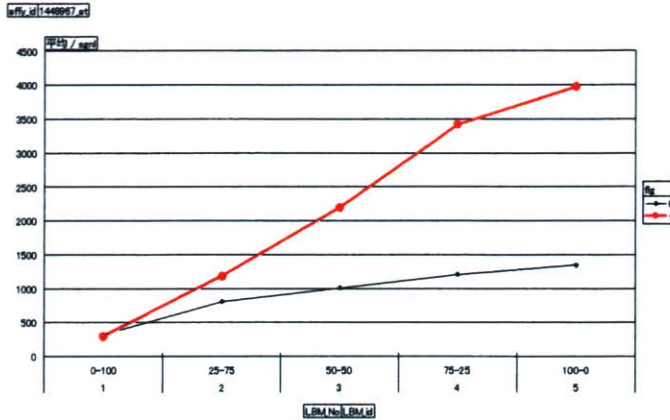


Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③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
- 線形に引き延ばされ、高い値で不安定になっており、補正がある程度成功したと考えられる例(1448967_at:Nipsnap3a)



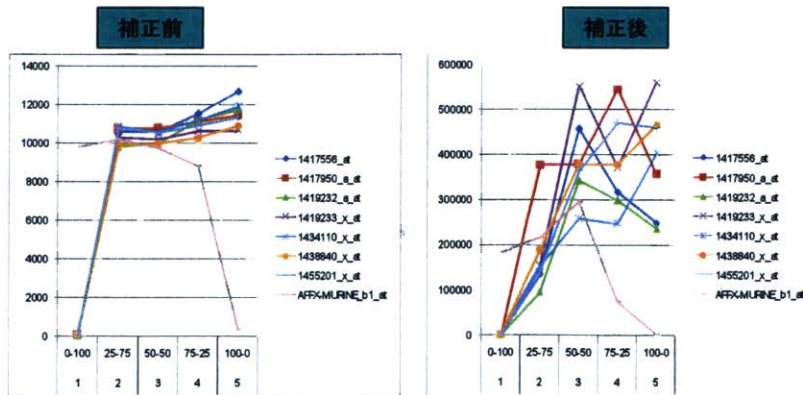
Liver 側飽和のプローブは直線に引き延ばされ、高い値で、不安定になっている。

Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③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



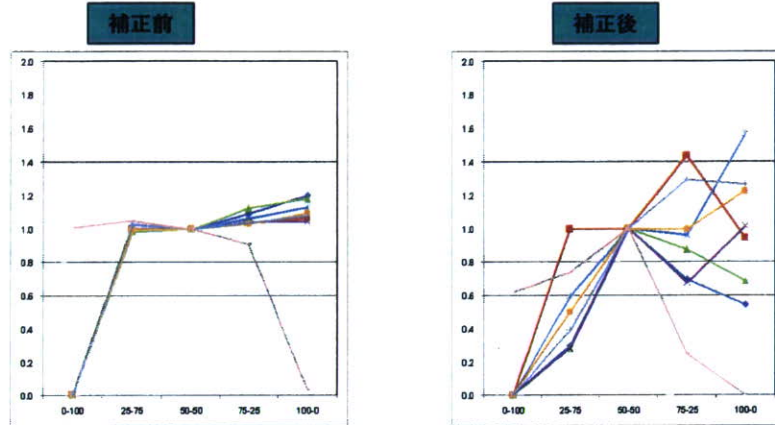
0%から50%まではすこし上向きだが直線に見えなくもない。しかし、50%を超えると、崩れている。Lngmirの最大値推定に問題がありそうである

Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③検証①



- 50%:50%において、1に正規化



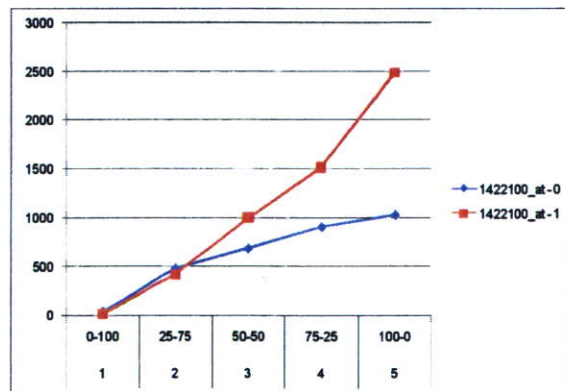
50%を超えるものでは、適切な補正ができているとは言い難い

Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③検証①蝶の羽(7)



- Cyp7a1(1422100_at)で検証した



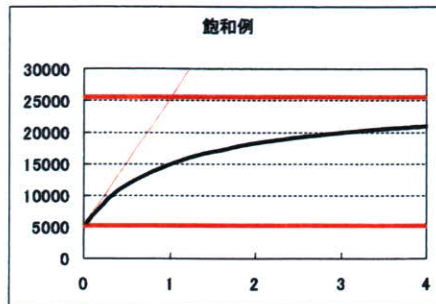
直線に引き延ばされたと考えられる。しかし、100%-0%では、行き過ぎた感がある

Copyright © NTT COMWARE 2008

1.7.6. 全プローブ計算③検証①課題



- 高い値を示す場合に、Langmuirの最大値を超える状況が発生した。推定されるLangmuirの最大値を引き上げれば、状況は少し改善する。



27000の場合にどうするか

- 現在、誤差項の計算を対数領域で行っており、同一条件内で平均をとると、算術平均よりも小さな値となる。算術平均に置き換えるだけで大きな値側にぶれる。

Copyright © NTT COMWARE 2008

1.7.7. 全プローブ計算③ 検証②



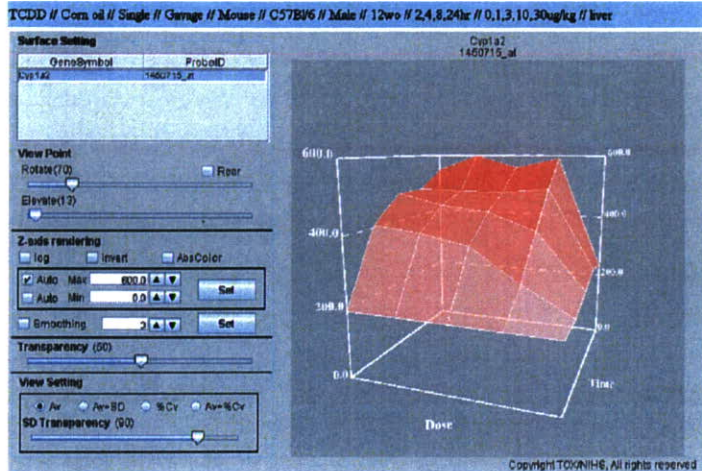
- 基本的確認
 - Percellome論文(BMC Genomics 2006, 7:64 doi:10.1186/1471-2164-7-64)において、QPCR計測を行っている遺伝子で飽和していると考えられる遺伝子が、本方法と、結果が合致しているかを確認する。
- 課題
 - QPCR対象のプローブセットがLangmuirモデルになっているか？
 - Cyp7a1は、Langmuirモデル対象となっている。
 - サーフェースまで描いた遺伝子はLangmuir対象は他になし。
- 懸念事項
 - Percellome自身が飽和の影響を受けていて、高発現域でBiasを発生していないか？
-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

Copyright © NTT COMWARE 2008

1.7.7. 全プローブ計算③検証②Cyp1a2



- Cyp1a2(1450715_at)のTTG20の結果(ToxicOmics Database)



Copyright © NTT COMWARE 2008

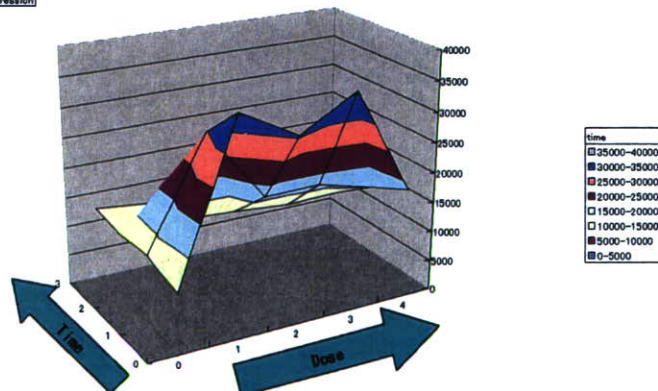
1.7.7. 全プローブ計算③検証②Cyp1a2



- Cyp1a2(1450715_at)のTTG20の変換結果

PM(CVC)

平均 / expression



前ページの図とは、Dose軸の正負が異なるグラフである。大きな値に対して、小さな値を生み出している

Copyright © NTT COMWARE 2008

1.7.8. 全プローブ計算③見つかった課題: 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



LBMにおけるRNAの濃度範囲を逸脱している



本来のLangmuir方程式より低い値で飽和しているような値が推定されている



誤差を対数領域で計算する(幾何平均)と
線形領域で計算する(算術平均)よりも小さな値を示す

Copyright © NTT COMWARE 2008

1.7.8. 全プローブ計算③見つかった課題: Summarize (Probe → ProbeSet)



- MAS5は、Tukey's biweight algorithmを使用している
- GCOSとSDK(=Expression Console)の間で違う結果となる。
 - Absentで多少の値が違うというレベルではなく、Presentで大きく値がずれているものが存在した。
 - 調査の結果、コンパイラのバージョンの違いによる計算誤差が原因と判明した。
- アルゴリズム
 1. ProbePair値計算: $x = PM - MM$
 2. Medianを求める
 3. Median absolute deviation: 標準偏差計算の中で、2乗計算と平均を絶対値とMedianで置き換えたもの
 4. 標準化: MedianとMADを用いたZScore化: $t = (x - M) / (c * MAD)$
 5. 重み $w = ((1 - t^2)^2)$
 6. 重み付き平均
 - Median近辺の平均値を用いることで、極端な値の影響を排除している
- 今後対処方法を検討する

Copyright © NTT COMWARE 2008

1.8. 全プローブ計算④



- 数値計算のテクニックを付加することで、安定した補正を行えるか検討する

Copyright © NTT COMWARE 2008

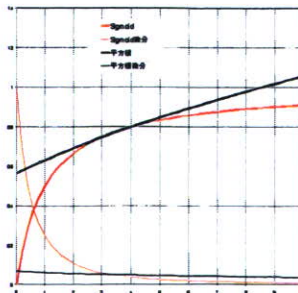
1.8.1. 全プローブ計算④数値計算上の技巧



- 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、平方根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。

$$f(x) = \begin{cases} \frac{x}{x+1} & \text{for } x \leq 4 \\ \frac{1}{5\sqrt{0.5}}\sqrt{x+4} & \text{for } x \geq 4 \end{cases}$$

$$I_N = \begin{cases} I_p \frac{k_p c_j}{k_p c_j + 1} + b g_p & \text{for } k_p c_j \leq 4 \\ \frac{I_p}{5\sqrt{0.5}}\sqrt{k_p c_j + 4} + b g_p & \text{for } k_p c_j \geq 4 \end{cases}$$



Copyright © NTT COMWARE 2008

1. 8. 2. 全プローブ計算④ Teradataによる線形領域モデル計算を用いた絞り込み

- 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	101,432	82,443
移動量判定収束	35,281	23,624
局所的最小値	13,051	13,989
繰り返し上限を超えた	5	0
総計	149,769	120,056

AICでLangmuirは選
択されない

Copyright © NTT COMWARE 2008

1. 8. 3. 全プローブ計算④非線形最適化の課題 (Langmuirモデル)

- Rによる収束計算における課題について記す
- 収束結果判定
 - 収束判定係数 $Steptol=1.0e-17$ として、完全不動の状態でしか収束させないようにする。
- 初期値
 - データから計算して求め、収束しやすいようにする。
 - 計算方法
 - I_{pl} は、最大値と最小値の差を与える
 - 増加の場合、100:0の値に100:0と25:75の差を加えて、最大値のおおよその値を見つける
 - 数値計算上の技巧を付け加えたことにより、若干小さいほうが収束しやすい可能性がある
 - B_{gpl} は、最小値の半分

Copyright © NTT COMWARE 2008

1.8.4. 全プローブ計算④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た

終了状態	Brain側飽和	B側採用個数	Liver側飽和	L側採用個数
勾配ベクトル判定収束	101,432	364	82,443	29,109
移動量判定収束	35,281	155	23,624	8,390
局所的最小値	13,051	109	13,989	7,775
繰り返し上限を超えた	5	0	0	0
総計	149,769	628	120,056	45,274

Copyright © NTT COMWARE 2008

1.8.4. 全プローブ計算④ 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を検討した
- Brain側飽和

合計 / Count(*) 行ラベル	MM				総計
	列ラベル 0	1	2	11	
0		152	3		155
1	255	6			261
2	30				30
3	5	2			7
4	-	-	-	-	-
5	3	1	1		5
6	2	3	1		6
11	1				1
20				1	1
総計	297	185	5	1	468

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

Copyright © NTT COMWARE 2008

1.8.4. 全プローブ計算④ Langmuir Model採用個数

- Liver側飽和

MM

合計 / Count(*) 行ラベル	列ラベル											総計	
	0	1	2	3	4	5	6	7	8	9	10		11
0		3565	191	15									3771
1	3231	930	112	12									4285
2	546	285	55	7									893
3	263	133	52	9	2								459
4	160	101	47	20	7								335
5	120	94	53	26	15	2			1				311
6	57	58	58	32	18	9	1						263
7	34	75	65	56	17	15	2	1					265
8	32	57	45	51	32	19	15	5					256
9	8	31	35	59	49	41	33	24	5				285
10	1	17	16	54	61	45	58	56	40	24	14	3	389
11	2	6	14	18	22	40	75	72	75	85	110	96	615
16								1					1
総計	4454	5382	743	359	223	171	184	159	121	109	125	99	12129

PM

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

Copyright © NTT COMWARE 2008

1.8.5. 全プローブ計算④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

Copyright © NTT COMWARE 2008

1.8.5. 全プローブ計算④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

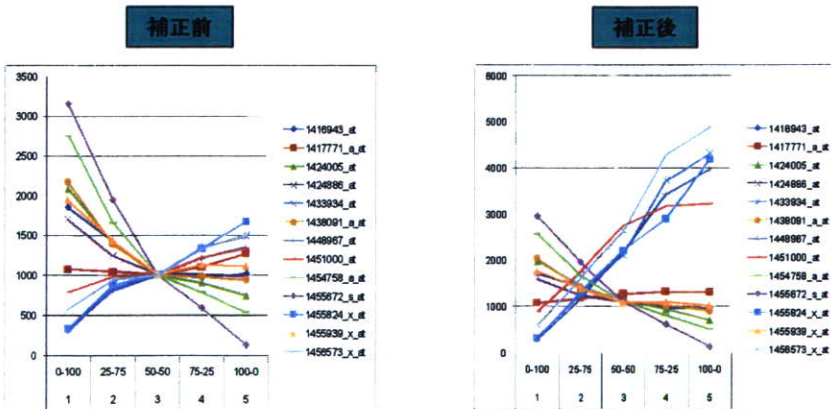
affy id	Intensity
1424886_at	1008.53
1433934_at	1004.20
1424005_at	1002.17
1417771_a_at	1005.90
1451000_at	1000.37
1456573_x_at	1005.87
1455824_x_at	1006.07
1416943_at	1007.17
1454758_a_at	1007.60
1438091_a_at	1008.33
1448967_at	1003.97
1455939_x_at	1006.53
1455672_s_at	1004.00

Copyright © NTT COMWARE 2008

1.8.5. 全プローブ計算④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 側値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。

Copyright © NTT COMWARE 2008