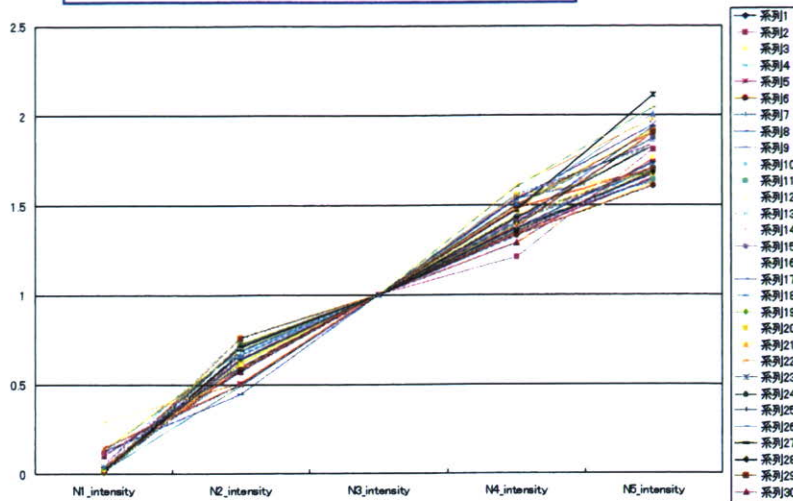


1.2.2. データ状況把握(線形的増加)



線形性が認められるプローブセットを見つけ出す



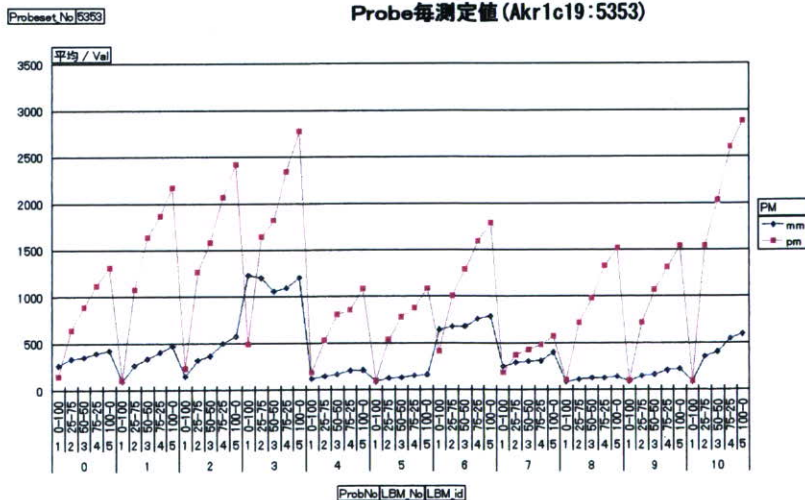
Copyright © NTT COMWARE 2008

1.2.2. データ状況把握(線形的増加)



最もきれいな直線を描くProbesetでも、Probe単位ではブレがある。

Probe毎測定値 (Akr1c19:5353)

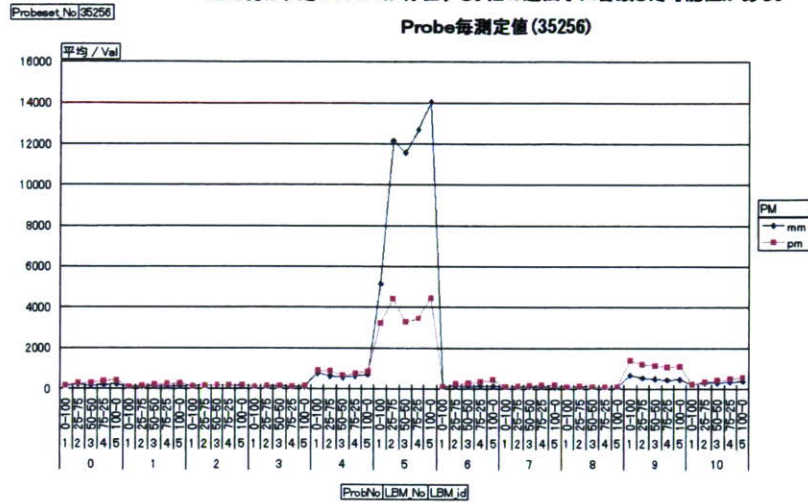


Copyright © NTT COMWARE 2008

1.2.2. データ状況把握(線形的増加)



■の方が大きいProbeが存在する。他の遺伝子に合致した可能性がある。

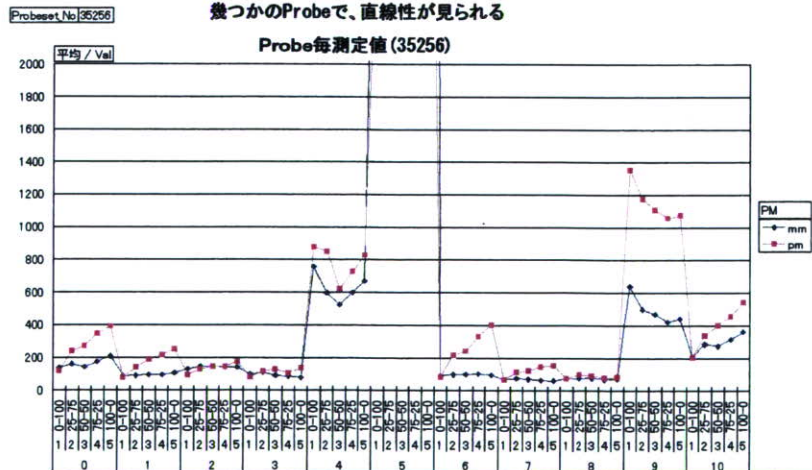


Copyright © NTT COMMWARE 2008

1.2.2. データ状況把握(線形的増加)



幾つかのProbeで、直線性が見られる



■の大きなプローブは無視され、小さい値の線形性がみられるプローブから、線形性の強い遺伝子として検出された。

Copyright © NTT COMMWARE 2008

1.3.1. 基本的アイデア①:

Langmuir isotherm equation

- Irving Langmuirによって1918年に導出された理論的な吸着等温式である。以下のような仮定を持っている。
 - 吸着媒には有限な数Nの吸着サイトがあり、そこだけで吸着質分子と結合する。
 - すべての吸着サイトは等価である。
 - 1つの吸着サイトは1つの吸着質分子としか結合しない。
 - 空の吸着サイトM、気相中の吸着質S、吸着サイトに結合した吸着質M-Sの間に $M + S \leftrightarrow M-S$ の化学平衡が成立する。
- 各プローブが飽和している場合には、Langmuirの方程式に従うと仮定する

$$\log(PM_{p,l}) = \log\left(I_p \frac{k_p c_j}{1 + k_p c_j} + bg_p\right) + \varepsilon_{p,l}$$

$p = 1, \dots, P$: probe_index

j : concentration_index

c : concentration

$l = 1, \dots, L$: replicate_index

I : Saturation_Intensity

k : equilibrium_constant

bg : background_component

Copyright © NTT COMWARE 2008

1.3.2. 基本的アイデア②

AIC(Akaike Information Criteria)

- すべてのプローブが飽和しているとは考えられない。また、飽和していたとしても一定の値を示しているため、飽和しているかを判別できないことが考えられる。これらの状況を判断するため、Langmuirモデル、線形モデル、定数モデルを選択するためにAICを使用する
- AICとは、元統計数理研究所所長の赤池弘次によって1971年に考案された統計モデルの良さを示す指標である
 - モデルは複雑にすれば細かく適合させることができるが、偶然の影響を受ける可能性も高くなる
 - Akaike, Hirotugu (1974). "A new look at the statistical model identification". IEEE Transactions on Automatic Control 19 (6): 716-723.

$$AIC = -2 \log(L) + 2k$$

L : 尤度

k : パラメータ数

$$\text{確率密度関数: } P_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\text{尤度関数: } L_i(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

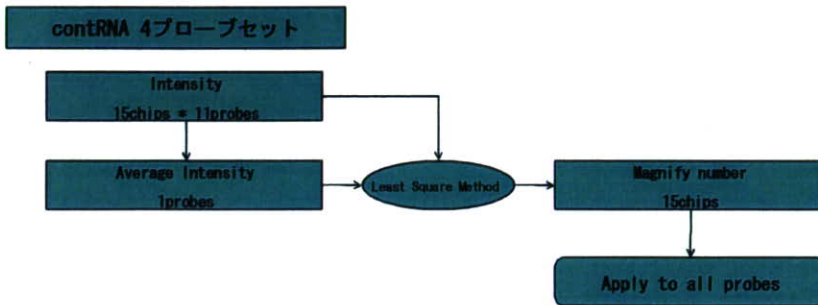
$$\begin{aligned} L_n(\mu, \sigma) &= \prod_{i=1}^n L_i(\mu, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ \log L_n(\mu, \sigma) &= \log\left(\prod_{i=1}^n L_i(\mu, \sigma)\right) = \log\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) \\ &= \sum_{i=1}^n \log\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) \\ &= \sum_{i=1}^n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \sum_{i=1}^n \log\left(e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) \\ &= -n \log \sqrt{2\pi}\sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

Copyright © NTT COMWARE 2008

1.3.3. チップごとの標準化(contRNA)



- チップごとの標準化のためにControl用RNAのデータ(contRNA)を使用する
 - 各プローブに対する平均値を求め、それらの平均値に近づくように(最小二乗法)にチップごとの変換倍率を求め、その値をチップの全プローブにかける。
 - 対数領域で計算する
 - ContRNAはハイブリ液に一定量を加えたE.Coli.由来のmRNAの測定データを示す



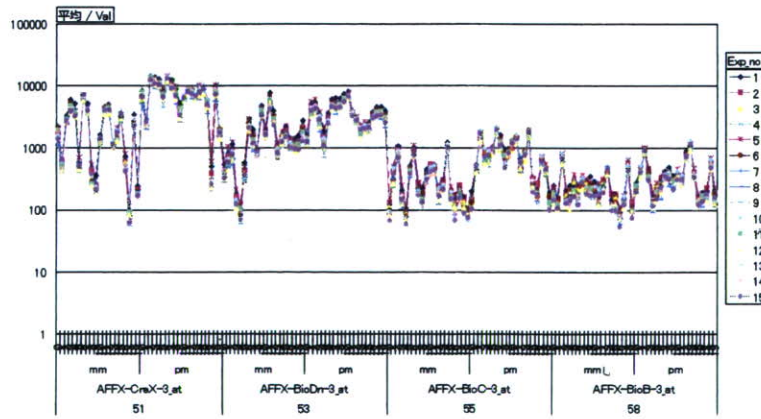
Copyright © NTT COMWARE 2008

1.3.3. チップごとの標準化(contRNA)



チップごとの標準化(contRNA)

- 露光等の問題により、チップ毎にバイアス値が異なる。このような測定誤差の変動を補正し、標準化を行う。
- contRNAのプローブセットを使用する。



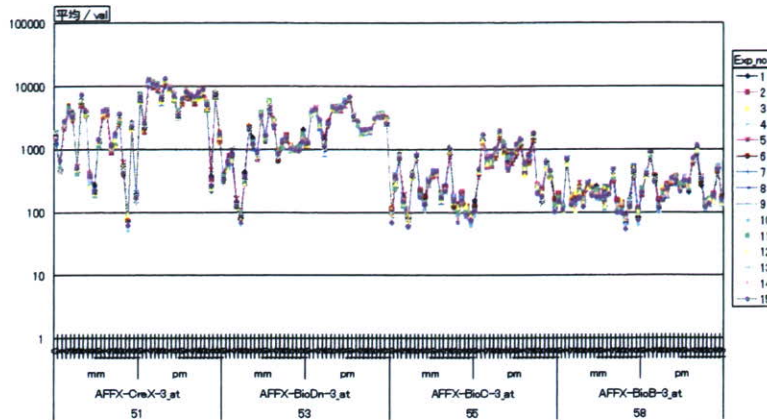
プローブ毎に変動しているが、チップ (Exp. no) 毎に似たパターンを示している。これらの平均値への偏差が最小になるように、定数倍することで標準化を行う。

Copyright © NTT

1.3.3. チップごとの標準化 (contRNA)



チップごとの標準化 (contRNA) の実施後のグラフ



わずかでであるが、線の重なりが増え、チップ間のばらつきが小さくなったことが読み取れる

Copyright © NTT COMWARE 2008

1.4.1. Langmuir式フィッティング検証



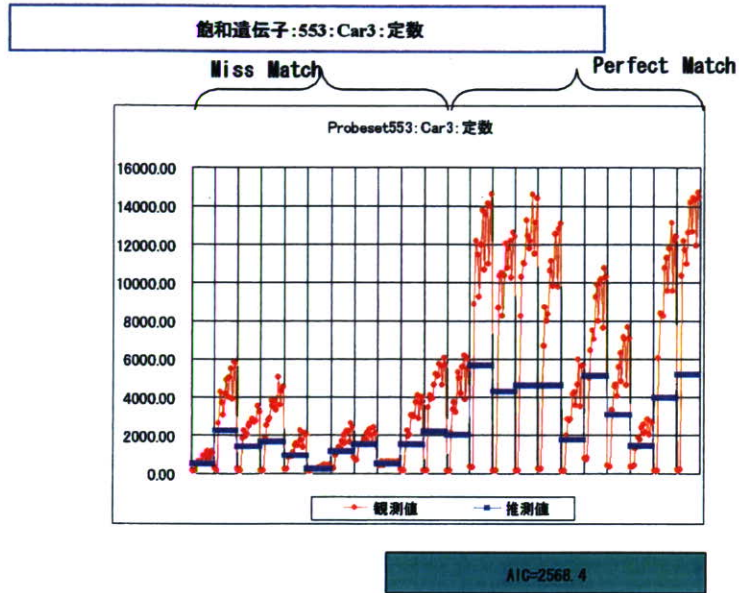
Langmuir式にフィッティング可能かいくつかのプローブセットで試してみる 実験対象プローブセット

- 計算を試みるプローブセット
 - 飽和、線形と思われたプローブセット
 - LBM、50%:50%における、発現値順位、最上位、10%、50%、80%
 - Percollome Spike Cocktail (内2プローブセットは、プローブ数が異なるので今回は省略)

	No	PctNo	Avg	SD	AffID	GeneSymbol	CV
飽和	0.1%	49	553	7700.28	638.10	1480256_at	2%
	0.2%	96	1201.0	4636.40	230.28	1448792_at	3%
	0.1%	60	439.62	7389.56	278.33	1416808_at	4%
線形	0.0%	16	40331	8888.57	284.42	1420456_at	3%
	6.0%	2684	5353	777.05	27.06	1455454_at	3%
	38.1%	17178	35249	53.56	3.83	1425540_at	11%
高発現線形	0.0%	11	42921	110781.10	223.24	1417350_at	2%
	0.0%	2	4321.5	10550.20	405.28	1417358_at	4%
	0.0%	3	28686	10519.77	445.44	1434110_x.at	4%
中発現線形	0.0%	4	5808	10504.17	346.16	1456201_x.at	3%
	0.0%	5	41538	10193.04	234.14	1419233_x.at	2%
	10.0%	4613	45063	470.25	49.31	1415708_at	10%
低発現線形	10.0%	4614	81.7	4311.8	10.17	1439890_at	2%
	10.0%	4615	38943	463.79	15.76	1421853_at	3%
	10.0%	4616	32355	463.60	68.73	1429441_at	15%
Percollome Spike	10.0%	4617	6059	469.59	6.77	1454748_at	1%
	50.0%	22551	36922	49.64	4.75	1423874_at	10%
	50.0%	22552	6806	48.64	11.65	1454001_at	2%
Percollome Spike	50.0%	22553	43844	49.62	11.99	1417127_at	2%
	50.0%	22554	16763	49.61	10.41	1444041_at	21%
	50.0%	22555	19740	49.61	11.62	1441056_at	23%
Percollome Spike	80.0%	36083	38463	13.48	3.65	1422303_at	27%
	80.0%	36083	20485	13.47	2.67	1440311_at	20%
	80.0%	36084	4163	13.47	0.71	1456644_at	5%
Percollome Spike	80.0%	36085	37240	13.47	9.07	1435659_at	67%
	80.0%	36086	14458	13.47	11.53	1445338_at	5%
	18.9%	7638	3	265.81	4.64	AFFX-ChiX-3_at	2%
Percollome Spike	6.8%	2865	30	702.92	27.91	AFFX-2-Ba-lac-3_at	4%
	2.4%	1089	37	1832.62	77.83	AFFX-ChiX-3_at	4%
	0.5%	218	27	4901.25	362.87	AFFX-2-Ba-lac-3_at	7%
Percollome Spike	0.0%	15	21	8929.21	348.22	AFFX-2-Ba-lac-3_at	4%

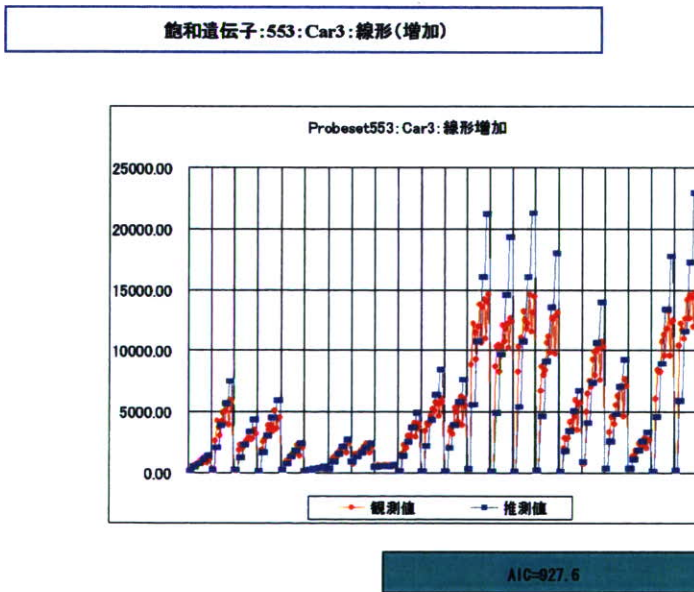
Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子

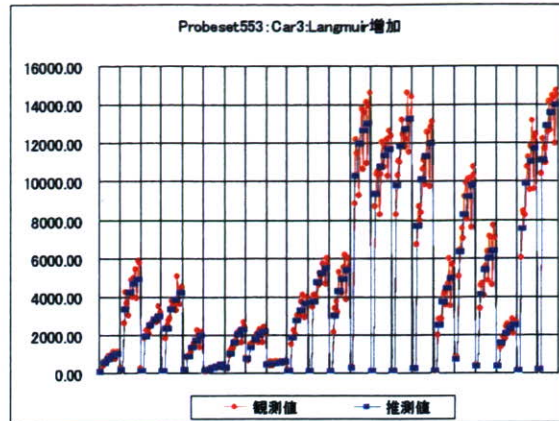


Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子:553:Car3:Langmuir(増加)



最もAICの小さい最適モデル

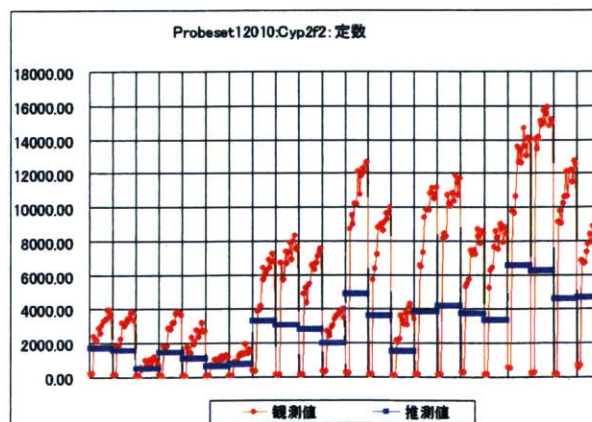
AIC=714.3

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子:12010:Cyp2f2:定数



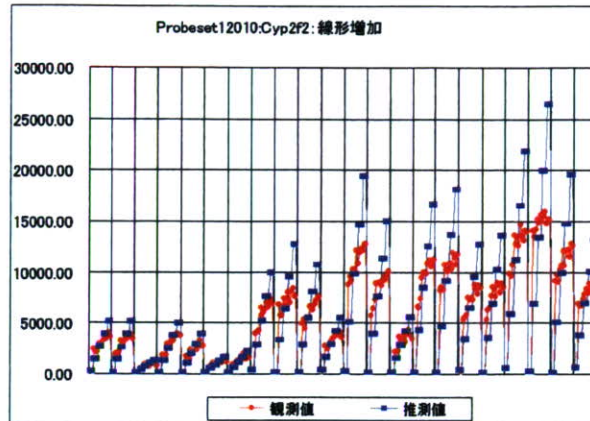
AIC=2768.1

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子:12010:Cyp2f2:線形増加



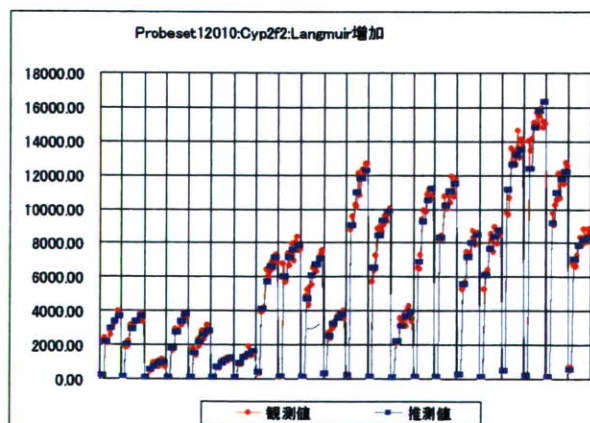
AIC=930.2

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子:12010:Cyp2f2:Langmuir増加



最もAICの小さい最適モデル

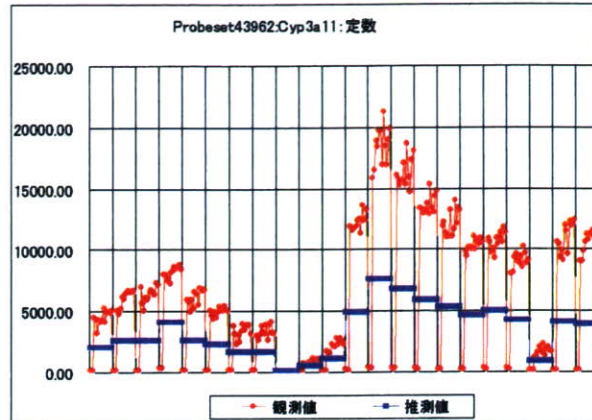
AIC=590.1

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子: 43962: Cyp3a11: 定数



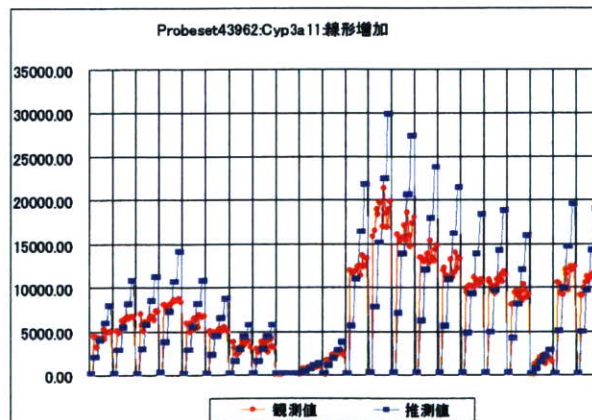
AIC=2985.2

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子: 43962: Cyp3a11: 線形(増加)



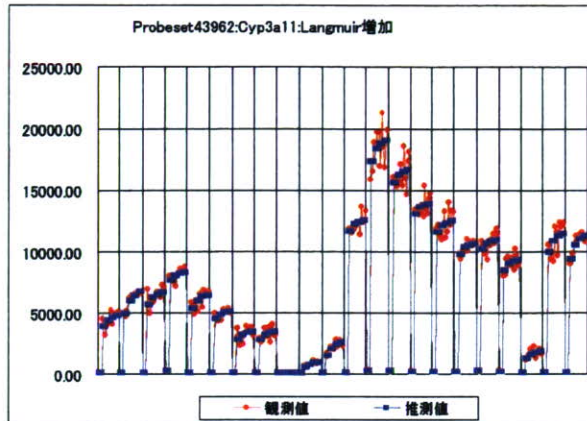
AIC=1076.8

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子:43962:Cyp3a11:Langmuir(増加)



最もAICの小さい最適モデル

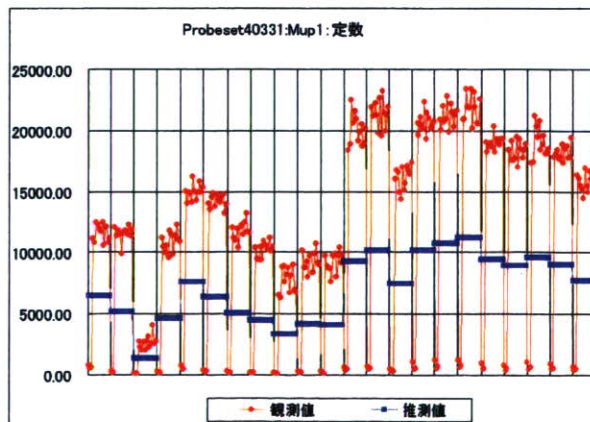
AIC=622.9

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子:40331:Mup1:定数



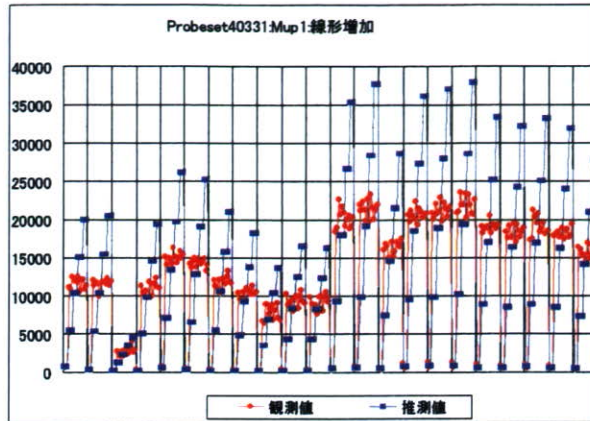
AIC=2865.0

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子:40331:Mup1:線形(増加)



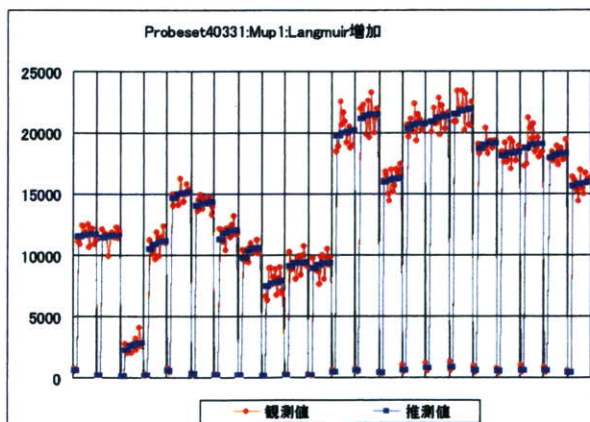
AIC=1181.2

Copyright © NTT COMWARE 2008

1.4.2. Langmuir式フィッティング検証 飽和遺伝子



飽和遺伝子:40331:Mup1:Langmuir(増加)



AIC=684.7

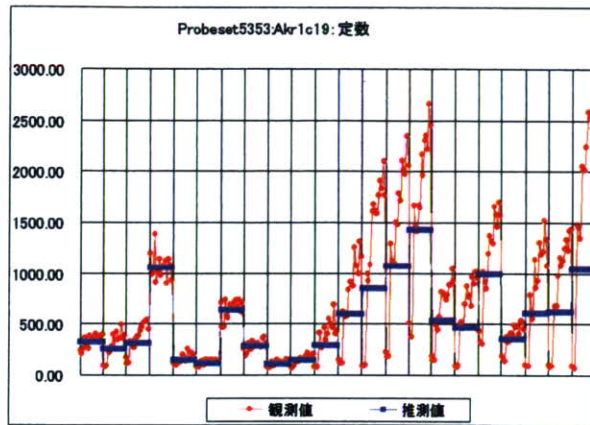
最もAICの小さい最速モデル

Copyright © NTT COMWARE 2008

1.4.3. Langmuir式フィッティング検証 線形遺伝子



線形遺伝子:5353:Akr1c19:定数



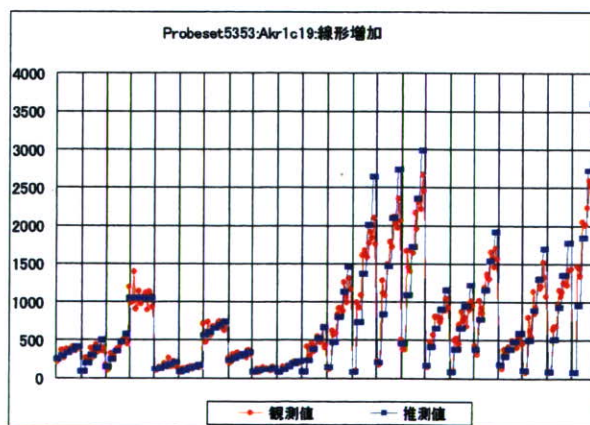
AIC=1543.4

Copyright © NTT COMWARE 2008

1.4.3. Langmuir式フィッティング検証 線形遺伝子



線形遺伝子:5353:Akr1c19:線形(増加)



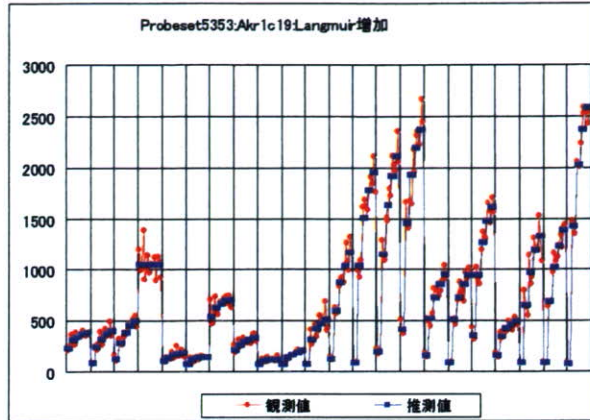
AIC=723.6

Copyright © NTT COMWARE 2008

1.4.3. Langmuir式フィッティング検証 線形遺伝子



線形遺伝子:5353:Akr1c19:Langmuir(増加)



最もAICの小さい最適モデル

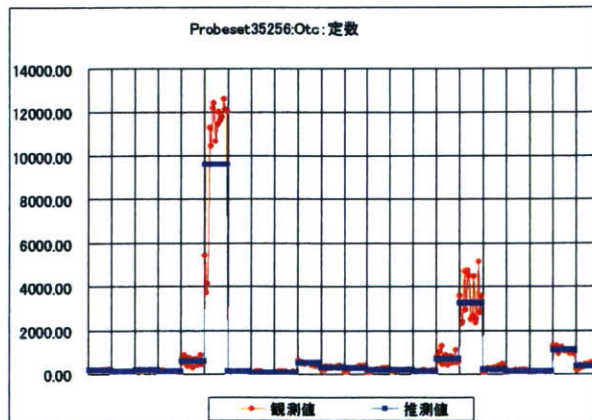
AIC=642.4

Copyright © NTT COMWARE 2008

1.4.3. Langmuir式フィッティング検証 線形遺伝子



線形遺伝子:35256:Otc:定数



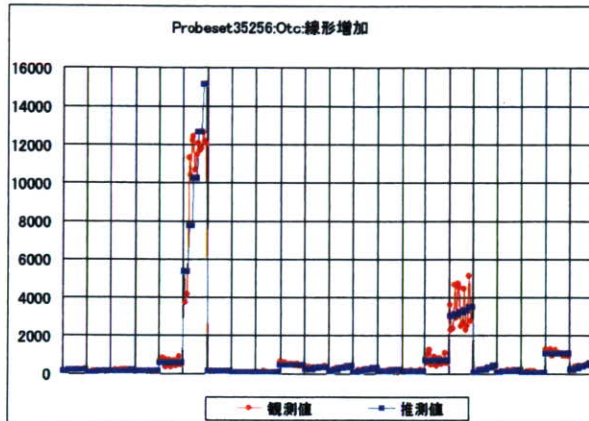
AIC=658.1

Copyright © NTT COMWARE 2008

1.4.3. Langmuir式フィッティング検証 線形遺伝子



線形遺伝子:35256:Otc:線形(増加)



最もAICの小さい最適モデル

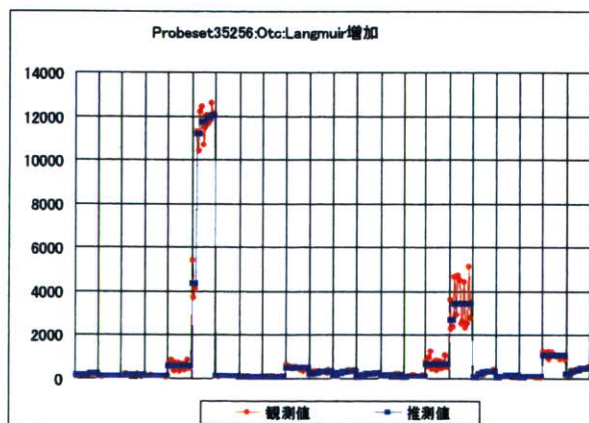
AIC=723.5

Copyright © NTT COMWARE 2008

1.4.3. Langmuir式フィッティング検証 線形遺伝子



線形遺伝子:35256:Otc:Langmuir(増加)

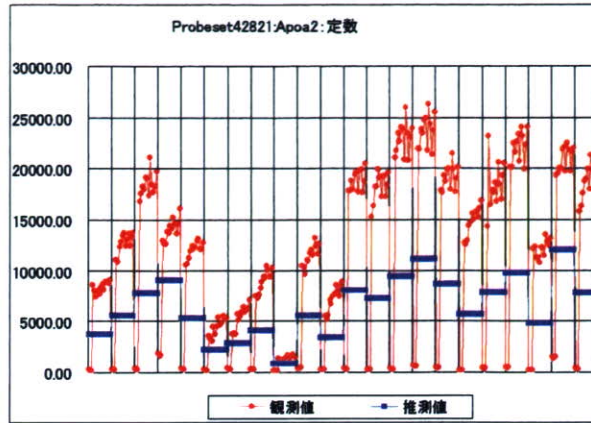


AIC=734.6

Copyright © NTT COMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子

高発現遺伝子:42821:Apoa2:定数

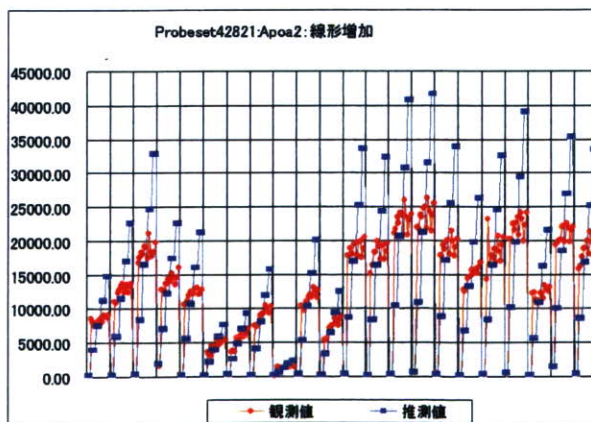


AIC=2986.3

Copyright © NTT COMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子

高発現遺伝子:42821:Apoa2:線形(増加)



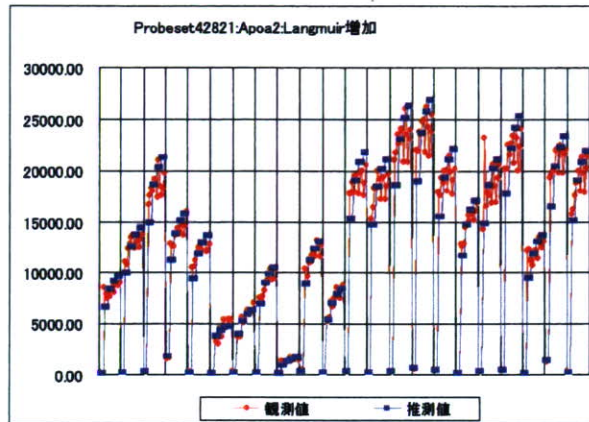
AIC=1083.0

Copyright © NTT COMWARE 2008

1. 4. 4. Langmuir式フィッティング検証 高発現遺伝子



高発現遺伝子:42821:Apoa2:Langmuir(増加)



最もAICの小さい最適モデル

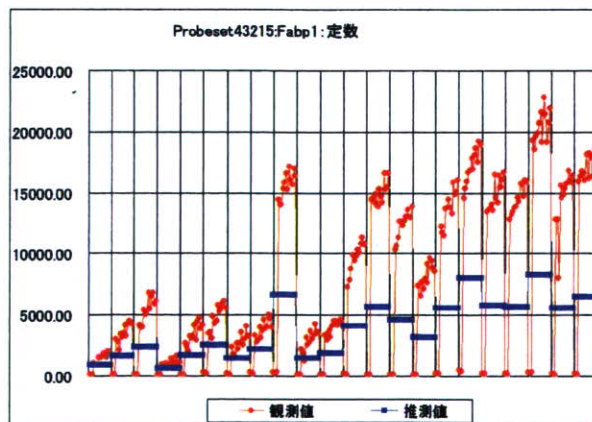
AIC=659.3

Copyright © NTT COMWARE 2008

1. 4. 4. Langmuir式フィッティング検証 高発現遺伝子



高発現遺伝子:43215:Fabp1:定数

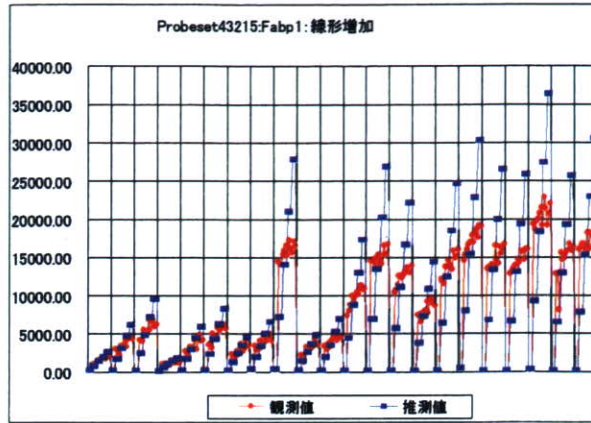


AIC=3103.6

Copyright © NTT COMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子

高発現遺伝子:43215:Fabp1:線形増加

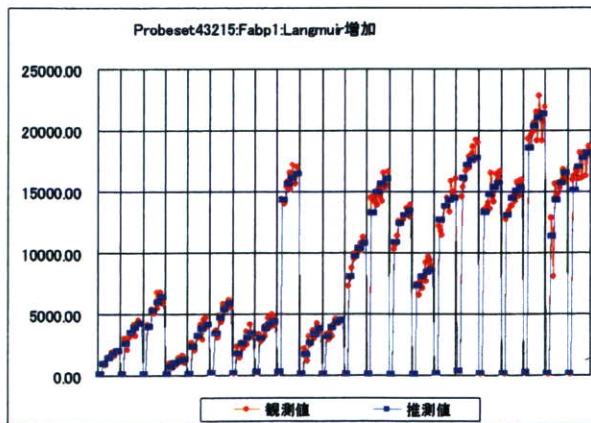


AIC=998.5

Copyright © NTT COMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子

高発現遺伝子:43215:Fabp1:Langmuir増加



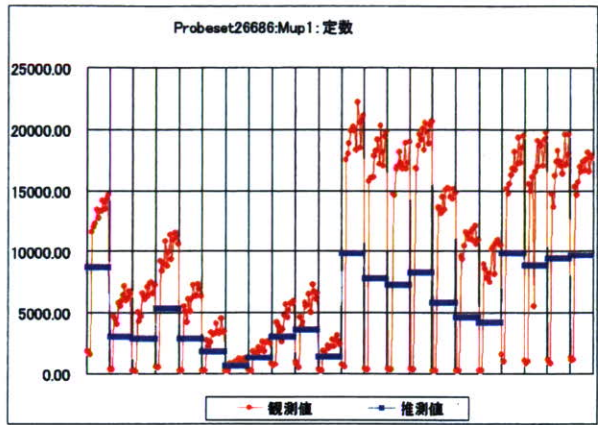
最もAICの小さい最適モデル

AIC=619.0

Copyright © NTT COMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子

高発現遺伝子:26686:Mup1:定数

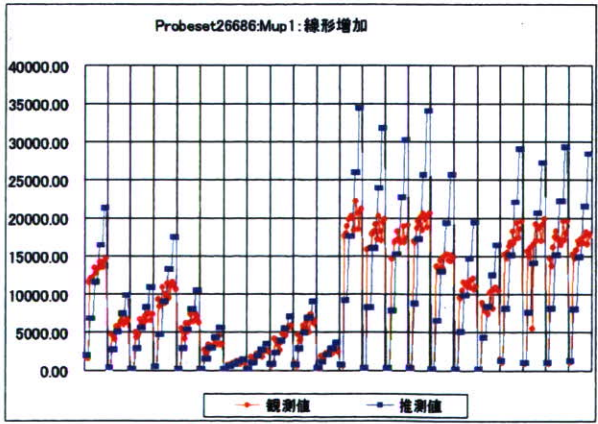


AIC=2571.6

Copyright © NTT COMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子

高発現遺伝子:26686:Mup1:線形(増加)

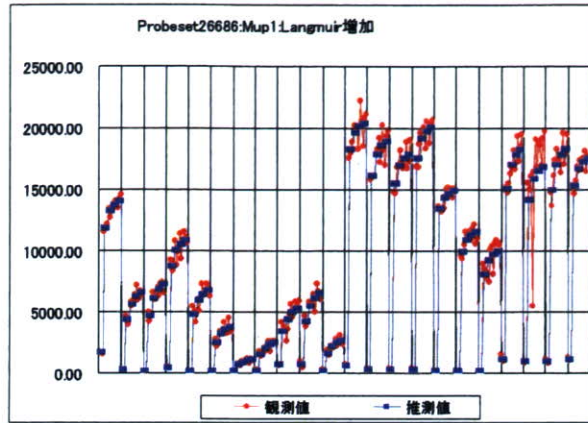


AIC=1015.8

Copyright © NTT COMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子

高発現遺伝子:26686:Mup1:Langmuir(増加)



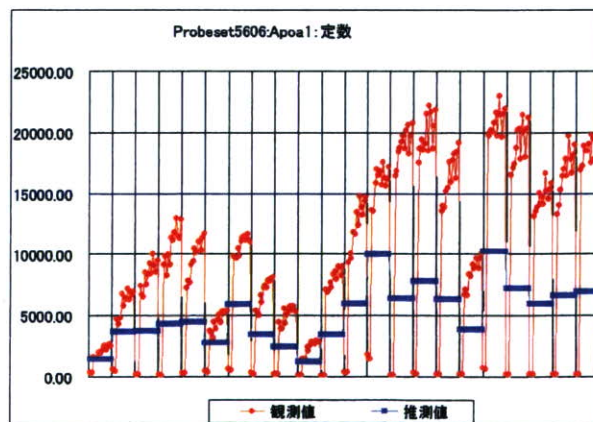
最もAICの小さい最適モデル

AIC=647.6

Copyright © NTT COMMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子

高発現遺伝子:5606:Apoa1:定数



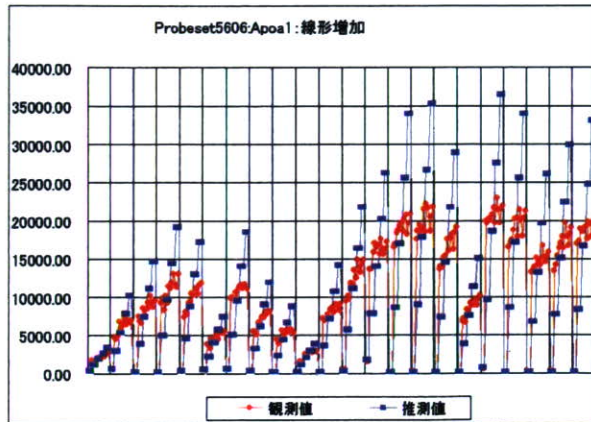
AIC=2987.8

Copyright © NTT COMMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子



高発現遺伝子:5606:Apoa1:線形(増加)



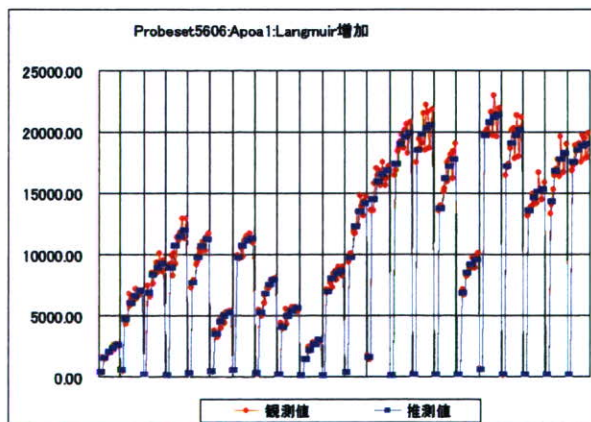
AIC=1020.1

Copyright © NTT COMWARE 2008

1.4.4. Langmuir式フィッティング検証 高発現遺伝子



高発現遺伝子:5606:Apoa1:Langmuir(増加)



最もAICの小さい最適モデル

AIC=578.5

Copyright © NTT COMWARE 2008