

BIC モデル比較による DNA マイクロアレイデータ正規化変換

岡崎 威生[†] 神谷 円^{††} 安富祖 仁^{††} 名嘉村盛和[†]

[†] 琉球大学工学部情報工学科 〒903-0213 沖縄県中頭郡西原町字千原 1

^{††} 琉球大学大学院理工学研究科情報工学専攻 〒903-0213 沖縄県中頭郡西原町字千原 1

E-mail: †{okazaki,morikazu}@ie.u-ryukyu.ac.jp, ††{mado,limbay}@ms.ie.u-ryukyu.ac.jp

あらまし DNA マイクロアレイデータは、実験に伴うノイズにより観測値が影響を受けやすいことが知られている。複数実験間の比較やデータ統合のためには、適切なバイアス補正による標準化が必要となる。これまで、理想データにおける分布の正規性を根拠に、特定の発現データに対し特定のバイアスを考慮した補正法が提案されてきた。本研究では想定される複数バイアスとその補正法的全組み合わせをモデル化し、情報量基準 BIC により最も適合したモデルを選出した上で正規化変換する手法を提案した。提案法を Stanford MicroArray Database で提供されている酵母菌、マウス、ヒトのマイクロアレイデータに適用し、従来正規化法との比較実験結果を示した。

キーワード DNA マイクロアレイ, 遺伝子発現データ, 正規化, BIC

Normalization of DNA Microarray data with BIC Model Comparison

Takeo OKAZAKI[†], Madoka KAMIYA^{††}, Hitoshi AFUSO^{††}, and Morikazu NAKAMURA[†]

[†] Information Engineering, Faculty of Engineering, University of the Ryukyus

1 Senbaru, Nishihara-cho, Okinawa, 903-0213 JAPAN

^{††} Information Engineering, Graduate School of Science and Engineering, University of the Ryukyus

1 Senbaru, Nishihara-cho, Okinawa, 903-0213 JAPAN

E-mail: †{okazaki,morikazu}@ie.u-ryukyu.ac.jp, ††{mado,limbay}@ms.ie.u-ryukyu.ac.jp

Abstract DNA Microarray data, which are efficient for estimation and identification of genetic network, have a large variation due to those experimental environment and measurement. Standardization by an appropriate bias correction is needed for the comparison and the data integration between two or more experiments. On grounds of normality of distribution in the ideal expression data, some adjustment methods that considers a specific bias for specific expression data has been proposed. In this research, after all combinations of assumed multiple biases and the adjustment methods were modeled, and the appropriate model by BIC was selected for normalization of expression data. The proposal method was applied to a yeast, mouse, and human microarray data from Stanford MicroArray Database, and the comparison experiment result with previous methods was shown so far.

Key words DNA Microarray, Gene Expression data, Normalization, BIC

1. はじめに

遺伝子ネットワークモデルの推定・同定のために有効な実験方法として、マイクロアレイ法があげられる。マイクロアレイ法は、スライドガラス上に数千から数万個の遺伝子を異なるスポットとして固定させたもので、細胞由来の mRNA もしくは mRNA を鋳型にして合成した cDNA をハイブリダイズさせ、ゲノムスケールで全ての遺伝子の発現パターンを同時に知ることができる。

一方で、この実験における観測値は実験環境・測定法などによる変動が大きく、再現性が低いことから、繰り返し実験を行

なうことが求められる。しかし、コスト面の問題や、大量の実験パターンが計画される中での優先順位の関係から、実際に十分な回数の繰り返し実験を行なうことは稀である。そのため、正確なゲノム解析や複数実験の比較、データの統合化が困難である。

比較・統合に向けた標準化のための指針としては、各スポットの観測蛍光比を対数変換した値が正規性を持つようにすることが一般的である。この指針に基づき、特定のデータおよびバイアスに対して提案されてきた補正法は、次の通りである。Yang *et al.* [1], [2] はマウスに対して、染料の変動の差によるバ

バイアス、シグナル強度に依存したバイアス、print-tip の変動によるバイアス、print-tip group 間の scale のばらつきを取り上げ、補正法を提案した。Uchida *et al.* [4] は酵母菌に対して、染料の変動の差によるバイアス、シグナル強度に依存したバイアス、print-tip の変動によるバイアス、print-tip でスポットをする順序によるバイアスを取り上げ、補正法を提案した。Smyth *et al.* [5] はマウスに対して、アレイ間の scale のばらつきを取り上げ、補正法を提案した。これらの補正法は、特定の発現データおよび特定のバイアスを対象としたものであるため、他のデータの補正に用いた場合に、同様に補正されるか保証がないといえる。

本研究では、特定の実験に依存しない補正法 (正規化法) を構築するために、従来のマイクロアレイデータに対するバイアス発生要因とその補正法を体系化した。その上で、系における大域的な補正を、各発生要因の有無・適用補正法選択・補正順序の情報量基準による判定で実現した。

2. 遺伝子発現データのバイアスと補正法

同一の実験系に対してマイクロアレイの繰り返し実験を行なったとしても、mRNA もしくは遺伝子発現量は一致するわけではなく、再現性の低さは、バイアスによるものが大きいとされている。ここでのバイアスは、Kohane [6] による定義のうち、“対象となる生体システムとは無関係に、外部から加わる物理的な効果で、測定感度を越えるレベルに達すると、生体システム側もしくは測定機器側の計測値に影響を及ぼすもの”に該当する。

生体システムに関する実験に影響するパラメータの範囲は広く、それらをすべて完全に制御することは非常に困難である。そのためマイクロアレイ実験には、生体システム側、測定機器側の両者から生ずるあらゆるバイアスが複雑に交錯する。

バイアスは、同一アレイ内の各スポットに見られるものと、異なるマイクロアレイ間での実験環境などの違いによるものに大別でき、従来補正法で取り上げられた各バイアスを、その発現箇所の視点で整理すると以下ようになる。

表 1 マイクロアレイ実験におけるバイアス発生箇所

	マイクロアレイ作製	試薬の調整	蛍光標識特性	ハイブリダイズ	スキャン
アレイ内	○	○	○	○	○
アレイ間	○			○	

発生要因毎の従来補正法は次の通りである。

2.1 染料の変化の差により発生するバイアスの補正

使用する蛍光色素間での標識と検出効率に起因するバイアスに対しては、染料 Cy3 と Cy5 のシグナル強度の平均比率が 1 になるよう補正するために、2 方法が提案されている。

- Total intensity normalization [7]

j 番目の遺伝子に対する Cy3, Cy5 のシグナル強度がそれぞれ G_j , R_j で、 N をマイクロアレイ上にスポットされた遺伝子数とすると、

$$T_{total} = \frac{\sum_{j=1}^N R_j}{\sum_{j=1}^N G_j} \quad (1)$$

を用いて、

$$G'_j = T_{total} G_j \quad R'_j = R_j \quad (2)$$

と補正する。

- Global normalization [1]

Cy3 と Cy5 のシグナルの強度が $R = kG$ のように恒常的要因の関係を持つものと仮定し、次の式で補正を行う。

$$M' = M - c \quad (3)$$

$M = \log_2 \frac{R}{G}$ で、 c は M の中央値または平均値が用いられる。

2.2 シグナル強度の絶対量に依存したバイアスの補正

発現レベルが小さいスポットは、実験中に発生するバイアスの影響をより大きく受けやすいため、発現レベルに対する回帰式を推定し補正を行う。

$$M' = M - c(A) \quad (4)$$

回帰式 $c(A)$ の構成法として、次の 2 方法が提案されている。

- loess 法 [1]

局所的な重み付けを行なう線形回帰を使ってデータの平滑化を行なう。データセットを局所的に平滑化した後に、外れ値を除いた上で再び回帰を行う。

- loess 法 [5]

loess 法と同じく局所的な重み付けを行なう線形回帰を使ってデータの平滑化を行なう。loess 法との違いは、重みつき線形最小二乗回帰を行なう際に 2 次多項式を用いることである。

2.3 print-tip の変動によるバイアスの補正

同じ print-tip によってスポットされた遺伝子グループを、print-tip group と呼ぶ。print-tip の長さや穴の開き、時間の経過による状態の差は、発現量に系統的な差を生じるため、次の 3 補正法が提案されている。

- Within-print-tip-group normalization [1]

アレイの全てのブロックで同じ print-tip によりスポットされていることから、系統的なバイアスが存在すると考えられる。各 print-tip group ごとに次式を用いて補正する。

$$M'_i = M_i - c_i(A) \quad i = 1, \dots, I \quad (5)$$

$c_i(A)$ は i 番目のブロック (i 番目の print-tip group) に対する loess 法を用いた回帰式で、 I は print-tip の総数である。

- Linear regression line equations [4]

Within-print-tip-group normalization と同様で、print-tip group 毎に線形回帰式 $c_i(A)$ を適用する。

- Robust fitting of linear models [8]

最小二乗法は線形回帰の標準的仮定のずれに敏感で、1 つの外れ値によっても大きな影響を受けてしまうので、最尤法を拡張した M 推定法を用いたロバストな線形回帰式により補正する。

2.4 print-tip 内のスポット順序によるバイアスの補正
Print-tip 由来のバイアスには上述とは別に、スポットされる順序により生じるものがあることが報告されており [5]、次の方法による補正が提案されている。

• Print-order normalization [4]

各プローブがマイクロアレイ上にスポットされた段階で生じるバイアスは、使用されたスポットピンに関係するとの仮定より、次のように補正される。

$$M'_{kn} = M_{kn} - C_k(n) \quad (6)$$

M_{kn} は k 番目の print-tip における第 n 番目のスポットに対する M 値である。 $C_k(n)$ は k 番目の print-order group における第 n 番目のスポットの線形回帰式である。

2.5 Print-tip group 間での分布の広がりへのばらつき

異なる Print-tip group 間では、分布の広がりへのばらつきが出る場合がある。これを補正するために、次の補正法が提案されている。

• Within-slide scale normalization [1]

α_i^2 は第 i 番目の Print-tip group に対するスケール要因とし、第 i 番目の Print-tip group の M 値が、正規分布 $N(0, \alpha_i^2 \sigma^2)$ に従うと仮定する。このとき α_i の最尤推定値は次式で与えられる。

$$\hat{\alpha}_i = \frac{\sum_{j=1}^{n_i} M_{ij}^2}{\sqrt{\prod_{i=1}^I \sum_{j=1}^{n_i} M_{ij}^2}} \quad i = 1, \dots, I \quad (7)$$

M_{ij} は i 番目の print-tip グループの j 番目のシグナル強度の log 比で、 I はアレイ上の print-tip の総数である。

2.6 マイクロアレイ間の補正

異なるマイクロアレイ間でのデータの比較には、次の2つの視点が挙げられる。

- (1) 同じ制御での繰り返し実験によるマイクロアレイ間比較
- (2) 異なる制御実験によるマイクロアレイ間比較

繰り返し実験では、実験環境のばらつきによるバイアスが生じる可能性がある。異なる制御実験では、実験環境のばらつきに加え、ハイブリダイゼーションのプロトコルなどの違いからバイアスが生じる可能性がある。それぞれのマイクロアレイ内で補正を行なった後は、各スライドのシグナル強度の log 比の中央値はほぼ同じになっている。しかし分布の広がりへのばらつきがあるため、その補正法として次の方法が提案されている。

• マイクロアレイ間 scale normalization [5]

$MAD_i = \text{median}_i \{ |M_{ij} - \text{median}_j(M_{ij})| \}$ の $\text{median}_j(M_{ij})$ を 0 とし、 M 値の絶対値の中央値を他のマイクロアレイと等しくすることで、scale の補正を行う。

3. 遺伝子発現データの正規化

マイクロアレイ実験はその過程においてさまざまな要因からバイアスが発生する。従来の正規化手法と対象要因・補正法の対応をまとめたものが表 2 である。

この表が示すように、これまで提案されてきた正規化手法は、特定の生体に対するものであり、また全てのバイアス要因を考

慮していないことがわかる。更に、各バイアス要因の有無の判定が明確でないため、データに依存した補正になっている。そのため、他のデータの補正に用いた場合に、同様に補正されるか保証がないといえる。

本研究では、提案する正規化の条件として次の3つを挙げた。

- 全てのバイアス要因を考慮できること
- 各補正法の統合モデルであること
- 各バイアスの有無判定が行なわれていること

これらを満足することにより、観測データに依存しない正規化手法が構成できると考えた。

3.1 各補正法の統合モデルにおける定式化

大域的な補正を実行するためには、すべてのバイアスを統合的に表現する必要がある。そこで観測データを次のように再定義した。

i : print-tip group の添字 (総数 I)

k : print-order group の添字 (総数 K)

j : マイクロアレイの添字

n : ある print-order group における、スポットされた順序

α_i : 第 i 番目の print-tip group におけるスケール要因

α_{*j} : 第 j 番目のマイクロアレイにおけるスケール要因

$G_{(ikn)}$: スポットの緑のシグナル強度

$R_{(ikn)}$: スポットの赤のシグナル強度

$$M_{(ikn)} = \log_2 \left(\frac{R_{(ikn)}}{G_{(ikn)}} \right)$$

$$A_{(ikn)} = \log_2 \sqrt{G_{(ikn)} R_{(ikn)}}$$

$$T_{\text{total}} = \frac{\sum_{j=1}^N R_j}{\sum_{j=1}^N G_j}$$

この定義により各補正法を次のように関数として再定式化できる。

- (1) 染料の変化の差により発生するバイアス補正式

Global normalization :

$$f_1^{(G)}(M_{(ikn)}) = M_{(ikn)} - c$$

c は $M_{(ikn)}$ の中央値または平均値を用いる。

Total intensity normalization :

$$f_1^{(T)}(G_{(ikn)}, R_{(ikn)}) = \log_2 \left(\frac{R_{(ikn)}}{T_{\text{total}} G_{(ikn)}} \right)$$

- (2) シグナル強度の絶対値に依存したバイアス補正式

線形回帰: $f_2^{(l)}(M_{(ikn)}, A_{(ikn)}) = M_{(ikn)} - c^{(l)}(A_{(ikn)})$

lowess 法: $f_2^{(lw)}(M_{(ikn)}, A_{(ikn)}) = M_{(ikn)} - c^{(lw)}(A_{(ikn)})$

loess 法: $f_2^{(le)}(M_{(ikn)}, A_{(ikn)}) = M_{(ikn)} - c^{(le)}(A_{(ikn)})$

ロバスト線形回帰:

$$f_2^{(r)}(M_{(ikn)}, A_{(ikn)}) = M_{(ikn)} - c^{(r)}(A_{(ikn)})$$

- (3) Print-tip の変動により発生するバイアス補正式

線形回帰: $f_3^{(l)}(M_i, A_i) = M_i - c_i^{(l)}(A_i)$

lowess 法: $f_3^{(lw)}(M_i, A_i) = M_i - c_i^{(lw)}(A_i)$

loess 法: $f_3^{(le)}(M_i, A_i) = M_i - c_i^{(le)}(A_i)$

ロバスト線形回帰: $f_3^{(r)}(M_i, A_i) = M_i - c_i^{(r)}(A_i)$

- (4) Print-tip 内スポット順序によるバイアス補正式

線形回帰: $f_4^{(l)}(M_{(kn)}, n_k) = M_{(kn)} - C_k^{(l)}(n)$

lowess 法: $f_4^{(lw)}(M_{(kn)}, n_k) = M_{(kn)} - C_k^{(lw)}(n)$

loess 法: $f_4^{(le)}(M_{(kn)}, n_k) = M_{(kn)} - C_k^{(le)}(n)$

表 2 バイアスと従来の補正法の対応表

	Yang et al. [2] (マウス)	Uchida et al. [4] (酵母菌)	Smyth et al. [5] (マウス)
染料の変動	Global normalization	Total intensity normalization	
シグナル強度依存	lowess 法	lowess 法	
print-tip 変動	lowess 法	線形回帰	loess 法
print-tip 内スポット順序		線形回帰	
print-tip group 間 scale	Within-slide scale normalization		
アレイ間 scale			アレイ間 scale normalizations

ロバスト線形回帰: $f_4^{(r)}(M_{(kn)}, n_k) = M_{(kn)} - C_k^{(r)}(n)$

(5) Print-tip group 間の scale 補正式

$$f_5(M_i) = (1/a_i)M_i$$

(6) マイクロアレイ間の scale 補正式

$$f_6(M_j) = (1/a_j)M_j$$

3.2 統合表現

各補正は関数として表現されたので、補正の組み合わせは関数の合成として表現できる。この時、関数の合成順序はバイアス補正の適用順序と対応することになる。

$$\text{例) } f_3^{(1)} \circ f_2^{(1)} \circ f_1^{(0)} = M_{(ikn)} - c^{(1)}(A_{(ikn)}) - c_1^{(1)}(A) - c$$

補正を適用する順序は、マイクロアレイ実験において各バイアスが発生する構造に対応する。しかし下記の場合、発生順序を限定できない。

(1) 染料はマイクロアレイがスキャンされるまで劣化し続けるため、染料の変動の差によるバイアスと、シグナル強度に依存したバイアスの発生順序が限定できない。

(2) Print-tip の変動によるバイアスと、Print-tip でスポットする順序によるバイアスは、どちらもマイクロアレイ作製時に起こるため、発生順序が限定でない。

従って、全バイアスが存在するときの補正の順序は次の 4 通りが考えられる。

Adjustment Order 1

1. 染料の変動によるバイアスの補正
2. シグナル強度依存のバイアスの補正
3. Print-tip の変動により発生するバイアスの補正
4. Print-tip 内でスポット順序によるバイアスの補正
5. Print-tip 間での scale の補正

Adjustment Order 2

1. 染料の変動によるバイアスの補正
2. シグナル強度依存のバイアスの補正
3. Print-tip 内でスポット順序によるバイアスの補正
4. Print-tip の変動により発生するバイアスの補正
5. Print-tip 間での scale の補正

Adjustment Order 3

1. シグナル強度依存のバイアスの補正
2. 染料の変動によるバイアスの補正
3. Print-tip の変動により発生するバイアスの補正
4. Print-tip 内でスポット順序によるバイアスの補正
5. Print-tip 間での scale の補正

Adjustment Order 4

1. シグナル強度依存のバイアスの補正
2. 染料の変動によるバイアスの補正
3. Print-tip 内でスポット順序によるバイアスの補正
4. Print-tip の変動により発生するバイアスの補正
5. Print-tip 間での scale の補正

これまでに示したバイアスは必ず発生するものではない。また、Adjustment Order 3 と Adjustment Order 4 の場合、シグナル強度依存のバイアスを補正した後では、Total intensity normalization を用いることはできない制約を考慮すると、補正モデル候補の総数は 5000 となる。

各バイアスの有無と補正の順序の判定は、補正モデル候補の中で、観測値に最も適合したものを選べばよいと考えられる。そこで適合度判定の基準として BIC(Baysian information criterion) [9] を用いることとした。

$$BIC = -2\log(\text{最大尤度}) + p\log(n) \quad (p: \text{パラメータ数})$$

BIC 値が最小の補正モデル候補が、最も観測データに適合したものとなり、各バイアスの有無の判定、補正順序、補正に用いる回帰式を決定することと同じ意味を持つ。

以上より、遺伝子発現データ正規化の手続きを以下に示す。

(Step1.) 各補正モデル候補においてマイクロアレイデータの補正を行なう。

(Step2.) 各補正モデル候補の BIC を算出する。

(Step3.) BIC 値が最小なモデルを選択し、その補正値を正規化データとして出力する。

従来補正法も補正モデル候補に含まれていることから、より汎用性の高い正規化手続きが構成できるといえる。

4. 検証実験

提案した正規化手法により、実際の遺伝子発現データを正規化できるか検証するために、Stanford MicroArray Database [10] で公開されている酵母菌・大腸菌・ヒトの発現データを利用し実験を行なった。

補正後の正規性検証は、正規分布との適合度で測ることが妥当である。総合的指標となる BIC 値が最小であることは自明なので、分布形状の特徴である歪度と尖度を付加した。正確な正規分布の下では、歪度 0、尖度 3 である。

4.1 酵母菌

得られた 10752 個の遺伝子データから、無効値を除いた 9851 個を適用した。BIC 値最小で選出されたモデルは次の補正によるものであった。

$$f_6 \circ f_4^{(te)} \circ f_3^{(1)} \circ f_2^{(1)} \circ f_1^{(T)}$$

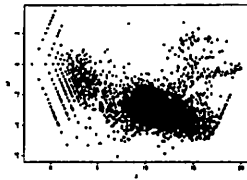


図1 補正前の MA-plot(酵母菌)

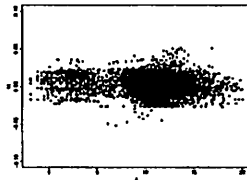


図2 提案法による正規化後の MA-plot(酵母菌)

1. Total intensity normalization を用いて染料のバイアスを補正
2. loess 法を用いてシグナル強度依存のバイアスを補正
3. 線形回帰を用いて print-tip group の変動のバイアスを補正
4. loess 法を用いて print-tip でスポット順序によるバイアスを補正
5. print-tip group 間の scale の補正

表3 正規化指標 (酵母菌)

	補正前	提案法	Yang et al. [2]	Uchida et al. [4]
BIC	25633.44	-61290.67	30195.46	16492.63
歪度	1.8018	-0.0693	0.8420	1.9520
尖度	9.4975	3.4720	8.4624	12.6900

歪度・尖度ともに理想値に十分近く、正規化されていることが確認できた。従来法と比較しても、良い結果が得られた。

4.2 大腸菌

得られた 6384 個の遺伝子データから、無効値を除いた 5603 個を適用した。BIC 値最小で選出されたモデルは次の補正によるものであった。

$$f_5 \circ f_4^{(le)} \circ f_3^{(l)} \circ f_2^{(le)} \circ f_1^{(T)}$$

1. Total intensity normalization を用いて染料のバイアスを補正
2. loess 法を用いてシグナル強度依存のバイアスを補正
3. 線形回帰を用いて print-tip group の変動のバイアスを補正
4. loess 法を用いて print-tip でスポット順序によるバイアスを補正
5. print-tip group 間の scale の補正

酵母菌の場合と同様に、歪度・尖度ともに理想値に十分近く、正規化されていることが確認できた。

表4 正規化指標 (大腸菌)

	補正前	提案法	Yang et al. [2]	Uchida et al. [4]
BIC	14201.22	-20742.05	13684.78	9532.93
歪度	0.0994	-0.0569	-0.4820	-0.4378
尖度	6.7203	4.8814	7.8463	7.4394

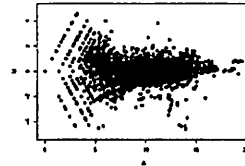


図3 補正前の MA-plot(大腸菌)

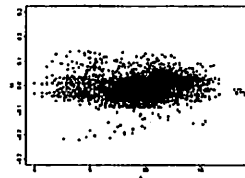


図4 提案法で補正後 MA-plot(大腸菌)

4.3 ヒト

得られた 44536 個の遺伝子データから、無効値を除いた 43715 個を適用した。BIC 値最小で選出されたモデルは次の補正によるものであった。

$$f_5 \circ f_4^{(le)} \circ f_3^{(r)} \circ f_2^{(le)} \circ f_1^{(T)}$$

1. Total intensity normalization を用いて染料のバイアスを補正
2. loess 法を用いてシグナル強度依存のバイアスを補正
3. ロバストな線形回帰を用いて print-tip group の変動のバイアスを補正
4. loess 法を用いて print-tip でスポットする順序によるバイアスを補正
5. print-tip group 間の scale の補正

表5 正規化指標 (ヒト)

	補正前	提案法	Yang et al. [2]	Uchida et al. [4]
BIC	76867.86	-326368.60	77814.04	42957.59
歪度	0.1470	-0.1864	0.0046	-0.2322
尖度	6.7113	4.6328	6.3400	6.6507

歪度に関しては若干劣化しており、Yang et al の補正法が有効であった。一方で尖度に関しては理想値に近づいており、他の補正法より改善が大きい。図7に示すヒストグラムからわかるように、オリジナルデータにおいて既に正規分布に近い対称性が得られているため、尖度の改善が歪度に影響したものと考えられる。

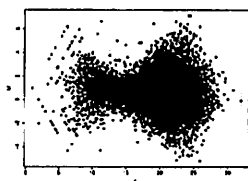


図5 補正前の MA-plot(ヒト)

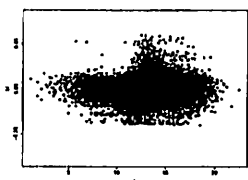


図6 提案法で補正後の MA-plot(ヒト)

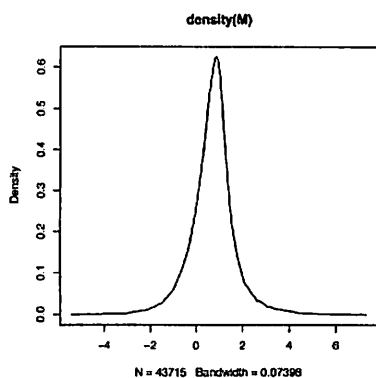


図7 補正前の遺伝子発現データのヒストグラム(ヒト)

5. まとめと課題

本研究では、従来のマイクロアレイデータ実験における問題であった、信頼性の確立を解決すべく、観測データの正規化手法を提案した。そのために、測定値を得るまでのバイアス発生要因とその補正法を体系化した。従来の補正法は特定のデータおよびバイアスに対する補正であったのに対し、提案法は大域的に各バイアスの有無を情報量基準により判定することで、特定データに依存しない正規化を実現した。

実際の酵母菌、大腸菌、ヒトの遺伝子発現データに対して提案法を適用し、オリジナルデータおよび従来の補正法と比較した結果、情報基準量、正規性を示す統計的指標において、提案法の優位性が確認された。

提案手法は、主目的のゲノム解析のための前処理に位置づけられる。そこで、提案法の正規化により複数実験結果を統合し、疑似的な大規模実験を構成することが重要である。この統合さ

れたデータをもとに各種ゲノム解析を実施し、正規化の実効性を検証することが課題である。

文 献

- [1] Y.H.Yang, S.Dudoit, P.Luu, and T.P.Speed, Normalization for cDNA Microarray Data, In M.L.Bittner, Y.Chen, A.N.Dorsel, and E.R.Dougherty, editors, *Microarrays: Optical Technologies and Informatics*, Vol 4266 of Proceedings of SPIE, 2001.
- [2] Y.H.Yang, S.Dudoit, P.Luu, D.M.Lin, V.Peng, J.Ngai, and T.P.Speed, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, 30(4):e15, 2002.
- [3] Y.H.Yang and N.Thorne, Normalization for Two-color cDNA Microarray Data. *Science and Statistics: A Festschrift for Terry Speed, D. Goldstein (eds.)*, IMS Lecture Notes, Monograph Series, Vol 40, pp.403-418, 2003.
- [4] S.Uchida, Y.Nishida, K.Satou, S.Muta, K.Tashiro, and S.Kuhara, Detection and Normalization of Biases Present in Spotted cDNA Microarray Data: A Composite Method Addressing Dye, Intensity-Dependent, Spatially-Dependent, and Print-Order Biases., *DNA RESEARCH* 12, 1, pp.1-7, 2005.
- [5] G.K.Smyth and T.Speed, Normalization of cDNA microarray data, *Methods*, 31(4), pp.265-273, 2003.
- [6] I.S.Kohane, A.T.Kho, A.J.Butte, 統合ゲノミクスのためのマイクロアレイデータアナリシス, シュプリンガー・フェアラーク東京, 2004.
- [7] J.Quackenbush, Microarray data normalization and transformation, *Nature Genetics Supplement*, 32, pp.496-501, 2002.
- [8] Y.Xiao, M.R.Segal, Y.H.Yang, Stepwise normalization of two-channel spotted microarrays, *Statistical Applications in Genetics and Molecular Biology*, 4(1):4., 2004.
- [9] Schwarz,G., Estimating the dimension of a model, *Ann. Statist.*, Vol 6, pp.461-464, 1978.
- [10] Stanford MicroArray Database, <http://genome-www5.stanford.edu/>
- [11] Steen Knudsen, わかる 使える DNA マイクロアレイデータ解析, 羊土社, 2002.
- [12] 岡崎康司, 必ずデータが出る DNA マイクロアレイ実験マニュアル 基本原理, チップ作製技術からバイオインフォマティクスまで, 羊土社, 2000.