

C.elegans のマイクロアレイ解析における
ワンサイクル法とエクспレスキット法の比較
Comparison of *C.elegans* microarray data
between One-Cycle and Express-Kit methods

岡野 文香† 物部 香奈美‡
Ayaka Okano Kanami Monobe

伊藤 将弘‡ 真栄城 哲也†
Masahiro Ito Tetsuya Maeshiro

要旨

Affymetrix 社のマイクロアレイ実験を行う際に必要な試料調製手法 2 種類から得られる発現量の実験データを、線虫 *C.elegans* を対象に比較した。これまで使用されていた One Cycle 法と、新しく使われるようになった Express Kit 法で、マイクロアレイデータにおける遺伝子発現量を比較し、手法の違いによってどのように実験結果が異なるのかを分析した。One Cycle 法と Express Kit 法の全体的なデータ間の相関係数は 0.93 と高く、両者の実験データは類似していると判断できる。また、線虫が持つ遺伝子約 2 万個それぞれの実験間の発現変動について t 検定を行い、One Cycle 法と Express Kit 法で発現量に差が出る遺伝子を検出した。さらに、One Cycle 法と Express Kit 法の発現量検出の傾向についても考察した。

1. はじめに

本研究の目的は、線虫 *Caenorhabditis elegans* を対象として、マイクロアレイ解析のための試料調整方法である One Cycle 法と Express Kit 法による実験結果の違いを比較し分析することである。

我々は、線虫 *C.elegans* の初期胚における全遺伝子の発現量を解析するため、Affymetrix 社のマイクロアレイを用いたゲノムワイドなマイクロアレイ実験を実施している [1-4]。従来の研究には One Cycle 法による実験データを使用していたが、この手法のキットの提供終了に伴い、今後は Express Kit 法による実験データに切り替わる。従って、線虫 *C.elegans* の初期胚の発現量データが、実験手法によって受ける影響を詳細に分析する必要がある。どのような遺伝子の発現量が変動するのか、One Cycle 法によるデータと Express Kit 法によるデータが比較できるのか、そして比較に必要な処理について詳しく知る必要がある。これらの詳細が明らかになることで、One Cycle 法による実験データを用いた研究と Express Kit 法による実験データを用いた研究を比較する際の注意点、また両データを同時に用いる際にどのようにデータを修正すべきかが明らかになる。

One Cycle 法および Express kit 法とは、Affymetrix

社が販売するマイクロアレイ実験用の RNA ターゲット調整キットである。Express kit 法は、One Cycle 法の後継のキットであり、微量発現量の検出能力がその特徴である。1 回の増幅に必要な RNA は Express Kit 法の方が少量であり、One Cycle 法は 100 μ g、Express kit 法は 50 ng である。Express Kit 法の精度について、Affymetrix 社が One Cycle 法と Express Kit 法についておおまかな比較を公表 [5-6] しているが、詳細は不明である。Affymetrix 社は Express kit 法と One Cycle 法の性能は同レベルであると説明しており、One Cycle 法と Express Kit 法の相関係数は 0.97 であり、高い相関があると述べている。しかしながら、個々の遺伝子についての記述は見られず、詳細は不明である。特に、我々の実験対象である *C.elegans* については役立つ情報は無い。そのため、線虫 *C.elegans* の初期胚に対して、キットの違いの影響を検討する必要がある。

2. 方法

実験に用いるマイクロアレイは、Affymetrix 社 GeneChip One-Cycle Target Labeling and Control Reagents と GeneChip 3' IVT Express Kit である。線虫 *C.elegans* のゲノムワイドな発現量を検出し、全 22,548 遺伝子を扱う。

実験に用いるサンプルは、*C.elegans* の同一サンプルを 6 分注し、One Cycle 法と Express Kit 法で 3 サンプルずつ調整した。このようにして、得た発現量のデータを分析する。サンプルは以下のように調整した。

卵を一例持った頃の線虫に対し、ナイロンメッシュで濾過し、標的となる線虫を集める。ナイロンメッシュ上に集めた線虫を 4 °C の滅菌蒸留水でよく洗い、500ml のビーカーに線虫を移す。移した線虫を 15ml チューブ (gainer bio-one 社製) に移し、4 °C、2000r.p.m. で 1 分間遠心分離 (KUBOTA 社製卓上小型冷却遠心機 2810) を行う。上清を取り除き、12.5ml まで滅菌蒸留水を加え、アルカリブリーチ液 (滅菌蒸留水 19.5ml, 5M KOH (Wako 和光一級) 2.5ml, NACIO (Wako 化学用) 3.0ml) を加え、静かに混ぜ 7 分間氷上に放置する。シリンジ (TERUMO 社製テルモシリンジ SS-50ESZ) に針 (TERUMO 社製テルモカラン針 NN-2360C) を通し、チューブの液をシリンジに移し吐出させる。4 °C の M9 緩衝液 (3g/L KH PO (Wako 試薬特級), 6g/L Na HPO (Wako 試薬特級), 5g/L NaCl (Wako 試薬特級), 1mM MgSO (Wako 試薬特級)) を少量加えながらナイロンメッシュで濾過し、濾過液をビーカーに回収する。ビーカー内に回収された卵を 50ml チューブに移し、

†筑波大学 図書館情報メディア研究科

‡立命館大学 理工学研究科

表 1: Poly-A Controls の 4 項目 (lys, phe, thr, dap) の希釈率

項目	希釈率
lys	1:100,000
phe	1:50,000
thr	1:25,000
dap	1:7,500

4 °C,3000r.p.m. で 1 分間遠心分離を行う.上清を捨て, M9 を 40ml まで加え, 4 °C,2000r.p.m. で 1 分間遠心分離を行う. 上清を捨て, M9 を 30ml まで加え, 4 °C,2000r.p.m. で 1 分間遠心分離を行う. 上清を捨て, 沈殿した卵を 15ml チューブに移し, 4 °C,2000r.p.m. で 1 分間遠心分離を行う.

すべてのサンプルを同じ基準にならすため, コントロールによる発現量を標準化した. 標準化に用いるコントロールはすべてのデータにおいて Poly-A Controls の 4 項目 (lys, phe, thr, dap) を使用した. 表 1 は, Poly-A Controls の各項目の希釈率である.

実験データ毎に, Poly-A コントロールの発現量を基準値として, 全遺伝子 (22,548 個) の発現量を Poly-A の発現量に対する比率に変換した. こうして実験毎の変動を吸収し, すべてのサンプルを比較可能にする. その後, 発現量の値を対数 (底 2) で変換した.

$$\bar{c} = \frac{c_{lys} + c_{phe} + c_{thr} + c_{dap}}{4} \quad (1)$$

$$w_i = \log_2 \frac{v_i}{\bar{c}} \quad (2)$$

ここで, c_{lys} , c_{phe} , c_{thr} , c_{dap} は lys, phe, thr, dap のコントロールの単位当たりの発現量, \bar{c} は Poly-A コントロールの平均発現量, v_i は遺伝子 i のマイクロアレイ実験の発現量の値, w_i は遺伝子 i の正規化された発現量の値である. 分析には, このようにして得た One Cycle 法 3 サンプル, Express Kit 法 3 サンプルの, 22,548 遺伝子からなる実験データ (22,548 遺伝子×6 サンプルの遺伝子発現量の行列) を取り扱う.

まず, 全体的な類似性を調べるために, ピアソンの積率相関係数を用いて評価する. One Cycle 法の 3 サンプル, Express Kit 法の 3 サンプルの全ペア間の相関係数を計算する.

22,548 個の全遺伝子に対する発現変動の有意性の判断には, Welch t 検定による二群の有意差検定を行う. これによって, One Cycle 法と Express Kit 法で発現量が大きく異なる遺伝子を抽出する. 抽出する遺伝子は p 値が 0.001 以下の遺伝子とする. 帰無仮説を「二群間の平均値に差が無い」に設定し, p 値が 0.001 以下の遺伝子に対して対立仮説「差が無いとはいえない」を採択する. 対立仮説が採択された遺伝子を, One Cycle 法と Express Kit 法の実験手法の違いによる発現変動遺伝子とする.

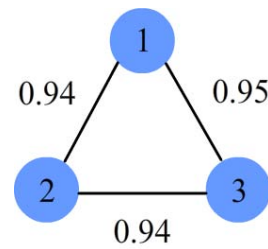


図 1: One Cycle 法による実験データ (3 回分) の間の相関係数. 数値は相関係数を表す.

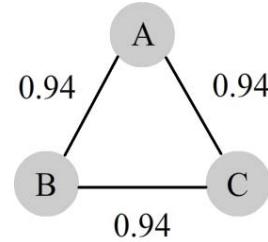


図 2: Express Kit 法による実験データ (3 回分) の間の相関係数. 数値は相関係数を表す.

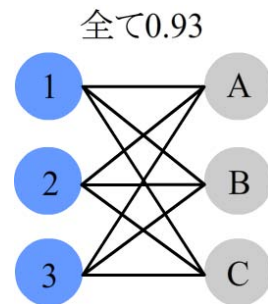


図 3: One Cycle 法による実験データ (3 回分) と Express Kit 法による実験データ (3 回分) の間の相関係数. 数値は相関係数を表す.

3. 結果および考察

One Cycle 法による実験データ (3 回分) 間の相関係数 (図 1). 数値はデータ間の相関係数を表す. 図中の 1, 2, 3 は One Cycle 法の 1 回目のデータ, 2 回目のデータ, 3 回目のデータに相当する. 相関係数平均 0.94 の, 高い正の相関が見られた.

Express Kit 法による実験データ (3 回分) 間の相関係数 (図 2). 数値はデータ間の相関係数を表す. 図中の A, B, C は Express Kit 法の 1 回目のデータ, 2 回目のデータ, 3 回目のデータに相当する. すべてのデータ間で, 相関係数 0.94 の, 高い正の相関が見られた.

One Cycle 法による実験データ (3 回分) と Express Kit 法による実験データ (3 回分) の間の相関係数 (図 3). 数値はデータ間の相関係数を表す. 図中の 1, 2, 3 は One Cycle 法の 1 回目のデータ, 2 回目のデータ, 3 回目のデータに相当し, A, B, C は Express Kit 法の 1 回目のデータ, 2 回目のデータ, 3 回目のデータに相当する. 異なる手法同士のすべてのデータ間で相関係数 0.93 の高い正の相関が見られた.

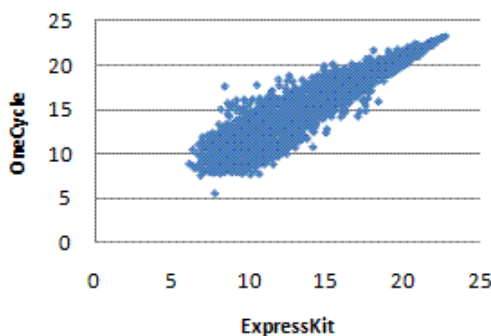


図 4: One Cycle 法の各遺伝子の発現量の平均値と Express kit 法の各遺伝子の発現量の平均値でプロットした散布図. 縦軸は One Cycle 法の発現量 (対数), 横軸は Express Kit 法の発現量 (対数) である.

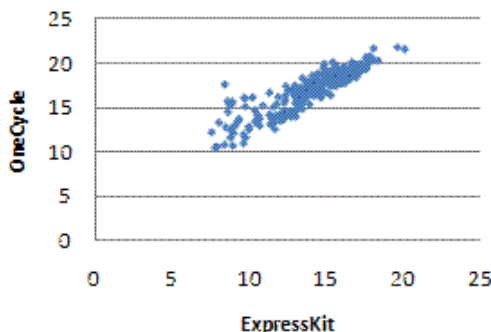


図 5: p 値 0.001 以下を満たす One Cycle 法の各遺伝子の発現量の平均値と Express Kit 法の各遺伝子の発現量の平均値でプロットした散布図. 縦軸は One Cycle 法の発現量 (対数), 横軸は Express Kit 法の発現量 (対数) である.

図 1-3 から明らかなように, One Cycle 法と Express Kit 法の全体的な相関は高い. 特に, One Cycle 法と Express Kit 法のデータ間の相関係数は, 同じ手法のデータ間より 0.01 低かったものの, ほぼ同程度の高い相関がみられた. 従って, One Cycle 法と Express Kit 法によるマイクロアレイ実験で得られる発現量は比例しており, 両者による実験データは比較可能であると言える. しかし, この値は Affymetrix 社が公表している 0.97 という値より小さい.

また, 図 4 から, One Cycle 法と Express Kit 法で検出された発現量の値に違いがある遺伝子群が確認できる. 発現量の平均 (対数) でプロットしてあり, データの最小値および最大値は, One Cycle 法が最小値 5.48, 最大値 23.20 であり, Express Kit 法が最小値 6.12, 最大値 22.77 であった.

大部分の遺伝子は, ほぼ直線上に位置するが, 少数の遺伝子の値にずれが見られる. これらの遺伝子は, One Cycle 法と Express Kit 法で発現が変動する遺伝子と考えられる.

One Cycle 法と Express Kit 法の発現量のデータ間の相関は高いものの, 生成される発現量の値については, One Cycle 法よりも Express Kit 法の方が発現量が低い傾向があった. 今回用いたデータでは, Poly A コントロー

表 2: 発現変動遺伝子として抽出された上位 20 個. 順位はより p 値が高い遺伝子, Probe Set ID はマイクロアレイのスポット名である.

順位	Probe Set ID	順位	Probe Set ID
1	192546_at	11	186199_at
2	172699_x_at	12	182981_at
3	178257_at	13	192802_s_at
4	177309_at	14	176088_at
5	190529_at	15	193805_at
6	188421_at	16	193503_s_at
7	191354_s_at	17	186225_at
8	190888_s_at	18	188130_s_at
9	193235_at	19	181177_at
10	192964_at	20	189103_at

ルの発現量で標準化されたデータの場合, One Cycle 法と Express Kit 法の発現量の比は 1.66 : 1.00 であった. 対数変換されたデータでは, この比は 1.08 : 1.00 であった.

従って, One Cycle 法と Express Kit 法のデータを発現の有無ではなく, 発現量の値を同時に使用する場合, ここで算出した比で値を変換する必要がある.

t 検定の結果, 発現変動遺伝子として認められた遺伝子 (p 値 0.001 以下) が, 線虫の遺伝子 22,548 個中 258 個 (全体の 1.14%) があった.

これらの発現変動遺伝子の中で, FDR (False Discovery Rate) は 8.74 であったため, 統計的には誤りが 22.55 個含まれていることになる. 経験的に FDR=5% の閾値になるように p 値を設定するのが一般的だが, 今回は n が少ないため FDR の閾値を緩めに設定した.

図 5 は発現変動遺伝子の散布図である. 発現変動遺伝子の発現量は, One Cycle 法が 10.37-21.75, Express Kit 法が 7.57-20.13 である.

図 4 に示したような, 手法の平均同士をプロットした散布図と比べ, p 値が 0.001 以下のものだけ取り出した遺伝子は, One Cycle 法と Express Kit 法両方において発現量が高いものが発現変動遺伝子となる傾向があった. Affymetrix のマイクロアレイ実験で生成されるデータには, スポット (遺伝子) 毎に値の信頼度 (p 値) が計算される. これらは, マイクロアレイによって検出可能な発現量の下限値と関連している. 発現変動遺伝子の発現量の信頼度を調べた (One Cycle と Express Kit 両方で信頼度があるスポットを抽出した) ところ, 167 個は信頼度があるスポットであった. 従って, 発現変動遺伝子の値は, 発現量の検出精度が原因ではなく, One Cycle 法と Express Kit 法の違いと, 遺伝子の発現形態や配列に起因すると考えられる.

検定で得られた発現変動遺伝子を, p 値が低い順にソートし, 発現変動遺伝子をランキング付けした. p 値がより低い遺伝子ほど One Cycle 法と Express Kit 法で発現が変動する遺伝子と仮定している.

発現変動遺伝子として検出された遺伝子の上位 20 個を表 2 に示す. 発現変動遺伝子として検出された遺伝子の機能をみると, 発生に関わる機能を持つものが多く, 特に, 検出された遺伝子の多くが, 孵化後の胚発生に関する機能

を持つという特徴が見られた。

4. おわりに

本研究では、線虫 *C.elegans* を対象として、マイクロアレイ解析のためのツール One Cycle 法と Express Kit 法による実験結果の違いを比較し、高い正の相関と発現変動遺伝子が見られた。しかし、一部の遺伝子の発現量に差があり、両手法の実験データの比較には、値の標準化が必要である。

参考文献

- [1] 岡野文香, 中山伸一, 伊藤将弘, 真栄城哲也, “線虫の初期胚発生に関連する遺伝子ネットワークの推定”. 情報処理学会第73回全国大会講演論文集.
- [2] T. Maeshito, H. Hemmi, K. Shimohara, “Ultra-fast Genome wide Simulation of Biological Signal Transduction Networks -- Starpack”, *Frontiers of Computational Science*, pp.177-180, Springer, 2007.
- [3] T. Maeshiro, K. Shimohara, S. Nakayama, “Fitness function for evolutionary system to predict unknown gene regulatory networks”, *Proceedings of the International Conference on Instrumentation, Control and Information Technology*, pp.2722-2727, 2009.
- [4] T. Maeshiro, Shin-ichi Nakayama, “Harmonic Pulse Analysis to detect biologically plausible gene regulatory networks”, *SICE Annual Conference 2010*, pp.3233-3239, 2010.
- [5] “Comparison between the GeneChip®3' IVT Express Kit and GeneChip One-Cycle Target Labeling and Control Reagents”
http://media.affymetrix.com/support/technical/technote/3_ivtexpresskit_technote.pdf
- [6] “White Paper: GeneChip®3' IVT Express Kit”
http://media.affymetrix.com/support/technical/whitepapers/3_ivtexpresskit_whitepaper.pdf