Vol.2012-BIO-29 No.26
2012/6/29

時系列発現プロファイルのための遺伝子機能グループ解析手法

大熊 祐太[1]　瀬尾茂人[1]　竹中要一[1]　松田秀雄[1]

概要：遺伝子の機能は研究され, 遺伝子に付加される機能情報は日々増加し続けている. こうした遺伝子機能情報を利用した解析手法の一つに, 2 つの実験条件の比較を目的とした遺伝子機能グループ解析がある. しかし, この解析手法では遺伝子機能の時間変化を解析することができない. そこで本研究では, 時系列データをスライディングウィンドウ方式で分割し, すべての分割期間に対して遺伝子機能グループ解析手法を実行することで時系列に対応できる遺伝子機能グループ解析を提案する. その結果, ある遺伝子機能が特定の期間で有意に発現していることを示した.

キーワード：時系列発現プロファイル, 遺伝子機能グループ解析, スライディングウィンドウ

Gene Set Enrichment Analysis for Time-series Gene Expression Profiles

Yuta Okuma[1]　Shigeto Seno[1]　Yoichi Takenaka[1]　Hideo Matsuda[1]

Abstract: Gene function is researched and gene functional information which is annotated on gene is increasing continuously. Gene Set Analysis is one of a method using gene functional information, and we use it when we want to compare two groups. However, this method can not be applied to time-series gene expression profile. In this reserch, I propose a method to analyze gene function groupsand handle time-series data. The method extracts a time period in which works from a time-series gene expression profile.

Keywords: Time-series Gene Expression Profile, Gene Set Enrichment Analysis, sliding window

1. はじめに

生物は体内に遺伝子を持っており, 遺伝子の発現量を調節することで必要量に応じた RNA やタンパク質を生成している. そこで, 遺伝子がどういった RNA やタンパク質を生成しその生成物によりどのような変化が生じるのかということを調査するために遺伝子発現解析が行われている. この遺伝発現解析を行う手法には, 有意差解析 [1], クラスタリング [2], ネットワーク解析 [3] など目的に応じて様々なものが提案されている. 遺伝子発現解析の結果, 遺伝子は単独で働くのではなく複数でネットワークを形成し, そのネットワーク内で互いの発現を促進・抑制することで複雑な生命機能を実現していることが判明した. また, 時系列に沿って測定された発現データを解析できる手法も登場し, 遺伝子発現がどのように進んでいくかを現実的に解析することが可能になっている.

このように遺伝子発現解析の手法は多く存在するが, その中の一つに遺伝子機能グループ解析 [4] がある. この手法では, 入力データを 2 つの群に分け, それぞれの群において各遺伝子機能ごとに発現解析し結果として対象においてどのような遺伝子機能が発現しているかを知ることができる手法である. この手法の実行結果は遺伝子機能がどちらの群でよく発現しているかを明らかにするもので, それがどの時期に発現したかを知ることはできない. 遺伝子機能グループ解析には様々なバリエーションがあり, PAGE[5], PGSEA[5], Gene Trail[6], SAM-GS[7], GSEA-P[8] などが提案されている. これらの手法に共通することは, 遺伝子

[1] 大阪大学大学院情報科学研究科
Osaka University Graduate Scool of information Science and Technology

のアノテーションという既知の情報を用いることで遺伝子グループというまとまった単位を形成し, そのグループごとに遺伝子の発現を解析することで遺伝子単体での解析結果では見つけることができなかった”生命機能としての発現変化”が結果として得られるということである. 遺伝子機能グループ解析のバリエーションは利用する統計量や実行環境の違いによって区別される.

遺伝子機能グループ解析では, 2 群のうち一方で発現量が大きく, もう一方で発現量が小さいというような機能を見つけることができる. しかし, 遺伝子機能グループ解析だけでは時系列に沿った解析を行うことができない. つまり, 遺伝子機能がどの時期に働いているかということまでは分からないということである. ところが遺伝子発現解析において時系列というのは非常に重要な要素であり, 時系列を考慮した上で既知の情報を用いて新たな知見を得ることが重要視されている. そこで本研究では時系列を考慮した上でデータベースから遺伝子機能情報を取得し, それらを各遺伝子機能ごとに解析する手法を提案する.

2. 遺伝子機能グループ解析

2.1 概要

遺伝子機能グループ解析では, 遺伝子アノテーションという既知の情報を利用し, 遺伝子が持つ機能を決定する. そのため, 最初に利用するデータベースを決定する必要がある. データベースが決定されるとその情報と発現データの遺伝子を元に遺伝子機能グループを作成する. この解析の目的は例えば, ”癌患者群 X と健常者群 Y の比較”, ”組織 X と組織 Y の比較”などのある群とある群の比較である. 遺伝子はある基準のもとにランク付けされ, 各遺伝子機能グループはそのグループ中の遺伝子のランクに応じたスコアをもつ. このスコアを比較することで X 群で発現しており, Y 群では発現していない遺伝子機能は機能 A である, といったような解析を行うことができる. 次節から, 遺伝子機能グループ解析で用いられる”遺伝子機能グループ”, ”遺伝子ランキング”, ”スコア”の 3 つについて説明を行う.

2.2 遺伝子機能グループ

遺伝子機能グループ解析は既知の知識として遺伝子機能グループを用いる. 遺伝子機能グループとは, 遺伝子を機能に応じてグループに分けたものであり, いくつかのデータベースが利用可能である. 最もよく利用されるのは登録数が多く, 用語の分類の細かい Gene Ontology である. Gene Ontology で定義された用語は GO term と呼ばれ, 3 つのカテゴリーに分かれている. 3 つのカテゴリーとは,

- biological process(生体内作用)
- cellular component(分子機能)
- molecular function(細胞内構成要素)

である. 解析の目的やデータとして用いる遺伝子に応じてこれらのカテゴリーから利用するものを決める. また, Gene Ontology は階層構造によって成り立っている. 上位層の GO Term は抽象度が高く, 下位層の GO Term は低くなっている. こうした遺伝子機能グループやパスウェイを用いて遺伝子発現解析を行おうという試みは盛んに行われている. [9][10][11]

統計量	ランキング方法
signal to noise	”二群間の平均の差が大きい”, ”群内のばらつきが小さい”遺伝子が上位
t 統計量	t 検定の統計量の絶対値が大きい遺伝子が上位
weighted average difference(WAD)	log 比を基本とし, 全体的にシグナル強度の高い遺伝子が上位
発現変動率	群間の発現量の変化が大きい遺伝子が上位

表 1: GSEA で使う統計量

2.3 遺伝子ランキング

遺伝子機能グループがどの程度発現しているかを評価する基準としてスコアを計算する. そのスコアを計算する際に基準となるのが遺伝子ランキングである. ランキングを作成する際には 2 群間で統計量の比較を行い, その統計量の値で遺伝子の順位を決める. 比較に用いる統計量とランキングの方法については表 1 を参照していただきたい. 例えば, signal to noise における遺伝子 i の統計量は,

$$R(i) = \frac{\overline{X^i} - \overline{Y^i}}{U_{X^i} + U_{Y^i}} \tag{1}$$

と表される. 従って, この統計量では二群間の平均の差大きく, 各群でのばらつきが小さい遺伝子の統計量が大きくなる. t 統計量は t 検定の統計量であり,

$$R(i) = \frac{\overline{X^i} - \overline{Y^i}}{\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\sqrt{\frac{(n_A-1)U^2_{A^i} + (n_B-1)U^2_{B^i}}{n_A+n_B-2}}} \tag{2}$$

のように表される. t 統計量では|R(i)|が大きいものを発現変動遺伝子の候補とする. 統計量には, 正と負があるものと正だけのものがある. signal to noise や発現変動率では正負があり, 統計量の変化の度合いを表している. 一方, t 統計量や WAD では絶対値をとることで変化の大きさを表している.

2.4 Enrichment Score と標準化

2.4.1 Enrichment Score

Enrichment Score(以下, ES) とは, 2.3 節で述べた遺伝子のランキングリストに基づいて計算されたスコアである. 全遺伝子を N 個, ある遺伝子機能グループを S, ランキングされた遺伝子リストを L とする. 遺伝子リスト L の上位から順に遺伝子をみて, S 中の遺伝子が登場すれば S(”hit”), 登場しなければ S(”miss”) としそれぞれの重みを計算する.

重みの計算式は,

$$P_{hit}(S,i) = \sum_{\substack{g_j \in S \\ j \leq i}} \frac{|r_j|^p}{N_R}, where N_R = \sum_{g_j \in S} |r_j|^p \qquad (3)$$

$$P_{miss}(S,i) = \sum_{\substack{g_j \notin S \\ j \leq i}} \frac{1}{(N - N_H)} \qquad (4)$$

である. ここで, 指数 p は各ステップにおける重みを調整するための値である. ES は $P_{hit} - P_{miss}$ のゼロからの最大のずれとして定義される. つまり, ある遺伝子機能グループに含まれている遺伝子の多くが上位にランクインしていれば図 1 のように遺伝子機能グループの ES は高くなる. 逆に, 遺伝子機能グループ内の遺伝子の発現変動率にばらつきがあれば, 図 2 のように ES は低い値にしかならない.

2.4.2 Normarized Enrichment Score

遺伝子機能グループに含まれる遺伝子数は多いものから少ないものまで様々である. ES はグループに含まれる遺伝子の数が多ければ高くなり, 少なければ低くなる傾向がある. そのため ES のみを用いて遺伝子機能グループ間での比較を行うことはできない. そこで用いるのが ES を標準化したスコアである Normalized Enrichment Score(以下, NES) である. この NES はランダムサンプリングによって求められる. ここで, M 回のランダムサンプリングを行うとし i 回目に算出された ES を ES_i とすると,

$$\mathrm{NES} = \frac{ES}{\frac{1}{M}\sum_{i=1}^{M} ES_i} \qquad (5)$$

と表される. この方法で標準化することでグループ内の遺伝子数の影響を受けることがなくなり, グループ間での比較を行うことができる.

3. 時系列遺伝子発現プロファイルのための遺伝子機能グループ解析手法

従来の遺伝子機能グループ解析は時系列を考慮した解析を行うことができない. そこで, 本研究ではスライディングウィンドウ方式を用いて時系列データを初めから順に一定期間で切り取り, 切り取った期間とそれ以外の期間を比較することで 2 群比較を行う. 切りだされたすべての期間についてこの 2 群比較を行うことで, 遺伝子機能グループ解析を時系列データへと適用する. それぞれの 2 群比較結果として, 発現変動率に基づくスコアが求まる. それを”時間について”と”遺伝子について”の 2 種類の検定を行い, 発現が有意であると判定された期間を, ”ある機能がよく発現している期間”として出力する. 以下が提案手法の流れ, 図 3 が提案手法の全体図である.

STEP1:
解析領域 (以下, ウィンドウ) を決定し, 入力された時系列遺伝子発現プロファイルの始めから終わりまでウィンドウをスライドさせることで注目する期間を決定する. 次に, 発現変動率のよる遺伝子ランキングを作成し, そのランキングに基づいて ES を計算する.

STEP2:
ある期間について, 全ての遺伝子機能グループを参照し, STEP1 で求めたスコアが他の期間に比べて有意であるかを判定する.

STEP3:
STEP2 で有意であると判定された期間内で他の遺伝子機能グループと比較を行い, 有意であるかどうかを判定する. STEP2 と STEP3 で有意と判定できた遺伝子機能グループの期間を出力する.

以下では, 全遺伝子数を G 個, 個々の遺伝子を g, ある遺伝子機能グループを S, 選択した期間を T, 全時点数を timepoint, ウィンドウサイズを w として説明を行う.

3.1 STEP1:時系列データへの適用

データの切り出し方はには, スライディングウィンドウ方式を用いる. データの切り出し方は以下の通りである.

(1) 切り出すサイズ (ウィンドウサイズ)w を決定する.
(2) データの始めにウィンドウを設定し, ウィンドウのある部分を選択期間, それ以外の部分を非選択期間とする.
(3) 以降ウィンドウを 1 時点づつスライドさせ期間 T を決定する.
(4) 遺伝子ランキングを作成し, 切りだされたすべての期間についてランキングに基づいた ES の計算を行う.

ここで切り出された期間の数は timepoint-w+1 個であり, それぞれの期間は T_1, T_2, ..., $T_{timepoint-w+1}$ と表される. (4) で切り出した全ての期間について発現変動率に基づくスコアが計算されているので, そのすべてについて有意差検定を行う.

3.2 有意差の検定

本研究では, ES による比較を行うことで発現している期間を特定するが, 計算された ES が偶然算出された可能性もある. そのため, ES がどれぐらい有意なのか, つまりどの程度偶然に起こりにくいかを判定する必要がある. 有意性の判定に必要な項目は,

- 時間についての有意性判定
- 遺伝子間での有意性判定

の 2 種類である. 時間についての有意性判定では, 全時間

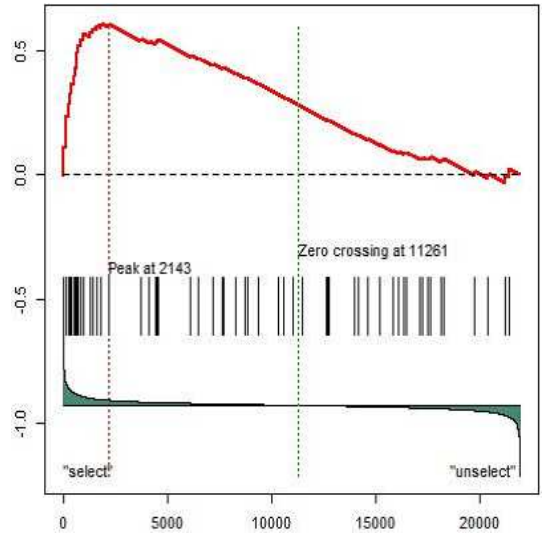

図 1: ES が高くなる場合

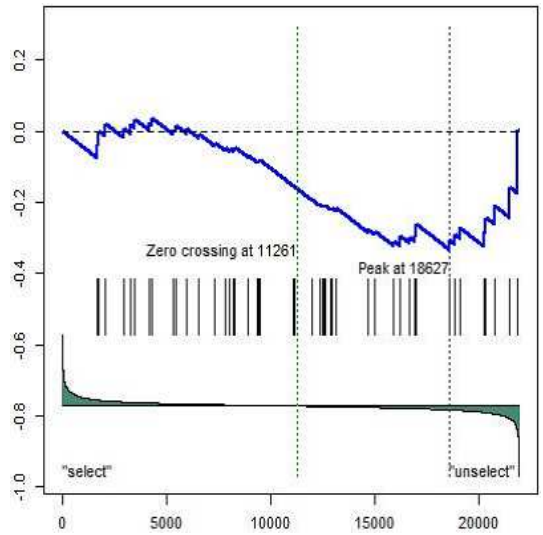

図 2: ES が高くならない場合

からランダムに選択された期間で ES を計算し, 分布をかく. そして, 元の ES がその分布上でどこに位置するかで有意性を検定する. 遺伝子間での有意性判定では, 比較対象の遺伝子機能グループ中の遺伝子と同数の遺伝子を全ての遺伝子からランダムに選出する操作を M 回行い, それぞれについて ES を計算し分布を描く. その分布において元の ES がどの程度有意なのかを検定する. 検定方法には Kolmogorov-Smirnov 検定を用いた.

3.2.1 STEP2:時間に関する有意差の検定

ある期間 T_j について判定を行う場合を考える. 検定は以下の手順で行う.

(1) すべての期間から M 回のランダムサンプリングを行い, ES_1, ES_2, ..., ES_M を計算
(2) (1) で得られた ES の分布を作成
(3) 分布から元の ES が出現する確率 p と False Discovery Rate(FDR) を算出

(1) でのランダムサンプリングでは同一遺伝子の別の期間を選択することでランダムサンプリングとしている. また, ここで選択する期間は連続する期間でなくてもよく, 合計が選択期間の幅と同じになればよい.

また, (3) で算出される p 値は以下の式で計算される.

$$Pr(ES(N, N_H) \leq \lambda) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2\lambda^2 n),$$

$$n = \frac{(N - N_H)N_H}{N} \qquad (6)$$

式中の λ は ES を, N_H は遺伝子機能グループ H 中の遺伝子の数を表している. 仮に期間 T_j が他の期間に比べて有意に発現している期間であれば, 期間についてのランダムサンプリング行った場合 ES の平均値は元の ES の値に比べて小さくなる. 逆に T_j での発現が有意でない期間であれば, ランダムサンプリングを行なっても ES の値はそれほど変わらず, ES の変動は小さくなる.

3.2.2 STEP3:遺伝子機能間での有意差の検定

STEP2 で有意に発現している期間が見つかった場合, 注目している遺伝子セットが有意に発現している期間が分かったことになる. しかし, これだけでは不十分である. なぜなら, 計測時のミスによって発現量に一定の偏りが生じており, この期間内であれば全ての遺伝子機能グループが他の期間に比べて有意に発現していると判定される可能性もあるからである. 従って, STEP2 で得られた結果をさらに遺伝子間で比較することが必要となる. ここでは, STEP2 で有意と判定された期間 T_j において遺伝子機能グループ S の ES が遺伝子機能グループ中の遺伝子を変えても有意であるかどうかの検定を行う.

(1) 期間 T_j において遺伝子 g_1, ..., g_G について M 回のランダムサンプリングを行い, ES_1, ES_2, ..., ES_M を計算
(2) (1) で得られた ES の分布を作成
(3) 分布から元の ES が出現する確率 p と False Discovery Rate(FDR) を算出

ここで, M 回のランダムサンプリングとは遺伝子機能グループ中の遺伝子と同じ数の遺伝子を全遺伝子からランダムで抽出し, それらを遺伝子機能グループ S の遺伝子として考えることを言う. この結果有意であると言えた場合, 期間 T_j において遺伝子機能グループ S が有意に発現しているということが言える. つまり, STEP2 と STEP3 の検定を行い双方で有意と判定された場合のみ本研究の目的である”時系列遺伝子発現プロファイルからある遺伝子機能がもっとも活発に発現する時期を特定”することができたとする.

4. 実験と考察

4.1 実験の目的と概要

本実験では, 提案手法と従来手法の比較を行い, それらの

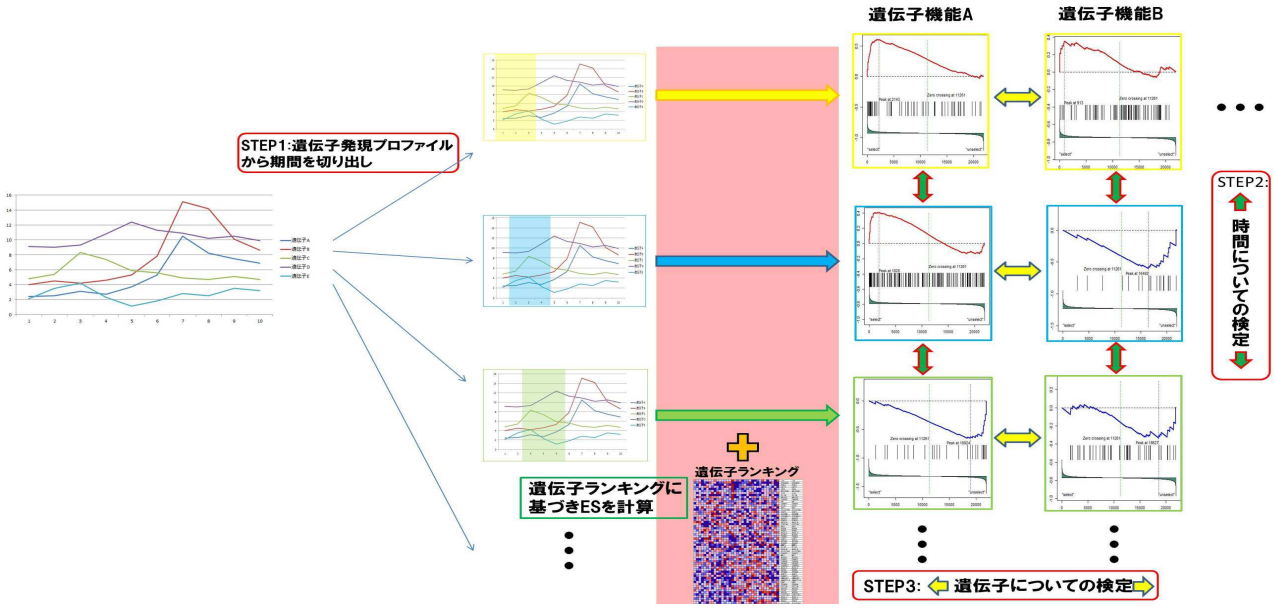

図 3: 時間分割と検定の流れ

実行結果に違いがあることを確かめる. また, 提案手法を実行して遺伝子機能が発現していると判定された期間を出力し, その中からどのような遺伝子機能が発現しており, それらが既知の情報と合致するかどうかを議論する.

4.2 実験条件

使用データ

・時系列遺伝子発現発現プロファイル

本研究で用いる時系列遺伝子発現プロファイルは, マウスの脂肪細胞分化と骨芽細胞分化の時系列遺伝子発現プロファイルである. 測定時点数は, 分化開始直後の 0 分, 5 分, 15 分, 30 分, 45 分の 5 時点. 1 時間間隔で 1, 2, ..., 29, 30 時間までの 30 時点, 6 時間刻みで 36, 42, ..., 186, 192 時間までの 27 時点の計 62 時点存在している. また, 実験データに用いる対象遺伝子は 21947 個である.

・遺伝子機能グループ

使用する機能遺伝子グループは, Gene Ontology の biological processes に属する遺伝子セットである.

・ウィンドウサイズ

今回の実験では, 注目する期間を 10 時点とし, 開始時点から終了時点までスライドさせて各期間について解析を行う

・遺伝子のランキング方法

今回は発現量の変化の大きさだけではなく, 発現量の正負と変化量の 2 つを基準とするために発現変動率によるランキングを採用した.

・閾値

閾値として検定で算出された FDR を用いる. 時間についての検定, 遺伝子についての検定の両方で FDR が 25%を下回った期間のみを出力する.

4.3 結果

4.3.1 従来手法との比較

従来手法として, ウィンドウで区切らずに時系列遺伝子発現プロファイルを 31 時点と 31 時点の 2 群に分割し遺伝子機能グループ解析を行う. 時系列を考慮することにより, 提案手法で脂肪細胞分化に関わりがあると思われる遺伝子機能をどの程度判定でき, 従来手法との差がどれぐらいあるのかを調べる. 脂肪細胞分化に関わりがあると思われる遺伝子機能は表 2 であり, それぞれの手法で発現していると判定できた場合は"○", 判定できなかった場合は"-"としている. 提案手法では表に挙げた遺伝子機能のすべてが検出できているのに対して, 従来手法で検出できたのは Fatty acid metabolic process と Fatty acid oxidation の 2 つのみであることがわかる. したがって, 従来手法では検出できなかった遺伝子機能の発現期間が提案手法を実行することで検出できるようになったと考えられる.

遺伝子機能名	提案手法	従来手法
Fatty acid metabolic process	○	○
Fatty acid oxidation	○	○
Cellular lipid metabolic process	○	-
Phospholipid biosynthetic process	○	-
Lipid transport	○	-
Phospholipid metabolic process	○	-
Lipid metabolic process	○	-
Lipid catabolic process	○	-
Cellular lipid catabolic process	○	-

表 2: 脂肪細胞分化に関わりがあると思われる遺伝子機能

4.3.2 実行結果と発現期間

図 5 が脂肪細胞分化時のデータでの結果, 図 6 が骨芽細

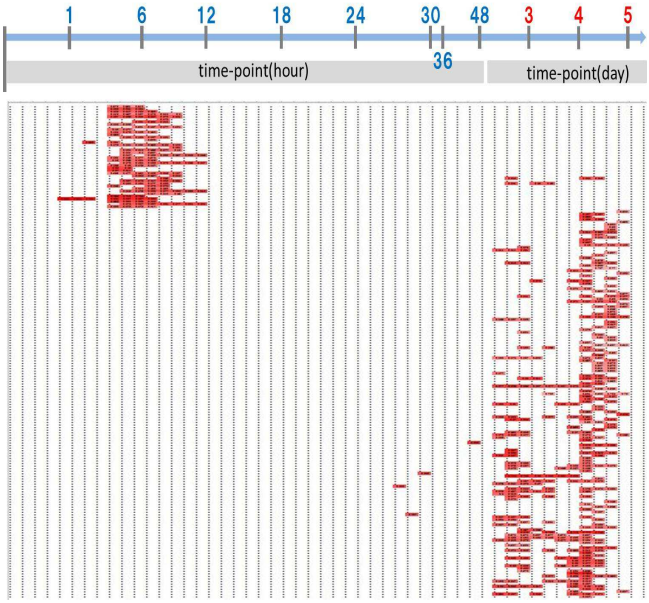

図 5: 脂肪細胞についての結果 (283 遺伝子機能グループ)

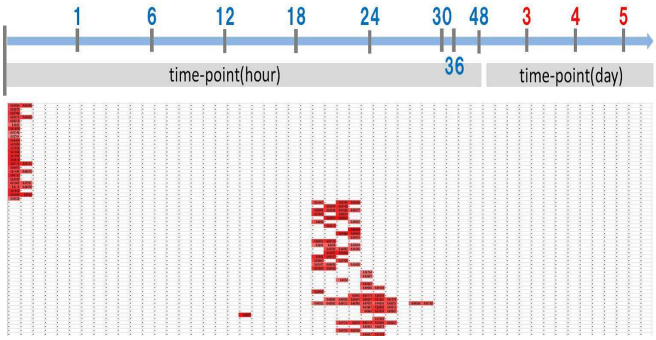

図 6: 骨芽細胞についての結果 (60 遺伝子機能グループ)

胞分化時のデータでの結果である. ES が高く, 他の期間や他の遺伝子機能に比べて有意に発現していると判定できた期間を赤で表している. この結果から, 脂肪細胞, 骨芽細胞共に強く発現している時期が 2 つに分かれていることが分かる. それぞれ, 脂肪細胞では 1 時間〜12 時間の間と 2 日目以降, 骨芽細胞では 1 時間までと 18 時間〜30 時間の間において強く発現している期間が続いている.

さらに, 脂肪細胞で発現していると判定されたものから表 2 に挙げた遺伝子機能を抽出したものが図 4 である. これらの機能についても 2 日目以降で強く発現していることが分かる.

4.4 考察

細胞や構成組織の変化や DNA から mRNA への転写など, 分化そのものや生命維持に関係のある機能については早い時期に, 表 5 に挙げた脂肪酸や脂質に関する化学反応と代謝経路など脂肪細胞に特有な遺伝子機能や骨芽細胞に特有な遺伝子機能など, 特定の状況において利用される機能は遅い時期に発現することが知られている. そこで, 提案手法で検出できた脂肪細胞分化での結果について検証する. 図 5 において初期に発現している遺伝子機能には, "translational initiation"(DNA の翻訳開始時に働く機能) や"RNA splicing"(RNA のスプライシング), "RNA processing"(未完成な RNA がより完全な RNA へと変換される過程) などの細胞の維持や変化そのものに関わる遺伝子機能が多い. さらに, これらは 1〜12 時間の間で強く発現しており, 一度発現しなくなると以降再び強く発現することはない. 後半に強く発現している遺伝子機能には表 2 で挙げた脂肪細胞に特有な遺伝子機能がある. これらの遺伝子機能は前半に発現しておらず, 後半に発現していることが図 4 から分かる. 以上のことより, 提案手法で得られた結果は既知の情報に合致する部分が多く, 信頼出来る解析結果だといえる. さらに, 後期で発現している遺伝子機能が初期にあまり発現していないのは, 分化という現象を進めるための機能が強く働く際に関係のない機能を抑制しているためだと思われる.

骨芽細胞の結果では, 脂肪細胞の場合と比べ多くの遺伝子機能において発現していると判定できる期間が存在しないという結果となった. これは, 脂肪細胞と骨芽細胞の発現期間の差によるものであると考えられる.

5. おわりに

本研究では, 遺伝子機能グループが有意に発現している期間の特定を目的として, ある遺伝子機能グループが他の遺伝子機能グループよりも有意に発現している時期を見つけることができるということを示した. さらに, 実験の章では, 従来手法よりも提案手法のほうが多くの遺伝子機能を検出できており, 時系列データへの対応ができていることを示した. さらに, 提案手法で検出した遺伝子機能の発現期間が生物学的な観点から見ても矛盾のないものであることを示した.

参考文献

[1] Tusher, V. G., Tibshirani, R. and Chu, G.: *SIGNIFICANCE ANALYSIS OF MICROARRAYS*, 2001.

[2] Daxin J., Chun T. and Aidong Z.: *Cluster analysis for gene expression data: a survey*, Knowledge and Data Engineering, IEEE Transactions on, Vol.16, p. 1370 - 1386 , 2004.

[3] Hiroyuki T. and Katsuhisa H.: *Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling*, Bioinformatics, Vol.18, p. 287 - 297 , 2002.

[4] Aravind S., Pablo T., Vamsi K. M., Sayan M., Benjamin L. E., Michael A. G., Amanda P., Scott L. P., Todd R. G., Eric S. L. and Jill P. M.: *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*, Proc Natl Acad Sci U S A, Vol.102, p. 15545–15550, 2005.

[5] Seon-Young K. and David J. V.: *PAGE: parametric anal-*

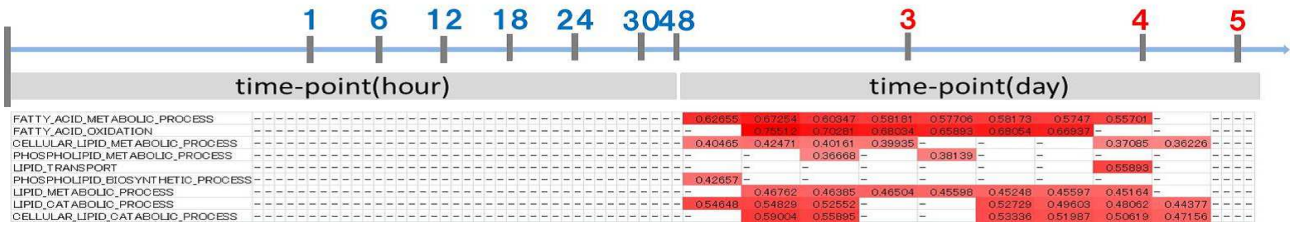

図 4: 抽出した遺伝子機能

ysis of gene set enrichment, BMC Bioinformatics, Vol.6, p. 144, 2005.

[6] Christina B., Andreas K., Jan K., Benny K., Nicole C., Yasser A. E., Rolf M., Eckart M. and Hans-Peter L.: *GeneTrail-advanced gene set enrichment analysis*, Nucleic Acids Research, Vol.35, p. 186-192, 2007.

[7] Irina D., John D. P., Thomas M., Qi L., Adeniyi J. A., Gian S. J., Gunilla E., Konrad S. F., Philip H. and Yutaka Y.: *Improving gene set analysis of microarray data by SAM-GS*, BMC Bioinformatics, Vol.8, p. 242, 2007.

[8] Aravind S., Heidi K., Joshua G., Pablo T. and Jill P. M.: *GSEA-P: a desktop application for Gene Set Enrichment Analysis*, Bioinformatics, Vol.23, p. 3251–3253, 2007.

[9] Doniger S. W., Salomonis N., Dahlquist K. D., Vranizan K., Lawlor S. C. and Conklin B. R.: *MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data*, Genome Biol, Vol.4, p. R7, 2003.

[10] Zhong S., Storch K. F., Lipan O., Kao M. C., Weitz C. J. and Wong W. H..: *GoSurfer: a graphical interactive tool for comparative analysis of large gene sets in Gene Ontology space*, Appl Bioinformatics, Vol.3, p. 261-264, 2004.

[11] Gabriel F. B., Oliver D. K., Barbara B., Chris S. and Frederick P. R.: *Characterizing gene sets with FuncAssociate*, Bioinformatics, Vol.19, p. 2502-2504, 2003.