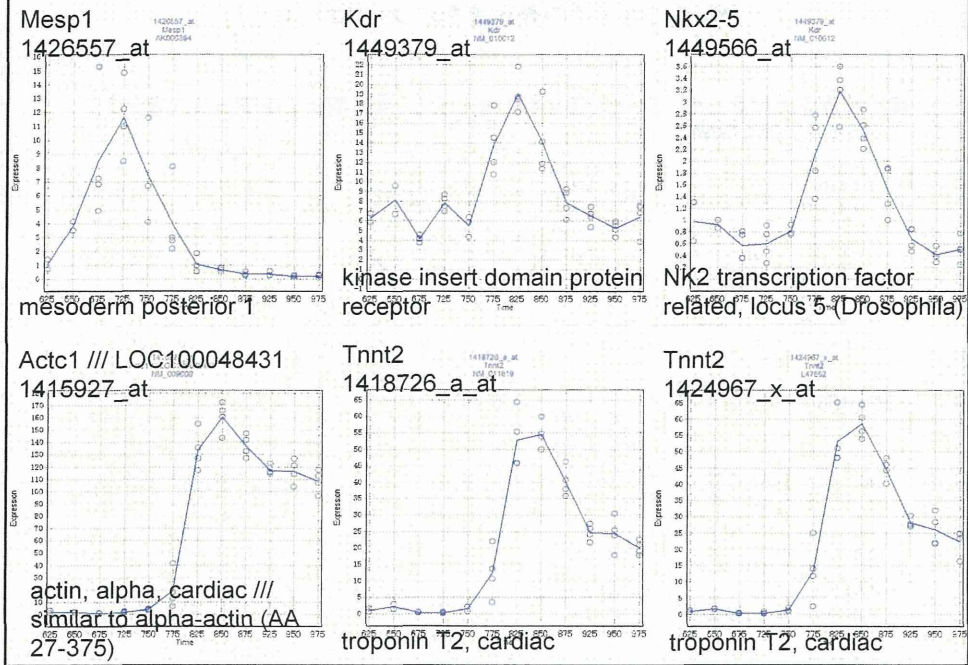


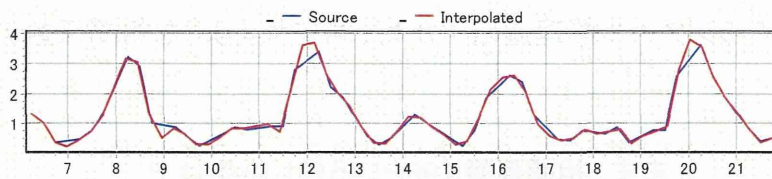
Nkx2-5シグナル関連遺伝子の発現変動



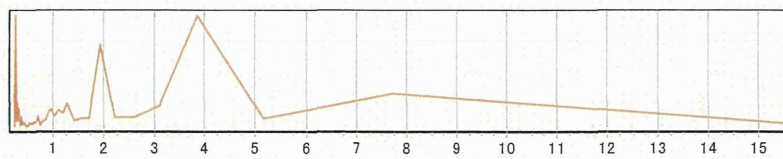
周期性を検討する為に MF Wave analyzer を利用し、
Nkx2-5 遺伝子の発現変動をフーリエ変換した

胎生 6.25、6.50、6.75、7.25、7.50、7.75、8.25、8.50、8.75、9.25、9.50、9.75 = [Time point : 12点 (6時間毎、n = 2~4)]

→ 解析精度を上げるため、データ(各点4例)を四分し、4周期 (= 14日間) として扱った
スプライン計算 [補関数: 64]



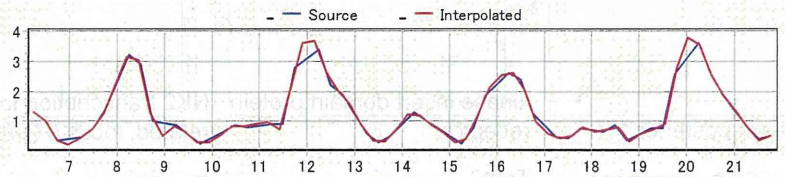
波長分布



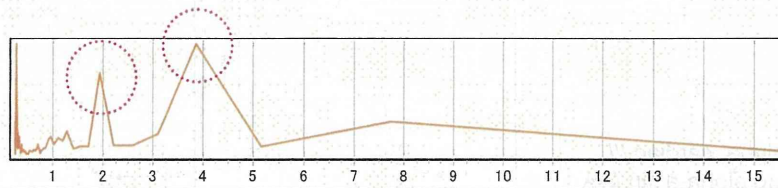
周期性を検討する為に MF Wave analyzer を利用し、
Nkx2-5 遺伝子の発現変動をフーリエ変換した

胎生 6.25、6.50、6.75、7.25、7.50、7.75、8.25、8.50、8.75、9.25、9.50、9.75 = [Time point : 12 点 (6 時間毎、
n = 2~4)]

→ 解析精度を上げるため、データ(各点4例)を四分し、4周期(= 14日間)として扱った
スプライン計算 [補関数: 64]



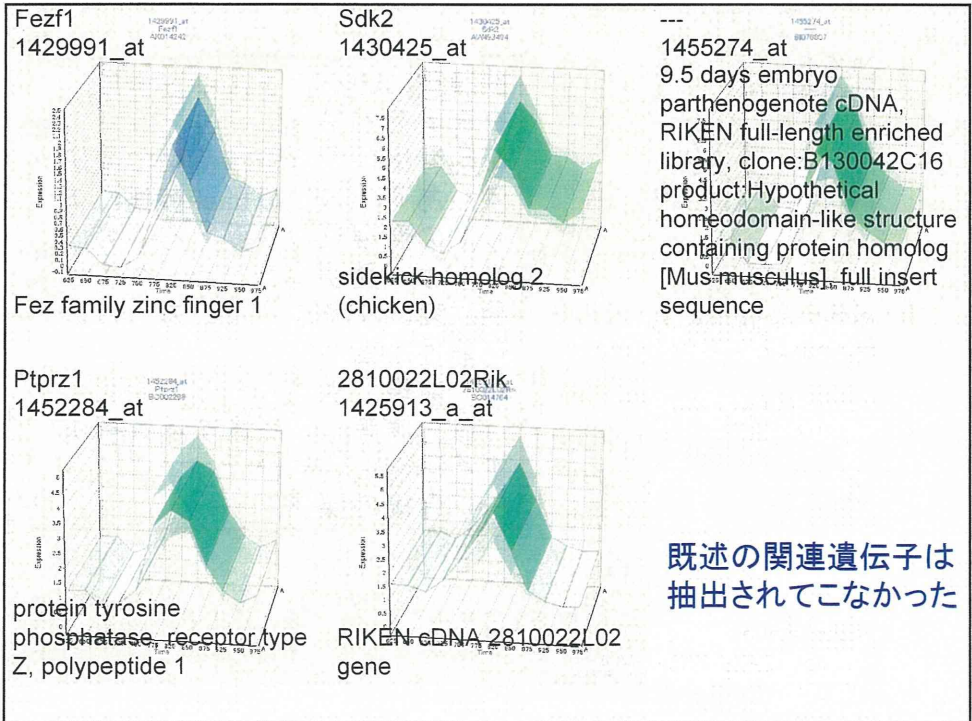
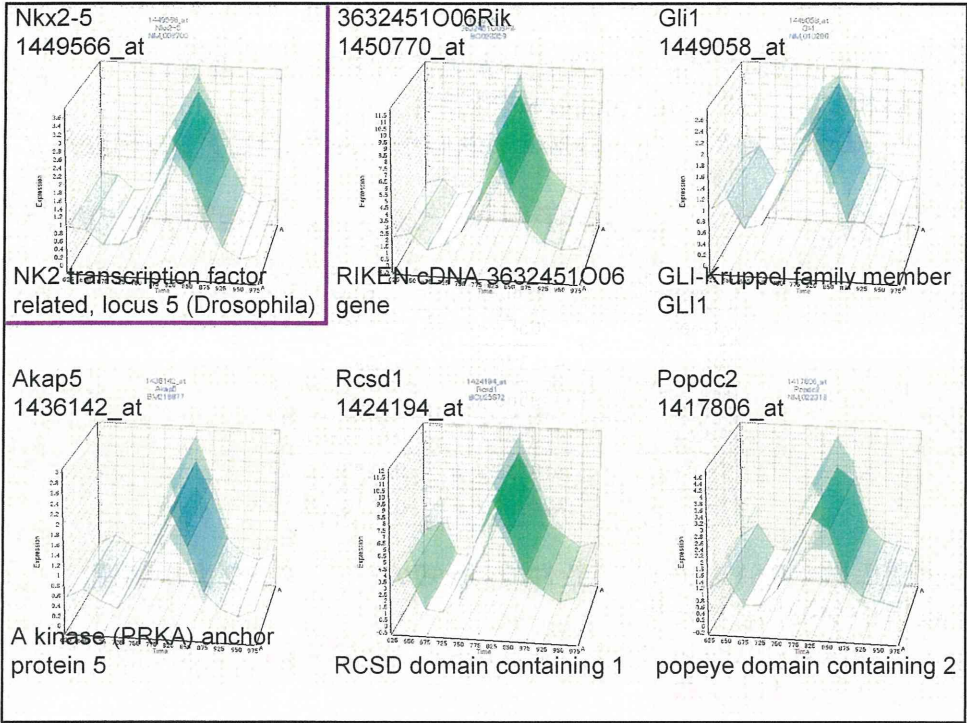
波長分布



- ・周期性を検討する為に、遺伝子発現変動をフーリエ変換した
[解析ソフト: MF Wave analyzer を利用]
→ この波長分布について Nkx2-5 遺伝子を基としたピアソン相関解析

■ Nkx2-5 シグナル関連遺伝子が皆、含まれるような相関係数の検討

まず、相関係数 0.98 という厳しい条件で
抽出されてきた遺伝子 (11 ps) の 3 次元表示
(目視による選別なし)



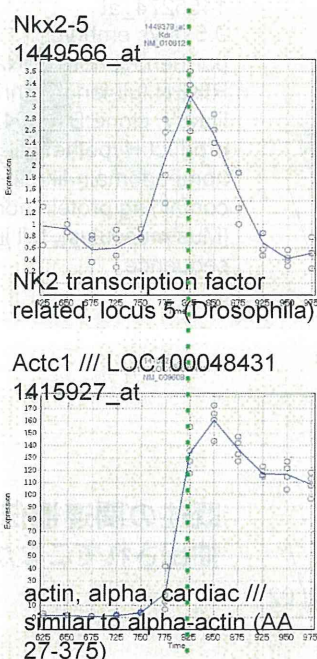
- ・周期性を検討する為に、遺伝子発現変動をフーリエ変換した
[解析ソフト: MF Wave analyzerを利用]
→この波長分布についてNkx2-5遺伝子を基としたピアソン相関解析

■ Nkx2-5シグナル関連遺伝子が皆、含まれるような相関係数の検討

| 遺伝子名 | ProbeID | RSQ |
|--------|--------------|-------------|
| Mesp1 | 1426557_at | 0.930527432 |
| Kdr | 1449379_at | 0.939394755 |
| Nkx2-5 | 1449566_at | 1 |
| Actc1 | 1415927_at | 0.811926939 |
| Tnnt2 | 1418726_a_at | 0.922013317 |
| Tnnt2 | 1424967_x_at | 0.911670319 |

相関係数が0.81以上という条件で、既知のNkx2-5シグナル関連遺伝子が抽出されたきた

→相関係数が0.81以上という条件により、25,254 psが抽出されてきた
※全ps中、ほぼ半数



発現誘導のシグナルネットワークを考える場合、細胞骨格等の細胞構造物の遺伝子、あるいはfeed back機構により別の発現制御を受ける等の場合には、発現の最大ピーク後の発現変動を考慮しない方が良い可能性が高い

→ 発現の立ち上がりの勾配 (微分値等) を考慮すれば解決する可能性が高い

→ 今後の課題

- ・周期性を検討する為に、遺伝子発現変動をフーリエ変換した
[解析ソフト: MF Wave analyzerを利用]
→この波長分布についてShh遺伝子を基としたピアソン相関解析

■ Nkx2-5シグナル関連遺伝子が皆、含まれるような相関係数の検討

| 遺伝子名 | ProbeID | RSQ |
|--------|--------------|-------------|
| Mesp1 | 1426557_at | 0.930527432 |
| Kdr | 1449379_at | 0.939394755 |
| Nkx2-5 | 1449566_at | 1 |
| Actc1 | 1415927_at | 0.811926939 |
| Tnnt2 | 1418726_a_at | 0.922013317 |
| Tnnt2 | 1424967_x_at | 0.911670319 |

相関係数が0.91以上という条件で、Actc1以外の既知のNkx2-5シグナル関連遺伝子が抽出されたきた

→ 相関係数が0.91以上という条件により、3,319 psが抽出されてきた

Nkx2-5遺伝子を基としフーリエ変換して得られた波長分布についてピアソン相関解析の結果得られた (相関係数>0.91)

3,319 psについて、

RSortで経時的に並びかえ(3.5日間のフォーマットで)、発現の最大ピークを示す各時点毎に抽出 (目視の検討はなし)

Nkx2.5遺伝子: 胎生8.25日に発現ピークを持つ遺伝子として抽出されてきた

結果 ← 3,319 ps

胎生6.25日: 292 ps
 胎生6.50日: 1,150 ps
 胎生6.75日: 8 ps
 胎生7.25日: 3 ps
 胎生7.50日: 16 ps
 胎生7.75日: 1,364 ps
 胎生8.25日: 280 ps [含、Nkx2-5遺伝子1 ps]
 胎生8.50日: 139 ps
 胎生8.75日: 0 ps
 胎生9.25日: 14 ps
 胎生9.50日: 3 ps
 胎生9.75日: 50 ps

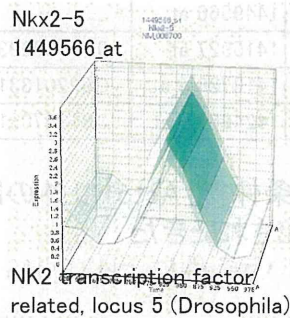
Setting:
 Explore Direction:
 Peak>>>Valley
 Posi,Intermed,Nega:全てcheck

周波数解析は周波数に着目した解析であり、直接、遺伝子発現の経時パターンを反映したものではない

Q1: どの程度、発現パターンを反映しているのか?

→

- Nkx2-5遺伝子が含まれる、発現の最大ピークが胎生8.25日である280 psについての検討
- 各時点での代表例について検討

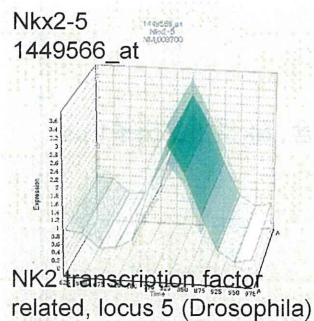


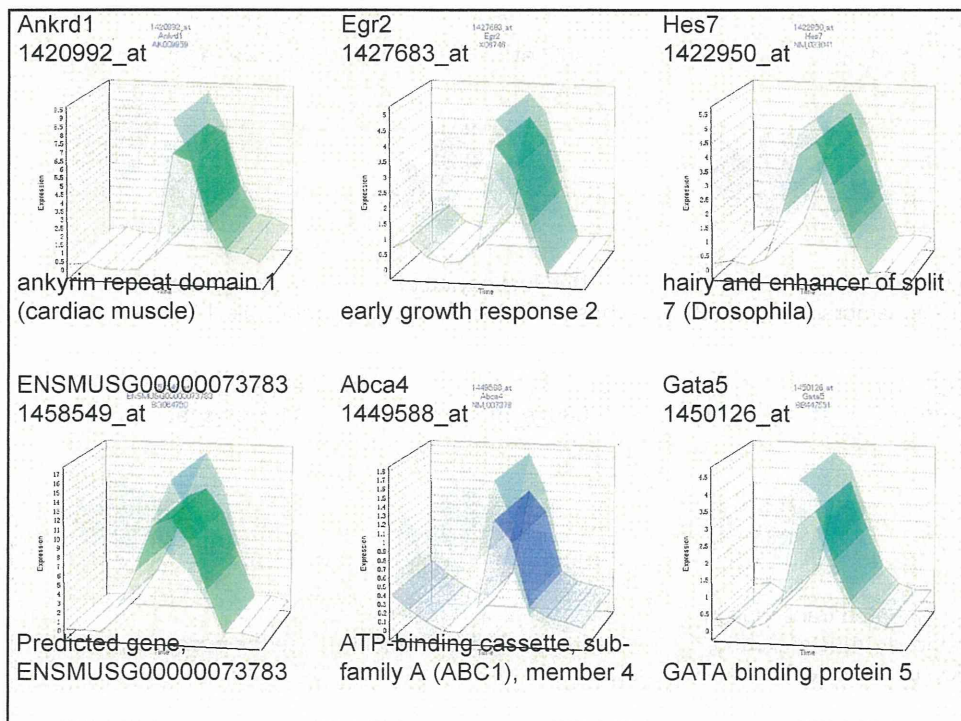
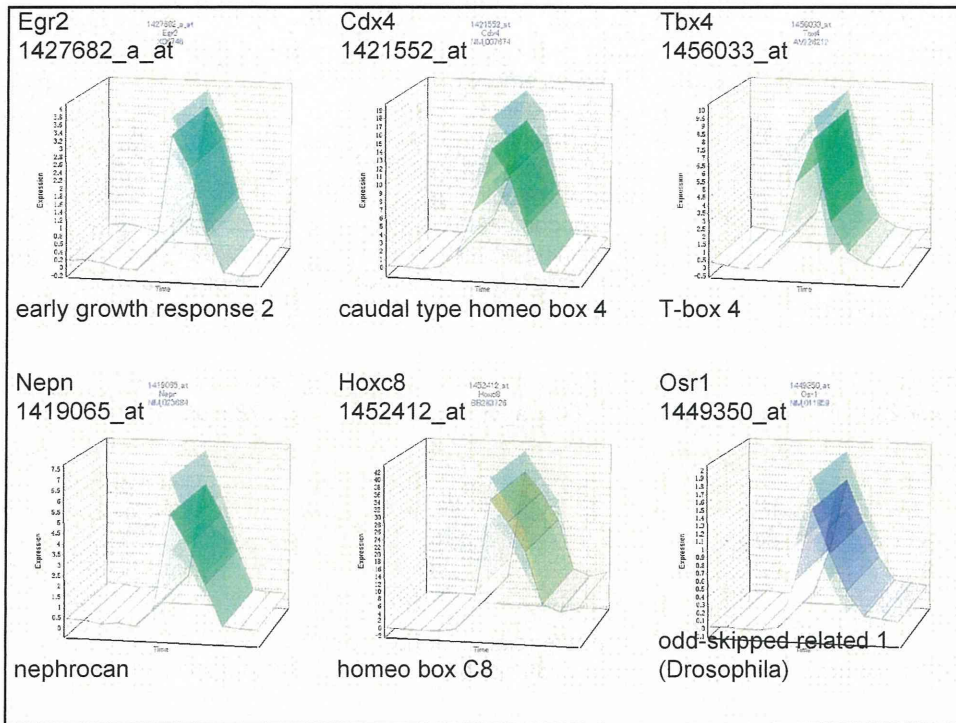
周波数解析は周波数に着目した解析であり、直接、遺伝子発現の経時パターンを反映したものではない

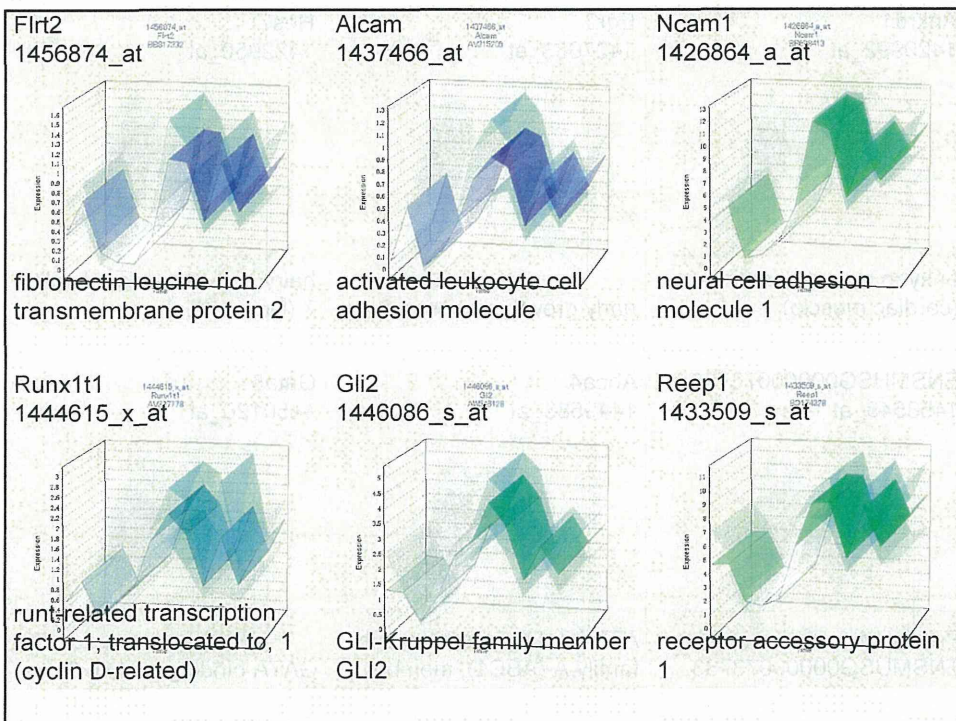
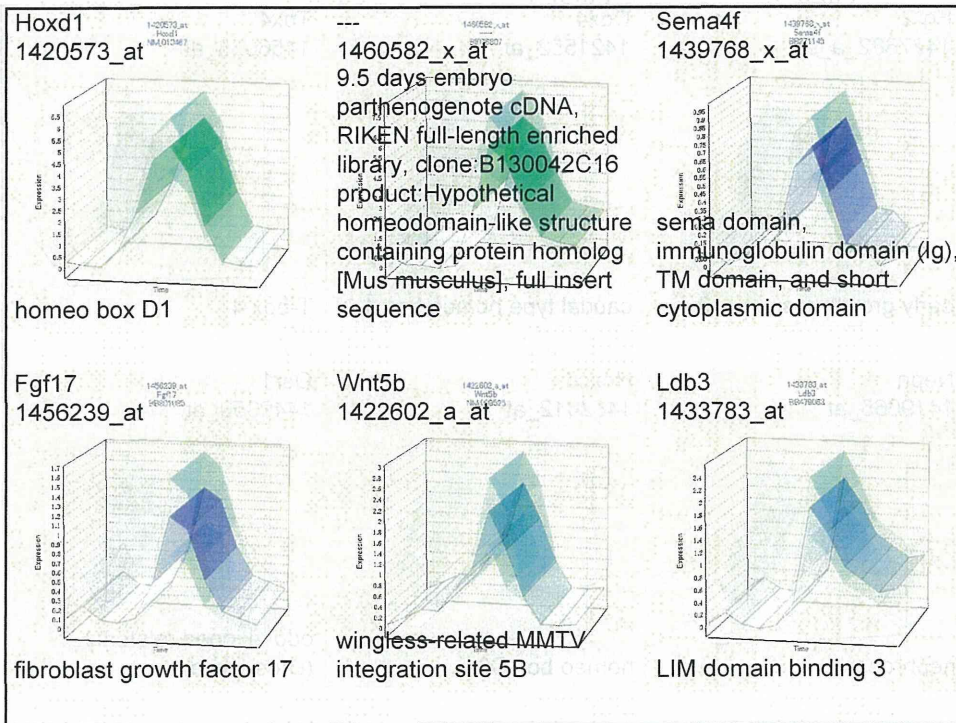
Q1: どの程度、発現パターンを反映しているのか?

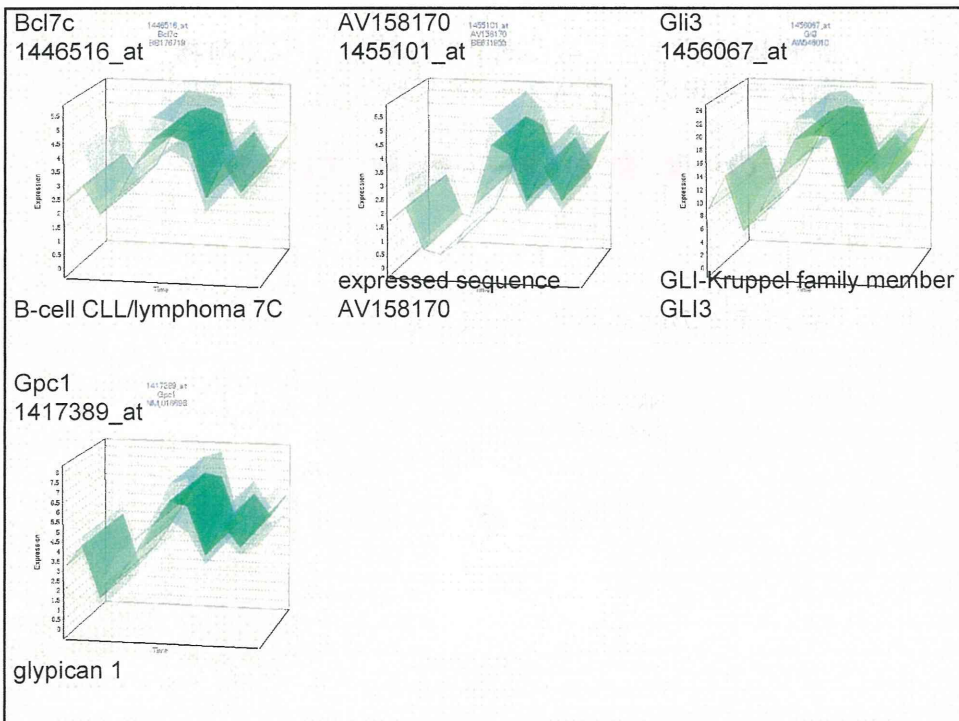
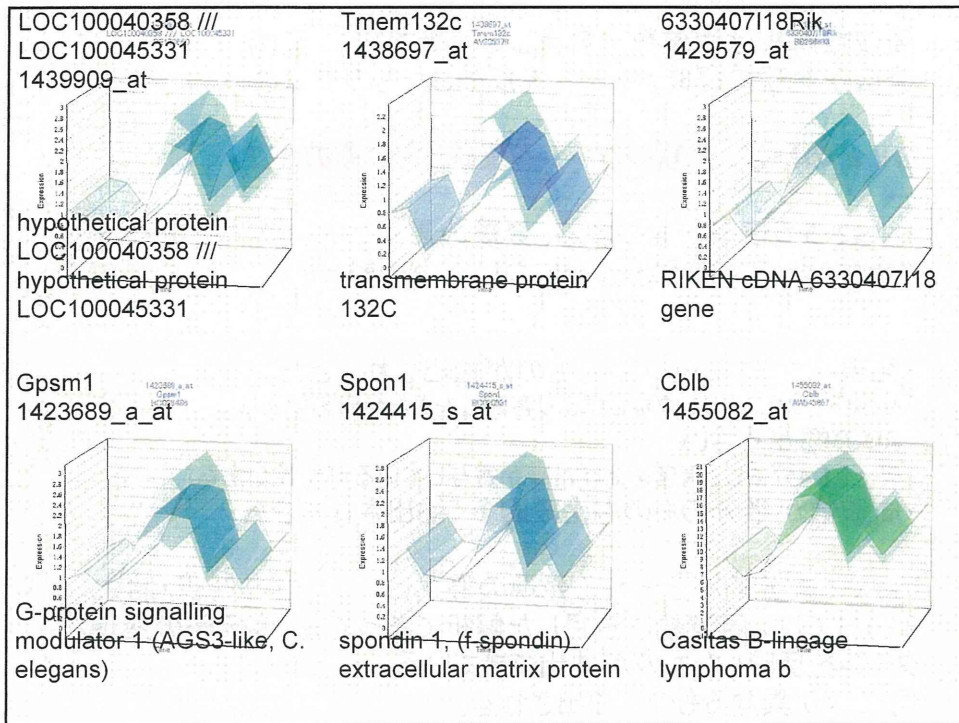
→

- Nkx2-5遺伝子が含まれる、発現の最大ピークが胎生8.25日である280 psについての検討
- 各時点での代表例について検討









周波数解析は周波数に着目した解析であり、直接、遺伝子発現の経時パターンを反映したものではない

Q1: どの程度、発現パターンを反映しているのか？

→

- ・Nkx2-5遺伝子が含まれる、発現の最大ピークが胎生8.25日である280 psfについての検討
- ・各時点での代表例について検討

- ・発現パターンが似ているものが抽出されてきた
(目視下。クラスタリングなどにより客観性をもたせる必要あり)
- ・SDの幅が小さい
(データを4分割し、精度を上げた事が寄与している可能性がある)
- ・周波数が半分のものも比較的多く抽出されてきた

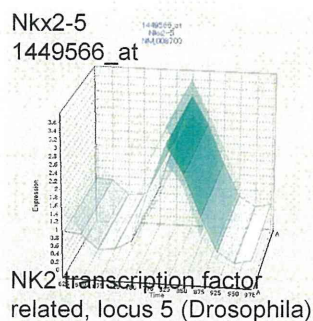
Nkx2-5遺伝子が含まれる発現の最大ピークを含む時点・胎生8.25日では、周波数に着目した解析であっても、遺伝子の発現パターンを反映したものが抽出できている。
周波数が異なるものも抽出される。

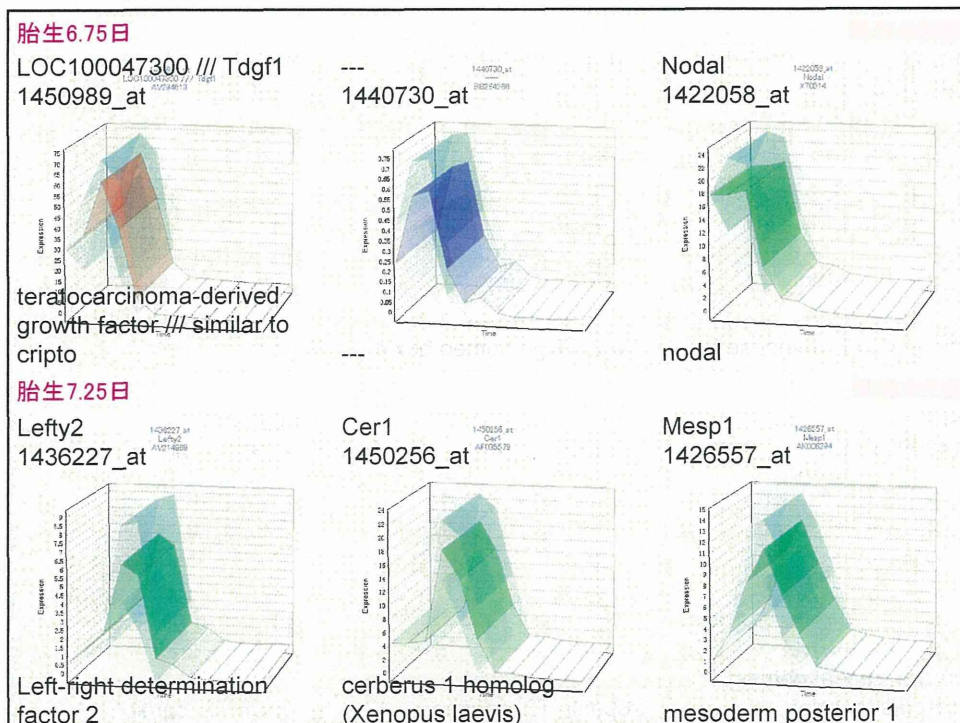
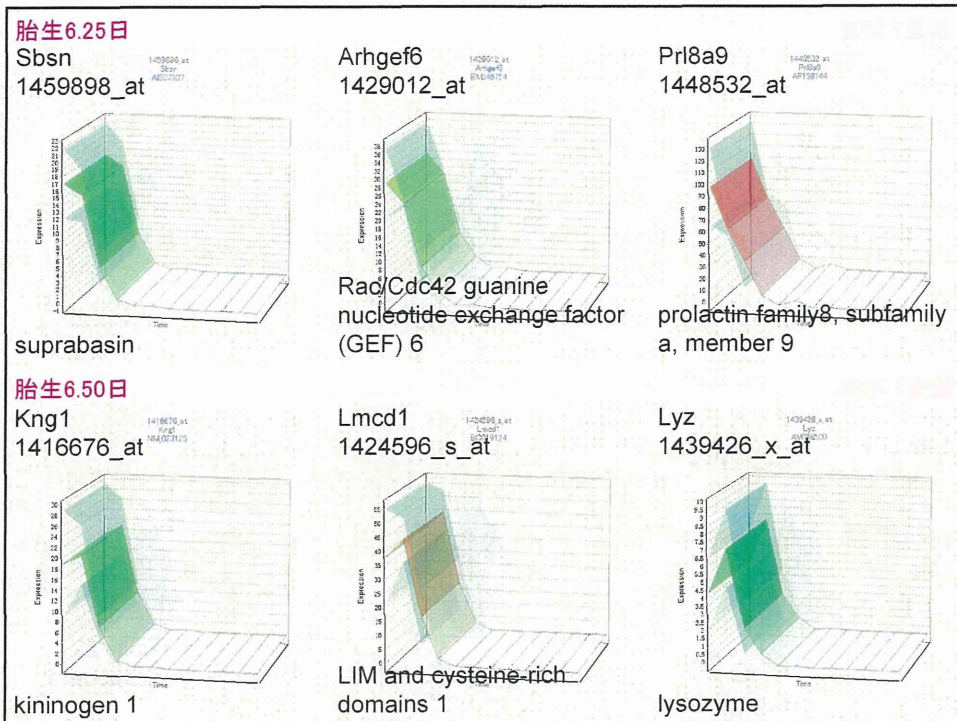
周波数解析は周波数に着目した解析であり、直接、遺伝子発現の経時パターンを反映したものではない

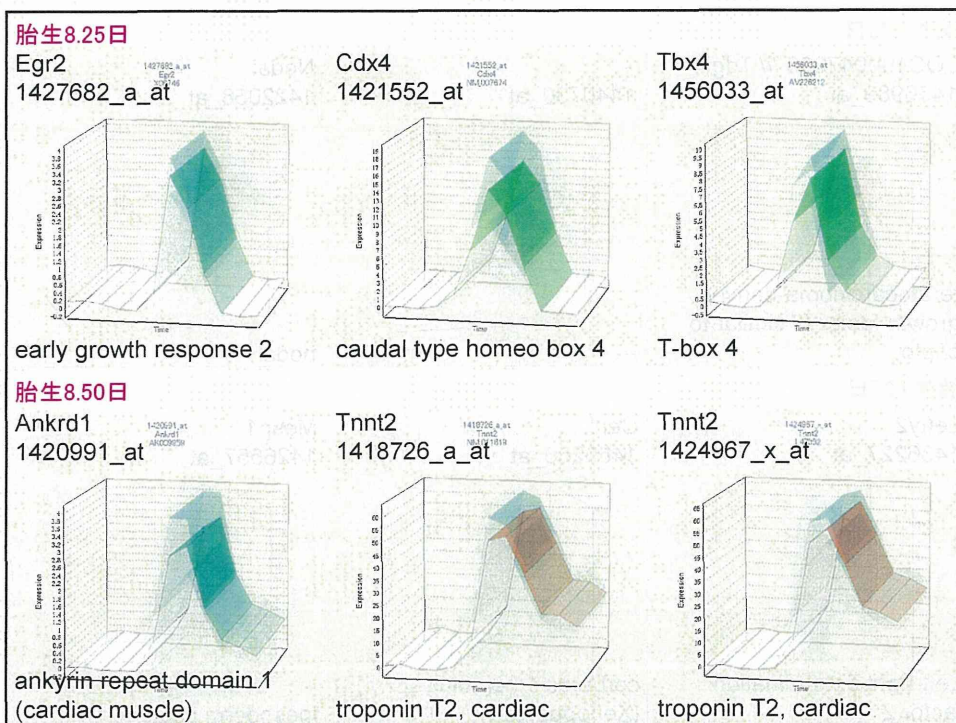
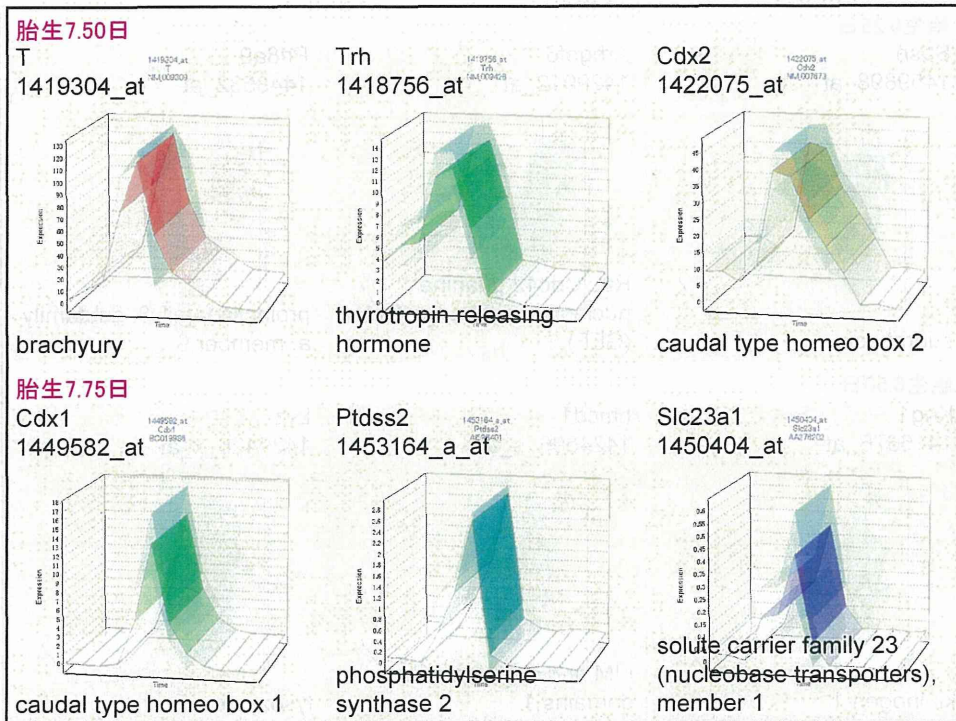
Q1: どの程度、発現パターンを反映しているのか？

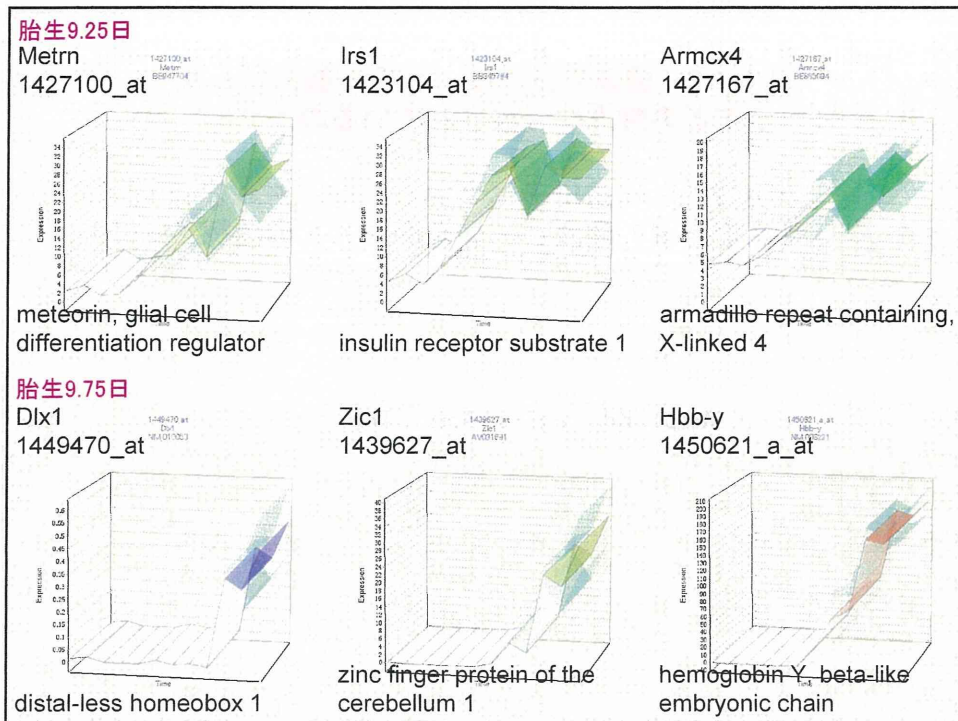
→

- ・Nkx2-5遺伝子が含まれる、発現の最大ピークが胎生8.25日である280 psfについての検討
- ・各時点での代表例について検討









周波数解析は周波数に着目した解析であり、直接、遺伝子発現の経時パターンを反映したものではない

Q1: どの程度、発現パターンを反映しているのか？

→

- ・Nkx2-5遺伝子が含まれる、発現の最大ピークが胎生8.25日である280 psについての検討
- ・各時点での代表例について検討

・Nkx2-5遺伝子が含まれる発現の最大ピークを含む時点・胎生8.25日に近い時点では発現パターンが比較的似ているものが抽出されてきたが、両端に近い時点(胎生6.25、6.50、9.50、9.75日)の場合、発現パターンがかなり異なるものが抽出されてきた。

目視による確認にかわる、自動認識手法*の導入を検討する。

*: 発現パターンを標準化しピーク特性を評価する

Q2: Nkx2-5関連遺伝子は、発現の最大ピークを示すどの時点に分類されているのか？

→Nkx2-5関連遺伝子について、それぞれの時点で発現の最大ピークを示すのかの確認

Mesp1、Kdr、Nkx2-5、Actc1、Tnnt2

Nkx2-5遺伝子を基としフーリエ変換して得られた波長分布についてピアソン相関解析の結果得られた（相関係数>0.91）

3,319 psについて、

RSortで経時的に並びかえ(3.5日間のフォーマットで)、発現の最大ピークを示す各時点毎に抽出（目視の検討はなし）

Nkx2.5遺伝子：胎生8.25日に発現ピークを持つ遺伝子として抽出されてきた

結果 ← 3,319 ps

| | | |
|----------|----------|-------------|
| 胎生6.25日： | 292 ps | |
| 胎生6.50日： | 1,150 ps | |
| 胎生6.75日： | 8 ps | |
| 胎生7.25日： | 3 ps | Mesp1 |
| 胎生7.50日： | 16 ps | |
| 胎生7.75日： | 1,364 ps | |
| 胎生8.25日： | 280 ps | Nkx2-5、Kdr1 |
| 胎生8.50日： | 139 ps | Tnnt2 |
| 胎生8.75日： | 0 ps | |
| 胎生9.25日： | 14 ps | |
| 胎生9.50日： | 3 ps | |
| 胎生9.75日： | 50 ps | |

Setting:

Explore Direction:

Peak>>>Valley

Posi,Intermed,Nega:全てcheck

Q2: Nkx2-5関連遺伝子は、発現の最大ピークを示すどの時点に分類されているのか？

→Nkx2-5関連遺伝子について、それぞれどの時点で発現の最大ピークを示すのかの確認

Mesp1、Kdr、Nkx2-5、Actc1、Tnnt2

発現の最大ピークだけを指標に分類し、解析していく手法では不十分

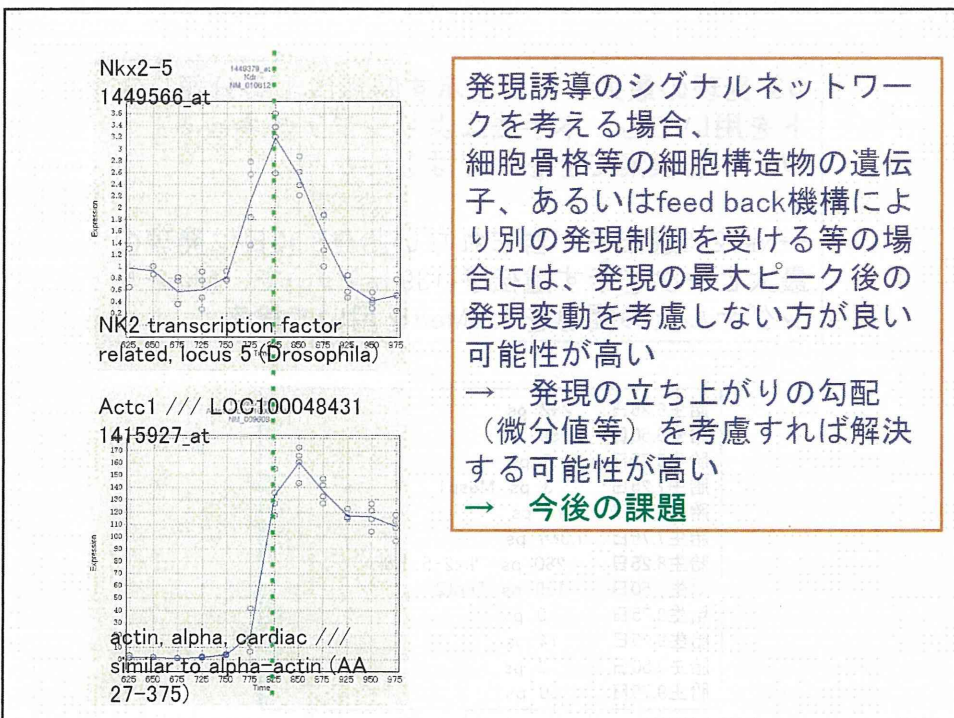
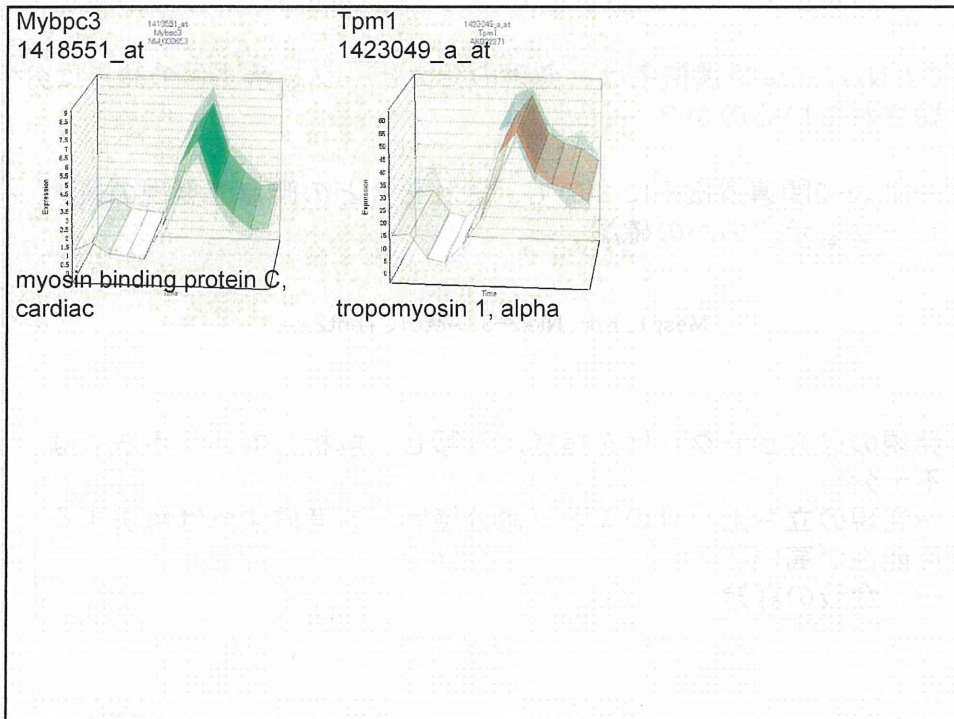
→発現の立ち上がりの勾配（微分値等）を考慮すれば解決する可能性が高い

→ 今後の課題

Q3: 発現の最大ピークを示す時点による分類リストを用いても、本手法によりシグナルネットワークを検索できるのではないか？

→Nkx2-5遺伝子が含まれない胎生8.50日に発現の最大ピークを示す遺伝子(139 ps)について、Nkx2-5シグナルとの関連をPubMedを用いて検索

| | | |
|----------|----------|-------------|
| 胎生6.25日: | 292 ps | |
| 胎生6.50日: | 1,150 ps | |
| 胎生6.75日: | 8 ps | |
| 胎生7.25日: | 3 ps | Mesp1 |
| 胎生7.50日: | 16 ps | |
| 胎生7.75日: | 1,364 ps | |
| 胎生8.25日: | 280 ps | Nkx2-5、Kdr1 |
| 胎生8.50日: | 139 ps | Tnnt2 |
| 胎生8.75日: | 0 ps | |
| 胎生9.25日: | 14 ps | |
| 胎生9.50日: | 3 ps | |
| 胎生9.75日: | 50 ps | |



周波数解析は周波数に着目した解析であり、直接、遺伝子発現の経時パターンを反映したものではない

Q1: どの程度、発現パターンを反映しているのか？

・Nkx2-5遺伝子が含まれる発現の最大ピークを含む時点・胎生8.25日に近い時点では発現パターンが比較的似ているものが抽出されてきたが、両端に近い時点(胎生6.25、6.50、9.50、9.75日)の場合、発現パターンがかなり異なるものが抽出されてきた。

目視による確認にかわる、自動認識手法*の導入を検討する。*: 発現パターンを標準化しピーク特性を評価する

Q2: Nkx2-5関連遺伝子は、発現の最大ピークを示すどの時点に分類されているのか？

発現の最大ピークだけを指標に分類し、解析していく手法では不十分
→発現の立ち上がりの勾配(微分値等)を考慮すれば解決する可能性が高い
→ 今後の課題

Q3: 発現の最大ピークを示す時点による分類リストを用いても、本手法によりシグナルネットワークを検索できるのではないか？

少なくとも2つの心筋関連遺伝子を見いだす事ができた。
抽出精度の向上のため、以下の方法を組み合わせる
・wholemound ISHでの
→胚でのISHのデータベースが存在しており、この充実に期待
・in silicoのプロモーター解析

目的

胎児発生過程におけるマスター遺伝子を基軸とした
遺伝子発現ネットワークの網羅的解析

- ・胎児発生過程 = 自律的な遺伝子発現が多い
連続性のある遺伝子発現変動現象として長時間にわたるものであり、
遺伝子発現シグナルの流れを解析するのに適している
→ 活発な細胞増殖や多様な細胞分化が同時並行で起こるため、
要素分析が困難
- 先行研究にて利用した周波数解析による局所ネットワーク描出技術の
適用拡大を検討

まとめ

- ・既知のシグナルネットワークをモデルに、本解析手法の性能評価をしたところ、
既知関連遺伝子の多くは検出可能であった。
- ・経時的に離れた関連遺伝子の場合や、遺伝子の機能によっては、発現パターンが
かなり異なる傾向があるため、本解析手法そのままでは効率の良い抽出ができな
かった。
- この方策として、発現の立ち上がりの勾配に着目すれば解決する可能性を見いだ
した。来年度、そのためのアルゴリズムの開発(微分値の利用)を予定している。

厚生労働科学研究補助金（化学物質リスク研究事業）

分担研究報告書

化学物質の有害性評価手法の迅速化、高度化に関する研究
ー網羅的定量的大規模トキシコゲノミクスデータベースの維持・拡充と
毒性予測評価システムの実用化の為のインフォマティクス技術開発ー

分担研究課題：「Percellome 3次元データ等の為の専用解析ソフトウェアの開発研究」

研究分担者 相崎 健一

国立医薬品食品衛生研究所・安全性生物試験研究センター・毒性部・第一室長

研究要旨

先行研究に於いて遺伝子発現データの絶対量化技術である Percellome 手法を実用化し、既に100余の化学物質による遺伝子発現変動情報を収録して現在もさらに拡張しつつある網羅的遺伝子発現情報データベース（Percellome データベース）を構築している。この高精度大規模データベースの有効活用には、時間×暴露用量×遺伝子発現量からなる遺伝子発現変動情報を生物学者によるデータ把握が容易な3次元波動面（Surface グラフ）に変換して解析する直感的で効率のよい手法（“Millefeuille”（ミルフィーユ、MF）データ処理）を採用している。本分担研究では、Percellome 手法や Surface グラフの特徴を活かし、遺伝子発現反応の網羅的解析を効率良く行うためのプログラムの開発・改良を行い、トキシコゲノミクス技術を利用した毒性予測・評価システムの実用化を推進する。

平成24年度は Percellome 法の適用を介して、異種由来の遺伝子発現データであってもシームレスな統合データベースとして解析するための基本アルゴリズムを生成した。これを用いて、Percellome 法を適用して実施された医薬基盤研・トキシコゲノミクスプロジェクト（TGP）のラットデータを、Percellome 変換（遺伝子発現値の絶対量化とデータ精度確認）した上で、マウスデータからなる Percellome データベースに結合し、統合トキシコゲノミクスデータベースの構築を進めた。また Percellome データベース及び解析プログラムの公開に伴い、Percellome 法を適用していないマイクロアレイデータであっても Percellome データベースとの直接比較を可能とする絶対量推定ソフトウェア SnCalc.exe を開発した。これらの研究開発の結果、より広範囲の化学物質-遺伝子発現情報が利用可能となり、データベースや解析ソフトウェアの有用性・汎用性を高めた。併せて Percellome データベースシステムの WebAPI を開発し、ライフサイエンス研究用ソフトウェアの国際共通プラットフォーム GARUDA プロジェクト等、外部の学術・行政システムからのオンラインアクセスを可能にする本格的な一般公開の準備を整えた。

A. 研究目的

本研究班が化学物質による生体影響の分子メカニズムに依拠した毒性評価手法の実用化の為にインフォマティクス開発を目指す中であって特に、100余の化学物質による遺伝子発現変動情報を絶対量化して収録している高精度大規模遺伝子発現情報データベース (Percellome データベース) を活用し、遺伝子発現反応の網羅的解析を効率良く行うためのプログラムの開発・改良を行って、トキシコゲノミクス技術を利用した毒性予測・評価システムの実用化を推進する。

B. 研究方法

独自の遺伝子発現データの絶対量化技術である Percellome 手法 (*BMC Genomics*, 7, 64, 2006 特許 4415079 号) をベースに、これに最適化したアルゴリズムや”Millefeuille” (ミルフィーユ) データ処理システムおよび解析用ソフトウェア群を開発している。

開発・改良したアルゴリズムは、研究者本人がソフトウェアを作成し、実際に本研究計画で得られた生データをサンプルに用いて、ワークステーション上で性能検証を行う。これは本研究におけるプログラム開発はその根幹となる解析アルゴリズムの開発・改良と同時に行う必要があり、職業プログラムに開発委託する際に必要となる仕様の事前決定が出来ないためである。開発言語には、研究的なプログラミング (研究成果のフィードバックを重視した柔軟な仕様設計や機能拡張の実施) に適した、つまり構造化・オブジェクト化といったプログラム言語仕様や生成したプログラムの実行速度だけでなく、プログラム実行ファイル

の生成 (コンパイル) 速度も速い Delphi 言語 (=object pascal 言語, 開発環境としては Borland Delphi2012 もしくは Delphi XE2) を使用した。また一般的なグラフ描画には TeeChart VCL コンポーネント (Steema 社) を、1 億件以下の小～中規模のデータベース処理には組み込み型リレーショナルデータベースコンポーネントである DBISAM (ElevateSoftware 社) を、WebAPI の構築には DataSnap (Embarcadero Technologies 社) を利用して、プログラム開発の効率化を図った。

大規模計算が必要な場合は、高速データベースエンジンである Teradata データベース (日本 Teradata 社) を装備する研究計算用サーバーシステム (MF サーバーシステム) にソフトウェアを移植して、解析処理を行った。

計算結果の検証は必要に応じて Excel (Microsoft 社) や R 言語 (オープンソース R Development Core Team) で実施し、浮動小数点誤差以上の乖離がないことを確認した。

C. 研究結果

平成 24 年度はデータベースの拡充と公開に向けた取り組みを重点的に実施した。

まずトキシコゲノミクスによる将来の毒性予測/評価システムに於いて基盤となる Percellome データベースを拡充するために、Percellome 法を適用して実施された医薬基盤研・トキシコゲノミクスプロジェクト (TGP) のラットデータの統合を検討した。マウスとラットの組み合わせの様に異種由来のデータを一括して扱う場合、相同遺伝子 (オルソログ遺伝子) 対を利用して結合