

図2 時間遅延を考慮したベイズ推定による推定結果

図2の推定結果を、次の方針により検証を行った。

- (1) PPAR γ , SP7(Osterix)のsiRNAによるノックダウンと発現ベクターによる強制発現の実験結果との整合性が取れている。
- (2) 文献データベースを検索しても、脂肪細胞分化や骨芽細胞分化に関与しているという報告がされていない。
- (3) PPAR γ のプロモータレイ実験の結果と整合性が取れている。

この検証により、図2の推定結果で得られた22個の遺伝子のうち、8個が実験結果との整合性が取れていないことがわかり、その8個を除いた14個の遺伝子のうち8個が既に脂肪細胞分化または骨芽細胞分化に関与しているという報告が見つかった。これらを除くと結果的に、脂肪細胞または骨芽細胞の分化に関与した新規の遺伝子が6個見つかった。

7. クロストーク遺伝子による両細胞分化ネットワークの統合と特性解析 (阪大)

6. では、脂肪細胞分化と骨芽細胞分化の転写制御ネットワークをそれぞれ独立に推定しているが、両者の推定結果で共通に現れる遺伝子はクロストーク遺伝子の候補とみなすことができる。実際に図2の推定結果のうち実験と整合性が取れた新規の遺伝子6個のうち3個の遺伝子で、両細胞分化双方の既知ネットワーク中の遺伝子からの転写制御関係が示唆された。

今後は、まず、2つの分化系列における時系列発現データ間に相関がない等の条件で絞り込みを行い、得られたクロストーク候補遺伝子をもとに、2つの細胞分化ネットワークを統合することで、フィードフォワードなど転写制御ネットワークの構造的な特性を解明する(図3参照)。さらに、統合したネットワークを基に、各遺伝子の時系列発現データを照合することで、両細胞分化の転写制御関係を微分方程式等で表現した動的ネットワークモデルを構築する予定である。

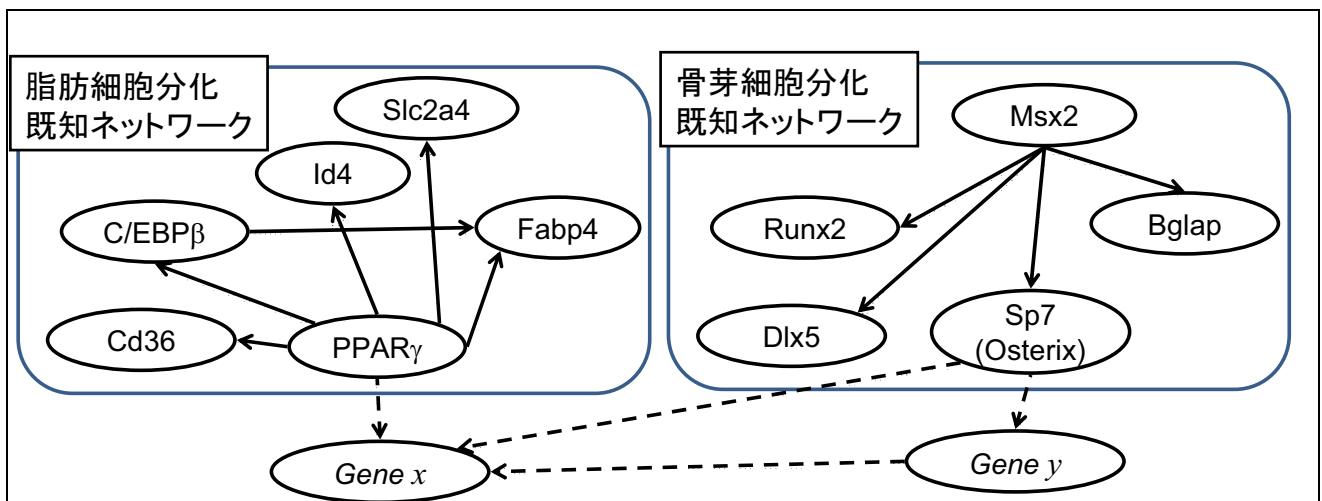


図3 クロストーク遺伝子による両細胞分化ネットワークの統合

8. 分子生物学的実験による解析結果の検証 (埼玉医大)

6. で得られた新規の転写因子候補である6個の遺伝子と既知の細胞分化ネットワーク中の遺伝子との間の転写制御関係を、ルシフェラーゼアッセイやsiRNAによるノックダウン実験等により検証しつつある。

【GNPとの関連・GNPへの貢献】

a) 転写・制御の解明に関連する研究成果

本研究でこれまでに得られた研究成果は次の通りである。

- (1) 脂肪細胞分化または骨芽細胞分化の既知の転写制御ネットワークに含まれる転写因子から制御を受けている細胞分化関連の転写因子が合わせて22個得られ、それらのうちで6個はこれまでに文献で報告されていない新規の転写因子であった。
- (2) (1)の6個の転写因子のうち3個は、脂肪細胞分化と骨芽細胞分化の両方の既知転写因子から何らかの制御関係が示唆されており、クロストーク遺伝子の候補と考えられる。
- (3) 複数の細胞分化系列間の転写制御のクロストークと、クロストークを考慮して両細胞分化にまたがる転写制御ネットワークの推定と動的数理モデルを構築するための解析方法を確立した。

b) 縦軸研究 (または横軸研究) への貢献に繋がる研究成果

縦軸研究である脂肪細胞・骨芽細胞分化過程の解析研究の実施機関と連携して、情報科学的なネットワーク解析と、解析に必要な時系列発現データの産生、解析結果の実験的な検証を、同一プロジェクト内で同時並行的に行うことで、従来の縦軸研究のアプローチを補完し、これまでにない新規の細胞分化関連因子の解明につながる研究成果が見込まれる。

【当該研究に関して活用した研究データ・リソース】

縦軸研究である脂肪細胞・骨芽細胞分化過程の解析研究から取得されたマイクロアレイの時系列発現データに加えて、タイムポイントを非常に細かく設定した時系列発現データを新たに取得した。また、遺伝子発現量の定量的な比較等のためCAGEデータも統合して利用している。さらに、転写因子間の相互作用の検討のため、PPIデータを利用する予定である。

【主な研究論文とその概要】

「4. 探索対象とする転写因子の選定」について関連論文発表1件、国際会議発表1件がある。

1. 論文: S. Maekawa, A. Matsumoto, Y. Takenaka, H. Matsuda, Tissue-specific functions based on information content of gene ontology using cap analysis gene expression, *Medical and Biological Engineering and Computing*, 45(11):1029-1036, 2007.

概要:

CAGE(Cap Analysis Gene Expression)法により計測した加算可能な遺伝子発現プロファイルを用いて、遺伝子発現量全体と各組織における発現量の差分情報量を求めることで、組織特異的な遺伝子機能を探索する方法を提案した。遺伝子機能を表現するにはGene ontologyの機能体系を用い、差分情報量を求めることで組織間での相対的な特徴量を定義している。

2. 国際会議: Y. Takenaka, A. Matsumoto, H. Matsuda, GO based tissue specific functions of mouse using countable gene expression profiles, *Genome Informatics* 19:154-165, 2007.

概要:

上記論文の手法をさらに発展させて、多数の組織で特徴的な遺伝子機能の発現量を表現できるようにした。これにより、肝臓、肺臓、心臓、筋肉などの組織に特徴的な機能の相関や、特定の機能に特徴的な組織を探索できることが示された。

今後、クロストーク遺伝子の探索や、脂肪細胞・骨芽細胞分化ネットワークの統合・特性解析について、結果が得られ次第、論文投稿を進める予定である。

【特許出願数の実績】

本研究課題は昨年10月から始まったばかりであり、まだ特許出願の実績はない。

新規でかつ細胞分化に重要な役割を果たす転写因子が検証され次第、特許化を推進する予定である。

ゲノムネットワークプロジェクトの
課題の報告

プログラム名	(1) ゲノム機能情報の解析 (2) ヒトゲノムネットワークプラットフォームの構築 (3) 次世代ゲノム解析技術の開発 (4) 個別生命機能の解析 (5) <u>動的ネットワーク解析技術開発</u>			
課題の分類	(1) 中核機関 (2) 指定課題 (3) <u>公募課題</u>			
課題名 (和英併記)	動的ネットワーク抽出のためのイン・シリコパイプラインの構築 (Development of <i>in Silico</i> Pipeline for Extracting Dynamical Networks)			
代表研究者名 (所属機関・職名)	宮野 悟 (国立大学法人東京大学・教授)			
年度別研究費 (千円)	19年度	20年度	合計	
	41,000	65,000	106,000	
	研究者名	所属機関・部局・職	専門分野	
研究組織	代表機関	☆ 宮野 悟	東京大学・医科学研究所・教授	バイオインフォマティクス
		(分担研究者)		
	分担機関	☆ (代表研究者)		
		(分担研究者)		
【研究目的】				
<p>マクロファージの分化・活性化ネットワークに関する遺伝子発現情報等の大規模・総合的データから動的転写制御ネットワークを抽出し、利用するための計算戦略をイン・シリコパイプラインとして構築することを目的とする。</p>				
【課題の概要】				
<p>状態空間モデルと次元圧縮技術により転写モジュールネットワーク構造を抽出し、さらに、複数ソース情報を利用した数理統計手法により転写制御ネットワーク構造を精緻化する方式を構築する。次に、データ同化技術により時系列データを利用してネットワーク構造へダイナミクスを導入する技術、及び文献キュレーション情報を用いて、動的な転写制御ネットワークを抽出するパイプラインを構築する。同時に、この動的転写制御ネットワークを開発・活用・発展させるためのソフトウェアプラットフォームを開発する。</p>				

【平成20年8月末までの研究成果】

1. イン・シリコパイプライン設計の基本方針を以下の5項目として決定した。また、図1に示しているように、イン・シリコパイプライン建築のためのタスク群を整理し、各タスクについて開発を進めた。

- ・情報を CSML (Cell System Markup Language, <http://www.csml.org/>) 形式で統一
- ・ネットワーク推定は、既存開発方式を CSML 形式に対応するように開発。Javaでコード化
- ・データ同化は、既存及び現在開発中の方式を CSML 形式に対応するようにコードを開発
スパコン利用
- ・可視化解析パイプラインツール CSMLPipeline を開発
- ・可視化・シミュレーションツールとして、既存ソフトウェアCell Illustrator Online (参考：<http://cionline.hgc.jp/>) を利用

2. CSMLデータベースとそのAPIの開発

マクロファージデータベーススキーマ及びデータベースにアクセスするためのAPIライブラリ (Macrophage CSMLDB) の開発を行った。イン・シリコパイプラインのためのデータベース管理システムの基礎的整備を行った。マクロファージデータの詳細なサーベイによって明らかとなったデータ群をイン・シリコパイプラインで効率的に利用できるよう、データベーススキーマ (Macrophage DB Schema

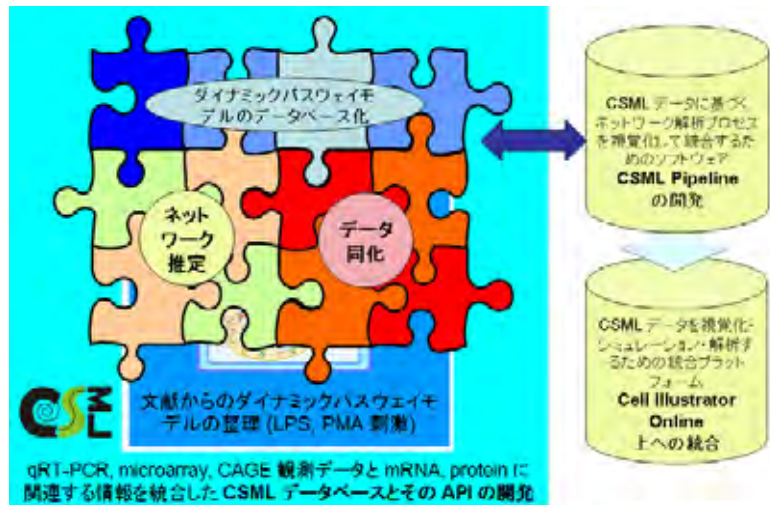


図1：パイプライン建築のタスク群

Version 2 (MDBS Ver 2.0)を開発した。2008年8月31日の時点でのそのスキーマを図2に示している。そして、MΦ、qRT-PCR、microarray、CAGE 観測データとmRNA、protein、TFなどの情報を統合したデータベースとそれにアクセスして利用するためのAPIを開発した(図3)。対象データ及びデータ数は、表1及び表2のようになっており、イン・シリコパイプラインに関連する以下の約1000万エントリのデータを含むヒトマクロファージデータベースを作成した。そして、CSML データベースと API を使って解析プログラムを開発できる環境を整備した(利用実績：図4)。

データタイプ	遺伝子発現	遺伝子発現 転写開始点	
使用した技術や試料	qRT-PCR	Illumina microarray (custom design)	CAGE
データ内容	転写開始点遺伝子 (mouse 12,159 遺伝子、human は 2,915 遺伝子) のプライマーを使用し、リアルタイム RT-PCR まで測定を行った。qRT-PCR 及びヒト、マウスのマクロファージから得られた細胞の遺伝子発現データを集計した。CAGE タグの配列、CAGE タグを付与したサンプルはそれぞれに計測した遺伝子のコピー数の相対平均を用いて normalize した。	Human は Illumina の Sentrix Human-6 Expression BeadChip を使用し、mouse は Mouse-6 Expression BeadChip を使用し、THP-1 及びヒト、マウスのマクロファージから得られた細胞のゲノムを基に、CAGE タグの配列、CAGE タグを付与したサンプルの標的 (tissue, devstage, cellline など) の遺伝子発現のプロファイルを作成した。	mRNA の 5' 端から 20 ベースを切り出した CAGE タグのデータ、データ重複したデータベース (CAGE base データベース) の各データの CAGE タグを基に、CAGE タグの配列、CAGE タグを付与したサンプルの標的 (tissue, devstage, cellline など) の遺伝子発現のプロファイルを作成した。
サンプルデータまたは 実験方法の説明	http://esomene.tebionk.ac.jp/public/download/cage-express_qrt-pcr.html	http://www.illumina.com/pdf/Human6/ExpressionBeadChipDataSheet.pdf	http://esomene.tebionk.ac.jp/public/download/cage-database.html
データの数	2,315 遺伝子のプライマーについて THP-1 の PMA 刺激後 9 time point のデータ及び LPS 刺激後の 11 time points のデータ	47,293 プロローブについて THP-1 の PMA 刺激後 9 time point のデータ及び LPS 刺激後の 12 time point のデータ	主要組織や THP-1 などの 119 サンプルについての CAGE タグ、59,193,600 タグのデータ
THP-1 PMA time points	control	○	○
	1hr	○	○
	2hr	○	○
	4hr	○	○
	6hr	○	○
	12hr	○	○
	24hr	○	○
	48hr	○	○
	72hr	○	○
96hr	○	○	

表1：マクロファージ (THP-1) の計測データ

Table Name	Columns
Experiment	98
ExperimentCategory	8
ExperimentGroup	9,016,478
Expression	9,314,723
Gene	68,530
GroupName	6
Probe	10,014,873
ProbeMatchedSequenceInfo	14,611,983
Promoter	101,438

表2：MDBS Ver. 2.0 (Beta) のデータエントリ数

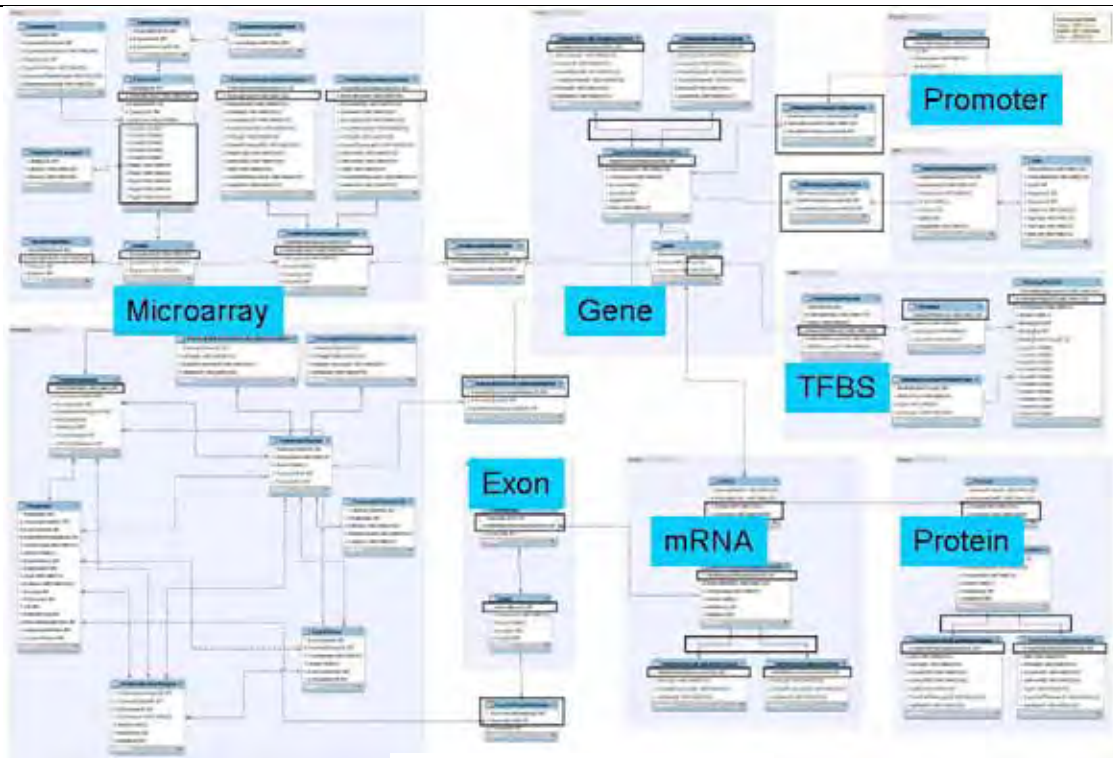
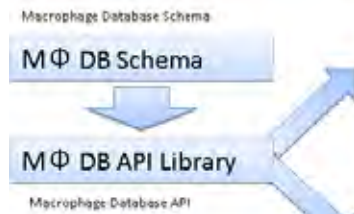


図 2 : MDB Schema Ver. 2



図 3 : MDB API Ver. 2 (Beta)の抜粋

図4: MDB Schema とその API を利用したプログラム開発



利用実績: ゲノムの配列位置情報, TSS の情報, qRT-PCR, Microarray, CAGE のデータを DB から取得して, 一連の統計解析によりデータの質的・量的比較に用いた

利用実績: qRT-PCR, Microarray, CAGE のデータを DB から取得して, データ解析ソフトウェア群 (CSML Pipeline) にプラグインするプログラムでの利用

3. 文献からのダイナミックパスウェイモデルの整理 (LPS, PMA 刺激)

本業務では、設計として、イン・シリコパイプラインの統合プラットフォームとして Cell Illustrator Online (本業務担当者等が1999年以来開発してきたパスウェイのモデリングとシミュレーションソフトウェア。CSML DBMS付、Javaで開発されている。) を利用している。これまで、マクロファージのLPS及びPMA刺激に関連する文献を調べあげ、本業務担当者等が本業務の開始に先立って開発した文献キュレーションプロトコルに基づき、複数の文献からモデルを作成するのではなく、1文献1モデルの指針に基づき、シミュレーション可能なCSML形式の転写制御・シグナル伝達パスウェイの動的モデルを開発した。

	Dynamic Pathway Model	Static Pathway Model
2008/3/31	38	45
2008/8/31	101	110

表2: MΦダイナミックパスウェイモデルの文献からの抽出と整理状況

そして、これらを統合プラットフォーム Cell Illustrator Online (CIO) for Macrophage (Macrophage CIO) のサーバ上のデータベース上で利用・管理できるようにした。2008年8月31日現在の状況を表2に示している。図4はCSML形式で記述した動的モデルをCIOで表示した一例で、CSML ReporterによりHTML形式にも変換できるようになっている (図5)。



図4: CSML形式で表現した動的モデルを CIO で表示したもの

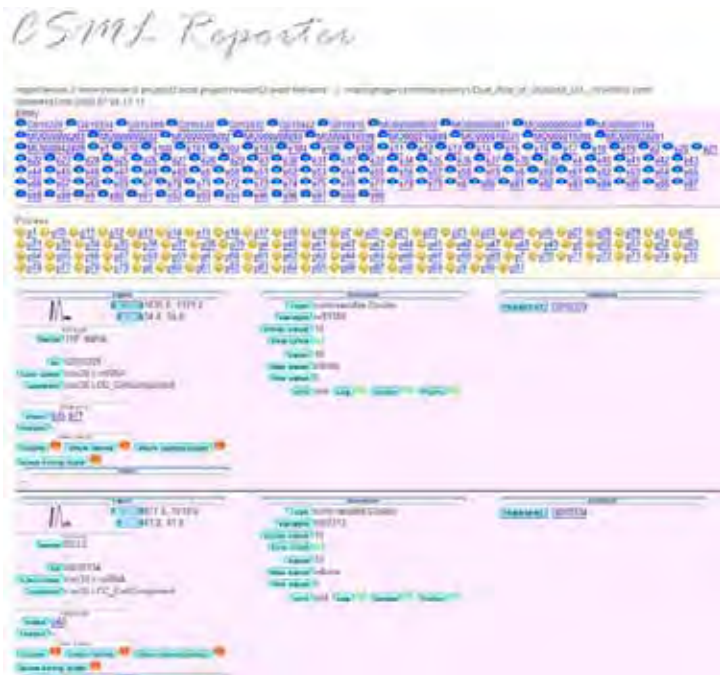


図5: CSML Reporter により HTML に変換

4. ネットワーク推定とデータ同化

アルゴリズム群のJava実装とCSML形式の視覚化技術の適用・開発が以下のように進行した。

4.1. 状態空間モデルによるネットワーク推定

状態空間モデルと次元圧縮による転写モジュールネットワーク構造の抽出プログラムに関する開発を完了した。マイクロアレイ等により計測された、繰り返し計測時系列データから、転写モジュールネットワークおよび遺伝子間ネットワークを抽出することが可能になった。このためのプログラムをJavaで開発し、以上のプログラムに加え、遺伝子発現時系列観測データから推定された制御マトリクス群をテキストで出力するアプリケーションを開発した。後述のCSMLPipelineへの導入まで終了した（図6参照）。プログラムの機能等について、以下に説明する。

通常の時系列解析では、オブジェクト間の制御関係を推定するためのモデルとして、まずオブジェクト対オブジェクトの関係を直接的に表現する、一次のベクトル自己回帰モデルの適用を考えることが普通である。しかしながら、生命科学研究で通常得られる時系列遺伝子発現データには、考慮する遺伝子の数（数百から数千）に比べて、計測される時点数（10点前後）が少ない（超高次元短時系列）という特殊性があるため、そのまま自己回帰モデルに含まれるパラメータを推定することは技術的・理論的に困難であることが判明している。状態

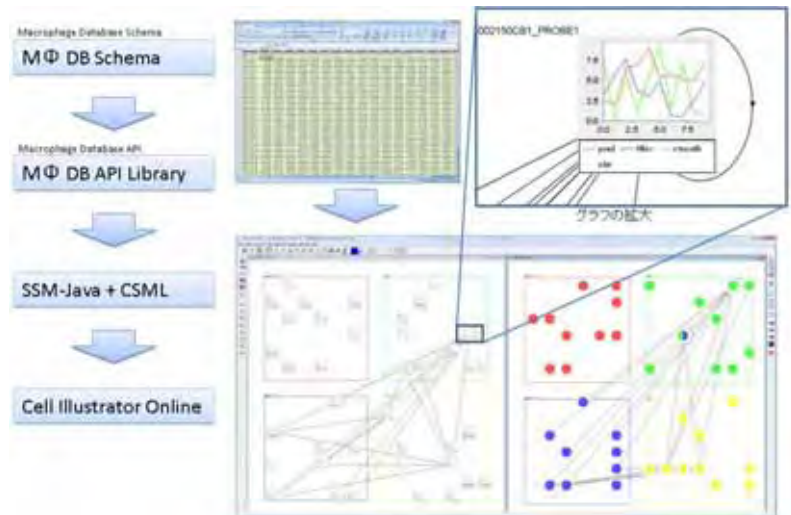


図6：状態空間モデルによるネットワーク推定のパイプラインが開通

空間モデル(State Space Model)とは、統計的時系列モデルの一種であるが、本業務では、この状態空間モデルにおいて、多くの遺伝子の発現状態を、少数の潜在変数の発現状態へ縮約（次元圧縮）することにより、パラメータの推定の困難さを克服する一連の技術が本業務参加者により考案されており、それをプログラム化してイン・シリコパイプラインに導入した。その潜在変数は、比較的少数の因子から制御を受け共発現する遺伝子を代表するメタ遺伝子として解釈が可能であり、これを転写モジュールと呼んでいる。その機能として、次の2つが特徴的である。

- (1) 推定されたパラメータは、観測された時系列遺伝子発現を生成した動的システムの構造に関する情報を有しており、それらを利用して、転写モジュール間制御ネットワークの推定および、各転写モジュールに所属する遺伝子群の推定が可能となっている。
- (2) パラメータの変換を通じて、パラメータ数を節約した一次のベクトル自己回帰モデルを得ることができ、通常のベクトル自己回帰モデルでは直接の推定が困難であった、遺伝子間ネットワークの推定が可能となっている。

(利用技術)

- Hirose, O., Yoshida, R., Imoto, S., Yamaguchi, R., Higuchi, T., Charnock-Jones, D.S., Print, C., Miyano, S. Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models. *Bioinformatics*. 24(7): 932-942, 2008.
- Yamaguchi, R., Yoshida, R., Imoto, S., Higuchi, T., Miyano, S. Finding module-based gene networks with state-space models? Mining high-dimensional and short time-course gene expression data. *IEEE Signal Processing Magazine*, 24(1): 37-46, 2007. 他.

4.2. ダイナミックベイジアンネットワークによるネットワーク推定

ダイナミックベイジアンネットワークに基づいた複数ソースデータの利用による転写制御ネットワーク構造の精緻化プログラムの整備を進めた。業務担当者等は、本業務の開始に先立って、遺伝子制御の親子関係をB-splineノンパラメトリック回帰により非線形モデル化したベイジアンネットワークによる転写制御ネットワーク構造推定プログラムをC言語で開発している。これは、C言語により実装され、ソースコードの総行数はおよそ5万行に達するものである。これまでの業務で、このプログラムのダイナミックベイジアンネットワークによる推定プログラムを改良するとともに、CSMLPipeline及びCell Illustrator Online環境で利用できるようプログラム開発を進め、CSML形式による出力機能を実装した。また、現在CSML形式で記述された動的パスウェイモデルを直接、事前情報として利用できるようにプログラムの拡張している。

(利用技術)

- Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S., Miyano, S. Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks. *J. Bioinformatics and Computational Biology*. 2(1):77-98, 2004.
- Kim, S., Imoto, S., Miyano, S. Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. *Biosystems*, 75(1-3), 57-65, 2004.

4.3. データ同化によるシミュレーションモデルの高精度化

他プロジェクトで既に開発されたコード及び他プロジェクトで開発された以下の技術を利用して、次の2つの機能を実装したプログラムの開発が進んでいる。“SSM-Java+CSML”と同様に、C言語版、FORTRAN版をCSML形式対応させ、可能な限り順次Java言語に移植し、CSML Pipelineに乗せる。Nagasaki et al. (2006)の技術については基本実装が済んでいる。

(1) 文献から抽出したダイナミックパスウェイモデルを素モデルとし、計測データをモデルに同化

(2) 状態空間モデル及びダイナミックベイジアンネットワークモデルに計測データを同化

(利用技術)

- Nakamura, K., Yoshida, R., Nagasaki, M., Miyano, S., Higuchi, T. Parameter estimation of in silico biological pathways with particle filtering towards a petascale computing. *PSB*. 14, 2009 (in press).
- Yoshida, R., Nagasaki, M., Yamaguchi, R., Imoto, S., Miyano, S., Higuchi, T. Bayesian learning of biological pathways on genomic data assimilation. (submitted for publication) <http://daweb.ism.ac.jp/~yoshidar/GDA/>
- Tasaki, S., Nagasaki, M., Oyama, M., Hata, H., Ueno, K., Yoshida, R., Higuchi, T., Sugano, S., Miyano, S. Modeling and estimation of dynamic EGFR pathway by data assimilation approach using time series proteomic data. *Genome Informatics*. 17(2): 226-228, 2006.
- Nagasaki, M., Yamaguchi, R., Yoshida, R., Imoto, S., Doi, A., Tamada, Y., Matsuno, H., Miyano, S., Higuchi, T. Genomic data assimilation for estimating hybrid functional Petri net from time-course gene expression data. *Genome Informatics*. 17(1). 46-61, 2006.

5. ダイナミックパスウェイモデルのデータベース化

本業務では、イン・シリコパイプラインの統合プラットフォームとして、CSML DBMSの付いているCell Illustrator Onlineを利用している。このCell Illustrator Onlineの“Project Manager”機能を通して、CSML データベース利用を実現した。また、Cell Illustrator Onlineの高速シミュレーションモジュール、パラメータ探索モジュール、TRANSPATH DBモジュール、パスウェイレイアウト機能、個人用データベース機能、遺伝子ネットワーク解析機能などを利用して、高度なデータベース利用を実現した。図7は、Project Managerを使ってMacrophage Pathway Databaseを構築した部分と、Cell Illustrator Onlineによる表示及びシミュレーションのスナップショットである。

6. CSML Pipeline の開発及び

Cell Illustrator Online 上への統合

CSML データに基づくネットワーク解析プロセスを視覚化して統合するためのソフトウェア CSML Pipeline を開発中で、CUIは完成した。現在、GUI及びネットワーク表示のためのCSMLフィルタリングソフトウェアを開発している(図8)。これらが完成すれば、CSML データを視覚化・シミュレーション・解析するための統合プラットフォームCell Illustrator Online上へ統合される。

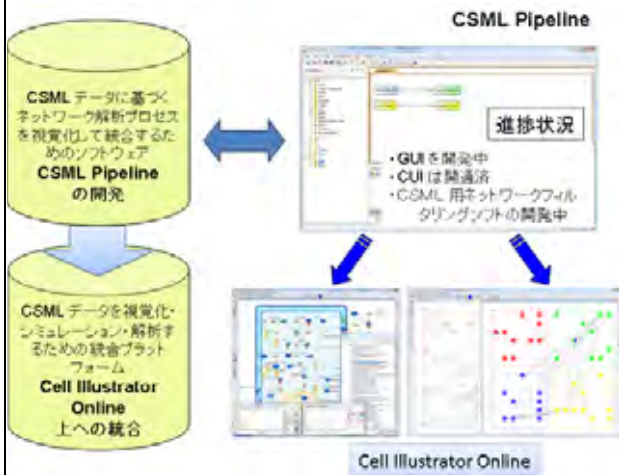
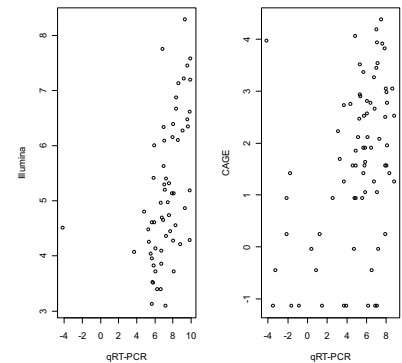


図8 : CSML Pipeline と Cell Illustrator Online への統合



図7 : 右下が Project Manager を使った Macrophage Pathway Database。青くハイライトされている動的パスウェイモデルを Cell Illustrator で表示している。下の2つのグラフはそのシミュレーションの状況。



7. データサーベイ及び時系列データの計測時点についての検討

より高精度のネットワーク推定を行うには、CAGE、Illuminaマイクロアレイ、qRT-PCRの各データの特徴を理解する必要があり、理化学研究所から提供された時系列データ(CAGEデータ、qRT-PCRによる遺伝子発現データ、Illuminaマイクロアレイによる遺伝子発現データ)の詳細なサーベイを行い、統計的解析を行い、得られた知見に基づいてイン・シリコパイプラインに導入する技術の検討を行った。具体的には、各遺伝子に対し、qRT-PCRの計測結果を標準データとし、IlluminaマイクロアレイデータとCAGEデータの精度調査を行った。CAGEデータを用いることで各遺伝子レベルではなく、転写レベルでの詳細なデータが取得できることから、

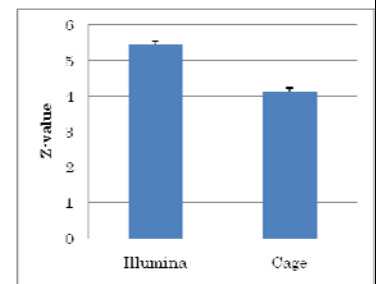


図9 : データ比較結果

こうしたレベルでの解析においてCAGEデータの優位性があるため、遺伝子レベルでの単純な比較はそのデータの品質の良し悪しの判断にはならないが、データの特徴を理解することがある程度できた。まず、qRT-PCR、マイクロアレイ、CAGEデータから外れ値をTukeyらの手法を用い除去した後、HoeffdingのD measureによる計測を行った。qRT-PCRとマイクロアレイ、qRT-PCRとCAGEの発現量の相関解析を行った。図9に、その解析例として、各発現量に対数を取った散布図、図9にqRT-PCRとマイクロアレイ、qRT-PCRとCAGEとの各Z-valueを示している(Z-valueが高いと相関が高い)。これらの解析結果を、データ生産を行っている連携チームにフィードバックした。

また、ネットワーク推定に必要とされる時点数及びその計測時点について連携して検討を行い、新たに、24時点のCAGE及びマイクロアレイデータの計測を検討することとなった。今後、そのデータに基づいてイン・シリコパイプライン解析を行い、得られた知見に基づいて、計測従事者の労働条件を考慮した、現実運用可能な時点数及び計測時点の設計法について検討する方向付けを行った。

【GNPとの関連・GNPへの貢献】

a) 転写・制御の解明に関連する研究成果

CSMLDB、Cell Illustrator Online、CSMLPipelineから構成される動的ネットワーク抽出イン・シリコパイプラインに、時系列計測されたトランスクリプトームデータを流すことにより、動的構造も含め、転写・制御に関する情報が抽出可能となる。

b) 縦軸研究（又は横軸研究）への貢献に繋がる研究成果

本イン・シリコパイプラインは、マクロファージトランスクリプトームデータに限らず、その他のトランスクリプトームデータやプロテオームデータについても対応可能である。

【当該研究に関して活用した研究データ・リソース】

データタイプ	使用技術	データ内容	サンプル	データ数
遺伝子発現	qRT-PCR	2,315 遺伝子のプライマーについてTHP-1のPMA刺激後9time pointのデータ 及びLPS刺激後の11time pointsのデータ	THP-1	2,315 遺伝子のプライマーについてTHP-1のPMA刺激後9time pointのデータ 及びLPS刺激後の11time pointsのデータ
遺伝子発現	Illumina Microarray (custom design)	HumanはIllumina のSentrix® Human-6 Expression BeadChip を使用し、mouse は、Mouse-6 Expression BeadChip を使用。THP-1及びヒト、マウスのマクロファージ についてPMA刺激及びLPS刺激後のwhole genomeの遺伝子発現のプロファイルを作成した。データはnormalizeされている。Normalization の方法は以下に述べる通り。0.01以下は0.01にセット、各遺伝子の値を全値の中央値で割る。中央値が10以下の場合、10以上の値は10で割る。10以下の値は使わない。	THP-1	47,293 プローブについて THP-1のPMA刺激後9 time pointのデータ及びLPS刺激後12 time pointのデータ
遺伝子発現 転写開始点	CAGE	mRNAの5' 端から20ベースを切り出したCAGEタグのデータ。データを蓄積したデータベース (CAGE basic データベース) の各テーブルのダンプデータを提供する。CAGEタグの配列、CAGEタグを作成したサンプルの情報 (tissue名、dev_stage、cell_line名など)、ゲノムマッピングされた位置、マッピングされた遺伝子名などが含まれる。	THP-1	主要組織やTHP-1などの19サンプルについてのCAGEタグ、59,193,600タグのデータ

【主な研究論文とその概要】

1. Fujita, A., Sato, J.R., Garay-Malpartida, H.M., Sogayar, M.C., Ferreira, C.E., Miyano, S. Modeling nonlinear gene regulatory networks from time series gene expression data. J. Bioinformatics and Computational Biology. In press.

(概要) 病気や細胞システムの解明には、実験データだけではなく、統計的な遺伝子ネットワークモデルが必要とされる。遺伝子ネットワーク推定の分野では、ほとんどのモデルにおいて線形構造が仮定されている。しかし、分子生物学においては線形性を妥当とする根拠はいまだ示されていない。この論文では、Hoeffding の D 相関係数を使い、遺伝子間の非線形関係の存在を証明し、非線形ベクトル自己回帰モデルの開発により非線形 Granger 因果律を特定する遺伝子ネットワーク推定手法を構築した。

【特許出願数の実績】

該当なし。